**Chapter1 Introduction**

According to estimates from the World Health Organization (WHO), non-communicable diseases are now responsible for the majority of global deaths, and cancer is expected to rank as the leading cause of death and the most important obstacle to increasing life expectancy in every country of the world in the 21st century. In addition, Cancer incidence and mortality are rapidly growing worldwide. Its etiology is involved many and various biological signal cascades and multifactorial aberrations of genetic, epigenetic and microenvironment. WHO represents that breast cancer is one of common types of all the cancer in 2018.

Breast cancer is one of the most frequently diagnosed malignancies and the leading cause of cancer death in women worldwide [1]. According to the previous studies, breast cancer is an extremely heterogeneous disease resulting from complex combination of genetic, epigenetic and microenvironment factors. Among all the breast cancers, triple-negative breast cancer (TNBC) is a more heterogeneous subtype of breast cancer immunohistochemically defined by lack of expression of estrogen receptor (*ER*), progesterone receptor (*PR*) and human epidermal growth factor receptor-2 (*HER2*) [2, 3]. TNBC not only constitutes approximately 10-15% of all breast cancer but also tends to be more common in young women than other subtypes of breast cancer [4]. Only 30–45% of TNBC patients can achieve a pathological complete response (pCR) and survival rates similar to other types of breast cancer [5]. In addition, patients with TNBC have the worst prognosis and mortality risk in five years than other subtypes of breast cancer and can’t benefit from hormone or trastuzumab-based therapy because of the loss of target receptors such as *ER*, *PR*, and *HER-2* [6]. Hence, surgery and chemotherapy are the only available modalities [7].

As the previous discussion, TNBC has a more biologically aggressive behavior, higher recurrence rate, higher frequency of metastases and worse survival than other subtypes of breast cancers. That is why finding new biomarkers for TNBC becomes increasingly difficult to ignore. In order to gain an insight into the molecular mechanisms of TNBC and non-TNBC and find out the differences between TNBC and non-TNBC. The molecular mechanism of TNBC is extremely difficult to analyze, because the molecular mechanism of cancer involved many and various cascades of spatial and temporal changes in pathways always caused by gene, miRNA, long non-coding RNA (lncRNA), epigenetic modification and microenvironment factors.

To begin with, microRNAs (miRNAs) are short non-coding RNAs with sizes of 17-25 nucleotides that function in regulation of gene expression. In cancer, miRNAs play roles in almost all aspects of cancer biology such as proliferation, cell cycle, apoptosis, invasion, metastasis, and angiogenesis. Given that miRNAs have been observed to be seriously dysregulated in cancer but have not yet been comprehensively understood. It is expected that more miRNAs will be found as significant roles in the etiology and progression of cancer. As a matter of fact, that recent studies have made significant breakthroughs about miRNAs in cancers. For instance, the microRNA-200 (miR-200) plays a critical role in regulating epithelial–mesenchymal transition (EMT) and invasion through inhibition of *ZEB1* in breast cancer.

Then, LncRNAs are defined as non-protein-coding RNAs with size more than 200 nucleotides. The lncRNAs play a versatile role in various biological processes and disease states through interacting with DNA, RNA and other transcriptional factors, such as gene expression regulation, histone modification, gene silencing and DNA methylation [8]. Previous studies suggest that the mutation in lncRNA is highly associated with cancer [9, 10]. Any dysregulation in lncRNAs influences the normal cellular functions, including cell proliferation, apoptosis, angiogenesis and tumor progression [11]. For example, MALAT1 facilitates cell proliferation, tumor progression and metastasis of TNBC cells [12]. Obviously, finding potential lncRNAs as biomarkers for cancer diagnosis and prognosis will make great contribution to development of new therapeutic strategies.

Furthermore, it is becoming clear that microenvironment-mediated epigenetic modification plays a necessary role in cancer [13]. Epigenetic modification is the study of heritable and reversible change in gene expression that occurs without a change in DNA sequence [14]. It includes DNA methylation and histone modifications. DNA methylation is characterized by epigenetic DNA modifications which are sufficiently powerful to regulate the dynamics of gene expression [15]. Histone modifications play an important role in gene regulation and carcinogenesis which influence chromatin structure [16]. Several types of histone modifications are known, such as acetylation, methylation, phosphorylation, and ubiquitination are the most important emphasized in terms of the regulation of chromatin structure and transcriptional activity [16-20].

Finally, molecular microenvironment known as the stroma is made of the cells surrounding a tumor. The stroma can be influenced and can in turn influence the growth and formation of tumors and new metastases. For instance, the extracellular matrix (ECM) plays a critical role in the tumor microenvironment. In order to form a tumor, cancer cells must form attachments to the ECM and communicate with other cells, such as tumor-associated neutrophils, cancer-associated fibroblasts, and myeloid-derived suppressor cells. Then, the tumor cells can colonize the microenvironment and form a metastasis. The ECM is especially important in tumor formation and invasion as cells respond and adapt to the local microenvironment to progress to malignancy. This involves both deregulated proliferation of tumor cells and modification of the immediate environment to favor cell survival, angiogenesis, and spread of the tumor.

In conclusion, understanding the comprehensive genome-wide genetic and epigenetic regulation is the most significant for molecular mechanism. In order to get a comprehensive understanding of TNBC and investigate the differences between TNBC and non-TNBC, we constructed the stochastic models of the genome-wide genetic epigenetic network (GWGEN) composed of protein-protein interaction networks (PPINs) and gene regulatory networks (GRNs) through data mining and systems biology method. By applying a system identification method and a system order detection scheme to prune the false positives from the candidate GWGENs, including protein-protein interaction and regulations among proteins, receptors, transcription factors (TFs), miRNAs and lncRNAs, we identified the real GWGENs of TNBC and non-TNBC. Although the insignificant edges of networks have been pruned out from candidate GWGENs through system identification method and a system order detection scheme, the real GWGENs of TNBC and non-TNBC are still too complex and large to find the core pathways in them. Then, we applied the principal network projection (PNP) method to the real GWGENs to extract the core network from the energy of connections of real GWGEN perspective. Finally, we totally compared two core GWGENs to further extract two core pathways of TNBC and non-TNBC, and then investigate the major differences between TNBC and non-TNBC to systematize the cellular mechanisms of TNBC and non-TNBC and propose network biomarkers, such as proteins and miRNAs, as potential diagnostic biomarker or new therapeutic drug targets.

**Chapter 2. Result**

To identify and investigate mechanism in TNBC and non-TNBC, we have to extract the core signaling pathways from GWGEN. However, the GWGENs of TNBC and non-TNBC are too complex to extract the core signaling pathways from GWGEN, it is necessary for us to prune the false-positives of candidate GWGENs by NGS data and system order detection method, Akaike Information Criterion (AIC). The flowchart in Figure1 has shown how to construct the candidate GWGENs, real GWGENs, core GWGENs of TNBC and non-TNBC to extract the core pathways from core GWGENs in TNBC and non-TNBC. First, we constructed candidate GWGEN made of candidate PPIN and candidate GRN by big data mining. After using the system order detection method and principle network projection method (PNP) to identify the real GWGENs and core GWGENs, we applied the network visualization software Cytoscape to visualize the real GWGENs and core GWGENs of TNBC and non-TNBC shown in Figure 2~Figure 3, respectively. However, the real GWGENs of TNBC and non-TNBC are still too complicated to investigate the molecular mechanism of TNBC and non-TNBC; therefore, we selected the top 2000 node according to the significant projection energy on the 85% significant network structure of real GWGENs of TNBC and non-TNBC as shown in Figure 4? and Figure 5, respectively. In addition, according to the candidate GWGENs and real GWGENs, the numbers of nodes of proteins, TFs, miRNAs, lncRNAs and edges of interaction and regulation were counted out as shown in Table 1. To further investigate the difference of mechanism between TNBC and non-TNBC, we extracted the core signaling pathways from core GWGENs and compared the core signaling pathways in TNBC and non-TNBC as shown in [Figure~ Figure]. Finally, according to the core signal pathways in TNBC and non-TNBC, we investigate the mechanism of TNBC and non-TNBC by comparing their interaction and regulation of ligand, core membrane proteins to core TFs and their target gene in microenvironment. We considered not only the signaling pathway but also some effects and factors in microenvironment including cytokines, chemokines, angiogenic factors etc., for the core signaling pathways are incomplete. The effects and factors in microenvironment play a crucial role in breast tumor progression according to previous studies.

**Chapter 3. Discussion**

**Chapter 4. Conclusion**

In this study, we focused on constructing the core signaling pathways involved in the core GWGENs to further investigate cellular mechanisms of TNBC and non-TNBC in [Figure 1], based on protein-protein interactions, transcriptional regulations, miRNA/lncRNA regulations and epigenetic regulation in GWGENs. To begin with, we proposed a approach to constructing the candidate GWGEN by big database mining. Then, we pruned the false-positive interactions and regulations through a system identification method and a system order detection scheme to identify the real GWGEN. We also used PNP to extract core GWGENs for TNBC and no-TNBC.

**Chapter 5. Method and Materials**

**5.1 Overview of the construction for core genome-wide genetic and epigenetic networks (GWGENs) of TNBC and non-TNBC.**

For the purpose of understanding more comprehensive molecular mechanism of TNBC and finding the difference between TNBC and non-TNBC, we totally compared two core genome-wide genetic and epigenetic networks (GWGENs) to further extract two core pathways of TNBC and non-TNBC from the perspective of systems biology, and then found the difference core pathways between TNBC and non-TNBC. The flowchart of constructing candidate GWGENs, real GWGENs, core GWGENs and core pathways of TNBC and non-TNBC is shown in [Figure 1]. According to the flowchart, we constructed the candidate GWGENs, real GWGENs and core GWGENs of TNBC and non-TNBC through big data mining and system model identification. The procedure of constructing core GWGENs can be divided into following four steps:

(1) Using big data mining to construct candidate GWGENs combined of candidate protein-protein interaction networks (PPIN) and candidate gene/miRNA/lncRNA regulatory networks (GRNs) by some interaction and regulatory databases;

(2) Identify real GWGENs of TNBC and non-TNBC via system identification scheme, system order detection method and their microarray data to prune false positives in candidate GWGENs;

(3) Apply the principal network projection (PNP) method to extract the core elements such as core protein, genes, miRNAs, TFs, receptors, and lncRNAs from real GWGENs to construct core GWGENs.

(4) By comparing the core GWGENs of TNBC and non-TNBC, we extracted the core pathways of TNBC and non-TNBC to gain insight into the genetic and epigenetic mechanisms of carcinogenic development of TNBC and non-TNBC in microenvironment and find the difference core pathways between TNBC and non-TNBC.

**5.2** **Big data mining and preprocessing of microarray data for constructing candidate GWGENs.**

In the study, we totally choose three microarray gene expression datasets of breast cancer with accession number GSE41998, GSE32646 and GSE25066 obtained from the NCBI gene expression omnibus (GEO) [21-23]. Their corresponding platforms are GPL571, GPL570 and GPL96, respectively. These datasets contain the mRNAs, miRNAs and lncRNA microarray data. The samples of datasets which we selected are classified as TNBC and non-TNBC according to definition of TNBC. In general, TNBC is described triple-negative breast cancer that has worse prognosis, while no-TNBC described the other subtypes of breast cancer. We established candidate GWGEN for both of TNBC and non-TNBC in this study. We obtained 284 and 544 samples for TNBC and non-TNBC, respectively. To avoid overfitting in network construction, the maximum degree of the proteins in the PPIN and GRN should be less than the cancer sample number. In addition, we considered that the candidate GWGEN in human cells composed of the candidate PPIs and gene, miRNA and lncRNA regulations. According to the candidate PPIs from the Biological General Repository for Interaction Datasets (BioGRID) database [24], the gene regulations from the Integrated Transcription Factor Platform (ITFP) database , the Human Transcriptional Regulation Interactions (HTRI) database [25], and the TRANScription FACtor database (TRANSFAC), the miRNA regulations from TargetScanHuman database [26], and the lncRNA from CircuitDB2. Because the candidate GWGEN contains all possible interaction and regulations obtained from experimental and computational results, which contains many false positives, we need to construct the stochastic interaction and regulatory models of the candidate GWGEN to identify the real GWGEN to characterize the molecular mechanisms in microenvironment. The real GWGENs of TNBC and non-TNBC can be obtained by pruning false positives in the candidate GWGEN through a system identification method and a system order detection scheme in the stochastic models.

**5.3 Constructing the Stochastic Models of the GWGEN to identify real GWGEN of TNBC and non-TNBC.**

In order to identify the real GWGEN to characterize the molecular mechanisms in microenvironment, we constructed the stochastic interaction and regulatory models of the candidate GWGEN, including protein-protein interactions, transcriptional regulations, miRNA/lncRNA regulations, basal level and stochastic noises due to model instability and data noise.

For the protein interactive model of the candidate protein-protein interaction networks (PPINs) in candidate GWGENs, the protein-protein interaction of the i-th protein in cell for sample n as given by the following equations:



Where  denotes the interaction ability between the *i-*th protein and the *j-*th interactive protein;  represents the expression level of the *i-*th protein for the sample n;  represents the expression level of the *j-*th protein for the sample n;  indicates the total number of proteins interacting with the *i-*th protein and  is the total number of proteins in candidate PPIN; N denotes the total number of data samples;  represents the basal level of *i-*th protein expression;  is the stochastic noise of the *i-*th protein for the sample n due to model instability and data noise.

For the transcription factor regulatory model of candidate GRN in candidate GWGENs, the transcriptional regulation *x-*th gene of breast cells in sample n can be described by the following equations:

 (2) 

Where  represents the expression level of the *x-*th gene;  indicates the total number of TFs binding to the x-th gene;  represents the total number of lncRNAs binding to the x-th gene;  denotes the total number of miRNAs inhibiting the x-th gene;  denotes the transcription regulatory ability from the *u-*th TF to the *x-*th gene;  is the transcription regulatory ability from the *k-*th lncRNA to the *x-*th gene;  represents the post-transcription regulatory ability which the *v-*th miRNA inhibit the *x-*th gene; , , and  indicate the expression of the *u-*th TF, the *k-*th lncRNA, and *v-*th miRNA, respectively.  is the total number of genes and  denotes the total number of data samples;  represents the basal level of the *x-*th gene expression;  is the stochastic noise of gene expression in the *x-*th gene for the sample n owing to model uncertainty and data noise.

In addition, the regulation of TFs, lncRNA and miRNAs has a great influence on the expression level of lncRNA. The candidate lncRNA regulatory network (LRN) model of candidate GENs by the regulatory equations as follow:

 (3)



Where  represents the expression level of the *y-*th lncRNA;  indicates the total number of TFs binding to the *y-*th lncRNA;  represents the total number of lncRNAs binding to the *y-*th lncRNA;  denotes the total number of miRNAs inhibiting the *y-*th lncRNA;  denotes the transcription regulatory ability from the *u-*th TF to the *y-*th lncRNA;  is the transcription regulatory ability from the *k-*th lncRNA to the *y-*th lncRNA;  represents the post-transcription regulatory ability which the *v-*th miRNA inhibit the *y-*th lncRNA; , , and  indicate the expression of the *u-*th TF, the *k-*th lncRNA, and *v-*th miRNA, respectively.  is the total number of lncRNAs and  denotes the total number of data samples;  represents the basal level of the *y-*th lncRNA expression;  is the stochastic noise of gene expression in the *y-*th lncRNA for the sample n owing to model uncertainty and data noise.

On the top of that the expression of miRNA is also affected by the regulation of TFs, lncRNA and miRNAs, the candidate miRNA regulatory network (MRN) model of candidate GWGENs by the regulatory equations as follow:

 (4)

Where  represents the expression level of the *z-*th miRNA;  indicates the total number of TFs binding to the *z-*th miRNA; represents the total number of lncRNAs binding to the *z-*th miRNA; denotes the total number of miRNAs inhibiting the *z-*th miRNA; denotes the transcription regulatory ability from the *u-*th TF to the *z-*th miRNA; is the transcription regulatory ability from the *k-*th lncRNA to the *z-*th miRNA; represents the post-transcription regulatory ability which the *v-*th miRNA inhibit the *z-*th miRNA; , , and  indicate the expression of the *u-*th TF, the *k-*th lncRNA, and *v-*th miRNA, respectively.  is the total number of miRNAs and  denotes the total number of data samples; represents the basal level of the *z-*th miRNA expression;  is the stochastic noise of gene expression in the *z-*th miRNA for the sample n owing to model uncertainty and data noise.

**5.4 Parameter estimation of the models of candidate GWGENs through system identification scheme, system order detection method and microarray data.**

As mentioned previously, we constructed the candidate protein-protein interaction network (PPIN) model (1) and gene regulatory network (2), lncRNA regulatory network (3), miRNA regulatory network (4) models in candidate GRN to combine the candidate GWGEN. We applied system identification scheme, system order detection method and their microarray data to identify the protein-protein interactive parameters of protein-protein interaction network model and regulatory parameters of gene, miRNA and lncRNA regulatory network model to identify the real GWGENs of TNBC and non-TNBC. To identify these interactive and regulatory parameters, these equations (1), (2), (3), (4) can be described by the linear regression form as following:

 (5)



As described above, the linear regression equations (5), (6), (7), (8) can be easily described as the following regression forms, respectively:









Where , , ,  represent regression vectors of the expression data of protein, gene, miRNA, lncRNA for the sample n, respectively;  denotes the parameter vectors associated to protein-protein interaction abilities;,  and  denote the parameter vector related to gene, miRNA and lncRNA transcriptional regulations including transcriptional regulatory abilities, post-transcriptional regulatory abilities and basal levels, respectively; ,  and  represent the stochastic noise of gene expression in the *x-*th gene, *y-*th lncRNA and *z-*th miRNA for the sample n owing to model uncertainty and data noise.

In addition, there are the equations of parameter estimation for the all N sample, which are described as the following forms:

 





These equations of parameter estimation are easily shown as follows:









Moreover, we need to estimate the parameter vectors , ,  and  by NGS data, we could estimate the parameter vectors via least square estimation as follows:





subject to 



subject to 



subject to 

As mentioned above, the equations of constrained least square parameter estimation in (21), (22), (23), (24) could guarantee the estimated post-transcriptional regulatory abilities of miRNA on the regulatory genes, lncRNAs and miRNAs are always negative. According to the above, using constrained least square parameter estimation method could always obtain protein interactive estimation parameters  and gene, lncRNA and miRNA regulatory estimation parameters , and  by MATLAB optimization toolbox.

We could not only know , , , , , ,  and  by the gene, lncRNA and miRNA expression level data of TNBC and non-TNBC downloaded from NCBI website but also identify the interactive abilities, regulatory abilities and basal levels in corresponding estimated parameters , ,  and  thorough the constrained least square parameter estimation problem in (21), (22), (23), (24). As a matter of fact, there are still a lot of false-positives from different database or some experimental data. That is why, we apply system order detection method to prune false-positives protein interactive abilities, transcriptional regulatory abilities and post-transcriptional regulatory abilities from candidate GWGEN to obtain real GWGEN.

There is one of system order detection scheme, Akaike Information Criterion (AIC) could detect the system order (the number of regulations) of real GRN through the identification system scheme. The system order detection criteria AICs including the i-th protein of protein interaction, the x-th gene, the y-th lncRNA and the z-th miRNA are respectively shown as follows:



were 

 and  represent the estimated residual error and number (order) of parameters of the *i-*th protein in the estimation parameter problem (21) of the PPIN, respectively;  denotes the estimated parameters of the *i-*th protein by (21).



were 

 and  represent the estimated residual error and the number of parameters of the *x-*th gene in the estimation parameter problem (22) of the GRN, respectively;  is the estimated parameters of the *x-*th gene by (22) in the regulatory model.



were 

 and  represent the estimated residual error and the number of parameters of the *y-*th lncRNA in the estimation parameter problem (23) of the GRN, respectively;  is the estimated parameters of the *y-*th lncRNA by (23) in the regulatory model.



were 

 and  denote the estimated residual error and the number of parameters of the *z-*th miRNA in the estimation parameter problem (24) of the GRN, respectively;  is the estimated parameters of the *z-*th miRNA by (24) in the regulatory model.

According to system identification theory of system modeling [ ], the real system order of system model could minimize AIC. In order to obtain real system order of GWGENs, we need to minimize AICs (i.e.,,  and ) of system order detection method to obtain the real number (order) of system model, i.e.  for the *i-*th protein,  for the *x-*th gene,  for the *y-*th lncRNA and  for the *z-*th miRNA.

As a consequence, we could remove the insignificant interactions and regulations out of true number identified by AIC as false-positives of interactions and regulations in the candidate GWGENs to extracted real GWGENs of TNBC and non-TNBC. In the end, we extracted the real GWGENs of TNBC and non-TNBC through NGS data as shown in [Figure~Figure].

**5.5 Applying the principal network projection (PNP) method to extract core GWGENs in the real GWGENs.**

After the system identification, the real GWGENs are still extremely difficult to investigate and compare the difference of genetic and epigenetic mechanisms between TNBC and non-TNBC. For the reason that we extract the corresponding core GWGENs of TNBC and non-TNBC in the real GWGENs of TNBC and non-TNBC through Using the principal network projection (PNP) method on GRNs. In addition, it is necessary that we integrated a system network matrix  of real GWGEN before we apply the PNP method. The system network matrix  which involves the whole estimated system parameters in the real GWGENs is described as follows:



where the sub-network matrix  represents the system matrix associated with interactive abilities of proteins; the sub-network matrix ,  and  represent the corresponding system matrices associated with TFs transcriptional regulatory abilities, lncRNAs transcriptional regulatory abilities and miRNAs post-transcriptional regulatory abilities on genes, respectively; the sub-network matrix ,  and  denote the system matrices associated with TFs transcriptional regulatory abilities, lncRNAs transcriptional regulatory abilities and miRNAs post-transcriptional regulatory abilities on lncRNAs, respectively; the sub-network matrix ,  and  represent the system matrices associated with TFs transcriptional regulatory abilities, lncRNAs transcriptional regulatory abilities and miRNAs post-transcriptional regulatory abilities on miRNAs, respectively.

In the network matrix , the corresponding component is zero if a link does not appear in the candidate GWGEN or has been pruned by AIC.

The Principal network projection method (PNP) which applied to the system network matrix  to extract core GWGEN in real GWGEN of TNBC and non-TNBC is based on singular value decomposition in the following:



where  and  are the unitary matrix;

 represents diagonal matrix which composed of  singular values of  in descending order (i.e., ).denotes diagonal matrix of  and .



In addition, we define expression fraction  for the normalization of singular values as follows:



We choose the top  singular vectors of network matrix  which consist of 85% network matrix ,such that  with minimum  to find 85% of principal network structure from the energy perspective of the real GWGEN. the projection of network matrix to the top singular vectors of is given as follows:



where  is the *c-*th row vector of ;  is the *r-*th row vector of . After we defined the 2-norm projection value of each node of proteins, genes, miRNAs and

lncRNAs in the real GWGEN to the top right singular vectors in the following:



where  is the 2-norm projection value of the *c-*th node of core network to the 85% principal network structure of GWGEN. Based on the projection values.

For example, if *B*(*c*) is close to zero, the corresponding node is insignificant and nearly independent to the top 85% network structure of core network. Conversely, if the projection value of a corresponding node is higher, the node is important in the GEN principal network structure. The core GWGENs of TNBC and non-TNBC through PNP method are shown Figure respectively.

We projected core GWGEN to KEGG pathways to construct the corresponding core signaling pathways to investigate and compare the difference of genetic and epigenetic mechanisms between TNBC and non-TNBC. From two mechanisms of TNBC and non-TNBC, we could identify significant biomarkers as drug targets. Further, from the changes of the core protein basal level in TNBC and non-TNBC more than a threshold, we could predict the core protein whether it interacts with acetyltransferase proteins, deacetylase proteins, ubiquitin proteins and deubiquitinase proteins directly, while the interacting epigenetic enzyme expression level is higher than the standard threshold or not. Finally, if the core gene basal level change between two consecutive stages is higher than a threshold, then it may be mainly due to DNA methylation expression change.

**Chapter 2. Results**

**2.1 Investigation of core signaling pathways from identified GWGEN in TNBC and non-TNBC.**

To identify and investigate mechanism in TNBC and non-TNBC, we have to extract the core signaling pathways from GWGEN. However, the GWGENs of TNNC and non-TNBC are too complex to extract the core signaling pathways from GWGEN, it is necessary for us to prune the false-positives of candidate GWGENs by NGS data and system order detection method, Akaike Information Criterion (AIC). The flowchart in [Figure1] has shown how to construct the candidate GWGENs, real GWGENs, core GWGENs of TNBC and non-TNBC to extract the core pathways from core GWGENs in TNBC and non-TNBC. First, we constructed candidate GWGEN made of candidate PPIN and candidate GRN by big data mining. After using the system order detection method and principle network projection method (PNP) to identify the real GWGENs and core GWGENs, we applied the network visualizing software Cytoscape to visualize the real GWGENs and core GWGENs of TNBC and non-TNBC shown in [Figure 2~Figure 3], respectively. However, the real GWGENs of TNBC and non-TNBC are still too complicated to investigate the molecular mechanism of TNBC and non-TNBC, we selected the top 2000 node according the significant projection energy on the 85% significant network structure of real GWGENs of TNBC and non-TNBC as shown in [Figure and Figure ], respectively. In addition, according to the candidate GWGENs and real GWGENs, the numbers of nodes of proteins, TFs, miRNAs, lncRNAs and edges of interaction and regulation were counted out as shown in [Table 1]. To further investigate the difference of mechanism between TNBC and non-TNBC, we extracted the core signaling pathways from core GWGENs and compared the core signaling pathways in TNBC and non-TNBC as shown in [Figure ~ Figure]. Finally, according to the core signal pathway in TNBC and non-TNBC, we investigate the mechanism of TNBC and non-TNBC by comparing their interaction and regulation of ligand, core membrane proteins to core TFs and their target gene in microenvironment. We considered not only the signaling pathway but also some effects and factors in microenvironment including of cytokines, chemokines, angiogenic factors etc., for the core signaling pathways are incomplete. The effects and factors in microenvironment play an import role around tumor on the development of breast cancer according to previous studies.

**2.2 Core pathway from core GWGEN in triple-negative breast cancer.**

Focusing on the core pathways in triple-negative breast cancer as shown in Figure. *S100A7*, a member of the S100 family of proteins, has been indicated one of initiation factors for tumor cell proliferation and metastasis in triple-negative breast cancer cell by our results. After *S100A7* translocated to intracellular, the TF *AR* has been induced by *S100A7*. TF *AR* overexpression upregulated *MMP9* [27]. The activation of *MMP9* has the ability to promote cell proliferation and metastasis in *TNBC* [28]. In addition, *S100A7* is also received by receptor *RAGE*. Once *RAGE* binging to *S100A7* enhanced *RAGE* expression, *RAGE* activates TFs *STAT3, NF-kB, P53* and *c-Myc* through multiple protein-protein interactions as shown in Figure[29-31]. The TF *STAT3,* a member of the STAT protein family, not only upregulates *MMP9* and *ZEB1* but also indirectly activates *PARP1*. First and foremost, *MMP9* upregulated by phosphorylation of *STAT3* promotes cell proliferation and metastasis [32]. Besides, TF *ZEB1* upregulated by *BCL6* which not only identified an import role in TNBC but also activated by *STAT3* is involved in promoting EMT [33-35]. Last but not least, *PARP1* automatically promotes DNA repair as soon as overexpression of *STAT3* increase DNA damage accumulation [36, 37]. The TF *NF-kB* also activates *MMP9* promoting cell proliferation. *BCL-xL* (encoded by the *BCL2-like 1* gene) activated by *NF-kB* plays a role of anti-apoptosis factor [38, 39]. In addition, *CD44*, a cell surface glycoprotein, upregulated by *NF-kB* has ability to promote cell proliferation and metastasis [40, 41]. The TF *p53* plays a role of nuclear transcription factors having ability to promote apoptosis in cancer cell [42]. Not only *AIFm2* targeted by *p53* induces apoptosis but also some studies suggest tumor suppressive effect of *AIFm2* [43, 44]. In addition, *p53* keeps patients from tumor cell proliferation, angiogenesis and metastasis via inhibiting *PAI-1* [45, 46]. In particular, *p53* suppress one of the most frequently altered pathways in human cancer, *PI3K/AKT* pathway, through increasing expression levels of *PTEN* to inhibit *PI3K* [47, 48]. The TF *c-Myc* increase expression levels of *FOXF2* to inhibit cell proliferation, EMT and metastasis [49].

**2.3 Core pathway from core GWGEN in non-triple-negative breast cancer.**

**Reference:**

1. Torre, L.A., et al., *Global cancer statistics, 2012.* CA Cancer J Clin, 2015. **65**(2): p. 87-108.

2. Cleator, S., W. Heller, and R.C. Coombes, *Triple-negative breast cancer: therapeutic options.* Lancet Oncol, 2007. **8**(3): p. 235-44.

3. Anders, C. and L.A. Carey, *Understanding and treating triple-negative breast cancer.* Oncology (Williston Park), 2008. **22**(11): p. 1233-9; discussion 1239-40, 1243.

4. Rakha, E.A., et al., *Prognostic markers in triple-negative breast cancer.* Cancer, 2007. **109**(1): p. 25-32.

5. Kalimutho, M., et al., *Targeted Therapies for Triple-Negative Breast Cancer: Combating a Stubborn Disease.* Trends Pharmacol Sci, 2015. **36**(12): p. 822-846.

6. Goncalves, H., Jr., et al., *Survival Study of Triple-Negative and Non-Triple-Negative Breast Cancer in a Brazilian Cohort.* Clin Med Insights Oncol, 2018. **12**: p. 1179554918790563.

7. Wahba, H.A. and H.A. El-Hadaad, *Current approaches in treatment of triple-negative breast cancer.* Cancer Biol Med, 2015. **12**(2): p. 106-16.

8. Birney, E., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.

9. Zhao, M., et al., *MALAT1: A long non-coding RNA highly associated with human cancers.* Oncol Lett, 2018. **16**(1): p. 19-26.

10. Sanchez Calle, A., et al., *Emerging roles of long non-coding RNA in cancer.* Cancer Sci, 2018. **109**(7): p. 2093-2100.

11. Gutschner, T. and S. Diederichs, *The hallmarks of cancer: a long non-coding RNA point of view.* RNA Biol, 2012. **9**(6): p. 703-19.

12. Jadaliha, M., et al., *Functional and prognostic significance of long non-coding RNA MALAT1 as a metastasis driver in ER negative lymph node negative breast cancer.* Oncotarget, 2016. **7**(26): p. 40418-40436.

13. Kelly, T.K., D.D. De Carvalho, and P.A. Jones, *Epigenetic modifications as therapeutic targets.* Nat Biotechnol, 2010. **28**(10): p. 1069-78.

14. Rodenhiser, D. and M. Mann, *Epigenetics and human disease: translating basic biology into clinical applications.* Cmaj, 2006. **174**(3): p. 341-8.

15. Moore, L.D., T. Le, and G. Fan, *DNA methylation and its basic function.* Neuropsychopharmacology, 2013. **38**(1): p. 23-38.

16. Bannister, A.J. and T. Kouzarides, *Regulation of chromatin by histone modifications.* Cell Res, 2011. **21**(3): p. 381-95.

17. !!! INVALID CITATION !!! {}.

18. Sawicka, A. and C. Seiser, *Histone H3 phosphorylation - a versatile chromatin modification for different occasions.* Biochimie, 2012. **94**(11): p. 2193-201.

19. Rossetto, D., N. Avvakumov, and J. Cote, *Histone phosphorylation: a chromatin modification involved in diverse nuclear events.* Epigenetics, 2012. **7**(10): p. 1098-108.

20. Swygert, S.G. and C.L. Peterson, *Chromatin dynamics: interplay between remodeling enzymes and histone modifications.* Biochim Biophys Acta, 2014. **1839**(8): p. 728-36.

21. Horak, C.E., et al., *Biomarker analysis of neoadjuvant doxorubicin/cyclophosphamide followed by ixabepilone or Paclitaxel in early-stage breast cancer.* Clin Cancer Res, 2013. **19**(6): p. 1587-95.

22. Miyake, T., et al., *GSTP1 expression predicts poor pathological complete response to neoadjuvant chemotherapy in ER-negative breast cancer.* Cancer Sci, 2012. **103**(5): p. 913-20.

23. Itoh, M., et al., *Estrogen receptor (ER) mRNA expression and molecular subtype distribution in ER-negative/progesterone receptor-positive breast cancers.* Breast Cancer Res Treat, 2014. **143**(2): p. 403-9.

24. Chatr-Aryamontri, A., et al., *The BioGRID interaction database: 2015 update.* Nucleic Acids Res, 2015. **43**(Database issue): p. D470-8.

25. Bovolenta, L.A., M.L. Acencio, and N. Lemke, *HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions.* BMC Genomics, 2012. **13**: p. 405.

26. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.

27. Zhang, B.G., et al., *Androgen receptor promotes gastric cancer cell migration and invasion via AKT-phosphorylation dependent upregulation of matrix metalloproteinase 9.* Oncotarget, 2014. **5**(21): p. 10584-95.

28. Mehner, C., et al., *Tumor cell-produced matrix metalloproteinase 9 (MMP-9) drives malignant progression and metastasis of basal-like triple negative breast cancer.* Oncotarget, 2014. **5**(9): p. 2736-49.

29. Nasser, M.W., et al., *S100A7 enhances mammary tumorigenesis through upregulation of inflammatory pathways.* Cancer Res, 2012. **72**(3): p. 604-15.

30. Nasser, M.W., et al., *RAGE mediates S100A7-induced breast cancer growth and metastasis by modulating the tumor microenvironment.* Cancer Res, 2015. **75**(6): p. 974-85.

31. Kwak, T., et al., *Targeting of RAGE-ligand signaling impairs breast cancer cell invasion and metastasis.* Oncogene, 2016. **36**: p. 1559.

32. Zhang, F., et al., *Elevated STAT3 Signaling-Mediated Upregulation of MMP-2/9 Confers Enhanced Invasion Ability in Multidrug-Resistant Breast Cancer Cells.* Int J Mol Sci, 2015. **16**(10): p. 24772-90.

33. Walker, S.R., et al., *The transcriptional modulator BCL6 as a molecular target for breast cancer therapy.* Oncogene, 2015. **34**(9): p. 1073-82.

34. Wu, Q., et al., *B-cell lymphoma 6 protein stimulates oncogenicity of human breast cancer cells.* BMC Cancer, 2014. **14**: p. 418.

35. Yu, J.M., et al., *BCL6 induces EMT by promoting the ZEB1-mediated transcription repression of E-cadherin in breast cancer cells.* Cancer Lett, 2015. **365**(2): p. 190-200.

36. Ko, H.L. and E.C. Ren, *Functional Aspects of PARP1 in DNA Repair and Transcription.* Biomolecules, 2012. **2**(4): p. 524-48.

37. Barry, S.P., et al., *STAT3 modulates the DNA damage response pathway.* Int J Exp Pathol, 2010. **91**(6): p. 506-14.

38. Lin, A. and M. Karin, *NF-kappaB in cancer: a marked target.* Semin Cancer Biol, 2003. **13**(2): p. 107-14.

39. Chen, C., L.C. Edelstein, and C. Gelinas, *The Rel/NF-kappaB family directly activates expression of the apoptosis inhibitor Bcl-x(L).* Mol Cell Biol, 2000. **20**(8): p. 2687-95.

40. Smith, S.M., Y.L. Lyu, and L. Cai, *NF-kappaB affects proliferation and invasiveness of breast cancer cells by regulating CD44 expression.* PLoS One, 2014. **9**(9): p. e106966.

41. Herrlich, P., et al., *CD44 acts both as a growth- and invasiveness-promoting molecule and as a tumor-suppressing cofactor.* Ann N Y Acad Sci, 2000. **910**: p. 106-18; discussion 118-20.

42. Yoshida, K. and Y. Miki, *The cell death machinery governed by the p53 tumor suppressor in response to DNA damage.* Cancer Sci, 2010. **101**(4): p. 831-5.

43. Wu, M., et al., *AMID is a p53-inducible gene downregulated in tumors.* Oncogene, 2004. **23**(40): p. 6815-9.

44. Lu, J., et al., *Activation of AIFM2 enhances apoptosis of human lung cancer cells undergoing toxicological stress.* Toxicol Lett, 2016. **258**: p. 227-236.

45. Kortlever, R.M., P.J. Higgins, and R. Bernards, *Plasminogen activator inhibitor-1 is a critical downstream target of p53 in the induction of replicative senescence.* Nat Cell Biol, 2006. **8**(8): p. 877-84.

46. Shetty, S., et al., *Regulation of plasminogen activator inhibitor-1 expression by tumor suppressor protein p53.* J Biol Chem, 2008. **283**(28): p. 19570-80.

47. Lien, E.C., C.C. Dibble, and A. Toker, *PI3K signaling in cancer: beyond AKT.* Curr Opin Cell Biol, 2017. **45**: p. 62-71.

48. Stambolic, V., et al., *Regulation of PTEN transcription by p53.* Mol Cell, 2001. **8**(2): p. 317-25.

49. Zhang, J., et al., *FOXF2 inhibits proliferation, migration, and invasion of Hela cells by regulating Wnt signaling pathway.* Biosci Rep, 2018. **38**(5).