Single-cell RNA-seq dissects the intratumoral heterogeneity of triple-negative breast cancer based on gene regulatory networks

Shunheng Zhou,[1] Yu-e Huang,[1] Haizhou Liu,[1] Xu Zhou,[1] Mengqin Yuan,[1] Fei Hou,[1] Lihong Wang,[2] and Wei Jiang[1]

[1]Department of Biomedical Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; [2]Department of Pathophysiology, School of Medicine, Southeast University, Nanjing 210009, China

Triple-negative breast cancer (TNBC) is a subtype of breast cancer with high intratumoral heterogeneity. Recent studies revealed that TNBC patients might comprise cells with distinct molecular subtypes. In addition, gene regulatory networks (GRNs) constructed based on single-cell RNA sequencing (scRNA-seq) data have demonstrated the significance for decoding the key regulators. We performed a comprehensive analysis of the GRNs for the intrinsic subtypes of TNBC patients using scRNA-seq. The copy number variations (CNVs) were inferred from scRNA-seq data and identified 545 malignant cells. The subtypes of the malignant cells were assigned based on the PAM50 model. The cell-cell communication analysis revealed that the macrophage plays a dominant role in the tumor microenvironment. Next, the GRN for each subtype was constructed through integrating gene co-expression and enrichment of transcription-binding motifs. Then, we identified the critical genes based on the centrality metrics of genes. Importantly, the critical gene *ETV6* was ubiquitously upregulated in all subtypes, but it exerted diverse roles in each subtype through regulating different target genes. In conclusion, the construction of GRNs based on scRNA-seq data could help us to dissect the intratumoral heterogeneity and identify the critical genes of TNBC.

INTRODUCTION

Breast cancer, which is always accompanied with high intratumoral heterogeneity, is one of the most commonly diagnosed carcinomas in women.[1,2] Traditionally, breast cancer patients could be classified into luminal A (LumA), luminal B (LumB), human epidermal growth factor receptor-2+ (HER2+), and a basal-like subtype, according to the immunohistochemical character of the ER (estrogen receptor), PR (progesterone receptor), and HER2 status.[3] The pathological characteristics and prognostic outcomes of the four subtypes showed great differences.[4] Most basal-like breast cancer exhibits the triple-negative phenotype.[5] Triple-negative breast cancer (TNBC) is always with higher metastatic probability and poorer prognostic than other subtypes.[6] Accumulative evidence showed that the intratumoral heterogeneity of breast cancer was the leading cause of drug resistance for tumor therapy.[7] Recently, Yeo and Guan[8] found that a tumor might be comprised of multiple breast cancer subtypes. What's more, Kim et al.[2] revealed that the

subtype composition of the TNBC patients would fluctuate under chemotherapy pressure. The intratumoral heterogeneity of breast cancer should be investigated for the precision medicine.

With the development of single-cell RNA sequencing (scRNA-seq) technology, we can gain insight into the genetic heterogeneity at the single-cell resolution for cancer research.[9] For instance, Patel et al.[10] adopted scRNA-seq to investigate genetic heterogeneity of glioblastoma and revealed the diversity of cell state and functional characteristics of tumor cells. The malignant cells of breast cancer were demonstrated, originating from epithelial cells. In addition, the malignant cells could interact with immune cells and stroma cells, which constituted the complex tumor microenvironment.[11] Karaayvaz et al.[12] revealed that the TNBC was comprised of a distinct subpopulation of epithelial cells. The signatures of one epithelial cell subpopulation could be associated with long-term outcomes of TNBC patients. Hence, traditional analysis of breast cancer based on bulk RNA-seq would be quite insufficient. Recently, Iacono et al.[13] applied a global regulatory model to construct the gene regulatory networks (GRNs) based on scRNA-seq data, which could help us to recognize the critical regulators in disease. Thus, network analysis of TNBC based on scRNA-seq data could help us to dissect the risk genes that play critical roles in tumorigenesis.

In this study, we proposed a comprehensive analysis of the GRNs for the breast cancer subtypes and dissected the critical genes for each molecular subtype. First, the malignant cells were identified according to the inferred copy number variation (CNV) of epithelial cells. Next, the subtypes for malignant cells were classified by using the PAM50 model. The ligand-receptor interaction analysis revealed the essential role of the macrophage in the tumor microenvironment. Then, the transcription factor (TF)-gene pairs that co-expressed and enriched transcription-
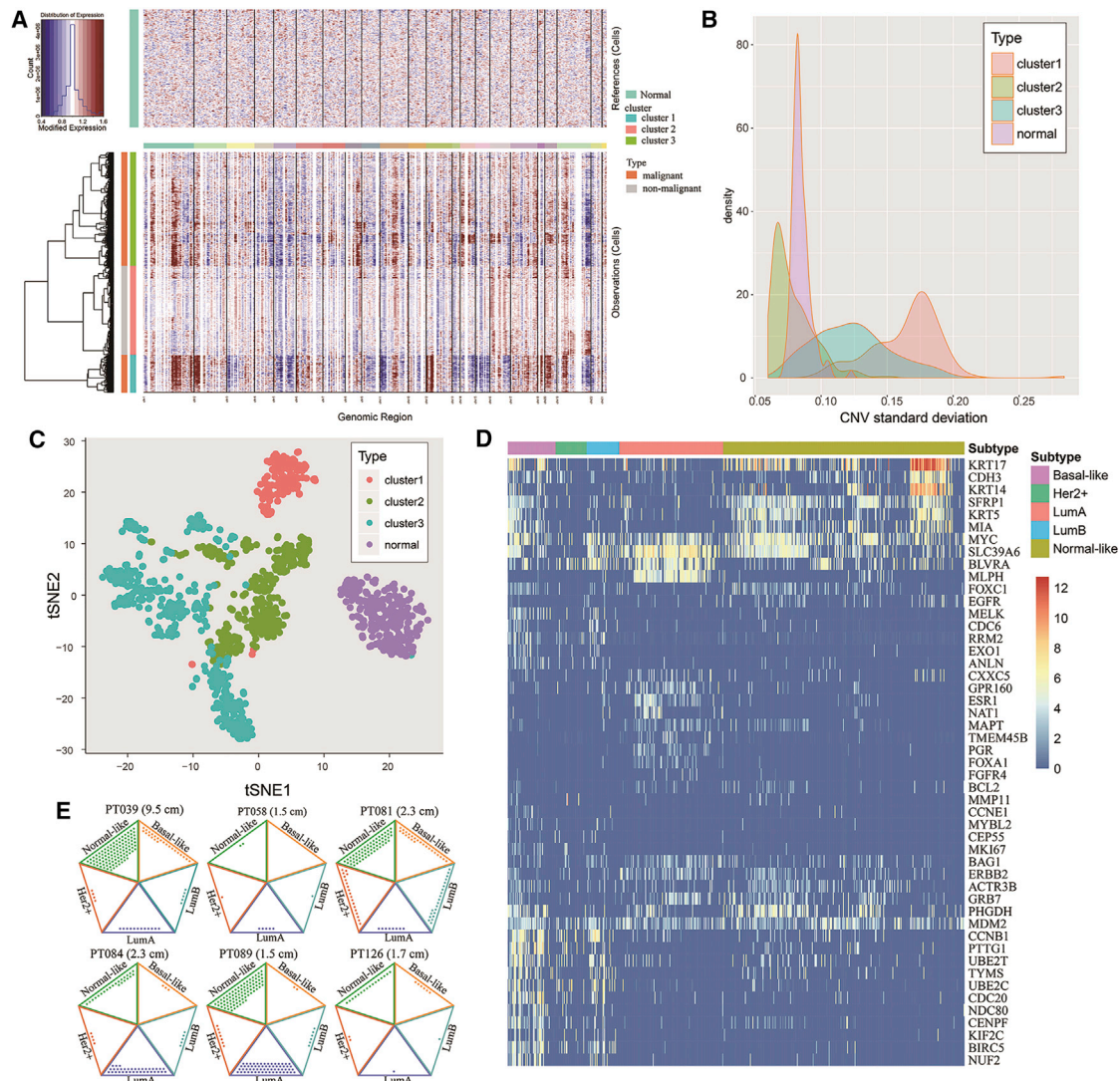
Figure 1. Cell composition of TNBC patients

(A) Heatmap of the inferred CNV across 868 epithelial cells, in which genes were sorted by genomic location. (B) Standard deviation of the CNV of the cells in each cluster. Cluster 1 and cluster 3 showed a relatively high variable of CNV. (C) tSNE plot of the epithelial cells using the inferred CNV. (D) Heatmap of the gene-expression profiles of 545 malignant epithelial cells. (E) Breast cancer subtype composition of each TNBC patient based on the PAM50 model. The point in the petal represents a single cell. The number beside the patient identification (ID) is the tumor size.

binding motifs were retained for GRN construction. Furthermore, critical regulators for each subtype were pinpointed by using the centrality metrics. At last, we found that *ETV6* was ubiquitously activated in the five molecular subtypes of breast cancer. The dysregulated *ETV6* could regulate diverse genes in each subtype, which could be associated with intratumoral heterogeneity and progression of TNBC.

RESULTS

Intrinsic molecular subtypes of TNBC patients

We obtained 868 epithelial cells and inferred their CNVs using the gene-expression profiles and genomic information. The epithelial cells were clustered into 3 distinct subgroups using the inferred CNVs. We

observed that cluster 1 and cluster 3 had obviously a high variable of CNVs. Thus, we speculated that the two clusters might be malignant cells, which comprised of 134 and 411 cells, respectively. The cells in cluster 2 had a lower variable of CNVs, which might be non-malignant cells and consisted of 323 cells (Figures 1A and 1B). Next, we employed tSNE (t-Distributed Stochastic Neighbor Embedding) to perform dimension reduction. The result showed that the epithelial cells were well partitioned into four distinct subpopulations (Figure 1C), which confirmed the results of hierarchy clustering.

To identify the molecular subtype of each malignant epithelial cell, the PAM50 model was employed to classify the malignant epithelial
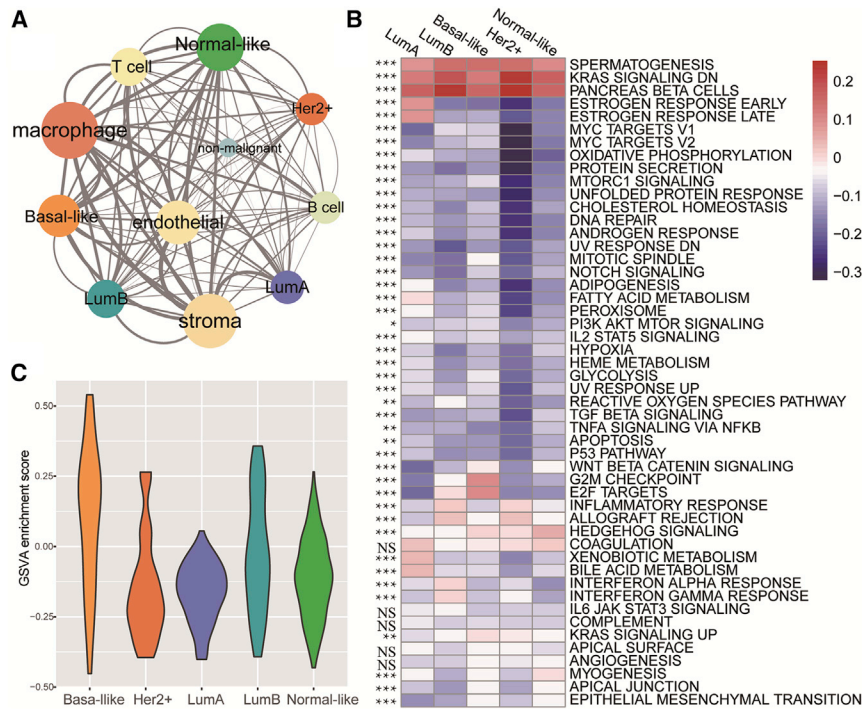
Figure 2. Cell-cell interaction and hallmark gene set activity analysis

(A) Cell-cell interaction network of different cell types. The node size represents the number of interactions. The width of edge represents the number of significant ligand-receptor interactions in two cell types. (B) Differences of the enrichment of the hallmark gene sets across the five molecular subtypes. The colors are encoded by the mean values of the GSVA enrichment scores in the molecular subtypes (one-way ANOVA, ***p < 0.001; **p < 0.01; *p < 0.05; NS, p > 0.05). (C) Violin plots of GSVA enrichment scores of the G2/M checkpoint for five molecular subtypes.

cells to five molecular subtypes, named as normal-like, basal-like, Her2+, LumA, and LumB (Figure 1D; Table S1). We found that the subtype composition of six patients showed great diversity (Figure 1E). PT039 and PT081 were comprised of a relatively high fraction of basal-like subtype cells. Through inspecting the clinical information of 6 patients, we found that PT039 and PT081 showed the relative larger tumor size (Figure 1E), which demonstrated that the high composition of the basal-like subtype could be associated with progression of breast cancer patients.

To investigate the cell-cell interaction network in the tumor microenvironment of TNBC, the python CellPhoneDB package was used to analyze the ligand-receptor interaction in different cell types (Table S2). The ligand-receptor pair with p < 0.05 was considered as significantly interacted in the two cell types. We found that the macrophage played a dominant role in the cell-cell communication network (Figure 2A), which revealed the important role of the macrophage in the tumor microenvironment.[14] For example, the epidermal growth factor receptor (EGFR)-amphiregulin (AREG) ligand-receptor pair was significantly interacted in the basal-like subtype and macrophage (p < 0.05). The modulation of autocrine signaling AREG was demonstrated to contribute to recruitment of the macrophage in the basal-like breast cancer cell line.[15] We then performed a gene set variation analysis (GSVA) for the five molecular subtypes to identify the cells' activity in 50 hallmark gene sets (Table S3). We found that the five molecular subtypes showed great diversity (Figure 2B), which revealed the intratumoral heterogeneity of different molecular subtypes. For example, the basal-like subpopulation was significantly enriched

in the G2/M checkpoint (Figure 2C; one-way ANOVA, p-value = 3.6×10^{-24}). A previous study demonstrated that regulating the G2/M checkpoint could improve the efficacy of cancer treatment.[16]

Molecular subtype-specific GRNs

The normal epithelial cells and five molecular subtype-specific GRNs were constructed based on gene expression profiles and TF-binding motifs information. First, we calculated the significantly co-expressed TF-gene pairs (Table 1). Second, the target gene significantly enriched of a TF-binding motif was retained for further study. Finally, the enrichment score (ES) >1 was considered as significant TF-target pairs for GRN construction (Tables S4 and 1). The constructed GRNs were shown in Figures 3A–3F. We found that the degree distributions of GRNs of both the normal epithelial cells and the five molecular subtypes approximately fitted the power-law distribution (Figures S1A–S1F), which indicated that the 6 networks could exist highly connected hub nodes. The word cloud map illustrated the degree of the TFs in subtype-specific GRNs (Figures S2A–S2F). We found that the degree of TFs shows great divergence in each subtype-specific GRN. Furthermore, we found that there were no regulations (TF-target pairs) conserved in five molecular subtype-specific GRNs (Figure S3A), but 10 regulators (TFs) were conserved in the five GRNs (Figure S3B). We proposed that these conserved TFs could exert diverse roles in various molecular subtypes, which could be associated with intratumoral heterogeneity.

We then identified the critical genes in the 6 GRNs. First, we employed the centrality metrics to assess the importance of the gene. The centrality metrics included degree, PageRank, betweenness, eigenvalue, and closeness (described in Materials and methods). Then, the Q statistic was employed to integrate the five centrality metrics. The centrality metrics and Q value for each node were presented in Table S5. The top 1% genes ranked by the integrated Q statistic in each network were considered as critical genes. Next, to evaluate whether the critical genes were enriched in cancer genes, we obtained the known cancer genes from the Cancer Gene Census (https://cancer.sanger.ac.uk/census).[17] We found that the critical genes in

Table 1. Number of candidate TF-target regulations in six GRNs

	Normal-like	Her2+	LumA	LumB	Basal-like	Normal epithelial
Co-expressed	352,255	197,997	311,071	277,920	350,142	176,041
TF-binding motif	7,212	2,098	9,838	4,550	7,718	113,678
ES > 1	5,421	1,467	7,491	3,133	5,329	2,944

each molecular subtype-specific GRN were significantly enriched in the known cancer genes (hypergeometric test p-value: normal-like, 2.55×10^{-4}; basal-like, 3.27×10^{-4}; Her2+, 8.97×10^{-5}; LumA, 4.12×10^{-4}; LumB, 7.94×10^{-5}). The critical genes involved in the known cancer genes in each subtype showed great diversity (Figure S4). For example, we found that the oncogenic gene *MYC* was a critical gene in basal-like and LumB subtype-specific networks with a high node degree (basal-like: degree = 323; LumB: degree = 116). The elevated expression of *MYC* was demonstrated to be associated with the poor prognosis of breast cancer patients.[18] Besides, the oncogenic gene *ELK*4 was another critical gene in the subtype-specific network. Mesquita et al.[19] found the copy number gain of *ELK*4 was associated with tumorigenesis. The diversity of critical nodes in each subtype might elucidate the intratumoral heterogeneity in the network perspective.

Diverse function of critical genes

To investigate the functions of the critical genes in the five molecular subtypes, we focused on the conserved critical genes in these molecular subtypes. We found that *ETV6* was ubiquitously activated in five molecular subtype-specific GRNs rather than normal epithelial cell GRN (Figures 3B–3F). We speculated that *ETV6* could play important roles in the five molecular subtypes. Then we focused on the functional diversity of *ETV6*. We found that *ETV6* regulated diverse genes in different subtypes. There were several subtype-specific genes in each subtype, which could be associated with intratumoral heterogeneity of breast cancer (Figure 4A). Besides, *ETV6* showed the highest degree in the basal-like subtype-specific GRNs (degree = 411), which could be critical for the malignant character of the basal-like subtype. Next, we employed Gene Ontology (GO) annotation scenario to assess the functions of genes regulated by *ETV6* in each subtype. Here, we considered the hallmark-associated GO terms for visualization.[20] We found that the genes regulated by *ETV6* participated in diverse roles in different subtypes, which showed a high level of intratumoral heterogeneity (Figure 4B). For example, the basal-like subtype was specifically annotated in the vasculogenesis biological process, which was owing to the specific regulation of *ETV6* to *SGPL*1 in the basal-like subtype. Engel et al.[21] found that the expression of *SGPL*1 in breast cancer might be a prognostic marker and served as a potential drug target for breast cancer treatment.

Then, we investigated the expression level of *ETV6* in different subtypes. At the single cell resolution, we found that *ETV6* was significantly upregulated in five molecular subtypes than the normal epithelial group (Figure 4C). Furthermore, we inspected the expression of *ETV6* in The Cancer Genome Atlas (TCGA) basal-like patients and normal breast samples. We found that *ETV6* was significantly upregulated in basal-like patients compared with normal control (Figure 4D). In addition, Osmanbeyoglu et al.[22] demonstrated that the strong nuclear *ETV6* staining could be associated with poor prognosis of uterine serous tumor. We suspected that the activation of *ETV6* could be associated with the survival of TNBC patients. Based on the METABRIC dataset, we found that the patients with higher expression of *ETV6* were associated wshangdddith a relatively poorer clinical outcome (Figure 4E). Thus, the activation of *ETV6* in the five molecular subtypes might cause intratumoral heterogeneity through regulating diverse genes. Thus, the dysregulated *ETV6* might be a potential target for tumor therapy.

DISCUSSION

TNBC is a subtype of breast cancer in women accompanied with high intratumoral heterogeneity. Network analysis leveraging the scRNA-seq data could help us to dissect the critical genes in tumorigenesis. In this study, we concentrated on the GRNs of molecular subtypes for TNBC patients. We proposed a comprehensive analysis of the GRNs in five molecular subtypes of TNBC patients using scRNA-seq data. With the integration of the gene-expression profiles and TF-binding motif information, we constructed the GRNs of the five molecular subtypes and normal epithelial cells. In addition, the critical genes in each subtype were identified in the constructed GRNs. The results showed that *ETV6* was specifically activated in the five molecular subtypes rather than in the normal epithelial cells, which indicated the crucial regulatory roles in the five molecular subtypes. Functional analysis further indicated that *ETV6* might exert various functions through regulating diverse target genes in each subtype. In addition, the survival analysis suggested that the overexpression of *ETV6* could be associated with poorer prognostic in TNBC patients. Thus, we speculated that the critical genes might play diverse roles in different subtypes, and the dysregulated critical genes might serve as potential drug targets for chemotherapy.

Recently, several studies investigated the tumor microenvironment of breast cancer.[23,24] The tumor microenvironment could be involved in paracrine interactions with malignant cells, leading to evading the immune surveillance and tumor progression. In this study, we focused on the malignant epithelial cells for network construction and revealed the essential roles of the critical genes. Although we did not take the impact of the immune cells and stromal cells into consideration, the results could still reveal the critical genes that play key roles in TNBC patients. The combination of the impacts of the tumor microenvironment, such as immune infiltration, might be meaningful
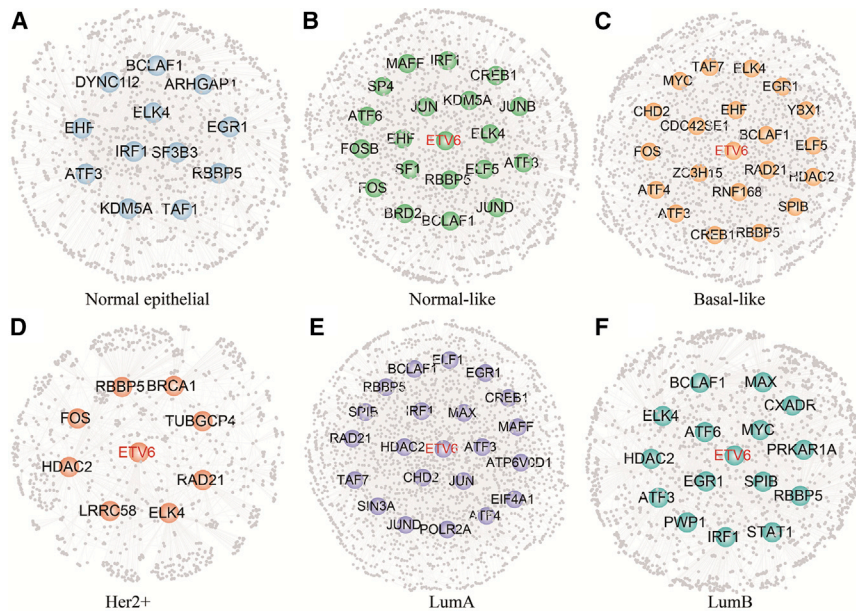
Figure 3. GRNs of normal epithelial cells and five molecular subtypes

Colored nodes imply the critical genes. (A–F) The GRNs of normal epithelial (A), normal-like (B), basal-like (C), Her2+ (D), LumA (E), and LumB (F).

genes for the corresponding GRNs. Undoubtedly, further external independent data validation could make our results more reliable. The dysregulation of *ETV6* has been found to be associated with several cancer types, such as leukemia and prostate cancer,[33,34] but the functional diversity and the prognostic value of *ETV6* in TNBC have not been investigated. In this study, we found that the expression of *ETV6* was elevated in the basal-like subtype based on both scRNA-seq and RNA-seq data (Figures 4C and 4D). Moreover, the TNBC patients with higher expression of *ETV6* showed a worse outcome (Figure 4E), which indicated that *ETV6* might play an oncogenic role in TNBC. In previous studies, the TFs, such as p53, and the non-coding RNAs (ncRNAs), such as microRNA (miR)-21, miR-10b, and *NEAT*1, could regulate different target genes in different subpopulations or subtypes of cancers.[35–37] In this study, we found that *ETV6* could play diversified roles through regulating different genes in the five molecular subtypes, which might be associated with intratumoral heterogeneity. To explore the potential drugs related to the *ETV6* gene, we used the ChEMBL database to scrutinize the candidate compounds.[38] As a result, we supposed that ceritinib might be a potential drug for breast cancer treatment. In this study, we proposed that the dysregulation of *ETV6* in TNBC patients might be a potential drug target for breast cancer treatment. Further experimental verification is also necessary to make our results more reliable.

As well, in this study, we employed the pySCENIC package for network construction. In addition, the AnimalTFDB and hTFtarget databases also curated the comprehensive TFs and TF-target regulations.[25,26] Construction of GRNs through integrating more information would be more reliable. In addition, the cells expressing ligand-receptor genes implied that cell-cell communications across the numerous cell types could be fundamental for the tumor microenvironment.[27] The cell-cell interaction quantified by gene-expression profiles in the tumor could be used to study the ligand-receptor interaction inside the tumor.[28] Further study focused on the cell-cell interaction network, which could be valuable for elucidating the crosstalk among different cell types. Besides, the GRN constructed for a single cell is another strategy for dissecting the intratumoral heterogeneity. Dai et al.[29] developed a pipeline for cell-specific network construction at single-cell resolution, which would be used for accurate cell clustering and pseudo-trajectory inference. Although our study focused on the GRNs in the holistic view, we could dissect the intratumoral heterogeneity in the view of an intrinsic subtype. Moreover, we could also reveal the critical genes that play important roles in the molecular subtypes of TNBC.

As we all know, the "zero dropouts" of scRNA-seq data could be inevitable for the small fraction of mRNA in single cells. Several data imputation methods were developed[30,31] to impute the zero value. In this study, the detected genes of the Smart-seq2 platform are sufficient for network construction. What's more, the co-expression-based scenario used in this study was verified to overcome this constraint.[32] Besides, there might be a patient's specific factors, but the correlation-based method for network construction could overcome these confounding factors. In this study, we only picked up the top 1% rank genes as critical genes in each GRN. Albeit this scenario would be arbitrary, but it enables us to explore the functional

MATERIALS AND METHODS
Data collection and processing
We downloaded the scRNA-seq data of 6 TNBC patients (PT039, PT058, PT081, PT084, PT089, and PT126) from GEO (https://www.ncbi.nlm.nih.gov/geo/) database (GEO: GSE118390), which comprised 1,189 high-quality cells with annotated cell types.[12] As described in a previous study, malignant cells of breast cancer were originated from epithelial cells.[11] Thus, only the 868 epithelial cells were retained for identification of malignant cells. In addition, the gene-expression profiles of 240 normal breast epithelial cells were downloaded from previous research,[39] which served as reference (normal) cells for copy number estimation and normal breast epithelial cell GRN construction. In addition, the gene-expression profiles of 105 normal and 115 TNBC samples were obtained from TCGA database[40] for differential gene-expression analysis. The METABRIC database[41] collected 186 TNBC patients with clinical information, which was used for survival analysis.
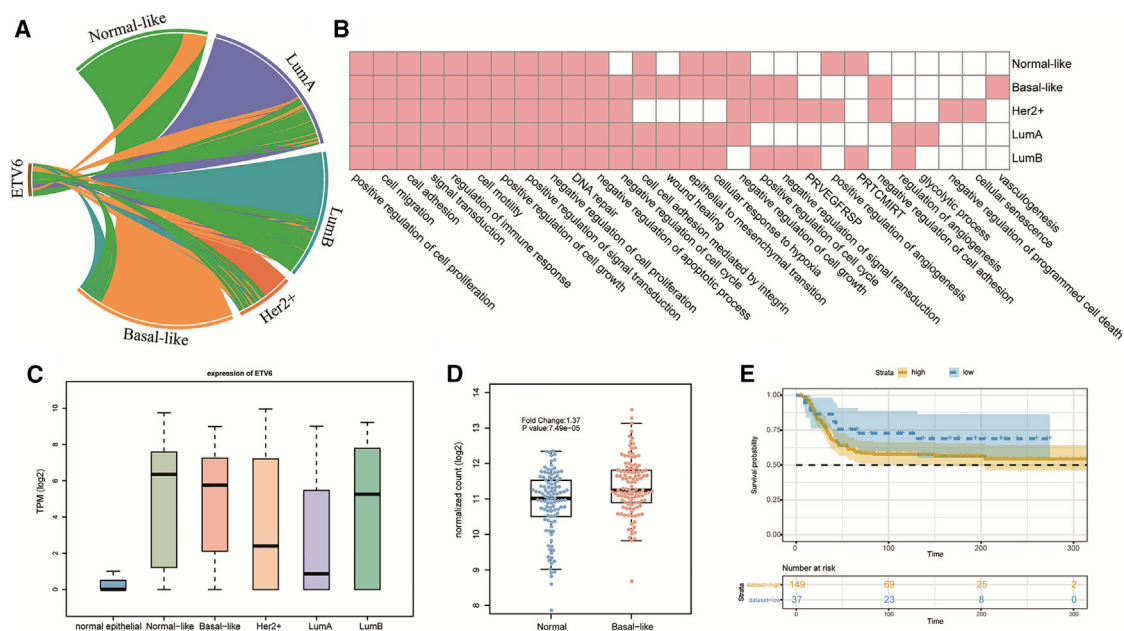
Figure 4. Functional diversity of *ETV*6

(A) *ETV*6 regulates diverse genes in each subtype. The line colors indicate the *ETV*6 regulations in different subtypes. (B) GO annotation of *ETV*6-regulated genes in each subtype. The colored lattice showed the *ETV*6-regulated genes annotated in the corresponding GO term. (C) Expression of *ETV*6 in normal epithelial cells and five molecular subtypes. (D) Expression of *ETV*6 in TCGA normal and TNBC samples. (E) Survival analysis of TNBC patients in the METABRIC dataset based on *ETV*6 expression. The lower expression of *ETV*6 showed better clinical outcome.

Identification of malignant epithelial cells and breast cancer subtypes

As mentioned in a previous study,[42] malignant cells always accompany a high variable of CNV. Here, we inferred the CNV for the 868 epithelial cells from the gene-expression profiles. The 240 normal epithelial cells were set as a reference set for CNV inference. The genes were sorted by the genomic location, and then the average gene expression of upstream and downstream genes was considered as putative CNV. In this study, the infercnv R package was used for CNV estimation.[43] After inferring the CNV from gene-expression profiles, hierarchy clustering was performed for identification of malignant cell clusters. The clusters with a high variable of CNV were considered as malignant cells for subtype identification.

The malignant cells were then classified into five molecular subtypes (normal-like, basal-like, Her2+, LumA, and LumB), according to their gene-expression profiles by using the PAM50 model,[2] which was widely used for identification of breast cancer subtypes based on 50 marker genes (PAM50 gene signatures).[44] The gene-expression value was transcripts per million (TPM) value. We then performed log2(TPM + 1) transformation for the gene-expression profiles. The subtype for each breast cancer cell was assigned by using the Sub-Pred_pam50 function of the genefu R package.[45]

Cell-cell communication and GSVA analysis

To investigate the cell-cell communication in different cell types, including the five molecular subtypes, first, the cell-type annotations were obtained from a previous study, including B cell, T cell, macrophage, stroma, and endothelial cell.[12] Then the python CellPhoneDB (https://www.cellphonedb.org/) package was used to estimate the significance of ligand-receptor pairs in different cell types.[46] The ligand-receptor pairs with p < 0.05 were considered as significantly interacted in two cell types.

To test the differentially activated gene sets in the five molecular subtypes, we first downloaded the 50 hallmark gene sets from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb). Then, the GSVA analysis for the five molecular subtypes was performed using the R GSVA package.[47]

Construction of GRNs and identification of critical nodes

We used the gene co-expression and transcription-binding motif information to build the GRN for each subtype of breast cancer and the normal epithelial cells. First, the TFs were obtained from a previous study, which contains 1,797 unique TFs.[32] Next, the co-expression analysis was performed to estimate the potential regulation strength between TFs and target genes.[48] Then, the TF-target gene pairs, which were significantly enriched with TF-binding motifs, were retained as real regulations in each subtype and normal epithelial cells. The co-expression and TF-binding motif enrichment analyses were performed by using the python pySCENIC package.[32]

After we obtained the GRNs of normal epithelial cells and five molecular subtypes, the node centrality metrics were employed
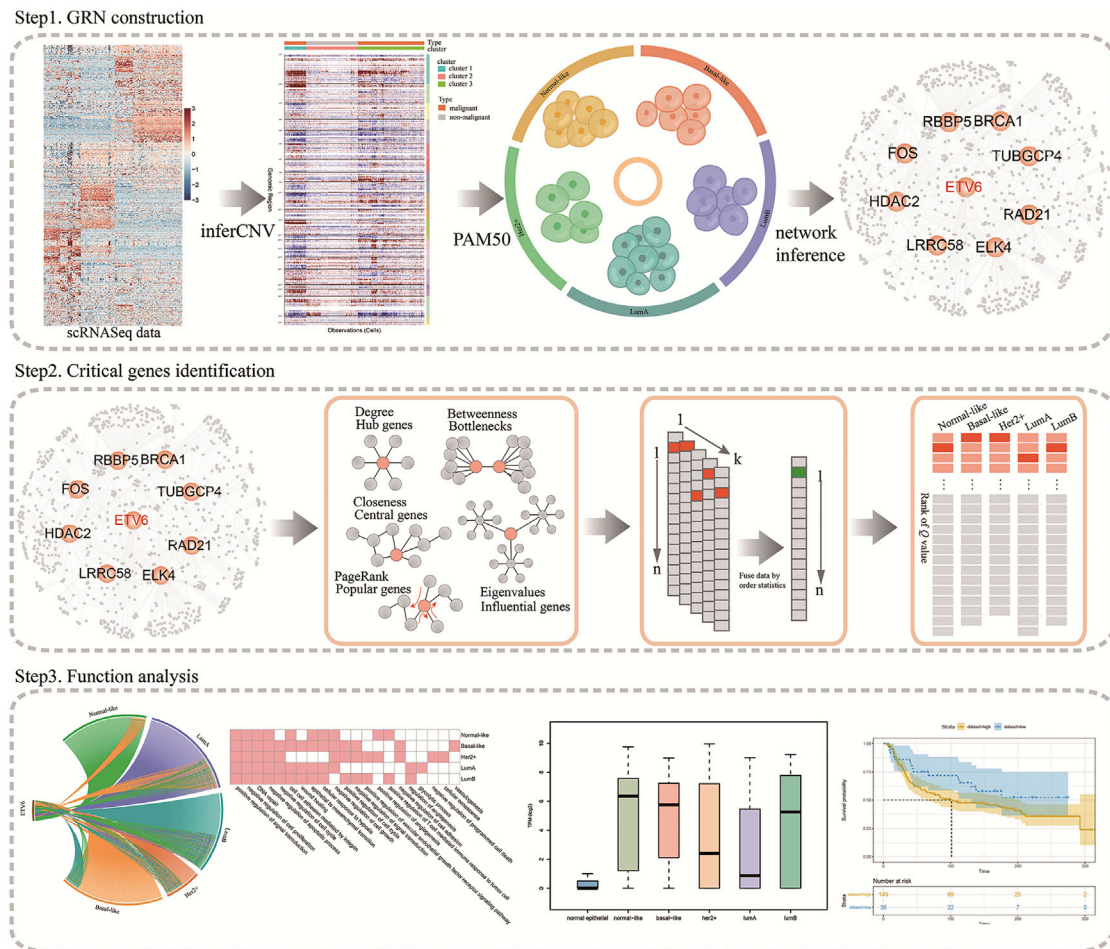
Figure 5. Schematic diagram of this study

First, the scRNA-seq data were used to infer CNV. The malignant cells were identified based on the inferred CNV. Then the subtype of each cell was assigned by using the PAM50 model, and the GRN was constructed for each subtype. Second, five centrality metrics were calculated to measure the importance of nodes. Then the Q statistic was used to integrate these centrality metrics. Finally, the diverse roles of the common critical genes were assessed. Differential expression analysis and survival analysis were performed for the critical gene.

to measure the importance of nodes in the constructed networks. In this study, we used degree, betweenness, eigenvalue, PageRank, and closeness to evaluate the centrality of nodes.[13] Degree is the number of adjacent nodes of the corresponding node. Node with high degree is usually considered a hub with essential functions. Betweenness is calculated based on the number of shortest paths passed through the corresponding node. Node with high betweenness could be the bottleneck of the GRN. The eigenvalue measures the node importance by taking the importance of neighbors into consideration. PageRank is the probability of the random walk of the corresponding node. PageRank is similar to eigenvalue, whereas PageRank introduces the damping factor (default 0.85). Closeness measures the average distance to all nodes for the corresponding node. Node with high closeness denotes that the node is located in the central location of network. All

of these centrality metrics were calculated by using the R igraph package.

After calculating the five centrality metrics for each node of each GRN, we employed the Q statistic[49] to integrate the five centrality metrics of the nodes for each GRN.

$$Q(r_1, r_2, \ldots r_i, \ldots, r_N) = N! V_N$$

$$V_k = \sum_{j=1}^{k} (-1)^{j-1} \frac{V_{k-j}}{j!} r_{N-k+1}^{j}$$

where $V_0 = 1$, $N = 5$, r_i is the rank ratio of the ith centrality metric, and r^j is r to the power of j in this study. Finally, we obtained the integrated Q statistic for each node in the corresponding GRN. The genes with

the top 1% Q statistic were considered as critical nodes in each GRN, respectively.

Function enrichment analysis and survival analysis

To obtain the function of TFs in the five molecular subtype-specific GRNs, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) online tool to perform GO annotations.[50] In addition, we downloaded the cancer hallmark-associated GO terms from previous research.[20] Only these cancer hallmark-associated GO terms were retained for visualization. In this study, we also investigated the expression of *ETV6* in the five molecular subtypes and the clinical outcome of TNBC patients. The R survival and survminer packages were used to evaluate the gene expression and prognostic significance. The flowchart of this study was shown in Figure 5.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.omtn.2020.12.018.

AUTHOR CONTRIBUTIONS

Conceptualization, methodology, and writing, S.Z.; data curation and formal analysis, Y.H.; data processing, H.L.; investigation, X.Z.; visualization, M.Y.; validation, F.H.; conceptualization, reviewing, editing, and supervision, L.W. and W.J.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J. Clin. *68*, 394–424.

2. Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., and Navin, N.E. (2018). Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell *173*, 879–893.e13.

3. Harbeck, N., and Gnant, M. (2017). Breast cancer. Lancet *389*, 1134–1150.

4. Heimes, A.S., Madjar, K., Edlund, K., Battista, M.J., Almstedt, K., Elger, T., Krajnak, S., Rahnenführer, J., Brenner, W., Hasenburg, A., et al. (2017). Subtype-specific prognostic impact of different immune signatures in node-negative breast cancer. Breast Cancer Res. Treat. *165*, 293–300.

5. Lachapelle, J., and Foulkes, W.D. (2011). Triple-negative and basal-like breast cancer: implications for oncologists. Curr. Oncol. *18*, 161–164.

6. Waks, A.G., and Winer, E.P. (2019). Breast Cancer Treatment: A Review. JAMA *321*, 288–300.

7. McGranahan, N., and Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell *168*, 613–628.

8. Yeo, S.K., and Guan, J.L. (2017). Breast Cancer: Multiple Subtypes within a Tumor? Trends Cancer *3*, 753–760.

9. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat. Methods *6*, 377–382.

10. Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science *344*, 1396–1401.

11. McCuaig, R., Wu, F., Dunn, J., Rao, S., and Dahlstrom, J.E. (2017). The biological and clinical significance of stromal-epithelial interactions in breast cancer. Pathology *49*, 133–140.

12. Karaayvaz, M., Cristea, S., Gillespie, S.M., Patel, A.P., Mylvaganam, R., Luo, C.C., Specht, M.C., Bernstein, B.E., Michor, F., and Ellisen, L.W. (2018). Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. Nat. Commun. *9*, 3588.

13. Iacono, G., Massoni-Badosa, R., and Heyn, H. (2019). Single-cell transcriptomics unveils gene regulatory network plasticity. Genome Biol. *20*, 110.

14. Ngambenjawong, C., Gustafson, H.H., and Pun, S.H. (2017). Progress in tumor-associated macrophage (TAM)-targeted therapeutics. Adv. Drug Deliv. Rev. *114*, 206–221.

15. Nickerson, N.K., Mill, C.P., Wu, H.J., Riese, D.J., 2nd, and Foley, J. (2013). Autocrine-derived epidermal growth factor receptor ligands contribute to recruitment of tumor-associated macrophage and growth of basal breast cancer cells in vivo. Oncol. Res. *20*, 303–317.

16. Choi, C., Cho, W.K., Park, S., Shin, S.W., Park, W., Kim, H., and Choi, D.H. (2020). Checkpoint Kinase 1 (CHK1) Inhibition Enhances the Sensitivity of Triple-Negative Breast Cancer Cells to Proton Irradiation via Rad51 Downregulation. Int. J. Mol. Sci. *21*, 2691.

17. Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183.

18. Horiuchi, D., Kusdra, L., Huskey, N.E., Chandriani, S., Lenburg, M.E., Gonzalez-Angulo, A.M., Creasman, K.J., Bazarov, A.V., Smyth, J.W., Davis, S.E., et al. (2012). MYC pathway activation in triple-negative breast cancer is synthetic lethal with CDK inhibition. J. Exp. Med. *209*, 679–696.

19. Mesquita, B., Lopes, P., Rodrigues, A., Pereira, D., Afonso, M., Leal, C., Henrique, R., Lind, G.E., Jerónimo, C., Lothe, R.A., and Teixeira, M.R. (2013). Frequent copy number gains at 1q21 and 1q32 are associated with overexpression of the ETS transcription factors ETV3 and ELF3 in breast cancer irrespective of molecular subtypes. Breast Cancer Res. Treat. *138*, 37–45.

20. Plaisier, C.L., Pan, M., and Baliga, N.S. (2012). A miRNA-regulatory network explains how dysregulated miRNAs perturb oncogenic processes across diverse cancers. Genome Res. *22*, 2302–2314.

21. Engel, N., Adamus, A., Frank, M., Kraft, K., Kühn, J., Müller, P., Nebe, B., Kasten, A., and Seitz, G. (2018). First evidence of SGPL1 expression in the cell membrane silencing the extracellular S1P siren in mammary epithelial cells. PLoS ONE *13*, e0196854.

22. Osmanbeyoglu, H.U., Shimizu, F., Rynne-Vidal, A., Alonso-Curbelo, D., Chen, H.A., Wen, H.Y., Yeung, T.L., Jelinic, P., Razavi, P., Lowe, S.W., et al. (2019). Chromatin-informed inference of transcriptional programs in gynecologic and basal breast cancers. Nat. Commun. *10*, 4369.

23. Yin, J., Li, Z., Yan, C., Fang, E., Wang, T., Zhou, H., Luo, W., Zhou, Q., Zhang, J., Hu, J., et al. (2018). Comprehensive analysis of immune evasion in breast cancer by single-cell RNA-seq. bioRxiv, 368605.

24. Azizi, E., Carr, A.J., Plitas, G., Cornish, A.E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kiseliovas, V., Setty, M., et al. (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell *174*, 1293–1308.e36.

25. Hu, H., Miao, Y.R., Jia, L.H., Yu, Q.Y., Zhang, Q., and Guo, A.Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. *47* (D1), D33–D38.

26. Zhang, Q., Liu, W., Zhang, H.M., Xie, G.Y., Miao, Y.R., Xia, M., and Guo, A.Y. (2020). hTFtarget: A Comprehensive Database for Regulations of Human Transcription Factors and Their Targets. Genomics Proteomics Bioinformatics *18*, 120–128.

27. Ramilowski, J.A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V.P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., and Forrest, A.R. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. Nat. Commun. 6, 7866.

28. Kumar, M.P., Du, J., Lagoudas, G., Jiao, Y., Sawyer, A., Drummond, D.C., Lauffenburger, D.A., and Raue, A. (2018). Analysis of Single-Cell RNA-Seq Identifies Cell-Cell Communication Associated with Tumor Characteristics. Cell Rep. 25, 1458–1468.e4.

29. Dai, H., Li, L., Zeng, T., and Chen, L. (2019). Cell-specific network constructed by single-cell RNA sequencing data. Nucleic Acids Res. 47, e62.

30. Li, W.V., and Li, J.J. (2018). An accurate and robust imputation method scImpute for single-cell RNA-seq data. Nat. Commun. 9, 997.

31. van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell 174, 716–729.e27.

32. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods 14, 1083–1086.

33. Tsai, Y.C., Zeng, T., Abou-Kheir, W., Yeh, H.L., Yin, J.J., Lee, Y.C., Chen, W.Y., and Liu, Y.N. (2018). Disruption of ETV6 leads to TWIST1-dependent progression and resistance to epidermal growth factor receptor tyrosine kinase inhibitors in prostate cancer. Mol. Cancer 17, 42.

34. Vijayakrishnan, J., Qian, M., Studd, J.B., Yang, W., Kinnersley, B., Law, P.J., Broderick, P., Raetz, E.A., Allan, J., Pui, C.H., et al. (2019). Identification of four novel associations for B-cell acute lymphoblastic leukaemia risk. Nat. Commun. 10, 5348.

35. Park, S.R., Namkoong, S., Friesen, L., Cho, C.S., Zhang, Z.Z.Z., Chen, Y.C., Yoon, E., Kim, C.H., Kwak, H., Kang, H.M., and Lee, J.H. (2020). Single-Cell Transcriptome Analysis of Colon Cancer Cell Response to 5-Fluorouracil-Induced DNA Damage. Cell Rep. 32, 108077.

36. Zhou, S., Wang, L., Yang, Q., Liu, H., Meng, Q., Jiang, L., Wang, S., and Jiang, W. (2018). Systematical analysis of lncRNA-mRNA competing endogenous RNA network in breast cancer subtypes. Breast Cancer Res. Treat. 169, 267–275.

37. Ochoa, S., de Anda-Jáuregui, G., and Hernández-Lemus, E. (2020). Multi-Omic Regulation of the PAM50 Gene Signature in Breast Cancer Molecular Subtypes. Front. Oncol. 10, 845.

38. Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Félix, E., Magariños, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M., et al. (2019). ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 47 (D1), D930–D940.

39. Gao, R., Kim, C., Sei, E., Foukakis, T., Crosetto, N., Chan, L.K., Srinivasan, M., Zhang, H., Meric-Bernstam, F., and Navin, N. (2017). Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer. Nat. Commun. 8, 228.

40. Cancer Genome Atlas Research Network (2012). Comprehensive genomic characterization of squamous cell lung cancers. Nature 489, 519–525.

41. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al.; METABRIC Group (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346–352.

42. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science 352, 189–196.

43. Tickle, T., Tirosh, I., Georgescu, C., Brown, M., and Haas, B. (2019). inferCNV of the Trinity CTAT Project. (Klarman Cell Observatory, Broad Institute of MIT and Harvard), https://github.com/broadinstitute/inferCNV.

44. Chia, S.K., Bramwell, V.H., Tu, D., Shepherd, L.E., Jiang, S., Vickery, T., Mardis, E., Leung, S., Ung, K., Pritchard, K.I., et al. (2012). A 50-gene intrinsic subtype classifier for prognosis and prediction of benefit from adjuvant tamoxifen. Clin. Cancer Res. 18, 4465–4472.

45. Gendoo, D.M.A., Ratanasirigulchai, N., Schröder, M.S., Paré, L., Parker, J.S., Prat, A., and Haibe-Kains, B. (2016). Genefu: an R/Bioconductor package for computation of gene expression-based signatures in breast cancer. Bioinformatics 32, 1097–1099.

46. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat. Protoc. 15, 1484–1506.

47. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14, 7.

48. Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PLoS ONE 5, e12776.

49. Aerts, S., Lambrechts, D., Maity, S., Van Loo, P., Coessens, B., De Smet, F., Tranchevent, L.C., De Moor, B., Marynen, P., Hassan, B., et al. (2006). Gene prioritization through genomic data fusion. Nat. Biotechnol. 24, 537–544.

50. Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4, 44–57.