EPH- International Journal of Medical and Health Sciences
Volume 12 Issue 4 October 2026
ISSN- 2456 - 6063
DOI: https://doi.org/10.67772/b85k9v50

COMPARATIVE TRANSCRIPTOMIC PROFILING REVEALS CONSISTENT GENE EXPRESSION SIGNATURES ASSOCIATED WITH BREAST CANCER ACROSS INDEPENDENT COHORTS

Michael T. Ardizone[1]

[1]Department of Education, Sta. Catalina Integrated National High School, Majayjay, Laguna 4005 Philippines Orcid i'd: https://orcid.org/0009-0001-7151-6077, Email: michael.ardizone@deped.gov.ph

Received:- 19/07/2026	Revised:- 25/09/2026	Accepted:- 06/10/2026	Published:- 24/10/2026

Abstract
Breast cancer exhibits substantial molecular heterogeneity that can limit the reproducibility of transcriptomic biomarkers across independent patient cohorts. This study performed a cross-cohort transcriptomic analysis to identify robust gene-expression signatures associated with breast cancer. Three independent datasets, BC-TCGA, GSE2034, and GSE25066, comprising 1,368 samples were analyzed after gene-identifier harmonization, yielding 10,351 genes shared across all cohorts. Principal component analysis demonstrated strong transcriptional separation in BC-TCGA, moderate separation in GSE25066, and substantial overlap in GSE2034. Differential-expression analysis identified 12,856, 32, and 2,889 FDR-significant genes in BC-TCGA, GSE2034, and GSE25066, respectively. Applying an additional standardized effect-size threshold (|Hedges' g| $\geq$ 0.50) identified 374 genes supported by at least two cohorts, of which 164 exhibited concordant directional effects. Most concordant genes (163) were reproduced between BC-TCGA and GSE25066, whereas no gene fulfilled stringent reproducibility criteria across all three cohorts. Prominent candidates included SFRP1, NFIB, GATA3, CCND1, FOXC1, KRT18, and UGP2. These findings demonstrate substantial cohort-dependent transcriptional heterogeneity and highlight the importance of effect-size-aware cross-cohort validation for prioritizing robust breast cancer molecular signatures for subsequent functional and clinical investigation.

Keywords: Breast cancer, Cross-cohort analysis, Gene expression, Transcriptomics, Molecular signatures

Introduction

Breast cancer is a very heterogeneous disease with significant interclonal differences in genetic changes, patterns of gene expression, cell phenotypes, and treatments. There is also interpatient and intratumoral heterogeneity, which leads to the clinical variability in the progression and behaviour of disease (Guo et al., 2023). Tumors of the same type may have significant functional and transcriptomic differences (Turner et al., 2021). The plasticity of tumor cells adds to this complexity, as cancer cells can change their phenotypes in progress and adaptation to new microenvironmental conditions (Lüönd et al., 2021). Molecular heterogeneity and clonal evolution are thus key aspects of breast cancer biology and a critical hurdle for the identification of widely shared molecular features (Caswell-Jin et al., 2021).

A powerful method to explore this complexity is using transcriptomic profiling to identify genome-wide alterations in the expression of genes that are linked to malignant transformation and disease progression. The advent of transcriptomics has enabled molecular classification, identification of potential biomarkers, characterization of dysregulated pathways and investigation of mechanisms of therapeutic response (Supplitt et al., 2021). The molecular signatures associated with diseases can be further linked to biologically relevant genes and pathways using functional genomic approaches that enhance interpretation of the molecular signatures of diseases (Ghoussaini et al., 2021). Multiple genomic resources have shown how valuable it is to draw conclusions from molecular and clinical data to explore cancer biology and generate testable hypotheses (Zhao et al., 2021). Likewise, the integration of transcriptomic information with other molecular layers can aid in prioritizing candidate genes that are associated with disease (Chan et al., 2020).

The molecular heterogeneity of breast cancer is relevant for the biological characterization and for response to therapy. Different breast cancer phenotypes may have unique genomic and transcriptomic features, as evidenced by the genomic and transcriptomic differences between inflammatory and non-inflammatory breast cancers (Zhou et al., 2024). Additionally, the composition and diversity of the tumor cells could cause distinct responses to treatment, which may help resistant sub-populations to survive and grow in response to therapy (Baliu-Piqué et al., 2020). Therefore, the characterization of molecular heterogeneity is becoming more and more important for precision treatment approaches (Carlino et al., 2024). Investigating the transcriptome in independent patients populations may consequently offer the opportunity to pinpoint molecular changes that are shared between or across different populations of breast cancer patients, but not specific to any one cohort. Although there are significant efforts to study the expression changes in genes across breast cancer patients, it is difficult to reproduce results across independent sets of data and studies due to influences from cohort composition, biological variability, experimental platforms and analysis methods. Analyses of individual cohorts can therefore reveal signals that are not consistently observed across all of the data, and do not reflect the molecular picture of breast cancer overall. Comparative analysis of independent transcriptomic cohorts offers a way to identify cohort-specific changes that are reproducible across cohorts and to see if consistently altered genes overlap in biological processes. This cross-cohort investigation would be especially important for a disease with high molecular and phenotypic heterogeneity (Guo et al., 2023).

To this end, this study compares independent breast cancer transcriptomic cohorts to find common and strong gene-expression signatures and to explore their biological implication. The specific goals are:

1. To characterize gene-expression differences between breast cancer and non-tumor samples across independent transcriptomic cohorts.
2. To identify reproducible differentially expressed genes shared across independent breast cancer datasets.
3. To determine the biological functions, molecular pathways, and interaction networks associated with consistently altered genes.

2. Materials and Methods

2.1 Study Design and Dataset Acquisition

A secondary-data comparative transcriptomic study was performed on publicly available datasets of gene expression for breast cancer (Xie et al., 2017). The BC-TCGA, GSE2034 and GSE25066 cohorts were added to examine transcriptomic changes in different sets of patients. Because the experimental platform and expression properties vary between cohorts, each cohort was evaluated separately. The simulated data in the repository were not included, as all analyses included were of biological origin.

2.2 Data Preprocessing and Normalization

Gene-expression matrices were analyzed for sample IDs, gene annotations, missing data, duplicates and expression distributions. Where appropriate, genes with low expression were filtered and multiple gene identifiers were merged prior to analysis. Count data from RNA-sequencing were normalised using TMM and then log2 transformed for exploration analyses. Preliminary normalization was performed on microarray-derived expression data, and, as necessary, it was quantile normalized and log2 transformed.

2.3 Exploratory and Differential Gene Expression Analysis

Global transcriptional variation was analyzed by principal component analysis, correlation analysis, hierarchical clustering and expression heat map. Differential expressions were evaluated separately in each cohort: edgeR (negative binomial modelling) for RNA-sequencing count data and limma (moderated empirical Bayes) for microarray data. To control for multiple testing, the Benjamini–Hochberg false discovery rate (FDR) procedure was used. Genes that were significantly differentially expressed had an adjusted p value < 0.05 and $|\log2 \text{ fold change}| \geq 1$.

2.4 Cross-Cohort and Functional Enrichment Analysis

Genes were harmonized for differential expressions using common gene identifiers and compared in independent cohorts. Genes that showed comparable and consistently up or down expression patterns in multiple data sets were selected as breast-cancer-associated candidates that were reproducible. Enriched biological processes and curated pathways were identified using Gene Ontology and KEGG and Reactome. After correction for multiple hypothesis testing, enrichment results were shown as being significant.

2.5 Protein Interaction Network and Statistical Analysis

Protein–protein Interaction (PPI) analysis was conducted in STRING to characterize the molecular interactions and to identify highly connected candidate hub genes of the reproducible candidate genes. Cross-cohort synthesis was done after statistical analysis of each cohort in isolation, at which point adjusted $p < 0.05$ was defined as statistically significant. Key discoveries were summarized in PCA plots, heatmaps, volcano plots,

gene-overlap diagrams and enrichment plots. Candidate genes were selected by their differential expression, direction of expression, reproducibility across cohorts, connectivity to the network, and biological relevance.

3. Results

3.1 Dataset Characteristics and Cross-Cohort Harmonization

Three independent breast cancer transcriptomic cohorts (total 1,368 samples) were included in the study. BC-TCGA had 590 samples, GSE2034 had 286 samples and GSE25066 had 492 samples. After the harmonization and removal of duplications, BC-TCGA had 17,813 unique genes and GSE2034 and GSE25066 had 12,632 and 13,632 unique genes, respectively. Each cohort shared a common set of 10351 genes, which was used for comparative cross-cohort analyses. The demographics and gene harmonization results are summarized in Table 1.

Table 1. Dataset characteristics and gene harmonization

Cohort	Normal/reference samples	Tumor/comparison samples	Total samples	Harmonized genes	Common genes
BC-TCGA	61	529	590	17,813	10,351
GSE2034	107	179	286	12,632	10,351
GSE25066	100	392	492	12,632	10,351

3.2 Global Transcriptomic Variation Across Cohorts

The results of the PCA showed significant differences between the extent of class separation in each of the data sets. BC-TCGA was the most obvious, accounting for 14.98% and 12.15% of variance with PCs 1 and 2, respectively. The standardized difference of the PC1 was 3.529 ($p < 0.001$). By comparison, GSE2034 had significant overlap among the groups, and GSE25066 was moderately separated, with more separation along PC2. The variance and group-separation statistics for each cohort in Figure 1 are provided in Table 2.

Table 2. Principal component analysis and class-separation statistics

Cohort	PC1 variance (%)	PC2 variance (%)	PC1 + PC2 (%)	Strongest separation component	Standardized difference	p value
BC-TCGA	14.98	12.15	27.13	PC1	3.529	<0.001
GSE2034	7.76	4.54	12.30	PC2	-0.300	0.0088
GSE25066	7.76	5.67	13.43	PC2	0.835	<0.001

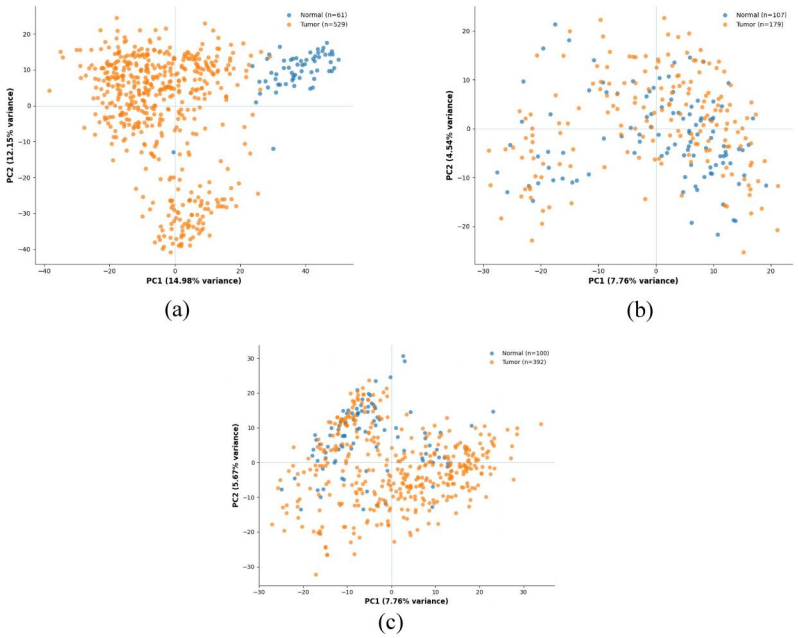


Figure 1. Principal Component Analysis of Transcriptomic Profiles Across Three Breast Cancer Cohorts

PCA illustrates group-wise transcriptomic variation in (a) BC-TCGA, (b) GSE2034, and (c) GSE25066. BC-TCGA shows pronounced separation, GSE2034 exhibits substantial overlap, while GSE25066 demonstrates intermediate separation, highlighting cohort-specific differences in global expression patterns.

3.3 Differential Gene Expression Profiles

The magnitude of transcriptional change was found to be highly variable between cohorts, in a differential-expression analysis. BC-TCGA found a total of 12,856 significant genes, with 7,160 genes being more highly expressed and 5,696 genes being less highly expressed in the comparison group (FDR<0.05). GSE25066 was much more successful with 2,889 significant genes, while GSE2034 was much less successful with only 32. Table 3 summarizes the transcriptional signal produced by each array, showing that BC-TCGA was the array with the greatest transcriptional signal, followed by GSE2034, which had limited differential expression.

Table 3. Differential gene-expression summary

Cohort	Genes tested	FDR-significant genes	Higher in comparison	Lower in comparison	Median |Hedges' g|
BC-TCGA	17,813	12,856	7,160	5,696	0.753
GSE2034	12,632	32	14	18	0.497
GSE25066	12,632	2,889	1,279	1,610	0.388

3.4 Effect-Size-Filtered and Reproducible Gene Signatures

To further enhance the biological relevance, significant genes were also filtered by FDR < 0.05 and |Hedges' g| ≥ 0.50. BC-TCGA kept 5,716 genes, GSE2034 kept 15 genes and GSE25066 kept 583 genes within the common universe of 10,351 genes. Cross-cohort comparison revealed 374 genes with at least two cohorts that supported them. Of those, 164 genes showed concordant directionality, while no gene met the stringent criterion in all three cohorts. These reproducibility results are summarized in Table 4.

Table 4. Effect-size filtering and cross-cohort reproducibility

Criterion	Number of genes
BC-TCGA: FDR < 0.05 and |Hedges' g| ≥ 0.50	5,716
GSE2034: FDR < 0.05 and |Hedges' g| ≥ 0.50	15
GSE25066: FDR < 0.05 and |Hedges' g| ≥ 0.50	583
Supported in at least two cohorts	374
Supported in at least two cohorts with concordant direction	164
Supported in all three cohorts with concordant direction	0

3.5 Cohort-Pair Reproducibility of Candidate Genes

In order to further increase the biological relevance, significant genes were filtered again, this time by FDR < 0.05 and |Hedges' g| ≥ 0.50. BC-TCGA kept 5,716 genes, GSE2034 kept 15 genes, and GSE25066 kept 583 genes in the common space of 10,351 genes. The cross-cohort comparison identified 374 genes that were supported by at least two cohorts. Of the 164, 164 genes showed concordant directionality with no gene meeting the stringent criterion across all three cohorts. These reproducibilities are summarized in Table 4.

Table 5. Cohort-pair support for reproducible genes

Supported cohorts	Concordant higher	Concordant lower	Total genes
BC-TCGA + GSE25066	80	83	163
BC-TCGA + GSE2034	0	1	1
GSE2034 + GSE25066	0	0	0

3.6 Cross-Cohort Effect-Size Patterns of Candidate Genes

We found significant directional consistency in the strongest reproducible genes between the BC-TCGA and GSE25066. While METRN, GATA3, KRT18 and CCND1 had positive standardized effects on both cohorts, NFIB, SFRP1, UGP2 and GPM6B had negative standardized effects on both cohorts. Overall, the effect sizes of GSE2034 were small, and in some cases near zero. Table 6 shows the representative genes and their standardized effects matched directly with the cross-cohort heatmap shown in Figure 2.

Table 6. Representative reproducible genes and cross-cohort standardized effects

Gene	Direction	BC-TCGA Hedges' g	GSE2034 Hedges' g	GSE25066 Hedges' g
NFIB	Concordant lower	-1.5845	0.0941	-0.9351
METRN	Concordant higher	1.6865	-0.0840	0.9073
GATA3	Concordant higher	0.8311	0.0059	0.7824
SFRP1	Concordant lower	-1.9046	0.2057	-0.7046
UGP2	Concordant lower	-1.9964	0.1643	-0.6999
GPM6B	Concordant lower	-1.7778	0.1095	-0.7091
KRT18	Concordant higher	1.4585	-0.1032	0.7071
FOXC1	Concordant lower	-0.8090	0.0955	-0.6783
CCND1	Concordant higher	0.6766	-0.1758	0.7877
EEF1A2	Concordant higher	1.0790	-0.4762	0.6747

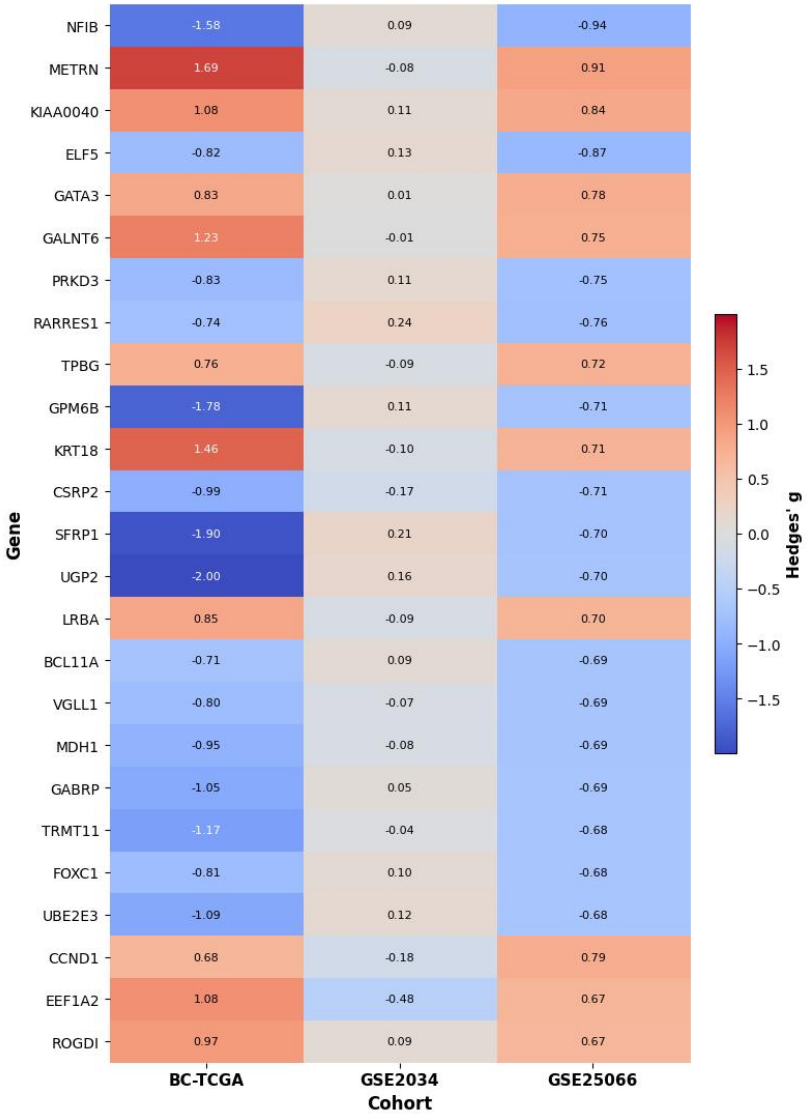


Figure 2. Cross-Cohort Standardized Effect Sizes of the Top 25 Reproducible Gene Signatures

The heatmap shows Hedges' g values for the top reproducible genes across three cohorts. BC-TCGA and GSE25066 demonstrate strong directional concordance, particularly for NFIB, SFRP1, GATA3, UGP2, and

CCND1, whereas GSE2034 generally exhibits weaker effects. This pattern highlights substantial cohort-dependent heterogeneity and stronger transcriptomic reproducibility between BC-TCGA and GSE25066.

4. Discussion

The present cross-cohort transcriptomic analysis revealed substantial heterogeneity in gene-expression patterns among the three breast cancer datasets and demonstrated that statistical significance within an individual cohort does not necessarily indicate reproducibility across independent populations. BC-TCGA exhibited the strongest differential-expression signal, with 12,856 FDR-significant genes, compared with 2,889 in GSE25066 and only 32 in GSE2034. Principal component analysis similarly demonstrated strong group separation in BC-TCGA, intermediate separation in GSE25066, and extensive overlap in GSE2034. These findings are consistent with previous evidence that breast cancer progression involves extensive transcriptional reprogramming and considerable molecular heterogeneity (Abba et al., 2004; Eswaran et al., 2012). Integrated genomic-transcriptomic investigations have likewise demonstrated that breast tumors possess diverse molecular architectures shaped by genomic alterations and distinct transcriptional programs (Smid et al., 2016). Applying both FDR < 0.05 and $|\text{Hedges' } g| \geq 0.50$ considerably refined the initial differential-expression results. BC-TCGA retained 5,716 genes, GSE25066 retained 583, and GSE2034 retained only 15. This effect-size-aware approach reduced the likelihood of prioritizing genes solely because of highly significant p values associated with large sample sizes. Combining statistical evidence with measures of biological magnitude is particularly relevant to high-throughput transcriptomic investigations, where thousands of simultaneous comparisons can produce extensive candidate lists (Siavoshi et al., 2022). Multi-cohort strategies can further improve robustness by distinguishing reproducible molecular signals from dataset-specific associations (Haynes et al., 2017; Sweeney et al., 2017).

Cross-cohort comparison identified 374 genes supported by at least two datasets, of which 164 displayed concordant directionalities, including 80 concordantly higher and 84 concordantly lower genes. Importantly, no gene satisfied the stringent statistical, effect-size, and directional criteria across all three cohorts. Instead, 163 of the 164 concordant genes were reproduced between BC-TCGA and GSE25066, while PCNP was the only concordant gene shared between BC-TCGA and GSE2034. This pattern demonstrates that the identified transcriptomic signature was predominantly cohort-dependent rather than universally conserved. Such variability is compatible with the recognized genomic and functional diversity of breast cancer (Korkola & Gray, 2010; Marcotte et al., 2016).

Several reproducible genes were particularly notable. SFRP1, NFIB, UGP2, GPM6B, and FOXC1 exhibited concordant negative effects, whereas GATA3, CCND1, KRT18, METRN, and GALNT6 demonstrated concordant positive effects between BC-TCGA and GSE25066. These genes encompass transcriptional regulation, epithelial differentiation, cell-cycle control, and other processes relevant to tumor biology. Their diversity supports previous observations that breast cancer development involves coordinated disruption of multiple molecular pathways rather than a single regulatory mechanism (Vishnubalaji et al., 2019). Network-based approaches have similarly shown that interactions among functionally connected genes can provide greater biological insight than isolated markers (Zhuang et al., 2015).

Overall, the findings demonstrate that effect-size filtering and independent cohort comparison can substantially refine large differential-expression profiles into a smaller set of reproducible candidates. However, limited replication in GSE2034 emphasizes the need for cautious interpretation and additional functional and clinical validation. Integration of these candidates with pathway, interaction-network, and functional genomic evidence could further prioritize biologically meaningful genes and potential therapeutic targets in breast cancer (Marcotte et al., 2016; Yoshimaru et al., 2021).

5. Conclusion

In this study, we have shown that, by combining transcriptional data from multiple cohorts, one can get a solid framework to distinguish which signals of the expression changes are reproducible with breast cancer or specific cohort. Within individual datasets, significant differential expressions were observed, but not all the differences were reproducible nor was the intensity of the expression differences large. To narrow the gene sets and include only those with potential biological significance, false-discovery-rate (FDR) control was combined with standardized effect-size filtering. These resulted in a gene set of 374 genes supported by at least 2 cohorts and 164 genes with concordant directional effects. Of these 163, only PCNP was supported by both BC-TCGA and GSE2034 and no gene was supported by both BC-TCGA and GSE25066 under stringent criteria for gene reproduction. The presence of prominent candidates such as SFRP1, NFIB, GATA3, CCND1, FOXC1, KRT18 and UGP2 points towards potentially relevant transcriptional programs in breast cancer biology. Importantly, there was no three-cohort concordance, highlighting molecular heterogeneity and the importance of not believing signatures from one dataset to be widely applicable as biomarkers. In conclusion, the results

presented here form a strong effect-size driven transcriptomic framework and offer a prioritized gene set of reproducible candidate genes to be followed up for pathway characterization, interaction network analysis, functional validation and evaluation of their diagnostic, prognostic and therapeutic value in breast cancer.

References

1. Guo, L., Kong, D., Liu, J., Zhan, L., Luo, L., Zheng, W., ... & Sun, S. (2023). Breast cancer heterogeneity and its implication in personalized precision therapy. Experimental hematology & oncology, 12(1), 3.
2. Turner, K. M., Yeo, S. K., Holm, T. M., Shaughnessy, E., & Guan, J. L. (2021). Heterogeneity within molecular subtypes of breast cancer. American Journal of Physiology-Cell Physiology, 321(2), C343-C354.
3. Lüönd, F., Tiede, S., & Christofori, G. (2021). Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. British journal of cancer, 125(2), 164-175.
4. Supplitt, S., Karpinski, P., Sasiadek, M., & Laczmanska, I. (2021). Current achievements and applications of transcriptomics in personalized cancer medicine. International journal of molecular sciences, 22(3), 1422.
5. Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E. M., Hercules, A., ... & Dunham, I. (2021). Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. Nucleic acids research, 49(D1), D1311-D1320.
6. Zhao, Z., Zhang, K. N., Wang, Q., Li, G., Zeng, F., Zhang, Y., ... & Jiang, T. (2021). Chinese Glioma Genome Atlas (CGGA): a comprehensive resource with functional genomic data from Chinese glioma patients. Genomics, proteomics & bioinformatics, 19(1), 1-12.
7. Chan, M. Y., Efthymios, M., Tan, S. H., Pickering, J. W., Troughton, R., Pemberton, C., ... & Richards, A. M. (2020). Prioritizing candidates of post–myocardial infarction heart failure using plasma proteomics and single-cell transcriptomics. Circulation, 142(15), 1408-1421.
8. Zhou, K., Zhang, M., Zhai, D., Wang, Z., Liu, T., Xie, Y., ... & Lin, Y. (2024). Genomic and transcriptomic profiling of inflammatory breast cancer reveals distinct molecular characteristics to non-inflammatory breast cancers. Breast Cancer Research and Treatment, 208(2), 441-459.
9. Caswell-Jin, J. L., Lorenz, C., & Curtis, C. (2021). Molecular heterogeneity and evolution in breast cancer. Annual Review of Cancer Biology, 5(1), 79-94.
10. Baliu-Piqué, M., Pandiella, A., & Ocana, A. (2020). Breast cancer heterogeneity and response to novel therapeutics. Cancers, 12(11), 3271.
11. Carlino, F., Solinas, C., Orditura, M., Bisceglia, M. D., Pellegrino, B., & Diana, A. (2024). Heterogeneity in breast cancer: clinical and therapeutic implications. Frontiers in oncology, 14, 1321654.
12. Abba, M. C., Drake, J. A., Hawkins, K. A., Hu, Y., Sun, H., Notcovich, C., ... & Aldaz, C. M. (2004). Transcriptomic changes in human breast cancer progression as determined by serial analysis of gene expression. Breast Cancer Research, 6(5), R499.
13. Eswaran, J., Cyanam, D., Mudvari, P., Reddy, S. D. N., Pakala, S. B., Nair, S. S., ... & Kumar, R. (2012). Transcriptomic landscape of breast cancers through mRNA sequencing. Scientific reports, 2(1), 264.
14. Smid, M., Rodríguez-González, F. G., Sieuwerts, A. M., Salgado, R., Prager-Van der Smissen, W. J., Vlugt-Daane, M. V. D., ... & Martens, J. W. (2016). Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration. Nature communications, 7(1), 12910.
15. Vishnubalaji, R., Sasidharan Nair, V., Ouararhni, K., Elkord, E., & Alajez, N. M. (2019). Integrated transcriptome and pathway analyses revealed multiple activated pathways in breast cancer. Frontiers in oncology, 9, 476320.
16. Haynes, W. A., Vallania, F., Liu, C., Bongen, E., Tomczak, A., Andres-Terrè, M., ... & Khatri, P. (2017). Empowering multi-cohort gene expression analysis to increase reproducibility. In Pacific symposium on biocomputing 2017 (pp. 144-153).
17. Sweeney, T. E., Haynes, W. A., Vallania, F., Ioannidis, J. P., & Khatri, P. (2017). Methods to increase reproducibility in differential gene expression via meta-analysis. Nucleic acids research, 45(1), e1-e1.
18. Siavoshi, A., Taghizadeh, M., Dookhe, E., & Piran, M. (2022). Gene expression profiles and pathway enrichment analysis to identification of differentially expressed gene and signaling pathways in epithelial ovarian cancer based on high-throughput RNA-seq data. Genomics, 114(1), 161-170.

19. Zhuang, D. Y., Jiang, L. I., He, Q. Q., Zhou, P., & Yue, T. (2015). Identification of hub subnetwork based on topological features of genes in breast cancer. International Journal of Molecular Medicine, 35(3), 664-674.
20. Marcotte, R., Sayad, A., Brown, K. R., Sanchez-Garcia, F., Reimand, J., Haider, M., ... & Neel, B. G. (2016). Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. Cell, 164(1), 293-309.
21. Korkola, J., & Gray, J. W. (2010). Breast cancer genomes—form and function. Current opinion in genetics & development, 20(1), 4-14.
22. Yoshimaru, T., Nakamura, Y., & Katagiri, T. (2021). Functional genomics for breast cancer drug target discovery. Journal of human genetics, 66(9), 927-935.
23. Xie, Haozhe; Li, Jie; Jatkoe, Tim; Hatzis, Christos (2017), "Gene Expression Profiles of Breast Cancer", Mendeley Data, V1, doi: 10.17632/v3cc2p38hb.1