HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 2 (2024)

Multi-Omics Review of GATA-3

Yashas Revanakara

Irvington High School, USA

ABSTRACT

Breast cancer is a prevalent mutation-caused disease that affects millions of people worldwide. One such gene that is related to causing the cancer is GATA3. This gene is a regulator of t-cell development and is also involved with causing luminal breast cancer when altered. The type of gene which GATA3 was never clear, displaying both properties of an oncogene and a tumor suppressor gene. The TCGA dataset provided by the Cancer Genome Atlas program provides relevant data through the numerous patients who have submitted data in the program. GATA3 has most of its mutations as truncation in both alleles with the X308_splice mutation. This recessive property suggests it is a tumor suppressor gene. Other mutations like the M293, a missense mutation, aren't well researched but provide signs of being oncogenic. GATA# also initially increases survivability when highly altered proposing it to be a tumor suppressor gene. The role of GATA3 is still not well defined but shows most properties of a tumor suppressor gene.

Introduction

Breast cancer, a disease where cells in the breast grow out of control, is diagnosed in over 240,000 people and kills over 42,000 women and 500 men every single year. [1] Being the most diagnosed cancer for women in 2023, and also casing the 4th most amount of deaths in types of cancer, breast cancer is prevalent in modern society. [2] This cancer can develop from DNA damage and mutations increasing the familial risk of passing down the disease. Breast cancer can be split into two different categories, invasive and non-invasive. [3] The non-invasive type can be lobular carcinoma in situ (LCIS) and ductal carcinoma in situ (DCIS). LCIS is defined as a breast change in the milk glands, or lobules. This change is usually the abnormal growth of cells within the area. DCIS on the other hand is known as a non-invasive breast cancer or a precursor at stage 0. [4] In this condition all the cells that are on the milk ducts have become cancer cells but have not yet spread to other tissue in the area. If it is not treated it can usually spread and then form into an invasive cancer. Invasive cancer can also be divided into groups, ductal and lobular types. Invasive ductal carcinoma accounts for up to 75% of all breast cancers. This type originates in the lobules and spreads around to the other tissues. Invasive ductal cancer tends to appear as a compound allowing it to be more present in mammograms and easier to diagnose. Invasive lobular cancer tends to be in a more linear fashion, allowing it to escape being detected in mammograms or even during physical inspection by doctors. This cancer accounts for 10% of all breast cancer.[5] This can eventually develop to utilize lymph nodes in the body and the bloodstream to spread making it metastatic breast cancer. [6] Breast cancer has been a focus of research for many, leasing to mass amounts of information already being known about the cancer. One such topic of research has been causation as cancer is a gene-related disease.

Oncogene

One known oncogene in the field of cancer is GATA3. This gene belongs to the GATA family of transcription factors and the protein contains two GATA-type zinc fingers which are regulators of t-cell development. [7] It regulates the development, proliferation, and maintenance of T-cells and has also recently been discovered in

lymphoid cells.[8] Gata3 is involved in many different signaling pathways were discovered with the tool String. [9] It's involved with various other proteins such as FoxA1 which it is coexpressed with, LSB1, TAL1, and others. Through these interactions, lots of research has been shown to prove GATA3's implication in cancer, breast cancer specifically. GATA3 is a marker in the luminal pattern of gene expression. This means that when GATA3 is altercated it causes luminal breast cancer which means the cells that are lined along the mammary gland are affected. If it is not expressed then there is invasive growth and a low chance of recovery for the disease. Two properties of GATA3 cause this. The first is that it participates within a positive feedback loop with the gene ER alpha and the GATA3 gene turns on the estrogen receptor gene which creates estrogen. This then binds to the actual Estrogen Receptor A which causes an overgrowth in the cells. The next is the features that the altercation of this gene causes are related to poor prognosis. It interrupts the cell's routing of undergoing epithelial-to-mesenchymal transition (EMT) and metastasis. [10] GATA3 also displays many hallmarks of cancer including proliferative signaling, escaped programmed cell death, suppressed cell growth, and more. The most noticeable ones are overexpression-causing murine in breast cancer. Other traits displayed by the gene are its homozygous deletion relating to the apoptosis in cells of mice. These properties root it as an even more important gene in cancer displaying many properties as an oncogene. [11]

Tools

CBioPortal

An analysis of the gene GATA3 can be done using the mass amounts of data already produced. However many tools need to be used in order to interpret all this data. One such important tool is the cBioPortal for Cancer Genomics repository. It was used to analyze the multi-omics data that was provided by the Cancer Genome Atlas, also known as TCGA Pancancer. It was created to make larger cancer genomics datasets more digestible for the cancer research community. [12] It contains over 5,000 tumor samples from over 20 datasets on cancer genomics.

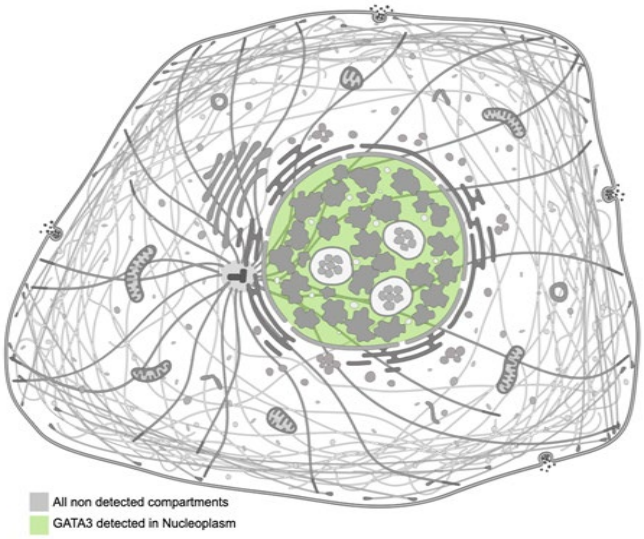


Figure 1. Representation of the location of gene GATA3 in the cell

UCSC Genome Browser

Another tool is the UCSC Genome Browser which displays portions of the genome in a visual manner. [13] Its data collection, or tracks and annotations, are endless and constantly updated to reflect connections that have been made. [14] The visualization of the DNA has been expanded with the BLAT and PCR tools.

Analysis

In Figure 1, the location of GATA3 is displayed as it is only found in the nucleoplasm. This location is where all the suspended DNA and RNA are located in the nucleus. The location suggests that it is mainly involved in the transcription, repair, modification, and development of DNA in the cell. [15] The gene is located on chromosome 10 at p14 and starts at position 8,054,688 and ends at position 8,075,198. The length of the gene is 20,511 base pairs. This gene contains 55 somatic mutations which were mostly in exons 5 and 6. Mutations can be grouped into 3 categories: splice site mutations at the exon 4/5 junction and the exon 5/6 junction, frameshift mutations in exon 6, and frameshift mutations in zinc finger 2. [16]

GATA3 has been profiled in the TCGA dataset accessible through cBioPortal. With it, 13% of patients had the amplification of the Gata3 which included missense mutations, truncating mutations, and splice mutations. The missense mutation means that the amino acid sequence was changed due to an alteration of the encoding which creates a whole different protein. Truncating mutations simply represent the shortening of the sequence of genes cutting off a part of the sequence which typically has large effects on the transcription of the protein. Lastly, the splice mutations which are insertions or deletions of nucleotide at a specific site in the DNA.

In Figure 2, the types of cancer are detailed and in the TCGA database. With the chart detailed, GATA3 is only present in breast cancer with 13% mutations as discussed before. However, breast cancer, it was most mutated in the mixed ductal and lobular carcinoma type of cancer. This type of cancer is still a grey area in the field of cancer research being stated as displaying symptoms of both subtypes of cancer. It displays at least 10% as ductal and over 50% as the lobular part of the tumor. [17] The other types of cancer, ranging from highest frequency to lowest, are invasive ductal carcinoma, metaplastic, invasive mixed mucinous carcinoma, and ILC. Two types that stand out as uncommon within the general scheme of breast cancer are metaplastic breast cancer and mixed mucinous breast cancer. Metaplastic breast cancer accounts for less than 1% of all breast cancers therefore reflecting the minute amounts of research that has been done on it. This is immune to most drugs because of its lack of estrogen receptors and progesterone receptors. [18] Mucinous carcinoma are cancer cells in mucin, the main component of mucus which means the mucus is now involved with the tumour. [19]

GATA3 is moderately mutated in the data on cBioPortal and this lollipop plot shows the difference in the mutations.

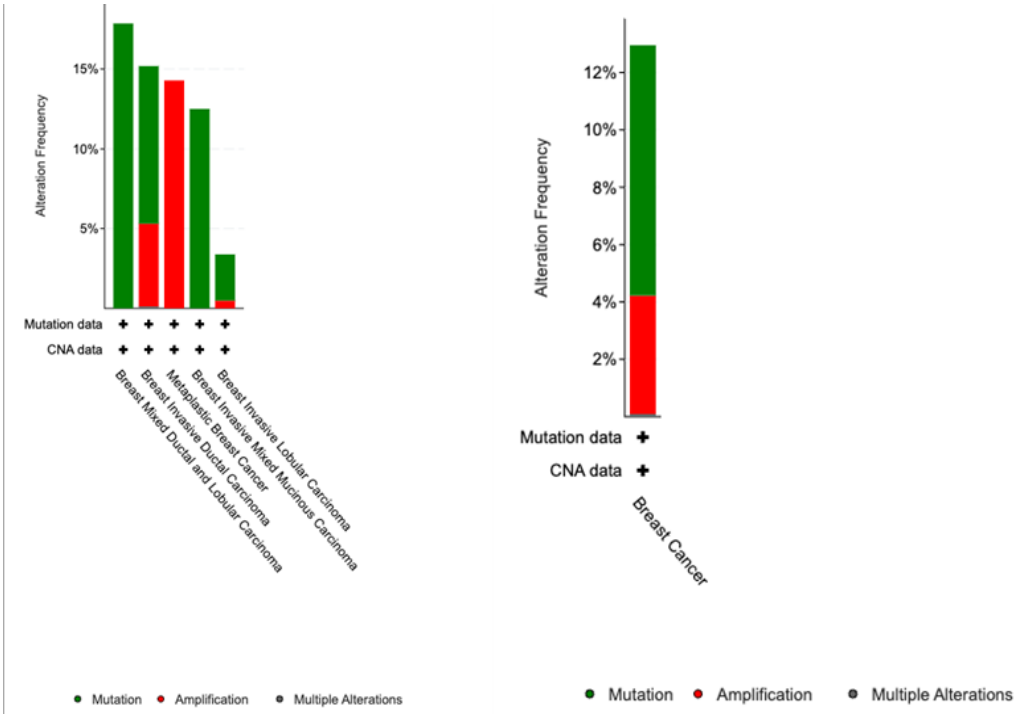
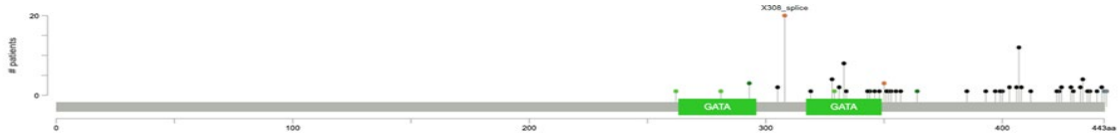


Figure 2. Pan-Cancer analysis of the gene GATa3 modeling cancer types and subtypes of cancer where GATA3 is expressed

The lollipop plot is a resource from cBioPortal's mutation mapper which displayed how prevalent certain mutations are on a gene. This helps to decipher how widespread certain mutations are and their type of mutations. [20] In Figure 3, the two zinc fingers are displayed for GATA3 with their stop and start also labeled on the x- axis which represents the position on the protein. Furthermore, the y axis represents the number of patients within the study that have displayed a mutation of the genes on the lollipop plot. For GATA3 there are two hotspots on the gene, M293 and R364, both of which are missense mutations. Even though the most repeated mutations are missense, a majority of mutations for the GATA3 are truncating mutations. The M293 mutation has not shown a clear effect on the gene and but it is likely to be oncogenic. [21] The same applies with the mutation R364 which has been proven to be repeating but the definitive result has not been yet researched. [21] The mutation with the most prevalence is a splice mutation also noted as X308_splice. A splice mutation means that the actual cause of the mutation is unknown, like if it is an insertion, deletion, etc. As stated previously, the most common type of mutation was truncation mutations which are in the range of 308 to 443. These mutations occur on the c-terminal location of GATA3 and are thus responsible for the production of functional proteins. These mutations have been heavily linked to the prognosis of breast cancer. Specifically, the truncations have been linked with causing increased tumor growth, upregulation of genes of peptidyl-tyrosine modification and epithelial-to-mesenchymal transition. The increased tumor growth is an obvious contributor to breast cancer suggesting that its an oncogene. Furthermore, the upregulation of genes for the peptidyl-tyrosine modification means the mutation changes the expression of those genes that are involved with the tyrosine residues. These changes can include a post-translational modification through phosphorylation which causes problems in the cell-signlaing pathway solidifying its name as an oncogene. Lastly a loss of function, or mutation in GATA3 results in epithelial-to-mesenchymal transition (EMT). [22]



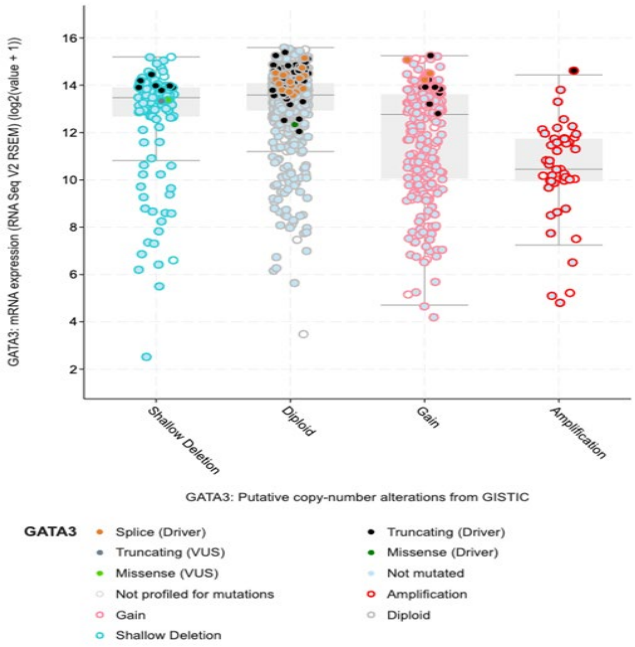


Figure 3. is a lollipop plot of gene GATA3 describing the mutations on the gene.
Figure 4. Is a mRNA vs Copy number alterations plot also displaying types of mutations.

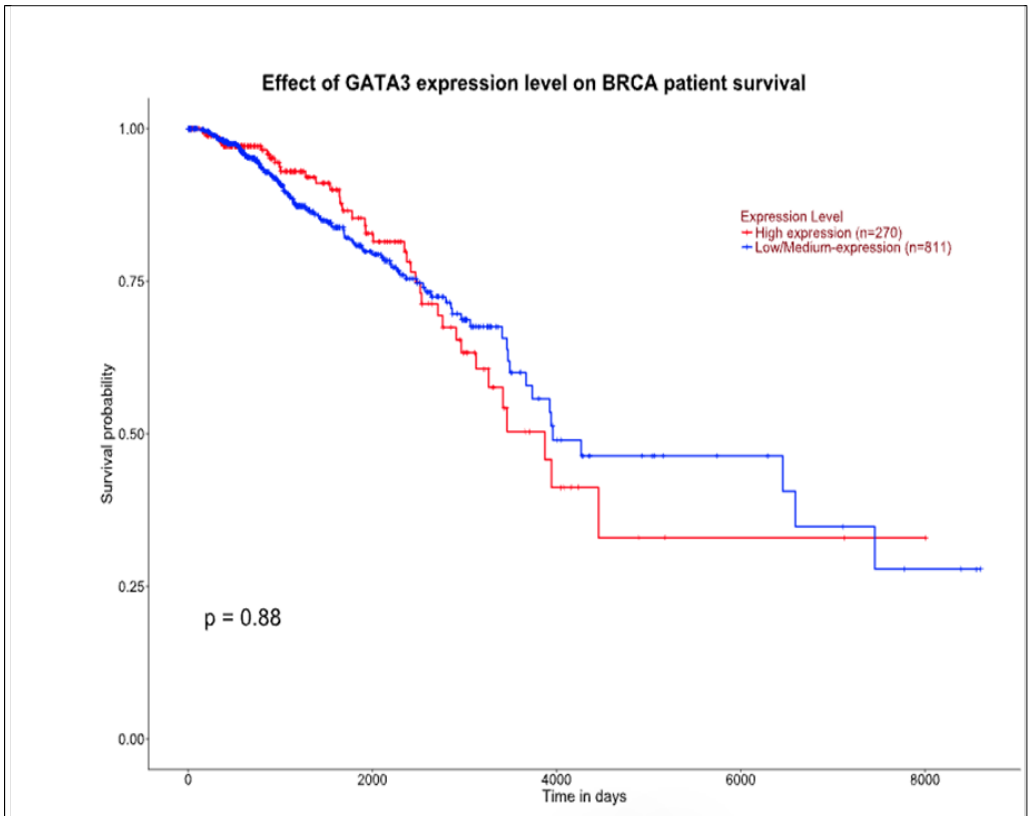


Figure 5. GATA3 expression level related to the survival after a number of days.

mRNA vs CNA

Figure 4 displays a graph of the copy number alterations relation to the mRNA expression. First off copy number alterations are changes in the structure of the chromosome that result in the change of the DNA through gain or loss. [23] Through the graph, it is visible that through the amplification of the gene there is a slight reduction in the expression of mRNA. Most of the truncating mutations that were profiled to be heavily linked to breast cancer can be found in the diploid category. This suggests that both alleles of the gene are experiencing the mutation X308_splice. Because both alleles are affected, there is a more significant impact on cellular function as proteins will stop getting produced from that gene. There is no backup gene to still have the protein created if one of the alleles are affected. Similarly, the shallow deletion and gain category both represent the same amount of expression in the graph. Both types of mutations are small in context compared to an amplification thus suggesting that they are similar in their loss of genetic material.

Figure 5 displays a Kaplan-Meir curve which shows the survivability of patients as time goes on depending on how GATA3 is expressed. There are two graphs, one for a high expression of the gene, and one for a lower expression of the gene. Clearly, the High expression of the gene leads to a lower survivability after reaching approximately 2000 days. The steeper slope on the high expression of the gene relates to a higher death rate suggesting that it has a worse survival than the control, low expression. The corresponding p-value is high suggesting that this data is convincing and that the higher expression of the gene does lead to a higher mortality rate. Since the p-value is high it is logical to assume that random chance did not play a factor in these results and it is an accurate representation of the effect of expression on the gene. However, the initial survivability is better with high expression because of GATA3's role in the cel. It needs to maintain the state of luminal cells so with higher expression the tumor is initially less aggressive. Not much research has been done to find

the reason for supposedly having more aggressive tumors after a period of 2000 days leading to lower survival probability.

Conclusion

To solve the debate of oncogene or tumor suppressor, GATA3 has displayed many properties suggesting it is on one side. Oncogenes are defined as genes that will change the cell when its expression or function is changed. And the normal version of these genes, proto-oncogenes, perform regular functions that are vital to cell homeostasis. [24] GATA3 displays little to none of these properties declaring it as a tumor suppressor. As displayed through relevant discussion, the presence of GATA3 inhibits the EMT process. This has also been supported by its involvement with estrogen-receptor. [25] Furthermore, with the truncation of GATA3, the most common mutation, there was an upregulation of genes that contribute to peptidyl-tyrosine modification further increasing tumor size and growth. However, on the other hand, some properties, such as the missense mutation M293, display oncogenic properties which may throw off the assertion that GATA3 is a tumor suppressor. Tumor suppressor genes also require inactivation of both alleles to leave for any type of effect within the cell which is common through deletion. [26] GATA3 displays these properties in the diploid mutation most of the mutations, and truncations were found in that category. While there is no definitive role of GATA3 the signs strongly imply its main role is as a tumor suppressor in breast cancer.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

1. Centers for Disease Control and Prevention. (2023, July 25). *Basic information about breast cancer.* Centers for Disease Control and Prevention. https://www.cdc.gov/cancer/breast/basic_info/index.htm#:~:text=Each%20year%20in%20the%20United,What%20Is%20Breast%20Cancer%3F
2. *Common cancer sites - cancer stat facts.* SEER. (n.d.). https://seer.cancer.gov/statfacts/html/common.html
3. Breast cancer - statpearls - NCBI bookshelf. (n.d.). https://www.ncbi.nlm.nih.gov/books/NBK482286/
4. *Ductal carcinoma in situ (DCIS).* Ductal Carcinoma in Situ (DCIS) | American Cancer Society. (n.d.). https://www.cancer.org/cancer/types/breast-cancer/about/types-of-breast-cancer/dcis.html
5. Depolo, J. (n.d.). *Invasive Lobular Carcinoma.* Invasive lobular carcinoma: What it is, diagnosis, and treatments. https://www.breastcancer.org/types/invasive-lobular-carcinoma
6. Penn Medicine. (n.d.). Pennmedicine.org. https://www.pennmedicine.org/cancer/types-of-cancer/breast-cancer/types-of-breast-cancer/invasive-lobular-carcinoma
7. *Gene Cards* . Gene Cards -Human Gene Databse. (n.d.). https://www.genecards.org/cgi-bin/carddisp.pl?gene=GATA3
8. Wan, Y. Y. (2014, June). *GATA3: A master of many trades in immune regulation.* Trends in immunology. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4045638/
9. Szklarczyk , D. (n.d.). *The string database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest.* Nucleic acids research. https://pubmed.ncbi.nlm.nih.gov/36370105/

10. Liu, Y., Ma, H., & Yao, J. (2020). ERα, A Key Target for Cancer Therapy: A Review. OncoTargets and Therapy, Volume 13, 2183–2191. https://doi.org/10.2147/ott.s236532
11. Zbysław Sońdka, Nidhi Bindal Dhir, Carvalho-Silva, D., Jupe, S., Madhumita, McLaren, K., Starkey, M., Ward, S., Wilding, J. L., Ahmed, M., Argasinska, J., Beare, D., Chawla, M. L., Duke, S. O., Fasanella, I., Avirup Guha Neogi, Haller, S., Balázs Hetényi, Hodges, L. K., & Holmes, A. (2023). COSMIC: a curated database of somatic variants and clinical data for cancer. Nucleic Acids Research, 52(D1), D1210–D1217. https://doi.org/10.1093/nar/gkad986
12. Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., & Schultz, N. (2012). The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. Cancer Discovery, 2(5), 401–404. https://doi.org/10.1158/2159-8290.cd-12-0095
13. Karolchik, D., Hinrichs, A. S., & Kent, W. J. (2009). The UCSC Genome Browser. Current Protocols in Bioinformatics, 28(1). https://doi.org/10.1002/0471250953.bi0104s28
14. Lee, B. T., Barber, G. P., Benet-Pagès, A., Casper, J., Clawson, H., Diekhans, M., Fischer, C., Gonzalez, J. N., Hinrichs, A., Lee, C., Muthuraman, P., Nassar, L., Nguy, B., Pereira, T., Perez, G., Raney, B., Rosenbloom, K., Schmelter, D., Speir, M., & Wick, B. (2021b). The UCSC Genome Browser database: 2022 update. Nucleic Acids Research, 50(D1), D1115–D1122. https://doi.org/10.1093/nar/gkab959
15. Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C. A.-K. ., Odeberg, J., Djureinovic, D., Takanen, J. O., Hober, S., & Alm, T. (2015). Tissue-based map of the human proteome. Science, 347(6220), 1260419–1260419. https://doi.org/10.1126/science.1260419
16. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. The human genome browser at UCSC. Genome Res. 2002 Jun;12(6):996-1006.
17. McCart Reed, A. E., Kutasovic, J. R., Nones, K., Saunus, J. M., Da Silva, L., Newell, F., Kazakoff, S., Melville, L., Jayanthan, J., Vargas, A. C., Reid, L. E., Beesley, J., Chen, X. Q., Patch, A.-M., Clouston, D., Porter, A., Evans, E., Pearson, J. V., Chenevix-Trench, G., & Cummings, M. C. (2018). Mixed ductal-lobular carcinomas: evidence for progression from ductal to lobular morphology. The Journal of Pathology, 244(4), 460–468. https://doi.org/10.1002/path.5040
18. Metaplastic Breast Cancer. (n.d.). MD Anderson Cancer Center. Retrieved January 23, 2024, from https://www.mdanderson.org/cancer-types/breast-cancer/metaplastic-breast-cancer.html#:~:text=Metaplastic%20breast%20cancer%20makes%20up
19. Mucinous Carcinoma: Definition, Pathology & Treatment. (n.d.). Cleveland Clinic. https://my.clevelandclinic.org/health/diseases/22975-mucinous-carcinoma
20. Jay, J. J., & Brouwer, C. (2016). Lollipops in the Clinic: Information Dense Mutation Plots for Precision Medicine. PLOS ONE, 11(8), e0160519. https://doi.org/10.1371/journal.pone.0160519
21. Suehnholz et al., Cancer Discovery 2023 and Chakravarty et al., JCO PO 2017
22. Liu, X., Bai, F., Wang, Y., Wang, C., Ho Lam Chan, Zheng, C., Fang, J., Zhu, W., & Pei, X.-H. (2023). Loss of function of GATA3 regulates FRA1 and c-FOS to activate EMT and promote mammary tumorigenesis and metastasis. Cell Death and Disease, 14(6). https://doi.org/10.1038/s41419-023-05888-9
23. QIAGEN. (n.d.). Somatic mutations and copy number changes in cancer: finding the right targets - QIAGEN. Www.qiagen.com. Retrieved January 23, 2024, from https://www.qiagen.com/us/spotlight-pages/newsletters-and-magazines/articles/reviews-online-copy-number-alteration/

24. Shortt, J., & Johnstone, R. W. (2012). Oncogenes in Cell Survival and Cell Death. Cold Spring Harbor Perspectives in Biology, 4(12), a009829–a009829. https://doi.org/10.1101/cshperspect.a009829
25. Yu, W., Huang, W., Yang, Y., Qiu, R., Zeng, Y., Hou, Y., Sun, G., Shi, H., Leng, S., Feng, D., Chen, Y., Wang, S., Teng, X., Yu, H., & Wang, Y. (2019). GATA3 recruits UTX for gene transcriptional activation to suppress metastasis of breast cancer. Cell Death & Disease, 10(11), 1–16. https://doi.org/10.1038/s41419-019-2062-7
26. Brenner, S., & Miller, J. (2002). Encyclopedia of genetics / 3. M - R. Acad. Press.