Gene expression analysis of TNBC, Non-TNBC, and HER2 breast cancer types

Adi Falach and Assaf Lovton

Abstract

Not only that breast cancer is the most common cancer it is also the most common type of cancer both in women in Israel and the United States women.[1] Creating a genetic profile could help to improve patient classification, predict prognosis, evaluate drug resistance, and identify drug targets. More importantly, current advancements in breast cancer genomics discoveries could be translated into therapeutic advances.[2] Today there are five intrinsic molecular subtypes. Immunohistochemical (IHC) techniques allow the measurement of expression of progesterone receptor (PR), estrogen receptor (ER), and overexpression of human epidermal growth factor receptor 2 (HER2/neu). Breast cancers can be classified by the presence or absence of these receptors. Breast tumors lacking expression of all three receptors are defined as triple-negative breast cancer (TNBC).[3] In this study, we investigated gene expression between three cancer types: HER2, TNBC, and Non-TNBC. We checked whether the current typing system matching the gene expression analysis. We found out that PCA clustered the samples respectively to the known types. We also found common down-regulated genes (CFS3, CSN3, IL6, CSN1S1, CSN2) between all three types and one upregulated gene (MTRNR2L1). We decided to investigate the relation between IL-6 and breast cancer, the bottom line was that studies have yet to decide its role regarding breast cancer and classify it as an appropriate approach for breast cancer therapy.

Introduction

Breast cancer is cancer that develops in breast cells. Breast cancer occurs almost entirely in women, yet men can get breast cancer, as well.[4] Usually, cancer develops in both the lobules and the ducts of the breast. Lobules are the glands that account for producing milk, and ducts are the pathways that convey the milk from the glands to the nipple. Cancer may also develop in the fatty tissue or the fibrous connective tissue within the breast. Uncontrolled cancer cells usually invade other healthy breast tissue

1 "האגודה למלחמה בסרטן - סרטן השד." Accessed August 14, 2021. https://www.cancer.org.il/template/publications.aspx?maincat=12.

2 Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer. 2007;7:169-181.

3 Hon JD, Singh B, Sahin A, et al. Breast cancer molecular subtypes: from TNBC to QNBC. Am J Cancer Res. 2016;6(9):1864-1872. Published 2016 Sep 1

4 "What Is Breast Cancer? | Breast Cancer Definition." Accessed August 15, 2021. https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html.

and can spread to the lymph nodes under the armpit. Lymph nodes are the main way to provide the cancer cells to spread to other parts of the body.[5] Breast cancer is the most common type of cancer both in women in Israel and the United States women.[6] In 2021, it's estimated that about 30% of newly diagnosed cancers in women will be breast cancers (among American women). According to data from the World Health Organization, it is not surprising that breast cancer has become the most common cancer in the world as of 2021, accounting for 12% of new cancer cases worldwide each year.[7]

Risk factors

The main factors are being a female and older age. In addition, lack of childbearing or breastfeeding, genetics, higher levels of hormones – estrogen and Progesterone, obesity. The risk factors are divided into two groups, modifiable and fixed. Lifestyle, obesity, heavy consumption of alcohol, and smoking tobacco are all common modifiable risk factors. While genetics, family history, gender, and age lay under fixed risk factors.

Aging

As findings suggest, the incidence of breast cancer is highly correlated with age growth. In 2016, it was reported that approximately 99.3% and 71.2% of breast cancer-related deaths in the United States occurred in women over 40 and 60 years old, respectively.[8]

Family history

Almost 25% of all breast cancer cases are linked to family history. Women, whose mother or sister has breast cancer, are prone to this disease. A cohort study involving more than 113,000 women in the UK showed that women with one first-degree relative were at 1.75 higher risks of developing this disease than women without any affected relatives.[9] Moreover, the risk becomes 2.5 times or higher in women with at least more first-degree relatives with breast cancer.[10] Hereditary susceptibility to breast cancer is partially attributed to mutations of breast cancer-related genes like BRCA1 and BRCA2.

---

5 “Breast Cancer: Symptoms, Stages, Types and More.” Accessed August 18, 2021. https://www.healthline.com/health/breast-cancer.

6 “סרטן השד - סרטן בסרטן למלחמה האגודה.” Accessed August 26, 2021.
https://www.cancer.org.il/template/publications.aspx?maincat=12.

7 “U.S. Breast Cancer Statistics | Breastcancer.Org.” Accessed August 26, 2021.
https://www.breastcancer.org/symptoms/understand_bc/statistics.

8 Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. CA Cancer J Clin. 2017; 67:7–30.

9 Sun YS, Zhao Z, Yang ZN, et al. Risk Factors and Preventions of Breast Cancer. Int J Biol Sci. 2017;13(11):1387-1397. Published 2017 Nov 1. doi:10.7150/ijbs.21635

10 Brewer HR, Jones ME, Schoemaker MJ. et al. Family history and risk of breast cancer: an analysis accounting for family structure. Breast Cancer Res Treat. 2017; 165:193–200.

Reproductive factors

Reproductive factors such as early menarche, late age at first pregnancy, low parity and late menopause, may increase the risk of breast cancer. Every one-year delay in menopause increases the risk of breast cancer by 3%. Every one-year delay in menarche or each additional birth decreases the risk of breast cancer by 5% or 10%, respectively.[11]

Estrogen

The endogenous estrogen is usually produced by the ovary in premenopausal women and ovariectomy can reduce the risk of breast cancer.[12] The main sources of exogenous estrogen are oral contraceptives and hormone replacement therapy (HRT). HRT involves the administration of exogenous estrogen or other hormones for menopausal or postmenopausal women. Oral contraceptives (birth control pills) have been widely used since the 1960s and the formulations have been upgraded to reduce side effects. Nevertheless, oral contraceptives do not increase the risk of breast cancer in women who stop using them for more than 10 years. Several studies have shown that the use of HRT can increase breast cancer risk. However, the risk of breast cancer has been shown to significantly decrease after two years of stopping HRT.[13]

Lifestyle

Modern lifestyles such as excessive alcohol consumption and eating high-fat food might increase the risk for breast cancer. However, the correlation between these factors and breast cancer is linear. Alcohol consumption can raise the level of estrogen-related hormones in the blood and trigger the estrogen receptor pathways. The modern western diet contains too much fat and excess intake of fat, particularly saturated fat, is associated with mortality and bad prognosis in breast cancer patients.[14] Studies have shown that those who rapidly gained weight in adulthood are at higher risk than those who have been overweight since childhood.

Smoking tobacco appears to increase the risk of breast cancer, while the earlier in life smoking began the greater the risk. Mutagens from cigarette smoke have been detected in the breast fluid from non-lactating women. The risk of breast cancer is also elevated in women who both smoke and drink.[15]

---

11 Washbrook E. Risk factors and epidemiology of breast cancer. Women's Health Medicine. 2006; 3:8–14.

12 Endogenous H, Breast Cancer Collaborative G, Key TJ. et al. Sex hormones and risk of breast cancer in premenopausal women: a collaborative reanalysis of individual participant data from seven prospective studies. Lancet Oncol. 2013; 14:1009–1019.

13 Narod SA. Hormone replacement therapy and the risk of breast cancer. Nature reviews. Clinical oncology. 2011; 8:669–676.

14 Makarem N, Chandran U, Bandera EV. et al. Dietary fat in breast cancer survival. Annu Rev Nutr. 2013; 33:319–348.

15 Knight JA, Fan J, Malone KE. et al. Alcohol consumption and cigarette smoking in combination: A predictor of contralateral breast cancer risk in the WECARE study. Int J Cancer. 2017; 141:916–924.

Genes related to breast cancer

Lots of genes have been identified in relation to breast cancer. Mutations and abnormal amplification of both oncogenes and antioncogenes play key roles in the processes of tumor initiation and progression.

BRCA1/2

Breast cancer-associated genes 1 and 2 (BRCA1 and BRCA2) are two famous anti-oncogenes for breast cancer risk. BRCA1 and BRCA2 mutations account for up to 90% of the total genetic influence with a risk of breast cancer of 60–80% in those affected.[16] BRCA1 is located on chromosome 17q21 and BRCA2 on 13q12. Both genes encode tumor suppressor proteins. BRCA1 deficiency leads to the dysregulation of cell cycle checkpoint, abnormal centrosome duplication, genetic instability, and eventually apoptosis. BRCA2 protein regulates recombinational repair in DNA double-strand breaks by interacting with RAD51 and DMC1. BRCA2-associated breast cancers are more likely to be high-grade invasive ductal carcinomas, but with a luminal phenotype.[17]

A deletion mutation in either BRCA1 or BRCA2 genes increases the risk of breast cancer greatly. About a quarter of hereditary breast cancers and 5-10% of all breast cancers are caused by BRCA1 and BRCA2 mutations.[18]

HER2

Human epidermal growth factor receptor 2 (also known as c-erbB-2) is an important oncogene in breast cancer and is located on the long arm of human chromosome 17 (17q12). The expression of the HER2 gene is activated mainly through gene amplification and re-arrangement. The HER2 protein is an epidermal growth factor receptor (EGFR) of the tyrosine kinase family and forms heterodimers with other ligand-bound EGFR family members such as Her3 and Her4, thus, to activate downstream signaling pathways.[19]

Breast cancer types:

Breast cancer is recognized as a heterogeneous disease. Whereas numerous factors have been investigated to stratify patients by risk and treatment options (parity, age, family genetics, etc.), it has been demonstrated that the receptor status is the

16 Inoue K, Fry EA. Aberrant expression of cyclin D1 in cancer. Signal transduction insights. 2015; 4:1–13.

17 Bane AL, Beck JC, Bleiweiss I. et al. BRCA2 mutation-associated breast cancers exhibit a distinguishing phenotype based on morphology and molecular profiles from tissue microarrays. Am J Surg Pathol. 2007; 31:121–128.

18 Interpreting Signs and Symptoms. Lippincott Williams & Wilkins. 2007. pp. 99–. ISBN 978-1-58255-668-0.

19 "Am I Sick?" Breast Cancer Care. 23 February 2018. Archived from the original on 25 October 2013. Retrieved 22 October 2013.

foremost valuable in predicting prognosis and responsiveness to treatment. Immunohistochemical (IHC) techniques allow the measurement of expression of progesterone receptor (PR), estrogen receptor (ER), and overexpression of human epidermal growth factor receptor 2 (HER2/neu). Breast cancers can be classified by the presence or absence, of these receptors. Breast tumors lacking expression of all three receptors are defined as triple-negative breast cancer (TNBC).[20] Cancer types that express ER, PR or Her2-neu are amenable to targeted therapies directed at these receptors, while traditional chemotherapeutic reagents are applied on TNBC patients.[21]

Breast cancer intrinsic molecular subtypes

Based on expression array analysis that resulted in a classification of five distinct breast cancer intrinsic molecular subtypes: luminal A, luminal B, HER2-enriched, and basal-like breast cancer (BLBC) and normal-like. Based on these findings, a molecular signature containing 50 genes (PAM50), has been devised for clinical use in prognosis and treatment decisions, using IHC determination of the receptor status we are able to classify the type as follows: [22]

Figure 1: Intrinsic subtyping by Immunohistochemical techniques:

Intrinsic subtype	IHC* status	Prognosis	Prevalence
Luminal A	ER+/PR+;HER2−;Ki67 low	Good	40%
Luminal B	ER+/PR+;HER2−;Ki67 high	Intermediate	20%
	ER+/PR+;HER2+;Ki67 any	Intermediate	
HER2 overexpression	ER−/PR−;HER2+;Ki67 any	Poor	12 to 21%
Basal	ER−/PR−;HER2−;basal marker+	Poor	11 to 23%
Normal-like	ER+/PR+;HER2−Ki67 any	İntermediate	3 to 10%

Moreover, the type of treatment can be chosen based on the patient subtypes since patients with luminal A and B, and HER2-enriched subtypes are sensitive to targeted treatments, while patients with BLBC currently have only chemotherapy as an option.

Triple-negative breast cancer

TNBC makes up 10-30% of all breast cancers. TNBC is associated with a higher stage at diagnosis, younger age, higher nuclear grade and mitotic activity, and poorer prognosis. Within the TNBC designation are heterogeneous characteristics. TNBC can be

20 Hon JD, Singh B, Sahin A, et al. Breast cancer molecular subtypes: from TNBC to QNBC. Am J Cancer Res. 2016;6(9):1864-1872. Published 2016 Sep 1

21 Eswaran, J., Cyanam, D., Mudvari, P. et al. Transcriptomic landscape of breast cancers through mRNA sequencing. Sci Rep 2, 264 (2012). https://doi.org/10.1038/srep00264

22 Güler EN. Gene Expression Profiling in Breast Cancer and Its Effect on Therapy Selection in Early-Stage Breast Cancer. Eur J Breast Health. 2017;13(4):168-174. Published 2017 Oct 1. doi:10.5152/ejbh.2017.3636

categorized by its morphological appearance. Currently, the only treatment for patients with TNBC is chemotherapy.

Survival rates

Statistics were taken from the SEER (Surveillance, Epidemiology, and End Results) database, maintained by the National Cancer Institute (NCI). Based on women diagnosed with breast cancer between 2010 and 2016. The SEER database tracks 5-year relative survival rates for breast cancer in the United States, based on how far cancer has spread. The SEER database, groups cancers into localized, regional, and distant stages:

- **Localized:** There is no sign that cancer has spread outside of the breast.
- **Regional:** Cancer has spread outside the breast to nearby structures or lymph nodes.
- **Distant:** Cancer has spread to distant parts of the body such as the lungs, liver, or bones.

Figure 2: 5-year survival rates for different cancer types:

SEER Stage	5-year Relative Survival Rate [23]	5-year relative survival rates for triple-negative breast cancer (TNBC) [24]	5-year relative survival rates for inflammatory breast cancer (IBC) [25]
Localized	99%	91%	X
Regional	86%	65%	56%
Distant	28%	12%	19%
All SEER stages combined	90%	77%	41%

(Compares women with the same type and stage of breast cancer to women in the overall population).

23 “Survival Rates for Breast Cancer.” Accessed August 26, 2021. https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html.

24 “Triple-Negative Breast Cancer | Details, Diagnosis, and Signs.” Accessed August 16, 2021. https://www.cancer.org/cancer/breast-cancer/about/types-of-breast-cancer/triple-negative.html.

25 “Inflammatory Breast Cancer | Details, Diagnosis, and Signs.” Accessed August 16, 2021. https://www.cancer.org/cancer/breast-cancer/about/types-of-breast-cancer/inflammatory-breast-cancer.html.

Diagnosis

Breast cancer screening and diagnostic techniques can be divided into three groups. The first one is the clinical breast exam – physical examination of breast and armpit lymph nodes. Checking for any lumps or abnormalities. The second is breast imaging modalities – mammography (based on low dose x-rays technology), MRI and MRE (based on magnetic fields technologies), and PET, PET-CT (based on nuclear medicine technologies). Under the third group falls biopsy- removal, and testing of a sample of breast cells using a specialized needle device.[26]

Treatment of different cancer subtypes

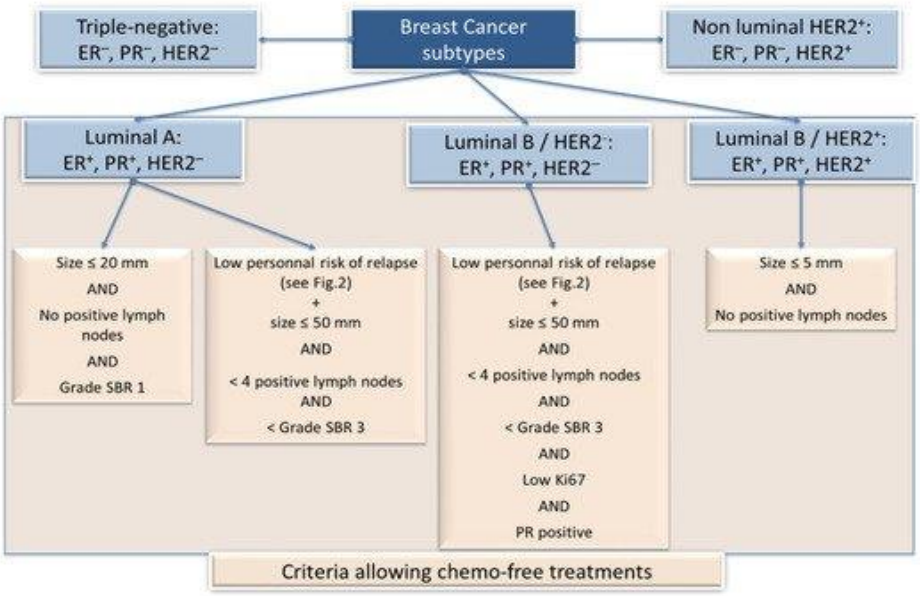

Figure 3: Breast cancer histological subtypes (**blue** color) and clinical criteria allowing to avoid chemotherapy (**orange** color). Among the five intrinsic subtypes (that either luminal, non-luminal, or triple-negative), three of them (the luminal ones) can be associated with chemo-free treatments following specific clinical criteria (tumor size, lymph nodes, Scarff–Bloom–Richardson (SBR) grade, Ki67, and progesterone receptor (PR) status).

Gene expression analysis and cancer

Cancer is the most common genetic disease that results from the accumulation of genetic alterations. These genetic alterations are divided into 2 major categories:

26 "Pharmaceutics | Free Full-Text | Current State of Breast Cancer Diagnosis, Treatment, and Theranostics | HTML." Accessed August 26, 2021. https://www.mdpi.com/1999-4923/13/5/723/htm.

1. Germline: germline alterations are found in the germ cell; hence, this type of alteration can be inherited from parents to offspring.

2. Somatic: somatic mutations are cellular alterations that are randomly acquired throughout the lifetime after exposure to various carcinogens or aging that damage the DNA.

Both germline and somatic alterations play pivotal roles in predisposing individuals to cancer and the initiation as well as to the progression of cancer. Therefore, genetic alterations could serve as effective biomarkers for early detection, monitoring, and prognosis of cancer.[27]

Particularly in breast cancer, as we saw earlier- genetic risk factors play a significant role. For the past two decades, genomic research has advanced remarkably, evolving from single-gene to whole-genome screening by using genome-wide association study and next-generation sequencing that contributes to big genomic data. Cancer precision medicine aims to provide the right dose of the right drug for the right patient at the right time, based on the genetic profiles of cancer and the individual.[28]

Today researchers can use genomic data by several open-access databases such as the GWAS catalog, NCI Genomic Data Commons, ClinVar, and ClinGen. Through those databases, it is possible to build a genetic profile, see levels of gene expression, find mutations, and create a specific treatment for breast cancer. For example, a great number of meta-analyses were carried out through international consortium networks with the purpose of identifying shared genetic susceptibilities among different populations for various complex diseases.[29] For breast cancer, the Breast Cancer Association Consortium (BCAC) and Asia Breast Cancer Consortium were started to assess the associations of common genetic variations with breast cancer.

After looking at the genetic risk factors of breast cancer we can infer that building a genetic profile and looking at down-regulated and upregulated genes could help to improve patient classification, predict prognosis, evaluate drug resistance, and identify drug targets. More importantly, current advancements in breast cancer genomics discoveries could be translated into therapeutic advances.[30]

---

27 “Breast Cancer: The Translation of Big Genomic Data to Cancer Precision Medicine - Low - 2018 - Cancer Science - Wiley Online Library.” Accessed August 26, 2021. https://onlinelibrary.wiley.com/doi/full/10.1111/cas.13463.

28 Peck RW. The right dose for every patient: a key step for precision medicine. Nat Rev Drug Discov. 2016;15:145-146.

29 “Breast Cancer: The Translation of Big Genomic Data to Cancer Precision Medicine - Low - 2018 - Cancer Science - Wiley Online Library.” Accessed August 26, 2021. https://onlinelibrary.wiley.com/doi/full/10.1111/cas.13463.

30 Sharma SV, Bell DW, Settleman J, Haber DA. Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer. 2007;7:169-181.

Relevant Datasets:

Link to the studies	Number of samples	Tissue/ cell type was profiled	Main findings
https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-52194/files/processed/ Title: E-GEOD-52194 - RNA-seq of 17 breast tumor samples of three different subtypes and normal human breast organoids samples	20	human breast cancer tissue	1.There are subtype specific differentially spliced genes and splice isoforms not previously recognized in human transcriptome. 2. It was found that differential expression of primary transcripts and promoter switching are significantly deregulated in breast cancer compared to normal breast.
https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9428/files/processed/ Title: High accuracy gene expression profiling of sorted cell subpopulations from breast cancer PDX model tissue	40	human breast cancer tissue	Each subpopulation shows distinct expression patterns suggestive of distinct function of genes within breast cancer tissue.
https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-8107/files/processed/ Title: Single-cell RNA sequencing of ovarian, colorectal and breast cancer	49	human breast, ovarian and colorectal cancer tissue	Profile single cells of colorectum, ovary and breast cancer. including Chromium 3' and 5' single-cell RNA-sequencing. .
https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-56022/files/processed/ Title: E-GEOD-56022 - Ligand-dependent genomic function of glucocorticoid receptor in triple-negative breast cancer	22	human breast cancer tissue	dexamethasone (Dex, a synthetic GC) regulated genes in triple-negative breast cancer (TNBC) cells are associated with drug resistance.
https://www.ebi.ac.uk/arrayexpress/files/E-GEOD-64590/E-GEOD-64590.processed.1.zip	18	human breast cancer tissue	Estrogens play an important role in breast cancer development and progression, where the two isoforms of the estrogen receptor (ERα and ERβ) are generally co-

Title: Estrogen Receptor Beta Impacts Hormone-Induced Alternative mRNA Splicing in Breast Cancer Cells			expressed and mediate the effects of these hormones in cancer cells.
https://www.ebi.ac.uk/arrayexpress/files/E-GEOD-71960/E-GEOD-71960.processed.1.zip Title: E-GEOD-71960 - Research resource: global identification of estrogen receptor β target genes in triple negative breast cancer cells	12	human breast cancer tissue	Identify all estrogen receptor beta target genes using RNA sequencing in MDA-MB-468 triple negative breast cancer cells engineered with inducible expression of full-length estrogen receptor beta.
ftp://ftp.sra.ebi.ac.uk/vol1/fastq/SRR164/009/SRR1648589/SRR1648589.fastq.gz Title: E-GEOD-63189 - Identification and functional characterization of long noncoding RNAs in breast cancer	25	human breast cancer tissue	1.The researchers create a comprehensive catalog of polyadenylated lncRNAs in MCF-7 cells, about half of which have not been annotated previously and about a quarter of which are estrogen-regulated. 2.characterization of long noncoding RNAs
https://www.ebi.ac.uk/arrayexpress/files/E-GEOD-71012/E-GEOD-71012.processed.1.zip Title: LIN28A modulates splicing and gene expression programs in breast cancer cells [RNA-Seq]	8	human breast cancer tissue	LIN28 regulates alternative splicing and steady state mRNA expression of genes implicated in aspects of breast cancer biology.
https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-59531/files/processed/ Title: TNFα Signaling Exposes Latent Estrogen Receptor Binding Sites in Breast Cancer Cells [GRO-seq]	12	human breast cancer tissue	The interplay between mitogenic and proinflammatory signaling pathways play key roles in determining the phenotypes and clinical outcomes of breast cancers.
https://www.ebi.ac.uk/arrayexpress/files/E-GEOD-45732/E-GEOD-45732.processed.1.zip Title: Gene expression analysis of breast cancer cell-lines	14	human breast cancer tissue	Recurrent mutations in histone modifying enzymes in multiple cancer types imply key roles in tumorigenesis. However, the functional relevance of these mutations remains unknown.

Results:

Our goal was to try classifying cancer types using gene expression analysis, we first started with general classification based on distances between the samples counts as presented in figure 4. Our analysis revealed a strong similarity between the different cancer types compare to the control group, supported by both clustering of the overall count's distances (figure 4), the top 20 highest variance genes (figure 6), and the top down-regulated genes (figure 8). Yet, PCA (figure 5) analysis has shown clustering almost identical to the sample's cancer types looking at both PC1 and PC2. PC1 distinguished between NBS (normal) and cancer, and PC2 distinguished between the cancer types. We also did not find many common genes between the top upregulated genes (figure 8) beside one gene.

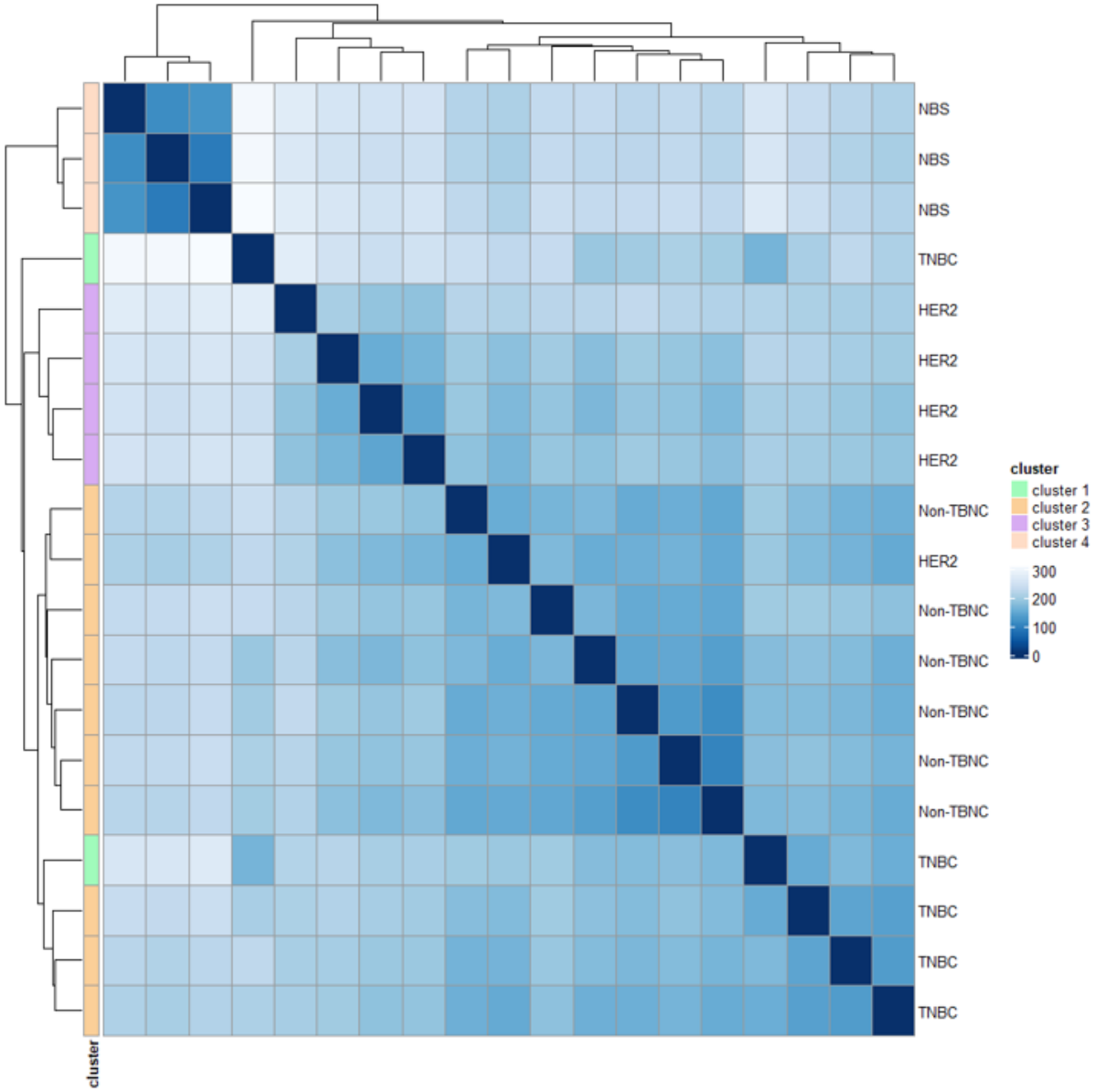

Figure 4: count's distance between NBS (normal), TNBC, Non-TNBC, and HER2 after applying the regularized-logarithm transformation.

We started with this analysis since we wanted to get a general perspective considering all the data available on the samples.

We ran regularized logarithm on the count's matrix, then created a distance matrix between each sample. The distance matrix is represented as a heatmap (figure 4). We also added a dendrogram clustering of the samples, we decided to examine a four-group clustering since there were four groups, to begin with (three cancer types and one control group). We can see that the NBS samples clustered together (cluster 4), and most of the HER2 samples also clustered together (Cluster 3, with one sample out of five that clustered with cluster 2). The TNBC and Non-TNBC types formed 2 mixed clusters (clusters 1 and 2, with one sample of HER2 present in cluster 2). Figure 4 allowed us to validate that there are similarities inside the sample groups (sanity check) and more interestingly to find out that there are similarities across them as mentioned above.

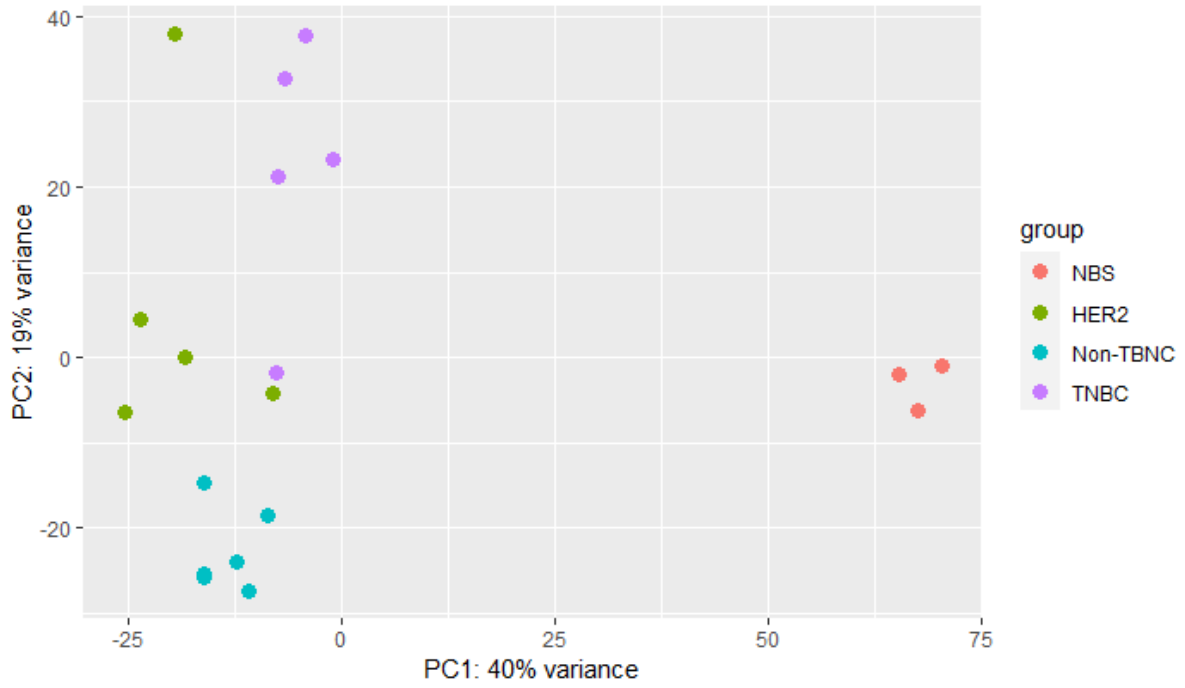

Figure 5: PCA based on the distance matrix between counts of NBS (normal), TNBC, Non-TNBC, and HER2 after applying the regularized logarithm transformation.

We ran PCA on the count's matrix after applying variance stabilizing transformation – VST function for negative binomial data with a dispersion-mean, to get a different perspective on the possible clustering, as we can see in figure 5. PC1 (x-axis) separated the control groups from the cancer groups. while PC2 (y-axis) suggested a division into three different clusters. We found a small mixture between clusters, with one HER2 sample and TNBC sample, but the rest of the samples clustered with the samples with the same type. To conclude, PCA provided classification similar up to almost identical to the original classification by cancer types.

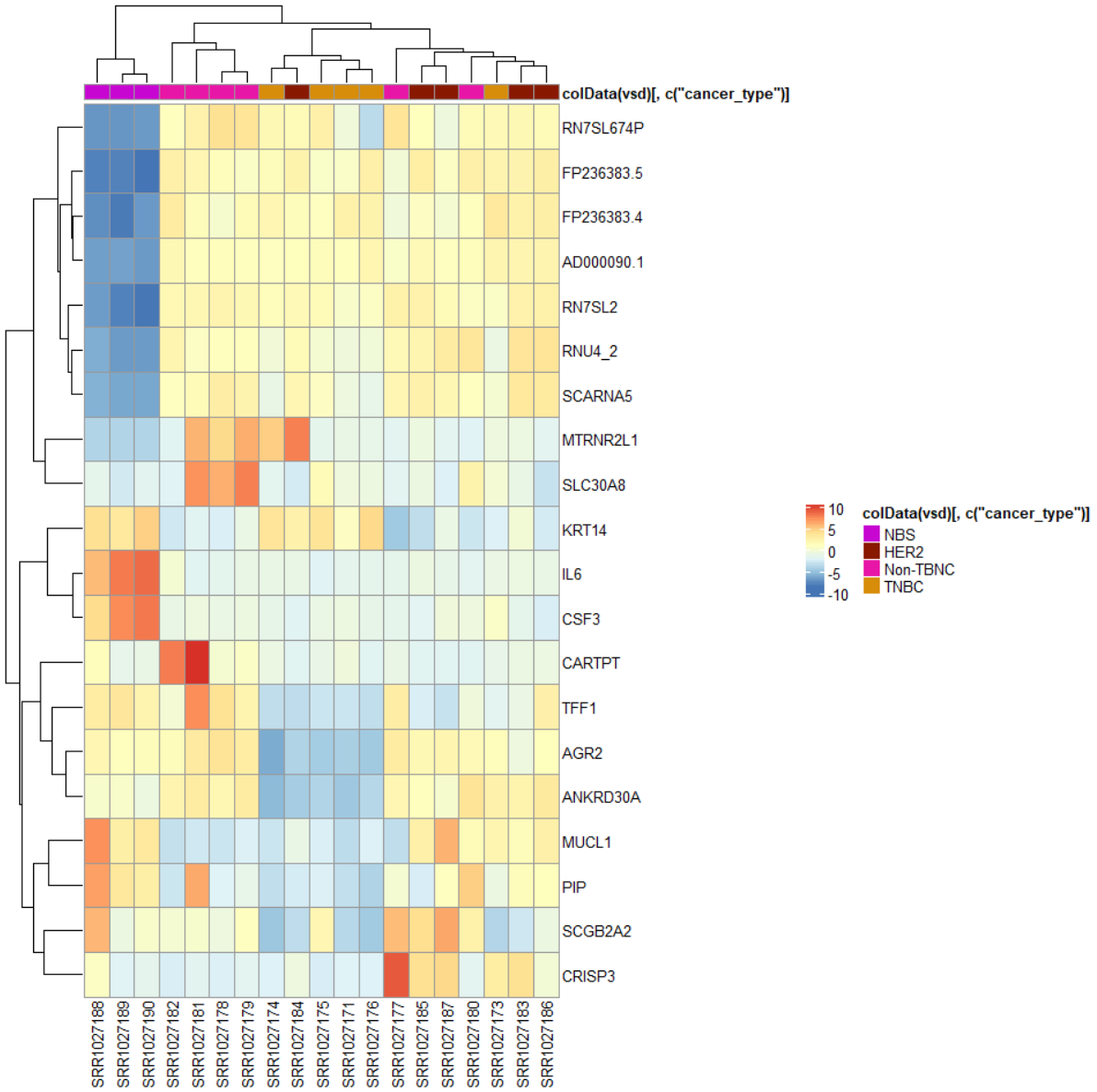

figure 6: Top 20 highest variance genes across NBS (normal), TNBC, Non-TNBC, and HER2.

We performed clustering for genes, choosing the 20 highest variance across samples. This analysis (figure 6) allowed focusing on the data in the garrulity of genes and not the entire sample. Results presented in a heatmap created from a distance matrix of the highest variance genes, after applying the VST function. From the dendrogram we can see that the control group remained as one cluster, four of the Non-TNBC samples clustered together and the rest of the samples mixed and formed another 2 clusters. We found that clustering based on this analysis is valid enough to distinguish between the normal samples and the cancer samples but not to differentiate between the types.

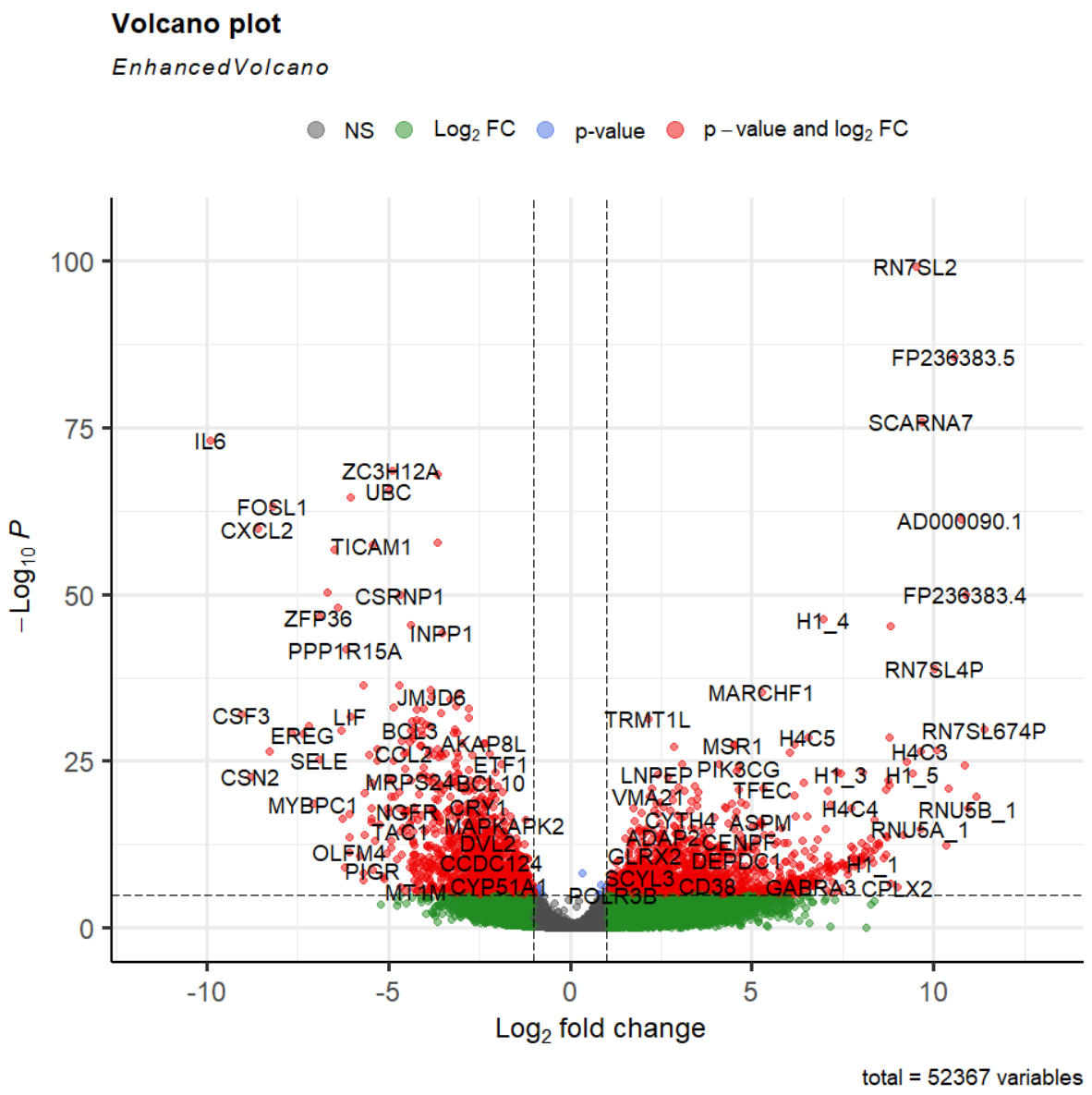

figure 7: P-value and log2 fold change volcano plot -Y-axis is p-value on a scale of -log10 and X-axis is fold change on a log2 scale

We decided to make a volcano plot (figure 7) of the log2foldchange to -log10p to get a visualization of the entire data set in order to find which genes upregulated and down-regulated the most. We used the enhanced volcano library. On the left side of figure 7, we can find the down-regulated genes and on the right side of the figure we can find the up-regulated genes, we can also see that most of the genes are trapped between -6<log2foldchange<7.5 and 0<-log10p<25 leaving us with only a few prominence genes. This plot allows us to assure the difference between the up/down-regulated genes to the other genes is significant and can be an appropriate way of investigating the data.

Figure 8: Up/Down-regulated genes due to cancer comparison:

	HER2 vs NBS	**TNBC vs NBS**	**Non.TBNC vs NBS**	**cancer vs normal (NBS)**	**common expressed genes**
Down-Regulation	CSN3 CSF3 IL6 NEFM CSN1S1 CSN2	IL6 CSN1S1 SCGB3A1 CSF3 MT1A CSN2	CSN1S1 CSN2 CSN3 IL6 CSF3 CXCL2	IL6 CSN3 CSN1S1 CSF3 CSN2 CXCL2	**CFS3** **CSN3** **IL6** **CSN1S1** **CSN2**
Up-Regulation	MTRNR2L1 H2BC1 TBX10 PAGE2 TACC1P1 GABRAS	IGKV1_35 RNU1_88P MTRNR2L1 FP236383.4 RN7SL674AP RN7SL507P	FP671120.5 AC104984.5 MTRNR2L1 RN7SL674AP SCARNA5 RNUSB_1	RNU5B_1 MTRNR2L1 SCARNA5 RN7SL674P RNU4_2 MT_TM	**MTRNR2L1** **RN7SL674AP** (Only in TNBC and Non-TNBC)

We applied filtering by p-value <= 0.05 and set the log fold change threshold to 1, and calculated the top up-regulated, and down-regulated genes (comparing the control group and the cancer groups). We ran 4 different tests: NBS vs. Non-TNBC, NBS vs. TNBC, NBS vs. HER2, and normal (NBS) vs. cancer (all samples). We summarized the top up-regulated and down-regulated genes in the figure 8 table (full test results can be found in the methods section). We saw a strong similarity between the cancer types in the down-regulated genes group as you can see in the rightmost column. The results have shown the commonly expressed genes - CFS3, CSN3, IL6, CSN1S1, CSN2 as down-regulated in all four comparisons. Moreover, we found that MTRNR2L1 is a common gene that was up-regulated in all four tests, and RN7SL674AP (Common only between TNBC and Non-TNBC).

Discussion:

A short overview of the common down-regulated genes:

- **CFS3:**
  CSF3 (Colony Stimulating Factor 3) is a Protein Coding gene. Related diseases are Neutropenia and Mucositis. It is related to PEDF Induced Signaling and Hematopoietic Stem Cell Differentiation-signaling pathways. Gene Ontology reveals that it is associated with enzyme binding and growth factor activity.

- **CSN3:**
  CSN3 (Casein Kappa) is a Protein Coding gene. Related diseases are include Milk Allergy and Food Allergy.

- **IL6**:
  IL6 (Interleukin 6) is a Protein Coding gene. Related diseases are Kaposi Sarcoma and Rheumatoid Arthritis, Systemic Juvenile. It is related to TNFR1 Pathway and PEDF Induced Signaling. Gene Ontology reveals that this gene includes signaling receptor binding and growth factor activity.

- **CSN1S1:**
  CSN1S1 (Casein Alpha S1) is a Protein Coding gene. Related diseases include Milk Allergy and Food Allergy. Gene Ontology reveals that this gene includes transporter activity.

- **CSN2:** This gene is a part of the beta-casein family. It can be divided into two types of casein protein, beta (encoded by this gene) and kappa, both are secreted in human milk. Beta and kappa casein proteins acting together form spherical micelles which bind within them important dietary minerals. Beta casein is the principal protein in human milk and the primary source of essential amino acids for a suckling infant.

A short overview of the common up-regulated gene:

- MTRNR2L1: MTRNR2L1 (MT-RNR2 Like 1) is a Protein Coding gene. Related diseases include Dystonia 23. An important paralog of this gene is MTRNR2L6.[31]

Interleukin-6 (IL-6)

Interleukin-6 (IL-6) affects the neoplastic process through its action on cancer cell adhesion, motility, proliferation, tumor-specific antigen expression, and thrombopoiesis. IL-6 trigger its activity by binding to a high-affinity receptor complex consisting of two membrane glycoproteins: the 80 kDa IL-6 a-receptor subunit (IL-6R)

---

31 "CSF3 Gene - GeneCards | CSF3 Protein | CSF3 Antibody." Accessed August 17, 2021. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CSF3.

and the 130 kDa signal-transducing protein (GP130).[32]

Our analysis has shown opposite results to known studies, as IL6 was downregulated in our analysis, and we found many articles pointing out that overexpression of interleukin-6 (IL-6) cytokine in the tumor microenvironment has been demonstrated in numerous tumors including breast cancer.[33] Tumor cells and tumor-associated fibroblasts are the major sources of IL-6 secretion in the tumor microenvironment. Several studies have demonstrated the immunopathogenic function of IL-6 and it is signaling in tumor growth, metastasis, and therapeutic resistance in breast cancer. Therefore, it seems that targeting IL-6 and/or its receptor in combination with other potent anticancer therapies may be a potent therapeutic approach for breast cancer therapy. We were also able to find a study regarding down-regulation of expression of interleukin-6 and its receptor results in growth inhibition of MCF-7 breast cancer cells. That suggests that although IL-6 inhibits cell proliferation in human colon carcinoma, melanoma cell lines, and M1 leukemia cells. As well as promotes cell proliferation of prostate carcinoma, colon carcinoma, leukemia, melanoma, and renal cell carcinoma. In human breast carcinoma, IL-6 has been shown to have contradictory effects on cell proliferation.

On one hand, it has been reported that IL-6 inhibited the proliferation of T47D breast cancer cells through STAT3 activation.[34] Furthermore, it has been shown that IL-6 additively inhibited the growth of MCF-7 breast cancer cells.[35]
On the other hand, it was reported that IL-6 induced multidrug resistance in MCF-7 cells and had no significant effect on cell proliferation.[36] Likewise, another research found that IL-6 alone did not significantly affect cell proliferation of MCF-7 cells, but significantly increased the estrone sulfate-induced proliferation when IL-6 and estrone sulfate were simultaneously added to the culture media. They hypothesized that IL-6 regulated the proliferation of breast cancer cells through estrogen production by steroid-catalyzing enzymes in the tissue.[37] In agreement with the last two findings mentioned above, it is hard to determine whether the addition of human recombinant IL-6 significantly stimulated or inhibited cell proliferation in MCF-7 breast cancer cells. Thus, the molecular mechanism of IL-6 action on MCF-7 cells remains unclear therefore these results suggest careful modulation of IL-6 and IL-6R expression of cells as a potential approach for breast cancer therapy.

---

32 Down-regulation of Expression of Interleukin-6 and its Receptor Results in Growth Inhibition of MCF-7 Breast Cancer Cells XIAN-PENG JIANG, DING CHENG YANG, ROBERT L. ELLIOTT, JONATHAN F. HEAD Anticancer Research Sep 2011, 31 (9) 2899-2906.

33 Dethlefsen C, Højfeldt G, Hojman P. The role of intratumoral and systemic IL-6 in breast cancer. Breast Cancer Res Treat. 2013 Apr;138(3):657-64. doi: 10.1007/s10549-013-2488-z. Epub 2013 Mar 27. PMID: 23532539.

34 Badache A, Hynes NE: Interleukin-6 inhibits proliferation and, in cooperation with an epidermal growth factor receptor autocrine loop, increases migration of T47D breast cancer cells. Cancer Res 61: 383-391, 2001. Abstract/FREE Full TextGoogle Scholar

35 Danforth DN Jr.., Sgagias MK: Interleukin-1 alpha and interleukin-6 act additively to inhibit growth of MCF-7 breast cancer cells in vitro. Cancer Res 53: 1538-1545, 1993. Abstract/FREE Full TextGoogle Scholar

36 Conze D, Weiss L, Regan PS, Bhushan A, Weaver D, Johnson P, Rincon M: Autocrine production of interleukin-6 causes multidrug resistance in breast cancer cells. Cancer Res 61: 8851-8858, 2001. Abstract/FREE Full TextGoogle Scholar

37 Honma S, Shimodaira K, Shimizu Y, Tsuchiya N, Saito H, Yanaihara T, Okai T: The influence of inflammatory cytokines on estrogen production and cell proliferation in human breast cancer cells. Endocr J 49: 371-377, 2002. PubMedGoogle Scholar

Further research

We think that in order to be able to classify breast cancer using gene expression analysis we need much more samples. Since we had only a few samples of each type, we don't have enough data to consider errors coming from abnormality and measurement errors. We believe that a larger number of samples would provide a more accurate and more distinguishing analysis revealing more precise clusters and similarities between different known types. We also believe that a timestamp of the cancer progression stat would have contributed a lot since we might be able not only to classify cancer types using gene expression but also classify the progression level. Furthermore, age and family history, lifestyle, and the rest of the factors that we reviewed in the introduction would provide more accurate and detailed analysis at least for the PCA direction that can combine all factors. We believe that investigating the down-regulated genes can be an interesting direction of research. In this research, we chose to focus on IL-6 but the rest of the genes that we found also have been proven to relate to breast cancer.

Conclusion

We saw that PCA was a good way to classify cancer types accordingly to the known types today. Moreover, we found out that there are strong similarities in counts expression between the down-regulated genes as opposed to normal breast cells (CFS3, CSN3, IL6, CSN1S1, CSN2) and MTRNR2L1 as a common upregulated gene. For a more general and accurate clustering system, we will need more samples and more general information about the patients. There is still a debate whether the IL6's expression levels an appropriate classifier for cancer and treatment. Therefore, we think that applying a different type of analysis might enlighten this unresolved problem.

Methods:

The data set was taken from array express named: E-GEOD-52194 - RNA-seq of 17 breast tumor samples of three different subtypes and normal human breast organoids samples.

Analysis was made using the library DESeq2. Firstly, we created DESeq data set (named dds) from the count's matrix (sorted.csv), after releveling the control group as the first level (NBS), we also created a coldata file containing information about the samples (coldata.csv –one column was the cancer type and the second classified the samples to normal vs caner). Using the coldata and count data and choosing design = ~ cancer_type (We chose cancer_type column as we wanted it to be investigated factor), we finally created the data set (dds). Then we filtered rows with no counts. We applied two transformations on the dds object:

-The variance stabilizing transformation (VST) for negative binomial data with a dispersion-mean trend implemented in the vst function. (Named vsd).

-The regularized-logarithm transformation implemented in rlog() function. (Named rld).

We called the transformation with blind = FALSE, which means that differences between cell lines and treatment (the variables in the design) will not contribute to the expected variance-mean trend of the experiment.

Figure 4:

We ran regularized logarithm transformation (using the DESeq2 - R library with rlog function) then created a distance matrix based on the different cancer types (TNBC, Non-TNBC, HER2) and the control group (denoted as NBS) counts. Then we visualized the distance matrix with the pheatmap function from the pheatmap library and using Complex heatmap and dendextend libraries we added a dendrogram clustering the samples into four clusters.

Figure 5:

We ran Principal component analysis using the plotPCA function on the vsd object.

Figure 6:

Using the order function, we choose the 20 highest variances across samples (from vsd). We visualized the results using pheatmap function.

Figure 7:

we created another DESeq data set (named dds) from the count's matrix (sorted.csv), this time choosing the condition column in the coldata file (coldata.csv), when creating

the data set (design = ~ condition) then we filtered rows with no counts. Using this DESeq data set and EnhancedVolcano library we were able to draw the figure.

Figure 8:

To create this table, we created 4 data set objects, one of them was the one mentioned above in Figure 4 section (that represents Cancer vs. NBS-normal). And the other three were built the same way to test Non-TNBC Vs. NBS, TNBC Vs. NBS and HER2 Vs. NBS. We performed filtering (of p-value < 0.05 and fold change threshold = 1). Then we sorted the results by log2 fold change to get the top up/down-regulated genes. We also exported the results for all four data sets analysis (to files named: resultsCombinedNormalVs.Cancer.csv, resultsNBSVs.Non-TNBC.csv, resultsNBSVs.TNBC.csv, and resultsNBSVs.HER2.csv respectively).

References:

“Am I Sick?” Breast Cancer Care. 23 February 2018. Archived from the Original on 25 October 2013. Retrieved 22 October 2013.,” n.d.

“BRCA2 Mutation-Associated Breast Cancers Exhibit a Distinguishing Phenotype Based on Morphology and Molecular Profiles from Tissue Microarrays - PubMed.” Accessed August 19, 2021. https://pubmed.ncbi.nlm.nih.gov/17197928/.

“Breast Cancer: Symptoms, Stages, Types and More.” Accessed August 18, 2021. https://www.healthline.com/health/breast-cancer.

“Breast Cancer: The Translation of Big Genomic Data to Cancer Precision Medicine - Low - 2018 - Cancer Science - Wiley Online Library.” Accessed August 14, 2021. https://onlinelibrary.wiley.com/doi/full/10.1111/cas.13463.

Brewer, Hannah R., Michael E. Jones, Minouk J. Schoemaker, Alan Ashworth, and Anthony J. Swerdlow. “Family History and Risk of Breast Cancer: An Analysis Accounting for Family Structure.” *Breast Cancer Research and Treatment* 165, no. 1 (August 2017): 193–200. https://doi.org/10.1007/s10549-017-4325-2.

“CSF3 Gene - GeneCards | CSF3 Protein | CSF3 Antibody.” Accessed August 17, 2021. https://www.genecards.org/cgi-bin/carddisp.pl?gene=CSF3.

“Dethlefsen C, Højfeldt G, Hojman P. The Role of Intratumoral and Systemic IL-6 in Breast Cancer. Breast Cancer Res Treat. 2013 Apr;138(3):657-64. Doi: 10.1007/S10549-013-2488-z. Epub 2013 Mar 27. PMID: 23532539. - Google Search.” Accessed August 9, 2021. https://www.google.com/search?q=Dethlefsen+C%2C+H%C3%B8jfeldt+G%2C+Hojman+P.+The+role+of+intratumoral+and+systemic+IL-6+in+breast+cancer.+Breast+Cancer+Res+Treat.+2013+Apr%3B138%283%29%3A657-64.+doi%3A+10.1007%2Fs10549-013-2488-z.+Epub+2013+Mar+27.+PMID%3A+23532539.&rlz=1C1GCEU_enIL957IL957&sxsrf=ALeKk004tTPjp7QNnuh0jdY9H40y0p787w%3A1630011220000&ei=U_8nYavBPM6ua4yJnHg&oq=Dethlefsen+C%2C+H%C3%B8jfeldt+G%2C+Hojman+P.+The+role+of+intratumoral+and+systemic+IL-6+in+breast+cancer.+Breast+Cancer+Res+Treat.+2013+Apr%3B138%283%29%3A657-64.+doi%3A+10.1007%2Fs10549-013-2488-z.+Epub+2013+Mar+27.+PMID%3A+23532539.&gs_lcp=Cgdnd3Mtd2l6EANKBAhBGABQ21VY21VgnVxoAHAAeACAAQCIAQCSAQCYAQCgAQKgAQHAAQE&sclient=gws-wiz&ved=0ahUKEwjrtv39yM_yAhVO1xoKHYwEBw8Q4dUDCA8&uact=5.

Dethlefsen, Christine, Grith Højfeldt, and Pernille Hojman. “The Role of Intratumoral and Systemic IL-6 in Breast Cancer.” *Breast Cancer Research and Treatment* 138, no. 3 (April 2013): 657–64. https://doi.org/10.1007/s10549-013-2488-z.

“Dietary Fat in Breast Cancer Survival - PubMed.” Accessed August 14, 2021. https://pubmed.ncbi.nlm.nih.gov/23701588/.

“Gene Expression Profiling in Breast Cancer and Its Effect on Therapy Selection in Early-Stage Breast Cancer.” Accessed August 7, 2021. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5648272/.

Hon, Jane Date C., Baljit Singh, Aysegul Sahin, Gang Du, Jinhua Wang, Vincent Y. Wang, Fang-Ming Deng, David Y. Zhang, Marie E. Monaco, and Peng Lee. “Breast Cancer Molecular Subtypes: From TNBC to QNBC.” *American Journal of Cancer Research* 6, no. 9 (2016): 1864–72.

“Hormone Replacement Therapy and the Risk of Breast Cancer | Nature Reviews Clinical Oncology.” Accessed August 13, 2021. https://www.nature.com/articles/nrclinonc.2011.110.

“Inflammatory Breast Cancer | Details, Diagnosis, and Signs.” Accessed August 16, 2021. https://www.cancer.org/cancer/breast-cancer/about/types-of-breast-cancer/inflammatory-breast-cancer.html.

Inoue, Kazushi, and Elizabeth A. Fry. “Aberrant Expression of Cyclin D1 in Cancer.” *Signal Transduction Insights* 4 (2015): 1–13. https://doi.org/10.4137/STI.S30306.

“Interleukin 6 Inhibits Proliferation and, in Cooperation with an Epidermal Growth Factor Receptor Autocrine Loop, Increases Migration of T47D Breast Cancer Cells - PubMed.” Accessed August 7, 2021. https://pubmed.ncbi.nlm.nih.gov/11196191/.

“Interleukin-1 Alpha and Interleukin-6 Act Additively to Inhibit Growth of MCF-7 Breast Cancer Cells in Vitro - PubMed.” Accessed August 21, 2021. https://pubmed.ncbi.nlm.nih.gov/8453620/.

Interpreting Signs and Symptoms. Lippincott Williams & Wilkins. 2007. Pp. 99–. ISBN 978-1-58255-668-0., n.d.

Jiang, Xian-Peng, Ding Cheng Yang, Robert L Elliott, and Jonathan F Head. “Down-Regulation of Expression of Interleukin-6 and Its Receptor Results in Growth Inhibition of MCF-7 Breast Cancer Cells.” *ANTICANCER RESEARCH*, 2011, 8.

Knight, Julia A., Jing Fan, Kathleen E. Malone, Esther M. John, Charles F. Lynch, Rikke Langballe, Leslie Bernstein, et al. “Alcohol Consumption and Cigarette Smoking in Combination: A Predictor of Contralateral Breast Cancer Risk in the WECARE Study.” *International Journal of Cancer* 141, no. 5 (September 1, 2017): 916–24. https://doi.org/10.1002/ijc.30791.

“Masjedi A, Hashemi V, Hojjat-Farsangi M, Ghalamfarsa G, Azizi G, Yousefi M, Jadidi-Niaragh F. The Significant Role of Interleukin-6 and Its Signaling Pathway in the Immunopathogenesis and Treatment of Breast Cancer. Biomed Pharmacother. - Google Search.” Accessed August 19, 2021. https://www.google.com/search?q=Masjedi+A%2C+Hashemi+V%2C+Hojjat-Farsangi+M%2C+Ghalamfarsa+G%2C+Azizi+G%2C+Yousefi+M%2C+Jadidi-Niaragh+F.+The+significant+role+of+interleukin-6+and+its+signaling+pathway+in+the+immunopathogenesis+and+treatment+of+breast+cancer.+Biomed+Pharmacother.&rlz=1C1GCEU_enIL957IL957&oq=Masjedi+A%2C+Hashemi+

V%2C+Hojjat-Farsangi+M%2C+Ghalamfarsa+G%2C+Azizi+G%2C+Yousefi+M%2C+Jadidi-Niaragh+F.+The+significant+role+of+interleukin-6+and+its+signaling+pathway+in+the+immunopathogenesis+and+treatment+of+breast+cancer.+Biomed+Pharmacother.&aqs=chrome..69i57.734j0j4&sourceid=chrome&ie=UTF-8.

"Pharmaceutics | Free Full-Text | Current State of Breast Cancer Diagnosis, Treatment, and Theranostics | HTML." Accessed August 12, 2021. https://www.mdpi.com/1999-4923/13/5/723/htm.

"Sex Hormones and Breast Cancer Risk in Premenopausal Women: Collaborative Reanalysis of Seven Prospective Studies." Accessed August 6, 2021. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4056766/.

Sharma, Sreenath V., Daphne W. Bell, Jeffrey Settleman, and Daniel A. Haber. "Epidermal Growth Factor Receptor Mutations in Lung Cancer." *Nature Reviews. Cancer* 7, no. 3 (March 2007): 169–81. https://doi.org/10.1038/nrc2088.

Siegel, Rebecca L., Kimberly D. Miller, and Ahmedin Jemal. "Cancer Statistics, 2017." *CA: A Cancer Journal for Clinicians* 67, no. 1 (January 2017): 7–30. https://doi.org/10.3322/caac.21387.

"Significance of Interleukin-6 (IL-6) in Breast Cancer (Review) | SpringerLink." Accessed August 26, 2021. https://link.springer.com/article/10.1007/s10549-006-9328-3.

Sun, Yi-Sheng, Zhao Zhao, Zhang-Nv Yang, Fang Xu, Hang-Jing Lu, Zhi-Yong Zhu, Wen Shi, Jianmin Jiang, Ping-Ping Yao, and Han-Ping Zhu. "Risk Factors and Preventions of Breast Cancer." *International Journal of Biological Sciences* 13, no. 11 (November 1, 2017): 1387–97. https://doi.org/10.7150/ijbs.21635.

"Survival Rates for Breast Cancer." Accessed August 15, 2021. https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html.

"The Right Dose for Every Patient: A Key Step for Precision Medicine - PubMed." Accessed August 13, 2021. https://pubmed.ncbi.nlm.nih.gov/26669674/.

"Transcriptomic Landscape of Breast Cancers through MRNA Sequencing | Scientific Reports." Accessed August 2, 2021. https://www.nature.com/articles/srep00264.

"Triple-Negative Breast Cancer | Details, Diagnosis, and Signs." Accessed August 7, 2021. https://www.cancer.org/cancer/breast-cancer/about/types-of-breast-cancer/triple-negative.html.

"U.S. Breast Cancer Statistics | Breastcancer.Org." Accessed August 17, 2021. https://www.breastcancer.org/symptoms/understand_bc/statistics.

Washbrook, Elinor. "Risk Factors and Epidemiology of Breast Cancer." *Women's Health Medicine*, Breast disorders, 3, no. 1 (January 1, 2006): 8–14. https://doi.org/10.1383/wohm.2006.3.1.8.

“What Is Breast Cancer? | Breast Cancer Definition.” Accessed August 1, 2021. https://www.cancer.org/cancer/breast-cancer/about/what-is-breast-cancer.html.

“האגודה למלחמה בסרטן - סרטן השד.” Accessed August 9, 2021. https://www.cancer.org.il/template/publications.aspx?maincat=12.