A single nucleotide variant on chromosome 5 residing within NDUFS4 distinguishes patients with basal-like human breast cancer.

Shahan Mamoor, MS
shahanmamoor@gmail.com
East Islip, NY 11730

Patients diagnosed with basal-like (also known as basal) subtype breast cancer face a more aggressive disease course and more dismal prognosis than patients diagnosed with luminal A and luminal B breast cancer molecular subtypes (1-4). We mined published microarray data (5, 6) to understand in an unbiased fashion the most distinguishing genetic and transcriptional features of tumors from patients with basal or basal-like subtype breast cancer. We identified a single nucleotide polymorphism, rs2124948, residing within NDUFS4, as among the most significant genetic differences in the tumors of patients with basal-like breast cancer. In a separate cohort of patients with basal subtype breast cancer, we observed transcriptome-wide differential expression of a NDUFS4 transcript. Analysis of patient survival data revealed that NDUFS4 primary tumor expression was correlated with recurrence-free survival in patients with basal-like and luminal B subtype breast cancer. Sequence variation in the NDUFS4 gene may be important in understanding differences in genetic background that favor development of basal subtype human breast cancer.

Keywords: basal subtype, basal-like breast cancer, NDUFS4, systems biology of breast cancer, targeted therapeutics in breast cancer, breast cancer, molecular subtype.

Classification of disease types and disease subtypes enables development of enhanced diagnostic capabilities, prognostic information, as well as a nuanced understanding of human disease that can support superior medical, surgical and radiological treatment of complex diseases such as cancer. Molecular subtypes in human breast cancer include luminal A, luminal B, basal or basal-like, and HER2+. Systems-level analyses, such as comparative tumor transcriptome analyses, can inform such approaches. Importantly, differences in genetic background that can influence disease frequency or progression in basal subtype (1, 2) or triple negative breast cancer (which shares significant overlap with basal subtype breast cancer (3, 4)) are difficult to discern without whole-genome sequencing data. However, in the case of whole genome sequencing, data indicating that changes at the level of the genome at single nucleotide resolution are accompanied by differences in gene expression, between molecular subtypes and at the level of the transcriptome, are still lacking, but could be ascertained based on combined analysis of microarray data that encompasses both transcript and SNP-level data. We mined published microarray data (5, 6) to identify the most significant genetic and transcriptional differences when comparing human breast cancer molecular subtypes. Integrative microarray dataset analysis has provided data illuminating genes whose sequence and expression defines basal subtype breast cancers, and we present evidence here.

Methods

We utilized datasets GSE10893 (5) and GSE87049 (6) for this global differential gene expression analysis of basal subtype human breast cancer in conjunction with GEO2R. GSE10893 was generated using Illumina Human-1 Genotyping BeadChip technology with $n=37$ basal-like breast tumors and $n=132$ primary tumors of the breast from patients diagnosed with breast cancer of different, non-basal molecular subtypes (luminal A, luminal B, HER2+, claudin, claudin-low, normal-like); analysis was performed using platform GPL8864. GSE87049 was generated using Affymetrix Human Gene 1.0 ST Array [transcript (gene) version] technology with $n=14$ basal-like breast tumors and $n=109$ breast tumors of other molecular subtypes (luminal A, luminal B, HER2-enriched, and normal-like); analysis was performed using platform GPL6244. The Benjamini and Hochberg method of p -value adjustment was used for ranking of differential expression but raw p -values were used to assess statistical significance of global differential expression. Log-transformation of data was auto-detected, and the NCBI generated category of platform annotation was used. A statistical test was performed to evaluate whether NDUF54 gene expression was significantly different between the tumors of patients with basal and non-basal molecular subtype breast cancer using a two-tailed t-test with Welch's correction. For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter tool (7) for correlation of NDUF54 mRNA expression levels with recurrence-free survival in $n=4929$ breast cancer patients, $n=953$ basal subtype breast cancer patients, $n=1809$ luminal A subtype breast cancer patients, $n=1353$ luminal B subtype breast cancer patients, $n=695$ HER2+ subtype breast cancer patients, and $n=119$ normal subtype breast cancer patients.

Results

SNP array genotyping identifies rs2124948 as a defining sequence variation in human basal breast cancer.

We used published microarray data (5) to compare the whole tumor genomes of 37 patients with basal or basal-like subtype breast cancer, with that of 132 patients with disease of other molecular subtype, including luminal A, luminal B, HER2+, claudin, claudin-low and normal-like, to identify single nucleotide polymorphisms that describe the most significant variations in the tumor genomes of patients with basal-like subtype breast cancer (BLBC).

SNP genotyping arrays utilize large differences in signal, after hybridization to DNA rather than RNA, to identify polymorphic sites in the genome at single nucleotide resolution. We identified a single nucleotide polymorphism (rs2124948) residing within the NDUF54 gene, on chromosome 5 (5q11.2), as

among the most significant genetic differences in the tumors of patients with basal subtype breast cancer as compared to that of other molecular subtypes. Polymorphism at rs2124948 was the 158th most significant difference, genome-wide, when comparing the tumor genomes of basal subtype patients with that of tumors from patients with other molecular subtypes (Rank: 158th out of 108250 total SNPs, equating to 99.9% (p -value: $3.97E-06$)).

NDUFS4 is differentially expressed in basal-like subtype human breast cancer.

Analysis of a second published microarray dataset, comparing tumor transcriptome data from a separate cohort of Mexican patients with basal subtype breast cancer (14 tumors from patients with basal-like breast cancer and 109 tumors from patients with breast cancer of other molecular subtype [luminal A, luminal B, HER2-enriched, and normal-like]) (6) revealed transcriptome-wide differential expression of a NDUFS4 transcript in patients with basal subtype human breast cancer (Table 2). When sorting each of the transcripts whose expression was measured globally, NDUFS4 expression ranked 9750 out of 33297 total transcripts, equating to 70.7% differential expression; this was statistically significant (p -value: $4.95E-02$).

NDUFS4 expression correlates with patient survival in basal and luminal B subtype breast cancer patients.

We asked whether primary tumor expression of NDUFS4 was correlated with patient survival outcomes in patients with breast cancer when analyzing outcomes based on PAM50 molecular subtype. We observed a correlation between primary tumor expression of NDUFS4 and recurrence-free survival (RFS) in patients with basal subtype human breast cancer (Figure 2; p -value: 0.043; hazard ratio: 1.25 (1.01-1.55)). NDUFS4 primary tumor expression was a negative prognostic indicator in patients with basal subtype cancer. Patients with low tumor expression of NDUFS4 possessed median RFS of 29 months, while patients with high tumor expression of NDUFS4 possessed median RFS of 21 months.

We also observed a correlation between primary tumor expression of NDUFS4 and recurrence-free survival (RFS) in patients with luminal B subtype human breast cancer (Figure 2; p -value: 0.0035; hazard ratio: 1.3 (1.09-1.54)). NDUFS4 primary tumor expression was a negative prognostic indicator in patients with luminal B subtype cancer. Luminal B patients with low tumor expression of NDUFS4 possessed median RFS of 44.5 months, while luminal B patients with high tumor expression of NDUFS4 possessed median RFS of 37 months.

NDUFS4 primary tumor expression was not correlated with recurrence-free survival in all patients (Figure 2; p -value: 0.95; hazard ratio: 1 (0.91-1.11)), in patients with luminal A subtype breast cancer (Figure 2; p -value: 0.66; hazard ratio: 1.05 (0.85-1.29)), in patients with HER2+ subtype breast cancer (Figure 2; p -value: 0.26; hazard ratio: 1.15 (0.9-1.46)), and in patients with normal subtype breast cancer (Figure 2; p -value: 0.17; hazard ratio: 0.63 (0.32-1.24)).

Discussion

We found through analysis of published microarray datasets (5, 6) that a genomic position marked by SNP rs2124948 was among the most significant sequence variations in the tumor genomes of patients with basal subtype human breast cancer when compared with that of patients with breast cancer of other molecular subtypes. rs2124948 resides within the NDUFS4 gene. Similar comparative tumor transcriptome study based on molecular subtype in a separate patient dataset revealed differential expression of a NDUFS4 transcript when comparing tumors of other molecular subtypes against that of tumors from patients of the basal subtype. Kaplan-Meier analysis of patient survival data revealed that NDUFS4 expression was negatively correlated with recurrence-free survival in patients with basal-like

1 and luminal B breast cancer. Changes in NDUF54 expression, or structural properties imparted to the
2 human genome by alternative allele usage at this polymorphic site, may be important events to understand
3 as part of the myriad processes that likely govern transformation and tumor maintenance in basal or
4 basal-like human breast cancer.
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

References

1. Rakha, E.A., Reis-Filho, J.S. and Ellis, I.O., 2008. Basal-like breast cancer: a critical review. *Journal of clinical oncology*, 26(15), pp.2568-2581.
2. Milioli, H.H., Tishchenko, I., Riveros, C., Berretta, R. and Moscato, P., 2017. Basal-like breast cancer: molecular profiles, clinical features and survival outcomes. *BMC medical genomics*, 10(1), pp.1-17.
3. Foulkes, W.D., Smith, I.E. and Reis-Filho, J.S., 2010. Triple-negative breast cancer. *New England journal of medicine*, 363(20), pp.1938-1948.
4. Dietze, E.C., Sistrunk, C., Miranda-Carboni, G., O'regan, R. and Seewaldt, V.L., 2015. Triple-negative breast cancer in African-American women: disparities versus biology. *Nature Reviews Cancer*, 15(4), pp.248-254.
5. Weigman, V.J., Chao, H.H., Shabalin, A.A., He, X., Parker, J.S., Nordgard, S.H., Grushko, T., Huo, D., Nwachukwu, C., Nobel, A. and Kristensen, V.N., 2012. Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast cancer research and treatment*, 133(3), pp.865-880.
6. Romero-Cordoba, S.L., Salido-Guadarrama, I., Rebollar-Vega, R., Bautista-Piña, V., Dominguez-Reyes, C., Tenorio-Torres, A., Villegas-Carlos, F., Fernández López, J.C., Uribe-Figueroa, L., Alfaro-Ruiz, L. and Hidalgo-Miranda, A., 2021. Comprehensive omic characterization of breast cancer in Mexican-Hispanic women. *Nature communications*, 12(1), pp.1-19.
7. Györfy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q. and Szallasi, Z., 2010. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast cancer research and treatment*, 123(3), pp.725-731.

Rank: 158
SNP: rs2124948
 p -value: 3.97E-06
 t : -4.76391563
 B : 4.052798

Table 1: A genomic position marked by rs2124948 (residing within the NDUFS4 gene) is among the most significant genetic differences in the tumor genomes of basal subtype human breast cancer when compared with that of other molecular subtypes.

The rank of global differential expression, SNP ID, the p -value with respect to differential expression transcriptome-wide, t , a moderated t -statistic, and B , the log-odds of differential expression between the groups compared are listed in this chart.

Figure 1: rs2124948 is significantly depleted in the primary tumors of patients with basal subtype human breast cancer.

A quantitative measure of NDUFS4 polymorphism at rs2124948 in the tumors of patients with basal subtype breast cancer (left) and in the tumors of patients with breast cancer of non-basal molecular subtype (right) is graphically depicted; the result of a statistical test evaluating significance of difference in this signal between the tumors of basal subtype patients and the tumors of patients with cancer of other molecular subtype is $p < 0.0001$. Depletion here indicates that in the basal subtype, probe signal was significantly decreased as compared to other molecular subtypes considered together, demonstrating a difference in allelic frequency in basal subtype patients at this genomic site.

Rank: 9750
probe: 8105318
p-value: 4.95E-02
t: -1.9839781
B: -4.7943
Gene: NDUFS4
Gene name: NADH:ubiquinone oxidoreductase subunit S4

Table 2: NDUFS4 is among the genes most differentially expressed, transcriptome-wide, in a separate cohort of patients with basal subtype breast cancer.

The rank of global differential expression, probe/transcript ID, the *p*-value with respect to differential expression transcriptome-wide, *t*, a moderated *t*-statistic, *B*, the log-odds of differential expression between the groups compared, the gene and gene name are listed in this chart.

Figure 2: NDUFS4 primary tumor expression correlates with recurrence-free survival in basal and luminal B subtype breast cancer patients.

Depicted in this Kaplan-Meier plot is the probability of recurrence-free survival for $n=4929$ breast cancer patients (above, left), $n=953$ basal subtype breast cancer patients (above, middle), $n=695$ HER2+ subtype breast cancer patients (above, right), $n=1809$ luminal A subtype breast cancer patients (below, left), $n=1353$ luminal B subtype breast cancer patients (below, middle), and $n=119$ normal subtype breast cancer patients (below, right) stratified into two groups, based on low or high expression of NDUFS4 in patient primary tumors. The log rank p -value denoting statistical significance of difference in recurrence-free survival when comparing the two groups, as well as hazard ratio for this comparison is listed above. Listed below is the number of patients at risk (number of patients alive) per interval, after stratification based on NDUFS4 expression; in the first interval, number at risk is number of patients alive; in each subsequent interval, number at risk is the number at risk less those who have expired or are censored.

Basal subtype

Low NDUFS4 expression: 29 months

High NDUFS4 expression: 21 months

Luminal B subtype

Low NDUFS4 expression: 44.5 months

High NDUFS4 expression: 37 months

Table 3: Median RFS is superior in basal subtype breast cancer patients with low primary tumor expression of NDUFS4.

The median RFS (recurrence-free survival) of $n=953$ basal subtype breast cancer patients and $n=1353$ luminal B subtype breast cancer patients, with low primary tumor expression of NDUFS4 and high primary tumor expression of NDUFS4 is listed in this chart.