A single nucleotide variant on chromosome 5 residing within MSH3 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, rs40408, residing within MSH3, 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 MSH3 transcript. Analysis of patient survival data revealed that MSH3 primary tumor expression was correlated with overall survival in patients with normal-like breast cancer. Sequence variation in the MSH3 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, BLBC, systems biology of breast cancer, targeted therapeutics in breast cancer, breast cancer, triple negative breast cancer.

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 MSH3 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 (8) for correlation of MSH3 mRNA expression levels with overall survival in $n=1879$ breast cancer patients, $n=431$ basal subtype breast cancer patients, $n=596$ luminal A subtype breast cancer patients, $n=439$ luminal B subtype breast cancer patients, $n=362$ HER2+ subtype breast cancer patients, and $n=51$ normal subtype breast cancer patients.

Results

SNP array genotyping identifies rs40408 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 (rs40408) residing within the MSH3 gene, on chromosome 5 (5q14.1), 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 rs40408 was the 58th 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: 58 out of 108250 total SNPs, equating to 99.95% (p -value: $6.07E-07$)).

MSH3 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 MSH3 transcript in patients with basal subtype human breast cancer (Table 2). When sorting each of the transcripts whose expression was measured globally, MSH3 expression ranked 2064 out of 33297 total transcripts, equating to 93.8% differential expression; this was statistically significant (p -value: $1.69E-05$).

MSH3 expression correlates with patient survival in normal-like subtype breast cancer patients.

We asked whether primary tumor expression of MSH3 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 MSH3 and overall survival (OS) in patients with normal-like subtype human breast cancer, and this correlation trended towards statistical significance (Figure 2; p -value: 0.08; hazard ratio: 0.4 (0.14-1.16)). MSH3 primary tumor expression was a positive prognostic indicator in patients with normal-like subtype cancer. Normal-like patients with low tumor expression of MSH3 possessed, on average, markedly shorter OS as compared to normal-like patients with high tumor expression of MSH3.

MSH3 primary tumor expression was not correlated with overall survival in all patients (Figure 2; p -value: 0.34; hazard ratio: 0.91 (0.76-1.1)), in patients with basal subtype breast cancer (Figure 2; p -value: 0.92; hazard ratio: 0.98 (0.67-1.44)), in patients with luminal A subtype breast cancer (Figure 2; p -value: 0.52; hazard ratio: 1.15 (0.75-1.76)), in patients with luminal B subtype breast cancer (Figure 2; p -value: 0.77; hazard ratio: 1.05 (0.74-1.5)), and in patients with HER2+ subtype breast cancer (Figure 2; p -value: 0.4; hazard ratio: 1.18 (0.8-1.74)).

Discussion

We found through analysis of published microarray datasets (5, 6) that a genomic position marked by SNP rs40408 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. rs40408 resides within the MSH3 gene. Similar comparative tumor transcriptome study based on molecular subtype in a separate patient dataset revealed differential expression of a MSH3 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 MSH3 expression was positively correlated with overall survival in patients with normal-like breast cancer. Changes in MSH3 expression may be important events to understand as part of the myriad processes that likely govern transformation and tumor maintenance in basal or basal-like human breast cancer.

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: 58
SNP: rs40408
p-value: 6.07E-07
t: -5.17921265
B: 5.752402

Table 1: A genomic position marked by rs40408 (residing within the MSH3 gene) is among the most significant genetic differences in the tumor genomes of basal subtype human breast cancer.

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: rs40408 is significantly depleted in the primary tumors of patients with basal subtype human breast cancer.

A quantitative measure of MSH3 polymorphism at rs40408 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: 2064
probe: 8106633
p-value: 1.69E-05
t: -4.4772582
B: 2.50448
Gene: MSH3
Gene name: mutS homolog 3

Table 2: MSH3 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: MSH3 primary tumor expression correlates with overall survival in normal subtype breast cancer patients.

Depicted in this Kaplan-Meier plot is the probability of overall survival for $n=1879$ breast cancer patients (above, left), $n=431$ basal subtype breast cancer patients (above, middle), $n=362$ HER2+ subtype breast cancer patients (above, right), $n=596$ luminal A subtype breast cancer patients (below, left), $n=439$ luminal B subtype breast cancer patients (below, middle), and $n=51$ normal subtype breast cancer patients (below, right) stratified into two groups, based on low or high expression of MSH3 in patient primary tumors. The log rank p -value denoting statistical significance of difference in overall 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 MSH3 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.