A long, intergenic non-coding RNA with differential expression in lymph node metastasis in human breast cancer.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY USA

Metastasis to the brain is a clinical problem in patients with breast cancer¹⁻³. Between the breast and the brain reside the secondary lymphoid organ, the lymph nodes. We mined published microarray data^{4,5} to compare primary and metastatic tumor transcriptomes for the discovery of genes associated with metastasis to the lymph nodes in humans with metastatic breast cancer. We found that long intergenic non-protein coding RNA 1521, LINC01521, was among the genes whose expression was most different in the lymph node metastases of patients with metastatic breast cancer as compared to primary tumors of the breast. Analysis of a separate microarray dataset revealed that LINC01521 was also differentially expressed in brain metastatic tissues. LINC01521 mRNA was present at increased quantities in lymph node metastases as compared to primary tumors of the breast. Importantly, expression of LINC01521 in primary tumors of the breast was correlated with patient overall survival, in lymph node positive patients but not in lymph node negative patients. Modulation of LINC01521 expression may be relevant to the biology by which tumor cells metastasize from the breast to the lymph nodes and the brain in humans with metastatic breast cancer.

Keywords: breast cancer, metastasis, brain metastasis, central nervous system metastasis, lymph node metastasis, long intergenic non-protein coding RNA 1521, LINC01521, systems biology of breast cancer, targeted therapeutics in breast cancer.

One report described a 34% incidence of central nervous system metastases in patients treated with trastuzumab for breast cancer². More recently, the NEfERT-T clinical trial⁶ which compared administration of either neratinib or trastuzumab in conjunction with paclitaxel demonstrated that in a randomized, controlled setting, in breast cancer patients treated with neratinib, not only was the incidence of central nervous system recurrence significantly lower, the time to central nervous system metastasis was significantly delayed as compared to patients administered trastuzumab⁶. The alarmingly high rate of central nervous system metastasis described, as well as data, both anecdotal² and from a randomized, controlled setting⁶ illustrating that treatment with trastuzumab may be associated with these events demands an enhanced understanding of the transcriptional makeup of brain and lymph node metastatic tissues to support identification of therapeutic targets, whether they are treatment related or not. The lymph nodes reside between the breast and the brain. We performed a global comparative analysis of lymph node metastatic tumor tissues in patients with metastatic breast cancer, compared to primary tumors and normal breast tissues^{4,5}. We discovered transcriptome-wide differential expression of the gene encoding long intergenic non-protein coding RNA 1521, LINC01521, in lymph node metastatic tissues of patients with metastatic breast cancer.

Methods

We used datasets GSE10893⁴ and GSE52604⁵ for this global differential gene expression analysis of brain metastatic breast cancer in conjunction with GEO2R. GSE10893 was generated using Agilent-011521 Human 1A Microarray G4110A technology with $n=71$ primary breast tumors and $n=7$ lymph node metastases from patients with breast cancer; analysis was performed using platform GPL887. GSE52604 was generated using Agilent-014850 Whole Human Genome Microarray 4x44K G4112F with $n=10$ normal breast tissues and $n=35$ brain metastases from patients with breast cancer; analysis was performed using platform GPL6480. 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 LINC01521 gene expression was significantly different between primary tumors of the breast and lymph node metastases in humans with breast cancer using a two-tailed t-test. For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter online tool⁷ for correlation of LINC01521 mRNA expression levels with overall survival in $n=230$ lymph node positive patients and $n=180$ lymph node negative breast cancer patients.

Results

We performed global comparative transcriptome analysis of metastatic tumor tissues of patients with metastatic breast cancer using published microarray data^{4,5} to describe the transcriptional landscape of metastasis in human breast cancer in an unbiased fashion and for the discovery of novel therapeutic targets.

LINC01521 is differentially expressed in the lymph node metastases of patients with metastatic breast cancer.

Through blind, systems-level analysis of published microarray data⁴, we identified long intergenic non-protein coding RNA 1521, encoded by LINC01521, as a differentially expressed gene in the lymph node metastatic tissues of humans with breast cancer (Table 1). When sorting each of the genes expressed in lymph node metastases based on significance of difference as compared to primary tumors of the breast in patients with breast cancer, LINC01521 ranked 239 out of 18034 total transcripts (Chart 1), equating to 98.7% differential expression. Differential expression of LINC01521 in the lymph

node metastases of patients with metastatic breast cancer was statistically significant (Chart 1; $p=2.95e-03$).

Differential gene expression analyses that utilize primary and metastatic tumor tissues to discover genes associated with metastasis can be challenging due to the relatively heterogeneous cellular composition and constitution of the tumor as compared to normal, untransformed (benign) tissues. Thus, to attempt to validate differential transcriptome-wide differential expression of LINC01521 in metastatic tissues in human breast cancer, we queried a second microarray dataset⁵, here comparing normal breast tissues rather than primary tumors of the breast, to brain metastases. Again, we identified LINC01521 as a differentially expressed gene in the metastatic tissues of humans with breast cancer, here, in the brain metastatic tissues of patients with breast cancer (Chart 2). When sorting each of the genes expressed in brain metastases based on significance of difference as compared to normal breast tissues, LINC01521 ranked 9125 out of 41093 total transcripts (Chart 2), equating to 77.8% differential expression. Differential expression of LINC01521 in the brain metastases of patients with metastatic breast cancer was statistically significant (Chart 2; $p=3.33e-03$). Thus, differential expression of LINC01521, transcriptome-wide, in the metastatic tissues of women with metastatic breast cancer was conserved across two independent microarray datasets, both in lymph node and brain metastases. This suggests that LINC01521 differential expression in human metastatic breast cancer is a *bona fide* biological process and that differential expression of LINC01521 in metastatic breast cancer is shared between the brain and the lymph nodes.

LINC01521 is expressed at higher levels in the lymph node metastases of patients with metastatic breast cancer.

We obtained exact mRNA expression levels for LINC01521, in primary tumors of the breast and in lymph node metastases of patients with metastatic breast cancer to determine direction and statistical significance of change in LINC01521 expression in lymph node metastatic tissues. We observed significantly increased expression of LINC01521 in the lymph node metastases of patients with breast cancer as compared to primary tumors of the breast: LINC01521 was expressed at 0.16 ± 0.32 arbitrary units (AU) in primary tumors of the breast, while it was expressed at 0.60 ± 0.62 AU in lymph node metastatic tissues (Figure 1). The difference in LINC01521 mRNA levels between primary tumors of the breast and lymph node metastases was statistically significant (Figure 1; $p=0.002774$).

LINC01521 expression is significantly correlated with survival outcomes in human breast cancer.

We performed Kaplan-Meier survival analysis⁷ in $n=230$ lymph node positive breast cancer patients and $n=180$ lymph node negative breast cancer patients, to evaluate whether LINC01521 tumor expression was correlated with survival outcomes in breast cancer and to evaluate whether lymph node status influenced correlation of LINC01521 tumor expression with survival outcomes in human breast cancer. We observed a correlation between primary tumor expression of LINC01521 and overall survival (OS) in lymph node positive breast cancer (Figure 2). Lymph node positive patients whose primary tumors expressed low levels of LINC01521 possessed median OS of 70.68 months, while lymph node positive patients whose tumors expressed high levels of LINC01521 possessed median OS of 40.15 months. This difference in OS based on LINC01521 tumor expression in lymph node positive patients with breast cancer was statistically significant (Figure 2, Chart 3; logrank p -value: 0.02; hazard ratio: 1.77 (1.09-2.89)). LINC01521 primary tumor expression was not correlated with overall survival outcomes in lymph node negative patients (Figure 2, Chart 3; logrank p -value: 0.91; hazard ratio: 1.04 (0.48-2.29)).

1 Thus, by mining published microarray data^{4,5} in an unbiased and systematic fashion, we identified
2 LINC01521 as among the genes whose expression was most different, transcriptome-wide, in the lymph
3 node metastases of patients with breast cancer, when compared to primary tumors; we observed
4 significantly increased expression of LINC01521 in lymph node metastases as compared to primary
tumors of the breast. Further, we found correlation between LINC01521 expression and overall survival.

5 **Discussion**

6 We identified here a non-coding RNA whose expression is significantly perturbed at the
7 transcriptome level in metastasis to the lymph nodes in humans with breast cancer, and provided evidence
8 that its expression is significantly correlated with overall survival in human breast cancer, in lymph node
positive patients but not in lymph node negative patients. LINC01521 is poorly characterized and its
potential functions in lymph node metastasis *in vivo* remain to be defined.

References

1. Lin, N.U., Amiri-Kordestani, L., Palmieri, D., Liewehr, D.J. and Steeg, P.S., 2013. CNS metastases in breast cancer: old challenge, new frontiers.
2. Bendell, J.C., Domchek, S.M., Burstein, H.J., Harris, L., Younger, J., Kuter, I., Bunnell, C., Rue, M., Gelman, R. and Winer, E., 2003. Central nervous system metastases in women who receive trastuzumab-based therapy for metastatic breast carcinoma. *Cancer*, 97(12), pp.2972-2977.
3. Tsukada, Y., Fouad, A., Pickren, J.W. and Lane, W.W., 1983. Central nervous system metastasis from breast carcinoma autopsy study. *Cancer*, 52(12), pp.2349-2354.
4. 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.
5. Salhia, B., Kiefer, J., Ross, J.T., Metapally, R., Martinez, R.A., Johnson, K.N., DiPerna, D.M., Paquette, K.M., Jung, S., Nasser, S. and Wallstrom, G., 2014. Integrated genomic and epigenomic analysis of breast cancer brain metastasis. *Plos one*, 9(1), p.e85448.
6. Awada, A., Colomer, R., Inoue, K., Bondarenko, I., Badwe, R.A., Demetriou, G., Lee, S.C., Mehta, A.O., Kim, S.B., Bachelot, T. and Goswami, C., 2016. Neratinib plus paclitaxel vs trastuzumab plus paclitaxel in previously untreated metastatic ERBB2-positive breast cancer: the NEfERT-T randomized clinical trial. *JAMA oncology*, 2(12), pp.1557-1564.
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: 239
Probe ID: 3160
p-value: 2.95e-03
t: -3.0677779
B: -1.7231
Gene: LINC01521
Gene name: long intergenic non-protein coding RNA 1521

Chart 1: LINC01521 is differentially expressed in lymph node metastatic breast cancer when comparing lymph node metastases to primary tumors of the breast.

The rank of global differential expression, the 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.

Rank: 9125
Probe ID: A_23_P57480
p-value: 3.33e-03
t: -3.0933264
B: -2.347498
Gene: LINC01521
Gene name: long intergenic non-protein coding RNA 1521

Chart 2: LINC01521 is differentially expressed in brain metastatic breast cancer when comparing brain metastases to normal breast tissues.

The rank of global differential expression, the 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 1: LINC01521 is expressed at significantly higher levels in the lymph node metastases of patients with metastatic breast cancer when compared to primary tumors of the breast.

The mRNA expression level of LINC01521 in primary tumors of the breast (left) and in lymph node metastases of women with metastatic breast cancer (right) is graphically depicted; the result of a statistical test evaluating significance of difference in LINC01521 expression between primary tumors of the breast and lymph node metastases, a p -value, is $p=0.002774$.

Figure 2: LINC01521 primary tumor expression is correlated with overall survival in patients with breast cancer, in lymph node positive patients but not in lymph node negative patients.

Depicted in these Kaplan-Meier plots is the probability of overall survival in $n=230$ lymph node positive patients (**left**) and $n=180$ lymph node negative patients (**right**), based on low or high expression of LINC01521 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 LINC01521 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.

Lymph node positive patients

Low LINC01521 expression: 70.68 months

High LINC01521 expression: 40.15 months

Chart 3: In lymph node positive breast cancer patients, median overall survival is inferior in patients with high primary tumor expression of LINC01521.

The median OS (overall survival) of patients with low primary tumor expression of LINC01521 and high primary tumor expression of LINC01521 is listed in this chart, for $n=230$ lymph node positive breast cancer patients.