Computational systems analysis of differential gene expression in brain metastatic breast cancer.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY USA

We utilized readily available computational tools (1-5) and published microarray data (6-9) to understand in an unbiased fashion the most distinguishing molecular features of differentially expressed genes in brain metastatic tissues from humans with metastatic breast cancer, and their genomic sites. We present our findings here, which reveal enrichment for function at the Golgi apparatus, an overlap of molecular signatures with development of the lip and the craniofacial structure, and demonstrate that the genomic sites of the most differentially expressed genes in metastasis to the brain in human breast cancer are enriched for binding sites for Barx2 and IKZF1, may be occupied by Aebp2 and the histone lysine methyltransferase KMT2H, and marked at the chromatin at H3K79me2 and H3K4me3.

Keywords: brain metastasis, brain metastatic breast cancer, metastatic breast cancer, central nervous system metastasis, systems biology of human breast cancer.

Metastasis to the central nervous system, including metastasis to the brain is a serious complication of invasive breast cancer in humans (10-12). Limited options exist for the treatment of CNS metastatic breast cancer (13-16). We used a transcriptome-level differential gene expression approach for the identification of differentially expressed genes in human brain metastatic breast cancer, and then studied the molecular nature of the genomic sites of these genes - occupancy at regulatory regions and the likely presence of specific chromatin marks - using basic computational tools.

Methods

For systems-level pathway analysis of enriched gene sets, including computational analysis of enrichment of consensus transcription factor motifs and nuclear factor binding sites at the genomic sites of conserved differentially expressed genes we used the Broad Institute / UC San Diego online tool Molecular Signatures Database (MSigDB) (1, 2). For additional systems-level functional molecular classification of differentially expressed genes in brain metastatic human breast cancer we used Reactome pathway database (3). For computational analysis of enrichment of chromatin marks at the genomic sites of conserved differentially expressed genes we used ENRICHr (4,5).

GSE10893 (6) was generated using Agilent-011521 Human 1A Microarray G4110A technology with $n=11$ primary breast tumors and $n=3$ brain metastases from patients with breast cancer; analysis was performed using platform GPL885. GSE52604 (7) 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. GSE125989 (8) was generated using Affymetrix Human Genome U133A 2.0 array technology with $n=16$ primary breast tumors and $n=16$ matched brain metastases from patients with breast cancer; analysis was performed using platform GPL571. GSE43837 (9) was generated using Affymetrix Human X3P Array technology with $n=19$ primary tumors of the breast and $n=19$ brain metastases from patients with breast cancer; analysis was performed using platform GPL6480.

Results

We performed basic discovery research using published microarray data (6-9) and computational tools (1-5) to understand the molecular nature of the transcriptional fingerprint of brain metastasis in human breast cancer.

We utilized global gene expression data from primary tumors of the breast and from metastases to the brain in human breast cancer (basal subtype) to identify the most significant gene expression changes, transcriptome-wide, in brain metastatic breast cancer. Among these genes, we then assembled a list of genes whose expression was up-regulated in metastasis to the brain, and whose transcriptome-level differential expression could be validated in a separate microarray dataset, in metastasis to the brain or otherwise in human breast cancer. This list of genes was utilized as a unique set alongside basic computational tools including the Molecular Signatures Database to identify any distinguishing molecular features or “signatures”, Reactome pathway database analysis, and ENRICHr evaluation of chromatin modification.

The transformative potential of this gene set is made obvious by the identification of a set of genes transcriptionally activated by mutant Kras in the Kras2LA murine lung cancer model “SWEET_LUNG_CANCER_KRAS_UP” (Figure 1; $p=0.0000018$), including carbonic anhydrase 8 (CA8), ERH mRNA splicing and mitosis factor (ERH), and GDP-L-fucose synthase (GFUS), as well as a set of genes whose expression is down-regulated upon depletion of ALK in a medulloblastoma cell line (ALK_DN.V1_DN) ($p=8.98E-07$), including the VPS37B subunit of ESCRT-I (VPS37B) and proline dehydrogenase 1 (PRODH).

We utilized for this analysis genes whose expression was up-regulated in brain metastasis (as opposed to down-regulated). Interestingly, we observed enrichment for a set of genes whose expression is down-regulated in basal or basal-like breast cancer (SMID_BREAST_CANCER_BASAL_DN), including GPC3, NR2F1, and ATP1A2. It appears that at least some of the genes that are up-regulated at the transcriptome-level in metastasis to the brain had previously been down-regulated between transition from normal breast to primary tumor of the breast in basal-like breast cancer ($p=4.96E-13$).

We observed enrichment of multiple gene sets describing function in craniofacial development, including development of the lip - “HP_ABNORMALITY_OF_UPPER_LIP” ($p=1.76E-07$), HP_ABNORMAL_LIP_MORPHOLOGY” ($p=1.41E-06$), and “HP_ABNORMALITY_OF_UPPER_LIP_VERMILLION” ($p=1.75E-05$). These genes included CNTNAP2, CANT1, and ABCC8. Other gene sets describing perturbations in craniofacial development include “HP_ABNORMAL_FACIAL_SHAPE” ($p=2.00E-05$), “HP_LONG_PHILTRUM” ($p=1.45E-05$), and “HP_ABNORMALITY_OF_THE_NASAL_ALAE” ($p=1.25E-05$).

The binding sites of two separate transcription factors - Barx2 (BARX2_TARGET_GENES; $p=1.23E-08$) and IKZF1, also known as Lyf1 (LYF1_01; $p=4.08E-05$) were found to be enriched for at the genomic sites of the most differentially expressed genes in metastasis to the brain in human breast cancer. Barx2 is described to be important for craniofacial development, with both activating and repressive transcriptional activity (17), and is expressed during murine development in cells residing within indentations surrounding the maxillo-nasal groove (17). Barx2 has been reported to function in concert with the estrogen receptor alpha (ESR α) and co-occupy ESR α genomic targets in breast cancer cells (18). Genes that we predict to be Barx2 genomic targets include TOM1L2, DLGAP4, and PDK2, and genes predicted to be IKZF1 genomic targets include GPC3, DMPK, and ALKBH6. A third transcription factor-like protein, ZNF592, was also predicted to target the promoter region of many genes in this set (ZNF592_TARGET_GENES; $p=5.32E-08$) (19).

We also observed predicted enrichment of other nuclear factors at the genomic sites of the most differentially expressed genes in metastasis to the brain in human breast cancer, most probably at the regulatory regions of these genes. This included putative occupancy by KMT2H (Ash1l) (ASH1L_TARGET_GENES; $p=5.76E-06$) and Aebp2 (AEBP2_TARGET_GENES; $p=2.66E-05$). MSigDB also predicted enrichment of binding of methyl-binding domain-containing proteins, predicted by data evaluating gene expression changes following simultaneous depletion of all three MBD proteins MBD1, MBD2, and MeCP2 in HeLa cells (LOPEZ_MBD_TARGETS; $p=3.05E-07$), including genes STAP2, PFDN6, and EOLA2.

_____Enrichment of genes with function at or localized to the Golgi apparatus was observed (GOCC_GOLGI_APPARATUS) (Figure 1; $p=1.89E-05$), including genes TRAPPC9, NECAB1, and NCAN.

A separate computational tool (Reactome) detected pathway enrichment of Rab regulation of trafficking - trafficking within the cell (Figure 2; $p=2.65E-03$), including genes RAB5C, NIBP and GDI1.

Finally, we used Enrichr to understand chromatin modification at the genomic sites of the most differentially expressed genes in human brain metastatic breast cancer in an unbiased manner. We observed predicted enrichment of two separate chromatin marks at the genomic sites of differentially expressed genes in CNS metastatic breast cancer: H3K4me3 -at genes including TIMM17B, TMEM205, and ALDH16A1, as well as H3K79me2, at genes including PSENEN, TFPT, and CXorf40B (Figure 3).

Discussion

We identified here gene signatures which describe the molecular nature of the most differentially expressed genes in metastasis to the brain in humans, in addition to nuclear proteins that might reside at their genomic regions, and chromatin marks which might decorate their sequences. We observed enrichment of the H3K79me2 mark on a substantial fraction of genes in the gene set we identified. Enhanced di-methylation at lysine 79 of histone H3 is a feature of gene expression in mixed lineage leukemia (MLL) fusion-driven leukemias in mice and in humans, including MLL-AF4 acute lymphoblastic leukemias (ALL) (20-22), and the H3K79 methyltransferase Dot1l is required for transformation of the bone marrow in murine leukemia driven by both MLL-AF9 (23-25) and MLL-AF10 (26). It would be of value to evaluate the viability of brain metastases in murine xenografts or of human brain metastatic cell lines *in vitro* in the presence of the Dot1L inhibitor, pinometostat (27).

References

1. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), pp.15545-15550.
2. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P., 2015. The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6), pp. 417-425
3. Fabregat, A., Sidiropoulos, K., Viteri, G., Marin-Garcia, P., Ping, P., Stein, L., D'Eustachio, P. and Hermjakob, H., 2018. Reactome diagram viewer: data structures and strategies to boost performance. *Bioinformatics*, 34(7), pp.1208-1214.
4. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R. and Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14(1), p.128.
5. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A. and McDermott, M.G., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1), pp.W90-W97.
6. 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.
7. 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.
8. Iwamoto, T., Niikura, N., Ogiya, R., Yasojima, H., Watanabe, K.I., Kanbayashi, C., Tsuneizumi, M., Matsui, A., Fujisawa, T., Iwasa, T. and Shien, T., 2019. Distinct gene expression profiles between primary breast cancers and brain metastases from pair-matched samples. *Scientific reports*, 9(1), pp.1-8.
9. McMullin, R.P., Wittner, B.S., Yang, C., Denton-Schneider, B.R., Hicks, D., Singavarapu, R., Moulis, S., Lee, J., Akbari, M.R., Narod, S.A. and Aldape, K.D., 2014. A BRCA1 deficient-like signature is enriched in breast cancer brain metastases and predicts DNA damage-induced poly (ADP-ribose) polymerase inhibitor sensitivity. *Breast Cancer Research*, 16(2), pp.1-10.
10. Altundag, K., Bondy, M.L., Mirza, N.Q., Kau, S.W., Broglio, K., Hortobagyi, G.N. and Rivera, E., 2007. Clinicopathologic characteristics and prognostic factors in 420 metastatic breast cancer patients with central nervous system metastasis. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 110(12), pp.2640-2647.
11. Arslan, U.Y., Oksuzoglu, B., Aksoy, S., Harputluoglu, H., Turker, I., Ozisik, Y., Dizdar, O., Altundag, K., Alkis, N. and Zengin, N., 2011. Breast cancer subtypes and outcomes of central nervous system metastases. *The Breast*, 20(6), pp.562-567.

12. Rostami, R., Mittal, S., Rostami, P., Tavassoli, F. and Jabbari, B., 2016. Brain metastasis in breast cancer: a comprehensive literature review. *Journal of neuro-oncology*, 127(3), pp.407-414.
13. Murthy, R.K., Loi, S., Okines, A., Paplomata, E., Hamilton, E., Hurvitz, S.A., Lin, N.U., Borges, V., Abramson, V., Anders, C. and Bedard, P.L., 2020. Tucatinib, trastuzumab, and capecitabine for HER2-positive metastatic breast cancer. *New England Journal of Medicine*, 382(7), pp.597-609.
14. Bachelot, T., Romieu, G., Campone, M., Diéras, V., Cropet, C., Dalenc, F., Jimenez, M., Le Rhun, E., Pierga, J.Y., Gonçalves, A. and Leheurteur, M., 2013. Lapatinib plus capecitabine in patients with previously untreated brain metastases from HER2-positive metastatic breast cancer (LANDSCAPE): a single-group phase 2 study. *The lancet oncology*, 14(1), pp.64-71.
15. Niwińska, A., Murawska, M. and Pogoda, K., 2010. Breast cancer brain metastases: differences in survival depending on biological subtype, RPA RTOG prognostic class and systemic treatment after whole-brain radiotherapy (WBRT). *Annals of Oncology*, 21(5), pp.942-948.
16. Mahmoud-Ahmed, A.S., Suh, J.H., Lee, S.Y., Crownover, R.L. and Barnett, G.H., 2002. Results of whole brain radiotherapy in patients with brain metastases from breast cancer: a retrospective study. *International Journal of Radiation Oncology* Biology* Physics*, 54(3), pp.810-817.
17. Jones, F.S., Kiuoussi, C., Copertino, D.W., Kallunki, P., Holst, B.D. and Edelman, G.M., 1997. Barx2, a new homeobox gene of the Bar class, is expressed in neural and craniofacial structures during development. *Proceedings of the National Academy of Sciences*, 94(6), pp.2632-2637.
18. Stevens, T.A. and Meech, R., 2006. BARX2 and estrogen receptor- α (ESR1) coordinately regulate the production of alternatively spliced ESR1 isoforms and control breast cancer cell growth and invasion. *Oncogene*, 25(39), pp.5426-5435.
19. Nicolas, E., Poitelon, Y., Chouery, E., Salem, N., Levy, N., Mégarbané, A. and Delague, V., 2010. CAMOS, a nonprogressive, autosomal recessive, congenital cerebellar ataxia, is caused by a mutant zinc-finger protein, ZNF592. *European journal of human genetics*, 18(10), pp.1107-1113.
20. Guenther, M.G., Lawton, L.N., Rozovskaia, T., Frampton, G.M., Levine, S.S., Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E. and Young, R.A., 2008. Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. *Genes & development*, 22(24), pp.3403-3408.
21. Krivtsov, A.V., Feng, Z., Lemieux, M.E., Faber, J., Vempati, S., Sinha, A.U., Xia, X., Jesneck, J., Bracken, A.P., Silverman, L.B. and Kutok, J.L., 2008. H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer cell*, 14(5), pp.355-368.
22. Bernt, K. M., Zhu, N., Sinha, A. U., Vempati, S., Faber, J., Krivtsov, A. V., ... & Armstrong, S. A. (2011). MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer cell*, 20(1), 66-78.
23. Chang, M.J., Wu, H., Achille, N.J., Reisenauer, M.R., Chou, C.W., Zeleznik-Le, N.J., Hemenway, C.S. and Zhang, W., 2010. Histone H3 lysine 79 methyltransferase Dot1 is required for immortalization by MLL oncogenes. *Cancer research*, 70(24), pp.10234-10242.
24. Jo, S.Y., Granowicz, E.M., Maillard, I., Thomas, D. and Hess, J.L., 2011. Requirement for Dot1l in murine postnatal hematopoiesis and leukemogenesis by MLL translocation. *Blood, The Journal of the American Society of Hematology*, 117(18), pp.4759-4768.

- 1
- 2 25. Nguyen, A.T., Taranova, O., He, J. and Zhang, Y., 2011. DOT1L, the H3K79 methyltransferase, is
- 3 required for MLL-AF9-mediated leukemogenesis. *Blood, The Journal of the American Society of*
- 4 *Hematology*, 117(25), pp.6912-6922.
- 5
- 6 26. Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G. and Zhang, Y., 2005.
- 7 hDOT1L links histone methylation to leukemogenesis. *Cell*, 121(2), pp.167-178.
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
27. Stein, E.M., Garcia-Manero, G., Rizzieri, D.A., Tibes, R., Berdeja, J.G., Savona, M.R.,
Jongen-Lavrenic, M., Altman, J.K., Thomson, B., Blakemore, S.J. and Daigle, S.R., 2018. The
DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult
acute leukemia. *Blood, The Journal of the American Society of Hematology*, 131(24),
pp.2661-2669.

Gene Set	k/K	p-value
SWEET_LUNG_CANCER_KRAS_UP	9/481	1.80E-06
ALK_DN.V1_DN	6/139	8.98E-07
SMID_BREAST_CANCER_BASAL_DN	17/700	4.96E-13
HP_ABNORMALITY_OF_UPPER_LIP	12/754	1.76E-07
HP_ABNORMAL_LIP_MORPHOLOGY	12/919	1.41E-06
HP_ABNORMALITY_OF_UPPER_LIP_VERMILLION	7/350	1.75E-05
HP_ABNORMAL_FACIAL_SHAPE	11/1001	2.00E-05
HP_LONG_PHILTRUM	7/340	1.45E-05
HP_ABNORMALITY_OF_THE_NASAL_ALAE	9/612	1.25E-05
BARX2_TARGET_GENES	19/1723	1.23E-08
LYF1_01	6/271	4.08E-05
ZNF592_TARGET_GENES	19/1890	5.32E-08
ASH1L_TARGET_GENES	14/1446	5.76E-06
AEBP2_TARGET_GENES	11/1033	2.66E-05
LOPEZ_MBD_TARGETS	13/954	3.05E-07
GOCC_GOLGI_APPARATUS	14/1608	1.89E-05

Table 1: Gene sets that comprise part of a molecular signature of brain metastasis in human breast cancer.

Where k denotes the number of genes in the signature we defined and where K denotes the total number of genes in the signature the Molecular Signatures database identified as overlapping.

Reaction	p-value
Rab regulation of trafficking	2.65E-03

Table 2: Reactome pathway analysis highlights regulation of cellular transport involving Rab.

Chromatin modification	p-value
H3K4me3 spleen	0.0046
H3K4me3 MEL cell line	0.0095
H3K4me3 olfactory bulb	0.0215
H3K4me3 limb	0.0425
H3K4me3 bone marrow macrophage	0.0425
H3K4me3 Caco-2	0.0472
H3K79me2 astrocyte	0.0102
H3K79me2 T-cell acute lymphoblastic leukemia	0.0102
H3K79me2 fibroblast of lung	0.0102
H3K79me2 K562	0.0425
H3K79me2 endothelial cell of umbilical vein	0.0425
H3K79me2 fibroblast of dermis	0.0425

Table 3: ENCODE predicts that in metastasis to the brain, the genomic regions of many of the genes up-regulated in human brain metastatic breast cancer are marked by H3K4me3 and H3K79me2.

In each of these cellular contexts, the promoter regions of a statistically significant number of genes in the set we derived are methyl marked at either H3K4 or H3K79.