A shared transcriptional landscape between human breast cancers and breast cancers induced by genetic modeling of viral infection in mice.

Shahan Mamoor, MS¹
¹shahanmamoor@gmail.com
East Islip, NY 11730

We recently demonstrated, through analysis of published transcriptome data from human and murine breast cancer that eight genes whose expression was most different in tumors that arise in the mouse breast following transgenic expression of the SV40 large T-antigen or modeling of integration following infection with multiple mammary tumor viruses, including CD36, ASPM, PDGFD and CEP55 were also among the genes most differentially expressed in the primary tumors of humans with breast cancer¹. We report here that these transcriptional similarities are a much larger phenomena, with 45-53% of the top one hundred most differentially expressed genes in murine breast cancer resulting from genetic modeling of viral infection with the murine mammary tumor virus MMTV or transgenic expression of SV40 large T also among the most differentially expressed genes in human breast cancer. Despite inherent differences in transcription due to species divergence, blind whole transcriptome²⁻⁴ analyses reveals striking transcriptional similarity in the landscape of human breast cancer and murine breast cancers resulting from genetic modeling of viral infection, providing further evidence to support the hypothesis that one or more viral agents with transformation capacity may contribute to the development of human breast cancer.

Keywords: breast cancer, breast cancer etiology, transformation, virally-induced transformation.

1 Familial, or inherited breast cancer caused by mutations in genes including BRCA1, p53
2 and RAD51C are found in 5-10% of patients with breast cancer⁵⁻⁹. In the remaining patients
3 with breast cancer, which in the United States numbered 268,600 in total in 2019¹⁰, the
4 etiology, or cause of the disease remains unknown. It has been suggested that viral infection
5 can confer susceptibility to or cause human breast cancer but evidence to support this claim
6 has thus far been largely limited to data indicating the presence of viral transcripts or viral
7 antigens in tissues from patients with breast cancer, with a high degree of variability based on
8 virus evaluated and laboratory from which the study was conducted¹¹⁻²⁷. Demonstration of a
9 pathology resulting from an infectious source requires fulfillment of Koch's postulate²⁸; It has
10 been demonstrated that viral infection or transgenic expression of a viral oncoprotein can result
11 in breast cancer in mice²⁹⁻³¹; while performing such experimentation in humans is not feasible
12 due to ethical concerns, data connecting these findings to human breast cancer³² are limited
13 relative to our understanding of the genomic landscape (e.g., copy number variation and
14 mutation status) of human breast cancer³³⁻³⁷. We provide evidence here from analysis of
15 differentially expressed genes using published microarray data that the overlap between the
16 most significant transcriptional features of human breast cancer and breast tumors resulting
17 from genetic modeling of viral infection in mice is pervasive; the extent of this overlap is likely
18 larger than previously understood.

19 Methods

20 We utilized datasets GSE124648², GSE109169³, GSE23938⁴ and for this differential
21 gene expression analysis of female breast cancer. GSE124648 was generated using Affymetrix
22 Human Genome U133A Array technology with $n=10$ normal breast tissue and $n=126$ tumors of
23 the breast and; analysis was performed using platform GPL96. GSE109169 was generated
24 using Affymetrix Human Exon 1.0 ST Array technology with $n=25$ normal breast tissue and $n=25$
25 tumors of the breast and; analysis was performed using platform GPL5175. The tissues from
26 this dataset are paired tissues (25 tumors matching 25 breast tissues from 25 patients).
27 GSE23938 was generated using Affymetrix Mouse Genome 430A 2.0 Array technology with $n=5$
28 normal breast tissue and $n=6$ breast tumors from MMTV-PyMT mice, $n=4$ breast tumors from
MMTV-Wnt mice, $n=5$ breast tumors from SV40 large T-antigen-expressing mice (C3TA_g), $n=4$
breast tumors from MMTV-C-Myc mice, $n=5$ breast tumors from MMTV-HRas mice, and $n=5$
breast tumors from BRCA1 $-/-$ p53 $-/+$ transgenic mice; analysis was performed using platform
GPL8321. 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.

Results

We performed blind, comparative transcriptome analysis using multiple published microarray datasets⁷⁻⁹ to identify conserved differentially expressed genes when comparing primary tumor to normal breast tissue, in humans, in mice with breast cancers induced by genetic modeling of integration events or transgenic expression of oncogenes during viral infection, and in mice genetically engineered to model familial breast cancer.

Nearly half of the most significantly differentially expressed genes in murine breast cancers resulting from genetic modeling of viral infection are conserved differentially expressed genes in human breast cancer.

We compared global gene expression profiles of primary tumors of the breast resulting from viral infection in mice to benign mammary tissue to identify the 100 most significantly differentially expressed genes in the mammalian breast upon transformation induced by genetic modeling of multiple sources of viral infection in mice (Table 1). These included the mouse mammary tumor virus (MMTV) MMTV-PyMT harboring the polyomavirus middle T antigen. Of the most differentially expressed genes transcriptome-wide in murine models of breast cancer, depending on the dataset queried, 45 of the top 100, or 53 of the top 100 were also differentially expressed to a statistically significant extent in human breast cancer when comparing human breast tumors to normal breast tissue from patients with breast cancer (Table 2 and Table 3). Thus, between 45% and 53% of the most differentially expressed genes in murine models of breast cancer induced by viral infection, including MMTV-PyMT and MMTV-Wnt, or transgenic expression of a viral oncoprotein (SV40 large T antigen) were conserved differentially expressed genes in human breast cancer.

Remarkably, canonical genes found to be mutated in familial breast cancer, Rad51C and BRCA1, were both recovered as genes most differentially expressed in breast cancer, both in primary tumors of mice resulting from multiple variants of viral infection (Table 1) as well as in human breast cancer (Table 2 and Table 3) (Table 1: $p=4.43E-04$ for Rad51C and $p=3.78E-04$ for BRCA1 and Table 2: $p=8.86E-05$ Rad51C and $p=2.72E-07$ for BRCA1; Table 3: $p=4.19E-15$ for Rad51C and $p=1.15E-11$ for BRCA1). In human breast cancer, differential expression of BRCA1 ranked 2002 out of 22283 total transcripts in one dataset (91% differential expression, or change in expression between tumor and normal breast tissue more significant than 91% of the measured transcriptome; Table 2), and 1732 out of 19076 total transcripts in a second microarray dataset (90.9% differential expression; Table 3). In murine breast cancer resulting from viral infection, differential expression of BRCA1 ranked 51 out of 18586 total transcripts (Table 1; $p=1.15E-11$).

We identified multiple components of transcriptional and epigenetic machinery as differentially expressed both in virally-induced murine breast cancer and in human breast cancer. The DNA methyltransferase DNMT1 was ranked 68 out of 18586 transcripts when comparing breast tumors resulting from viral infection in mice to benign mammary tissue; in human breast cancer, differential expression of DNMT1 ranked 432 out of 19076 total transcripts (Table 2: $p=1.19E-07$; 98.1% differential expression) and 375 out of 22283 in a second dataset of primary tumors from patients with breast cancer (Table 3; $p=1.4E-11$; 98% differential expression). TAF4 was differentially expressed in breast cancers arising from

genetic modeling of viral infection (rank 52 out of 18586 total transcripts; $p=1.49E-11$; 99.7% differential expression), and we found significant differential expression of TAF4 in both primary tumor datasets in human breast cancer (Table 2: rank 2914 out of 19076 total transcripts; $p=2.19E-03$; 84.7% differential expression; Table 3: rank 1560 out of 22283 total transcripts; $p=1.21E-07$; 93.0% differential expression). Differential expression of the PHD finger protein PHF20 was also conserved across species, with differential expression of PHF20 ranking 73 out of 18586 total transcripts in murine breast cancer resulting from genetic modeling of viral infection (Table 1; $p=6.21E-11$; 99.6% differential expression), and ranking 5037 out of 19076 total transcripts (Table 2; $p=1.92E-02$; 73.6% differential expression) and 2598 out of 22283 total transcripts (Table 3; $p=4.28E-06$; 88.3% differential expression) in the primary tumors of humans with breast cancer in two separate datasets.

Genes involved in cellular metabolism were also differentially expressed in a conserved fashion across mouse models of breast cancer induced by genetic modeling of viral infection and in human breast cancer. Differential expression of the acetyl co-A carboxylase alpha, ACACA, ranked 64 out of 18586 total transcripts (Table 1; $p=3.68E-11$; 99.7% differential expression); in the primary tumors of humans with breast cancer, ACACA differential expression ranked 3552 out of 19076 total transcripts (Table 2; $p=5.39E-03$; 84.1% differential expression) and 5902 out of 22283 total transcripts (Table 3; $p=1.9E-03$; 69.1% differential expression). The fatty acid synthase FASN was also differentially expressed in breast cancer in a conserved fashion across man and mouse, with differential expression of FASN ranking 72 out of 18586 total transcripts in mice with breast cancers resulting from genetic modeling of viral infection (Table 1; $p=5.65E-11$; 99.6% differential expression), and 4139 out of 19076 total transcripts (Table 2; $p=9.2E-03$; 78.3% differential expression) and 8441 out of 22283 total transcripts in human breast cancer (Table 3; $p=2.33E-02$; 62.1% differential expression). Adenine phosphoribosyltransferase (APRT) was among the genes most differentially expressed in the murine breast cancer resulting from genetic modeling of viral infection, ranking 100 out of 18586 total transcripts (Table 1; $p=3.36E-10$; 99.5% differential expression), and also differentially expressed in the primary tumors of humans with breast cancer: APRT differential expression ranked 2078 out of 19076 total transcripts (Table 2; $p=4.71E-04$; 89.1% differential expression) and 6682 out of 22283 total transcripts (Table 3; $p=4.67E-03$; 70.0% differential expression). Dihydrolipoamide S-acetyltransferase (DLAT) was another cellular metabolism gene identified as differentially expressed in a conserved fashion in breast cancer. Differential expression of DLAT ranked 49 out of 18586 total transcripts when comparing benign mammary tissue to murine breast tumors resulting from genetic modeling of viral infection (Table 1; $p=1.04E-11$; 99.7% differential expression); DLAT differential expression ranked 5027 out of 19076 transcripts in human breast cancer (Table 2; $p=1.91E-02$; 73.6% differential expression) and 7606 out of 22283 transcripts in a separate human breast cancer microarray dataset (Table 3; $p=1.14E-02$; 65.9% differential expression).

We identified genes classically associated with cancers across tissues, including the oncogenes Myc/Mycn and HRas. Differential expression of Myc ranked 98 out of 18586 total transcripts when comparing benign mammary tissue to tumors derived from mice genetically engineered to mimic viral integration or oncogene expression (Table 1; $p=3.04E-10$; 99.5% differential expression). In primary tumors from humans with breast cancer, differential expression of Myc ranked 357 out of 19076 total transcripts when comparing normal breast tissue to primary tumors from females (Table 2; $p=5.41E-08$; 98.1% differential expression), and Mycn ranked 6602 out of 22283 total transcripts in a second dataset from patients with breast cancer (Table 3; $p=4.32E-03$; 70.4% differential expression). Differential expression of HRas ranked 75 out of 18586 total transcripts when comparing benign mammary tissue to tumors

1 derived from mice genetically engineered to mimic viral integration or oncogene expression
2 (Table 1; $p=6.78E-11$; 99.6% differential expression). In primary tumors from humans with
3 breast cancer, differential expression of HRas ranked 3278 out of 19076 total transcripts when
4 comparing normal breast tissue to primary tumors from females (Table 2; $p=3.81E-03$; 82.8%
differential expression), and 6227 out of 22283 total transcripts in a second dataset from
patients with breast cancer (Table 3; $p=2.82E-03$; 72.1% differential expression).

5 The progestin and adipoQ receptor family member IV, PAQR4 was also among the
6 genes whose expression changed most significantly when comparing primary tumors from
7 mice resulting from genetic engineering of viral integration events or transgenic expression of a
8 viral oncogene. Differential expression of PAQR4 ranked 61 out of 18586 total transcripts when
9 comparing benign mammary tissue to tumors derived from mice genetically engineered to
10 mimic viral integration or oncogene expression (Table 1; $p=3.18E-11$; 99.7% differential
11 expression). In primary tumors from humans with breast cancer, differential expression of
PAQR4 ranked 571 out of 19076 total transcripts when comparing normal breast tissue to
primary tumors from females (Table 2; $p=5.19E-07$; 97.4% differential expression), and 540 out
of 22283 total transcripts in a second dataset from patients with breast cancer (Table 3;
 $p=1.53E-10$; 97.2% differential expression).

12 We described here just 12 of 45-53 genes out of 100 whose expression is most
13 significantly different, transcriptome-wide, when comparing non-transformed breast tissues to
14 primary tumors from mice genetically engineered to mimic viral integration or oncogene
expression, or to primary tumors from humans with breast cancer in two separate patient
microarray datasets.

15 Thus, nearly half of the genes whose expression changes most significantly in breast
16 cancers resulting from genetic modeling of viral infection at the level of integration or
17 expression of a viral oncoprotein in mice are also among those whose expression changes
18 most significantly in the tumors of humans with breast cancer. These findings were replicated
19 across two separate patient tumor datasets. The transcriptional signature of virally-induced
transformation in the breast as ascertained by genetic modeling with elements from
polyomavirus, SV40, and mouse mammary tumor virus are, despite inherent differences in
transcription due to species divergence, remarkably similar to that of human breast cancer.

20 Discussion

21
22 Despite the fact that only 5-10% of the greater than quarter of a million of women in the
23 United States alone that develop breast cancer are thought to harbor mutations that result in
24 the familial or inherited form of breast cancer, it remains unclear what are the major causative
25 factors resulting in the frequency of breast cancer observed in women today. We harnessed
26 the power of blind, transcriptome-level analyses comparing tumors from mice resulting from
genetic modeling of viral integration and transgenic expression events, and from humans with
breast cancer to ask whether we could observe any similarities in gene expression between
human breast cancers and those causing breast cancer and induced during viral infection in
mice.

27 We identified multiple genes with important functions in cancer cell biology⁴⁴⁻⁴⁸ as
28 transcriptionally perturbed in breast cancers, in humans and in mice resulting from genetically

defined integration events and expression of viral oncogenes that mimic those found in viral infections. This was not limited to a select number of genes differentially expressed; nearly half of the top 100 most differentially expressed genes when comparing benign tissue of the breast to primary breast tumors induced by genetically defined events that model viral infection of the murine mammary gland were found to be significantly differentially expressed in the primary tumors of humans with breast cancer. These data demonstrate that the transcriptional fingerprint of viral transformation in the mammalian breast is of significant similarity with that of breast cancer in females, and provide further evidence to support the hypothesis that one or more viral agents may contribute to transformation leading to development of non-inherited forms of breast cancer in humans. It will be of interest to evaluate, using next-generation RNA sequencing methods, the presence of viral transcripts, the open reading frames they represent and viruses from which they are derived, as well as the possibility of integration sites in tissues from the female breast, map these sites using whole genome sequencing, and determine whether modeling these integration events experimentally can result in transformation in human cells. It also remains a possibility that endogenous proviral sequences can be mobilized during retroviral infection of the breast; these methods can ascertain the existence of these sequences in the breast cancer genome, determine whether transcripts produced from these elements are found in the human breast cancer transcriptome, and provide tools to evaluate the possibility that infectious triggers can mobilize these elements to support or cooperate with other factors to support transformation.

References

1. Mamoor, S., 2020. Transcriptional similarities between human breast cancers and breast cancers induced by viral infection in mice.
2. Sinn, B.V., Fu, C., Lau, R., Litton, J., Tsai, T.H., Murthy, R., Tam, A., Andreopoulou, E., Gong, Y., Murthy, R. and Gould, R., 2019. SET ER/PR: a robust 18-gene predictor for sensitivity to endocrine therapy for metastatic breast cancer. *NPJ breast cancer*, 5(1), pp. 1-8.
3. Chang, J.W., Kuo, W.H., Lin, C.M., Chen, W.L., Chan, S.H., Chiu, M.F., Chang, I.S., Jiang, S.S., Tsai, F.Y., Chen, C.H. and Huang, P.H., 2018. Wild-type p53 upregulates an early onset breast cancer-associated gene GAS7 to suppress metastasis via GAS7–CYFIP1-mediated signaling pathway. *Oncogene*, 37(30), pp.4137-4150.
4. Zhu, M., Yi, M., Kim, C.H., Deng, C., Li, Y., Medina, D., Stephens, R.M. and Green, J.E., 2011.
5. miRNA and mRNA expression profiling of mouse mammary tumor models identifies miRNA signatures associated with mammary tumor lineage. *Genome biology*, 12(8), p.R77.
6. American Cancer Society, 2019. Breast cancer facts & figures 2019–2020. *Am. Cancer Soc*, pp.1-44.
7. Board, P.C.G.E., 2002. Genetics of Breast and Gynecologic Cancers (PDQ®): Health Professional Version. *PDQ Cancer Information Summaries [Internet]*.
8. King, M.C., Marks, J.H. and Mandell, J.B., 2003. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science*, 302(5645), pp.643-646.
9. Coles, C., Condie, A., Chetty, U., Steel, C.M., Evans, H.J. and Prosser, J., 1992. p53 mutations in breast cancer. *Cancer research*, 52(19), pp.5291-5298.
10. Meindl, A., Hellebrand, H., Wiek, C., Erven, V., Wappenschmidt, B., Niederacher, D., Freund, M., Lichtner, P., Hartmann, L., Schaal, H. and Ramser, J., 2010. Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. *Nature genetics*, 42(5), pp.410-414.
11. DeSantis, C.E., Ma, J., Gaudet, M.M., Newman, L.A., Miller, K.D., Goding Sauer, A., Jemal, A. and Siegel, R.L., 2019. Breast cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(6), pp.438-451.
12. Moore DH, Charney JE, Kramarsky BE, Lasfargues EY, Sarkar NH, BRENNAN MJ, BURROWS JH, SIRSAT SM, Paymaster JC, Vaidya AB. Search for a human breast cancer virus. *Nature*. 1971 Feb;229(5287):611-5.

- 1 13. Glenn, W.K., Heng, B., Delprado, W., Iacopetta, B., Whitaker, N.J. and Lawson, J.S., 2012.
2 Epstein-Barr virus, human papillomavirus and mouse mammary tumour virus as multiple
3 viruses in breast cancer. *PloS one*, 7(11), p.e48788.
- 4 14. Joshi, D. and Buehring, G.C., 2012. Are viruses associated with human breast cancer?
5 Scrutinizing the molecular evidence. *Breast cancer research and treatment*, 135(1), pp.
6 1-15.
- 7 15. Kan, C.Y., Iacopetta, B.J., Lawson, J.S. and Whitaker, N.J., 2005. Identification of human
8 papillomavirus DNA gene sequences in human breast cancer. *British Journal of*
9 *Cancer*, 93(8), pp.946-948.
- 10 16. Hogan, T.F., Padgett, B.L., Walker, D.L., Borden, E.C. and Frias, Z., 1983. Survey of human
11 polyomavirus (JCV, BKV) infections in 139 patients with lung cancer, breast cancer,
12 melanoma, or lymphoma. *Progress in clinical and biological research*, 105, pp.311-324.
- 13 17. Martini, F., Mazzoni, E., Corallini, A., Taronna, A., Querzoli, P., Magri, E., Marci, R., Dolcetti,
14 R., Rezza, G., Barbanti-Brodano, G. and Tognon, M., 2013. Breast cancer and simian virus
15 40 infection. *Epidemiology*, 24(3), pp.464-465.
- 16 18. Sitas, F., Carrara, H., Beral, V., Newton, R., Reeves, G., Bull, D., Jentsch, U., Pacella-
17 Norman, R., Bourboulia, D., Whitby, D. and Boshoff, C., 1999. Antibodies against human
18 herpesvirus 8 in black South African patients with cancer. *New England Journal of*
19 *Medicine*, 340(24), pp.1863-1871.
- 20 19. Labrecque, L.G., Barnes, D.M., Fentiman, I.S. and Griffin, B.E., 1995. Epstein-Barr virus in
21 epithelial cell tumors: a breast cancer study. *Cancer research*, 55(1), pp.39-45.
- 22 20. Richardson, A.K., Walker, L.C., Cox, B., Rollag, H., Robinson, B.A., Morrin, H., Pearson,
23 J.F., Potter, J.D., Paterson, M., Surcel, H.M. and Pukkala, E., 2020. Breast cancer and
24 cytomegalovirus. *Clinical and Translational Oncology*, 22(4), pp.585-602.
- 25 21. Wang, Y., Melana, S.M., Baker, B., Bleiweiss, I., Fernandez-Cobo, M., Mandeli, J.F.,
26 Holland, J.F. and Pogo, B.G.T., 2003. High prevalence of MMTV-like env gene sequences
27 in gestational breast cancer. *Medical Oncology*, 20(3), pp.233-236.
- 28 22. Wang, Y., Pelisson, I., Melana, S.M., Go, V., Holland, J.F. and Pogo, B.T., 2001. MMTV-like
env gene sequences in human breast cancer. *Archives of virology*, 146(1), pp.171-180.
23. Wang, Y., Pelisson, I., Melana, S.M., Holland, J.F. and Pogo, B.G.T., 2001. Detection of
MMTV-like LTR and LTR-env gene sequences in human breast cancer. *International journal*
of oncology, 18(5), pp.1041-1044.

24. Liu, B., Wang, Y., Melana, S.M., Pelisson, I., Najfeld, V., Holland, J.F. and Pogo, B.G., 2001. Identification of a proviral structure in human breast cancer. *Cancer research*, 61(4), pp. 1754-1759.
25. Naushad, W., Surriya, O. and Sadia, H., 2017. Prevalence of EBV, HPV and MMTV in Pakistani breast cancer patients: a possible etiological role of viruses in breast cancer. *Infection, Genetics and Evolution*, 54, pp.230-237.
26. Wang-Johanning, F., Frost, A.R., Jian, B., Epp, L., Lu, D.W. and Johanning, G.L., 2003. Quantitation of HERV-K env gene expression and splicing in human breast cancer. *Oncogene*, 22(10), pp.1528-1535.
27. Wang-Johanning, F., Frost, A.R., Johanning, G.L., Khazaeli, M.B., LoBuglio, A.F., Shaw, D.R. and Strong, T.V., 2001. Expression of human endogenous retrovirus k envelope transcripts in human breast cancer. *Clinical Cancer Research*, 7(6), pp.1553-1560.
28. Dalton-Griffin, L. and Kellam, P., 2009. Infectious causes of cancer and their detection. *Journal of biology*, 8(7), p.67.
29. Kim, C.M., Koike, K., Saito, I., Miyamura, T. and Jay, G., 1991. HBx gene of hepatitis B virus induces liver cancer in transgenic mice. *Nature*, 351(6324), pp.317-320.
30. Connolly, D.C., Bao, R., Nikitin, A.Y., Stephens, K.C., Poole, T.W., Hua, X., Harris, S.S., Vanderhyden, B.C. and Hamilton, T.C., 2003. Female mice chimeric for expression of the simian virus 40 TAG under control of the MISIR promoter develop epithelial ovarian cancer. *Cancer research*, 63(6), pp.1389-1397.
31. Lin, E.Y., Jones, J.G., Li, P., Zhu, L., Whitney, K.D., Muller, W.J. and Pollard, J.W., 2003. Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *The American journal of pathology*, 163(5), pp.2113-2126.
32. Rennhack, J.P., To, B., Swiatnicki, M., Dulak, C., Ogradzinski, M.P., Zhang, Y., Li, C., Bylett, E., Ross, C., Szczepanek, K. and Hanrahan, W., 2019. Integrated analyses of murine breast cancer models reveal critical parallels with human disease. *Nature communications*, 10(1), pp.1-12.
33. Angus, L., Smid, M., Wilting, S.M., van Riet, J., Van Hoeck, A., Nguyen, L., Nik-Zainal, S., Steenbruggen, T.G., Tjan-Heijnen, V.C., Labots, M. and van Riel, J.M., 2019. The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nature genetics*, pp.1-9.

34. Marcotte, R., Sayad, A., Brown, K.R., Sanchez-Garcia, F., Reimand, J., Haider, M., Virtanen, C., Bradner, J.E., Bader, G.D., Mills, G.B. and Pe'er, D., 2016. Functional genomic landscape of human breast cancer drivers, vulnerabilities, and resistance. *Cell*, 164(1-2), pp.293-309.
35. Yates, L.R., Knappskog, S., Wedge, D., Farmery, J.H., Gonzalez, S., Martincorena, I., Alexandrov, L.B., Van Loo, P., Haugland, H.K., Lilleng, P.K. and Gundem, G., 2017. Genomic evolution of breast cancer metastasis and relapse. *Cancer cell*, 32(2), pp. 169-184.
36. Bertucci, F., Ng, C.K., Patsouris, A., Droin, N., Piscuoglio, S., Carbuccia, N., Soria, J.C., Dien, A.T., Adnani, Y., Kamal, M. and Garnier, S., 2019. Genomic characterization of metastatic breast cancers. *Nature*, 569(7757), pp.560-564.
37. Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R. and Yates, L.R., 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature*, 486(7403), pp.400-404.
38. Rhee, I., Bachman, K.E., Park, B.H., Jair, K.W., Yen, R.W.C., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W. and Baylin, S.B., 2002. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature*, 416(6880), pp.552-556.
39. Pathania, R., Ramachandran, S., Elangovan, S., Padia, R., Yang, P., Cinghu, S., Veeranan-Karmegam, R., Arjunan, P., Gnana-Prakasam, J.P., Sadanand, F. and Pei, L., 2015. DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. *Nature communications*, 6(1), pp.1-11.
40. Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y. and Li, E., 2007. Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nature genetics*, 39(3), pp.391-396.
41. Zhang, X., Peng, D., Xi, Y., Yuan, C., Sagum, C.A., Klein, B.J., Tanaka, K., Wen, H., Kutateladze, T.G., Li, W. and Bedford, M.T., 2016. G9a-mediated methylation of ERα links the PHF20/MOF histone acetyltransferase complex to hormonal gene expression. *Nature communications*, 7(1), pp.1-12.
42. Cui, G., Park, S., Badeaux, A.I., Kim, D., Lee, J., Thompson, J.R., Yan, F., Kaneko, S., Yuan, Z., Botuyan, M.V. and Bedford, M.T., 2012. PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. *Nature structural & molecular biology*, 19(9), p.916.

43. Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T. and Chen, F., 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer cell*, 10(6), pp.529-541.
44. Menendez, J.A. and Lupu, R., 2017. Fatty acid synthase (FASN) as a therapeutic target in breast cancer. *Expert opinion on therapeutic targets*, 21(11), pp.1001-1016.
45. Alwarawrah, Y., Hughes, P., Loiselle, D., Carlson, D.A., Darr, D.B., Jordan, J.L., Xiong, J., Hunter, L.M., Dubois, L.G., Thompson, J.W. and Kulkarni, M.M., 2016. Fasnall, a selective FASN inhibitor, shows potent anti-tumor activity in the MMTV-Neu model of HER2+ breast cancer. *Cell chemical biology*, 23(6), pp.678-688.
46. Berns, E.M., Klijn, J.G., van Putten, W.L., van Staveren, I.L., Portengen, H. and Foekens, J.A., 1992. c-myc amplification is a better prognostic factor than HER2/neu amplification in primary breast cancer. *Cancer research*, 52(5), pp.1107-1113.
47. Geyer, F.C., Li, A., Papanastasiou, A.D., Smith, A., Selenica, P., Burke, K.A., Edelweiss, M., Wen, H.C., Piscuoglio, S., Schultheis, A.M. and Martelotto, L.G., 2018. Recurrent hotspot mutations in HRAS Q61 and PI3K-AKT pathway genes as drivers of breast adenomyoepitheliomas. *Nature communications*, 9(1), pp.1-16.
48. Pedersen, L., Panahandeh, P., Siraji, M.I., Knappskog, S., Lønning, P.E., Gordillo, R., Scherer, P.E., Molven, A., Teigen, K. and Halberg, N., 2020. Golgi-Localized PAQR4 Mediates Antiapoptotic Ceramidase Activity in Breast Cancer. *Cancer Research*, 80(11), pp. 2163-2174.

Rank	ID	p-value	F	Gene	Gene name
1	1420539_a_at	3.36E-26	378.18071	Chrdl2	chordin-like 2
2	1453076_at	6.9E-26	359.61882	Batf3	basic leucine zipper transcription factor, ATF-like 3
3	1451417_at	7.81E-18	96.35439	Brca1	breast cancer 1, early onset
4	1451510_s_at	6.62E-17	82.42435	Olah	oleoyl-ACP hydrolase
5	1424855_at	1.81E-16	76.54317	Olah	oleoyl-ACP hydrolase
6	1421048_a_at	4.11E-16	72.03452	Ypel1	yippee-like 1 (Drosophila)
7	1451173_at	1.15E-15	66.74479	Lrrc49	leucine rich repeat containing 49
8	1455579_at	2.6E-15	62.78362	Csn1s2a	casein alpha s2-like A
9	1419491_at	3.02E-15	62.07817	Defb1	defensin beta 1
10	1438453_at	4.19E-15	60.57993	Rad51c	RAD51 paralog C
11	1417940_s_at	8.03E-15	57.68112	Rad51ap1	RAD51 associated protein 1
12	1418818_at	8.34E-15	57.51566	Aqp5	aquaporin 5
13	1416031_s_at	8.96E-15	57.20793	Mcm7	minichromosome maintenance complex component 7
14	1448397_at	2.32E-14	53.23537	Gjb6	gap junction protein, beta 6
15	1455834_x_at	2.34E-14	53.20092	Tacc3	transforming, acidic coiled-coil containing protein 3
16	1427495_at	7.06E-14	48.89494	Scn7a	sodium channel, voltage-gated, type VII, alpha
17	1417257_at	1.34E-13	46.54823	Cel	carboxyl ester lipase
18	1418221_at	1.6E-13	45.91904	Csn1s2b	casein alpha s2-like B
19	1451824_at	1.68E-13	45.73887	Btn1a1	butyrophilin, subfamily 1, member A1
20	1424143_a_at	1.78E-13	45.53115	Cdt1	chromatin licensing and DNA replication factor 1
21	1425439_a_at	2.52E-13	44.33724	Slc41a3	solute carrier family 41, member 3
22	1436872_at	2.56E-13	44.28346	Tacc3	transforming, acidic coiled-coil containing protein 3
23	1419140_at	2.98E-13	43.75724	Acvr2b	activin receptor IIB
24	1460316_at	3.06E-13	43.66939	Acsl1	acyl-CoA synthetase long-chain family member 1
25	1417841_at	3.06E-13	43.66603	Pxmp2	peroxisomal membrane protein 2

Table 1: Genes most differentially expressed in breast tumors from mice genetically engineered to model integration or expression of viral oncogenes when compared to benign mammary tissue. Rank of differential expression, probe ID, p-value with respect to global differential expression, F-statistic, gene and gene name are listed in this chart.

Rank	ID	p-value	F	Gene	Gene name
26	1448949_at	3.25E-13	43.46664	Car4	carbonic anhydrase 4
27	1426697_a_at	4.13E-13	42.66092	Lrpap1	low density lipoprotein receptor-related protein associated protein 1
28	1425004_s_at	4.45E-13	42.41997	Mocs1	molybdenum cofactor synthesis 1
29	1417545_at	6.07E-13	41.40516	Trpv4	transient receptor potential cation channel, subfamily V, member 4
30	1418719_at	9.46E-13	39.98956	Haus8	4HAUS augmin-like complex, subunit 8
31	1418094_s_at	1.38E-12	38.82271	Car4	carbonic anhydrase 4
32	1438320_s_at	1.53E-12	38.51087	Mcm7	minichromosome maintenance complex component 7
33	1438694_at	2.04E-12	37.64074	Csn1s2a	casein alpha s2-like A
34	1460168_at	2.31E-12	37.27139	Slbp	stem-loop binding protein
35	1425737_at	2.33E-12	37.25105	1500002F19Rik	RIKEN cDNA 1500002F19 gene
36	1417650_at	3.59E-12	35.9944	Lao1	L-amino acid oxidase 1
37	1417403_at	3.68E-12	35.92678	Elovl6	ELOVL family member 6, elongation of long chain fatty acids (yeast)
38	1453614_a_at	3.86E-12	35.78806	Nfe2l3	nuclear factor, erythroid derived 2, like 3
39	1418525_at	4.08E-12	35.62793	Pcm1	pericentriolar material 1
40	1428010_at	4.33E-12	35.45995	Timm9	translocase of inner mitochondrial membrane 9
41	1435275_at	6.13E-12	34.49379	Cox6b2	cytochrome c oxidase subunit VIb polypeptide 2
42	1422007_at	6.51E-12	34.32464	Aqp3	aquaporin 3
43	1418481_at	6.63E-12	34.27847	Pkmyt1	protein kinase, membrane associated tyrosine/threonine 1
44	1428117_x_at	6.9E-12	34.16898	Dynlt1f	dynein light chain Tctex-type 1F
45	1423846_x_at	6.98E-12	34.13641	Gm5620///Tuba1b	tubulin, alpha 1B pseudogene///tubulin, alpha 1B
46	1421907_at	8.47E-12	33.61158	Med1	mediator complex subunit 1
47	1418979_at	8.52E-12	33.59308	Akr1c14	aldo-keto reductase family 1, member C14
48	1417830_at	9.55E-12	33.28906	Smc1a	structural maintenance of chromosomes 1A
49	1426265_x_at	1.04E-11	33.04822	Dlat	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)
50	1421922_at	1.07E-11	32.98106	Sh3bp5	SH3-domain binding protein 5 (BTK-associated)

Table 1: Genes most differentially expressed in breast tumors from mice genetically engineered to model integration or expression of viral oncogenes when compared to benign mammary tissue.

Rank of differential expression, probe ID, p-value with respect to global differential expression, F-statistic, gene and gene name are listed in this chart.

Rank	ID	p-value	F	Gene	Gene name
51	1424630_a_at	1.15E-11	32.79382	Brca1	breast cancer 1, early onset
52	1427460_at	1.49E-11	32.12144	Taf4	TATA-box binding protein associated factor 4
53	1416934_at	1.63E-11	31.8873	Mtm1	X-linked myotubular myopathy gene 1
54	1439269_x_at	1.75E-11	31.70363	Mcm7	minichromosome maintenance complex component 7
55	1420658_at	1.77E-11	31.6708	Ucp3	uncoupling protein 3 (mitochondrial, proton carrier)
56	1423760_at	1.89E-11	31.50752	Cd44	CD44 antigen
57	1422970_at	2.17E-11	31.14883	Mxd3	Max dimerization protein 3
58	1424509_at	2.19E-11	31.13501	Cd177	CD177 antigen
59	1428485_at	2.38E-11	30.92431	Car12	carbonic anhydrase 12
60	1417586_at	3.16E-11	30.21293	Timeless	timeless circadian clock 1
61	1423101_at	3.18E-11	30.20209	Paqr4	progesterone and adipoQ receptor family member IV
62	1426028_a_at	3.23E-11	30.16464	Cit	citron
63	1425499_at	3.59E-11	29.90119		
64	1427595_at	3.68E-11	29.84226	Acaca	acetyl-Coenzyme A carboxylase alpha
65	1416049_at	3.69E-11	29.83625	Gldc	glycine decarboxylase
66	1418560_at	3.85E-11	29.73033	Pdha1	pyruvate dehydrogenase E1 alpha 1
67	1435122_x_at	4.24E-11	29.49695	Dnmt1	DNA methyltransferase (cytosine-5) 1
68	1448382_at	4.58E-11	29.31035	Ehhadh	enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase
69	1448361_at	4.68E-11	29.26315	Ttc3	tetratricopeptide repeat domain 3
70	1460716_a_at	4.7E-11	29.24828	Cbfb	core binding factor beta
71	1418743_a_at	4.76E-11	29.22238	Tesc	tescalcin
72	1423828_at	5.65E-11	28.81026	Fasn	fatty acid synthase
73	1452259_at	6.21E-11	28.58799	Phf20	PHD finger protein 20
74	1416292_at	6.72E-11	28.40523	Prdx3	peroxiredoxin 3
75	1422407_s_at	6.78E-11	28.38373	Hras	Harvey rat sarcoma virus oncogene

Table 1: Genes most differentially expressed in breast tumors from mice genetically engineered to model integration or expression of viral oncogenes when compared to benign mammary tissue. Rank of differential expression, probe ID, p-value with respect to global differential expression, F-statistic, gene and gene name are listed in this chart.

Rank	ID	p-value	F	Gene	Gene name
76	1417930_at	6.99E-11	28.31166	Nab2	Ngfi-A binding protein 2
77	1429240_at	7.06E-11	28.28749	Stard4	StAR-related lipid transfer (START) domain containing 4
78	1423302_a_at	7.92E-11	28.02132	Paxip1	PAX interacting (with transcription-activation domain) protein 1
79	1460247_a_at	8.15E-11	27.95485	Skp2	S-phase kinase-associated protein 2 (p45)
80	1452459_at	8.36E-11	27.8954	Aspm	asp (abnormal spindle)-like, microcephaly associated (Drosophila)
81	1423123_at	8.91E-11	27.75021	Rad54l	RAD54 like (S. cerevisiae)
82	1460652_at	1.02E-10	27.44954	Esrra	estrogen related receptor, alpha
83	1432538_a_at	1.03E-10	27.41343	Rfc3	replication factor C (activator 1) 3
84	1449964_a_at	1.12E-10	27.22002	Mlycd	malonyl-CoA decarboxylase
85	1434537_at	1.14E-10	27.19728	Slco3a1	solute carrier organic anion transporter family, member 3a1
86	1424629_at	1.17E-10	27.12862	Brca1	breast cancer 1, early onset
87	1448221_at	1.22E-10	27.03415	Ddx39b	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39B
88	1437194_x_at	1.24E-10	27.00654	Lrwd1	leucine-rich repeats and WD repeat domain containing 1
89	1428306_at	1.38E-10	26.7667	Ddit4	DNA-damage-inducible transcript 4
90	1418514_at	1.38E-10	26.7624	Mtf2	metal response element binding transcription factor 2
91	1424715_at	1.44E-10	26.66192	Retsat	retinol saturase (all trans retinol 13,14 reductase)
92	1451838_a_at	1.61E-10	26.4126	Tc2n	tandem C2 domains, nuclear
93	1429527_a_at	1.66E-10	26.35632	Plscr1	phospholipid scramblase 1
94	1426264_at	1.72E-10	26.26776	Dlat	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)
95	1415802_at	1.93E-10	26.0204	Slc16a1	solute carrier family 16 (monocarboxylic acid transporters), member 1
96	1424132_at	2.02E-10	25.92339	Hras	Harvey rat sarcoma virus oncogene
97	1423714_at	2.07E-10	25.87433	Asf1b	anti-silencing function 1B histone chaperone
98	1424942_a_at	3.04E-10	25.04658	Myc	myelocytomatosis oncogene
99	1455730_at	3.14E-10	24.98169	Dlgap5	discs, large (Drosophila) homolog-associated protein 5
100	1451703_s_at	3.36E-10	24.8357	Aprt	adenine phosphoribosyl transferase

Table 1: Genes most differentially expressed in breast tumors from mice genetically engineered to model integration or expression of viral oncogenes when compared to benign mammary tissue.

Rank of differential expression, probe ID, p-value with respect to global differential expression, F-statistic, gene and gene name are listed in this chart.

Rank	ID	p-value	t	B	Gene	Gene name	% DE
53	207803_s_at	4.48E-15	-8.8436986	23.79174	CSN3	casein kappa	99.8%
48	218662_s_at	2.64E-15	8.9368446	24.30524	NCAPG	non-SMC condensin I complex subunit G	99.8%
157	219304_s_at	5.48E-12	-7.5617051	16.89095	PDGFD	platelet derived growth factor D	99.3%
179	218542_at	1.11E-11	7.4299541	16.20361	CEP55	centrosomal protein 55	99.2%
207	219918_s_at	2.35E-11	7.2900218	15.47901	ASPM	abnormal spindle microtubule assembly	99.1%
250	222037_at	6.66E-11	7.0935796	14.47174	MCM4	minichromosome maintenance complex component 4	98.9%
273	218308_at	1.02E-08	1.25E-10	6.9731386	TACC3	transforming acidic coiled-coil containing protein 3	98.8%
357	202431_s_at	5.41E-08	8.66E-10	-6.5988382	MYC	v-myc avian myelocytomatosis viral oncogene homolog	98.4%
389	203764_at	7.99E-08	1.4E-09	6.5048977	DLGAP5	DLG associated protein 5	98.3%
432	201697_s_at	1.19E-07	2.31E-09	6.4048693	DNMT1	DNA (cytosine-5-)-methyltransferase 1	98.1%
578	207275_s_at	5.57E-07	1.44E-08	-6.0348651	ACSL1	acyl-CoA synthetase long-chain family member 1	97.4%
571	212858_at	5.19E-07	1.33E-08	6.0513797	PAQR4	progesterin and adipoQ receptor family member 4	97.4%
954	209704_at	1.1E-05	4.7E-07	5.2950421	MTF2	metal response element binding transcription factor 2	95.7%
818	220358_at	4.7E-06	1.72E-07	5.5137315	BATF3	basic leucine zipper ATF-like transcription factor 3	93.4%
2002	211851_x_at	3.78E-04	3.4E-05	4.2887732	BRCA1	BRCA1, DNA repair associated	91%
2050	209849_s_at	4.43E-04	4.08E-05	4.2425118	RAD51C	RAD51 paralog C	90.8%
2078	203219_s_at	4.71E-04	4.39E-05	-4.223569	APRT	adenine phosphoribosyltransferase	90.7%
2140	204146_at	5.42E-04	5.21E-05	4.1799244	RAD51AP1	RAD51 associated protein 1	90.4%
2535	201589_at	1.21E-03	1.37E-04	3.9255202	SMC1A	structural maintenance of chromosomes 1A	88.6%
2718	204702_s_at	1.63E-03	1.98E-04	3.8261076	NFE2L3	nuclear factor, erythroid 2 like 3	87.8%
2835	204267_x_at	2E-03	2.54E-04	3.7579179	PKMYT1	protein kinase, membrane associated tyrosine/threonine 1	87.3%
2914	213090_s_at	2.19E-03	2.87E-04	3.7242758	TAF4	TATA-box binding protein associated factor 4	86.9%
3081	207395_at	2.92E-03	4.04E-04	-3.6282644	BTN1A1	butyrophilin subfamily 1 member A1	86.2%
3278	212983_at	3.81E-03	5.61E-04	3.5342148	HRAS	HRas proto-oncogene, GTPase	85.3%
3431	218316_at	4.62E-03	7.12E-04	-3.4648503	TIMM9	translocase of inner mitochondrial membrane 9	84.6%

Table 2: Differential expression of genes in breast tumors from mice genetically engineered to model integration or expression of viral oncogenes is conserved in primary tumors from patients with human breast cancer.

Rank of differential expression, probe ID, p-value with respect to global differential expression, F-statistic, gene, gene name and percent differential expression (DE) are listed in this chart.

3552	212186_at	5.39E-03	8.58E-04	3.4095024	ACACA	acetyl-CoA carboxylase alpha	84.1%
4065	218872_at	8.64E-03	1.58E-03	-3.2257684	TESC	tescalcin	81.8%
4139	217006_x_at	9.2E-03	1.71E-03	3.2006802	FASN	fatty acid synthase	81.4%
4249	203046_s_at	1.02E-02	1.95E-03	3.159724	TIMELESS	timeless circadian clock	80.9%
4346	221250_s_at	1.13E-02	2.2E-03	3.1223586	MXD3	MAX dimerization protein 3	80.5%
4442	202887_s_at	1.24E-02	2.47E-03	-3.0847831	DDIT4	DNA damage inducible transcript 4	80.1%
4449	212825_at	1.25E-02	2.49E-03	3.0824109	PAXIP1	PAX interacting protein 1	80%
5027	213149_at	1.91E-02	4.3E-03	2.9045673	DLAT	dihydrolipoamide S-acetyltransferase	77.4%
5037	206567_s_at	1.92E-02	4.34E-03	2.9015049	PHF20	PHD finger protein 20	77.4%
5027	213149_at	1.91E-02	4.3E-03	2.9045673	DLAT	dihydrolipoamide S-acetyltransferase	77.4%
5117	208073_x_at	2.03E-02	4.66E-03	-2.877534	TTC3P1/// TTC3	tetratricopeptide repeat domain 3 pseudogene 1/// tetratricopeptide repeat domain 3	77%
5194	209766_at	2.13E-02	4.96E-03	-2.8565602	PRDX3	peroxiredoxin 3	76.7%
5471	218124_at	2.54E-02	6.25E-03	-2.7780445	RETSAT	retinol saturase	75.4%
5865	219076_s_at	3.27E-02	8.6E-03	2.6665849	PXMP2	peroxisomal membrane protein 2	73.7%
6018	36920_at	3.56E-02	9.6E-03	2.6272728	MTM1	myotubularin 1	73%
6083	207864_at	3.69E-02	1.01E-02	-2.6104017	SCN7A	sodium voltage-gated channel alpha subunit 7	72.7%
6811	210916_s_at	5.37E-02	1.64E-02	-2.4301468	CD44	CD44 molecule (Indian blood group)	69.4%
7127	204558_at	6.19E-02	1.98E-02	2.3583937	RAD54L	RAD54-like (S. cerevisiae)	68%
7574	203625_x_at	7.46E-02	2.54E-02	2.2607651	SKP2	S-phase kinase-associated protein 2, E3 ubiquitin protein ligase	66%
7604	210542_s_at	7.51E-02	2.56E-02	-2.2567138	SLCO3A1	solute carrier organic anion transporter family member 3A1	65.9%

Table 2: Differential expression of genes in breast tumors from mice genetically engineered to model integration or expression of viral oncogenes is conserved in primary tumors from patients with human breast cancer.

Rank of differential expression, probe ID, p-value with respect to global differential expression, F-statistic, gene, gene name and percent differential expression (DE) are listed in this chart.

Rank	ID	p-value	t	B	Gene	Gene name	% DE
80	2953139	4.09E-15	-1.1E+01	24.209175	MOCS1	Molybdenum cofactor synthesis 1	99.6%
171	3010503	1.78E-13	-9.91	20.496206	CD36	CD36 molecule (thrombospondin receptor)	99.1%
224	3473436	7.48E-13	-9.49	19.078092	TESC	Tescalcin	98.8%
275	3457824	2.41E-12	9.15	17.925394	TIMELESS	timeless homolog	98.6%
313	3474104	5E-12	8.94	17.205446	CIT	citron (rho-interacting, serine/threonine kinase 21)	98.4%
321	3852565	5.89E-12	8.90	17.042783	ASF1B	ASF1 anti-silencing function 1 homolog B (S. cerevisiae)	98.3%
375	3850069	1.4E-11	8.65	16.191591	DNMT1	DNA (cytosine-5-)-methyltransferase 1 (DNMT1)	98%
540	3645477	1.53E-10	7.99	13.832578	PAQR4	progesterone and adiponectin receptor family member IV	97.2%
635	3677315	4.18E-10	7.71	12.839389	PKMYT1	Protein kinase, membrane associated tyrosine/threonine 1	96.7%
822	2714955	2.16E-09	7.25	11.222962	TACC3	Transforming, acidic coiled-coil containing protein 3	95.7%
889	3401804	3.29E-09	7.14	10.807745	RAD51AP1	RAD51 associated protein 1	95.3%
1526	2334646	1.05E-07	6.18	7.398644	RAD54L	RAD54-like	92%
1560	3912718	1.21E-07	6.14	7.258535	TAF4	TAF4 RNA polymerase II, TBP-associated factor	91.8%
1628	2585476	1.69E-07	-6.05	6.935013	SCN7A	Sodium channel, voltage-gated, type VII, alpha	91.5%
1732	3758317	2.72E-07	5.92	6.469422	BRCA1	Breast cancer 1, early onset	90.9%
2562	4009238	3.99E-06	5.17	3.844297	SMC1A	Structural maintenance of chromosomes 1A	86.6%
2598	3883547	4.28E-06	5.15	3.775966	PHF20	PHD finger protein 20	86.3%
2690	3016791	5.4E-06	5.08	3.550697	LRWD1	leucine-rich repeats and WD repeat domain containing 1	85.9%
2807	3608787	7.49E-06	-4.99	3.232472	SLCO3A1	solute carrier organic anion transporter family, member 3A1	85.3%
2844	3299661	8.29E-06	-4.96	3.133265	SLC16A2	solute carrier family 16, member 12	85.1%
3134	3835035	1.67E-05	4.75	2.453418	CD177	CD177 molecule	83.6%
3437	3080843	3.28E-05	4.56	1.800431	PAXIP1	PAX interacting (with transcription-activation domain) protein 1	82%
3553	2757319	4.35E-05	4.47	1.527362	SLBP	Stem-loop binding protein	81.4%
3945	3728776	8.86E-05	4.26	0.8428	RAD51C	RAD51 homolog C	79.3%
4166	3671482	1.34E-04	-4.13	0.447595	MLYCD	malonyl-CoA decarboxylase	78.2%
4444	3063685	2.18E-04	3.98	-0.021744	MCM7	Minichromosome maintenance complex component 7	76.7%

Table 3: Differential expression of genes in breast tumors from mice genetically engineered to model integration or expression of viral oncogenes is conserved in primary tumors from patients with human breast cancer.

Rank of differential expression, probe ID, p-value with respect to global differential expression, F-statistic, gene, gene name and percent differential expression (DE) are listed in this chart.

Rank	ID	p-value	t	B	Gene	Gene name	% DE
4841	2781813	4.31E-04	3.77	-0.669114	ELOVL	ELOVL family member 6, elongation of long chain fatty acids	74.6%
5249	2796553	7.88E-04	-3.57	-1.242008	ACSL1	Acyl-CoA synthetase long-chain family member 1	72.5%
5802	3309383	1.65E-03	3.32	-1.939571	PRDX3	peroxiredoxin 3	69.6%
5902	3754469	1.9E-03	3.28	-2.069569	ACACA	acetyl-CoA carboxylase alpha	69.1%
6084	3673684	2.38E-03	3.20	-2.279566	CDT1	Homo sapiens chromatin licensing and DNA replication factor 1	68.1%
6222	3087813	2.8E-03	-3.14	-2.433128	PCM1	Pericentriolar material 1	67.4%
6227	3358090	2.82E-03	3.14	-2.439295	HRAS	Harvey rat sarcoma viral oncogene homolog	67.4%
6475	3600212	3.79E-03	-3.03	-2.713731	LRRC49	Leucine-rich repeat-containing protein 49	66.1%
6602	2470805	4.32E-03	-2.99	-2.834405	MYCN	v-myc myelocytomatosis viral oncogene, neuroblastoma derived (avian)	65.4%
6682	3704495	4.67E-03	2.96	-2.905933	APRT	adenine phosphoribosyltransferase (APRT)	65%
6714	3994795	4.82E-03	2.95	-2.934906	MTM1	Myotubularin	64.8%
6724	3417809	4.85E-03	2.95	-2.940996	NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)	64.8%
6735	2708720	4.91E-03	-2.94	-2.951967	EHHADH	enoyl-CoA, hydratase/3-hydroxyacyl CoA dehydrogenase	64.7%
6756	3665116	5.01E-03	2.93	-2.970377	CBFB	core-binding factor, beta subunit	64.6%
6973	3414315	6.17E-03	-2.86	-3.162765	AQP5	Aquaporin 5	63.4%
6974	3914114	6.18E-03	-2.86	-3.164111	ZBTB46	Zinc finger and BTB domain containing 46	63.4%
7038	3485074	6.55E-03	2.84	-3.216873	RFC3	replication factor C (activator 1) 3	63.1%
7606	3348852	1.14E-02	2.62	-3.724434	DLAT	dihydrolipoamide S-acetyltransferase	60.1%
7645	3854218	1.18E-02	2.61	-3.752901	HAUS8	HAUS augmin-like complex, subunit 8	59.9%
7691	3504213	1.22E-02	2.60	-3.782772	GJB6	Gap junction protein, beta 6	59.7%
8184	2806517	1.89E-02	2.43	-4.175328	SKP2	S-phase kinase-associated protein 2 (p45)	57.1%
8339	2562233	2.33E-02	-2.34	-4.362082	RETSAT	retinol saturase (all-trans-retinol 13,14-reductase)	56.3%
8441	3774635	2.33E-02	2.34	-4.364349	FASN	fatty acid synthase	55.8%
8699	2870828	2.81E-02	2.26	-4.526454	STARD4	StAR-related lipid transfer (START) domain containing 4	54.4%
8789	3439178	2.98E-02	2.24	-4.578655	PXMP2	Peroxisomal membrane protein 2	53.9%
8805	3197955	3.01E-02	2.23	-4.589719	GLDC	glycine dehydrogenase (decarboxylating)	53.8%
9434	3566652	4.61E-02	2.04	-4.958476	TIMM9	Translocase of inner mitochondrial membrane 9	50.5%

Table 3: Differential expression of genes in breast tumors from mice genetically engineered to model integration or expression of viral oncogenes is conserved in primary tumors from patients with human breast cancer.

Rank of differential expression, probe ID, p-value with respect to global differential expression, F-statistic, gene, gene name and percent differential expression (DE) are listed in this chart.