1 **Transcriptional similarities between human breast cancers and breast cancers resulting**
2 **from genetic modeling of viral infection in mice.**

3 Shahan Mamoor, MS¹

4 ¹shahanmamoor@gmail.com

5 East Islip, NY 11730

6 Breast cancer affects women in the United States and worldwide at relatively high frequency¹.
7 Mutations in BRCA1 have been identified as a cause of breast cancer in a minority of patients
8 with the familiar or inherited form of breast cancer^{2,3}, but it remains unknown what factor or
9 factors trigger development of breast cancer in the vast majority of women. It has been
10 suggested that infection with a spectrum of viruses, including DNA tumor viruses like human
11 papillomavirus and SV40, and retroviruses like the mouse mammary tumor virus (MMTV) can
12 be associated with human breast cancer⁴⁻⁶ but experimental evidence to support the claim that
13 viral infections cause human breast cancer is lacking. We evaluated the hypothesis that
14 primary tumors from humans with breast cancers will share transcriptional similarity at the
15 systems level with experimental breast cancers induced by genetic modeling of viral infection or
16 transgenic expression of a viral oncoprotein by mining published microarray data⁷⁻⁹. We
17 identified, in a blind fashion, eight independent genes with crucial functions in cancer cell
18 survival among the most differentially expressed genes transcriptome-wide, including genes not
19 previously implicated in human breast cancer, both when comparing normal breast tissue and
20 primary tumors from women with breast cancer, and when benign mammary tissue to breast
21 tumors from 5 independent strains of mice genetically engineered to model viral induction of
22 transformation in the mammalian breast, including MMTV-PyMT, MMTV-Wnt1 and the large T-
23 antigen of SV40. These marked transcriptional changes between breast and tumor were not
24 only shared by humans with breast cancer and breast cancer resulting from genetic modeling of
25 viral integration or oncogene expression events in mice, but also by mice genetically engineered
26 to model familial breast cancer, lacking BRCA1 and one copy of the tumor suppressor p53.
27 These data demonstrate that key changes in the human breast cancer tumor transcriptome are
28 also found in mice that develop breast cancer as a direct result of viral infection, and are one
piece of experimental evidence, obtained through unbiased methods, to support the assertion
that viral infection can contribute to breast cancer in humans.

Keyword: breast cancer, virally-induced transformation, mouse mammary tumor virus, MMTV,
breast cancer etiology.

Studies evaluating the claim that viral infection can be the cause of breast cancer in humans have until now relied for the large part on analyses of viral transcription in patient tumors by standard PCR using oligonucleotide primers targeting a region or region(s) of the viral genome, or immunohistochemical methods to evaluate translation of viral proteins using antibodies specific to viral antigens. These studies, which have evaluated the association of viruses including human papillomavirus, the cause of cervical cancer, the polyomaviruses JC, BK and SV40, human herpesviruses (HHV) such as the cause of Burkitt's lymphoma, Epstein-Barr virus, HHV-8 the cause of Kaposi's Sarcoma in patients with acquired immunodeficiency virus (AIDS), and cytomegalovirus, retroviruses including mouse mammary tumor virus (MMTV) for which there is no identified human counterpart, as well as proviral elements such as HERV-K with the development of breast cancer, have been mostly correlative, demonstrating existence of viral transcription or viral antigen expression in tumor tissues and with great variation in frequency, ranging from 0 to 100%^{4-6,10-22}. Extensive laboratory investigation has demonstrated that infection with viruses can result in tumor development^{23,24}. However, evidence demonstrating that transformation in mice caused by viral infection, such as that of the mouse mammary gland caused by engineering of the MMTV virus promoter driving expression of the polyomavirus middle T antigen, which induces tumor formation in the breast with metastasis to the lung, is transcriptionally related or similar, and by analogy related in etiology, is relatively limited. Here we demonstrate through blind, comparative transcriptome analysis of human and mouse breast cancers that many of the most significantly differentially expressed genes in the primary tumors of human females with breast cancer, when compared to normal breast tissue, are also the most differentially expressed genes in the primary tumors of five separate mouse models of breast cancer specifically induced by integration events that mimic viral infection. These genes, with functions in the cell cycle, DNA replication and growth factor signaling²⁵⁻³⁰, are also among the most differentially expressed genes following transformation of the mammary gland induced by deletion of BRCA1 and haploinsufficiency of p53. Thus, primary tumors of patients with breast cancer, breast tumors induced by modeling of viral infection in mice, and tumors in a genetically engineered mouse model of familial breast cancer share strong transcriptional identity, providing one piece of evidence that a virus or multiple viruses might contribute to the development of breast cancers in humans.

Methods

We utilized datasets GSE124648⁷, GSE109169⁸, GSE23938⁹ and for this differential gene expression analysis of female breast cancer in conjunction with GEO2R. GSE124648 was generated using Affymetrix Human Genome U133A Array technology with $n=10$ normal breast tissue and $n=126$ tumors of the breast and; analysis was performed using platform GPL96. GSE109169 was generated using Affymetrix Human Exon 1.0 ST Array technology with $n=25$ normal breast tissue and $n=25$ tumors of the breast and; analysis was performed using platform GPL5175. The tissues from this dataset are paired tissues (25 tumors matching 25 breast tissues from 25 patients). GSE23938 was generated using Affymetrix Mouse Genome 430A 2.0 Array technology with $n=5$ benign mammary tissue from pregnant females and $n=6$ breast tumors from MMTV-PyMT mice expressing the middle T antigen of polyomavirus under the control of the MMTV promoter on an FVB strain, $n=4$ breast tumors from MMTV-Wnt1 mice on an FVB strain, $n=5$ breast tumors from SV40 large T-antigen-expressing mice (C3TA_g), driving

expression of the large T-antigen of SV40 under the control of the 5' flanking region of the C3(1) component of the rat prostate steroid binding protein on an FVB strain, $n=4$ breast tumors from MMTV-C-Myc mice on an FVB strain, $n=5$ breast tumors from MMTV-HRas mice on an FVB strain, and $n=5$ breast tumors from BRCA1 $-/-$ p53 $-/+$ MMTV-Cre transgenic mice on a 129B6/FVB strain; 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. A statistical test was performed to evaluate whether target gene expression was significantly between primary breast tumors and breast tissue in humans with breast cancer using a two-tailed, unpaired t-test with Welch's correction. A statistical test was performed to evaluate whether target gene expression was significantly between primary breast tumors and breast tissue in mice with breast cancer induced by genetic engineering of viral integration events or in a genetically engineered mouse model of breast cancer using a one-way analysis of variance (ANOVA) with Dunnett's multiple comparisons test. We used PRISM for all statistical analyses of differential gene expression in human and murine breast cancer (Version 8.4.0)(455).

Results

We performed a 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 developing breast cancers following genetic engineering of integration events or oncogene expression during viral infection, and in mice genetically engineered to model familial breast cancer.

The scavenger receptor CD36 is among the most differentially expressed genes in human breast cancer, its expression is significantly decreased in patient tumors as compared to normal breast, these transcriptional features are essentially identical in five separate mouse models of breast cancer induced by transgenic modeling of viral infection or transgenic expression of a viral oncoprotein, and this is a conserved transcriptional feature of a genetically engineered mouse model of BRCA1-deficient inherited breast cancer.

We identified CD36 as among the genes most differentially expressed in the primary tumors of females with breast cancer. When sorting each of the genes expressed in humans with breast cancer based on significance of difference between primary tumors and normal breast tissue, CD36 ranked 171 out of 19076 (Table 1A). Differential expression of CD36 in human breast cancer was statistically significant (Table 1A; $p=1.78E-13$). Next we compared benign mammary tissue to tumors from five separate models of breast cancer induced in mice by genetic modeling of integration events during viral infection by the mouse mammary tumor virus (MMTV) or by transgenic expression of a viral oncoprotein under the control of endogenous regulatory elements: MMTV-PyMT, MMTV-Wnt1, MMTV-C-Myc, MMTV-HRas, and the C3Tag mouse model, which utilizes transgenic expression of the large T-antigen of SV40 under the control of the rat prostate steroid binding gene C3(1). Included in this

analysis were primary tumors from a sixth mouse model, genetically engineered to harbor deletions in the breast cancer susceptibility gene BRCA1 as well as loss of one copy of p53. CD36 was also among the most differentially expressed genes transcriptome-wide when comparing normal mammary tissues to primary tumors from each of these virally-induced or genetically engineered mouse models (Table 1B). When sorting each of the genes expressed in mice based on significance of change in expression between normal murine mammary tissue and tumors from the mouse breast experimentally induced by viral infection, expression of the large T-antigen, or combined BRCA1 deletion and p53 haploinsufficiency, CD36 ranked 230 out of 18586 total transcripts. Differential expression of CD36 in primary murine breast tumors was statistically significant (Table 1B; $p=5.38E-09$). CD36 was expressed at significantly lower levels in primary tumors from females with breast cancer as compared to normal breast tissue (Figure 1; $p<0.0001$). Decreased expression of CD36 was conserved in each of five mouse models of breast cancer induced by viral infection or expression of a viral oncoprotein (Figure 1; $p<0.0001$ for MMTV-PyMT, $p<0.0001$ for MMTV-Wnt1, $p<0.0001$ for C3TA_g, $p<0.0001$ for MMTV-C-Myc, $p<0.0001$ for MMTV-HRas). Differential, decreased expression of CD36 in primary breast tumors from BRCA1 $-/-$ p53 $-/+$ mice was also statistically significant (Figure 1; $p<0.0001$).

Abnormal spindle-like microcephaly associated, ASPM, is among the most differentially expressed genes in human breast cancer, its expression is significantly increased in patient tumors as compared to normal breast, these transcriptional features are shared by some mouse models of breast cancer induced by transgenic modeling of viral infection or transgenic expression of a viral oncoprotein, and this is a conserved transcriptional feature of a genetically engineered mouse model of BRCA1-deficient inherited breast cancer.

Next we identified the abnormal spindle-like microcephaly associated, ASPM, as among the genes most differentially expressed in the primary tumors of females with breast cancer. When sorting each of the genes expressed in humans with breast cancer based on significance of difference between primary tumors and normal breast tissue, ASPM ranked 207 out of 22283 (Table 2A). Differential expression of ASPM in human breast cancer was statistically significant (Table 2A; $p=2.35E-11$). Next we compared benign mammary tissue to tumors from five separate models of breast cancer induced in mice by genetic modeling of integration events during viral infection by the mouse mammary tumor virus (MMTV) or by transgenic expression of a viral oncoprotein. ASPM was also among the most differentially expressed genes transcriptome-wide when comparing normal mammary tissues to primary tumors from each of these virally-induced or genetically engineered mouse models (Table 2B). When sorting each of the genes expressed in mice based on significance of change in expression between normal murine mammary tissue and tumors from the mouse breast experimentally induced by viral infection, expression of the large T-antigen, or combined BRCA1 deletion and p53 haploinsufficiency, ASPM ranked 80 out of 18586 total transcripts. Differential expression of ASPM in primary murine breast tumors was statistically significant (Table 2B; $p=8.36E-11$). ASPM was expressed at significantly higher levels in primary tumors from females with breast cancer as compared to normal breast tissue (Figure 2; $p<0.0001$). Increased expression of ASPM was conserved in the C3TA_g mouse models (Figure 2; $p<0.0001$) and in the MMTV-C-Myc mouse model (Figure 2; $p=0.0369$) but not in the MMTV-PyMT mouse model (Figure 2;

p=0.9998) nor in the MMTV-Wnt1 (Figure 2; p=0.9997) or MMTV-HRas mouse models (Figure 2; p=0.9445). Increased expression of ASPM in primary breast tumors from BRCA1 -/- p53 +/- mice was conserved and statistically significant (Figure 2; p=0.0267).

The platelet-derived growth factor D, PDGFD, is among the most differentially expressed genes in human breast cancer, its expression is significantly decreased in patient tumors as compared to normal breast, these transcriptional features are essentially identical in five separate mouse models of breast cancer induced by transgenic modeling of viral infection or transgenic expression of a viral oncoprotein, and this is a conserved transcriptional feature of a genetically engineered mouse model of BRCA1-deficient inherited breast cancer.

We also identified the platelet-derived growth factor D, PDGFD as among the genes most differentially expressed in the primary tumors of females with breast cancer. When sorting each of the genes expressed in humans with breast cancer based on significance of difference between primary tumors and normal breast tissue, PDGFD ranked 157 out of 22283 (Table 3A). Differential expression of PDGFD in human breast cancer was statistically significant (Table 3A; $p=5.48E-12$). Next we compared benign mammary tissue to tumors from five separate models of breast cancer induced in mice by genetic modeling of integration events during viral infection by the mouse mammary tumor virus (MMTV) or by transgenic expression of a viral oncoprotein. PDGFD was also among the most differentially expressed genes transcriptome-wide when comparing normal mammary tissues to primary tumors from each of these virally-induced or genetically engineered mouse models (Table 3B). When sorting each of the genes expressed in mice based on significance of change in expression between benign murine mammary tissue and tumors from the mouse breast experimentally induced by genetic modeling of viral infection, expression of the large T-antigen, or combined BRCA1 deletion and p53 haploinsufficiency, PDGFD ranked 121 out of 18586 total transcripts. Differential expression of PDGFD in primary murine breast tumors was statistically significant (Table 3; $p=6.54E-10$). PDGFD was expressed at significantly lower levels in primary tumors from females with breast cancer as compared to normal breast tissue (Figure 3; $p<0.0001$). Decreased expression of PDGFD was conserved in each of five mouse models of breast cancer induced by viral infection or expression of a viral oncoprotein (Figure 3; $p<0.0001$ for MMTV-PyMT, $p<0.0001$ for MMTV-Wnt1, $p<0.0001$ for C3TA_g, $p<0.0001$ for MMTV-C-Myc, $p<0.0001$ for MMTV-HRas). Differential, decreased expression of PDGFD in primary breast tumors from BRCA1 -/- p53 +/- mice was also statistically significant (Figure 3; $p<0.0001$).

Minichromosome maintenance 4, MCM4, is among the most differentially expressed genes in human breast cancer, its expression is significantly increased in patient tumors as compared to normal breast, these transcriptional features are essentially identical in five separate mouse models of breast cancer induced by transgenic modeling of viral infection or transgenic expression of a viral oncoprotein, and this is a conserved transcriptional feature of a genetically engineered mouse model of BRCA1-deficient inherited breast cancer.

We also identified the minichromosome maintenance 4, MCM4, as among the genes most differentially expressed in the primary tumors of females with breast cancer. When sorting each of the genes expressed in humans with breast cancer based on significance of difference between primary tumors and normal breast tissue, MCM4 ranked 250 out of 22283 (Table 4A). Differential expression of MCM4 in human breast cancer was statistically significant (Table 4A; $p=6.66E-11$). Next we compared benign mammary tissue to tumors from five separate models of breast cancer induced in mice by genetic modeling of integration events during viral infection by the mouse mammary tumor virus (MMTV) or by transgenic expression of a viral oncoprotein. MCM4 was also among the most differentially expressed genes transcriptome-wide when comparing normal mammary tissues to primary tumors from each of these genetically engineered mouse models mimicking viral infection or genetic changes found in familial breast cancer (Table 4B). When sorting each of the genes expressed in mice based on significance of change in expression between normal murine mammary tissue and tumors from the mouse breast experimentally induced by viral infection, expression of the large T-antigen, or combined BRCA1 deletion and p53 haploinsufficiency, MCM4 ranked 122 out of 18586 total transcripts. Differential expression of MCM4 in primary murine breast tumors was statistically significant (Table 4B; $p=7.07E-10$). MCM4 was expressed at significantly lower levels in primary tumors from females with breast cancer as compared to normal breast tissue (Figure 4; $p=p<0.0001$). Decreased expression of MCM4 was conserved in each of five mouse models of breast cancer induced by viral infection or expression of a viral oncoprotein (Figure 4; $p<0.0001$ for MMTV-PyMT, $p=0.0002$ for MMTV-Wnt1, $p<0.0001$ for C3Tag, $p=0.0004$ for MMTV-C-Myc, $p=0.0661$ for MMTV-HRas). Differential, decreased expression of MCM4 in primary breast tumors from BRCA1 -/- p53 +/- mice was also statistically significant (Figure 4; $p=0.0004$).

Kappa-casein, CSN3, is among the most differentially expressed genes in human breast cancer, its expression is significantly decreased in patient tumors as compared to normal breast, these transcriptional features are essentially identical in five separate mouse models of breast cancer induced by transgenic modeling of viral infection or transgenic expression of a viral oncoprotein, and this is a conserved transcriptional feature of a genetically engineered mouse model of BRCA1-deficient inherited breast cancer.

Next we identified the abnormal spindle-like microcephaly associated, CSN3, as among the genes most differentially expressed in the primary tumors of females with breast cancer. When sorting each of the genes expressed in humans with breast cancer based on significance of difference between primary tumors and normal breast tissue, CSN3 ranked 138 out of 22283 (Table 5A). Differential expression of CSN3 in human breast cancer was statistically significant (Table 5A; $p=1.16E-09$). Next we compared benign mammary tissue to tumors from five separate models of breast cancer induced in mice by genetic modeling of integration events during viral infection by the mouse mammary tumor virus (MMTV) or by transgenic expression of a viral oncoprotein. CSN3 was also among the most differentially expressed genes transcriptome-wide when comparing normal mammary tissues to primary tumors from each of these virally-induced or genetically engineered mouse models (Table 5B). When sorting each of the genes expressed in mice based on significance of change in expression between normal

murine mammary tissue and tumors from the mouse breast experimentally induced by viral infection, expression of the large T-antigen, or combined BRCA1 deletion and p53 haploinsufficiency, CSN3 ranked 53 out of 18586 total transcripts. Differential expression of CSN3 in primary murine breast tumors was statistically significant (Table 5B; $p=4.48E-15$). CSN3 was expressed at significantly lower levels in primary tumors from females with breast cancer as compared to normal breast tissue (Figure 5; $p=0.0034$). Decreased expression of CSN3 was conserved in each of mouse models of breast cancer induced by viral infection or expression of a viral oncoprotein except for MMTV-PyMT (Figure 5; $p=0.7640$ for MMTV-PyMT, $p<0.0001$ for MMTV-Wnt1, $p<0.0001$ for C3TA γ , $p=0.0054$ for MMTV-C-Myc, and $p=0.0471$ for MMTV-HRas). Differential, decreased expression of CSN3 in primary breast tumors from BRCA1 $-/-$ p53 $-/+$ mice was also statistically significant (Figure 5; $p<0.0001$).

Centrosomal protein 55, CEP55, is among the most differentially expressed genes in human breast cancer, its expression is significantly increased in patient tumors as compared to normal breast, these transcriptional features are essentially identical in five separate mouse models of breast cancer induced by transgenic modeling of viral infection or transgenic expression of a viral oncoprotein, and this is a conserved transcriptional feature of a genetically engineered mouse model of BRCA1-deficient inherited breast cancer.

Next we identified the abnormal spindle-like microcephaly associated, CEP55, as among the genes most differentially expressed in the primary tumors of females with breast cancer. When sorting each of the genes expressed in humans with breast cancer based on significance of difference between primary tumors and normal breast tissue, CEP55 ranked 247 out of 22283 (Table 6A). Differential expression of CEP55 in human breast cancer was statistically significant (Table 6A; $p=6.31E-09$). Next we compared benign mammary tissue to tumors from five separate models of breast cancer induced in mice by genetic modeling of integration events during viral infection by the mouse mammary tumor virus (MMTV) or by transgenic expression of a viral oncoprotein. CEP55 was also among the most differentially expressed genes transcriptome-wide when comparing normal mammary tissues to primary tumors from each of these virally-induced or genetically engineered mouse models (Table 6B). When sorting each of the genes expressed in mice based on significance of change in expression between normal murine mammary tissue and tumors from the mouse breast experimentally induced by viral infection, expression of the large T-antigen, or combined BRCA1 deletion and p53 haploinsufficiency, CEP55 ranked 179 out of 18586 total transcripts. Differential expression of CEP55 in primary murine breast tumors was statistically significant (Table 6B; $p=1.11E-11$). CEP55 was expressed at significantly higher levels in primary tumors from females with breast cancer as compared to normal breast tissue (Figure 6; $p<0.0001$). Increased expression of CEP55 was conserved in each of mouse models of breast cancer induced by viral infection or expression of a viral oncoprotein except for MMTV-PyMT (Figure 6; $p<0.0001$ for MMTV-PyMT, $p<0.0001$ for MMTV-Wnt1, $p<0.0001$ for C3TA γ , $p=0.0003$ for MMTV-C-Myc, and $p<0.0001$ for MMTV-HRas). CEP55 was also expressed at significantly higher levels in primary breast tumors from BRCA1 $-/-$ p53 $-/+$ mice as compared to normal mammary tissue and this was also statistically significant (Figure 6; $p<0.0001$).

The non-SMC condensin I complex, subunit G, NCAPG is among the most differentially expressed genes in human breast cancer, its expression is significantly increased in patient tumors as compared to normal breast, these transcriptional features are essentially identical in five separate mouse models of breast cancer induced by transgenic modeling of viral infection or transgenic expression of a viral oncoprotein, and this is a conserved transcriptional feature of a genetically engineered mouse model of BRCA1-deficient inherited breast cancer.

We also identified the non-SMC condensin I complex, NCAPG, as among the genes most differentially expressed in the primary tumors of females with breast cancer. When sorting each of the genes expressed in humans with breast cancer based on significance of difference between primary tumors and normal breast tissue, NCAPG ranked 48 out of 22283 (Table 7A). Differential expression of NCAPG in human breast cancer was statistically significant (Table 7A; $p=2.64E-15$). Next we compared benign mammary tissue to tumors from five separate models of breast cancer induced in mice by genetic modeling of integration events during viral infection by the mouse mammary tumor virus (MMTV) or by transgenic expression of a viral oncoprotein. NCAPG was also among the most differentially expressed genes transcriptome-wide when comparing normal mammary tissues to primary tumors from each of these virally-induced or genetically engineered mouse models (Table 7B). When sorting each of the genes expressed in mice based on significance of change in expression between normal murine mammary tissue and tumors from the mouse breast experimentally induced by viral infection, expression of the large T-antigen, or combined BRCA1 deletion and p53 haploinsufficiency, NCAPG ranked 250 out of 18586 total transcripts. Differential expression of NCAPG in primary murine breast tumors was statistically significant (Table 7B; $p=6.56E-09$). NCAPG was expressed at significantly higher levels in primary tumors from females with breast cancer as compared to normal breast tissue (Figure 7; $p=0.0016$). Increased expression of NCAPG was conserved and statistically significant in each of mouse models of breast cancer induced by viral infection or expression of a viral oncoprotein except for MMTV-Wnt1 and MMTV-HRas (Figure 7; $p<0.0001$ for MMTV-PyMT, $p=0.3543$ for MMTV-Wnt1, $p<0.0001$ for C3TA, $p<0.0001$ for MMTV-C-Myc, and $p=0.0802$ for MMTV-HRas). Differential, decreased expression of NCAPG in primary breast tumors from BRCA1 $-/-$ p53 $-/+$ mice was also statistically significant (Figure 7; $p=0.0009$).

Molybdenum cofactor synthesis 1, MOCS1 is among the most differentially expressed genes in human breast cancer, its expression is significantly decreased in patient tumors as compared to normal breast, these transcriptional features are essentially identical in five separate mouse models of breast cancer induced by transgenic modeling of viral infection or transgenic expression of a viral oncoprotein, and this is a conserved transcriptional feature of a genetically engineered mouse model of BRCA1-deficient inherited breast cancer.

Molybdenum cofactor synthesis, MOCS1, as among the genes most differentially expressed in the primary tumors of females with breast cancer. When sorting each of the genes expressed in humans with breast cancer based on significance of difference between primary

tumors and normal breast tissue, MOCS1 ranked 80 out of 19076 (Table 8A). Differential expression of MOCS1 in human breast cancer was statistically significant (Table 8A; $p=1.1E+01$). Next we compared benign mammary tissue to tumors from five separate models of breast cancer induced in mice by genetic modeling of integration events during viral infection by the mouse mammary tumor virus (MMTV) or by transgenic expression of a viral oncoprotein. MOCS1 was also among the most differentially expressed genes transcriptome-wide when comparing normal mammary tissues to primary tumors from each of these virally-induced or genetically engineered mouse models (Table 8B). When sorting each of the genes expressed in mice based on significance of change in expression between normal murine mammary tissue and tumors from the mouse breast experimentally induced by viral infection, expression of the large T-antigen, or combined BRCA1 deletion and p53 haploinsufficiency, MOCS1 ranked 28 out of 18586 total transcripts. Differential expression of MOCS1 in primary murine breast tumors was statistically significant (Table 8B; $p=4.45E-13$). MOCS1 was expressed at significantly lower levels in primary tumors from females with breast cancer as compared to normal breast tissue (Figure 8; $p<0.0001$). Decreased expression of MOCS1 was conserved in each mouse models of breast cancer induced by viral infection or expression of a viral oncoprotein (Figure 8; $p<0.0001$ for MMTV-PyMT, $p<0.0001$ for MMTV-Wnt1, $p<0.0001$ for C3TA γ , $p<0.0001$ for MMTV-C-Myc, and $p=0.0005$ for MMTV-HRas). MOCS1 expression in BRCA1 $-/-$ p53 $-/+$ mice was decreased as compared to normal mammary tissue and this was also statistically significant (Figure 8; $p<0.0001$).

Thus, by performing blind, whole transcriptome analysis of primary tumors of the breast to normal breast tissue, we found recurrent transcriptional changes that were not only among the most significant transcriptional changes transcriptome-wide in breast cancer, but of importance, these changes were found in the primary breast tumors of mice who developed breast cancer as the result of experimental viral infections, and were also among the most significant transcriptional changes transcriptome-wide in mice with breast cancer caused by viral infection. Thus, critical transcriptional changes induced by viral infection in mice, in five separate mouse models of virally-induced breast cancer, are the same transcriptional changes found in the primary tumors of humans with breast cancer. Moreover, in most cases, these changes were found in the primary breast tumors of mice who developed breast cancer due to deletion of the breast cancer susceptibility gene BRCA1 in conjunction with loss of one copy of the tumor suppressor p53, demonstrating that some of the most significant transcriptional features of breast tumors in an experimental mouse model of familial breast cancer are found not only in human breast cancer but also shared among breast cancer resulting from genetic modeling of integration events viral infection in mice. We conclude that human breast cancer and breast cancer resulting from integration events and oncogene expression that occur during viral infection in mice are transcriptionally similar.

Discussion

Multiple studies have attempted to provide evidence of a viral cause for breast cancer in humans by assessing the presence of viral transcription or of viral antigen production. In this

1 study, we harnessed the power of three separate published microarray datasets⁷⁻⁹ to determine
2 in an unbiased fashion and at the transcriptome level what were the most significant
3 transcriptional in the primary tumors of mice that develop breast cancer as a result of genetic
4 modeling of integration events during viral infection or transgenic expression of a viral
oncoprotein, and whether any of these changes could be identified in the primary tumors of
humans with breast cancer.

5 We identified eight critical changes that were among the most significant transcriptional
6 changes in the tumors of mice with breast cancer resulting from genetic modeling of infection
7 with a mammary tumor virus or transgenic expression of a viral oncogene, that were recurrently
8 found in the tumors of humans with breast cancer. These included changes expression of the
genes CD36, ASPM, MCOS1, NCAPG, CEP55, PDGFD, MCM4, and CSN3. Several of these
genes have established roles as important molecules in human breast cancer²⁵⁻³⁰.

9 We provided evidence in this study that at the level of transcription, human breast
10 cancers share significant similarity with the tumors of mice with breast cancer resulting from
11 genetic modeling of integration events or oncogene expression during viral infection.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

References

1. 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.
2. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M. and Ding, W., 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science*, 266(5182), pp.66-71.
3. 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.
4. 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.
5. Glenn, W.K., Heng, B., Delprado, W., Iacopetta, B., Whitaker, N.J. and Lawson, J.S., 2012. Epstein-Barr virus, human papillomavirus and mouse mammary tumour virus as multiple viruses in breast cancer. *PloS one*, 7(11), p.e48788.
6. Joshi, D. and Buehring, G.C., 2012. Are viruses associated with human breast cancer? Scrutinizing the molecular evidence. *Breast cancer research and treatment*, 135(1), pp. 1-15.
7. 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.
8. 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.
9. Zhu, M., Yi, M., Kim, C.H., Deng, C., Li, Y., Medina, D., Stephens, R.M. and Green, J.E., 2011. Integrated miRNA and mRNA expression profiling of mouse mammary tumor models identifies miRNA signatures associated with mammary tumor lineage. *Genome biology*, 12(8), p.R77.

10. Kan, C.Y., Iacopetta, B.J., Lawson, J.S. and Whitaker, N.J., 2005. Identification of human papillomavirus DNA gene sequences in human breast cancer. *British Journal of Cancer*, 93(8), pp.946-948.
11. Hogan, T.F., Padgett, B.L., Walker, D.L., Borden, E.C. and Frias, Z., 1983. Survey of human polyomavirus (JCV, BKV) infections in 139 patients with lung cancer, breast cancer, melanoma, or lymphoma. *Progress in clinical and biological research*, 105, pp.311-324.
12. Martini, F., Mazzoni, E., Corallini, A., Taronna, A., Querzoli, P., Magri, E., Marci, R., Dolcetti, R., Rezza, G., Barbanti-Brodano, G. and Tognon, M., 2013. Breast cancer and simian virus 40 infection. *Epidemiology*, 24(3), pp.464-465.
13. Sitas, F., Carrara, H., Beral, V., Newton, R., Reeves, G., Bull, D., Jentsch, U., Pacella-Norman, R., Bourboulia, D., Whitby, D. and Boshoff, C., 1999. Antibodies against human herpesvirus 8 in black South African patients with cancer. *New England Journal of Medicine*, 340(24), pp.1863-1871.
14. Labrecque, L.G., Barnes, D.M., Fentiman, I.S. and Griffin, B.E., 1995. Epstein-Barr virus in epithelial cell tumors: a breast cancer study. *Cancer research*, 55(1), pp.39-45.
15. Richardson, A.K., Walker, L.C., Cox, B., Rollag, H., Robinson, B.A., Morrin, H., Pearson, J.F., Potter, J.D., Paterson, M., Surcel, H.M. and Pukkala, E., 2020. Breast cancer and cytomegalovirus. *Clinical and Translational Oncology*, 22(4), pp.585-602.
16. Wang, Y., Melana, S.M., Baker, B., Bleiweiss, I., Fernandez-Cobo, M., Mandeli, J.F., Holland, J.F. and Pogo, B.G.T., 2003. High prevalence of MMTV-like env gene sequences in gestational breast cancer. *Medical Oncology*, 20(3), pp.233-236.
17. 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.
18. 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.
19. 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.
20. 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.

- 1
- 2 21. Wang-Johanning, F., Frost, A.R., Jian, B., Epp, L., Lu, D.W. and Johanning, G.L., 2003.
- 3 Quantitation of HERV-K env gene expression and splicing in human breast
- 4 cancer. *Oncogene*, 22(10), pp.1528-1535.
- 5 22. Wang-Johanning, F., Frost, A.R., Johanning, G.L., Khazaeli, M.B., LoBuglio, A.F., Shaw,
- 6 D.R. and Strong, T.V., 2001. Expression of human endogenous retrovirus k envelope
- 7 transcripts in human breast cancer. *Clinical Cancer Research*, 7(6), pp.1553-1560.
- 8 23. Martin, G.S., 1970. Rous sarcoma virus: a function required for the maintenance of the
- 9 transformed state. *Nature*, 227, pp.1021-1023.
- 10 24. Kim, C.M., Koike, K., Saito, I., Miyamura, T. and Jay, G., 1991. HBx gene of hepatitis B
- 11 virus induces liver cancer in transgenic mice. *Nature*, 351(6324), pp.317-320.
- 12 25. Pascual, G., Avgustinova, A., Mejetta, S., Martín, M., Castellanos, A., Attolini, C.S.O.,
- 13 Berenguer, A., Prats, N., Toll, A., Hueto, J.A. and Bescós, C., 2017. Targeting metastasis-
- 14 initiating cells through the fatty acid receptor CD36. *Nature*, 541(7635), pp.41-45.
- 15 26. Jiang, L., Ren, L., Chen, H., Pan, J., Zhang, Z., Kuang, X., Chen, X., Bao, W., Lin, C., Zhou,
- 16 Z. and Huang, D., 2020. NCAPG confers trastuzumab resistance via activating SRC/STAT3
- 17 signaling pathway in HER2-positive breast cancer. *Cell death & disease*, 11(7), pp.1-12.
- 18 27. Liu, J., Liao, S., Huang, Y., Samuel, R., Shi, T., Naxerova, K., Huang, P., Kamoun, W., Jain,
- 19 R.K., Fukumura, D. and Xu, L., 2011. PDGF-D improves drug delivery and efficacy via
- 20 vascular normalization, but promotes lymphatic metastasis by activating CXCR4 in breast
- 21 cancer. *Clinical Cancer Research*, 17(11), pp.3638-3648.
- 22 28. Tang, J., Lu, M., Cui, Q., Zhang, D., Kong, D., Liao, X., Ren, J., Gong, Y. and Wu, G., 2019.
- 23 Overexpression of ASPM, CDC20, and TTK confer a poorer prognosis in breast cancer
- 24 identified by gene co-expression network analysis. *Frontiers in oncology*, 9, p.310.
- 25 29. Shima, N., Alcaraz, A., Liachko, I., Buske, T.R., Andrews, C.A., Munroe, R.J., Hartford, S.A.,
- 26 Tye, B.K. and Schimenti, J.C., 2007. A viable allele of Mcm4 causes chromosome instability
- 27 and mammary adenocarcinomas in mice. *Nature genetics*, 39(1), pp.93-98.
- 28 30. Kalimutho, M., Sinha, D., Jeffery, J., Nones, K., Srihari, S., Fernando, W.C., Duijf, P.H.,
- Vennin, C., Raninga, P., Nanayakkara, D. and Mittal, D., 2018. CEP 55 is a determinant of
- cell fate during perturbed mitosis in breast cancer. *EMBO molecular medicine*, 10(9),
- p.e8566.

Rank	ID	p-value	t	B	Gene	Gene name
171	3010503	1.78E-13	-9.91	20.496206	CD36	CD36 molecule (thrombospondin receptor)

Table 1A: CD36 in differentially expressed in human breast cancer.

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

Rank	ID	p-value	F	Gene	Gene name
230	1450884_at	5.38E-09	19.54694	Cd36	CD36 antigen

Table 1B: CD36 in differentially expressed in mouse models of breast cancer induced by genetic modeling of viral infection.

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

Rank	ID	p-value	t	B	Gene	Gene name
207	219918_s_at	2.35E-11	7.2900218	15.47901	ASPM	abnormal spindle microtubule assembly

Table 2A: ASPM in differentially expressed in human breast cancer.

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

Rank	ID	p-value	F	Gene	Gene name
80	1452459_at	8.36E-11	27.8954	Aspm	asp (abnormal spindle)-like, microcephaly associated (Drosophila)

Table 2B: ASPM in differentially expressed in mouse models of breast cancer induced by genetic modeling of viral infection.

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

Rank	ID	p-value	t	B	Gene	Gene name
157	219304_s_at	5.48E-12	-7.5617051	16.89095	PDGFD	platelet derived growth factor D

Table 3A: PDGFD in differentially expressed in human breast cancer.

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

Rank	ID	p-value	F	Gene	Gene name
121	1426319_at	6.54E-10	23.47491	Pdgfd	platelet-derived growth factor, D polypeptide

Table 3B: PDGFD in differentially expressed in mouse models of breast cancer induced by genetic modeling of viral infection.

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

Rank	ID	p-value	t	B	Gene	Gene name
250	222037_at	6.66E-11	7.0935796	14.47174	MCM4	minichromosome maintenance complex component 4

Table 4A: MCM4 in differentially expressed in human breast cancer.

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

Rank	ID	p-value	F	Gene	Gene name
122	1416214_at	7.07E-10	23.31907	Mcm4	minichromosome maintenance complex component 4

Table 4B: MCM4 in differentially expressed in mouse models of breast cancer induced by genetic modeling of viral infection.

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

Rank	ID	p-value	F	Gene	Gene name
138	1451851_a_at	1.16E-09	22.35453	Csn3	casein kappa

Table 5A: CSN3 in differentially expressed in human breast cancer.

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

Rank	ID	p-value	t	B	Gene	Gene name
53	207803_s_at	4.48E-15	-8.8436986	23.79174	CSN3	casein kappa

Table 5B: CSN3 in differentially expressed in mouse models of breast cancer induced by genetic modeling of viral infection.

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

Rank	ID	p-value	F	Gene	Gene name
247	1452242_at	6.31E-09	19.27119	Cep55	centrosomal protein 55

Table 6A: CEP55 in differentially expressed in human breast cancer.

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

Rank	ID	p-value	t	B	Gene	Gene name
179	218542_at	1.11E-11	7.4299541	16.20361	CEP55	centrosomal protein 55

Table 6B: CEP55 in differentially expressed in mouse models of breast cancer induced by genetic modeling of viral infection.

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

Rank	ID	p-value	t	B	Gene	Gene name
48	218662_s_at	2.64E-15	8.9368446	24.30524	NCAPG	non-SMC condensin I complex subunit G

Table 7A: NCAPG in differentially expressed in human breast cancer.

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

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

Rank	ID	p-value	F	Gene	Gene name
250	1429172_a_at	6.56E-09	19.20425	Ncapg	non-SMC condensin I complex, subunit G

Table 7B: NCAPG in differentially expressed in mouse models of breast cancer induced by genetic modeling of viral infection.

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

Rank	ID	p-value	t	B	Gene	Gene name
80	2953139	4.09E-15	-1.1E+01	24.209175	MOCS1	Molybdenum cofactor synthesis 1

Table 8A: MOCS1 in differentially expressed in human breast cancer.

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

Rank	ID	p-value	F	Gene	Gene name
28	1425004_s_at	4.45E-13	42.41997	Mocs1	molybdenum cofactor synthesis 1

Table 8B: MOCS1 in differentially expressed in mouse models of breast cancer induced by genetic modeling of viral infection.

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

Figure 1: Conserved down-regulation of CD36 in human breast cancer and in mouse models of breast cancer resulting from genetic modeling of viral infection.

The expression of CD36 in normal breast tissue and in primary breast tumors from human patients (left) and in normal mammary tissue and in five models of breast cancer induced by genetic modeling of viral infection or expression of a viral oncoprotein, as well as in a genetically engineered mouse model of familial breast cancer (below) is graphically represented here with the results of a statistical test evaluating the significance of difference in CD36 expression when comparing normal breast tissue to that of breast tumors in humans and in mice.

Figure 2: Semi-conserved up-regulation of ASPM in human breast cancer and in mouse models of breast cancer induced by genetic modeling of viral infection.

The expression of ASPM in normal breast tissue and in primary breast tumors from human patients (left) and in normal mammary tissue and in five models of breast cancer induced by genetic modeling of viral infection or expression of a viral oncoprotein, as well as in a genetically engineered mouse model of familial breast cancer (below) is graphically represented here with the results of a statistical test evaluating the significance of difference in ASPM expression when comparing normal breast tissue to that of breast tumors in humans and in mice.

Figure 3: Conserved down-regulation of PDGFD in human breast cancer and in mouse models of breast cancer induced by genetic modeling of viral infection.

The expression of PDGFD in normal breast tissue and in primary breast tumors from human patients (left) and in normal mammary tissue and in five models of breast cancer induced by genetic modeling of viral infection or expression of a viral oncoprotein, as well as in a genetically engineered mouse model of familial breast cancer (below) is graphically represented here with the results of a statistical test evaluating the significance of difference in PDGFD expression when comparing normal breast tissue to that of breast tumors in humans and in mice.

Figure 4: Conserved up-regulation of MCM4 in human breast cancer and in mouse models of breast cancer induced by genetic modeling of viral infection.

The expression of MCM4 in normal breast tissue and in primary breast tumors from human patients (left) and in normal mammary tissue and in five models of breast cancer induced by genetic modeling of viral infection or expression of a viral oncoprotein, as well as in a genetically engineered mouse model of familial breast cancer (below) is graphically represented here with the results of a statistical test evaluating the significance of difference in MCM4 expression when comparing

normal breast tissue to that of breast tumors in humans and in mice.

Figure 5: Conserved down-regulation of CSN3 in human breast cancer and in mouse models of breast cancer induced by genetic modeling of viral infection.

The expression of CSN3 in normal breast tissue and in primary breast tumors from human patients (left) and in normal mammary tissue and in five models of breast cancer induced by genetic modeling of viral infection or expression of a viral oncoprotein, as well as in a genetically engineered mouse model of familial breast cancer (below) is graphically represented here with the results of a statistical test evaluating the significance of difference in CSN3 expression when comparing normal breast tissue to that of

breast tumors in humans and in mice.

Figure 6: Conserved up-regulation of CEP55 in human breast cancer and in mouse models of breast cancer induced by genetic modeling of viral infection.

The expression of CEP55 in normal breast tissue and in primary breast tumors from human patients (left) and in normal mammary tissue and in five models of breast cancer induced by genetic modeling of viral infection or expression of a viral oncoprotein, as well as in a genetically engineered mouse model of familial breast cancer (below) is graphically represented here with the results of a statistical test evaluating the significance of difference in CEP55 expression when comparing normal breast tissue to that of breast tumors in humans and in mice.

Figure 7: Semi-conserved up-regulation of NCAPG in human breast cancer and in mouse models of breast cancer induced by genetic modeling of viral infection.

The expression of NCAPG in normal breast tissue and in primary breast tumors from human patients (left) and in normal mammary tissue and in five models of breast cancer induced by genetic modeling of viral infection or expression of a viral oncoprotein, as well as in a genetically engineered mouse model of familial breast cancer (below) is graphically represented here with the results of a statistical test evaluating the significance of difference in NCAPG expression when comparing normal breast tissue to that of breast tumors in humans and in mice.

Figure 8: Conserved down-regulation of MOCS1 in human breast cancer and in mouse models of breast cancer induced by genetic modeling of viral infection.

The expression of MOCS1 in normal breast tissue and in primary breast tumors from human patients (left) and in normal mammary tissue and in five separate models of breast cancer induced by genetic modeling of viral infection or expression of a viral oncoprotein, as well as in a genetically engineered mouse model of familial breast cancer (below) is graphically represented here with the results of a statistical test evaluating the significance of difference in MOCS1 expression when comparing normal breast tissue

to that of breast tumors in humans and in mice.