A DREAM complex gene expression program involving E2F4, E2F6, SIN3A and NFYB in human breast cancer and in murine breast cancers arising from genetic modeling of transformation induced by viral infection.

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

We recently reported that nearly half of the most differentially expressed genes in the the breast tumors of mice resulting from genetic modeling of integration events or oncogene expression occurring during viral infection by viruses including the murine mammary tumor virus, SV40 and polyomavirus are also significantly differentially expressed in human breast cancer, the majority of which are of unknown etiology^{1,2}. Here we present evidence of the existence of a gene expression program operating through the DREAM complex, E2F transcription factors E2F4 and E2F6, as well as NFYB and SIN3A, both in the tumors of mice resulting from genetic modeling of viral infection and in the primary tumors of human patients with breast cancer through blind analysis of published whole transcriptome data³⁻⁵. The similarity between human breast cancer and cancers induced by genetic engineering of integration events and oncogene expression in mice is not limited to differential expression of single genes in isolation, nor multiple numbers of single genes; a centrally regulated, major transcriptional network is perturbed in a mirrored fashion in human breast cancer and virally-induced breast cancers in mice.

Keywords: breast cancer, breast cancer etiology, transformation, transformation induced by viral infection, DREAM complex, lin37, E2F4, E2F6, SIN3A, NFYB.

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 cancer⁶⁻¹⁰. In the remaining patients
3 with breast cancer, which in the United States numbered 268,600 in total in 2019¹¹, 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 conducted¹²⁻²⁶. Studies evaluating
9 the claim that viral infection can be the cause of breast cancer in humans have until now relied
10 for the large part on analyses of viral transcription in patient tumors by standard PCR using
11 oligonucleotide primers targeting a region or region(s) of the viral genome, or
12 immunohistochemical methods to evaluate translation of viral proteins using antibodies
13 specific to viral antigens. These studies, which have evaluated the association of viruses
14 including human papillomavirus, the cause of cervical cancer, the polyomaviruses JC, BK and
15 SV40, human herpesviruses (HHV) such as the cause of Burkitt's lymphoma, Epstein-Barr
16 virus, HHV-8 the cause of Kaposi's Sarcoma in patients with acquired immunodeficiency virus
17 (AIDS), and cytomegalovirus, retroviruses including mouse mammary tumor virus (MMTV) for
18 which there is no identified human counterpart, as well as proviral elements such as HERV-K
19 with the development of breast cancer, have been mostly correlative, demonstrating existence
20 of viral transcription or viral antigen expression in tumor tissues and with great variation in
21 frequency, ranging from 0 to 100%¹²⁻²⁶. Extensive laboratory investigation has demonstrated
22 that infection with viruses can result in tumor development²⁷⁻²⁹. However, evidence
23 demonstrating that transformation in mice caused by viral infection, such as that of the mouse
24 mammary gland caused by engineering of the MMTV virus promoter driving expression of the
25 polyomavirus middle T antigen, which induces tumor formation in the breast with metastasis to
26 the lung, is transcriptionally related or similar, and by analogy related in etiology, is relatively
27 limited.

28 We recently provided three lines of evidence to support, but not prove, the hypothesis
that a viral infection or multiple viral infections may be causative agents in breast cancer:
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^{1,2}; in addition, the BK polyomavirus, in the setting of a
foreign tissue, can transcriptionally modulate the expression of genes also transcriptionally
modulated in human breast cancer³⁰. Here we utilize computational tools to provide further
evidence supporting this hypothesis, demonstrating that this transcriptional similarity is not
limited to one or several or even dozens of transcripts differentially expressed in a conserved
fashion; a centrally regulated, organized transcriptional network operates in a mirrored fashion,
both in human breast cancers and in murine breast cancers resulting from genetic modeling of
viral infection through engineering of integration events or oncogene expression, like that of the
SV40 large T-antigen.

24 **Methods**

25 We utilized datasets GSE124648³, GSE109169⁴, GSE23938⁵ and for this differential
26 gene expression analysis of female breast cancer. GSE124648 was generated using Affymetrix
27 Human Genome U133A Array technology with $n=10$ normal breast tissue and $n=126$ tumors of
28 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 $n=5$ 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. For analysis of enriched gene sets, we used the Broad Institute / UC San Diego online tool Molecular Signatures Database (MSigDB)^{31,32}. For computational analysis of enrichment of consensus transcription factor motifs in conserved differentially expressed genes we used ENRICHr^{33,34}.

Results

Breast cancer etiology (cause of the disease) is a large gap in scientific understanding. Deficiencies in understanding of the basic principles by which the disease is initiated can result in serious consequences resulting from poorly guided treatment regimens. We recently discovered evidence of significant conserved differential gene expression when comparing the global gene expression profiles³⁻⁵ of breast cancers arising in mice due to genetic modeling of virus infection, such as expression of the large T antigen of SV40 polyomavirus and genomic integration of the murine mammary tumor virus (MMTV), and in human breast cancers, the majority of which are of unknown etiology^{1,2}. From this analysis we identified (depending on the dataset queried) 45 or 53 genes out of the top 100 most differentially expressed in murine breast cancers arising from genetic modeling of viral infection, when compared to benign mammary tissue that were conserved differentially expressed genes in human breast cancer when compared to normal human breast tissue.

To describe the molecular nature of the shared transcriptional profiles between human breast cancer and cancers arising in the mouse breast as a result of genetic modeling of viral infection, we utilized computational tools to interrogate this list of 45 or 53 genes and further ascertain in an unbiased fashion the existence of functional, coordinated gene expression networks.

Identification of a DREAM complex gene expression signature.

Analysis of both sets of differentially expressed genes (conserved between human breast cancers and murine breast cancers induced by genetic modeling of viral infection) using the Broad Institute computational tool MSigDB^{31,32} revealed that targets of the DREAM complex were significantly enriched for in both sets of differentially expressed genes (Fischer Dream Targets: Table 1 and Table 2; $p=7.39\text{E-}20$ and $p=1.32\text{E-}16$, respectively). This came of significant interest to us as we recently demonstrated that the DREAM complex component lin37 was among the genes most differentially expressed in human breast cancers, and expressed at significantly higher levels in human breast tumors as compared to normal breast tissue³⁵. We further detected elements of gene expression regulated by E2F4

(Marson Bound By E2F4 Unstimulated: Table 1 and Table 2: $p=0.0174$ and $p=0.0189$) and of genes displaying cell cycle stage-specific gene expression (Benporath Cycling Genes: Table 1 and Table 2: $p=0.0154$ and $p=0.017$). Curiously, we detected enrichment of a set of genes down-regulated during week 4 and week 5 of pubertal breast development in one human breast cancer dataset analyzed (McBryan Pubertal Breast 4 5wk DN: Table 2: $p=2.34E-07$).

DREAM complex gene expression targets are markedly enriched for among conserved differentially expressed genes between human breast cancers and murine breast cancers resulting from genetic modeling of viral infections.

Since generation of a DREAM complex gene signature implies that this list should contain some fraction of targets of the DREAM complex, we manually cross-referenced both lists of conserved differentially expressed genes with a list of DREAM complex gene expression targets to determine exactly what portion of these gene sets were targets of the DREAM complex. Depending on the microarray dataset queried, either 34% or 42% of conserved differentially expressed genes in human breast cancers and murine breast cancers induced by genetic modeling of viral infection were established targets of the DREAM complex (Table 3 and Table 4, DREAM complex targets in bold). These genes include: NCAPG, TACC3, MYC, CEP55, ASPM, MTF2, BRCA1, RAD51C, RAD51AP1, SMC1A, PKMYT1, DLGAP5, DNMT1, ACACA, TIMELESS, MXD3, CIT, ASF1B, RAD54L, SLBP, MCM4, MCM7, SKP2, and CDT1 (Table 3 and Table 4). As the list of DREAM complex targets we cross-referenced was only one, we argue this likely underestimates the size of this DREAM complex gene expression network. Nevertheless, we identified a DREAM complex gene expression signature and a striking fraction of direct DREAM complex gene expression targets in this set of genes through computational analysis.

Together with our recent findings demonstrating that the lin37 DREAM complex component is among the genes most differentially expressed in human breast cancer and present at significantly higher levels in breast tumors than in normal breast tissue³⁵, we argue that the DREAM complex is likely active in human breast cancer and that this is a shared feature between murine breast cancers resulting from genetic modeling of viral infection and human breast cancers of largely unknown etiology.

Enrichment of consensus motifs for E2F4, E2F6, NFYB and SIN3A.

We further probed the molecular nature of conserved differentially expressed genes in human breast cancer and murine breast cancers resulting from genetic modeling of viral infection using ENRICH^{34,35}. We identified the presence and enrichment of four consensus transcription factor motifs at the genomic loci of conserved differentially expressed genes. These included the E2F transcription factors E2F4 and E2F6, as well as NFYB and SIN3A (Table 5 and Table 6). Enrichment of E2F4, E2F6, NFYB and SIN3A consensus motifs was statistically significant in both datasets (E2F4: $p=5.238E-14$ and $p=2.346E-11$, E2F6: $p=8.265E-05$ and $p=4.104E-04$; NFYB: $p=1.005E-05$ and $p=2.842E-04$; SIN3A: $p=3.939E-06$ and $p=0.00936$; Table 5 and Table 6, respectively). We identified between 17 and 16 E2F4 gene expression targets, 19 and 18 E2F6 gene expression targets, 21 NFYB gene expression targets, and 12-8 SIN3A gene expression targets.

Genes whose expression could be controlled by all four factors based on the presence of consensus motifs at their genomic loci included RAD51AP1, MYC, BRCA1, CEP55 and MCM4 in one dataset (Table 5) and RAD51AP1, SLBP, MCM7, BRCA1 and HAUS8 (Table 6). Importantly, each of these genes is also a direct target of the DREAM complex.

Thus, through computational analysis of conserved genes differentially expressed in human breast cancer and in murine breast cancers arising from genetic modeling of viral infections when compared to benign mammary tissues, we identified the existence of a DREAM complex gene expression network involving E2F4, E2F6, NFYB and SIN3A and targeting a core set of genes including RAD51AP1, MYC, BRCA1, CEP55, MCM4, MCM7, SLBP, and HAUS8.

Discussion

We report here the existence of a pervasive gene expression network operating in human breast cancers and in murine breast cancers resulting from genetic modeling of viral infection. This gene expression network is modulated by the DREAM complex, and is enriched for gene expression targets of E2F4, E2F6, NFYB and SIN3A. We recently reported significant differential and over-expression of the DREAM complex component lin37, providing further evidence to support the existence of an organized DREAM complex gene expression program in human breast cancer. Importantly, we arrived at these findings through study of murine models of breast cancer induced by genetic engineering to mimic events transpiring during viral infection, such as viral integration in the host genome and expression of a viral oncogene. Thus, not only does this gene expression program exist in humans, it exists in murine breast tumors resulting from hallmark events during infection with viruses capable of inducing transformation in the breast

References

1. Mamoor, S., 2020. Transcriptional similarities between human breast cancers and breast cancers resulting from genetic modeling of viral infection in mice.
2. Mamoor, S., 2020. A shared transcriptional landscape between human breast cancers and breast cancers induced by genetic modeling of viral infection in mice.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. American Cancer Society, 2019. Breast cancer facts & figures 2019–2020. *Am. Cancer Soc*, pp.1-44.
10. Board, P.C.G.E., 2002. Genetics of Breast and Gynecologic Cancers (PDQ®): Health Professional Version. *PDQ Cancer Information Summaries [Internet]*.
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.

- 1 12. 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 13. 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 14. 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 15. 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 16. 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 17. 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 18. 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 19. 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 20. 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 21. 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.

22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
27. Martin, G.S., 1970. Rous sarcoma virus: a function required for the maintenance of the transformed state. *Nature*, 227, pp.1021-1023.
28. 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.
29. 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.
30. Mamoor, S., 2020. Infection of the kidneys with the BK polyomavirus is associated with transcriptional induction of PSMB9 and IGFLR, genes differentially expressed in human breast cancer.
31. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. and Mesirov, J.P., 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43), pp.15545-15550.

32. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P., 2015. The molecular signatures database hallmark gene set collection. *Cell systems*, 1(6), pp. 417-425
33. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R. and Ma'ayan, A., 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14(1), p.128.
34. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A. and McDermott, M.G., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research*, 44(W1), pp.W90-W97.
35. Mamoor, S., 2020. Differential expression of the DREAM MuvB core complex component Lin37 in cancers of the breast.

Gene Set Name	K	Description	k	k/K	p-value	FDR q-value
FISCHER_DREAM_TARGETS	968	Target genes of the DREAM complex.	19	0.0196	7.39E-20	2.3E-15
HALLMARK_E2F_TARGETS	200	Genes encoding cell cycle related targets of E2F transcription factors.	11	0.0550	1.81E-16	7.06E-13
E2F4DP2_01	237	Genes having at least one occurrence of the transcription factor binding site V\$E2F4DP2_01 in the regions spanning up to 4 kb around their transcription starting sites.	9	0.0380	3.6E-12	4.15E-09
E2F4DP1_01	241	Genes having at least one occurrence of the transcription factor binding site V\$E2F4DP1_01 in the regions spanning up to 4 kb around their transcription starting sites.	9	0.0373	4.18E-12	4.65E-09
MARSON_BOUND_BY_E2F4_UNSTIMULATED	689	Genes with promoters bound by E2F4 in unstimulated hybridoma cells.	12	0.0174	5.73E-12	5.75E-09
BENPORATH_CYCLING_GENES	648	Genes showing cell-cycle stage-specific expression.	10	0.0154	1.35E-09	6.02E-07

Table 1: Identification of a DREAM complex gene expression signature.

Listed in this chart is the gene set name, number of genes in each gene set (K), description of the gene set, genes overlapping between the computationally identified gene set and set of genes we identified as differentially expressed expressed between human breast cancers and murine cancers resulting from genetic modeling of viral infection (k), a metric measuring the ratio of genes in each set and number of overlapping genes (k/K), the p-value for each gene set, the q-value of false discovery rate for each gene set (FDR q-value).

Gene Set Name	K	Description	k	k/K	p-value	FDR q-value
HALLMARK_E2F_TARGETS	200	Genes encoding cell cycle related targets of E2F transcription factors.	12	0.0600	2.97E-17	3.08E-13
FISCHER_DREAM_TARGETS	968	Target genes of the DREAM complex.	18	0.0186	1.32E-16	1.03E-12
MARSON_BOUND_BY_E2F4_UNSTIMULATED	689	Genes with promoters bound by E2F4 in unstimulated hybridoma cells.	13	0.0189	3.52E-12	7.83E-09
BENPORATH_CYCLING_GENES	648	Genes showing cell-cycle stage-specific expression.	11	0.0170	6.02E-10	6.69E-07
ISHIDA_E2F_TARGETS	52	Genes up-regulated in MEF cells (embryonic fibroblast) after expression of E2F1 or E2F2.	5	0.0962	7.5E-09	5.18E-06
MCBRYAN_PUBERTAL_BREAST_4_5WK_DN	199	Genes down-regulated during pubertal mammary gland development between week 4 and 5.	6	0.0302	2.34E-07	7.43E-05

Table 2: Identification of a DREAM complex gene expression signature.

Listed in this chart is the gene set name, number of genes in each gene set (K), description of the gene set, genes overlapping between the computationally identified gene set and set of genes we identified as differentially expressed expressed between human breast cancers and murine cancers resulting from genetic modeling of viral infection (k), a metric measuring the ratio of genes in each set and number of overlapping genes (k/K), the p-value for each gene set, the q-value of false discovery rate for each gene set (FDR q-value).

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 paralogue 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%

Table 3: DREAM complex gene expression targets in human breast cancer.

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; DREAM complex gene expression targets are in bold.

Rank	ID	p-value	t	B	Gene	Gene name	% DE
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%
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 3: DREAM complex gene expression targets in human breast cancer.

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; DREAM complex gene expression targets are in bold.

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 4: DREAM complex gene expression targets in human breast cancer.

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; DREAM complex gene expression targets are in bold.

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 4: DREAM complex gene expression targets in human breast cancer.

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; DREAM complex gene expression targets are in bold.

Term	Overlap	P-value	Odds Ratio	Combined Score	Genes
E2F4 ENCODE	17/710	5.238E-14	10.883	332.820	PAQR4; NCAPG; BRCA1; PKMYT1; SMC1A; RAD51AP1; ASPM; RAD51C; MYC; MTF2; TIMELESS; RAD54L; MCM4; SKP2; CEP55; DLGAP5; MXD3
SIN3A ENCODE	12/1131	3.939E-06	4.823	60.017	RAD51AP1; TIMM9; MYC; MTF2; DDIT4; PXMP2; TACC3; NCAPG; MCM4; BRCA1; CEP55; DLGAP5
NFYB ENCODE	21/3715	1.005E-05	2.569	29.568	RETSAT; PAQR4; PHF20; TIMM9; BRCA1; MTM1; APRT; RAD51AP1; ASPM; MYC; FASN; DDIT4; TACC3; RAD54L; MCM4; SKP2; TAF4; HRAS; CEP55; DLGAP5; MXD3
E2F6 ENCODE	18/3245	8.265E-05	2.521	23.703	BATF3; PHF20; NCAPG; BRCA1; SMC1A; ACACA; APRT; RAD51AP1; RAD51C; MYC; FASN; TTC3; TIMELESS; RAD54L; MCM4; DLAT; CEP55; MXD3

Table 5: Enrichment of E2F4, E2F6, NFYB and SIN3A consensus motifs among conserved differentially expressed genes in human breast cancer and murine breast cancers arising from genetic modeling of viral infection.

Listed in this chart is the transcription factor whose consensus motif was enriched for among input genes, overlap between number of genes in the gene set we identified as differentially expressed in a conserved fashion between human breast cancers and murine breast cancers arising from genetic modeling of murine infection containing this consensus motif and total number of occurrences of this motif in the genome, p-value for enrichment of these motifs, odds ratio of these occurrences, combined score for these occurrences, and the names of the genes whose genomic loci contain these consensus motifs.

Term	Overlap	P-value	Odds Ratio	Combined Score	Genes
E2F4 ENCODE	16/710	2.346E-11	8.504	208.134	SLBP; CDT1; RFC3; MCM7; PAQR4; BRCA1; SMC1A; PKMYT1; CIT; RAD51AP1; RAD51C; HAUS8; TIMELESS; RAD54L; SKP2; ASF1B
NFYB ENCODE	21/3715	2.842E-04	2.133	17.418	SLBP; STARD4; MCM7; PAQR4; RETSAT; PHF20; GLDC; TIMM9; BRCA1; CIT; APRT; MTM1; RAD51AP1; FASN; TACC3; RAD54L; MLYCD; SKP2; TAF4; HRAS; ASF1B
E2F6 ENCODE	19/3245	4.104E-04	2.209	17.230	SLBP; CDT1; RFC3; MCM7; PHF20; NAB2; BRCA1; SMC1A; ACACA; APRT; RAD51AP1; RAD51C; HAUS8; FASN; TIMELESS; RAD54L; MLYCD; DLAT; ASF1B
SIN3A ENCODE	8/1131	0.00936	2.669	12.466	RAD51AP1; SLBP; MCM7; HAUS8; TIMM9; PXMP2; TACC3; BRCA1

Table 6: Enrichment of E2F4, E2F6, NFYB and SIN3A consensus motifs among conserved differentially expressed genes in human breast cancer and murine breast cancers arising from genetic modeling of viral infection.

Listed in this chart is the transcription factor whose consensus motif was enriched for among input genes, overlap between number of genes in the gene set we identified as differentially expressed in a conserved fashion between human breast cancers and murine breast cancers arising from genetic modeling of murine infection containing this consensus motif and total number of occurrences of this motif in the genome, p-value for enrichment of these motifs, odds ratio of these occurrences, combined score for these occurrences, and the names of the genes whose genomic loci contain these consensus motifs.