

Article

# In BRCA1 and BRCA2 breast cancers, chromosome breaks occur near herpes tumor virus sequences

Bernard Friedenson <sup>1\*</sup>

<sup>1</sup> Dept. of Biochemistry and Molecular Genetics, College of Medicine  
University of Illinois Chicago; bernief@UIC.edu

\* Correspondence: bernief@UIC.edu;

**Abstract:** Inherited mutations in BRCA1 and BRCA2 genes increase risks for breast, ovarian, and other cancers. Both genes encode proteins for accurately repairing chromosome breaks. If mutations inactivate this function, chromosome fragments may not be restored correctly. Resulting chromosome rearrangements can become critical breast cancer drivers. Because I had data from thousands of cancer structural alterations that matched viral infections, I wondered whether infections contribute to chromosome breaks and rearrangements in hereditary breast cancers. There are currently no interventions to prevent chromosome breaks because they are thought to be unavoidable. However, if chromosome breaks come from infections, they can be treated or prevented. I used bioinformatic analyses to test publicly available breast cancer sequence data around chromosome breaks for DNA similarity to all known viruses. Human DNA flanking breakpoints usually had the strongest matches to Epstein-Barr virus (EBV) tumor variants HKHD40 and HKNPC60. Many breakpoints were near sites that anchor EBV genomes, human EBV tumor-like sequences, EBV-associated epigenetic marks, and fragile sites. On chromosome 2, sequences near EBV genome anchor sites accounted for 90% of breakpoints ( $p < 0.0001$ ). On chromosome 4, 51/52 inter-chromosomal breakpoints were close to EBV-like sequences. Five EBV genome anchor sites were near breast cancer breakpoints at precisely defined, disparate gene or LINE locations. Breakpoint flanking regions resembled known EBV-cancers. Twenty-five breakpoints in breast cancers were within 1.25% of EBV cancer breakpoints. In addition to BRCA1 or BRCA2 mutations, all the breast cancers had mutated genes essential for immune responses. Because of this immune compromise, herpes viruses can activate and produce nucleases that break chromosomes. Alternatively, anchored viral episomes can obstruct break repairs, whatever the cause. The results, therefore, imply proactive treatment and prevention of herpes viral infections may prevent some chromosome breaks and benefit BRCA mutation carriers.

**Keywords:** Breast cancer infection, breast cancer immunity, breast cancer virus, nasopharyngeal cancer, hereditary breast cancer, BRCA1, BRCA2, Burkitt's lymphoma.

## 1. Introduction

Inherited mutations in *BRCA1* and *BRCA2* genes increase risks for breast, ovarian and other cancers. As one of their many functions, the two genes encode proteins that work with many other proteins to restore broken DNA by homologous recombination. Homologous recombination repairs double-strand breaks in chromosomal DNA with high fidelity because the pathway uses the sister chromatid as a template. Mutations that inactivate either the *BRCA1* or *BRCA2* gene force DNA break repairs into less accurate, error-producing pathways that do not use a template [1]. Without template guidance during repairs, chromosomes can undergo visible insertions, deletions, and rearrangements as large DNA fragments do not reattach at their original positions [2, 3]. *BRCA1* also participates in the S-phase and G2/M checkpoints [4-6], so signals from broken fragile sites in *BRCA1*-associated breast cancers become unable to stop cell replication so that damaged fragile sites can be repaired. *BRCA2* may also participate in G2/M checkpoint control [6].

Carriers of *BRCA1* or *BRCA2* gene mutations have breast cancers that show chromosome rearrangements and shifts [7]. Landmark studies of Nik-Zainal and colleagues report hundreds of DNA rearrangements and characteristic rearrangement signatures [8-11]. Chromosome rearrangements may be critical driver events leading to hereditary breast cancers. There are currently no interventions to prevent them because they are thought to be unavoidable and spontaneous in response to replication stress or mutagens.

I wondered whether infections contributed to chromosome changes in hereditary breast cancers. The main reason for asking this question is that if infections are involved in causing fundamental chromosome changes in cancer, it becomes possible to treat the infections and produce vaccines. I had tested thousands of insertions and deletions in lung, colon, cervix, and breast cancers for potential relationships to infections in unpublished experiments. There were often retroviral-like sequences in insertions or deletions, but flanking sequences at breakpoints had better agreement with human herpesvirus 4 (EBV) variants.

The literature helps understand my results. About 8% of human DNA probably originated from retroviruses [12-15]. Both endogenous and exogenous retroviruses are significant sources of human genome instability because they insert their DNA or cause deletions. Unlike retroviruses, EBV is an exogenous DNA virus that does not have an integrase enzyme to incorporate its DNA into humans. However, EBV can still integrate its DNA into human DNA at fragile site breaks by micro-homology mediated repairs [16]. These relatively arbitrary micro-homology repairs should become more frequent if *BRCA1* and *BRCA2* mutations disable the more accurate templated repair system. In other cancers, DNA tumor viruses such as HPV16 in cervical cancer [17] and HBV in liver cancer [18], preferentially integrate their DNA at common fragile sites.

EBV infects over 90% of all adults worldwide. Most hosts remain healthy because their immune systems control the infection, but EBV stays for life as a latent circular episome in host memory B-cells. EBV produces EBNA1, a viral protein that binds specific sites in the human genome. A circular viral episome attaches to EBNA1 and persists as stubborn parasite

chromatin that follows host chromosomes even through cell division. The episome establishes premalignant latent gene expression programs with most viral genes shut off by epigenetic marks. On B-cell activation, EBV instructs its host to remove the inhibitory epigenetic marks on viral chromatin. EBV then induces massive changes in host chromosomal architecture, redirecting host transcription into viral production and inhibiting host antiviral responses [19]. The active virus uses all its genes, producing nucleases (BGLF5 and BALF3)[20]. The viral nucleases cause further host chromosome aberrations, micronuclei, [21] and they inhibit host protein synthesis, blocking the production of host DNA repair proteins. Relatively unopposed EBV replication then drives EBV-mediated malignancies, [22] especially when cells survive because viral production does not go to completion to cause host cell lysis [23].

Its ubiquitous presence and potential for genome damage and perversion associate EBV with a diverse group of human malignancies: including lymphomas, Hodgkin's disease, gastric cancer, nasopharyngeal cancers, and lymphoproliferative disorders. The strongest association is with undifferentiated nasopharyngeal cancer (NPC) [24-26]. Downregulation or mutations in DNA repair genes linked to *BRCA1-BRCA2* mediated repair pathways are common in NPC [27, 28]. NPC in human epithelial cells includes inappropriate DNA repairs causing gene fusions at DNA breakpoints, such as *YAP1-MAML2*, *PTPLB-RSRC1*, and *SP3-PTK2* [29].

The same EBV-sensitive, *BRCA*-related pathway related to NPC is also essential to prevent hereditary breast cancers, suggesting a role for EBV infection in hereditary breast cancers. Evidence exists supporting this relationship. Epidemiological associations between breast cancer and EBV infection exist in different geographical locations [30, 31]. EBV infection of breast epithelial cell models facilitates malignant transformation and tumor formation [32]. Breast cancer cells from biopsies express gene products from latent EBV infection (LMP-1, -2, EBNA-, and EBER) [33] even after excluding the possibility that the virus comes from lymphocytes [34]. Evidence for the EBV lytic form in breast cancer associates with a worse outcome [35].

Carriers of hereditary mutations in *BRCA1* and *BRCA2* genes should be especially likely to sustain damage from viral infection. The mutations cause problems and unreliability in break restoration at fragile sites or virus-induced breakpoints [36]. Immune deficits in *BRCA1* and *BRCA2* mutation carriers [37] may allow the reactivation of retroviruses or latent EBV infections.

Hereditary breast cancers with their increase in chromosome breaks occur in the context of nearly universal EBV infection of humans. Associations between this infection and the origins of chromosome breaks in hereditary breast cancer have not been studied. The aim of the present study was to determine whether chromosome breakages associate with human EBV genome anchoring sites and human virus-like sequences.

## 2. Materials and Methods

### *Breast cancer genomic sequences*

**Characteristics of breast cancers compared to viral cancers.** The selection of breast cancer genomes for this study required a known, typed *BRCA1* or *BRCA2* gene mutation. Breast cancer genome sequences were from the COSMIC database curated from original publications [10, 11]. 15/25 breast cancers were stage III, four were stage II, and three had no data. Six breast cancers had a typed germ-line *BRCA1* mutation, and nineteen had a typed germ-line *BRCA2* mutation. Blood was the source of normal genes for comparison [11]. All 25 cancers were female ductal breast cancers.

The 25 ductal breast cancers contained many DNA missense mutations, with an average of 1.55 mutations per gene analyzed (range 1.0-2.75). A total of 275,730 mutations had a mean value of 11,029 per cancer, ranging from 1-2.75 per gene examined. The 4316 DNA breakpoints varied from 33 to 396 per different individual cancer with a mean of 173. For all 25 breast cancers, chromosomes 1 and 2 were the most frequent sites of intra-chromosome rearrangements, but the distributions of DNA breakpoints among the various chromosomes were markedly different (Table 1).

Table 1 *BRCA1* and *BRCA2* breast cancers studied

Sample	<i>BRCA1</i> / <i>BRCA2</i> mutation status	Analyzed genes	Mutations	Breaks	Mutations per gene	Mutations per break	Chromosome with most breaks "From"	Chromosome with most breaks "To"	Most often intra-chromosomal breaks
PD3890	<i>BRCA1</i>	5978	13401	269	2.24	49.8	3=4,1	11, 16,21	1
PD3904	<i>BRCA1</i>	6337	13559	247	2.14	54.9	6	17,18	8
PD3945	<i>BRCA1</i>	9203	23132	114	2.51	202.9	1	X	2
PD4005	<i>BRCA1</i>	6943	15093	185	2.17	81.6	3	5	1
PD4006	<i>BRCA1</i> R1835X	7433	20453	396	2.75	51.6	1 (only 1)	22,16	2 (387 intra)
PD4115*	<i>BRCA1</i>	9069	22662	180	2.50	125.9	2	17	8
PD4116	<i>BRCA2</i>	7831	19958	359	2.55	55.6	11	17	3
PD4836	<i>BRCA2</i>	4782	5381	80	1.13	67.3	10,11,12	18,20	10
PD4872	<i>BRCA2</i>	5095	5906	90	1.16	65.6	2	18	13
PD4874	<i>BRCA2</i>	8201	10668	226	1.30	47.2	6	20	10
PD4875	<i>BRCA2</i>	6262	8056	140	1.29	57.5	5	19	7
PD4876	<i>BRCA2</i>	5718	8317	95	1.45	87.5	5	9	12
PD4951	<i>BRCA2</i>	3178	3283	33	1.03	99.5	6	8	4
PD4952	<i>BRCA2</i>	10806	15155	332	1.40	45.6	3	14	14
PD4953	<i>BRCA2</i>	6690	8090	143	1.21	56.6	4	18	18
PD4954	<i>BRCA2</i>	5904	6605	126	1.12	52.4	1	8	2
PD4955	<i>BRCA2</i>	6654	7920	100	1.19	79.2	12	15	1
PD4956	<i>BRCA2</i>	11036	14860	238	1.35	62.4	7	10	3
PD4957	<i>BRCA2</i>	3521	3520	74	1.00	47.6	7=10	17=11	8

PD4958	<i>BRCA2</i>	8244	10956	102	1.33	107.4	1	1	4
PD6406	<i>BRCA2</i>	4976	5485	191	1.10	28.7	2,4	6,8	2
PD6416	<i>BRCA2</i>	3364	3609	71	1.07	50.8	1	20	7
PD7217	<i>BRCA2</i>	8841	11215	124	1.27	90.4	3	18	3
PD8621	<i>BRCA2</i>	7964	9617	275	1.21	35.0	1	X	1
PD8969	<i>BRCA2</i>	7335	8829	126	1.20	70.1	1	8	9

\*Uncertain significance variant

The genes analyzed in *BRCA1*-associated breast cancers (Table 1) were on average almost twice as likely to be mutated as the genes analyzed in *BRCA2* associated breast cancers ( $p < 0.0001$ ). Moreover, an unpaired t-test assuming equal variances found that the means were different ( $BRCA1=94.5$ ,  $BRCA2=63.5$ ,  $p=0.03$ ). Some tests added breast cancer genome data from 29 presumptive mutation carriers when the number of identified *BRCA1,2* gene mutation carriers became limiting. Their breast cancers occurred under age 36, but the patients had not been tested for BRCA mutations.

**Breast cancer DNA sequence data.** Gene breakpoints for inter-chromosomal and intra-chromosomal translocations were obtained from the COSMIC catalog of somatic mutations as curated from original publications [11] and converted to the GrCH38 human genome version. DNA flanking sequences at breakpoints were downloaded primarily using the UCSC genome browser, but did not differ from sequences obtained using the Ensembl genome browser.

Fragile site sequences used positions of genes containing fragile sites from the fragile site database [38] and from original publications [36]. The location of repetitive di- and trinucleotides was used to exactly locate likely fragile sites.

**Comparisons of DNA sequences.** The NCBI BLAST program (MegaBLAST) and database [39, 40] compared DNA sequences around breakpoints in *BRCA1*- and *BRCA2*- mutation-positive breast cancers to all available viral DNA sequences. Virus DNA was from BLAST searches using "viruses (taxid:10239)" with homo sapiens and uncharacterized sample mixtures excluded. EBV DNA binding locations on human chromosomes were obtained from publications [41-43], from databases, by interpolation from published figures, or by determining the location of genes within EBNA1 binding sites. EBNA1 binding data was based on lymphoblastoid and nasopharyngeal cancer cell lines. When necessary, genome coordinates were all converted to the GrCH38 version. Breaks in hereditary breast cancers were compared to EBV DNA binding sites, epigenetic marks on chromatin, genes, and copy number variations. The MIT Integrated Genome Viewer (IGV) with ENCODE data loaded and from the UCSC genome browser provided locations of H3K9Me3 chromatin epigenetic modifications. The ENCODE website also provided positions of H3K9Me3 marks (www.ENCODEproject.org).

Homology among viruses was determined by the method of Needleman and Wunsch [44].

**Data analyses.** Microsoft Excel, OriginPro, Visual basic, and Python scripts provided data analysis. Excel worksheets were often imported into Python Jupyter notebooks for extended graphical analysis. Chromosome

annotation software was from the NCBI Genome Decoration page and the Ritchie lab using the standard algorithm for spacing [45]. Statistical analyses used StatsDirect statistical software. Linear correlation, Kendall, and Spearman tests compared distributions of the same numbers of chromosome locations that matched viral DNA. Because the comparisons require the same numbers of sites, comparisons truncated data down to a minimum value of maximum homology (human DNA vs. viral DNA) of at least 400. Excel compared the positions of breast cancers vs. midpoints of genes containing repetitive DNA fragile sites on the same chromosome.

**Genes associated with the immune response damaged in breast cancer.** Breast cancer somatic mutations in genes were compared to genes in the immune metagenome [46-49]. Sets of genes involved in immune responses also mediate other functions and represent a vast and growing dataset (geneontology.org). Genes involved in cancer control by immune surveillance and immunoediting are not well characterized. An extensive validation was conducted to include both direct and indirect effects of gene mutations. In addition, the Online Mendelian Inheritance in Man database (www.OMIM.org) was routinely consulted to determine gene function with frequent further support obtained through PubMed, Google scholar, GeneCards, and UniProtKB. The "interferome" was also sometimes used [www.interferome.org].

### 3. Results

Virus-human homology comparisons around inter-chromosomal breakpoints produced many results like those in Fig.1. Many DNA segments are virtually identical to EBV variants (Human gamma-herpesvirus 4 variants, HKNPC60 or HKHD40) at or near the different inter-chromosomal breakpoints shown. A deleted fragment in panel (C) includes virus-like sequences and has viral matching sequences in its flanking regions. Maximum homology scores between human DNA vs. herpes viral DNA are over 4000 for breast cancer PD3945 and just under 4000 for PD4874. These scores correspond to about 97% identity for up to 2462 base pairs, with E "expect" values (essentially p-values) equal to 0.

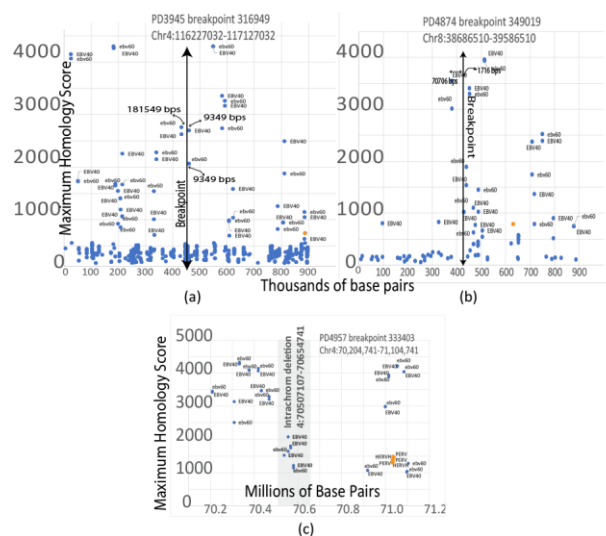




Figure 1. Examples of viral homologies at breast cancer breakpoints. The blue dots represent start points of human DNA sequences that match EBV tumor variants and orange dots are retroviruses. The numbers near breakpoints are the distances from the homology start points to the breakpoint. (a) A breakpoint on a BRCA1 associated breast cancer PD3945 is surrounded by DNA homologous to HKHD40 (EBV40) and HKNPC60 (ebv60). (b) A breakpoint in the BRCA2-associated breast cancer PD4874 is also near regions of homology to HKHD40 and HKNPC60. In the bottom panel (c), an intra-chromosome deletion on chromosome 4 (gray area) in BRCA2 associated breast cancer (PD4957) is near EBV40 and ebv60 sequences and also includes them.

Inter-chromosomal breakpoint distribution in BRCA2-Associated breast cancers.

There are significant differences in where inter-chromosomal breaks occur on 19 BRCA2-associated breast cancers (Fig. 2). Breakpoints distribute almost uniformly throughout chromosome 8, which has the most numerous breakpoints, but the breaks generally do not occur at random. Inter-chromosomal breakpoints tend to cluster in specific chromosome regions for individual breast cancers. Breast cancer PD4956 has a large breakpoint cluster on chromosome 11, and PD4952 has a group on chromosome 14. A few chromosome areas rearrange in only one cancer. Large numbers of base pairs often separate the breakpoint positions

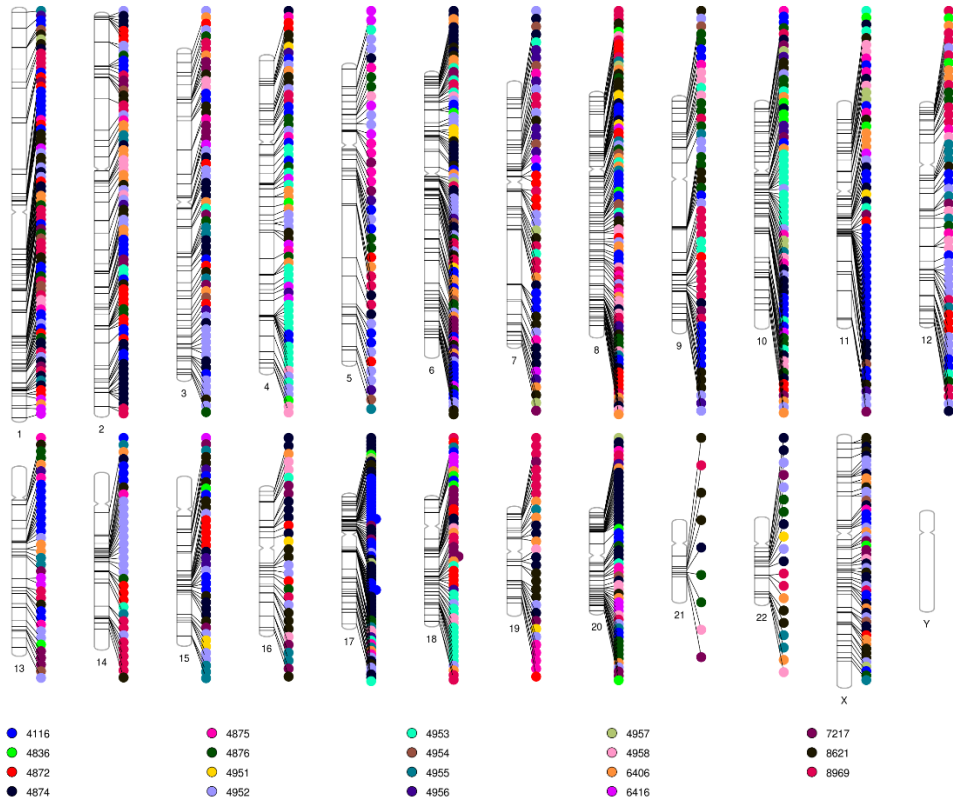


Figure 2. Distribution of chromosome breakpoints in BRCA2 mutation associated breast cancers over all human female chromosomes. Each BRCA2 associated breast cancer (Table 1) is indicated by a four-digit number and labeled with a different color. Chromosome numbers are given below the chromosome drawing. The lines on the chromosome depictions indicate positions where the break occurs. Since all the breast cancers occurred in females, there are no breaks on the Y chromosome.

### Breast cancer breakpoint flanking sequences often resemble EBV tumor variants.

EBV genomes can be anchored at many sites on chromosome 4, which was arbitrarily selected as a representative chromosome to estimate how often breakpoint regions resemble herpes viruses. Nearly all (51/52) breakpoints had statistically significant homology to herpes viruses within about 200k base pairs from the breakpoint (Fig. 3).

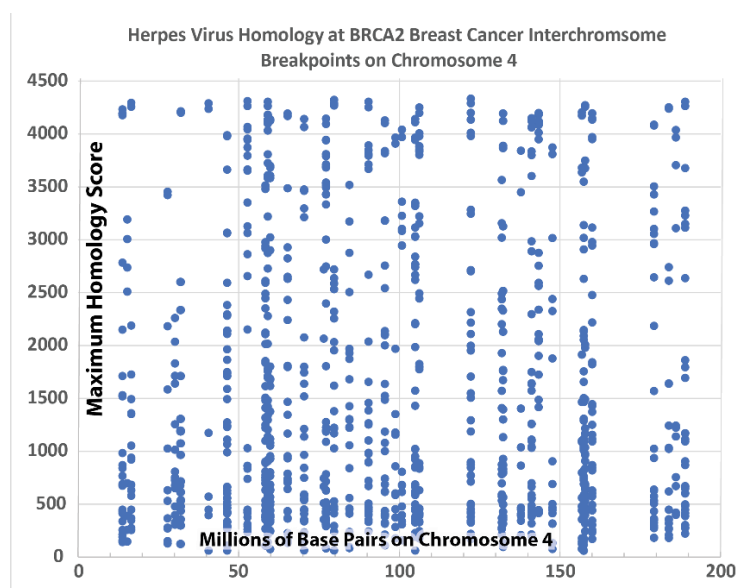


Figure 3. Statistically significant maximum homology scores (blue dots) for herpes virus DNA vs. human DNA at inter-chromosome breakpoints. The graph represents the entire length of chromosome 4 and only shows homology scores above 100.

### Breast cancer breakpoint homologies to EBV are near to known EBV genome anchor sites: global comparisons on two chromosomes.

Most breakpoints on chromosomes 2 and 12 are near EBV genome anchor sites. On a 21 Mb section of chromosome 2, (Fig. 4a), the  $R^2$  statistic was calculated to estimate correlation between EBV genome anchoring sites and breast cancer breakpoint positions. According to  $R^2$ , EBV anchor sites account for about 91% of chromosome 2 breakpoints ( $p < 0.0001$ ). The normal plot of the residuals from this analysis is approximately linear, suggesting that the residuals are distributed relatively randomly about zero. Breakpoints on a 14 Mb section of chromosome 12 also correlate with viral



genome anchor sites ( $r^2 = 0.88$ ,  $p < 0.0001$ ) and share many features with the chromosome 2 breakpoint region (Figs. 4b vs. 4a).

All breakpoints on both chromosome sections are near regions of human DNA sequence homology to the EBV variant tumor viruses HKNPC60 and HKHD40. Breakpoints are all accessible as indicated by DNase hypersensitivity and many breakpoints disrupt gene regulation, gene interaction, and transcription. Many breaks affect cancer-associated (COSMIC) genes. Breakpoints all go through ENCODE candidate cis-regulatory elements (cCREs). Some breakpoints disrupt the epigenetic stimulator H3K27Ac, an enhancer mark on histone packaging proteins associated with increased transcription. Most breast cancer breakpoints are near inhibitory epigenetic H3K9Me3 peaks in CD14+ primary monocytes (RO-01946). These markings are known to occur around EBV genome anchor sites where they contribute to viral latency and repress transcription [42]. Both regions on chromosomes 2 and 12 are rich in these sites (Figs. 4a and 4b). Both chromosome 2 and 12 sections appear to be a focus for structural variation such as CNV's, inversions, and short insertion/deletions.

Multiple breakpoints on either chromosome 2 or 12 disrupt reference genes (bottom of figures). Human reference sequence genes in the chromosome 2 region include *KYNU*, *GTDC1*, *ACVR2A*, *KIF5C*, *STAM2*, *KCNJ3*, *ERMN*, *PKP4*, *BAZ2B*, *TANK*, and *DPP4* (Fig. 4a, bottom). Breast cancer breaks near at least some of these genes interrupt functions essential for immunity and preventing cancer. For example, *KYNU* mediates the response to IFN-gamma. *TANK* is necessary for *NFkB* activation in the innate immune system. *DPP4* is essential for preventing viral entry into cells. Reference gene functions in the breakpoint region of chromosome 12 include vesicle trafficking (*RASSF9*), endocytosis (*EEA1*), blood cell formation (*KITLG*), interferon response control (*SOC2*) and nerve cell patterning (*NR2C1*).

Potential retrovirus-like contributions are different in Figs. 4a vs. 4b. On chromosome 2, porcine endogenous retrovirus (PERV) [50], human endogenous retrovirus (HERV) sequences [51] and a pseudogene (pHERV), also have significant homology to human DNA. The PERV-like sequence lies within a retroposed area on chromosome 2 within 28 Kbps 5' and 80 Kbps 3' of EBV sequences. EBV-like ends potentially generate homologies for retro-positioning and inserting PERV sequences. In contrast, significant retroviral sequence homologies are outside the breakpoint-rich stretch on chromosome 12 (Fig. 4b).

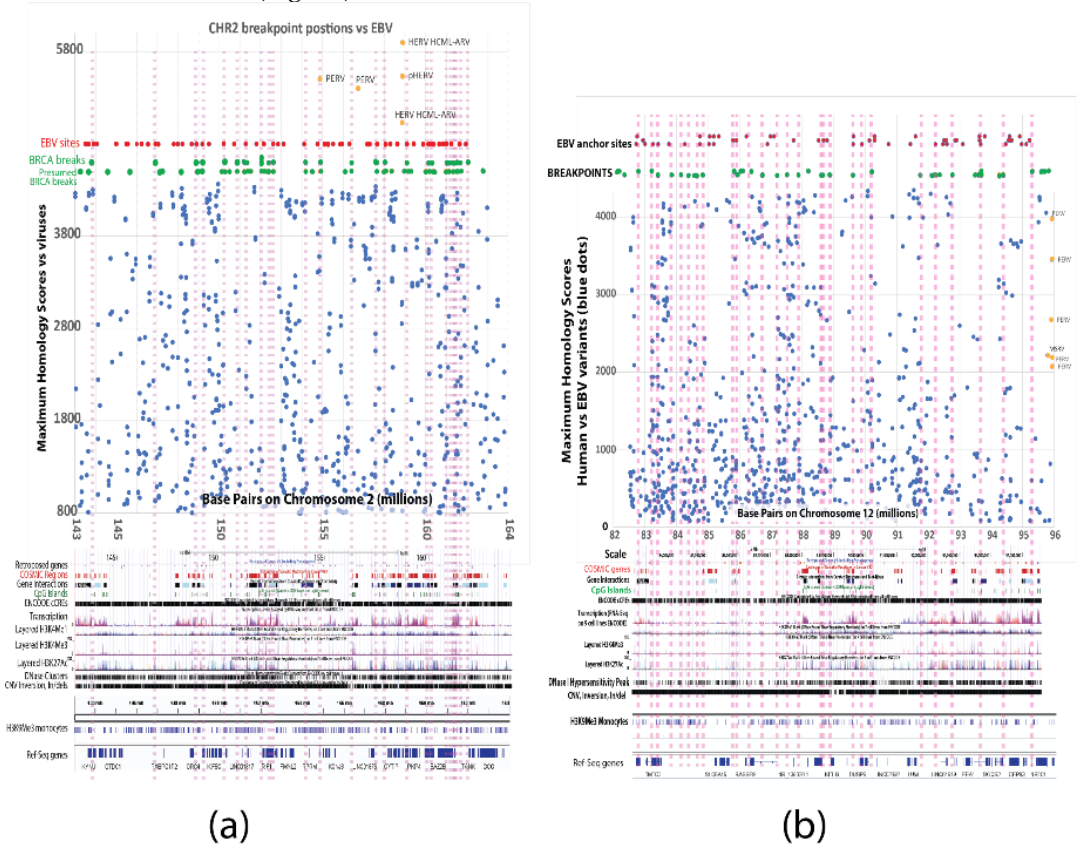


Figure 4. (a), Breakpoints on a 21 million base pair segment in human chromosome 2 and (b) on a 14 million base pair segment of chromosome 12 in BRCA1, BRCA2, and presumptive BRCA-associated breast cancers (green dots). Blue dots represent significant homology scores between human chromosome sequences vs. EBV variants HKNPC60 and HKHD40. Dashed lines align EBV genome anchor sites with breast cancer breakpoints. EBV anchor locations (red dots) and breast cancer breakpoints (green dots) are shown to allow comparisons. PERV and HERV variants (orange) may also contribute to breakpoints, but they distribute very differently on the two chromosome sections.

**Identified EBV genome anchors near known genes match breast cancer breakpoints.**

Fig. 5 further tests the relationship of breast cancer breaks to EBV genome anchor sites, precisely identified at disparate chromosome or gene locations [41]. In early-onset breast cancer PD23566, episomal anchors and EBV variant homologies fall within a deleted segment of chromosome 1. Breakpoints on chromosome 6 from three different breast cancers surround

EBV genome anchor sites and viral homologies. Even a chromosome 6 LINE retrotransposon has these homologies (Fig. 5a). A primary EBV genome binding site on chromosome 11 [41] matches a breakpoint in an early-onset breast cancer. A region of chromosome 5 containing an anchor site near HDAC3 has only distant EBV homologies. Instead, there are much shorter and weaker nearby similarities to retroviruses.

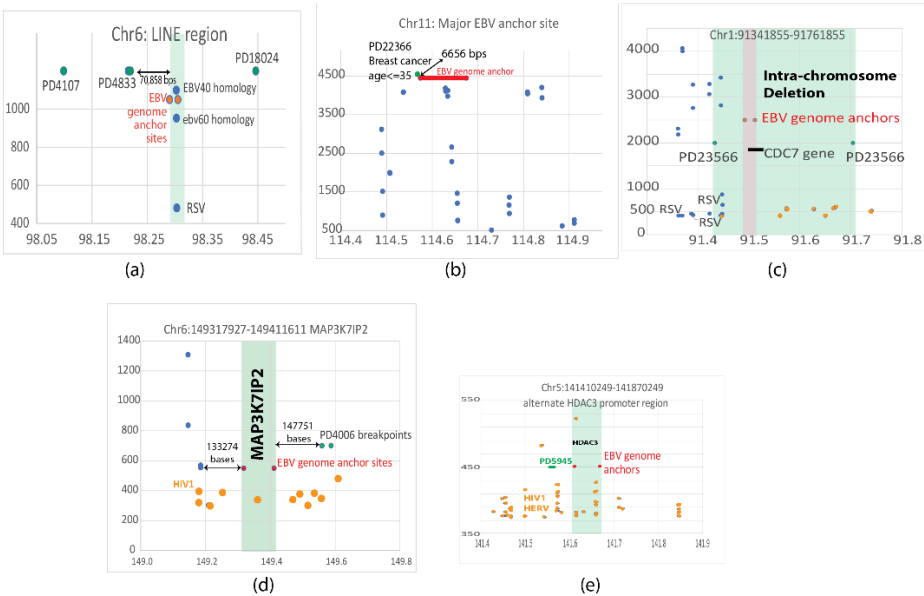


Figure 5. Maximum homology to human DNA for all viruses (y axis) are plotted for precisely known EBV genome anchor sites in relation to LINE or gene coordinates. (a) A direct EBV binding site is near the region Chr6: 98,850,000-98,860,000, which contains a transposable LINE element (within the red dots) [26]. The area is near breakpoints in presumptive BRCA-associated breast cancers (green dots). Blue dots indicate significant HKHD40 and HKNPC60 homologies. Weaker homology to respiratory syncytial virus (RSV) also exists. In (b), a primary EBV binding site is within 6656 base pairs of a breakpoint in breast cancer PD22366. (c) EBV variant homologies near CDC7 fall within a deleted segment in an intra-chromosomal rearrangement on chromosome 1. In (d), EBV anchors are within 148K bps of a breast cancer breakpoint. The MAP3K7IP2 gene has EBV anchor sites at its boundaries and nearby viral homologies at 133,274 bases. In (e), breakpoints in breast cancer PD5945 are close to EBV-genome binding near the HDAC3 gene, but EBV variant homologies are more distant. Weaker homologies exist with HIV1 variants and endogenous retroviruses (orange dots).

Five of six localized EBV binding sites were near HKHD40 / HKNPC60 human viral homologies. The known EBV binding site at the *HDAC3* region on chr5:141610249-141660249 [41] was close to breakpoints in breast cancer PD5945 (at 141,557,433 and 141,564,233), but did not contain nearby significant EBV variant homologies. Instead, there were lower maximum homology matches to retroviruses. The *HDAC3* results suggest there are alternate mechanisms involving retroviral participation in breast cancer breakpoints (Fig. 5).

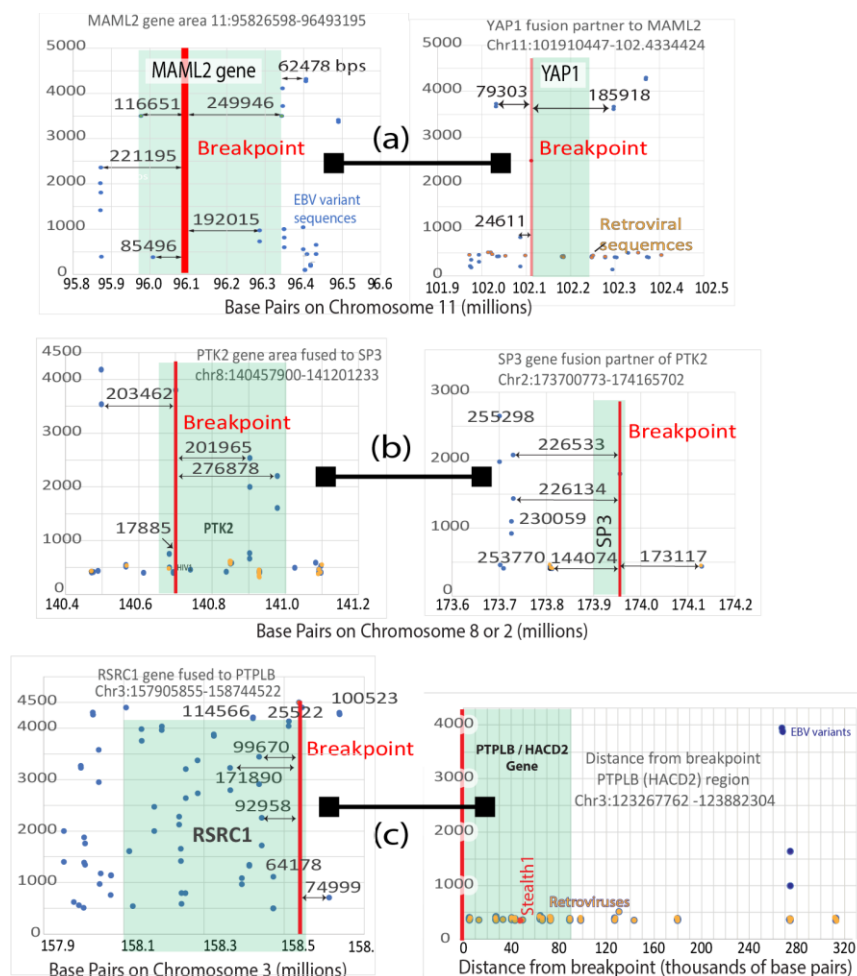


Figure 6. Areas around most gene fusions in nasopharyngeal cancers resemble breast cancer chromosome break regions. Homology to EBV tumor variants (blue dots) is similar to that found near breast cancer chromosome breaks. Gene fusions are shown as left to right pairs (a) MAML2-YAP1, (b) PTK2-SP3, and (c) RSRC1-PTPLB). Maximum homology scores are plotted against chromosome locations for the first five panels. The PTPLB panel (lower right) graphs maximum homology vs. absolute values for distance from viral homologies to the breakpoint. Orange dots indicate the start point of retroviral homologies.

### Viral homologies around breakpoints in *BRCA* - associated breast cancers resemble model cancers caused by EBV.

DNA sequences at recurrent breakpoints in known EBV mediated cancers (nasopharyngeal cancer and Burkitt's lymphoma) have homologies to EBV variants, just like the breast cancers (Fig. 6). For example, the *PTK2* and *MAML* genes targeted by rearrangements in nasopharyngeal cancer (NPC) include strong similarities to EBV. The *RSRC1* breakpoint in a gene fusion of *RSRC1* to *PTPLB* in NPC is close to regions of strong EBV homology. However, the flanking sequences around the *PTPLB* breakpoint have homology to retroviral sequences.

The known EBV cancers and hereditary breast cancers break at some of the same chromosome positions (Table 2). Comparable breakpoint positions in BRCA-breast vs. known EBV cancers (NPC and. BL) differ by less than 1.25%). Normality plots of the breakpoint positions in Table 2 for all three types of cancers are identical, go through zero, and follow the line  $y=x$ . One set of data can calculate the other using the equation Breast Cancer Breaks=1.00021(NPC\_BL Breaks) +9012 ( $p<0.0001$ ). An unpaired t-test did not find significant differences between the two sets of data,  $p<0.001$ .

Table 2 Comparisons of Cancer Chromosome Breakpoints

Chromosome	NPC /BL Break coordinate	Breast cancer break coordinate	% Difference
1	151,928,027	151,922,525	0.00
1	155,119,764	155,020,623	0.06
1	154,117,249	154,111,293	0.00
1	197,593,291	197,867,477	0.14
1	203,179,378	203,428,861	0.12
1	204,435,983	204,386,502	0.02
2	66,324,118	66,473,128	0.22
3	158,537,092	158,531,593	0.00
4	101,138,921	101,009,819	0.13
4	152804117	152,852,058	0.03
5	43086276	42897159	0.44
6	74142321	74016919	0.17
6	104668449	104,676,699	0.01
8	98073546	98,164,150	0.09
8	128,468,417	128,300,594	0.13
8	128,481,840	128,300,594	0.14
11	102,114,930	102,384,027	0.26
11	102,115,180	102,384,027	0.26
11	101,983,850	101,895,690	0.09
13	74133903	74,125,809	0.01
15	41381710	41,532,838	0.37
18	43243754	43103981	0.32
19	14770552	14,943,490	1.17
19	33307575	33,086,916	0.66
X	3,996,343	4,044,529	1.21

Supplementary File Fig. S1 shows distances to EBV homologous regions starting from breakpoints in nasopharyngeal cancer NPC-5989 calculated from data in reference 12. Just as in the hereditary breast cancers, all breakpoints in nasopharyngeal cancer NPC-5989 are close to human genome regions that resemble the EBV variants.

### Damage to genes needed for the immune system in *BRCA1* and *BRCA2* associated breast cancers.

All 25 hereditary breast cancers had significant damage to genes needed for immune system functions (Table 3). A total of at least 1307 immune-related genes had mutations. Of the top 20 mutated genes, 8 (40%) belong to the innate immune system, and the remaining 12 participate in adaptive immunity. A variety of cells express the genes, including macrophages, B-cells, mast cells, helper- and memory-T-cells.

Significant differences exist in distributions and numbers of mutations among the breast cancer genomes [8, 9, 11, 52] but mutations commonly cripple some aspect of innate immunity, its regulation, or its connections to adaptive immunity (Table 3). The damage interferes with responses to antigens, pathogens, and the ability to remove abnormal cells.

Table 3 The top 20 most commonly mutated genes in hereditary breast cancers are associated with the immune and nervous systems

Gene	Mutations in 25 <i>BRCA1</i> or <i>BRCA2</i> breast cancers	Type of Immunity listed in database [53]	Example of function in immunity	Known or likely connection to the nervous system
<i>AKT3</i>	20	Innate	AKT3 amplifies innate immune responses to DNA or RNA virus infections [54]. AKT3 binds interferon response factor 3, enhancing its activation and stimulating interferon responses.	✓
<i>FRMD4A</i>	20	Innate	Depends on Interferon Regulatory Factor 5 [55]. Affects antigen presentation and subsequent effects on dendritic cells. Mediated by cytoskeletal actin and endocytosis	✓
<i>ARHGAP15</i>	19	Innate	Negative regulator of neutrophil function. Affects mast cell function	✓
<i>CAMTA1</i>	19	Adaptive	Pattern of methylation distinguishes subsets of T-cells [56]	✓
<i>DPYD</i>	19	Innate	Interferon response [57]. Affects infection severity in mice. Natural killer cell function	✓
<i>CHRM3</i>	18	Adaptive	Cholinergic receptor muscarinic 3 Deficiency associated with immune	✓



			impairment [58] Targeting by miRNAs related to inflammation [59].	
<i>ADAMTS12</i>	17	Adaptive	Member of a group of related proteins with many connections to the immune system and immune disorders [60]	✓
<i>DAB1</i>	17	Adaptive	Associated with microbial profile [61]	✓
<i>ADAMTS3</i>	16	Innate	Member of a group of related proteins with many connections to the immune system and immune disorders [60] Mast cell function	✓
<i>COL23A1</i>	16	Adaptive	Enriched in plasmacytoid dendritic cells	✓
<i>DLC1</i>	16	Adaptive	Immune regulation of stem cells through interactions with <i>NOTCH-1</i> [62] Type 2 T helper cell	✓
<i>ADAM12</i>	15	Adaptive	Controls response to infectious disease [63]. Central memory CD8 t-cells	✓
<i>CD36</i>	15	Adaptive	<i>TLR4-TLR6-Cd36</i> activation is a common molecular mechanism by which atherogenic lipids and amyloid-beta stimulate sterile inflammation. Gamma delta T cell function	✓
<i>CDH2</i>	15	Innate	Maintains stem cell quiescence [64]. Natural killer cell function	✓
<i>DACH1</i>	15	Innate	Downregulated in lymphoid cell progenitor [65]	✓
<i>CREB5</i>	14	Innate	CREB factors limit proinflammatory responses, provide anti-apoptotic signal, regulate Th1, Th2, Th3 responses, generate and maintain regulatory T cells [66]	✓
<i>ANK1</i>	13	Adaptive	Type 17 T helper cell	✓
<i>HDAC9</i>	13	Adaptive	Known to alter B-cell responses in autoimmune diseases Immature B cell function.	✓
<i>COL4A1</i>	12	Adaptive	Central memory CD4 T cell	✓
<i>F13A1</i>	12	Adaptive	Phagocytes at CNS borders [67]	✓

### Comparisons of viral variants to other viruses

Both the EBV tumor viruses (HKHD40 and HKNPC60) were typical of many other herpesvirus isolates, with some haplotypes conferring a high NPC risk [26]. About 100 other gamma herpes viral variants strongly matched both HKHD40 and HKNPC60 in regions with enough data to make comparisons possible. HKNPC60 was 99% identical to the EBV reference sequence at bases 1-7500 and 95% identical at bases 1,200,000-1,405,000. HKHD40 gave values of 99 and 98% identity for comparisons to the same regions.

Comparisons of breast cancer mutations to known or likely viral cancers

Sets of genes mutated in breast cancers are similar to known or likely viral cancers. A set of breast cancers share 1143 identical genes mutated with viral cervical cancer and about 50% of genes mutated in Burkitt's lymphoma and viral liver cancers.

Fragile site breaks are near BRCA1/2 breast cancer breakpoints

Lu et al. found 4785 EBNA1 binding sites with over 50% overlapping a repetitive sequence element [41] such as runs of consecutive AT bases. Kim et al reported that EBNA1 anchor sites have A-T rich flanking sequences [42]. Based on the fragile site database, chromosome 1 contains 658 fragile site genes, the most of any chromosome [38]. Over 200 DNA breast cancer break positions differ from the midpoints of fragile site sequences on chromosome 1 by about the number of base pairs in EBV or less. Distances from breakpoints to fragile site midpoints ranged from 12 to 172000 base pairs (Fig. 7). Flanking sequences were rich in A-T repeats. About 60% of the breakpoints were within 66,000 base pairs of a fragile site.

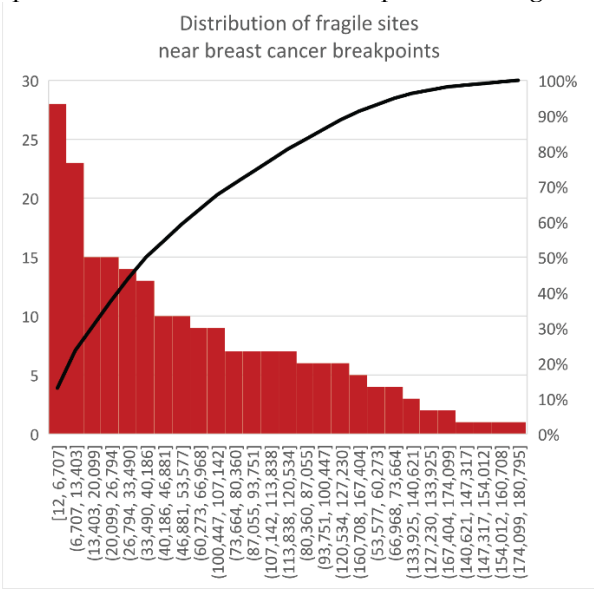


Figure 7. Distances from breast cancer breakpoints to the midpoints of fragile sites on chromosome 1. The vertical left axis shows the number of breakpoints at the distance ranges (in base pairs) given on the x-axis. The black curve gives the cumulative total that can be read on the right vertical axis.

Discussion

This work produces a working model for the participation of EBV variants in causing chromosome breaks in *BRCA1* and *BRCA2* associated

breast cancers. Multiple lines of evidence support the model given below. For example, positions of viral anchors (EBNA1) are near chromosome breakpoints; human herpes virus-like sequences are frequently found in sequences flanking breast cancer breakpoints; patterns of viral homology are challenging to distinguish from cancers known to be caused by EBV; breast cancer breakpoints on chromosome 1 are within the number of bases in EBV from genes containing fragile sites.

Fig. 8 summarizes the model. Pathogenic mutations in either BRCA gene increase DNA breaks caused by mutagens, DNA damaging agents, and inflammation. Inactive BRCA genes also compromise homologous recombination pathways needed to restore fragmented chromosomes to their original condition. In this context, almost every human harbors a latent EBV infection which exists as a circular viral DNA parasite genome attached to human chromosomes. The immune response normally controls EBV, allowing its latent form to express only a few genes that do not trigger immunity. One of these viral genes expresses EBNA1, which anchors EBV genomes to host chromosomes at hundreds of sites, where the virus may still alter control of host cellular genes. Viral genomes remain attached as parasites even during human cell division. This attachment maintains stable numbers of episome copies in multiplying host cells [68, 69]. If the DNA near these attachment sites breaks, attached EBNA1 and viral DNA obstruct access to machinery required for repair.

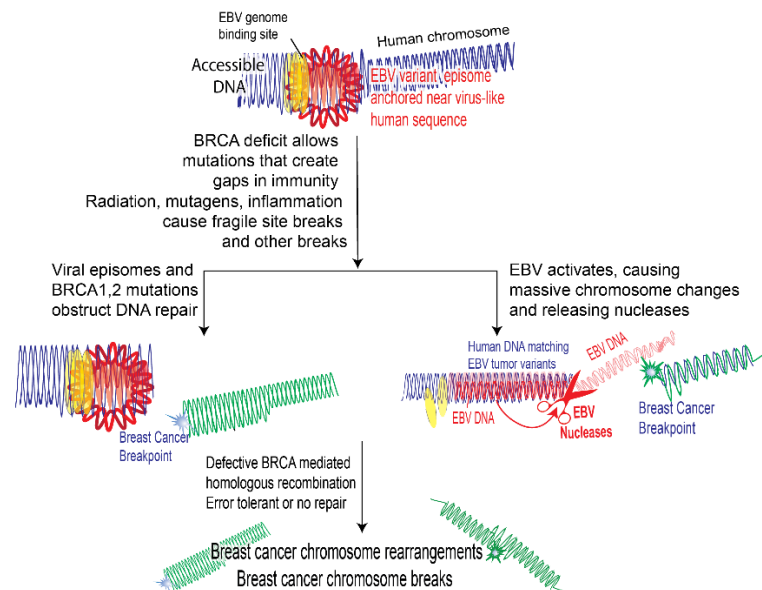


Figure 8. Model for the role for EBV variants in chromosome breaks in breast cancers associated with BRCA1 and BRCA2 gene mutations

This interference becomes so significant in mouse models that B-cell tumors develop [22]. Flanking the viral anchor sites on human chromosomes are regions that match the viral sequence, are accessible to DNases, and contain fragile site, repetitive A-T rich DNA sequences. Many breast cancer breaks occur near these regions. Mutations in *BRCA1* and *BRCA2* associated breast cancers compromise host immune responses, favoring the conversion of

inactive circular EBV to active linear forms. On activation, the virus instructs host cells to remove inhibitory epigenetic marks from the viral DNA [19] and the dissociation of EBNA1 oligomers occurs [70]. The activated viral forms cause massive host chromosome changes as they force host cells into viral production. The infection spreads and more host cells become infected with latent forms. Viral production that does not go to completion allows viral enzymes to persist as cancer drivers in abnormal cells [23, 71]. Latent circular viral episomes obstruct the host cell's ability to access a broken human chromosome for repair. Activated EBV in tumors produces nucleases BGLF5 and BALF3, which cause additional DNA breaks and inhibit host protein synthesis needed for immunity [21]. Whatever repairs are made produce abnormal chromosome structures.

There are countless agents and mechanisms known to break DNA and chromosomes that do not involve viruses. Accordingly, chromosome breaks could easily follow some random distribution. In contrast, inter-chromosome BRCA1 and BRCA2 breakpoints are not distributed randomly among chromosomes, and there is some selectivity among breast cancers in where these breakpoints occur and in partners for repair (Fig. 2).

EBV contributes to this selectivity because nearly all the inter-chromosome breakpoints on the entire length of chromosome 4 occur within the approximate number of base pairs in lytic EBV variant sequences (roughly 200,000 in Fig. 3). Many breast cancer breakpoints are separated from viral homologous sequences by numbers of base pairs consistent with circular viral episome lengths (roughly 65,000 base pairs, Fig. 7). Circular viral episomes could also be a dominant carcinogen by blocking access to proteins and nucleic acids needed for repair.

Comparisons of chromosome break sites in hereditary cancers to actual sites of anchoring by EBNA1 further supports their relationship. Comparisons on chromosome 2 (21 million base pairs) and chromosome 12 (24 million base pairs) give similar results (Fig. 4). DNA comparisons for a few EBV binding sites at precisely known LINE transposable element or gene sequences (Fig. 5) also show that viral genome binding, breast cancer breaks, and viral human homologies are often close together.

EBV episomes bind to human chromosomes using EBNA1-dimers as anchors for subsequent viral attachment [70, 72]. The dimerized anchor itself can alter human chromosome structure and mediate distant interactions related to transcription during infection. EBNA1 anchor sites favor some repetitive elements, especially LINE 1 retrotransposons (Fig. 6), and correlate weakly with histone modifications [41].

Retroviruses and retrotransposons [73] may also participate in breast cancer breaks. Participation from porcine endogenous retroviruses is actionable by thoroughly cooking pork products. However, despite assertions that xenotransplantation with pig cells is safe, as many as 6500 bps in human chromosome 11 are virtually identical to pig DNA is concerning (Fig. 4a).

Nasopharyngeal cancer (NPC) and Burkitt's lymphoma (BL) have known links to EBV. At least 25 breakpoints in these known EBV mediated cancers are probably within experimental error [74] of breakpoint positions in breast cancers (Table 2). Running these comparisons for such accepted

EBV-related cancers gave results that were difficult to distinguish from breast cancers (Figs. 6-7, S1, and Table 2). Five of six genes involved in gene fusions are near EBV variant sequences and include EBV-like variant sequences within the gene boundaries (Fig. 6).

Breast cancer mutations cripple the immune system's ability to control cancer-causing infections, remove cells damaged by disease, and regulate inflammation. Many genes mutated in human breast cancers link to some variable function within the immune system or structural barrier defenses [37, 75, 76]. Antigen and viral recognition, signaling essential to transmit the immune responses, and immune regulation can all be impaired. Many mutations independently link to various infections, including all known cancer-causing microorganisms. The top 20 mutated genes in BRCA1 and BRCA2 breast cancers are all associated with some function of immunity (Table 3). The relationships of the top 20 mutated genes to the nervous system may also favor herpes viral infections.

Multiple components of immune defenses mutate in the 25 hereditary breast cancer genomes. The mutations affect processes such as cytokine production, autophagy, etc. These functions depend on many genes dispersed throughout the genome, so any cancer needs only to damage one gene to cripple an immune function. Each breast cancer genome has a different set of these mutations, with the same gene only occasionally damaged. Damage affecting the nervous system was also universal. Some herpes viruses establish permanent infection within the central nervous system even after other sites become virus-free [77].

Viral variants can be significant breast cancer drivers without producing large numbers of active viral particles in the tumor. The association of EBV variant sequences with chromosome aberrations does not require the continuing presence of active virus anywhere within the resulting tumor. Because virtually everyone carries EBV infection, viral participation in breast cancer is very difficult to distinguish from its role in seemingly normal cells. In general, electron microscopy of EBV associated cancer cells does not detect EBV particles, but malignant cell nuclei do show viral DNA [78]

The model based on the present work is potentially actionable. The present evidence adds support for developing EBV treatment and a childhood herpes vaccine. EBV causes about 200,000 cancers per year of multiple different types. The prospects for producing an EBV vaccine are promising, but the most appropriate targets are still not settled. Some immunotherapy strategies rely on augmenting the immune response, but this approach may need modification because mutations create additional holes in the immune response.

## References

1. Sun Y, McCorvie TJ, Yates LA, Zhang X: **Structural basis of homologous recombination.** *Cellular and molecular life sciences : CMLS* 2020, **77**(1):3-18.
2. Murashko MM, Stasevich EM, Schwartz AM, Kuprash DV, Uvarova AN, Demin DE: **The Role of RNA in DNA Breaks, Repair and Chromosomal Rearrangements.** *Biomolecules* 2021, **11**(4).
3. Garcia-de-Teresa B, Rodriguez A, Frias S: **Chromosome Instability in Fanconi Anemia: From Breaks to Phenotypic Consequences.** *Genes (Basel)* 2020, **11**(12).

4. Arlt MF, Xu B, Durkin SG, Casper AM, Kastan MB, Glover TW: **BRCA1 is required for common-fragile-site stability via its G2/M checkpoint function.** *Molecular and cellular biology* 2004, **24**(15):6701-6709.
5. Voutsinos V, Munk SHN, Oestergaard VH: **Common Chromosomal Fragile Sites-Conserved Failure Stories.** *Genes (Basel)* 2018, **9**(12).
6. Xu B, Kim S, Kastan MB: **Involvement of Brca1 in S-phase and G(2)-phase checkpoints after ionizing irradiation.** *Molecular and cellular biology* 2001, **21**(10):3445-3450.
7. Venkitaraman AR: **Tumour suppressor mechanisms in the control of chromosome stability: insights from BRCA2.** *Molecules and cells* 2014, **37**(2):95-99.
8. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL *et al*: **Signatures of mutational processes in human cancer.** *Nature* 2013, **500**(7463):415-421.
9. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, Jones D, Hinton J, Marshall J, Stebbings LA *et al*: **Mutational Processes Molding the Genomes of 21 Breast Cancers.** *Cell* 2012, **149**(5):979-993.
10. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC *et al*: **Author Correction: Landscape of somatic mutations in 560 breast cancer whole-genome sequences.** *Nature* 2019, **566**(7742):E1.
11. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, Alexandrov LB, Martin S, Wedge DC *et al*: **Landscape of somatic mutations in 560 breast cancer whole-genome sequences.** *Nature* 2016, **534**(7605):47-54.
12. Griffiths DJ: **Endogenous retroviruses in the human genome sequence.** *Genome Biol* 2001, **2**(6):REVIEWS1017.
13. Ueda MT, Kryukov K, Mitsuhashi S, Mitsuhashi H, Imanishi T, Nakagawa S: **Comprehensive genomic analysis reveals dynamic evolution of endogenous retroviruses that code for retroviral-like protein domains.** *Mob DNA* 2020, **11**:29.
14. Xiang Y, Liang H: **The Regulation and Functions of Endogenous Retrovirus in Embryo Development and Stem Cell Differentiation.** *Stem Cells Int* 2021, **2021**:6660936.
15. Goke J, Ng HH: **CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome.** *EMBO Rep* 2016, **17**(8):1131-1144.
16. Xu M, Zhang WL, Zhu Q, Zhang S, Yao YY, Xiang T, Feng QS, Zhang Z, Peng RJ, Jia WH *et al*: **Genome-wide profiling of Epstein-Barr virus integration by targeted sequencing in Epstein-Barr virus associated malignancies.** *Theranostics* 2019, **9**(4):1115-1124.
17. Thorland EC, Myers SL, Gostout BS, Smith DI: **Common fragile sites are preferential targets for HPV16 integrations in cervical tumors.** *Oncogene* 2003, **22**(8):1225-1237.
18. Jiang S, Yang Z, Li W, Li X, Wang Y, Zhang J, Xu C, Chen PJ, Hou J, McCrae MA *et al*: **Re-evaluation of the carcinogenic significance of hepatitis B virus integration in hepatocarcinogenesis.** *PloS one* 2012, **7**(9):e40363.
19. Buschle A, Mrozek-Gorska P, Cernilogar FM, Ettinger A, Pich D, Krebs S, Mocanu B, Blum H, Schotta G, Straub T *et al*: **Epstein-Barr virus inactivates the transcriptome and disrupts the chromatin architecture of its host cell in the first phase of lytic reactivation.** *Nucleic acids research* 2021, **49**(6):3217-3241.
20. Chiu SH, Wu CC, Fang CY, Yu SL, Hsu HY, Chow YH, Chen JY: **Epstein-Barr virus BALF3 mediates genomic instability and progressive malignancy in nasopharyngeal carcinoma.** *Oncotarget* 2014, **5**(18):8583-8601.
21. Wu CC, Liu MT, Chang YT, Fang CY, Chou SP, Liao HW, Kuo KL, Hsu SL, Chen YR, Wang PW *et al*: **Epstein-Barr virus DNase (BGLF5) induces genomic instability in human epithelial cells.** *Nucleic acids research* 2010, **38**(6):1932-1949.



22. Munz C: **Latency and lytic replication in Epstein-Barr virus-associated oncogenesis.** *Nature reviews Microbiology* 2019, **17**(11):691-700.
23. Morales-Sanchez A, Fuentes-Panana EM: **The Immunomodulatory Capacity of an Epstein-Barr Virus Abortive Lytic Cycle: Potential Contribution to Viral Tumorigenesis.** *Cancers (Basel)* 2018, **10**(4).
24. Germini D, Sall FB, Shmakova A, Wiels J, Dokudovskaya S, Drouet E, Vassetzky Y: **Oncogenic Properties of the EBV ZEBRA Protein.** *Cancers (Basel)* 2020, **12**(6).
25. Hau PM, Lung HL, Wu M, Tsang CM, Wong KL, Mak NK, Lo KW: **Targeting Epstein-Barr Virus in Nasopharyngeal Carcinoma.** *Front Oncol* 2020, **10**:600.
26. Xu M, Yao Y, Chen H, Zhang S, Cao SM, Zhang Z, Luo B, Liu Z, Li Z, Xiang T *et al*: **Genome sequencing analysis identifies Epstein-Barr virus subtypes associated with high risk of nasopharyngeal carcinoma.** *Nature genetics* 2019, **51**(7):1131-1136.
27. Lung RW, Tong JH, Ip LM, Lam KH, Chan AW, Chak WP, Chung LY, Yeung WW, Hau PM, Chau SL *et al*: **EBV-encoded miRNAs can sensitize nasopharyngeal carcinoma to chemotherapeutic drugs by targeting BRCA1.** *Journal of cellular and molecular medicine* 2020, **24**(22):13523-13535.
28. Dai W, Chung DL, Chow LK, Yu VZ, Lei LC, Leong MM, Chan CK, Ko JM, Lung ML: **Clinical Outcome-Related Mutational Signatures Identified by Integrative Genomic Analysis in Nasopharyngeal Carcinoma.** *Clinical cancer research : an official journal of the American Association for Cancer Research* 2020, **26**(24):6494-6504.
29. Valouev A, Weng Z, Sweeney RT, Varma S, Le QT, Kong C, Sidow A, West RB: **Discovery of recurrent structural variants in nasopharyngeal carcinoma.** *Genome research* 2014, **24**(2):300-309.
30. Fina F, Romain S, Ouafik L, Palmari J, Ben Ayed F, Benharkat S, Bonnier P, Spyrtos F, Foekens JA, Rose C *et al*: **Frequency and genome load of Epstein-Barr virus in 509 breast cancers from different geographical areas.** *British journal of cancer* 2001, **84**(6):783-790.
31. Peng J, Wang T, Zhu H, Guo J, Li K, Yao Q, Lv Y, Zhang J, He C, Chen J *et al*: **Multiplex PCR/mass spectrometry screening of biological carcinogenic agents in human mammary tumors.** *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology* 2014, **61**(2):255-259.
32. Hu H, Luo ML, Desmedt C, Nabavi S, Yadegarynia S, Hong A, Konstantinopoulos PA, Gabrielson E, Hines-Boykin R, Pihan G *et al*: **Epstein-Barr Virus Infection of Mammary Epithelial Cells Promotes Malignant Transformation.** *EBioMedicine* 2016, **9**:148-160.
33. Ayee R, Ofori MEO, Wright E, Quaye O: **Epstein Barr Virus Associated Lymphomas and Epithelia Cancers in Humans.** *Journal of Cancer* 2020, **11**(7):1737-1750.
34. Lorenzetti MA, De Matteo E, Gass H, Martinez Vazquez P, Lara J, Gonzalez P, Preciado MV, Chabay PA: **Characterization of Epstein Barr virus latency pattern in Argentine breast carcinoma.** *PloS one* 2010, **5**(10):e13603.
35. Marrao G, Habib M, Paiva A, Bicout D, Fallecker C, Franco S, Fafi-Kremer S, Simoes da Silva T, Morand P, Freire de Oliveira C *et al*: **Epstein-Barr virus infection and clinical outcome in breast cancer patients correlate with immune cell TNF-alpha/IFN-gamma response.** *BMC cancer* 2014, **14**:665.
36. Maccaroni K, Balzano E, Mirimao F, Giunta S, Pelliccia F: **Impaired Replication Timing Promotes Tissue-Specific Expression of Common Fragile Sites.** *Genes (Basel)* 2020, **11**(3).
37. Friedenson B: **Mutations in components of antiviral or microbial defense as a basis for breast cancer.** *Functional & integrative genomics* 2013, **13**(4):411-424.
38. Kumar R, Nagpal G, Kumar V, Usmani SS, Agrawal P, Raghava GPS: **HumCFS: a database of fragile sites in human chromosomes.** *BMC genomics* 2019, **19**(Suppl 9):985.

39. Mount DW: **Using the Basic Local Alignment Search Tool (BLAST)**. *CSH Protoc* 2007, **2007**:pdb top17.
40. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *Journal of molecular biology* 1990, **215**(3):403-410.
41. Lu F, Wikramasinghe P, Norseen J, Tsai K, Wang P, Showe L, Davuluri RV, Lieberman PM: **Genome-wide analysis of host-chromosome binding sites for Epstein-Barr Virus Nuclear Antigen 1 (EBNA1)**. *Virology journal* 2010, **7**:262.
42. Kim KD, Tanizawa H, De Leo A, Vladimirova O, Kossenkova A, Lu F, Showe LC, Noma KI, Lieberman PM: **Epigenetic specifications of host chromosome docking sites for latent Epstein-Barr virus**. *Nat Commun* 2020, **11**(1):877.
43. Xiao K, Yu Z, Li X, Li X, Tang K, Tu C, Qi P, Liao Q, Chen P, Zeng Z *et al*: **Genome-wide Analysis of Epstein-Barr Virus (EBV) Integration and Strain in C666-1 and Raji Cells**. *Journal of Cancer* 2016, **7**(2):214-224.
44. Needleman SBaW, C.D.: **A general method applicable to search for similarities in the amino acid sequence of two proteins**. *Journal of molecular biology* 1970, **48**:453-453.
45. Wolfe D, Dudek S, Ritchie MD, Pendergrass SA: **Visualizing genomic information across chromosomes with PhenoGram**. *BioData Min* 2013, **6**(1):18.
46. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, Hackl H, Trajanoski Z: **Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade**. *Cell reports* 2017, **18**(1):248-262.
47. Lynn DJ, Winsor GL, Chan C, Richard N, Laird MR, Barsky A, Gardy JL, Roche FM, Chan TH, Shah N *et al*: **InnateDB: facilitating systems-level analyses of the mammalian innate immune response**. *Molecular systems biology* 2008, **4**:218.
48. Ortutay C, Siemala M, Vihinen M: **ImmTree: database of evolutionary relationships of genes and proteins in the human immune system**. *Immunome Res* 2007, **3**:4.
49. Ortutay C, Siemala M, Vihinen M: **Molecular characterization of the immune system: emergence of proteins, processes, and domains**. *Immunogenetics* 2007, **59**(5):333-348.
50. Denner J: **The porcine virome and xenotransplantation**. *Virology journal* 2017, **14**(1):171.
51. Prusty BK, zur Hausen H, Schmidt R, Kimmel R, de Villiers EM: **Transcription of HERV-E and HERV-E-related sequences in malignant and non-malignant human haematopoietic cells**. *Virology* 2008, **382**(1):37-45.
52. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR: **Deciphering signatures of mutational processes operative in human cancer**. *Cell reports* 2013, **3**(1):246-259.
53. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, Lo R, Winsor GL, Hancock RE, Brinkman FS, Lynn DJ: **InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation**. *Nucleic acids research* 2013, **41**(Database issue):D1228-1233.
54. Xiao J, Li W, Zheng X, Qi L, Wang H, Zhang C, Wan X, Zheng Y, Zhong R, Zhou X *et al*: **Targeting 7-Dehydrocholesterol Reductase Integrates Cholesterol Metabolism and IRF3 Activation to Eliminate Infection**. *Immunity* 2020, **52**(1):109-122 e106.
55. Chow KT, Driscoll C, Loo YM, Knoll M, Gale M, Jr.: **IRF5 regulates unique subset of genes in dendritic cells during West Nile virus infection**. *J Leukoc Biol* 2019, **105**(2):411-425.
56. Minskaia E, Saraiva BC, Soares MMV, Azevedo RI, Ribeiro RM, Kumar SD, Vieira AIS, Lacerda JF: **Molecular Markers Distinguishing T Cell Subtypes With TSDR Strand-Bias Methylation**. *Front Immunol* 2018, **9**:2540.
57. Zhu WP, Liu ZY, Zhao YM, He XG, Pan Q, Zhang N, Zhou JM, Wang LR, Wang M, Zhan DH *et al*: **Dihydropyrimidine dehydrogenase predicts survival and response to interferon-alpha in hepatocellular carcinoma**. *Cell Death Dis* 2018, **9**(2):69.

58. Grondin JA, Kwon YH, Far PM, Haq S, Khan WI: **Mucins in Intestinal Mucosal Defense and Inflammation: Learning From Clinical and Experimental Studies.** *Frontiers in immunology* 2020, **11**:2054.
59. Ibanez-Cabellos JS, Seco-Cervera M, Osca-Verdegal R, Pallardo FV, Garcia-Gimenez JL: **Epigenetic Regulation in the Pathogenesis of Sjogren Syndrome and Rheumatoid Arthritis.** *Front Genet* 2019, **10**:1104.
60. Mead TJ, Apte SS: **ADAMTS proteins in human disorders.** *Matrix Biol* 2018, **71-72**:225-239.
61. Reverter A, Ballester M, Alexandre PA, Marmol-Sanchez E, Dalmau A, Quintanilla R, Ramayo-Caldas Y: **A gene co-association network regulating gut microbial communities in a Duroc pig population.** *Microbiome* 2021, **9**(1):52.
62. Na T, Zhang K, Yuan BZ: **The DLC-1 tumor suppressor is involved in regulating immunomodulation of human mesenchymal stromal /stem cells through interacting with the Notch1 protein.** *BMC cancer* 2020, **20**(1):1064.
63. Aljohmani A, Yildiz D: **A Disintegrin and Metalloproteinase-Control Elements in Infectious Diseases.** *Front Cardiovasc Med* 2020, **7**:608281.
64. Florez MA, Matatali KA, Jeong Y, Ortinau L, Shafer PW, Lynch AM, Jaksik R, Kimmel M, Park D, King KY: **Interferon Gamma Mediates Hematopoietic Stem Cell Activation and Niche Relocalization through BST2.** *Cell reports* 2020, **33**(12):108530.
65. Amann-Zalcenstein D, Tian L, Schreuder J, Tomei S, Lin DS, Fairfax KA, Bolden JE, McKenzie MD, Jarratt A, Hilton A *et al*: **A new lymphoid-primed progenitor marked by Dach1 downregulation identified with single cell multi-omics.** *Nature immunology* 2020, **21**(12):1574-1584.
66. Wen AY, Sakamoto KM, Miller LS: **The role of the transcription factor CREB in immune function.** *Journal of immunology* 2010, **185**(11):6413-6419.
67. Ivan DC, Walther S, Berve K, Steudler J, Locatelli G: **Dwellers and Trespassers: Mononuclear Phagocytes at the Borders of the Central Nervous System.** *Frontiers in immunology* 2020, **11**:609921.
68. Dheekollu J, Wiedmer A, Ayyanathan K, Deakyne JS, Messick TE, Lieberman PM: **Cell-cycle-dependent EBNA1-DNA crosslinking promotes replication termination at oriP and viral episome maintenance.** *Cell* 2021, **184**(3):643-654 e613.
69. De Leo A, Calderon A, Lieberman PM: **Control of Viral Latency by Episome Maintenance Proteins.** *Trends in microbiology* 2020, **28**(2):150-162.
70. Jiang L, Lung HL, Huang T, Lan R, Zha S, Chan LS, Thor W, Tsoi TH, Chau HF, Borestrom C *et al*: **Reactivation of Epstein-Barr virus by a dual-responsive fluorescent EBNA1-targeting agent with Zn(2+)-chelating function.** *Proceedings of the National Academy of Sciences of the United States of America* 2019.
71. Kirchner EA, Bornkamm GW, Polack A: **Transcriptional activity across the Epstein-Barr virus genome in Raji cells during latency and after induction of an abortive lytic cycle.** *The Journal of general virology* 1991, **72** ( Pt 10):2391-2398.
72. **Correction for Jiang et al., Reactivation of Epstein-Barr virus by a dual-responsive fluorescent EBNA1-targeting agent with Zn(2+)-chelating function.** *Proceedings of the National Academy of Sciences of the United States of America* 2020, **117**(10):5542.
73. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M: **Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing.** *Genome research* 2014, **24**(7):1053-1063.
74. Pfeiffer F, Grober C, Blank M, Handler K, Beyer M, Schultze JL, Mayer G: **Systematic evaluation of error rates and causes in short samples in next-generation sequencing.** *Scientific reports* 2018, **8**(1):10950.

75. Friedenson B: **Many Breast Cancer Mutations Parallel Mutations in Known Viral Cancers.** *Journal of Genomes and Exomes* 2014, **3**(4437-JGE-Many-Breast-Cancer-Mutations-Parallel-Mutations-in-Known-Viral-Cancers-.pdf):17-35.
76. Friedenson B: **Mutations in Breast Cancer Exome Sequences Predict Susceptibility to Infections and Converge on the Same Signaling Pathways.** *J Genomes and Exomes* <http://mrcrossreforg/iPage?doi=104137%2FJGES30058> 2015, **4**:1-28.
77. Bhela S, Mulik S, Reddy PB, Richardson RL, Gimenez F, Rajasagi NK, Veiga-Parga T, Osmand AP, Rouse BT: **Critical role of microRNA-155 in herpes simplex encephalitis.** *Journal of immunology* 2014, **192**(6):2734-2743.
78. Oker NKN-I, and Busson, P: **Review: Biological and Pharmacological Basis for Cytolytic Viral Activation in EBV-Associated Nasopharyngeal Carcinoma.** *InTech Open* 2016:149-178.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1), Figure S1: title

**Funding:** Please add: This research received no external funding

**Conflicts of Interest:** The author declares no conflict of interest.