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

Bernard Friedenson

Dept. of Biochemistry and Molecular Genetics

College of Medicine

University of Illinois Chicago

Correspondence to: Bernard Friedenson

email: bernief@uic.edu / molmeddoc@gmail.com

Telephone/text 847-827-1958/847-912-4216

Fax +18474433718

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, broken chromosome fragments get lost or reattach indiscriminately. These mistakes are characteristic of hereditary breast cancer. We tested the hypothesis that mistakes in reattaching broken chromosomes preferentially occur near viral sequences on human chromosomes. We tested millions of DNA bases around breast cancer breakpoints for similarities to all known viral DNA. DNA around breakpoints often closely matched the Epstein-Barr virus (EBV) tumor variants HKHD40 and HKNPC60. Almost all breakpoints were near EBV anchor sites, EBV tumor variant homologies, and EBV-associated regulatory marks. On chromosome 2, EBV binding sites accounted for 90% of breakpoints ($p < 0.0001$). On chromosome 4, 51/52 inter-chromosomal breakpoints were close to EBV variant sequences. Five viral anchor sites at critical genes were near breast cancer breakpoints. Twenty-five breast cancer breakpoints were within 1.25% of breakpoints in model EBV cancers. EBV-like sequence patterns around breast cancer breakpoints resemble gene fusion breakpoints in model EBV cancers. All BRCA1 and BRCA2 breast cancers had mutated genes essential for immune responses. Because of this immune compromise, herpes viruses can attach and produce nucleases that break chromosomes. Alternatively, anchored viruses can retard break repairs, whatever the causes. The results imply proactive treatment and prevention of herpes viral infections may benefit BRCA mutation carriers.

Keywords

Breast cancer infection, breast cancer immunity, breast cancer virus, genome biochemistry, nasopharyngeal cancer, herpes viruses, hereditary breast cancer, BRCA1, BRCA2, Burkitt's lymphoma, homologous recombination, DNA repair, viral cancer, chromosome breaks

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 needed to restore broken DNA by homologous recombination. If mutations inactivate either the BRCA1 or BRCA2 gene, then the broken DNA can become pathogenic. Pieces of DNA get lost or reattach at the wrong positions on the original or different chromosomes. In carriers of BRCA1 or BRCA2 gene mutations, these errors lead to chromosome rearrangements and shifts typical of hereditary breast cancers. Landmark studies of Nik-Zainal and colleagues report hundreds of DNA rearrangements and characteristic rearrangement signatures[1-4]. Chromosome rearrangements may be critical events leading to hereditary breast cancers, but we know very little about what causes these events.

About 8% of human DNA probably originated from retroviruses [5] and both endogenous and exogenous retrovirus sequence insertions are a significant source of genome instability. Exogenous human herpesvirus 4 (EBV) can destabilize the genome by inducing DNA breaks. 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 associates with a diverse group of malignancies. EBV infection is always present in undifferentiated nasopharyngeal cancer (NPC) [6-8]. Downregulation or mutations in DNA repair genes related to BRCA1-BRCA2 mediated repair pathways are common in NPC [9,10]. Because of repair defects, NPC shows chromosomal shifts and rearrangements, like those seen in hereditary breast cancers.

HKNPC60 and HKHD40 are two EBV tumor virus variants 100% associated with NPC [11]. NPC in human epithelial cells arises after viral DNases induce DNA strand breaks followed by inappropriate gene fusions, such as YAP1-MAML2, PTPLB-RSRC1, and SP3-PTK2 [12]. EBV produces an exonuclease (BGLF5) and an endonuclease (BALF3). These viral products cause DNA strand breaks and genome instability [13,14], characteristic of cancers in EBV infected lymphoid cells [15]. Fortunately, a signaling pathway dependent on BRCA1 and BRCA2 genes prevents EBV from causing malignant transformation [16].

The same EBV-sensitive, BRCA-related pathway is also essential to prevent hereditary breast cancers, suggesting a role for EBV infection in hereditary breast cancers. Some previous literature supports this relationship. EBV infection of breast epithelial cell models

facilitates malignant transformation and tumor formation [17]. Epidemiological associations between breast cancer and EBV infection exist in different geographical locations [18,19]. Breast cancer cells from biopsies express gene products from latent EBV infection (LMP-1, -2, EBNA-, and EBER) [20] even after excluding the possibility that the virus comes from lymphocytes [21]. The EBV lytic form in breast cancer associates with a worse outcome [22]. Even though breast cancer malignancies may not have active EBV infection, the virus may still leave its footprints. Carriers of hereditary mutations in BRCA1 and BRCA2 genes would be especially susceptible to damage from viral infection.

Immune deficits in BRCA1 and BRCA2 mutation carriers [23] may allow the reactivation of latent EBV infections or new herpes viral infections. DNA breaks induced by EBV nucleases are then challenging to repair correctly and may become pathogenic. The availability of breast cancer genomic sequences permits testing the possibility that EBV contributes to the incorrect reattachment of broken chromosomes in hereditary breast cancer.

Materials and Methods

Breast cancer genomic sequences

Characteristics of breast cancers compared to viral cancers. Breast cancer genome sequences were from the COSMIC database curated from an original publication [3,4]. 15/25 breast cancers were stage III, four were stage II, and three had no data. Six breast cancers had typed germ-line BRCA1 mutations, and nineteen had typed germ-line BRCA2 mutations. Blood was the source of normal genes for comparison [4]. 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	BRCA1 / BRCA2 mutation status	Analyzed genes	Muta- tions	Breaks	Muta- tions per gene	Muta- tions per break	Chro- some with most breaks "From"	Chro- some with most breaks "To"	Most often intra-chro- mosomal breaks
PD3890	BRCA1	5978	13401	269	2.24	49.8	3=4,1	11, 16,21	1
PD3904	BRCA1	6337	13559	247	2.14	54.9	6	17,18	8
PD3945	BRCA1	9203	23132	114	2.51	202.9	1	X	2
PD4005	BRCA1	6943	15093	185	2.17	81.6	3	5	1
PD4006	BRCA1 R1835X	7433	20453	396	2.75	51.6	1 (only 1)	22,16	2 (387 intra)
PD4115	BRCA1*	9069	22662	180	2.50	125.9	2	17	8
PD4116	BRCA2	7831	19958	359	2.55	55.6	11	17	3
PD4836	BRCA2	4782	5381	80	1.13	67.3	10,11,12	18,20	10
PD4872	BRCA2	5095	5906	90	1.16	65.6	2	18	13
PD4874	BRCA2	8201	10668	226	1.30	47.2	6	20	10
PD4875	BRCA2	6262	8056	140	1.29	57.5	5	19	7
PD4876	BRCA2	5718	8317	95	1.45	87.5	5	9	12
PD4951	BRCA2	3178	3283	33	1.03	99.5	6	8	4
PD4952	BRCA2	10806	15155	332	1.40	45.6	3	14	14
PD4953	BRCA2	6690	8090	143	1.21	56.6	4	18	18
PD4954	BRCA2	5904	6605	126	1.12	52.4	1	8	2
PD4955	BRCA2	6654	7920	100	1.19	79.2	12	15	1
PD4956	BRCA2	11036	14860	238	1.35	62.4	7	10	3
PD4957	BRCA2	3521	3520	74	1.00	47.6	7=10	17=11	8
PD4958	BRCA2	8244	10956	102	1.33	107.4	1	1	4
PD6406	BRCA2	4976	5485	191	1.10	28.7	2,4	6,8	2
PD6416	BRCA2	3364	3609	71	1.07	50.8	1	20	7
PD7217	BRCA2	8841	11215	124	1.27	90.4	3	18	3
PD8621	BRCA2	7964	9617	275	1.21	35.0	1	X	1
PD8969	BRCA2	7335	8829	126	1.20	70.1	1	8	9

*Mutation of uncertain significance

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, $p = 0.03$)

In some cases, we also used breast cancer genome data from presumptive mutation

carriers. These women were all under age 36 but had not been tested for BRCA mutations.

DNA sequence data. We used the NCBI BLAST program (MegaBLAST) and database [24,25] to compare 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 excluded. Gene break-points for inter-chromosomal and intra-chromosomal translocations were obtained from the COSMIC catalog of somatic mutations as curated from original publications [4] and converted to the GrCH38 human genome version. DNA sequences at breakpoints were downloaded primarily from the UCSC genome browser. The Ensembl genome browser provided control comparison tests.

EBV genome comparisons. EBV binding locations on human chromosomes were obtained from publications [26-28], from databases, by interpolation from published figures, or by determining the location of genes within EBV binding sites. EBV binding data was based on lymphoblastoid and nasopharyngeal cancer cell lines. The NCBI BLAST program (megaBLAST) determined maximum homology scores between human DNA and all sequenced viruses in genome regions where EBNA1 binding occurred before anchoring EBV. When necessary, genome coordinates were all converted to the GrCH38 version. We further compared EBV binding sites to hereditary breast cancers with respect to chromosome break positions, epigenetic marks on chromatin, genes, and copy number variations. Locations of H3K9Me3 chromatin epigenetic modifications were from the MIT Integrated Genome Viewer (IGV) with ENCODE data loaded and from the UCSC genome browser. Data from the ENCODE website. (www.ENCODEproject.org) also provided positions of H3K9Me3 marks. Homology among viruses was determined by the method of Needleman and Wunsch [29].

Data analyses. Microsoft Excel, OriginPro, Visual basic, and Python scripts provided data analysis. Chromosome annotation software was from the NCBI Genome Decoration page and the Ritchie lab using the standard algorithm for spacing [30]. 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.

Genes associated with the immune response damaged in breast cancer. We compared breast cancer somatic mutations to the immune metagenome [31-34]. However, computerized lists of immune system gene functions vs. breast cancer mutations contain

many potential uncertainties. Sets of genes involved in immune responses also mediate other functions and represent a vast and growing dataset (geneontology.org). Immune defense genes now extend to three different immune systems (intrinsic, innate, and adaptive). Sets of genes involved in cancer control by immune surveillance and immunediting are not well characterized. To address these uncertainties, we determined whether mutated genes fit into known immune defense pathways. 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]. This extensive validation permitted including both direct and indirect effects of gene mutations.

Results

Fig. 1 shows virus-human homology comparisons of one inter-chromosomal breakpoint in PD3945, a BRCA1 breast cancer (a) and one in a BRCA2 associated PD4874 a breast cancer (b). Panel (c) shows an intra chromosomal breakpoint. The comparisons include 450,000 bases extending in both directions from each breakpoint. The mega-BLAST program then compared the resulting 900,000 human base pairs around the breakpoint to all known viral sequences. Fig. 1 shows many DNA segments that are virtually identical to EBV variants (Human gamma-herpesvirus 4 variants, “HKNPC60” or “HKHD40”) at or near the different inter-chromosomal breakpoints shown. A chromosome fragment deleted in panel (C) includes virus-like sequences and has nearby viral matching sequences. Maximum homology scores between human DNA vs. herpes viral DNA are over 4000 for breast cancer PD3945 and just under 4000 for PD4874. Matches between human and viral sequences of about 97% for up to 2462 base pairs, with E “expect” values (essentially p-values) equal to 0. Many matches of shorter length in Fig.1 were also statistically significant.

In Fig. 1, many human herpes viral sequences were within the viral DNA length (~175,000 base pairs) of the linear lytic infection form. Bound latent EBV circular episomes could also interfere with the large complex apparatus required for complete DNA restoration by homologous recombination. Viral homologies suggest a large number of EBV binding sites with a substantial percentage at repetitive sequences. Lu et al. found 4785 EBV binding sites with over 50% overlapping a repetitive sequence element [26].

Fig.1. Examples of viral homologies at breast cancer breakpoints. Chromosome coordinates for locating the region with viral homologies are at the top of each panel. The lower horizontal axis represents distance from the breakpoint in thousands or millions of base pairs. The “y” coordinates are the maximum homology scores from megaBLAST analysis. The blue dots represent the start points of the homologous human-viral sequences. 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 ebv 60 sequences and includes them. Weaker agreement with retroviral sequences is at greater distances at the lower right of the panel.

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

Fig. 2 shows the positions where inter-chromosomal breaks occur on 19 BRCA2 associated breast cancers, listed numerically at the bottom of the figure. There are significant differences in how the breakpoints distribute on different chromosomes. Spaces between breakpoints are roughly the same throughout chromosome 8. Breakpoints are much sparser on some other chromosomes, e. g., chromosomes 9, 12, and 21. Chromosome 8

Fig. 2 Distribution of chromosome breakpoints in BRCA2 mutation associated breast cancers over all human 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 chromosome positions where the break occurs. The distributions are not random, with large regions of chromosomes that do not have breaks. Some breast cancer breaks appear to focus on a few areas and others are more widely distributed. Since all the breast cancers occurred in females, there are no breaks on the Y chromosome.

was the site of the most numerous chromosome breaks. Inter-chromosomal breakpoints tend to cluster in specific chromosome regions for individual breast cancers. A few chromosome areas in Fig 2 show inter-chromosome rearrangements occur in only one cancer. Large numbers of base pairs often separate the breakpoint positions, indicated by horizontal lines on the chromosome diagrams. Breast cancer PD4956 has a large breakpoint cluster on chromosome 11, and PD4952 has a group on chromosome 14. PD8621 is one of only a few BRCA2-related breast cancers with breaks on chromosome 21.

Fig. 3. Statistically significant maximum homology scores (blue dots) for herpes virus DNA vs human DNA at inter-chromosome breakpoints on chromosome 4. Nearly all (51/52) inter - chromosome breakpoints on chromosome 4 in all BRCA2 associated breast cancers occur within 200K base pairs of EBV-like sequences. The value of 200K base pairs is approximately the length of the lytic EBV sequences allowing for some error and for tethering by EBNA1 dimers. The horizontal x axis represents the entire length of chromosome 4 and the Y axis shows the maximum homology scores. Values above 4000 represent about 97% homology over about 2500 base pairs. E (expect) values (essentially p values) = 0. Only homology scores over 100 are shown.

Breast cancer breakpoint homologies to EBV are frequent.

Chromosome 4 is one of the preferred landing sites for EBV infection. Breast cancer breakpoints cluster in inter-chromosome rearrangements on chromosome 4. To obtain an estimate of how often breakpoint regions have homology to herpes viruses, we tested every inter-chromosomal BRCA2 donor breakpoint on chromosome 4 for homology to herpes viruses. Nearly all (51/52) breakpoints had statistically significant homology to herpes viruses within about 200k base pairs (Fig. 3).

Breast cancer breakpoint homologies to EBV are near to known EBV binding sites.

Chromosome 2. Fig. 4 shows viral homologies in a 21 million base pair section of chromosome 2. The pink dotted lines in Fig. 4 indicate positions where EBV binding sites align with breast cancer breakpoints. We calculated R^2 by comparing the coordinates of EBV binding sites to breakpoints. The R^2 value of 0.91 estimates that EBV homology accounts for about 91% of the breakpoint position differences. The normal plot of the residuals from this analysis is approximately linear, suggesting that the residuals are distributed relatively randomly about zero.

Human reference sequence genes in this region of chromosome 2 include KYNU, GTDC1, ACVR2A, KIF5C, STAM2, KCNJ3, ERMN, PKP4, BAZ2B, TANK, and DPP4 (bottom of figure). Chromosome breaks near 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.

Most homologies are again to EBV variants, but porcine endogenous retrovirus ("PERV," isolate PERV_VIRES_GX05_C1 env) [35] has two high scores. Three HERV sequences [36] including a pseudogene (pHERV), also have significant homology to human DNA. The porcine PERV sequence lies within a retroposed area on chromosome 2 close to EBV sequences, within 28 Kbps 5' and 80 Kbps 3' of EBV sequences. These EBV-like ends potentially generate homologies for retro-positioning and inserting the porcine sequences.

DNA surrounding many breakpoints had the EBV variant sequences (HKHD40 and HKNPC60) in both the donor and acceptor sequences. Even DNA sequences within

Fig.5. Breast cancer breakpoints on 14 million base pairs on chromosome 12 (green dots) in BRCA1, BRCA2, and presumptive BRCA associated (breast cancer at age ≤ 35) occur in close proximity to EBV anchor sites. Statistically significant homologies to EBV viruses occur for all but one breakpoint within 200k base pairs of an anchor site. Green dots show the breast cancer breakpoints. Dashed vertical lines show that almost all breakpoints are near EBV tethering sites (red dots at top of the graph). The blue dots indicate positions of homology to the EBV variants HKHD40 and HKNPC60, but a few dots at the right match human and pork endogenous retroviruses. The bottom of the figure shows that breakpoints disrupt ENCODE candidate cis-regulatory elements (cCREs), H3K27Ac regulatory marks, and some regions of active transcription. Cancer-associated (COSMIC) genes and gene interactions are also disrupted. The breaks all occur in chromosome regions with accessible DNA (DNase hypersensitivity). Breakpoint positions also agree with positions of the EBV-associated epigenetic mark H3K9Me3 in CD14+ primary monocytes (RO-01946). The breaks usually disrupt one or more reference sequence genes, and the region gives rise to structural variation, i.e., CNV's, inversions, and Insertion/deletions. Endogenous human and porcine retroviruses are less likely to affect the area and lie at the right.

intra-chromosomal deletions contained these variant EBV sequences (Fig. 1c)

Chromosome 12. On a 14 million base-pair section of chromosome 12, we aligned EBV binding sites on chromosome 12, breast cancer breakpoints, and homology between human sequences and tumor-related herpes viruses (Fig 5). All breakpoints are near regions of homology to the EBV variant tumor viruses HKNPC60 and HKHD40. Again, the maximum homology scores (over 4000) correspond to about 2500 bps that are 97% identical. All but one of the breakpoints is near an EBV anchoring site. A few homologies to human and porcine retroviruses are outside this region of chromosome 12. All the breakpoints appear to be accessible as indicated as DNase hypersensitivity.

All breakages disrupt regulatory properties shown in Fig 5. Results are consistent with data from epithelial cell lines showing EBNA1 binds cellular promoters to control cellular gene expression [37]. In the present work, the breakpoints all go through ENCODE candidate cis-regulatory elements (cCREs). Some breakpoints disrupt cancer-related (COSMIC) genes, transcription, gene interactions, and epigenetic marks such as H3K27Ac. At least 11 of the breakpoints disrupt reference genes (bottom of figure). The region also appears to be a focus for structural variation such as CNV's, inversions, and short Insertion/deletions. The graph shows H3K9Me3 chromatin mark positions based on ChIP-Seq signals for monocytes. These markings are known to occur around EBV sites and support the positioning of EBV anchor sites. H3K9Me3 marks are thought to repress host transcription at EBNA1 binding sites [27], and the chromosome 12 region shown is rich in these sites. H3K9Me3 distributions on the host cell might also contribute significantly to EBV latency [27].

Distributions of viral matches around different chromosome breakpoints correlate.

We tested six different breast cancer breakpoints on different chromosomes to determine whether their separations from viral sequences had any correlation. The distribution of viral sequences from breakpoints along different chromosomes in various breast cancers is strongly correlated. The correlation held for linear regression ($p=0.0093-0.000$), Kendall ($p<0.0001$), and Spearman tests ($p<0.0001$).

Identified EBV binding sites match breast cancer breakpoints.

After converting the binding site coordinates to the same system as breast cancers, EBV binding sites [26] are close to early-onset breast cancer breakpoints and HKHD40 / HKNPC60 viral-human homologies (Fig 6). Different EBV binding sites on chromosome 6 and

chromosome 1 (Including the CDC7 gene) were near an early-onset breast cancer breakpoint and HKHD40 / HKNPC60 viral-human homologies (Fig. 6). The known EBV binding site on chr5:141610249-141660249 was close to chromosome 5 breakpoints in breast cancer PD5945 (at 141,557,433 and 141,564,233). Intra-chromosomal breakpoints, which caused a deletion on chromosome 1, matched both the EBV variant sequences and the known anchor site. Fig. 6b shows the high-affinity anchor site for the EBV nuclear antigen (EBNA1) on chromosome 11[26]. The HDAC3 region on chromosome 5 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 perhaps involving retroviral participation in breast cancer breakpoints. Nonetheless, viral binding sites were always near at least one breast cancer breakpoint. Five of the six localized EBV binding sites were near HKHD40 / HKNPC60 human viral homologies.

Viral homologies around breakpoints in BRCA - associated breast cancers are difficult to distinguish from cancers caused by EBV.

To compare breakpoints in BRCA1,2 associated breast cancers with model cancers known to be caused by EBV, we tested sequences around known recurrent gene fusions in nasopharyngeal cancer and Burkitt's lymphoma (which has a 30-50% association with EBV). We compared the DNA within about 200k base pairs around breakpoints in these known EBV mediated cancers.

Fig. 7 shows gene fusions in nasopharyngeal cancer (as models for known EBV mediated cancers). These models resemble breast cancers in that they have homologies to EBV variants at breakpoints where fusions between genes occur. For example, the MAML2 gene, which rearranges in some NPC cancers, has homology to the EBV variants. The YAP1 breakpoint in its fusion to MAML2 occurs just within YAP1 gene boundaries. (Fig 7a). In Fig 7b, the PTK2 gene has strong similarities to EBV variants within its sequence. In Fig 7c, the RSRC1 breakpoint in the gene fusion RSRC1 to PTPLB is close to regions of strong EBV homology. In contrast, the breakpoint in PTPLB has weaker homology to endogenous and exogenous retroviral sequences. Five of the six gene fusions breakpoints are within the length of EBV variant number of base pairs from regions of EBV homology. In the PTPLB region of chromosome 3, there are numerous retroviral sequences near the breakpoint. Base pair distances from the EBV variant sequences are just outside the absolute values of the numbers of base pairs in the EBV variants.

Fig. 8 shows distances to EBV homologous regions starting from breakpoints in

Fig. 6. (a) A direct EBV binding site is near the region chr6: 98850000-98860000, which contains a LINE sequence (within the red dots) [26]. The area is near breakpoints in presumptive BRCA-associated breast cancers (green dots). Significant HKHD40 and HKNPC60 homologies overlap the LINE sequence. The break in breast cancer PD4833 is 70,858 bps from the EBV binding site. Weaker homology to the 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. Weaker retrovirus homologies are also shown. 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), EBV binding near the HDAC3 gene does not have EBV variant homologies within 200K base pairs, but breakpoints in breast cancer PD5945 are still close to the region. Somewhat weaker homologies exist with HIV1 variants and endogenous retroviruses.

Fig. 7. Gene fusions in nasopharyngeal cancers are near human sequences with strong homology to EBV tumor variants and sequences with weaker homology to retroviruses. 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 the total distance of viral homologies from the breakpoint. Orange dots indicate the start point of retroviral homologies.

Fig. 8. Distributions of EBV variant homologies around eight breakpoints (indicated by red lines) in nasopharyngeal cancer NPC-5989 [12] resemble distributions around hereditary breast cancers. Burkitt's lymphoma break coordinates are also near DNA sequences with EBV variant homologies (lower right panel). The numbers near each of the blue dots are the number of base pairs between the breakpoint and the start of the homologous sequences. The numbers on the horizontal axis are the numbers of base pairs on the chromosome in millions. The vertical axes are again the maximum homology scores vs. all viruses

nasopharyngeal cancer NPC-5989 calculated from data from reference 12. In NPC-5989, all the breakpoints on chromosomes 1, 2, 8, and 11 are close to human genome regions that resemble the EBV variants (Fig. 8).

Table 2 shows that some breakpoints in NPC and BL cancers are near hereditary breast cancer breakpoints. Most (18/25) cancer breaks are within the number of bases in the EBV reference sequence (~175,000). Comparable breakpoint positions in BRCA-breast vs. NPC vs. BL cancers differ by less than 1.25%. Table 2 also includes five values ranging from 181,246 to 274,186 base pairs. If the breakpoint's actual value is closer to the average of the breast and other cancer breakpoints, then the break position still lies within the number of bases in EBV. Normality plots of the breakpoint positions for nasopharyngeal cancer and Burkitt's lymphoma vs. breast cancers are identical, go through zero, and follow the line $y=x$. One set of data can calculate the other using the equation $\text{Breast Cancers} = 1.00021 \text{ NPC_BL} + 9012$ ($p < 0.0001$). An unpaired t-test showed differences between the two sets of data were not significant, $p < 0.001$.

Table 2 Comparisons of Cancer Chromosome Breakpoints

Chromosome	NPC /BL Break coordinate	Breast cancer break coordinate	Difference (number base pairs)	% Difference
1	151,928,027	151,922,525	5,502	0.00
1	155,119,764	155,020,623	99,141	0.06
1	154,117,249	154,111,293	5,956	0.00
1	197,593,291	197,867,477	274,186	0.14
1	203,179,378	203,428,861	249,483	0.12
1	204,435,983	204,386,502	49,481	0.02
2	66,324,118	66,473,128	149,010	0.22
3	158,537,092	158,531,593	4002	0.00
4	101,138,921	101,009,819	129,102	0.13
4	152804117	152,852,058	47,941	0.03
5	43086276	42897159	189,117	0.44
6	74142321	74016919	125,402	0.17
6	104668449	104,676,699	8,250	0.01
8	98073546	98,164,150	90,604	0.09
8	128,468,417	128,300,594	167,823	0.13
8	128,481,840	128,300,594	181,246	0.14
11	102,114,930	102,384,027	269,097	0.26
11	102,115,180	102,384,027	268,847	0.26
11	101,983,850	101,895,690	88,160	0.09

13	74133903	74,125,809	8,094	0.01
15	41381710	41,532,838	151,128	0.37
18	43243754	43103981	139,773	0.32
19	14770552	14,943,490	172,938	1.17
19	33307575	33,086,916	220,659	0.66
X	3,996,343	4,044,529	48,186	1.21

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

All 25 of the hereditary breast cancers had significant damage to genes needed for immune system functions (Table 3). A total of 1307 immune-related genes had mutations. Table 3 shows the 20 most frequently mutated genes in the breast cancers belong to the “immune genome”. Of these top 20 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.

Multiple components of immune defenses mutate in every one of 25 hereditary breast cancer genomes. Mutations also deregulate the immune response and contribute to cancer, e.g., by causing chronic inflammation. The mutations affect processes such as cytokine production, autophagy, etc. Every breast cancer had mutations affecting these functions or their regulation. The impaired functions depend on many genes dispersed throughout the genome, so mutation in only one gene can compromise a complex process. Each breast cancer genome has a different set of these mutations, with many genes only occasionally damaged in more than one breast cancer. Antigen and viral recognition, signaling essential to transmit the immune responses, and immune regulation can all be impaired. Damage affecting the nervous system was also universal. Some herpes viruses establish permanent infection within the central nervous system even after other sites clear [38].

Significant differences exist in distributions and numbers of mutations among the breast cancer genomes. Table 1 shows considerable differences in the numbers of DNA breaks [1,2,4,39]. Despite these differences, the significant common properties of many mutations are that they cripple some aspect of innate immunity, its regulation, or its connections to adaptive immunity. Damage to some genes can potentially deregulate immunity and inflammation in breast cancers (Table 3), damaging specific responses to antigens or pathogens. Deregulation could conceivably result from mutations alone

without exposure to antigens or pathogens, impairing host response to breast cancer whatever its cause.

Controls.

Viral variants match other herpes viruses. Both HKHD40 and HKNPC60 viruses were typical of many other herpesvirus isolates, with some haplotypes conferring a high NPC risk [8]. 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 1200000-1405000. HKHD40 gave values of 99 and 98% identity for comparisons to the same regions.

If the infection is active and lytic, many breakpoints lie within the length of bases in the EBV genome and EBV anchor sites. Alternatively, viral episome binding may block homologous recombination machinery from precisely correcting DNA breaks if EBV infection remains latent.

Breast cancers have similarities to known or likely viral cancers. Positive controls found 1143 identical genes mutated in breast cancers vs. viral cancers of the cervix. About 50% of the genes mutated in a set of breast cancers were identical to genes mutated in the viral cancer Burkitt's lymphoma and virally associated liver cancers. Human telomeric sequences did not have homology to EBV, e.g., DNA bases 1-1,000,000 on chromosome 11 had no homology to EBV.

The UCSC genome browser and the Ensembl genome browser gave indistinguishable results.

Discussion

These results implicate variants of EBV in hereditary breast cancers because viral anchors match chromosome breakpoints; because many breast cancer chromosome breakpoints are near human sequences that match viruses closely; and because patterns of viral homology are difficult to distinguish from cancers known to be caused by EBV. Viral variants are 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. Our comparisons of all virus sequences to the human genome reveal widespread, repeated EBV-like variant sequences dispersed throughout the human genome. In general, electron microscopy of EBV associated cancer cells does not detect EBV particles, but malignant cell nuclei do show viral DNA [40]

HKHD40 and HKNPC60 herpes viruses can exist as covalently closed circular episomes with multiple copies. These viral genomes assemble into chromatin, undergoing epigenetic histone and DNA modifications analogous to host genomes. This epigenetic programming of the viral episome controls the viral life cycle. The viruses switch from latent to lytic replication when signals from host cells indicate cell stress, DNA damage, another infection, or differentiation. In latently infected cells, the EBV protein EBNA1 binds host DNA and anchors viral episomes to host chromosomes at hundreds of sites. Viral genomes remain attached to host chromosomes as parasites during mitotic cell division. This attachment maintains stable numbers of episome copies in multiplying host cells [41,42]. However, attached EBNA1 and viral particles also interfere with host chromosome replication. This interference becomes so significant in mouse models that B-cell tumors develop [43].

In infected B-cells, EBV establishes premalignant latent gene expression programs that also exist in lymphomas. On B-cell activation, the virus switches to its lytic replication program. During an early stage of EBV lytic replication, the virus generates strong nicks in host cell DNA, causing host chromosome aberrations [14]. EBV encodes a DNase (BGLF5) that causes these nicks and induces genomic instability in human epithelial cells, inducing chromosome micronuclei formation in epithelial cells [14]. BGLF5 also shuts off host protein synthesis and interferes with host cell DNA repair. The enzyme is a potent inducer of genomic instability. Lytic EBV replication was previously thought to destroy infected cells, which would prevent linear, complete viral DNA from causing tumors. However, recent work establishes that lytic, full-length EBV replication helps drive EBV-mediated malignancies [43]. Lytic replication also produces BALF3, another EBV nuclease and potential culprit [13].

The present study really began when we found EBV-like sequences very near many breakpoints in BRCA1 and BRCA2-associated breast cancer genomes. Fig. 1 shows three examples of breast cancer breakpoint positions close to human genome sequences that strongly resemble viral variants. These viral similarities also exist within intra-chromosome regions that get deleted in individual breast cancers (Fig.1 bottom). BRCA1 and BRCA2 mutation carriers are likely to be highly susceptible to EBV variant nucleases, interference with replication, and recombinase activity. Mutations in BRCA1 and BRCA2 genes

inactivate the high-fidelity repair of EBV-induced chromosome damage.

Two explanations must be explored for the role of EBV-like variants in chromosome breaks in hereditary breast cancers. (1) Chromosome breaks in hereditary cancers occur because human DNA homologies to EBV-variants help guide viral variant binding in the epithelium of breast ducts, where viral nucleases damage the host chromosome. Alternatively, (2) pre-existing viral variant binding interferes with access to repair machinery so that breaks, whatever the cause, are slow to repair. In either event BRCA1 and BRCA2 gene deficits make repairs less accurate.

Of course, there are countless unrelated 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 are marked differences among breast cancers in the positions where these breakpoints occur (Fig. 2). These different distributions suggest that inter-chromosome breaks followed by incorrect repairs do not result from random processes but have some selectivity.

A consequence of this selectivity is that 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 200k in Fig. 3). Moreover, many breast cancer breakpoints are separated from viral homologous sequences by numbers of base pairs consistent with circular viral episomes (roughly 65000). Circular viral episomes could also be a dominant carcinogen.

Comparisons to actual sites of EBV anchoring by EBNA1 further supports a relationship between EBV and chromosome breaks in hereditary cancers. Comparisons on chromosome 2 (21 million base pairs) and chromosome 12 (24 million base pairs) give similar results (Figs 4-5). Short DNA comparisons for a few known EBV binding sites also show that viral binding, breast cancer breaks, and viral human homologies are near each other (Fig. 6). EBV anchors and EBV variant homologies fall within a deleted segment of chromosome 1 (shaded area) in early-onset breast cancer PD23566. Breakpoints on chromosome 6 from three breast cancers (PD4107, PD4833, and PD18024) surround EBV anchor sites and EBV variant homologies. A primary binding site on chromosome 11 [26] matches a breakpoint in an early-onset breast cancer (Fig. 6). Weaker EBV binding sites on chromosome 11 also agree with chromosome break positions in BRCA1 and BRCA2 breast cancers (data not shown).

Figs. 4-6 are consistent with EBV variants interfering with transcription. Chromosome sequences near EBV anchor sites are enriched in transcribed DNA, DNase accessible

sequences, and regulatory sequences. EBNA1 has linking modules that form aggregates, but the linkers also bind EBV episomes to metaphase chromosomes. In many cases, the positions of EBNA1 anchors, breast cancer breakpoints, and human regions of EBV homology all agree.

EBV binds to human chromosomes using EBNA1-dimers as anchors for subsequent viral attachment [44,45]. 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 [26]. EBNA1 binds close to transcription start sites for cellular genes and positively regulates them. These genes include MAP3K71, CDC7, and HDAC3. Fig. 6 shows that most of the EBV anchor sites near these genes are relatively close to chromosome breaks in breast cancers. HDAC3 on chromosome 5 is an exception. A region of chromosome 5 containing an EBV anchor site near HDAC3 does not have significant EBV variant homologies near an EBV binding site. Instead, there are much shorter and weaker nearby similarities to retroviruses.

Chromosome 12 shows strong homologies to porcine endogenous retroviruses just outside the test region. Data from chromosome 5 and chromosome 12 suggest that retroviruses may also participate in breast cancer breaks in agreement with much previous literature. Thoroughly cooking pork products should avoid danger from porcine endogenous retroviruses. 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. This extensive homology challenges the safety of transplantation from porcine donors.

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 [46] of breakpoint positions in breast cancers (Table 2). Running viral human homology comparisons for such accepted EBV-related cancers gave results that were difficult to distinguish from those obtained for breast cancers (Figs. 7-8). Five of six genes involved in gene fusions are near EBV variant sequences and include EBV-like variant sequences within the gene boundaries. The PTPLB gene has these EBV variant homologies at about 265-275,000 base pair distances with short, weak retroviral homologies much closer to the fusion breakpoint (Fig 7).

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 [23,47,48]. 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).

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 BRCA1 or BRCA2 breast cancers	Type of Immunity in database [49]	Example of function in immunity	Known or likely connection to the nervous system
AKT3	20	Innate	AKT3 amplifies innate immune responses to DNA or RNA virus infections [50]. AKT3 binds Interferon Response Factor 3, enhancing its activation and stimulating interferon responses.	✓
FRMD4A	20	Innate	Depends on Interferon Regulatory Factor 5 [51]. Affects antigen presentation and subsequent effects on dendritic cells. Mediated by cytoskeletal actin and endocytosis	✓
ARHGAP15	19	Innate	Negative regulator of neutrophil function. Affects mast cell function	✓
CAMTA1	19	Adaptive	Pattern of methylation distinguishes subsets of T-cells [52]	✓
DPYD	19	Innate	Interferon response [53]. Affects infection severity in mice. Natural killer cell function	✓
CHRM3	18	Adaptive	Cholinergic receptor muscarinic 3 Deficiency associated with immune impairment [54] Targeting by miRNAs related to inflammation [55].	✓
ADAMTS12	17	Adaptive	Member of a group of related proteins with many connections to the immune system and immune disorders [56]	✓
DAB1	17	Adaptive	Associated with microbial profile [57]	✓

Gene	Mutations in 25 BRCA1 or BRCA2 breast cancers	Type of Immunity in database [49]	Example of function in immunity	Known or likely connection to the nervous system
ADAMTS3	16	Innate	Member of a group of related proteins with many connections to the immune system and immune disorders [56] Mast cell function	✓
COL23A1	16	Adaptive	Enriched in plasmacytoid dendritic cells	✓
DLC1	16	Adaptive	Immune regulation of stem cells through interactions with NOTCH-1 [58] Type 2 T helper cell	✓
ADAM12	15	Adaptive	Controls response to infectious disease [59]. Central memory CD8 t-cells	✓
CD36	15	Adaptive	TLR4-TLR6-Cd36 activation is a common molecular mechanism by which atherogenic lipids and amyloid-beta stimulate sterile inflammation. Gamma delta T cell function	✓
CDH2	15	Innate	Maintains stem cell quiescence [60]. Natural killer cell function	✓
DACH1	15	Innate	Downregulated in lymphoid cell progenitor [61]	✓

Gene	Mutations in 25 BRCA1 or BRCA2 breast cancers	Type of Immunity in database [49]	Example of function in immunity	Known or likely connection to the nervous system
CREB5	14	Innate	CREB factors limit proinflammatory responses, provide anti-apoptotic signal, regulate Th1, Th2, Th3 responses, generate and maintain regulatory T cells [62]	✓
ANK1	13	Adaptive	Type 17 T helper cell	✓
HDAC9	13	Adaptive	Known to alter B-cell responses in autoimmune diseases Immature B cell function.	✓
COL4A1	12	Adaptive	Central memory CD4 T cell	✓
F13A1	12	Adaptive	Phagocytes at CNS borders [63]	✓

Mutational damage to genes essential for an immune response and structural barriers to tumor viruses facilitates infection. The relationships of the top 20 mutated genes to the nervous system may also favor herpes viral infections. Damage to the immune system in hereditary cancers suggests DNA breakages might result when viruses escape from control. For example, escape of EBV from latency would release nucleases that fragment human DNA. Moreover, abundant viral DNA might form heteroduplexes with pieces of human DNA that resemble viral sequences. EBNA1 anchor sites might facilitate interactions that lead to chromosomal translocations.

The finding of a relationship between breast cancer chromosome breaks and viruses is potentially actionable. Some immunotherapy strategies rely on augmenting the immune response, but this approach may need modification because mutations create additional holes in the immune response. 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 controversial.

References

1. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Aparicio, S.A.; Behjati, S.; Biankin, A.V.; Bignell, G.R.; Bolli, N.; Borg, A.; Borresen-Dale, A.L.; et al. Signatures of mutational processes in human

- cancer. *Nature* **2013**, 500, 415-421, doi:10.1038/nature12477.
2. Nik-Zainal, S.; Alexandrov, L.B.; Wedge, D.C.; Van Loo, P.; Greenman, C.D.; Raine, K.; Jones, D.; Hinton, J.; Marshall, J.; Stebbings, L.A.; et al. Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* **2012**, 149, 979-993, doi:10.1016/j.cell.2012.04.024.
3. Nik-Zainal, S.; Davies, H.; Staaf, J.; Ramakrishna, M.; Glodzik, D.; Zou, X.; Martincorena, I.; Alexandrov, L.B.; Martin, S.; Wedge, D.C.; et al. Author Correction: Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **2019**, 566, E1, doi:10.1038/s41586-019-0883-2.
4. Nik-Zainal, S.; Davies, H.; Staaf, J.; Ramakrishna, M.; Glodzik, D.; Zou, X.; Martincorena, I.; Alexandrov, L.B.; Martin, S.; Wedge, D.C.; et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **2016**, 534, 47-54, doi:10.1038/nature17676.
5. Griffiths, D.J. Endogenous retroviruses in the human genome sequence. *Genome Biol* **2001**, 2, REVIEWS1017, doi:10.1186/gb-2001-2-6-reviews1017.
6. Germini, D.; Sall, F.B.; Shmakova, A.; Wiels, J.; Dokudovskaya, S.; Drouet, E.; Vassetzky, Y. Oncogenic Properties of the EBV ZEBRA Protein. *Cancers (Basel)* **2020**, 12, doi:10.3390/cancers12061479.
7. Hau, P.M.; Lung, H.L.; Wu, M.; Tsang, C.M.; Wong, K.L.; Mak, N.K.; Lo, K.W. Targeting Epstein-Barr Virus in Nasopharyngeal Carcinoma. *Front Oncol* **2020**, 10, 600, doi:10.3389/fonc.2020.00600.
8. Xu, M.; Yao, Y.; Chen, H.; Zhang, S.; Cao, S.M.; 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, 1131-1136, doi:10.1038/s41588-019-0436-5.
9. Lung, R.W.; Tong, J.H.; Ip, L.M.; Lam, K.H.; Chan, A.W.; Chak, W.P.; Chung, L.Y.; Yeung, W.W.; Hau, P.M.; Chau, S.L.; et al. EBV-encoded miRNAs can sensitize nasopharyngeal carcinoma to chemotherapeutic drugs by targeting BRCA1. *Journal of cellular and molecular medicine* **2020**, 24, 13523-13535, doi:10.1111/jcmm.16007.
10. Dai, W.; Chung, D.L.; Chow, L.K.; Yu, V.Z.; Lei, L.C.; Leong, M.M.; Chan, C.K.; Ko, J.M.; Lung, M.L. 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, 6494-6504, doi:10.1158/1078-0432.CCR-20-2854.
11. Hui, K.F.; Chan, T.F.; Yang, W.; Shen, J.J.; Lam, K.P.; Kwok, H.; Sham, P.C.; Tsao, S.W.; Kwong, D.L.; Lung, M.L.; et al. High risk Epstein-Barr virus variants characterized by distinct polymorphisms in the EBER locus are strongly associated with nasopharyngeal carcinoma. *International journal of cancer. Journal international du cancer* **2019**, 144, 3031-3042, doi:10.1002/ijc.32049.
12. Valouev, A.; Weng, Z.; Sweeney, R.T.; Varma, S.; Le, Q.T.; Kong, C.; Sidow, A.; West, R.B. Discovery of recurrent structural variants in nasopharyngeal carcinoma. *Genome research* **2014**, 24, 300-309, doi:10.1101/gr.156224.113.
13. Chiu, S.H.; Wu, C.C.; Fang, C.Y.; Yu, S.L.; Hsu, H.Y.; Chow, Y.H.; Chen, J.Y. Epstein-Barr virus BALF3 mediates genomic instability and progressive malignancy in nasopharyngeal carcinoma. *Oncotarget* **2014**, 5, 8583-8601, doi:10.18632/oncotarget.2323.
14. Wu, C.C.; Liu, M.T.; Chang, Y.T.; Fang, C.Y.; Chou, S.P.; Liao, H.W.; Kuo, K.L.; Hsu, S.L.; Chen, Y.R.; Wang, P.W.; et al. Epstein-Barr virus DNase (BGLF5) induces genomic instability in human epithelial cells. *Nucleic acids research* **2010**, 38, 1932-1949, doi:10.1093/nar/gkp1169.
15. Lacoste, S.; Wiechec, E.; Dos Santos Silva, A.G.; Guffei, A.; Williams, G.; Lowbeer, M.; Benedek,

- K.; Henriksson, M.; Klein, G.; Mai, S. Chromosomal rearrangements after ex vivo Epstein-Barr virus (EBV) infection of human B cells. *Oncogene* **2010**, 29, 503-515, doi:10.1038/onc.2009.359.
16. Nikitin, P.A.; Yan, C.M.; Forte, E.; Bocedi, A.; Tourigny, J.P.; White, R.E.; Allday, M.J.; Patel, A.; Dave, S.S.; Kim, W.; et al. An ATM/Chk2-mediated DNA damage-responsive signaling pathway suppresses Epstein-Barr virus transformation of primary human B cells. *Cell host & microbe* **2010**, 8, 510-522, doi:10.1016/j.chom.2010.11.004.
17. Hu, H.; Luo, M.L.; Desmedt, C.; Nabavi, S.; Yadegarynia, S.; Hong, A.; Konstantinopoulos, P.A.; 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, doi:10.1016/j.ebiom.2016.05.025.
18. Fina, F.; Romain, S.; Ouafik, L.; Palmari, J.; Ben Ayed, F.; Benharkat, S.; Bonnier, P.; Spyrtos, F.; Foekens, J.A.; 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, 783-790, doi:10.1054/bjoc.2000.1672.
19. 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, 255-259, doi:10.1016/j.jcv.2014.07.010.
20. Ayee, R.; Ofori, M.E.O.; Wright, E.; Quaye, O. Epstein Barr Virus Associated Lymphomas and Epithelia Cancers in Humans. *Journal of Cancer* **2020**, 11, 1737-1750, doi:10.7150/jca.37282.
21. Lorenzetti, M.A.; De Matteo, E.; Gass, H.; Martinez Vazquez, P.; Lara, J.; Gonzalez, P.; Preciado, M.V.; Chabay, P.A. Characterization of Epstein Barr virus latency pattern in Argentine breast carcinoma. *PloS one* **2010**, 5, e13603, doi:10.1371/journal.pone.0013603.
22. 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, doi:10.1186/1471-2407-14-665.
23. Friedenson, B. Mutations in components of antiviral or microbial defense as a basis for breast cancer. *Functional & integrative genomics* **2013**, 13, 411-424, doi:10.1007/s10142-013-0336-1.
24. Mount, D.W. Using the Basic Local Alignment Search Tool (BLAST). *CSH Protoc* **2007**, 2007, pdb top17, doi:10.1101/pdb.top17.
25. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *Journal of molecular biology* **1990**, 215, 403-410, doi:10.1016/S0022-2836(05)80360-2.
26. Lu, F.; Wikramasinghe, P.; Norseen, J.; Tsai, K.; Wang, P.; Showe, L.; Davuluri, R.V.; Lieberman, P.M. Genome-wide analysis of host-chromosome binding sites for Epstein-Barr Virus Nuclear Antigen 1 (EBNA1). *Virology journal* **2010**, 7, 262, doi:10.1186/1743-422X-7-262.
27. Kim, K.D.; Tanizawa, H.; De Leo, A.; Vladimirova, O.; Kossenkov, A.; Lu, F.; Showe, L.C.; Noma, K.I.; Lieberman, P.M. Epigenetic specifications of host chromosome docking sites for latent Epstein-Barr virus. *Nat Commun* **2020**, 11, 877, doi:10.1038/s41467-019-14152-8.
28. 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, 214-224, doi:10.7150/jca.13150.
29. Needleman, S.B.a.W., 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.

30. Wolfe, D.; Dudek, S.; Ritchie, M.D.; Pendergrass, S.A. Visualizing genomic information across chromosomes with PhenoGram. *BioData Min* **2013**, *6*, 18, doi:10.1186/1756-0381-6-18.
31. 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*, 248-262, doi:10.1016/j.celrep.2016.12.019.
32. Lynn, D.J.; Winsor, G.L.; Chan, C.; Richard, N.; Laird, M.R.; Barsky, A.; Gardy, J.L.; Roche, F.M.; Chan, T.H.; Shah, N.; et al. InnateDB: facilitating systems-level analyses of the mammalian innate immune response. *Molecular systems biology* **2008**, *4*, 218, doi:10.1038/msb.2008.55.
33. 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, doi:10.1186/1745-7580-3-4.
34. Ortutay, C.; Siemala, M.; Vihinen, M. Molecular characterization of the immune system: emergence of proteins, processes, and domains. *Immunogenetics* **2007**, *59*, 333-348, doi:10.1007/s00251-007-0191-0.
35. Denner, J. The porcine virome and xenotransplantation. *Virology journal* **2017**, *14*, 171, doi:10.1186/s12985-017-0836-z.
36. Prusty, B.K.; zur Hausen, H.; Schmidt, R.; Kimmel, R.; de Villiers, E.M. Transcription of HERV-E and HERV-E-related sequences in malignant and non-malignant human haematopoietic cells. *Virology* **2008**, *382*, 37-45, doi:10.1016/j.virol.2008.09.006.
37. Canaan, A.; Haviv, I.; Urban, A.E.; Schulz, V.P.; Hartman, S.; Zhang, Z.; Palejev, D.; Deisseroth, A.B.; Lacy, J.; Snyder, M.; et al. EBNA1 regulates cellular gene expression by binding cellular promoters. *Proceedings of the National Academy of Sciences of the United States of America* **2009**, *106*, 22421-22426, doi:10.1073/pnas.0911676106.
38. Bhela, S.; Mulik, S.; Reddy, P.B.; Richardson, R.L.; Gimenez, F.; Rajasagi, N.K.; Veiga-Parga, T.; Osmand, A.P.; Rouse, B.T. Critical role of microRNA-155 in herpes simplex encephalitis. *Journal of immunology* **2014**, *192*, 2734-2743, doi:10.4049/jimmunol.1302326.
39. Alexandrov, L.B.; Nik-Zainal, S.; Wedge, D.C.; Campbell, P.J.; Stratton, M.R. Deciphering signatures of mutational processes operative in human cancer. *Cell reports* **2013**, *3*, 246-259, doi:10.1016/j.celrep.2012.12.008.
40. Oker, N.K.N.-I., and Busson, P. Review: Biological and Pharmacological Basis for Cytolytic Viral Activation in EBV-Associated Nasopharyngeal Carcinoma. *InTech Open* **2016**, 149-178, doi:<http://dx.doi.org/10.5772/64738>.
41. Dheekollu, J.; Wiedmer, A.; Ayyanathan, K.; Deakyne, J.S.; Messick, T.E.; Lieberman, P.M. Cell-cycle-dependent EBNA1-DNA crosslinking promotes replication termination at oriP and viral episome maintenance. *Cell* **2021**, *184*, 643-654 e613, doi:10.1016/j.cell.2020.12.022.
42. De Leo, A.; Calderon, A.; Lieberman, P.M. Control of Viral Latency by Episome Maintenance Proteins. *Trends in microbiology* **2020**, *28*, 150-162, doi:10.1016/j.tim.2019.09.002.
43. Munz, C. Latency and lytic replication in Epstein-Barr virus-associated oncogenesis. *Nature reviews. Microbiology* **2019**, *17*, 691-700, doi:10.1038/s41579-019-0249-7.
44. 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*, 5542, doi:10.1073/pnas.2002334117.

45. Jiang, L.; Lung, H.L.; Huang, T.; Lan, R.; Zha, S.; Chan, L.S.; Thor, W.; Tsoi, T.H.; Chau, H.F.; 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**, doi:10.1073/pnas.1915372116.
46. Pfeiffer, F.; Grober, C.; Blank, M.; Handler, K.; Beyer, M.; Schultze, J.L.; Mayer, G. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific reports* **2018**, 8, 10950, doi:10.1038/s41598-018-29325-6.
47. Friedenson, B. Many Breast Cancer Mutations Parallel Mutations in Known Viral Cancers. *Journal of Genomes and Exomes* **2014**, 3, 17-35, doi:10.4137/JGE.S18944.
48. Friedenson, B. Mutations in Breast Cancer Exome Sequences Predict Susceptibility to Infections and Converge on the Same Signaling Pathways. *J Genomes and Exomes* <http://mr.crossref.org/iPage?doi=10.4137%2FJGE.S30058> **2015**, 4, 1-28, doi: 10.4137/JGE.S30058.
49. Breuer, K.; Foroushani, A.K.; Laird, M.R.; Chen, C.; Sribnaia, A.; Lo, R.; Winsor, G.L.; Hancock, R.E.; Brinkman, F.S.; Lynn, D.J. InnateDB: systems biology of innate immunity and beyond-recent updates and continuing curation. *Nucleic acids research* **2013**, 41, D1228-1233, doi:10.1093/nar/gks1147.
50. 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, 109-122 e106, doi:10.1016/j.immuni.2019.11.015.
51. Chow, K.T.; Driscoll, C.; Loo, Y.M.; Knoll, M.; Gale, M., Jr. IRF5 regulates unique subset of genes in dendritic cells during West Nile virus infection. *J Leukoc Biol* **2019**, 105, 411-425, doi:10.1002/JLB.MA0318-136RRR.
52. Minskaia, E.; Saraiva, B.C.; Soares, M.M.V.; Azevedo, R.I.; Ribeiro, R.M.; Kumar, S.D.; Vieira, A.I.S.; Lacerda, J.F. Molecular Markers Distinguishing T Cell Subtypes With TSDR Strand-Bias Methylation. *Front Immunol* **2018**, 9, 2540, doi:10.3389/fimmu.2018.02540.
53. Zhu, W.P.; Liu, Z.Y.; Zhao, Y.M.; He, X.G.; Pan, Q.; Zhang, N.; Zhou, J.M.; Wang, L.R.; Wang, M.; Zhan, D.H.; et al. Dihydropyrimidine dehydrogenase predicts survival and response to interferon-alpha in hepatocellular carcinoma. *Cell Death Dis* **2018**, 9, 69, doi:10.1038/s41419-017-0098-0.
54. Grondin, J.A.; Kwon, Y.H.; Far, P.M.; Haq, S.; Khan, W.I. Mucins in Intestinal Mucosal Defense and Inflammation: Learning From Clinical and Experimental Studies. *Frontiers in immunology* **2020**, 11, 2054, doi:10.3389/fimmu.2020.02054.
55. Ibanez-Cabellos, J.S.; Seco-Cervera, M.; Osca-Verdegai, R.; Pallardo, F.V.; Garcia-Gimenez, J.L. Epigenetic Regulation in the Pathogenesis of Sjogren Syndrome and Rheumatoid Arthritis. *Front Genet* **2019**, 10, 1104, doi:10.3389/fgene.2019.01104.
56. Mead, T.J.; Apte, S.S. ADAMTS proteins in human disorders. *Matrix Biol* **2018**, 71-72, 225-239, doi:10.1016/j.matbio.2018.06.002.
57. Reverter, A.; Ballester, M.; Alexandre, P.A.; 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, 52, doi:10.1186/s40168-020-00994-8.
58. Na, T.; Zhang, K.; Yuan, B.Z. The DLC-1 tumor suppressor is involved in regulating immuno-modulation of human mesenchymal stromal /stem cells through interacting with the Notch1 protein. *BMC cancer* **2020**, 20, 1064, doi:10.1186/s12885-020-07542-5.

59. Aljohmani, A.; Yildiz, D. A Disintegrin and Metalloproteinase-Control Elements in Infectious Diseases. *Front Cardiovasc Med* **2020**, *7*, 608281, doi:10.3389/fcvm.2020.608281.
60. Florez, M.A.; Matatall, K.A.; Jeong, Y.; Ortinau, L.; Shafer, P.W.; Lynch, A.M.; Jaksik, R.; Kimmel, M.; Park, D.; King, K.Y. Interferon Gamma Mediates Hematopoietic Stem Cell Activation and Niche Relocalization through BST2. *Cell reports* **2020**, *33*, 108530, doi:10.1016/j.celrep.2020.108530.
61. Amann-Zalcenstein, D.; Tian, L.; Schreuder, J.; Tomei, S.; Lin, D.S.; Fairfax, K.A.; Bolden, J.E.; McKenzie, M.D.; 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*, 1574-1584, doi:10.1038/s41590-020-0799-x.
62. Wen, A.Y.; Sakamoto, K.M.; Miller, L.S. The role of the transcription factor CREB in immune function. *Journal of immunology* **2010**, *185*, 6413-6419, doi:10.4049/jimmunol.1001829.
63. Ivan, D.C.; 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, doi:10.3389/fimmu.2020.609921.