

# **Evidence of lesions from Epstein-Barr virus infection in human breast cancer genomes**

**Bernard Friedenson**

**Department of Biochemistry and Molecular Genetics**

**and Illinois Cancer Center**

**College of Medicine**

**University of Illinois Chicago**

**Chicago, Illinois**

***bernief@uic.edu***

## Abstract

**Epstein-Barr virus (EBV) infects essentially all humans and provides no benefit. EBV can cause nasopharyngeal cancer (NPC), Burkitt's lymphoma (BL), and perhaps breast cancer. Breast tissues from patients with breast cancer are more likely to be EBV-positive than tissues from healthy controls. However, EBV is not a proven cause of breast cancer because the tissues are not consistently EBV-positive. If EBV causes breast cancer, it would have to do it without an active infection. Other cancers with known viral origins do not require continuing presence of the virus. However, the "hit and run" theory is difficult to test for breast cancer without a proven EBV connection.**

**Here, I test this theory with multiple independent bioinformatic analyses. First, hundreds of breast cancer genomes contained characteristic methylation scars that indicate a cleared EBV infection. The genomes had further differential hypermethylation near positions where EBV reprograms normal cells into malignancy. Second, genomes from EBV cancers and breast cancers inactivated the same tumor-suppressive mechanisms. Third, deletions were identified on chromosome 3p in EBV cancers that shift cells to oxidative glycolysis, a prominent breast cancer phenotype known as the Warburg effect. Similar 3p deletions were found in breast cancer genomes. Fourth, somatic hypermutation clusters in EBV-cancers marked genome positions in breast cancers near translocations and focal oncogene amplification. EBV deregulation of deaminase and estrogen-induced topoisomerase explain these translocation breakpoints. Fifth, several alternate explanations for these results were ruled out. Finally, only limited segments of EBV DNA matched the human genome, making it possible that a childhood vaccine would end breast cancer.**

## Introduction

Saliva, blood, breast milk [1], and semen transmit Epstein-Barr virus (EBV). This transmission is so effective that virtually all humans acquire EBV infection at some point in their lives [2]. The infection has no known benefit to the host.

The virus creates reservoirs in the throat and lymph nodes, sometimes targeting the back of the throat, where it can cause nasopharyngeal cancer (NPC) and the lymphoid tissue, where it can cause Burkitt's lymphoma (BL). NPC is the best studied example of an EBV cancer in epithelial tissue. Detection of EBV DNA in plasma samples is a nearly 100% accurate screen for NPC [3]. In lymphoid tissue, EBV transforms B-cells in culture [4] and antibodies against viral capsid antigen predict BL [5]. Both NPC and BL are therefore unequivocal examples of cancers caused by EBV. In both cancers, EBV attaches to human chromosomes as a circular episome and replicates along with host DNA. The EBV episome is exceedingly difficult to maintain in NPC cell culture because malignant cells lose the virus [6]. BL cell lines also lose or inactivate EBV episomes, but they are more stable than in NPC [6, 7]. The spontaneous loss of viral episomes shows that EBV can predispose to malignancy but is not required after transformation occurs.

Even though the virus can disappear, it still leaves evidence of infection behind. EBV-infected oral keratinocytes that expressed telomerase but were not malignant provide an example of this evidence. These cells have hypermethylated CpG islands as epigenetic scars. The methyl groups remain even if repeated host cell divisions cause the loss of the virus. These epigenetic alterations change gene expression, delay differentiation and can lead to an invasive phenotype [8, 9]. The mechanism for these changes included EBV infection upregulating DNA methyltransferases (DNMT3A, DNMT3B, and DNMT1). Non-malignant keratinocytes were more stable than carcinoma cells, but EBV infection reiterated methylation patterns found in NPC and gastric carcinoma [9]. Thus, patterns of hypermethylation provide a marker for past EBV infection, even if the infection disappears.

EBV reprogramming of the host epigenetic system causes massive changes in methylation characteristic of NPC [10, 11]. NPC has hypermethylation in at least 80% of cases [12]. In the remaining 20% of NPC, subtle changes in specificity and activity of methyltransferases DNMT3A vs. DNMT3B may account for hypomethylation [13]. EBV also suppresses the removal of methyl groups by demethylation pathways TET-TDG-BER and APOBEC/AID-MBD4. These changes alter chromatin exposure [12]. Like NPC, BL causes epigenetic reprogramming by silencing at least one critical transcription factor [14]. BL and NPC both have inappropriate activation of the inflammatory NFkB pathway

[15, 16], causing massive immune deregulation. Both these EBV cancers have significant off-target effects of activation induced deaminase (AID) which causes local somatic hypermutation (SHM) [15]. BL includes gene rearrangements which often deregulate the MYC gene. Most viral genes disappear from BL, perhaps by hijacking and perverting functions from host genes [17]. Both NPC and BL have forms in which EBV is undetectable, compatible with loss of virus as the cancer progresses. Thus, both NPC and BL provide data that support one current model for how EBV causes cancer. The model suggests that inactivation of tumor suppressors and immune defenses [16] occurs in premalignant lesions. These deficits then allow latent EBV in reservoirs to trigger critical epigenetic changes throughout the lesions. EBV reprograms the human regulatory epigenome, changes chromatin accessibility, then cellular differentiation and proliferation [12]. The continuing presence of the virus is not required once cells become malignant.

In contrast to its role in NPC and BL, a role for EBV in breast cancer is not broadly accepted [18]. Epidemiologic studies have shown that evidence of current EBV infection is about five-fold more likely in breast cancer tissues than in non-malignant controls [19]. However, only 4 of 30 studies convincingly demonstrated the presence of EBV while 14 studies convincingly demonstrated its absence [20]. The inconsistent detection of EBV in breast cancers leads to suggestions that EBV predisposes mammary epithelial cells to breast cancer, but the virus is not required after the cells become malignant [21]. Breast cancer initiation is more likely in individuals when near-universal EBV infection escapes from control such as by immunocompromise, but because replicating cells can clear EBV, active EBV infection cannot always be detected. Cancer in the absence of its origins describes several other viral or microbial cancers [22, 23]. However, there is currently no way to determine whether this hit and run mechanism occurs in breast cancer.

The aim of the present study was to determine whether oncogenic lesions left by EBV infection tie a lost EBV infection to breast cancer. The experiments were multi-pronged bioinformatic comparisons of breast cancer genomes to EBV infected cells and model EBV cancers. Characteristic epigenetic, metabolic, and chromosome changes that EBV leaves behind in infected host cells converge on the idea that EBV mediated damage is a significant driver for human breast cancer. Five independent lines of evidence support the conclusion that EBV predisposes breast tissue to malignancy but is not required once transformation occurs. Only short EBV-like DNA sequences exist in the human genome and only match a small area of the EBV genome. This limited

homology makes it possible that a childhood vaccine against products expressed by other segments of EBV DNA would effectively end breast cancer as we know it.

## Results

Oral keratinocytes previously infected with EBV and breast cancer cells exhibit similar differential DNA methylation profiles.

The aim of this experiment was to determine whether prior EBV infection may have left remnants of hypermethylation on breast cancer genomes. EBV-infected oral keratinocytes have abnormal methylation that remains even when the infection clears [9]. I compared the distributions of retained hypermethylation in these formerly infected cells to epigenetic marks in breast cancer cells [24] that are likely breast cancer drivers. The epigenetic marks in once-infected keratinocyte chromosomes globally predict which chromosomes are involved in breast cancer cells (Fig. 1A). The frequencies of which chromosomes have these epigenetic drivers correlate according to linear regression analysis ( $r=0.93, r^2=0.87, p<0.0001$ ), rejecting the null hypothesis that there is no relationship between chromosomes in EBV cancers and breast cancers. Two- or one-sided Kolmogorov-Smirnov tests gave high p values (0.89 or 0.5 to 0.44), suggesting any differences in the two distributions are likely due to chance. Similarly, Spearman's rank correlation gave a correlation coefficient of 0.88 (CI=0.72-0.95). Although these results are not a detailed comparison of methylation sites, they were the first hints of a potential relationship between EBV epigenetic reprogramming and breast cancer drivers.

To examine the relationship between breast cancer drivers and EBV infected oral keratinocytes in greater detail, I compared the exact genome coordinates of abnormal methylation in both systems. I found breast cancers have at least eight chromosomes with methylation marks in about the same position as the formerly infected keratinocytes (Fig. 1B). Initially, the screening window was 200,000 bp. Then, the exact distances were calculated more closely for the ten genes methylated in breast cancer that were nearest to sites of abnormal methylation in oral keratinocytes that had lost their EBV infection (Table 1). Functional descriptions of the ten genes link them all to breast cancer development and metastasis (Table 1, last column, and supplementary references to Table 1). This result shows EBV infection has left characteristic marks on breast cancer cells, and these epigenetic scars can predispose to cancer.

## Sites of methylation in the EBV-related cancer NPC cluster around the sites of driver gene methylation in breast cancer

The next step was to determine whether the relationship between EBV infected cells and breast cancer extends to a relationship with EBV-mediated cancers. To test breast cancer for these EBV-type changes, I compared the positions of hypermethylation in EBV-positive NPC [12] to positions of hypermethylated driver mutations [24] in breast cancers. NPC hypermethylation positions agree with the positions of hypermethylation in breast cancer within or near genes that are being regulated for thirteen chromosomes (Fig. 2A). Not only did the positions of many NPC and breast cancer hypermethylated groups match closely, but it was possible that most chromosomes had the same distribution of methyl groups in chromosomes in these two different cancers. The frequency distributions of chromosome positions with abnormal methylated groups on NPC's and breast cancers overlap at many loci on example chromosomes (Fig. 2B). These example chromosomes were selected because Mann-Whitney results ( $P > 0.05$ ) could not rule out identical distributions of differential hypermethylation.

In cases where the distributions do not agree ( $P < 0.05$ , Fig 2C), the distributions of hypermethylation are the same in chromosome areas where matches occur ( $P$  values approach 1.0 in Fig. 2C). Both cancer sites involve a significant upregulation of DNMT3A, which catalyzes changes in the epigenome. To verify that the 10,000 base pairs window was not too large, breast cancer methylation sites were run against a random distribution with the same mean and standard deviation as the NPC data. Instead of nearly 100 positions that agreed, trials against two different random distributions gave values of 1 and 11 (Fig 2A). This result means that the agreement between NPC and breast cancer methylation positions is unlikely to occur purely by chance. These results show that massive methylation changes caused by EBV in NPC also occur in breast cancer.

To determine if this agreement is more general, I tested BL, another cancer driven by EBV infection. Differential methylation positions near breast cancer drivers were the largest peak on all chromosomes. Again Mann-Whitney testing could not rule out the same hypermethylation distribution in BL and in breast cancer ( $P > 0.05$ , Fig. 3). This result shows that many positions of differential methylation in a second EBV cancer are similar to those in breast cancers. These results support the possibility that EBV infection has driven the epigenetic changes in breast cancer.

Genes that control stem cell properties and antiviral immunity are key targets of hypermethylation that breast cancers share with EBV-driven cancers.

To test whether the methylation changes had a plausible role in causing breast and EBV cancers, the functions of hypermethylated genes were researched. Hypermethylated genes in NPC were found to cluster around homeobox (HOX) on chromosome 17, which are also hypermethylated breast cancer driver genes (Fig. 4). These genes affect oncogenesis because they regulate stem cell differentiation and self-renewal [25]. Data from chromosome 18 (Fig. 4, bottom) show additional effects on chromatin, invasion, and hint at estrogen interaction (Table 2). These results provide examples of mechanisms that explain how differential methylation caused by EBV infection can predispose to breast cancer.

If EBV has a role in driving breast cancer, then one would expect that antiviral defense mechanisms would also be disabled. On chromosome 11, breast cancers and NPCs both had methylation clustered around genes that respond to interferons (65 to 161-fold enrichment,  $p=1.4e-4$  to  $9e-4$ ). Methylation affects induced transmembrane proteins (IFITM1, IFITM-2, and IFITM-3), which cooperate in interferon-stimulated cells. IFITM3 fuses with incoming virus particles to send them to lysosomes [26]. Deficiencies in the breast cancer antiviral immune responses were related to defects caused by EBV infection in NPC (Table 2). This result shows that interference with antiviral defenses in breast cancer resembles the interference in an EBV mediated cancer.

Chromosome 3p deletions in EBV cancers can create BRCA1/2-like deficits and cause the Warburg effect.

I noticed that almost all NPC genomes (80% [27], 81-100% [28], 96% [29]) have deletions in at least one locus on chromosome 3p. These 3p deletions strongly associate with EBV: 100% of NPC with chromosome 3p deletions from a high-risk region of China were positive for EBV infection [28]. Precancerous lesions also lose heterozygosity (LOH) on chromosome 3p, implying chromosome 3p deletions are early events [28, 29]. To determine the effect of these deletions, known protein coding transcripts from reference genome sequence data were tabulated. Transcripts essential for HR and anti-EBV immune responses cluster together (Fig. 5A). During normal high-fidelity DNA repair, DNA damage directs FANCD2 to nuclear foci where BRCA1 and BRCA2 proteins operate [30]. NPC deletions reported [31] remove FANCD2 from chromosome 3p, effectively disabling HR and allowing chromosome breaks to persist (supplementary Fig. 1). Loss of the nearby VHL tumor suppressor to the deletion further

impairs HR repair [32]. These considerations show that EBV cancer loses protective mechanisms from chromosome 3p. These losses approximate inheriting a defective BRCA1 or BRCA2 gene because these genes depend on transcripts encoded on chromosome 3p (supplementary Fig.1).

*Chromosome 3p deletions in breast cancers are similar to those that occur in NPC.* As in NPC, deletions affecting chromosome 3p are found in up to 87% of breast cancer cases, and nearly 20% of breast ductal tumors completely lose FANCD2 expression due to deletions on chromosome 3p [33-44]. The frequency and extent of 3p losses increase with the severity of pathological changes [45]. The fragile site FRA3B including FHIT is nearby [46] and is another source of frequent breakages on chromosome 3 (Fig 5A), with a potential connection to EBV infection [47].

Chromosome 3p includes the gene encoding the catalytic portion of pyruvate dehydrogenase (PDHB at chr3:58,427,630-58,433,832 ). EBV tumors without PDHB cannot metabolize pyruvate by oxidative phosphorylation [48, 49], so these tumors cannot use available oxygen to produce energy most efficiently from glucose (the Warburg effect). TCGA data finds deletions, including PDHB occur in about 30% of breast cancers when comparing results from multiple datasets (Fig 5B).

To test for additional changes in chromosome 3p in breast cancer that might be responsible for the Warburg effect, I plotted frequencies of known short deletions on chromosome 3p in COSMIC data from 476 breast cancers (Fig. 5C). These deletions are typically caused by AID [50], which EBV induces prior to lymphomas [51, 52]. The deletions (Fig 5C) mainly target transcript variants of CTNNB1, which then deregulate WNT/CTNNB1/TCFL2/PDHB signaling, causing the Warburg effect. EBV-positive gastric cancer targets the same pathway [8]. Furthermore, breast cancers also have hypermethylation of PDHB. Because glucose metabolism is complex and tightly controlled, other genetic abnormalities probably contribute to the Warburg effect. A breast cancer breakpoint approximately the same as an NPC breakpoint is only 6743 base pairs from the PDHX gene, a different subunit of the pyruvate dehydrogenase complex. PDHX is within a fragile site on chromosome 11 [46].

The microenvironment of early breast cancer strongly selects for the Warburg phenotype [53], and almost all breast cancers have it. The Warburg phenotype associates with progression in triple-negative breast cancer [49]. These results show that EBV infection can contribute to both defective DNA repair and metabolic reprogramming in breast cancer.



## Breakpoints in EBV-associated NPC and breast cancer occur at similar positions

The HR deficits observed in NPC are like those in hereditary breast cancer, so it is possible that breakpoints are also similar, especially those associated with translocations and oncogene amplification. To test this possibility, every breast cancer translocation breakpoint in a dataset from 780 breast cancers [54] was compared to all breakpoints on corresponding chromosomes from 70 NPC's [16]. The results revealed that both breast cancer and NPC had many similar breakpoints on all chromosomes (Fig. 6A), but chromosome 11 had the largest number of similar breakpoint positions. On some chromosomes, Mann-Whitney testing gave P values  $>0.05$  that could not rule out identical overall breakpoint distributions in NPC and breast cancer (Fig. 6A).

To verify this result, breakpoint positions on chromosome 11 in NPC and breast cancer were compared further (Fig 6B). Breast cancers and NPC's shared a large cluster of breakpoints telomeric to the centromere between 55 - 80 million bps. This region includes the CCND1 gene, which is known to be amplified in NPC [16] and is a site for breast cancer translocations and oncogene amplifications [54]. Further comparisons revealed the frequencies of breakpoint positions in breast cancers and NPC were similar on chromosomes 8, 11 and 17 (Fig. 6B).

The numbers of breakpoints in breast cancer that agreed with NPC depended on the window size, but smaller windows did not change the conclusions. For chromosome 11, the use of a window of 2.5k, 1k, 0.5k, or 0.25k instead of either 5k or 10k produced fewer breakpoints that approximately coincided, but did not change the result that some breast cancer breakpoint were near NPC breakpoints (Fig. 6C). Normality tests indicated that breakpoint positions in breast cancer, NPC, or their frequency distributions on chromosome 11 were unlikely to represent normal distributions (Fig 6D). Based on the above results, breakpoints in an EBV mediated cancer occur near some of the same chromosome translocations that drive breast cancer.

## Breakpoints in the EBV cancer BL are also similar to breast cancer breakpoints.

To determine whether the resemblance between breakpoints in EBV cancer and breast cancer depends on whether the tissue is epithelial or lymphoid, I compared published copy number and structural variants (CNV and SV) in chromosomes from 94 EBV-positive children with BL [15] to translocation breakpoints in 780 breast cancers [54]. Many breast cancers have chromosome breaks closely related to breakpoints in EBV-positive BL (Fig. 7A). Polynomial regression analysis found that breakpoint

numbers on BL chromosomes predicted breakpoint numbers in breast cancer, albeit according to a polynomial curve (Fig 7B). Specifically, the coefficient of determination (r-squared) value suggested that the breakpoint numbers in EBV-positive BL could account for approximately 48.6% of the variability in breast cancer breakpoint numbers. The polynomial curve suggests that the distance between BL and breast cancer breakpoints somehow affects the total numbers of breakpoints. In summary, the relationship between breast cancer breakpoints in BL, an EBV cancer in lymphoid tissue, is consistent with EBV influencing or causing some of the chromosome breakpoints.

### Breast cancer translocations and focal amplifications are related to somatic hypermutations in EBV cancers

Identifying a plausible mechanism that EBV could use to drive breast cancer began by reviewing the relationship between EBV and AID. EBV-positive BL has increased AID activity in vivo [15] and NPC also has AID footprints [12]. AID converts cytosine to uracil that mimics dT during DNA replication. Uracil causes mutations if it pairs with G because the U:G mismatch triggers error-prone DNA repair. Then uracil DNA glycosylase and/or DNA mismatch repair enzymes produce local mutations and double-strand breaks as they remove DNA mismatches [55]. These events cause the DNA double-strand breaks that produce c-myc/IgH translocations characteristic of BL. If EBV is involved in breast cancer, then breakpoints at breast cancer translocations and focal amplifications could also correspond to positions where AID deregulation causes SHM. The MYC locus is a prototypical example of translocations and amplifications accompanying AID deregulation.

The first test of this hypothesis was to determine whether EBV cancers and breast cancers have similar breakpoints on chromosome 8 around the MYC gene locus. The results revealed that the distance between NPC and BL breakpoints plotted against the number of breakpoints was approximately linear up to about 11,000 bp (Fig 8A). The first 11,000 base pair interval included about 75 breakpoints (Fig. 8B). This result implies that EBV induces a set of related breakpoints in NPC and BL. Consistent with this conclusion, various breast cancers had over 400 breakpoints near the MYC locus (Fig. 8C). These results show that MYC locus breakpoints characteristic of EBV cancers are near breakpoints in breast cancer.

EBV is capable of inducing AID [56-58]. If EBV mediates AID deregulation in breast cancer, then SHM in EBV related cancers might also be near translocation and focal amplification sites. To test this hypothesis, the positions of SHM in BL on chromosome 8 were compared to breast cancer breakpoints. The results confirmed that

the majority of EBV associated SHMs in BL on chromosome 8 (685/1221) were near a breast cancer breakpoint (Fig 8D). In translocations characteristic of BL, chromosome 14 and chromosome 2 are significant partners with MYC, so BL somatic mutations were then compared to breast cancer breakpoints on chromosomes 14 and 2. Breast cancer breakpoints were again related to positions of SHM (Fig 8E). Chromosome 17 had the largest number of breakpoints related to BL (Fig 7A) and was used to further test for a relationship between EBV-associated SHM in BL and breast cancer breakpoints. Nearly 140 positions of somatic hypermutation in BL were within 5k bps of a breast cancer breakpoint (Fig. 8E) and the relationship fit a complex polynomial (Fig 8F). These results establish that SHM in EBV cancers, induced by AID deregulation, occurs near breakpoints in breast cancer and includes the characteristic MYC locus.

Estrogen has been proposed as directly generating initial translocations in estrogen receptor targets in human breast cancer [54]. However, BL in mostly male pediatric patients has breakpoints [59] like those in breast cancer so BL data conflicts with the role of estrogen in initiating translocations. Instead, transcription at elements that respond to estrogen-receptor complexes requires topoisomerase. Topoisomerase causes transient DNA breaks that are normally repaired by HR-proteins [60]. In breast cancer, topoisomerase is a strong predictor for amplification at translocation boundaries [54]. EBV-infected cells also lose normal control of aromatase, an enzyme which converts estrogen precursor into estrogen. Transcription in response to high estrogen levels induces topoisomerase mediated DNA breaks but herpesviruses can also hijack topoisomerase [61]. The expression of ZEBRA (the EBV lytic switch in NPC) increased levels of aromatase and ER- $\alpha$  [62]. According to this hypothesis, it is EBV mediated deregulation of AID, estrogen production from EBV control of aromatase, and topoisomerase that drives breast cancer translocations. If so, then the distribution of breakpoints and focal abnormalities in breast cancers should be relatively independent of ER status.

To test these explanations, I compared 10736 breakpoints in 271 ER-negative breast cancer cases to 12242 breakpoints in 473 ER-positive breast cancer cases. The results revealed hundreds of breakpoints near each other on some chromosomes (Fig. 9A). On chromosome 17, 29 breakpoints in ER-negative breast cancers were within ten bp of a breakpoint in ER-positive breast cancer. In contrast, a random breakpoint control had only 0-3 breakpoints within 10 bp of an ER positive breakpoint. Chromosome 8 had over 70 breakpoints within 1500 bp of one another (Fig. 9A, right). Chromosome 11 had about 130 breakpoints within 5k bp of one another (Fig. 9B) and their frequency

distributions had a region of local overlap (Fig 9C). The proximal breakpoints focused on chromosomes 17, 8, and 11, but all chromosomes had them (Fig 9A). The distributions of breakpoints in ER-positive vs. ER-negative breast cancer could not be confidently distinguished ( $P>0.05$ ).

The effect of estrogen on translocations and focal oncogene amplifications in breast cancer [54] is undeniable, but the above evidence shows that ER status does not determine the location of chromosome breaks. The results instead show EBV fits breast cancer and BL data better than differences in ER status because of AID deregulation, inappropriate estrogen production from aromatase with herpes virus mediated deregulation of topoisomerase.

Many chromosome breakpoints occur at approximately the same positions in HR-deficient and HR-proficient breast cancers.

If EBV creates deficits in HR, then breakpoints in HR-deficient (HRD) BRCA1/2-type breast cancers should resemble those in HR-proficient (HRP) breast cancers. To test this hypothesis, I compared the breakpoint positions in HRD vs. HRP breast cancers. The exact coordinates of all breakpoints in HRP vs HRD breast cancers were then compared to measure this apparent similarity more precisely (Fig. 10). All chromosomes with the possible exception of chromosome 22 showed some breakpoints on HRP breast cancers that were near (within 5000 bp) those in HRD breast cancers. On chromosome 17, >180 breakpoints were within 5000 bp of one another and 50 breakpoints were within 50 bp. For chromosomes 3, 6, 19, 22, and X, the Mann-Whitney test could not exclude the possibility that the entire distribution of breakpoints was the same (Table 3).

Next, the idea was tested that there are local areas of similarity even when the overall distributions of breakpoints differs. Local breakpoint distributions in HRD breast cancers matched local breakpoint distributions in HRP breast cancers in at least some chromosome sections, even if the overall distributions differed (Table 3 and Fig. 10). The breakpoints in HRD breast cancer nearest those in HPR breast cancers probably follow the same distribution. Based on the above statistical comparisons, many breakpoint positions in HRD and HRP breast cancers are similar. These results are consistent with the idea that EBV inhibits HR to create a deficit in HRP breast cancers equivalent to hereditary pathogenic BRCA1 or BRCA2 mutations (HRD breast cancers).

Breakpoints in EBV-positive BL genomes approximate the ability of BRCA1 or BRCA2 mutation to block a cancer suppressive mechanism.

To test the idea that EBV can mimic hereditary BRCA gene mutations, I assessed breakpoint positions in EBV-positive BL in the context of well-known genes required for

HR [63] or identified as HR genes from KEGG pathway analysis. BL breakpoints were found to affect HR-related genes mainly on chromosomes 1, 11, and 17 (Fig. 11). The most frequently targeted gene was H2AX, which is located on chromosome 11. On chromosome 13, BRCA2 was affected in 3 BL samples, and other breakpoints clustered near BRCA1 on chromosome 17. Although the data are limited, this finding related to HR damage echoes the results in NPC, showing that EBV-related cancers cause damage to genes associated with hereditary breast cancer risk.

### Alternate Explanations.

The binding of the viral tethering protein EBNA1 binding to a DNA consensus sequence does not initiate chromosome breakage in breast cancers.

An alternative explanation for EBV-related carcinogenesis involves the docking of EBNA1 virus tethering protein at specific EBV binding sequences clustered on chromosome 11 [64]. The host binding sequences consist of repeating imperfect 18-base pair palindromes. In this mechanism, the bound EBNA1 then crosslinks EBV DNA to human chromosome 11 [64]. To test this mechanism, existing literature data was first compared to the specific human EBNA1 binding site [65-67]. The results (Table 4) are not compatible with a single host sequence that binds EBNA1.

I next measured the probability of finding the imperfect 18 bp palindrome among the 3.2 billion bases of the human genome purely by chance (BLAST analysis "E" (Expect) value). BLAST analysis produced 4,352 matches, from 12 to 18 base pairs in length with E values between 16 and 964 (Supplementary Table S1). Chromosome 11 contained only 197 of these 4352 matches, and none were within or near the palindromic region. The prototype DNA palindrome (Table 1, line 2) produced 7074 matches with E values ranging from 0.25 to 964. Further BLAST analyses of the docking sequence in the EBNA1-DNA complex which was crystallized (Table 4, line 1) against other genome assemblies [68] revealed matches on chromosomes 2, 19, 4, and 12. Various isolates of human immunodeficiency viruses had 52 matching sequences.

Subsequent experiments tested whether breakpoints at the chromosome 11 palindromic locus occur consistently in EBV-positive lymphoid cancers. I searched for breakpoints in CNVs or SVs in EBV-positive cancers near the palindromic locus (within 10,000 base pairs). The first sequences were from pediatric BL cases. All except one breakpoint were millions of base pairs from the palindromic locus (Fig. 12). In a subgroup of 94 EBV-positive BL cases, the palindromic locus was nearly 100 million bp away from the actual major breakpoint coordinates (Fig. 12B). Most breakpoints

occurred in a narrow region of approximately 36,000 bp (Chr11: 18,915,804-18,951,521). At least 130,000 base pairs still separated the six breakpoints closest to the palindromic hotspot (Fig 12B). In 94 BL samples from EBV-positive patients, breakpoints were concentrated within chromosomes 2, 8, 13, 14, and 22 (Fig. 12C).

Chromosome 14 contained 610 breakpoints (IgVH regions), and chromosome 2 (IgVK regions) contained 522 breakpoints. EBV hijacks activation induced cytidine deaminase (AICDA) which typically generates antibody gene variants that respond to myriad antigens.

Only 19 of the 94 EBV-positive patients [15] had breakpoints anywhere on chromosome 11 (Fig 12C). In BL cells, EBV docks on chromosome 11 near the FAM-D and FAM-B genes [66], but other EBV docking sites on chromosome 11 are over 100 million and 21 million base pairs from the palindrome locus, near the LUZR2 and FAT3 genes, respectively. Therefore, chromosome 11 fusions in the palindromic regions probably do not initiate structural variation in most cases of EBV-positive BL. To test the involvement of the palindromic locus in other cancers, CNVs were compared for diverse cancers from 8227 patients [69]. Neither amplifications nor deletions included the palindromic sequence (Fig 12C). The chromosome breakpoints closest to the palindrome locus in 70 NPC cases [16] occur at positions 112,803,697 and 115,316,621. These breaks are at least 691,000 bp from the palindromic locus. Although the proximity of breast cancer breakpoints to those in NPC on chromosome 11 was clear, the coinciding breakpoints were not near the palindromic locus (Fig. 6). All of these results are incompatible with the notion that EBV docking and EBV cancers originate from the crosslinking of EBNA1 and imperfect palindromic repeats on human chromosome 11.

Tumor-infiltrating lymphocytes do not cause the breaks observed in breast cancer but may influence where they occur.

Comparisons of a limited number of breast cancer cases characterized by a large number of tumor-infiltrating lymphocytes (TILs) or no TILs did not find a significant difference in how breakpoints were distributed on the chromosomes ( $P=0.9073$ ) (Fig 13).

The abundant EBV-like sequences in human chromosomes represent a small section of EBV DNA.

To study the feasibility of a future vaccine against EBV, human chromosomes 3, 8, 11, and 17 DNA were arbitrarily selected for comparison to all known and available viral sequences (Fig. 14). Many positions along the entire lengths of the four

**chromosomes resemble EBV and retroviruses, but the distributions of virus-like sequences differ. Chromosome 8 had the strongest resemblance to EBV with 6014 sequences that matched EBV (homology scores over 500, indicating 355-500 identical bases). Because EBV-like sequences are abundant on human chromosomes, EBV can mimic human sequences. The following five EBV-like human sequences have oncogenic properties in NPC [70]:**

- LMP2A/LMP2B regulate the host aryl hydrocarbon pathway [71] and maintain latent infection.
- A73 correlates with NPC occurrence [72]
- BALF4 is a virion envelope glycoprotein in spikes that attaches to host cell surface proteoglycans. The viral and cellular membranes then fuse, enabling virus entry. BALF4 dramatically enhances the ability to infect human cells [73]
- BALF3 is an endonuclease that mediates mature virion production and packaging during the EBV lytic cycle[74]. The BALF3 protein causes double-strand breaks and micronuclei in NPC [75].
- LMP-1 reprograms infected cells, inducing proliferation, inflammation, and preventing differentiation [76]

**This result shows that thousands of full or partial copies of EBV-like sequences relate to instability of the human genome (Fig. 14A -14C) and emphasizes how dangerous EBV infection can become. Fortunately, only a few areas of EBV genomes were found in human DNA and these sequences were confined to about 2500 bases out of the approximately 173,000 bases in EBV genomes.**

## Discussion

**This study identified at least five separate lines of evidence that implicate EBV infection in breast cancer (Fig.15). First, genome scars left by EBV on breast cancer genomes are characteristic evidence of prior EBV infection, regardless of whether the infection is still active. Breast cancer methylation was observed near the same positions of differential methylation in keratinocytes that had cleared previous EBV infections. This persistent methylation does not occur in never-infected cells. Functional analyses of the genes near these continuing positions suggest that their inhibition could predispose the cells to becoming malignant. Sites of aberrant methylation in NPC and BL, both known EBV-related cancers, occur near sites that are also aberrantly methylated in breast cancers. Stem cell properties, immune responses, and high-fidelity DNA repair are foci of EBV-related methylation that are shared among cancers.**

**In NPC, near-universal deletions on chromosome 3p remove FANCD2 genes. FANCD2 protein is essential for the function of the breast cancer genes BRCA1 and**

**BRCA2. In NPC, EBV causes deletions, mutations and rearrangements in the host that affect the same HR pathway as hereditary deficiencies in BRCA1 and BRCA2 genes[77-79]. EBV infection in the breast then approximates inheriting a defect in BRCA1 or BRCA2, causing an increased risk for breast cancer. The virus leaves this evidence behind as broken chromosomes, gene rearrangements, deletions, and insertions. The genome positions where these events take place in breast cancer often resemble those in EBV cancers. This result is consistent with the subsequent finding that breakpoints in HRD vs. HRP breast cancers match.**

**The 3p deletion in EBV cancer also disables normal glucose metabolism via oxidative phosphorylation, causing the Warburg effect. The 3p deletion omits code for PDHB, a key enzyme needed for normal glucose metabolism. Absence of PDHB makes oxidative phosphorylation impossible, so glucose metabolism mimics anerobic conditions in the presence of plentiful oxygen.**

**Alternate explanations for chromosome breaks and breast cancer invoke initiation by viral tethering at an expandable sequence on chromosome 11. The effect of estrogen is consistent with viral activation of AID because most AID-induced mutations on chromosome 8 in BL are near breast cancer breakpoints. EBV can also deregulate estrogen production, facilitating topoisomerase induced breaks. Finally, abundant EBV-like sequence fragments on at least four chromosomes matched an oncogenic protein from EBV and support the idea that EBV and human chromosomes have extensive interactions.**

**A strength of this study is its use of alterations to the genome as markers of the presence of a past virus infection. Analysis of data from EBV-infected keratinocytes suggested that EBV could be cleared after multiple generations of breast cancer cell division. The methods and tests developed here should be applied to test other cancers for roles of latent infections. Other cancers also show the Warburg effect and have 3p deletions. For example, small cell lung cancers universally delete part of chromosome 3p [80]. Deletion of chromosome 3p is common in squamous cell cancers [81]. Viral involvement in these cancers should also be tested. A limitation of the current study is the relative rarity of NPC and BL cancers.**

**The current findings explain inconsistencies in epidemiological studies that evaluated EBV infection in breast cancers. The results of the current study show how EBV can nonetheless contribute to breast cancer even if the infection is ultimately cleared. This contribution becomes independent of the presence of virus because EBV infection leaves marks on the human genome and causes chromosomal aberrations.**



The fact that 100% of NPC cases were positive for EBV infection in a selected area of China suggests differences in the role of EBV infection in NPC and breast cancer, because 100% of breast cancer cells do not test positive for EBV infection. These findings may represent geographic differences in NPC prevalence. Geographic differences are also characteristic of breast cancer. Over 2.3 million new cases and 685,000 deaths from breast cancer occurred in 2020. Typical of an infectious disease, the 2.3 million new breast cancer cases reported in 2020 show large variations in different countries and world regions (<40/100,000 females in parts of Asia and Africa to >80 per 100,000 in North America, some areas of Europe, and Australia/New Zealand)[82]. Areas of Asia (prone to NPC) can have over a fifteen fold differences in incidence [83]. However, geographic variation is complex and subject to additional variables such as population density, genetic backgrounds, and environmental factors.

The results presented here predict that vaccination against near-universal EBV infection could markedly reduce the incidence of breast cancer. Breast cancer treatment regimen design may also consider remnants of past EBV infection. However, the significant homology between the human and EBV genomes complicates the development of a vaccine or treatment protocol. The finding that only a few of all possible EBV genetic regions exist in the human genome suggests products of these regions should be excluded from vaccines or treatment protocols. Future research should extend these experiments to the remaining chromosomes.

## Methods

***BL breakpoint data.*** Breakpoints in endemic BL were obtained from published data from endemic BL (eBL) and sporadic BL (sBL) [15]. Both ends of every CNV and every fusion were counted as breakpoints. Data for comparisons often came from two different genome versions. In these cases, data were converted to GRCh38 coordinates using UCSC LiftOver methods. The length of whole chromosomes was from CHM13 coordinates. Data for 780 breast cancers were obtained from genome sequences compiled from five published studies [54]

***Calculation of distances between breakpoints in different cancers.*** The break position in a breast cancer genome located closest to a break in an EBV cancer genome (NPC or BL) was taken as the XLOOKUP value for the number of base pairs from the closest NPC breakpoint 5' or 3' to the BL break or the NPC breakpoint (Supplementary Table S2). The distance from the breast cancer breakpoint was then the absolute value of the difference between the closest EBV cancer break and the tested break in breast cancer.

Differences in the amount of data available for NPC, BL, and breast cancers complicated the calculations, especially near telomeres. Several methods of handling these end regions were tried and made no meaningful difference in the outcomes.

Several different window sizes were tried for making comparisons of positions in breast cancer vs. EBV-related cancers. A 10,000 base pair window with an overflow window of 10,000,000 was used for initial screening, and often proved to be adequate. Later, an optimum window size used was calculated from a least squares plot of distances of base pairs vs numbers of breakpoints. The window size was taken as the last point that gave a consistent slope. These results gave a value of approximately 5000 for both chromosome 1 and chromosome 11, so 5000 was used as the default window size. For these two chromosomes, the ratio of chromosome breaks in breast cancers nearest NPC to the ratio of chromosome breaks nearest randomly generated values were maximum near a window size of 5,000. In contrast, translocation breakpoints on chromosome 8 comparing MYC and BL were linear up to a value of 10,000 bp and then leveled off. Alternative methods of calculation using array formulas always gave the same results as the XLOOKUP methods. The minimum distances separating breast cancer breaks from EBV cancer breaks in these arrays were calculated as the absolute value of breast cancer break position minus the array of EBV cancer break positions. The millions of calculations required were repeated at least twice and often many times (supplementary Table 2)

***Comparisons of distributions.*** Tests for normality routinely showed that the data was unlikely to be normally distributed. Therefore, Mann-Whitney nonparametric tests were regularly used to compare the distributions of breakpoints and methylated promoters. Correlation between epigenetic modifications on breast vs EBV cancers was by linear regression analyses with a power for 5% significance of 99.99%. Kolmogorov-Smirnov and Spearman's tests were also used. Frequencies were plotted to verify the distributions of coinciding areas.

***DNA sequence homology.*** The NCBI BLASTn program (MegaBLAST) and database [84-86] were used to compare the consensus DNA palindrome sequence to human DNA sequences and to determine the homology of human DNA to viral DNA. BLAST corrects for short input sequences by using a smaller word size and a different scoring algorithm. E(expect) values are related to p values and represent the probability that a given homology bit score occurs by chance. A strict cutoff was used because the 3.2 billion bases in the human genome increase the odds that identical sequences can occur by chance. Only E values <1e-10 were considered to indicate significant homology. In cases with significant homology, the E values were represented as "0" (<1e-180) and were always far below 1e-10.

## References

1. Daud, II, Coleman CB, Smith NA, Ogolla S, Simbiri K, Bukusi EA, Ng'ang'a ZW, Sumba PO, Vulule J, Ploutz-Snyder R *et al*: **Breast Milk as a Potential Source of Epstein-Barr Virus Transmission Among Infants Living in a Malaria-Endemic Region of Kenya**. *The Journal of infectious diseases* 2015, **212**(11):1735-1742.
2. Kerr JR: **Epstein-Barr virus (EBV) reactivation and therapeutic inhibitors**. *J Clin Pathol* 2019, **72**(10):651-658.
3. **Analysis of Plasma Epstein-Barr Virus DNA to Screen for Nasopharyngeal Cancer**. *The New England journal of medicine* 2018, **378**(10):973.
4. Munz C: **Latency and lytic replication in Epstein-Barr virus-associated oncogenesis**. *Nature reviews Microbiology* 2019, **17**(11):691-700.
5. Coghill AE, Proietti C, Liu Z, Krause L, Bethony J, Prokunina-Olsson L, Obajemu A, Nkrumah F, Biggar RJ, Bhatia K *et al*: **The Association between the Comprehensive Epstein-Barr Virus Serologic Profile and Endemic Burkitt Lymphoma**. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2020, **29**(1):57-62.
6. Dittmer DP, Hilscher CJ, Gulley ML, Yang EV, Chen M, Glaser R: **Multiple pathways for Epstein-Barr virus episome loss from nasopharyngeal carcinoma**. *International journal of cancer Journal international du cancer* 2008, **123**(9):2105-2112.
7. Kolman JL, Kolman CJ, Miller G: **Marked variation in the size of genomic plasmids among members of a family of related Epstein-Barr viruses**. *Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**(16):7772-7776.
8. Birdwell CE, Prasai K, Dykes S, Jia Y, Munroe TGC, Bienkowska-Haba M, Scott RS: **Epstein-Barr virus stably confers an invasive phenotype to epithelial cells through reprogramming of the WNT pathway**. *Oncotarget* 2018, **9**(12):10417-10435.
9. Birdwell CE, Queen KJ, Kilgore PC, Rollyson P, Trutschl M, Cvek U, Scott RS: **Genome-wide DNA methylation as an epigenetic consequence of Epstein-Barr virus infection of immortalized keratinocytes**. *Journal of virology* 2014, **88**(19):11442-11458.
10. Zhang L, Wang R, Xie Z: **Correction to: The roles of DNA methylation on the promotor of the Epstein-Barr virus (EBV) gene and the genome in patients with EBV-associated diseases**. *Appl Microbiol Biotechnol* 2022, **106**(21):7373.
11. Zhang L, Wang R, Xie Z: **The roles of DNA methylation on the promotor of the Epstein-Barr virus (EBV) gene and the genome in patients with EBV-associated diseases**. *Appl Microbiol Biotechnol* 2022, **106**(12):4413-4426.
12. Ka-Yue Chow L, Lai-Shun Chung D, Tao L, Chan KF, Tung SY, Cheong Ngan RK, Ng WT, Wing-Mui Lee A, Yau CC, Lai-Wan Kwong D *et al*: **Epigenomic landscape study reveals molecular subtypes and EBV-associated regulatory epigenome reprogramming in nasopharyngeal carcinoma**. *EBioMedicine* 2022, **86**:104357.
13. Gao L, Emperle M, Guo Y, Grimm SA, Ren W, Adam S, Uryu H, Zhang ZM, Chen D, Yin J *et al*: **Comprehensive structure-function characterization of DNMT3B and DNMT3A reveals distinctive de novo DNA methylation mechanisms**. *Nat Commun* 2020, **11**(1):3355.
14. Manara F, Jay A, Odongo GA, Mure F, Maroui MA, Diederichs A, Sirand C, Cuenin C, Granai M, Mundo L *et al*: **Epigenetic Alteration of the Cancer-Related Gene TGFBI in B Cells Infected with Epstein-Barr Virus and Exposed to Aflatoxin B1: Potential Role in Burkitt Lymphoma Development**. *Cancers* 2022, **14**(5).

15. Grande BM, Gerhard DS, Jiang A, Griner NB, Abramson JS, Alexander TB, Allen H, Ayers LW, Bethony JM, Bhatia K *et al*: **Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma.** *Blood* 2019, **133**(12):1313-1324.
16. Bruce JP, To KF, Lui VWY, Chung GTY, Chan YY, Tsang CM, Yip KY, Ma BBY, Woo JKS, Hui EP *et al*: **Whole-genome profiling of nasopharyngeal carcinoma reveals viral-host co-operation in inflammatory NF-kappaB activation and immune escape.** *Nat Commun* 2021, **12**(1):4193.
17. Hutcheson RL, Chakravorty A, Sugden B: **Burkitt Lymphomas Evolve to Escape Dependencies on Epstein-Barr Virus.** *Front Cell Infect Microbiol* 2020, **10**:606412.
18. Farahmand M, Monavari SH, Shoja Z, Ghaffari H, Tavakoli M, Tavakoli A: **Epstein-Barr virus and risk of breast cancer: a systematic review and meta-analysis.** *Future Oncol* 2019, **15**(24):2873-2885.
19. Lawson JS, Glenn WK: **Catching viral breast cancer.** *Infectious agents and cancer* 2021, **16**(1):37.
20. Joshi D, Buehring GC: **Are viruses associated with human breast cancer? Scrutinizing the molecular evidence.** *Breast cancer research and treatment* 2012, **135**(1):1-15.
21. 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.
22. Shen Y, Zhu H, Shenk T: **Human cytomegalovirus IE1 and IE2 proteins are mutagenic and mediate "hit-and-run" oncogenic transformation in cooperation with the adenovirus E1A proteins.** *Proceedings of the National Academy of Sciences of the United States of America* 1997, **94**(7):3341-3345.
23. Hatakeyama M: **Helicobacter pylori CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis.** *Cell host & microbe* 2014, **15**(3):306-316.
24. Batra RN, Lifshitz A, Vidakovic AT, Chin SF, Sati-Batra A, Sammut SJ, Provenzano E, Ali HR, Dariush A, Bruna A *et al*: **DNA methylation landscapes of 1538 breast cancers reveal a replication-linked clock, epigenomic instability and cis-regulation.** *Nat Commun* 2021, **12**(1):5406.
25. Bhatlekar S, Fields JZ, Boman BM: **Role of HOX Genes in Stem Cell Differentiation and Cancer.** *Stem Cells Int* 2018, **2018**:3569493.
26. Spence JS, He R, Hoffmann HH, Das T, Thinon E, Rice CM, Peng T, Chandran K, Hang HC: **IFITM3 directly engages and shuttles incoming virus particles to lysosomes.** *Nature chemical biology* 2019, **15**(3):259-268.
27. Liu X, Li Y, Zhou X, Zhu S, Kaya NA, Chan YS, Ma L, Xu M, Zhai W: **An Integrative Analysis of Nasopharyngeal Carcinoma Genomes Unraveled Unique Processes Driving a Viral-Positive Cancer.** *Cancers* 2023, **15**(4).
28. Chan AS, To KF, Lo KW, Mak KF, Pak W, Chiu B, Tse GM, Ding M, Li X, Lee JC *et al*: **High frequency of chromosome 3p deletion in histologically normal nasopharyngeal epithelia from southern Chinese.** *Cancer research* 2000, **60**(19):5365-5370.
29. Lo KW, Teo PM, Hui AB, To KF, Tsang YS, Chan SY, Mak KF, Lee JC, Huang DP: **High resolution allelotyping of microdissected primary nasopharyngeal carcinoma.** *Cancer research* 2000, **60**(13):3348-3353.
30. Garcia-Higuera I, Taniguchi T, Ganesan S, Meyn MS, Timmers C, Hejna J, Grompe M, D'Andrea AD: **Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway.** *Molecular cell* 2001, **7**(2):249-262.
31. Ding RB, Chen P, Rajendran BK, Lyu X, Wang H, Bao J, Zeng J, Hao W, Sun H, Wong AH *et al*: **Molecular landscape and subtype-specific therapeutic response of nasopharyngeal carcinoma revealed by integrative pharmacogenomics.** *Nat Commun* 2021, **12**(1):3046.

32. Metcalf JL, Bradshaw PS, Komosa M, Greer SN, Stephen Meyn M, Ohh M: **K63-ubiquitylation of VHL by SOCS1 mediates DNA double-strand break repair.** *Oncogene* 2014, **33**(8):1055-1065.
33. Chen LC, Matsumura K, Deng G, Kurisu W, Ljung BM, Lerman MI, Waldman FM, Smith HS: **Deletion of two separate regions on chromosome 3p in breast cancers.** *Cancer research* 1994, **54**(11):3021-3024.
34. Kerangueven F, Noguchi T, Wargniez V, Allione F, Geneix J, Jacquemier J, Theillet C, Sobol H, Birnbaum D: **Multiple sites of loss of heterozygosity on chromosome arms 3p and 3q in human breast carcinomas.** *Oncol Rep* 1996, **3**(2):313-316.
35. Matsumoto S, Kasumi F, Sakamoto G, Onda M, Nakamura Y, Emi M: **Detailed deletion mapping of chromosome arm 3p in breast cancers: a 2-cM region on 3p14.3-21.1 and a 5-cM region on 3p24.3-25.1 commonly deleted in tumors.** *Genes Chromosomes Cancer* 1997, **20**(3):268-274.
36. Maitra A, Wistuba II, Washington C, Virmani AK, Ashfaq R, Milchgrub S, Gazdar AF, Minna JD: **High-resolution chromosome 3p allelotyping of breast carcinomas and precursor lesions demonstrates frequent loss of heterozygosity and a discontinuous pattern of allele loss.** *Am J Pathol* 2001, **159**(1):119-130.
37. Martinez A, Walker RA, Shaw JA, Dearing SJ, Maher ER, Latif F: **Chromosome 3p allele loss in early invasive breast cancer: detailed mapping and association with clinicopathological features.** *Mol Pathol* 2001, **54**(5):300-306.
38. Yang Q, Yoshimura G, Mori I, Sakurai T, Kakudo K: **Chromosome 3p and breast cancer.** *J Hum Genet* 2002, **47**(9):453-459.
39. Tang XL, Yao GY, Chen LR, Yang ZR, Li SL: **[Loss of heterozygosity on chromosome 3p in breast cancer and precancerous lesion].** *Zhonghua Wai Ke Za Zhi* 2006, **44**(19):1314-1317.
40. Fu Q, Yao GY, Tang XL, Chen LR, Zheng ZX: **[Microsatellite instability and allele-specific chromosome 3p deletion in breast cancer and precancerous lesions].** *Zhonghua zhong liu za zhi [Chinese journal of oncology]* 2007, **29**(1):34-40.
41. Loginov VI, Bazov IV, Khodyrev DS, Pronina IV, Kazubskaja TP, Ermilova VD, Gar'kavtseva RF, Zbarovskii ER, Braga EA: **[Human chromosome 3P regions of putative tumor-suppressor genes in renal, breast, and ovarian carcinomas].** *Genetika* 2008, **44**(2):250-256.
42. Ismail HM, Medhat AM, Karim AM, Zakhary NI: **FHIT gene and flanking region on chromosome 3p are subjected to extensive allelic loss in Egyptian breast cancer patients.** *Mol Carcinog* 2011, **50**(8):625-634.
43. de Oliveira MM, de Oliveira SF, Lima RS, de Andrade Urban C, Cavalli LR, de Souza Fonseca Ribeiro EM, Cavalli IJ: **Differential loss of heterozygosity profile on chromosome 3p in ductal and lobular breast carcinomas.** *Hum Pathol* 2012, **43**(10):1661-1667.
44. van der Groep P, Hoelzel M, Buerger H, Joenje H, de Winter JP, van Diest PJ: **Loss of expression of FANCD2 protein in sporadic and hereditary breast cancer.** *Breast cancer research and treatment* 2008, **107**(1):41-47.
45. Angeloni D: **Molecular analysis of deletions in human chromosome 3p21 and the role of resident cancer genes in disease.** *Brief Funct Genomic Proteomic* 2007, **6**(1):19-39.
46. 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.
47. Xiao YP, Han CB, Mao XY, Li JY, Xu L, Ren CS, Xin Y: **Relationship between abnormality of FHIT gene and EBV infection in gastric cancer.** *World journal of gastroenterology : WJG* 2005, **11**(21):3212-3216.
48. Kalezic A, Udicki M, Srdic Galic B, Aleksic M, Korac A, Jankovic A, Korac B: **Tissue-Specific Warburg Effect in Breast Cancer and Cancer-Associated Adipose Tissue-Relationship between AMPK and Glycolysis.** *Cancers* 2021, **13**(11).

49. Liu S, Li Y, Yuan M, Song Q, Liu M: **Correlation between the Warburg effect and progression of triple-negative breast cancer.** *Front Oncol* 2022, **12**:1060495.
50. Larson ED, Maizels N: **Transcription-coupled mutagenesis by the DNA deaminase AID.** *Genome Biol* 2004, **5**(3):211.
51. Arai A, Horino M, Imadome K-I, Imai K, Komatsu H, Wang L, Matsuda G, Hamazaki K, Kurata M, Kurosu T *et al*: **Epstein-Barr Virus Induces Activation-Induced Cytidine Deaminase Expression in T or NK Cells Leading to Mutagenesis and Development of Lymphoma.** *Blood* 2013, **122**(21):1765-1765.
52. Nagata K, Kumata K, Nakayama Y, Satoh Y, Sugihara H, Hara S, Matsushita M, Kuwamoto S, Kato M, Murakami I *et al*: **Epstein-Barr Virus Lytic Reactivation Activates B Cells Polyclonally and Induces Activation-Induced Cytidine Deaminase Expression: A Mechanism Underlying Autoimmunity and Its Contribution to Graves' Disease.** *Viral Immunol* 2017, **30**(3):240-249.
53. Damaghi M, West J, Robertson-Tessi M, Xu L, Ferrall-Fairbanks MC, Stewart PA, Persi E, Fridley BL, Altrock PM, Gatenby RA *et al*: **The harsh microenvironment in early breast cancer selects for a Warburg phenotype.** *Proceedings of the National Academy of Sciences of the United States of America* 2021, **118**(3).
54. Lee JJ, Jung YL, Cheong TC, Espejo Valle-Inclan J, Chu C, Gulhan DC, Ljungstrom V, Jin H, Viswanadham VV, Watson EV *et al*: **ERalpha-associated translocations underlie oncogene amplifications in breast cancer.** *Nature* 2023, **618**(7967):1024-1032.
55. Ramiro AR, Jankovic M, Eisenreich T, Difilippantonio S, Chen-Kiang S, Muramatsu M, Honjo T, Nussenzweig A, Nussenzweig MC: **AID is required for c-myc/IgH chromosome translocations in vivo.** *Cell* 2004, **118**(4):431-438.
56. Xie W, Duan R, Safe S: **Estrogen induces adenosine deaminase gene expression in MCF-7 human breast cancer cells: role of estrogen receptor-Sp1 interactions.** *Endocrinology* 1999, **140**(1):219-227.
57. Gil Y, Levy-Nabot S, Steinitz M, Laskov R: **Somatic mutations and activation-induced cytidine deaminase (AID) expression in established rheumatoid factor-producing lymphoblastoid cell line.** *Molecular immunology* 2007, **44**(4):494-505.
58. Pauklin S, Sernandez IV, Bachmann G, Ramiro AR, Petersen-Mahrt SK: **Estrogen directly activates AID transcription and function.** *The Journal of experimental medicine* 2009, **206**(1):99-111.
59. Burkhardt B, Michgehl U, Rohde J, Erdmann T, Berning P, Reutter K, Rohde M, Borkhardt A, Burmeister T, Dave S *et al*: **Clinical relevance of molecular characteristics in Burkitt lymphoma differs according to age.** *Nat Commun* 2022, **13**(1):3881.
60. Ju BG, Lunyak VV, Perissi V, Garcia-Bassets I, Rose DW, Glass CK, Rosenfeld MG: **A topoisomerase IIbeta-mediated dsDNA break required for regulated transcription.** *Science* 2006, **312**(5781):1798-1802.
61. Wang P, Rennekamp AJ, Yuan Y, Lieberman PM: **Topoisomerase I and RecQL1 function in Epstein-Barr virus lytic reactivation.** *Journal of virology* 2009, **83**(16):8090-8098.
62. Dochi H, Kondo S, Murata T, Fukuyo M, Nanbo A, Wakae K, Jiang WP, Hamabe-Horiike T, Tanaka M, Nishiuchi T *et al*: **Estrogen induces the expression of EBV lytic protein ZEBRA, a marker of poor prognosis in nasopharyngeal carcinoma.** *Cancer science* 2022, **113**(8):2862-2877.
63. D'Andrea AD: **Susceptibility pathways in Fanconi's anemia and breast cancer.** *The New England journal of medicine* 2010, **362**(20):1909-1919.
64. Rawlins DR, Milman G, Hayward SD, Hayward GS: **Sequence-specific DNA binding of the Epstein-Barr virus nuclear antigen (EBNA-1) to clustered sites in the plasmid maintenance region.** *Cell* 1985, **42**(3):859-868.

65. Li JSZ, Abbasi A, Kim DH, Lippman SM, Alexandrov LB, Cleveland DW: **Chromosomal fragile site breakage by EBV-encoded EBNA1 at clustered repeats.** *Nature* 2023, **616**(7957):504-509.
66. 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.
67. Bochkarev A, Barwell JA, Pfuetzner RA, Bochkareva E, Frappier L, Edwards AM: **Crystal structure of the DNA-binding domain of the Epstein-Barr virus origin-binding protein, EBNA1, bound to DNA.** *Cell* 1996, **84**(5):791-800.
68. Altemose N, Logsdon GA, Bzikadze AV, Sidhwani P, Langley SA, Caldas GV, Hoyt SJ, Uralsky L, Ryabov FD, Shew CJ *et al*: **Complete genomic and epigenetic maps of human centromeres.** *Science* 2022, **376**(6588):eabl4178.
69. Kim TM, Xi R, Luquette LJ, Park RW, Johnson MD, Park PJ: **Functional genomic analysis of chromosomal aberrations in a compendium of 8000 cancer genomes.** *Genome research* 2013, **23**(2):217-227.
70. Tsang CM, Lui VWY, Bruce JP, Pugh TJ, Lo KW: **Translational genomics of nasopharyngeal cancer.** *Semin Cancer Biol* 2020, **61**:84-100.
71. Jiang Y, Xiao H, Sun L, Zhang Y, Liu S, Luo B: **LMP2A suppresses the role of AHR pathway through ERK signal pathway in EBV-associated gastric cancer.** *Virus research* 2021, **297**:198399.
72. Shen JJ, Niu WN, Zhou M, Zhou F, Zhang HY, Wang L: **Association of Epstein Barr virus A73 gene polymorphism with nasopharyngeal carcinoma.** *Genet Test Mol Biomarkers* 2015, **19**(4):187-190.
73. Neuhierl B, Feederle R, Hammerschmidt W, Delecluse HJ: **Glycoprotein gp110 of Epstein-Barr virus determines viral tropism and efficiency of infection.** *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(23):15036-15041.
74. Chiu SH, Wu MC, Wu CC, Chen YC, Lin SF, Hsu JT, Yang CS, Tsai CH, Takada K, Chen MR *et al*: **Epstein-Barr virus BALF3 has nuclease activity and mediates mature virion production during the lytic cycle.** *Journal of virology* 2014, **88**(9):4962-4975.
75. Liu X, Deng Y, Huang Y, Ye J, Xie S, He Q, Chen Y, Lin Y, Liang R, Wei J *et al*: **Nasopharyngeal Carcinoma Progression: Accumulating Genomic Instability and Persistent Epstein-Barr Virus Infection.** *Curr Oncol* 2022, **29**(9):6035-6052.
76. Singh DR, Nelson SE, Pawelski AS, Kansra AS, Fogarty SA, Bristol JA, Ohashi M, Johannsen EC, Kenney SC: **Epstein-Barr virus LMP1 protein promotes proliferation and inhibits differentiation of epithelial cells via activation of YAP and TAZ.** *Proceedings of the National Academy of Sciences of the United States of America* 2023, **120**(20):e2219755120.
77. Hau PM, Tsao SW: **Epstein-Barr Virus Hijacks DNA Damage Response Transducers to Orchestrate Its Life Cycle.** *Viruses* 2017, **9**(11).
78. Kudoh A, Iwahori S, Sato Y, Nakayama S, Isomura H, Murata T, Tsurumi T: **Homologous recombinational repair factors are recruited and loaded onto the viral DNA genome in Epstein-Barr virus replication compartments.** *Journal of virology* 2009, **83**(13):6641-6651.
79. Chatterjee K, Das P, Chattopadhyay NR, Mal S, Choudhuri T: **The interplay between Epstein-Bar virus (EBV) with the p53 and its homologs during EBV associated malignancies.** *Heliyon* 2019, **5**(11):e02624.
80. Whang-Peng J: **3p deletion and small cell lung carcinoma.** *Mayo Clin Proc* 1989, **64**(2):256-260.
81. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J *et al*: **Genomic and Functional Approaches to Understanding Cancer Aneuploidy.** *Cancer cell* 2018, **33**(4):676-689 e673.

82. Arnold M, Morgan E, Rungay H, Mafra A, Singh D, Laversanne M, Vignat J, Gralow JR, Cardoso F, Siesling S *et al*: **Current and future burden of breast cancer: Global statistics for 2020 and 2040.** *Breast* 2022, **66**:15-23.
83. Dhakal R, Noula M, Roupa Z, Yamasaki EN: **A Scoping Review on the Status of Female Breast Cancer in Asia with a Special Focus on Nepal.** *Breast Cancer (Dove Med Press)* 2022, **14**:229-246.
84. Mount DW: **Using the Basic Local Alignment Search Tool (BLAST).** *CSH Protoc* 2007, **2007**:pdb top17.
85. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *Journal of molecular biology* 1990, **215**(3):403-410.
86. Zhang Z, Schwartz S, Wagner L, Miller W: **A greedy algorithm for aligning DNA sequences.** *J Comput Biol* 2000, **7**(1-2):203-214.
87. Kleinberger I, Sanders E, Staes K, Van Troys M, Hirano S, Hochepped T, Lemeire K, Martens L, Ampe C, van Roy F: **Innovative mouse models for the tumor suppressor activity of Protocadherin-10 isoforms.** *BMC cancer* 2022, **22**(1):451.
88. Liu W, Wu J, Shi G, Yue X, Liu D, Zhang Q: **Aberrant promoter methylation of PCDH10 as a potential diagnostic and prognostic biomarker for patients with breast cancer.** *Oncol Lett* 2018, **16**(4):4462-4470.
89. Kim MS, Lee HS, Kim YJ, Lee DY, Kang SG, Jin W: **MEST induces Twist-1-mediated EMT through STAT3 activation in breast cancers.** *Cell death and differentiation* 2019, **26**(12):2594-2606.
90. Lin J, Li Q, Lei X, Zhao H: **The emerging roles of GPR158 in the regulation of the endocrine system.** *Front Cell Dev Biol* 2022, **10**:1034348.
91. Pointer TC, Gorelick FS, Desir GV: **Renalase: A Multi-Functional Signaling Molecule with Roles in Gastrointestinal Disease.** *Cells* 2021, **10**(8).
92. Luo Z, Gao X, Lin C, Smith ER, Marshall SA, Swanson SK, Florens L, Washburn MP, Shilatifard A: **Zic2 is an enhancer-binding factor required for embryonic stem cell specification.** *Molecular cell* 2015, **57**(4):685-694.
93. Liu ZH, Chen ML, Zhang Q, Zhang Y, An X, Luo YL, Liu XM, Liu SX, Liu Q, Yang T *et al*: **ZIC2 is downregulated and represses tumor growth via the regulation of STAT3 in breast cancer.** *International journal of cancer Journal international du cancer* 2020, **147**(2):505-518.
94. Ear J, Dunkel Y, Mittal Y, Lim BBC, Liu L, Holda MK, Nitsche U, Barbazan J, Goel A, Janssen KP *et al*: **Two Isoforms of the Guanine Nucleotide Exchange Factor, Daple/CCDC88C Cooperate as Tumor Suppressors.** *Sci Rep* 2019, **9**(1):12124.
95. Kim YJ, Zhao Y, Myung JK, Yi JM, Kim MJ, Lee SJ: **Suppression of breast cancer progression by FBXL16 via oxygen-independent regulation of HIF1alpha stability.** *Cell reports* 2021, **37**(8):109996.
96. Lee H, Kwon MJ, Koo BM, Park HG, Han J, Shin YK: **A novel immune prognostic index for stratification of high-risk patients with early breast cancer.** *Sci Rep* 2021, **11**(1):128.
97. Zhao T, Wang C, Zhao N, Qiao G, Hua J, Meng D, Liu L, Zhong B, Liu M, Wang Y *et al*: **CYB561 promotes HER2+ breast cancer proliferation by inhibiting H2AFY degradation.** *Cell Death Discov* 2024, **10**(1):38.
98. Chen W, Li X, Jiang Y, Ni D, Yang L, Wu J, Gao M, Wang J, Song J, Shi W: **Pancancer analysis of the correlations of HS6ST2 with prognosis, tumor immunity, and drug resistance.** *Sci Rep* 2023, **13**(1):19209.



## TABLES

**Table 1. Methylation sites that predispose to breast cancer are near aberrant methylation sites in infected keratinocytes that have lost their EBV infection.**

Gene showing persistent methylation after the clearance of EBV	Gene name	Chromosome coordinates (Hg19)	Distance (bp) from sites of hyper-methylation in breast cancers	Function, relation to breast cancer
PCDH10	Protocadherin-10	4: 134,070,449-134,129,761	945	Cell-cell adhesion [87]. Inhibits cancer cell motility and cell migration. PCDH10 methylation is common in primary breast cancer and associated with poor survival [88].
MEST	Mesoderm-specific transcript	7: 130,126,016-130,146,306	4360	Loss of imprinting has been linked to promoter switch and breast cancer metastasis [89]
GPR158	Signal transduction, G-protein-coupled receptor	10:25,463,930-25,891,158	465	Estrogen withdrawal or antagonist exposure down-regulates GPR158 expression in estrogen-sensitive breast cancer. GPR158 may predict prognosis in breast and ovarian cancer [90]
RNLS	Renalase, FAD-dependent amine oxidase	10:89,931,280-90,343,075	137	Promotes survival with anti-inflammatory effects. Overexpressed in breast cancer [91]
ZIC2	Zinc finger protein AIC family member 2	13:100,632,108 - 100,632,305; 100,632,306-100,632,466; 100,632,478-100,632,573; 100,633,028-100,633,249	789	Regulates chromatin and transcription. Embryonic cell differentiation [92] Downregulated in breast cancer [93]
CCDC88C	coiled-coil domain containing 88C	14:91,737,667-91,884,164	4658	Highly mutated in breast cancer. A tumor suppressor that is related to Wnt signaling [94]
FBXL16	F-box and leucine-rich repeat protein 16	16:742,500-755,801	53	Suppression is associated with aggressive, lethal breast cancers [95]
IL21R	interleukin 21 receptor	16:27,413,495-27,463,363	8168	Activates NK cells. Part of an immune prognostic gene panel that predicts breast cancer survival [96].
CYB561	cytochrome b561	17:61,509,665-61,524,000	8543	Increases breast cancer cell proliferation, especially HER2+ breast cancer [97].
HS6ST2	heparan sulfate 6-O-sulfotransferase	X:131,760,043-132,095,398	199	HS6ST2 encodes a heparan sulfate (HS) sulfotransferase gene. The encoded protein catalyzes the transfer of sulfate to HS. HS6ST2 regulates cell adhesion,

	2			migration, growth, and differentiation. Low expression is related to poor breast cancer survival [98]
--	---	--	--	---

**Table 2. Hypermethylation in breast cancers inhibits genes that are essential for protection against EBV-related cancer (references listed in Table S2)**

Chromosome	Gene that are hyper-methylated in both NPC and BRC	Function related to EBV immunity or EBV infection and reference
1	PRDX1	The encoded protein is essential to protect cells from replication stress and DNA damage. Linked to cancer development and the antiviral activity of T-cells [1, 2]
1	NFIA	Associated with common variable immunodeficiency-like B-cell disease, recurrent pulmonary infections, and EBV lymphoproliferative disease [3].
1	CASZ1	A target of EBV-encoded miRNAs [4]. Regulates the Th17/Th1/regulatory cell differentiation program, potentially influencing the immune response to EBV. [5]
1	MAD2L2	Associated with immune cells, including CD4+T cells, CD8+T cells, B-cells, natural killer cells, and myeloid dendritic cells
1	AK4	Regulates the proliferation of EBV-infected B-cells, which is normally controlled by T-cells as part of an antiviral immune response.
1	CAMK2N1	Associated with poor prognosis and tumor immune infiltration of gastric cancer, sometimes caused by EBV [6].
2	GPC1	Involved in EBV infection and immune cell infiltration [7]
2	CYP1B1	Modulates anti-EBV immune responses [8]. Several connections to EBV oncogenesis
3	PARP9	DNA damage repair drives interferon-mediated antiviral defenses and other immune responses [9, 10].
5	CXXC5	An epigenetic regulator that recruits Tet2 demethylase to regulate IFN responses [11]
5	FAM134B	Limits the replication of several viruses, including dengue virus, zika, WNV, and Ebola virus [12].
5	IQGAP2	Upregulated in EBV-infected cells and induced by the lytic protein rta. Knockdown of IQGAP2 affects the clumping of infected cells. Varied functions related to its binding partner. Modulates the cytoskeleton [13].
7	HOXA10-HOXA9	EBV-driven DNA hypermethylation downregulates the HOXA10 and HOXA9 genes. HOXA10 has antiviral activity against both EBV and KSHV2 [14].
8	PUF-60	Interacts with the EBV nuclear protein EB2 (Mta/SM/BMLF1) which causes the cytoplasmic accumulation of viral mRNAs from genes without introns [15]
9	PKN3	Regulates NFkB activation, a pathway activated by EBV [16, 17]
9	NFIB	Associated with NFkB1 and important for NFkB signaling in both B-cells and cytotoxic T-cells [3]
10	GATA3	Key regulator of T-cell development, its expression in breast cancer is a favorable prognostic indicator [18]
10	CTBP2	A target of EBV transcription factors [19]
11	MUC5AC	A member of a family of proteins that form a protective mucus barrier on epithelial surfaces to protect against infections
11	IFITM1,M2,M3	Encode proteins that restrict cellular entry by diverse viral pathogens [20, 21]
11	NRXN	On chromosome 11, NRXN2 forms a binding site for EBNA1 which then attaches to the viral episome [22].

12	KLRF1	Encodes an activating costimulatory receptor for NK cells
12	FAIM2	Marker of immune infiltration [23]
12	HSPB8	Cell stress response, regulates cell division. HSPB8 shares functions that regulate cell life and death with its homologue in herpes viruses [24]
13	DACH1	Enhances cell cycle arrest by binding p53. Prevents breast cancer invasion and metastasis. DACH1 is a cytokine hijacked by a prototype foamy virus [25].
14	FOXA1	Overexpression of FOXA1 suppresses IFN.
15	NR2F2	A long non-coding RNA downregulated in EBV infection [26]
16	PGP	Required for cell-mediated cytotoxicity and other immune functions [27].
17	ADORA2B	An adenosine receptor that activates adenylyl cyclase signals, present on immune cells, e.g. macrophages, dendritic cells, killer and NK-cells, T-, and B-cells. Signaling controls the recruitment, trafficking, and function of these cells [28].
17	RAB34	Essential for immune responses. Involved in primary cilium membrane formation [29] to receive and transduce environmental signals. Phagosome maturation [30]
17	SARM1	Negative regulator of innate immunity [31].
17	SOX9	Immune cell infiltration into tumors [32].
18	SPIRE1	A virus restriction factor that stimulates innate immunity [33].
18	ANKRD12	Inhibits nuclear receptors via histone deacetylases. ANKRD12 silencing promotes invasion [34].
18	CHD7	Chromatin remodeling.
18	DTNA	Upregulated in viral cancer [35], controls apoptosis. Localized to cell junctions
18	GREB1L	Associated with good clinical outcomes in breast cancer, interacts with estrogen, and promotes immune cell invasion [36, 37]
19	NFIX	Involved in the oxidative stress response and cell fate decisions [38]. Binds viral and cellular promoters.
19	BST2	Natural immune factor in viral infections [39].
20	TAF4	Interacts with the EBV protein rta to activate transcription [40].
21	SUMO3	Equivalent to SUMO2 which causes changes in response to EBV reactivation. The EBV protein BZLF1 disables this defense. The downregulation of sumoylation is essential for lytic infection and progression through the EBV lytic cycle [41].

#### Supplementary References in Table 2

1. Cao R, Zhang W, Zhang H, Wang L, Chen X, Ren X, Cheng B, Xia J: **Comprehensive Analysis of the PRDXs Family in Head and Neck Squamous Cell Carcinoma.** *Front Oncol* 2022, **12**:798483.
2. Moretton A, Kourtis S, Ganez Zapater A, Calabro C, Espinar Calvo ML, Fontaine F, Darai E, Abad Cortel E, Block S, Pascual-Reguant L *et al*: **A metabolic map of the DNA damage response identifies PRDX1 in the control of nuclear ROS scavenging and aspartate availability.** *Molecular systems biology* 2023, **19**(7):e11267.
3. Cohen JI, Meyts I: **Editorial: EBV Infection and Human Primary Immune Deficiencies.** *Frontiers in immunology* 2020, **11**:130.
4. Skalsky RL, Cullen BR: **EBV Noncoding RNAs.** *Curr Top Microbiol Immunol* 2015, **391**:181-217.

5. Dalton T, Doubrovina E, Pankov D, Reynolds R, Scholze H, Selvakumar A, Vizconde T, Savalia B, Dyomin V, Weigel C *et al*: **Epigenetic reprogramming sensitizes immunologically silent EBV+ lymphomas to virus-directed immunotherapy.** *Blood* 2020, **135**(21):1870-1881.
6. Lin JH, Lin JY, Chou YC, Chen MR, Yeh TH, Lin CW, Lin SJ, Tsai CH: **Epstein-Barr virus LMP2A suppresses MHC class II expression by regulating the B-cell transcription factors E47 and PU.1.** *Blood* 2015, **125**(14):2228-2238.
7. Xiong J, Dai YT, Wang WF, Zhang H, Wang CF, Yin T, Cheng S, Zhong HJ, Yu SH, Jiang L *et al*: **GPCR signaling contributes to immune characteristics of microenvironment and process of EBV-induced lymphomagenesis.** *Sci Bull (Beijing)* 2023, **68**(21):2607-2619.
8. Torti MF, Giovannoni F, Quintana FJ, Garcia CC: **The Aryl Hydrocarbon Receptor as a Modulator of Anti-viral Immunity.** *Frontiers in immunology* 2021, **12**:624293.
9. Long HM, Meckiff BJ, Taylor GS: **The T-cell Response to Epstein-Barr Virus-New Tricks From an Old Dog.** *Frontiers in immunology* 2019, **10**:2193.
10. Thirunavukkarasu S, Ahmed M, Rosa BA, Boothby M, Cho SH, Rangel-Moreno J, Mbandi SK, Schreiber V, Gupta A, Zuniga J *et al*: **Poly(ADP-ribose) polymerase 9 mediates early protection against Mycobacterium tuberculosis infection by regulating type I IFN production.** *The Journal of clinical investigation* 2023, **133**(12).
11. Ma S, Wan X, Deng Z, Shi L, Hao C, Zhou Z, Zhou C, Fang Y, Liu J, Yang J *et al*: **Epigenetic regulator CXXC5 recruits DNA demethylase Tet2 to regulate TLR7/9-elicited IFN response in pDCs.** *The Journal of experimental medicine* 2017, **214**(5):1471-1491.
12. Guan K, Su Q, Kuang K, Meng X, Zhou X, Liu B: **MiR-142-5p/FAM134B Axis Manipulates ER-Phagy to Control PRRSV Replication.** *Frontiers in immunology* 2022, **13**:842077.
13. Lin KM, Weng LF, Chen SJ, Lin SJ, Tsai CH: **Upregulation of IQGAP2 by EBV transactivator Rta and its influence on EBV life cycle.** *Journal of virology* 2023, **97**(8):e0054023.
14. Fiches GN, Zhou D, Kong W, Biswas A, Ahmed EH, Baiocchi RA, Zhu J, Santoso N: **Profiling of immune related genes silenced in EBV-positive gastric carcinoma identified novel restriction factors of human gammaherpesviruses.** *PLoS pathogens* 2020, **16**(8):e1008778.
15. Juillard F, Bazot Q, Mure F, Tafforeau L, Macri C, Rabourdin-Combe C, Lotteau V, Manet E, Gruffat H: **Epstein-Barr virus protein EB2 stimulates cytoplasmic mRNA accumulation by counteracting the deleterious effects of SRp20 on viral mRNAs.** *Nucleic acids research* 2012, **40**(14):6834-6849.

16. Gewurz BE, Towfic F, Mar JC, Shinnars NP, Takasaki K, Zhao B, Cahir-McFarland ED, Quackenbush J, Xavier RJ, Kieff E: **Genome-wide siRNA screen for mediators of NF-kappaB activation.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(7):2467-2472.
17. Lavorgna A, Harhaj EW: **EBV LMP1: New and shared pathways to NF-kappaB activation.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(7):2188-2189.
18. Albergaria A, Paredes J, Sousa B, Milanezi F, Carneiro V, Bastos J, Costa S, Vieira D, Lopes N, Lam EW *et al*: **Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours.** *Breast cancer research : BCR* 2009, **11**(3):R40.
19. McClellan MJ, Wood CD, Ojeniyi O, Cooper TJ, Kanhere A, Arvey A, Webb HM, Palermo RD, Harth-Hertle ML, Kempkes B *et al*: **Modulation of enhancer looping and differential gene targeting by Epstein-Barr virus transcription factors directs cellular reprogramming.** *PLoS pathogens* 2013, **9**(9):e1003636.
20. Narayana SK, Helbig KJ, McCartney EM, Eyre NS, Bull RA, Eltahla A, Lloyd AR, Beard MR: **The Interferon-induced Transmembrane Proteins, IFITM1, IFITM2, and IFITM3 Inhibit Hepatitis C Virus Entry.** *The Journal of biological chemistry* 2015, **290**(43):25946-25959.
21. Shi G, Kenney AD, Kudryashova E, Zani A, Zhang L, Lai KK, Hall-Stoodley L, Robinson RT, Kudryashov DS, Compton AA, Yount JS: **Opposing activities of IFITM proteins in SARS-CoV-2 infection.** *The EMBO journal* 2021, **40**(3):e106501.
22. 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.
23. Cai J, Ye Z, Hu Y, Wang Y, Ye L, Gao L, Sun Q, Tong S, Sun Z, Yang J, Chen Q: **FAIM2 is a potential pan-cancer biomarker for prognosis and immune infiltration.** *Front Oncol* 2022, **12**:998336.
24. Aurelian L, Laing JM, Lee KS: **H11/HspB8 and Its Herpes Simplex Virus Type 2 Homologue ICP10PK Share Functions That Regulate Cell Life/Death Decisions and Human Disease.** *Autoimmune Dis* 2012, **2012**:395329.
25. Ma Y, Wei J, Song J, Hu Z, Zhang R, Li Z, Sun Y: **The DACH1 Gene Transcriptional Activation and Protein Degradation Mediated by Transactivator Tas of Prototype Foamy Virus.** *Viruses* 2023, **15**(9).
26. Zhang J, Li X, Hu J, Cao P, Yan Q, Zhang S, Dang W, Lu J: **Long noncoding RNAs involvement in Epstein-Barr virus infection and tumorigenesis.** *Virology journal* 2020, **17**(1):51.

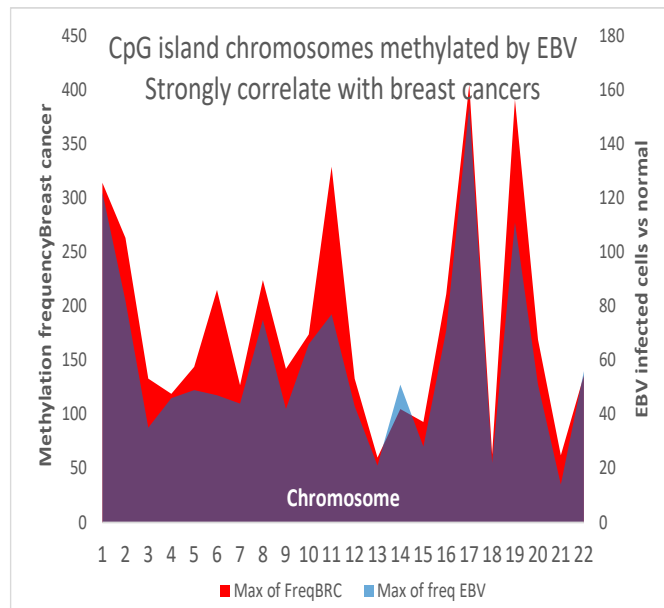
27. Sankatsing SU, Beijnen JH, Schinkel AH, Lange JM, Prins JM: **P glycoprotein in human immunodeficiency virus type 1 infection and therapy.** *Antimicrobial agents and chemotherapy* 2004, **48**(4):1073-1081.
28. Strickland LN, Faraoni EY, Ruan W, Yuan X, Eltzschig HK, Bailey-Lundberg JM: **The resurgence of the Adora2b receptor as an immunotherapeutic target in pancreatic cancer.** *Frontiers in immunology* 2023, **14**:1163585.
29. Bruel AL, Ganga AK, Noskova L, Valenzuela I, Martinovic J, Duffourd Y, Zikanova M, Majer F, Kmoch S, Mohler M *et al*: **Pathogenic RAB34 variants impair primary cilium assembly and cause a novel oral-facial-digital syndrome.** *Human molecular genetics* 2023, **32**(18):2822-2831.
30. Kasmapour B, Gronow A, Bleck CK, Hong W, Gutierrez MG: **Size-dependent mechanism of cargo sorting during lysosome-phagosome fusion is controlled by Rab34.** *Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**(50):20485-20490.
31. Hopkins EL, Gu W, Kobe B, Coleman MP: **A Novel NAD Signaling Mechanism in Axon Degeneration and its Relationship to Innate Immunity.** *Front Mol Biosci* 2021, **8**:703532.
32. Liu S, Yang L, Fu J, Li T, Zhou B, Wang K, Wei C, Fu J: **Comprehensive analysis, immune, and cordycepin regulation for SOX9 expression in pan-cancers and the matched healthy tissues.** *Frontiers in immunology* 2023, **14**:1149986.
33. Torres AA, Macilwee SL, Rashid A, Cox SE, Albarnaz JD, Bonjardim CA, Smith GL: **The actin nucleator Spir-1 is a virus restriction factor that promotes innate immune signalling.** *PLoS pathogens* 2022, **18**(2):e1010277.
34. Ribeiro V, Martins SG, Lopes AS, Thorsteinsdottir S, Zilhao R, Carlos AR: **NFIXing Cancer: The Role of NFIX in Oxidative Stress Response and Cell Fate.** *International journal of molecular sciences* 2023, **24**(5).
35. Cao Y, Xie L, Shi F, Tang M, Li Y, Hu J, Zhao L, Zhao L, Yu X, Luo X *et al*: **Targeting the signaling in Epstein-Barr virus-associated diseases: mechanism, regulation, and clinical study.** *Signal Transduct Target Ther* 2021, **6**(1):15.
36. Yang YC, Chang LK: **Role of TAF4 in transcriptional activation by Rta of Epstein-Barr Virus.** *PLoS one* 2013, **8**(1):e54075.
37. De La Cruz-Herrera CF, Tatham MH, Siddiqi UZ, Shire K, Marcon E, Greenblatt JF, Hay RT, Frappier L: **Changes in SUMO-modified proteins in Epstein-Barr virus infection identifies reciprocal regulation of TRIM24/28/33 complexes and the lytic switch BZLF1.** *PLoS pathogens* 2023, **19**(7):e1011477.

**Table 3. Relationships of breakpoints in HRD breast cancers to those in HRP breast cancers**

Chromosome	Fraction of HRD breast cancers with breaks matching (+-10k) those of HRP breast cancers	P values that unfiltered breakpoints have the same distribution (all HRP breast cancer breaks vs all HRD breast cancer breaks)	Number of HRD breakpoints/HRP breakpoints	P value assuming that all HRD breakpoints come from the same breakpoint distribution as that of the nearest HRPs (column 2)
1	22/34	<0.0001	391/1420	P=0.8836
2	11/30	0.0014	299/301	P=0.8732
3	16/55	0.68	654/414	P=0.6455
4	12/35	<0.0001	222/466	P=0.9622
5	8/28	<0.0001	168/258	P=0.9548
6	19/42	0.44	364/1090	P=0.9982
7	7/30	<0.0001	196/382	P=0.9113
8	46/60	<0.0001	354/2230	P=0.994
9	9/108	<0.0001	108/330	P=0.9012
10	27/33	0.0002	204/538	P=0.9987
11	35/47	<0.0001	322/2251	P=0.9821
12	16/33	0.0003	170/987	P=0.9658
13	19/27	<0.0001	150/432	P=0.9862
14	15/30	<0.0001	208/369	P=0.9828
15	32/32	0.001	226/207	P=0.9716
16	17/19	0.0311	108/276	P=0.9566
17	34/34	<0.0001	254/3462	P=0.9899

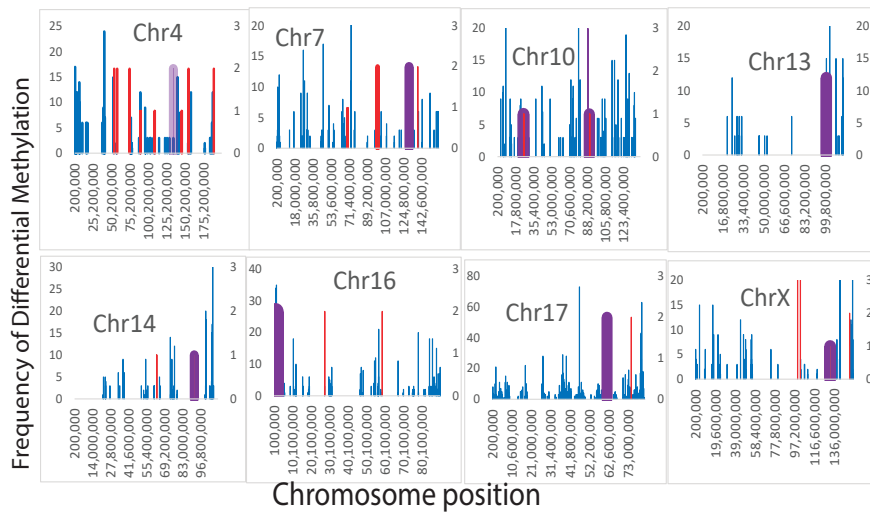




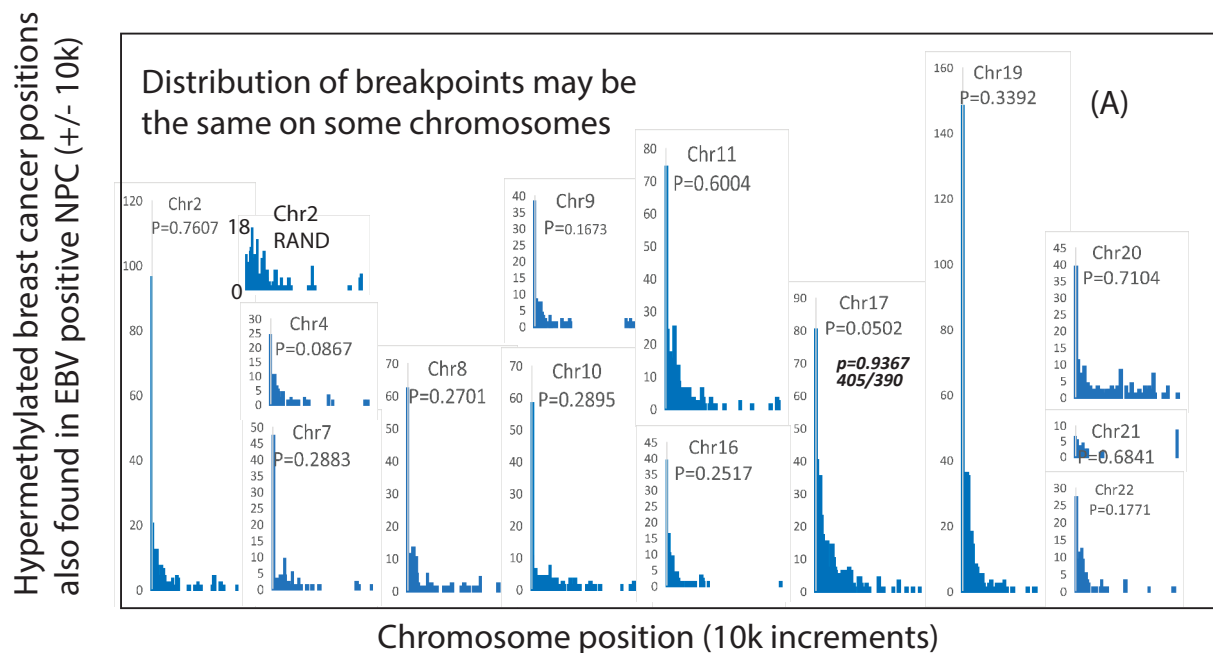


**Fig. 1A. The frequency of differential methylation of chromosomes in oral keratinocytes that have cleared EBV correlates with the frequency of differential methylation in breast cancer chromosomes.**

**Fig.1B**



**Fig. 1B. Differential methylations that persist after EBV infection clears from oral keratinocytes also occur in breast cancers. The frequencies of differential methylation in breast cancers cells (blue) and keratinocytes that have lost their EBV infection were compared. Chromosome positions that overlap are indicated in purple.**



**Fig 2A.** Positions of hypermethylation in NPC, an EBV-mediated cancer, relate to cis-hypermethylated positions in breast cancers. P values do not exclude the possibility that the hypermethylation positions in breast cancer and NPC are the same. Comparison to a normal random distribution with the same mean and standard deviation as the NPC data for chr2 is to the right of the chr2 graph and labeled “chr2RAND”.

Fig 2B

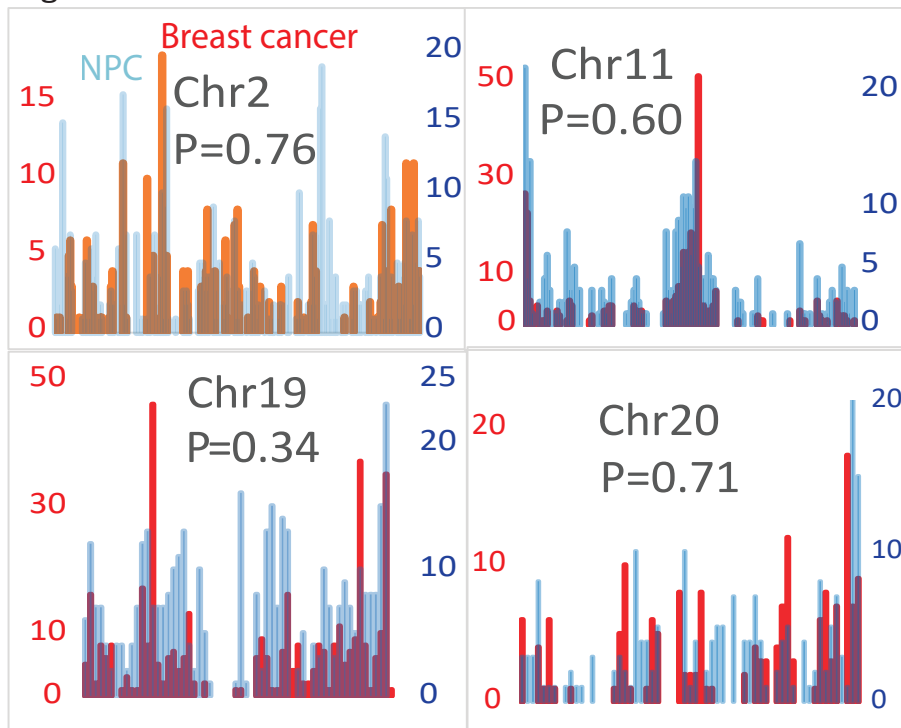


Fig 2B. The frequency distributions of methyl groups overlap in NPC and breast cancer when the chromosomes are divided into bins of 1 million bp.

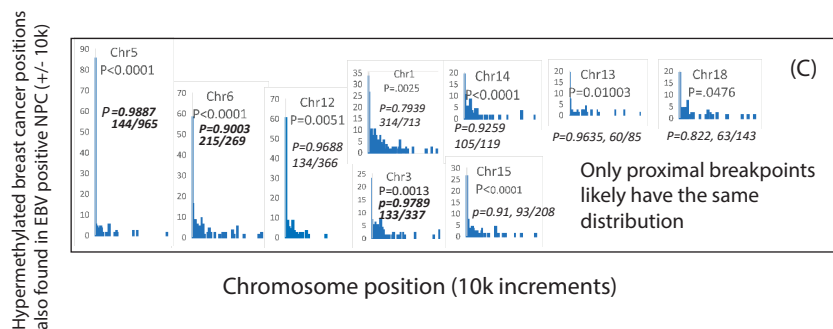
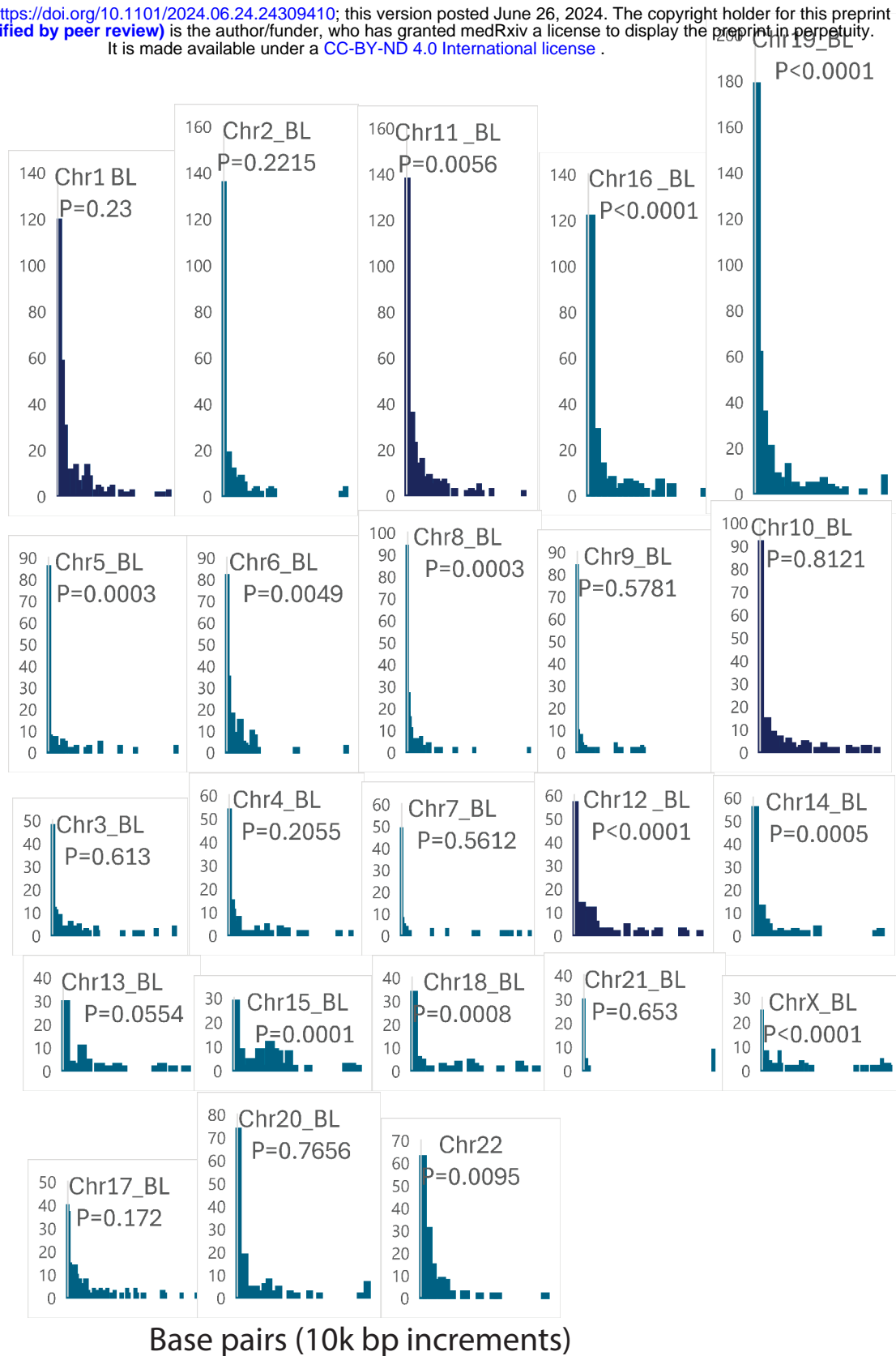
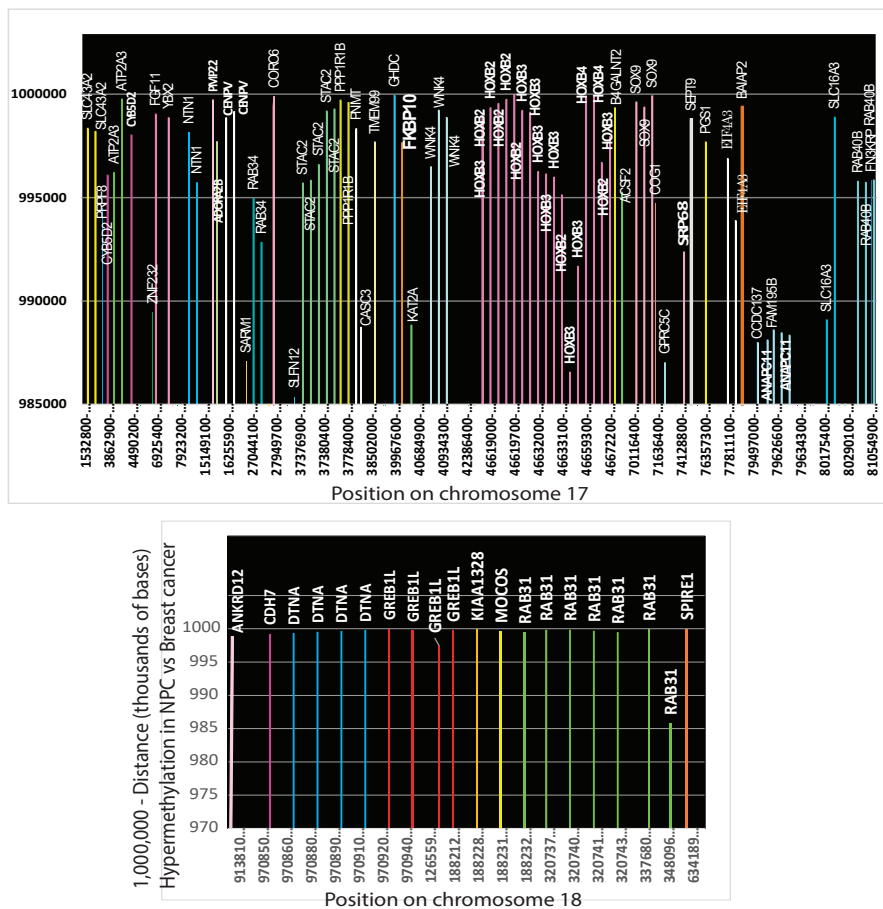


Fig. 2C. If P values suggested that the distributions of methylation positions in breast cancer and NPC were unrelated ( $P < 0.05$ ), probabilities were calculated that methylation positions within 10,000 base pairs of each other (fractions in bold type) came from the same distributions.

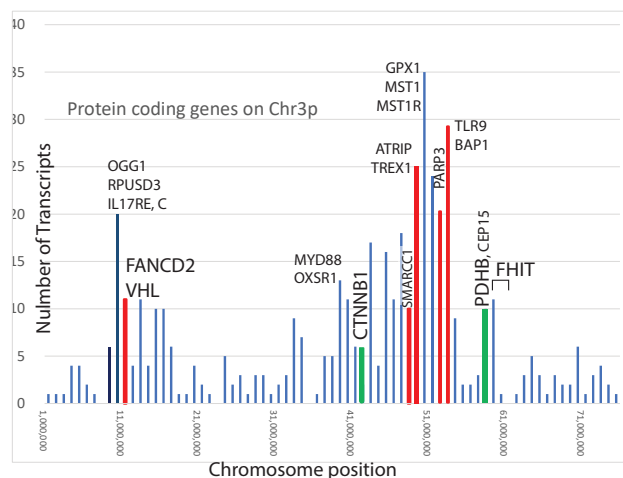
Number of differentially methylated breast cancer chromosome positions  
<10000 bases from differential methylation in Burkitt's Lymphoma



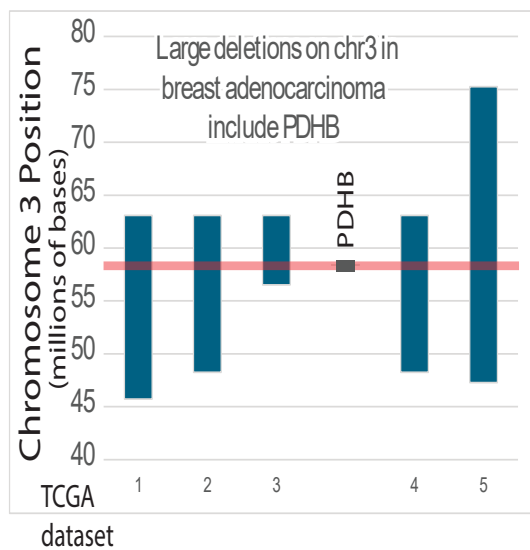
**Fig. 3. Differential methylation occurs on every chromosome at approximately the same positions in breast cancer and BL. Positions of differential methylation on every chromosome in 10k base pair increments (breast cancer vs. normal compared to BL vs. normal). P values are probabilities that the values come from the same distribution.**



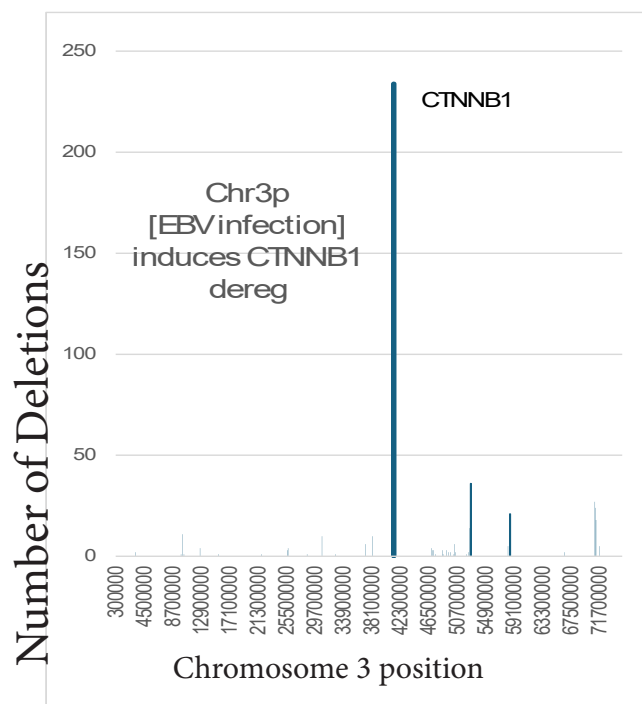
**Fig 4. Genes hypermethylated in NPC cluster around cis-regulated genes hypermethylated in breast cancer. Data from chromosome 17 and chromosome 18 are shown as examples.**



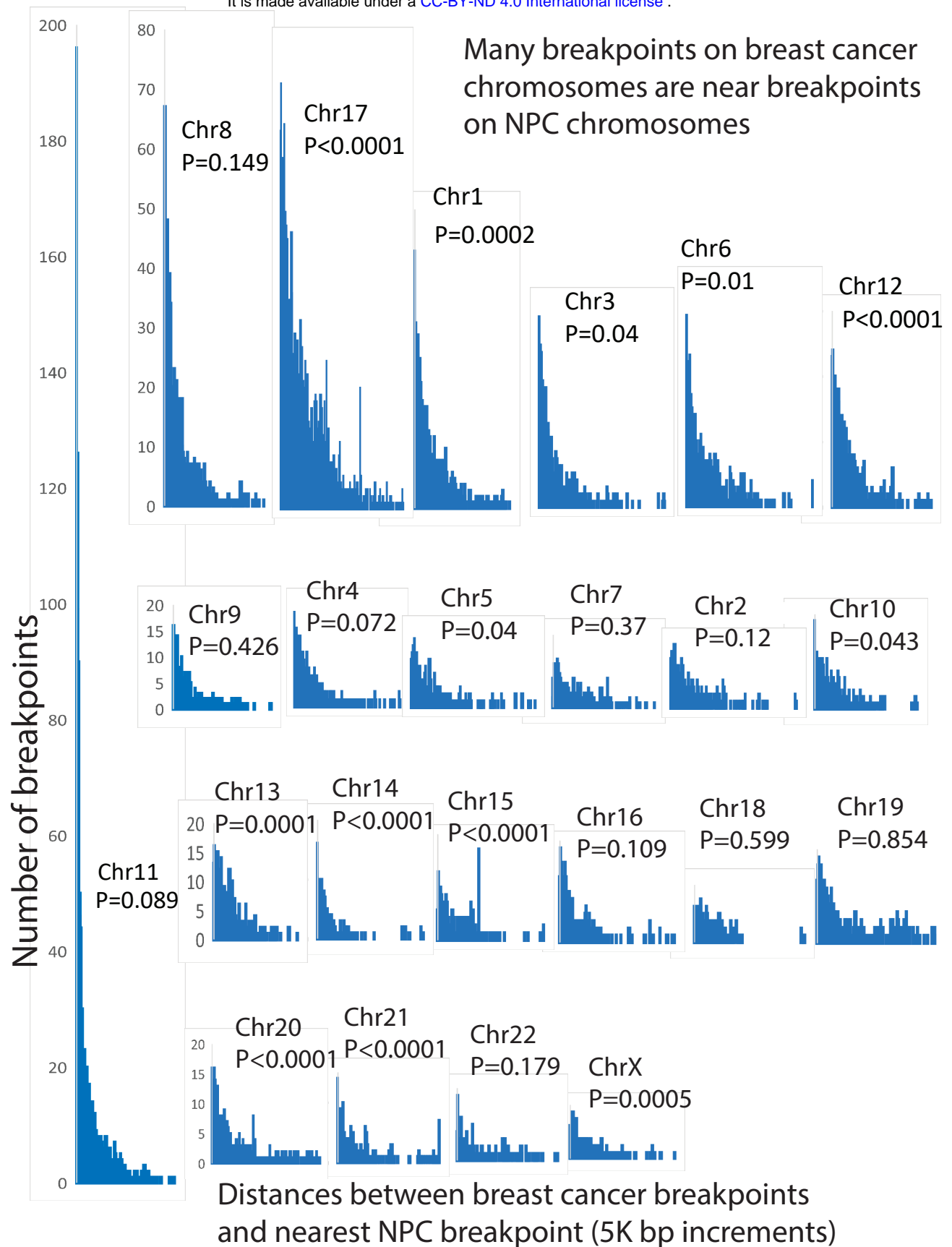
**Fig. 5A. Selected protein-coding genes on chromosome 3p. Transcripts essential for HR are colored red, and transcripts necessary for oxidative phosphorylation are colored green.**



**Fig. 5B. Deletions on chromosome 3p from five TCGA datasets include the pyruvate dehydrogenase catalytic subunit (PDHB).**



**Fig 5C. The short deletions on chromosome 3p in breast cancers listed in the COSMIC database focus on CTNNB1. This damage should deregulate PDHB.**



**Fig. 6A.** Distances separating the nearest breakpoints in breast cancer from breakpoints in the EBV cancer NPC. P values are from Mann-Whitney calculations and represent probabilities that the distributions are the same.  $P > 0.05$  indicates that the possibility that the breakpoint distributions are identical cannot be excluded.



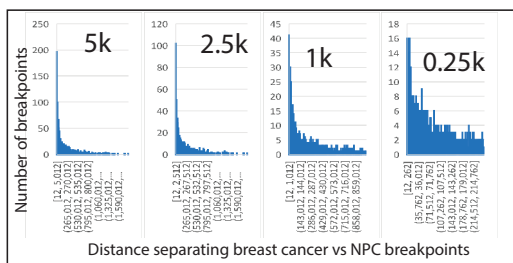
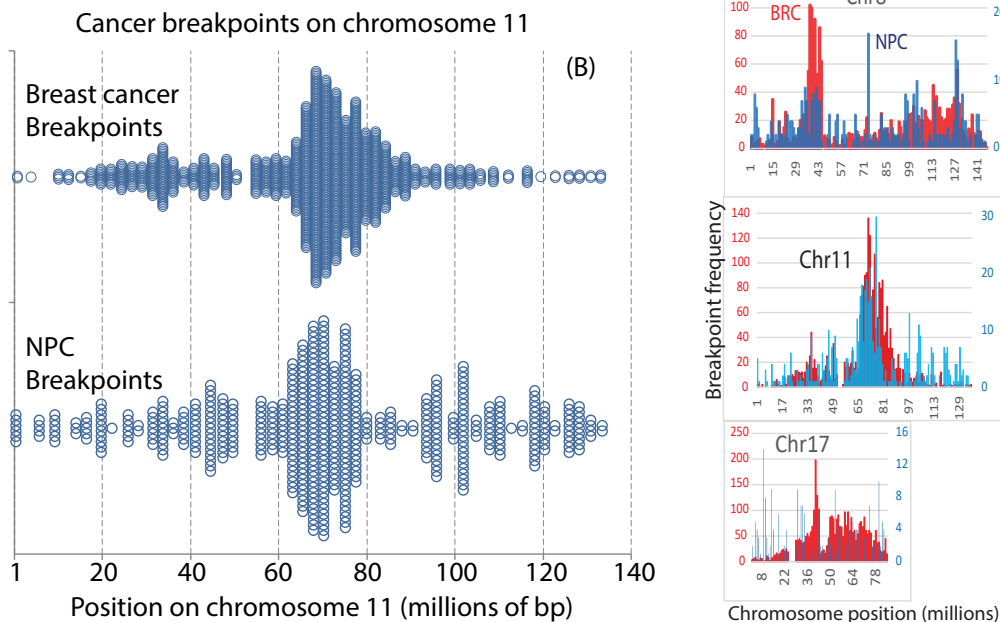
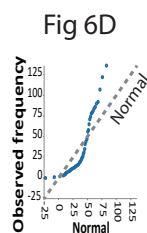


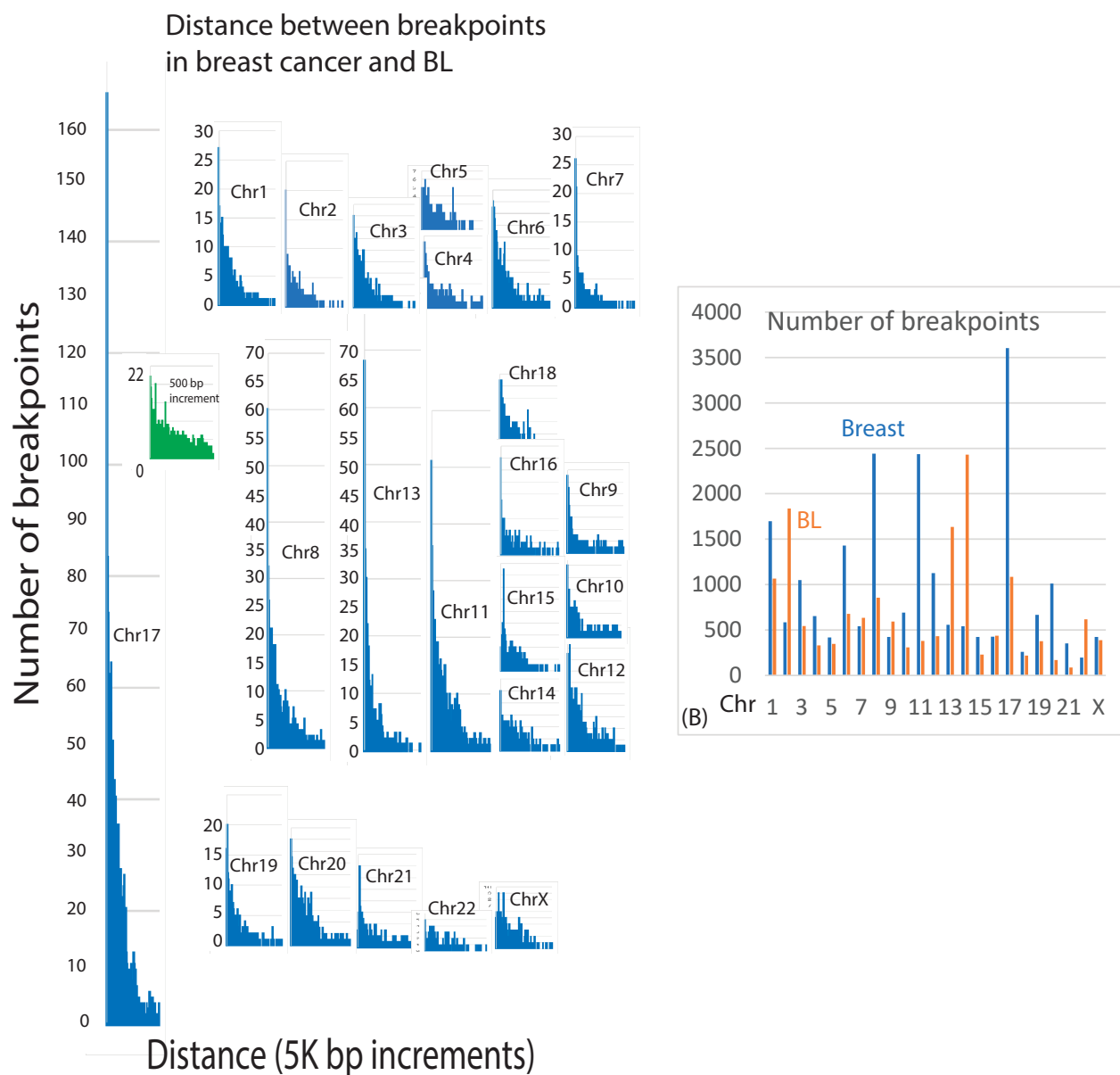
Fig 6C



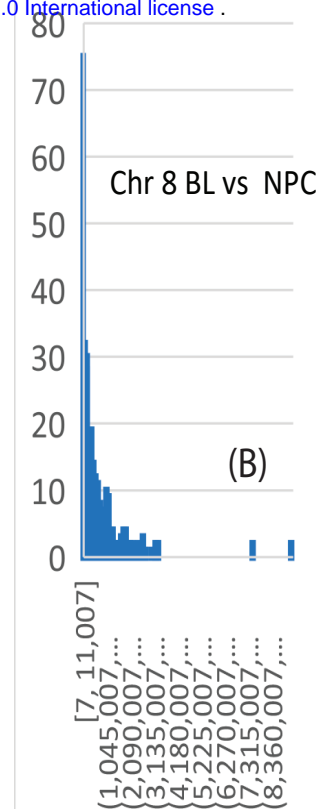
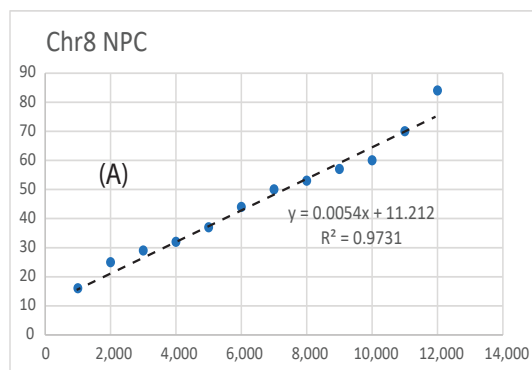
**Fig. 6B.** Plotting the number of breakpoints or the frequencies of break-points vs. chromosome positions in breast cancer and NPC illustrates chromosome positions where breakpoints are near each other.

**Fig. 6C.** The relationship between some NPC and breast cancer break-points on chromosome 11 is evident down to a window of 250 base pairs.

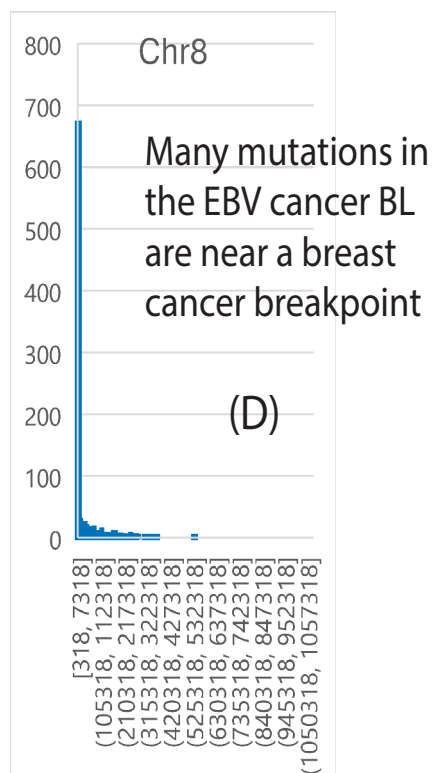
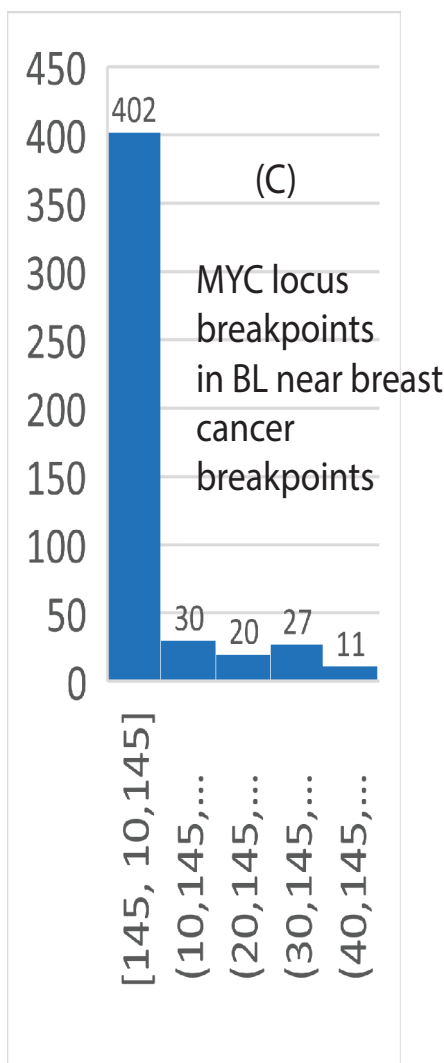
**Fig 6D.** Tests for normality of the distribution of nearest breakpoints show that the values are unlikely to be normally distributed.



**Fig. 7A.** Distances between breakpoints in breast cancers nearest those in EBV-positive BL. The proximity of breast cancer and BL breakpoints on chromosome 17 is still apparent down to a value of about 500 base pairs. **Fig. 7B.** Numbers of breakpoints for each chromosome in breast cancer vs. BL. The numbers are statistically related by polynomial regression analysis.



Chr8 breakpoints in BL are near those in NPC



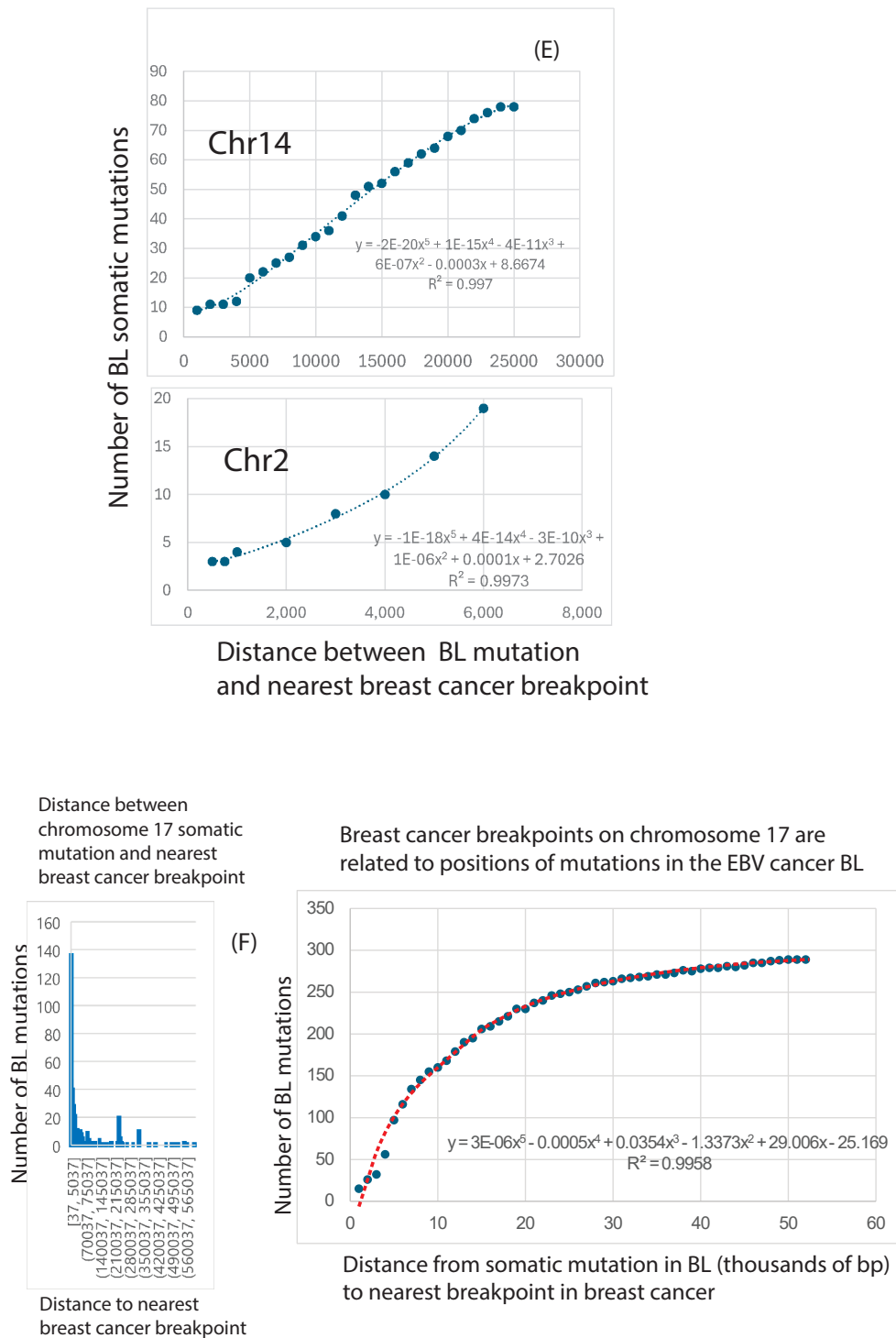
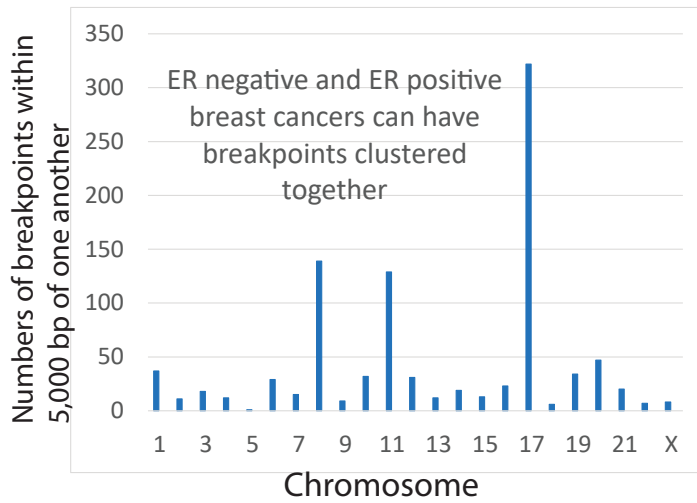
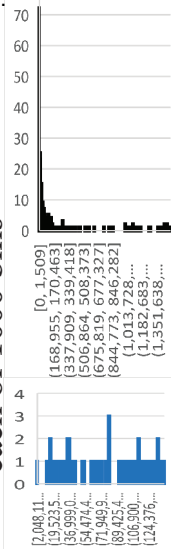


Fig 8A. Breakpoints on chromosome 8 in BL are near those in NPC. The numbers of breakpoints increase linearly as the window size increases up to about 11,000 bp. (B). About 75 breakpoints in NPC are near a BL breakpoint within the 11,000 bp window. (C). Various breast cancers had over 400 breakpoints near MYC locus breakpoints on chromosome 8 in BL. (D). Nearly 700 BL somatic mutations are within 7318 bp of a breast cancer breakpoint on chromosome 8. (E). Chromosome 14 and chromosome 2 translocation partners to chromosome 8 also show associations between AID induced SHMs in BL and breast cancer breakpoints. (F). Chromosome 17 has the largest number of breast cancer breakpoints related to BL. Nearly 140 BL mutations on chromosome 17 are within 5k bp of a breast cancer breakpoint. The overall distribution of BL mutations has a complex polynomial relationship to breast cancer breakpoints. This shows that the effect of somatic mutation on breast cancer breakpoints changes at different separation distances.

Numbers of breakpoints within 10,000 base pairs of each other

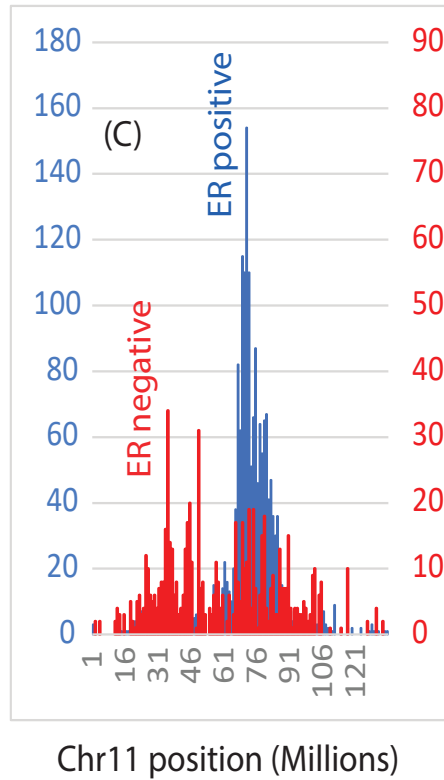
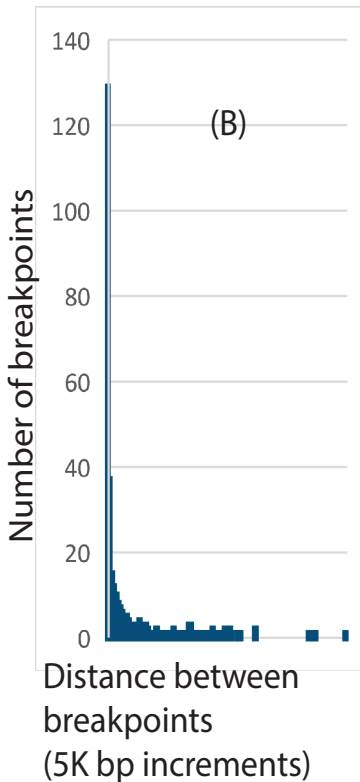


Numbers of breakpoints in each of 1000 bins

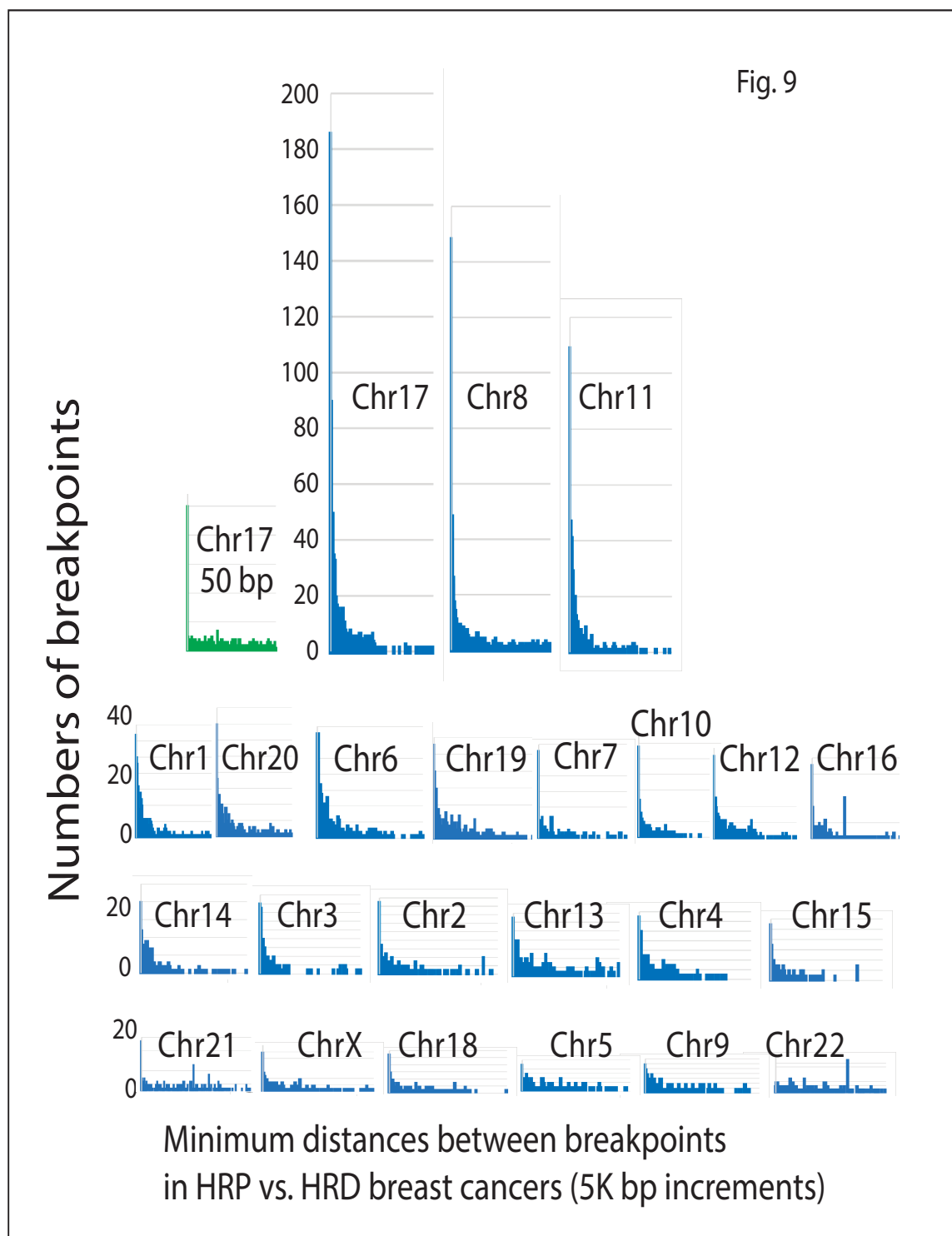


Breakpoints on chr8 in ER-neg vs ER-pos in 1000 bins

Breakpoints on chr8 in ER-neg vs Random in 1000 bins



**Fig. 9.** ER status in breast cancers is not sufficient to predict breakpoint positions because ER-negative breakpoints could be near ER-positive breakpoints. **Fig 9A.** Summary of the numbers of breakpoints within 10,000 base pairs of each other in ER-positive vs ER-negative breast cancers. At the right of Fig. 9A is a comparison of the breakpoint distributions between ER negatives vs. ER-positive samples (top) vs. a comparison between ER negative samples and a random distribution (bottom). Both comparisons are from chromosome 8 divided into 1000 bins. **Fig 9B.** About 130 breakpoints on chromosome 11 in ER-positive and ER-negative breast cancers are within 5K bp of each other. **Fig. 9C.** Comparisons of the frequency distributions of breakpoints in ER-positive vs ER-negative breast cancers on chromosome 11.



**Fig. 10. Comparisons of breakpoint distances between breast cancers that are HR proficient vs HR deficient (Increments of 5K bp). The first peak is the numbers of breakpoints within 5k bp and is the largest for all chromosomes except chromosome 22.**

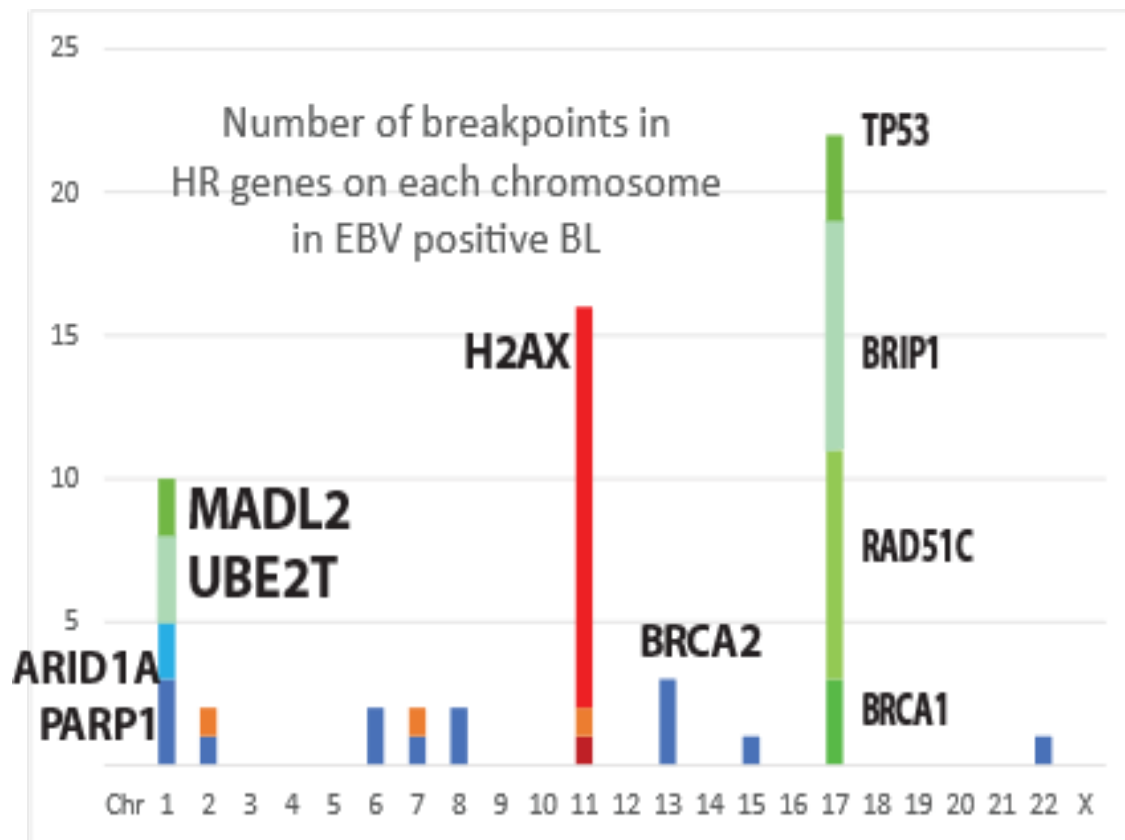
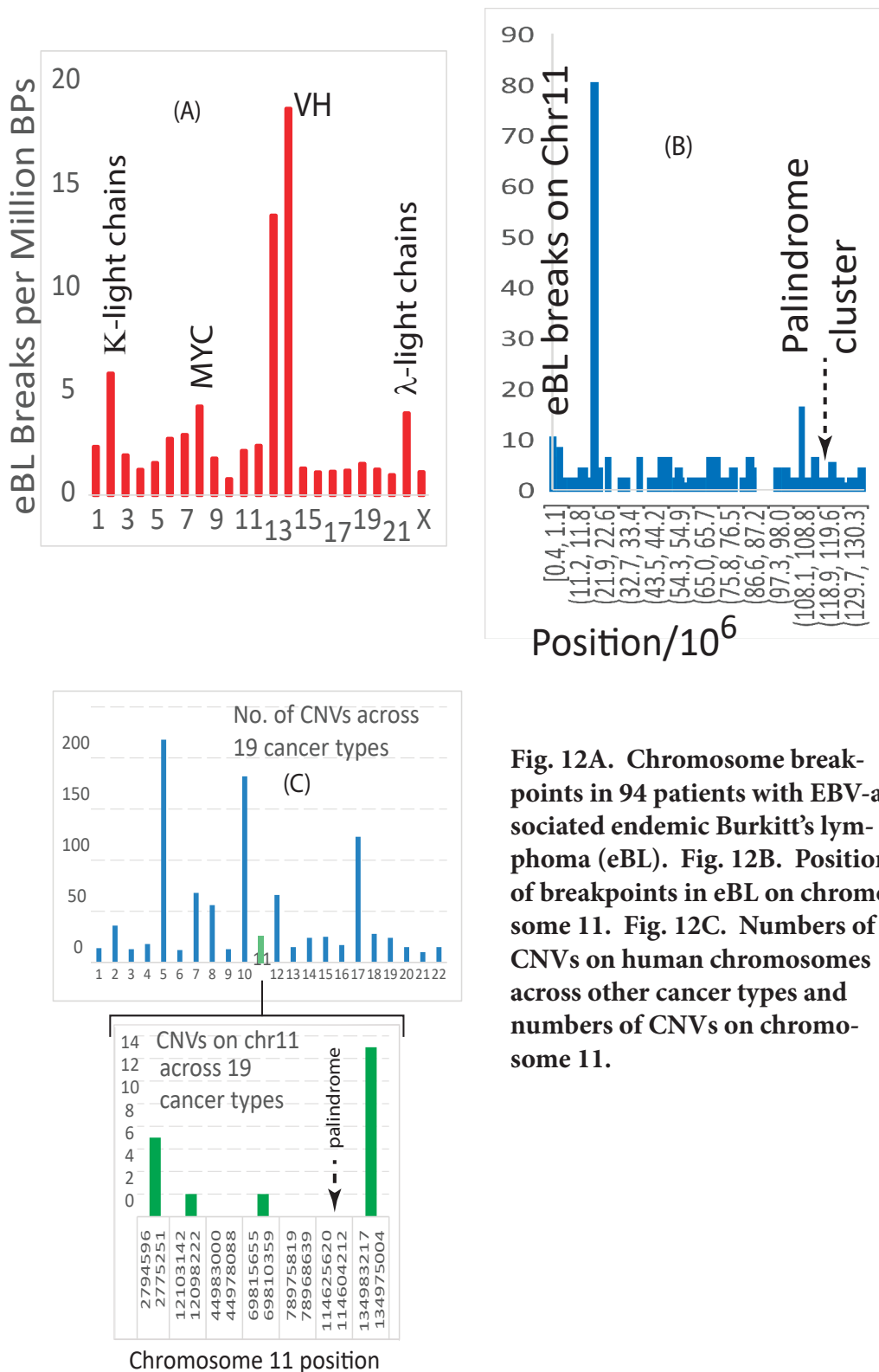
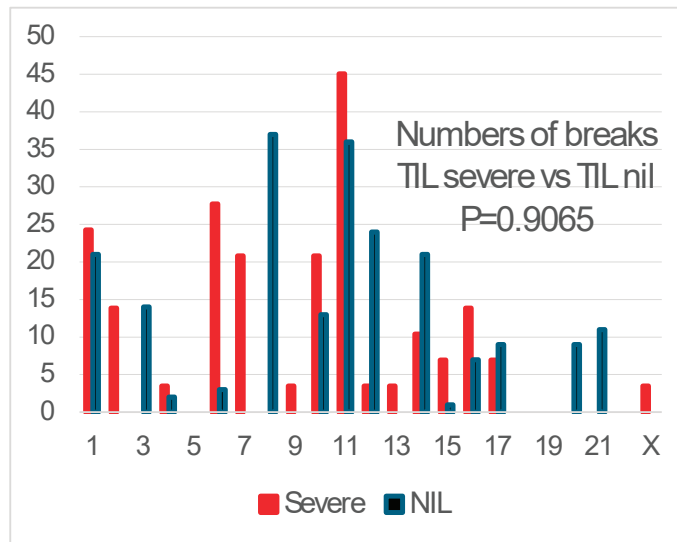


Fig 11. Breakpoints in EBV-positive BL affect genes essential for HR. A minimal list of these HR genes is compiled for each chromosome and indicated by different colored bars.



**Fig. 12A.** Chromosome break-points in 94 patients with EBV-associated endemic Burkitt's lymphoma (eBL). **Fig. 12B.** Positions of breakpoints in eBL on chromosome 11. **Fig. 12C.** Numbers of CNVs on human chromosomes across other cancer types and numbers of CNVs on chromosome 11.





**Fig.13. Numbers of breakpoints on chromosomes from breast cancers with severe tumor lymphocyte infiltrates vs. no lymphocyte infiltrates. The normalized values are likely from the same distribution (P=0.9065)**

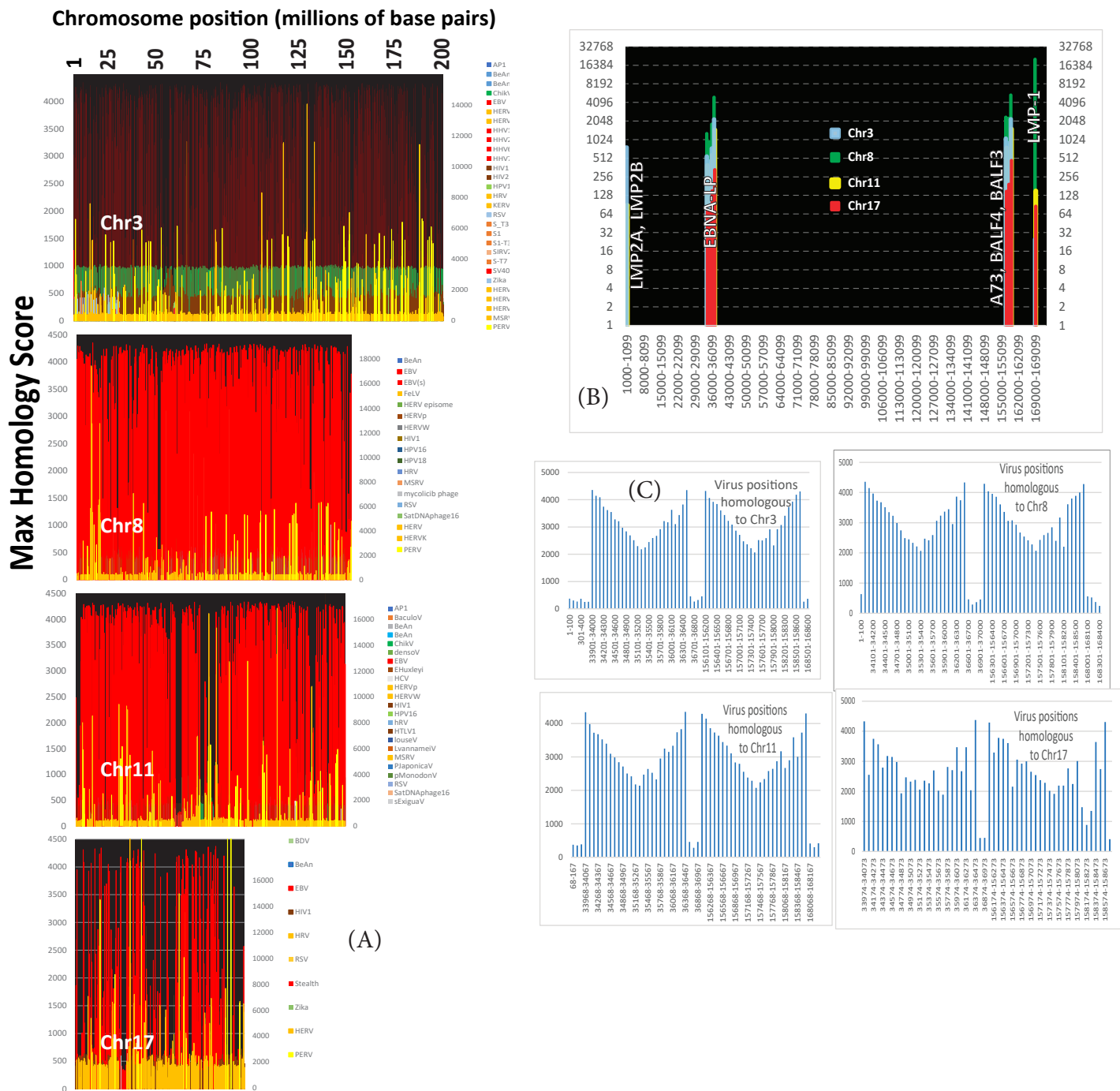
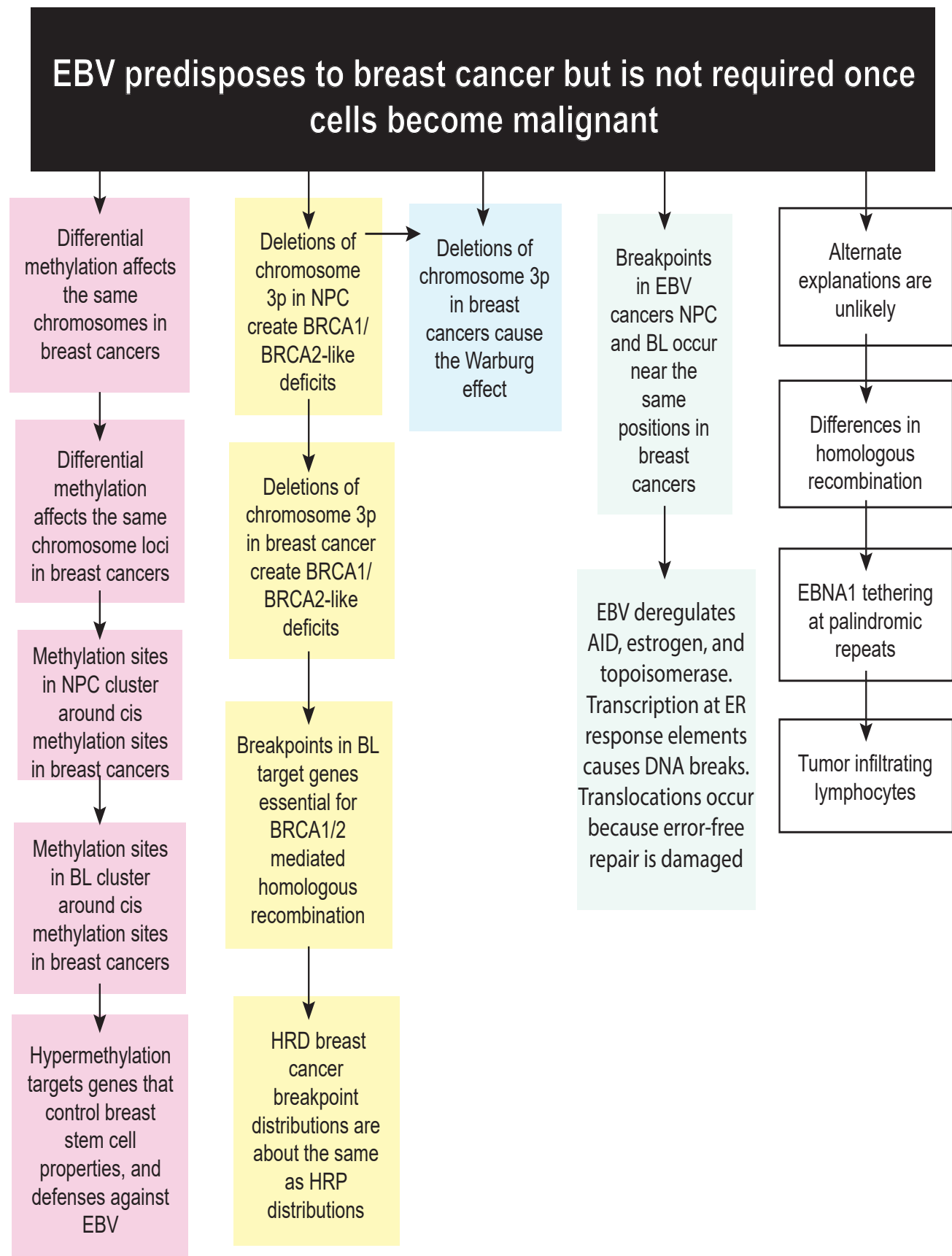
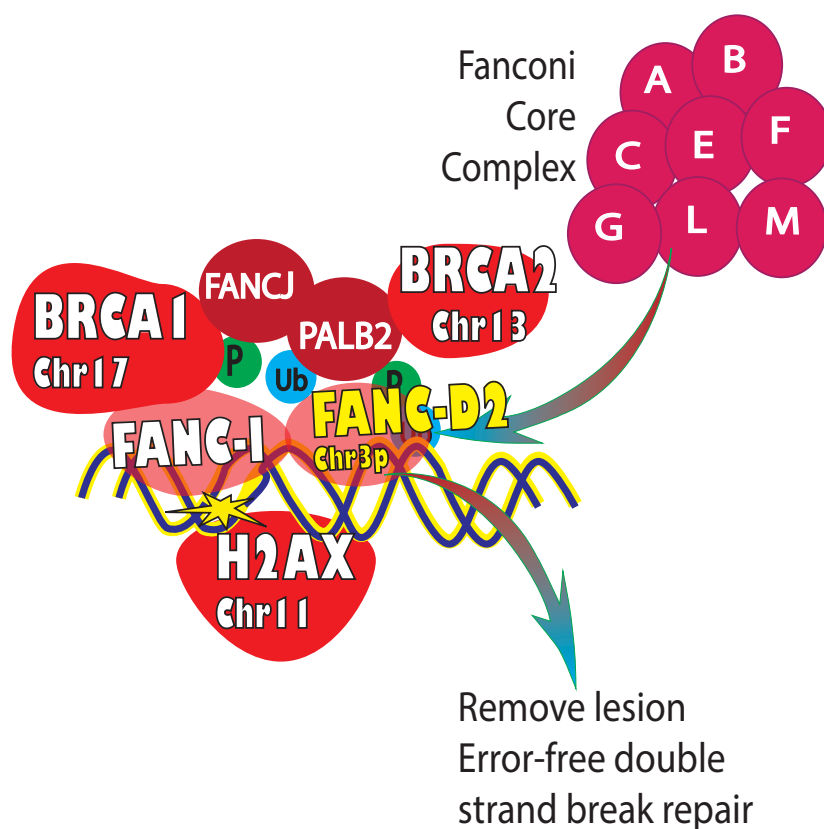


Fig 14A. Positions on chromosomes 3, 8, 11, and 17 with significant homology to a known virus. Fig 14B. Genes on the EBV chromosome that correspond to the positions of homology in Fig. 14A. Fig. 14C. Number of homologous sequences at each position in the EBV genome.



**Fig. 15. Five independent lines of evidence support the conclusion that EBV predisposes its hosts to breast cancer, but the virus is not required once malignant transformation occurs.**



Supplementary Fig. 1. Summary of selected gene product interactions in the pathway for error-free double strand break repair. FANCD2 is a critical intermediate encoded on chromosome 3p. The pathway is based on many publications of D'andrea et al. (e.g. references 30, 62)