https://doi.org/10.1038/d41586-026-00038-5

News & views

Genetics

DNA scraps from biobanks reveal virus's secrets

Kristen J. Wade & Jill A. Hollenbach

Viral DNA that is usually dismissed when sequencing the human genome could help to uncover useful information about complex diseases.

When whole-genome sequencing samples are collected for biobanks, any non-human DNA found in the sample is usually ignored. Writing in *Nature*, Nyeo *et al.*[1] describe an ingenious method that uses these discarded sequences to detect and quantify viral DNA that lingers in human cells.

Epstein–Barr virus (EBV) is a herpesvirus that infects more than 90% of the global population[2]. Some individuals develop an active, symptomatic infection leading to mononucleosis (also known as glandular fever), but many do not experience symptoms at all. EBV is a DNA virus that infects mainly white blood cells called B cells. The viral genome tethers itself to the host cell's chromatin (the packaged form of genomic DNA) without integrating into it directly, forming an external viral chromosome known as an episome[3]. The extent to which EBV genes in this episome are expressed determines whether the virus is in an active 'lytic' state, or a dormant 'latent' state.

In nearly all infected individuals, EBV persists in a latent state. This enables it to maintain a lifelong infection, which generally remains benign. However, in rare cases, EBV can induce cancers such as lymphomas and carcinomas[2], and it can trigger dysregulation of the immune system[4,5]. Surprisingly, for a virus that has infected most of the human population and can cause serious health complications[3], there is limited understanding of the factors that affect its ability to establish a persistent infection.

Making use of discarded DNA-sequence content from massive human biobanking efforts, Nyeo *et al.* quantify EBV DNA, revealing variants in the human genome that affect how EBV persists. They also identify associations between EBV viral load and a variety of human diseases. Importantly, this approach should be applicable to any DNA virus that infects humans.

Biobanks have been a key advance in realizing the goals of precision medicine. These large-scale databases store information about human populations, including biometric data that are relevant to health. This allows researchers to identify and study risk factors for a wide range of diseases. For each participant, DNA is generally collected from a blood sample and sequenced. Sequenced DNA is usually aligned to a human reference sequence, which enables the comparison of disease-associated genetic variation across the population.

Nyeo *et al.* draw on a crucial but often-overlooked feature of this sampling scheme: all of the DNA in an individual's white blood cells is collected and sequenced, and this includes any viral DNA that might be present. Using the EBV reference genome, the authors identified viral DNA sequences and extracted them along with genetic information about the virus's human host (Fig. 1a). By accessing DNA from two of the world's largest biobanks — the UK Biobank[6] (490,560 participants) and the All of Us biobank[7] (245,394 participants) — the authors generated a sample size that is 100 times larger than those of previous studies of EBV in human health, which relied on collecting and analysing blood samples.

Using this method, the authors identified

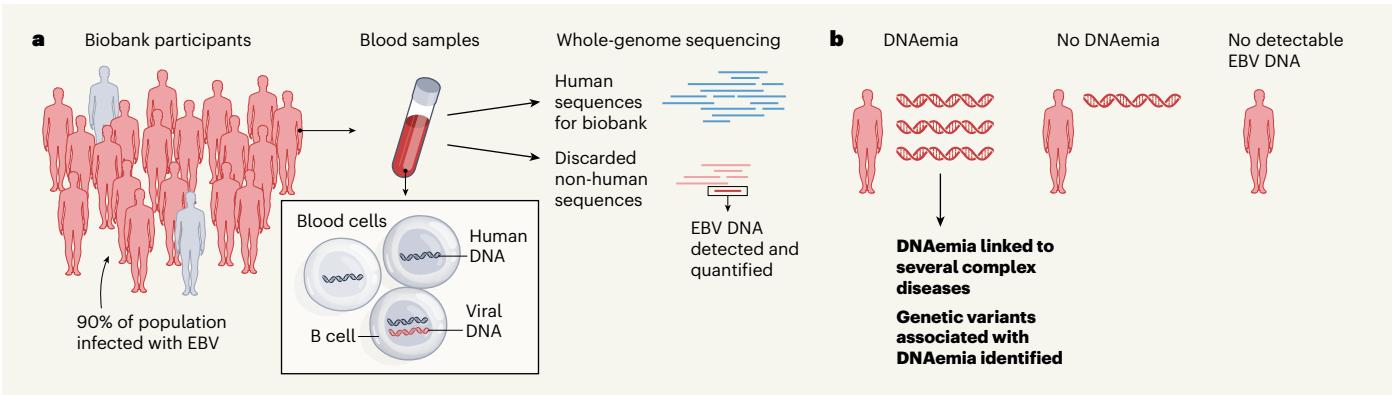

Figure 1 | Overlooked viral DNA in human biobanks. **a**, Epstein–Barr virus (EBV) infects around 90% of the human population, resulting in a lifelong infection owing to viral DNA persisting in B cells in the blood. Although usually benign, EBV infection is a risk factor for complex diseases, including cancer and autoimmune conditions. To examine EBV persistence in human populations, Nyeo *et al.*[1] accessed data from human biobanking efforts that involve sequencing DNA from blood samples. Any non-human sequencing reads are usually discarded, but the authors used these discarded reads to identify and quantify EBV DNA in the blood of thousands of individuals. **b**, The authors distinguished individuals whose blood contained large amounts of EBV DNA (DNAemia) from those with barely detectable levels of EBV DNA or no EBV DNA. DNAemia was then correlated with disease history, revealing associations between EBV persistence and a number of complex conditions. Genetic information from individuals with DNAemia was also evaluated to identify risk variants associated with EBV persistence.

associations between having a high load of EBV DNA in the blood (which the authors refer to as 'DNAemia') and complex diseases such as Hodgkin's lymphoma, rheumatoid arthritis, chronic pulmonary disease, systemic lupus erythematosus and fatigue. The associations suggest that EBV infection can cause systemic dysregulation in humans. Furthermore, many of these diseases have been linked previously to EBV through indirect evidence, so this work is crucial in confirming those links.

The authors also describe new genetic risk factors — in both human and viral genomes — that can be used to understand how EBV drives the emergence of disease. For example, they identified specific variants in 'HLA' genes as the main human genetic risk markers for EBV DNAemia. HLA genes encode a set of proteins that are essential for mediating adaptive immune responses to infection.

The authors also discovered four variants in the EBV genome that were enriched in individuals with a type of cancer called nasopharyngeal carcinoma. Using computational methods, the authors predicted that these EBV variants are unlikely to affect the host's ability to target the virus, but might instead be involved in the viral life cycle. This provides insights into how EBV infects people, and why some EBV infections lead to poor outcomes whereas others remain benign.

Nyeo and colleagues' work opens the door to understanding how human genetic variation affects viral persistence. By linking EBV DNA load with disease status across large human populations, the authors reveal human and viral genetic risk factors for numerous diseases (Fig. 1b). These results provide much-needed clues to understanding disease mechanisms and to begin identifying treatments that block EBV infection from persisting[8,9].

Importantly, the authors demonstrate that EBV DNAemia strongly correlates with EBV serostatus (an indication of whether a person has made antibodies against the virus), which is a well-established clinical metric of infection. This supports using EBV DNAemia as a biomarker when serostatus data might not be available. Furthermore, the authors have created a reproducible computational method that can be used to improve understanding of other DNA viruses, such as the *Anelloviridae* and *Polyomaviridae* families, which similarly establish life-long infections in humans but remain understudied.

Interpreting viral load from DNA quantification is an essential step forwards in genomic methods for studying EBV persistence. However, the EBV life cycle is dynamic, and the DNA quantity at a single time point cannot provide a complete picture of it. For instance, Nyeo and colleagues' results provide information only about individuals who had a high level of EBV DNA at the time of sampling. More than 90% of humans are expected to be infected, but substantial EBV DNAemia was detected in only around 10% of the authors' study sample. Although Nyeo *et al.* demonstrate that it is unlikely that EBV DNAemia is attributable to lytic activation of the virus, these data cannot fully describe the latent state either. This means that the authors' findings relate to a subset of infected individuals, and imply that the traits they uncovered are relevant only to individuals who exhibit high levels of EBV DNA.

This might explain why the authors did not find an association with multiple sclerosis, an autoimmune disease for which EBV infection is a known risk factor[4]. Such a result suggests that the role of EBV in driving the risk of this disease might not be linked to the quantity of EBV DNA in the cell. It is likely that there are other EBV-associated diseases that work through similar alternative mechanisms. Therefore, more metrics should be developed and applied in tandem with this approach to fully characterize the relationship of EBV infection with outcomes in the host. Ultimately, this study provides a necessary and reproducible method with which to begin tackling these lines of enquiry.

Kristen J. Wade and **Jill A. Hollenbach** are in the Department of Neurology, University of California, San Francisco, San Francisco, California 94143, USA. J.A.H. is also in the Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California 94158, USA.
e-mail: jill.hollenbach@ucsf.edu

1. Nyeo, S. S. *et al. Nature* https://doi.org/10.1038/s41586-025-10020-2 (2026).
2. Meirhaeghe, M. R. & Balfour, H. H. Jr. *J. Clin. Virol.* **180**, 105854 (2025).
3. Damania, B., Kenney, S. C. & Raab-Traub, N. *Cell* **185**, 3652–3670 (2022).
4. Bjornevik, K. *et al. Science* **375**, 296–301 (2022).
5. Robinson, W. H., Younis, S., Love, Z. Z., Steinman, L. & Lanz, T. V. *Nature Rev. Rheumatol.* **20**, 729–740 (2024).
6. The UK Biobank Whole-Genome Sequencing Consortium. *Nature* **645**, 692–701 (2025).
7. The All of Us Research Program Genomics Investigators. *Nature* **627**, 340–346 (2024).
8. Soldan, S. S. *et al. Nature Microbiol.* **9**, 1540–1554 (2024).
9. Guo, R. *et al. Mol. Cell* **78**, 653–669 (2020).

The authors declare no competing interests.