Table S1. Nominally significant (p<0.05) associations between viral read counts and AD

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Virus | Dataset | Odds Ratio | p-value | Adj p-value | Direction |
| Human alphaherpesvirus 1 (HSV-1) | **Meta Analysis** | **3.69** | **6.71 x10-5** | **6.71 x10-4** | **++--** |
| WES Blood | 4.10 | 3.58x10-5 | 3.58x10-4 | + |
| WGS Blood | 0.52 | 0.67 | 1.00 | - |
| WES Brain | 4.85 | 0.44 | 1.00 | + |
| WGS Brain | 4.48x10-4 | 0.43 | 1.00 | - |
| Human papillomavirus 71 (HPV-71) | **Meta Analysis** | **3.55** | **3.41x10-3** | **0.03** | **+---** |
| WES Blood | 3.90 | 1.97x10-3 | 0.02 | + |
| WGS Blood | 1.90x10-101 | 0.92 | 1.00 | - |
| WES Brain | 0.16 | 0.47 | 1.00 | - |
| WGS Brain | 0.50 | 1.00 | 1.00 | - |
| Torque teno virus 10 (TTV-10) | **Meta Analysis** | **30.33** | **1.07x10-3** | **0.06** | **+?+?** |
| WES Blood | 22.20 | 3.59x10-3 | 0.21 | + |
| WGS Blood | 8.67x104 | 0.03 | 1.00 | + |
| WES Brain | — | — | — | — |
| WGS Brain | — | — | — | — |

**Table S2**. Sample demographics of the AD cases and controls by tissue source and experiment type

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset Stratum | | | Controls | | | | | AD Cases | | | | | |
|  | %  APOE ε4 | Mean Age | % Female | MeanViral Read Count | N | %  APOE ε4 | Mean Age | % Female | Mean Viral ReadCount |
| WES | Ancestry | African American | 1793 | 31.6 | 75.6 | 72.4 | 6.4 | 1038 | 55.5 | 76.2 | 70.1 | 6.7 |
| Caribbean Hispanic | 949 | 20.1 | 80.1 | 67.3 | 4.8 | 482 | 32.4 | 81.1 | 69.7 | 4.8 |
| European Descent | 2228 | 16.7 | 83.5 | 57.7 | 4.9 | 2036 | 39.8 | 76.7 | 57.8 | 4.6 |
| Body Tissue Source | Blood | 4846 | 22.9 | 79.9 | 65.1 | 5.4 | 3132 | 42.6 | 77.6 | 64.4 | 5.3 |
| Brain | 124 | 13.7 | 84.4 | 54.0 | 4.4 | 424 | 49.1 | 73.9 | 52.8 | 4.6 |
| WGS | Ancestry | African American | 1354 | 30.1 | 74.8 | 74.8 | 12.4 | 893 | 59.4 | 74.6 | 70.1 | 10.4 |
| Caribbean Hispanic | 1374 | 18.6 | 75.7 | 69.4 | 18.8 | 327 | 34.6 | 76.8 | 64.2 | 13.0 |
| European Descent | 1459 | 31.6 | 80.5 | 60.2 | 8.1 | 2351 | 51.7 | 75.2 | 56.8 | 9.7 |
| Body Tissue Source | Blood | 3929 | 27.2 | 76.7 | 68.4 | 13.4 | 2681 | 49.8 | 73.9 | 59.3 | 11.3 |
| Brain | 258 | 15.9 | 83.6 | 60.9 | 6.9 | 890 | 58.9 | 79.1 | 65.4 | 6.7 |

A picture containing text, writing implement, stationary, pencil

Description automatically generated

**Figure S1. Viral load detected by sequencing center, tissue source, and ethnicity**. Distribution of viral reads as grouped boxplots for (a) blood samples in WES, (b) brain samples in WGS, (c) blood samples in WGS, (d) brain samples in WGS. Each subplot is first grouped by ancestry and then within each of the three ancestry groups is color-coded by sequencing center. The two boxplots of WES share the same y-axis and the two boxplots in WGS share the same y-axis. The y-scale is normalized by the total number of reads with each mark equal to one log unit.

Graphical user interface, application

Description automatically generated

**Figure S2. Top predictors of AD among African Americans in (a) whole exome and (b) whole genome sequence data**. Each feature within each subset is assigned a score created by summing the accuracy of the ML prediction model in which it improved the prediction of AD. The top 15 features are shown in each bar chart though several other viruses improved the prediction models.Graphical user interface, application

Description automatically generated

**Figure S3. Top predictors of AD among Caribbean Hispanics in (a) whole exome and (b) whole genome sequence data.** Each feature within each subset is assigned a score created by summing the accuracy of the ML prediction model in which it improved the prediction of AD. The top 15 features are shown in each bar chart though several other viruses improved the prediction models.

Graphical user interface, chart

Description automatically generated

**Figure S4. Top viral predictors of AD among persons of European ancestry in (a) whole exome and (b) whole genome sequence data**. Each feature within each subset is assigned a score created by summing the accuracy of the ML prediction model in which it improved the prediction of AD. The top 15 features are shown in each bar chart though several other viruses improved the prediction models.