Article

Neuron

Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus

Highlights

- Common viral species frequently detected in normal, aging brain
- Increased HHV-6A and HHV-7 in brains of subjects with Alzheimer's disease (AD)
- Findings were replicated in two additional, independent cohorts
- Multiscale networks reveal viral regulation of AD risk, and APP processing genes

Authors

Ben Readhead, Jean-Vianney Haure-Mirande, Cory C. Funk, ..., Michelle E. Ehrlich, Sam Gandy, Joel T. Dudley

Correspondence

joel.dudley@mssm.edu

In Brief

Readhead et al. construct multiscale networks of the late-onset Alzheimer's disease (AD)-associated virome and observe pathogenic regulation of molecular, clinical, and neuropathological networks by several common viruses, particularly human herpesvirus 6A and human herpesvirus 7.

Readhead et al., 2018, Neuron *99*, 1–19
July 11, 2018
https://doi.org/10.1016/j.neuron.2018.05.023

CellPress

Please cite this article in press as: Readhead et al., Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus, Neuron (2018), https://doi.org/10.1016/j.neuron.2018.05.023

Neuron
Article

Multiscale Analysis of Independent Alzheimer's Cohorts Finds Disruption of Molecular, Genetic, and Clinical Networks by Human Herpesvirus

Ben Readhead,[1,2,3,4,17] Jean-Vianney Haure-Mirande,[5,17] Cory C. Funk,[6] Matthew A. Richards,[6] Paul Shannon,[6] Vahram Haroutunian,[7,8] Mary Sano,[8,15] Winnie S. Liang,[9,10] Noam D. Beckmann,[1,2] Nathan D. Price,[6] Eric M. Reiman,[9,10,11,12] Eric E. Schadt,[1,2,13] Michelle E. Ehrlich,[1,2,5,14] Sam Gandy,[5,8,15,16,17] and Joel T. Dudley[1,2,3,4,17,18,*]

[1]Departments of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[2]Icahn Institute of Genomic Sciences and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[3]Institute for Next Generation Healthcare, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[4]ASU-Banner Neurodegenerative Disease Research Center, Arizona State University, Tempe, AZ 85287-5001, USA
[5]Department of Neurology, Alzheimer's Disease Research Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[6]Institute for Systems Biology, Seattle, WA, 98109-5263, USA
[7]Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[8]James J. Peters VA Medical Center, 130 West Kingsbridge Road, New York, NY 10468, USA
[9]Arizona Alzheimer's Consortium, Phoenix, AZ 85014, USA
[10]Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ 85004, USA
[11]Department of Psychiatry, University of Arizona, Phoenix, AZ 85721, USA
[12]Banner Alzheimer's Institute, Phoenix, AZ 85006, USA
[13]Sema4, Stamford, CT 06902, USA
[14]Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[15]Department of Psychiatry, Alzheimer's Disease Research Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[16]Center for NFL Neurological Care, Department of Neurology, New York, NY 10029, USA
[17]These authors contributed equally
[18]Lead Contact
*Correspondence: joel.dudley@mssm.edu
https://doi.org/10.1016/j.neuron.2018.05.023

SUMMARY

Investigators have long suspected that pathogenic microbes might contribute to the onset and progression of Alzheimer's disease (AD) although definitive evidence has not been presented. Whether such findings represent a causal contribution, or reflect opportunistic passengers of neurodegeneration, is also difficult to resolve. We constructed multiscale networks of the late-onset AD-associated virome, integrating genomic, transcriptomic, proteomic, and histopathological data across four brain regions from human post-mortem tissue. We observed increased human herpesvirus 6A (HHV-6A) and human herpesvirus 7 (HHV-7) from subjects with AD compared with controls. These results were replicated in two additional, independent and geographically dispersed cohorts. We observed regulatory relationships linking viral abundance and modulators of APP metabolism, including induction of *APBB2*, *APPBP2*, *BIN1*, *BACE1*, *CLU*, *PICALM*, and *PSEN1* by HHV-6A. This study elucidates networks linking molecular, clinical, and neuropathological features with viral activity and is consistent with viral activity constituting a general feature of AD.

INTRODUCTION

Important roles for microbes and antimicrobial defenses in the pathogenesis of Alzheimer's disease (AD) have been postulated or evaluated for at least six decades, beginning with Sjögren in 1952 (Sjogren et al., 1952). "Slow virus" was one of the early names used for the illness that eventually came to be known as prion disease, referring to the hypothesis that conventional viruses might be capable of acting to cause not only acute encephalitis, but also a progressive neuronal destruction process that might engender less inflammation because of its slowly progressive nature (Sigurðsson, 1954). Measles (MV) is a conventional virus that can act through acute inflammatory and slow neurodegenerative processes, occasionally re-emerging as a fatal brain disease known as subacute sclerosing panencephalitis (SSPE) up to a decade after a typical acute MV infection (Murphy and Yunis, 1976).

Beginning with Crapper McLachlan in 1980 (Middleton et al., 1980), several investigators have proposed that AD is an SSPE-like illness, caused by a slow virus form of herpes simplex (Itzhaki, 2014). Hundreds of reports have associated AD with diverse bacterial and viral pathogens (Itzhaki et al., 2016; Mastroeni et al., 2018), most frequently implicating *Herpesviridae* (particularly HSV-1 [Lövheim et al., 2015a, 2015b], EBV, HCMV, and HHV-6 [Westman et al., 2017; Carbone et al., 2014]). The results of these studies, taken in aggregate, are

suggestive of a viral contribution to AD, though findings offer little insight into potential mechanisms, and a consistent association with specific viral species has not emerged.

Recent reports demonstrate that diverse classes of microbes can stimulate amyloid-beta (Aβ) aggregation and deposition as part of an intra-CNS anti-microbial innate immune response whereby the amyloidosis triggered by various microbes results in the coating of the infectious particles by the growing amyloid aggregate (Soscia et al., 2010; Kumar et al., 2016). These microbes coated with aggregated Aβ become unable to interact with cell surfaces, thereby arresting the infectious process.

We designed this study to map and compare biological networks underlying two distinct AD-associated phenotypes using multiple independent datasets collected from human subjects. We began with a computational network characterization of a specific endophenotype of AD: brains meeting neuropathological criteria for AD from individuals who were cognitively intact at the time of death (Liang et al., 2010), which we refer to here as "preclinical AD" (Sperling et al., 2011). We presumed that a network model of preclinical AD (and its comparison with networks built from cognitively intact persons without neuropathology) might provide novel insights into the molecular context of neuropathology in the absence of clinical symptoms. Alternatively, since these individuals had eluded cognitive decline despite significant AD pathology, we reasoned that this might illuminate protective or resilience mechanisms. Functional genomic analysis of preclinical AD network alterations revealed multiple lines of evidence consistent with viral activity. We then directly evaluated viral activity in a multiscale network analysis of four large, multi-omic datasets (comprising samples from individuals with "clinical AD" as well as neuropathologically and cognitively normal controls) that included next generation sequencing data, enabling direct examination of viral DNA and RNA sequences.

This study presents novel evidence linking the activity of specific viruses with AD. This has been enabled by comprehensive molecular profiling of large patient cohorts, facilitating the integration of diverse biomedical data types into an expansive view spanning multiple disease stages, brain regions, and -omic domains. This has also allowed us to direct our analysis in an entirely data-driven manner and benefit from a form of data capable of implicating specific viral species. Our results offer evidence of complex viral activity in the aging brain, including changes specific to AD, particularly implicating *Herpesviridae*, HHV-6, and HHV-7. Taken together, these data provide compelling evidence that specific viral species contribute to the development of neuropathology and AD.

RESULTS

Preclinical AD Networks Indicate Early Changes in G-Quadruplex and C2H2 Zinc Finger Activity

We constructed, mapped, and compared differences between distinct gene regulatory networks to investigate functional molecular changes underlying the etiology of AD (Figure 1). We used laser-captured neuronal gene expression data to construct probabilistic causal networks (PCN) representing preclinical AD and also healthy control (individuals without cognitive impairment or neuropathology) (Liang et al., 2007, 2008) states. We focused on samples derived from brain regions associated with the most profound neuronal loss: the entorhinal cortex (Gómez-Isla et al., 1996) (EC) and the hippocampus (Hyman et al., 1984) (HIP). We built each PCN using a modified "inductive causation with latent variables" procedure (Pearl, 2009), constructing separate preclinical AD and control (CON) networks including paired samples from EC and HIP for each donor, and for each network, nominated genes that regulate the expression of unexpectedly large subnetworks as "network drivers" (see STAR Methods and Table S1). Edge reproducibility in PCNs has been demonstrated to be strongly dependent on sample size; however, detection of highly connected nodes is more robust across a range of sample sizes (Cohain et al., 2017). We therefore focused initially on characterizing the set of drivers present *only* in the CON network ("Lost in preclinical AD") and those present *only* in the preclinical AD network ("Gained in preclinical AD") as a means to prioritize differences between preclinical AD and CON states. We found that promoters of both sets of drivers are strongly enriched (compared with the rest of the network) for a shared set of C2H2 zinc finger transcription factor (C2H2-TF) binding motifs (Figure 1E), especially SP1, MAZ, NRF1, and EGR1. This prompted us to evaluate other co-regulatory features associated with C2H2-TF activity that might explain a general shift in their collective activity. We found that G-quadruplex (G4) sequences are strongly enriched among the promoters of both sets of drivers in the network, but that the "Lost in preclinical AD" drivers have significantly more G4 motifs within their introns, exons, and 3′-UTR on both coding and non-coding strands (Figures 1F and 1G). We concordantly found a strong negative correlation between gene G4 density (G4 motif count, normalized by gene length), and gene expression in the EC in preclinical AD and clinical AD samples (Figure 1H). Given the complex roles for G4 in dynamically regulating mRNA transcription, stability, translation, and localization (Rhodes and Lipps, 2015), we hypothesized that a global shift in G4 regulation or stability could explain the differences in C2H2-TF regulatory programs—for example, changes in expression and network influence of genes with especially high G4 density in locations like the 3′-UTR that are associated with alternative polyadenylation and miRNA-mediated regulation (Beaudoin and Perreault, 2013).

Functional Analysis of Network Patterns Suggests Roles for Viral Mediators in AD

Identification of strong C2H2-TFs and G4 functional patterns in the differential network analysis suggested a potential role for virus-mediated network activities in AD (Figure 1D, see Table S1). Enrichments among the network drivers (gained *and* lost) in preclinical AD implicated viral infection susceptibility risk genes, as well as host differential gene expression changes associated with viral infection. We also noted findings around C2H2-TF and G4 sequences that are implicated in a range of proviral and antiviral contexts, including SP1: (1) binding with Epstein-Barr Virus (EBV) protein Rta to regulate host and viral SP1 target genes (Chang et al., 2005); (2) regulating Human Immunodeficiency Virus 2 (HIV-2) LTR transcription (Harrich et al., 1989); and (3) mediating antiviral effects against Human

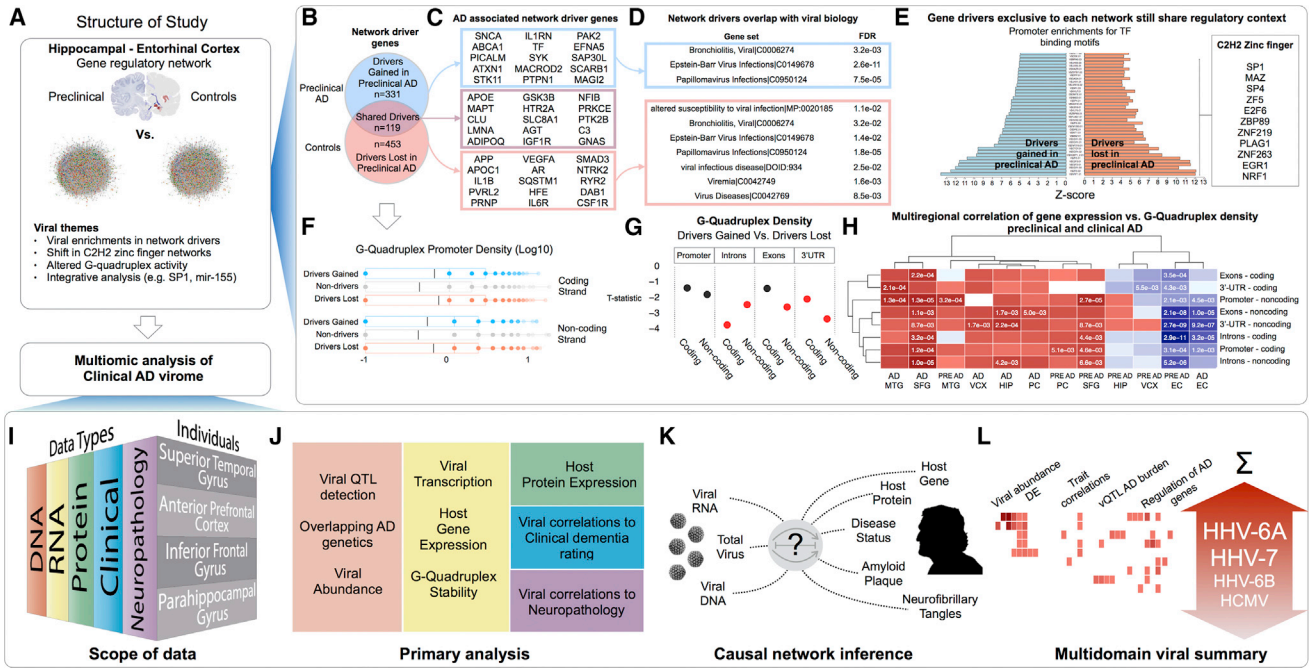


Figure 1. Multiomic Evaluation of Alzheimer's Disease-Associated Virome

(A) This study was performed in two phases, an initial exploration into the earliest network changes associated with preclinical AD, which identified multiple shifts in network biology consistent with viral perturbations and prompted a systematic, multiomic evaluation of viral biology on larger datasets focused on clinical AD.

(B–E) Regulatory networks built across EC and HIP samples showed differences in gene drivers for preclinical AD versus control networks (B) including many AD associated genes (C), although drivers exclusive to each network shared functional characteristics, such as association with viral biology (D), and promoter enrichments for the same C2H2 zinc finger transcription factors binding motifs (E).

(F and G) G-quadruplex sequence motifs were strongly enriched in the promoters of the drivers "lost in preclinical AD," and those "gained in preclinical AD" (F); however, drivers "lost in preclinical AD" had much higher G-Quadruplex density in introns, exons, and 3′-UTR (G).

(H) Genes with high G-Quadruplex density in these locations were uniformly downregulated in the EC of preclinical AD and AD samples, suggesting some significant alteration of G-Quadruplex regulation.

(I–L) To investigate the possibility of an AD-associated virome, we incorporated clinical, neuropathological, RNA, DNA, and proteomic data from individuals with clinical AD (and controls) (I) across four brain regions to identify and characterize viral activity associated with AD biology (J and K), highlighting *Roseoloviruses* HHV-6A and HHV7 (L).

Region abbreviations: EC, entorhinal cortex; HIP, hippocampus; MTG, medial temporal gyrus; PC, posterior cingulate cortex; SFG, superior frontal gyrus; VCX, visual cortex.

Cytomegalovirus (HCMV) (Scholz et al., 2004). In addition, G4 sequences are recognized as important regulatory features for viral pathogens such as EBV and HSV-1(Artusi et al., 2016), dynamically impacting translation of viral proteins (Murat et al., 2014).

Patterns of miRNA target enrichments identified by the differential network analysis of preclinical AD versus CON, as well as multiregional differential gene expression changes in clinical AD samples, offered an additional line of evidence for virally mediated network activity. We looked for significant overlap between experimentally validated gene targets of human miRNAs (Hsu et al., 2011) and multiregional differential gene expression signatures from preclinical AD samples (Liang et al., 2010), clinical AD samples (Liang et al., 2008), as well as "Lost in preclinical AD" and "Gained in preclinical AD" drivers. This identified a number of miRNA with gene networks overlapping many of these AD contexts, particularly miR-155, a multifunctional miR with associations to malignancy, innate immunity, and DNA virus activity. Interactions between miR-155 and viral biology are well established, including perturbation of miR-155 by EBV to stabilize viral latency (Gatto et al., 2008), inhibition of miR-155 by HHV-6A, (Caselli et al., 2017) and coding of a miR-155 functional ortholog by Kaposi's sarcoma-associated herpesvirus (KSHV) (Gottwein et al., 2007) and Marek's Disease Virus (Zhao et al., 2009). Considering these findings, we sought to directly evaluate viral DNA and RNA sequences in the context of clinical AD. We performed this investigation using four large, independent multiomic datasets from individuals with clinical AD as well as neuropathologically and cognitively normal controls.

Human Herpesvirus 6A and 7 Are More Abundant in Alzheimer's Disease in Multiple Brain Regions across Three Independent Clinical Cohorts

To evaluate differential viral abundance in AD, we initially performed RNA-seq on samples from a cohort of AD and Control brains from the Mount Sinai Brain Bank (MSBB) and quantified differences in viral sequences between groups. We began by profiling MSBB sample transcriptomes across four brain regions, (superior temporal gyrus [STG], n = 137, anterior prefrontal cortex [APFC], n = 213, inferior frontal gyrus [IFG], n = 186, and parahippocampal gyrus [PHG], n = 107), which we used to

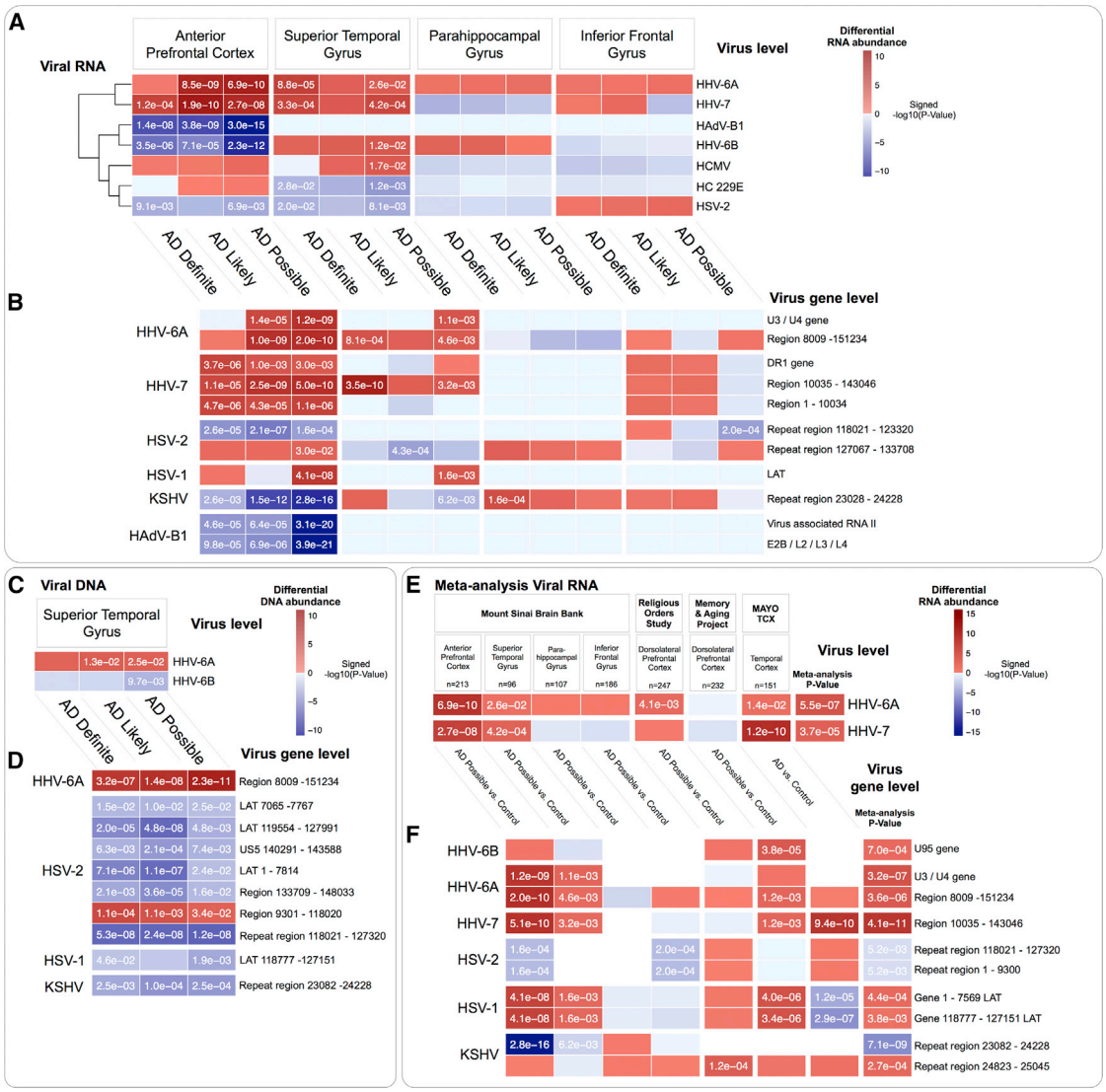


Figure 2. Differential Abundance of Viral RNA and DNA in AD

Multiregional comparison between AD versus controls of viral RNA and DNA, summarized to the level of full viral sequences (A and C) and viral genomic features (B and D). Meta-analysis of differential abundance of viral RNA in AD (E and F), incorporating post-mortem brain RNA-seq data from the MSBB, ROS, MAP, and MAYO TCX consortium studies also revealed increased HHV-6A and HHV-7 in AD. p values shown in cells with FDR < 0.1, features with a meta-analysis FDR < 0.1 shown in (E) and (F).

quantify the presence and abundance of 515 viral species known or suspected to infect humans as a primary host (Brister et al., 2015). We applied a viral mapping approach (Figure S1), based on a modified ViromeScan workflow (Rampelli et al., 2016), optimized for detection specificity, rather than sensitivity to allow us to discriminate between viruses with highly homologous regions, and to ensure we were not falsely including human-derived transcripts when summarizing viral abundance.

We estimated viral abundance at two levels. We summarized RNA reads to the level of the entire viral sequence, with the aim of estimating "total viral transcription." We then summarized the RNA reads to the level of individual genomic features based on counting reads that overlap any of the genomic features included in the NCBI annotations for that viral sequence (Brister et al., 2015). Throughout this study, we have evaluated viral abundance according to these two levels separately.

We identified differential abundance of multiple viral species in the APFC and STG (Figure 2A, Table S2). The most consistent difference we saw was an AD-associated increase in the abundance of two closely related *Roseoloviruses*, HHV-6A and HHV-7, across the APFC and STG. The third *Roseolovirus*, HHV-6B had a discordant profile with increased abundance in AD in the STG, and reduced in the APFC. Viral gene-level differential expression (Figure 2B, Table S2) identified increased abundance across the APFC and STG regions of the HHV-6A U3/U4 genes (positional homologs of human cytomegalovirus

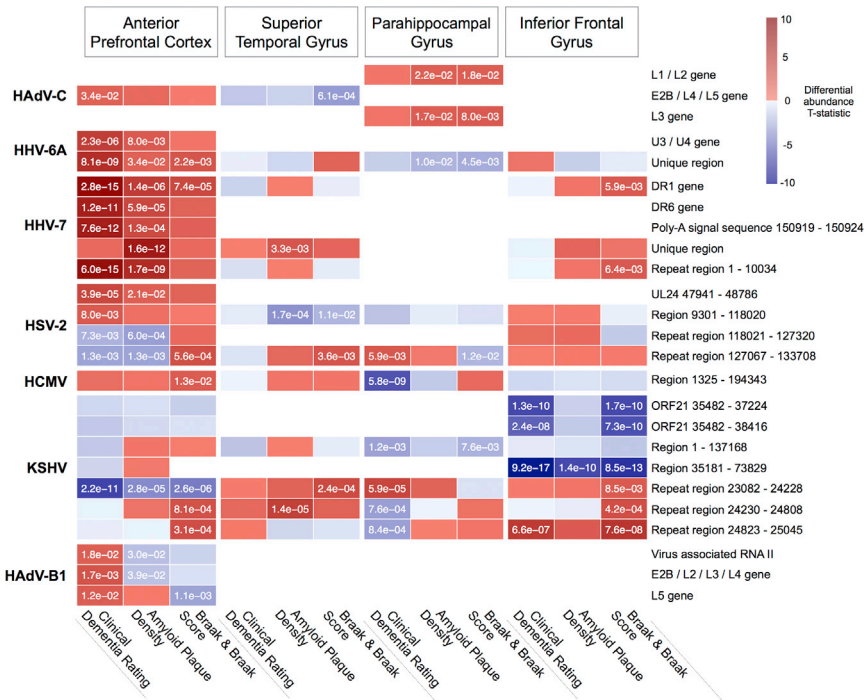


Figure 3. Viral Abundance Associates with AD Clinical and Neuropathology Traits

Multiregional associations between abundance of viral genomic feature RNA and clinical and neuropathological traits. Viral features with significant (FDR < 0.1) associations across multiple traits or brain regions are shown. p values shown in cells with FDR < 0.1.

[HCMV] US22, with transactivating effects on other viral species [Mori et al., 1998]) and the HHV-7 direct repeat terminal gene, DR1.

To understand whether our observations of altered viral abundance in AD would also be preserved in additional cohorts, we incorporated post-mortem brain RNA-seq data from three additional, independent consortium studies: (1) Religious Orders Study[47] (ROS), a longitudinal clinico-pathological study comprising 300 samples from the dorsolateral prefrontal cortex (DLPFC) of individuals with AD and healthy controls, (2) Memory and Aging Project (Bennett et al., 2012a, 2012b) (MAP) comprising 298 samples from the DLPFC of individuals with AD, as well as healthy controls, and (3) Mayo Temporal Cortex (Allen et al., 2016) (MAYO TCX) comprising 278 samples from the temporal cortex of individuals with AD, pathological aging (PA) (Dickson et al., 1992), progressive supranuclear palsy (PSP), and healthy controls. Our goal was to process these additional samples using the procedure described above and integrate the results (Table S2) into a meta-analysis of viral abundance in AD (Figures 2E and 2F). Our main finding was a consistently increased abundance of HHV-6A and HHV-7, driven mainly by the "unique" region of each virus (the full viral sequence excluding the ~15 kb flanking DR_L and DR_R terminal repeats). We also observed an increased abundance of the HSV-1 latency-associated transcript (LAT), supporting an increased rate of HSV-1 infection (latent or otherwise) in AD. These additional validations support the main findings originally observed in the MSBB cohort, that multiple viruses are at increased abundance in AD, across multiple brain tissues, with prominent roles for *Roseoloviruses* HHV-6A and HHV-7.

Increased HHV-6A and HHV-7 Are Not Ubiquitous Features of Neurodegeneration

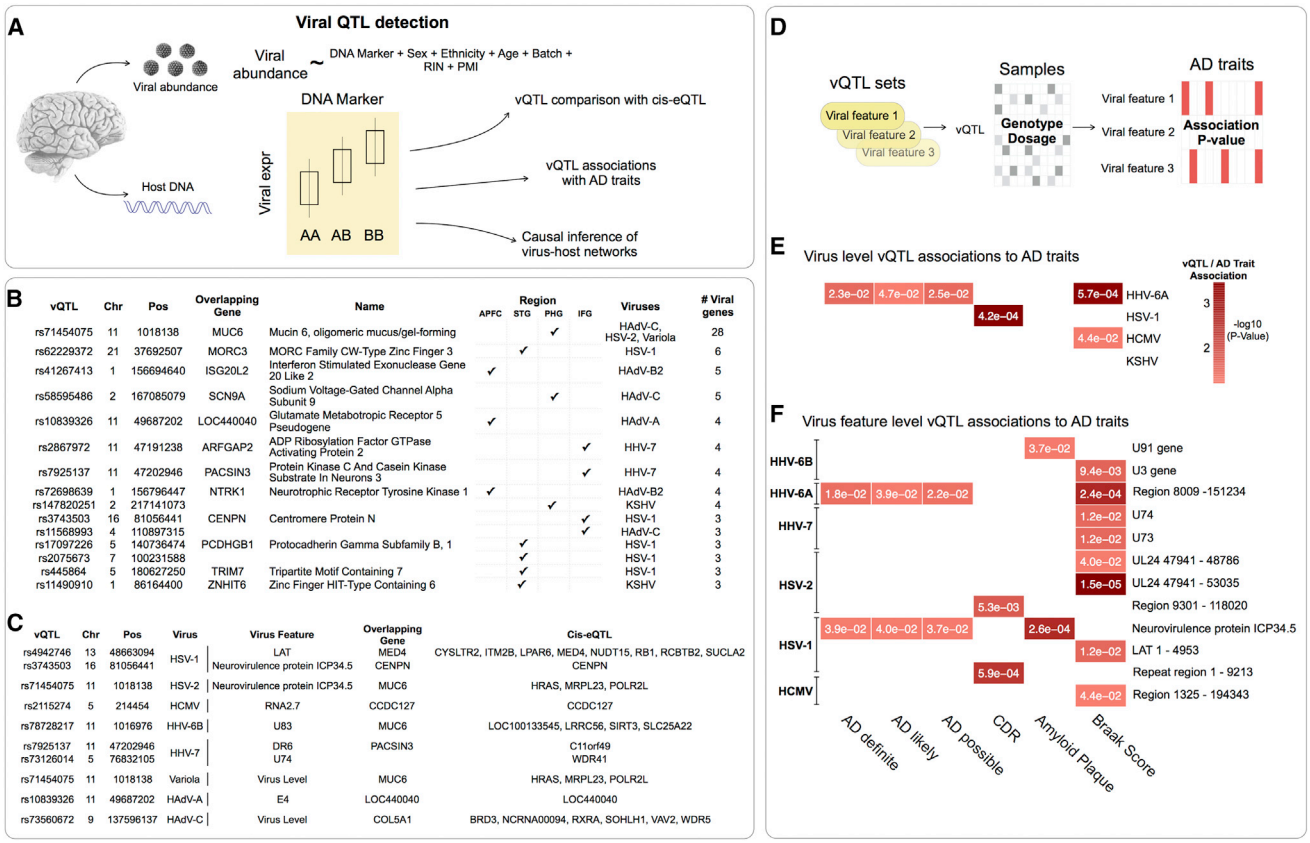
We utilized the PA and PSP samples available within the MAYO TCX cohort to perform comparisons against these additional neurodegenerative disorders, to understand whether our findings were specific to AD or perhaps reflect a more general feature of neurodegenerative processes. In a comparison of AD versus PA, we found an increased abundance of HHV-6A and HHV-7 (Figure S2). When we compared AD versus PSP, we found increased HHV-7 in AD, but *reduced* HHV-6A. Taken together, these observations suggest that elevated HHV6-A and HHV-7 are not ubiquitous features of neurodegenerative disease, although HHV-6A may also be relevant to other diseases such as PSP.

Increased Abundance of HHV-6A DNA in Alzheimer's Disease

We looked for evidence of viral DNA in whole-exome sequencing (WES) data that was generated on STG samples (n = 286) in the MSBB cohort, applying a similar procedure to that used in evaluating viral RNA abundance (Figure S1). We detected viral DNA for multiple viruses, and identified an increased abundance of HHV-6A (Figures 2C and 2D, Table S2). This was primarily due to reads mapping to HHV-6A Region 8009-151234, which comprises the "unique" region of HHV-6A (and is consistent with our findings in the RNA sequences). Chromosomal integration of HHV-6A into host subtelomeric regions is well described (Arbuckle et al., 2010), occurring via a mechanism involving homologous recombination between telomeric repeats and the DR_R. Excision and reactivation of integrated HHV-6A is associated with preferential loss of the entire DR_L, facilitating viral circularization and rolling circle replication (Prusty et al., 2013). This may indicate that the HHV-6A DNA that we find as more abundant in AD reflects HHV-6A that has undergone reactivation from a chromosomally integrated form, although we have not evaluated this directly.

Viral RNA Abundance Associates with Clinical Dementia and Neuropathology Traits

We extended our analysis to identify significant associations between virus level and viral gene level RNA abundance and AD-relevant clinical and neuropathological traits ("AD traits")



B

vQTL	Chr	Pos	Overlapping Gene	Name	Region APFC	STG	PHG	IFG	Viruses	# Viral genes
rs71454075	11	1018138	MUC6	Mucin 6, oligomeric mucus/gel-forming			✓		HAdV-C, HSV-2, Variola	28
rs62229372	21	37692507	MORC3	MORC Family CW-Type Zinc Finger 3		✓			HSV-1	6
rs41267413	1	156694640	ISG20L2	Interferon Stimulated Exonuclease Gene 20 Like 2	✓				HAdV-B2	5
rs58595486	2	167085079	SCN9A	Sodium Voltage-Gated Channel Alpha Subunit 9			✓		HAdV-C	5
rs10839326	11	49687202	LOC440040	Glutamate Metabotropic Receptor 5 Pseudogene	✓				HAdV-A	4
rs2867972	11	47191238	ARFGAP2	ADP Ribosylation Factor GTPase Activating Protein 2				✓	HHV-7	4
rs7925137	11	47202946	PACSIN3	Protein Kinase C And Casein Kinase Substrate In Neurons 3				✓	HHV-7	4
rs72698639	1	156796447	NTRK1	Neurotrophic Receptor Tyrosine Kinase 1	✓				HAdV-B2	4
rs147820251	2	217141073					✓		KSHV	4
rs3743503	16	81056441	CENPN	Centromere Protein N				✓	HSV-1	3
rs11568993	4	110897315						✓	HAdV-C	3
rs17097226	5	140736474	PCDHGB1	Protocadherin Gamma Subfamily B, 1		✓			HSV-1	3
rs2075673	7	100231588				✓			HSV-1	3
rs445864	5	180627250	TRIM7	Tripartite Motif Containing 7		✓			HSV-1	3
rs11490910	1	86164400	ZNHIT6	Zinc Finger HIT-Type Containing 6		✓			KSHV	3

C

vQTL	Chr	Pos	Virus	Virus Feature	Overlapping Gene	Cis-eQTL
rs4942746	13	48863094	HSV-1	LAT	MED4	CYSLTR2, ITM2B, LPAR6, MED4, NUDT15, RB1, RCBTB2, SUCLA2
rs3743503	16	81056441		Neurovirulence protein ICP34.5	CENPN	CENPN
rs71454075	11	1018138	HSV-2	Neurovirulence protein ICP34.5	MUC6	HRAS, MRPL23, POLR2L
rs2115274	5	214454	HCMV	RNA2.7	CCDC127	CCDC127
rs78728217	11	1016976	HHV-6B	U83	MUC6	LOC100133545, LRRC56, SIRT3, SLC25A22
rs7925137	11	47202946	HHV-7	DR6	PACSIN3	C11orf49
rs73126014	5	76832105		U74		WDR41
rs71454075	11	1018138	Variola	Virus Level	MUC6	HRAS, MRPL23, POLR2L
rs10839326	11	49687202	HAdV-A	E4	LOC440040	LOC440040
rs73560672	9	137596137	HAdV-C	Virus Level	COL5A1	BRD3, NCRNA00094, RXRA, SOHLH1, VAV2, WDR5

Figure 4. Viral QTL Detection and Association with AD Genetics

(A) Host DNA markers significantly associated (FDR < 0.25) with viral abundance were classified as vQTL for that feature-tissue combination.

(B) Top multi-viral vQTL associations, with overlapping gene symbols and associated viruses.

(C) Top vQTL associations where vQTL is also a cis-eQTL for at least one host gene.

(D) Sequence kernel association test to evaluate whether vQTL markers are also associated with AD traits.

(E and F) Virus level vQTL sets for HHV-6A were highlighted most strongly (E), whereas viral feature level associations implicated features from HHV-6A and HSV-1 (F).

Panels (E) and (F) show associations for any viruses/viral features that were implicated in viral differential RNA abundance, p values < 0.05 shown in cells.

(Figure 3, Table S3), including a consensus-based clinical dementia rating score (Morris, 1993) (CDR), multi-regional neuritic amyloid plaque density (Haroutunian et al., 1998) (APD), and Braak and Braak score (Braak et al., 2006) (Braak score). We identified several viral genes that significantly associate (FDR < 0.1) with multiple AD traits, including the HHV-7 DR1 gene and HHV-6A unique region, both demonstrating a positive association with all three AD traits within the APFC.

Human DNA Variants that Associate with Viral Abundance Also Associate with AD Status, Clinical Dementia, and Neuropathological Features of AD

Given the prolonged preclinical course of AD, a primary question for us was whether these viral species represent a truly informative, causal component of AD or instead reflect an "opportunistic passenger" of a neurodegenerative process driven by other factors. To help address this, we integrated WES data with RNA-seq for each donor within the MSBB. Our goal was to identify host DNA variants that significantly associate with viral abundance, which we refer to as "viral quantitative trait loci" (vQTL). These vQTL might then be used in a causal inference paradigm (Millstein et al., 2009) to resolve directed regulatory interactions in virus-host networks and, for the fraction of vQTLs that are also AD risk-associated, evaluate whether viral abundance is an authentic risk factor for AD. Causal inference approaches like this have been used to elucidate molecular networks impacted by DNA loci across biological contexts as diverse as cardiometabolic disease (Franzén et al., 2016), COPD (Yoo et al., 2015), and meditation (Epel et al., 2016).

We identified host DNA markers (Shabalin, 2012) across all four brain regions with paired DNA and RNA samples (APFC: n = 174, STG: n = 86, PHG: n = 80, IFG: n = 147) that significantly associated with normalized viral abundance, for any viral species detected within that region (Figure 4, Table S4). DNA markers with a permutation-based FDR < 0.25 were classified as vQTLs for that specific viral gene in the context of that region.

We identified 1,672 vQTL associations across the four regions assayed (APFC: 883, STG: 479, PHG: 175, IFG: 135). This represented 747 non-independent vQTL markers that collectively associate with 16 separate viruses. The viruses with the largest

number of separate vQTL markers were human adenovirus-A (HAdV-A), HHV-6A, HSV-2, and HSV-1 (222, 103, 91, and 87 markers, respectively). The vQTL associated with the most viruses and regions (Figure 4B) (rs71454075) falls within the glycoprotein Mucin 6, Oligomeric Mucus/Gel-Forming gene (*MUC6*), expressed particularly in gastric epithelium, and with cytoprotective roles against pathogens, acids, and proteases (Toribara et al., 1993). The genes that collectively overlapped vQTLs across the most regions and viruses (Figure 4C) indicated biological themes that plausibly relate to individual variability in virome composition, including mucosal immunity (*MUC6* and *NTRK1*), innate immunity, and anti-viral sensing (*ISG20L2*, *MORC3*, *NTRK1*, and *TRIM7*).

We hypothesized that cis-eQTL associations could represent a potential mechanism linking vQTL with viral abundance, whereby a vQTL might alter the expression of a nearby host gene that has the potential to impact, or be impacted by viral abundance. We performed a cis-eQTL analysis across this same cohort, detecting significant associations (FDR < 0.1) between host gene expression, and markers within 1 MB of gene boundaries. Thirty-five percent of the unique vQTL markers (263 of 747) were associated with at least one cis-gene, including several with associations to AD and other dementias, for instance, rs4942746 is a vQTL for HSV-1 LAT and is also a cis-eQTL for Integral Membrane Protein 2B (*ITM2B*), an endogenous inhibitor of Aβ aggregation and associated with familial British (Vidal et al., 1999) and Danish (Vidal et al., 2000) dementia.

We employed the sequence kernel association testing (Lee et al., 2012; Wu and Pankow, 2016) approach to understand whether vQTLs are genetic risk markers for AD status, clinical dementia rating or neuropathology (AD traits) (Figure 4D, Table S5). Iterating over each viral gene feature, we combined vQTL from all regions into a single set, and calculated a P value associating each vQTL set with each AD trait. We detected multiple viral features with significant associations to AD traits (Figures 4E and 4F), supporting the hypothesis that DNA variants that predispose to AD also predispose to abundance of key viral species—most notably HHV-6A. We also found multiple viral genes with vQTL sets that associate with multiple AD traits, including the HHV-6A unique region and the HSV-1 Neurovirulence protein ICP34.5.

These results indicate a significant overlap between the genetic basis for AD traits, and the abundance of specific viral species and viral genomic features. This supports our broader hypothesis that viral activity plays a role in the development and progression of AD and is consistent with a role for HHV-6A in particular.

Viral Regulation of AD Associated Host Networks

We generated virus-host gene regulatory networks to understand the biological context of viral activity across samples. We constructed informative models by detecting the set of host genes that are correlated with vQTL/virus pairs within each tissue, iterating over each candidate trio (each trio comprising a vQTL marker, virus feature abundance, and host gene expression), and then testing the mathematical conditions required to demonstrate causal mediation of an association between a DNA marker and a trait (Millstein et al., 2009) (Figure 5A).

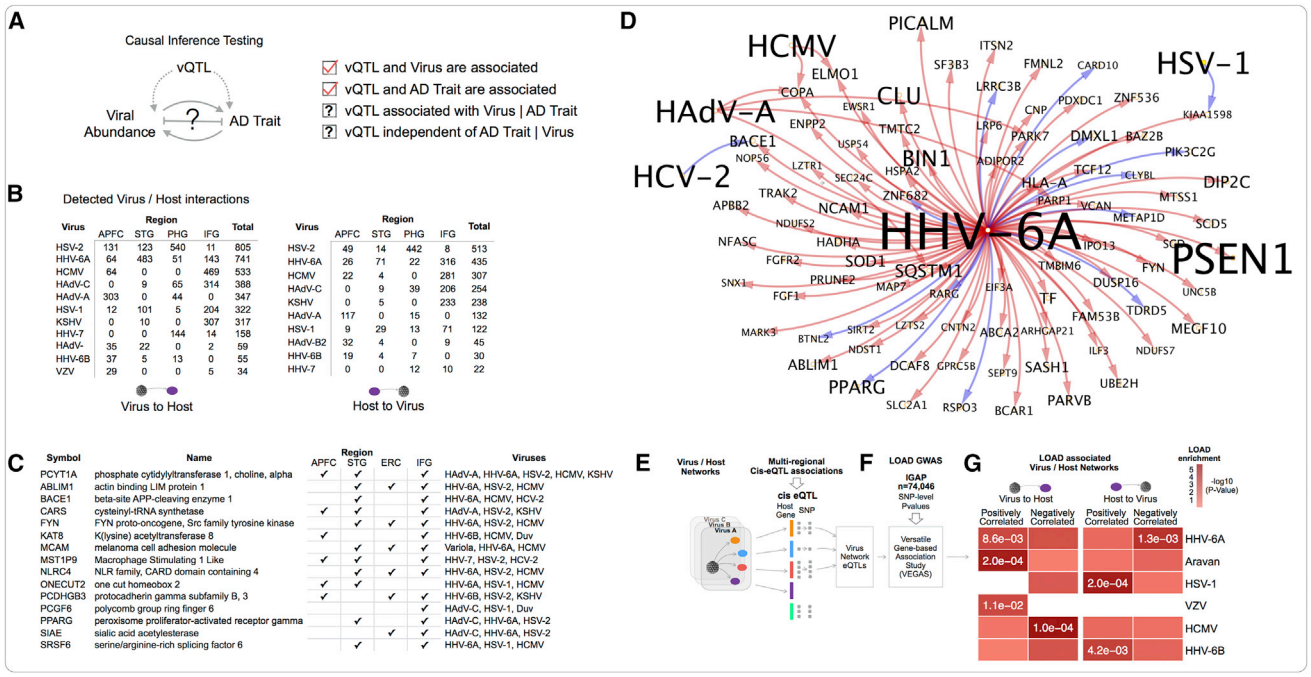
We built tissue-specific virus-host gene networks for all detected viral features (Table S6) identifying interactions in all four brain tissues (Figure 5B), comprising a total of 4,110 "virus to host" interactions, and 2,255 "host to virus" interactions, collectively associating 16 different viral species with 4,929 host genes. Only three viruses had detected interactions in all four tissues: HSV-1, HSV-2, and HHV-6A. Several viruses only had interactions detected in a single tissue: Aravan Virus, HCV-2, Variola Virus, and Wesselbron Virus, which may reflect higher regional tropism for the species or the potential for contamination to be driving the detection of these viruses in a subset of samples.

Host genes that are most commonly perturbed by viruses are shown (Figure 5C), ranked according to the number of unique viruses that we detected perturbing that gene across all four regions surveyed. This includes several genes implicated in regulation of APP processing and AD, including β-site amyloid precursor protein cleaving enzyme 1 (*BACE1*), FYN Proto-Oncogene, Src Family Tyrosine Kinase (*FYN*), and Peroxisome Proliferator Activated Receptor Gamma (*PPARG*). Several of these genes are also associated with pro- and anti-viral signaling, including positive regulation of interferon-λ1 genes by *FYN* during viral infection (Nousiainen et al., 2013), negative regulation of viral replication by *PPARG* (Bernier et al., 2013), and promotion of viral translation by *SRSF6* (Swanson et al., 2010).

We evaluated the set of genes causally regulated by each virus, against a set of known AD-associated genes, including risk genes for early- and late-onset AD, as well as AD associated traits (such as β-amyloid plaque density, rate of disease progression, neurofibrillary tangle density) from multiple human genetics disease resources (Rouillard et al., 2016). We found that multiple viruses interact with AD risk genes. HHV-6A stood out as notable with significant overlap (FDR < 3e-3) between the set of host genes it collectively induces across all tissues and AD-associated genes (Figure 5D, Table S7). This includes several regulators of APP processing and AD risk-associated genes, including gamma-secretase subunit presenilin-1 (*PSEN1*), *BACE1*, amyloid beta precursor protein binding family B member 2 (*APBB2*), Clusterin (*CLU*), Bridging Integrator 1 (*BIN1*), and Phosphatidylinositol Binding Clathrin Assembly Protein (*PICALM*). We also found that several other viruses regulate, or are regulated by, AD risk genes, including: (1) HAdV-C-induced expression of Complement Receptor 1 (*CR1*), and inhibition of Solute Carrier Family 24 Member 4 (*SLC24A4*), (2) inhibition of KSHV by Fermitin Family Member 2 (*FERMT2*), and (3) inhibition of HSV-2 by Translocase of Outer Mitochondrial Membrane 40 (*TOMM40*). These findings indicate multiple points of overlap between virus-host interactions and AD risk genes.

Expression QTL of Virus-Host Networks Are Enriched for AD GWAS Risk Loci

To investigate the broader context of how virus-host interactions might overlap with AD genetic risk, we integrated cis-eQTL data with AD GWAS summary statistics (Figures 5E and 5F). Our hypothesis was that identifying virus-host networks that are enriched for AD GWAS loci could provide a natural means to prioritize the relevance of individual viruses to AD, as well as provide useful functional context for specific viruses. Our approach was to use cis-eQTL identified in any of the four brain tissues within



Virus	Region	APFC	STG	PHG	IFG	Total
HSV-2		131	123	540	11	805
HHV-6A		64	483	51	143	741
HCMV		64	0	0	469	533
HAdV-C		0	9	65	314	388
HAdV-A		303	0	44	0	347
HSV-1		12	101	5	204	322
KSHV		0	10	0	307	317
HHV-7		0	0	144	14	158
HAdV-		35	22	0	2	59
HHV-6B		37	5	13	0	55
VZV		29	0	0	5	34

Virus	Region	APFC	STG	PHG	IFG	Total
HSV-2		49	14	442	8	513
HHV-6A		26	71	22	316	435
HCMV		22	4	0	281	307
HAdV-C		0	9	39	206	254
KSHV		0	5	0	233	238
HAdV-A		117	0	15	0	132
HSV-1		9	29	13	71	122
HAdV-B2		32	4	0	9	45
HHV-6B		19	4	7	0	30
HHV-7		0	0	12	10	22

Symbol	Name	Region APFC	STG	ERC	IFG	Viruses
PCYT1A	phosphate cytidylyltransferase 1, choline, alpha	✓	✓		✓	HAdV-A, HHV-6A, HSV-2, HCMV, KSHV
ABLIM1	actin binding LIM protein 1		✓	✓	✓	HHV-6A, HSV-2, HCMV
BACE1	beta-site APP-cleaving enzyme 1		✓		✓	HHV-6A, HCMV, HCV-2
CARS	cysteinyl-tRNA synthetase	✓	✓		✓	HAdV-A, HSV-2, KSHV
FYN	FYN proto-oncogene, Src family tyrosine kinase		✓	✓	✓	HHV-6A, HSV-2, HCMV
KAT8	K(lysine) acetyltransferase 8	✓			✓	HHV-6B, HCMV, Duv
MCAM	melanoma cell adhesion molecule		✓	✓	✓	Variola, HHV-6A, HCMV
MST1P9	Macrophage Stimulating 1 Like	✓	✓		✓	HHV-7, HSV-2, HCV-2
NLRC4	NLR family, CARD domain containing 4		✓	✓	✓	HHV-6A, HSV-2, HCMV
ONECUT2	one cut homeobox 2	✓	✓			HHV-6A, HSV-1, HCMV
PCDHGB3	protocadherin gamma subfamily B, 3	✓		✓	✓	HHV-6B, HSV-2, KSHV
PCGF6	polycomb group ring finger 6				✓	HAdV-C, HSV-1, Duv
PPARG	peroxisome proliferator-activated receptor gamma		✓		✓	HAdV-C, HHV-6A, HSV-2
SIAE	sialic acid acetylesterase			✓	✓	HAdV-C, HHV-6A, HSV-2
SRSF6	serine/arginine-rich splicing factor 6		✓		✓	HHV-6A, HSV-1, HCMV

Figure 5. Viral Regulation of AD-Associated Host Networks

(A) Integration of vQTL with viral abundance, and host gene expression was used to infer directed virus/host subnetworks.

(B) Virus/host gene network sizes for viruses with interactions detected in multiple tissues.

(C) Host genes that are most frequently perturbed by viruses.

(D) Host genes upregulated by HHV-6A are enriched for a heterogeneous set of AD risk and biomarker-associated genes. Shown here are the HHV-6A/AD-associated gene subnetwork and detected interactions with additional viruses.

(E–G) Brain cis-eQTLs associated with expression of specific virus/host networks are enriched for AD GWAS risk loci.

p values shown in cells with FDR < 0.1.

the MSBB data, to define the set of markers that are associated with the host genes in each virus/host network ("virus network eQTLs") and then determine whether virus network eQTLs are enriched for AD risk-associated loci (Lambert et al., 2013) using the versatile gene-based association study (VEGAS) approach (Liu et al., 2010b; Mishra and Macgregor, 2015). We observed that multiple viruses have host networks that are enriched for AD risk-associated eQTLs (Figure 5G), most consistently for HHV-6A, but also HCMV, HSV-1, Aravan, HHV-6B, and VZV. This suggests that loci that alter the expression of genes that regulate, or are regulated by specific viruses, are in aggregate associated with AD risk.

Viral Mediation of Neuronal Loss in AD

Progressive neuronal dysfunction and eventual death is a hallmark feature of AD and is the primary driver of the striking cortical atrophy associated with the disease. Diverse findings implicate aberrant regulation of apoptosis (Su et al., 1994), necroptosis (Caccamo et al., 2017), and autophagy (Nixon et al., 2005) although findings are conflicting and a unified understanding of the mechanisms underlying neuronal loss is lacking. Given the potential for viral infection to modulate cellular death pathways through these mechanisms and others (Upton and Chan, 2014), we were interested in understanding whether viral abundance might be associated with specific brain cell-type fractions within our dataset, particularly neurons. We used CIBERSORT (Newman et al., 2015) to deconvolute each MSBB RNA-seq sample into estimated fractions for major brain cell types (neurons, astrocytes, microglia, endothelial cells, and oligodendrocytes), based on comparison with a reference panel of single-cell RNA-seq in cortical samples from neurologically normal, middle-aged adults (Darmanis et al., 2015) (Figure 6A). Our goal was to identify significant associations between estimated cell fractions, AD traits, and viral abundance.

The most consistent AD-associated change (Table S8) was a significant (FDR < 0.1) decrease in neuronal fraction, and an increased endothelial cell fraction in the PHG, IFG, and STG, which we confirmed using an alternative cell signature enrichment method, xCell (Aran et al., 2017) (Table S8). We observed strong negative correlation between neuronal fractions and CDR in all four brain tissues, and with Braak score and neuritic plaque density in three (PHG, IFG, STG) (Figure 6B, Table S8).

We found multiple correlations between cell fractions and viral RNA abundance (Figure 6C, Table S9), with associations detected in the APFC, STG, and IFG, collectively implicating all cell types with five viruses. Only a single virus (HHV-6A) was associated with cell fraction changes in multiple tissues (APFC and STG; the same two tissues in which HHV-6A was observed at increased abundance in AD), although the specific cell profiles were different between tissues. We were most interested in

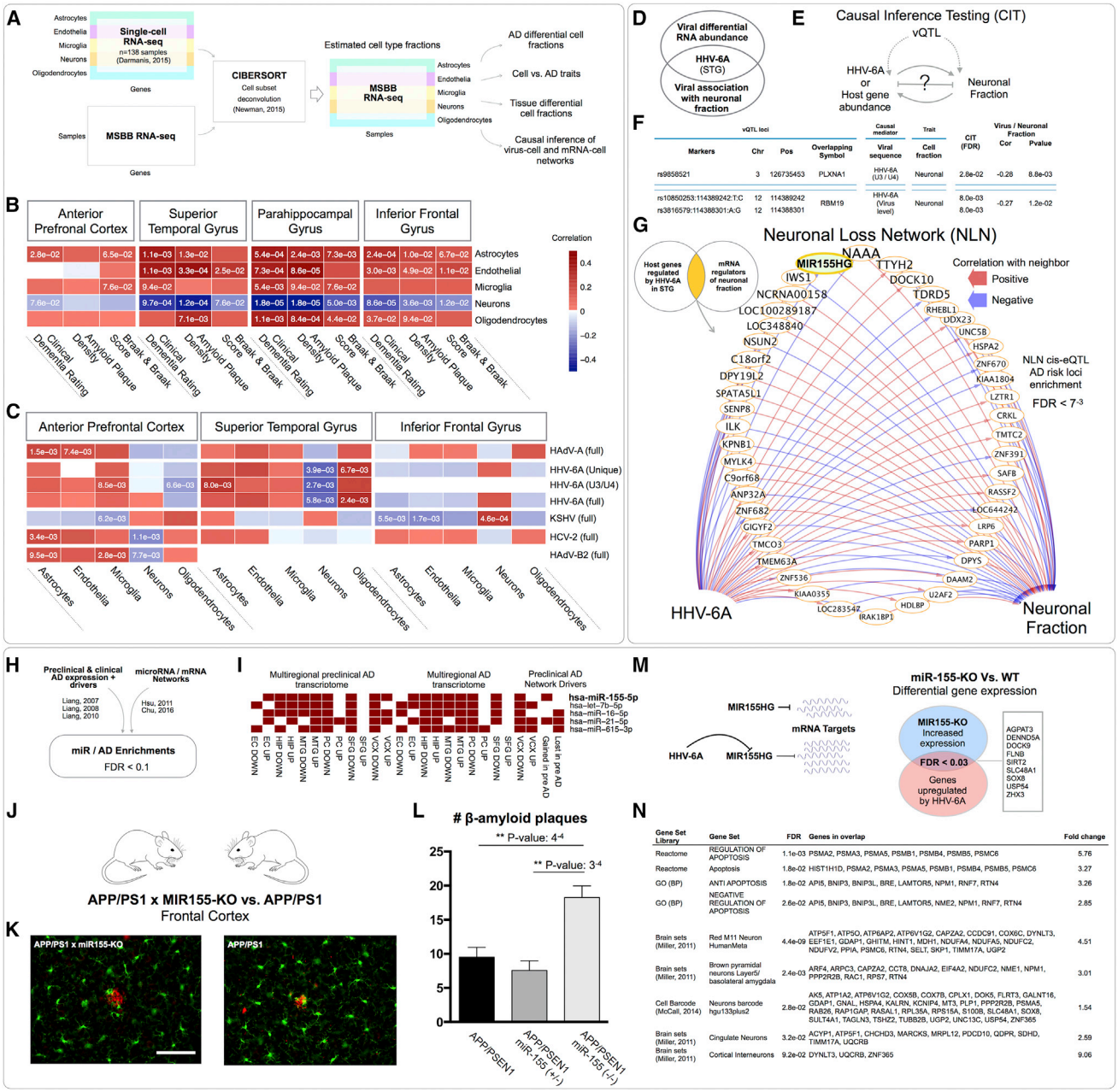


Figure 6. Viral Mediation of Neuronal Loss in AD

(A–C) Cell fraction estimates in the MSBB RNA-seq data (A), linking cell fractions with AD traits (B), and viral RNA abundance (C).

(D) HHV-6A was negatively correlated with neuronal fraction, while also at increased abundance in the STG.

(E and F) Causal inference testing identified HHV-6A vQTL associated with neuronal fraction, which also have their effect mediated by HHV-6A sequences.

(G) Neuronal loss network of host genes regulated by HHV-6A, which also exerts an effect on neuronal fraction, including *MIR155HG*. Gene label size is inversely proportional to the smallest AD risk p value for its associated cis-eQTL.

(H and I) We initially prioritized miR-155 during our analysis of the preclinical AD networks, due to strong associations between known miR-155 gene targets and a variety of multiregional AD and preclinical AD transcriptomic changes, as well as differential drivers of the preclinical AD network.

(J–L) Four-month-old miR-155-KOxAPP/PS1 produce increased cortical β-amyloid plaques compared with APP/PS1 mice.

(M) We hypothesized that HHV-6A inhibition of miR-155 might cause a disinhibition of miR-155 targets. Cortical RNA-seq of miR-155-KO versus WT mice demonstrated significant overlap between genes upregulated (FDR < 0.1) in the absence of miR-155, and host genes detected as upregulated by HHV-6A in the virus/host networks.

(N) Gene set enrichments for miR-155-KO versus WT differentially expressed genes identified multiple enrichments (FDR < 0.1) consistent with a role for miR-155 in neuronal loss.

Error bars shown in (L) denote mean ± SEM.

focusing on viruses or viral genes that are negatively correlated with neuronal fraction, particularly if they were also detected as differentially abundant in that same brain tissue (Figure 6D). Only HHV-6A met these criteria demonstrating strong negative correlation with neuronal fraction in the STG, as well as being highlighted in the viral RNA analyses of the MSBB, MAP, MAYO TCX cohorts, and meta-analysis.

We applied causal inference testing to evaluate the hypothesis that abundance of HHV-6A exerts a significant effect on neuronal fraction in the STG and detected multiple instances consistent with the HHV-6A virus and the HHV-6A U3/U4 gene mediating such an effect (Figures 6E and 6F, Table S9). We reasoned that part of this effect might be through the direct impact of viral products on biological mechanisms underlying the neuronal loss; however, there might also be informative, indirect effects mediated through a host subnetwork (i.e., host genes regulated by HHV-6A, and which *also* causally regulate neuronal fraction). We iterated over the set of host genes we detected as regulated by each HHV-6A sequence in the STG, and for each gene, tested whether it also correlated with neuronal fraction, and whether the vQTL marker for that host gene was associated with neuronal fraction in the STG. We performed causal inference testing and detected genes that are regulated by HHV-6A virus (121 genes), HHV-6A unique region (86 genes), and HHV-6A U3/U4 (6 genes) and that exert an effect on neuronal fraction (Figure 6G, Table S9). To determine whether these "neuronal loss networks" (NLN) might overlap with AD GWAS risk loci, we used the VEGAS approach described above and found a significant enrichment within the cis-eQTL of the HHV-6A virus level NLN (FDR < 7e-3).

The HHV-6A virus level NLN includes many genes with roles in cellular viral response, as well as associations to biological mechanisms that could help account for neuronal death, including upregulation of Poly(ADP-Ribose) Polymerase 1 (*PARP1*) by HHV-6A (and negative regulation of neuronal fraction by *PARP1*), consistent with its reported activation by EBV (Mattiussi et al., 2007), HIV (Ha et al., 2001), and HSV-1 (Grady et al., 2012), and induction of caspase-independent apoptosis in the latter (Grady et al., 2012). AD risk-associated haplotypes for PARP1 have also been reported (Liu et al., 2010a) and suggested as a mechanism mediating cell death following cytotoxic response (Liu et al., 2010a). The NLN gene with the lowest eQTL AD risk p values (Figure 6G) is N-acylethanolamine acid amidase (*NAAA*), a close homolog of Acid Ceramidase (*AC*), suggested to regulate neuronal apoptosis in AD (Huang et al., 2004).

We were also interested to find that MIR155 Host Gene (*MIR155HG*) was within the HHV-6A NLN and associated with multiple low AD risk p value eQTLs. We had prioritized miR-155 during our analysis of the preclinical AD networks (Figures 6H and 6I, also Table S1) due to strong associations between its mRNA targets and a variety of multiregional AD and preclinical AD transcriptomic changes as well as "gained in preclinical AD" drivers. In addition to diverse associations with viral biology including EBV (Gatto et al., 2008), HHV-6A (Caselli et al., 2017), KSHV (Gottwein et al., 2007), and MDV (Zhao et al., 2009), miR-155 has also been reported as a regulator of T cell response in AD (Song and Lee, 2015), a mediator of inflammation-induced neurogenic dysfunction and apoptosis (Woodbury et al., 2015), and an effector of TREM2-APOE regulation of a microglial neurodegenerative phenotype (Krasemann et al., 2017). Reports of viral tropism have also demonstrated that HHV-6 and HHV-7 can infect microglia (Albright et al., 1998), and macrophages (Zhang et al., 2001). In the NLN, we found that HHV-6A suppressed miR-155, as described in recent reports of miR-155 inhibition by HHV-6A in infected T cells (Caselli et al., 2017).

These findings show that computational deconvolution of RNA-seq in AD recapitulates the expected neuronal loss. The negative correlations between STG neuronal fraction and HHV-6A sequences (that are also at increased abundance in the STG), in combination with the findings of causal testing, and the association of the NLN cis-eQTLs with AD risk are consistent with HHV-6A exerting an effect on neuronal fraction in AD.

miR-155 Is Suppressed by HHV-6A and Alters Aβ42 Levels and Amyloid Plaque Density in a Genetically Manipulated Mouse Model

Given the convergence of multiple analyses upon miR-155 (Figures 6G and 6I), we crossed miR-155-KO mice (Thai et al., 2007) with a standard APP/PS1 amyloidosis strain (Jankowsky et al., 2004) to evaluate the effect of miR-155 depletion on molecular and tissue pathology. At age 4 months, we observed that the brains of miR-155-KO/APP/PS1 mice displayed larger, more frequent cortical amyloid plaques (Figures 6K and 6L) and a higher level of Aβ42 (p value: 0.03, Mann-Whitney test) as compared with APP/PS1 mice. We hypothesized that if HHV-6A was inhibiting the expression of miR-155, then this might cause the de-repression of certain miR-155 mRNA targets (Figure 6M). We generated RNA-seq on cortical samples comparing miR-155-KO versus wild-type mice to identify the differentially expressed genes (DEG, FDR < 0.1, Table S10) and found a significant overlap between upregulated DEGs, and the set of host genes we had identified as upregulated by HHV-6A (FDR < 0.03), suggesting that some component of our detected HHV-6A network is mediated through an effect on miR-155, or a viral ortholog of miR-155 (which has not been described for HHV-6A).

Given the context of HHV-6A-induced inhibition of miR-155 within the NLN, we wondered whether we would observe transcriptomic evidence for mechanisms associated with neuronal loss. Molecular and functional enrichments of the miR-155-KO DEGs (Table S10) highlighted multiple enrichments for apoptosis regulatory pathways and neuronal signature gene sets (Figure 6N).

Collectively, these findings support the role of miR-155 as a key node in host response to AD-relevant viral perturbation, and as a potential mediator of neuronal loss. This is also consistent with a contribution of viral perturbation in driving the preclinical AD transcriptional phenotype given that our prioritization of miR-155 was informed by findings in the preclinical AD networks. The finding that miR-155-KO causes increased Aβ plaque deposition in the presence of APP/PS1 mutations also suggests a pathway linking viral perturbation with AD-associated neuropathology. This line of evidence is also consistent with findings linking viral infection with antimicrobial innate immune response, proamyloidogenesis, and microbe entombment (Kumar et al., 2016).

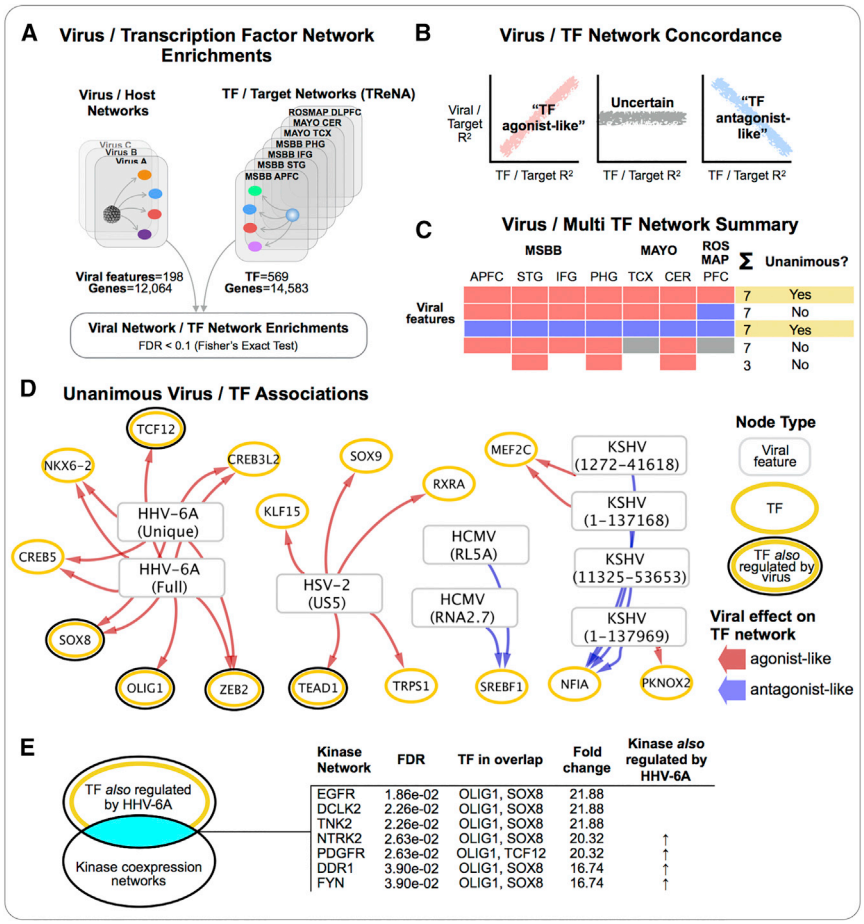


Figure 7. Viral Perturbation of Transcription Factor Regulatory Networks

(A) Comparison of virus/host networks with diverse TF-Target networks built from multiple independent AD datasets, to identify virus / TF network enrichments.

(B–D) Examination of the concordance of effect exerted by associated virus and TF upon target genes (B), and summarization of results across all TF-target networks (C) to identify unanimous virus/TF associations (D).

(E) Kinase enrichment analysis of the most strongly implicated TF identified several kinases that we *also* detected as regulated by HHV-6A, with known associations to AD and HHV-6A, indicating a potential mechanism for viral co-option of TF networks in AD.

Viral Perturbation of Transcription Factor Regulatory Networks

Viruses demonstrate remarkable strategies to effectively co-opt endogenous host factors necessary to their survival. Through context-dependent transcription of viral gene products, viruses hijack host transcription, signaling networks, and cellular machinery to orchestrate all aspects of their life cycle.

Given our earlier findings of widespread changes in C2H2-TF in the preclinical networks, we were interested in determining whether the virus-host networks might indicate viral targeting (directly or indirectly) of specific TFs (Figure 7). We constructed a diverse collection of TF-target networks, generated from the MSBB, MAYO, and ROSMAP consortium data, using the Transcriptional Regulatory Network Analysis (TReNA) approach comprising seven TF-target networks in total (reflecting tissue-specific networks within each cohort), which collectively model transcriptional relationships between 569 unique TFs and 14,583 target genes.

Within each virus-host network, we compared the "virus to host" genes with the target genes for each TF, across all seven TF-target networks to identify significantly (FDR < 0.1) associated virus/TF pairs (Figure 7A). For each enriched pair, we examined the concordance of effect that the virus and TF demonstrate on the genes driving the enrichment (Figure 7B). We calculated the Pearson's correlation between the individual virus-host gene correlations and the TF/target gene correlations to identify instances where the virus and TF both exerted a similar effect ("TF agonist-like," Pearson Corr > 0, FDR < 0.1), and instances where they exerted an opposing effect ("TF antagonist-like," Pearson Corr < 0, FDR < 0.1). We then identified the set of virus/TF pairs that were associated in all seven TF-networks and that demonstrated a consistent status as TF agonist-like/antagonist-like in each case (Figure 7C), reasoning that these "unanimous virus/TF associations" would represent the most robust candidates for considering viral perturbation of a TF network.

This unanimous virus/TF network (Figure 7D) includes 26 associations between four different viruses (HSV-2, HHV-6A, HCMV, and KSHV) and 14 TFs. The majority of connections (20 of 26) indicate associations where a viral feature exerts an agonist-like effect on TF targets. We found multiple virus-TF interactions where the TF was *also* directly detected as being regulated by the virus, including agonist-like interactions between HHV-6A and *OLIG1*, *TCF12*, *SOX8*, and *ZEB2*. In each of these cases, the TF was upregulated by the virus, consistent with the agonist-like associations inferred from the TF network enrichments. We hypothesized that a potential mechanism to link viral perturbation of multiple TF networks might be mediated through viral mimicry or modulation of host kinase activity (Shugar, 1999). We compared the set of four HHV-6A regulated TF (that were also directly detected as regulated by HHV-6A) with a library of kinase coexpression networks (Lachmann et al., 2018) and identified significant (FDR < 0.1) enrichments with seven different human kinases (Figure 7E). This includes four kinases that were *also* detected as directly upregulated by HHV-6A. For example, Neurotrophic Receptor Tyrosine Kinase 2 (*NTRK2*) expression is linked with neuronal survival in AD (Wong et al., 2012) and associated with AD risk (Chen et al., 2008). *FYN* tyrosine kinase, linked with synaptic dysfunction in AD via a range of Aβ- and/or tau-related mechanisms (Nygaard, 2018), is also bound directly by the HHV-6A U24 protein

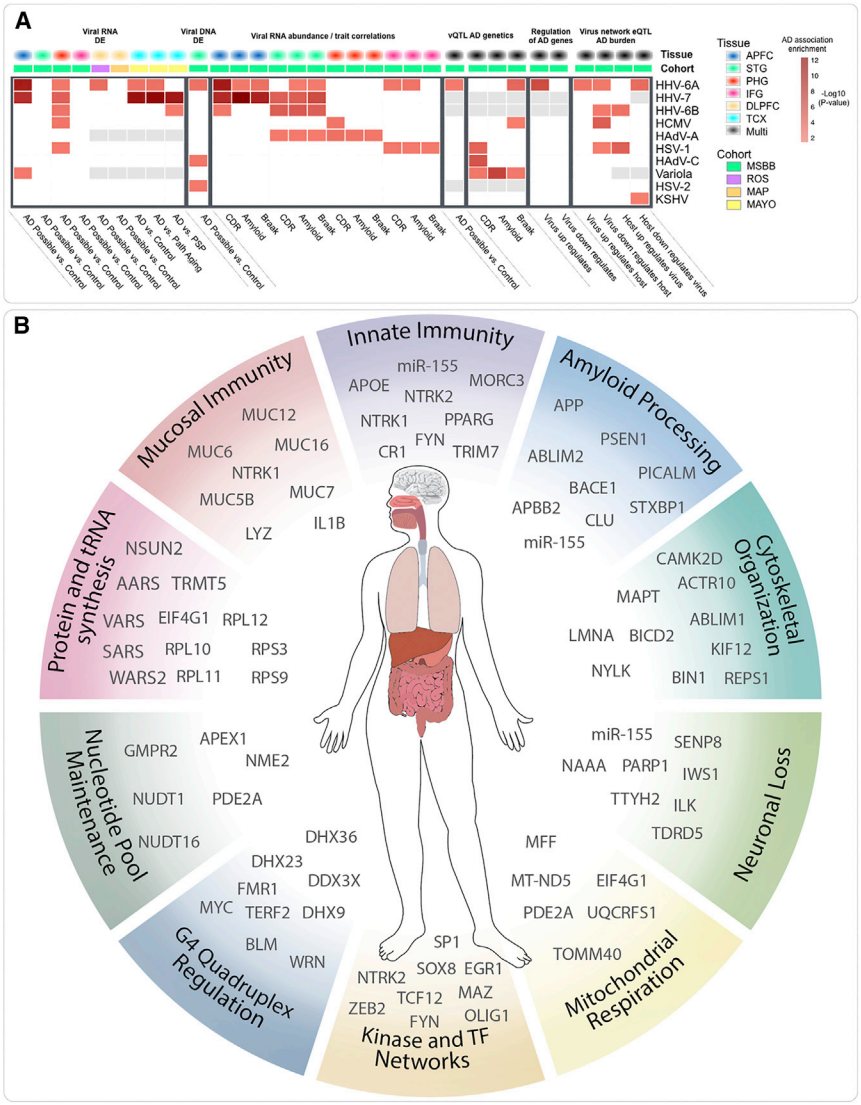


Figure 8. Multiomic Evaluation of AD-Associated Virome Implicates *Roseoloviruses* HHV-6 and HHV-7

(A) Summarized associations for each virus to diverse aspects of AD biology. Multiple viruses appear to have significant impacts on AD associated biology, particularly HHV-6A, HHV-7, and HHV-6B.

(B) Findings from this study indicate complex relationships between viral and host factors that are likely to be relevant across a range of timescales and organ systems. Key biological processes that have been highlighted are shown, along with top candidate molecular mediators.

potentially disrupting interactions with endogenous ligands (Sang et al., 2014).

These observations indicate the potential for viral perturbation of host TFs and TF-Target networks in the context of AD and offer an explanatory mechanism that could account for the large number of virus-host interaction detected for species such as HHV-6A and HSV-2. The described kinase-TF enrichments represent a potential upstream mechanism whereby HHV-6A modulation of kinase activity could alter the activity of specific endogenous TFs, thus perturbing host regulatory programs in the manner reflected in the HHV-6A host networks.

Virus-Host Protein Networks Indicate Perturbation of Cellular Nucleotide Pools, tRNA Synthesis, and Protein Translation

We constructed virus-host protein networks to evaluate proteomic consequences of viral activity in AD. We generated protein expression profiles for a subset of APFC samples (n = 152), performing liquid-chromatography-mass spectrometry (LC-MS) and using MaxQuant (Cox and Mann, 2008) for quantifying label free protein. Using a similar procedure as outlined for generation of virus-host RNA networks, we identified proteins that are associated with vQTLs and that are regulated by (or that regulate) each virus. Of the viruses that we found as differentially abundant in the APFC RNA data, we detected interactions with host proteins for HSV-1 (14 proteins), HSV-2 (34 proteins), and HHV-6A (28 proteins) (Figure S3, Table S11). Protein regulators of cellular nucleotide pools, especially purine biosynthesis (NUDT16 and GMPR2), guanine nucleotide binding proteins (GNAS, GNAO1, GNG3, and GNG5), aminoacyl-tRNA synthetases (SARS, VARS, and AARS), mitochondrial function (MT-ND5 and MFF), nuclear organization (NCL and LMNA), and cytoskeletal disruption (CAMK2D, LMNA, MYLK, PRKCB, and TF) are among the most prominent biological themes of the networks. For instance, we found that several viruses alter expression of nucleotide regulating proteins, including HSV-2 induction of reductase enzyme GMPR2, which catalyzes conversion of G to A nucleotides, and HHV-6A induction of inosine diphosphatase NUDT16, which depletes the cellular pool of non-canonical purines IDP and ITP. Collectively, this suggested a picture of virally induced dysregulation of nucleotide pool metabolism, especially purine bases, consistent with several metabolomics studies in AD (Kaddurah-Daouk et al., 2011, 2013). This is notable, given our observations of G4 activity as key regulatory features among preclinical AD driver genes, which are primarily features of guanine-enriched sequences. Nucleotide pool depletion or imbalance is known to induce replicative stress, mediated through inadequate unwinding of stabilized G4 sequences, with associated genomic and epigenetic instability (Papadopoulou et al., 2015).

We also found that HHV-6A-induced expression of multiple AARS enzymes (VARS and SARS), responsible for charging their

cognate tRNAs with valine and serine, respectively. Increased d-serine has been reported in the CSF of patients with AD, as well cellular and mouse AD models (Madeira et al., 2015). Some viruses possess endogenous tRNA and AARS sequences, which appear critical to viral protein synthesis throughout the infectious cycle (Nishida et al., 1999). Multiple viruses recruit host tRNAs into virions for use as primers in reverse transcription (Mak and Kleiman, 1997) and some such as HIV also selectively incorporate host AARSs as well (Cen et al., 2001). Recent works have also demonstrated the role of tRNA in shifting the G4 conformational equilibrium toward a hairpin conformer (Rode et al., 2016).

These results indicate novel molecular mediators and associated pathways that might help shed light on mechanisms of viral pathogenicity, such as a potential role for AARSs in HHV-6A co-option of host protein synthesis machinery. These findings may also suggest novel molecular mediators and mechanisms for more widespread changes we have observed, such as dysregulation of G4 activity.

Impact of Viral Activity on Genetic, Transcriptomic, Clinical, and Neuropathology Networks in AD

We aggregated the results across several analyses to summarize the systems-level impact of individual viral species on AD biology (Figure 8A). This included multiregional viral RNA differential abundance, multiregional correlations of RNA abundance with AD traits, associations between vQTL sets and AD genetics, viral DNA differential abundance, and AD risk gene enrichment scores for virus-host subnetworks. Viral species implicated in any of these analyses were assigned a combined score based on the summed -log10(p value) of each individual association. This view indicated that multiple viruses impact on AD associated biology, across multiple -omic domains. The most strongly implicated viruses, were *Roseoloviruses* HHV-6A, HHV-7, and HHV-6B. HHV-6A in particular was robustly prioritized on the basis of: (1) RNA and (2) DNA differential abundance, (3) RNA abundance association with AD traits, (4) vQTL markers association with AD traits, (5) HHV-6A/host network eQTL enrichments for AD risk loci, and (6) inducing the expression of a significant fraction of AD risk genes within the HHV-6A/host network.

DISCUSSION

Developing a sophisticated understanding of the causal basis for AD is complicated by its protracted preclinical course, and the inability to routinely sample brain tissue. Distinguishing the earliest drivers of disease from the "opportunistic passengers" of a multi-decade neurodegenerative process is especially formidable given the profound changes in transcriptomic, proteomic, and histopathological profiles (Zhang et al., 2013).

We report a multi-stage study that aims to reconcile nascent changes in preclinical AD with findings made in the context of AD. Our strategy began by examining transcriptomes from brain regions that undergo the earliest changes in AD with the goal of identifying novel biology that could offer a frame for understanding the more dramatic changes seen in later stages of AD. Exploration of the preclinical AD networks sensitized us to pervasive dysregulation of C2H2-TF, G4 sequences, and a possible role for viral perturbations in driving network changes. This informed a focused evaluation of viral perturbations in clinical AD. We examined four, large multi-omic datasets that included next-generation sequence data that enabled direct examination of viral sequences. We observed the presence of many viral species in the aging brain and linked multiple viral species with AD biology, including regulation of AD genetic risk networks, AD gene expression changes, and association with clinical dementia rating and neuropathology burden. We found prominent roles for *Roseoloviruses* HHV-6A and HHV-7, both implicated across multiple domains, and in 3 independent cohorts. Importantly, the inclusion of the MAYO TCX dataset allowed us to perform comparisons between AD and other neuropathological controls. These additional comparisons suggest that HHV-6A and HHV-7 are *not* ubiquitous features of neuropathology and appear at least partly specific to AD. Comparison against additional neuropathological diseases and brain regions would help clarify this specificity, for instance although PSP is associated with accumulation of cortical neurofibrillary tangles, the most severe manifestations are typically seen throughout the basal ganglia, brainstem, and cerebellar structures, regions that were not profiled in this study, yet which might harbor different viral species and abundances than those observed.

Additional focused sequencing efforts are required to gain further resolution on the role of HHV-6 and HHV-7 in AD. Previous reports suggest an increased prevalence of HHV-6 in AD based on PCR of post-mortem brain tissue (Lin et al., 2002) and seroprevalence (Carbone et al., 2014), though subsequent serological studies did not find increased HHV-6 seropositivity (Agostini et al., 2016). Importantly, not all of the methods employed in these studies distinguished between HHV-6A and HHV-6B, which may account for some discrepancy. We found considerable inter-regional variability in AD-associated viral abundance, finding differences in four of the six regions assayed. Additionally, we found a discordant profile for HHV-6B (increased in STG, reduced in APFC), indicating the importance of distinguishing HHV-6B from HHV-6A (increased in the STG, APFC, DLPFC, and TCX). Given the near universal seropositivity for HHV-6 in the general population, seropositivity is likely too non-specific to reliably distinguish AD-relevant states of viral activity within the brain.

The miR-155 network offers a targeted area for further investigation of viral activity and neuronal loss in AD. We first identified miR-155 during investigation of the preclinical AD networks and later identified that HHV-6A negatively regulates neuronal fraction in the STG and that the cis-eQTLs of the host genes that mediate this are enriched for AD risk loci, leading to the re-emergence of miR-155 (via *MIR155HG*) in our analysis. Our finding that 4-month-old miR-155-KO x APP/PS1 mice develop increased cortical amyloid plaque density and increased levels of Aβ oligomers provides further evidence linking miR-155 to AD pathology. These findings support the view of miR-155 as a regulator of complex anti- and pro-viral actions, offer a mechanism linking viral activity with AD neuropathology, and support the hypothesis that viral activity contributes to AD.

The integrated findings of this study suggest that AD biology is impacted by a complex constellation of viral and host factors acting across different timescales and physiological systems

(Figure 8B). This includes host mucosal defense and modulation of innate immune response by virus *and* host. It also includes disturbance of core biological processes, including some that are well described in AD (e.g., APP processing, cytoskeletal organization, mitochondrial respiration, protein synthesis, and cell-cycle control) and some that are less well characterized (e.g., widespread shifts in G4 activity and C2H2-TF regulatory programs). We note potential mechanisms (and candidate molecular mediators) that we find perturbed by viral species and that have known impacts on these altered processes, for instance, virally driven changes in protein synthesis machinery, tRNA synthetase activity, and nucleotide pool maintenance, which collectively exert complex effects on G4 regulation and C2H2-TF activity.

Our interpretation of the changes seen in the preclinical AD networks rests partly on the true disease relevance of neuropathology in the absence of cognitive impairment. We cannot readily discriminate molecules involved in disease progression from molecules that are responsible for resilience and for maintaining brain function in the face of advanced AD pathology. Despite the uncertainty around the eventual health trajectory of these donors, our reasoning was that by that conditioning our analysis on changes in AD vulnerable brain regions, we might still find instructive biological themes, even if uncertainty remained around whether those changes were disease associated, opportunistic, or somehow adaptive. Despite this reasoning, and supportive circumstantial evidence (e.g., miR-155), we have not yet confirmed in an equivalent dataset whether the viral findings associated with clinical AD have predecessors in the preclinical AD context.

Investigating the subcellular distribution of viral DNA, especially for the key species of interest (HHV-6A and HHV-7) would add valuable context to these findings. Unlike most viruses discussed in this study, HHV-6 (Tanaka-Taya et al., 2004) and HHV-7 (Prusty et al., 2017) can integrate into subtelomeric regions of host chromosomes during latent infection and are excised into an episomal form during lytic infection and replication. Characterizing the extent and distribution of integrated versus episomal *Roseolovirus* in AD would be an important step in further understanding the mechanisms that connect viral abundance with molecular aspects of AD biology and would have implications for therapeutic targeting of latent viral reservoirs relevant to AD.

It is important to note that the findings reported in this study are not sufficient to definitively demonstrate that viral activity causally contributes to the onset or progression of AD, which would be most naturally established in a prospective, intervention-based study. We do report on multiple streams of indirect evidence, however, that enabled us to partially address this with the available data, including: (1) causal inference testing that supports a role for HHV-6A in contributing to neuronal loss in AD, (2) AD GWAS risk loci enrichments in virus-host network eQTLs, (3) emergence of molecules such as miR-155 from preclinical AD networks and virus-host networks, and (4) relative specificity of HHV-6A and HHV-7 for AD, compared with other neurodegenerative diseases. Follow-up studies that evaluate the onset and progression of AD phenotypes in virally infected AD model systems would be one approach to better delineate the causal and mechanistic relationships that link pathogen activity with the evolution of AD-associated behavioral, molecular, and neuropathological changes.

In summary, we find evidence that links the activity of specific viral species with molecular, genetic, clinical, and neuropathological aspects of AD. Interpretation of these findings in light of the disturbances in G4 and C2H2-TF regulation in the preclinical AD samples that prompted our evaluation of viral activity is supportive of an important role for viral activity, especially *Roseolovirus* HHV-6A and HHV-7, in the development and progression of AD.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Animals
- METHOD DETAILS
  - RNA sequencing human gene expression
  - Whole exome sequencing
  - Liquid Chromatography tandem mass spectrometry
  - Immunohistochemistry
  - Mouse RNA isolation and library preparation
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Construction of entorhinal cortex and hippocampal preclinical AD and control regulatory networks
  - G-quadruplex sequence prediction
  - Generation of viral 31-mer database
  - Detection of viral transcription in RNA and whole exome sequencing
  - Differential abundance of viral transcription in RNA sequencing
  - Viral QTL detection
  - AD GWAS enrichments for virus network eQTLs
  - Estimating cell type fractions from RNA-seq
  - Molecular and functional enrichment analysis
  - Inferring causal relationships between viral abundance and biomolecular, clinical and neuropathological traits
  - Transcription Factor Regulatory Network Analysis
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures and 11 tables and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.05.023.

ACKNOWLEDGMENTS

MSBB RNA-seq, WES, and proteomics data used for the evaluation of viral sequences in AD were generated from post-mortem brain tissue collected through the Mount Sinai VA Medical Center Brain Bank and were provided by Dr. Eric Schadt and Dr. Mary Sano from Mount Sinai School of Medicine (P50 AG005138). Proteomics data were also provided by Dr. Levey from Emory University. Construction of TReNA networks was supported by NIA grant U01AG046139 and NIH grant U54EB020406. The MAYO TCX RNAseq study data was led by Dr. Nilüfer Ertekin-Taner, Mayo Clinic, Jacksonville,

FL as part of the multi-PI U01 AG046139 using samples from the following sources. (1) The Mayo Clinic Brain Bank. Data collection was supported through funding by NIA grants P50 AG016574, R01 AG032990, U01 AG046139, R01 AG018023, U01 AG006576, U01 AG006786, R01 AG025711, R01 AG017216, R01 AG003949, NINDS grant R01 NS080820, CurePSP Foundation, and support from Mayo Foundation. (2) Sun Health Research Institute Brain and Body Donation Program of Sun City, Arizona. The Brain and Body Donation Program is supported by the National Institute of Neurological Disorders and Stroke (U24 NS072026 National Brain and Tissue Resource for Parkinson's Disease and Related Disorders), the National Institute on Aging (P30 AG19610 Arizona Alzheimer's Disease Core Center), the Arizona Department of Health Services (contract 211002, Arizona Alzheimer's Research Center), the Arizona Biomedical Research Commission (contracts 4001, 0011, 05-901, and 1001 to the Arizona Parkinson's Disease Consortium), and the Michael J. Fox Foundation for Parkinson's Research. The ROS/MAP RNAseq study data were provided by the Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago. Data collection was supported through funding by NIA grants P30AG10161, R01AG15819, R01AG17917, R01AG30146, R01AG36836, U01AG32984, U01AG46152, the Illinois Department of Public Health, and the Translational Genomics Research Institute. S.G. and M.E.E. acknowledge the support of U01 AG046170 from the NIA. B.R., S.G., M.E.E., and J.T.D. acknowledge the support of 1R56AG058469 from the NIA. Philanthropic financial support was provided by Katherine Gehl. The computational resources and staff expertise provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai also contributed to the performance of this research.

AUTHOR CONTRIBUTIONS

B.R., S.G., M.E.E., and J.T.D. designed the study. B.R. performed the computational analysis. M.E.E. and J.-V.H.-M. carried out the murine miR-155 experiments. B.R., S.G., J.-V.H.-M., M.E.E., and J.T.D. wrote the paper. C.C.F., M.A.R., P.S., and N.D.P. contributed the TReNA analyses to this paper. V.H., M.S., W.S.L., N.D.B., E.M.R., and E.E.S. revised the analysis critically for important intellectual content. All authors read and approved the final manuscript.

DECLARATIONS OF INTERESTS

The authors declare that they have no competing financial interests in relation to the work described.

Received: December 12, 2017
Revised: March 5, 2018
Accepted: May 15, 2018
Published: June 21, 2018

REFERENCES

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Agostini, S., Mancuso, R., Baglio, F., Cabinio, M., Hernis, A., Guerini, F.R., Calabrese, E., Nemni, R., and Clerici, M. (2016). Lack of evidence for a role of HHV-6 in the pathogenesis of Alzheimer's disease. J. Alzheimers Dis. *49*, 229–235.

Albright, A.V., Lavi, E., Black, J.B., Goldberg, S., O'Connor, M.J., and González-Scarano, F. (1998). The effect of human herpesvirus-6 (HHV-6) on cultured human neural cells: oligodendrocytes and microglia. J. Neurovirol. *4*, 486–494.

Allen, M., Carrasquillo, M.M., Funk, C., Heavner, B.D., Zou, F., Younkin, C.S., Burgess, J.D., Chai, H.S., Crook, J., Eddy, J.A., et al. (2016). Human whole genome genotype and transcriptome data for Alzheimer's and other neurodegenerative diseases. Sci. Data *3*, 160089.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Ament, S., Shannon, P., and Richards, M. (2017). TReNA: Fit transcriptional regulatory networks using gene expression, priors, machine learning. R package version 0.99.10.

Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220.

Arbuckle, J.H., Medveczky, M.M., Luka, J., Hadley, S.H., Luegmayr, A., Ablashi, D., Lund, T.C., Tolar, J., De Meirleir, K., Montoya, J.G., et al. (2010). The latent human herpesvirus-6A genome specifically integrates in telomeres of human chromosomes in vivo and in vitro. Proc. Natl. Acad. Sci. USA *107*, 5563–5568.

Artusi, S., Perrone, R., Lago, S., Raffa, P., Di Iorio, E., Palù, G., and Richter, S.N. (2016). Visualization of DNA G-quadruplexes in herpes simplex virus 1-infected cells. Nucleic Acids Res. *44*, 10343–10353.

Beaudoin, J.D., and Perreault, J.P. (2013). Exploring mRNA 3′-UTR G-quadruplexes: evidence of roles in both alternative polyadenylation and mRNA shortening. Nucleic Acids Res. *41*, 5898–5911.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. J. R. Stat. Soc. B *57*, 289–300.

Bennett, D.A., Schneider, J.A., Arvanitakis, Z., and Wilson, R.S. (2012a). Overview and findings from the religious orders study. Curr. Alzheimer Res. *9*, 628–645.

Bennett, D.A., Schneider, J.A., Buchman, A.S., Barnes, L.L., Boyle, P.A., and Wilson, R.S. (2012b). Overview and findings from the rush Memory and Aging Project. Curr. Alzheimer Res. *9*, 646–663.

Bernier, A., Cleret-Buhot, A., Zhang, Y., Goulet, J.P., Monteiro, P., Gosselin, A., DaFonseca, S., Wacleche, V.S., Jenabian, M.A., Routy, J.P., et al. (2013). Transcriptional profiling reveals molecular signatures associated with HIV permissiveness in Th1Th17 cells and identifies peroxisome proliferator-activated receptor gamma as an intrinsic negative regulator of viral replication. Retrovirology *10*, 160.

Boyle, A.P., Guinney, J., Crawford, G.E., and Furey, T.S. (2008). F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics *24*, 2537–2538.

Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H., and Del Tredici, K. (2006). Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. Acta Neuropathol. *112*, 389–404.

Brister, J.R., Ako-Adjei, D., Bao, Y., and Blinkova, O. (2015). NCBI viral genomes resource. Nucleic Acids Res. *43*, D571–D577.

Caccamo, A., Branca, C., Piras, I.S., Ferreira, E., Huentelman, M.J., Liang, W.S., Readhead, B., Dudley, J.T., Spangenberg, E.E., Green, K.N., et al. (2017). Necroptosis activation in Alzheimer's disease. Nat. Neurosci. *20*, 1236–1246.

Carbone, I., Lazzarotto, T., Ianni, M., Porcellini, E., Forti, P., Masliah, E., Gabrielli, L., and Licastro, F. (2014). Herpes virus in Alzheimer's disease: relation to progression of the disease. Neurobiol. Aging *35*, 122–129.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., and Werner, T. (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics *21*, 2933–2942.

Caselli, E., D'Accolti, M., Soffritti, I., Zatelli, M.C., Rossi, R., Degli Uberti, E., and Di Luca, D. (2017). HHV-6A in vitro infection of thyrocytes and T cells alters the expression of miRNA associated to autoimmune thyroiditis. Virol. J. *14*, 3.

Cen, S., Khorchid, A., Javanbakht, H., Gabor, J., Stello, T., Shiba, K., Musier-Forsyth, K., and Kleiman, L. (2001). Incorporation of lysyl-tRNA synthetase into human immunodeficiency virus type 1. J. Virol. *75*, 5043–5048.

Chang, L.K., Chung, J.Y., Hong, Y.R., Ichimura, T., Nakao, M., and Liu, S.T. (2005). Activation of Sp1-mediated transcription by Rta of Epstein-Barr virus via an interaction with MCAF1. Nucleic Acids Res. *33*, 6528–6539.

Chen, Z., Simmons, M.S., Perry, R.T., Wiener, H.W., Harrell, L.E., and Go, R.C. (2008). Genetic association of neurotrophic tyrosine kinase receptor type 2 (NTRK2) With Alzheimer's disease. Am. J. Med. Genet. B. Neuropsychiatr. Genet. *147*, 363–369.

Chen, E.Y., Xu, H., Gordonov, S., Lim, M.P., Perkins, M.H., and Ma'ayan, A. (2012). Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers. Bioinformatics *28*, 105–111.

Cohain, A., Divaraniya, A.A., Zhu, K., Scarpa, J.R., Kasarskis, A., Zhu, J., Chang, R., Dudley, J.T., and Schadt, E.E. (2017). Exploring the reproducibility of probabilistic causal molecular network models. Pac. Symp. Biocomput. *22*, 120–131.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Darmanis, S., Sloan, S.A., Zhang, Y., Enge, M., Caneda, C., Shuer, L.M., Hayden Gephart, M.G., Barres, B.A., and Quake, S.R. (2015). A survey of human brain transcriptome diversity at the single cell level. Proc. Natl. Acad. Sci. USA *112*, 7285–7290.

Dickson, D.W., Crystal, H.A., Mattiace, L.A., Masur, D.M., Blau, A.D., Davies, P., Yen, S.H., and Aronson, M.K. (1992). Identification of normal and pathological aging in prospectively studied nondemented elderly humans. Neurobiol. Aging *13*, 179–189.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Epel, E.S., Puterman, E., Lin, J., Blackburn, E.H., Lum, P.Y., Beckmann, N.D., Zhu, J., Lee, E., Gilbert, A., Rissman, R.A., et al. (2016). Meditation and vacation effects have an impact on disease-associated molecular phenotypes. Transl. Psychiatry *6*, e880.

Franzén, O., Ermel, R., Cohain, A., Akers, N.K., Di Narzo, A., Talukdar, H.A., Foroughi-Asl, H., Giambartolomei, C., Fullard, J.F., Sukhavasi, K., et al. (2016). Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. Science *353*, 827–830.

Garant, J.M., Luce, M.J., Scott, M.S., and Perreault, J.P. (2015). G4RNA: an RNA G-quadruplex database. Database (Oxford) *2015*, bav059.

Gatto, G., Rossi, A., Rossi, D., Kroening, S., Bonatti, S., and Mallardo, M. (2008). Epstein-Barr virus latent membrane protein 1 trans-activates miR-155 transcription through the NF-kappaB pathway. Nucleic Acids Res. *36*, 6608–6619.

Gómez-Isla, T., Price, J.L., McKeel, D.W., Jr., Morris, J.C., Growdon, J.H., and Hyman, B.T. (1996). Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. J. Neurosci. *16*, 4491–4500.

Gottwein, E., Mukherjee, N., Sachse, C., Frenzel, C., Majoros, W.H., Chi, J.T., Braich, R., Manoharan, M., Soutschek, J., Ohler, U., and Cullen, B.R. (2007). A viral microRNA functions as an orthologue of cellular miR-155. Nature *450*, 1096–1099.

Grady, S.L., Hwang, J., Vastag, L., Rabinowitz, J.D., and Shenk, T. (2012). Herpes simplex virus 1 infection activates poly(ADP-ribose) polymerase and triggers the degradation of poly(ADP-ribose) glycohydrolase. J. Virol. *86*, 8259–8268.

Gusmao, E.G., Allhoff, M., Zenke, M., and Costa, I.G. (2016). Analysis of computational footprinting methods for DNase sequencing experiments. Nat. Methods *13*, 303–309.

Ha, H.C., Juluri, K., Zhou, Y., Leung, S., Hermankova, M., and Snyder, S.H. (2001). Poly(ADP-ribose) polymerase-1 is required for efficient HIV-1 integration. Proc. Natl. Acad. Sci. USA *98*, 3364–3368.

Haroutunian, V., Perl, D.P., Purohit, D.P., Marin, D., Khan, K., Lantz, M., Davis, K.L., and Mohs, R.C. (1998). Regional distribution of neuritic plaques in the nondemented elderly and subjects with very mild Alzheimer disease. Arch. Neurol. *55*, 1185–1191.

Harrich, D., Garcia, J., Wu, F., Mitsuyasu, R., Gonazalez, J., and Gaynor, R. (1989). Role of SP1-binding domains in in vivo transcriptional regulation of the human immunodeficiency virus type 1 long terminal repeat. J. Virol. *63*, 2585–2591.

Hauw, J.J., Daniel, S.E., Dickson, D., Horoupian, D.S., Jellinger, K., Lantos, P.L., McKee, A., Tabaton, M., and Litvan, I. (1994). Preliminary NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy). Neurology *44*, 2015–2019.

Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.

Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. G3 (Bethesda) *1*, 457–470.

Hsu, S.D., Lin, F.M., Wu, W.Y., Liang, C., Huang, W.C., Chan, W.L., Tsai, W.T., Chen, G.Z., Lee, C.J., Chiu, C.M., et al. (2011). miRTarBase: a database curates experimentally validated microRNA-target interactions. Nucleic Acids Res. *39*, D163–D169.

Huang, Y., Tanimukai, H., Liu, F., Iqbal, K., Grundke-Iqbal, I., and Gong, C.X. (2004). Elevation of the level and activity of acid ceramidase in Alzheimer's disease brain. Eur. J. Neurosci. *20*, 3489–3497.

Hyman, B.T., Van Hoesen, G.W., Damasio, A.R., and Barnes, C.L. (1984). Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. Science *225*, 1168–1170.

Itzhaki, R.F. (2014). Herpes simplex virus type 1 and Alzheimer's disease: increasing evidence for a major role of the virus. Front. Aging Neurosci. *6*, 202.

Itzhaki, R.F., Lathe, R., Balin, B.J., Ball, M.J., Bearer, E.L., Braak, H., Bullido, M.J., Carter, C., Clerici, M., Cosby, S.L., et al. (2016). Microbes and Alzheimer's Disease. J. Alzheimers Dis. *51*, 979–984.

Jankowsky, J.L., Fadale, D.J., Anderson, J., Xu, G.M., Gonzales, V., Jenkins, N.A., Copeland, N.G., Lee, M.K., Younkin, L.H., Wagner, S.L., et al. (2004). Mutant presenilins specifically elevate the levels of the 42 residue beta-amyloid peptide in vivo: evidence for augmentation of a 42-specific gamma secretase. Hum. Mol. Genet. *13*, 159–170.

Kaddurah-Daouk, R., Rozen, S., Matson, W., Han, X., Hulette, C.M., Burke, J.R., Doraiswamy, P.M., and Welsh-Bohmer, K.A. (2011). Metabolomic changes in autopsy-confirmed Alzheimer's disease. Alzheimers Dement. *7*, 309–317.

Kaddurah-Daouk, R., Zhu, H., Sharma, S., Bogdanov, M., Rozen, S.G., Matson, W., Oki, N.O., Motsinger-Reif, A.A., Churchill, E., Lei, Z., et al.; Pharmacometabolomics Research Network (2013). Alterations in metabolic pathways and networks in Alzheimer's disease. Transl. Psychiatry *3*, e244.

Kersey, P.J., Allen, J.E., Armean, I., Boddu, S., Bolt, B.J., Carvalho-Silva, D., Christensen, M., Davis, P., Falin, L.J., Grabmueller, C., et al. (2016). Ensembl Genomes 2016: more genomes, more complexity. Nucleic Acids Res. *44* (D1), D574–D580.

Krasemann, S., Madore, C., Cialic, R., Baufeld, C., Calcagno, N., El Fatimy, R., Beckers, L., O'Loughlin, E., Xu, Y., Fanek, Z., et al. (2017). The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. Immunity *47*, 566–581 e9.

Kumar, D.K., Choi, S.H., Washicosky, K.J., Eimer, W.A., Tucker, S., Ghofrani, J., Lefkowitz, A., McColl, G., Goldstein, L.E., Tanzi, R.E., and Moir, R.D. (2016). Amyloid-β peptide protects against microbial infection in mouse and worm models of Alzheimer's disease. Sci. Transl. Med. *8*, 340ra72.

Lachmann, A., Xu, H., Krishnan, J., Berger, S.I., Mazloom, A.R., and Ma'ayan, A. (2010). ChEA: transcription factor regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics *26*, 2438–2444.

Lachmann, A., Torre, D., Keenan, A.B., Jagodnik, K.M., Lee, H.J., Wang, L., Silverstein, M.C., and Ma'ayan, A. (2018). Massive mining of publicly available RNA-seq data from human and mouse. Nat. Commun. *9*, 1366.

Lambert, J.C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., DeStafano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., et al.; European Alzheimer's Disease Initiative (EADI); Genetic and Environmental Risk in Alzheimer's Disease; Alzheimer's Disease Genetic Consortium; Cohorts for Heart and Aging Research in Genomic Epidemiology (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. *45*, 1452–1458.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al.; International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Lee, S., Wu, M.C., and Lin, X. (2012). Optimal tests for rare variant effects in sequencing association studies. Biostatistics *13*, 762–775.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Liang, W.S., Dunckley, T., Beach, T.G., Grover, A., Mastroeni, D., Walker, D.G., Caselli, R.J., Kukull, W.A., McKeel, D., Morris, J.C., et al. (2007). Gene expression profiles in anatomically and functionally distinct regions of the normal aged human brain. Physiol. Genomics *28*, 311–322.

Liang, W.S., Dunckley, T., Beach, T.G., Grover, A., Mastroeni, D., Ramsey, K., Caselli, R.J., Kukull, W.A., McKeel, D., Morris, J.C., et al. (2008). Altered neuronal gene expression in brain regions differentially affected by Alzheimer's disease: a reference data set. Physiol. Genomics *33*, 240–256.

Liang, W.S., Dunckley, T., Beach, T.G., Grover, A., Mastroeni, D., Ramsey, K., Caselli, R.J., Kukull, W.A., McKeel, D., Morris, J.C., et al. (2010). Neuronal gene expression in non-demented individuals with intermediate Alzheimer's Disease neuropathology. Neurobiol. Aging *31*, 549–566.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Res. *41*, e108.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Lin, W.R., Wozniak, M.A., Cooper, R.J., Wilcock, G.K., and Itzhaki, R.F. (2002). Herpesviruses in brain and Alzheimer's disease. J. Pathol. *197*, 395–402.

Liu, H.P., Lin, W.Y., Wu, B.T., Liu, S.H., Wang, W.F., Tsai, C.H., Lee, C.C., and Tsai, F.J. (2010a). Evaluation of the poly(ADP-ribose) polymerase-1 gene variants in Alzheimer's disease. J. Clin. Lab. Anal. *24*, 182–186.

Liu, J.Z., McRae, A.F., Nyholt, D.R., Medland, S.E., Wray, N.R., Brown, K.M., Hayward, N.K., Montgomery, G.W., Visscher, P.M., Martin, N.G., and Macgregor, S.; AMFS Investigators (2010b). A versatile gene-based test for genome-wide association studies. Am. J. Hum. Genet. *87*, 139–145.

Lövheim, H., Gilthorpe, J., Adolfsson, R., Nilsson, L.G., and Elgh, F. (2015a). Reactivated herpes simplex infection increases the risk of Alzheimer's disease. Alzheimers Dement. *11*, 593–599.

Lövheim, H., Gilthorpe, J., Johansson, A., Eriksson, S., Hallmans, G., and Elgh, F. (2015b). Herpes simplex infection and the risk of Alzheimer's disease: A nested case-control study. Alzheimers Dement. *11*, 587–592.

Madeira, C., Lourenco, M.V., Vargas-Lopes, C., Suemoto, C.K., Brandão, C.O., Reis, T., Leite, R.E., Laks, J., Jacob-Filho, W., Pasqualucci, C.A., et al. (2015). d-serine levels in Alzheimer's disease: implications for novel biomarker development. Transl. Psychiatry *5*, e561.

Mak, J., and Kleiman, L. (1997). Primer tRNAs for reverse transcription. J. Virol. *71*, 8087–8095.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770.

Mastroeni, D., Nolz, J., Sekar, S., Delvaux, E., Serrano, G., Cuyugan, L., Liang, W.S., Beach, T.G., Rogers, J., and Coleman, P.D. (2018). Laser-captured microglia in the Alzheimer's and Parkinson's brain reveal unique regional expression profiles and suggest a potential role for hepatitis B in the Alzheimer's brain. Neurobiol. Aging *63*, 12–21.

Mattiussi, S., Tempera, I., Matusali, G., Mearini, G., Lenti, L., Fratarcangeli, S., Mosca, L., D'Erme, M., and Mattia, E. (2007). Inhibition of Poly(ADP-ribose)polymerase impairs Epstein Barr Virus lytic cycle progression. Infect. Agent. Cancer *2*, 18.

Middleton, P.J., Petric, M., Kozak, M., Rewcastle, N.B., and McLachlan, D.R. (1980). Herpes-simplex viral genome and senile and presenile dementias of Alzheimer and Pick. Lancet *1*, 1038.

Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R., and Horvath, S. (2011). Strategies for aggregating gene expression data: the collapseRows R function. BMC Bioinformatics *12*, 322.

Millstein, J., Zhang, B., Zhu, J., and Schadt, E.E. (2009). Disentangling molecular relationships with a causal inference test. BMC Genet. *10*, 23.

Mirra, S.S., Heyman, A., McKeel, D., Sumi, S.M., Crain, B.J., Brownlee, L.M., Vogel, F.S., Hughes, J.P., van Belle, G., and Berg, L. (1991). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. Neurology *41*, 479–486.

Mishra, A., and Macgregor, S. (2015). VEGAS2: Software for More Flexible Gene-Based Testing. Twin Res. Hum. Genet. *18*, 86–91.

Mori, Y., Yagi, H., Shimamoto, T., Isegawa, Y., Sunagawa, T., Inagi, R., Kondo, K., Tano, Y., and Yamanishi, K. (1998). Analysis of human herpesvirus 6 U3 gene, which is a positional homolog of human cytomegalovirus UL 24 gene. Virology *249*, 129–139.

Morris, J.C. (1993). The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology *43*, 2412–2414.

Murat, P., Zhong, J., Lekieffre, L., Cowieson, N.P., Clancy, J.L., Preiss, T., Balasubramanian, S., Khanna, R., and Tellam, J. (2014). G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. Nat. Chem. Biol. *10*, 358–364.

Murphy, J.V., and Yunis, E.J. (1976). Encephalopathy following measles infection in children with chronic illness. J. Pediatr. *88*, 937–942.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Nishida, K., Kawasaki, T., Fujie, M., Usami, S., and Yamada, T. (1999). Aminoacylation of tRNAs encoded by Chlorella virus CVK2. Virology *263*, 220–229.

Nixon, R.A., Wegiel, J., Kumar, A., Yu, W.H., Peterhoff, C., Cataldo, A., and Cuervo, A.M. (2005). Extensive involvement of autophagy in Alzheimer disease: an immuno-electron microscopy study. J. Neuropathol. Exp. Neurol. *64*, 113–122.

Nousiainen, L., Sillanpää, M., Jiang, M., Thompson, J., Taipale, J., and Julkunen, I. (2013). Human kinome analysis reveals novel kinases contributing to virus infection and retinoic-acid inducible gene I-induced type I and type III IFN gene expression. Innate Immun. *19*, 516–530.

Nygaard, H.B. (2018). Targeting Fyn Kinase in Alzheimer's Disease. Biol. Psychiatry *83*, 369–376.

Papadopoulou, C., Guilbaud, G., Schiavone, D., and Sale, J.E. (2015). Nucleotide pool depletion induces G-quadruplex-dependent perturbation of gene expression. Cell Rep. *13*, 2491–2503.

Pearl, J. (2009). Causality (Cambridge University Press).

Pearl, J.R., Bergey, D.E., Funk, C.C., Basu, B., Oshone, R., Shannon, P., Hood, L., Price, N.D., Colantuoni, C., and Ament, S.A. (2017). Genome-scale transcriptional regulatory network models of psychiatric and neurodegenerative disorders. bioRxiv. https://doi.org/10.1101/190959.

Piper, J., Elze, M.C., Cauchy, P., Cockerill, P.N., Bonifer, C., and Ott, S. (2013). Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. Nucleic Acids Res. *41*, e201.

Prusty, B.K., Krohne, G., and Rudel, T. (2013). Reactivation of chromosomally integrated human herpesvirus-6 by telomeric circle formation. PLoS Genet. *9*, e1004033.

Prusty, B.K., Gulve, N., Rasa, S., Murovska, M., Hernandez, P.C., and Ablashi, D.V. (2017). Possible chromosomal and germline integration of human herpesvirus 7. J. Gen. Virol. *98*, 266–274.

Rampelli, S., Soverini, M., Turroni, S., Quercia, S., Biagi, E., Brigidi, P., and Candela, M. (2016). ViromeScan: a new tool for metagenomic viral community profiling. BMC Genomics *17*, 165.

Readhead, B., Haure-Mirande, J.V., Zhang, B., Haroutunian, V., Gandy, S., Schadt, E.E., Dudley, J.T., and Ehrlich, M.E. (2016). Molecular systems evaluation of oligomerogenic APP(E693Q) and fibrillogenic APP(KM670/671NL)/PSEN1(Δexon9) mouse models identifies shared features with human Alzheimer's brain molecular pathology. Mol. Psychiatry *21*, 1099–1111.

Rhodes, D., and Lipps, H.J. (2015). G-quadruplexes and their regulatory roles in biology. Nucleic Acids Res. *43*, 8627–8637.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rode, A.B., Endoh, T., and Sugimoto, N. (2016). tRNA shifts the G-quadruplex-hairpin conformational equilibrium in RNA towards the hairpin conformer. Angew. Chem. Int. Ed. Engl. *55*, 14315–14319.

Rotmistrovsky, K., and Agarwala, R. (2011). BMTagger: Best Match Tagger for removing human reads from metagenomics datasets. ftp://ftp.ncbi.nlm.nih.gov/pub/agarwala/bmtagger/.

Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) *2016*, 2016.

Sang, Y., Tait, A.R., Scott, W.R., Creagh, A.L., Kumar, P., Haynes, C.A., and Straus, S.K. (2014). Probing the interaction between U24 and the SH3 domain of Fyn tyrosine kinase. Biochemistry *53*, 6092–6102.

Scholz, M., Vogel, J.U., Höver, G., Prösch, S., Kotchetkov, R., Cinatl, J., Koch, F., Doerr, H.W., and Cinatl, J., Jr. (2004). Thrombin induces Sp1-mediated antiviral effects in cytomegalovirus-infected human retinal pigment epithelial cells. Med. Microbiol. Immunol. (Berl.) *193*, 195–203.

Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics *28*, 1353–1358.

Shugar, D. (1999). Viral and host-cell protein kinases: enticing antiviral targets and relevance of nucleoside, and viral thymidine, kinases. Pharmacol. Ther. *82*, 315–335.

Sigurðsson, B. (1954). Rida, a chronic encephalitis of sheep: with general remarks on infections which develop slowly and some of their special characteristics. Brit. Vet. J. *110*, 341–344.

Sjogren, T., Sjogren, H., and Lindgren, A.G. (1952). Morbus Alzheimer and morbus Pick; a genetic, clinical and patho-anatomical study. Acta Psychiatr. Neurol. Scand., Suppl. *82*, 1–152.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, e3.

Song, J., and Lee, J.E. (2015). miR-155 is involved in Alzheimer's disease by regulating T lymphocyte function. Front. Aging Neurosci. *7*, 61.

Soscia, S.J., Kirby, J.E., Washicosky, K.J., Tucker, S.M., Ingelsson, M., Hyman, B., Burton, M.A., Goldstein, L.E., Duong, S., Tanzi, R.E., and Moir, R.D. (2010). The Alzheimer's disease-associated amyloid beta-protein is an antimicrobial peptide. PLoS ONE *5*, e9505.

Sperling, R.A., Aisen, P.S., Beckett, L.A., Bennett, D.A., Craft, S., Fagan, A.M., Iwatsubo, T., Jack, C.R., Jr., Kaye, J., Montine, T.J., et al. (2011). Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement. *7*, 280–292.

Su, J.H., Anderson, A.J., Cummings, B.J., and Cotman, C.W. (1994). Immunohistochemical evidence for apoptosis in Alzheimer's disease. Neuroreport *5*, 2529–2533.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550.

Swanson, C.M., Sherer, N.M., and Malim, M.H. (2010). SRp40 and SRp55 promote the translation of unspliced human immunodeficiency virus type 1 RNA. J. Virol. *84*, 6748–6759.

Tanaka-Taya, K., Sashihara, J., Kurahashi, H., Amo, K., Miyagawa, H., Kondo, K., Okada, S., and Yamanishi, K. (2004). Human herpesvirus 6 (HHV-6) is transmitted from parent to child in an integrated form and characterization of cases with chromosomally integrated HHV-6 DNA. J. Med. Virol. *73*, 465–473.

Thai, T.H., Calado, D.P., Casola, S., Ansel, K.M., Xiao, C., Xue, Y., Murphy, A., Frendewey, D., Valenzuela, D., Kutok, J.L., et al. (2007). Regulation of the germinal center response by microRNA-155. Science *316*, 604–608.

Toribara, N.W., Roberton, A.M., Ho, S.B., Kuo, W.L., Gum, E., Hicks, J.W., Gum, J.R., Jr., Byrd, J.C., Siddiki, B., and Kim, Y.S. (1993). Human gastric mucin. Identification of a unique species by expression cloning. J. Biol. Chem. *268*, 5879–5885.

Upton, J.W., and Chan, F.K. (2014). Staying alive: cell death in antiviral immunity. Mol. Cell *54*, 273–280.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 1–33, 33.

Vidal, R., Frangione, B., Rostagno, A., Mead, S., Révész, T., Plant, G., and Ghiso, J. (1999). A stop-codon mutation in the BRI gene associated with familial British dementia. Nature *399*, 776–781.

Vidal, R., Revesz, T., Rostagno, A., Kim, E., Holton, J.L., Bek, T., Bojsen-Møller, M., Braendgaard, H., Plant, G., Ghiso, J., and Frangione, B. (2000). A decamer duplication in the 3′ region of the BRI gene originates an amyloid peptide that is associated with dementia in a Danish kindred. Proc. Natl. Acad. Sci. USA *97*, 4920–4925.

Westman, G., Blomberg, J., Yun, Z., Lannfelt, L., Ingelsson, M., and Eriksson, B.M. (2017). Decreased HHV-6 IgG in Alzheimer's disease. Front. Neurol. *8*, 40.

Wong, J., Higgins, M., Halliday, G., and Garner, B. (2012). Amyloid beta selectively modulates neuronal TrkB alternative transcript expression with implications for Alzheimer's disease. Neuroscience *210*, 363–374.

Woodbury, M.E., Freilich, R.W., Cheng, C.J., Asai, H., Ikezu, S., Boucher, J.D., Slack, F., and Ikezu, T. (2015). miR-155 is essential for inflammation-induced hippocampal neurogenic dysfunction. J. Neurosci. *35*, 9764–9781.

Wu, B., and Pankow, J.S. (2016). Sequence kernel association test of multiple continuous phenotypes. Genet. Epidemiol. *40*, 91–100.

Yoo, S., Takikawa, S., Geraghty, P., Argmann, C., Campbell, J., Lin, L., Huang, T., Tu, Z., Foronjy, R.F., Spira, A., et al. (2015). Integrative analysis of DNA methylation and gene expression data identifies EPAS1 as a key regulator of COPD. PLoS Genet. *11*, e1004898.

Zaharia, M., Bolosky, W.J., Curtis, K., Fox, A., Patterson, D., Shenker, S., Stoica, I., Karp, R.M., and Sittler, T. (2011). Faster and more accurate sequence alignment with SNAP. arXiv, arXiv:1111.5572, https://arxiv.org/abs/1111.5572.

Zhang, Y., de Bolle, L., Aquaro, S., van Lommel, A., De Clercq, E., and Schols, D. (2001). Productive infection of primary macrophages with human herpesvirus 7. J. Virol. *75*, 10511–10514.

Zhang, B., Gaiteri, C., Bodea, L.G., Wang, Z., McElwee, J., Podtelezhnikov, A.A., Zhang, C., Xie, T., Tran, L., Dobrin, R., et al. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell *153*, 707–720.

Zhao, Y., Yao, Y., Xu, H., Lambeth, L., Smith, L.P., Kgosana, L., Wang, X., and Nair, V. (2009). A functional MicroRNA-155 ortholog encoded by the oncogenic Marek's disease virus. J. Virol. *83*, 489–492.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-Iba1	Wako	Cat#019-19741; lot# WDK2121; RRID: AB_2665520
Mouse monoclonal anti-β-Amyloid (6E10)	Covance	Cat#SIG-39320; lot#D14FF01323
anti-rabbit Alexa 488	Thermo Fisher Scientific	Cat#A-11008; RRID: AB_143165
anti-mouse Alexa 594	Thermo Fisher Scientific	Cat#A-11005; RRID: AB_2534073
Deposited Data		
Preclinical AD gene expression microarrays	Liang et al., 2010	GEO: GSE9770
Control gene expression microarrays	Liang et al., 2007	GEO: GSE5281
Viral nucleotide sequences and genomic feature annotations	Brister et al., 2015	https://www.ncbi.nlm.nih.gov/genome/viruses/
GRCh37 ensembl reference sequences and gene annotations	Kersey et al., 2016	https://uswest.ensembl.org/info/data/ftp/index.html
MSBB RNA-sequences	https://www.synapse.org/#!Synapse:syn3159438	syn3157743
MAYO TCX RNA-sequences	Allen et al., 2016	syn8612203
ROSMAP RNA-sequences	Bennett et al., 2012a, Bennett et al., 2012b	syn8612097
MSBB Whole exome sequences	https://www.synapse.org/#!Synapse:syn3159438	syn4645334
MSBB Proteomics	https://www.synapse.org/#!Synapse:syn3159438	syn5759470
IGAP Alzheimer's disease GWAS summary statistics	Lambert et al., 2013	http://web.pasteur-lille.fr/en/recherche/u744/igap/igap_download.php
Brain single cell RNA-sequencing data	Darmanis et al., 2015	GSE67835
1000 Genomes Phase 3 reference genotypes	1000 Genomes Phase 3 reference genotypes	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase3/
UCSC hg19 "Known genes" table	Lander et al., 2001	http://hgdownload.cse.ucsc.edu/goldenPath/hg19
Viral abundance estimates and processing scripts to reproduce results	This paper	https://www.synapse.org/#!Synapse:syn12177270
Experimental Models: Organisms/Strains		
Mouse: B6.Cg-Mir155tm1Rsky/J	The Jackson Laboratory	Cat#: 007745; RRID: IMSR_JAX:007745
Mouse: B6.Cg-Tg(APPswe,PSEN1dE9) 85Dbo/Mmjax	The Jackson Laboratory	Cat#: 34832-JAX; RRID: MMRRC_034832-JAX
Mouse: C57BL/6J	The Jackson Laboratory	Cat #000664 RRID: IMSR_JAX:000664
Software and Algorithms		
MatInspector	Cartharius et al., 2005	https://www.genomatix.de/online_help/help_matinspector/matinspector_help.html
R	R Core Team	https://cran.r-project.org
Viromescan	Rampelli et al., 2016	https://sourceforge.net/projects/viromescan/
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2
STAR-RNaseq (2.4.0g1) read aligner	Dobin et al., 2013	https://github.com/alexdobin/STAR
CIT	Millstein et al., 2009	https://cran.r-project.org/package=cit
VEGAS	Liu et al., 2010b; Mishra and Macgregor, 2015	https://bioconductor.org/packages/cpvSNP/
CIBERSORT	Newman et al., 2015	https://cibersort.stanford.edu/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
miRTarBase	Hsu et al., 2011	http://mirtarbase.mbc.nctu.edu.tw/php/download.php
SKAT	Lee et al., 2012; Wu and Pankow, 2016	https://cran.r-project.org/package=SKAT
Limma	Law et al., 2014; Ritchie et al., 2015	https://bioconductor.org/packages/limma/
MatrixEQTL	Shabalin, 2012	https://cran.r-project.org/package=MatrixEQTL
Subread	Liao et al., 2013; Liao et al., 2014	https://sourceforge.net/projects/subread/
Jellyfish	Marçais and Kingsford, 2011	http://www.genome.umd.edu/jellyfish.html
BMtagger	Rotmistrovsky and Agarwala, 2011	https://sourceforge.net/projects/viromescan/
Samtools	Li et al., 2009	https://github.com/samtools/samtools
BLAST	Altschul et al., 1990	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/
edgeR	Robinson et al., 2010	https://bioconductor.org/packages/edgeR/
IMPUTE2	Howie et al., 2009, 2011	https://mathgen.stats.ox.ac.uk/impute/impute_v2.html#download

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Joel T. Dudley (joel.dudley@mssm.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

APP$^{KM670/671NL/PSEN1\Delta exon9}$ (*APP/PS1*) (Jankowsky et al., 2004), and miR-155-KO (Thai et al., 2007) mice were obtained from Jackson Laboratories. *APP*$^{KM670/671NL/PSEN1\Delta exon9}$ were crossed with miR-155-KO mice to obtain *APP*$^{KM670/671NL/PSEN1\Delta exon9}$ heterozygous or KO for miR-155. Mouse lines were maintained on a C57Bl6/J background. 4-month-old male and female mice were sacrificed by decapitation. Brains were dissected into right and left hemispheres. One hemisphere was collected and fixed in 4% paraformaldehyde for immunohistochemistry analysis. The second hemisphere was dissected and prefrontal cortex (PC) was collected for transcriptomic analysis. PC and hemibrains were then snap-frozen and stored at −80°C prior to RNA isolation or biochemistry analysis. Male and female mice were used in the experiments. The experimental procedures were conducted in accordance with NIH guidelines for animal research and were approved by the Institutional Animal Care and Use Committee (IACUC) at Icahn School of Medicine at Mount Sinai.

METHOD DETAILS

RNA sequencing human gene expression

RNA sequencing data was obtained from the Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) Knowledge Portal (MSSB synapse ID: syn3157743, MAYO TCX: syn8612203, ROS/MAP: syn8612097). For the MSBB cohort, post-mortem samples were collected from the STG (Brodmann Area 22), APFC (Brodmann Area 10), PHG (Brodmann Area 36) and IFG (Brodmann Area 44) by the Mount Sinai NIH Brain and Tissue Repository. MAYO TCX cohort samples were collected from the temporal cortex, as previously described (Allen et al., 2016). ROS and MAP cohort samples were collected from the DLPFC as previously described (Bennett et al., 2012a, 2012b). For all cohorts, RNA-seq samples with RIN less than 6, were removed from the analysis.

For host gene expression (MSSB cohort), single end reads were aligned to human genome reference (GRCh37 ensembl version 70 (Kersey et al., 2016)), using STAR-RNaseq (2.4.0g1) read aligner (Dobin et al., 2013), and accepted mapped reads were summarized to gene level counts using the featureCounts function of the subread software package (Liao et al., 2013, 2014). Genes with at least 1 count per million mapped reads (Robinson et al., 2010) in at least half of the sample libraries were retained, and normalized using the voom function in the Limma package (Law et al., 2014; Ritchie et al., 2015).

Whole exome sequencing

Whole exome sequencing data (MSBB) used in this study can be obtained from the Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) Knowledge Portal (synapse ID: syn4645334). Reads were aligned to human genome hg19 using BWA aligner (Li and Durbin, 2009). DNA sequence variants were called using the DNaseq Variant Analysis workflow of GATK Best Practices version 3 (Van der Auwera et al., 2013). Variants with a minor allele frequency <0.05, or with missing calls in >10 samples were removed from further analysis. Common variants were imputed using IMPUTE2 (Howie et al., 2009, 2011) using 1000 Genomes Phase 3 reference genotypes (1000 Genomes Project Consortium et al., 2015).

Liquid Chromatography tandem mass spectrometry

Proteomics data (MSBB) used in this study can be obtained from the Accelerating Medicines Partnership - Alzheimer's Disease (AMP-AD) Knowledge Portal (synapse ID: syn5759470). All samples were from a single brain region (APFC), and underwent Liquid Chromatography tandem Mass Spectrometry (LC-MS/MS) and MaxQuant (Cox and Mann, 2008) was used to quantify label free protein counts. Protein counts were normalized by the total counts detected for that sample, and log transformed (log2) using an offset of 0.25 for null values.

Immunohistochemistry

Paraformaldehyde-fixed mouse brains were cut (30μm thick) with a vibratome VT1000S (Leica Microsystems, Germany). Sagittal free-floating sections were pre-treated with 70% formic acid for 15min at room temperature. Sections were then blocked for 1h at room temperature (PBS with 0.1% v/v Tween-20 and 10% goat serum) and incubated overnight with anti-Iba1 (1:500;Wako, Richmond, VA) and 6E10 (1:1000;Covance, Princeton, NJ) antibodies (PBS with 0.1% v/v Tween-20 and 1% goat serum). Sections were then incubated for 1h with fluorescent conjugated secondary antibodies (PBS with 0.1% v/v Tween-20 and 1% goat serum) (anti-rabbit Alexa 488 (1:400) for Iba1 and anti-mouse Alexa 568 (1:400) for 6E10) and mounted in Superfrost Plus slides. Images were acquired on an Olympus BX61 microscope with an attached Olympus DP71 camera.

Mouse RNA isolation and library preparation

RNA isolation and library preparation were performed as previously described (Readhead et al., 2016). Briefly, snap frozen samples from male and female mice were homogenized in QIAzol Lysis Reagent (QIAGEN). Total RNA purification was performed with the miRNeasy Mini kit (QIAGEN), according to the manufacturer's instructions. RNA quantification and quality were evaluated by Agilent BioAnalyzer and processed for RNA library preparation. RNA integrity was checked by either the Fragment Analyzer (Advanced Analytical, IA, USA) or the 2100 Bioanalyzer using the RNA 6000 Nano assay (Agilent, CA, USA). All processed total RNA samples had RQN/RIN value of 8.8 or greater.

QUANTIFICATION AND STATISTICAL ANALYSIS

Construction of entorhinal cortex and hippocampal preclinical AD and control regulatory networks

Preclinical AD and Control networks were generated from post-mortem microarray gene expression data downloaded from Gene Expression Omnibus (Preclinical AD: accession GSE9770 (Liang et al., 2010), Control accession: GSE5281 (Liang et al., 2007)). In both datasets, samples were derived from layer 2 entorhinal cortex neurons, and hippocampal CA1 pyramidal neurons. For each group, we built single-tissue networks (HIP and EC), as well as a separate cross-tissue HIP-EC network, with the goal of combining these three individual networks afterward into a single union network that is able to capture intra-tissue, as well as inter-tissue connectivity.

Networks were constructed using a modification of the "inductive causation with latent variables" procedure (Pearl, 2009). This approach is a constraint-based method for building causal networks, using conditional dependence tests between nodes to first identify a network skeleton (the set of undirected edges), learn v-structures (a triplet of nodes where two non-adjacent nodes share a target node), and output a partially directed graph (PDG). The final PDG contains four types of edges: undirected (indicating uncertain causation), bi-directed (spurious or latent causation), directed (that directed edge is present in at least one Markov equivalent graph, and none contain the opposite arrow for that edge), and marked-directed edges (that directed edge exists in every Markov-equivalent graph). For each network, we retained all edges marked as directed, or marked-directed in the PDG.

Each of the networks we generated was constructed across the top 15,000 most differentially expressed probes within that brain region (Preclinical AD versus Controls). For the HIP-EC network, we combined HIP and EC samples from each donor, into a single expression set, comprising the top 7,500 most differentially expressed probes from each tissue. We constructed the final preclinical AD and control networks by combining discovered edges from the EC, HIP and HIP-EC networks into a final structure comprising 30,000 tissue-annotated probes. We only included HIP-EC cross-tissue edges if the edge was not present as an intra-tissue edge in either of the HIP or EC networks.

We then identified network probes with unusually high downstream influence, iterating over Gene Ontology (GO) biological process terms, and inducing connected subgraphs from our network, according to their annotation with that specific term. Within that subgraph (containing a minimum of two connected probes) for each probe, we identified the number of probes within the network diameter of the induced subgraph that can be reached via incoming connections, and outgoing connections respectively. A GO term score

was calculated for each probe, based on the difference between Z-scores of total network connectivity. A final network driver score was calculated based on the scaled sum of individual scores associated for all GO biological process terms. More formally, the driver score for probe (P) is the sum of individual differences between upstream neighborhood size and downstream neighborhood size, for each subgraph (G) induced according to each GO biological process term:

$$Driver\ Score(P) = \sum_{G=1}^{N} (Z_{Downstream}(PG) - Z_{Upstream}(PG))$$

We converted these driver scores to standard scores, and nominated probes with a Z-score $\geq$ 2 as network drivers. Drivers that were seen in the context of the preclinical AD network (but not the control network) were classified as "Gained in preclinical AD," and those seen only in the control network, were classified as "Lost in preclinical AD."

G-quadruplex sequence prediction

G4 sequence prediction within human genes was based on pattern matching against the human reference genome (hg19). Genomic coordinates for each gene in the UCSC hg19 "Known genes" were used to download nucleotide sequences for gene promoter regions, 5′-UTR, introns, exons and 3′-UTR. Promoters were defined as 2000 base pairs upstream and downstream of the transcription start site. We used regular expressions to identify the genomic coordinates corresponding to occurrences of four runs of at least three guanine bases, interspersed with between one to seven other bases (including guanine) based on similar approaches(Garant et al., 2015). We used separate patterns to retrieve matches located on the coding strand (G(3,).(1,7)?G(3,).(1,7)?G(3,).(1,7)?G(3,)) and non-coding strand (C(3,).(1,7)?C(3,).(1,7)?C(3,).(1,7)?C(3,)) respectively. Normalized G4 motif density within a particular feature was calculated by dividing the number of predicted motifs by the length (in base pairs) of that feature.

Generation of viral 31-mer database

We downloaded nucleotide sequences for 515 unique viruses with humans as known or suspected hosts (Brister et al., 2015). We segmented each viral sequence into its set of unique 31mers using Jellyfish count and dump commands (Marçais and Kingsford, 2011), followed by removal of any 31mers present in multiple viral species. We then generated bowtie2 indices (Langmead and Salzberg, 2012) for the viral sequences, and the corresponding 31-mer database for use in subsequent viral mapping steps.

Detection of viral transcription in RNA and whole exome sequencing

We quantified viral transcription through a modified workflow based on ViromeScan (Rampelli et al., 2016), which proceeds as follows: preliminary alignment of fastq files to a viral reference database using bowtie2 (Langmead and Salzberg, 2012), identifying candidate viral sequences. Mapped reads were filtered through BMtagger (Rotmistrovsky and Agarwala, 2011) to remove likely human reads. Any putatively non-human reads with low quality scores were trimmed, and reads with a trimmed read length < 60 bases were discarded. Reads were again filtered using BMtagger (Rotmistrovsky and Agarwala, 2011) to remove likely bacterial reads. Filtered, trimmed, non-human reads are then mapped to the viral 31-mer database with bowtie2 (Langmead and Salzberg, 2012), using a very sensitive, local alignment, outputting all valid alignments. Bam files were then sorted and indexed, using Samtools (Li et al., 2009). BAM files were processed manually to output a single best alignment for each read (randomly outputting a single alignment in cases where multiple best alignments were found). To minimize misclassification of human reads as viral, we then performed an additional BLAST (Altschul et al., 1990) search for any 31mers with homology to the combined hg19 and cDNA sequences (Lander et al., 2001) (blastn e-value < 1e-3), and removed those 31mers from further analysis.

We generated a 31-mer count matrix for all samples, and then summarized these separately to the level of viral species, and also for individual viral genomic features. Genomic features and coordinates were derived from NCBI gene transfer format files for each sequence. Although each viral 31-mer is unique to a single viral sequence, some 31mers might occur multiple times within that sequence. In these instances, counts for 31mers were assigned to all genomic features containing that 31-mer. We merged any overlapping genomic features that also shared the same total counts in the genomic feature count matrix.

Differential abundance of viral transcription in RNA sequencing

We performed differential viral abundance analysis at the level of viral species, and separately, at the level of individual genomic features. Due to differences in the study design and availability of covariates in each of the three AD cohorts, there were slight differences in the procedure for estimating differential viral abundance, reflecting differences in approaches for the (a) classification of AD versus Control status, and (b) availability of technical covariates to incorporate into linear modeling. In the MSBB, ROS and MAP cohorts, for each of the brain regions assayed, we compared normalized viral abundance between AD cases and controls, using three different definitions of AD within each comparison. Definitions of AD were based on the multiple levels of CERAD neuropathology classification (Mirra et al., 1991), specifically we performed comparisons between "Definite AD Vs. Controls" (NP score: 2 Vs. NP score: 1), "Likely AD Vs. Controls" (NP score: 2 / 3 Vs. NP score: 1) and "Possible AD Vs. Controls" (NP score: 2 / 3 / 4 Vs. NP score: 1).

Diagnosis within the MAYO TCX cohort was based on neuropathological evaluation and classification as follows:

(1) AD was based on "Definite AD" diagnosis according to the NINCDS-ADRDA criteria and a Braak NFT stage of IV or greater.
(2) Control subjects had Braak NFT stage of $\leq$ III, CERAD neuritic and cortical plaque densities of 0 (none) or 1 (sparse) and the absence of any of the following pathologic diagnoses: AD, Parkinson's disease, dementia with Lewy Bodies, vascular dementia, PSP, motor neuron disease, corticobasal degeneration, Pick's disease, Huntington's disease, frontotemporal lobar degeneration, hippocampal sclerosis, or dementia lacking distinctive histology.
(3) PA subjects had a Braak NFT stage of $\leq$ III, CERAD neuritic and cortical plaque densities of > = 2, and the absence of the above diagnoses, as well as the absence of any dementia or mild cognitive impairment (MCI).
(4) PSP subjects were identified via semiquantitative distribution of NFT (Hauw et al., 1994) at autopsy.

Within each comparison, we retained viral features with multiple mapped reads in at least 10 samples. Viral feature counts adjusted for the total number of reads in each fastq file, and quantile normalized using the Voom function in the Limma package (Law et al., 2014; Ritchie et al., 2015). Linear models were fit for each of the viral features, for analysis of the MSBB cohort, we included covariates for: AD status, age of death, sex, ethnicity, RIN, post-mortem interval (PMI), and batch. For the ROS and MAP cohorts, we included covariates for AD status, age of death, sex, ethnicity, RIN, PMI, batch and years of education. For the MAYO TCX cohort, we included covariates for diagnosis, age of death, sex, RIN, batch and study center. Differential abundance between AD and control groups were estimated using the eBayes function (Smyth, 2004), setting robust = TRUE to minimize the effect of outliers in variance.

Viral QTL detection

Host DNA markers that associate with viral abundance (vQTL) were identified separately for each brain region. Within each region, data was included from donors with paired whole exome sequencing and RNA-seq data. Viral genomic feature counts were normalized using the same procedure used in analyzing differential viral abundance, and we used the Matrix eQTL (Shabalin, 2012) software package to identify DNA variants associated with normalized viral abundance (controlling for age, sex, ethnicity, RNA-seq batch, RIN and PMI as covariates), assuming an additive linear model for associating genotype dosage with viral abundance. We generated a distribution of null association P values for each viral feature by shuffling sample labels for viral abundance (1000 permutations), retaining the minimum association P value for that feature, across all markers. We estimated the empirical FDR by comparison of each observed association P value with a distribution of null association P values generated separately for each viral feature (1000 permutations of viral abundance labels). DNA markers with an empirical association FDR < 0.25 were classified as vQTLs for that specific viral feature in the context of that brain region.

AD GWAS enrichments for virus network eQTLs

We identified *cis*-acting (<1 MB) DNA markers that associate with gene expression (cis-eQTL) separately for each brain region. Within each region, we included data from donors with paired whole exome sequencing and RNA-seq data. We used the Matrix eQTL (Shabalin, 2012) software package to identify DNA variants associated with normalized gene expression (controlling for age, sex, ethnicity, RNA-seq batch, RIN and PMI as covariates), assuming an additive linear model for associating genotype dosage with gene expression. We classified DNA markers with an association FDR < 0.1 as cis-eQTL, and used these to define the set of markers that are collectively associated with each virus/host subnetwork. For each virus, we pooled host interactions from all four tissues, classified each according to direction (i.e., "virus to host" or "host to virus"), and sign of correlation (i.e., positively or negatively correlated with viral abundance), and then calculated enrichments for AD risk-associated loci (Lambert et al., 2013) using the versatile gene-based association study (VEGAS) approach (Liu et al., 2010b; Mishra and Macgregor, 2015).

Estimating cell type fractions from RNA-seq

We used CIBERSORT (Newman et al., 2015) to deconvolute RNA-seq samples (MSBB) into estimated fractions for major brain cell types (neurons, astrocytes, microglia, endothelial cells, oligodendrocytes and oligodendrocyte precursor cells). This requires an independent reference panel of presumed relevant cell-types, usually derived from transcriptomic studies of individual cell types. We utilized a single cell RNA-seq dataset generated by Darmanis et al. (Darmanis et al., 2015), using 138 individual samples derived from the cortex of middle-aged adults (age range: 47 – 63 years) representing a variety of brain cell types (astrocytes n = 52, neurons n = 43, oligodendrocytes n = 19, endothelial n = 15, microglia n = 7, oligodendrocyte precursor cells n = 2). Genes with at least 1 count per million mapped reads (Robinson et al., 2010) in at least n sample libraries were retained, where n = the size of the smallest group of cell types. We then subset this single cell type expression, and the MSBB expression to the set of unique gene identifiers present in both, and supplied these as inputs to CIBERSORT, identifying signature genes according to default parameters: kappa: 999, qvalue: 0.3, min: 50, max: 150, quantile normalization = disabled, and specifying 1000 permutations.

Molecular and functional enrichment analysis

Gene set enrichments for discrete groups of genes (i.e., virus-host subnetworks networks) were calculated using Fisher's exact text, and one-sided P values (to identify over-representation of genesets) were adjusted using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Gene sets used throughout the enrichment analysis were derived from a combination of publicly available

sources, such as the molecular signatures database (Subramanian et al., 2005), brain specific gene sets curated from publicly available data (Miller et al., 2011), protein-protein hubs interactor sets (Chen et al., 2012), Mirtarbase (Hsu et al., 2011) and ChipSeq based transcription factor target sets (Lachmann et al., 2010).

Transcription factor motif enrichments in the promoters of preclinical AD network drivers were calculated based on position weight matrix matching within the gene promoter for each driver (defined as the region within 2000 basepairs of the transcription start site) using MatInspector (Cartharius et al., 2005).

Inferring causal relationships between viral abundance and biomolecular, clinical and neuropathological traits

We applied a causal inference paradigm to multiple aspects of the analysis performed in this study, to determine directed relationships between viral abundance, with a variety of host traits, including molecular (e.g., gene or protein expression), clinical (clinical dementia rating) and neuropathological (Neuritic plaque density, Braak and Braak score) traits. We used a statistical framework introduced by Millstein et al. (Millstein et al., 2009), which offers a causal inference test (CIT) that tests the hypothesis that a molecule (such as the normalized abundance of a viral species, or of a specific viral genomic feature) is mediating a causal association between a DNA locus (such as a vQTL for that specific virus), and some other quantitative trait (such as the expression of host genes that are correlated with the vQTL and the viral abundance). Causal relationships can be inferred from a chain of mathematical conditions, requiring that for a given trio of loci (L), a potential causal mediator (G) and a quantitative trait (T), the following conditions must be satisfied to establish that G is a causal mediator of the association between L and T:

$$(a) \text{L and G are associated}$$

$$(b) \text{L and T are associated}$$

$$(c) \text{L is associated with G, given T}$$

$$(d) \text{L is independent of T, given G}$$

Although CIT includes tests for linkage (conditions a and b), to control the number of candidate L / G / T trios that are submitted to the CIT function, we perform multiple pre-filtering steps, which are aimed at establishing association between L and G, and L and T, before we submit a particular trio for CIT. Association between L and G is established in the course of the viral QTL analysis (described above), where we classify variants with an association FDR < 0.1 as a vQTL for that specific viral feature. If T is a molecular species (gene expression, protein expression), nominal association between L and T is established using matrix eQTL (Shabalin, 2012) retaining candidate T molecules (for that specific vQTL) with an association P value < 0.05.

Although CIT outputs what is ostensibly a P value, it is actually the highest P value of the four constituent hypothesis tests, reflecting each of the conditions required to establish causal mediation. This results in a non uniform CIT P value distribution under null conditions, which can make appropriate multiple test correction unreliable. To overcome this, we employed a permutation based approach to assess the significance of candidate causal relationships, where candidate traits (T) are randomly shuffled, separately within each genotype dosage group (0, 1 or 2) for each permutation. The false discovery rate was estimated by counting the proportion of permutations (1000 per trio) with a CIT P value lower than the test CIT p value.

To minimize the number of false positive inferences, we performed two separate tests for each candidate trio. We tested models that include the viral feature (G) as causal for the host trait (T) ("causal model") and separately, the G being regulated by the T ("reactive model"). We required that for G to be classified as regulating T, its permutation based FDR for the causal model be < 0.05, and reactive model be > 0.05. Conversely, for a G to be classified as being regulated by a T, we required that the FDR for the reactive model be < 0.05, and > 0.05 for the causal model.

Transcription Factor Regulatory Network Analysis

The conceptual framework for the transcriptional regulatory network using TReNA was described previously (Pearl et al., 2017). Briefly, DNase Hypersensitivity (DHS) fastq files from ENCODE for all available brain samples were downloaded and aligned using the SNAP method (Zaharia et al., 2011). Two alignments were performed using seed size 16 and 20 as the sequence data was typically > 50 bp in length. The peak calling algorithm F-seq was used to identify regions of open chromatin (Boyle et al., 2008). Footprinting algorithms for Wellington (Piper et al., 2013) and HINT (Gusmao et al., 2016) were generated using default parameters. For each individual gene model, footprints within the proximal promoter (+/−5 kb of the transcription start site) were considered as priors in assessing the relationship between the expression of the transcription factor and target gene. Using the R package *trena* (Ament et al., 2017), which utilizes several LASSO regression techniques, Pearson and Spearman correlation, and random forest to

prioritize a list of putative transcription factor regulators for each gene. Scores from all these approaches were scaled and projected into PCA space and their principle components added together to produce a single composite score (pcaMax). This approach was applied to each of the RNA-seq datasets generated within the AMP-AD consortium to generate networks used for the virus / host analyses herein.

Kinase enrichment analysis of candidate virus-TF associations were performed as described in "Molecular and functional enrichment analysis," while subsetting the gene background to the set of 569 TF within the scope of the TF networks.

DATA AND SOFTWARE AVAILABILITY

Viral abundance estimates, and scripts to reproduce results can be accessed at: https://www.synapse.org/#!Synapse:syn12177270.