The transcriptional landscape of Venezuelan equine encephalitis virus infection

Zhiyuan Yao[1, 2,*], Fabio Zanini[3,4,*], Sathish Kumar[1, 2], Nuttada Panpradist[1,2,3,5], Avery Muniz[1, 2], Stephen R. Quake[3,6,7], and Shirit Einav[1, 2,]

*: these authors contributed equally

[1]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA

[2]Division of Infectious Diseases and Geographic Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

[3]Department of Bioengineering, Stanford University, Stanford, CA, USA

[4]Lowy Cancer Research Centre, University of New South Wales, Sydney, Australia

[5]Department of Bioengineering, University of Washington, Seattle, WA, USA

[6]Chan Zuckerberg Biohub, San Francisco, CA, USA

[7]Department of Applied Physics, Stanford University, Stanford, CA, USA

Corresponding authors: Fabio Zanini (fabio.zanini@unsw.edu.au) and Shirit Einav (email: seinav@stanford.edu).

Abstract

No vaccines or antivirals are approved against Venezuelan equine encephalitis virus (VEEV) infection in humans. To improve our understanding of VEEV-host interactions, we simultaneously profiled host transcriptome and viral RNA (vRNA) in thousands of single cells during infection of human astrocytes. Host transcription was suppressed, and “superproducer cells” with extreme vRNA abundance and altered transcriptome emerged during the first viral life cycle. Cells with increased structural-to-nonstructural transcript ratio demonstrated upregulation of trafficking genes at later time points. Loss- and gain-of-function experiments confirmed pro- and antiviral host factors. Single-cell deep sequencing analysis identified a viral E3 protein mutation altering host gene expression. Lastly, comparison with data from other viruses highlighted common and unique pathways perturbed by infection across evolutionary scales. This study provides a high-resolution characterization of the cellular response to VEEV infection, identifies candidate targets for antivirals, and establishes a comparative single-cell approach to study the evolution of virus-host interactions.

Introduction

For more than a century, Venezuelan Equine Encephalitis Virus (VEEV), a member of the *Alphavirus* genus, has been the causative agent of outbreaks of febrile neurological disease in both animals and humans in Central and South America (Sharma and Knollmann-Ritschel 2019; Aguilar et al. 2011). The incidence of VEEV infection is underestimated since early symptoms are non-specific (Aguilar et al. 2011). While typically transmitted via a mosquito bite, VEEV is also infectious as an aerosol, hence it is considered a major bioterrorism threat (Hawley and Eitzen 2001). To date, no US FDA approved drugs or vaccines against VEEV are available. A deeper understanding of VEEV biology in human cells is required to advance the development of effective countermeasures against VEEV.

Because VEEV is a biosafety level 3 pathogen, TC-83, a live-attenuated vaccine strain, is commonly used for research purposes (Berge et al. 1961). Although attenuated, VEEV TC-83 replicates rapidly: viral protein production is observed as early as 6 hours postinfection (hpi) of human astrocytoma cells (U-87 MG) at multiplicity of infection (MOI) of 2, and over 10^{10} copies of intracellular viral RNA (vRNA) can be detected by 24 hpi (Keck et al. 2018). It remains unknown, however, whether a large number of cells, each producing a small number of virions, or a few “superproducer” cells drive this effective virus production. Productive replication is associated with profound shutdown of host gene transcription (Garmashova, Atasheva, et al. 2007), however, since the virus relies on cellular machineries, it is important to understand what host factors are “spared” from complete shutdown and needed for virus production.

The genome of VEEV is an ~11.5 kb single-stranded positive-sense RNA. The genomic RNA contains two domains. The 5’ two-thirds of the genome constitutes the first open reading frame (ORF), which encodes the nonstructural (ns) proteins required for viral RNA synthesis (nsP1-4). The 3’ one-third of the genome is the structural protein domain. The structural proteins (capsid, envelope glycoproteins E1-3, 6k, and transframe (TF) protein) are translated from a second

ORF that is expressed through the production of a subgenomic mRNA from an internal promoter in the negative-strand RNA replication intermediate and function in the assembly of new virions and their attachment and entry into cells (Strauss and Strauss 1994). While the stoichiometry of the genomic and subgenomic transcripts in the setting of VEEV infection has not been characterized, the transcription of the subgenomic RNA of a related alphavirus, Sindbis virus (SINV), was shown to be ~3-fold higher than the genomic RNA during late stages of the viral life cycle (Shirako and Strauss 1990; Lemm et al. 1994), supporting a switch towards increased synthesis of structural proteins required for virion formation over nonstructural proteins required primarily for viral RNA replication (Raju and Huang 1991; Levis, Schlesinger, and Huang 1990).

The understanding of the alphavirus life cycle is largely based on studies conducted with the non-pathogenic SINV and Semliki forest virus (SFV). Alphaviruses enter their target cells via clathrin-mediated endocytosis and release their nucleocapsid into the cytoplasm via fusion with endosomal membranes, followed by translation and processing of the nonstructural polyprotein (Kielian, Chanel-Vos, and Liao 2010). Viral RNA replication occurs within membrane invaginations called spherules that are thought to be derived from the plasma membrane, endoplasmic reticulum and late endosomes and are subsequently incorporated into type 1 cytopathic vacuoles (CPV)-I composed of modified endosomes and lysosomes (Spuul et al. 2010; Grimley, Berezesky, and Friedman 1968; Kujala et al. 2001; Garoff et al. 1994). Production of genomic RNA and subsequently subgenomic RNA are followed by polyprotein translation and processing. The current model of infectious alphavirus production suggests that the genomic RNA is packaged by the capsid in the cytoplasm and that the viral glycoproteins traffic via membrane structures, presumed to be *trans*Golgi-derived (CPV-II), to budding sites on the plasma membrane, followed by membrane curving and scission facilitating envelopment of the nucleocapsid (Garoff et al. 1994; Griffiths, Quinn, and Warren 1983; Soonsawad et al. 2010).

Although VEEV is predicted to extensively interact with cellular factors to effectively replicate and evade cellular immune responses, like other small RNA viruses, little is known about these interactions. A recent small interfering RNA (siRNA) screen revealed a requirement for actin-remodeling pathway proteins including ARF1, RAC1, PIP5K1-α, and ARP3 in VEEV infection and specifically in promoting viral glycoprotein transport to the plasma membrane (Radoshitzky et al. 2016). Various other cellular factors, such as DDX-1 and -3 (Amaya et al. 2016), have been reported to have proviral functions, whereas IFITM3 (Gupta et al. 2017) and members of the PARP protein family (Atasheva et al. 2012), were shown to be antiviral factors. Nevertheless, to the best of our knowledge, the interplay between VEEV and the human host has not been studied to date via an unbiased, genome-wide approach.

Single cell RNA sequencing (scRNA-Seq) has demonstrated utility for understanding the heterogeneity of both viral and cellular transcriptome dynamics at a high resolution. We have recently developed virus-inclusive single-cell RNA-Seq (viscRNA-Seq), an approach to simultaneously profile host and viral gene expression in thousands of single cells (Zanini, Pu, et al. 2018). The studies we and others have conducted in cell lines infected with dengue (DENV), Zika (ZIKV), influenza A (IAV) (Russell, Trapnell, and Bloom 2018; Russell et al. 2019) and West Nile (WNV) viruses (O'Neal et al. 2019) and our results in samples from DENV-infected patients (Zanini, Robinson, et al. 2018) revealed a tremendous cell-to-cell heterogeneity in both vRNA abundance and levels of host factors that support or restrict infection. Moreover, we have demonstrated the utility of this approach in identifying novel cellular factors that support or restrict viral infection (Zanini, Pu, et al. 2018). We have therefore hypothesized that studying VEEV transcriptome dynamics at a single cell resolution may overcome challenges related to the high viral replication rate, thereby highlighting specific transcriptomic signatures above the suppressed transcriptional landscape and identifying novel cellular factors that support or restrict VEEV replication.

We conducted a longitudinal study of virus-host cell interactions across 24 hours of VEEV infection in U-87 MG cells via viscRNA-Seq. We detected extreme heterogeneity in vRNA abundance and host transcriptome across cells from the same culture. To overcome the challenge presented by this uneven and rapid viral replication, we stratified cell populations based on vRNA abundance rather than time postinfection and correlated cellular gene expression with both (i) total vRNA and (ii) the ratio of total (genomic + subgenomic) to genomic vRNA. These approaches enabled identification of genes whose expression is altered during VEEV infection, many of which were then confirmed via loss-of-function and gain-of-function experiments to have pro- and antiviral roles, respectively. Moreover, we revealed a small population of “superproducer cells” that drives the rapid increase in vRNA in the first replication cycle and a cell population that harbors excess of the structural over nonstructural viral ORFs at late stages of viral infection, both associated with distinct host gene expression patterns. We identified single nucleotide polymorphisms (SNPs) that are correlated with a specific host response and structural variants within the VEEV genome. Lastly, comparison of the VEEV dataset with published data on other RNA viruses revealed unique and overlapping host gene responses across viral clades, highlighting the utility of comparative single-cell transcriptomics.

Results

viscRNA-Seq reveals cell-to-cell heterogeneity in VEEV and host gene expression.

To characterize the relation between viral and host cell transcriptional dynamics over the course of VEEV infection, human astrocytoma cells (U-87 MG) (PontÉN and Macintyre 1968) were infected with VEEV (TC-83, attenuated vaccine strain) conjugated to GFP (Sun et al. 2014) at MOIs of 0.1 and 1 or mock infected and harvested at six time points: 0.5, 1.5, 4, 6, 12, and 24 hpi **(Figure 1A).** Single cells were then isolated and processed by viscRNA-Seq, as described previously (Zanini, Pu, et al. 2018). Since the VEEV RNA is polyadenylated, it can be captured by the standard poly-T oligonucleotide that hybridizes with host transcripts. Nevertheless, to

improve vRNA capture and ensure coverage at the 5' end of the viral genome, two specific viral capture oligonucleotides, at positions 352 and 1,742 of the VEEV genome, were added to the reaction (see Methods). In total, 4608 cells were processed, of which 2301 cells were sequenced with approximately 1 million reads/cell **(Figure S1A).** 2004 cells passed quality controls and were analyzed (see Methods).

The fraction of VEEV-infected cells, defined by the presence of 10 or more viral reads, increased with both time and MOI and saturated at 12 and 24 hpi with MOI 1 and 0.1, respectively **(Figure 1B and Figure S1B).** A rapid increase in the ratio of viral/total reads was observed within single cells over time **(Figure 1C)**. Starting at 6 hpi, the duration of a single cycle of VEEV replication (Jose, Taylor, and Kuhn 2017; Strauss and Strauss 1994), a fraction of the cells (5% at 6 hours and 32% at 12 hours in cells infected with MOI 0.1 and 1, respectively) contained a large number of vRNA reads (up to 10% of the total reads), indicating that once initiated, viral replication proceeded extremely fast in these "superproducer" cells **(Figure 1C)**. At later time points, particularly 24 hpi, the increase in vRNA reads was associated with a decline in cellular transcripts. The normalized cellular mRNA reads (calculated by dividing the absolute number of reads by the sum of External RNA Controls Consortium (ERCC) spike-in reads) declined in infected cells (**Figure 1D**), but not in uninfected bystander cells derived from the infected samples (**Figure 1E**) or in mock-infected control cells (**Figure S1C**). To avoid an artificial decline in host gene reads in cells with high vRNA abundance, rather than normalizing host gene reads by the total reads, we normalized by ERCC reads for most downstream analyses: this transformation is akin to an estimate of the actual number of mRNA molecules for each gene (up to a constant factor).

The distributions of viral/total reads at 6 and 12 hpi suggested that infected cells can be roughly divided into two populations based on vRNA abundance with a cutoff of 0.001 virus/total reads (**Figure 1C and Figure S1D**). Independently, dimensionality reduction followed by unsupervised

Louvain clustering (Blondel et al. 2008) performed on the combined host gene expression data in infected cells harvested at all time points (ignoring vRNA abundance) also revealed two cell populations, albeit with a partial overlap (**Figure 1F**). Notably, cluster 1 was enriched for cells with high vRNA abundance, whereas cluster 0 included mostly cells with low vRNA abundance, largely overlapping with the two cell populations defined by the cutoff of 0.001 virus/total reads in the distribution analysis (**Figure 1G**). These results confirm that a cutoff of 0.001 virus/total reads can effectively divide infected cells into two, potentially functionally different, populations and support its use over time postinfection as a grouping strategy for differential gene expression analysis.

Altered expression of host factors and pathways during VEEV infection. To identify host genes whose expression is altered during VEEV infection, we integrated differential gene expression and correlation analyses. First, we combined cells harvested at different time points and divided them into three groups based on the cutoff defined above: infected cells with high vRNA (>0.001 virus/total reads), infected cells with low vRNA (<0.001 virus/total reads), and mock-infected controls. Computing differential expression at the distribution level (Mann-Whitney U test) revealed 4608 and 548 host genes whose expression level significantly differed between the high or low vRNA groups versus the mock-infected controls, respectively. We predicted that differential expression of some genes might be related to batch effect resulting from differences in incubation time rather than from viral infection. To control for such confounders, we calculated Pearson correlation coefficients between gene expression and time postinfection. We then used two-dimensional scatterplots to visualize the distribution of these correlation coefficients for each gene in the infected group (with high or low vRNA) versus the mock-infected group (**Figure S2A**). Genes demonstrating greater than 2-fold difference between infected and mock-infected cells were considered differentially expressed genes (DEGs), whereas genes whose expression was similarly altered over time between infected and mock-infected cells were thought to represent batch effect. 2559 and 144 genes passed this

additional filter as DEGs between the high and low vRNA groups vs. mock-infected group, respectively. The magnitude of the differences in expression levels of most DEGs was smaller (<0.5 log2 fold change) between the low vRNA vs. mock-infected groups than between the high vRNA vs. mock-infected groups. Representative genes that were overexpressed (TNFAIP3), underexpressed (TAF3), or unaltered (RND3) in the high vRNA vs. mock-infected groups are shown in **Figure S2B**. The expression level of these genes did not change over time in mock-infected cells, supporting that the altered levels represent actual differences between the groups rather than batch effects (**Figure S2C**).

To better understand the host gene dynamics in VEEV infection, we also computed Spearman's rank correlation coefficients between gene expression and vRNA abundance across all cells, as done previously for flaviviruses (Zanini, Pu, et al. 2018). Examples of positively correlated (TNFAIP3), negatively correlated (TAF3), and uncorrelated (RND3) genes are shown in **Figure 1H**. Our data indicate that the majority of host genes are negatively correlated with vRNA abundance (**Figure S2D)**. Stratifying host genes by expression level in mock-infected cells indicated a stronger negative correlation for highly expressed genes (**Figure S2E**), suggesting that cellular functions relying on highly expressed genes are more vulnerable to VEEV infection.

To identify genes that are both differentially expressed between infected and uninfected cells and are correlated with vRNA (beyond TNFAIP3 and TAF3), we computed the intersection between the 2559 DEGs with the top 600 genes that positively (n=300) or negatively (n=300) correlated with vRNA. 206 overlapping genes emerged from this analysis, of which 5 were positively correlated and 201 were negatively correlated with vRNA. Gene Ontology (GO) enrichment analysis of these 206 genes via metascape (Zhou et al. 2019) highlighted negative regulation of intracellular signal transduction as the most enriched molecular function term (**Figure 1I**).

Early infected "superproducer'' cells show distinct patterns of host gene expression. To elucidate whether the "superproducer" cells (harboring > 0.001 vRNA/total reads at 6 hpi, **Figure 1C**) exhibit a distinct gene expression pattern, we conducted differential gene expression analysis between these 13 cells (5% of total cell population at 6 hpi) and an equal number of bystander cells harboring no vRNA reads derived from the same culture at the same time point. A total of 32 DEGs were identified, with representative overexpressed and underexpressed genes shown in **Figure 1J**. Among the overexpressed genes are MEIS2, a master transcriptional regulator, WWP1, an E3 Ubiquitin ligase involved in membrane trafficking, and ADAMTS6, a disintegrin and metalloproteinase involved in epithelial cell-cell junctions and O-linked glycosylation of proteins. Among the underexpressed genes are ZMAT5, an RNA-binding protein belonging to the CCCH zinc finger family of proteins implicated in antiviral immune regulation (Fu and Blackshear 2017), and PARP1, which is a poly(ADP-ribosyl) transferase related to PARP7, 10 and 12, previously shown to have antiviral functions in VEEV infection (Atasheva, Frolova, and Frolov 2014; Atasheva et al. 2012). These findings provide evidence that a small subset of "superproducer'' cells largely drives VEEV replication during the first viral life cycle and demonstrates a distinct gene expression pattern. These results also point to MEIS2, WWP1, and ADAMTS6 as candidate proviral factors, and to ZMAT5 and PARP1 as potential antiviral factors.

The expression of genes involved in intracellular membrane trafficking correlates with the ratio of 3' to 5' vRNA reads. By including both a poly-T and a 5'-end specific capture oligonucleotides in the viscRNA-Seq, good read coverage at both ends of the VEEV genome was obtained (**Figure 2A**). We defined 5' RNA reads as those corresponding to the first 1,700 bases (encoding nonstructural proteins) and thus derived from the genomic vRNA only, and 3' RNA reads as those corresponding to the last third of the genome (encoding structural proteins), derived from both the genomic and subgenomic vRNAs (**Figure 2B**). The stoichiometry of the 3' and 5' RNAs was highly heterogeneous between cells. While at early

263 stages of infection the 3’ to 5’ (structural to nonstructural) vRNA read ratio (3’/5’ read ratio), as
264 defined by these genomic regions, was below or around 1, at late stages, it reached up to 4 and
265 was correlated with total vRNA abundance (**Figure 2C**). In contrast, the read ratio between two
266 segments we selected as internal controls at the 5’ end of the vRNA (5’a/5’b read ratio) and
267 between two segments at the 3’ end (3’a/3’b read ratio) did not correlate with the cellular vRNA
268 abundance (**Figures 2D-E**). To test the hypothesis that differences in vRNA stoichiometry are
269 associated with distinct host responses, we measured the Spearman correlation coefficients of
270 all host genes with the 3’/5’ read ratio in the same cell. The resulting histogram distribution
271 curve revealed a long tail of host genes whose expression increased with the 3’/5’ read ratio
272 (**Figure 2F**), in contrast to the distribution of host genes in correlation with the total vRNA reads
273 (**Figure S2D**). Positively correlated genes were mostly involved in various aspects of
274 intracellular trafficking and included factors previously reported to be required for VEEV infection
275 via an siRNA screen including ARP3 (Radoshitzky et al. 2016), RAC2, a paralog of RAC1
276 (Radoshitzky et al. 2016), and DDX5, a member of the DEAD box family of RNA
277 helicases (Amaya et al. 2016). Novel factors among the positively correlated genes included
278 factors involved in late endosomal trafficking (RAB7A (Verhoeven et al. 2003), the accessory
279 ESCRT factor (BROX) (Mu et al. 2012), and the SNARE protein VAMP7 (Pryor et al. 2008)), ER
280 to Golgi trafficking (SEC22B) (Zhang et al. 1999), regulation of secretion (PIP4K2A) (Rozenvayn
281 and Flaumenhaft 2003), lysosome function and autophagy (LAMP2) (Hubert et al. 2016), actin
282 polymerization (PFN2) (Honoré et al. 1993), and acidification of intracellular organelles for
283 protein sorting (ATP6V1B2) (Bernasconi et al. 1990) (**Figure 2G**). Accordingly, pathway
284 analysis on the top 300 correlated genes identified macroautophagy, membrane trafficking,
285 vesicle organization, exosomal secretion, and regulated exocytosis as the highly enriched
286 functions (**Figure 2H**). Notably, these genes were only positively correlated with the 3’/5’ read
287 vRNA ratio and not with the total vRNA reads. These findings indicate that the late stages of
288 VEEV infection are characterized by heterogeneous stoichiometry of structural (3’) and
289 nonstructural (5’) vRNAs and upregulation of intracellular trafficking pathways previously

implicated in assembly and egress of various RNA viruses in cells with an excess of structural vRNA. Moreover, these results highlight the unique opportunity to discover candidate proviral factors for VEEV infection by correlating gene expression with specific viral genome stoichiometry via viscRNA-Seq.

Validation of candidate proviral and antiviral factors. Next, we probed the functional relevance of 24 genes that either strongly or moderately correlated with vRNA abundance to viral infection. We first conducted loss-of-function screens by measuring the effect of siRNA-mediated depletion of the 24 individual genes on VEEV infection and cellular viability in U-87 MG cells (**Figures 3A and S3**). Depletion of 4 and 10 genes suppressed or increased VEEV infection by more than 40%, respectively, as measured by luciferase assays 18 hpi with a nano-luciferase reporter TC-83 virus and normalized to cellular viability in two independent screens. Depletion of CXCL3, ATF3, TNFAIP3, and CXCL2, four out of five genes tested that positively correlated with vRNA abundance via viscRNA-Seq (orange bars), reduced VEEV infection, suggesting that they are proviral factors. In contrast, depletion of 10 of 19 genes tested that negatively correlated with vRNA (grey bars) enhanced infection, suggesting that these proteins may function as antiviral factors. Suppression of four underexpressed genes demonstrated no effect on VEEV infection, suggesting that they were either non-essential or not restricting (possibly due to redundancy in host factors requirement) or that the level of knockdown was insufficient to trigger a phenotype.

Next, we conducted gain-of-function screens by ectopically expressing the 24 individual gene products in U-87 MG cells followed by VEEV infection (**Figure 3B**). Using a cutoff of greater than 40% change in viral infection normalized to cell viability in two independent screens, overexpression of most genes resulted in an inverse effect to that observed with the siRNA, i.e. if knockdown inhibited viral infection, overexpression enhanced it and vice versa. Overexpression of CXCL3, ATF3, TNFAIP3 and CXCL2 increased VEEV infection, indicating

rate limitation associated with these candidate proviral factors. In contrast, overexpression of the majority of the anticorrelated gene products reduced VEEV infection, consistent with an antiviral phenotype.

While the transcriptional level of EIF4A3, SRSF1, TRMT10C, PSMD14 and PSMC5 anticorrelated with vRNA abundance, their gene products demonstrated a proviral phenotype. This may either result from regulation of these genes at the translational level or from downstream effects of these multifunctional genes. ARRDC3, a member of the arrestin family (Qi et al. 2014), was positively correlated with vRNA abundance, yet its depletion increased infection and its overexpression decreased infection, in contrast with the other four positively correlated genes tested. To probe this discrepancy, we measured the correlation of ARRDC3 expression with the 5' and 3' vRNA reads separately. Notably, ARRDC3 reads positively correlated with the 3' vRNA reads but negatively correlated with the 5' vRNA reads. In contrast, the other four proviral candidates positively correlated with both the 5' and 3' vRNA reads (**Figure S3C**). This finding suggests that ARRDC3 might have a dual function during VEEV infection. These findings highlight the utility of viscRNA-Seq in identifying candidate proviral and antiviral factors.

Viral polymorphisms and rare structural viral read variants correlate with expression of specific host genes. The high coverage corresponding to almost the entire viral genome obtained via viscRNA-Seq enabled computation of the frequencies of SNPs at all sites within single cells. We identified one position with high variance in SNP frequency across cells, site 9227 (**Figure 4A**), where a G > A nonsynonymous change caused a substitution from Alanine to Threonine in the E3 protein. To determine whether the presence of this mutation in the viral genome is associated with an altered expression of host genes, we correlated the frequency of the 9227A allele with host gene expression. We identified several genes including BROX, TMED and SRSF6 that negatively correlated with the frequency of this mutation (**Figure 4B**).

Only weak positive correlation was calculated between the frequency of the 9227A allele and vRNA abundance within the same cell, suggesting that this mutation was not under positive selection within the viral pool infecting the cell culture.

Additionally, among millions of viral reads detected, we observed 14,956 gap reads (~0.1% of total viral reads), defined by having a deletion within read 1 or read 2 (we used Illumina paired-end sequencing, see Methods), not including reads with gaps between the two reads (**Figure S4A**). These gap reads were present in 271 cells and their abundance strongly correlated with vRNA abundance in the same cells, indicating that deep viral coverage was required for detection (**Figure S4B**). The length of these gaps ranged from 20 to over 10,000 nucleotides, with the majority being shorter than 1000 nucleotides (**Figure S4C**). The most common was a 36-base gap located within the coding region of the 6K protein (black arrow in Figure S4A). This gap was found in a total of 1,226 reads derived from 55 different cells. Prediction of the RNA structure via RNAfold web server(Gruber et al. 2008) revealed that in the presence of the 36-base gap, there is formation of a hairpin with a free energy of -21.23 kcal/mol, indicating a very stable structure (**Figure S4D**). Although the biological function of this hairpin is unknown, stable RNA structures play essential roles in viral replication and tropism across multiple viruses. Alternatively, we cannot currently exclude that this gap could be a result of polymerase errors during the library preparation.

While further studies are needed to clarify the function of the observed SNPs and gaps, these findings highlight the utility of viscRNA-Seq in studying virus evolution and structural genome variants.

Comparative viscRNA-Seq analysis across five RNA viruses reveals distinct and common cellular pathways affected by viral infection.

To define which elements of the host response are unique to VEEV or common across multiple viruses, we first compared the VEEV dataset with our previously published viscRNA-Seq data on human hepatoma (Huh7) cells infected with DENV and ZIKV (Zanini, Pu, et al. 2018). Since the baseline gene expression levels in astrocytes (VEEV) are different from those in hepatocytes (DENV, ZIKV), we limited the analysis to genes that were similarly expressed (within a 10-fold change) in uninfected Huh7 and U-87 MG cells. We selected cells with greater than 2 vRNA reads per million joint (viral + host) reads and monitored how the expression of host genes changes with increasing vRNA abundance across the three infections. In all three viral infections, the majority of host genes were not correlated with vRNA abundance. Nevertheless, a number of host genes exhibited correlations with one or more viruses. Three robust patterns were identified (**Figures 5A-C**): genes, such as HSPA5, that were upregulated in DENV infection and downregulated in ZIKV and VEEV infections (**Figure 5A**); genes like NRBF2 that were upregulated only during ZIKV infection (**Figure 5B**); and genes, such as SERP1, that were downregulated only in VEEV infection (**Figure 5C**). No genes that are upregulated only in VEEV infection could be identified. Beyond these general categories, the resulting patterns of viral and host expression were, however, quite complex. To better elucidate the large differences in correlation with vRNA abundance between each virus and the other two demonstrated by these representative genes, we plotted the data using a ternary plot (**Figure 5D**). In this plot, individual lines indicate the proportion of reads of a specific host gene at increasing vRNA abundance normalized by the sum of reads of that gene in all three experiments (each with different virus), and stars indicate cells with the highest vRNA abundance. Proximity to either the lower right (HSPA5) or top corner (NRBF2) of the triangle indicate high expression of the gene during DENV or ZIKV infection, respectively, whereas proximity to the midpoint of the ZIKV axis (SERP1) indicates reduced expression specifically during VEEV infection (**Figure 5D**).

To circumvent the masking effect of VEEV transcriptional shutdown, we then compared the genes that positively correlated with the 3'/5' VEEV RNA ratio with those positively or negatively correlating with DENV or ZIKV vRNA (**Figure 5E**). This analysis revealed genes, such as BROX, GEM, and RNF114 that are positively correlated with the respective vRNA in all three viral infections, genes, such as CTSB and SPTLC1 that are positively correlated with 3'/5' VEEV RNA and ZIKV but not DENV vRNA, and genes that are positively correlated with 3'/5' VEEV RNA but negatively correlated with DENV and ZIKV vRNA, such as PFN2 and DPYSL2. In contrast, no large correlations were observed when a comparable number of random genes were similarly analyzed (**Figure 5F**). Pathway analysis on genes that are positively correlated with both the 3'/5' VEEV RNA ratio and the two flaviviral RNAs identified ER processing, glycosylation, SELK (part of Endoplasmic-Reticulum-Associated Degradation), tRNA synthesis, protein folding, virion assembly, and intracellular transport as the highly enriched functions (**Figure S5A**). In contrast, cell cycle and apoptosis control were the most highly enriched functions in pathway analysis on genes that were positively correlated with 3'/5' VEEV RNA ratio but negatively correlated with the two flaviviral RNA (**Figure S5B**). These results provide evidence that complex temporal dynamics exist across different RNA viral infections, and highlight both common and unique cellular pathways that are altered by VEEV and flaviviruses.

Next, we expanded our comparative analysis by including published datasets derived from single-cell transcriptomic studies on different cell lines infected with IAV (Russell, Trapnell, and Bloom 2018) and WNV (O'Neal et al. 2019) generated via 10x Genomics and Smart-seq2, respectively. Because different cell lines were used for different viruses, we calculated the ranks of the correlation coefficients between the expression of each host gene and vRNA for each virus, restricted the selection to the top and bottom 200 genes, and normalized the results between -1 and 1 for each virus. We then calculated the network of similarities between genes. t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten and Hinton 2008) and Leiden clustering (Traag, Waltman, and van Eck 2019) of the genes highlighted nine gene clusters with

different expression patterns during various viral infections (**Figure 5G**). To understand the meaning of these clusters, we performed double hierarchical clustering and observed that clusters 0, 4, 7, and 2 are upregulated, while clusters 1, 3, 8, 6, and 5 are mostly downregulated during viral infection (**Figure 5H**). The dendrogram of the five viruses was qualitatively consistent with the known phylogeny as derived from viral genomic sequences, which could indicate ancestral phenotypic signatures. While upregulation was largely virus-specific, gene downregulation was sometimes shared across many viruses, particularly in cluster 3. Pathway analysis on individual clusters revealed that cluster 3 is greatly enriched in cell cycle genes (**Figure 5I**), providing evidence that unrelated RNA viruses commonly downregulate at least part of the cell cycle machinery during replicative infection. Notably, although the list of genes that were exclusively correlated with 3'/5' VEEV read ratio was also enriched in cell cycle genes (see Figures S5B), over 80% of those genes did not overlap with cluster 3 genes, indicating that only a minority of host factors within this pathway are associated with VEEV virus production. The pathway analysis of cluster 0 confirmed previously identified pathways of ER stress and vesicle transport in flaviviral infection (**Figure S6**). The signatures obtained in other clusters were less clear and would require more sophisticated analyses to be deciphered.

Overall, our analysis indicates that although comparing single cell viral infection data across species, cell lines, and technologies still presents challenges, this approach is informative in highlighting host genes and pathways that are commonly affected across very different viral families.

Discussion

We and others have recently characterized the cellular response in virally infected cell lines (Zanini, Pu, et al. 2018; Russell, Trapnell, and Bloom 2018), primary cells (Gorman et al. 2018; O'Neal et al. 2019) and patient samples (Zanini, Robinson, et al. 2018) via single-cell RNA-seq

approaches. Moreover, we reported unique and overlapping determinants in the host response to two related flaviviruses at a single cell resolution (Zanini, Pu, et al. 2018). Nevertheless, the host transcriptomic response to infection by alphaviruses, which induces a profound transcriptional shutdown of host genes, has not been previously characterized at a single cell level, and the single-cell transcriptomic responses of unrelated viruses have not been compared. By applying viscRNA-Seq to study the temporal infection dynamics of VEEV (TC-83) in human astrocytes, we revealed large cell-to-cell heterogeneity in VEEV and host gene expression, transcriptomic signatures in distinct cell subpopulations, and candidate proviral and antiviral factors, some of which we then validated. Additionally, we established a role for viscRNA-Seq in comparative evolutionary virology by demonstrating SNPs that are correlated with a specific host response, structural variants within the VEEV genome, as well as unique and overlapping host gene responses across multiple RNA viral clades. These findings provide insights into the virus-host determinants that regulate VEEV infection and highlight the utility of virus-inclusive RNA-seq approaches and comparative single-cell transcriptomics.

A prominent feature of VEEV infection is a profound suppression of cellular transcription (Garmashova, Atasheva, et al. 2007). Nevertheless, it remained unknown whether this transcriptional shutdown globally affects all cells in virus-infected cell culture and all host mRNAs. We provide evidence that the suppression of host mRNAs occurs only in infected but not bystander cells derived from the same infected samples. This finding indicates that previously reported direct effects of VEEV (Yin et al. 2009; Garmashova, Gorchakov, et al. 2007), but not indirect mechanisms, mediate this suppression. Additionally, computing the distributions of vRNA expression in correlation with 5 groups of genes, distinguished by the level of gene expression in uninfected cells, demonstrated that highly expressed genes are more likely to be negatively correlated with vRNA abundance than genes that are expressed at a lower level. The cellular energy and machinery required to maintain a high level of gene

expression likely play a role in increasing the vulnerability of highly expressed cellular genes to VEEV-induced transcriptional shutdown.

We have previously reported the utility of viscRNA-Seq in discovering functional transcriptomic signatures and candidate pro- and antiviral factors of DENV and ZIKV infections (Zanini, Pu, et al. 2018; Zanini, Robinson, et al. 2018). Nevertheless, the high replication rate of VEEV and the transcriptional shutdown it induces challenged our ability to detect alterations in gene expression and identify pro- and antiviral factors. To overcome these challenges, we used several strategies. First, since the viscRNA-Seq analysis revealed large differences in vRNA abundance between cells infected with the same MOI and harvested at the same time point, we stratified cell populations based on vRNA abundance rather than time postinfection. Integrating differential gene expression and correlation analyses of vRNA abundance with gene expression across the entire human transcriptome facilitated the discovery of 206 genes that were both differentially expressed between the high and low vRNA groups and correlated with total vRNA. siRNA-mediated depletion and overexpression of a subset of these genes revealed that overall, genes involved in cytokine production, plus ATF3, a transcription factor commonly expressed in response to cellular stress, and TNFAIP3, an inhibitor of NFκB signaling, demonstrated a phenotype consistent with a rate-limiting proviral function, whereas a variety of regulatory genes, such as XBP1, TAF7 and DUSP14, were rate-limiting antiviral factors. ARRDC3, one of 5 genes that were both differentially expressed and positively correlated with total vRNA, demonstrated a phenotype consistent with antiviral rather than a proviral effect. Interestingly, when studied in correlation with the individual vRNA transcripts, ARRDC3, a signaling arrestin family protein and a cargo-specific endosomal adaptor, was positively correlated with the 3’ vRNA but negatively correlated with the 5’ vRNA, suggesting that it may have a proviral effect during later stages and an antiviral effect in earlier stages of replication. By capturing such complex dynamics and not relying on averaging signals at distinct time points postinfection for

stratification, the viscRNA-Seq approach has an advantage over bulk sample knockdown or knockout approaches in identifying factors required for or restrictive of VEEV infection.

The high resolution provided by viscRNA-Seq enabled us to further focus on distinct cell populations, which facilitated identification of additional transcriptomic signatures. We discovered a subpopulation of cells consisting of 5% of total cells at 6 hpi that demonstrated unusually high viral replication upon completion of a single cycle of viral replication. Importantly, this cell subpopulation is associated with host cell gene expression that is distinct from cells harboring lower vRNA at the same time. It is intriguing to speculate that overexpression of the identified hits (e.g. the NEDD4-like E3 ubiquitin ligase that mediates trafficking, WWP1; the disintegrin, ADAMTS6; and the transcription regulator, MEIS2P) concurrently with underexpression of factors implicated in antiviral immune responses (such as ZMAT5 and PARP1) in this cell population drive the rapid increase in viral replication during the first viral life cycle.

To further increase the resolution of our analysis, we took advantage of the ability of viscRNA-Seq to detect the two VEEV transcripts. A prior study on IAV has detected different levels of various segments of the viral genome across cells and investigated how this finding relates to successful virion production (Russell, Trapnell, and Bloom 2018). Similarly, analysis of the stoichiometry of the 5’ and 3’ RNA reads of VEEV, a non-segmented virus, revealed a large cell-to-cell heterogeneity. Moreover, the 3’/5’ vRNA ratio substantially increased at late stages of infection, consistent with a previous report in another alphavirus, SINV (Raju and Huang 1991). Remarkably, the histogram distribution curve of the Spearman correlation coefficients of all host genes with the 3’/5’ read ratio in the same cell revealed a long tail of host genes whose expression increased with the 3’/5’ read ratio. Our findings indicate that these changes in stoichiometry of the vRNA transcripts during late stages of VEEV infection are associated with upregulation of distinct genes, particularly those involved in intracellular trafficking pathways.

Notably, detection of these factors was only possible by correlating their expression specifically with the 3'/5' vRNA ratio and not the total vRNA reads. The involvement of these factors specifically in cells harboring high 3'/5' vRNA read ratio thus makes it experimentally challenging to further study them via bulk sample approaches. Nevertheless, it is tempting to speculate that some of the discovered late endosomal trafficking and lysosomal proteins (RAB7A (Verhoeven et al. 2003), BROX (Mu et al. 2012), VAMP7 (Pryor et al. 2008) and LAMP2 (Hubert et al. 2016)) may be involved in forming the CPV-I composed of modified endosomes and lysosomes in which VEEV RNA replication occurs (Spuul et al. 2010; Grimley, Berezesky, and Friedman 1968; Kujala et al. 2001; Salonen et al. 2003; Friedman et al. 1972; Pietilä, van Hemert, and Ahola 2018; Pietilä, Hellström, and Ahola 2017), and that ATP6V1B2 (Bernasconi et al. 1990) may mediate the acidification of this acidic intracellular compartment (Jose, Taylor, and Kuhn 2017). Moreover, the positive correlation of proteins involved in ER to Golgi trafficking (SEC22B) (Zhang et al. 1999), regulation of secretion (PIP4K2A) (Rozenvayn and Flaumenhaft 2003), autophagy (LAMP2) (Hubert et al. 2016), actin polymerization (PFN2) (Honoré et al. 1993), and ESCRT machinery (BROX, a Bro1 domain-containing protein like ALIX) (Mu et al. 2012; Zhai et al. 2011), TSG101 and STAM2) with the 3'/5' vRNA read ratio proposes roles for these factors in late stages of the VEEV lifecycle, such as trafficking of the CPV-IIs to the plasma membrane, virion assembly, and/or budding (Garoff et al. 1994; Griffiths, Quinn, and Warren 1983; Soonsawad et al. 2010). These results propose a model wherein specific genes are upregulated within the profound transcriptional downregulation in a stoichiometry-dependent manner, and further illuminate the utility of viscRNA-Seq in identifying candidate proviral and antiviral factors, including druggable candidates for host-targeted antiviral approaches.

Comparative evolutionary virology is an ideal application for single cell technologies because of the degree of genomic and functional diversity of infections. At the microevolutionary scale, we observed a negative correlation between the minor allele frequency of a nonsynonymous SNP

in the E3 protein and expression of the host genes BROX, TMED2, and SRSF6. Analysis of vRNA reads also revealed structural genome variants, the most common of which harbors a 36-base gap and is predicted to form a stable RNA structure. In the setting of IAV infection, an association between defective viral genomes and host responses was recently reported (Wang et al. 2020). Our findings highlight the unique opportunity provided by virus-inclusive single cell transcriptomic approaches to study genomic interactions between viruses and the host. Similar strategies could be used on engineered mutant viral libraries to study host-pathogen interactions in a massively parallel scale, by combining deep mutational scanning (Sourisseau et al. 2019) with host perturbation libraries (Hein and Weissman 2019).

At the macroevolutionary scale, we compared the effect of unrelated human RNA viruses on the host cell. To address the confounding effect of different host cell lines, we restricted the analyses in **Figures 5A-F** to genes with a similar baseline expression level across cell lines. When comparing the VEEV to the DENV and ZIKV datasets, we identified genes that were downregulated but not upregulated only in VEEV. We then compared genes that positively correlated with the 3'/5' VEEV RNA ratio with those correlating with DENV or ZIKV vRNA and found concordant signal for genes involved in protein processing and transport, whereas some cell cycle and apoptosis genes appeared to be specific to VEEV. When comparing data on five different viruses derived using different cell lines and technologies, we observed that downregulated host factors include a shared subset of cell cycle and nucleoside metabolism genes, whereas upregulated factors are more virus-specific. These findings indicate that while some cell cycle components are generally inhibited by many viruses, others might interact specifically with VEEV. The "correct" phylogeny of the five viruses could be recovered purely from the host transcriptome perturbation, i.e. without using viral genomic information, which is intriguing. More viruses across the viral phylogeny should be assessed to evaluate whether this signal is the result of conserved ancestral function or, alternatively, of convergent functional evolution.

Overall, our study uncovered global and gene-specific host transcriptional dynamics during VEEV infection at single cell resolution and presented a novel approach to elucidate the evolution of virus-host interactions.

Acknowledgements

This work was supported by HDTRA11810039 from the Defense Threat Reduction Agency (DTRA)/Fundamental Research to Counter Weapons of Mass Destruction, by the Chan Zuckerberg Biohub, and by a Stanford Bio-X Interdisciplinary Initiative Program Award. ZY was supported by the Maternal and Child Health Research Institute, Lucile Packard Foundation for Children's Health). NP was supported by the University of Washington School of Medicine Guy Tribble and Susan Barnes Graduate Discovery Fellowship. We thank investigators who have provided plasmids (see Methods).The opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the U.S. Army or the other funders.

Competing interests: The authors declare that no competing interests exist.

References:

Aguilar, Patricia V., Jose G. Estrada-Franco, Roberto Navarro-Lopez, Cristina Ferro, Andrew D. Haddow, and Scott C. Weaver. 2011. “Endemic Venezuelan Equine Encephalitis in the Americas: Hidden under the Dengue Umbrella.” *Future Virology* 6 (6): 721–40.

Amaya, Moushimi, Taryn Brooks-Faulconer, Tyler Lark, Forrest Keck, Charles Bailey, Venu Raman, and Aarthi Narayanan. 2016. “Venezuelan Equine Encephalitis Virus Non-Structural Protein 3 (nsP3) Interacts with RNA Helicases DDX1 and DDX3 in Infected Cells.” *Antiviral Research* 131 (July): 49–60.

Anders, S., P. T. Pyl, and W. Huber. 2015. “HTSeq—a Python Framework to Work with High-Throughput Sequencing Data.” *Bioinformatics* . https://academic.oup.com/bioinformatics/article-abstract/31/2/166/2366196.

Atasheva, Svetlana, Maryna Akhrymuk, Elena I. Frolova, and Ilya Frolov. 2012. “New PARP Gene with an Anti-Alphavirus Function.” *Journal of Virology* 86 (15): 8147–60.

Atasheva, Svetlana, Elena I. Frolova, and Ilya Frolov. 2014. “Interferon-Stimulated poly(ADP-Ribose) Polymerases Are Potent Inhibitors of Cellular Translation and Virus Replication.” *Journal of Virology* 88 (4): 2116–30.

Berge, Trygve O., Isaac S. Banks, W. D. Tigertt, and Others. 1961. “Attenua-Lion of Venezuelan Equine Encephalomyelitis Virus by Ire Vitro Cultivation in Guinea-Pig Heart Cells.” *American Journal of Hygiene* 73 (2): 209–18.

Bernasconi, Paul, T. Rausch, I. Struve, Louis Morgan, and L. Taiz. 1990. “An mRNA from Human Brain Encodes an Isoform of the B Subunit of the Vacuolar H (+)-ATPase.” *The Journal of Biological Chemistry* 265 (29): 17428–31.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. “Fast Unfolding of Communities in Large Networks.” *Journal of Statistical Mechanics: Theory and Experiment*. https://doi.org/10.1088/1742-5468/2008/10/p10008.

Butler, Andrew, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. 2018. “Integrating Single-Cell Transcriptomic Data across Different Conditions, Technologies, and Species.” *Nature Biotechnology* 36 (5): 411–20.

Coutant, Eloi P., Glwadys Gagnot, Vincent Hervin, Racha Baatallah, Sophie Goyard, Yves Jacob, Thierry Rose, and Yves Janin. 2019. “Bioluminescence Profiling of NanoKAZ/NanoLuc Luciferase Using a Chemical Library of Coelenterazine Analogues.” *Chemistry – A European Journal*. https://doi.org/10.1002/chem.201904844.

Coutant, Eloi P., Sophie Goyard, Vincent Hervin, Glwadys Gagnot, Racha Baatallah, Yves Jacob, Thierry Rose, and Yves L. Janin. 2019. “Gram-Scale Synthesis of Luciferins Derived from Coelenterazine and Original Insights into Their Bioluminescence Properties.” *Organic*

& Biomolecular Chemistry 17 (15): 3709–13.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21.

Friedman, R. M., J. G. Levin, P. M. Grimley, and I. K. Berezesky. 1972. “Membrane-Associated Replication Complex in Arbovirus Infection.” *Journal of Virology* 10 (3): 504–15.

Fu, Mingui, and Perry J. Blackshear. 2017. “RNA-Binding Proteins in Immune Regulation: A Focus on CCCH Zinc Finger Proteins.” *Nature Reviews. Immunology* 17 (2): 130–43.

Garmashova, Natalia, Svetlana Atasheva, Wenli Kang, Scott C. Weaver, Elena Frolova, and Ilya Frolov. 2007. “Analysis of Venezuelan Equine Encephalitis Virus Capsid Protein Function in the Inhibition of Cellular Transcription.” *Journal of Virology* 81 (24): 13552–65.

Garmashova, Natalia, Rodion Gorchakov, Eugenia Volkova, Slobodan Paessler, Elena Frolova, and Ilya Frolov. 2007. “The Old World and New World Alphaviruses Use Different Virus-Specific Proteins for Induction of Transcriptional Shutoff.” *Journal of Virology* 81 (5): 2472–84.

Garoff, H., J. Wilschut, P. Liljeström, J. M. Wahlberg, R. Bron, M. Suomalainen, J. Smyth, A. Salminen, B. U. Barth, and H. Zhao. 1994. “Assembly and Entry Mechanisms of Semliki Forest Virus.” *Archives of Virology. Supplementum* 9: 329–38.

Gorman, Matthew J., Elizabeth A. Caine, Konstantin Zaitsev, Matthew C. Begley, James Weger-Lucarelli, Melissa B. Uccellini, Shashank Tripathi, et al. 2018. “An Immunocompetent Mouse Model of Zika Virus Infection.” *Cell Host & Microbe* 23 (5): 672–85.e6.

Griffiths, G., P. Quinn, and G. Warren. 1983. “Dissection of the Golgi Complex. I. Monensin Inhibits the Transport of Viral Membrane Proteins from Medial to Trans Golgi Cisternae in Baby Hamster Kidney Cells Infected with Semliki Forest Virus.” *The Journal of Cell Biology* 96 (3): 835–50.

Grimley, P. M., I. K. Berezesky, and R. M. Friedman. 1968. “Cytoplasmic Structures Associated with an Arbovirus Infection: Loci of Viral Ribonucleic Acid Synthesis.” *Journal of Virology* 2 (11): 1326–38.

Gruber, Andreas R., Ronny Lorenz, Stephan H. Bernhart, Richard Neuböck, and Ivo L. Hofacker. 2008. “The Vienna RNA Websuite.” *Nucleic Acids Research* 36 (Web Server issue): W70–74.

Gupta, Paridhi, Anuj Sharma, Jing Han, Amy Yang, Manish Bhomia, Barbara Knollmann-Ritschel, Raj K. Puri, and Radha K. Maheshwari. 2017. “Differential Host Gene Responses from Infection with Neurovirulent and Partially-Neurovirulent Strains of Venezuelan Equine Encephalitis Virus.” *BMC Infectious Diseases* 17 (1): 309.

Hawley, Robert J., and Edward M. Eitzen Jr. 2001. “Biological Weapons—a Primer for Microbiologists.” *Annual Review of Microbiology*. https://doi.org/10.1146/annurev.micro.55.1.235.

Hein, Marco Y., and Jonathan S. Weissman. 2019. “Functional Single-Cell Genomics of Human Cytomegalovirus Infection.” *bioRxiv*. https://doi.org/10.1101/775080.

Honoré, B., P. Madsen, A. H. Andersen, and H. Leffers. 1993. “Cloning and Expression of a Novel Human Profilin Variant, Profilin II.” *FEBS Letters* 330 (2): 151–55.

Hoyer, Stephan, and Joe Hamman. 2017. “Xarray: ND Labeled Arrays and Datasets in Python.” *Journal of Open Research Software* 5 (1). https://openresearchsoftware.metajnl.com/articles/148/.

Hubert, Virginie, Andrea Peschel, Brigitte Langer, Marion Gröger, Andrew Rees, and Renate Kain. 2016. “LAMP-2 Is Required for Incorporating Syntaxin-17 into Autophagosomes and for Their Fusion with Lysosomes.” *Biology Open* 5 (10): 1516–29.

Hunter, John D. 2007. “Matplotlib: A 2D Graphics Environment.” *Computing in Science & Engineering* 9 (3): 90–95.

Jose, Joyce, Aaron B. Taylor, and Richard J. Kuhn. 2017. “Spatial and Temporal Analysis of Alphavirus Replication and Assembly in Mammalian and Mosquito Cells.” *mBio* 8 (1). https://doi.org/10.1128/mBio.02294-16.

Keck, Forrest, Stephanie Kortchak, Allison Bakovic, Brian Roberts, Nitin Agrawal, and Aarthi Narayanan. 2018. “Direct and Indirect pro-Inflammatory Cytokine Response Resulting from TC-83 Infection of Glial Cells.” *Virulence*. https://doi.org/10.1080/21505594.2018.1509668.

Kielian, Margaret, Chantal Chanel-Vos, and Maofu Liao. 2010. “Alphavirus Entry and Membrane Fusion.” *Viruses* 2 (4): 796–825.

Kujala, P., A. Ikäheimonen, N. Ehsani, H. Vihinen, P. Auvinen, and L. Kääriäinen. 2001. “Biogenesis of the Semliki Forest Virus RNA Replication Complex.” *Journal of Virology* 75 (8): 3873–84.

Lemm, J. A., T. Rümenapf, E. G. Strauss, J. H. Strauss, and C. M. Rice. 1994. “Polypeptide Requirements for Assembly of Functional Sindbis Virus Replication Complexes: A Model for the Temporal Regulation of Minus- and plus-Strand RNA Synthesis.” *The EMBO Journal*. https://doi.org/10.1002/j.1460-2075.1994.tb06587.x.

Levis, R., S. Schlesinger, and H. V. Huang. 1990. “Promoter for Sindbis Virus RNA-Dependent Subgenomic RNA Transcription.” *Journal of Virology* 64 (4): 1726–33.

Maaten, Laurens van der, and Geoffrey Hinton. 2008. “Visualizing Data Using T-SNE.” *Journal of Machine Learning Research: JMLR* 9 (Nov): 2579–2605.

Mu, Ruiling, Vincent Dussupt, Jiansheng Jiang, Paola Sette, Victoria Rudd, Watchalee Chuenchor, Nana F. Bello, Fadila Bouamr, and Tsan Sam Xiao. 2012. “Two Distinct

Binding Modes Define the Interaction of Brox with the C-Terminal Tails of CHMP5 and CHMP4B." *Structure* 20 (5): 887–98.

O'Neal, Justin T., Amit A. Upadhyay, Amber Wolabaugh, Nirav B. Patel, Steven E. Bosinger, and Mehul S. Suthar. 2019. "West Nile Virus-Inclusive Single-Cell RNA Sequencing Reveals Heterogeneity in the Type I Interferon Response within Single Cells." *Journal of Virology* 93 (6). https://doi.org/10.1128/JVI.01778-18.

Pietilä, Maija K., Kirsi Hellström, and Tero Ahola. 2017. "Alphavirus Polymerase and RNA Replication." *Virus Research*. https://doi.org/10.1016/j.virusres.2017.01.007.

Pietilä, Maija K., Martijn J. van Hemert, and Tero Ahola. 2018. "Purification of Highly Active Alphavirus Replication Complexes Demonstrates Altered Fractionation of Multiple Cellular Membranes." *Journal of Virology* 92 (8). https://doi.org/10.1128/JVI.01852-17.

PontÉN, J. A. N., and Elizabeth H. Macintyre. 1968. "Long Term Culture of Normal and Neoplastic Human Glia." *Acta Pathologica et Microbiologica Scandinavica* 74 (4): 465–86.

Pryor, Paul R., Lauren Jackson, Sally R. Gray, Melissa A. Edeling, Amanda Thompson, Christopher M. Sanderson, Philip R. Evans, David J. Owen, and J. Paul Luzio. 2008. "Molecular Basis for the Sorting of the SNARE VAMP7 into Endocytic Clathrin-Coated Vesicles by the ArfGAP Hrb." *Cell* 134 (5): 817–27.

Qi, Shiqian, Morgan O'Hayre, J. Silvio Gutkind, and James H. Hurley. 2014. "Structural and Biochemical Basis for Ubiquitin Ligase Recruitment by Arrestin-Related Domain-Containing Protein-3 (ARRDC3)." *The Journal of Biological Chemistry* 289 (8): 4743–52.

Radoshitzky, Sheli R., Gianluca Pegoraro, Xi Olì Chī, Lián D Ng, Chih-Yuan Chiang, Lucas Jozwick, Jeremiah C. Clester, et al. 2016. "siRNA Screen Identifies Trafficking Host Factors That Modulate Alphavirus Infection." *PLoS Pathogens* 12 (3): e1005466.

Raju, R., and H. V. Huang. 1991. "Analysis of Sindbis Virus Promoter Recognition in Vivo, Using Novel Vectors with Two Subgenomic mRNA Promoters." *Journal of Virology* 65 (5): 2501–10.

Rozenvayn, Nataliya, and Robert Flaumenhaft. 2003. "Protein Kinase C Mediates Translocation of Type II Phosphatidylinositol 5-Phosphate 4-Kinase Required for Platelet α-Granule Secretion." *The Journal of Biological Chemistry* 278 (10): 8126–34.

Rual, Jean-François, Tomoko Hirozane-Kishikawa, Tong Hao, Nicolas Bertin, Siming Li, Amélie Dricot, Ning Li, et al. 2004. "Human ORFeome Version 1.1: A Platform for Reverse Proteomics." *Genome Research* 14 (10B): 2128–35.

Russell, Alistair B., Elizaveta Elshina, Jacob R. Kowalsky, Aartjan J. W. Te Velthuis, and Jesse D. Bloom. 2019. "Single-Cell Virus Sequencing of Influenza Infections That Trigger Innate Immunity." *Journal of Virology* 93 (14). https://doi.org/10.1128/JVI.00500-19.

Russell, Alistair B., Cole Trapnell, and Jesse D. Bloom. 2018. "Extreme Heterogeneity of

Influenza Virus Infection in Single Cells." *eLife* 7 (February). https://doi.org/10.7554/eLife.32303.

Salonen, Anne, Lidia Vasiljeva, Andres Merits, Julia Magden, Eija Jokitalo, and Leevi Kääriäinen. 2003. "Properly Folded Nonstructural Polyprotein Directs the Semliki Forest Virus Replication Complex to the Endosomal Compartment." *Journal of Virology* 77 (3): 1691–1702.

Sharma, Anuj, and Barbara Knollmann-Ritschel. 2019. "Current Understanding of the Molecular Basis of Venezuelan Equine Encephalitis Virus Pathogenesis and Vaccine Development." *Viruses*. https://doi.org/10.3390/v11020164.

Shirako, Y., and J. H. Strauss. 1990. "Cleavage between nsP1 and nsP2 Initiates the Processing Pathway of Sindbis Virus Nonstructural Polyprotein P123." *Virology* 177 (1): 54–64.

Soonsawad, Pan, Li Xing, Emerson Milla, Juan M. Espinoza, Masaaki Kawano, Michael Marko, Chyongere Hsieh, et al. 2010. "Structural Evidence of Glycoprotein Assembly in Cellular Membrane Compartments prior to Alphavirus Budding." *Journal of Virology* 84 (21): 11145–51.

Sourisseau, Marion, Daniel J. P. Lawrence, Megan C. Schwarz, Carina H. Storrs, Ethan C. Veit, Jesse D. Bloom, and Matthew J. Evans. 2019. "Deep Mutational Scanning Comprehensively Maps How Zika Envelope Protein Mutations Affect Viral Growth and Antibody Escape." *Journal of Virology* 93 (23). https://doi.org/10.1128/JVI.01291-19.

Spuul, P., G. Balistreri, L. Kaariainen, and T. Ahola. 2010. "Phosphatidylinositol 3-Kinase-, Actin-, and Microtubule-Dependent Transport of Semliki Forest Virus Replication Complexes from the Plasma Membrane to Modified Lysosomes." *Journal of Virology*. https://doi.org/10.1128/jvi.00477-10.

Strauss, J. H., and E. G. Strauss. 1994. "The Alphaviruses: Gene Expression, Replication, and Evolution." *Microbiological Reviews* 58 (3): 491–562.

Sun, Chengqun, Christina L. Gardner, Alan M. Watson, Kate D. Ryman, and William B. Klimstra. 2014. "Stable, High-Level Expression of Reporter Proteins from Improved Alphavirus Expression Vectors to Track Replication and Dissemination during Encephalitic and Arthritogenic Disease." *Journal of Virology* 88 (4): 2035–46.

Traag, V. A., L. Waltman, and N. J. van Eck. 2019. "From Louvain to Leiden: Guaranteeing Well-Connected Communities." *Scientific Reports*. https://doi.org/10.1038/s41598-019-41695-z.

Verhoeven, Kristien, Peter De Jonghe, Katrien Coen, Nathalie Verpoorten, Michaela Auer-Grumbach, Jennifer M. Kwon, David FitzPatrick, et al. 2003. "Mutations in the Small GTP-Ase Late Endosomal Protein RAB7 Cause Charcot-Marie-Tooth Type 2B Neuropathy."

American Journal of Human Genetics 72 (3): 722–27.

Wang, Chang, Christian V. Forst, Tsui-Wen Chou, Adam Geber, Minghui Wang, Wissam Hamou, Melissa Smith, et al. 2020. “Cell-to-Cell Variation in Defective Virus Expression and Effects on Host Responses during Influenza Virus Infection.” *mBio* 11 (1). https://doi.org/10.1128/mBio.02880-19.

Waskom, Michael L., Dharshan Kumaran, Alan M. Gordon, Jesse Rissman, and Anthony D. Wagner. 2014. “Frontoparietal Representations of Task Context Support the Flexible Control of Goal-Directed Cognition.” *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34 (32): 10743–55.

Wickham, Hadley. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Yin, Jun, Christina L. Gardner, Crystal W. Burke, Kate D. Ryman, and William B. Klimstra. 2009. “Similarities and Differences in Antagonism of Neuron Alpha/beta Interferon Responses by Venezuelan Equine Encephalitis and Sindbis Alphaviruses.” *Journal of Virology* 83 (19): 10036–47.

Zanini, Fabio, Szu-Yuan Pu, Elena Bekerman, Shirit Einav, and Stephen R. Quake. 2018. “Single-Cell Transcriptional Dynamics of Flavivirus Infection.” *eLife* 7 (February). https://doi.org/10.7554/eLife.32942.

Zanini, Fabio, Makeda L. Robinson, Derek Croote, Malaya Kumar Sahoo, Ana Maria Sanz, Eliana Ortiz-Lasso, Ludwig Luis Albornoz, et al. 2018. “Virus-Inclusive Single-Cell RNA Sequencing Reveals the Molecular Signature of Progression to Severe Dengue.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (52): E12363–69.

Zhai, Qianting, Michael B. Landesman, Howard Robinson, Wesley I. Sundquist, and Christopher P. Hill. 2011. “Structure of the Bro1 Domain Protein BROX and Functional Analyses of the ALIX Bro1 Domain in HIV-1 Budding.” *PloS One* 6 (12): e27466.

Zhang, T., S. H. Wong, B. L. Tang, Y. Xu, and W. Hong. 1999. “Morphological and Functional Association of Sec22b/ERS-24 with the Pre-Golgi Intermediate Compartment.” *Molecular Biology of the Cell* 10 (2): 435–53.

Zhou, Yingyao, Bin Zhou, Lars Pache, Max Chang, Alireza Hadj Khodabakhshi, Olga Tanaseichuk, Christopher Benner, and Sumit K. Chanda. 2019. “Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets.” *Nature Communications* 10 (1): 1523.

Materials and methods

Cells

U-87 MG and BHK-21 cell lines were obtained from ATCC (Manassas, VA). Cells were grown in Dulbecco's Modified Eagle's medium (DMEM, Mediatech, Manassas, VA), supplemented with 1% Penicillin-Streptomycin solution, 1% L-glutamine 200 mM (Thermo Fisher Scientific, Waltham, MA) and 10% Fetal Bovine Serum (FBS, Omega Scientific, INC, Tarzana, CA). Cells were maintained in a humidified incubator with 5% CO2 at 37 ˚C. Cells were tested negative for mycoplasma by the MycoAlert mycoplasma detection kit (Lonza, Morristown, NJ).

Plasmids and virus constructs

The plasmids encoding infectious VEEV TC-83 with a GFP reporter (VEEV TC-83-Cap-eGFP-Tav, hereafter VEEV-TC-83-GFP) or a nanoluciferase reporter (VEEV TC-83-Cap-NLuc-Tav, hereafter VEEV-TC-83-nLuc), were a gift from Dr. William B. Klimstra (Department of Immunology, University of Pittsburgh) (Sun et al. 2014). Open reading frames (ORFs) encoding 24 hits were selected from the Human ORFeome library of cDNA clones (Open Biosystems) (Rual et al. 2004) and recombined into a FLAG (for FLAG tagging) vector using Gateway technology (Invitrogen).

Virus production

Viral RNA (vRNA) (VEEV-TC-83-GFP or nLuc) was transcribed *in vitro* from cDNA plasmid templates linearized with MluI via MegaScript Sp6 kit (Invitrogen #AM1330) and electroporated into BHK-21 cells. VEEV was harvested from the supernatant 24 hours postelectroporation, clarified from cell debris by centrifugation, and stored at -80 °C. Virus stock titers were determined by standard BHK-21 cell plaque assay, and titers were expressed as PFU/ml.

Infection assays

U-87 MG cells were infected with VEEV-TC-83-GFP at various MOIs (0, 0.1, and 1) and harvested at various time points postinfection. For the functional screens, U-87 MG cells were infected with VEEV-TC83-nLuc in 8 replicates at MOI of 0.01. Overall infection was measured at 18 hpi via a nanoluciferase assay using a luciferin solution obtained from the hydrolysis of its O-acetylated precursor, hikarazine-103 (prepared by Dr. Yves Janin, Pasteur Institute, France) as a substrate (Coutant, Goyard, et al. 2019; Coutant, Gagnot, et al. 2019).

Loss-of-function assays

siRNAs (1 pmol) were transfected into cells using lipofectamine RNAiMAX transfection reagent (Invitrogen) 96 hours prior to infection with VEEV-TC-83-nLuc at MOI of 0.01. Custom Cherry-Pick ON-TARGETplus siRNA library against 24 genes was purchased from Dharmacon (see Supplementary Table 1 for gene and siRNA sequence details).

Gain-of-function assays

Individual plasmids encoding 24 human genes or empty control vector were transfected individually into U-87 MG cells with lipofectamine-3000 (Invitrogen) 48 hours prior to infection with VEEV-TC-83-nLuc at MOI of 0.01.

Viability assays

Viability was measured using alamarBlue reagent (Invitrogen) according to the manufacturer's protocol. Fluorescence was detected at 560 nm on an Infinite M1000 plate reader (Tecan).

Detection of infected cells using VEEV-specific capture oligo

To optimize the viscRNA-Seq protocol for a wide dynamic range of vRNA amount per VEEV-infected cells, we designed and screened eight oligo capture as follows. The sequences that hybridize to VEEV sequence are underlined.

Oligo name	Sequence (5’ to 3’)	Position on VEEV genome
VEEV_1	AAGCAGTGGTATCAACGCAGAGTACTTCCTTATCAGTTATTTCCTTACAG	353 - 377
VEEV_2	AAGCAGTGGTATCAACGCAGAGTACAGATAATTTTTCACTCTTGAGTACA	1742 - 1766
VEEV_3	AAGCAGTGGTATCAACGCAGAGTACTTTTAGGTCTTATAATGGCTATGAG	2442 - 2466
VEEV_4	AAGCAGTGGTATCAACGCAGAGTACTGCTGATAGTGATGGTATTTATATG	3700 - 3724
VEEV_5	AAGCAGTGGTATCAACGCAGAGTACCTACTGACTTGTAATTGTTATCGTT	4320 - 4344
VEEV_6	AAGCAGTGGTATCAACGCAGAGTACGTAGTAATTCTTCTTTTTCTTGGTC	5823 - 5847
VEEV_7	AAGCAGTGGTATCAACGCAGAGTACTTCATTATTACACGCATATTTCTTG	6383 - 6407
VEEV_8	AAGCAGTGGTATCAACGCAGAGTACGCATCTATAATCTTGACTTCCATAT	7162 - 7286

To screen these capture oligo, we first generated cDNA from VEEV-infected cells in the presence of each or combinations of VEEV-specific capture oligo. Specifically, 30 pg of both vRNA and cellular RNA purified from VEEV-infected cells was reverse-transcribed to cDNA in a reaction containing SuperScript™ IV reverse transcriptase, 1X First Strand buffer (Invitrogen), 5 mM DTT, 1 M betaine, 6 mM $MgCl_2$, 1 µM oligo dT and each or combinations of 100 nM reverse

VEEV oligo capture. Subsequently, cDNA underwent 21-cycle PCR amplification using ISPCR primers. cDNA was then purified using Ampure XP beads (Beckman Coulter) at the ratio of 0.8 and eluted in 15 µL EB buffer. Fragments of purified, concentrated cDNA were visualized and quantified using bioanalyzer (DNA High Sensitivity kit, Agilent Technologies). To quantify the amount of vRNA captured by each or combinations of oligo capture, these purified cDNA were also subjected to qPCR (Hot-start OneTaq (New England Biolabs), 1x Standard Taq buffer, 1x Evagreen (Biotium), forward primer: ATTCTAAGCACAAGTATCATTGTAT and reverse primer: TTAGTTGCATACTTATACAATCTGT located upstream of all the capture oligos. VEEV_1 and VEEV_2 yielded the highest copies of viral cDNA and did not generate significant primer dimers. Therefore, this combination of the capture oligo was selected for downstream experiments.

Single cell sorting

At each time point, cells were trypsinized for 10 min, spun and resuspended in 1 mL fresh media. Within 15 min, cells were pelleted again and resuspended in 2 ml 1X phosphate-buffered saline (PBS) buffer at a concentration of 10^6 cells per ml. Cells were filtered through a 40 µm filter into a 5 ml FACS tube and sorted on a Sony SH800 sorter using SYTOX™ Blue dead cell stain (ThermoFisher) to distinguish living cells from dead cells and debris. VEEV harboring cells were sorted based on GFP signal. Cells were sorted into 384-well PCR plates containing 0.5 µl of lysis buffer using ‘Single cell’ purity mode. A total of 12 384-well plates of single cells were sorted for the VEEV time course.

Lysis buffer, reverse transcription, and PCR

To capture and amplify both mRNA and vRNA from the same cell, the Smart-seq2 protocol was adapted (Picelli et al., 2014). All volumes were reduced by a factor of 12 compared to the original protocol to enable high-throughput processing of 384-well plates. ERCC spike-in RNA was added at a concentration of 1:10 of the normal amount. The lysis buffer contained 100nM

of oligo-dT primer, 100 mM of virus specific capture oligo mix (i.e. VEEV_1 and VEEV_2) to capture the positive-stranded virus RNA.

Other virus-specific primers and higher primer concentrations were tested but resulted in a large fraction of primer dimers. In order to reduce interference between the virus-specific primer and the Template Switching Oligo (TSO) used to extend the RT products, a 5'-blocked biotinylated TSO was used at the standard concentration. Reverse transcription (RT) and PCR of the cDNA were performed in a total volume of 1 µl and 2.5 µl for each well respectively. The resulting cDNAs were amplified for 21 cycles. Lambda exonuclease was added to the PCR buffer at a final concentration of 0.0225 U/µl and the RT products were incubated at 37 ˚C for 30 min before melting the RNA-DNA hybrid (as it was observed that this reduced the amount of low-molecular weight bands from the PCR products). The cDNA was then diluted 1 to 7 in EB buffer for a final volume of 17.5 µl. All pipetting steps were performed using a Mosquito HTS robotic platform (TTP Labtech).

cDNA quantification

To quantify the amount of cDNA in each well after PCR, a commercial fluorometric assay was used (ThermoFisher Quant-It™ Picogreen). Briefly, 1 µl of cDNA and 50 µl of 1:200 dye-buffer mix were pipetted together into a flat-bottom 384-well plate (Corning 3711). For each plate, six wells were used as standard wells. 1 µl dd H2O was added into one standard well as blank. The standard solutions were diluted into 5 concentrations (0.1, 0.2, 0.4, 0.8, 1.6 ng/µl) and added 1µl into the remaining 5 standard wells. The plate was vortexed for 2 min, centrifuged, incubated in the dark for 5 min, and measured on a plate reader at wavelength 550 nm. cDNA concentrations were calculated via an affine fit to the standard wells.

Library preparation and sequencing

For each time point, one plate was sent for library preparation and sequencing. In total, 6 plates (2304 cells) were prepared. Sequencing libraries were prepared using the illumina Nextera XT kit according to the manufacturer's instructions, with the following exceptions: (1) we used a smaller reaction volume (around 1 µl per cell); (2) we chose a slightly higher cDNA concentration (0.4 ng/µl) as input, to compensate for the lack of bead purification upstream; (3) we used the commercial 24 i7 barcodes and the 64 new i5 barcode sequences. We noticed a low level of cross-talk between these barcodes, indicated by up to five virus reads found in a few uninfected cells. However, considering that a sizeable fraction of cells in the same sequencing run (late infected and high MOI) had thousands of virus reads, the amount of cross-talk between barcodes appears to be of the order of 1 in 10,000 or less. We used Illumina Novaseq sequencer for sequencing.

Bioinformatics pipeline

Sequencing reads were mapped against the human GRCh38 genome with supplementary ERCC sequences and TC-83-VEEV-GFP genome using STAR Aligner (Dobin et al. 2013) . Genes were counted using htseq-count (Anders, Pyl, and Huber 2015). The Stanford high performance computing cluster Sherlock 2.0 was used for the computations. Once the gene/virus counts were available, the downstream analysis was performed on laptops using the packages Seurat (Butler et al. 2018) and singlet (https://github.com/iosonofabio/singlet), as well as custom R and Python scripts. Ggplot2 (Wickham 2016), matplotlib (Hunter 2007) and seaborn (Waskom et al. 2014) were used for plotting.

For the mutational analysis, all reads mapping to VEEV were extracted from all cells with a unique identifier of the cell of origin, and all four possible alleles at each nucleotide were counted by custom scripts based on pysam (https://github.com/pysam-developers/pysam) and wrapped in an xarray Dataset (Hoyer and Hamman 2017). The analysis was restricted to infected cells with an average of 100 or more reads per viral genomic site to reduce shot noise.

Comparison with flaviviruses was performed as follows. First, host genes with similar expression (within a factor of 10) in counts per millions (cpm) were identified. Within that class, correlations with vRNA for VEEV, DENV, ZIKV were computed separately. Host factors with the highest discrepancies between pairs of viruses were identified. For Figures 5A-C, a gene was chosen from the most discrepant genes exemplifying the different behaviors observed and the cells were scattered using vRNA abundance and gene expression axes, and colored by virus. For Figure 5D, the host counts for each gene from all three experiments (in cpm) were added and fractions belonging to each experiment were computed. Because the sum is constrained to be 100%, ternary plots could be used for plotting the three different fractions in two dimensions. For figures 5E-F, for each gene shown we computed its percentile in correlation with DENV and ZIKV vRNA, i.e. the percentage of other host genes with a correlation less than this focal gene. This transformation emphasizes the top correlates/anticorrelates against batch effects and different multiplicities of infection in the DENV and ZIKV experiments. For figures 5G-I, published tables of counts and metadata were downloaded from links present in each publication, normalized to counts per millions, and filtered for low-quality cells. We computed the correlation of host gene expression and vRNA in each experiment, then features were selected that had a high rank in at least one virus and the selected correlation coefficients were centered and normalized between -1 and 1 for each virus to enable meaningful cross-experiment comparison. Principal Component Analysis (PCA), t-SNE, similarity graphs, and Leiden clustering (Traag et al. 2019) were computed and plotted

Cell selection and normalization

The criteria to select cells were as follows: total reads > 300,000, gene counts > 500 and a ratio of ERCC spike-in RNA to total reads ratio < 0.05. Based on these criteria, 2004 out of 2301 cells were selected for downstream analysis. Due to the high viral copies of VEEV in cells infected for 12 and 24 hrs (more than 10%), traditional normalization (dividing by total reads)

caused a bias which underestimated the expression of host genes. To avoid this, we normalized gene counts to ERCC total reads, since these are not affected by the virus. Each gene count column (including virus reads) was thus divided by ERCC total reads and then log transformed.

Data and code availability

The single cell RNA-Seq data for this study is available on GEO at submission number: XXX. The code used in the computational analyses can be found at https://github.com/saberyzy/VEEV-single_cell.

1014 **Figures and legends:**

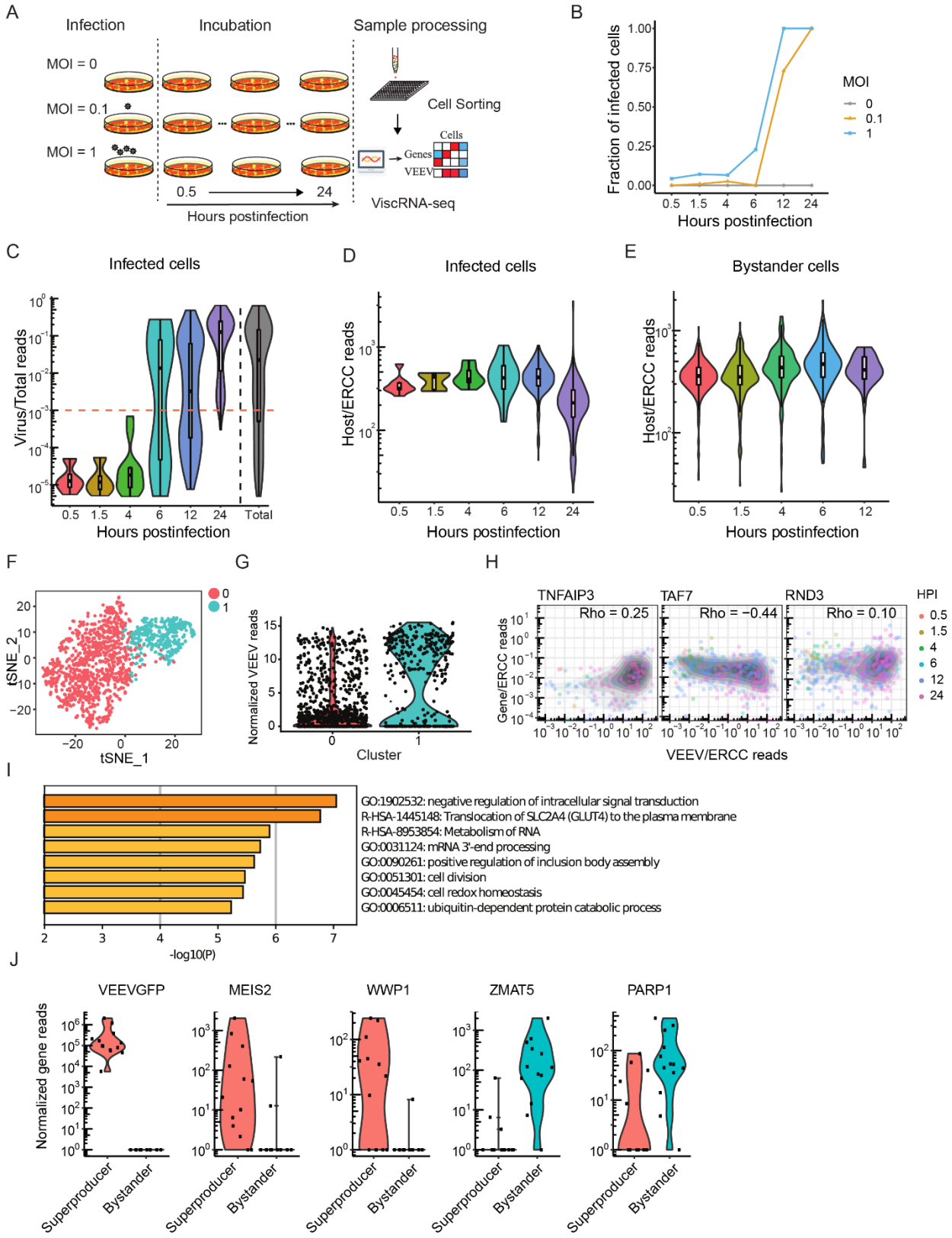
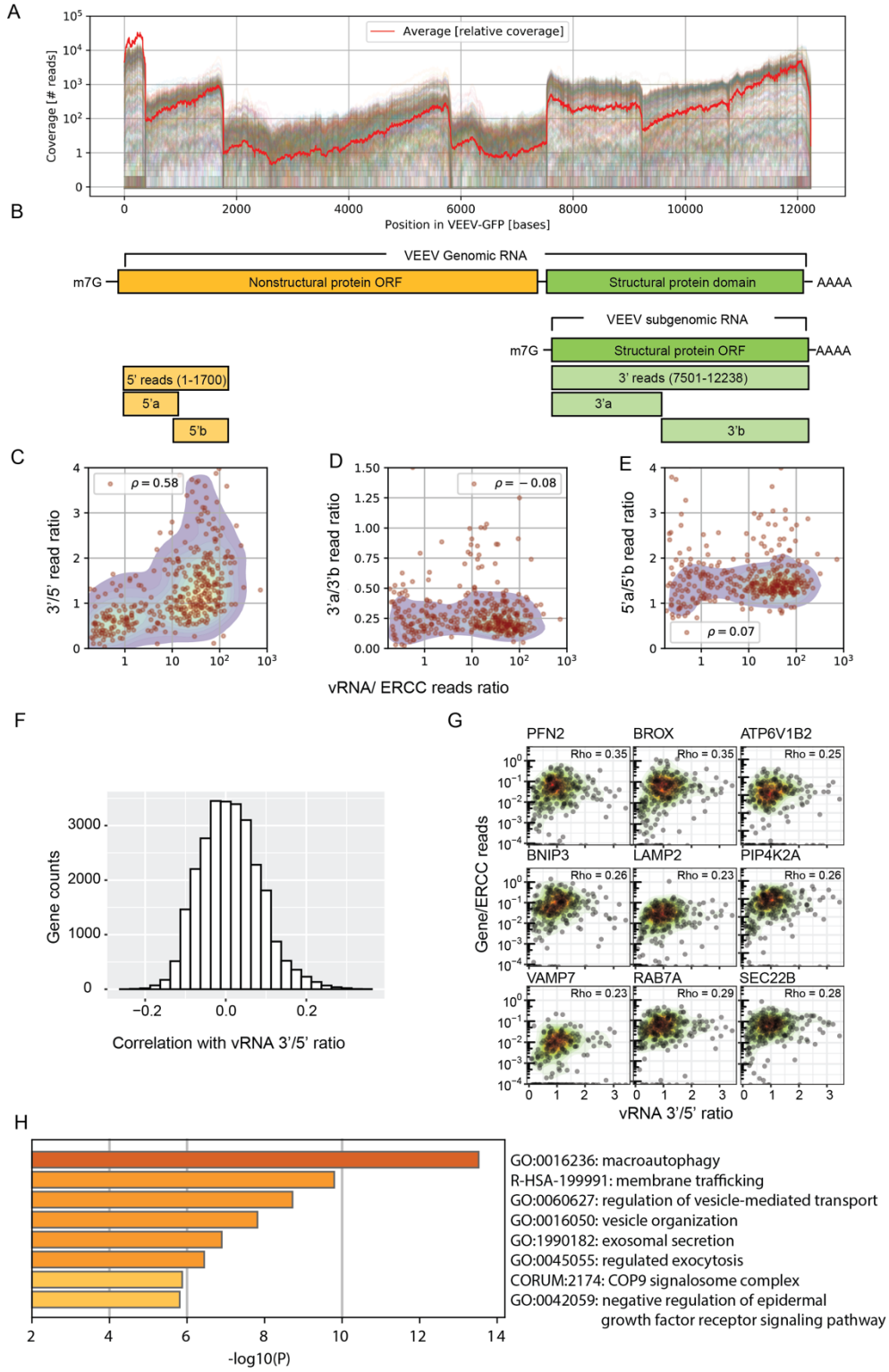


Figure 1. Cell-to-cell heterogeneity of VEEV and host gene expression.

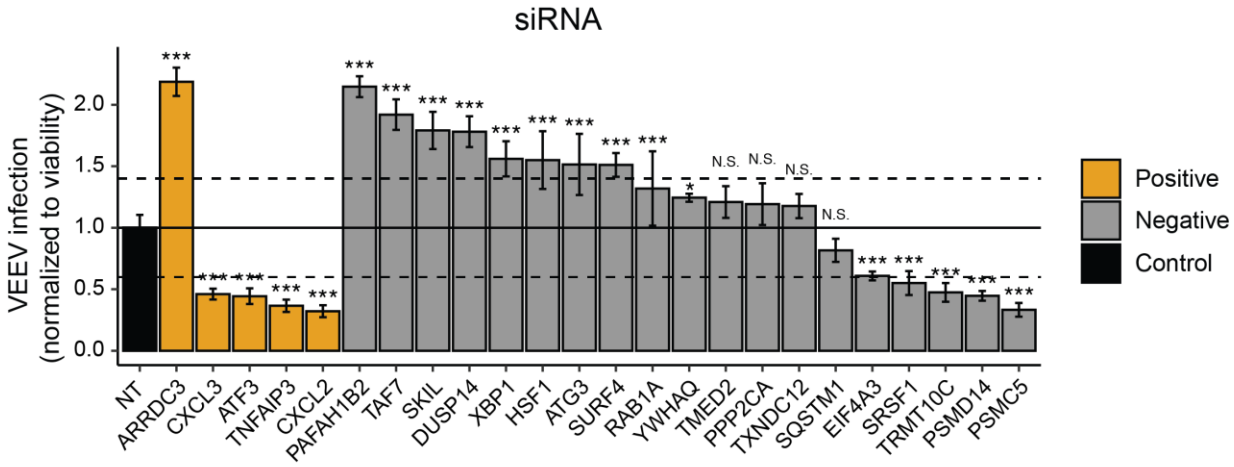
(A) Schematic of the experimental setup. (B) The fraction of VEEV-TC-83-infected U-87 MG cells over time for two MOIs. (C) Violin plots depicting the ratio of virus to total cDNA reads over time. Total (on the right) represents the distribution of cells from all time points. The horizontal dotted line represents the threshold dividing cells to “low vRNA” and “high vRNA” harboring cells. (D-E) Violin plots showing host cDNA to ERCC reads in infected (D) and bystander (E) cells derived from the same samples. (F) t-SNE plot of cells from infected cultures (MOIs 0.1 and 1) on the combined host gene expression data independently of vRNA abundance. Each color refers to a Louvain cluster (0 and 1). (G) The distribution of vRNA abundance in the two cell clusters defined in (F). (H) Representative scatter plots of host gene expression versus vRNA abundance and corresponding Rho Spearman correlation coefficients. Each dot is a single cell colored by the time postinfection, and the shaded contours indicate cell density (greyscale, darker is higher). (I) Molecular function terms and P values derived from Gene Ontology (GO) enrichment analysis of 206 genes that are both differentially expressed between high vRNA and mock-infected cells and correlated with vRNA. (J) Violin plots of vRNA (VEEV-GFP) abundance and expression of representative genes that are differentially expressed between “superproducer” cells (n=13) and bystander cells (n=13) at 6 hpi with an MOI of 1. HPI, hours postinfection; MOI, multiplicity of infection; ERCC, External RNA Controls Consortium.


A
Average [relative coverage]
Coverage [# reads]
Position in VEEV-GFP [bases]
B
VEEV Genomic RNA
m7G
Nonstructural protein ORF
Structural protein domain
AAAA
VEEV subgenomic RNA
Structural protein ORF
3' reads (7501-12238)
3'a
3'b
5' reads (1-1700)
5'a
5'b
C
ρ = 0.58
3'/5' read ratio
D
ρ = −0.08
3'a/3'b read ratio
E
ρ = 0.07
5'a/5'b read ratio
vRNA/ ERCC reads ratio
F
Gene counts
Correlation with vRNA 3'/5' ratio
G
PFN2
BROX
ATP6V1B2
BNIP3
LAMP2
PIP4K2A
VAMP7
RAB7A
SEC22B
Rho = 0.35
Rho = 0.35
Rho = 0.25
Rho = 0.26
Rho = 0.23
Rho = 0.26
Rho = 0.23
Rho = 0.29
Rho = 0.28
Gene/ERCC reads
vRNA 3'/5' ratio
H
GO:0016236: macroautophagy
R-HSA-199991: membrane trafficking
GO:0060627: regulation of vesicle-mediated transport
GO:0016050: vesicle organization
GO:1990182: exosomal secretion
GO:0045055: regulated exocytosis
CORUM:2174: COP9 signalosome complex
GO:0042059: negative regulation of epidermal growth factor receptor signaling pathway
-log10(P)

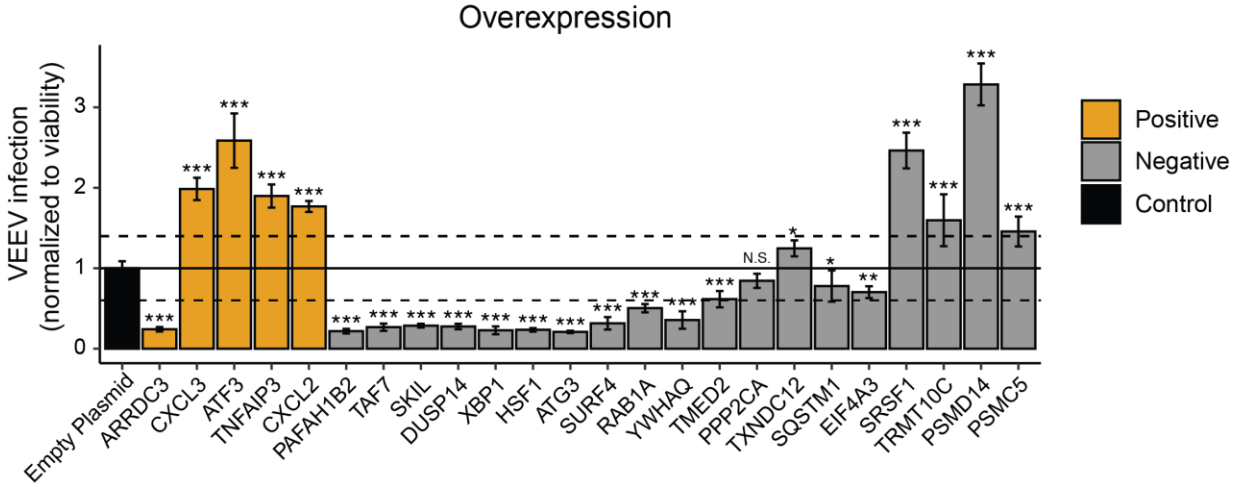
1036

Figure 2. The expression of genes involved in intracellular membrane trafficking correlates with the ratio of 3' to 5' vRNA reads. (A) Coverage of viral reads over the entire VEEV genome. Each line is a cell, and the red line is a scaled average across all cells. (B) Genome architecture of VEEV highlighting the nonstructural (yellow) and structural (green) protein domains. (C) Scatter plot showing positive correlation of VEEV 3'/5' read ratio with cellular vRNA abundance. Each dot is an infected cell. (D-E) Scatter plots showing no correlation between the 3'a/3'b read ratio (D) and 5'a/5'b read ratio (E) and cellular vRNA abundance. (F) Histogram of Spearman correlation coefficients between all host genes and the 3'/5' read ratio. (G) Representative scatter plots of host gene expression versus vRNA 3'/5' read ratio and corresponding Rho Spearman correlation coefficients. Each dot is a cell and contour plots indicate cell density (low to high, green to red). (H) Gene enrichment analysis of top 300 genes positively correlated with the 3'/5' read ratio.

A


siRNA
VEEV infection
(normalized to viability)
0.0
0.5
1.0
1.5
2.0
NT
ARRDC3
CXCL3
ATF3
TNFAIP3
CXCL2
PAFAH1B2
TAF7
SKIL
DUSP14
XBP1
HSF1
ATG3
SURF4
RAB1A
YWHAQ
TMED2
PPP2CA
TXNDC12
SQSTM1
EIF4A3
SRSF1
TRMT10C
PSMD14
PSMC5
N.S.
Positive
Negative
Control

B


Overexpression
VEEV infection
(normalized to viability)
0
1
2
3
Empty Plasmid
ARRDC3
CXCL3
ATF3
TNFAIP3
CXCL2
PAFAH1B2
TAF7
SKIL
DUSP14
XBP1
HSF1
ATG3
SURF4
RAB1A
YWHAQ
TMED2
PPP2CA
TXNDC12
SQSTM1
EIF4A3
SRSF1
TRMT10C
PSMD14
PSMC5
N.S.
Positive
Negative
Control

1057

Figure 3. Validation of candidate VEEV proviral and antiviral genes. VEEV infection relative to non-targeting (NT) siRNA (A) or empty plasmid (B) controls following siRNA-mediated knockdown (A) or overexpression (B) of the indicated host factors measured by luminescence assays at 18 hpi (MOI = 0.01) of U-87 MG cells with VEEV-TC-83-NLuc and normalized to cell viability. Columns are color-coded based on the correlation of the respective gene with vRNA abundance via viscRNA-Seq: yellow for genes that are positively correlated with vRNA and grey for genes that are negatively correlated with vRNA. Both data sets are pooled from two independent experiments with six replicates each. Shown are means ± SD; *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$ relative to the respective control by 1-way ANOVA followed by Dunnett's post hoc test. The dotted lines represent the cutoffs for positivity. Cellular viability measurements are shown in supplemental Figure S3.

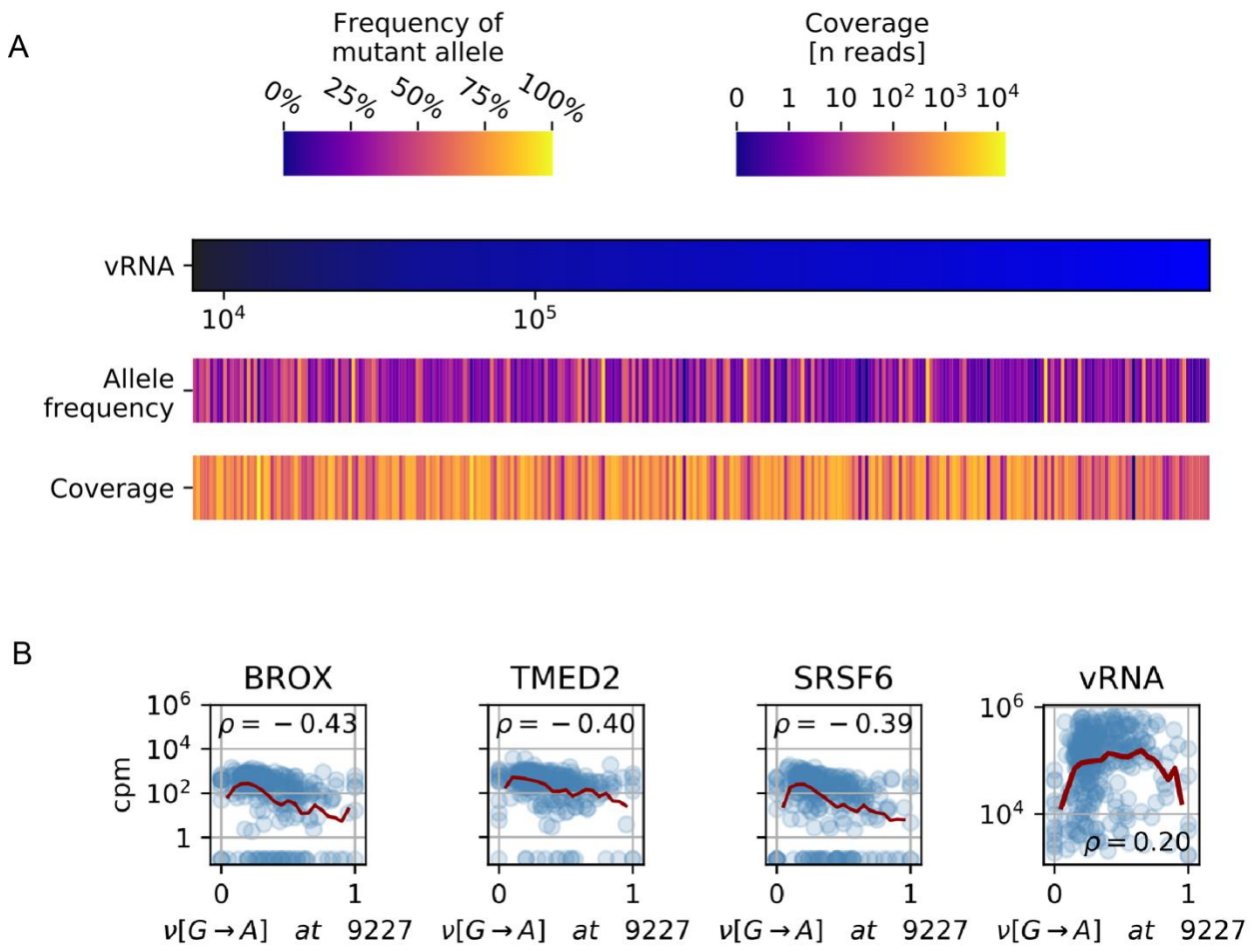
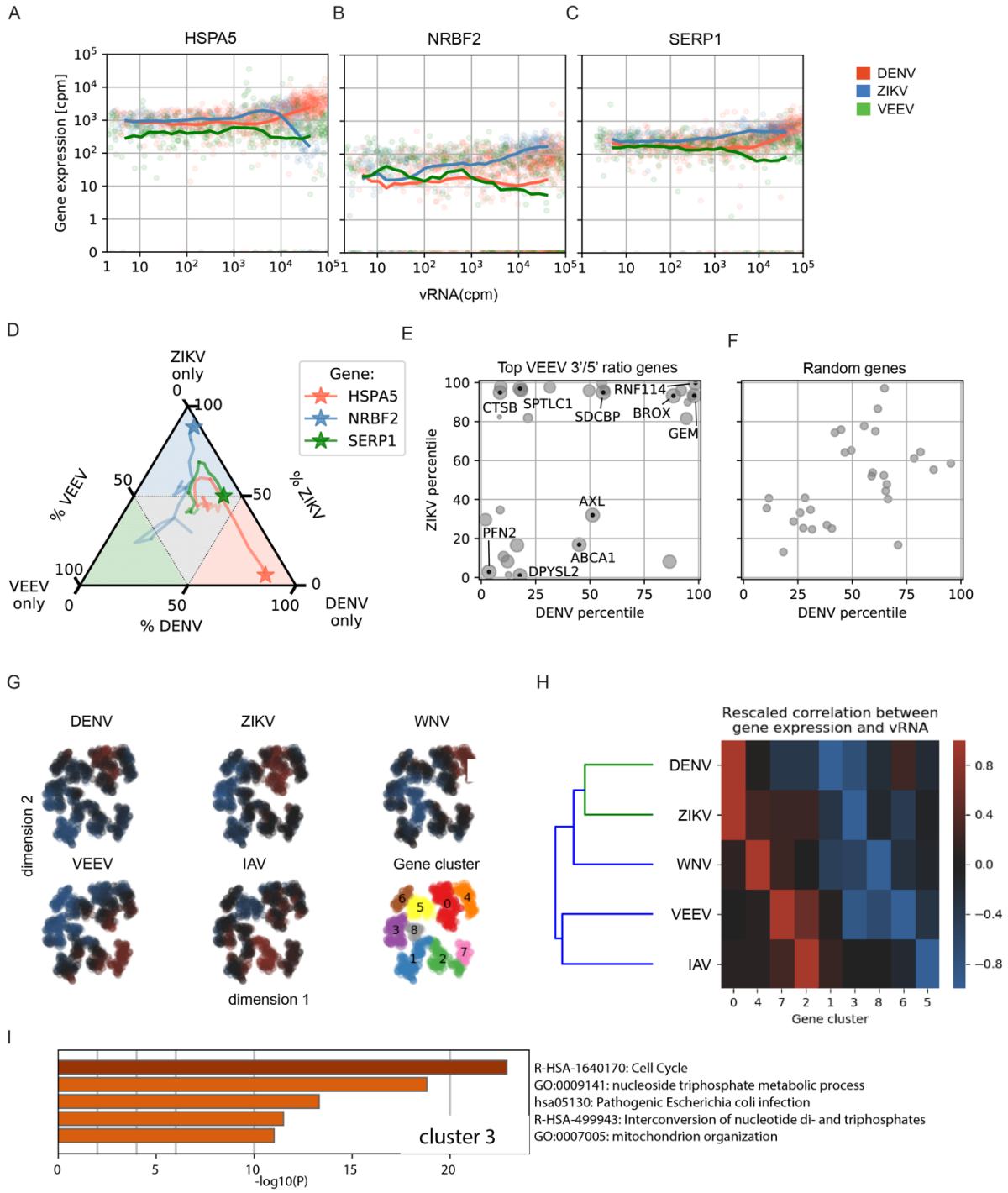


Figure 4. Viral polymorphisms correlate with expression of specific host genes. (A) Heat maps of SNP (allele) frequencies and number of reads (coverage) at the most variable sites in the VEEV-TC-83-GFP genome. Each vertical bar is a cell and the x axis orders cells by increasing vRNA abundance. (B) Correlation between the frequency of the A allele at the most variable site (position 9227) in the VEEV-TC-83-GFP genome and the expression of representative host genes and vRNA.


A
HSPA5
B
NRBF2
C
SERP1
Gene expression [cpm]
vRNA(cpm)
DENV
ZIKV
VEEV
D
ZIKV only
% VEEV
% ZIKV
VEEV only
DENV only
% DENV
Gene:
HSPA5
NRBF2
SERP1
E
Top VEEV 3'/5' ratio genes
ZIKV percentile
DENV percentile
CTSB
SPTLC1
RNF114
SDCBP
BROX
GEM
AXL
PFN2
ABCA1
DPYSL2
F
Random genes
DENV percentile
G
DENV
ZIKV
WNV
VEEV
IAV
Gene cluster
dimension 2
dimension 1
H
Rescaled correlation between gene expression and vRNA
DENV
ZIKV
WNV
VEEV
IAV
Gene cluster
I
cluster 3
-log10(P)
R-HSA-1640170: Cell Cycle
GO:0009141: nucleoside triphosphate metabolic process
hsa05130: Pathogenic Escherichia coli infection
R-HSA-499943: Interconversion of nucleotide di- and triphosphates
GO:0007005: mitochondrion organization

1079

Figure 5. Comparative viscRNA-Seq analysis across five RNA viruses. (A-C) Scatter plots of representative gene expression versus vRNA in single cells during DENV (orange), ZIKV (blue), and VEEV (green) infection. Dots indicate single cells, lines are running averages at increasing vRNA abundances. (D) A ternary plot for the three genes shown in (A-C). Lines indicate the percentage of host gene reads from each experiment at increasing vRNA abundance, divided by the number of reads across all three experiments. Stars indicate the cells with the highest vRNA abundance. Proximity to a corner of the triangle suggests that the expression of the gene is much higher during late infection by the corresponding virus than the other two. Colored triangles in the background indicate areas of upregulation specific to each virus. (E, F) Correlation between expression and vRNA during DENV versus ZIKV infection of the top genes that positively correlate with the VEEV 3'/5' read ratio (E) or a similar number of random genes (F). Each dot is a gene and the axis coordinate is the percentage of genes with a correlation with vRNA smaller than the the gene of interest. For (E), size of the dot increases with the correlation with VEEV 3'/5' read ratio (top correlated gene is largest). (G) t-SNE embedding of host genes correlation with vRNA during infection by 5 individual RNA viruses. Blue and red indicate downregulation and upregulation during infection, respectively. Several clusters of genes are observed (0-8). (H) Hierarchical clustering of host gene clusters highlights that upregulation is more virus-specific and is consistent with the known phylogeny. (I) Gene set enrichment analysis of cluster 3, which contains commonly downregulated genes. cpm, count per million; WNV, West Nile virus; IAV, influenza A virus.

Supplementary materials

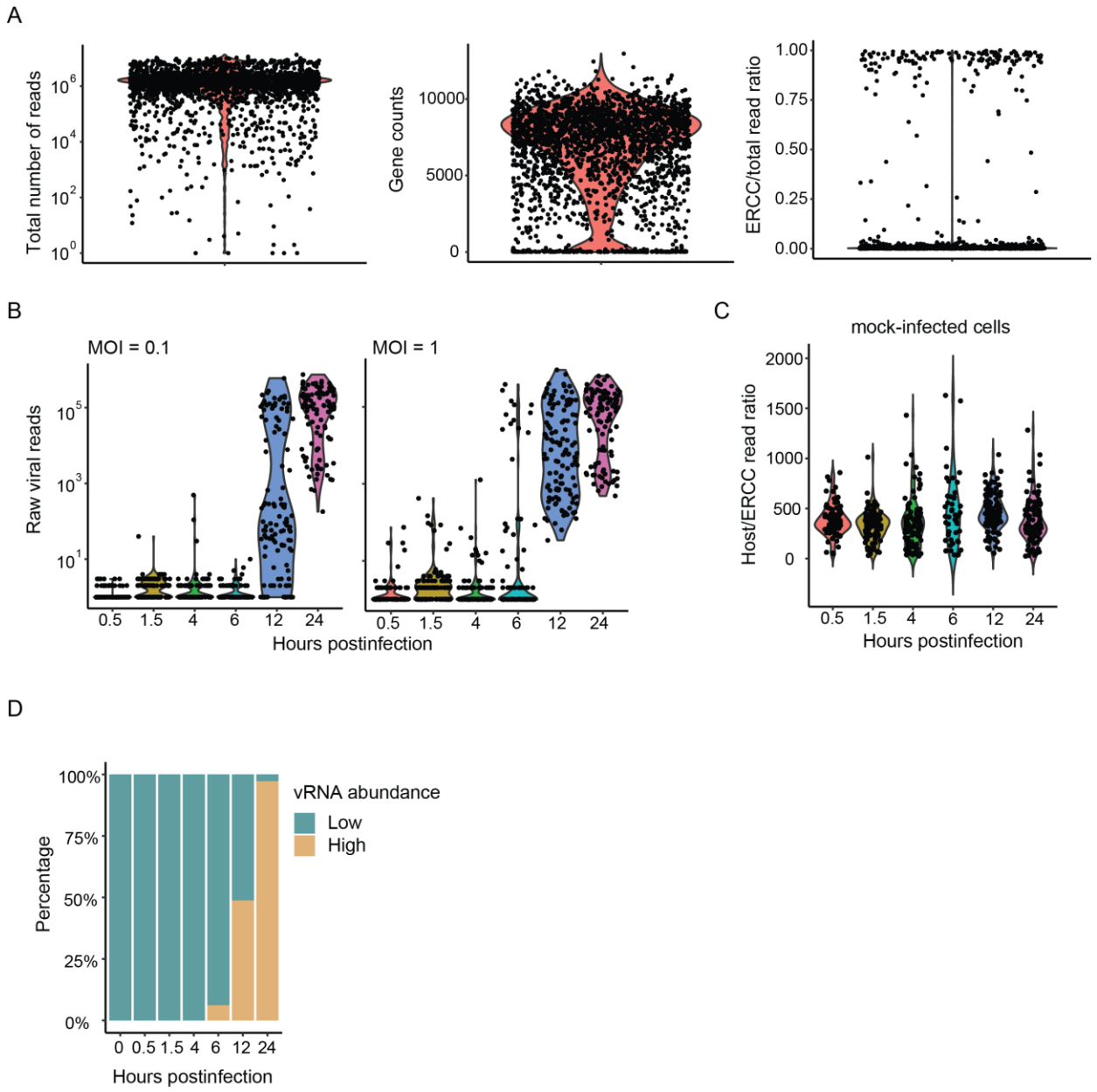


Figure S1. Quality control, normalization and viral load distribution of the sample. (A) Quality control of the VEEV dataset. Shown are the total number of reads (left panel), gene counts (middle panel) and ratio of ERCC spike-in RNA reads to total reads (right panel). Cell selection criteria included: total reads > 300,000, gene counts > 500 and a ratio of ERCC spike-in RNA to total reads < 0.05. (B) Raw viral reads over time in cells infected with MOI of 0.1 (left panel) and 1 (right panel). (C) Host to ERCC read ratio over time in mock-infected cells.

Individual dots in panels A, B, and C represent single cells. (D) Percentage of low and high vRNA-harboring cells at each time point. MOI, multiplicity of infection; ERCC, External RNA Controls Consortium.

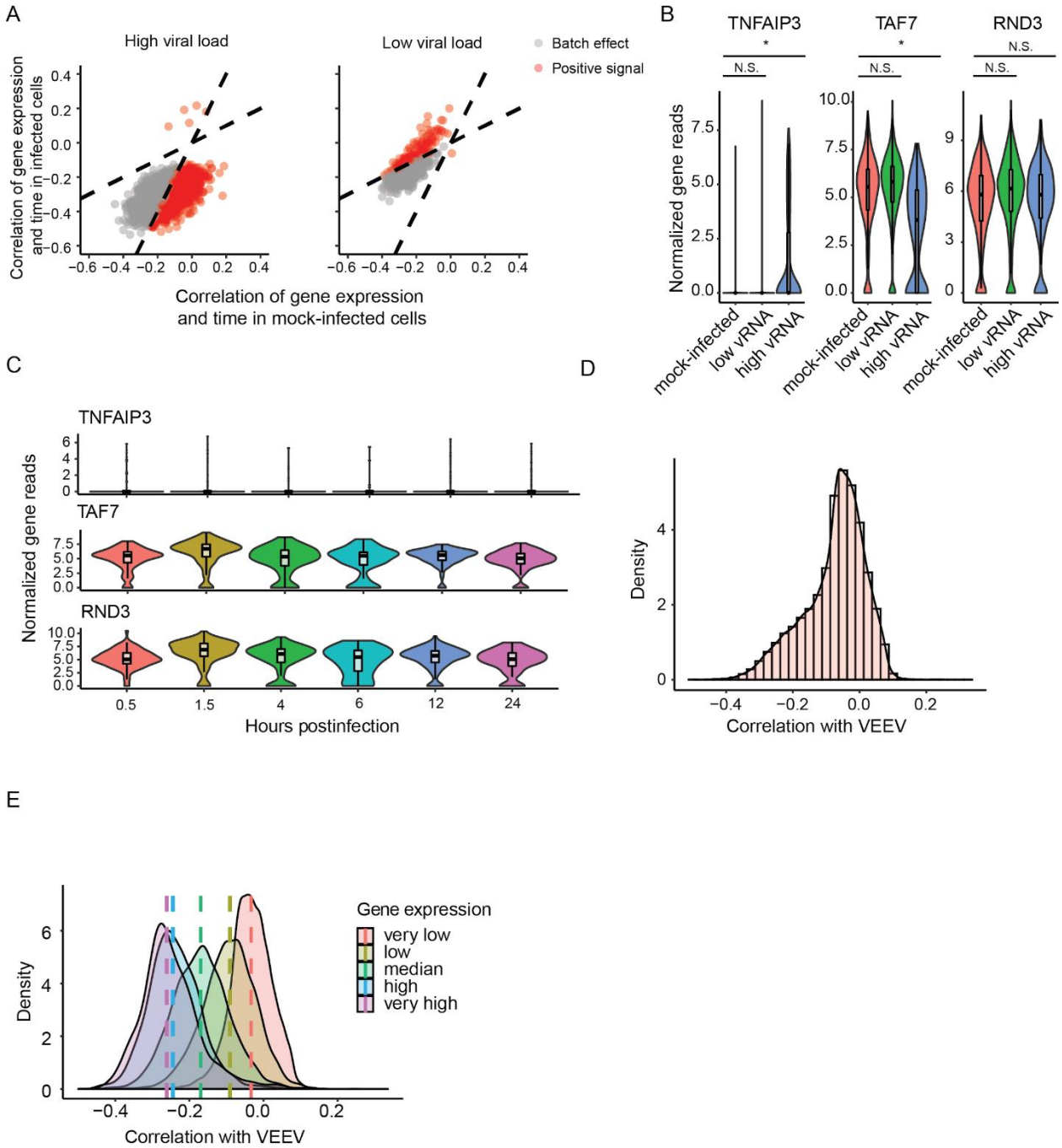
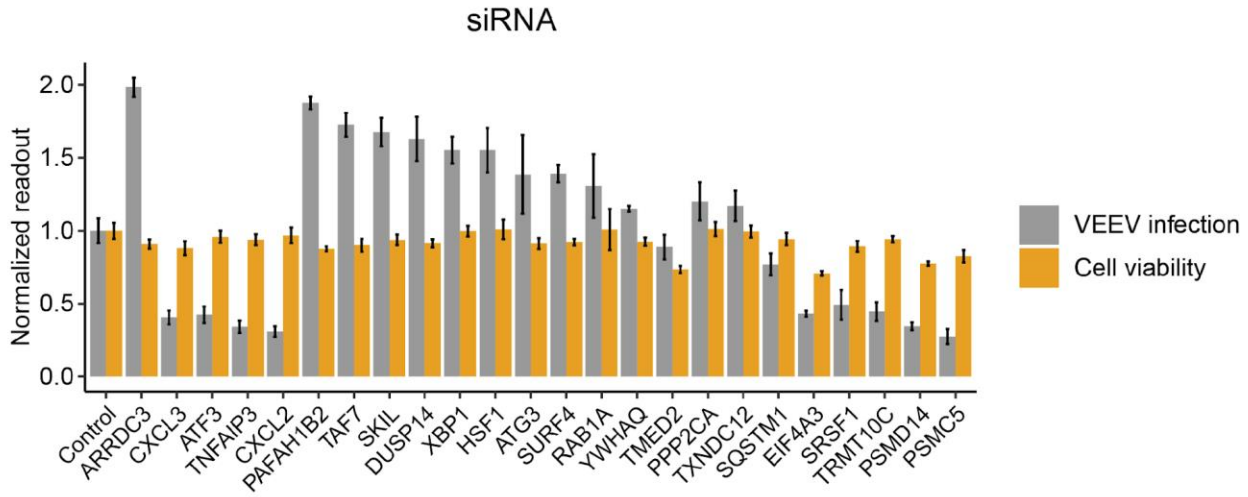


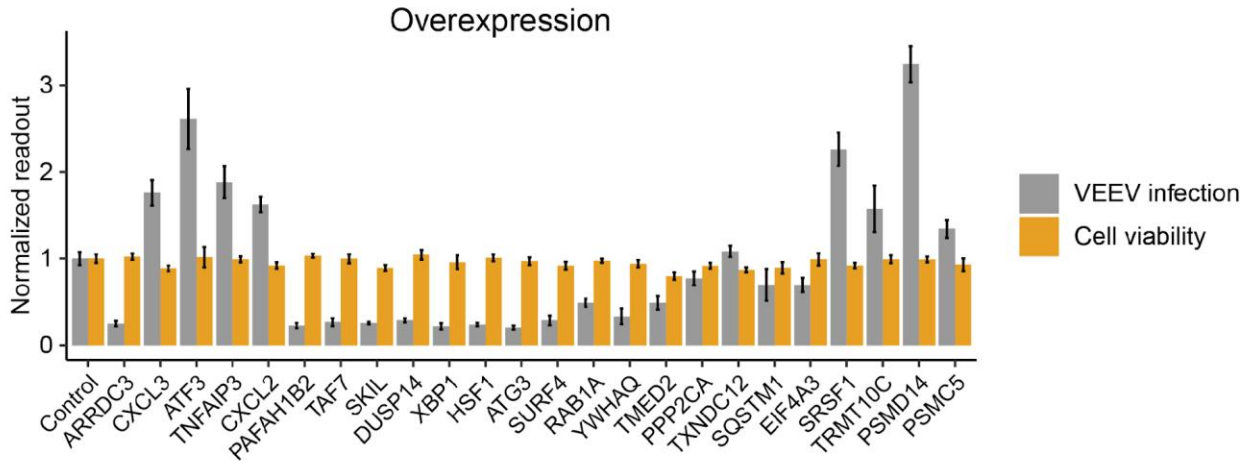
Figure S2. Batch effect removal, representative differentially expressed genes (DGEs) and correlation analysis. (A) Two-dimensional scatter plots showing correlations between gene expression and time postinfection in VEEV-infected (left: high vRNA abundance, right: low vRNA abundance) and mock-infected cells. The dotted lines represent cut-off edge of 2-fold differences. Each dot represents a gene. Orange, genes demonstrating greater than 2-fold

difference between infected and mock-infected cells and thus considered DEGs; grey, genes whose expression is similarly altered over time between infected and mock-infected cells and thus likely represent batch effect. (B) Representative genes with distinct expression patterns among different cell groups. N.S., not significant. (C) The expression of genes shown in (B) does not significantly change over time in mock-infected cells. (D) Distribution of Spearman correlation coefficients between VEEV vRNA abundance and ~55,000 host genes. (E) Distributions of Spearman correlation coefficients shown in D stratified by the average expression level of the gene in mock-infected cells.

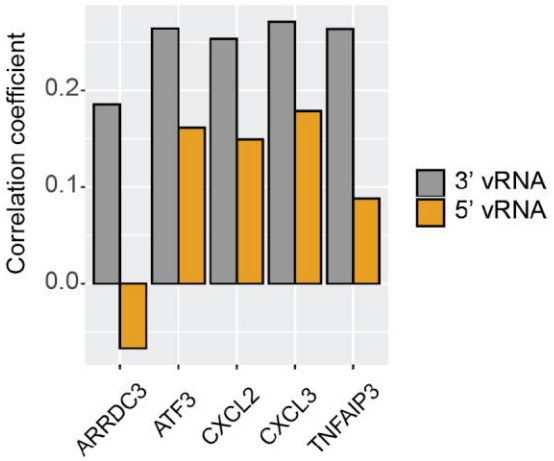
A


siRNA
Normalized readout
2.0
1.5
1.0
0.5
0.0
Control
ARRDC3
CXCL3
ATF3
TNFAIP3
CXCL2
PAFAH1B2
TAF7
SKIL
DUSP14
XBP1
HSF1
ATG3
SURF4
RAB1A
YWHAQ
TMED2
PPP2CA
TXNDC12
SQSTM1
EIF4A3
SRSF1
TRMT10C
PSMD14
PSMC5
VEEV infection
Cell viability

B


Overexpression
Normalized readout
3
2
1
0
Control
ARRDC3
CXCL3
ATF3
TNFAIP3
CXCL2
PAFAH1B2
TAF7
SKIL
DUSP14
XBP1
HSF1
ATG3
SURF4
RAB1A
YWHAQ
TMED2
PPP2CA
TXNDC12
SQSTM1
EIF4A3
SRSF1
TRMT10C
PSMD14
PSMC5
VEEV infection
Cell viability

C


Correlation coefficient
0.2
0.1
0.0
ARRDC3
ATF3
CXCL2
CXCL3
TNFAIP3
3' vRNA
5' vRNA

1164

Figure S3. Validation of proviral and antiviral factors. siRNA (A) and ectopic expression screens (B) testing the involvement of the indicated host factors in VEEV infection. Overall VEEV infection (grey) measured by luminescence assays and cell viability (orange) measured by alamarBlue assays in U-87 MG cells transfected with the indicated siRNAs or ectopically expressing the indicated cellular factors at 18 hpi with VEEV-TC-83-NLuc (MOI = 0.01). Data are expressed relative to NT siRNA or empty plasmid control. Both data sets are pooled from two independent experiments with six replicates each. Shown are means ± SD. (C) Correlation coefficients between proviral candidates with the 3’ (grey) and 5’ (orange) vRNA reads.

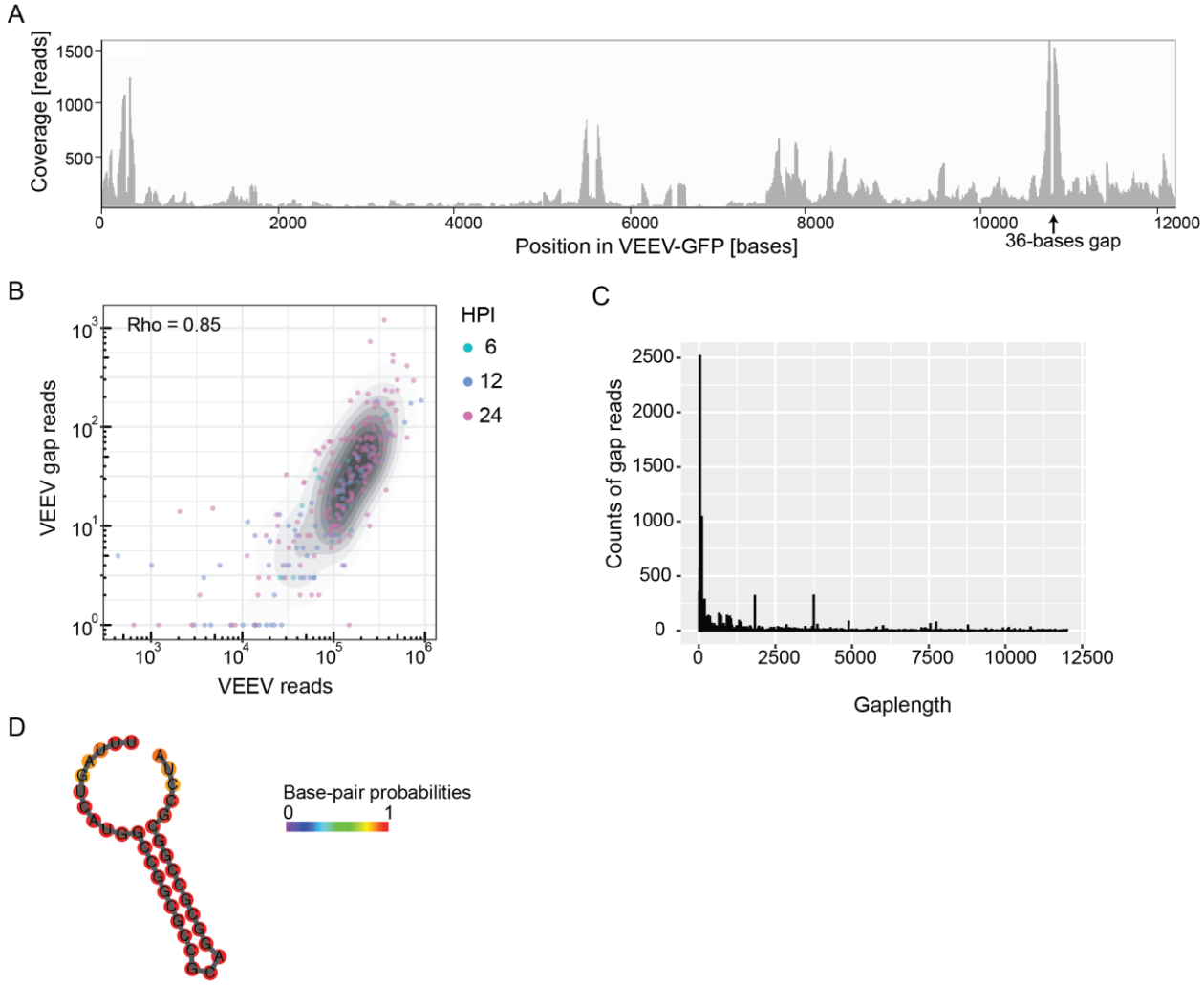
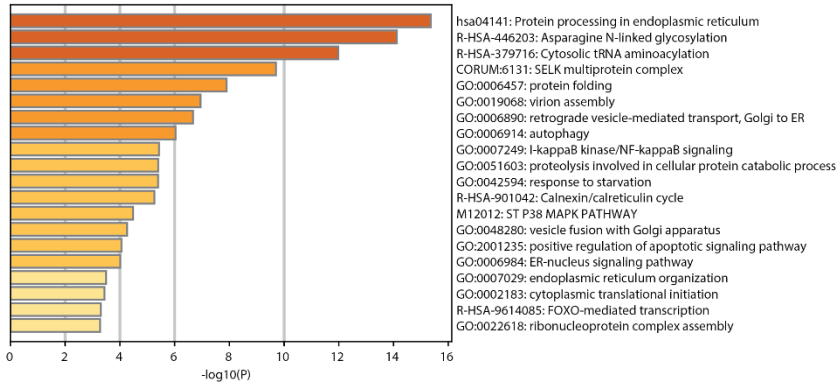


Figure S4. VEEV gap reads identified via viscRNA-Seq. (A) Coverage of VEEV gap reads over the VEEV-TC-83-GFP genome. (B) Scatter plot of number of VEEV gap reads and VEEV total reads within cells with detected gap reads. Each dot represents a cell and colored by hpi. (C) Histogram of gap lengths indicating that the majority of gaps are shorter than 1000 nucleotides. (D) RNA structural prediction of the most common 36-base gap (arrow in A) via RNAfold web server. Scale bar indicates the possibilities of base pairing.

A



B

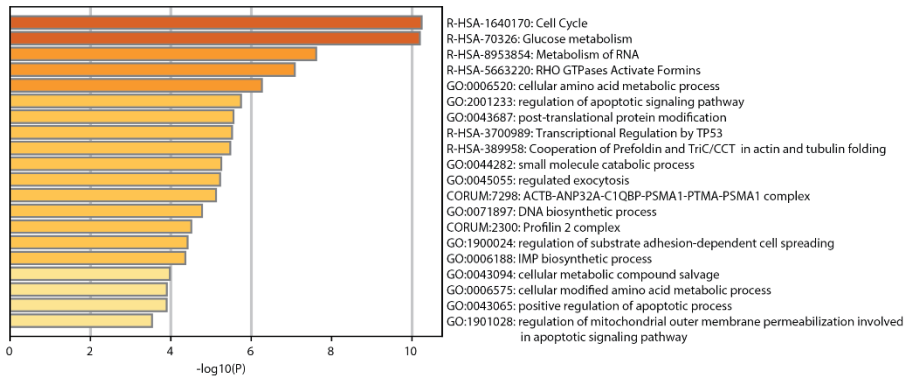
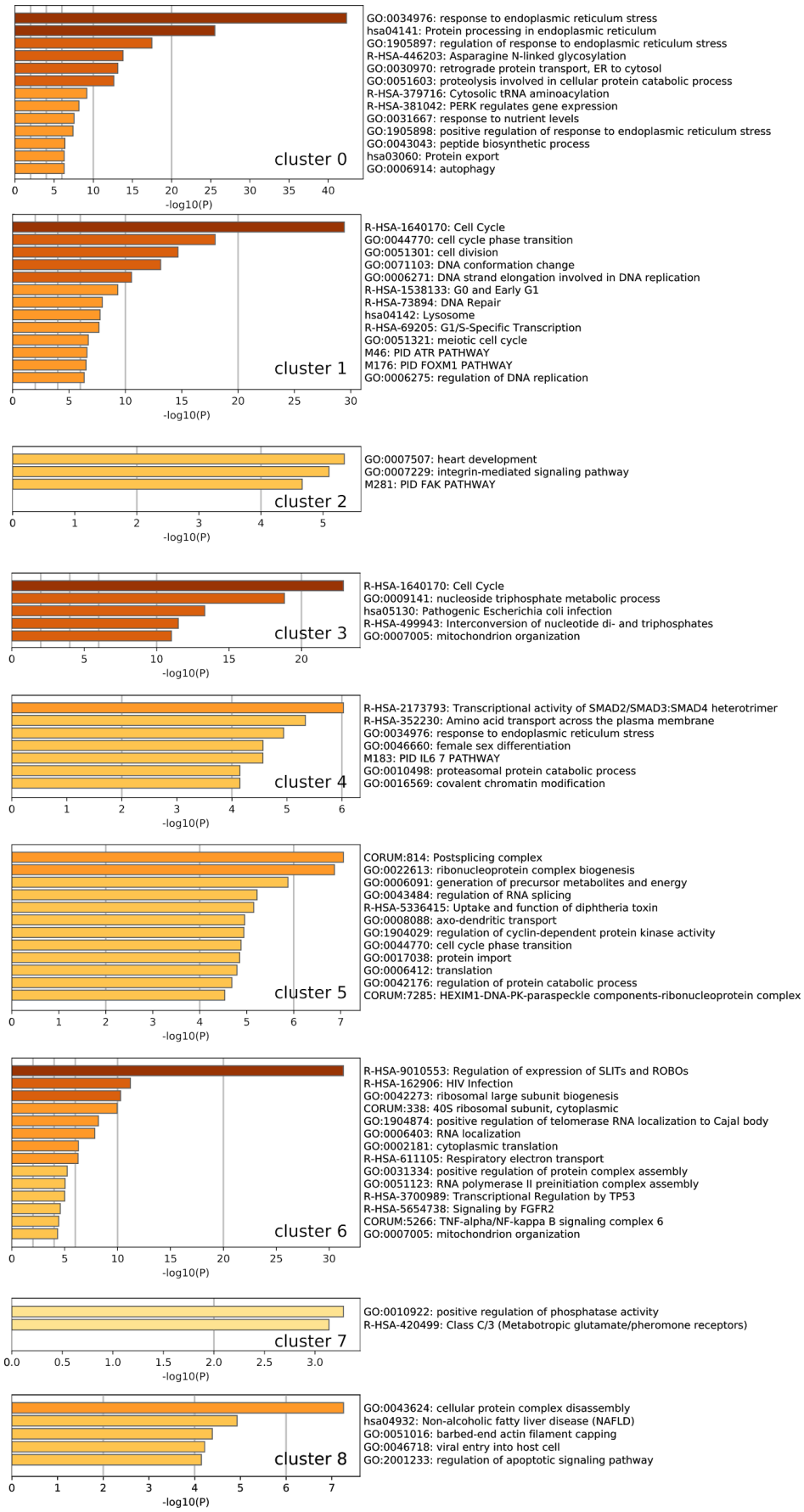


Figure S5. Pathway analysis for genes that positively correlated with VEEV 3'/5' read ratio and positively (A) or negatively (B) correlated with DENV and ZIKV. Each bar represents a group of genes according to Gene Ontology, KEGG, or other databases of biological function. The plot was made using metascape (Zhou et al. 2019).


cluster 0
GO:0034976: response to endoplasmic reticulum stress
hsa04141: Protein processing in endoplasmic reticulum
GO:1905897: regulation of response to endoplasmic reticulum stress
R-HSA-446203: Asparagine N-linked glycosylation
GO:0030970: retrograde protein transport, ER to cytosol
GO:0051603: proteolysis involved in cellular protein catabolic process
R-HSA-379716: Cytosolic tRNA aminoacylation
R-HSA-381042: PERK regulates gene expression
GO:0031667: response to nutrient levels
GO:1905898: positive regulation of response to endoplasmic reticulum stress
GO:0043043: peptide biosynthetic process
hsa03060: Protein export
GO:0006914: autophagy
-log10(P)
cluster 1
R-HSA-1640170: Cell Cycle
GO:0044770: cell cycle phase transition
GO:0051301: cell division
GO:0071103: DNA conformation change
GO:0006271: DNA strand elongation involved in DNA replication
R-HSA-1538133: G0 and Early G1
R-HSA-73894: DNA Repair
hsa04142: Lysosome
R-HSA-69205: G1/S-Specific Transcription
GO:0051321: meiotic cell cycle
M46: PID ATR PATHWAY
M176: PID FOXM1 PATHWAY
GO:0006275: regulation of DNA replication
-log10(P)
cluster 2
GO:0007507: heart development
GO:0007229: integrin-mediated signaling pathway
M281: PID FAK PATHWAY
-log10(P)
cluster 3
R-HSA-1640170: Cell Cycle
GO:0009141: nucleoside triphosphate metabolic process
hsa05130: Pathogenic Escherichia coli infection
R-HSA-499943: Interconversion of nucleotide di- and triphosphates
GO:0007005: mitochondrion organization
-log10(P)
cluster 4
R-HSA-2173793: Transcriptional activity of SMAD2/SMAD3:SMAD4 heterotrimer
R-HSA-352230: Amino acid transport across the plasma membrane
GO:0034976: response to endoplasmic reticulum stress
GO:0046660: female sex differentiation
M183: PID IL6 7 PATHWAY
GO:0010498: proteasomal protein catabolic process
GO:0016569: covalent chromatin modification
-log10(P)
cluster 5
CORUM:814: Postsplicing complex
GO:0022613: ribonucleoprotein complex biogenesis
GO:0006091: generation of precursor metabolites and energy
GO:0043484: regulation of RNA splicing
R-HSA-5336415: Uptake and function of diphtheria toxin
GO:0008088: axo-dendritic transport
GO:1904029: regulation of cyclin-dependent protein kinase activity
GO:0044770: cell cycle phase transition
GO:0017038: protein import
GO:0006412: translation
GO:0042176: regulation of protein catabolic process
CORUM:7285: HEXIM1-DNA-PK-paraspeckle components-ribonucleoprotein complex
-log10(P)
cluster 6
R-HSA-9010553: Regulation of expression of SLITs and ROBOs
R-HSA-162906: HIV Infection
GO:0042273: ribosomal large subunit biogenesis
CORUM:338: 40S ribosomal subunit, cytoplasmic
GO:1904874: positive regulation of telomerase RNA localization to Cajal body
GO:0006403: RNA localization
GO:0002181: cytoplasmic translation
R-HSA-611105: Respiratory electron transport
GO:0031334: positive regulation of protein complex assembly
GO:0051123: RNA polymerase II preinitiation complex assembly
R-HSA-3700989: Transcriptional Regulation by TP53
R-HSA-5654738: Signaling by FGFR2
CORUM:5266: TNF-alpha/NF-kappa B signaling complex 6
GO:0007005: mitochondrion organization
-log10(P)
cluster 7
GO:0010922: positive regulation of phosphatase activity
R-HSA-420499: Class C/3 (Metabotropic glutamate/pheromone receptors)
-log10(P)
cluster 8
GO:0043624: cellular protein complex disassembly
hsa04932: Non-alcoholic fatty liver disease (NAFLD)
GO:0051016: barbed-end actin filament capping
GO:0046718: viral entry into host cell
GO:2001233: regulation of apoptotic signaling pathway
-log10(P)

1212

Figure S6. Pathway analysis for all gene clusters in Figures 5G and H. Each bar represents a group of genes according to Gene Ontology, KEGG, or other databases of biological function. The plot was made using metascape (Zhou et al. 2019). In cluster 6, the most enriched pathway (Regulation of expression of SLITs and ROBOs) is related to the astrocyte nature of U87 cells used for VEEV infection, as this cell lines express these genes at higher levels than Huh7 cells used for flavivirus infection (around 6 times higher) as well as the cells used for influenza virus infection (around 100 times higher).