# Title: Determinants of Zika Virus Host Tropism Uncovered by Deep Mutational Scanning

**Authors:** Yin Xiang Setoh[1,†,*], Alberto A. Amarilla[1,†], Nias Y.G. Peng[1], Rebecca E. Griffiths[2], Julio Carrera[3], Morgan E. Freney[1], Eri Nakayama[5], Shinya Ogawa[6], Daniel Watterson[1], Naphak Modhiran[1], Faith Elizabeth Nanyonga[1], Francisco J. Torres[1], Andrii Slonchak[1], Parthiban Periasamy[1], Natalie A. Prow[4], Bing Tang[4], Jessica Harrison[1], Jody Hobson-Peters[1], Thom Cuddihy[7], Justin Cooper-White[2], Roy A. Hall[1], Paul R. Young[1], Jason M Mackenzie[3], Ernst Wolvetang[2], Jesse D. Bloom[8,9], Andreas Suhrbier[4,#], Alexander A. Khromykh[1,#,*]

*Correspondence to: Alexander A. Khromykh alexander.khromykh@uq.edu.au or

Yin Xiang Setoh y.setoh@uq.edu.au.

## Supplementary Information

Supplementary Notes
Supplementary Tables 1 – 5
Supplementary Figures 1 – 9
Supplementary Data 1
Supplementary Data 2

## Supplementary Notes

**Flavivirus genome organization and virion assembly and maturation.** Flavivirus genome encodes a single open reading frame flanked by 5' and 3' untranslated regions. Flavivirus polyproteins are processed by host and viral proteases into three structural proteins; capsid (C), pre-membrane (prM), envelope (E), and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, NS5). Flavivirus virions are composed of structural proteins, host lipid membrane bilayer and viral RNA [1]. The E protein is the major viral protein interacting with cell receptors and consists of three ectodomains (E-DI, E-DII, and E-DIII) and two trans-membrane domains (E-TM1 and E-TM2). The E protein is N-glycosylated in most ZIKV strains at position 154 located in E-DI. The prM protein is a small glycoprotein that forms heterodimers with E to produce immature viral particles. Flavivirus virions assemble and bud into the lumen of the endoplasmic reticulum (ER) as non-infectious immature particles, with prM-E arranged as heterotrimeric spikes. Upon transit of the immature virus particles through the Golgi network, the low pH environment induces the 'tightening' of the unfolded portion of pr, which brings down the folded pr head which interacts with the domain II of E, forming a flattened prME heterodimeric structure [2]. This conformation exposes the furin cleavage site on prM and allows the cleavage of prM to pr and M [3]. Despite being cleaved, pr remains associated with E in the trans-Golgi network through proposed interactions between pr and E [2,4,5], and this prevents premature fusion within the cell while the virions are exiting. Upon release into the neutral pH extracellular environment, pr dissociates from the E and the mature infectious particles consisting of M and E proteins are then formed. For some flaviviruses (e.g. dengue) and under certain conditions (e.g. virus propagation in

mosquito cells) prM-E cleavage by furin is inefficient which results in the release of significant proportion of immature or partially mature particles containing prM and E proteins (reviewed in [6]).

**A S66L mutation arises in the 316Q substituted virus.** Although the 316Q virus initially replicated slower than WT virus in mammalian cells, the growth of this virus improved in the course of infection and ultimately reached titers similar to WT virus (Fig 2b, c, d). The 316Q virus also accumulated to similar, and even higher, levels than WT virus in IFNAR$^{-/-}$ MEFs later in the course of infection (Supplementary Figure 2c). Similarly, 316Q virus accumulated to significantly higher titers than WT virus later during infection of C6/36 and Aag2 mosquito cells (Fig 2e and Supplementary Figure 2f). Deep sequencing of the 316Q virus collected from Vero cells at 5 days post infection detected a mutation, S66L. The S66L mutation was present in 73% of this virus population, while the 316Q substitution was retained in 100% of this virus population. Deep sequencing of the original 316Q stock virus (generated in C6/36 cells) showed that the S66L mutation was already present in 19% of this population. This suggested that the 316Q substitution has a detrimental effect on virus replication, with the S66L mutation partially rescuing this replication defect. To test this hypothesis, the 316Q virus (generated in C6/36 cells with 19% of the population containing the S66L mutation) was used to infect C6/36 cells at MOI=0.01. On day 8 post infection the virus in the supernatant was deep sequenced, with the S66L mutation now present in 50% of the virus population, while the 316Q substitution was again retained in 100% of the virus population. The result confirmed that that S66L mutation is able to partially compensate for the defect caused by the K316Q substitution.

**The 316 and 461 residues in arboviruses and insect specific viruses of the Flavivirus genus.** An amino acid alignment of selected arboviruses in the genus Flavivirus showed that 316K and 461S residues are widely conserved (Supplementary Table 2), although West Nile virus has 316G with a conserved 461S residue, and Yellow fever virus has a conserved 316K residue with a 461N residue (Supplementary Table 2). None of these arboviruses have the 316Q or 461G substitution (Supplementary Table 2).

When the E protein amino acid sequence of WT ZIKV was aligned to lineage I insect-specific flaviviruses (ISF), alignment scores were poor (13-14%). Nevertheless, the alignment suggested the residue corresponding to 316K in lineage I ISFs was often Q, with the residue corresponding to 461S also sometimes S (Table S2). The lineage I ISFs are thought to be the ancestors for all flaviviruses [7].

Lineage II ISFs are thought to have encountered a vertebrate host at some stage of their evolution, and later reverted back to being ISFs [7]. Alignment of WT ZIKV E protein with lineage II ISFs showed much higher alignment scores (41-44%) and most Lineage II ISFs retain K in the position corresponding with 316K (Supplementary Table 2). Three Lineage II ISFs have a G in the position corresponding to 461S, whereas the rest have S (Supplementary Table 2).

## References


1 Roby, J. A., Funk, A. & Khromykh, A. A. in *Molecular virology and control of flaviviruses* (ed Pei-Yong Shi) Ch. 3, 21-49 (Caister Academic Press, 2012).

2 Zhang, X. *et al.* Cryo-EM structure of the mature dengue virus at 3.5-A resolution. *Nat Struct Mol Biol* **20**, 105-110, doi:10.1038/nsmb.2463 (2013).

3 Roby, J. A., Setoh, Y. X., Hall, R. A. & Khromykh, A. A. Post-translational regulation and modifications of flavivirus structural proteins. *J Gen Virol* **96**, 1551-1569, doi:10.1099/vir.0.000097 (2015).

4 Yu, I. M. *et al.* Association of the pr peptides with dengue virus at acidic pH blocks membrane fusion. *J Virol* **83**, 12101-12107, doi:10.1128/JVI.01637-09 (2009).

5 Yu, I. M. *et al.* Structure of the immature dengue virus at low pH primes proteolytic maturation. *Science* **319**, 1834-1837, doi:10.1126/science.1153264 (2008).

6 Pierson, T. C. & Diamond, M. S. Degrees of maturity: the complex structure and biology of flaviviruses. *Curr Opin Virol* **2**, 168-175, doi:10.1016/j.coviro.2012.02.011 (2012).

7 Blitvich, B. J. & Firth, A. E. Insect-specific flaviviruses: a systematic review of their discovery, host range, mode of transmission, superinfection exclusion potential and genomic organization. *Viruses* **7**, 1927-1959, doi:10.3390/v7041927 (2015).

**Supplementary Table 1.** DMS mutagenesis primers.

120

Provided as separate Supplementary Table 1.xlsx

**Supplementary Table 2.** Amino acid sequence alignments for different flaviviruses for the sequences adjacent to 316K and 461S residues in ZIKV E protein.

125

| Vertebrate-infecting flaviviruses* | Residue 316 | Residue 461 |
|---|---|---|
| Zika virus (316Q/461G) | $^{312}$FTFT **Q** IPAE$^{320}$ | $^{457}$FGGM **G** WFSQ$^{465}$ |
| Zika virus (WT) | $^{312}$FTFT **K** IPAE$^{320\#}$ | $^{457}$FGGM **S** WFSQ$^{465\#}$ |
| Murray Valley encephalitis virus | $^{309}$FTFS **K** NPAD$^{317}$ | $^{454}$FGGM **S** WISP$^{462}$ |
| Japanese encephalitis virus | $^{308}$FSFA **K** NPAD$^{316}$ | $^{453}$FGGM **S** WITQ$^{461}$ |
| Saint Louis encephalitis virus | $^{309}$FTFS **K** NPAD$^{317}$ | $^{454}$FGGM **S** WITQ$^{462}$ |
| Rocio virus | $^{309}$FAFA **K** NPVD$^{317}$ | $^{454}$FGGM **S** WVTQ$^{462}$ |
| Dengue virus 1 | $^{306}$FKLE **K** EVAE$^{314}$ | $^{448}$FSGV **S** WTMK$^{456}$ |
| Dengue virus 2 | $^{306}$FKIV **K** EIAE$^{314}$ | $^{448}$FSGV **S** WTMK$^{456}$ |
| Dengue virus 3 | $^{304}$FVLK **K** EVSE$^{312}$ | $^{446}$FSGV **S** WVMK$^{454}$ |
| Dengue virus 4 | $^{306}$FSID **K** EMAE$^{314}$ | $^{448}$FGGV **S** WMVR$^{456}$ |
| West Nile virus (NY99) | $^{309}$FKFL **G** TPAD$^{317}$ | $^{454}$FGGM **S** WITQ$^{462}$ |
| West Nile virus (NSW2011) | $^{309}$FRFL **G** TPAD$^{317}$ | $^{454}$FGGM **S** WITQ$^{462}$ |
| Yellow fever virus (17D) | $^{304}$MFFV **K** NPTD$^{312}$ | $^{446}$FGGL **N** WITK$^{454}$ |
| **Insect-specific flaviviruses (Lineage I)**** | | |
| Cell fusing agent virus | $^{279}$-T-FS **Q** -DG-$^{284}$ | $^{392}$VG-- **-** -AFF-SN$^{398}$ |
| Culex theileri flavivirus | $^{276}$-M-FV **Q** -DG-$^{281}$ | $^{389}$FG-- **-** -ALF-SN$^{395}$ |
| Quang Binh virus | $^{276}$-V-FV **Q** -SG-$^{281}$ | $^{389}$LG-- **-** -NLI-TN$^{395}$ |
| Culex flavivirus | $^{276}$-V-FV **Q** -SG-$^{281}$ | $^{389}$L--- **-** FSLL-SS$^{395}$ |
| Nienokoue virus | $^{276}$-V-FV **Q** -NG-$^{281}$ | $^{389}$M--- **-** WKLM-M-$^{394}$ |
| Nakiwogo virus | $^{275}$-AYFT **Q** -SG-$^{281}$ | $^{387}$FN-- **P** WKLL-ST$^{395}$ |
| Palm Creek virus | $^{276}$-T-FT **Q** -SG-$^{281}$ | $^{387}$LN-- **P** WSFF-ST$^{395}$ |
| Hanko virus | $^{276}$-V-FS **Q** -SP-$^{281}$ | $^{393}$WD-- **S** WKVI---$^{399}$ |
| Parramatta River virus | $^{277}$---FI **Q** -SS-$^{281}$ | $^{393}$WD-- **S** WKII---$^{399}$ |
| Aedes flavivirus | $^{273}$GV-FV **Q** -NG-$^{279}$ | $^{389}$W--- **S** W--IPG-$^{394}$ |
| Kamiti River virus | $^{273}$GT-FI **Q** -NG-$^{279}$ | $^{389}$W--- **S** W--FPS-$^{394}$ |
| Karumba virus | $^{277}$---FM **-** -AGP$^{281}$ | $^{393}$---- **-** -RL--E-$^{395}$ |
| Mac Peak virus | $^{277}$---FT **-** -PGP$^{281}$ | $^{393}$---- **-** -RI--E-$^{395}$ |
| **Insect-specific flaviviruses (Lineage II)***** | | |
| Binjari virus | $^{306}$FTTE **T** RPAD$^{314}$ | $^{448}$FGGM **G** WLTK$^{456}$ |
| Chaoyang virus | $^{308}$FTFA **K** RPVD$^{316}$ | $^{450}$FGGI **G** WMAK$^{458}$ |
| Nounane virus | $^{316}$VVLT **K** EPVD$^{324}$ | $^{458}$FGGL **S** WISQ$^{466}$ |
| Nhumirim virus | $^{314}$YSFV **K** VPSD$^{322}$ | $^{456}$FGGM **S** WVTK$^{464}$ |
| Lammi virus | $^{308}$FTFS **K** RPVD$^{316}$ | $^{450}$FGGI **S** WIAK$^{458}$ |
| Ilomantsi virus | $^{308}$FTFE **K** RPVA$^{316}$ | $^{450}$FGGM **S** WISK$^{458}$ |
| Hidden Valley virus | $^{306}$FNME **T** RPAD$^{314}$ | $^{448}$FGGM **G** WLTK$^{460}$ |
| Long Pine Key virus | $^{308}$FTFV **K** RPTE$^{316}$ | $^{450}$FGGM **S** WITK$^{458}$ |
| Donggang virus | $^{310}$FSIL **K** RPTA$^{318}$ | $^{452}$FGGM **S** WISK$^{460}$ |
| Marisma mosquito virus | $^{311}$FTLL **K** RPTA$^{319}$ | $^{453}$FGGM **S** WISK$^{461}$ |
| Barkedji virus | $^{314}$FSFS **K** QPSD$^{322}$ | $^{456}$FGGM **S** WLTK$^{464}$ |
| Nanay virus | $^{307}$FTFE **K** IPTD$^{315}$ | $^{448}$FGGV **G** IMVR$^{456}$ |
| Kampung Karu virus | $^{307}$FTFS **R** APAD$^{315}$ | $^{448}$FGGL **S** WMTR$^{456}$ |

Alignment scores for E protein alignments (verses ZIKV WT), were calculated using ClustalW on default settings (https://www.genome.jp/tools-bin/clustalw) – 39%-55% for the flaviviruses; 13-14% for the Insect-specific flaviviruses (Lineage I); and 41-

44% for the Insect-specific flaviviruses (Lineage II). #Amino acid positions of the first and the last residues in the aligned regions of E protein of respective viruses are indicated as superscript numbers. Dashes (-) indicate deletions/missing amino acids. Alignments were performed using CLC Main Workbench 6.9.2. *Vertebrate infecting viruses were aligned using the default parameters – open gap cost = 10, end gap cost = 1. **Lineage I ISFs were aligned using the parameters – open gap cost = 1, end gap cost = 1. ***Lineage II ISFs were aligned using the default parameters – open gap cost = 10, end gap cost = 1.

**Supplementary Table 3.** Interaction analysis data by MOE for Fig. 4b

| Type | ChainA | SetA | ChainB | SetB | Dist (Å) | BB | Freq |
|---|---|---|---|---|---|---|---|
| **d** | 5IZ7.C | Ser461 | 5IZ7.F | Ser8 | 3.869 | ** | 4 |
| **d** | 5IZ7.C | Ser461 | 5IZ7.F | Lys11 | 4.077 | *- | 6 |
| **d** | 5IZ7.C | Ser461 | 5IZ7.F | Glu24 | 3.791 | -- | 4 |
| **d** | 5IZ7.C | Ser461 | 5IZ7.F | His28 | 4.254 | -- | 2 |
| **d** | 5IZ7.C | Ser461 | 5IZ7.D | Tyr74 | 3.879 | -b | 2 |
| **d** | 5IZ7.C | Ser461 | 5IZ7.D | Ser75 | 3.533 | *b | 13 |

Data generated using "Protein Contacts" algorithm of MOE. The “Force” field was set to Amber10-EHT, “Aggregate” field was set to residue, and all other settings were kept at default. **Type:** Contact interaction type. d:distance, h:hydrogen bond. **ChainA:** Chain for SetA residue 461 in E. **SetA:** Residue 461 in E set as one side of the contact. **ChainB:** Chain for SetB residue in M. **SetB:** Residue identified as the other side of the contact in M. **Dist:** Distance (in Å) between interacting residues. **BB:** Indicates whether a backbone atom is involved in the interaction. First character is SetA and second character is SetB, “b” indicates all atoms are backbone, “-” non-backbone, “*”both backbone and non-backbone atoms are represented. **Freq:** Number of interactions between atoms of SetA and SetB residues. The chains used are **5IZ7.C:** E protein (colored green, in Fig. 4b,c). **5IZ7.F:** M protein (Chain F, in Fig. 4b,c). **5IZ7.D:** The second M protein (Chain D, in Fig. 4b,c) of ME dimer.

**Supplementary Table 4.** Interaction analysis data by MOE for Fig. 4c.

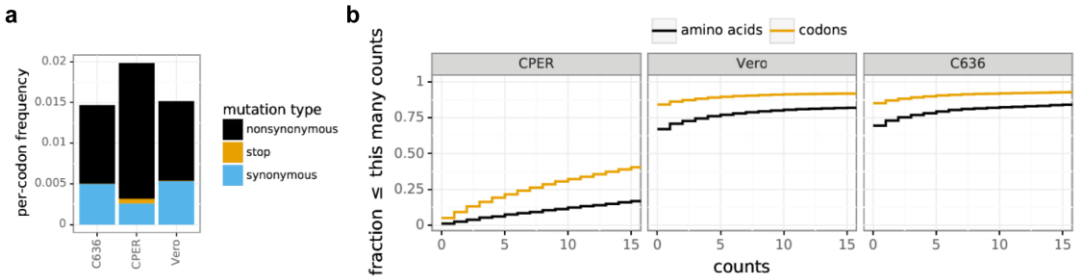
| Type | ChainA | SetA | ChainB | SetB | Dist (Å) | BB | Freq |
|---|---|---|---|---|---|---|---|
| **d** | 5IZ7.C | Gly461 | 5IZ7.F | Ser8 | 4.128 | bb | 2 |
| **d** | 5IZ7.C | Gly461 | 5IZ7.F | Lys11 | 3.854 | b- | 5 |
| **dh** | 5IZ7.C | Gly461 | 5IZ7.D | Ser75 | 3.673 | bb | 7 |

Data generated using "Protein Contacts" algorithm of MOE. The “Force” field was set to Amber10-EHT, “Aggregate” field was set to residue, and all other settings were kept at default. **Type:** Contact interaction type. d:distance, h:hydrogen bond. **ChainA:** Chain for SetA residue 461 in E. **SetA:** Residue 461 in E set as one side of the contact. **ChainB:** Chain for SetB residue in M. **SetB:** Residue identified as the other side of the contact in M. **Dist:** Distance (in Å) between interacting residues. **BB:** Indicates whether a backbone atom is involved in the interaction. First character is SetA and second character is SetB, “b” indicates all atoms are backbone, “-” non-backbone, “*”both backbone and non-backbone atoms are represented. **Freq:** Number of interactions between atoms of SetA and SetB residues. The chains used are **5IZ7.C:** E protein (colored green, in Fig. 4b,c). **5IZ7.F:** M protein (Chain F, in Fig. 4b,c). **5IZ7.D:** The second M protein (Chain D, in Fig. 4b,c) of ME dimer.

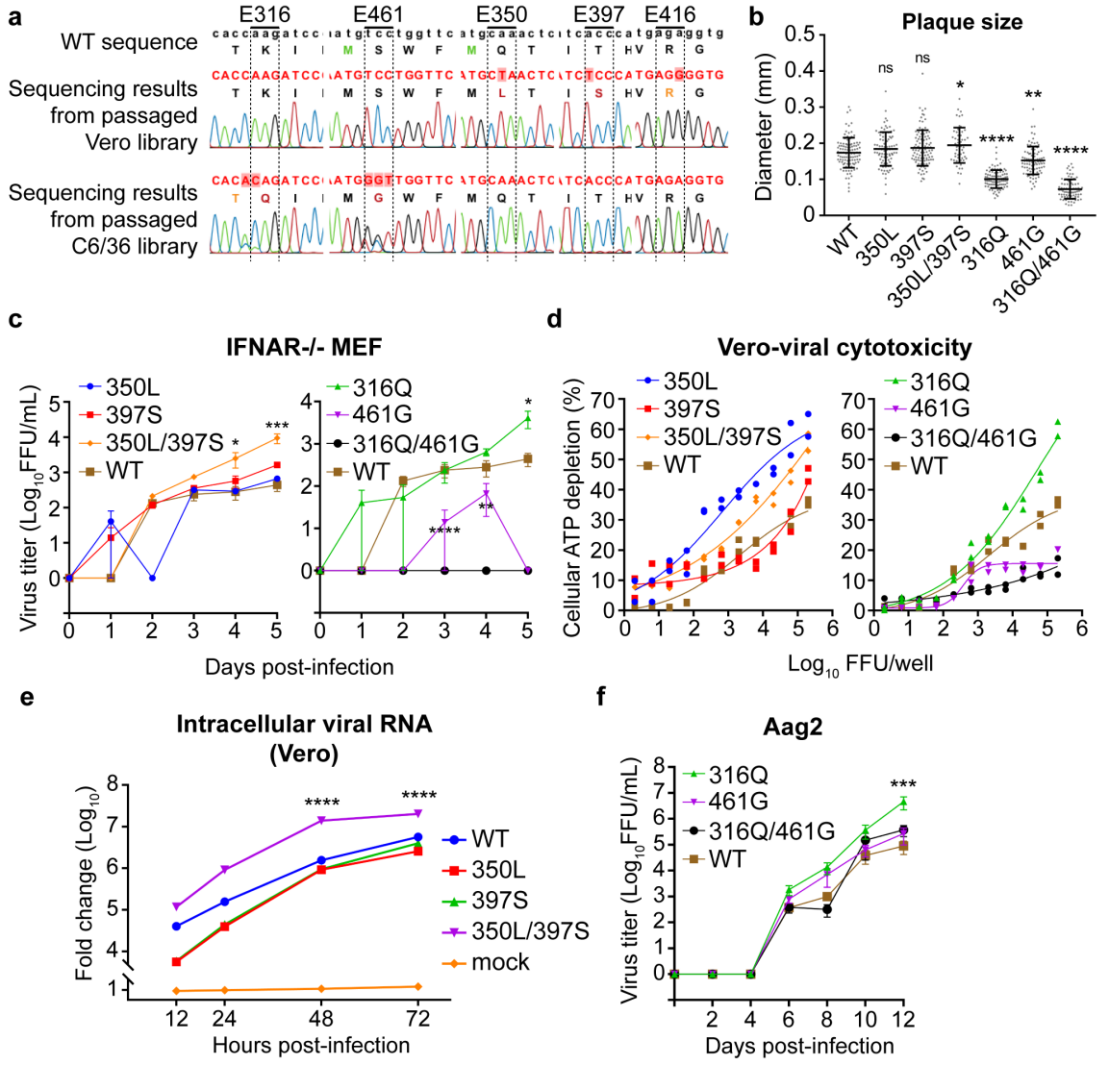
**Supplementary Table 5.** Free energy calculations of interaction between E and M determined by MOE.

| Mutation | ΔG (kcal/mol) |
|---|---|
| **S461A** | 1.6 |
| **S461R** | -5.18 |
| **S461N** | -3.33 |
| **S461D** | -2.48 |
| **S461C** | 0.94 |
| **S461Q** | -2.79 |
| **S461E** | -3.67 |
| **S461G** | 3.59 |
| **S461H** | -2.33 |
| **S461I** | 1.77 |
| **S461L** | -1.53 |
| **S461K** | -0.21 |
| **S461M** | -4.89 |
| **S461F** | 1.58 |
| **S461P** | 2.31 |
| **S461S** | 0 |
| **S461T** | -0.4 |
| **S461W** | 3.41 |
| **S461Y** | -1.07 |
| **S461V** | 1.55 |

The S461G substitution provides the biggest change in free energy and thus the lowest interaction strength between E and M (highlighted in green).

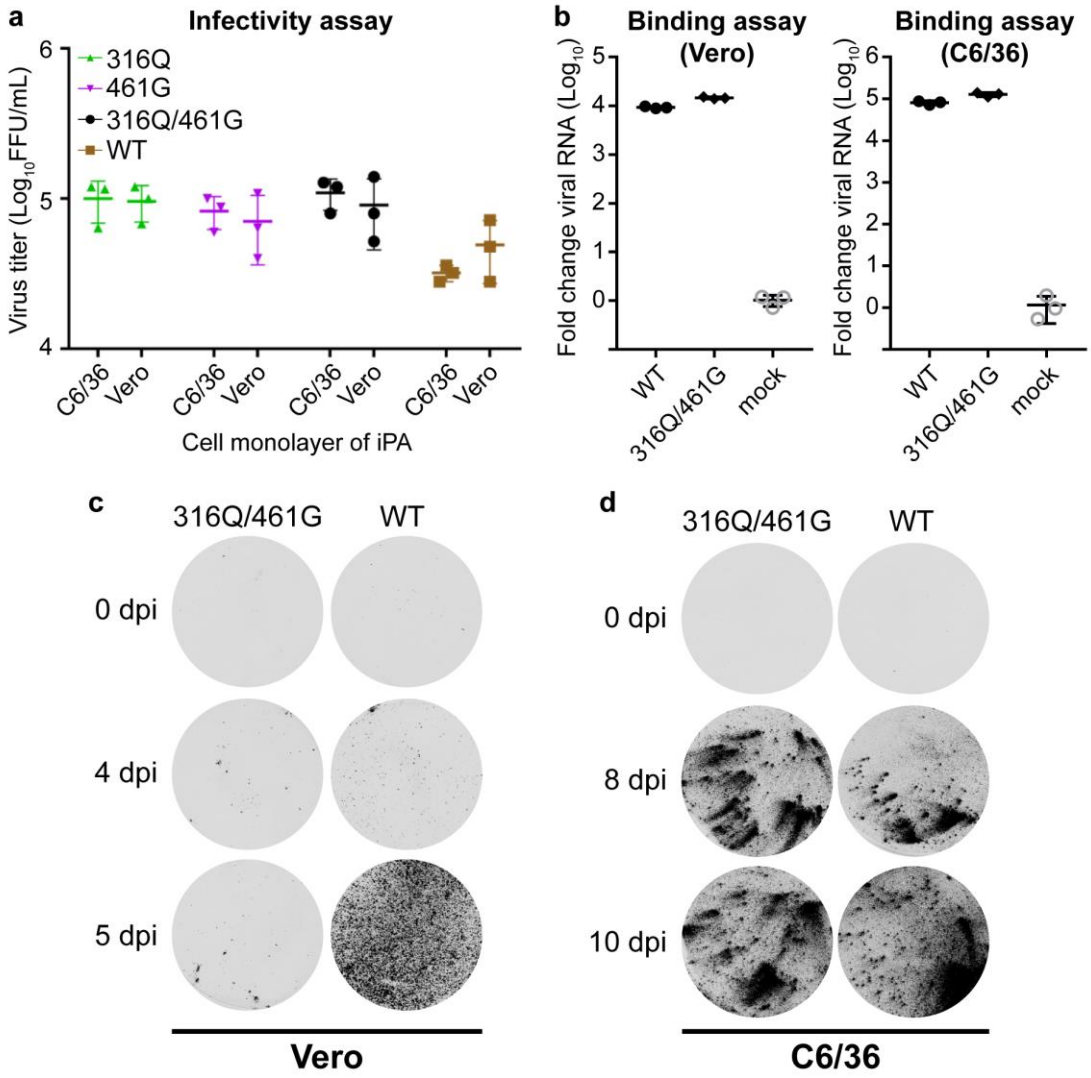



**Supplementary Figure 1.** Codon mutagenesis sequencing data of CPER DNA in E-DIII stem-anchor. **a)** Mutation types. Nonsynonymous mutations are predominant in CPER cDNA amplicon library and in both viral libraries; all stop codons were purged in viral libraries, as expected. **b)** cumulative counts for codons and amino acid variants. The y-axis ("Fraction ≤ this many counts") refers to the percentage of mutations that are found less than or equals to the indicated number of times ("counts"). For example, in the CPER DNA, ~20% of codon mutants are found less than or equals to 5 times at that site in all sequenced molecules, and at the same site, ~6% of amino acid mutants are found less than or equals to 5 times at that site in all sequenced molecules. Overall, viral libraries contained lower representation of variants compared to CPER cDNA amplicon library, as anticipated from the selection of smaller numbers of viable viruses.



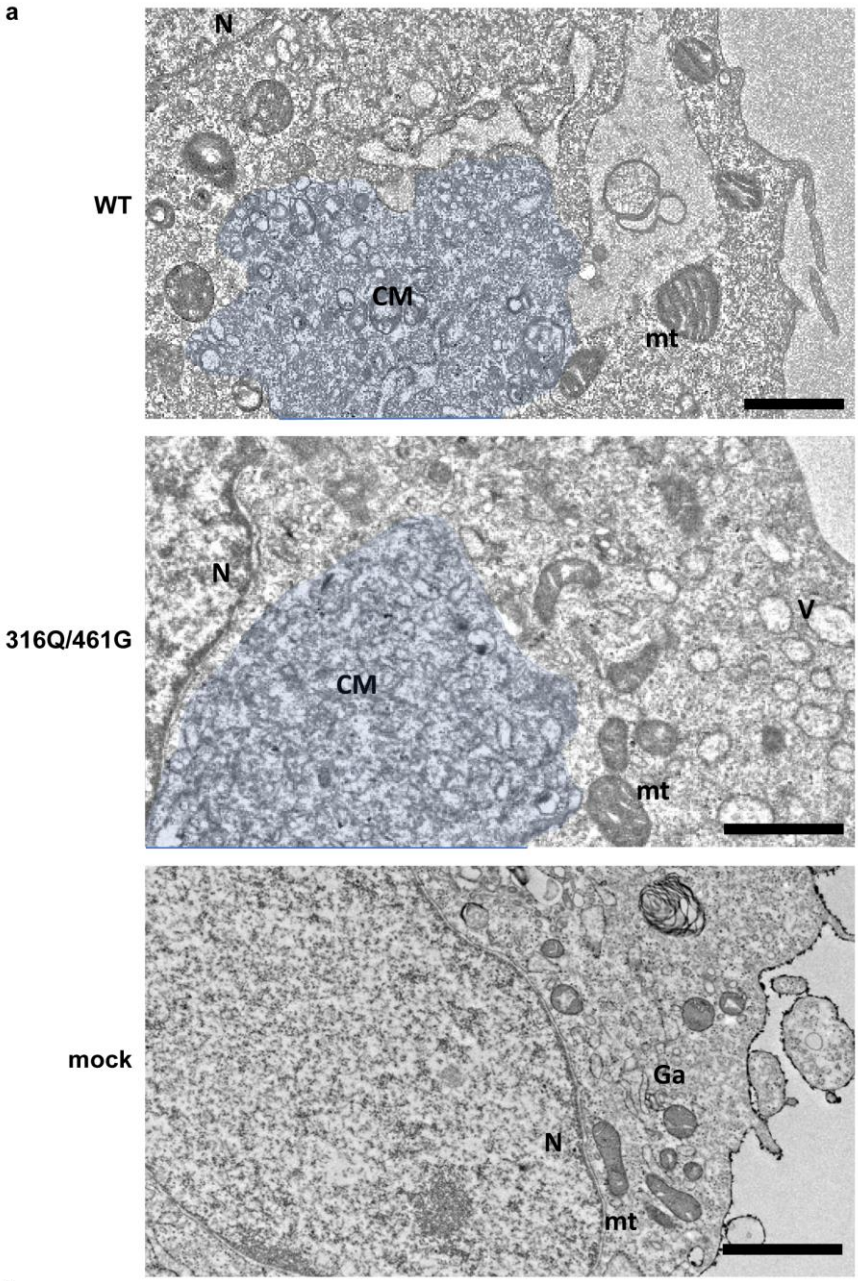

**Supplementary Figure 2.** Characterization of Vero- and C6/36-derived virus mutants. **a)** Sequencing of C6/36- and Vero-derived virus libraries passaged once at a low multiplicity-of-infection (MOI=0.01) in C6/36 and Vero cells, respectively. **b)** Plaque size of the WT and respective mutant viruses was quantified by measurement of plaque diameter using ImageJ software. WT (n=106 individual plaques), 350L (n=87 individual plaques), 397S (n=99 individual plaques), 350L/397S (n=56 individual plaques), 316Q (n=90 individual plaques), 461G (n=92 individual plaques), 316Q/461G (n=77 individual plaques). Mean ± standard error of the mean. Statistical analysis performed by one-way ANOVA with multiple comparisons. * P = 0.0109, ** P = 0.0014, **** P = 0.0001 **c)** Growth kinetics of WT and mutant viruses in IFNAR$^{-}$

$^{/-}$ mouse embryonic fibroblasts (MEF) infected at a MOI=0.1. Culture supernatants were harvested at the indicated time points after infection and virus titres were determined by iPA in Vero cells. Mean ± standard error of the mean. Limit of detection for iPA is 1.6 $Log_{10}$FFU/mL. Three independent experiments were conducted and statistical analysis was performed by two-way ANOVA with Tukey's multiple comparisons test against WT. * P = 0.0103 (350L/397S@4dpi), *** P = 0.0004 (350L/397S@5dpi), **** P = <0.0001 (461G@3dpi), ** P = 0.0045 (461G@4dpi), * P = 0.0125 (316Q@5dpi). **d)** Virus-induced cell cytotoxicity was determined by infecting Vero cells with WT and mutant viruses and assaying for cellular ATP at 5 dpi. **e)** Vero cells were infected with WT and the respective mutant viruses at an MOI of 0.1, and intracellular replication of viral RNA was determined using qRT-PCR at 12, 24, 48 and 72 hours post infection. Three independent experiments were conducted and statistical analysis was performed by two-way ANOVA with Tukey's multiple comparisons test against WT. **** P = <0.0001. **f)** Growth kinetics of WT and mutant viruses in *Aedes aegypti* cells, Aag2, infected at MOI=0.1. Culture supernatants were harvested at the indicated time points after infection and virus titres were determined by iPA in Vero cells. Three independent experiments were conducted and statistical analysis was performed by two-way ANOVA with Tukey's multiple comparisons test against WT. *** P = 0.0008. Mean ± standard error of the mean. Limit of detection for iPA is 1.6 $Log_{10}$FFU/mL.

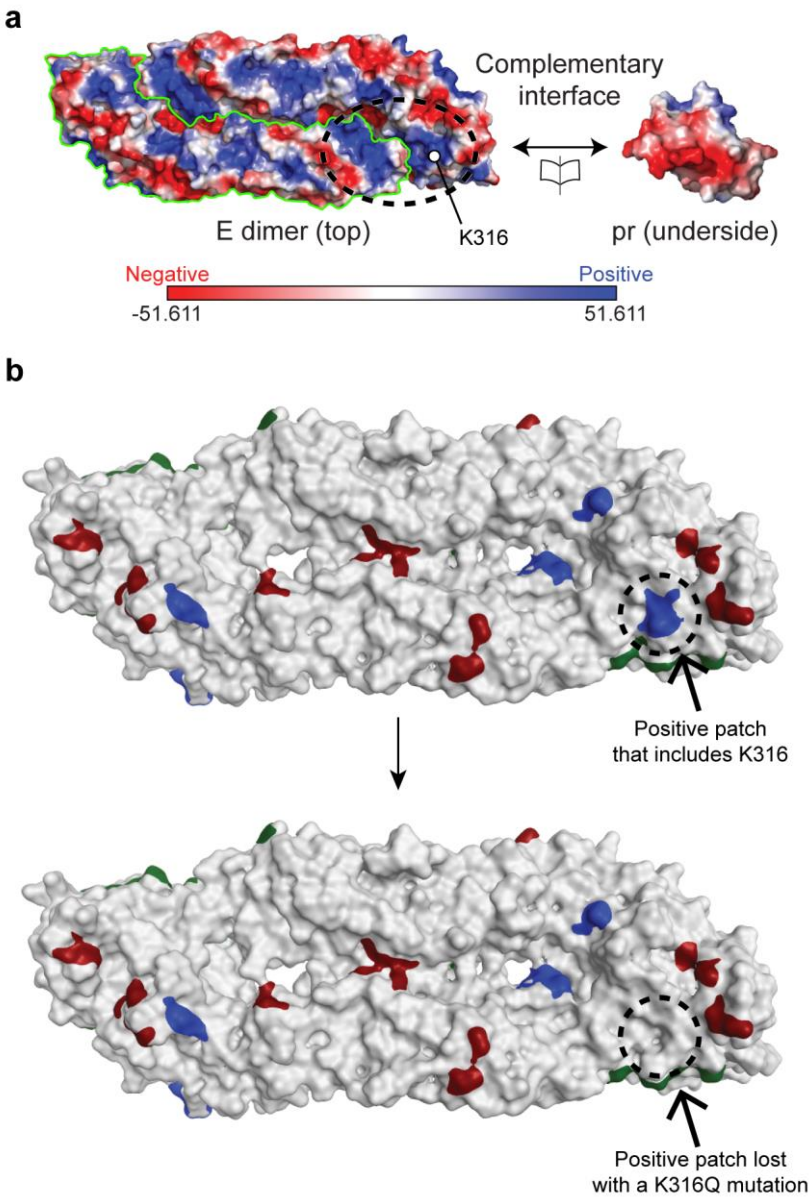



**Supplementary Figure 3.** The K316Q/S461G mutations do not affect viral binding and entry, but attenuate viral spread. **a)** Viral infectivity assay in C6/36 and Vero cells. Stocks of 316Q, 461G, 316Q/461G or WT viruses were normalized to $10^5$ FFU/mL (using titres determined in C6/36 cells) and then titered by iPA on C6/36 or Vero cell monolayers. Three independent experiments were conducted and statistical analysis performed by t-test comparing C6/36 and Vero titers for the same viral stock showed no significant differences. Mean ± standard error of the mean. **b)** Bindings assay. Vero or C6/36 cells were incubated for 2 hours with WT or 316Q/461G viruses, after which cells were washed 10 times with fresh media to remove unbound viruses. Cells were then harvested in TRI Reagent, total RNA was purified and viral

RNA was quantified by qRT-PCR. Y-axis shows fold change of viral RNA against TBP-1 (Vero) and RPL-11 (C6/36) house-keeping genes. Three independent experiments were conducted and statistical analysis performed by one way ANOVA with multiple comparisons of fold changes for viral RNA between WT and 316Q/461G mutant virus infections in C6/36 and Vero cells showed no significant differences. Mean ± standard error of the mean. Viral spread assay performed in **c)** Vero or **d)** C6/36 cells infected with 316Q/461G or WT viruses at MOI=0.0001 and incubated in liquid media to allow virus spread. At indicated time points after infection, cells were fixed and immuno-stained with 4G2 anti-E antibody to visualize viral spread. Representative images of n=2 biologically independent samples.

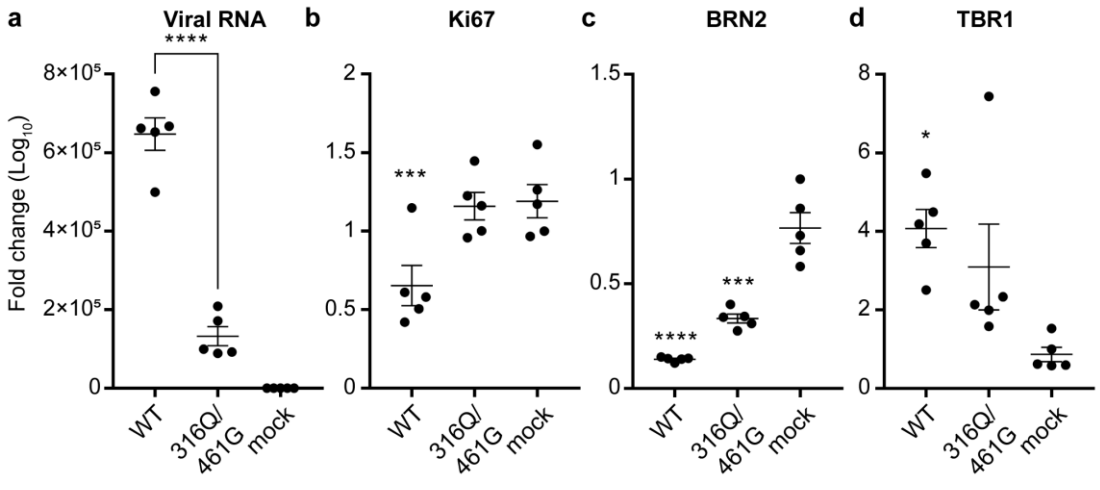

a
WT
N
CM
mt
316Q/461G
N
CM
V
mt
mock
Ga
N
mt
b
***
Number of particles / slide
80
25
20
15
10
5
0
WT
316Q/
461G

**Supplementary Figure 4.** Transmission electron microscopy of Vero cells infected with WT virus, 316Q/461G mutant virus, or uninfected (mock). **a)** WT virus- or 316Q/461G virus-infected Vero cells presents convoluted membranes (highlighted in blue) that are not observed in uninfected cells. CM – convoluted membranes (blue highlight), V – vesicles, mt – mitochondria, N – nucleus, Ga – golgi apparatus. Scale bars are 1 µm. Representative images of n=3 biologically independent samples. **b)** Virion quantification in Vero cells infected with WT or 316Q/461G mutant viruses. A total of n=30 individual slides per virus infection were imaged, and number of virions per slide were tabulated. Mean ± standard error of the mean. Statistical analysis was performed using the Kolmogorov-Smirnov test, two-sided, *** P = 0.0004.

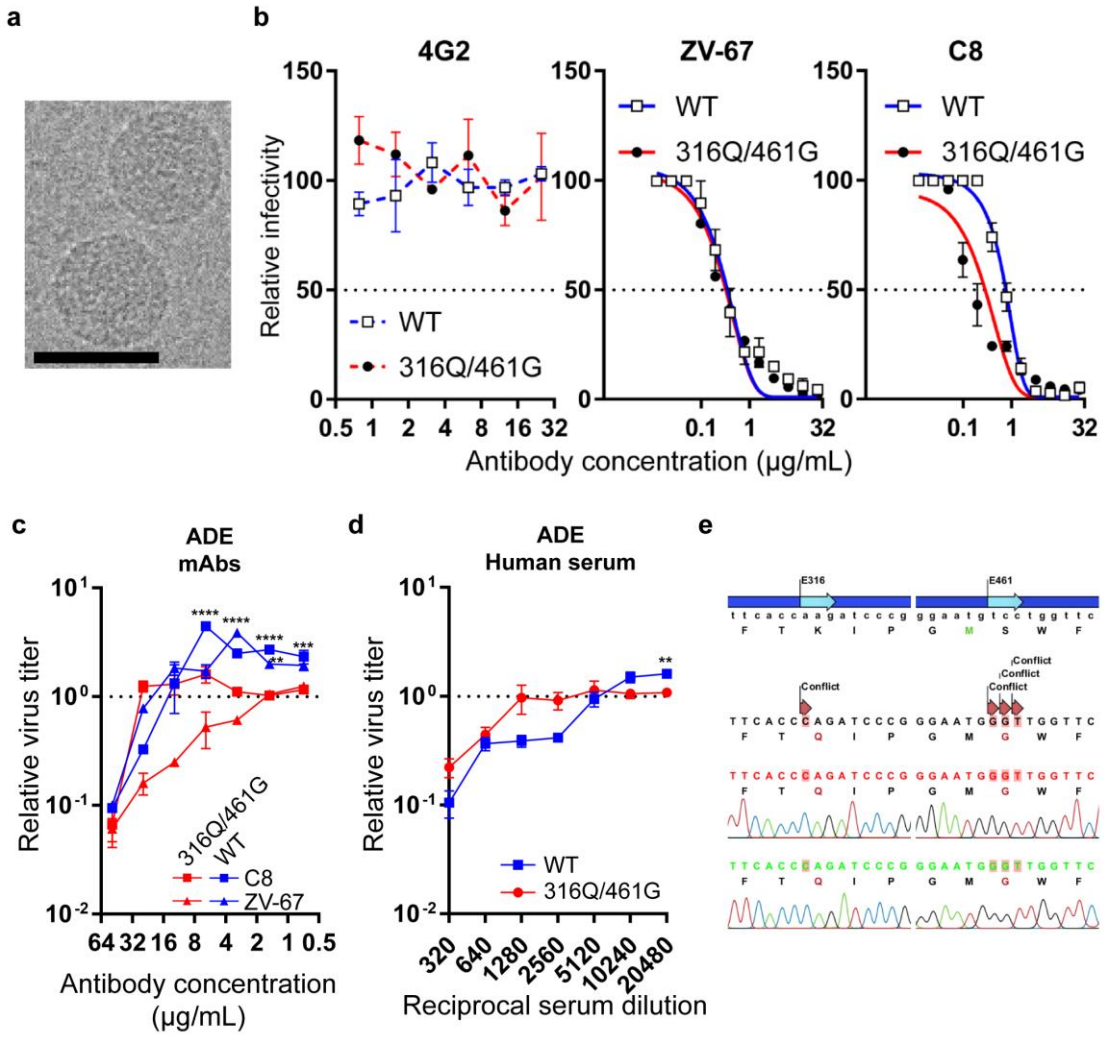



**Supplementary Figure 5**. Modelling of surface charge. **a)** Protein contact potential of dimeric E and pr interface calculated using Pymol reveals complementary charged surfaces. The pr protein underside is largely negative and interacts with positively charged regions that surround fusion loop on the E dimer and span both E monomers (dotted circle). One of the two monomers of E outlined in green for clarity. Red – negative, blue – positive, white – neutral. Position of K316 within the interaction site is indicated. **b)** Protein surface patch analysis showing the positively charged patch (dotted circle) comprising W101 from one E protein chain and H148, S149, G150,

M151, T315, K316 and K373 from the second E protein chain (default setting >50 Å2). Green – hydrophobic patch, Blue – positively charged patch. Red – negatively charged patch. The K316Q mutation results in the loss of the positively charged patch (dotted circle on bottom panel).

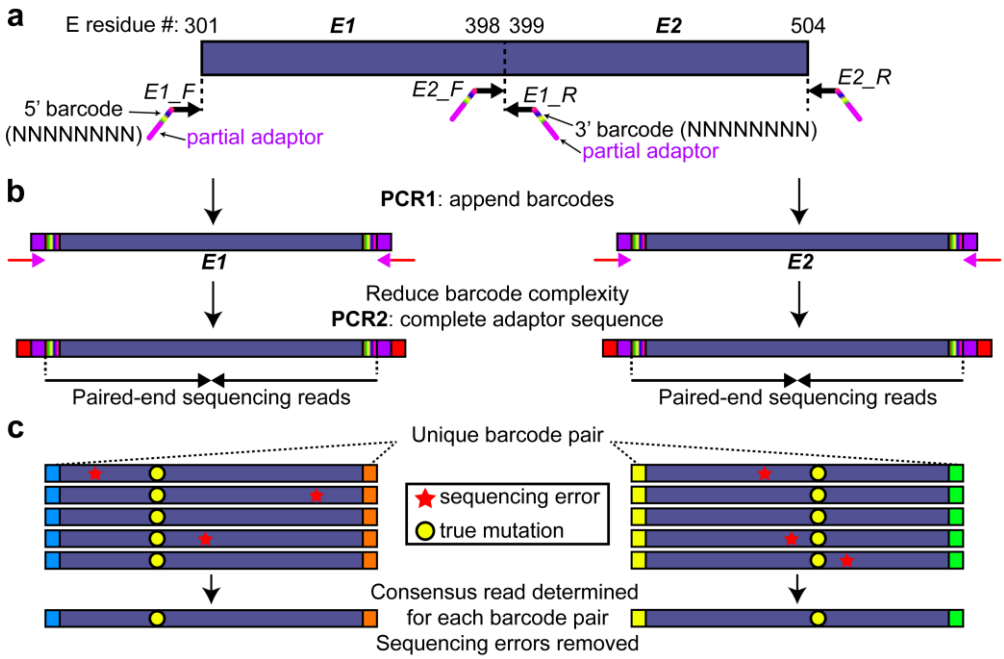



**Supplementary Figure 6.** qRT-PCR analysis of mRNAs for developmental markers in infected human iPSC-derived human brain organoids. Human iPSC-derived human brain organoids were infected with WT virus, 316Q/461G virus, or mock-infected (30 organoids per each infection/mock) and at 18 dpi organoids were harvested in TRI Reagent (split into 5 groups of 6 organoids per each infection/mock), total RNA purified and qRT-PCR performed to quantify relative levels of **a)** viral RNA, **b)** Ki67 mRNA, **c)** BRN2 mRNA, and **d)** TBR1 mRNA, compared to ETFA house keeping gene. Mean ± standard error of the mean. Statistical analysis was performed using one-way ANOVA with multiple comparisons against mock-infected. **** P = < 0.0001 (viral RNA-WT vs.316Q/461G), *** P = 0.0008 (Ki67-WT vs. 316Q/461G), *** P = 0.0006 (Ki67-WT vs. mock), **** P = <0.0001 (WT vs. mock), *** P = 0.0001 (316Q/461G vs. mock).

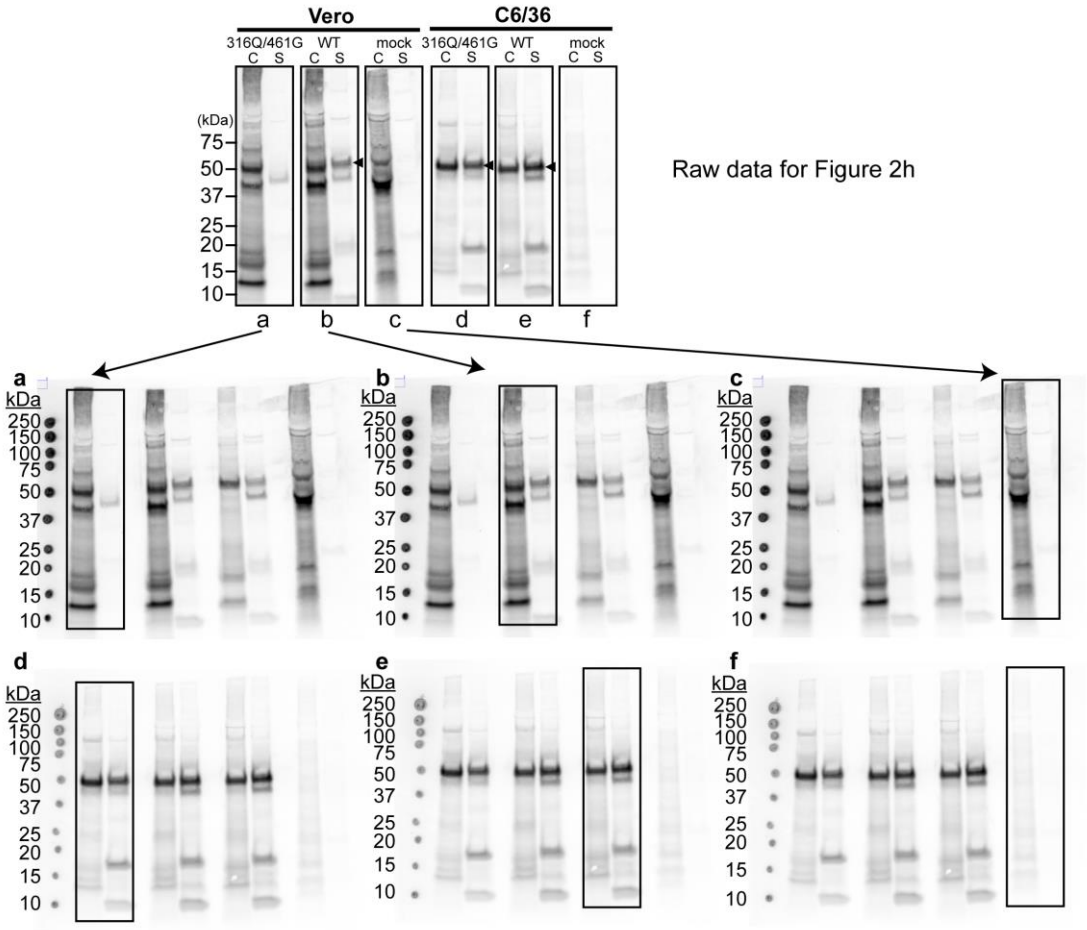



**Supplementary Figure 7.** Structural and antigenic characteristics of 316Q/461G mutant. **a)** Cryo-EM micrograph of 316Q/461G mutant virus particles produced from C6/36 cells. Scale bar is 50 nm. Representative image of n=2 independent experiments. **b)** Antigenic characterization of 316Q/461G mutant virus. PRNT assay using 4G2 mAb (negative control, does not neutralize ZIKV), or ZV-67 and C8 mAb. PRNT was performed on WT or 316Q/461G mutant viruses incubated with indicated mAbs. Relative infectivity was determined by normalizing viral titers against no antibody control. Three independent experiments were conducted. Mean ± standard error of the mean. Antibody dependent enhancement (ADE) assay using **c)** C8 and ZV-67 monoclonal antibodies (mAb), or **d)** ZIKV antibody-positive human serum.

Relative titers are calculated by normalizing to no antibody/serum controls (indicated by dotted line). n=3 independent experiments, mean ± standard error of the mean. Two-way ANOVA with multiple comparisons was performed. **** P = <0.0001, *** P = 0.0001, ** P = 0.0056. **e)** Sanger sequencing of 316Q/461G virus after 5 repeated passages in Vero cells.




**Supplementary Figure 8.** Barcoded subamplicon sequencing. **a)** Region of DMS indicated by E amino acid residue numbers (301 – 504). E DMS region amplified in two amplicons (E1 and E2). **b)** The first round of PCR (PCR1) appends random barcodes (8 × Ns, rainbow color) and part of the Illumina adaptor to each subamplicon (purple box). PCR2 is performed after complexity of barcoded subamplicons is diluted to less than the sequencing depth, and a second round of PCR adds the remaining adaptor (red box). **c)** Paired-end sequencing is performed and reads are grouped by their unique barcode pairs, distinguishing sequencing errors (red star) from true mutations (yellow circle) that occur in all reads.




315 **Supplementary Figure 9. Raw data for main figure 2h.** Each individual blot is shown in full, unmodified in any way other than to note which lanes are relevant to the corresponding main text figure and to label the molecular weight markers.

**Supplementary Data 1. Data analysis Jupyter Notebook.** The Jupyter notebook rendered as a html file describes the step-by-step analysis of the deep sequencing data.

**Supplementary Data 2. Raw data files.** The zip file contains all the raw data and results files from analysis, as well as template fasta files and ipython notebook file for running the full analysis.