Multiple roles of the non-structural protein 3 (nsP3) alphavirus unique domain (AUD) during Chikungunya virus genome replication and transcription

Yanni Gao, Niluka Goonawardane, Andrew Tuplin and Mark Harris*

School of Molecular and Cellular Biology, Faculty of Biological Sciences and Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK

* Corresponding author: e-mail: m.harris@leeds.ac.uk

Running Title: Chikungunya virus nsP3 AUD function

Word Counts:

Abstract: 243, author summary: 143, main text (excluding figure legends): 7015

Twelve figures, one table, two supplementary figures.

12 # Abstract

13 Chikungunya virus (CHIKV) is a re-emerging *Alphavirus* causing fever, joint pain, skin
14 rash, arthralgia, and occasionally death. Antiviral therapies and/or effective vaccines
15 are urgently required. CHIKV biology is poorly understood, in particular the functions
16 of the non-structural protein 3 (nsP3). Here we present the results of a mutagenic
17 analysis of the alphavirus unique domain (AUD) of nsP3. Informed by the structure of
18 the Sindbis virus AUD and an alignment of amino acid sequences of multiple
19 alphaviruses, a series of mutations in the AUD were generated in a CHIKV sub-
20 genomic replicon. This analysis revealed an essential role for the AUD in CHIKV RNA
21 replication, with mutants exhibiting species- and cell-type specific phenotypes. To test
22 if the AUD played a role in other stages of the virus lifecycle, the mutant panel was
23 also analysed in the context of infectious CHIKV. Results indicated that, in addition to
24 a role in RNA replication, the AUD was also required for virus assembly. Further
25 analysis revealed that one mutant (P247A/V248A) specifically blocked transcription of
26 the subgenomic RNA leading to a dramatic reduction in synthesis of the structural
27 proteins and concomitant reduction in virus production. This phenotype could be
28 explained by both a reduction in the binding of the P247A/V248A mutant nsP3 to viral
29 genomic RNA *in vivo*, and the reduced affinity of the mutant AUD for the subgenomic
30 promoter RNA *in vitro*. We propose that the AUD is a pleiotropic protein domain, with
31 multiple functions during CHIKV RNA synthesis.

32

Author summary

Chikungunya virus (CHIKV) is an emerging threat to world health. It is transmitted by Aedes species mosquitos, and has caused massive epidemics across the globe. The virus causes fever, rash, arthritis and can sometimes be fatal. The biology of CHIKV is poorly understood, to address this deficiency we aimed to identify functions of one of the viral proteins, nsP3. We focussed on the central part of this protein, termed the alphavirus unique domain (AUD) because it is unique to the genus of viruses to which CHIKV belongs – the *Alphaviruses* – and not present in other related viruses. By making changes (mutations) in the AUD and analysing the effects of these changes we show that it is involved in multiple stages of the virus lifecycle. These observations identify nsP3 and the AUD in particular as a potential target for antiviral therapy or rational vaccine design.

Introduction

Chikungunya virus (CHIKV; family *Togaviridae*, genus *Alphavirus*) [1] is an arbovirus that causes fever, rash and arthralgia with an infrequent fatal outcome [2]. It was first isolated in Tanzania in 1952-1953 [3, 4]. During the last 50 years, numerous CHIKV re-emergences have been documented across the world, including Africa, Asia, Europe and America [5, 6]. CHIKV is transmitted to humans by mosquitoes, mainly *Aedes aegypti* and *Ae. albopictus.* The latter can reproduce in more moderate climates which means that CHIKV has spread from Southern Africa and is now present across the Americas and parts of Southern Europe (including France and Italy). Increasing global temperatures resulting from climate change raise the concern that CHIKV will spread further. In this regard, there are no antiviral therapies or safe, effective vaccines available to treat CHIKV infection.

CHIKV has an 11.5 kilobase positive-sense, single-stranded RNA genome that is both capped and polyadenylated, and contains two open reading frames (ORFs). The first ORF is translated directly from full-length genomic RNA and encodes the non-structural proteins nsP1 to nsP4. These four proteins are required for RNA synthesis – generating both negative and positive full-length genomic RNA and a smaller subgenomic RNA from which the second ORF is translated to yield the structural proteins (capsid, envelope glycoproteins E1-3 and the 6K viroporin). Biochemical functions have been ascribed to 3 of the nsPs: nsP1 exhibits methyl- and guanyl-transferase activities, nsP2 is a helicase/protease, and nsP4 is the RNA-dependent

RNA polymerase. Although nsP3 plays an essential role in RNA replication, its biochemical functions remain largely undefined [7, 8]. It is proposed to comprise three domains: at the N-terminus is a macro-domain which exhibits both ADP-ribose and RNA binding, and ADP-ribosylhydrolase capabilities [9-11]. This is followed by the alphavirus unique domain (AUD), so called as it is only present in the *Alphavirus* genus and is absent from the closely related Rubella virus (the sole member of the *Rubivirus* genus within the *Togaviridae*), and a C-terminal hypervariable region (Fig. 1A). The latter plays an important role in virus-host interactions and may be a significant determinant of pathogenesis through interactions with cell-type-specific factors [12, 13]. The AUD is located in the centre of nsP3, and despite a high level of sequence homology across the alphaviruses, the function of this domain remains elusive. Of note, the structure of the Sindbis virus (SINV) AUD has been determined in the context of a pre-cleavage fragment of the polyprotein spanning the C-terminus of nsP2 (protease and methyl-transferase-like domains), and the N-terminus of nsP3 (macro-domain and AUD) [14]. This revealed that the AUD presents a unique protein fold containing a zinc coordination site. In this study we sought to investigate the function of AUD during the virus lifecycle in cells derived from both the vertebrate host and the mosquito vector, to identify targets for antiviral intervention and means of rational attenuation for vaccine development. By mutagenic analysis we demonstrate that the AUD exhibits both species- and cell-type specific phenotypes, and plays roles in both virus genome replication and structural protein expression.

Results

Construction of CHIKV subgenomic replicons with AUD mutations.

To identify residues within the AUD that are conserved across the *Alphavirus* genus we first aligned the AUD amino acid sequences of a range of both Old World and New World alphaviruses (Supplementary Fig S1A). As the AUD sequences between SINV and CHIKV are highly conserved (118 of 243 residues are identical), the nsP2/nsP3 protein structure of SINV [14] was referenced to identify the putative location of each of the conserved residues. Following from the above analysis, 10 residues were chosen for further study as they were located on the surface of the protein (Supplementary Fig S1B) and were either absolutely conserved throughout the alphaviruses, or in other cases were substituted by residues with similar physical characteristics (specifically the corresponding residue for both Met219 and Val260 in CHIKV is leucine in SINV) (Figs 1B and S1A). We chose to make two single substitutions (nsP3 amino acid numbering: M219A and E225A), and four double substitutions of adjacent or closely located residues (R243A/K245A, P247A/V248A, V260A/P261A and C262A/C264A). The latter two residues were shown to be involved in zinc coordination in the SINV AUD domain [14]. These mutations were cloned into a CHIKV subgenomic replicon (CHIKV-D-Luc-SGR) (Fig. 1C), derived from the ECSA strain (ICRES) (kind gift from Andres Merits, University of Tartu). This construct contains two luciferase reporter genes, a renilla luciferase (RLuc) is fused in frame within the C-terminal hypervariable domain of nsP3 in ORF1 and a firefly

luciferase (FLuc) replaces the structural protein encoding region of ORF2, allowing simultaneous assessment of both input translation and genome replication. As a negative control we also created a polymerase-inactive mutant (GDD-GAA in the active site of nsP4).

CHIKV subgenomic replicons exhibited different phenotypes in human, mammalian and mosquito cells.

To analyse the effects of the AUD mutations on CHIKV genome replication, the panel of mutant CHIKV-D-Luc-SGR RNAs were transfected into a range of cell lines. As both liver and muscle are target organs for CHIKV infection we used three human cell lines. The human hepatoma cell line Huh7 are well characterised and we have previously shown [15] that they efficiently support CHIKV replication. To test any potential role of the AUD in protecting CHIKV from innate immune sensing we also used the Huh7 derivative cell line Huh7.5, which have a defect in innate immunity due to a mutation in one allele of the retinoic acid-inducible gene I (RIG-I) [16]. To investigate the role of the AUD in infection of muscle cells we used a human rhabdomyosarcoma cell line, RD. Additionally, two other mammalian (non-human) cell lines were used: C2C12 (a murine myoblast cell line) and BHK-21 (baby hamster kidney cells) based on their ability to support high levels of CHIKV replication [15]. Lastly, we used two mosquito (*Ae. albopictus*) derived cell lines: U4.4 and C6/36. Of note C6/36 have a defect in RNA interference (RNAi) due to a frameshift mutation in the Dcr2 gene, leading to production of a truncated and inactive Dicer-2 protein [17]. Again, use of these cells

was intended to allow us to assess any role of the AUD in counteracting mosquito innate immunity.

We first tested replication in the human hepatoma cell line, Huh7 (Fig. 2A). Wildtype CHIKV-D-Luc-SGR exhibited robust replication in these cells with FLuc levels (a measure of genome replication) increasing approx. 30-fold between 4-12 h post-transfection. Consistent with this, RLuc levels (reflecting both input translation and replication) increased between 4-12 h but then declined at 24 h to input levels, possibly due to preferential transcription of the sub-genomic RNA at later times. Of the mutants M219A exhibited a modest, non-significant, reduction in replication, E225A replicated as wild type, but the other four nsP3 mutants and the nsP4 GAA mutant failed to replicate. The replication defect was indicated by either a reduction or a minimal increase in both RLuc and FLuc values from 4-24 h. A similar picture emerged when the mutant panel was screened in Huh7.5 cells (Fig. 2B). Consistent with the defect in cytosolic RNA sensing, replication of wildtype, M219A and E225A was higher in Huh7.5 cells compared to Huh7, however, this did not allow replication of the inactive mutants. For RD human rhabdomyosarcoma cells, a slightly different picture emerged (Fig. 2C): firstly both M219A and E225A replicated to a similar level as wildtype. Secondly, the P247A/V248A mutant, which was unable to replicate in Huh7 or Huh7.5 cells, was able to replicate to a low level in RD cells. The other 3 mutants and nsP4 GAA again failed to replicate.

We then evaluated the mutant panel in two other mammalian cell lines: C2C12 murine myoblasts (Fig 3A) and BHK-21 (Fig 3B). Wildtype CHIKV-D-Luc-SGR replicated to very high levels in both cell lines, with FLuc levels increasing ~1000-fold between 4-24 h. Overall the phenotypes of the panel of mutants were similar to those observed in RD, however two noticeable differences were observed. Firstly, R243A/K245A showed a low replication level in C2C12 cells. Secondly, P247A/V248A was capable of robust replication in both (albeit nearly 10-fold lower than wildtype). Interestingly, although FLuc levels for P247A/V248A were reduced, the concomitant RLuc levels were higher than wildtype, suggesting that although this mutant is replication competent there may be a defect in translation of ORF2. These data suggested that P247 and V248 were required for CHIKV genome replication in liver-derived cells, whilst enhancing but not essential for replication in cells derived from muscle or kidney, implying some cell type specific interactions of nsP3. V261A/P261A (adjacent to the zinc–binding site), and the zinc-coordinating cysteine mutant C262A/C264A were unable to replicate in either cell line, being indistinguishable from the GAA nsP4 control.

As a mosquito transmitted virus, CHIKV must replicate in both mammalian and mosquito cells. We therefore proceeded to evaluate the replicative capacity of the mutant panel in cells derived from the *Ae. albopictus* mosquito. Two cell lines were used: U4.4 and C6/36. The major difference between these two cell lines is that C6/36 have a defect in the RNAi response due to a frameshift mutation in the Dcr2 gene [17].

Consistent with this, although both mosquito cell lines supported robust replication, C6/36 supported higher levels than U4.4 (up to 1000-fold increase at 48 h). As described below, we observed remarkable differences in the mutant phenotypes in these cells compared to the mammalian cells (Fig. 4). The first difference was that M219A failed to replicate in U4.4 cells (Fig. 4A) but exhibited wildtype levels of replication in C6/36 cells (Fig. 4B), suggesting that M219 might be involved in interacting with, and inhibiting, the mosquito cell RNAi pathway. Secondly, R243A/K245A, which was unable to replicate in human cell lines and only showed low replication level in the C2C12 cells, was fully replication competent in both mosquito cell lines. Mutant P247A/V248A was partially replication competent in both cell lines, whereas as seen in mammalian cell lines neither V261A/P261A nor C262A/C264A replicated in mosquito cells.

The striking phenotypic difference between mammalian and mosquito cell lines for R243A/K245A led us to investigate this further. We considered that a simple explanation might be that the mutations had reverted in mosquito cell lines. To test this we extracted cytoplasmic RNA from C2C12, U4.4 and C6/36 cells at various time post transfection, and subjected them to RT-PCR and sequence analysis. In C2C12 cells at 48 h.p.t. we did not observe any sign of reversion (Fig 4C) – the sequence remained the same as the input RNA. However, for both U4.4 at 48 h.p.t., and C6/36 samples the sequence traces revealed the presence of a mixed population of mutant and wildtype. Notably we observed a sequential accumulation of revertants in the

C6/36 samples: At 24 h.p.t. a very low proportion of revertants at the first position in the two codons was seen, at 48 h.p.t. the proportion increased and at 72 h.p.t. the sequences were almost entirely wildtype. These data are consistent with a requirement for R243 and K245 for efficient CHIKV genome replication.

Role of the AUD in the context of infection of mammalian cells with CHIKV.

We then sought to determine if the AUD played any role in other stages of the virus lifecycle. To test this, a subset of mutations that were able to replicate in all, or some of, the mammalian cells tested (M219A, E225A and P247A/V248A), together with the nsP4 GAA mutant as negative control, were introduced into an infectious CHIKV construct (ICRES-CHIKV). Capped *in vitro* transcribed virus RNA was electroporated into C2C12 cells and production of infectious virus was assessed by plaque assay of cell supernatants at 8, 24 and 48 h.p.e. C2C12 cells were chosen as our previous analysis had revealed that CHIKV grew to very high titres in these cells [15], and they are physiologically relevant, being muscle derived. As expected (Fig 5A), wildtype CHIKV produced a high titre of infectious virus following electroporation of C2C12 cells whereas the nsP4 GAA mutant did not produce any infectious virus. M219A and E225A were indistinguishable from wildtype, but P247A/V248A showed a significantly lower titre (approx. 10-fold reduced). Importantly, we showed by RT-PCR and sequence analysis of the infectious virus stocks, that they all retained the mutant sequence and had not undergone reversion to wildtype (Supplementary Fig S2).

During the course of these assays we noted that the P247A/V248A mutant uniquely exhibited a much smaller plaque size than the wildtype (Fig. 5B). We reasoned that this might reflect a defect in either virus production or spread that could be masked in the electroporation procedure due to the high level of input RNA. To test this we performed a one-step growth assay by infecting C2C12 cells at an MOI of 0.1 with either wildtype CHIKV or the 3 mutants (M219A, E225A and P247A/V248A). Cell supernatants were harvested at various times post infection and analysed for genomic RNA by qRT-PCR (Fig 5C, and infectivity by plaque assay (Fig 5D). Wildtype, M219A and E225A showed a rapid increase in both genomic RNA and infectivity between 8-48 h.p.i., reaching very high titres (for wildtype: $3.4x10^{10}$ RNA copies/ml and $4.7x10^{8}$ pfu/ml). In contrast, levels of P247A/V248A accumulated very slowly, reaching a maximum of $4.6\ x10^{6}$ RNA copies/ml and $2.8x10^{5}$ pfu/ml at 48 h.p.i. However, direct comparison of the genomic RNA quantification with the infectivity revealed that the specific infectivity of all four viruses were indistinguishable (Fig 5E). We conclude that, although P247A/V248A exhibited a defect in production of virus particles, the virions produced were equally infectious as wildtype.

We then asked whether the reduced virus titre exhibited by P247A/V248A resulted from a defect in virus assembly or release from infected cells. To address this question we first analysed levels of viral genomic RNA (by qRT-PCR) and infectious virus (by plaque assay) present at 24 h.p.i. within cells infected with wildtype or the 3 mutants (at an MOI of 1). This analysis (Fig 6A) revealed that the levels of intracellular genomic

RNA for wildtype, M219A and E225A were comparable whereas P247A/V248A showed a 1000-fold reduction. This was consistent with the replicon data – for P247A/V248A genomic RNA levels were reduced from a mean of 7.9 x 10^9 RNA copies/ml to 7.4x10^6. Levels of infectious virus were similar for wildtype, M219A and E225A, however, uniquely P247A/V248A showed a dramatic 10^7-fold reduction in the amount of intracellular infectious virus compared to wildtype: from 2.4 x 10^8 pfu/ml to 9.8 x 10^1 pfu/ml. This was reflected in a dramatic change in the ratio of genomic RNA:infectivity (Fig 6B), suggesting that although P247A/V248A exhibited a defect in genome replication there was an additional, more substantial phenotype in the production of infectious virus particles. The difference in magnitude of these two phenotypes suggests that they represent different functions of the AUD. To provide further support for this observation, we performed a similar experiment in which C2C12 cells were electroporated with wildtype or the 3 mutant virus RNAs (Fig 6C, D). This analysis revealed that the ratio of extra- to intracellular virus titres was significantly higher for P247A/V248A compared to wildtype and the other two mutants. We conclude that, although P247A/V248A produces less infectious virus, this can be released from the infected cells more efficiently than wildtype and the other two mutants.

The P247A/V248A mutation selectively impairs subgenomic RNA synthesis.

We considered that the reason for the reduction in virus assembly exhibited by P247A/V248A could be due to a direct role of nsP3 in this process, or some defect in

the production of the structural proteins. To test this we analysed electroporated cells by western blot for the presence of both nsP3 and the capsid protein. P247A/V248A exhibited a modest reduction in nsP3 expression but a much greater reduction in the level of capsid expression. Indeed the ratio of capsid to nsP3 expression determined from the western blot analysis was approximately 10-fold lower for P247A/V248A (Fig 7A). During alphavirus replication the non-structural proteins (including nsP3) are translated from the full-length genomic RNA (gRNA), whereas the capsid and other structural proteins are translated from a subgenomic RNA (sgRNA). Transcription of both positive sense RNAs is mediated by a complex of the four nsPs using the full-length negative strand as template. This complex either initiates transcription from the 3′end of the negative strand or from the sub-genomic promoter. We hypothesised that the reduction in capsid expression for P247A/V248A could result from a defect in sgRNA transcription. To test this, C2C12 cells were electroporated and treated with actinomycin D (ActD) to block cellular RNA synthesis, prior to labelling with [^{3}H]-uridine. Cellular RNA was extracted and analysed by MOPS-formaldehyde gel electrophoresis and autoradiography. As shown in Fig 7B, for WT, M219A and E225A, 2 radiolabelled species corresponding to gRNA and sgRNA were detected. However, for P247A/V248A, the sgRNA was present at very low levels, almost undetectable. The corresponding ratio of gRNA:sgRNA for P247A/V248A (25.3:1) was significantly higher than that of WT (1.5:1). As controls, mock electroporated cells treated with ActD contained no [^{3}H]-labelled RNA species, whereas in the absence of ActD the expected

smear of [^{3}H]-labelled RNAs with predominant bands corresponding to 18S and 28S ribosomal RNAs were observed (Fig 7B). To confirm these results, the harvested RNAs were also analysed by sucrose gradient centrifugation. Consistent with the electrophoretic analysis, wildtype, M219A and E225A showed two peaks corresponding to sgRNA and gRNA, whereas P247A/V248A exhibited a dramatically reduced sgRNA peak (Fig 7C).

Effect of the P247A/V248A mutation on the RNA-binding activity of the AUD.

Our data thus far are consistent with the hypothesis that the P247A/V248A mutation results in a reduction in the ability of the nsP complex to recognise and initiate transcription from the subgenomic promoter. As the AUD has been predicted to possess RNA-binding activity [14], and initiation of gRNA or sgRNA transcription by the nsP complex will require specific recognition of cognate RNA sequences on the CHIKV negative strand, we further postulated that the phenotype of the P247A/V248A might be explained by a defect in RNA binding activity. To test this we expressed wildtype and P247A/V248A AUD mutants in *E.coli* as His-Sumo fusion proteins. The AUDs were cleaved from the fusion proteins by Sumo-protease and analysed by SDS-PAGE. As shown in Fig 8A, the AUDs could be purified to a high degree of homogeneity. Circular dichroism (CD) analysis (Fig 8B) revealed that, as expected, the AUDs comprised predominantly α-helix with no significant differences in the overall structure as a result of the mutations. To test for RNA-binding activity we performed a filter binding assay [18] using purified AUDs and a radiolabelled RNA corresponding

295 to the 3′ end of the CHIKV genome (3′UTR(+)), the 3′ end of the genomic negative
296 strand RNA (5′UTR(-)), or the negative strand subgenomic promoter (sg-prom(-)) (Fig
297 8C). Both wildtype and P247A/V248A AUD were able to bind the 3′UTR(+) (Fig 8D)
298 and 5′UTR(-) (Fig 8E), however P247A/V248A exhibited a significant increase in K_d
299 values and decrease in maximal binding levels (endpoints) compared to wildtype. As
300 3′UTR(+) and 5′RNA(-) are involved in the initiation of negative and positive strand
301 genome RNA synthesis, respectively; impaired binding of the P247A/V248A mutant
302 AUD may explain the observed defect in CHIKV genome replication (Fig 3). For
303 binding to the sg-prom(-) RNA (Fig 8F), P247A/V248A AUD showed a different
304 phenotype with both a higher endpoint and K_d than wildtype. K_d and endpoint values
305 are listed in Fig 8G. This result suggests that the P247A/V248A defect in sgRNA
306 synthesis may be in part explained by a reduction in the ability to specifically bind the
307 subgenomic promoter.

308 To explore the RNA binding activity of nsP3 to CHIKV genomic RNA during virus
309 replication, we exploited a previously generated derivative of the ICRES infectious
310 clone in which a twin-strep tag (TST) was introduced in frame near the C-terminus of
311 nsP3, allowing efficient affinity purification of nsP3 by streptactin chromatography.
312 We had previously used this experimental approach to investigate protein-protein and
313 protein-RNA interactions of the hepatitis C virus NS5A protein [19-21]. C2C12 cells
314 were electroporated with either wildtype or P247A/V248A mutant TST-nsP3 CHIKV
315 RNAs. NsP3 proteins were purified from cell lysates on streptactin beads and analysed

316 by western blot for nsP3 (Fig 9A) and qRT-PCR to determine the amount of gRNA
317 associated with nsP3 (Fig 9B). Consistent with the *in vitro* RNA filter binding assay
318 data, P247A/V248A bound approximately 10-fold less gRNA compared to wildtype
319 (Fig 9C).

320 **Sub-cellular localisation of nsP3, capsid and dsRNA during CHIKV replication.**

321 The effect of P247A/V248A on the nsP3:gRNA interaction suggested that this
322 mutation might also disrupt the subcellular localisation of nsP3 in relation to both
323 replication complexes and sites of virion assembly. To test this we exploited another
324 derivative of the ICRES infectious CHIKV clone in which ZsGreen was inserted into
325 nsP3 at the same position as the TST tag [22, 23]. C2C12 cells were electroporated
326 with ICRES-nsP3-ZsGreen-CHIKV RNAs (wildtype or P247A/V248A), and cells were
327 analysed by confocal laser scanning microscopy (CLSM) with Airyscan for the
328 distribution of nsP3, capsid (as a marker for virion assembly sites) and dsRNA (as a
329 marker of genome replication) at different times post-electroporation. For wildtype at
330 4 h.p.e. (Fig 10), small clusters of nsP3, capsid and dsRNA appeared in the cytoplasm
331 but there was little co-localisation. By 8 h.p.e., nsP3, capsid and dsRNA co-localised
332 in larger clusters, these appeared to accumulate at the plasma membrane at 12 and
333 16 h.p.e., by which time the majority of nsP3, capsid and dsRNA were co-localised on
334 plasma membrane. By 24 h.p.e., it was clear that the infection cycle was complete
335 as there was a reduction in levels of nsP3, capsid and dsRNA. Interestingly, capsid
336 and dsRNA were still co-localised at the plasma membrane while most nsP3 was

perinuclear. In contrast, P247A/V248A exhibited a very different distribution pattern of all three markers throughout the infection cycle (Fig 11). Consistent with the western blot data (Fig 7A) levels of capsid and dsRNA were lower than wildtype at all timepoints, but in addition the co-localisation of nsP3, capsid and dsRNA was markedly reduced and the three markers never accumulated at the plasma membrane as seen for wildtype. Consistent with the delay in virus release shown in Fig 5, it was clear that, unlike wildtype, the infection cycle was not complete by 24 h.p.e. as levels of nsP3 and capsid were highest at this timepoint.

To provide a quantitative assessment of the differences between wildtype and P247A/V248A we quantified the percentage co-localisation of nsP3 with either dsRNA (Fig 12A) or capsid (Fig 12B) from 5 cells at each timepoint. As shown in Fig 12A, at all timepoints the percentage co-localisation of nsP3 with dsRNA was significantly lower for P247A/V248A. The results for nsP3 co-localisation with capsid were less clear-cut: for wildtype there was a gradual increase from 8-24 h.p.e., however for P247A/V248A levels remained fairly constant with a transient drop at 16 h.p.e., consistent with the confocal images in Fig 11. These data are consistent with a role for the AUD, and residues P247/V248 in particular, in mediating interactions between nsP3 and both genome replication complexes and virus assembly sites. They also suggest that nsP3 is not only involved in virus genome replication, but may also play a role in the trafficking of capsid to plasma membrane.

Discussion

Of the four alphavirus non-structural proteins, nsP3 remains the least well understood [24]. The protein consists of three domains, the N-terminal of which has been identified as a macrodomain that binds to ADP-ribose and possesses ADP-ribosylhydrolase activity [9]. Recent studies have proposed a role for this enzymatic activity in virus pathogenesis but as yet the underlying mechanisms remain elusive [11]. The C-terminal hypervariable domain differs dramatically in amino acid sequence between different alphaviruses and is intrinsically disordered. It has been shown to interact with a range of cellular proteins, including components of stress granules [25], and is implicated in the assembly of virus genome replication complexes.

In contrast, we know virtually nothing about the function of the central AUD domain. The fact that this domain is highly conserved between different alphaviruses suggests that it plays a fundamental role in the virus lifecycle. Detailed structural information about the AUD is available however, as the partial structure of the SINV nsP2-nsP3 precursor, including the C-terminal protease and methyltransferase-like domains of nsP2 and the macro and AUD domains of nsP3, has been determined [14]. This analysis revealed that the AUD contained an unique zinc-binding fold with four cysteine residues coordinating a zinc molecule, this formed part of a putative RNA binding surface. Mutagenesis of two of these cysteines revealed an essential role in virus replication. Our data agree with this observation, as the C262A/C264A mutant failed to replicate in any cell type tested.

Mutation of two residues adjacent to the zinc-binding cysteines, V260A/P261A, also completely abrogated CHIKV genome replication. Although adjacent in the primary amino acid sequence, these residues are located on the distal face of the AUD (Fig 1B), suggesting that they are not involved in zinc binding, but may instead interact with key cellular factor(s) or play an alternative structural role.

In contrast, the other mutants generated during this study exhibited a number of distinct cell-type and species-specific phenotypes (summarised in Table 1). Mutation of two surface exposed basic residues (R243 and K245) abrogated replication in all mammalian cells but showed full replication capability in mosquito cells. However, this apparent discrepancy could be explained by the observation that these two mutations rapidly reverted to wildtype in mosquito cells but failed to do so in C2C12 cells (Fig 4C). These data indicate that R243 and K245 are required for CHIKV genome replication. We do not have an explanation for why R243A/K245A, but not the other lethal mutations, was able to revert in mosquito cells. However it is noteworthy that the reversion to the wildtype sequence took 72 h in C6/36 cells, suggesting that perhaps the lack of cytopathology of CHIKV replication in mosquito cells could facilitate the replication of a minority species. Interestingly, the sequence trace at 24 h in C6/36 shows the presence of a such minority species that would encode a Thr at 243 and 245, suggesting that the two basic residues are not absolutely required.

Table 1. AUD mutant replication phenotypes in different cell types.

CHIKV-D-*luc* SGR	Human			Rodent		Mosquito	
	Huh7	Huh7.5	RD	C2C12	BHK	U4.4	C6/36
Wildtype							
M219A							
E225A							
R243A/K245A							
P247A/V248A							
V260A/P261A							
C262A/C264A							

Key: Wildtype replication, impaired replication, no replication. reversion

M219 was also of particular interest as mutation of this residue had no significant effect on genome replication in any cell type apart from U4.4 mosquito cells. M219A replicated well in C6/36 mosquito cells and the key difference between these two cell lines is that C6/36 have a defect in the RNAi response due to a Dcr2 mutation [17]. We propose therefore that M219 may interact with a component of the mosquito RNAi response to inhibit this key mosquito antiviral pathway. We are currently undertaking proteomic and functional analysis to test this hypothesis.

In the second part of this study we focussed on P247A/V248A to address the molecular mechanism underpinning the phenotype of this mutant. In the context of the subgenomic replicon P247A/V248A showed a variety of phenotypes from

411 complete lack of replication in Huh7 and Huh7.5 cells (Fig 2), to a 10-fold reduction in
412 other mammalian and mosquito cells (Fig 3). Virus data in C2C12 cells were consistent
413 with replicon results: following electroporation of viral RNA P247A/V248A showed a
414 modest but significant defect in virus infectivity and exhibited a small plaque
415 phenotype. Further study indicated that P247A/V248A was competent in virus entry
416 and virus release, however, it exhibited a major defect in assembly of infectious virus
417 particles. This defect led to both a delay and a reduction in the release of infectious
418 virus, consistent with the small plaque phenotype. The molecular mechanism
419 underpinning the P247A/V248A defect was shown to be a reduction in subgenomic
420 RNA synthesis, leading to a concomitant reduction in the expression of the structural
421 proteins. It is noteworthy that when analysed in the context of the replicon in C2C12
422 cells, the P247A/V248A mutant exhibited a 10-fold reduction in FLuc, but RLuc was
423 higher than wildtype. This is also consistent with a defect in transcription of the
424 subgenomic mRNA. The reduced affinity of P247A/V248A AUD for the CHIKV sg-
425 prom(-) RNA as demonstrated by the RNA filter binding assay may help to explain
426 this, but it is possible that other effects of P247A/V248A, such as aberrant host protein
427 recruitment, may also help to explain the defect in subgenomic RNA synthesis. Of
428 note, P247A/V248A AUD also exhibited impaired binding to the 3′UTR(+) and 5′RNA(-
429) *in vitro*, and also bound less genomic RNA *in vivo*, compared to wildtype, consistent
430 with an overall defect in RNA replication. Previous studies have also shown that nsP3
431 is important for initial replication complex formation and negative strand RNA

synthesis [26]. Taken together we propose that the AUD plays a critical role in all stages of CHIKV RNA synthesis, but particularly in the transcription of the sgRNA.

Analysis of the distribution of nsP3, capsid and dsRNA during CHIKV replication by confocal microscopy revealed further insights into the P247A/V248A phenotype. Wildtype nsP3 exhibited a high level of co-localisation with dsRNA at all time points up to 24 h.p.e., consistent with the role of nsP3 in genome replication. At 12/16 h.p.e. both nsP3 and dsRNA also co-localised with capsid and were concentrated at the plasma membrane. In contrast, P247A/V248A nsP3 showed not only a significant reduction in the co-localisation with dsRNA, but also a loss of the plasma membrane accumulation. These observations are consistent with a role for nsP3 (and the AUD in particular) in coordinating the processes of genome replication and virus assembly to facilitate production of infectious virus particles at the plasma membrane. This is in agreement with early evidence for a juxtaposition of sites of genome replication, viral protein translation and nucleocapsid assembly in the case of Sindbis virus [27]. In addition it may be that nsP3 has a role in the trafficking of nucleocapsids from these sites (cytoplasmic vacuoles (CPVs)) to the plasma membrane.

In conclusion we propose that the nsP3 AUD is a multi-functional domain. It is not only a critical determinant of both cell and species-specificity, but also plays roles in virus genome replication and assembly. The rational mutagenesis of the CHIKV AUD described here is the first detailed structure-function analysis of this domain and raises many questions. In particular, we need to determine what cellular and viral proteins

453 are interaction partners for the AUD, and investigate how these interactions shed light
454 on the phenotypes of AUD mutants? We expect that our current proteomic analysis,
455 exploiting both a novel SNAP-tagged nsP3 recently developed in our laboratory [28]
456 and a One-Strep tag (OST) approach which we recently used to identify interacting
457 partners of the hepatitis C virus NS5A protein [19], will provide some of the answers
458 to these questions. We also hope that these studies will help to identify targets for
459 antiviral intervention and means of rational attenuation for vaccine development.

460 ## Acknowledgements

461 We thank Andres Merits (Institute of Technology, University of Tartu, Estonia) for the
462 CHIKV subgenomic replicons, ICRES constructs and antibodies to nsP3 and capsid.
463 We thank Raymond Li (University of Leeds) for the OST-tagged ICRES constructs,
464 and Joseph Ward (University of Leeds) for help and advice with the ^{3}H-labelling.

465 ## Materials and Methods

466 **Sequence alignment.** AUD amino acid sequences of different alphaviruses (Sindbis,
467 Ockelbo, O'Nyong-Nyong, Ross River, Semliki Forest, Fort Morgan, Venezuelan,
468 Eastern and Western Equine Encephalitis, Highlands-J and CHIKV) were obtained
469 from NCBI and aligned by Clustal Omega. The predicted locations of conserved
470 residues were then identified by Pymol, taking the Sindbis nsP2/3 protein structure
471 (PDB ID code 4GUA) [14] as a reference.

Cell culture. Mammalian cells Huh7, Huh7.5, RD, C2C12 and BHK-21 were maintained at 37°C with 5% CO_2 in DMEM supplemented with 10% FCS, 0.5 mM non-essential amino acids and penicillin-streptomycin (100 units/mL). Huh7 cells were obtained from John McLauchlan (Centre for Virus Research, Glasgow), Huh7.5 cells from Charles Rice (Rockefeller University, New York), RD cells from Nicola Stonehouse (University of Leeds), C2C12 cells from Michelle Peckham (University of Leeds) and BHK-21 cells from John Barr (University of Leeds). Mosquito cell lines (U4.4 and C6/36) were obtained from Susan Jacobs (The Pirbright Institute) and maintained in Leibovitz's L-15 media supplemented with 10% FBS, 10% tryptose phosphate broth and penicillin-streptomycin (100 units/mL). Mosquito cells were incubated at 28°C without CO_2.

Construction of CHIKV subgenomic replicons and infectious viruses with AUD mutations. A fragment including the AUD was excised from the CHIKV-D-Luc-SGR plasmid, and inserted into pcDNA3.1 to generate pcDNA3.1-AUD. This was used as a template for site-directed (Quikchange) mutagenesis using specific primers (primer sequences available upon request) to produce the required AUD mutations. Finally, the AUD mutated fragments were excised and religated into the CHIKV-D-Luc-SGR. AUD fragments were subsequently excised from CHIKV-D-Luc-SGRs and ligated into ICRES-CHIKV-WT. For the One-Strep tag (OST) derivatives of wildtype and P247A/V248A synthetic oligonucleotides (sequence available upon request) were used to substitute the appropriate coding sequence [19] for the RLuc in nsP3.

493 **Transfection and Dual-luciferase Assay.** Capped RNAs were generated from the
494 CHIKV-D-Luc-SGR for transfection using the mMACHINE SP6 transcription kit
495 (ThermoFisher Scientific) and purified with PureLink RNA Mini Kit (Life Technologies).
496 Transfection of the CHIKV-D-Luc-SGR RNAs was performed in different cells using
497 Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions.
498 At 4, 12, 24 and 48 h post transfection (h.p.t.), cells were harvested and both Renilla
499 and Firefly luciferase activity measured using the Dual-luciferase Assay System
500 (Promega) according to the manufacturer's instructions. Each sample had three
501 repeats and the data shown in this study represent the mean of three experimental
502 replicates.

503 **Sequence analysis of subgenomic or viral RNA.** Cytoplasmic RNA from
504 electroporated or infected cells were Trizol extracted, prior to reverse transcription with
505 random primers using SuperScript IV Reverse Transcriptase (Invitrogen) according to
506 the manufacturer's instructions. cDNAs were then used as a template to amplify part
507 of nsP3 sequence (macro domain and AUD) with specific primers (primer sequences
508 available upon request). PCR products were subjected to sequencing analysis.

509 **Titration of infectious CHIKV by plaque assay.** ICRES-CHIKV RNAs were
510 produced and purified as described above. C2C12 cells were electroporated with
511 ICRES-CHIKV RNAs and incubated at 37ºC. Cell supernatants were collected at 8,
512 24 and 48 hours post electroporation (h.p.e.), diluted with cell medium and applied to
513 monolayers of BHK-21 cells for 1 h at 37ºC. The inoculum was aspirated and plates

were overlaid with 0.8% methylcellulose for 48 h at 37ºC for plaque formation. For titration of intracellular viruses, cells were freeze/thawed 3 times and supernatants were collected by centrifugation at 12000×g for 10 min., prior to application to BHK-21 cells All virus work was performed in a Biological Containment Level 3 (BSL3) laboratory. Plaques were visualised by photography with a Canon EOS 80D.

One-step virus growth curve. Infectious CHIKV was harvested from C2C12 cells electroporated with ICRES RNAs at 48 h.p.e. Cell supernatants were titrated by plaque assay in BHK-21 cells and stored at -80ºC. For virus growth kinetic analysis, C2C12 cells were infected with wildtype or AUD mutant CHIKV at an MOI of 0.1 for 1 h. Infected cells were washed three times with PBS and incubated with fresh complete medium at 37ºC. For RNA quantification, RNAs were extracted from supernatants Trizol (ThermoFisher Scientific). qRT-PCR for CHIKV genome RNA was performed with One-step MESA GREEN qRT-PCR MasterMix Plus for SYBR assay (Eurogentec) following the manufacturer's instructions. Primer sequences available upon request. Plaque assay was performed as described above.

CHIKV RNA synthesis. C2C12 cells were electroporated with ICRES RNAs for 10 h. Actinomycin D (1 µg/ml) was added and the cells were incubated for 2 h. [^{3}H]-uridine (20 µCi/ml) was then added and the cells were incubated for a further 3 h., at which time the monolayers were washed 3 times with ice-cold PBS, lysed and RNA extracted with TRIzol reagent.

For measurement of viral RNA synthesis, the harvested RNAs were separated on a MOPS-Formaldehyde gel. The gel was fixed (15% methanol, 10% acetic acid and 75% dH_2O) for 30 min followed by fluorography (Fluorographic reagent amplify, GE HEALTHCARE) for another 30 min. Gels were dried for 2 h. before exposure to autoradiographic film at -80 ºC for 4 days.

For gradient analysis, equal volumes of harvested RNAs were loaded onto 14 ml 5-25% sucrose gradients in 100mM sodium acetate and 0.1% SDS followed by centrifugation at 150,000×g for 5 h. at room temperature . Gradients were fractionated into 350 µl fractions, and radioactivity of each fractions was determined by liquid scintillation counting.

Analysis of CHIKV protein expression. C2C12 cells were electroporated with ICRES RNAs and incubated for 36 h. Cells were washed 3 times with PBS, lysed by resuspension in Glasgow lysis buffer (GLB) [1% Triton X-100, 120 mM KCl, 30mM NaCl, 5mM $MgCl_2$, 10% glycerol (v/v), and 10 mM piperazine-N,N'-bis (2-ethanesulfonic acid) (PIPES)-NaOH, pH 7.2] supplemented with protease inhibitors and phosphatase inhibitors (Roche Diagnostics), and incubated on ice for 15 min. Following separation by SDS-PAGE, proteins were transferred to a polyvinylidene fluoride (PVDF) membrane and blocked in 50% (v/v) Odyssey blocking buffer (LiCor) in Tris-buffered saline (TBS) [50 mM Tris, 150 mM NaCl, pH 7.4]. The membrane was incubated with primary antibody in 25% (v/v) Odyssey blocking buffer overnight at 4ºC, then incubated with fluorescently labelled anti-rabbit (800nm) secondary antibodies

for 1 h at room temperature (RT) before imaging on a LiCor Odyssey Sa fluorescence imager.

Expression and purification of AUD proteins. The wildtype and P247A/V248A AUD (nsP3 residues Ile141 to Gly374) were cloned into pET-28a-His-sumo for expression in *Escherichia coli* and subsequent analysis. His-sumo tagged AUD expression plasmids were transformed into Rosetta 2 and cultures were grown in Luria-Bertani (LB) medium supplemented with 50 µg/µl ampicillin and 1% (wt/vol) glucose. The cells were grown at 37°C to an optical density at 600 nm (OD_{600}) of 0.5 to 0.7 and then induced with IPTG (isopropyl-D-thiogalactopyranoside) (0.5 mM) for 5 h at 18°C. The cells were harvested by centrifugation at 7,000 rpm for 10 min. Wildtype or mutant AUD protein was purified by sequential His-tag affinity purifications. Briefly, cell pellets were suspended in 20 ml AUD lysis buffer (100 mM Tris-HCl pH 7, 200 mM NaCl, 20 mM imidazole) supplemented with 2 µg/µl DNase and EDTA-free protease inhibitor cocktail tablets (Roche). The cell suspension was lysed by sonication on ice at an amplitude of 10 µm for six pulses of 20 s separated by 20 s and the extract clarified by centrifugation at 16000×g for 30 min at 4°C. The supernatant was filtered through a 0.45 µm filter and applied to a Ni^{2+} His-tag column for purification. Purified proteins were dialyzed to remove imidazole and sumo-protease was added to cleave the His-sumo tag. After dialysis, the proteins were applied again to the His-tag purification column, and the flow-throughs were collected as purified AUD proteins.

575 **Circular Dichroism (CD) Spectroscopy.** Far-UV CD spectroscopy was performed
576 on an APP Chirascan CD spectropolarimeter to obtain the secondary structure of
577 AUDs. Spectra (190-260 nm) were recorded using 200 µl protein solution (at a
578 concentration of 0.2 mg/ml) in a 1 mm path-length cuvette. Protein CD spectra
579 deconvolution was analysed by DichroWeb.

580 **RNA filter binding assay.** Radiolabelled RNA transcripts and AUD proteins were
581 diluted in binding buffer (40 mM Tris-HCl [pH 7.5], 5 mM $MgCl_2$, 10 mM DTT, 50 µg/ml
582 bovine serum albumin, 10 µg/ml yeast tRNA [Ambion]) and pre-incubated separately
583 for 10 min at 4ºC. The binding reaction was initiated by mixing 1 nM radio-labelled
584 RNA and AUD proteins (0 to 500 nM) in a 200 µl final volume at 4ºC for 30 min.
585 Membranes were pre-soaked in binding buffer supplemented with 5% (v/v) glycerol
586 and assembled from bottom to top as follows in a slot-blot apparatus (Bio-Rad): filter
587 paper, Hybond-N nylon (Amersham Biosciences) to bind free RNA molecules, and
588 nitrocellulose (Schleicher & Schuell) to trap soluble protein-RNA complexes. After
589 assembly, 200 µl of each binding reaction mixture was applied to each slot and filtered
590 through the membranes. Each slot was washed with 0.5 ml of binding buffer and air
591 dried, and quantification of radioactivity was performed using an image plate, BAS
592 1000 Bioimager (Fuji), and Aida Image Analyser v4.22 software. Fitting was performed
593 using GraphPad Prism 5 software. In each case, the data were fitted to the hyperbolic
594 equation $R = R_{max} \times R/(K_d + [P])$, where R is the percentage of bound RNA, R_{max} is the

maximal percentage of RNA competent for binding, [P] is the concentration of AUD, and K_d is the apparent dissociation constant.

Precipitation of nsP3 and viral RNA. Co-precipitation experiments were performed in C2C12 cells electroporated with ICRES One-Strep-tag (OST) RNAs using Streptactin-agarose (Thermo Fisher Scientific), following the manufacturers protocol. Precipitated proteins were subjected to immunoblotting and co-precipitated RNAs were extracted by TRIzol and quantified by qRT-PCR.

Distribution of nsP3, capsid protein and dsRNA in cells during CHIKV replication. Wildtype and nsP3-P247A/V248A CHIKV were introduced into ICRES-nsP3-ZsGreen-CHIKV where ZsGreen was fused into the hypervariable domain of nsP3. C2C12 cells were electroporated and harvested at defined times post electroporation, fixed with 4% paraformaldehyde (PFA), permeabilised by treatment with methanol, blocked with 2% BSA, and incubated with capsid protein antibody (gift from Andres Merits) or dsRNA antibody (J2 antibody, Scicons) at 4ºC overnight, followed by secondary antibodies (Alexa Fluor 633 conjugated chicken anti-rabbit IgG and Alexa Fluor 594 conjugated donkey anti-mouse IgG) for 1h at room temperature. Distribution of nsP3, capsid protein and dsRNA were detected using a Zeiss LSM880 with Airyscan. Post-acquisition analysis was conducted using Zen software (Zen version 2015 black edition 2.3, Zeiss) or Fiji (v1.49) software [29].

Co-localisation analysis. For co-localisation analysis, Manders' overlap coefficient was calculated using Fuji ImageJ software with Just Another Co-localisation Plugin

(JACoP) (National Institutes of Health) [30]. Coefficient M1 indicated here reports the fraction of the nsP3 signal that overlaps either the anti-dsRNA or anti-capsid signal. Coefficient values range from 0 to 1, corresponding to non-overlapping images and 100% co-localisation images, respectively. Co-localisation calculations were performed on >5 cells from at least two independent experiments.

Statistical analysis. Statistical analysis was performed using unpaired two-tailed Student's t tests, unequal variance to determine statistically significant differences from the results for the wild type ($n \geq 3$). Data in bar graphs are displayed as the means ± S.E.

References

1. Chen R, Mukhopadhyay S, Merits A, Bolling B, Nasar F, Coffey LL, et al. ICTV Virus Taxonomy Profile: Togaviridae. Journal of General Virology. 2018;99(6):761-2. doi: 10.1099/jgv.0.001072.

2. Schwartz O, Albert ML. Biology and pathogenesis of Chikungunya virus. Nature Reviews Microbiology. 2010;8(7):491-500. doi: 10.1038/nrmicro2368.

3. Robinson MC. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952-53. I. Clinical features. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1955;49(1):28-32.

4. Lumsden WH. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952-53. II. General description and epidemiology. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1955;49(1):33-57.

5. Schuffenecker I, Iteman I, Michault A, Murri S, Frangeul L, Vaney MC, et al. Genome microevolution of Chikungunya viruses causing the Indian Ocean outbreak. PLoS Med. 2006;3(7):e263. doi: 10.1371/journal.pmed.0030263.

6. Sergon K, Njuguna C, Kalani R, Ofula V, Onyango C, Konongoi LS, et al. Seroprevalence of Chikungunya virus (CHIKV) infection on Lamu Island, Kenya, October 2004. American Journal of Tropical Medicine and Hygiene. 2008;78(2):333-7.

7. LaStarza MW, Lemm JA, Rice CM. Genetic analysis of the nsP3 region of Sindbis virus: evidence for roles in minus-strand and subgenomic RNA synthesis. Journal of Virology 1994;68(9):5781-91.

8. Rupp JC, Sokoloski KJ, Gebhart NN, Hardy RW. Alphavirus RNA synthesis and non-structural protein functions. Journal of General Virology. 2015;96(9):2483-500. doi: 10.1099/jgv.0.000249.

9. Malet H, Coutard B, Jamal S, Dutartre H, Papageorgiou N, Neuvonen M, et al. The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. Journal of Virology. 2009;83(13):6534-45. doi: 10.1128/JVI.00189-09.

655 10. Eckei L, Krieg S, Butepage M, Lehmann A, Gross A, Lippok B, et al. The
656 conserved macrodomains of the non-structural proteins of Chikungunya virus and
657 other pathogenic positive strand RNA viruses function as mono-ADP-
658 ribosylhydrolases. Scientific Reports. 2017;7:41746. doi: 10.1038/srep41746.

659 11. McPherson RL, Abraham R, Sreekumar E, Ong SE, Cheng SJ, Baxter VK, et al.
660 ADP-ribosylhydrolase activity of Chikungunya virus macrodomain is critical for virus
661 replication and virulence. Proc Natl Acad Sci U S A. 2017;114(7):1666-71. doi:
662 10.1073/pnas.1621485114.

663 12. Foy NJ, Akhrymuk M, Shustov AV, Frolova EI, Frolov I. Hypervariable domain of
664 nonstructural protein nsP3 of Venezuelan equine encephalitis virus determines cell-
665 specific mode of virus replication. Journal of Virology 2013;87(13):7569-84. doi:
666 10.1128/JVI.00720-13.

667 13. Foy NJ, Akhrymuk M, Akhrymuk I, Atasheva S, Bopda-Waffo A, Frolov I, et al.
668 Hypervariable domains of nsP3 proteins of New World and Old World alphaviruses
669 mediate formation of distinct, virus-specific protein complexes. Journal of Virology
670 2013;87(4):1997-2010. doi: 10.1128/JVI.02853-12.

671 14. Shin G, Yost SA, Miller MT, Elrod EJ, Grakoui A, Marcotrigiano J. Structural and
672 functional insights into alphavirus polyprotein processing and pathogenesis. Proc Natl
673 Acad Sci U S A. 2012;109(41):16534-9. doi: 10.1073/pnas.1210418109.

674 15. Roberts GC, Zothner C, Remenyi R, Merits A, Stonehouse NJ, Harris M.
675 Evaluation of a range of mammalian and mosquito cell lines for use in Chikungunya
676 virus research. Scientific Reports. 2017;7(1):14641. doi: 10.1038/s41598-017-15269-
677 w.

678 16. Sumpter R, Jr., Loo YM, Foy E, Li K, Yoneyama M, Fujita T, et al. Regulating
679 intracellular antiviral defense and permissiveness to hepatitis C virus RNA replication
680 through a cellular RNA helicase, RIG-I. Journal of Virology 2005;79(5):2689-99. doi:
681 10.1128/JVI.79.5.2689-2699.2005.

682 17. Morazzani EM, Wiley MR, Murreddu MG, Adelman ZN, Myles KM. Production of
683 virus-derived ping-pong-dependent piRNA-like small RNAs in the mosquito soma.
684 PLoS pathogens. 2012;8(1):e1002470. doi: 10.1371/journal.ppat.1002470.

685 18. Foster TL, Belyaeva T, Stonehouse NJ, Pearson AR, Harris M. All three domains
686 of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. Journal
687 of Virology. 2010;84(18):9267-77. doi:10.1128/JVI.00616-10

688 19. Goonawardane N, Gebhardt A, Bartlett C, Pichlmair A, Harris M. Phosphorylation
689 of Serine 225 in Hepatitis C Virus NS5A Regulates Protein-Protein Interactions.
690 Journal of Virology. 2017;91(17). doi: 10.1128/JVI.00805-17.

691 20. Ross-Thriepland D, Harris M. Insights into the complexity and functionality of
692 hepatitis C virus NS5A phosphorylation. Journal of Virology. 2014;88(3):1421-32. doi:
693 10.1128/JVI.03017-13.

694 21. Yin C, Goonawardane N, Stewart H, Harris M. A role for domain I of the hepatitis
695 C virus NS5A protein in virus assembly. PLoS Pathogens. 2018;14(1):e1006834. doi:
696 10.1371/journal.ppat.1006834.

697 22. Remenyi R, Gao Y, Hughes RE, Curd A, Zothner C, Peckham M, et al. Persistent
698 Replication of a Chikungunya Virus Replicon in Human Cells is Associated with
699 Presence of Stable Cytoplasmic Granules Containing Non-structural Protein 3.
700 Journal of Virology. 2018. doi: 10.1128/JVI.00477-18.

701 23. Pohjala L, Utt A, Varjak M, Lulla A, Merits A, Ahola T, et al. Inhibitors of alphavirus
702 entry and replication identified with a stable Chikungunya replicon cell line and virus-
703 based assays. PLoS One. 2011;6(12):e28923. doi: 10.1371/journal.pone.0028923.

704 24. Gotte B, Liu L, McInerney GM. The Enigmatic Alphavirus Non-Structural Protein 3
705 (nsP3) Revealing Its Secrets at Last. Viruses. 2018;10(3). doi: 10.3390/v10030105.

706 25. Kim DY, Reynaud JM, Rasalouskaya A, Akhrymuk I, Mobley JA, Frolov I, et al.
707 New World and Old World Alphaviruses Have Evolved to Exploit Different
708 Components of Stress Granules, FXR and G3BP Proteins, for Assembly of Viral
709 Replication Complexes. PLoS Pathogens. 2016;12(8):e1005810. doi:
710 10.1371/journal.ppat.1005810.

711 26. Wang YF, Sawicki SG, Sawicki DL. Alphavirus nsP3 functions to form replication
712 complexes transcribing negative-strand RNA. Journal of Virology. 1994;68(10):6466-
713 75.

714 27. Froshauer S. Alphavirus RNA replicase is located on the cytoplasmic surface of
715 endosomes and lysosomes. Journal of Cell Biology. 1988;107(6):2075-86. doi:
716 10.1083/jcb.107.6.2075.

717 28. Remenyi R, Roberts GC, Zothner C, Merits A, Harris M. SNAP-tagged
718 Chikungunya Virus Replicons Improve Visualisation of Non-Structural Protein 3 by
719 Fluorescence Microscopy. Scientific Reports. 2017;7(1):5682. doi: 10.1038/s41598-
720 017-05820-0.

721 29. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, et al.
722 Fiji: an open-source platform for biological-image analysis. Nat Methods.
723 2012;9(7):676-82. doi: 10.1038/nmeth.2019.

724 30. Bolte S, Cordelieres FP. A guided tour into subcellular colocalization analysis in
725 light microscopy. J Microsc. 2006;224(Pt 3):213-32. doi: 10.1111/j.1365-
726 2818.2006.01706.x.

727

728

Figure Legends

Figure 1. (A) Three domain structure of the alphavirus nsP3 protein. (B) Surface representation of the Sindbis virus nsP3 AUD structure (PDB ID code 4GUA) [14] (residues 161-320), including the 40 amino acid flexible linker between the macrodomain and the AUD. The locations of the mutated residues in nsP3 are indicated. The two images show opposite faces of the structure, rotated 180^{o} along the vertical axis. (C) Structure of CHIKV-D-Luc-SGR. (RLuc: Renilla luciferase, FLuc: firefly luciferase) RLuc is expressed as an internal fusion with nsP3 and thus is produced following translation of the input RNA. RLuc activity is therefore an indirect measure of both input translation and replication. FLuc is expressed from the subgenomic promoter and thus is only produced after RNA replication has occurred.

Figure 2. CHIKV AUD mutant replication in human cells. The indicated cells were transfected with CHIKV-D-luc-SGR wildtype and mutant RNAs and harvested for both RLuc and FLuc assays at the indicated time points. Luciferase values of wildtype and each mutant were normalized to 4 h values. (GAA: inactive mutant of nsP4 polymerase). Significant differences denoted by * ($P<0.05$), and ** ($P<0.01$), compared to wildtype.

Figure 3. CHIKV AUD mutant replication in murine cells. The indicated cells were transfected with CHIKV-D-luc-SGR wildtype and mutant RNAs and harvested for both

RLuc and FLuc assays at the indicated time points. Luciferase values of wildtype and each mutant were normalized to 4 h values. (GAA: inactive mutant of nsP4 polymerase). Significant differences denoted by * (P<0.05), and ** (P<0.01), compared to wildtype.

Figure 4. CHIKV AUD mutant replication in *Aedes albopictus* mosquito cells. (A, B) The indicated cells were transfected with CHIKV-D-luc-SGR wildtype and mutant RNAs and harvested for both renilla and firefly luciferase assay at the indicated time points. Luciferase values of wildtype and each mutant were normalized to 4 h values. (GAA: inactive mutant of nsP4 polymerase). Significant differences denoted by ** (P<0.01), compared to wildtype. (C) RT-PCR and sequencing analysis of CHIKV-D-luc-SGR-R243A/K245A. RNA was harvested at the indicated times, amplified by RT-PCR and sequenced. The wildtype and mutated sequences are shown below the sequence traces for reference. Nucleotide ambiguity codes used: R (A/G), S (G/C) and M (A/C).

Figure 5. Phenotype of AUD mutations in the production of infectious virus. (A) ICRES-RNAs were electroporated into C2C12 cells and supernatants were collected at 48 h.p.e. Virus was titrated by plaque assay in BHK-21 cells. (B) Plaques for wildtype and P247A/V248A were visualised illustrating the small plaque phenotype for this mutant. (C-E) C2C12 cells were infected with CHIKV (wildtype and mutants) at an MOI of 0.1. Supernatants were collected at the indicated times for genome RNA quantification (qRT-PCR) (C) and virus titration by plaque assay (D). The ratios of

genome RNA:infectivity were determined from (C) and (D) at 16, 24 and 48 h.p.i. and presented graphically (E).

Figure 6. Phenotype of AUD mutations on virus entry, release and assembly. (A) C2C12 cells were infected with CHIKV at MOI of 1. At 24 h.p.i, cells were washed with PBS and resuspended in 1 ml fresh medium. Cell suspensions were freeze/thawed 3 times to release intracellular virus. Genome RNA was quantified by qRT-PCR, and virus titrated by plaque assay. (B) Graphical representation of the ratio of infectivity to genomic RNA. (C) Intracellular and extracellular viruses were collected at 36 h.p.e from C2C12 cells electroporated with the indicated ICRES RNA, and titrated by plaque assay. (D) Graphical representation of the ratio of extracellular to intracellular virus titres. Significant difference denoted by * ($P<0.05$) compared to wildtype.

Figure 7. Effect of AUD mutations on CHIKV protein expression and RNA synthesis. (A) C2C12 cells were electroporated with ICRES-RNAs and cell lysates were collected at 36 h.p.e. Expression of nsP3 and capsid was analysed by western blot. Multiple western blots were quantified using a LiCor Odyssey Sa fluorescence imager and the graph on the right shows the ratio of capsid to nsP3 expression. (B) C2C12 cells were electroporated with the indicated ICRES RNA, cellular RNA synthesis was inhibited by actinomycin D and nascent viral RNAs were labelled with [^{3}H]-uridine. The graph on the right shows the ratio of gRNA to sgRNA. (C) The same RNAs were fractionated on a sucrose gradient and [^{3}H]-labelled RNAs were detected by scintillation counting of individual fractions.

Figure 8. AUD RNA-binding activity to different viral RNAs. (A) *E. coli* expressed AUD (wildtype and P247A/V248A) analysed by SDS-PAGE and Coomassie blue staining. (B) Circular Dichroism analysis of purified AUD. (C) Schematic of the CHIKV genome showing the location of the various RNAs used in subsequent filter binding analysis. (D-F) Filter binding analysis of the interaction between AUD and the indicated RNA species. Purified AUD at the indicated concentrations was incubated with radiolabelled RNA (1 nM) before application to a slot blot apparatus, filtering through nitrocellulose (protein-RNA complex) and Hybond-N (free RNA) membranes, and visualization by phosphoimaging. The negative control is wildtype AUD with an 80-mer aptamer raised against the foot-and-mouth disease virus 3D RNA-dependent RNA polymerase [18]. The percentage of RNA bound to the nitrocellulose membrane was quantified and plotted as a function of the AUD concentration. The data was fitted to a hyperbolic equation. (G) Endpoint (% of total RNA bound) and K_d values derived from the graphs in (D-F).

Figure 9. CHIKV genome RNA association with nsP3 during virus replication. C2C12 cells were electroporated with ICRES nsP3-TST RNAs. Cell lysates were collected at 60 h.p.e. and nsP3-TST was precipitated with Streptactin-sepharose beads. Bound proteins were subjected to western blotting (A) and co-precipitated RNAs were extracted by TRIzol and quantified by qRT-PCR (B). The ratio of gRNA to nsP3 is depicted graphically (C).

Figure 10. Fluorescence analysis of nsP3, capsid and dsRNA distribution during infection of C2C12 cells with wildtype CHIKV. C2C12 cells were electroporated with ICRES-nsP3-ZsGreen-CHIKV RNA. Cells were fixed at the indicated time points post-infection and stained with antibodies to capsid protein (white) and dsRNA (red). Green: nsP3-ZsGreen fusion, blue: nuclear DAPI counterstain. The scale bars are 5 μm and 1 μm, respectively.

Figure 11. Fluorescence analysis of nsP3, capsid and dsRNA distribution during infection of C2C12 cells with P247A/V248A mutant CHIKV. C2C12 cells were electroporated with ICRES-nsP3-ZsGreen-CHIKV-P247A/V248A RNA. Cells were fixed at the indicated time points post-infection and stained with antibodies to capsid protein (white) and dsRNA (red). Green: nsP3-ZsGreen fusion, blue: nuclear DAPI counterstain. The scale bars are 5 μm and 1 μm, respectively.

Figure 12. Co-localisation of nsP3 with capsid protein and dsRNA during virus replication. (A) Quantification of the percentages of nsP3 colocalised with dsRNA. Co-localisation of nsP3 with dsRNA. (B) Quantification of the percentages of nsP3 colocalised with capsid. Co-localisation analysis (green blocks) were determined from 5 cells for each construct using Fiji. ****, * indicates significant difference ($P<0.0001$, $P<0.0479$, respectively) from the results for WT.

Supporting Information

Supplementary Figure 1. Alignment of AUD amino acid sequences (nsP3 residues 210-276) and ribbon structure of Sindbis virus AUD showing location of mutated residues (PDB ID code 4GUA).

Supplementary Figure 2: Sequence analysis of virus passage P0.

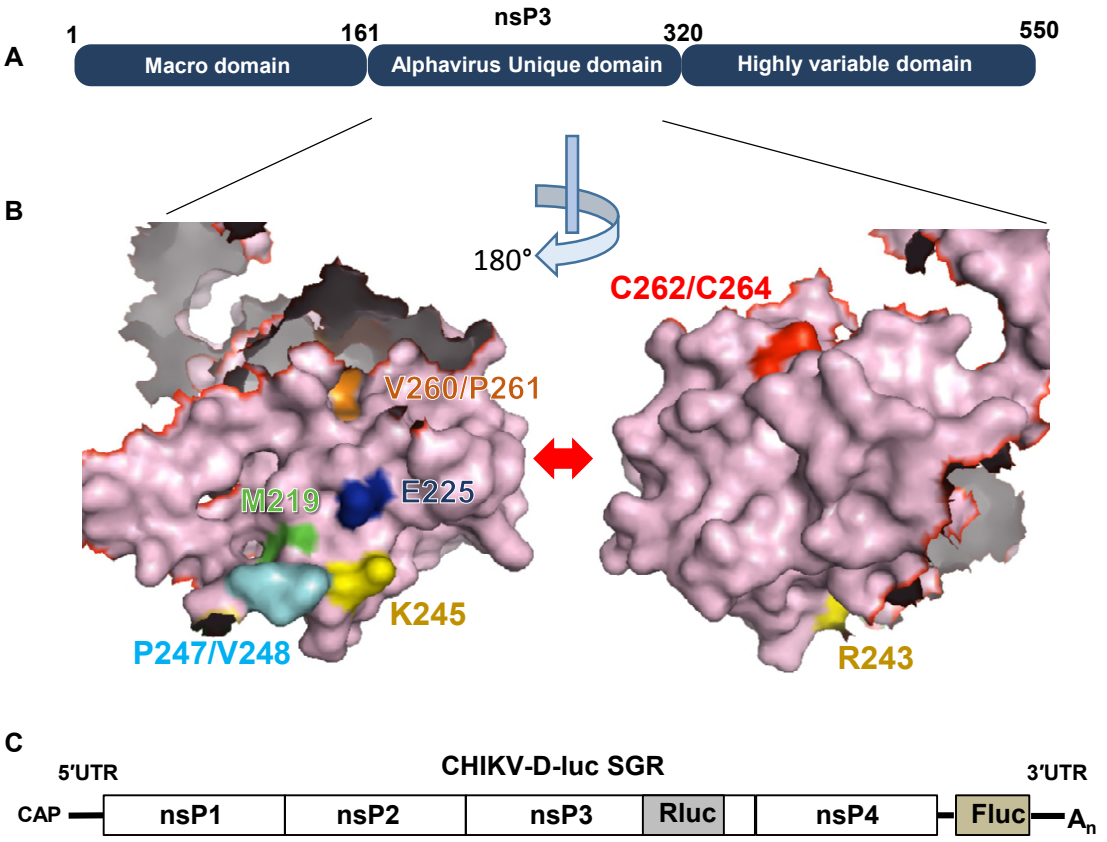
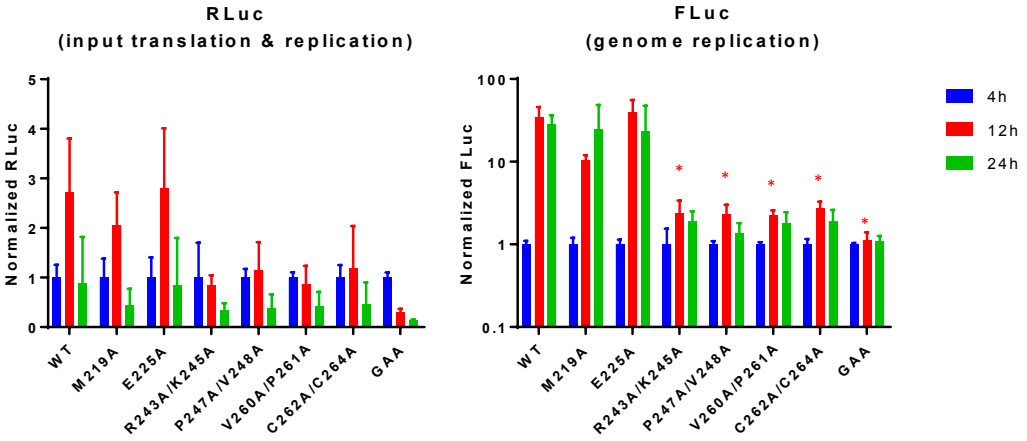
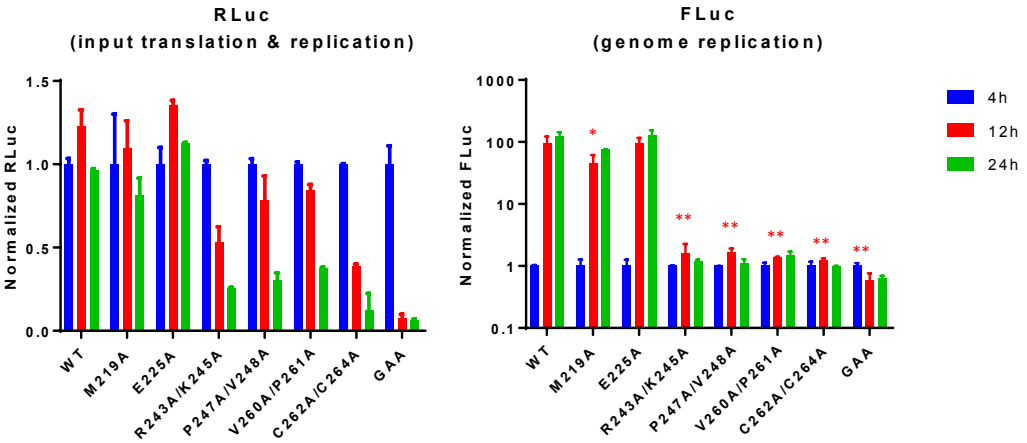

A
nsP3
1
161
320
550
Macro domain
Alphavirus Unique domain
Highly variable domain
B
180°
C262/C264
V260/P261
M219
E225
K245
P247/V248
R243
C
CHIKV-D-luc SGR
5'UTR
3'UTR
CAP
nsP1
nsP2
nsP3
Rluc
nsP4
Fluc
An

Figure 1

A Huh7



B Huh7.5



C RD

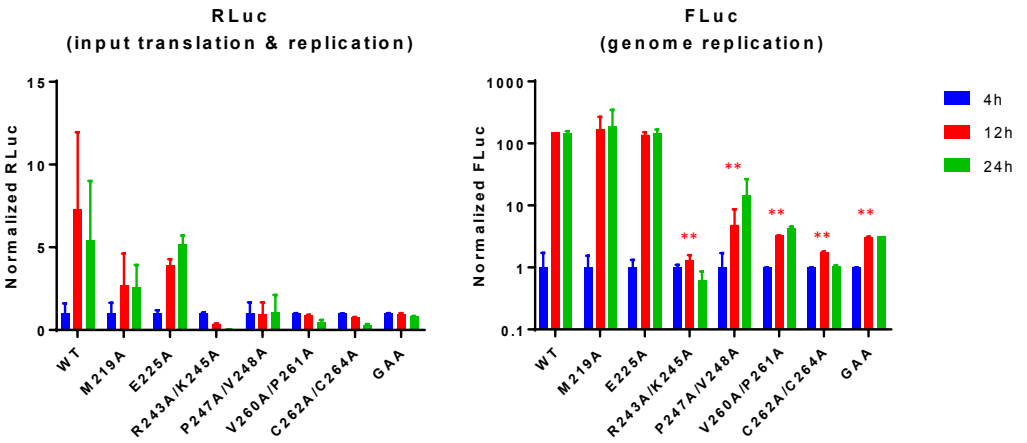
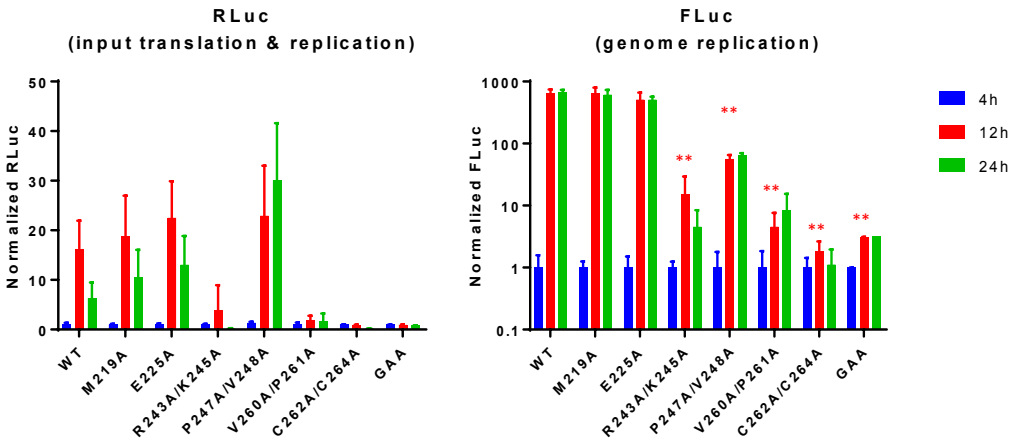


Figure 2

A C2C12



B BHK

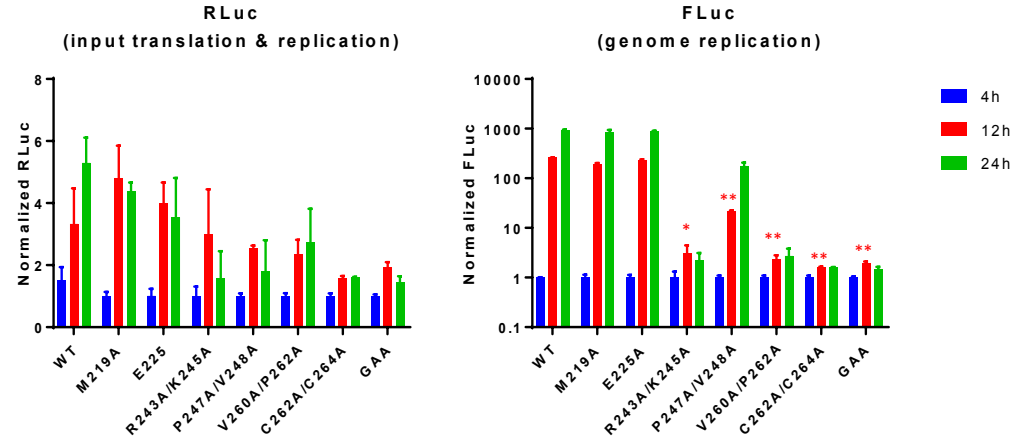


Figure 3

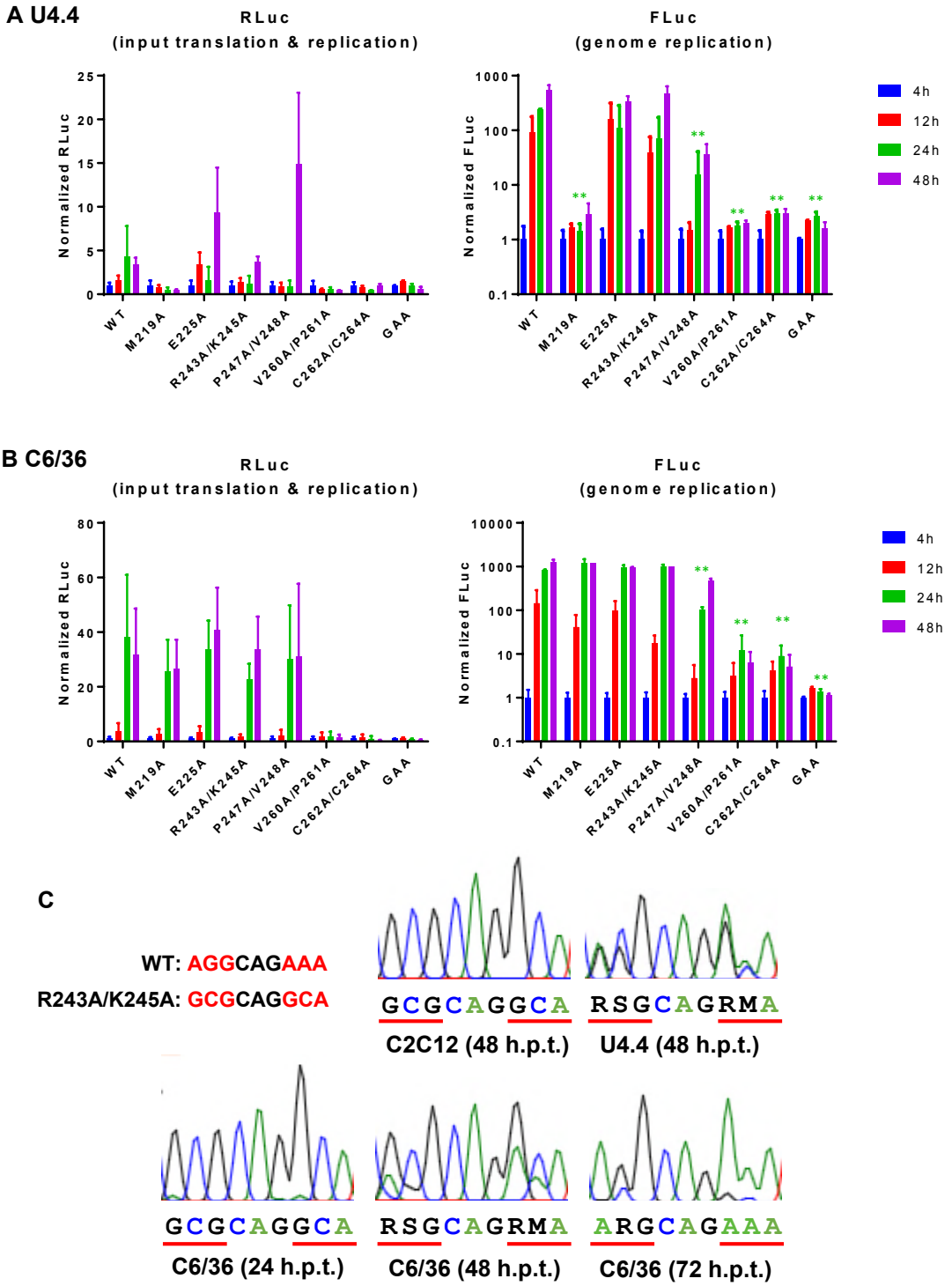

A U4.4
RLuc
(input translation & replication)
FLuc
(genome replication)
Normalized RLuc
Normalized FLuc
4h
12h
24h
48h
WT
M219A
E225A
R243A/K245A
P247A/V248A
V260A/P261A
C262A/C264A
GAA
B C6/36
C
WT: AGGCAGAAA
R243A/K245A: GCGCAGGCA
GCGCAGGCA
RSGCAGRMA
C2C12 (48 h.p.t.)
U4.4 (48 h.p.t.)
GCGCAGGCA
RSGCAGRMA
ARGCAGAAA
C6/36 (24 h.p.t.)
C6/36 (48 h.p.t.)
C6/36 (72 h.p.t.)

Figure 4

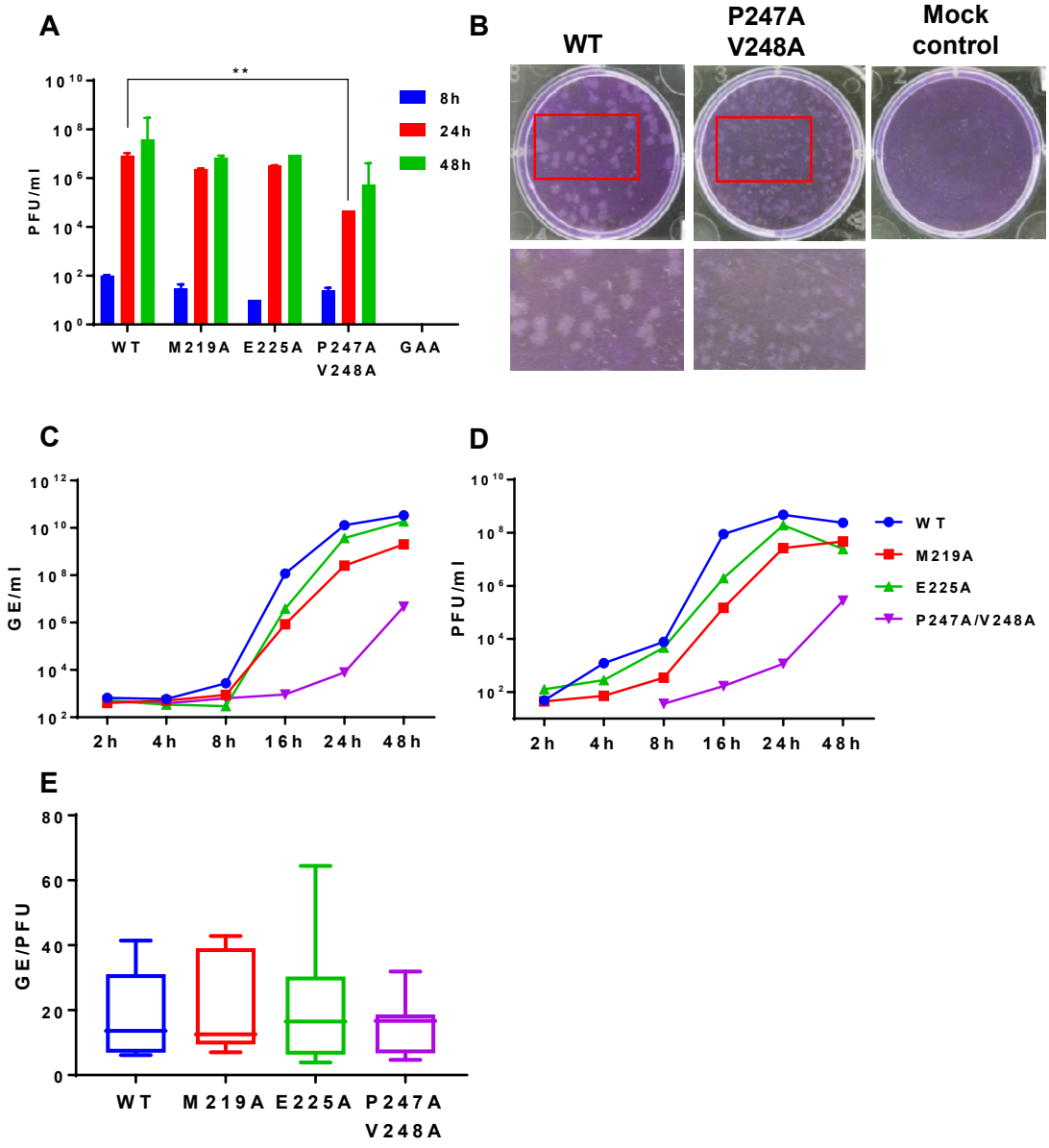

A
B
C
D
E
WT
P247A
V248A
Mock
control
PFU/ml
8h
24h
48h
**
WT
M219A
E225A
P247A
V248A
GAA
GE/ml
2h
4h
8h
16h
24h
48h
WT
M219A
E225A
P247A/V248A
GE/PFU

Figure 5

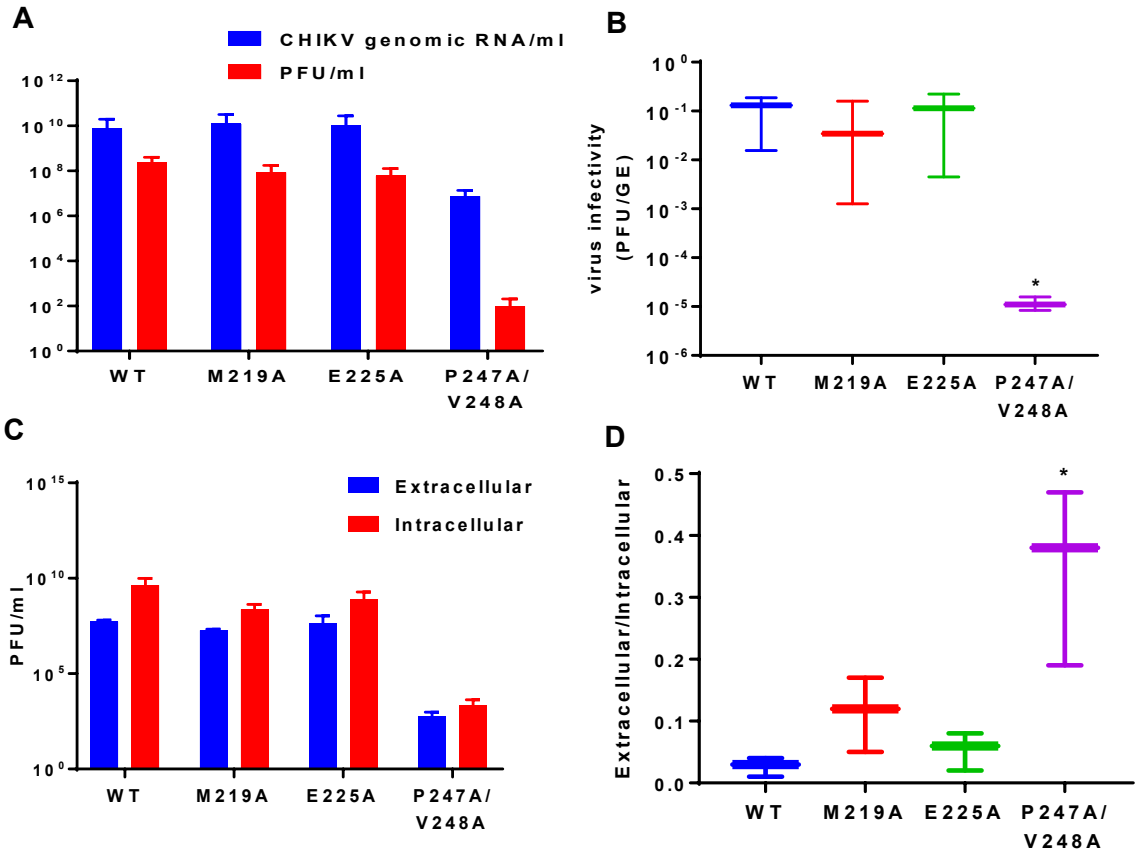

A
CHIKV genomic RNA/ml
PFU/ml
10 12
10 10
10 8
10 6
10 4
10 2
10 0
WT
M219A
E225A
P247A/
V248A
B
virus infectivity
(PFU/GE)
10 0
10 -1
10 -2
10 -3
10 -4
10 -5
10 -6
*
WT
M219A
E225A
P247A/
V248A
C
Extracellular
Intracellular
PFU/ml
10 15
10 10
10 5
10 0
WT
M219A
E225A
P247A/
V248A
D
Extracellular/Intracellular
0.5
0.4
0.3
0.2
0.1
0.0
*
WT
M219A
E225A
P247A/
V248A

Figure 6

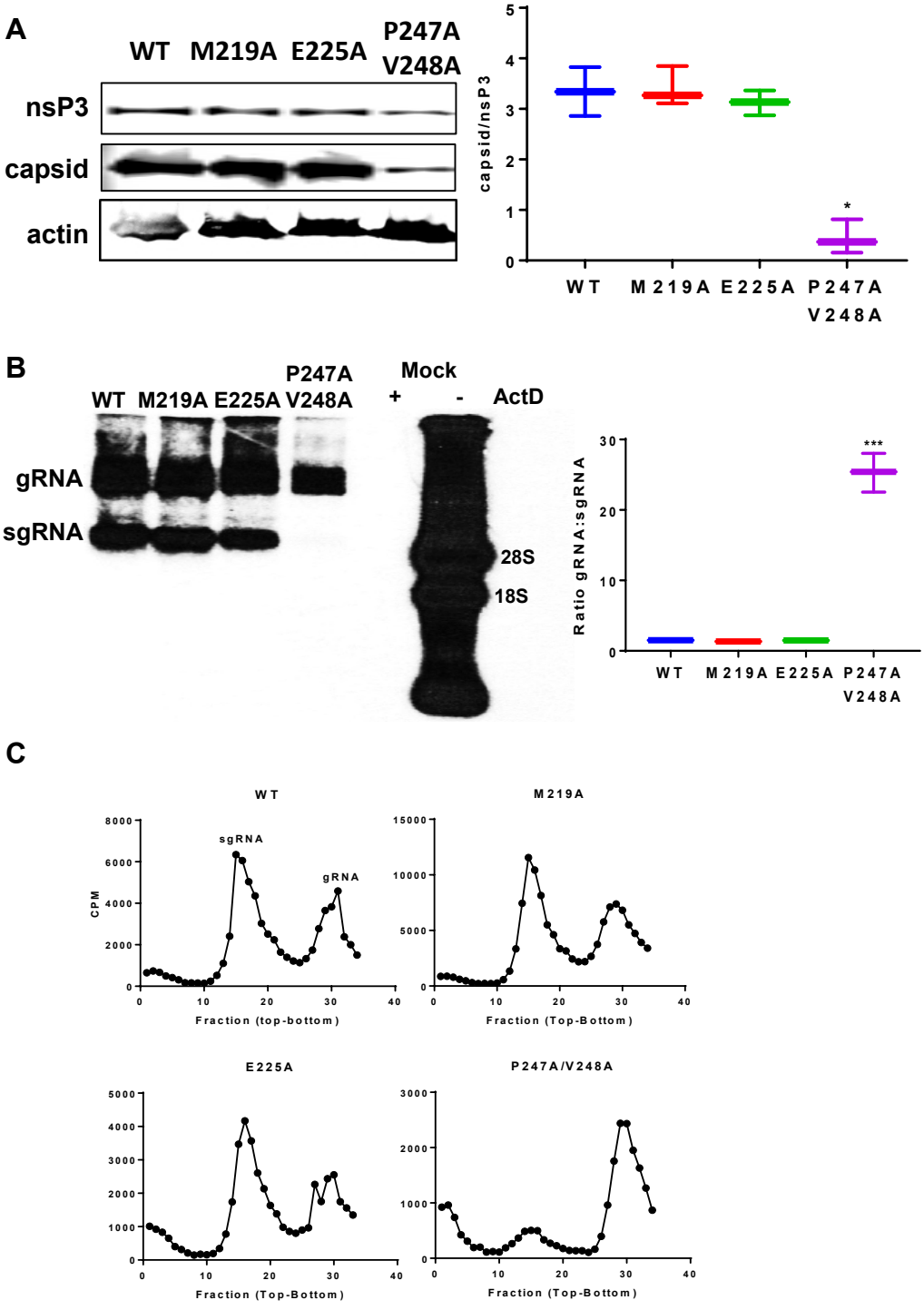
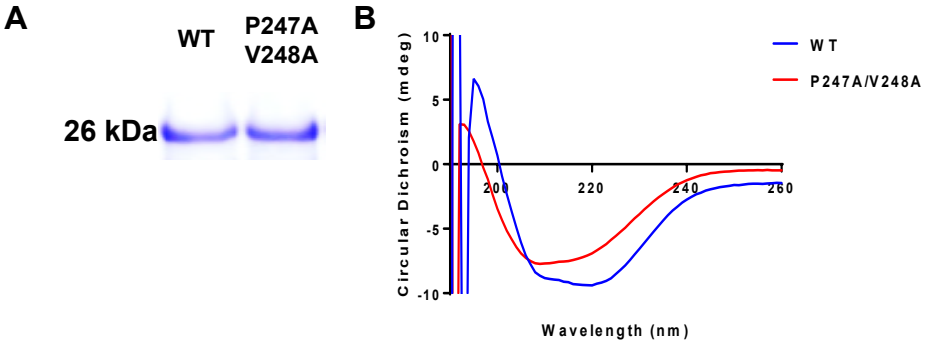
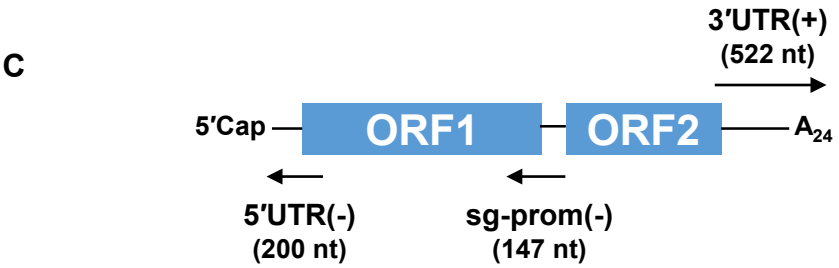
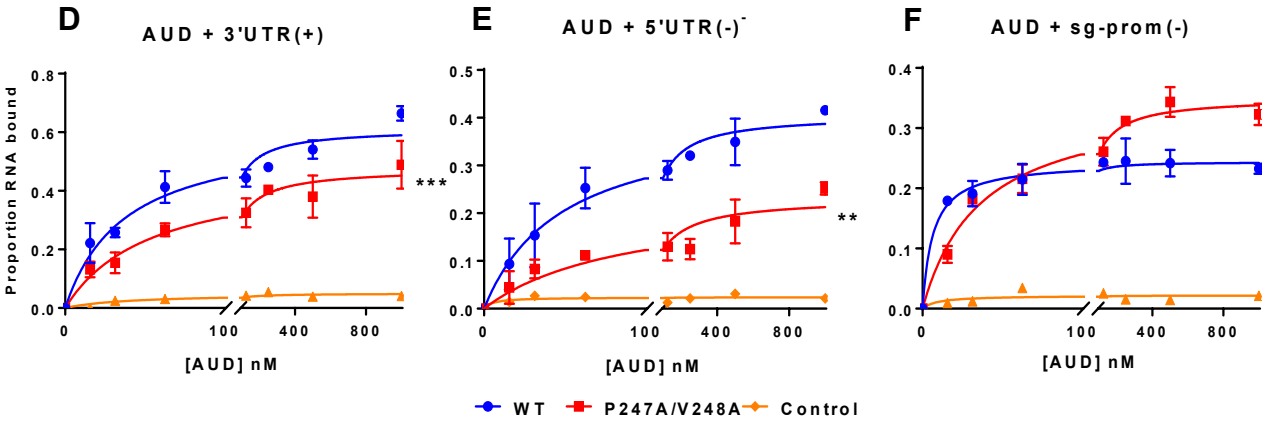

A
WT
M219A
E225A
P247A
V248A
nsP3
capsid
actin
capsid/nsP3
WT
M219A
E225A
P247A
V248A
B
WT
M219A
E225A
P247A
V248A
Mock
+
-
ActD
gRNA
sgRNA
28S
18S
Ratio gRNA:sgRNA
C
WT
sgRNA
gRNA
CPM
Fraction (top-bottom)
M219A
Fraction (Top-Bottom)
E225A
Fraction (Top-Bottom)
P247A/V248A
Fraction (Top-Bottom)

Figure 7







G

	3′UTR(+)		5′RNA(-)		sg-prom(-)	
	WT	P247A V248A	WT	P247A V248A	WT	P247A V248A
Endpoint % (mean ± SE)	61±2	48±2	41±2	23±2	24±1	35±1
K_d nM (mean ± SE)	37.2±5.8	53.8±10.0	48.4±8.3	87.9±25.4	6.5±1.4	36.4±2.8

Figure 8

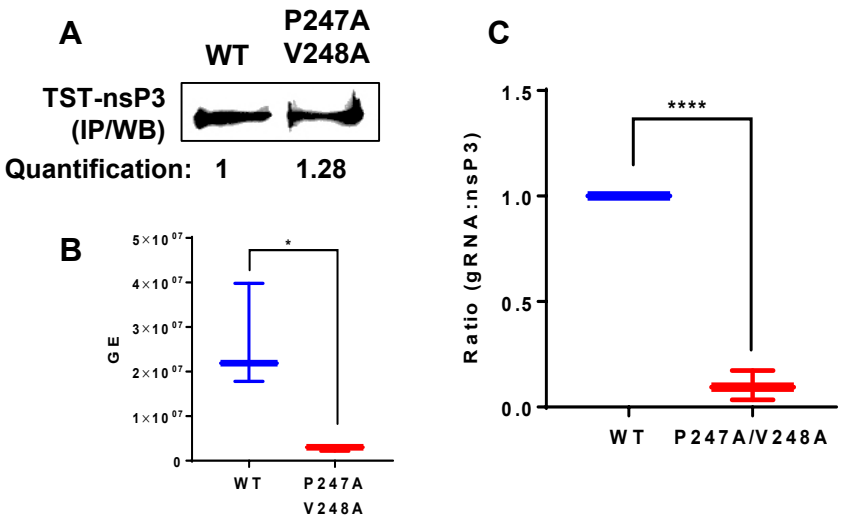
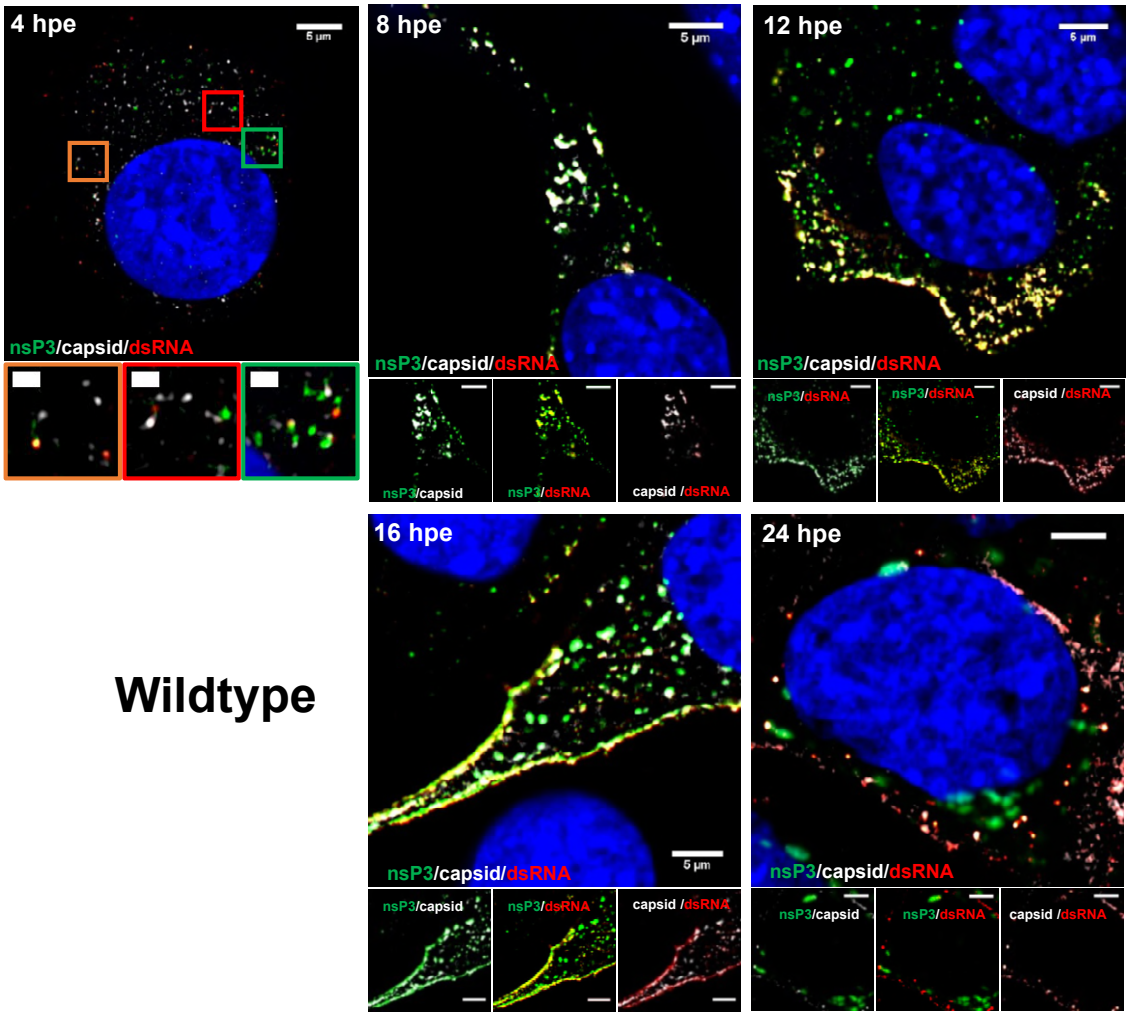

A
WT
P247A
V248A
TST-nsP3
(IP/WB)
Quantification: 1 1.28
B
GE
WT
P247A
V248A
C
Ratio (gRNA:nsP3)
WT
P247A/V248A
****

Figure 9



Wildtype

Figure 10

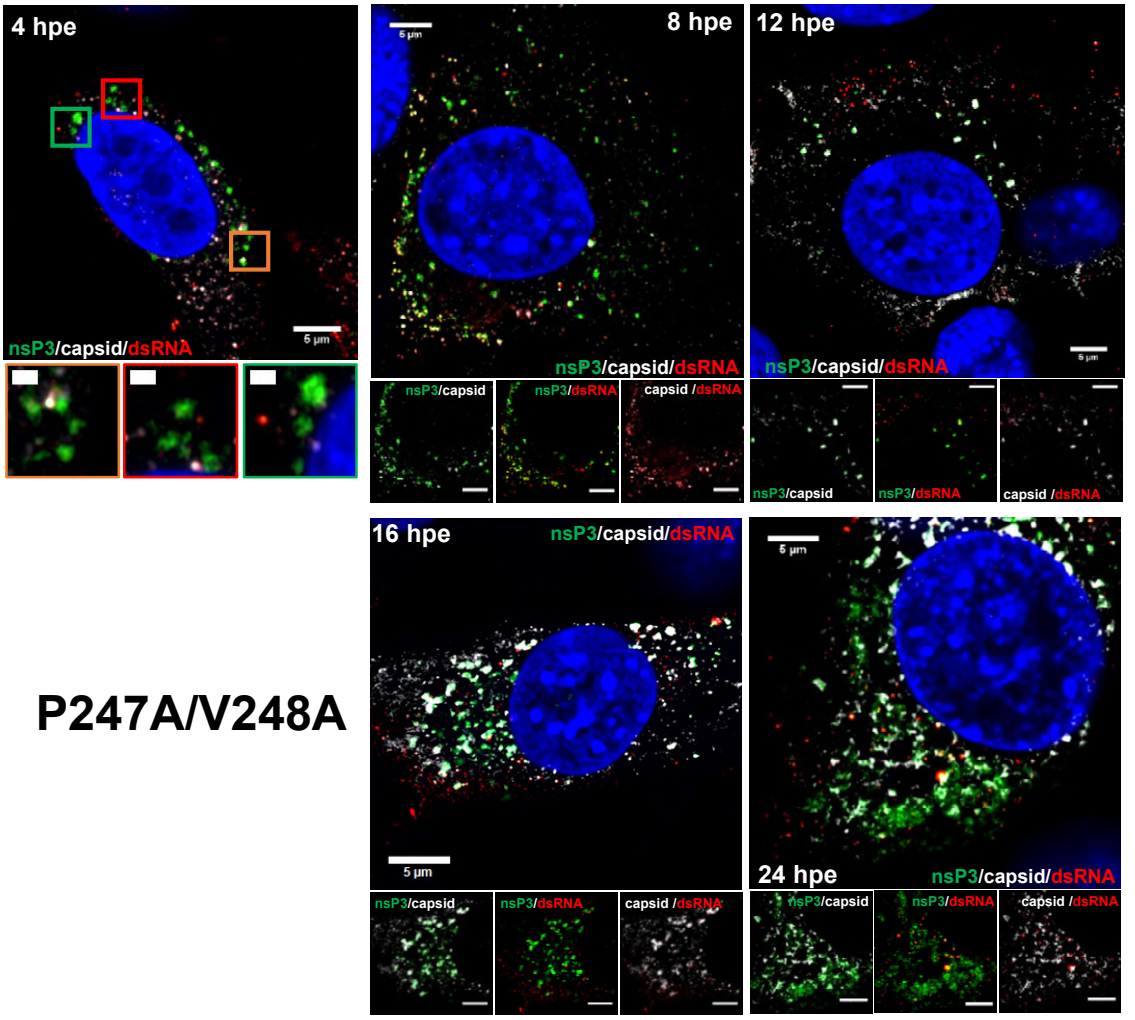

4 hpe
nsP3/capsid/dsRNA
5 µm
8 hpe
nsP3/capsid/dsRNA
nsP3/capsid
nsP3/dsRNA
capsid /dsRNA
12 hpe
nsP3/capsid/dsRNA
nsP3/capsid
nsP3/dsRNA
capsid /dsRNA
P247A/V248A
16 hpe
nsP3/capsid/dsRNA
5 µm
nsP3/capsid
nsP3/dsRNA
capsid /dsRNA
24 hpe
nsP3/capsid/dsRNA
nsP3/capsid
nsP3/dsRNA
capsid /dsRNA

Figure 11

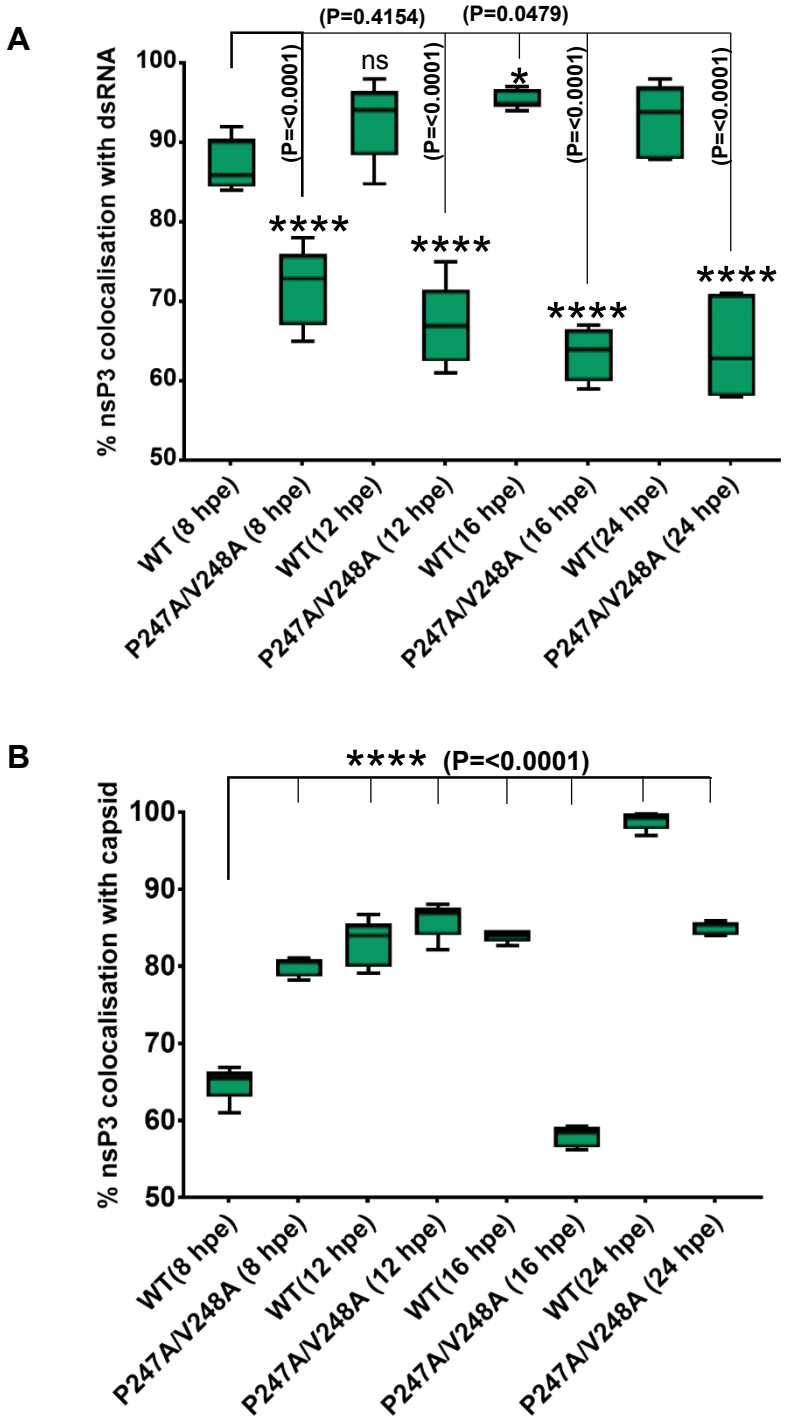

A
(P=0.4154)
(P=0.0479)
(P=<0.0001)
ns
*
****
% nsP3 colocalisation with dsRNA
100
90
80
70
60
50
WT (8 hpe)
P247A/V248A (8 hpe)
WT(12 hpe)
P247A/V248A (12 hpe)
WT(16 hpe)
P247A/V248A (16 hpe)
WT(24 hpe)
P247A/V248A (24 hpe)
B
**** (P=<0.0001)
% nsP3 colocalisation with capsid
100
90
80
70
60
50
WT(8 hpe)
P247A/V248A (8 hpe)
WT(12 hpe)
P247A/V248A (12 hpe)
WT(16 hpe)
P247A/V248A (16 hpe)
WT(24 hpe)
P247A/V248A (24 hpe)

Figure 12