Neurodevelopmental protein Musashi 1 interacts with the Zika genome and promotes viral replication

Pavithra L. Chavali[1,2,†], Lovorka Stojic[1,†], Luke W. Meredith[3], Nimesh Joseph[1], Michael S. Nahorski[4], Thomas J. Sanford[3], Trevor R. Sweeney[3], Ben A. Krishna[5], Myra Hosmillo[3], Andrew E. Firth[3], Richard Bayliss[6], Carlo L. Marcelis[7], Susan Lindsay[8], Ian Goodfellow[3], C. Geoffrey Woods[4] and Fanni Gergely[1*]

[1] Cancer Research UK Cambridge Institute, Li Ka Shing Centre, University of Cambridge, Robinson Way, Cambridge CB2 0RE, UK

[2] Current address: MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK

[3] Division of Virology, Department of Pathology, University of Cambridge, Addenbrookes Hospital, Cambridge, UK.

[4] Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge, CB2 0XY, UK.

[5] Department of Medicine, University of Cambridge, Hills Road, Cambridge, CB2 2QQ, UK.

[6] Faculty of Biological Sciences, Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK.

[7] Department of Human Genetics, Radboud University Medical Centre, Nijmegen, The Netherlands.

[8] Institute of Genetic Medicine, Newcastle University, International Centre for Life , Central Parkway , Newcastle upon Tyne , NE1 3BZ, UK

† these authors contributed equally

* Corresponding author. E-mail:
Fanni.Gergely@cruk.cam.ac.uk

Abstract:

A recent outbreak of Zika virus in Brazil has led to a simultaneous increase in reports of neonatal microcephaly. Zika targets cerebral neural precursors, a cell population essential for cortical development, but the cause of this neurotropism remains obscure. Here we report that the neural RNA-binding protein Musashi-1 (MSI1) interacts with the Zika genome and enables viral replication. Zika infection disrupts the binding of MSI1 to its endogenous targets, thereby deregulating expression of factors implicated in neural stem cell function. We further show that MSI1 is highly expressed in neural progenitors of the human embryonic brain, and is mutated in individuals with autosomal recessive primary microcephaly. Selective MSI1 expression in neural precursors could therefore explain the exceptional vulnerability of these cells to Zika infection.

Main text:

Zika virus (ZIKV) recently emerged as a major public health risk because of its devastating effect on fetal neurodevelopment (*1-3*). ZIKV was first isolated in Uganda in 1947, and the virus subsequently spread through Asia, and from there to the Americas (*4*). A causal link between ZIKV infection and congenital brain malformations became apparent in 2016 following an outbreak in Brazil (*1*). Brazilian ZIKV is closely related to the Asian-lineage strain, which affected New Caledonia and French Polynesia, where cases of microcephaly were reported retrospectively (*5*).

Intrauterine infections can impair neurodevelopment (*6*), but ZIKV is highly neurotropic and interferes specifically with fetal brain development causing microcephaly, cortical malformations and intracranial calcifications (*7-10*). We hypothesized that the single-stranded RNA flavivirus ZIKV may hijack RNA-binding factors present in the developing central nervous system (*11*). Host RNA-binding proteins are known to interact with untranslated regions (UTRs) to regulate replication, translation and stabilization of viral genomes (*11*). In *silico* analysis of the genomic RNA of the Brazilian ZIKV strain, PE243, revealed three consensus binding sites in the 3'UTR for the highly conserved Musashi family of RNA binding proteins, Musashi-1 (MSI1) and Musashi-2 (MSI2), both important translational regulators in stem cells (*12-15*). Two sites were conserved between PE243 and the Ugandan MR766 strains (Sites 1, 2), whereas the third (Site 3) was found only in the Asian-lineage strains including PE243 (Fig. 1A; Fig. S1A, B). By mapping these sites onto a predicted secondary structure of ZIKV 3'UTR, we found all the three to be present on stem-loop structures, which are considered optimal for MSI binding (*16, 17*). Moreover, a recent study revealed nucleotide substitutions proximal to Sites 1 and 2 in the Asian-lineage strains, which could indicate positive selection for MSI1 binding during ZIKV evolution (*18*).

To address if the Musashi proteins interacted with ZIKV, we first tested their binding to ZIKV 3'UTR. RNA pull-downs identified binding of MSI1, but not MSI2, to the 3'UTR of PE243 (Fig. 1B) (*15*). Mutating the three consensus MSI1 sites in the 3'UTR of PE243 significantly weakened this interaction (Fig. 1C, D; Fig. S1C). We also confirmed binding between MSI1 and the 3'UTR of MR766 (Fig. 1C). To investigate whether MSI1 also binds ZIKV 3'UTR *in vivo,* ultraviolet (UV) crosslinking and RNA immunoprecipitation (CLIP) was performed from PE243-infected U-251 glioblastoma cells, revealing a robust direct interaction between MSI1 and PE243 ZIKV RNA (Fig. 1E). Consistently, in ZIKV-infected cells, MSI1 co-localised with double-stranded RNA, a viral replication intermediate, as visualised by confocal and STED super-resolution microscopy (Fig. 1F, G). These data confirm an interaction between MSI1 and ZIKV RNA, which is at least in part mediated by the 3'UTR of the virus.

To investigate whether MSI1 had an impact on the life cycle of ZIKV, we used RNA interference to deplete the protein in U-251 glioblastoma, SK-N-BE2c neuroblastoma and H9 derived neural stem cells (NSC) and performed PE243 viral infections. In all three cell types MSI1 depletion led to a marked reduction in viral RNA levels (Fig. 2A, B). We then generated MSI1 knockouts (KO) in U-251 cells by CRISPR/Cas9-mediated targeting of exons 8 or 6 of *MSI1* (KO1 and KO2, respectively; Fig. 2C; Fig. S2). Control cells were obtained through clonal expansion of cells transfected with Cas9 alone. By measuring viral RNA at different times following PE243 infection, a marked reduction of viral load was seen in KO1 and KO2 cells at 24 and 48 hours (Fig. 2D). Whereas extensive cell death precluded RNA analysis in the controls at 72 hours, viral RNA was comparable between the 48- and 72-hour timepoints in the KO cells. Consistently, levels of the viral dsRNA and flavivirus E protein as well as the infectious titre were reduced in the KOs (Fig. 2E; Fig S3). Because MSI2 levels were similar between control and KO cells (Fig. 2C), MSI1 and MSI2 are unlikely to have complete functional redundancy in ZIKV replication. Replication of the MR766 strain was also impaired in the KO cells (Fig. 2F). In summary, we identify MSI1 as an important factor for ZIKV replication, both in primary and transformed neural cell lines.

Since there was no discernible difference between ZIKV binding and entry into control and KO cells (Fig. 2G, H), we asked if MSI1 could regulate translation through ZIKV UTRs. To this end, luciferase RNA flanked by the 5’ and 3’UTRs of PE243, was transfected with increasing amounts of MSI1 into HEK293T cells, which do not normally express MSI1 (Fig. S4). We observed a modest MSI1-driven increase in luciferase expression. The ability of MSI1 to promote ZIKV UTR-driven translation *in vitro* raises the possibility that it performs a similar function *in vivo*. Alternatively, MSI1 might stabilise the viral RNA genome and/or regulate its cyclization or synthesis. In addition, given the pleiotropic roles of MSI1 in cellular pathways, it is plausible that MSI1-dependent regulation of gene expression contributes to the ZIKV life cycle (*19*). However, MSI1 is unlikely to act through general pro- or anti-viral pathways as infection with H1N1 influenza virus was unaffected by MSI1 expression levels (Fig. S5). In line with published work, we also find that MSI1 KO cells exhibited defective migration, increased doubling time and cell cycle delay (Fig. S6) (*17, 20, 21*). Because ZIKV replication requires cyclin-dependent kinase activity, such pro-proliferative effects exerted by MSI1 might contribute to virus production (*22*). Nevertheless, the direct interaction between MSI1 and the ZIKV genome is consistent with the hypothesis that the protein promotes some aspect of the viral life cycle.

ZIKV predominantly infects neural progenitors in human fetal brain. We find MSI1 to be highly enriched in neural precursors of the ventricular and subventricular zones of the human embryonic brain, but absent from mature neurons (Fig. 3A; Fig. S7). Owing to its high levels in neural progenitors, and its ability to stimulate ZIKV replication, MSI1 could be instrumental to ZIKV-induced cytopathicity in the fetal brain. In addition, MSI1 is required for neurodevelopment in both invertebrates and vertebrates, with MSI1-depleted zebrafish displaying microcephaly, and mutant mice exhibiting a thin cerebral cortex and reduced number of mature neural cell types among other morphological brain abnormalities (*14, 19, 23-25*). We have identified a consanguineous Turkish family in which two siblings displayed clinical features suggestive of autosomal primary microcephaly (MCPH), a condition associated with a significant reduction in cerebral cortex size, but a structurally normal brain (Fig. S8A, B) (*26, 27*). Exome sequencing uncovered potentially deleterious homozygous mutations in *MSI1*, *ACACB*, *DKK4* and *DTX3L* (Fig. S8C-F, Tables S1 and S2). Of these only MSI1 is known to have neural functions, but because mutations were present in four genes, the p. Ala184Val point mutation in MSI1 may not be the sole cause of MCPH in these individuals (referred to as MSI1^{A184V}). Nevertheless, we have three lines of evidence showing that A184V mutant MSI1 is functionally impaired. First, MSI1^{A184V} patient cells exhibit premature chromosome condensation (PCC), the same phenotype as MSI1-deficient glioblastoma cells. Second, we show that the A184V mutation impedes RNA binding of MSI1, leading to deregulated expression of its endogenous targets. Third, we find that the A184V mutant MSI1 is unable to support ZIKV replication.

The PCC phenotype seen in MSI1^{A184V} patient cells has been previously described in cells deficient of the MCPH-associated protein, Microcephalin (MCPH1) (Fig. S8G) (*28, 29*). Since the *MCPH1* locus is unaffected in MSI1^{A184V} patients, we speculated that MSI1 could control chromosome condensation by regulating *MCPH1* expression. To determine if MCPH1 was a MSI1 target, we performed RNA immunoprecipitation under native conditions. In addition to its known targets *NUMB, p21^{WAF}* and the MCPH-associated gene, *CDK6/MCPH12,* MSI1 co-precipitated with two isoforms of *MCPH1* (*MCPH1_S* and *MCPH_L*) despite their divergent 3' UTRs, but did not interact with other *MCPH* genes tested (Fig. S9A, B) (*12, 30, 31*). CLIP and RNA mobility shift assays suggest a direct interaction between MSI1 and *MCPH1_L* (Fig. 3B, C). MSI1 can act as a translational suppressor (i.e. for *NUMB* and *p21^{WAF1}*) or activator (i.e. for *CDK6*) (*12, 21, 30*). Consistent with a role for MSI1 in translational activation of MCPH1, we observed low MCPH1 protein levels in MSI1^{A184V} patient and MSI1-deficient U-251 cells and a reduction in polysome-associated *MCPH1* transcripts in KO cells (Fig. 3D, E; Figs. S9C, S10). Given that MSI1 interacts with ZIKV RNA, and viral RNA is abundant in the infected cell, it could compete with endogenous targets for binding MSI1. Indeed, upon ZIKV infection of U-251 cells, we observed a marked reduction in the interaction between MSI1 and its target RNAs including *MCPH1* and *NUMB*, accompanied by changes in their protein levels that mirrored those of MSI1 KO cells (Fig. 3F,G).

MSI1 interacts with target transcripts via its two RNA recognition motifs (RRM) (*19, 30, 32*). NMR studies show that the conserved Ala184 within the RRM2 is an RNA-binding residue (*33*). Modeling based on the crystal structure of the RNA-binding protein HRP1 with RNA, indicates that the A184V mutation impairs the MSI1-RNA interaction (Fig. 4A). To evaluate the effect of A184V on MSI1 function in cells, transgenes encoding wild-type and A184V MSI1 (MsiWT and MsiAV, respectively) were randomly integrated into KO1 or KO2 U-251 cells and single clones were isolated (KO1/2-MsiWT or KO1/2-MsiAV). MSI1 protein expression in these clones was quantified with respect to endogenous protein levels in parental cells (Fig. 4B). When compared to KO2-MsiWT cells, MSI1 RIP recovered 2-4 fold fewer target transcripts from KO2-MsiAV, indicative of reduced RNA binding by the A184V mutant (Fig. 4C). NUMB and MCPH1 protein levels changed accordingly (Fig. 4B). These results were recapitulated in HEK293T cells expressing Msi1WT or Msi1AV (Fig. S11). MSI1^{A184V}, MSI1 KO, KO1/2-MsiAV and MSI1-depleted cells all exhibited suboptimal MCPH1 protein levels and PCC, prompting us to investigate if a functional link existed between these phenotypes (Fig. S9C-F). MCPH1 overexpression reduced PCC frequency in MSI1-depleted cells, thereby confirming a role for MSI1 in chromosome condensation via translational control of MCPH1 (Fig. S8G, H). Therefore, the A184V mutation impairs binding of MSI1 to RNA, which leads to reduced MCPH1 expression and a concomitant increase in PCC frequency. Remarkably, defective chromosome condensation has been recently found to cause MCPH (*34*).

Given that the A184V mutation impedes binding of MSI1 to RNA, we next probed the effect of A184V mutant MSI1 on ZIKV replication. To this end, viral RNA levels and cell viability were assayed in KO1/2, KO1/2-MsiWT and KO1/2-MsiAV U-251 cells infected with PE243. Complementation of KO1/2 cells with MsiWT increased both ZIKV RNA levels and cell death (Fig. 4D, E). By contrast, MsiAV was unable to support ZIKV replication, and showed minimal impact on cell viability. Additionally, in HEK293T cells, expression of MsiWT, but not MsiAV, increased viral RNA and cell death upon infection (Fig. 4F, G; Fig. S12). These findings also imply that MSI1 expression increases susceptibility of HEK293T cells to ZIKV infection (*35*). Furthermore, we noted an apparent dose-dependent effect of MSI1 on viral replication; those U-251 and HEK293T clones that express higher levels of MSI1 displayed greater viral RNA levels and increased cell death (Fig. 4B, D-G; Figs. S11 and S12).

Our study raises the question whether MSI1 could have functions in other flaviviruses. We have surveyed putative MSI1 binding sites in a number of flaviviruses by mapping the consensus (A/GU$_{(1-3)}$AG) onto predicted secondary structures of flaviviridae 3'UTRs obtained from a recent publication (Table S3) (*16*). Although several flaviviruses harbour consensus MSI1 sites within appropriate structural landscapes, whether MSI1 is relevant to the biology of these viruses remains to be established. Furthermore, while our data are consistent with a role for MSI1 in ZKV neurotropism and pathology, multiple factors must collude in ZIKV infection of the fetal brain, not least viral entry receptors that allow the virus to cross the placental barrier (*36*). Viruses such as human cytomegalovirus and Rubella can also access the developing fetal brain, but whether MSI1 contributes to their replication or pathogenesis is unknown and would require further study.

This work suggests that high MSI1 expression levels in neural precursors could be a key contributor to the fetal neurotropism exhibited by ZIKV (*2, 10, 37*) (Fig. S13). Intriguingly, MSI1 is also highly expressed in the retina and testis, other tissues deemed vulnerable to ZIKV infection (*38-41*). While our study provides new insight into the potential pathogenic mechanisms of ZIKV, further work will be required to determine if the modification or interference of the MSI1-ZIKV interaction results in neuronal attenuation of ZIKV.

References and notes

1. J. Mlakar *et al.*, Zika Virus Associated with Microcephaly. *The New England journal of medicine* **374**, 951-958 (2016).
2. A. S. Oliveira Melo *et al.*, Zika virus intrauterine infection causes fetal brain abnormality and microcephaly: tip of the iceberg? *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology* **47**, 6-7 (2016).
3. F. R. Cugola *et al.*, The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* **534**, 267-271 (2016).
4. S. N. Slavov, K. K. Otaguiri, S. Kashima, D. T. Covas, Overview of Zika virus (ZIKV) infection in regards to the Brazilian epidemic. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.]* **49**, e5420 (2016).
5. S. Cauchemez *et al.*, Association between Zika virus and microcephaly in French Polynesia, 2013-15: a retrospective study. *Lancet* **387**, 2125-2132 (2016).
6. K. M. Adams Waldorf, R. M. McAdams, Influence of infection during pregnancy on fetal development. *Reproduction* **146**, R151-162 (2013).
7. C. Li *et al.*, Zika Virus Disrupts Neural Progenitor Development and Leads to Microcephaly in Mice. *Cell stem cell* **19**, 120-126 (2016).
8. H. Li *et al.*, Zika Virus Infects Neural Progenitors in the Adult Mouse Brain and Alters Proliferation. *Cell stem cell*, (2016).
9. J. J. Miner, M. S. Diamond, Understanding How Zika Virus Enters and Infects Neural Target Cells. *Cell stem cell* **18**, 559-560 (2016).
10. J. Dang *et al.*, Zika Virus Depletes Neural Progenitors in Human Cerebral Organoids through Activation of the Innate Immune Receptor TLR3. *Cell stem cell* **19**, 258-265 (2016).
11. Z. Li, P. D. Nagy, Diverse roles of host RNA binding proteins in RNA virus replication. *RNA biology* **8**, 305-315 (2011).
12. T. Imai *et al.*, The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. *Molecular and cellular biology* **21**, 3888-3900 (2001).
13. C. L. Donald *et al.*, Full Genome Sequence and sfRNA Interferon Antagonist Activity of Zika Virus from Recife, Brazil. *PLoS neglected tropical diseases* **10**, e0005048 (2016).
14. S. Sakakibara *et al.*, Mouse-Musashi-1, a neural RNA-binding protein highly enriched in the mammalian CNS stem cell. *Developmental biology* **176**, 230-242 (1996).
15. S. Sakakibara, Y. Nakamura, H. Satoh, H. Okano, Rna-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **21**, 8091-8107 (2001).
16. S. M. Villordo, C. V. Filomatori, I. Sanchez-Vargas, C. D. Blair, A. V. Gamarnik, Dengue virus RNA structure specialization facilitates host adaptation. *PLoS Pathog* **11**, e1004604 (2015).
17. P. J. Uren *et al.*, RNA-Binding Protein Musashi1 Is a Central Regulator of Adhesion Pathways in Glioblastoma. *Molecular and cellular biology* **35**, 2965-2978 (2015).
18. Z. A. Klase *et al.*, Zika Fetal Neuropathogenesis: Etiology of a Viral Syndrome. *PLoS neglected tropical diseases* **10**, e0004877 (2016).
19. R. G. Fox, F. D. Park, C. S. Koechlein, M. Kritzik, T. Reya, Musashi signaling in stem cells and cancer. *Annual review of cell and developmental biology* **31**, 249-267 (2015).
20. P. R. de Araujo *et al.*, Musashi1 Impacts Radio-Resistance in Glioblastoma by Controlling DNA-Protein Kinase Catalytic Subunit. *The American journal of pathology* **186**, 2271-2278 (2016).
21. C. Battelli, G. N. Nikopoulos, J. G. Mitchell, J. M. Verdi, The RNA-binding protein Musashi-1 regulates neural development through the translational repression of p21WAF-1. *Molecular and cellular neurosciences* **31**, 85-96 (2006).
22. M. Xu *et al.*, Identification of small-molecule inhibitors of Zika virus infection and induced neural cell death via a drug repurposing screen. *Nature medicine* **22**, 1101-1107 (2016).
23. Y. Kaneko *et al.*, Musashi1: an evolutionally conserved marker for CNS progenitor cells including neural stem cells. *Dev Neurosci* **22**, 139-153 (2000).
24. S. Sakakibara *et al.*, RNA-binding protein Musashi family: roles for CNS stem cells and a subpopulation of ependymal cells revealed by targeted disruption and antisense ablation. *Proc Natl Acad Sci U S A* **99**, 15194-15199 (2002).
25. J. T. Eppig *et al.*, The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. *Nucleic acids research* **43**, D726-736 (2015).
26. E. C. Gilmore, C. A. Walsh, Genetic causes of microcephaly and lessons for neuronal development. *Wiley Interdiscip Rev Dev Biol* **2**, 461-478 (2013).
27. P. L. Chavali, M. Putz, F. Gergely, Small organelle, big responsibility: the role of centrosomes in development and disease. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **369**, (2014).
28. M. Trimborn *et al.*, Mutations in microcephalin cause aberrant regulation of chromosome condensation. *American journal of human genetics* **75**, 261-266 (2004).
29. A. P. Jackson *et al.*, Identification of microcephalin, a protein implicated in determining the size of the human brain. *American journal of human genetics* **71**, 136-142 (2002).

30. F. M. Cambuli *et al.*, A Mouse Model of Targeted Musashi1 Expression in Whole Intestinal Epithelium Suggests Regulatory Roles in Cell Cycle and Stemness. *Stem Cells* **33**, 3621-3634 (2015).
31. M. S. Hussain *et al.*, CDK6 associates with the centrosome during mitosis and is mutated in a large Pakistani family with primary microcephaly. *Human molecular genetics* **22**, 5199-5214 (2013).
32. H. Kawahara *et al.*, Neural RNA-binding protein Musashi1 inhibits translation initiation by competing with eIF4G for PABP. *The Journal of cell biology* **181**, 639-653 (2008).
33. T. Nagata *et al.*, Structure, backbone dynamics and interactions with RNA of the C-terminal RNA-binding domain of a mouse neural RNA-binding protein, Musashi1. *J Mol Biol* **287**, 315-330 (1999).
34. C. A. Martin *et al.*, Mutations in genes encoding condensin complex proteins cause microcephaly through decatenation failure at mitosis. *Genes & development* **30**, 2158-2172 (2016).
35. R. Hamel *et al.*, Biology of Zika Virus Infection in Human Skin Cells. *Journal of virology* **89**, 8880-8896 (2015).
36. A. S. Richard *et al.*, AXL-dependent infection of human fetal endothelial cells distinguishes Zika virus from other pathogenic flaviviruses. *Proc Natl Acad Sci U S A* **114**, 2024-2029 (2017).
37. J. B. Brault *et al.*, Comparative Analysis Between Flaviviruses Reveals Specific Neural Stem Cell Tropism for Zika Virus in the Mouse Developing Neocortex. *EBioMedicine* **10**, 71-76 (2016).
38. H. M. Lazear *et al.*, A Mouse Model of Zika Virus Pathogenesis. *Cell Host Microbe* **19**, 720-730 (2016).
39. S. Salinas *et al.*, Zika Virus Efficiently Replicates in Human Retinal Epithelium and Disturbs Its Permeability. *Journal of virology* **91**, (2017).
40. K. Susaki *et al.*, Musashi-1, an RNA-binding protein, is indispensable for survival of photoreceptors. *Experimental eye research* **88**, 347-355 (2009).
41. W. Ma *et al.*, Zika Virus Causes Testis Damage and Leads to Male Infertility in Mice. *Cell* **167**, 1511-1524 e1510 (2016).
42. C. G. Woods, E. M. Valente, J. Bond, E. Roberts, A new method for autozygosity mapping using single nucleotide polymorphisms (SNPs) and EXCLUDEAR. *J Med Genet* **41**, e101 (2004).
43. E. S. Lander, D. Botstein, Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567-1570 (1987).
44. R. F. Mueller, D. T. Bishop, Autozygosity mapping, complex consanguinity, and autosomal recessive disorders. *J Med Genet* **30**, 798-799 (1993).
45. S. Ogawa, Y. Nanya, G. Yamamoto, Genome-wide copy number analysis on GeneChip platform using copy number analyzer for affymetrix GeneChip 2.0 software. *Methods Mol Biol* **396**, 185-206 (2007).
46. C. G. Woods *et al.*, Quantification of homozygosity in consanguineous individuals with autosomal recessive disease. *American journal of human genetics* **78**, 889-896 (2006).
47. E. Bochukova *et al.*, A mutation in the thyroid hormone receptor alpha gene. *The New England journal of medicine* **366**, 243-249 (2012).
48. J. A. Miller *et al.*, Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206 (2014).
49. C. Genomes Project *et al.*, A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
50. W. Fu *et al.*, Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216-220 (2013).
51. V. Gandin *et al.*, Polysome fractionation and analysis of mammalian translatomes on a genome-wide scale. *Journal of visualized experiments : JoVE*, (2014).
52. R. A. Urbanowicz *et al.*, Human Adaptation of Ebola Virus during the West African Outbreak. *Cell* **167**, 1079-1087 e1075 (2016).
53. L. Stojic *et al.*, Transcriptional silencing of long noncoding RNA GNG12-AS1 uncouples its transcriptional and product-related functions. *Nat Commun* **7**, 10406 (2016).
54. J. H. Yoon, M. Gorospe, Cross-Linking Immunoprecipitation and qPCR (CLIP-qPCR) Analysis to Map Interactions Between Long Noncoding RNAs and RNA-Binding Proteins. *Methods Mol Biol* **1402**, 11-17 (2016).
55. J. M. Perez-Canadillas, Grabbing the message: structural basis of mRNA 3'UTR recognition by Hrp1. *The EMBO journal* **25**, 3167-3178 (2006).
56. A. Sali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815 (1993).

Acknowledgements

The authors are indebted to Alain Kohl (Centre for Virus Research, University of Glasgow) and Lindomar J. Pena and Rafael Oliveira de Freitas França, Fiocruz Recife, Pernambuco, Brazil, for the provision of PE243 ZIKV RNA used to generate the virus stock. We would like to thank and acknowledge Steve Lisgo for the expert provision of human embryonic histology sections through the Human Developmental Biology Resource (HDBR) at the University of Newcastle funded by a joint UK MRC/Wellcome Trust grant (099175/Z/12/Z). We would like to thank Leanna Smith for her assistance with homology modeling, Guillaume van Zande for his help and the patients' families for their participation. The National Research Ethics Service Committee, East of England - Cambridge Central, UK (C.G. Woods, REC 05/Q0108/402) approved the informed consent to enter the study. We would like to further thank Tanweer Hussain for help with polysome fractionations, and John Sinclair and Anna Git for technical advice. We thank KJ Patel and members of the Gergely lab for useful discussions and comments. We are grateful for expert help by the CRUK CI Core Facilities. I.G. and A. E. F are Wellcome Trust Senior Fellows. I.G. was supported by research grants 097997/Z/11/A and 097997/Z/11/Z, whereas A. E. F by grant 106207. M.S.N was funded by the Wellcome Trust (200183/Z/15/Z) and T. R. S is a Wellcome Trust Henry Dale Fellow (202471/Z/16/Z). This work was made possible by funding from Cancer Research UK C14303/A17197 to FG and C24461/A12772 to R.B. F.G. and C.G.W. acknowledge support from NIHR Cambridge Biomedical Research Centre, the University of Cambridge and Hutchison Whampoa Ltd. All data to understand and assess the conclusions of this research are available in the main paper and supplementary materials.

Fig. 1. MSI1 interacts directly with the ZIKV RNA genome.

A) Schematic diagram of PE243 ZIKV containing three putative MSI1 binding sites in its 3’UTR. Sites shared with MR766 are red, site unique to PE243 is blue.

B) RNA pull-down assays performed with the 3’UTR of PE243. Increasing concentrations of *in vitro* transcribed biotinylated PE243 RNA were incubated with U-251 cell extracts and RNA-protein complexes were captured on streptavidin beads. Representative western blots were probed with antibodies against MSI1 and the RNA binding proteins, Musashi-2 (MSI2) and hnRNP Q/R. Corresponding protein and RNA inputs are shown on right.

C) RNA pull-down assays performed with the wild type (WT) or triple mutant (Δ123) 3’UTR of PE243. Note that PE243-3’UTR_Δ123 lacks all three MSI1 binding sites depicted in Fig. 1A (see Fig. S1C for further details). Increasing concentrations of *in vitro* transcribed biotinylated RNA were incubated with U-251 cell extracts and RNA-protein complexes were captured on streptavidin beads. Representative western blots probed with antibody against MSI1 are shown together with corresponding protein and RNA inputs.

D) Densitometric analysis of MSI1 levels from western blots of RNA pull-down assays, an example for which is shown in Fig. 1C. The amount of MSI1 precipitated by PE243-3’UTR_Δ123 is expressed as a percentage of MSI1 precipitated by the same concentration of PE243-3’UTR_WT. n=3 biological replicates. P-values were obtained from Student’s t-test, unpaired, two-tailed: * p<0.05, *** p<0.0005. Bar charts depict mean±s.e.m.

E) CLIP analysis from mock- or ZIKV PE243-infected U-251 cells. Western blot shows immunoprecipitatations (IP) by rabbit IgG and MSI1 antibodies from mock and PE243-infected U-251 cells following UV crosslinking. Input (5%) represents whole cell extract. Western blot was probed with MSI1 antibody. Graph below shows qPCR performed on bound RNA from IP. CLIP values are presented as a percentage of input following subtraction of the IgG background. *GAPDH* serves as negative control. n=3 biological replicates. Note that a primer pair against 9519-9681 bp of ZIKV genome was used in these qPCRs (Table S4).

F) Immunofluorescence of mock- or PE243-infected U-251 cells. MSI1 (green) and dsRNA (red) signals are detected by confocal microscopy. DNA is detected by Hoechst stain (blue). Framed area is shown at higher magnification below.

G) Immunofluorescence of a PE243-infected U-251 cell. MSI1 (green) and dsRNA (red) signals are detected by STED super resolution microscopy. Outlines of the cell and nucleus are indicated in white and blue, respectively. Framed area is shown at higher magnification.

Fig. 2. MSI1 is required for effective replication of ZIKV.

A) Graph on left depicts viral RNA copies in control siRNA- (siCon) and MSI1 siRNA (siMSI1)-treated SK-N-BE2c and U-251 cells following infection with PE243 (MOI: 1 FFU/cell, 72h). Graph on right shows viral RNA copies in MSI1 siRNA-treated H9-derived neural stem cells (NSC) following infection with PE243 (MOI: 1 FFU/cell, 48h). Note that in all viral replication assays ZIKV was quantified by TaqMan assay as described in Materials and Methods. n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s.).

B) Representative western blots of cell lines treated with control and MSI1 siRNAs from Fig. 2A. Blots were probed with antibodies against MSI1 or p150 as loading control.

C) Location of the guide RNAs used for CRISPR/Cas9-mediated editing of the *MSI1* locus in U-251 cells. For further details see text and Fig. S2. Western blots of parental, control, KO1 and KO2 cell lines probed with antibodies against N- and C-termini of MSI1 (NT or CT) and MSI2. p150 serves as loading control.

D) Kinetics of PE243 viral RNA copies following infection in U-251 cells of different genotypes at the indicated time points. Note that cell death precluded collection of RNA from parental and control cells at 72h (MOI: 3 FFU/cell, 72h).

E) Confocal microscopy images of PE243-infected control and KO1 U-251 cells immunostained with antibodies against dsRNA (green) and Hoechst DNA stain (red) following mock or PE243 infection (MOI: 3 FFU/cell, 48h). Graph on top shows percentage of cells containing dsRNA signal, whereas box plot on bottom depicts total dsRNA staining volume per cell. Note that only cells with detectable dsRNA signal were included in the latter analysis. Boxes: 25th to 75th percentile; whiskers: 5-95% range; line: median. P-value represents Student's t-test, unpaired, two-tailed: **** $p<0.0001$.

F) Viral RNA copies in U-251 cells of different genotypes following infection with MR766 (MOI: 3 FFU/cell, 48 h). n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s.).

G) Virus-binding assays performed under conditions that prevent internalisation. PE243 infection was performed in U-251 cells of different genotypes (MOI: 3 FFU/cell, 1h). n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s.). Bar charts depict mean±s.e.m.

H) Pseudotyped particle infectivity assay in U-251 cells of different genotypes. HIV (pNL4-3.luc.R-E-) or MoMLV pseudotyped virus expressing a luciferase reporter, with either PE243 ZIKV, VSVg or a negative control envelope used to determine viral entry events. n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s.). Bar charts depict mean±s.e.m.

Fig. 3. MSI1 is enriched in neural progenitors, regulates Microcephalin (MCPH1) expression and is mutated in MCPH patients.

A) Immunohistochemistry of human embryonic brain at post conception week (pcw) 10 and 12. Tissue sections stained with antibodies against MSI1 (red) combined with neuron-specific β-III tubulin (green), or the apical neural progenitor marker Nestin (green). DNA is detected by DAPI (blue). Note that MSI1 is enriched in neural progenitors at the ventricular and subventricular zones (VZ and SVZ), but is absent from the cortical plate (CP).

B) MSI1 CLIP from U-251 cells with genotypes as indicated. CLIP was performed with rabbit IgG or MSI1 antibodies. Graphs show qPCRs of bound transcripts. n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s).

C) RNA EMSA analysis to detect binding between MCPH1_L 3'UTR and purified GST-MSI1 recombinant protein. Coomassie staining of corresponding purified proteins is shown below.

D) Western blots of parental, control, KO1 and KO2 cell lines. Blots were probed with antibodies as indicated. p150 serves as loading control.

E) Western blots of whole cell lysates from parent-of-patient- and patient-derived primary lymphocytes. Blots were probed with antibodies as indicated. MSI1 levels are unchanged.

F) MSI1 RIP from mock- and ZIKV-infected U-251 cells (MOI: 1 FFU/cell). Rabbit IgG or MSI1 antibodies were used for immunoprecipitation. Input corresponds to 10% of whole cell extract. Western blot was probed with MSI1 antibody. Graphs below show amounts of bound RNAs including ZIKV genome and endogenous target transcripts. RIP values are presented as a percentage of input following subtraction of the IgG background. MR766 and PE243 were quantitated by TaqMan assay, whereas endogenous transcripts with SYBR qPCR. Bar charts depict mean±s.e.m. n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s.).

G) Western blots of PE243-infected U-251 cells. (MOI: 1 FFU/cell, 72h). Blots were probed with antibodies as indicated. α-tubulin serves as loading control. Note that MSI1 positively regulates MCPH1 and CDK6, and negatively regulates NUMB and p21 protein levels.

Fig. 4. The A184V MCPH mutation disrupts RNA binding by MSI1 and impairs the ability of MSI1 to drive ZIKV replication.

A) Structural model of MSI1 (blue) and HRP1/RNA complex (grey/orange) predicts that the A184V mutation impairs the interaction with RNA due to a steric clash.

B) Western blots of cell lines stably expressing WT or A184V MSI1 transgenes. Note that cell lines were derived from either KO1 or KO2 cells as specified. Blots were probed with antibodies as indicated. p150 serves as loading control. Graph shows signal intensities of each protein normalized to parental cells.

C) MSI1 RIP from U-251 cells with genotypes as indicated. Rabbit IgG or MSI1 antibodies were used for immunoprecipitation. Input corresponds to 10% of whole cell extract. Western blot was probed with MSI1 antibody. Graphs below show qPCRs of bound transcripts. n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s).

D) Quantification of viral RNA copies in U-251 cells with the indicated genotypes following infection with PE243 (MOI: 3 FFU/cell, 48h). n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: ** $p<0.005$, not significant (n.s).

E) Changes in survival of U-251 cells with different genotypes following infection with PE243 (MOI: 3 FFU/cell, 48h). n=3 biological replicates. P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s).

F) ZIKV infection of HEK293T cells as measured by FACS analysis of flavivirus protein E. FACS was performed on control, MsiWT or MsiAV transgene-expressing cells following infection with PE243 (MOI: 1 FFU/cell, 48h). n=2 biological replicates. Bar charts depict mean±s.e.m.

G) Changes in survival of control, MsiWT or MsiAV transgene-expressing HEK293T cells following infection with PE243 (MOI: 3 FFU/cell, 48h). n=3 biological replicates, and P-values were obtained from Student's t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s).

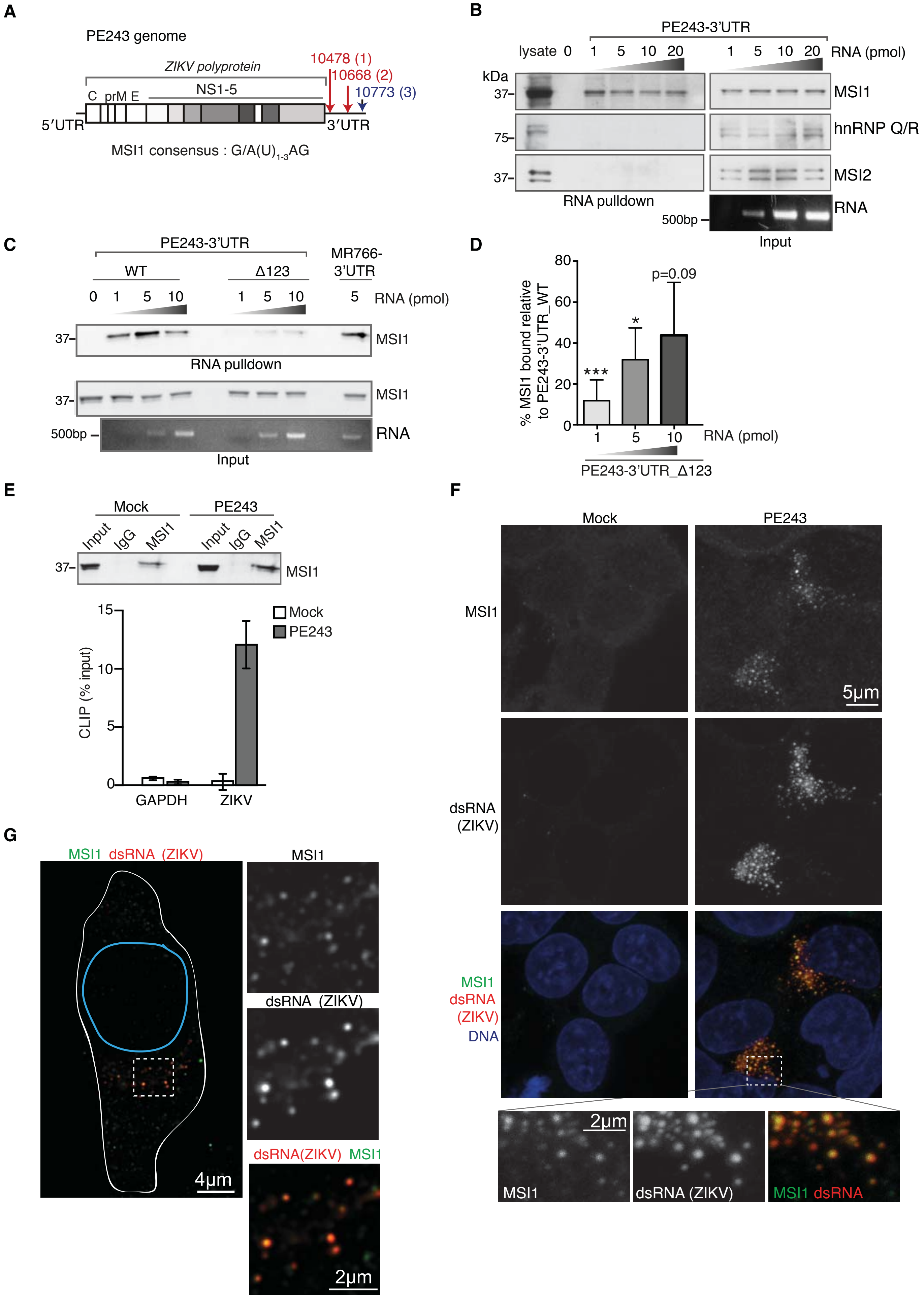
Supplementary Material Content

Materials and Methods

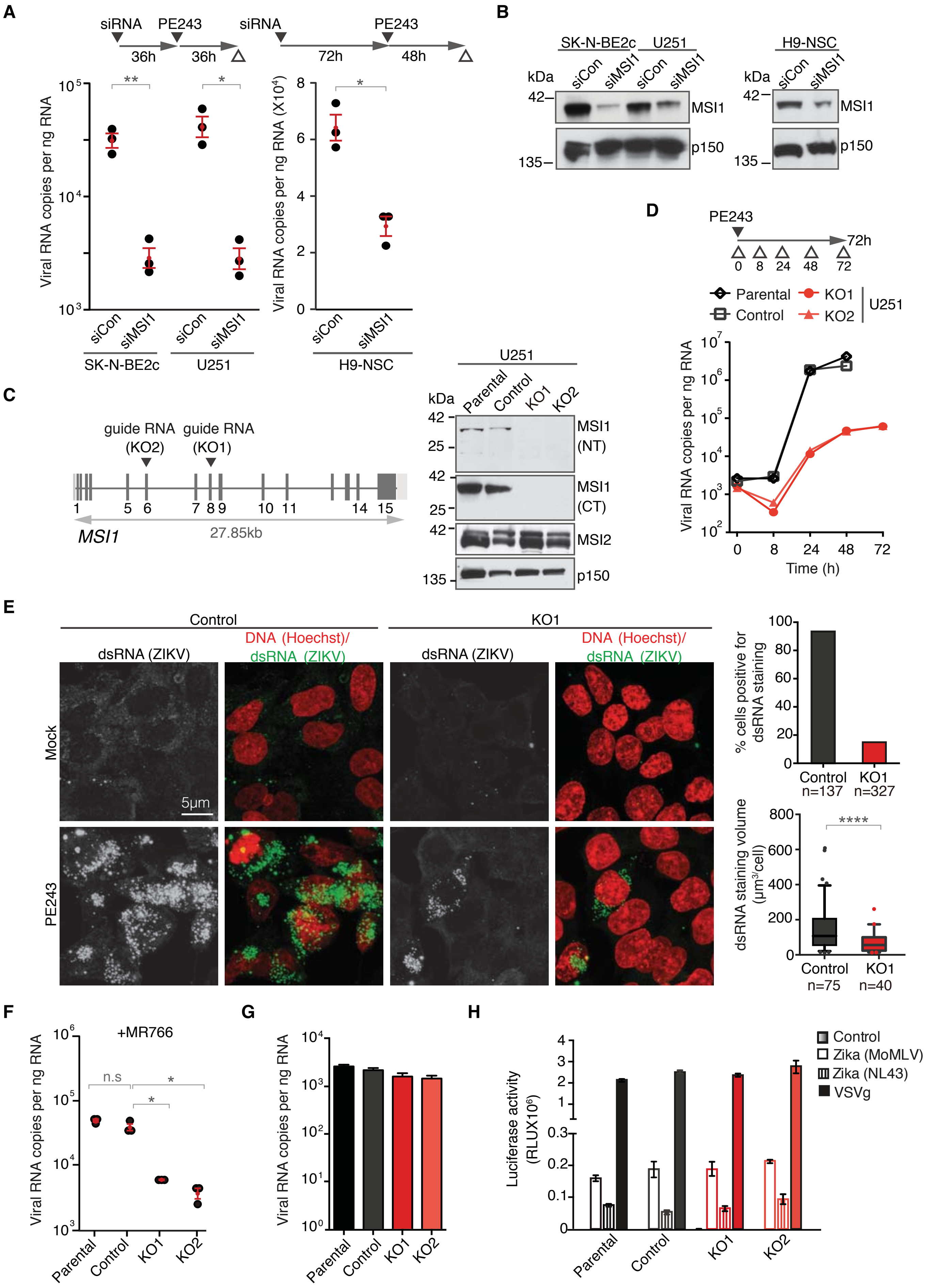
Figures S1-S13

Tables S1-S4

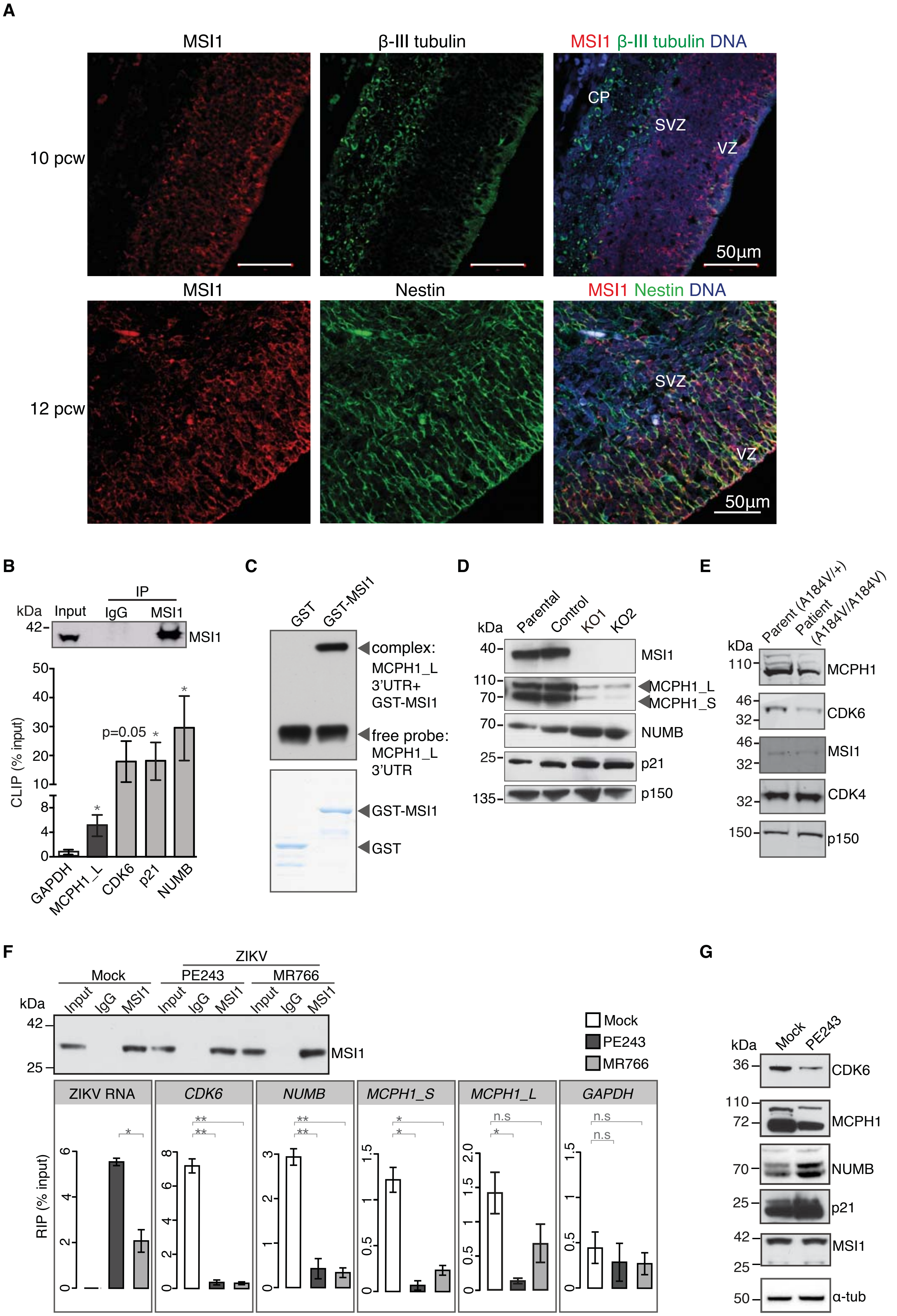
References *42-56*



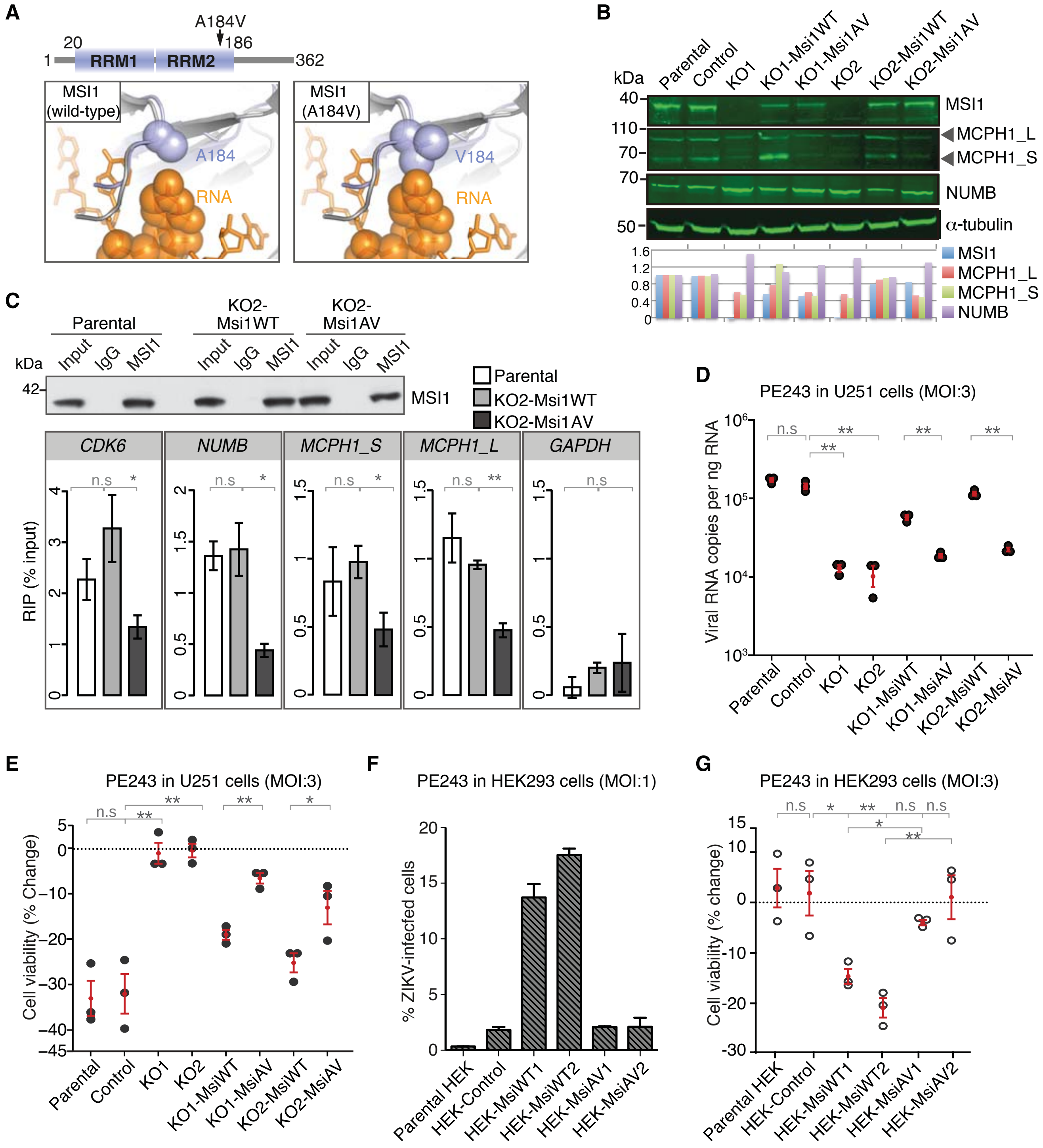
Chavali et al., Fig.1



Chavali et al., Fig. 2



Chavali et al., Figure 3



Chavali et al., Figure 4

Supplementary Materials for

Neurodevelopmental protein Musashi 1 interacts with the Zika genome and promotes viral replication

Pavithra L. Chavali[1,2,†], Lovorka Stojic[1,†], Luke W. Meredith[3], Nimesh Joseph[1], Michael S. Nahorski[4], Thomas J. Sanford[3], Trevor R. Sweeney[3], Ben A. Krishna[5], Myra Hosmillo[3], Andrew E. Firth[3], Richard Bayliss[6], Carlo L. Marcelis[7], Susan Lindsay[8], Ian Goodfellow[3], C. Geoffrey Woods[4] and Fanni Gergely[1*]

Correspondence to: fanni.gergely@cruk.cam.ac.uk

This PDF file includes:

Materials and Methods

Figs. S1 to S13

Tables S1 to S4.

Materials and Methods

Clinical details

The index case was the first born to healthy Turkish parents, who were first cousins (see Fig. S8). He was noted to be microcephalic at birth, but had no significant growth or health problems in his first year. At four years of age he was assessed to have mild cognitive delay, but had no motor delay nor fits and had a normal neurological examination. At that age his height was 95.3cm (-3 standard deviations (SDS)) and OFC 43.5 (-4 SDS for age and sex). With the exception of microcephaly he had no dysmorphic features, was in good health, did not suffer from excess infections and had a normal tolerance to sunlight, vision and hearing. He was seen again at age 8, when it was clear he had definite mild/moderate developmental delay. He attended a special school. His height was 117cm (-3 SDS) and OFC 45 (-5 SDS). On investigation he had no structural chromosomal anomalies, but cytogenetic spreads did show increased premature chromosome condensation (PCC). The patient had a significantly increased number of cells exhibiting PCC (14%) with respect to three control patients (6%). His youngest sister was seen at age of 7 months and had been noted to be microcephalic at birth. At 7 months her psychomotor development seemed normal. Her length was 63 cm (-2 SDS) and OFC 37.4 (-4 SDS). Both parents have normal intelligence, a normal height (both 163cm, -2 SDS for the father and 0 SD for the mother) and head circumferences; father 53cm (-1 SD) and mother 52 (-2 SDS). There is a healthy second sister with normal development and growth parameters. We concluded that the phenotype was inherited as an autosomal recessive trait because i) the male and female children were similarly affected, ii) a third sibling was unaffected, and iii) both parents were unaffected. The family gave informed consent to enter the study, which was approved by the National Research Ethics Service Committee, East of England - Cambridge Central, UK (C.G. Woods, REC 05/Q0108/402).

Molecular genetics

Standard methods were used to extract DNA from the two affected children and their parents. Given the phenotype of primary microcephaly, premature chromosome condensation and possible short stature, the MCPH1 gene was sequenced (*28*, *29*). Surprisingly, no mutations were found and examination of the MCPH1 genotypes showed that they had discordant heterozygosity for MCPH1 (subsequent exome sequencing confirmed this finding). Taken together these results eliminated MCPH1 as a cause of this recessive phenotype. As the phenotype was recessive and the parents consanguineous, we performed homozygosity analysis seeking concordant homozygous segments of > 3cM(*42*). Autozygosity mapping is likely to identify the regions of the genome containing recessive gene mutations in consanguineous pedigrees (*43*, *44*). Genomic DNA from the affected male was hybridized to the Affymetrix GeneChip® Human Mapping 250 K Nsp Array. Results of the alleles present at each SNP for each individual were generated using Affymetrix software. Regions of shared allele-concordant homozygosity were sought using CNAG2.0 (http://www.genome.umin.jp/ and (*45*)), and further analysed with ExcludeAR(*42*). This revealed fourteen regions of conserved homozygosity totaling 374Mb, a tenth of the genome - as expected for the child of first cousins (*46*), (see Table S1). Next, exome sequencing was performed on the affected male and female using the SureSelect Human All Exon 50Mb Kit (Agilent Technologies UK, Cat. No G3370A), enabling targeted capture of 50Mb sequence of exonic regions and non-coding RNAs in the human genome. Sequencing was performed with the SOLiD™4 System (Applied Biosystems) with 50bp fragment reads, to generate 4.3Gb of sequence achieving 83% of bases coverage by >10 reads (sufficient for the detection of homozygous mutations) of the mappable targeted 50Mb exome. Initially, the raw sequencing reads were mapped to the GRCh37 reference human genome and changes compared to this reference sequence identified (*47*). First, we checked all reported MCPH genes, but found no homozygous or potential compound heterozygous mutations. Next, we focused our analyses on non-synonymous coding, nonsense, splice site variants and indels (insertions–deletions) involving exons (1542 before further filtering analysis). We filtered against: (i) known variations where the rare allele frequency was >1% (derived from dbSNP and 1000 Genomes); (ii) high (all mammals) or complete evolutionary conservation of the encoded amino acid, by use of the Human Genome Browser full conservation track; (iii) examination of the sequence reads containing potential mutations using the Integrated Genome Viewer (which shows misaligned, poorly sequenced and recurrent artefact changes); (iv) presence within a homozygous region and (v) known phenotype of recessive gene mutations. We found four homozygous mutations (see Table S2), of which, three were not expressed in the foetal/ embryonic brain of humans or mice (Genepaint: http://www.genepaint.org/Frameset.html and

Brainspan Atlas of the Developing Human Brain (*48*)). The remaining mutation was on chromosome 12 at position 120794806G>A. This causes a missense mutation c.551C>T in MSI1 (MUSASHI-1) changing a GCT codon to GTT and resulting in p.Ala184Val, predicted to be potentially pathogenic in PolyPhen and SIFT (see Table S2). The p.(Ala184Val) mutation has not been reported in the 1000 Genomes project, EVS or ExAC (http://browser.1000genomes.org/index.html (*49*); http://evs.gs.washington.edu/EVS/#tabs-7 (*50*); and http://exac.broadinstitute.org). The mutation segregated as expected in the affected children and their parents. MSI1 Ala184 is present in all reported vertebrates; in *D. melanogaster* Musashi Ala184 was equivalent to Ala337 and in *C. elegans* Ala209. We found no further bi-allelic or homozygous MSI1 mutations in our own cohort, and those of colleagues, of approximately 100 MCPH families, which had no mutations in known MCPH genes.

Cell lines and reagents

U-251 MG, Vero, HEK293T cells were maintained in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% FBS, 1% penicillin/streptomycin, 1% non-essential amino acids and 1% L-glutamine. H9-derived neural stem cells (hNSC) were obtained from Invitrogen (Cat. No. N7800-200) and were cultured following the manufacturer's instructions. SK-N-BE2c cells were maintained in DMEM:F12 media containing 10% FBS, 1% penicillin/streptomycin and 1% L-glutamine. All cell lines were tested to be mycoplasma free. C6/36 insect cells were supplied by A. Kohl (Centre for Virus Research, University of Glasgow) and were maintained in Leibowitz-15 (L-15) media containing 10% FBS, 1% penicillin/streptomycin, 1% L-glutamine, 1% non-essential amino acids, and 20% tryptose phosphate broth (TPB, Sigma).

Genome editing of the *MSI1* locus by CRISPR-Cas9

The strategy and exons targeted with respective guides to generate MSI1 knockout in U-251 cells are depicted Fig. S2. Briefly, guides were generated as single stranded oligos and annealed to be cloned into Bbs1 linearised pSpCas9(BB)-2A-Puro (pX459). Positive clones containing guides were identified by sequencing and 0.5 μg of the guide RNA vectors were transfected into a single well of 12 well plate. Twenty-four hours post transfection, 2.5 μg/ml puromycin was added and left for 48 h. The surviving cells were plated onto a 15cm plate and single colonies were picked after 2 weeks. Clones were expanded and screened using western blot analysis using both the N- and the C-terminal MSI1 antibodies to detect complete knockout. Positive knockout clones were then genotyped to confirm biallelic modification of the *MSI1* locus.

Antibodies

Primary antibodies used in this study are: MSI1 (rabbit - ab21628: Abcam; rat - D270-3: MBL), MCPH1 (HPA008238, Cambridge Bioscience), CDK6 (Cell Signaling), NUMB (N6539: Sigma), p150 (612708: BD Bioscience), MSI2 (107701AP, Proteintech), hnRNP Q/R (Cell Signaling), PABP (Abcam, ab21060), Lamin B1 (Cell Signaling), Nestin (Abcam, ab22035), β-tubulin (Sigma), β-tubulin III (Abcam, ab18207), dsRNA-J2 (Scicons), p21 (2936, CST; ab109199, Abcam), flavivirus envelope E protein (D1-4G2-4-15, Absolute antibodies, Millipore). DNA was stained with Hoechst 33258 (Sigma-Aldrich) or DAPI. Isotype-matched Alexa-conjugated secondary antibodies (Thermo Fischer Scientific) were used for immunofluorescence and secondary antibodies conjugated to HRP (Amersham) or IRDye (LI-COR) were used for immunoblotting. For detection of MSI1 (rat), goat anti-rat HRP (Abcam, 97057) was used.

Western blotting and quantitation

U-251 cell pellets were lysed in modified TNN buffer (25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1.5 mM EDTA, 0.5% NP-40, 0.1% Triton-X100, and protease inhibitor cocktail [Sigma-Aldrich]) on ice for 15 min. Whole cell extracts were obtained by pelleting (16,000 *g* for 15 min at 4°C) and discarding the insoluble debris. H9-hNSC cells were lysed in NP40 lysis buffer (50 mM Tris-HCl, pH 8, 125 mM NaCl, 1% NP-40, 2 mM EDTA, 1 mM phenylmethylsulphonyl fluoride and protease inhibitor cocktail (Roche) and incubated on ice for 25 min.

Extracts were separated by SDS-PAGE using 4-12% Bis-Tris NuPAGE gels (Invitrogen) and blotted onto nitrocellulose membrane using iBlot1.0 (Invitrogen). Blots were incubated with primary antibodies at 1 μg/ml final concentration. Secondary antibodies conjugated to HRP (Amersham) were used at 1:5,000 dilution. Antibody binding was detected using an ECL system (Lumilight: Roche; Femto: Pierce) according to the manufacturer’s instructions. Densitometric analysis of Western blots was performed by ImageJ. For LiCOR, blots were processed as above and probed with mouse or rabbit specific IRDye 800 (LiCOR) and acquired using Odyssey CLx. Obtained fluorescent images were quantitated using Imagestudio.

Plasmid transfections and RNA interference

Cells were seeded the day before to achieve 60% confluency on the day of transfection. To generate stable cell lines expressing *MSI1* transgene, cells were co-transfected with pCDNA-Msi1WT or pCDNA-Msi1A184V and pBSK-PuromycinR in HEK293 and pBSK-BlasticidinR in U-251 cells in a single well of 6 well plate. 24 h post transfection, cells were split into 15 cm plates with 2.5 μg/ml puromycin or 8 μg/ml blasticidin. After 2 weeks, single colonies were picked and expanded. Positive clones were selected based on MSI1 reactivity on western blot, with parental U-251 cells serving as positive control. In case of siRNAs in U-251 cells, 25nM siRNA per well was transfected in a 24-well plate using Viromer green (Lipocalyx) as recommended by the manufacturer. 6h after transfection, the medium was changed. Cells were processed 72h after transfection and analysed by western blot to check the knock-down efficiency. siRNAs used in this study were MSI1 (SR302970; Origene), MCPH1 (SR312491; Origene), universal scrambled negative control (SR30004; Origene). For siRNA treatments followed by viral infections, 25nM MSI1 siRNA was transfected using RNAiMax following manufacturer’s instructions (Invitrogen). hNSC cells were seeded in 6 or 24 well and transfected with 1μM of control non targeting pool (Thermo Fisher, D-001910-10-20) or MSI1 Accell siRNA smart pool (Thermo Fisher, E-011338-00) according to the manufacturer’s instructions. After 3 days of transfection, Accell medium was replaced with hNSC medium and the cells were subjected to viral infections for additional 2 days before collecting RNA and proteins.

ZIKV 5’/3’ UTR luciferase reporter assay

Firefly luciferase reporter genes flanked by the 5’UTR and 3’UTR sequences of the PE243 strain of Zika virus and a T7 promoter sequence were synthesized by Integrated DNA Technologies. 5’ XbaI and 3’ HindIII restriction enzyme sites were included to facilitate cloning into pUC57. Reporter RNAs were produced by in vitro T7 transcription of

HindIII linearised plasmids and capped using the ScriptCap system (Cellscript). HEK293T cells were initially transfected with Lipofectamine 2000 (Thermo Fisher) in 6-well plates with 1.5 μg total nucleic acid comprised of pcDNA Msi1 and/or empty pcDNA as indicated in Fig. S4. Following expression for 24h, cells were re-seeded in 96 well plates and subsequently transfected with 145ng of PE243 capped firefly luciferase reporter mRNA and 5ng of pRL-TK Renilla luciferase reporter plasmid (Promega). Cells were incubated for a further 24h before luciferase production was assayed.

Polysome fractionation

Polysome fractionation from parental or MSI1 KO2 U-251 cells was performed as described in (*51*). Briefly, cells were washed with PBS containing cycloheximide (CHX) and lysed in hypotonic buffer (5 mM Tris-HCl (pH 7.5), 2.5 mM MgCl2, 1.5 mM KCl and 1x protease inhibitor cocktail (EDTA-free, 50 μg CHX, 1μM DTT, 100 units of RNaseOUT (Thermo Fisher Scientific), 0.5% Triton X-100 and 0.5% sodium deoxycholate). Clarified lysates (OD_{260} of 15) were loaded onto a continuous sucrose gradient (5-50% containing CHX) and subjected to ultracentrifugation. Fractions were collected and UV absorbance profiles recorded. 1 ml of trizol was added to 300μl fraction and processed for mRNA isolation. RNA from each of the 12 fractions was reverse transcribed using SuperScript III First-Strand Synthesis SuperMix and qPCR performed to determine distributions of *GAPDH* and *MCPH1_L*. For protein analysis 50μl fractions were subjected to TCA precipitation and processed for western blotting as described before.

Plasmid mutagenesis

pJET1.2 vector containing 3’-UTR of Zika virus PE243 (428bp) were subjected to site-directed mutagenesis following the standard sense-antisense method using Phusion DNA polymerase (NEB) in GC buffer with 2% DMSO followed by DpnI (NEB) digestion. The primers used for the mutagenesis are listed in Table S4.

Cell cycle analysis

The U-251 control and MSI1 knock-out cells were dissociated by trypsin digestion, washed with PBS and fixed with cold 70% ethanol for 30 min on ice. The cells were then washed twice with PBS, incubated in PBS containing RNaseA (100 µg/ml Life Technologies) for 30 min at 37 °C followed by staining with propidium iodide (20 µg/ml, Life Technologies) in PBS on ice in the dark for 30 min. DNA content was analysed by FACS Calibur (Beckton Dickinson) using BD CellQuest Pro Software V6. Cell cycle analysis was performed using FlowJo software V9 (TreeStar Inc).

Wound healing assay

Wound healing assay for U-251 control and MSI1 knock-out cells was performed in triplicates on 24-well (Essen Imagelock) plates seeded with 300,000 cells/well and imaged the second day onwards. The cell migration was followed for 78 h using IncuCyte ZOOM (Essen Bioscience), making measurements in triplicate every 3 h. The IncuCyte ZOOM version 2016B software was used to capture and analyse the images.

Zika virus genesis and infection

Zika virus stocks were generated by transfecting PE243 RNA (supplied by A. Kohl, Centre for Virus Research, University of Glasgow, Lindomar J. Pena and Rafael Oliveira de Freitas França, Fiocruz Recife, Pernambuco, Brazil) and MR766 RNA (PHE Culture Collections) into C6/36 cells using Lipofectamine 2000 (Life Technologies). Viral supernatants were collected 72h post-transfection and at subsequent 24h intervals. Virus stocks were pooled, aliquoted and stored at -80^{o}C. Viral titres were determined by immunofocal assays with serial viral dilutions onto Vero cells. Following 72h incubation, cells were fixed and stained for Zika flavivirus E-protein, which was then quantified by counting fluorescent foci (immunofluorescence described below) to determine focus-forming units per mL (FFU/mL). Zika viral RNA loads were also determined by qPCR (as described below). For infections, various cell lines (treated or untreated, as described in the text) were inoculated with virus at defined MOI (FFU/cell), incubated for 6h, then washed thoroughly with PBS, before media was replaced and incubated for a further 48 or 72h as stated. Viral cells or supernatants were then harvested for either RNA extraction or imaging. For TCID50 assay, U-251 control and MsiKO cells were infected with PE243 virus at an MOI of 0.01 and 0.001 FFU/cell. 72 hours post-infection, the cultures were frozen and thawed, the supernatant clarified and the virus yield determined by TCID50 using the Reed–Muench method in Vero cells by staining for the presence of viral antigen.

Zika pseudoparticle entry assay

Pseudoparticles were generated by transfecting HEK293T cells with plasmids encoding HIV-1 provirus expressing luciferase (pNL4-3-Luc-R-E-), or Moloney murine leukemia virus (MoMLV) with a luciferase reporter (pTG126) with either ZIKV-Env, vesicular stomatitis virus G (VSVg), or a no-envelope (NE) control as previously reported (*52*). Supernatants were harvested 48h post transfection and clarified by centrifugation. Virus containing medium was added to target cells plated in 96-well plates, incubated for 4h, before cells were washed and media replaced. 48h post-infection, media was removed, cells lysed and luciferase activity measured. ZIKVpp infectivity was calculated by expressing the ZIKV or VSVg luciferase signal (relative light units, RLU) relative to the NE control.

Virus Binding Assay

U-251 parental, control and Msi1-KO cells were incubated on ice for 30 mins, then equivalent MOI of PE243 virus was added to the cells, while still on ice. Cells were then incubated for 1h, before washing 5 times with PBS, then lysing and extracting RNA from cells. RNA loads were then assessed by qPCR using the Genesig Zika qPCR kit (PrimerDesign).

Flu virus genesis and infection

Human influenza virus A/Puerto Rico/8/34 (H1N1) strain, designated as PR8, was provided by Paul Digard (University of Edinburgh), amplified in A549 cells, then titered on MDCK cells. For infections, equal virus titres were treated with trypsin (0.1%) for 5 mins, then added to target cells for 4h. Virus was then removed and the cells incubated for a further 48h before RNA was extracted and quantified using Genesig H1N1 Influenza qPCR kit (PrimerDesign).

Zika RNA quantification

To quantify viral replication, RNA was extracted from cells using the GenElute Mammalian RNA Miniprep Kit (Roche) following the manufacturer's protocol. Total RNA was then quantified by NanoDrop 1000 quantification (Thermo Fisher Scientific). qPCR was carried out using TaqMan® chemistry, with a commercial quantification kit (Genesig Standard Zika Virus Quantification Kit, PrimerDesign), using One-Step 2X qPCR Reagent (PrimerDesign).

Flow cytometry: quantification of flavivirus E protein

For flow cytometry, infected or control cells were fixed in 4% paraformaldehyde for 20 mins, then washed thoroughly, followed by blocking in PBS containing 0.5% bovine serum albumin (BSA), 0.1% Triton X-100 (PBST). Cells were stained overnight at 4ºC, using 4G2 anti-flavivirus antibody in PBST plus sodium azide, followed by isotype-matched Alexa 488-conjugated mouse IgG antibody (Invitrogen) for 1h at room temperature. Cells were then analysed on a FACSCalibur (BD Biosciences), counting 20,000 cells of each condition in duplicate. Results were then analyzed with FlowJo® Flow Cytometry software.

Immunofluorescence analysis

For fluorescent imaging of flaviviral particles and dsRNA, cells were fixed using 1% formaldehyde for 10 min, then washed thoroughly with PBS and permeabilised with PBS containing 0.5% Triton-X-100 for 10 min, followed by blocking in PBS containing 0.5% bovine serum albumin (BSA), 0.1% Triton X-100 (PBST). Cells were stained overnight at 4ºC, using 4G2 anti-flavivirus antibody or dsRNA-J2 antibody in BSA-PBST, followed by isotype-matched Alexa 488-conjugated mouse IgG antibody (Invitrogen) for 1h at room temperature. To analyse premature chromatin condensation, cells were fixed in ice-cold methanol for 5 min, followed by washing with 0.1% Tween containing 1XPBS. Cells were then blocked in PBS containing 5% BSA for 10 min and subjected to overnight staining with Lamin B1 antibody (1:1000, Cell Signalling) at 4ºC. Isotype specific secondary antibody was added for an hour at 37ºC and DNA was stained with DAPI or Hoescht. Images were acquired on a Leica SP8 STED-equipped microscope in confocal or STED modes. Image quantification was performed by Volocity 6.3 (Perkin Elmer). dsRNA-positive staining volumes were selected by intensity thresholding. Identical settings were used on all cells regardless of genotype or treatment.

Immunohistochemistry of fetal brain

Human paraffin-embedded brain sections (10pcw and 12pcw), fixed with 4%

paraformaldehyde were obtained from the MRC-Wellcome Trust Human Developmental Biology Resource. Sections were boiled for 10 min in 10 mM sodium citrate buffer (pH 6.0) and permeabilized in 0.3% Triton-X-100 for 30 min. Sections were blocked in 3% BSA and 0.3% Triton-X-100, and incubated with primary antibodies overnight in blocking buffer. Alexa-Fluor-conjugated secondary antibodies (Invitrogen) were added on the sections for 1 hour at room temperature, DAPI stained (Invitrogen) for 3 minutes and finally mounted onto ProLong Gold antifade reagent (Invitrogen). Sections were visualised on both a LSM880 Confocal Microscope with AiryScan, and a Zeiss AxioImager Z2 Upright Wide-field Microscope. Primary antibodies and their concentrations used on tissue sections are rat anti-MSI1 at (1:200), anti-beta III tubulin antibody (1:2000) and anti-Nestin antibody (1:200).

Cell viability assay

Cell viability was determined using the CellTitre Blue™ Kit (Promega), following manufacturer’s instructions. Briefly, cells were seeded at equal densities and infected for 48h with PE243. Virus was removed, cells washed and 1X CellTitre Blue reagent was then added to the cells and then the plates were incubated for a further 4h at 37ºC. Absorbance was then measured at 570 nm, and viability compared to uninfected control cells.

RNA immunoprecipitation (RIP)

RIP was performed from the parental and clonal U-251 cells, or PE243/ MR766/mock-infected U-251 cells using rabbit IgG (Sigma) and MSI1 (Abcam, ab21628) antibodies as previously reported (*53*). Quality of immunoprecipitations was assessed by immunoblot analysis with rat MSI1 antibody (MBL, D270-3). Briefly, 1.5 mg of total protein lysates were incubated with 8μg of for 2h at 4°C, followed by addition of Protein A/G beads (Pierce) for a further 2h at 4°C. Beads were washed 3 times (5 min each) with lysis buffer and RNA extracted by addition of 1 ml TRIzol (Life Technologies) to the beads. 10% of the lysate was used as input RNA pellet obtained after phenol chloroform extraction and ethanol precipitation. Finally, RNA was dissolved in 60µl of RNase-free water (Life Technologies). Power SYBR Green RNA-to-CT 1-Step Kit (Life Technologies) was used to set up for qPCR. Primers used are listed in Table S4. To detect ZIKV in RIP (Fig. 4H) qPCR was carried out using TaqMan® chemistry, with a commercial quantification kit (Genesig Standard Zika Virus Quantification Kit), using One-Step 2X qPCR Reagent (PrimerDesign).

Cross-Linking Immunoprecipitation and qPCR (CLIP-qPCR)

CLIP was performed as described previously (*54*) with slight modifications. The U-251 cells were crosslinked with 300mJ/cm2 of UVA (365nm) on ice using Stratalinker device. Subsequently lysates were subjected to partial RNaseI digestion (Thermo Fischer Scientific, AM2295, 10µl of 1:500 dilution per 1ml of cell lysate) for 3 min at 37°C and verified on agarose gel. This lysate was incubated with MSI1 coated Protein G Dynabeads (Life Technologies). Efficiency of immunoprecipitations was verified by western blot analysis. Beads were then washed in NP40 lysis buffer and DNaseI (20 U, Promega) was added for 15 min at 37°C following addition of 0.1% SDS and 0.5 mg/ml Proteinase K (Thermo Fischer Scientific) for 15 min at 55°C. The supernatant was collected and RNA was extracted using phenol/chloroform extraction and ethanol precipitation. Finally, RNA was dissolved in 15µl of RNase free water and qPCR was performed using Power SYBR Green RNA-to-CT 1-Step Kit (Life Technologies). Primers are listed in Table S4.

Electrophoretic mobility shift assays (EMSA)

MCPH1 (long) 3’UTR (NM_001322042.1) was amplified from U-251 cDNA and cloned in pJET1.2 vector, linearised with Xba1, purified and 1µg was used as template for *in vitro* transcription with 10X biotinylation mix (Roche) using T7 polymerase (NEB). Biotinylated RNA was recovered by phenol:chloroform:isoamyl alcohol extraction followed by ethanol precipitation. Resulting RNA pellet was resuspended in 50µl DEPC water and concentration estimated with nanodrop. Approximately 2 nM biotinylated RNA was mixed with 1µg of GST or GST-MSI1 and EMSA performed using LightShift Chemiluminescence EMSA kit (Pierce). The reaction mixes were then run on a 6% PAGE gel and transferred on to a nylon membrane. The transferred complexes were UV crosslinked at 120 mJ/cm2 for 45 seconds, blocked and probed with streptavidin HRP.

In vitro–biotinylated RNA pulldown

5µg of the wild type and mutant PE243 3’UTRs in pJet1.2 vector were linearised and subjected to *in vitro* transcription using T7 RNA polymerase buffer, T7 RNA polymerase (20U, Thermo Fisher Scientific, 18033-019), 8mM GTP (GE Healthcare), 5mM ATP (GE Healthcare), 5mM CTP (GE Healthcare), 1.3mM UTP (GE Healthcare), 0.7mM Bio-16-UTP

(Epicentre), 5mM DTT (Sigma), 20mM MgCl2, 80U/ml RNaseOUT (Thermo Fisher Scientific). The mixture was incubated for 3h at 37°C and EDTA (60mM) was added and precipitated overnight at -20°C. After centrifugation at 14000g for 30 min, the precipitate was dissolved in 50µl of water and cleaned with Bio-Spin 30 columns (Bio-Rad, 732-6231). Following DNaseI treatment (Promega) for 15 min at 37°C the reaction was cleaned once again with Bio-Spin 30 columns and RNA was precipitated overnight at -20°C and dissolved in 30µl of RNase free water. Purity of RNA was analysed on agarose gel using RNA Gel Loading Dye (2X, Thermo Fisher Scientific, R0641) after denaturation at 72°C for 10 min. For the pulldown, three hundred micrograms of total protein from U-251 cells was used (lysis buffer: 25mM Tris HCL, pH 7.4, 150mM KCl, 5mM EDTA, 5mM MgCl2, 1% NP40, 0.5mM DTT, protease inhibitor cocktail (Roche), 100U/ml RNAse OUT) and first precleread with magnetic MyOne Streptavidin T1 Beads (Thermo Fischer Scientific, 65601). The precleared lysate was diluted 2x in lysis buffer and supplemented with tRNA (0.1µg/µl, Thermo Fisher Scientific) and incubated with different concentrations of biotinylated RNA (1-20 pmol) for 2h at 4°C. Biotinylated RNA was previously heated up to 60°C for 5 min and slowly cooled down to room temperature. Subsequently, 40µl of MyOne Streptavidin T1 Beads were added for an hour and the beads were washed three times with lysis buffer containing 300mM KCl. Finally, the beads were resuspended in 50µl of sample buffer (4X, Bio-Rad) and reducing agent (20X, Bio-Rad), vortexed and boiled for 5 min at 95°C. After a short spin, the supernatant was collected and subjected to western blot analysis.

Statistical analysis and validations

Statistical significance of data was determined by two tailed t-tests in all experiments using R, after Shapiro-Wilk test for normality. P-values for each comparison are indicated in corresponding legends. In certain cases Fisher's exact test was used as indicated in legend. A minimum of 3 biological replicates was performed for each experiment, unless otherwise stated. All viral infections and qPCR experiments were performed double blind.

Structural modelling

The NMR structure of human MSI1 RNA binding domain 2 in the absence of RNA (PDB code 2MSS, residues 110-184) and the crystal structure of *S. cerevisiae* HRP1/RNA complex

(PDB code 2CJK) were used to build a homology model of the MSI1 residues 110-187 in a predicted RNA-bound conformation(*55*). The model was generated using Modeller 9.15 and the structure figure was generated using PyMOL (www.pymol.org)(*56*).

Figures and Figure legends

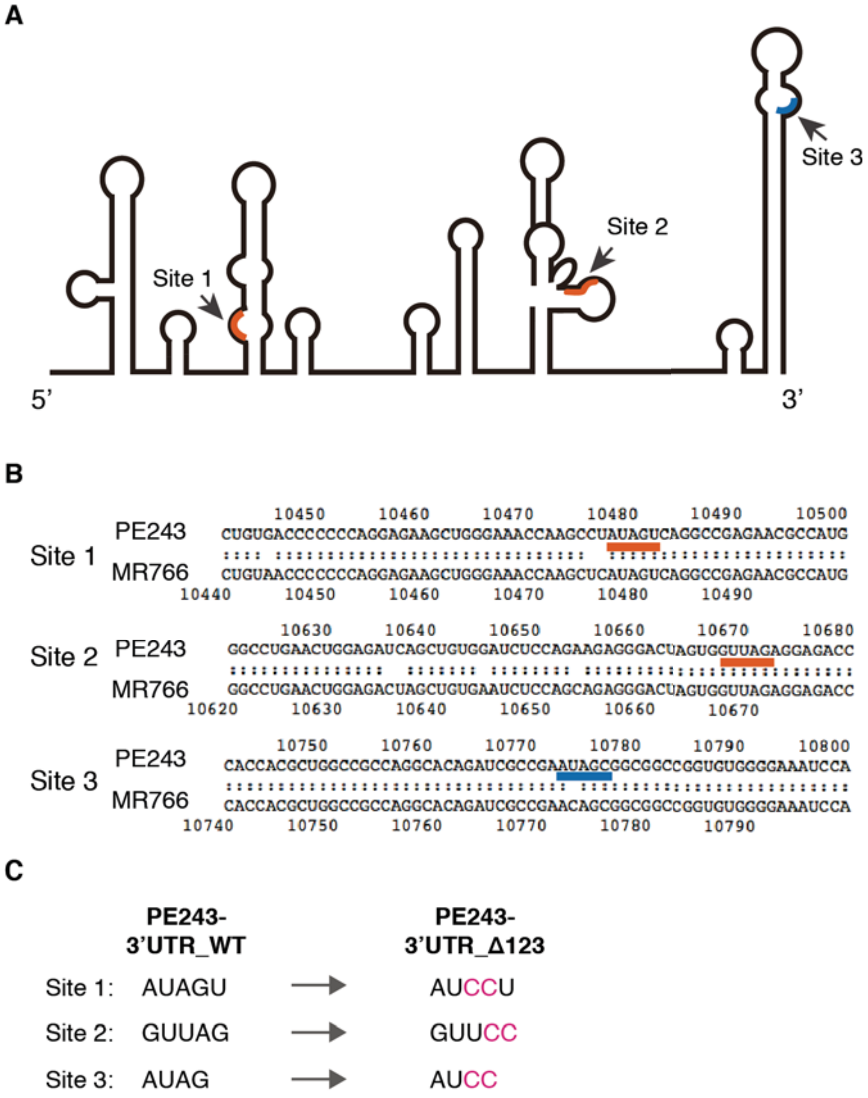


Fig. S1. Consensus MSI1 binding sites in the 3'UTR of ZIKV.

A) Location of consensus MSI1 binding sites in a structural model of ZIKV 3'UTR (adapted from (*16*), http://sergio14.github.io/Flaviviruses_RNA_Structures/Flavivirus-3UTR.svg#1). Sites common between MR766 and PE243 are in orange, site unique to PE243 is in blue. Note that there are additional putative MSI1 binding sites in the coding region of ZIKV RNA. However, a lack of detailed structural information makes it difficult to assess whether these map to stem-loop structures.

B) Alignment of sequences proximal to MSI1 binding sites in MR766 and PE243.

C) Mutagenesis of MSI1 binding sites in 3’UTR of PE243. All three MSI1 binding sites were mutated as indicated in the 3’UTR of PE243.

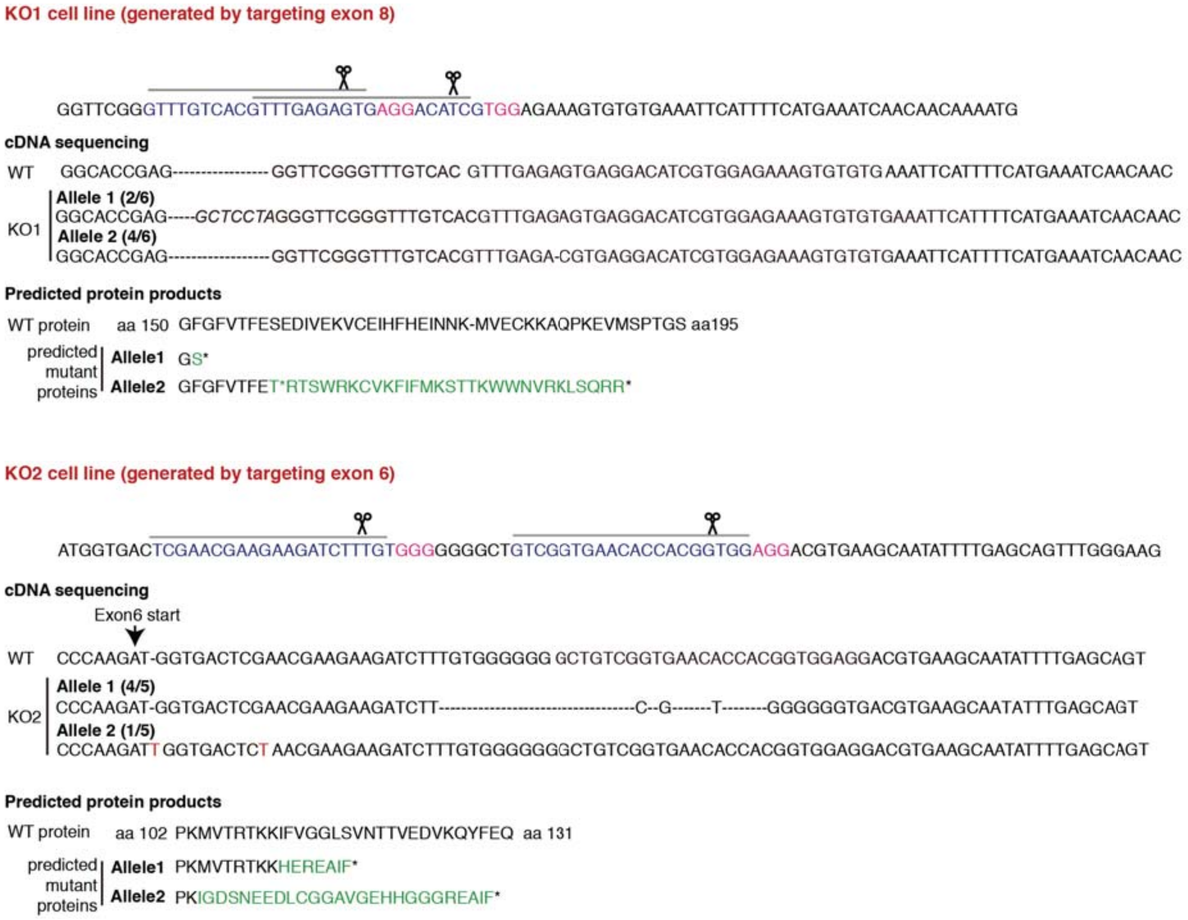


Fig. S2. Characterisation of *MSI1* cDNA sequences from MSI1 KO1 and KO2 U-251 cells.

Sanger sequencing reveals differential modifications of the *MSI1* alleles in each knockout, but all lead to frameshift mutations and hence protein null cell lines (Fig. 2C). PAM sites are in pink, guide RNA sequences in blue, whereas base insertions are in italics. Amino acids different from wild type sequence are shown in green.

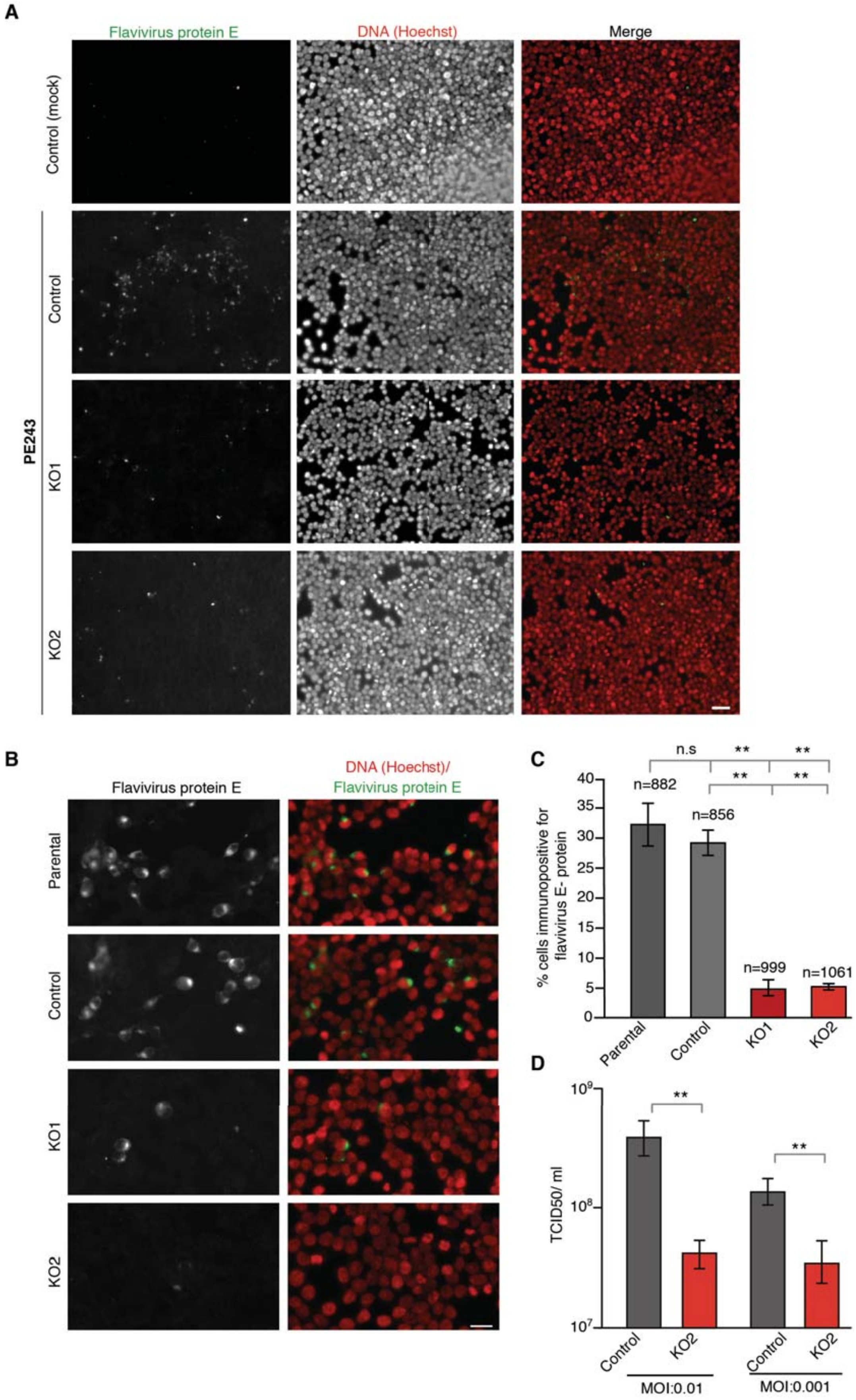

A
Flavivirus protein E
DNA (Hoechst)
Merge
Control (mock)
Control
KO1
KO2
PE243
B
Flavivirus protein E
DNA (Hoechst)/
Flavivirus protein E
Parental
Control
KO1
KO2
C
% cells immunopositive for flavivirus E- protein
0
5
10
15
20
25
30
35
40
n=882
n=856
n=999
n=1061
n.s
**
**
**
**
Parental
Control
KO1
KO2
D
TCID50/ ml
10^9
10^8
10^7
**
**
Control
KO2
Control
KO2
MOI:0.01
MOI:0.001

Fig. S3. ZIKV replication is impaired in MSI1 KO U-251 cells.

A) Images show U-251 cells with genotypes as indicated following mock or PE243 infection. MOI: 3 FFU/cell, 48h. Cells were stained with an antibody against flavivirus E protein. DNA is shown with Hoechst. Scale bar=50µm.

B) Higher magnification images from same experiment as Fig. S3A. Scale bar=25µm.

C) Quantification of flavivirus E protein-positive U-251 cells following PE243 infection. Genotypes as indicated. MOI: 1 FFU/cell, 24h. Data presented as mean ± s.e.m, n=3. p-values were obtained from Student t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s.).

D) U-251 cells were infected with PE243 at an MOI of 0.01 and 0.001 FFU/cell for 72h. Infectious virus titer was then quantitated by TCID50 assay and expressed as TCID50 per ml values by Reed-Muench method. Genotypes as indicated. Data presented as mean ± s.e.m, n=4, p-values were obtained from two-way ANOVA, **$p < 0.005$.

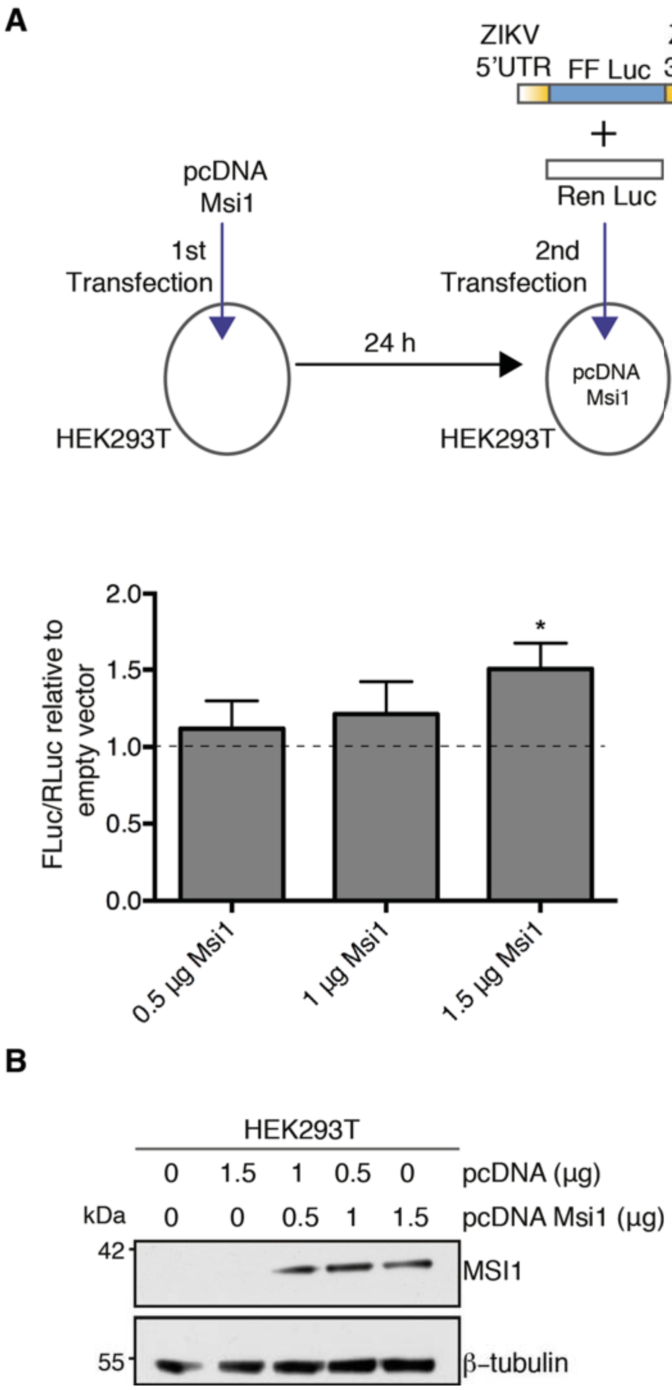


Fig. S4. MSI1 increases luciferase expression in a ZIKV UTR-dependent manner.

A) Experimental procedure is depicted in schematics. Following transfection with pcDNA or pcDNA-Msi1, Firefly (FLuc) luciferase flanked by the 5’ and 3’UTRs of PE243 ZIKV was introduced along with a Renilla luciferase reporter (RLuc) into HEK293T cells. Chart below shows the mean±s.e.m. (n=3 biological repeats, Mann-Whitney test, *=$p<0.05$) of the fold change in the ratio of firefly to renilla luciferase expression relative to the empty pcDNA transfection for each experiment. Note that unlike the 3’UTR of ZIKV, the 5’UTR does not contain consensus MSI1 binding sites.

B) A representative western blot of HEK293T cells transfected with defined concentrations of an empty vector (pcDNA) and a MSI1 transgene (pcDNA-Msi1) as in Fig. S4A. Blots were probed with antibodies against MSI1. β-tubulin serves as loading control.

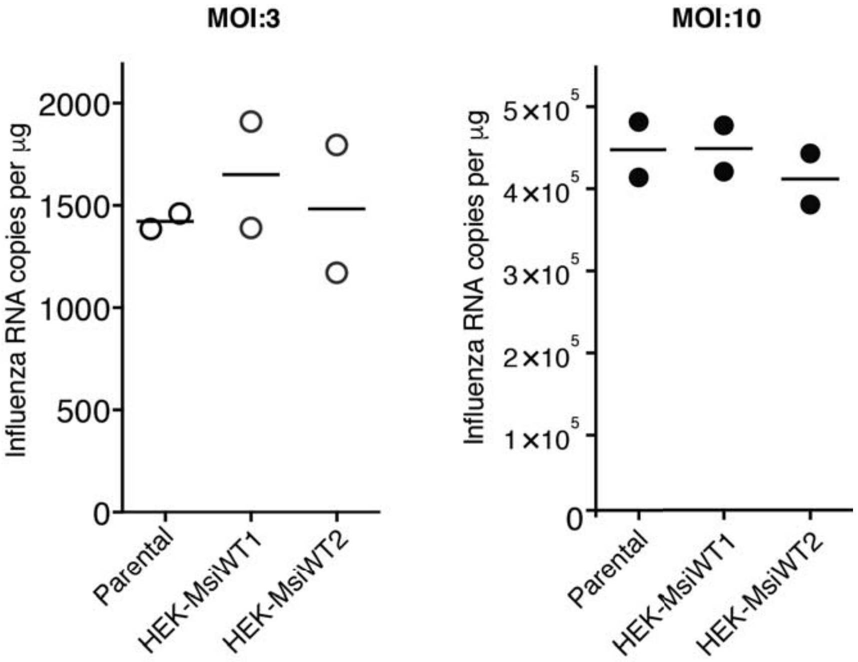


Fig. S5. MSI1 does not affect replication of H1N1 influenza virus. Viral RNA in HEK293T cells clones stably expressing a MSI1 transgene following infection with influenza virus (MOI:3 or MOI:10 FFU/cell, 48h). Refer to Figs. S11 and S12 for molecular characterization and ZIKV infection of the same HEK-MsiWT clones, respectively. n=2 biological replicates at each MOI.

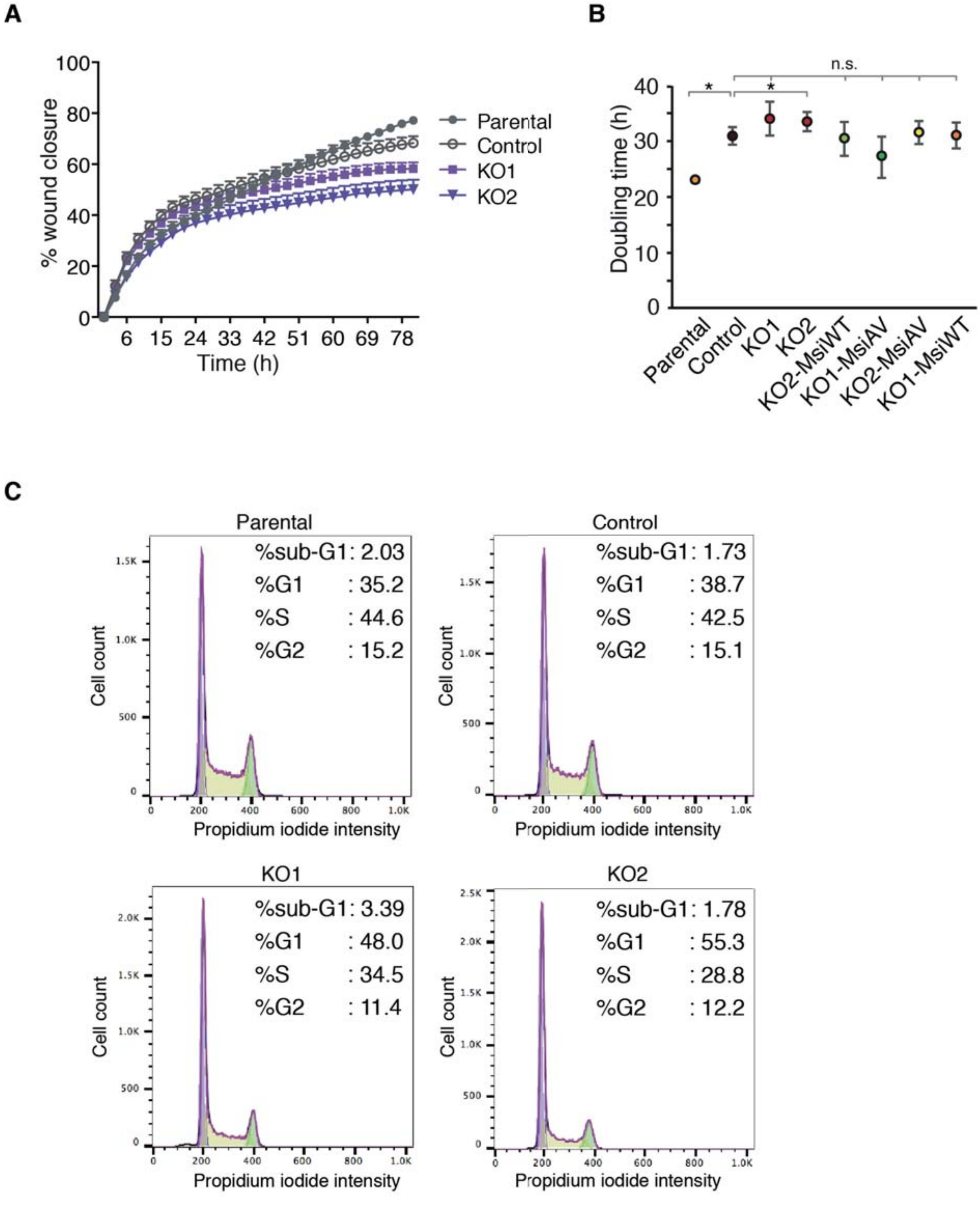


Fig. S6. Characterisation of MSI1 KO U-251 cells.

A) Wound closure assay of cells with indicated genotypes. Wound closure was measured by time-lapse imaging in an automated fashion by the IncuCyte system. n=3 biological replicates are plotted as mean ± s.e.m. P-values: $p=0.2$ for parental vs. control, $p<0.05$ for control vs. KO1 and $p<0.005$ for control vs. KO2. Values were obtained from Student's t-test, unpaired, two-tailed.

B**)** Doubling times of cells with indicated genotypes were obtained by an IncuCyte live-cell imaging and analysis system. n=3 replicates are plotted as mean ± s.e.m. p values were obtained from Student's t test, unpaired, two-tailed: $*p<0.05$, n.s= not significant.

C) Representative plots of cell cycle profiles of KO1 and KO2 cells as determined by FACS.

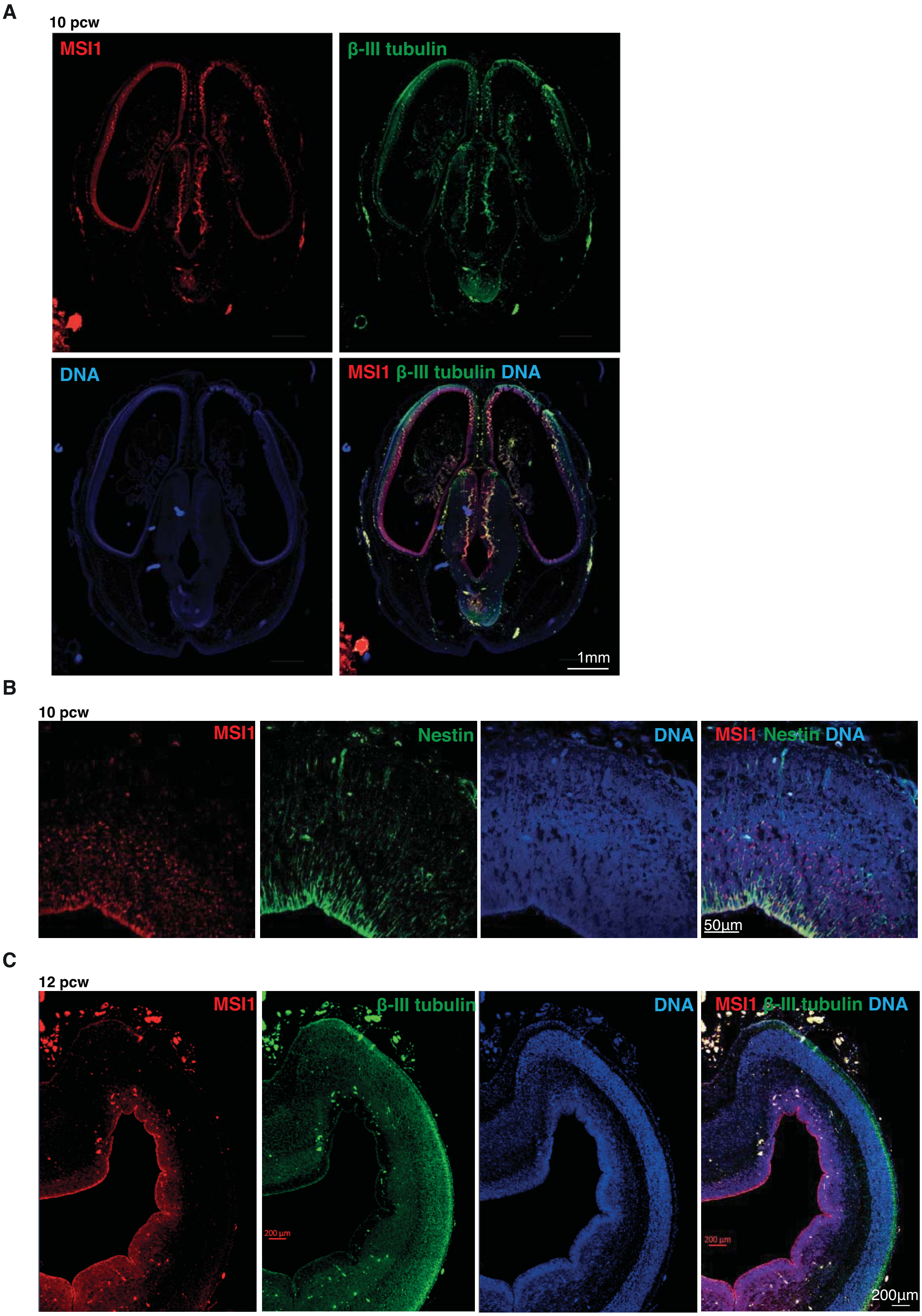

A
10 pcw
MSI1
β-III tubulin
DNA
MSI1 β-III tubulin DNA
1mm
B
10 pcw
MSI1
Nestin
DNA
MSI1 Nestin DNA
50µm
C
12 pcw
MSI1
β-III tubulin
200 µm
DNA
MSI1 β-III tubulin DNA
200 µm
200µm

Fig. S7. MSI1 expression is restricted to neural progenitors during human embryonic brain development.

A) Immunohistochemistry of human embryonic brain at post conception week (pcw) 10. Tissue sections were stained with antibodies against MSI1 (red) and the neuron-specific β-III tubulin (green). DNA is stained with DAPI (blue).

B) Immunohistochemistry of human embryonic brain at pcw 10. Tissue sections were stained with antibodies against MSI1 (red) and the apical neural progenitor marker Nestin (green). DNA is stained with DAPI (blue).

C) Tissue sections were stained with antibodies against MSI1 (red) and the neuron-specific β-III tubulin (green). DNA is stained with DAPI (blue).

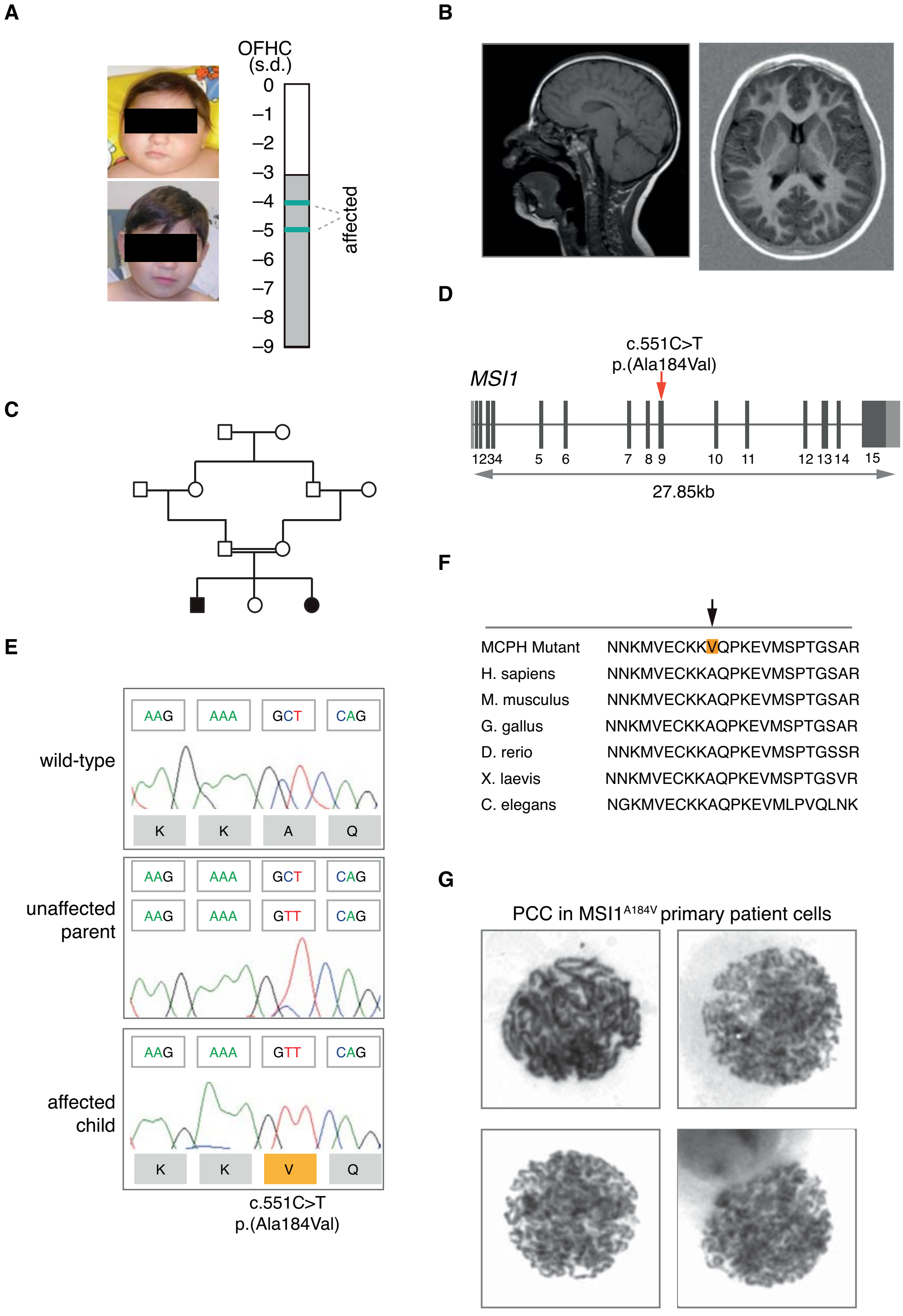

[redacted]
[redacted]

Fig. S8. Clinical and molecular genetic data of MSI1^{A184V} patients.

A) Images of affected children (published with consent from the family). Box on right indicates occipito-frontal head circumference (OFHC). Normal range is white, range typical for MCPH is grey with blue lines indicating the affected children.

B) Images show two representative T1-weighted MRI scans of patient 1 at age 4y. The brain architecture is normal. Note measurements of the cerebrum - reduction in brain volume is most evident for the cerebral cortices, and there are no congenital anomalies such as defects of neuronal migration. The left image is mid-sagittal and illustrates the normal structure and proportions of the brain regions and the right image is coronal at the level of the basal ganglia and shows normal cerebral cortex structure and proportions. Cerebral white and grey matter are normal for age.

C) Pedigree of affected family. Consanguineous relationship is marked with double line.

D) Location of point mutation in the *MSI1* gene. Light and dark grey boxes correspond to UTRs and exons, respectively.

E) Sanger sequencing traces of wild-type, parent (heterozygous) and affected patient (homozygous) samples.

F) Alanine 184 in MSI1 is a highly conserved residue across species.

G) Examples for premature chromosome condensation (PCC) in Giemsa-stained metaphase spreads prepared from patient 1’s primary lympocytes. Note condensed chromosomes in prophase-like cells. 14% of metaphases in patient cells and 6% in control cells were scored as PCC at the clinic.

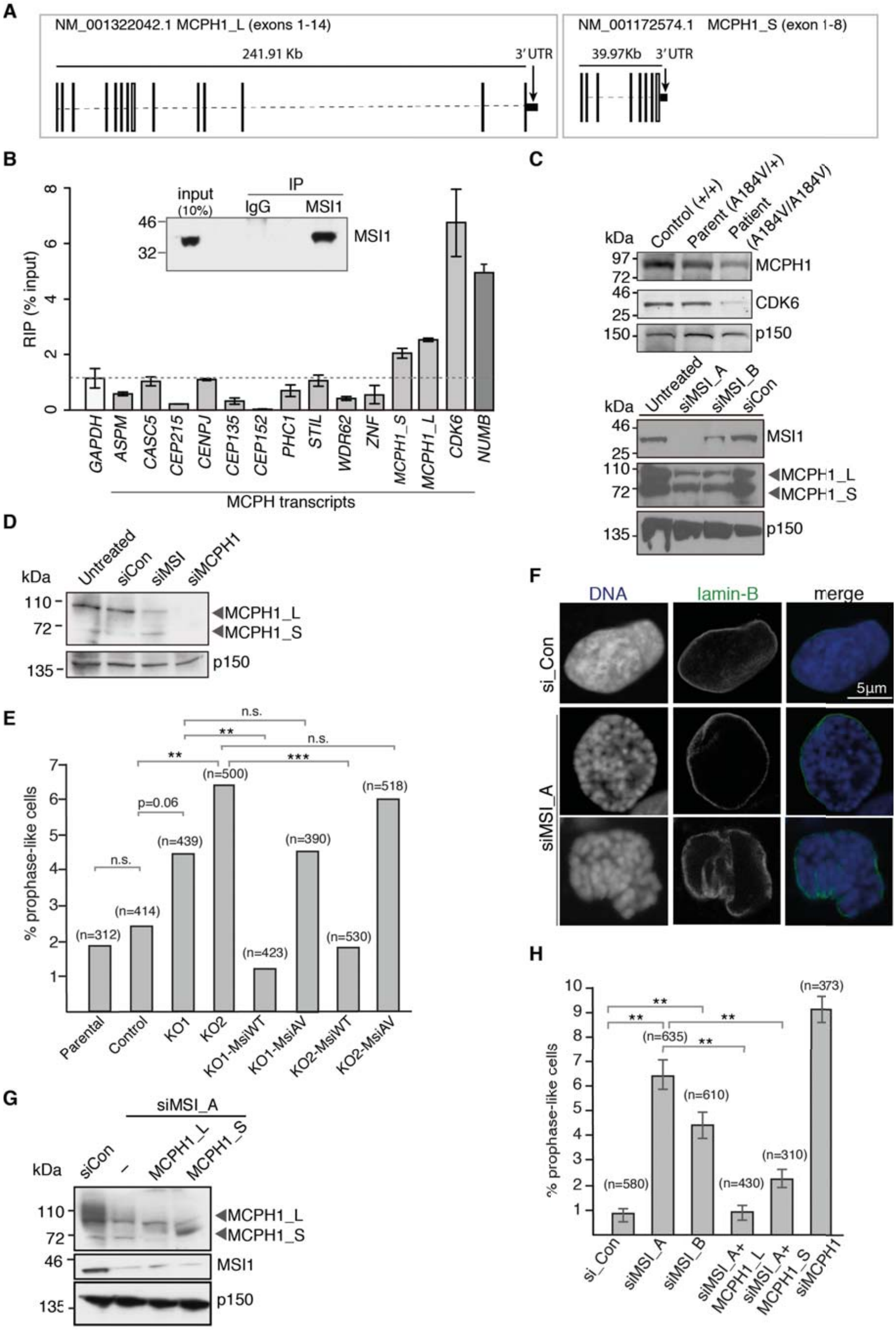

A
NM_001322042.1 MCPH1_L (exons 1-14)
241.91 Kb
3′UTR
NM_001172574.1 MCPH1_S (exon 1-8)
39.97Kb
3′UTR
B
RIP (% input)
0
2
4
6
8
input (10%)
IP
IgG
MSI1
46
32
MSI1
GAPDH
ASPM
CASC5
CEP215
CENPJ
CEP135
CEP152
PHC1
STIL
WDR62
ZNF
MCPH1_S
MCPH1_L
CDK6
NUMB
MCPH transcripts
C
kDa
Control (+/+)
Parent (A184V/+)
Patient (A184V/A184V)
97
72
MCPH1
46
25
CDK6
150
p150
kDa
Untreated
siMSI_A
siMSI_B
siCon
46
25
MSI1
110
72
MCPH1_L
MCPH1_S
135
p150
D
kDa
Untreated
siCon
siMSI
siMCPH1
110
72
MCPH1_L
MCPH1_S
135
p150
E
% prophase-like cells
1
2
3
4
5
6
7
n.s.
**
n.s.
**
***
p=0.06
n.s.
(n=312)
(n=414)
(n=439)
(n=500)
(n=423)
(n=390)
(n=530)
(n=518)
Parental
Control
KO1
KO2
KO1-MsiWT
KO1-MsiAV
KO2-MsiWT
KO2-MsiAV
F
DNA
lamin-B
merge
si_Con
siMSI_A
5μm
G
siMSI_A
kDa
siCon
–
MCPH1_L
MCPH1_S
110
72
MCPH1_L
MCPH1_S
46
MSI1
135
p150
H
% prophase-like cells
1
2
3
4
5
6
7
8
9
10
**
**
**
**
(n=580)
(n=635)
(n=610)
(n=430)
(n=310)
(n=373)
si_Con
siMSI_A
siMSI_B
siMSI_A+ MCPH1_L
siMSI_A+ MCPH1_S
siMCPH1

Fig. S9. MCPH1 downregulation is responsible for the PCC phenotype of MSI1-deficient cells.

A) Schematic representation of the long and short splice isoforms of MCPH1.

B) MSI1 RNA immunoprecipitation (RIP) from U-251 cells with different genotypes. Western blot shows immunoprecipitations (IP) by rabbit IgG and MSI1 antibodies from U-251 cells. Input represents 10% of whole cell extract. Western blot was probed with MSI1 antibody. Graph below shows qPCR performed on bound RNA from IP. RIP values are presented as a percentage of input following subtraction of the IgG background. *GAPDH* serves as negative and *NUMB* as positive control. n=3 biological replicates. Bar chart depicts mean ± s.e.m.

C) MCPH1 protein levels are reduced in MSI1-deficient cells. Western blot on top shows whole cell lysates from control, parent-of-patient- and patient-derived primary lymphocytes. Blots were probed with antibodies as indicated. Blot below shows whole cell extracts from cells treated with two different siRNAs targeting MSI1. Antibodies used for immunoblotting are indicated.

D) Western blot shows whole cell extracts from cells treated with siRNAs targeting MSI1 or MCPH1. Antibodies used for immunblotting are indicated.

E) Frequency of premature chromosome condensation (PCC) in U-251 cells with indicated genotypes. PCC refers to an increase in prophase-like cells, which can be distinguished by their condensed chromatin and intact nuclear envelope (by lamin-B staining) (as in Fig. S9F). p values were obtained from Fisher's exact test.

F) PCC in MSI1-depleted cells. Note intact nuclear envelope (as detected by Lamin-B antibody, green) and condensed chromatin (DNA is stained by Hoechst, blue).

G) Western blot shows whole cell lysates of cells treated with siRNAs targeting MSI1 in combination with transgenes expressing MCPH1_S or MCPH1_L. Antibodies used for immunblotting are indicated. Western blot is representative of experiments shown in Fig. S9H.

H) Graph depicts PCC frequency in MSI1-depleted cells expressing MCPH1_S or MCPH1_L transgenes. MCPH1_L and MCPH1_S were introduced 36h after addition of

MSI1 siRNAs and cells were fixed 36h later. Cells with indicated treatments and genotypes were scored by combining lamin-B antibody with DNA staining as in Fig. S9F. n=3 biological replicates. p values were obtained from Student’s t test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$.

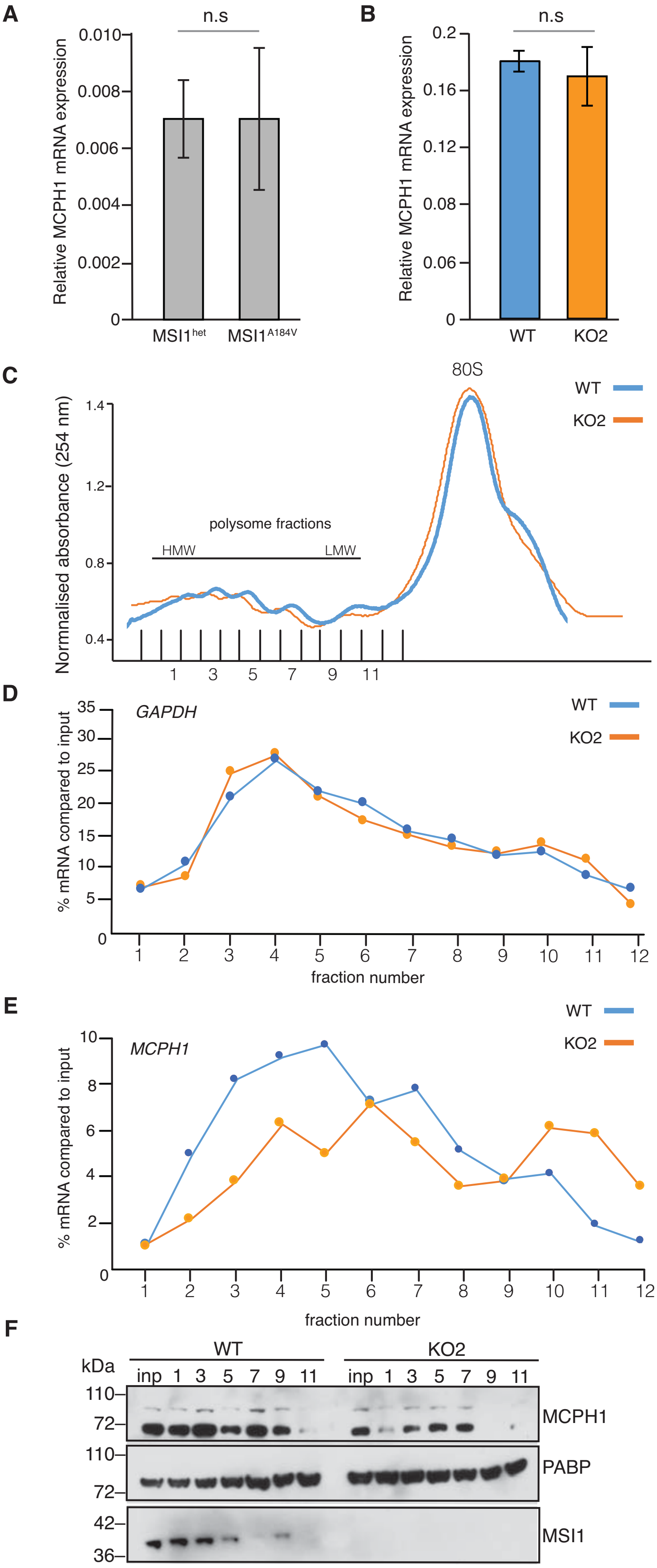

A
n.s
Relative MCPH1 mRNA expression
0.010
0.008
0.006
0.004
0.002
0
MSI1het
MSI1A184V
B
n.s
Relative MCPH1 mRNA expression
0.2
0.16
0.12
0.08
0.06
0
WT
KO2
C
80S
WT
KO2
Normalised absorbance (254 nm)
1.4
1.2
0.8
0.4
polysome fractions
HMW
LMW
1
3
5
7
9
11
D
GAPDH
WT
KO2
% mRNA compared to input
35
30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
11
12
fraction number
E
MCPH1
WT
KO2
% mRNA compared to input
10
8
6
4
2
0
1
2
3
4
5
6
7
8
9
10
11
12
fraction number
F
kDa
WT
KO2
inp 1 3 5 7 9 11
inp 1 3 5 7 9 11
110
72
MCPH1
110
72
PABP
42
36
MSI1

Fig. S10. MSI1 regulates translation of *MCPH1*.

A) Relative levels of total *MCPH1_L* transcript in parent-of-patient (Msi1^{het}) or patient-derived lymphocytes (Msi1^{A184V}) normalized to *GAPDH*. Bar chart depicts mean ± s.e.m. p values were obtained from Student's t test, unpaired, two-tailed. The differences were not significant (n.s).

B) Relative levels of total *MCPH1_L* transcript in WT and MSI1 KO U-251 cells normalized to *GAPDH*. Bar chart depicts mean ± s.e.m. p values were obtained from Student's t test, unpaired, two-tailed. The differences were not significant (n.s).

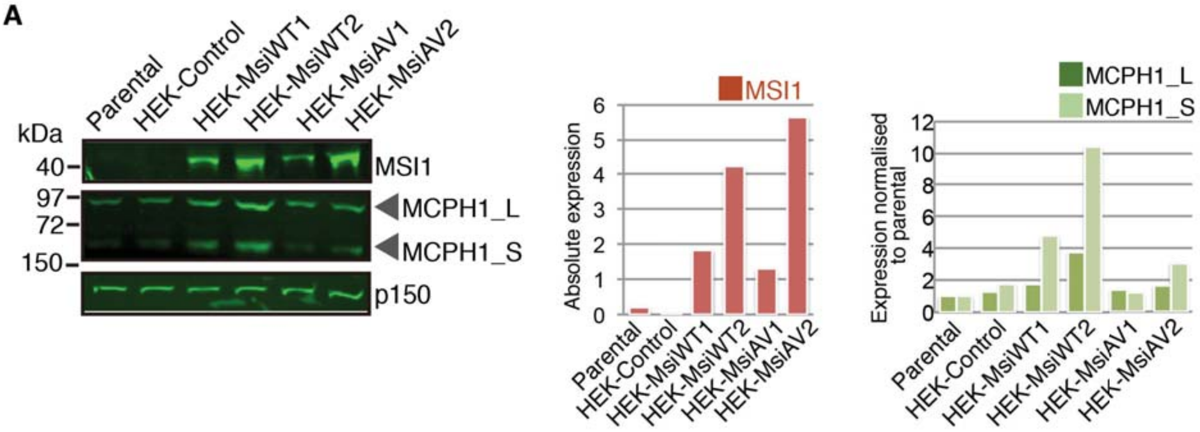
C) Absorbance (A254) profiles for continuous gradient fractionation of WT and MSI1 KO U-251 cells.

D) Distribution of sedimented *GAPDH* mRNAs throughout the gradients as quantified by qPCR. mRNA signal is normalized to cytosolic RNA content before fractionation.

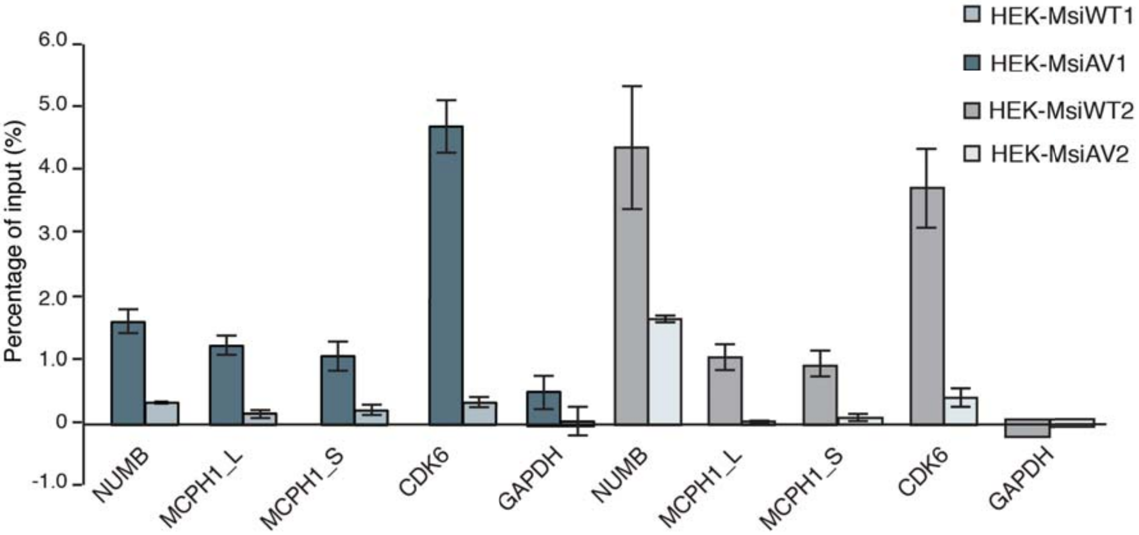
E) Distribution of sedimented *MCPH1_L* mRNAs throughout the gradients as quantified by qPCR. mRNA signal is normalized to cytosolic RNA content before fractionation.

F) Western blot analysis of TCA-precipitated fractions (corresponding to Fig. S10C) probed with antibodies against MCPH1 and MSI1. Poly A-binding protein (PABP) was used to confirm the polysomal fractions.

A


Parental
HEK-Control
HEK-MsiWT1
HEK-MsiWT2
HEK-MsiAV1
HEK-MsiAV2
kDa
40
97
72
150
MSI1
MCPH1_L
MCPH1_S
p150
MSI1
Absolute expression
0
1
2
3
4
5
6
Parental
HEK-Control
HEK-MsiWT1
HEK-MsiWT2
HEK-MsiAV1
HEK-MsiAV2
MCPH1_L
MCPH1_S
Expression normalised to parental
0
2
4
6
8
10
12
Parental
HEK-Control
HEK-MsiWT1
HEK-MsiWT2
HEK-MsiAV1
HEK-MsiAV2

B


HEK-MsiWT1
HEK-MsiAV1
HEK-MsiWT2
HEK-MsiAV2
Percentage of input (%)
6.0
5.0
4.0
3.0
2.0
1.0
0
-1.0
NUMB
MCPH1_L
MCPH1_S
CDK6
GAPDH
NUMB
MCPH1_L
MCPH1_S
CDK6
GAPDH

C

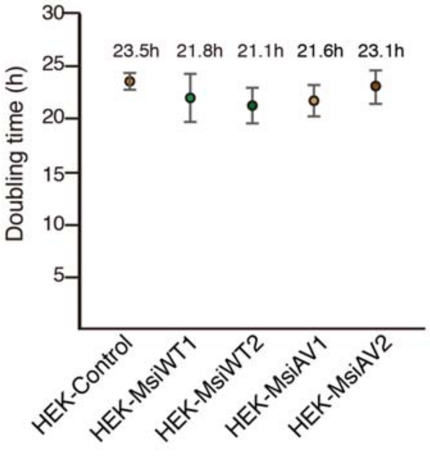

Doubling time (h)
30
25
20
15
10
5
23.5h
21.8h
21.1h
21.6h
23.1h
HEK-Control
HEK-MsiWT1
HEK-MsiWT2
HEK-MsiAV1
HEK-MsiAV2

Fig. S11. The A184V mutation impairs RNA binding by MSI1 and deregulates its targets in HEK293T cells.

A) Western blots show MSI1 and MCPH1 levels in whole cell extracts of HEK293T cell lines. Multiple stable, single cell-derived clones of HEK293T expressing MsiWT or MsiAV transgenes were generated. WT and AV clones were matched according to expression levels based on the western blot (WT1, AV1: low; WT2, AV2: high). Western blots were probed with antibodies against MSI1 and MCPH1. p150 serves as loading control. Quantitation of MSI1 and MCPH1 signals are shown on graphs to the right. Note that due to its absence in parental and control cells, absolute values are shown for MSI1, whereas MCPH1 levels were normalized against levels in parental cells.

B) qPCR from a MSI1 RNA immunoprecipitation (RIP) performed from the different HEK293T cell lines. RIP values are presented as a percentage of input following subtraction of the IgG background. *GAPDH* serves as negative and *NUMB* as positive control. Graph depicts mean±s.d. n=3 biological replicates.

C) Doubling times of cells with indicated genotypes were obtained by an IncuCyte live-cell imaging and analysis system. n=3 replicates are plotted as mean ± s.e.m. p values were obtained from Student's t test, unpaired, two-tailed. The differences were not significant for any of the genotypes.

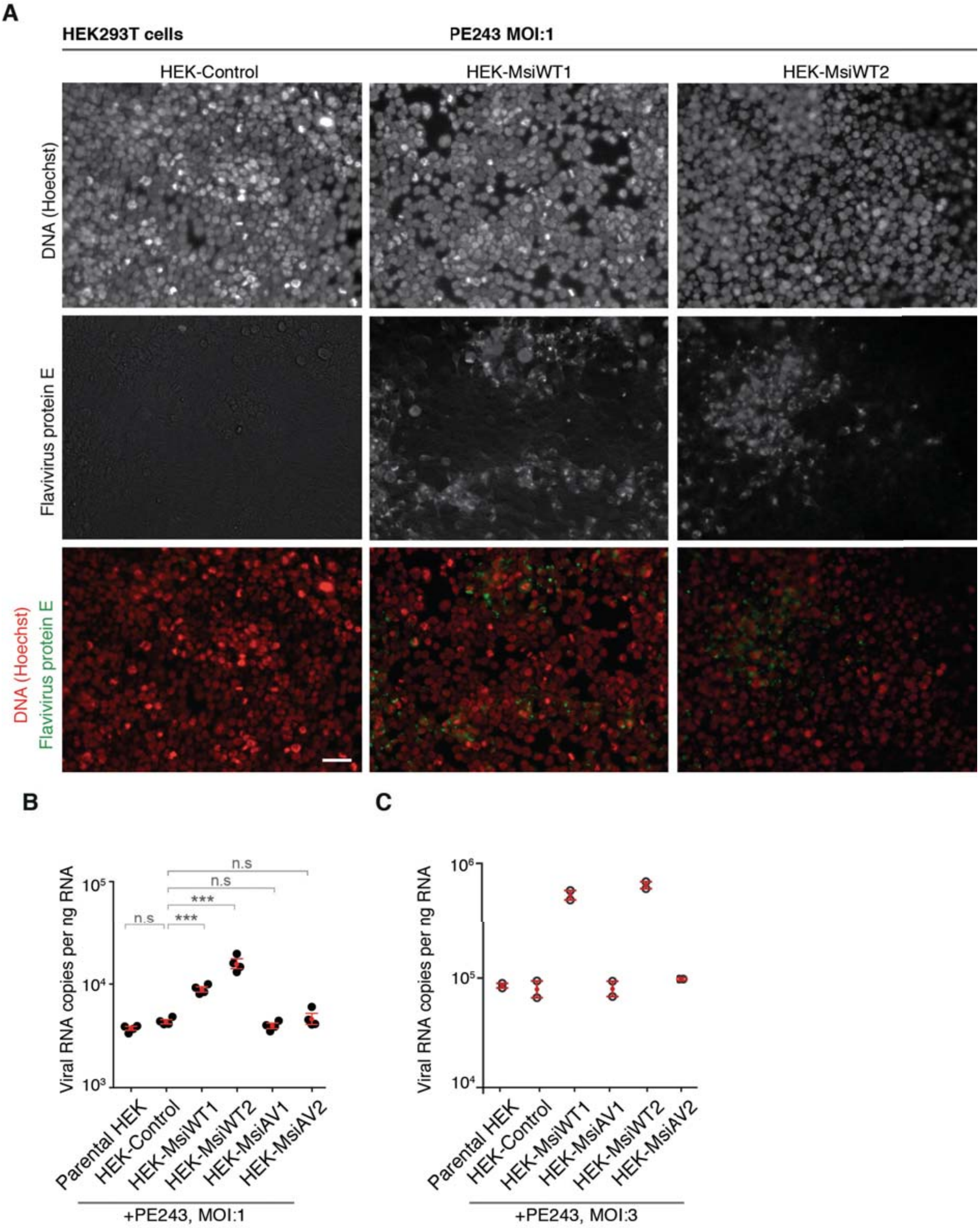


Fig. S12. MSI1 expression increases susceptibility of HEK293T cells to ZIKV infection.

A) Images show parental HEK293T cells, which lack endogenous MSI1, and HEK293T cells expressing a wild type MSI1 transgene (clones reported in Fig. S11) following PE243 ZIKV infection. MOI:1 FFU/cell, 36h. Cells were stained with an antibody against flavivirus protein E. DNA is stained with Hoechst. Scale bar=50µm.

B) Viral RNA levels in HEK293T cells following infection with PE243 (MOI: 1 FFU/cell, 48h). n=3 biological replicates. P-values were obtained from Student t-test, unpaired, two-tailed: * $p<0.05$, ** $p<0.005$, not significant (n.s.).

C) Viral RNA load in HEK293T cells following infection with PE243 (MOI: 3 FFU/cell, 48h). n=2 biological replicates.

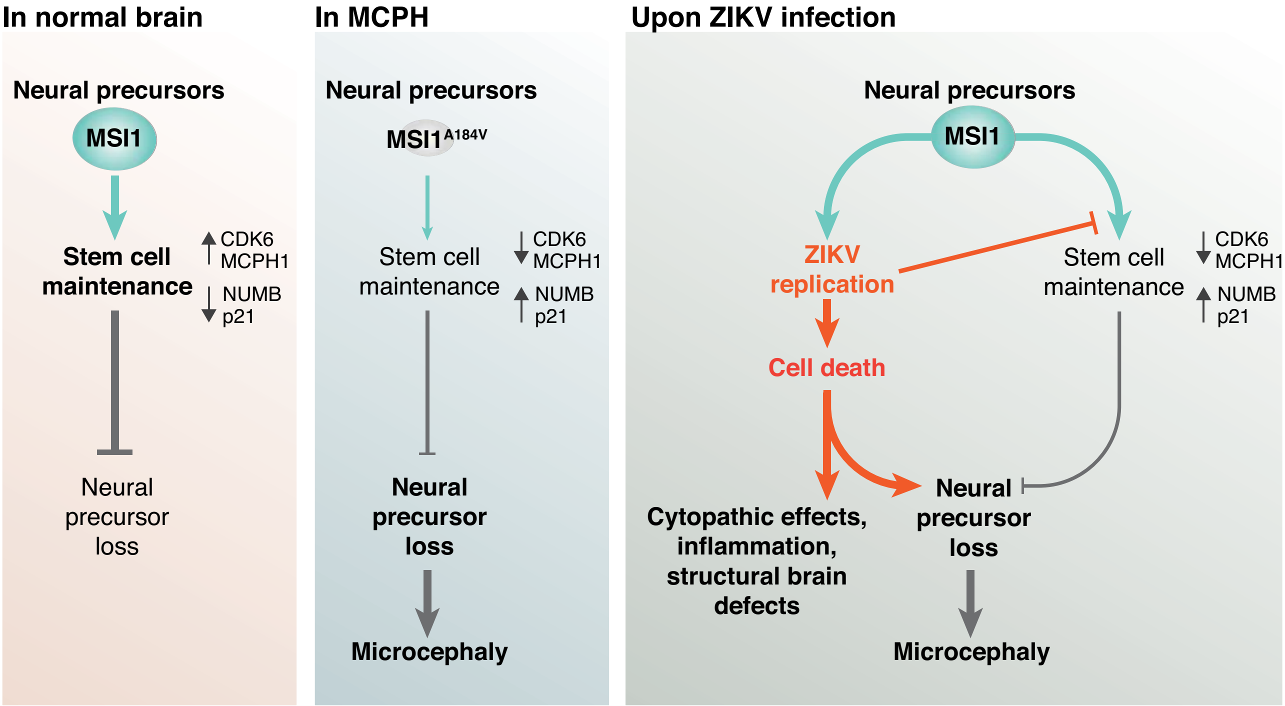


Fig. S13. Proposed model to explain the role of MSI1 in MCPH and ZIKV-induced microcephaly. During normal fetal development (left panel) MSI1 maintains proliferation of neural precursors, a role that involves translational control of its cellular targets. It promotes expression of the MCPH proteins, CDK6 and MCPH1, and represses expression of NUMB and p21. In MSI1^{A184V} patients (middle panel) mutant MSI1 does not bind its RNA targets effectively, thereby deregulating their expression. This impedes normal proliferation of neural precursors, produces fewer neurons, which causes microcephaly. Upon ZIKV infection (right panel), MSI1 binds the viral genome in neural precursors and stimulates its replication directly and/or indirectly, thus decimating this irreplaceable stem cell population, which leads to microcephaly. Moreover, in those rare cases when stem cells survive the infection, transient sequestration of MSI1 by ZIKV RNA may deregulate MSI1 targets and alter cell fate. In addition, the virus also causes a wide range of cytopathic effects in neural progenitors and surrounding cells, leading to inflammation and structural brain defects.

Table S1. TCNAG2.0 results listing the shared homozygous and concordant regions present in both affected children. All chromosome locations based on NCBI36/hg18 human genome assembly. The homozygous segment containing the pathogenic mutation is shown in bold.

Chromosome bands	Size of region	Start of region	End of region
1p21.3 – p13.1	19.5Mb	97,640,328	117,188,024
2q31.3 – q34	30Mb	180,421,369	210,469,387
2q37.2 - q37.3	6.6Mb	236,067,214	242,717,659
3q13.11 - q25.2	48.6Mb	107,683,064	156,303,776
4q22.1 - q24	18.4Mb	88,866,848	107,281,230
5q11.2 - q12	10.4Mb	53,499,547	63,910,619
5q13.3 - q22.2	36.7Mb	75,279,191	112,004,899
8p11.23 – q12.3	24.7Mb	38,886,941	38,886,941
12q15 - q21.1	4.6Mb	67,133,823	71,772,466
12q23.2 - q24.31	**23.3Mb**	**101,470,766**	**124,807,539**
13q31.1 – q32.2	16.2Mb	81,857,944	98,119,056
17q23.2 – q24.2	10Mb	53,183,878	63,168,794
18p11.32 – q1	19.5Mb	833,971	20,311,736
18q21.33 – q22.3	11.8Mb	59,322,883	71,109,557

Table S2. The four homozygous mutations considered after autozygosity mapping and exome sequencing. All mutations were non-synonymous. Position as in GRCh37/hg19. The transcripts reported were NM_014420.2, NM_138287.3, NM_002442.2 and NM_001093.3. PolyPhen and SIFT were used to predict effects of amino acid substitutions. Brainspan data from http://www.brainspan.org indicates expression levels during human fetal brain development.

Position	**Change**	**Depth**	**Gene**	**AA change**	**PolyPhen**	**SIFT**	**Brainspan**	**Mouse model**
8:42233261	G>A	20	DKK4	p.Arg67 Cys	deleterious/0	probably damaging/1	very low	no phenotype
3:122287775	C>T	33	DTX3L	p.Thr280 Ile	deleterious/0.01	possibly damaging/0.467	very low	n/a
12:120794806	G>A	26	MSI1	p.Ala184 Val	deleterious/0	probably damaging/0.997	high	defects in neuro-development
12:109704152	C>G	23	ACACB	p.Ser2454 Arg	deleterious/0.01	possibly damaging/0.269	very low	n/a

Table S3. List of putative MSI1 binding sites in flavivirus 3’ UTRs. G/AU(1-3)AG consensus sites were mapped onto structural prediction of 3’UTRs from various flaviviruses from http://sergio14.github.io/Flaviviruses_RNA_Structures/Flavivirus-3UTR.svg#19 and (*16*). Only sites on accessible (i.e. not blocked by predicted pseudoknots) loop structures are listed. MBV:mosquito-borne viruses; TBV: tick-borne viruses; ISV: insect-specific viruses; NKV: no known vector viruses.

Virus	**Msi1 consensus**	**Number of sites**
MBV		
ZKV	AUAG(2), GUUAG	3
DENV1	GUAG, GUUAG(2)	3
DENV2	GUAG, GUUAG(2)	3
DENV3	GUAG, GUUAG(2)	3
DENV4	GUUAG	2
KEDV	GUUAG	1
ROCV	GUUAG(2)	2
MVEV	AUAG, GUAG, GUUAG(3)	5
KUNV	GUAG,GUUAG(2)	3
ALFV	GUUAG(2)	2
JEV	GUAG,GUUAG(2)	3
KOKV	GUUAG(2), AUUAG	3
IGUV	GUUAG	1
USUV	AUUUAG(2),GUUUAG(2)	2
BAGV	AUAG	1
WNV	AUAG, GUAG	2
SEPV	GUUAG	1
YFV	GUUAG	1
TBV		
TBEV(n)	GUAG	1
AHFV		0
LIV	AUAG	1
LGTV	AUAG	1
POWV	GUUAG	1
TBEV(V)		0
ISV		
CxFV	AUAG,AUAG,GUUAG	3
CFAV	AUAG,GUAG,GUUAG,AUAG	4
NKV		
MMLV	AUUAG,GUAG,GUUAG	3

APOIV	GUUAG	2
RBV	AUAG,GUUAG	2
MODV	AUAG,AUUAG,GUUAG	3

Table S4. Table of primers used to generate plasmids and perform qPCR in this study.

Plasmid	Forward primer	Reverse primer
PCDNA MSI WT	CCAGGATCCATGGAGACTGACGCGCCCCAG	GCGGCCGCTAGTGGTACCCATTGGTGAAGC
PCDNA MSI1A>V	TGGAATGTAAGAAAGTTCAGCCAAAGGAGGT	ACCTCCTTTGGCTGAACTTTCTTACATTCCA
Gst-MSI1 (Gateway)	GGGGACAAGTTTGTACAAAAAAGCAGGCT TCATGGAGACTGACGCGCCCCAG	GGGGACCACTTTGTACAAGAAAGCTGGG TCCTAGTGGTACCCATTGGTGAAGC
MCPH1_L (Gateway)	GGGGACAAGTTTGTACAAAAAAGCAGGCTT CAATGGCGGCCCCCATCCTGAAA	GGGGACCACTTTGTACAAGAAAGCTGGGT CCTATTGTGACAATAGGTAGTTTTCAGG
MCPH1_S	-	TCCTATCACATACTTCCACTGTATC
Pjet MCPH1_L UTR	CAGTGACCTCACTGGCCTG	CTGTGCTCTAATGACAAAGGT
Pjet PE243 NCR	CACCAATCTTAATGTTGTC	AGACCCATGGATTTCCCC
Pjet MR766 NCR	CACCAATTTTAGTGTTGTC	AGAAACCATGGATTTCCC
Pjet Mut_NCR	S1F: GAAACCAAGCCTATCCTCAGGCCGAGAACG S2F: GACTAGTGGTTCCAGGAGACCC S3F: GATCGCCGAATCCCGGCGGCCGG	S1R:CGTTCTCGGCCTGAGGATAGGCTTGGTTTC S2R: GGGTCTCCTGGAACCACTAGTC S3R: CCGGCCGCCGGGATTCGGCGATC
qPCR primers		
MCPH1_1 CDK6 GAPDH NUMB p21	Quantitect primer assays (Qiagen)	QT00073395 QT00019985 QT00079247 QT00078442 QT00062090
MCPH1_S	ACTTGCTGTCTTGTGGAACTTCTA	ACTTCACTCCCCTCAGTTATTCAT
GAPDH	CAACAGCCTCAAGATCATCAG	ATGGACTGTGGTCATGAGTC
ZIKV (9519—9681)	GTGGTGCAACTCATTCGGAA	TTGGCTTCACAACGCAATCA