Supplemental Data
Cell Host & Microbe, *Volume 5*

A Functional Genomic Screen Identifies Cellular Cofactors of Hepatitis C Virus Replication

Andrew W. Tai, Yair Benita, Lee F. Peng, Sun-Suk Kim, Naoya Sakamoto, Ramnik J. Xavier, and Raymond T. Chung

Supplemental Experimental Procedures

Cell Culture

Huh7, Huh7.5.1 and 293T cells were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS), 100 U/mL of penicillin, and 100 μg/mL of streptomycin. Huh7/Rep-Feo and OR6 replicon cells were grown in the same medium with 400 μg/mL of G418. 293GPG cells were grown in DMEM with 10% FBS, 300 μg/mL G418, 1 μg/mL tetracycline, and 2 μg/mL puromycin. UHCVcon57.3 cells were maintained in DMEM, 10% FBS, 500 μg/mL G418, 1 μg/mL tetracycline, and 1 μg/mL puromycin.

High-Throughput Screening: Hit Selection

Luminescence values were log-transformed and then subtracted by the median value of the experimental and negative control wells. These values were then divided by the standard deviation of the plate well values to yield Z-scores. Luciferase Z-scores on duplicate plates were averaged and then converted to p-values with the assumption that the values were normally distributed. The uncorrected p-values were converted to q-values using the software package QVALUE (Storey, 2002). A q-value of 0.10, corresponding to a false-discovery rate of 0.10, was used as the threshold for selecting hit siRNA pools from the primary screen for siRNA deconvolution.

Gene Ontology

Gene ontology terms and gene associations were obtained from the Gene Ontology web site (www.geneontology.org; ontologies revision: 5.778; gene associations revision: 5/27/08) (Ashburner et al., 2000). Uniprot and VEGA gene identifiers were mapped to NCBI gene identifiers. In cases where multiple IDs matched the same NCBI gene, all gene ontology terms from these IDs were combined and assigned to the NCBI gene. We accepted biological process terms that were assigned to less than 500 genes each. Statistical analysis was performed using a hypergeometric distribution as described in the GOHyperGAll module of BioConductor (Gentleman et al., 2004). The Gene Ontology terms are arranged in a tree structure with a single root node. To simplify the representation of terms, terms that were significantly enriched with a p-value < 0.05 and connected in the tree hierarchy were combined to form an over-represented cluster of connected terms. All the genes annotated within that cluster of terms were represented by the most significant term in the cluster. To further reduce the redundancy within the Gene Ontology tree, the clusters were ordered based on p-values and if the genes in one cluster were fully contained within another more significant cluster, only the most significant cluster was accepted. Finally, we accepted clusters for which at least two genes were assigned.

Subcellular Localization and Gene Annotation

Annotation information for each gene was obtained from UniProt (Bairoch et al., 2005), Gene Ontology (Ashburner et al., 2000), NCBI Reference Sequence (Pruitt et al., 2005), NCBI OMIM (McKusick, 2007), NCBI GeneRIF (Wheeler et al., 2007), KEGG Pathways (Kanehisa and Goto, 2000), NetPath (http://www.netpath.org), and Reactome (Joshi-Tope et al., 2005). For each data source genes were mapped to NCBI gene IDs. Statistical significance was calculated for each KEGG, NetPath and Reactome pathway

using the hypergeometric distribution as described for GO terms. For each gene the homolog proteins in other species were identified using NCBI HomoloGenes (Wheeler et al., 2007) and annotation information for the homologs was obtained from NCBI Reference Sequence. PubMed publications linked to each HomoloGene cluster were reviewed to assess gene function and localization.

Protein-protein interactions were obtained from the NCBI Entrez Gene database (Wheeler et al., 2007). Non-human protein interactions were mapped to matching human homologs through HomoloGene. For each of the genes that scored in the screen, the protein interactors were reviewed to identify interactions with HCV proteins or with human proteins that were reported to interact with HCV proteins. In some cases where a gene had multiple functions, the function that was most relevant to HCV infection was accepted.

Supplemental References

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T.*, et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M.*, et al.* (2005). The Universal Protein Resource (UniProt). Nucleic Acids Res *33*, D154-159.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J.*, et al.* (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome biology *5*, R80.

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G.R., Wu, G.R., Matthews, L.*, et al.* (2005). Reactome: a knowledgebase of biological pathways. Nucleic Acids Res *33*, D428-432.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

McKusick, V.A. (2007). Mendelian Inheritance in Man and its online version, OMIM. Am J Hum Genet *80*, 588-604.

Pruitt, K.D., Tatusova, T., and Maglott, D.R. (2005). NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res *33*, D501-504.

Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., Dicuccio, M., Edgar, R., Federhen, S.*, et al.* (2007). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res.

Figure S1

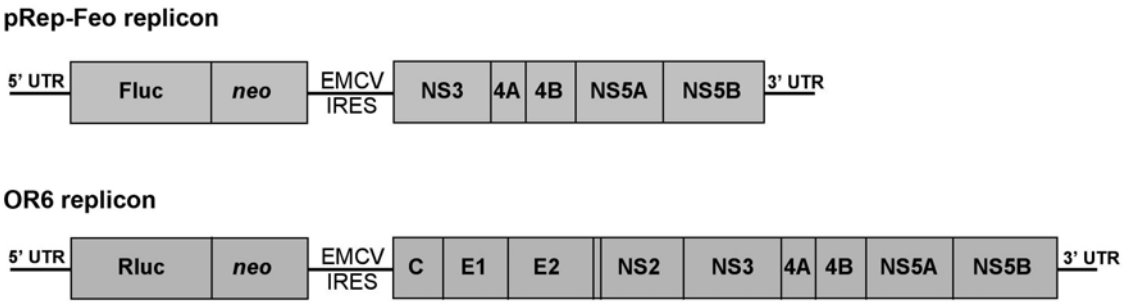


Figure S1. Schematic of the pRep-Feo Subgenomic Replicon and the OR6 Full-Length Replicon
pRep-Feo encodes a firefly luciferase reporter, while OR6 encodes a Renilla luciferase reporter upstream of HCV sequences. C, HCV core; E1 and E2, envelope proteins; NS2-NS5B, nonstructural proteins.

Figure S2

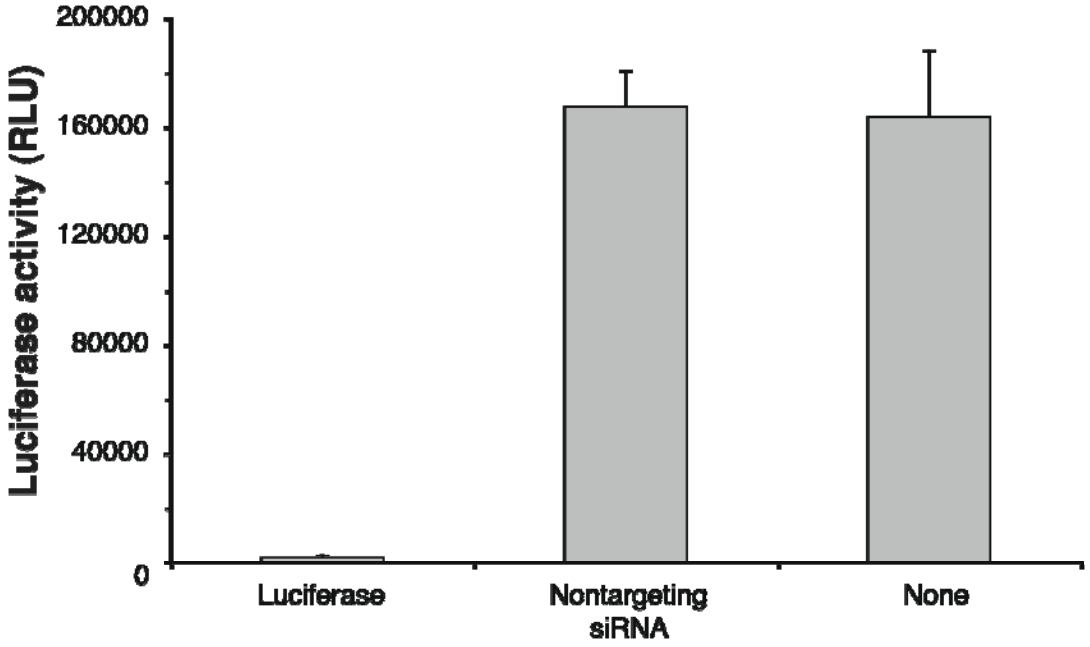


Figure S2. Validation of High-Throughput siRNA Transfection
Huh7/Rep-Feo cells were transfected in 384-well format with no siRNA, a nontargeting control siRNA, or a positive control luciferase siRNA. Firefly luciferase activity was assayed 72 hr after transfection. Values represent mean of 96 wells ± SD.

Figure S3

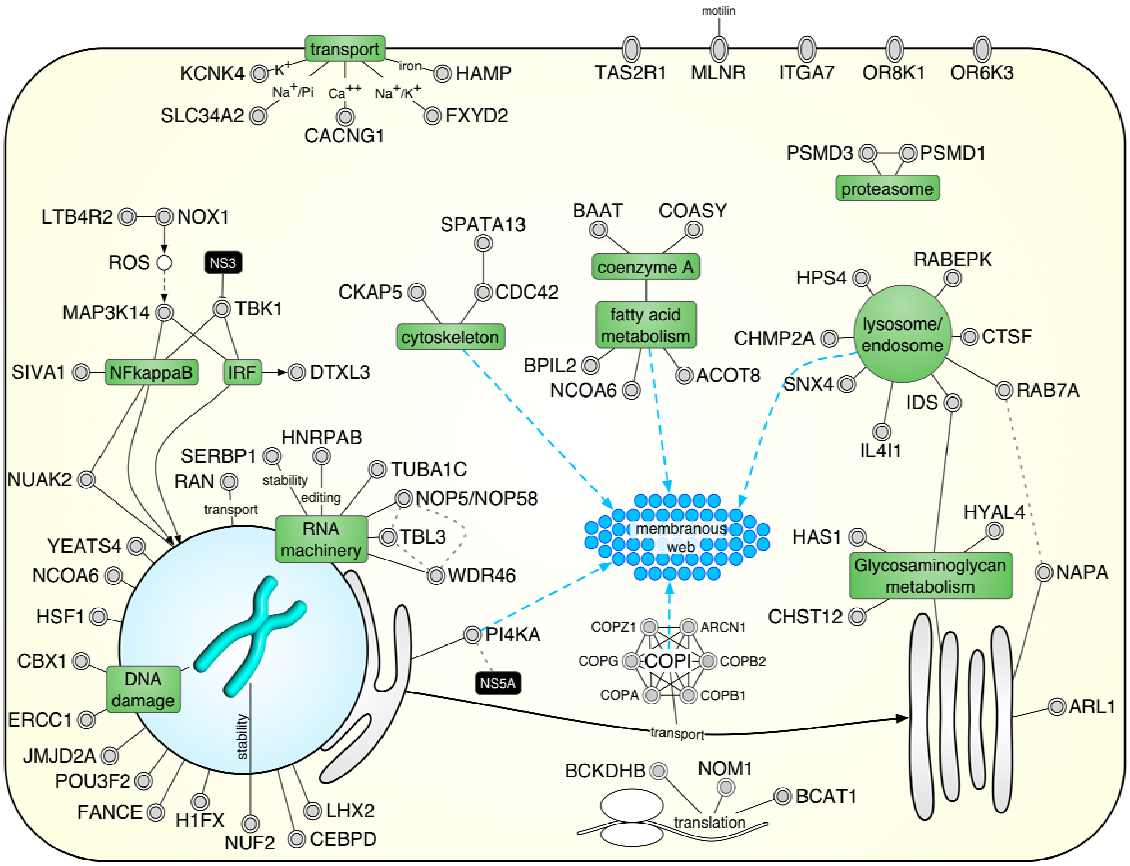


Figure S3. Functional and Subcellular Localization Mapping of HCV Cofactors
The scientific literature and gene annotation for each of the 96 HCV cofactors that were identified in the screen was reviewed (see Experimental Procedures). Genes were preferentially placed according to their functional role that was most relevant in the context of HCV replication. Black arrows indicate activation; dotted gray lines indicate protein-protein interactions; dotted blue lines indicate processes that impinge on membranous web structures; green boxes indicate cellular processes or compartments and black boxes indicate HCV proteins. Supporting references are listed in Table S4.

Figure S4.

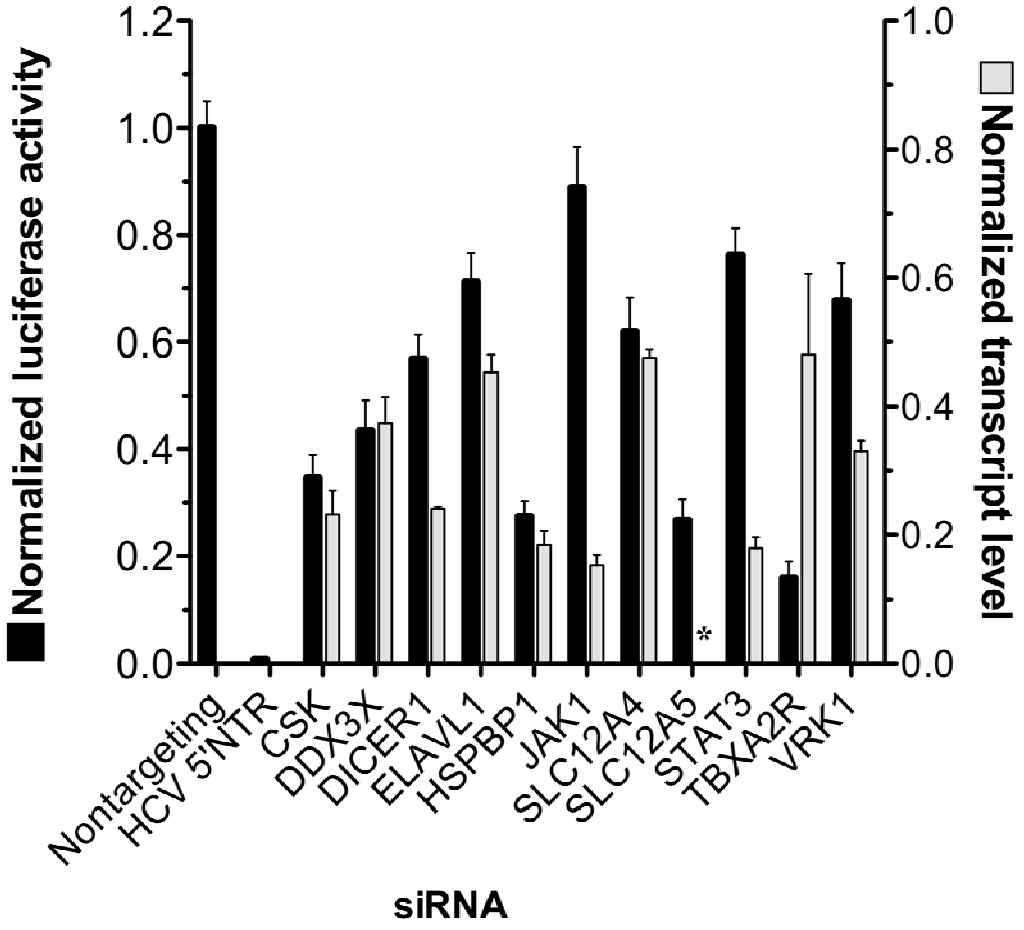


Figure S4. Retesting of HCV Replication Cofactors Identified in Prior siRNA Screens in Huh7/Rep-Feo Cells

Huh7/Rep-Feo cells were transfected in 96-well format with the indicated siRNA pools at 50 nM for 72 hr. Wells were assayed for luciferase activity (black bars) and the target transcript levels were assayed by qPCR (gray bars). A nontargeting siRNA and HCV 5'NTR siRNA were used as a negative and positive control, respectively. The asterisk for *SLC12A5* mRNA levels indicates that we were unable to identify a *SLC12A5* transcript by RT-PCR in Huh7/Rep-Feo cells using two independent primer sets. The luciferase data were obtained from quadruplicate assays in three independent experiments and are expressed as mean ± SEM. The qPCR data are from triplicate reactions from duplicate wells and are expressed as mean ± SD.

Figure S5.

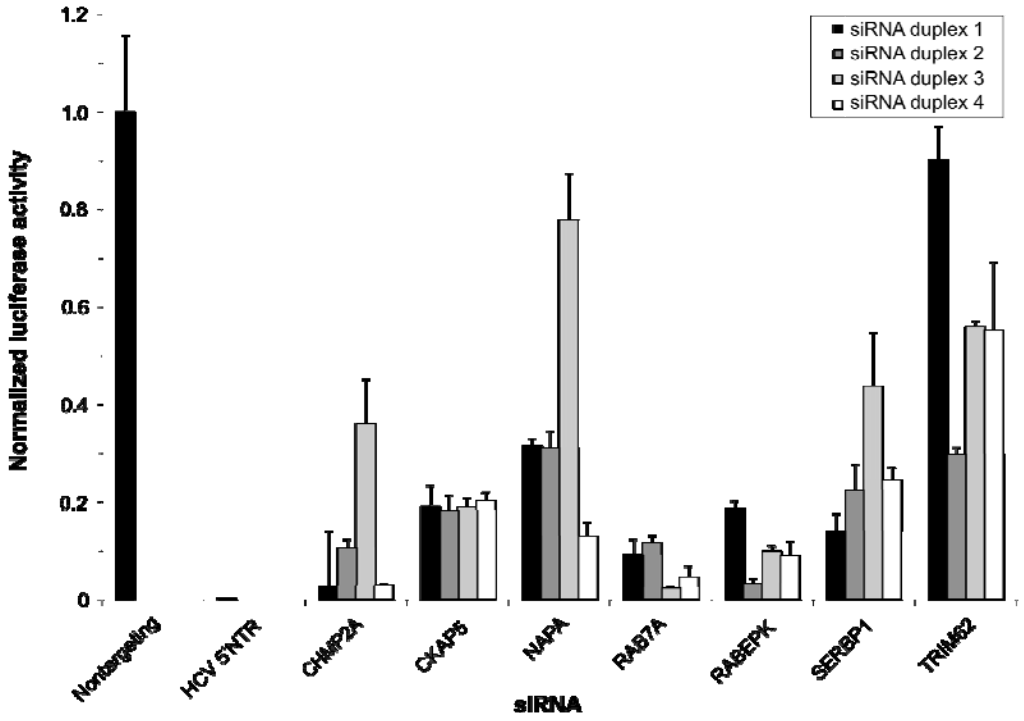
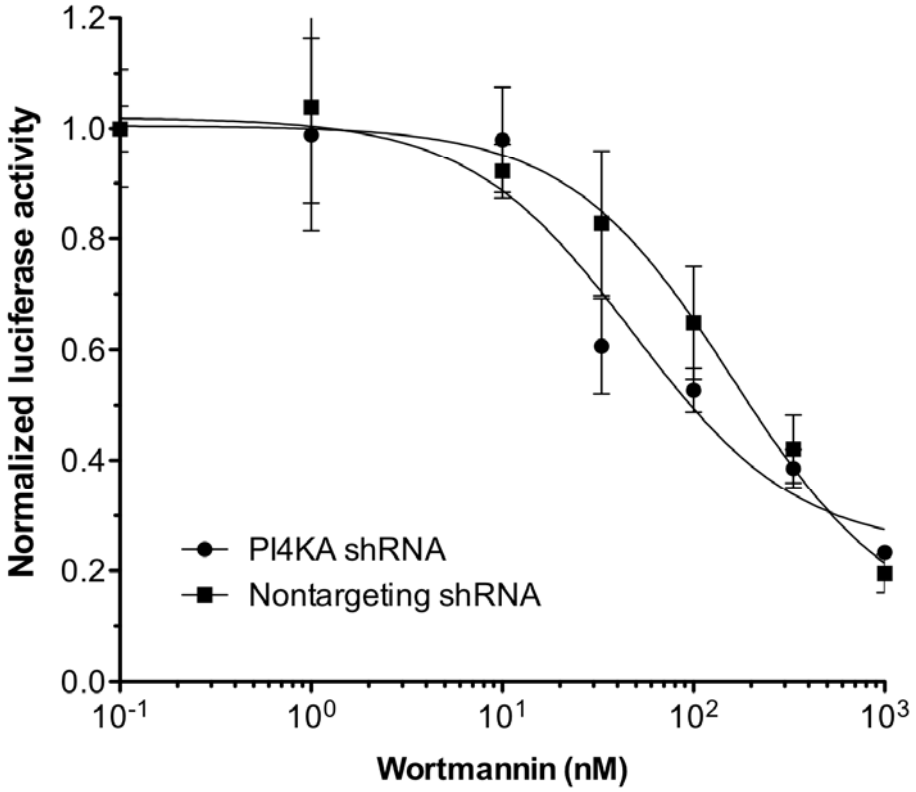


Figure S5. Reconfirmation of Candidate HCV Replication Cofactors in the OR6 Full-Length Replicon Model

The indicated genes were silenced with four individual siRNA duplexes in the OR6 full-length replicon cell line. OR6 cells were transfected in 96-well format with the indicated siRNA duplexes at 50 nM for 72 hr. Wells were assayed for *Renilla* luciferase activity. A nontargeting siRNA and HCV 5'NTR siRNA were used as a negative and positive control, respectively. Assays were performed in quadruplicate wells and are expressed as mean ± SD.

Figure S6.



Figures S6. PI4KA Silencing Reduces the IC50 of Wortmannin on HCV Replication
OR6 replicon cells were transduced with a lentiviral vector encoding a PI4KA shRNA (circles) or a nontargeting shRNA (squares). Seventy-two hours after transduction, wortmannin was added at the indicated concentrations for 24 hr, at which time luciferase activity was measured. The observed IC50 was calculated by nonlinear regression with a Hill slope of -1.0 using the Prism 5 software package (GraphPad Software, La Jolla, CA) and was 162 nM for the nontargeting shRNA and 49 nM for the PI4KA shRNA.. All values are mean ± SD.