METHOD ARTICLE

# Integrative vectors for regulated expression of SARS-CoV-2 proteins implicated in RNA metabolism [version 1; peer review: 2 approved]

Stefan Bresson*, Nic Robertson*, Emanuela Sani, Tomasz W Turowski, Vadim Shchepachev, Michaela Kompauerova, Christos Spanos, Aleksandra Helwak, David Tollervey

Wellcome Centre for Cell Biology, University of Edinburgh, Michael Swann Building, Kings Buildings, Mayfield Road, Edinburgh, EH9 3BF, UK

* Equal contributors

**v1** **First published:** 03 Nov 2020, **5**:261
https://doi.org/10.12688/wellcomeopenres.16322.1
**Latest published:** 03 Nov 2020, **5**:261
https://doi.org/10.12688/wellcomeopenres.16322.1

## Abstract

Infection with SARS-CoV-2 is expected to result in substantial reorganization of host cell RNA metabolism. We identified 14 proteins that were predicted to interact with host RNAs or RNA binding proteins, based on published data for SARS-CoV and SARS-CoV-2. Here, we describe a series of affinity-tagged and codon-optimized expression constructs for each of these 14 proteins. Each viral gene was separately tagged at the N-terminus with Flag-$His_8$, the C-terminus with $His_8$-Flag, or left untagged. The resulting constructs were stably integrated into the HEK293 Flp-In T-REx genome. Each viral gene was expressed under the control of an inducible Tet-On promoter, allowing expression levels to be tuned to match physiological conditions during infection. Expression time courses were successfully generated for most of the fusion proteins and quantified by western blot. A few fusion proteins were poorly expressed, whereas others, including Nsp1, Nsp12, and N protein, were toxic unless care was taken to minimize background expression. All plasmids can be obtained from Addgene and cell lines are available. We anticipate that availability of these resources will facilitate a more detailed understanding of coronavirus molecular biology.

## Keywords

COVID-19, fusion proteins, UV crosslinking, RNA

**Open Peer Review**

**Reviewer Status** ✓ ✓

| | Invited Reviewers | |
|---|---|---|
| | **1** | **2** |
| **version 1** 03 Nov 2020 | ✓ report | ✓ report |

1. **Sheila Graham**, University of Glasgow, Glasgow, UK

2. **Chad Petit**, University of Alabama at Birmingham, Birmingham, USA

Any reports and responses or comments on the article can be found at the end of the article.

This article is included in the Coronavirus (COVID-19) collection.

**Corresponding authors:** Aleksandra Helwak (olahelwak@yahoo.com), David Tollervey (D.Tollervey@ed.ac.uk)

**Author roles: Bresson S**: Conceptualization, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Robertson N**: Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Sani E**: Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Review & Editing; **Turowski TW**: Conceptualization, Methodology, Writing – Review & Editing; **Shchepachev V**: Investigation, Methodology, Writing – Review & Editing; **Kompauerova M**: Formal Analysis, Investigation, Validation, Writing – Review & Editing; **Spanos C**: Data Curation, Formal Analysis, Investigation; **Helwak A**: Conceptualization, Formal Analysis, Investigation, Methodology, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Tollervey D**: Conceptualization, Formal Analysis, Funding Acquisition, Project Administration, Supervision, Writing – Original Draft Preparation, Writing – Review & Editing

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was supported by Wellcome through a Principle Research Fellowship to D.T. [077248] and an ECAT Fellowship to NR [213011]. TWT was supported by the Polish Ministry of Science and Higher Education Mobility Plus program [1069/MOB/2013/0]. VS was funded by the Swiss National Science Foundation [P2EZP3_159110]. Work in the Wellcome Centre for Cell Biology is supported by a Centre Core grant [203149].
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*



**How to cite this article:** Bresson S, Robertson N, Sani E *et al.* **Integrative vectors for regulated expression of SARS-CoV-2 proteins implicated in RNA metabolism [version 1; peer review: 2 approved]** Wellcome Open Research 2020, **5**:261 https://doi.org/10.12688/wellcomeopenres.16322.1

**First published:** 03 Nov 2020, **5**:261 https://doi.org/10.12688/wellcomeopenres.16322.1

## Introduction

SARS-CoV-2 is a large positive-sense, single-stranded RNA virus that encodes four structural proteins, several accessory proteins, and sixteen nonstructural proteins (nsp1-16) (Figure 1). The latter are mainly engaged in enzymatic activities important for translation and replication of the RNA genome. In addition, nonstructural proteins also target host RNA metabolism in order to manipulate cellular gene expression or facilitate immune evasion (Gordon *et al.*, 2020; Thoms *et al.*, 2020; Yuen *et al.*, 2020). Understanding these pathways in greater detail will be an important step in the development of antiviral therapies.

Our lab has previously developed techniques to map the RNA-bound proteome, including crosslinking and analysis of cDNA (CRAC), to identify binding sites for proteins on RNA, and total RNA-associated proteome purification (TRAPP) to identify and quantify the RNA-bound proteome (Granneman *et al.*, 2009; Shchepachev *et al.*, 2019). Among other things, TRAPP has given insights into the mechanism of stress-induced translation shutdown in yeast (Bresson *et al.*, 2020b). In the context of SARS-CoV-2, CRAC is expected to identify host RNAs that are targeted by viral factors, while TRAPP will reveal how host cell RNA-protein interactions are globally remodeled in response to specific viral proteins.

To facilitate application of these techniques to SARS-CoV-2, we generated a series of synthetic, codon-optimized constructs for 14 different viral proteins that are expected to interact with RNA or RNA binding proteins. To remove the need for error-prone PCR steps, we devised a cloning scheme in which a single synthetic construct could be used to generate untagged, N-, or C-terminally tagged versions of the protein. Using these vectors, we generated and tested a series of human cell lines with the viral open reading frames (ORFs) stably integrated, under the control of an inducible Tet-On promoter. We expect that broad availability of this collection will enable a deeper understanding of the coronavirus life cycle and its impact on host RNA biology.

## Results and discussion

### Design and cloning of viral expression constructs

We selected 14 proteins for initial analysis, based on putative roles in viral or host RNA metabolism (Figure 1 and Table 1). Two of the selected proteins, Nsp7 and Nsp8, reportedly form a stable heterodimer *in vivo* (Gao *et al.*, 2020; Hillen *et al.*, 2020; te Velthuis *et al.*, 2012), so we also designed constructs in which Nsp7 and Nsp8 were expressed as a fusion protein, connected by a short, unstructured linker.

For cloning, we selected pcDNA5-FRT/TO (Thermo Fisher) as the backbone vector (Figure 2A). This vector can be used for transient transfection or flippase (Flp) recombinase-mediated integration into the genome of cells with a pre-inserted Flp Recombination Target (FRT). It also carries a hygromycin resistance gene to allow selection for integration (Figure 2). As host cells, we used HEK293 Flp-In T-REx cells (293FiTR), but other cell lines that carry an FRT site could also be used.

The expression levels of viral proteins vary substantially during infection, so we used constructs that, in addition to stably integrating, were expressed under the control of a tetracycline-regulated human cytomegalovirus (CMV)/TetO2 promoter, which is induced by the addition of doxycycline to the medium (Figure 2). Viral protein expression can then be titrated by varying either doxycycline concentration or induction time.

Using pcDNA5-FRT/TO as a starting point, we generated two additional parental vectors with pre-inserted tandem-affinity purification tags, either an N-terminal, FH-tag (FLAG-$Ala_4$-$His_8$) or C-terminal HF-tag ($His_8$-$Ala_4$-FLAG) (Figure 2A). We have recently shown that these tags work well for tandem affinity purification, including in the denaturing conditions used for CRAC (Bresson *et al.*, 2020b).

Each synthetic construct was codon-optimized and included a consensus Kozak sequence upstream of the start codon

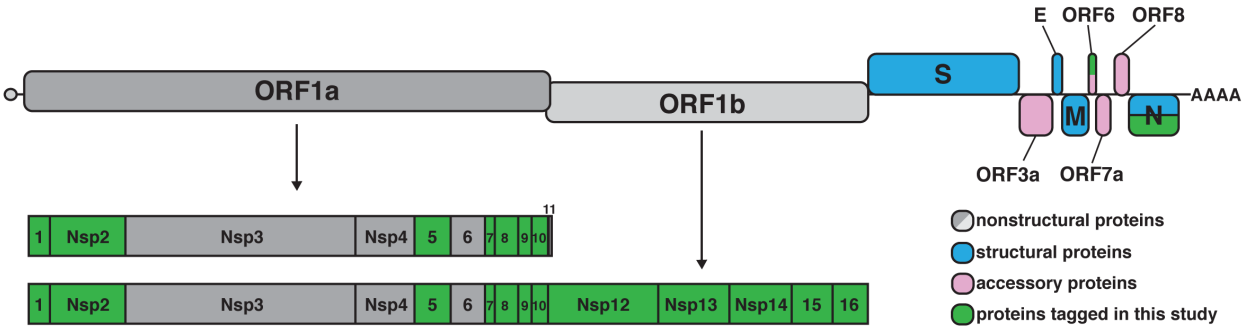


**Figure 1. SARS-CoV-2 genome organization.** The viral genome consists of a ~30 kb positive-sense transcript that is capped and polyadenylated. Two overlapping open reading frames (ORFs) are translated from the genomic RNA, ORF1a and ORF1ab. Translation of the ORF1b region is mediated by a -1 frameshift allowing readthrough of the stop codon at the end of ORF1a. ORF1a and ORF1ab encode large polyproteins that are cleaved into 16 nonstructural proteins. Structural and accessory proteins (shown in blue and pink, respectively) are separately translated from subgenomic RNAs. ORFs selected for tagging are highlighted in green.

**Table 1. Predicted molecular weights and putative functions of selected open reading frames (ORFs).**

| ORF | Size (aa) | Putative function | References |
|---|---|---|---|
| Nsp1 | 180 | Suppresses host translation; binds 40S ribosomal subunit | (Huang *et al.*, 2011; Lokugamage *et al.*, 2012; Tanaka *et al.*, 2012; Thoms *et al.*, 2020) |
| Nsp2 | 638 | Unknown function; binds eIF4E | (Gordon *et al.*, 2020) |
| Nsp5 | 306 | CL3 $M^{pro}$; Protease (3C-like) | (Kneller *et al.*, 2020; Muramatsu *et al.*, 2016) |
| Nsp7 | 83 | RNA polymerase cofactor | (Gao *et al.*, 2020; Hillen *et al.*, 2020) |
| Nsp8 | 198 | RNA polymerase cofactor; poly(A) polymerase | (Gao *et al.*, 2020; Hillen *et al.*, 2020; Tvarogova *et al.*, 2019) |
| Nsp9 | 113 | ssRNA binding | (Egloff *et al.*, 2004; Littler*et al.*, 2020; Sutton *et al.*, 2004) |
| Nsp10 | 139 | Cap methylation complex | (Chen *et al.*, 2011; Krafcikova *et al.*, 2020; Viswanathan *et al.*, 2020) |
| Nsp12 | 932 | RNA polymerase catalytic subunit | (Gao *et al.*, 2020; Hillen *et al.*, 2020) |
| Nsp13 | 601 | RNA helicase; cap tri-phosphatase | (Jang *et al.*, 2020) (Tanner *et al.*, 2003) |
| Nsp14 | 527 | 3′ exoribonuclease (NTD); RNA cap N7 methyltransferase (CTD) | (Ma *et al.*, 2015) (Eckerle *et al.*, 2010; Ogando *et al.*, 2020) |
| Nsp15 | 346 | Uridine-specific RNA endoribonuclease | (Bhardwaj *et al.*, 2004) |
| Nsp16 | 298 | RNA 2′-*O*-methyltransferase | (Chen *et al.*, 2011; Krafcikova *et al.*, 2020; Viswanathan *et al.*, 2020) |
| Orf6 | 61 | Blocks nuclear-cytoplasmic transport; significantly diverged between SARS-CoV and SARS-CoV-2 | (Frieman *et al.*, 2007; Gussow *et al.*, 2020; Li *et al.*, 2020; Sims *et al.*, 2013) |
| N | 418 | Nucleocapsid phosphoprotein | (Cascarina & Ross, 2020; Chang *et al.*, 2014; Cubuk *et al.*, 2020) |

(Figure 2B). The sequences initially used for all ORFs were generated by the algorithms used by Integrated DNA Technologies (IDT; Coralville, Iowa). The prevalence of G-C base pairs, particularly in the third position of codons (GC3), is strongly correlated with increased protein accumulation (Kudla *et al.*, 2006; Mordstein *et al.*, 2020). To potentially enhance protein synthesis, we ordered alternative ORFs for Nsp8, Nsp13, and N, using the algorithms from GeneArt (Thermo-Fisher Scientific), which have a higher G-C content, particularly in third codon positions.

The insert sequences were designed such that a single synthetic construct could be used to generate an untagged, or N- or C-terminal fusion protein (Figure 2B). BamHI and EcoRV restriction sites were placed on either side of the open reading frame, together with an AvrII site overlapping the stop codon (important for C-terminal cloning, as discussed below).

Generating the untagged and N-terminal tagged constructs was straightforward. The synthetic constructs were cut with BamHI and EcoRV and ligated into pre-cut vector, either pcDNA5-FRT/TO (generating an untagged construct), or pcDNA5-FRT/TO-N-Flag-His8 (generating an N-terminally tagged construct) (Figure 2C). The resulting plasmids were verified by colony PCR and Sanger sequencing.

Generating the C-terminal tagged construct required additional steps to remove the in-frame stop codon upstream of the EcoRV restriction site. Cleavage of the AvrII restriction site, which overlaps the stop codon in the synthetic constructs (Figure 2B), left a 5′ overhang containing the stop codon. Subsequently, the overhang (and thus the stop codon) was removed by treatment with Mung Bean Nuclease, an ssDNA endonuclease (Figure 2C). The resulting fragment possessed a blunt 3′ end, compatible with the EcoRV restriction site in the target vector. Subsequently, the 5′ end of the insert was prepared by digestion with BamHI, and the resulting fragment was ligated into pcDNA5-FRT/TO-C-His8-Flag pre-cut with BamHI and EcoRV.

In total, we generated 54 viral protein expression vectors, and three additional GFP expression vectors as controls. These constructs, together with the parental tagging vectors, are listed in Table 2.

## Expression of fusion proteins

Each construct was introduced into Flp-In T-REx cells (Thermo Fisher) by transfection, followed by hygromycin treatment for 10–16 days to select for chromosomal integration (Figure 2E). In the initial experiments, all constructs except the Nsp1 series, and the GC3 (high-expression) optimized versions of untagged N and FH-N protein yielded stable hygromycin-resistant cells (Table 2, column 4). Resistant clones were obtained for FH-N by co-transfecting this construct with a plasmid encoding a tetracycline repressor protein (pcDNA6/TR), to suppress possible high levels of FH-N expression at initial transfection.

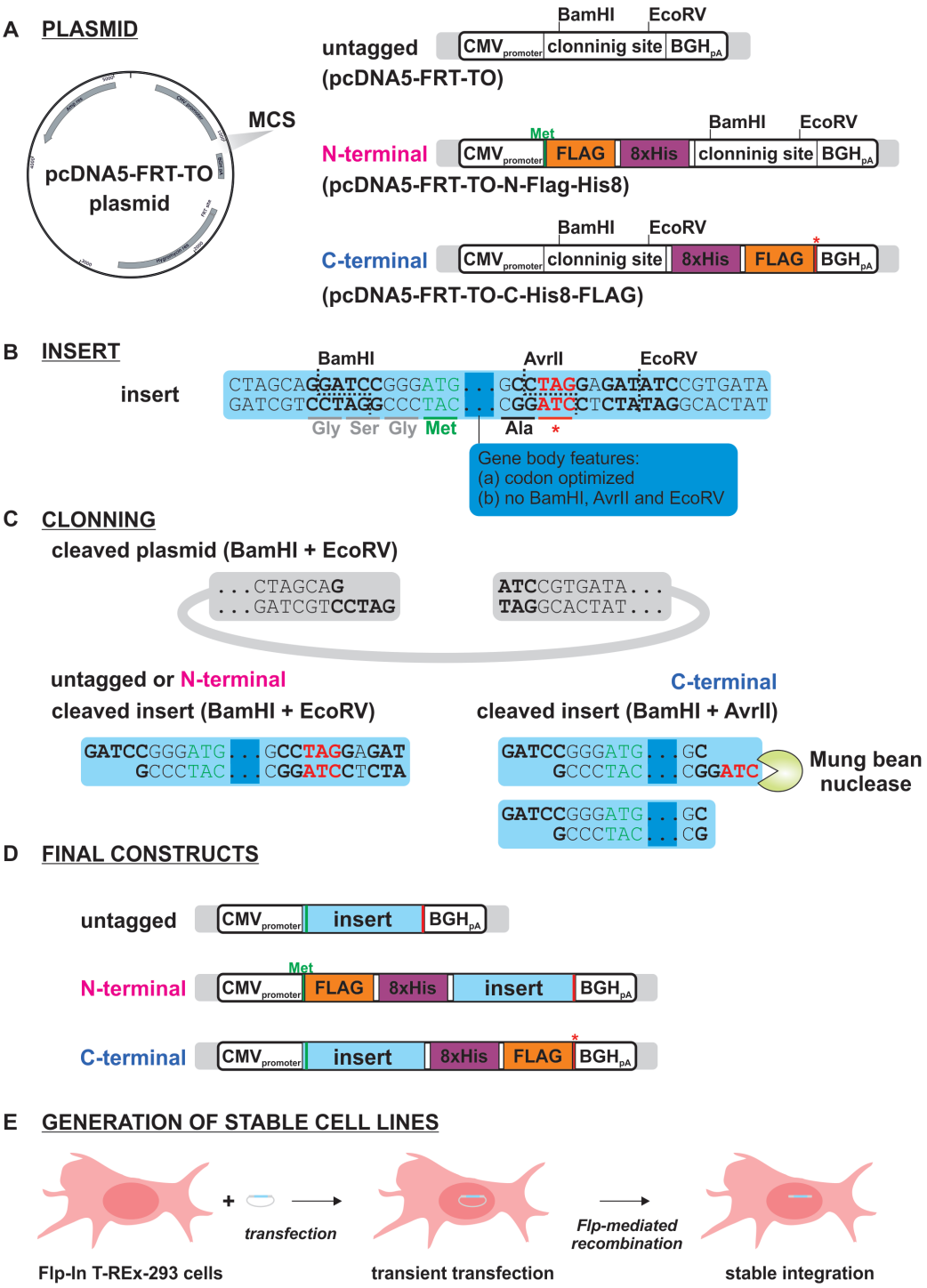


**Figure 2. Schematics illustrating the cloning strategy.** (**A**) The three parental vectors used for generating untagged, N-, and C-terminally tagged constructs. (**B**) Features common to all synthesized insert sequences. Each insert included BamHI and EcoRV sites at either end to facilitate cloning into the three parental vectors. To allow for C-terminal cloning, an AvrII site was inserted such that it overlapped the stop codon (see text for details). In order to accommodate the AvrII site, an alanine residue was added to the end of each expression construct. The viral open reading frames (ORFs) were codon-optimized for moderate or high expression, and lacked BamHI, AvrII, and EcoRV sites. (**C**) For untagged and N-terminal tagging, inserts were digested with BamHI and EcoRV and ligated directly into plasmid precut with the same enzymes. For C-terminal cloning, the inserts were first digested with AvrII, blunted with Mung Bean Nuclease, and then cut with BamHI. The resulting fragment was ligated into plasmid cut with BamHI and EcoRV. (**D**) Untagged, N-, and C-terminally tagged expression constructs. (**E**) Strategy for generating stable cell lines (see text for details).

**Table 2. List of available plasmids and cell lines.**

| Protein | Plasmid number | Tag | Stable cell line | Induction |
|---|---|---|---|---|
| pcDNA5-FRT-TO | | untagged | | |
| pcDNA5-FRT-TO-N-Flag-His8 | pSB084 | N-terminal FH | | |
| pcDNA5-FRT-TO-C-His8-Flag | pSB085 | C-terminal HF | | |
| eGFP | **56** | untagged | yes | |
| eGFP | **55** | N-terminal FH | yes | +++ |
| eGFP | **57** | C-terminal HF | yes | +++ |
| N protein | **10** | untagged | yes | |
| N protein | **22** | N-terminal FH | yes | ++ |
| N protein (GC3 opt.) | **28** | untagged | no | |
| N protein (GC3 opt.) | **31** | N-terminal FH | yes | +++ |
| N protein (GC3 opt.) | **54** | C-terminal HF | yes | +++ |
| Nsp1 | **1** | untagged | no | |
| Nsp1 | **12** | N-terminal FH | no | N/A |
| Nsp1 | **39** | C-terminal HF | no | N/A |
| Nsp2 | **2** | untagged | yes | |
| Nsp2 | **33** | N-terminal FH | yes | ++ |
| Nsp2 | **40** | C-terminal HF | yes | +++ |
| Nsp5 | **3** | untagged | yes | |
| Nsp5 | **13** | N-terminal FH | yes | +++ |
| Nsp5 | **41** | C-terminal HF | yes | +++ |
| Nsp7 | **4** | untagged | yes | |
| Nsp7 | **14** | N-terminal FH | yes | - |
| Nsp7 | **42** | C-terminal HF | yes | + |
| Nsp7-Nsp8 (GC3 opt.) | **26** | untagged | yes | |
| Nsp-Nsp8 (GC3 opt.) | **29** | N-terminal FH | yes | + |
| Nsp7-Nsp8 (GC3 opt.) | **53** | C-terminal HF | yes | +++ |
| Nsp8 | **5** | untagged | yes | |
| Nsp8 | **15** | N-terminal FH | yes | + |
| Nsp8 | **43** | C-terminal HF | yes | - |
| Nsp8 (GC3 opt.) | **35** | untagged | yes | |
| Nsp8 (GC3 opt.) | **37** | N-terminal FH | yes | + |
| Nsp8 (GC3 opt.) | **50** | C-terminal HF | yes | ++ |
| Nsp8-Nsp7 (GC3 opt.) | **27** | untagged | yes | |
| Nsp8-Nsp7 (GC3 opt.) | **30** | N-terminal FH | yes | ++ |
| Nsp8-Nsp7 (GC3 opt.) | **52** | C-terminal HF | yes | +++ |
| Nsp9 | **6** | untagged | yes | |

| Protein | Plasmid number | Tag | Stable cell line | Induction |
|---|---|---|---|---|
| Nsp9 | **16** | N-terminal FH | yes | - |
| Nsp9 | **44** | C-terminal HF | yes | + |
| Nsp10 | **7** | untagged | yes | |
| Nsp10 | **17** | N-terminal FH | yes | - |
| Nsp10 | **45** | C-terminal HF | yes | - |
| Nsp12 | **18** | N-terminal FH | yes | |
| Nsp12 | **25** | untagged | yes | + |
| Nsp12 | **46** | C-terminal HF | yes | + |
| Nsp13 | **32** | untagged | yes | |
| Nsp13 | **34** | N-terminal FH | yes | + |
| Nsp13 (GC3 opt.) | **36** | untagged | yes | |
| Nsp13 (GC3 opt.) | **38** | N-terminal FH | yes | + |
| Nsp13 (GC3 opt.) | **51** | C-terminal HF | yes | +++ |
| Nsp14 | **23** | untagged | yes | |
| Nsp14 | **19** | N-terminal FH | yes | + |
| Nsp14 | **47** | C-terminal HF | yes | + |
| Nsp15 | **8** | untagged | yes | |
| Nsp15 | **20** | N-terminal FH | yes | ++ |
| Nsp15 | **48** | C-terminal HF | yes | ++ |
| Nsp16 | **9** | untagged | yes | |
| Nsp16 | **21** | N-terminal FH | yes | ++ |
| Nsp16 | **49** | C-terminal HF | yes | +++ |
| Orf6 | **11** | untagged | yes | |
| Orf6 | **24** | N-terminal FH | yes | - |

For expression analyses, doxycycline was added to the cell culture medium to a final concentration of 1 µg $ml^{-1}$, and time-points were collected for analysis by western blot. A prominent cross-reacting band visible in all samples was used as a loading control for most analyses. The exceptions were Nsp15 and Nsp12, for which anti-M2-Flag and GAPDH were used, as they have migration positions close to the cross-reacting band. Representative data are shown in Figure 3B–C; all other western scans are shown in Figure 4. Quantitation is shown in Figure 5, and associated raw data are provided as *Underlying data*. Clear protein expression was observed for 29 of the 34 tagged constructs (Table 2, column 5). In general, C-terminal tagged proteins were more highly expressed than their N-terminally tagged counterparts.

As described above, we were initially unable to generate stable cell lines for any of the Nsp1 constructs. To confirm that Nsp1 could be expressed in cells, we transiently transfected each construct into 293FiTR cells, and confirmed protein expression by western blot. Only the C-terminally tagged Nsp1 showed robust expression (Figure 6).

To allow absolute quantification of SARS-CoV-2 protein expression, we used the N-terminally tagged N protein (integrated from plasmid 22; moderate expression) as a reference standard. Cells containing integrated FH-N were treated with doxycycline for 0, 6, and 18 hours. Protein was extracted, separated by SDS-PAGE and analyzed using mass spectrometry with label-free quantification. To compare proteins, their abundance was expressed as a percentage of the total proteome. This value was calculated using the relative, intensity-based absolute quantification (riBAQ) score for each protein, which represents the iBAQ score for a given protein divided by the iBAQ scores for all proteins. After induction for six hours, N protein comprised 0.2% of the cellular proteome (Figure 3A).

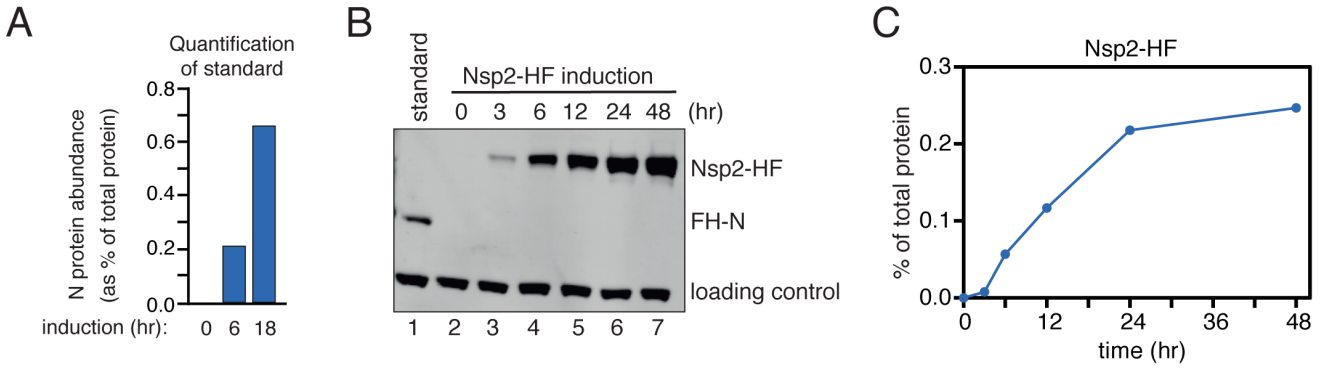


**Figure 3. Viral protein induction.** (**A**) Quantification of N protein abundance following a 0, 6, or 18 h induction. Protein levels were assessed by mass spectrometry and label-free quantification, and expressed as a percentage of the total cellular proteome. (**B**) A representative blot showing induction for C-terminally tagged Nsp2 (lanes 2–7). A prominent cross-reacting band was used as a loading control. To quantify protein abundance, each blot included a normalization standard (lane 1), consisting of lysate from cells expressing N-terminally tagged N protein for 6 h derived from (**A**). (**C**) Quantification of the protein bands in (**B**). Expression was normalized to both the loading control and the N protein standard.

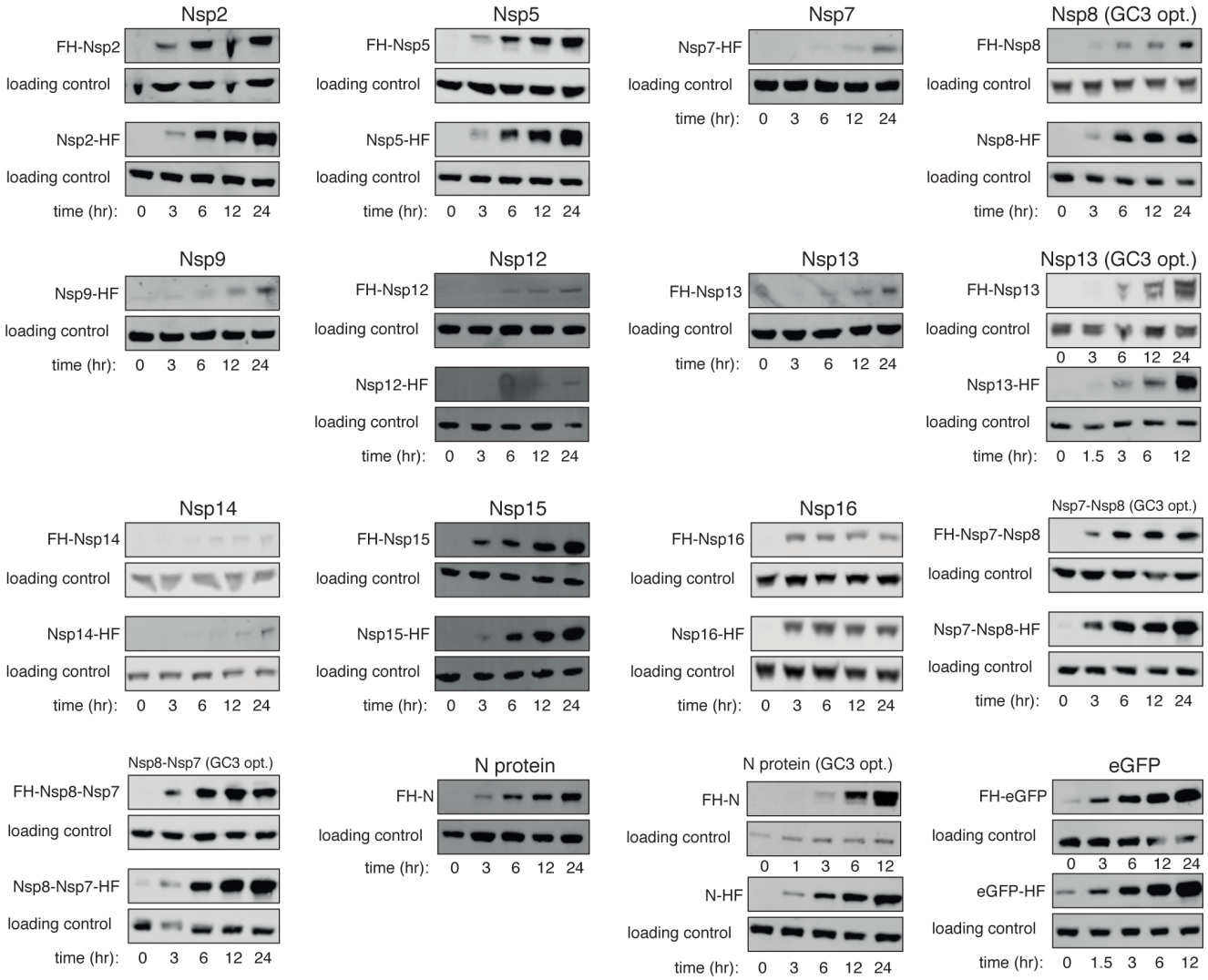


**Figure 4. Viral protein induction western blots.** As Figure 3B, but showing the proteins indicated.

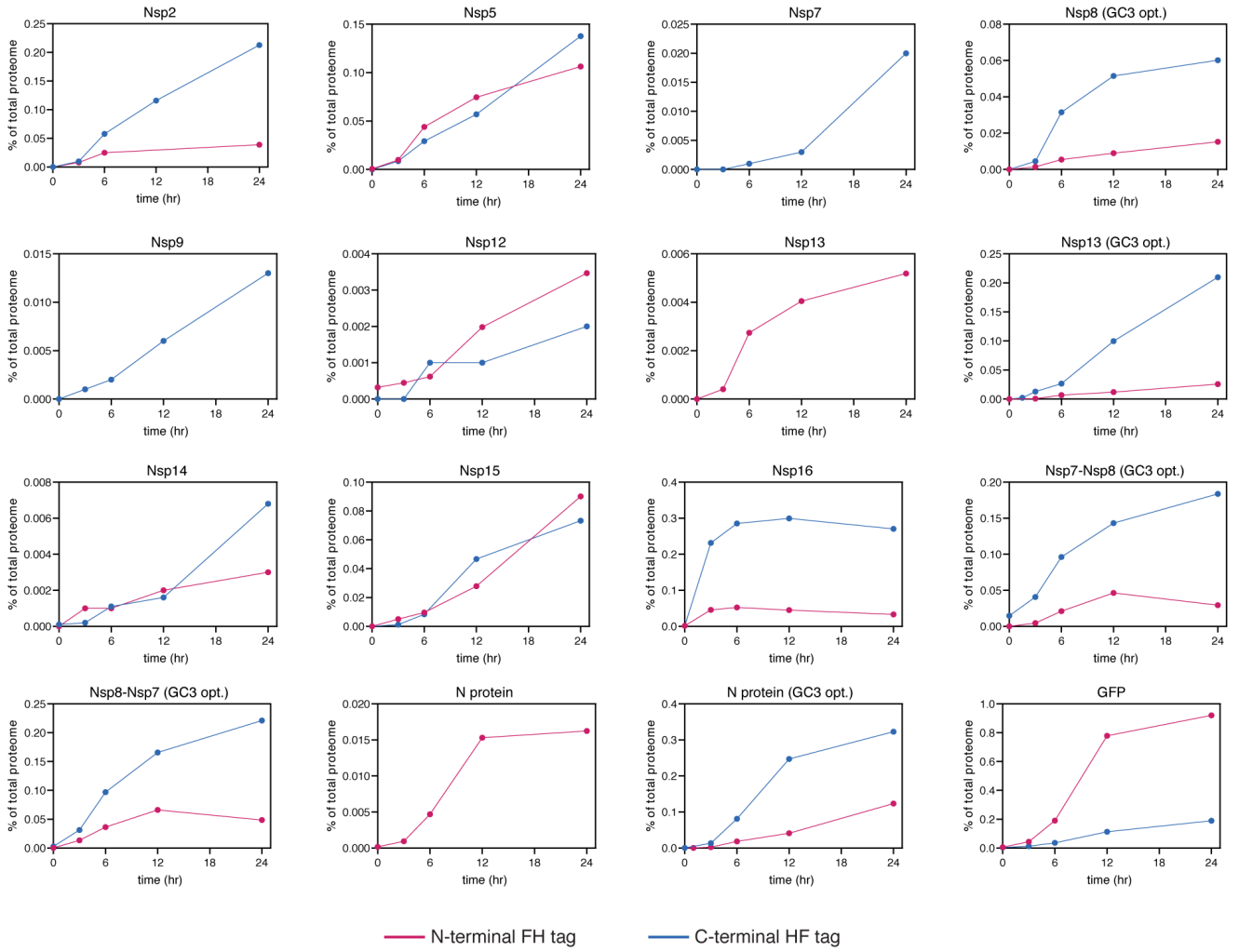


**Figure 5. Viral protein induction western blot quantitation.** As Figure 3C, but showing the proteins indicated.

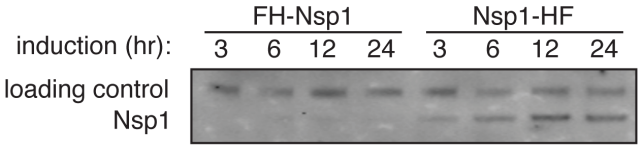


**Figure 6. Transient transfection of FH-Nsp1 (derived from plasmid 12) and Nsp1-HF (derived from plasmid 39).** Protein expression was assessed following induction from 3 to 24 h.

Because all of the tagged proteins possessed an identical FLAG epitope, aliquots of this 6h N protein sample could then be used as a reference standard on all subsequent western blots (e.g. lane 1 in Figure 3B), to allow similar abundance estimates for the other viral proteins (Figure 4; see *Underlying data*). This approach allows viral protein levels to be carefully titrated, so that protein induction approximately matches physiological conditions. This is important because different viral proteins show extreme differences in expression during infection. The nucleocapsid (N) protein can represent 2% of total protein, whereas the non-structural proteins may be 2 to 3 orders of magnitude less abundant (Finkel *et al.*, 2020; Grenga *et al.*, 2020). Moreover, all proteins will vary in their abundance throughout the course of infection.

## Conclusions

There has been a large amount of research activity surrounding COVID-19 but, understandably, this has predominately focused on epidemiology and the development of anti-viral strategies. The fundamental biology of SARS-CoV-2 infection has been relatively less addressed, and we aim to help fill this gap. To facilitate this process, we report the construction of 57 expression vectors and cell lines. The plasmids can be obtained from Addgene and cell lines are available. We expect that these collections will be a valuable resource for future research into the mechanisms by which coronavirus exploits the genetic machinery of its host to facilitate its own replication.

## Methods

### Construction of expression vectors

All viral genes were cloned into each of three parental vectors: pcDNA5-FRT-TO (to generate an untagged version of the protein), pcDNA5-FRT-TO-N-Flag-His8 (N-terminal tag), and pcDNA5-FRT-TO-C-His8-Flag (C-terminal tag). The N-terminal tag consisted of a single Flag motif, a four-alanine spacer, eight consecutive histidine residues, and a short unstructured linker (DYKDDDDKAAAAHHHHHHHHGSG). The C-terminal tag was essentially the same but in reverse (SGGHHHHHHHHAAA ADYKDDDDK). Oligos used and sequences of fusion constructs are provided as *Underlying data*.

***Construction of pcDNA5-FRT-TO-N-Flag-His8.*** To generate the Flag-His8 tag, we first designed partially complementary DNA oligonucleotides (oSB707 and oSB708) containing the Flag-His sequence. These oligos contained 20 nucleotides of complementarity at their 3′ ends, and an additional 35–36 nucleotides at their 5′ ends. The oligos were annealed in a reaction consisting of 12.5 µM forward oligo and 12.5 µM reverse oligo in a 40 µL reaction volume. The hybridization reaction was initially incubated at 95°C for 6 min, and gradually decreased to 25°C at the rate of 1.33°C/min. Subsequently, the annealed oligos were incubated in the same buffer supplemented with 250 µM dNTPs and 5U Klenow exo- (NEB) in a 50µL reaction at 37°C to fill in the single-stranded regions. After 1 hour, the insert sequence was purified on a silica column (Oligo Clean & Concentrator, Zymo Research, Cat No. D4060).

The insert fragment was then digested with restriction enzymes to facilitate cloning into the acceptor plasmid pcDNA5-FRT-TO. The insert was digested at 37°C for 2 hours in a reaction consisting of 800 ng of DNA, 1X Cutsmart buffer (NEB), 40U HindIII-HF (NEB, Cat No. R3104), and 40U BamHI-HF (NEB, Cat No. R3136). Subsequently, the DNA was purified, as above. In parallel, pcDNA5-FRT-TO was digested under identical conditions but with 2 µg of DNA. The digested DNA was subsequently purified on a silica column (QIAquick PCR Purification Kit, Qiagen, Cat No. 28104). To prevent self-ligation, the vector DNA was phosphatase treated in a reaction consisting of ~2 µg of DNA, 50mM Bis-Tris-Propane-HCl pH 6, 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, and 5U of Antarctic phosphatase (NEB, Cat No. M0289) at 37°C for 40 minutes. The DNA was again purified.

Finally, the digested Flag-His insert sequence was ligated into the digested and phosphatase-treated pcDNA5-FRT-TO acceptor plasmid. The ligation reaction consisted of 0.013 pmol vector, 0.042 pmol insert, 50 mM Tris-HCl 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, and 400U of T4 DNA ligase (NEB, Cat No. M0202) in a 20 µL reaction volume. The ligation mix was transformed into homemade DH5α *E. coli*, and plated overnight on LB-Amp. DNA was isolated from several colonies and sequenced to ensure correct insertion of the Flag-His sequence.

***Construction of pcDNA5-FRT-TO-C-His8-Flag.*** The pcDNA5-FRT-TO-C-His8-Flag sequence was generated as described above for the N-terminal tagging vector, with the following changes: 1) the oligos used for hybridization were oSB709 and oSB710, and 2) the insert and pcDNA5-FRT-TO were digested with 40U of EcoRV-HF (NEB, Cat No. R3195) and 40U of XhoI (NEB, Cat No. R0146).

***Cloning of untagged and N-terminally tagged viral genes.*** pcDNA5-FRT-TO and pcDNA-FRT-TO-N-Flag-His8 were each digested in a 50 µL reaction consisting of 2 µg DNA, 1X Cutsmart buffer, 20U of BamHI-HF, and 20U of EcoRV-HF at 37°C for two hours. In parallel, 2 µg of plasmid ($Kan^R$) containing the desired viral gene was digested under identical conditions. All three reactions were purified using a PCR cleanup kit (QIAquick PCR Purification Kit, Qiagen, Cat No. 28106) Subsequently, the acceptor plasmids were phosphatase treated as described above and again purified using a PCR cleanup kit.

Digested vector and insert were ligated together in a reaction consisting of 40 ng vector, 120 ng insert, 50 mM Tris-HCl 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, and 400 U of T4 DNA ligase (NEB) in a 10 reaction volume. The ligation mix was transformed into homemade DH5α *E. coli*, and plated overnight on LB-Amp. Colony PCR was performed using oAH195-196 to verify the presence of the insert: colonies were picked into 100 µL of Lysogeny broth (LB), and 1 µL of this mix used as input in a 12.5 µL reaction containing 0.5 µM of each primer and a final 1x dilution of Q5 High-Fidelity 2X Master Mix (NEB, Cat No. M0492). PCR cycling was done in a Kyratec SC300 SuperCycler with the following programme: 98°C for 30 s, then 30 cycles of 98°C for 5 s, 65°C for 20 s and 72°C for 30 s. Final extension was done for 2 minutes at 72°C. Products were run on a 2% agarose gel and visualised with SYBR Safe DNA Gel Stain (ThermoFisher, Cat No. S33102).

A positive colony for each construct was grown overnight in LB-Amp, and plasmids purified with QIAprep Spin Miniprep Kit (Qiagen, Cat No. 27106). 30 ng of plasmid was submitted for Sanger sequencing at the Medical Research Council Protein Phosphorylation and Ubiquitylation Unit (MRC PPU, Dundee, UK).

***Cloning of C-terminally tagged viral genes.*** To prepare the backbone, 2 µg of pcDNA5-FRT-TO-C-His8-Flag was digested with BamHI-HF and EcoRV-HF, as above, followed by DNA purification. To prepare viral gene inserts, 1 µg of pUC containing the relevant gene was initially digested with AvrII (NEB, Cat No. R0174S) in the same reaction conditions followed by purification. The 5' overhang, encoding the stop codon, was then removed by digestion with Mung Bean Nuclease (NEB, Cat No. M0250): 1 U of Mung Bean Nuclease was added to 1 µg of digested vector in 30 µl of 1x Cutsmart buffer, and incubated for 30 minutes at 30 °C. After purification, the final insert was produced by digestion with BamHI-HF. The insert was then ligated into the backbone after purification, following procedure described above with a 4:1 molar ratio of insert to backbone.

***Cloning of eGFP controls.*** The eGFP insert was amplified using pEGFP-N2 (Clontech) as a template and the DNA oligonucleotides oAH211 and oAH212 (untagged and N-terminal cloning) or oAH211 and oAH213 (C-terminal cloning). Subsequently,

the PCR-generated inserts were cloned into pcDNA5-FRT-TO, pcDNA-FRT-TO-N-Flag-His8, and pcDNA5-FRT-TO-C-His8-Flag as described above for the viral constructs.

### Cell culture and transfection

***Generation of stable cell lines.*** HEK293 Flp-In T-REx cells (Thermo Fisher, Cat No. R78007) were cultured at 37°C with 5% $CO_2$ in DMEM (Thermo-Fisher, Cat No.10566016) supplemented with 10% tetracycline-tested FBS (Sigma, Cat No. F2442), 100 µg/mL Zeocin (Alfa Aesar, Cat No. J67140.8X), and 15 µg/mL Blasticidin S (Sigma, Cat No.15205). Approximately $1\times10^6$ cells were seeded without antibiotics on six-well plates 24 hours prior to transfection. The following day, the viral expression constructs and pOG44 (the FRT recombinase) were co-transfected in a 1:9 ratio (1 µg total DNA/well) using Lipofectamine 2000 (3 µl/well) (Thermo-Fisher, Cat No. 11668030) according to the manufacturer's protocol.

The medium was replaced approximately five hours later to remove the transfection reagents. The next day, the cells were split to a 10 cm plate, and after an additional 24 hours, hygromycin B (150 µg/mL) and blasticidin S (15 µg/mL) were added to the medium. Stable integrants were selected over the course of 10–16 days, with medium replacement at regular intervals. Thereafter, stable cell lines were maintained in hygromycin B and blasticidin S.

For the FH-N constructs that initially yielded no colonies, transfection was repeated with the addition of pcDNA6/TR (Invitrogen, Cat No. V102520), encoding the tetracycline repressor protein: total transfected DNA was kept at 1 µg, with pcDNA5/FRT-TO, pcDNA6/TR and pOG44 used at a ratio of 1:4.5:4.5.

***Transient transfection of Nsp1.*** Approximately $2\times10^5$ cells were seeded without antibiotics on 24-well plates. The following day, 0.2 µg of FH-Nsp1 or Nsp1-HF was transfected into cells using Lipofectamine 2000 (Thermo-Fisher, Cat No. 11668030) according to the manufacturer's protocol.

### Induction tests

For induction tests, $1–2\times10^5$ cells were plated into six wells each of a 24-well plate. The following day, the cells were induced with 1 µg/mL of doxycycline, and harvested at varying timepoints by resuspending the cells in 50–100 µL of 1X Passive Lysis Buffer (Promega, Cat No. E1941). Inductions were staggered so that all timepoints could be harvested simultaneously. For the transient transfection experiments, cells were induced two hours post-transfection with 1 µg/mL of doxycycline. As with the stable cell lines, inductions were staggered and all timepoints were harvested at once.

For proteins larger than 20 kDa, cell lysates were resolved on 4–12% Bis-Tris gels, run at 180V for 1h in 1X MOPS (ThermoFisher, Cat No. NP0001). For smaller proteins, lysates were run for 25 min in MES buffer (Thermo Fisher, Cat No. NP0002). Proteins were wet transferred onto PVDF membrane (Millipore, Cat No. IPFL00010), blocked with 5% skimmed milk and probed successively with anti-Flag (Agilent, 200474-21; rat monoclonal antibody) and anti-Rat (Licor, 926-68076; goat polyclonal antibody). This anti-Flag antibody generated a prominent cross-reacting band at around 50 kDa which was used for loading normalization. For proteins running at the same size of the cross-reacting band (Nsp15 and Nsp12), an anti-M2-FLAG antibody (Sigma-Aldrich, F1804) was used in combination with anti-GAPDH (BioRad, VPA00187) as the loading control. For these, anti-Mouse (Licor, 926-68070) and anti-Rabbit (Licor, 926-32211) were then used. Protein bands were visualized and quantified by scanning in a Licor Odyssey CLX. Normalization was performed using the appropriate loading control and the N protein expression standard which was included on each gel.

### Mass spectrometry

HEK293 cells with stably integrated N-terminal tagged N protein (from plasmid 22) were induced with 1 µg/ml doxycyline for 0, 6, and 18 hours. Cells were harvested in lysis buffer containing 0.1% Rapigest (Waters, Cat No. 186001861) and sonicated (10 cycles of 30 seconds on, 30 seconds off in Bioruptor Pico, Diagenode). Approximately 30 µg of protein was resolved on a 4–20% Miniprotean TGX gel (Bio-Red, Cat No. 4561093), run in Tris-Glycine running buffer at 100 V. Subsequently, the gel was rinsed with water, stained for 1 hour with Imperial Protein Stain (Thermo Scientific, Cat No. 10006123), rinsed several times, and destained in water for three hours. Each lane was cut into four fractions and processed using in-gel digestion and the STAGE tip method, as previously described (Bresson *et al.*, 2020b; Rappsilber *et al.*, 2007).

### Availability of materials

The vectors are available from Addgene: Deposit number 78322

(https://www.addgene.org/search/catalog/plasmids/?q=tollervey).

Cell lines are available upon request by emailing the corresponding authors.

The datasets generated and analysed during the current study are available in the Figshare repository https://doi.org/10.6084/m9.figshare.13013492 (Bresson *et al.*, 2020a).

An earlier version of this article can be found on bioRxiv (doi: https://doi.org/10.1101/2020.07.20.211623).

## Data availability

### Underlying data

Mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (Perez-Riverol *et al.*, 2019), Accession number PXD020339: https://identifiers.org/pride.project:PXD020339.

Figshare: Data on integrative vectors for the regulated expression of SARS-CoV-2 proteins implicated in RNA metabolism. https://doi.org/10.6084/m9.figshare.13013492 (Bresson *et al.*, 2020a).

This project contains the following underlying data:

- Table 3.xlsx (Raw data for Figures 3C and 5)
- Table 4.xlsx (Mass-spectrometry data for N expression)

- Table 5.xlsx (Oligonucleotides used in this work)
- Table 6.xlsx (Sequences of fusion protein ORFs)
- Parental vectors Sanger sequencing (folder): original Sanger sequencing files for confirming the parental tagging vectors. Files in AB1 and SEQ formats.
- C terminal constructs Sanger sequencing (folder): original Sanger sequencing files for confirming C terminally tagged constructs, provided embedded in SnapGene files as well as original files. Files in DNA, AB1 and SEQ formats.
- Untagged and N terminal constructs Sanger sequencing (folder): original Sanger sequencing files for confirming untagged and N-terminally tagged constructs. Files in DNA, AB1 and SEQ formats.
- Licor imaging (folder): original scan of the western blot performed. Files in TIF, JPG and TXT formats.

## Extended data

Figshare: Data on integrative vectors for the regulated expression of SARS-CoV-2 proteins implicated in RNA metabolism. https://doi.org/10.6084/m9.figshare.13013492 (Bresson *et al.*, 2020a).

This project contains the following extended data:

- Cloning FH and HF containing-vectors.docx (protocol and lab notes for cloning and screening colonies for the parental tagging vectors)
- C terminal tagged and eGFP construct cloning.docx (protocol and original gel images for cloning and screening colonies for C terminally tagged constructs. Untagged, FH and HF eGFP constructs gel images are also included in this file)
- Untagged and N-terminal tagged construct cloning.docx (protocol and original gel images for cloning and screening colonies for constructs)



## Acknowledgements

We thank members of the Tollervey lab and Grzegorz Kudla for helpful discussions.

## References

Bhardwaj K, Guarino L, Kao CC: **The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as a cofactor.** *J Virol.* 2004; **78**(22): 12218–12224.
PubMed Abstract | Publisher Full Text | Free Full Text

Bresson S, Robertson N, Sani E, *et al.*: **Integrative Vectors for Regulated Expression of SARS-CoV-2 Proteins Implicated in RNA Metabolism.** *figshare.* 2020a.
http://www.doi.org/10.6084/m9.figshare.13013492

Bresson S, Shchepachev V, Spanos C, *et al.*: **Stress-induced translation inhibition through rapid displacement of scanning initiation factors.** *Mol Cell.* 2020b. S1097-2765(20)30651-1.
PubMed Abstract | Publisher Full Text

Cascarina SM, Ross ED: **A proposed role for the SARS-CoV-2 nucleocapsid protein in the formation and regulation of biomolecular condensates.** *FASEB J.* 2020.
PubMed Abstract | Publisher Full Text | Free Full Text

Chang CK, Hou MH, Chang CF, *et al.*: **The SARS coronavirus nucleocapsid protein--forms and functions.** *Antiviral Res.* 2014; **103**: 39–50.
PubMed Abstract | Publisher Full Text | Free Full Text

Chen Y, Su C, Ke M, *et al.*: **Biochemical and structural insights into the mechanisms of SARS coronavirus RNA ribose 2'-O-methylation by nsp16/nsp10 protein complex.** *PLoS Pathog.* 2011; **7**(10): e1002294.
PubMed Abstract | Publisher Full Text | Free Full Text

Cubuk J, Alston JJ, Incicco JJ, *et al.*: **The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA.** *bioRxiv.* 2020. 2020.06.17.158121.
PubMed Abstract | Publisher Full Text | Free Full Text

Eckerle LD, Becker MM, Halpin RA, *et al.*: **Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing.** *PLoS Pathog.* 2010; **6**(5): e1000896.
PubMed Abstract | Publisher Full Text | Free Full Text

Egloff MP, Ferron F, Campanacci V, *et al.*: **The severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world.** *Proc Natl Acad Sci U S A.* 2004; **101**(11): 3792–3796.
PubMed Abstract | Publisher Full Text | Free Full Text

Finkel Y, Mizrahi O, Nachshon A, *et al.*: **The coding capacity of SARS-CoV-2.** *bioRxiv.* 2020; 2020.2005.2007.082909.
Publisher Full Text

Frieman M, Yount B, Heise M, *et al.*: **Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane.** *J Virol.* 2007; **81**(18): 9812–9824.
PubMed Abstract | Publisher Full Text | Free Full Text

Gao Y, Yan L, Huang Y, *et al.*: **Structure of the RNA-dependent RNA polymerase from COVID-19 virus.** *Science.* 2020; **368**(6492): 779–782.
PubMed Abstract | Publisher Full Text | Free Full Text

Gordon DE, Jang GM, Bouhaddou M, *et al.*: **A SARS-CoV-2 protein interaction map reveals targets for drug repurposing.** *Nature.* 2020; **583**(7816): 459–468.
PubMed Abstract | Publisher Full Text | Free Full Text

Granneman S, Kudla G, Petfalski E, *et al.*: **Identification of protein binding sites on U3 snoRNA and pre-rRNA by UV cross-linking and high throughput analysis of cDNAs.** *Proc Natl Acad Sci USA.* 2009; **106**(24): 9613–9818.
PubMed Abstract | Publisher Full Text | Free Full Text

Grenga L, Gallais F, Pible O, *et al.*: **Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines.** *Emerg Microbes Infect.* 2020; **9**(1): 1712–1721.
PubMed Abstract | Publisher Full Text | Free Full Text

Gussow AB, Auslander N, Faure G, *et al.*: **Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses.** *Proc Natl Acad Sci U S A.* 2020; **117**(26): 15193–15199.
PubMed Abstract | Publisher Full Text | Free Full Text

Hillen HS, Kokic G, Farnung L, *et al.*: **Structure of replicating SARS-CoV-2 polymerase.** *Nature.* 2020; **584**(7819): 154–156.
PubMed Abstract | Publisher Full Text

Huang C, Lokugamage KG, Rozovics JM, *et al.*: **SARS coronavirus nsp1 protein induces template-dependent endonucleolytic cleavage of mRNAs: viral mRNAs are resistant to nsp1-induced RNA cleavage.** *PLoS Pathog.* 2011; **7**(12): e1002433.
PubMed Abstract | Publisher Full Text | Free Full Text

Jang KJ, Jeong S, Kang DY, *et al.*: **A high ATP concentration enhances the cooperative translocation of the SARS coronavirus helicase nsP13 in the unwinding of duplex RNA.** *Sci Rep.* 2020; **10**(1): 4481.
PubMed Abstract | Publisher Full Text | Free Full Text

Kneller DW, Phillips G, O'Neill HM, *et al.*: **Structural plasticity of SARS-CoV-2 3CL $M^{pro}$ active site cavity revealed by room temperature X-ray crystallography.** *Nat Commun.* 2020; **11**(1): 3202.
PubMed Abstract | Publisher Full Text | Free Full Text

Krafcikova P, Silhan J, Nencka R, *et al.*: **Structural analysis of the SARS-CoV-2 methyltransferase complex involved in coronaviral RNA cap creation.**

*bioRxiv.* 2020; 2020.2005.2015.097980.
**Publisher Full Text**

Kudla G, Lipinski L, Caffin F, *et al.*: **High guanine and cytosine content increases mRNA levels in mammalian cells.** *PLoS Biol.* 2006; **4**(6): e180.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Li JY, Liao CH, Wang Q, *et al.*: **The ORF6, ORF8 and nucleocapsid proteins of SARS-CoV-2 inhibit type I interferon signaling pathway.** *Virus Res.* 2020; **286**: 198074.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Littler DR, Gully BS, Colson RN, *et al.*: **Crystal Structure of the SARS-CoV-2 Non-structural Protein 9, Nsp9.** *iScience.* 2020; **23**(7): 101258.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Lokugamage KG, Narayanan K, Huang C, *et al.*: **Severe acute respiratory syndrome coronavirus protein nsp1 is a novel eukaryotic translation inhibitor that represses multiple steps of translation initiation.** *J Virol.* 2012; **86**(24): 13598–13608.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Ma Y, Wu L, Shaw N, *et al.*: **Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex.** *Proc Natl Acad Sci U S A.* 2015; **112**(30): 9436–9441.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Mordstein C, Savisaar R, Young RS, *et al.*: **Codon Usage and Splicing Jointly Influence mRNA Localization.** *Cell Syst.* 2020; **10**(4): 351–362.e8.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Muramatsu T, Takemoto C, Kim YT, *et al.*: **SARS-CoV 3CL protease cleaves its C-terminal autoprocessing site by novel subsite cooperativity.** *Proc Natl Acad Sci U S A.* 2016; **113**(46): 12997–13002.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Ogando NS, Zevenhoven-Dobbe JC, Posthuma CC, *et al.*: **The enzymatic activity of the nsp14 exoribonuclease is critical for replication of Middle East respiratory syndrome-coronavirus.** *bioRxiv.* 2020; 2020.2006.2019.162529.
**Publisher Full Text**

Perez-Riverol Y, Csordas A, Bai J, *et al.*: **The PRIDE database and related tools and resources in 2019: improving support for quantification data.** *Nucleic Acids Res.* 2019; **47**(D1): D442–D450.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Rappsilber J, Mann M, Ishihama Y: **Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips.** *Nat Protoc.* 2007; **2**(8): 1896–1906.
**PubMed Abstract** | **Publisher Full Text**

Shchepachev V, Bresson S, Spanos C, *et al.*: **Defining the RNA Interactome by Total RNA-Associated Protein Purification.** *Mol Sys Biol.* 2019; **15**(4): e8689.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Sims AC, Tilton SC, Menachery VD, *et al.*: **Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells.** *J Virol.* 2013; **87**(7): 3885–3902.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Sutton G, Fry E, Carter L, *et al.*: **The nsp9 Replicase Protein of SARS-Coronavirus, Structure and Functional Insights.** *Structure.* 2004; **12**(2): 341–353.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Tanaka T, Kamitani W, DeDiego ML, *et al.*: **Severe acute respiratory syndrome coronavirus nsp1 facilitates efficient propagation in cells through a specific translational shutoff of host mRNA.** *J Virol.* 2012; **86**(20): 11128–11137.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Tanner JA, Watt RM, Chai YB, *et al.*: **The severe acute respiratory syndrome (SARS) coronavirus NTPase/helicase belongs to a distinct class of 5′ to 3′ viral helicases.** *J Biol Chem.* 2003; **278**(41): 39578–39582.
**PubMed Abstract** | **Publisher Full Text**

te Velthuis AJ, van den Worm SH, Snijder EJ: **The SARS-coronavirus nsp7+nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension.** *Nucleic Acids Res.* 2012; **40**(4): 1737–1747.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Thoms M, Buschauer R, Ameismeier M, *et al.*: **Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2.** *Science.* 2020; **369**(6508): 1249–1255.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Tvarogova J, Madhugiri R, Bylapudi G, *et al.*: **Identification and Characterization of a Human Coronavirus 229E Nonstructural Protein 8-Associated RNA 3′-Terminal Adenylyltransferase Activity.** *J Virol.* 2019; **93**(12): e00291–19.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Viswanathan T, Arya S, Chan SH, *et al.*: **Structural Basis of RNA Cap Modification by SARS-CoV-2 Coronavirus.** *bioRxiv.* 2020; 2020.2004.2026.061705.
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

Yuen CK, Lam JY, Wong WM, *et al.*: **SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists.** *Emerg Microbes Infect.* 2020; **9**(1): 1418–1428
**PubMed Abstract** | **Publisher Full Text** | **Free Full Text**

# Open Peer Review

## Current Peer Review Status: ✔ ✔

**Version 1**

Reviewer Report 03 December 2020

https://doi.org/10.21956/wellcomeopenres.17943.r41430



✔ **Chad Petit**
Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Birmingham, AL, USA

The manuscript is well written with the development of the methodology clearly stated and presented to allow maximum impact for the scientific community. The etiological agent of the COVID-19 pandemic is the Severe Acute Respiratory Syndrome (SARS) coronavirus (CoV) 2. Unfortunately, the specific pathways and host-pathogen interactions that facilitate the viral lifecycle for this novel virus are largely not understood. The authors of this manuscript have selected 14 SARS-CoV-2 proteins that are expected to interact with RNA or RNA binding proteins for construction of reagents that can be used for analyzing their role in the coronavirus lifecycle. Briefly, numerous cell lines were established that express various forms (e.g. amino tagged, carboxyl tagged, etc) of each of the viral proteins. Furthermore, the expression of the viral proteins can be tuned to their relative expression levels during viral replication using different concentrations of doxycycline or incubation time.

The rationale for the development of these reagents is sound in that they could be used for future research to understand how these viral proteins affect the host cell. The description of the synthesis of the reagents is clear and provided at a level of details that would allow the replication of the method by other labs. The conclusions about the method and their relevance for future deployment in other research programs are more than adequately supported by the data presented in the manuscript.

**Is the rationale for developing the new method (or application) clearly explained?**
Yes

**Is the description of the method technically sound?**
Yes

**Are sufficient details provided to allow replication of the method development and its use by others?**

Yes

**If any results are presented, are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions about the method and its performance adequately supported by the findings presented in the article?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Structural virologist focusing on influenza and SARS-CoV-2.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

Reviewer Report 19 November 2020

https://doi.org/10.21956/wellcomeopenres.17943.r41258



**Sheila Graham**
MRC - University of Glasgow Centre for Virus Research, Institute of Infection, Immunity and Inflammation, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow, UK

This is a methods paper that creates, and experimentally assesses, 54 inducible expression constructs encoding SARS-CoV2 (SCV2) proteins which are predicted to bind RNA. Bresson *et al.* carefully describe generation of this important resource, which will facilitate studies into how the host RNA proteome is affected by SCV2 infection and how viral proteins might alter RNA-protein interactions. Such a resource is essential for any thorough analysis of the SCV2 life cycle.

In this method, an important innovation was a fix to get around use of RT-PCR, which uses an error-prone polymerase, to produce mutation-free constructs. Expression plasmids were designed to be multifunctional. They are suitable for transient expression, or hygromycin-selectable stable expression, following flippase integration into the genome of cells engineered to contain a flippase recombination target sequence. Finally, the constructs express proteins that are tandem affinity-tagged (untagged, or N- or C-terminally tagged) in order to facilitate purification. Importantly, the constructs were also codon optimised to ensure optimal expression levels.

The well-thought out cloning strategy is well illustrated in Figure 2 allowing the method to be easily replicated. Quantification of expression levels as a percentage of total cellular protein is shown using a mass spectroscopy approach while western blots show the quality of expression

and a time course of tetracycline induction. This is important to allow the reader to assess usefulness of the constructs.

The article is scientifically sound. However, it would be helpful if the authors could suggest the nature of the cross reacting protein that they have used as a loading control in many of the western blots. They do not state the antibody used for the western blots in the figure legend. I imagine it is an anti-FLAG antibody?

It would be good to be clear about the FH/HF nomenclature in the text and in Table 2 and the figure legends. Also correcting the misspelled “cloning” in the Figure would be good.

It is a pity that Nsp12, encoding the catalytic subunit of the viral RdRp, a key RNA binding viral protein, was relatively poorly expressed, especially the C-terminal-tagged version. Can the authors give any information on their attempts to improve expression of this or any other viral proteins?

**Is the rationale for developing the new method (or application) clearly explained?**
Yes

**Is the description of the method technically sound?**
Yes

**Are sufficient details provided to allow replication of the method development and its use by others?**
Yes

**If any results are presented, are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions about the method and its performance adequately supported by the findings presented in the article?**
Yes

***Competing Interests:*** No competing interests were disclosed.

***Reviewer Expertise:*** Virology, RNA biology.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**