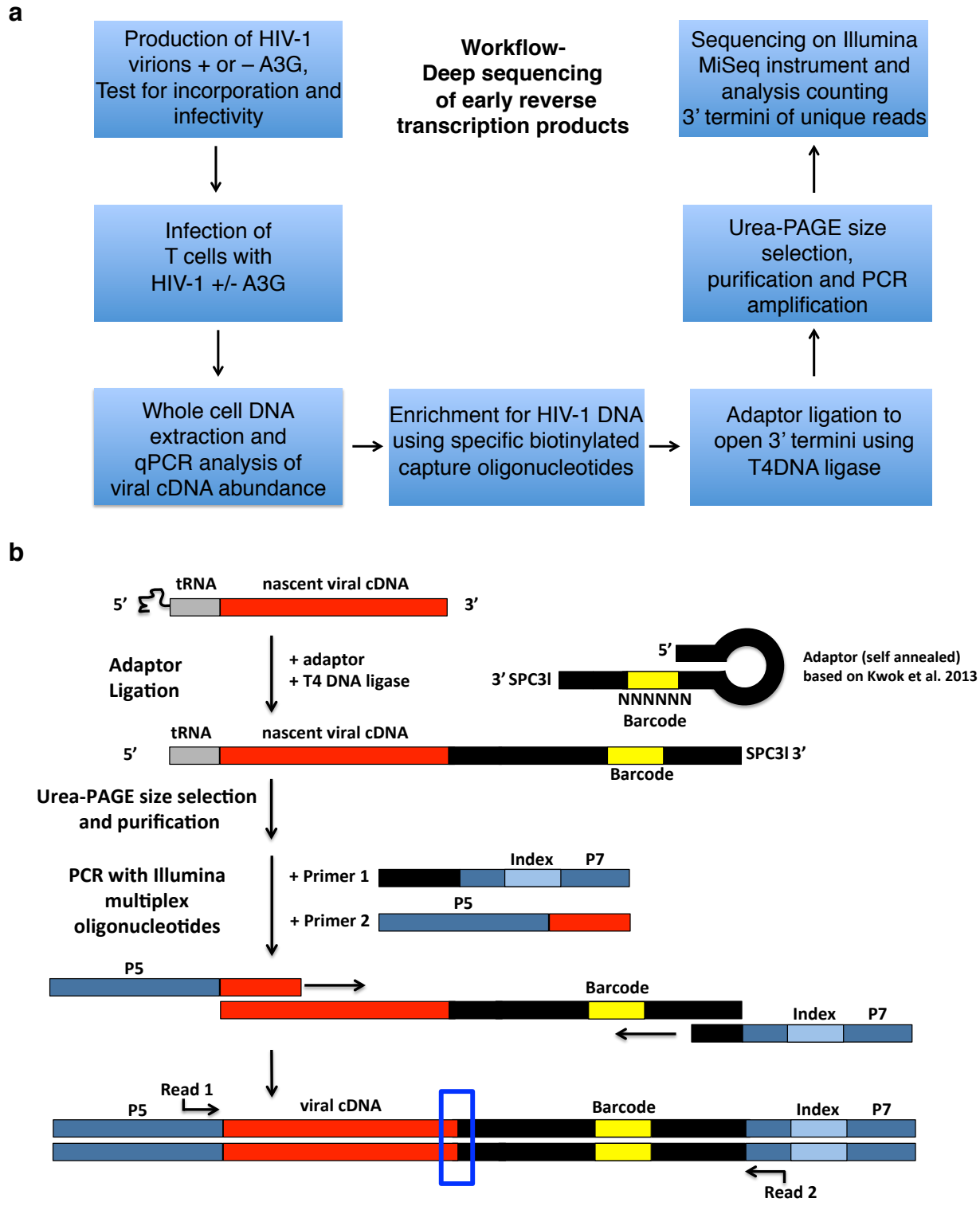
Supplementary Figures 1 through 11 for manuscript:

Deep sequencing of HIV-1 reverse transcripts reveals the multifaceted anti-viral functions of APOBEC3G

Darja Pollpeter, Maddy Parsons, Andrew E. Sobala, Sashika Coxhead, Rupert D. Lang, Annie M. Bruns, Stelios Papaioannou, James M. McDonnell, Luis Apolonia, Jamil A. Chowdhury, Curt M. Horvath, Michael H. Malim

Supplemental Figure 1



Supplemental Figure 1- continued

c

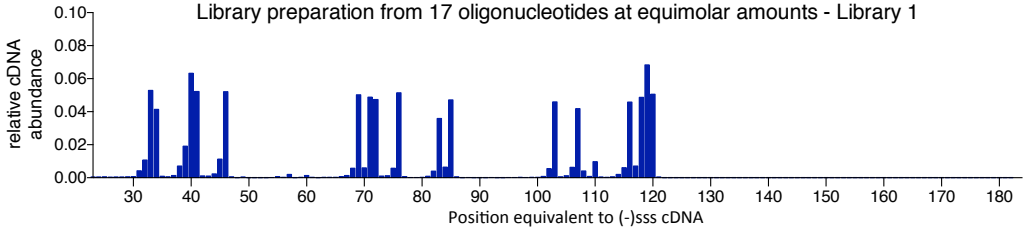
Oligo name	Length in nt	Sequence
full Kwok + MiSeq	61	5'-PHO-tgaagagcctagtcgctgttcannnnnnctgcccatagagagatcggaagagcacacgtct-SpC3-3'
2xBiotin SS bait	40	5'-biotin-cagtgtggaaaatctctagcagtggcgcccgaacagggac-biotin-3'
Biotin 1-16 ss	22	5'-cagtgtggaaaatctctagcag-BiTEG-3'
Biotin tRNA + CTG	16	5'-cagtggcgcccgaaca-BITEG-3'
MP1.0+22HIV	82	5'-aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatctcactgctagagattttccacactg-3'

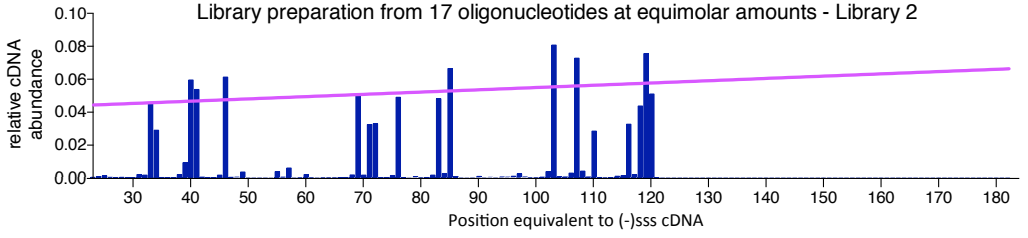
Supplementary Figure 1: Diagram of HIV-1 reverse transcripts deep sequencing strategy and adaptor sequences

a) Workflow of sequencing library protocol. **b)** Diagram of library creation. Following the extraction of whole cell DNA from infected cells, HIV-1 DNA is enriched over cellular DNA by hybridization of biotinylated oligonucleotides and strepdavidin magnetic beads separation; single stranded nascent HIV-1 cDNAs with open 3'-termini are captured by this method. cDNA is ligated to a single stranded DNA adaptor using T4 DNA ligase (NEB). The adaptor (named full Kwok+MiSeq) design was inspired by Kwok et al. (Anal Biochem. 2013 Apr 15;435(2):181-6. doi: 10.1016/j.ab.2013.01.008.), such that it is stabilized by folding back on itself. Of note, we explored the use of other ligases and ligation strategies, especially thermostable 5'App DNA/RNA ligase (NEB). However, we noticed strong biases in ligation efficiency of control oligonucleotides, even with single base differences, suggesting considerable structure based bias for this ligase. In contrast, biases were not identified for the ligation strategy used here. The adaptor carries a random barcode sequence (BC), which simultaneously allows for base-pairing to facilitate ligation. The 3'-end of the adaptor carries a spacer (SpC3) to prevent self-ligation. After enzyme inactivation, ligations are separated by denaturing PAGE. The gel is stained with SYBR Gold and cut into three separate, equal-size pieces according to position in the gel from above the adaptor to the well. This enables removal of unligated adaptor and addresses a potential size-based bias in subsequent PCR steps. After elution, precipitation and resuspension the products are PCR amplified with primers annealing to the known sequence of the adaptor (primer 1, NEBNext Multiplex Oligos for Illumina) and a primer carrying the first 22 nt of the HIV-1 5'-LTR sequence immediately following the tRNA (primer 2, MP1.0+22HIV). The 5' ends of the chosen primers carry adaptors for the MiSeq platform (P5 and P7) as well as an index sequence to distinguish individual samples run in the same library. **c)** Table of oligonucleotides used as capture probes for enrichment, adaptors and PCR primers.

Supplemental Figure 2

a





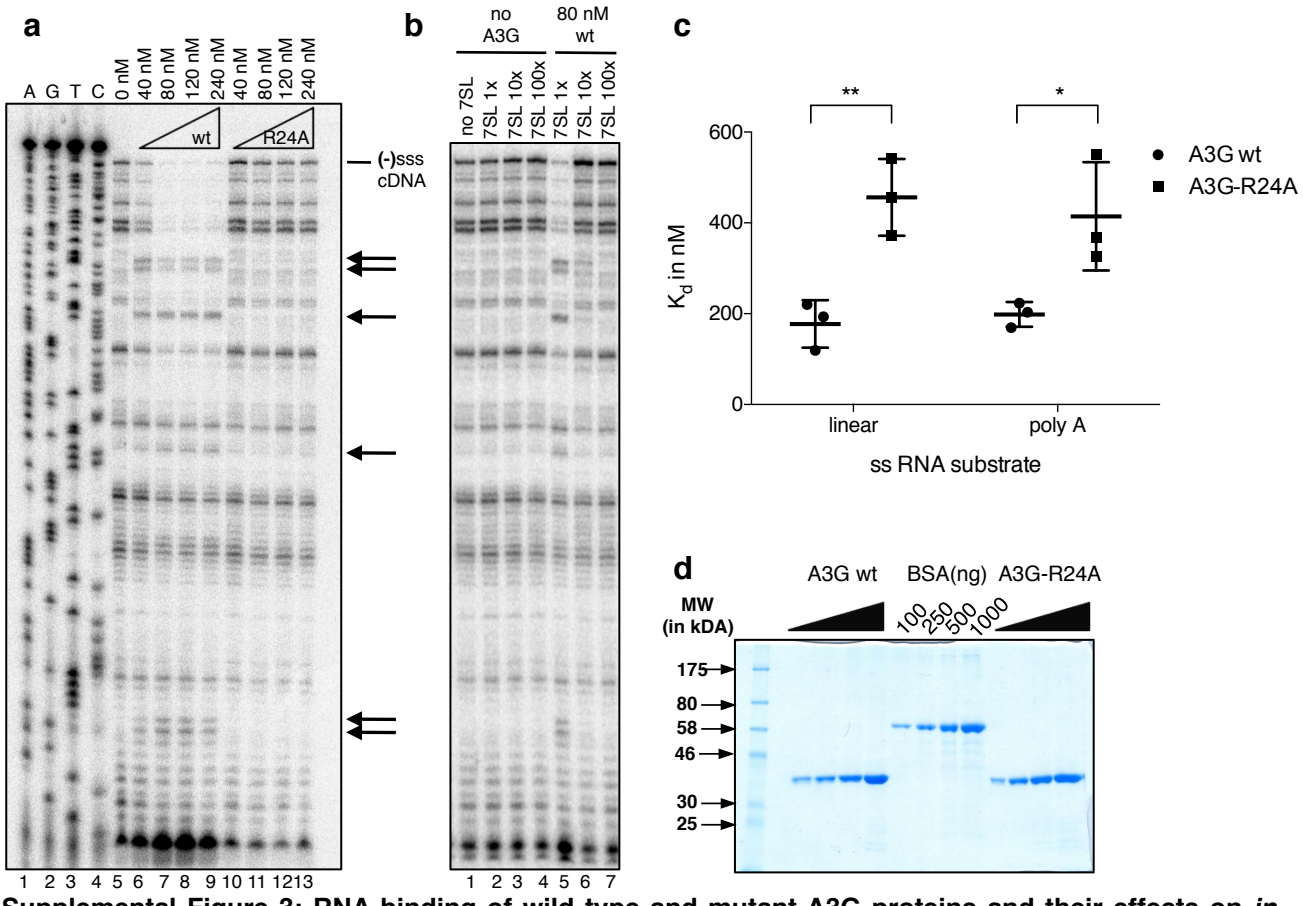
b

Oligo name	Length (in nt)	Sequence
HTP con long C	120	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcacacactactttgagcactcaaggcaagctttattgaggcttaagc-3'
HTP con long G	119	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcacacactactttgagcactcaaggcaagctttattgaggcttaag-3'
HTP con long T	116	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcacacactactttgagcactcaaggcaagctttattgaggctt-3'
HTP con long A	118	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcacacactactttgagcactcaaggcaagctttattgaggcttaa-3'
HTP con mid C	76	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcac-3'
HTP con mid G (a)	71	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacagacg-3'
HTP con mid G (b)	72	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacagacgg-3'
HTP con mid A	69	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacaga-3'
HTP con mid T	85	5'-ctgctagagattttccacactgactaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcacacactactt-3'
HTP con short A	40	5'-ctgctagagattttccacactgactaaaagggtctgaggga-3'
HTP con short T	33	5'-ctgctagagattttccacactgactaaaagggt-3'
HTP con short G	41	5'-ctgctagagattttccacactgactaaaagggtctgaggg-3'
HTP con short C	34	5'-ctgctagagattttccacactgactaaaagggtc-3'
HTP Con 46 (T)	46	5'-ctgctagagattttccacactg actaaaagggtctgagggatctct-3'
HTP Con 83 (C)	83	5'-ctgctagagattttccacactg actaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcacacactac-3'
HTP Con 103 (C)	103	5'-ctgctagagattttccacactg actaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcacacactactttgagcactcaaggcaagc-3'
HTP Con 107 (A)	107	5'-ctgctagagattttccacactg actaaaagggtctgagggatctctagttaccagagtcacacaacagacgggcacacactactttgagcactcaaggcaagcttta-3'

Supplemental Figure 2: Control libraries and sequences of control oligonucleotides

a) In order to account for potential bias in library preparation, particularly ligation steps and library sequencing, we included control library samples within all multiplexed libraries. Shown are the sequencing profiles for samples comprising pools containing equimolar amounts of 17 different length oligonucleotides. These are the natural HIV-$1_{NL4.3}$ sequence and were selected based on their relative length and 3'-nucleotide. We demonstrate that our optimized protocol has no significant bias towards the open 3'-end. In one library run, which included sample data presented in this article (library 2, Figures 3,8,S4,S6), we observed a minor length bias in sequencing. This may be do to clustering bias on the MiSeq™ Chip or PCR bias. The latter might occur in spite of being addressed by size selection and separate PCR reaction (see Materials and Methods and Supplementary Fig 1). In this case, we therefore applied a normalization factor reflecting the slope (shown in pink) that represents size bias. **b)** Table of oligonucleotides used for the control library as presented in a). Oligonucleotides were chosen based on size (short (33-41nt), mid (69-85nt), long (116-134nt)) and include 4 oligonucleotides corresponding to A3G-induced pause sites seen in reconstitution assays (refer to Supplementary Fig 3a).

Supplemental Figure 3

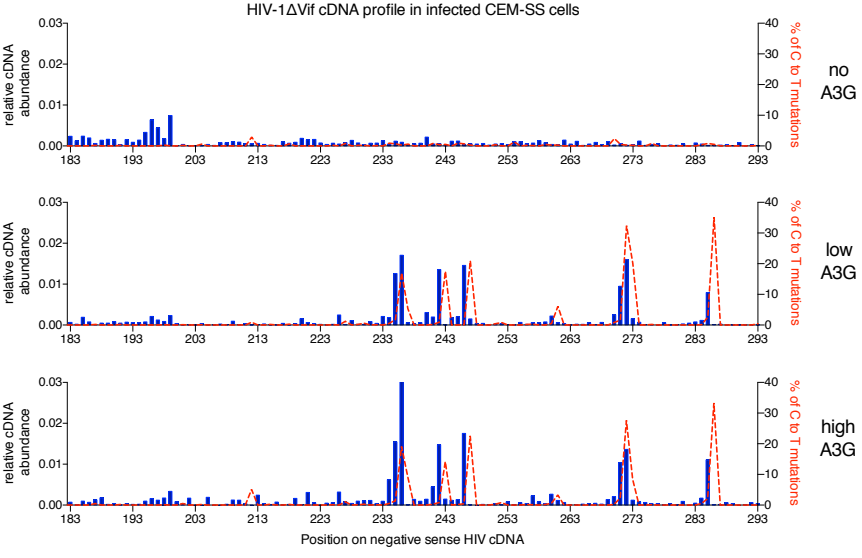


Supplemental Figure 3: RNA binding of wild type and mutant A3G proteins and their effects on *in vitro* primer extension by HIV-1 RT

a) Reconstituted *in vitro* assay of purified HIV-1 RT catalyzing (-)sss cDNA synthesis, visualized using sequencing urea-acrylamide gels and autoradiography. Extension of a radiolabelled, single stranded, 18 nt DNA oligonucleotide primer, which annealed to the single stranded HIV-1 RNA template, by HIV-1 RT (lane 5). Reactions in lanes 6-9 and 10-13 include increasing amounts of wild type A3G or the R24A mutant, respectively. Lane 1-4 show product ladders, which were generated by Sanger-sequencing of the HIV-1 template cDNA. Arrows indicate A3G induced specific RT pause sites, which are seen with the wild type protein but not with R24A. **b)** Reconstituted, *in vitro* primer extension assay in the presence of a competitor RNA, 7SL RNA. Single stranded RNA template with increasing amounts of 7SL RNA was pre-incubated with constant amounts of wild type A3G and the reactions initiated by addition of HIV-1 RT. The presence of excess amounts of 7SL RNA relieves the inhibitory effect of A3G. **c)** Dor plot with mean dissociation constants (K_D) for wild type and R24A using two different single stranded RNA substrates, as measured by observing single molecule RNA binding/unbinding events of unlabelled A3G in TIRF microscopy and exploiting the photophysical effect of protein induced fluorescent enhancement (PIFE). Substrates were designed to have low secondary structure (linear) or carry a known structure hairpin, the polyA loop of the HIV-1 5'-LTR. Error bars represent standard deviation of 3 independent experiments. An unpaired t-test was performed in GraphPad Prism® and **:p=0.0083; *:p=0.0377 **d)** Coomassie gel of purified proteins used in primer extension and RNA binding assays.

Supplemental Figure 4

a



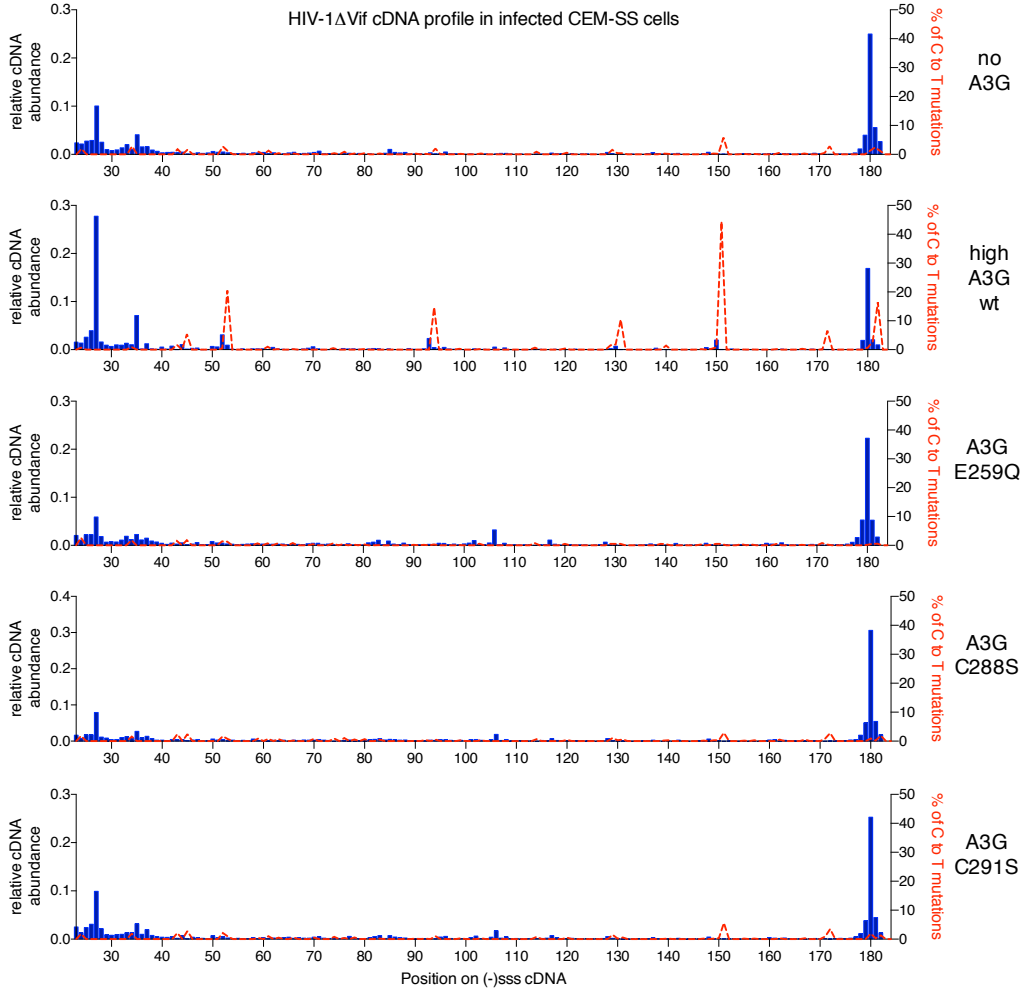
b

Library	Figure	Sample	Reads	ENA - sample accession number
Library 1	Figure 2/S4	no A3G	81272	ERS1871624
		low A3G	46953	ERS1871625
		high A3G	26592	ERS1871626
		uninfected control	1858	ERS1871628
		Oligo control library 1	124107	ERS1871627
Library 2	Figure 3	no A3G ctr(P)-ctr(T)	112647	ERS1903608
		high A3G ctr(P)-ctr(T)	15681	ERS1903594
		high A3G hUGI(P)-hUGI(T)	73385	ERS1903598
		high A3G hUGI(P)-ctr(T)	52025	ERS1903597
		high A3G ctr(P)-hUGI(T)	36499	ERS1903595
	Figure 8	no A3G	72006	ERS1903607
		low A3G wt	9440	ERS1903606
		A3G-L35A	9890	ERS1903604
		A3G-R24A	12740	ERS1903605
	Figure S7	no A3G hUGI(P)-hUGI(T)	58996	ERS1878244
		low A3G ctr(P)-ctr(T)	63021	ERS1903599
		low A3G hUGI(P)-hUGI(T)	94968	ERS1903603
		low A3G hUGI(P)-ctr(T)	95634	ERS1903602
		low A3G ctr(P)-hUGI(T)	73865	ERS1903600 ERS1903601
		Oligo control library 2	84823	ERS1878245
	Figure S5	no A3G	16366	ERS1878243
		high A3G	4784	ERS1903596
		A3G-E259Q	48223	ERS1903593
		A3G-C288S	24246	ERS1903591
		A3G-C291S	21316	ERS1903592

Supplemental Figure 4: cDNA profiles after first strand transfer and total read counts

a) Shown in blue histograms are cDNA profiles of negative sense HIV-1 cDNA continuing beyond the (-)sss cDNA sequence profiles that are presented in main Fig 1. 'Low' A3G refers to a producer cell transfection ratio of 1:10 (A3G expression plasmid to NL4.3/ΔVif), whereas high 'A3G' packaging levels derive from a 1:4 ratio. The initial 110 nt that are reverse transcribed following 1st strand transfer (nt 183-293 of HIV-$1_{NL4.3}$) are plotted on the x-axis. The left y-axis shows the relative abundance of cDNA molecules calculated from the entire dataset, including reads terminating within the first 182 nt of HIV-1 sequence. Shown in dashed red lines is the percentage of reads which carried C to T/U mutations at the indicated position (right y-axis). The same correspondence between the A3G induced mutational profile (C to U mutations) and accumulation of truncated molecules, as described in Fig 1, is observed. **b)** List of all samples presented in main or supplemental figures, including total read numbers of each sample and individual accession codes at the European Nucleotide Archive (ENA, Study accession code PRJEB22170). P: in HEK293T producer cells, T: In CEM-SS target cells.

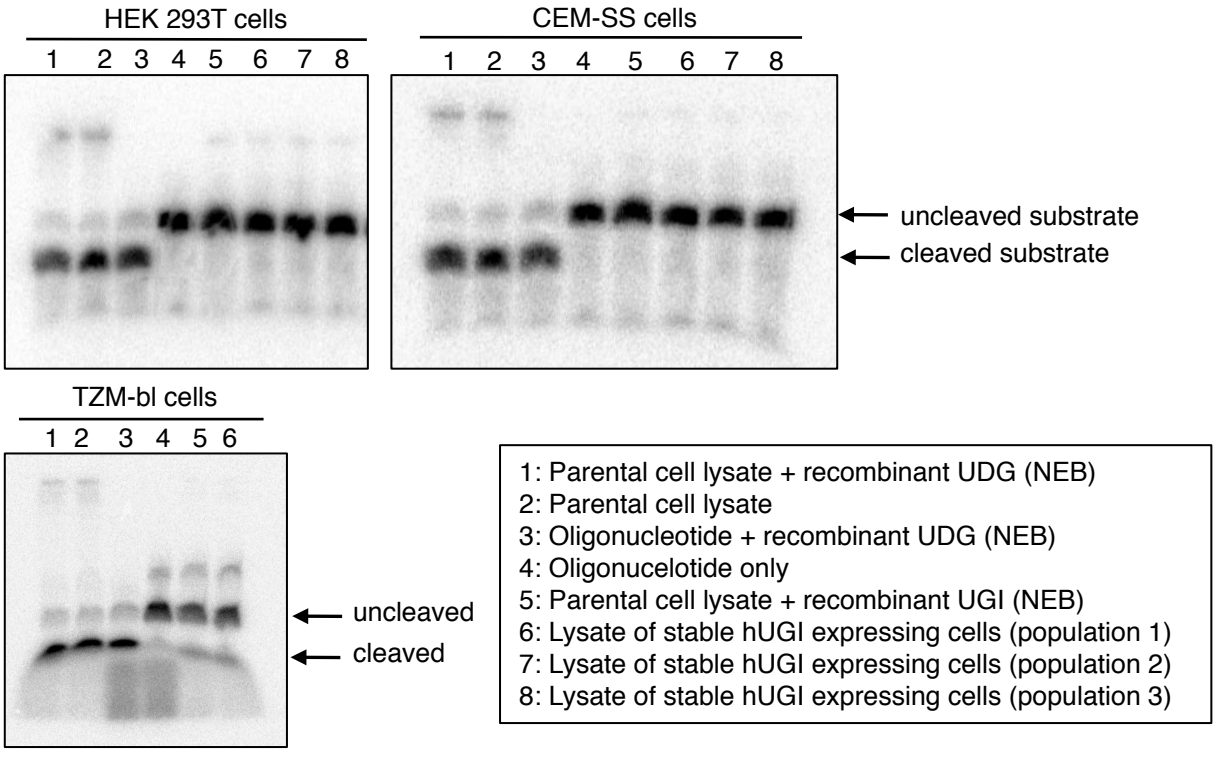
Supplemental Figure 5



Supplemental Figure 5: cDNA profiles in cells infected with virions containing wild type or catalytically inactive A3G proteins

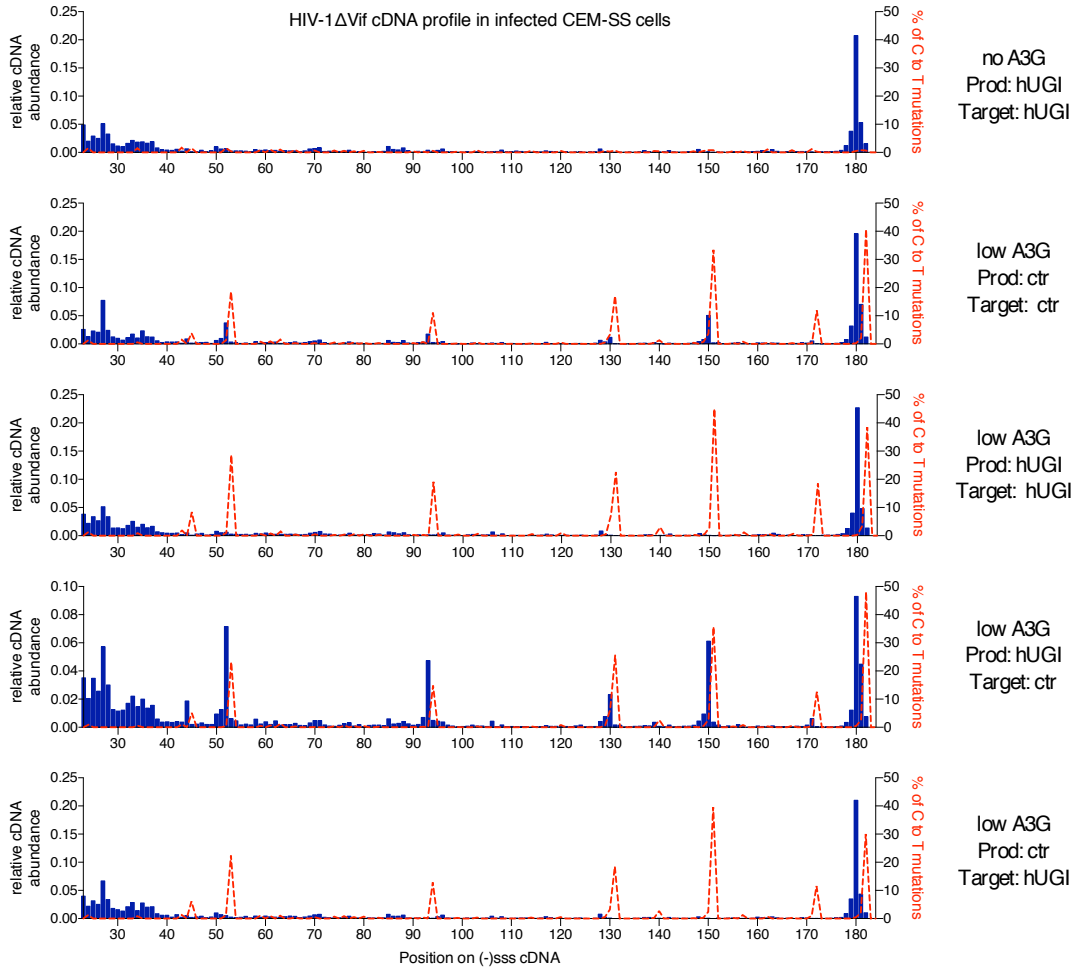
The graphs show the relative abundance of HIV-1 cDNA molecules for each length between nucleotides 23 and 182 of the HIV-$1_{NL4.3}$ (-)sss product. Shown in dashed red lines is the percentage of reads which carried C to T/U mutations at the indicated positions. The labeling to the right of the graphs refers to the form of A3G that was present. The absence of significant levels of C to U mutations in the presence of catalytically inactive A3G proteins leads to the absence of cDNA species foreshortened by 1 nt relative to A3G editing sites. 'High' A3G refers to a producer cell transfection ratio of 1:4 (A3G expression plasmid to NL4.3/Δvif).

Supplemental Figure 6



Supplemental Figure 6: Stable expression of hUGI eliminates UDG activity in cell lysates
HEK293T, CEM-SS and TZM-bl cell lines were transduced with a retroviral vector encoding hUGI. A radiolabelled, single stranded 34 nt oligonucleotide containing a central uracil was incubated with cell lysates from the parental or hUGI expressing lines (lane 2 and 6, respectively;. oligonucleotide cleavage indicates UDG activity. Controls were the parental cell lysates or the oligonucleotide by itself with exogenous recombinant UDG (NEB) (lane 1,3) and parental cell lysate with exogenous recombinant UGI (NEB) (lane 5). Samples were run on an 8% denaturing urea-polyacrylamide without molecular weight markers. For HEK293T and CEM-SS we created three cell populations and one cell population for TZM-bl cells. All hUGI expressing cell lines show a complete inhibition of UDG activity.

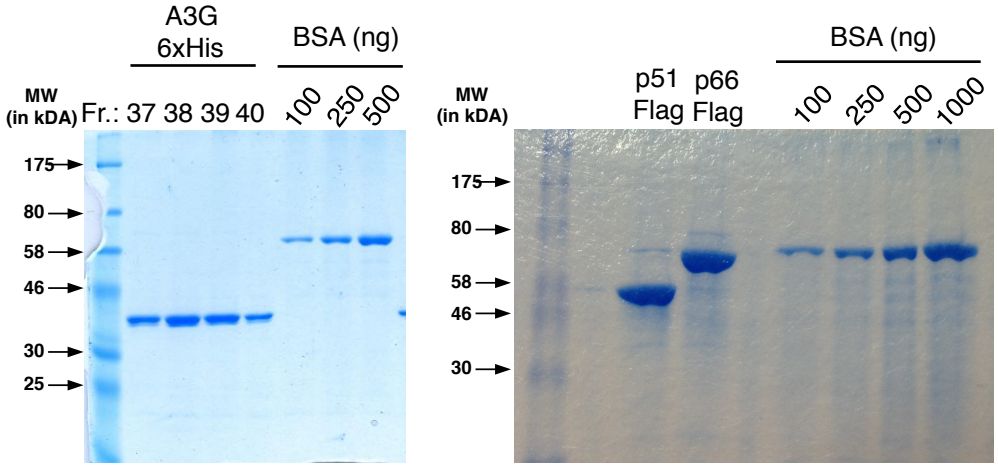
Supplemental Figure 7



Supplemental Figure 7: cDNA profiles in cells infected with virions containing A3G when hUGI is present in producer cells, target cells or both

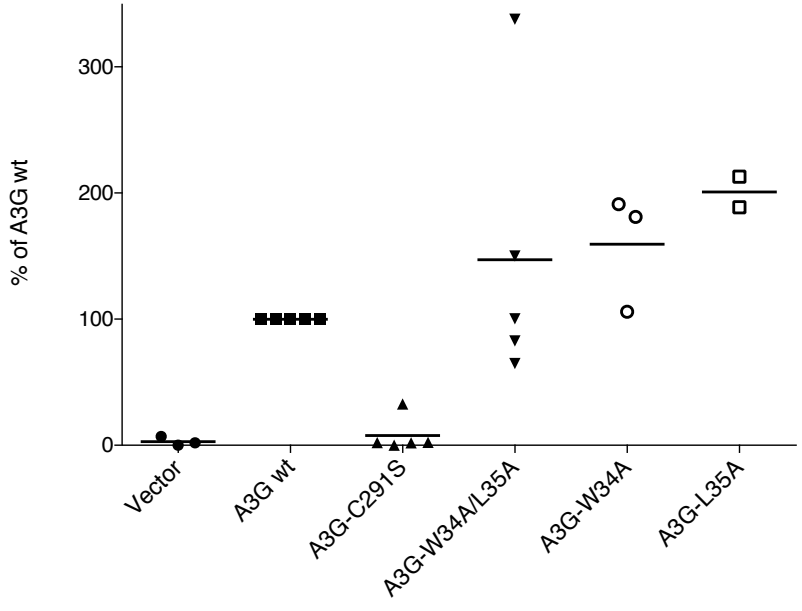
The graphs show the relative abundance of HIV-1 cDNA molecules for each length between nucleotides 23 and 182 of the HIV-$1_{NL4.3}$ (-)sss product, as presented in Supplementary Fig 5. Labels indicate the presence or absence of hUGI and low A3G refers to relative A3G content in the producer cells. The corresponding data for 'high' A3G expression are shown in Fig 3. 'Low' A3G refers to a producer cell transfection ratio of 1:10 (A3G expression plasmid to NL4.3/Δvif).

Supplemental Figure 8



Supplemental Figure 8: Purified proteins used in surface plasmon resonance experiments
Coomassie stained SDS polyacrylamide gels showing purification of A3G, p51 and p66. Fraction 38 of A3G_6xHis and p51_Flag were used in surface plasmon resonance experiments shown in Fig 4.

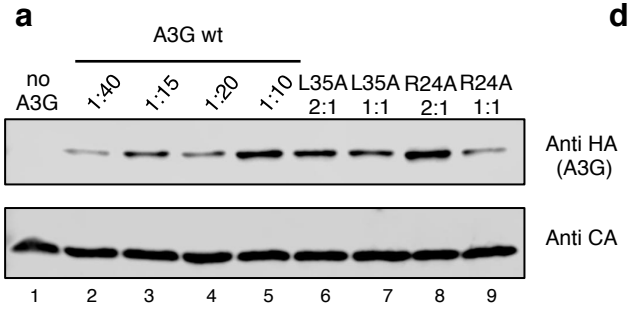
Supplemental Figure 9

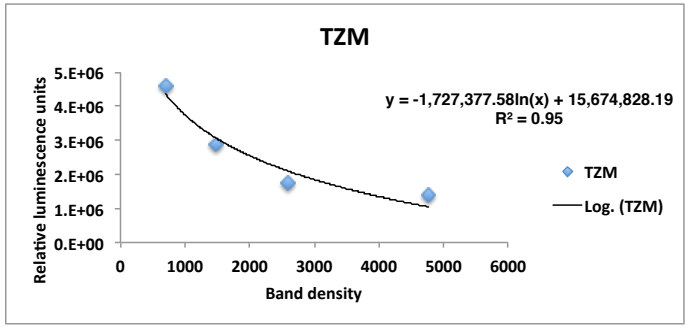


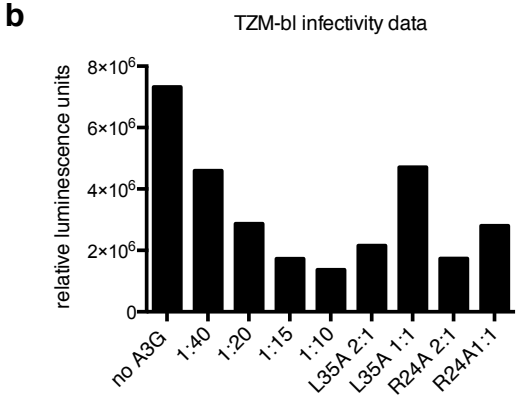
Supplemental Figure 9: Deaminase activity of wild type and mutant A3G proteins measured in a bacterial mutator assay

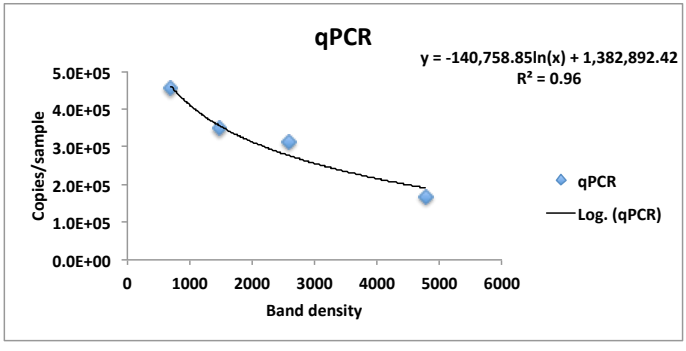
Relative DNA editing ability of A3G proteins were determined by *E. coli* colony forming on rifampicin versus ampicillin selection. Data of at least two replicates are presented relative to wild type function, with means indicated. C291S is a catalytically inactive mutant and served as a negative control.

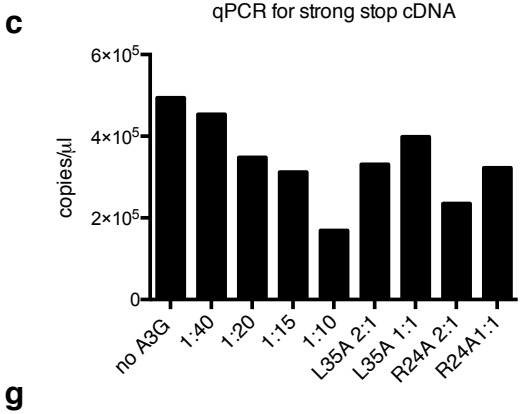
Supplemental Figure 10





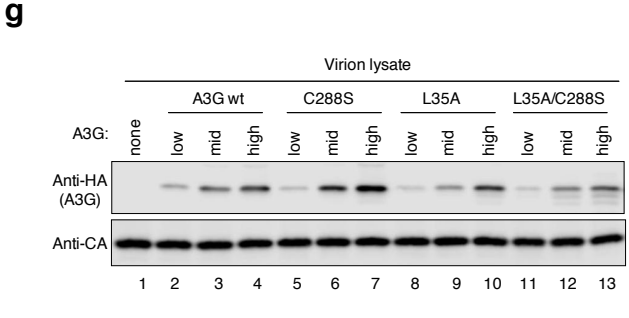


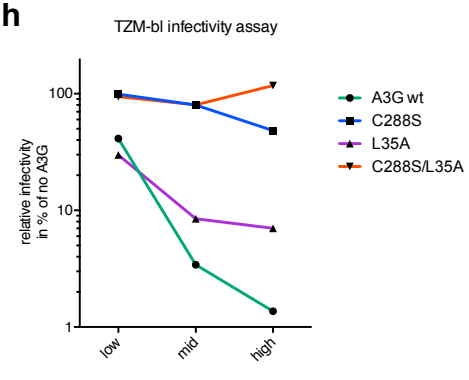




f

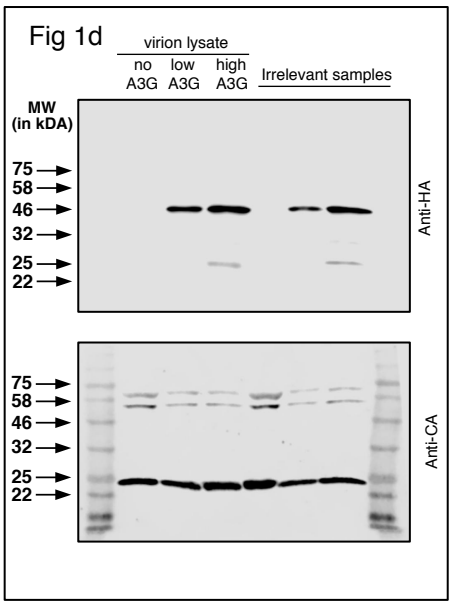
	Band Density	TZM		qPCR	
		Actual Value	Fitted value	Actual Value	Fitted value
no A3G		7320354		494071	
A3G 1:40	700	4595145	4358840	453957	460791
A3G 1:20	1479	2871627	3066444	348119	355478
A3G 1:15	2591	1729111	2098274	312148	276585
A3G 1:10	4770	1371255	1043597	169250	190643
L35A 2:1	3719	2157774	1473782	331052	225698
L35A 1:1	2922	4708594	1890324	398594	259640
R24A 2:1	5493	1733374	800114	235458	170803
R24A1:1	1401	2806662	3159582	322855	363068

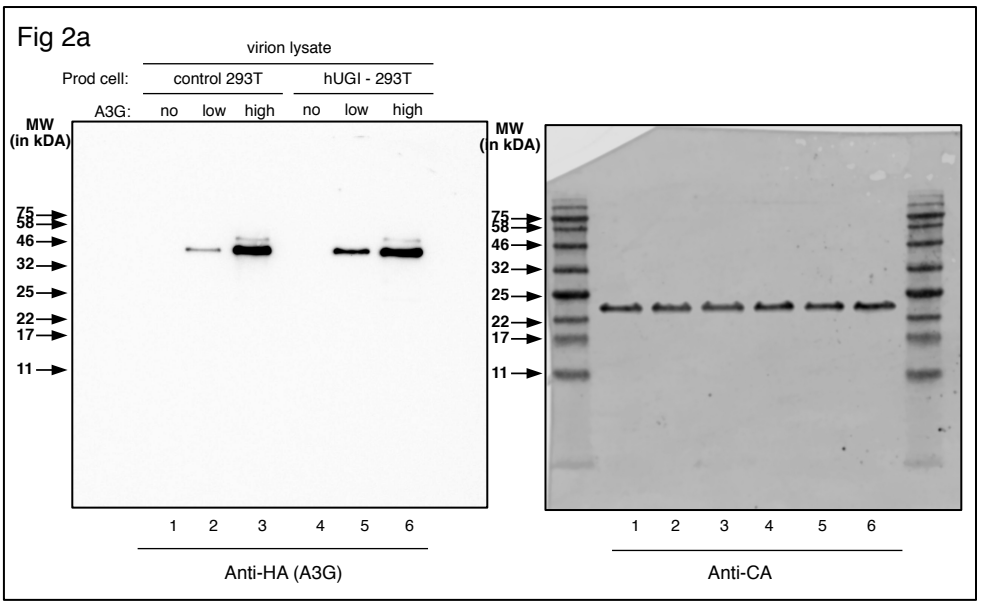


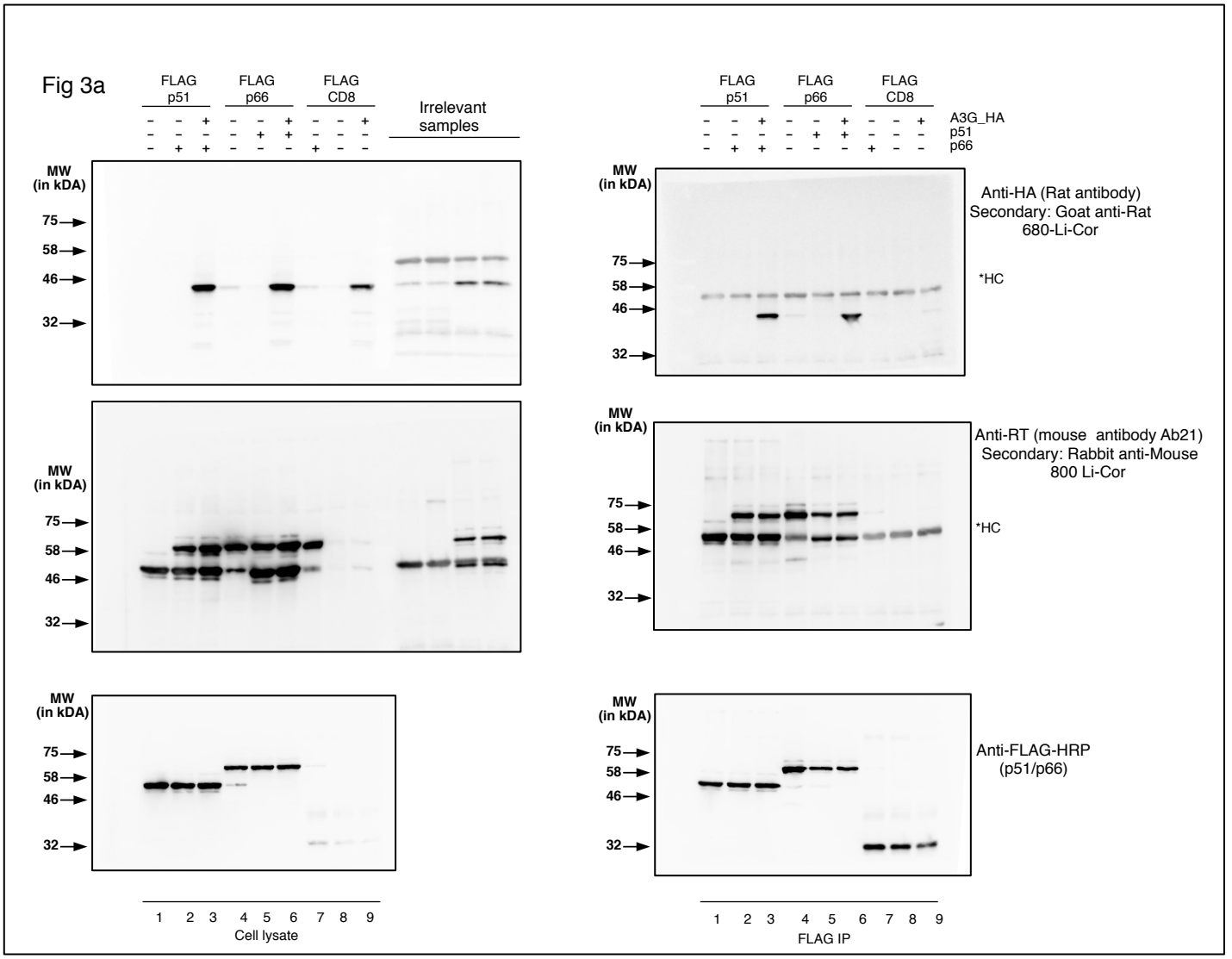


Supplemental Figure 10: Representative examples of data calculation for main Figure 6
a) Immunoblot of virion lysates showing relative packaging of wild type or mutant A3G proteins. Ratios (e.g., 1:40) refer to ratios of µg transfected A3G expression plasmid to proviral plasmid. Note that samples 1:20 and 1:15 are loaded out of order. **b)** Results of TZM-bl infectivity assays with virions from a). **c)** Results of qPCR measuring (-)sss cDNA abundance in CEM-SS cells infected with virions from a). **d)** and **e)** Fitted curves for graphs plotting immunoblot band intensity, representing packaging, over infectivity or cDNA levels respectively. **f)** Actual and standard curve fitted values for each band intensity. Data were then calculated as percent relative to no A3G, as described in the legend for main Fig 6. Note that this experiment has two different concentrations of mutant protein. All other datapoints presented in main Fig 6 are from separate biological replicates, i.e. independent virus preparations in the presence of wild type or mutant A3G. **g)** and **h)** One representative immunoblot and infectivity assay of 5 independent repeats demonstrating titrated packaging levels of wild type and mutant A3G protein. For titration purposes, the transfection ratios of A3G expression plasmid to proviral plasmid for A3G wt and C288S were low: 1:40; mid: 1:10, high: 1:4. Ratios used for L35A and C288S/L35A double mutant were low: 1:2, mid: 1:1 and high: 1:2. The infectivity data of all 'high' A3G levels of all 5 experiments are shown in main Figure 6d.

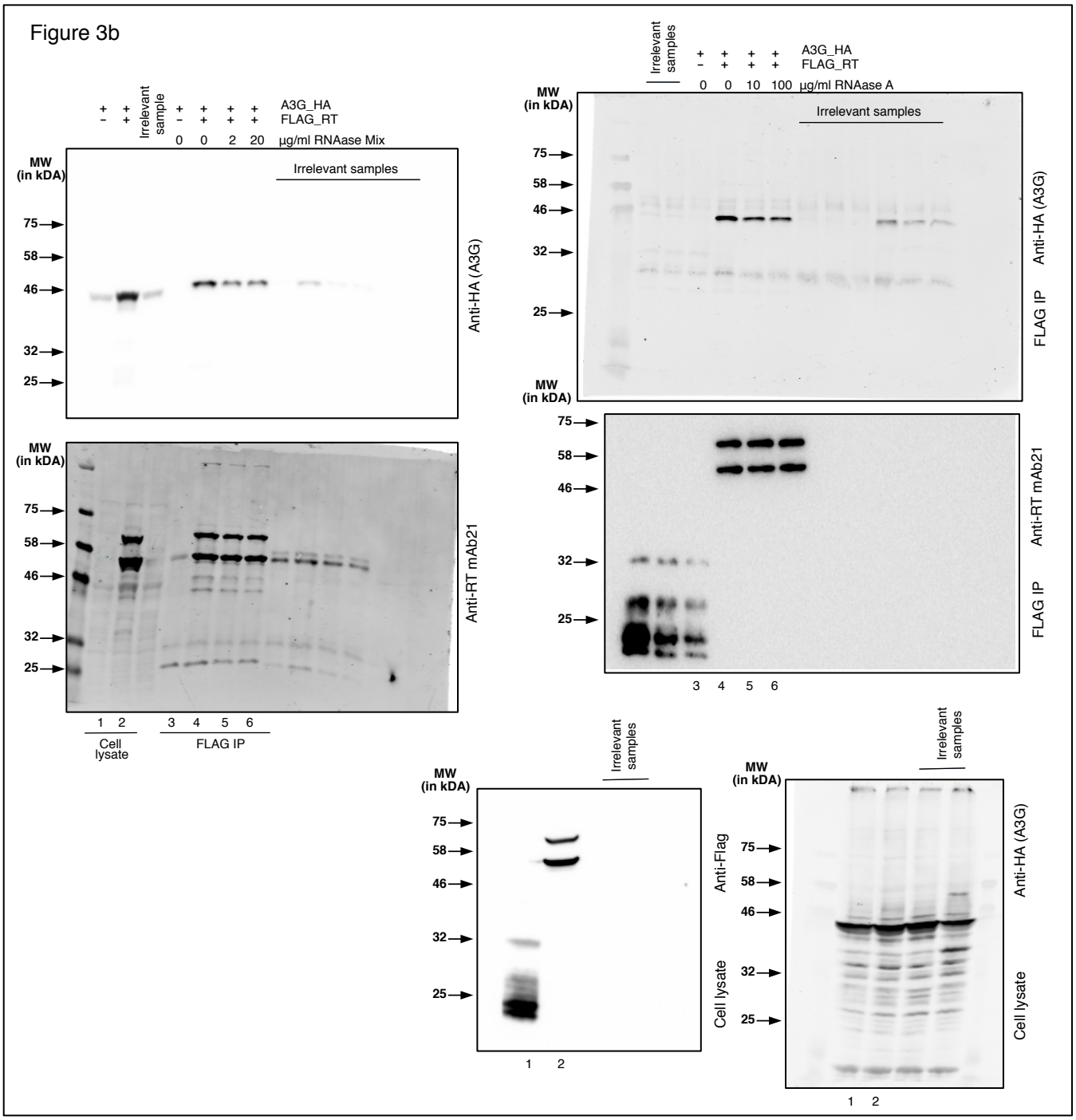
Supplemental Figure 11
Full length versions of all immunoblots presented in main and supplementary figures



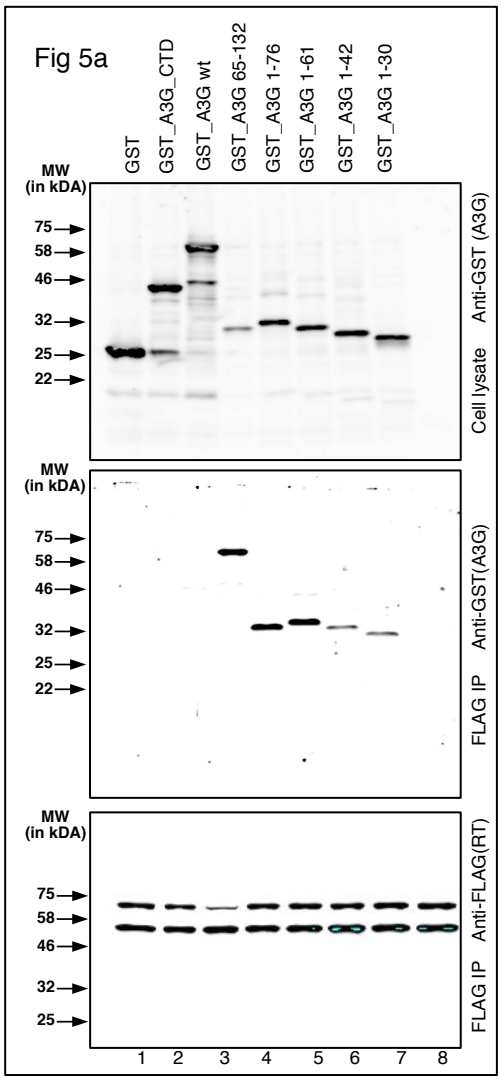


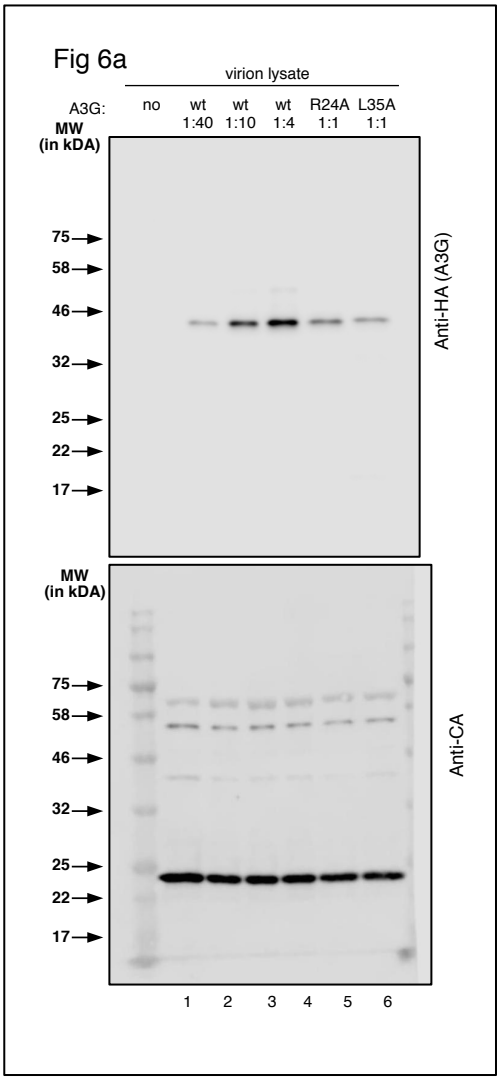


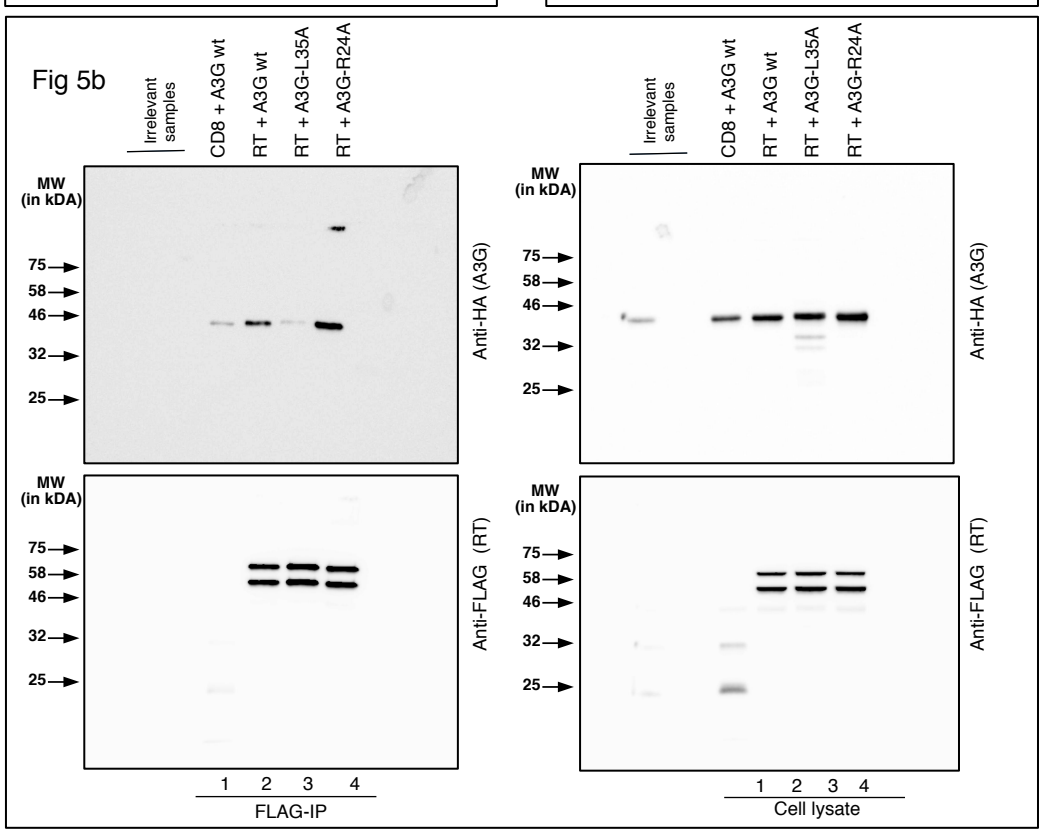
Supplemental Figure 11 - continued
Full length versions of all immunoblots presented in main and supplementary figures



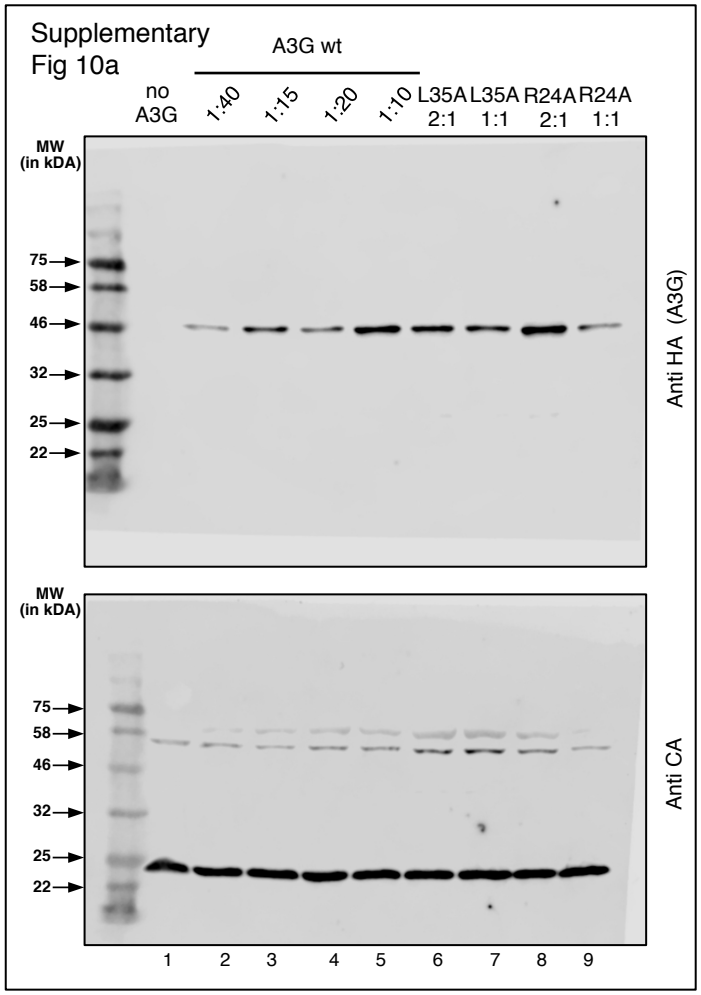
Supplemental Figure 11 - continued
Full length versions of all immunoblots presented in main and supplementary figures



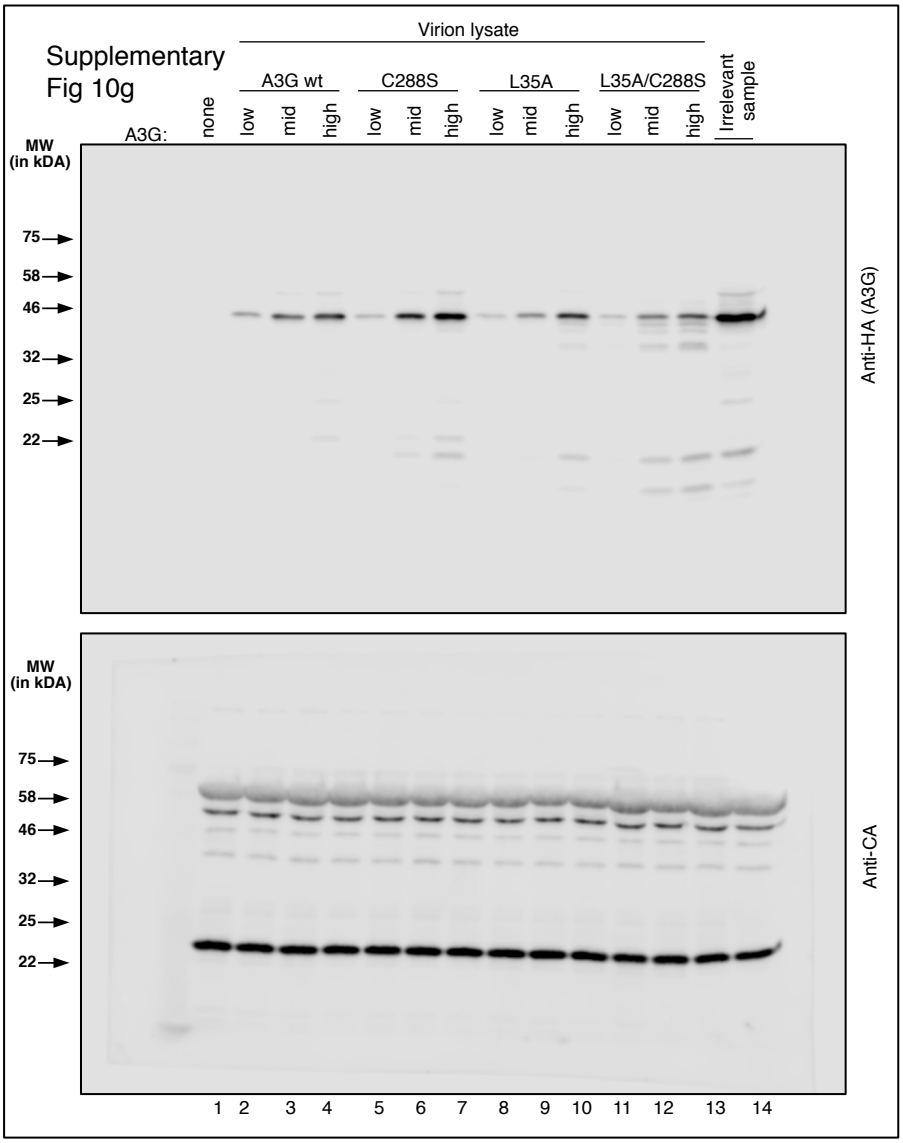




Supplemental Figure 11
Full length versions of all immunoblots presented in main and supplementary figures



Supplemental Figure 11 - continued
Full length versions of all immunoblots presented in main and supplementary figures



Supplemental Figure 11: Full length immunoblots from all main and supplementary figures

Approximate molecular weights are indicated as judged by protein standards (New England Biolabs). Standards are only clearly visible on the blots when exposure was conducted with the Li-cor system and the marker channel overlay did not add background to the remaining blot area. On all other blots molecular weight size marker positions were taken from parallel images (not shown) displaying either the marker channel (Li-cor) or a white light channel (ImageQuant, chemiluminescence system).