Comprehensive characterisation of transcriptional activity during influenza A virus infection reveals biases in cap-snatching of host RNA sequences.

Comprehensive characterisation of transcriptional activity during influenza A virus infection reveals biases in cap-snatching of host RNA sequences

Sara Clohisey[a,b], Nicholas Parkinson[a,b], Bo Wang[a,b], Nicolas Bertin[c], Helen Wise[b,d], Andru Tomoiu[a,b], FANTOM5 Consortium, Kim M. Summers[e], Ross W. Hendry[a], Piero Carninci[g], Alistair R. R. Forrest[f], Yoshihide Hayashizaki[g], Paul Digard[b]*, David A. Hume[e]*#, J. Kenneth Baillie[a]*#

[a] Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

[b] Division of Infection and Immunity, The Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom

[c] RIKEN Center for Life Sciences Technologies, RIKEN Yokohama Campus, Japan

[d] Clinical Biochemistry, Clock Tower Building, Western General Hospital, Edinburgh

[e] Mater Research Institute-University of Queensland, Translational Research Institute, Brisbane, Australia

[f] Harry Perkins Institute of Medical Research, Nedlands, Australia

[g] RIKEN Preventive Medicine and Diagnosis Innovation Program, Japan

Running Title: Biases in cap-snatching of host RNA by IAV

* Joint senior authors

Address correspondence to J. Kenneth Baillie, j.k.baillie@ed.ac.uk , and Prof. David Hume, david.hume@uq.edu.au

Abstract: 216 words, Importance: 150 words

Text Word Count: 5533 words

Abstract

Macrophages in the lung detect and respond to influenza A virus (IAV), determining the nature of the immune response. Using terminal depth 5'-RNA sequencing (CAGE) we quantified transcriptional activity of both host and pathogen over a 24-hour timecourse of IAV infection in primary human monocyte-derived macrophages (MDM). This method allowed us to observe heterogenous host sequences incorporated into IAV mRNA, "snatched" 5' RNA caps, and corresponding RNA sequences from host RNAs. In order to determine whether cap-snatching is random or exhibits a bias , we systematically compared host sequences incorporated into viral mRNA ("snatched") against a complete survey of all background host RNA in the same cells, at the same time. Using a computational strategy designed to eliminate sources of bias due to read length, sequencing depth, multi-mapping, we were able to quantify over-representation of host RNA features among the sequences that were snatched by IAV. We demonstrate biased snatching of numerous host RNAs, particularly snRNAs, and avoidance of host transcripts encoding host ribosomal proteins, which are required by IAV for replication. We then used a systems approach to describe the transcriptional landscape of the host response to IAV, observing many new features, including a failure of IAV-treated MDMs to induce feedback inhibitors of inflammation, seen in response to other treatments.

Importance

Infection with influenza A virus (IAV) infection is responsible for an estimated 500,000 deaths and up to 5 million cases of severe respiratory illness each year. In this study we looked at human primary immune cells, macrophages, infected with influenza A. Our method allows us to look at both the host and the virus in parallel. We used this data to explore a process known as 'cap-snatching', where influenza A snatches a short nucleotide sequence from capped host RNA. This process was believed to be random. We demonstrate biased snatching of numerous host RNAs, including those associated with snRNA transcription, and avoidance of host transcripts encoding host ribosomal proteins, which are required by IAV for replication. We then describe the transcriptional landscape of the host

response to IAV, observing new features, including a failure of IAV-treated MDMs to induce feedback inhibitors of inflammation, seen in response to other treatments.

54 **Introduction**

55 Infection with influenza A virus (IAV) infection is responsible for an estimated 500,000 deaths and up
56 to 5 million cases of severe respiratory illness each year (WHO) (1). The abundant macrophages of
57 the airway and lung interstitium detect and respond to the virus, determining both the nature and
58 the magnitude of the innate and acquired immune response (2), and contribute to systemic
59 inflammatory cytokine production in severe influenza (3).
60 As an obligate intracellular parasite, IAV is reliant on host cellular machinery for replication. The IAV
61 genome comprises 8 negative-sense RNA segments that are transcribed and replicated in the nucleus
62 of the host cell. In order to co-opt host translational machinery, and to evade detection of non-self
63 RNAs by host cells, IAV "snatches" 5' RNA caps from host RNAs. The IAV polymerase binds directly to
64 the 5' 7-methylguanylate cap of a nascent host RNA and cleaves it roughly 10-14 nucleotides
65 downstream. The snatched "leader" sequence is employed as a primer for efficient transcription of
66 the viral mRNA (4) and subsequently, the host cap facilitates translation of viral mRNAs by host
67 ribosomes. Previous large-scale studies of this process (5–8) have produced evidence that host-
68 derived RNA caps are frequently snatched from non-coding RNAs, particularly small nuclear RNAs
69 (snRNAs), due to their high abundance in infected cells. This has led to the conclusion that cap-
70 snatching is not a selective process – that is, that capped host RNAs are snatched at random (8, 9).
71 Previous RNA-seq studies have detected snatched leaders, but have been unable observe the
72 complete pool of *unsnatched* sequences, because of limited sequencing depth and resolution at the
73 5' end, both of which are necessary to accurately quantify the background distribution of each host
74 transcript.
75 To overcome these limitations, we utilised single molecule, terminal-depth cap analysis of gene
76 expression (CAGE) to sequence all capped RNA from primary monocyte-derived macrophages
77 (MDMs) from 4 human donors in vitro at 4 time points over the course of a 24 hour, productive

infection with IAV. The CAGE RNA sequencing method captures both host and virus-derived transcripts and, importantly, does not require a PCR amplification step, thus eliminating PCR bias. By comparing the sequences of the snatched population to the sequences of the total capped RNA background, we observed biases in the snatching of transcripts encoding spliceosome components and avoidance of transcripts encoding host ribosomes.

This methodology allowed us to observe the transcriptional response to IAV infection over time in unprecedented molecular detail. We previously used CAGE to quantify transcript expression, promoter and enhancer activity in human MDM and produced a detailed time course profiling their response to bacterial lipopolysaccharide (LPS) (10). In a comprehensive analysis of the host macrophage transcriptome during IAV exposure, we used a similar systems approach, using co-expression to identify key biological processes (11, 12), and compare the response of MDMs to both IAV and LPS, revealing IAV-specific features of the host response.

93 **Results**

94 **Transcriptional activity of IAV in human MDMs**

95 To observe IAV transcriptional dynamics in human MDMs in vitro, we infected MDMs from four
96 different donors with influenza A/Udorn/72 (H3N2; hereafter, IAV) at a multiplicity of infection (MOI)
97 of 5 (Figure 1 A). RNA libraries were prepared from cells at 0, 2, 7, and 24 hours post-infection and
98 from two uninfected-infected samples at 0 and 24 hours. Libraries were sequenced using HeliScope
99 CAGE as previously described (11, 13). A minority of cells were positive for viral antigen (IAV
100 nucleoprotein) by immunofluorescence after 2 hours and the large majority after 7 hours (Figure 1 B),
101 suggesting that viral mRNA molecules were being transcribed and translated. We confirmed a
102 previous report (14) that IAV-infected MDM cells release infectious virus (Figure 1 C), albeit at
103 approximately 10-fold lower levels compared to published results for permissive cancer cell lines (15,
104 16), and with little evidence of cell death up to 7 hours (Figure 1 D).
105 IAV mRNAs contain a conserved 12-base long 5'-adjacent non-coding region present in all 8 segments
106 ('AGCAAAAGCAGG') derived from template-dependent transcription of the viral promoter (9). This
107 sequence was used to identify viral transcripts. Similar to results seen elsewhere (7, 17, 18), the A at
108 the 5' end of the IAV promoter was not always present and so sequences which contained the 11
109 nucleotide sequence 'GCAAAAGCAGG' (IAV promoter) were brought forward for analysis (Figure 1 E).
110 Most (74%) of the leader sequences, those preceding the promoter, were between 10 and 14
111 nucleotides long (Figure 1 F). Published studies of IAV-infected A549 cells have reported that, within
112 8 hours post-exposure, >50% of total cellular mRNA was viral (19). In contrast, IAV RNA constituted a
113 relatively small proportion (4 -11%) of total capped RNA in MDMs, even at the peak of viral
114 replication (Figure 1 G).
115 The relative proportion of IAV mRNA arising from each viral segment was also consistent across the 4
116 donors at each time point (Figure 2 A), consistent with previous evidence that transcription of each
117 segment is a highly controlled process (20). By 24 hours, the pattern was less defined, which may be

a consequence of mRNA decay and/or potential reinfection of the minor fraction of cells not infected at time 0.

Potential alternative splice variants in IAV

Spicing has been observed in segments 7 and 8 of IAV. In particular, segment 7 contains the splice donor site for the mRNA3/M3 transcript, which is found at the end of the promoter sequence (27) . Over 400,000 reads contained the IAV promoter sequence and a leader sequence, but did not originate from the genome sequence proximal to the promoter in any of the 8 segments. The leader and promoter sequences were removed and the sequences aligned throughout the Udorn genome. In order to quantify RNA expression at these loci, we summed the weighted abundances of reads originating at the same position. This revealed 6,902 putative capped IAV RNA sequences from the IAV genome, including the known splice variant of segment 7, the mRNA3 transcript (Figure 2 B). The alignments observed (Table S1) are likely to include previously unidentified splice variants. However, in a systematic search, no putative IAV splice variant RNA was preceded by a canonical major spliceosome acceptor site, apart from the mRNA3 transcript. It is possible these represent variants that are expressed in such low amounts they are not detectable by other means, for example northern blot or radioactive primer extension. It is of interest to determine if these putative mRNAs are true transcription products and if their transcription and translation contributes to viral pathogenesis.

Characterisation of host leader sequences incorporated into viral capped RNA

We identified 4,575,918 unique leader sequences, heterogeneous in both sequence and length, snatched from the host and incorporated into viral mRNA. Contrary to previous reports (5, 8), we observed no difference in leader lengths between different viral segments. 18.8% (859,789) of leader sequences appeared more than once and 1.5% (69,443) appeared ten times or more across all samples, indicating the presence of a highly snatched population.

142 We sought to determine whether there was over-representation of particular sequences, host
143 transcripts, or biological pathways in the population of leader sequences compared to the
144 background population of CAGE reads. In order to eliminate the risk of bias due to the different rates
145 of successful mapping for sequences of different lengths, we restricted our analysis to the first 10
146 bases of every CAGE tag (10mers), including both IAV and host sequences. The number of times a
147 10mer was followed by the IAV promoter, i.e. incorporated into viral mRNA ("snatched"), was
148 compared to the number of times a 10mer was not followed by an IAV promoter ("unsnatched")
149 using Fisher's Exact test (FDR <0.05) at each time point. Of 29,195 10mers meeting our minimum
150 count threshold of 1000 reads, we assigned a host transcript identity to 12,992 (44.5%). The
151 remainder are a mixture of alternative host promoters, lncRNAs, eRNAs and other RNA species (21).
152 Within these named 10mers, 6,353 mapped ambiguously to more than one transcription initiation
153 site so a single identity was chosen at random from the possible sites. This approach decreased
154 discovery power, but was necessary to avoid bias that might be introduced into the identification
155 based on a quantitative measure, such as abundance. The 1,000 most significantly enriched named
156 genes in the snatched and unsnatched sets are reported in Table S2.

157 **Host snRNA is targeted by the cap-snatching mechanism**

158 Key spliceosome snRNAs (RNU1, RNU11, RNU12, RNU4ATAC, RNU5A, RNU5E, RNU5F, RNU5D, RNU7)
159 and their variants/pseudogenes were among the most significantly enriched named genes. This is
160 consistent with previous observations that snRNAs are snatched frequently (5, 18) and shows that
161 this may represent a true preference for these RNAs. In view of this apparent preferential snatching
162 of multiple snRNAs, we considered whether specific classes of capped host RNAs might be targeted.
163 Of the RNA types considered, only snRNAs were strongly over-represented in the snatched
164 population (Figure 3 A, B). It is unclear at this time if both mature snRNA (incorporated into the
165 spliceosome) and immature snRNA (prior to nuclear export and processing) are snatched as leader
166 sequences.

167 This sequencing method also allows the observation of histone mRNA which enabled us to observe
168 that 10mers corresponding to histone mRNAs were also significantly over-represented.
169 The 10mer corresponding to the transcript encoding the largest subunit of RNA polymerase II
170 (*POLR2A*), was 5.83-fold over-represented in snatched sequences (OR = 5.83; FDR <0.05). *PABPN1*,
171 which encodes poly(A) binding protein, was also over-represented (OR = 2.28; FDR <0.05). These
172 comprise key elements of both transcription and polyadenylation of host mRNAs.
173 Taken together these observations might imply that cap-snatching interferes with regulation of
174 transcription and splicing in the infected cell. However, *POLR2B*, another subunit of RNA polymerase
175 II, was 7.77-fold under-represented (OR = 0.13; FDR < 0.05) making it difficult to draw simple
176 conclusions from enrichment analysis that rely primarily on overlap statistics. To rectify this, we
177 performed further gene set enrichment analyses that take into account background gene expression
178 to determine statistically over- and under-represented pathways affected by the cap-snatching
179 mechanism.
180 **Specific ribosome-associated transcripts are avoided by the cap-snatching mechanism**
181 Identified transcripts from all time points and donors were collated and gene set enrichment analysis
182 (FGSEA) performed by querying various pathway/gene ontology datasets (listed in Table S3).
183 Querying Reactome identified a single over-represented pathway: RNA Polymerase II transcribes
184 snRNA genes (Figure 3 C). A volcano plot highlighting the distribution of pathway members shows
185 that many mRNA pathway members were under-represented and its enrichment as an over-
186 represented pathway was driven by snRNA transcripts (Figure 3 E), particularly snRNA members of
187 the minor spliceosome. This is consistent with the observed apparent preferential snatching of
188 snRNAs.
189 Pathway enrichment also allowed us to look for pathways that were *avoided* by the cap-snatching
190 mechanism. We identified pathways associated with translation and ribosome formation as
191 significantly under-represented in the cap-snatched pool (Figure 3 D). Although multiple pathways

were identified, these were not independent: these associations were largely driven by presence of a group of transcripts encoding the same set of ribosomal proteins (Table S3). These data show that IAV avoids snatching caps from ribosomal mRNA transcripts. Interestingly, not all mRNAs encoding ribosomal subunits were avoided. We compared our results to a recent study reporting the effect of targeted knockdown of specific ribosomal subunit mRNAs in the context of IAV infection (22), but saw no clear relationship between cap-snatching preference and viral protein production, host protein production, or antigen presentation.

Enrichment of specific RNA motifs in the snatched and unsnatched sequence populations

Leader sequences are known to commonly have 'GCA' at the interface between the host sequence and the IAV promoter (23, 24) introduced partially through the "prime and realign" mechanism (17, 24, 25). More recently, an 'AG' at the 5' end of the leader sequence has also been shown to be prevalent in snatched sequences (6).

Our analysis of 10mers enables a statistically powerful comparison of snatched and unsnatched sequences in which the position of sequence motifs can be compared without reference to distance from the 5' or 3' ends. We used Pysster (26) to train convolutional neural networks using the sequence data to explore sequence and positional features, of a length of 4 nucleotides, for pools of highly-significantly over-represented snatched and unsnatched 10mers (0.3 >= OR >= 3, -log(FDR < 10). This stringency was introduced to eliminate potential noise. The snatched 10mers showed an enrichment of two motifs, A[G/C][T/A][C/G] and the similar sequence AGNN, both beginning at the first base (position 0) (Figure 4 A). These motifs were most apparent 2 hours post infection, coinciding with levels of high transcription by the virus and are consistent with previous reports of an 'AG' preference at the 5' end of the leader (6).

The unsnatched 10mers also showed an enrichment of two distinct motifs, CTAG and [T/C][A/T][T/G/A]A, most evident at 7 hours post infection (Figure 4 B). While the CTAG motif was unsnatched primarily when it began in the first position (position 0), there was also an association

between this motif at any position in the 10mer and unsnatched status. Similarly, the [T/C][A/T][T/G/A]A motif was avoided by cap snatching if it occurred at any position within the 10mer (Figure 4 B). To our knowledge this is the first evidence for the avoidance of particular sequences as priming leaders by the IAV polymerase.

Network analysis of the response to IAV virus infection in MDM

Temporal changes in host cell transcription are likely to occur both in recognition of viral infection and as a consequence of viral lifecycle progression. IAV can dysregulate host transcription, in a manner which leaves transcription initiation apparently unaffected (28). The advantage of CAGE in this scenario is the snapshot of transcription initiation it provides, in contrast to other technique, such as RNA-Seq which sequence the entire mRNA molecule, including downstream-of-gene transcripts (29).

We utilised the network analytical tool, Graphia (30), to identify sets of co-regulated transcripts in the MDM response to IAV (Table S4). For simplicity, we restricted the analysis to the dominant (most frequently used) promoters (p1) and used averaged data from the 4 donors. We have summarised the GO term enrichment and pathway enrichment in the 10 largest clusters using GATHER (31) (Table S5) and Enrichr (32, 33) (Table S6) respectively.

Figure 5 A shows the sample-to-sample correlation graph for each of the averaged data sets. Although there was a global alteration in transcript induction that progressed with time, the profile at 7 hours remained correlated with the profiles in uninfected cells at both early and late time points. This suggests that the virus did not cause a selective, or global, loss of host transcription initiation. In keeping with that conclusion, the largest cluster, Cluster 1, contained more than 4,500 genes (Figure 5 B) whose shared pattern was continuous induction across the time course with particularly high initiation at 24 hours. This cluster contained genes encoding the interferon-responsive transcription factors, *IRF1, 2, 4, 7, 8*, and *9* and numerous known interferon-responsive antiviral effector genes (e.g. *APOBEC3G, RSAD2, DDX58, ISG15, MX1, OAS1, TRIM25*).

242 We observed that the response of MDMs to viral infection was immediate. IL1B was rapidly and
243 strongly induced by IAV at 0 hours (effectively 1 hour post virus addition) and peaked at 2 hours (3
244 hours post virus addition) (Figure 5 C). Other early response genes that were detected early after IAV
245 exposure included those encoding immediate early transcription factors such as EGR1, the
246 proinflammatory cytokine TNFα and the neutrophil chemoattractant CXCL2 (Figure 5 D-F).
247 Rapidly-induced genes are concentrated in clusters 3 and 4, including *interferons IFNB1, IFNA1,*
248 *IFNA2, IFNA8, IFNA14, IFNE* and further known IFN-regulated targets such as *IFI6, IFIT2,*
249 *IFITM3, IRG1, GBP1 and MNDA*. Also enriched in these clusters are genes involved in protein
250 synthesis, including 46 ribosomal protein subunit genes, which are avoided by IAV cap-snatching (see
251 above).
252 **Comparative analysis of the response of MDMs to treatment with IAV and with LPS**
253 The response of MDMs to IAV and LPS was compared at equivalent time points, uncovering some
254 common transcripts that were induced in both treatments (Figure 5 H, top row). Transcripts induced
255 specifically by LPS but not by IAV were revealed by differential expression analysis (Figure 5 G, Table
256 S7) and included classical inflammatory cytokines IL12B (although not IL12A) and IL6, and the
257 feedback regulator of inflammation, IL10 (Figure 5 H, central row; Figure 6 A, B). Conversely,
258 induction of genes associated with interferon signalling was more substantial and prolonged in IAV-
259 treated MDM than those treated with LPS. IAV induced *IFNB1* mRNA some 10-fold more than
260 observed in response to LPS in MDM, and sustained this expression throughout the time course
261 (Figure 5 H, bottom row). IAV also induced multiple IFNA genes (*IFNA1, A2, A8, A14, A22*, Figure 6 C-
262 E) and the type III interferon genes, *IFNL1* (aka I*L28A*) and I*FNL2* (aka *IL29*), which were not induced
263 at all by LPS (Figure 5 H, bottom row; and Figure 6 F).
264

Discussion

This comprehensive analysis of host and viral transcripts reveals key features of the host-pathogen interaction at a molecular level. We demonstrate that IAV cap-snatching is biased towards host transcripts associated with splicing, and avoids host ribosomal subunit transcripts. Additionally, we provide a comprehensive analysis of transcripts initiated as part of the host response to IAV in a vital innate immune cell.

Elimination of bias for accurate quantification of leader sequences and 5' RNA ends

Our choice of sequencing methodology and analytical approach eliminated numerous sources of bias that have limited the interpretation of previous studies of cap-snatching preference.

The HeliScope single molecule CAGE sequencing methodology sequences transcripts from the 5' end without internal segment-specific primers, and without PCR amplification (13). In contrast, previous studies of IAV virus transcripts used internal primers for the viral segments (5, 8) or performed library amplification on cDNA derived from capped RNA (6). A key difference from previous work is the quantification of background transcription, which enables the first accurate quantification of the transcripts not snatched by IAV.

In addition, our use of terminal-depth sequencing limits noise and sampling error, in both the snatched sequences and the background distribution. Since CAGE reads sequences directly from the 5' end, we can be confident that we have quantified the background pool of potential leader sequences that were available to be snatched. By limiting our analysis to sequences of a specific length (10mers), we eliminate bias that may occur due to differential mapping or identification of sequences of different lengths.

The cap-snatching mechanism is not entirely random

Although random cap-snatching does occur, 18.8% of leaders are snatched multiple times and our analysis shows that many are snatched more frequently than one would expect from the level of background RNA expression. Non-coding RNAs, particularly snRNAs, have been identified as the

source of the most frequently snatched leader sequences (5, 6). However, it was unclear whether this frequency reflected the high abundance of these transcripts in infected cells, or a true over-representation of this RNA type among leaders. Our analysis, enables an unbiased, accurate quantification of the abundance of each sequence in both the snatched, and unsnatched, sequence sets.

Differential expression analysis revealed that all snRNAs, apart from RNU1, were upregulated in IAV-treated MDMs compared to LPS (Table S7). Notably, snRNA components of the minor spliceosome, (*RNU11, RNU12, RNU4ATAC, RNU5A* and *RNU5E*), a molecular machine that splices <1% of introns in the human genome, were highly snatched compared to background expression, particularly at 2 and 7 hours.

If cap-snatching were only determined by abundance, as previously thought (8, 9), we would expect to see leader sequences derived from ribosomal genes prominently among the snatched sequences. Our comparison of LPS and IAV-treated cells shows that genes encoding ribosomal subunits are highly transcribed in IAV-treated cells. Although we do see a minority of mRNAs encoding ribosomal proteins in the snatched set, IAV cap-snatching exhibited a surprisingly strong avoidance of most mRNAs encoding ribosomal proteins, which is particularly evident in pathway enrichment analysis.

Characteristics of MDM response to IAV

The snapshot of transcription initiation provided by CAGE analysis allowed us to examine co-expression clusters and observe the consistent similarity in global transcription initiation between uninfected and early post-infection time point MDMs, suggesting that most basic cellular processes are maintained during infection in this model. The largest co-expression cluster, Cluster 1 (Table S4), included genes encoding the ubiquitin-proteasome complex, oxidative phosphorylation, cell cycle and transcriptional regulation including mRNA splicing and binding. In A549 cells, IAV infection causes cell cycle arrest (34), and down-regulation of cell-cycle associated genes. Since MDM are not actively proliferative, the apparent induction by IAV infection of many cell cycle-related genes, including

cyclin genes and 19 genes encoding multiple cyclin-dependent kinases (CDK) is unlikely to be associated with cellular proliferation.

The comparison between the host response in IAV and LPS treated MDMs

Like LPS, IAV strongly induced TNFα, IL1B, multiple chemokine genes (e.g. *CCL2, CCL3, CXCL1, CXCL2, CCL20*) and many genes for immediate early transcription factors (e.g. *EGR* family). However, the global transcript initiation-based analysis of the response to IAV reveals a clear contrast to the LPS response in MDMs. In LPS-treatment, levels of many inflammatory transcripts are subject to control by a complex network of rapidly-induced feedback regulators (10). The sustained induction of proinflammatory transcripts in response to IAV contrasts with this transient induction in response to LPS.

Following LPS treatment, MDM have low initiation of *IL12A* (p35) mRNA (Figure S 4 B), instead inducing *IL23A* and *IL12B* mRNA, which together encode the heterodimeric proinflammatory cytokine IL23. These were not detected in IAV-infected cells. Similarly, there was no detectable induction of the anti-inflammatory cytokine *IL10* mRNA by IAV, while transcript initiation was massive and sustained in LPS treated cells.

The type III interferons were specific to IAV-treated MDMs. These were recently shown to mediate a key mechanism preventing viral spread to the lower respiratory tract in mice (35), which is believed to cause life-threatening disease in humans (36). The profound difference in induction of IFN-responsive genes in this cell type between LPS and IAV stimulation is reflective of blood transcriptome profiles of patients with severe IAV compared to those with bacterial sepsis (37).

Limitations of this study

Our study is to our knowledge the most comprehensive systems-level evaluation of both host and viral transcriptional activity for IAV replication, and the first study to perform an unbiased quantification of cap-snatching preference compared with accurate measurement of background transcription. It is, however, limited to a single cell type and one strain of IAV. It is possible that the

observed apparent preference and avoidance of specific capped RNA are specific to MDMs. The observation that snRNAs RNU1 and RNU2 are the most frequently snatched sequences in H1N1 infected A549 cells in other studies (17, 18), indicates that it is reasonable to speculate this mechanism is generalizable across other types. Finally, our method measures total capped-RNA and does not differentiate between nuclear and cytoplasmic RNA molecules.

Future work is needed to explore the mechanisms underlying the preference and avoidance of specific mRNAs, and to determine cap-snatching preferences of other IAV strains.

348 **Materials and Methods**

349 **Ethics, cell culture, virus propagation and infections**

350 Cells were isolated from fresh blood of volunteer donors under ethical approval from Lothian
351 Research Ethics Committee (11/AL/0168). Primary CD14+ human monocytes were isolated from
352 whole blood as described previously (38) from 4 human donors. Monocytes were plated for 7 days in
353 RPMI-1640 supplemented with 10% (vol/vol) FBS, 2 mM glutamine, 100 U/ml penicillin, 100 µg/ml
354 streptomycin (Sigma Co.), and 104 U/ml (100 ng/ml) recombinant human colony-stimulating factor 1
355 (rhCSF1; a gift from Chiron, Emeryville, CA, USA) for differentiation into macrophages. Cells were
356 maintained at 37°C with 5% CO2. A/Udorn/72 (H3N2) was generated as described previously (14).
357 Differentiated macrophages were infected on day 8. Cells were washed in serum free media after
358 which they were infected at MOI 5 in a volume of 200µl infection media. Cells were incubated for 1
359 hour at 37°C then washed three times with serum-free media and incubated in RPMI-1640
360 supplemented with 1µg/ml TPCK-trypsin, 0.7% BSA, 2mM glutamine, 100 U/ml penicillin, 100 µg/ml
361 streptomycin (Sigma Co.), and 104 U/ml (100 ng/ml) rhCSF1. Samples were collected at 4 time points
362 post infection/media change: 0 hour (1 hour after addition of the virus), 2 hours, 7 hours and 24
363 hours. Uninfected samples were also collected at 0 and 24 hours. LPS treatments were carried out
364 as described previously (10). Only time points with corresponding IAV treated time points were used
365 in this analysis.

366 **Immunofluorescence**

367 Primary human monocyte derived macrophages were differentiated, as described above, on glass
368 coverslips. Cells were infected as described. At 0, 2, 7, and 24 hours post infection cells were fixed for
369 20 min in 4% formaldehyde in PBS. After permeabilisation with 0.2% Triton X124 100 in PBS for 5 min
370 at room temperature, cells were incubated with mouse monoclonal influenza A NP AA5H (BioRad) at
371 1:500. After 1 hour cells were washed three times with PBS

and incubated with goat anti-mouse Alexa Fluor 488 at 1:1000 (ThermoFisher). After 1 hour cells were washed three times with PBS and incubated in DAPI (ThermoFisher) for ten minutes after which they were washed three times with PBS and mounted on slides using VECTASHIELD® Antifade Mounting Medium. Cells were viewed on a Leica fluorescence upright microscope and imaged using a Hamamatsu Orca-ER low light mono camera. Scale bars were added using ImageJ.

Cell viability and Virus Titration

Cell viability was measured using Cell Titre Glo ® at 0, 2, 7, and 24 hours post infection. Virus produced was titrated by plaque assay on MDCK cells. Virus titres in cell supernatants were determined by plaque titration using ten-fold serial dilutions of virus stocks. Confluent MDCK cells in 6 well plates were inoculated with cell supernatant for 1 hour in serum-free medium. An overlay (mixture of equal volume of DMEM and 2.4% Avicel (Sigma-Aldrich,UK) supplemented with 1 µg/ml TPCK-treated trypsin and 0.14% BSA fraction V) was then put onto the wells. After 48 hours, cells were fixed using 3.5% formaldehyde and stained with 0.1% crystal violet. Virus titres were calculated by (plaque count*dilution factor/(volume of inoculum)) and expressed as plaque forming units per millilitre of supernatant (pfu/ml).

CAGE

RNA was extracted using the Qiagen miRNeasy mini kit (217004). RNA quality was assessed and CAGE was performed as described previously (39) as part of the FANTOM5 project. Virus genome information is available in Table S8.

Data analysis and identification of IAV mRNA

Computational analysis was performed using custom Python scripts and as described previously (11). Capped IAV RNAs were identified by the conserved 11 base promoter sequence expected to be in all viral mRNA ('GCAAAAGCAGG'), as described in the text. Sequences that contained the promoter were classified as capped viral mRNA and aligned to the Udorn sequence.

Unbiased analysis of leader sequence preference

397 The first ten nucleotides of each CAGE tag (10mers) that reached the abundance threshold in our
398 dataset were extracted and this set of unique 10mers were used in subsequent analysis. The
399 abundance threshold was set to 1,000 occurrences across all samples. To determine the 10mer
400 sequences that were over- and under-represented in the snatched population based on background
401 abundance, the number of times a 10mer was associated with the IAV promoter was counted
402 ("snatched") along with the number of times the 10mer occurred without the promoter
403 ("unsnatched"). These were analysed using Fisher's Exact test. Benjamini-Hochberg correction was
404 applied to p-values. Significance was determined by an FDR < 0.05. The number of times a 10mer
405 was snatched was compared to the number of times it occurred unsnatched at the previous time
406 point by Fisher's Exact test.

407 **Assignment of transcript identity to 10mer sequences**

408 CAGE tags were mapped to the human reference genome (hg19) as described (11). We extracted
409 every possible chromosomal location for a 10mer that met the abundance threshold of 1000 across
410 all samples from the original alignment BAMfiles created as part of the Fantom5 project. 10mers
411 containing a 6mer from within the IAV promoter ('GCAAAA', 'CAAAAG', 'AAAAGC', 'AAAGCA',
412 'AAGCAG', 'AGCAGG') were removed. Reference transcription start sites were downloaded from
413 Fantom5. Promoter identity was assigned first using BEDtools 2.25.0 with a window of +/- 5 bases
414 and exact strand match only. For each possible promoter identity the 10mer sequence was mapped
415 to the genomic sequence with a window of +/- 5 bases directly and exact matches only were used to
416 assign promoter identity.

417 A promoter identity was chosen at random from the list of mapped sites, to avoid any effect of
418 abundance that may bias transcript identification. Promoter identities were converted to HGNC
419 format. To determine preference of promoters and genes in leader sequences, all 10mers that were
420 assigned to that promoter or gene name were counted and the Fisher's Exact test was performed.
421 Benjamini-Hochberg FDRs were calculated using the scipy.stats v 0.18.1

statsmodels.stats.multitest.mutlipletests function with method = 'fdr_bh'. Significance was determined by an FDR < 0.05. RNA type was assigned to named transcripts using reference data downloaded from Biomart (http://www.ensembl.org).

Pathway and Gene Set Enrichment Analysis

GO term assignment and pathway analysis for coexpression clusters were performed using Enrichr (mp.pharm.mssm.edu/Enrichr) (32, 33) and GATHER (31). Pathway databases queried were: Reactome 2016, KEGG 2016, WikiPathways 2016 and GO Molecular Function 2015, GO Cellular Component 2015 and GO Biological Process 2015. Gene Set Enrichment analysis on ranked cap-snatching preference data was performed using R package FGSEA (40), in R version 3.5.1, with the following parameters: set.seed = 42, min set size = 5, max size = 5000, nproc = 1, nperm = 1000000. Gene set libraries KEGG 2016, BioCarta 2016, Reactome 2016, WikiPathways 2016, NCI Nature 2016, GO Biological Process 2018, GO Molecular Function 2018, and GO Cellular Component 2018 were used. Genes were ranked by -log10(p-value), and log10(OR). Benjamini-Hochberg correction was applied to p-values. All named genes that appeared significant were included in this analysis.

Analysis of leader motifs using convolutional neural networks

A sub-set of 10mers that reached the following threshold: 0.3 < (OR) > 3, -log(FDR < 10) were brought forward for analysis of motif preference using convolutional neural networks. We optimised an existing network (26) for our use by using altering the parameters to find suitable settings. Optimisation experiments demonstrated that a kernel length of 4 gave us relatively high, and relatively consistent, precision and recall. by using the grid search to explore various kernel lengths (2, 3, 4, and 5) and drop rate (0, 0.1, and 0.5); for other parameters, we used the default settings of Pysster (kernel number: 20, convolutional layer number: 2) apart from learning rate at 0.0001 and patience, stopping at 100. Since our analysis was restricted to 10mers, we did not use the pooling method. We randomly selected the training set and validation set in the proportion of 60% and 30%

446 independently. Motifs were considered if they reached a score of at least 50% the maximum score
447 for that time point.

448 **Identification of potential alternative splice variants**

449 CAGE tags containing a leader sequence and an IAV promoter sequence followed by a sequence that
450 did not align proximal to the IAV promoter sequence in the Udorn genome were extracted. These
451 novel 'promoter proximal' sequences were hypothesised to be derived from putative 5'UTR
452 sequences internal to a segment arising from mRNA from splice variants. These sequences were
453 aligned throughout the Udorn genome using custom Python scripts. The abundance of each
454 sequence was divided by the number of locations in the Udorn genome it could map to. The
455 weighted abundances at each position were then summed and graphed. Segment 7 mRNA3 was used
456 as a proof of principle

457 **Network Analysis of the MDM transcriptome during infection.**

458 Network analysis of the MDM transcriptome during infection was carried out using Graphia
459 Professional (Kajeka Ltd., United Kingdom; http://www.kajeka.com) -formerly Biolayout Express3D.
460 Results were filtered to exclude any transcript where the maximum value across all samples did not
461 reach 10 tags per million (TPM). The sample-to-sample analysis was performed at a Pearson
462 correlation coefficient of ≥ 0.70. The gene-to-gene analysis was performed at a Pearson correlation
463 coefficient of ≥ 0.94 and used a relatively coarse Markov cluster algorithm inflation value of 1.7 to
464 avoid excessive cluster fragmentation. We restricted the analysis to the dominant promoters (p1)
465 and used averaged data from the 4 donors.

466 **EdgeR analysis of LPS treated versus IAV treated samples.**

467 Differential expression between groups of genes was analysed using the EdgeR package (41) in R
468 version 3.5.1. CAGE data for LPS and IAV datasets were processed as described previously (10).
469 Clustered transcription start site (CTSS) with a minimum expression level of 10 tags per million in at
470 least one comparable time point, and with a coefficient of variation > 0.5, were included in

expression analysis. Samples corresponding to 7 hours post-treatments were carried forward for analysis. We used the glmFit function to fit the models and glmLRT to perform testing between the LPS and IAV treated samples. Benjamini-Hochberg correction was applied to p-values. A significance threshold of FDR < 0.05 was used.

Data Availability

Custom Python scripts are available at: https://github.com/baillielab/influenza_cage.

CAGE data is available to download from http://fantom.gsc.riken.jp/5/data/.

480 **Acknowledgements**

481 JKB gratefully acknowledges funding support from a Wellcome Trust Intermediate Clinical Fellowship
482 (103258/Z/13/Z), a Wellcome-Beit Prize (103258/Z/13/A), and the UK Intensive Care Society. JKB,
483 DAH, PD, and HW acknowledge support from BBSRC Institute Strategic Programme Grants
484 BB/J004324/1 and BB/P013740/1. KMS and DAH are supported by the Mater Foundation, Brisbane,
485 Australia.

486

487

488

489

490 References

491 1. WHO. http://www.who.int/mediacentre/factsheets/fs211/en/. WHO.

492 2. Cline TD, Beck D, Bianchini E. 2017. Influenza virus replication in macrophages: Balancing
493 protection and pathogenesis. J Gen Virol 98:2401–2412.

494 3. Short KR, Veeris R, Leijten LM, Van Den Brand JM, Jong VL, Stittelaar K, Osterhaus ADME,
495 Andeweg A, Van Riel D. 2017. Proinflammatory Cytokine Responses in Extra-Respiratory Tissues
496 during Severe Influenza. J Infect Dis 216:829–833.

497 4. Plotch SJ, Bouloy M, Ulmanen I, Krug RM. 1981. A unique cap(m7GpppXm)-dependent
498 influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA
499 transcription. Cell 23:847–858.

500 5. Koppstein D, Ashour J, Bartel DP. 2015. Sequencing the cap-snatching repertoire of H1N1
501 influenza provides insight into the mechanism of viral transcription initiation. Nucleic Acids Res 43:1–
502 13.

503 6. Gu W, Gallagher GR, Dai W, Liu P, Li R, Trombly MI, Gammon DB, Mello CC, Wang JP, Finberg
504 RW. 2015. Influenza A virus preferentially snatches noncoding RNA caps. RNA 21:2067–75.

505 7. Sikora D, Rocheleau L, Brown EG, Pelchat M. 2014. Deep sequencing reveals the eight facets
506 of the influenza A/HongKong/1/1968 (H3N2) virus cap-snatching process. Sci Rep.

507 8. Sikora D, Rocheleau L, Brown EG, Pelchat M. 2017. Influenza A virus cap-snatches host RNAs
508 based on their abundance early after infection. Virology 509:167–177.

509 9. De Vlugt C, Sikora D, Pelchat M. 2018. Insight into Influenza: A Virus Cap-Snatching. Viruses
510 10:641.

511 10. Baillie JK, Arner E, Daub C, De Hoon M, Itoh M, Kawaji H, Lassmann T, Carninci P, Forrest ARR,
512 Hayashizaki Y, Faulkner GJ, Wells CA, Rehli M, Pavli P, Summers KM, Hume DA. 2017. Analysis of the
513 human monocyte-derived macrophage transcriptome and response to lipopolysaccharide provides
514 new insights into genetic aetiology of inflammatory bowel disease. PLoS Genet 13:1–36.

515 11. Forrest ARR, Kawaji H, Rehli M, Baillie JK, de Hoon MJL, Lassmann T, Itoh M, Summers KM,
516 Suzuki H, Daub CO, Kawai J, Heutink P, Hide W, Freeman TC, Lenhard B, Bajic VB, Taylor MS, Makeev
517 VJ, Sandelin A, Hume D a, Carninci P, Hayashizaki Y. 2014. A promoter-level mammalian expression
518 atlas. Nature 507:462–70.
519 12. Baillie JK, Bretherick A, Haley CS, Clohisey S, Gray A, Stahl EA, Tenesa A, Andersson R, Brown J
520 Ben, Faulkner GJ, Schaefer U, Daub C, Itoh M, Kondo N, Lassmann T, Kawai J, Consortium I,
521 Consortium F, Bajic VB, Heutink P, Rehli M, Kawaji H, Sandelin A, Suzuki H, Satsangi J, Wells CA,
522 Hacohen N, Freeman TC, Hayashizaki Y, Carninci P, Forrest ARR, Hume DA. 2016. Shared activity
523 patterns arising at genetic susceptibility loci reveal underlying genomic and cellular architecture of
524 human disease . bioRxiv 1–24.
525 13. Kanamori-katayama M, Itoh M, Kawaji H, Lassmann T, Katayama S, Kojima M, Bertin N, Kaiho
526 A, Ninomiya N, Daub CO, Carninci P, Forrest ARR, Hayashizaki Y. 2011. Unamplified cap analysis of
527 gene expression on a single-molecule sequencer. Cold Spring Harb Genome 1150–1159.
528 14. Hoeve MA, Nash AA, Jackson D, Randall RE, Dransfield I. 2012. Influenza virus A infection of
529 human monocyte and macrophage subpopulations reveals increased susceptibility associated with
530 cell differentiation. PLoS One 7.
531 15. Simpson-Holley M, Ellis D, Fisher D, Elton D, McCauley J, Digard P. 2002. A functional link
532 between the actin cytoskeleton and lipid rafts during budding of filamentous influenza virions.
533 Virology 301:212–225.
534 16. Roberts PC, Compans RW. 1998. Host cell dependence of viral morphology. Proc Natl Acad Sci
535 U S A 95:5746–5751.
536 17. Koppstein D, Ashour J, Bartel DP. 2015. Sequencing the cap-snatching repertoire of H1N1
537 influenza provides insight into the mechanism of viral transcription initiation. Nucleic Acids Res
538 43:5052–5064.

539 18. Gu W, Gallagher GR, Dai W, Liu P, Li R, Trombly MI, Gammon DB, Mello CC, Wang JP, Finberg
540 RW. 2015. Influenza A virus preferentially snatches noncoding RNA caps. Rna 21:2067–2075.
541 19. Bercovich-Kinori A, Tai J, Gelbart IA, Shitrit A, Ben-Moshe S, Drori Y, Itzkovitz S, Mandelboim
542 M, Stern-Ginossar N. 2016. A systematic view on influenza induced host shutoff. Elife 5:1–20.
543 20. McCauley JW, Mahy BW. 1983. Structure and function of the influenza virus genome.
544 Biochem J 211:281–94.
545 21. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X,
546 Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N,
547 Bagger FO, Jørgensen M, Andersen PR, Bertin N, Rackham O, Burroughs a. M, Baillie JK, Ishizu Y,
548 Shimizu Y, Furuhata E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H,
549 Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume D a., Jensen TH, Suzuki H, Hayashizaki Y,
550 Müller F, Consortium TF, Forrest ARR, Carninci P, Rehli M, Sandelin A. 2014. An atlas of active
551 enhancers across human cell types and tissues. Nature 507:455–461.
552 22. Wei J, Kishton RJ, Angel M, Conn CS, Dalla-Venezia N, Marcel V, Vincent A, Catez F, Ferré S,
553 Ayadi L, Marchand V, Dersh D, Gibbs JS, Ivanov IP, Fridlyand N, Couté Y, Diaz J-J, Qian S-B, Staudt LM,
554 Restifo NP, Yewdell JW. 2019. Ribosomal Proteins Regulate MHC Class I Peptide Generation for
555 Immunosurveillance. Mol Cell 1162–1173.
556 23. Rao P, Yuan W, Krug RM. 2003. Crucial role of CA cleavage sites in the cap-snatching
557 mechanism for initiating viral mRNA synthesis. EMBO J 22:1188–1198.
558 24. Geerts-Dimitriadou C, Goldbach R, Kormelink R. 2011. Preferential use of RNA leader
559 sequences during influenza A transcription initiation in vivo. Virology 409:27–32.
560 25. Beaton AR, Krug RM. 1981. Selected host cell capped RNA fragments prime influenza viral
561 RNA transcription in vivo. Nucleic Acids Res 9:4423–4436.
562 26. Budach S, Marsico A. 2018. Pysster: Classification of biological sequences by learning
563 sequence and structure motifs with convolutional neural networks. Bioinformatics 34:3035–3037.

564 27. Lamb RA, Lai CJ, Choppin PW. 1981. Sequences of mRNAs derived from genome RNA segment
565 7 of influenza virus: colinear and interrupted mRNAs code for overlapping proteins. Proc Natl Acad Sci
566 U S A 78:4170–4.
567 28. Bauer DLV, Tellier M, Martínez-Alonso M, Nojima T, Proudfoot NJ, Murphy S, Fodor E. 2018.
568 Influenza Virus Mounts a Two-Pronged Attack on Host RNA Polymerase II Transcription. Cell Rep
569 23:2119-2129.e3.
570 29. Vilborg A, Steitz JA. 2017. Readthrough transcription: How are DoGs made and what do they
571 do? RNA Biol 14:632–636.
572 30. Freeman TC, Goldovsky L, Brosch M, Van Dongen S, Mazière P, Grocock RJ, Freilich S,
573 Thornton J, Enright AJ. 2007. Construction, visualisation, and clustering of transcription networks
574 from microarray expression data. PLoS Comput Biol 3:2032–2042.
575 31. Chang JT, Nevins JR. 2006. GATHER: A systems approach to interpreting genomic signatures.
576 Bioinformatics 22:2926–2933.
577 32. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G V., Clark NR, Ma'ayan A. 2013. Enrichr:
578 Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14.
579 33. Kuleshov M V., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL,
580 Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. 2016. Enrichr:
581 a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res 44:W90–
582 W97.
583 34. He Y, Xu K, Keiner B, Zhou J, Czudai V, Li T, Chen Z, Liu J, Klenk H-D, Shu YL, Sun B. 2010.
584 Influenza A Virus Replication Induces Cell Cycle Arrest in G0/G1 Phase. J Virol 84:12832–12840.
585 35. Klinkhammer J, Schnepf D, Ye L, Schwaderlapp M, Gad HH, Hartmann R, Garcin D, Mahlakõiv
586 T, Staeheli P. 2018. IFN-λ prevents influenza virus spread from the upper airways to the lungs and
587 limits virus transmission. Elife 7:e33354.

588 36. Van Riel D, Munster VJ, De Wit E, Rimmelzwaan GF, Fouchier RAM, Osterhaus ADME, Kuiken
589 T. 2007. Human and avian influenza viruses target different cells in the lower respiratory tract of
590 humans and other mammals. Am J Pathol 171:1215–1223.
591 37. Ramilo O, Allman W, Chung W, Mejias A, Ardura M, Glaser C, Wittkowski KM, Piqueras B,
592 Banchereau J, Palucka AK, Chaussabel D. 2007. Gene expression patterns in blood leukocytes
593 discriminate patients with acute infections. Blood 109:2066–2077.
594 38. Irvine KM, Andrews MR, Fernandez-Rojo M a, Schroder K, Burns CJ, Su S, Wilks AF, Parton RG,
595 Hume D a, Sweet MJ. 2009. Colony-stimulating factor-1 (CSF-1) delivers a proatherogenic signal to
596 human macrophages. J Leukoc Biol 85:278–288.
597 39. Takahashi H, Lassmann T, Murata M, Carninci P. 2012. 5' end-centered expression profiling
598 using cap-analysis gene expression and next-generation sequencing. TL - 7. Nat Protoc 7 VN-re:542–
599 561.
600 40. Sergushichev A. 2016. An algorithm for fast preranked gene set enrichment analysis using
601 cumulative statistic calculation. bioRxiv 60012.
602 41. Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: A Bioconductor package for differential
603 expression analysis of digital gene expression data. Bioinformatics 26:139–140.
604
605

606 **Figure Legends**

607 Figure 1: Characterisation of human monocyte derived macrophages productively infected
608 with IAV.
609 (A) Experimental outline. Blood was taken from 4 human donors, with appropriate ethical approval.
610 CD14+ monocytes were extracted using magnetic beads and cultured in CSF1 for 8-10 days. MDMs
611 were infected with A/Udorn/72 (H3N2) at a multiplicity of infection of 5. At 4 time points (0, 2, 7, and
612 24 hours after medium change) the cells were collected and RNA isolated. (B) Human MDMs were
613 stained using antibodies specific for viral nucleoprotein to confirm infection at 0, 2, 7, and 24 hours
614 post infection. Scale bars 10µm. (C) Viral titre was measured by plaque assay at 0, 2, 7, and 24 hours
615 post infection (n = 3 independent experiments) and shown in pfu/ml supernatant. (D) Cell viability
616 was measured using Cell Titre Glo™ at 0, 2, 7, and 24 hours post infection (n = 3 independent
617 experiments). (E) Schematic showing the structure of the capped 5' end of IAV mRNAs. (F) Length of
618 leader sequences across segments. Segments are coloured as shown in the legend. (G) Frequency, as
619 percentage, of IAV promoter-containing CAGE tags in each IAV infected sample.
620
621 Figure 2: IAV segment transcriptional dynamics during infection of MDM
622 (A) The relative amount, compared to the total amount of viral mRNA, of mRNA from each viral
623 segment was calculated for individual donors at each of the four timepoints. Height of the bar
624 represents the mean frequency between donors. Error bars show standard deviation. (B) The
625 positions of potential splice variant sequences aligned to the Udorn genome are shown as adjusted
626 abundance (AA). The known mRNA3 splice variant in segment 7 is shown (blue arrow). Time points
627 and donors have been collated to increase signal
628
629 Figure 3. Pathways enrichment in snatched and unsnatched tenmer sequences.

(A, B) RNA type was assigned to 10mers based on transcript identity. Only 10mers with transcript identity were included. The significance of RNA type snatching was compared using ANOVA. RNA types were plotted against –log10(FDR) for 10mers of that type. The box denotes the interquartile range. The within the box represents the average and the whiskers represent standard deviation. The individual data-point for each 10mer is also plotted. The number of 10mers attributed to each RNA type is given as n above the box. (C) The 10 most under-represented pathways (negative enrichment score, blue) and single significantly over-represented pathway (positive enrichment score, orange) in the Reactome 2016 database are shown. N represents the number of genes associated with that pathway detectable in the dataset. p-values shown are Benjamini-Hochberg FDR-adjusted p-values. (D) Volcano plot showing the significance as -log10(FDR) and odds ratio of snatched versus unsnatched 10mers with members of the Reactome pathway 'RNA Polymerase transcribes snRNA genes' highlighted (snRNA, green diamonds, mRNA orange circles). (E) The same volcano plot as in (D) with members of the Reactome pathway 'Viral mRNA Translation' highlighted (blue circles).

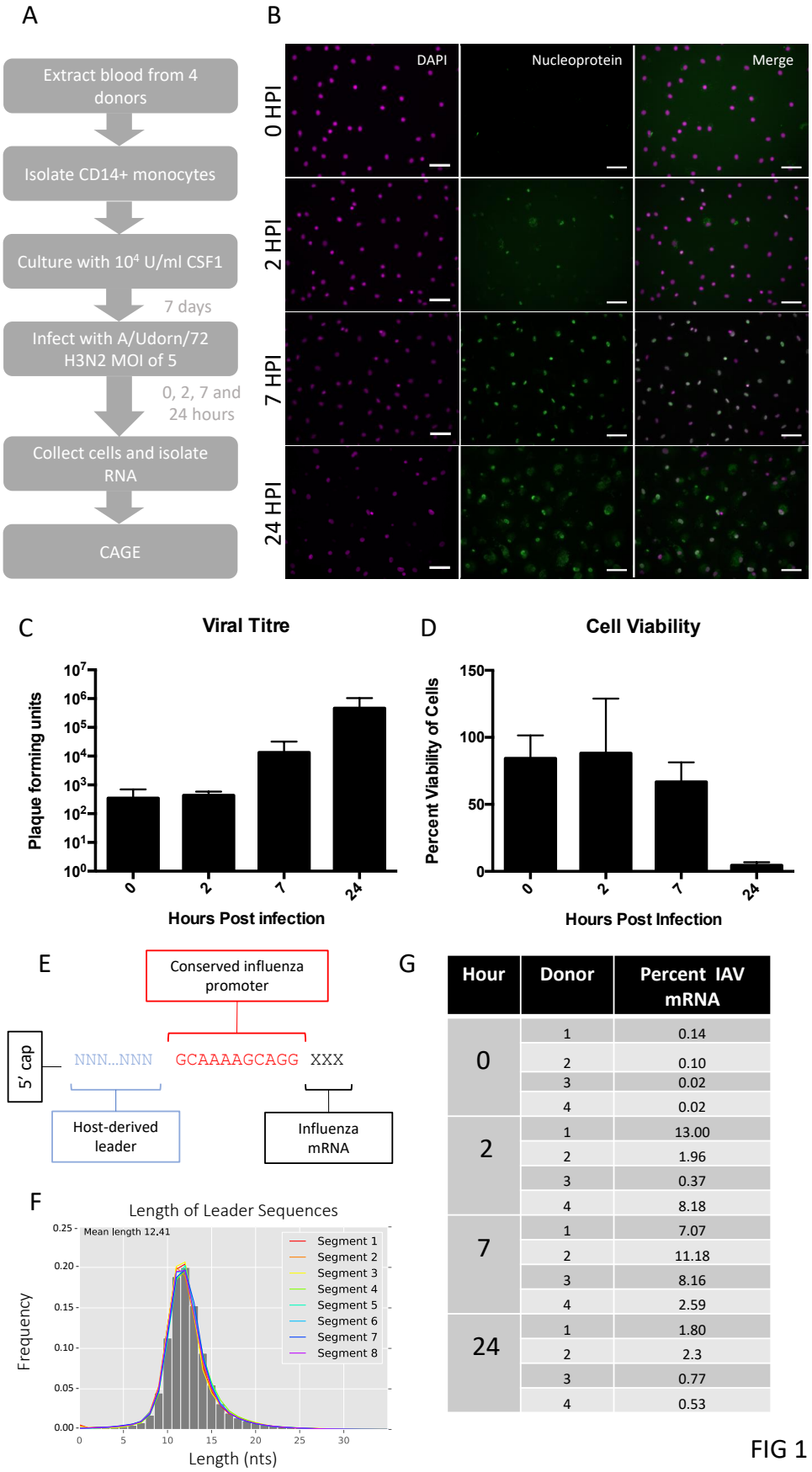
Figure 4: Nucleotide motifs associated with snatched and unsnatched tenmer sequences

The first ten nucleotides of each CAGE tag were extracted and the abundance of each sequence associated with IAV was compared to the background abundance by Fisher's Exact test (FDR < 0.05). Identification of motifs associated with snatched (A) and unsnatched (B) sequences. Violin plots show the maximum activation distributions for snatched (S) and unsnatched (U) sequence categories in arbitrary units. The four-nucleotide long motifs associated with each category are visualised as position weight matrices. The positional enrichment of the four-nucleotide motifs across the 10mer sequences is shown. The number of sequences is given as n above each bar chart. (C) Position weight matrices for all unsnatched sequences at each of four timepoints.

655 Figure 5. Network analysis of the co-expressed genes during IAV infection in MDMs.

656 (A) Sample-to-sample network. A correlation coefficient of ≥ 0.7 was used to include all samples in
657 the network. Analysis was restricted to the dominant promoters (p1) and data were averaged across
658 the 4 donors. Blue – uninfected; pink – infected; darker colours show later time points. (B) Gene-to-
659 gene correlation profile of transcripts. Network analysis identified the sets of co-regulated
660 transcripts in the MDM response to IAV. Analysis was restricted to the dominant promoters (p1) and
661 data were averaged across the 4 donors. Lines represent connections at Pearson correlation
662 coefficient ≥ 0.94 and spheres represent genes (promoters). The clustering procedure used a
663 relatively coarse Markov clustering algorithm of 1.7 to avoid excessive cluster fragmentation. The
664 four largest clusters, along with their average expression profiles, are shown. Y axis in the expression
665 profiles shows the expression level in tags per million (TPM). (C-F) Abundance of transcripts for IL1B
666 (C) EGR1 (D), TNFα (E) and CXCL2 (F) at the indicated time points. y-axis shows expression in tags per
667 million (TPM). (G) Differential gene expression analysis comparing expression of transcripts in LPS-
668 treated and IAV- treated monocyte derived MDMs. Transcripts with a relative log fold change
669 (log2FC) >=2 and a -log10(FDR) >=3 are shown in red (higher in LPS treated) and blue (higher in IAV
670 infection). Genes with greatest difference in expression are labelled. Genes referenced in the text
671 are shown in black. (H) Comparison of the temporal response of genes between IAV- and LPS-
672 treated MDMs. Expression (TPM) of selected genes in LPS- treated (red) and IAV- infected (blue)
673 human MDMs at 0, 2, 7, and 24 hours post treatment is shown in tags per million (TPM). Solid lines
674 show the mean expression of all donors (n = 3 for LPS, n = 4 for IAV). Filled-in area shows standard
675 deviation between donors.
676
677 Figure 6: Comparative analysis of the response of MDMs to treatment with IAV and with LPS.
678 (A-B) Comparison of the temporal response of transcripts between IAV- and LPS- treated MDMs.
679 Relative expression of selected genes in LPS- treated (red) and IAV- infected (blue) human MDMs at

680 0, 2, 7, and 24 hours post treatment is shown in tags per million (TPM). Solid lines show the mean
681 expression of all donors, filled-in area shows standard deviation between donors (n = 3 for LPS, n = 4
682 for IAV).



G

Hour	Donor	Percent IAV mRNA
0	1	0.14
	2	0.10
	3	0.02
	4	0.02
2	1	13.00
	2	1.96
	3	0.37
	4	8.18
7	1	7.07
	2	11.18
	3	8.16
	4	2.59
24	1	1.80
	2	2.3
	3	0.77
	4	0.53

FIG 1

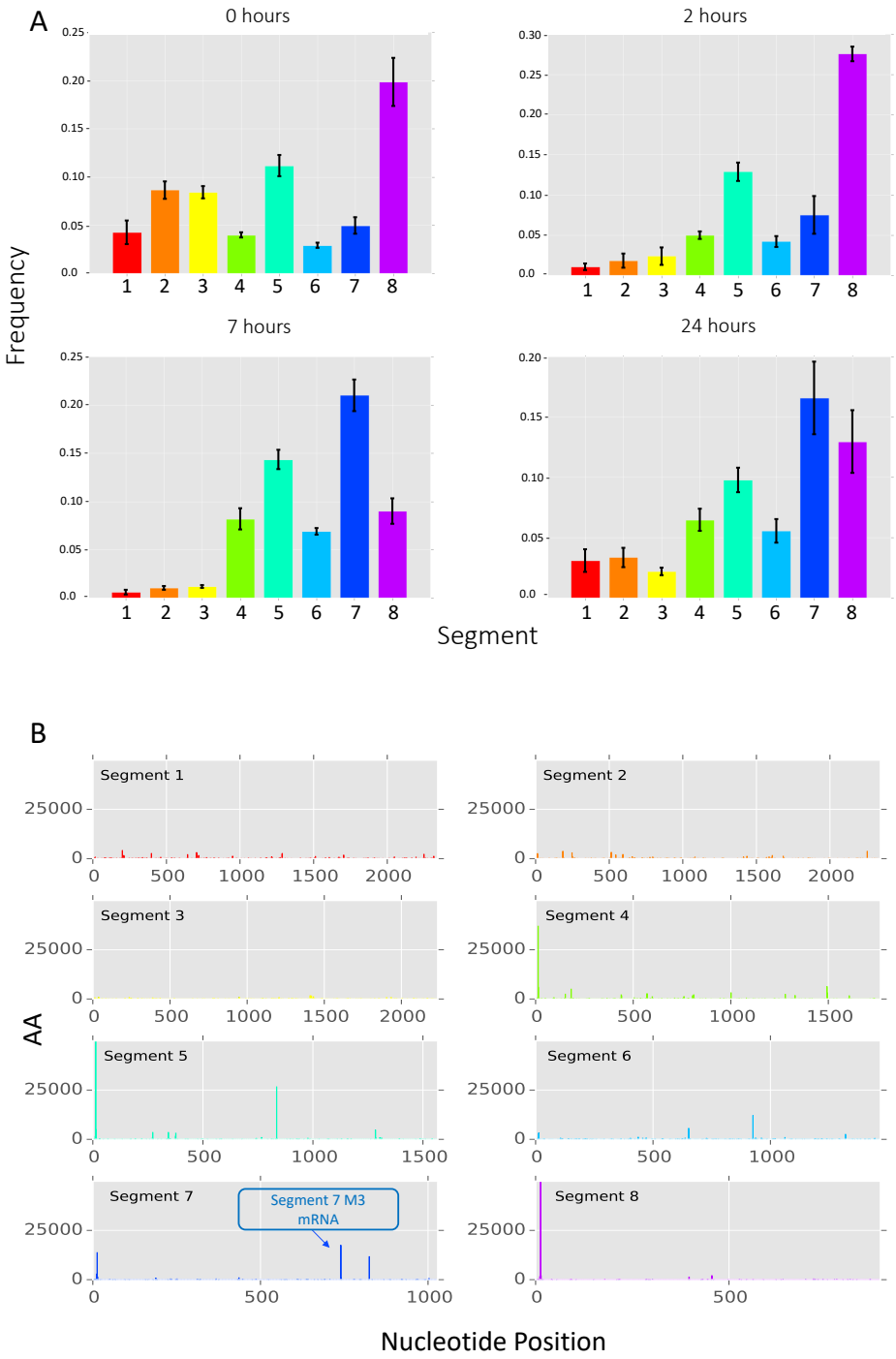


FIG 2

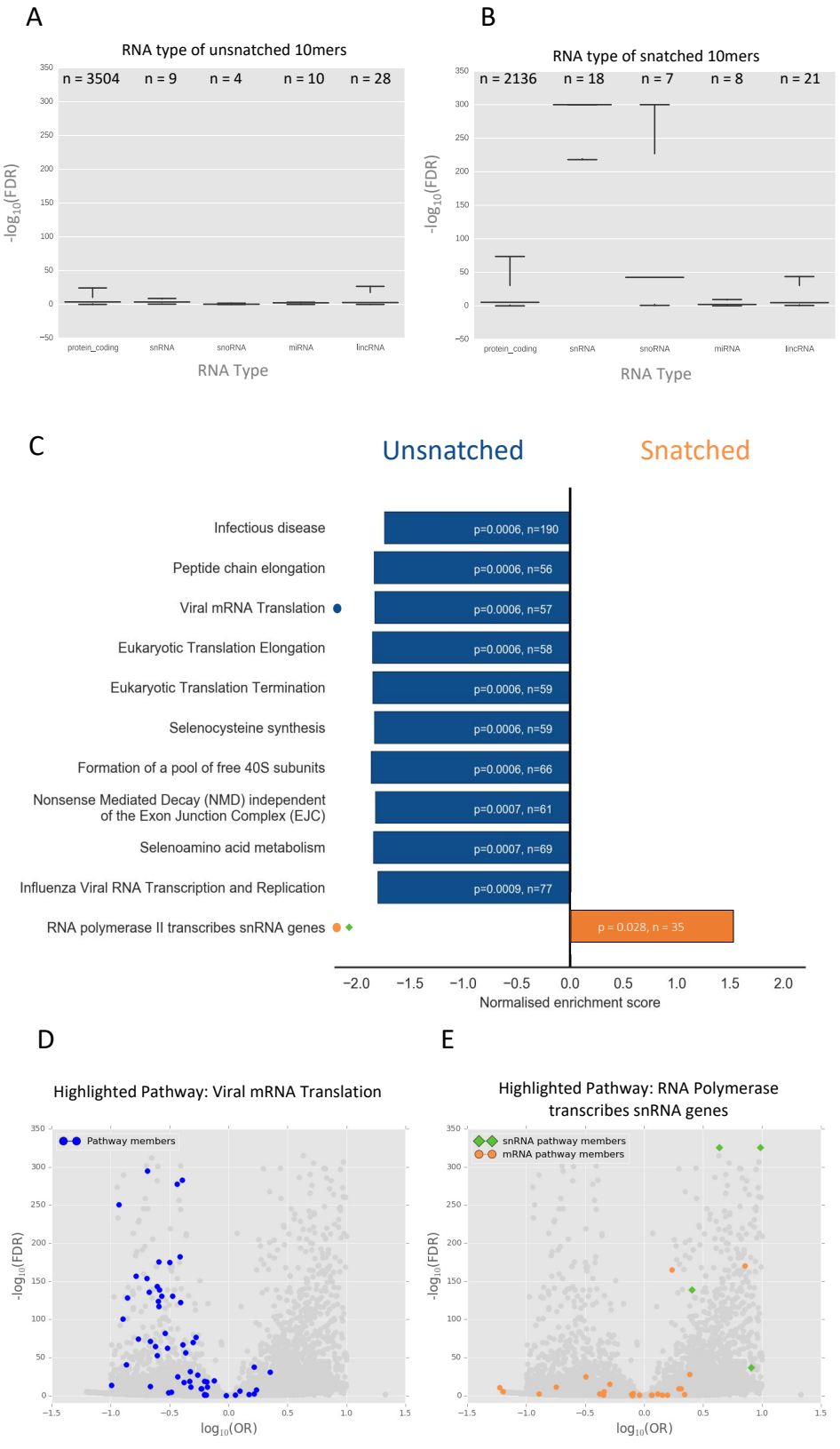


FIG 3

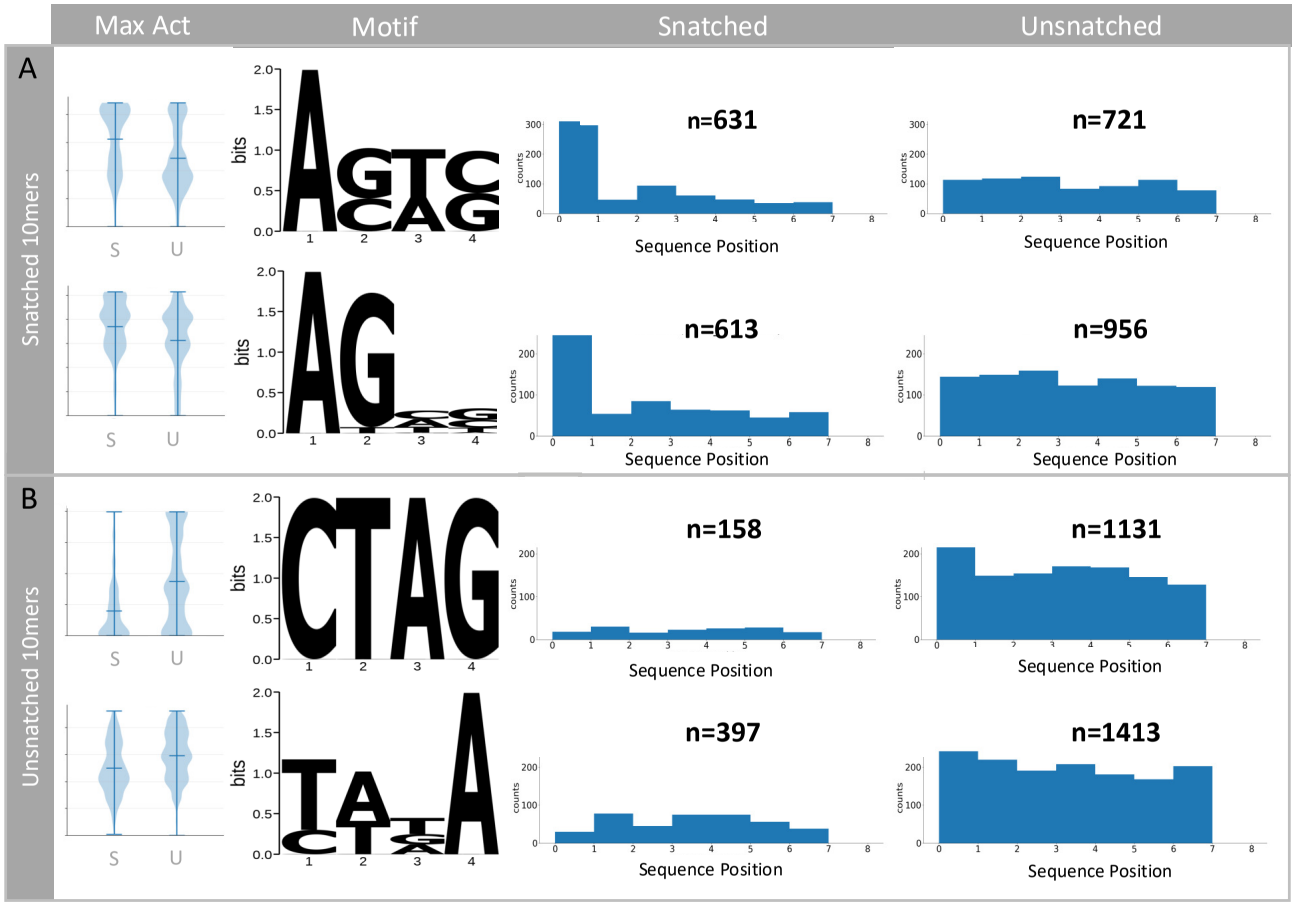


FIG 4

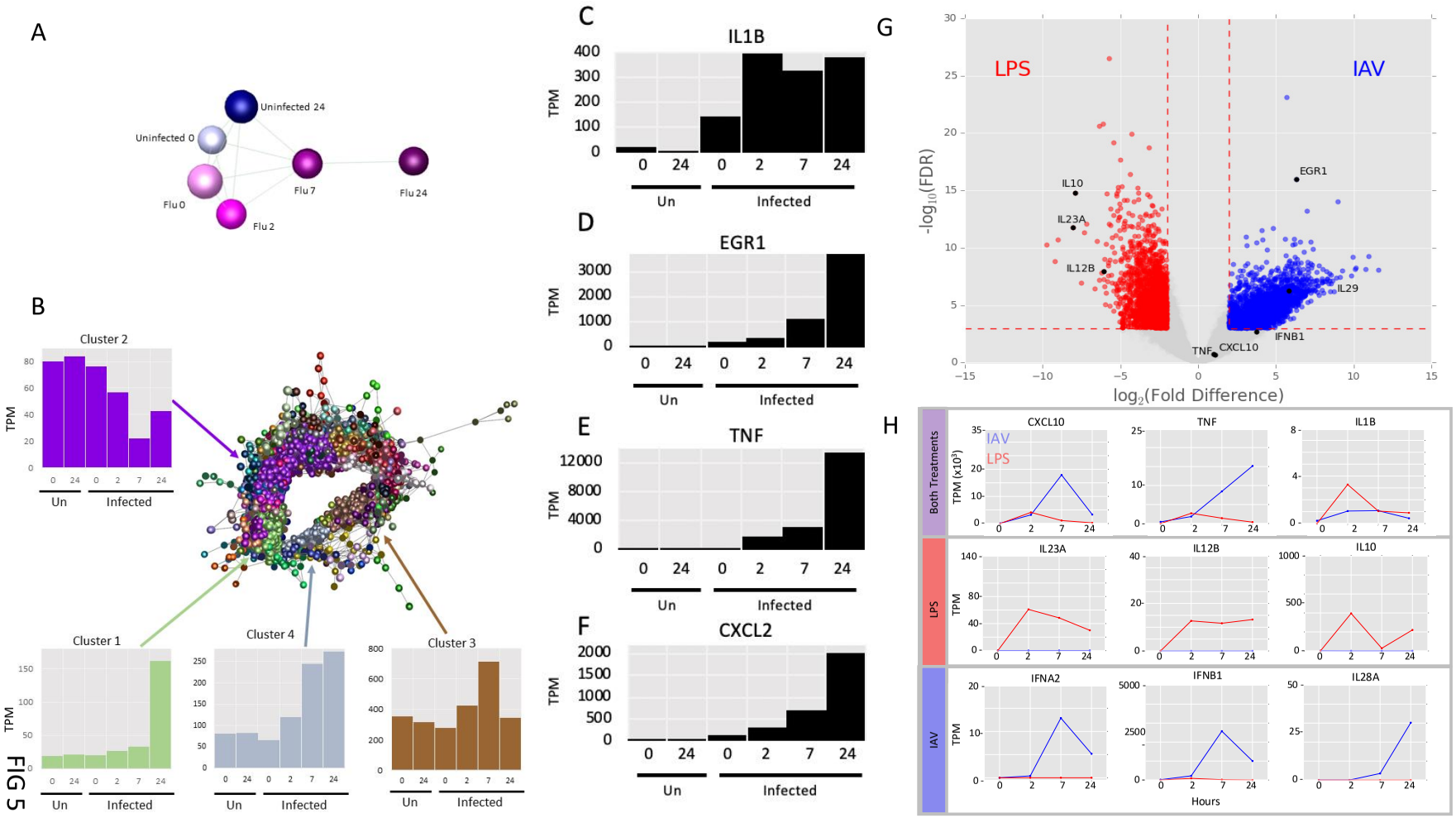


FIG 5

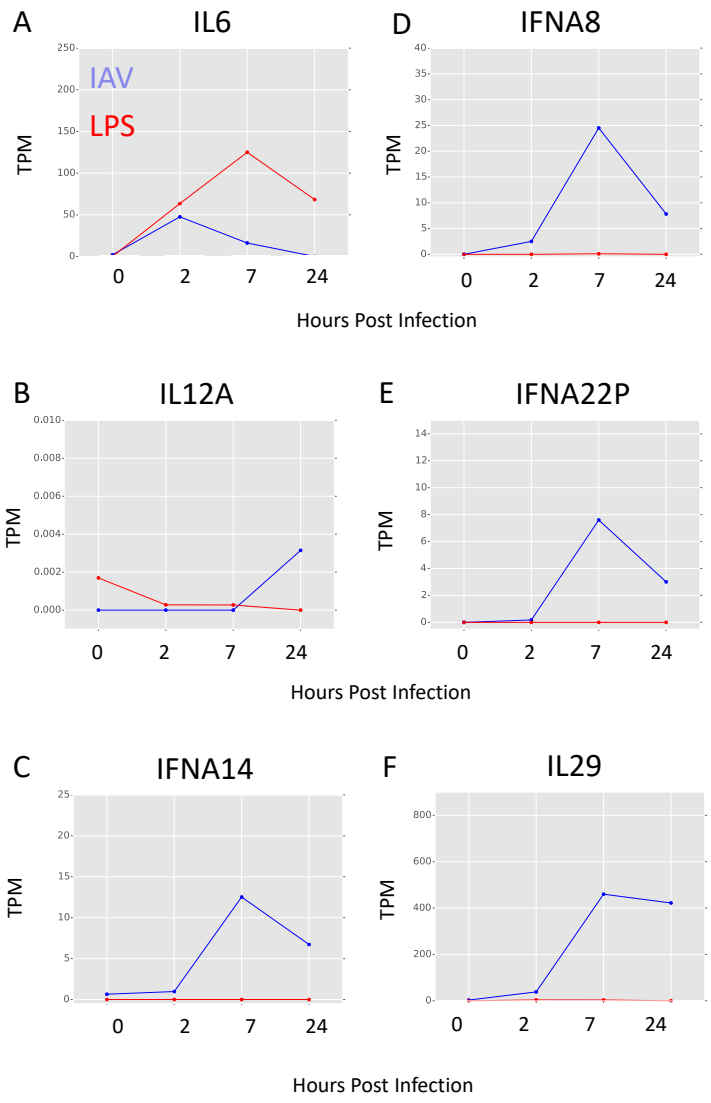


FIG 6