1 # Progression of whole blood transcriptional signatures from
2 interferon-induced to neutrophil-associated patterns in patients
3 with severe influenza

4 Jake Dunning[1, 4], Simon Blankley[3], Long T Hoang[1], Mike Cox[2], Christine M Graham[3], Philip L James[2],
5 Chloe I Bloom[3], Damien Chaussabel[5], Jacques Banchereau[6], Stephen J. Brett[1], the MOSAIC
6 Investigators[6], Miriam F Moffatt[2], Anne O'Garra[1, 3], Peter JM Openshaw[1]

7 Corresponding authors: Anne.OGarra@crick.ac.uk or p.openshaw@imperial.ac.uk

8 **Affiliations**

9 [1]Respiratory Sciences and [2]Genomic Medicine, National Heart and Lung Institute, Imperial College
10 London, W2 1PG, UK

11 [3]Laboratory of Immunoregulation & Infection, The Francis Crick Institute, 1 Midland Road, London
12 NW1 1AT, UK

13 [4]Current address: National Infection Service, Public Health England, 61 Colindale Avenue, London
14 NW9 5EQ, UK

15 [5]Sidra Medical and Research Centre, PO Box 26999, Doha, Qatar

16 [6]The Jackson Laboratory for Genomic Medicine, 263 Farmington Ave; Farmington, CT 06030,
17 Connecticut, USA

18 [6]**MOSAIC investigators:**

19 **Chelsea and Westminster NHS Foundation Trust:** B.G. Gazzard**. Francis Crick Institute, 1 Midland**
20 **Road, London, World Influenza Centre:** A. Hay, J. McCauley. **Francis Crick Institute, 1 Midland Road,**
21 **London: Laboratory of Immunoregulation & Infection:** A. O'Garra, C.M. Graham, C.I. Bloom, S.
22 Blankley. **Imperial College London, UK:** P. Aylin, D. Ashby, W.S. Barclay, W.O. Cookson, L.N.
23 Drumright, R.A. Elderfield, L. Garcia-Alvarez, M.J. Griffiths, M.S. Habibi, T.T. Hansel, J.A. Herberg, A.H.
24 Holmes, S.L. Johnston, O.M. Kon, M. Levin, S. Nadel, P.J. Openshaw, R.S Thwaites, J.O. Warner.
25 **Liverpool School of Tropical Medicine, UK:** S.J. Aston, S.B. Gordon. **Manchester Collaborative**
26 **Centre for Inflammation Research (MCCIR):** T. Hussell. **Public Health England (formerly Health**
27 **Protection Agency), UK:** C. Thompson, M.C. Zambon. **The Roslin Institute, University of Edinburgh:**
28 D.A. Hume. **University College London, UK**: A. Hayward. **UCL Institute of Child Health:** R.L. Smyth;
29 **University of Edinburgh, UK:** J.K. Baillie, P. Simmonds **University of Liverpool, UK:** P.S. McNamara;
30 M.G. Semple; **University of Nottingham, UK**: J.S. Nguyen-Van-Tam; **University of Oxford, UK:** L-P.
31 Ho, A. J. McMichael **Wellcome Trust Sanger Institute, UK:** P. Kellam **West of Scotland Specialist**
32 **Virology Centre, Glasgow, UK:** W.E. Adamson, W.F. Carman.

33 **Abstract**

34 Transcriptional profiles are increasingly used to investigate the severity, subtype and pathogenesis
35 of disease. We now describe whole blood RNA signatures and local and systemic immune mediator
36 levels in a large cohort of adults hospitalised with influenza from which extensive clinical and
37 investigational data was obtained. Signatures reflecting interferon-related antiviral pathways were
38 common up to day 4 of symptoms in cases not requiring mechanical ventilatory support; in those
39 needing mechanical ventilation, an inflammatory, activated neutrophil and cell stress/death
40 ('bacterial') pattern was seen, even early after disease onset. Identifiable bacterial co-infection was
41 not necessary for this 'bacterial' signature but could enhance its development while attenuating the
42 early 'viral' signature. Our findings emphasise the importance of timing and severity in the
43 interpretation of transcriptomic profiles and soluble mediator levels, and identify specific patterns of
44 immune activation that may enable the development of novel diagnostics and therapeutics.
45

46 Influenza viruses present a continuous threat to global health, mutating and spreading within and
47 between species. It is estimated that one billion cases of human influenza occur worldwide each
48 year, causing 3-5 million cases of severe illness and 300,000 to 500,000 deaths[1]. While most deaths
49 and illnesses are attributable to seasonal influenza, pandemics caused by novel viruses regularly
50 pose an unpredictable challenge to public health.

51 Infection with pandemic 2009 H1N1 influenza A virus (pH1N1) resulted in generally mild disease[2],
52 but still caused an estimated 250,000 – 500,000 additional deaths during the first 12 months of
53 global circulation[3]. Whereas seasonal influenza commonly causes severe disease in the old and
54 infirm, serious pH1N1 disease mostly occurred in infants and younger adults, presenting as viral
55 pneumonia and sometimes complicated by multi-organ failure[4, 5]. It has been suggested that severe
56 influenza may in part result from an over-exuberant host reaction to infection (sometimes termed
57 "cytokine storm"), but is also driven by a high viral load in affected persons[6, 7, 8]. Although analysis of
58 transcriptional signatures and mediator levels has helped to clarify the pathogenesis of severe
59 influenza, the relationship between severity, timing and complications of infection remains unclear.
60 Previous studies of gene expression patterns in influenza have typically involved small numbers of
61 individuals, healthy subjects undergoing experimental challenge or patients suffering from mild
62 disease[9, 10, 11, 12, 13, 14, 15]. Transcriptomic analysis has also been used to study a variety of acute and
63 chronic infections, including bacterial sepsis, dengue virus infection and tuberculosis[16] and to
64 examine differences and similarities between infectious and non-infectious inflammatory disorders,
65 such as systemic lupus erythematosus[17].

66 To further elucidate influenza pathogenesis, the Mechanisms of Severe Acute Influenza Consortium
67 (MOSAIC) recruited 255 hospitalised patients with suspected influenza in England over two
68 consecutive seasons (2009/10 and 2010/11). By analysing biological samples taken at multiple time-
69 points and correlating this with extensive clinical data, MOSAIC aimed to define the contributions
70 made by influenza virus sequence variation, co-pathogens (non-influenza viruses and bacteria) and
71 host factors (genetic and transcriptional differences, soluble mediator responses and cellular
72 immune responses) to disease pathogenesis. Sample analysis resulted in a cumulative total of 2.1 x
73 10^7 data items on this population, a dataset that we now describe in outline and provide as a
74 resource. To date, MOSAIC has reported enrichment for a host genetic variant, the interferon-
75 inducible transmembrane protein 3 (*IFITM3)* allele SNP rs12252-C in some hospitalised patients with
76 influenza[18], and that viral sequence changes that accumulate over time may contribute to the
77 variation in disease severity[19, 20, 21, 22]. The exceptional size and depth of the MOSAIC study provides a
78 unique database to allow these complex issues to be resolved.

79 In the present study, we used whole blood transcriptional RNA analysis and data from soluble
80 immune mediator (cytokines and chemokine) measurements from mucosal fluids and serum to
81 define associations between individual responses to infection and clinical and laboratory findings in
82 hospitalised adults with influenza. The richness of clinical and ancillary data in the MOSAIC study
83 allowed us to examine the extent to which specific patterns of inflammation arise from progression
84 of antiviral and inflammatory responses induced simply by viral infection, or whether they reflect a
85 response to pathogenic or commensal bacteria. We found that transcriptomic signatures and
86 mediator levels are strongly associated with both severity and duration of illness, indicating a phased
87 and graded activation of interferon-related and inflammatory genes. Clinically-evident bacterial co-
88 infection influenced the pattern of gene expression, but effects were superimposed on patterns
89 governed by the duration and severity of influenza.

90

91 **Results**

92 **Influenza is characterized by an overabundance of interferon and inflammation-related transcripts**
93 Principal Component Analysis (PCA) of the 18,974 most abundant transcripts from whole blood RNA
94 at enrolment (T1) showed that influenza patients from the 2010/11 season (n=109) formed a distinct
95 cluster that was clearly differentiated from matched healthy controls (n=130). PCA indicated no
96 discernible difference between patients with influenza caused by diverse influenza A or influenza B
97 viruses (Fig. 1a).

98 Microarray profiles of whole blood RNA from the first and second acute illness sampling time-points
99 (T1 vs. T2) were indistinguishable except in that they reflected different times after symptom onset
100 (see below). Results from samples from the final time-point (T3, at least 4 weeks after T1) were
101 similar to those from healthy controls in cases that were clinically resolved, but were highly
102 abnormal in patients who remained unwell and in hospital (data not shown). Since T3 samples were
103 highly diverse, they are not described further in the present report.

104 Modular analysis[23] of the 2010/11 samples revealed a marked overabundance of transcripts within
105 the interferon-inducible (M3.1) and neutrophil (M2.2) genes relative to healthy controls (Fig. 1b).
106 Transcripts representing plasma cells (M1.1), a subset of myeloid lineage genes (M2.6) and two
107 inflammation modules (M3.2 and M3.3) were also increased. There was a decrease in expression of
108 T- (M2.8) and B-cell (M1.3) modules (Fig. 1b). The calculated index termed 'molecular distance to
109 health' (MDTH, derived from analysis of 4526 transcripts significantly detected from background
110 filtered for low expression[24]) was increased in most cases of influenza compared to healthy controls

111 (*P* < 0.0001; Fig. 1c), although this varied according to the disease stage and severity (see below). In
112 the 2010/11 cohort, a combination of expression-level and statistical filtering identified 1255
113 differentially-expressed transcripts compared to healthy controls. Supervised hierarchical clustering
114 of these transcripts (expressed as a heat-map) revealed transcripts that were relatively over- or
115 under-expressed in influenza patients, relative to healthy controls (Fig. 1d). When the 1255
116 transcript signature from the 2010/11 cohort was applied to the 2009/10 cohort (22 influenza
117 patients and 25 matched healthy controls), the 2009/10 profiles appeared to be the same as the
118 2010/11 profiles (Supplementary Fig. 1a), indicating that viral variation between the two seasons[22]
119 did not appreciably affect transcriptomic patterns.

120 Ingenuity Pathway Analysis (IPA) identified the top five canonical pathways associated with up-
121 regulated and down-regulated transcripts (*P* < 0.05, Fisher's Exact Test; Fig. 1d). Transcripts that
122 were up-regulated in influenza patients were associated with 'interferon signalling genes' (including
123 *IFITM1, IFI35*, *IFIT1, OAS1, IFIT3* and *IFI35*; Fig. 1e), 'activation of pattern recognition receptors by
124 bacteria and/or viruses', 'activation of IRF by cytosolic pattern recognition receptors', 'hepatic
125 fibrosis/hepatic stellate cell activation', and 'IL-6 signalling'. Transcripts that were down-regulated in
126 influenza patients were those associated with 'iCOS-ICOSL signalling in T helper cells', 'primary
127 immunodeficiency signalling', 'role of NFAT in regulation of the immune response', 'OX40 signalling
128 pathway', and 'T cell receptor signalling' (Fig. 1d).

129 **Heterogeneity in gene expression of the 25 most significant transcripts in hospitalised adults with**
130 **influenza**

131 The heat-map generated by hierarchical clustering of the top 25 most significant transcripts in the
132 2010/11 influenza patients (identified by mean fold-change over healthy controls) showed that the
133 influenza cases clustered into two major groups, albeit with two further sub-clusters (Fig. 1f). The
134 transcripts for the IFN-stimulated gene *IFI27* were over-expressed in almost all influenza patients,
135 but most also showed decreased *FCER1A* transcription. Independent analysis of the same 25
136 transcripts applied to data from the 2009/10 comparison cohort showed similar clustering effects
137 (Supplementary Fig. 1b). Patients with type I interferon-induced gene activation typically did not
138 express neutrophil-associated and bacterial response-associated transcripts, and those samples with
139 an overabundance of neutrophil-associated and bacterial response-associated transcripts (e.g.
140 DEFA4, ELANE, MMP-8) did not simultaneously show consistent overabundance of antiviral
141 response-associated transcripts (e.g. RSAD2, IFI6, IFI44L); these patterns were generally mutually
142 exclusive (Fig. 1f and Supplementary Fig. 1b).

143 **Modular transcriptional analysis and disease severity**

144 To investigate whether severity of illness might explain the sub-clustering of the transcriptional
145 responses seen in influenza patients and the heterogeneity of influenza transcriptional responses
146 suggested by data presented in Figure 1f, 2010/11 cases were grouped according to their severity of
147 illness at the first sampling time point (T1) using a three-point severity score based on treatment
148 decisions in relation to the presence and severity of respiratory failure (severity 1: no supplemental
149 oxygen requirement; 2: oxygen by mask; 3: mechanical ventilation). Relative to healthy controls, a
150 similar increase in mean MDTH was seen in patients with severity 1 and 2 illness, but a greater
151 increase was seen in severity 3 patients (4526 transcripts; Fig. 2a). Using modular analysis[23], we
152 noted an over-abundance of plasma cell (M1.1), neutrophil activation (M2.2), and myeloid lineage
153 (M2.6) transcripts in influenza-infected patients that was most marked in those with the greatest
154 severity. Severity 3 cases also showed the greatest abundance of transcripts in the inflammation
155 modules M3.2 and M3.3. By contrast, increased abundance of interferon-related transcripts (M3.1)
156 was most clearly evident in cases with severity 1 or 2, but was less evident in patients with severity 3
157 disease (Fig. 2b).

158 **Relationship between GO Term clusters and severity of illness**

159 Semi-supervised hierarchical clustering of 231 differentially expressed transcripts was performed
160 and results were expressed as a heat-map (with transcripts retained if there was greater than two-
161 fold change between severity 3 and severity 1 and 2; Fig. 3a). This heat-map suggested that severity
162 1 and 2 patients had similar over- and under-abundance patterns, and there was marked over-
163 abundance of transcripts associated with 'response to virus' identified by GO Terms analysis
164 (Supplementary Table 1). By contrast, severity 3 patients had less marked abundance of 'response to
165 virus' transcripts, but much more marked over-abundance of 'response to bacterium' transcripts
166 that are often associated with (but not exclusive to) bacterial infection (Supplementary Table 2), as
167 compared to patients with severity 1 or severity 2 illness. Additionally, patients with severity 3 illness
168 demonstrated greater under-abundance of transcripts associated with 'cellular defence response',
169 relative to patients with severity 1 or severity 2 illness.

170 The same 231 transcript list was tested by hierarchical clustering analysis of the 2009/10 comparison
171 cohort (Supplementary Fig. 1c). Influenza patients within the cluster were characterised again by an
172 over-abundance of transcripts associated with 'response to virus' by GO terms analysis, and these
173 patients had either severity 1 or severity 2 illness. By contrast, the cluster of patients with over-
174 abundance of transcripts associated with 'response to bacterium' by GO terms analysis, but much

175 less marked abundance of transcripts associated with 'response to virus', included all three patients
176 with severity 3 illness (Supplementary Fig. 1c).

177 A molecular score was calculated for each individual for the 51 'response to virus' and the 112
178 'response to bacterium' transcripts identified by GO Terms analysis across the 2010/11 influenza
179 patients of different severities at the first time of sampling (T1 only; n=109; Fig. 3b; Supplementary
180 Tables 1 and 2 respectively). Influenza patients with high 'viral' molecular scores (> 500) were
181 exclusively from the severity 1 and severity 2 groups. The majority of patients with high 'bacterial'
182 scores (> 500) had severity 3 illness, and these patients had low 'viral' scores, in keeping with the
183 modular analysis. A small minority of severity 1 and 2 patients had relatively low 'viral' molecular
184 scores with moderately high 'bacterial' molecular scores. Six patients with known bacteraemia were
185 included in the analysis but removal of cases with known bacterial co-infection did not influence the
186 appearance of the 'bacterial' molecular signature (data not shown). A similar effect was observed in
187 the influenza patients from 2009/10 (Supplementary Fig. 1d).
188

189 Reciprocal expression of activated and repressed biofunctions of identified genes was observed in
190 severity 3 patients, compared to severity 1 and 2 patients combined (Fig. 3c, 3d). Nine genes
191 associated with neutrophil activation were shown to be upregulated, including the genes *MPO*,
192 *DEFA1*, and *ELANE*. Additionally, three genes associated with activation of leukocyte influx were
193 upregulated in severity 3 patients: *MPO*, *MMP9* and *LCN2* (Fig. 3c). The associated repressed
194 biofunctions in severity 3 patients were 'activation of cytotoxic T cells', 'adhesion of immune cells'
195 and 'quantity of leukocytes' (Fig. 3d).
196

197 **Effect of illness duration on molecular signatures**

198 Patients with influenza symptoms of up to 4 days' duration at the time of sampling typically had
199 elevated 'viral' molecular scores, but not if they required mechanical ventilation (Severity 3); in such
200 cases, the 'viral' score was low, even early in the disease (Fig. 4a). 'Bacterial' molecular scores were
201 low in patients with severity 1 and 2 illness regardless of the time of sampling, whereas patients with
202 severity 3 illness showed higher 'bacterial' molecular scores than patients with less severe disease,
203 irrespective of illness duration (Fig. 4b).

204 Focussing on influenza patients (2010/11 cohort) with repeat samples (T1 and T2; n=59) separated
205 by 2-5 days, the 'viral' molecular score usually (but not always) decreased between T1 and T2
206 (Supplementary Fig. 2a). In those cases where T2 samples were obtained 48 hours after T1 (n=41,
207 2010/11 cohort), the reduction in 'viral' molecular score was statistically significant ($P = 0.0002$;

208 Supplementary Fig. 2b). Changes in 'bacterial' molecular scores between T1 and T2 were much
209 more heterogeneous than changes in 'viral' scores, irrespective of the actual timing of T2 sample
210 collection in relation to T1 (Supplementary Fig. 2c, 2d).
211

212 **Effect of confirmed infection and treatment on transcriptomic signatures**

213 We next sought correlations between influenza viral load (estimated from the samples of mucus
214 obtained from the nasopharyngeal space by suction catheter at T1 and T2, Supplementary Fig 2e)
215 and the transcriptional 'viral' and 'bacterial' scores at T1 and T2. No association was found between
216 viral load and the 'viral' transcriptomic score (Supplementary Fig. 2f, and additional data not
217 depicted).

218 To investigate whether bacterial infection is necessary for the observed activation of neutrophils and
219 'response to bacterium' by GO terms analysis, we analysed a subgroup of influenza-infected patients
220 that had been thoroughly investigated for bacterial infection by analysis of five sample types: T1
221 nasopharyngeal aspirate (NPA) for PCR detection of bacterial pathogens; T1 NPA for culture; T1
222 throat swab for culture; blood for culture; urine for pneumococcal antigen testing. To account for
223 incomplete bacteriological sampling in some patients, we excluded 36/109 (33%) patients for whom
224 two or more sample-types were not available for analysis. Of the remaining 73 patients, 39 (53%)
225 had potentially pathogenic bacteria detected in at least one sample type, and 34 (47%) patients
226 provided at least four out of five sample-types and did not have significant bacteria detected by
227 review of all the data available to the clinical panel. All patients thus categorised as 'bacteria not
228 detected' had provided NPA and throat swab samples; blood cultures were not obtained from 10
229 patients and urine for pneumococcal antigen testing was not obtained from 11 patients.

230 Comparing those cases of influenza in which significant bacterial infection was confirmed (Bac+) with
231 those in whom no bacterial infection was found despite extensive investigation (Bac-), the average
232 'viral' molecular score was lower in those with bacterial infection at all times up to day 12 after
233 illness onset (Fig. 4c). The average 'bacterial' score was greater in those with bacterial co-infection
234 between day 3 and 14 (Fig. 4d), but the time-trends in either case showed a similar pattern
235 regardless of the presence or absence of significant bacteria. Similar trends were observed when
236 stricter exclusion criteria were applied to the subgroup analysis, excluding patients from the
237 'bacteria not detected' group if they had not provided all five sample-types (data not shown); in this
238 case, statistical analysis could not be performed due to the low sample size (only 13 patients
239 provided all five sample types and did not have bacteria detected).

240 To investigate further whether bacterial infection drives the transcriptional response, serum (or
241 plasma) levels of procalcitonin (PCT) were used as a possible discriminant of invasive bacterial
242 infection[25, 26, 27]. Serum procalcitonin levels showed no relationship to 'viral' molecular score (Fig.
243 4e), and there was no correlation between 'viral' molecular scores at T1 and T2 and PCT levels
244 measured at respective time-points (data not shown). However, 'bacterial' molecular scores tended
245 to be raised in those cases with the highest PCT levels (Pearson r = 0.44, P <0.001) regardless of the
246 presence or absence of significant detectable bacteria (Fig. 4f).

247 Therefore, the 'viral' molecular score was greatest early in disease, being lost after about 5 days.
248 Even early in disease, cases needing mechanical ventilation had low 'viral' scores; this was especially
249 true in those with bacterial co-infection. On the other hand, expression of 'bacterial' response genes
250 is seen only in the most severe cases of influenza; bacterial infection enhances this signal, but the
251 'bacterial' score was increased in those with influenza regardless of measurable bacterial co-
252 infection, especially if the disease had lasted over one week. Taken together, our data show that
253 viral infection alone can induce the up-regulation of neutrophil-related genes, but induction of these
254 genes is enhanced in severe disease or by detectable bacterial co-infection.

255 We further examined the possible influence of treatment of bacterial infection on the observed
256 'viral' and 'bacterial' responses, by stratifying T1 and T2 'bacterial' and 'viral' scores in 2010/11
257 influenza patients according to receipt of antibiotics. Almost all patients recruited to the MOSAIC
258 study (92%; 234/255) were treated with antibiotics on clinical grounds at some time. Antibiotics
259 prior to first sampling had no demonstrable effect on transcriptomic patterns (Supplementary Fig.
260 3a). Comparing patients who were not given antibiotics (n=7) with those given sustained antibiotic
261 treatment following T1 (n=24) or throughout illness (including when T1 and T2; n=27), there was no
262 statistically significant effect of antibiotic administration on the 'bacterial' molecular scores
263 (Supplementary Fig. 3b). We next examined the levels of the 16S rRNA gene (bacterial load) in the
264 throat swab and NPA samples from cases that were classified as 'viral and bacterial co-infection' or
265 'viral infection, no bacterial infection'. The levels of the 16S rRNA gene were no different between
266 these groups on throat swabs, but the NPA bacterial load was greater in those cases with confirmed
267 bacterial co-infection (Supplementary Fig. 3c).

268 **Effect of illness duration, severity and bacterial co-infection on soluble mediators**

269 We next sought to validate the 'bacterial' and 'viral' transcriptional signatures observed in blood
270 with and protein-level data of mediators in the blood, NPA and anterior nasal fluid (SAM) at three
271 time points (35 mediators, data available online).

272 Serum levels of the pro-inflammatory cytokine IL-1β, which has limited anti-viral activity, showed no
273 trend against severity (Fig. 5a), but was significantly increased in the NPA in those with the most
274 severe influenza disease (Fig. 5b); nasal-absorption samples similarly showed an increase in IL-1β
275 only in those with greatest severity (Fig. 5c). Serum IL-6 levels were raised in those with influenza
276 and especially in those with most severe disease (Fig. 5d); in the NPA, IL-6 was undetectable in most
277 of the healthy controls but increased in most of the cases of influenza and especially in those with
278 severe disease (Fig. 5e). The levels of IL-6 in nasal-absorption samples paralleled those in serum and
279 NPA (Fig. 5f). CXCL8 serum levels tended to be higher in cases of flu than in controls, again increasing
280 with disease severity (Fig. 5g). CXCL8 levels in NPA samples were highly variable but increased
281 alongside influenza severity (Fig. 5h); however many NPA and nasabsorption CXCL8 measurements
282 were so high as to be unquantifiable without dilution, even in a proportion of healthy controls (Fig.
283 5i). By contrast to these predominantly inflammatory/bacterially-driven mediators, IFNα2a, which
284 was measurable only in a proportion of individuals, was raised in serum in milder (severity 1 or2)
285 rather than severe (severity 3) disease (Fig. 5j). IFNα2a levels in NPA and nasal-absorption samples
286 were similarly higher in some milder influenza cases, relative to severe cases, though such
287 differences were not significant. (Fig. 5k, 5l). The analysis of inflammatory mediators generally
288 supports the association in the transcriptomics data between severe disease at T1 and increased
289 inflammatory/'bacterial' markers, along with decreased 'viral' markers, in keeping with their known
290 role in infection and disease pathogenesis[7].

291 Serum levels of IL-17 were increased with severity at T1 (Supplementary Fig. 4a) and were elevated
292 in the BAL of eight patients in whom samples were available, relative to healthy controls
293 (Supplementary Fig. 4b). By inter-relating transcriptomic findings with mediator levels, we found a
294 significant positive correlation between serum IL-17 (a marker of bacterial inflammation, acting on
295 stromal cells to drive production of antimicrobial peptides and neutrophil chemoattractants), and
296 the bacterial MDTH (Supplementary Fig. 4c). A similar trend was seen between TNFα and MDTH
297 (Supplementary Fig. 4d). Since 'bacterial load' (as measured as NPA 16S rRNA copy number) was
298 raised in cases of significant bacterial infection (Supplementary Fig. 3c), we regressed this parameter
299 against viral or bacterial MDTH. High levels of viral MDTH only occurred in those with low bacterial
300 load in the NPA (Supplementary Fig. 4e). By contrast, high bacterial MDTH were seen only in those
301 with raised 16S bacterial load in the NPA (Supplementary Fig. 4f). These data supported a strong
302 'bacterial' transcriptomic signature to be associated with a neutrophilic/anti-bacterial inflammatory
303 response and higher respiratory tract bacterial load. This signature was in turn associated with
304 severe influenza, or later disease time points.

305 Considering the importance of time in the transcriptomics data, we next accounted for symptom
306 duration in this protein mediator data. CXCL10, IL-6 and CCL2 were elevated in serum from severe
307 cases of influenza (especially between days 5 and 10: Fig. 6a, 6b and not depicted). Proven bacterial
308 co-infection had no evident additional effect on CXCL10 (Fig. 6c), but serum IL-6 was more abundant
309 not only in severe cases (even early in disease, Fig 6b) but especially in cases of bacterial co-infection
310 (especially between days 5 and 10, Fig. 6d). In the NPA, most mediators (e.g. CXCL10, IL-6, CCL2 and
311 CXCL8) were markedly increased in severe disease, especially after day 4 (e.g. Figs. 6e and 6f, and
312 not depicted). NPA CXCL10 was again unaffected by confirmed bacterial disease (Fig. 6g) whereas
313 levels of IL-6 (and CCL2 and CXCL8, data not shown) were particularly increased in patients with
314 bacterial co-infection (Fig. 6h). In the anterior nasal fluid (SAM), mediator levels declined slowly with
315 time even in less severe disease; CXCL10 was depressed by bacterial co-infection but IL-6, CCL2 and
316 CXCL8 levels at this site were unaffected by severity or bacterial status (not depicted; data online).

317 **Discussion**

318 The MOSAIC study is exceptional in presenting data from a large number of exceptionally well-
319 characterised hospitalised patients with influenza, studied prospectively. We found that whole-
320 blood RNA expression profiles of patients hospitalised with influenza evolve over time and that the
321 speed and scale of this evolution reflects severity. Patients with mild (or early) disease typically
322 showed a pattern dominated by interferon-inducible genes and type 1 interferon, but this 'viral
323 response' signature was replaced during severe (or late) disease by a pattern reflecting inflammation
324 and neutrophil activation, more typically associated with the GO term 'response to bacteria'. The
325 'viral' response was rarely seen in patients beyond day 4 of symptoms; the inflammation/neutrophil
326 activation signal dominated during the second week, but resolved during clinical recovery. In
327 patients providing multiple samples during the first 3 days of illness, the 'viral' molecular score
328 decreased rapidly between the first and subsequent sample, whereas the 'bacterial' score showed
329 less consistent change.

330 The decline in viral response and the dominance of inflammation and neutrophil activation seen in
331 severe disease was enhanced by proven bacterial co-infection, but did not depend on it. However,
332 the bacterial load in the nasopharynx (quantified by 16S copy number) needed to be low for a 'viral'
333 signature to be high, and high for an inflammatory/cell activation response to be evident. This
334 finding suggests an interaction between viral and bacterial sensing and response mechanisms. In
335 terms of soluble protein mediators, serum and nasopharyngeal levels were generally highest in cases
336 of severe disease, even during the early stages. Levels of inflammatory mediators (e.g. IL-1β and IL-6)
337 were augmented in those with clinically significant bacterial co-infections. Interferon α levels tended
338 to be low or undetectable in most compartments in those with very severe influenza, but interferon-
339 related mediators (e.g. CXCL10 in serum) were generally most abundant in severe cases.

340 We were unable to show any qualitative difference in the response to different infecting viruses (e.g.
341 influenza virus A vs. B). From previous studies, it seems that viral load does, in part, drive disease
342 severity in patients infected with highly pathogenic strains of influenza[6, 7, 8], but we were unable to
343 show a correlation between the viral load in the nasopharyngeal aspirate and the degree of
344 abnormality of whole-blood RNA signatures. Measurement of viral load in influenza is technically
345 difficult and depends on variations in sampling of nasopharyngeal mucus. For ethical and practical
346 reasons we were unable to obtain routine samples from the lower respiratory tract; our findings
347 should not be cited as evidence that viral load is irrelevant to disease severity but only that we were
348 unable to demonstrate a relationship using the methods available to us.

349 The linked nature of MOSAIC cohort data allowed us to adopt an integrated approach to data
350 exploration. Cases with early or mild influenza showed a transcriptional signature typical of viral
351 infection (up-regulation of Type I interferon related genes *IFIT1, IFIT3, OAS1, IFITM1 and IFI35,* and
352 type II interferon-stimulated genes *IFITM1* and *IFI35*) in the JAK-STAT activation pathways. This was
353 independently evident in samples collected in 2009/10, but in the larger 2010/11 cohort there was a
354 distinct subgroup of cases expressing both IFNα-inducible protein 27 (*IFI27*) transcripts and antiviral
355 response genes (e.g. *RSAD2, IFI6* and *IFI44L*). *IFI27* encodes an IFN-regulated mitochondrial protein
356 that has antiviral effects *via* sensitization of cells to pro-apoptotic stimuli[28]. Although *IFI27*
357 expression has been proposed as a potential biomarker for influenza infection[11], overexpression of
358 this transcript is not unique to influenza; it is also strongly upregulated in human airway epithelial
359 cells and peripheral blood after infection with respiratory syncytial virus[9]. In our study, its translation
360 was not associated with the stage, severity, or complications of influenza but persisted in the
361 absence of up-regulation of other ISGs.

362 Following the initial interferon-dominated phase, patients with severe or prolonged symptoms
363 activate a broad range of genes in addition to those classically associated with viral responses
364 (reviewed elsewhere[16]). After the first four days of illness, these genes include those that encode
365 inflammatory cytokines and chemokines, classical 'antibacterial' effector molecules (especially from
366 neutrophils), and regulators of apoptosis and anaerobic metabolism[29]. We find that this occurs
367 irrespective of identifiable bacterial co-infection, mirroring studies in macaques in which
368 administration of recombinant IFNα2a initially up-regulates the expression of antiviral genes and
369 prevents viral infection, but continued IFNα2a treatment subsequently causes desensitization and a
370 decrease in antiviral gene expression[30]. In animal studies, it has been shown that IFNα/β is not only
371 antiviral but can also promote inflammation and disease. This occurs *via* immunosuppressive effects
372 that impede viral control[31] or by triggering inflammation and tissue damage[32]. In mice, influenza
373 infection also causes an early influx of neutrophils into the lung followed by a virus specific $CD8^+$ T-
374 cell response[33, 34, 35]. Neutrophils might facilitate the development of this antigen-specific response
375 as they may serve as antigen-presenting cells in influenza infection in mice[35, 36] and guide influenza-
376 specific CD8+ T cells into sites of infection by laying chemokine trails containing CXCL12[37]. In animal
377 models of viral lung disease, dysregulated host immune responses[38] and interferon production[32] can
378 lead to complex inflammatory responses which contribute to pathogenesis[39, 40]. We recognise that
379 transcriptomic data do not always reflect protein data, although genes down-stream of cytokine
380 signalling may remain over-expressed[41]. This is expected, since it is well established that cytokines
381 are under tight regulatory mechanisms and their expression at the level of mRNA and protein is
382 short-lived[42], limiting such correlations.

Given our observed ‘bacterial’ transcriptomic signature, we sought to define the incidence of clinically significant bacterial infections in our patients. To optimise the characterisation, respiratory tract sampling was supplemented with blood cultures for significant bacterial species and urinary pneumococcal antigen, where available. This sampling protocol and analysis went well beyond that normally used in clinical practice, possibly leading to detection of inconsequential bacterial carriage, rather than just those contributing to disease. Additionally, clinicians did not request blood cultures or urinary pneumococcal antigen tests for all patients, restricting the numbers of cases in which we could definitively determine presence or absence of significant bacterial infection. Clearly, we cannot exclude the possibility of bacterial co-infection in patients who were categorised as ‘no significant bacteria detected’; we base our conclusions on the best evidence available to us and exhaustive case-by-case analysis, selecting only those cases which our expert panel felt could confidently be classified as ‘bacterially infected’ or ‘uninfected’. The difficulties that we encountered in confirming or excluding the presence of bacterial infection in patients with respiratory tract infections have been highlighted by others[43, 44].

When measuring total bacterial loads in nasopharyngeal samples, we were surprised to find them of some predictive value: only in cases with low bacterial loads did we see high levels of ‘viral’ response gene activation, while low bacterial load was almost never associated with ‘bacterial’ response gene expression patterns. Indeed this ‘bacterial’ MDTH signature was also associated with serum IL-17 levels; along with higher IL-6 and IL-1β levels (but lower IFN-α2a) at T1, these data were indicative of an immune response biased toward anti-bacterial immunity early in severe influenza. We also examined the possible use of serum or plasma procalcitonin as an indicator of invasive bacterial disease. The strength of the transcriptomic ‘bacterial’ signature showed a significant positive correlation with procalcitonin levels. If procalcitonin were a true marker of bacterial invasive disease this would support the contention that the ‘bacterial’ transcriptional pattern indicates bacterial infection. However, procalcitonin release is suppressed by high levels of type II interferon (which can result from viral infection) and elevated levels of procalcitonin are seen in some non-bacterial inflammatory conditions[26, 45]. In our study, elevated levels of procalcitonin were indicative of severe lung inflammation and did not help in deciding the presence or absence of significant bacterial infection.

To narrow the focus still further, we identified cases with pathogenic bacteria found in blood culture as a subgroup with definite and unequivocal bacterial sepsis. Three of these six cases had markedly elevated ‘bacterial’ molecular score without any increase in ‘viral’ molecular score; one patient had elevated ‘bacterial’ and ‘viral’ scores. All of these patients needed invasive mechanical ventilation.

416 The remaining two cases with bacteraemia did not have marked elevations in their 'bacterial' scores,
417 despite detection of relevant bacteria in blood (*Streptococcus pneumoniae* and Group B
418 *Streptococcus*); both had mild (Severity 1) disease. Interestingly, one of these patients had a high
419 'viral' score but the other did not. Early intervention with antibiotics was also considered as a
420 potential explanation of transcriptomic changes, but in the wider cohort, prior administration of
421 antibiotics had no effect on transcript abundance or the presence or absence of 'viral' or 'bacterial'
422 signatures. From careful analysis of these cases, we conclude that transcriptomic scores are not
423 invariably a reflection of the presence or absence of bacterial co-infection, as far as it has been
424 possible for us to determine.

425 We next used stringent criteria to identify influenza cases that were extensively investigated for
426 bacterial co-infection and yet found not to be infected, and cases in which pathogenic bacteria were
427 identified with certainty. Progression of the transcriptomic signatures observed over time was
428 similar in these two groups, but patients with confirmed bacterial infection had higher 'bacterial'
429 molecular scores overall, compared to those in whom bacteria were not detected, reaching
430 statistical significance at some day-of-illness time-points. We therefore conclude that influenza virus
431 infection alone can drive what has been referred to in the literature and by GO terms as the
432 'bacterial' signature relating to neutrophil and inflammation-associated genes in patients with
433 severe influenza, but that this response is enhanced by bacterial co-infection. The mechanism by
434 which this occurs is open to speculation but might include alterations of innate sensitivity to resident
435 microbiota in the gut, activation of Th17 pathways and leakage of endotoxins from the intestinal
436 lumen[46].

437 Our study has important limitations. Despite its ambition, scope and intensity we had limited
438 numbers of repeat samples in individual patients. Our description of trends over time since onset of
439 illness depends in large part on summative data and on subjective reporting of disease onset. Ideally,
440 the findings need validation in time-series studies simple and complicated acute viral disease with
441 frequent sampling at multiple sites. For ethical and practical reasons, we were unable to plan more
442 than 3 sampling points and not all samples were available from all patients. In addition, we were
443 unable to study the early or preclinical phases but were limited to investigation of symptomatic
444 cases presenting with disease of sufficient severity to reach hospital. For logistical and practical
445 reasons we could not recruit mild cases seen in the community. The prodromal early stage of
446 infection can only be studied with ease in infection challenge of volunteers (in whom severe disease
447 is not expected), but our ongoing studies of experimental infections with pH1N1 in volunteers will
448 allow us to resolve some of these limitations in the future.

449 We here present only selected results of an extended study of soluble immune mediator data from
450 the MOSAIC cohort. Our main findings were of decreased IFNα2a and increased IL-1β, IL-6 and CXCL8
451 levels in the nasal and/or serum compartments in patients with severe disease. This apparent
452 reciprocity may relate to the known cross-regulatory functions of IL-1 and type I IFNs in experimental
453 models[29, 47]. Our results generally fit with the proposal that responses seen in severe influenza are
454 strongly influenced by bacterial co-infection, which contributes to driving high levels of mediators
455 such IL-1β, IL-6 and IL-17. However, there are many additional possible analyses to be performed.
456 We chose only to illustrate those most relevant to the transcriptomic analysis and the question of
457 bacterial superinfection, and we invite readers to explore additional correlations using our online
458 data as a resource. We will welcome discussion with respect to additional analyses.

459 In summary, virus-induced type I interferon-related pathways are activated during the first four days
460 of symptomatic influenza in hospitalised patients. These 'viral' pathways are then down-regulated,
461 to be replaced by inflammatory, activated neutrophil and apoptosis-related pathways associated
462 with IL-17 abundance, host-mediated tissue damage and 'response to bacteria', particularly in cases
463 with a high 16S bacterial load in the nasopharyngeal secretions. In severe cases of influenza, this
464 'viral' response may be depressed early in disease and is accompanied by an increase in IL-1β and IL-
465 17. These findings emphasise that the stage and severity of disease need to be taken into account in
466 interpreting host responses to infection and in the development of potential diagnostic tests to
467 differentiate between treatable causes.

468 Online Methods

469 **Study population and inclusion criteria**

470 Patients ≥ 16 years of age were recruited during two successive winters (01 December 2009 to 03
471 March 2011). Patients with suspected influenza were identified by medical or nursing staff, or
472 notified to investigators by hospital diagnostic laboratories. Patients in London were recruited from
473 four Imperial College Healthcare NHS Trust hospitals, the Chelsea and Westminster Hospital, and the
474 intensive care unit at the Royal Brompton Hospital (a national referral centre for severe respiratory
475 failure). In Liverpool, patients were recruited from the Royal Liverpool, Liverpool Women's and
476 Arrowe Park Hospitals. Patients were included irrespective of prior or concurrent comorbidity (most
477 commonly asthma, pregnancy, immunocompromising conditions, or co-infection with other
478 respiratory pathogens), to reflect the populations known to be at greatest risk of severe influenza.
479 Adult healthy controls were recruited and matched to the patient cohorts for age, sex and ethnicity
480 and were screened to exclude known illnesses or current use of medications.

481 **Research Ethics Committees' Approvals**

482 The study was approved by the NHS National Research Ethics Service, Outer West London REC
483 (09/H0709/52, 09/MRE00/67). Patients or their legally authorised representatives provided
484 informed consent. Additional adult healthy controls were recruited as part of a separate study and
485 consented to their samples being used in additional studies (Central London 3 Research Ethics
486 Committee, 09/H0716/41).

487 **Biological sampling**

488 Research samples were obtained at three time points: T1 (recruitment); T2 (approximately 48h after
489 T1); T3 (at least 4 weeks after T1). Only T1 and T2 samples were included in this report. Whole blood
490 samples for transcriptomics were collected during the two recruitment periods, 2009/10 and
491 2010/11. Of 85 MOSAIC participants presenting with influenza-like illness in 2009/10, 23 (27%) were
492 adults with confirmed influenza, and T1 transcriptomic samples were available from 22 adults. Of
493 171 MOSAIC participants presenting with influenza-like illness in 2010/11, 111 (65%) were adults
494 with confirmed influenza, and T1 transcriptomics samples were available from 109/111 (98%). RNA
495 extraction and microarray were successful for all available patient samples from both cohorts.
496 Microarrays were also performed on samples from adult healthy controls of similar age, sex and
497 ethnicity to the study patients (Table 1). One healthy control sample for the 2009/10 cohort was not
498 included in final analysis because it failed quality control assessments.

499 Of the 109 adult patients recruited in 2010/11 and included in this analysis, 94 (86%) were infected
500 with A(H1N1)pdm09 influenza virus, the remainder being infected with influenza A(H3N2) virus, non-
501 subtyped influenza A virus, or influenza B virus. One of 22 adult patients recruited during 2009/10
502 was infected with A(H3N2) virus; remaining patients were infected with A(H1N1)pdm09 virus. Due to
503 the natural evolution of influenza activity during the 2009-10 pandemic in the UK, the 2009/10
504 cohort was smaller than originally anticipated. Therefore, to assess the host response in the blood
505 transcriptional signature as thoroughly as possible, we focussed our analysis on the larger 2010/11
506 cohort and then compared findings with the smaller 2009/10 cohort.

507 **Influenza virus infection status**

508 For each participant, influenza virus infection status was determined by reverse transcription
509 polymerase chain-reaction (RT-PCR) testing of an appropriate respiratory tract sample by local
510 clinical virology laboratories, as part of routine clinical care. Clinical laboratories followed nationally
511 agreed and validated PCR protocols, and a panel of experts reviewed all results.

512 **Clinical data collection and severity of illness scoring**

513 Clinical data were extracted from hospital case notes and recorded in the Flu-CIN data collection
514 tool[48] by trained researchers. Prescription charts were examined to determine whether antibiotics
515 were being administered before, during or after sampling time points.

516 Severity of illness was graded at T1 and T2 according to the following criteria: (1) no significant
517 respiratory compromise, with blood oxygen saturation >93% whilst breathing room air; (2) oxygen
518 saturation ≤ 93% whilst breathing room air, justifying or requiring supplemental oxygen by face mask
519 or nasal cannulae (with or without continuous positive airway pressure support or non-invasive
520 mechanical ventilation); (3) respiratory compromise requiring invasive mechanical ventilation with
521 or without ECMO. All clinical data underwent extensive validation and quality checking by
522 independent data collection staff.

523 **Detection of bacteria**

524 Nasopharyngeal aspirates and swabs collected at T1 underwent microscopy and culture for bacteria.
525 Additionally, multiplex PCR was performed to detect the following common respiratory bacteria in
526 these samples: *Staphylococcus aureus*, *Chlamydia pneumoniae*, *Haemophilus influenzae*,
527 *Streptococcus pneumoniae*, *Pneumocystis pneumoniae*, *Legionella* species, *Klebsiella pneumoniae*,
528 *Salmonella* species, *Moraxella catarrhalis*, *Mycoplasma pneumoniae,* and *Bordetella pertussis*.
529 Throat swab samples taken at T1 also underwent culture and microscopy. When available, urine
530 samples collected between T1 and T2 underwent pneumococcal antigen testing (BinaxNow, Allere,

531 Stockport, UK). Clinical microbiology data were obtained from hospital laboratory databases,
532 including results of blood cultures (when taken 48 hours either side of T1) and urinary pneumococcal
533 antigen results (for patients who did not have a researcher-requested urinary antigen sample). An
534 independent microbiologist assessed the significance and validity of positive blood culture results, in
535 an attempt to exclude cases of pseudobacteraemia caused by commensal contamination.

536 **Soluble immune mediators**

537 Serum, nasopharyngeal aspirate (NPA) and nasal-absorption fluid were collected at recruitment (T1)
538 from participants with confirmed influenza and from adult healthy controls. Clotted blood was
539 centrifuged at 1000 x g at 4°C and aliquots of serum supernatant were stored at -80°C. Each NPA was
540 collected using a 10F Argyle suction catheter, inserted to reach the posterior nasopharyngeal wall;
541 moderate suction was applied while the catheter was withdrawn over 5 seconds. The catheter was
542 flushed through with 5 mL of sterile normal saline and the total contents were collected in a
543 universal container. Aliquots of NPA were stored at -80°C. Nasal-absorption fluid was collected from
544 the lateral wall of the nasal cavity using a synthetic absorptive matrix (SAM) strips (Leukosorb, Pall,
545 UK) and stored at -80°C until analysis. On the day of analysis, 500 µl Milliplex assay buffer (Millipore,
546 UK) was added to each thawed SAM strip before being placed in a Costar Spin-X centrifuge filter of
547 pore size 0.22 µm held within an Eppendorf tube. Samples were centrifuged at 16,000 x g for 5
548 minutes at 4°C and eluates were kept on ice.

549 IL-1β, IL-6 and CXCL8 were quantified in each sample type using a 10-plex inflammatory soluble
550 immune mediator electrochemiluminescence assay analysed on an MSD SECTOR instrument (Meso
551 Scale Discovery, USA). For each mediator, a percentage coefficient variation cut-off of 10% was used
552 to set the lower limit of detection. Sample results below the GM-LLOD were assigned half the value
553 of the respective GM-LLOD.

554 **Blood procalcitonin assay**

555 Procalcitonin (PCT) in plasma or serum (collected at T1 and T2) was quantified using the Elecsys
556 BRAHMS PCT assay on a calibrated Cobas e602 platform. Samples with a PCT value at the upper limit
557 of detection (ULOD) were arbitrarily assigned the value of 100 ng/mL (the ULOD). Results may be
558 interpreted as follows: <0.5 ng/mL, low probability of significant bacterial infection; 0.5-2.0 ng/mL,
559 medium probability of significant bacterial infection; >2.0 ng/mL, high probability of significant
560 bacterial infection.

561 **16S rRNA gene bacterial load measurement**

562 The 16S rRNA gene was targeted with 0.3 µl each of 10 µM universal primers 520F 5'-AYT GGG YDT
563 AAA GNG and 802R 5'-TAC NVG GGT ATC TAA TCC added to 7.5 µl of SYBR Fast qPCR Kit Master Mix
564 (KapaBio) and 5 µl of a 1:5 dilution of sample DNA extract and 1.9 µl of PCR Clean water (Mobio).
565 Reactions were prepared in triplicate and thermal cycling carried out on a VIIA-7 Real-Time PCR
566 System. Thermal-cycling conditions were 90°C for 3 mins, then 40 cycles of 95 °C for 20 s, 50 °C for
567 30 s, 72 °C for 30 s with default melt conditions. A standard curve of cloned (TOPO TA, Invitrogen)
568 full length *Vibrio natriegens* DSMZ 749 16S rRNA gene was included in order to be able to calculate
569 an absolute abundance from C_T values together with no template controls. The resulting 16S rRNA
570 gene copy number (bacterial load) was log transformed prior to using analytically.

571 **Microarray Gene Expression Profiling**

572 At each time point, 3 ml of whole blood were collected into each of two Tempus tubes (Applied
573 Biosystems/Ambion) by trained research staff following a standard phlebotomy protocol. Blood was
574 vigorously mixed immediately following collection and stored at -80°C before RNA extraction. For
575 each patient, the contents of one tube were used for analysis and the other tube was retained in
576 case of assay failure. RNA was isolated using 1.5 ml whole blood and the MagMAX-96 Blood RNA
577 Isolation Kit (Applied Biosystems/Ambion), as per the manufacturer's instructions. 250 µg of isolated
578 total RNA was globin-reduced using the GLOBINclear 96-well format kit (Applied
579 Biosystems/Ambion) according to the manufacturer's instructions. Total and globin-reduced RNA
580 integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA yield was
581 assessed using a NanoDrop8000 spectrophotometer (NanoDrop Products, Thermo Fisher Scientific).
582 High-quality (>6.5 RIN) whole blood RNA was successfully obtained and processed by microarray in
583 all cases. Biotinylated, amplified antisense complementary RNA (cRNA) targets were prepared from
584 200-250 ng of globin-reduced RNA using the Illumina CustomPrep RNA amplification kit (Applied
585 Biosystems/Ambion). For each sample, seven hundred and fifty nanograms of labelled cRNA were
586 hybridised overnight to Illumina Human HT12 V4 BeadChip arrays (Illumina), which contained
587 greater than 47,000 probes. The arrays were washed, blocked, stained and scanned on an Illumina
588 iScan, as per the manufacturer's instructions. GenomeStudio (Illumina) was used to perform quality
589 control and generate signal intensity values.

590 **Microarray Data Processing**

591 Raw microarray data were processed using GeneSpring GX version 12.5 (Agilent Technologies).
592 Following background subtraction, each probe was attributed a flag to denote its signal intensity
593 detection *P* value. Filtering on flags removed probe sets that did not result in a 'present' call in at

594 least 10% of the samples, where the 'present' lower cut-off = 0.99. Signal values were then set to a
595 threshold level of 10, log^2 transformed, and per-chip normalised using a 75th percentile-shift
596 algorithm. Each gene was normalised by dividing each mRNA transcript by the median intensity of all
597 samples. Statistical analysis was performed after these steps had been performed.

598 **Microarray Data Analysis**

599 Transcripts significantly detected from background hybridisation were filtered for low expression in
600 GeneSpring GX 12.5, whereby the only transcripts retained were those with at least two-fold change
601 from the median normalised intensity value in at least 10% of all samples. Principal component
602 analysis of all transcripts significantly above background in at least 10% of samples (18974
603 transcripts) was performed using R 3.3.2 (R Development Core Team). To derive the 1255 transcript
604 list, non-parametric statistical filters (Mann-Whitney unpaired test with Bonferroni family-wise error
605 rate multiple testing correction, $p < 0.01$) were applied, followed by fold-change filtering between
606 groups (transcripts were retained if greater than two-fold change between any two groups). For
607 severity analysis, 231 normalised intensity value transcripts were obtained by filtering for low
608 expression and then applying statistic filters (Kruskal-Wallis test with Bonferroni FWER, $P < 0.01$),
609 followed by fold change filtering between groups (transcripts were retained if greater than two-fold
610 change between those from severity 3 patients and severity 1 and 2 patients). All heat-maps were
611 generated in GeneSpring GX 12.5 (semi-supervised analysis, clustered by Pearson's un- centred
612 method with average linkage rule).

613 Comparison Ingenuity Pathway Analysis (IPA) (Ingenuity Systems Inc., Redwood, CA) was used to
614 determine the most significant canonical pathways for up-regulated and down-regulated transcripts
615 (P <0.05 Fishers Exact test). Additionally, IPA was used to generate graphical representations of
616 selected canonical pathways, and network diagrams. For the 231 transcript list, significantly
617 activated (z score >2) and significantly repressed (z score <2) biofunctions were identified in IPA and
618 expressed in gene network diagrams. GO Term (Gene Ontology Consortium) analysis integrated with
619 GeneSpring GX12.5 was used to identify biological processes, according to GO annotations[49].

620 The molecular distance to health (MDTH) and molecular scored were calculated using methods
621 described previously[24] and applied to different signatures. Transcriptional modular analysis was
622 applied as described previously[23]. Briefly, raw expression levels of all transcripts significantly
623 detected from background were compared between each sample and all the controls present in a
624 given data set. The percentage of significantly expressed genes in each module was represented by
625 the colour intensity, with red indicating over-expression and blue indicating under-expression.
626 Statistical testing was performed using Student's t-test ($P < 0.05$). The mean percentage of

627 significant genes and the mean fold change of these genes compared to the controls in specific
628 modules were shown in graphical form (unpaired t-test, $P < 0.00001$). MDTH and modular analysis
629 were calculated in Microsoft Excel 2010 (Microsoft Corp.). GraphPad Prism V5 for Windows
630 (GraphPad Software Inc., La Jolla, CA, USA) and R 3.3.2 (R Development Core Team) were used to
631 generate graphs and perform additional statistical analyses.

Figure 1. Transcriptional signature of influenza compared to healthy controls. (a) Principal component analysis of all transcripts significantly above background in at least 10% of samples (130 healthy controls (green squares), 97 influenza A (red circles: H1N1; green triangles, H3N2), and 12 influenza B (purple squares); all from 2010/11). **(b)** Modular analysis of influenza patients relative to healthy controls. The expression of the modules is shown on the left according to the colour intensity display; the corresponding modules are identified in the key to the right. **(c)** Weighted 'molecular distance to health' (MDTH[24] of Influenza patients compared to healthy controls, undertaken on 4526 transcripts that were significantly detected above background, filtered for low expression (transcripts retained if >2 fold-change (FC) from median normalised intensity value in more than 10% of all samples). Box whisker plot with min/max lines; statistical test: Mann-Whitney $P< 0.0001$. **(d)** Heat-map of 1255 normalised intensity value transcripts, filtered for low expression then statistically filtered (Mann-Whitney with Bonferroni multiple testing correction $P<0.01$) followed by fold change filter between groups (transcripts retained if >2FC between any 2 groups). Listed next to the heat-map are the top five IPA canonical pathways (by significance $P<0.05$, Fisher's Exact test) for upregulated and downregulated transcripts. **(e)** IPA canonical pathway for Interferon Signalling. Red shading represents up-regulated genes, blue represents down-regulated genes. **(f)** Heat-map of normalised intensity values of the top 25 significant transcripts by mean fold-change between healthy controls and influenza groups clustered on entities and by individuals (Pearson's uncentered (cosine) with averaged linkage).

Figure 2. Severity of disease is associated with diminished expression of interferon-related modules and over-expression of inflammation modules. (a) Weighted molecular distance to health (MDTH) of Influenza patients grouped by severity of illness score (1: normoxic; 2: hypoxia requiring correction by mask oxygen; 3: mechanical ventilation), compared to healthy controls (HC), based on 4526 transcripts that were significantly expressed above background and filtered for low expression (transcripts retained if >2FC from median normalised intensity value in more than 10% of all samples). Box whisker plots are shown with min/max lines. **(b)** Modular analysis of influenza patients grouped by severity, relative to healthy controls. The colour intensity correlates with the percentage of genes in that module that are significantly differentially expressed.

Figure 3. Severe disease is associated with lower expression of "viral" response genes, compared to early and less severe influenza. (a) Heat-map of 231 normalised intensity value transcripts, obtained by filtering for low expression followed by statistical filtering (Kruskal-Wallis with Bonferroni multiple testing correction $P<0.01$) followed by fold change filter between groups (restricted to initial T1 samples, transcripts retained if >2FC between severity 3 and severity 1&2).

665 Listed next to the heat-map are the top GO terms for the 3 major subdivisions of the dendrogram
666 (clustered by Pearson's uncentered (cosine) with average linkage rule). **(b)** Weighted molecular
667 score (relative to healthy controls) of the 112 'bacterial response' transcripts plotted against
668 molecular score of the 51 'viral response' transcripts for the 109 influenza individuals at the T1 time
669 point. Severity of illness is indicated by different colours of dots: severity 1, black dots; severity 2,
670 blue dots; severity 3, red dots. Circled dots identify patients with confirmed bacteraemia. **(c)** IPA
671 significantly activated (*z* score >2) or **(d)** repressed (*z* score <2) biofunctions, identified by analysis of
672 231 transcript list; selected networks of biofunctional genes are shown.

673 **Figure 4. Relationship between severity of illness, bacterial infection, procalcitonin and molecular**
674 **scores.** 'Viral' and 'Bacterial' response scores (according to GO terms, as described in Fig. 3)
675 calculated for individual cases of confirmed influenza according to clinical categories at both the first
676 and second sampling time-points (T1 and T2). Loess fitting was used to interpolate and estimate
677 mean values non-parametrically from the data (solid lines); dashed lines show the estimated 95%
678 confidence interval values of the mean; statistical significance of differences were calculated using
679 Chi-squared tests to compare the deviance of generalized linear models. **(a)** Separating cases into
680 grades of clinical severity, cases without need for mechanical ventilatory support (i.e. severity 1 and
681 2), showing high 'viral' molecular score at the early stage of the disease. **(b)** Those requiring
682 mechanical ventilatory support (severity 3) had higher 'bacterial' molecular scores regardless of time
683 since onset. **(c)** When classified according to the presence of absence of clinically significant
684 bacterial co-infections, those with proven bacterial co-infection (39 subjects, 63 samples) had lower
685 'viral' molecular scores than those without identifiable co-infection (34 subjects, 52 samples)
686 regardless of time since disease onset. **(d)** Those with proven bacterial co-infection (39 subjects, 63
687 samples) had higher 'bacterial' molecular scores than those without identifiable co-infection (34
688 subjects, 52 samples) regardless of time since disease onset. **(e)** The 'viral' molecular score was
689 unrelated to serum procalcitonin; the 'bacterial' molecular score **(f)** tended to be high in those with
690 raised procalcitonin, but was unaffected by proven bacterial co-infection.

691 **Figure 5. Levels of selected mediators in different compartments according to severity of illness**
692 **and clinical designation of probable bacterial co-infection status.** Serum, nasopharyngeal aspirate
693 (NPA) and nasabsorption eluates from participants with confirmed influenza were obtained at
694 recruitment (T1) and compared with samples from adult healthy controls. Individual values (pg/mL,
695 log scale) are shown with median and interquarterile ranges. Zero values and values below the lower
696 limit of detection were assigned half the geometric mean lower limit of detection for display
697 purposes. The upper limit of detection for all assays shown was 2500 pg/mL. Kruskal-Wallis test with

698 Dunn's post test was used to assess significance (*** p<0.001; ** p<0.01; * p<0.05; NS = not
699 significant). Severity of illness at T1 is shown. HC = healthy controls. Serum samples for HCs and
700 participants with severity 1, 2, and 3 illness **(a, d, g, and j)**: n = 36, 58, 43, and 31, respectively. NPA
701 samples for healthy controls and participants with severity 1, 2, and 3 illness **(b, e, h and k)**: n = 35,
702 50, 32, and 27, respectively. Nasabsorbtion eluate samples for healthy controls and participants with
703 severity 1, 2, and 3 illness **(c, f, i and l)**: n = 36, 60, 43, and 30, respectively.

704 **Figure 6: Relationships between severity of illness, bacterial infection, and selected mediators.**
705 Levels of CXCL10 and IL-6 in serum **(a-d)** and NPA **(e-h)** according to day of illness at both the first
706 and second sampling time-points (T1 and T2). Loess fitting was used to demonstrate time trends of
707 mean values interpolated non-parametrically from the data (solid lines); dashed lines show the
708 estimated 95% confidence interval values of the mean. Statistical significance of differences was
709 calculated using Chi-squared tests to compare the deviance of generalized linear models. Patients
710 requiring mechanical ventilation (severity 3 illness) had significantly higher mean levels of CXCL10
711 and IL-6 in both serum **(a, b)** and NPA **(e, f)** than patients with less severe diseases (severity 1 and 2).
712 Classifying cases according to the presence or absence of clinically significant bacterial co-infections,
713 those with proven bacterial infection (39 subjects, 63 samples) had higher mean IL-6 levels in both
714 serum **(d)** and NPA **(h)** compared to patients who did not (34 subjects, 52 samples), regardless of
715 time since disease onset. Bacterial infection had no discernible effect on serum **(c)** or NPA **(g)** levels
716 of CXCL10.

717

718 **Supplementary Figure 1. Validation of transcriptional signatures in an independent cohort. (a)**
719 2009/2010 cohort clustered on individuals and transcripts (Pearson's uncentered with averaged
720 linkage) using 1255 transcript list (from Figure 1). **(b)** 2009/2010 cohort clustered on individuals and
721 transcripts (Pearson's uncentered with averaged linkage) using 25 transcript list (from Figure 1). **(c)**
722 2009/2010 cohort clustered on individuals and transcripts (Pearson's uncentered with averaged
723 linkage) using 231 transcript list of severity (from Figure 2, transcripts retained if >2FC between
724 severity 3 and severity 1&2). GO Terms analysis of 3 major branches of the transcripts dendrogram
725 was undertaken and listed next to the heat-map. **(d)** Using 51 and 112 transcripts lists (from Figure
726 3) 'viral response' and 'bacterial response' molecular scores were calculated and plotted for each
727 influenza patient (relative to healthy controls). Cases were coded according to severity of illness,
728 indicated by the colour of the respective dots (severity 1, black; severity 2, blue; severity 3, red).

729 **Supplementary Figure 2. Change of 'viral' and 'bacterial' molecular scores over time and**
730 **association with viral load.** **(a)** 'Viral' molecular scores plotted for 59 influenza patients (2010/11

731 cohort) who provided T1 and T2 samples, plotted against respective day of illness at time of
732 sampling. **(b)** Change in 'viral' molecular score between first (T1) and precise second time point (48
733 hours after T1) in 41 patients with appropriate samples available (P = 0.0002, Mann-Whitney test).
734 **(c)** 'Bacterial' molecular score plotted for 59 influenza patients who had both a T1 and a T2 sample,
735 shown plotted against respective day of illness. **(d)** Change in 'bacterial' molecular score between T1
736 and precise T2 (48h post T1), in 41 patients with appropriate samples available (*NS*, Mann-Whitney
737 test). **(e)** Influenza viral load estimation (Pfu/ml) in nasopharyngeal samples obtained at T1 and T2.
738 **(f)** Relationship between influenza viral load (Pfu/ml) at T1 or T2 and the simultaneous 'viral'
739 molecular score on whole blood.

740 **Supplementary Figure 3. Administration of antibiotics does not affect 'bacterial' or 'viral'
741 molecular scores. (a)** 109 Influenza patients (2010/11 cohort) presenting within the first 14 days of
742 illness grouped by reciept of any antibiotic in the 24 hours prior to T1 sample. There was no
743 difference (*NS,* Mann-Whitney test) in either bacterial or viral molecular scores between the two
744 groups. **(b)** Prescription of antibiotics after T1 did not significantly influence 'bacterial' molecular
745 score. Fifty-nine influenza patients who had both T1 and T2 samples were grouped by those who did
746 not receive antibiotics (n=7), those whose antibiotics were stopped at T1 (n=1), those who had
747 antibiotics prescribed after T1 but before T2 (n=24), and those who were receiving antibiotics at
748 both T1 and T2 (N=27). **(c)** Total 16S copies at T1 in throat swabs and NP aspirate in patients
749 adjudicated to be with or without bacterial co-infection. Those with confirmed bacterial infections
750 had greater levels of total 16S copies in NP aspirate than those deemed to be without co-infection (P
751 = 0.036).

752 **Supplementary figure 4. Correlation of serum cytokines and bacterial load in nasophaynx with
753 'viral' and 'bacterial' molecular distance to health.** Levels of IL-17 in the serum of healthy controls
754 (HC) and influenza infected patients (severity 1-3) **(4a)**. Concentration of IL-17 in broncoalveolar
755 lavage (BAL) of HC and from BAL, NPA, SAM and Serum of patients **(4b)**. Serum levels of IL-17 **(4c)**
756 and TNFα **(4d)** correlate with the bacterial MDTH (IL-17, Spearman R =0.39, P value <0.001; TNFα R =
757 0.4, P < 0.01). The total 16S rRNA gene copies in NP aspirate samples were inversely correlated with
758 the viral MDTH (Spearman R = -0.28, P value < 0.05) **(4e)** but positively correlated with bacterial
759 MDTH (Spearman R = 0.37, P value < 0.05) **(4f)**.

760 **Author Contributions**

761 Conceived and designed the study: PJMO, AOG, CMG, DC, JB, SB, and JD. Performed the microarray
762 experiments: CMG. Developed clinical protocols, recruited subjects and collated clinical data: JD.
763 Analysed the microarray data: JD, CIB, LH and SB with supervision by AOG and PJMO. Wrote and
764 revised the manuscript: JD, SB, LH, AOG, PJMO.

765 **Acknowledgments**

766 MOSAIC Study was supported by a joint award from the Wellcome Trust and the Medical Research
767 Council (090382/Z/09/Z). We gratefully acknowledge the support of the MOSAIC administrative
768 team (Mary Cross, Lindsey-Anne Cumming, Matthew Minns, Tom Ford, Barbara Cerutti, Denise
769 Gardner and Zoe Williams) and the generosity of our patients and their families, healthy volunteers,
770 and staff at participating National Health Service (NHS) hospitals (Alder Hey Children's Hospital;
771 Brighton & Sussex University Hospitals NHS Trust; Central Manchester University Hospitals NHS
772 Foundation Trust; Chelsea and Westminster Hospital NHS Foundation Trust; Imperial College
773 Healthcare NHS Trust; Liverpool Women's NHS Foundation Trust; Royal Liverpool and Broadgreen
774 University Hospitals NHS Trust; Royal Brompton and Harefield NHS Foundation Trust; University
775 Hospitals Coventry and Warwickshire NHS Trust).

776 In particular, we thank Alshafi, K.; Ashton, S.; Bailey, E.; Bermingham, A.; Berry, M.; Bloom, C.; Booth,
777 A.; Brannigan, E.; Bremang, S.; Clark, J.; Cross, M.; Cumming, L. A.; Dyas, S.; England-Smith, J.;
778 Enstone, J.; Ferreira, D.; Goddard, N.; Godlee, A.; Gormley, S.; Guiver, M.; Hassan-Ibrahim, M.O.; Hill,
779 H.; Holloway, P.; Hoschler, K.; Houghton, G.; Hughes, F.; Israel, R.R.; Jepson, A.; Jones, K.D.; Kelleher,
780 W.P.; Kidd, M.; Knox, K.; Lackenby, A.; Lloyd, G.; Longworth, H.; Mookerjee, S.; Mt-Isa, S.; Muir, D.;
781 Paras, A.; Pascual, V.; Rae, L.; Rodenhurst, S.; Rozakeas, F.; Scott, E.; Sergi, E.; Shah, N.; Sutton, V.;
782 Vernazza, J.; Walker, A.W.; Wenden, C.; Wotherspoon, T.; Wright, A.D. and Wurie, F. We also thank
783 E. Anguiano and members of the Genomic Core, BIIR, Dallas, for assisting with the microarray
784 analysis and to M. Berry, Imperial College London, for guidance. We especially thank the MOSAIC
785 Data Team (Lydia Drumright, Laura Garcia-Alvarez, Judith Lieber, Sid Mookerjee, and Beth Pamba)
786 for assistance in collating and validating clinical data, Rosalind Smyth for careful review of the
787 manuscript and Dr. Korbinian Strimmer (of Imperial College London) for statistical advice in revisions
788 of the manuscript.

789 The MOSAIC consortium was supported by several Comprehensive Local Research Networks
790 (CLRNs), the National Institute for Health Research (NIHR) Comprehensive Local Research Network
791 and Biomedical Research Centre (NIHR Imperial BRC) and Unit (BRU) support in both London and

792 Liverpool, the Health Protection Agency (latterly Public Health England) Microbiology Services,
793 Colindale and the staff of the Roslin Institute, Edinburgh, Scotland. Anne O'Garra and Christine
794 Graham were supported by the Medical Research Council, United Kingdom (U117565642), The
795 Francis Crick Institute, London (AOG10126, which receives its core funding from Cancer Research
796 U.K. (FC001126), the U.K. Medical Research Council (FC001126), the Wellcome Trust (FC001126),
797 and the U.K. Medical Research Council (MR/U117565642/1). SB was in part jointly funded by the UK
798 Medical Research Council (MRC) as above and the UK Department for International Development
799 (DFID) under the MRC/DFID Concordat agreement (MR/J010723/1); Chloe Bloom was funded by an
800 MRC CRTF. The views expressed are those of the authors and not necessarily those of the NHS, NIHR,
801 Public Health England or the Department of Health (UK). The funders had no role in study design,
802 data collection and analysis, decision to publish, or preparation of the manuscript.

803 **References**

804 1. Stöhr, K. Preventing and treating influenza. *BMJ* **326**, 1223-1224 (2003).

805
806 2. Hayward, A.C. *et al.* Comparative community burden and severity of seasonal and pandemic
807 influenza: results of the Flu Watch cohort study. *The Lancet. Respiratory medicine* **2**, 445-454
808 (2014).

809
810 3. Dawood, F.S. *et al.* Estimated global mortality associated with the first 12 months of 2009
811 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet infectious*
812 *diseases* **12**, 687-695 (2012).

813
814 4. Bautista, E. *et al.* Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection. *The*
815 *New England journal of medicine* **362**, 1708-1719 (2010).

816
817 5. Hui, D.S., Lee, N. & Chan, P.K. Clinical management of pandemic 2009 influenza A(H1N1)
818 infection. *Chest* **137**, 916-925 (2010).

819
820 6. Peiris, J.S., Cheung, C.Y., Leung, C.Y. & Nicholls, J.M. Innate immune responses to influenza A
821 H5N1: friend or foe? *Trends in immunology* **30**, 574-584 (2009).

822
823 7. Tisoncik, J.R. *et al.* Into the eye of the cytokine storm. *Microbiology and molecular biology*
824 *reviews : MMBR* **76**, 16-32 (2012).

825
826 8. de Jong, M.D. *et al.* Fatal outcome of human influenza A (H5N1) is associated with high viral
827 load and hypercytokinemia. *Nat Med* **12**, 1203-1207 (2006).

828
829 9. Ioannidis, I. *et al.* Plasticity and virus specificity of the airway epithelial cell immune response
830 during respiratory virus infection. *Journal of virology* **86**, 5422-5436 (2012).

831
832 10. Ramilo, O. *et al.* Gene expression patterns in blood leukocytes discriminate patients with
833 acute infections. *Blood* **109**, 2066-2077 (2007).

834
835 11. Zaas, A.K. *et al.* Gene expression signatures diagnose influenza and other symptomatic
836 respiratory viral infections in humans. *Cell Host Microbe* **6**, 207-217 (2009).

837
838 12. Woods, C.W. *et al.* A host transcriptional signature for presymptomatic detection of
839 infection in humans exposed to influenza H1N1 or H3N2. *PloS one* **8**, e52198 (2013).

840
841 13. Nakaya, H.I. *et al.* Systems biology of vaccination for seasonal influenza in humans. *Nat*
842 *Immunol* **12**, 786-795 (2011).

843

14. Li, S. *et al.* Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* **15**, 195-204 (2014).

15. Obermoser, G. *et al.* Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity* **38**, 831-844 (2013).

16. Blankley, S. *et al.* The application of transcriptional blood signatures to enhance our understanding of the host response to infection: the example of tuberculosis. *Philos Trans R Soc Lond B Biol Sci* **369**, 20130427 (2014).

17. Pascual, V., Chaussabel, D. & Banchereau, J. A genomic approach to human autoimmune diseases. *Annual review of immunology* **28**, 535-571 (2010).

18. Everitt, A.R. *et al.* IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* **484**, 519-523 (2012).

19. Ruggiero, T. *et al.* A(H1N1)pdm09 hemagglutinin D222G and D222N variants are frequently harbored by patients requiring extracorporeal membrane oxygenation and advanced respiratory assistance for severe A(H1N1)pdm09 infection. *Influenza and other respiratory viruses* **7**, 1416-1426 (2013).

20. Rykkvin, R., Kilander, A., Dudman, S.G. & Hungnes, O. Within-patient emergence of the influenza A(H1N1)pdm09 HA1 222G variant and clear association with severe disease, Norway. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* **18** (2013).

21. Herfst, S. *et al.* Introduction of virulence markers in PB2 of pandemic swine-origin influenza virus does not result in enhanced virulence or transmission. *Journal of virology* **84**, 3752-3758 (2010).

22. Elderfield, R.A. *et al.* Accumulation of human-adapting mutations during circulation of A(H1N1)pdm09 influenza virus in humans in the United Kingdom. *Journal of virology* **88**, 13269-13283 (2014).

23. Chaussabel, D. *et al.* A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity* **29**, 150-164 (2008).

24. Pankla, R. *et al.* Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol* **10**, R127 (2009).

25. Simon, L., Gauvin, F., Amre, D.K., Saint-Louis, P. & Lacroix, J. Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-

887 analysis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society*
888 *of America* **39**, 206-217 (2004).

889
890 26. Wu, M.H. *et al.* Can procalcitonin tests aid in identifying bacterial infections associated with
891 influenza pneumonia? A systematic review and meta-analysis. *Influenza and other*
892 *respiratory viruses* **7**, 349-355 (2013).

893
894 27. Falsey, A.R. *et al.* Bacterial complications of respiratory tract viral illness: a comprehensive
895 evaluation. *The Journal of infectious diseases* **208**, 432-441 (2013).

896
897 28. Rosebeck, S. & Leaman, D.W. Mitochondrial localization and pro-apoptotic effects of the
898 interferon-inducible protein ISG12a. *Apoptosis* **13**, 562-572 (2008).

899
900 29. McNab, F., Mayer-Barber, K., Sher, A., Wack, A. & O'Garra, A. Type I interferons in infectious
901 disease. *Nat Rev Immunol* **15**, 87-103 (2015).

902
903 30. Sandler, N.G. *et al.* Type I interferon responses in rhesus macaques prevent SIV infection and
904 slow disease progression. *Nature* **511**, 601-605 (2014).

905
906 31. Biron, C.A. Interferons alpha and beta as immune regulators--a new look. *Immunity* **14**, 661-
907 664 (2001).

908
909 32. Davidson, S., Crotta, S., McCabe, T.M. & Wack, A. Pathogenic potential of interferon
910 alphabeta in acute influenza infection. *Nat Commun* **5**, 3864 (2014).

911
912 33. Narasaraju, T. *et al.* Excessive neutrophils and neutrophil extracellular traps contribute to
913 acute lung injury of influenza pneumonitis. *The American journal of pathology* **179**, 199-210
914 (2011).

915
916 34. Lukens, M.V. *et al.* A systemic neutrophil response precedes robust CD8(+) T-cell activation
917 during natural respiratory syncytial virus infection in infants. *Journal of virology* **84**, 2374-
918 2383 (2010).

919
920 35. Tate, M.D., Brooks, A.G., Reading, P.C. & Mintern, J.D. Neutrophils sustain effective CD8(+) T-
921 cell responses in the respiratory tract following influenza infection. *Immunology and cell*
922 *biology* **90**, 197-205 (2012).

923
924 36. Hufford, M.M. *et al.* Influenza-infected neutrophils within the infected lungs act as antigen
925 presenting cells for anti-viral CD8(+) T cells. *PloS one* **7**, e46581 (2012).

926
927 37. Lim, K. *et al.* Neutrophil trails guide influenza-specific CD8(+) T cells in the airways. *Science*
928 *(New York, N.Y.)* **349**, aaa4352 (2015).

929
930 38. Openshaw, P.J. & Chiu, C. Protective and dysregulated T cell immunity in RSV infection. *Curr*
931 *Opin Virol* **3**, 468-474 (2013).

932
933 39. Gregory, D.J. & Kobzik, L. Influenza lung injury: mechanisms and therapeutic opportunities.
934 *Am J Physiol Lung Cell Mol Physiol* **309**, L1041-1046 (2015).

935
936 40. Short, K.R., Kroeze, E.J., Fouchier, R.A. & Kuiken, T. Pathogenesis of influenza-induced acute
937 respiratory distress syndrome. *Lancet Infect Dis* **14**, 57-69 (2014).

938
939 41. Berry, M.P. *et al.* An interferon-inducible neutrophil-driven blood transcriptional signature in
940 human tuberculosis. *Nature* **466**, 973-977 (2010).

941
942 42. Anderson, P. Post-transcriptional regulons coordinate the initiation and resolution of
943 inflammation. *Nat Rev Immunol* **10**, 24-35 (2010).

944
945 43. Jain, S. & Pavia, A.T. Editorial Commentary: The Modern Quest for the "Holy Grail" of
946 Pneumonia Etiology. *Clin Infect Dis* **62**, 826-828 (2016).

947
948 44. Dickson, R.P. The microbiome and critical illness. *Lancet Respir Med* **4**, 59-72 (2016).

949
950 45. Maier, M. *et al.* Serum procalcitonin levels in patients with multiple injuries including visceral
951 trauma. *The Journal of trauma* **66**, 243-249 (2009).

952
953 46. Marshall, J.C. Iatrogenesis, inflammation and organ injury: insights from a murine model. *Crit*
954 *Care* **10**, 173 (2006).

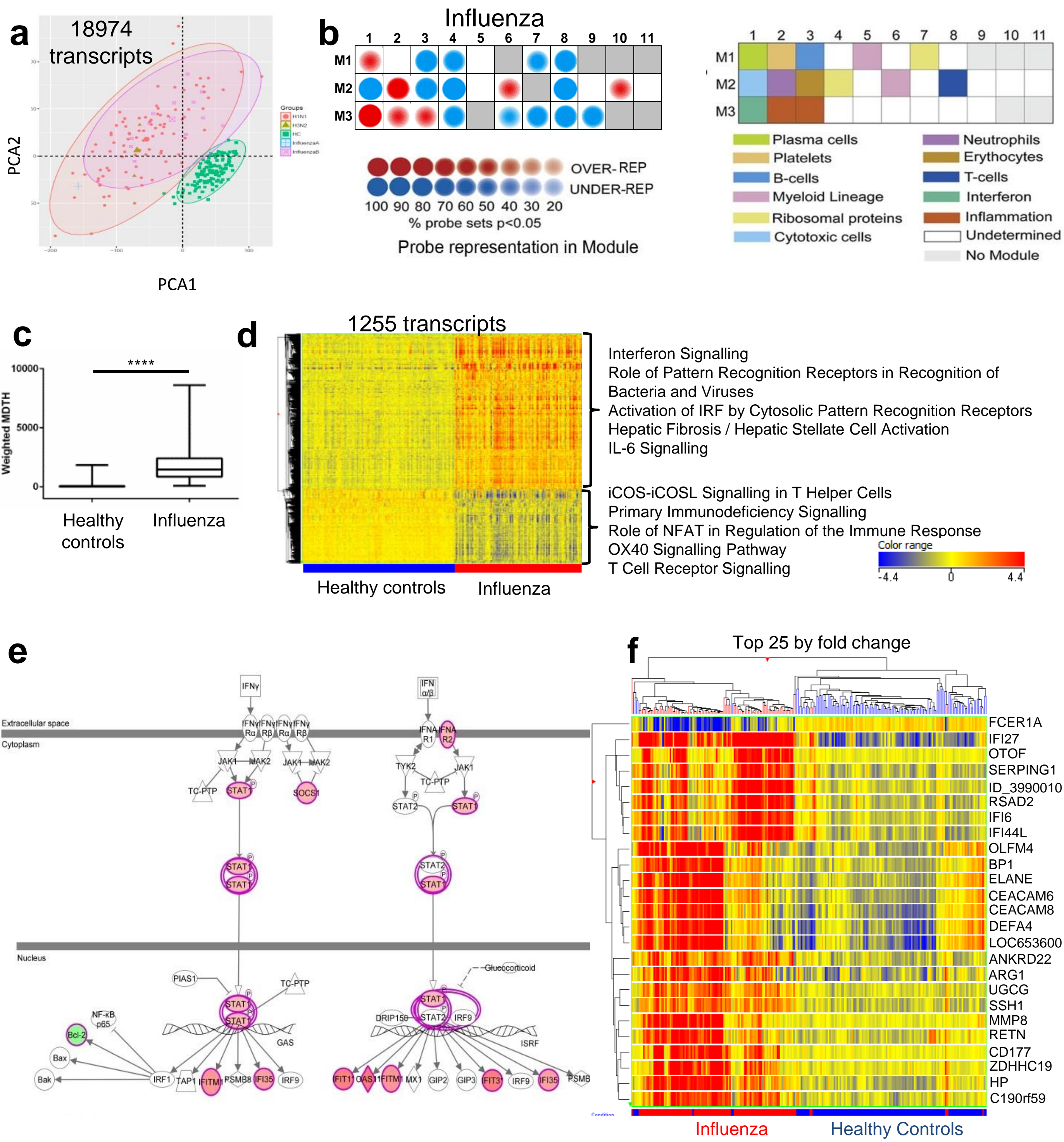
955
956 47. Mayer-Barber, K.D. *et al.* Host-directed therapy of tuberculosis based on interleukin-1 and
957 type I interferon crosstalk. *Nature* **511**, 99-103 (2014).

958
959 48. Nguyen-Van-Tam, J.S. *et al.* Risk factors for hospitalisation and poor outcome with pandemic
960 A/H1N1 influenza: United Kingdom first wave (May-September 2009). *Thorax* **65**, 645-651
961 (2010).

962
963 49. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology
964 Consortium. *Nat Genet* **25**, 25-29 (2000).

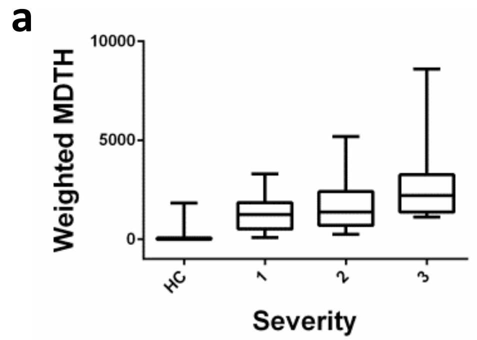
965

966

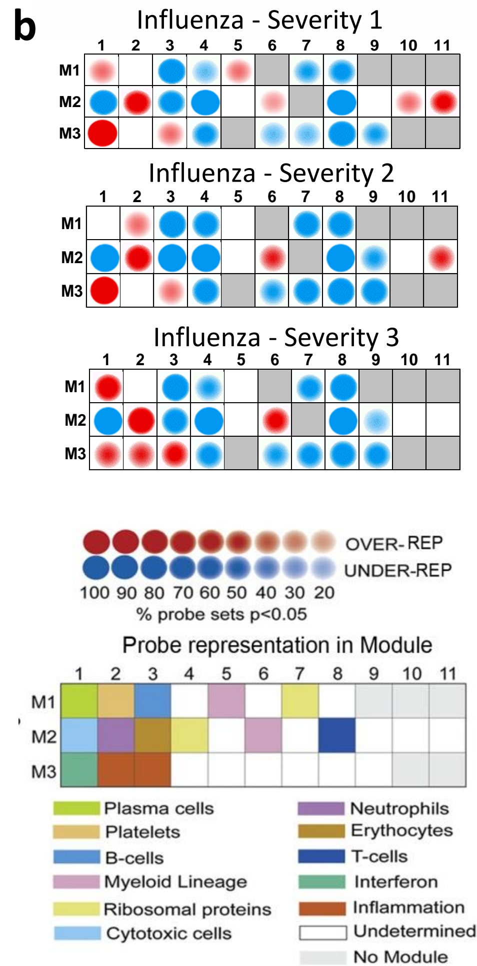


Dunning *et al* **Figure 1**

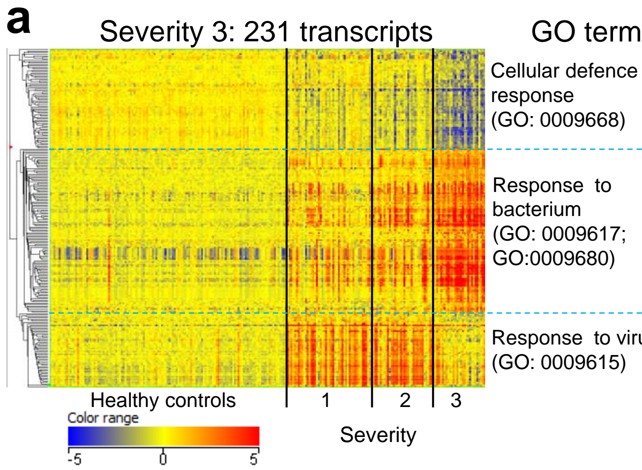
a

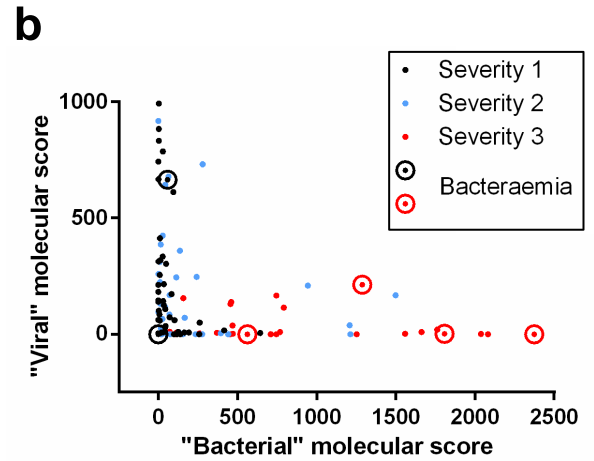


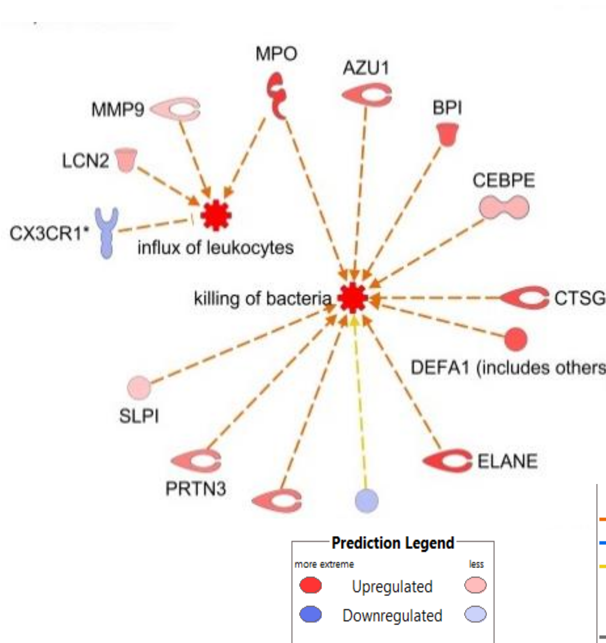
b

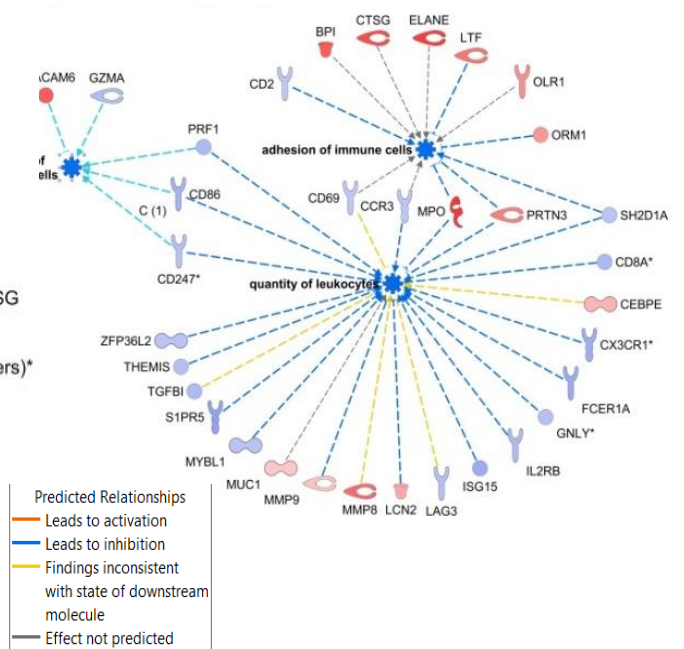


Dunning *et al* **Figure 2**

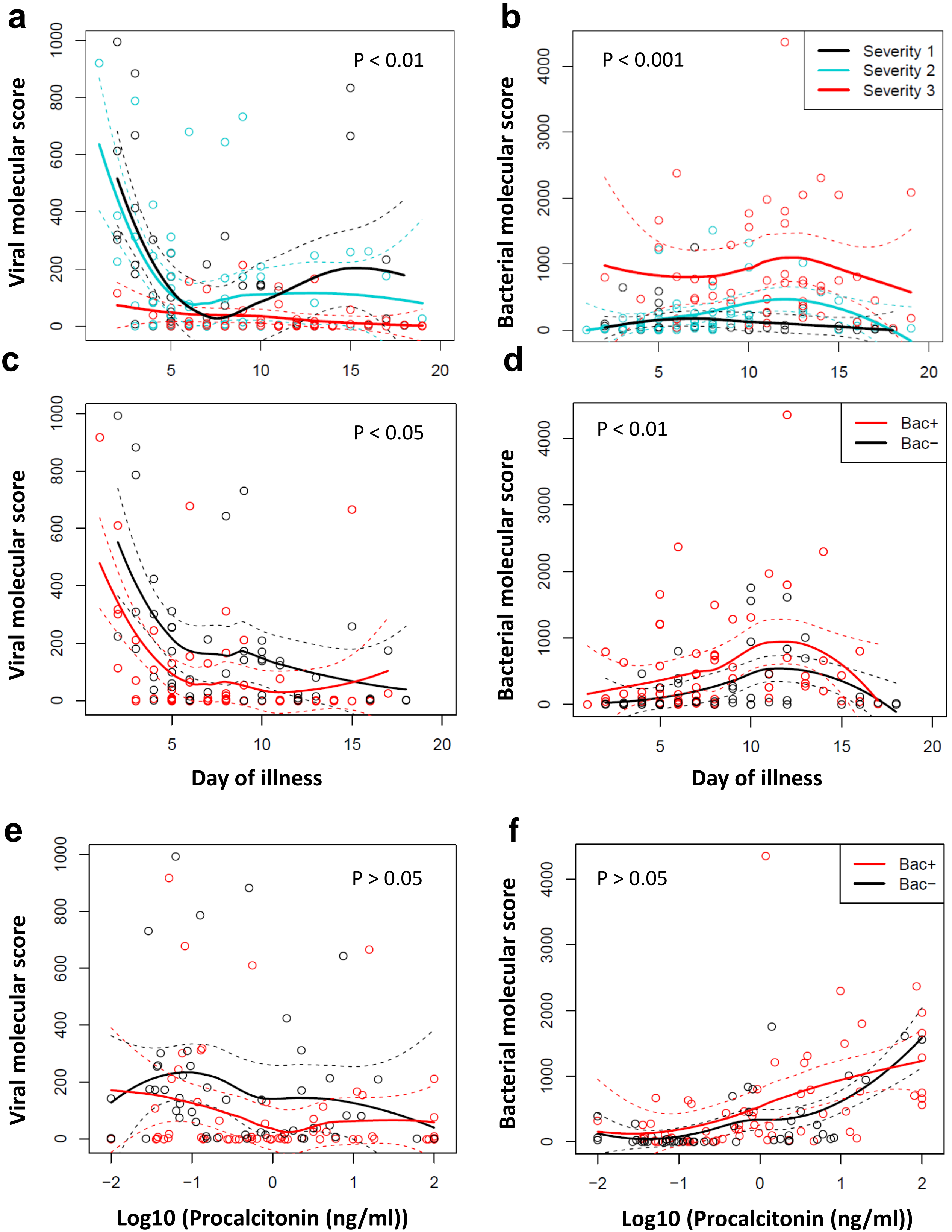




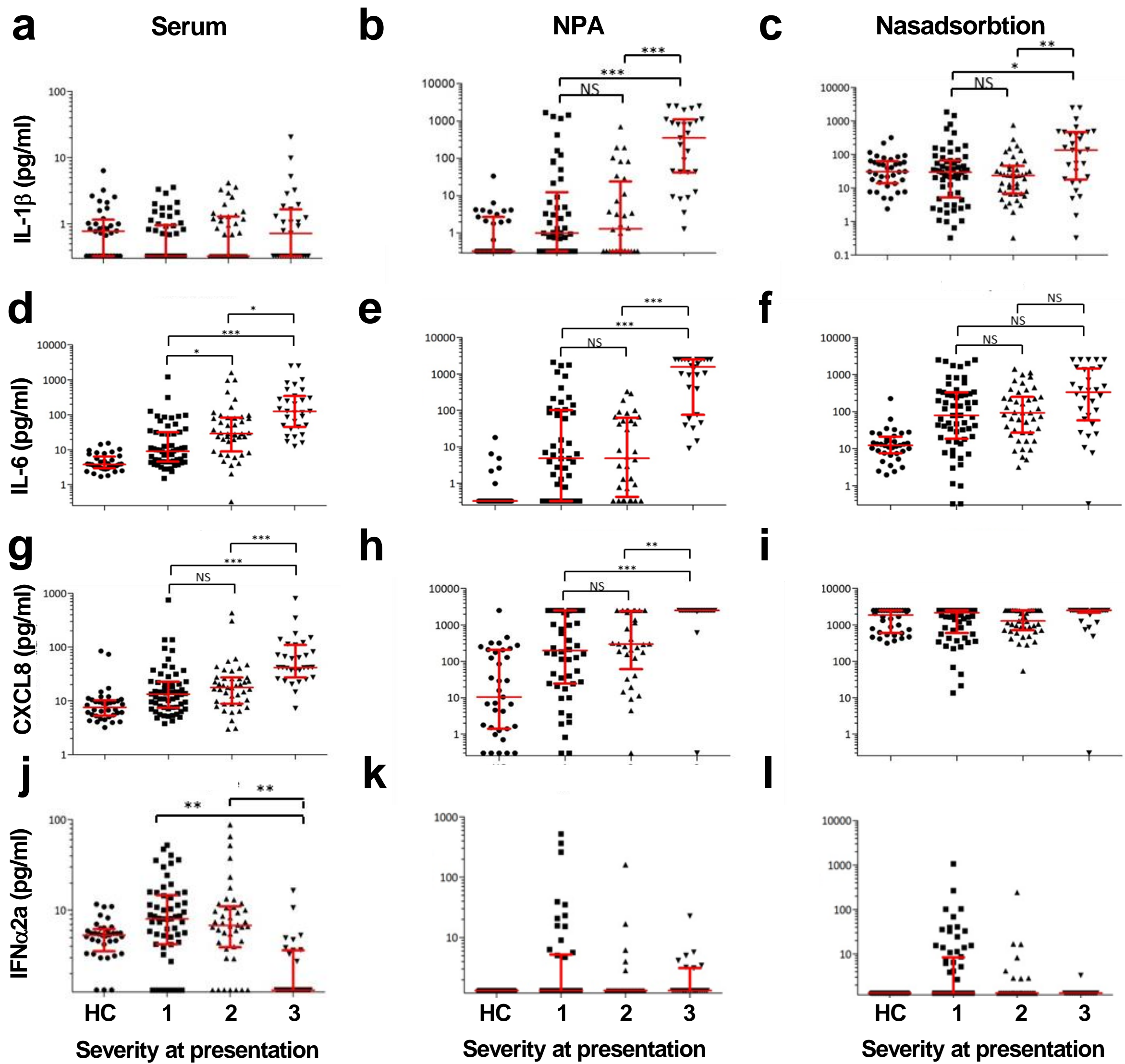




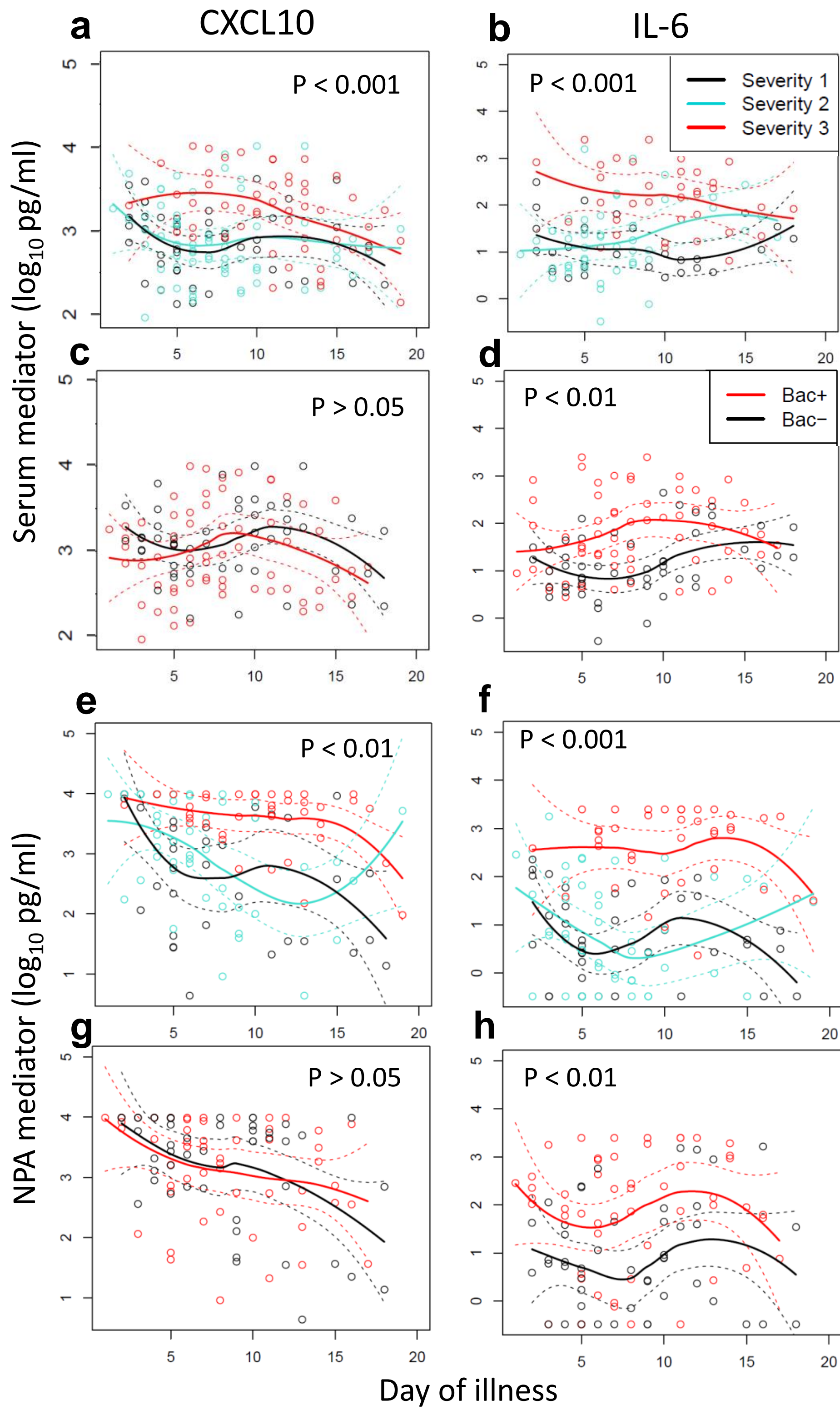
Dunning *et al* **Figure 3**



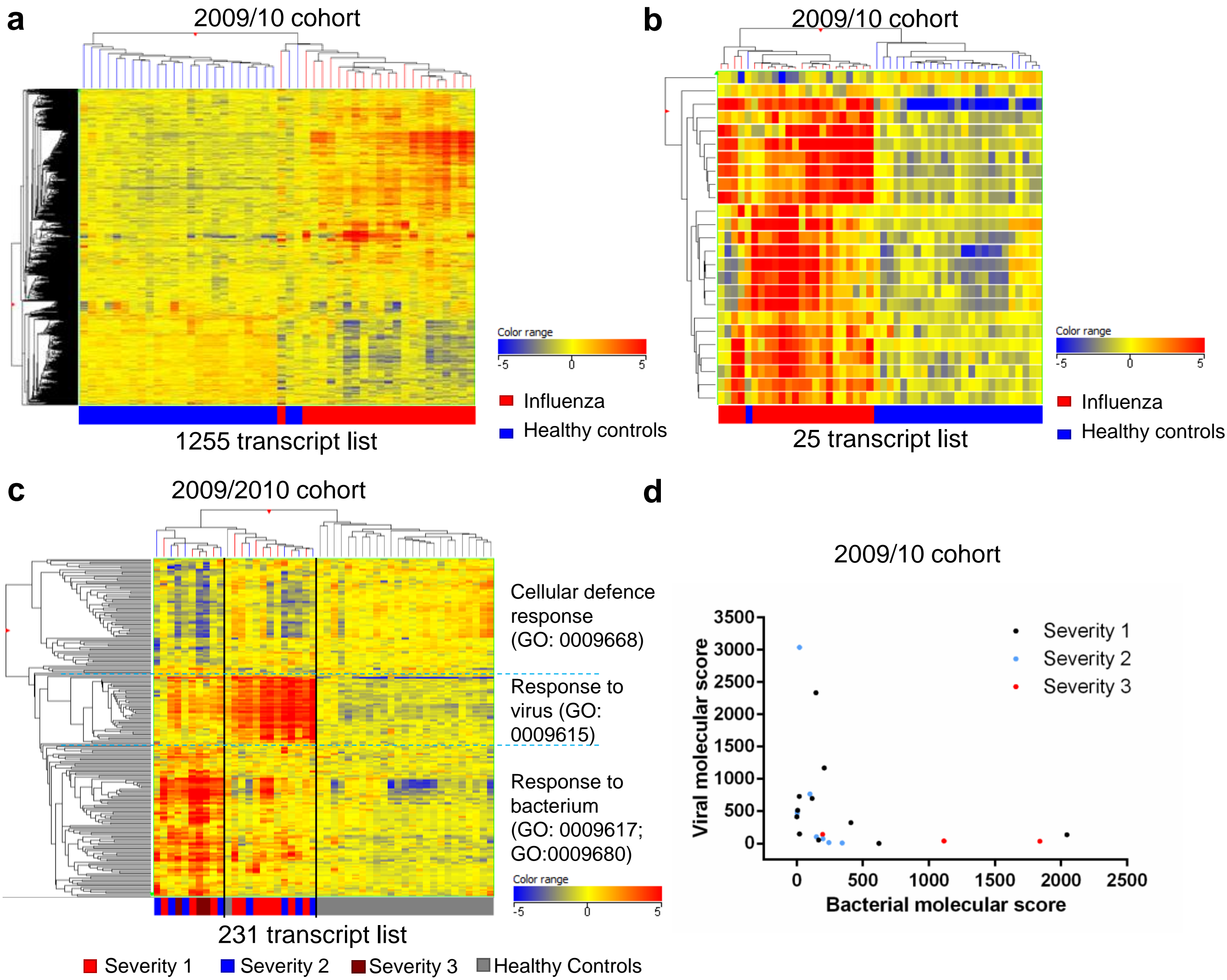
Dunning *et al* **Figure 4**



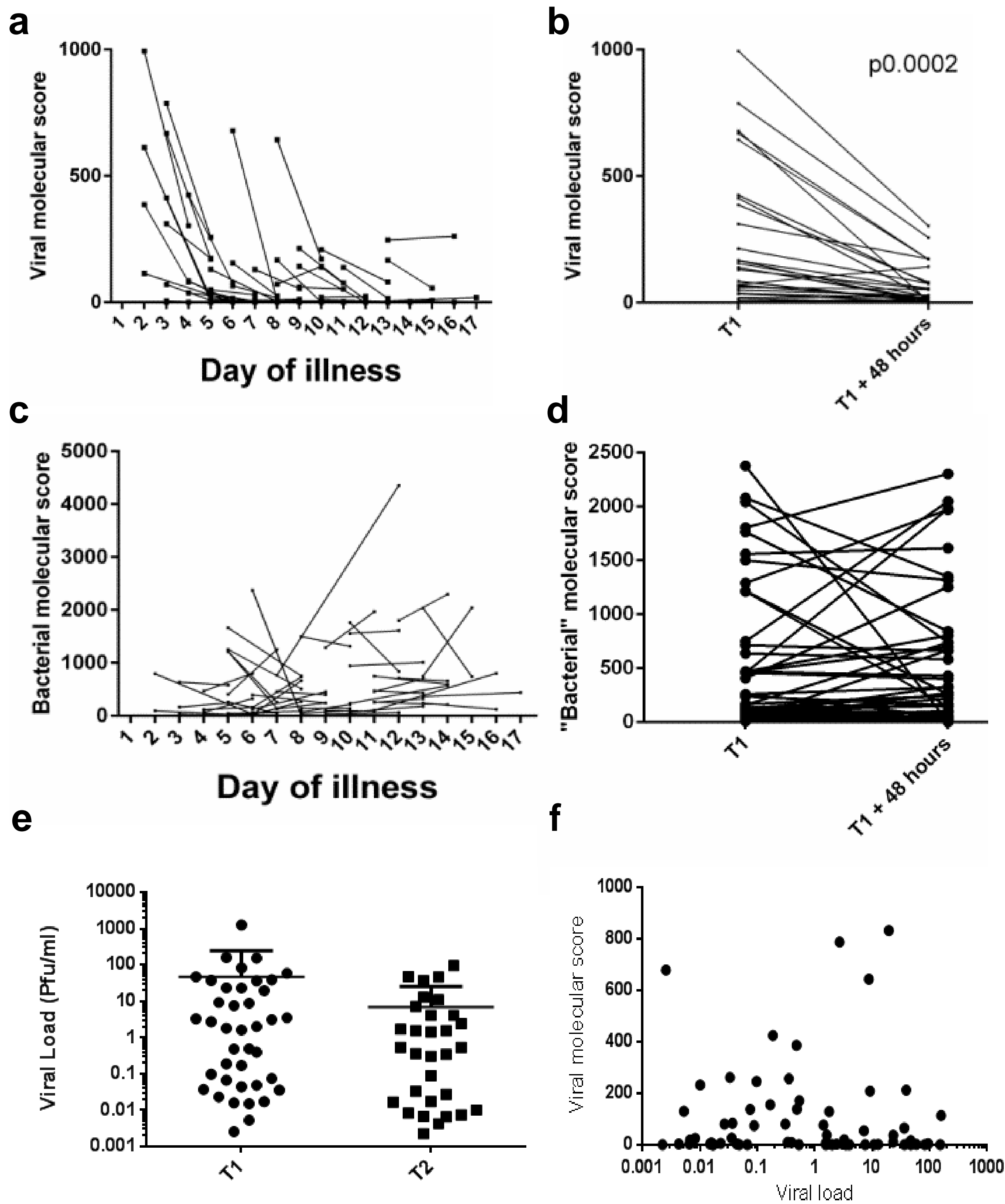
Dunning *et al* **Figure 5**



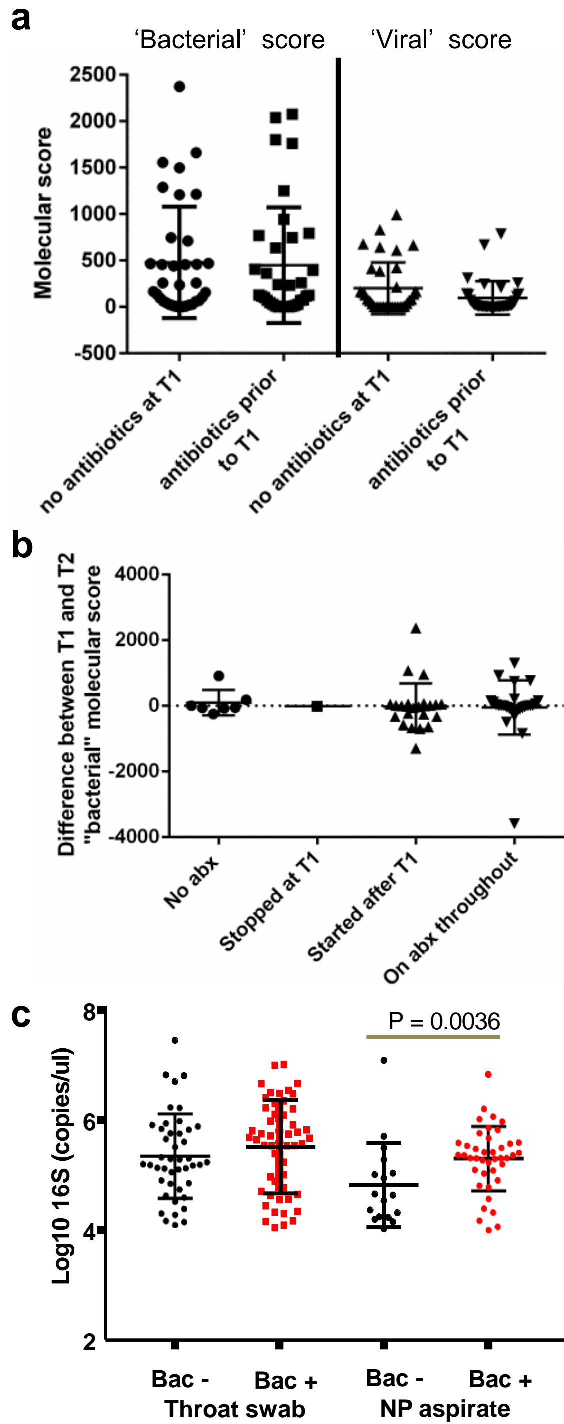
Dunning *et al* **Figure 6**



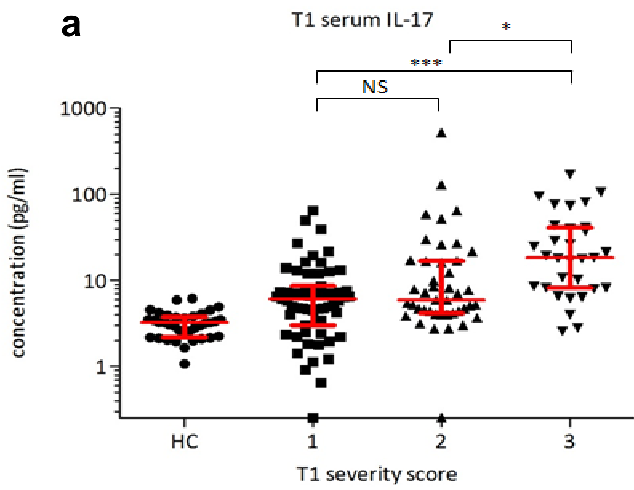
Dunning *et al* **Supplementary Figure 1**

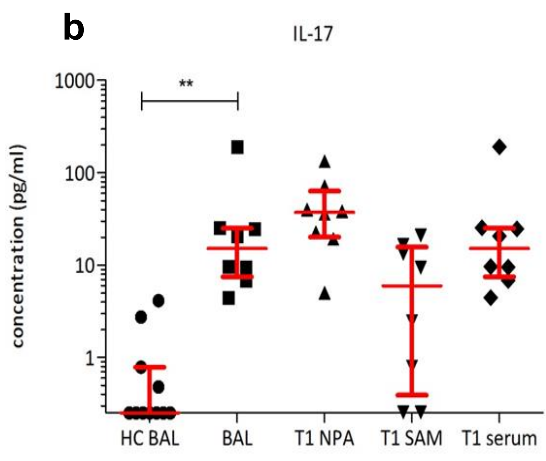


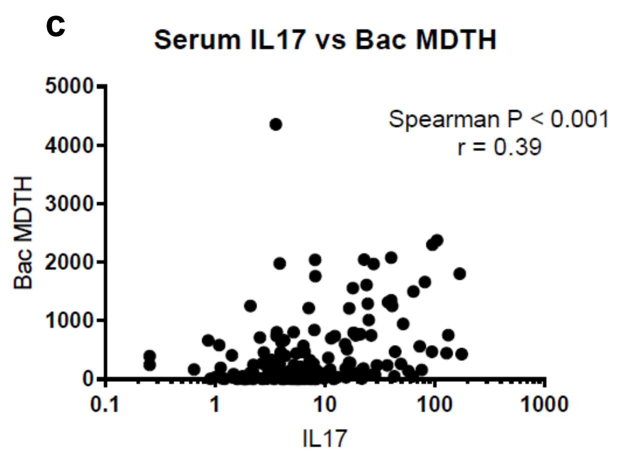
Dunning *et al* **Supplementary Figure 2**

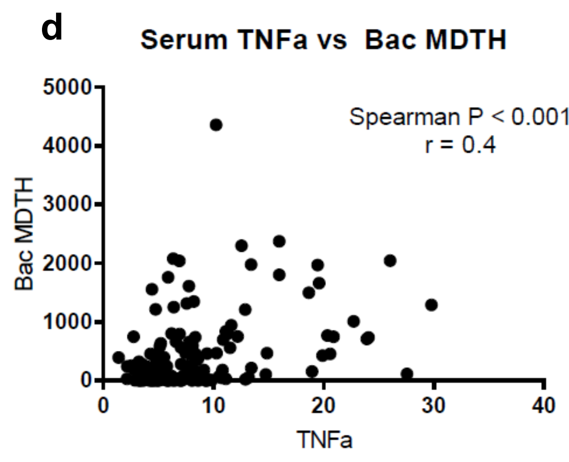


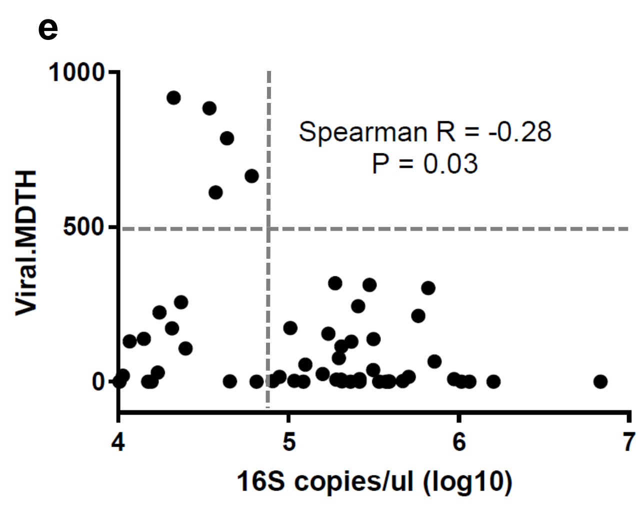
Dunning *et al* **Supplementary Figure 3**

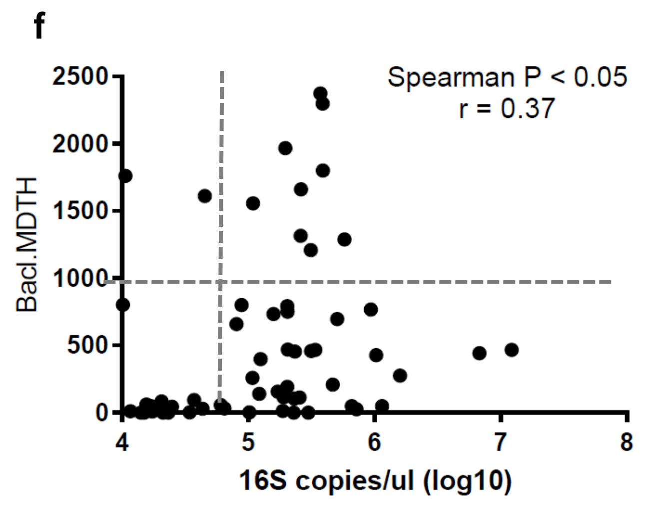












Dunning *et al* **Supplementary Figure 4**

Table 1. Characteristics of recruited patients and healthy controls

	2010/11 Cohort (n=109)	2010/11 Healthy Controls (n=130)	2009/10 Cohort (n=22)	2009/10 Healthy Controls (n=25)
Mean age in years (range)	41 (17-71)	35 (20-68)	44 (23-74)	37 (21-54)
Female (%)	53 (48.6)	75 (57.7)	10 (45.5)	14 (56)
Ethnicity (%)				
White	78 (71.6)	90 (69.2)	10 (45.5)	14 (56)
Black	17 (15.6)	23 (17.7)	5 (22.7)	5 (20)
Asian	9 (8.3)	15 (11.5)	0	6 (24)
Other	5 (4.6)	2 (1.5)	7 (31.8)	0
Comorbidities				
None	28 (25.7)	130 (100)	4 (18.2)	25 (100)
1	31 (28.4)	0	12 (54.5)	0
2	28 (25.7)	0	3 (13.6)	0
≥ 3	22 (20.2)	0	3 (13.6)	0
Women age 15-49y who were pregnant	10/43 (23.3)	1/75 (1.3)	2/8 (25)	0
Influenza type				
pH1N1	94 (86.2)	NA	21 (95.5)	NA
A (H3N2)	2 (1.8)	NA	1 (4.5)	NA
A (unknown)	1 (0.9)	NA	0	NA
B	12 (11)	NA	0	NA
Severity of illness at T1 (%)				
Severity 1	47 (43.1)	NA	11 (50)	NA
Severity 2	34 (31.2)	NA	8 (36.4)	NA
Severity 3	28 (25.7)	NA	3 (13.6)	NA
Peak severity for illness episode (%)				
Severity 1	35 (32.1)	NA	6 (27.3)	NA
Severity 2	44 (40.4)	NA	12 (54.5)	NA
Severity 3	30 (27.5)	NA	4 (18.2)	NA

Note that percentages may not add up to 100 for all variables due to rounding. NA: not applicable.