

Published in final edited form as:

*Nat Immunol.* 2018 June ; 19(6): 625–635. doi:10.1038/s41590-018-0111-5.

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

Jake Dunning<sup>1,4</sup>, Simon Blankley<sup>3</sup>, Long T Hoang<sup>1</sup>, Mike Cox<sup>2</sup>, Christine M Graham<sup>3</sup>, Philip L James<sup>2</sup>, Chloe I Bloom<sup>3</sup>, Damien Chaussabel<sup>5</sup>, Jacques Banchereau<sup>6</sup>, Stephen J. Brett<sup>7</sup>, the MOSAIC Investigators<sup>8</sup>, Miriam F Moffatt<sup>2</sup>, Anne O'Garra<sup>1,3</sup>, and Peter JM Openshaw<sup>1</sup>

<sup>1</sup>Respiratory Sciences, National Heart and Lung Institute, Imperial College London, W2 1PG, UK

<sup>2</sup>Genomic Medicine, National Heart and Lung Institute, Imperial College London, W2 1PG, UK

<sup>3</sup>Laboratory of Immunoregulation & Infection, The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK

<sup>5</sup>Sidra Medical and Research Centre, PO Box 26999, Doha, Qatar

<sup>6</sup>The Jackson Laboratory for Genomic Medicine, 263 Farmington Ave; Farmington, CT 06030, Connecticut, USA

<sup>7</sup>Surgery and Cancer, National Heart and Lung Institute, Imperial College London, W2 1PG, UK

### Abstract

Transcriptional profiles and host response biomarkers are used increasingly to investigate the severity, subtype and pathogenesis of disease. We now describe whole blood mRNA signatures, local and systemic immune mediator concentrations in 131 adults hospitalised with influenza from which extensive clinical and investigational data were obtained by the MOSAIC consortium.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

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

<sup>4</sup>Current address: National Infection Service, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK

<sup>8</sup>MOSAIC investigators:

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

### Author Contributions

Conceived and designed the study: P.J.M.O., A.O'G. and J.D. with input from D.C. and J.B. Performed the microarray experiments: C.M.G. Performed and analysed the microbiome experiments: M.C., P.L.J. and M.F.M. Developed clinical protocols, recruited subjects and collated clinical data: J.D., S.J.B., P.J.M.O. Analysed the microarray data: J.D., C.I.B., L.T.H., and S.B. with supervision by A.O'G. and P.J.M.O. and input from D.C. and J.B. Wrote and revised the manuscript: P.J.M.O., J.D., S.B., L.T.H. and A.O'G.

### Competing Financial Interests

The authors declare no competing financial interests.

Signatures reflecting interferon-related antiviral pathways were common up to day 4 of symptoms in cases not requiring mechanical ventilatory support; in those needing mechanical ventilation an inflammatory, activated neutrophil and cell stress/death ('bacterial') pattern was seen, even early in disease. Identifiable bacterial co-infection was not necessary for this 'bacterial' signature but could enhance its development, while attenuating the early 'viral' signature. Our findings emphasise the importance of timing and severity in the interpretation of host responses to acute viral infection, and identify specific patterns of immune activation that may enable the development of novel diagnostic and therapeutic tools for severe influenza.

## Introduction

Influenza viruses present a continuous threat to global health, mutating and spreading within and between species. It is estimated that one billion cases of human influenza occur worldwide each year, causing 3-5 million cases of severe illness and 300,000 to 500,000 deaths<sup>1</sup>. Infection with pandemic 2009 H1N1 influenza A virus (pH1N1) resulted in generally mild disease<sup>2</sup>, but still caused an estimated 250,000 – 500,000 additional deaths during the first 12 months of global circulation<sup>3</sup>. Whereas seasonal influenza commonly causes severe disease in the old and infirm, serious pH1N1 disease mostly occurred in infants and younger adults, presenting as viral pneumonia and sometimes complicated by multi-organ failure<sup>4, 5</sup>. It has been suggested that severe influenza may in part result from an over-exuberant host reaction to infection (sometimes termed 'cytokine storm'), but is also driven by a high viral load in affected persons<sup>6, 7, 8</sup>.

Although analysis of transcriptional signatures and mediator levels has helped to clarify the pathogenesis of severe influenza, the relationship between severity, timing and complications of infection remains unclear. Previous studies of gene expression patterns in influenza have typically involved small numbers of individuals, healthy subjects undergoing experimental challenge or patients suffering from mild disease<sup>9, 10, 11, 12, 13, 14, 15</sup>. Transcriptomic analysis has also been used to study a variety of acute and chronic infections, including bacterial sepsis, dengue virus infection and tuberculosis<sup>16</sup> and to examine differences and similarities between infectious and non-infectious inflammatory disorders, such as systemic lupus erythematosus<sup>17</sup>.

To further elucidate influenza pathogenesis, the Mechanisms of Severe Acute Influenza Consortium (MOSAIC; <https://goo.gl/kyY2Eu>) recruited 255 hospitalised patients with suspected influenza in England over two consecutive seasons (2009/10 and 2010/11) and 155 healthy adult controls. By analysing biological samples taken at multiple time-points and correlating this with extensive clinical data, MOSAIC aimed to define the contributions made by influenza virus sequence variation, co-pathogens (non-influenza viruses and bacteria) and host factors (genetic and transcriptional differences, soluble mediator responses and cellular immune responses) to disease pathogenesis. Sample analysis resulted in a cumulative total of  $2.1 \times 10^7$  data items on this population, a dataset that we now describe in outline and provide as a resource. To date, MOSAIC has reported enrichment for a host genetic variant, the interferon-inducible transmembrane protein 3 (*IFITM3*) allele SNP rs12252-C in some hospitalised patients with influenza<sup>18</sup>, and that viral sequence

changes that accumulate over time may contribute to the variation in disease severity<sup>19, 20, 21, 22</sup>. The exceptional size and depth of the MOSAIC study provides a unique database to allow these complex issues to be resolved.

We now describe the use of whole blood transcriptional mRNA and soluble mediator data to define associations between individual responses to infection and clinical and laboratory findings in adult MOSAIC patients in whom influenza virus infection was confirmed. Transcriptomic patterns and mediator levels were strongly associated with both severity and duration of illness, indicating a phased and graded activation of interferon-related and inflammatory genes; the effects of clinically evident bacterial co-infection were superimposed on these patterns, which were however primarily related to the duration and severity of influenza.

## Results

### Clinical cohorts

22 and 109 adult patients with laboratory-confirmed influenza were recruited in 2009/10 and 2010/11, respectively. The majority had pH1N1 influenza virus infection (95.5% in 2009/10 and 86.2% in 2010/11). In each cohort, the majority of patients had at least one comorbidity (81.8% in 2009/10 and 74.3% in 2010/11). 13.6% participants in 2009/10 and 25.7% patients in 2010/11 had level 3 severity of illness at the first sampling time-point (Table 1).

### Whole blood transcriptomics

Principal Component Analysis (PCA) of the 18,974 most abundant transcripts from whole blood RNA at enrolment (T1, 2010/11 season, 109 cases) showed a clustering distinct from matched healthy controls (n=130). There was no discernible difference between patients infected with influenza A vs. B viruses (Fig. 1a). Samples from the final time-point (T3, >4 weeks after T1) were similar to those from healthy controls in cases that were clinically resolved, but remained abnormal in patients who remained unwell (data not shown).

Modular analysis<sup>23</sup> of the 2010/11 samples showed abundant transcripts within the interferon-inducible (M3.1) and neutrophil (M2.2) genes relative to healthy controls (Fig. 1b). Transcripts representing plasma cells (M1.1), a subset of myeloid lineage genes (M2.6) and two inflammation modules (M3.2 and M3.3) were also increased, while expression of T (M2.8) and B cell (M1.3) modules (Fig. 1b) decreased. The calculated index termed 'molecular distance to health' (MDTH, derived from analysis of 4526 transcripts significantly detected from background filtered for low expression<sup>24</sup>) increased in most cases of influenza compared to healthy controls ( $P < 0.0001$ ; Fig. 1c), although affected by disease stage and severity (see below). A combination of expression-level and statistical filtering identified 1255 differentially expressed transcripts compared to healthy controls. Supervised hierarchical clustering revealed transcripts that were relatively over- or under-expressed in influenza patients relative to controls (Fig. 1d); applying this same 1255 transcript set to the 2009/10 cohort (22 influenza patients and 25 matched healthy controls) replicated these profiles (Supplementary Fig. 2) indicating that viral variation between the two seasons<sup>22</sup> did not appreciably affect transcriptomic patterns.

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

Hierarchical clustering of the top 25 most significant transcripts in the 2010/11 influenza patients showed two major clusters (Fig. 1e). Transcripts for the IFN-stimulated gene *IFI27* were over-expressed in almost all cases, while *FCER1A* transcription usually decreased. Independent analysis of the 25 transcripts from the 2009/10 dataset showed similar clustering (Supplementary Fig. 2b). Patients with type I IFN-induced gene activation (e.g. *RSAD2*, *IFI6*, *IFI44L*) typically did not express neutrophil-associated and bacterial response-associated transcripts (e.g. *DEFA4*, *ELANE*, *MMP8*) and *vice versa* (Fig. 1e and Supplementary Fig. 2b). Taken together, these results show that patients with acute influenza show activation of whole blood transcriptomic pathways indicating responses to Type 1 and Type 2 interferon, inflammatory markers and possibly combined with the effects of depletion of some cell types from the blood.

### Transcriptomics and disease severity

The 2010/11 cases were grouped according to their severity of illness at T1 using a three-point scale: Severity 1: no supplemental oxygen required; 2: oxygen by mask; 3: mechanical ventilation. Relative to healthy controls, transcriptomic abnormality (expressed as mean MDTH) was increased in severity 1-2 illness but was further increased in severity 3 cases (4526 transcripts; Fig. 2a). Using modular analysis<sup>23</sup>, there was an over-abundance of plasma cell (M1.1), neutrophils (M2.2), and myeloid lineage (M2.6) transcripts in influenza-infected patients, most evident in those with the greatest severity. Severity 3 cases also showed an abundance of transcripts in inflammation modules M3.2 and M3.3. By contrast, IFN-related transcripts (M3.1) were most evident in cases with severity 1 or 2 (Fig. 2b). Therefore, patients with the most severe illness have transcriptomic patterns that differ from those with less severe illness, with an increased abundance of inflammation-related transcripts and a decrease in IFN-related transcripts.

Semi-supervised hierarchical clustering of 231 transcripts with a more than two-fold change between severity 1-2 and severity 3 showed that transcripts associated with GO term 'response to virus' were typical of cases with milder disease (Fig. 3a; Supplementary Table 1), whereas patients needing mechanical ventilation showed a marked abundance of 'response to bacterium' transcripts (Supplementary Table 2). Severe cases typically showed a relative under-abundance of transcripts associated with 'cellular defence response' (Fig. 3a).

The same 231 transcript list was verified by hierarchical clustering analysis of the 2009/10 cohort. Influenza patients severity 1 or 2 were again characterised by ‘response to virus’ transcripts, whereas three cases with severity 3 disease instead showed ‘response to bacterium’ transcripts (Supplementary Fig. 2c). The relationship between total molecular score for the 51 GO Term defined ‘response to virus’ (Supplementary Table 1) and the 112 ‘response to bacterium’ (Supplementary Table 2) transcripts from the 2010/11 patients at T1 (n=109) are shown in Fig. 3b. Influenza patients with high ‘viral’ responses (> 500) were exclusively from the severity 1 or 2 groups, whereas most patients with high ‘bacterial’ scores (> 500) had severity 3 illness and had low ‘viral’ scores (reflecting the modular analysis). However, a few severity 1 or 2 patients had low ‘viral’ molecular scores and moderately high ‘bacterial’ molecular scores. Removal of 6 patients with known bacterial co-infection did not eliminate this subgroup. Similar findings were observed in the 2009/10 cohort (Supplementary Fig. 2d).

Reciprocal expression of activated and repressed biofunctions of identified genes was observed in severity 3 patients, compared to severity 1 or 2 patients (Fig. 3c, 3d). Nine genes associated with neutrophil activation were upregulated (e.g. *MPO*, *DEFA1*, and *ELANE*) along with 3 associated with leukocyte influx (*MPO*, *MMP9* and *LCN2*; Fig. 3c). The repressed biofunctions in severity 3 patients were ‘activation of cytotoxic T cells’, ‘adhesion of immune cells’ and ‘quantity of leukocytes’ (Fig. 3d). These results show that patients with the most severe disease show upregulation of genes associated with neutrophil activation and leukocyte influx, not seen in those with less severe influenza.

### Effect of illness duration, severity and viral load on transcriptomic patterns

Patients with symptoms of up to 4 days’ duration at the time of sampling typically had elevated ‘viral’ molecular scores, but not if they required mechanical ventilation (Severity 3); in such cases, the ‘viral’ score was low, even early in the disease (Fig. 4a). Patients with severity 3 illness showed higher ‘bacterial’ molecular scores than patients with less severe disease even at first presentation, whereas ‘bacterial’ molecular scores were low in patients with severity 1 or 2 illness regardless of the time of sampling (Fig. 4b).

In patients in the 2010/11 cohort with repeat samples (T1 and T2 separated by 2-5 days; n=59), the ‘viral’ molecular score usually (but not always) decreased between T1 and T2 (Supplementary Fig. 3a). In cases where T2 samples were 2 days after T1 (n=41), the reduction in ‘viral’ score was significant ( $P=0.0002$ ; Supplementary Fig. 3b). Changes in ‘bacterial’ molecular scores between T1 and T2 were more variable (Supplementary Fig. 3c and d). A decline in viral load (measured in mucus obtained by nasopharyngeal suction) was observed between T1 and T2 (Supplementary Fig. 3e), but there was no clear relationship between viral load and ‘viral’ transcriptomic score (Supplementary Fig. 3f). Taken together, the results show that the relative dominance of ‘viral’ or ‘bacterial’ transcriptomic responses is influenced by both severity of illness and duration of illness.

### Effect of bacterial infection and carriage on transcriptomic patterns

To examine the role of bacterial infection in driving ‘bacterial’ GO terms, we identified a subgroup of influenza-infected patients that had been thoroughly investigated for bacterial

infection by testing the T1 NPA and throat swabs by culture and T1 nasopharyngeal aspirate by PCR detection of bacterial pathogens (in addition to blood culture and urine pneumococcal antigen testing in most cases). Incomplete bacteriological sampling excluded 36/109 (33%) patients; 34 (47%) patients provided at least four out of five sample-types. Of the 73 cases with adequate samples, 39 (53%) were deemed to have potentially pathogenic bacteria detected in at least one sample type and were classified by an expert clinical review panel to have significant bacterial co-infections.

Comparing those cases of influenza in which significant bacterial co-infection was identified (Bac<sup>+</sup>) with those in whom no bacterial infection was found despite adequate investigation (Bac<sup>-</sup>), the average 'viral' molecular score was lower in those with bacterial infection at all times up to day 12 after illness onset (Fig. 4c), and the average 'bacterial' score was greater in those with bacterial co-infection between day 3 and 14 (Fig. 4d). However, the transcriptomic scores showed similar time trends regardless of the presence or absence of bacterial co-infection. Similar findings were observed when stricter exclusion criteria were applied to the subgroup analysis, excluding patients from the 'bacteria not detected' group if they had not provided all five sample-types (data not shown). In this case, statistical analysis could not be performed due to the low sample size (only 13 patients provided all five sample types and did not have bacteria detected).

To examine the influence of treatment of bacterial infection on the observed 'viral' and 'bacterial' responses, we stratified T1 and T2 'bacterial' and 'viral' scores in 2010/11 influenza patients according to antibiotic prescription. In the MOSAIC study, 92% (234/255) of cases were treated with antibiotics at some time. Antibiotics prior to T1 had no demonstrable effect on transcriptomic patterns (Supplementary Fig. 4a). Comparing influenza patients who were not given antibiotics (n=7) with those given sustained antibiotic treatment after T1 (n=24) or throughout illness (including T1 and T2; n=27), there was no discernible or statistically significant effect of antibiotic administration on the 'bacterial' molecular scores (Supplementary Fig. 4b).

We next examined the levels of the 16S rRNA gene (bacterial load) in the throat swab and NPA samples from cases that were classified as 'bacterial co-infection' or 'viral infection without bacterial infection'. The levels of the 16S rRNA genes were no different between these groups on throat swabs, but the NPA bacterial load was greater in those cases with confirmed bacterial co-infection (Supplementary Fig. 4c).

Finally, we investigated the utility of procalcitonin (PCT) as a possible guide to the presence of significant bacterial infection<sup>25, 26, 27</sup>. Procalcitonin levels showed no relationship to 'viral' molecular score (Fig. 4e), and there was no correlation between 'viral' molecular scores at T1 and T2 and PCT levels measured at respective time-points (data not shown). However, 'bacterial' molecular scores tended to be raised in those cases with the most severe disease and the highest PCT levels (Pearson  $r = 0.44$ ,  $P < 0.001$ ) regardless of the presence or absence of significant detectable bacteria (Fig. 4f).

In summary, 'viral' molecular scores were seen in disease of up to 5 days duration. Even during this early phase, cases needing mechanical ventilation had low 'viral' scores and this

was especially true in those with clinically determined bacterial co-infection. On the other hand, expression of ‘bacterial’ response genes was seen only in the most severe cases of influenza; significant bacterial infection enhances this signal, but the ‘bacterial’ score was evident in those with severe influenza regardless of the presence of bacterial co-infection (especially if the disease had lasted a week or more).

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

An advantage of the MOSAIC study is that it provides linked data of whole blood transcriptomic signatures and, for example, levels of 35 soluble mediators in the blood, NPA and anterior nasal fluid (‘nasadsorption’ samples using synthetic adsorptive matrices) at up to 3 time points.

The observed changes depended on the mediator and compartment. For example, serum concentrations of IL-1 $\beta$  showed no trend against severity (Fig. 5a), but NPA or nasadsorption IL-1 $\beta$  levels increased in those with severe disease (Fig. 5b, c). By contrast, serum IL-6 levels increased with severity (Fig. 5d); in the NPA, IL-6 was undetectable in most of the healthy controls but was detected in most of the cases of influenza and increased in with severity (Fig. 5e). The levels of IL-6 in nasadsorption samples were more consistent than in the NPA being measureable in normal subjects, raised in most cases of influenza but not reflecting severity (Fig. 5f).

CXCL8 serum levels tended to be higher in cases of influenza than in controls, again increasing with severity (Fig. 5g); in NPA, CXCL8 levels were variable but generally increased with influenza severity and tended to saturate the assay (Fig. 5h). Nasadsorption CXCL8 levels were even higher, often saturating the assay even in some healthy controls (Fig. 5i). IFN- $\alpha$ 2a was measurable only in a proportion of individuals but was raised in serum in milder (severity 1 or 2) cases but not in severe disease (Fig. 5j). In NPA or nasadsorption samples, IFN- $\alpha$ 2a levels were raised in some milder cases (not statistically significant, Fig. 5k, 5l).

Serum concentrations of IL-17 increased with severity at T1 (Supplementary Fig. 5a) and were elevated in the BAL of 8 patients (in whom samples were taken for clinical indications) relative to experimental healthy controls (Supplementary Fig. 5b). In addition, we found a significant positive correlation of serum IL-17 (Supplementary Fig. 5c) and TNF $\alpha$  with MDTH (Supplementary Fig. 4d).

Regarding the effects of timing, CXCL10, IL-6 and CCL2 were elevated in serum from severe cases of influenza, especially between days 5 and 10 (Fig. 6a, 6b and not depicted). Proven bacterial co-infection had no evident additional effect on CXCL10 (Fig. 6c), but serum IL-6 was abundant not only in severe cases (even early in disease, Fig 6b) but especially in cases of bacterial co-infection (especially between days 5 and 10, Fig. 6d).

In the NPA, most mediators (e.g. CXCL10, IL-6, CCL2 and CXCL8) were markedly increased in severe disease and especially after day 4 (e.g. Figs. 6e and 6f, and not depicted). NPA CXCL10 was again unaffected by confirmed bacterial disease (Fig. 6g) whereas levels of IL-6 (and CCL2 and CXCL8, data not shown) were particularly increased in patients with

bacterial co-infection (Fig. 6h). In the nasalsorption samples mediator concentrations declined slowly with time even in less severe disease; CXCL10 was depressed by known bacterial co-infection but IL-6, CCL2 and CXCL8 concentrations at this site were unaffected by severity or bacterial status (not depicted; data online).

Since ‘bacterial load’ (as measured as NPA 16S rRNA copy number) was raised in cases of significant bacterial infection (Supplementary Fig. 4c), we regressed this parameter against viral or bacterial MDTH. High values of viral MDTH only occurred in those with lower bacterial loads in the NPA (Supplementary Fig. 5e), and high values of bacterial MDTH were seen only in a subset of those with higher 16S bacterial load in the NPA (Supplementary Fig. 5f).

These data are consistent with bacterial load in the inflamed respiratory tract playing a part in driving levels of soluble mediators in mucosal fluids and serum as well as transcriptomic signatures in the blood; however, the effects of influenza severity and time after disease onset remain the dominant determinants of host responses.

## Discussion

The MOSAIC study is exceptional in including a large number of well-characterised hospitalised patients with influenza, studied prospectively and sampled intensively. We found that whole-blood RNA expression profiles of patients hospitalised with influenza evolve over time and that the pattern reflects severity. Patients with mild (or early) disease typically showed responses dominated by interferon-inducible genes and type 1 interferon, but this ‘viral’ signature was replaced during severe (or late) disease by a pattern reflecting inflammation and neutrophil activation, more typically associated with the GO term ‘response to bacteria’, including genes encoding regulators of apoptosis and anaerobic metabolism<sup>28</sup>. The ‘viral’ response was rarely seen in patients beyond day 4, whereas the inflammation/neutrophil activation signal peaked during the second week.

In severe disease, the early ‘viral’ response was typically absent whereas the ‘bacterial’ signature was present at study enrolment; this was especially so in cases with proven bacterial co-infection, but did not depend on it. In addition, the bacterial load in the nasopharynx (quantified by 16S copy number) tended to be low if the ‘viral’ signature was evident, and high in those cases in which an inflammatory/cell activation pattern was seen.

Serum and nasopharyngeal soluble protein mediators were generally abundant in cases of severe disease, even early after onset. Inflammatory mediators (e.g. IL-17, IL-1 $\beta$  and IL-6) were augmented in those with clinically significant bacterial co-infections, whereas interferon  $\alpha$  levels tended to be low or undetectable in most compartments in those with very severe influenza; however, interferon-related secondary mediators (e.g. CXCL10 in serum) were generally most abundant in severe cases. These findings suggest complex interactions between viral and bacterial sensing and response mechanisms, evolving over time.

To examine the issue of bacterial co-infection specifically, we identified cases in which pathogenic bacteria were found in mucosal samples or blood culture as a subgroup with



clinically confirmed bacterial sepsis. Three of these six cases needed mechanical ventilation and had markedly elevated 'bacterial' signature without any increase in 'viral' score; one patient had elevated 'bacterial' and 'viral' scores. The remaining two cases with bacteraemia did not have marked elevations in their 'bacterial' scores; both had mild (Severity 1) disease. We next used stringent criteria to identify influenza cases that were extensively investigated for bacteria and found not to be co-infected, comparing them to cases in which pathogenic bacteria were identified with certainty. Patients with confirmed bacterial co-infection had higher 'bacterial' molecular scores overall, but progression of the transcriptomic signatures was similar over time. Therefore, severe influenza virus infection alone appears able to drive the 'bacterial' signature but this response is enhanced by bacterial co-infection. We conclude that transcriptomic data from the blood are an unreliable guide to the presence or absence of bacterial co-infection, but need careful interpretation in relation to timing and severity of disease. We cannot determine the extent to which these changes might be driven by injury caused by influenza virus, or by innate sensitivity to resident microbiota leading to activation of Th17 pathways triggered by endotoxins from mucosal surfaces<sup>29</sup>.

In animal models of viral lung disease, dysregulated host immune responses<sup>30</sup> and interferon production<sup>31</sup> can lead to complex inflammatory responses which contribute to pathogenesis<sup>32, 33</sup>. In macaques, administration of recombinant IFN $\alpha$ 2a initially up-regulates the expression of antiviral genes and prevents viral infection but continued treatment causes desensitization and a paradoxical decrease in antiviral gene expression<sup>34</sup>. These paradoxical immunosuppressive effects may impede viral control<sup>35</sup> or trigger inflammation and tissue damage<sup>31</sup>. In mice, influenza causes an early local influx of neutrophils followed by a virus specific CD8<sup>+</sup> T-cell response<sup>36, 37, 38</sup>. Neutrophils might facilitate the development of this antigen-specific response by guiding influenza-specific CD8<sup>+</sup> T cells into sites of infection by laying chemokine trails<sup>39</sup>. Our findings in human influenza are generally compatible with these animal studies.

We here present only selected results of an extended study of soluble immune mediator data from the MOSAIC cohort. Our main findings were of decreased IFN $\alpha$ 2a and increased IL-1 $\beta$ , IL-6 and CXCL8 levels in the nasal and/or serum compartments in patients with severe disease. This apparent reciprocity may relate to the known cross-regulatory functions of IL-1 and type I IFNs in experimental models<sup>28, 40</sup>. Our results generally fit with the proposal that levels of mediators such IL-1 $\beta$ , IL-6 and IL-17 are influenced by bacterial co-infection in severe influenza, but not driven by it. However, there are many additional possible analyses to be performed. We chose only to illustrate those most relevant to the transcriptomic analysis and the question of bacterial superinfection. We invite readers to explore additional correlations using our online data as a resource, and will welcome discussion with respect to additional interpretations.

Our study has important limitations. Despite its ambition, scope and intensity we had limited numbers of repeat samples from individual patients. Our description of trends over time depends largely on summative data and on subjective reporting of the time of onset of disease. Ideally, our findings need validation in other time-series studies of simple and complicated acute viral disease with frequent sampling at multiple sites. We were unable to study the early or preclinical phases but were limited to investigation of symptomatic cases

presenting with disease of sufficient severity to reach hospital. Ongoing studies of experimental infection with pH1N1 in volunteers should allow us to overcome some of these limitations.

In summary, virus-induced type I interferon-related pathways are activated during the first four days of symptomatic influenza in hospitalised patients. These 'viral' pathways are then down-regulated, to be replaced by inflammatory, activated neutrophil and apoptosis-related pathways associated with IL-17 abundance, host-mediated tissue damage and expression of 'response to bacteria' gene clusters, particularly in cases with a high 16S bacterial load in the nasopharyngeal secretions. In severe cases the 'viral' response was depressed even early in disease, accompanied by an increase in IL-1 $\beta$  and IL-17. These findings emphasise that the stage and severity of disease need to be taken into account in interpreting host responses to infection and in the development of potential diagnostic tests to differentiate between possible causes and appropriate therapies.

## Online Methods

### Study population and inclusion criteria

Patients 16 years of age were recruited during two successive winters (01 December 2009 to 03 March 2011). Patients with suspected influenza were identified by medical or nursing staff, or notified to investigators by hospital diagnostic laboratories. Patients in London were recruited from four Imperial College Healthcare NHS Trust hospitals, the Chelsea and Westminster Hospital, and the intensive care unit at the Royal Brompton Hospital (a national referral centre for severe respiratory failure). In Liverpool, patients were recruited from the Royal Liverpool, Liverpool Women's and Arrowe Park Hospitals. Patients were included irrespective of prior or concurrent comorbidity (most commonly asthma, pregnancy, immunocompromising conditions, or co-infection with other respiratory pathogens), to reflect the populations known to be at greatest risk of severe influenza. Adult healthy controls were recruited and matched to the patient cohorts for age, sex and ethnicity and were screened to exclude known illnesses or current use of medications (Registered Clinical Trial NCT00965354).

### Research Ethics Committees' Approvals

The study was approved by the NHS National Research Ethics Service, Outer West London REC (09/H0709/52, 09/MRE00/67). Patients or their legally authorised representatives provided informed consent. Additional adult healthy controls were recruited as part of a separate study and consented to their samples being used in additional studies (Central London 3 Research Ethics Committee, 09/H0716/41). Informed consent was obtained from all participants and we complied with all relevant ethical regulations.

### Biological sampling

Research samples were obtained at three time points: T1 (recruitment); T2 (approximately 48h after T1); T3 (at least 4 weeks after T1). Only T1 and T2 samples were included in this report. Whole blood samples for transcriptomics were collected during the two recruitment periods, 2009/10 and 2010/11. Of 85 MOSAIC participants presenting with influenza-like

illness in 2009/10, 23 (27%) were adults with confirmed influenza, and T1 transcriptomic samples were available from 22 adults. Of 171 MOSAIC participants presenting with influenza-like illness in 2010/11, 111 (65%) were adults with confirmed influenza, and T1 transcriptomics samples were available from 109/111 (98%). RNA extraction and microarray were successful for all available patient samples from both cohorts. Microarrays were also performed on samples from adult healthy controls of similar age, sex and ethnicity to the study patients (Table 1). One healthy control sample for the 2009/10 cohort was not included in final analysis because it failed quality control assessments.

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

### **Influenza virus infection status**

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

### **Influenza virus quantification**

Nasopharyngeal secretions were collected into sterile universal sputum traps by suction catheterisation. After 5 seconds of suctioning, any contents remaining within the catheter were flushed through with 5 ml normal saline. Samples were stored at -80°C until analysis. Viral nucleic acid was extracted using the Qiagen MDx Biorobot automated extractor with the QIAamp Virus MDx Kit according to the manufacturer's instructions. qRT-PCR reactions were set up to a total volume of 15 µl using the Qiagen One-Step RT-PCR kit, using primers (influenza A matrix [M] or pH1N1 neuraminidase [NA]) as described previously<sup>41</sup> on an ABI Prism 7500 SDS real time platform (Applied Biosystems). For viral load quantitation, we first derived the crossing threshold (Ct) value (at the inflexion spot of the sigmoid amplification curve to capture the point at which DNA amplification is exponential) performed in a batched assay as a relative expression of viral burden against each sample. Subsequently, this was measured against a standard curve of Ct value to plaque-forming units (pfu)/ml, generated by measuring pfu when MDCK cells were inoculated with a known amount of pH1N1.

## Clinical data collection and severity of illness scoring

Clinical data were extracted from hospital case notes and recorded in the Flu-CIN data collection tool<sup>42</sup> by trained researchers. Prescription charts were examined to determine whether antibiotics were being administered before, during or after sampling time points.

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

## Detection of bacteria

Nasopharyngeal aspirates and swabs collected at T1 underwent microscopy and culture for bacteria. Additionally, multiplex PCR was performed to detect the following common respiratory bacteria in these samples: *Staphylococcus aureus*, *Chlamydia pneumoniae*, *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Pneumocystis pneumoniae*, *Legionella* species, *Klebsiella pneumoniae*, *Salmonella* species, *Moraxella catarrhalis*, *Mycoplasma pneumoniae*, and *Bordetella pertussis*. Throat swab samples taken at T1 also underwent culture and microscopy. When available, urine samples collected between T1 and T2 underwent pneumococcal antigen testing (BinaxNow, Allere, Stockport, UK). Clinical microbiology data were obtained from hospital laboratory databases, including results of blood cultures (when taken 48 hours either side of T1) and urinary pneumococcal antigen results (for patients who did not have a researcher-requested urinary antigen sample). An independent microbiologist assessed the significance and validity of positive blood culture results, in an attempt to exclude cases of pseudobacteremia caused by commensal contamination.

## Soluble immune mediators

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

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

### Blood procalcitonin assay

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

### 16S rRNA gene bacterial load measurement

The 16S rRNA gene was targeted with 0.3  $\mu$ l each of 10  $\mu$ M universal primers 520F 5'-AYT GGG YDT AAA GNG and 802R 5'-TAC NVG GGT ATC TAA TCC added to 7.5  $\mu$ l of SYBR Fast qPCR Kit Master Mix (KapaBio) and 5  $\mu$ l of a 1:5 dilution of sample DNA extract and 1.9  $\mu$ l of PCR Clean water (Mobio). Reactions were prepared in triplicate and thermal cycling carried out on a VIIA-7 Real-Time PCR System. Thermal-cycling conditions were 90°C for 3 mins, then 40 cycles of 95 °C for 20 s, 50 °C for 30 s, 72 °C for 30 s with default melt conditions. A standard curve of cloned (TOPO TA, Invitrogen) full length *Vibrio natriegens* DSMZ 749 16S rRNA gene was included in order to be able to calculate an absolute abundance from C<sub>T</sub> values together with no template controls. The resulting 16S rRNA gene copy number (bacterial load) was log transformed prior to using analytically.

### Microarray Gene Expression Profiling

At each time point, 3 ml of whole blood were collected into each of two Tempus tubes (Applied Biosystems/Ambion) by trained research staff following a standard phlebotomy protocol. Blood was vigorously mixed immediately following collection and stored at -80°C before RNA extraction. For each patient, the contents of one tube were used for analysis and the other tube was retained in case of assay failure. RNA was isolated using 1.5 ml whole blood and the MagMAX-96 Blood RNA Isolation Kit (Applied Biosystems/Ambion), as per the manufacturer's instructions. 250  $\mu$ g of isolated total RNA was globin-reduced using the GLOBINclear 96-well format kit (Applied Biosystems/Ambion) according to the manufacturer's instructions. Total and globin-reduced RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA yield was assessed using a NanoDrop8000 spectrophotometer (NanoDrop Products, Thermo Fisher Scientific). High-quality (>6.5 RIN) whole blood RNA was successfully obtained and processed by microarray in all cases. Biotinylated, amplified antisense complementary RNA (cRNA) targets were prepared from 200-250 ng of globin-reduced RNA using the Illumina CustomPrep RNA amplification kit (Applied Biosystems/Ambion). For each sample, seven hundred and fifty nanograms of labelled cRNA were hybridised overnight to Illumina Human HT12 V4 BeadChip arrays (Illumina), which contained greater than 47,000 probes.

The arrays were washed, blocked, stained and scanned on an Illumina iScan, as per the manufacturer's instructions. GenomeStudio (Illumina) was used to perform quality control and generate signal intensity values.

### Microarray Data Processing

Raw microarray data were processed using GeneSpring GX version 12.5 (Agilent Technologies). Following background subtraction, each probe was attributed a flag to denote its signal intensity detection  $P$  value. Filtering on flags removed probe sets that did not result in a 'present' call in at least 10% of the samples, where the 'present' lower cut-off = 0.99. Signal values were then set to a threshold level of 10,  $\log^2$  transformed, and per-chip normalised using a 75<sup>th</sup> percentile-shift algorithm. Each gene was normalised by dividing each mRNA transcript by the median intensity of all samples. Statistical analysis was performed after these steps had been performed.

### Microarray Data Analysis

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

Comparison Ingenuity Pathway Analysis (IPA) (Ingenuity Systems Inc., Redwood, CA) was used to determine the most significant canonical pathways for up-regulated and down-regulated transcripts ( $P < 0.05$  Fishers Exact test). Additionally, IPA was used to generate graphical representations of selected canonical pathways, and network diagrams. For the 231 transcript list, significantly activated (z score  $> 2$ ) and significantly repressed (z score  $< -2$ ) biofunctions were identified in IPA and expressed in gene network diagrams. GO Term (Gene Ontology Consortium) analysis integrated with GeneSpring GX12.5 was used to identify biological processes, according to GO annotations<sup>43</sup>.

The molecular distance to health (MDTH) and molecular scored were calculated using methods described previously<sup>24</sup> and applied to different signatures. Transcriptional modular analysis was applied as described previously<sup>23</sup>. Briefly, raw expression levels of all transcripts significantly detected from background were compared between each sample and all the controls present in a given data set. The percentage of significantly expressed genes in

each module was represented by the colour intensity, with red indicating over-expression and blue indicating under-expression. Statistical testing was performed using Student's t-test ( $P < 0.05$ ). The mean percentage of significant genes and the mean fold change of these genes compared to the controls in specific modules were shown in graphical form (unpaired t-test,  $P < 0.00001$ ). MDTH and modular analysis were calculated in Microsoft Excel 2010 (Microsoft Corp.). GraphPad Prism V5 for Windows (GraphPad Software Inc., La Jolla, CA, USA) and R 3.3.2 (R Development Core Team) were used to generate graphs and perform additional statistical analyses.

The raw and normalized microarray data that support the findings of this study have been deposited in GEO with the accession code GSE111368 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE111368>)

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

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

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

The MOSAIC consortium (ClinicalTrials.gov identifier: NCT00965354) was supported by UK National Institute for Health Research (NIHR) Comprehensive Local Research Networks (CLRNs), the Biomedical Research Centre (NIHR Imperial BRC) and Unit (NIHR Liverpool BRU), the Health Protection Research Unit in Respiratory Infections in partnership with Public Health England (PHE) at Imperial College London (NIHR HPRU RI), the Health Protection Agency (latterly PHE) Microbiology Services, Colindale and the staff of the Roslin Institute, Edinburgh, Scotland. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, PHE or the Department of Health (UK).

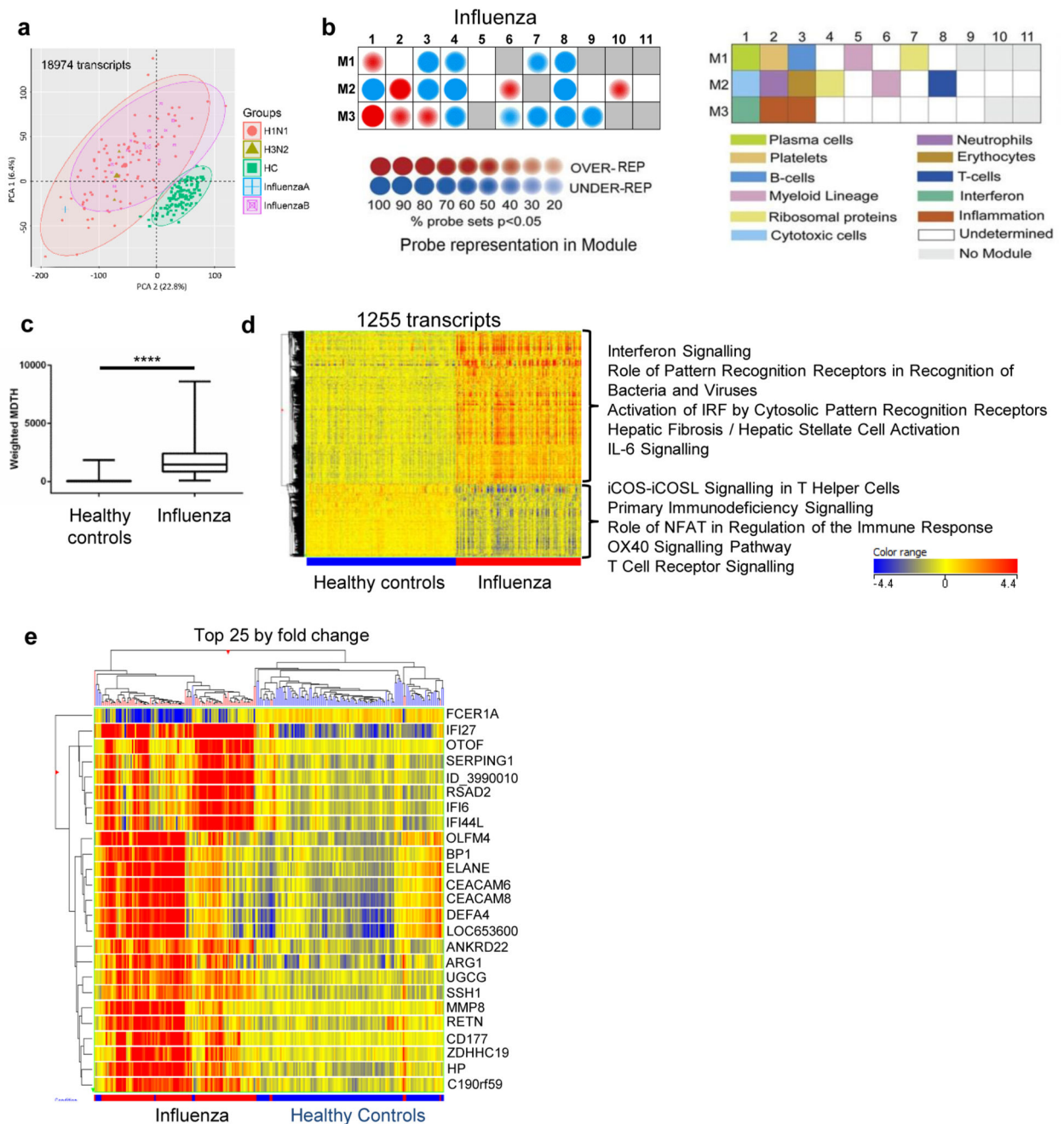
Anne O'Garra and Christine Graham were supported by the Medical Research Council, United Kingdom (U117565642), The Francis Crick Institute, London (AOG10126, which receives its core funding from Cancer Research U.K. (FC001126), the U.K. Medical Research Council (FC001126), the Wellcome Trust (FC001126),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

## References

1. Stöhr K. Preventing and treating influenza. *BMJ*. 2003; 326:1223–1224. [PubMed: 12791711]
2. Hayward AC, et al. Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study. *The Lancet Respiratory medicine*. 2014; 2:445–454. [PubMed: 24717637]
3. Dawood FS, et al. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *The Lancet infectious diseases*. 2012; 12:687–695. [PubMed: 22738893]
4. Bautista E, et al. Clinical aspects of pandemic 2009 influenza A (H1N1) virus infection. *The New England journal of medicine*. 2010; 362:1708–1719. [PubMed: 20445182]
5. Hui DS, Lee N, Chan PK. Clinical management of pandemic 2009 influenza A(H1N1) infection. *Chest*. 2010; 137:916–925. [PubMed: 20022969]
6. Peiris JS, Cheung CY, Leung CY, Nicholls JM. Innate immune responses to influenza A H5N1: friend or foe? *Trends in immunology*. 2009; 30:574–584. [PubMed: 19864182]
7. Tisoncik JR, et al. Into the eye of the cytokine storm. *Microbiology and molecular biology reviews* : MMBR. 2012; 76:16–32. [PubMed: 22390970]
8. de Jong MD, et al. Fatal outcome of human influenza A (H5N1) is associated with high viral load and hypercytokinemia. *Nat Med*. 2006; 12:1203–1207. [PubMed: 16964257]
9. Ioannidis I, et al. Plasticity and virus specificity of the airway epithelial cell immune response during respiratory virus infection. *Journal of virology*. 2012; 86:5422–5436. [PubMed: 22398282]
10. Ramilo O, et al. Gene expression patterns in blood leukocytes discriminate patients with acute infections. *Blood*. 2007; 109:2066–2077. [PubMed: 17105821]
11. Zaas AK, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host Microbe*. 2009; 6:207–217. [PubMed: 19664979]
12. Woods CW, et al. A host transcriptional signature for presymptomatic detection of infection in humans exposed to influenza H1N1 or H3N2. *PloS one*. 2013; 8:e52198. [PubMed: 23326326]
13. Nakaya HI, et al. Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol*. 2011; 12:786–795. [PubMed: 21743478]
14. Li S, et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol*. 2014; 15:195–204. [PubMed: 24336226]
15. Obermoser G, et al. Systems scale interactive exploration reveals quantitative and qualitative differences in response to influenza and pneumococcal vaccines. *Immunity*. 2013; 38:831–844. [PubMed: 23601689]
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*. 2014; 369 20130427.
17. Pascual V, Chaussabel D, Banchereau J. A genomic approach to human autoimmune diseases. *Annual review of immunology*. 2010; 28:535–571.
18. Everitt AR, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature*. 2012; 484:519–523. [PubMed: 22446628]
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*. 2013; 7:1416–1426. [PubMed: 23927713]
20. Rykkvin R, Kilander A, Dudman SG, Hungnes O. Within-patient emergence of the influenza A(H1N1)pdm09 HA1 222G variant and clear association with severe disease, Norway. *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin*. 2013; 18
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*. 2010; 84:3752–3758. [PubMed: 20130063]



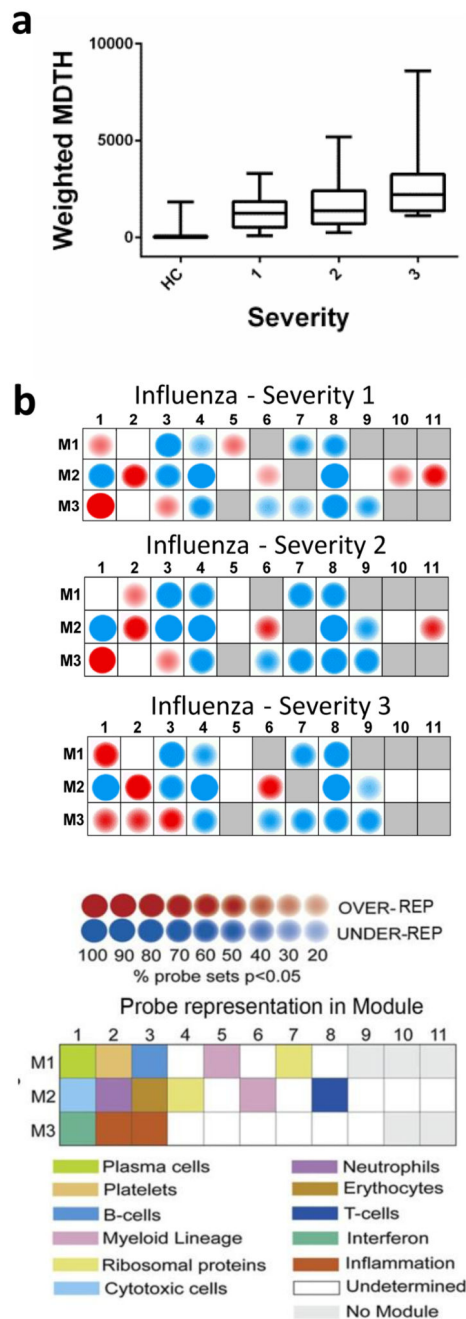
22. Elderfield RA, et al. Accumulation of human-adapting mutations during circulation of A(H1N1)pdm09 influenza virus in humans in the United Kingdom. *Journal of virology*. 2014; 88:13269–13283. [PubMed: 25210166]
23. Chaussabel D, et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. *Immunity*. 2008; 29:150–164. [PubMed: 18631455]
24. Pankla R, et al. Genomic transcriptional profiling identifies a candidate blood biomarker signature for the diagnosis of septicemic melioidosis. *Genome Biol*. 2009; 10:R127. [PubMed: 19903332]
25. Simon L, Gauvin F, Amre DK, Saint-Louis P, Lacroix J. Serum procalcitonin and C-reactive protein levels as markers of bacterial infection: a systematic review and meta-analysis. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2004; 39:206–217. [PubMed: 15307030]
26. Wu MH, et al. Can procalcitonin tests aid in identifying bacterial infections associated with influenza pneumonia? A systematic review and meta-analysis. *Influenza and other respiratory viruses*. 2013; 7:349–355. [PubMed: 22672284]
27. Falsey AR, et al. Bacterial complications of respiratory tract viral illness: a comprehensive evaluation. *The Journal of infectious diseases*. 2013; 208:432–441. [PubMed: 23661797]
28. McNab F, Mayer-Barber K, Sher A, Wack A, O'Garra A. Type I interferons in infectious disease. *Nat Rev Immunol*. 2015; 15:87–103. [PubMed: 25614319]
29. Marshall JC. Iatrogenesis, inflammation and organ injury: insights from a murine model. *Crit Care*. 2006; 10:173. [PubMed: 17123432]
30. Openshaw PJ, Chiu C. Protective and dysregulated T cell immunity in RSV infection. *Curr Opin Virol*. 2013; 3:468–474. [PubMed: 23806514]
31. Davidson S, Crotta S, McCabe TM, Wack A. Pathogenic potential of interferon alpha in acute influenza infection. *Nat Commun*. 2014; 5:3864. [PubMed: 24844667]
32. Gregory DJ, Kobzik L. Influenza lung injury: mechanisms and therapeutic opportunities. *Am J Physiol Lung Cell Mol Physiol*. 2015; 309:L1041–1046. [PubMed: 26408556]
33. Short KR, Kroeze EJ, Fouchier RA, Kuiken T. Pathogenesis of influenza-induced acute respiratory distress syndrome. *Lancet Infect Dis*. 2014; 14:57–69. [PubMed: 24239327]
34. Sandler NG, et al. Type I interferon responses in rhesus macaques prevent SIV infection and slow disease progression. *Nature*. 2014; 511:601–605. [PubMed: 25043006]
35. Biron CA. Interferons alpha and beta as immune regulators--a new look. *Immunity*. 2001; 14:661–664. [PubMed: 11420036]
36. Narasaraju T, et al. Excessive neutrophils and neutrophil extracellular traps contribute to acute lung injury of influenza pneumonitis. *The American journal of pathology*. 2011; 179:199–210. [PubMed: 21703402]
37. Lukens MV, et al. A systemic neutrophil response precedes robust CD8(+) T-cell activation during natural respiratory syncytial virus infection in infants. *Journal of virology*. 2010; 84:2374–2383. [PubMed: 20015982]
38. Tate MD, Brooks AG, Reading PC, Mintern JD. Neutrophils sustain effective CD8(+) T-cell responses in the respiratory tract following influenza infection. *Immunology and cell biology*. 2012; 90:197–205. [PubMed: 21483446]
39. Lim K, et al. Neutrophil trails guide influenza-specific CD8(+) T cells in the airways. *Science (New York, N.Y.)*. 2015; 349:aaa4352.
40. Mayer-Barber KD, et al. Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. *Nature*. 2014; 511:99–103. [PubMed: 24990750]
41. Carr MJ, et al. Development of a real-time RT-PCR for the detection of swine-lineage influenza A (H1N1) virus infections. *J Clin Virol*. 2009; 45:196–199. [PubMed: 19540799]
42. Nguyen-Van-Tam JS, et al. Risk factors for hospitalisation and poor outcome with pandemic A/H1N1 influenza: United Kingdom first wave (May–September 2009). *Thorax*. 2010; 65:645–651. [PubMed: 20627925]
43. Ashburner M, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000; 25:25–29. [PubMed: 10802651]



**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 2010/11 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’ (MDTH24) of 2010/11 influenza patients compared to healthy controls, undertaken on 4526

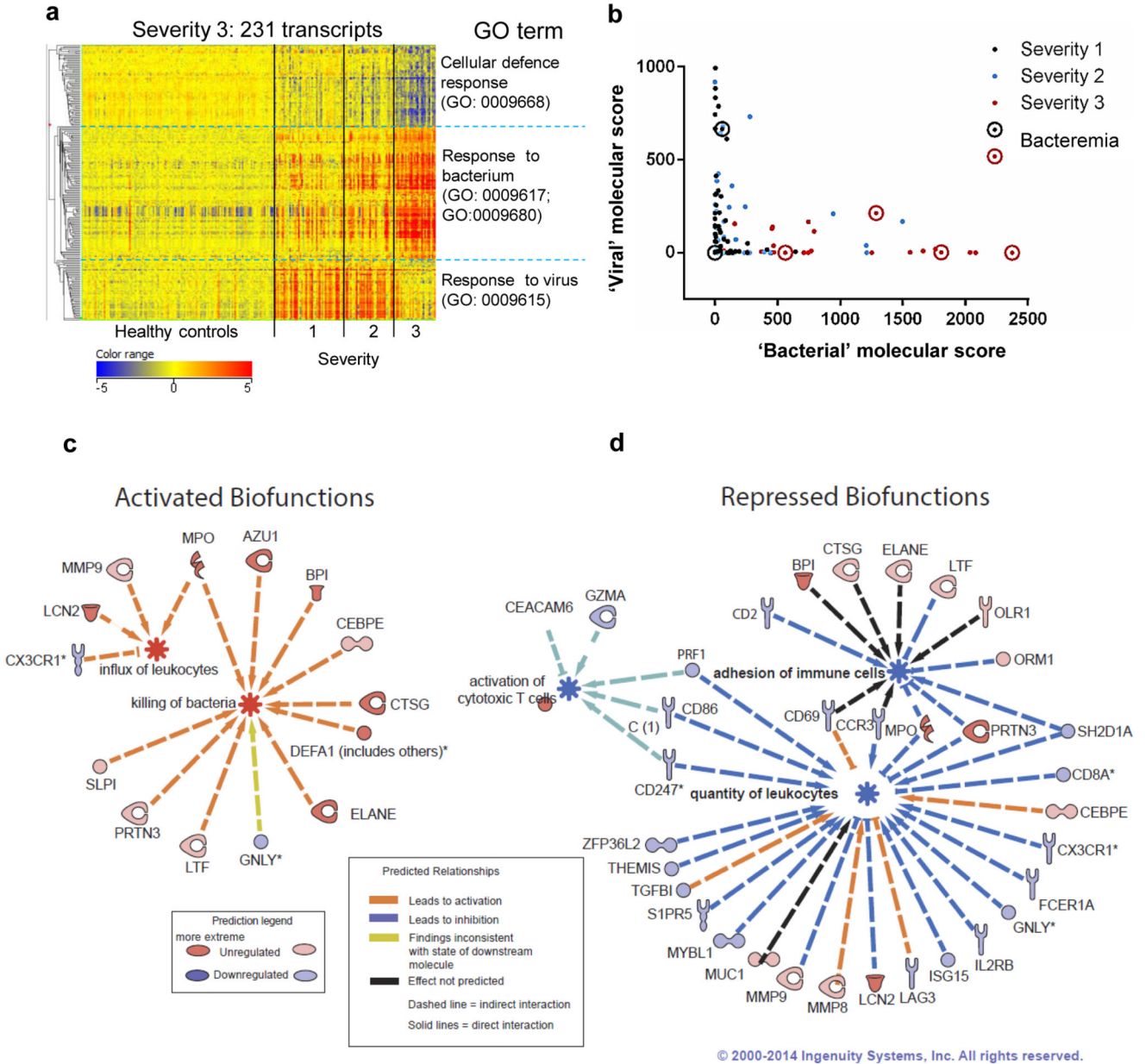
transcripts that were significantly detected above background, filtered for low expression (transcripts retained if  $>2$  fold-change (FC) from median normalised intensity value in more than 10% of all samples). Box whisker plot with min/max lines; statistical test: Mann-Whitney  $P < 0.0001$ . **(d)** 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  $>2$ FC 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)** 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 MDTH of 2010/11 influenza patients (n=109) grouped by severity of illness score (1: normoxic (n=47); 2: hypoxia requiring correction by mask oxygen (n=34); 3: mechanical ventilation (n=28)), compared to healthy controls (HC; n=130), 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

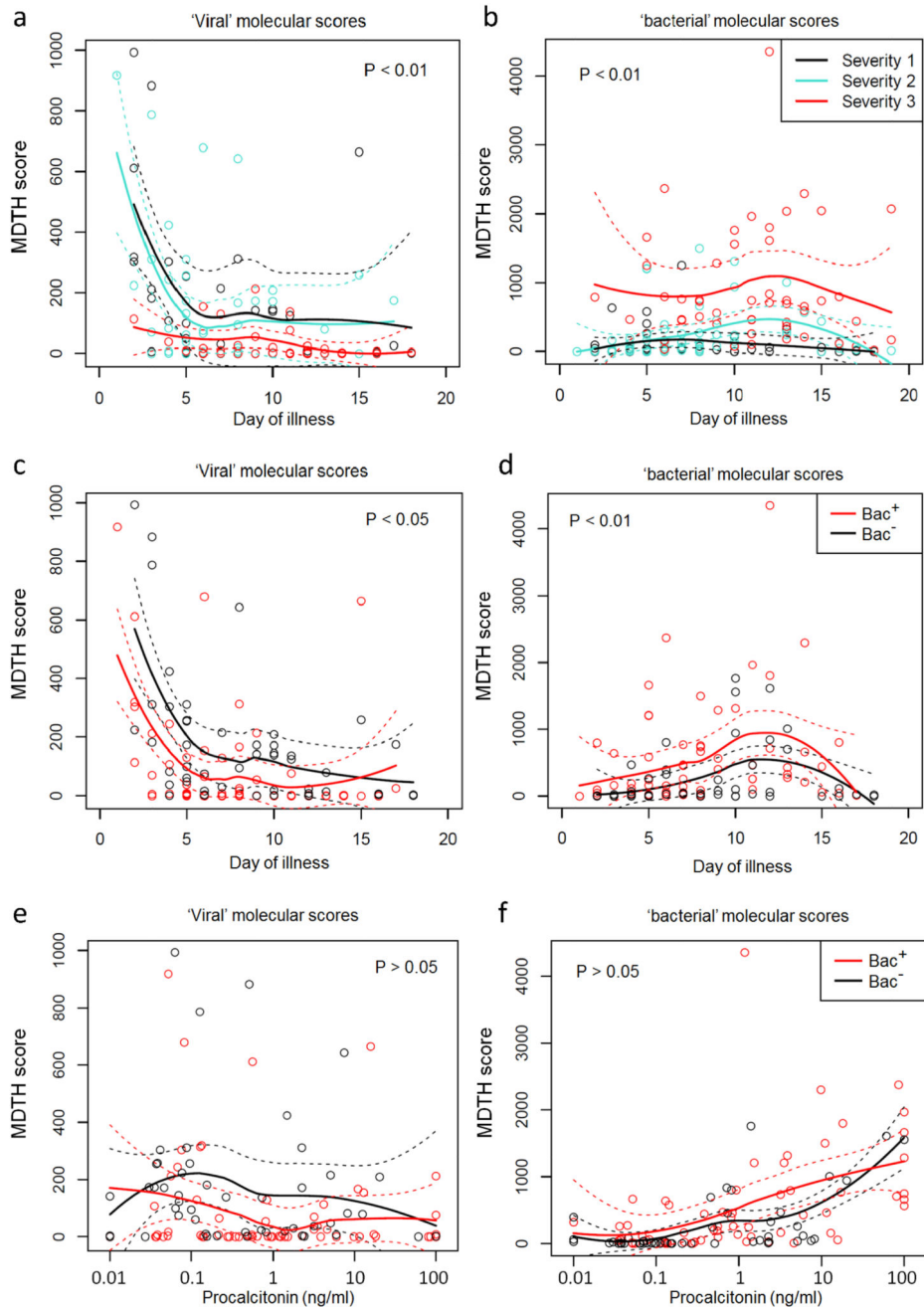
analysis of 2010/11 influenza patients (n=109) grouped by severity, relative to healthy controls (n=130). The colour intensity correlates with the percentage of genes in that module that are significantly differentially expressed.



**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). Listed next to the heat-map are the top GO terms for the 3 major subdivisions of the dendrogram (clustered by Pearson’s uncentered (cosine) with average linkage rule). 2010/11 cohort: severity 1,  $n=47$ ; severity 2,  $n=34$ ; severity 3,  $n=28$ ; HC,  $n=130$ . (b) Weighted molecular score (relative to healthy controls,  $n=130$ ) of the 112 ‘bacterial response’ transcripts plotted against molecular score of the 51 ‘viral response’ transcripts for the 109 influenza individuals at the

T1 time point. Severity of illness is indicated by different colours of dots: severity 1, black dots; severity 2, blue dots; severity 3, red dots. Circled dots identify patients with confirmed bacteraemia. **(c)** IPA significantly activated ( $z$  score  $>2$ ) or **(d)** repressed ( $z$  score  $<2$ ) biofunctions, identified by analysis of 231 transcript list; selected networks of biofunctional genes are shown.

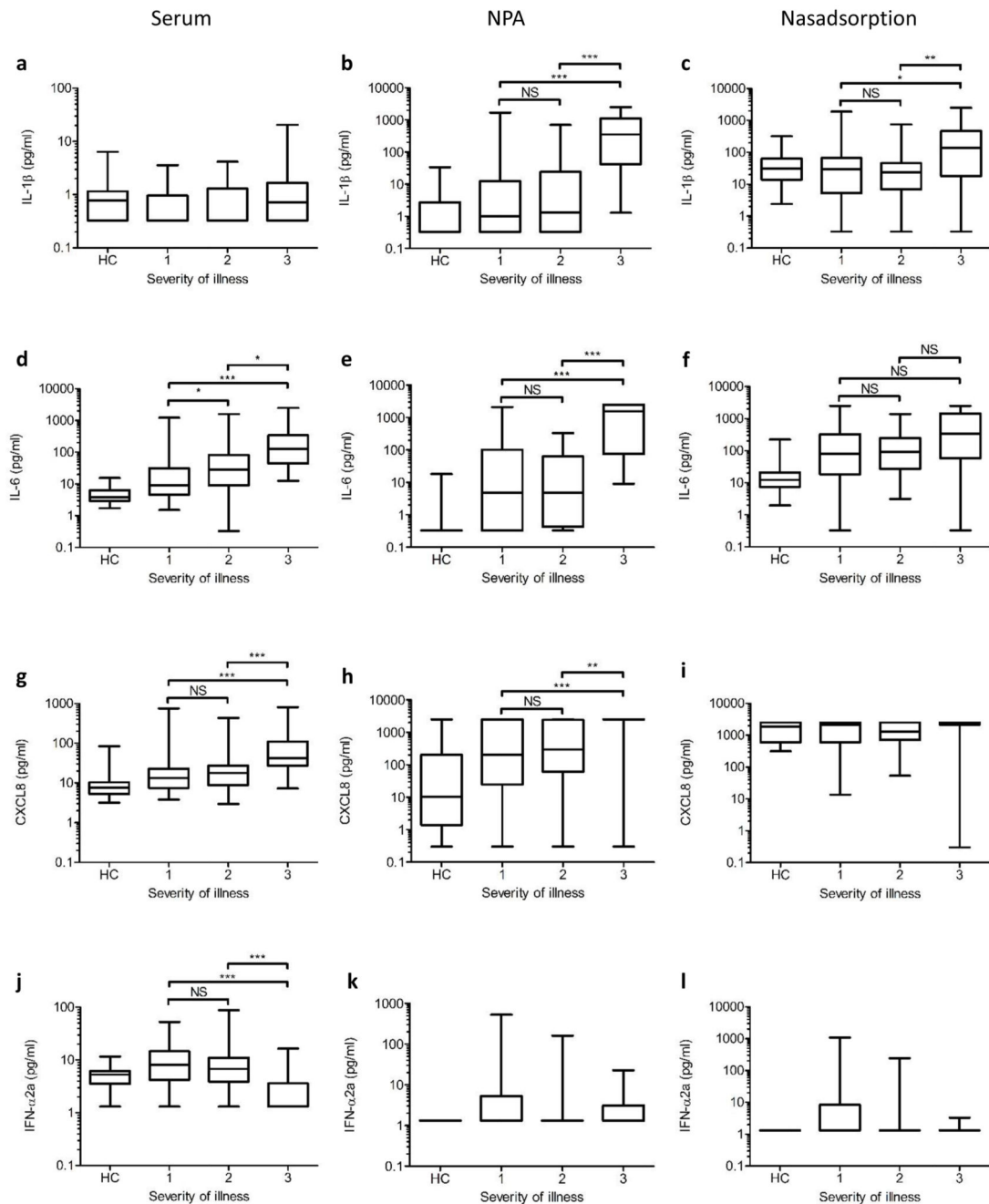


**Figure 4. Relationship between severity of illness, bacterial infection, procalcitonin and molecular scores.**

'Viral' and 'bacterial' MDTH scores (according to GO terms, as described in Fig. 3) calculated for patients with confirmed influenza (2010/11 cohort, n=109) according to clinical categories at both the first and second sampling time-points (T1 and T2). Loess fitting was used to interpolate and estimate mean values non-parametrically from the data (solid lines); dashed lines show the estimated 95% confidence interval values of the mean; statistical significance of differences were calculated using Chi-squared tests to compare the deviance of generalized linear models. **(a)** Viral MDTH according to day of illness, for



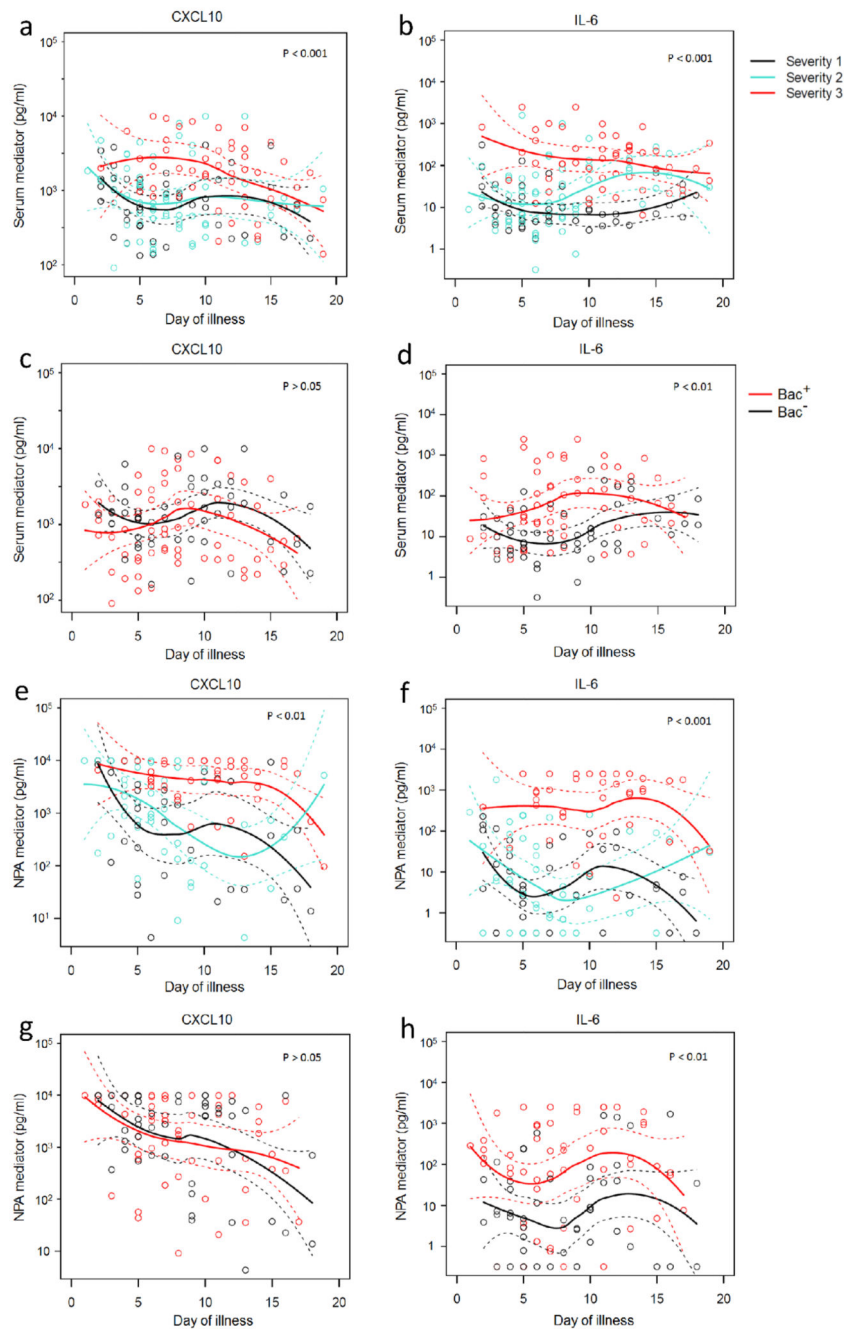
influenza patients stratified by severity of illness. **(b)** Bacterial MDTH according to day of illness, for influenza patients stratified by severity of illness. **(c)** Viral MDTH according to day of illness, for influenza patients stratified by presence ( $Bac^+$ , n=39; 63 samples) or absence ( $Bac^-$ , n=34; 52 samples) of clinically significant bacterial co-infections. **(d)** Bacterial MDTH according to day of illness, for influenza patients stratified by presence ( $Bac^+$ , n=39; 63 samples) or absence ( $Bac^-$ , n=34; 52 samples) of clinically significant bacterial co-infections. **(e)** Viral MDTH according to blood procalcitonin levels, for influenza patients stratified by presence ( $Bac^+$ , n=39; 63 samples) or absence ( $Bac^-$ , n=34; 52 samples) of clinically significant bacterial co-infections. **(f)** Bacterial MDTH according to blood procalcitonin levels, for influenza patients stratified by presence ( $Bac^+$ , n=39; 63 samples) or absence ( $Bac^-$ , n=34; 52 samples) of clinically significant bacterial co-infections.



**Figure 5. Levels of selected mediators in different compartments according to severity of illness and clinical designation of probable bacterial co-infection status.**

Serum, nasopharyngeal aspirate (NPA) and nasadsorption eluates were obtained from influenza patients at T1, and from adult healthy controls. Results are shown for IL-1 $\beta$  (a, b, c), IL-6 (d, e, f), CXCL8 (g, h, i), and IFN- $\alpha$ 2a (j, k, l). For each mediator box plot, the central line shows the median mediator level (pg/mL, log scale) and the box margins show the interquartile range; outer bars show the range. Zero values and values below the lower limit of detection were assigned half the geometric mean lower limit of detection for display

purposes. The upper limit of detection for all assays was 2500 pg/mL. Kruskal-Wallis test with Dunn's post test was used to assess significance (\*\* $p < 0.001$ ; \* $p < 0.01$ ; \* $p < 0.05$ ; NS = not significant). Severity of illness at T1 is shown. HC = healthy controls. Serum samples for HCs and participants with severity 1, 2, and 3 illness (**a, d, g, and j**):  $n = 36, 58, 43,$  and  $31,$  respectively. NPA samples for healthy controls and participants with severity 1, 2, and 3 illness (**b, e, h and k**):  $n = 35, 50, 32,$  and  $27,$  respectively. Nasabsorption eluate samples for healthy controls and participants with severity 1, 2, and 3 illness (**c, f, i and l**):  $n = 36, 60, 43,$  and  $30,$  respectively.



**Figure 6. Relationships between severity of illness, bacterial infection, and selected mediators.** Levels (pg/ml) of CXCL10 and IL-6 in serum (a-d) and NPA (e-h) according to day of illness at both the first and second sampling time-points (T1 and T2). Patients were stratified according to severity of illness (a, b, e, f) and by presence (Bac<sup>+</sup>; 39 subjects, 63 samples) or absence (Bac<sup>-</sup>; 34 subjects, 52 samples) of proven bacterial infection (c, d, g, h). Loess fitting was used to demonstrate time trends of mean values interpolated non-parametrically from the data (solid lines); dashed lines show the estimated 95% confidence interval values

of the mean. Statistical significance of differences was calculated using Chi-squared tests to compare the deviance of generalized linear models.

**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)
<b>Mean age in years (range)</b>	41 (17-71)	35 (20-68)	44 (23-74)	37 (21-54)
<b>Female (%)</b>	53 (48.6)	75 (57.7)	10 (45.5)	14 (56)
<b>Ethnicity (%)</b>				
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
<b>Comorbidities</b>				
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
<b>Women age 15-49y who were pregnant</b>	10/43 (23.3)	1/75 (1.3)	2/8 (25)	0
<b>Influenza type</b>				
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
<b>Severity of illness at T1 (%)</b>				
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
<b>Peak severity for illness episode (%)</b>				
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 variable due to rounding. NA: Not Applicable.