Online Supplement for

NTHi infection of pulmonary macrophages drives neutrophilic inflammation in severe asthma

Jodie Ackland\textsuperscript{a}, Clair Barber\textsuperscript{a,c}, Ashley Heinson\textsuperscript{a}, Adnan Azim\textsuperscript{a,c}, David W. Cleary\textsuperscript{a,c}, Myron Christodoulides\textsuperscript{a}, Ramesh J. Kurukulaaratchy\textsuperscript{a,c}, Peter Howarth\textsuperscript{a,c}, Tom M.A. Wilkinson\textsuperscript{a,b,c}, Karl J. Staples\textsuperscript{a,b,c}\# On behalf of the WATCH study investigators.

\textsuperscript{a}Clinical and Experimental Sciences, University of Southampton Faculty of Medicine, Southampton, UK

\textsuperscript{b}Wessex Investigational Sciences Hub, University of Southampton Faculty of Medicine, Southampton General Hospital, Southampton, UK

\textsuperscript{c}NIHR Southampton Biomedical Research Centre, University Hospital Southampton NHS Foundation Trust, Southampton, UK

**Short Title:** NTHi drives neutrophilic inflammation in asthma

\#Address correspondence to Karl J. Staples, K.Staples@soton.ac.uk
Supplementary Material and Methods

Lung sample processing from WATCH study

BAL was collected from patients and stored on ice until processing by members of the WATCH study team. Briefly, samples were filtered using a 100 μm cell strainer and centrifuged at 790 $g$ for 10 min at 4°C. The cell pellet was resuspended in PBS and a cell count using the trypan blue exclusion method was performed to assess cell viability, which identified a median BAL cell viability of 77%. Subsequently, the cell suspension was adjusted to $5 \times 10^5$ cells per ml in PBS. Cell cytospins were generated, with 75 μl of the cell suspension added to a cytospin funnel and subjected to centrifugation using a Shandon Cytospin® centrifuge at 450 rpm for 6 min. Slides were left to air dry at room temperature for at least 2 h before use or stored at -20°C until required for Fluorescence in situ hybridization (FISH, as described in 2.24). The remaining cell suspension was resuspended to allow for a minimum of $5 \times 10^5$ cells per ml and centrifuged at 790 $g$ for 10 min at 4°C. The sample was resuspended in 500 μl QIAzol and vortexed for 30 s. Samples were stored at -80°C until RNA isolation was performed.

Sputum was also collected and processed for each patient, and differential sputum cell counts were performed in order to determine patient inflammatory phenotypes. Briefly, sputum was obtained by members of the WATCH study team and was stored on ice until processing. Sputum was weighed and diluted with 2 volumes of 0.2% (w/v) dithiothreitol (DTT) to processed sputum weight (g) and vortexed for 15 s. Sample was placed on a bench roller in ice for 30 min. The sample was then passed through a 100 μm cell strainer and centrifuged at 790 $g$ for 10 min at 4°C. The cell pellet was resuspended and a cell count using the trypan blue exclusion method was performed to assess cell viability, which indicated a median sputum cell viability of 70%. Cells were resuspended to a concentration of $5 \times 10^5$ cells per ml to generate cytospin slides, with 75 μl of the cell suspension added to a cytospin funnel and subjected to centrifugation using a Shandon Cytospin® centrifuge at 450 rpm for 6 min. Slides were left to air dry at room temperature for at least 2 h before use or stored at -20°C until required. Once dry, slides were manually counted under a light microscope, with a total of 400
cells manually counted to determine the proportion of macrophages, neutrophils, eosinophils, lymphocytes and epithelial cells present in each sample. Sputum cell counts were used to determine the inflammatory phenotype of patients using published cut off values:\textsuperscript{50,68,69}:

- neutrophilic = $\geq 40\%$ neutrophils,
- eosinophilic = $\geq 3\%$ eosinophils,
- mixed granulocytic = $\geq 40\%$ neutrophils and $\geq 3\%$ eosinophils,
- paucigranulocytic = $< 40\%$ neutrophils and $< 3\%$ eosinophils.

No data regarding bacteria or viral culture were available for analysis.

**Fluorescence in situ hybridisation (FISH)**

Stored BAL cytospin slides were thawed and allowed to come to room temperature. Slides were washed in PBS before being fixed in a paraformaldehyde-acetic acid fixative solution (36.5M paraformaldehyde and 5% acetic acid in PBS) for 15 min. Slides were subsequently washed twice in PBS for 5 min, before two final quick washes in water. Slides were dehydrated using an ethanol gradient of 70% ethanol for 5 min, 90% ethanol for 5 min and 100% ethanol for 5 min before being left to completely air dry. The NTHi specific probe HAIN16S1251 and the pan-bacteria probe EUB338A were diluted to a concentration of 50 $\mu$g/μl in hybridization buffer (0.9 M NaCl, 20 mM Tris/HCl, 0.01% SDS, and 30% formamide pH 7.5) and 20 $\mu$l added to each slide. Hybridization was performed for 2 h at 46$^\circ$C using a ThermoBrite StatSpin humid chamber before being washed for 15 min in wash buffer (80 mM NaCl, 20 mM Tris- HCl pH 8, 5 mM EDTA, and 0.01% SDS) at 48$^\circ$C. Slides were again washed in PBS before addition of 150 $\mu$l of Vector® TrueVIEW® Autofluorescence Quenching reagent (Vectorlabs) for 4 min. Slides were washed in PBS before glass coverslips were added using Vectashield mounting solution containing 4',6-Diamidino-2-Phenylindole (DAPI) nuclear stain (Vectorlabs). Slides were stored at 4$^\circ$C until visualisation.

**Quantification of NTHi colonisation in BAL samples**
Quantification of the presence of NTHi was performed in ImageJ/FIJI. The number of macrophages in each image, represented by DAPI staining of single cell nucleus, was quantified using the Analyse Particles in-built ImageJ tool. This was achieved by setting the threshold for each image and running the ‘Convert to Mask’ and ‘Erode’ functions before the ‘Analyse Particles’ function to count the individual nuclei. The CY3 NTHi-specific signal was next quantified using the Find Maxima in-built ImageJ tool. Firstly, each image was adjusted using the ‘Enhance Contrast’, ‘Smooth’ and ‘Gaussian Blur’ in-built ImageJ functions. The ‘Find Maxima’ function was then used to identify individual CY3 signal per host cell which passed a set ‘prominence’ level, which was set to reduce potential background or autofluorescence being counted as NTHi. The output of DAPI and CY3 signal quantification was used to determine the number of DAPI-stained cells associated with the CY3-NTHi signal, which was expressed as a percentage of cells colonised with NTHi.

**Weighted gene correlation network analysis (WGCNA)**

Weighted gene correlation network analysis (WGCNA) was performed using the WGCNA R package\(^{42}\) on the normalised, log transformed CPM MDM data set. Construction of the gene network was performed using the WGCNA automatic network construction method. A soft thresholding power of 18 was chosen based on the scale-free topology fit output of the `pickSoftThreshold` function using default values provided by the WGCNA package, with a power of 18 the lowest power to completely intersect the high value red line (\(R^2 = 0.9\)) on the scale independence plot (Figure S9).

**Identification of modules and genes of interest**

Correlation analyses were performed using Pearson correlation as described in the WGCNA package. Correlation analysis between module eigengene and sample trait information was performed to identify which modules were significantly association with NTHi infection. The
significant modules were assessed for correlations between gene significance (GS, representing the correlation between each gene and the trait of interest) and module membership (MM, representing the significance of gene assignment to the module) to determine module-trait relationships. To visualise the gene network within the significant blue module, the network file was exported to Cytoscape. Due to the large size of the module (2333 genes), a topological overlap threshold of 0.34 was used to extract a meaningful network of highly interconnected genes for further exploration. The cytoHubba plugin in Cytoscape was used to rank genes present in the blue module gene network using the Maximal Clique Centrality (MCC) scoring method.

**Gene list enrichment analysis**

Gene list enrichment analysis and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis was performed using ToppFunn, which is part of the online ToppGene Suite, using default parameter settings (FDR multiple correction method and enrichment significance cut off level 0.05).

**Quantification of MDM mediator release by Luminex**

Analysis of MDM cell culture supernatants for quantification of mediator release also performed using a customised 14-plex Luminex Human Magnetic Assay according to the manufacturer’s instructions (R&D Systems). Briefly, 50 µl of the premixed microparticle cocktail was added to each well of a 96 well plate and incubated with 50 µl of neat MDM culture supernatant or the pre-prepared standard at 4°C overnight. The plate was then washed with 100 µl Wash Buffer (R&D Systems) whilst attached to a Handheld Magnetic Washer, which held the microparticles and bound sample analytes to the bottom of the plate to prevent them washing away. The biotin antibody cocktail provided was diluted with the provided diluent concentrate (R&D Systems) and 50 µl was added to each well and incubated for 1 h on a plate
shaker. The plate was again washed as described above, before 50 µl of Streptavidin-PE was added and again incubated for 30 min on a plate shaker. The plate was read using a Bio-Plex 200 system (Bio-Rad), with the concentration of each analyte calculated from the concentration of the 1 in 3 diluted standard curve for each analyte. The lower and upper limit of quantitation (LLOQ and ULOQ) for each analyte is indicated in Table S4.
### Supplementary Tables

Table S1. Additional clinical characteristics and demographic data for the 25 severe asthma patients used in this study

|                          | Total (n=25) | NTHi + (n=14) | NTHi – (n=11) | p-value |
|--------------------------|--------------|---------------|---------------|---------|
| Total IgE$^a$            | 147.6 [32.7 – 366.1] | 213.2 [21.2-399.1] | 139.7 [44.3 – 239.3] | 0.8718  |
| Atpy$^b$                 | 10 (40%)     | 6 (42.4%)     | 4 (36.4%)     | 0.6084  |
| Blood eosinophils        | 0.2 [0.1 – 0.4] | 0.15 [0.1 – 0.3] | 0.2 [0.1 – 0.4] | 0.4963  |
| Comorbidities (Y/N)      | 16/9         | 9/5           | 7/4           | >0.9999 |

The 25 severe asthma patients were split based on the outcome of the FISH analysis into NTHi+ (14) and NTHi- (11) groups. Values reported are medians [IQR] or (%). # indicates missing patient data; Total IgE was not available for 1 patient, and atopy was determined as a positive skin prick test to at least one common aeroallergen, however this was not performed or data were missing for 10 patients. Comorbidities indicate the number of patients reporting at least 1 or more comorbidities (Y) or no comorbidities (N). Statistical analysis was performed between NTHi+ and NTHi- groups. Continuous data were analysed by Mann-Whitney U test. **Bold** indicates p-values determined as statistically significant (p<0.05).
Table S2. Ranking scores of the 62 blue module network genes.

| Rank | Gene                                      | Score |
|------|-------------------------------------------|-------|
| 1    | CASP4                                     | 28    |
| 2    | PNRC1                                     | 21    |
| 3    | SGPP2                                     | 9     |
| 4    | GBP5, GBP4                                | 3     |
| 6    | ABTB2, UPB1, GBP1, RIPK2, GCH1, PLAT, NAMPT, SLC1A2, CCL20, CCR7, STK26, IL6, TRAF1 | 2     |
| 19   | CXCL8, PTGS2, PFKFB3, LAMP3, CXCL2, MARCKS, IL23A, MX1, IL1B, ISG15, SOD2, IL12B, IL1A, SLAMF1, PPP3CC, IL15RA, SESN2, GFPT2, RSAD2, CMPK2 | 1     |
| 39   | IFI35, ISG20, PDE4B, CD40, BATF3, CSF2, AIM2, BATF2, TAP1, PKIG, CSF3, IL27, RNF19B, CD274, TNFAIP6, IRF1, CMTR1, RNF144B, LAP3, HAPLN3, SLAMF7, NLRC5, WTP, TNIP3 | 0     |

Genes were ranked using Maximal Clique Centrality (MCC) method using the cytoHubba plugin in Cytoscape. **CASP4**, **SGPP2** and **PNRC1** were the three highest ranked genes in the blue module. **Bold** indicates genes chosen for qPCR and/or ELISA/Luminex validation in MDM or BAL samples.
Table S3. List of TaqMan gene expression assays (all from Thermo Fisher) used for validation of macrophage gene expression by qPCR.

| Gene | Protein                                         | Assay ID             |
|------|------------------------------------------------|----------------------|
| B2M  | Beta-2-microglobulin                            | Hs00984230_m1        |
| CASP4| Caspase-4                                       | Hs01031951_m1        |
| GBP1 | Guanylate Binding Protein 1                     | Hs00977005_m1        |
| IL1B | Interleukin 1 Beta                              | Hs01555410_m1        |
| LAMP3| Lysosomal Associated Membrane Protein 3         | Hs01111316_m1        |
| PNRC1| Proline Rich Nuclear Receptor Coactivator 1     | Hs00199095_m1        |
| SGPP2| Sphingosine-1-Phosphate Phosphatase 2            | Hs00544786_m1        |
Table S4. The lower and upper limit of quantitation (LLOQ and ULOQ) for each analyte (pg/ml) measured by a customised Luminex Human Magnetic Assay (R&D).

| Analyte | Lower Limit of Quantification | Upper Limit of Quantification |
|---------|-------------------------------|------------------------------|
| IL-23   | 45.919                        | 34817.527                    |
| CCL20   | 8.558                         | 2169.997                     |
Supplementary Figures

Figure S1. Proportion of BAL cell types isolated from the airways of severe asthma patients. (A) The cells recovered from BAL were predominantly macrophages (91.6%). (B) No differences in BAL macrophage proportions were detected between NTHi+ (92.2%) and NTHi- (91.6%) groups. Data were analysed by Mann Whitney U test. Lines indicate median values. Key shows cell types included for differential cell counting.
Figure S2. Stratification of the 25 severe asthma BAL samples by NTHi colonisation and inflammatory phenotype status. Patients were stratified into neutrophilic (≥40% neutrophils, <3% eosinophils), eosinophilic (<40% neutrophils, ≥3% eosinophils), mixed granulocytic (≥40% neutrophils, ≥3% eosinophils) or paucigranulocytic (<40% neutrophils, <3% neutrophils) inflammatory phenotypes. Accompanying table (bottom) indicates the significance between NTHi+ and NTHi- groups as performed by Fisher’s exact test.
Figure S3. Individual module membership and gene significance correlation plots for the ten modules significantly associated with NTHi infection. Identification of gene significance in the ten modules (A) blue, (B) brown, (C) red, (D) black, (E) yellow, (F) pink, (G) green, (H) turquoise, (I) salmon and (J) tan found to be significantly correlated with the NTHi infection. Module membership indicates the correlation of each individual gene to the specific module eigengene. The correlation and corresponding p-value for each comparison is present above each scatterplot and modules are ordered based on level of correlation (high to low).
Figure S4. Gene ontology and KEGG pathway analysis indicates intracellular immune response pathways are modulated in the cluster associated with infection. The top 5 significantly enriched terms for Biological Processes (BP), Cellular Component (CC), Molecular Function (MF) and KEGG pathway are shown for each cluster, with Pathway/category IDs ordered by enrichment significance (FDR), of which all categories were deemed as statistically significantly functionally enriched (FDR p<0.05). (A) Cluster 1 indicating enrichment of immune response pathways, (B) Cluster 2 showing enrichment of metabolic and cell activation processes, (C) Cluster 3 is enriched in gene regulation and metabolic processes, (D) Cluster 4 is enriched in cell cycle processes, however no significant KEGG pathways were identified for this cluster. Gene ontology performed using ToppGene.
Figure S5. The blue module drives the enrichment of immune responses in cluster I and is associated with NTHi infection. (A) Enrichment analysis performed on each of the modules assigned to cluster I individually showed enrichment of different process for each module. The total number of enriched Biological Process terms for each category were: Cluster I: 1022, Blue module: 786, Yellow module: 9 and Salmon module: 31. No enriched terms for the pink module were found. Venn diagram shows the overlapping terms between each of the modules and cluster I, with the blue module and cluster I sharing the highest proportion of terms, indicating the contribution of the genes within this module to the enrichment of terms within Cluster I. (B) Barplot shows the top 5 most significantly enriched (adjusted p-val <0.05) Biological Process terms for the blue module, salmon module and yellow module respectively, further highlighting the immune processes specific to the blue module. (C) Heatmap of the 2333 genes present in the blue module constructed using Euclidean distance and Ward linkage. Samples cluster according to infection status (uninfected or infected).
Figure S6. The most highly interconnected blue module genes are more highly expressed in response to NTHi infection and are enriched in immune responses. (A) Heatmap visualising the expression profile of the 62 most highly connected blue module network genes. All genes were more highly expressed in response to NTHi infection compared to uninfected controls. Clustering was performed using Euclidean distance and Ward linkage methods. Gene enrichment analysis performed on the 62 genes using ToppGene demonstrated significant enrichment of (B) Biological Processes and (C) KEGG pathways (adjusted p-value <0.05 for all).
Figure S7. Validation of inflammatory mediator release. To validate the upregulation of genes for inflammatory mediators present in the blue module gene network resulted in increased protein level release, macrophage release of T17 mediators (A) IL-1β, (B) IL-8, (C) IL-23 and (D) CCL20 into cell culture supernatants at 6 h and 24 h in response to NTHi infection of MDM (N=5) was measured by Luminex (IL-23 and CCL20) or ELISA (CXCL8 and IL-1β). Horizontal dotted lines on graphs show the upper limit of quantification (ULOQ) or lower limit of quantification (LLOQ) of the Luminex assay where required. CCL20 was released into cell culture supernatants at such high amounts that the concentration of this chemokine was unable to be accurately quantified, so is represented here as fluorescence intensity. (E) Levels of CXCL8 in severe asthma BAL supernatants were assessed by ELISA (n=25) Graphs A-D show paired data and data were analysed by Wilcoxon signed rank test. Graph E shows unpaired data and data were analysed by Mann Whitney U test. *p<0.05. Lines indicate medians.
Figure S8. Validation of macrophage gene expression in the MDM model. To validate the expression of the genes in the blue module network, the top three ranked genes (A) CASP4, (B) PNRC1 and (C) SGPP2 and three other immune related genes in the blue module gene network (D) IL1B, (E) GBP1 and (F) LAMP3 were validated by qPCR using the MDM model \((N=6)\). Gene expression was normalised to B2M and is expressed as \(2^{-\Delta\Delta Ct}\). All genes were more highly upregulated in response to NTHi infection compared to uninfected controls. Data were analysed by Wilcoxon signed rank test, \(*p<0.05\). All graphs show paired data and bars indicate median values. Data were analysed by Wilcoxon signed rank test, \(*p<0.05\).
Figure S9. Network topology analysis for determining soft-threshold analysis power to be used for automatic network construction. (A) scale-free fit index (y axis) as a function of the soft-thresholding power (x axis). (B) mean connectivity degree (y axis) as a function of the soft thresholding power (x axis).