# Epigenetic and transcriptional plasticity of myeloid cells in Cystic Fibrosis

Adam M. Dinan[1,2]‡, Odiri Eneje[1,3]‡, Karen P. Brown[1,3], Frances Burden[4,5], Mary Morse[6], Rab K. Prinjha[6], Mattia Frontini*[4,5,7,8] & R. Andres Floto*[1,2,3]

1. Molecular Immunity Unit, Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Cambridge, UK.
2. University of Cambridge Centre for AI in Medicine, Cambridge, UK.
3. Cambridge Centre for Lung Infection, Royal Papworth Hospital, Cambridge, UK.
4. Department of Haematology, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK.
5. National Health Service (NHS) Blood and Transplant, Cambridge Biomedical Campus, Cambridge, UK.
6. Immunology Research Unit, GlaxoSmithKline Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, UK.
7. British Heart Foundation, Cambridge Centre for Research Excellence, University of Cambridge, Cambridge Biomedical Campus, Cambridge, UK.
8. Clinical and Biomedical Science, Faculty of Health and Life Sciences, University of Exeter Medical School, Exeter, UK.

‡ Contributed equally

*Correspondence to: Andres Floto (arf27@cam.ac.uk) or Mattia Frontini (M.Frontini@exeter.ac.uk)

## Abstract

Recent *in vitro* and *in vivo* studies suggest that epigenetic training in innate immune cells can alter cellular function over extended time periods. It is unclear to what extent such training persists in human myeloid cells during microbial infections and alters clinical outcomes. We therefore examined longitudinal transcriptional and epigenetic changes in patients with Cystic Fibrosis (CF), a disease characterised by temporal fluctuations in lung infection and inflammation. We find that sudden clinical deteriorations in lung health, termed Acute Pulmonary Exacerbations (APEs), are linked to a robust innate immune response (triggered in part by pattern recognition receptor (PRR) activation) and associated changes in phagocytic function. Treatment of patients with intravenous antibiotics results in rapid modification of myeloid cell gene expression and epigenetic state, towards that of healthy volunteers, and suggests that CF inflammatory lung damage is driven by repeated acute inflammatory episodes rather than a distinct chronic inflammatory programme.

## Main Text

Myeloid cell exposure to microbial cellular components has been found to lead to functional training via the deposition of epigenetic marks at gene promoters and transcriptional enhancers[1,2]. In particular, marks such as histone 3 lysine 27 acetylation (H3K27ac) and histone 3 lysine 4 trimethylation (H3K4me3) are believed to facilitate a more rapid recruitment of regulatory effectors at previously activated genes upon subsequent stimulation. Novel therapeutic strategies exploiting the induction of epigenetic memory to alter immune cell phenotypes have therefore been proposed[3].

It remains unclear to what extent infection-induced epigenetic alterations persist *in vivo* in mature myeloid cells such as neutrophils and monocytes, which are relatively short-lived in circulation[4,5]. Recent work has suggested that the peripheral innate immune system may play a role in epigenetic training, as acute stimulation of mouse haematopoietic stem cells (HSCs) with lipopolysaccharide (LPS), a component of the gram-negative bacterial cell wall, causes epigenetic imprinting protecting against future infection with *Pseudomonas aeruginosa* (PsA)[6]. Similarly, Bacillus Calmette-Guerin (BCG) vaccination of humans has been shown to induce long-term epigenetic and transcriptional changes in neutrophils, associated with altered functional capacity, presumably through functional changes in the bone marrow precursors of these cells[7].

Here, we sought to define the epigenetic and transcriptional landscapes in the myeloid cells of patients with CF, a disease associated with well-defined cycles of infection and inflammation[8]. CF patients frequently suffer from chronic bacterial infections of the lungs, and have episodes of rapid clinical deterioration (APEs) associated with decreased lung function and increased lung and systemic inflammation, that often require prolonged (usually 14 days) treatment with intravenous antibiotics. APEs are thought to lead to the cumulative inflammatory lung damage that remains the major cause of morbidity and mortality in this patient group. CF therefore provides a unique opportunity to understand the temporal dynamics of epigenetic training in humans in the context of chronic bacterial infection.

We collected whole blood samples from adult CF patients ($n$ = 13) chronically infected with PsA at 3 timepoints: at an APE onset, at the end of intravenous antibiotic treatment, and when returned to stable clinical baseline (at least 30 days after the APE onset; **Fig. 1a**; **Supplementary Table 1**). Treatment with intravenous antibiotics was, as expected, associated with resolution of systemic inflammation and an improvement in lung function (as measured by circulating C-reactive protein levels and FEV1 respectively; **Fig. 1b**). We confirmed that the study group was broadly representative of the clinical characteristics of the entire cohort attending the adult CF centre from which they were recruited (**Fig. 1c; Supplementary Fig. 1**).

Gene expression (bulk RNA-seq) of peripheral neutrophils and monocytes isolated from CF patients on each sampling day was directly compared with samples from age- and sex-matched healthy volunteers (HV; $n$ = 8; **Fig. 1d; Supplementary Fig. 2**). We observed a marked decrease in differential gene expression between CF and HV across the time series, with more differentially

expressed (DE) genes at the onset of exacerbation (day 0), than on subsequent sampling days (days 14 and 30+) (**Supplementary Table 2**). Only 8 genes (including *MAPK14*, *MAP3K20* and *TLR5*) were upregulated, and only 2 genes (*NUP210L* and *TSPAN13*) were downregulated, in CF neutrophils on all sampling days; while no genes were upregulated and only 2 genes (*ARL17A* and *LRRC37A2*) were downregulated on all sampling days in CF monocytes (**Fig. 1d & e; Supplementary Table 3**).

We used unsupervised hierarchical clustering (based on median expression levels per time point) to identify 7 DE gene clusters with different temporal dynamics in neutrophils (**Fig. 1f**). Notably, the median expression levels of all upregulated genes in CF were highest at day 0 (clusters 1-5, **Fig. 1f**) and the median expression levels of all downregulated genes in CF were lowest at day 0 (in clusters 6 and 7; **Fig. 1f**). A similar trend was observed in monocytes (**Supplementary Fig. 3**). GO term analysis showed that clusters of upregulated genes in CF were enriched for processes involved in innate immune response (**Fig. 1g**), while downregulated clusters showed no immune-specific enrichment (**Supplementary Fig. 4**).

We next assessed changes at the epigenetic level, in both cell types, through genome-wide profiling of H3K27ac by chromatin immunoprecipitation (ChIP-seq). Peaks were assigned to genes based on genomic proximity, using Hi-C data, and by correlation with expression levels across samples (**Methods**). Differentially acetylated regions (DAR) were identified for each time point by direct comparison with HV samples (**Supplementary Tables 4 & 5**). Again, fewest DAR were identified in both neutrophils and monocytes at Day 14 (**Supplementary Fig. 5**), indicating that IV treatment causes a reduction of the effects observed at day 0. In keeping with these changes, we also found differences in the functional properties of monocyte-derived macrophages between day 0, day 14, and day 30+, with decreased intracellular killing and greater inflammatory cytokine production observed in Day 14 cells (**Supplementary Fig. 6**).

Two genes (*TLR5* and *MAP3K20*) had coordinate changes in transcription and H3K27ac at all time points in neutrophils (**Fig. 2a; Supplementary Table 6**), but not in monocytes (**Supplementary Table 7**). In total, four peaks of increased acetylation were assigned to *TLR5* (see **Methods; Fig. 2a**), spanning the promoter region of the gene and a distal site located 57 kb upstream on chromosome 1 (**Fig. 2b**). This latter site has not previously been associated with *TLR5*, but H3K27Ac levels here were strongly correlated with *TLR5* transcription (**Fig. 2b; Supplementary Fig. 7**).

To assess sources of variation in gene expression and epigenetics among CF patients at the individual level, we integrated all data layers for neutrophils and monocytes using multi-omics factor analysis (MOFA)[9] and identified eight latent factors accounting for 30-60% of the variance across data types (**Fig. 2c, Supplementary Fig. 8**). Correlation with phenotypic data from the longitudinal clinical dataset showed that neutrophil count, monocyte count, immunoglobulin G (IgG), serum iron (Fe), and alanine transaminase (ALT) levels were all significantly correlated with latent MOFA factors

(**Fig. 2c**), indicating that phenotypic variation among CF patients is reflected at gene expression level and in the epigenetics of innate immune cells.

As all patients in this study were infected with PsA, we sought to assess the extent to which direct detection of PsA by myeloid cells was responsible for the observed chnages. Neutrophils from three healthy volunteers were exposed to PsA or purified PsA flagellin (a known ligand for TLR5[10]) and their transcriptional response was determined using RNA-seq (**Supplementary Fig. 9**).

In total, we identified 1,810 DE genes in PsA-exposed neutrophils vs unexposed controls (894 up, 916 down) and 328 DE genes in flagellin-exposed neutrophils vs unexposed (246 up; 82 down). As expected, the majority (86%) of the genes responsive to flagellin were also differentially expressed in PsA-exposed cells ($P$ = 8.5 × 10-207) (**Fig. 2d**).

PsA-upregulated genes overlapped with genes overexpressed significantly in CF neutrophils at day 0 (86 genes, $P$ = 5.0 × 10-13) and at day 30+ (7 genes, $P$ = 0.01), but not at day 14 (2 genes, $P$ = 0.31; **Fig. 2**; **Supplementary Table 8**). In total, approximately 13% of all CF-upregulated genes could be directly attributed to PsA exposure. No statistically significant overlaps among downregulated genes were observed (P > 0.05 in each case) (**Supplementary Fig. 10**).

Genes upregulated in PsA-exposed neutrophils and CF patient neutrophils formed a highly connected functional network (STRING[11] protein-protein interaction enrichment $P$ value < 1 × 10-16). The largest connected component of this network included a subset of genes directly responsive to flagellin exposure, including *CD59*, *IL1RN*, *IRAK2*, *IRAK3*, *MAP4K4* and *TNFAIP6* (**Fig. 2e**), implicating this bacterial product as an important inflammatory driver.

In summary, we have shown that epigenetic and transcriptional changes in the myeloid cells of CF patients fluctuate temporally during the infection cycle and influence phagocyte function. Reducing bacterial burden through antibiotic treatment causes rapid ablation of the epigenetic marks that distinguish CF from HV during APE. Moreover, we defined a functionally connected network of genes, which is directly responsive to PsA exposure. Targeting this network could therefore offer a rational strategy for reducing inflammation-associated pathology in CF.

## References

1. Netea, M. G. *et al.* Trained immunity: A program of innate immune memory in health and disease. *Science* **352**, aaf1098 (2016).

2. Netea, M. G. *et al.* Defining trained immunity and its role in health and disease. *Nat. Rev. Immunol.* **20**, 375–388 (2020).

3. Mulder, W. J. M., Ochando, J., Joosten, L. A. B., Fayad, Z. A. & Netea, M. G. Therapeutic targeting of trained immunity. *Nat. Rev. Drug Discov.* **18**, 553–566 (2019).

4. Patel, A. A. *et al.* The fate and lifespan of human monocyte subsets in steady state and systemic inflammation. *J. Exp. Med.* **214**, 1913–1923 (2017).

5. Pillay, J. *et al.* In vivo labeling with 2H2O reveals a human neutrophil lifespan of 5.4 days. *Blood* **116**, 625–627 (2010).

6. de Laval, B. *et al.* C/EBPβ-Dependent Epigenetic Memory Induces Trained Immunity in Hematopoietic Stem Cells. *Cell Stem Cell* **26**, 657–674.e8 (2020).

7. Moorlag, S. J. C. F. M. *et al.* BCG Vaccination Induces Long-Term Functional Reprogramming of Human Neutrophils. *Cell Rep.* **33**, 108387 (2020).

8. Goss, C. H. & Burns, J. L. Exacerbations in cystic fibrosis. 1: Epidemiology and pathogenesis. *Thorax* **62**, 360–367 (2007).

9. Argelaguet, R. *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).

10. Yoon, S.-I. *et al.* Structural basis of TLR5-flagellin recognition and signaling. *Science* **335**, 859–864 (2012).

11. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

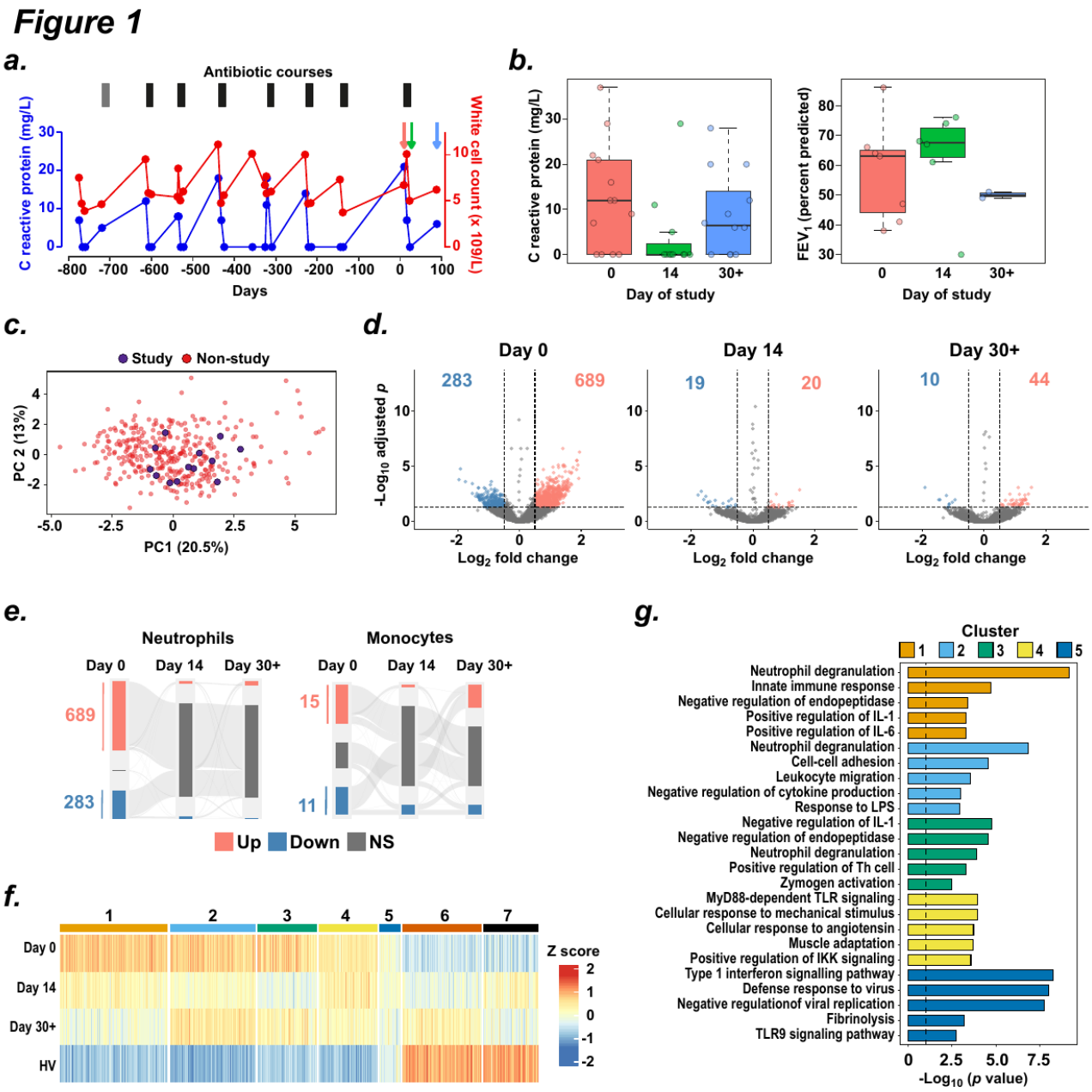



**Figure 1. (a)** Longitudinal fluctuation in c-reactive protein (CRP) and total white blood cell (WBC) count during periods of acute disease exacerbation over a three-year period in a patient recruited for this study. Red arrow indicates Day 0; green arrow indicates day 14; blue arrow indicates Day 30+. **(b)** Levels of CRP and percent predicted forced expiratory volume in one second ($FEV_1$ % predicted) for study patients at each sampling day. **(c)** Principal component analysis (PCA) of curated longitudinal clinical data set for 354 CF patients (Papworth hospital cohort). Patients selected for inclusion in this study are indicated in purple. **(d)** Volcano plots show neutrophils differential gene expression in CF relative to HV at the indicated time points. Genes with increased expression in CF are shown in red, while genes with decreased expression are shown in blue. **(e)** Gene expression dynamics in neutrophils and monocytes. Parallel set diagrams show the changes in numbers of differentially expressed genes over sampling days. The width of the connector reflects the number

of genes in each set. **(f)** Unsupervised hierarchical clustering of neutrophils DE genes. The heatmap depicts relative expression levels of all DE genes between CF and HV in at least one time point. Median z-scores per sample group are plotted for each gene. Gene clusters are numbered and indicated with coloured bars above heatmap. **(g)** GO terms functional enrichment of neutrophil DE genes by cluster. Bar plot shows the −log10(*P* value) per enriched term. Dashed line represents *P* value = 0.05.

# Figure 2

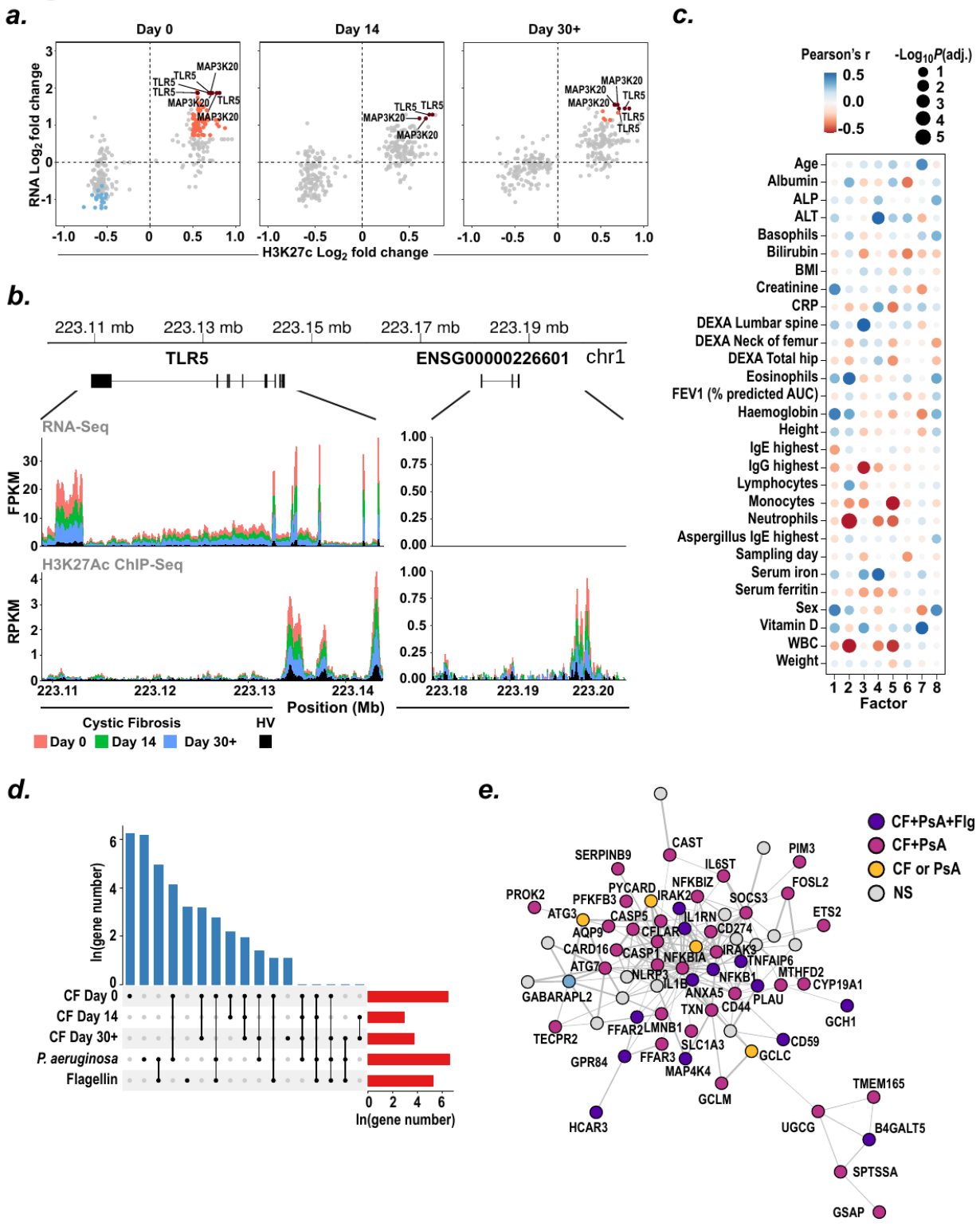



**Figure 2. (a)** Comparison of RNA-Seq and ChIP-Seq log2 fold change values by sampling day. All differentially acetylated H3K27Ac peaks assigned to genes are shown as dots. Red dots

correspond to peaks and transcripts for genes significantly increased in CF relative to HV, blue dots correspond to peaks and transcripts significantly decreased in CF relative to HV, and grey dots represent peaks and transcripts for genes which are not significantly different from HV in one or both data sets. **(b)** RNA-Seq and H3K27Ac ChIP-Seq coverage of the *TLR5* gene (left panel) and distal region of acetylation (right panel). The total coverage is shown as the median fragments per kilobase per million (FPKM) for RNA-Seq, and median reads per kilobase per million (RPKM) for ChIP-Seq, in 40 bp bins. Red bars, CF day 0 samples; green bars, CF day 14 samples; blue bars, CF day 30+ samples; black bars, HV samples. **(c)** Correlation of clinical data with MOFA factors. Colour scale represents the strength of correlation (Pearson *r*) and point size represents the associated *P* value, with multiple testing correction performed using the Benjamini-Hochberg method. Abbreviations: ALP, alkaline phosphatase; ALT, alanine transaminase; Asp, Aspergillus; AUC, area under curve; BMI, body mass index; CRP, C-reactive protein; FEV1 % pred., forced expiratory volume in one second (FEV1) percent of predicted; IgE, immunoglobulin E; IgG, immunoglobulin G; RAST, radioallergosorbent; WBC, total white blood cell count. **(d)** UpSet plot shows number of neutrophils upregulated genes in *P. aeruginosa*-infected HV neutrophils, flagellin-exposed HV neutrophils, and in CF at Day 0, Day 14, and Day 30+. **(e)** Functional network of genes commonly upregulated in *P. aeruginosa*-infected HV neutrophils (PsA) and CF neutrophils, constructed using interactions from the STRING database. A subset of genes in this network are also upregulated upon flagellin (flg) exposure. The edge width reflects the combined score of the interaction from STRING, and the node colour indicates the contrasts in which a gene is upregulated. NS, not significantly differentially expressed.

## Methods

### Study design and participants

Patients with Cystic Fibrosis (CF) were enrolled from Royal Papworth Hospital, UK, after written informed consent (ethical approval REC 19/EE/0241). Peripheral blood samples were obtained at pre-defined time-points during an infective acute pulmonary exacerbation (APE) of CF. An APE was defined as a change in the patient's symptoms from baseline. These symptoms include increased cough, increased sputum production, change in sputum colour, change in sputum thickness and fevers or temperatures. A fall in forced expiratory vital capacity in 1 second (FEV1) was used as an additional indicator. Patients who reported 2 or more of the above symptoms and where a clinical decision was made to start intravenous (IV) antibiotics, were eligible for the study. Blood samples were taken at the start of an APE (Day 0), at the end of an APE (Day 14) and at a later time point during a period of clinical stability (Day 30+). Patient selection was based on their colonising organism (*Pseudomonas aeruginosa*), stability on current medications for 3 months with no planned changes during the study period, and their FEV1. Age and sex matched healthy volunteers were enrolled.

### Sample collection and processing

Peripheral blood samples were collected and processed using established protocols from the BLUEPRINT consortium[1]. In brief, peripheral blood was collected in citrate tubes. Plasma was separated out by centrifugation. Peripheral blood mononuclear cells (PBMC) were extracted using a density gradient (Ficoll, GE healthcare) and purified for CD14 positive monocytes using magnetic microbeads (Militenyi Biotec). The remaining granulocytes underwent ammonia red cell lysis with additional washing stages. Cells were checked for their purity by using cytospins for morphology in combination with flow cytometry analysis for cell surface markers. Serum was collected in separate serum blood tubes and extracted via centrifugation.

### RNA processing and sequencing

Purified CD 14+ monocytes and neutrophils were stored in TRIzol. RNA was extracted from TRIzol, using BLUEPRINT protocols[1]. One hundred nanograms of RNA was converted to rRNA-depleted c-DNA libraries using the KAPA Stranded RNA-Seq Kit with RiboErase (Roche). Samples were indexed with Tru-seq adapters (Illumina), and 150-bp paired end sequencing was performed on Illumina's NovaSeq platform.

238 **CHiP processing and sequencing**

239 Purified CD 14+ monocytes and neutrophils were cross-linked with formaldehyde and processed
240 using BLUEPRINT protocols[1]. Cells were lysed, nuclei prepared and sonicated using the Bioruptor
241 Pico (Diagenode). Sonicated chromatin was pre-cleared using Dynabeads, Protein A magnetic
242 beads (Invitrogen) before proceeding to automated chromatin immunoprecipitation (SX-8G IP Star
243 Compact, Diagenode). Antibodies used were H3K27ac (Diagenode). Samples underwent reverse
244 cross linking and DNA capture, using the ChIP DNA clean and concentrator kit (Zymo Research).
245 DNA libraries were prepared using the Diagenode microplex library kit. Samples were dual indexed
246 with MicroPlex adaptors (Diagenode), and 50-bp single read sequencing was performed on
247 Illumina's HiSeq 4000 platform.

248

249 ***Pseudomonas* Experiment: Cell preparation**

250 Neutrophils were isolated from whole blood using the EasySep™ Direct Human Neutrophil Isolation
251 Kit (Stem Cell technologies) as per the manufacturer's protocol. Freshly isolated neutrophils were
252 re-suspended in Iscove's Modified Dulbecco's Medium (IMDM) without phenol red (Gibco) and 10%
253 autologous serum, at a concentration of 5x106 neutrophils/ml. Autologous serum was prepared by
254 collecting 5-10 mls of whole blood in a 10 ml syringe, gently transferring to a sterile 15ml Falcon tube
255 and allowing the blood to clot prior to centrifugation to obtain the serum layer.

256

257 ***Pseudomonas* Experiment: Reagent preparation**

258 100 µL of nuclease free water (Qiagen) was added to 50 µg purified flagellin from Pseudomonas
259 aeruginosa (FLA-PA; InvivoGen), giving a stock concentration of 500 µg/ml. 100 µL of nuclease free
260 water was added to 50 µg of a soluble ectodomain of human TLR5 (hTLR5-Fc; InvivoGen), giving a
261 stock concentration of 500 µg/ml.

262

263 ***Pseudomonas* Experiment: Bacterial preparation**

264 Pseudomonas aeruginosa (PsA) PAO1 (ATCC) was cultured on a Columbia blood agar plate
265 overnight at 37 oC at 5 % CO2. A single colony-forming unit (CFU) from the plate was removed and
266 cultured in 10 mLs of Luria broth (LB), in a 50ml falcon tube, overnight to an OD600 of 0.5-0.8. A
267 multiplicity of infection (MOI) of 1:1 was used.

268

269

270

271

### *Pseudomonas* Experiment: Cell culture

180 µl of the neutrophil suspension was gently placed in a 2 mL round-bottom DNA LoBind Eppendorf tube (Eppendorf). Neutrophils were cultured with either phosphate buffered saline (PBS; Sigma), FLA-PA, hTLR5-Fc, PsA, hTLR5-Fc followed by PsA or hTLR5-Fc followed by FLA-PA. Cells were placed in a shaking incubator at 180 rpm at 37oC. In the case of hTLR5-Fc with PsA or FLA-PA, the hTLR5-Fc was added 10 minutes prior to the addition of PsA or FLA-PA.

Cells were returned to the shaking incubator. After 1 hour 200 µg/ml Gentamicin and Strepatmycin (Sigma) was added to the neutrophil suspension. The neutrophils were returned to the shaking incubator. After 4 hours, cells were pelleted by centrifugation, placed on ice, the supernatant removed and cells suspended in Trizol (Invitrogen). RNA processing and sequencing were carried out as outlined above.

### Monocyte derived macrophages: Cytokine production

Experiments were conducted using previously established protocols[2]. Frozen monocytes collected at Day 0, 14 and 30+ and stored in recovery media (Gibco) were thawed as per the manufacturer's protocol and differentiated into macrophages in autologlous serum for the corresponding time points. Monocytes for CF005, CF007, CF008 and Day 0 for their matched HV samples (HV005, HV007 and HV004) were selected. Cells were differentiated in 24-well tissue culture plates at a cell density of 0.2 x106 cells per well using granulocyte-macrophage colony-stimulating factor (GM-CSF; PeproTech EC Ltd) 200 ng/ml, 10% autologous serum from each time point, 100 U/ml penicillin, and 100 µg/ml streptomycin (Sigma) in DMEM media.

Cells were maintained at 37°C with 5% CO2. On day 5 interferon-γ (IFN γ; PeproTech EC Ltd) 50ng/ml was added. On day 7, antibiotics were removed by washing in PBS, the macrophages were infected with PsA (PAO1; ATCC) at an MOI of 1:1 in DMEM media. The supernatant was removed at 4 hours post-infection. Cytokine concentration for IL-6, IL-8 and TNF α, of the supernatant, was measured using the human cytokine magnetic kit (Milliplex) as per the manufacturer's protocol. Samples were analysed on the Luminex200. The average of three independent replicates was analysed.

### Processing of clinical data

C-reactive protein (CRP) measures of 4 ml/L or less were considered to be 0 mg/L. For principal component analysis (PCA), median values of each variable were calculated across the time series of measurements per patient, except for the ratio of FEV1/FVC, for which the area under the curve

(AUC) was estimated using the trapezoid rule. Missing values were imputed using the multivariate imputation via chained equations (MICE) package in R (https://doi.org/10.18637/jss.v045.i03).

**RNA-Seq data analysis**

Adaptors were trimmed from reads using Trim Galore v0.3.7 (https://github.com/FelixKrueger/TrimGalore), with the following parameters: -q 15 --stringency 3 --length 20 -e 0.05 --trim1 -a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA; -q 15 --stringency 3 -a AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT. Adaptor-trimmed reads were aligned to transcripts from the human reference genome GRCh38 using bowtie v1.0.1[3], with the following parameters: -a --best --strata -S -m 100 -X 500 --chunkmbs 256 --nofw –fr.

Transcript and gene abundance estimates were generated from bowtie alignments using mmseq v1.0.10[4]. For differential expression analysis, only patients with complete data at each time point (patients CF004, CF005, CF007, CF008, CF010, CF011, CF012, CF013) and only genes with an FPKM of at least 1 in at least 1 sample were included. Differentially expressed (DE) genes were identified using the Wald test function in the DESeq2 package[5]. Sex and age were used as covariates in the model. Genes with adjusted P-value < 0.05 and absolute log2 fold change > 0.5 were considered differentially expressed.

Hierarchical clustering of DE genes was carried out using median z-scores calculated from log2(FPKM) values per time point. The number of gene clusters was chosen using the approach of Marriott[6]. Hierarchical clustering was performed using Manhattan distances and the Ward method. Enrichment of gene ontology (GO) biological processes (BP) in clusters was assessed using the topGO package in R, using the weight01 algorithm, and in each taking case the remaining set of expressed non-cluster genes as the background set. GO terms containing a minimum of 10 genes were included.

**ChIP-Seq data analysis**

Adaptors were trimmed from reads using Trim Galore v0.3.7 (https://github.com/FelixKrueger/TrimGalore), with the following parameters: -q 15 --stringency 3 --length 25 -e 0.05 --trim1 -a CTGTCTCTTATACACATC;-q 15 --stringency 3 -a AAGCAGTGGTATCAACGCAGAGT. Adaptor-trimmed reads were aligned to the human reference genome GRCh38 using bwa v0.7.12[7]. Duplicate aligned reads were marked using Picard v2.0.1 (https://github.com/broadinstitute/picard), and alignments were de-duplicated and quality-filtered using samtools v1.3.1[8], by using the following parameters: view -u -F 1024 -q 15. Peaks were called

on de-duplicated, quality-filtered alignments using MACS2 v2.1.1[9] with parameters -q 1e-2 -g 3049315783. Peaks were annotated based on distance to the nearest TSS using the R package ChIPseeker[10] with Ensembl gene annotations for GRCh38.

Peaks of H3K27Ac were categorised as either promoter (≤ 2 kb from the nearest annotated transcription start site [TSS]) or distal (> 2 kb from the nearest TSS). From the union set of peaks detected across all samples, high-confidence peaks were selected for downstream analysis as those with an RPKM of at least 5 in at least 1 sample (for neutrophils) and an RPKM of at least 7 in at least 1 sample (for monocytes). These thresholds were chosen because peaks passing them were detected relatively evenly across samples, while peaks at lower RPKM thresholds were only detected in a subset of the samples. In total, 32,688 high-confidence peaks (16,745 promoter, 15,943 distal) were identified in neutrophils and 37,048 high-confidence peaks (17,808 promoter, 19,240 distal) were identified in monocytes.

Differential acetylation was assessed using the Wald test function in the DESeq2 package[5]. For differential acetylation analysis, only patients with complete data at each time point (patients CF004, CF005, CF007, CF008, CF010, CF011, CF012, CF013) and only high-confidence peaks were included. Sex and age were used as covariates in the model. Peaks with adjusted P-value < 0.05 and absolute log2 fold change > 0.5 were considered differentially acetylated.

To assign peaks to genes, we combined multiple data sets as follows. All distal peaks located within 10 kb of the nearest TSS were assigned to the gene transcribed from that TSS. To assign peaks located more than 10 kb from the nearest TSS, we assessed the degree of correlation between the normalised ChIP-Seq coverage in each peak and the normalised RNA-Seq coverage for all genes within 100 kb upstream or downstream of the peak, across all available samples, and if a significant correlation (adjusted P-value < 0.05) was found, we assigned peaks to the correlated genes. Finally, peaks that were located in a region known to be distally connected to a promoter from publically available HiC data[11] were assigned to the connected genes.

**Data integration (MOFA)**

For integration of the combined transcriptomic and epigenetic data sets using MOFA[12], the top 10,000 most variable features were selected based on median absolute deviation (MAD) for protein-coding genes and ChIP-Seq peaks. For non-coding RNA, the top 1,000 genes based on MAD were used.

377 **Analysis of *Pseudomonas* data**

378 RNA-Seq data were processed exactly as for CF patient data. For principal component analysis, the
379 effect of donor sex was removed from raw gene expression values using the sva package in R. The
380 significance of overlaps of DE gene lists was determined using Fisher's exact test. For overlapping
381 DE genes with CF patient data, only genes expressed in both data sets (FPKM ≥ 1 in ≥ 1 sample
382 from each data set) were included. Gene Ontology (GO) term biological process (BP) enrichment
383 was assessed using the topGO package in R, with the weight01 algorithm, in each case taking as
384 the background gene list the remaining set of genes expressed in both data sets. GO BP terms
385 containing a minimum of 10 genes were included.

386

387 To construct a functional network of shared DE genes, interactions for the 87 genes commonly
388 upregulated in CF (at any time point) and in PsA-stimulated cells were retrieved from the STRING
389 database[13], with at least medium confidence (minimum combined score = 0.4) and allowing a
390 maximum of 20 additional directly interacting genes. The largest connected component of the
391 network was visualised using Cytoscape[14].

392

393 **References**

394 1. Adams, D. et al. BLUEPRINT to decode the epigenetic signature written in blood. *Nat.*
395 *Biotechnol.* 30, 224–226 (2012).

396 2. Hepburn, L. et al. Innate immunity. A Spaetzle-like role for nerve growth factor β in vertebrate
397 immunity to Staphylococcus aureus. *Science* 346, 641–646 (2014).

398 3. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment
399 of short DNA sequences to the human genome. *Genome Biol.* 10, R25 (2009).

400 4. Turro, E. et al. Haplotype and isoform specific expression estimation using multi-mapping RNA-
401 seq reads. *Genome Biol.* 12, R13 (2011).

402 5. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
403 RNA-seq data with DESeq2. *Genome Biol.* 15, 550 (2014).

404 6. Marriott, F. H. Practical problems in a method of cluster analysis. *Biometrics* 27, 501–514
405 (1971).

406 7. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.
407 *Bioinformatics* 25, 1754–1760 (2009).

408 8. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079
409 (2009).

410 9. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137 (2008).

10. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31, 2382–2383 (2015).

11. Javierre, B. M. et al. Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* 167, 1369–1384.e19 (2016).

12. Argelaguet, R. et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* 21, 111 (2020).

13. Szklarczyk, D. et al. The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49, D605–D612 (2021).

14. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504 (2003).

## Data availability

Sequence data have been deposited at the European Genome-phenome Archive (EGA) under accession number EGAS00001006421.

## Code availability

Publicly available software was used for all data processing steps, and a detailed description is available in Methods. Data files and code used for clustering and functional enrichment of differentially expressed genes can be accessed at https://doi.org/10.5281/zenodo.7113651.

## Acknowledgements

This work was supported by: University of Cambridge/GSK Varsity research award (A.M.D., O.E., M.F. & R.A.F.); UK Cystic Fibrosis Trust Innovation Hub award (001) (O.E., K.P.B., R.A.F.), NIHR Cambridge Biomedical Research Centre (R.A.F., K.P.B.), British Heart Foundation Senior Basic Science Fellowship (FS/18/53/33863) and BHF Cambridge Centre for Research excellence (RE/18/1/34212) (M.F.); and The Wellcome Trust (107032AIA) (R.A.F., K.P.B.), The authors would like to thank Victoria Higgins for administrative support and Uche Agwo for contracting support.

**Figure 1**

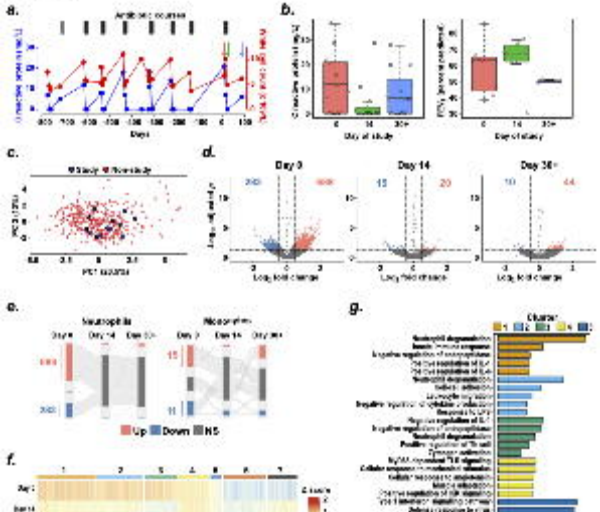
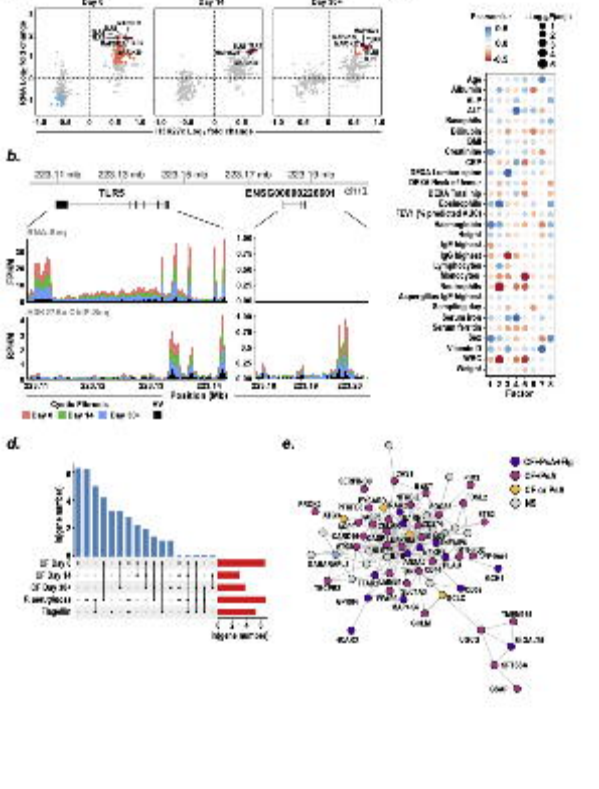

a.
Antibiotic courses
Days
b.
Day of study
Day of relapse
c.
PC1 (22.2%)
d.
Day 0
Day 14
Day 28+
Log2 fold change
e.
Neutrophils
Monocytes
Up
Down
NS
f.
Day 0
Day 14
Day 28+
NS
Z score
g.
Cluster

*Figure 2*


a.
Day 0
Day 14
Day 21+
b.
TLR5
ENSG00000223609
Day 0
Day 14
Day 21+
Position (Mb)
c.
Factor
d.
e.
Day 0+Day 14
Day 0+Day 21+
Day 0 or Day 14
NS