Spatial transcriptomics identifies novel *P. aeruginosa* virulence factors

Hao Zhou[1], Oscar Negrón[1*], Serena Abbondante[2*], Michaela Marshall[2*], Brandon Jones[1], Edison Ong[1], Nicole Chumbler[1], Christopher Tunkey[1], Groves Dixon[1], Haining Lin[1], Obadiah Plante[1**], Eric Pearlman[2**], and Mihaela Gadjeva[1]

Affiliations

[1]Moderna, Inc., 325 Binney St, Cambridge, MA 02142

[2]University of California Irvine, School of Medicine, 843 Health Sciences Rd, Irvine, CA 92697

***,**These authors have contributed equally**

Corresponding author

Mihaela Gadjeva, PhD

Moderna Inc., 325 Binney St, Cambridge, MA 02142

Email: mgadjeva@modernatx.com

Abstract

The interactions of a pathogen with the host take place within the confines of infected tissues. However, current technologies do not allow for contextual characterization of infection. We report a unified spatial transcriptomic analysis detecting pathogen- and host-transcripts simultaneously. We showcase our approach using a *P. aeruginosa*-induced ocular infection, where we observed differential enrichment of host and bacterial transcripts at specific anatomical sites. The ridge regression model, trained using a minimum of 150 host gene features, accurately predicted bacterial burden in tissue regions. The tissue-in-depth gene expression enrichment analysis identified bacterial transcript PA2590, encoding a currently uncharacterized gene as deeply penetrant. Infection experiments with the PA2590 transposon deletion mutant strain caused less disease than the WT strain. Comparative structural analysis identified the PA2590 gene product as an iron-scavenging and cobalamin transporter. Cumulatively, our data highlight coordinated spatial interplay between the host and the pathogen representing an approach to identify novel virulence traits.

Introduction

Pseudomonas aeruginosa is an opportunistic Gram-negative bacterium that belongs to the group of ESKAPE pathogens, posing a growing challenge for treatment due to the rapidly increasing acquisition of antimicrobial resistance (AMR) (1). Our preliminary analysis points to a 70% increase in antimicrobial resistance acquisition among *P. aeruginosa* clinical isolates globally, highlighting it as a major health concern. *P. aeruginosa* causes a diverse range of infections including cystic fibrosis, wound and burn related skin infections, urinary tract infections, and acute pneumonia (2). *P. aeruginosa* is also a major cause of contact-lens and trauma-induced corneal infections worldwide. While intact corneas are typically resistant to infection, ocular trauma and poor contact lens hygiene are risk factors that can allow bacterial penetration to the corneal stroma (3). Infection progresses very rapidly and can lead to tissue damage that manifests as severe ocular pain, visual impairment, and can cause permanent blindness if left untreated (4).

Currently, there are no current available vaccines against *P. aeruginosa.* Most of the preclinical studies leveraged individual virulence factors to generate subunit-based vaccine candidates including T3SS components (e.g, PcrV), polysaccharides (e.g, alginate, psl, LPS), outer membrane proteins (e.g., OprF), detoxified toxins (e.g., ExoA), or flagella (5), (6), (7). Despite the promising initial preclinical data, none of the clinical trials successfully showed protection, highlighting the need for new discovery pipelines to generate protective immunity in susceptible populations.

Despite the increase in the availability of genomic data from laboratory strains and clinical isolates, there are only two studies using reverse vaccinology pipelines. Bianconi et al. reported the complete PAO1 proteome and predicted 718 surface-exposed and conserved

antigens, identifying 52 potential vaccine candidates (8). That study used only a small number of clinical isolates limiting its impact given the genomic diversity of *P. aeruginosa*. In contrast, Solanaki et.al. utilized 1191 predicted proteomes in their vaccine nomination pipeline and generated a concatemeric vaccine candidate based on a series of *in silico* analysis including location, hydrophilicity, essentiality, antigenicity, immunogenicity, and MHC class I and II epitope predictions (9). However, there have been no follow-up reports addressing the efficacy of the vaccine candidates.

Another limitation of the reverse vaccinology pipelines is the lack of consideration for bacterial transcriptional plasticity, functional redundancy of metabolically adaptative mechanisms, and temporal transcriptional changes that the pathogen undergoes during infection. Bacterial transcriptome analysis demonstrated that *P. aeruginosa* transcriptional profiles depend on the host and/or the site of infection. For example, *in vitro* bulk transcriptional responses of bacteria taken from growth media were different from the *in vivo* responses in the infected lungs (10,11). Exposure to Cystic Fibrosis sputum induced a distinct transcriptional profile, illustrating significant transcriptional plasticity (12). Cumulatively, a detailed understanding of the transcriptional behavior of the pathogen may identify new antigens for vaccine development and offer an approach for antigen prioritization.

To achieve global profiling of host-pathogen interactions, we leveraged the resources of spatial transcriptomics technology and simultaneously characterized transcriptional responses of both the host and the pathogen *in situ*. We based these studies on the hypothesis that by gaining insights into the host-pathogen specific temporal and spatial transcriptional gradients, new patho-adaptative mechanisms can be defined.

Results

Spatial transcriptomics reveals tissue-segment specific host transcript enrichment.

To generate a spatial transcriptional profile of infection, we used a well-characterized murine model of *P. aeruginosa*- corneal infection. We utilized the Nanostring GeoMx platform for non-biased characterization of the host transcriptome (Fig. 1A). Sections from paraffin-embedded eyes were processed and we examined the responses in the corneal epithelium, the corneal stroma, and the anterior chamber as regions of interest (ROIs). Workflow is shown in Fig. 1A.

Corneal epithelial cells were identified as keratin 6^{+}ve; neutrophils were identified as $Ly6G^{+}$ve. Uninfected (contralateral eyes) and infected ocular tissues were examined (Fig. 1B). Transcriptome analysis revealed common and segment-specific enrichment of host and bacterial transcripts. Consistent with the rich neutrophil infiltration into the infected tissues, neutrophil-specific S100a8 transcripts were highly upregulated in the epithelium and stroma (Fig. 1C). Also upregulated were neutrophil and monocyte chemokines Cxcl2, and Ccl3, pro-inflammatory cytokines Il-1a and Il-1b, and the neutrophil iron chelating gene Lcn2. Cxcl2, Ccl3, Csf3, Cxcr4, and Il-1b, Il-1a, Il-12a cytokine transcripts were higher in the epithelium relative to the stroma (Fig. 1C). These data were reflected in the Gene Ontology (GO) enrichment analysis showing stronger granulocyte chemotaxis enrichment in epithelium versus stroma.

In contrast, transcripts associated with RNA metabolism, processing, ncRNA processing, and chromatin remodeling were predominantly upregulated in the stromal layer when compared to epithelium. Surprisingly, the infected stroma showed elevated anti-viral responses (Fig. 1D), including Type I IFN, IFNAR, and others (Suppl. Table 1).

After 48 hours of infection rapid disease progression was observed, as illustrated by elevated corneal opacity of the infected corneas (Suppl. Fig. 1A and B). The epithelial layer displayed enrichment of transcripts associated with metabolic alterations including glucose catabolism and cellular responses to glucose starvation, translation, and RNA processing (Suppl. Fig. 1D and Suppl. Table 2). In the stroma, we found sustained enrichment of transcripts associated with neutrophil recruitment (chemotaxis and migration), glucose catabolism, pyruvate metabolism and apoptosis (Suppl. Fig. 1C and Suppl. Table 2). Cumulatively, our data provide a temporal transcriptional and spatial record of tissue alterations during infection that are associated with neutrophil cytokine production and metabolic changes.

Spatial analysis of bacterial transcript abundance in infected corneas reveals distinct enrichment profiles between the epithelium and stroma

Probes were generated for 100 *P. aeruginosa* genes that were encoded outer membrane proteins, extracellular proteins, reported virulence factors, and unkown, predicted genes. Probe development was based on the criteria of ORF frequency in >11,000 clinical isolates, evolutionary conservation, and their relation to known virulence. Three different gene-specific probes were designed for each of the bacterial genes of interest and used simultaneously to identify bacterial transcripts in the infected tissues (Suppl. Table 3).

Transcript expression was examined in PAO1 infected eyes in the positive corneal epithelium compared with the stroma and anterior chamber (AC). Representative images showed that infection caused loss of the cytokeratin staining in some of the surface layers of the cornea, consistent with the corneal perforation (ulceration) that occurs following infection. The underlying stromal layers showed rich $Ly6G^{+}$ staining, indicative of neutrophil infiltration to

infected cornea (Fig. 2B). Overall, we could distinguish ROIs with high and low bacterial transcript presence. This was reflective of the differences in the number of bacteria at distinct sites in the infected eye. While 97 transcripts were detected in almost half of all examined surface segments, only 37 transcripts were detected in most of the stromal segments. Interestingly, nine transcripts were consistently detected in the cytokeratin^{+} and cytokeratin^{-} surface layers, Ly6G^{+} and Ly6G^{-} stromal regions, and AC segments, including AhpC, TufB, PhzD1, ArcA, FpvB, PA2590, IcmP, LipA, and FemA (Fig. 2A and B).

Bacterial transcript enrichment analysis, normalized to the housekeeping RplL transcript abundance, reflected transcript increases in the infected stromal segments when compared to the infected surface layers (Fig. 2C). Nineteen transcripts showed significant upregulation in the infected tissues (Suppl. Table 4). Those included transcripts with known functions such as ArcA, PhzD1, IcmP, OprC, AhpC, LipA, ExoT, PhuR, OpdH, FpvB, and FemA, and transcripts encoding proteins that are predicted but have not been characterized: PA2590, PA0439, PA2204, PA2291, and PA4881. In contrast, transcripts for AlgE, PopD, and Piv were decreased in the stroma, while ExoS and FecA showed no enrichment (Fig. 2C and Suppl. Table 4).

To determine if bacterial transcript tissue penetrance is strain-specific, corneas were infected with the cytotoxic (ExoU expressing) PA14 strain. No significant enrichment of bacterial transcripts was observed at 24h post-challenge (data not shown); however, at 48h post-challenge, deeper bacterial transcript penetrance and significant changes were detected. Fourteen transcripts were similarly enriched between the PA14 and PAO1 data sets. These included PhzD1, FpvB, PA2590, IcmP, ArcA, FemA, LipA (Fig. 3B, C, and D). The transcriptional heat map comparisons between the sham and infected tissues confirmed the specificity of bacterial transcript detection (Fig. 3A and Suppl. Table 5). Cumulatively, these results highlight the

appearance of bacterial transcriptional patterns in the infected tissues irrespective of the strain types and thereby illustrate spatial organization during infection.

Harnessing host-bacteria dynamics for accurate prediction of infection levels in tissue regions.

Next, leveraging the host transcriptional data, we examined its functional relationship with the *P. aeruginosa* housekeeping gene RplL, which we used as a correlate for the number of bacteria in the infected tissues. Using Ridge Regression analysis (Fig. 4A) we demonstrated significant predictive accuracy, achieving an R^2 score of 0.923. The selection of 600 key host gene transcript features (Fig. 4B) refined the model, enhancing our understanding of the spatial distribution of bacterial burden across infection sites.

We found that 150 host features were sufficient to accurately predict bacterial burden measured based on RplL transcript abundance, with an R^2 score of >0.8. We generated a host-bacterial RplL transcript correlation network and examined it for known host protein interactions (Fig. 4C). We found multiple signaling mediators associated with type I IFN, IL-1a, and TNF-driven inflammatory responses, which were highly interconnected for protein-protein interactions, further indicating the observation that the identified host response network reflected highly coordinated immune mechanisms (Fig. 4C). Examples for individual strong correlations between *P. aeruginosa* RplL and TNF-α, Trim13, and phorbol-12-myristate-13-acetate-induced protein1 (Pmaip1) are shown (Fig. 4D), and demonstrate that this approach can identify known (e.g., Il-1α, TNF-α) and unexpected immune determinants of infection including Trim13 and Pmaip1. Employing ‘leave-one-out' validation on our model with a set of mouse hosts infected by PAO1 or PA14 strain, as illustrated in Suppl. Figure 2, showcases its adaptability and

potential for broad applicability. This approach involves training the model on data from all but one host and assessing its predictive accuracy on the excluded host. The successful application of this technique across biological replicates demonstrates the model's robust predictive capacity.

Cumulatively, our approach offers a detailed perspective on the variability of bacterial load across tissue regions and demonstrated the ability to predict bacterial tissue burden by select host transcript feature abundance, contributing to the field of precision medicine in diagnosing bacterial infections.

Infection with the PA2590Tn deletion mutant strain is associated with reduced pathology and bacterial burden.

Bacterial enrichment analysis identified known and unknown *P. aeruginosa* transcripts that were enriched in the infected tissues. To determine the functional importance of the predicted PA2590 gene that showed significant penetrance in the infected tissues from PAO1 and PA14-challenged mice, corneas were infected with *P. aeruginosa* PA14 WT parental strain or the PA2590 transposon mutant (PA2590Tn). Infected eyes were imaged, and corneal opacity, was imaged and quantified (Fig. 5). We also assessed viable bacteria by CFU. Since we established similar transcriptional enrichment for the PA2590 in the infected tissues of PAO1 and PA14-induced infections, infection studies were carried out with the PA14 strain.

There were no significant differences in corneal opacity or bacterial count at 24h post-challenge between the WT PA14 strain and PA2590Tn, consistent with the lack of enrichment for PA2590 at this data point (data not shown). However, the PA2590Tn mutant showed decreased virulence at 48h post-challenge as we recovered significantly less bacteria when compared to WT infected tissues (Fig. 5B, C, and D). As bacterial burden at 48h post-challenge is typically

predictive of the disease outcomes in this model, we concluded that infection with the PA2590Tn mutant causes less severe disease.

Structural comparison analysis suggests that PA2590 transports iron or vitamin B12.

To understand how PA2590 might affect *P. aeruginosa* virulence, we examined its prevalence and conservation, and conducted structural studies. Analysis for PA2590 prevalence in 11,447 publicly available *P. aeruginosa* genomes derived from clinical isolates revealed that the PA2590 ORF is widely distributed and present in at least 97% of all clinical isolates (Fig. 6A). There were no significant differences in the PA2590 distribution based on geographical location of the isolates. Amino acid conservation analysis was also performed using a library of genomes derived from clinical isolates from different clinical presentations such as wound, blood, respiratory, eye, and genitourinary tract infections. A high degree of conservation was identified in all isolates at the amino acid stretches encoding the extracellular loops of the predicted structure of PA2590 with AlphaFold2 (Fig. 6B) (13). Cumulatively, these data suggested an important biological function for PA2590.

Comparative structural analysis established a high degree of conformational similarities between PA2590 and *the E. coli* BtuB protein (PDB: 3m8d; R-Value Free: 0.275) present in *E. coli* (Fig. 6C). BtuB is a vitamin B12 (VB12) transporter, suggesting that PA2590 might have a similar function. Docking experiments revealed that PA2590 is likely to bind VB12 (Fig. 6E, Vina docking: -9.1 kcal/mol). For comparison, the docking score for BtuB and VB12 was -8.2 kcal/mol. In addition to BtuB, our analysis identified structural homology between PA2590 and *E. coli* FepA (PDB: 1fep; R-Value Free: 0.282) (Fig. 6D). FepA is an iron and VB12 transporter (14). The acquisition of iron depends on ferric enterobactin binding to FepA where the

transporter preferentially recognizes the catecholate part of the complex (15). Unlike *E. coli*, our docking experiments did not indicate a potential interaction between enterobactin and PA2590. However, mammalian siderophores can be synthesized by eukaryotic homologues to the *E. coli* enzyme generating enterobactin (16). Therefore, we examined the potential binding of PA2590 to an abundant host siderophore 2,5-dihydroxybenzoic acid (2,5-DHBA). The *in silico* docking experiments confirmed the predicted interaction (Suppl. Fig. 3C, Vina docking: -6.0 kcal/mol).

P. aeruginosa uses two main siderophores for iron uptake: pyoverdine and pyochelin (17). As both known receptors were upregulated in the infected tissues and to determine if there might be redundancy in iron uptake, the potential interaction of PA2590 with Fe(III)-pyochelin was examined. The experiments showed a significant docking score of -8.2 kcal/mol, indicative of a role for PA2590 in the internalization of iron complexed with *Pseudomonas*-derived siderophores (Fig. 6F). Cumulatively, our data indicate that PA2590 mediates both iron and VB12 transport.

Discussion

Spatial transcriptomic profiling of infected tissues provides specific information on gene expression at specific sites of infection. Several different approaches were developed in the last decade to examine spatial proteomics and transcriptomics focusing on characterizing eukaryotic tissues. These tools have not been applied to study microbial pathogenesis. To the best of our knowledge, this report is the first to use dual spatial RNA-seq, combining bacterial pathogen-specific and host transcriptome data with spatial resolution into a unified, multidimensional dataset. This experimental approach represents a significant advancement over previous studies that used either sequential analysis of eukaryotic and prokaryotic features or reported bacterial

analysis based only on 16S RNA seq (18,19). Our data illuminate differential enriched bacterial transcripts at the host-pathogen interface, thereby identifying potentially novel patho-adaptative mechanisms. Acquisition of this type of data enable machine learning approaches to connect host transcriptional networks with bacterial profiles, thereby identifying key nodes for targeted therapies.

Our study uncovered a strong, predictive correlation between a network of interconnected host features and the *P. aeruginosa* house-keeping transcript (RplL) during infection, offering insights into the biological processes underpinning infection dynamics. The abundance of RplL transcripts was used as a measure for bacterial burden during infection, and successfully predicted bacterial abundance in each tissue with a significant degree of accuracy using at least 150 host features. Correlation analysis identified a network of IFN-inducible features: IL-1α, Pmaip1, TNFα, and Trim 13. The network comprises mediators with known contributors to *P. aeruginosa*-induced infection (e.g., IL-1α and TNF) and the previously unrecognized mediators (TRIM13 and Chit1).

P. aeruginosa is a potent inducer of IL-1α and TNFα, cytokines that orchestrate neutrophil infiltration to the infected site (20,21,22). Loss of IL-1α, unlike the loss of IL-1β, promotes resistance to infection (23,24), suggesting that the network identifies key features mediating infection outcomes. Type I IFN induced mediators inhibit IL-1α responses, and, conversely, IL-1α potently antagonizes Type I IFN by regulating IFNβ transcription and translation (25). In the context of infection, when inflammation onset is acute, it is likely that disease outcomes are determined by IL-1α/Type I IFN balance, which are reflected in the predictive analysis.

Type I IFN responses are induced by the cGAS/STING pathway during *P. aeruginosa*-induced infection and deficiencies in either cGAS or STING promote susceptibility to infection (26,27). This pathway controls autophagy, a key determinant of disease outcome (16,28). Functionally, type I IFN prevents *P. aeruginosa*-induced tissue inflammation, thereby preserving tissue integrity and facilitating bacterial clearance (26,29), further, reduced Type I IFN responsiveness by cystic fibrosis (CF) epithelial cells promotes bacterial colonization (30). In our analysis, increase in TRIM13 and Pmaip1 is consistent with known downstream events from Type I IFN signaling (31). TRIM13 inhibits Type I IFNs (26), whereas Pmaip1 mediates IFN-driven keratocyte and neutrophil apoptosis by mcl-1 degradation (32). Cumulatively, these data identified a complex interplay between IL-1 and Type I IFN responses which predict pathogenesis.

Pathogen biofilm gene expression is a function of bacterial location within the biofilm (33). Therefore, we predicted that the bacterial transcriptional characteristics will show distinct features depending on location. Consistent with this hypothesis, bacterial enrichment analysis exhibited a higher tissue penetrance of transcripts associated with iron acquisition. *P. aeruginosa* requires iron to detoxify reactive oxygen species (ROS), and nitric oxide (NO) and facilitates pathogen survival. *P. aeruginosa* makes two major siderophores: pyoverdine and pyochelin, and the latter is essential for acute infection (34). Transcripts for the pyoverdine receptor, FpvB, and ferric mycobactin receptor, FemA were significantly enriched, although the pyochelin receptor FptA did not show significant enrichment in our analysis and suggesting that there are other mediators of pyochelin import.

Interestingly, we observed a significant penetrance of the PA2590 transcript encoding a predicted protein. The structural analysis indicated that PA2590 belongs to a 22-stranded

antiparallel β-barrel domain containing Ton B-dependent transporters and the high degree of structural similarity to BtuB and FemA suggested a possible role for PA2590 in mediating iron intake. Our docking experiments reinforce this prediction by indicating its possible binding to pyochelin. The significance of these data is highlighted by the reduction in the *P. aeruginosa* virulence in the absence of PA2590. Cumulatively, our spatial transcriptome enrichment analysis approach allowed us to identify novel virulence mechanisms associated with key requirements for pathogen survival during infection.

In conclusion, we demonstrate the feasibility to spatially monitor pathogen transcriptome profiles to identify interconnected host-pathogen networks and novel virulence effectors.

Methods

Bacterial strains

P. aeruginosa strains PAO1, PA14, PA14_30590_Tn (PA2590 transposon mutant) were grown in Luria–Bertoni broth (LB) broth (BD-Difco) overnight at 37°C with agitation (200rpm). Cultures were then passaged in fresh LB broth (BD-Difco) until an OD_{600} of 0.20 (midlog phase) was reached. The PAO1 strain was obtained from Dr. A. Rietsch (Case Western Reserve University). The PA14 and PA14_2590_Tn were generated during the PA14 transposon insertion mutant library generation ((http://ausubellab.mgh.harvard.edu/cgi-bin/pa14/home.cgi)) and obtained from Prof. G. Pier (BWH, HMS) (35,36). The PA2590 transposon mutant was sequence validated (data not shown).

Corneal infection model

Male C57BL/6 ("WT") mice were purchased from The Jackson Laboratory (Bar Harbor, ME) and were housed in a B-level UCI vivarium room. Mice were maintained by the Pearlman Lab and UCI ULAR staff. Infected mice were transferred to D-level holding rooms until the experimental endpoint was reached. All protocols were approved by UCI Institutional Animal Care and Use Committee under IACUC protocol AUP-21-123.

Overnight cultures of *P. aeruginosa* were grown to an OD_{600} of 0.20 in LB broth, and cultures were then washed and resuspended in 1x sterile PBS (Corning). C57BL/6J mice were anesthetized with ketamine/xylazine (MWI Animal Health) and corneal epithelium was abraded with three, parallel 10mm scratches, performed with a 26 ½ -gauge needle (BD). *Pseudomonas* was then applied to the abraded cornea topically with 2µl containing $5x10^5$ CFU in PBS in the PA14 infection studies and with 5ml of $1x10^6$ CFU in the PAO1 infection studies. For use in the infection experiments, 100µl of overnight culture was added to 10ml of fresh LB and was subcultured for 2.5-3h to reach OD_{600}=0.20. 1ml of culture was spun down at 3,000 x g (rotor, centrifuge) for 10 min and diluted to a desired inoculum. Experiments were carried out predominantly with male C57Bl6/J mice. After 24h or 48h, mice were euthanized, and corneas were imaged by brightfield microscopy to observe opacification. Bacterial CFU was quantified at 24h and 48h by serial dilutions of whole eye homogenates onto LB plates and were counted manually.

Corneal opacity

Corneal images were taken using a Leica MZ10 F Modular Stereo Microscope with a Leica DFC450 C camera attachment. Image brightness was adjusted manually using ImageJ software (NIH) and was kept consistent for all images. Images were then converted to 8-bit for

quantification. Using ImageJ, ROIs were drawn around the corneal area only (excluding the sclera) and mean intensity was recorded for each ROI. Images were corrected by subtracting intensity caused by glare of dissection lights (glare values were based on naïve corneas). Percent corneal opacity was calculated by dividing corrected mean intensity by the intensity of a completely white cornea.

Bacterial quantification

Once images were taken, eyes were resected and placed into a 2 ml centrifuge tube containing 1ml of sterile 1x PBS and a sterile steel ball bearing (5 mm diameter). Tubes were placed in a Qiagen Tissue Lyser II and were homogenized at 30Hz for three minutes. After lysis, homogenates were serially diluted up to 1:10,000 and 10µl of each dilution was plated onto one LB agar (BD-Difco) plate (concentrations plated: undiluted, 1:10, 1:100, 1:1,000, 1:10,000). Dilutions were streaked down the plate and plates were incubated at 37°C with 5% CO_2 overnight and colony forming units (CFU) were counted manually the next day. CFU was then quantified as CFU/ml.

Ocular tissue processing

After imaging the corneas, eyes had 15µl of 4% formalin (ThermoFisher) dropped onto them and were left to fix at room temperature for 15 minutes. Eyes were then carefully resected and placed in 1ml of iced 4% formalin (ThermoFisher) and were fixed at 4°C for 48-72hrs. After fixation, formalin was removed, and tissue was washed twice with sterile 1x PBS. Samples were then placed in 2ml of 70% ethanol and were kept at 4°C until shipment to UMass Histopathology was

complete. Tissue embedding was done by UMass Histopathology. All tissues were processed using UMass Chan Medical School Morphology Core services following standard protocols.

Target selection and probe designs

Probes targeting 100 selected genes in the PAO1 strain of *P. aeruginosa* were designed using Nanostring in-house scripts (Supplementary Table 1). These probes were designed to specifically bind to *Pseudomonas* genes, avoiding cross-reactivity with mouse genes and other microbes.

Spatial transcriptomics workflow

5um thick sections from formalin-fixed paraffin embedded (FFPE) tissue blocks were mounted onto Superfrost Plus slides and prepped for GeoMx spatial transcriptomic profiling. Briefly, slides were baked at 37° C for 1 h prior to loading on a BOND RXm (Leica) for further baking, rehydration, heat-induced epitope retrieval (ER2 for 20 min at 100° C) and enzymatic digestion (0.1 ug/mL proteinase K for 15 min at 37° C). Tissue sections were then hybridized with the Nanostring mouse Whole Transcriptome Atlas (WTA) and custom *P. aeruginosa* probe sets at 37° overnight. Following hybridization, two, 25min stringency washes were performed to remove off-target probes and sections were subsequently blocked and stained with the following morphology marker antibodies: DNA (Syto13; 488 channel; Nanostring), PanCK (532 channel; Nanostring), OprI (Opr17_VHHFC; 594 channel; Thermo Fisher Scientific) and Ly6G (1A8; 647 channel; Novus Biologicals). Slides were then loaded into the GeoMx DSP instrument where the entire sections were scanned and imaged. Regions of interest (ROIs) were manually drawn on the section images and approved for collection. The GeoMx DSP then directed UV light to each ROI to cleave the barcodes (RNA ID and unique molecular identifier-containing oligonucleotide)

from the bound probes. Illumina i5 and i7 dual-indexing primers (Illumina) were added to the collected barcodes during polymerase chain reaction (8 µl collected oligonucleotides/ROI) to uniquely index each ROI. AMPure XP beads (Beckman Coulter) were used for purification of PCR. Library concentration was measured using a Qubit fluorimeter (Thermo Fisher Scientific), and quality was assessed using a Tape Station (Agilent Technologies).

Illumina Sequencing and Processing

The sequencing depth requirement was determined by aggregating the areas of illumination (measured in μm^2) and multiplying the total area by a factor of 100, as recommended in the GeoMx DSP Library and Sequencing Guide (nanostring.com). Sequencing was executed using the NovaSeq 6000 System. The resultant FASTQ files were processed to Digital Count Conversion (DCC) files utilizing the GeoMxNGSPipeline v.2.3.3.10, a pipeline publicly available through NanoString Technologies. This pipeline facilitated individual probe alignment by identifying read tag sequences (RTS). Additionally, it eliminated duplicates in probe counts generated during the library preparation's polymerase chain reaction, through a meticulous analysis of the unique molecular identifiers (UMIs) associated with each probe.

Quantification and Statistical Analysis

The DCC files underwent analysis using NanoString Technologies' GeoMx™ Digital Spatial Profiler. Key metrics such as the number of raw reads, the alignment rate of these reads, and the sequencing saturation for each area of illumination were scrutinized to evaluate the quality of library preparation. Criteria for inclusion in the analysis were set: more than 10^6 raw reads per ROI, an alignment rate exceeding 80%, and a sequencing saturation greater than 80%. A "no

template control" sample, devoid of target probes, was integrated into the library preparation and subsequent sequencing to gauge potential contamination. To estimate background noise and determine the limit of quantification, negative control probes employing External RNA Controls Consortium (ERCC) sequences were included in each GeoMx™ RNA assay. Target probe data were filtered based on frequency within each ROI, followed by normalization against the highest negative control counts using background subtraction. Differential gene expression analysis across various groups was performed using DESeq2, a statistical method designed to identify significant changes in normalized gene expression levels (37). The Python-based library GOATools were used for the functional enrichment analysis (38).

Analysis of PA2590

To identify the PA2590 gene, we analyzed more than 11,000 genome sequences from clinical isolates collected between 2000 and 2023. These sequences, sourced from the NCBI database, were subjected to BlastP analysis with the criteria of >85% identity and >70% query and >70% subject coverage, to identify homologous genes of PA2590. Following this, Clustal Omega was utilized for sequence alignment (39). The conservation of amino acid residues within PA2590 was quantitatively evaluated using Shannon Entropy (40). This approach effectively highlights conserved regions by examining the variability of amino acid positions across the aligned sequences, offering insights into the evolutionary stability of the gene. MembraneFold was employed to predict membrane topology of PA2590 (41).

Structural homology analysis and molecular docking

For structural insights, we employed Foldseek to align PA2590 (www.uniprot.org/uniprotkb/Q9I0P7/entry#structure) with existing structures in the Protein Data Bank (PDB) in 01/2024 (42). This alignment was instrumental in hypothesizing the functional structure of PA2590, based on similarities with known protein structures. Structural superimposition was conducted using the cmd.align function in PyMOL v2.5.4, involving 1000 alignment cycles with the transform parameter set to 1. *In silico* docking experiments were conducted using Autodock Vina (43). The pose with the lowest energy score (kcal/mol) was chosen for representation. Pymol scripts was used to highlight residues potentially interacting with PA2590, focusing on polar contacts within a 3.5Å range.

Predictive Models and Correlative Analysis

We employed Random Forest (RF), Ridge Regression (RR), and Support Vector Machine (SVM) algorithms for predictive analysis, with these models trained on a complete set of host features (as detailed in Suppl. Fig. 2A). Among these, Ridge Regression emerged as the most accurate model and was selected for further in-depth analysis and optimization.

The Ridge Regression model was trained using host response profiles and subsequently used to predict the expression data of the PA RplL gene. For normalization of data values, a log transformation ($\log_2$(values + 1)) was applied. To assess the model's predictive accuracy, we implemented a five-fold cross-validation strategy, utilizing Regions of Interest (ROIs) from mice that had been infected. Additionally, cross-validation was carried out across different hosts to validate the model's capacity to predict outcomes in hosts at similar infection stages, as detailed in Supplemental Figure 2. The effectiveness of the model was thoroughly evaluated based on the average Mean Absolute Error (MAE), Explained Variance Score (EVS), and R^2 scores.

Further, we assessed the model's performance when trained with varying numbers of mouse gene features. For this, we selected the top-performing features based on their coefficient values in the model trained with the complete dataset. The primary metric for assessing these subset-trained models was the average R^2 score. To optimize these models, based on different sets of features, we applied a grid search technique for hyperparameter tuning, ensuring that each model was fine-tuned for the best possible predictive performance.

We assessed correlations between bacterial RplL and host gene expressions using Spearman's rank correlation, with p-value adjustments made through the Benjamini-Hochberg FDR method.

References

1. Mladenovic-Antic S, Kocic B, Velickovic-Radovanovic R, Dinic M, Petrovic J, Randjelovic G, et al. Correlation between antimicrobial consumption and antimicrobial resistance of Pseudomonas aeruginosa in a hospital setting: a 10-year study. J Clin Pharm Ther. 2016;41(5):532–7.

2. Gellatly SL, Hancock REW. Pseudomonas aeruginosa : new insights into pathogenesis and host defenses. Pathog Dis. 2013 Apr 1;67(3):159–73.

3. Fleiszig SMJ, Kroken AR, Nieto V, Grosser MR, Wan SJ, Metruccio MME, et al. Contact lens-related corneal infection: Intrinsic resistance and its compromise. Prog Retin Eye Res. 2020 May 1;76:100804.

4. Sy A, Srinivasan M, Mascarenhas J, Lalitha P, Rajaraman R, Ravindran M, et al. Pseudomonas aeruginosa Keratitis: Outcomes and Response to Corticosteroid Treatment. Invest Ophthalmol Vis Sci. 2012 Jan 25;53(1):267–72.

5. Inoue K, Kinoshita M, Muranishi K, Ohara J, Sudo K, Kawaguchi K, et al. Effect of a Novel Trivalent Vaccine Formulation against Acute Lung Injury Caused by Pseudomonas aeruginosa. Vaccines. 2023 Jun;11(6):1088.

6. Killough M, Rodgers AM, Ingram RJ. Pseudomonas aeruginosa: Recent Advances in Vaccine Development. Vaccines. 2022 Jul 8;10(7):1100.

7. Döring G, Meisner C, Stern M, Flagella Vaccine Trial Study Group. A double-blind randomized placebo-controlled phase III study of a Pseudomonas aeruginosa flagella vaccine in cystic fibrosis patients. Proc Natl Acad Sci U S A. 2007 Jun 26;104(26):11020–5.

8. Bianconi I, Alcalá-Franco B, Scarselli M, Dalsass M, Buccato S, Colaprico A, et al. Genome-Based Approach Delivers Vaccine Candidates Against Pseudomonas aeruginosa. Front Immunol. 2018;9:3021.

9. Solanki V, Tiwari M, Tiwari V. Prioritization of potential vaccine targets using comparative proteomics and designing of the chimeric multi-epitope vaccine against Pseudomonas aeruginosa. Sci Rep. 2019 Mar 27;9(1):5240.

10. Rossi E, Falcone M, Molin S, Johansen HK. High-resolution in situ transcriptomics of Pseudomonas aeruginosa unveils genotype independent patho-phenotypes in cystic fibrosis lungs. Nat Commun. 2018 Aug 27;9(1):3459.

11. Damron FH, Oglesby-Sherrouse AG, Wilks A, Barbier M. Dual-seq transcriptomics reveals the battle for iron during Pseudomonas aeruginosa acute murine pneumonia. Sci Rep. 2016 Dec 16;6(1):39172.

12. Tata M, Wolfinger MT, Amman F, Roschanski N, Dötsch A, Sonnleitner E, et al. RNASeq Based Transcriptional Profiling of Pseudomonas aeruginosa PA14 after Short- and Long-Term Anoxic Cultivation in Synthetic Cystic Fibrosis Sputum Medium. PLOS ONE. 2016 Jan 28;11(1):e0147811.

13. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature. 2021 Aug;596(7873):583–9.

14. Barnard TJ, Watson ME, McIntosh MA. Mutations in the Escherichia coli receptor FepA reveal residues involved in ligand binding and transport. Mol Microbiol. 2001 Aug;41(3):527–36.

15. Buchanan SK, Smith BS, Venkatramani L, Xia D, Esser L, Palnitkar M, et al. Crystal structure of the outer membrane active transporter FepA from Escherichia coli. Nat Struct Biol. 1999 Jan;6(1):56–63.

16. Moretti J, Roy S, Bozec D, Martinez J, Chapman JR, Ueberheide B, et al. STING Senses Microbial Viability to Orchestrate Stress-Mediated Autophagy of the Endoplasmic Reticulum. Cell. 2017 Nov 2;171(4):809-823.e13.

17. Ghssein G, Ezzeddine Z. A Review of Pseudomonas aeruginosa Metallophores: Pyoverdine, Pyochelin and Pseudopaline. Biology. 2022 Dec;11(12):1711.

18. Galeano Niño JL, Wu H, LaCourse KD, Kempchinsky AG, Baryiames A, Barber B, et al. Effect of the intratumoral microbiota on spatial and cellular heterogeneity in cancer. Nature. 2022;611(7937):810–7.

19. Lyu L, Li X, Feng R, Zhou X, Guha TK, Yu X, et al. Simultaneous profiling of host expression and microbial abundance by spatial metatranscriptome sequencing. Genome Res. 2023 Mar 1;33(3):401–11.

20. Mizgerd JP. Competing Benefits of Tumor Necrosis Factor-α for Bacteria and for Host Defense. Am J Respir Crit Care Med. 2003 Dec 15;168(12):1410–1.

21. Minns MS, Liboro K, Lima TS, Abbondante S, Miller BA, Marshall ME, et al. NLRP3 selectively drives IL-1β secretion by Pseudomonas aeruginosa infected neutrophils and regulates corneal disease severity. Nat Commun. 2023 Sep 20;14(1):5832.

22. Lagoumintzis G, Xaplanteri P, Dimitracopoulos G, Paliogianni F. TNF-alpha induction by Pseudomonas aeruginosa lipopolysaccharide or slime-glycolipoprotein in human monocytes is regulated at the level of Mitogen-activated Protein Kinase activity: a distinct role of Toll-like receptor 2 and 4. Scand J Immunol. 2008 Feb;67(2):193–203.

23. Ratitong B, Marshall ME, Dragan MA, Anunciado CM, Abbondante S, Pearlman E. Differential Roles for IL-1α and IL-1β in Pseudomonas aeruginosa Corneal Infection. J Immunol Baltim Md 1950. 2022 Aug 1;209(3):548–58.

24. Kugadas A, Christiansen SH, Sankaranarayanan S, Surana NK, Gauguet S, Kunz R, et al. Impact of Microbiota on Resistance to Ocular Pseudomonas aeruginosa-Induced Keratitis. PLOS Pathog. 2016 Sep 22;12(9):e1005855.

25. Mayer-Barber KD, Yan B. Clash of the Cytokine Titans: counter-regulation of interleukin-1 and type I interferon-mediated inflammatory responses. Cell Mol Immunol. 2017 Jan;14(1):22–35.

26. Chen K, Fu Q, Liang S, Liu Y, Qu W, Wu Y, et al. Stimulator of Interferon Genes Promotes Host Resistance Against Pseudomonas aeruginosa Keratitis. Front Immunol. 2018 Jun 5;9:1225.

27. Woznica A, Kumar A, Sturge CR, Xing C, King N, Pfeiffer JK. eLife. eLife Sciences Publications Limited; 2021 [cited 2024 Apr 2]. STING mediates immune responses in the closest living relatives of animals. Available from: https://elifesciences.org/articles/70436/figures

28. Yuan K, Huang C, Fox J, Laturnus D, Carlson E, Zhang B, et al. Autophagy plays an essential role in the clearance of Pseudomonas aeruginosa by alveolar macrophages. J Cell Sci. 2012 Jan 15;125(2):507–15.

29. Zhou C min, Wang B, Wu Q, Lin P, Qin S gang, Pu Q qin, et al. Identification of cGAS as an innate immune sensor of extracellular bacterium Pseudomonas aeruginosa. iScience. 2020 Dec 10;24(1):101928.

30. Parker D, Cohen TS, Alhede M, Harfenist BS, Martin FJ, Prince A. Induction of Type I Interferon Signaling by Pseudomonas aeruginosa Is Diminished in Cystic Fibrosis Epithelial Cells. Am J Respir Cell Mol Biol. 2012 Jan;46(1):6–13.

31. Knowlton JJ, Dermody TS, Holm GH. Apoptosis Induced by Mammalian Reovirus Is Beta Interferon (IFN) Independent and Enhanced by IFN Regulatory Factor 3- and NF-κB-Dependent Expression of Noxa. J Virol. 2012 Feb;86(3):1650–60.

32. Kirschnek S, Vier J, Gautam S, Frankenberg T, Rangelova S, Eitz-Ferrer P, et al. Molecular analysis of neutrophil spontaneous apoptosis reveals a strong role for the pro-apoptotic BH3-only protein Noxa. Cell Death Differ. 2011 Nov;18(11):1805–14.

33. Amieva M, Peek RM. Pathobiology of Helicobacter pylori-Induced Gastric Cancer. Gastroenterology. 2016 Jan;150(1):64–78.

34. Takase H, Nitanai H, Hoshino K, Otani T. Impact of siderophore production on Pseudomonas aeruginosa infections in immunosuppressed mice. Infect Immun. 2000 Apr;68(4):1834–9.

35. Kulasakara H, Lee V, Brencic A, Liberati N, Urbach J, Miyata S, et al. Analysis of Pseudomonas aeruginosa diguanylate cyclases and phosphodiesterases reveals a role for bis-(3'-5')-cyclic-GMP in virulence. Proc Natl Acad Sci U S A. 2006 Feb 21;103(8):2839–44.

36. Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E, Wu G, et al. An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. Proc Natl Acad Sci U S A. 2006 Feb 21;103(8):2833–8.

37. Muzellec B, Teleńczuk M, Cabeli V, Andreux M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. Bioinformatics. 2023 Sep 1;39(9):btad547.

38. Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi A, et al. GOATOOLS: A Python library for Gene Ontology analyses. Sci Rep. 2018 Jul 18;8(1):10872.

39. Sievers F, Higgins DG. Clustal Omega. Curr Protoc Bioinforma. 2014;48(1):3.13.1-3.13.16.

40. Capra JA, Singh M. Predicting functionally important residues from sequence conservation. Bioinformatics. 2007 Aug 1;23(15):1875–82.

41. Gutierrez S, Tyczynski WG, Boomsma W, Teufel F, Winther O. MembraneFold: Visualising transmembrane protein structure and topology [Internet]. bioRxiv; 2022 [cited 2024 Apr 12]. p. 2022.12.06.518085. Available from: https://www.biorxiv.org/content/10.1101/2022.12.06.518085v1

42. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, et al. Fast and accurate protein structure search with Foldseek. Nat Biotechnol. 2024 Feb;42(2):243–6.

43. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010;31(2):455–61.

Figure legend

Figure 1. Non-biased whole mouse transcriptome analysis of *P. aeruginosa-infected* tissues reveal differential enrichment for host transcripts depending on the anatomical site.

a) A schematic diagram of the workflow. Ocular tissues were harvested at 24h post-challenge with $1x10^6$ CFU PAO1. Tissues were paraffin-embedded, sectioned, stained for morphological markers and hybridized with whole host transcriptome probe library. Regions of interest were selected to encompass epithelial layer, stromal layer, and anterior chamber of the eye. Probes were cleaved post-hybridization, collected with GeoMx, amplified, and generated libraries sequenced. Image generated with BioRender.

b) Representative immunohistochemical images of noninfected (contralateral, control eyes) and *P. aeruginosa*-infected ocular tissues stained for cytokeratin (green), Ly6G (red), DNA (blue), and c-KIT (yellow) were visualized with GeoMx (300 □m size bar, N=2 biological replicas). Robust neutrophil infiltration (red) in the corneal stroma and epithelium is readily observed in the infected site.

c) Volcano plots illustrating differentially expressed transcripts in epithelium and stroma tissue layers of the eye following infection, compared to tissues of the non-infected contralateral eye at 24 hours post-challenge. (N=2, biological replicas). The x-axis indicates fold change (log2 scale), and the y-axis shows the statistical significance (-log10 FDR-adjusted p-values). Sample metadata included.

d) Bar plot of GO (Gene Ontology) functional enrichment analysis, highlighting tissue site-specific gene transcript clustering at level 7. The x-axis displays the FDR-adjusted p-values (-log10 scale), indicating the statistical significance of the enrichment. The y-axis

lists the top level 7 GO categories with enriched transcripts. This analysis incorporates data from ROI profiles comprising 24 epithelial and 24 stromal samples from ocular tissues harvested at 24 hours post-challenge and 8 epithelial and 8 stromal samples from control samples. (N=2, biological replicas). GO enrichment analysis listed in Suppl. Table 2.

Cumulatively these data offer a comprehensive tissue map of infection indicating region-specific alterations.

Figure 2. Bacterial transcript profiling of infected with *P. aeruginosa* PAO1 corneas reveal differential tissue enrichment.

a) Heatmap showing the distribution of 100 bacterial transcripts in infected eyes. Each column represents data from individual ROIs. Surface epithelial ROIs (N=14) (green), stromal ROIs (16) (yellow), and Anterior Chamber (AC) ROIs (N=16) (purple) are grouped together. The curved lines on the top of the heatmap connect juxtaposing ROIs. Data are presented cumulatively from ROI selections derived from two biological replicas.

b) Immunohistochemistry analysis of infected corneal section stained for cytokeratin (green) and Ly6G^{+} neutrophils (red). Size bar: 0.□mm. Overlay plot of host and bacterial transcript levels in PAO1-infected tissues. Logarithmic transcript abundances (log2(count + 1)) of host TNFα, and bacterial PA2590, FpvB, RplL, and AlgE are represented as circles, whose diameter reflects transcript burden in the individual ROIs.

c) Bacterial enrichment was calculated as relative transcript abundance in the surface segments compared to the stromal segments. Bacterial transcript levels were normalized

to the housekeeping bacterial RplL transcripts. FDR values and fold change (FC) are annotated per each gene. Representative bacterial enrichment data depicting: 1) transcripts that were enriched in the stroma FpvB, PA2590, FemA; 2) transcripts showing no enrichment, and 3) transcripts showing decreased abundances such as AlgE and Piv. Complete enrichment analysis is provided in Suppl Table 4. Data are representative of two experiments (Suppl. Fig. 2)

Figure 3. Bacterial transcript profiling of PA14-infected tissues exhibits similar tendencies as the PAO1-induced infections.

a) Heat map showing the distribution of 100 bacterial transcripts in shams and PA14-infected eyes. The shams and the infected tissues were harvested and processed 48h post-challenge. Each column represents data from individual ROIs. Transcript abundances were compared between surface epithelial sham ROIs (N=7) (green), stromal sham ROIs (N=4) (yellow), and infected surface ROIs (N=11) (green), infected stromal ROIs (N=11) (yellow), infected AC ROIs (N=8) (purple) are grouped together. Data are presented cumulatively from ROI selections derived from three biological replicates.
b) Bacterial enrichment was calculated as relative transcript abundances in the surface segments compared to the stromal segments. Bacterial transcript levels were normalized to the housekeeping bacterial RplL transcripts. FDR values and fold change (FC) are annotated for FpvB, PA2590, and FemA. Complete enrichment analysis is provided in Suppl Table 5.
c) Immunohistochemistry analysis of shams and PA14-infected corneal section stained for OprI (yellow), cytokeratin (green), Ly6G^{+} neutrophils (red), and DNA (blue). Size bar:

0.1mm . Arrows indicate bacterial spread from the infected surface lesion into the stroma. Shams, N=3, infected tissues N=11 biological replicates.

d) Venn diagram depicting differentially present and shared bacterial transcripts that are enriched within the PAO1 and PA14-infected tissues.

Cumulatively, data point to commonalities in infection architecture that are not strain-specific.

Figure 4. Harnessing host responses for highly predictive analysis of infection levels in tissue regions.

a) Residual plot of a Ridge Regression model trained on host expression profiles to predict bacterial RplL gene expression levels. Data values are transformed using $\log_2$(values + 1). The model underwent five-fold cross-validation using ROIs from tissues derived from 24h post-challenge. Model performance was evaluated based on average MAE, EVS, and R^2 scores.

b) Evaluation of model performance when trained with varying numbers of mouse gene features. Top features were chosen for subset training based on the coefficient values from the model trained using the complete dataset. Average R^2 scores provided the metric of assessment. Models, based on different feature sets, underwent optimization using grid search for hyperparameter tuning.

c) Host-bacteria gene expression correlation network, including features with absolute rho values greater than 0.7 and an FDR less than 0.05, determined using Spearman's correlation.

d) Correlation plot showcasing the relationships between RplL and Pmaip1, TNF, and Trim13 in ROIs from infected mice.

Cumulatively, these data offer a path towards predicting pathogen burden through the surrogate measurements of RplL abundance.

Figure 5. Infection with the PA2590Tn deletion mutant strain is associated with reduced pathology and bacterial burden.

a) Representative images from infected with PA14 and PA14 PA2590Tn deletion mutant mice from two independent experiments. Images were collected at 48h post-challenge.
b) Corneal opacity scores reflective of disease severity were quantified by image analysis at 24h and 48h post-challenge. Data are presented cumulatively from two independent experiments. The individual symbols present a single infected cornea. The filled symbols represent infected C57Bl6/J mice with the PA14 strain (N=10) and the open circles represent C57Bl6/J mice infected with the transposon deletion mutant for PA2590 (N=10). Unpaired Student's t-test with Welch's correction (P=0.01).
c) C57Bl6/J mice were infected with $5x10^5$ CFU of PA14 (N=19) (filled circles) or the PA14 PA2590Tn deletion mutant (N=9) (open circles), and bacterial burden was quantified at 24h and 48h post-challenge. Data are presented cumulatively from two independent experiments. The individual symbols present individual infected corneas. Unpaired Student's t-test with Welch's correction (P=0.0007).

Cumulatively, data point to decreased disease severity in the absence of the PA2590 gene product.

Figure 6. Structural homology-analysis predicts PA2590 gene to function as a vitamin B12 and iron transporter.

a) Prevalence of the PA2590 gene in the genomes of *P. aeruginosa* clinical isolates sourced from various regions: Africa (N=216), Asia (N=2278), Europe (N=1680), North America (N=5763), Oceania (N=366), and South America (N=253). Here, 'N' denotes the total number of clinical isolates analyzed from each region.

b) Amino acid conservation analysis for the PA2590 gene, illustrating data from clinical isolates of *P. aeruginosa* obtained from different infection types: genitourinary (blue line, N=1991), respiratory (orange line, N=6132), blood (green line, N=1202), and wound/skin infections (red line, N=1302). The x-axis displays the residue position, while the y-axis shows the 1-Shannon entropy values. Regions shaded indicate the extracellular loops of the protein.

c) Structural homology modeling highlights parallels between the PA2590 structure predicted by AlphaFold2 and the experimentally determined structure of the *E. coli* BtuB protein. The figure contrasts the predicted model (light green) with the Btub structure (dark green), with VB12 depicted in red. Structural superimposition was conducted using the cmd.align function in PyMOL, involving 1000 alignment cycles with the transform parameter set to 1. This analysis resulted in a Root Mean Square Deviation (RMSD) of 1.774 Å for residues 126-884 of PA2590 and 11-595 of BtuB, indicating a significant level of structural similarity.

d) Structural homology modeling highlights parallels between the PA2590 structure predicted by AlphaFold2 and the experimentally determined structure of the *E. coli* FepA protein. The figure depicts the predicted model (light green) with the FepA structure

(dark blue). The RMSD is 1.426 Å over amino acid residues 126-884 of PA2590 and 11-725 of FepA, indicating a significant degree of structural similarity.

e) *In silico* docking experiments were conducted between PA2590 and VB12 using Autodock Vina. The pose with the lowest energy (kcal/mol) was chosen for representation. Pymol scripts were used to highlight residues potentially interacting with PA2590, focusing on polar contacts within a 3.5Å range.

f) *In silico* docking experiments were conducted between PA2590 and pyochelin using Autodock Vina. The pose with the lowest energy (kcal/mol) was chosen for representation, indicating a docking score of -8.2kcal/mol.

Cumulatively, data indicate that the PA2590 might function as either an iron imported utilizing a host siderophore or VB12 importer.

Supplementary Figures

Supplementary Figure 1. Temporal profiling of host response confirms sustained transcriptional remodeling and tissue-specific gene enrichment.

a) Representative images from infected with $5x10^5$ CFU *P. aeruginosa* PAO1 strain C57Bl6/J mice (N=5 biological replicas). Images are collected at 24h and 48h post-challenge.
b) Corneal opacity scores reflective of disease severity were estimated at 24h and 48h post-challenge. (N=5 biological replicas).
c) Volcano plots illustrating differentially expressed transcripts in epithelium and stroma tissue layers of the eye following infection, compared to tissues of the non-infected control eyes at 48h post-challenge. Mice were challenged with $5x10^5$ CFU *P. aeruginosa* PAO1 and tissues were harvested at 24h (N=1) and 48h post-infection (N=2). Data are presented cumulatively. The x-axis indicates fold change (log^2 scale), and the y-axis shows the statistical significance ($-log^{10}$ FDR-adjusted p-values). Detailed sample metadata is available in Suppl. Table 2.
d) Bar plot of GO (Gene Ontology) functional enrichment analysis, highlighting tissue site-specific gene transcript clustering at level 7. The x-axis displays the FDR-adjusted p-values (-log10 scale), indicating the statistical significance of the enrichment. The y-axis lists the top level 7 GO categories with enriched transcripts. This analysis incorporates data from ROI profiles comprising 24 epithelial and 24 stromal samples from ocular tissues harvested at 24 hours post-challenge and eight epithelial and eight stromal samples from non-infected eyes, control samples.

Supplementary Figure 2. Using host responses to cross validate predictions of infection burden in tissue regions.

a) Ridge Regression model residual plot derived from host response profiles to predict pathogen-specific RplL gene expression levels. Data values are transformed using $\log_2$(values + 1). The model was trained on ROIs selections from tissues derived from 24h post-challenge and cross validated against ROI data set derived from a different biological replica. Model performance was evaluated based on average MAE, EVS, and R^2 scores.
b) Evaluation of model performance when trained on a single biological data set and validated against a second data set derived from a biological replica. Average R^2 scores provided the metric of assessment. Models, based on different feature sets, underwent optimization using grid search for hyperparameter tuning.
c) The assessment of model performance involved training on multiple biological replicas and validating on a single excluded replica, which included samples infected with PA14 and PAO1. Average R2 scores served as the evaluation metric. The models, based on various feature sets, underwent optimization through grid search for hyperparameter tuning.

Supplementary Figure 3. Structural homology-analysis predicts PA2590 gene to function as an iron transporter.

a) Genomic data from 11,447 *P. aeruginosa* clinical isolates was sourced from NCBI reflecting strain diversity for the past 23 years. The bar graph depicts the global

distribution of the number of isolates plotted on the y-axis, and the types of associated infections plotted on the x-axis.

b) Membrane topology of PA2590 structure predicted by MembraneFold.

c) *In silico* docking experiments were conducted between PA2590 and 2,5-DHBA. The pose with the lowest energy (kcal/mol) was chosen for representation.

Supplementary Tables

Supplementary Table 1. Host-specific transcriptional changes induced by infection at 24h post-challenge (Supplementary data to Figure 1.). List of differentially expressed genes in infected epithelium ROIs when compared to contralateral, control ROIs. List of differentially expressed genes in the infected stroma ROIs when compared to contralateral, control stroma ROIs. GO enrichment analysis. Metadata to Figure 1.

Supplementary Table 2. List of genes differentially expressed in infected epithelial and stromal ROIs when comparing 48-hour infection to 24-hour infection.

Supplementary Table 3. List of 100 *P. aeruginosa* transcripts profiled in the GeoMx experiments and corresponding Nanostring detection probes.

Supplementary Table 4. Enrichment analysis reflecting *P. aeruginosa* PAO1 bacterial transcript fold changes at 24h post-challenge.

Supplementary Table 5. Enrichment analysis reflecting *P. aeruginosa* PA14 bacterial transcript fold changes at 48h post-challenge.

Data and Code Availability

Upon acceptance of the manuscript, the raw sequencing data from our genomics and spatial transcriptomics studies will be made available in the NCBI database. Additionally, the code employed for training predictive models and performing correlative analysis will be shared openly on our Github page.

Acknowledgments

We would like to acknowledge the outstanding contributions of the NanoString team: Maxine Macclain, Espy Anguiano, Tim Riordan, Ozge Cagsal-Getkin, Mary Hentschel, and Ankush Tyagi who worked tirelessly to help us conduct the feasibility studies and establish the technology. We would like to thank University of Massachusetts Histopathology core, Yu Liu and Jayme Haywosz, for embedding and processing services. We would like to thank Prof. G. Pier (MGB, HMS) for providing the PA14 and PA14 PA2590 strains.

Author contributions

HZ conducted all computational analyses. SB and MM carried out all in vivo experiments and analyzed their results. ON was responsible for data acquisition within the GeoMx workflow. BJ primarily handled the sequencing analysis portion of the GeoMx workflow. EO and HZ selected the PA genes. CT advised on the sequencing aspects of the GeoMx workflow. GD and HL

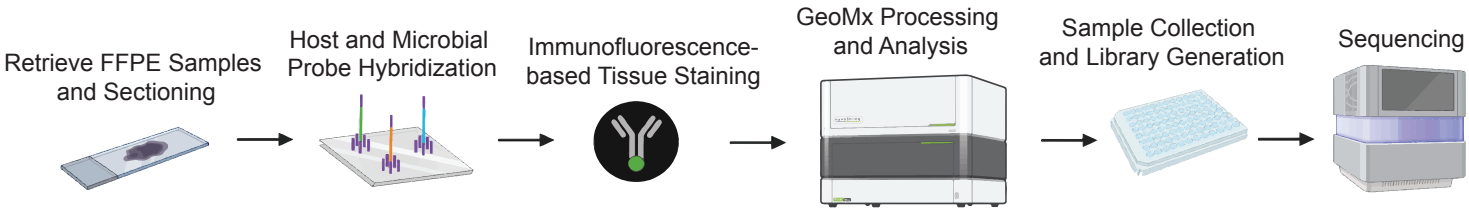
validated the PA2590 Tn mutant. HZ, NC, EP, and SB played key roles in study design and data interpretation. MG devised the study concept, and contributed to data collection, analysis, and interpretation. HZ, ON, NC, EP, and MG drafted the initial manuscript. All authors discussed the data, reviewed, and revised the manuscript.

Ethics declarations

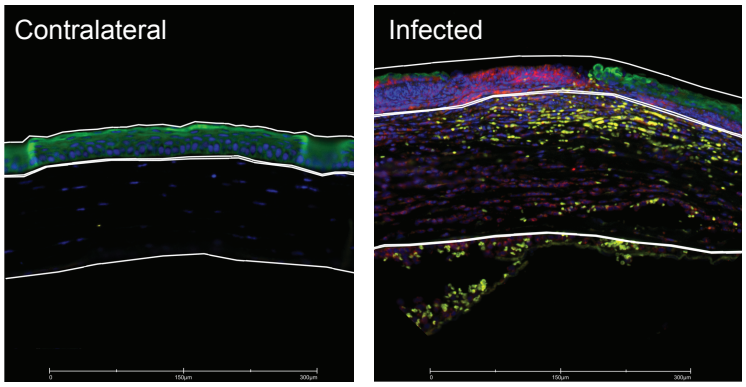
HZ, ON, NC, BJ, EO, CT, GD, HL, OP, and MG are employee of Moderna Inc. and may hold stock/stock options in Moderna, Inc. SA, MM, and EP have no conflicts to declare.

Figure 1

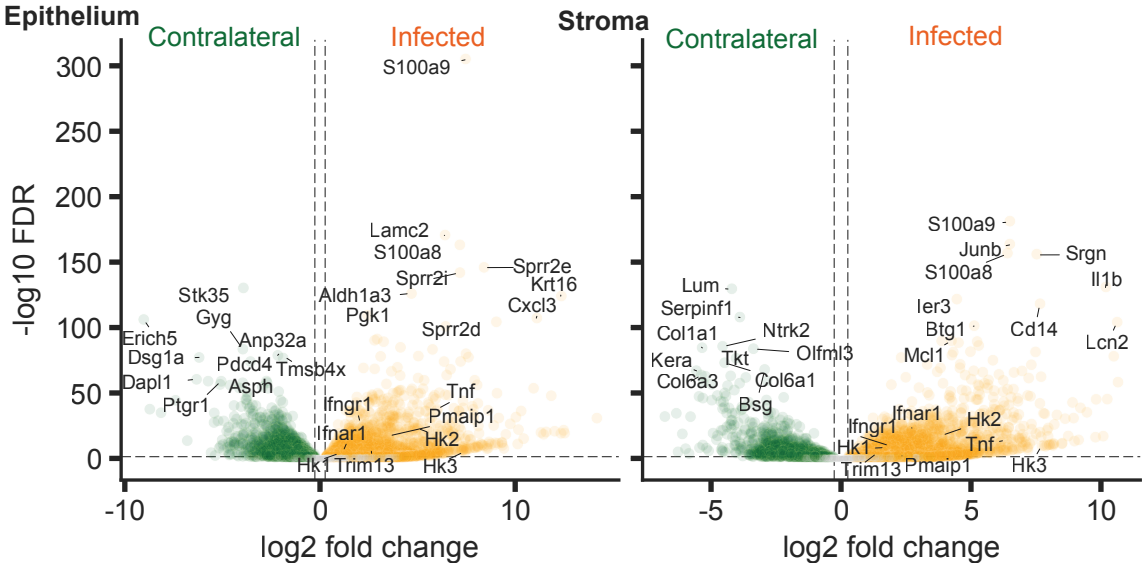
A



B



C



D

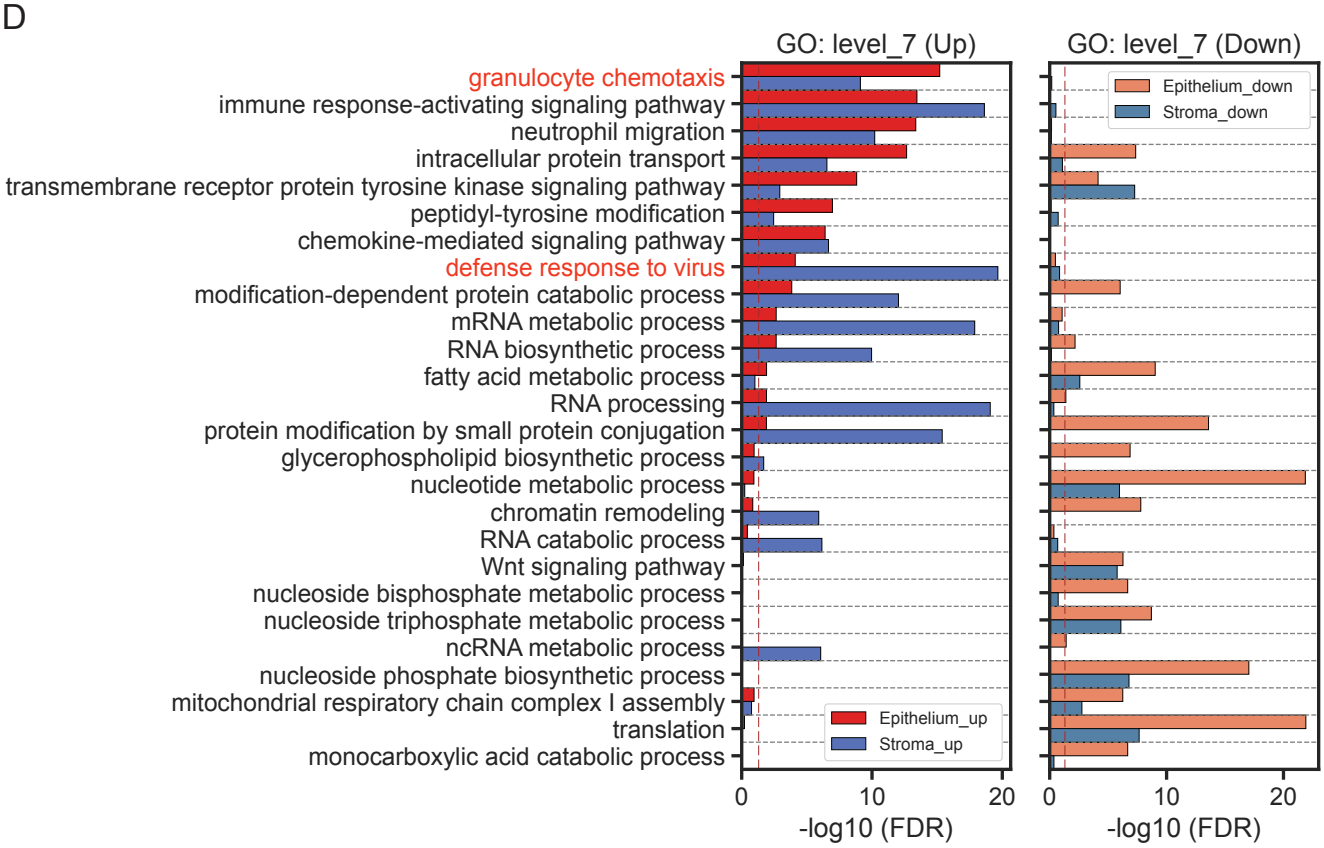


Figure 2

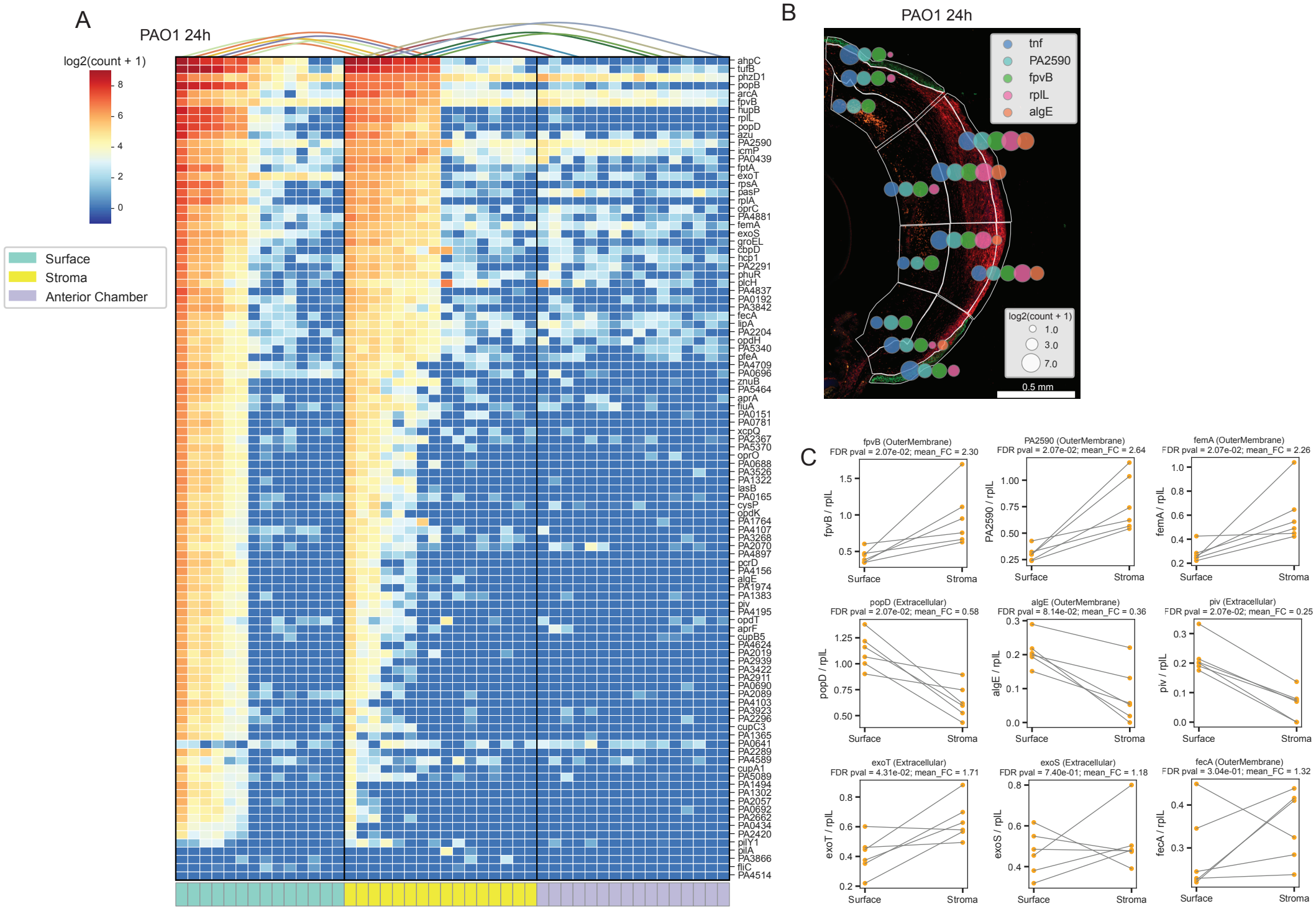

A
PAO1 24h
log2(count + 1)
8
6
4
2
0
Surface
Stroma
Anterior Chamber
ahpC
tufB
phzD1
popB
arcA
fpvB
hupB
rplL
popD
azu
PA2590
icmP
PA0439
fptA
exoT
rpsA
pasP
rplA
oprC
PA4881
femA
exoS
groEL
cbpD
hcp1
PA2291
phuR
plcH
PA4837
PA0192
PA3842
fecA
lipA
PA2204
opdH
PA5340
pfeA
PA4709
PA0696
znuB
PA5464
aprA
fiuA
PA0151
PA0781
xcpQ
PA2367
PA5370
oprO
PA0688
PA3526
PA1322
lasB
PA0165
cysP
opdK
PA1764
PA4107
PA3268
PA2070
PA4897
pcrD
PA4156
algE
PA1974
PA1383
piv
PA4195
opdT
aprF
cupB5
PA4624
PA2019
PA2939
PA3422
PA2911
PA0690
PA2089
PA4103
PA3923
PA2296
cupC3
PA1365
PA0641
PA2289
PA4589
cupA1
PA5089
PA1494
PA1302
PA2057
PA0692
PA2662
PA0434
PA2420
pilY1
pilA
PA3866
fliC
PA4514
B
PAO1 24h
tnf
PA2590
fpvB
rplL
algE
log2(count + 1)
1.0
3.0
7.0
0.5 mm
C
fpvB (OuterMembrane)
FDR pval = 2.07e-02; mean_FC = 2.30
fpvB / rplL
PA2590 (OuterMembrane)
FDR pval = 2.07e-02; mean_FC = 2.64
PA2590 / rplL
femA (OuterMembrane)
FDR pval = 2.07e-02; mean_FC = 2.26
femA / rplL
popD (Extracellular)
FDR pval = 2.07e-02; mean_FC = 0.58
popD / rplL
algE (OuterMembrane)
FDR pval = 8.14e-02; mean_FC = 0.36
algE / rplL
piv (Extracellular)
FDR pval = 2.07e-02; mean_FC = 0.25
piv / rplL
exoT (Extracellular)
FDR pval = 4.31e-02; mean_FC = 1.71
exoT / rplL
exoS (Extracellular)
FDR pval = 7.40e-01; mean_FC = 1.18
exoS / rplL
fecA (OuterMembrane)
FDR pval = 3.04e-01; mean_FC = 1.32
fecA / rplL
Surface
Stroma

Figure 3

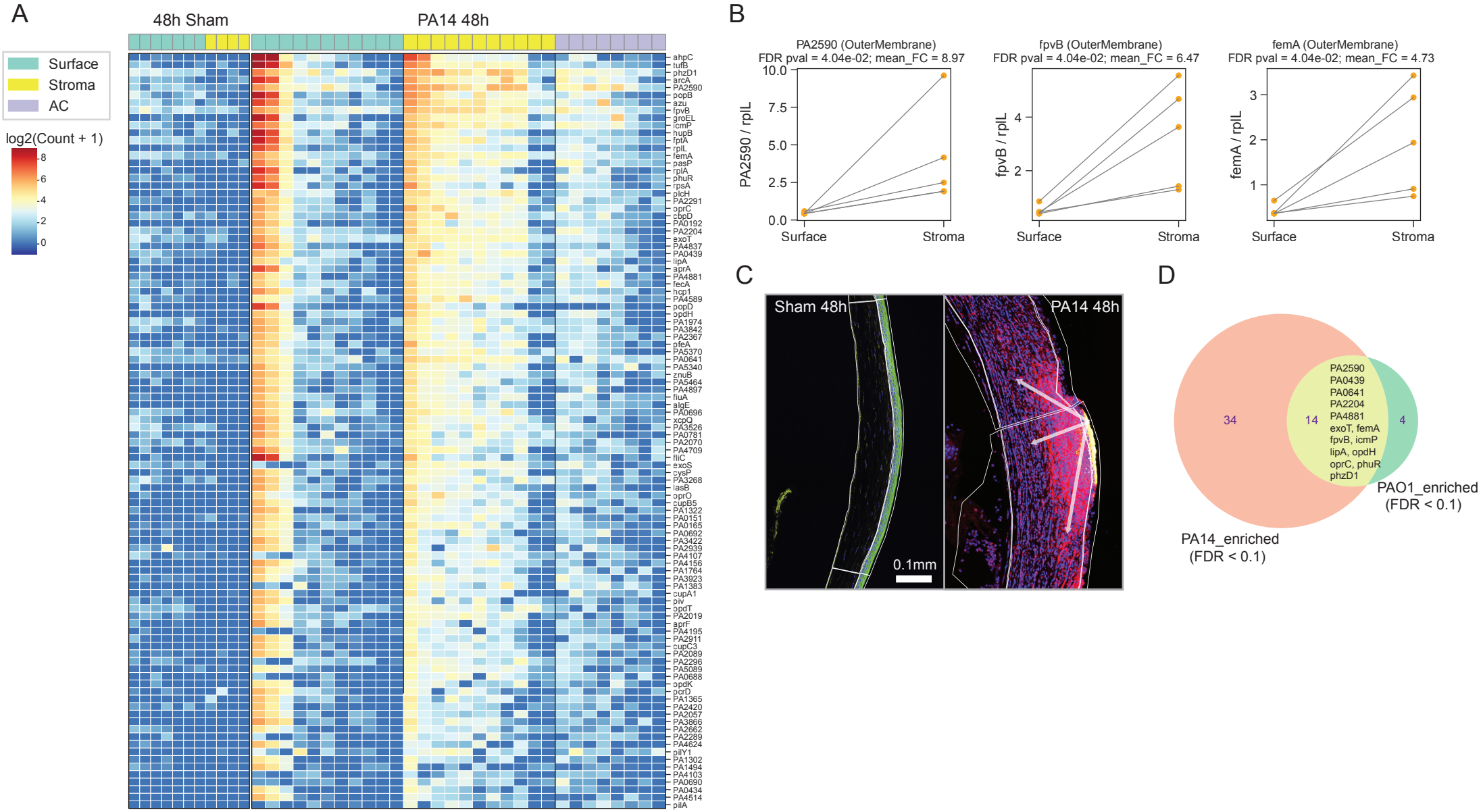

A
48h Sham
PA14 48h
Surface
Stroma
AC
log2(Count + 1)
8
6
4
2
0
ahpC
tufB
phzD1
arcA
PA2590
popB
azu
fpvB
groEL
icmP
hupB
fptA
rplL
femA
pasP
rplA
phuR
rpsA
plcH
PA2291
oprC
cbpD
PA0192
PA2204
exoT
PA4837
PA0439
lipA
aprA
PA4881
fecA
hcp1
PA4589
popD
opdH
PA1974
PA3842
PA2367
pfeA
PA5370
PA0641
PA5340
znuB
PA5464
PA4897
fiuA
algE
PA0696
xcpQ
PA3526
PA0781
PA2070
PA4709
fliC
exoS
cysP
PA3268
lasB
oprO
cupB5
PA1322
PA0151
PA0165
PA0692
PA3422
PA2939
PA4107
PA4156
PA1764
PA3923
PA1383
cupA1
piv
opdT
PA2019
aprF
PA4195
PA2911
cupC3
PA2089
PA2296
PA5089
PA0688
opdK
pcrD
PA1365
PA2420
PA2057
PA3866
PA2662
PA2289
PA4624
pilY1
PA1302
PA1494
PA4103
PA0690
PA0434
PA4514
pilA
B
PA2590 (OuterMembrane)
FDR pval = 4.04e-02; mean_FC = 8.97
PA2590 / rplL
10.0
7.5
5.0
2.5
0.0
Surface
Stroma
fpvB (OuterMembrane)
FDR pval = 4.04e-02; mean_FC = 6.47
fpvB / rplL
4
2
Surface
Stroma
femA (OuterMembrane)
FDR pval = 4.04e-02; mean_FC = 4.73
femA / rplL
3
2
1
Surface
Stroma
C
Sham 48h
PA14 48h
0.1mm
D
34
14
4
PA2590
PA0439
PA0641
PA2204
PA4881
exoT, femA
fpvB, icmP
lipA, opdH
oprC, phuR
phzD1
PAO1_enriched
(FDR < 0.1)
PA14_enriched
(FDR < 0.1)

Figure 4

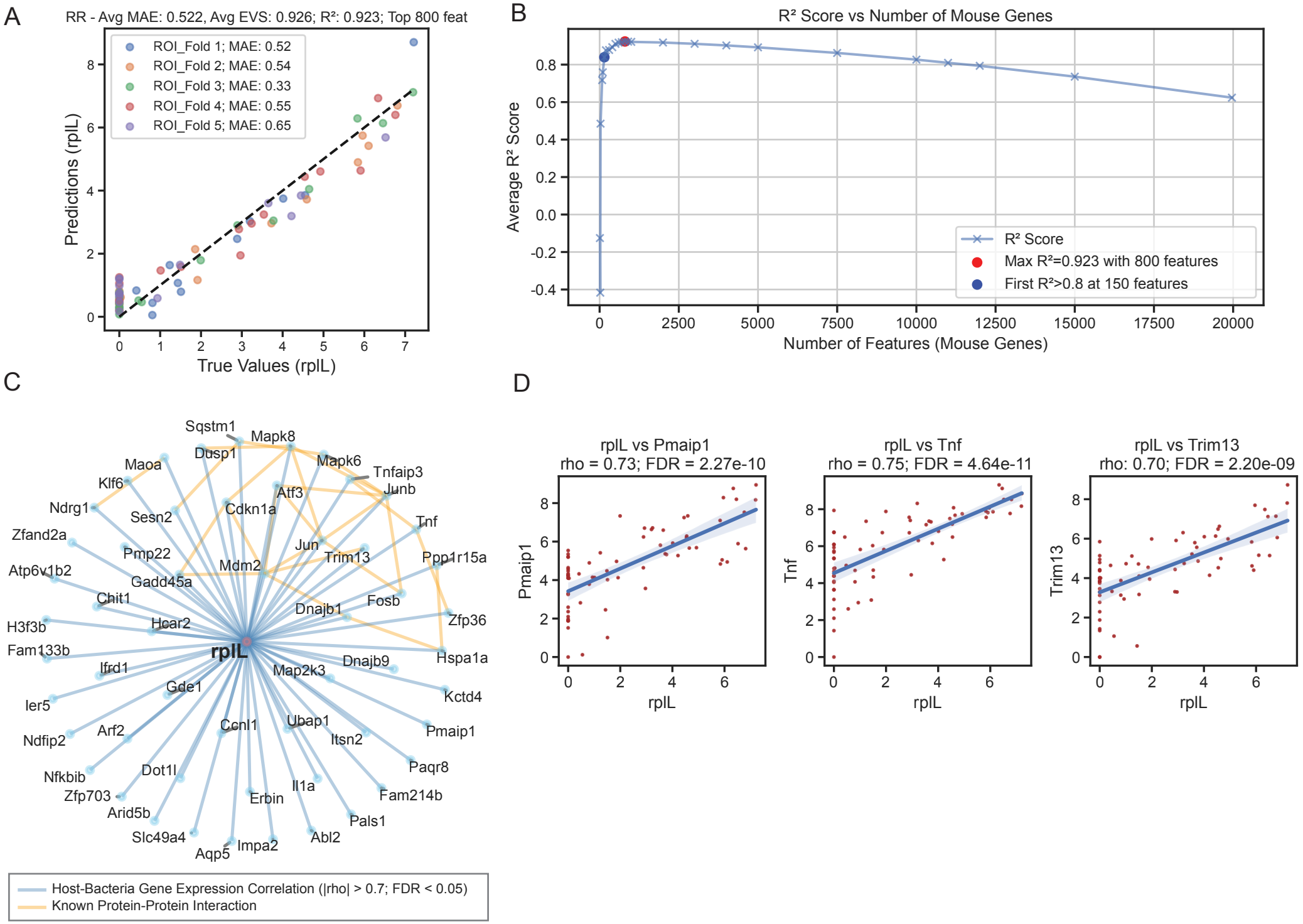
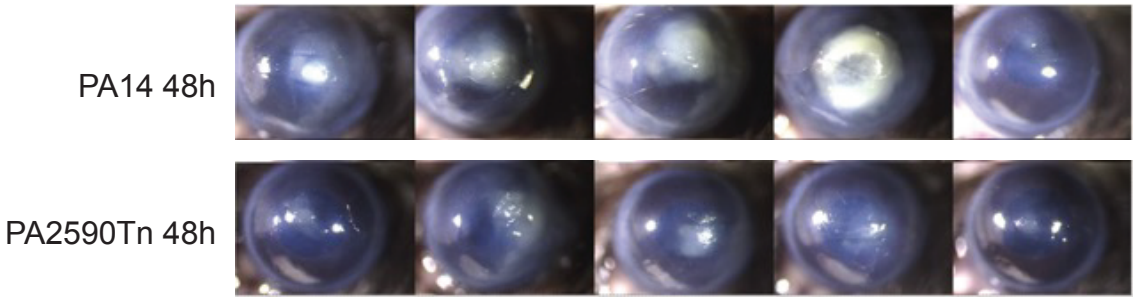
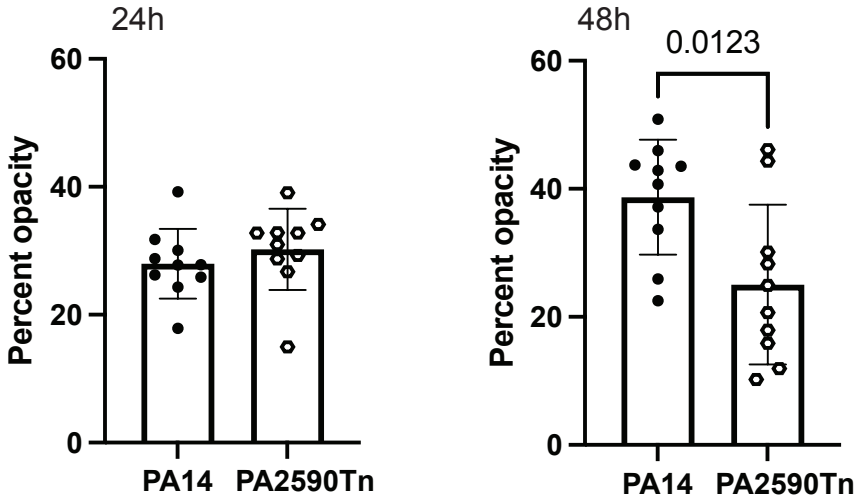

A
RR - Avg MAE: 0.522, Avg EVS: 0.926; R²: 0.923; Top 800 feat
ROI_Fold 1; MAE: 0.52
ROI_Fold 2; MAE: 0.54
ROI_Fold 3; MAE: 0.33
ROI_Fold 4; MAE: 0.55
ROI_Fold 5; MAE: 0.65
Predictions (rplL)
True Values (rplL)
B
R² Score vs Number of Mouse Genes
Average R² Score
Number of Features (Mouse Genes)
R² Score
Max R²=0.923 with 800 features
First R²>0.8 at 150 features
C
rplL
Host-Bacteria Gene Expression Correlation (|rho| > 0.7; FDR < 0.05)
Known Protein-Protein Interaction
D
rplL vs Pmaip1
rho = 0.73; FDR = 2.27e-10
rplL vs Tnf
rho = 0.75; FDR = 4.64e-11
rplL vs Trim13
rho: 0.70; FDR = 2.20e-09
Pmaip1
Tnf
Trim13
rplL

Figure 5

A


PA14 48h
PA2590Tn 48h

B


24h
48h
0.0123
Percent opacity
60
40
20
0
PA14
PA2590Tn

C

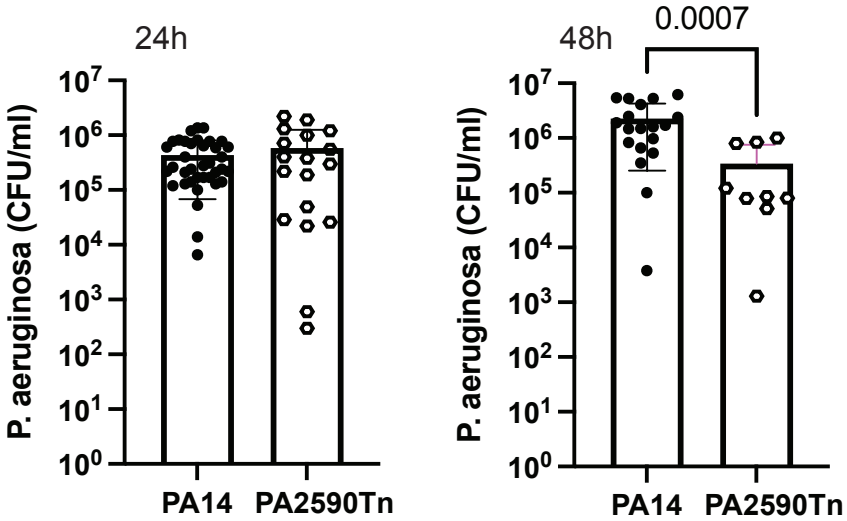
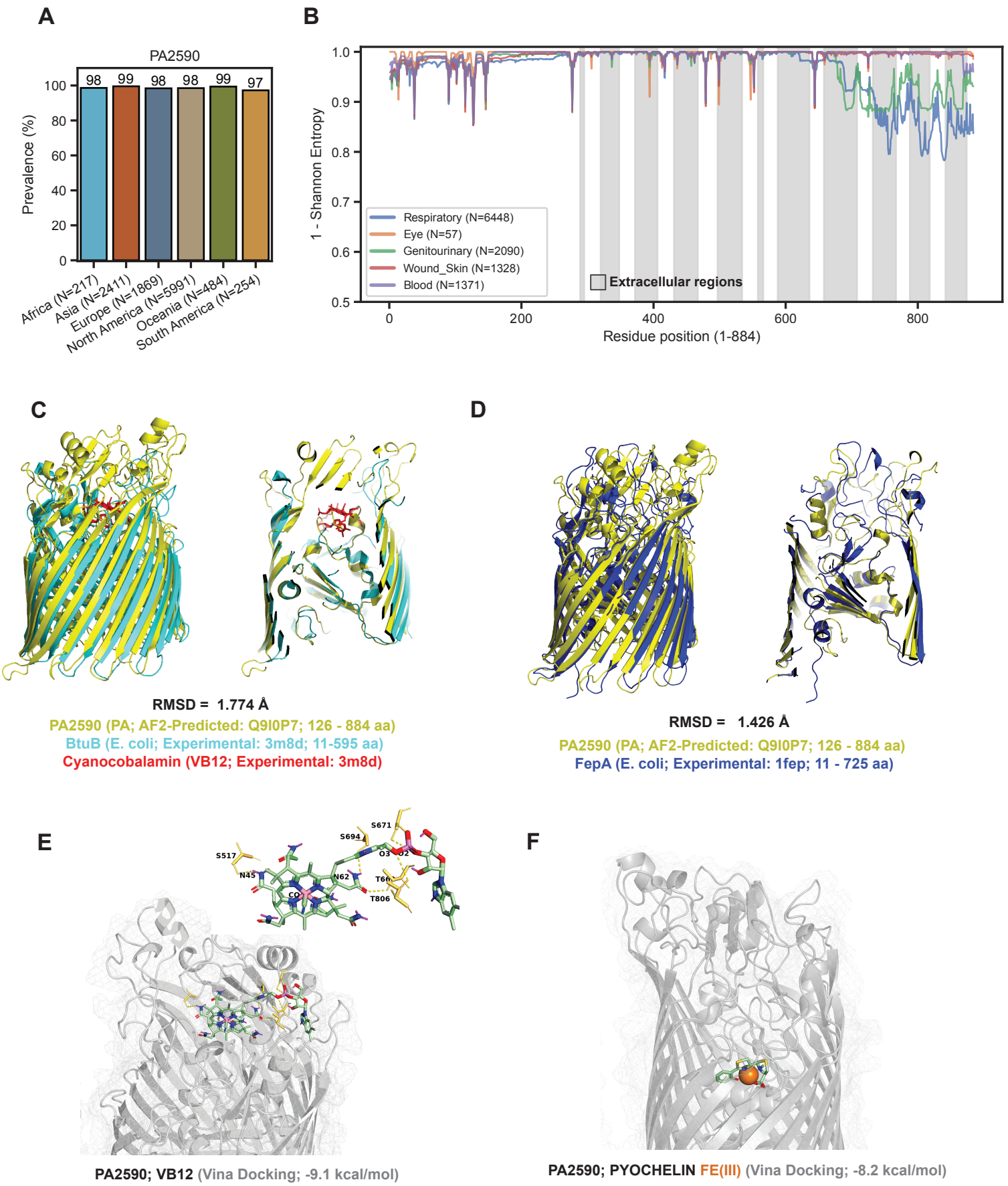
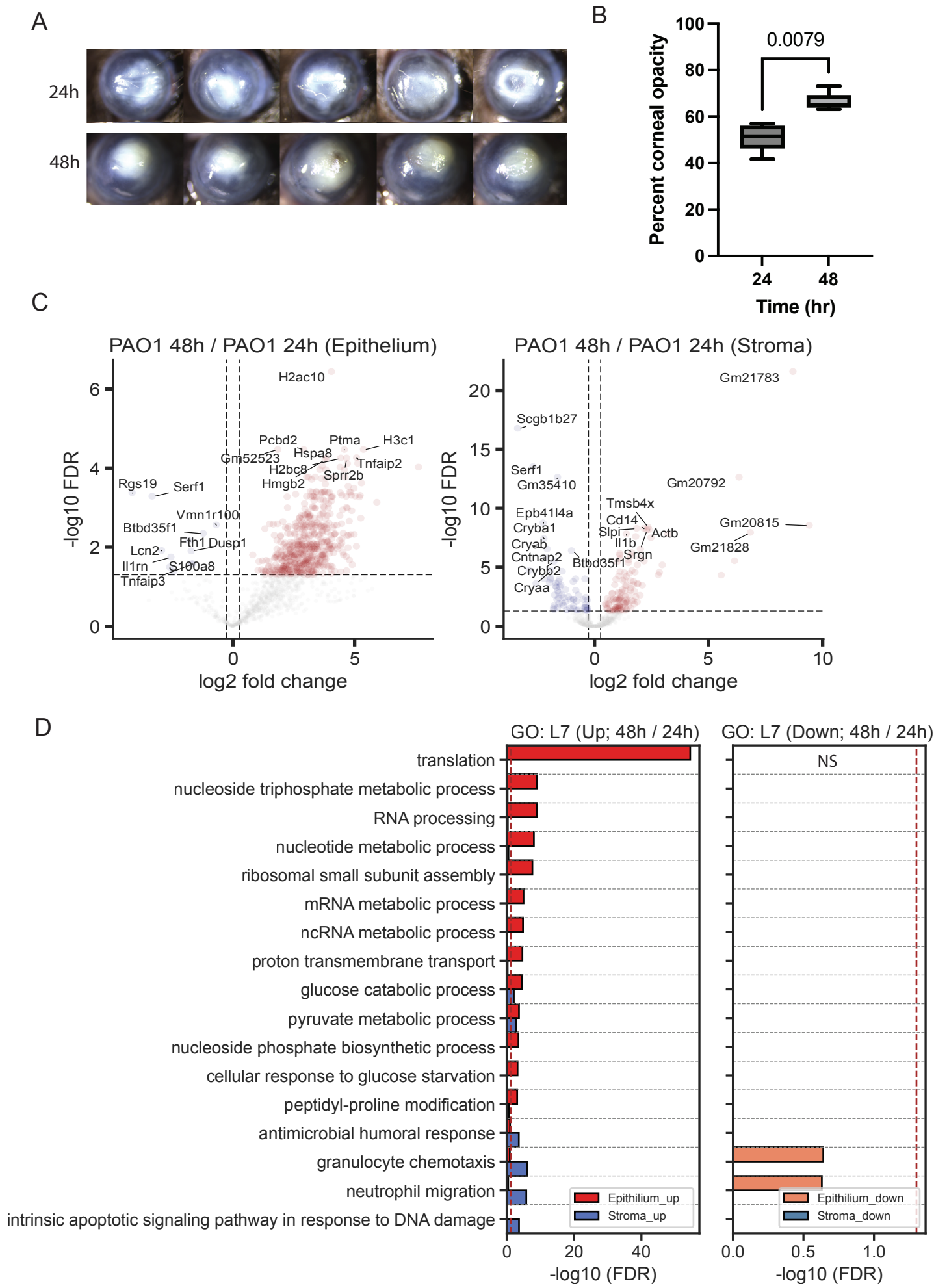

24h
48h
0.0007
P. aeruginosa (CFU/ml)
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
PA14
PA2590Tn

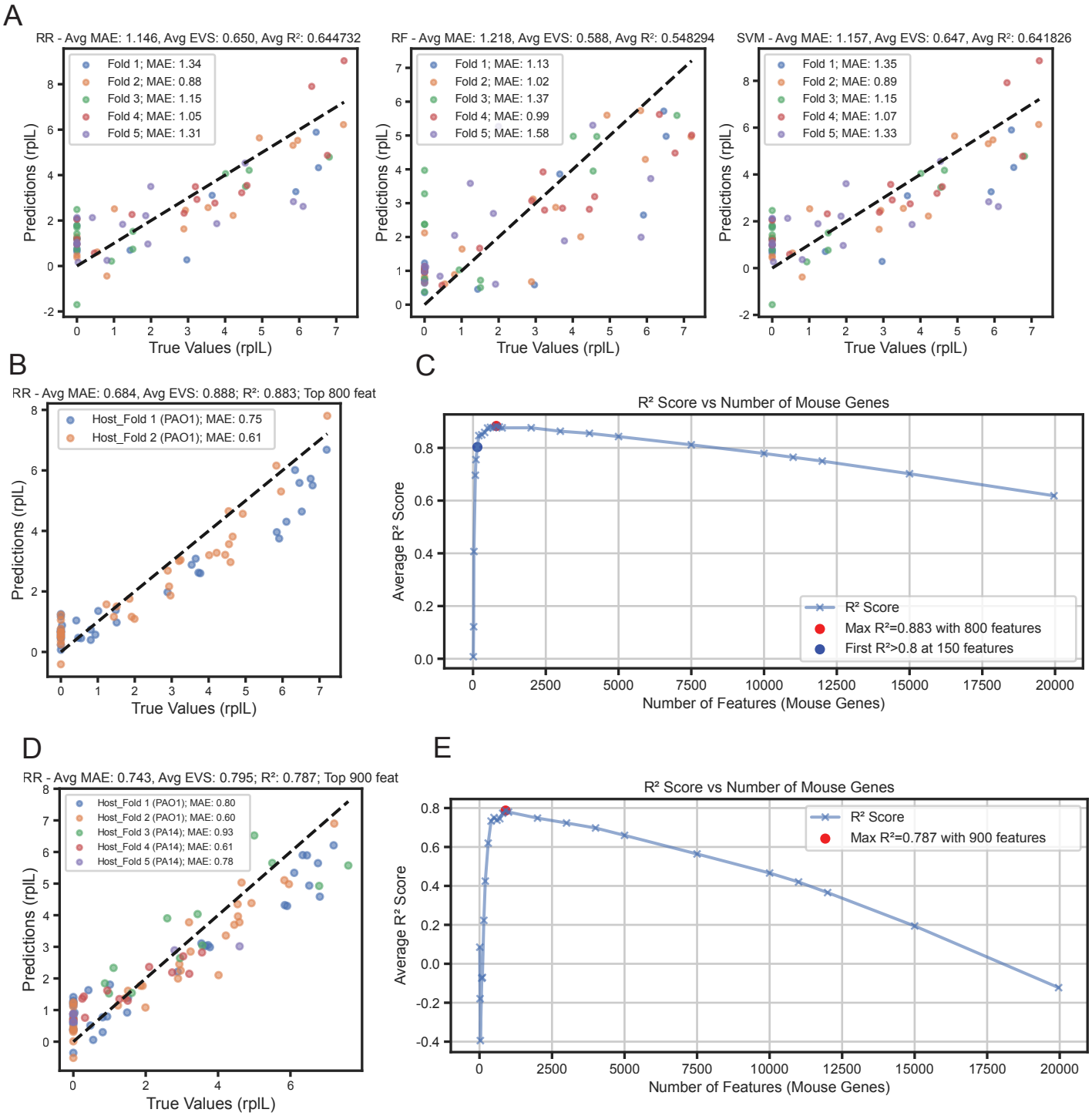
Figure 6


A
PA2590
Prevalence (%)
98
99
98
98
99
97
Africa (N=217)
Asia (N=2411)
Europe (N=1869)
North America (N=5991)
Oceania (N=484)
South America (N=254)
B
1 - Shannon Entropy
Respiratory (N=6448)
Eye (N=57)
Genitourinary (N=2090)
Wound_Skin (N=1328)
Blood (N=1371)
Extracellular regions
Residue position (1-884)
C
RMSD = 1.774 Å
PA2590 (PA; AF2-Predicted: Q9I0P7; 126 - 884 aa)
BtuB (E. coli; Experimental: 3m8d; 11-595 aa)
Cyanocobalamin (VB12; Experimental: 3m8d)
D
RMSD = 1.426 Å
PA2590 (PA; AF2-Predicted: Q9I0P7; 126 - 884 aa)
FepA (E. coli; Experimental: 1fep; 11 - 725 aa)
E
S671
S694
S517
N45
N62
T66
T806
PA2590; VB12 (Vina Docking; -9.1 kcal/mol)
F
PA2590; PYOCHELIN FE(III) (Vina Docking; -8.2 kcal/mol)

Suppl. Fig. 1

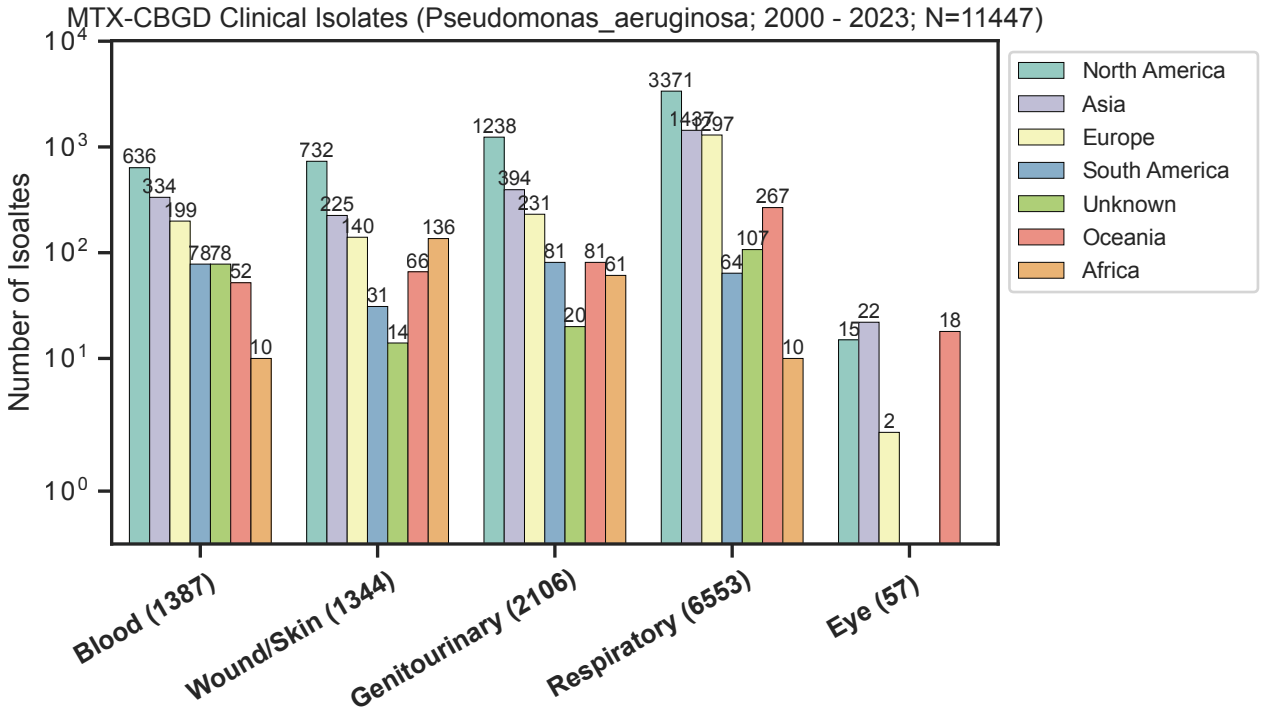

A
24h
48h
B
Percent corneal opacity
0.0079
100
80
60
40
20
0
24
48
Time (hr)
C
PAO1 48h / PAO1 24h (Epithelium)
-log10 FDR
log2 fold change
H2ac10
Pcbd2
Ptma
H3c1
Gm52523
Hspa8
H2bc8
Tnfaip2
Sprr2b
Hmgb2
Rgs19
Serf1
Vmn1r100
Btbd35f1
Fth1
Dusp1
Lcn2
Il1rn
S100a8
Tnfaip3
PAO1 48h / PAO1 24h (Stroma)
Gm21783
Scgb1b27
Serf1
Gm35410
Gm20792
Tmsb4x
Epb41l4a
Cryba1
Cd14
Slpi
Actb
Gm20815
Cryab
Il1b
Gm21828
Cntnap2
Srgn
Btbd35f1
Crybb2
Cryaa
D
GO: L7 (Up; 48h / 24h)
GO: L7 (Down; 48h / 24h)
translation
nucleoside triphosphate metabolic process
RNA processing
nucleotide metabolic process
ribosomal small subunit assembly
mRNA metabolic process
ncRNA metabolic process
proton transmembrane transport
glucose catabolic process
pyruvate metabolic process
nucleoside phosphate biosynthetic process
cellular response to glucose starvation
peptidyl-proline modification
antimicrobial humoral response
granulocyte chemotaxis
neutrophil migration
intrinsic apoptotic signaling pathway in response to DNA damage
NS
Epithilium_up
Stroma_up
Epithilium_down
Stroma_down
-log10 (FDR)
-log10 (FDR)

Suppl. Fig. 2

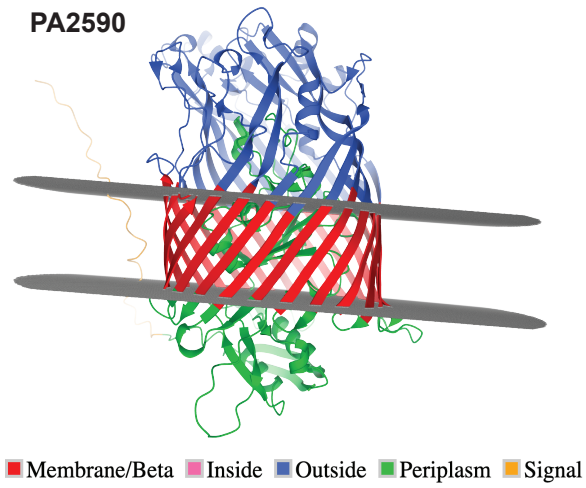

A
RR - Avg MAE: 1.146, Avg EVS: 0.650, Avg R²: 0.644732
Fold 1; MAE: 1.34
Fold 2; MAE: 0.88
Fold 3; MAE: 1.15
Fold 4; MAE: 1.05
Fold 5; MAE: 1.31
RF - Avg MAE: 1.218, Avg EVS: 0.588, Avg R²: 0.548294
Fold 1; MAE: 1.13
Fold 2; MAE: 1.02
Fold 3; MAE: 1.37
Fold 4; MAE: 0.99
Fold 5; MAE: 1.58
SVM - Avg MAE: 1.157, Avg EVS: 0.647, Avg R²: 0.641826
Fold 1; MAE: 1.35
Fold 2; MAE: 0.89
Fold 3; MAE: 1.15
Fold 4; MAE: 1.07
Fold 5; MAE: 1.33
Predictions (rplL)
True Values (rplL)
B
RR - Avg MAE: 0.684, Avg EVS: 0.888; R²: 0.883; Top 800 feat
Host_Fold 1 (PAO1); MAE: 0.75
Host_Fold 2 (PAO1); MAE: 0.61
Predictions (rplL)
True Values (rplL)
C
R² Score vs Number of Mouse Genes
Average R² Score
Number of Features (Mouse Genes)
R² Score
Max R²=0.883 with 800 features
First R²>0.8 at 150 features
D
RR - Avg MAE: 0.743, Avg EVS: 0.795; R²: 0.787; Top 900 feat
Host_Fold 1 (PAO1); MAE: 0.80
Host_Fold 2 (PAO1); MAE: 0.60
Host_Fold 3 (PA14); MAE: 0.93
Host_Fold 4 (PA14); MAE: 0.61
Host_Fold 5 (PA14); MAE: 0.78
Predictions (rplL)
True Values (rplL)
E
R² Score vs Number of Mouse Genes
Average R² Score
Number of Features (Mouse Genes)
R² Score
Max R²=0.787 with 900 features

Suppl. Fig. 3

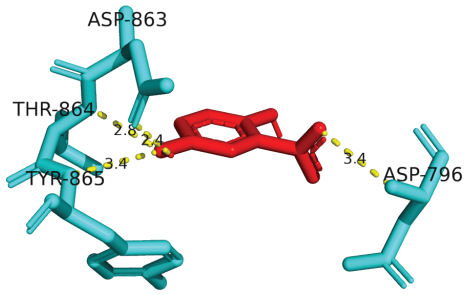
A



B



C



2,5-DHBA (Vina Docking; -6.0 kcal/mol)