



Application of a quantitative framework to improve the accuracy of a bacterial infection model

Gina R. Lewin^{a,b,1} , Ananya Kapur^{c,1}, Daniel M. Cornforth^{a,b,1}, Rebecca P. Duncan^{b,d} , Frances L. Diggle^{a,b}, Dina A. Moustafa^{b,d} , Simone A. Harrison^{e,f,g}, Eric P. Skaar^h, Walter J. Chazin^{e,f,g} , Joanna B. Goldberg^{b,d} , Jennifer M. Bomberger^{c,2,3}, and Marvin Whiteley^{a,b,3}

Edited by Caroline Harwood, University of Washington, Seattle, WA; received December 20, 2022; accepted April 7, 2023

Laboratory models are critical to basic and translational microbiology research. Models serve multiple purposes, from providing tractable systems to study cell biology to allowing the investigation of inaccessible clinical and environmental ecosystems. Although there is a recognized need for improved model systems, there is a gap in rational approaches to accomplish this goal. We recently developed a framework for assessing the accuracy of microbial models by quantifying how closely each gene is expressed in the natural environment and in various models. The accuracy of the model is defined as the percentage of genes that are similarly expressed in the natural environment and the model. Here, we leverage this framework to develop and validate two generalizable approaches for improving model accuracy, and as proof of concept, we apply these approaches to improve models of *Pseudomonas aeruginosa* infecting the cystic fibrosis (CF) lung. First, we identify two models, an in vitro synthetic CF sputum medium model (SCFM2) and an epithelial cell model, that accurately recapitulate different gene sets. By combining these models, we developed the epithelial cell-SCFM2 model which improves the accuracy of over 500 genes. Second, to improve the accuracy of specific genes, we mined publicly available transcriptome data, which identified zinc limitation as a cue present in the CF lung and absent in SCFM2. Induction of zinc limitation in SCFM2 resulted in accurate expression of 90% of *P. aeruginosa* genes. These approaches provide generalizable, quantitative frameworks for microbiological model improvement that can be applied to any system of interest.

Pseudomonas aeruginosa | calprotectin | preclinical model | epithelial cell model | cystic fibrosis

Laboratory models are a bedrock of microbiology research, and their importance has been noted since the time of Pasteur (1). The goal of laboratory models is to provide versatile, highly reproducible systems that allow detailed studies of individual microbes and microbial communities, with the assumption that the experimental results provide insight into microbial behavior in the natural environment. Thus, significant resources, time, and effort have been devoted to designing models to enhance their applicability to natural systems. For example, to improve model relevance, researchers developed humanized mice to allow infection by certain microbes (2, 3) and designed standardized environments, such as fabricated microbial ecosystems or “EcoFABs”, for understanding natural microbial communities (4, 5). However, model development has primarily relied on the intuition of scientists, as a systematic, quantitative framework for assessing the accuracy of laboratory models did not exist. As a result, a comprehensive understanding of the relevance of models to the natural environment is generally unknown.

In previous work, we established a quantitative framework to assess the accuracy of model systems using bacterial gene expression as a proxy for microbial function. This approach directly compares gene expression of microbe(s) in a model of interest relative to gene expression in the native environment (6, 7). Genes are defined as “accurate” based on similarity in expression in the model system and the natural environment, providing a genome-wide assessment of the similarities in gene expression in the natural environment and laboratory models. These studies revealed that model systems vary widely in their accurate recapitulation of microbial functions, and many intuitively designed models were less accurate overall compared with traditional microbiology culture conditions (8). This global comparison of microbial functions in models and the natural environment provided a straightforward method for model choice, allowing scientists to choose models that accurately recapitulate their function(s) of interest.

In this study, we leveraged our ability to quantify the accuracy of model systems to develop a framework to improve models. As proof of concept, we focused on the improvement of model systems to study *Pseudomonas aeruginosa* behavior during stable, chronic infection of the lungs of people with cystic fibrosis (CF) (7). CF is a genetic disease that results in inflammation and accumulation of mucus (sputum) in the lungs (9–11). People

Significance

Model systems are a cornerstone of microbiology. However, despite microbiology's heavy reliance on laboratory models, these systems are typically not analyzed systematically to improve their relevance. This limitation is a primary challenge to understand microbes' physiology in natural environments. We provide a proof of concept for generalizable approaches for model improvement using transcriptomic data of the pathogen *Pseudomonas aeruginosa* from sputum of patients with cystic fibrosis. We quantitatively improve experimental model systems by 1) combining two models with different accuracies and 2) leveraging publicly available data to identify a condition (low zinc) that corrects the accuracy of target genes. These rationalized frameworks are broadly applicable and have the potential to reshape how we understand the role of microbes across ecosystems.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

¹G.R.L., A.K., and D.M.C. contributed equally to this work.

²Present address: Department of Microbiology and Immunology, Geisel School of Medicine at Dartmouth, Hanover, NH 03756.

³To whom correspondence may be addressed. Email: Jennifer.M.Bomberger@dartmouth.edu or mwhiteley3@gatech.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2221542120/-DCSupplemental>.

Published May 1, 2023.

with CF are chronically colonized with several bacteria including *P. aeruginosa*, and these infections are a primary cause of morbidity and mortality (12). We previously identified models that accurately capture expression of approximately 85% of the *P. aeruginosa* genes, indicating the need for improved models that capture the hundreds of inaccurate genes (7). Here, we first show that through the combination of two differentially accurate model systems, we could create a new model with improved accuracy. Second, we mined publicly available transcriptome data to identify a condition to improve the accuracy of ten genes of interest, which resulted in a model with 90% accuracy. Finally, we emphasize that this process is iterative, as we show with additional CF models and other growth conditions. In sum, this work presents a new, generalizable conceptual framework for overcoming the important challenge of model system design, which will lead to a better understanding of bacterial physiology across ecosystems.

Results

Human Metatranscriptomes and *P. aeruginosa* Core Genome. The key to our approach is the acquisition of high-quality transcriptomes of the bacterium of interest in the native environment. Here, we used 24 bacterial metatranscriptomes obtained from human expectorated CF sputum that was immediately preserved after collection (Dataset S1) (13). These samples originated from 21 patients with stable, chronic infections (i.e., not undergoing an acute exacerbation) from Atlanta, Georgia and Copenhagen, Denmark. All samples were collected prior to September 2019 so the TRIKAFTA modulator therapy was not prescribed to any patients at the time of collection. These metatranscriptomes were chosen as they map with high specificity and high read coverage to *P. aeruginosa* PAO1 (Fig. 1A and Dataset S1), which is the primary strain used in our model systems. Across the 24 sputum samples, between 1.6×10^5 and 1.0×10^7 reads were assigned to *P. aeruginosa* PAO1 protein-coding genes after first mapping metatranscriptomes to a set of “decoy” bacterial genomes (Dataset S1). Of the 5586 protein-coding genes in *P. aeruginosa* PAO1, reads mapped to between 3,956 and 5,391 PAO1 genes across samples (Fig. 1A).

One important consideration is that when a gene is not detected in a metatranscriptome, it may be because the gene is not present in the *P. aeruginosa* strain(s) infecting the patient or because the gene is not expressed. Thus, it was important to identify PAO1 genes that were likely present in CF-infecting strains to avoid describing a gene as inaccurate when it was not present in the infecting strain. To identify this gene set, we performed two analyses. First, we identified PAO1 genes that were also present in at least 95% of 291 complete *P. aeruginosa* genomes, with the rationale that these genes have high probability of being present in CF-infecting strains. This analysis identified that 4,975 (89%) *P. aeruginosa* PAO1 protein-coding genes were also present in the other 290 strains (Fig. 1B and Dataset S2). Second, we identified PAO1 genes that were expressed in at least 95% (23 of 24) of the human CF metatranscriptomes. This analysis revealed that 3,777 PAO1 protein-coding genes were expressed in 95% of the human CF metatranscriptomes, including 172 that were present in less than 95% of *P. aeruginosa* genomes. Based on these findings, we limited our analyses to the 5,147 (4,975 + 172) genes that were either present in 95% of *P. aeruginosa* genomes and/or expressed in 95% of human sputum metatranscriptomes. Together, the 24 high-quality sputum metatranscriptomes and this “core set” of 5,147 genes allowed us to analyze *P. aeruginosa* gene expression during infection of the CF lung.

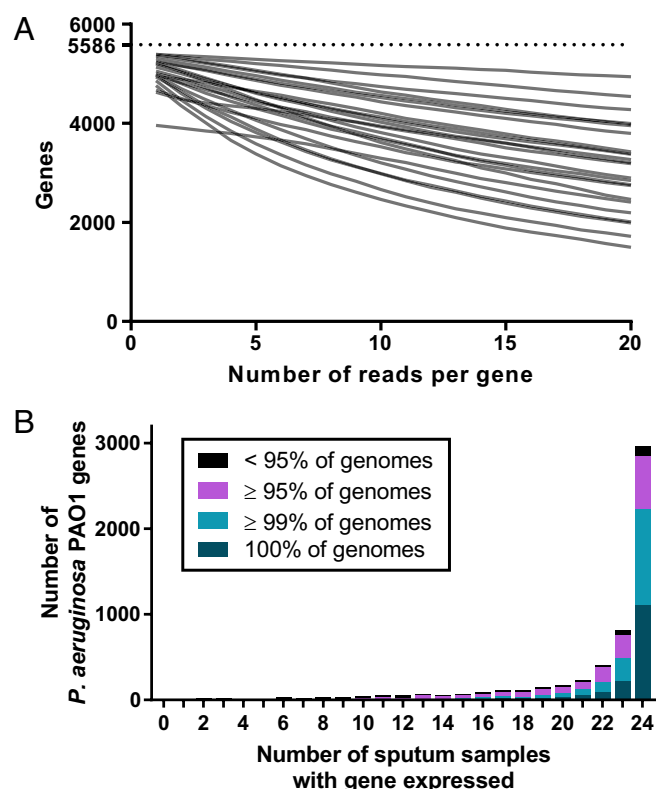


Fig. 1. *P. aeruginosa* gene expression in human CF sputum metatranscriptomes. (A) Coverage of the *P. aeruginosa* PAO1 genome. Each line represents the depth of coverage for PAO1 genes for an individual human sputum metatranscriptome, demonstrating sufficient coverage for all metatranscriptomes used in this study. The dashed line shows the 5,586 total protein coding genes in the PAO1 genome. (B) PAO1 gene expression across the 24 human sputum metatranscriptomes, relative to their prevalence in the *P. aeruginosa* pangenome. The pangenome was built using 291 complete *P. aeruginosa* genomes. The 5,147 genes included in all downstream analyses include those present in at least 95% of the *P. aeruginosa* genomes and/or expressed in 95% (23 or 24) of the metatranscriptomes.

Accuracy.

Rationale for improving models. The primary goal of this study was to develop approaches to improve the accuracy of experimental models to better mimic gene expression in an environment of interest. Our model accuracy framework is based on a direct comparison of the expression of each microbial gene in the natural environment and a laboratory model (7). The first step involves calculating the mean and SD of normalized read counts for each gene in the natural environment. Next, the average expression of each gene is calculated in the model and a z-score is determined. The z-score is defined as the number of SDs the mean expression of a gene is in the model from the mean in the natural environment. We use the absolute value of each gene's z-score as an indication of how similarly the gene is expressed between the model and natural environment. We define a model's accuracy score (AS_2) as the percentage of a microbe's genes that are expressed within two SDs of that in the natural environment. For example, if a model has an AS_2 of 90%, then the expression of 90% of a microbe's genes in the model fall within two SDs of the mean expression in the natural environment. Of note, we can also calculate the AS_2 of a natural environment relative to itself to understand the variation in gene expression across the samples. This was performed for *P. aeruginosa* gene expression in the 24 CF sputum transcriptomes by randomly choosing three samples as the “model” and comparing gene expression to the remaining 21 sputum metatranscriptomes. This analysis was repeated 1,200 times to obtain a mean AS_2 of 98.6%, which was similar to that previously calculated (7). This

value functions as an upper limit in AS_2 for this sample set and indicates high conservation in *P. aeruginosa* gene expression across the sputum metatranscriptomes.

In this study, we detail two generalizable approaches for improving the accuracy of experimental models. The first approach leverages our ability to choose two models that are inaccurate in different functional categories and then test the hypothesis that combining these models will be more accurate than the individual models. The second approach capitalizes on the vast resource of *P. aeruginosa* publicly available transcriptomes as a discovery tool to identify modifications that improve the accuracy of specific genes of interest. Together, the accuracy score framework and these two approaches for improving models are highly transferable to any experimental system.

Approach 1. Combining models to improve accuracy. We previously found that the two CF preclinical models with the highest AS_2 were an in vitro synthetic CF sputum medium (SCFM2) and a CF airway epithelial cell infection model (7). SCFM2 is a defined medium designed to mimic the chemistry and viscosity of expectorated CF sputum and has been used to advance our understanding of *P. aeruginosa* aggregate (biofilm) growth and physiology (14–18). The CF airway epithelial cell model involves addition of planktonic *P. aeruginosa* to the apical surface of immortalized CF airway epithelial cells that have been differentiated at the air–liquid

interface in vitro; it has been used extensively to study host–pathogen interactions (19–22). As SCFM2 and the CF airway epithelial cell model have similar AS_2 scores but are accurate in many different functional categories (SI Appendix, Fig. S1) (7), we hypothesized that combining these models would produce a model with increased accuracy compared to the individual models. To test this hypothesis, we developed a new model: the CF airway epithelial cell–SCFM2 model (epiSCFM2). In this model, *P. aeruginosa* PAO1 was grown for 16 h in SCFM2, aggregates were diluted into fresh SCFM2, and then added to the apical side of differentiated CF airway epithelial cells at a multiplicity of infection of 0.05 (Fig. 2A).

We first characterized this model over an 8-h time course, focusing on epithelial cell viability and *P. aeruginosa* growth, with the rationale that the utility of this model for studying host–pathogen interactions requires both viable epithelial cells and bacterial growth. Epithelial cell integrity, assessed by transepithelial electrical resistance (TEER), was not compromised by the addition of *P. aeruginosa* PAO1 (Fig. 2B) or *P. aeruginosa* CF clinical strains (SI Appendix, Fig. S2) at 8 h post infection. *P. aeruginosa* PAO1 grew well in the model with over 10^9 colony forming units/mL/cm² present at 8 h post infection, primarily in the apical supernatant (Fig. 2C). At 1 h post infection, *P. aeruginosa* was present in the apical supernatant and associated with the epithelial cell surface as

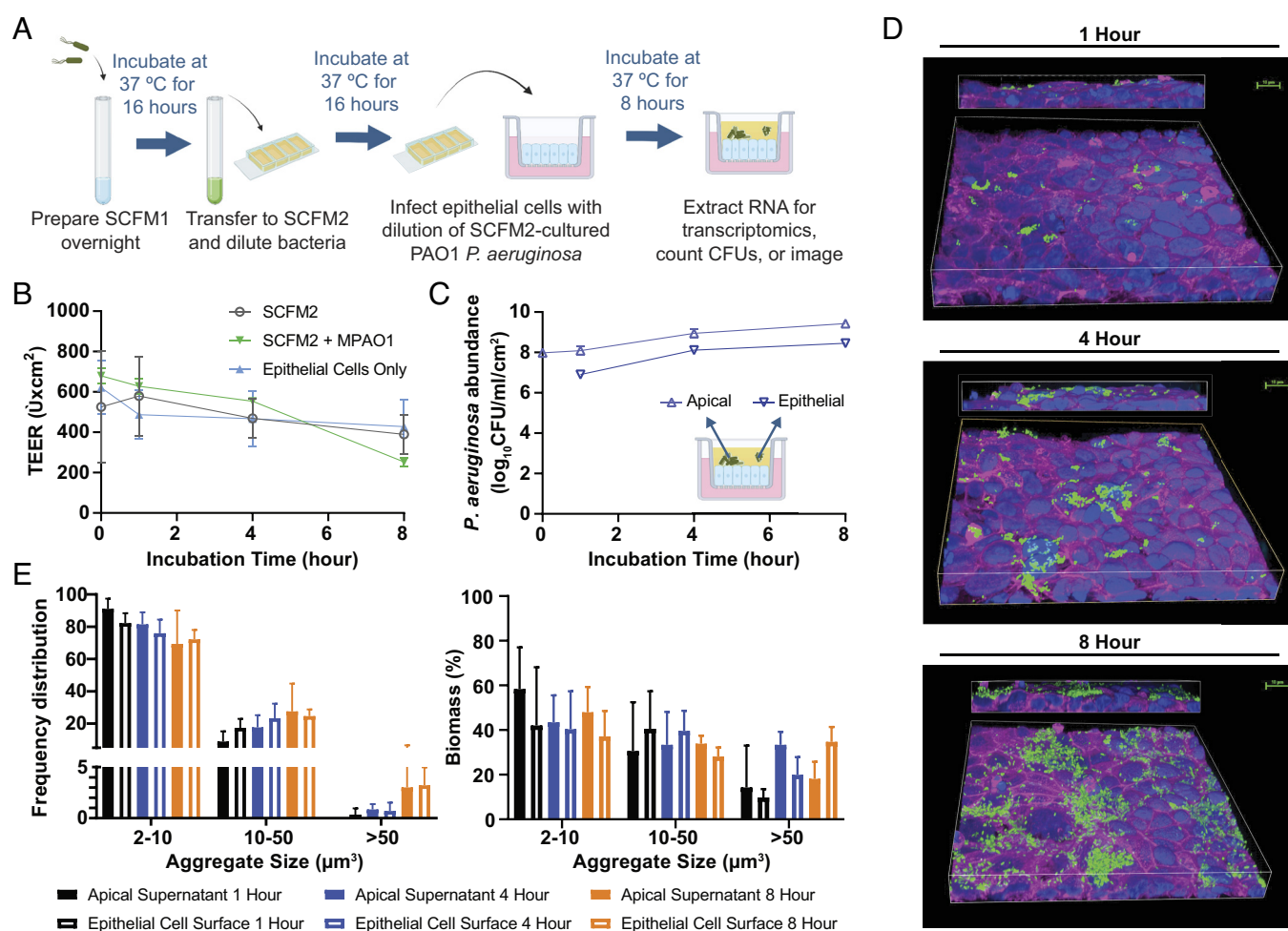


Fig. 2. *P. aeruginosa* growth and epithelial cell integrity in the epiSCFM2 model. (A) Experimental timeline for epiSCFM2. (B) Epithelial integrity measurement in epiSCFM2 model over 8-hr time course, as assessed by TEER. N = 3. (C) *P. aeruginosa* biofilm growth in epiSCFM2 over time. Apical, SCFM2 aggregates in airway lumen; epithelial, aggregates associated with CF airway epithelial cells at end of assay. N = 4. (D) Fluorescence imaging of bacterial aggregates in epiSCFM2. Blue, Hoechst staining of CF airway epithelial nuclei; purple, phalloidin staining of CF airway epithelial cell actin cytoskeleton; green, PAO1 *P. aeruginosa* expressing GFP. Representative images of 3 biological replicates are shown. (E) Distribution of *P. aeruginosa* aggregates in epiSCFM2. (Left) frequency distribution; (Right) % of total biomass for each aggregate size category. N = 3. Abbreviation: CFU = colony forming unit.

single bacterial cells and bacterial aggregates smaller than $50\ \mu\text{m}^3$ (Fig. 2 *D* and *E*). At 8 h post infection, aggregate size increased significantly with many larger than $50\ \mu\text{m}^3$. These data reveal that this model both supports epithelial cell health and *P. aeruginosa* growth.

We next quantified *P. aeruginosa* gene expression in the epiSCFM2 model with the goal of assessing model accuracy. Overall, the AS_2 for the epiSCFM2 model (87.8%) was significantly improved compared to the CF airway epithelial cell infection model (84.7%), but not SCFM2 (86.4%) (Fig. 3*A* and [Dataset S3](#)). Of note, there were three gene categories that increased in accuracy in the epiSCFM2 model (Fig. 3 *B* and *C*). The first are 171 genes that are accurate in both epithelial cell models, but not in SCFM2, indicating that the epithelial cells contribute to the accuracy of these genes. These genes encode housekeeping genes such as ribosomal proteins and elongation factors as well as iron uptake functions such as ferric enterobactin transport proteins and heme transport proteins. The second set is 338 genes that are accurate in SCFM2 and the epiSCFM2 model but not in the airway epithelial model, indicating the importance of SCFM2 for the accuracy of these genes. These genes encode functions such as pyochelin biosynthesis, type III secretion, and stress-related proteins. Finally, 70 genes are accurate only in the epiSCFM2 model, including at least ten genes involved in the synthesis and export of the siderophore, pyoverdine. In addition, although 107 and 128 genes lose accuracy in the epiSCFM2 model relative to SCFM2 and the airway epithelial cell model, respectively, many of these are borderline, with z-scores falling only slightly outside of 2 (Fig. 3 *B* and *D*). The ability of the epiSCFM2 model to capture aspects of *P. aeruginosa* gene expression in SCFM2 and the airway epithelial model is also visible using a principal component analysis (PCA; [SI Appendix, Fig. S4](#)). Thus, combining two models together into the epiSCFM2 model created a highly reproducible growth environment for *P. aeruginosa* and increased the accuracy of *P. aeruginosa* gene expression.

Approach 2. Targeting specific genes for improvement. Although a long-term goal for model development may be to improve expression of as many genes as possible, it is likely that in many cases, researchers will want to improve specific sets of genes/functions in a model of interest. Here, we focused on a set of genes whose expression we previously identified as effective in distinguishing between in vitro and human transcriptomes by conducting feature (gene) selection followed by ranking the selected genes by importance in distinguishing between the sample types; the top ten genes were reported (Fig. 4*A*) (6). Eight of these genes were inaccurate in SCFM2, including several genes that had the largest z-scores in SCFM2 and other in vitro models compared to human CF sputum (7). Our approach was to leverage publicly available *P. aeruginosa* transcriptomes to identify datasets in which the genes of interest are expressed at levels similar to that in human infection, with the rationale that the growth conditions for these transcriptomes will provide insight into an improvement strategy. We compared normalized gene expression of the ten genes of interest across 88 in vitro *P. aeruginosa* transcriptomes (6, 23). This analysis showed that in one condition, low zinc (23), five of the ten genes were accurately expressed (i.e., $|z\text{-score}| < 2$) when compared to expression in expectorated CF sputum (Fig. 4*A* and [Dataset S3](#)). Low zinc also increased expression of PA0781, although to a level substantially greater than that in human CF sputum.

Based on these findings, we hypothesized that low zinc is the missing cue in SCFM2 that impacts the accuracy of these genes. Although zinc is not an ingredient of SCFM2, it is likely present due to contamination of the chemicals used to construct this medium. Thus, it was not possible to simply reduce the amount

of zinc added to the medium. Therefore, to test our hypothesis, we reduced bioavailable zinc through chelation using the human protein complex calprotectin. Calprotectin sequesters transition metals including zinc, manganese, iron, and copper and is known to induce a zinc starvation response in *P. aeruginosa* (24, 25). Calprotectin is produced during inflammation and is present in CF sputum at high levels (150 to 1,000 $\mu\text{g}/\text{mL}$) (26). Thus, we added 400 $\mu\text{g}/\text{mL}$ calprotectin to SCFM2, grew *P. aeruginosa* PAO1 in this new model, and performed transcriptomics. Addition of calprotectin to SCFM2 resulted in accurate expression of eight out of ten target genes: all five of the low zinc responsive genes, PA0781, and the two genes accurate in SCFM2 (Fig. 4*B* and [Dataset S3](#)). As a control, a mutant calprotectin incapable of chelating zinc (27) was added to SCFM2 and did not alter the accuracy of these genes relative to the standard SCFM2 medium, indicating that calprotectin metal binding is essential for improved gene expression accuracy (Fig. 4*B*).

We next analyzed the impact of calprotectin on SCFM2 accuracy at the genome-wide scale. The addition of calprotectin significantly increased the AS_2 of SCFM2 from 86.4 to 89.9%, while addition of the mutant calprotectin had a similar AS_2 as SCFM2 alone (86.6%) (Fig. 4*C*). Thus, calprotectin addition to SCFM2 not only increased the accuracy of eight of our target genes but also 223 additional genes (Fig. 4 *D–F* and [Dataset S3](#)). Many of these genes, including those involved in cation binding and transport, have a large increase in expression in the presence of active calprotectin (Fig. 4*E*). Other functional categories that increase in accuracy with the addition of calprotectin include porins, energy metabolism, and ribosomal proteins (Fig. 4*F*). Although a small number of genes were accurate in SCFM2 or SCFM2 with the mutant calprotectin, but not accurate in SCFM2 with active calprotectin, their z-scores with calprotectin were borderline, falling just above 2 (Fig. 4*E*). The PCA also showed that the addition of calprotectin shifted the *P. aeruginosa* gene expression profile to be more similar to the profile in CF sputum ([SI Appendix, Fig. S4](#)). Of the hundred genes whose expression levels contributed most to the sample location along principal component 1, only five were accurate in SCFM2, but 39 were accurate in SCFM2 with calprotectin. These data indicate that a targeted approach to improve the accuracy of select genes is successful and can result in significant changes in genome-wide accuracy.

Identification of sixteen elusive genes that are not accurate across experimental models and strains. The epiSCFM2 and SCFM2-calprotectin models are two new models that together capture the expression of all but 281 *P. aeruginosa* PAO1 genes (6%) during CF lung infection (Fig. 5*A* and [Dataset S3](#)). Thus, with two in vitro models, it is possible to study most *P. aeruginosa* genes/functions in the context of human CF infection. To identify conditions that could be leveraged to capture these 281 genes and further improve our model systems, we analyzed gene accuracy in additional CF-specific models and control conditions (7). Addition of only two models—a mouse acute lung infection model and the CF clinical strain *P. aeruginosa* LESB58-SED21 grown in SCFM2—captured all but 54 *P. aeruginosa* PAO1 genes (1%) ([SI Appendix, Fig. S5](#)). We further asked which of these 54 genes were still elusive, even among 14 additional laboratory models commonly used for *P. aeruginosa*. This analysis identified sixteen *P. aeruginosa* PAO1 genes that were not accurately expressed in any of these experimental models (Fig. 5*B* and [Dataset S3](#)). These elusive genes include two genes involved in alginate biosynthesis (*alg44* and *algD*), a porin (*oprQ*), a pili protein (*pilF*), the post-transcriptional regulator *rsmN*, and seven hypothetical proteins. Many of these genes have large differences in expression between the experimental models and human CF sputum. In addition, four of these genes (PA3237, *algD*, *alg44*, and PA4883) show extremely

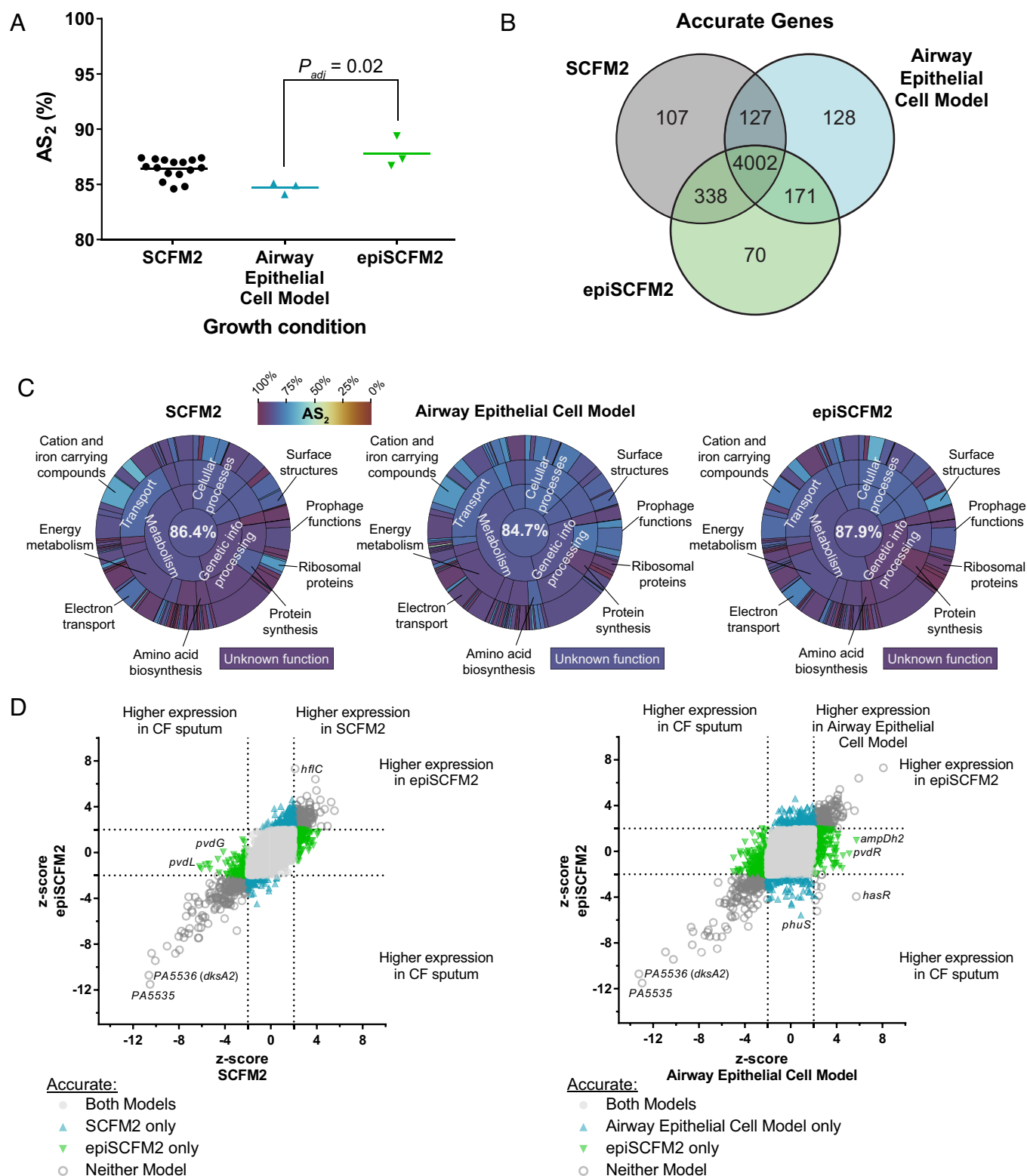


Fig. 3. Increased accuracy by combining the airway epithelial model and SCFM2 to make epiSCFM2. (A) AS_2 for each replicate of each model. Significant differences are shown using Kruskal-Wallis and Dunn's Multiple Comparison Tests. (B) Venn diagram showing the shared and unique accurate genes in each model. (C) AS_2 for TIGRFAM subcategories for *P. aeruginosa* PAO1 in SCFM2, the airway epithelial cell model, and epiSCFM2. The color in the middle represents the average AS_2 across individual replicates for all PAO1 genes (those with and without TIGRFAM designations). The next level out from the middle of the circle contains TIGRFAM "meta roles", the next contains TIGRFAM "main roles", and the outer-most layer contains TIGRFAM "sub roles". The area of each category is proportional to the number of genes in that category. See *SI Appendix, Fig. S3* for plots *P. aeruginosa* gene expression in the apical supernatant and epithelial surface of epiSCFM2. (D) The average z-score for each gene in each condition is shown, and genes are colored based on the conditions in which they are accurate. Genes are considered accurate when their z-score is between -2 and 2 . Annotations are shown for genes of interest.

low expression in all experimental models. Finally, some of these sixteen genes, such as *algD*, have high variance in CF sputum, emphasizing that the lack of accurate expression in laboratory

models is not necessarily because their in vitro expression has to fall within a narrow window to be called accurate (Fig. 5B, *SI Appendix, Fig. S6*, and *Dataset S3*). The identification of sixteen

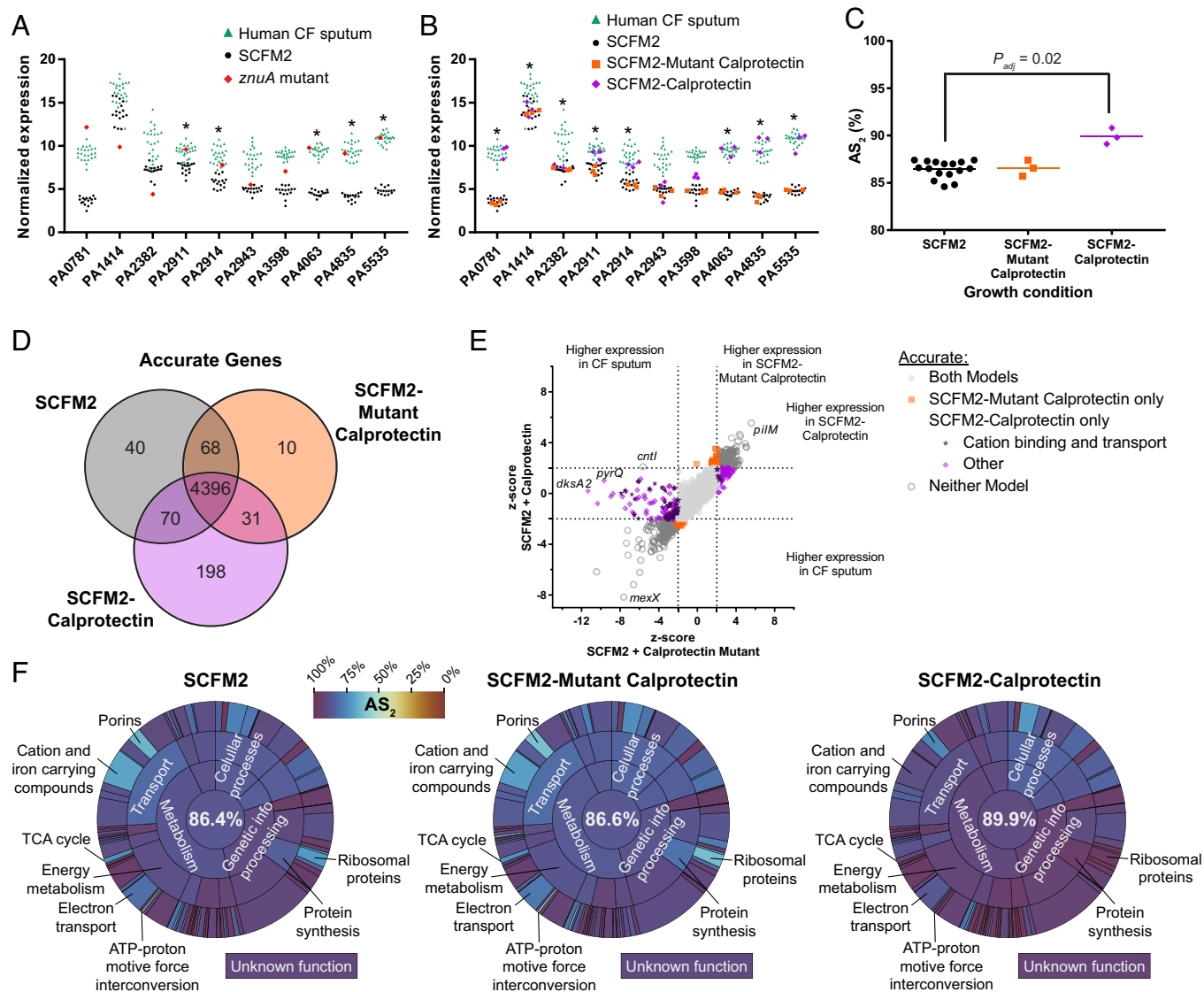


Fig. 4. Addition of calprotectin improves the accuracy of SCFM2. (A) Variance stabilizing transformation (VST)-normalized expression during growth in CF sputum, SCFM2, and low zinc (*znuA* mutant) for 10 genes previously identified using a supervised learning model as the most diagnostic for *P. aeruginosa* growth in humans (6, 23). Asterisks indicate genes with accurate gene expression during zinc-limitation ($|z\text{-score}| < 2$). (B) VST-normalized expression during growth in CF sputum, SCFM2, SCFM2-Mutant Calprotectin, and SCFM2-Calprotectin for 10 genes previously identified using a supervised learning model as the most diagnostic for *P. aeruginosa* growth in humans (6). Asterisks indicate genes with accurate gene expression in SCFM2-Calprotectin ($|z\text{-score}| < 2$). (C) AS_2 for each replicate of each model. Significant differences are shown using Kruskal-Wallis and Dunn's Multiple Comparison Tests. (D) Venn diagram showing the shared and unique accurate genes in each model. (E) The average z-score for each gene in SCFM2-Mutant Calprotectin and SCFM2-Calprotectin, demonstrating how the addition of metal-binding calprotectin alters the accuracy of PAO1 genes. Genes are considered accurate when their z-score is between -2 and 2 . Genes are colored based on the conditions in which they are accurate, and annotations are shown for select genes. (F) AS_2 for TIGRFAM subcategories for *P. aeruginosa* PAO1 genes in SCFM2, SCFM2-Mutant Calprotectin, and SCFM2-Calprotectin. The color in the middle represents the average AS_2 across individual replicates for all PAO1 genes (those with and without TIGRFAM designations). The next level out from the middle of the circle contains TIGRFAM meta roles, the next contains TIGRFAM main roles, and the outer-most layer contains TIGRFAM sub roles. The area of each category is proportional to the number of genes in that category.

P. aeruginosa PAO1 genes that are not accurately expressed in any experimental model emphasizes the potential for iterative refinement of model systems using the approaches outlined here to capture these genes.

Discussion

One of the biggest challenges to understanding a microbe's role in the natural environment is the lack of relevant experimental models. Significant focus and resources are put toward improving laboratory models, but this work is largely based on prior experience and intuition. Here, we develop and validate two quantitative frameworks that can be used across microbial systems for assessing and improving the accuracy of experimental models.

Both frameworks aim to improve known inaccuracies in models, one by combining experimental models and the other by targeting specific genes of interest. We show that these two approaches lead to the accurate expression of up to 90% of genes in a single model relative to the native environment and accurate expression of all but 281 genes across these two new in vitro models, allowing for the investigation of most microbial functions. These validated, straightforward model improvement frameworks are generalizable and can be implemented for any environment and any model microbial system. While in this work we have defined accuracy as two SDs from the mean expression in the CF lung, other more stringent criteria (i.e., one SD) could be used for both model selection and refinement (SI Appendix, Fig. S7). Also, as more *P. aeruginosa* CF sputum transcriptomes are obtained, the

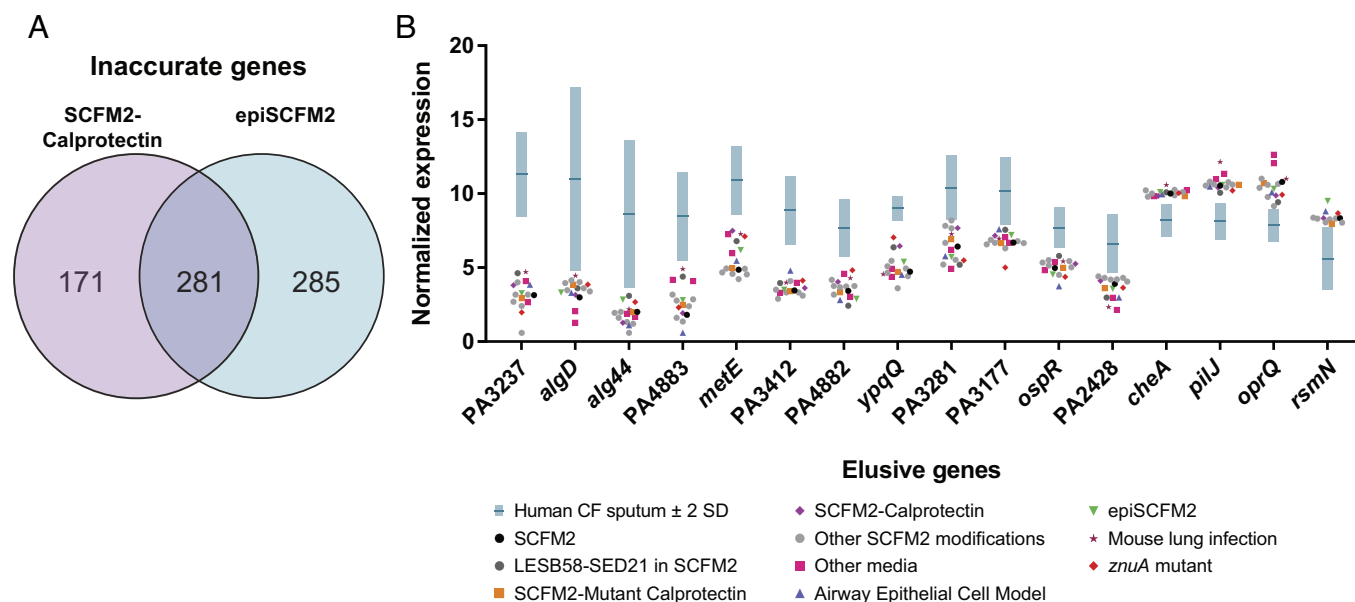


Fig. 5. Elusive genes across 18 growth conditions and two strains. (A) Venn diagram showing the shared and unique inaccurate genes across SCFM2-Calprotectin and epiSCFM2 (7). (B) The mean VST-normalized expression of *P. aeruginosa* genes in human CF sputum and a range of experimental models for the 16 genes that are inaccurate in all conditions. The error bars for human CF sputum represent \pm two SDs, which is the range considered accurate. Other SCFM2 modifications include SCFM1, use of Casamino acids instead of defined amino acids, addition of yeast extract, addition of hemin and removal of iron, addition of polymyxin B, and addition of vitamins. Other media includes MOPS-succinate and LB. Note that there are no data for *rsmN* for the strain *P. aeruginosa* LESB58-SED21 in SCFM2, MOPS-succinate, LB, or the mouse lung model, as this gene was not included in the analyses in ref. 7.

confidence in calling a gene accurate or not will likely improve for many genes, particularly for genes whose expression is not normally distributed.

Combining models is a common technique in the development of new laboratory models. Traditionally, the models that are combined are often chosen because they are thought to build the complexity of the system to better capture multiple aspects of the natural environment (28). Our approach is novel because it leverages a quantitative framework to identify the models to combine based on the rationale that a new, more accurate model would result from the combination of models that are complimentary in their accuracies. In our proof-of-concept demonstrated here, the combined epiSCFM2 model was largely additive from the component models (Fig. 3). However, the new model was also unique from both SCFM2 and the airway epithelial model in surprising ways, emphasizing that we do not fully understand the biology of *P. aeruginosa* or the CF infection environment. It should be noted that epiSCFM2 does not contain immune cells, and we propose that their addition to this model could potentially improve the overall accuracy and is the subject of future study.

We also showed that model improvement can be accomplished by leveraging publicly available transcriptomes to improve the accuracy of target genes. This approach relies on the massive repositories of published RNA sequencing (RNA-seq) datasets. Transcriptomes are most valuable when they are from diverse but defined environments or from known mutants, as these should capture the range of bacterial gene expression and allow for clear inferences of how to modify laboratory models. Here, we used data from the deletion of a *P. aeruginosa* gene, *znuA*, encoding a high-affinity zinc-binding protein that improved the accuracy of many of our target genes and emphasized the importance and challenge of capturing the native metal environment in experimental models (23). Addition of the zinc-binding calprotectin to SCFM2 resulted in improved accuracy of the target genes as well as an additional 223 genes. Unlike in the combination of models approach, the increased AS_2 in SCFM2 with calprotectin was driven by a small subset of genes with large shifts in z-score.

As calprotectin, but not the mutant calprotectin, binds manganese and iron in addition to zinc, these additional genes may also be due to manganese/iron limitation. Regardless, it is clear that the targeted approach to improve models is powerful for improving the accuracy of specific genes. The continued production of diverse, well-documented transcriptomes will be important for the wide-spread application of this targeted approach, especially in taxa that are not as deeply studied as *P. aeruginosa*.

There are additional straightforward approaches to model improvement, most notably the use of different bacterial strains. This work focused on the use of the laboratory strain *P. aeruginosa* PAO1, originally isolated in 1954 (referred to as strain 1) from a human wound in Melbourne, Australia (29). We chose this strain as it and *P. aeruginosa* PA14 are the most well-studied strains. In addition, PAO1 grows well and is easily genetically manipulated in the lab, and there exist significant community resources for this strain including an ordered transposon mutant library (30, 31). Thus, we propose that developing highly accurate models for PAO1 is of utmost importance, particularly for laboratories focused broadly on evolution and ecology that have limited molecular microbiology expertise. However, it is well known that during long-term colonization of the CF lung, *P. aeruginosa* can acquire dozens, or even hundreds, of mutations (32–36) that can alter gene expression. Thus, CF-adapted strains would be more appropriate if the focus is on these commonly mutated genes/functions, evidenced by the high accuracy of these genes when the CF-adapted strain LESB58-SED21 was grown in SCFM2 (7). An additional approach is to assess the impact of other microbes on the accuracy of these models using *P. aeruginosa* as the focal species. While one might predict that addition of other CF microbes will increase the accuracy of *P. aeruginosa* gene expression, it is important to point out that the outcome of microbe–microbe interactions is dependent on the environment and the biogeography of the infection (37–39). Thus, these experiments may not be as straightforward as some may predict. However, we propose that this approach is essential to uncover the importance of microbe–microbe interactions in the CF lung.

In our previous analysis, we found 211 elusive genes across five CF model systems (7). Here, using our improved model systems, we capture the accuracy of all but 281 genes using two improved in vitro models (Fig. 5A) and all but 54 genes using four models (SI Appendix, Fig. S5). The identification of 16 genes that were inaccurate in 18 models (Fig. 5B) shows the utility of the model improvement framework, and we propose that the targeted approach for model improvement could be powerful to improve the accuracy of these genes. Alternatively, the accuracy of these genes could be improved in any model using genetic approaches to control their expression. To study their ecological function, it is especially important to identify accurate growth conditions for these genes, as many had negligible expression across experimental models but were highly expressed in human infections. It should be noted that the identification of 16 elusive genes is likely conservative, as the use of large numbers of models increases the likelihood that a gene may not be deemed inaccurate even though it is. Regardless, our data indicate that there are clearly *P. aeruginosa* genes that are not captured by any of the commonly used CF models.

This iterative approach for model improvement leads to the question, what is the maximum accuracy possible for a single model? With the specialist oral pathogen, *Porphyromonas gingivalis*, we found that laboratory growth can result in an AS₂ of 96% (8). However, the *P. aeruginosa* genome contains over twice as many genes as *P. gingivalis*, and the gene regulatory architecture in *P. aeruginosa* is significantly more complex. Thus, it is not surprising that it is more difficult to reach this high accuracy and that as the accuracy of some functions is improved, others are decreased. Future modification of *P. aeruginosa* and similar studies with other microbes will be important for understanding the limits of model accuracy across diverse organisms.

While this paper focused on gene expression similarities between laboratory models and human infection, other aspects of host-microbe or microbe-microbe interactions are also important to consider. It would be highly valuable to improve models to better capture gene expression of the entire microbial community, not just individual species. For example, as calprotectin likely modifies functions important for microbe-microbe interactions (25, 40), we hypothesize that it would also alter the accuracy of a multispecies model. In addition, it is also important to consider the micron-scale biogeography of a microbial community, which impacts the function(s) of that community and interactions with the host (37–39, 41). Measurements of the *P. aeruginosa* aggregates in epiSCFM2 (Fig. 2D and E) revealed that while they overlap in size with aggregates in expectorated CF sputum, there were fewer aggregates larger than 50 μm^3 in epiSCFM2 than observed in CF sputum (42–45). Thus, a similar quantitative approach to that used here could be used to quantify the size and spatial patterning of aggregates in model systems, assess accuracy using data from CF sputum (45), and potentially refine biogeographical characteristics of these models. We propose that other important aspects of model systems, such as host response or single-cell heterogeneity, could also be quantified and refined using a similar approach.

We anticipate that future work will apply these model improvement frameworks broadly across microbial taxa and environments. In addition to studies of other human infections, accuracy scores can also be applied to improve models of other important microbial systems, such as the oceans and soils, facilitating more accurate analyses of the microbial contribution to global warming, nutrient cycling, and agriculture. While we have focused on RNA-seq here because of its applicability to natural samples, our framework could be applied to other functional measurements such as proteomics data. Together, this work will be critical for improving our understanding of microbes across their natural environments.

Finally, it is important to point out that the goal of this work is not to define a model as “good” or “bad”. All models, including common laboratory test-tube growth and animal models of infection, have strengths and weaknesses. Instead, we view these frameworks as benefiting the needs of individual labs and fields, providing guidance for model choice as well as a means of quantitatively addressing weaknesses in new or well-established models. In addition, we have demonstrated the value of integrating data from multiple sources and perspectives. There is a clear need in biology to develop robust models that capture important characteristics of the system studied, and we propose that the straightforward approaches developed here will be highly impactful in multiple fields.

Materials and Methods

CF Sputum Collection. Expectorated CF sputum was collected in RNeasy lysis buffer from the Emory-Children’s Center for Cystic Fibrosis and Airways Disease Research by the Cystic Fibrosis Biospecimen Laboratory as previously described (6). Patients had stable infections, were not undergoing acute exacerbation, and had not been given the TRIKAFTA modulator therapy. Collections were approved by Georgia Tech IBC protocol H18220.

***P. aeruginosa* In Vitro Growth Conditions.** *P. aeruginosa* MPA01 (30) was used for all in vitro experiments. SCFM1 and SCFM2 were prepared as previously described (14, 15). SCFM2 was also modified with the following additions or alterations: 1) Calprotectin and mutant calprotectin were purified as described (27, 46) and added to SCFM2 at 400 $\mu\text{g}/\text{mL}$; 2) Casamino acids were substituted for the individual amino acids in SCFM2 at a ratio of 3.33 g to 250 mL SCFM2; 3) 0.05% yeast extract was added to SCFM2; 4) the iron in SCFM2 was substituted with 5 μM hemin; 5) 0.8 $\mu\text{g}/\text{mL}$ polymyxin B was added to SCFM2; 6) and the 1 \times vitamin mix from chemically defined medium (47) was added to SCFM2.

For growth conditions other than those involving the epithelial cells, cells were grown overnight in the SCFM1 medium. Overnight cultures were inoculated into the indicated medium at a final OD of 0.05 and grown statically at 37 °C. The culture conditions are indicated in Dataset S1 for each condition. Both deep-well 96-well plates and 4-well chamber slides had a total volume of 500 μL of medium per sample. After 6 h, cultures were preserved in RNeasy lysis buffer (ThermoFisher) for RNA extraction.

CF Airway Epithelial Cell-SCFM2 Model. Immortalized homozygous CFTR $\Delta\text{F508CFBe410-}$ human bronchial epithelial cells (obtained from J.P. Clancy, Cincinnati Children’s Hospital) were maintained in a humidified incubator at 37 °C and 5% CO₂. The cells were fed with growth media containing minimum essential medium (MEM) with phenol red supplemented with 10% fetal bovine serum (Gemini Bio-Products), 2 mM L-glutamine, 5 U/mL penicillin, and 5 $\mu\text{g}/\text{mL}$ streptomycin (Sigma). For biofilm assays, TEER, and confocal imaging, CFBe410- epithelial cells were seeded at 2×10^5 cells per 6.5-mm transwell permeable membrane supports (Corning). CFBe410- epithelial cells were seeded at 1.5×10^6 cells per 24-mm transwell permeable-membrane supports (Corning) for RNA isolation. Cells were used 18 to 21 d after seeding. The day before inoculation, cells were washed in MEM lacking phenol red and fed with antibiotics-free media containing MEM with phenol red supplemented with 10% fetal bovine serum (Gemini Bio-Products) and 2 mM L-glutamine (antibiotic-free).

P. aeruginosa MPA01 was grown planktonically in SCFM1 shaking at 250 rpm at 37 °C overnight. On the following day, the culture was diluted to an OD of 0.05 in SCFM2 in chamber slides (ThermoFisher) and grown statically in humidified incubator at 37 °C and 5% CO₂ for 16 h. CFBe410- were inoculated in duplicate with *P. aeruginosa* and SCFM2 coculture at a multiplicity of infection of 0.05. Apical growth media was removed from the transwell supports and 495 μL of SCFM2 was placed on the apical side. Then, 5 μL of the 16-h SCFM2-grown *P. aeruginosa* culture was pipetted directly into 495 μL SCFM2 on top of the epithelial cells. The bacterial and epithelial cells were incubated in a humidified incubator at 37 °C for 8 h and then collected for bacterial enumeration or added to 1 mL RNABee (Amsbio) and frozen at –80 °C for RNA extraction.

Imaging of the epiSCFM2 Model. For imaging analysis using confocal laser scanning microscopy, the infection experiment was conducted using the following methods. CFBe410- epithelial cells were seeded at 2×10^5 per 6.5-mm transwell

permeable-membrane supports (Corning). At the day of the experiment, the epithelial nuclei were prestained with Hoechst 33342 (H3570, Sigma) and inoculated as described above with a SCFM2-grown culture of *P. aeruginosa* MPAO1 carrying pMQ361g, a tdTomato fluorescent plasmid. At 1, 4, and 8 h, the SCFM2 apical supernatant was removed, and the bacteria were fixed with 4% paraformaldehyde at 4 °C overnight. Fixed cultures on permeable membrane supports were then washed with phosphate-buffered saline, stained with Phalloidin 647, and fixed to microscope slides with Prolong Gold (ThermoFisher). Each biological replicate was imaged with eight fields of view using a Nikon C2 confocal microscope. Images were rendered and bacterial biomass was quantified using Nikon Elements software.

RNA Extraction and Sequencing. RNA was extracted from human CF sputum samples and experimental model systems following the protocol in ref. 7. Briefly, samples were thawed, RNAlater was removed, and samples were resuspended in RNase-free TE buffer containing lysozyme and lysostaphin. Samples were incubated for 30 min at 37 °C for enzymatic lysis, RNA-Bee (Amsbio) was added, and samples were bead beat 3 × for 30 s, placing on ice in between each round. Chloroform was added to each sample, samples were mixed and centrifuged to separate phases. The aqueous phase was removed to a new tube, and the RNA was precipitated with isopropanol. When necessary, RNA was fragmented using the NEBNext Magnesium fragmentation module (New England Biosciences) per the manufacturer's instructions. Sequencing libraries were prepared with the NEBNext Small RNA Library prep kit (New England Biosciences), and adapter dimers were removed by size selection on a 5% TBE polyacrylamide gel. rRNAs were removed using the Illumina RiboZero Gold (epidemiology) kit, the Qiagen QIAseq FastSelect Kits (HMR and 5S/16S/23S), the Invitrogen MICROExpress Bacterial mRNA Enrichment Kit, or the New England Biosciences NEBNext rRNA Depletion Kit with Human/Mouse/Rat and Bacterial probes mixed 1:1, as indicated in [Dataset S1](#). All samples were sequenced at the Molecular Evolution Core at the Georgia Institute of Technology on an Illumina NextSeq500 using 75-bp single-end runs.

***P. aeruginosa* Genome Annotation and Pangenome Analysis.** To identify genes that are present in most *P. aeruginosa* strains, the 291 genomes designated as "complete" on NCBI as of April 29, 2021 were downloaded and analyzed with Roary v3.13.0 using 90% as the minimum percentage identity for blastp ([Dataset S2](#)) (48). Core genes were defined as the 5,147 PAO1 genes that were either present in 95% of *P. aeruginosa* genomes and/or expressed in 95% of human sputum samples ([Dataset S2](#)). Hierarchical TIGRFAM annotations were expanded from those used in Cornforth et al., 2020 (7) to provide hand-curated annotations for all PAO1 core genes, with "unknown function" being used in all cases where the annotation was unknown (49).

RNA-seq Analysis. RNA-seq read quality was confirmed with FastQC v0.11.8 (50). Reads were trimmed using Cutadapt v2.6 to remove Illumina adapters from the 3' end, and reads were retained that were at least 22 base pairs long (51). To remove non-*P. aeruginosa* reads, trimmed reads were mapped with bowtie2 v2.3.5 using default parameters to a metagenome of 105 decoy strains from 59 species, based on species previously identified in sputum samples ([Dataset S1](#)) (7, 52). The reads that did not map to the decoy genomes were then mapped to *P. aeruginosa* PAO1 (Accession number GCF_000006765.1) using bowtie2. featureCounts v2.0.1 was used to assign mapped reads to PAO1 genes with the flags -s 1 (stranded) and -O (allowMultiOverlap) so that each read was assigned to a single locus or to neighboring genes (53). The exception was the *znuA* mutant RNA-seq data, which is reverse-stranded, so mapped reads were assigned with

featureCounts flags -s 2 (reverse stranded) and -O. At each step, MultiQC v1.10 was used to track analysis quality (54).

All downstream RNA-seq analyses were performed using count data for the 5,147 core genes, with VST-normalization performed on all samples together in DESeq2 v1.28.1 with blind = TRUE in R (55).

AS₂ Analyses. AS₂ values were calculated following the approach developed in ref. 7, and all analysis scripts are available at https://github.com/glew8/PA_ModelAccuracy. First, we calculated the mean and SD of expression of VST-normalized count data for the human expectorated CF sputum samples. Then, z-scores and AS₂ values were calculated for each experimental model system using two approaches. In the first approach, which was used to calculate the overall AS₂ for each model and the AS₂ for the resampled sputum metatranscriptomes, z-scores for each gene were calculated for each replicate of an experimental model, the z-scores were rounded to 4 decimal places, and then the AS₂ for each replicate was calculated as the percentage of genes with a z-score between -2 and 2. The ROUT method in GraphPad Prism was used to identify outliers, which identified sample Rn_PAO1_SCFM2_15 as an outlier; this sample was excluded from all further analyses (56). Then, the overall AS₂ for the experimental model was calculated using the mean AS₂ score across replicates. In the second approach, which was used to compare across models at the gene and functional levels (e.g., in Venn Diagrams and TIGRFAM analyses), the z-score of each gene was calculated as the mean z-score across replicates, rounded to 4 decimal places. Then, the AS₂ for each functional category was calculated as the percentage of genes with a z-score between -2 and 2. "Sunburst plots" showing the TIGRFAM annotations were constructed using ggsunburst and the scico palette "roma0" (57, 58).

Data, Materials, and Software Availability. All RNA-seq reads are available in the NCBI Sequence Read Archive, as shown in [Dataset S1](#), including new datasets from this paper which are available under BioProject [PRJNA909326](#) (13). Raw and normalized count data and z-scores are available in [Dataset S3](#). The hand curated annotation of the *P. aeruginosa* MPAO1 genome and the pangenome data are available in [Dataset S2](#).

ACKNOWLEDGMENTS. We would like to acknowledge members of the Whiteley, Bomberger, and Goldberg labs for data analysis discussions. This study was supported by grants from the Cystic Fibrosis Foundation (WHITEL20AO to J.B.G., J.M.B., and M.W. and WHITEL22GO to M.W.), the Shurl and Kay Curci Foundation, the Cystic Fibrosis Trust Foundation SRC017 (to M.W.), and the NIH (K99DE031018 to G.R.L., R01AI101171 and R01AI127793 to W.J.C.). M.W. is a Burroughs Wellcome Investigator in the Pathogenesis of Infectious Disease.

Author affiliations: ^aSchool of Biological Sciences and Center for Microbial Dynamics and Infection, Georgia Institute of Technology, Atlanta, GA 30332; ^bEmory-Children's Cystic Fibrosis Center, Atlanta, GA 30332; ^cDepartment of Microbiology and Molecular Genetics, University of Pittsburgh, Pittsburgh, PA 15219; ^dDepartment of Pediatrics, Division of Pulmonary, Asthma, Cystic Fibrosis, and Sleep, Emory University School of Medicine, Atlanta, GA 30322; ^eDepartment of Biochemistry, Vanderbilt University, Nashville, TN 37232; ^fDepartment of Chemistry, Vanderbilt University, Nashville, TN 37232; ^gCenter for Structural Biology, Vanderbilt University, Nashville, TN 37232; and ^hDepartment of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN 37232

Author contributions: G.R.L., A.K., D.M.C., R.P.D., J.B.G., J.M.B., and M.W. designed research; G.R.L., A.K., R.P.D., F.L.D., D.A.M., S.A.H., J.M.B., and M.W. performed research; S.A.H., E.P.S., and W.J.C. contributed new reagents/analytic tools; G.R.L., A.K., D.M.C., R.P.D., D.A.M., J.B.G., J.M.B., and M.W. analyzed data; and G.R.L., A.K., D.M.C., R.P.D., D.A.M., E.P.S., W.J.C., J.B.G., J.M.B., and M.W. wrote the paper.

1. J. G. Hanley, Models and microbiology: Pasteur and the body. *Can Bull Med. Hist* **20**, 419–435 (2003).
2. D. Masemann, S. Ludwig, Y. Boergeling, Advances in transgenic mouse models to study infections by human pathogenic viruses. *Int. J. Mol. Sci.* **21**, 9289 (2020).
3. J. R. Swearingen, Choosing the right animal model for infectious disease research. *Animal Model Exp. Med.* **1**, 100–108 (2018).
4. K. Zengler et al., EcoFABs: Advancing microbiome science through standardized fabricated ecosystems. *Nat Methods* **16**, 567–571 (2019).
5. K. Zhalnina, K. Zengler, D. Newman, T. R. Northen, Need for laboratory ecosystems to unravel the structures and functions of soil microbial communities mediated by chemistry. *mBio* **9**, e01175-18 (2018).
6. D. M. Cornforth et al., *Pseudomonas aeruginosa* transcriptome during human infection. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E5125–E5134 (2018).
7. D. M. Cornforth, F. L. Diggle, J. A. Melvin, J. M. Bomberger, M. Whiteley, Quantitative framework for model evaluation in microbiology research using *Pseudomonas aeruginosa* and cystic fibrosis infection as a test case. *mBio* **11**, e03042-19 (2020).
8. G. R. Lewin, K. S. Stocke, R. J. Lamont, M. Whiteley, A quantitative framework reveals traditional laboratory growth is a highly accurate model of human oral infection. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2116637119 (2022).
9. K. De Boeck, Cystic fibrosis in the year 2020: A disease with a new face. *Acta Paediatr.* **109**, 893–899 (2020).
10. M. Shteinberg, I. J. Haq, D. Polineni, J. C. Davies, Cystic fibrosis. *Lancet* **397**, 2195–2211 (2021).
11. N. L. Turcios, Cystic fibrosis lung disease: An overview. *Respir Care* **65**, 233–251 (2020).
12. S. Malhotra, D. Hayes Jr., D. J. Wozniak, Cystic fibrosis and *Pseudomonas aeruginosa*: The host-microbe interface. *Clin. Microbiol. Rev.* **32**, e00138-18 (2019).
13. G. R. Lewin et al., Application of a quantitative framework to improve the accuracy of a bacterial infection model. NCBI Sequence Read Archive. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA909326>. Deposited 6 December 2022.
14. K. L. Palmer, L. M. Aye, M. Whiteley, Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J. Bacteriol.* **189**, 8079–8087 (2007).

15. K. H. Turner, A. K. Wessel, G. C. Palmer, J. L. Murray, M. Whiteley, Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 4110–4115 (2015).
16. S. E. Darch *et al.*, Spatial determinants of quorum signaling in a *Pseudomonas aeruginosa* infection model. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4779–4784 (2018).
17. K. L. Palmer, S. A. Brown, M. Whiteley, Membrane-bound nitrate reductase is required for anaerobic growth in cystic fibrosis sputum. *J. Bacteriol.* **189**, 4449–4455 (2007).
18. K. L. Palmer, L. M. Mashburn, P. K. Singh, M. Whiteley, Cystic fibrosis sputum supports growth and cues key aspects of *Pseudomonas aeruginosa* physiology. *J. Bacteriol.* **187**, 5267–5277 (2005).
19. M. R. Hendricks *et al.*, Extracellular vesicles promote transkingdom nutrient transfer during viral-bacterial co-infection. *Cell Rep.* **34**, 108672 (2021).
20. M. R. Hendricks *et al.*, Respiratory syncytial virus infection enhances *Pseudomonas aeruginosa* biofilm growth through dysregulation of nutritional immunity. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 1642–1647 (2016).
21. S. Moreau-Marquis *et al.*, The DeltaF508-CFTR mutation results in increased biofilm formation by *Pseudomonas aeruginosa* by increasing iron availability. *Am. J. Physiol. Lung Cell Mol. Physiol.* **295**, L25–37 (2008).
22. A. C. Zemke *et al.*, Dispersal of epithelium-associated *Pseudomonas aeruginosa* biofilms. *mSphere* **5**, e00630-20 (2020).
23. V. G. Pederick *et al.*, ZnuA and zinc homeostasis in *Pseudomonas aeruginosa*. *Sci. Rep.* **5**, 13139 (2015).
24. D. M. Vermilyea, A. W. Crocker, A. H. Gifford, D. A. Hogan, Calprotectin-mediated zinc chelation inhibits *Pseudomonas aeruginosa* protease activity in cystic fibrosis sputum. *J. Bacteriol.* **203**, e0010021 (2021).
25. C. A. Wakeman *et al.*, The innate immune protein calprotectin promotes *Pseudomonas aeruginosa* and *Staphylococcus aureus* interaction. *Nat. Commun.* **7**, 11951 (2016).
26. R. D. Gray *et al.*, Sputum and serum calprotectin are useful biomarkers during CF exacerbation. *J. Cyst. Fibros.* **9**, 193–198 (2010).
27. T. E. Kehl-Fie *et al.*, Nutrient metal sequestration by calprotectin inhibits bacterial superoxide defense, enhancing neutrophil killing of *Staphylococcus aureus*. *Cell Host Microbe.* **10**, 158–164 (2011).
28. N. Vogt, Modeling multi-organ systems on a chip. *Nat. Methods* **19**, 641 (2022).
29. B. W. Holloway, Genetic recombination in *Pseudomonas aeruginosa*. *J. Gen. Microbiol.* **13**, 572–581 (1955).
30. M. A. Jacobs *et al.*, Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 14339–14344 (2003).
31. S. Lewenza *et al.*, Construction of a mini-Tn5-luxCDABE mutant library in *Pseudomonas aeruginosa* PAO1: A tool for identifying differentially regulated genes. *Genome Res.* **15**, 583–589 (2005).
32. L. Freschi *et al.*, The *Pseudomonas aeruginosa* pan-genome provides new insights on its population structure, horizontal gene transfer, and pathogenicity. *Genome Biol. Evol.* **11**, 109–120 (2019).
33. R. L. Marvig *et al.*, Within-host microevolution of *Pseudomonas aeruginosa* in Italian cystic fibrosis patients. *BMC Microbiol.* **15**, 218 (2015).
34. R. L. Marvig, L. M. Sommer, L. Jelsbak, S. Molin, H. K. Johansen, Evolutionary insight from whole-genome sequencing of *Pseudomonas aeruginosa* from cystic fibrosis patients. *Future Microbiol.* **10**, 599–611 (2015).
35. R. L. Marvig, L. M. Sommer, S. Molin, H. K. Johansen, Convergent evolution and adaptation of *Pseudomonas aeruginosa* within patients with cystic fibrosis. *Nat. Genet.* **47**, 57–64 (2015).
36. L. M. Sommer *et al.*, Is genotyping of single isolates sufficient for population structure analysis of *Pseudomonas aeruginosa* in cystic fibrosis airways? *BMC Genomics* **17**, 589 (2016).
37. S. Azimi, G. R. Lewin, M. Whiteley, The biogeography of infection revisited. *Nat. Rev. Microbiol.* **20**, 579–592 (2022).
38. J. P. Barraza, M. Whiteley, A *Pseudomonas aeruginosa* antimicrobial affects the biogeography but not fitness of *Staphylococcus aureus* during coculture. *mBio* **12**, e00047-21 (2021).
39. C. B. Ibberson, J. P. Barraza, A. L. Holmes, P. Cao, M. Whiteley, Precise spatial structure impacts antimicrobial susceptibility of *S. aureus* in polymicrobial wound infections. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2212340119 (2022).
40. J. Baishya, J. A. Everett, W. J. Chazin, K. P. Rumbaugh, C. A. Wakeman, The innate immune protein calprotectin interacts with and encases biofilm communities of *Pseudomonas aeruginosa* and *Staphylococcus aureus*. *Front. Cell Infect. Microbiol.* **12**, 898796 (2022).
41. A. Stacy *et al.*, Bacterial fight-and-flight responses enhance virulence in a polymicrobial infection. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7819–7824 (2014).
42. T. Bjarnsholt *et al.*, *Pseudomonas aeruginosa* biofilms in the respiratory tract of cystic fibrosis patients. *Pediatr. Pulmonol.* **44**, 547–558 (2009).
43. L. Jackson *et al.*, Visualization of *Pseudomonas aeruginosa* within the sputum of cystic fibrosis patients. *J. Vis. Exp.* **161**, e61631 (2020), 10.3791/61631.
44. W. H. DePas *et al.*, Exposing the three-dimensional biogeography and metabolic states of pathogens in cystic fibrosis sputum via hydrogel embedding, clearing, and rRNA labeling. *mBio* **7**, e00796-16 (2016).
45. M. Kolpen *et al.*, Bacterial biofilms predominate in both acute and chronic human lung infections. *Thorax* **77**, 1015–1022 (2022), 10.1136/thoraxjnl-2021-217576.
46. M. J. Hunter, W. J. Chazin, High level expression and dimer characterization of the S100 EF-hand proteins, migration inhibitory factor-related proteins 8 and 14. *J. Biol. Chem.* **273**, 12427–12435 (1998).
47. S. S. Socransky, J. L. Dzink, C. M. Smith, Chemically defined medium for oral microorganisms. *J. Clin. Microbiol.* **22**, 303–305 (1985).
48. A. J. Page *et al.*, Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
49. W. Li *et al.*, RefSeq: Expanding the prokaryotic genome annotation pipeline reach with protein family model curation. *Nucleic Acids Res.* **49**, D1020–D1028 (2021).
50. S. Andrews, FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010). Accessed 16 November 2021.
51. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 3 (2011).
52. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
53. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
54. P. Ewels, M. Magnusson, S. Lundin, M. Kaller, MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
55. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
56. H. J. Motulsky, R. E. Brown, Detecting outliers when fitting data with nonlinear regression—A new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7**, 123 (2006).
57. D. Santesmasses, ggsunburst: Adjacency diagrams with ggplot2 (Version 0.3.0, 2020).
58. F. Crameri, Scientific colour maps (Version 1.3.0, Zenodo, 2018).