E.PathDash, pathway activation analysis of publicly available pathogen gene expression data

Abstract

E.PathDash facilitates re-analysis of gene expression data from pathogens clinically relevant to chronic respiratory diseases, including a total of 48 studies, 548 samples, and 404 unique treatment comparisons. The application enables users to assess broad biological stress responses at the KEGG pathway or Gene Ontology level and also provides data for individual genes. E.PathDash reduces the time required to gain access to data from multiple hours per dataset to seconds. Users can download high quality images such as volcano plots and boxplots, differential gene expression results and raw count data, making it fully interoperable with other tools. Importantly, users can rapidly toggle between experimental comparisons and different studies of the same phenomenon, enabling them to judge the extent to which observed responses are reproducible. As a proof of principle, we invited two cystic fibrosis scientists to use the application to explore scientific questions relevant to their specific research areas. Reassuringly, pathway activation analysis recapitulated results reported in original publications, but it also yielded new insights into pathogen responses to changes in their environments, validating the utility of the application. All software and data are freely accessible and the application is available at scangeo.dartmouth.edu/EPathDash.

Importance

Chronic respiratory illnesses impose a high disease burden on our communities and people with respiratory diseases are susceptible to robust bacterial infections from pathogens, including *Pseudomonas aeruginosa* and *Staphylococcus aureus*, that contribute to morbidity and mortality. Public gene expression datasets generated from these and other pathogens are abundantly available and an important resource for synthesizing existing pathogenic research, leading to interventions that improve patient outcomes. However, it can take many hours or weeks to render publicly available datasets usable; significant time and skills are needed to clean, standardize, and apply reproducible and robust bioinformatic pipelines to the data. Through collaboration with two microbiologists we have shown that E.PathDash addresses this problem, enabling them to elucidate pathogen responses to a variety of over 400 experimental conditions and generate mechanistic hypotheses for cell-level behavior in response to disease-relevant exposures, all in a fraction of the time.

Authors & Affiliations

Lily Taub[1]
Thomas H. Hampton[1]
Sharanya Sarkar[1]

Georgia Doing[2]
Samuel L. Neff[3]
Carson E. Finger[1]
Kiyoshi Ferreira Fukutani[1]
Bruce A. Stanton[1]

1. Geisel School of Medicine at Dartmouth
2. The Jackson Laboratory for Genomic Medicine
3. Lewis-Sigler Institute for Integrative Genomics, Princeton University

Introduction

In a review of the 2017 Global Burden of Disease report conducted by the Institute for Health Metrics and Evaluation, Soriano et al. found that the prevalence of chronic respiratory diseases worldwide increased by 39.8%[1] and recent research has identified an association between multiple chronic respiratory diseases and changes in the airway microbiome[2]. Therefore, in developing treatments for chronic respiratory diseases it is important to understand pathogenic behavior in response to experimental conditions such as growth medium, drug treatment, or genetic mutation[3–5]. Although there have been thousands of studies in this area of research, only a small percentage of the archived data is readily accessible for re-analysis. The goal of the application described here is to facilitate both gene and pathway-level re-analysis of publicly available gene expression data from pathogens clinically relevant to a variety of respiratory diseases, including *Pseudomonas aeruginosa*, *Bacteroides thetaiotaomicron*, *Staphylococcus aureus*, and *Streptococcus sanguinis*[6–15].

The rise in transcriptomic data production began with the invention of the microarray in 1995[16] and has continued through the past three decades, with data accumulating ever more rapidly after the development of sequencing methodologies that utilize the computational and hardware advances of modern computing[17]. Public repositories such as the European Bioinformatics Institute's ArrayExpress[18] and the National Center for Biotechnology Information's Gene Expression Omnibus (GEO)[19], have come online in concert with the rise of high-throughput sequencing data. As the infrastructure for data collection and hosting has become more robust, the data science and research communities have adopted the mission of developing FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) data practices[20]. While public repositories further the goals of findability and accessibility, the microbiology research community is without tools for the reuse of RNA-sequencing (RNA-seq) data that do not require computationally expensive and labor-intensive data cleaning, formatting, and statistical expertise.

The goal of the application described in this paper is to facilitate the reuse of publicly available data. Our application reduces the time required to gain access to data in 48 studies, 548 samples, and 404 unique treatment comparisons from multiple hours per dataset to seconds. The application primarily serves the microbiologist research community interested in pathogens associated with chronic respiratory diseases. In this paper we chose cystic fibrosis (CF) as a case study to demonstrate the value of the application to researchers. People with CF (pwCF)

are subject to chronic lung infections, which are responsible for 90% of the disease's morbidity and mortality[21–24]. Two pathogens that commonly dominate the lungs of pwCF, *Pseudomonas aeruginosa* (*P. aeruginosa)* and *Staphylococcus aureus* (*S. aureus*), are included in over 80% of the datasets in the application presented here. *S. aureus* is a dominant pathogen in lung infection in children with CF; the 2022 CF Patient Registry reports 60-80% prevalence of *S. aureus* infection in age cohorts younger than 18 years old[25]. Research has shown a strong association between *S. aureus* infection and poor clinical outcomes like decreased lung function and increased inflammation[26,27]. *P. aeruginosa,* dominant in late stage lung infections, has been shown to be multi-drug resistant, and is adept at forming biofilms that inhibit the immune response to bacterial infection[28].

The role of *S. aureus* and *P. aeruginosa* extend beyond CF. As members of the ESKAPE pathogen group (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, *Enterobacter* sp.) they are classified as highly virulent organisms that are adept at developing antibiotic resistance[29,30]. There is a demonstrated association between *S. aureus* and persistent asthma[14] and methicillin-resistant *S. aureus* (MRSA) is a common cause of hospital-acquired pneumonia infections[31]; in chronic obstructive pulmonary disease (COPD) *P. aeruginosa* infection has been shown to increase the risk of exacerbation and hospitalization events[32].

E.PathDash (ESKAPE Act PLUS Pathway Dashboard) streamlines the re-analysis of publicly available RNA-seq data relevant to respiratory diseases by enabling users to run pathway activation analysis for all possible treatment comparisons within each dataset in its compendium. This analytical approach uses transcriptomic data to identify differentially induced biological pathways across treatment comparisons. In doing this the program contributes to the body of tools that promote FAIR data principles and allows users to derive biological insights and formulate new hypotheses for future experiments. It achieves this without requiring investigators to be familiar with statistical methods for pathway analysis, or forcing them to normalize raw gene counts, establish consistent gene identifier encodings, or perform differential gene expression analysis. E.PathDash dramatically reduces the time required to access and analyze the data included in its compendium from many hours per dataset to seconds.

E.PathDash can be accessed freely at scangeo.dartmouth.edu/EPathDash.

Addressing the Limitations of Existing Applications

A number of bioinformatic tools that perform pathway analysis have been developed in response to the increased use of omics data in biological research[33] . Most of the applications require user-supplied data and therefore do not readily promote the reuse of publicly available datasets[34–37]. An exception to note is iDEP (integrated Differential Expression and Pathway analysis), but the datasets included in iDEP do not include data from bacterial species of interest to respiratory disease researchers[37].

We have published several applications that facilitate computational tasks for microbiologists and CF researchers[36,38–40]. Two of these applications, ESKAPE Act PLUS (Activation Analysis for ESKAPE Pathogens and other Prokaryotes Labs Usually Study)[36] and CF-Seq[38], inspired the development of the application described here. ESKAPE Act PLUS performs pathway activation analysis on user-uploaded differential gene expression data. CF-Seq is a platform for exploring gene-level analyses of public experimental data from clinically relevant CF pathogens, utilizing RNA-seq datasets from the GEO. By combining the statistical backend of ESKAPE Act PLUS and the processed RNA-seq data from CF-Seq, E.PathDash gives users a platform to explore pathway activation across a compendium of RNA-seq datasets without having to perform any data cleaning, formatting, or analysis tasks themselves, saving a considerable amount of time even if they have the skills to perform the steps required to access the data.

Results

E.PathDash is a R Shiny web application that facilitates the re-analysis of publicly available pathogen gene expression data relevant to respiratory diseases and includes a total of 48 studies, 548 samples, and 404 unique treatment comparisons. Specifically, the application automates pathway activation analysis of its compendium of RNA-seq datasets sourced from the GEO. Figure 1 provides an overview of the application flow, capturing the decisions users make as they navigate the application (orange branch nodes) and the analysis products that are generated (purple leaf nodes).

Workflow

After launching the application in an internet browser and reading an application overview (Figure 2, screen 1), the user begins interacting with the compendium data by filtering the cataloged datasets on bacterial species and strain(s) of interest (Figure 2, screen 2). From here the user can navigate between four different dashboard pages (Figure 2, screens 3-5). Individual screenshots of the dashboard pages are included in the supplementary material (Figures S1-S9).

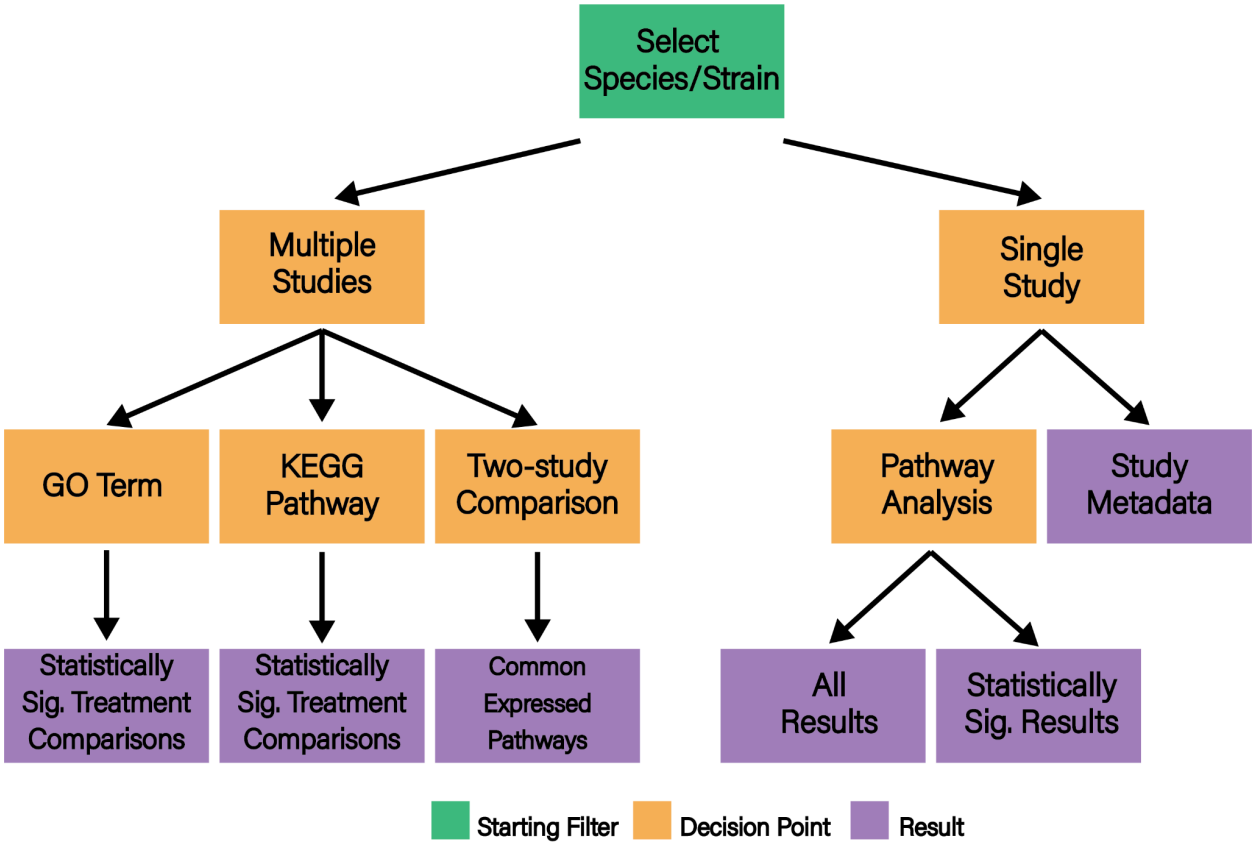


Figure 1. Conceptual flowchart for E.PathDash. The root (green) represents the bacterial species/strain filter the user must set to start exploring the data. Orange nodes represent decision points the user makes while navigating the application and the purple leaves are results the different decision paths generate.

Study Explorer

Selecting a single dataset (labeled by GEO accession number) from a list filtered by bacterial species and strain (Figure 2, label C) activates the **Study Explorer** page (Figure 2, screen 3). This page contains metadata about the dataset (a link to the GEO entry, study title, study description), downloadable raw count data and study design matrix, and four pathway activation analysis components controlled by a treatment comparison drop down (Figure 2, screen 3). Significantly activated or repressed KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways[41,42] and GO (Gene Ontology) terms[43,44] are shown in two boxplots, which use the distribution of $\log_2$ fold change (logFC) values for genes within the pathway to capture how the pathway is activated or repressed (Figure 2, label E). Two tables, one for KEGG pathways and one for GO terms, contain additional statistical information regarding the pathway activation analysis: binomial test statistic (which represents the proportion of activated pathway genes, i.e., those with a positive logFC value), median gene logFC, p-value, and FDR corrected p-value (Figure 2, label F). Changing the treatment comparison (Figure 2, label D) updates the data in the plots and tables to show pathway activation analysis for the specified comparison. For example, treatment 1 vs. treatment 2. Treatment 2 is the reference and the results can be interpreted as follows: pathways are activated or repressed in treatment 1 compared to treatment 2.

KEGG Pathway Explorer and GO Term Explorer

Users can search the datasets based on a KEGG pathway of interest in the **KEGG Pathway Explorer** page. The bacterial species and strain filters control the pathway options available; users select from pathways that were significantly activated or repressed in any of the treatment comparisons captured in the filtered compendium data. Selecting a pathway returns a table that lists all datasets for which that pathway was significantly activated or repressed (Figure 2, label G). In addition to the GEO accession number, the table contains the specific treatment comparison that showed activation or repression and the median logFC value for the genes within the pathway. Selecting a row in the results table renders a volcano plot of the genes within the pathway of interest. The volcano plot compares a gene's logFC value on the x-axis to the negative $\log_{10}$ transformed p-value on the y-axis. The p-value transformation plots statistically significant genes at higher y-axis values. This functionality is duplicated in the **GO Term Explorer** page, where users can search the data by GO term.

Study Comparison

The **Study Comparison** page allows users to compare pathway activation analyses for two selected datasets (Figure 2, screen 5). The page contains a table of KEGG pathways and GO terms that were significantly activated or repressed in both study datasets (Figure 2, label J). Selecting an entry in the table shows how the KEGG pathway or GO term was expressed across the different treatment comparisons within each dataset (Figure 2, label K), utilizing the bar plot described in the Application Design section of the Methods.

Downloadable Content

Downloadable content is an important feature of the application because it provides full interoperability with other tools. Users can download all generated graphs, tables, raw gene count data, and differential gene expression results for each treatment comparison within each dataset. Furthermore, users can download the differential gene expression data (by study and sample comparison) for just the genes within a KEGG pathway or GO term of interest. A complete list of downloadable content is included in the supplementary material (Table S1). Users can leverage this clean and well-formatted data to jump start additional analyses they wish to perform in other bioinformatic pipelines, eliminating the significant pre-processing time necessary to make much publicly available transcriptomic data usable.

To demonstrate the capabilities of the application we gave it to two CF scientists, who used the application to investigate questions relevant to their research. They contributed user stories detailing how they used the application, their findings, and their overall experience with E.PathDash.

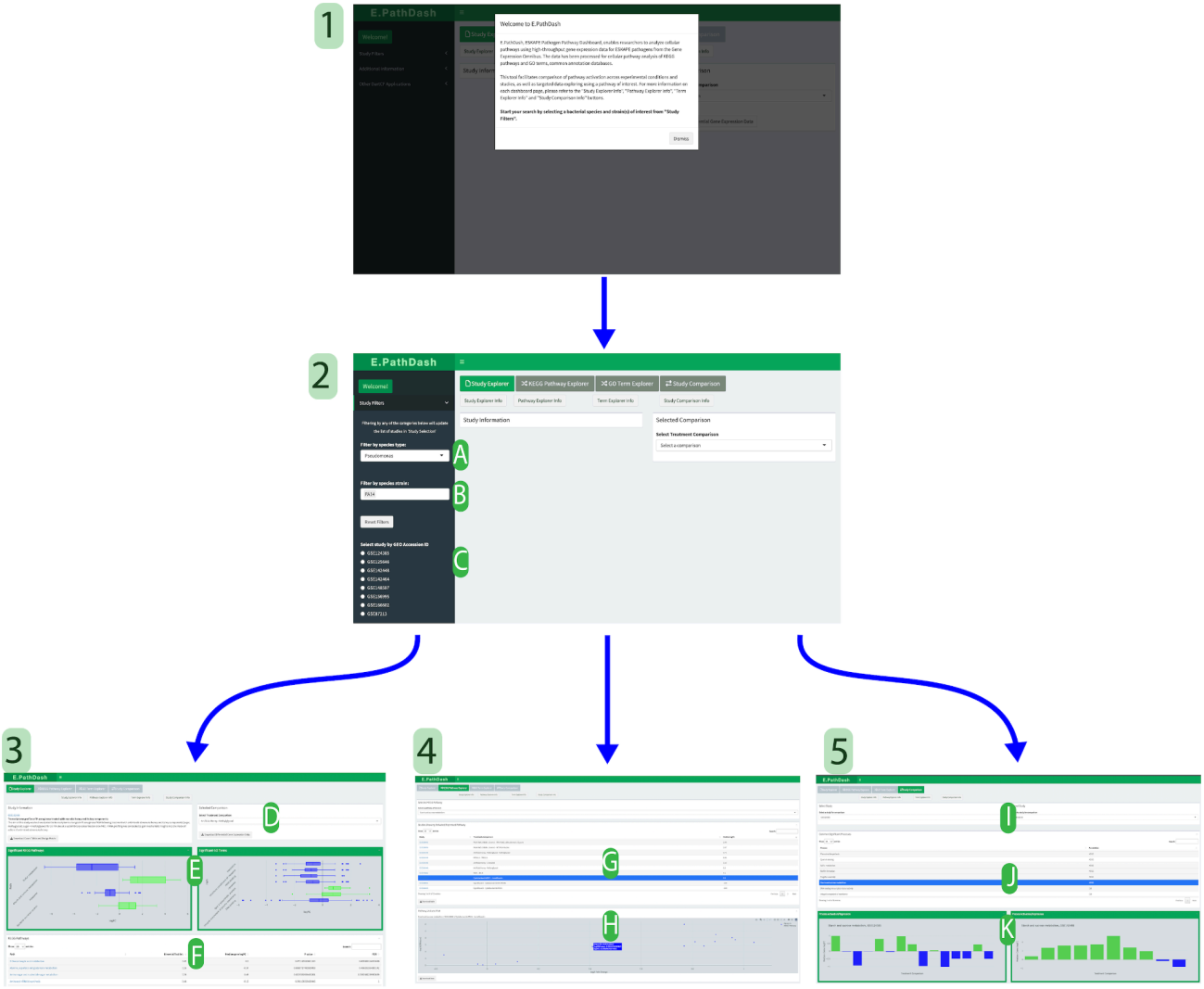
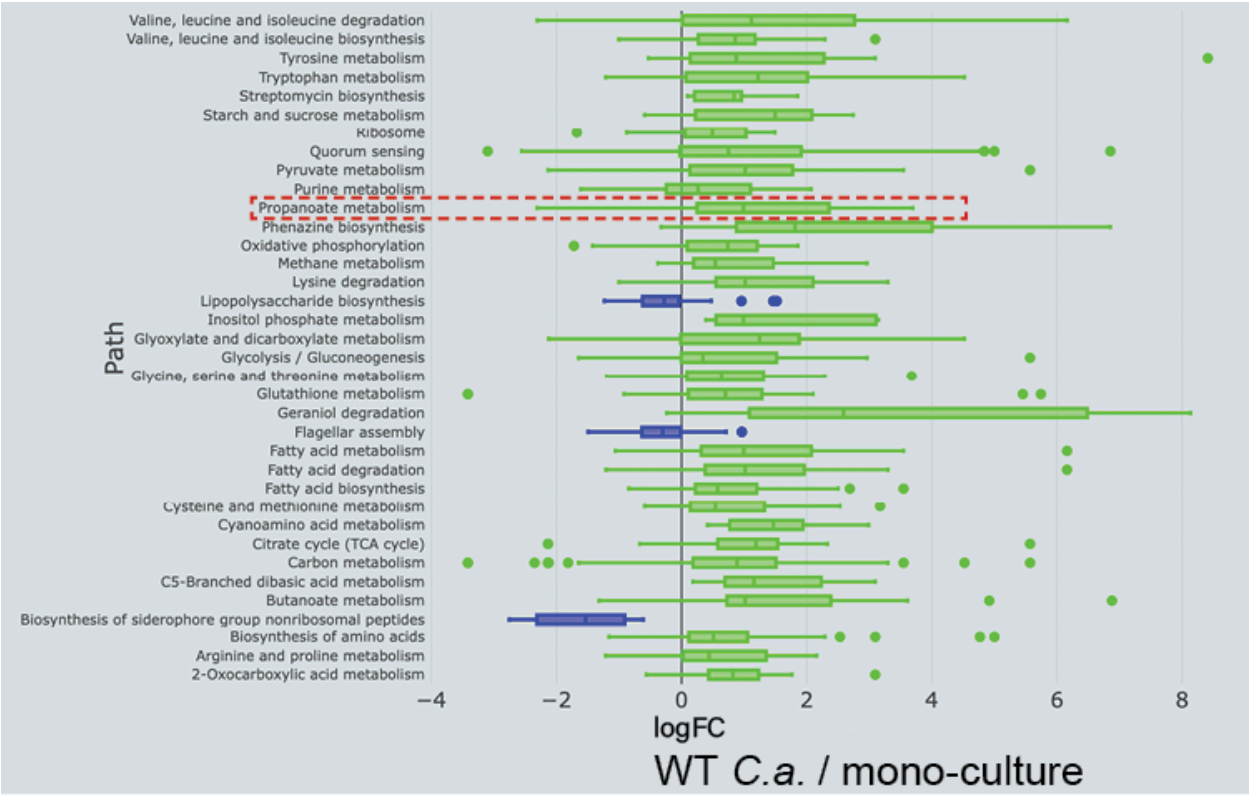


Figure 2. *Screen 1:* Landing page *Screen 2*: Filtering side panel where users select species **(a)** and strain **(b)**. Studies that meet filter requirements are shown in **(c)**. After filtering users can move to screen 3, 4 or 5 by selecting the corresponding button. *Screen 3:* The Study Explorer page shows information about a selected study. Selecting a treatment comparison **(d)** populates boxplots of gene expression in statistically significant activated or repressed KEGG pathways and GO terms **(e)** and a table of statistical information for all pathways/terms analyzed **(f)**. *Screen 4:* The KEGG Pathway Explorer and GO Term Explorer pages search the data by a pathway or term of interest and return a table of studies in which the pathway was significantly expressed **(g)**. Selecting a study shows a volcano plot for the genes within the pathway **(h)**. *Screen 5:* In the Study Comparison page users select two studies **(i)** to see commonly activated/repressed KEGG pathways and GO terms **(j)**. Selecting a pathway renders bar charts of median gene logFC value for each treatment comparison **(k)**. There is an enlarged version of this figure in the supplementary material (Figure S10).

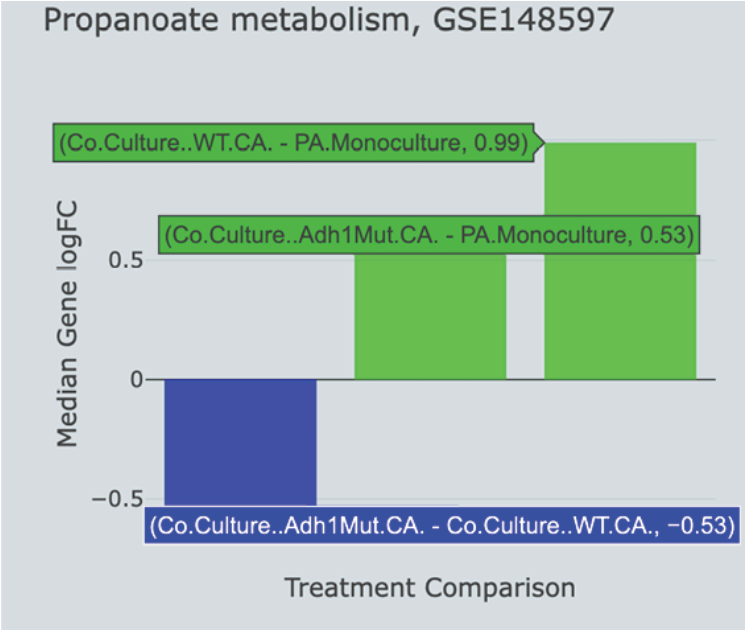
Case Study #1: Propanoate metabolism in *P. aeruginosa* during microbial interactions. Dr. Georgia Doing, The Jackson Laboratory for Genomic Medicine

To expand upon conclusions drawn from a previously published dataset of *P. aeruginosa* gene expression from *in vitro* co-culture with *Candida albicans* (GEO: GSE148597)[45], I used E.PathDash to explore a comparison between *P. aeruginosa* in co-culture and mono-culture (in the application the treatment comparison selected was “co-culture WT - PA Monoculture”). This comparison is complementary to those presented in the original publication and reveals exciting new hypotheses. Re-analyzing the data using the E.PathDash Study Explorer shows that the propanoate metabolism KEGG pathway (pau00640) had statistically significant activation (binomial test estimate 0.8, median gene fold-change 0.99 and FDR < 0.001). Specifically, *P. aeruginosa* genes for the metabolism of propanoate have higher mRNA levels when *P. aeruginosa* is in co-culture with *C. albicans* compared to when *P. aeruginosa* is in mono-culture (Figure 3a). This suggests *P. aeruginosa* may be metabolizing propanoate or inducing the 2-methylcitrate cycle[46].

a)



b)



c)

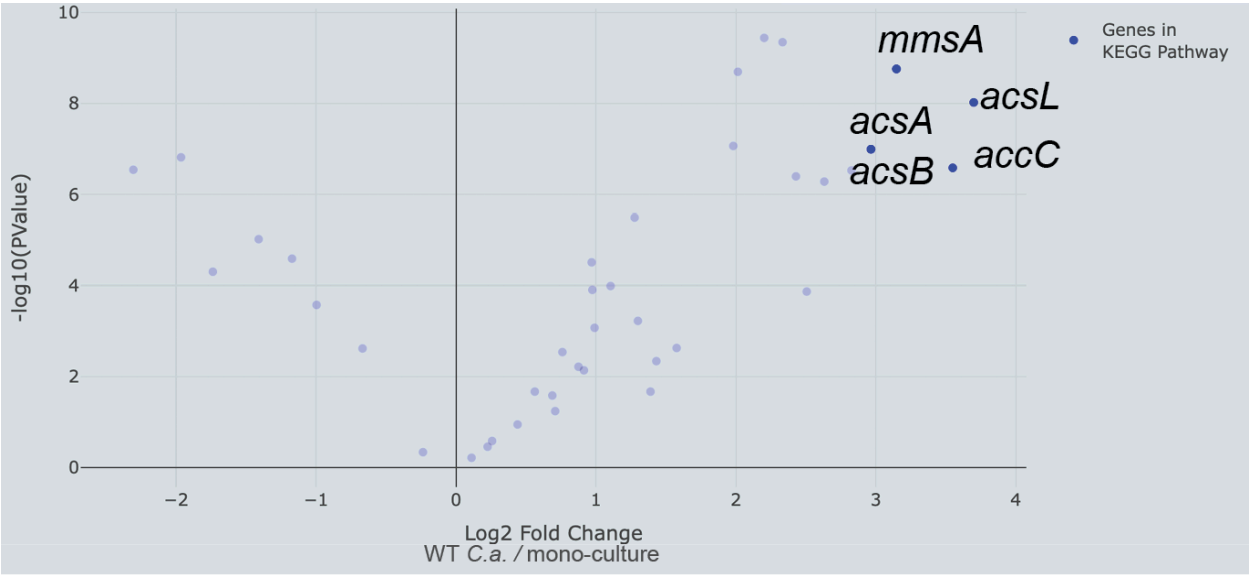


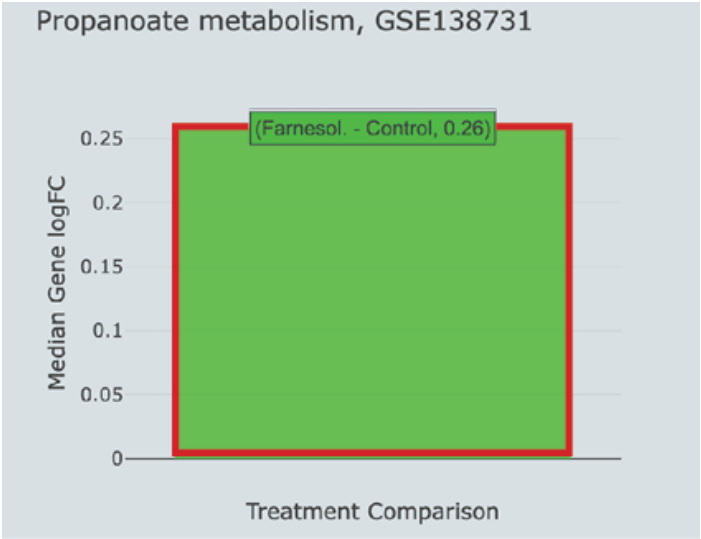
Figure 3. Use of E.PathDash to analyze the expression of propanoate metabolism genes in *P. aeruginosa*. **(a)** The Study Explorer of GSE148597 shows that the KEGG pathways significantly activated in *P. aeruginosa* co-cultured with *C. albicans* compared to mono-culture include propanoate metabolism. Box and whisker plots show log fold-change distributions of genes in each pathway. **(b)** The Study Comparison tool shows the other comparisons in the study and suggest *C. albicans ADH1*-dependent ethanol production plays a role in stimulating expression of the propanoate metabolism KEGG pathway. **(c)** Volcano plot from the Study Explorer shows individual genes in the propanoate metabolism KEGG pathway. Data are plotted as fold-change values and negative log transformed p-values.

Changing the treatment comparison in the Study Explorer showed that the KEGG propanoate metabolism pathway is also activated, though to a lesser extent, when *P. aeruginosa* is grown in co-culture with WT *C. albicans* compared to *P. aeruginosa* grown with the *adh1*Δ/Δ *C. albicans* mutant, which is deficient in ethanol production, (Figure 3b) (binomial test estimate 0.88, median log fold-change 0.53 and FDR < 0.0001). The trend toward increased expression of propanoate metabolism genes when *P. aeruginosa* is grown with WT *C. albicans* compared to *adh1*Δ/Δ *C. albicans*, as well as the lower magnitude fold-change when *P. aeruginosa* is grown with *adh1*Δ/Δ *C. albicans* compared to *P. aeruginosa* grown in mono-culture (binomial test estimate 0.63, median log fold-change 0.53 and FDR 0.24), suggests that *C. albicans* ethanol production may be contributing to the response of increased propanoate metabolism in *P. aeruginosa*. This is consistent with a report in *E. coli* that showed ethanol induced propanoate metabolism activated the expression of *prpD*, a gene also present in *P. aeruginosa*[47].

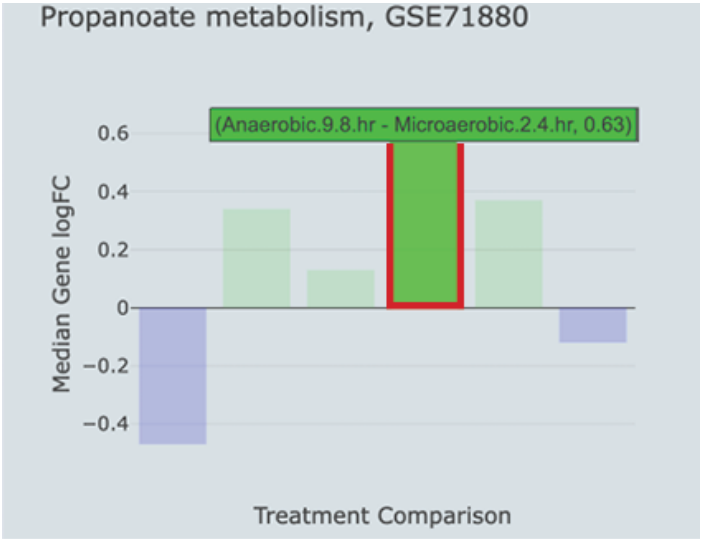
Downloading the log fold-change values of the genes in the propanoate metabolism KEGG pathway using the E.PathDash KEGG Pathway Explorer interface (Table S2) provided the nuanced information on the top five genes in the propanoate KEGG pathway. The highest log fold-change values in co-culture were *acsL*, *accC*, *mmsA*, *acsA*, and *acsB* (Figure 3c). Notably, *acsL*, *acsA* and *acsB* are acetyl-coA synthetases that can convert propanoate to propionyl-adenylate and then propanoyl-coA, but *acsA* and *acsB* can also convert acetate to acetyl-coA during ethanol oxidation[48]. This suggests that ethanol may induce ErdR responsive genes[49] which, in addition to regulation through *prpD*, may also facilitate the metabolism of propanoate.

The E.PathDash KEGG Pathway Explorer shows that propanoate metabolism is also activated in a study wherein *P. aeruginosa* was treated with farnesol (GEO: GSE138731, median gene log fold-change 0.26, Figure 4a), a *C. albicans*-produced quorum-sending molecule known to alter *P. aeruginosa* metabolism[50]. Furthermore, propanoate metabolism is also activated in studies in which *P. aeruginosa* is transitioned from anaerobic conditions to microaerobic conditions (GEO: GSE71880, median gene log fold-change 0.63, Figure 4b) and treated with artificial honey and methylglyoxal (GEO: GSE142448, median gene log fold-change -1.22, Figure 4c)[51]. The E.PathDash Study Comparison feature made the specific directionality of each comparison clear: propanoate was activated upon the addition of farnesol and micro-oxia and repressed upon the addition of methylglyoxal (Figure 4, bars highlighted in red). These conditions may mimic those of metabolic imbalance such that intermediates of glycolysis or the TCA cycle accumulate. Notably, methylglyoxal can be detoxified into intermediates that converge with those of propanoate metabolism[52] and the TCA intermediate succinate can stimulate propanoate metabolic genes[53]. While these connections are likely indirect, they provide preliminary evidence suggesting co-culture with *C. albicans* alters *P. aeruginosa* central metabolism in a manner that leads to altered pools of metabolic intermediates and perhaps a convergence on propanoate metabolism.

a)


Propanoate metabolism, GSE138731
(Farnesol. - Control, 0.26)
Median Gene logFC
0.25
0.2
0.15
0.1
0.05
0
Treatment Comparison

b)


Propanoate metabolism, GSE71880
(Anaerobic.9.8.hr - Microaerobic.2.4.hr, 0.63)
Median Gene logFC
0.6
0.4
0.2
0
−0.2
−0.4
Treatment Comparison

c)

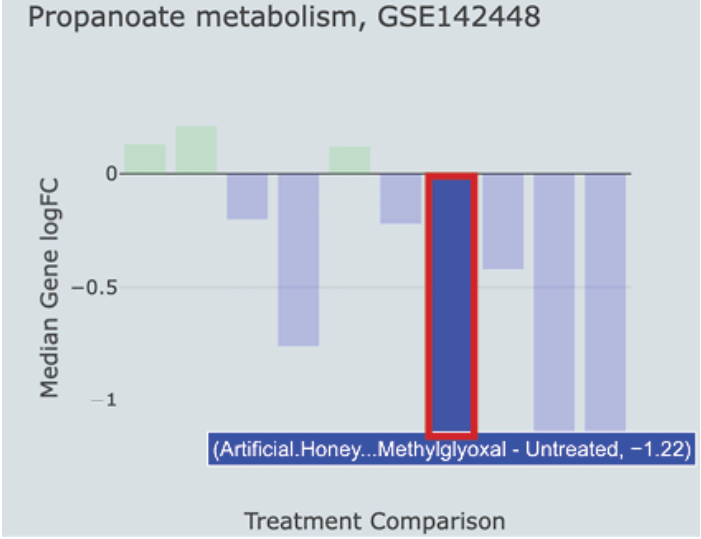

Propanoate metabolism, GSE142448
Median Gene logFC
0
−0.5
−1
(Artificial.Honey...Methylglyoxal - Untreated, −1.22)
Treatment Comparison

Figure 4. The KEGG Pathway explorer showed that propanoate metabolism genes are also increased in expression in *in vitro* studies of **(a)** *P. aeruginosa* treated with farnesol, **(b)** transitioned between anaerobic and micro-oxic conditions and **(c)** treated with artificial honey and methylglyoxal. Highlighted bars show relevant comparisons of each study.

Analysis of previously published datasets with E.PathDash suggests a hypothesis wherein *P. aeruginosa* propanoate metabolism may be stimulated by *C. albicans*-produced ethanol through (i) direct regulation of propanoate catabolic gene *prpD* by ethanol-sensing ErdR and through (ii) metabolic convergence of ethanol oxidation and propanoate catabolism by dual-function acetyl-coA synthetases *acsA* and *acsB* (Figure 5). This model could be readily tested with *erdR* mutations and assays of *acsA* and *acsB* expression as well as growth on ethanol and propanoate. Additionally, the broad activation of the propanoate KEGG pathway in other studies that looked at *in vitro* conditions relevant to co-culture suggest propanoate and ethanol metabolism may be further stimulated by (iii) other factors present in co-culture such as farnesol, oxygen limitation and metabolic intermediate methylglyoxal. The pathway-level interactions could be untangled with epistasis experiments on relevant environment-sensing genes such as oxygen-sensing *anr* and metabolic genes like methylglyoxal reductases. E.PathDash allowed for the rapid re-analysis of data in approximately 1/100th of the time (hours rather than weeks) it would take to analyze multiple studies manually. This tool facilitated hypothesis generation resulting in a testable model of the key role propanoate metabolism may play in medically relevant microbial interactions of *P. aeruginosa*.

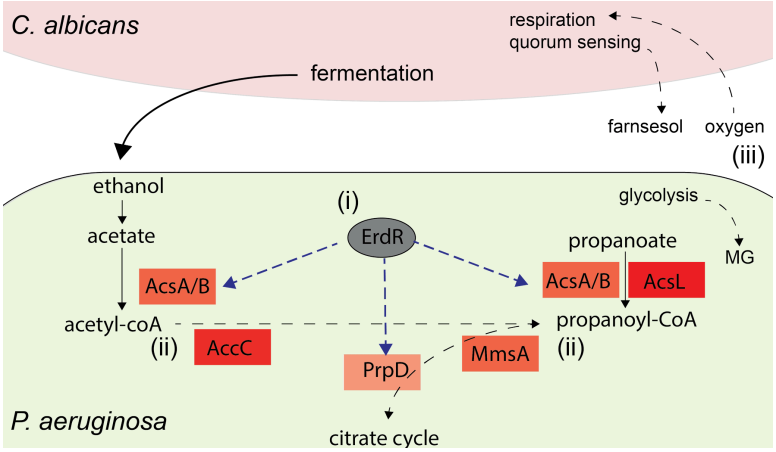


Figure 5. Hypothesized model synthesizing the results found using E.PathDash: ethanol may stimulate propanoate metabolism directly through ErdR regulation of *prpD* and indirectly through activity of AcsA and AcsB acting upon both acetate and propanoate. Furthermore, metabolic products of ethanol and propanoate catabolism may converge on propanoyl-CoA and feed into the citrate cycle.

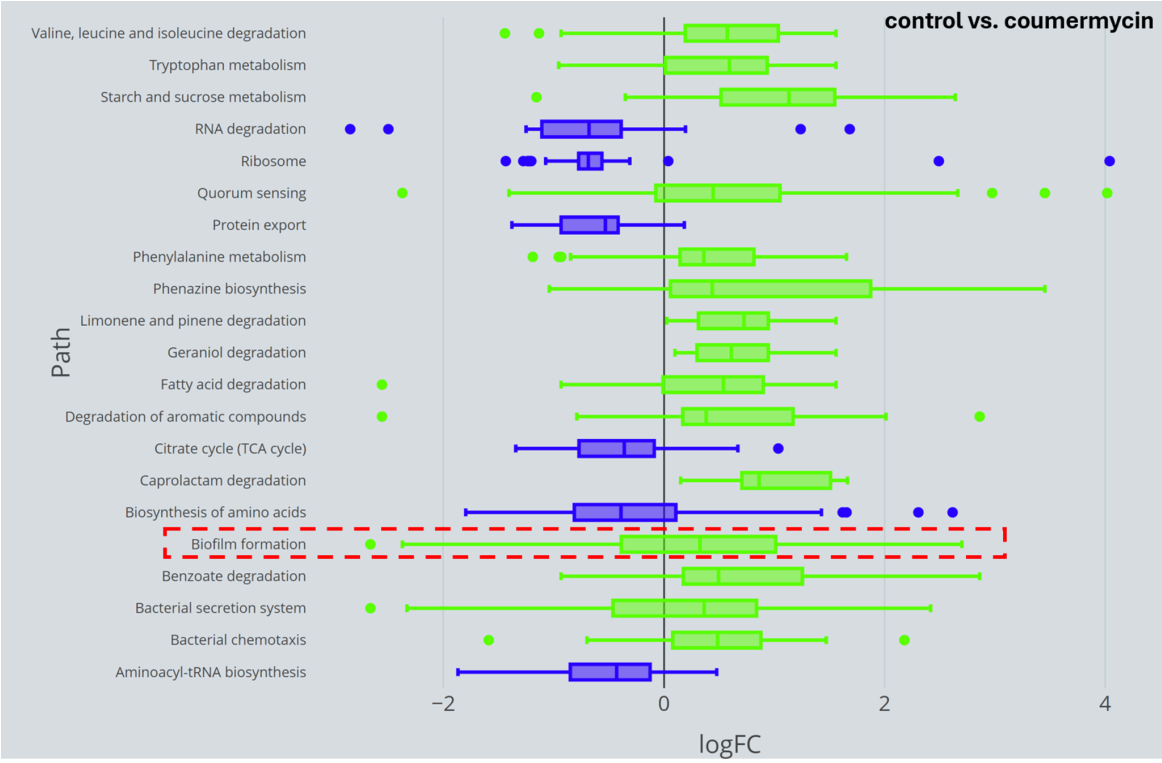
Case Study #2: Differential effects of DNA-gyrase inhibitor classes on biofilm formation in *P. aeruginosa*. Sharanya Sarkar, Ph.D. candidate, Geisel School of Medicine at Dartmouth

Cystic Fibrosis (CF) is a genetic disease that is associated with chronic pulmonary infections[54]. Although the infections are polymicrobial in nature, comprising several bacterial genera,

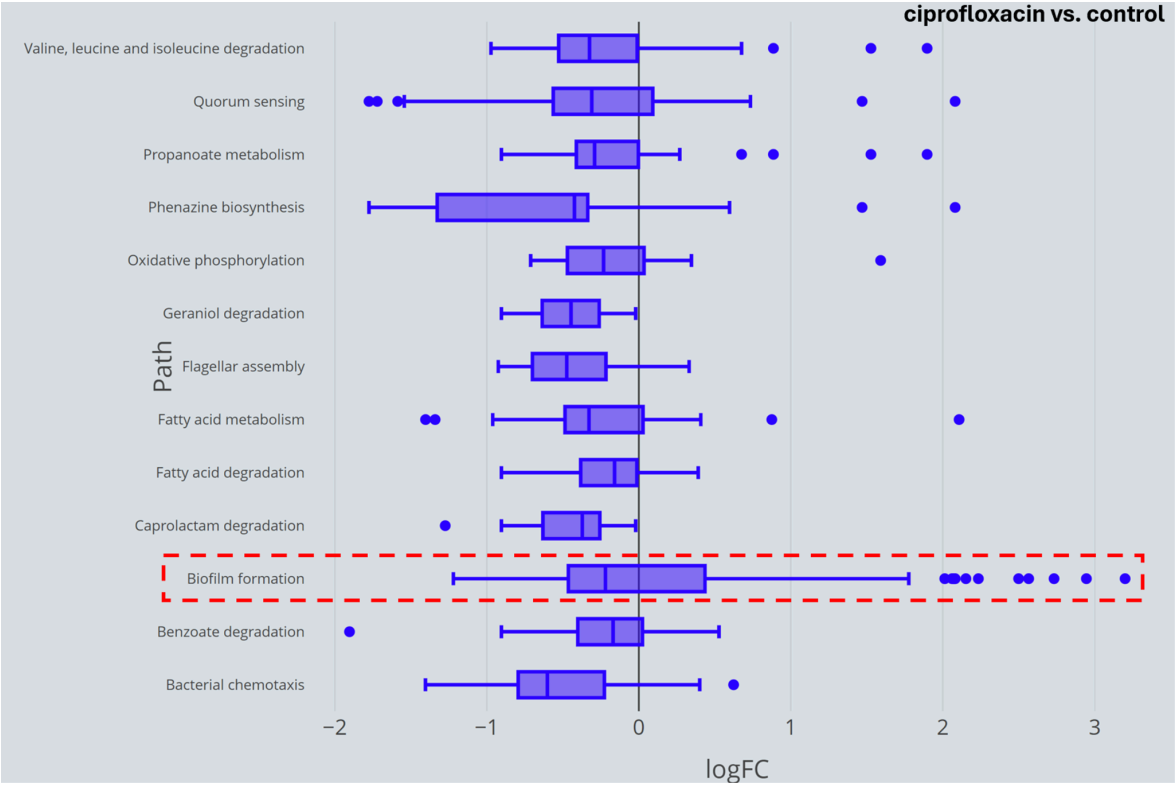
P. aeruginosa is the most common pathogen found in adult people with CF (pwCF)[25]. *P. aeruginosa* expresses distinct virulence factors (beta-lactamases, efflux pumps, etc.) via acquiring adaptive mutations over the course of chronic infection and becomes resistant to host factors and rigorous antimicrobial therapy[28]. Furthermore, *P. aeruginosa* forms robust, antibiotic-resistant biofilms within the respiratory tract, making it exceedingly challenging to eliminate once established[55,56]. To maintain control of the infections, pwCF are prescribed antibiotics as a part of their treatment regimen[57] and DNA gyrase inhibitors, particularly the fluoroquinolones, constitute a standard treatment for *P. aeruginosa* infections[58].

Our lab investigates antimicrobial treatments that interfere with biofilm formation in *P. aeruginosa*[59–61]. Gyrase inhibitors can inhibit biofilm formation[62,63], so we used E.PathDash to analyze how DNA gyrase inhibitors affected this process in experimental settings. Additionally, we wanted to determine if distinct classes of DNA-gyrase inhibitors impacted biofilm formation differentially. Utilizing the species filter in E.PathDash to search across study data for *P. aeruginosa* datasets, we investigated dataset GSE166602, which looked at molecular signatures in *P. aeruginosa* with different gyrase inhibitors[64]. Using different treatment comparisons in the Study Explorer page, we found that untreated *P. aeruginosa* had significantly upregulated biofilm formation compared to *P. aeruginosa* treated with coumermycin, with a median gene logFC value of 0.33 (Figure 6a, FDR 0.015). In other words, the median expression of biofilm genes in the coumermycin condition was 0.79 times that observed in the untreated condition. Subsequently, we aimed to compare the inhibitory effects of ciprofloxacin, an antibiotic belonging to the fluoroquinolone group (the second class of gyrase inhibitors), with a control condition (untreated). The results showed that ciprofloxacin repressed biofilm formation in *P. aeruginosa* compared to untreated *P. aeruginosa*, with a median gene logFC of -0.22 (Figure 6b, FDR 0.042). In this comparison, gene expression within the biofilm formation pathway was 0.85 times that of the control condition. Considering these two comparisons we hypothesized that *P. aeruginosa* treated with ciprofloxacin will have greater biofilm formation compared to those treated with coumermycin. As anticipated, when comparing ciprofloxacin-treated *P. aeruginosa* to coumermycin-treated *P. aeruginosa,* biofilm formation was upregulated in the former group (Figure 6c, FDR 0.001). This indicates that ciprofloxacin does not substantially reduce biofilm formation, unlike coumermycin.

a)


control vs. coumermycin
Valine, leucine and isoleucine degradation
Tryptophan metabolism
Starch and sucrose metabolism
RNA degradation
Ribosome
Quorum sensing
Protein export
Phenylalanine metabolism
Phenazine biosynthesis
Limonene and pinene degradation
Geraniol degradation
Fatty acid degradation
Degradation of aromatic compounds
Citrate cycle (TCA cycle)
Caprolactam degradation
Biosynthesis of amino acids
Biofilm formation
Benzoate degradation
Bacterial secretion system
Bacterial chemotaxis
Aminoacyl-tRNA biosynthesis
Path
−2
0
2
4
logFC

b)


ciprofloxacin vs. control
Valine, leucine and isoleucine degradation
Quorum sensing
Propanoate metabolism
Phenazine biosynthesis
Oxidative phosphorylation
Geraniol degradation
Flagellar assembly
Fatty acid metabolism
Fatty acid degradation
Caprolactam degradation
Biofilm formation
Benzoate degradation
Bacterial chemotaxis
Path
−2
−1
0
1
2
3
logFC

c)

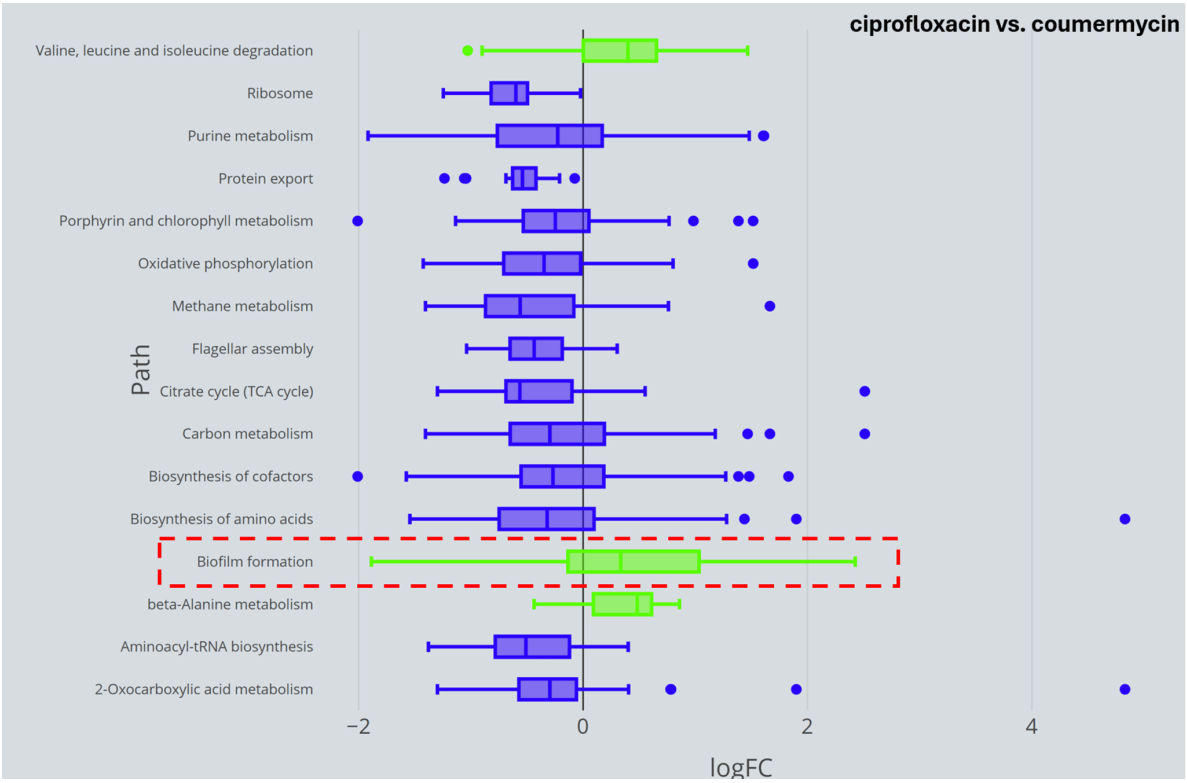


Figure 6 (a) Coumermycin downregulates biofilm formation in *P. aeruginosa* compared to control, with median expression of biofilm genes 0.79 times that of control **(b)** Ciprofloxacin represses biofilm formation in *P. aeruginosa* as compared to control with median expression of biofilm genes 0.85 times that of control **(c)** Coumermycin is more efficient than ciprofloxacin in reducing biofilm formation in *P. aeruginosa*

Next we explored responses of genes in the biofilm formation KEGG pathway using the KEGG Pathway Explorer page. Our specific focus was on examining significantly upregulated biofilm-related genes in control *P. aeruginosa* exhibiting a minimum log2 fold change of 0.5, as compared to coumermycin. One particular gene (UniProt: A0A0H2ZEE6, gene name: *rhlI*) met these criteria (Figure 7a) and was chosen for further scrutiny. A brief search using the UniProt ID revealed that this gene encodes an Acyl-homoserine-lactone synthase. Previous research has demonstrated that biofilm formation in *P. aeruginosa* can be influenced by targeting Acyl-homoserine-lactones (AHLs)[65]. Taken together, the analysis suggests that one of the mechanisms through which coumermycin significantly downregulates biofilm formation in *P. aeruginosa* is by targeting AHL synthase. Given that coumermycin demonstrated superior efficacy in preventing biofilm formation compared to ciprofloxacin, we postulate that the central biofilm gene, AHL synthase, might be more highly expressed in the ciprofloxacin group than in the coumermycin group. We referred to the downloaded gene expression data for the ciprofloxacin-coumermycin comparison from the KEGG Pathway Explorer page and easily located the log2 fold change value for this gene. It was evident that this gene was significantly upregulated in the ciprofloxacin group (Figure 7b), implying that ciprofloxacin may not be as proficient as coumermycin in targeting AHL synthase.

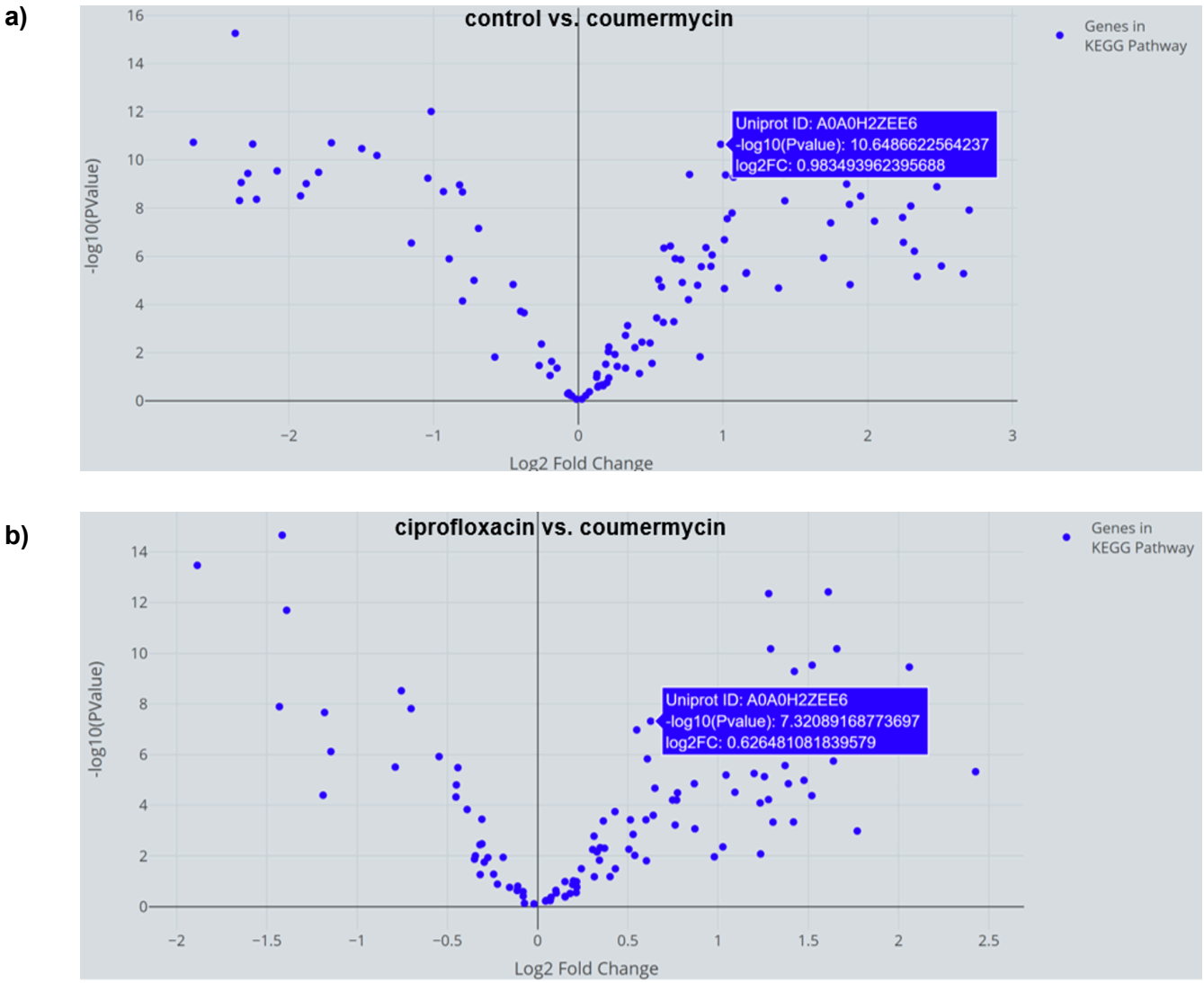


Figure 7 AHL synthase, a key biofilm gene in *P. aeruginosa* is significantly upregulated in **(a)** control, as compared to coumermycin-treated *P. aeruginosa,* and in **(b)** ciprofloxacin-treated *P. aeruginosa* as compared to coumermycin-treated *P. aeruginosa*

In summary, the application provided a highly visual and efficient means to re-evaluate a public dataset. It indicated that coumermycin could potentially be a more effective option for treating biofilm-forming microorganisms such as *P. aeruginosa*, in comparison to ciprofloxacin, and this can be tested experimentally. Our re-analysis of the data extends the findings of the original publication by focusing on quorum-sensing mediators.

This case study underscores an important lesson in understanding how antibiotics, even when binding to the same target in the pathogen, but belonging to distinct structural classes, can have varying effects on pathogenic processes. Our research team has demonstrated that when combined with antibiotics, let-7b-5p exhibits additive effects in suppressing biofilm formation in *P. aeruginosa*[59]. The findings from this analysis, facilitated by E.PathDash, guide our research group in conducting preliminary studies on *P. aeruginosa* to assess the efficacy of various

antibiotics. The evaluation will help us identify the most potent antibiotic to maximize this additive effect of let-7b and antibiotics.

Discussion

E.PathDash facilitates re-analysis of gene expression data from pathogens clinically relevant to respiratory diseases, including a total of 48 studies, 548 samples, and 404 unique treatment comparisons. E.PathDash has several key features that make it a valuable resource for hypothesis generation that also enhance data reusability and generate reproducible analyses. First, E.PathDash provides multiple entry points into the datasets within its compendium; the user can drill down into a single dataset related to a bacterial species of interest, search for activated pathways across datasets, or compare pathway expression between datasets and treatments. The case studies highlighted in this paper demonstrate how this allows researchers to follow an investigative path through multiple questions, all within a single interface, and identify research datasets that show activation of a pathway of interest. Second, E.PathDash makes the cleaned raw data and analysis results available for download. This promotes transcriptomic analyses beyond the capabilities built into the application and cuts down on the time investigators need to spend preparing publicly available data for reuse (from many hours to seconds). Third, the application links gene-level and pathway-level analyses, which allows researchers to quickly extend the conclusions of pathway activation to expression data at the individual gene level. In both case studies E.PathDash was used to identify how genes within a significantly activated pathway were expressed, which led to mechanistic hypotheses regarding pathway activation. Therefore, insights facilitated by E.PathDash can help generate ideas for future experiments.

E.PathDash was developed as a tool that expands upon ESKAPE Act PLUS and CF-Seq, other applications developed by our group that facilitate re-analysis of omics data. Given this, the application was restricted to pathogens included in the analytical pipeline for ESKAPE Act PLUS. This decision restricted the number of RNA-seq datasets included in the application. An area of future work is to increase the number of datasets and compatible pathogens, which could be done by separating the analytical pipeline of E.PathDash from ESKAPE Act PLUS or by expanding the number of species and strains for which ESKAPE Act PLUS has pathway data.

In its implementation of pathway activation analysis, ESKAPE Act PLUS assumes that under the null hypothesis the split of up and down regulated genes is 50%[36]. This assumption ignores inherent differences between expression levels across RNA-seq datasets and could be made more sophisticated by using the overall rate of induction within the dataset as the null hypothesis ratio. Subsequently, the magnitude of deviation between the overall induction rate and the induction rate among genes within a pathway could determine significant pathway activation or repression.

It is important to acknowledge that E.PathDash and other methods to interrogate archived data are hypothesis generating and should be confirmed by wet bench experiments. For example,

we have published several studies where we mined publicly available data, generated new hypotheses, and confirmed the hypotheses by laboratory experiments[3,66–68].

Users of E.PathDash should also note that datasets in the compendium were not generated from experiments that shared wet bench or computational protocols in their data collection pipeline. Thus, any direct comparisons between datasets should consider the underlying study design differences that could have contributed to the analysis results.

Methods

Data Collection and Cleaning

Figure 8 shows an overview of the data collection, cleaning, and analysis pipelines. Data collection for E.PathDash started with a compendium of RNA-seq datasets from the GEO that were originally compiled for the development of CF-Seq. This included 128 datasets covering 13 different pathogens. The datasets used for CF-Seq were restricted to clinically relevant CF pathogens and those for which investigators uploaded tables of raw gene counts in formats compatible with R's file functions and the differential expression analysis package edgeR. Furthermore, the study information had to define sample groups for the purpose of differential gene expression analysis. Differential gene expression analysis results, conducted within the CF-Seq analysis pipeline, were extracted for use in E.PathDash.

The original 128 studies and the corresponding differential gene expression data were stored in a R data file. R scripts run on R version 4.2.1[69] filtered the datasets to those pathogens compatible with the ESKAPE Act PLUS application[36], which was leveraged to conduct pathway activation analysis. Those pathogens were *P. aeruginosa* (strains PA14 and PAO1), *Bacteroides thetaiotaomicron* (strain VPI-5482), *Staphylococcus aureus* (strains Newman and USA300), and *Streptococcus sanguinis* (strain SK36). These pathogens have been shown to play an important role in a variety of respiratory diseases[6–15] and be effected in the airway by pertinent environmental exposures like air pollution[70].

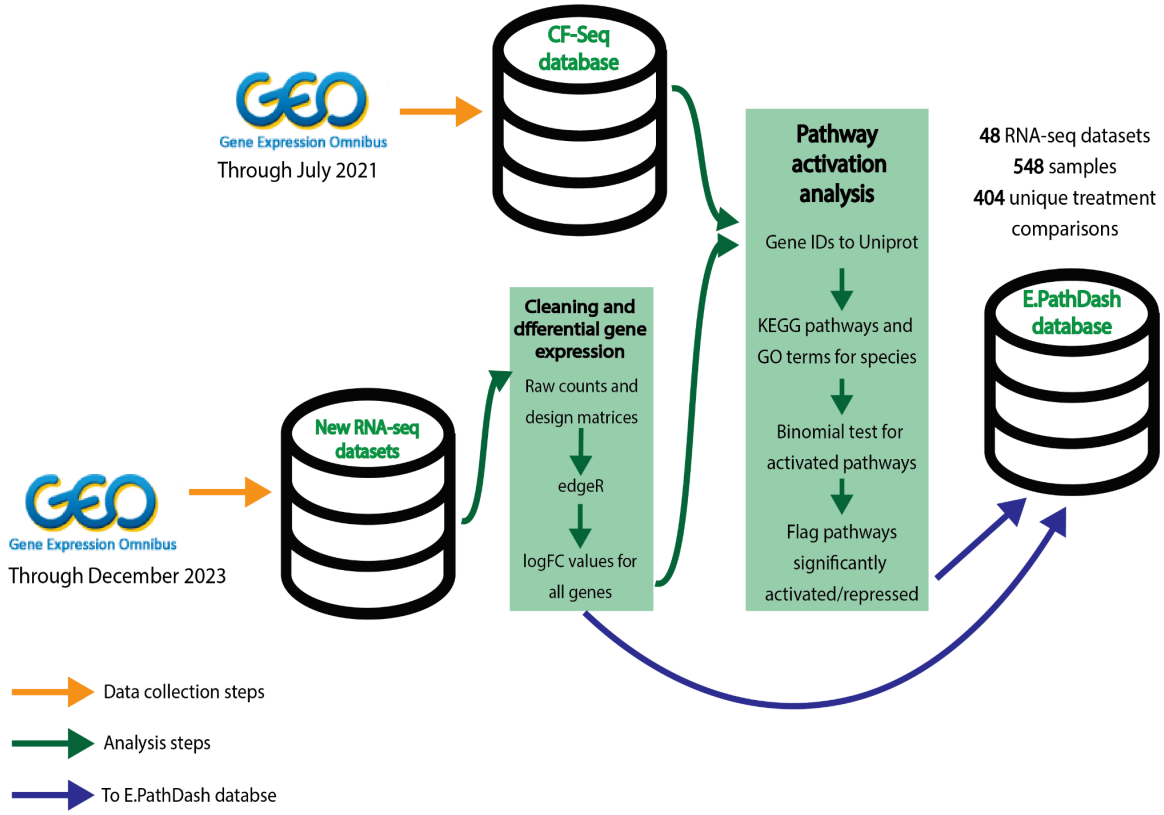


Figure 8. Data collection, cleaning, and analysis pipeline for the creation of the E.PathDash application database. All RNA-seq datasets originated from GEO. Data uploaded after July 2021 (not included in the CF-Seq application database) was run through the CF-Seq differential expression analysis pipeline. Subsequently, all datasets were run through the ESKAPE pathway activation analysis pipeline, and both differential analysis and pathway activation results were pulled into the E.PathDash application database.

The CF-Seq application database included RNA-seq datasets published on the GEO through July 2021. Datasets published on the GEO from August 2021 through December 2023 that met the same organism, file format, content, and sample group requirements outlined above were pulled into the database for E.PathDash. The scripts used to do this can be found in the Git repository associated with this publication [https://github.com/DartCF/EPathDash]. Differential gene expression analysis for the new datasets was conducted using the CF-Seq analysis scripts. The final compendium consisted of 48 studies, 548 samples, and 404 unique treatment comparisons across the four different bacterial species.

ESKAPE Act PLUS, used to conduct pathway activation analysis, uses the KEGGREST R package[71] from Bioconductor to map genes to KEGG pathways for the purpose of the analysis[36]. The KEGGREST mappings use UniProtKB (UniPort Knowledgebase) gene encodings, requiring all gene identifiers to be translated to UniProtKB . Metadata for each study

was reviewed to identify the gene encoding schema and the UniProt Consortium's web-based gene ID mapping tool was used to translate gene IDs to UniProtKB[72]. Gene IDs in their native encoding were extracted from the RNA-seq datasets and written to CSV files, which were uploaded to the UniProt ID mapping tool. Batches for each encoding schema were run separately because the translation tool does not dynamically detect the gene encoding schema and requires the input schema to be manually specified. Results consisted of the original gene identifier and the corresponding UniProt ID. This dictionary was used to translate gene IDs for pathway activation analysis. The scripts used to conduct the data pre-processing can be found in the Git repository cited previously.

Within the 48 studies included in the final compendium, 31 of the studies had multiple rows with the same UniProt identifier but different logFC values for differential expression. This inconsistency was a result of there being rows in the raw count data for an ordered locus name and gene name corresponding to the same protein. To resolve this we chose to keep only the information for the identifier associated with the smaller p-value for differential gene expression, optimizing the sensitivity of the downstream pathway activation analysis. In general p-values for these unique rows were similar.

After data collection and cleaning, the data structure used for pathway activation analysis consisted of differential gene expression results for each unique treatment comparison within each dataset. Each row of the dataframe consisted of UniProt gene identifier, logFC value, and adjusted p-value (corrected for multiple hypotheses using the FDR method).

Data Analysis

Pathway activation analysis was conducted using ESKAPE Act PLUS functionality. ESKAPE Act PLUS uses a binomial test to identify significantly activated or repressed cellular pathways between treatment groups. The binomial test uses logFC values calculated from differential expression analysis of RNA-seq data[36]. Under the null hypothesis there is an equal proportion of genes with positive and negative logFC values. ESKAPE Act Plus uses the binom.test function in R's stats package[69] to determine if the proportion of genes with positive logFC values within a given pathway differs significantly from 50%. If this difference has an FDR corrected p-value < 0.05 the pathway is labeled as significantly activated or repressed.

An advantage of this method of pathway analysis is that it accounts for all genes within a pathway, not just those that meet a threshold of differential expression between treatment groups. Therefore, information from genes on the edges of a threshold for statistical significance or logFC value is not lost. Over-representation analysis (ORA) for pathway enrichment, a common methodology that identifies functional gene sets that have coordinated expression beyond what is expected by chance, restricts the analyzed genes to those with statistically significant differential expression between samples[73].

Pathway activation analysis was conducted for KEGG pathways and GO terms. KEGG and GO are databases annotated by domain experts that map genes to functional groups[41–44]. Pathway activation analysis results for each treatment comparison within each dataset were stored in the

final database for E.PathDash. The stored results include 1) KEGG or GO path name 2) raw and FDR-corrected p-values 3) binomial test statistic and 4) median logFC value of genes in the pathway. The script to run pathway activation analysis can be found in the Git repository specified previously.

In addition to differential gene expression and pathway activation analyses, E.PathDash has data that links each KEGG pathway and GO term to their constituent genes, which was retrieved using the EnrichmentBrowser package[74]. Using this information, the E.PathDash interface links pathway activation analysis and differential gene expression analysis results.

The data collection, cleaning and analysis were performed offline in order to minimize the amount of processing the application needed to do in real time, facilitating a better user experience. The collection and analysis steps were split into the two R scripts mentioned previously, which were run in succession on the compiled datasets. The scripts utilize several R packages in addition to the KEGGREST and EnrichmentBrowser packages: 'stringr' for string manipulation, 'readxl' for file ingestion, and 'tidyverse' for data frame manipulation[75–77].

Application Design

All the application code for E.PathDash is contained in the app.R file, which has two components, "ui" and "server", that define a reactive web application. The code in the "ui" component is responsible for implementation of the interface. In E.PathDash this consists of a menu sidebar, a header, and four main dashboards (Figure 2). The interface is rendered as a single HTML page and elements are hidden or shown based on the application state. The interface components are implemented by the 'shinydashboard' R package[78].

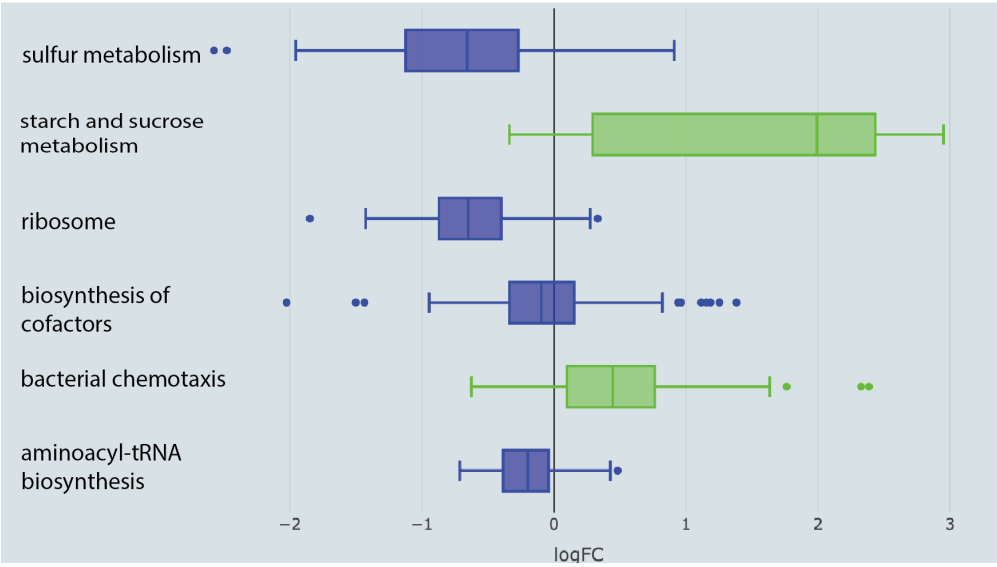
The "server" component loads the backend database and defines the results that are returned to the UI in response to how a user interacts with the application. The server code is split into sections that correspond to the four dashboards. Together with the "ui" component, this code utilizes several R libraries for plotting and design: 'plotly' for all application plots and 'shinycssloaders', 'shinyBS' and 'shinyjs' for design functionality[79–82]. For additional design customization the app.R file loads a custom CSS design file.

Pathway Visualizations

In addition to displaying pathway analysis results in a table format, E.PathDash shows visualizations that enhance the user's ability to compare different pathway activation or repression patterns within and across studies.

Boxplots show KEGG pathway and GO term activation and repression within a single study (Figure 9a). The boxplots capture the distribution of the logFC values for genes within each statistically significant activated/repressed KEGG pathway and GO term (adjusted p-value < 0.05). Showing the distribution of logFC values for genes within the pathway communicates repression/activation strength between treatment groups. The distributions are colored based on whether the pathway is activated or repressed (positive or negative median logFC value).

a)



b)

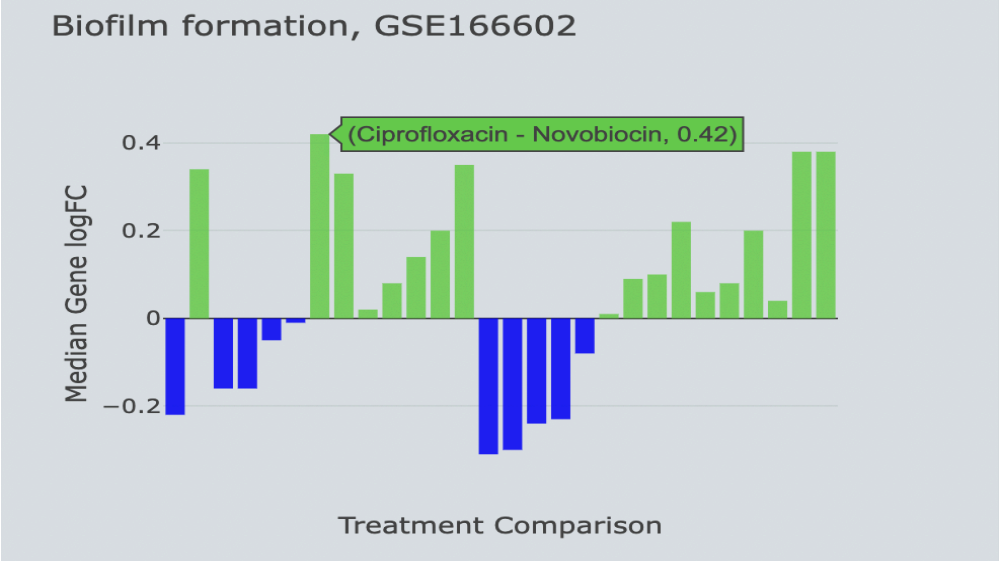


Figure 9. **(a)** Boxplot showing pathway analysis results for GEO dataset GSE142448, which measures responses of *P. aeruginosa* to manuka honey and its components. The plot shows distributions of the gene logFC values for KEGG pathways differentially expressed between artificial honey and manuka honey, which has a well-known antibacterial effect. **(b)** Bar plot showing how the KEGG biofilm formation pathway is expressed across all treatment comparisons in GEO dataset GSE166602, which exposed *P. aeruginosa* to different gyrase inhibitors.

When comparing pathway activation between studies, a bar chart shows pathway activation patterns across all possible treatment comparisons within each study (Figure 9b). Plots show the median logFC value for the genes within the KEGG pathway or GO term and the treatment comparison and median logFC value are displayed in a popup message upon hovering over each bar in the chart.

Data Availability

Raw count tables, constructed design matrices, and additional study metadata files for all GEO datasets included in this version of E.PathDash are available in the Git repository [https://github.com/DartCF/EPathDash/tree/main/GEO_Datasets].

Code Availability

All application code and data processing and analysis scripts are available in the Git repository [https://github.com/DartCF/EPathDash]. The application is hosted in a cloud computing environment maintained by Dartmouth Research Computing Infrastructure and can be accessed at scangeo.dartmouth.edu/EPathDash.

In its current version, E.PathDash utilizes the following R packages: *KEGGREST* (v 1.32.0), *EnrichmentBrowser* (v 2.28.0), *stringr* (v 1.5.0), *readxl* (v 1.4.3), *tidyverse* (v 1.3.2), *shinydashboard* (v 0.7.2), *plotly* (v 4.10.1), *shinycssloaders* (v 1.0.0), *shinyBS* (v 0.61.1), *shinyjs* (v 2.1.0). The application runs on R version 4.3.3.

Author Contributions

L.T wrote the publication except for the user stories and developed the E.PathDash application. L.T., S.N. and T.H. conceived of the application and T.H. provided guidance throughout the development and publication drafting processes. S.S., G.D., and C.F. tested the application and provided feedback on its utility, and S.S. and G.D. wrote user stories for this publication. B.S. contributed valuable feedback during publication drafting and provided primary funding for this project. All authors reviewed drafts of the publication.

Competing Interests

The authors declare no competing interests.

References

1. Soriano, J. B. *et al.* Prevalence and attributable health burden of chronic respiratory diseases, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet Respiratory Medicine* **8**, 585–596 (2020).
2. Marsland, B. J. & Gollwitzer, E. S. Host–microorganism interactions in lung diseases. *Nat Rev Immunol* **14**, 827–835 (2014).
3. Neff, S. L. *et al.* Pseudomonas aeruginosa transcriptome analysis of metal restriction in ex

vivo cystic fibrosis sputum. *Microbiol Spectr* e0315723 (2024) doi:10.1128/spectrum.03157-23.

4. Barrack, K. E. *et al.* An in vitro medium for modeling gut dysbiosis associated with cystic fibrosis. *Journal of Bacteriology* **206**, e00286-23 (2024).
5. Vieira, J. *et al.* Supplemental Oxygen Alters the Airway Microbiome in Cystic Fibrosis. *mSystems* **7**, e0036422 (2022).
6. Murphy, T. F. *et al.* Pseudomonas aeruginosa in Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* **177**, 853–860 (2008).
7. Fischer, A. J. *et al.* Sustained Coinfections with Staphylococcus aureus and Pseudomonas aeruginosa in Cystic Fibrosis. *Am J Respir Crit Care Med* **203**, 328–338 (2021).
8. Li, K., Gifford, A. H., Hampton, T. H. & O'Toole, G. A. Availability of Zinc Impacts Interactions between Streptococcus sanguinis and Pseudomonas aeruginosa in Coculture. *Journal of Bacteriology* **202**, 10.1128/jb.00618-19 (2020).
9. Filkins, L. M. *et al.* Prevalence of Streptococci and Increased Polymicrobial Diversity Associated with Cystic Fibrosis Patient Stability. *Journal of Bacteriology* **194**, 4709–4717 (2012).
10. Willner, D. L. *et al.* Reestablishment of recipient-associated microbiota in the lung allograft is linked to reduced risk of bronchiolitis obliterans syndrome. *Am J Respir Crit Care Med* **187**, 640–647 (2013).
11. Erb-Downward, J. R. *et al.* Analysis of the Lung Microbiome in the "Healthy" Smoker and in COPD. *PLOS ONE* **6**, e16384 (2011).
12. Dickson, R. P. *et al.* Enrichment of the lung microbiome with gut bacteria in sepsis and the acute respiratory distress syndrome. *Nat Microbiol* **1**, 1–9 (2016).
13. Guo, Y. *et al.* Association Studies on Gut and Lung Microbiomes in Patients with Lung Adenocarcinoma. *Microorganisms* **11**, 546 (2023).
14. Hilty, M. *et al.* Disordered Microbial Communities in Asthmatic Airways. *PLOS ONE* **5**,

e8578 (2010).

15. Borewicz, K. *et al.* Longitudinal analysis of the lung microbiome in lung transplantation. *FEMS Microbiology Letters* **339**, 57–65 (2013).
16. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science* **270**, 467–470 (1995).
17. Gondane, A. & Itkonen, H. M. Revealing the History and Mystery of RNA-Seq. *Curr Issues Mol Biol* **45**, 1860–1874 (2023).
18. Athar, A. *et al.* ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Research* **47**, D711–D715 (2019).
19. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic acids research* **41**, D991–D995 (2012).
20. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).
21. Kälin, N., Claaß, A., Sommer, M., Puchelle, E. & Tümmler, B. ΔF508 CFTR protein expression in tissues from patients with cystic fibrosis. *J Clin Invest* **103**, 1379–1389 (1999).
22. Painter, R. G. *et al.* CFTR Expression in human neutrophils and the phagolysosomal chlorination defect in cystic fibrosis. *Biochemistry* **45**, 10260–10269 (2006).
23. Tousson, A., Van Tine, B. A., Naren, A. P., Shaw, G. M. & Schwiebert, L. M. Characterization of CFTR expression and chloride channel activity in human endothelia. *Am J Physiol* **275**, C1555-1564 (1998).
24. Lyczak, J. B., Cannon, C. L. & Pier, G. B. Lung Infections Associated with Cystic Fibrosis. *Clinical Microbiology Reviews* **15**, 194–222 (2002).
25. *2022 Annual Data Report*. https://www.cff.org/medical-professionals/patient-registry.
26. Gangell, C. *et al.* Inflammatory Responses to Individual Microorganisms in the Lungs of Children With Cystic Fibrosis. *Clinical Infectious Diseases* **53**, 425–432 (2011).
27. Pillarisetti, N. *et al.* Infection, inflammation, and lung function decline in infants with cystic

fibrosis. *Am J Respir Crit Care Med* **184**, 75–81 (2011).

28. Malhotra, S., Hayes, D. & Wozniak, D. J. Cystic Fibrosis and Pseudomonas aeruginosa: the Host-Microbe Interface. *Clin Microbiol Rev* **32**, e00138-18 (2019).
29. Venkateswaran, P. *et al.* Revisiting ESKAPE Pathogens: virulence, resistance, and combating strategies focusing on quorum sensing. *Front Cell Infect Microbiol* **13**, 1159798 (2023).
30. De Oliveira, D. M. P. *et al.* Antimicrobial Resistance in ESKAPE Pathogens. *Clinical Microbiology Reviews* **33**, 10.1128/cmr.00181-19 (2020).
31. Rubinstein, E., Kollef, M. H. & Nathwani, D. Pneumonia Caused by Methicillin-Resistant Staphylococcus aureus. *Clinical Infectious Diseases* **46**, S378–S385 (2008).
32. Eklöf, J. *et al. Pseudomonas aeruginosa* and risk of death and exacerbations in patients with chronic obstructive pulmonary disease: an observational cohort study of 22 053 patients. *Clinical Microbiology and Infection* **26**, 227–234 (2020).
33. Werner, T. Bioinformatics applications for pathway analysis of microarray data. *Current Opinion in Biotechnology* **19**, 50–54 (2008).
34. Reyes, A. L. P. *et al.* GENAVi: a shiny web application for gene expression normalization, analysis and visualization. *BMC Genomics* **20**, 745 (2019).
35. Mubeen, S. *et al.* DecoPath: a web application for decoding pathway enrichment analysis. *NAR Genomics and Bioinformatics* **3**, lqab087 (2021).
36. Koeppen, K., Hampton, T. H., Neff, S. L. & Stanton, B. A. ESKAPE Act Plus: Pathway Activation Analysis for Bacterial Pathogens. *Msystems* **7**, e00468-22 (2022).
37. Ge, S. X., Son, E. W. & Yao, R. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**, 534 (2018).
38. Neff, S. L. *et al.* CF-Seq, an accessible web application for rapid re-analysis of cystic fibrosis pathogen RNA sequencing studies. *Scientific Data* **9**, 343 (2022).
39. Koeppen, K., Stanton, B. A. & Hampton, T. H. ScanGEO: parallel mining of high-throughput

gene expression data. *Bioinformatics* **33**, 3500–3501 (2017).

40. Neff, S. L. *et al.* Rocket-miR, a translational launchpad for miRNA-based antimicrobial drug development. *mSystems* **8**, e00653-23 (2023).
41. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27–30 (2000).
42. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic acids research* **51**, D587–D592 (2023).
43. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research* **32**, D258–D261 (2004).
44. Consortium, T. G. O. The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).
45. Doing, G., Koeppen, K., Occipinti, P., Harty, C. E. & Hogan, D. A. Conditional antagonism in co-cultures of Pseudomonas aeruginosa and Candida albicans: An intersection of ethanol and phosphate signaling distilled from dual-seq transcriptomics. *PLoS Genetics* **16**, (2020).
46. Dolan, S. K. *et al.* Systems-Wide Dissection of Organic Acid Assimilation in Pseudomonas aeruginosa Reveals a Novel Path To Underground Metabolism. *mBio* **13**, e02541-22 (2022).
47. Zheng, H. *et al.* Ethanol effects on the overexpression of heterologous catalase in Escherichia coli BL21 (DE3). *Appl Microbiol Biotechnol* **103**, 1441–1453 (2019).
48. Kretzschmar, U., Schobert, M. & Görisch, H. The Pseudomonas aeruginosa acsA gene, encoding an acetyl-CoA synthetase, is essential for growth on ethanol. *Microbiology* **147**, 2671–2677 (2001).
49. Kretzschmar, U., Khodaverdi, V. & Adrian, L. Transcriptional regulation of the acetyl-CoA synthetase gene acsA in Pseudomonas aeruginosa. *Arch Microbiol* **192**, 685–690 (2010).
50. Cugini, C. *et al.* Farnesol, a common sesquiterpene, inhibits PQS production in Pseudomonas aeruginosa. *Molecular Microbiology* **65**, 896–906 (2007).

51. Bouzo, D. *et al.* Characterizing the Mechanism of Action of an Ancient Antimicrobial, Manuka Honey, against Pseudomonas aeruginosa Using Modern Transcriptomics. *mSystems* **5**, e00106-20 (2020).
52. Sukdeo, N. & Honek, J. F. Pseudomonas aeruginosa contains multiple glyoxalase I-encoding genes from both metal activation classes. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1774**, 756–763 (2007).
53. Cui, G. *et al.* PmiR senses 2-methylisocitrate levels to regulate bacterial virulence in Pseudomonas aeruginosa. *Science Advances* **8**, eadd4220 (2022).
54. Petrocheilou, A., Moudaki, A. & Kaditis, A. G. Inflammation and Infection in Cystic Fibrosis: Update for the Clinician. *Children (Basel)* **9**, 1898 (2022).
55. Rasamiravaka, T., Labtani, Q., Duez, P. & El Jaziri, M. The formation of biofilms by Pseudomonas aeruginosa: a review of the natural and synthetic compounds interfering with control mechanisms. *Biomed Res Int* **2015**, 759348 (2015).
56. Oluyombo, O., Penfold, C. N. & Diggle, S. P. Competition in Biofilms between Cystic Fibrosis Isolates of Pseudomonas aeruginosa Is Shaped by R-Pyocins. *mBio* **10**, e01828-18 (2019).
57. Antibiotics | Cystic Fibrosis Foundation. https://www.cff.org/managing-cf/antibiotics.
58. Morita, Y., Tomida, J. & Kawamura, Y. Responses of Pseudomonas aeruginosa to antimicrobials. *Front Microbiol* **4**, 422 (2014).
59. Koeppen, K. *et al.* Let-7b-5p in vesicles secreted by human airway cells reduces biofilm formation and increases antibiotic sensitivity of P. aeruginosa. *Proc Natl Acad Sci U S A* **118**, e2105370118 (2021).
60. Moreau-Marquis, S., O'Toole, G. A. & Stanton, B. A. Tobramycin and FDA-approved iron chelators eliminate Pseudomonas aeruginosa biofilms on cystic fibrosis cells. *Am J Respir Cell Mol Biol* **41**, 305–313 (2009).
61. Moreau-Marquis, S., Coutermarsh, B. & Stanton, B. A. Combination of hypothiocyanite and

lactoferrin (ALX-109) enhances the ability of tobramycin and aztreonam to eliminate Pseudomonas aeruginosa biofilms growing on cystic fibrosis airway epithelial cells. *J Antimicrob Chemother* **70**, 160–166 (2015).

62. Masadeh, M. M., Alzoubi, K. H., Ahmed, W. S. & Magaji, A. S. In Vitro Comparison of Antibacterial and Antibiofilm Activities of Selected Fluoroquinolones against Pseudomonas aeruginosa and Methicillin-Resistant Staphylococcus aureus. *Pathogens* **8**, 12 (2019).
63. Dewangan, R. P., Singh, M., Ilic, S., Tam, B. & Akabayov, B. Cell-penetrating peptide conjugates of indole-3-acetic acid-based DNA primase/Gyrase inhibitors as potent anti-tubercular agents against planktonic and biofilm culture of Mycobacterium smegmatis. *Chem Biol Drug Des* **98**, 722–732 (2021).
64. Franke, R., Overwin, H., Häussler, S. & Brönstrup, M. Targeting Bacterial Gyrase with Cystobactamid, Fluoroquinolone, and Aminocoumarin Antibiotics Induces Distinct Molecular Signatures in Pseudomonas aeruginosa. *mSystems* **6**, e0061021 (2021).
65. Khalid, S. J. *et al.* Targeting Acyl Homoserine Lactones (AHLs) by the quorum quenching bacterial strains to control biofilm formation in Pseudomonas aeruginosa. *Saudi J Biol Sci* **29**, 1673–1682 (2022).
66. Stanton, B. A., Hampton, T. H. & Ashare, A. SARS-CoV-2 (COVID-19) and cystic fibrosis. *Am J Physiol Lung Cell Mol Physiol* **319**, L408–L415 (2020).
67. Koeppen, K. *et al.* Tobramycin reduces key virulence determinants in the proteome of Pseudomonas aeruginosa outer membrane vesicles. *PLoS One* **14**, e0211290 (2019).
68. Li, Z. *et al.* GAUGE-Annotated Microbial Transcriptomic Data Facilitate Parallel Mining and High-Throughput Reanalysis To Form Data-Driven Hypotheses. *mSystems* **6**, 10.1128/msystems.01305-20 (2021).
69. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2022).
70. Purves, J. *et al.* Air pollution induces Staphylococcus aureus USA300 respiratory tract

colonization mediated by specific bacterial genetic responses involving the global virulence gene regulators Agr and Sae. *Environmental Microbiology* **24**, 4449–4465 (2022).

71. Tenenbaum, D. & Maintainer, B. P. *KEGGREST: Client-Side REST Access to the Kyoto Encyclopedia of Genes and Genomes (KEGG)*. (2022). doi:10.18129/B9.bioc.KEGGREST.
72. Pundir, S., Martin, M. J., O'Donovan, C. & Consortium, U. UniProt tools. *Current protocols in bioinformatics* **53**, 1–29 (2016).
73. Entropy | Free Full-Text | Fifteen Years of Gene Set Analysis for High-Throughput Genomic Data: A Review of Statistical Approaches and Future Challenges. https://www.mdpi.com/1099-4300/22/4/427.
74. Geistlinger, L., Csaba, G. & Zimmer, R. Bioconductor's EnrichmentBrowser: seamless navigation through combined results of set- & network-based enrichment analysis. *BMC Bioinformatics* **17**, 45 (2016).
75. Wickham, H. *Stringr: Simple, Consistent Wrappers for Common String Operations*. (2022).
76. Wickham, H. & Bryan, J. *Readxl: Read Excel Files*. (2023).
77. Wickham, H. *et al.* Welcome to the tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).
78. Chang, W. & Ribeiro, B. B. *Shinydashboard: Create Dashboards with 'Shiny'*. (2021).
79. Sievert, C. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny*. (Chapman and Hall/CRC, 2020).
80. Sali, A. & Attali, D. *Shinycssloaders: Add Loading Animations to a 'shiny' Output While It's Recalculating*. (2020).
81. Bailey, E. *shinyBS: Twitter Bootstrap Components for Shiny*. (2022).
82. Attali, D. *Shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds*. (2021).