

# 1 Plasma Cell-free RNA Signatures of Inflammatory Syndromes in Children

2 **Authors:** Conor J. Loy<sup>1</sup>, Venice Servellita<sup>2</sup>, Alicia Sotomayor-Gonzalez<sup>2</sup>, Andrew Bliss<sup>1</sup>, Joan Lenz<sup>1</sup>, Emma  
3 Belcher<sup>1</sup>, Will Suslovic<sup>3</sup>, Jenny Nguyen<sup>2</sup>, Meagan E. Williams<sup>3</sup>, Miriam Oseguera<sup>2</sup>, Michael A. Gardiner<sup>4</sup>, Pediatric  
4 Emergency Medicine Kawasaki Disease Research Group (PEMKDRG), The CHARMS Study Group, Jong-Ha  
5 Choi<sup>5</sup>, Hui-Mien Hsiao<sup>5</sup>, Hao Wang<sup>6</sup>, Jihoon Kim<sup>7</sup>, Chisato Shimizu<sup>6</sup>, Adriana Tremoulet<sup>4,6</sup>, Meghan Delaney<sup>3</sup>,  
6 Roberta L. DeBiasi<sup>3</sup>, Christina A. Rostad<sup>5</sup>, Jane C. Burns<sup>4,6\*</sup>, Charles Y. Chiu<sup>2,8,9\*</sup>, Iwijn De Vlaminc<sup>1\*</sup>

## 8 Affiliations:

9 <sup>1</sup> Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY 14850, USA

10 <sup>2</sup> Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA 94143, USA

11 <sup>3</sup> Children's National Hospital, Washington D.C. 20010, USA

12 <sup>4</sup> Rady Children's Hospital-San Diego, San Diego, CA, USA

13 <sup>5</sup> Department of Pediatrics and Center for Childhood Infections at Vaccines, Emory University School of Medicine  
14 and Children's Healthcare of Atlanta, Atlanta, GA 30322, USA

15 <sup>6</sup> Kawasaki Disease Research Center, Department of Pediatrics, Rady Children's Hospital San Diego and University  
16 of California San Diego, La Jolla, CA, USA

17 <sup>7</sup> Biomedical Informatics & Data Science, Yale School of Medicine, New Haven, CT 06510 USA

18 <sup>8</sup> Department of Medicine, Division of Infectious Diseases, University of California, San Francisco, San Francisco,  
19 CA 94158, USA

20 <sup>9</sup> Chan-Zuckerberg Biohub, San Francisco, CA 94158, USA

21  
22 \*Co-corresponding authors:

23 Jane Burns, [jcburns@health.ucsd.edu](mailto:jcburns@health.ucsd.edu)

24 Charles Chiu, [charles.chiu@ucsf.edu](mailto:charles.chiu@ucsf.edu)

25 Iwijn De Vlaminc, [vlaminc@cornell.edu](mailto:vlaminc@cornell.edu)

## 27 ABSTRACT

28 Inflammatory syndromes, including those caused by infection, are a major cause of hospital  
29 admissions among children and are often misdiagnosed because of a lack of advanced molecular  
30 diagnostic tools. In this study, we explored the utility of circulating cell-free RNA (cfRNA) in plasma as an  
31 analyte for the differential diagnosis and characterization of pediatric inflammatory syndromes. We  
32 profiled cfRNA in 370 plasma samples from pediatric patients with a range of inflammatory conditions,  
33 including Kawasaki disease (KD), Multisystem Inflammatory Syndrome in Children (MIS-C), viral  
34 infections and bacterial infections. We developed machine learning models based on these cfRNA  
35 profiles, which effectively differentiated KD from MIS-C — two conditions presenting with overlapping  
36 symptoms — with high performance (Test Area Under the Curve (AUC) = 0.97). We further extended this  
37 methodology into a multiclass machine learning framework that achieved 81% accuracy in distinguishing  
38 among KD, MIS-C, viral, and bacterial infections. We further demonstrated that cfRNA profiles can be  
39 used to quantify injury to specific tissues and organs, including the liver, heart, endothelium, nervous  
40 system, and the upper respiratory tract. Overall, this study identified cfRNA as a versatile analyte for the  
41 differential diagnosis and characterization of a wide range of pediatric inflammatory syndromes.

## 43 INTRODUCTION

44 The differential diagnosis of inflammatory syndromes in children is complex owing to their  
45 overlapping clinical manifestations, non-specific symptoms and developmental age-related barriers to  
46 communication. These challenges often result in delayed or inaccurate diagnoses, thereby impeding

47 effective clinical management and increasing the risk of long-term adverse health effects. A key example  
48 is Kawasaki disease (KD), an inflammatory syndrome of unknown etiology that primarily affects children  
49 under five years of age. KD is often misdiagnosed because significant clinical signs overlap between KD  
50 and other inflammatory and/or infectious conditions. Accurate diagnosis is critical, as KD patients who do  
51 not receive intravenous immunoglobulin (IVIG) early in the course of illness have a substantially  
52 increased risk of developing coronary artery aneurysms, making KD the leading cause of acquired heart  
53 disease in children. Thus, there is a clear need for accurate molecular tests for inflammatory conditions  
54 such as KD to inform timely and appropriate treatment.

55 Currently, the differential diagnosis of pediatric inflammatory syndromes relies on clinical  
56 assessment of signs and symptoms and results from a broad array of laboratory tests. Culture-based  
57 and molecular assays are routinely used to identify viral and bacterial pathogens, but these tests do not  
58 interrogate the host response and hence are unable to differentiate between infectious and non-infectious  
59 conditions. Serologic metabolic, and antigen biomarkers used for diagnosis often lack specificity<sup>1,2</sup>. To  
60 address these limitations, recent studies have explored the use of whole blood RNA transcriptome  
61 profiling of the human host response to assess disease severity and differentiate among inflammatory  
62 conditions including KD, MIS-C, viral, and bacterial infection<sup>3-6</sup>. However, while the whole blood profile is  
63 indicative of the host immune response, it provides little information regarding the extent of inflammation-  
64 related cell injury or death in solid organ tissues. In contrast, cell-free nucleic acids in plasma, including  
65 cell-free DNA (cfDNA) and cell-free RNA (cfRNA) are promising analytes for evaluating inflammation as  
66 they are released by dead or dying cells originating from both the bloodstream and solid tissues<sup>7</sup>. Many  
67 recent studies have explored the use of cell-free nucleic acids for monitoring of pregnancy<sup>8,9</sup>, cancer<sup>10,11</sup>,  
68 transplantation<sup>12,13</sup> and infection<sup>5,14-23</sup>, yet the potential of cell-free nucleic acids for the differential  
69 diagnosis of inflammatory syndromes remains largely unexplored.

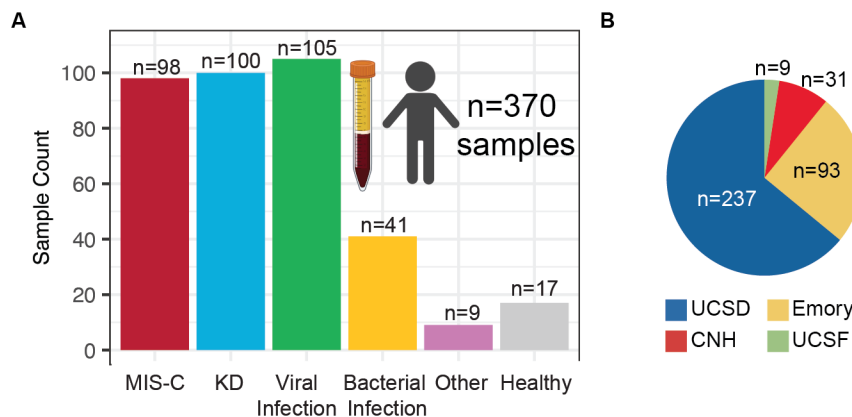
70 In this study, we applied plasma cell-free RNA (cfRNA) profiling by RNA sequencing to compare  
71 host immune and cellular injury responses associated with four different inflammatory and/or infectious  
72 syndromes in children. We identified shared signatures across multiple conditions, highlighting the  
73 importance of incorporating multiple comparison groups for the development of disease-specific  
74 biomarkers. Using cfRNA profiles, we built a machine learning model with high accuracy in differentiating  
75 between MIS-C and KD, two pediatric inflammatory syndromes with overlapping clinical presentations.  
76 We then expanded this methodology to construct a multi-class diagnostic classifier capable of  
77 distinguishing among KD, MIS-C, viral, and bacterial infections. In addition, we demonstrate that the  
78 cfRNA profile can be correlated with markers of tissue damage and may be able to differentiate among  
79 different viral infections. We propose an application of cfRNA profiling as a decision support tool for  
80 differential diagnosis of inflammatory syndromes in the clinical setting.

81

## 82 **RESULTS**

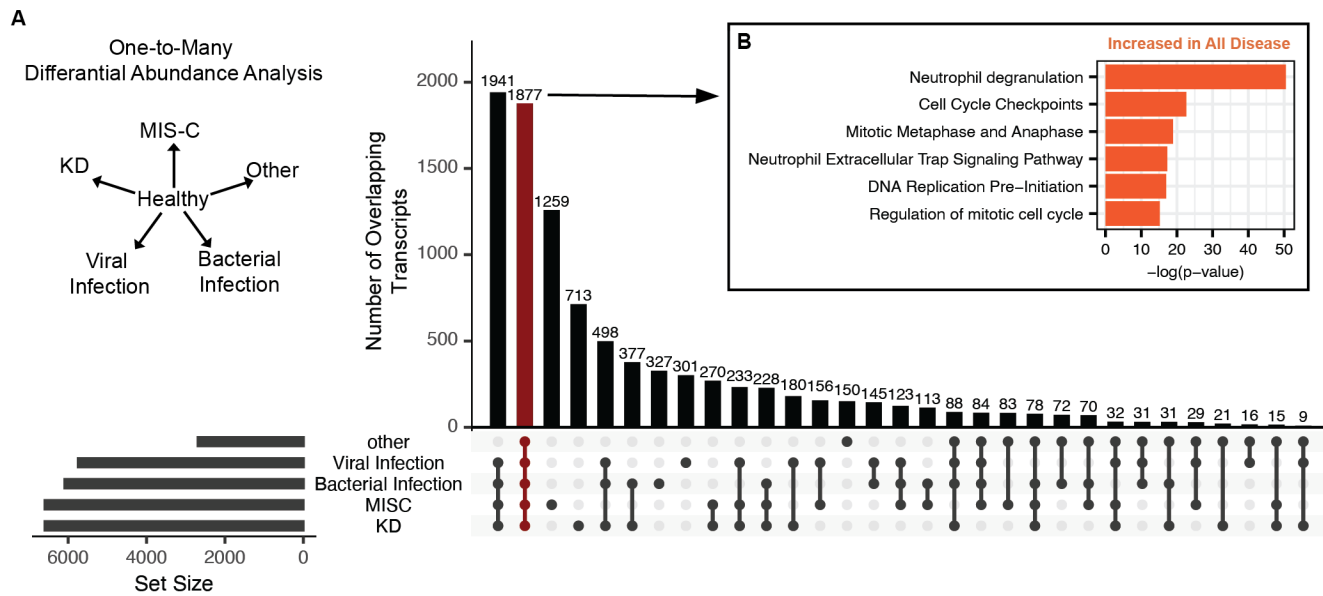
83 **Clinical cohort.** We collected and analyzed 370 plasma samples from pediatric patients with  
84 inflammatory and infectious conditions at four hospitals in the US, including Rad Children's Hospital San  
85 Diego (UCSD), Emory, Children's National Hospital (CNH), and University of California San Francisco  
86 (UCSF) (**Figure 1A and B**). This sample set included patients diagnosed with KD, MIS-C, viral infection,  
87 bacterial infection, and other hospitalized pediatric controls, some but not all with inflammatory disease  
88 (for example, arthralgia, Crohn's disease flare, parenchymal lung disease, chronic lung disease, toxic  
89 shock syndrome, and post-vaccine myocarditis), as well as healthy children (**Figure 1A and Table 1**).  
90 Included cases of bacterial and viral infections were heterogeneous with respect to infection site and

91 pathogen (**Supplementary File 1**). All patient samples were collected within 4 days of hospitalization  
92 during the acute phase of disease (Methods). RNA next-generation sequencing was used to quantify the  
93 genes and cell types of origin of cfRNA in each sample, with an average of 20.7 million reads sequenced  
94 per sample (Methods). Machine learning models were built to identify biosignatures associated with  
95 different diseases.



**Figure 1. Sample Overview. A)** Sample counts for each disease group. “Other” indicates other hospitalized pediatric controls. **B)** Distribution of collected samples based on hospital site.

96 **cfRNA signatures of disease.** Our initial focus was on characterizing changes in cfRNA profiles in  
97 plasma that are common across the various inflammatory conditions. We performed pairwise differential  
98 gene abundance analysis between the healthy control group and each disease group (Methods, **Figure**  
99 **2A** and **Supplementary File 2**). This analysis identified differentially abundant genes (DAGs) for all  
100 comparisons (BH adjusted p-value < 0.05). The smallest number of DAGs was identified when comparing  
101 healthy controls to patients in the other hospitalized control group (n=2,686) and the greatest number  
102 when comparing healthy controls to patients with KD (n=6,591, **Figure 2A**). Further analysis of the DAGs  
103 for all conditions revealed a significant number of shared genes (1,877 DAGs, **Figure 2A**). Chief among  
104 these were histone protein coding genes that were elevated for all disease groups, indicating that an  
105 increase in histone transcripts is a universal indicator of inflammation (**Figure S1A**). The analysis also  
106 identified immune related transcripts MPO, ELANE, CD53, and CXCR2 elevated in each disease group,  
107 homeostasis related transcripts CDK19, ANAPC5, and 32 mitochondrial protein coding RNAs in the  
108 control group (**Figure S2B**). Pathway analysis of the 1,877 overlapping DAGs revealed an enrichment of  
109 neutrophil and cell replication transcripts related to inflammation (**Figure 2B** and **Supplementary File**  
110 **3**). These findings point to shared signatures of inflammation among disease groups and emphasize the  
111 need for inter-group comparisons to develop disease-specific biosignatures.

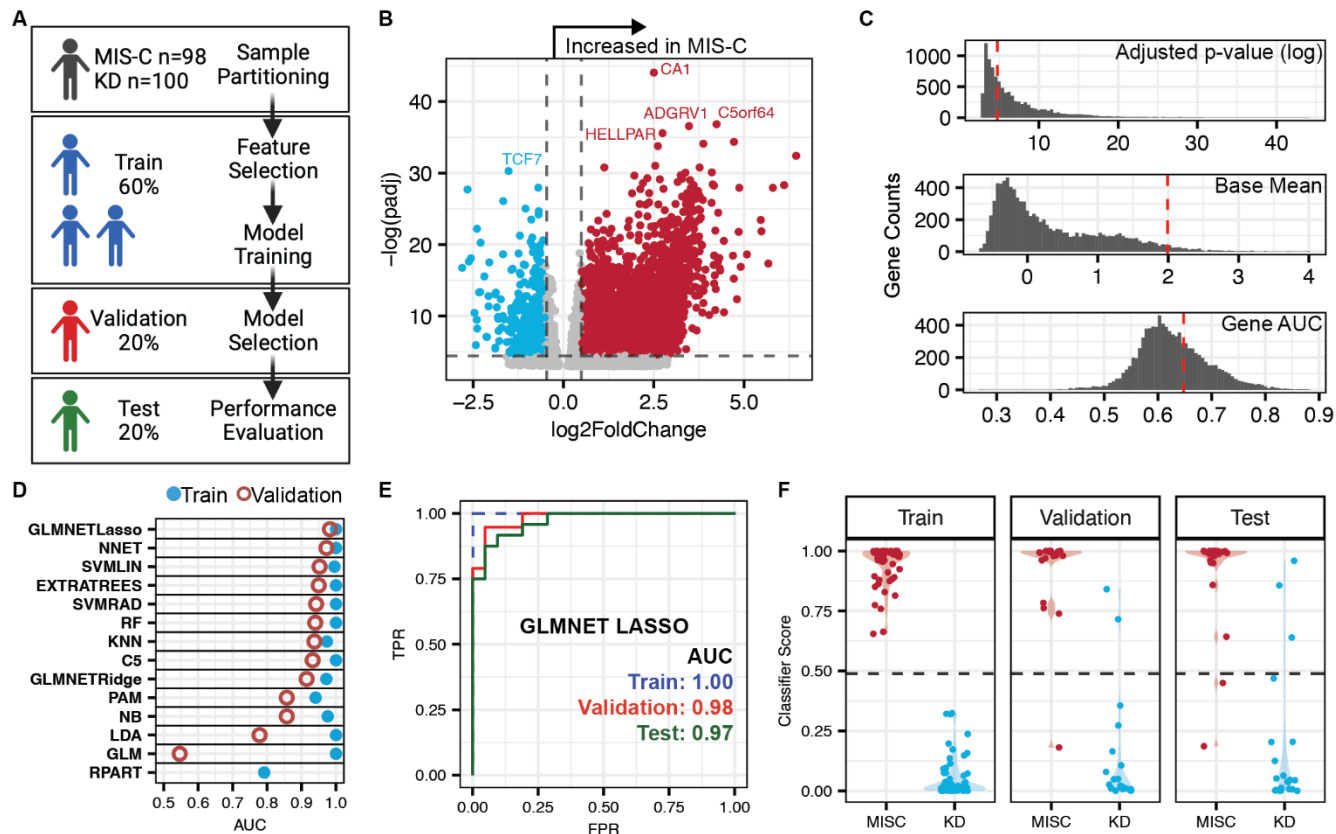


**Figure 2. Shared cfRNA signatures among different inflammation syndromes. A)** Differential abundance analysis using DESeq2 was performed between healthy controls and each other group individually. Vertical columns indicate the number of overlapping genes that are significantly differentially abundant between groups and controls (Benjamini-Hochberg (BH) adjusted p-values < 0.05). Dots below bars indicate the groups being intersected. Horizontal columns indicate the total number of DAGs between groups and controls. **B)** Significantly enriched pathways in the set of genes found to be differentially abundant between healthy controls and each disease group. Average p-value and fold change used for pathway analysis (Qiagen, IPA).

112 **cfRNA can differentiate MIS-C and Kawasaki Disease.** KD and MIS-C share many clinical  
 113 characteristics: they are highly inflammatory, present with endothelial dysfunction, and have multiple  
 114 overlapping signs, including skin rashes, mucosal involvement, and fever. There are currently no  
 115 molecular tests to distinguish between KD and MIS-C. We therefore assessed if cfRNA could be used to  
 116 differentiate these severe pediatric inflammatory syndromes. We divided the KD (n=100) and MIS-C  
 117 (n=98) samples into training (60%), validation (20%), and test (20%) sets, ensuring a roughly equal  
 118 representation of hospital of origin and disease subclassification across all 3 sets (**Figure 3A**). We used  
 119 the training data for feature selection and to train machine learning models. We then used the validation  
 120 set to select the final model, based on the model with the highest receiving operator characteristic Area  
 121 Under the Curve (AUC). Last, we evaluated the performance of the final model using the test set. To  
 122 prevent influence of the test set on the training data and to ensure unbiased results, we normalized each  
 123 set separately using a variance stabilizing transformation (Methods).

124 In our initial analysis using the training set, we identified 1,242 differentially abundant genes  
 125 between KD and MIS-C (DESeq2, Benjamini-Hochberg (BH) adjusted p-value < 0.01 and base mean >  
 126 10; **Figure 3B** and **Supplementary File 4**). We refined this gene list based on adjusted p-value, base  
 127 mean, individual gene AUC, and fold change, resulting in a final tally of 132 genes for model input (**Figure**  
 128 **3C**). We then trained 14 machine learning classification models, evaluating their performance on the  
 129 validation set (Methods). The GLMNET model with LASSO regression exhibited the highest validation  
 130 set AUC and was selected as our final model (**Figure 3D**). A unique feature of the GLMNET LASSO  
 131 algorithm is the feature selection step, which selected 25 genes for the modeling (**Table S1**). Using this  
 132 trained model, we tested the classification performance using the test set (ROC-AUC train=1.00,

133 validation=0.98, test=0.97) (**Figure 3E**). We also observed similar distributions in the classification scores  
 134 across sample sets, further confirming that there was little to no overfitting of the training and test sets  
 135 (**Figure 3F**).



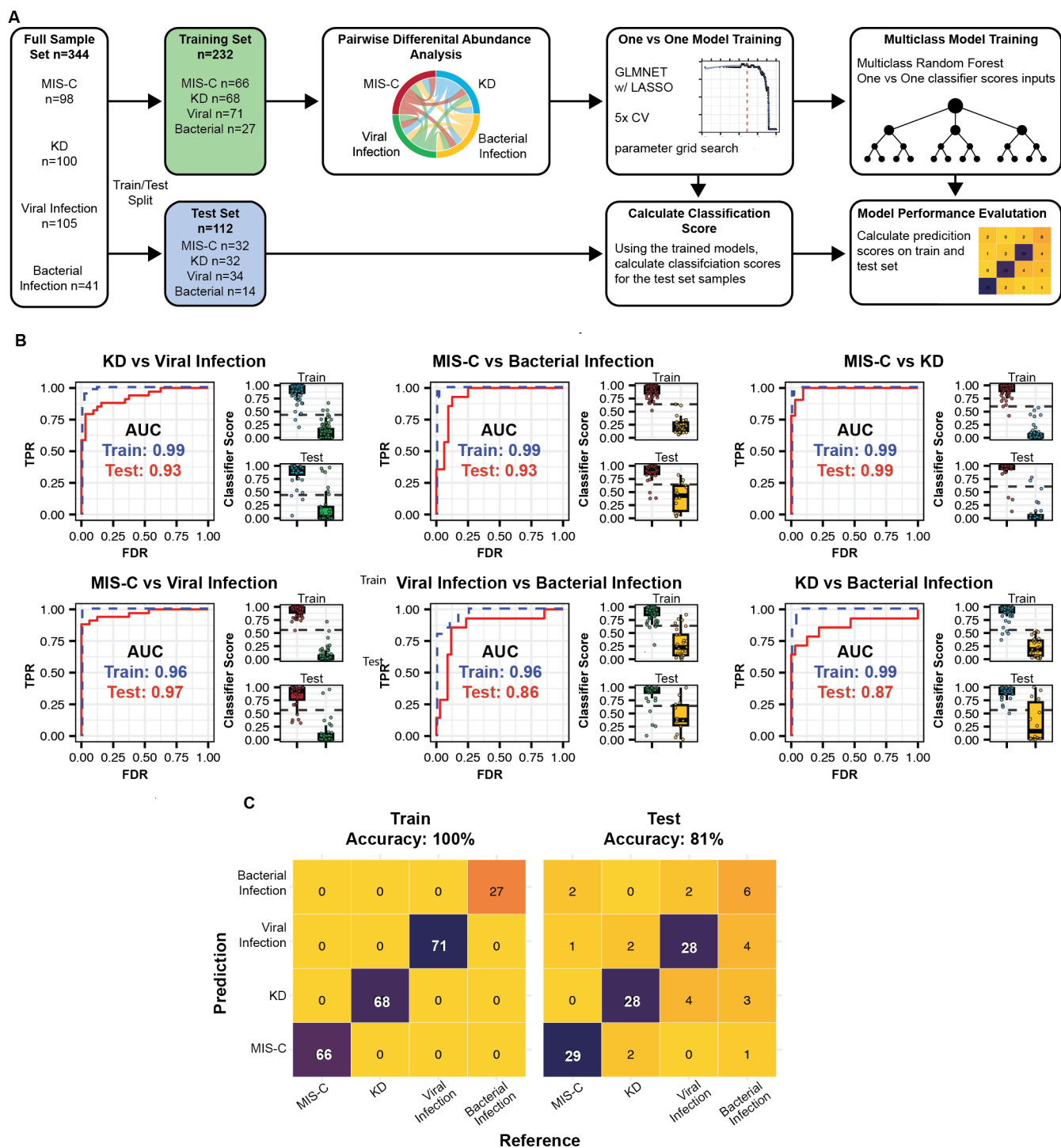
**Figure 3. cfRNA distinguishes KD and MIS-C. A)** Overview of sample set and modeling scheme. **B)** Volcano plot of differentially abundant transcripts between MIS-C and KD. Analysis was performed using the training data set (DESeq2). **C)** Adjusted p-value, base mean, and gene ROC AUC distributions for all genes from the training KD vs MIS-C comparison. **D)** ROC AUC values for the 14 machine learning classification models applied to training and validation sets. **E)** ROC-AUC curves of the training, validation, and test sets using the GLMNET with LASSO regression algorithm. **F)** Violin plots of the classifier scores from the GLMNET with LASSO regression algorithm.

136  
 137 **Multi-disease classification using cfRNA.** We next asked if cfRNA could be used in the more  
 138 challenging scenario of multi-disease classification. For this, we developed a machine learning framework  
 139 that combines the outputs of one-vs-one models using a random forest multiclass algorithm (Methods,  
 140 **Figure 4A**). We first split our dataset into training (70%) and test (30%) sets, with roughly comparable  
 141 proportions of samples from patients with KD, MIS-C, viral infection, or bacterial infection, while also  
 142 stratifying the groups evenly with respect to hospital of origin and disease subclassification. We  
 143 performed differential abundance analysis for each pairwise comparison using only the training data  
 144 (**Supplementary File 5**). Next, we trained individual one-vs-one GLMNET models with LASSO  
 145 regression for each sample group (MIS-C, KD, viral infection, or bacterial infection), using the top 100  
 146 genes identified in the differential abundance analysis. These genes were selected based on gene

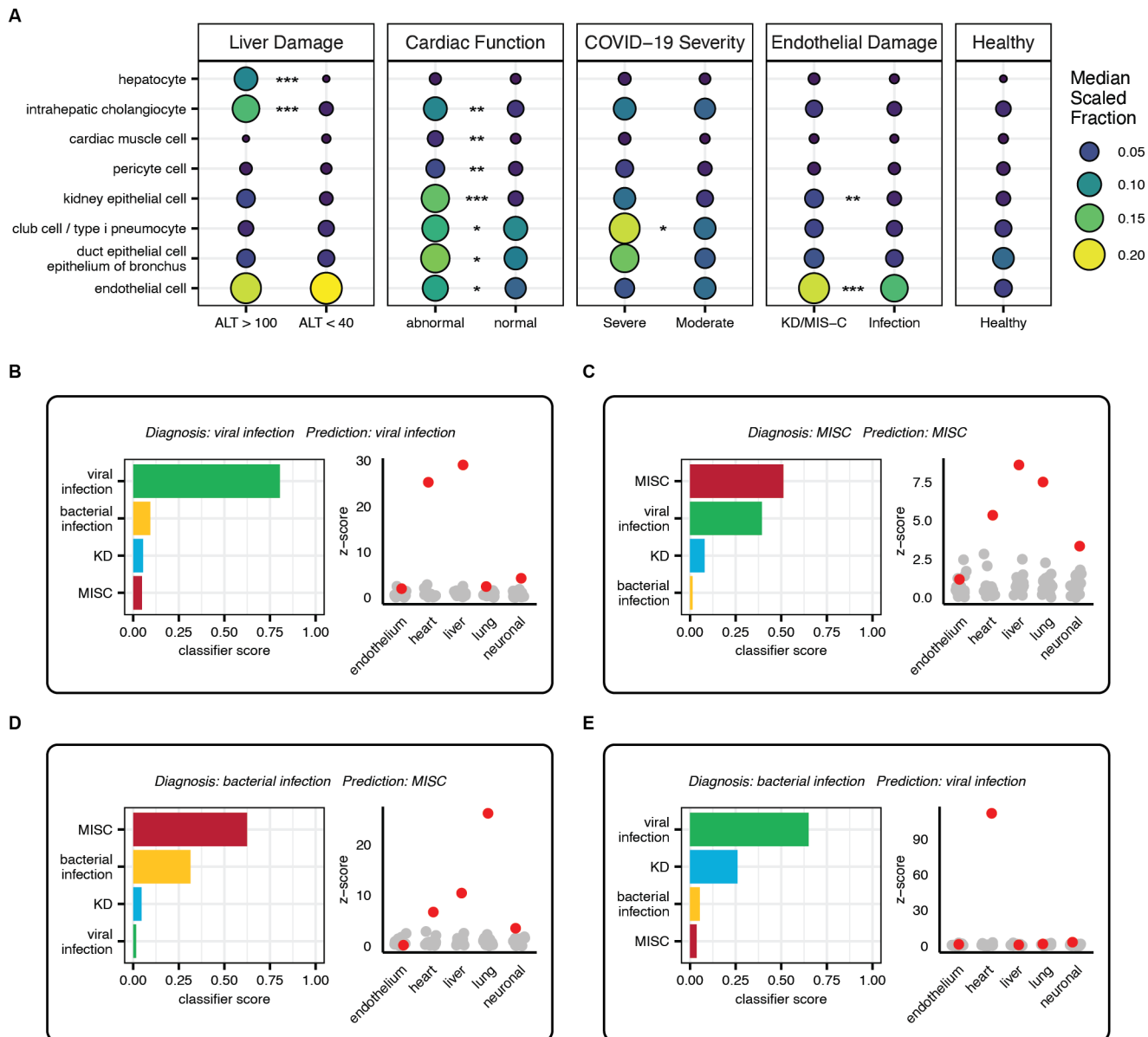
147 abundance AUC and level of significance (Methods). Each of the one-vs-one models demonstrated high  
148 performance with individual model scores classifying samples with high accuracy (Test AUC: min=0.86,  
149 max=0.99; Test Accuracy: min=0.83, max=0.92; **Figure 4B**). The union of genes used by each GLMNET  
150 model with LASSO regression generated a 109-gene panel (**Supplementary File 6**). To combine the  
151 outputs of the individual models, we trained a multiclass random forest classifier using the classification  
152 scores from the one-vs-one models. Using this framework, our multiclass machine learning model  
153 achieved high accuracy in both the training and test sets (Accuracy, train=100% and test=81%), with the  
154 highest rate of misclassified samples in the bacterial infection group. The high performance of the  
155 multiclass machine learning model on a relatively small number of genes points to the potential utility of  
156 cfRNA in differentiating complex inflammatory conditions in a clinical setting.

157 **Characterization of disease using cfRNA.** We further investigated whether cfRNA could be employed  
158 not only for classification but also for disease characterization. Since inflammation and/or infection can  
159 impact multiple organ systems, understanding organ-specific damage and function is crucial for guiding  
160 clinical management. Previous work from our group and others has shown that quantifying the cfRNA  
161 cell-type-of-origin (CTO) is a viable non-invasive method to assess cell, tissue and organ-specific  
162 injury<sup>17,24</sup>. Here, we employed deconvolution of the cell types of origin of cfRNA using BayesPrism and  
163 the Tabula Sapiens human cell atlas as a reference. The CTO estimates were compared with known  
164 biomarkers and other indicators of specific organ injury or dysfunction (Methods, **Figure 5A**). We first  
165 compared hepatocyte and intrahepatic cholangiocyte contributions to the cfRNA in plasma to alanine  
166 transaminase measurements (ALT, n=141). We observed significantly elevated levels of cfRNA from  
167 hepatocytes and intrahepatic cholangiocytes in patients with high ALT (ALT > 100 IU/L) compared to  
168 patients with normal ALT (ALT < 40 IU/L, **Figure 5A**). To explore cardiac function and damage, samples  
169 were categorized as having either normal or abnormal cardiac function (Methods). In the abnormal group,  
170 we observed increased levels of cfRNA from cardiac muscle cells and pericytes, likely indicative of  
171 increased cardiac cell injury or death. Interestingly, this group also exhibited elevated cfRNA levels from  
172 kidney epithelial cells, intrahepatic cholangiocytes, club cells and type I pneumocytes, and bronchial  
173 epithelium, suggesting other end organ damage associated with impaired cardiac function (**Figure 5A**).  
174 Of note, patients with abnormal cardiac function did not have elevated levels of smooth or skeletal muscle  
175 cell derived cfRNA (**Figure S2A**). Last, to evaluate lung function and damage we focused on samples  
176 from healthy individuals and viral infection patients with COVID-19, further stratifying the COVID-19  
177 samples by disease severity (Methods). We observed elevated levels of cfRNA from club cells and type  
178 I pneumocytes in moderate COVID-19 cases compared to healthy individuals, with a greater increase for  
179 severe cases (**Figure 5A**). We observed similar trends in the levels of bronchial epithelium derived  
180 cfRNA; however, the difference was not statistically significant.

181 We next assessed if cfRNA could be used to further stratify sample groups by distinguishing  
182 between COVID-19 infection from SARS-CoV-2 and infections from other viral pathogens (Influenza,  
183 RSV, EBV, Adenovirus, etc). We separated viral infection samples into COVID-19 and non-COVID-19  
184 viral infections and randomly split the data into training (70%) and testing (30%) data sets. Next, we  
185 trained a GLMNET with LASSO regression model to differentiate between these groups (Methods). The  
186 trained model had high performance on the training and test sets despite including only seven genes  
187 (train AUC = 0.99, test AUC = 0.93), demonstrating the potential of cfRNA for respiratory viral species  
188 differentiation (**Figure S2B-C**).



**Figure 4. Multiclass Classification of pediatric disease using cfRNA. A)** Overview of machine learning framework used for multiclass classification. **B)** ROC-AUC plots for each one-vs-one model trained, along with train (top) and test (bottom) classifier score distributions. **C)** Confusion matrix of reference and predicted diagnoses for train and test samples. Color indicates number of samples in each category.



**Figure 5. cfRNA as a clinical decision support tool. A)** Median scaled cell type of origin fractions for samples separated by liver damage, cardiac function, COVID-19 severity, and endothelial damage, and healthy controls (Methods). Stars indicate statistically significant differences between groups in comparison (Wilcoxon rank sum test, BH adjusted p-value < 0.05). **B-E)** Case studies of patients with example clinical decision support tool results from the multiclass algorithm, along with measurements of endothelium, heart, liver, lung, and neuronal damage from the deconvolution data. Sample shown in red and healthy donor samples in gray. Z-scores calculated relative to healthy donor sample distributions. Endothelium refers to endothelial cell, heart to cardiac muscle cells, liver to hepatocyte, lung to club cell and type I pneumocyte, and neuronal to Schwann cell cfRNA cell type of origin fractions.

189 **cfRNA profiling as a clinical decision support tool.** We next simulated a cfRNA “report” for each  
 190 patient in the multiclass test set (**Supplementary File 7**) and discuss four patient cases in detail to  
 191 illustrate how cfRNA can be integrated to support clinical decision making. Each report includes



192 diagnostic predictions from the multiclass classification algorithm and predicted organ involvement levels  
193 as z-scores of CTO fractions in samples from patients with disease compared to healthy patients.

194 The first case is a 10–15-year-old male with very-high-risk B-cell acute lymphoblastic leukemia in  
195 delayed intensification chemotherapy who presented with a 1-day history of fever (**Figure 5B**). His  
196 laboratory assessments were notable for leukopenia, anemia, thrombocytopenia, elevated C-reactive  
197 protein (CRP), mildly elevated alanine transaminase, normal kidney function, and developed hypotension  
198 following admission. He had known household contacts with COVID-19, and his nasopharyngeal PCR  
199 was positive for SARS-CoV-2, but he did not develop respiratory symptoms. The final diagnosis was  
200 COVID-19 in an immunocompromised host. Despite the immunocompromised state of the patient and  
201 lack of respiratory symptoms common in COVID-19, the cfRNA algorithm correctly identified the patient  
202 as having a viral infection. Furthermore, the elevated levels of liver and heart derived cfRNA are  
203 interesting given the clinical manifestations of mildly elevated ALT and hypotension.

204 In the second case, the cfRNA model correctly predicted the patient as having MIS-C, but the  
205 classifier score for MIS-C and viral infection were very comparable, producing what could be considered  
206 a “borderline” result (**Figure 5C**). The patient is a 0–5-year-old male who presented with a 1-week history  
207 of abdominal pain, vomiting, and watery diarrhea, with subsequent development of fever, conjunctival  
208 injection and swelling of the hands and feet. The patient had normal white blood cell counts (WBC), with  
209 lymphopenia, thrombocytopenia, hyponatremia, acidosis, and elevated brain natriuretic peptide (BNP),  
210 C-reactive protein (CRP) and erythrocyte sedimentation rate (ESR), with normal liver and kidney function.  
211 He was SARS-CoV-2 PCR negative but SARS-CoV-2 immunoglobulin G positive. No additional viral  
212 testing was performed. His course was complicated by hypotension, but an echocardiogram was grossly  
213 normal. The final diagnosis was MIS-C, which was the top prediction from the cfRNA model. Interestingly,  
214 viral infection was the second most probable diagnosis from the cfRNA model. MIS-C is derived from a  
215 viral infection (SARS-CoV-2), and MIS-C and viral infection can be difficult to differentiate. There are at  
216 least 3 possible interpretations for these results: (1) the model is correctly identifying MIS-C along with a  
217 viral signature from SARS-CoV-2, (2) the patient is misdiagnosed with MIS-C and has only a viral  
218 infection, or (3) the detection of a viral signature is due to errors in the model. We also observed elevated  
219 liver cfRNA despite the patient having normal liver enzymes along with elevated heart tissue derived  
220 cfRNA, which is consistent with the elevated BNP.

221 In the third case, the model incorrectly predicted the diagnosis in an immunocompromised patient,  
222 but the correct diagnosis of bacterial infection was the second most likely prediction (**Figure 5D**). The  
223 patient is a 0-5-year-old male with a history of Wiskott-Aldrich syndrome requiring a bone marrow  
224 transplant. He was admitted with fever, fatigue, pancytopenia with severe thrombocytopenia, elevated  
225 CRP, and blood cultures positive for *Klebsiella pneumoniae*. The source of the infection is unknown but  
226 was believed to be secondary to bacterial gut translocation or central line infection. His pancytopenia and  
227 thrombocytopenia were attributed to his history of Wiskott-Aldrich syndrome. He was treated with  
228 cefepime and responded clinically to a full course of antibiotics. The final diagnosis was bacterial  
229 infection, which was the second ranked prediction by the cfRNA model. Interestingly, the patient  
230 presented with some symptoms characteristic of MIS-C, the apparently erroneous top prediction made  
231 by the cfRNA model; specifically with severe thrombocytopenia and elevated CRP, which are part of the  
232 MIS-C case definition<sup>25</sup>.

233 In the last case, the model predicted an incorrect diagnosis, but the cfRNA organ damage scores  
234 in the patient were consistent with what was observed clinically (**Figure 5E**). The patient is a 0-5-year-  
235 old female presenting with 2 months of fever, progressive fatigue, and diffuse muscle pain. On admission,

236 the patient was found to have peripheral leukocytosis (WBC count of 24.6) with >20% blasts, relative  
237 neutropenia and tumor lysis syndrome. The leukemia was subtyped as B-cell acute lymphocytic leukemia  
238 by flow cytometry. The patient was positive for methicillin-sensitive *Staphylococcus aureus* (MSSA)  
239 bacteremia with 1 of 1 culture positive. Hazy opacities observed on the chest x-ray raised concerns about  
240 pulmonary edema, while the echocardiogram indicated coronary dilation and cardiac hypertrophy which  
241 was attributed to chronic anemia. The patient responded to a full course of antibiotics, and the final  
242 diagnosis was MSSA bacterial infection, which the cfRNA model did not correctly predict. However,  
243 despite the incorrect diagnosis, organ damage scores clearly show elevated levels of heart derived  
244 cfRNA, consistent with the abnormal cardiac hypertrophy observed in this patient.

245

## 246 **DISCUSSION**

247 In this work, we established a proof-of-concept that cfRNA can be used to differentiate among  
248 inflammatory syndromes in children across both one-on-one and multiclass comparisons. Interest in the  
249 use of blood-borne nucleic acids for diagnosis of inflammatory conditions is driven by continual  
250 advancements in sequencing technology, the ease of drawing blood, and the critical need for better  
251 diagnostic tools. Coote et al. recently introduced whole blood RNA multi-classification models for pediatric  
252 illnesses, distinguishing 6 broad and 18 granular categories. However, whole blood RNA biosignatures  
253 are primarily derived from immune cells, and do not provide information on cell, tissue, and organ  
254 damage. In contrast, cfRNA signals derive from dead and dying cells in blood and tissues, thus providing  
255 insight into immune dynamics and underlying tissue involvement<sup>8-10,17-19,24,26,27</sup>. Kalantar et al. used a  
256 combination of cfRNA, cfDNA, and whole blood RNA for differentiating sepsis and other causes of critical  
257 illness<sup>5</sup>. However, the approach used by Kalantar et al. lacks consolidation into a unified multi-class  
258 model, requires the use of two different analytes (whole blood and plasma), and did not analyze cfRNA  
259 to characterize tissue and organ injury. Our work expands on these recent studies by presenting a  
260 multiclass framework using only cfRNA to differentiate, diagnose, and better characterize inflammatory  
261 syndromes in children.

262 The data from this study provides opportunities to better understand cfRNA profiles in health and  
263 disease. We observed cfRNA signatures that are common to all inflammatory conditions, including  
264 elevated histone related RNAs, neutrophil extracellular trap (NET) components MPO and ELANE<sup>28</sup>, as  
265 well as immune markers ILF2, IFI16, CD53, and CXCR2. NETs are composed of DNA, histones, and  
266 other proteins, and act to trap and neutralize pathogens while minimizing host cell damage<sup>29</sup>. Our  
267 observation of cfRNA signatures of NET formation is consistent with the reported role of NETs in KD and  
268 MIS-C<sup>30,31</sup>.

269 To test the ability of cfRNA profiles to distinguish among pediatric inflammatory syndromes, we  
270 compared cfRNA profiles for patients diagnosed with KD and MIS-C, conditions that are very similar  
271 clinically. There is a need for improved molecular tools to discriminate between these two conditions as  
272 exemplified by a recent study from Day-Lewis et al. which reported significant overlap in signs and  
273 symptoms between KD and MIS-C (based on the 2023 case definition), with an estimated false positive  
274 rate of 8%<sup>25</sup>. The cfRNA signature presented here has near-perfect accuracy and therefore has high  
275 potential for translation into a useful clinical molecular test. Furthermore, the identified gene signature  
276 provides new mechanistic insight into KD and MIS-C. For example, two genes included in the machine  
277 learning signature are EEF2, elevated in KD, and FKBP5, elevated in MIS-C. EEF2 is an elongation  
278 factor that has previously been implicated in senescence<sup>32</sup> and exposure to bacterial toxins<sup>33</sup>. This is  
279 compelling given that it has been suggested that an inhaled toxin or pathogen may be one of the triggers

280 of KD. FKBP5 is a member of the immunophilin family and has been implicated in immune-stress  
281 response and other cellular processes in the brain and peripheral nervous system<sup>34</sup>. This observation is  
282 relevant given the neuronal involvement reported for MIS-C<sup>35</sup>. Beyond KD and MIS-C, we show that  
283 cfRNA can similarly differentiate between COVID-19 and other viral infections, opening the door for future  
284 studies to create more granular classification algorithms that can differentiate among pathogens at the  
285 species level based on cfRNA host profiling.

286 The case studies focusing on both correct and incorrect predictions made by the classification  
287 model highlight the potential utility of cfRNA profiling in the clinical setting. They also underscore the  
288 importance of inclusion of as many “real-life” cases as possible during the development of classification  
289 models to maximize the specificity of their predictions. In particular, immunocompromised patients exhibit  
290 altered immune and tissue/organ damage responses, and more data will likely need to be collected and  
291 classifications models tailored to optimize performance of cfRNA profiling in this subgroup. However, we  
292 are encouraged that our results suggest that cell, tissue, and/or organ injury based on cfRNA levels can  
293 be detected even for those cases for which the model disease classification is incorrect.

294 Here we present the first multiclass model for differential diagnosis of inflammatory syndromes  
295 using plasma cfRNA. The final model includes just 109 genes, suggesting that translating the model to a  
296 multiplexed qRT-PCR based platform with a rapid turnaround time of a few hours is likely practical.  
297 Furthermore, one of the most compelling aspects of cfRNA profiling is the ability to quantify the extent of  
298 organ involvement. We demonstrate this concept with simultaneous quantification of injury to multiple  
299 organ systems (liver, cardiac, lung, endothelium). Available clinical tests for evaluating tissue injury (for  
300 example, alanine aminotransferase levels in the liver) enabled us to confirm the accuracy and cfRNA-  
301 based detection of organ injury. Even more compelling is the potential clinical utility of cfRNA in  
302 quantifying injury to tissues where current tests are not available or lack adequate sensitivity, such as  
303 endothelial and neuronal injury. In the future, simultaneous diagnosis and characterization of tissue injury  
304 may be critical in informing clinicians on the optimal management and treatment of their patients with  
305 inflammatory syndromes of unknown etiology.

306

## 307 **METHODS**

308 **Ethics Statement.** The University of California, San Francisco (UCSF) Institutional Review Board (IRB)  
309 (#21-33403), San Francisco, CA; Emory University IRB (STUDY00000723), Atlanta, GA; Children’s  
310 National Medical Center IRB (Pro00010632), Washington, DC; and Cornell University IRB for Human  
311 Participants (2012010003), New York, NY each approved the protocols for this study. All samples and  
312 patient information were de-identified for analysis and shared with collaborating institutions. At Emory  
313 University, the IRB approved protocol was a prospective enrollment study under which parents provided  
314 consent and children assent as appropriate for age. At Children’s National Medical Center and UCSF,  
315 the IRB protocols were “no subject contact” sample biobanking protocols under which consent was not  
316 obtained and data was extracted from medical charts. At University of California, San Diego (UCSD),  
317 the IRB reviewed and approved collection and sharing of samples and data (IRB #140220). Signed  
318 consent and assent were provided by the parent(s) and pediatric patient, respectively.

319

320 **Sample Acquisition.** Samples were acquired<sup>17</sup> from UCSF as previously described. Briefly, hospitalized  
321 pediatric patients were identified as having COVID-19 by testing positive with SARS-CoV-2 real-time  
322 PCR (RT-PCR). Residual whole blood samples were collected in EDTA lavender top tubes and diluted  
323 1:1: in DNA/RNA shield (Zymo Research). The remaining blood was centrifuged at 2500 rpm for 15 min

324 and the available plasma was retained. All samples were stored at -80°C freezer until used. Samples  
325 were acquired from Emory and Children’s Healthcare of Atlanta as previously described<sup>17</sup>. Briefly,  
326 pediatric patients were classified as having COVID-19 by SARS-CoV-2 RT-PCR and as having MIS-C if  
327 they met the CDC case definition. Controls were healthy outpatients with no known history of COVID-19  
328 who volunteered for specimen collection. Whole blood was collected in EDTA lavender top tubes and  
329 aliquoted for plasma extraction via centrifugation at 2500 rpm for 15 min. Samples were stored at -80°C  
330 and shipped on dry ice to either UCSF or Cornell for analysis. Samples were acquired from UCSD prior  
331 to any treatment in all subjects in EDTA lavender top tubes and centrifuged at 2,000g for 10 minutes.  
332 Plasma was collected and stored at -80F. Samples were acquired from Children’s National Hospital as  
333 previously described<sup>17</sup>. Briefly, pediatric patients were classified as having MIS-C if they met the CDC  
334 case definition. Whole blood samples were collected and centrifuged at 1300 xG for 5 minutes at room  
335 temperature. Plasma was aliquoted into a cryovial and frozen at -80°C.

336  
337 **Clinical Data.** Patients were stratified as previously described<sup>17</sup>. For the purposes of this study, patients  
338 were classified as having MIS-C by multidisciplinary teams that adjudicated whether a patient met the  
339 CDC case definition of MIS-C. COVID-19 was defined as any patient with PCR-confirmed SARS-CoV-2  
340 infection within the preceding 14 days who did not also meet the MIS-C case definition. Kawasaki Disease  
341 patients at UCSD met the AHA definition for complete or incomplete KD. Viral and bacterial infection  
342 patients enrolled at UCSD were adjudicated and a final diagnosis assigned by the research team (one  
343 ED clinician and one pediatric infectious disease expert) 2-3 months after the acute illness to allow time  
344 for serologies, recurrence, and clinical recovery to be assessed. Patients with a self-limited illness that  
345 resolved without treatment and for whom viral studies were either negative or not done were classified  
346 as having a “viral syndrome”. Clinical data was abstracted from the medical record and submitted into a  
347 REDCap databases housed at UCSF or UCSD.

348  
349 **Sample processing and sequencing.** Samples were processed as described previously<sup>17</sup>. Briefly,  
350 samples were received on dry ice, RNA was extracted, and libraries prepared and sequenced on a  
351 NextSeq or NovaSeq Illumina sequencer. Reads were trimmed to 61 bp and sequencing data was  
352 processed using a custom bioinformatics pipeline which included quality filtering and trimming, alignment  
353 to the human GRCh38 reference genome, and counting of gene features.

354  
355 **Sample quality filtering.** Using the sequencing data, quality control was performed by analyzing DNA  
356 contamination, rRNA contamination, total counts, and RNA degradation. DNA contamination was  
357 estimated by calculating the ratio of reads mapping to introns and exons. rRNA contamination was  
358 measured using SAMtools (v1.14). Total counts were calculated using featureCounts<sup>30</sup> (v2.0.0).  
359 Degradation was estimated by calculating the 5’-3’ bias using Qualimap<sup>31</sup> (v2.2.1). Samples were  
360 removed from analysis if either the intron to exon ratio was greater than 3, if a sample had less than  
361 75,000 total feature counts, or if the 5’-3’ read alignment ration bias was greater than 2.

362  
363 **Differential abundance analysis.** Gene transcript abundances were compared using a negative  
364 binomial model implemented using the DESeq2 R package<sup>36</sup>. Gene transcript base mean abundance,  
365 adjusted p-value, and log2 fold change were taken from the DESeq2 Results output. Gene transcript  
366 AUC was calculated using VST transformed counts and the pROC R package<sup>37</sup>.

367

368 **Sample partitioning.** Samples were partitioned for machine learning applications taking into  
369 consideration diagnosis, hospital of origin, and disease subclassification. MIS-C and COVID-19 samples  
370 were subclassified by severity, as previously defined<sup>17</sup>. KD samples were subclassified by phenotypic  
371 subclusters, as previously defined<sup>38</sup>. Non-COVID-19 Viral and bacterial infection samples were evenly  
372 partitioned by diagnosis and hospital of origin only.

373

374 **Machine learning: MIS-C vs KD.** Samples were partitioned into training, validation, and test sets at a  
375 ratio of 60:20:20, partitioning evenly based on factors such as diagnosis, severity/subgroup, and hospital  
376 of origin. Subsequently, differential abundance analysis was conducted on the training data. Genes were  
377 filtered based on specific criteria (adjusted p-value < 0.01, base mean abundance > 100, gene transcript  
378 AUC > 0.65, and absolute log2 fold change > 0.25) and the top 150 genes, as ordered by gene transcript  
379 AUC, were selected as inputs for machine learning analyses.

380 Raw counts for the training, validation, and test sets were individually normalized using variance  
381 stabilizing transformation, and subsets were created based on the chosen transcript features. Fourteen  
382 machine learning classification algorithms were employed using the R package Caret  
383 (10.18637/jss.v028.i05), including generalized linear models with Ridge and LASSO feature selection  
384 (GLMNETRIDGE and GLMNETLASSO), support vector machines with linear and radial basis function  
385 kernel (SVMlin and SVMRAD), random forest (RF), random forest ExtraTrees (EXTRATREES), neural  
386 networks (NNET), linear discriminant analysis (LDA), nearest shrunken centroids (PAM), C5.0 (C5), k-  
387 nearest neighbors (KNN), naive bayes (NB), CART (RPART), and generalized linear models (GLM).  
388 Training was performed using 5-fold cross-validation and grid search hyperparameter tuning.

389 For each model, classification score thresholds were determined using Youden's index on the  
390 training data. The trained models were then employed to predict labels for the validation set, and  
391 performance was assessed using the area under the receiver operating characteristic curve area under  
392 the curve (ROC-AUC). The model achieving the highest AUC on the validation set was selected as the  
393 final model and subsequently applied to the test set, which had not been utilized in any phase of model  
394 training or selection. Prediction on both the validation and test sets utilized the Youden's index threshold  
395 derived from the training set.

396

397 **Machine learning: Multi-Classification.** Samples were partitioned into train and test sets at a ratio of  
398 70:30, considering factors such as diagnosis, severity/subgroup, and hospital of origin. One-vs-one  
399 GLMNET LASSO models were trained for each pairwise comparison of samples groups (eg. KD vs Viral  
400 Infection) using the top 150 significant features (adjusted p-value <0.05, base mean abundance > 50,  
401 and absolute log2 fold change > 1) ordered by gene transcript AUC calculated using the training data.  
402 Trained models were used to calculate classification scores for all samples in the train and test data set,  
403 resulting in six classification scores for each sample. Classification scores were used to train a multiclass  
404 Random Forest algorithm which assigned probability scores for each condition. The final condition with  
405 the highest probability score from the Random Forest was assigned as the prediction for each sample.

406

407 **CTO analysis.** Cell type deconvolution was performed using BayesPrism (v1.1)<sup>39</sup> with the Tabula  
408 Sapiens single-cell RNA-seq atlas (Release 1)<sup>27</sup> as a reference. Cells from the Tabula Sapiens atlas were  
409 grouped as previously described in Vorperian et al.<sup>24</sup>. Cell types with more than 100,000 unique molecular  
410 identifiers (UMIs) were included in the reference and subsampled to 300 cells using ScanPy (v1.8.1)<sup>40</sup>.  
411 Deconvolution values were scaled from 0-1 for each cell type and medians calculated for plotting.

412 For liver damage assessment, samples were separated as having normal or high ALT  
413 measurements (normal: ALT < 40 IU/L, high: ALT > 100 IU/L). The ALT measurements were taken from  
414 the same blood draw as the plasma for cfRNA processing. For cardiac function, patients were categorized  
415 as abnormal if they had abnormal EKG/ECG and/or echocardiogram results. EKG/ECG and  
416 echocardiogram results were categorized as abnormal in the context of the patient narrative and final  
417 interpretation of the studies. For endothelial damage, samples were separated as either having KD/MIS-  
418 C or bacterial/viral infection. COVID-19 patients were determined to have moderate or severe disease  
419 using the following criteria:

420 *Moderate:* The patient must have been hospitalized due to COVID-19 respiratory disease and/or  
421 any systemic/non-respiratory symptoms attributed to COVID-19 (e.g., neonatal fever, dehydration, new  
422 diagnosis diabetes, acute appendicitis, necrosis of extremities, diarrhea, encephalopathy, renal  
423 insufficiency, mild coagulation abnormalities, etc.).

424 *Severe:* The patient must have been hospitalized for COVID-19 with either high-flow oxygen  
425 requirement (high-flow nasal cannula (NC), continuous positive airway pressure (CPAP), bilevel positive  
426 airway pressure (BIPAP), intubation with mechanical ventilation, or extracorporeal membranous  
427 oxygenation (ECMO)) and/or evidence of end-organ failure (acute renal failure requiring dialysis,  
428 coagulation abnormalities resulting in bleeding or stroke, diabetic ketoacidosis (DKA), hemodynamic  
429 instability requiring vasopressors) and/or dying from COVID-19. These patients were almost always  
430 admitted to the ICU.

431  
432 **Machine learning: COVID-19 vs non-COVID19 viral infection.** Training was done using the same  
433 method as the KD vs MIS-C model. Briefly, viral infection samples were partitioned into train and test  
434 sets at a ratio of 70:30, considering factors such as COVID-19 status, severity/subgroup, and hospital of  
435 origin. Subsequently, differential abundance analysis was conducted on the training data. Genes were  
436 filtered based on specific criteria (adjusted p-value < 0.01, base mean abundance > 100, gene transcript  
437 AUC > 0.65, and absolute log<sub>2</sub> fold change > 0.25) and the top 150 genes, as ordered by gene transcript  
438 AUC, were selected as inputs for machine learning analysis.

439 Raw counts for the train and test sets were individually normalized using variance stabilizing  
440 transformation, and subsets were created based on the chosen transcript features. A GLMNET with  
441 LASSO regression was trained on the training set using 5-fold cross-validation and grid search  
442 hyperparameter tuning. Classification score thresholds were determined using Youden's index on the  
443 training data. The trained models were then employed to predict labels for the test set using the Youden's  
444 index threshold derived from the training set.

445  
446 **Quantification and statistical analyses.** The programming language R (v4.1.0) was utilized for all  
447 statistical analyses. Statistical significance was assessed through two-sided Wilcoxon signed-rank tests  
448 and Mann-Whitney U tests, unless specified otherwise. Machine learning algorithms were trained using  
449 the Caret R package and pipelines were run using the Snakemake workflow management system<sup>41,42</sup>. In  
450 boxplots, boxes denote the 25th and 75th percentiles, the band within the box signifies the median, and  
451 whiskers extend to 1.5 times the interquartile range of the hinge. The alignment of all sequencing data  
452 was performed against the GRCh38 Gencode v38 Primary Assembly, with feature counting conducted  
453 using the GRCh38 Gencode v38 Primary Assembly Annotation<sup>43</sup>.

454  
455 **DATA AVAILABILITY**

456 Raw sequencing data in this study cannot be deposited in a public repository due to patient privacy  
457 concerns and lack of consent for a subset of the patient samples. Instead, de-identified RNA-seq count  
458 matrices have been uploaded to the NCBI (National Center for Biotechnology Information) GEO (Gene  
459 Expression Omnibus) database and will be publicly available upon publication (GSE255555). All code  
460 has been deposited on GitHub and will be available upon publication.

461

#### 462 **AUTHOR CONTRIBUTIONS**

463 C.J.L., R.L.D, C.A.R, J.C.B, C.Y.C, and I.D.V conceived and designed the study. C.J.L. A.B.,J.L., and  
464 E.B. performed sequencing experiments. W.S, J.N, M.E.W, M.O, M.A.G, J.C, H.H, C.S, A.T, M.D, R.L.D,  
465 C.A.R, J.C.B, C.Y.C, PEMKDRG, and the CHARMS Study Group identified and collected patient samples  
466 and clinical metadata. V.S, A.S.G, A.B, W.S, H.W, and J.K provided input for data analysis. C.J.L., A.B.,  
467 R.L.D., C.A.R., J.C.B, C.Y.C., and I.D.V. analyzed sequencing data. M.D., R.L.D., C.A.R., J.C.B, C.Y.C.,  
468 and I.D.V. supervised the study. C.J.L., C.Y.C., and I.D.V. wrote the manuscript and prepared the figures.  
469 All authors read and edited the manuscript and agreed to its contents.

470

#### 471 **COMPETING INTERESTS**

472 C.J.L and I.D.V are inventors on submitted patents pertaining to cell-free nucleic acids (US patent  
473 applications 63/237,367 and 63/429,733). I.D.V. is a member of the Scientific Advisory Board of Karius  
474 Inc., Kanvas Biosciences and GenDX. I.D.V. is listed as an inventor on submitted patents pertaining to  
475 cell-free nucleic acids (US patent applications 63/237,367, 63/056,249, 63/015,095, 16/500,929, 41614P-  
476 10551-01-US) and receives consulting fees from Eurofins Viracor. C.A.R. has received institutional  
477 support from ModernaTX, Inc., Pfizer Inc., BioFire Inc., GSK plc, MedImmune, Micron Technology Inc.,  
478 Janssen Pharmaceuticals, Merck & Co., Inc., Novavax, PaxVax, Regeneron, and Sanofi Pasteur. She is  
479 co-inventor of patented RSV vaccine technology which has been licensed to Meissa Vaccines, Inc.  
480 C.Y.C. receives grant funding for research unrelated to this work from the Bay Area Lyme Disease  
481 Foundation, the Chan-Zuckerberg Biohub, and Abbott Laboratories, Inc. C.y.C. is on the scientific  
482 advisory board for Mammoth Biosciences, Poppy Health, BiomeSense, BioMeme, FlightPath  
483 Biosciences, and Delve Bio, and is a co-founder of Delve Bio.

484

#### 485 **ACKNOWLEDGEMENTS**

486 We would like to acknowledge staff members at the UCSF Clinical Laboratories and the UCSF Clinical  
487 Microbiology Laboratories for their help in identifying and retrieving patient whole blood samples. We  
488 thank the Cornell Genomics Center and the UCSF Center for Advanced Technology for help with  
489 sequencing libraries. At Emory, we thank Nadine Baida, Amrita Banerjee, Julia Bartol, Kushmita Bhakta,  
490 Caroline Ciric, Khalel De Castro, Khadijah Francois, Theda Gibson, Laila Hussaini, Grace Li, Wensheng  
491 Li, Austin Lu, Lisa Macoy, Clair Martin, Molly Morrison, Amy Muchinsky, Heather Nurse, Maria A. Perez,  
492 Etza Peters, Brianna Rice, Anna Siaw-Anim, Kathy Stephens, Elizabeth Grace Taylor, Ashley Tippet,  
493 Katie Zaks, and the Children's Healthcare of Atlanta Research Laboratory for their contributions to  
494 specimen and data collection. At UCSD we thank the members of the PEMKDRG: Lukas Austin-Page,  
495 MD, Amy Bryl, MD, Joelle Donofrio-Ödmann, MD, Atim Ekpenyong, MD, David Gutglass, MD, Scott  
496 Herskovitz, MD, Paul Ishimine, MD, John Kanegaye, MD, Margaret Nguyen, MD, Mylinh Nguyen, MD,  
497 Kristy Schwartz, MD, Stacey Ulrich, MD, Tatyana Vayngortin, MD, and Elise Zimmerman, MD. We thank  
498 the members of the CHARMS study that contributed samples: Jocelyn Ang, MD, Margalit Rosenkranz,  
499 MD, Joseph Bochini, MD, Michelle Sykes, MD, Lerraughn Morgan, MD, Laura D'Addese, MD, and Maria

500 Pilar Gutierrez, MD. We thank the patients and their families for their help to further our understanding  
501 of pediatric inflammatory conditions. This work was funded by National Institutes of Health (NIH) /  
502 National Institute of Child Health and Human Development (NICHD) grants R61HD105618,  
503 R33HD105618, and R33HD105593 (R.D., C.A.R., I.D.V., JCB, AHT, CS, and C.Y.C.). The funder had  
504 no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.



505 **REFERENCES**

- 506 1. Hisamuddin, E., Hisam, A., Wahid, S. & Raza, G. Validity of C-reactive protein (CRP) for diagnosis  
507 of neonatal sepsis. *Pak. J. Med. Sci.* **31**, 527–531 (2015).
- 508 2. Tan, M., Lu, Y., Jiang, H. & Zhang, L. The diagnostic accuracy of procalcitonin and C-reactive  
509 protein for sepsis: A systematic review and meta-analysis. *J. Cell. Biochem.* **120**, 5852–5859  
510 (2019).
- 511 3. Servellita, V. *et al.* A diagnostic classifier for gene expression-based identification of early Lyme  
512 disease. *Commun. Med.* **2**, 92 (2022).
- 513 4. Habgood-Coote, D. *et al.* Diagnosis of childhood febrile illness using a multi-class blood RNA  
514 molecular signature. *Med N. Y. N* **4**, 635-654.e5 (2023).
- 515 5. Kalantar, K. L. *et al.* Integrated host-microbe plasma metagenomics for sepsis diagnosis in a  
516 prospective cohort of critically ill adults. *Nat. Microbiol.* **7**, 1805–1816 (2022).
- 517 6. Jackson, H. R. *et al.* Diagnosis of Multisystem Inflammatory Syndrome in Children by a Whole-  
518 Blood Transcriptional Signature. *J. Pediatr. Infect. Dis. Soc.* **12**, 322–331 (2023).
- 519 7. Loy, Conor J., Ahmann, Lauren, De Vlaminc, I. & Gu, W. Liquid Biopsy Based on Cell-Free DNA  
520 and RNA. (2024) doi:<https://doi.org/10.1146/annurev-bioeng-110222-111259>.
- 521 8. Ngo, T. T. M. *et al.* Noninvasive blood tests for fetal development predict gestational age and  
522 preterm delivery. *Science* **360**, 1133–1136 (2018).
- 523 9. Moufarrej, M. N. *et al.* Early prediction of preeclampsia in pregnancy with cell-free RNA. *Nature*  
524 **602**, 689–694 (2022).
- 525 10. Larson, M. H. *et al.* A comprehensive characterization of the cell-free transcriptome reveals tissue-  
526 and subtype-specific biomarkers for cancer detection. *Nat. Commun.* **12**, 2357 (2021).
- 527 11. Mathios, D. *et al.* Detection and characterization of lung cancer using cell-free DNA fragmentomes.  
528 *Nat. Commun.* **12**, 5060 (2021).
- 529 12. Cheng, A. P. *et al.* Cell-free DNA profiling informs all major complications of hematopoietic cell  
530 transplantation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2113476118 (2022).

- 531 13. De Vlaminck, I. *et al.* Circulating cell-free DNA enables noninvasive diagnosis of heart transplant  
532 rejection. *Sci. Transl. Med.* **6**, 241ra77 (2014).
- 533 14. Cheng, A. P. *et al.* A cell-free DNA metagenomic sequencing assay that integrates the host injury  
534 response to infection. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 18738–18744 (2019).
- 535 15. Burnham, P. *et al.* Urinary cell-free DNA is a versatile analyte for monitoring infections of the urinary  
536 tract. *Nat. Commun.* **9**, 2412 (2018).
- 537 16. Cheng, A. P. *et al.* Cell-free DNA tissues of origin by methylation profiling reveals significant cell,  
538 tissue, and organ-specific injury related to COVID-19 severity. *Med N. Y. N* **2**, 411-422.e5 (2021).
- 539 17. Loy, C. J. *et al.* Nucleic acid biomarkers of immune response and cell and tissue damage in  
540 children with COVID-19 and MIS-C. *Cell Rep. Med.* **4**, 101034 (2023).
- 541 18. Wang, Y. *et al.* Plasma cell-free RNA characteristics in COVID-19 patients. *Genome Res.* **32**, 228–  
542 241 (2022).
- 543 19. Chang, A. *et al.* Circulating Cell-Free RNA in Blood as a Host Response Biomarker for the  
544 Detection of Tuberculosis. *MedRxiv Prepr. Serv. Health Sci.* 2023.01.11.23284433 (2023)  
545 doi:10.1101/2023.01.11.23284433.
- 546 20. De Vlaminck, I. *et al.* Temporal response of the human virome to immunosuppression and antiviral  
547 therapy. *Cell* **155**, 1178–1187 (2013).
- 548 21. De Vlaminck, I. *et al.* Noninvasive monitoring of infection and rejection after lung transplantation.  
549 *Proc. Natl. Acad. Sci. U. S. A.* **112**, 13336–13341 (2015).
- 550 22. Burnham, P. *et al.* Separating the signal from the noise in metagenomic cell-free DNA sequencing.  
551 *Microbiome* **8**, 18 (2020).
- 552 23. Mzava, O. *et al.* A metagenomic DNA sequencing assay that is robust against environmental DNA  
553 contamination. *Nat. Commun.* **13**, 4197 (2022).
- 554 24. Vorperian, S. K., Moufarrej, M. N., Tabula Sapiens Consortium & Quake, S. R. Cell types of origin  
555 of the cell-free transcriptome. *Nat. Biotechnol.* **40**, 855–861 (2022).

- 556 25. Day-Lewis, M. *et al.* Updated Case Definition of MIS-C and Implications for Clinical Care. *Pediatrics*  
557 **153**, e2023063259 (2024).
- 558 26. Toden, S. *et al.* Noninvasive characterization of Alzheimer's disease by circulating, cell-free  
559 messenger RNA next-generation sequencing. *Sci. Adv.* **6**, eabb1654 (2020).
- 560 27. Tabula Sapiens Consortium\* *et al.* The Tabula Sapiens: A multiple-organ, single-cell transcriptomic  
561 atlas of humans. *Science* **376**, eabl4896 (2022).
- 562 28. Rizo-Téllez, S. A., Sekheri, M. & Filep, J. G. Myeloperoxidase: Regulation of Neutrophil Function  
563 and Target for Therapy. *Antioxid. Basel Switz.* **11**, 2302 (2022).
- 564 29. Kaplan, M. J. & Radic, M. Neutrophil extracellular traps: double-edged swords of innate immunity.  
565 *J. Immunol. Baltim. Md 1950* **189**, 2689–2695 (2012).
- 566 30. Yoshida, Y. *et al.* Enhanced formation of neutrophil extracellular traps in Kawasaki disease. *Pediatr.*  
567 *Res.* **87**, 998–1004 (2020).
- 568 31. Boribong, B. P. *et al.* Neutrophil profiles of pediatric COVID-19 and multisystem inflammatory  
569 syndrome in children. *Cell Rep. Med.* **3**, 100848 (2022).
- 570 32. Liu, S., Matsui, T. S., Kang, N. & Deguchi, S. Analysis of senescence-responsive stress fiber  
571 proteome reveals reorganization of stress fibers mediated by elongation factor eEF2 in HFF-1 cells.  
572 *Mol. Biol. Cell* **33**, ar10 (2022).
- 573 33. Mateyak, M. K. & Kinzy, T. G. ADP-ribosylation of translation elongation factor 2 by diphtheria toxin  
574 in yeast inhibits translation and cell separation. *J. Biol. Chem.* **288**, 24647–24655 (2013).
- 575 34. Zannas, A. S., Wiechmann, T., Gassen, N. C. & Binder, E. B. Gene-Stress-Epigenetic Regulation of  
576 FKBP5: Clinical and Translational Implications. *Neuropsychopharmacol. Off. Publ. Am. Coll.*  
577 *Neuropsychopharmacol.* **41**, 261–274 (2016).
- 578 35. LaRovere, K. L. *et al.* Neurologic Involvement in Children and Adolescents Hospitalized in the  
579 United States for COVID-19 or Multisystem Inflammatory Syndrome. *JAMA Neurol.* **78**, 536–547  
580 (2021).

- 581 36. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-  
582 seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- 583 37. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves.  
584 *BMC Bioinformatics* **12**, 77 (2011).
- 585 38. Wang, H. *et al.* Subgroups of children with Kawasaki disease: a data-driven cluster analysis. *Lancet*  
586 *Child Adolesc. Health* **7**, 697–707 (2023).
- 587 39. Chu, T., Wang, Z., Pe'er, D. & Danko, C. G. Cell type and gene expression deconvolution with  
588 BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in  
589 oncology. *Nat. Cancer* **3**, 505–517 (2022).
- 590 40. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data  
591 analysis. *Genome Biol.* **19**, 15 (2018).
- 592 41. Köster, J. & Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinforma. Oxf.*  
593 *Engl.* **28**, 2520–2522 (2012).
- 594 42. Kuhn, M. Building Predictive Models in R Using the **caret** Package. *J. Stat. Softw.* **28**, (2008).
- 595 43. Frankish, A. *et al.* GENCODE reference annotation for the human and mouse genomes. *Nucleic*  
596 *Acids Res.* **47**, D766–D773 (2019).
- 597