communications medicine

A Nature Portfolio journal

Article

https://doi.org/10.1038/s43856-026-01489-7

Minimal correlation but complementary diagnostic utility for plasma cell-free RNA and proteins

Andrew Bliss[1], Conor J. Loy[1], Jihoon Kim[2], Chisato Shimizu[3], Joan S. Lenz[1], Emma Belcher[1], Adriana H. Tremoulet[3,4], Jane C. Burns[3,4] & Iwijn De Vlaminck[1]

Abstract

Background Proteins and RNA circulate in plasma and can offer insights into human physiology. Yet, despite their clinical importance, direct comparisons between these analytes remain unexplored.

Methods: Here, we measure and compare plasma cell-free RNA (cfRNA) and protein levels for 263 children diagnosed with inflammatory diseases, specifically either Kawasaki disease (KD) or Multisystem Inflammatory Syndrome in Children (MIS-C), by RNA sequencing (n = 108 KD and n = 47 MIS-C, mean age=4.2 years) and SomaScan proteomics (n = 70 KD and n = 101 MIS-C, mean age=6.8 years).

Results Here we show that cell-free RNA and protein levels are largely uncorrelated across samples (feature-by-sample correlation coefficient 0.052; median feature-level correlation coefficient 0.009). Nonetheless, machine learning models based on either modality distinguish KD from MIS-C with similar high accuracy (median area under the curve greater than 0.93). Analysis of KD subtypes reveals distinct cell-free RNA and protein signatures, with one group showing molecular similarity to MIS-C.

Conclusions These findings underscore the complementary nature of cell-free RNA and protein profiling and highlight the utility of integrating multiple plasma analytes to improve disease classification and deepen our understanding of complex inflammatory conditions.

Plain language summary

Proteins and cell-free RNA are molecules in the blood that can be measured in plasma, but it is unknown whether they contain overlapping or non-overlapping information about disease. We compared these two plasma analytes in children with Kawasaki disease and MIS-C, two serious pediatric inflammatory illnesses. We collected plasma samples and used laboratory tests to measure thousands of proteins and cell-free RNA and applied statistical and machine-learning methods to compare them. We found that protein and cell-free RNA levels were weakly related, but both can be used to differentiate Kawasaki Disease and MIS-C with high accuracy. These results suggest that combining several blood analytes may improve diagnosis and deepen understanding of inflammatory diseases.

Blood-borne bio-analytes such as cfRNA and proteins in plasma are increasingly used in diagnostic medicine and therapeutic development. Protein biomarkers are frequently used in translational research and clinical practice due to their biological origin, their direct or indirect roles as therapeutic targets, and their accessibility in plasma. Recent advances in high-throughput proteomics, including mass spectrometry and affinity-based assays, have expanded the quantifiable proteome to thousands of distinct proteins[1]. Despite this promise, high-throughput proteomics assays still face challenges. Mass spectrometry often requires extensive pre-fractionation and complicated sample preparation, which can limit throughput, increase costs, and introduce variability in large-scale analyses[1,2]. Affinity-based assays, including Olink and SomaScan, circumvent some of these issues but are constrained by antibody or aptamer specificity and remain expensive, limiting their widespread use in routine diagnostics and drug screening[1]. Recent comparisons of SomaScan and Olink have shown limited agreement[3] (median rho = 0.26), drawing into question the best platform for plasma-based proteomics. Moreover, the large dynamic range of protein concentrations in plasma adds further complexity for detection and quantification irrespective of the assay platform[4].

cfRNA is an emerging circulating bio-analyte that may address some of these challenges. cfRNA can be measured by RNA sequencing, which tends to be highly reproducible and benefits from a continued reduction in cost[5]. Its origins in both cell death and damage and active excretion make it highly dynamic and responsive to immune, infectious, and inflammatory disease[6–8]. However, cfRNA suffers from low biomass and is susceptible to preanalytical factors that impact reproducibility. It is also relatively new to clinical research and thus has not yet been implemented extensively at the

[1]Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA. [2]Department of Biomedical Informatics and Data Science, Yale School of Medicine, New Haven, CT, USA. [3]Department of Pediatrics, Kawasaki Disease Research Center, University of California San Diego, La Jolla, CA, USA. [4]Department of Pediatrics, Rady Children's Hospital-San Diego, San Diego, CA, USA. e-mail: vlaminck@cornell.edu

point of care[9]. Further evaluation is therefore needed to confirm its diagnostic performance relative to more established proteomic approaches.

To address these gaps, we investigated the agreement and diagnostic performance of high-throughput analysis of protein and RNA in plasma. Previous studies have compared protein and RNA profiles in other tissues[10,11], but no direct comparisons of cfRNA and protein levels in plasma, where these analytes have high clinical relevance, have been performed. We used cfRNA sequencing and proteomics profiling by SomaScan for two overlapping pediatric inflammatory syndromes—KD and MIS-C. Our aims were i) to compare the abundance and correlation of cfRNA and proteins measured in matched plasma samples, ii) to identify both unique and shared molecular signatures associated with each condition and its subtypes, and iii) to evaluate the performance of cfRNA- and protein-based models in classifying KD and MIS-C.

We find that cell-free RNA and protein levels measured by RNA sequencing and SomaLogic proteomics are largely uncorrelated across these samples. Nonetheless, despite this lack of correlation, we find that highly accurate classifiers can nonetheless be developed to elucidate disease pathogenesis and heterogeneity based on both modalities. This study underscores the potential utility of differing molecular assays to refine diagnostic and therapeutic strategies.

Methods

Ethics Statement

The Cornell University IRB for Human Participants (2012010003), New York, NY, approved the protocols for this study. All samples and patient information were deidentified for analysis and shared with collaborating institutions. At UCSD, the IRB reviewed and approved collection and sharing of samples and data (IRB #140220). Signed consent and assent as appropriate were provided by the parent(s) and pediatric patient, respectively.

Clinical cohort. We examined a cohort of 263 pediatric patients diagnosed with KD or MIS-C at Rady Children's Hospital–San Diego between 2006 and 2021. Two experienced KD clinicians (AHT and JCB) confirmed each diagnosis in accordance with American Heart Association criteria for both complete and incomplete KD[12], and clinical data were prospectively recorded at the time of diagnosis. Among these patients, plasma samples for 155 ($n = 108$ KD and $n = 47$ MIS-C) were analyzed by cfRNA sequencing. In a previous study[13], KD cases were further categorized into four subgroups ($n = 26$ Subgroup 1, $n = 26$ Subgroup 2, $n = 26$ Subgroup 3, $n = 23$ Subgroup 4). Additionally, we performed SomaScan proteomic analysis on samples from 171 patients ($n = 70$ KD and n = 101 MIS-C), which included 54 KD patients who had been categorized by subgroups ($n = 9$ Subgroup 1, $n = 13$ Subgroup 2, $n = 17$ Subgroup 3, $n = 15$ Subgroup 4). A total of 63 patient samples were included in both the cfRNA and proteomic analysis.

Sample collection and cell-free RNA sequencing. Plasma samples were collected at UCSD in lavender-top EDTA tubes prior to treatment. Samples were centrifuged at 2,000 g for 10 min and stored at −80 °C prior to shipping. cfRNA was extracted and cDNA libraries were prepared for sequencing. Sequencing data was analyzed as previously described in ref. 6.

Sample Quality Control. Quality control was performed using the sequencing data by assessing DNA contamination, rRNA contamination, total counts, and RNA degradation. DNA contamination was estimated based on the intron-to-exon read mapping ratio, while rRNA contamination was measured with SAMtools (v1.14). Total counts were computed using featureCounts30 (v2.0.0), and degradation was estimated by calculating the 5′-to-3′ read alignment bias with Qualimap31 (v2.2.1). Samples were excluded if the intron-to-exon ratio exceeded 3, if total counts were below 75,000, or if the 5′-to-3′ bias exceeded 2.

SomaScan analysis. Plasma samples from the same collection used for cfRNA were sent to SomaLogic for proteomic profiling using the SomaScan 7 K assay. Raw and ANML SMP normalized data using a pediatric normalization reference were provided by SomaLogic.

Differential abundance analysis. We assessed transcript abundance using a negative binomial model and applied variance stabilizing transformation (VST) via the DESeq2 R package[14].

Machine learning. The data were split into training and test sets at a 70:30 ratio over 250 iterations to minimize random seed variability. This splitting was performed for each sample subset (all data, matched samples, matched features, and both matched) for a total of 1000 iterations across all subsets. Class proportion was preserved across splits in all subsets to minimize the effects of class imbalance. Differential abundance analysis was then performed on the training data, filters by adjusted P-value < 0.05, base mean abundance > 50, and an absolute log2 fold change > 1. The test set was excluded from threshold determination to prevent overfitting.

Raw counts for the training, validation, and test sets were each normalized via variance-stabilizing transformation (VST), and transcript subsets were formed based on the selected features. Variance stabilization was performed using DESeq2, with the dispersion function from the training set applied to transform the validation and test sets. Model training incorporated fivefold cross-validation and grid search hyperparameter tuning using GLMNET with Lasso regularization[15], selected for its previous high performance with RNA[6], which reduced irrelevant feature coefficients to zero and selecting only the most informative transcripts, a critical feature in establishing an accessible clinical diagnostic tool.

For each model, classification score thresholds were established using Youden's index on the training data. The trained models were subsequently used to predict labels in the validation set, and performance was measured by the area under the ROC-AUC. Predictions on the test set employed the same Youden's index threshold derived from the training set.

To quantify model stability, classification performance using 5-fold cross-validation was computed. Pooled 95% confidence interval across seeds (computed as the 2.5th and 97.5th percentiles of the 250 seed-level mean accuracies), mean seed-level confidence interval bounds, and distribution of confidence interval widths for cross-validation are shown in Supplementary Table 3.

Cell type enrichment analysis. Cell enrichment analysis was performed via single sample gene set enrichment analysis (ssGSEA) using the GSVA package in R[16] (ssgsea.norm = FALSE, tau = 0.75, parallel.sz=10, min.sz=10), instead using Tabula Sapiens v2 as the reference gene set to annotate enrichments to cells rather than pathways. We performed pathway and network analysis using Ingenuity Pathway Analysis[17] (IPA, QIAGEN).

Ingenuity pathway analysis. We performed pathway and network analysis using Ingenuity Pathway Analysis (IPA, QIAGEN) separately on cfRNA and protein expression datasets. The top five enriched pathways were determined by successive removal of pathways sharing greater than 75% of annotated features with the top enriched pathway. Pathways annotated as cancer-specific were excluded from the results to avoid overrepresentation.

Statistical analyses. All statistical analyses were conducted using R (v4.1.0). We determined statistical significance with two-sided Wilcoxon signed-rank tests, two-sided Mann-Whitney U tests, and Welch two-sample t tests. In boxplots, the box edges represent the 25th and 75th percentiles, the central line indicates the median, and whiskers extend to 1.5 times the interquartile range beyond the hinge.

Results

Clinical cohort

We enrolled and analyzed plasma samples from 263 pediatric patients diagnosed with Kawasaki Disease (KD) or Multisystem Inflammatory Syndrome in Children (MIS-C) at Rady Children's Hospital San Diego (RCHSD) / University of California San Diego (UCSD). Of these, 155 samples were profiled by cfRNA sequencing (108 KD, 47 MIS-C) and 171 were analyzed using the SOMALogic SomaScan platform (70 KD, 101 MIS-C), with 63 samples profiled by both methods (24 KD, 39 MIS-C) (Fig. 1A). KD cases were further stratified into clinical subtypes according to classification criteria defined by Wang et al.[13]. All samples were collected during the acute phase of illness and prior to treatment with intravenous immunoglobulin (IVIG).

Lack of correlation of cfRNA and protein levels

To assess how plasma cfRNA and proteins compare as potential diagnostic biomarkers, we first compare the biomolecular features captured by each assay. We conduct untargeted cfRNA profiling by RNA sequencing and assign the reads to a total of 59,428 annotated features. For proteomics, we use the SOMALogic SomaScan affinity-based platform, which contains 7,310 targeted aptamers to 6,397 unique proteins[3]. Of these, 6,292 overlap with the cfRNA features (Fig. 1B). We evaluate global similarities by measuring the correlation of all molecules in paired samples. We observe no overall correlation (Pearson $r = 0.052$, Spearman rho = 0.131) when comparing counts per million (cfRNA) to adaptive normalization by maximum likelihood relative fluorescence units (ANML RFUs, protein, Fig. 1C). At the individual feature level, protein and RNA levels are uncorrelated (median Pearson $r = -0.008$), suggesting that cfRNA and protein capture largely distinct plasma biomolecular landscapes. Randomly shuffling sample identifiers and feature labels produces similar distributions centered around zero correlation (Fig. 1D).

Differentially abundant features and pathways

Given the observed lack of correlation between cfRNA and proteins in plasma, we asked how each analyte distinguishes between KD and MIS-C—two clinically overlapping yet pathologically distinct pediatric inflammatory conditions[18]. We perform differential abundance and pathway analyses for both cfRNA and proteins and compare their significant features (Methods). Overall, we identify 439 differentially abundant cfRNA transcripts and 63 differentially abundant proteins (BH adjusted p-value < 0.05 and |log2 fold change| > 1, Fig. 2A).

While cfRNA and proteins show similar ranges of fold changes (log_2: cfRNA = [−2.71, 1.68], protein = [−1.75, 2.75]), 68 cfRNA features—versus only one protein—exhibit a log_2 fold change greater than 2 (Fig. 2A). We observe more differentially abundant proteins upregulated in KD than cfRNA (protein $n = 40$, RNA $n = 27$, log fold change > 0), with only two features, COL10A1 and CCL5, upregulated across both datasets. The differences between cfRNA and protein become more pronounced in MIS-C, where 412 features were upregulated in cfRNA but only 23 in protein (Fig. 2B). Among these, 5 (IL1R1, LAP3, CXCL10, CNDP1, VSIG4) are upregulated in both analytes. Several of these (IL1R1, CXCL10, VSIG4) are known to contribute to immune activation and response[19–21], suggesting that despite limited overall overlap at the molecule level, cfRNA and proteins may converge on immune-related features.

To contextualize the differentially abundant cfRNA and proteins within larger-scale cellular functions, we perform pathway analyses using Ingenuity Pathway Analysis (IPA) (Methods, Fig. 2C). In KD, the top-enriched pathways in cfRNA highlight active immune and inflammatory responses, most notably hematoma resolution and cellular effects of Sildenafil[22], suggesting transcriptional regulation of acute vascular pathology. In contrast, protein-based pathways in KD include IL-4 signaling, myelination signaling, and collagen chain trimerization, which reflect downstream effects on inflammation, cellular function, and tissue

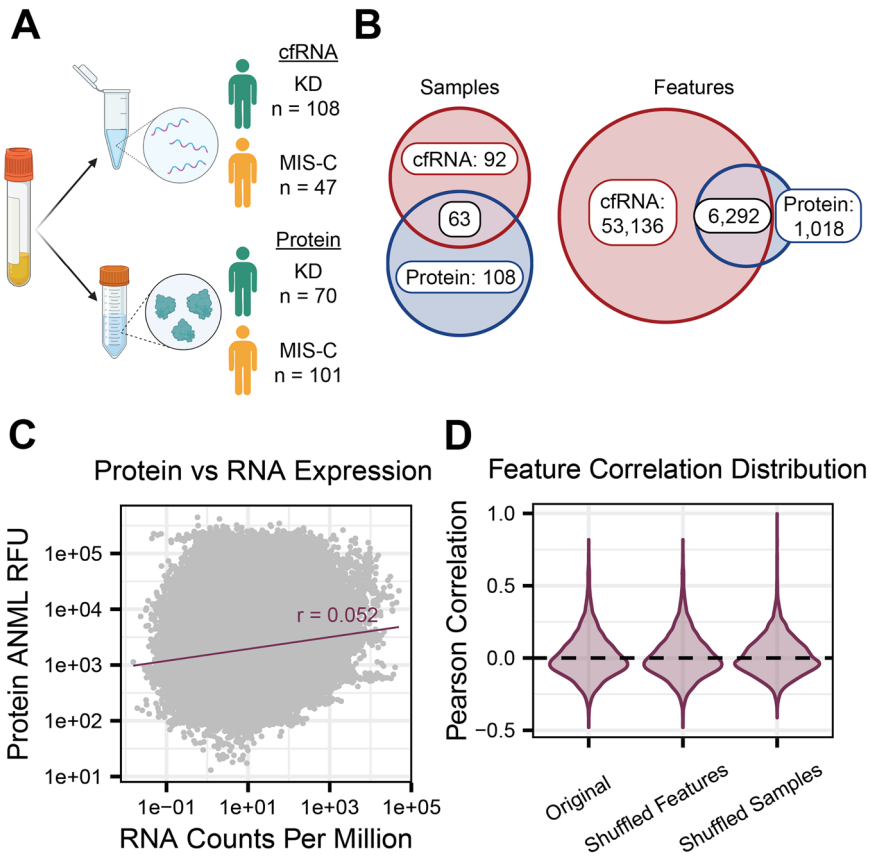

Fig. 1 | Overview of study analytes and samples. **A** cfRNA sequencing and SomaScan from plasma sample composition by condition and analyte. **B** Sample and feature overlap across cfRNA and protein assays. **C** Correlation of feature expression in 63 matched samples. **D** Distributions of individual feature correlations compared to correlations of randomized samples and features.

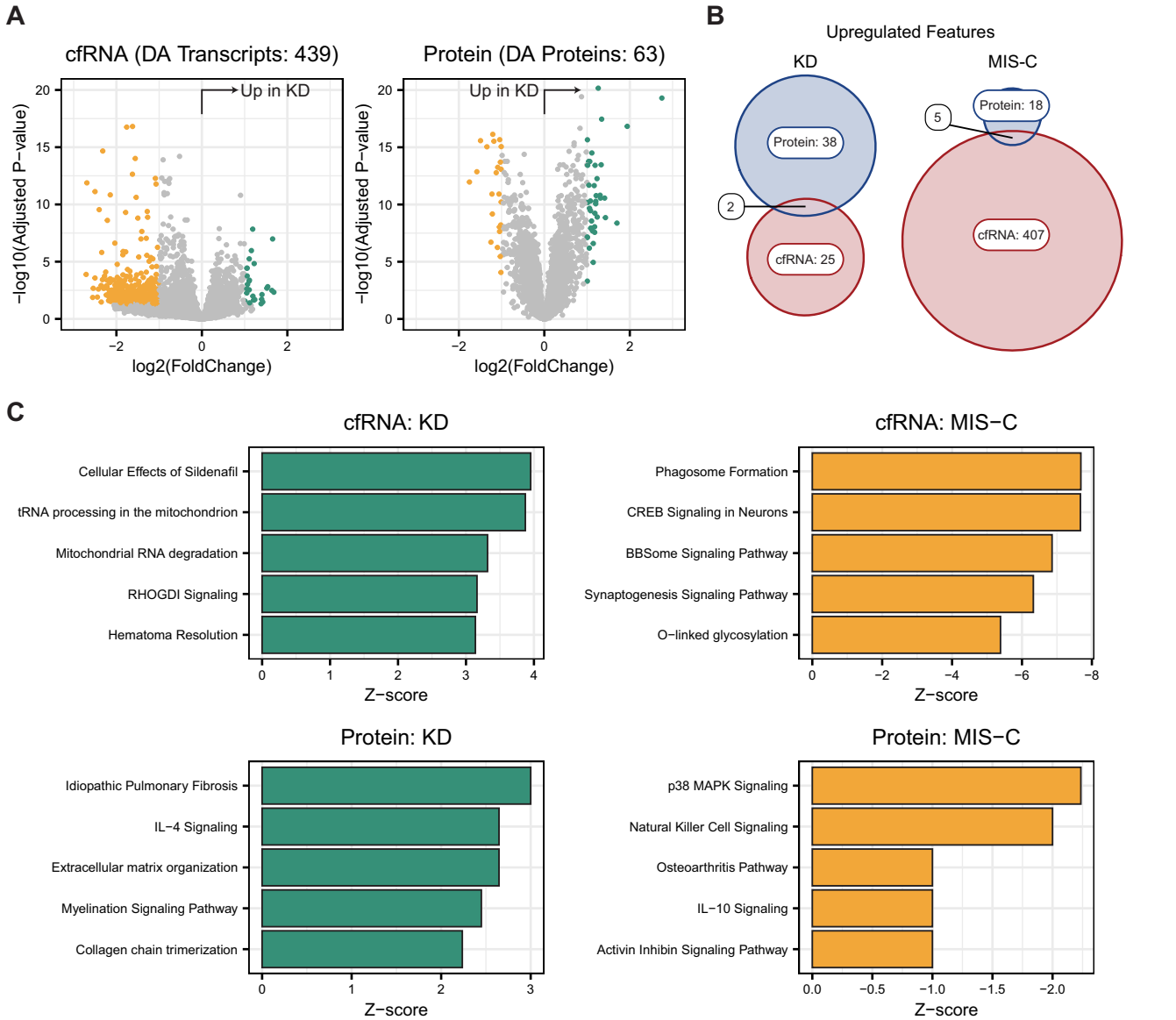

Fig. 2 | Comparisons in differential abundance and Ingenuity Pathway Analysis (IPA). A Volcano plot of differentially abundant features found in cfRNA and protein. **B** Overlap in differentially abundant features by condition found in cfRNA and protein. **C** Top 5 upregulated pathways found in IPA for each condition and analyte.

remodeling. In MIS-C, cfRNA analysis reveals disruptions in neuronal and intracellular signaling pathways (e.g., phagosome formation and synaptogenesis). Conversely, protein analysis points to inflammatory and metabolic dysregulation, such as p38 MAPK signaling, natural killer cell signaling, and IL-10 signaling, highlighting active immune processes not captured at the transcript level. These findings suggest that cfRNA and protein analyses provide complementary insight, with cfRNA capturing upstream regulatory processes and protein revealing downstream functionality. This analysis reinforces the complementary nature of both analytes to gain a more complete understanding of pathophysiology.

Both cfRNA and protein can classify KD and MIS-C

Because no molecular tests currently exist for KD or MIS-C, we evaluate whether plasma cfRNA, proteins, or both can be used to develop diagnostic signatures. To do so, we train machine learning classifiers using repeated cross-validation (Methods). Specifically, we train a generalized linear model with LASSO regression (GLMNETLasso) to optimize biomarker panel size via regularization. Across 250 iterations, we observe similar performance for models trained on cfRNA and protein profiles, with a median test AUC across iterations of 0.940 for cfRNA and 0.937 for protein ($p = 0.8602$, Fig. 3C). While performance is similar for both analytes, the cfRNA classifiers achieve this high performance using fewer features (Fig. 3D). To dissect whether these outcomes are influenced by sample composition or the number of available features, we repeat the cross-validation analysis for three scenarios: matched samples ($n = 63$), matched features ($n = 6{,}292$), and matched samples and features (Fig. 3E). Matching samples reduces performance for both cfRNA (median AUC = 0.896) and protein (median AUC = 0.870), highlighting the impact of sample size on classification accuracy ($p = 6.528\text{e-}06$). When only features are matched, proteomics (median AUC = 0.933) outperforms cfRNA (median AUC = 0.877, $p = 2.2\text{e-}16$), suggesting that the broader biomarker repertoire of cfRNA is a key driver of its classification power. With both samples and features matched, cfRNA (median AUC = 0.883) and protein (median AUC = 0.857) perform similarly with a slight improvement in cfRNA ($p = 3.148\text{e-}4$), again indicating comparable performance in distinguishing KD from MIS-C. Finally, we examine which biomarkers were most frequently selected across the full dataset and target space. Six gene transcripts (RN7SL752P, *TCF7, CA1, IFIT2, GBP5, MAP3K7CL*) appear in more than

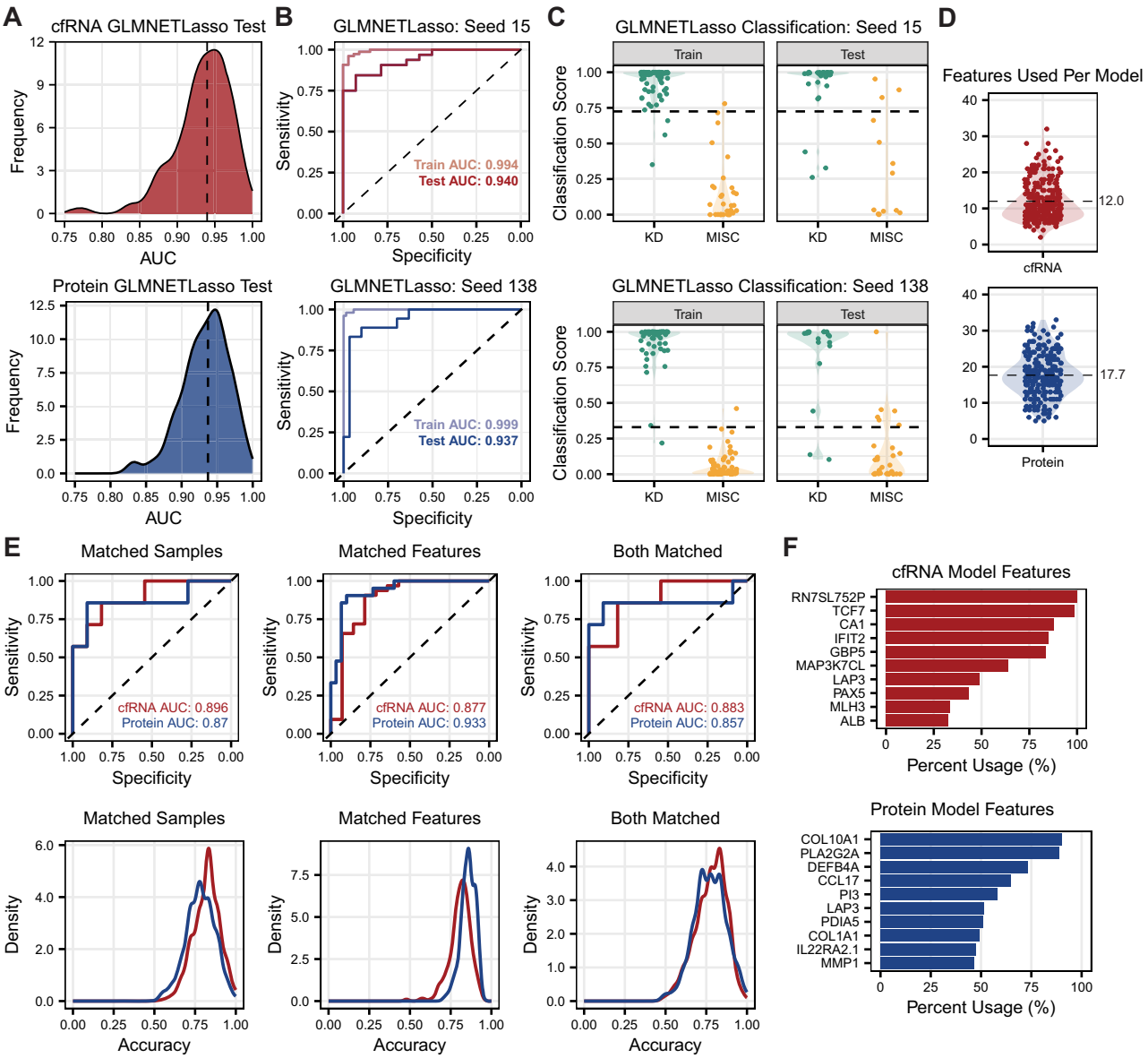

Fig. 3 | Machine learning classifiers constructed using a GLMNET model with Lasso regression. A Distribution of AUC values across 250 iterations for cfRNA- and protein-based classifiers. **B** AUC curves and classification results for train and test splits for the seed with the median AUC (cfRNA seed 15, protein seed 138). **C** Classification results of the median-performing seed for cfRNA and protein. **D** Distribution of total features used in the final model across 250 iterations for cfRNA and protein. **E** AUC curves of the median AUC seed and test accuracy distribution for cfRNA and protein when using only matched samples, matched features, and both matched samples and features. **F** Top 10 cfRNA and protein features by usage in total percentage of final models.

50% of cfRNA-based models, while seven proteins (COL10A1, PLA2GA, DEFB4A, CCL17, PI3, LAP3, PDIA5) are selected in the majority of protein-based models. Figure 3F) Notably, LAP3 is one of the most frequent features selected in both analytes, which suggests a more ubiquitous expression across RNA and protein that is relevant to KD.

KD subtyping based on cfRNA and protein profiles

A recent study demonstrated that KD patients can be stratified into four subtypes based on 14 clinical variables, each associated with distinct clinical outcomes and molecular profiles[13]. These subtypes align with specific hallmark features: subtype 1 is linked to elevated liver involvement, subtype 2 to increased band neutrophil activity, subtype 3 to pronounced lymphadenopathy, and subtype 4 to a lower average age of onset of disease. Given the different molecular perspectives offered by cfRNA and protein, we examine whether these more granular phenotypic differences are more clearly captured by one analyte compared to the other, and whether the three clinical variables present in the data (ALT, GGT, and CRP) are differentially expressed in any subtype for each analyte, and if they correlate with the measured clinical lab values. We perform differential abundance analyses of cfRNA and proteins for each KD subtype relative to MIS-C. All KD subtypes exhibit both differentially abundant cfRNA transcripts and proteins, except subtype 2, in which only one protein, CCL17, is differentially abundant when compared to MIS-C (Fig. 4A). This suggests that subtype 2 may share greater molecular similarity with MIS-C than with other KD subtypes. Consistent with these findings, repeated cross-validation analyses reveal comparably high accuracy for cfRNA- and protein-based models in distinguishing the other KD subtypes from MIS-C; however, classification accuracy for subtype 2 is notably lower in the protein-based models at only 52%, suggesting the model is incapable of distinguishing subtype 2 from MIS-C (Fig. 4B). Together, these results support an elevated molecular similarity between subtype 2 and MIS-C that may influence clinical presentation.

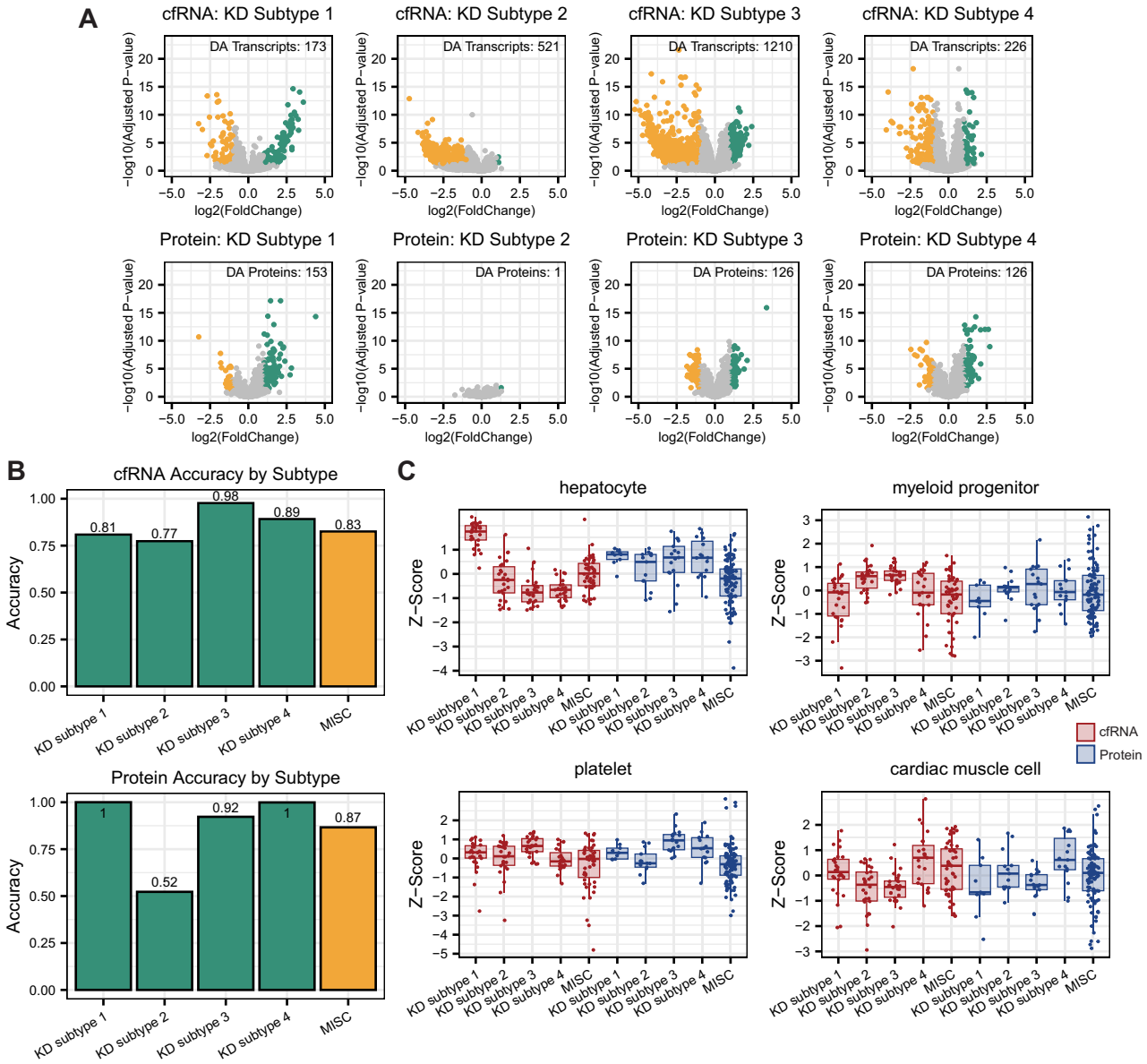

Fig. 4 | Comparative analysis of KD subtypes using cfRNA and protein. **A** Volcano plots for each subtype versus MIS-C in both cfRNA and protein. **B** Rate of correct classification of each subtype and MIS-C across 250 iterations of machine learning classifiers. **C** Boxplot of differential cell types for each KD subtype by Z-score using SSGSEA.

Analysis of the clinical biomarkers partially recapitulates the known features of each KD subtype: in cfRNA, *CRP* expression is significantly elevated in subtypes 1 (padj = 0.01), 3 (padj = 5.79e-04), and 4 (padj = 5.09e-08), consistent with the cardiac-associated phenotype of subtype 4, which shows the greatest fold change (|log FC| = 4.06). cfRNA for neither *CRP* ($r = 0.032$) nor *ALT* ($r = -0.043$) is well correlated with the respective clinical lab value, suggesting a unique measurement captured in cfRNA. In the protein data, ALT is the only significantly differentially expressed marker, elevated only in the liver-associated subtype 1 (padj = 7.54e-06), and while CRP ($r = 0.121$) is not well correlated with the clinical lab value, ALT is ($r = 0.722$). These results highlight that cfRNA and protein data capture different aspects of subtype-specific biology, each reflecting distinct dimensions of proximity to the underlying clinical phenotype. Notably, ALT is the only marker that showed a strong correlation with its clinical counterpart, consistent with its role as a liver-derived enzyme whose protein abundance more directly reflects hepatocellular injury and leakage into circulation.

To further explore subtype-specific differences, we perform single-sample gene set enrichment analysis (ssGSEA) using the Tabula Sapiens cell type reference (Methods, Fig. 4C). The top-enriched cell type differs across KD subtypes and often diverges between cfRNA and protein data. For instance, subtype 1—known for liver-enriched pathophysiology—exhibits increased liver-associated transcripts in cfRNA, yet a similar pattern is not seen in the protein data despite hepatocytes being the most enriched cell type in the protein data. In contrast, immune cell types—particularly granulocytes in subtype 2 and CD4-positive helper T cells in subtype 4—display more consistent enrichment patterns across both cfRNA and protein, underscoring a shared underlying immune activation. These results suggest that cfRNA may better capture the subtype-specific clinical phenotypes, but better concordance between cfRNA and protein may be in detecting immune-related molecular features and pathways.

Discussion

We compared two high-throughput platforms for plasma bio-analyte analysis, RNA sequencing of cell-free RNA (cfRNA) and SomaScan

proteomics of circulating proteins, to evaluate both their concordance and their utility in disease classification. Previous studies of mRNA–protein relationships in tissues and whole blood have reported only modest correlations[10,11] (0.3–0.6), reflecting the multiple biological mechanisms that decouple transcript and protein levels. Post-transcriptional regulation, protein turnover, and dynamic transitions from transcription to translation[23–26] may all contribute to this discordance in addition to technical differences in the measurement platforms. Future studies incorporating additional independent cohorts will be essential to assess the generalizability of these findings and establish the biological mechanisms underlying cfRNA–protein discordance.

Here, we found that cfRNA transcripts and proteins are essentially uncorrelated on a feature-by-feature basis, suggesting that each platform captures a distinct aspect of underlying biology. Indeed, the correlations between plasma RNA and proteins observed here are considerably weaker than the correlations reported in tissues and whole blood, indicating that additional factors such as release mechanisms, stability, and clearance may further disrupt any residual tissue-level concordance. The greater abundance and dynamic range of plasma proteins also may contribute to increased difficulty in detecting differential protein expression compared to RNA[4]. Notably, in MIS-C, there were substantially more differentially abundant transcripts than proteins, which could reflect increased cfRNA release triggered by tissue damage during systemic inflammation. Thus, it is likely that these differences arise from biological processes rather than technical artifacts. Despite minimal overlap at the individual molecular level, both cfRNA and protein converged on immune pathways, with cfRNA often revealing upstream regulators (e.g., interferon-inducible genes) and protein data highlighting downstream effectors (e.g., inflammatory mediators). Together, these findings suggest that combining cfRNA and protein measurements can lead to a more comprehensive perspective on pathophysiological processes.

Previous work has established the ability of cfRNA to correctly classify KD and MIS-C[6]. When we used machine-learning models to classify KD vs. MIS-C in both cfRNA and protein, both classifiers achieved high accuracy but selected different sets of diagnostic features. The cfRNA models emphasized immune-related transcripts (e.g., *TCF7*, which regulates the development of immune cells and inflammation[27], *IFIT2*, which regulates interferon expression[28], *GBP5*, which contributes to inflammasome assembly and is induced by interferon[29], and *MAP3K7CL*, which has been implicated in pediatric sepsis[30]. Of note, the models all included RN7SL752P, a pseudogene, suggesting additional diagnostic information contained within non-protein-coding transcripts. Protein models also chose immune-related features (e.g., DEFB4A, a defensin that can be and found in vascular endothelial cells[31], COL10A1 and COL1A1, involved in immune cells recruitment[32,33], PI3, an antimicrobial peptide[34], and LAP3, a metallopeptidase implicated in T cell depletion[35]. CCL17, a cytokine implicated in immune processes including arthritic disease and skin inflammation[36,37], was the most differentially abundant protein we observed in KD, suggesting a possible connection between its expression and the clinical presentation of skin inflammation seen in KD patients.

Our targeted evaluation of clinical biomarkers previously used to establish subtypes of KD[13] showed that cfRNA CRP and protein ALT measurements exhibited subtype-specific differences that partially mirrored known clinical patterns, yet these did not uniformly translate into parallel shifts across both analytes in accordance with serum marker expression. Together, these observations suggest that a biomarker's position along the transcriptional and translational pathway does not straightforwardly determine its concordance with clinical measurements and phenotypes, allowing cfRNA, proteins, and standard laboratory values to diverge while still offering complementary views of the same disease process.

Further analysis of KD subtypes revealed that subtype 2 proved the most challenging to distinguish from MIS-C, with few or no differentially abundant proteins identified. This may stem from an infectious trigger common to both conditions, as suggested by a higher incidence of KD shock syndrome (KDSS) in subtype 2. A final point of interest was the discrepancy in liver-derived signals for KD subtype 1: cfRNA revealed elevated liver transcripts, yet proteomic data did not mirror this trend, even though classical protein-based liver function tests align with clinical pathology. This highlights the ability of cfRNA to capture dynamic changes that manifest clinically on a finer scale. These findings illustrate how clinical assays, affinity-based proteomics, and cfRNA profiling each capture unique facets of disease.

High-throughput proteomics platforms continue to evolve, with debate ongoing over their optimal role in large-scale research and clinical settings. While mass spectrometry has been a gold standard for protein discovery, affinity-based methods such as SomaScan and Olink can profile thousands of targets from minimal volumes but are susceptible to platform biases and remain expensive ($750-1000). Proteomic platforms have shown a lack of interplatform agreement, coverage, and reproducibility[1], underscoring the importance of platform choice in interpreting results[38]. Emerging lower-cost approaches, including magnetic bead–based enrichment workflows[39] (e.g., MagNet) and paramagnetic bead–based kits[40] (e.g., ENRICH-iST), have demonstrated the ability to identify thousands of plasma proteins with high reproducibility, automation, and tissue-enriched coverage, representing promising directions for scalable plasma proteomics. In contrast, cfRNA profiling offers a real-time snapshot of gene expression at comparatively lower costs (~$150 per sample) and the option to convert detection of cfRNA to a quantitative PCR platform. However, widespread clinical implementation of cfRNA assays will require new protocols, workflows, and infrastructure.

Taken together, our results underscore the complementary nature of cfRNA and protein as blood-based biomarkers. Ultimately, integrating cfRNA and protein analyses together could enhance molecular characterization, improve disease classification, and guide more targeted therapeutic approaches.

Data availability

De-identified RNA-seq count matrices and have been deposited in the NCBI Gene Expression Omnibus (GEO) database (GSE255555). All figure source data has been uploaded to GitHub. The source data for Figs. 1–4 are available in the corresponding figure source data files (Fig. 1A, Fig. 1B, Fig. 2A, etc.) within the same GitHub repository. (https://github.com/adb258/cfRNA_proteins_manuscript.git).

Code availability

All associated code and has been uploaded to GitHub (https://github.com/adb258/cfRNA_proteins_manuscript.git).

Received: 5 May 2025; Accepted: 20 February 2026;
Published online: 13 March 2026

References

1. Kirsher, D. Y. et al. The Current Landscape of Plasma Proteomics: Technical Advances, Biological Insights, and Biomarker Discovery. 2025.02.14.638375 Preprint at https://doi.org/10.1101/2025.02.14.638375 (2025).
2. Guo, T., Steen, J. A. & Mann, M. Mass-spectrometry-based proteomics: from single cells to clinical applications. *Nature* **638**, 901–911 (2025).
3. Wang, B. et al. Comparative studies of 2168 plasma proteins measured by two affinity-based platforms in 4000 Chinese adults. *Nat. Commun.* **16**, 1869 (2025).
4. Hortin, G. L. & Sviridov, D. The dynamic range problem in the analysis of the plasma proteome. *J. Proteom.* **73**, 629–636 (2010).
5. Hu, T., Chitnis, N., Monos, D. & Dinh, A. Next-generation sequencing technologies: An overview. *Gener. Seq. Its Appl. Med. Lab. Immunol.* **82**, 801–811 (2021).
6. Loy, C. J. et al. Plasma cell-free RNA signatures of inflammatory syndromes in children. *Proc. Natl. Acad. Sci*. **121**, e2403897121 (2024).

7. Loy, C. J. et al. Nucleic acid biomarkers of immune response and cell and tissue damage in children with COVID-19 and MIS-C. *Cell Rep. Med*. **4**, (2023).
8. Chang, A. et al. Circulating cell-free RNA in blood as a host response biomarker for detection of tuberculosis. *Nat. Commun.* **15**, 4949 (2024).
9. Loy, C., Ahmann, L., Vlaminck, I. D. & Gu, W. Liquid Biopsy Based on Cell-Free DNA and RNA. *Annu. Rev. Biomed. Eng.* **26**, 169–195 (2024).
10. Buccitelli, C. & Selbach, M. mRNAs, proteins and the emerging principles of gene expression control. *Nat. Rev. Genet.* **21**, 630–644 (2020).
11. Ponomarenko, E. A. et al. Workability of mRNA Sequencing for Predicting Protein Abundance. *Genes* **14**, 2065 (2023).
12. McCrindle, B. W. et al. Diagnosis, Treatment, and Long-Term Management of Kawasaki Disease: A Scientific Statement for Health Professionals From the American Heart Association. *Circulation* **135**, e927–e999 (2017).
13. Wang, H. et al. Subgroups of children with Kawasaki disease: a data-driven cluster analysis. *Lancet Child Adolesc. Health* **7**, 697–707 (2023).
14. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
15. Friedman, J. et al. *glmnet: Lasso Elastic-Net. Regularized Generalized Linear Models* **4**, 1–10, https://doi.org/10.32614/CRAN.package.glmnet (2008).
16. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* **14**, 7 (2013).
17. Krämer, A., Green, J., Pollard, J. & Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinforma. Oxf. Engl.* **30**, 523–530 (2014).
18. Day-Lewis, M., Son, M. B. F. & Lo, M. S. Kawasaki disease: contemporary perspectives. *Lancet Child Adolesc. Health* **8**, 781–792 (2024).
19. Song, A. et al. Salient type 1 interleukin 1 receptor expression in peripheral non-immune cells. *Sci. Rep.* **8**, 723 (2018).
20. Elemam, N. M., Talaat, I. M. & Maghazachi, A. A. CXCL10 Chemokine: A Critical Player in RNA and DNA Viral Infections. *Viruses* **14**, 2445 (2022).
21. Liu, B. et al. The biology of VSIG4: Implications for the treatment of immune-mediated inflammatory diseases and cancer. *Cancer Lett.* **553**, 215996 (2023).
22. Brindis, R. G. & Kloner, R. A. Sildenafil in patients with cardiovascular disease. *Am. J. Cardiol.* **92**, 26M–36M (2003).
23. Leyes Porello, E. A., Trudeau, R. T. & Lim, B. Transcriptional bursting: stochasticity in deterministic development. *Dev. Camb. Engl.* **150**, dev201546 (2023).
24. Miceli, R. T. et al. Extracellular vesicles, RNA sequencing, and bioinformatic analyses: Challenges, solutions, and recommendations. *J. Extracell. Vesicles* **13**, e70005 (2024).
25. Nation, G. K., Saffold, C. E. & Pua, H. H. Secret messengers: Extracellular RNA communication in the immune system. *Immunol. Rev.* **304**, 62–76 (2021).
26. Wittich, H. et al. Transcriptome-wide association study of the plasma proteome reveals cis and trans regulatory mechanisms underlying complex traits. *Am. J. Hum. Genet.* **111**, 445–455 (2024).
27. Ma, Z. et al. TCF7 is highly expressed in immune cells on the atherosclerotic plaques, and regulating inflammatory signaling via NFκB/AKT/STAT1 signaling. *Biosci. Rep.* **42**, BSR20212064 (2022).
28. Siegfried, A. et al. IFIT2 Is an Effector Protein of Type I IFN–Mediated Amplification of Lipopolysaccharide (LPS)-Induced TNF-α Secretion and LPS-Induced Endotoxin Shock. *J. Immunol.* **191**, 3913–3921 (2013).
29. Li, X. et al. Critical role of guanylate binding protein 5 in tumor immune microenvironment and predictive value of immunotherapy response. *Front. Genet.* **13**, 984615 (2022).
30. Zhang, L., Chu, Q., Jiang, S. & Shao, B. Integration of Mendelian Randomization to explore the genetic influences of pediatric sepsis: a focus on RGL4, ATP9A, MAP3K7CL, and DDX11L2. *BMC Pediatr.* **25**, 66 (2025).
31. Kawsar, H. I. et al. Expression of human β-defensin-2 in intratumoral vascular endothelium and in endothelial cells induced by transforming growth factor β. *Peptides* **31**, 195–201 (2010).
32. Wang, X. et al. Prognostic value of COL10A1 and its correlation with tumor-infiltrating immune cells in urothelial bladder cancer: A comprehensive study based on bioinformatics and clinical analysis validation. *Front. Immunol.* **14**, 955949 (2023).
33. Zhang, C. et al. COL1A1 Is a Potential Prognostic Biomarker and Correlated with Immune Infiltration in Mesothelioma. *BioMed. Res. Int.* **2021**, 5320941 (2021).
34. Okkenhaug, K. Signalling by the phosphoinositide 3-kinase family in immune cells. *Annu. Rev. Immunol.* **31**, 675–704 (2013).
35. Li, L. et al. LAP3 contributes to IFN-γ-induced arginine depletion and malignant transformation of bovine mammary epithelial cells. *BMC Cancer* **22**, 864 (2022).
36. Lee, K. M.-C. et al. CCL17 in Inflammation and Pain. *J. Immunol. Baltim. Md 1950* **205**, 213–222 (2020).
37. Gilet, J. et al. Role of CCL17 in the Generation of Cutaneous Inflammatory Reactions in Hu-PBMC-SCID Mice Grafted with Human Skin. *J. Invest. Dermatol.* **129**, 879–890 (2009).
38. Upadhya, S. R. & Ryan, C. J. Antibody reliability influences observed mRNA-protein correlations in tumour samples. *Life Sci. Alliance* **6**, e202201885 (2023).
39. Wu, C. C. et al. Mag-Net: Rapid enrichment of membrane-bound particles enables high coverage quantitative analysis of the plasma proteome. *BioRxiv Prepr. Serv. Biol*. 2023.06.10.544439 https://doi.org/10.1101/2023.06.10.544439.(2024).
40. Soni, R. K. Frontiers in plasma proteome profiling platforms: innovations and applications. *Clin. Proteom.* **21**, 43 (2024).

Acknowledgements

We thank the patients and their families for their help to further our understanding of pediatric inflammatory conditions. We thank the Cornell Genomics Center for help with sequencing libraries. We thank the members of the PEMKDRG: Lukas Austin-Page, MD, Amy Bryl, MD, Joelle Donofrio-Ödmann, MD, Atim Ekpenyong, MD, David Gutglass, MD, Scott Herskovitz, MD, Paul Ishimine, MD, John Kanegaye, MD, Margaret Nguyen, MD, Mylinh Nguyen, MD, Kristy Schwartz, MD, Stacey Ulrich, MD, Tatyana Vayngortin, MD, and Elise Zimmerman, MD. This work was funded by NIH/National Institute of Child Health and Human Development grants R61HD105618, R33HD105618, and R33HD105593 (C.S., A.H.T., J.C.B., and I.D.V.) and the Macklin foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

A.B., C.J.L., J.C.B., and I.D.V. designed research; A.B., C.J.L., J.L., E.B. processed samples; A.B., C.J.L., J.K., C.S., J.L., E.B., A.H.T., J.C.B., C.Y.C., and I.D.V. analyzed data; C.S., A.H.T., and J.C.B. identified and collected patient samples and clinical metadata; and A.B., C.J.L., and I.D.V. wrote the paper.

Competing interests

C.J.L. and I.D.V. are inventors on submitted patents pertaining to cell-free nucleic acids (US patent applications 63/237,367 and 63/429,733). C.J.L. and I.D.V. are co-founders of Romix. I.D.V. is a co-founder of Kanvas Biosciences. I.D.V. is a member of the Scientific Advisory Board of Karius Inc., Kanvas Biosciences and GenDX. I.D.V. is listed as an inventor on submitted patents pertaining to cell-free nucleic acids (US patent

applications 63/237,367, 63/056,249, 63/015,095, 16/500,929, 14P-47510551-01-US) and receives consulting fees from Eurofins Viracor. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43856-026-01489-7.

Correspondence and requests for materials should be addressed to Iwijn De Vlaminck.

Peer review information *Communications Medicine* thanks Zhimin Geng, Julia Ive, Estefanía Núñez, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.