Immunology & Cell Biology 2024; 1–13
www.wileyonlinelibrary.com/journal/icb

ORIGINAL ARTICLE

OPEN

Proteomics screening after pediatric allogenic hematopoietic stem cell transplantation reveals an association between increased expression of inhibitory receptor FCRL6 on γδ T cells and cytomegalovirus reactivation

Adam Alexandersson[1,2,3], Mikko S Venäläinen[4], Nelli Heikkilä[2,5], Xiaobo Huang[2], Mervi Taskinen[1], Pasi Huttunen[1,3], Laura L Elo[4,6], Minna Koskenvuo[1] & Eliisa Kekäläinen[2,7]

1 Division of Pediatric Hematology, Oncology and Stem Cell Transplantation, New Children's Hospital, Helsinki University Hospital, Helsinki, Finland
2 Translational Immunology Research Program, Faculty of Medicine, University of Helsinki, Helsinki, Finland
3 Children and Adolescents, Pediatric Research Center, New Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
4 Turku Bioscience Centre, University of Turku and Åbo Akademi University, Turku, Finland
5 Center of Vaccinology, University of Geneva, Geneva, Switzerland
6 Institute of Biomedicine, University of Turku, Turku, Finland
7 HUS Diagnostic Center, Clinical microbiology, Helsinki University Hospital, Helsinki, Finland

Keywords
biomarkers, FCRL6, gd T cells, HSCT, pediatric, proteomics, γδ T cells

Correspondence
Adam Alexandersson, Translational Immunology Research Program, PB 21, 00014 University of Helsinki, Helsinki, Finland.
E-mail: adam.alexandersson@helsinki.fi

Received 23 November 2023;
Revised 7 and 10 April 2024;
Accepted 10 April 2024

doi: 10.1111/imcb.12762

Immunology & Cell Biology 2024; 1–13

Abstract

We studied the associations between inflammation-related proteins in circulation and complications after pediatric allogenic hematopoietic stem cell transplantation (HSCT), to reveal proteomic signatures or individual soluble proteins associated with specific complications after HSCT. We used a proteomics method called Proximity Extension Assay to repeatedly measure 180 different proteins together with clinical variables, cellular immune reconstitution and blood viral copy numbers in 27 children (1–18 years of age) during a 2-year follow-up after allogenic HSCT. Protein profile analysis was performed using unsupervised hierarchical clustering and a regression-based method, while the Bonferroni-corrected Mann–Whitney *U*-test was used for time point–specific comparison of individual proteins against outcome. At 6 months after allogenic HSCT, we could identify a protein profile pattern associated with occurrence of the complications such as chronic graft-*versus*-host disease, viral infections, relapse and death. When protein markers were analyzed separately, the plasma concentration of the inhibitory and cytotoxic T-cell surface protein FCRL6 (Fc receptor-like 6) was higher in patients with cytomegalovirus (CMV) viremia [$\log_2$-fold change 1.5 ($P = 0.00099$), 2.5 ($P = 0.00035$) and 2.2 ($P = 0.045$) at time points 6, 12 and 24 months]. Flow cytometry confirmed that FCRL6 expression was higher in innate-like γδ T cells, indicating that these cells are involved in controlling CMV reactivation in HSCT recipients. In conclusion, the potentially druggable FCRL6 receptor on cytotoxic T cells appears to have a role in controlling CMV viremia after HSCT. Furthermore, our results suggest that system-level analysis is a useful addition to the studying of single biomarkers in allogenic HSCT.

INTRODUCTION

Allogenic hematopoietic stem cell transplantation (HSCT) has the potential to cure several malignant and benign conditions of the bone marrow, but is burdened by risk of complications.[1] Routinely used clinical parameters and laboratory tests are insufficient to predict the occurrence of potentially lethal and debilitating complications such as graft-*versus*-host disease (GVHD)[2] and bacterial, fungal and viral infections.[3–5] Reactivation of latent cytomegalovirus (CMV) is one of the main complications after allo-HSCT.[6] Tools to predict complications are needed because, for example, antivirals used to treat CMV reactivations are potentially toxic, and their use should be limited only to patients that clearly benefit from them. Adoptive transfer of pathogen-specific allogenic lymphocytes is still mostly an experimental treatment or prophylaxis option for reactivation of latent viruses; the optimal cellular composition of the adoptive transfer to balance the risk for GVHD and effective antiviral clearance remains unknown.[7–9]

Many studies have focused on predicting CMV reactivation but validated biomarkers are still lacking.[10] Delayed immune reconstitution (IR) is a risk for CMV reactivations.[10] CMV-specific $CD8^+$ T-cell cytokine signatures can predict the risk of CMV reactivation, but testing of cellular responses is complicated and not suitable for routine clinical laboratories.[11] Testing for individual biomarkers or using a panel of soluble plasma biomarkers to predict complications after HSCT would better suit clinical practice. Proteomics approaches after HSCT show promising results regarding the understanding of pathogenesis as well as predicting relapse, nonrelapse mortality, acute GVHD (aGVHD) and chronic GVHD (cGVHD), including therapy response.[2,12,13] No biomarker or panel of biomarkers is, however, yet widely in clinical use.[2]

A benefit of the proteomics and other systems immunology approaches is their hypothesis-generating attribute, as opposed to testing *a priori* formulated research questions.[14] To our knowledge, no proteomic studies have been performed prospectively on a well-defined cohort of children receiving HSCT with myeloablative conditioning. To better predict the onset and to understand the mechanisms of severe complications after HSCT in children, we performed a longitudinal deep analysis on the plasma proteome in relation to clinically relevant outcomes such as cGVHD, viral infections, relapse and death.

RESULTS

Unsupervised clustering of plasma proteome identifies patients with HSCT-related complications and poor immune reconstitution

We first classified the patients based on the complications that they experienced during the IR period into complication-free and complication-prone groups. The "Complication-free" status was attributed to patients who did not suffer any severe complications (death, relapse, latent virus reactivations, severe aGVHD or cGVHD) and the "complication-prone" status was correspondingly designated if any of those complications were present during the follow-up. Sparkline plots were generated with locally estimated scatterplot smoothing (Figure 1) to illustrate how individual protein levels measured with the Proximity Extension Assay and plotted over the IR time change in complication-free and complication-prone individuals. However, almost all measured proteins changed over time with varying patterns with no clear patterns emerging between groups. Therefore, we decided to next investigate a possible association between a global plasma protein profile and clinical outcome measures, by performing an unsupervised clustering. A heatmap (Figure 2a), created using hierarchical clustering based on protein levels at 6 months after HSCT, revealed two clusters with distinctly different outcome profiles. A similar but less clearly separating pattern for two main clusters was visible in other time points (Supplementary figure 4). We chose to focus on the 6-month time point in the following analyses, because by this time most complications had occurred, but the reconstitution of the immune system would not yet be finalized. At this point, all aGVHD had occurred, every viral infection was diagnosed[15] and most cases of cGVHD had begun. Some cases of death and relapse had yet to occur, but all patient-specific complications were registered and included in analysis independent of time of onset, because relevant immunological patterns might be seen before, during and after debut.

At 6 months after HSCT, the complication-prone status was clearly more frequent in cluster 2, where 14/15 patients had complication-prone status *versus* 4/10 in cluster 1 (93% *versus* 40%, $P = 0.007$ by Fisher's exact test). The separation was mainly explained by higher CMV infection incidence in cluster 2 (53% *versus* 10%, $P = 0.04$) as well as the major complications of relapse, death and cGVHD only occurring in cluster 2. Other viral reactivations were distributed evenly in both clusters: human herpesvirus 6 reactivation affected 40% of the patients in cluster 2 *versus* 20% in cluster 1 ($P = 0.4$ by

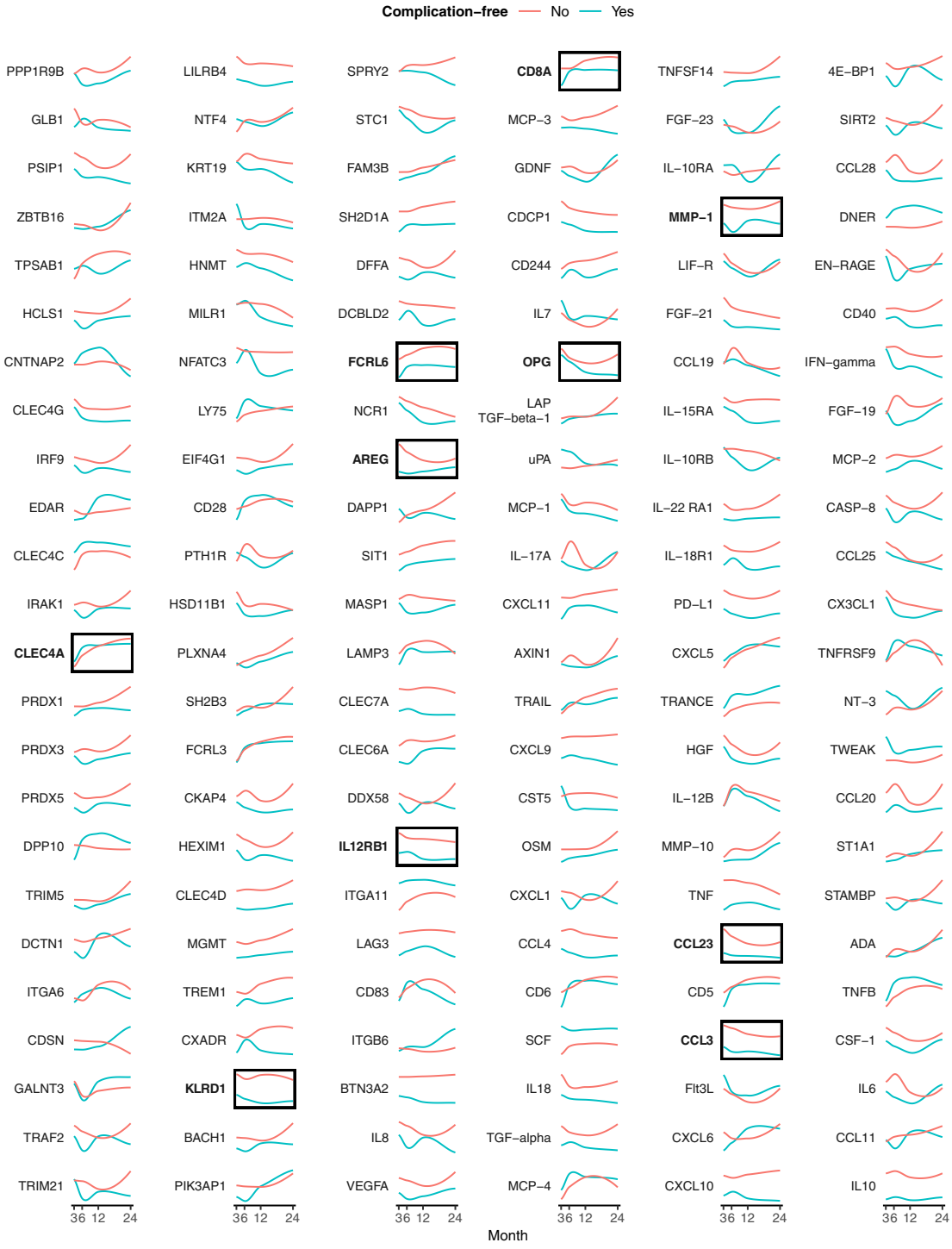

Figure 1. Trajectories of individual proteins. Sparkline plot, generated with locally estimated scatterplot smoothing (LOESS), of individual protein kinetics over time according to patient grouping by complication-free status. The "complication-free" status (red lines) was attributed to patients who did not suffer any severe complications (death, relapse, latent virus reactivations, severe aGVHD or cGVHD) and the "complication-prone" status (blue lines) was correspondingly designated if any of those complications were present during the follow-up. Highlighted proteins were found to be clearly associated with complications. aGVHD, acute graft-*versus*-host disease; cGVDH, chronic graft-*versus*-host disease.

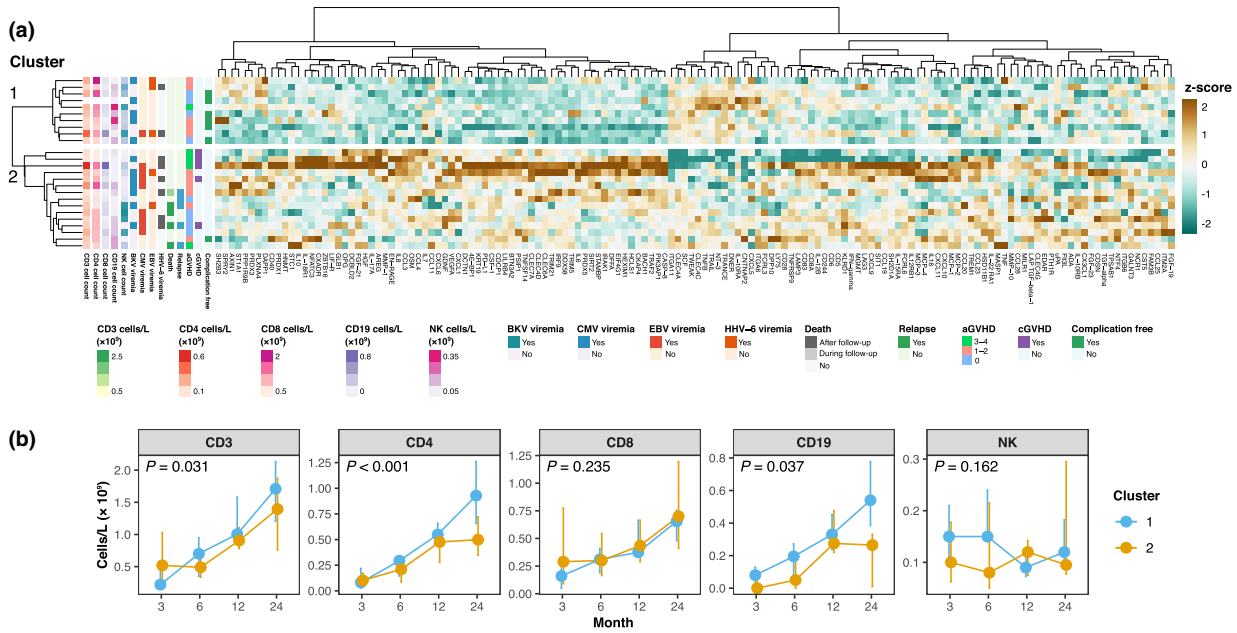

Figure 2. Protein profile in relation to outcome. **(a)** Heatmap generated by hierarchical clustering of protein levels at 6 months after hematopoietic stem cell transplantation (HSCT). One patient per row is displayed, with patient characteristics to the left. **(b)** Comparison of immune reconstitution between the two major clusters from the heatmap. The *P*-values were obtained for cluster × time interaction terms in linear mixed effect models with time and cluster as additional fixed effects and including a random intercept for patients. aGVHD, acute graft-*versus*-host disease; BKV, BK virus; cGVDH, chronic graft-*versus*-host disease; CMV, cytomegalovirus; EBV, Epstein–Barr virus; HHV, human herpesvirus; NK natural killer.

Fisher's exact test), Epstein–Barr virus and BK virus viremias occurred with similar frequencies [13% *versus* 30% ($P = 0.4$) and 47% *versus* 50% ($P > 0.99$), respectively]. The frequencies of aGVHD were similar (67% *versus* 60%, $P > 0.99$). The clinically more relevant aGVHD grades 3 and 4 were more common in cluster 2, but the difference was insignificant (23% *versus* 10%, $P = 0.6$). The corresponding heatmaps for other time points also showed association with major clinical outcomes (Supplementary figure 4).

We next analyzed how these two clusters differed in their overall cellular IR parameters. When the whole follow-up was included in the statistical analysis, the amount of circulating T cells (defined as $CD3^+CD19^-$), $CD4^+$ T cells (defined as $CD3^+CD19^-CD4^+$), and B cells (defined as $CD3^-CD19^+$) were significantly lower from the 6-month time point onwards in the complication-enriched cluster 2, compared to cluster 1 (Figure 2b). In cluster 2, total T cells and $CD4^+$ T cells were lower in the time point 6 months onward, while B cells were lower during the whole follow-up. Interestingly, CD8 T and natural killer cell counts did not show differences. We can thus conclude that follow-up-wide IR of CD4 T cells and B cells correlated with protein profile status at the 6-month time point (Figure 2b).

In summary, our unsupervised analysis highlighted that CMV reactivation was one of the most defining clinical features for the patients segregated by a specific plasma inflammatory proteome profile. Moreover, this proteome profile was evident before failure in specific parts of cellular IR.

Protein coexpression levels are connected to outcome

We analyzed the global evolution of the proteome profile by studying patterns of correlating proteins between our complication-free and complication-prone groups (Figure 3a, b). Positive and negative correlation coefficients, respectively, indicate a correlated upregulation or a correlated downregulation of two proteins, whereas a correlation coefficient of about 0 indicates no coordinated expression of two proteins. The complication-prone patients had a narrow distribution of correlation coefficients peaking at 0, indicating little protein coexpression in the early time points of 3 and 6 months when compared with complication-free patients.

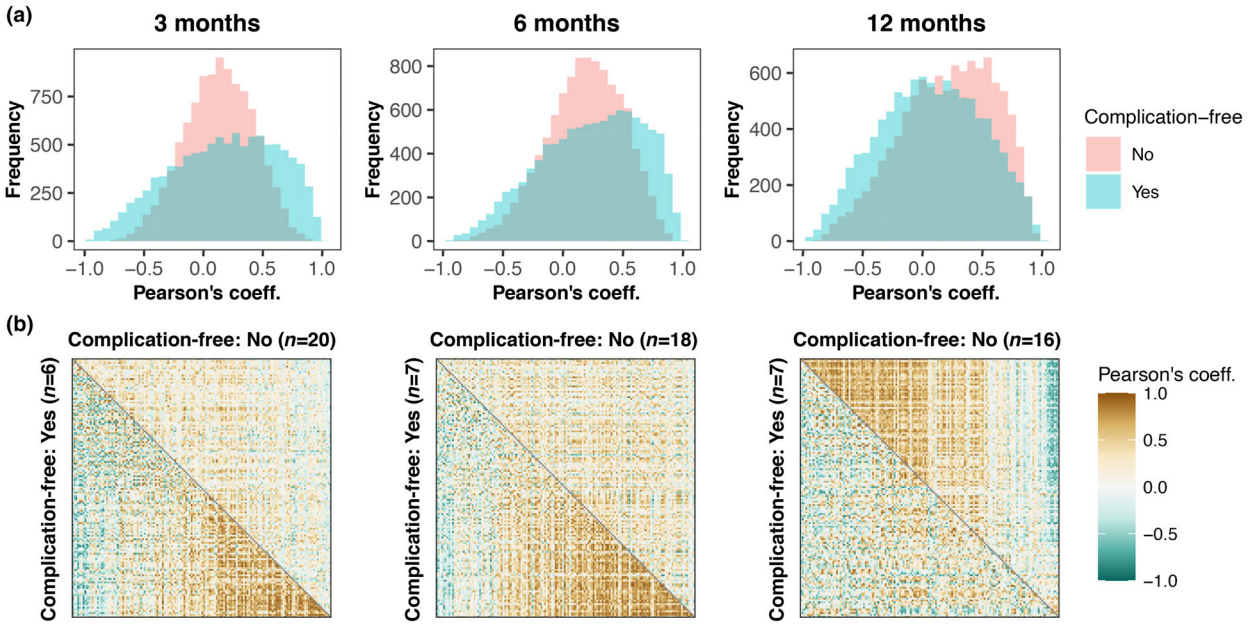

Figure 3. Interprotein correlations within complication-free and complication-prone groups. **(a)** Histograms of Pearson's correlation coefficients grouped according to complication status. **(b)** Pearson's correlation matrices for correlation-free and correlation-prone groups. The proteins were ordered according to the differences in correlation between the two groups by the median strength of correlation differences for each protein.

By contrast, complication-free patients had a positively skewed pattern of correlation coefficients, indicating correlated (i.e. coordinated) upregulation of proteins (Figure 3a). At 12 months after HSCT, at the peak of thymopoiesis, the protein correlations of the complication-prone patients become positively skewed, while complication-free patients had an even Gaussian distribution of the correlation coefficients, reflecting the homeostasis acquired (Figure 3a). This is illustrated in the correlation matrix of different correlations between proteins: at the early time points in the complication-free patients, there is a clear module of coexpressed proteins, whereas in complication-prone patients, the protein expression is disorganized. By contrast at the 12-month time point, modules of coexpressed proteins were observed only in the complication-prone patients (Figure 3b).

In summary, the protein correlation patterns over the IR period after HSCT are different depending on the outcome. Graphically, the Pearson's correlation matrices show that patients with more complications have less correlating protein pairs, that is, a more random protein profile at early time points, whereas the complication-free patients showed more synchronized inflammatory protein kinetics. This was also shown statistically (Supplementary table 5), where the protein pair correlation level was higher in the complication-free compared with the complication-prone group at early time points. The heterogeneity of the complications included in the complication-prone group likely explains the disorganized inflammatory protein kinetics. The positively skewed protein correlation pattern in complication-prone individuals 1 year after HSCT could indicate that achieving immune homeostasis was delayed in the complication-prone individuals.

Individual protein trajectories are associated with specific outcomes

Our analysis thus far has established that our complication-free and complication-prone patients had qualitatively different proteome profiles over the IR after allo-HSCT. As large-scale proteome analysis is not feasible for use in clinical practice, we continued to analyze specific proteins in relation to specific outcomes over the whole course of our prospective sample set. Here, we applied the maSigPro regression-based analysis which follows a two-step regression strategy to find protein expression values with significant temporal expression changes. We reasoned that the proteins with significantly different expression values over the whole IR have the highest potential to be clinically relevant biomarkers. In addition, we tested the protein levels against clinical outcomes at every time point using the

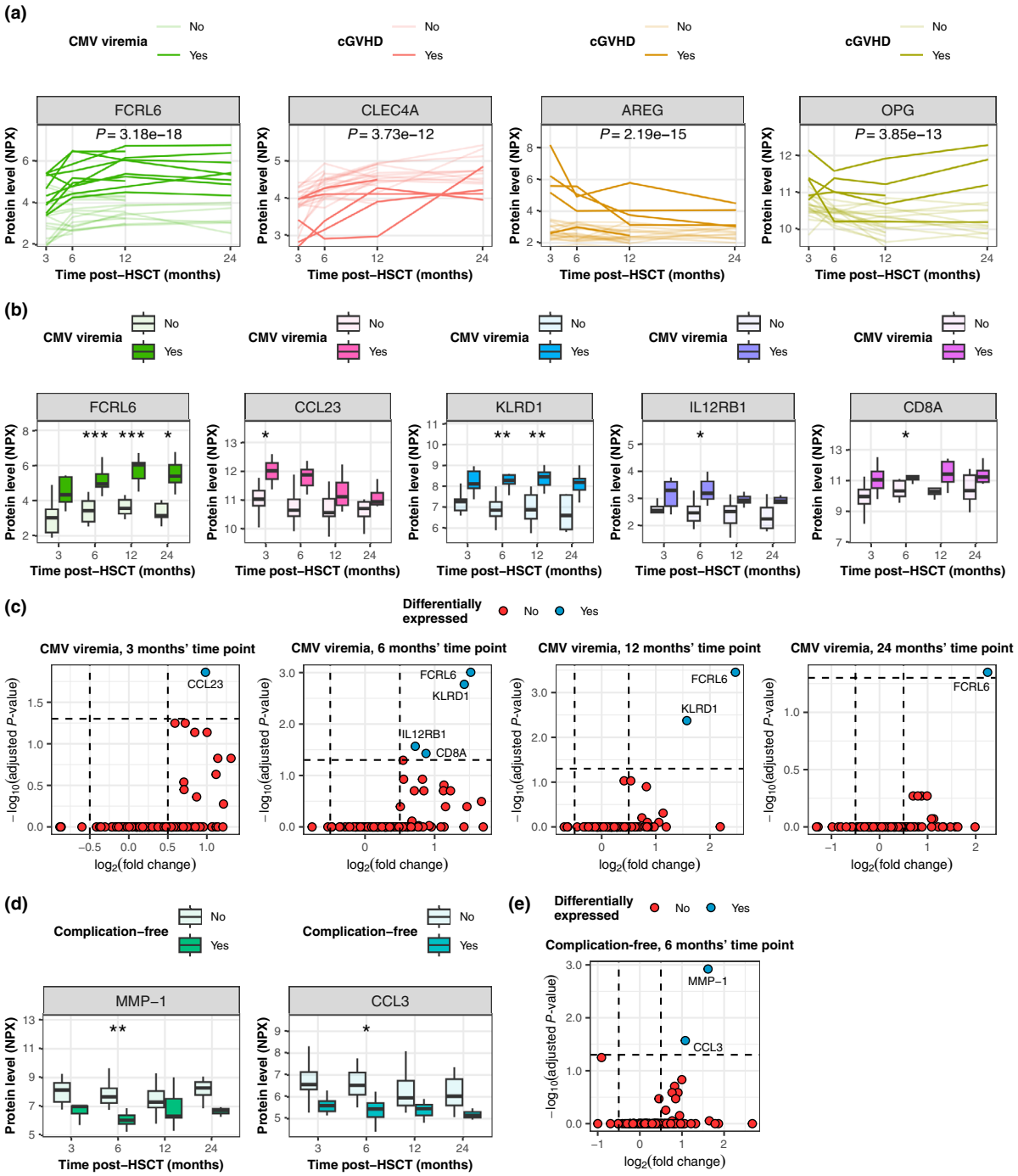

Figure 4. Individual proteins and outcomes. **(a)** Kinetics of maSigPro-identified significant protein level differences over time, depending on outcome. One line represents one patient. **(b)** Time point–specific comparison of protein levels depending on cytomegalovirus (CMV) status. **(c)** Levels of differentiation according to outcome and time point depending on CMV status. **(d)** Time point–specific comparison of protein levels depending on complication status. **(e)** Levels of differentiation according to outcome and time point depending on complication status. AREG, amphiregulin; CCL3, C–C motif chemokine ligand 3; CCL23, C–C motif chemokine ligand 23; CD8A, T-cell surface glycoprotein CD8 alpha chain; cGVDH, chronic graft-*versus*-host disease; CLEC4A, C-type lectin domain family 4 member A; CMV, cytomegalovirus; FCRL6, Fc receptor-like 6; HSCT, hematopoietic stem cell transplantation; IL12RB1, interleukin 12 receptor subunit beta 1; KLRD1, killer cell lectin-like receptor D1; MMP-1, matrix metalloproteinase-1; NPX, normalized protein expression; OPG, osteoprotegerin. $*P < 0.05$; $**P < 0.01$, $***P < 0.001$.

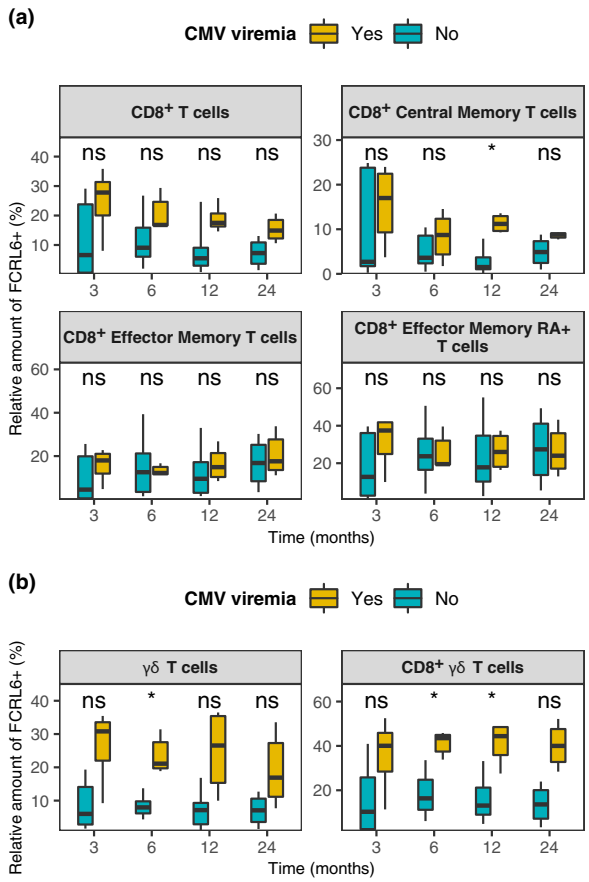

Figure 5. FCRL6 (Fc receptor-like 6)-positive cytotoxic T-cell subsets in patients with cytomegalovirus (CMV) viremia. **(a)** The box plots show the fraction of FCRL6$^+$ cells of CD8$^+$ T cells, CD8$^+$ central memory T cells, CD8$^+$ effector memory T cells and CD8$^+$ effector memory RA$^+$ T cells at different time points, depending on CMV status. **(b)** The box plots show the fraction of FCRL6$^+$ cells of all γδ T cells and of CD8$^+$ γδ T cells, depending on CMV status. ns, not significant. *$P < 0.05$.

Mann–Whitney *U*-test to identify the clinical correlates for each identified potential biomarker (Figure 4).

Our analysis with maSigPro showed that proteins CLEC4A [C-type lectin domain family 4 member A; also known as DCIR (DC immunoreceptor)], AREG (amphiregulin), OPG (osteoprotegerin; also known as tumor necrosis factor receptor superfamily member 11B) and FCRL6 (Fc receptor-like 6) levels showed significantly different expression values over the whole IR (Figure 4a). When we correlated these proteins with clinical outcomes, we found correlations between cGVHD and low CLEC4A ($P = 3.73 \times 10^{-12}$), high AREG ($P = 2.19 \times 10^{-15}$) and high OPG ($P = 3.85 \times 10^{-13}$; Figure 4a). Moreover, there was a highly significant association between CMV viremia and high FCRL6 ($P = 3.18 \times 10^{-18}$; Figure 4a).

Of the proteins that were significantly different over the whole follow-up period, FCRL6 showed the clearest connection with outcome: for CMV infection at time points 6, 12 and 24 months, $\log_2$-fold change was 1.52 ($P = 9.9 \times 10^{-4}$), 2.47 ($P = 3.5 \times 10^{-4}$) and 2.25 ($P = 0.045$; Figure 4b), respectively. In time point–specific analysis, CMV viremia was also associated with high CCL23 (C–C motif chemokine ligand 23), high KLRD1 (killer cell lectin-like receptor D1), high IL12RB and high CD8A (T-cell surface glycoprotein CD8 alpha chain), but these correlations were statistically significant in only one or two time points and the fold-change differences were lower than those of FCRL6 (Figure 4c). For example, higher CCL23 expression levels were significantly associated with CMV reactivation only in the earliest time point, after which its expression trajectory was not significantly different from the individuals who did not have CMV reactivation. Only FCRL6 showed significant differences in all time points from the 6-month time point onward. By contrast, the overall complication-prone status was associated with high MMP-1 (matrix metalloproteinase-1) and CCL3 (C–C motif chemokine ligand 3; also known as macrophage inflammatory protein 1 alpha) levels (Figure 4d, e).

γδ T cells expressing FCRL6 associates with CMV infection

FCRL6 appeared to be the most significant protein differentiating complication-free and complication-prone patients from 6 months after HSCT. As it was associated with CMV reactivation, we used flow cytometric analysis to further focus on its expression on cytotoxic T cells with known roles to control CMV. Overall, there was a higher trend for FCRL6 expression in CD8$^+$ T cells in patients with CMV viremia than in those without, but the results remained statistically nonsignificant (Figure 5a). No statistically significant differences of FCRL6 expression between patients with or without CMV viremia were observed in effector-memory, central memory or RA$^+$ effector-memory subsets of CD8$^+$ T cells except for a slight difference at 12-month time point for the central memory subset. We then focused on another cytotoxic T-cell population, the γδ T cells (defined as CD3$^+$CD8$^+$γδ$^+$), where we observed significantly higher levels of FCRL6 expression in CD8$^+$ γδ T cells at time points 6 and 12 months after HSCT in patients with CMV viremia (Figure 5b). In addition, the expression of FCRL6 in all γδ T cells was significantly higher in patients with CMV at 6 months (Figure 5b).

DISCUSSION

In this single-center prospective study of children undergoing allogenic HSCT after myeloablative conditioning as the treatment for a hematologic malignancy, we found a clear association between overall plasma protein profile, as well as certain specific proteins, and the complication occurrence.

Hierarchical clustering, an inherently unsupervised process, revealed distinctly different protein profiles with regard to the following outcomes: death, relapse, cGVHD, CMV infection and IR. Even though the predictive value and clinical utility of the method are challenging to establish, hierarchical clustering illustrates the difficulties in denoting specific biomarkers for separate complications because the pathophysiologic processes are greatly intertwined. Delayed IR is a well-known risk factor for HSCT complications.[3] When considering the individual clinical outcomes, the two clusters generated by hierarchical clustering were the most clearly separated by CMV reactivation status, which is also a risk factor for delayed IR. After analyzing how different protein expression levels correlated during the IR period, we can conclude that the complication-prone individuals showed a clearly divergent pattern of protein correlations, while the good outcome–associated cluster 1 displayed a more synchronized pattern. This chaotic appearance is, by our interpretation, a representation of dysregulated IR in individuals that developed complications. These findings were in line with the protein pair analysis, where a good outcome was associated with a clear abundance of positively correlating protein pairs at 3 and 6 months (Supplementary table 5).

Among proteins that were higher in CMV-reactivation patients were KLRD1, IL12RB1 (interleukin 12 receptor subunit beta 1), CD8A and CCL23. All these molecules are associated with cytotoxic responses (i.e. their elevated levels were not surprising in patients experiencing CMV reactivation). The known proinflammatory proteins chemokine CCL3 and metalloproteinase MMP-1 were higher in complication-prone patients, unsurprising because they are known contributors of pathological inflammatory processes.[16–19] AREG, an autocrine growth factor and a mitogen for astrocytes, Schwann cells and fibroblasts,[20] was higher in patients with cGVHD at 3 months after HSCT, probably reflecting the verified relevance of AREG in aGVHD[21] and late aGVHD.[22] Other cGVHD-associations were higher levels of OPG (osteoprotegerin or TNF receptor superfamily member 11b) and lower levels of CLEC4A.

The most noticeable change between patients with CMV viremia and those without was the protein levels of FCRL6 (Figure 4a–c), an immunoregulatory receptor whose expression is restricted to cytotoxic lymphocytes.[23,24] The mechanism of action of FCRL6 is largely unknown, but its interaction with HLA-DR molecules led to reduced cytotoxicity of natural killer cells toward tumor cells, similar to LAG-3, which is a known marker for effector cell exhaustion.[25,26] Hence, FCRL6 has been suggested to be a possibly druggable check-point molecule to reverse the effector cell exhaustion. It could also serve as a potential biomarker for relapses.[25] To our knowledge, no association between FCRL6 expression and CMV infection has been shown before. Although FCRL6 has been shown to be upregulated in chronic inflammatory states[27] and HIV infection,[28] we found no connection with, for example, cGVHD or other viral infections. The difference in FCRL6 levels was not explained by standard lymphocyte cell counts, because its expression is restricted to cytotoxic $CD8^+$ and natural killer cells which showed comparable IR between clusters 1 and 2 at the 6-month time point. As the FCRL6 has no known soluble function, its measurable plasma levels are likely a result of increased receptor shedding from FCRL6-positive cells.

The fraction of FCRL6 expressing $CD8^+$ γδ cells was consistently higher in patients that had CMV reactivation (Figure 5). γδ T cells have a central role in defense against CMV reactivation after HSCT, and higher levels are associated with higher survival and lower relapse rates.[29] The important role of CMV control by γδ T cells has also been established in the αβ T-cell–depleted haploidentical HSCT setting.[30] However, FCRL6 expressing γδ T cells have not been studied before. Furthermore, the role of the $CD8^+$ γδ T-cell subset in allo-HSCT is mostly unknown, but it has been suggested to be more autoreactive based on higher proliferative and cytotoxic response in mixed-lymphocyte reaction.[31] It is thus possible that $CD8^+$ γδ cells are more potent effectors in containing CMV reactivation. Recent findings indicate that γδ T cells can control CMV not through direct CMV–antigen recognition but instead by sensing upregulated HLA molecules by antigen-presenting cells during the CMV infection.[32] Specifically, HLA-DR has been identified as one TCR ligand of γδ T cells responding to CMV infection.[33] One receptor involved in this sensing of activation could be FCRL6 whose known ligand is HLA-DR. Our results thus highlight the role of γδ T cells in controlling CMV and reveal a new receptor potentially involved in activation or exhaustion of γδ T cells. As γδ cells do not cause GVHD, they offer an interesting cellular source for adoptive cellular therapies to treat CMV reactivation.[34] Future studies should focus on understanding the functional significance of FCRL6 on γδ T-cell subsets.

Table 1. Cohort demographics.

Demographics	*n*	%	Complication prone	% of complication prone	Complication free	% of complication free	*P*-value[a]
Patients, total	27	100	20	100	7	100	
Male/female	18/9	67/33	13/7	65/35	5/2	71/29	> 0.99
Age	3–18 (mean 10.5)		3–18 (mean 11.3)		3–14 (mean 8.3)		0.13
Donor							> 0.99
Sibling	9	33	7	35	2	29	
MUD	18	67	13	65	5	71	
HLA match							0.46
9/10	2	7	1	5	1	14	
10/10 or higher	25	93	19	95	6	86	
Blood group mismatch							0.18
None	14	52	11	55	3	43	
Major/minor	5/8	18/30	5/4	25/20	0/4	0/57	
Diagnosis							0.76
Pre-B ALL	16	59	12	60	4	57	
AML	7	26	5	25	2	29	
MDS	1	4	1	5	0	0	
MPAL	1	4	1	5	0	0	
CML	1	4	0	0	1	14	
ETP-ALL	1	4	1	5	0	0	
Conditioning							
TBI	17	63	15	75	2	29	0.06
Serotherapy	18	67	13	65	5	71	> 0.99
GVHD prophylaxis							> 0.99
Prophylaxis	26	96	19	95	7	100	
No prophylaxis	1	4	1	5	0	0	

ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; CML, chronic myeloid leukemia; ETP-ALL, early T-cell precursor acute lymphoblastic leukemia; GVHD, graft-*versus*-host disease; HLA, human leukocyte antigen; MDS, myelodysplastic syndrome; MPAL, mixed phenotype acute leukemia; MUD, matched unrelated donor; serotherapy, alemtuzumab or antithymocyte globulin; TBI, total body irradiation.
[a]Differences between the complication-free and complication-prone groups were tested using the Mann–Whitney *U*-test for continuous variables (age) and the chi-square test or Fisher's exact test ($n < 5$) for categorical variables.

The relatively small sample size of our cohort did not allow for division into training and validation groups, but our set up with multiple time point measurements in addition to a great sample cover make the data robust. Ideally, time points would have been more numerous and closer to each other, yielding yet greater information on sequence of events as well as more power to time-dependent analyses. Even though not a weakness *per se*, our multiple parameter data were under significant strain of multiple testing correction, which might have masked some meaningful true associations. Unfortunately, our limited flow cytometry panel did not allow us to fully explore all possible FCRL6-expressing cytotoxic innate lymphocyte populations such as natural killer cells that could also be the source for the increased soluble FCRL6 in patients with CMV reactivation.

Proteomics screening methods offer a powerful tool to reveal new biomarkers or targets for novel therapies. In our pediatric cohort, the proteomic approach revealed several potential biomarkers that showed different kinetics in patients with and without complications after HSCT. Although not directly clinically translatable, plasma protein profiles seem to tell us more than levels of individual proteins alone. Moreover, the inhibitory FCRL6 receptor on cytotoxic T cells emerged as a potentially relevant biomarker or therapy target for CMV reactivation.

METHODS

Patient cohort, sample collection and clinical outcomes

This single-center study from the New Children's Hospital, Helsinki University Hospital, Finland, followed 27 children (1–18 years of age) with hematological cancer for 2 years after HSCT with myeloablative conditioning. We have previously

published the cohort's clinical characteristics, including timing of complications.[15] In addition to registering clinical disease– and transplant-related features, symptoms and signs, we collected blood samples at time points 3, 6, 12 and 24 months after HSCT. Plasma was isolated and stored frozen. Peripheral blood mononuclear cell isolation was done with gradient centrifugation (Ficoll-Paque PLUS, Cytiva, Buckinghamshire, UK) and cells were frozen in 10% dimethyl sulfoxide in fetal calf serum and stored frozen in liquid nitrogen or at –140°C. We measured viral copy numbers of CMV, human herpesvirus 6, BK virus and Epstein–Barr virus at 3, 6, 12 and 24 months, and additionally extracted levels of the same viruses from the patient records, when they had been requested by the clinician. IR was followed by clinical lymphocyte subtype testing (total $CD3^+$ T cells, $CD4^+$ T helper cells, $CD8^+$ cytotoxic T cells, $CD19^+$ B cells and natural killer cells). Lymphocytes and subclasses were quantified using accredited flow cytometry for clinical use and viral levels with real-time PCR at an accredited clinical laboratory (HUS Diagnostic Center, Helsinki, Finland).

As the spectrum of HSCT complications is diverse and numerous outcome subgroups strain statistical analyses, we created a composite outcome called "complication free". The "complication-free" status was attributed to patients who did not suffer any of the following: death during follow-up, relapse during follow-up, cGVHD per the 2014 National Institutes of Health (NIH) criteria,[35] aGVHD grade 3 or 4,[36] CMV viremia, BK virus viremia over 10 000 IU mL^{-1}, human herpesvirus 6 viremia in combination with BK virus or Epstein–Barr virus viremia. The "complication-prone" status was correspondingly designated if any of the aforementioned findings were present during the follow-up.

The study was approved by the Ethics Committee of the Helsinki University Hospital (HUS/2306/2016), and written informed consent was obtained from all parents and patients above 6 years of age. Baseline demographics are presented in Table 1.

Quality control and descriptive statistics

All 90 samples of the 27 patients collected at time points 3, 6, 12 and 24 months after HSCT passed proteomics quality control, yielding a total of 16 560 measurements excluding technical controls. Values below the validated limit of detection (LOD) ranged from 0% to 94% per protein, with 20% of proteins having more than 50% of measurements below LOD. We performed statistical analyses separately both with and without values below LOD, and no relevant or noticeable differences were found when comparing results presented with or without values below LOD. Therefore, presented results are based on actual data, including values below LOD. Four individual proteins were included in both panels used. To avoid their overrepresentation, only the proteins from panels containing fewer values below LOD were used in downstream analyses. Results of the shared proteins showed good correlation between panels (Supplementary figure 3), further providing evidence of good technical quality of the data obtained.

Proteome analysis

For proteome analysis we used the Proximity Extension Assay (Olink Proteomics, Uppsala, Sweden) on freshly thawed frozen plasma samples at time points 3, 6, 12 and 24 months after HSCT. In short, Proximity Extension Assay is a proteomics method where two target-specific oligo-conjugated antibodies bind their respective epitopes that are very closely situated on the target, allowing the oligos to hybridize. After this, only the hybridized oligonucleotide sequences are detected with quantitative PCR. The method is highly specific and enables a relative quantification of large amounts of protein targets simultaneously in a longitudinal sample series.[37] We used two Olink panels ("Inflammation" and "Immune response", Olink Proteomics, Uppsala, Sweden) consisting of 92 proteins each (Supplementary table 1). The Olink panels are commercially available, and designed and optimized by the Olink company.[38] Samples (1 µL) were incubated overnight at 4°C with Olink's Incubation mix (Olink Proteomics, Uppsala, Sweden), which contains probes against 92 biomarkers and 4 control reactions. The following day, the extension reaction was performed by incubating the reagents for 5 min at room temperature with the PEA enzyme (Olink Proteomics, Uppsala, Sweden), followed by performing PCR (PEA program) using a PCR polymerase (Olink Proteomics, Uppsala, Sweden). The concentration of biomarkers were detected using high-throughput quantitative PCR using Fluidigm's 96.96 Dynamic array and Biomark HD. Analysis service was provided by the Biomedicum Functional Genomics Unit at the Helsinki Institute of Life Science and Biocenter Finland at the University of Helsinki. Sample availability is displayed in Supplementary table 4.

Flow cytometry

For FCRL6 flow cytometric analysis, peripheral blood mononuclear cells from 17 patients collected at 3, 6, 12 and 24 months after transplantation were used as available (Supplementary table 2). Frozen aliquots of peripheral blood mononuclear cells were thawed in +37°C water bath and transferred to 50-mL conical tubes. Then, 10 mL of +37°C thawing solution [Roswell Park Memorial Institute (RPMI) supplemented with 10% CTL-WASH (Cellular Technology Limited, Shaker Heights, OH, USA) and 10 mg mL^{-1} DNase I] was added slowly, the tube was centrifuged at 350 *g*, 10 min in room temperature, and the supernatant was discarded. Cells were washed again with 10 mL of thawing solution at 350 *g* for 10 min, adjusting the temperature to +4°C during the centrifugation and resuspended in +4°C staining buffer (phosphate-buffered saline supplemented with 10% fetal calf serum and 2 mol L^{-1} ethylenediaminetetraacetic acid).

For each individual, 1–2 million cells were washed with 500 µL of staining buffer. The fluorescent-labeled antihuman antibodies are listed in Supplementary table 3. The optimal staining concentrations for antibodies were titrated with live peripheral blood mononuclear cells. Cells were incubated with a mix of antibodies and 1:1000 LIVE/DEAD Fixable Green Dead Cell Stain (Thermo Fisher Scientific, Waltham, MA, USA) diluted in Brilliant Stain Buffer (BD Biosciences,

Franklin Lakes, NJ, USA) in a final volume of 50 µL for 30 min in +4°C. Cells were washed with 1 mL of staining buffer and run on LSR II Fortessa (BD Biosciences, Franklin Lakes, NJ, USA). A median of 78 000 (range 8000–990 000) viable lymphocytes were registered per each run. Flow cytometric data were analyzed using FlowJo software (version 10.8; BD Biosciences, Franklin Lakes, NJ, USA) using fluorescence-minus-one staining for FCRL6 expression, and biological negative controls when applicable. The full gating strategy is displayed in the Supplementary figure 1.

Statistical methods

Protein concentrations were expressed as normalized protein expression values, an arbitrary unit on $\log_2$-scale. Values below quantitative PCR LOD were treated both by using actual obtained data and, because of the risk of bias and the risk of missing out on relevant associations, by setting values below LOD to missing. However, proteins with over 50% of measured values below LOD were excluded completely from further analyses. The multidimensional protein data were visualized by applying principal component analysis (Supplementary figure 2) on data at each individual time point. Time-course trends of individual proteins were visualized using sparkline plots generated by applying locally estimated scatterplot smoothing regression smoothing for the complication-free and complication-prone groups.

Unsupervised hierarchical clustering of protein levels was performed using the Euclidean distance and Ward's method independently at each time point. Before clustering, the normalized protein expression values were scaled using *z*-score normalization. Differences in outcome rates between identified clusters were compared using the chi-square or Fisher's exact test ($n < 5$). Differences in trends of IR between the clusters over time were analyzed using linear mixed effect models with cluster, time and cluster × time interaction term as fixed effects and including a random intercept for patients.

Global correlation matrices were computed and prepared for visualization using the R package DGCA (version 1.0.3). For each time point, the top 10 protein pairs with the greatest absolute difference in correlation between complication-free and complication-prone patients were reported. Differentially correlating protein pairs were identified using the R package DiffCorr (version 0.4.3).

Differentially expressed markers for outcomes of interest were identified using R package maSigPro (version 1.70.0), a regression-based method designed for identifying differential expression profiles between experimental groups in time-course experiments.[39] Here, up to third-degree polynomial models were considered and the obtained *P*-values were corrected for multiple comparisons by applying the Bonferroni correction. For each model, we used $R^2 = 0.5$ as a threshold for acceptance into the group of significant proteins. In addition, differential expression for outcomes of interest at each individual time point was tested using the Bonferroni-corrected Mann–Whitney *U*-test.

All statistical analyses and modeling were carried out using the R statistical computing environment version 4.0.3 (R Core Team, 2016, Vienna, Austria; https://www.R-project.org/). The level of significance was set at $P < 0.05$.

ACKNOWLEDGMENTS

We thank research nurse Pia Valle, and lab technicians Marjo Rissanen, Tamas Bazsinka and Turku University for indispensable assistance.

AUTHOR CONTRIBUTIONS

Adam Alexandersson: Conceptualization; data curation; formal analysis; investigation; methodology; project administration; resources; visualization; writing – original draft; writing – review and editing. **Mikko S Venäläinen:** Formal analysis; methodology; software; visualization; writing – original draft; writing – review and editing. **Nelli Heikkilä:** Formal analysis; investigation; visualization; writing – original draft; writing – review and editing. **Xiaobo Huang:** Formal analysis; writing – review and editing. **Mervi Taskinen:** Resources; writing – original draft; writing – review and editing. **Pasi Huttunen:** Conceptualization; resources; writing – original draft; writing – review and editing. **Laura L Elo:** Methodology; resources; supervision; writing – review and editing. **Minna Koskenvuo:** Project administration; resources; supervision; writing – review and editing. **Eliisa Kekäläinen:** Conceptualization; funding acquisition; project administration; supervision; writing – original draft; writing – review and editing.

CONFLICT OF INTEREST

AA reports funding from the foundations Finska Läkaresällskapet and Lastentautien Tutkimussäätiö. MSV reports funding from the Academy of Finland (grant number 322123). EK reports funding from Lastentautien Tutkimussäätiö, University of Helsinki research funds and Helsinki University Hospital Governmental Funding for Research. LLE reports funding from the European Research Council ERC (677943), European Union's Horizon 2020 Research and Innovation Programme (955321), Academy of Finland (310 561, 314 443, 329 278, 335 434, 335 611 and 341 342) and Sigrid Jusélius Foundation. LLE's research is also supported by Biocenter Finland and ELIXIR Finland. NH, XH, MT, MK and PH report no competing financial interests.

DATA AVAILABILITY STATEMENT

The proteomic data that support the findings of this study can be found in Supplementary tables 6–13.

REFERENCES

1. Wingard JR, Majhail NS, Brazauskas R, *et al.* Long-term survival and late deaths after allogeneic hematopoietic cell transplantation. *J Clin Oncol* 2011; **29**: 2230–2239.

2. Adom D, Rowan C, Adeniyan T, Yang J, Paczesny S. Biomarkers for allogeneic HCT outcomes. *Front Immunol* 2020; **11**: 673.
3. de Koning C, Plantinga M, Besseling P, Boelens JJ, Nierkens S. Immune reconstitution after allogeneic hematopoietic cell transplantation in children. *Biol Blood Marrow Transplant* 2016; **22**: 195–206.
4. Blennow O, Ljungman P, Sparrelid E, Mattsson J, Remberger M. Incidence, risk factors, and outcome of bloodstream infections during the pre-engraftment phase in 521 allogeneic hematopoietic stem cell transplantations. *Transpl Infect Dis* 2014; **16**: 106–114.
5. Zhang P, Jiang EL, Yang DL, *et al*. Risk factors and prognosis of invasive fungal infections in allogeneic stem cell transplantation recipients: a single-institution experience. *Transpl Infect Dis* 2010; **12**: 316–321.
6. Haidar G, Boeckh M, Singh N. Cytomegalovirus infection in solid organ and hematopoietic cell transplantation: state of the evidence. *J Infect Dis* 2020; **221**: S23–S31.
7. Gottlieb DJ, Clancy LE, Withers B, *et al*. Prophylactic antigen-specific T-cells targeting seven viral and fungal pathogens after allogeneic haemopoietic stem cell transplant. *Clin Transl Immunology* 2021; **10**: e1249.
8. Greenberg PD, Reusser P, Goodrich JM, Riddell SR. Development of a treatment regimen for human cytomegalovirus (CMV) infection in bone marrow transplantation recipients by adoptive transfer of donor-derived CMV-specific T cell clones expanded *in vitro*. *Ann N Y Acad Sci* 1991; **636**: 184–195.
9. Shafat MS, Mehra V, Peggs KS, Roddie C. Cellular therapeutic approaches to cytomegalovirus infection following allogeneic stem cell transplantation. *Front Immunol* 2020; **11**: 1694.
10. Stern L, Withers B, Avdic S, *et al*. Human cytomegalovirus latency and reactivation in allogeneic hematopoietic stem cell transplant recipients. *Front Microbiol* 2019; **10**: 1186.
11. Camargo JF, Wieder ED, Kimble E, *et al*. Deep functional immunophenotyping predicts risk of cytomegalovirus reactivation after hematopoietic cell transplantation. *Blood* 2019; **133**: 867–877.
12. Chen S, Zeiser R. Novel biomarkers for outcome after allogeneic hematopoietic stem cell transplantation. *Front Immunol* 2020; **11**: 1854.
13. Giaccone L, Faraci DG, Butera S, *et al*. Biomarkers for acute and chronic graft versus host disease: state of the art. *Expert Rev Hematol* 2021; **14**: 79–96.
14. Davis MM. Systems immunology. *Curr Opin Immunol* 2020; **65**: 79–82.
15. Alexandersson A, Koskenvuo M, Tiderman A, *et al*. Viral infections and immune reconstitution interaction after pediatric allogenic hematopoietic stem cell transplantation. *Infect Dis* 2019; **51**: 772–778.
16. Burrage PS, Mix KS, Brinckerhoff CE. Matrix metalloproteinases: role in arthritis. *Front Biosci* 2006; **11**: 529–543.
17. Wang YD, Yan PY. Expression of matrix metalloproteinase-1 and tissue inhibitor of metalloproteinase-1 in ulcerative colitis. *World J Gastroenterol* 2006; **12**: 6050–6053.
18. Montaño M, Sansores RH, Becerril C, *et al*. FEV1 inversely correlates with metalloproteinases 1, 7, 9 and CRP in COPD by biomass smoke exposure. *Respir Res* 2014; **15**: 74.
19. Rumbaugh J, Turchan-Cholewo J, Galey D, *et al*. Interaction of HIV tat and matrix metalloproteinase in HIV neuropathogenesis: a new host defense mechanism. *FASEB J* 2006; **20**: 1736–1738.
20. Uhlén M, Karlsson MJ, Hober A, *et al*. The human secretome. *Sci Signal* 2019; **12**: eaaz0274.
21. Holtan SG, DeFor TE, Panoskaltsis-Mortari A, *et al*. Amphiregulin modifies the Minnesota acute graft-versus-host disease risk score: results from BMT CTN 0302/0802. *Blood Adv* 2018; **2**: 1882–1888.
22. Holtan SG, Khera N, Levine JE, *et al*. Late acute graft-versus-host disease: a prospective analysis of clinical outcomes and circulating angiogenic factors. *Blood* 2016; **128**: 2350–2358.
23. Schreeder DM, Cannon JP, Wu J, Li R, Shakhmatov MA, Davis RS. Cutting edge: FcR-like 6 is an MHC class II receptor. *J Immunol* 2010; **185**: 23–27.
24. Kulemzin SV, Zamoshnikova AY, Yurchenko MY, *et al*. FCRL6 receptor: expression and associated proteins. *Immunol Lett* 2011; **134**: 174–182.
25. Davis RS. Roles for the FCRL6 immunoreceptor in tumor immunology. *Front Immunol* 2020; **11**: 575175.
26. Johnson DB, Nixon MJ, Wang Y, *et al*. Tumor-specific MHC-II expression drives a unique pattern of resistance to immunotherapy via LAG-3/FCRL6 engagement. *JCI Insight* 2018; **3**: e120360.
27. Schreeder DM, Pan J, Li FJ, Vivier E, Davis RS. FCRL6 distinguishes mature cytotoxic lymphocytes and is upregulated in patients with B-cell chronic lymphocytic leukemia. *Eur J Immunol* 2008; **38**: 3159–3166.
28. Wilson TJ, Presti RM, Tassi I, Overton ET, Cella M, Colonna M. FcRL6, a new ITIM-bearing receptor on cytolytic cells, is broadly expressed by lymphocytes following HIV-1 infection. *Blood* 2007; **109**: 3786–3793.
29. Arruda LCM, Gaballa A, Uhlin M. Impact of γδ T cells on clinical outcome of hematopoietic stem cell transplantation: systematic review and meta-analysis. *Blood Adv* 2019; **3**: 3436–3448.
30. Park M, Im HJ, Lee YJ, *et al*. Reconstitution of T and NK cells after haploidentical hematopoietic cell transplantation using αβ T cell-depleted grafts and the clinical implication of γδ T cells. *Clin Transpl* 2018; **32**: e13147.
31. Gaballa A, Stikvoort A, Önfelt B, *et al*. T-cell frequencies of $CD8^+\gamma\delta$ and $CD27^+\gamma\delta$ cells in the stem cell graft predict the outcome after allogeneic hematopoietic cell transplantation. *Bone Marrow Transplant* 2019; **54**: 1562–1574.
32. Prinz I, Koenecke C. Antigen-specific γδ T cells contribute to cytomegalovirus control after stem cell transplantation. *Curr Opin Immunol* 2023; **82**: 102303.
33. Deseke M, Rampoldi F, Sandrock I, *et al*. A CMV-induced adaptive human $V\delta1^+\gamma\delta$ T cell clone recognizes HLA-DR. *J Exp Med* 2022; **219**: e20212525.

34. Handgretinger R, Schilbach K. The potential role of γδ T cells after allogeneic HCT for leukemia. *Blood* 2018; **131**: 1063–1072.
35. Jagasia MH, Greinix HT, Arora M, *et al.* National Institutes of Health consensus development project on criteria for clinical trials in chronic graft-versus-host disease: I. The 2014 diagnosis and staging working group report. *Biol Blood Marrow Transplant* 2015; **21**: 389–401.
36. Glucksberg H, Storb R, Fefer A, *et al.* Clinical manifestations of graft-versus-host disease in human recipients of marrow from HL-A-matched sibling donors. *Transplantation* 1974; **18**: 295–304.
37. Assarsson E, Lundberg M, Holmquist G, *et al.* Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* 2014; **9**: e95192.
38. Assarsson E, Lundberg M. Development and Validation of Customized PEA Biomarker Panels with Clinical Utility, In "Advancing Precision Medicine: Current and Future Proteogenomic Strategies for Biomarker Discovery and Development", produced by the *Science*/AAAS Custom Publishing Office, Washington, DC. *Science* Supplement; 2017:32–36.
39. Conesa A, Nueda MJ, Ferrer A, Talón M. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* 2006; **22**: 1096–1102.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.