


RESEARCH

Open Access



Whole blood RNA sequencing reveals a unique transcriptomic profile in patients with ARDS following hematopoietic stem cell transplantation

Joshua A. Englert^{1*†} , Michael H. Cho^{2,3†}, Andrew E. Lamb², Maya Shumyatcher⁴, Diana Barragan-Bradford³, Maria C. Basil³, Angelica Higuera³, Colleen Isabelle³, Mayra Pinilla Vera³, Paul B. Dieffenbach³, Laura E. Fredenburgh³, Joyce B. Kang⁵, Ami S. Bhatt⁵, Joseph H. Antin⁶, Vincent T. Ho⁶, Robert J. Soiffer⁶, Judie A. Howrylak⁷, Blanca E. Himes^{4†} and Rebecca M. Baron^{3†}

Abstract

Background: The acute respiratory distress syndrome (ARDS) is characterized by the acute onset of hypoxemia and bilateral lung infiltrates in response to an inciting event, and is associated with high morbidity and mortality. Patients undergoing allogeneic hematopoietic stem cell transplantation (HSCT) are at increased risk for ARDS. We hypothesized that HSCT patients with ARDS would have a unique transcriptomic profile identifiable in peripheral blood compared to those that did not undergo HSCT.

Methods: We isolated RNA from banked peripheral blood samples from a biorepository of critically ill ICU patients. RNA-Seq was performed on 11 patients with ARDS (5 that had undergone HSCT and 6 that had not) and 12 patients with sepsis without ARDS (5 that had undergone HSCT and 7 that had not).

Results: We identified 687 differentially expressed genes between ARDS and ARDS-HSCT (adjusted p -value < 0.01), including *IFI44L*, *OAS3*, *LY6E*, and *SPATS2L* that had increased expression in ARDS vs. ARDS-HSCT; these genes were not differentially expressed in sepsis vs sepsis-HSCT. Gene ontology enrichment analysis revealed that many differentially expressed genes were related to *response to type I interferon*.

Conclusions: Our findings reveal significant differences in whole blood transcriptomic profiles of patients with non-HSCT ARDS compared to ARDS-HSCT patients and point toward different immune responses underlying ARDS and ARDS-HSCT that contribute to lung injury.

Keywords: Acute respiratory distress syndrome (ARDS), RNA sequencing, RNA-Seq, Transcriptome profiling, Hematopoietic stem cell transplantation, Respiratory failure, Bone marrow transplant

* Correspondence: joshua.englert@osumc.edu

[†]Joshua A. Englert, Michael H. Cho, Blanca E. Himes and Rebecca M. Baron contributed equally to this work.

¹Division of Pulmonary, Critical Care, and Sleep Medicine, The Ohio State Wexner Medical Center, 201 Davis Heart and Lung Research Institute, 473 West 12th Avenue, Columbus, OH 43210, USA

Full list of author information is available at the end of the article



Background

ARDS is a syndrome with high morbidity and mortality characterized by the acute onset of bilateral lung infiltrates and hypoxemia that frequently results in acute respiratory failure [1]. A recent study demonstrated that the worldwide burden of ARDS is greater than previously appreciated and accounts for up to 10% of intensive care unit (ICU) admissions [2]. Despite decades of research into the pathophysiology underlying ARDS, supportive care with mechanical ventilation remains the mainstay of therapy, and there are no effective targeted pharmacologic treatments for these patients. One barrier to developing effective therapies is that ARDS is a heterogeneous syndrome that encompasses a wide variety of patients with lung injury from many different causes. We and other groups have postulated that the pathophysiology of lung injury in ARDS subgroups may differ substantially which, in turn, may lead to variable responses in ARDS clinical trials [3–6]. Recently, several studies have used clinical data and previously identified biomarkers to define several ARDS subphenotypes with distinct clinical outcomes, inflammatory profiles, and response to ARDS therapies [3, 7, 8] underscoring the need for a better understanding of these different populations.

One important subgroup of patients at increased risk for ARDS that experience particularly high mortality are those with hematologic malignancies undergoing hematopoietic stem cell transplantation. While allogeneic hematopoietic stem cell transplantation (allo-HSCT) is a life-saving treatment for patients with hematologic malignancies [9], a variety of pulmonary complications, both infectious and non-infectious, can complicate the post-transplant course [10, 11]. One report found that ARDS occurs in over 15% of patients undergoing allo-HSCT with ICU mortality rates of nearly 50% [12]. It remains unclear whether the pathobiology of ARDS following HSCT is similar to ARDS in other patient populations, even though ARDS post HSCT is often clinically indistinguishable from ARDS in non-HSCT patients. A better understanding of the biology of ARDS following allo-HSCT may lead to novel strategies for prevention and treatment in this population and, in addition, might identify broad treatment approaches for ARDS.

Previous transcriptomic studies have used expression microarrays to search for ARDS gene-expression signatures. Using whole blood mRNA, Howrylak et al. identified a signature of 8 genes in ICU patients with sepsis that was associated with the development of ARDS [13]. A microarray study of septic patients found increased expression of neutrophil related genes early in ARDS [14], and our group demonstrated the critical role of IL18, a prominent inflammasome-related cytokine, in the pathogenesis of sepsis-induced ARDS [15]. In a study by Juss et al., peripheral blood neutrophils from

ARDS patients were compared to those of healthy volunteers, and patients with ARDS had significant increases in immune response pathways [16]. Published ARDS transcriptomic studies have not characterized different subgroups of patients with ARDS, an approach that may reveal distinct gene expression changes that underlie similar clinical symptoms. Here, we used RNA-Seq, a powerful and unbiased approach to characterize transcriptomes, to compare gene expression in the blood of patients with ARDS following allo-HSCT compared to patients with ARDS that had not undergone allo-HSCT.

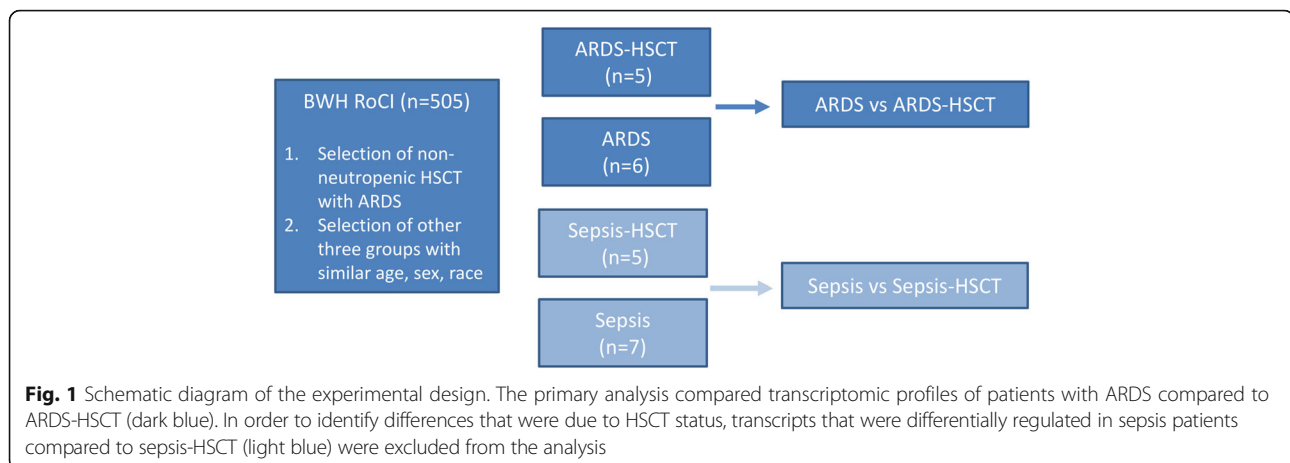
Methods

Study design and sample collection

Patients with ARDS following HSCT who contributed samples to the IRB-approved Brigham and Women's Registry of Critical Illness (RoCI) from 2008 to 2013 were identified [6, 17, 18]. All patients were admitted to the medical intensive care unit at Brigham and Women's Hospital at the time of enrollment in the RoCI. Subjects with ARDS that did not undergo HSCT from the same 5 year period were then identified as a comparator group. To reduce sample heterogeneity, we selected subjects of similar age, time in the ICU, male sex, and white race. ARDS was defined by the Berlin [1] definition for cases after 2012 and the American-European Consensus (AECC) definition [19] for cases prior to 2012. While our primary interest was identifying differences in gene expression in ARDS subjects that had undergone HSCT compared those without HSCT, we also sequenced 12 subjects with sepsis (HSCT and non-HSCT) without ARDS in order to identify differences between the ARDS groups that were not due to HSCT status. Sepsis was defined by the presence of the systemic inflammatory response syndrome (SIRS) with suspected or confirmed infection [20]. Whole blood samples from the beginning of the ICU course (within the first 5 days of ICU admission) were used for gene expression profiling. Our overall study design is shown in Fig. 1.

RNA-Seq library construction and sequencing

RNA was extracted and globin-cleared using the Ambion GLOBINclear kit (Thermo Fisher, Waltham, MA). RNA-Seq libraries were prepared with 0.1–1 µg total RNA using Illumina TruSeq RNA-Seq v2 kit (Illumina, San Diego, CA) according to manufacturer protocol. Ambion External RNA Controls Consortium (ERCC) RNA Spike-In Control Mix 1 (Thermo Fisher, Waltham, MA) was added to the samples. Quality control of libraries involved picogreen and size analysis on an Agilent Bioanalyzer or TapeStation 2200 (Agilent, Santa Clara, CA) and qPCR quantitation against a standard curve. Samples with an initial low RIN (i.e. < 6) were re-extracted; any samples with a persistent RIN < 6 were run in duplicate. Samples



were randomized to one of 4 pools of 5–6 subjects, and each pool was sequenced in two different lanes. Sequencing of 75 base pair, paired-end reads was performed with an Illumina HiSeq 2500 instrument at Partners Personalized Medicine (Boston, MA) under rapid mode using PhiX spike-in and processed using Real-Time Analysis v1.18.64, Control Software v2.2.58 (Illumina, San Diego, CA).

RNA-Seq data analysis

Preliminary processing of raw reads was performed using Casava 1.8 (Illumina, San Diego, CA). Reads were aligned to the human genome build 38 with STAR v2.4.0. [21] Taffeta scripts (<https://github.com/blancahimes/taffeta>) were used to assess quality of aligned reads, which included quantifying number of mapped reads, junction spanning reads, assessing 5' and 3' bias, and insert size distribution. For each sample, ERCC Spike-in dose response curves (i.e. plots of ERCC transcript FPKM vs. ERCC transcript molecules) were created following the manufacturer's protocol [22]. Because samples with low RIN (<6) had characteristics similar to those of higher quality and duplicate samples were highly correlated, we combined data for duplicate samples and included them in further analyses. Raw read plots were created by displaying bigwig files for each sample in the UCSC Genome Browser. The RNA-Seq data is available at the Gene Expression Omnibus Web site (<http://www.ncbi.nlm.nih.gov/geo/>) under accession GSE84439. Gene counts were obtained with HTSeq v0.6.1. Differential expression analysis was performed using DESeq2 v1.12.3 [23], after excluding reads mapped to hemoglobin genes and while adjusting for RIN scores as a covariate [24]. The NIH Database for Annotation, Visualization and Integrated Discovery (DAVID) was used to perform gene ontological category enrichment analysis using *Homo sapiens* as background, and default options and annotation categories [25].

qPCR validation of differentially expressed genes

RNA (200 ng per subject) was used to prepare cDNA utilizing the Applied Biosciences High Capacity cDNA Reverse Transcriptase Kit in 20 μ l reactions (Thermo Fisher, Waltham, MA). The cDNA was then used for qPCR analysis with the following Taqman/Applied Biosciences (Thermo Fisher, Waltham, MA) primers (all human and all primers chosen to span an exon junction): *IFI44L* (Hs00915292_m1), *OAS2* (Hs00942643_m1), *OAS3* (Hs00196324_m1), *SPATS2L* (Hs91916364_m1). Briefly, qPCR was performed with a 1:5 dilution of cDNA and run for 40 cycles. All samples were run in triplicate, and GAPDH was used as a housekeeping control. All samples had SD < 0.5 between triplicates. Controls without cDNA templates, primers, and reverse transcriptase were included and were all appropriately negative. Delta CT values between genes of interest and GAPDH were analyzed for statistical significance using a t-test [26].

Analysis of unmapped reads

Fastq files containing paired-end reads were merged for each sample (i.e., forward and reverse sequence files were combined). Reads were trimmed with Trim Galore using a minimum Phred quality score of 20 and minimum final read length of 60 bp [27]. Trimmed reads were de-duplicated to remove PCR artifacts and sequences containing at least one "N" was performed with Super-Deduper [28]. Reads of potential human origin were removed by aligning resultant fastq files to GRCh37 with STAR using the recommended default parameters [21]. The unmapped (non-human) reads were taxonomically classified based on mapping all *k*-mers (*k* = 31) within each sequence to the lowest common ancestor in a database of > 83,000 bacterial, viral, fungal, and protozoan genomes in the OneCodex reference database (database version as of February 26, 2018). Default OneCodex alignment parameters were used [29].

Results

Subject characteristics

Characteristics of all subjects are shown in Table 1. To reduce sources of gene expression heterogeneity not related to disease status, selected subjects were age-matched, non-Hispanic white males with similar WBC, APACHE scores, and severity of hypoxemia (PaO₂/FiO₂ ratios) (Wilcoxon $P > 0.1$). For the HSCT-ARDS group, all patients had undergone HSCT for the treatment of various forms of acute leukemia and median time from transplant to development of ARDS was 117 days. All patients with ARDS received supportive care with low tidal volume ventilation and average tidal volumes did not differ between

the ARDS and ARDS-HSCT at the time that blood samples were collected (Table 1). None of the ARDS subjects were treated with neuromuscular blockade or prone positioning at the time the blood sample used for RNA sequencing was collected. All HSCT-ARDS subjects were treated with corticosteroids (median dose 80 mg methylprednisolone daily), as were half of ARDS subjects (median dose 16 mg methylprednisolone daily). Similarly, in the non-ARDS control groups with sepsis, all of the sepsis-HSCT and half of the sepsis subjects were treated with corticosteroids. Etanercept was given to 4 of the 5 ARDS-HSCT patients as treatment for idiopathic pneumonia syndrome. Two ARDS subjects were confirmed to

Table 1 Baseline characteristics of the subjects

Characteristics	ARDS	ARDS-HSCT	Sepsis	Sepsis-HSCT	P (all groups)	P (ARDS vs ARDS-HSCT)
N	6	5	7	5		
Age	46 (37–60)	39 (25–52)	50 (26–59)	55 (61–42)	0.27	0.41
WBC	6.9 (3.1–16.8)	11 (2.8–19.1)	9 (4–42.1)	8.6 (6.5–29.9)	0.88	0.86
% Neutrophils	87 (71–98)	95 (88–98)	86 (63–95)	80 (65–96)	0.14	0.10
% Lymphocytes	6 (1–18)	2 (0–10)	10 (1–21)	10 (1–21)	0.26	0.12
% Monocytes	4 (1–15)	2 (0–4)	4 (1–11)	9 (0–11)	0.39	0.46
Sepsis	33.3%	40.0%	100%	100%		1
Vasopressors	50%	60%	29%	40%	0.8	1
Days since HSCT		117 (50–490)		148 (50–1322)		
APACHE II	27.5 (21–47)	30.0 (24–38)	18 (14–40)	19 (14–32)	0.10	0.58
ICU Mortality	67%	100%	14%	0%	0.001	0.45
Mechanical Ventilation	100%	100%	29%	40%	0.04	1
P/F ratio	70.0 (41–205.0)	81.3 (66.0–109.2)				0.83
Tidal volume (mL/kg)	5.6 (5.2–7.6)	6.5 (4.5–16.5)				0.46
Corticosteroids	50%	100.0%	57%	100%	0.04	0.18
Etanercept	0.0%	80.0%	0%	0%		0.02
Other Immunosuppression	50.0%	80%	0%	60%	0.02	0.5
ARDS etiology						
Pneumonia, influenza	33%	0%				
Pancreatitis	17%	0%				
Idiopathic pneumonia syndrome	0%	100%				
Idiopathic/CVD	50%	0%				
Infection						
Type						
Bacterial	0%	0%	100%	80%		
Viral	33%	0%	0%	20%		
Source						
Blood	0%	0%	43%	60%		
Lung	33%	0%	14%	20%		
GI	0%	0%	57%	0%		

Idiopathic/collagen-vascular disease (CVD) diagnoses include acute interstitial pneumonia, cryptogenic organizing pneumonia, and collagen-vascular disease associated interstitial lung disease. Values given as median (range)

have influenza infection (Table 1). Patients with sepsis were managed with source control, early antibiotic administration, and volume resuscitation in accordance with sepsis guidelines at the time [30, 31].

Immune response and interferon-stimulated genes are upregulated in ARDS compared to ARDS-HSCT

We identified 687 differentially expressed genes between ARDS and ARDS-HSCT (false-discovery rate (FDR) corrected p -value < 0.01). In ARDS-HSCT vs. ARDS, we found 687 differentially expressed genes (520 were upregulated, 167 were downregulated, Fig. 2). To identify which differences in gene expression reflected unique features of ARDS-HSCT, rather than the history of HSCT itself, we filtered out genes that were also nominally differentially expressed (unadjusted p -value < 0.1) in a consistent direction in patients with sepsis-HSCT vs. sepsis. Of the 687 differentially expressed genes, 483 were not present in the sepsis-HSCT vs. sepsis comparison, suggesting that the majority of expression changes reflected processes unique to ARDS-HSCT. Top differentially expressed genes upregulated in ARDS vs. ARDS-HSCT included immune response and interferon signaling genes (e.g. *IFI44L*, *OAS3*, *LY6E*, *OAS2*, *USP18*) as shown in Table 2. Gene ontological category enrichment analysis with all 687 differentially expressed genes found that *immune response*- and *interferon*

signaling-related pathways were statistically overrepresented (Table 3 and Additional file 1: Tables S1 to Table S3). To confirm top RNA-Seq results, we selected several of the most differentially expressed genes and performed qRT-PCR using an aliquot of whole blood RNA from the subjects that were sequenced. The qRT-PCR results were consistent with the RNA-Seq results and showed significant differences in expression for the transcripts that were tested (Additional file 2: Figure S1).

Differences in transcriptomic profiles are not due to influenza infection in ARDS subjects

To determine whether increased expression of immune response pathway genes in ARDS was due to viral infection, we repeated the differential expression analysis after excluding two patients with confirmed influenza infection (remaining ARDS subjects did not have any evidence of active viral infection while in the ICU). Results of this sub-analysis were consistent with those of the full set of subjects: of the top 100 genes differentially expressed in ARDS, 99 out of 100 were also differentially expressed after removal these two patients, and top results of the ontological category enrichment analysis were unchanged (Additional file 1: Table S2). Thus, it appears that increased expression of interferon-related genes in ARDS was not due to influenza infection.

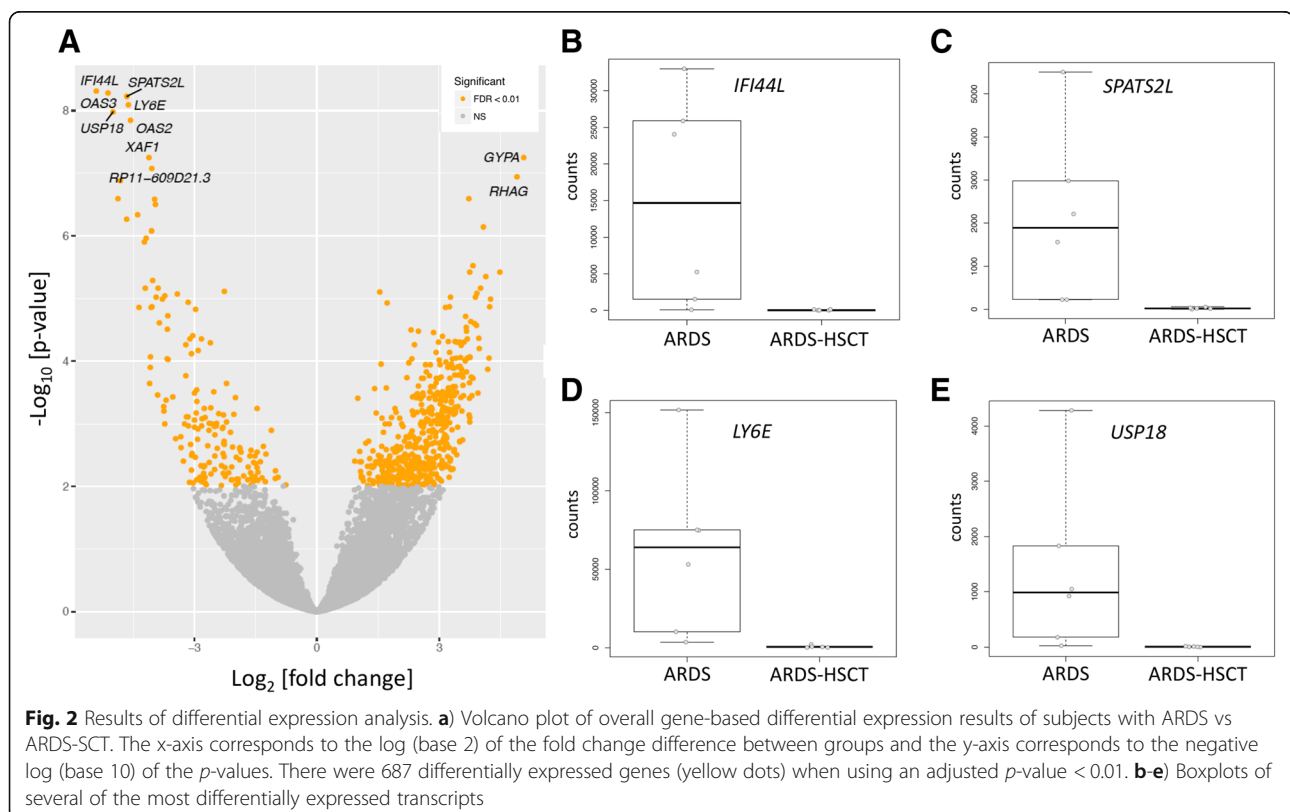


Table 2 Top 10 most differentially expressed transcripts from ARDS vs ARDS-HSCT

Gene	feature_name	log2FoldChange	Adjusted P
<i>IFI44L</i>	ENSG00000137959	-5.4	5.0E-09
<i>OAS3</i>	ENSG00000111331	-5.1	5.3E-09
<i>SPATS2L</i>	ENSG00000196141	-4.6	6.0E-09
<i>LY6E</i>	ENSG00000160932	-4.6	8.1E-09
<i>USP18</i>	ENSG00000184979	-5.0	1.1E-08
<i>OAS2</i>	ENSG00000111335	-4.6	1.4E-08
<i>XAF1</i>	ENSG00000132530	-4.1	5.7E-08
<i>GYPA</i>	ENSG00000170180	5.1	5.7E-08
<i>RP11-609D21.3</i>	ENSG00000279296	-4.0	8.4E-08
<i>RHAG</i>	ENSG00000112077	4.9	1.2E-07

Analysis of non-human transcripts did not show evidence of undiagnosed infection in patients with ARDS after HSCT

Some experts have speculated that lung injury following stem cell transplant may be due to infection with pathogens that are not identified using microbiologic techniques routinely used in clinical laboratories [32]. For example, next-generation sequencing was used to discover a novel bacterium as a candidate pathogen in cord colitis syndrome, a complication of allogeneic cord blood HSCT [33]. To investigate the possibility that undiagnosed infection accounts for differences observed in ARDS-HSCT, we used RNA-Seq reads that did not map to the human genome to perform genus-level taxonomic classification. After removing possible duplicate reads, we analyzed the remaining 8–19% of unmapped reads. Based on the genera present and relative similarity of microbial profiles across all 23 subjects, there were no significant differences between samples both within and among categories (Additional file 2: Figure S2). Additional file 1: Table S3 shows the number of reads per sample assignable to a virus. We detected the presence of virus in four subjects (with reads mapping to EBV and torquetenovirus). Only one subject had >10 reads assignable to a virus (with 97 reads aligning to EBV). The lack of recurrent abundant pathogens in the ARDS HSCT

Table 3 Top results from DAVID Functional Annotation Clustering

Cluster	Enrichment score	Representative category	Count	Adjusted P-value
1	11.05	Antiviral defense	30	3.9×10^{-19}
2	5.51	Response to virus	17	1.3×10^{-5}
3	3.84	Heme biosynthesis	7	2.8×10^{-5}
4	3.29	Hereditary hemolytic anemia	9	4.2×10^{-5}
5	3	Herpes simplex infection	16	1.8×10^{-2}

population is consistent with clinical data and suggested that these patients did not have undiagnosed bacterial or viral bloodstream infections as a proximate cause of their ARDS.

Discussion

We performed RNA-Seq using whole blood of patients with ARDS after HSCT and compared the transcriptomic profile to that of non-HSCT ARDS patients. Our analysis identified a set of differentially expressed genes with the strongest differences in pathways related to innate immune responses to pathogens and type I interferon signaling, all of which were upregulated in ARDS. Our findings raise the question of whether there may be an immune dysregulation post HSCT that predisposes subjects to lung injury or ARDS. Recently, Nick et al. found that high- and low- blood expression of a set of interferon-stimulated genes (*MX1*, *IFIT1*, and *ISG15*) identified a set of ARDS patients with worse outcomes. Morrell et al., analyzed peripheral blood monocyte gene expression by microarray and found increased interferon signaling in ARDS patients with a favorable outcome (i.e. more ventilator-free days) compared to ARDS patients with fewer ventilator-free days [34]. Thus, it is enticing to hypothesize that modulation of interferon signaling pathways represents a treatment approach for HSCT subjects, but additional data are required to support such a strategy. Interestingly, a recently completed phase 3 trial of human interferon beta 1a for moderate-to-severe ARDS did not find a significant difference in the primary composite endpoint of mortality and ventilator free days [35, 36]. Although the administration of interferon beta 1a was not efficacious in this population of unselected patients, it is possible that assessing whether interferon signaling pathways are activated in ARDS patients might provide a more targeted approach to treatment.

Our study is the first to use RNA-Seq in order to determine whether the specific ARDS subphenotype of HSCT-ARDS has a unique peripheral blood gene expression profile. Our patients were carefully selected from a single center and although sample size was small, this proof-of-concept study provides insights that are valuable to design future studies. We chose age- and sex-matched subjects to limit heterogeneity for this initial study, though we note that this may limit the generalizability of our findings. Another limitation of our study is the use of peripheral blood RNA, which is derived from a mixture of cell types. Although there were no significant differences in differential cell counts between groups, it is possible that gene expression profiles could be affected by subtle changes in cell maturity between groups. For example, blood gene expression profiles of patients with ARDS may have been

dominated by systemic inflammation when compared to subjects without ARDS [37]. We focused our study on two subgroups of ARDS patients, presumably with similar degrees of inflammation, in order to identify differences in gene expression that may explain heterogeneity amongst different types of ARDS patients. Our two groups were similar in demographic characteristics and severity of lung injury and overall illness, but still demonstrated clinical heterogeneity. While our results suggest that gene expression differences were not due to HSCT itself, we are unable to rule out the effect of potential confounders such as immunosuppressant medications. Notably, all of the patients in the ARDS-HSCT group were being treated with moderate-to-high dose corticosteroids at the time of inclusion, whereas half of the non-HSCT ARDS patients were on moderate dose corticosteroids. Another potential confounder is the possibility of undiagnosed infections. To address this issue, we analyzed reads from our RNA-Seq data that did not map to the human genome and performed taxonomic classification to look for the presence of common pathogens. Although our data did not suggest the presence of an undiagnosed bloodstream infection, this analysis was limited by the sample size, the fact that some respiratory pathogens might not exhibit a bloodstream signature, and the fact that the majority of bacterial RNA are not polyadenylated and may have been eliminated during the library preparation process. Finally, our samples were collected over the course of five years during which a new definition of ARDS was adopted, [1] and thus, it is possible that multiple definitions could have unintentionally impacted our results.

Patients who have undergone HSCT are at heightened risk for a number of types of lung injury including pulmonary infections as well as non-infectious lung injury such as drug-induced pneumonitis, diffuse alveolar hemorrhage, and idiopathic pneumonia syndrome. Given the lack of well-defined diagnostic criteria for many of these entities, other investigators have focused on the epidemiology of the most severe form of lung injury (i.e. ARDS) regardless of the etiology and found that ARDS after allo-HSCT is common and has high mortality [12]. We used a similar strategy for our study choosing not to focus on the proximal cause of lung injury. One limitation of this strategy is that it is difficult to determine whether the transcriptomic changes in the ARDS-HSCT population reflect the different underlying etiologies of ARDS, differential host responses, or most likely both. Interestingly, the majority of patients in the HSCT-ARDS group were suspected of having idiopathic pneumonia syndrome and received treatment with etanercept. While the effect of etanercept on peripheral blood gene expression in critical illness is not known, studies in autoimmune diseases have found differential

regulation of interferon pathways. For example, *OAS1*, *OAS2*, *OAS3*, and *IFI44L* appear to be down-regulated by etanercept in responders to treatment for psoriasis [38]. However, in a set of patients treated with infliximab for rheumatoid arthritis, treatment induced an increase in interferon response genes including *OAS1* and *OAS2* in a subset of patients who had poor clinical response; [39] a subsequent study including patients on a variety of anti-TNF therapies including etanercept also found divergent responses in interferon profiles, though did not find an association with clinical response [40].

Conclusions

ARDS is a heterogeneous syndrome, with a diverse set of risk factors, etiologies, and outcomes. Our study is the first to identify a potential difference in gene expression that distinguishes ARDS-HSCT from ARDS, representing changes in immune response and interferon signaling pathways. Our data, along with existing literature, suggest that there is a spectrum of dysregulated interferon signaling underlying lung injury and further supports the notion that improved phenotyping of ARDS subgroups is critical for designing targeted treatments for this devastating syndrome. Importantly, our data raises the intriguing question of whether there exists a modifiable immune deficiency post-HSCT that ameliorates lung injury. Future studies to examine these areas are critically important.

Additional files

Additional file 1: Table S1. Individual gene names from David functional clustering analysis. **Table S2.** Top results from DAVID Functional Annotation Clustering, excluding 2 subjects with diagnosed influenza. **Table S3.** Number of reads per sample assignable to a virus. (DOCX 24 kb)

Additional file 2: Figure S1. Transcript levels of several of the most differentially expressed genes were measured by qPCR to confirm changes found by RNA-seq. $n = 3/\text{group}$, $*p < 0.05$ vs ARDS by unpaired t-test. **Figure S2.** Taxonomic classification of unmapped reads within each patient sample after human read removal. Relative read count indicates the fraction of reads belonging to a given genus out of all reads classified to at least genus-level specificity. Genera with minimum read fraction of 2% within a sample are displayed (otherwise collapsed into "Other"). No dramatic classwise differences are observed, and the lack of a recurrent abundant pathogen unique to ARDS-SCT patients suggests that undiagnosed bloodstream infection is an unlikely cause of ARDS in these SCT patients. (DOCX 549 kb)

Abbreviations

allo HSCT: Allogeneic hematopoietic stem cell transplant; ARDS: Acute respiratory distress syndrome; ICU: Intensive care unit; RNA-seq: RNA sequencing

Acknowledgements

The authors would like to acknowledge the following members of the BWH Registry of Critical Illness: Samuel Ash, MD; Carolina Quintana, MD; Augustine M.K. Choi, MD, and Anthony Massaro, MD.

Funding

Supported, in part, by NIH grants K08 GM102695 (JAE), R56 HL142767 (JAE), R01 HL135142 (MHC), R01 HL133433 (BEH), R01 HL141992 (BEH), R01

HL112747 (RMB). This study was also supported in part by the Marshall Neilson Fund for GVHD Research. Dr. Cho was supported in part by a grant from Glaxo Smith Kline. The funding agencies did not have any role in the design of the study, the collection and analysis of data, or the preparation of the manuscript.

Availability of data and materials

The datasets generated and analyzed during the current study have been deposited in the Gene Expression Omnibus database under accession GSE84439.

Authors' contributions

JAE, MHC, BEH, and RMB conceived and designed the study. DBB, MCC, AH, CI, PBD, MPV, and LEF collected the biospecimens and clinical data and performed the qPCR validation. JAE, MHC, MS, BEH, and RMB analyzed and interpreted the RNA sequencing data. JHA, VTO, RJS, and JAH assisted with study design and interpretation of gene expression data. JBK and ASB performed the analysis of the non-human reads. JAE, MHC, BEH, and RMB drafted and revised the manuscript. All authors read and approved the final manuscript.

Ethics approval and consent to participate

The samples used for this study were obtained from the Registry of Critical Illness which is an IRB approved biorepository at Brigham and Women's Hospital. Consent is obtained from each subject or their legally authorized representative at the time of enrollment. The Partners Healthcare Human Research Committee originally provided initial approval in June 2008 and the protocol was most recently approved in October 2017 (Protocol number 2008P000495).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Division of Pulmonary, Critical Care, and Sleep Medicine, The Ohio State Wexner Medical Center, 201 Davis Heart and Lung Research Institute, 473 West 12th Avenue, Columbus, OH 43210, USA. ²Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115, USA. ³Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA. ⁴Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, 402 Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, USA. ⁵Departments of Medicine and Genetics, Stanford University, CCSR1155b, 269 Campus Drive, Palo Alto, CA 93405, USA. ⁶Division of Hematologic Malignancies, Dana-Farber Cancer Institute, Harvard Medical School, 450 Brookline Avenue, Boston, MA 02215, USA. ⁷Division of Pulmonary and Critical Care Medicine, Penn State Milton S. Hershey Medical Center, 500 University Drive, Hershey, PA 17033, USA.

Received: 27 July 2018 Accepted: 8 January 2019

Published online: 21 January 2019

References

- Force ADT, Ranieri VM, Rubenfeld GD, Thompson BT, Ferguson ND, Caldwell E, Fan E, Camporota L, Slutsky AS. Acute respiratory distress syndrome: the Berlin definition. *JAMA*. 2012;307:2526–33.
- Bellani G, Laffey JG, Pham T, Fan E, Brochard L, Esteban A, Gattinoni L, van Haren F, Larsson A, McAuley DF, et al. Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries. *JAMA*. 2016;315:788–800.
- Calfee CS, Delucchi K, Parsons PE, Thompson BT, Ware LB, Matthay MA, Network NA. Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. *Lancet Respir Med*. 2014;2:611–20.
- Calfee CS, Janz DR, Bernard GR, May AK, Kangelaris KN, Matthay MA, Ware LB, Network NNA. Distinct molecular phenotypes of direct vs indirect ARDS in single-center and multicenter studies. *Chest*. 2015; 147:1539–48.
- Reilly JP, Bellamy S, Shashaty MG, Gallop R, Meyer NJ, Lanken PN, Kaplan S, Holena DN, May AK, Ware LB, Christie JD. Heterogeneous phenotypes of acute respiratory distress syndrome after major trauma. *Ann Am Thorac Soc*. 2014;11:728–36.
- Putman RK, Hunninghake GM, Dieffenbach PB, Barragan-Bradford D, Serhan K, Adams U, Hatabu H, Nishino M, Padera RF, Fredenburgh LE, et al. Interstitial lung abnormalities are associated with acute respiratory distress syndrome. *Am J Respir Crit Care Med*. 2017;195:138–41.
- Delucchi K, Famous KR, Ware LB, Parsons PE, Thompson BT, Calfee CS, Network A. Stability of ARDS subphenotypes over time in two randomised controlled trials. *Thorax*. 2018;73:439–45.
- Famous KR, Delucchi K, Ware LB, Kangelaris KN, Liu KD, Thompson BT, Calfee CS, Network A. Acute respiratory distress syndrome subphenotypes respond differently to randomized fluid management strategy. *Am J Respir Crit Care Med*. 2017;195:331–8.
- Copelan EA. Hematopoietic stem-cell transplantation. *N Engl J Med*. 2006; 354:1813–26.
- Afessa B, Litzow MR, Tefferi A. Bronchiolitis obliterans and other late onset non-infectious pulmonary complications in hematopoietic stem cell transplantation. *Bone Marrow Transplant*. 2001;28:425–34.
- Chi AK, Soubani AO, White AC, Miller KB. An update on pulmonary complications of hematopoietic stem cell transplantation. *Chest*. 2013;144:1913–22.
- Yadav H, Nolan ME, Bohman JK, Cartin-Ceba R, Peters SG, Hogan WJ, Gajic O, Kor DJ. Epidemiology of acute respiratory distress syndrome following hematopoietic stem cell transplantation. *Crit Care Med*. 2016.
- Howrylak JA, Dolinay T, Lucht L, Wang Z, Christiani DC, Sethi JM, Xing EP, Donahoe MP, Choi AM. Discovery of the gene signature for acute lung injury in patients with sepsis. *Physiol Genomics*. 2009;37:133–9.
- Kangelaris KN, Prakash A, Liu KD, Aouizerat B, Woodruff PG, Erle DJ, Rogers A, Seeley EJ, Chu J, Liu T, et al. Increased expression of neutrophil-related genes in patients with early sepsis-induced ARDS. *Am J Physiol Lung Cell Mol Physiol*. 2015;308:L1102–13.
- Dolinay T, Kim YS, Howrylak J, Hunninghake GM, An CH, Fredenburgh L, Massaro AF, Rogers A, Gazourian L, Nakahira K, et al. Inflammation-regulated cytokines are critical mediators of acute lung injury. *Am J Respir Crit Care Med*. 2012;185:1225–34.
- Juss JK, House D, Amour A, Begg M, Herre J, Storisteanu DM, Hoenderdos K, Bradley G, Lennon M, Summers C, et al. Acute respiratory distress syndrome neutrophils have a distinct phenotype and are resistant to phosphoinositide 3-kinase inhibition. *Am J Respir Crit Care Med*. 2016;194:961–73.
- Dalli J, Colas RA, Quintana C, Barragan-Bradford D, Hunwitzer S, Levy BD, Choi AM, Serhan CN, Baron RM. Human Sepsis eicosanoid and Proresolving lipid mediator temporal profiles: correlations with survival and clinical outcomes. *Crit Care Med*. 2017;45:58–68.
- Langley RJ, Tipper JL, Bruse S, Baron RM, Tsalik EL, Huntley J, Rogers AJ, Jaramillo RJ, O'Donnell D, Mega WM, et al. Integrative "omic" analysis of experimental bacteremia identifies a metabolic signature that distinguishes human sepsis from systemic inflammatory response syndromes. *Am J Respir Crit Care Med*. 2014;190:445–55.
- Bernard GR, Artigas A, Brigham KL, Carlet J, Falke K, Hudson L, Lamy M, Legall JR, Morris A, Spragg R. The American-European consensus conference on ARDS. Definitions, mechanisms, relevant outcomes, and clinical trial coordination. *Am J Respir Crit Care Med*. 1994;149:818–24.
- Levy MM, Fink MP, Marshall JC, Abraham E, Angus D, Cook D, Cohen J, Opal SM, Vincent JL, Ramsay G. International Sepsis definitions C: 2001 SCCM/ESICM/ACCP/ATS/SIS international Sepsis definitions conference. *Intensive Care Med*. 2003;29:530–8.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
- Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5:621–8.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
- Gallego Romero I, Pai AA, Tung J, Gilad Y. RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol*. 2014;12:42.

25. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protocols*. 2008;4:44–57.
26. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 2008;3:1101–8.
27. Martin M: Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011 2011, 17.
28. Petersen KR, Streett DA, Gerritsen AT, Hunter SS, Settles ML. Super deduper, fast PCR duplicate detection in fastq files. In: In Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics. Atlanta, Georgia: ACM; 2015. p. 491–2.
29. Minot SS, Krumm N, Greenfield NB. One Codex: A Sensitive and Accurate Data Platform for Genomic Microbial Identification. *bioRxiv*. 2015. <https://www.biorxiv.org/content/early/2015/09/25/027607>.
30. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Sevransky JE, Sprung CL, Douglas IS, Jaeschke R, et al. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med*. 2013;41:580–637.
31. Dellinger RP, Levy MM, Carlet JM, Bion J, Parker MM, Jaeschke R, Reinhart K, Angus DC, Brun-Buisson C, Beale R, et al. Surviving Sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2008. *Crit Care Med*. 2008;36:296–327.
32. Seo S, Renaud C, Kuypers JM, Chiu CY, Huang ML, Samayoa E, Xie H, Yu G, Fisher CE, Gooley TA, et al. Idiopathic pneumonia syndrome after hematopoietic cell transplantation: evidence of occult infectious etiologies. *Blood*. 2015;125:3789–97.
33. Bhatt AS, Freeman SS, Herrera AF, Pedamallu CS, Gevers D, Duke F, Jung J, Michaud M, Walker BJ, Young S, et al. Sequence-based discovery of *Bradyrhizobium enterica* in cord colitis syndrome. *N Engl J Med*. 2013;369:517–28.
34. Morrell ED, Radella F 2nd, Manicone AM, Mikacenic C, Stapleton RD, Gharib SA, Wurfel MM. Peripheral and alveolar cell transcriptional programs are distinct in acute respiratory distress syndrome. *Am J Respir Crit Care Med*. 2018;197:528–32.
35. Faron Pharmaceuticals Ltd. Efficacy and Safety of FP-1201-lyo (Interferon Beta-1a) in Patients Having Acute Respiratory Distress Syndrome (ARDS) (INTEREST). NIH website. <https://clinicaltrials.gov/ct2/show/NCT02622724>.
36. Bellingan G, Brealey D, Mancebo J, Mercat A, Patroniti N, Pettita V, Quintel M, Vincent JL, Maksimov M, Jalkanen M, et al. Comparison of the efficacy and safety of FP-1201-lyo (intravenously administered recombinant human interferon beta-1a) and placebo in the treatment of patients with moderate or severe acute respiratory distress syndrome: study protocol for a randomized controlled trial. *Trials*. 2017;18:536.
37. Sweeney TE, Thomas NJ, Howrylak JA, Wong HR, Rogers AJ, Khatri P. Multicohort analysis of whole-blood gene expression data does not form a robust diagnostic for acute respiratory distress syndrome. *Crit Care Med*. 2018;46:244–51.
38. Zaba LC, Suarez-Farinas M, Fuentes-Duculan J, Nograles KE, Guttman-Yassky E, Cardinale I, Lowes MA, Krueger JG. Effective treatment of psoriasis with etanercept is linked to suppression of IL-17 signaling, not immediate response TNF genes. *J Allergy Clin Immunol*. 2009;124:1022–10 e1021–1395.
39. van Baarsen LG, Wijbrandts CA, Rustenburg F, Cantaert T, van der Pouw Kraan TC, Baeten DL, Dijkmans BA, Tak PP, Verweij CL. Regulation of IFN response gene activity during infliximab treatment in rheumatoid arthritis is associated with clinical response to treatment. *Arthritis Res Ther*. 2010;12:R11.
40. Reynier F, Petit F, Paye M, Turrel-Davin F, Imbert PE, Hot A, Mouglin B, Miossec P. Importance of correlation between gene expression levels: application to the type I interferon signature in rheumatoid arthritis. *PLoS One*. 2011;6:e24828.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

