Title: Single cell transcriptomics identifies immunologic priming related to extra corporeal life support survival.

Authors: Eric J. Kort MD MS[1,2], Matthew Weiland MS[1], Edgars Grins MD[3], Emily Eugster MS[1], Hsiao-yun Milliron PhD[1], Catherine Kelty MS[1], Nabin Manandhar Shrestha PhD[1], Tomasz Timek MD[3], Marzia Leacche MD[3], Stephen J Fitch MD[3], Theodore J Boeve MD[3], Greg Marco MD[3], Michael Dickinson MD[3], Penny Wilton MD[3], Stefan Jovinge MD PhD[1,3,4,*]

Affiliations:

[1]DeVos Cardiovascular Research Program, Spectrum Health & Van Andel Institute, Grand Rapids, Michigan, USA.

[2]Department of Pediatrics & Human Development, Michigan State University, Grand Rapids, Michigan, USA.

[3]Fredrik Meijer Heart and Vascular Institute, Spectrum Health, Grand Rapids, Michigan, USA.

[4]Cardiovascular Institute, Stanford University, Palo Alto, CA USA

[*]Correspondence to Stefan Jovinge: Stefan.jovinge@vai.org

Abstract:

As both sentinels and effectors of disease response, peripheral blood mononuclear cells are an accessible and attractive target for clinical application of high throughput, fluidics based single cell RNASeq (scRNASeq). However, new analytic tools required by unique characteristics of scRNASeq data lack validation in acutely ill patients. We report scRNASeq analysis of ~1,000 cells from each of 38 patients requiring veno-arterial extracorporeal life support (VA-ECLS)—a diverse group of critically ill patients experiencing circulatory collapse as a common endpoint to wide ranging diseases. We established an analysis pipeline capturing major biological signals from theses samples, confirmed by flow cytometry. Further, we found that these patients appeared immunologically poised at the outset of treatment as either reactive or permissive, and this balance predicted their survival. Annotated code detailing the analysis is available at https://github.com/vanandelinstitute/va_ecls.

One Sentence Summary: Patterns in single cell gene expression reveal aspects of cell biology associated with patient survival, even in complex clinical settings.

Main Text:

Microfluidics based single cell transcriptional profiling enables study of cell biology at higher resolution and throughput than previously possible (*1, 2*). This approach has led to numerous studies deepening our understanding of development (*3, 4*), as well as animal models of disease (*5*). Initial efforts to apply this technology clinically have shown promise in the contexts of infectious disease (*6*) and tumor biology (*7, 8*). Moreover, transcriptomics has been shown to be a powerful tool in repurposing drugs (*9, 10*) and single cell transcriptomics may reveal novel therapeutic targets in this clinical context. The extent to which high throughput, microfluidics based single cell RNASeq (scRNASeq) can be applied to complex clinical questions—including patient survival—remains largely undefined, as does the optimal approach to such analysis.

As both sentinels and effectors of disease response, peripheral blood mononuclear cells (PBMCs) are an accessible and attractive target for clinical application of scRNASeq. However, PBMCs represent a hetergenous and dynamic cell population. Further complicating such analysis is the fact that the data produced by microfluidics scRNASeq represents a "sparse matrix" with only a fraction of known genes exhibiting detectable expression. This represents a combination of genes that are not expressed—biological dropouts—as well as genes that are expressed but failed to be detected—technical dropouts. As a result of this and other unique features of the data produced by high throughput scRNASeq, specialized tools have been developed (*11-15*). But it remains to be determined to what extent an analytical pipeline can be identified that can extract clinically useful information from single cell RNASeq analysis of circulating PBMCs in a dynamic clinical setting.

In an effort to address this question, we performed fluidics based scRNASeq analysis of PBMCs from critically ill patients being initiated on venous-arterial extracorporeal life support

(VA-ECLS). Patients requiring VA-ECLS represent a heterogeneous population with a wide range of disease states including myocardial infarction, primary graft dysfunction, myocarditis, and acutely deteriorated heart failure (*16*). However, these patients exhibit a common pathophysiologic endpoint: circulatory collapse (*17*). For some patients, VA-ECLS support provides life support while failing organs recover and regain function, while in others it provides a life support bridge until definitive intervention—such as organ transplantation or mechanical assist devices—can be achieved. VA-ECLS continues to be associated with mortality rates of around 60% (*18*), due mainly to unrelenting progression of the underlying disease process, but also complication of VA-ECLS itself (*19*).

We performed scRNASeq based transcriptional profiling of a total of 40,000 peripheral blood mononuclear cells from 38 patients at the time of initiation of VA-ECLS (mean time between sample acquisition and VA-ECLS, ±79 minutes) on the inDrop microfluidics encapsulation platform (*1*). The study design and approach to analysis are summarized in Figure 1, and the clinical characteristics of these patients are summarized in Supplemental Table S1. Not surprisingly, non-surviving patients exhibited significantly more acidosis and poorer renal function at baseline than survivors, but no other clinical parameters were predictive. We were interested in identifying whether scRNASeq analysis could provide additional predictors of—and potentially mechanistic insights into—survival in these patients.

While unsupervised, genome-wide clustering of the expression data demonstrated a number of cell clusters comprised solely of cells from deceased patients, it was apparent that this was an artefact of batch effect as these clusters were each comprised of cells from a single patient (Fig 2A-C). It is well established that microfluidics based scRNASeq approaches are subject to technical dropouts—failure to detect transcripts that are present due to technical

factors including low efficiency of RNA capture by the oligo-dT primers in the droplets. Imputation of technical dropouts via the ALRA algorithm (*14*) ameliorated the batch effects to some extent (Fig 2D-F). Explicit regression of the batch effects via the Harmony algorithm (*15*) homogenized the data with respect to donor/batch, while improving clustering by PBMC subtype (Fig 2G-I) as determined by expression of genes encoding canonical surface markers (supplemental Table S2).

Unsupervised clustering based on genome wide gene expression after removal of batch effects and imputation of technical dropouts demonstrated that the cells tended to cluster by major PBMC subtypes (Fig 3A), based on expression of established markers of these cell populations (Fig 3B). This observation suggests that scRNASeq can capture the major themes of cellular biology even in this clinically complex and dynamic setting—provided appropriate steps are taken to account for technical dropouts and batch effects.

Previous work has indicated that the expression of genes encoding surface markers is highly correlated to protein levels of those surface markers (*20*). To verify that this was the case in this clinical context of patients under extreme physiological stress, we also analyzed these samples by conventional flow cytometry (FC). We compared the proportions of cells in the major lymphocyte subtypes as defined by the gene expression data vs. direct measurement of surface markers by FC. Cell assignments between the two methodologies correlated well (Fig 3C). Prior to imputation of technical dropouts, only 54% of cells could be unambiguously assigned to a specific PBMC subtype, compared to 81% after imputation. Notably, imputation had no significant effect on either the slope of the regression line or the R^2 between the proportion of lymphocytes assigned to each subclass by either FACS or scRNASeq (in fact, both values increased slightly after imputation—supplemental Table S3). The fact that the correlation

between lymphocyte subset proportions as defined by FACS and scRNASeq did not deteriorate following imputation suggests that the substantial increase in cell assignment achieved by imputation identified the true biological type of these cells as opposed to random noise.

We next sought to determine whether conventional PBMC classification could predict outcome. Using our gene expression-based cell type assignments, we looked for differences in proportions of these cell types in surviving and non-surviving patients. Although some trends were apparent, none of the observed differences were statistically significant after adjustment for multiple comparisons (Fig. 4A). This suggested we needed to look deeper into each major cell type to try to identify novel sub-populations that may relate to survival.

We identified the set of highly variable genes within our dataset based on normalized dispersion (*2*). For each patient, we then quantified the proportion of cells of each subtype that expressed each of these genes. For each gene, the proportions among surviving patients were compared to the proportions among non-surviving patients using the Wilcoxon rank sum test. This allowed us to identify genes within each PBMC subtype that were associated with survival (blue bars in Fig 4B) or non-survival (red bars in Fig 4B). We clustered the genes based on this measure of differential expression, and annotated the resulting gene clusters based on their enrichment (*21*) for Gene Ontology biological process terms (Fig 4B, left panel, and Supplemental Table 4). Within multiple PBMC subtypes, patients who died exhibited higher proportions of cells expressing genes associated with GO terms related to inflammation including antigen binding, cytokine activity, and heat shock protein binding. This suggests a wide spread increase in inflammatory response across cell types in these non-surviving patients.

However, other processes appeared to have more discreet patterns of activation within specific cell types. Genes associated with S100 protein and actin binding—perhaps related to the

cytoskeletal rearrangement essential to antigen presentation and effector cell activation (*22*)—exhibited an inverse relationship between $CD8^+$ naïve T cells compared to CD8+ natural killer T (NKT) cells. Non-surviving patients exhibited higher proportions of $CD8^+$ naïve T cells expressing genes related to these biological processes at the time of initiation of VA-ECLS, while the inverse was true of CD8+ NKT cells. Although the CD8+ naïve T cells constituted a small cell population, these cells can proliferate rapidly when they encounter their target antigens. The extensive synthetic surfaces of the VA-ECLS circuit can induce significant inflammatory responses (*19*). The presence of CD8+ naïve T cells with activating gene signatures among non-surviving patients—as identified here—may be indicative of an immune system poised to have a maladaptive inflammatory response to VA-ECLS.

Future measurement and/or isolation of novel subpopulations of survival-related cells would be facilitated by identifying surface markers that identify these cells. Therefore, we next focused our differential gene analysis on surface markers (Fig 4B, left panel). After control for multiple comparisons, two surface markers stood out. Surviving patients had significantly higher proportions of $CD8^+$ NKT cells ($CD3^+/CD8^+/CD19^-/CD56^+$) that were $CD52^+$ or $CD3G^+$ ($p = 0.001$, FDR < 0.05 for both markers). $CD52^+$ cells have been implicated in autoimmune disease, and this surface marker is the target of a therapeutic agent used to treat both some leukemias and multiple sclerosis (*23, 24*). However, there is evidence suggesting that a there is a subset of $CD8^+$ cells that are protective against autoimmune disease (*25*), and that these protective cells may be $CD52^+$ (*26*). Similarly, CD3G has been demonstrated to play an essential role in restraining auto-immune reactivity (*27*). Thus, the $CD52^+$ NKT cells we identified here may be indicative of an immune system poised for a more permissive response to the inflammatory stimulation of VA-ECLS.

We stratified the patients based on their CD52 +/- proportions and CD3G+/- proportions (among all $CD8^+$ NKT cells), using the median proportion as the cutoff in each case. This stratification identified patients with significantly different mortality risks on VA-ECLS during the first 72 hours of support (Fig 4C,D). We hypothesized that immune response at VA-ECLS initiation would be predictive of early survival, and this was confirmed by this analysis. After this initial period, however, we expected that other factors related to the underlying disease process would be more important than the initial inflammatory response. It was surprising, therefore, that CD3G expression levels were predictive not only of 72-hour survival, but also of 30-day survival (Fig. 4E).

These results indicate that, with the aid of available data processing tools, high throughput microfluidics based scRNASeq can capture significant aspects of cell biology even in complex clinical settings. Our analysis further suggests that patients on VA-ECMO may be immunologically primed at the outset to follow one of two inflammatory response paths. While it remains to be determined whether this priming is cause or effect, the patients studied here exhibited what may represent either a reactive (CD8+ naïve T exhibiting activating gene signatures) or permissive ($CD8^+$ NKT cells that are $CD52^+$ or $CD3G^+$) predominant phenotype, and that balance was predictive of outcome.

References and Notes:

1. A. M. Klein *et al.*, Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201 (2015).
2. E. Z. Macosko *et al.*, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161**, 1202-1214 (2015).

3. A. Sebe-Pedros *et al.*, Cnidarian Cell Type Diversity and Regulation Revealed by Whole-Organism Single-Cell RNA-Seq. *Cell* **173**, 1520-1534 e1520 (2018).

4. D. E. Wagner *et al.*, Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981-987 (2018).

5. D. Aran *et al.*, Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019).

6. Z. Wang *et al.*, Clonally diverse CD38(+)HLA-DR(+)CD8(+) T cells persist during fatal H7N9 disease. *Nat Commun* **9**, 824 (2018).

7. R. Sen *et al.*, Single-Cell RNA Sequencing of Glioblastoma Cells. *Methods Mol Biol* **1741**, 151-170 (2018).

8. M. D. Young *et al.*, Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *Science* **361**, 594-599 (2018).

9. K. Stegmaier *et al.*, Gene expression-based high-throughput screening(GE-HTS) and application to leukemia differentiation. *Nature genetics* **36**, 257-263 (2004).

10. K. Stegmaier *et al.*, Gefitinib induces myeloid differentiation of acute myeloid leukemia. *Blood* **106**, 2841-2848 (2005).

11. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).

12. C. Trapnell *et al.*, The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* **32**, 381-386 (2014).

13. P. V. Kharchenko, L. Silberstein, D. T. Scadden, Bayesian approach to single-cell differential expression analysis. *Nat Methods* **11**, 740-742 (2014).

14. G. C. Linderman, J. Zhao, Y. Kluger, Zero-preserving imputation of scRNA-seq data using low-rank approximation. *bioRxiv*, https://doi.org/10.1101/397588 (2018).

15. I. Korsunsky *et al.*, Fast, sensitive, and accurate integration of single cell data with Harmony. *bioRxiv*, https://doi.org/10.1101/461954 (2018).

16. D. M. Ouweneel *et al.*, Extracorporeal life support during cardiac arrest and cardiogenic shock: a systematic review and meta-analysis. *Intensive Care Med* **42**, 1922-1934 (2016).

17. J. L. Vincent, D. De Backer, Circulatory shock. *N Engl J Med* **369**, 1726-1734 (2013).

18. R. R. Thiagarajan *et al.*, Extracorporeal Life Support Organization Registry International Report 2016. *ASAIO J* **63**, 60-67 (2017).

19. J. E. Millar, J. P. Fanning, C. I. McDonald, D. F. McAuley, J. F. Fraser, The inflammatory response to extracorporeal membrane oxygenation (ECMO): a review of the pathophysiology. *Crit Care* **20**, 387 (2016).

20. M. Stoeckius *et al.*, Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868 (2017).

21. G. Yu, L. G. Wang, Y. Han, Q. Y. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).

22. Y. Yu, A. A. Smoligovets, J. T. Groves, Modulation of T cell signaling by the actin cytoskeleton. *J Cell Sci* **126**, 1049-1058 (2013).

23. B. Dumitriu *et al.*, Alemtuzumab in T-cell large granular lymphocytic leukaemia: interim results from a single-arm, open-label, phase 2 study. *Lancet Haematol* **3**, e22-29 (2016).

24. J. Dorr, K. Baum, Alemtuzumab in the treatment of multiple sclerosis: patient selection and special considerations. *Drug Des Devel Ther* **10**, 3379-3386 (2016).

25. S. Sinha, A. W. Boyden, F. R. Itani, M. P. Crawford, N. J. Karandikar, CD8(+) T-Cells as Immune Regulators of Multiple Sclerosis. *Front Immunol* **6**, 619 (2015).

26. S. von Kutzleben, G. Pryce, G. Giovannoni, D. Baker, Depletion of CD52-positive cells inhibits the development of central nervous system autoimmune disease, but deletes an immune-tolerance promoting CD8 T-cell population. Implications for secondary autoimmunity of alemtuzumab in multiple sclerosis. *Immunology* **150**, 444-455 (2017).

27. B. Gokturk *et al.*, CD3G gene defects in familial autoimmune thyroiditis. *Scand J Immunol* **80**, 354-361 (2014).

Acknowledgments: We would like to acknowledge Jennifer Schuitema for overseeing recruitment, consent, and collection of blood samples, the ECMO nursing staff for collecting blood samples and processing plasma after-hours, and David Chesla and Donald Daley from the Spectrum Health Biorepository for sample processing during after-hours. The authors thank the Van Andel Genomics Core for providing sequencing facilities and services. **Funding:** This work was made possible by the generosity of the Helen and Richard DeVos foundation. **Author contributions:** EJK and SJ conceptualized and conceived the project. EJK also performed the analysis, developed methodology, and wrote the manuscript. MW performed the flow cytometry work described in this work, analyzed data, and reviewed and edited the manuscript. HM, and EE performed the single cell encapsulation, and reviewed and edited the manuscript. CK assisted with patient enrollment and data collection. EG, ML, SF, TB, GM, TT, MD, and PW provided clinical resources and supervision for the project and reviewed the manuscript. NMS performed analysis of clinical data. SJ also acquired funding, developed methodology, supervised the project, and wrote the manuscript. **Competing interests:** The authors declare no competing

interests. **Data and materials availability:** The expression data is available as a matrix of raw counts from GEO (GSE127221). Code used to perform the analysis and generate the figures is available from Github at http://github.com/vanandelinstitute/va_ecls.

Supplementary Materials:

Materials and Methods

Tables S1-S5

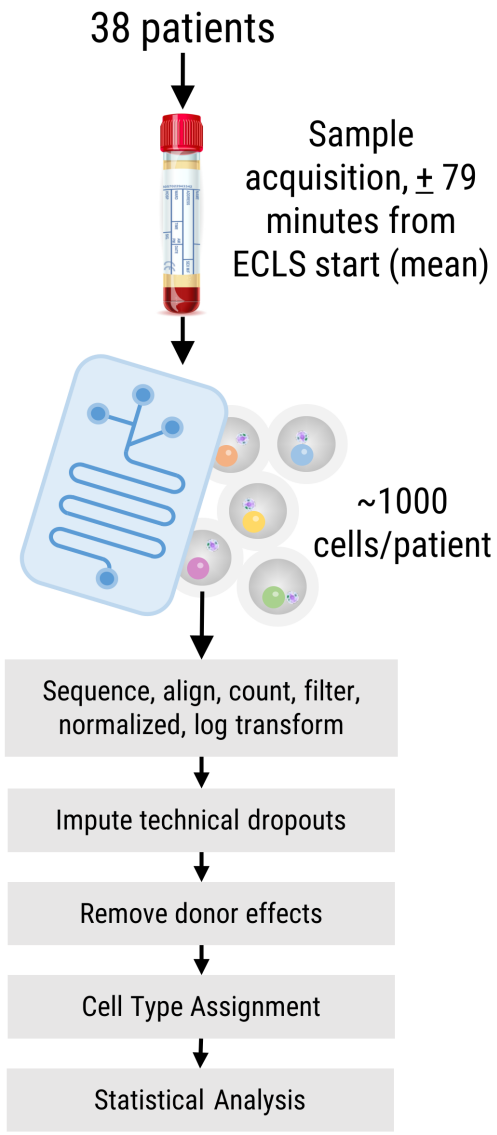

Fig. 1. Overview of study design, including the work flow used for data processing.

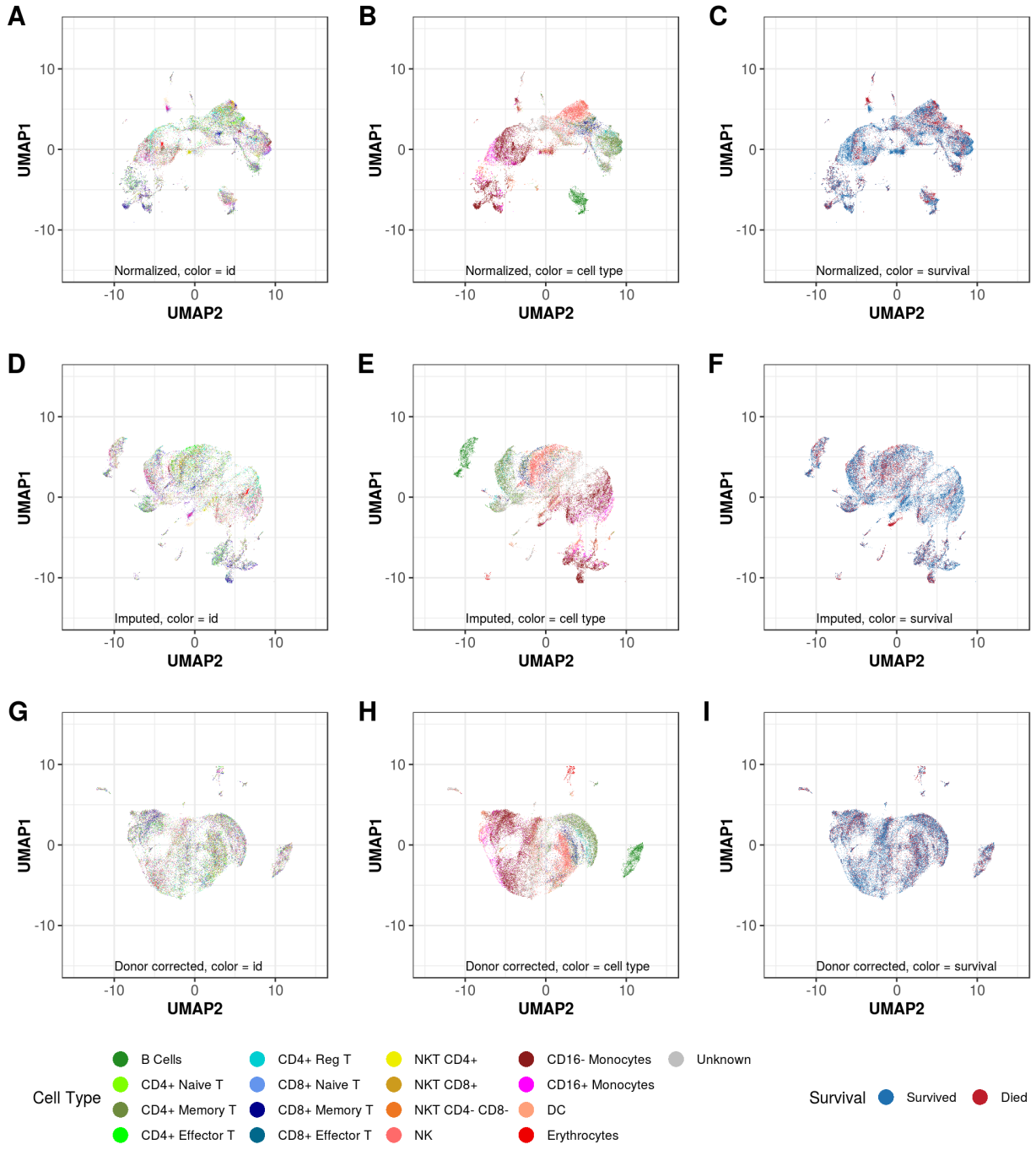

Fig. 2. Unsupervised clustering of the profiled cells in two dimensions by Uniform Manifold Approximation and Projection (UMAP). Cells are color coded by patient study identifier (A,D,G), cell type as assigned by expression of surface marker genes (B,E,H) or 72 hour mortality status (C,F,I). Systematic removal of artefactual clustering is demonstrated by imputation of technical dropouts by the ALRA algorithm (*14*) (D-F), followed by batch (donor) effect removal by Harmony (*15*) (G-I).

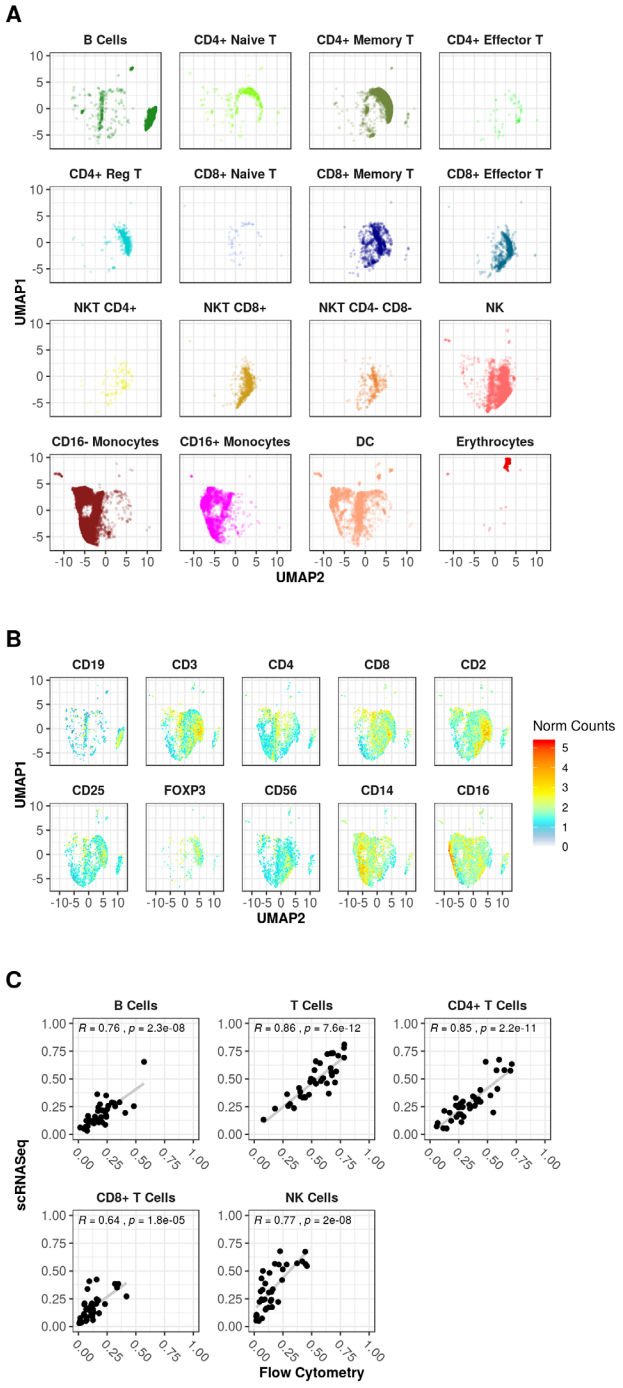

Fig. 3. After data processing, cells clustered predominantly by cell type as defined by expression of surface marker genes (A). In addition to major PBMC populations, a small number of contaminating erythrocytes can also be seen. The expression of surface markers themselves show distinct patterns of expression (B) suggesting that the cells are clustering by biological class rather than other technical or clinical factors. Proportions of cells in each major lymphocyte subgroup as defined by scRNASeq correlate well with results of flow cytometric analysis of these same samples (C).

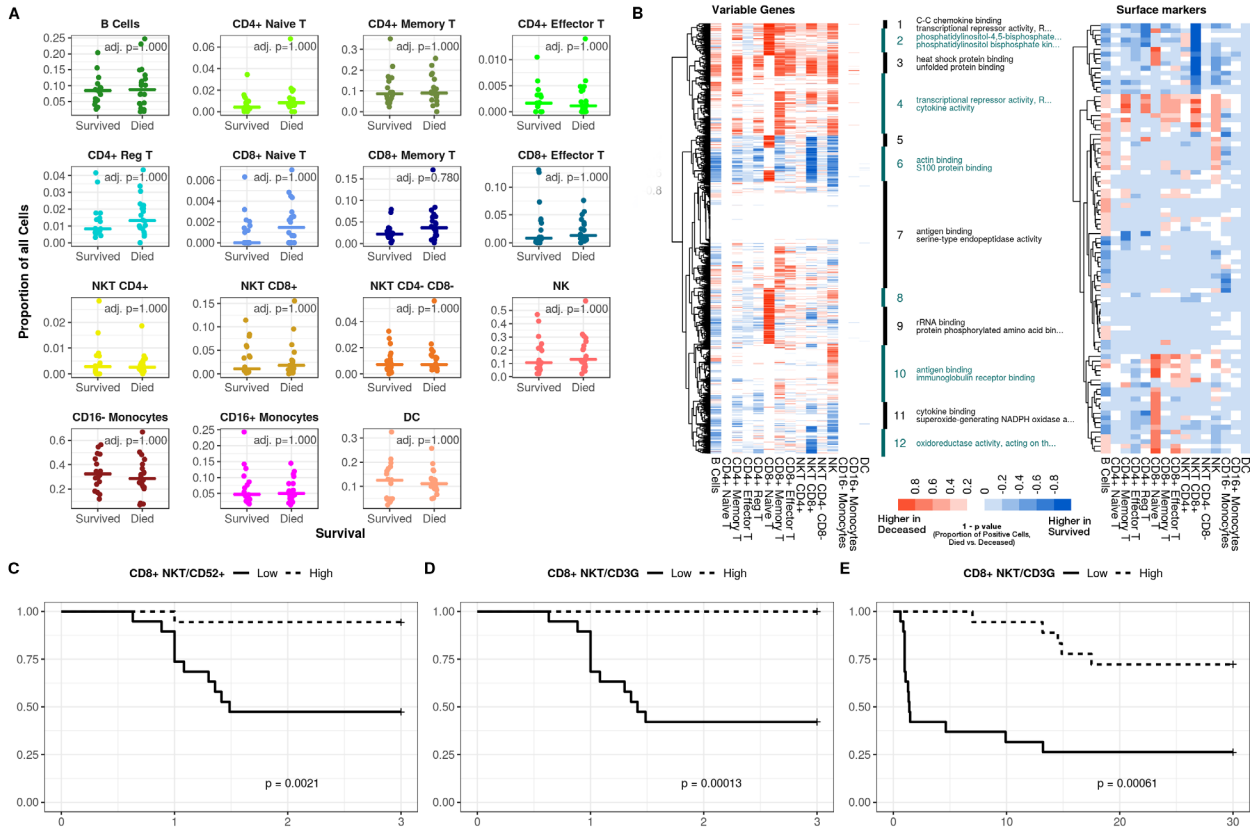

Fig. 4. (A) Proportions of all cells in each patient that were classified as each of the major PBMC subtypes, stratified by patent outcome at 72 hours. After adjustment for multiple comparisons, none of these conventional subtypes could significantly predict survival. **(B)** Expression patterns of highly variable genes (defined by normalized dispersion (*2*)) and cell surface markers. For each patient, the proportion of cells in each PBMC subtype that expressed each gene was calculated. The median proportion was then compared between survived vs. deceased patients by Wilcoxon rank sum test. Red bars indicate the genes that were expressed more frequently in cells of a given subtype from deceased patients. Blue bars indicate the genes that were expressed more frequently in cells for a given subtype in patients who survived. The intensity of the color corresponds to 1-FDR, such that the darkest colors indicate the most significant genes. Those clusters containing genes enriched for specific Gene Ontology Biological Process terms are annotated. Up to two significant terms are listed for each cluster. A complete list of enriched terms is provided in Supplemental Table S4. **(C-E)** 72 hour (C,D) or 30 day (E) survival based expression of surface markers. Patients were stratified using median expression level as the cutoff based on proportion of CD8+ NKT cells that were positive for CD52 (C) or CD3G (D,E).