Supplemental Appendix

Table of Contents

3.6 Supplemental Table 6. List of differentially expressed genes by TCS versus TCE (6A) or carHLH versus no carHLH (6B).

4.0 Supplemental Figures

4.1 Supplemental Figure 1. Clinical characteristics of carHLH in patients receiving CD22 CAR T-cells, stratified by CRS grade (1-2 vs 3-4)
4.2 Supplemental Figure 2. Anemia, thrombocytopenia and transfusion requirements
4.3 Supplemental Figure 3. Clinical characteristics of carHLH in patients receiving CD19 CAR T-cells.
4.4 Supplemental Figure 4. Serial cytokine profiling in patients receiving CD22 CAR T-cells relative to CRS onset (top) and relative to day of CAR T-cell infusion (bottom). Concurrent data/statistical analysis shown in Supplemental Tables 2 and 3.
4.5 Supplemental Figure 5 Cytokine profiling of IL-6 with and without tocilizumab in CD22 CAR patients. Soluble IL2R and CXCL9 levels
4.6 Supplemental Figure 6. Serial lymphocyte phenotyping following product infusion in patients with vs without carHLH.
4.7 Supplemental Figure 7. NK-cell subset analysis.
4.8 Supplemental Figure 8. Impact of TCS on carHLH and transcriptional profiling of CAR T-cell infusion products.

1.0 Supplemental Methods

We retrospectively analyzed patient and CAR T-cell product characteristics based on subjects who enrolled on a phase I trial testing CD22 CAR T-cells in patients with relapsed/refractory CD22+ leukemia or lymphoma. The results from the first 58 subjects treated were recently reported.[1] The primary objective was to identify factors associated with HLH-like toxicities (carHLH), as distinguished from those without them, amongst all those with CRS.

1.1 Subjects

Only those patients who received a CAR product were analyzed for clinical outcomes (n=60); with exclusion of 1 subject who received chemotherapy on day +4 following infusion due to rapid disease progression, whose outcome measures regarding toxicity could not be accurately assessed since CAR T-cell response was abrogated prior to expansion. All subjects provided written informed consent, or parental permission with minor assent obtained when appropriate. All patients were treated in the Pediatric Oncology Branch of the National Cancer Institute (NCI), and the protocol was approved by the NCI Institutional Review Board and the NIH Recombinant DNA Advisory Committee. This trial is registered at ClinicalTrials.gov (NCT02315612). Methods regarding CD22 CAR T-cell manufacturing have been previously reported.[2]

CD22 CAR T-Cell Products

All products were manufactured at the NIH Clinical Center, Center for Cellular Engineering (CCE). Initial manufacturing strategies incorporated elutriation (n=6) and flask adhesion methods with utilization of anti-CD3/anti-CD28 magnetic beads (n=19) as an enrichment step for T-cell purification (TCE). Following the initial experience, as previously described, CD4/CD8 Miltenyi® beads were incorporated for T-cell selection (TCS) of all starting apheresis products without any other downstream changes in manufacturing (Supplemental Figure 6C).[2] This change was introduced to enhance feasibility of CAR T-cell manufacturing, as well as improve the consistency of and reduce inherent interpatient variability between starting apheresis product materials. Subsequent subjects (n=35), with rare exception, were treated with a product manufactured utilizing TCS. Following TCS or TCE, the intermediate CAR product was either cryopreserved or put directly into culture for further manufacture. CAR products were assessed by flow cytometry immediately following apheresis, post-selection (or enrichment), after culture for 7 days, and on the final day of product culture. Additional information regarding manufacturing process is described in the Supplemental Appendix.

1.2 Toxicity and Efficacy Evaluations

CRS was graded by Lee *et al*,[3] which may not be congruent with how all studies grade CRS. For this reason, we reconciled our data with the current ASTCT CRS grading criteria,[4] and found that the incidence of carHLH was equally distributed amongst all patients who were up or downgraded in CRS grade, and would not impact our findings. CRS was defined by the period of time starting from the first onset of fevers through a 24-hour afebrile period and at least a 50% decrease in C-reactive protein (CRP) from its peak.

CarHLH criteria was derived from HLH-2004 criteria, but required further modification due to patient population-specific considerations. Fever was excluded to avoid confounding with CRS-associated fevers. Splenomegaly was excluded as it is often observed in leukemic patients at baseline (and generally improves with CAR treatment). Cytopenias were excluded as tri-lineage decrease in blood cell count is a known common complication of CAR treatment. The cutoff for

ferritin was re-assigned as >100,000 ug/L rather than >500 ug/L as all patients in the cohort would satisfy the latter criteria. Triglycerides, fibrinogen and sCD25 were initially not routinely collected, and the former two were also frequently impacted by prior chemotherapy. The IL-6R antagonist tocilizumab was administered as clinically indicated for CRS.

Disease evaluations included peripheral blood flow cytometry, CSF sampling, and bone marrow evaluation with aspirate, biopsy and flow cytometry. Disease burden was assessed at pre-CAR infusion baseline, and restaging was performed at day 28 +/- 4 days post CD22 CAR T-cell infusion with rare exception (n=1 had a delayed restaging due to being unable to undergo a sedated procedure).

Pre-CAR infusion baseline bone marrow was obtained for all patients to assess disease burden (n=59). However, characterization of T-, B-, and NK-cells (TBNKs) were not performed for n=4 patients for the following reasons: a limited flow panel was performed for n=3 (n=2 carHLH-, n=1 carHLH+) due to low aspirate cellularity; analysis was not performed for n=1 (CRS-) without comment (subject #1). “Pre-CAR Baseline” bone marrow is defined as bone marrow aspirate flow taken closest to CAR infusion (day 0) but prior to lymphodepleting chemotherapy, with a median of 15 days pre-infusion (range: 5 - 42).

Restaging bone marrow aspirate flow was performed at day 28 +/- 4 days post CD22 CAR T-cell infusion for n=48 subjects to assess disease burden. Deviations were made for the following reasons (n=11 deviations total): subject #24 (carHLH-) died at day +9 following infusion; subject #63 (carHLH+) suffered a large intracranial hemorrhage at day +17 and only underwent minimally invasive procedures; n=3 subjects (#29, #59, #62) without clinical signs of CRS underwent early restaging at days +20, 22, and 20 respectively; subject #42 (carHLH+) underwent early restaging at day +20 due to clinical concern about disease status; n=4 (#31 (carHLH+), #39 (carHLH-), #51 (carHLH+), and #56 (carHLH+)) had restaging at day 28 +/- 4 days but required repeat restaging due to hypocellular marrow which were performed at day +48, +56, +40, and +41 respectively; and subject #27 (carHLH+) was deemed medically unsafe for sedation at day 28 and had the first restaging bone marrow at day +52.

Notably, a limited flow panel was performed for subjects #16 (carHLH+), #45 (carHLH-), and #53 (carHLH+) at day 28+/-4 days because of hypocellularity, which was sufficient for assessment of leukemic burden but precluded TBNK analysis; these were not repeated. Additionally, subject #42 was hypocellular on restaging at day +20 which allowed for assessment

of leukemic burden but precluded TBNK analysis; restaging bone marrow aspirate flow was not repeated.

1.3 Bone Marrow Analysis

Specimens were processed within 12 hours of collection. Briefly, red blood cell lysis was performed with lysing solution (150 mmol/L NH4Cl, 10 mmol/L KHCO3, and 0.1 mmol/L EDTA) before staining for 30 min at room temperature with a panel of antibodies: CD3 PerCP (clone SK7, BD Biosciences), CD13 PE-Cy7 (L138, BD Biosciences), CD14 APC-H7 (MphiP9, BD Biosciences), CD16 FITC (clone DJ130c, Dako), CD19 PE (clone SJ25C1, BD Biosciences), CD34 APC (clone 8G12, BD Biosciences), CD45 v500 (clone HI30, BD Biosciences) and CD56 v450 (clone B159, BD Biosciences). Specimens were acquired on a 3-laser FACS Canto II (BD Biosciences, San Jose, CA) with DiVa 6.1.1 software. Data acquired in list mode were analyzed with FCS Express v.6 (De Novo Software, Glendale, CA). The T/B/NK-cell populations were gated by forward and side scatter light properties consistent with lymphoid cells, in conjunction with antigen gating to select T cells (CD3+, CD14-), B cells (CD19+, CD3-) and NK-cells (CD16+ and/or CD56+, CD45 bright, CD3-, CD13-, CD14-, CD19-, CD34-).

1.4 Correlative Studies

1.4.1 CAR T-cell expansion:

Analysis of CAR T-cell expansion and persistence by flow cytometry were performed weekly for the first month post-infusion and then at 2, 3, 6 and 12 months. Serial lymphocyte phenotyping was ascertained by flow cytometric analysis of peripheral blood (PB) and bone marrow (BM) samples at pre-determined timepoints to evaluate CAR T-cell expansion and T-, B, and NK-cell counts (TBNK).

1.4.2 Serial cytokine profiling:

Serial plasma cytokine levels, performed using ELISA based Mesoscale V-Plex platform, included IL-1β, IL-2, IL-4, IL-6, IL-8, IL-10, IL-12p70, IL-13, IL-15, TNFα, MIP-1α, IFNγ and GM-CSF and were resulted on 59 subjects. Ferritin and C-reactive protein (CRP) were captured serially through day 28 for all subjects, with additional sampling as clinically indicated. IL-18 and IL-18bp were captured separately from other cytokines, and obtained only for patients with CRS (n=52).

Each test plate (cytokines and IL-18bp) included addition of 3 (QC) controls, either purchased from BioTechne or were prepared from standard proteins included in the kits as spikes in assay diluent. The measured QC values were within pre-established concentration ranges ("expected") on all plates having the patient samples. Using anonymized plasma samples, the total IL-18 ELISA kit (MBL International) and IL18bp values (BioTechne Quantikine ELISA kit (cat. no. DBP180)) were performed on a single (respective) manufacturing lot. The dilution factor for the latter used was 1:50. Two additional markers, MIG-1(CXCL9) and IP-10(CXCL10) were tested in a 2-plex using the MSD U-plex platform, the dilution factor used was 1:40.

The Immunology Flow Cytometry Laboratory used a fully integrated system with BD FACS™ Sample Prep Assistant (SPA) for automated lyse/no-wash sample preparation for routine T/B/NK-cell studies in a single tube using BD Multitest™ 6-color TBNK reagent. Samples were then loaded on the BD FACS™ Loader for automated sample acquisition on a BD FACSCanto™ instrument.

The BD Multitest 6-color TBNK reagent contained the following antibodies to identify and enumerate the different (T, B, and NK) lymphocyte subsets:

- CD3 FITC for the identification of T lymphocytes
- CD16 and CD56 PE for identifying NK lymphocytes
- CD45 PerCP-Cy™5.5 to allow for gating on the lymphocyte populations
- CD4 PE-Cy™7 for detecting T helper/inducer lymphocytes - CD19 APC to identify B lymphocytes
- CD8 APC-Cy7 for the identification of the suppressor/cytotoxic T lymphocyte subset

The 13-Plex for cytokine assessment was comprised of the 10-plex Proinflammatory Panel 1, the 2-plex GMCSF & IL-15, and a single-plex MIP-1a. These kits (MesoScale Discovery) were of the V-Plex format, which is validated by the manufacturer and comes with reagents that have been bridged lot-to-lot to yield nearly the same standard curves; necessary for consistent combining of results from samples tested over time. The IL-18 ELISA kit was from MBL International, and all tests were performed on a single manufacturing lot. Standard proteins included in the kits were spiked in assay diluent as 3 QC (high, intermediate and low) levels. All controls were evaluated for accuracy with recoveries (%Expected) that performed in the range of ±25% variation.

Soluble IL2R/CD25 was a send out lab sent to the Mayo Clinic. Methods are as follow: Analyte-specific antibodies are pre-coated onto color-coded magnetic microparticles. Samples are diluted 1:2 in a mixing plate and then standards, samples, and microparticles are pipetted into wells and the immobilized antibodies capture the analytes of interest. Unbound substances are washed away while the magnetic microparticles are immobilized. Next, a biotinylated analyte specific antibody cocktail is added to each well. Following a wash to remove any unbound biotinylated antibody, streptavidin-phycoerythrin conjugate (Streptavidin-PE), is added to each well. After removal of unbound Streptavidin-PE and resuspension of the microparticles in buffer, the plate is analyzed using a Luminex FLEXMAP 3D analyzer. A charged-coupled device (CCD) camera captures an image of each well and data reduction is performed using the XPONENT software. (Package insert: Human Luminex Assay. R and D Systems, Inc; 05/2020)

1.4.3 NK-cell studies:

NK-cell subset analysis:

Whole blood specimens for flow cytometry were processed within 12 hours of collection. Briefly, red blood cell lysis was performed using ammonium chloride before staining for 30 min at room temperature with the following cocktail (antibody concentration according to manufacturer's recommendations): CD16 FITC (clone DJ130c, Dako), CD19PE (clone SJ25C1, BD Biosciences), CD3 PerCP (clone SK7, BD Biosciences), CD13 PE-Cy7 (L138, BD Biosciences), CD34 APC (clone 8G12, BD Biosciences), CD14 APC-H7 (MP9, BD Biosciences), CD56 V450 (clone B159, BD Biosciences), CD45 V500 (clone HI30, BD Biosciences). One million cells were acquired per tube on a 3-laser FACS Canto II cytometer (BD Biosciences, San Jose, CA) and were analyzed with FCS Express 6 software (DeNovo Software, Los Angeles, CA).

NK-cells were identified by gating on the CD45-positive, CD13-negative lymphoid cells, excluding B lymphocytes, T lymphocytes, monocytes, and immature blasts by the expression of CD19, CD3, CD14 and CD34, respectively. The NK-cells were further subclassified based on the differential expression of CD56 and CD16 into 6 subgroups (NK1 to NK6), representing CD56brightCD16−, CD56brightCD16dim, CD56dimCD16−, CD56dimCD16dim, CD56dimCD16bright, and CD56−CD16bright subsets, as previously described in the literature.[5,6] Among these subsets, NK1 and NK2 display bright CD56 expression, while subgroups NK3 to

NK5 are characterized by dim CD56 expression. NK6 represents only a minor fraction of NK-cells in the majority of cases in this study and was not included in subsequent data analysis.

NK functional studies:

Prior to *in vitro* stimulation, thawed PBMC cells were rested overnight in 10% FBS-containing cell culture medium. For determination of challenged NK-cells' CD107a expression along with IFN-g, CD107a- PerCP/Cy5.5 (BD Biosciences) was added to cells at the start of the *in vitro* stimulated condition of 10,000 NK effectors : 10,000 K562 Targets (1:1) (or without). After 1 hr, monensin (Golgi Stop, BD Biosciences) was added to cells for the next 5 hrs. Cells were harvested at 6 hr.

PMA/ ionomycin was included as a separate test of each patient's sample as a positive assay control. This was used to aid in interpreting the extent of intracellular staining possible, as these agents trigger a vigorous stimulation of NK-cell activation. See the Flow Cytometry section below for gates utilized (CD3–, CD56+, CD14–, CD20–) in order to focus only upon the NK-cells and then assess CD107a surface expression and IFN-g production.

Flow cytometry:

At appropriate time points pre- and post- CAR-T infusion, CD3-negative NK-cells within the PBMC population (also CD20- and CD14-negative) were examined for activation surface markers (CD69, CD54, HLA-DR) and intracellular Perforin, GrzB, CD107a and IFN-g expression by flow cytometry.

For phenotyping, thawed cells were rested 2 hours. Cells were washed with PBS and incubated in Live/dead dye (Thermo Fisher) for 30 minutes (RT); cells were then washed twice in PBS/1% BSA before monoclonal antibody (mab) labeling with CD56(ICAM-1)-PE (BD Biosciences), CD69- BV421(Biolegend), CD3-PE/Cy7 (BioLegend), CD20-PE/Cy7(BioLegend), CD14-PE/Cy7 (BioLegend), and HLA-DR-BV510 (BioLegend) for 30 mins at 4°C. Cells were washed twice in PBS/1% BSA before Cytofix/Cytoperm (BD) was added for 20 min at 4°C. Prior to subsequent intracellular staining (ICS), cells were washed twice in Perm/Wash Buffer (BD Biosciences) and then mab to Perforin-FITC or GrzB-Alexa Fluor 647 (BD Biosciences) or IFN-g - BV421 mab (BioLegend) were added to appropriate tubes for 30 mins at 4 °C. Cells were

washed with Perm/Wash solution, fixed in 2% Formaldehyde and analyzed with flow cytometer (BD FACS Canto II).

Data acquired with BD Canto-II flow cytometer were analyzed with FCS Express (De Novo Software). Cell staining parameters were verified at the outset of the experiments, with use of appropriate FMO controls to assess nonspecific binding. Residual RBC contamination in some patient vials was gated so as to be removed from the counts.

1.4.4 RNA sequencing analysis

RNA sequencing read quality was assessed per sample using FastQC (version 0.11.5)[7], Preseq (version 2.0.3)[8], Picard tools (version 1.119)[9] and RSeQC (version 2.6.4)[10]. Sequencing reads were trimmed of low-quality bases and adapter sequences were removed using Cutadapt (version 1.18)[11]. The trimmed reads were aligned the GRCh38 (GENCODE hg38, version 30). Gene and isoform expression levels are quantified using RSEM (version 1.3.0). In summary, 33 individual patient samples remained after removing samples with poor library quality (low overall % of mapped reads or low library complexity).

Genes with a log2(CPM) > 2.5 in at least 5 samples were selected for analysis. Only protein coding genes were used for further downstream analysis. DEseq2 (version 1.28.1)[12] was used to normalize read counts, correct for batch effect between sequencing runs, and identify differentially expressed genes. Genes with an FDR qvalue < 0.01 were considered significantly expressed.

1.5 Timepoints

Lymphocyte phenotyping of peripheral blood was performed by flow cytometry to assess TBNKs at the following timepoints: “pre-CAR baseline,” “day 0,” “day 7,” “day 14,” and “day 28.” “Pre-CAR Baseline peripheral blood” TBNK is defined as peripheral blood flow taken closest to “Pre-CAR Baseline bone marrow” date, within a +/- 14-days so long as it remains prior to lymphodepleting chemotherapy. Median “Pre-CAR Baseline peripheral blood” reading was 1 day prior to “Pre-CAR Baseline bone marrow” (range 14 days prior – 11 days after). All other timepoints used closest available data to specified timepoint relative to actual day of CAR infusion, within the following windows: “day 0 timepoint”, 1 day before to 2 days after to CAR infusion;

"day 7 timepoint", 5 to 9 days post CAR infusion; "day 14 timepoint", 10 to 18 days post CAR infusion; "day 28 timepoint", 24 to 32 days post CAR infusion.

CAR T-cell expansion was assessed at "day 0," "day 7," "day 14," and "day 28" timepoints as described above for lymphocyte phenotyping of peripheral blood.

Soluble IL-2Rα was assessed on a limited cohort in a 25-day period prior to lymphodepleting (pre-LD) chemotherapy (n=33).

1.6 Logistic Regression Model Generation and Validation

Based on factors found to be associated with carHLH, we sought to develop multiparameter logistic regression models predictive of future development of carHLH. An initial model was generated using 55 of 59 patients with full data available. To generate this model, following an initial screening by univariate methods, for those parameters for which $p<0.05$, multivariable logistic regression analysis using both backward and stepwise selection was used to identify a set of factors which could jointly impact development of carHLH. Because all but one of the carHLH occurrences were after day 7, prediction models concentrated on factors which were known at or before day 7, with exception for the maximum grade of CRS. Patients with missing values were excluded from analysis. Levels of clinically actionable cytokines (which could readily be obtained after CRS onset but prior to carHLH onset, i.e., IFNγ, IL-1β, IL-6, and TNFα) on day of CRS+1 or day of CRS+2, and baseline BM T:NK ratio were included in the analysis.

However, since no patients developed carHLH without first having CRS, a second model (Model A) was generated with a goal of predicting carHLH after CRS onset, using features available no later than day +2 after CRS onset (earliest day of carHLH onset). This model utilized 39 of 52 patients developing CRS who had full data available. To assess whether features would impact carHLH development, p-values were calculated using a Wald chi-square test, and parameters for which $p<0.05$ were included in model development. Patients without CRS and those with missing values were excluded from the analysis. Features were log-transformed (natural logarithm) before model fitting. A logistic regression function *Logit()* from the python *statsmodels* package was fit to the data.

Although the primary objective of this analysis was to identify factors associated with development of carHLH, an additional analysis was also performed restricted to the subset of patients in the discovery cohort whose products underwent TCS during CAR manufacturing

process (n=34), since all patients will undergo TCS moving forward. Additional analysis resulted in development of an predictive models, Model B. This model was otherwise generated as described for Model A, with exception that cytokines were not put into this model due to not enough data given the smaller numbers.

Model Cross-Validation:

Following model development, model performance was assessed using the main (discovery) cohort of the CD22 CAR T-cell trial and via randomly shuffled 3-fold cross-validation. For cross-validation, the data was partitioned into parts of equal size (folds). For each fold, a model was fitted to the data on the training cohort and tested on the validation cohort. Repeating this process, we obtained an AUC value for each fold. AUC values for each fold are averaged (mean AUC) to give an estimate for the overall model performance. Cross-validation and model selection was performed using the python *sklearn* package.

1.7 CAR T-cell Product Manufacturing

All products were manufactured at the NIH Clinical Center, Center for Cellular Engineering (CCE). Initial manufacturing strategies incorporated elutriation (n=6) and flask adhesion methods with utilization of anti-CD3/anti-CD28 magnetic beads (n=19) as an enrichment step for T-cell purification (TCE). Following the initial experience, as previously described, CD4/CD8 Miltenyi® beads were incorporated for T-cell selection (TCS) of all starting apheresis products without any other downstream changes in the manufacturing process (Supplemental Figure 6C).

CAR product analysis by flow cytometry was performed at the following timepoints: “apheresis bag” refers to flow performed immediately after lymphocyte collection by lymphapheresis. CD4/CD8 T-cell section (TCS) was performed on apheresis bag collection prior to cryopreservation. CD3/CD28 T-cell enrichment (TCE) was performed following thawing of cryopreserved products. The “post-selection” timepoint refers to flow cytometry performed after thawing of TCS products, or directly after TCE (which had undergone freeze/thaw prior to enrichment). Enriched/selected products were then put into culture for further manufacture at “day 0”. “Culture day 7” timepoint refers to a flow taken 7 days after culturing of CAR product, and “culture day final” timepoint refers to a flow taken on the final day the CAR product was in culture,

at which time it was either infused fresh from culture into the patient, or cryopreserved for infusion at a later date.

The vast majority of products (n=56) were cultured for 9 days. The remaining products were cultured for 11 (n=1) or 8 (n=2) days.

Flows were captured for the vast majority of products at all timepoints with the following exception: due to limited ability to capture parameters early on in the trial, NK-cell percent was not included as a parameter for "culture day final" for the first n=18 subjects, after which time ability to capture additional variables was expanded. All 18 subjects with missing NK-cell percent data underwent TCE, 14 of whom went on to have CRS, and 2 of whom went on to have carHLH.

2.0. Supplemental Results

2.1 Timing and clinical characteristics of carHLH varies between different CAR T-cell constructs

CD19 CAR T-cells

To closely evaluate for carHLH beyond the CD22 CAR T-cell trial, a cohort of 88 patients who received a CD19/4-1BB CAR T-cell product on a phase I/II clinical trial (NCT02028455) was assessed. Outcomes have recently been reported.[13,14] Importantly, on this trial, CD19/BB products were manufactured using T-cell selection (TCS), as were a subset of our CD22/BB products.

The first notable difference was that amongst those with CRS, patients on the CD22 CAR T-cell trial were more likely to have a ferritin greater than 10,000 µg/L than those receiving CD19 CAR T-cells. Amongst those with CRS, 37 of 45 (82.2%) patients on the CD22 CAR T-cell trial had a peak ferritin greater than 10,000 µg/L versus 34 of 78 (43.6%) patients on the CD19/4-1BB trial, $p< 0.0001$. (Supplemental Figure 2) Due to both overall lower ferritin values and limitation on quantification of ferritin values above 10,000 µg/L on the phase I/II CD19/4-1BB trial, the standard definition from Neelapu et al. was used to identify those with carHLH, understanding the limitations of this approach.[15] Based on this definition, only 1 patient met criteria for carHLH. Employing less stringent definitions, 11 of 78 (14.1%) patients with CRS had evidence of a hyperinflammatory process akin to carHLH, however did not meet the full criteria. The median time to CRS onset was 5 days (range, 2-11 days) post CD19 CAR T-cell infusion, with HLH-like toxicities occurring at a median of 9 days post-infusion (range, 5-15 days). Coincidentally, intra-patient cross-trial comparison was facilitated through 5 patients who were treated with both

CD19/4-1BB CAR T-cells and CD22 CAR T-cells. Amongst these 5, two patients had HLH-like manifestations with CD19 CAR and no patient had carHLH following CD22 CAR T-cells.

In an additional cohort of patients with B-ALL (n=50) receiving a different CAR T-cell construct (CD19/28ζ CAR T-cells) on a phase I clinical trial (NCT01593696), based on clinical parameters there was no evidence of an extended or secondary inflammatory process in patients who developed CRS, although ferritin was not routinely measured. These comparisons suggest that the incidence, presentation and timing of carHLH differ amongst CAR T-cell construct and antigen targeted.

2.2 Cytokine profiling reveals temporally divergent patterns of inflammation amongst those receiving CD22 CAR T-cells (Figure 3, Supplemental Figure 3, Supplemental Table 3)

Baseline soluble IL-2Rα was assessed on a limited cohort prior to lymphodepleting (pre-LD) chemotherapy (n=33), and was generally higher in subjects who developed carHLH amongst those with CRS. Serial profiling of cytokines relative to CRS onset revealed unique patterns in patients with carHLH. (Supplemental Figure 2A, Supplemental Table 2A) As described in the main text, several cytokines, including interferon-γ (IFNγ), IL-1β, IL-8, IL-10, and MIP-1alpha peaked 2-4 days after CRS onset, and subsequently remained highly differentially elevated after adjustment for multiple testing in those with carHLH compared to those without it. Comparable analysis of cytokines in patients with low grade (grade 1-2) versus high-grade (grade 3-4) CRS did not reveal as stark differences, with differences in select cytokines limited to higher peak values without persistent elevation. However, after adjustment for multiple testing, no substantial differences between cytokine levels were observed, likely in part due to the small n of those with high-grade CRS. (Supplemental Table 2B) IL-18bp was not differentially expressed in those with higher grade CRS versus lower grade CRS, but numbers were limited.

Similar patterns of cytokine elevation were observed in those with carHLH when levels were plotted relative to CAR infusion. (Supplemental Figure 2B, Supplemental Table 3A) IFNγ, IL-1β, IL-8, and MIP-1alpha levels peaked 8-12 days after CAR infusion, and levels were highly elevated after adjustment for multiple testing from days 11-15, overlapping with median day of carHLH onset. Differential elevation in those with carHLH were also observed in IL-4, IL-10, and IL-13, as well as IL-6 levels, though this was confounded by administration of tocilizumab. (Supplemental Figure 3) Few substantial differences were found when comparing cytokine levels

relative to CAR infusion in those with high-grade CRS compared to low-grade CRS, again limited by the low n of the former group. (Supplemental Table 3B).

The differential and *persistent* hypercytokinemia following CRS onset supports the notion of carHLH as a unique and divergent pathway emerging from CRS – to be distinguished from simply a more severe form of CRS - where differences, if any, seem limited to higher peak values. Plotting data relative to CAR infusion illustrates similarly higher cytokine levels in those with carHLH, but visualizing relative to CRS onset may improve understanding of the pathophysiological process underlying associated clinical changes given that patients have variable time to onset of CRS toxicity.

2.3 Tocilizumab use confounds IL-6 levels (Supplemental Figure 4)

CarHLH incidence was higher in subjects who received tocilizumab for CRS. Specifically, 17 of 24 (71%) patients who received tocilizumab and 4 of 28 (14%) subjects who did not receive tocilizumab developed carHLH ($p<0.0001$). All patients who received tocilizumab and had carHLH toxicity met criteria after initiating the drug, with a time to onset ranging from 0 days to 12 days after the first dose of tocilizumab.

2.4 Lymphocyte phenotyping reveals higher T-cell/NK-cell ratios in those with carHLH (Supplemental Table 4, Supplemental Figure 5)

Serial profiling of lymphocyte phenotyping, limited to a basic T-cell, B-cell, and NK-cell panel was routinely performed in patients at established timepoints pre and post CAR T-cell infusion.

- *Baseline*: Prior to CAR therapy, TBNK values did not differ between those with and without CRS; however, appreciable differences were seen in those with and without carHLH. Specifically, patients who developed carHLH had higher PB CD3% and CD8%, and lower NK%. Consequentially, PB CD3:NK and CD8:NK ratios were substantially higher at baseline in carHLH patients. The corresponding BM CD3:NK ratio was also notably higher at this timepoint.
- *Day 0*: Patients with subsequent carHLH continued to have lower PB NK% and higher CD8:NK ratios.

- *Day 7*: The previously noted disparities seemingly normalized with no obvious differences seen in lymphocyte phenotyping between the two populations. Low level CAR T-cell expansion (median <1%) was appreciated in both groups.
- *Day 14*: Overlapping with median day of carHLH onset, peak CAR expansion is noted and is CD8 predominant. Patients with carHLH had higher PB CAR T-cell%, CD8% and CD3:NK ratios, as well as lower NK%.
- *Day 28*: Those with carHLH had higher PB CAR T-cell%, CD3%, CD8%, CD3:NK and CD8:NK ratios; as well as lower NK% and CD4:CD8 ratio.
- *Restaging:* Those with carHLH had higher BM CAR T-cell%, CD3%, CD3:NK and CAR T-cell:NK ratios, as well as lower NK%. The change in CD3:NK ratio from pre-CAR baseline BM to restaging BM is shown in Supplemental Figure 4 for individual patients, stratified by carHLH toxicity.

2.5 Clinical characteristics of cohort treated with CD22 CAR T-cells

The median age of patients enrolled on the cohort was 17 years old (range: 4-30), and 20 of 59 (33.9%) were female. Patients enrolled on this study were heavily pretreated: 69.5% received at least one prior allogeneic hematopoietic stem cell transplantation, 40.7% received prior blinatumomab, 25.4% received prior inotuzumab, and 64.4% received prior CD19-targeted CAR. However, there was no difference in prior treatments amongst those with or without carHLH, amongst those with CRS. 74.1% of patients had an objective response following treatment with CAR T-cells, including 80% (24/30) who had CRS without carHLH, and 90.5% (19/21) who had carHLH. Full clinical characteristics of the first 58 patients treated on trial were recently described.[16]

There was no association with anti-leukemia response and development of carHLH (see below)

	CR (MRD neg)	CR (MRD +)	PR	SD/PD
Total	37	5	2	14
# with HLH	14 (37.8%)	4 (80%)	1 (50%)	2 (14.9%)
# without HLH	23 (62.2%)	1 (20%)	1 (50%)	12 (85.7%)
p-value (CR/CR-MRD+ versus PR/SD/PD)	0.13			

p-value (CR-MRD neg vs CR-MRD+ versus PR)	0.21

Patients with TCS, however, were more likely to experience carHLH

		DL1-TCE	DL2-TCE	DL3-TCE	DL2-TCS	DL1-TCS
n (treated)		6	17	2	7	25
n (with CRS)		3	16	2	6	23
Amongst those with CRS	n (with carHLH), %	0 (0%)	3 (18.8%)	0 (0%)	5 (83.3%)	13 (56.5%)
	n (without carHLH) %	3 (100%)	13 (81.3%)	2 (100%)	1 (16.7%)	10 (43.5%)

2.6 Development and validation of logistic regression models to predict carHLH (Figure 5, Supplemental Table 5)

Model Development

Comprehensive evaluation of the clinical characteristics of the CD22 CAR patient cohort revealed T- and NK-cell ratios, tocilizumab use and incorporation of CD4/CD8 T-cell selection (TCS) of the apheresis product, (which we recently described),[2] were all independently associated with increased risk of carHLH (Figure 5A).

Based on the aforementioned factors found to be associated with carHLH, we sought to develop a multiparameter logistic regression model predictive of future development of carHLH. An initial model based on univariate and multivariate analysis from all 59 patients revealed that only TCS and baseline bone marrow T/NK ratios were relevant for a model and led to the following classification rule: there was a high risk of carHLH if: "1.9937 x TCS (1=Yes; 0=No) + 0.2415 x baseline bone marrow T/NK ratio" is > 2.71625. Using this model, Thus, 16/20 (80%) of those with HLH can be predicted as well as 27/35 (77.1%) of those without HLH. 43/55 (78% accuracy).

However, since no patients developed carHLH without first having CRS, and with a goal to utilize the model with data readily available no later than day 7 post-CAR infusion, or day +2 after CRS onset (the earliest days of carHLH onset in any of the patients in our cohort), two additional models were generated and further explored. Figure 5B summarizes the best performing regression model. Model A was generated with a goal of predicting carHLH after CRS onset, using

features available no later than day +2 after CRS onset (earliest day of carHLH onset). This model utilized 39 of 52 patients developing CRS who had full data available. Model B incorporated baseline BM T:NK ratio (odds ratio=4.82 (95% CI: 1.3139, 17.71; p=0.018)), and IFNγ on day of CRS+2 (odds ratio=4.56 (95% CI: 1.43, 14.5; p=0.01)), achieving a sensitivity of 76% and a specificity of 77% on the discovery cohort.

Given that all CAR products moving forward will be generated using the TCS-based manufacturing process, we then sought to develop models predictive of future carHLH restricted to only those patients who received TCS-based manufacturing products (n=34). In this restricted cohort, bone marrow T:NK ratio remained predictive, and an additional model was generated using this limited cohort (Model B). Model B was generated using 31 patients with full data available (of n=34 with TCS in the discovery cohort). When applied to these subjects, it had sensitivity=77% and specificity=71%. It solely incorporated baseline BM T:NK ratio (odds ratio=1.23 (95% CI: 1.01, 1.50; p=0.043)) into its predictions.

Model Assessment using Cross-Validation, and Prospective Independent Validation Cohort

Following model development, model performance was assessed using cross-validation of the main (discovery) cohort of the CD22 CAR T-cell trial. (Supplemental Table 5) For model A, a mean AUC=0.87±0.09 in 4-fold cross-validation was achieved (Figure 5C), and model B a mean AUC=0.86±0.04 using 3-fold cross validation was achieved (Figure 5D). Models B performed less well, achieving a mean AUC=0.73±0.2 in 3-fold cross-validation.

To further interrogate model performance in a prospective clinical setting, models were applied to an independent cohort of 7 additional subjects on the CD22 CAR T-cell trial who were treated in the interim of model genesis. (Supplementary Table 5) However, of the 7 additional subjects of the independent cohort, only 1 developed carHLH—a considerably lower incidence of carHLH than was seen in the discovery cohort, additionally relevant cytokine data was available for 4 of 7 patients in the validation cohort. Model A could predict carHLH in these 4 patients with 75% (3/4) accuracy, 100% (1/1) sensitivity, and 66.7% (2/3) specificity. Model B performed less well; with 71.4% (5/7) accuracy, 100% (1/1) sensitivity, and 66.7% (4/6) specificity.

2.7 T-cell selection skews towards increased CD3% and decrease NK-cell% during CAR T-cell manufacturing (Supplemental Figure 7)

As illustrated in Supplemental Figure 6B, selection (via TCS or via CD3/CD28 T-cell enrichment (TCE)) was restricted to selection of the starting material post-collection with no further changes in manufacturing. Specifically, either CD3/CD28 magnetic Dynabeads T-cell enrichment (TCE; n=25, 42.4%) or CD4 and CD8 Miltenyi particle T-cell selection (TCS; n=34, 57.6%) was performed on the collected apheresis product.

3.0 Supplemental Tables:

Supplemental Table 1. Complete data and statistical analysis for CRP (1C), Ferritin (1D), temperature (1E), LDH (1F), ALT (1G), AST (1H), total bilirubin (1I), triglyceride level (1J), creatinine (1K), ANC (1L), platelet count (1M), and fibrinogen (1N) from CRS onset, stratified by presence and absence of carHLH. Data corresponds to Figure 1 (primary manuscript).

Supplemental Table 2. Complete data and statistical analysis for CRP (2C), Ferritin (2D), temperature (2E), LDH (2F), ALT (2G), AST (2H), total bilirubin (2I), triglyceride level (2J), creatinine (2K), ANC (2L), platelet count (2M), and fibrinogen (2N) from CRS onset, stratified by CRS grades 1-2 versus grades 3-4. Data corresponds to Supplemental Figure 1.

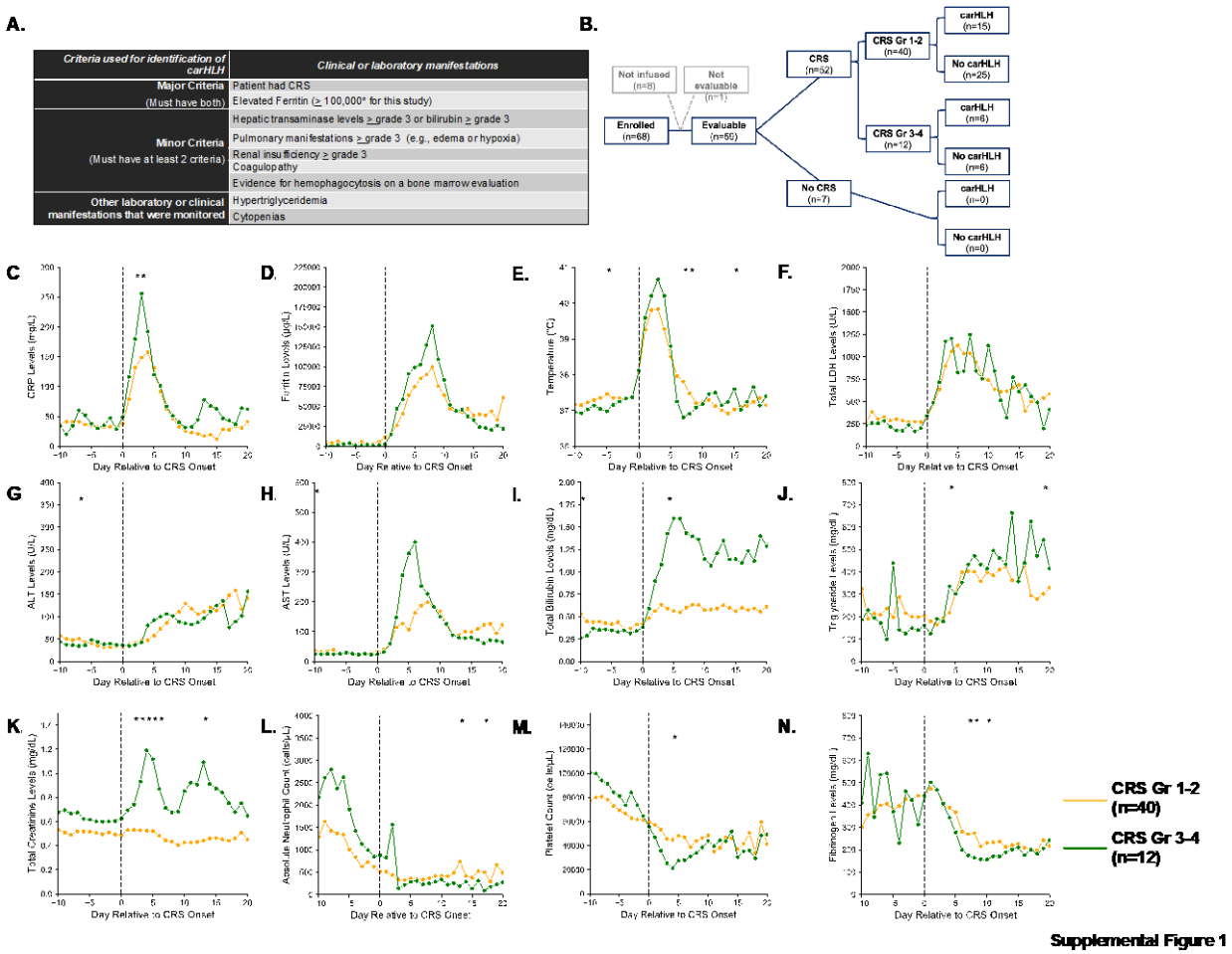
Supplemental Table 3. Complete data and statistical analysis for cytokine profiling from CRS onset, stratified by those with and without carHLH (3A) and those with CRS grades 1-2 versus grades 3- 4 (3B). Complete data and statistical analysis for cytokine profiling from CAR infusion, stratified by those with and without carHLH (3C) and those with CRS grades 1-2 versus grades 3-4 (3D). Data corresponds to Supplemental Figure 3.

Supplemental Table 4. CAR T-cell expansion and lymphocyte phenotyping post-infusion, distinguished by absence or presence of CRS and further stratified by carHLH toxicity

Supplemental Table 5. Logistic regression models for prediction of CarHLH.

Supplemental Table 6. List of differentially expressed genes by TCS versus TCE (6A) or carHLH versus no carHLH (6B).

4.0 Supplemental Figures



Supplemental Figure 1

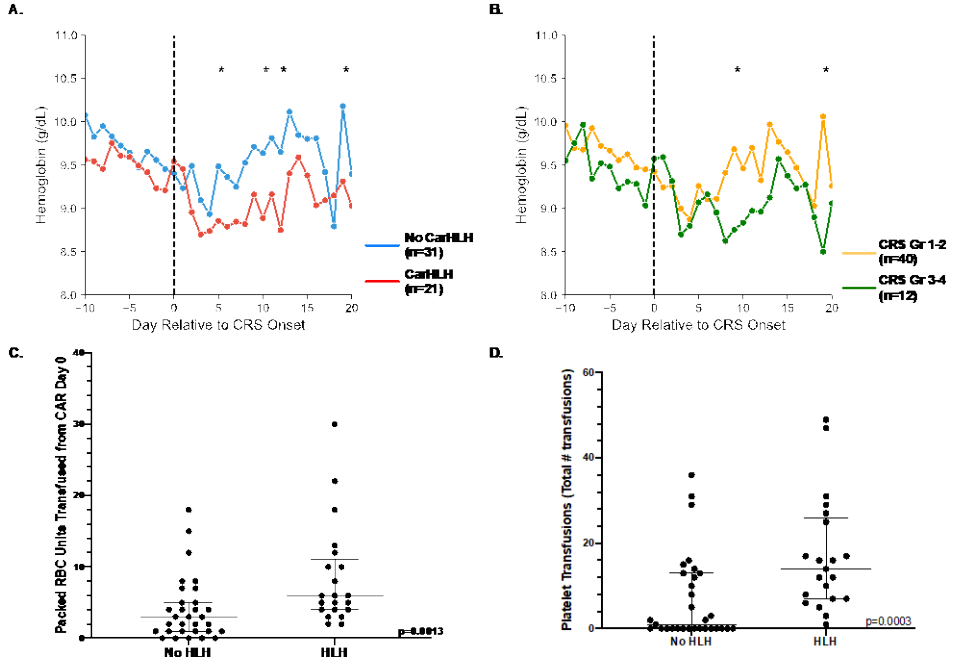
Figure 1. Clinical characteristics of carHLH in patients receiving CD22 CAR T-cells, stratified by CRS grade (1-2 vs 3-4); (See corresponding data file—Supplemental Table 2)
1A. Criteria used for identification of carHLH on this study

1B. Schematic of immune-associated toxicities of CD22 CAR T-cells stratified by CRS, CRS grade and subsequent development of carHLH. Amongst 68 enrolled subjects, 8 were not infused and 1 subject (subject #20) was deemed unevaluable for CAR toxicity due to the necessity to treat with high-dose chemotherapy due to rapid leukemia progression prior to CAR T-cell expansion and thus was not included in subsequent analysis. A total of 59 subjects were evaluable for CAR toxicity evaluation.

Subsequent figures all show laboratory values and clinical parameters in relationship to CRS onset, indicated by timepoint = 0 and stratified by CRS Grade 1-2 (yellow) to CRS grade 3-4 (green), with CRS graded by ASTCT criteria. Data are restricted to only patients who had CRS. Values which were blank or otherwise uninterpretable (e.g. “Hemolyzed”, “Lipemic”) were excluded. Dots indicate the aggregated mean for each timepoint, and the error bands show the estimated 68%

confidence interval. Dashed line indicates the date of CRS onset. Asterisks denote p<0.05 by Mann-Whitney U test.

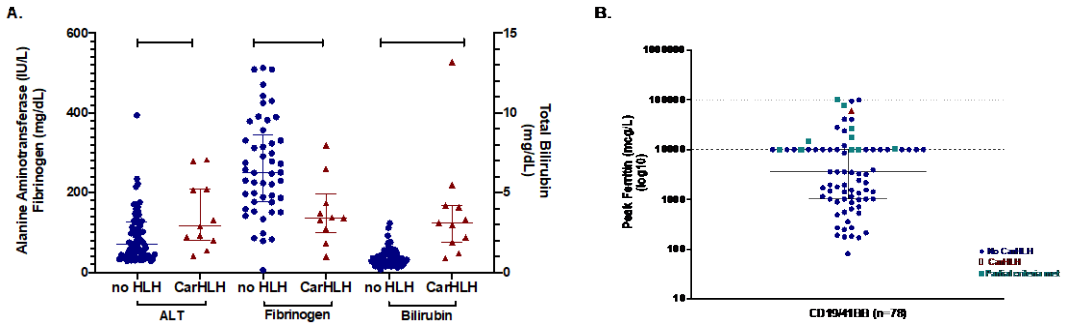
1C. C-reactive protein; statistics for each timepoint are found in Supplemental Table 2C
1D. Ferritin; Statistics for each timepoint are found in Supplemental Table 2D
1E. Temperature; Statistics for each timepoint are found in Supplemental Table 2E. Daily max temperature (Tmax) was obtained for each patient on each day relative to CRS onset, and aggregated for the figure and statistical analysis, stratified by carHLH status
1F. Total lactate dehydrogenase (LDH), Statistics for each timepoint are found in Supplemental Table 2F
1G. Alanine aminotransferase (ALT), Statistics for each timepoint are found in Supplemental Table 2G
1H. Aspartate aminotransferase (AST), Statistics for each timepoint are found in Supplemental Table 2D
1I. Bilirubin (total), Statistics for each timepoint are found in Supplemental Table 2I
1J. Triglycerides, Statistics for each timepoint are found in Supplemental Table 2J
1K. Creatinine, Statistics for each timepoint are found in Supplemental Table 2K
1L. Absolute neutrophil count, Statistics for each timepoint are found in Supplemental Table 2L
1M. Platelet count, Statistics for each timepoint are found in Supplemental Table 2M. Daily minimum platelet count was obtained for each patient on each day relative to CRS onset, and aggregated for the figure and statistical analysis, stratified by CRS severity.
1N. Fibrinogen, Statistics for each timepoint are found in Supplemental Table 2N



Supplemental Figure 2

Supplemental Figure 2. Anemia, thrombocytopenia and transfusion requirements
2A. Hemoglobin (g/dL) relative to CRS onset, stratified by those with carHLH (Red, n=21) and those without carHLH (Blue, n=31)

2B. Hemoglobin (g/dL) relative to CRS onset, stratified by those with Grades 1-2 CRS (Yellow, n=40) and those with Grades 3-4 CRS (Green, n=12)
2C. Total packed red blood cell units administered from Day 0 to day 30, restricted to those with CRS, stratified by those with and without carHLH.
2D. Total # of platelet transfusions (not individual platelet products) administered from Day 0 to day 30, restricted to those with CRS, stratified by those with and without carHLH.

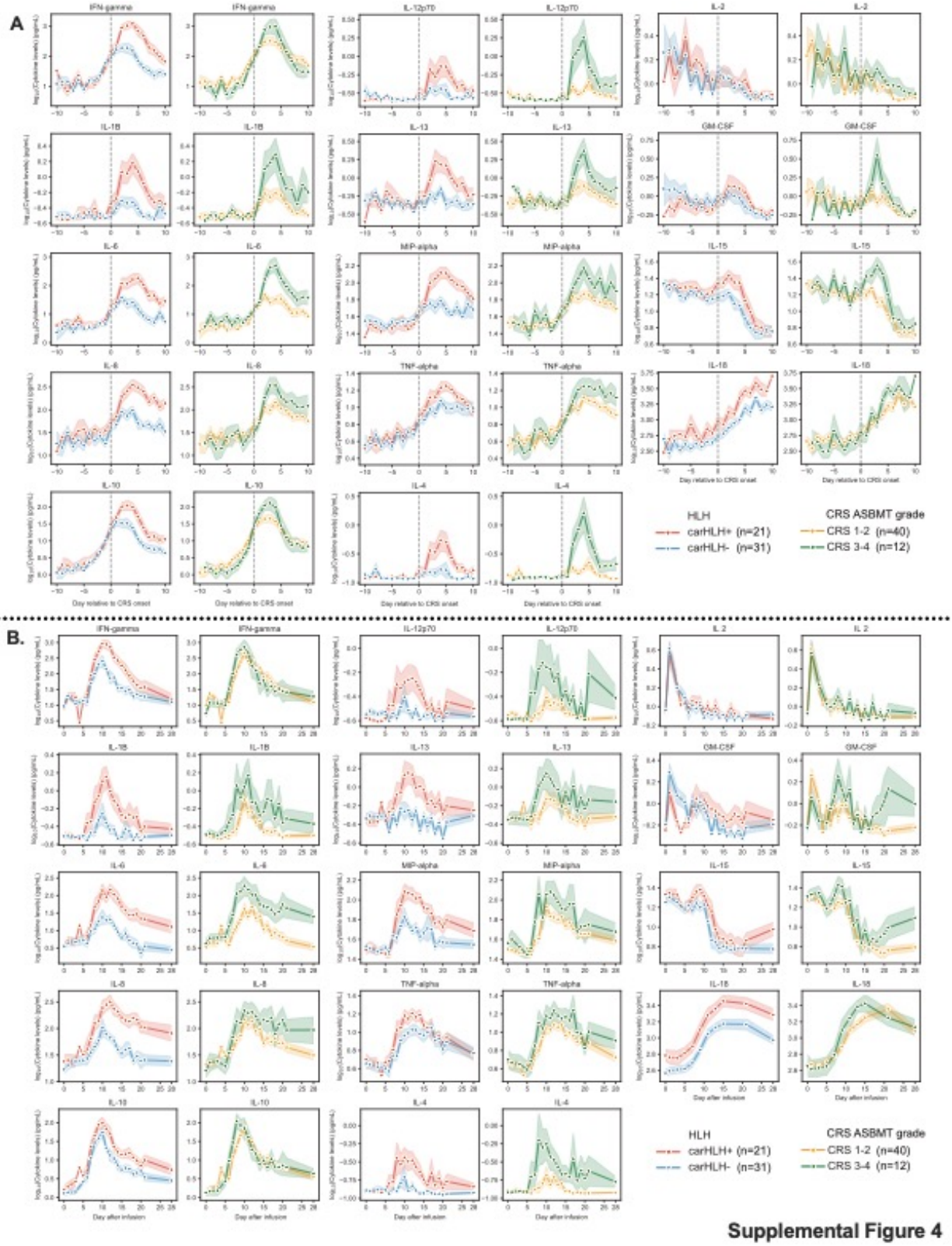


Supplemental Figure 3

Supplemental Figure 3. Clinical characteristics of carHLH in patients receiving CD19 CAR T-cells

3A. Clinical manifestations of peak hepatic transaminase levels (left y-axis) (ALT: alanine aminotransferase), fibrinogen nadir (left y-axis) and peak total bilirubin levels (right y-axis) amongst those with and without carHLH. Graph shows scatter plot of all values with lines at median and IQR. (*=p<0.05; **=p<0.01; ***=p<0.001; ****=p<0.0001)

3B. Comparison of peak ferritin levels amongst patients receiving CD22/41BB CAR T-cells and CD19/41BB CAR T-cells. Amongst those receiving CD22 CAR T-cells, n=45 because at the initiation of the study, ferritin was not routinely checked, thus there are 7 patients who do not have ferritin measurements. For those receiving CD19/41BB CAR T-cells, which were at Seattle Children's Hospital, the upper limit for quantification of ferritin was at 10,000 µg/L in the majority of patients, thus the true peak ferritin level in these patients is unknown. Ferritin levels > 10,000 µg/L were quantified in a limited cohort of patients, for whom the peak value is accounted for.

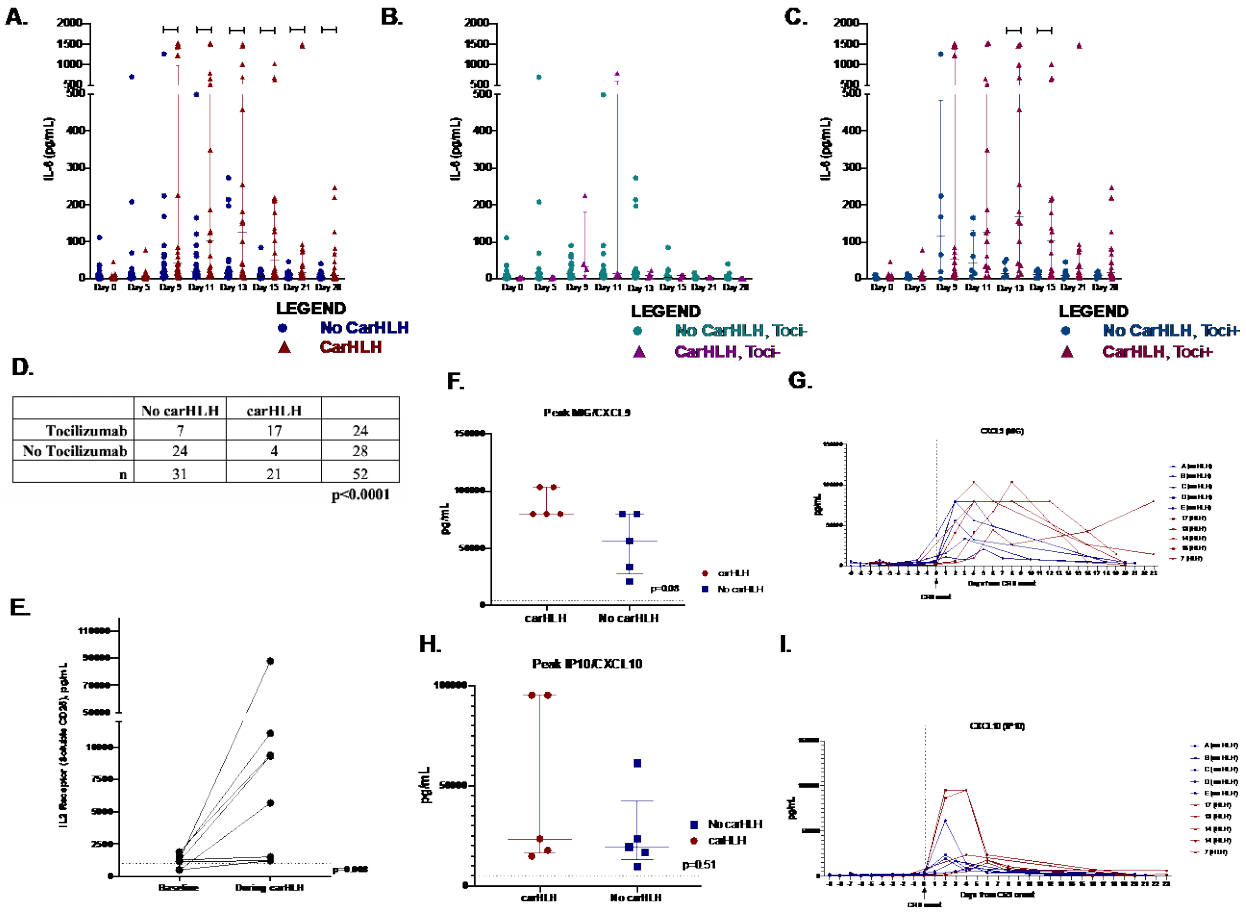


Supplemental Figure 4

Supplemental Figure 4. Serial cytokine profiling in patients receiving CD22 CAR T-cells relative to day of CRS onset (top) and relative to day of CAR T-cell infusion (bottom).

4A. Various cytokine values are shown in relationship to CRS onset, indicated by timepoint=0 (dashed line). For each timepoint we aggregated individual patient values stratified by carHLH status and also compared cytokine profiles amongst those with CRS grades 1-2 (indicated in yellow) versus CRS grades 3-4 (indicated in green). Dots indicate the aggregated mean and the error bands show the estimated 68% confidence interval. Cytokines with differential expression in patients with carHLH (as indicated in red) include, IFNγ, IL-1β, IL-6, IL-8, IL-10, IL-12p70, MIP-1α and IL-4. Cytokines with considerable overlap in values between those with and without carHLH (as indicated in blue) include IL-2, GM-CSF, and TNFα. IL-18 shows parallel increase in level over time. There was no differential expression amongst any cytokine when comparing CRS Grades 1-2 versus 3-4. Concurrent data/statistical analysis shown in Supplemental Tables 3A and 3B.

4B. Various cytokine values are shown in relationship to CAR T-cell infusion (Day 0). For each timepoint we aggregated individual patient values stratified by carHLH status and also compared cytokine profiles amongst those with CRS grades 1-2 (indicated in yellow) versus CRS grades 3-4 (indicated in green). Dots indicate the aggregated mean and the shaded error bands show the estimated 68% confidence interval. Substantial elevations were seen at multiple timepoints for the following cytokines in those with carHLH: IFNγ, IL-1β, IL-4, IL-6, IL-8, and MIP-1α. There was no substantial differential expression amongst any cytokine when comparing CRS Grades 1-2 versus 3-4. Concurrent data/statistical analysis shown in Supplemental Tables 3C and 3D. .



Supplemental Figure 5

Supplemental Figure 5. Cytokine profiling of IL-6 with and without tocilizumab in CD22 CAR patients, Soluble IL2R and CXCL9 levels

Plasma IL-6 concentration was measured serially following CAR T-cell administration.

5A. Substantially higher IL-6 concentrations are seen in the cohort with carHLH (compared to those without carHLH) from day 9 through day 28 timepoints.

5B. However, when the cohort is stratified by administration of tocilizumab, there are no differences in IL-6 concentration at any timepoint when comparing those without carHLH who did not receive tocilizumab (*turquoise circles)*, and those with carHLH who did not receive tocilizumab (*plum triangles)*.

5C. For those who did receive tocilizumab, higher levels are seen in the group with carHLH (*salmon-red triangles)* compared to those without carHLH (*blue-gray circles)* at days 13 and 15.

5D. Fisher's exact test for those who did and did not receive tocilizumab versus presence of carHLH versus no carHLH.

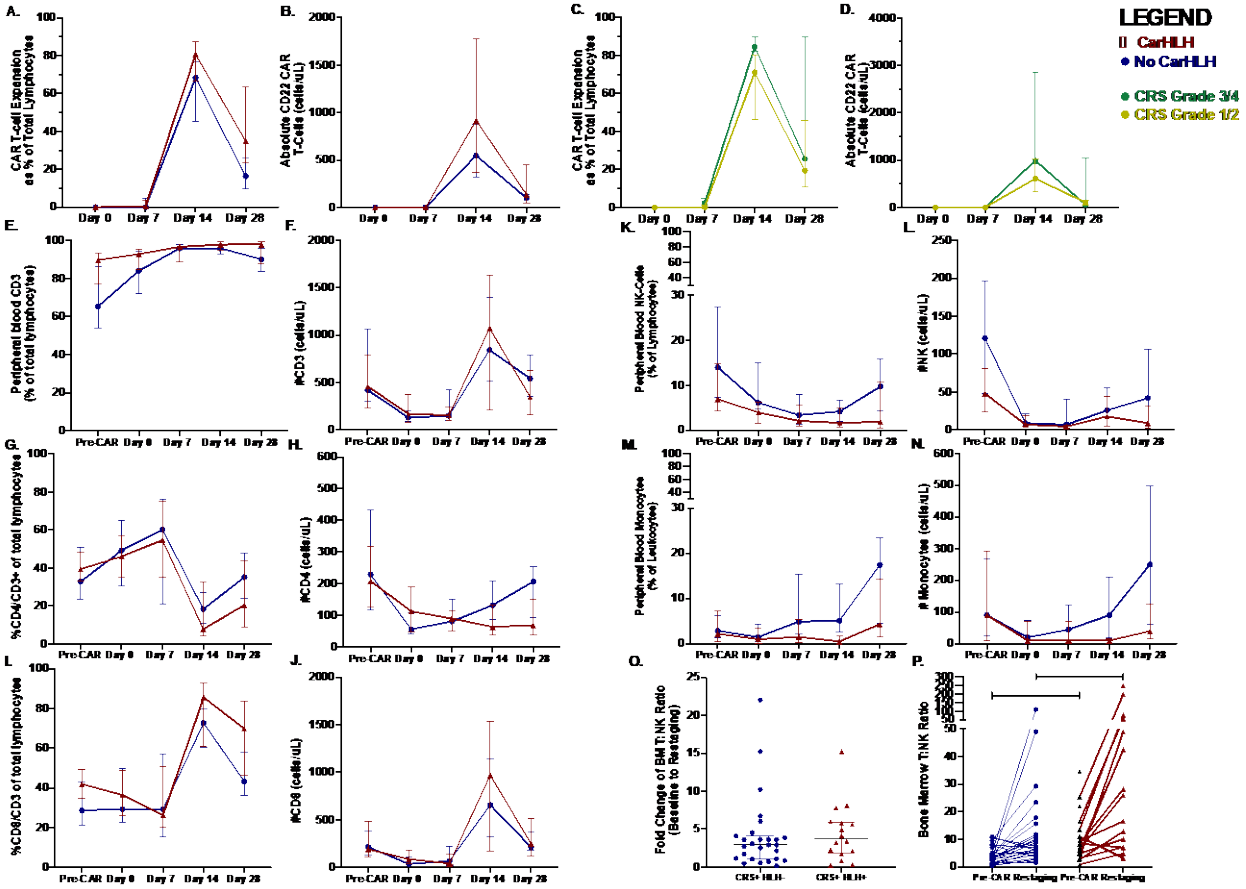
5E. Serial IL2 receptor (CD25) levels in select patients where levels were sent at various timepoints during carHLH.

5F. Peak CXCL9 (MIG) levels in 10 patients with CRS (5 with and 5 without carHLH)

5G. Serial CXCL9 (MIG) levels in 10 patients with CRS (5 with and 5 without carHLH)

5H. Peak CXCL10 (IP10) levels in 10 patients with CRS (5 with and 5 without carHLH)

5I. Serial CXCL10 (IP10) levels in 10 patients with CRS (5 with and 5 without carHLH)
[For CXCL9 levels, the upper limit of detection was set at >80,000 pg/mL, with two samples that were able to be diluted out to a slightly higher upper limit; For CXCL10, the upper limit was > 95400 pg/mL. These studies were conducted on an exploratory basis only and were not performed as part of the original analysis. Limitations based on sample availability] .



Supplemental Figure 6

Supplemental Figure 6. Serial lymphocyte phenotyping following product infusion in patients with vs without carHLH. (*=p<0.05; **=p<0.01; ***=p<0.001; ****=p<0.0001)
**Figures reproduced from Figure 3 for purposes of completeness and comparison
6A(Primary figure 4I).** CD22 CAR T-cells are first detectable in peripheral blood flow around the time of CRS onset (median day 8) and levels generally peak at the day 14 timepoint, which coincides with median day of carHLH onset. CD22 CAR T-cells comprise a higher percentage of total lymphocytes (as well as total T-cells, data not shown) in patients with carHLH at the day 14 timepoint, as well as at the day 28 timepoint. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.

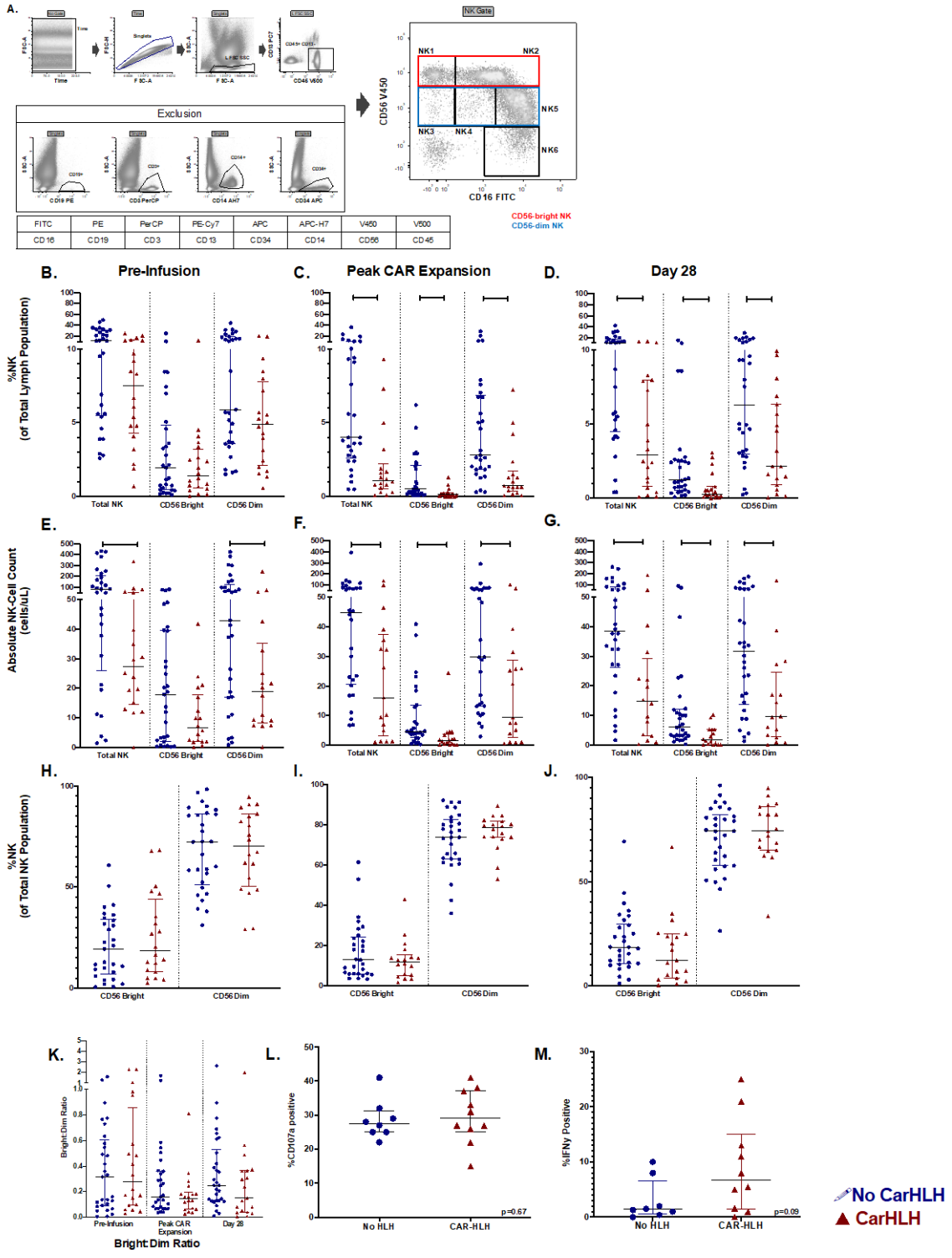
6B. Absolute CAR T-cell counts are not different at any time point between those who do and do not develop carHLH. Thus, the increased CAR expansion is a relative percentage with no significant differences in overall T-cell count at any time point. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6C. Patients with severe CRS (grade 3 or 4) have a higher percentage of T-cells which are CD22 CAR T-cell positive at day 14, but substantial differences do not persist out to the day 28 timepoint, in contrast with carHLH. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6D. Absolute CAR T-cell counts are not different at any time point between those with and without severe CRS. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6E. In carHLH patients, %CD3 cells (as a percentage of total lymphocytes) comprise a larger percentage of the cellular milieu at pre-CAR infusion baseline, and at the day 28 timepoint. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6F. Absolute CD3 counts are not different at any time point between those who develop and do not develop carHLH. Thus, the increased CAR expansion is a relative percentage with no significant differences in overall T-cell count at any time point. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6G. In carHLH patients, the percentage of CD4/CD3+ cells (as a percentage of total lymphocytes) in the cellular milieu does not significantly differ at any time point. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6H. Absolute CD4 count is higher in the cohort without carHLH at the day 14 and 28 timepoints, reflecting an absence of CD4 count recovery in the carHLH cohort at this time. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6I. In carHLH patients, the percentage of CD8/CD3+ cells (as a percentage of total lymphocytes) comprise a larger percentage of the cellular milieu at pre-CAR infusion baseline, and on day 14 and 28 timepoints, though not on day of CAR infusion itself. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6J. Absolute CD8 counts are not different at any time point between those who develop and do not develop carHLH. A notable spike in CD8 cells matches corresponding spike of absolute CAR T-cells without substantial increase in CD4 T-cells, indicating that CAR T-cell expansion is CD8 predominant. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6K (Primary figure 4J).** In subjects with carHLH, NK cells comprise a smaller percentage of the cellular milieu at pre-CAR infusion baseline, on the day of CAR infusion, as well as at the day 14 and day 28 timepoints. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6L. Absolute NK counts are lower at pre-CAR infusion baseline and on the day 28 timepoint. Patients with carHLH are characterized by an absence of NK-cell re-expansion at the latter timepoint, compared to a marked re-expansion towards baseline levels in those without carHLH. NK-cells are the only cell type which differ both in absolute and relative amounts, and patients with carHLH are NK-cell lymphopenic throughout the entirety of the study period. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.
6M (Primary Figure 4L).** In carHLH patients, monocytes do not differ at early timepoints but comprise a smaller percentage of the cellular milieu at day 7 timepoint onwards, roughly corresponding with median day CRS onset (day 8). Differences were most pronounced at day 14

timepoint, corresponding with median day of carHLH onset, but persisted through day 28 timepoint. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.

6N. Absolute monocyte counts are substantially lower in those with carHLH at day 14 and day 28 timepoints, but do not substantially differ at earlier timepoints. Line graphs are drawn through median value for each timepoint, with thin vertical lines denoting IQR.

6O. The fold change of CD3:NK cell ratio from pre-infusion baseline to restaging bone marrow (CD3:NK ratio at restaging / CD3:NK ratio pre-infusion) do not differ between those with and without carHLH. Graphs show scatter plot of all values with lines at median and IQR.

6P. Bone marrow CD3:NK cell ratio is denoted for patients with and without carHLH at pre-CAR infusion and restaging time points. Most patients have increased CD3:NK cell ratio at restaging compared to pre-CAR infusion timepoint. Patients with carHLH have a higher CD3:NK ratio at both pre-CAR infusion and restaging timepoints. Graphs show scatter plot of all values with lines at median and IQR. Patients who have readings at both time points are connected by a solid line.



Supplemental Figure 7

Supplemental Figure 7. NK-cell subset analysis

Graph shows scatter plot of all values with lines at median and IQR. (*=p<0.05; **=p<0.01; ***=p<0.001; ****=p<0.0001)

7A. Flow cytometry dot plots showing the gating and subgrouping of NK cells on a representative case. NK cells are gated on alive, single, CD45+ CD3- CD13- CD14- CD19- CD34- lymphoid cells (by FSC/SSC), and further subclassified based on the differential expression of CD56 and CD16.

7B. Total NK-cell, CD56bright, and CD56dim subset percentage of the total lymphocyte population Pre-CAR infusion in those who would and would not go on to develop carHLH. There is a trend towards lower total NK% (p=0.05), and CD56dim% (p=0.07) in those who go on to develop carHLH, without difference in the CD56bright%.

7C. Total NK-cell, CD56bright, and CD56dim subset percentage of the total lymphocyte population at Peak CAR Expansion in those who would and would not go on to develop carHLH. There are substantially lower relative populations of all three NK-cell groups in those with carHLH relative to those without carHLH.

7D. Total NK-cell, CD56bright, and CD56dim subset percentage of the total lymphocyte population at Day 28 timepoint following infusion in those who would and would not go on to develop carHLH. There are substantially lower relative populations of all three NK-cell groups in those with carHLH relative to those without carHLH.

7E. Total NK-cell, CD56bright, and CD56dim subset absolute cell count Pre-CAR infusion in those who would and would not go on to develop carHLH. There are substantially lower NK-cell counts in the total NK and the CD56dim subset in those who would go on to develop carHLH, without difference in the CD56bright subset.

7F. Total NK-cell, CD56bright, and CD56dim subset absolute cell count at Peak CAR Expansion in those who would and would not go on to develop carHLH. There are substantially lower populations of all three NK-cell groups in those with carHLH relative to those without carHLH.

7G. Total NK-cell, CD56bright, and CD56dim subset absolute cell count at Day 28 following infusion in those who would and would not go on to develop carHLH. There are substantially lower populations of all three NK-cell groups in those with carHLH relative to those without carHLH.

7H. CD56bright and CD56dim subsets of the total NK-cell population pre-infusion in those who would and would not go on to develop carHLH.

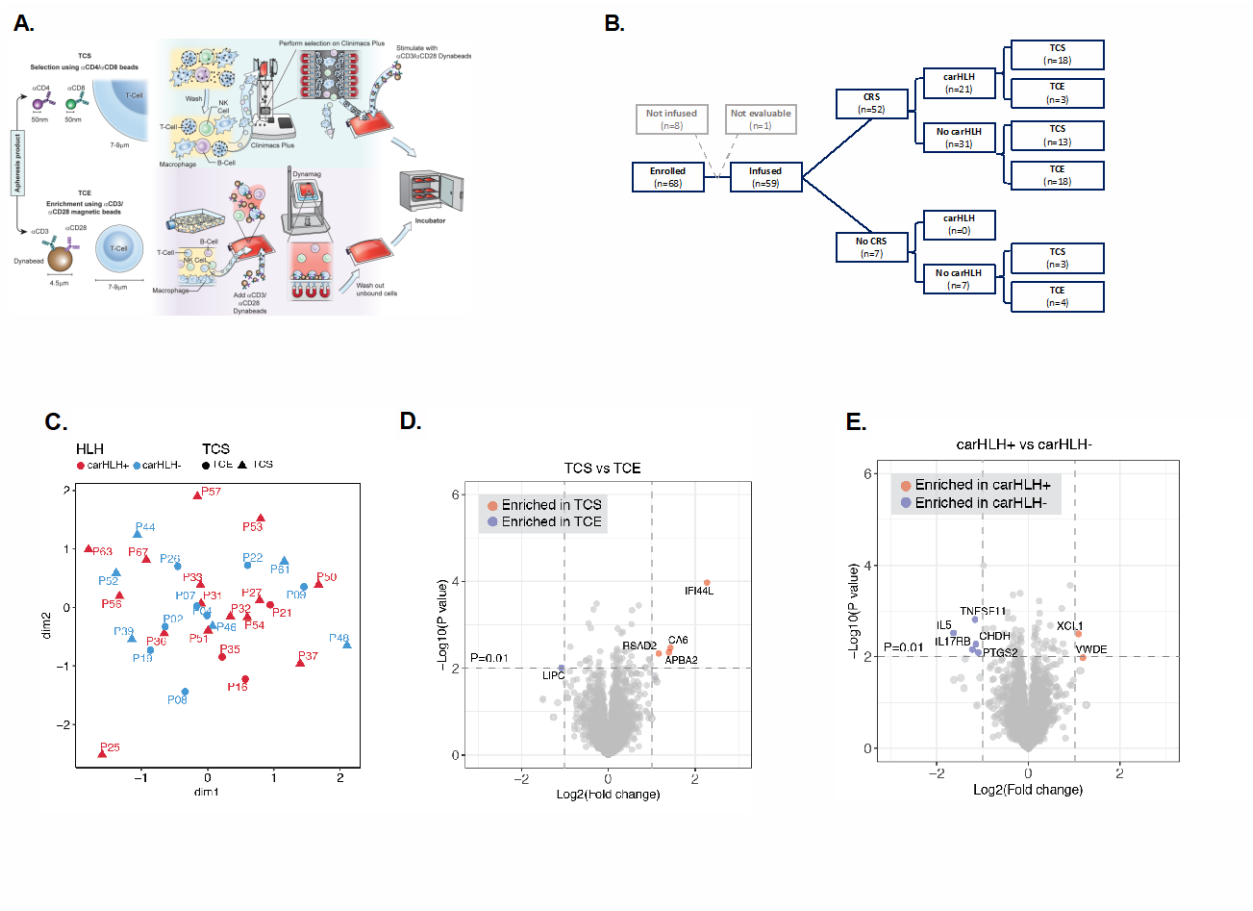
7I. CD56bright and CD56dim subsets of the total NK-cell population at Peak CAR Expansion in those who would and would not go on to develop carHLH.

7J. CD56bright and CD56dim subsets of the total NK-cell population at Day 28 timepoint following infusion in those who would and would not go on to develop carHLH.

7K. Ratio of CD56bright:CD56dim populations at all three timepoints. There were no substantial differences between the subsets in those with or without carHLH, consistent with NK% of the Total NK Population findings, and suggestive that the NK-cell deficit associated with carHLH is global.

7L. Degranulation, as measured by NK cell expression of CD107a after coculture with K562 leukemia target cells, did not differ in patients at peak CAR expansion between those with and without carHLH.

7M. Expression of IFNγ after coculture with K562 leukemia target cells, did not differ in patients at peak CAR expansion between those with and without carHLH.



Supplemental Figure 8

Supplemental Figure 8. Impact of TCS on carHLH and transcriptional profiling of CAR T-cell infusion products (*=p<0.05; **=p<0.01; ***=p<0.001; ****=p<0.0001)

8A. Illustration of CAR manufacturing process using either CD4/CD8 T-cell selection (4/8 TCS) (top) or CD3/CD28 enrichment (3/28 TCE) (bottom)

8B. Schematic of immune-associated toxicities of CD22 CAR T-cells. Following CD4/CD8 T-cell selection (TCS), there was a higher incidence of carHLH (18/34, 52.9%) compared to CD3/CD28 T-cell enrichment (TCE) (3/25, 12.0%) (p=0.002), though this did not impact the incidence of high grade CRS (Table 1).

8C. PCA plot using normalized gene expression values of the top 500 variable genes does not show distinct clustering of carHLH+ and carHLH- (color) or TCS vs TCE (symbol).

8D and 8E. Volcano plot of differentially expressed genes (only protein coding genes) between clinical features (6E – TCE vs TCS and 6F – carHLH+ vs carHLH-). Genes with a |FC|>1 and P-value < 0.01 are colored in either red (upregulated genes) or blue (downregulated genes) or grey (not significant). No genes reached significant threshold (FDR < 0.01).

References:

1. Shah NN, Highfill SL, Shalabi H, et al. CD4/CD8 T-Cell Selection Affects Chimeric Antigen Receptor (CAR) T-Cell Potency and Toxicity: Updated Results From a Phase I Anti-CD22 CAR T-Cell Trial. *J Clin Oncol*. 2020:Jco1903279.
2. Shah NN, Highfill SL, Shalabi H, et al. CD4/CD8 T-Cell Selection Affects Chimeric Antigen Receptor (CAR) T-Cell Potency and Toxicity: Updated Results From a Phase I Anti-CD22 CAR T-Cell Trial. *J Clin Oncol*. 2020;38(17):1938-1950.
3. Lee DW, Gardner R, Porter DL, et al. Current concepts in the diagnosis and management of cytokine release syndrome. *Blood*. 2014;124(2):188-195.
4. Lee DW, Santomasso BD, Locke FL, et al. ASTCT Consensus Grading for Cytokine Release Syndrome and Neurologic Toxicity Associated with Immune Effector Cells. *Biol Blood Marrow Transplant*. 2019;25(4):625-638.
5. Michel T, Poli A, Cuapio A, et al. Human CD56bright NK Cells: An Update. *J Immunol*. 2016;196(7):2923-2931.
6. Amand M, Iserentant G, Poli A, et al. Human CD56(dim)CD16(dim) Cells As an Individualized Natural Killer Cell Subset. *Front Immunol*. 2017;8:699.
7. Andrews S. FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/; 2012.
8. Daley T, Smith AD. Predicting the molecular complexity of sequencing libraries. *Nat Methods*. 2013;10(4):325-327.
9. Stroncek DF, Ren J, Lee DW, et al. Myeloid cells in peripheral blood mononuclear cell concentrates inhibit the expansion of chimeric antigen receptor T cells. *Cytotherapy*. 2016;18(7):893-901.
10. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*. 2012;28(16):2184-2185.
11. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *2011*. 2011;17(1):3.
12. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
13. Gardner RA, Ceppi F, Rivers J, et al. Preemptive mitigation of CD19 CAR T-cell cytokine release syndrome without attenuation of antileukemic efficacy. *Blood*. 2019;134(24):2149-2158.
14. Gardner RA, Finney O, Annesley C, et al. Intent-to-treat leukemia remission by CD19 CAR T cells of defined formulation and dose in children and young adults. *Blood*. 2017;129(25):3322-3331.
15. Neelapu SS, Tummala S, Kebriaei P, et al. Chimeric antigen receptor T-cell therapy - assessment and management of toxicities. *Nat Rev Clin Oncol*. 2018;15(1):47-62.
16. Nirali N. Shah SLH, Haneen Shalabi, Bonnie Yates, Jianjian Jin, Pamela L. Wolters, Amanda Ombrello, Seth M. Steinberg, Staci Martin, Cindy Delbrook, Leah Hoffman, Lauren Little, Anusha Ponduri, Haiying Qin, Haris Qureshi, Alina Dulau-Florea, Dalia Salem, Hao-Wei Wang, Constance Yuan, Maryalice Stetler-Stevenson, Sandhya Panch, Minh Tran, Crystal L. Mackall, David F. Stroncek, and Terry J. Fry. CD4/CD8 T-Cell Selection Affects Chimeric Antigen Receptor (CAR) T-Cell Potency and Toxicity: Updated Results From a Phase I Anti-CD22 CAR T-Cell Trial. *Journal of Clinical Oncology*. 2020.