Supplemental Appendix

Table of Contents

1.0 Supplemental Methods

1.1 Assessments and Definitions

1.1.1 Assessment Performed

Disease assessments were performed on all patients at baseline within 28 days prior to initiation of LD and at day 28 +/-4 days. Assessments included: bone marrow (BM) aspirate, BM biopsy, and BM Flow Cytometry (FC), cerebrospinal fluid (CSF) cell counts, CSF cytopathology, and CSF FC as feasible. The NCI Flow Cytometry Laboratory performed minimal residual disease (MRD) assessment with a validated limit of detection of ALL blasts in the bone marrow at 0.002% of total cells.[1,2]

1.1.2 Bone Marrow Definitions

M1 (< 5% blasts), M2 (5-25% blasts), and M3 (>25% blasts) were used for morphology. High-burden disease is defined as ≥5% bone marrow blasts. Minimal residual disease (MRD) negativity was defined as <0.01% detectable leukemic blasts of mononuclear cells by multiparametric FC. CD19+ was defined as having ≥ 90% CD19 expression by FC, or ≥15% by immunohistochemistry stains, along with any level of CD22 positivity.

1.1.3 CNS Definitions

CNS1: absence of blasts on cytospin preparation, regardless of the number of white blood cells (WBC); CNS2: presence of < 5/uL WBCs in the CSF and cytospin positive for blasts, CNS3: ≥ 5 WBC/uL in the CSF and cytospin positive for blasts or any clinical evidence of CNS leukemia (e.g., leptomeningeal enhancement or definitive CNS lesion).

1.2 Manufacturing of CAR T-cells

Automated T-cell transduction (TCT) was performed on the CliniMACS Prodigy. Where not specified otherwise, reagents and materials were obtained from Miltenyi Biotec. In brief, patient leukapheresis products were either obtained fresh or frozen and thawed for these studies. Fresh cells were washed once and frozen cells were washed twice in Plasmalyte-A+5% human AB serum (Valley Biomedical). Leukapheresis product targeting $3x10^9$ total CD3+ T-cells was selected using CD4 and CD8 GMP Reagent. Cultivation was initiated with 1×10^8 lymphocytes in

a total volume of 70 mL Tex CART-200 CM containing TexMACS medium, 3% Human AB serum (Valley Biomedical) + 200 IU/mL interleukin (IL)-2 (Clinigen) in the CentriCult-Unit. Cells were activated with 4ml MACS GMP TransAct.

Cells were transduced on day 1 using a multiplicity of infection (MOI) of 40 with a self-inactivating third-generation lentiviral vector encoding a bispecific CD19CD22 CAR, under the control of MSCV internal promoter (GMP-grade vector was manufactured for this investigator-initiated trial by Lentigen Technology Inc). In the original method (OM), transduction was terminated with a wash step at day 5. This however led to suboptimal product characteristics with decreased cell viability and decreased expansion (something not observed using healthy donor cells or other patient materials during validation). Consequently, a manufacturing change was investigated, and a modification to the process was made to perform the wash on day 3 rather than day 5, (modified method, MM). All other processes remained the same. On day 7, cells were collected and sampled for quality control testing and prepared for infusion.

1.3 Toxicity Assessments

1.3.1 Adverse Events

Adverse events (AE) $\geq$ grade 3 that were possibly or probably related to lymphodepletion and first CAR T-cell infusion were collected on all patients from protocol specified lymphodepletion through either day 28 or resolution, whichever occurred later. Max grade AE was captured per patient.

The neuro-symptom checklist (NSC) is a CAR T-cell specific observer-reported outcome measure that was developed to capture the severity (mild=1, moderate=2, severe=3) and duration (<24 hours, 24-48 hours, and >48 hours) of neurologic symptoms in the past week.[3] The primary caregiver who was with the subject throughout treatment was asked to complete the form at three timepoints: prior to infusion, at day 10 (+/-4 days) post-infusion, and at day 21-28 post infusion. The 12 symptoms listed on the checklist include visual and auditory hallucinations, unresponsiveness to commands, disorientation, depressed mood, distressed mood, drowsiness/sleepiness, difficulty speaking, pain, blurred vision, seizures, and other. We considered observed symptoms as related to CAR T-cell therapy if they were new or worsening from baseline

and occurred in the presence of either CRS or CAR T-cell expansion. In reporting neurologic symptoms, we did not include pain or fever/chills as they are not specific to neurotoxicity and already are known symptoms of CAR T therapy.

1.3.2 Cytopenias Infection, B-cell Aplasia Monitoring and Definitions

Complete blood counts with differential were obtained at least twice weekly during the first 28 days post-CAR infusion. Patients were monitored for signs of infection, with blood, fungal, and viral panels obtained in symptomatic patients. T-, B-, NK-cell panels (TBNK) were performed at baseline and day 28. Medical records were reviewed and data were abstracted in the peri-CAR setting, up until initiation of next line of therapy, or patient discharge from our institution.

Severe neutropenia was defined as grade 4 neutropenia (<500/mcL) and grade 3 and 4 thrombocytopenia was defined as platelet counts from 25-50K/mcL, and <25K/mcL respectively. B-cell aplasia (BCA) was defined by an absolute CD19+ cell count < 50/mcL in the peripheral blood.

1.4 Correlative Studies

1.4.1 Neurocognitive Testing

The Cogstate computerized battery and paper/pencil tests for processing speed and verbal fluency were administered to subjects ages 5 years and older pre- and post-infusion (one time between day 21-28). Tests used in the study are briefly described below:

Computerized Cogstate Test Battery

The Cogstate battery used for this CAR T-cell therapy trial consisted of computerized tests assessing core cognitive domains such as processing speed, attention, working memory, visual learning and memory, and executive functioning (https://cogstate.com). These tests have been validated, found to be sensitive to subtle cognitive changes, are easy to understand, have adequate test-retest reliability, exhibit minimal practice effects, are available in multiple languages, and have been used in clinical trials to assess cognitive functioning in children and adults with cancer.[4-8]

For this trial, Cogstate was administered using an iPad. All subjects were administered a brief practice session that consisted of a truncated version (approximately 1-2 minutes) of the five

tests to familiarize the subjects with the tasks before the actual scored modules. Speed of response is used to calculate z-scores (normative mean=0, SD=1) for Detection, Identification, and One-Back tests, accuracy of responses is used for the Once Card Learning Test, and total number of errors is used for the Groton Maze Learning Test. Cogstate has calculated the standardized scores so that higher z-scores indicate better performance on all tests. The five Cogstate tests administered on this trial are described below.

- *Detection Task (DT)* – This is a simple, reaction timed test that measures psychomotor and processing speed. Average reaction time measured in milliseconds is the primary outcome measure for this test.
- *Identification Task (IDEN)* – This is a timed task that uses a choice reaction paradigm and measures sustained visual attention. Average reaction time measured in milliseconds is the primary outcome measure for this test.
- *One Card Learning Task (OCLT)* – This task measures visual recognition memory and attention. Accuracy is the primary outcome measure for this test.
- *One Back Task* (OBT) – This task measures working memory and attention. Reaction time is the primary outcome measure for this test.
- *Groton Maze Learning Task* (GMLT) – This task is based on a well-validated maze learning paradigm and measures executive functioning (problem solving, working memory, and cognitive flexibility). The outcome measure for this test is the total number of errors in learning the pathway.

<u>Traditional Paper/Pencil Tests</u>

Processing Speed – Subjects were administered the Processing Speed subtests from the Wechsler Scales (WISC-IV for children 8-15 years, 11 months, or WAIS-IV for adults 16 years+) to assess speed of mental processing. The Symbol Search and Cancellation subtests are both paper-and-pencil timed tasks the required speed, visual scanning, and discrimination. Scaled scores (mean=10; SD=3) are derived for each subtest and combined to yield a composite Processing Speed Index (PSI; mean=100; SD=15).

Verbal Fluency – Subjects' verbal fluency was assessed using the F-A-S (B-H-R alternate form) Verbal Fluency subtest of the Delis-Kaplan Executive Function System (DKEFS) for 8 years and older or the Category Verbal Fluency subtest from the McCarthy Scales of Children's Abilities for younger children aged 5-7 years. Scaled scores are derived for the DKEFS subtest while z-scores were calculated for McCarthy Verbal Fluency and then transformed to scaled scores.

1.4.2 Expansion and Persistence of CD19/CD22 CAR T-cells

Samples were processed for 8-color flow cytometry evaluation within 24 hours of collection with cocktailed antibodies. Bispecific CAR T-cells were detected using the anti-idiotype monoclonal antibody 136.20.1 conjugated to Alexa 647 as previously described for detection of CD19 CAR-T cells.[9,10] If enough cells were present, detection of CD19/22 bispecific CAR T-cells were also detected with an additional antibody cocktail containing CD22 Fc fragment conjugated to Alexa 647 (R&D Systems) with a mouse serum pre-incubation step. If enough cells were obtained, a fluorescence minus one (FMO) cocktail was also included, which omits the CAR T-cell detection reagents.

The cell processing procedure included NH4Cl whole blood lysis, followed by phosphate-buffered saline wash and antibody incubation. Cells were then washed, pelleted, fixed in 1% formalin and stored at 4°C for <12 h prior to acquisition. A target of 1,000,000 cells per tube/cocktail was acquired on FACSCanto II flow cytometer (BD Biosciences). Regarding reagents, CD20-FITC was obtained from Dako/Agilent Technologies and CD19 PE-Cy7 was obtained from Beckman Coulter. All other antibody reagents were obtained from BD Biosciences.

Flow cytometry data analysis was performed using FCS Express Software (DeNovo Software, Glendale CA). Cells were gated by light scatter with antigen back-gating to verify relevant populations. Normal hematopoietic cells within the specimens served as internal positive/negative controls. T-cells were identified using CD3, CD45 and forward and side light scatter properties; concurrently, CD14(+) monocytes, CD34(+) blasts, and CD19(+) B-cells, including mature CD20(+) B-cells, were excluded from the analysis to ensure a clean and uniform T-cell population. The percent of T-cells showing a positive signal by mAb 136.20.1 and CD22 Fc fragment (if enough cells available) was reported. The FMO cocktail was analyzed to ensure that relevant cell populations (T-cells, monocytes, blasts, B-cells, etc.) showed appropriate antigen expression in the absence of CAR-T cell detection reagents.

CAR T-cell expansion, both for peak CAR T-cell expansion in the peripheral blood, and also CAR T-cell expansion in the D28 bone marrow evaluations, was also assessed as a function intensified versus standard LD.

1.4.3 Serial lymphocyte counts

Serial absolute lymphocyte counts from start of LD through day +14 were obtained from routine CBC performed as part of the patient course. Subjects were subsequently stratified by intensified LD (n=6) versus standard LD (n=18), and included all patients who received either first or second CAR T-cell infusion.

1.4.4 Cytokines

Serum cytokines were serially obtained during the first month and included IL1B, IL2, IL4, IL6, IL8, IL10, IL12p70, IL13, IL15, IL18, IFNγ, TNFα, GMCSF, and MIP1α. Cytokines excluding IL18 were measured using a multiplex format according to manufacturer's instructions (MesoScaleDiscovery V-plex, Gaithersburg, MD, USA) while IL18 was measured by ELISA (MBL Life Science). HAMA was measured using an ELISA kit from Eagle Biosciences.

1.4.4 CD19, CD22, CD19/22 Comparison Studies

Peak values of CAR T-cells in the peripheral blood and bone marrow were compared amongst responders from three trials conducted in the Pediatric Oncology Branch, CD19 (n=31), CD22 (n=51), and CD19/22 (n=12). Absolute peak CAR T-cells in the peripheral blood were calculated using the following formula: measured absolute lymphocyte counts x % T-cells x %CAR T-cells. Duration of CAR T-cell persistence was documented from the date of CAR infusion to the first day a patient had undetectable levels of circulating CAR T-cells measured by flow cytometry.

Peak CRP and serum cytokine levels were compared in all patients who had data available across the three trials. Serum cytokines that were performed on all three trials included IL1B, IL2, IL4, IL6, IL8, IL12p70, IFNγ, TNFα, and GMCSF. Additional serum cytokines IL10, IL18, MIP1α, IL15, and IL13 were assessed in patients treated on CD22 and CD19/22 CAR trials and therefore where only compared amongst responders on these two trials. Peak ferritin levels were

compared in all patients with data available from CD22 and CD19/22 as ferritin was not routinely assessed in the CD19 trial.

1.4.5 Human anti-mouse antibody (HAMA)

Given that the murine component of FMC63 is the most utilized of CAR T-cell constructs, we explored the question of anti-CAR immunogenicity, both as a function of prior CAR exposure and how it impacts disease response, by evaluating human anti-mouse antibody (HAMA) using Eagle Biosciences (Nashua, NH) HAMA ELISA (Cat. No. HAM31-K01). This "sandwich" ELISA captured human anti-mouse antibody between immobilized murine IgG and a horseradish peroxidase (HRP)- labeled murine IgG. After washes, detection of this immunocomplex attached to HRP was developed with a substrate solution before being measured on an absorbance microplate reader (Molecular Devices' SpectraMax).

The performed assays met the requirements for QC samples included on the plates to allow the reporting of the quantities of HAMA within the subjects' plasma samples. The 99% confidence normal cut-off is <25ng/mL, per the manufacturer's website, thus we used anything $\geq$ 25 ng/mL as our cut-off for positivity, with values < 25 and >0 ng/mL considered detectable.

1.4.6 Laboratory Investigations

For *ex vivo* cytokine analyses, cells were co-cultured at a 1:1 E/T ratio on a TECAN robotic platform programmed to perform automatic collection and replenishment of supernatants at frequent time points, from 1 to 72 hours post co-culture. Briefly, 10µl of culture supernatants were collected with minimal disruption of cell pellets and replenished with 10µl of fresh medium to keep total volume constant throughout the experiment. Collected supernatants were stored at -20°C until evaluation by cytokine bead array (BD Biosciences, San Diego, CA) and run on a Fortessa FACS cytometer (BD Biosciences) to measure cytokine concentrations over time.

Cell surface CD19 CAR expression was evaluated using either a PE-labeled monoclonal anti-FMC63 scFv antibody (Acro) or APC-labeled monoclonal anti-FMC63 scFv antibody.[10] CD22 CAR expression was monitored by staining with a recombinant human siglec-2/CD22 Fc chimera protein (R&D) followed by incubation with a PE- or APC-conjugated goat-anti-human IgG (Jackson ImmunoResearch).

1.5 Reinfusions

Reinfusions were available to subjects who had a response to the previous infusion with either a partial remission, stable disease with clinical benefit, or in patients who achieved a complete remission and had recurrence of disease, loss of BCA, and/or loss of CAR T-cell persistence. Subjects were required to have an additional dose of cryopreserved CD19.22.BBζ CAR T-cells.

2.0 Statistical Analyses

2.1 Neurotoxicity Statistical Analyses

Based on the Shapiro-Wilks test, all cognitive test scores were normally distributed except for the Cogstate Detection scores due to an outlier. Thus, the Detection scores were transformed to be normally distributed using log10 transformation. To examine differences from pre- to post-infusion, we used paired t-tests. To examine differences in test scores between groups (No/Low CRS [n=12; grades 0, 1] vs High CRS [n=5; grades 2, 3] with cognitive testing; Responders [n=11] vs. Not Responders [n=6] with cognitive testing) from pre- to post-infusion, we used repeated measures Analysis of Variance (ANOVA) with between (groups) and within (time) subject variables. We analyzed the neurocognitive data using SPSS version 21 with two-tailed comparisons and the alpha set at 0.05.

3.0 Supplemental Results

3.1 Product Characteristics

The median time from apheresis to cell infusion was 17 days (range: 7-148 days). This did not include two patients who had previously collected cryopreserved T-cells from a different CAR study. Nineteen of twenty (95%) infused patient products had a CD4 predominance, with the median CD4:CD8 ratio of products being 2.6 (range: 0.4-8.1). Interestingly, on the pre-apheresis lymphocyte phenotype analysis, only 12/20 (60%) patients had predominance of CD4+ T cells, with a median CD4:CD8 of 1.18 (range: 0.5-3.3). After T-cell selection, most products maintained consistency with the pre-apheresis analysis, with 12/20 patient products having CD4+

predominance. The median transduction efficiency was 71.9% (range: 54.2-86%; Supplemental Table 1).

3.2 Toxicity

3.2.1 Adverse Events

Headaches were notable on this study and were seen in 16/20 ALL patients treated. Except for one patient who had a grade 3 thunder-clap headache, fifteen patients had grade 1 or grade 2 headaches which were attributed to research and CAR T-cells. All headaches eventually resolved with supportive care. Other notable AEs that were seen on study included grade 3 febrile neutropenia (n=8) and grade 3 hypotension (n=4) (Supplemental Table 2).

New or worsening symptoms observed by caregivers of the 20 patients at the day 10 post-infusion evaluation, which appeared in the presence of CRS and/or CAR expansion, included drowsiness/sleepiness (n=8), distress (n=4), depressed mood (n=4), disorientation (n=1), and severe headache (n=1) (Supplemental Table 3).

3.2.2 Cytopenias, Infection, and B-cell Aplasia

Prior to lymphodepletion, the median ANC was 1.4 K/mcL (range: 0.20-8.75K/mcL), with four patients having severe neutropenia. After lymphodepletion and CAR T-cell infusion, all patients had some degree of neutropenia, with 16/20 (80%) patients having grade 4 neutropenia. The median ANC nadir was 0.17K/mcL (range: 0-1.2K/mcL) and occurred on day 12 (range: 1-18) post-CAR infusion. Seven patients continued to have severe neutropenia at day 28, three of whom had severe neutropenia prior to initiation of LD. The median ANC at day 28 was 0.62K/mcL (range: 0-3.7) and three patients received filgrastim, 2 requiring only intermittent dosing. Despite neutropenia occurring in all patients, only 2 patients had infections during the first 28 days of therapy, with one patient having both Bacteroides Fragilis bacteremia and influenza A at day 22, and another patient developing Staphylococcus Epidermis bacteremia at D+6. Before LD, the median platelet count was 117 K/mcL (range: 33-397), and 4 patients had baseline $\geq$ grade 3 thrombocytopenia. During the first 28 days post LD and CAR infusion, 10/20 (50%) patients had $\geq$grade 3 thrombocytopenia, with a median platelet nadir of 53K/mcL (range 5-176). At day 28, the median platelet count was 126K/mcL (range:11-262), and only 2 patients were transfusion

dependent, one of whom was transfusion dependent prior to CAR infusion. Median bone marrow cellularity at D28 was 30% (range: 0-70%) with most patients 17/20 having trilineage hematopoiesis.

B-cell aplasia (BCA) was present in 12/20 (60%) patients before receiving CD19/22 CAR infusion, 6 of whom had received prior CAR therapy. After CD19/22 CAR infusion, 8 additional patients achieved peripheral B-cell aplasia at a median of 14 days (range: 7-26) post-infusion, and all patients who had BCA at the time of CD19/CD22 CAR infusion remained in BCA in the first month post therapy. Loss of BCA was noted in 5 patients who had follow up labs and did not receive additional therapy (e.g HSCT), at a median time of 80 days post-CAR (range: 45-277).

3.3 Correlative Study Results

3.3.1 Neurocognitive Outcomes

Of the 20 subjects with ALL treated, two refused to do any testing, and one broke his dominant hand (non-research related traumatic fracture) so could only be administered the one verbal measure of the follow-up assessment; thus, 17 subjects completed the pre-and post-infusion test sessions. All 17 subjects completed all five of the Cogstate computerized tests at baseline and post-infusion; of these, only one Detection test did not meet data integrity standards, indicating that the subject did not perform the test according to test requirements, so it did not produce a score. All 17 subjects also completed the traditional paper/pencil processing speed tests except for one subject who was not administered one of these tests (Cancellation) in error; only 13 subjects were administered the verbal fluency measure because five subjects were non-English speakers.

When comparing the groups (CRS or Responder Groups), there were no significant differences in test scores found in the ‘group by time’ interaction, between the groups, or over time (p values >0.05).

When examining individual change in neurocognitive test scores from pre- to post-infusion, the majority remained stable or improved (Supplemental Table 6). In comparing the 146 paired test scores from pre- to post-infusion, 104 (71.2%) were stable, 28 (19.2%) significantly increased, and 14 (9.6%) significantly decreased over time. Six subjects exhibited a decline in only one test and four declined in two tests out of the eight computerized and traditional tests administered. Interestingly, the only tests that showed decline were those from the Cogstate

computerized battery, which suggests that these tests may be more sensitive to change and less prone to practice effects.

3.3.2 Absolute lymphocyte counts (ALC) and CAR T-cell expansion

The ALC was lower from day -2 through day +3 in patients receiving intensified LD compared to standard LD (Supplemental Figure 2A). CAR T-cell expansion, however, did not differ between those receiving intensified LD (n=6) or standard LD (n=18). Given that most patients receiving intensified LD were re-infusions, it is difficult to determine if the lack of robust responses with intensified LD were due to limitations of re-infusion. (Supplemental Figure 2B)

3.3.3 Cytokine profiling

Restricting to those with complete responses to CD19/22 CAR T-cells (n=12), patients with higher grade CRS (grade 3) had higher levels of peak serum IL2 (p=0.036), IL4 (p=0.009), IL12p70 (p=0.018), and IL6 (p=0.018; Data not shown).

3.3.4 Antigen Expression and Serial Profiling

All had CD19+/CD22+ disease with a median antigen binding capacity (ABC) of 9,346 sites/cell (range: 1,124-24,498) for CD19 and 2,719 sites/cell (range: 671-20,825) for CD22 at baseline. Stratified by response, CD19 and CD22 antigen expression did not differ in responders vs non-responders; median CD19 site density on bone marrow blasts for responders was 9,508 sites/cell vs 8,400 sites/cell in non-responders, p=0.596 and the median CD22 site density on bone marrow blasts for responders was 2,930 sites/cell vs 2,527 sites/cell, p=0.859 (Supplemental Figure 1B and 1C). A subset of patients treated on this study had serial profiling of CD19 and CD22 pre- and post-CAR infusion over time (Supplemental Figure 1F-G). Patient 5 received CD19/CD22 CAR T-cells and at relapse post-HSCT was noted to have CD19 negative disease. Additionally, patient 13 relapsed within 60 days post HSCT with CD19 positive disease, however quickly developed CD19 negative disease after receipt of blinatumomab in the post-HSCT setting. Interestingly, this patient also had CD22 antigen diminution after CD19/22 CAR, with continued further decline of CD22 antigen expression even before receiving additional CD22-directed therapy with inotuzumab

3.3.5 Immunogenicity

Of the 20 patients, HAMA was confirmed positive (≥25 ng/mL) in only one patient (who was CAR naïve) at any timepoint. HAMA was detectable at any level in 8 patients at baseline, in 8 patients at day 13, and in 5 patients at day 28 (Supplemental Figure 1G). There was no association between detectable HAMA at any point during therapy and either response to therapy or prior CAR T-cell exposure (Supplemental Figure 1G). A similar analysis was performed for detectable HAMA at baseline, day 7, day 13, and day 28, and again no statistically significant association was found (data not shown).

3.4 Associations with Response

There was no apparent impact on response by prior HSCT (p=0.07), presence of non-CNS EMD (p=0.36), or high bone marrow disease burden (p=0.65), although our sample size was small. Additionally, there were no differences in pretreatment CD19 ABC in the bone marrow between responders and non-responders (p=0.60) or those who were CAR-naïve vs CAR-pretreated (p=0.246). Favorable responses were noted in those who received DL ≥ 2, (12/16 vs 0/4, p= 0.014).

Of note, 3 patients treated at active dose levels (DL≥2) were CAR pre-treated, and 2/3 patients received increased lymphodepletion, both of whom achieved MRD-negative CR; the one patient who received standard lymphodepletion had stable disease.

3.5 Discrepant Disease Response:

Discrepancies in disease responses were defined if a patient had a complete response in one compartment, and had either a partial response or no response in another compartment, and occurred in 4 patients on the study (Supplemental Table 4).

3.5.1 Extramedullary disease response

Three patients had low level CNS1 $FC^{positive}$ classification and 2 of these patients had clearance of CNS ALL, and the other remained with residual CNS1 $FC^{positive}$. CSF analysis post-infusion was performed in 17/20 patients, with CAR T-cells detected in the CSF in 10/17 patients (58.8%) at a median of 0.7% T-cells that were CAR+ (range: 0-73%). Additionally, 8/20 (40%) patients had non-CNS extramedullary disease at the time of treatment, and three of these eight patients had an overall complete response to therapy. Sites of EMD resolution included bulky

lymphadenopathy, focal liver lesion, and right sided maxillary lesion. In the 5 patients who did not have an overall complete response to therapy, 2 had an overall partial response, 2 had SD, and 1 had progressive disease.

3.6 Reinfusions

Four patients received reinfusion with CD19/CD22 CAR T-cells; 2 on study and 2 treated on an individualized protocol, with all receiving increased lymphodepletion. Indications included antigen positive relapse (n=3), and treatment at dose level 1, deemed an ineffective dose, during the first infusion (n=1). The patient who was initially treated at dose level 1 had a suboptimal response and received an increased dose (DL2, $1x10^6$ CAR T-cells/kg) for re-infusion; all the other patients received the same CAR T-cell dose for the second infusion as they received for their initial infusion. Notably, all 4 patients had evidence for CAR T-cell expansion in the blood and/or bone marrow with reinfusion; in 2 patients it was at levels comparable to the first infusion and 2 others it was less than with the first infusion. Three of four patients who received a reinfusion had CR, including 2 patients who previously did not respond to the initial infusion. The one patient who did not respond emerged with $CD19^{dim/partial/neg}$ disease, so antigen escape likely played a role in suboptimal response to reinfusion. Despite achieving CR, response to reinfusions were short lived with relapse occurring at a median time of 59 days (range: 40-194).

3.6.1 Impact of intensified LD on CRS

We evaluated the incidence of severe CRS in patients who received intensified LD versus standard LD and found no obvious differences, which may be due in part to the fact that many patients receiving intensified LD were receiving it as part of a “reinfusion” strategy. Amongst 6 patients receiving intensified LD, 1 (16.7%) had $\geq$ grade 3 CRS; 1 (16.7%) had grade 1 CRS and 4 (66.7%) patients had no CRS at all. Across 18 patients receiving standard LD, 3 (16.7%) had $\geq$ grade 3 CRS; 6 had grade 1 or 2 CRS and 9 (50%) of patients had no CRS.

3.7 Experience in NHL

A 20-year-old HSCT and CAR naïve patient with refractory Burkitt’s lymphoma was treated at DL3. With disease localized to the mediastinum and chest wall, she tolerated therapy well without CRS or ICANs but had progressive disease following only transient CAR T-cell

expansion and no persistence. She subsequently died from progressive disease 5-months post-CAR therapy.

Supplemental Tables.

Supplemental Table 1. CAR T-cell final product characteristics and patient outcomes

ID	CD3%	CD4%	CD8%	Transduction Efficiency %	Overall Response Rate
1	95.7	67.1	23.8	74.2	SD
2	99.9	78.5	20.9	84.9	PR
3	99.2	87.9	10.8	75.4	SD
4	99.8	71.7	27.1	78.6	PR
5	99.7	74.4	23.3	77.4	MRD negative CR
6	99.6	68.4	30.5	54.2	MRD negative CR
7	99.3	60.6	37.8	76.3	SD
8	99.5	75.4	22.8	86.0	PD
9	99.4	29.6	68.9	73.0	MRD negative CR
10	99.7	71.6	26.0	77.8	MRD negative CR
12	99.5	50.1	47.4	54.9	SD
13	99.7	80.3	18.7	65.5	MRD negative CR
14	99.9	51.2	48.0	61.3	MRD negative CR
15	99.8	66.7	32.3	79.9	PR
16	99.4	61.3	36.7	60.3	MRD negative CR
17	99.5	78.2	20.8	53.6	PD
18	99.2	63.8	34.2	62.7	MRD negative CR
19	99.8	78.6	20.4	70.7	MRD negative CR
20	98.9	70.9	27.4	62.5	MRD negative CR
21	99.4	56.2	42.0	62.8	MRD negative CR
22	99.2	67.7	30.5	67.9	MRD negative CR

SD: stable disease; PR: partial response; MRD: minimal residual disease; CR: complete response; PD: progressive disease

Supplemental Table 2. Adverse events ≥ grade 3 attributed to IND following first infusion

	All Patients (n=20)	
Adverse Event	Grade 3	Grade 4
Cardiovascular		
Hypotension	3	0
Sinus tachycardia	1	0
Constitutional Symptoms		
Febrile Neutropenia	8	0
Fever	5	0
Electrolyte Derangements		
Hypokalemia	0	1
GI and Hepatic		
ALT Increase	2	0
AST Increase	1	0
Hypertriglyceridemia	4	0
Hematologic		
Anemia	7	1
Lymphopenia/ Leukopenia	5	7
Thrombocytopenia	2	5
Neutropenia	3	7
Immune-mediated		
Cytokine Release Syndrome	3	0
Neurologic		
Dysphagia	1	0
Encephalopathy	1	0
Respiratory		
Hypoxia	1	0
Total	47	21

Supplemental Table 3. Observer-reported symptoms on the neuro-symptom checklist

ID #	CRS Grade	CAR expansion	Related to CAR T	Baseline	Day 14*	Day 21-28
1^	0	No	No	-- Mild distress (<24 hrs) Mild sleepiness (<24 hrs)	-- Moderate distress (<24 hrs) Mild sleepinesss (<24 hrs)	Mild depressed mood (24-48 hrs) Mild distress (24-48 hrs) --
2	1	Yes	Yes	Moderate depressed mood (>48 hrs) Mild distress (>48 hrs) --	-- Mild distress (>48 hrs) **Moderate drowsiness/sleepiness (24-48 hrs)%**	Moderate depressed mood (<24 hrs) Severe distress (24-48 hrs) Mild drowsiness/sleepiness (<24 hrs)
3	0	Yes	Yes	Mild depressed mood (<24 hrs) Moderate distress (<24 hrs) Severe drowsiness/sleepiness (>48 hrs) Severe blurred vision (>48 hrs)	-- -- -- --	-- -- Mild drowsiness/sleepiness (<24 hrs) --
4	0	Yes	Yes	--	--	--
5	3	Yes	Yes	--	--	Mild drowsiness/sleepiness (24-48 hrs)
6	3	Yes	Yes	--	**Mild drowsiness/sleepiness (<24 hrs)%**	Mild drowsiness/sleepiness (>48 hrs)
7^	0	No	No	Mild depressed mood (<24 hrs) Mild drowsiness/sleepiness (<24 hrs) --	-- Moderate drowsiness/sleepiness --	Mild depressed mood (<24 hrs) Mild drowsiness/sleepiness (<24 hrs)
8	0	Yes	Yes	Mild distress (<24 hrs) --	Mild distress (<24 hrs) **Mild drowsiness/sleepiness (<24 hrs)%**	-- --
9	1	Yes	Yes	Mild depressed mood (<24 hrs) Mild distress (<24 hrs) Mild drowsiness/sleepiness (<24 hrs)	Mild depressed mood (24-48 hrs) **Moderate distress (24-48 hrs)** --	-- -- --
10	0	Yes	Yes	Mild depressed mood (>48 hrs) Mild distress (>48 hrs) Mild drowsiness/sleepiness (<24 hrs)	-- -- --	-- Mild distress (>48 hrs) --

12	0	Yes	Yes	Moderate depressed mood (<24 hrs) Mild drowsiness/sleepiness (<24 hrs)	Moderate depressed mood (24-48 hrs) **Severe drowsiness/sleepiness (>48 hrs)%**	Mild depressed mood (>48 hrs) Mild drowsiness/sleepiness (>48 hrs)
13	1	Yes	Yes	Moderate unresponsiveness (24-48 hrs) -- Moderate distress (<24 hrs) --	-- **Mild depressed mood (>48 hrs)** -- **Mild drowsiness/sleepiness (>48 hrs)**	-- -- -- --
14	3	Yes	Yes	Moderate depressed mood (<24 hrs) Moderate distress (<24 hrs) Mild drowsiness/sleepiness (<24 hrs)	Moderate depressed mood (<24 hrs) Moderate distress (<24 hrs) **Moderate drowsiness/sleepiness (<24 hrs)**	Mild depressed mood (<24 hrs) Moderate distress (<24 hrs) --
15	0	Yes	Yes	-- --	**Moderate depressed mood (>48 hrs)** **Moderate distress (>48 hrs)%**	-- --
16	1	Yes	Yes	-- -- -- --	**Mild disorientation (<24 hrs)** **Mild depressed mood (<24 hrs)** **Moderate drowsiness/sleepiness (<24 hrs)%** **Other: Severe headache (>48 hrs)%**	-- -- -- --
18	3	Yes	Yes	-- --	**Mild depressed mood** **Mild distress**	Mild depressed mood (<24 hrs) --
19	2	Yes	Yes	--	--	--
20	2	Yes	Yes	Mild depressed mood (<24 hrs) -- --	Mild depressed mood (<24 hrs) **Mild distress (<24 hrs)** **Moderate drowsiness/sleepiness (24-48 hrs)%**	-- -- Mild drowsiness/sleepiness (24-48 hrs)
21	0	Yes	Yes	Mild depressed mood (<24 hrs) Mild distress (<24 hrs)	-- --	-- --
22	0	Yes	Yes	-- Moderate distress (24-48 hrs) Moderate drowsiness/sleepiness (>48 hrs) --	-- -- -- --	Mild depressed mood Mild distress -- --

CRS=Cytokine Release Syndrome; CAR=Chimeric Antigen Receptor; ***Bolded symptoms post-infusion were considered related to CAR T-cell therapy if new or worsening and occurred with either CRS or CAR expansion**; ^The shaded rows indicate the two subjects who had CRS grade 0 and no CAR-T expansion; thus, their NSC symptoms were not considered related to CAR T-cell therapy; % indicates symptoms that were concurrently reported by medical team as adverse events

Supplemental Table 4. Discrepant disease responses after CD19.22.BBζ infusion

Patient ID	**Pre-Infusion Non-CNS EMD site(s)**	**Bone Marrow Disease Burden**	**Non-CNS EMD Response; Post Infusion Non-CNS EMD site(s)**	**Bone Marrow Response Post Infusion**
3	GI tract, gallbladder, sinuses, bones	M1, 0.35% blasts	*Mixed Response;* Significant improvement in the soft tissue components, and mixed response in bony lesions, some with improvement or complete resolution while others were larger or ha more avidity.	MRD negative
4	Breast, spine, and skin and soft tissue	M1, 0.004% blasts	*Partial Response;* Complete resolution extramedullary disease including the spine and soft tissue and skin, however had residual disease in the breast mass	MRD negative
8	Isolated pancreatic mass	M1, 4.96% blasts	*Progressive disease;* Enlarged pancreatic mass	MRD negative
15	Leukemia cutis, inguinal lymph nodes	M1, 0.01% blasts	*Partial Response;* partial resolution of leukemia cutis and complete resolution of inguinal lymphadenopathy	MRD negative

EMD: extramedullary disease; CNS: central nervous system; GI: gastrointestinal; M1: <5% bone marrow blasts; MRD negative: minimal residual disease negative complete response in the bone marrow compartment

Supplemental Table 5. Mean scores on neurocognitive measures pre- and post-infusion

Cogstate Tests*	N	Pre-Infusion Mean (SD)	Post-Infusion Mean (SD)	*p* value
Detection	16	0.423 (.24)	0.358 (.23)	0.245
Identification	17	-0.955 (1.7)	-0.780 (1.6)	0.315
One Card Learning	17	0.653 (.84)	0.945 (.93)	0.191
One Back	17	-0.715 (1.5)	-0.713 (1.0)	0.994
Groton Maze	17	-0.114 (1.0)	-0.134 (.70)	0.943
Traditional Tests				
Processing Speed Index^	16	92.56 (17.4)	95.38 (15.9)	0.052
Symbol Search[#]	17	9.00 (3.3)	9.47 (3.4)	0.280
Cancellation[#]	16	8.25 (3.9)	8.88 (3.4)	0.145
Verbal Fluency[#]	13	7.65 (3.0)	8.89 (2.6)	0.059

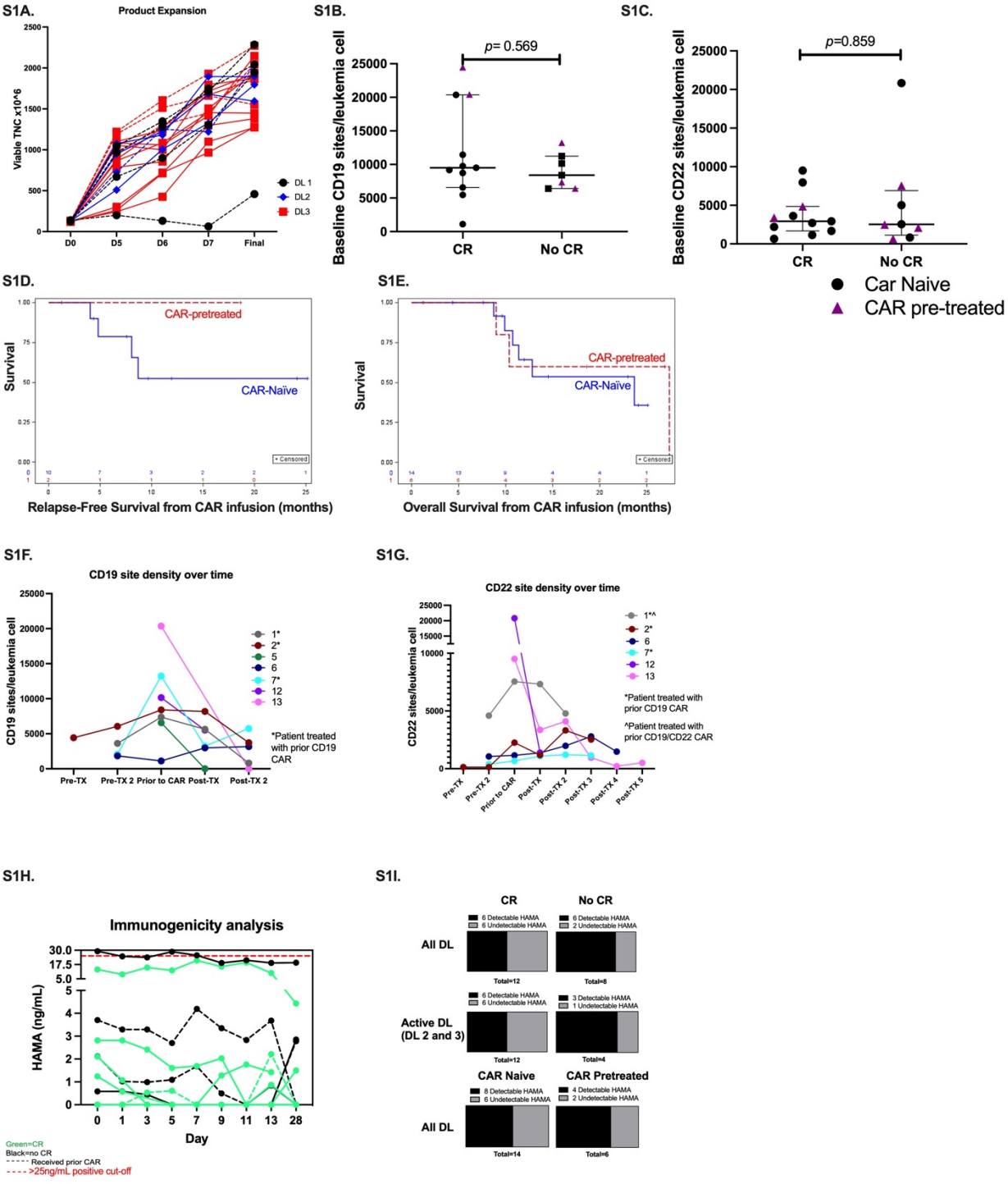
*z-scores (normative mean=0, SD=1.0); [#]scaled scores (normative mean=10, SD=3); ^standard scores (normative mean=100, SD=15); For all cognitive tests, higher scores indicate better performance.

Supplemental Table 6. Individual change in cognitive tests scores from pre- and post-infusion

Cogstate Computerized Tests						**Traditional Tests**			
Patient N=20	**Detection** Processing Speed N=16	**Identification** Attention N=17	**One Card Learning** Learning/ Memory N=17	**One Back** Working Memory N=17	**Groton Maze** Cognitive Flexibility N=17	**Processing Speed Index** (composite) N=16	**Symbol Search** Processing Speed N=17	**Cancellation** Processing Speed N=16	**Verbal Fluency** N=13
1	decrease	stable	stable	stable	increase	stable	stable	stable	stable
2[1]	-	-	-	-	-	-	-	-	-
3[1]	-	-	-	-	-	-	-	-	-
4[2]	increase	increase	stable	stable	stable	stable	stable	stable	-
5[2]	stable	decrease	increase	increase	increase	stable	stable	stable	-
6[2]	stable	stable	increase	stable	stable	stable	stable	stable	-
7	stable	stable	increase	decrease	stable	stable	stable	stable	stable
8	increase	stable	stable	stable	stable	stable	increase	stable	stable
9[3]	-	-	-	-	-	-	-	-	increase
10	stable	stable	decrease	decrease	stable	increase	stable	increase	stable
12	stable	increase	stable	stable	stable	stable	stable	stable	stable
13	decrease	stable	stable	stable	decrease	stable	stable	stable	stable
14	stable	stable	increase	increase	stable	stable	increase	stable	stable
15	stable	stable	increase	stable	decrease	-	stable	-	stable
16	decrease	stable	decrease	stable	increase	stable	stable	stable	increase
18	increase	stable	decrease	stable	increase	stable	stable	increase	stable
19[2]	stable	stable	stable	decrease	decrease	stable	stable	stable	-
20	stable	stable	increase	stable	decrease	stable	increase	stable	stable
21[2,4]	-	stable	stable	increase	increase	stable	stable	stable	-
22	stable	stable	stable	stable	stable	stable	stable	stable	increase

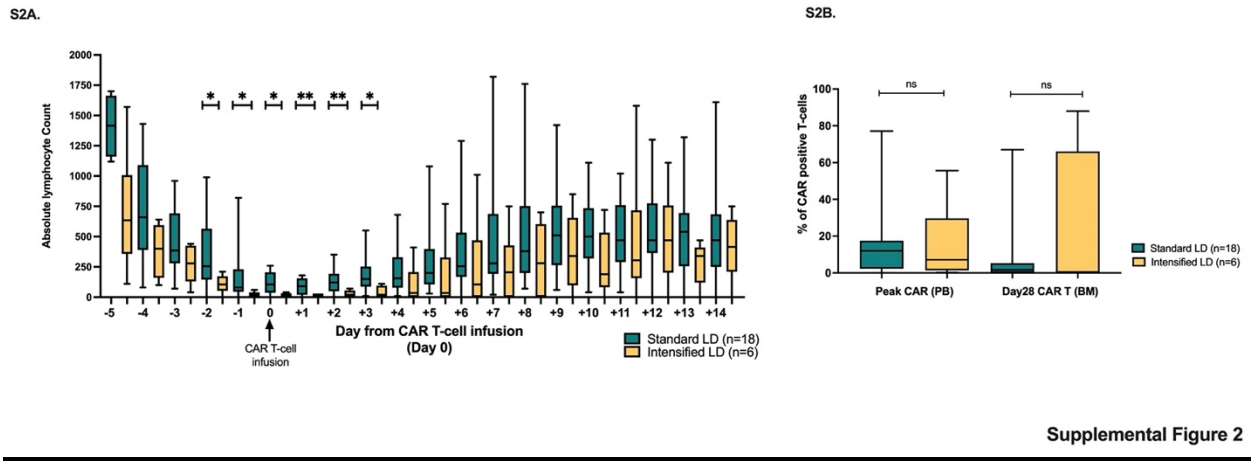
A significant change in neurocognitive test scores was defines as an increase or decrease of 3/4th of a standard deviation; The Processing Speed Index is a composite score of the Symbol Search and Cancellation subtests; Reasons for missing scores: [1] Refused all testing; [2] Verbal Fluency is a language based assessment and English was not Patients' primary language so test was not administered; [3] broken arm post-infusion; [4] detection test was not valid.

Supplemental Figures.

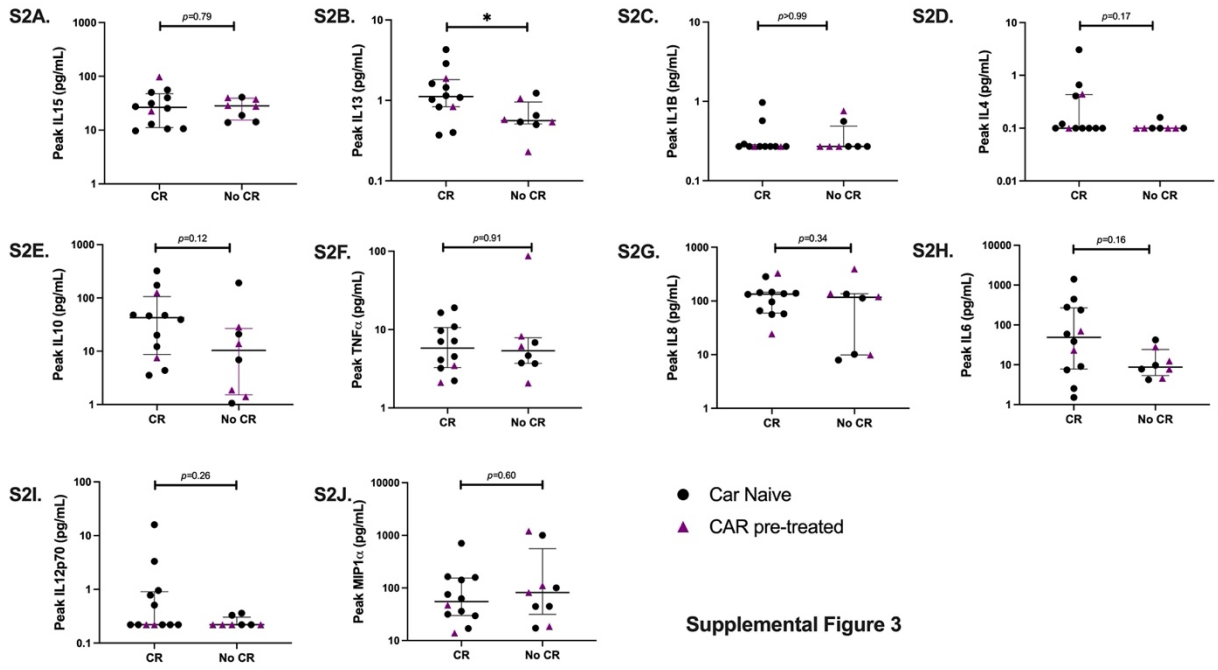


Supplemental Figure 1

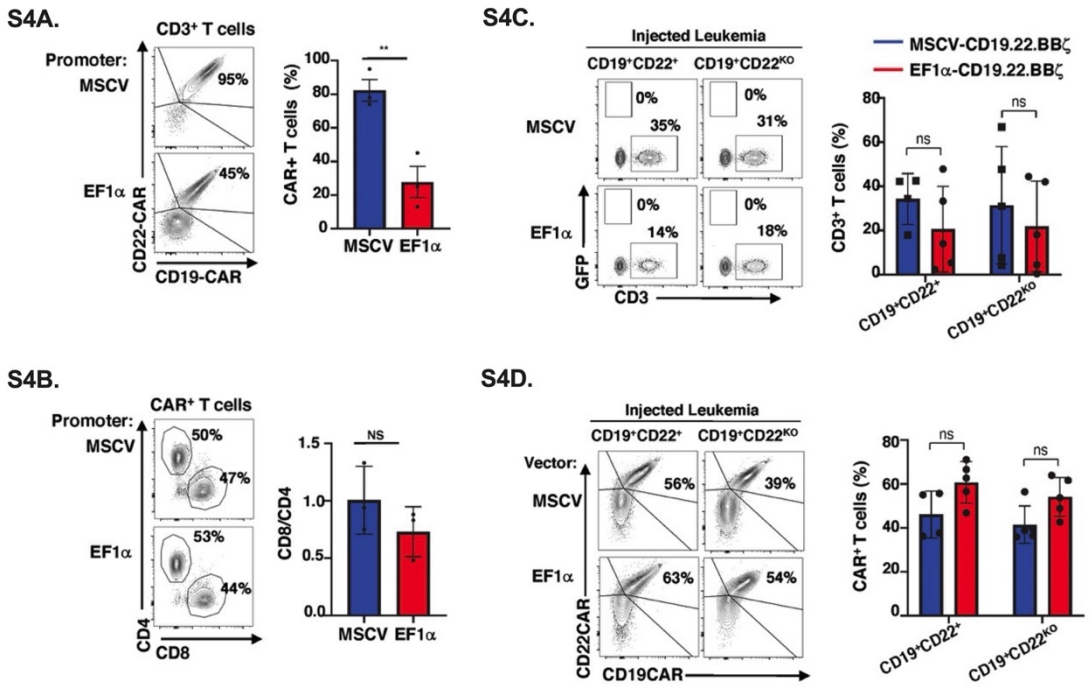
Supplemental Figure 1. Clinical correlative data. A. CAR T-cell expansion in all patients with ALL categorized by dose level. Dotted lines indicate CAR-pretreated patients. B. Baseline CD19 antigen density, stratified by responders and non-responders. C. Baseline CD22 antigen density, stratified by responders and non-responders. D. Overall survival stratified by CAR-naïve (blue) vs CAR-pretreated (red) demonstrating no significant difference (p=0.73). E. Relapse-free survival separated by CAR-naïve (blue) vs CAR-pretreated (red) demonstrating a trend towards improved RFS in patients who were CAR-pretreated, although numbers were small, and follow-up was short (p=0.44). F. CD19 antigen expression in select patients as seen over their treatment course. G. CD22 antigen expression in select patients as seen over their treatment course. H. Time course analysis of human anti-mouse antibody (HAMA) results. Patients who achieved complete response (CR) noted in green (n=12, all achieved MRD-negative CR) vs. patients who did not achieve CR in black (n=8). Six patients treated with prior chimeric antigen receptor T-cell therapy (CAR) (2 who achieved CR, 4 who did not achieve CR) denoted with dashed lines. 7 patients (5 who achieved CR, 2 who did not achieve CR) not visible as all results were 0. Accepted cutoff for positive value, >25ng/mL, depicted as red dashed line. I. Analysis of patients who achieved CR vs did not achieve CR, grouped by whether they had detectable HAMA at any timepoint and by patients in all dose levels (p=0.374) and patients at active dose levels (dose level 2 and 3, p=0.585), and those who were CAR-naïve vs CAR-pretreated and whether they had detectable HAMA at any timepoint (p=1). *p-values* calculated using Fisher's exact test.



Supplemental Figure 2. Serial absolute lymphocyte count and CAR T-cell expansion, stratified by lymphodepletion strategy. A. Serial absolute lymphocyte count from start of LD chemotherapy through day +14 post CAR T-cell infusion, stratified by standard LD (n=18, green) or intensified LD (n=6, yellow). *=p<0.05; **=p<0.01. Importantly, not all timepoints were represented by all patients, and an occasional value may not have been evaluated. In patients with a total white blood cell count where the ALC could not be calculate, the value of "0" was used. B. CAR T-cell expansion, represented as a % of T-cells that are CAR positive in the peripheral blood (PB), peak or at the D28 marrow are shown, stratified by LD strategy. There was no substantial difference seen between the two groups.

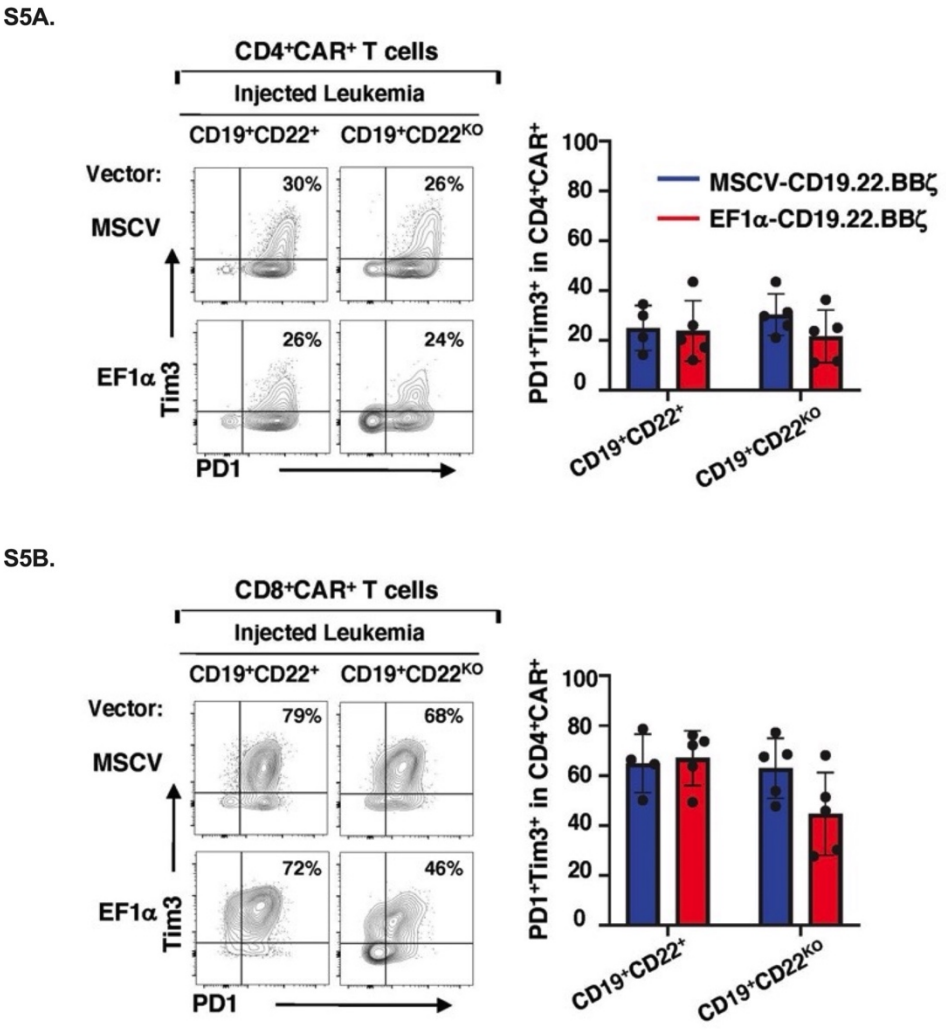


Supplemental Figure 3. Cytokine profiling with MSCV-CD19.22.BBζ CAR. Peak serum cytokine levels of IL15, IL1B, IL13, IL6, IL12p70, TNFα, IL8, IL4, IL10, MIP1α in CR vs no CR and categorized by those who are CAR naïve vs. CAR pre-treated. (ns=*p-value* >/= 0.05)



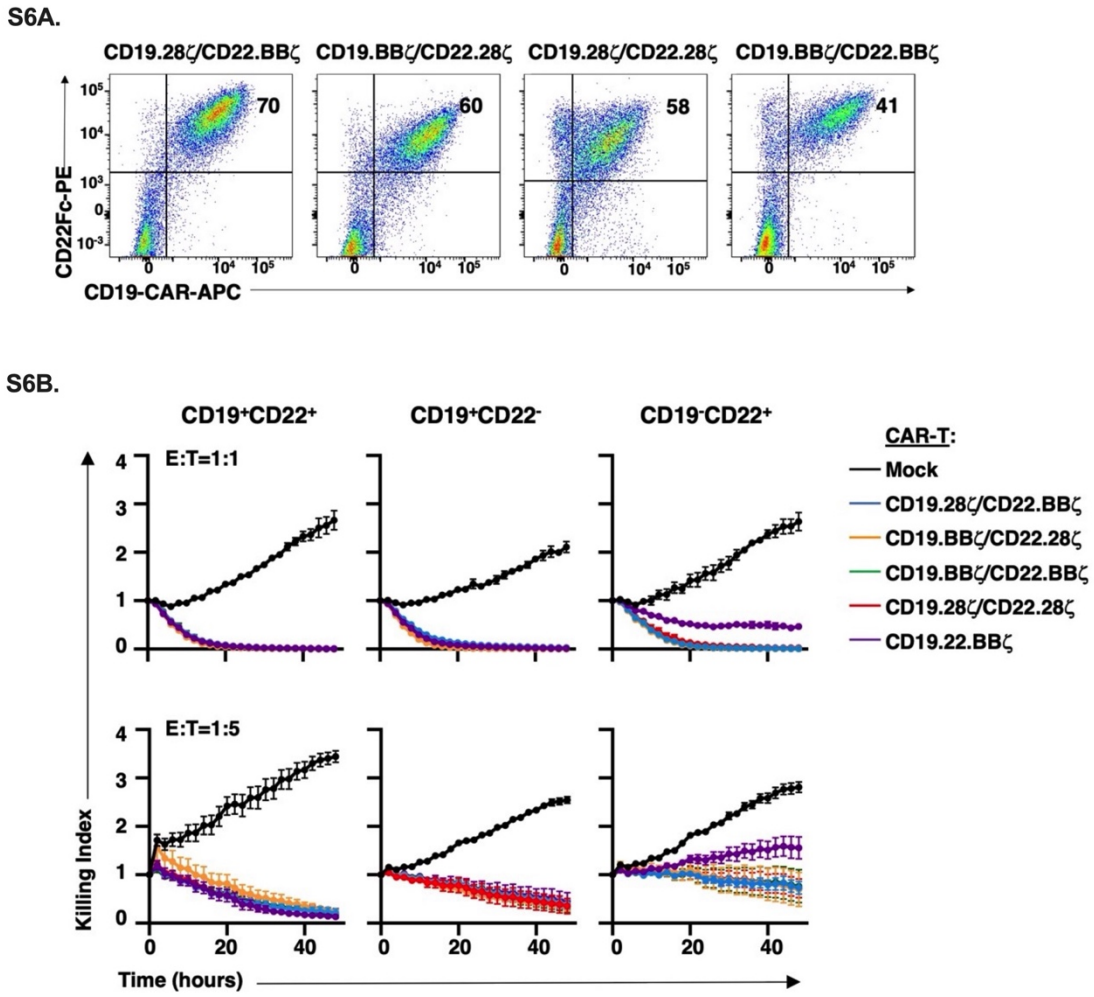
Supplemental Figure 4

Supplemental Figure 4. Persistence of T cells transduced with MSCV- and EF1α-CD19.22.BBζ CAR constructs. (A) T cells were transduced with MSCV-CD19.22.BBζ and EF1α-CD19.22.BBζ constructs and cell surface CD19/ CD22 CAR expression was evaluated at day 7 with a PE-conjugated anti-FMC63 mAb and a recombinant CD22-Fc chimeric protein. Representative dot plots are presented (left) and quantification of CAR expression in 3 individual donor T cells is shown (right). $^{**}p<0.01$ (B) Transduction of CD4 and CD8 T cell subsets was evaluated within gated CD19/CD22-CAR+ T cells for the MSCV and EF1α constructs. Representative plots (left) and quantification of CD8/CD4 ratios (n=3, right) are presented. (C) As presented in Figure 3B, luciferase transduced NALM6 cells (1e6) were injected intravenously (IV) via tail vein into NSG mice and CART were injected at day 3. The presence of GFP^+ leukemic cells and human $CD3^+$ T cells was evaluated in spleens of NSG mice at day 42 post CART treatment. Representative dot plots (left) and quantifications of the percentages of $CD3^+$ splenic T cells (n=4-5 per group, right) are shown. (D) Representative dot plots of cell surface CD19/CD22-CAR expression on splenic CD3+ T cells (left) and quantification of CAR+ T cells (right).



Supplemental Figure 5

Supplemental Figure 5. High expression of exhaustion markers in CD8$^+$ T cells transduced with MSCV- or EF1α-CD19.22.BBζ constructs in a NALM6 xenograft model. (A) As presented in Figure 3B, NALM6-bearing NSG mice (day 0) were injected with the indicated MSCV- or EF1α-CD19.22.BBζ CART and at day 42, expression of the PD1 and Tim3 exhaustion markers was evaluated in gated CD4$^+$CAR$^+$ T cells. The percentages of PD1$^+$Tim3$^+$ cells within the CD4$^+$CAR$^+$ subset are presented for NSG mice harboring CD19$^+$CD22$^+$ and CD19$^+$CD22$^-$ NALM6 as indicated (left). Quantifications of PD1$^+$Tim3$^+$ cells are shown (n=5 mice per group, right). (B) The percentages of PD1$^+$Tim3$^+$ cells were evaluated within the CD8$^+$CAR$^+$ subset. Representative dot plots (left) and quantifications (right) are shown ($p>0.05$).



Supplemental Figure 6

Supplemental Figure 6. Expression and function of CD19-CD22 bicistronic CAR constructs. (A) Cell surface CD19- and CD22-CAR expression from the different bicistronic constructs described in Figure 4A was evaluated on transduced primary T cells at day 7. Cells were stained with an APC-conjugated anti-FMC63 mAb and a recombinant CD22-Fc chimeric protein followed by a PE-conjugated goat-anti-human IgG. Representative dot plots with the percentages of CAR-expressing cells are shown. Data are representative of 1 of 6 individual T cell donors. (B) Cytotoxic potential of bicistronic CART was evaluated in comparison with the bivalent CART in an Incucyte assay. Killing indices of GFP^+ NALM6 leukemic cells co-cultured with the indicated CAR-T at 1:1 and 1:5 effector/target (E:T) ratios are shown.

6.0 References

1. Cherian S, Miller V, McCullouch V, Dougherty K, Fromm JR, Wood BL. A novel flow cytometric assay for detection of residual disease in patients with B-lymphoblastic leukemia/lymphoma post anti-CD19 therapy. *Cytometry B Clin Cytom*. Jan 2018;94(1):112-120. doi:10.1002/cyto.b.21482
2. Cherian S, Stetler-Stevenson M. Flow Cytometric Monitoring for Residual Disease in B Lymphoblastic Leukemia Post T Cell Engaging Targeted Therapies. *Curr Protoc Cytom*. Oct 2018;86(1):e44. doi:10.1002/cpcy.44
3. Shalabi H, Wolters PL, Martin S, et al. Systematic Evaluation of Neurotoxicity in Children and Young Adults Undergoing CD22 Chimeric Antigen Receptor T-Cell Therapy. *J Immunother*. Sep 2018;41(7):350-358. doi:10.1097/CJI.0000000000000241
4. Caine J, Mehta M, Deshmukh S, et al. CogState computerized memory tests in patients with br ain metastases: Secondary endpoint results of RTOG 0933. *Radiother Oncol*. 2014;111:S5-S5.
5. Heitzer AM, Ashford JM, Harel BT, et al. Computerized assessment of cognitive impairment among children undergoing radiation therapy for medulloblastoma. *J Neurooncol*. Jan 2019;141(2):403-411. doi:10.1007/s11060-018-03046-2
6. Sands SA, Harel BT, Savone M, et al. Feasibility of baseline neurocognitive assessment using Cogstate during the first month of therapy for childhood leukemia. *Support Care Cancer*. Feb 2017;25(2):449-457. doi:10.1007/s00520-016-3422-9
7. Bangirana P, Sikorskii A, Giordani B, Nakasujja N, Boivin MJ. Validation of the CogState battery for rapid neurocognitive assessment in Ugandan school age children. *Child Adol Psych Men*. Aug 14 2015;9doi:ARTN 38 10.1186/s13034-015-0063-6
8. Falleti MG, Maruff P, Collie A, Darby DG. Practice effects associated with the repeated assessment of cognitive function using the CogState battery at 10-minute, one week and one month test-retest intervals. *J Clin Exp Neuropsychol*. Oct 2006;28(7):1095-112. doi:10.1080/13803390500205718
9. Lee DW, Kochenderfer JN, Stetler-Stevenson M, et al. T cells expressing CD19 chimeric antigen receptors for acute lymphoblastic leukaemia in children and young adults: a phase 1 dose-escalation trial. *Lancet*. Feb 07 2015;385(9967):517-28. doi:10.1016/S0140-6736(14)61403-3
10. Jena B, Maiti S, Huls H, et al. Chimeric Antigen Receptor (CAR)-Specific Monoclonal Antibody to Detect CD19-Specific T Cells in Clinical Trials. *Plos One*. Mar 1 2013;8(3)doi:ARTN e57838 10.1371/journal.pone.0057838