Title: Identification of Predictive Biomarkers for Cytokine Release Syndrome after Chimeric Antigen Receptor T cell Therapy for Acute Lymphoblastic Leukemia

Supplemental Methods

Trial Design

We collected laboratory and clinical data from patients treated on 3 clinical trials designed to assess the safety and feasibility of CTL019 T cell therapy in relapsed/refractory CD19+ malignancies. These trials were conducted at the Children's Hospital of Philadelphia (CHOP) (Clinicaltrials.gov: NCT01626495) and the Hospital of the University of Pennsylvania (PENN) (NCT02030847 and NCT01029366). Patients treated at CHOP constituted the pediatric cohort and patients treated at PENN constituted the adult cohort. Written informed consent was obtained from all subjects or their legal guardian according to the Declaration of Helsinki and protocols were approved by institutional review boards at CHOP and PENN. Eligibility criteria for the three trials are included on the Clinicaltrials.gov website and are also previously published.(1) Of note, patients with active graft versus host disease, active CNS involvement with leukemia (CNS3), an uncontrolled infection, active hepatitis B or C, or HIV were excluded. Patients who had a prior allogeneic stem cell transplant were eligible provided it had been at least 6 months since the transplant and the patient did not require immunosuppression at the time of enrollment. Study procedures including details of leukapheresis and types of lymphodepleting chemotherapy are previously published (Maude, NEJM).(1) Patients were infused with 1-10 x 10^7 T cells/kg (5-50 x 10^8 T cells for patients over 50 kg) over 1-3 days as previously described.(1) Details on response to therapy for the first 30 subjects included in this report are also previously published.(1)

General Laboratory Statement

CTL019 T cells were produced in the Clinical Cell and Vaccine Production Facility at the University of Pennsylvania (UPENN) under principles of current Good Manufacturing Practices. Research sample processing, including cytokine, chemokine, and soluble receptor (hereafter called cytokine) was performed in the Translational and Correlative Studies Laboratory at UPENN. Both laboratories operate with established Standard Operating Procedures (SOPs) and/or protocols for sample receipt, processing, freezing, and analysis. Clinical laboratory studies, including ferritin, C-reactive protein (CRP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood urea nitrogen (BUN), creatinine (Cr), complete blood count (CBC) with differential, prothrombin time (PT), partial thromboplastin time (PTT), and fibrinogen were performed in the Clinical Laboratory Improvement Amendments (CLIA)-certified and College of American Pathologist (CAP)-accredited clinical laboratories at CHOP and PENN.

Production of CTL019 T cells(1)

Peripheral blood mononuclear cells (PBMCs) were collected by leukapheresis T cells, and were enriched washed, and expanded by addition of anti-CD3/CD28–coated paramagnetic beads for activation of T cells.(2) The lentiviral vector containing the previously described anti-CD19-BB-ζ transgene(2) (produced by the Vector Core at the Children's Hospital of Philadelphia) was added at the time of cell activation and was washed out on day 3 after culture initiation.(3) Cells were expanded on a rocking platform device (WAVE Bioreactor System) for 8 to 12 days. On the final day of culture, the beads were removed by passage over a magnetic field and the CTL019 cells were harvested and cryopreserved in infusible cryomedium. Final product release criteria in the IND included the following: cell viability ≥ 70%, CD3+ cells ≥ 80%, residual paramagnetic anti-CD3/CD28-coated paramagnetic beads ≤ 100 per $3x10^6$ cells, Endotoxin ≤ 3.5 EU/mL, Mycoplasma negative, Bacterial and fungal cultures negative, residual bovine serum albumin ≤ 1 μg/mL, VSV-G DNA as a surrogate marker for replication competent lentivirus ≤ 50

copies per μg DNA, transduction efficiency by flow cytometry ≥ 2%, transduction efficiency by vector DNA sequence 0.02 to 4 copies per cell.

Clinical Laboratory tests

Clinical laboratory tests included serial monitoring of ferritin, C-reactive protein (CRP), alanine aminotransferase (ALT), aspartate aminotransferase (AST), blood urea nitrogen (BUN), creatinine (Cr), complete blood count (CBC) with differential, lactate dehydrogenase (LDH), prothrombin time (PT), partial thromboplastin time (PTT), and fibrinogen. These tests were performed in CLIA and CAP certified clinical laboratories at CHOP and PENN.

Sample processing for non-clinical laboratory studies

Peripheral blood and bone marrow samples were collected in lavender top (K2EDTA) or red top (no additive) vacutainer tubes (Becton Dickinson). Lavender top tubes were delivered to the laboratory within 2 hours of the sample draw, or shipped overnight at room temperature in insulated containers as described.(4) Samples were processed within 16 hours of drawing according to the established SOP. Peripheral blood (PBMC) and bone marrow (BMMC) mononuclear cells were purified, processed, and stored at -140°C as described.(5) Red top tubes were processed within 2 hours of the draw, including coagulation time; serum was isolated by centrifugation, aliquoted in single use 130 μL aliquots and stored at -80°C.

Methods for Luminex 14 plex and 30 plex assays

Human cytokine magnetic 30-plex panel catalog number LHC6003M was purchased from Life Technologies (Carlsbad, CA). The following analytes are in the panel (IL1RA, FGF-Basic, MCP1, G-CSF, IFNγ, IL12, IL13, IL7, GM-CSF, TNFα, IL1β, IL2, IL4, IL5, IL6, IFNα, IL15, IL10, MIP1α, IL17, IL8, EGF, HGF,

VEGF, MIG, RANTES, Eotaxin, MIP1β, IP10, IL2R). Human soluble cytokine receptor magnetic bead 14-plex panel catalog number HSCRMA32KPX14 was purchased from EMD Millipore (Darmstadt, Germany). The following analytes are in the panel (sCD30, sEGFR, sgp130, sIL1RI, sIL1RII, sIL2Rα, sIL4R, sIL6R, sRAGE, sTNFRI, sTNFRII, sVEGFR1, sVEGFR2, sVEGFR3). Serum samples cryopreserved at -80°C from day -4 to day 28 were thawed and analyzed according to the manufacturers' protocols. Assay plates were measured using a FlexMAP 3D instrument (Luminex, Austin, TX), and data acquisition and analysis were done using xPONENT software (Luminex). Data quality was examined based on the following criteria. The standard curve for each analyte has a R^2 value > 0.95 with or without minor fitting using xPONENT software. Non-extrapolated data from low bead counts that also have CV (coefficient of variation) >20% were flagged. For the 14-plex kit, > 90% of the results for the two control samples included in the kit was required to be within the expected ranges provided by the manufacturer. No further tests were done on samples with results out of range low (<OOR) and substituted with half of the lower end of the standard curve for data analysis. Samples with results that were out of range high (>OOR) or greater than two times the standard curve maximum value (SC max) were re-tested at higher dilutions. Results that passed the above quality controls or retests were used in translational correlative studies. sIL2Rα was included in both Luminex arrays. Accordingly, analyses considered sIL2Rα the 14-plex platform and excluded the sIL2Rα from the 30-plex platform, because the 14-plex platform has a more accurate dynamic range for many of the high sIL2Rα values seen with CRS. In order to ensure tocilizumab did not interfere with quantification of cytokine levels, we added tocilizumab *ex vivo* to sera collected from ALL patients with CRS and found that tocilizumab effects detection of sIL6R, but these effects were very minimal using the Luminex kits studied herein (Chen, *et al*. in preparation, data not shown).

Collection Time points.

Cytokines, CBC, ALT, AST, BUN, Cr, ferritin, CRP, and LDH were collected on the following days based on study (Supplemental Table 3). Protocols also allowed for additional collection of samples when patients were ill. Of note, ferritin and CRP were not included when the study first opened. Baseline ferritin was not obtained in 2 children. Baseline CRP was not obtained in 3 children and 2 adults. Ferritin was collected on CHP959 with the first subject enrolled, but not until day 11 and this subject only had 3 time-points collected in the first 35 days. After learning the importance of ferritin, the second, third, and fourth CHP959 subjects had ferritin checked more frequently 8, 24, and 9 times, respectively over the first 35 days. By subject 5, ferritin was collected routinely on the same schedule as the other clinical labs and cytokines. CRP was not sent on the first two subjects at any time-point. By subject 3, CRP was tested on the same schedule as the other clinical labs. Supplemental Table 2 summarizes collection time-points for the different trials included herein. Details on coagulation studies are provided below in section: additional details on coagulopathy.

Baseline Disease Burden

The children on this study had a bone marrow aspirate and biopsy, as well as minimal residual disease testing sent immediately prior to infusion. Aspirates and biopsies were quantified for amount of disease burden by hematopathologists at CHOP. MRD was performed in the CLIA and CAP approved Children's Oncology Group Western Flow Cytometry Reference Laboratory in Seattle Washington as previously described.(6, 7) One child did not have a bone marrow biopsy performed and 4 children had bone marrow biopsies that disease could not be accurately quantified. All children had bone marrow aspirates for morphology and 5 children did not have central MRD. Disease burden was considered as both a continuous and dichotomous variable. Baseline disease burden was defined by taking the highest value (percent blasts) either by MRD or morphology on aspirate or biopsy.

Quantitative (q) PCR analysis

Genomic DNA was isolated directly from whole blood and qPCR analysis performed using ABI Taqman technology and a validated assay to detect the integrated CD19 CAR transgene sequence as described using 200 ng genomic DNA per time-point for peripheral blood and marrow samples.(8) To determine copy number per unit DNA, an 8- point standard curve was generated consisting of 5 to 10^6 copies CTL019 lentivirus plasmid spiked into 100 ng non-transduced control genomic DNA. The number of copies of plasmid present in the standard curve was verified using digital qPCR with the same CD19 CAR primer/probe set, and performed on a QuantStudio™ 3D digital PCR instrument (Life Technologies). Each data-point (sample, standard curve) was evaluated in triplicate with a positive Ct value in 3/3 replicates with % CV less than 0.95% for all quantifiable values. To control for the quality of interrogated DNA, a parallel amplification reaction was performed using 20 ng genomic DNA, and a primer/probe combination specific for a non-transcribed genomic sequence upstream of the CDKN1A (p21) gene as described.(5) These amplification reactions generated a correction factor to adjust for calculated versus actual DNA input. Copies of transgene per microgram DNA were calculated according to the formula: Copies/microgram genomic DNA = copies calculated from CTL019 standard curve x correction factor/amount DNA evaluated (ng) x 1000 ng. Supplemental Table 2 depicts time-points that qPCR was obtained.

Supplemental Statistical Methods

Per Protocol Analysis

To reduce potential confounding by more frequent measurement in more severe CRS cases, the association analysis of 1-month peak cytokine levels with CRS severity was repeated using data from a common measurement schedule applied across individuals. For each subject, a cytokine measurement was included in the analysis if it were taken during the target measurement schedule provided in study protocol. If there

were more than one visit per scheduled day within the allowable window, the visit closest to the scheduled day was used; if two within-window visits were equally close to the scheduled day, the two values were averaged. Measurements were assigned to visit windows in a sequential fashion. Once an observation was been accepted as the measurement for a given period, it is not eligible for use as the measurement observed in a subsequent period. Due to early deaths in 3 adults and some minor changes to the schedule over time, particularly for the pediatric protocol, this sensitivity analysis has some limitations. However, the target assessment windows shown in Supplemental Table 10 represented the overwhelming majority of individuals. Deaths occurred in the first month post infusion for three adults, after the development of severe CRS (Figure 4). The same biomarkers were tested in the pediatric and adult cohort except for baseline disease burden and which was not routinely collected in adults.

Methods to develop CRS Prediction Models

We sought to develop a predication model of severe CRS by selecting candidate predictors from baseline covariates, as well as clinical and laboratory factors measured within the first 3 days post-infusion. Cytokines were analyzed from the first 3 days after infusion and prior to severe CRS. Models were kept small due to the limited number of cases (14 overall and 11 in the pediatric cohort). First, we developed a logistic regression model of the probability of developing using forward selection with the Akaike information criterion (AIC) to select the best 2- and 3-variable models. Initial candidate variables included those factors for which the 3-day peak was missing in no more than 2/14 cases and 10% overall: ALT, AST, BUN, CR, Ferritin, LDH, the 43 cytokines markers, qPCR, as well CRS-defining symptoms in the first 2 days post-infusion (yes/no) and age of infusion. This model was fit in the combined cohort and separate modeling was considered for the pediatric cohort, in which baseline disease burden was an additional candidate variable. Similar models were fit using the 3-day peak for the fold change from baseline, instead of the absolute peak level. In the above models none of qPCR, ferritin, AST or ALT were ever selected. Thus

they were removed from the candidate variables and, since these variables had some missingness, the predictive selection for forward and tree models were refit with the reduced candidate variable list and a larger, more complete dataset.

We considered a few alternative modeling approaches with similar candidate covariates, including logistic regression models found using forward selection where the cytokine predictors were replaced with their principle components, as well as a classification tree models, which were fit with the `tree` package in R using the default deviance split method[9]. We did a sensitivity analysis for the selected fold-change models that added ½ to all cytokine values, so that fold change was stabilized. The models with peak fold change cytokine covariates did not appreciably increase accuracy over those using the peak values alone, but involved a more difficult to obtain covariate (using baseline+post-infusion measurements), and so were given limited consideration. The principle component approaches did appreciably worse that simple logistic regression models and were dropped from consideration. Finally, we fit models using the penalized regression method Lasso, fit with the `glmnet` package in R, that permits a larger dimension model, even with the small number of cases, and simultaneously considered all peak and fold change cytokine covariates as well as clinical predictors.[10,11] This model was more of theoretical interest, since not practical given the large number of cytokine predictors not routinely available.

For each logistic regression prediction model, a cut off for the predicted probability for being a case was chosen so that those with predicted probabilities greater than the cutoff were classified by the model as being a case. The cutoff was chosen to find the point of the receiver operating characteristic (ROC) curve based on the predicted probability that was closest to the point (0,1) (perfect specificity and sensitivity). Accuracy of the model is described using several measures: the area under the ROC curve (AUC), sensitivity and specificity. Case classification (predicted severe CRS: yes/no) was directly provided by the

classification tree models. Accuracy for the tree models was summarized also by the sensitivity and specificity. Exact 95% Clopper-Pearson binomial confidence intervals are provided for the sensitivity and specificity of the top models. The Delong 95% confidence intervals for the AUC, calculated using the pROC package in R, are also provided.

Application of the CRS Prediction Model: An Example

In this section we show how to apply each of the prediction models in Table 3 (and Figure 2) of the main manuscript to obtain a prediction for severe CRS status for Subject 136 from the pediatric protocol. This subject had severe (grade 4) CRS.

Model A: Top regression model for combined cohort

Subject 136 had a day 1-3 peak (Pk3) of 272,292.6, 59.9067, and 57.0324 for sgp130, IFNγ, and IL1RA, respectively. After taking the common logarithm (base 10) and using the respective estimates in the 1st row of Table 3, x = (13.8712 * 5.4350) + (2.4626 * 1.7775) + (-1.6559 * 1.7561) -75.3502 = 1.5095, exp(x) = 4.5245, and exp(x) / (1 + exp(x)) = the predicted probability =0.8190. Since this is greater than the cutoff of 0.3623 in Table 3, the model (correctly) predicts severe CRS.

Model B: Top tree model for combined cohort

sgp130Pk3 = 272,292.6 is greater than cutoff of 218,179, and MCP1Pk3 = 1,053.2294 is less than the cutoff of 4,636.52, and EotaxinPk3 = 18.0830 is less than the cutoff of 29.0902, so the model (correctly) predicts severe CRS.

Model C: top regression model for pediatric cohort

IFNγPk3 as above; IL13Pk3 was 1.556 and MIP1αPk3 was 25.886. After taking the common logarithm and using the estimates in the 3rd row of Table 3, x = (8.483 * 1.777) + (-5.599 * 0.192) + (-16.343 * 1.413) +15.742 = 6.652. exp(6.652) = 774.38, and exp(x) / (1 + exp(x)) = the predicted probability =0.999. Since this is greater than the cutoff of 0.3288, the model (correctly) predicts severe CRS.

Model D: top tree model for pediatric cohort

Subject 136 had IL10Pk3 = 37.4598, which was greater than the cutoff of 11.747; and a disease burden of 98.5 is greater than a cutoff of 51%, so model (correctly) predicts severe CRS.

Model E: Classifier using factors from top pediatric logistic regression model

IFNγPk3=59.9067 was greater than the cutoff of 27.6732 and MIP1αPk3= 25.8859 was less than the cutoff of 30.1591, so model (correctly) predicts severe CRS.

Supplemental Results and Discussion

Additional Details on Clinical Description of CRS

Start of CRS was defined as the day of first fever for patients who had fever relative to infusion of CTL019. Per protocol, for the pediatric cohort a fever was defined as 38.6°C (101.5°F) or higher. Per protocol, for the adult cohort a fever was defined as 38.0°C (100.5°F) or higher. For consistency, for the purposes of the analysis start of CRS was defined as the day with the first fever >=38.0°C for all subjects. Of note, the first day of CRS would not change for any pediatric subject using a definition of >=38.0°C or >=38.5°C. Stop of CRS was defined as the first day of no fever >38.0°C.

Regardless of grade, febrile patients with CRS were hospitalized for at least 24 hours with initial fever. All patients with fever were evaluated for co-morbid infections. For patients who developed fever, start of CRS was defined as the day with the first fever >= 38.0°C (100.5°F) relative to infusion of CTL019. Stop of CRS was defined as 24 hours without fever or vasoactive medications, indicating recovery from shock. Four children (and zero adults) had CRS that manifested with flu-like symptoms, but did not have documented fever. For these children, start and stop of CRS were defined based on the first day with flu-like symptoms and the first 24-hour period without symptoms, respectively.

Median time to first fever was earlier in patients with grade 4-5 CRS as compared to grade 1-3 CRS (Table 1). This difference was not statistically significant in the total cohort or the adults, but there was a significant difference in the children (median 4 days vs 1 day; $p < 0.05$). While the median time to first fever was statistically significant in the children, time to first fever is not a clinically meaningful marker of CRS severity, as there were too many outliers. There was no difference in total number of days febrile based on CRS grade in the children or adults. Patients with severe CRS were treated with the IL6R inhibitor tocilizumab, which typically caused rapid defervescence.

Nine patients (6 in the pediatric and 3 in the adult cohort) required mechanical ventilation, typically for pulmonary edema, secondary to capillary leak with volume resuscitation. Twenty patients required vasoactive medications for either distributive (19/20) or cardiogenic shock (1/20). Fourteen patients required high dose vasoactives as defined in the Supplemental Table 1B. For those needing vasoactives, the median time to starting vasoactive medications was not different for grade 3 (4 days) and grade 4-5 CRS (5 days). Supplemental Table 4 provides more details regarding the time to development of severe CRS for the subjects who developed grade 4-5 CRS. Despite a clinical picture that could also be consistent with sepsis, only 5 patients developed a documented co-morbid infection (Supplemental Table 6). Clinical

factors related to CRS are summarized in Table 1. Three adults died in the first 30 days after CTL019 treatment. Two of these adults died from grade 5 CRS events and one adult died from grade 5 intracranial hemorrhage with grade 4 CRS. Additional details regarding the 3 deaths are described in Frey, *et al*. submitted. Additional details on the intensive care management and course for the children with severe CRS are described in Fitzgerald, *et al*. submitted.

Clinically, we monitored for the development of organomegaly in the pediatric cohort treated with CTL019 as patients with HLH often develop organomegaly, including hepato- and splenomegaly.(9) 5 of 10 evaluable children with grade 4 CRS developed splenomegaly and 7/10 developed hepatomegaly by clinical exam. In contrast, no child with grade 0-3 CRS developed hepato- or splenomegaly (Table 1). One patient in the pediatric cohort was morbidly obese and physical exam was unreliable. Organomegaly was not identified by serial exams in adults.

Encephalopathy

Similar to other T-cell engaging therapies, patients treated with CAR T cells can develop neurologic manifestations that range from mild delirium to global encephalopathy with aphasia, confusion, seizure, and tremor.(1, 10-13) Encephalopathy can be caused by febrile delirium but also develops in a subset of patients after resolution of fever and CRS. The majority of children with high grade CRS developed some degree of encephalopathy (Table 1), but some patients with mild-moderate CRS also developed neurologic changes. All patients had a complete neurologic recovery with no long-term neurologic sequelae. Thirteen patients, all children, developed encephalopathy. Encephalopathy occurred in 8 children with grade 4 CRS and 5 children with either grade 2 or 3 CRS. Onset of encephalopathy relative to start of infusion occurred on a median of day 6 (range days 2-14) with no significant difference comparing children with CRS 0-3 (n = 5; median 6 days; range 4-13 days) vs. CRS 4-5 (n=8; median 7 days (range 2-14 days). End of

encephalopathy relative to start of infusion occurred on a median of day 12 (range 8-21) with no significant difference comparing children with CRS 0-3 (median 12 days; range 8-18 days) vs CRS 4-5 (median 14 days; range 9-21 days). The median total duration of encephalopathy was 6 days (range 4-18 days) with no statistical difference seen in CRS 0-3 (median 5 days; range 4-13 days) vs CRS 4-5 (median 8 days; range 4-18).

Details on comorbid infections

A total of 6 patients had co-morbid new infections that were identified in the first month after infusion with CTL019. In the pediatric cohort these did not appear to affect severity of CRS. In the adult cohort, infections in two patients likely contributed to mortality and were consistent with clinical sepsis. One adult had influenza and a second had pseudomonal bacteremia and pneumonia. The other infections included a pseudomonas colonization from a tracheal culture that did not grow in blood, a strep mitis bacteremia that had completely resolved prior to the development of severe CRS, an enterococcal bacteremia in a patient who had grade 2 CRS with fever as the only symptom, and a C. diff stool isolate that caused mild diarrhea. Additional details on the adult patients are provided in Frey, *et al.* in submission. Supplemental Table 6 provides additional details on the co-morbid infections.

Additional Details on Utility of CRP and ferritin in CRS prediction

Median CRP for the first 3 days after infusion was 4.8mg/dl (range: <0.50 - 36.3) in grade 0-3 and 13.6mg/dl (2.9- 33.2) in grade 4-5 CRS (p =0.022). There were patients with grade 0-3 CRS who had marked early elevation in CRP, including 10/34 patients with grade 0-3 CRS whose peak CRP within the first 3 days of infusion was >10mg/dl, while 6/11 patients with grade 4-5 CRS had a peak CRP >10mg/dl (PPV = 0.375). Early measurement of CRP was not routinely performed when the study first opened and a CRP between days 1-3 was not available for 3 patients with grade 4-5 CRS and 3 patients with grade 0-3

CRS. Early ferritin was available for 48 subjects. Grade 4-5 CRS was weakly associated with an elevated early ferritin ($p = 0.07$), with observed median (range) of 2300 (280, 15870) and 10660 (366, 53200) for patients with grade 0-3 versus grade 4-5 CRS, respectively; but we did not find early assessment of ferritin in the first three days following CTL019 infusion was useful in predicting severity of CRS (AUC=0.67) (Supplemental Figure 1).

Details on normal donor cohort

Serum was collected from 10 normal healthy volunteers under an IRB approved protocol through UPENN (Local IRB study number 705906). 5 were males and 5 were females. Age median (range): 40 (25-65).

Additional details on coagulopathy

PT, PTT, and fibrinogen were only obtained when clinically indicated. Accordingly, there is a definite selection bias toward the children and adults who became ill (see Supplemental Table 2 for additional details on sampling frequency). Unlike patients who develop DIC or liver failure, patients with HLH often develop a coagulopathy that is striking for the marked degree of hypofibrinogenemia relative to modest elevations in PT/INR or PTT. This pattern was also seen in the pediatric cohort with grade 4 CRS (Table 2 in main text).[12-14] Adults also developed coagulopathy; however, there were no differences when comparing CRS 4-5 vs CRS 0-3 (Table 2 in main text).

Additional analysis of grade 3 CRS

Our primary analyses divided patients into severe vs. not severe based on either requiring high-dose vasoactive medications and/or mechanical ventilation. Of note, all subjects who required mechanical ventilation also required high dose vasoactives, but there were five subjects (all children) who required

high dose vasoactives, but did not require mechanical ventilation. Patients with low grade (0-2) CRS only developed mild illness; however, using our CRS grading scale patients with grade 3 CRS had a more heterogeneous presentation. We performed additional analyses sub-dividing patients with grade 3 CRS into two groups based on the need for any vasoactive medications or significant oxygen requirement (>=40% FI02) (3a and 3b) and re-split our cohort into severe and not severe defined as CRS 0-3a vs CRS 3b-5. Grade 3a: Patients who did not require ANY vasoactive medications for CRS and who never required FIO_2 >= 40%. Grade 3b: Patients who either required vasoactive medications or >=40% FIO_2. We chose these cutoffs, because they are objective and also allow us to compare our CRS grading with other CRS grading scales.

Sixteen subjects in our main cohort had grade 3 CRS, including 8 in the pediatric cohort and 8 in the adult cohort. 3 of the 8 in the pediatric cohort and 5 of the 8 in the adult cohort met the criteria for CRS3b. Supplemental Table 18 presents all of the subjects in the main cohort and how they would be reclassified using the alternate CRS severity classification. We performed logistic regression and decision tree modeling to develop new models with this alternate categorization also tested the accuracy of our "original" models by recalibrating (refitting) the selected models using the alternate categorization of CRS outcome.

The new models developed using the alternative CRS severity classification (0-3a vs 3b-5) are depicted in Supplemental Table 19 and as Supplemental Figure 5. As expected, these models do include different cytokines than the models developed using the original CRS severity classification (0-3 vs 4-5). These additional models could be used in scenarios where the goal is to identify any patient likely to need ICU level monitoring, but who may or may not necessarily develop life-threatening CRS.

The original models that were recalibrated to the revised CRS severity classification remained quite accurate with high sensitivity and specificity when analyzing the pediatric cohort even with the alternate definition of severe vs not severe CRS (Supplemental Table 19). In contrast, the sensitivity remained high but the specificity dropped in the combined cohort when using our recalibrated models with the alternate definition of severe vs not severe (Supplemental Table 19). We provided 95% confidence intervals for the sensitivity and specificities for the model in the Discovery Cohort, which can be compared to the analogous values for our selected models in Table 3 of the main manuscript. The area under the ROC curve (95% confidence interval) is 0.93 (0.87, 0.99) for model A1 (Figure 5A) and 0.94 (0.86,1.0) for model C1 (Figure 5B).

Comparisons between different CRS grading scales

Common Terminology Criteria for Adverse Events (CTCAE) grading scales do not adequately or accurately define CRS after T-cell engaging therapies. Thus, different sites and different publications have used different grading scales, making comparisons between studies difficult. Lee and colleagues and Davila and colleagues also published CRS grading scales for patients treated with CAR T cells (reproduced in Supplemental Tables 16 and 17).(12, 14) The grading systems are similar enough that the predictive models developed herein are relevant in these other grading systems. Regardless of "numerical grade" these models identify patients who develop life-threatening complications of CRS (mechanical ventilation and/or decompensated shock). The Lee grading scale differentiates patients requiring high-dose vasoactives and mechanical ventilation into grades 3 and 4, respectively. Thus, grade 3-5 patients on the Lee scale would be the equivalent to grade 4-5 patients in this series. Davila and colleagues defined severe CRS by a combination of fever for over 3 days, two cytokines with a max fold change of at least 75 or one cytokine with a max fold change of 250, at least one clinical sign of toxicity such as hypotension requiring vasoactives, hypoxemia, or neurologic disorders. Accordingly, our alternate severe CRS definition (CRS 3b-

5) reflects approximately the same population.

Comparisons of the different models

Through extensive statistical analysis, we developed a large number of models that are listed in Supplemental Table 12. We selected the models with the highest accuracy in the discovery cohort and presented these models in Table 3. Five different models were selected because they have different uses. First, we had several models that were very accurate at predicting severe CRS severity. We wanted to consider models fit to our entire cohort, which would use all of our data and be the most broadly applicable, and models specific to children as those models had the potential to be the most accurate and preferred for that cohort. Unlike our pediatric cohort, we did not feel like we enough individuals in our adult cohort to fit a separate model limited to adults. In Table 3 for our combined and pediatric cohorts, we presented the top logistic regression model, for which prediction is determined by a mathematical equation, and the top classification tree model, for which prediction is determined by "following the logic rules of the tree". We provide both as scientists and clinicians may prefer different methods for determining risk, and accuracy was similar. We provided one extra model for the pediatric cohort that was a classifier model using only the covariates that were selected by the best logistic regression model, which was highly accurate and likely to be used in practice. Since tree models are sometimes considered easier to implement, we wanted to investigate whether the classifier built on predictors from the top regression model would retain the same accuracy in both the discovery and validation cohorts.

Many of the models were accurate at predicting severity of CRS in our cohorts. It is not surprising that our statistical procedures identified multiple different models that accurately predict risk, because many cytokines trend together. Future work will evaluate all of the models in additional cohorts of patients and

determine if these models are also valid for other T cell engaging therapies. The predictive model(s) selected as the "best" now may not hold up as the "best" model(s) as more data become available. Thus, it is important to present all of the models. Of particular note, one of the most accurate models using our original definition of severe CRS (grade 4-5) and our alternate definition of severe CRS (grade 3b-5) was a combination of IL-10 plus disease burden. We hypothesize this combination may be the most informative moving forward with prospective trials.

It is important to note that all of the cytokines included in logistic regression models are not independently predictive of severe CRS. As mentioned, only sgp130 and IFNγ were independently associated with the later development of severe CRS. Additional cytokines in the logistic regression models, identified by forward selection, helped improve the overall accuracy, using information in the multivariate structure of the cytokines. For example, one of the models that predicts severe CRS in the combined cohort includes sgp130, IFNγ and IL1RA. IL1RA is not independently predictive of severe CRS. In contrast, in the model that includes sgp130 and IFNγ, IL1RA is helpful at lowering the false positive rate compared to that of the 2 variable model only using sgp130 and IFNγ.

Peak vs Fold Change Cytokine Measurement

Cytokine level comparisons are sometimes reported as fold-changes without consideration of the absolute values, and this can be misleading, as statistically significant differences may not be biologically or clinically meaningful. Increases in interferon inducible cytokines (IP-10 and MIG), as well as, IL6 and TNFα have been reported with advancing age in adults.(15-17) Cytokines, related to T cell function, often decrease with advanced age.(18) These differences are generally relatively small when considering absolute changes, and these differences are significantly lower in magnitude than those changes seen with rheumatologic illness, cancer, infection, or after T-cell engaging therapies, including CTL019. Nevertheless,

the normal variations seen in healthy populations can seem large when cytokines are interpreted as fold changes instead of absolutes. As an example, Berdat and colleagues found a statistically significant 3-fold difference in median IL6 levels comparing healthy children aged 25-36 months vs 0-3 months (1.75pg/ml vs 5.78pg/ml; p = 0.0165), respectively.(19) In order to understand cytokines levels that are biologically relevant after CTL019, we considered values in the context of the degree of variation seen in healthy populations and the levels reported in patients with inflammatory diseases and/or infection. We also considered both the absolute and relative differences between groups.

Supplemental Tables

Supplemental Table 1A: CRS grading system (adapted from Porter et al. Sci Trans Med 2015)(20)

Grade	Description of toxicity
Grade 1	Mild reaction treated with supportive care only
Grade 2	Moderate reaction requiring intravenous therapies or parenteral nutrition. Mild signs of organ dysfunction (creatinine <=grade 2 or LFTs <=grade 3) related to CRS.
Grade 3	More severe reactions. Moderate signs of organ dysfunction (grade 3 creatinine or grade 4 LFTs) related to CRS. Hypotension treated with intravenous fluids or low-dose vasoactive medications. Hypoxemia requiring oxygenation, CPAP, or BIPAP.
Grade 4	Life-threatening complications including hypotension requiring high-dose vasoactives or hypoxemia requiring mechanical ventilation.
Grade 5	Death

Supplemental Table 1B: Definition of high dose vasoactive medications (adapted from Porter et al. Sci Trans Med 2015)(20)

Vasoactive	Dose for >=3 hours
Norepinephrine monotherapy	>=0.2 mcg/kg/min
Dopamine monotherapy	>=10 mcg/kg/min
Phenylephrine monotherapy	>=200 mcg/min
Epinephrine monotherapy	>=0.1 mcg/kg/min
If on vasopressin	High dose if vasopressin + NE equivalent of >= 0.1 mcg/kg/min (using VASST formula)
If on combination vasopressors (not vasopressin)	Norepinephrine equivalent of >=20mcg/min (using VASST formula)*

***VASST Trial Vasopressor Equivalent Equation:**
Norepinephrine equivalent dose = [norepinephrine (mcg/min)] + [dopamine (mcg/kg/min) ÷ 2] + [epinephrine (mcg/min)] + [phenylephrine (mcg/min) ÷10] (adapted from Russell et al NEJM 2008)

Supplemental Table 2. Patient Numbers by Time points

NCT01626495 (CHP-959): Trial for pediatric cohort. (n = 39)

Timepoint	Cytokines*	Cytokines**	ALT	AST	BUN	Cr	Ferritin	CRP	LDH	PCR	Fibrin-ogen	PT	PTT
Baseline	38	39	39	39	39	39	37	36	38	39	0	1	1
Day 1-3	38	37	39	39	39	39	37	34	38	39	12	13	12
Day 4-5	7	7	27	27	28	28	22	22	22	7	11	11	11
Day 6-8	39	37	38	38	39	39	38	36	37	39	19	20	21
Day 9-11	35	35	38	38	39	39	39	37	38	37	17	16	15
Day 12-15	38	37	38	38	38	38	36	34	37	39	12	13	13
Day 16-18	31	31	36	36	36	36	35	34	34	34	9	9	9
Day 19-24	38	36	38	38	38	38	37	36	37	39	8	7	6
Day 25-35	36	36	36	36	36	36	36	35	35	37	3	5	5

NCT02167360 (UPENN 04409): Trial for adult cohort. (n =6)

Timepoint	Cytokines*	Cytokines**	ALT	AST	BUN	Cr	Ferritin	CRP	LDH	PCR	Fibrinogen	PT/PT
Baseline	6	6	6	6	6	6	6	4	6	6	6	6
Day 1-3	6	6	6	6	6	6	5	5	6	6	4	5
Day 4-5	0	0	3	3	6	6	6	5	5	0	4	5
Day 6-8	4	4	6	6	6	6	6	6	5	4	4	6
Day 9-11	6	6	6	6	6	6	6	6	6	6	4	6
Day 12-15	5	5	6	6	6	6	5	5	5	5	4	5
Day 16-18	1	1	3	3	3	3	3	2	3	1	2	2
Day 19-24	5	5	5	5	5	5	3	2	4	5	2	2
Day 25-35	6	6	6	6	6	6	4	2	6	6	2	2

NCT0029366 (UPENN 21413): Second trial for adult cohort. (n = 6)

Timepoint	Cytokines*	Cytokines**	ALT	AST	BUN	Cr	Ferritin	CRP	LDH	PCR	Fibrin-ogen	PT	PT
Baseline	6	6	6	6	6	6	6	6	6	6	6	6	6
Day 1-3	5	5	5	5	6	6	6	6	6	5	6	6	6
Day 4-5	2	2	5	5	5	5	5	3	5	3	4	5	5
Day 6-8	5	5	5	5	5	5	5	5	5	5	5	5	5
Day 9-11	5	5	5	5	5	5	5	5	5	5	5	5	5
Day 12-15	5	5	5	5	5	5	5	5	5	5	5	5	5
Day 16-18	0	0	2	2	2	2	2	2	2	0	2	2	2
Day 19-24	3	3	3	3	3	3	3	3	3	3	1	2	2
Day 25-35	3	2	3	3	3	3	3	2	3	3	1	3	3

* 30-plex Luminex panel
** 14-plex Luminex panel

Table above lists number of patients who had a sample that was collected and processed at specific time-points. Time-points are relative to day of first infusion of CTL019 = Day 0. Baseline = Within 7 days prior to and closest to day of first infusion.

Supplemental Table 3. HLH Diagnostic Criteria (adapted from Henter, et. al. Pediatr. Blood Cancer, 2007) (21)

1. Fever 2. Splenomegaly 3. Cytopenias (affecting >= 2 of 3 cell lineages in the peripheral blood) A. hemoglobin <9gm/dl; B. platelets <100K/ul; C. absolute neutrophil count <1000/ul 4. Fibrinogen <150mg/dl (or 3 SD below the mean) and/or fasting triglycerides >=2 mmol/L (or 3 SD above mean) 5. Low or absent NK-cell activity 6. Ferritin >500mg/dl 7. Soluble CD25 (sIL2Rα) >2400 U/ml 8. Hemophagocytosis in bone marrow, spleen or lymph node
Diagnosis: 5 of 8 criteria above OR molecular diagnosis based on pathogenic mutation in HLH-causative gene

Supplemental Table 4. Timing of severe CRS relative to CRS onset and time from infusion

Subject	Day CRS starts	Day HD vasoactives	Day met CRS4 criteria	Day first toci	Day intubated
Pediatric severe CRS discovery cohort					
959-100	2	5	5	8	6
959-111	1	10	4	4	4
959-113	1	5	5	5	N/A
959-114	0	9	9	10	10
959-120	1	3	3	5	8
959-125	2	5	5	5	N/A
959-131	1	5	5	5	5
959-136	0	10	10	10	N/A
959-138	1	6	6	6	6
959-139	1	3	3	3	N/A
959-140	1	5	5	5	N/A
Pediatric severe CRS validation cohort					
959-156	1	6	6	6	N/A
959-160	1	5	5	5	13
Adult severe CRS discovery cohort					
21413-3	1	4	4	3	3
21413-5	6	13	13	12	13
21413-6	0	10	10	2	10

All dates are relative to time of infusion of CTL019 which is considered Day 0. HD = high dose.
CRS = cytokine release syndrome. Toci = tocilizumab.

Supplemental Table 5. Patient Demographics and baseline characteristics (N=51)

Patient Characteristics	Total (N = 51)	Age Group		Grade	
		<25 years (N = 39)	>25 years (N = 12)	0 - 3 (N = 37)	4 - 5 (N = 14)
Sex - N (%)					
Female	23 (45%)	19 (49%)	4 (33%)	16 (43%)	7 (50%)
Male	28 (55%)	20 (51%)	8 (67%)	21 (57%)	7 (50%)
Age at Infusion, years					
Median (Min - Max)	15 (5 - 72)	11 (5 - 23)	56 (25 - 72)	15 (5 – 72)	13 (7 – 67)
Race - N (%)					
Caucasian	44 (86%)	34 (87%)	10 (83%)	31 (84%)	13 (93%)
Black	4 (8%)	3 (8%)	1 (8%)	4 (11%)	0 (0%)
Asian	1 (2%)	1 (3%)	0 (0%)	1 (3%)	0 (0%)
Pacific Islander: N (%)	1 (2%)	1 (3%)	0 (0%)	1 (3%)	0 (0%)
Other	1 (2%)	0 (0%)	1 (8%)	0 (0%)	1 (7%)
Allogeneic Transplant - N (%)	30 (59%)	26 (67%)	4 (33%)	23 (62%)	7 (50%)
Number Relapses - N (%)					
0 (Primary Refractory Disease)	3 (6%)	1 (3%)	2 (17%)	3 (8%)	0 (0%)
1	12 (24%)	6 (15%)	6 (50%)	8 (22%)	4 (29%)
≥2	36 (71%)	32 (82%)	4 (33%)	26 (70%)	10 (71%)

Supplemental Table 6. Co-morbid infections

Study	Organism/ Type of infection	Day post-infusion	Maximum CRS grade	Contributed to morbidity or mortality
959	Pseudomonas aeruginosa from tracheal culture	D+8	4	Unlikely. Clinically consistent with colonization and not infection.
959	Strep mitis bacteremia	D+1	4	Unlikely. Cleared with antimicrobials before became ill with CRS
959	Clostridium difficile in stool	D+18	2	Unlikely. Resolved with antimicrobials
959	Enterococcal bacteremia	D+9	2	Unlikely. Resolved with antimicrobials
21413	Influenza B	D+3	5	Likely contributed to death
21413	Multi-drug resistant Pseudomonas bacteremia and pneumonia	PM	5	Likely contributed to death

PM = identified post-mortem;

Supplemental Table 7. Median (range) for baseline cytokine values (pg/ml) are presented for study subjects (N = 50) and normal controls (N = 10), as well as by adult status for study subjects.

	Normal Subjects (N = 10)	Study Subjects (N = 50)	Children (< 25 years) (N = 38)	Adults (≥25 years) (N = 12)
IL1β	1.02 (0.67 - 187)	0.80 (0.23 - 3,704)	0.80 (0.23 - 3,704)	0.74 (0.55 - 74.5)
IL2	1.22 (0.68 - 9.16)	0.83 (0.05 - 1,838)	0.83 (0.05 - 1,838)	0.83 (0.33 - 24.2)
IL4	4.25 (3.27 - 7.11)	5.87 (1.09 - 165)	6.07 (1.35 - 165)	5.23 (1.09 - 24.0)
IL5	0.80 (0.35 - 4.46)	0.89 (0.04 - 167)	0.86 (0.24 - 167)	0.96 (0.04 - 13.9)
IL6	1.52 (0.29 - 87.6)	7.48 (1.01 - 1,020)	7.68 (1.01 - 1,020)	5.29 (1.67 - 58.4)
IL7	**0.43 (0.43 - 8.05)**	**1.06 (0.12 - 158)**	1.44 (0.12 - 158)	0.55 (0.42 - 4.41)
IL8	18.5 (12.9 - 50.1)	46.6 (0.86 - 2,168)	42.9 (0.86 - 2,168)	80.1 (2.62 - 483)
IL10	1.87 (1.29 - 270)	4.33 (0.86 - 2,153)	5.11 (1.32 - 2,153)	3.98 (0.86 - 10.9)
IL12	**184 (166 - 243)**	**92.4 (8.63 - 413)**	103 (8.6 - 413)	61.0 (17.6 - 187)
IL13	**35.7 (19.8 - 47.6)**	**16.4 (0.68 - 43.1)**	13.9 (0.68 - 43.1)	18.3 (0.71 - 29.7)
IL15	1.94 (1.94 - 222)	2.67 (0.71 - 41,899)	2.73 (0.71 - 41,899)	2.51 (1.06 - 575)
IL17	1.46 (0.89 - 13.2)	1.67 (0.06 - 103)	1.66 (0.06 - 2.44)	1.67 (1.42 - 103)
IL1RA	54.9 (30.8 - 235)	192 (5.23 - 2,510)	186 (5.23 - 2,510)	208 (48.7 - 717)
sIL1RI	114 (94 - 146)	95.1 (6.52 - 184)	89.3 (6.52 - 184)	113 (67.1 - 155)
sIL1RII	7,341 (5,864 - 10,618)	5,531 (2,160 - 54,071)	5,531 (2,160 - 54,071)	5,916 (4,694 - 14,985)
sIL2Ra	**312 (191 - 527)**	**1,400 (187 -232,571)**	1,701 (302 - 232,571)	681 (187 -106,432)
sIL4R	342 (277 - 573)	395 (60.5 - 742)	407 (129 - 742)	305 (60.5 - 482)
sIL6R	22,065 (15,910 - 29,589)	22,799 (10,182 - 50,657)	23,066 (10,318 - 50,657)	18,613 (10,182 - 28,017)
sgp130	182,508 (129,781 - 218,309)	213,419 (128,062 - 379,764)	216,294 (138,997 - 379,764)	186,505 (128,062 - 271,466)
TNFα	1.32 (0.83 - 1.94)	1.35 (0.02 - 111)	1.19 (0.11 - 111)	1.40 (0.02 - 6.40)
IFNα	**23.3 (16.5 - 28.6)**	**37.1 (14.0 - 264)**	36.7 (14.0 - 264)	37.1 (17.1 -87.5)
IFNγ	**19.7 (16.4 - 28.1)**	**4.58 (0.08 - 22.9)**	4.58 (0.08 - 22.9)	3.07 (0.15 -10.2)
MCP1	**357 (272 - 864)**	**1,770 (185 - 7,211)**	1,794 (185 - 7,211)	1,608 (625 - 5,340)
MIP1α	30.9 (21.0 - 39.3)	23.9 (7.99 - 1,008)	23.8 (7.99 - 1,008)	24.4 (12.0 - 84.8)
MIP1β	48.7 (30.2 - 96.5)	59.3 (15.3 - 8,216)	53.2 (15.3 - 8,216)	66.1 (28.9 - 217)
MIG	17.6 (5.59 - 42.5)	38.3 (0.34 - 591)	34.6 (0.34 - 591)	59.1 (6.04 - 193)
GCSF	101 (83 - 122)	61.2 (0.23 - 725)	69.8 (0.23 - 582)	51.6 (4.43 - 725)
GM-CSF	1.36 (1.08 - 88.9)	1.60 (0.07 - 527)	1.47 (0.07 - 527)	1.64 (0.68 - 2.30)
IP10	**19.0 (10.3 - 27.8)**	**68.0 (12.7 - 746)**	59.4 (12.7 - 746)	107 (36.4 - 195)
sTNFRI	**517 (288 - 694)**	**1,325 (508 - 3,858)**	1,272 (561 - 3,858)	1,366 (508 - 2,622)
sTNFRII	**4,723 (3,435 - 6,068)**	**10,686 (3,120 - 36,318)**	10,686 (4,963 - 36,318)	10,943 (3,120 - 18,089)
FGF-Basic	0.23 (0.23 - 79.4)	0.76 (0.02 - 2,609)	0.91 (0.13 - 2,609)	0.27 (0.02 - 84.2)
RANTES	**9,827 (2,179 - 17,352)**	**4,029 (25 -100,081)**	4,135 (25 -100,081)	3,170 (1,113 - 7,943)
Eotaxin	59.5 (43.8 - 140)	59.9 (17.6 - 157)	57.0 (17.6 - 127)	60.7 (30.7 - 157)
EGF	2.49 (0.81 - 15.6)	4.51 (0.28 - 1,223)	7.88 (0.28 - 1,223)	2.64 (0.56 - 84.0)
HGF	217 (113 - 292)	233 (47.2 - 12,327)	244 (47.2 - 12,327)	179 (77.2 - 4,384)
VEGF	0.39 (0.22 - 1.64)	1.12 (0.03 - 84.8)	1.17 (0.03 - 84.8)	0.59 (0.41 - 34.9)
sCD30	**161 (111 - 255)**	**60.9 (0.74 - 269)**	63.6 (0.74 - 269)	31.9 (3.24 - 129)
sEGFR	55,128 (46,646 - 81,733)	58,434 (31,247 -117,269)	60,369 (32,711 -117,269)	39,711 (31,247 - 65,628)
sRAGE	66.0 (36.1 - 91.3)	77.5 (1.53 - 538)	137 (2.5 - 538)	23.6 (1.53 - 56.9)
sVEGFR1	438 (291 - 802)	838 (2.52 - 16,540)	1,195 (2.52 - 7,467)	106 (61.3 - 16,540)
sVEGFR2	13,892 (8,336 - 36,932)	15,637 (4,443 - 58,790)	17,094 (9,323 - 58,790)	8,019 (4,443 - 12,466)
sVEGFR3	1,491 (809 - 2,828)	2,571 (467 - 41,589)	2,967 (541 - 41,589)	1,442 (467 - 29,293)

*** Baseline cytokine values for study subjects (N=50) were compared to 10 normal controls using the exact Wilcoxon test. Significance testing was done at the 0.05 level, with the Holm adjustment for multiple comparisons. Significant findings are bolded. Baseline was defined as the value observed within 7 days prior to and closest to day of infusion.**

Supplemental Table 8. Median (range) baseline cytokine values (pg/ml) by baseline disease burden in children (N = 39).

	High Disease Burden (>25%) [N = 23]	Low disease Burden (≤25%) [N = 16]
IL1β	0.72 (0.23 - 2.40)	0.93 (0.63 - 3,704)
IL2	0.78 (0.19 - 1.86)	0.84 (0.05 - 1,838)
IL4	4.35 (1.35 - 18.9)	6.80 (2.78 - 165)
IL5	0.85 (0.24 - 119)	0.95 (0.43 - 167)
IL6	7.80 (1.94 - 137)	7.40 (1.01 - 1,020)
IL7	0.96 (0.42 - 29.5)	1.45 (0.12 - 158)
IL8	54.6 (11.1 - 2,168)	30.9 (0.86 - 113)
IL10	4.02 (1.66 - 24.1)	7.08 (1.32 - 2,153)
IL12	**80.3 (8.63 - 206)**	**168 (33 - 413)**
IL13	**4.30 (0.68 - 29.9)**	**19.5 (3.45 - 43.1)**
IL15	2.71 (0.71 - 78.9)	2.87 (1.76 - 41,899)
IL17	1.53 (0.06 - 2.44)	1.68 (0.19 - 1.94)
IL1RA	113 (5.2 - 897)	551 (8.0 - 2,510)
sIL1RI	89.3 (38.1 - 179)	90.4 (6.52 - 184)
sIL1RII	5,724 (2,160 - 54,071)	5,509 (3,142 - 9,099)
sIL2Rα	1,733 (302 - 232,571)	1,395 (348 - 2,911)
sIL4R	429 (214 - 742)	406 (129 - 624)
sIL6R	23,231 (10,318 - 50,657)	20,844 (16,658 - 37,882)
sgp130	228,740 (168,928 - 379,764)	205,931 (138,997 - 330,876)
TNFα	1.46 (0.11 - 8.79)	0.97 (0.12 - 111)
IFNα	35.7 (14.0 - 101)	39.6 (21.0 - 264)
IFNγ	3.51 (0.08 - 22.9)	5.59 (0.55 - 19.3)
MCP1	1,787 (185 - 7,211)	1,800 (251 - 4,326)
MIP1α	22.0 (7.99 - 50.4)	28.8 (17.1 - 1,008)
MIP1β	50.2 (15.3 - 205)	64.4 (27.7 - 8,216)
MIG	34.3 (0.34 - 330)	39.0 (1.13 - 591)
GCSF	49.2 (0.23 - 334)	78.3 (4.49 - 582)
GM-CSF	1.27 (0.07 - 3.02)	1.62 (0.73 - 527)
IP10	62.4 (12.9 - 254)	57.5 (12.7 - 746)
sTNFRI	1,476 (810 - 2,512)	1,199 (561 - 3,858)
sTNFRII	9,420 (4,963 - 24,739)	11,390 (5,341 - 36,318)
FGF-Basic	0.47 (0.23 - 15.1)	4.59 (0.13 - 2,609)
RANTES	4,013 (25 - 33,417)	4,225 (1,544 - 100,081)
Eotaxin	41.6 (17.6 - 127)	68.1 (31.0 - 100)
EGF	**1.37 (0.28 - 54.9)**	**40.0 (0.62 - 1,223)**
HGF	226 (47 - 880)	301 (102 - 12,327)
VEGF	0.63 (0.03 - 14.6)	4.25 (0.26 - 84.8)
sCD30	63.4 (0.88 - 218)	90.2 (0.74 - 269)
sEGFR	63,544 (32,711 - 117,269)	55,524 (41,769 - 69,750)
sRAGE	128 (2.5 - 538)	145 (39 - 324)
sVEGFR1	971 (2.5 - 7,467)	1,659 (38 - 6,245)
sVEGFR2	16,904 (9,323 - 58,790)	18,127 (10,878 - 28,491)
sVEGFR3	3,246 (1,359 - 41,589)	2,597 (541 - 8,245)

Baseline cytokine levels between disease burden groups were compared using the exact Wilcoxon test. Significance testing was done at the 0.05 level, with the Holm adjustment for multiple comparisons. Significant findings are bolded. Baseline cytokine and disease burden observations were measured within 7 days prior to and closest to day of infusion.

Supplemental Table 9. Median (Range) for **one-month-peak cytokine values (pg/ml) in children and adults by grade (N = 51).**

	Total Cohort (N = 51)		Children (<25 years) (N = 39)		Adults (≥25 years) (N = 12)	
	CRS 0-3 (N = 37)	**CRS 4-5 (N = 14)**	**CRS 0-3 (N = 28)**	**CRS 4-5 (N = 11)**	**CRS 0-3 (N = 9)**	**CRS 4-5 (N = 3)**
IL1β	1.67 (0.52 - 3,652)	8.81 (0.60 - 125)	1.20 (0.52 - 3,652)	16.3 (0.60 - 125)	3.27 (0.55 - 120)	8.07 (1.59 - 8.29)
IL2	5.47 (0.78 - 1,832)	32.9 (1.42 - 296)	**3.49 (0.78 - 1,832)**	**37.0 (2.27 - 296)**	22.6 (1.27 - 63.1)	12.9 (1.42 - 138)
IL4	**8.06 (2.19 - 216)**	**24.5 (4.03 - 50.5)**	7.66 (2.19 - 216)	24.6 (4.03 - 50.5)	10.9 (2.77 - 27.9)	22.2 (11.7 - 27.3)
IL5	4.25 (0.39 - 264)	15.3 (1.71 - 333)	2.57 (0.39 - 264)	8.92 (1.97 - 284)	16.2 (2.71 - 188)	186 (1.71 - 333)
IL6	**122 (0.40 - 20,892)**	**8,309 (580 - 102,476)**	**69.1 (0.40 - 1,728)**	**8,273 (580 - 102,476)**	530 (113 - 20,892)	8,345 (1,088 - 37,324)
IL7	19.6 (0.42 - 233)	46.6 (0.53 - 284)	19.0 (0.42 - 164)	45.7 (0.53 - 284)	27.6 (0.54 - 233)	47.5 (43.3 - 49.0)
IL8	**192 (10.7 - 21,948)**	**9,523 (1,469 - 35,811)**	**119 (10.7 - 16,911)**	**11,684 (1,469 - 35,811)**	737 (73.0 - 21,948)	6,742 (5,862 - 25,081)
IL10	**47.6 (4.62 - 2,272)**	**265 (53.8 - 1,521)**	26.9 (5.29 - 2,272)	301 (53.8 - 1,084)	151 (4.62 - 517)	228 (132 - 1,521)
IL12	217 (38.1 - 772)	162 (45.7 - 583)	246 (56.3 - 772)	176 (45.7 - 276)	166 (38.1 - 424)	147 (78.0 - 583)
IL13	25.2 (1.66 - 90.3)	41.7 (10.2 - 78.4)	25.8 (1.66 - 90.3)	38.2 (10.2 - 78.4)	24.8 (13.2 - 69.1)	61.8 (32.0 - 62.3)
IL15	135 (1.76 - 41,701)	511 (122 - 951)	80.9 (1.76 - 41,701)	511 (122 - 951)	310 (31.9 - 900)	589 (388 - 742)
IL17	1.68 (1.32 - 87.82)	1.88 (1.43 - 52.74)	1.68 (1.32 - 7.38)	1.95 (1.43 - 20.95)	1.68 (1.42 - 87.82)	1.81 (1.70 - 52.74)
IL1RA	**734 (55.0 - 7,883)**	**1,629 (253 - 40,127)**	684 (55.0 - 6,584)	1,514 (253 - 40,127)	871 (522 - 7,883)	5,671 (1,739 - 11,332)
sIL1RI	**167 (62.7 - 368)**	**327 (163 - 607)**	**156 (62.7 - 344)**	**362 (163 - 607)**	189 (135 - 368)	254 (212 - 383)
sIL1RII	**9,897 (5,145 - 52,277)**	**87,146 (9,081 - 795,940)**	**8,715 (5,511 - 31,885)**	**102,950 (9,272 - 795,940)**	11,495 (5,145 - 52,277)	71,342 (9,081 - 234,397)
sIL2Rα	**8,002 (553 - 109,501)**	**63,022 (15,757 - 268,469)**	**6,930 (553 - 35,854)**	**62,771 (15,757 - 268,469)**	14,218 (1,988 - 109,501)	72,635 (17,267 - 86,034)
sIL4R	484 (203 - 887)	522 (428 - 870)	522 (203 - 887)	573 (428 - 870)	429 (252 - 677)	445 (440 - 520)
sIL6R	**31,086 (15,777 - 229,219)**	**187,225 (26,355 - 350,276)**	**30,935 (21,413 - 229,219)**	**194,880 (144,081 - 350,276)**	103,294 (15,777 - 221,705)	117,068 (26,355 - 138,107)
sgp130	**251,414 (143,605 - 395,926)**	**344,709 (274,650 - 566,041)**	**254,768 (191,302 - 395,926)**	**350,337 (274,650 - 566,041)**	221,940 (143,605 - 362,475)	295,846 (282,440 - 348,658)
TNFα	**2.58 (0.66 - 105)**	**10.3 (1.01 - 47.0)**	**2.44 (0.66 - 105)**	**10.1 (1.01 - 24.7)**	2.62 (1.69 - 18.15)	10.5 (7.31 - 47.0)
IFNα	**54.9 (18.4 - 258)**	**208 (43.1 - 325)**	**45.4 (18.4 - 258)**	**210 (43.1 - 325)**	85.4 (33.8 - 241)	138 (89.9 - 217)
IFNγ	**34.1 (2.14 - 8,233)**	**3,119 (160 - 15,482)**	**27.1 (2.14 - 5,070)**	**3,361 (160 - 15,482)**	188 (8.58 - 8,233)	1,449 (171 - 4,839)
MCP1	**2,859 (554 - 236,606)**	**54,831 (3,482 - 216,717)**	**2,304 (554 - 20,051)**	**63,093 (4,743 - 216,717)**	5,452 (1,371 - 236,606)	24,303 (3,482 - 80,968)
MIP1α	**44.6 (21.2 - 1,164)**	**383 (72.4 - 656)**	**43.2 (23.5 - 1,164)**	**353 (72.4 - 656)**	79.1 (21.2 - 697)	413 (193 - 626)
MIP1β	**114 (44.8 - 8,074)**	**563 (123 - 1,610)**	**105 (44.8 - 8,074)**	**658 (123 - 1,610)**	162 (81.8 - 2,458)	468 (457 - 1,048)
MIG	**634 (7.68 - 35,033)**	**5,416 (1,461 - 24,038)**	**528 (7.68 - 24,435)**	**5,464 (1,461 - 24,038)**	1,733 (85.8 - 35,033)	4,806 (3,467 - 9,560)
GCSF	**130 (10 - 9,233)**	**438 (179 - 27,751)**	**127 (10 - 4,286)**	**425 (179 - 4,669)**	409 (35 - 9,233)	1,062 (429 - 27,751)
GM-CSF	**4.20 (0.82 - 622)**	**73.6 (2.28 - 606)**	**3.56 (0.82 - 556)**	**71.9 (5.97 - 606)**	17.0 (2.41 - 622)	154 (2.28 - 591)
IP10	**655 (34 - 22,490)**	**13,222 (2,248 - 151,579)**	**420 (34 - 4,618)**	**17,389 (2,541 - 151,579)**	4,378 (172 - 22,490)	6,085 (2,248 - 9,759)
sTNFRI	**2,935 (925 - 17,633)**	**10,301 (5,149 - 34,278)**	**2,116 (925 - 17,633)**	**10,111 (5,149 - 34,278)**	5,700 (1,855 - 9,923)	14,662 (8,829 - 27,354)
sTNFRII	**29,046 (10,130 - 71,883)**	**78,268 (49,003 - 547,637)**	**25,221 (10,130 - 71,883)**	**73,312 (53,030 - 547,637)**	41,287 (22,400 - 70,417)	86,803 (49,003 - 136,400)
FGF-Basic	9.31 (0.13 - 3,096)	8.65 (0.27 - 54.11)	8.37 (0.13 - 3,096)	12.5 (0.32 - 54.1)	11.3 (0.20 - 87.4)	1.16 (0.27 - 3.86)
RANTES	5,330 (455 - 59,342)	4,764 (2,288 - 13,583)	5,264 (455 - 59,342)	4,703 (2,288 - 13,583)	5,571 (2,799 - 29,600)	4,764 (4,764 - 6,421)
Eotaxin	86.8 (28.5 - 241)	130 (69.9 - 443)	84.0 (28.5 - 141)	128 (69.9 - 443)	131 (77.5 - 241)	150 (75.5 - 177)
EGF	39.1 (0.32 - 1,199)	24.6 (9.22 - 223)	43.8 (0.32 - 1,199)	35.9 (9.22 - 223)	28.3 (2.22 - 157)	18.3 (17.5 - 27.3)
HGF	668 (198 - 12,635)	1,968 (377 - 12,450)	566 (198 - 12,635)	2,007 (377 - 12,095)	1,077 (334 - 5,743)	1,517 (1,358 - 12,450)
VEGF	**15.9 (0.37 - 130)**	**83.4 (15.2 - 146)**	**12.0 (0.37 - 102)**	**81.0 (15.2 - 146)**	25.5 (15.8 - 130)	85.7 (75.5 - 115)
sCD30	125 (0.74 - 439)	168 (35.0 - 547)	124 (0.74 - 376)	183 (35.0 - 547)	138 (57.4 - 439)	120 (49.1 - 280)

sEGFR	70,463 (37,797 - 111,450)	69,001 (50,946 - 148,360)	72,061 (50,761 - 111,450)	74,375 (54,456 - 148,360)	53,645 (37,797 - 96,448)	50,972 (50,946 - 58,832)
sRAGE	**172 (40.6 - 501)**	**478 (97.5 - 1,465)**	**209 (48.7 - 501)**	**406 (104 - 853)**	52.8 (40.6 - 472)	550 (97.5 - 1,465)
sVEGFR1	3,363 (61.3 - 27,812)	4,183 (749 - 10,099)	3,739 (416 - 27,812)	4,214 (749 - 10,099)	1,813 (61.3 - 4,751)	2,598 (2,050 - 7,314)
sVEGFR2	21,612 (8,110 - 74,457)	22,439 (6,206 - 36,792)	23,050 (13,660 - 74,457)	24,539 (15,494 - 36,792)	11,369 (8,110 - 24,351)	6,880 (6,206 - 17,584)
sVEGFR3	3,767 (1,274 - 27,575)	5,542 (2,933 - 26,066)	4,224 (1,274 - 12,653)	5,577 (2,933 - 26,066)	2,480 (1,481 - 27,575)	4,060 (3,594 - 21,007)

One-month-peak cytokine values are compared between those who developed severe (grade 4-5) CRS versus not using the exact Wilcoxon test. Significance testing was done at the 0.05 level with the Holm adjustment for multiple comparisons separately by total, children, adults cohorts. Significant findings are bolded.

Supplemental Table 10. Time points of target cytokine assessment listed as day since first CTL019 infusion, with the acceptable windows provided by study protocol.

Scheduled Day	Protocol		
	959 (Pediatric, N=39)	**04409 (Adult, N=6)**	**21413 (Adult, N=6)**
Day 1	Day 1		
Day 2	-	2	2
Day 3	-	3	3 - 4
Day 4	3 – 5	-	-
Day 5	2 – 8	-	-
Day 6	-	-	5 – 7
Day 9	6 – 8	-	
Day 10	9 – 11	9 – 11	9 -11
Day 11	-	10 – 12	-
Day 12	-	11 – 13	12 – 14
Day 13	-	12 – 14	-
Day 14	12 – 16	11 – 17	-
Day 17	14 – 20	-	-
Day 20	-	-	17 - 23
Day 21	17 – 25	18 – 24	18 – 24
Day 27	-	-	24 - 30
Day 28	24 – 32	25 – 31	-

Supplemental Table 11. Median (range) day 1-3 peak cytokine values (pg/ml) for those with severe (grade 4-5) CRS and without (grade 0-3) for study subjects (N=50), and by adult status (<25 versus ≥25 years)

	Total Cohort (N = 50)		Children (<25 years) (N = 39)		Adults (≥25 years) (N = 12)	
	CRS 0-3 (N = 36)	CRS 4-5 (N = 14)	CRS 0-3 (N = 27)	CRS 4-5 (N = 11)	CRS 0-3 (N = 9)	CRS 4-5 (N = 3)
IL1β	0.72 (0.36 - 3,652)	0.71 (0.40 - 8.29)	0.72 (0.36 - 3,652)	0.69 (0.40 - 2.01)	0.68 (0.55 - 120)	1.59 (0.67 - 8.29)
IL2	3.71 (0.36 - 1,832)	26.5 (0.85 - 138)	1.52 (0.36 - 1,832)	26.8 (1.19 - 92.9)	20.5 (1.27 - 63.1)	12.9 (0.85 - 138)
IL4	6.31 (1.56 - 191)	5.78 (1.69 - 28.28)	6.00 (1.56 - 191)	4.10 (1.69 - 28.28)	7.42 (2.44 - 18.07)	11.7 (6.90 - 17.6)
IL5	3.40 (0.04 - 106)	2.95 (0.11 - 333)	2.41 (0.04 - 106)	2.84 (0.11 - 21.70)	8.58 (0.99 - 46.52)	133 (0.33 - 333)
IL6	35.1 (0.38 - 1,105)	63.2 (5.29 - 8,399)	19.1 (0.38 - 1,105)	50.4 (12.0 - 1,137)	149 (12.3 - 474)	1,088 (5.29 - 8,399)
IL7	1.44 (0.12 - 164)	1.45 (0.42 - 196)	1.45 (0.12 - 164)	1.44 (0.42 - 196)	0.55 (0.48 - 27.57)	43.3 (1.35 - 47.5)
IL8	103 (8.89 - 3,980)	181 (54.9 - 9,100)	95.1 (8.89 - 786)	167 (54.9 - 8,351)	247 (16.3 - 3,980)	5,862 (132 - 9,100)
IL10	10.8 (1.15 - 2,272)	27.6 (10.8 - 132)	9.94 (1.15 - 2,272)	23.8 (11.9 - 65.7)	25.1 (2.86 - 90.8)	129 (10.8 - 132)
IL12	165 (3.16 - 665)	87.3 (12.9 - 258)	212 (3.16 - 665)	91.3 (12.9 - 258)	66.2 (17.2 - 424)	83.4 (78.0 - 184)
IL13	13.6 (0.39 - 53.9)	13.7 (1.56 - 62.3)	16.1 (0.39 - 46.1)	6.28 (1.56 - 47.56)	13.2 (5.66 - 53.9)	61.8 (28.2 - 62.3)
IL15	30.2 (0.71 - 41,701)	120 (1.92 - 742)	6.34 (0.71 - 41,701)	117 (2.37 - 511)	164 (5.82 - 900)	589 (1.92 - 742)
IL17	1.67 (0.07 - 87.82)	1.67 (0.23 - 4.96)	1.67 (0.07 - 4.11)	1.66 (0.23 - 4.96)	1.67 (1.42 - 87.82)	1.70 (1.67 - 1.81)
IL1RA	415 (4.60 - 7,845)	306 (3.17 - 11,332)	258 (4.60 - 2,414)	285 (3.17 - 968)	472 (201 - 7,845)	472 (64.0 - 11,332)
sIL1RI	120 (12.4 - 210)	128 (57.7 - 248)	119 (12.4 - 210)	126 (57.7 - 228)	139 (92.5 - 189)	212 (108 - 248)
sIL1RII	6,143 (2,987 - 12,652)	7,503 (2,460 - 71,342)	5,948 (2,987 - 11,356)	5,368 (2,460 - 39,944)	7,598 (4,305 - 12,652)	10,766 (7,729 - 71,342)
sIL2Rα	3,267 (661 - 102,471)	7,525 (1,100 - 249,796)	2,499 (661 - 14,208)	5,813 (1,100 - 249,796)	5,010 (1,123 - 102,471)	41,685 (1,761 - 72,635)
sIL4R	398 (170 - 778)	446 (316 - 747)	448 (170 - 778)	478 (381 - 747)	364 (174 - 677)	401 (316 - 440)
sIL6R	21,473 (9,460 - 45,654)	21,299 (7,858 - 32,142)	22,294 (14,952 - 45,654)	20,979 (7,858 - 32,142)	15,212 (9,460 - 25,178)	26,355 (18,058 -27,445)
sgp130	**201,493 (117,065 - 319,392)**	**254,052 (201,232 - 02,946)**	211,976 (144,191-319,392)	249,111 (201,232-402,946)	170,828 (117,065-229,251)	282,279 (219,625-282,440)
TNFα	1.68 (0.13 - 105)	1.38 (0.51 - 9.69)	1.42 (0.13 - 105)	1.21 (0.51 - 4.10)	2.10 (0.83 - 6.25)	7.31 (1.22 - 9.69)
IFNα	44.3 (14.9 - 258)	64.2 (22.7 - 197)	40.3 (14.9 - 258)	62.3 (22.7 - 197)	72.3 (33.8 - 186)	138 (37 - 174)
IFNγ	**15.2 (0.13 - 404)**	**140 (10.5 - 1,621)**	**10.8 (0.13 - 341)**	**59.9 (10.8 - 544)**	77.8 (8.27 - 404)	1,449 (11 - 1,621)
MCP1	2,082 (210 - 22,754)	6,294 (921 - 26,254)	1,846 (210 - 7,655)	4,743 (1,053 - 15,479)	4,465 (554 - 22,754)	24,303 (921 - 26,254)
MIP1α	33.0 (7.74 - 992)	30.3 (20.0 - 194)	32.6 (7.74 - 992)	26.8 (20.0 - 66.1)	39.9 (18.8 - 125)	193 (33.2 - 194)
MIP1β	83.3 (24.0 - 8,074)	81.4 (21.9 - 457)	81.6 (24.0 - 8,074)	73.5 (21.9 - 207)	87.6 (53.1 - 424)	329 (68.6 - 457)
MIG	231 (5.22 - 4,232)	369 (37.7 - 9,560)	176 (5.220 - 1,188)	319 (53.9 - 1,824)	476 (69.0 - 4,232)	4,806 (37.7 - 9,560)
GCSF	85.7 (3.93 - 779)	104 (15 - 1,718)	86.0 (3.93 - 541)	85.9 (30.9 - 1,718)	45.2 (34.8 - 779)	707 (15.0 - 1,062)
GM-CSF	3.56 (0.82 - 556)	11.1 (0.80 - 591)	2.50 (0.82 - 556)	10.4 (0.80 - 137)	13.4 (1.64 - 107)	154 (0.83 - 591)
IP10	212 (14.4 - 4,378)	585 (82.9 - 9,759)	165 (14.4 - 1,163)	555 (82.9 - 4,741)	768 (163 - 4,378)	6,085 (106 - 9,759)
sTNFRI	1,938 (792 - 7,102)	2,914 (1,247 - 12,697)	1,742 (792 - 5,070)	2,547 (1,247 - 12,697)	2,677 (1,472 - 7,102)	7,432 (1,690 - 8,829)
sTNFRII	16,781 (6,572 - 41,287)	23,203 (13,302 - 66,627)	15,163 (6,572 - 35,066)	21,531 (13,611 - 66,627)	20,775 (11,337 - 41,287)	49,003 (13,302 -50,446)
FGF-Basic	3.19 (0.13 - 3,096)	1.90 (0.13 - 11.17)	3.87 (0.13 - 3,096)	2.77 (0.32 - 11.17)	1.38 (0.20 - 87.43)	0.27 (0.13 - 3.86)
RANTES	3,239 (1.57 - 29,600)	2,635 (558 - 13,583)	3,742 (1.57 - 28,476)	2,067 (558 - 13,583)	3,214 (1,889 - 29,600)	4,764 (3,757 - 6,421)
Eotaxin	56.6 (21.9 - 187)	41.3 (15.5 - 147)	54.2 (21.9 - 134)	34.4 (15.5 - 99.3)	76.3 (36.8 - 187)	75.5 (49.8 - 147)

EGF	13.4 (0.32 - 1,199)	4.91 (0.55 - 53.15)	12.2 (0.32 - 1,199)	4.84 (0.55 - 53.15)	16.1 (0.53 - 118)	17.5 (3.15 - 18.3)
HGF	352 (112 - 12,635)	453 (183 - 3,241)	358 (112 - 12,635)	451 (183 - 3,241)	334 (188 - 1,895)	688 (251 - 1,517)
VEGF	2.98 (0.26 - 87.11)	5.36 (0.43 - 99.29)	2.16 (0.26 - 87.11)	5.01 (0.43 - 36.64)	17.8 (0.52 - 27.5)	75.5 (0.56 - 99.3)
sCD30	60.1 (0.74 - 299)	49.3 (0.88 - 163)	62.8 (0.74 - 299)	63.3 (0.88 - 163)	57.4 (3.24 - 129)	49.1 (28.6 - 49.1)
sEGFR	55,033 (33,737 - 94,490)	61,023 (28,632 - 148,360)	60,427 (39,888 - 94,490)	68,253 (39,013 - 148,360)	39,911 (33,737 - 69,721)	47,031 (28,632 -58,832)
sRAGE	89.6 (0.36 - 311)	94.6 (18.0 - 419)	141 (0.12 - 311)	142 (43.3 - 419)	38.4 (1.53 - 51.4)	91.7 (18.0 - 97.5)
sVEGFR1	998 (61.3 - 24,142)	701 (61.31 - 3,547)	1,209 (193 - 24,142)	589 (105 - 3,547)	287 (61.3 - 3,224)	1,387 (61 - 2,050)
sVEGFR2	14,941 (5,332 - 58,390)	13,609 (4,509 - 22,812)	17,709 (8,291 - 58,390)	13,767 (10,764 - 22,812)	8,257 (5,332 - 10,998)	6,880 (4,509 - 10,025)
sVEGFR3	2,993 (1,075 - 27,575)	3,349 (1,290 - 26,066)	3,273 (1,512 - 6,312)	3,408 (1,290 - 26,066)	1,435 (1,075 - 27,575)	2,551 (1,719 - 21,007)

Day 1-3 peak cytokine values are compared between those who developed severe (grade 4-5) CRS versus not using the exact Wilcoxon test. Significance testing was done at the 0.05 level with the Holm adjustment for multiple comparisons separately by total, children, adults cohorts. Significant findings are bolded.

Supplemental Table 12. Summary of all predictive models generated by discovery cohort and validated

Model*	Equation+	Prediction Cut off	Model Accuracy	Validation accuracy
Combined cohort (pediatric and adult) models				
Top 2-variable logistic regression model using peak values	expit{12.8046*log10(sgp130Pk3) + 1.5578*log10(IFNγPk3) - 71.996}	0.2326	Sens=12/14 Spec=30/35	Sens=2/2 Spec=5/10
Top 3-variable logistic regression model using peak values	expit{13.8712*log10(sgp130Pk3) + 2.4626*log10(IFNγPk3) - 1.6559*log10(IL1RAPk3)-75.3502}	0.3623	Sens=12/14 Spec=31/35	sens=2/2 spec=6/10
Top 2-variable logistic regression model using fold change	expit{2.0616*log10(IL2PkFC3) -5.8483*log10(EotaxinPkFC3) - 3.1313}	0.3208	Sens=11/14 Spec=29/34	Sens=2/2 Spec=5/10
Top 3-variable logistic regression model using fold change	expit{1.9862*log10(IL2PkFC3) -7.1063*log10(EotaxinPkFC3) + 28.755*log10(sgp130PkFC3)-3.2407}	0.2817	Sens=13/14 Spec=30/34	Sens=2/2 Spec=5/10
Top 3-variable logistic regression model using peak and fold change	expit{19.360*log10(sgp130Pk3) + 3.720*log10(IFNγPk3) – 3.184*log10(IL5Pk3FC)-109.541}	0.4714	Sens=12/14 Spec=32/34	Sens=1/2 Spec=8/10
Top tree model using variables selected by logistic regression	Predict low risk if sgp130Pk3 < 218179; failing that if IFNγk3 < 10.4272; failing that if IL1RAPk3 >= 657.987	---	Sens=11/14 Spec=33/35;	Sens=2/2 Spec =5/10
Top tree model using peak values and fold change (same as if using peaks only)	Predict low risk if sgp130Pk3 < 218179 or failing that if MCP1Pk3<4636.52 & EotaxinPk3>29.09	---	Sens= 12/14 Spec=34/35	Sens=2/2 Spec =7/10
Lasso model (penalized logistic regression with cross validation)	expit{-28.869 -0.7200 log10(EotaxinPk3) + 0.55417 log10(IFNγPk3) +5.07982 log10(sgp130Pk3)+ 0.21308 log10(sIL2RAPk3) -1.013127 log10(EotaxinPkFC3) + 0.1599 log10(IL2PkFC3) + 11.2725 log10(sgp130PkFC3) -4.611926 log10(sIL6RPkFC3)}	0.2607	Sens=13/14 Spec=30/34;	Sens=2/2 Spec =5/10
Pediatric only models				
Top 2 variable logistic regression model using peak values	expit{ 4.305*log10(IFNγPk3) -4.428*log10(IL13Pk3) - 3.369 }	0.3413	Sens=10/11 Spec= 23/27	Sens=1/2 Spec=7/10
Top 3 variable logistic regression model using peak values	expit{ 8.483*log10(IFNγPk3) - 5.599*log10(IL13Pk3) - 16.343log10(MIP1αPk3) +15.742 }	0.3288	Sens=11/11 Spec=26/27	Sens=1/2 Spec=8/10
Top 2 variable logistic regression model using fold change	expit{ 5.185*log10(IL2PkFC3) -13.205*log10(EotaxinPkFC3) -6.630}	0.3370	Sens=10/11 Spec= 25/26	Sens=2/2 Spec=6/10
Top tree model using only variables selected by logistic regression	Predict low risk if IFGγk3<27.6732 or failing that if (MIP1αPk3 ≥30.1591 and IFGPk3 < 94.8873		Sens=9/11 Spec=25/27	Sens=1/2 Spec=10/10
Top tree model based on peak values	Predict low risk if IL10Pk3 < 11.7457, or failing that if burden <50%		Sens=10/11 Spec=26/27	Sens 2/2 Spec=9/10
Top tree model based on variables selected by regression+burden	Predict low risk if IFNγPk3 < 27.6732 or failing that if burden < 51%		Sens=9/11 Spec=26/27	Sens=2/2 Spec =10/10
Top tree model using peak values and fold change	Predict Low risk if IL2PKFC3 < 12.2154	---	Sens=9/11 Spec 24/26	Sens=1/2 Spec=7/10
Lasso model (penalized logistic regression with cross validation)	expit{6.8102- .65091 log10(EotaxiniPk3) +1.10064 log10(sIL2RαPk3) -2.7939 log10(sIL6Rpk3)-0.235124 log10(sVEGFR1pk3) -3.66899 log10(EotakinPkFC3)+1.6778 log10(IL2PkFC3)+0.103514 log10(IL6PkFC3)+0.820699 log10(IL8PkFC3) + 0.79372 log10(sEGFRPkFC3) +10.9054 log10(sgp130PkFC3)+0.0005840 burden}	0.4097	Sens=11/11 Spec=25/25	Sens=2/2 Spec=8/10

*Model name describes which cytokine variables were considered in the model selection procedure; the following clinical variables were also considered in the selection procedure: blood urea nitrogen (BUN), creatinine, C-reactive protein, age at infusion, whether or not CRS occurred in the first 2 days post infusion, gender and, for pediatric models, baseline disease burden.

+The *expit* function converts the logistic regression score to the modeled probability of being a case. expit(x) = exp(x)/{exp(x) + 1}; FC=fold change; Pk=peak;

Supplemental Table 13. Clinical details on validation cohort (N = 12)

Patient Characteristics	Total (N = 12)
Sex - N (%)	
Female	6 (50%)
Male	6 (50%)
Age at Infusion, years	
Median (Min - Max)	11 (10 - 16)
Race - N (%)	
Caucasian	9 (75%)
Black	1 (8%)
Asian	1 (8%)
Other	1 (8%)
Time to CRS, days (N = 10):	2.5
Median (Min – Max)	(1 – 10)
Received TOCI: N (%)	2/10* (20%)
Allogeneic Transplant - N (%)	12 (100%)
Number Relapses - N (%)	
0 (Primary Refractory Disease)	1 (8%)
1	4 (33%)
≥2	7 (58%)

*Number of validation cohort patients with CRS.

Supplemental Table 14: Therapies used for CRS

Drug	Cohort	Drug
Tocilizumab	Pediatric	13 of 39
	Adult	8 of 12
Corticosteroids	Pediatric	7 of 39
	Adult	5 of 12
Siltuximab	Pediatric	0 of 39
	Adult	0 of 12
Etanercept	Pediatric	1 of 39
	Adult	1 of 12

Supplemental Table 15. Tocilizumab (Toci) Use: Subjects with CRS Receiving Toci, by Grade (N = 21). Median Days (Range) (unless otherwise noted)

	Total Cohort (N = 21)	Total Cohort		Children		Adults	
		Grade 3 (N = 7)	Grade 4-5 (N = 14)	Grade 3 (N = 2)	Grade 4-5 (N = 11)	Grade 3 (N = 5)	Grade 4-5 (N = 3)
Number of subjects treated with >1 dose: N (%)	10 (48%)	4 (57%)	6 (43%)	0 (0%)	4 (36%)	4 (80%)	2 (67%)
Days to TOCI relative to Infusion	5 (2 - 12)	5 (3 - 9)	5 (2 - 12)	6 (5 - 7)	5 (3 - 10)	5 (3 - 9)	3 (2 -12)
Days to TOCI relative to fever [N = 20]	4 (1 - 10)	4.5 (2 - 7)	4 (1 - 10)	5.5 (4 - 7)	4 (2 - 10)	3.5 (2 - 6)	2 (1 – 6)
Days to TOCI relative to vasoactives [N = 18]	0 (-8 - 6)	0.5 (-1 - 3)	0 (-8 - 6)	1.5 (0 - 3)	0 (-6 - 6)	0 (-1 - 1)	-1 (-8 - -1)
Days to TOCI relative to intubation [N = 9]	0 (-8 - 2)		0 (-8 - 2)		0 (-3 - 2)		-1 (-8 – 0)
Days to fever stop relative to TOCI [N = 20]	2.5 (-3 - 11)	2.5 (0 - 5)	2.5 (-3 - 11)	0 (0 - 0)	2 (-3 - 8)	4 (1 - 5)	3 (2 – 11)
Days to vaso stop relative to TOCI [N = 18]	4 (1 - 13)	**3 (1 - 3)**	**4.5 (1 - 13)**	3 (3 - 3)	5 (1 - 9)	2 (1 - 3)	4 (2 – 13)
Days extubation relative to TOCI [N = 9]	7 (2 - 62)		7 (2 - 62)		7 (5 - 62)		4 (2 - 13)

p < 0.05 (bold); Wilcoxon exact test

Supplemental Table 16: Lee CRS grading scale (adapted from Lee et al. Blood 2014)[14]

Grade	Description of toxicity
Grade 1	Non-life threatening symptoms requiring symptomatic treatment only
Grade 2	Symptoms that require moderate intervention, including oxygen requirement <40% or hypotension that responds to IV fluids or low dose of one vasopressor or grade 2 organ toxicity
Grade 3	Symptoms that require and respond to aggressive intervention, including oxygen requirement >=40% or hypotension requiring high dose or multiple vasoactives* or grade 3 organ toxicity or grade 4 transaminitis.
Grade 4	Life-threatening symptoms including requirement for ventilator support or grade 4 organ toxicity (excluding transaminitis)
Grade 5	Death

*See supplemental Table 1B

Supplemental Table 17: Davila CRS grading scale (adapted from Davila, et. al Science Translational Med, 2014)[12]

Criteria for severe CRS
Fevers for at least three consecutive days Two cytokine max fold changes of at least 75 or one cytokine max fold change of at least 250 At least one clinical sign of toxicity such as hypotension (requiring at least one intravenous vasoactive pressor) or, Hypoxia (PO2 < 90%) or, Neurologic disorders (including mental status changes, obtundation, and seizures)

Supplemental Table 18. Reclassification of CRS severity by dividing grade 3 into 2 groups

Revised CRS grade	Original CRS grade	Revised CRS severe (Y/N)	Original CRS severe (Y/N)	Subjects (N)	Pediatric Subjects (N)	Adults (N)
0	0	No	No	3	3	0
1	1	No	No	3	2	1
2	2	No	No	15	15	0
3a	3	No	No	8	5	3
3b	3	Yes	No	8	3	5
4	4	Yes	Yes	12	11	1
5	5	Yes	Yes	2	0	2

Patients with grade 3 CRS were divided into two groups based on the need for any vasoactive medications or having a significant oxygen requirement (>=40% FIO_2).

Supplemental Table 19. Summary of Top Predictive Models with alternate determination of CRS severity

Model+	Equation	Discovery Accuracy	Validation accuracy
Revised Models Generated with Alternate CRS Severity Determination			
A1: Top Pk3 logistic regression model for the combined cohort, alternate CRS severity	Predict low risk if expit{4.109*log10(IFNγPk3) – 8.526*log10(VEGFPk3) + 6.010*log10(IL6Pk3)-11.678} ≤ 0.4663	Sens=18/22 0.82 (0.60,0.95) Spec=24/28 0.86 (0.67,0.96)	Sens=2/2 Spec=3/10
B1: Top tree model for combined cohort, alternate CRS severity	Predict low risk if IL12Pk3 >270.855 or failing that if (IL10Pk3 < 11.7457 and sVEGFR2Pk3>12945.3); failing that if (IL10pk3 ≥11.7457 and sgp130Pk3 <210402)	Sens= 18/22 0.82 (0.60,0.95) Spec=25/27 0.93 (0.76,0.99)	Sens=1/2 Spec=10/10
C1: Top logistic regression model for the pediatric cohort, alternate CRS severity	Predict low risk if expit{11.269*log10(MCP1Pk3) – 8.118*log10(EotaxinPk3) – 3.335*log10(VEGFPk3) -24.209} ≤ 0.3342	Sens=13/14 0.93 (0.66,1.0) Spec=21/24 0.88 (0.68,0.97)	Sens=2/2 Spec=8/10
D1: Top tree model for pediatric cohort , alternate CRS severity	Predict low risk if IL10Pk3 < 11.7457; or Failing that if burden < 47	Sens= 12/14 0.86 (0.57,0.98) Spec=24/24 1.0 (0.86, 1.0)	Sens=2/2 Spec= 9/10
E1: Best pediatric classifier using covariates from top pediatric regression model, alternate CRS severity	Predict low risk if MCP1pk3 < 4210.44 and Eotaxinpk3 > 29.251	Sens= 13/14 0.93 (0.66,1.0) Spec=21/24 0.88 (0.68,0.97)	Sens=2/2 Spec=4/10
Performance of models generated with original CRS severity but applying patients divided by alternate CRS severity definition			
A2: Logistic model using original Model A variables for combined cohort ; alternate CRS severity	Predict low risk if expit{5.6914*log10(sgp130Pk3) + 1.6654*log10(IFNγPk3) - 0.7002*log10(IL1RAPk3)-31.3060} ≤ 0.3623	Sens=19/22 0.86 (0.65,0.97) Spec=21/27 0.78 (0.58,0.91)	Sens=2/2 Spec=2/10
C2: Logistic model using original Model A variables for pediatric; alternate CRS severity	Predict low risk if expit{2.5467*log10(IFNγk3) – 0.9643*log10(IL13Pk3) – 5.2301*log10(MIP1αPk3) +4.7698} ≤ 0.4316	Sens= 12/14 0.86 (0.57,0.98) Spec=21/24 0.88 (0.68,0.97)	Sens=1/2 Spec=7/10

*The *expit* function converts the logistic regression score to the modeled probability of being a case. expit(x) = exp(x)/{exp(x) + 1}; Pk=peak; sens = sensitivity; spec = specificity. Original CRS severity: Severe = Grades 4-5; Not severe Grades 0-3. Alternate CRS severity: Severe = Grades 3b-5; Not Severe = 0-3a.

Supplemental Figure Legends

Supplemental Figure 1. CRP and ferritin are poor early predictors of severe CRS. CRP and ferritin were measured in subjects treated with CTL019. Sampling frequency varied between subjects (see Supplemental Table 2). 45 (48) out of 51 subjects had at least one measure of CRP (ferritin) in the first 3 days after infusion, prior to development of severe CRS. Top (bottom) left panel depicts ROC curve for CRP (ferritin) in first 3 days post infusion, demonstrating poor sensitivity and specificity for severe CRS. 49 (51) out of 51 subjects had at least one measure of CRP (ferritin) in the month following infusion. Top (bottom) right panel depicts ROC curve for CRP (ferritin) in the first month after infusion, demonstrating improved sensitivity and specificity.

Supplemental Figure 2. Early increases in INFγ and sgp130 after CTL019 infusion are associated with development of severe CRS. Cytokines were analyzed from the first 3 days after infusion, sent before patients developed severe CRS. Cytokine profiles were compared in patients who later developed severe CRS (grade 4-5) with patients who did not. Only two cytokines, IFNγ and sgp130 are differentially elevated in the first three days after infusion in patients who later develop severe CRS as compared with those who do not and significant by Holm's adjusted p-value.

Supplemental Figure 3. Severity of CRS correlates moderately with CAR T cell expansion but early expansion of CAR T cells does not predict CRS severity. CTL019 cells were serially measured in peripheral blood by quantitative PCR. Supplemental Figure 3 depicts CAR T cell expansion in copies/microgram qPCR in the first 3 days (panel A) and one month (panel B) after CTL019 infusion, comparing patients who developed severe CRS with those who did not.

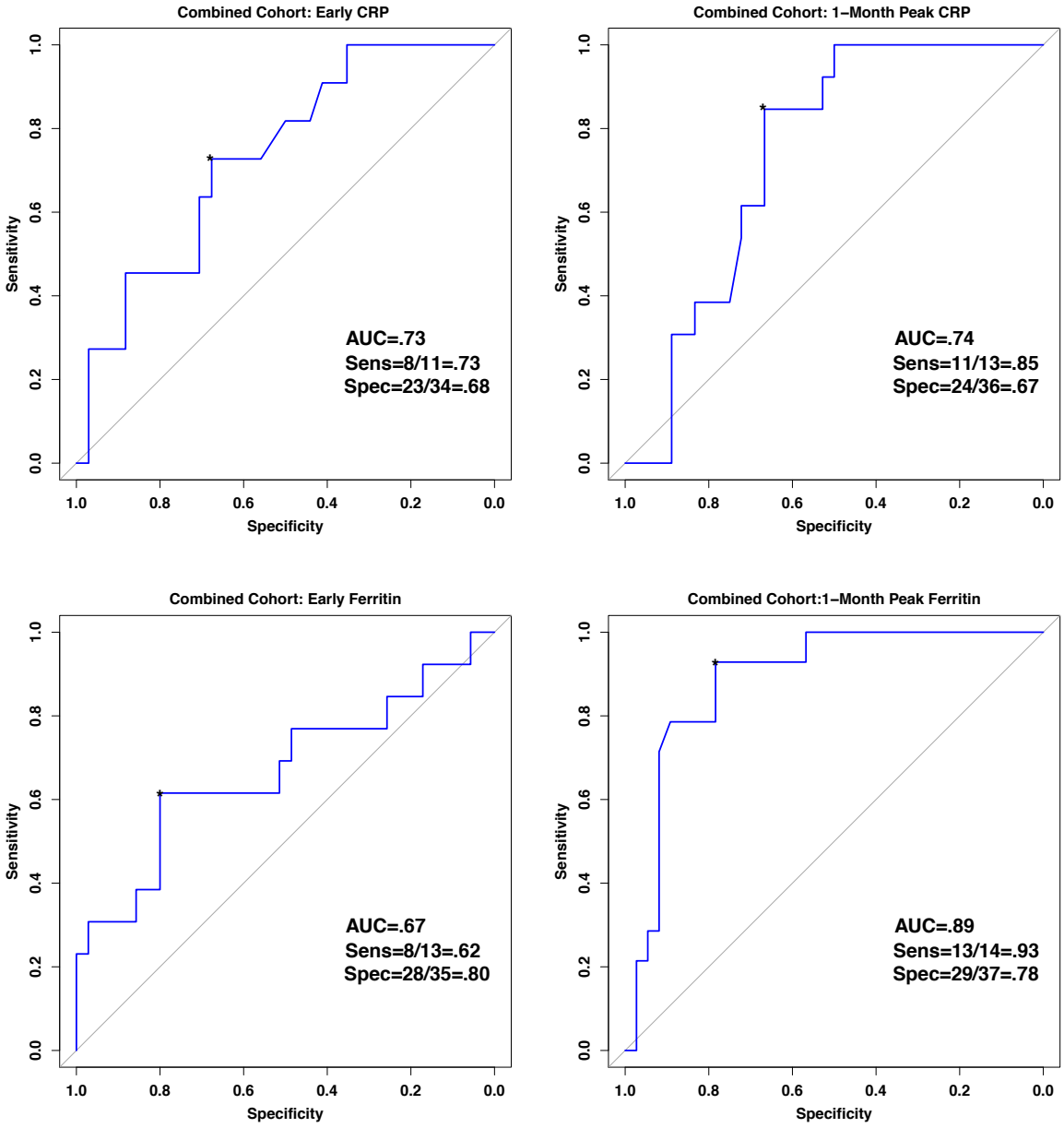
Supplemental Figure 4. Cytokine profiles predict CRS. Cytokines were analyzed from the first 3 days after infusion, sent before patients developed severe CRS. Logistic modeling was used to develop predictors of severe CRS. For the combined cohort with a single variable regression model, found by forward selection, the best cytokine for predicting later development of severe CRS was sgp130 (panel A). For the pediatric cohort, the best cytokine at predicting severe CRS was IFNγ (panel B). With a 2 variable regression model, found by forward selection, the models were almost as robust as the three variable regression models depicted in Figure 2 (main document). We could accurately predict which patients would develop severe CRS in the combined cohort with a combination of sgp130 and IFNγ (panel C) and in the pediatric cohort with a combination of INFγ and IL13 (panel D).

Supplemental Figure 5. Alternate CRS Predictive Models. Most of the models presented herein were designed to identify patients who would become critically ill with life-threatening complications, including respiratory failure and decompensated shock, eg grade 4-5 CRS. An alternative definition for severe CRS was developed that also included some grade 3 patients who became quite ill and required FIO_2 >=40% and/or lower dose vasoactives, eg they needed vasoactives but did not meet the high dose vasoactive definition described in Supplemental Table 1b (termed grade 3b). Cytokines were analyzed in the first 3 days after infusion, sent before patients developed severe CRS. Logistic and classification tree models were used to identify predictors of severe CRS, using the alternative definition (Grade 0-3a vs 3b-5). With a 3-variable regression model, found by forward selection, we predicted which patients developed severe CRS using IFNγ, VEGF, and IL6 (panel A) in the combined cohort and MCP1, Eotaxin, and VEGF (panel B) in the pediatric cohort. Using decision tree modeling, we needed a combination of 4 cytokines (panel C) for accurate prediction of severe CRS, using the alternate definition. In the pediatric cohort, we found that IL10 and disease burden could predict severe CRS, using the alternate definition. Of note, this combination was also the most accurate at predicting severe CRS using the original definition (grade 0-3 vs 4-5), but the

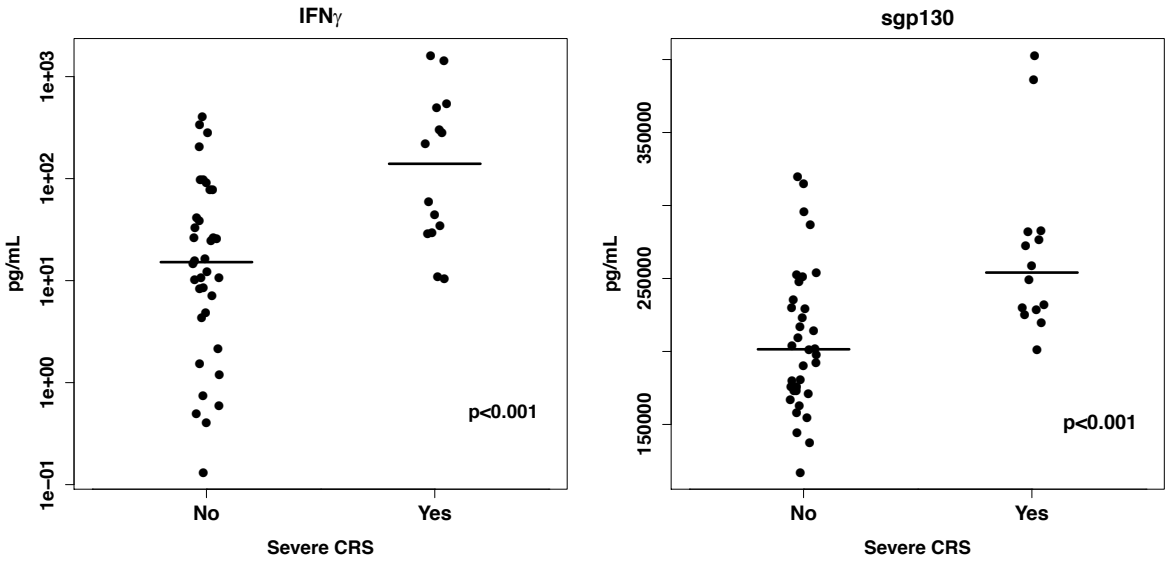
cut-offs were slightly different (panel D). Decision tree modeling built on the predictors selected for Model B, which excluded disease burden, a combination of MCP1 and Eotaxin was accurate in the pediatric only cohort (panel E).

Supplemental Figures

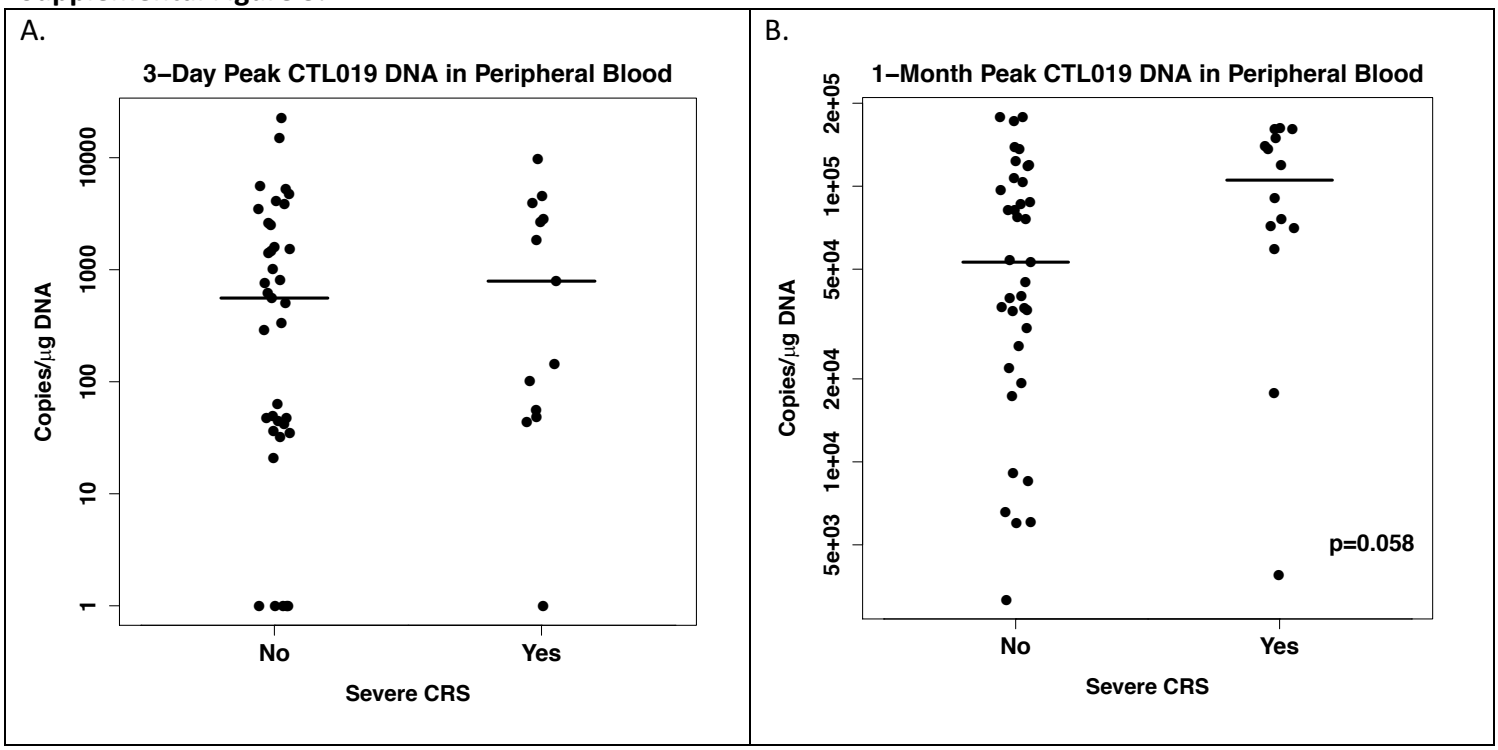
Supplemental Figure 1.



Supplemental Figure 2.

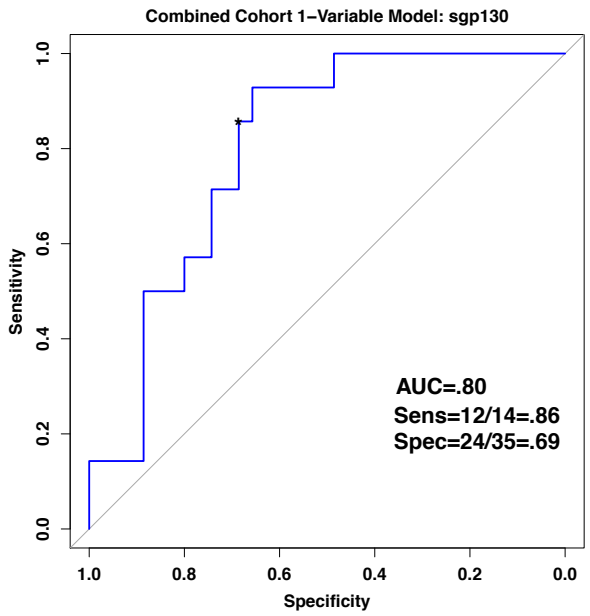


Supplemental Figure 3.

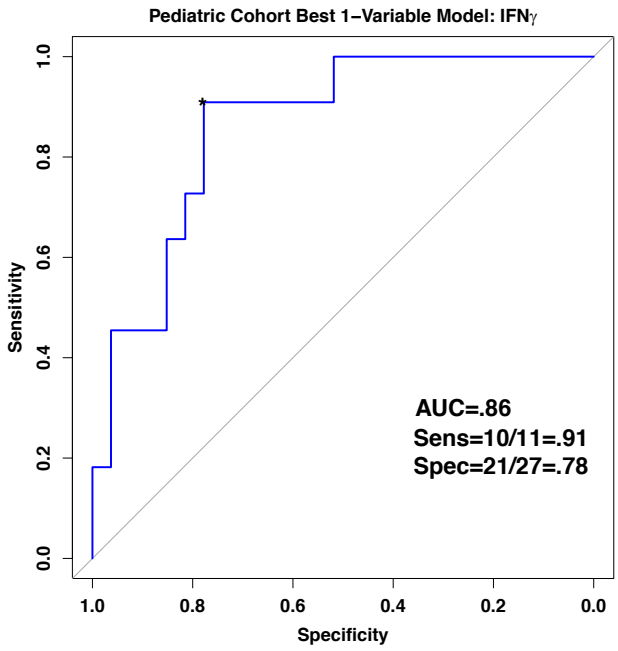


Supplemental Figure 4.

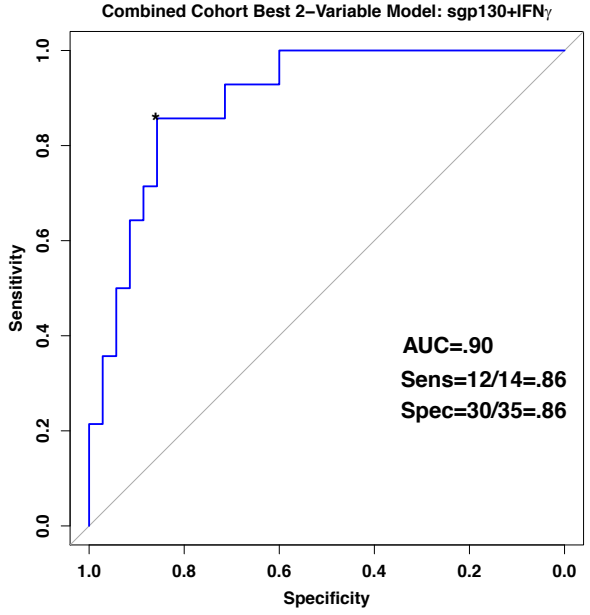
A.



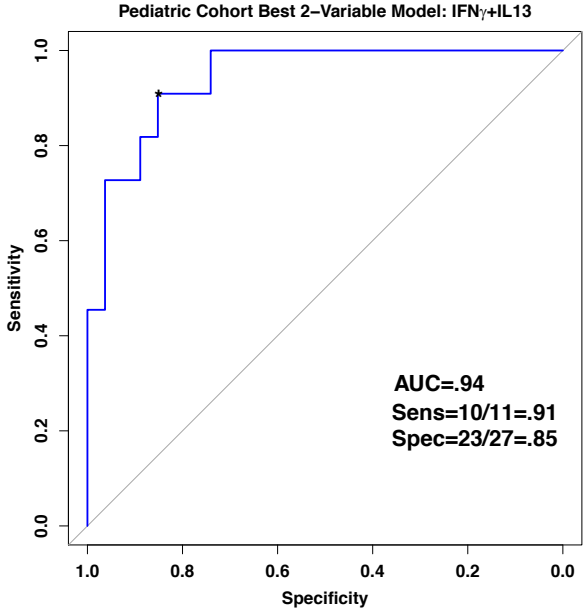
B.



C.

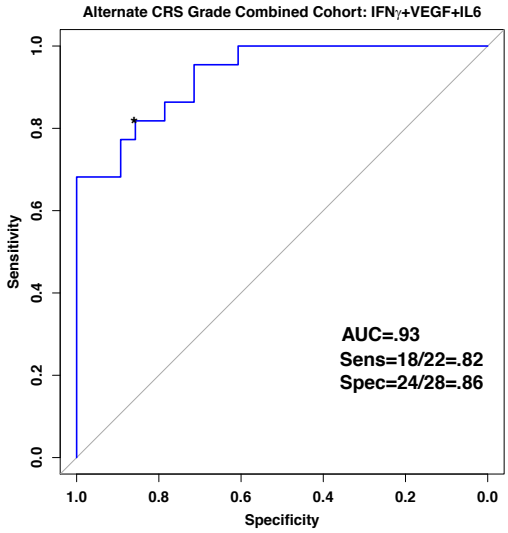


D.

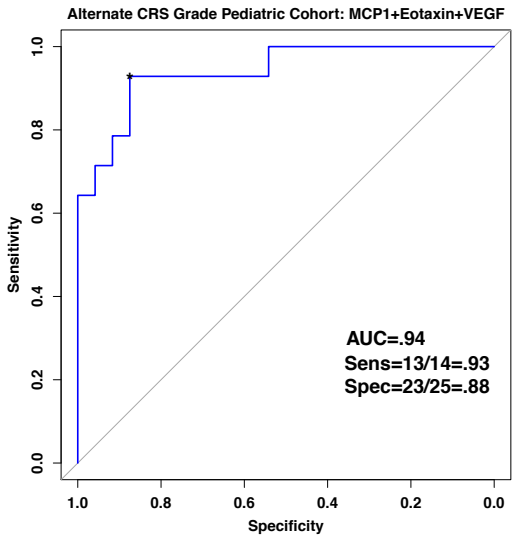


Supplemental Figure 5

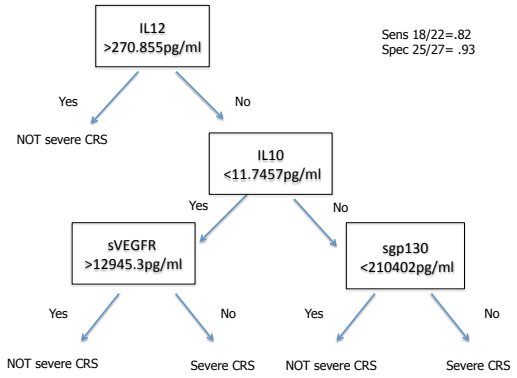
A.



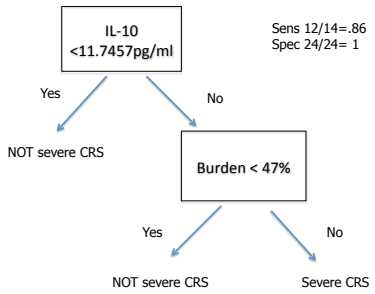
B.



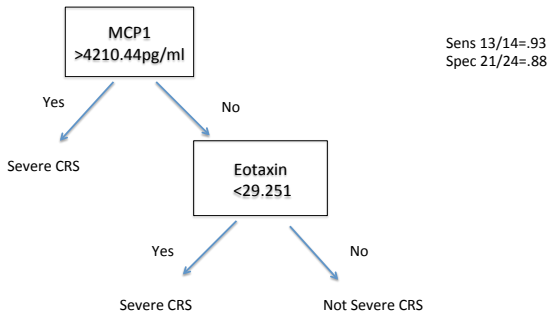
C.



D.



E.



Supplemental References

1. Maude SL, Frey N, Shaw PA, Aplenc R, Barrett DM, Bunin NJ, et al. Chimeric antigen receptor T cells for sustained remissions in leukemia. The New England journal of medicine 2014; 371: 1507-17.
2. Milone MC, Fish JD, Carpenito C, Carroll RG, Binder GK, Teachey D, et al. Chimeric receptors containing CD137 signal transduction domains mediate enhanced survival of T cells and increased antileukemic efficacy in vivo. Molecular therapy : the journal of the American Society of Gene Therapy 2009; 17: 1453-64.
3. Levine BL, Humeau LM, Boyer J, MacGregor RR, Rebello T, Lu X, et al. Gene transfer in humans using a conditionally replicating lentiviral vector. Proceedings of the National Academy of Sciences of the United States of America 2006; 103: 17372-7.
4. Olson WC, Smolkin ME, Farris EM, Fink RJ, Czarkowski AR, Fink JH, et al. Shipping blood to a central laboratory in multicenter clinical trials: effect of ambient temperature on specimen temperature, and effects of temperature on mononuclear cell yield, viability and immunologic function. J Transl Med 2011; 9: 26.
5. Kalos M, Levine BL, Porter DL, Katz S, Grupp SA, Bagg A, et al. T cells with chimeric antigen receptors have potent antitumor effects and can establish memory in patients with advanced leukemia. Sci Transl Med 2011; 3: 95ra73.
6. Borowitz MJ, Devidas M, Hunger SP, Bowman WP, Carroll AJ, Carroll WL, et al. Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. Blood 2008; 111: 5477-85.
7. Weir EG, Cowan K, LeBeau P, Borowitz MJ. A limited antibody panel can distinguish B-precursor acute lymphoblastic leukemia from normal B precursors with four color flow cytometry: implications for residual disease detection. Leukemia 1999; 13: 558-67.
8. Janetzki S, Britten CM, Kalos M, Levitsky HI, Maecker HT, Melief CJ, et al. "MIATA"-minimal information about T cell assays. Immunity 2009; 31: 527-8.
9. Risma K, Jordan MB. Hemophagocytic lymphohistiocytosis: updates and evolving concepts. Curr Opin Pediatr 2012; 24: 9-15.
10. Lee DW, Kochenderfer JN, Stetler-Stevenson M, Cui YK, Delbrook C, Feldman SA, 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 2015; 385: 517-28.
11. Grupp SA, Kalos M, Barrett D, Aplenc R, Porter DL, Rheingold SR, et al. Chimeric antigen receptor-modified T cells for acute lymphoid leukemia. The New England journal of medicine 2013; 368: 1509-18.
12. Davila ML, Riviere I, Wang X, Bartido S, Park J, Curran K, et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia. Sci Transl Med 2014; 6: 224ra25.
13. Portell CA, Wenzell CM, Advani AS. Clinical and pharmacologic aspects of blinatumomab in the treatment of B-cell acute lymphoblastic leukemia. Clin Pharmacol 2013; 5: 5-11.
14. Lee DW, Gardner R, Porter DL, Louis CU, Ahmed N, Jensen M, et al. Current concepts in the diagnosis and management of cytokine release syndrome. Blood 2014; 124: 188-95.
15. Bruunsgaard H, Pedersen M, Pedersen BK. Aging and proinflammatory cytokines. Current opinion in hematology 2001; 8: 131-6.
16. Shurin GV, Yurkovetsky ZR, Chatta GS, Tourkova IL, Shurin MR, Lokshin AE. Dynamic alteration of soluble serum biomarkers in healthy aging. Cytokine 2007; 39: 123-9.

17. Krabbe KS, Pedersen M, Bruunsgaard H. Inflammatory mediators in the elderly. Exp Gerontol 2004; 39: 687-99.
18. Fulop T, Larbi A, Douziech N, Levesque I, Varin A, Herbein G. Cytokine receptor signalling and aging. Mech Ageing Dev 2006; 127: 526-37.
19. Berdat PA, Wehrle TJ, Kung A, Achermann F, Sutter M, Carrel TP, et al. Age-specific analysis of normal cytokine levels in healthy infants. Clin Chem Lab Med 2003; 41: 1335-9.
20. Porter DL, Hwang WT, Frey NV, Lacey SF, Shaw PA, Loren AW, et al. Chimeric antigen receptor T cells persist and induce sustained remissions in relapsed refractory chronic lymphocytic leukemia. Sci Transl Med 2015; 7: 303ra139.
21. Henter JI, Horne A, Arico M, Egeler RM, Filipovich AH, Imashuku S, et al. HLH-2004: Diagnostic and therapeutic guidelines for hemophagocytic lymphohistiocytosis. Pediatric blood & cancer 2007; 48: 124-31.