RESEARCH ARTICLE

Predicting T-cell quality during manufacturing through an artificial intelligence-based integrative multiomics analytical platform

Valerie Y. Odeh-Couvertier[1] | Nathan J. Dwarshuis[2] | Maxwell B. Colonna[3] |
Bruce L. Levine[4] | Arthur S. Edison[3] | Theresa Kotanchek[5] | Krishnendu Roy[2] |
Wandaliz Torres-Garcia[1] [ID]

[1]Department of Industrial Engineering, University of Puerto Rico Mayagüez, Mayagüez, Puerto Rico, USA

[2]The Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA

[3]Departments of Genetics and Biochemistry & Molecular Biology, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia, USA

[4]Center for Cellular Immunotherapies, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[5]Evolved Analytics LLC, Rancho Santa Fe, California, USA

Correspondence
Wandaliz Torres-Garcia, Department of Industrial Engineering, University of Puerto Rico Mayagüez, Mayagüez, PR 00681, USA.
Email: wandaliz.torres@upr.edu

Abstract

Large-scale, reproducible manufacturing of therapeutic cells with consistently high quality is vital for translation to clinically effective and widely accessible cell therapies. However, the biological and logistical complexity of manufacturing a living product, including challenges associated with their inherent variability and uncertainties of process parameters, currently make it difficult to achieve predictable cell-product quality. Using a degradable microscaffold-based T-cell process, we developed an artificial intelligence (AI)-driven experimental-computational platform to identify a set of critical process parameters and critical quality attributes from heterogeneous, high-dimensional, time-dependent multiomics data, measurable during early stages of manufacturing and predictive of end-of-manufacturing product quality. Sequential, design-of-experiment-based studies, coupled with an agnostic machine-learning framework, were used to extract feature combinations from early in-culture media assessment that were highly predictive of the end-product CD4/CD8 ratio and total live CD4$^+$ and CD8$^+$ naïve and central memory T cells (CD63L$^+$CCR7$^+$). Our results demonstrate a broadly applicable platform tool to predict end-product quality and composition from early time point in-process measurements during therapeutic cell manufacturing.

KEYWORDS
artificial intelligence, bioprocess optimization, cell therapy manufacturing, cytokines, design of experiments, metabolomics, T-cell memory

1 | INTRODUCTION

T-cell-based immunotherapies have received great interest from clinicians and industry due to their potential to treat, and often functionally cure some hematological cancers and their potential applicability in many other diseases.[1,2] Since 2017, four genetically modified

autologous Chimeric Antigen Receptor (CAR) T-cell therapies (*Yescarta*™, *Kymriah*™, *Tecartus*™, and *Breyanzi*®) have received approval from the U.S. Food and Drug Administration to treat certain B-cell malignancies. Despite these successes, T-cell-based immuno-therapies are constrained by time-intensive, high cost, complex manufacturing processes that are time-intensive, expensive, and difficult to scale[3,4] and lack methods and tools to predict the end-product quality during manufacturing. Quality assessment is performed only at the end of manufacturing which takes many days. Identification of early putative critical quality attributes (CQAs) and the associated critical process parameters (CPPs) that can be measured nondestructively during culture and can predict end-product attributes early in the manufacturing timeline could be transformative for the cell therapy field.

Translating laboratory-scale T-cell expansion experiments into a large-scale manufacturing process is hindered by the incomplete understanding of cell properties and how they are affected by process variables, lack of detailed characterization, and high variability of materials during manufacturing.[5] These challenges of manufacturing a "living product" are further magnified since current chemistry, manufacturing, and control, analytics, regulations, and product specifications are designed for conventional chemical and biopharmaceutical manufacturing systems.[6] This underscores the need to develop innovative tools, methods, and standards to ensure appropriate quality controls, and new strategies involving quality by design and good manufacturing practices for cell-based therapies.[7–9] The intricate manufacturing process for T cells and other cell therapies must be deeply assessed and appropriately controlled to ensure scalability, predictability, and a high-quality manufacturing process at the most reasonable cost. A key step for reaching this goal is to identify putative CQAs and CPPs early in the manufacturing process that can predict the quality of the manufactured cell-therapy product. We hypothesized that rigorous characterization of process parameters along with longitudinal measurements of cell-secreted cytokine, chemokine, and metabolites from the culture media early during manufacturing will allow us to develop an artificial intelligence (AI)-based mathematical-computational framework for the identification of multivariate parameters that are predictive of the end-of-manufacturing product phenotypes.

Characterization studies of approved autologous anti-CD19 CAR-T cell therapies have recently revealed initial sets of candidate quality attributes, that is, percent transduction, vector copy number, and interferon-γ production for axicabtagene ciloleucel (Yescarta™),[10] while CAR expression and release of interferon-γ are a few of those identified for tisagenlecleucel (Kymriah™).[11] Many of these attributes are calculated as endpoint responses and thus a deeper understanding of the cell growth process impacted by starting conditions and performance during their manufacturing is essential. Hence, CQAs that enable early monitoring through real-time process measurements such as multiomics cell characterization can overcome current challenges in assessing product consistency. Yet, the computational complexity of dealing with the heterogeneity and multivariate nature of multiomics measurements to characterize T-cell quality, that is, high-definition phenotyping of naïve and memory subsets, remains a challenge.
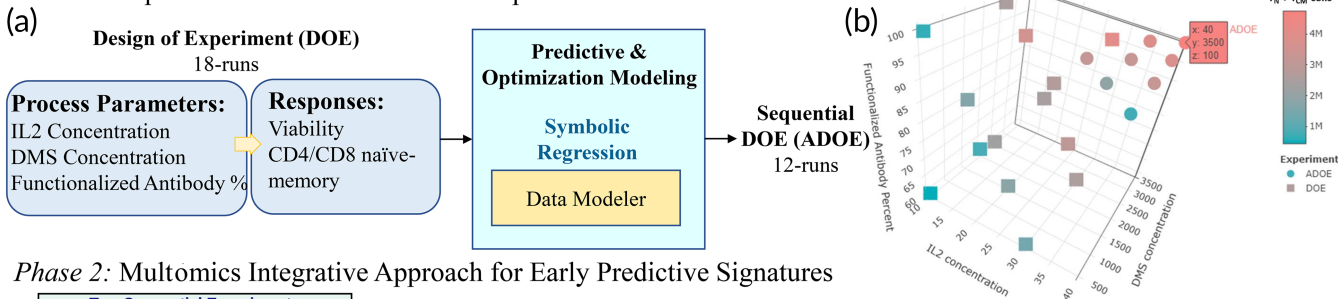
Generally, T cells with a lower differentiation state such as naïve and stem cell or central memory cells have been shown to provide superior anti-tumor potency, presumably due to their higher potential to replicate, migrate, and engraft, leading to a long-term, durable response.[12–15] Likewise, CD4 T cells are similarly important to anti-tumor potency due to their cytokine release properties and ability to resist exhaustion.[16,17] Our group has developed a novel degradable microscaffold (DMS)-based method using porous microcarriers functionalized with anti-CD3 and anti-CD28 mAbs for use in T-cell expansion cultures. We showed that compared to commercially available microbeads (Miltenyi), DMSs generated a higher number of migratory naïve (T_N) and central memory (T_{CM}) (CCR7$^+$CD62L$^+$) T cells and CD4$^+$ T cells across multiple donors.[18] We used this manufacturing process as an exemplar to develop an experimental-computational AI-based tool to predict product quality from early process measurements. This two-phase approach consists of (1) the optimization of process parameters through experimental designs, and (2) the extraction of early predictive signatures of T-cell quality by multiomics integration using regression models. This agnostic computational approach provides a platform to discover early predictive CQAs and CPPs to ensure consistent product quality that can be widely applicable for other cellular therapies.

2 | RESULTS

2.1 | Overall multiomics study design

T cells were expanded ex vivo for 14 days and 100 μl of supernatant media samples were collected at days 4, 6, 8, 11, and 14 to measure cytokine profiles and perform nuclear magnetic resonance (NMR) analysis. Endpoint responses on DMS-based T-cell extracts were measured for different combinations of DMS parameters: IL2 concentration, DMS concentration, and functionalized antibody percent. Two experimental regions were determined using a design-of-experiments (DOE) methodology to maximize the yields of CD62L$^+$CCR7$^+$ cells (i.e., naïve and central memory T cells, $T_N + T_{CM}$) as a function of these process parameters. The first DOE resulted in a randomized 18-run I-optimal custom design where each DMS parameter was evaluated at three levels. To further optimize this DOE in terms of total live CD4$^+$ $T_N + T_{CM}$ cells, a sequential adaptive design-of-experiment (ADOE) was designed with 12 additional samples (Figure 1b). All 30 runs from both experiments (DOE, ADOE) were molecularly characterized to model total live $T_N + T_{CM}$ (a) CD4$^+$, (b) CD8$^+$, and (c) their ratio (Supporting Figure S1). The extraction of early predictive CPPs and CQAs for the expansion of $T_N + T_{CM}$ cells during ex vivo culture was performed in two phases: (1) optimization of process parameters and (2) integration of multiomics for predictive modeling (Figure 1).

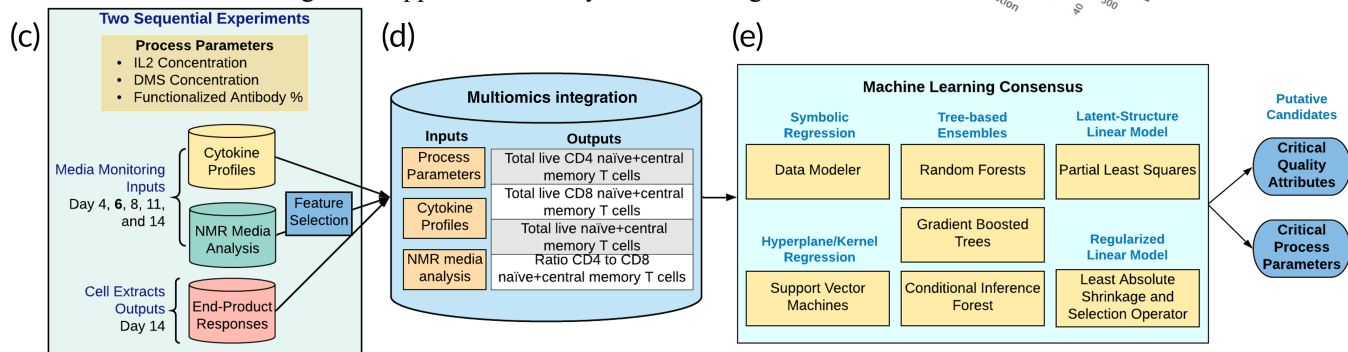*Phase 1:* Experimental Process Parameters Optimization

FIGURE 1   Two-phase approach to extract early predictive critical process parameters (CPPs) and critical quality attributes (CQAs) for CD4$^+$/CD8$^+$ T$_N$ + T$_{CM}$ cells. (a) Design-of-experiment (DOE) modeling and optimization of process parameters. (b) Experimental region studied and optimized for total live CD4$^+$ T$_N$ + T$_{CM}$ cells. (c) Total live CD4$^+$ T$_N$ + T$_{CM}$ cells across the overall study design (two experiments varying process parameters). (d) Integrative multiomics approach through (e) a machine learning consensus analysis to identify early predictive CPPs and CQAs putative candidates for both total live CD4$^+$ and CD8$^+$ T$_N$ + T$_{CM}$ cells

TABLE 1   LOO-R^2 prediction performance results for all machine learning (ML) models when evaluating process parameters, and features from cytokine and nuclear magnetic resonance (NMR) media analysis at day 6 or day 4

LOO-R^2	ML						
Response/predictors	SR	RF	GBM	CIF	LASSO	PLSR	SVM
Ratio of CD4 to CD8 T$_N$ + T$_{CM}$ cells							
PP + N4	**99%**	86.8%	96.3%	**84.5%**	88.6%	92.5%	88.5%
PP + N6	**99%**	73.6%	95.9%	70.1%	81.0%	95.8%	79.7%
PP + S6	**99%**	**87.1%**	**99.9%**	83.4%	87.2%	97.9%	86.8%
PP + S6 + N6	**99%**	85.5%	95.3%	83.4%	**92.9%**	**99.7%**	**90.5%**
Total live CD4$^+$ T$_N$ + T$_{CM}$ cells							
PP + N4	97%	67.0%	93.6%	69.3%	34.3%	90.1%	75.5%
PP + N6	96%	45.9%	92.6%	51.2%	42.8%	**92.1%**	**79.4%**
PP + S6	**98%**	**71.4%**	**99.9%**	**75.0%**	**74.9%**	80.0%	75.5%
PP + S6 + N6	**98%**	68.2%	95.6%	74.4%	72.5%	81.7%	77.0%
Total live CD8$^+$ T$_N$ + T$_{CM}$ cells							
PP + N4	93%	4.7%	**44.4%**	9.2%	1.2%	65.1%	9.1%
PP + N6	86%	2.0%	29.9%	**15.8%**	28.5%	63.3%	30.6%
PP + S6	**93%**	7.8%	28.0%	15.1%	**76.2%**	98.4%	**49.8%**
PP + S6 + N6	**93%**	0.3%	32.7%	9.8%	51.5%	96.4%	37.8%

Notes: ML models' prediction performance is measured as the leave-one-out cross-validated R^2 (LOO-R^2) while SR prediction performance is measured as R^2 of the ensemble prediction where the ensemble is composed of diverse models with complexity constrained. Predictors evaluated: (PP) Process parameters, (N) NMR, (S) Cytokines measured at day 4 or 6. Maximum R^2 within each ML method are shown in bold.

2.2 | Optimization of $T_N + T_{CM}$ cells as a function of process parameters

Using symbolic regression (Data Modeler software from Evolved Analytics LLC), we examined the interactive effects of the DMS parameters on yield to simultaneously predict and optimize both CD4$^+$ and CD8$^+$ $T_N + T_{CM}$. A model ensemble predicted 4.2×10^6 CD4$^+$ $T_N + T_{CM}$ cells at an optimum setting of 30 U/μl IL2, 2500 carriers/μl, and 100% functionalized mAbs (Supporting Figure S2). This result was consistent with the observed maximum value of 4.0×10^6, highlighting that CD4$^+$ $T_N + T_{CM}$ yield was maximized at high levels of DMS parameters (Figure 1b). In contrast, the predicted optimum yield for CD8$^+$ $T_N + T_{CM}$ was 1.9×10^7 cells at a setting of 30 U/μl IL2,

600 carriers/μl, and 100% functionalized mAbs (data not shown). Although this combination was not experimentally tested, the closest measured record (30 U/μl IL2, 500 carriers/μl, 100% functionalized mAbs) achieved the predicted maximum yield. Hence, the CD8$^+$ $T_N + T_{CM}$ yield was maximized at high IL2 concentration and functionalized mAbs percentage but low DMS concentration.

The DOE analysis highlighted the potential for further optimization of total live CD4$^+$ $T_N + T_{CM}$ cells, as well as the potential to optimize the CD4$^+$ to CD8$^+$ $T_N + T_{CM}$ cells ratio, at DMS levels greater than those originally evaluated (DOE). Therefore, to test and validate, a second adaptive design of experiment (ADOE) was designed to maximize the total live CD4$^+$ $T_N + T_{CM}$ cells. We expanded the parameter range, assessing IL2 concentration >30 U/μl and DMS concentration
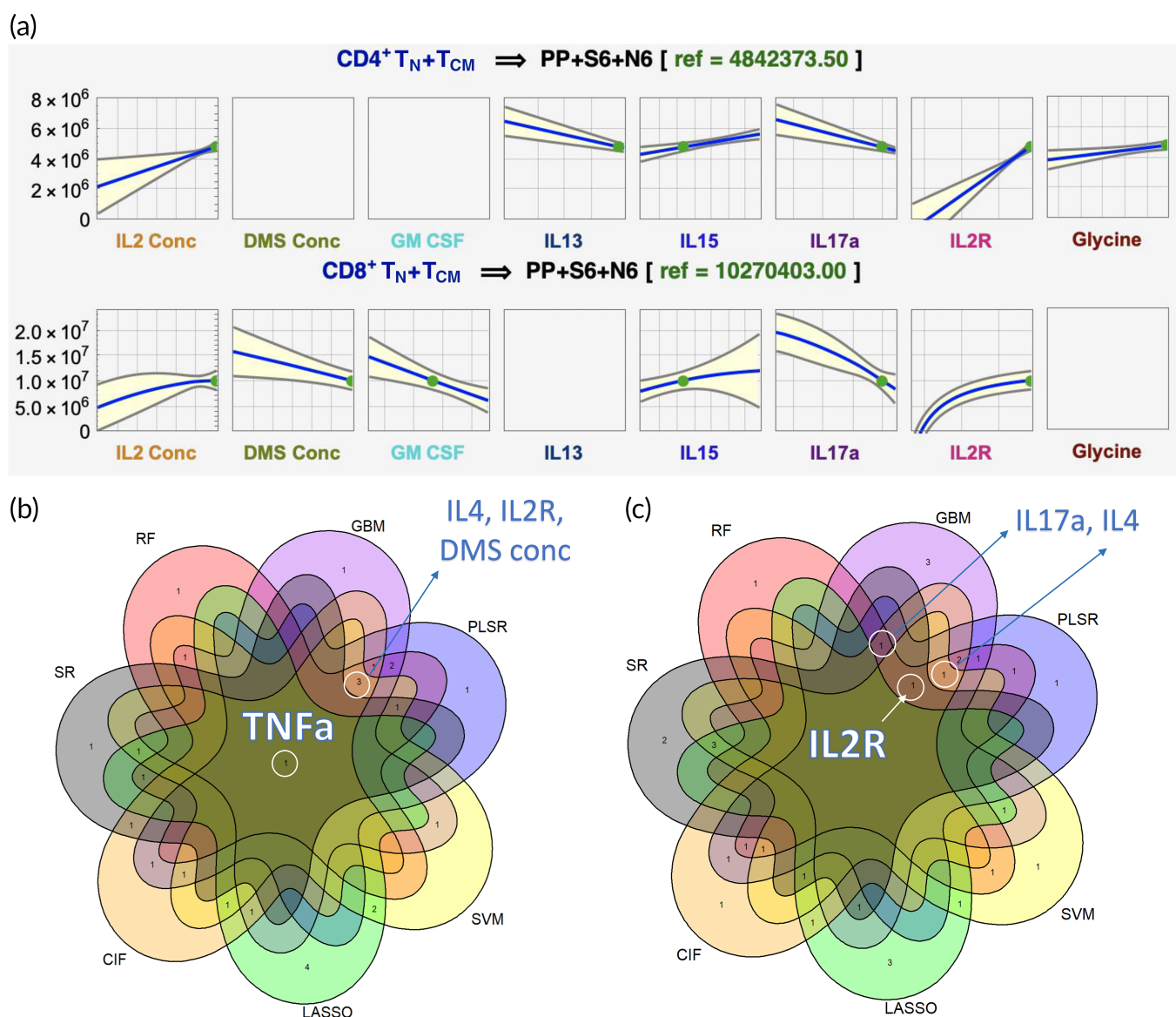

FIGURE 2 Multiomics culturing media prediction profiles of highly predictive features for early monitoring of T-cell manufacturing. (a) Prediction model profiles from day 6 cultured media monitoring where total live CD4$^+$ $T_N + T_{CM}$ is maximized. (b) Machine learning (ML) models consensus for ratio CD4$^+$ to CD8$^+$ $T_N + T_{CM}$ cells, and (c) ML models consensus for total live CD4$^+$ $T_N + T_{CM}$ cells. Feature names are shown for consensus with 5 or more ML models at the highest-ranking standing (see the Materials and Methods section)
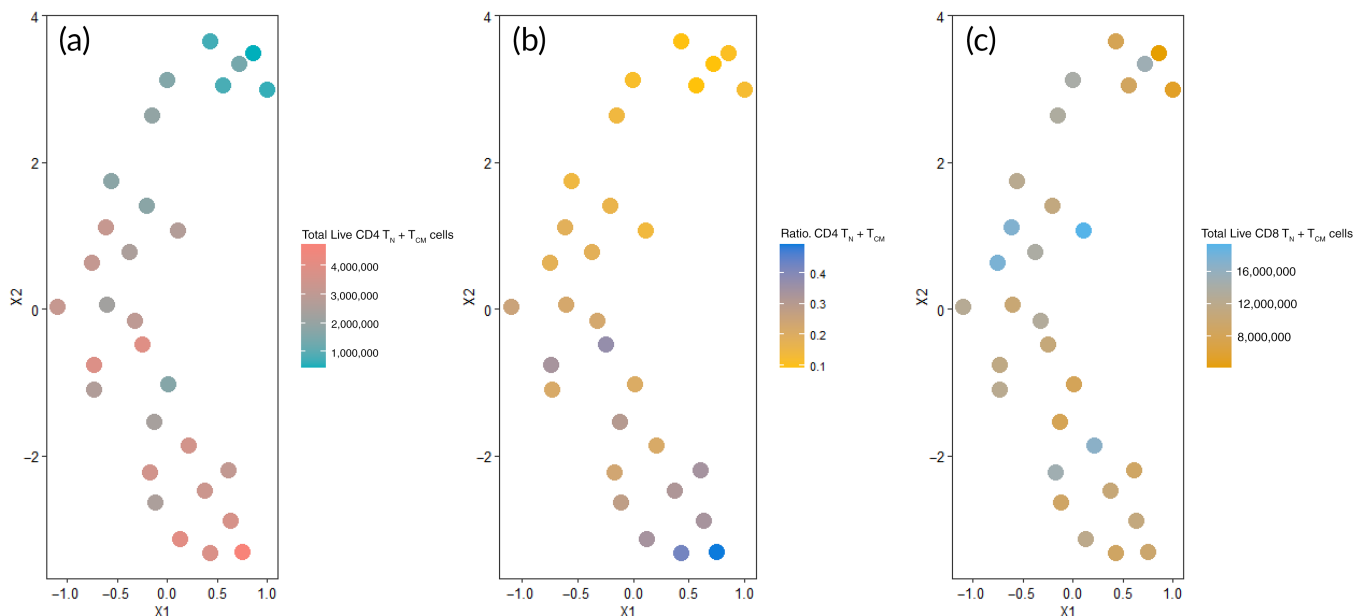
FIGURE 3 Uniform manifold approximation and projection (UMAP) clustering in 2D (X1, X2) of T-cell samples from early predictive from nuclear magnetic resonance (NMR) and cytokine media features at day 6 of T-cell culturing (formate, lactate, histidine, ethanol, dimethylamine, branch chain amino acids (BCAAs), glucose, glutamine, TNFα, IL2R, IL4, IL17a, IL13, IL15, and GM-CSF): for (a) ratio CD4$^+$ to CD8$^+$ T$_N$ + T$_{CM}$, (b) total live CD4$^+$ T$_N$ + T$_{CM}$ cells, and (c) total live CD8$^+$ T$_N$ + T$_{CM}$ cells

>2500 carriers/μl (Figure 1b). CD4$^+$ T$_N$ + T$_{CM}$ and its ratio to CD8$^+$ T$_N$+ T$_{CM}$, 4.7 × 10^6 cell and 0.49 respectively, were maximized when IL2 concentration (40 U/μl) and DMS concentration (3500 carriers/μl) were maximized (Figure 1b; Supporting Table S1; Figure S2). Utilizing the ADOE data set, new response ensembles were generated enabling more robust prediction over the expanded parameter space (↑IL2 and ↑DMS concentrations).

2.3 | Multiomic integrative analysis for early monitoring of T-cell manufacturing

Due to the heterogeneity of the multivariate data collected and knowing that no single model structure is perfect for all applications, we implemented an agnostic modeling approach to better understand these T$_N$ + T$_{CM}$ responses. To achieve this, a consensus analysis using seven machine learning (ML) techniques, random forest (RF), gradient boosted machine (GBM), conditional inference forest (CIF), least absolute shrinkage and selection operator (LASSO), partial least-squares regression (PLSR), support vector machine (SVM), and data modeler's symbolic regression (SR), was implemented to molecularly characterize T$_N$ + T$_{CM}$ cells and to extract predictive features of quality early on their expansion process (Figure 1d,e).

SR models achieved the highest predictive performance (R^2 > 93%) when using multiomics predictors for all endpoint responses (Table 1). SR achieved R^2 > 98%, while GBM tree-based ensembles showed leave-one-out cross-validated R^2 (LOO-R^2) >95% for CD4$^+$ and CD4$^+$/CD8$^+$ T$_N$ + T$_{CM}$ responses. Similarly, LASSO, PLSR, and SVM methods showed consistent high LOO-R^2, 92.9%,

99.7%, and 90.5%, respectively, to predict the CD4$^+$/CD8$^+$ T$_N$ + T$_{CM}$. Yet, about 10% reduction in LOO-R^2, 72.5%–81.7%, was observed for CD4$^+$ T$_N$ + T$_{CM}$ with these three methods. Lastly, SR and PLSR achieved R^2 > 90% while other ML methods exhibited exceedingly variable LOO-R^2 (0.3%, RF-51.5%, LASSO) for CD8$^+$ T$_N$ + T$_{CM}$ cells. The top-performing technique, SR, showed that the median aggregated predictions for total live CD4$^+$ and CD8$^+$ T$_N$ + T$_{CM}$ cells increases when IL2 concentration, IL15, and IL2R increase, while IL17a decreases in conjunction with other interactive features. These patterns combined with low values of DMS concentration and GM_CSF uniquely characterized maximum CD8$^+$ T$_N$ + T$_{CM}$. Meanwhile, higher glycine but lower IL13 in combination with others showed maximum CD4$^+$ T$_N$ + T$_{CM}$ predictions (Figure 2a).

Selecting CPPs and CQAs candidates consistently for T-cell memory across different models is desired. Here, TNFα was found in consensus across all seven ML methods for predicting CD4$^+$/CD8$^+$ T$_N$ + T$_{CM}$ when considering features with the highest importance scores across models (Figure 2b; Materials and Methods section). Other features, IL2R, IL4, IL17a, and DMS concentration, were commonly selected in ≥5 ML methods (Figure 2b,c). Moreover, IL13 and IL15 were found predictive in combination with these using SR (Table S2).

This integrative analysis of cytokine and NMR media analysis monitored at early stages of the T-cell process provided highly predictive feature combinations of end-product quality particularly for total live T$_N$ + T$_{CM}$ CD4$^+$ cells and CD4$^+$/CD8$^+$ ratio as shown in Figure 3a,b. However, when translating a real-time monitoring strategy to a large-scale manufacturing process, measuring both cytokine and NMR features from media can be difficult and expensive. To be
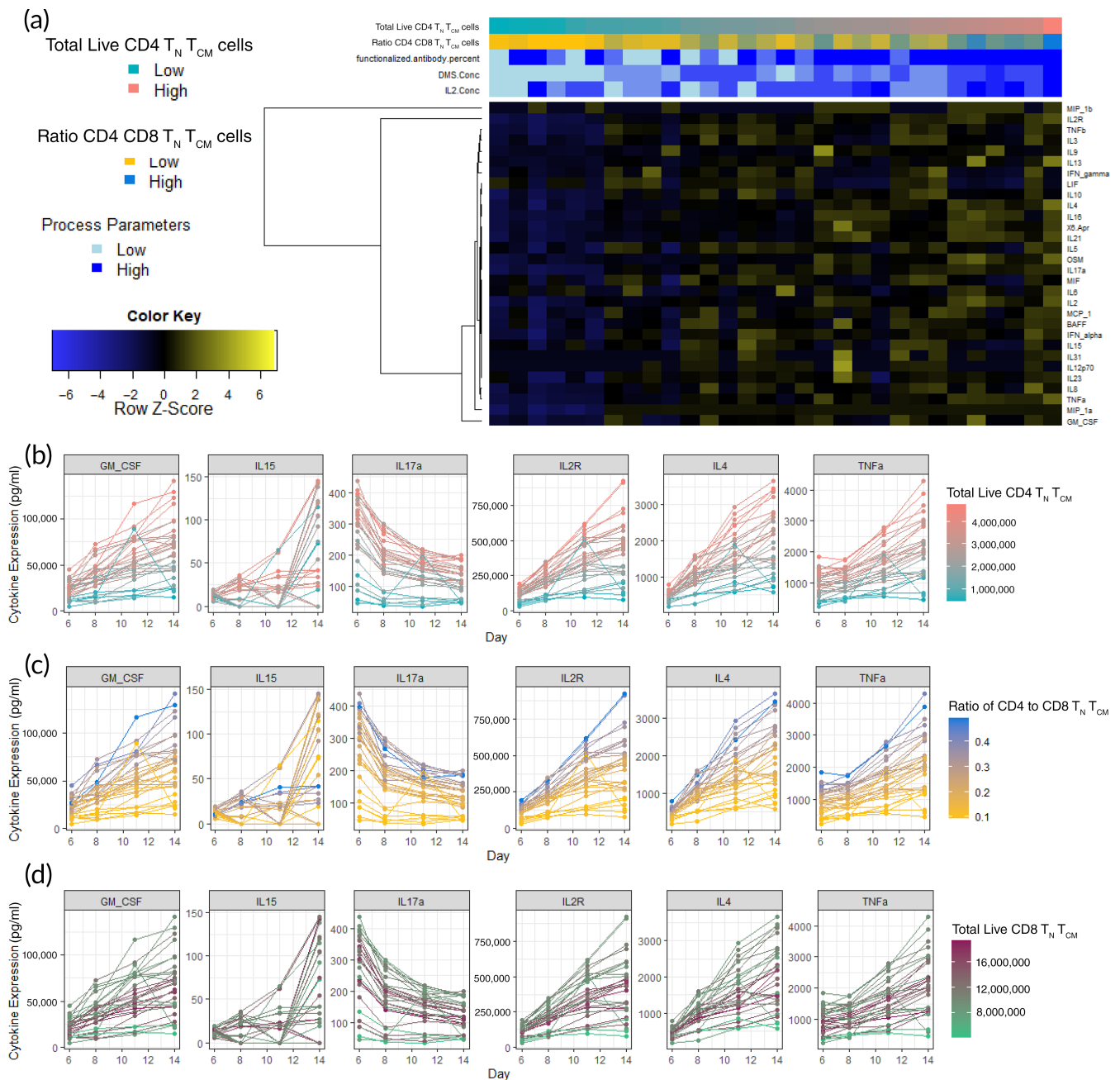
FIGURE 4 General characteristics of cytokine media profiles. (a) Heatmap for cytokine profiles from media samples on day 6. Expression in picograms/milliliter across time points for relevant cytokine features for (b) ratio CD4$^+$ to CD8$^+$ T$_N$ + T$_{CM}$ cells, (c) total live CD4$^+$ T$_N$ + T$_{CM}$ cells, and (d) total live CD8$^+$ T$_N$ + T$_{CM}$ cells

cost-efficient and translatable, we demonstrated that either cytokine profiles or NMR media analysis alone is sufficient to find predictive features without compromising prediction performance.

2.4 | Single-omics media analysis for early prediction

ML models using solely media cytokine profiles at day 6 reached similar or higher R^2 than those of the multiomics models (CD4$^+$ T$_N$ + T$_{CM}$:

71.4%–99.9%; CD4$^+$/CD8$^+$: 83.4%–99.7%). However, CD8$^+$ T$_N$ + T$_{CM}$ still had variable LOO-R^2, 7.8%–93%. Overall, higher cytokine media profiles showed higher CD4$^+$ T$_N$ + T$_{CM}$ and consequently its ratio with CD8$^+$ (Figure 4a). This behavior was evident, even beyond day 6, for TNFα, IL2R, IL17a, and IL4 which were frequently selected as predictive features across models (Figures 4b,c and S3g–i). A more complex behavior was detected for CD8$^+$ T$_N$ + T$_{CM}$ which cannot be explained by cytokine secretion alone (Figure 4d).

Models using only NMR media intensities on day 6 revealed an R^2 decrease of 8.8% and 11.1%, on average, compared with the
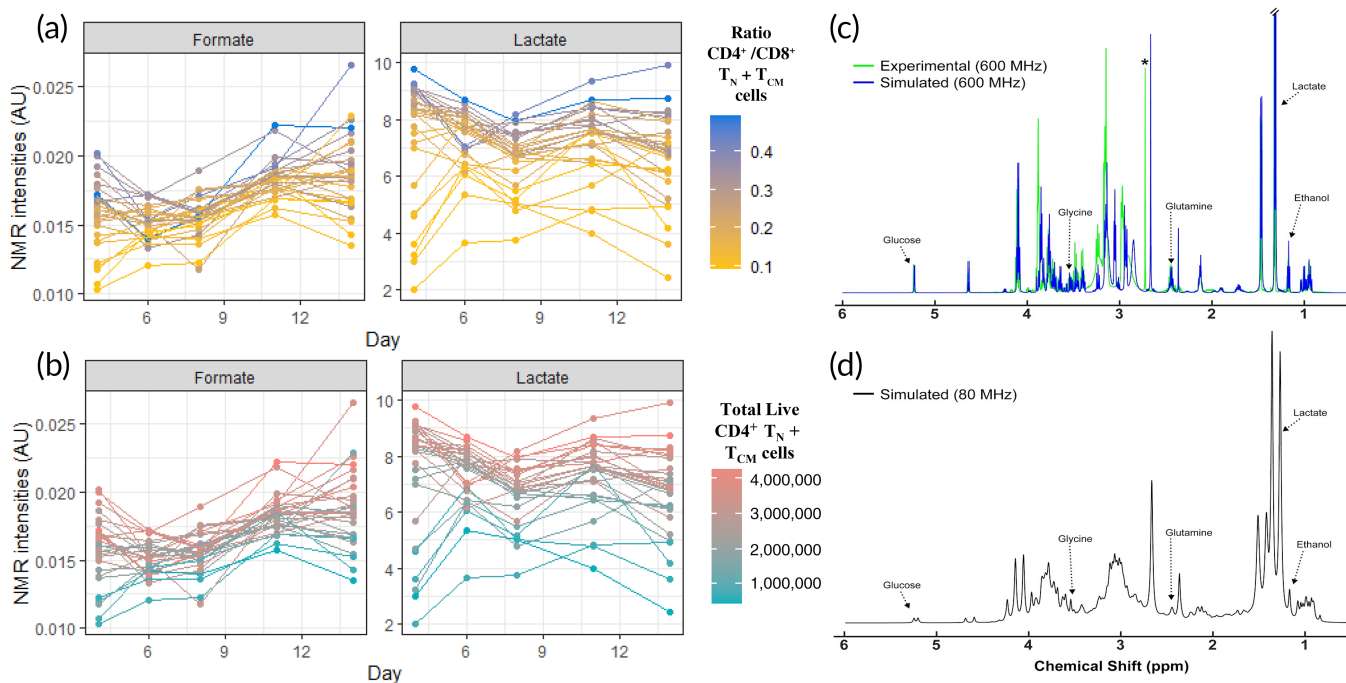
FIGURE 5 Top-performing features nuclear magnetic resonance (NMR) media analysis. NMR intensities in arbitrary units (AU) across time points for (a) ratio CD4$^+$/CD8$^+$ T$_N$ + T$_{CM}$ cells, and (b) total live CD4$^+$ T$_N$ + T$_{CM}$ cells. (c) Simulation of ^{1}H NMR spectrum shows the potential to detect multiple predictive features at lower magnetic fields. Overlay of a pooled experimental spectrum of T-cell culture medium (green) and GISSMO[19,20] simulated spectrum (blue), composed of 19 compounds that reasonably approximate the experimental spectrum acquired at 600 MHz. Asterisk indicates an unknown feature of high intensity that was simulated with 2,3-dimethylamine (blue feature to right). Annotated features in the spectrum correspond to those identified as being highly predictive of output responses across computational methods. (d) GISSMO[19,20] simulated spectrum at 80 MHz, corresponding to a field strength of commercially available benchtop NMR systems

multiomics and cytokine models, respectively. Yet, SR, GBM, and PLSR reached high LOO-R^2 (92.1%–99%), specifically for CD4$^+$/CD8$^+$ and CD4$^+$ T$_N$ + T$_{CM}$. Although good prediction was achieved with NMR media analysis on day 6, we obtain slightly better predictions with NMR media analysis on day 4 (Table 1). From these models, formate, lactate, DMS concentration were highly ranked to predict both, ratio CD4$^+$/CD8$^+$ and CD4$^+$ T$_N$ + T$_{CM}$ (Figure S3a–f). Some variable combinations also contained histidine, ethanol, dimethylamine, branch chain amino acids (BCAAs), glucose, and glutamine (Table S3). Lower intensity values for BCAAs, dimethylamine, glucose, and glutamine displayed higher CD4$^+$ T$_N$ + T$_{CM}$ cells across the different media monitoring times (Figure S5a). Inversely, higher intensities of formate and lactate showed higher CD4$^+$ T$_N$ + T$_{CM}$ and its ratio with CD8$^+$ consistently across time (Figure 5a,b).

The initial screening of a few samples from a different experimental batch shows much lower values of T$_N$ + T$_{CM}$ responses but maintains a similar NMR and cytokine media patterns as the DOE and ADOE experiments (lower value intensities/secretion, lower T$_N$ + T$_{CM}$ response) in terms of the total live T$_N$ + T$_{CM}$ cells for CD4$^+$ and CD8$^+$. However, the decay in total live T$_N$ + T$_{CM}$ cells for CD8$^+$ is much rapid than CD4$^+$ which makes the ratio behave in a more complex behavior (Figures S7 and S8).

3 | DISCUSSION

CPP's understanding is critical to new product development and, especially in cell therapy development, it can have life-saving implications. The challenges for effective modeling grow with the increasing complexity of processes due to high dimensionality, and the potential for process interactions and nonlinear relationships. Another critical challenge is the limited amount of available data, mostly small DOE data sets. SR has the necessary capabilities to resolve the issues of process effects modeling and has been applied across multiple industries.[21] SR discovers mathematical expressions that fit a given sample and differs from conventional regression techniques in that a model structure is not defined a priori.[22] Hence, a key advantage of this methodology is that transparent, human-interpretable models can be generated from small and large data sets with no prior assumptions.[23,24]

Since the model search process lets the data determine the model, diverse and competitive (e.g., accuracy and complexity) model structures are typically discovered. An ensemble of diverse models can be formed where its constituent models will tend to agree when constrained by observed data yet diverge in new regions. Collecting data in these regions helps to ensure that the target system is accurately modeled, and its optimum is accurately located.[23,24] Exploiting

these features allows adaptive data collection and interactive modeling. Consequently, this adaptive-DOE approach is useful in a variety of scenarios, including maximizing model validity for model-based decision making, optimizing processing parameters to maximize target yields, and developing emulators for online optimization and human understanding.[23,24]

An in-depth characterization of potential DMS-based T-cell CQAs includes a list of cytokine and NMR features from media samples that are crucial in many aspects of T-cell fate decisions and effector functions of immune cells. Cytokine features were observed to slightly improve prediction and dominated the ranking of important features and variable combinations when modeling together with NMR media analysis and process parameters (Figure 2a,b). Predictive cytokine features such as TNFα, IL2R, IL4, IL17a, IL13, and IL15 were biologically assessed in terms of their known functions and activities associated with T cells. T helper cells secrete more cytokines than T cytotoxic cells, as per their main functions, and activated T cells secrete more cytokines than resting T cells. It is possible that some cytokines simply reflect the CD4$^+$/CD8$^+$ ratio and the activation degree by proxy proliferation. However, the exact ratio of expected cytokine abundance is less clear and depends on the subtypes present, and thus examination of each relevant cytokine is needed.

IL2R is secreted by activated T cells and binds to IL2, acting as a sink to dampen its effect on T cells.[25] Since IL2R was much greater than IL2 in solution, this might reduce the overall effect of IL2, which could be further investigated by blocking IL2R with an antibody. In T cells, TNF can increase IL2R, proliferation, and cytokine production.[25] It may also induce apoptosis depending on concentration and alter the CD4$^+$ to CD8$^+$ ratio.[26] Given that TNF has both a soluble and membrane-bound form, this may either increase or decrease CD4$^+$ ratio and/or memory T cells depending on the ratio of the membrane to soluble TNF.[27] Since only soluble TNF was measured, membrane TNF is needed to understand its impact on both CD4$^+$ ratio and memory T cells. Furthermore, IL13 is known to be critical for Th2 response and therefore could be secreted if there are significant Th2 T cells already present in the starting population.[28] This cytokine has limited signaling in T cells and is thought to be more of an effector than a differentiation cytokine.[29] This feature might be emerging as relevant due to an initially large number of Th2 cells or because Th2 cells were preferentially expanded; indeed, IL4 is the conical cytokine that induces Th2 cell differentiation and was observed to be an important variable (Figure 2b,c). The role of these cytokines could be investigated by quantifying the Th1/2/17 subsets both in the starting population and longitudinally. Similar to IL13, IL17 is an effector cytokine produced by Th17 cells[30] thus may reflect the number of Th17 subset of T cells. GM-CSF has been linked with activated T cells, specifically Th17 cells, but it is not clear if this cytokine is inducing differential expansion of CD8$^+$ T cells or if it is simply a covariate with another cytokine inducing this expansion.[31] Finally, IL15 has been shown to be essential for memory signaling and effective in skewing CAR-T cells toward the Tscm phenotype when using membrane-bound IL15Ra and IL15R.[32] Its high predictive behavior goes with its ability to induce large numbers

of memory T cells by functioning in an autocrine/paracrine manner and could be explored by blocking either the cytokine or its receptor.

Similarly, literature suggests that many of the predictive metabolites found here are consistent with metabolic activity associated with T-cell activation and differentiation, yet it is not clear how the various combinations of metabolites relate with each other in a heterogeneous cell population and should be explored. Formate and lactate were found to be highly predictive and observed to positively correlate with higher values of total live CD4$^+$ T$_N$ + T$_{CM}$ cells (Figures 5a,b and S6). Formate is a byproduct of the one-carbon cycle implicated in promoting T-cell activation.[33] Importantly, this cycle occurs between the cytosol and mitochondria of cells and formate excreted.[34] Mitochondrial biogenesis and function have been shown necessary for memory cell persistence.[35,36] Therefore, increased formate in media could be an indicator of one-carbon metabolism and mitochondrial activity in the culture.

In addition to formate, lactate was found as a putative CQA of T$_N$ + T$_{CM}$. Lactate is the end-product of aerobic glycolysis, characteristic of highly proliferating cells and activated T cells.[37,38] Glucose import and glycolytic genes are immediately upregulated in response to T-cell stimulation and thus the generation of lactate. At earlier time points, this abundance suggests a more robust induction of glycolysis and higher overall T-cell proliferation. Interestingly, our models indicate that higher lactate predicts higher CD4$^+$, both in total and in proportion to CD8$^+$, seemingly contrary to previous studies showing that CD8$^+$ T cells rely more on glycolysis for proliferation following activation.[39] It may be that glycolytic cells dominate in the culture at the early time points used for prediction, and higher lactate reflects more cells.

Ethanol patterns are difficult to interpret since its production in mammalian cells is still poorly understood.[40] Fresh media analysis indicates ethanol presence in the media used, possibly utilized as a carrier solvent for certain formula components. However, this does not explain the high variability and trend of ethanol abundance across time (Figure S5). As a volatile chemical, variation could be introduced by sample handling throughout the analysis process. Nonetheless, it is also possible that ethanol excreted into media over time, impacting processes regulating redox and reactive oxygen species which have previously been shown to be crucial in T-cell signaling and differentiation.[41]

Metabolites that consistently decreased over time are consistent with the primary carbon source (glucose) and essential amino acids (BCAA, histidine) that must be continually consumed by proliferating cells. Moreover, the inclusion of glutamine in our predictive models also suggests the importance of other carbon sources for certain T-cell subpopulations. Glutamine can be used for oxidative energy metabolism in T cells without the need for glycolysis.[39] Overall, these results are consistent with existing literature that show different T-cell subtypes require different relative levels of glycolytic and oxidative energy metabolism to sustain the biosynthetic and signaling needs of their respective phenotypes.[42,43] It is worth noting that the trends of metabolite abundance here are potentially confounded by the partial

replacement of media that occurred periodically during expansion (see the Materials and Methods section), thus likely diluting some metabolic byproducts (i.e., formate and lactate) and elevating depleted precursors (i.e., glucose and amino acids). More definitive conclusions of metabolic activity across the expanding cell population can be addressed by a closed system, ideally with online process sensors and controls for formate, lactate, along with ethanol and glucose.

We demonstrated the ability to identify predictive markers using high-magnetic field NMR spectrometers. However, these are expensive, require a significant amount of resources to house and maintain, and would be the unlikely option for routine monitoring in industrial cell-manufacturing. Another common method, liquid chromatography (LC) coupled to mass spectrometry, has the advantage of a relatively smaller footprint and less upfront cost but it has other drawbacks such as destruction of the sample and difficulty with components in culture media that damage LC columns without extraction. Nevertheless, methods like continuous closed-loop sampling are being developed to address this and might be readily available in the future.[44] Recently, permanent magnet-based NMR spectrometers (benchtop-size) have become available at a lower cost. Many of these are readily configured for flow-through reaction monitoring, which can be leveraged in a closed-cell manufacturing process. To explore the feasibility of such system, we utilized a spectral simulation to evaluate if putative CQAs identified here could theoretically be observed and quantified at a magnetic field strength of 80 MHz (common commercial benchtop systems). First, the experimental data acquired at 600 MHz was approximated by creating a simulated mixture of identified metabolites (Figure 5c) and then simulated at 80 MHz (Figure 5d). While the spectral resolution is significantly reduced compared to a spectrum at high-field, there are still numerous features that can be attributed to unique metabolites, including those identified as highly predictive (Figure 5c,d). Although this is promising, there will be challenges to acquiring high-quality data in a closed bioreactor system, that is, cells/DMS-particles present in suspension, final media formulation dictated by the amount of spectral complexity/overlap, and accurate quantitation of features with high overlap from other signals. However, a dedicated benchtop NMR coupled to a bioreactor could provide a simple system for real-time monitoring of CQAs.

4 | CONCLUSIONS

Henceforth, this two-phase approach enabled in-depth characterization and identification of potential CQAs and CPPs for T cells. More sampling is needed to explore aspects like donor-to-donor variability or orthogonal behaviors from failed expansions when available it can be incorporated into this workflow which will be enriched due to its data-driven iterative design that fine-tunes model parameters as more data fit back into it, providing a powerful framework to optimize a complex experimental space during the cell-manufacturing process, and to facilitate the identification of CPPs and early predictive CQAs from multiomics, which can be used broadly in the cell therapy and regenerative medicine field to accurately predict end-of-manufacturing quality at early stages.

The workflow and methods developed here could eventually allow manufacturers to identify deviations and problems with a manufacturing batch early during the culture and potentially implement corrective in-process controls. This could provide a more thorough understanding of the process parameters and their influence on end-product quality, and allow manufacturers to reduce batch failures, and thus improve cost, reduce risk, and increase access to cell-based therapies.

5 | MATERIALS AND METHODS

5.1 | Microcarrier fabrication

DMSs were fabricated as previously described.[18] To vary the surface concentration of the antibodies, the anti-CD3/anti-CD28 mAb mixture was further combined with a biotinylated isotype control to reduce the overall fraction of targeted mAbs. All mAbs were low endotoxin azide-free (Biolegend custom, LEAF specification). The surface concentration of the antibodies was quantified as previously described using a bicinchoninic acid assay kit (Thermo Fisher 23227).[18] See Supplementary Methods.

5.2 | T-cell culture including sample collection

Cryopreserved primary human T cells were obtained as sorted CD3 subpopulations (Astarte Biotech). T cells were activated by adding DMSs (amount specified by the DOE) at day 0 of culture immediately after thaw. DMSs were not added or removed during the culture and had antibodies that were conjugated in proportions specified by the DOE. Initial cell density was 2.0×10^6 cells/ml in a 96-well plate with 300 μl volume. Media was serum-free TexMACS (Miltentyi Biotech 170-076-307) supplemented with recombinant human IL2 in concentrations specified by the DOE (Peprotech 200-02). Cell cultures were expanded for 14 days as counted from the time of initial seeding and activation. Cell counts and viability were assessed using acridine orange/propidium iodide (AO/PI) and a Countess Automated Cell Counter (Thermo Fisher). Media was added to cultures every 2 days to 3 days in a 3:1 ratio (new volume: old volume) or based on a 300 mg/dl glucose threshold. The ADOE was done using the same feeding schedule as the initial DOE to maintain consistency for validation. Media glucose was measured using a ChemGlass glucometer to confirm cell growth and activation.

5.3 | Flow cytometry

At the end of culture, at least 1e5 T cells from each run were washed with PBS once, resuspended in PBS, and stained with Zombie UV (Biolegend, 423107) for 30 min at room temperature in the dark at a 1:1000 dilution. Cells were spun and resuspended in FACS buffer

(1X PBS, 2% bovine serum albumin, 5 mM EDTA) and were stained with antibodies according to Table S1 for 60 min in the dark at 4°C. Cells were then resuspended in fresh FACS buffer, after which they were run on a BD LSR ortessa. All stained was performed in a 96 well v-bottom plate. See Supplementary Methods.

5.4 | Cytokine measurements

Cytokines were measured using a custom ProcartaPlex Luminex kit (Thermo Fisher). The assay was performed using media samples taken at various time points throughout the T-cell culture according to the manufacturer's instructions with modifications to half the reagent requirements. Data available at Supporting Dataset S1. See Supplementary Methods.

5.5 | NMR metabolomics sample preparation

Fifty microliter of media was collected from each culture at each time point (before media exchange, if applicable), flash-frozen in liquid nitrogen, and stored at −80°C. Samples were shipped to CCRC on dry ice for NMR analysis. Run order of samples was randomized. Samples were prepared in two batches for each rack of NMR samples to be run. For each rack, samples were pulled and sorted on dry ice, then thawed at 4°C for 1 h. Samples were then centrifuged at $2990 \times g$ at 4°C for 20 min to pellet any cells or debris that may have been collected with the media. 5 μl of 100/3 mM DSS-D6 in deuterium oxide (Cambridge Isotope Laboratories) were added to 1.7 mm NMR tubes (Bruker BioSpin), followed by 45 μl of media from each sample that was added and mixed, for a final volume of 50 μl in each tube. Samples were prepared on ice and in pre-determined, randomized order. The remaining volume from each sample in the rack (∼4 μl) was combined to create an internal pool. This material was used for internal controls within each rack as well as metabolite annotation.

5.6 | NMR data collection and processing

NMR spectra were collected on a Bruker Avance III HD spectrometer at 600 MHz using a 5-mm TXI cryogenic probe and TopSpin software (Bruker BioSpin). One-dimensional spectra were collected on all samples using the noesypr1d pulse sequence under automation using ICON NMR software. Two-dimensional (2D) HSQC and TOCSY spectra were collected on internal pooled control samples for metabolite annotation. One-dimensional spectra were manually phased and baseline corrected in TopSpin. 2D spectra were processed in NMRpipe.[45] One dimensional spectra were referenced, water/end regions removed, and normalized with the PQN algorithm[46] using an in-house MATLAB (The MathWorks, Inc.) toolbox (https://github.com/artedison/Edison_Lab_Shared_Metabolomics_UGA).

5.7 | NMR feature selection

To reduce the total number of spectral features from approximately 250 peaks and enrich for those that would be most useful for statistical modeling, a variance-based feature selection was performed within MATLAB. For each digitized point on the spectrum, the variance was calculated across all experimental samples and plotted. Clearly resolved features corresponding to peaks in the variance spectrum were manually binned and integrated to obtain quantitative feature intensities across all samples (Figure S4). In addition to highly variable features, several other clearly resolved and easily identifiable features were selected (glucose, BCAA region, etc.). Some features were later discovered to belong to the same metabolite but were included in further analysis. Data are available at Dataset S1.

5.8 | Metabolite annotation

2D spectra collected on pooled samples were uploaded to COLMARm web server,[47] where HSQC peaks were automatically matched to database peaks. HSQC matches were manually reviewed with additional 2D and proton spectra to confirm the match. Annotations were assigned a confidence score based upon the levels of spectral data supporting the match as previously described.[48] Annotated metabolites were matched to previously selected features used for statistical analysis. Several low abundance features selected for analysis did not have database matches and were not annotated.

5.9 | Low-field spectrum simulation

Using the list of annotated metabolites obtained above, an approximation of a representative experimental spectrum was generated using the GISSMO mixture simulation tool.[19,20] With the simulated mixture of compounds, generated at 600 MHz to match the experimental data, a new simulation was generated at 80 MHz to match the field strength of commercially available benchtop NMR spectrometers. The GISSMO tool allows visualization of signals contributed from each individual compound as well as the mixture, which allows annotation of features in the mixture belonging to specific compounds.

5.10 | ML modeling

Seven ML techniques were implemented to predict T_N and T_{CM} responses related to the memory phenotype of the cultured T cells under different process parameters conditions. The ML methods executed were RF, GBM, CIF, LASSO, PLSR, SVM, and SR. Primarily, SR models were used to optimize process parameter values based on $T_N + T_{CM}$ phenotype and to extract early predictive variable combinations from the multiomics experiments. SR was done using Evolved Analytics' Data Modeler software (Evolved Analytics LLC). While

nonparametric tree-based ensembles were done through the *randomForest*, *gbm*, and *cforest* regression functions in R, for RF, gradient boosted trees, and CIF models, respectively. Prediction performance was evaluated using LOO-R^2 and permutation-based variable importance scores assessing % increase of mean squared errors, relative influence based on the increase of prediction error, coefficient values for RF, GBM, and CID, respectively. Partial least squares regression was executed using the *plsr* function from the *pls* package in R while LASSO regression was performed using the *cv.glmnet* R package, both using leave-one-out cross-validation. Finally, the *kernlab* R package was used to construct the SVM regression models. Parameter tuning was done for all models in a grid search manner using the *train* function from the *caret* R package using LOO-R^2 as the optimization criteria. Prediction performance was measured for all models using the final model with LOO-R^2 tuned parameters. More details at Table S2. See Supplementary Methods.

5.11 | ML consensus analysis

Consensus analysis of the relevant variables extracted from each ML model was done to identify consistent predictive features of quality at the early stages of manufacturing. Using importance scores, key predictive variables were selected if their importance scores were within the 80th percentile ranking for the following ML methods: RF, GBM, CIF, LASSO, PLSR, SVM while for SR variables present in >30% of the top-performing SR models from Data Modeler ($R^2 \geq 90\%$, complexity ≤100) were chosen to investigate consensus. Only variables with those high percentile scoring values were evaluated in terms of their logical relation (intersection across ML models). See Supplementary Methods.

CONFLICT OF INTEREST

Bruce L. Levine declares financial interest intellectual property and patents in the field of cell and gene therapy (University of Pennsylvania Alliance with Novartis, licensing, and royalty fees). Bruce L. Levine is a consultant for Novartis, Terumo, and Lilly Asia Ventures and he is part of the Scientific Advisory Board for Avectas, Brammer Bio/TF Viral Vector Services, Immuneel, Incysus, Ori Biotech, and Vycellix. Moreover, Bruce L. Levine is the co-founder and equity holder Tmunity Therapeutics and all of his conflict of interest is managed in accordance with University of Pennsylvania policy and oversight. Theresa Kotanchek is the Chief Executive Officer of Evolved Analytics, LLC. The remaining authors declare no competing interests. Krishnendu Roy declares consulting, intellectual property, and patents in cell and gene therapy. Krishnendu Roy is a consultant to Terumo, Merck, LEK consulting, Mubadala Ventures, Anzu Partners, Decibio, and Clearview Healthcare Partners. Krishnendu Roy also serves on the advisory board of the MIT-Singapore Cell therapy Partnership.

AUTHOR CONTRIBUTIONS

Valerie Odeh-Couvertier: Conceptualization (equal); data curation (equal); formal analysis (lead); investigation (lead); methodology (lead); software (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Nathan J. Dwarshuis:** Conceptualization (equal); data curation (lead); formal analysis (equal); investigation (lead); methodology (lead); resources (equal); validation (lead); writing – original draft (equal); writing – review and editing (equal). **Maxwell B. Colonna:** Conceptualization (equal); data curation (lead); formal analysis (equal); investigation (lead); methodology (lead); resources (equal); software (equal); validation (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Bruce L. Levine:** Conceptualization (equal); supervision (equal); writing – original draft (equal); writing – review and editing (equal). **Arthur S. Edison:** Conceptualization (equal); funding acquisition (equal); investigation (equal); resources (lead); writing – original draft (equal); writing – review and editing (equal). **Theresa Kotanchek:** Conceptualization (equal); formal analysis (lead); methodology (lead); software (equal); supervision (equal); visualization (equal); writing – original draft (equal); writing – review and editing (equal). **Krishnendu Roy:** Conceptualization (lead); funding acquisition (lead); project administration (equal); resources (lead); supervision (equal); writing – original draft (equal); writing – review and editing (equal). **Wandaliz Torres-Garcia:** Conceptualization (equal); formal analysis (lead); methodology (equal); project administration (lead); resources (equal); software (lead); supervision (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead).

DATA SHARING AND DATA AVAILABILITY

The pre-processed set of the data used in this work is available in Supplementary Information (see Dataset S1). All NMR data are available at the Metabolomics Workbench[49] with DOI: https://doi.org/10.21228/M8F982. Machine learning implementation codes used in this work are available at GitHub (https://github.com/wandaliz/CMaT_TCell_MachineLearning/). DataModeler information can be requested at http://www.evolved-analytics.com/.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1002/btm2.10282.

ORCID

Wandaliz Torres-Garcia https://orcid.org/0000-0001-7729-5554

REFERENCES

1. Fesnak AD, June CH, Levine BL. Engineered T cells: the promise and challenges of cancer immunotherapy. *Nat Rev Cancer*. 2016;16(9): 566-581. doi:10.1038/nrc.2016.97

2. Rosenberg SA, Restifo NP. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science (New York, NY)*. 2015; 348(6230):62-68. doi:10.1126/science.aaa4967

3. Dwarshuis NJ, Parratt K, Santiago-Miranda A, Roy K. Cells as advanced therapeutics: state-of-the-art, challenges, and opportunities in large scale biomanufacturing of high-quality cells for adoptive immunotherapies. *Adv Drug Deliv Rev*. 2017;114:222-239. doi: 10.1016/j.addr.2017.06.005

4. Roddie C, O'Reilly M, Pinto JDA, Vispute K, Lowdell M. Manufacturing chimeric antigen receptor T cells: issues and challenges. *Cytotherapy*. 2019;21:327-340. doi:10.1016/j.jcyt.2018.11.009

5. Roh K-H, Nerem RM, Roy K. Biomanufacturing of therapeutic cells: state of the art, current challenges, and future perspectives. *Annu Rev Chem Biomol Eng*. 2016;7(1):455-478. doi:10.1146/annurev-chembioeng-080615-033559

6. Carmen J, Burger SR, McCaman M, Rowley JA. Developing assays to address identity, potency, purity and safety: cell characterization in cell therapy process development. *Regen Med*. 2012;7(1):85-100. doi: 10.2217/rme.11.105

7. Simon CG, Lin-Gibson S, Elliott JT, Sarkar S, Plant AL. Strategies for achieving measurement assurance for cell therapy products. *Stem Cells Transl Med*. 2016;5(6):705-708. doi:10.5966/sctm.2015-0269

8. Campbell A, Brieva T, Raviv L, et al. Concise review: process development considerations for cell therapy. *Stem Cells Transl Med*. 2015; 4(10):1155-1163. doi:10.5966/sctm.2014-0294

9. Lipsitz YY, Timmins NE, Zandstra PW. Quality cell therapy manufacturing by design. *Nat Biotechnol*. 2016;34(4):393-400. doi: 10.1038/nbt.3525

10. Better M, Chiruvolu V, Sabatino M. Overcoming challenges for engineered autologous T cell therapies. *Cell Gene Ther Insights*. 2018;4(4):173-186.

11. Tyagarajan S, Spencer T, Smith J. Optimizing CAR-T cell manufacturing processes during pivotal clinical trials. *Mol Ther Methods Clin Dev*. 2020;16:136-144. doi:10.1016/j.omtm.2019.11.018

12. Xu Y, Zhang M, Ramos CA, et al. Closely related T-memory stem cells correlate with in vivo expansion of CAR.CD19-T cells and are preserved by IL-7 and IL-15. *Blood*. 2014;123(24):3750-3759. doi: 10.1182/blood-2014-01-552174

13. Gattinoni L, Klebanoff CA, Restifo NP. Paths to stemness: building the ultimate antitumour T cell. *Nat Rev Cancer*. 2012;12(10):671-684. doi:10.1038/nrc3322

14. Fraietta JA, Lacey SF, Orlando EJ, et al. Determinants of response and resistance to CD19 chimeric antigen receptor (CAR) T cell therapy of chronic lymphocytic leukemia. *Nat Med*. 2018;24(5):563-571. doi:10.1038/s41591-018-0010-1

15. Gattinoni L, Lugli E, Ji Y, et al. A human memory T cell subset with stem cell–like properties. *Nat Med*. 2011;17(10):1290-1297. doi: 10.1038/nm.2446

16. Wang D, Aguilar B, Starr R, et al. Glioblastoma-targeted CD4+ CAR T cells mediate superior antitumor activity. *JCI Insight*. 2018;3(10): e99048. doi:10.1172/jci.insight.99048.

17. Yang Y, Kohler ME, Chien CD, et al. TCR engagement negatively affects CD8 but not CD4 CAR T cell expansion and leukemic clearance. *Sci Transl Med*. 2017;9(417):eaag1209. doi:10.1126/scitranslmed.aag1209

18. Dwarshuis NJ, Song HW, Patel A, Kotanchek T, Roy K. Functionalized microcarriers improve T cell manufacturing by facilitating migratory memory T cell production and increasing CD4/CD8 ratio. *bioRxiv*. 2019;646760. doi:10.1101/646760

19. Dashti H, Westler WM, Tonelli M, Wedell JR, Markley JL, Eghbalnia HR. Spin system modeling of nuclear magnetic resonance spectra for applications in metabolomics and small molecule screening. *Anal Chem*. 2017;89(22):12201-12208. doi:10.1021/acs.analchem.7b02884

20. Dashti H, Wedell JR, Westler WM, et al. Applications of parametrized NMR spin systems of small molecules. *Anal Chem*. 2018;90(18): 10646-10649. doi:10.1021/acs.analchem.8b02660

21. Kordon AK, Lue C-T. Symbolic regression modeling of blown film process effects. *Proceedings of the 2004 Congress on Evolutionary Computation (IEEE Cat. No.04TH8753)*. Vol 1. IEEE; 2004:561-568. doi: 10.1109/CEC.2004.1330907

22. Koza JR. Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*. 1994;4(2):87. doi:10.1007/bf00175355

23. Kotanchek M, Smits G, Vladislavleva E. Pursuing the Pareto paradigm: tournaments, algorithm variations and ordinal optimization. In: Riolo R, Soule T, Worzel B, eds. *Genetic Programming Theory and Practice IV. Genetic and Evolutionary Computation*. Springer; 2007:167-185. doi:10.1007/978-0-387-49650-4_11

24. Kotanchek M, Smits G, Vladislavleva E. Exploiting trustable models via Pareto GP for targeted data collection. *Genetic Programming Theory and Practice VI. Genetic and Evolutionary Computation*. Springer; 2009:1-18. doi:10.1007/978-0-387-87623-8_10

25. Witkowska AM. On the role of sIL-2R measurements in rheumatoid arthritis and cancers. *Mediators Inflamm*. 2005;2005(3):121-130. doi: 10.1155/MI.2005.121

26. Vudattu NK, Holler E, Ewing P, et al. Reverse signalling of membrane-integrated tumour necrosis factor differentially regulates alloresponses of CD4+ and CD8+ T cells against human microvascular endothelial cells. *Immunology*. 2005;115(4):536-543. doi:10.1111/j.1365-2567.2005.02190.x

27. Mehta AK, Gracias DT, Croft M. TNF activity and T cells. *Cytokine*. 2018;101:14-18. doi:10.1016/j.cyto.2016.08.003

28. Wong FS. Stimulating IL-13 receptors on T cells: a new pathway for tolerance induction in diabetes? *Diabetes*. 2011;60(6):1657-1659. doi:10.2337/db11-0353

29. Junttila IS. Tuning the cytokine responses: an update on interleukin (IL)-4 and IL-13 receptor complexes. *Front Immunol*. 2018;9:888. doi: 10.3389/fimmu.2018.00888.

30. Amatya N, Garg AV, Gaffen SL. IL-17 signaling: the Yin and the Yang. *Trends Immunol*. 2017;38(5):310-322. doi:10.1016/j.it.2017.01.006

31. Becher B, Tugues S, Greter M. GM-CSF: from growth factor to central mediator of tissue inflammation. *Immunity*. 2016;45(5):963-973. doi:10.1016/j.immuni.2016.10.026

32. Hurton LV, Singh H, Najjar AM, et al. Tethered IL-15 augments antitumor activity and promotes a stem-cell memory subset in tumor-specific T cells. *Proc Natl Acad Sci USA*. 2016;113(48):E7788-E7797. doi:10.1073/pnas.1610544113

33. Ron-Harel N, Santos D, Ghergurovich JM, et al. Mitochondrial biogenesis and proteome remodeling promotes one carbon metabolism for T cell activation. *Cell Metab*. 2016;24(1):104-117. doi:10.1016/j.cmet.2016.06.007

34. Pietzke M, Meiser J, Vazquez A. Formate metabolism in health and disease. *Mol Metab*. 2020;33:23-37. doi:10.1016/j.molmet.2019.05.012

35. van der Windt GJW, Everts B, Chang C-H, et al. Mitochondrial respiratory capacity is a critical regulator of CD8+ T cell memory development. *Immunity*. 2012;36(1):68-78. doi:10.1016/j.immuni.2011.12.007

36. Vardhana SA, Hwee MA, Berisa M, et al. Impaired mitochondrial oxidative phosphorylation limits the self-renewal of T cells exposed

to persistent antigen. *Nat Immunol*. 2020;13:1-12. doi:10.1038/s41590-020-0725-2

37. Lunt SY, Vander Heiden MG. Aerobic glycolysis: meeting the metabolic requirements of cell proliferation. *Annu Rev Cell Dev Biol*. 2011;27(1):441-464. doi:10.1146/annurev-cellbio-092910-154237

38. Chang C-H, Curtis JD, Maggi LB, et al. Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell*. 2013;153(6):1239-1251. doi:10.1016/j.cell.2013.05.016

39. Cao Y, Rathmell JC, Macintyre AN. Metabolic reprogramming towards aerobic glycolysis correlates with greater proliferative ability and resistance to metabolic inhibition in CD8 versus CD4 T cells. *PLoS One*. 2014;9(8):e104104. doi:10.1371/journal.pone.0104104

40. Antoshechkin AG. On intracellular formation of ethanol and its possible role in energy metabolism. *Alcohol Alcohol*. 2001;36(6):608. doi:10.1093/alcalc/36.6.608

41. Sena LA, Li S, Jairaman A, et al. Mitochondria are required for antigen-specific T cell activation through reactive oxygen species signaling. *Immunity*. 2013;38(2):225-236. doi:10.1016/j.immuni.2012.10.020

42. Almeida L, Lochner M, Berod L, Sparwasser T. Metabolic pathways in T cell activation and lineage differentiation. *Semin Immunol*. 2016;28(5):514-524. doi:10.1016/j.smim.2016.10.009

43. Wang R, Green DR. Metabolic checkpoints in activated T cells. *Nat Immunol*. 2012;13(10):907-915. doi:10.1038/ni.2386

44. Chilmonczyk MA, Kottke PA, Stevens HY, Guldberg RE, Fedorov AG. Dynamic mass spectrometry probe for electrospray ionization mass spectrometry monitoring of bioreactors for therapeutic cell manufacturing. *Biotechnol Bioeng*. 2019;116(1):121-131. doi:10.1002/bit.26832

45. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR*. 1995;6(3):277-293. doi:10.1007/BF00197809

46. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. *Anal Chem*. 2006;78(13):4281-4290. doi:10.1021/ac051632c

47. Bingol K, Li D-W, Zhang B, Brüschweiler R. Comprehensive metabolite identification strategy using multiple two-dimensional NMR spectra of a complex mixture implemented in the COLMARm web server. *Anal Chem*. 2016;88(24):12411-12418. doi:10.1021/acs.analchem.6b03724

48. Walejko JM, Chelliah A, Keller-Wood M, Gregg A, Edison AS. Global metabolomics of the placenta reveals distinct metabolic profiles between maternal and fetal placental tissues following delivery in non-labored women. *Metabolites*. 2018;8(1). doi:10.3390/metabo8010010

49. Sud M, Fahy E, Cotter D, et al. Metabolomics workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res*. 2016;44(D1):D463-D470. doi:10.1093/nar/gkv1042

50. Holmes E, Cloarec O, Nicholson JK. Probing latent biomarker signatures and in vivo pathway activity in experimental disease states via statistical total correlation spectroscopy (STOCSY) of biofluids: application to HgCl2 toxicity. *J Proteome Res*. 2006;5(6):1313-1320. doi:10.1021/pr050399w

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.