

On-chip 3D potency assay for prediction of clinical outcomes for cell therapy candidates for osteoarthritis

Received: 10 January 2025

Accepted: 15 May 2025

Published online: 27 May 2025

 Check for updates

Rebecca S. Schneider^{1,2}, Elisa B. Nieves^{2,3}, Bhavay Aggarwal³, Annie C. Bowles-Welch^{2,4}, Hazel Y. Stevens^{2,4}, Linda E. Kippner^{2,4}, Scott D. Boden⁵, Kenneth Mautner⁵, Hicham Drissi⁵, Krishnendu Roy^{2,3}, Wilbur A. Lam^{2,3}, Saurabh Sinha³ & Andrés J. García^{2,6} ✉

The lack of clinically predictive potency assays for cell products significantly impedes translation of these therapies. Here, we describe a microfluidic on-chip 3D system for rapid evaluation of a subset of patient-derived bone marrow aspirate concentrate (BMAC) samples used in a phase 3 multicenter trial (NCT03818737) evaluating autologous cells for relieving knee osteoarthritis pain. BMAC clinical samples cultured in the on-chip 3D system exhibit elevated levels of immunomodulatory and trophic proteins compared to 2D culture. Using analyte information from in vitro assays and patient-matched clinical data, we build linear regression prediction models for clinical outcomes. We demonstrate improved clinical prediction by cross-validation accuracy for the on-chip 3D platform compared to 2D culture. Additionally, on-chip 3D assay metrics display higher correlative power with patient pain scores compared to the 2D assay. This study establishes a potency assay with improved prediction power to accelerate translation of cell therapies.

Cytherapies involving bone marrow- or adipose tissue-derived mesenchymal cells are actively pursued to treat diverse inflammatory and immunological disorders, including myocardial infarcts, osteoarthritis, autoimmune disorders, and diabetic and gastrointestinal wounds^{1,2}. However, the lack of clinically predictive potency metrics for cell products represents a major bottleneck for the development and translation of these therapies^{3–7}. ‘Predictive validity’ is defined as the correlation between the output of a decision tool and clinical utility across therapeutic candidates, where ‘decision tools’ include any tool that drives R&D decisions that are postulated to correlate with a therapeutic candidate’s clinical utility⁸. Improvements in a decision tool’s predictive validity can significantly accelerate the translation of cellular therapies by providing more relevant safety and

efficacy data as well as informing clinical decisions. In addition, improvements in potency assay could be worth hundreds of thousands of dollars per industry phase 1 candidate⁸.

Regulatory agencies, including the US Food and Drug Administration (US FDA), require quantitative measures of biologic function (potency metrics) as metrics of a product’s quality and means for donor-to-donor or lot-to-lot comparability for late stages of clinical investigations and commercial applications⁹. Where direct measurement of biologic function is not feasible, surrogate markers of potency are recommended¹⁰, but the identification of robust surrogate potency markers for cell therapies with unclear and/or pleiotropic mechanisms of action has been particularly challenging^{6,7,10}. For instance, the US FDA rejected a biological license application (BLA 125706) by

¹School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA, USA. ²Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA, USA. ³Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA, USA. ⁴Marcus Center for Therapeutic Cell Characterization and Manufacturing, Georgia Institute of Technology, Atlanta, GA, USA. ⁵Department of Orthopaedics, Emory University, Atlanta, GA, USA. ⁶Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA, USA.

✉ e-mail: andres.garcia@me.gatech.edu

Mesoblast, Inc. in August 2020 for a mesenchymal stromal cell product due to the lack of potency attributes representative of clinical performance and lack of assurance in consistent manufacturing processes¹¹. Planar (2D) cell culture systems have been traditionally used as in vitro assays for discovery and potency testing, yet the practical value of 2D assay data in drug development and clinical decision-making remains limited. Due to the shortcomings of 2D assay data, there has been a notable shift towards development and validation of 3D models, including organoid models^{12–15} and ‘organ-on-a-chip’ microphysiological systems^{16,17} to better emulate the complexity of in vivo environments and enhance predictive accuracy in clinical decision-making. Indeed, the recently enacted US FDA Modernization Act 2.0 allows the use of alternatives to animal testing, including human cell-based assays, organoids, and organ-on-chip platforms¹⁸. Whereas these technologies provide valuable insights into toxicology and mechanism(s) of action, they still face significant adoption challenges due to high variability, complex and lengthy assay times, and lack of scalability. Importantly, to the best of our knowledge, none of these tools have demonstrated improved prediction of patient-matched clinical outcomes.

Here, we evaluated a microfluidic on-chip 3D system as a decision tool for rapid and efficient assessment of a subset of patient-derived bone marrow aspirate concentrate (BMAC) samples investigated in a phase 3 multicenter trial (NCT03818737, MILES trial¹⁹) evaluating cell therapy products for relieving knee osteoarthritis (OA) pain. The on-chip 3D potency assay consists of a simple, low-cost microfluidic device with media perfusion through a cell-laden synthetic hydrogel and multiplexed analysis of secreted cellular products²⁰. This system provides 3D structural, mechanical, and biochemical cues and fluid flow to the encapsulated cells. This potency assay exhibits higher fidelity to in vitro suppression of T cell proliferation and secreted cytokine/chemokine levels in immunocompromised mice compared to 2D culture for human mesenchymal stromal cells. BMAC samples cultured in the on-chip 3D system exhibited elevated secreted levels of immunomodulatory and trophic proteins (cytokines, chemokines, cell adhesive proteins, MMPs) compared to 2D culture. Using secreted analyte information from in vitro assays and patient-matched clinical outcome data, we built linear regression prediction models for clinical outcomes. We demonstrate improved clinical prediction for the on-chip 3D system compared to 2D culture assay. Additionally, on-chip 3D assay metrics displayed higher correlative power with patient pain scores compared to the 2D assay. This study establishes a 3D potency assay with improved prediction power that can accelerate the translation of cell therapies.

Results

On-chip 3D potency assay for patient-derived BMAC samples

We investigated the potency of a subset of patient-derived BMAC samples from a phase 3 multicenter trial evaluating cell therapy products for relieving knee OA pain (NCT03818737, MILES trial¹⁹). Bone marrow aspirates were collected from OA patients and concentrated into BMAC for administration as fresh cells in saline via intra-articular injection into the OA-afflicted knee joint (autologous treatment). The clinical trial co-primary endpoints were the Visual Analog Scale (VAS) pain score and Knee injury and Osteoarthritis Outcome Score (KOOS) pain score at 12 months versus baseline (before treatment). When sufficient BMAC beyond the amount needed for clinical treatment of an individual patient was collected, this BMAC sample was cryopreserved and subsequently used for either RNA sequencing (reported in ref. 21) or evaluation of in vitro cell potency in the current study. Only 22 patient samples were available for in vitro cell potency assessment, and all available BMAC clinical samples with paired blinded clinical outcome data (responder/non-responder, pain scores) were analyzed and included in this report. Donor characteristics are summarized in Supplementary Table 1. BMAC is a heterogeneous mixture of innate

and adaptive immune, hematopoietic progenitor, and mesenchymal and stromal cell populations. Chatterjee et al. reported single-cell RNA sequencing for the broader MILES BMAC donor cohort²¹, and transcriptomic results for four of the donors evaluated in our study were included in that study (see Supplementary Table 1 for donor ID for cross-referencing).

BMAC clinical samples were evaluated using an in vitro microfluidic on-chip 3D system²⁰. Supplementary Fig. 1 presents the poly(dimethylsiloxane) (PDMS) microfluidic device design and an overview of the experimental workflow. BMAC samples were thawed and encapsulated in 4-arm maleimide-functionalized poly(ethylene-glycol) (PEG-4MAL) hydrogel presenting cell-adhesive RGD peptide and cross-linked with protease-degradable peptide and non-degradable dithiothreitol (Supplementary Fig. 2). This hydrogel formulation was selected as we previously showed that its biochemical and biophysical properties support the viability and secretory activity of human mesenchymal stromal cells²⁰, a key secretory cell population within BMAC. We note that this hydrogel is engineered to encapsulate and support the viability and function of BMAC cells while allowing media perfusion in the microfluidic device; the hydrogel is not intended to fully replicate the microenvironment that BMAC cells experience when injected into the joint cavity. The cell-laden hydrogel was incorporated into the device and perfused with media (Fig. 1). We selected a perfusion rate of 1.0 $\mu\text{L}/\text{min}$ based on prior work demonstrating that this perfusion rate supports high viability and activities for mesenchymal stromal cells encapsulated within the synthetic hydrogel in the microfluidic device²⁰. As reviewed in Low and Tagle²², microphysiological systems (i.e., organ-on-a-chip assays) often require fluid flow through the system to deliver nutrients to cells and remove cellular waste. Based on the device dimensions, we estimate that this perfusion rate corresponds to $\sim 1 \mu\text{m}/\text{s}$; this velocity falls within the physiological range of interstitial fluid velocity ($0.1\text{--}4 \mu\text{m}/\text{s}$)^{23–25}. For this range of interstitial fluid velocities, surface shear stresses on cells in extracellular matrices are estimated to be on the order of $0.1 \text{ dyne}/\text{cm}^2$ ^{26,27}. We selected a 24-h perfusion duration because this time provided for high cell viability and significant secretion of analytes while shortening the time for analysis, which are important considerations for cell potency assays.

We hypothesized that BMAC secretory response is dependent on the composition and viscosity of the culture media or synovial fluid in the joint. We therefore synthesized an OA simulated synovial fluid (simSF) mimic, informed by analyte analysis of patient synovial fluid, that contains the most abundant proteins present in OA patient-derived synovial fluid (Supplementary Table 2) and is supplemented with glycosaminoglycans to match the viscosity and shear thinning properties of OA patient-derived synovial fluid (Supplementary Fig. 3). The simSF formulation provides a uniform test condition for BMAC samples and does not aim to recapitulate all of the biological activities of natural synovial fluid. BMAC samples in on-chip 3D and 2D culture exposed to basal control media (ctrl) or media supplemented with 10% simSF (simSF) were analyzed for the secretion of 24 immunomodulatory and trophic proteins (Supplementary Table 3) selected from the secretome of mesenchymal stromal cells, a driving secretory subpopulation within BMAC²⁸. The secreted analyte levels were used in a data-driven linear regression prediction model where cross-validation testing was performed for matched-patient clinical outcomes (Fig. 1).

On-chip 3D system has increased immunomodulatory and trophic protein levels

After thawing, BMAC cell viability was $75\% \pm 10\%$ (mean \pm SD) across all donors. This range in out-of-thaw viability is likely due to donor-specific differences in cellular composition of the BMAC, including lymphocytes which are sensitive to cryopreservation. However, the out-of-thaw viability results are not indicative of disease severity or patient-related characteristics such as age, weight/BMI, or levels of

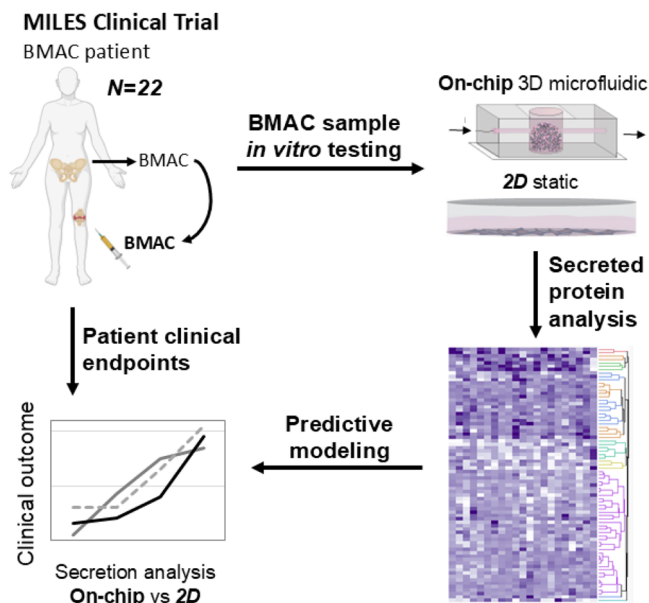


Fig. 1 | Schematic of approach and execution. BMAC donors ($n = 22$) were evaluated for secretion profiles in on-chip 3D microfluidic and conventional 2D culture models. Secreted analytes were assessed by data-driven predictive modeling in which accuracy in prediction for patient-matched clinical outcomes across models was compared. Created in BioRender. García, A. (2025) <https://BioRender.com/6crymr6>.

activity. We also examined the viability for cells recovered from the microfluidic device at the end of perfusion to confirm that culture within the device did not affect cell viability. BMAC viability after 24-h culture in the on-chip 3D and 2D assays was examined in 4 donors (randomly selected) and shown to be $77\% \pm 3\%$ with no differences between on-chip 3D and 2D for either ctrl or simSF conditions (Supplementary Fig. 4). Following 24-h culture in on-chip 3D or 2D assays, BMAC secreted analyte levels were compared. Equivalent number of cells were analyzed between the on-chip 3D platform and 2D culture. Across the 22 donors, there was strong separation by hierarchical clustering between on-chip 3D and 2D assays (Fig. 2A, analyte concentrations are provided in Source Data). This indicates that the secretory differences between assay platforms are dominated by the assay type rather than donor. Similarly, by discriminant analysis, outcomes for 2D cultures treated with simSF and ctrl media clustered together, whereas the on-chip 3D assay results clustered distinctly from 2D assay, with separation between simSF and ctrl treatments (Fig. 2B). Canonical scores 1 and 2 account for 97.0% and 2.9% of the overall variance of the data.

We previously showed that secretion values normalized to donor-matched 2D control samples (pg/pg) exhibited high consistency across independent experiments and higher correlation to in vitro immunomodulation (suppression of T cell proliferation) outcomes²⁰, thus, unless otherwise specified, all secretion values are normalized to the 2D control sample. Figure 3 presents normalized secreted levels for individual analytes. For all analytes, secreted values for the on-chip 3D group were significantly higher than the 2D assay. This includes potent immune modulators such as PD-L1 (Fig. 3A), TNF- α (Fig. 3B), IFN- γ (Fig. 3C), and IL-6 (Fig. 3D) as well as various matrix modulators including MMP-1 (Fig. 3E), MMP-3 (Fig. 3F), MMP-13 (Fig. 3G), and TIMP-1 (Fig. 3H). Of the 24 analytes measured, HGF (Fig. 3I), CCL2/MCP-1 (Fig. 3J), IL-8/CXCL8 (Fig. 3K), and CXCL9/MIG (Fig. 3L) showed strong upregulated response in the on-chip 3D simSF group compared to all other groups.

We postulated that the improved BMAC secretory responses for the on-chip 3D assay require both the 3D environment and fluid flow

conditions of the microfluidic chip. We conducted additional testing with on-chip 3D samples, 3D samples without simSF perfusion (3D static simSF) or ctrl media perfusion (3D static ctrl), and 2D simSF and ctrl groups for two BMAC donors. Hierarchical clustering again showed clustering by assay configuration (Supplementary Fig. 5A). For most analytes, higher secretion values were observed for the on-chip 3D (perfused) condition compared to the static 3D and 2D conditions (Supplementary Fig. 5B), indicating that both the 3D environment and fluid flow of the microfluidic assay are needed for the elevated secretory responses.

Correlation of secreted analytes with donor age and disease severity

We observed disparate responses for BMAC secreted analyte levels among on-chip 3D and 2D platforms. To gain insights into these platform-specific differences, we examined correlations between BMAC analyte secretion levels and matched donor age and disease severity prior to receiving cell therapy. We hypothesized that donor age or donor disease severity influences BMAC secretory responses to the stimuli provided by the in vitro platforms. The Kellgren-Lawrence (KL) classification represents a radiological score of OA disease severity²⁹. KL score ranges from 0 to 4, with a score of 0 representing a non-OA healthy joint, and a score of 4 representing the highest severity of OA. BMAC samples were collected from OA patients that ranged in age from 47 to 68 and KL score from 2 (less severe) to 4 (most severe). There were no differences in patient age across KL scores (Supplementary Fig. 6).

For each assay platform and condition, linear regression analyses between donor age (independent variable) and donor-matched normalized secreted analyte levels (dependent variable) were performed (Supplementary Fig. 7). We report two key metrics from the linear regression analyses: (1) P value for the non-zero regression slope test, and (2) R^2 for non-zero slope relationships. A statistically significant P value ($P < 0.05$) for the non-zero slope test indicates that the secreted analyte concentration correlates linearly with donor age. The R^2 statistic indicates the percentage of the variance in the dependent variable explained by the independent variable. Low R^2 values in these analyses are expected as other variables, such as the heterogeneity of the BMAC samples and donor-specific differences, contribute to the variance in secreted analyte concentration. Of the 24 analytes evaluated, only TNF R1 and IL-17E for on-chip 3D ctrl and TNF- α for 2D assay exhibited secretion levels that were linearly correlated with donor age ($P < 0.05$, Fig. 4A–C, Supplementary Fig. 7). For the on-chip 3D ctrl, secreted TNF R1 and IL-17E concentrations decreased with increasing donor age. The top correlative analyte to donor age, TNF R1, functions as an inhibitor of T cell immunoregulatory and proapoptotic responses^{30,31}. Unlike the on-chip 3D ctrl correlations, 2D simSF-treated cells showed that secretion of TNF- α increased with increasing donor age (Supplementary Fig. 7).

Linear regression analyses between donor KL grade prior to receiving cell therapy (independent variable) and donor-matched normalized secreted analyte levels (dependent variable) were also performed (Supplementary Fig. 8). Of the 24 analytes evaluated, only secreted concentration of MMP-3 for the on-chip 3D ctrl platform was linearly correlated with KL score (Fig. 4D–F). MMP-3 has been implicated in cartilage destruction and MMP-3 levels in synovial fluid is strongly positively correlated with pain scores in patients with knee OA^{32,33}.

On-chip 3D platform exhibits improved prediction power for clinical outcomes

To assess the clinical prediction power of the on-chip 3D system compared to 2D culture assay, we built linear regression prediction models for clinical outcomes. We received clinical outcomes (response to treatment) for 19 patients; 15 responded to the treatment

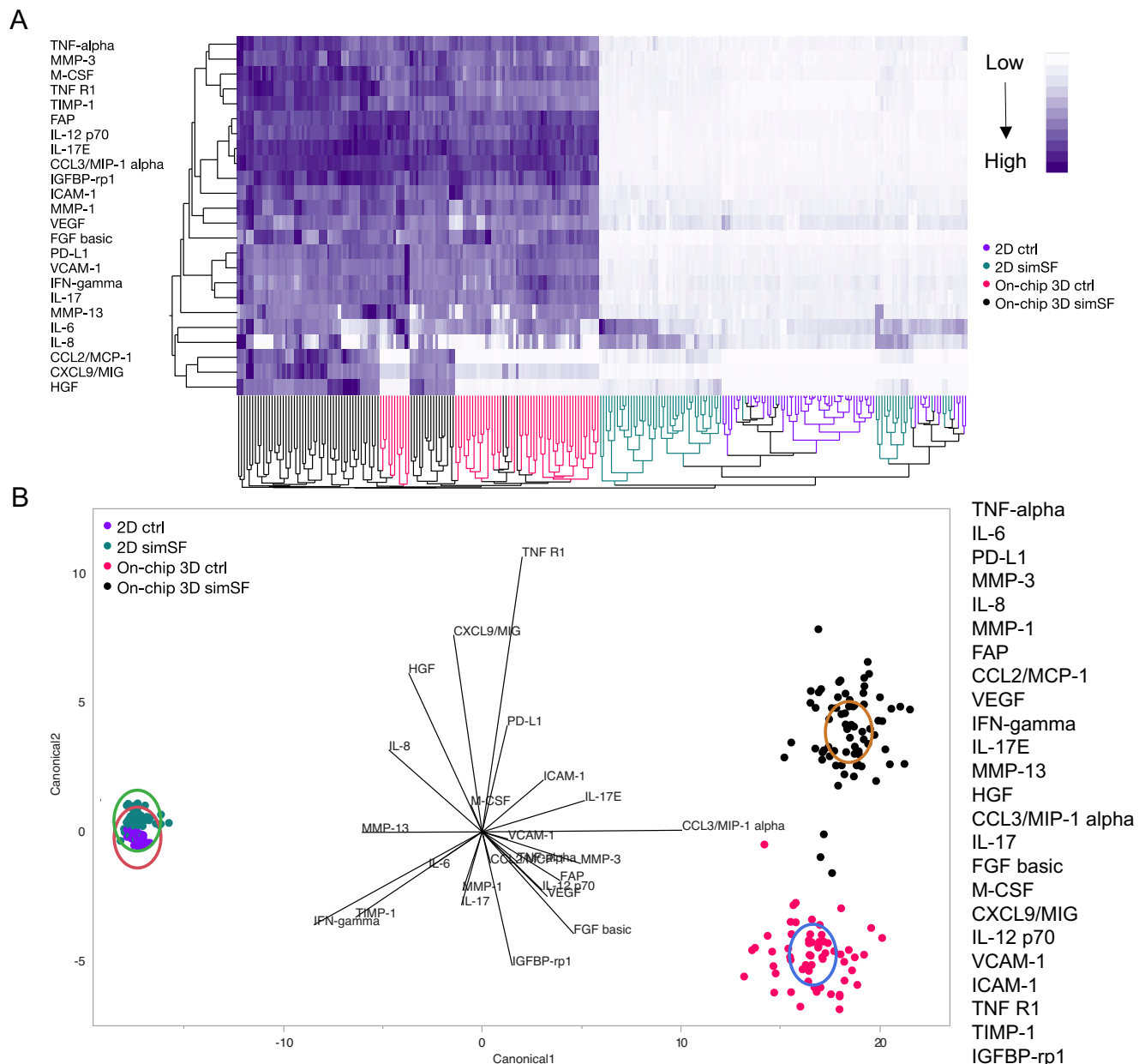


Fig. 2 | Higher separation of on-chip 3D secretion compared to 2D assay for BMAC donors. A Hierarchical clustering heatmap. Darker purple represents higher secretion value. **B** Discriminant analysis with biplot rays of normalized sample secretion (pg/pg) showing higher donor variability across on-chip 3D secretion not

captured in 2D assay. Canonical score 1 and 2 account for 97.0% and 2.9% of the overall variance of the data. Analysis represents 24 secreted proteins of on-chip 3D and 2D samples, with and without simSF.

(responders), and 4 did not respond (non-responders). These results are a small subset of the treated patient cohort and give no indication of the results of the clinical trial itself. We adopted a train-test split procedure to recreate a “test on unseen data” situation. A least absolute shrinkage and selection operator (LASSO) model was trained in repeats or folds of 4, using a semi-randomized 15 sample training set and 4 sample test set (Fig. 5A). The model was taught over the 4-folds by developing the model on the training set and testing for prediction accuracy on the test set. This accuracy is reported as an averaged cross-validation accuracy across the folds. Additionally, to determine model confidence, this model loop was repeated 100 times, and the cross-validation accuracy prediction confidence was measured as the standard error of the mean across the 100 repeats. The 24 analyte values for each BMAC donor are represented as the model ‘features’, whereas the donor-matched clinical outcomes (treatment response,

no treatment response) are characterized as the ‘classes’ for prediction. The ‘classes’ for this model are weighted for representation of responders and non-responders represented in the dataset.

We found a significant increase in clinical prediction by cross-validation accuracy for the on-chip 3D ctrl samples compared to 2D samples (0.61 vs. 0.48, $P < 0.0001$) (Fig. 5B). We found no increase in cross-validation accuracy between the on-chip 3D simSF and 2D samples. Interestingly, the on-chip 3D ctrl samples had increased clinical prediction compared to the on-chip 3D simSF (0.61 vs. 0.51, $P = 0.0009$). This result was unexpected as we hypothesized that the on-chip 3D simSF platform would have higher prediction of clinical outcomes as the simSF perfusate was synthesized to mimic the biochemical and biomechanical properties of OA synovial fluid. Although the on-chip platform is engineered to encapsulate cells in a 3D soft matrix with media perfusion to support the viability and function of

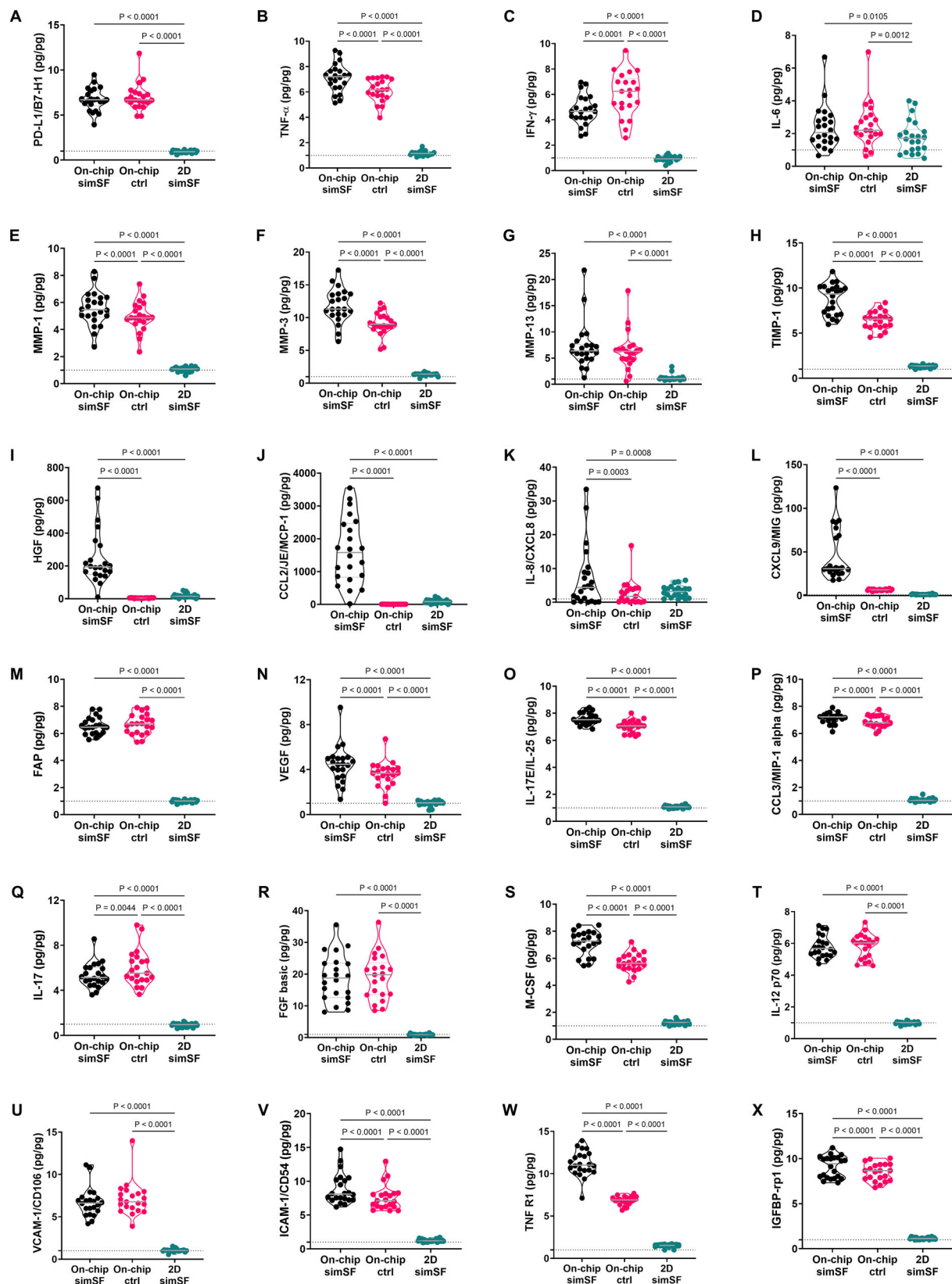


Fig. 3 | Individual analyte secretion of BMAC donors across on-chip 3D and 2D platforms. A PD-L1/B7-H1. **B** TNF- α . **C** IFN- γ . **D** IL-6. **E** MMP-1. **F** MMP-3. **G** MMP-13. **H** TIMP-1. **I** HGF. **J** CCL2/MCP-1. **K** IL-8/CXCL8. **L** CXCL9/MIG. **M** FAP. **N** VEGF. **O** IL-17E/IL-25. **P** CCL3/MIP-1 α . **Q** IL-17. **R** FGF basic. **S** M-CSF. **T** IL-12 p70. **U** VCAM-1/CD106. **V** ICAM-1/CD54. **W** TNF R1. **X** IGFBP-rp1. Values presented as normalized

(pg/pg) secretion values to donor-matched 2D controls across 24 analytes. Each point represents the mean of a unique donor ($n = 22$) across $n = 1-4$ technical replicates. The dotted line represents the baseline 2D control at 1 pg/pg. ANOVA performed across platforms with two-sided t -tests for post-hoc tests.

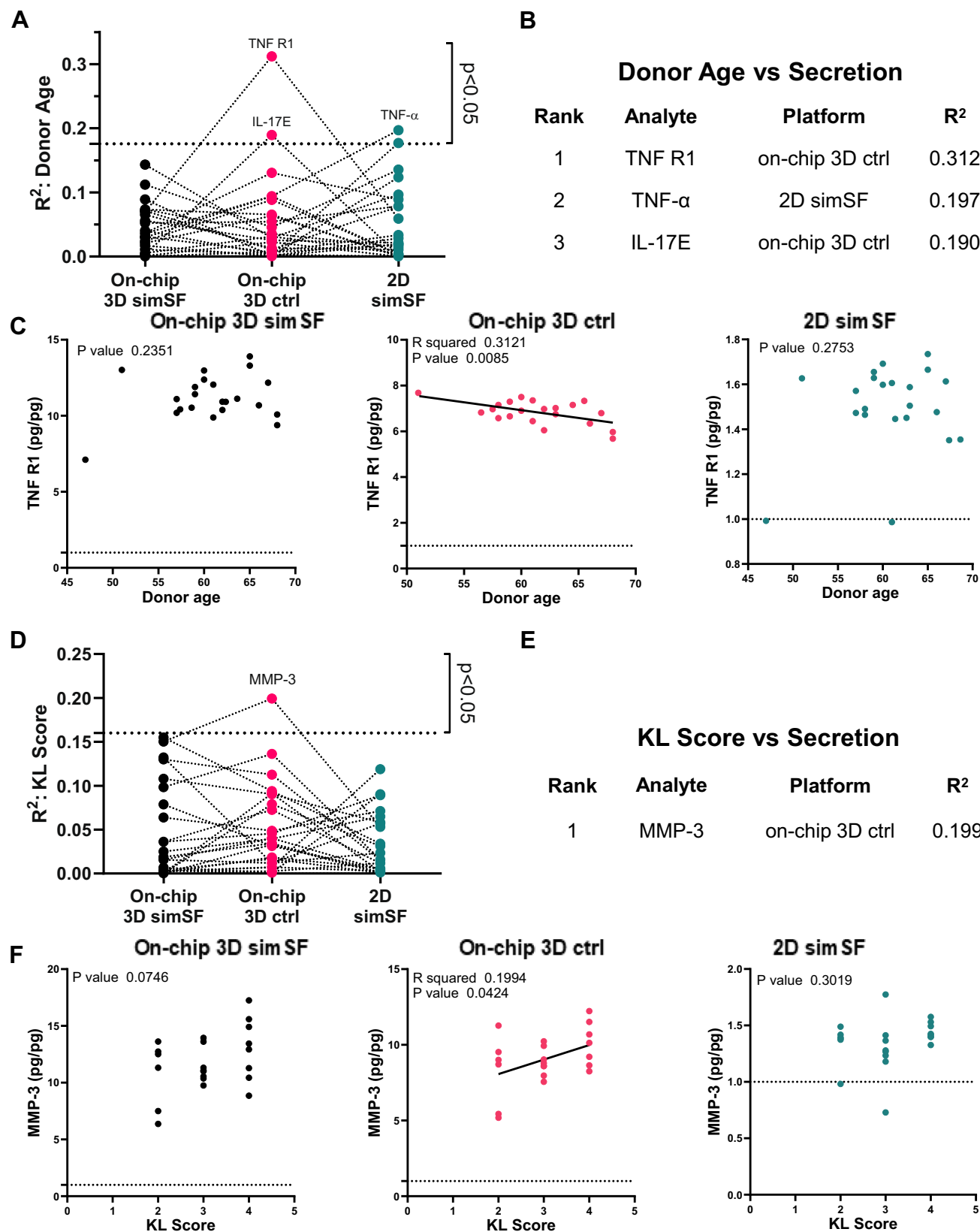


Fig. 4 | A subset of secreted analytes correlates with donor age and disease severity. **A** Donor age vs. secreted analyte concentration linear relationship across 24 analytes and 3 assay platforms for 22 donors. Linear regression analyses with non-zero slopes ($P < 0.05$) are indicated. **B** Top 3 best-fit analytes with non-zero regression slopes across assay platforms. **C** Linear regression plots for TNF R1. Each point represents the mean of a unique donor ($n = 22$) across $n = 1-4$ technical replicates. **D** OA disease severity by KL score vs. secreted analyte concentration across 24 analytes and 3 assay platforms for 22 donors. Linear regression analyses with non-zero slopes ($P < 0.05$) are indicated. **E** MMP-3 for on-chip 3D ctrl platform is the only analyte that exhibited a linear correlation with KL score. **F** Linear regression plots for MMP-3. Each point represents the mean of a unique donor ($n = 22$) across $n = 1-4$ technical replicates.

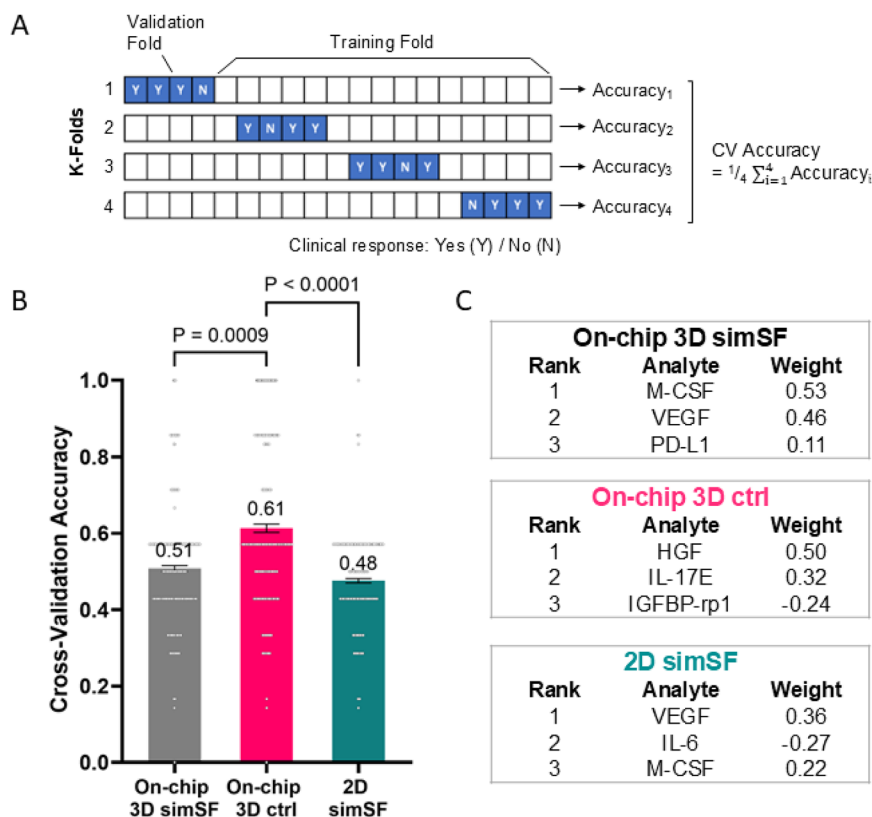


Fig. 5 | On-chip 3D assay exhibits improved prediction power for clinical outcomes. **A** LASSO model built from iterative learning of a logistic regression model with 24 secreted protein concentrations as the dependent variable and clinical outcomes as the independent variable. Outcomes are represented as responding and non-responding, or Yes (Y) and No (N), respectively. There are 4 k-folds, each consisting of a training fold ($n = 15$) and a validation fold ($n = 4$). Each validation fold has one unique non-responding (N) sample. To address this imbalance between

responders and non-responders, we use balanced accuracy rather than the usual accuracy as a measure of performance. **B** Cross-validation accuracy of on-chip 3D ctrl samples is greater than on-chip 3D simSF and 2D simSF. Cross-validation accuracy results comprise an average of 100 repeats of the prediction model for each fold (400 points per platform). ANOVA performed across platforms with two-sided t -tests for post-hoc tests. Error represented as SEM. **C** The top three contributing analytes to each model with listed model weights.

BMAC cells, this system may not fully replicate the microenvironment that BMAC cells experience when injected into the joint cavity. Thus, we cannot rule out that a different hydrogel formulation and/or flow rate will yield higher predictive power for the on-chip 3D simSF assay. Nevertheless, the improvement in cross-validation accuracy supports the adoption of on-chip 3D models for therapeutic development models.

The MILES clinical trial used the VAS pain score and KOOS pain score at 12 months versus baseline (before treatment) as primary clinical co-endpoints. Therefore, we next assessed correlations between donor-matched assay analyte secretion concentrations and changes in patient pain scores (VAS, KOOS) at 12 months post-treatment compared to baseline. For each of the 24 analytes, linear regression analyses were performed for changes (12 months vs. baseline) in VAS (Supplementary Fig. 9). VAS scores range from 0 to 100, with a higher VAS score indicating greater pain intensity. Significant linear correlations ($P < 0.05$) between assay secreted analyte concentration (independent variable) and change in VAS (dependent variable) were identified for PD-L1 in the on-chip 3D simSF platform and IL-17E and IL-12 p70 in the on-chip 3D ctrl platform; all three analytes exhibited a negative regression slope. No analytes were correlated with changes in VAS for the 2D simSF assay. KOOS pain scores range from 0 to 100, with 0 representing extreme knee problems and 100 representing no knee problems. Linear regression analyses between assay secreted concentrations (independent variable) and changes in KOOS (dependent variable) identified significant positive regression slopes for IL-17E and FAP in the on-chip 3D ctrl platform and

a negative regression slope for IL-8/CXCL8 for the 2D simSF assay (Supplementary Fig. 10). These evaluations identify in vitro assay metrics for cell products that correlate with matched patient outcomes.

Further analyses focused on the top k analytes that are most correlated with clinical outcomes and trained a multiple linear regression model using these k features. We repeated the analysis for $k = 1, 2, 3$ and for the outcome variable being change in VAS or change in KOOS, separately. The top k features (analytes), Pearson correlation coefficient, and P value of the trained linear model are listed in Tables 1, 2 for changes in VAS and KOOS as outcome, respectively. We observe that in all of the six cases (each value of $k \in \{1, 2, 3\}$) for the on-chip 3D ctrl platform and five out of six cases for the on-chip 3D simSF assay the markers for the on-chip 3D platform have a significant correlation ($P < 0.05$), whereas the 2D simSF markers do not correlate with pain scores ($P > 0.05$) for any of the six cases. The top features identified in this analysis have been previously identified as putative biomarkers in patient synovial fluid or implicated in OA progression. For instance, blockade of PD-L1 upregulates inflammatory responses, including TNF- α expression, and promotes the development of osteoarthritis in murine models³⁴. IL-17 exacerbates inflammation and cartilage destruction in OA and synovial fluid levels correlate with KL grade^{35,36}. VEGF, and HGF are implicated in angiogenesis and OA progression with elevated levels in the synovial fluid of OA patients^{33,37,38}. The TNF- α inflammatory pathway has been implicated in OA, and TNF- α has been classified as a biomarker in synovial fluid in various studies³⁹. Conversely, TNF R1 levels in synovial fluid are

Table 1 | Pearson correlation coefficient and *P* value (in parenthesis) for a multivariable regression model trained to predict the change in VAS scores (“outcome”) using the top *k* features most correlated with the outcome

<i>k</i>	On-chip 3D simSF	On-chip 3D ctrl	2D simSF
1	0.52 (0.0237) PD-L1	0.46 (0.0482) IL-17	0.20 (ns, 0.416) TNF R1
2	0.52 (0.0224) PD-L1, HGF	0.49 (0.0330) IL-17, VCAM-1	0.40 (ns, 0.0942) MMP-3, VEGF
3	0.56 (0.0130) TNF-α, IL-6, VCAM-1	0.53 (0.0197) HGF, IL-17, TNF R1	0.44 (ns, 0.0588) IL-6, MMP-3, VEGF

Results are shown for each of the assay platforms in separate columns and analytes corresponding to *k* features are listed below the Pearson correlation coefficient. A two-sided *t*-test was used. ns not significant. Statistically significant Pearson correlation coefficients are shown in bold.

Table 2 | Pearson correlation coefficient and *P* value (in parenthesis) for a multiple linear regression model trained to predict the change in KOOS scores (“outcome”) using the top *k* features most correlated with the outcome

<i>k</i>	On-chip 3D simSF	On-chip 3D ctrl	2D simSF
1	0.34 (ns, 0.149) VEGF	0.47 (0.0427) IL-17	0.40 (ns, 0.0920) IL-8
2	0.52 (0.0238) PD-L1, HGF	0.52 (0.0221) PD-L1, IL-17	0.42 (ns, 0.0731) MMP-3, IL-8
3	0.61 (0.0059) VEGF, CCL3/MIP-1α, VCAM-1	0.52 (0.0221) TNF-α, IL-17, VCAM-1	0.42 (ns, 0.0728) MMP-3, IL-8, VEGF

Results are shown for each of the assay platforms in separate columns and analytes corresponding to *k* features are listed below the Pearson correlation coefficient. A two-sided *t*-test was used; significant Pearson correlation coefficient are shown in bold. ns not significant.

negatively correlated with pain and self-reported physical function⁴⁰. In summary, the on-chip 3D platform provides improved predictive and correlative metrics of matched patient outcomes and identifies promising potency metrics for cell therapeutic products.

Discussion

The development of in vitro systems that better predict clinical outcomes would dramatically improve current efforts of therapeutic cell product development and translation. Regulatory agencies require potency metrics, i.e., quantitative measures of biologic function, as a measure of quality and consistency for product scale-up and release⁹. However, the identification of such potency metrics and surrogates is particularly challenging for cellular products because of unclear and/or pleiotropic mechanisms. Most potency assays involve evaluation of cell surface markers, secreted products, or transcriptomics for cells cultured on 2D supports (as reviewed in ref. 6). For example, in the case of mesenchymal stromal cells, the International Society for Cellular Therapy has endorsed suppression of T cell proliferation in mesenchymal stromal cell:peripheral blood mononuclear cell co-cultures as a functional metric for immunomodulatory potential⁴¹. This metric, however, has significant limitations as a scalable and reproducible potency assay due to the high variability of peripheral blood mononuclear cell donors and limitations on scalability of the assay. Alternative biomarkers, such as indoleamine 2,3-dioxygenase or programmed death-ligand 1, have been explored as surrogate metrics of potency in response to defined inflammatory stimuli, including IFN-γ^{6,7,42}. However, these surrogate metrics do not provide reliable predictors of clinical outcomes. Viswanathan and colleagues identified a panel of seven immunomodulatory markers in mesenchymal stromal cells that correlated with improved patient-reported outcomes in an OA trial with 12 patients⁴³, but other clinical studies failed to show a correlation between these cellular markers and graft-vs-host disease outcomes⁴⁴. Therefore, there still exists a significant need for rapid, reproducible, and scalable in vitro assays that correlate to clinical outcomes.

Because of the shortcomings of conventional 2D culture assays as predictive tools for cell potency, recent efforts have focused on the development and validation of 3D human organoid models to study the effects of genetic and functional background on the toxicity of anticancer therapy candidates^{12,13,45–49}. For example, Gjorevski et al. used co-cultures of patient-derived intestinal organoids and immune

cells to examine the in vitro toxicity of T-cell-engaging bispecific antibodies⁵⁰. This group demonstrated that bi-specific antibodies targeting epithelial cell antigens that showed acceptable safety profiles in simple in vitro and animal models but revealed off-tumor toxicities in phase 1 clinical trials induced immune cell-dependent killing of patient-derived intestinal organoids in the in vitro organoid platform. Although promising, these human organoid technologies are limited by high variability across patient-derived organoids, complex and lengthy assay times, and lack of scalability.

We previously engineered an on-chip microfluidic potency assay for mesenchymal stromal cells that maintains cells in 3D, within a soft hydrogel, and exposed to fluid flow²⁰. This on-chip microfluidic system provided improved recapitulation of secretory activities of human mesenchymal stromal cells implanted in immunodeficient mice compared to conventional 2D assays. Here, we adapted this on-chip microfluidic assay to evaluate BMAC used in a multicenter phase 3 clinical trial for OA. BMAC samples cultured in the on-chip 3D system exhibited elevated secreted levels of immunomodulatory and trophic proteins (cytokines, chemokines, cell adhesive proteins, MMPs) compared to 2D culture. Using analyte information from in vitro assays and patient-matched clinical data, we built linear regression prediction models for clinical outcomes. We demonstrate improved clinical prediction for the on-chip 3D system compared to the 2D culture assay. Furthermore, on-chip 3D assay metrics displayed higher correlative power with patient pain scores compared to the 2D assay. The on-chip microfluidic system has several advantages over other 2D culture and 3D organoid potency assays, including higher secreted product levels for increased sensitivity, faster analysis times, reduced costs, and fewer cells needed.

Scannell and Bosley proposed a quantitative decision-theoretic model for the drug discovery R&D process based on decision theory and decision analysis⁵¹. This model represents therapeutic candidates within a “measurement space” determined by the candidates’ performance on assays (decision tools) that correlate to a greater or lesser degree with clinical outcomes. The predictive validity of a decision tool is the degree to which its results correlate with metrics of clinical utility. These authors showed that, because of the compounding effects of true and false positive rates, decision tools with good predictive validity outperform simple, high-throughput methods in identifying successful therapeutic candidates. In this framework, predictive validity is operationalized as the Pearson coefficient between

the decision tool output and the measure of clinical utility. Notably, these authors found that the ability to identify good therapeutic candidates is extremely sensitive to predictive validity such that small changes in predictive validity (i.e., 0.1 absolute in correlation coefficient) can offset large (10–100X) changes in “brute-force” efficiency^{8,51}. In the present study, we demonstrate that the on-chip 3D assay exhibits increased predictive validity as determined by the Pearson correlation coefficient compared to the conventional 2D culture using BMAC secreted analyte levels as the decision tool results and VAS and KOOS patient scores as clinical outcomes. In particular, the on-chip 3D assay yielded statistically significant Pearson correlation coefficients ranging 0.46–0.61 whereas the 2D culture assay produced no statistically significant relationships. Furthermore, we applied trained linear regression prediction models and a train-test split procedure to assess the clinical prediction power of the in vitro 3D on-chip potency platform for responders and non-responders to the BMAC therapeutic intervention. When assessing cross-validation accuracy, we accounted for the imbalance between responders ($n = 15$) and non-responders ($n = 4$) by using balanced accuracy, ensuring that errors on the smaller non-responder group were appropriately weighted. Additionally, we repeated the cross-validation 100 times to provide confidence intervals and enable statistical comparison between platforms. We found a significant increase in clinical prediction by cross-validation accuracy for the on-chip 3D ctrl samples compared to 2D samples (0.61 vs. 0.48, $P < 0.0001$). These results motivate further development of the on-chip 3D platform as a decision tool for cell therapeutic candidates for OA.

The multicenter MILES clinical trial reported that cell therapies, including BMAC, were not superior to corticosteroid injection in alleviating OA knee pain at 1 year post-injection¹⁹. The primary outcomes of the MILES study were changes from baseline for VAS and KOOS pain scores at 12 months post-treatment, and we selected these outcomes for comparison with the in vitro potency platforms. Although many confounding factors influence pain sensation, VAS and KOOS are well accepted and widely used in the field. MRI is a powerful imaging modality to assess structural changes associated with OA, but the MILES study reported no significant changes in MRI scores (secondary endpoint) from baseline over 12 months for BMAC treatment. The authors of the MILES study posited that, because OA is a slow, progressive process, it is possible that a 1-year timeline was too short for this secondary measure. The lack of clinical superiority for cell therapies in the MILES study bolsters the motivation for developing more predictive assay metrics in order to assess patients, cell donors, and manufacturing techniques, prior to entering clinical studies. We envision wide-spread application of biomimetic systems that can be easily implemented with fast assay times for use at various steps through donor sourcing, development, and manufacturing. In vitro potency assays could replace animal studies to identify therapeutic candidates, inform manufacturers of not only of potency but also serve as a release assay for therapeutic products, or assist clinicians on making personalized decisions such as what type of cell therapy to use for different indications.

We note several limitations in the present study. First, our data set comprises only 22 BMAC donors with paired clinical outcomes. Increasing the number of donor samples is important to improving the sensitivity and predictive validity of the in vitro 3D on-chip potency platform, especially given the heterogeneity of BMAC across donors. Moreover, testing the prediction power of the model on an independent patient cohort not used to train the model will establish the robustness and general applicability of this decision tool. For clinical treatment in the MILES study, BMAC was collected and administered as fresh cells in saline via intra-articular injection into the OA-affected knee joint. The remaining BMAC (when available) was cryopreserved for later analyses. It is possible that there are differences in viability or secretory responses between the fresh BMAC sample delivered to the

patient and the thawed BMAC sample analyzed in the in vitro platforms. Despite these limitations, the major findings of our study are that the on-chip 3D potency assay has significantly improved clinical prediction power and higher correlative power with patient pain scores compared to the traditional 2D culture assay. The present study lays a solid foundation for follow-up studies with increased numbers of patient samples and a prospective study to assess the clinical predictive impact of this potency platform.

Methods

Ethical statement

This study complies with all relevant ethical regulations. Georgia Tech IRB issued a waiver (H19004) as no human subject work was performed at Georgia Tech for the present study. This study used de-identified BMAC samples from patients with varying severity of OA, all enrolled in the MILES clinical trial (NCT03818737; IRB001080460 from the Western Institutional Review Board, Emory University IRB, and Duke University IRB). All participants provided written informed consent. The clinical trial design for the MILES trial was performed independently and not affiliated with the present study. BMAC samples used for the present study were chosen by Marcus Center for Therapeutic Cell Characterization and Manufacturing staff based on the availability of BMAC samples not needed for the MILES trial and associated analyses.

Study design

The objective of this study was to evaluate an in vitro 3D microfluidic assay for clinical predictive power compared to conventional 2D culture methods for cell therapeutics. The overall hypothesis was that a 3D physiologically-relevant microfluidic assay elicits cellular responses (secreted factors) with improved predictive power of patient outcomes. We observed that the microfluidic on-chip 3D assay improved BMAC clinical prediction and correlations to OA outcomes over conventional 2D culture assays.

On-chip 3D set-up

Microfluidic chips (2.3 cm × 2.3 cm; 5 mm well diameter, Supplementary Fig. 1) of polydimethylsiloxane (PDMS, Sylgard 184 Dow Chemical) were cast in an aluminum mold with steel wire ($d = 0.02$ in) and bonded to glass slides with oxygen plasma. The steel wires were removed to form the fluidic channel. The chip and a thin top PDMS layer were treated by oxygen plasma, the cell-laden hydrogel was placed in the PDMS device, and the system was sealed. Control (“ctrl”) media (XSFM, Fujifilm) or media (XSFM, Fujifilm) supplemented with 10% simSF (“simSF”) was set up in bottles enclosed by rubber stoppers of inverted Eppendorf tubes. The microfluidic chip outlets were connected to PHD Ultra Pumps with attached 6/10 multi-racks (Harvard Apparatus) with PTFE tubing (#30 AWG, Cole Parmer), PE/PVC tubing adaptors (0.024 × 0.064 in, Instech Labs) and 20-gauge blunt tip needles (Industrial Dispensing Tips).

Hydrogel

20 kDa 4-arm maleimide-functionalized poly(ethylene-glycol) (PEG-4MAL, Laysan Bio) was dissolved in PBS at 6.8 mM. Adhesive peptide RGD (GRGDSPC, Genscript) was dissolved in 25 mM HEPES at 5 mM. Crosslinkers VPM peptide (GCRDVPMSMRGGDRCG, Genscript) and dithiothreitol (DTT, Sigma-Aldrich) were dissolved in 25 mM HEPES buffer at 24.8 mM and mixed 80:20 ratio by vol. Components were sterilized by filtration (Costar®, Spin-X®). PEG-4MAL and RGD components were mixed 2:1 vol ratio. Thawed BMAC cells were counted and added to the PEG-4MAL + RGD solution at 3:1 vol ratio for a final concentration of 80,000 cells/gel. Hydrogel gelation occurred by mixing PEG-4MAL + RGD + cell solution and crosslinking solution at 4:1 vol ratio for a 20 μ L hydrogel onto a sterilized hydrophobic surface.

On-chip 3D assay

After hydrogel synthesis, the hydrogel was placed into the microfluidic chip, sealed by oxygen plasma with a thin PDMS layer. Pumps were set to withdraw at $1.0 \mu\text{L min}^{-1}$ for 24 h. Cells were either treated with control media (XSF, Fujifilm) or media supplemented with 10% simSF. After 1 day of culture, media perfusate was collected for multiplex Luminex analysis, and cells were collected for flow cytometry analysis (where indicated). We note that the perfusate only has secreted factors; the cells remain encapsulated in the hydrogel. Each BMAC donor was run with 3–6 technical replicates, but on-chip samples with leaking and/or device clogging that resulted in $<0.5 \text{ mL}$ of collection were excluded from subsequent analysis (14/194, 7% of total on-chip samples).

2D culture assay

2D culture controls were run in parallel to on-chip 3D assays. Cells were seeded onto a tissue culture treated 96-well plate (Costar®) for an initial cell count of 80,000 cells/well. Cells were either treated with control basal media (XSF, Fujifilm) or media supplemented with 10% simSF. No media changes were performed to remove non-adherent cells. Following 24 h of culture, supernatant media was collected for multiplex Luminex analysis. Cells were isolated and analyzed by flow cytometry. Each BMAC donor was run in triplicate.

Viability assessment

The hydrogels were removed and placed in solution of collagenase I (0.2–0.5%), bovine serum albumin (0.1–0.2%), CaCl_2 (10 mM) dissolved in DI water for 30 min of incubation or until complete degradation with light agitation. Viability staining was performed using Zombie Live/Dead fixable stain (Biolegend). Human Fc-block (TruStain FcX, Biolegend) was added. BD Cytofix/Cytoperm was added per the manufacturer's instructions. Samples were run on a CytoFLEX S (Beckman Coulter; Pasadena, CA) and analyzed using FSC Express software v7 (De Novo Software; Glendale, CA).

Luminex assays

Custom made Luminex 24-plex panels were purchased from R&D Biotech and run per manufacturer's recommendations. Washing was performed using an automated 405 LS Washer (BioTek; Winooski, VT) and analyzed with MAGPIX® System (Luminex Corp; Austin, TX). Analyte-specific Luminex standards (R&D, Biotechne) were run in tandem on each run, resulting in a unique standard curve per analyte. Over 800 individual samples were tested, including nine 96-well plates of Luminex, and 4 unique lots of custom kits and/or standards. To best account for the minor differences in standard curves, analyte-specific curve fit types were matched across runs. Standard points were fit to five-point logarithmic curves, with exceptions of MMP-1 (cubic), CCL3 (five-point linear), IGFBP-rp1(cubic). Curve fits were chosen based on best-fit and run-to-run curve consistency.

Analyte normalization

We previously showed that normalization to experiment- and donor-matched 2D control samples yielded high consistency across repeated experiments compared to alternative methods investigated²⁰. To account for variations across experiments and analyses, the results (as indicated) are normalized to 2D controls based on expected total volume (on-chip 1.44 mL, 2D 0.2 mL) and divided by the average of 2D controls ($n = 3$ –6 replicates) for each experiment- and donor-matched sample.

Patient-derived synovial fluid (SF)

De-identified patient SF samples were thawed and centrifuged to $1000 \times g$ for 10 min remove any cells or debris. Following treatment with hyaluronidase (2 mg/mL) for 30 min, multiplex Luminex was performed using 30-plex inflammatory panel. For rheology, patient SF

samples were thawed and centrifuged to remove any cells or debris. Samples were tested using a 50 mm cone-and-plate 1° attachment (Anton Paar MCR302).

Statistics and reproducibility

De-identified BMAC samples were provided, and only donor sex, age, and KL score were provided with each patient sample. Clinical outcomes metrics were only made available following all sample testing and analyses. Clinical outcome metrics also remained blinded for the development of the predictive models. Each donor sample was tested across 3D on-chip simSF, 3D on-chip ctrl, 2D simSF, and 2D ctrl groups in tandem in 3 technical replicates or greater. 3D samples for which the microfluidic tubing clogged, and therefore had insufficient media collection, were excluded for secretion analyses. For linear regression analyses, data was analyzed using Prism 8 (GraphPad Software Inc., La Jolla, CA). P values represent confidence against a null hypothesis of slope deviation from zero. $P < 0.05$ was considered significant. Best fit lines are represented as solid lines. Hierarchical clustering, multivariate discriminant analysis, and correlation clusters were performed using JMP Pro 16.1 (JMP Software from SAS; Cary, NC). Hierarchical clustering used the Ward Method clustering method, with data standardized by analyte. Multivariate discriminant analysis used linear discriminant method with an ellipse representing estimated region to contain 50% of population.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data related to this paper are contained in the article, the Supplementary Information, and the Source Data file. Source data are provided with this paper.

Code availability

The script for Fig. 5 can be found at <https://github.com/bhavaygg/BMAC>.

References

- Pittenger, M. F. et al. Mesenchymal stem cell perspective: cell biology to clinical progress. *Regen. Med.* **4**, 22 (2019).
- Caplan, H. et al. Mesenchymal stromal cell therapeutic delivery: translational challenges to clinical application. *Front. Immunol.* **10**, 1645 (2019).
- Morgan, P. et al. Impact of a five-dimensional framework on R&D productivity at AstraZeneca. *Nat. Rev. Drug Discov.* **17**, 167–181 (2018).
- Emmerich, C. H. et al. Improving target assessment in biomedical research: the GOT-IT recommendations. *Nat. Rev. Drug Discov.* **20**, 64–81 (2021).
- Edwards, A. M. et al. Preclinical target validation using patient-derived cells. *Nat. Rev. Drug Discov.* **14**, 149–150 (2015).
- Robb, K. P., Fitzgerald, J. C., Barry, F. & Viswanathan, S. Mesenchymal stromal cell therapy: progress in manufacturing and assessments of potency. *Cytotherapy* **21**, 289–306 (2019).
- Yin, J. Q., Zhu, J. & Ankrum, J. A. Manufacturing of primed mesenchymal stromal cells for therapy. *Nat. Biomed. Eng.* **3**, 90–104 (2019).
- Scannell, J. W. et al. Predictive validity in drug discovery: what it is, why it matters and how to improve it. *Nat. Rev. Drug Discov.* **21**, 915–931 (2022).
- US Food and Drug Administration. Guidance for Industry Potency Tests for Cellular and Gene Therapy Products. (2011).
- Bravery, C. A. et al. Potency assay development for cellular therapy products: an ISCT review of the requirements and experiences in the industry. *Cytotherapy* **15**, 9–19 (2013).

11. FDA session on clinical evidence: BLA 125706: Remestemcel-L. Aug 20 (2020).
12. Ooft, S. N. et al. Patient-derived organoids can predict response to chemotherapy in metastatic colorectal cancer patients. *Sci. Transl. Med.* **11**, eaay2574 (2019).
13. Boretto, M. et al. Patient-derived organoids from endometrial disease capture clinical heterogeneity and are amenable to drug screening. *Nat. Cell Biol.* **21**, 1041–1051 (2019).
14. Beekhof, R. et al. Phosphoproteomics of patient-derived xenografts identifies targets and markers associated with sensitivity and resistance to EGFR blockade in colorectal cancer. *Sci. Transl. Med.* **15**, eabm3687 (2023).
15. Dekkers, J. F. et al. Characterizing responses to CFTR-modulating drugs using rectal organoids derived from subjects with cystic fibrosis. *Sci. Transl. Med.* **8**, 344ra384 (2016).
16. Chou, D. B. et al. On-chip recapitulation of clinical bone marrow toxicities and patient-specific pathophysiology. *Nat. Biomed. Eng.* **4**, 394–406 (2020).
17. Jang, K. J. et al. Reproducing human and cross-species drug toxicities using a Liver-Chip. *Sci. Transl. Med.* **11**, eaax5516 (2019).
18. Zushin, P. H., Mukherjee, S. & Wu, J. C. FDA Modernization Act 2.0: transitioning beyond animal models with human cells, organoids, and AI/ML-based approaches. *J. Clin. Investig.* **133**, e175824 (2023).
19. Mautner, K. et al. Cell-based versus corticosteroid injections for knee pain in osteoarthritis: a randomized phase 3 trial. *Nat. Med.* **29**, 3120–3126 (2023).
20. Schneider, R. S. et al. High-throughput on-chip human mesenchymal stromal cell potency prediction. *Adv. Healthc. Mater.* **11**, e2101995 (2022).
21. Chatterjee, P. et al. Single-cell transcriptome and crosstalk analysis reveals immune alterations and key pathways in the bone marrow of knee OA patients. *iScience* **27**, 110827 (2024).
22. Low, L. A. & Tagle, D. A. Tissue chips—innovative tools for drug development and disease modeling. *Lab Chip* **17**, 3026–3036 (2017).
23. Chary, S. R. & Jain, R. K. Direct measurement of interstitial convection and diffusion of albumin in normal and neoplastic tissues by fluorescence photobleaching. *Proc. Natl. Acad. Sci. USA* **86**, 5385–5385 (1989).
24. Schmid-Schonbein, G. W. Microlymphatics and lymph flow. *Physiol. Rev.* **70**, 987–987 (1990).
25. Polacheck, W. J., German, A. E., Mammoto, A., Ingber, D. E. & Kamm, R. D. Mechanotransduction of fluid stresses governs 3D cell migration. *Proc. Natl. Acad. Sci. USA* **111**, 2447–2452 (2014).
26. Pedersen, J. A., Boschetti, F. & Swartz, M. A. Effects of extracellular fiber architecture on cell membrane shear stress in a 3D fibrous matrix. *J. Biomech.* **40**, 1484–1492 (2007).
27. Tarbell, J. M. & Shi, Z. D. Effect of the glycocalyx layer on transmission of interstitial flow shear stress to embedded cells. *Biomech. Model. Mechanobiol.* **12**, 111–121 (2013).
28. Keeling, L. E. et al. Bone marrow aspirate concentrate for the treatment of knee osteoarthritis: a systematic review. *Am. J. Sports Med.* **50**, 2315–2323 (2022).
29. Liew, J. W. et al. A scoping review of how early-stage knee osteoarthritis has been defined. *Osteoarthr. Cartil.* **31**, 1234–1241 (2023).
30. Martire, A. et al. Mesenchymal stem cells attenuate inflammatory processes in the heart and lung via inhibition of TNF signaling. *Basic Res. Cardiol.* **111**, 54 (2016).
31. Van Hauwermeiren, F., Vandenbroucke, R. E. & Libert, C. Treatment of TNF mediated diseases by selective inhibition of soluble TNF or TNFR1. *Cytokine Growth Factor Rev.* **22**, 311–319 (2011).
32. Cuéllar, V. G., Cuéllar, J. M., Kirsch, T. & Strauss, E. J. Correlation of synovial fluid biomarkers with cartilage pathology and associated outcomes in knee arthroscopy. *Arthroscopy* **32**, 475–485 (2016).
33. Haraden, C. A., Huebner, J. L., Hsueh, M. F., Li, Y. J. & Kraus, V. B. Synovial fluid biomarkers associated with osteoarthritis severity reflect macrophage and neutrophil related inflammation. *Arthritis Res. Ther.* **21**, 146 (2019).
34. Liu, S., Mi, J., Liu, W., Xiao, S. & Gao, C. Blocking of checkpoint receptor PD-L1 aggravates osteoarthritis in macrophage-dependent manner in the mice model. *Int. J. Immunopathol. Pharm.* **33**, 2058738418820760 (2019).
35. Onishi, R. M. & Gaffen, S. L. Interleukin-17 and its target genes: mechanisms of interleukin-17 function in disease. *Immunology* **129**, 311–321 (2010).
36. Chen, B., Deng, Y., Tan, Y., Qin, J. & Chen, L. B. Association between severity of knee osteoarthritis and serum and synovial fluid interleukin 17 concentrations. *J. Int. Med. Res.* **42**, 138–144 (2014).
37. Anitua, E. et al. Relationship between investigative biomarkers and radiographic grading in patients with knee osteoarthritis. *Int. J. Rheumatol.* **2009**, 747432 (2009).
38. Mabey, T. et al. Angiogenic cytokine expression profiles in plasma and synovial fluid of primary knee osteoarthritis. *Int. Orthop.* **38**, 1885–1892 (2014).
39. Boffa, A. et al. Synovial fluid biomarkers in knee osteoarthritis: a systematic review and quantitative evaluation using BIPEDs criteria. *Cartilage* **13**, 82s–103s (2021).
40. Simão, A. P. et al. Soluble TNF receptors are produced at sites of inflammation and are inversely associated with self-reported symptoms (WOMAC) in knee osteoarthritis. *Rheumatol. Int.* **34**, 1759–1763 (2014).
41. Galipeau, J. et al. International Society for Cellular Therapy perspective on immune functional assays for mesenchymal stromal cells as potency release criterion for advanced phase clinical trials. *Cytotherapy* **18**, 151–159 (2016).
42. Levy, O. et al. Shattering barriers toward clinically meaningful MSC therapies. *Sci. Adv.* **6**, eaba6884 (2020).
43. Chahal, J. et al. Bone marrow mesenchymal stromal cell treatment in patients with osteoarthritis results in overall improvement in pain and symptoms and reduces synovial inflammation. *Stem Cell Transl. Med.* **8**, 746–757 (2019).
44. von Bahr, L. et al. Long-term complications, immunologic effects, and role of passage for outcome in mesenchymal stromal cell therapy. *Biol. Blood Marrow Transpl.* **18**, 557–564 (2012).
45. He, X. et al. Patient-derived organoids as a platform for drug screening in metastatic colorectal cancer. *Front. Bioeng. Biotechnol.* **11**, 1190637 (2023).
46. Dekkers, J. F. et al. Uncovering the mode of action of engineered T cells in patient cancer organoids. *Nat. Biotechnol.* **41**, 60–69 (2023).
47. Mo, S. et al. Patient-derived organoids from colorectal cancer with paired liver metastasis reveal tumor heterogeneity and predict response to chemotherapy. *Adv. Sci.* **9**, e2204097 (2022).
48. Ding, S. et al. Patient-derived micro-organospheres enable clinical precision oncology. *Cell Stem Cell* **29**, 905–917.e906 (2022).
49. Kim, M. et al. Patient-derived lung cancer organoids as in vitro cancer models for therapeutic screening. *Nat. Commun.* **10**, 3991 (2019).
50. Harter, M. F. et al. Analysis of off-tumour toxicities of T-cell-engaging bispecific antibodies via donor-matched intestinal organoids and tumouroids. *Nat. Biomed. Eng.* **8**, 345–360 (2024).
51. Scannell, J. W. & Bosley, J. When quality beats quantity: decision theory, drug discovery, and the reproducibility crisis. *PLoS ONE* **11**, e0147215 (2016).

Acknowledgements

Funding provided by the US Food and Drug Administration (EETWP#45), the National Institutes of Health (R01 ARO62368), and the National Science Foundation Engineering Research Center on Cell Manufacturing Technologies (EEC 1648035). R.S.S. and E.B.N. received funding from the National Science Foundation Graduate Fellowship Program. A.B.W.,

H.S., L.K., and K.R. were also supported by the Marcus Center for Therapeutic Cell Characterization and Manufacturing through generous funding from the Billie and Bernie Marcus Foundation and the Georgia Research Alliance. We thank the patients, clinicians, faculty, and study coordinators for the MILES clinical trial, which provided samples of BMAC and pdSF.

Author contributions

R.S.S. planned and performed experiments, performed all analysis, drafted manuscript. E.B.N. performed experiments. L.E.K. supported all sample collection and communication with MILES trial, coordinated transfer of patient samples and data, supported development of simSF, and performed patient SF Luminex assays. A.C.B.W., H.Y.S., L.K., K.R. supported development of the simSF, supported transfer of patient samples and data and all sample collection and communication with MILES trial, and performed patient SF Luminex assays and data analysis. K.R. provided communication with MILES trial as part of the study leadership. B.A., S.S. provided initial python code and further support and guidance for the clinical prediction model. H.D., K.M., S.D.B. oversaw clinical trial. W.A.L. provided guidance on the application of the microfluidic device and provided input on the experimental design. A.J.G. oversaw all activities, provided input on experimental design and analyses, and drafted manuscript.

Competing interests

A.J.G., R.S.S., and W.A.L. are co-inventors on a patent application filed by the Georgia Tech Research Corp. on the on-chip 3D potency assay. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-60158-w>.

Correspondence and requests for materials should be addressed to Andrés J. García.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025