

# An Appearance Data-Driven Model Visualizes Cell State and Predicts Mesenchymal Stem Cell Regenerative Capacity

Di Wu, Lu Zhao, Bingdong Sui, Lingping Tan, Lu Lu, Xueli Mao, Guiqing Liao, Songtao Shi, Yang Cao,\* Xiaobao Yang,\* and Xiaoxing Kou\*

Mesenchymal stem cells (MSCs) are widely used in treating various diseases. However, lack of a reliable evaluation approach to characterize the potency of MSCs has dampened their clinical applications. Here, a function-oriented mathematical model is established to evaluate and predict the regenerative capacity (RC) of MSCs. Processed by exhaustive testing, the model excavates four optimal fitted indices, including nucleus roundness, nucleus/cytoplasm ratio, side-scatter height, and ERK1/2 from the given index combinations. Notably, three of them except ERK1/2 are cell appearance-associated features. The predictive power of the model is validated via screening experiments of these indices by predicting the RC of newly enrolled and chemical inhibitor-treated MSCs. Further RNA-sequencing analysis reveals that cell appearance-based indices may serve as major indicators to visualize the results of integration-weighted signals in and out of cells and reflect MSC stemness. In general, this study proposes an appearance data-driven predictive model for the RC and stemness of MSCs.

in a variety of human tissues, including bone-marrow (BMMSC), adipose tissue (ASC), umbilical cord (UCMSC), gingiva (GMSC), cartilage tissue (CMSC), skin tissue (SSC), dental pulp (DPSC), and others, respectively. MSCs have self-renewal capabilities, and they are multipotent cells that can differentiate into several tissue-specific lineages, including osteoblasts, chondrocytes, adipocytes, hepatocytes, myocytes, and neuron-like cells among others.<sup>[1]</sup>

MSCs possess potent and broad therapeutic potential. There are currently over 1300 clinical trials reported worldwide at ClinicalTrials.gov, which apply transplanted MSCs to treat different diseases, including bone/cartilage repair, graft-versus-host disease, systemic lupus erythematosus, diabetes, as well as in regenerative medicine. An even greater translational promise is now achieving with MSCs becoming a growing global industry. The criteria for defining multipotent stem cells and the guidelines for clinical translation of human MSCs have been reviewed and

erative medicine. An even greater translational promise is now achieving with MSCs becoming a growing global industry. The criteria for defining multipotent stem cells and the guidelines for clinical translation of human MSCs have been reviewed and

## 1. Introduction

Mesenchymal stem cells (MSCs) are plastic-adherent, fibroblast-like, nonhematopoietic progenitor cells, which are present

D. Wu, L. Zhao, B. Sui, L. Tan, L. Lu, X. Mao, S. Shi, X. Kou  
Hospital of Stomatology  
Guanghua School of Stomatology  
South China Center of Craniofacial Stem Cell Research  
Guangdong Provincial Key Laboratory of Stomatology  
Sun Yat-sen University  
Guangzhou 510055, China  
E-mail: kouxiaoxing@mail.sysu.edu.cn


D. Wu, G. Liao  
Hospital of Stomatology  
Guanghua School of Stomatology  
Department of Oral and Maxillofacial Surgery  
Guangdong Provincial Key Laboratory of Stomatology  
Sun Yat-sen University  
Guangzhou 510055, China

D. Wu, Y. Cao  
Hospital of Stomatology  
Guanghua School of Stomatology  
Department of Orthodontics  
Guangdong Provincial Key Laboratory of Stomatology  
Sun Yat-sen University  
Guangzhou 510055, China  
E-mail: caoyang@mail.sysu.edu.cn

B. Sui  
State Key Laboratory of Military Stomatology  
National Clinical Research Center for Oral Diseases  
Shaanxi International Joint Research Center for Oral Diseases  
Center for Tissue Engineering  
School of Stomatology  
The Fourth Military Medical University  
Xi'an 710032, China

S. Shi, X. Kou  
Key Laboratory of Stem Cells and Tissue Engineering  
Ministry of Education  
Sun Yat-sen University  
Guangzhou 510080, China

X. Yang  
School of Physics and Optoelectronics  
South China University of Technology  
Guangzhou 510640, China  
E-mail: scxyang@scut.edu.cn

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/smt.202200087>.

DOI: 10.1002/smt.202200087

revised constantly with a deeper understanding of the characteristics of human MSCs.<sup>[2]</sup> However, the therapeutic efficacy of MSCs in the different donor batches or scales of clinical trials was always not quality-controlled, which hinders efficient translational application.<sup>[3]</sup> Furthermore, MSCs are heterogeneous cells reflected by cell morphology and functional properties. The heterogeneity of MSCs constitutes an important barrier to MSC-based clinical use at multiple levels, including but not limited to donor variance, tissues from different sources, cell passage numbers, differential cell isolation techniques, and diverse cell culture and preservation conditions.<sup>[3c,4]</sup> Current guidelines for clinical translation of human MSCs recommend that surrogate biomarkers related to target biological effects and the methods of rapid tests used for quality control of stem cell preparation should be proposed.

Recently studies have shown the correlation between cell morphological signatures with the differentiation or immunosuppressive capacity of MSCs, which indicates that cell morphological signatures possess the possibility to predict the running state of the cell system.<sup>[5,6]</sup> Cellular morphological features were modified by high-content imaging methodology or induced cell differentiation with interferon- $\gamma$  to predict immunosuppressive capacity or differentiation fate of BMMSCs.<sup>[5]</sup> MSCs from bone marrow were often selected as the only cell origin and cellular-level morphological analysis was considered as the main predictive parameter in previous studies. Nevertheless, considering that MSCs contain information about the complex living cell states, lack of a comprehensive understanding and deep evaluation of them is one of the leading reasons for failure to break through the bottleneck of current stem cell therapy.

In this study, we enrolled eleven different types of human MSCs to profile their basic characteristics by screening thirty biological indices. Processed by exhaustive testing, we construct an appearance data-driven mathematical model that excavated four indices, including nucleus roundness, nucleus/cytoplasm ratio, side-scatter height, and ERK1/2 for evaluating *in vitro* regenerative capacity (RC) of MSCs.

## 2. Results

### 2.1. Profiling Characteristics of Eleven Different Types of MSCs by Screening Thirty Biological Properties

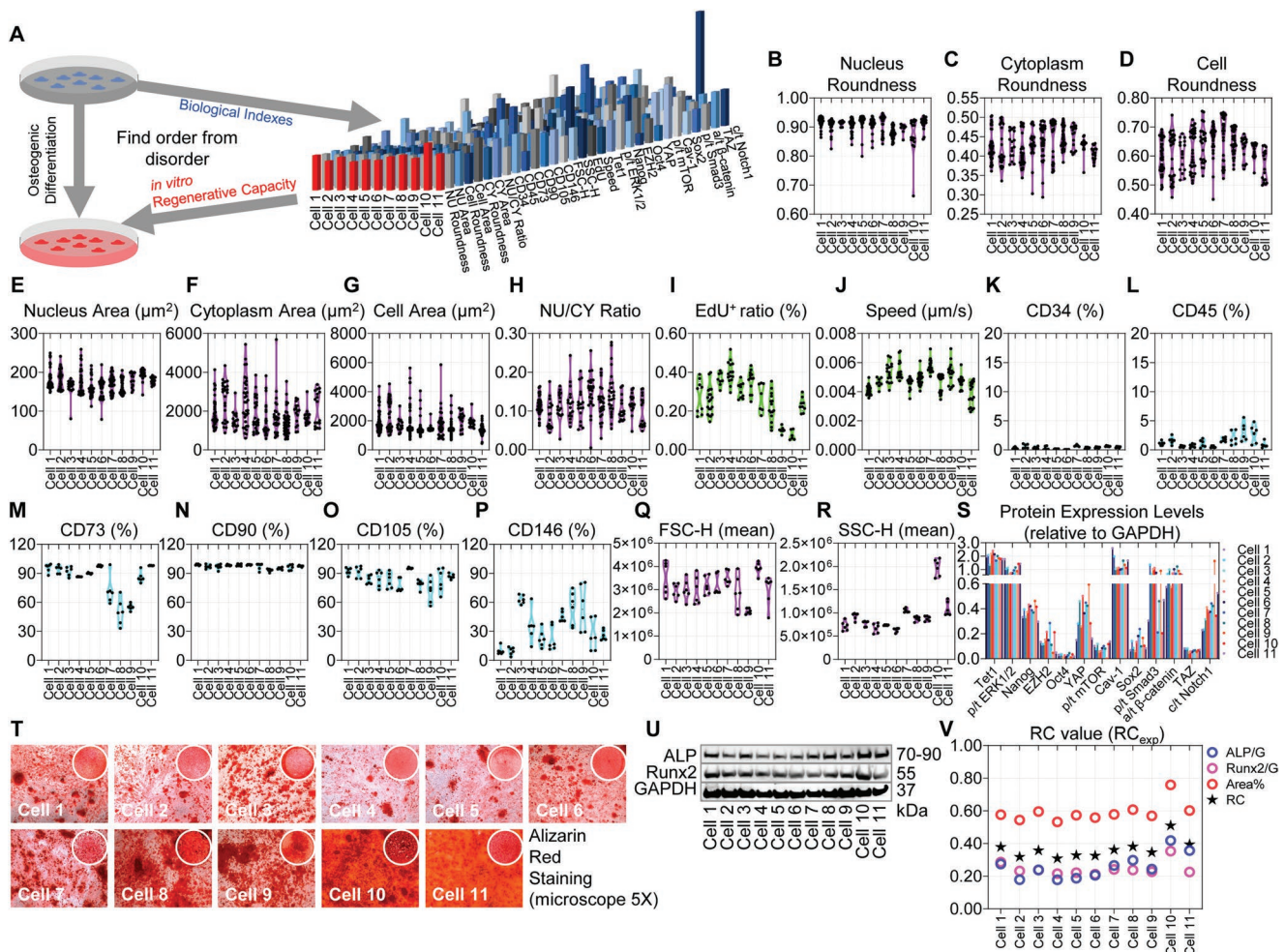
In order to develop a reliable and general evaluation model to characterize the potency of different types of MSCs, we considered the heterogeneity of MSCs and enrolled eleven different types of MSCs named Cell 1 to 11 based on three main aspects that led to heterogeneity, including donor variance, tissues from different sources, and different cell passage numbers. Specifically, these MSCs are hGMSC, SHED, DPSC, hUCMSC, and hSMSC from different donor sources, hBMMSC and hASC from commercial cell products, hGMSC from the same donor with different passages, and SHED or hUCMSC from different donors with the same R1passage (Figure 1A, see Table S1 of the Supporting Information for MSC details). First, cellular morphology is one of the significant indicators of cell status,<sup>[5,6]</sup> the morphological indicators (including cell area, cell roundness, nucleus area, nucleus roundness, cytoplasm area, cytoplasm

roundness, and the ratio of nucleus/cytoplasm) of eleven types of MSC in a living cell state were automatically extracted from images which were taken and analyzed using a high-content imaging system (Figure 1B–H). Compared to the morphology of dead MSCs after fixation (Figure S1A, Supporting Information), most of the living stem cells were homogeneous fibroblastoid cellular morphology with a smooth border and approximately circular nucleus (Figure S1B,C, Supporting Information). Cell proliferation rates were estimated by 5-ethynyl-20-deoxyuridine (EdU) incorporation (Figure 1I; Figure S1D, Supporting Information). The migration speed of label-free living MSCs was also monitored by a high-content imaging system with a digital phase contrast model and calculated as the total migrated distance divided by the duration of migration time, not relative displacement (Figure 1J). The relative displacement of all MSCs migration is less than 10  $\mu\text{m}$  within 24 h (Figure S1E, Supporting Information). Then, single-cell analysis of eleven types of MSC physical properties and surface markers was conducted by flow cytometry. All these eleven types of MSC were positive for MSC surface markers CD73, CD90, CD105, and CD 146 and negative for hematological markers CD34 and CD45 (Figure 1K–P; Figure S1F, Supporting Information). Parameters of flow cytometry reflecting the relative size and internal complexity of the suspension living cells were collected and recorded as forward-scatter height (FSC-H) and side-scatter height (SSC-H), respectively (Figure 1Q,R; Figure S1G, Supporting Information). Finally, protein expression levels of thirteen molecular signals essential for stem cell function, including Tet1, phospho/total ERK1/2, Nanog, EZH2, Oct4, YAP, phospho/total mTOR, Caveolin-1, Sox2, phospho/total Smad3, active/total  $\beta$ -catenin, TAZ, and cleaved/total Notch1 were detected in eleven types of MSC by western blot and calculated by relative expression levels to GAPDH (Figure 1S; Figure S2A, Supporting Information).

To evaluate the RC of MSCs, the alizarin red staining of mineralized nodules formation, and western blot analysis of relative osteogenic lineage protein (ALP and Runx2) were performed after osteogenic induction (Figure 1T,U). The RC values of eleven types of MSC were calculated by average values of alizarin red staining mineralized area percentages and osteogenic protein expression levels (Figure 1V).

### 2.2. Development of a Predictive Model for RC-Valued Assessment

Based on the results described above, characteristics of the eleven types of MSC were screened eligible and integrated into a final dataset, which consisted of thirty different indices being input  $x$  and actual RC values ( $RC_{\text{exp}}$ ) being output  $y$  (see Table S2, Supporting Information, for the raw data of  $x$  and  $y$ ). A given level of input  $x$  elicited corresponding output  $y$  indicating predictive RC values ( $RC_{\text{pred}}$ ). Our predictive equation (Eq.) model was constructed by considering the set of indices to reach the optimal fits. Specifically, eleven types of MSC were divided randomly with a split ratio of 8:3 into eight MSCs in the training set and the remaining three in the testing set. Relevant values of all thirty indices corresponding to 20 training/testing sets randomly selected from all sample-split

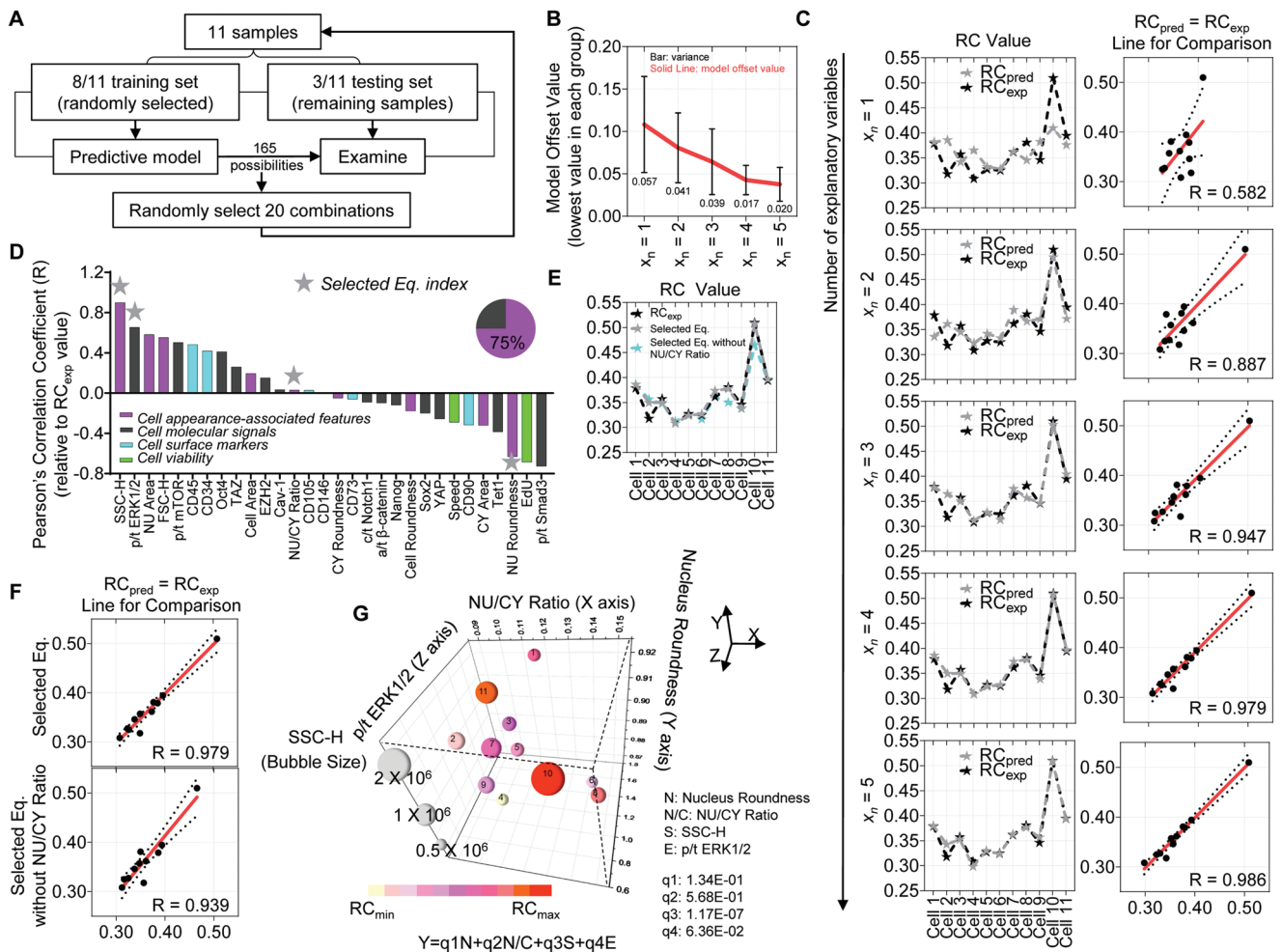


**Figure 1.** Profiling mesenchymal stem cells (MSCs) from different sources by experimental screening of thirty biological indices. A) Schematic diagram of the screening process. Eleven groups of MSCs, labeled as Cell 1 to 11, from different sources and stages, including bone marrow, adipose tissue, umbilical cord, dental pulp, and others (see Table S1, Supporting Information, for MSC details), were enrolled for the experimental screening of thirty biological indices. B–J) High-content imaging system-based screening of the morphology, migration speed, and proliferation of eleven MSCs (see Figure S1 for details). K–R) MSC surface markers, forward-scatter height (FSC-H), and side-scatter height (SSC-H) analysis by single-cell analysis (see Figure S1 for details). S) Screening of protein expression of multiple molecular signals (see Figure S2 for details). T) Alizarin red staining of eleven MSCs showing the capacity to form mineralized nodules after 12-day osteogenic induction. The upper right circles represent the entire field of alizarin red staining. U) Western blot analysis of eleven MSCs showing the expression levels of the osteogenic lineage protein ALP and Runx2. GAPDH was used as a protein loading control. V) Evaluation of the regenerative capacity (RC) of eleven MSCs. Alizarin red-positive area and osteogenic protein expression were analyzed using ImageJ software (NIH). The results were shown as a percentage of alizarin red-positive area over the total area (red circles) and relative density to the loading control (blue and purple circles), respectively. The RC value ( $RC_{exp}$ , black pentagrams) was calculated by average alizarin red-positive area and osteogenic protein expression. For all violin plots presented, each dot represented one replicate ( $n \geq 6$ ).

possibilities were calculated by fitting tests (Figure 2A). The offset values were reduced with the increasing number of explanatory variables  $x$ , which illustrated that increasing explanatory variables from  $x_n = 1$  to 5 improved the Eq. model fit quality (Figure 2B). However, if the number of explanatory variables was over half of the training set, in this case,  $x_n \geq 5$ , the model easily falls into overfitting. Specifically, in multiple linear regression, the variance range increased when  $x_n = 5$ , although the offset value slightly decreased (Figure 2B). As the  $x$  number increased from 4 to 5, the differences of their offset values were less than 0.01. Meanwhile, the smallest variance range of the predictive model appeared in the case of  $x_n = 4$  but not  $x_n = 5$  (Figure 2B). These data indicated that the explanatory

variables  $x_n = 4$  could serve as a potential combination for the predictive model.

For all the eleven types of MSC, their  $RC_{pred}$  values calculated by different optimal fit models with increasing explanatory variables  $x$  ( $x_n = 1$  to 5) were compared with their corresponding  $RC_{exp}$ . As the  $x$  number increased, the gray broken lines, connecting the  $RC_{pred}$  of Cell 1 to 11, closely overlapped the black broken lines representing their corresponding  $RC_{exp}$  (Figure 2C). Moreover, Pearson's correlation analysis showed that the coefficients between  $RC_{pred}$  versus  $RC_{exp}$  increased from 0.582 to 0.986 as the  $x$  number increased from 1 to 5 (Figure 2C). However, the coefficients between  $RC_{pred}$  versus  $RC_{exp}$  only increased from 0.979 to 0.986 as the number of input  $x$  was



**Figure 2.** A predictive model for RC-valued assessment fitted by eleven MSC biological properties as a training data set. A) Schematic diagram of the algorithm. B) Increasing the number of explanatory variable  $x$  (input indices) from 1 to 5 leads to a decrease of the corresponding mean offset values. C) Increasing the number of explanatory variable  $x$  from 1 to 5 enhances the overlaps of predictive regenerative capacity ( $RC_{pred}$ , gray pentagrams) values with actual osteogenic potential (experiment RC value,  $RC_{exp}$ , black pentagrams), and increases linear correlation of  $RC_{pred}$  versus  $RC_{exp}$ . Red lines represent linear regression fits for the RC data points and the  $R$  values of the fits represent Pearson correlation. Each black dot indicated one of the eleven MSCs. D) Pearson's correlation coefficients for pairs of thirty input  $x$  indices and output  $y$  values ( $RC_{exp}$  as the output  $y$ ). Gray pentagrams represent the selected optimal four indices combination for the selected equation (Eq.) model. Among numerous four input  $x$  combinations, the selected indices, including nucleus roundness, nucleus/cytoplasm ratio, SSC-H, and ERK1/2, are mainly cell appearance-associated indices. E) Excluding the NU/CY ratio from the selected Eq. model leads to an increased deviation between  $RC_{pred}$  and  $RC_{exp}$  values. The comparison of actual RC values versus predicted RC values was calculated using the selected Eq. model that fitted with (selected Eq.) or without NU/CY ratio (selected Eq. without NU/CY ratio). F) Excluding the NU/CY ratio from the selected Eq. model leads to the decrease of  $R$  values of the fit representing Pearson correlation. Each black dot indicated one of the eleven MSCs. G) A 3D graph represents the selected Eq. model using four Eq. indices to predict RC values. SSC-H, bubbles; nucleus roundness, Y-axis; NU/CY ratio, X-axis; p/t ERK1/2, Z-axis. The size of the bubbles represents the value of SSC-H. The color scale showed the value range of  $RC_{pred}$ .

added from four to five, suggesting that the increased predictive cost could not markedly increase the efficacy of the predictive model in this case (Figure 2C). These data indicated that the selected Eq. model fitted by four input  $x$  indices was optimal. Among numerous four input  $x$  sets, the optimal four indices combination for the selected Eq. model included nucleus roundness, nucleus/cytoplasm ratio, SSC-H, and ERK1/2 (Figure 2D; Figure S3, Supporting Information). Notably, three of the four selected indices except ERK1/2 are cell appearance-associated features rather than surface markers or molecular signals (Figure 2D).

Pearson's correlation coefficients for pairs of thirty input  $x$  indices and output  $y$  were showed to represent the degree of correlation (Figure 2D). Interestingly, NU/CY ratio, as one of four selected Eq. indices, showed a very low correlation with  $RC_{exp}$  value ( $R = 0.0298$ ). When the Eq. model was fitted by the other three indices excluding NU/CY ratio, the deviation between  $RC_{pred}$  and  $RC_{exp}$  values was larger than it between  $RC_{pred}$  obtained by the selected Eq. model and  $RC_{exp}$  values (Figure 2E). Pearson's correlation coefficients also decreased from 0.979 to 0.939 (Figure 2F). The above results obviously illustrated that although the low correlation of nucleus/cytoplasm

ratio index with  $RC_{exp}$  values was calculated, it was expected to have a non-negligible contribution to the accuracy of the selected Eq. model.

Overall, we developed a data-based multiple linear regression model that excavated four indices, including nucleus roundness, nucleus/cytoplasm ratio, side-scatter height, and ERK1/2, which were indispensable, balanced, and compensated for each other to predict the RC of MSCs in vitro. Our selected optimal combination is listed as Eq. below

$$Y = q1N + q2N/C + q3S + q4E(q1: 1.34E-01, q2: 5.68E-01, q3: 1.17E-07, q4: 6.36E-02) \quad (1)$$

The selected Eq. was demonstrated using a 5D graph (Figure 2G).

### 2.3. Experimental Validation of the Predictive Model via MSC Screening

To validate our predictive model of the RC value assessment, screening experiments of the selected four indices combination by predicting the RC of newly enrolled MSCs were conducted (Figure 3A). Three different types of MSCs from different donors or tissues were cultured and named Cell A, Cell B, and Cell C (see Table S1, Supporting Information, for details). To calculate the  $RC_{exp}$  values, the alizarin red assay and western blot analysis of ALP and Runx2 in Cell A, B, and C were performed under osteogenic inductive conditions. The results suggested that the osteogenic ability of Cell B was greater than Cell A and C (Figure 3C). According to our selected Eq., the ranking of the  $RC_{prep}$  values of Cell A, B, and C exhibited consistent with the actual ranking of  $RC_{exp}$  values, although the predicted values are a little smaller than the actual (Figure 3D). The  $RC_{exp}$ -to- $RC_{pred}$  value deviations may be due to the different conditions between different batches of experiments. The deviations of the  $RC_{exp}$ -to- $RC_{pred}$  value did not affect the ranking of the  $RC_{prep}$  values for the same batch of cells (Figure 3D,E).

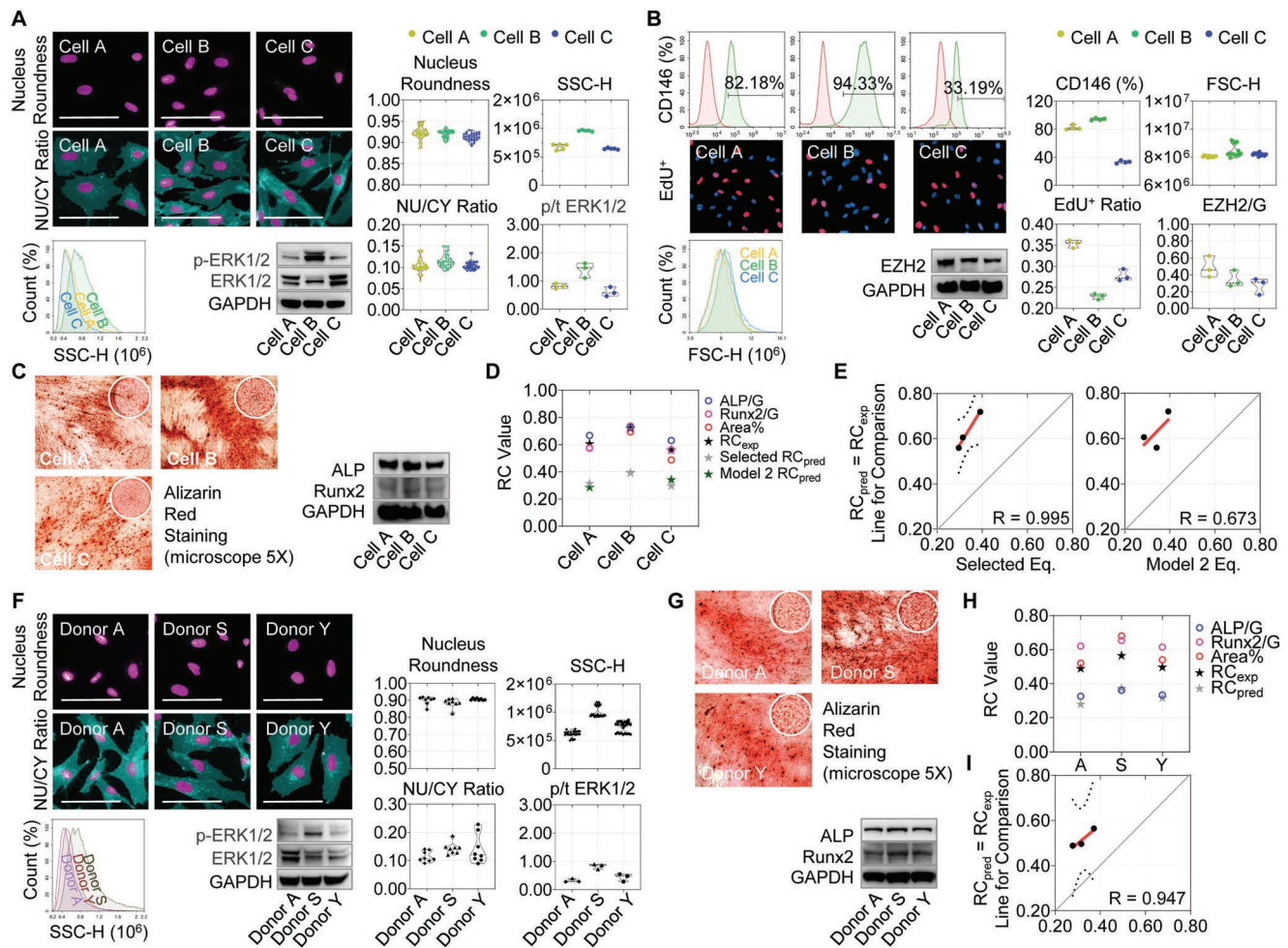
Meanwhile, all top thirty-four-index fitted combinations containing at least one index among nucleus roundness, NU/CY ratio, SSC-H, and ERK1/2, which almost specifically related to selected Eq. indices, except one combination of CD146, FSC, EDU, and EZH2 (Figure S3, Supporting Information). Then we named the model fitted by CD146, FSC, EDU, and EZH2 as Model 2 and it was also verified by Cell A, B, and C screens (Figure 3B). We found that the predicted ranking of Cell A and C was a little discrepancy compared with the actual result (Figure 3D). Furthermore, the linear correlation analysis showed that Pearson's correlation coefficients between  $RC_{exp}$  versus  $RC_{pred}$  obtained by our predictive model and Model 2 greatly decreased from 0.995 to 0.673 when the Model 2 without any selected Eq. indices was used (Figure 3E). From another perspective, the result suggested the importance of cell appearance indices for the predictive Eq. model.

As reviewed in the source information, Cell A was P3 UCMSCs derived from Donor Y and Cell B was P11 UCMSCs derived from Donor S (Table S1, Supporting Information).

Rather unexpectedly, Cell B with higher passage numbers (P11) marked higher osteogenic differentiation capability rather than Cell A with lower passage numbers (P3) (Figure 3C,D; Table S1, Supporting Information). These results indicate that stem cell donor conditions play an important role in the processes of self-renewal and differentiation. Therefore, a new type of UCMSC derived from a new donor, named Donor A, was enrolled to experimentally validate the accuracy of our selected Eq. model. UCMSCs derived from Donor A, S, and Y at the same passage numbers (P7) were used for the prediction, with donor source as the only variable. Then four selected indices were screened and substituted into the selected Eq. (Figure 3F). The result showed that the  $RC_{pred}$  value of Donor S was higher than Donor A/Y, and the predictive ranking matched perfectly with the actual ranking according to the results of the alizarin red assay and western blot (Figure 3G,H). The Pearson's correlation coefficient was 0.947 (Figure 3I). These results suggested that this appearance data-driven model can accurately predict the RC of MSCs rather than simply using the passage number.

### 2.4. Predicting the RC Outcomes of MSCs Treated by Chemical Inhibitors

It is well-known that Wnt/ $\beta$ -catenin signaling pathway is critically involved in regulating cell stemness and osteogenic ability.<sup>[7]</sup> It is worthy to explore whether the actual RC values were affected by inhibition of Wnt/ $\beta$ -catenin signaling pathway in vitro, and whether our prediction accuracy and sensitivity would be affected as well. We confirmed that treatment by XAV939, a specific inhibitor of Wnt/ $\beta$ -catenin signaling pathway, decreased the expression of  $\beta$ -catenin in Cell A, B, and C (Figure S2B, Supporting Information).<sup>[8]</sup> Most previous studies reported that the inhibition of Wnt/ $\beta$ -catenin resulted in a negatively regulated osteogenic differentiation. However, we could not ignore the different cell states that may respond inconsistently to the signaling pathway.<sup>[9]</sup> In our study, Cell A, B, and C were treated with increasing concentrations of XAV939 ( $1 \times 10^{-6}$ ,  $5 \times 10^{-6}$ , and  $10 \times 10^{-6}$  M) for 24 h and then tested the four selected Eq. indices to calculate the  $RC_{pred}$  values. As Wnt/ $\beta$ -catenin signaling pathway was inhibited, nucleus roundness and nucleus/cytoplasm ratio of Cell A/B/C decreased with increasing concentration of XAV939 (Figure 4A). The SSC-H of Cell A and Cell B decreased and Cell C increased with an increased inhibitor concentration (Figure 4A). We also found that ERK1/2 expression in Cell A and Cell C increased in a XAV939 dose-dependent manner, but ERK1/2 expression in Cell B decreased at first and then increased with increasing doses of XAV939 (Figure 4A). Substituting the corresponding values, the  $RC_{pred}$  values of Cell A and Cell B decreased and of Cell C showed no apparent decrease, even a slight increase with XAV939 concentration increasing (Figure 4C). The general changing trends of the  $RC_{exp}$  values of Cell A/B/C were consistent with the  $RC_{pred}$  values, except that Cell C death was observed at high concentration of  $10 \times 10^{-6}$  M XAV939 under prolonged osteoinductive culture (Figure 4B,C). The Pearson's correlation coefficients of Cell A/B/C were 0.979, 0.984, and 0.986, respectively (Figure 4D). These data indicate that even

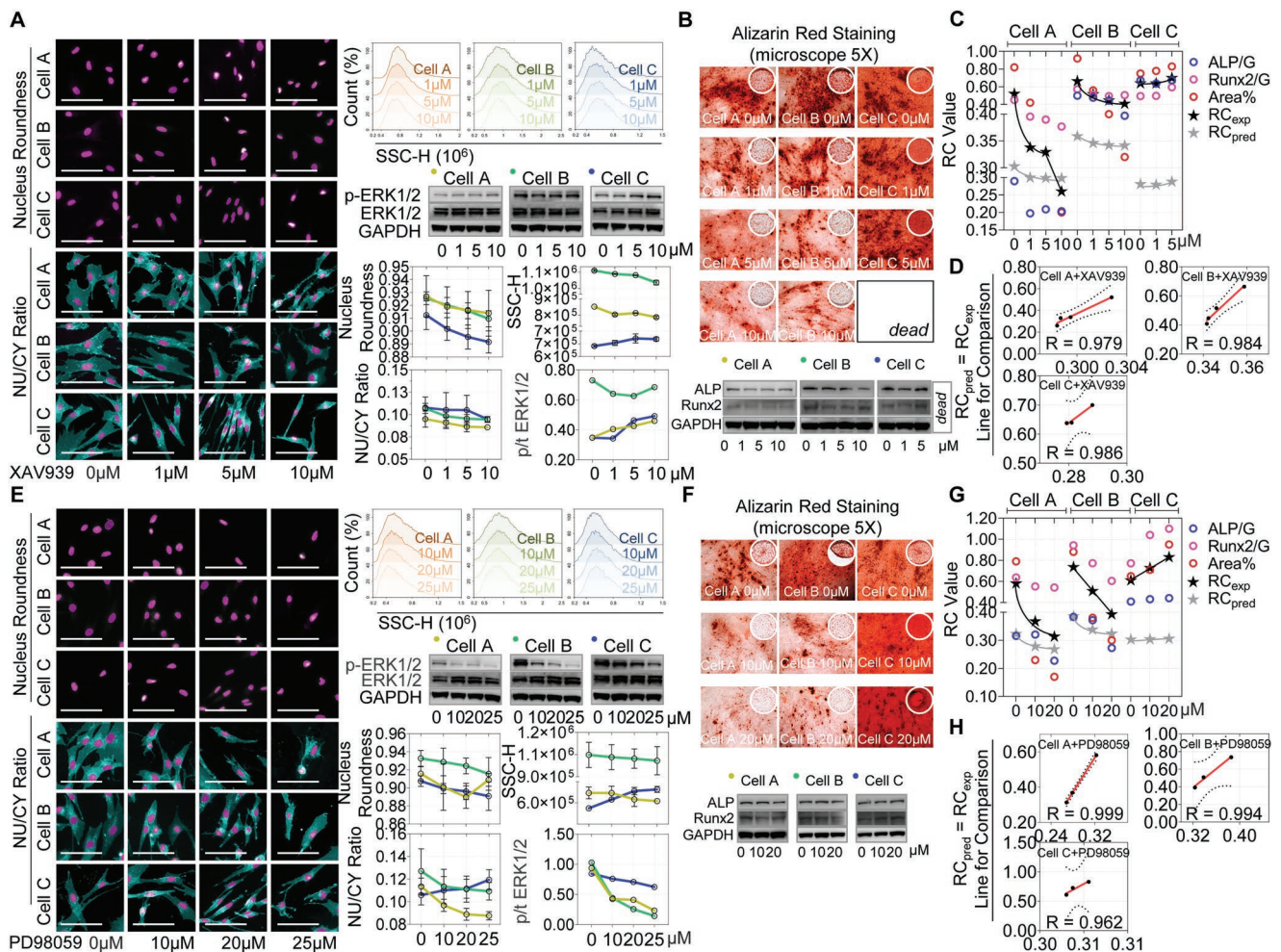


**Figure 3.** Experimental validation of the predictive model by predicting RC of newly enrolled MSCs. A) Experimental screening of the four selected Eq. indices in newly enrolled MSCs from different stages and tissues. Three different MSCs named Cell A, B, and C were enrolled in (A) to (E). Detailed information about these cells was provided in Table S1 of the Supporting Information. B) A model fitted by CD146, FSC, EDU, and EZH2 was named as Model 2 and also verified by Cell A, B, and C screens. These four indices, including CD146, FSC-H, EdU<sup>+</sup>, and EZH2, did not contain any index related to selected Eq. indices. See Figure S3 for details. C) Actual RC of newly enrolled MSCs were detected by alizarin red staining and western blot. D) Validation of the predictive model by predicting RC of newly enrolled MSCs. The accuracy of two models using four selected Eq. indices ( $RC_{pred}$ ) or four Model 2 indices (Model 2  $RC_{pred}$ ) is validated by predicting the RC of newly enrolled MSCs. E) Linear correlation of  $RC_{exp}$  versus  $RC_{pred}$  values obtained by selected Eq. model or Model 2 in newly enrolled MSCs. Red lines represent linear regression fits for the RC data points.  $R$  values of the fit represent Pearson correlation. F) Experimental screening of the four selected Eq. indices in newly enrolled MSCs derived from umbilical cord (UCMSCs) of three different donors. Three different UCMSCs at the same passage named Donor A, S, and Y were enrolled in (F) to (H). Detailed information about these cells was provided in Table S1 of the Supporting Information. G) Actual RC values of three different UCMSCs at the same passage were detected by alizarin red staining and western blot. H) Validation of the prediction model by predicting the RC of three UCMSCs from different donors. The accuracy of the prediction model using four selected Eq. indices is validated by comparing of the predicted RC value and actual RC value. I) Linear correlation of  $RC_{exp}$  versus  $RC_{pred}$  values in three UCMSCs from different donors. Red lines represent linear regression fits for the RC data points.  $R$  values of the fit represent Pearson correlation. Each black dot indicated Cell A, B, and C in panel (E) and Donor A/S/Y in the panel (I). Scale bars: 100  $\mu$ m (A–F).  $RC_{pred}$  represents predicted RC value,  $RC_{exp}$  represents actual RC value. For all violin plots presented, each dot represented one replicate ( $n \geq 3$ ).

variation of RC caused by the inconsistent drug-response of different types of MSCs can also be accurately predicted by this appearance data-driven model.

ERK1/2 was the only molecular signal Eq. index. To investigate the intrinsic associations among the ERK pathway and the other three selected Eq. indices, we tested the change of nucleus roundness, NU/CY ratio, and SSC of Cell A, B, and C by inhibiting ERK pathway with increasing concentrations of PD98059 ( $10 \times 10^{-6}$ ,  $20 \times 10^{-6}$ , and  $25 \times 10^{-6}$  M).<sup>[10]</sup> Overall, the nucleus roundness of all three cells decreased with increasing PD98059

concentrations, except for Cell A treated with  $25 \times 10^{-6}$  M inhibitor (Figure 4E). NU/CY ratio and SSC analysis showed that Cell A and Cell B presented a trend to decrease and Cell C mainly presented an increasing trend (Figure 4E). In addition, western blot analysis showed that ERK1/2 protein expression decreased in a concentration-dependent manner (Figure 4E). Collectively, based on the results of the Eq. indices, we predicted that  $RC_{prep}$  values of Cell A and B became smaller as the concentrations increased, but the  $RC_{prep}$  values became bigger in Cell C (Figure 4G). To obtain  $RC_{exp}$  values, Cell A, B, and



**Figure 4.** Predicting the RC of MSCs treated by different concentrations and types of chemical inhibitors. A,E) Experimental screening of the four selected Eq. indices in three different MSCs treated by increased dosages of XAV939 or PD98059. Three different MSCs named Cell A, B, and C were treated by Wnt/ $\beta$ -catenin pathway inhibitor XAV939 at dosages of  $0 \times 10^{-6}$ ,  $1 \times 10^{-6}$ ,  $5 \times 10^{-6}$ , and  $10 \times 10^{-6}$  M (A–D) or MEK/ERK pathway inhibitor PD98059 at dosages of  $0 \times 10^{-6}$ ,  $10 \times 10^{-6}$ ,  $20 \times 10^{-6}$ , and  $25 \times 10^{-6}$  M (E–H). Detailed information about these cells was provided in Table S1 of the Supporting Information. Scale bars: 100  $\mu$ m. B,F) Actual RC of three different MSCs treated by increased dosages of XAV939 or PD98059 were detected by alizarin red staining and western blot. The RC values of Cell C treated by  $10 \times 10^{-6}$  M XAV939 (B) or MSCs treated by  $25 \times 10^{-6}$  M of PD98059 (F) were excluded because of cell death under osteoinductive conditions. C,G) Validation of the prediction model by predicting RC of three different MSCs treated by increased dosages of XAV939 or PD98059. The accuracy of the prediction model using four selected Eq. indices is validated by comparing the predicted RC value and actual RC value. D,H) Linear correlation of  $RC_{exp}$  versus  $RC_{pred}$  values in three different MSCs treated with increased dosages of XAV939 or PD98059. Red lines represent linear regression fits for the RC data points.  $R$  values of the fit represent Pearson correlation. Each black dot indicated a group of MSCs treated by specific concentration of XAV939 or PD98059.  $RC_{pred}$  represents predicted RC value,  $RC_{exp}$  represents actual RC value. Data are presented as mean  $\pm$  SD in (A) and (E).

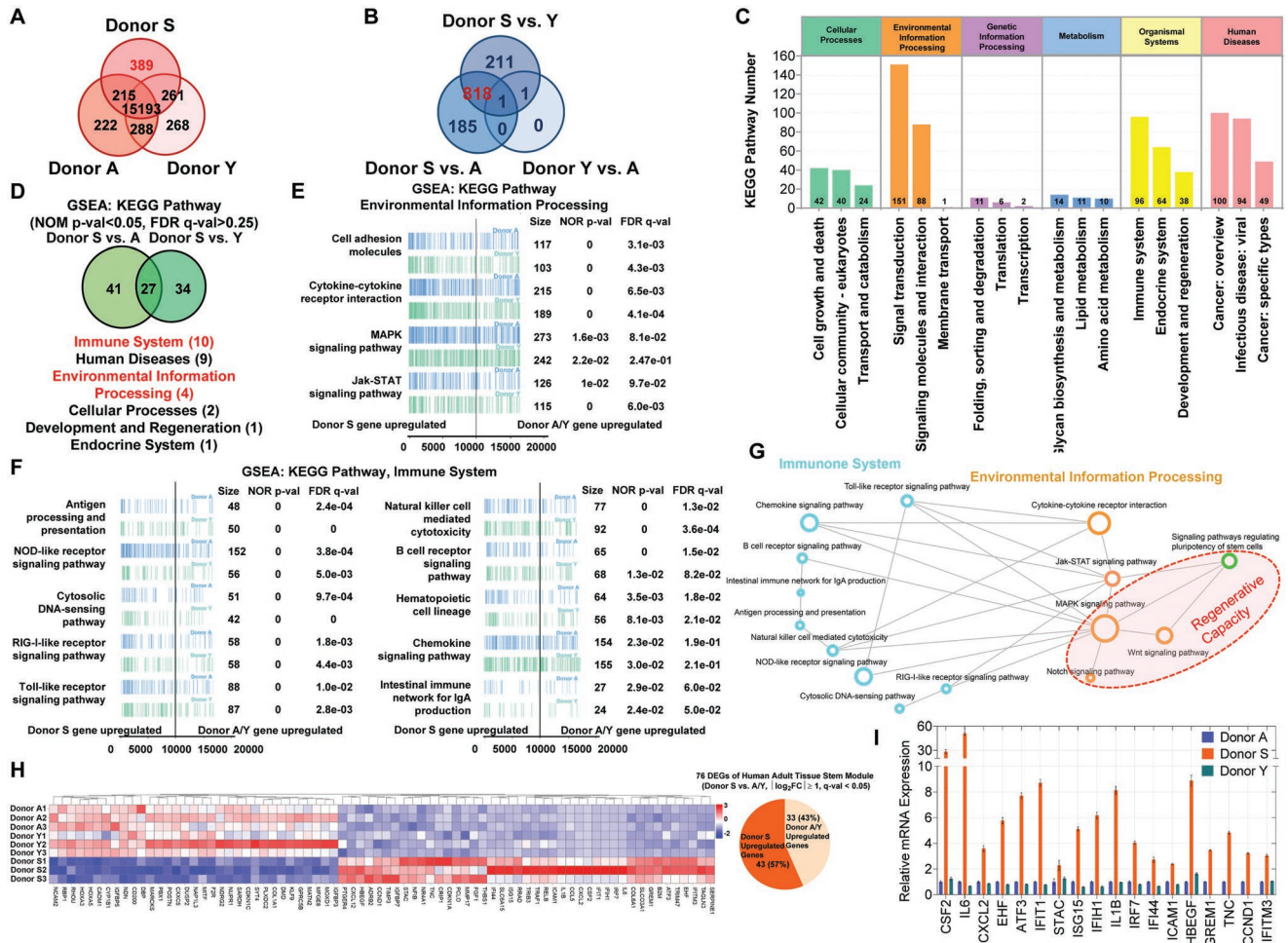
C were cultured for 12 days under the osteogenic inductive conditions, which were treated with PD98059 every two days and the  $RC_{exp}$  values were calculated by the alizarin red assay and western blot (Figure 4F). Although the majority of cells in high concentration of  $25 \times 10^{-6}$  M PD98059 treated groups died after 2 days for osteogenic induction. As expected, we found that the RC changing trends were consistent between  $RC_{exp}$  and  $RC_{pred}$  values in other groups of cells (Figure 4G). The Pearson's correlation coefficients of Cell A/B/C were 0.999, 0.994, and 0.962, respectively (Figure 4H).

Collectively, multidirection experimental validation showed that all Pearson  $R$  values between  $RC_{pred}$  and  $RC_{exp}$  were greater than 0.9, even reaching 0.999. It confirmed that our model

accurately captured the changing tendency of MSC life running state and predicted RC when applied to the complexity of MSC vital statuses.

## 2.5. Cell Appearance-Based Indices Combination May Serve as Major Indicators to Visualize Transcriptional States of MSCs

Among the four selected Eq. indices, including nucleus roundness, nucleus/cytoplasm ratio, SSC-H, and ERK1/2, none of them can solely serve as a sufficiently predictive index for MSC-RC. However, the SSC-H values showed consistent trends with  $RC_{exp}$  values in all the validation experiments



**Figure 5.** RNA-sequencing of three UCMSCs from different donors. A) Venn diagram of unique and shared gene numbers among preselected groups of MSCs. Three preselected UCMSCs from different donors named Donor A, S, and Y were enrolled for RNA-sequencing ( $n = 3$ ). The selected Eq. indices reflecting cell appearance of Donor S were distinguishable higher than Donor A and Y (related to Figure 3F). B) Venn diagram of differentially expressed genes (DEGs) among preselected groups of MSCs. The 818 DEGs of Donor S are common to both Donor A and Y cell populations.  $Q$ -value < 0.05 and  $|\log_2FC| \geq 1$  were set as the cut-off criteria. C) Functional classification of DEGs based on KEGG classification. D) Gene set enrichment analysis (GSEA) based on the KEGG database for gene expression comparisons of Donor S with Donor A or Y, respectively. Significantly enriched pathways with NOM  $p$ -value < 0.05 and FDR  $q$ -value > 0.25 were selected. Venn diagram showing the overlap of 27 KEGG pathways and their corresponding KEGG classification. E) GSEA showing four common KEGG pathways signature to environmental information processing. F) GSEA showing ten common KEGG pathways signature to immune system. G) The network-based on KEGG pathways mainly signature to environmental information processing and immune system showing latent relationships between the osteogenic differentiation-related pathways and significantly enriched pathways. H) Heatmap showing the DEGs of stemness-associated genes related to human adult tissue stem module among Donor A, Y, and S.  $Q$ -value < 0.05 and  $|\log_2FC| \geq 1$  were set as the cut-off criteria. The gene expression data was normalized by z-score. The pie chart illustrated that 57% of DEGs related to stemness-associated genes were Donor S upregulated. I) Dominant expression of stemness-associated genes in Donor S was confirmed by quantitative real-time polymerase chain reaction analysis (qPCR). Among the upregulated genes of Donor S in the heatmap (H), the top 10 genes and the other 8 genes with mean transcript per million (TPM) of Donor S larger than 30 are selected and subjected to qPCR analysis. Data are presented as mean  $\pm$  SD.

(Figures 3A,D,F,H and 4A,C,E,G). Moreover, the MSC population valued higher SSC-H was most significantly correlated with higher osteogenic differentiation capability ( $R = 0.89952$ ) (Figure 2D). These results inspired us to mine the underlined genetic information relevant to SSC-H. Thus, we compared the transcriptional profiles of the Donor A, Y, and S MSCs, which showed an increasing trend of SSC-H (Figure 3F). RNA-sequencing analysis showed that the distinct genes between Donor A, S, and Y were 222, 389, and 268, respectively. Donor S contained more than 100 distinct genes than the other two cell populations, implying Donor S with more abundant

gene expression (Figure 5A). In addition, a total of 818 genes were differentially expressed genes (DEGs) in Donor S compared to the other two cell populations.  $Q$ -value < 0.05 and  $|\log_2FC| \geq 1$  were set as the cut-off criteria (Figure 5B). According to the results above, the selected Eq. indices reflecting the cell appearance of Donor S were distinguishable from the other two cell populations, and its  $RC_{pred}$  values were also higher than Cell A/Y, which matched with the  $RC_{exp}$  (Figure 3F–H). Whatever the cell appearance and the intrinsic gene level, Donor S cells exhibited their distinctiveness. Then, the Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis of

818 DEGs was performed. Functional classification based on KEGG classification revealed that these DEGs were found to be mostly involved in signal transduction, signaling molecules and interaction, immune system, cancers, and infectious diseases (Figure 5C). These results could give us some hints that distinctive appearance profiles of MSCs from Donor S may be due to differences in the microenvironment information processing and immune system-associated pathways, which could be regarded as a special kind of environmental information capture.

Gene set enrichment analysis (GSEA) based on the KEGG database for gene expression comparisons of Donor S with Donor A or Y, respectively, was conducted and significantly enriched pathways with the nominal (NOM)  $p$ -value  $< 0.05$  and false discovery ratio (FDR)  $q$ -value  $> 0.25$  were selected. Venn diagram showed the overlap of 27 KEGG pathways and their corresponding KEGG classification (Figure 5D). Among the 27 commonly significant-enriched pathways, four environment information processing-related pathways, including mitogen-activated protein kinase (MAPK) signaling, Jak-STAT signaling, cell adhesion molecules, and cytokine–cytokine receptor interaction, and ten immune-related pathways were upregulated in Donor S (Figure 5E,F). To investigate the latent relationship between the Donor S upregulated pathways and osteogenic differentiation, a correlation network was constructed based on the significantly enriched KEGG pathway (Figure 5G). It is well-known that Wnt signaling pathway, MAPK signaling pathway, and Notch signaling pathway could govern MSC osteogenic differentiation and play an essential role in regulating stem cell self-renewal/proliferation.<sup>[11]</sup> KEGG correlation network determined that MAPK signaling pathway was a common node connecting other pathways, followed by Jak-STAT pathway and the immune-related pathways could affect MSC osteogenic differentiation by activating the signaling pathways of environment information processing (Figure 5G).

Then, we focused on 553 stemness-related genes that were previously shown to be consistently upregulated in human adult tissue stem cells,<sup>[12]</sup> and asked whether they showed the different expression profiles of Donor S and A/Y. The heatmap showed 76 DEGs with the cut-off criteria of  $q$ -value  $< 0.05$  and  $|\log_2FC| \geq 1$  and 57% of genes were significantly upregulated in Donor S, supporting the stemness of Donor S superior to Donor Y/A (Figure 5H; Table S3, Supporting Information). Among the upregulated genes of Donor S in the heatmap, the top 10 genes and the other 8 genes with mean transcript per million larger than 30 in Donor S were selected and subjected to quantitative real-time polymerase chain reaction (qPCR) analysis. The result confirmed that Donor S showed dominant upregulated mRNA expressions of these stemness-associated genes (Figure 5I). In summary, these data indicate that cell appearance-based indices may serve as major indicators to visualize the transcriptional states of MSCs and reflect MSC stemness.

### 3. Discussion

The capacity to produce therapeutically relevant, high-quality, and -quantities of MSCs by in vitro expanded culture is a common prerequisite for stem cell-based therapies.<sup>[13]</sup>

Outcomes of clinical trials are influenced by MSC heterogeneity, which is a multilevel issue of interdonor variations, inter-tissue differences, and various cell sub-populations.<sup>[14]</sup> Among the factors (donor variance, tissues from different sources, and different cell passage numbers) that led to MSC heterogeneity, our results cannot support us enough to explain the only key factor to mainly affect the regenerative capacity and stemness of MSCs. Nevertheless, our findings suggested that within the same tissue resource of MSCs, donor factors may be a more important determinant of the quality of stem cells than cell passaging numbers. In addition, the RC values decreased with the passage increasing for MSCs derived from the same donor and tissue sources. More importantly, our appearance data-driven model could predict RC ranking of different types of MSCs in each batch without considering the above factors. In addition, different protocols used by researchers for cryopreservation, expansion, and administration induce another level of MSC heterogeneity.<sup>[13]</sup> In this study, we enrolled different types of MSCs commonly used in clinical therapy to provide a reference for clinical application. We noticed that the selected indexes combinations, including nucleus roundness, nucleus/cytoplasm ratio, SSC-H, and ERK1/2 are common cellular parameters not only for MSCs. Further in-depth studies and inclusion of more stem cell data may extend this model to the evaluation of other stem cells, such as neural stem cells or induced pluripotent stem cells.

There are multiple approaches to determine the most important variables for predicting the RC of MSCs. The LASSO regression and principal component regression are useful in standard machine-learning and biomedical informatics. Usually, LASSO is applied to high-dimensional problems with a large number of parameters, for example, in the case of taking 20 out of 100.<sup>[15]</sup> For principal component regression analysis, the sample size is required to be much larger than the index size.<sup>[16]</sup> In this study, when  $x_n \geq 6$ , the offset values in LASSO regression began to increase, suggesting that the model with  $x_n \geq 6$  easily falls into overfitting (Figure S4A, Supporting Information). From  $x_n = 1$  to 5, the offset values of each combination of predictive indices in LASSO regression were larger than those in multiple linear regression. Thus, the multiple linear regression model was more practical due to our small sample size (Figure S4B, Supporting Information). In addition, this model processed by exhaustive testing has considered all of the possible combinations from  $x_n = 1$  to 5. The predictive coefficients only slightly increased (from 0.979 to 0.986) as the number of input  $x$  added from four to five. Reducing one explanatory variable could be beneficial for cost-saving and translational applications. Thus, the indices combination of  $x_n = 4$  could be optimal.

Among the four selected Eq. indices, nucleus roundness, NU/CY ratio, and SSC-H are indicators of cell appearance. Several studies have previously demonstrated that cellular morphology is associated with multipotential differentiation of MSCs and immunosuppression.<sup>[5,6d,17]</sup> We explored the combined explanatory variables from thirty diverse biological properties to predict the explained variable (RC). An equation was established as the mathematical model to evaluate the RC value of MSCs by simply detecting the selected optimal indices. Interestingly, three of four selected optimal indices were cell

appearance-associated indices. Among these, our findings suggested that SSC-H possesses the potential to serve as a single index to predict the RC of MSCs effectively. In addition, cell appearance-associated indices appeared with the highest frequency in relatively high-performing index combinations (roughly 50%) than cell molecular signal or surface marker indices (Figure S3, Supporting Information). These results suggested that cell appearance-associated indices seem superior to other indices for predicting the RC of MSCs. The construction of the predictive model in this study was performed by enrolling several different types of MSCs, which might provide a more general reference compared to prediction approaches based on a single type of MSC in previous studies.

Not just MSCs, the importance of morphology can best be generalized by the adage “form follows function.” The unique cell morphological phenotypes are always associated with different cell types. Cell morphological phenotypes are regulated to some extent by intrinsic and extrinsic factors. Intrinsic factors consist of cell-derived molecular and biophysical signals. Extrinsic factors consist of extracellular signals provided by the cell microenvironment that are defined by a variety of factors, including biological and chemical factors, cell–cell interactions, metabolic and mechanical cues.<sup>[18]</sup> BMMSCs actively changed their shapes by culturing the cells on micropatterned surfaces, 3D hydrogels, or stimulated with a cytokine, and all the alterations could determine the cell-fate outcome.<sup>[6b,c,19]</sup> An in-depth study of cellular geometry implies that mechanical forces at the plasma membrane control the flow of spatial signals through intracellular networks, including actin polymerization, focal adhesion formation, and actomyosin contractility.<sup>[19a,20]</sup> In addition, cell morphological cues could be translated into long-range physical forces that affect the nuclear morphology, mRNA concentration, and YAP/TAZ localization of cells, consistent with our results that two of four important fitting indices were related to nuclear morphology.<sup>[19a,20b,21]</sup>

Considering our four selected Eq. indices, none of them can solely be sufficiently predictive of MSC-RC. However, we found that MSC population valued higher SSC-H was most significantly correlated with higher osteogenic differentiation ( $R = 0.89952$ ). Interestingly, the predictive indices nucleus roundness, SSC-H, and ERK1/2 identified by multiple linear regression model were also found by LASSO regression analysis (Figure S4C, Supporting Information). Among which SSC-H is the most frequently appeared index. These results suggested that SSC-H might play a critical role in predicting the RC of MSCs. SSC-H is a light scattering method to measure the cell refractive index that depends on the cell granularity or internal complexity associated with cytoplasmic membrane wrinkling, the number, and shape of vesicles and mitochondria, development of the endoplasmic reticulum, and the nucleus structure. A previous study has reported that SSC signal measured by flow cytometry could reflect the heterogeneity of undifferentiated pluripotent stem cells, as suggested by the mitochondrial content, and predict clonogenic self-renewal capacity.<sup>[22]</sup> Previous studies indicated that different stem cell states showed distinct mitochondrial and metabolic profiles, indicating that mitochondrion are essential regulators of stem cell activation and fate decisions.<sup>[23]</sup> Intriguingly, we found that the SSC-H values showed consistent trends with  $RC_{exp}$  values

in all the validation experiments. SSC-H indicates a sustainably ordered state of cells and could be used to predict the potential against disorder. MSCs valued with high SSC-H may indicate more abundant protein with conformationally correct and stability and SSC-H monitoring might provide a simple but robust method to estimate the RC of MSCs.

As one of four predictive parameters, ERK1/2 is not directly linked to cell appearance. ERK1/2 is a serine/threonine protein kinase belonging to the MAPK family. It is well-known that ERK pathway is essential for osteogenic differentiation of MSCs, and it could be activated via mechanotransduction.<sup>[24]</sup> A previous study has revealed that ERK is an important endocytic signaling activated by plasma membrane tension gate to regulate the pluripotent cell fate.<sup>[25]</sup> In consideration of our RNA-sequencing data, the KEGG correlation network revealed that MAPK-ERK signaling pathway was a key node connecting other signaling pathways. According to previous studies, MAPK-ERK signaling is essential for lineage differentiation of osteogenic cells; however, the detailed role of ERK signaling in osteogenic cell differentiation is still controversial.<sup>[26]</sup> This linkage of the immune-related pathways could activate the key pathways governing osteogenic differentiation and cell stemness. Meanwhile, the disordered process could be reflected by cell appearance.<sup>[27]</sup> In addition, Donor S with a greater RC value might be attributed to the more enriched signal pathways for integrating environmental cues into changes in cell signaling according to GSEA.

This study established a quantitative relationship between cellular morphological cues and biologically RC. The innovation of our study is not limited to the individual cell or even the sole tissue-derived BMMSCs but considers the cell heterogeneity to establish an accurate and interpretable model to predict RC. Unexpectedly, the validation experiments showed that  $RC_{exp}$  values were greater than predicted values but consistent with the ranking. The  $RC_{exp}$ -to- $RC_{pred}$  value deviations may be due to the different experimental conditions between different batches of experiments, especially prolonged western blot exposure time may lead to the high value of  $RC_{exp}$  during validating. In addition, different cell states induced by single or continuous administration of chemical inhibitors may make it difficult to predict the RC values accurately. A more accurate and simplified model by including more stem cell data, reducing predictive index number, or optimizing the experimental methods for testing indices needs to be further explored to facilitate clinical application. In this study, at least the consistency between the actual and predicted dynamical trends of RC values demonstrated the sensitivity and applicability of our selected Eq.

## 4. Conclusion

In our study, we established a data-based multiple linear regression model that excavates four indices, including nucleus roundness, NU/CY ratio, SSC-H, and ERK1/2 to predict the RC of MSCs in vitro. The predictable result could advance the application of MSCs in basic and translational studies, and the model may provide a promising evaluation approach to characterize the other potencies of MSCs.

## 5. Experimental Section

**Ethical Approval:** Ethics approval for collection and use of human MSCs from umbilical cord, gingiva, skin, and dental pulp was approved by the Medical Ethics Committee of Hospital of Stomatology, Sun Yat-sen University (project ID: KQEC-2021-48-01). Informed consent was obtained from all donors and the material collection was without extra risk for them.

**Umbilical Cord (UC)-Derived MSC Isolation and Culture:** Four human umbilical cord tissues from pregnant women after baby delivery undergoing surgical resection at the Sixth Affiliated Hospital, Sun Yat-sen University, and Guangdong Provincial People Hospital. The UCs were washed three times with ethanol 70% and then three times with PBS to sterilize and remove blood. The UC was cut into small pieces, the blood vessels were removed and transferred into a 50 mL tube following by incubation with collagenase type I (3 mg mL<sup>-1</sup>, Worthington Biochemical Corporation, Lakewood, America) and dispase (4 mg mL<sup>-1</sup>, Roche Diagnostics GmbH, Mannheim, Germany) for 1 h at 37 °C with shaking. After digestion, the solution was diluted in PBS and centrifuge at 1500 × rpm/5 min at 4 °C to collect cell pellets. The cell pellet was resuspended in the  $\alpha$ -MEM culture medium (Gibco, Thermo Fisher Scientific, Waltham, America) containing 15% fetal bovine serum (Gibco, Thermo Fisher Scientific, Waltham, America), L-glutamine (2 × 10<sup>-3</sup> M, Gibco, Thermo Fisher Scientific, Waltham, America), penicillin and streptomycin (100 U mL<sup>-1</sup>, Invitrogen, Waltham, America), and L-ascorbic acid phosphate (10 × 10<sup>-3</sup> M, Wako, Tokyo, Japan) and transferred into cell culture dishes and incubated at 37 °C and 5% CO<sub>2</sub> for cell expansion. Culture medium was replaced by every 3 days. When the cells reached 90% confluence (after 10–14 days) to establish the passage (P) 0 MSCs, cells were digested with triple (Gibco, Thermo Fisher Scientific, Waltham, America) for the next passage. In the experiments, UCMSCs were used at P3, P7, and P12 *in vitro*.

**Dental Pulp-Derived MSC Isolation and Culture:** Stem cells from human exfoliated deciduous teeth (SHED) were isolated from normal exfoliated deciduous teeth collected from four children (aged from 5- to 10-year-old) and dental pulp stem cells (DPSCs) were obtained from teeth extracted of one youth donor for orthodontic reason at Hospital of Stomatology, Sun Yat-sen University. The method of cell culture was as reported previously.<sup>[28]</sup> Briefly, dental pulp was separated from root canals and then digested in a mixed solution of 3 mg mL<sup>-1</sup> collagenase type I and 4 mg mL<sup>-1</sup> dispase for 1 h at 37 °C with shaking. Single-cell suspensions were cultured at cell culture dishes with the  $\alpha$ -MEM culture medium mentioned above. Cells were expanded to 90% confluence and digested with triple for passaging. P11, P12, P13 of SHED, and P12 of DPSCs were prepared for *in vitro* experiments.

**Isolation and Culture of MSCs Derived from Human Gingiva, Skin, Bone Marrow, and Adipose Tissues:** Briefly, gingival and skin tissues were obtained from discarded tissues at the Hospital of Stomatology, Sun Yat-sen University. Gingival and skin tissues were treated aseptically into pieces and incubated with 3 mg mL<sup>-1</sup> collagenase type I and 4 mg mL<sup>-1</sup> dispase for 1 h at 37 °C with shaking. Then the dissociated cell suspension was diluted in PBS and centrifuged to get cell deposition. The cell deposition was suspended and plated on a 10 cm cell dish with the  $\alpha$ -MEM culture medium mentioned above and cultured at 37 °C in an incubator with 5% CO<sub>2</sub>. After cultured for 72 h, the nonadherent cells were removed. The plastic-adherent cells were treated with triple and passaged when they reached about 90% confluent density and subcultured continuously. Gingiva-derived MSCs at P6 and P11 were used in the *in vitro* experiments. Skin-derived MSCs at P8 were prepared for subsequent experiments. Additionally, two bone marrow-derived mesenchymal stem cells (BMSCs) and primary human adipose-derived stem cells (ASCs) were obtained from ScienCell (San Diego, CA, USA). BMSC at P15 or P16 and ASCs at P10 were maintained for the study under the recommended conditions and procedure.

**Antibodies and Reagents:** The antibody against EZH2 (#07-689), active- $\beta$ -Catenin (#05-665), and  $\beta$ -Catenin (#06-734) were purchased from Millipore (Merck Millipore, Billerica, MA, USA). The antibodies against Tet1 (#ab191698), Oct4 (#ab19857), SOX2 (#ab93689), and

Alkaline Phosphatase (ALP, #ab108337) were purchased from Abcam (Cambridge, MA, USA). The antibodies against Nanog (#4903T), TAZ (#83669S), YAP (#14074S), Phospho-Smad3 (#9520S), Smad3 (#9523S), Cleaved Notch1 (#4147S), Notch1 (#3608S), Phospho-p44/42 MAPK (Erk1/2) (#4370S), p44/42 MAPK (Erk1/2) (#4695S), Runx2 (#8486), Phospho-mTOR (#2971S), and mTOR (#2983S), were purchased from Cell Signaling Technology (CST, Danvers, MA, USA). The GAPDH were purchased from Sigma (St. Louis, MO, USA). Anti-Caveolin-1 antibodies (#sc-53564) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA). Anti-human-CD34, CD45, CD73, CD90, CD105, and CD146 antibodies were purchased from BD Biosciences (San Jose, CA, USA).

**High-Content Image Analysis:** The cells were seeded in 96-well black plates (CellCarrier-96, PERKInElmer) in cell culture media. After 24 h, the cells were washed with PBS and incubated with CellMask Deep Red Plasma Membrane Stain (dilution 1:2000, ThermoFisher Scientific) for 15 min at 37 °C. Then the cells were washed and incubated with Hoechst 33342 (dilution 1:2000, #H3570, Thermo Fisher Scientific) to visualize nuclei for 5 min at room temperature (RT). Finally, the cells were rinsed with PBS, cover-slipped in PBS and examined. The entire procedure should keep in dark. The plates with MSCs were subjected to high-content analysis for image acquisition and data processing (objective ×20 magnification, PERKInElmer, USA). Morphological indicators including cell area, cell roundness, nucleus area, nucleus roundness, cytoplasm area, cytoplasm roundness, and ratio of nucleus/cytoplasm were analyzed using Harmony 4.0 software by employing a custom-made image analysis building block. Each morphological parameter was quantified using an average of at least 5000 selected cells for each MSC.

**Flow Cytometry:** MSC-specific surface markers were evaluated and characterized by flow cytometry. Cells were rinsed, digested by triple, washed, and resuspended in PBS containing 0.5% bovine serum albumin (BSA) before incubation with the fluorochrome-conjugated antibodies against human antigens, including anti-CD34-PE, anti-CD45-PE, anti-CD73-PE, anti-CD90-PE, anti-CD105-PE, and anti-CD146-PE for 30 min in dark at 4 °C. After incubation, cells were rinsed with PBS containing 0.5% BSA, centrifuged at 1500 rpm for 5 min and resuspended for flow cytometry analysis. Samples were characterized using NovoCyte Flow Cytometer and analyzed using NovoExpress software (NovoCyte, ACEA Biosciences, USA).

**EdU Assay:** EdU assay was conducted using a kFluor488-EDU Kit (KeyGen Biotech, Jiangsu, China). 24-well plates were seeded with cells and cultured for 24 h at 37 °C. Cells were stained by 50 × 10<sup>-6</sup> M EdU in culture media for 2 h. Then, the cells were rinsed with PBS and fixed with 4% paraformaldehyde for 15 min at room temperature. EdU staining was performed according to the kit instructions. After the cells were counterstained with Hoechst 33342 (1:2000, 5 mg mL<sup>-1</sup>, H3570, ThermoFisher Scientific) for 15 min to stain the nuclei and washed with PBS. A high-content screening system (objective ×10 magnification, PERKInElmer, USA) was used to get the images of EdU positive cells.

**Western Blot:** Samples were lysed with a protein extraction kit (#78501, Thermo Fisher) supplemented with protease and phosphatase inhibitors. After quantification by the BCA kit, equal amounts of protein for each sample were loaded onto SDS-PAGE gels, then transferred to polyvinylidene difluoride membranes (Millipore). Membranes were blocked with 5% BSA in TBST for 2 h, then incubated with the following primary antibodies at 4 °C overnight with shaking: EZH2, active- $\beta$ -Catenin,  $\beta$ -Catenin, Tet1, Oct4, SOX2, Alkaline Phosphatase, Nanog, TAZ, YAP, Phospho-Smad3, Smad3, Cleaved Notch1, Notch1, Phospho-p44/42 MAPK (Erk1/2), p44/42 MAPK (Erk1/2), Runx2, Phospho-mTOR, mTOR, Caveolin-1 or GAPDH primary antibodies. Then the membranes were washed with TBST and incubated with the appropriate secondary antibody at room temperature for 1 h. After washing completely, protein bands were visualized using Supersignal West Pico Chemiluminescent Substrate (Thermo Fisher) and evaluated with a gel imaging system (Bio-Rad, USA). The relative density was calculated by NIH ImageJ software (Media Cybernetics, USA).

**Osteogenic Differentiation:** Cells were seeded in 12-well plates. At 90% confluence, the culture media were replaced with osteogenic

induction media containing  $1.8 \times 10^{-3}$  M monopotassium phosphate,  $10 \times 10^{-9}$  M dexamethasone,  $100 \text{ U mL}^{-1}$  penicillin/streptomycin,  $0.1 \times 10^{-3}$  M L-ascorbic acid phosphate,  $2 \times 10^{-3}$  M glutamine, 15% fetal bovine serum, and  $\alpha$ -MEM culture media. Osteogenic induction media were replaced every 2 days. After 5 to 12 days of induction, the cultured cells were stained with alizarin red, alkaline phosphatase, or lysed for protein isolation to assess osteogenic gene expression.

**Alizarin Red Staining:** After 4 weeks of osteogenic induction, the osteogenic differentiation was assessed by the ability of the cells to form mineralized nodules stained by alizarin red. Briefly, the cells were washed with PBS, fixed with 60% isopropanol for 30 min at RT, rehydrated with ddH<sub>2</sub>O for 5 min. Subsequently, cells were incubated with alizarin red staining solution (dissolved 1 g alizarin red S (#A5533, Sigma) in 100 mL ddH<sub>2</sub>O and filtrated) for 10 min at RT. Finally, the cells were washed with PBS to remove nonspecific binding, dried, and observed under a ZEISS Axio Observer microscope (Carl Zeiss Microscopy, White Plains, NY, USA).

**Calculation of RC Values:** The positively mineralized areas of Alizarin Red S staining were quantified as percentages of the total area using ImageJ software (ImageJ Software Inc., MD, USA). The relative density of osteogenic lineage protein was also calculated by ImageJ software. The RC values were calculated by the average value of positively mineralized area percentages and osteogenic lineage protein levels.

**Chemical Inhibitors Treatment:** Cells were seeded on a culture plate with 70–80% confluent and treated with XAV939 (13596, Cayman Chemical) (three concentration gradient ranges:  $1 \times 10^{-6}$ ,  $5 \times 10^{-6}$ ,  $10 \times 10^{-6}$  M), and PD98059 (#S1805 Beyotime, Shanghai, China) (three concentration gradient ranges:  $10 \times 10^{-6}$ ,  $20 \times 10^{-6}$ ,  $25 \times 10^{-6}$  M), respectively. Chemical efficiency was evaluated by western blotting. Treated cells were harvested and prepared for further experiments.

**RNA Sequencing Analysis:** UCMSCs collected from Donor A, Y, S, with three replicates of each sample were analyzed by mRNA sequencing (RNA-Seq, Huada Biotechnology Co., China). The BGISEQ-500 platform was used for RNA-seq libraries. The clean reads were stored in FASQC format and aligned to the reference genome and reference gene with HISAT and Bowtie2, respectively.<sup>[29]</sup> Then expression level of the gene was calculated by RSEM (v1.2.12).<sup>[30]</sup> Essentially, differential expression analysis was performed using the DESeq2(v1.4.5) with absolute log<sub>2</sub>-based fold change ( $|\log_2\text{FC}| \geq 1$  and  $Q$ -value  $< 0.05$ ).<sup>[31]</sup> To take insight into the change of phenotype, KEGG (<https://www.kegg.jp/>) and GSEA (<http://www.broadinstitute.org/gsea>) were performed.

**Quantitative Real-Time Polymerase Chain Reaction:** Transcription levels of the genes with the top ten DEGs of human adult tissue stem cell module were determined by RT-qPCR. Briefly, MSCs from Donor A, Y, and S groups were harvested and washed with PBS once. Total RNA was extracted using RNA-Quick Purification Kit (ESscience Cat# RN001). According to the manufacturer's protocol, 1  $\mu\text{g}$  of total RNA was transcribed in a 20  $\mu\text{L}$  reaction mixture by using the Primescript II RTase (Takara, Otsu, Japan) for complementary DNA (cDNA) synthesis. Amplification of target genes was performed by qPCR using the cDNA as a template, specific primers, and a TB Green Premix Ex Taq II kit (Takara, Otsu, Japan) on a LightCycler system (LightCycler, Roche Diagnostics). The gene expression was evaluated for three biological replicates and the relative transcription levels were calculated with  $2^{-\Delta\Delta\text{CT}}$  method. The housekeeping gene GAPDH was used to normalize the level of messenger RNA. Sequences of qPCR primers used in this study are shown in Table S4 of the Supporting Information.

**Excavation of Key Indices:** Multiple linear regression analysis method combined with the index enumeration was used to analyze the correlation between the RC values of eleven types of MSC and the thirty MSC-featured biological indices. In order to expand the sample size, eleven types of MSC were divided randomly with a split ratio of 8:3 into eight MSCs in the training set and the remaining three in the testing set. If the number of explanatory variables that were involved in the model development was over half of eight in this case ( $n \geq 5$ ), the model was easy to fall into overfitting. Thus, the enumeration of  $n$  ( $n = 1, 2, 3, 4, 5$ ) indices combination has been considered to determine the key indices for a predictive model. All 174436 combinations of  $n$  ( $n = 1, 2, 3, 4, 5$ )

explanatory variable were calculated. For a given index combination, the fitting tests were obtained from 20 random combinations out of the total 165 possible combinations by repeating the following process

Different parameters can be obtained for different combinations, as shown in Equation (2)

$$\begin{pmatrix} \gamma^1 \\ \dots \\ \gamma^8 \end{pmatrix} = \begin{pmatrix} x_1^1 & \dots & x_n^1 \\ \dots & \dots & \dots \\ x_1^8 & \dots & x_n^8 \end{pmatrix} \begin{pmatrix} b_1 \\ \dots \\ b_n \end{pmatrix} \quad (2)$$

where  $\gamma$  is the explained variable, representing the actual RC values;  $x$  is the explanatory variable, representing selected indices;  $b_1, b_2, \dots, b_n$  are the unknown parameters of the corresponding explanatory variables;  $n$  represents the number of selected indices. The multiple linear regression was performed to obtain  $b_1, b_2, \dots, b_n$ . Second, the difference between the experimental value and the predicted one was calculated. The predicted values can be obtained by Equation (3)

$$\begin{pmatrix} \gamma\gamma^9 \\ \gamma\gamma^{10} \\ \gamma\gamma^{11} \end{pmatrix} = \begin{pmatrix} x_1^9 & \dots & x_n^9 \\ x_1^{10} & \dots & x_n^{10} \\ x_1^{11} & \dots & x_n^{11} \end{pmatrix} \begin{pmatrix} b_1 \\ \dots \\ b_n \end{pmatrix} \quad (3)$$

$\gamma\gamma^9, \gamma\gamma^{10}$ , and  $\gamma\gamma^{11}$  were the predicted RC values calculated by Equation (3). Finally, the above process was repeated and the index combination with the most minor deviation was selected. The optimal multiple linear regression model as shown in Equation (4) can be constructed.

$$\gamma = b_1x_1 + \dots + b_nx_n \quad (4)$$

It is important to mention that the constant coefficients of  $b_1, b_2, \dots, b_n$  calculated were based on the data obtained by the instruments.

**Statistical Analysis:** The predictive model was developed by Octave 6.4.0 software (<http://www.gnu.org/software/octave/index.html>). Data were presented as mean  $\pm$  SD in Figures 4A,E and 5I. For measuring the association between variables of interest, giving information about the magnitude of the association, or correlation, as well as the direction of the relationship, the Pearson's correlation coefficient (also called Pearson's R) was analyzed with GraphPad Prism 8.3 software.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

This work was supported by grants from the National Key R&D Program of China (2021YFA1100600 to S.S.), the Pearl River Talent Recruitment Program (2019ZT08Y485, 2019QN01Y138, and 2019JC01Y182), the Guangdong Financial Fund for High-Caliber Hospital Construction (174-2018-XMZC-0001-03-0125, D-07 to S.S., C-03 and D-11 to X.K.), and the National Natural Science Foundation of China (82170924 to X.K., 82170987/ H1507 to Y.C.).

## Conflict of Interest

The authors declare no conflict of interest.

## Author Contributions

D.W. contributed to designing study plan, performing experimental procedures, and drafting and final approval of the manuscript. L.Z., B.S., L.T., L.L., G.L., and X.M. contributed to cell experiments, data acquisition, analysis, and manuscript writing. X.Y. contributed to building the multiple linear model. L.Z. and X.K. contributed to analyzing the RNA-sequencing data. S.S., C.Y., X.Y., and X.K. contributed to the project conception, experimental design, writing manuscript, and supervision. All authors approved the final version of the manuscript.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Keywords

cell appearance, mathematical models, mesenchymal stem cells, predictive models, regenerative capacities

Received: January 20, 2022

Revised: April 14, 2022

Published online: June 8, 2022

- [1] a) M. F. Pittenger, A. M. Mackay, S. C. Beck, R. K. Jaiswal, R. Douglas, J. D. Mosca, M. A. Moorman, D. W. Simonetti, S. Craig, D. R. Marshak, *Science* **1999**, *284*, 143; b) P. Bianco, P. G. Robey, P. J. Simmons, *Cell Stem Cell* **2008**, *2*, 313; c) R. Hass, C. Kasper, S. Bohm, R. Jacobs, *Cell Commun. Signaling* **2011**, *9*, 12.
- [2] a) M. Dominici, K. Le Blanc, I. Mueller, I. Slaper-Cortenbach, F. Marini, D. Krause, B. Deans, A. Keating, D. Prockop, E. Horwitz, *Cytotherapy* **2006**, *8*, 315; b) B. M. Dickens, *Med. Law* **2008**, *27*, 179; c) G. Q. Daley, I. Hyun, J. F. Apperley, R. A. Barker, N. Benvenisty, A. L. Bredenoord, C. K. Breuer, T. Caulfield, M. I. Cedars, J. Frey-Vasconcellos, H. E. Heslop, Y. Jin, R. T. Lee, C. McCabe, M. Munsie, C. E. Murry, S. Piantadosi, M. Rao, H. M. Rooke, D. Sipp, L. Studer, J. Sugarman, M. Takahashi, M. Zimmerman, J. Kimmelman, *Stem Cell Rep.* **2016**, *6*, 787.
- [3] a) U. Galderisi, G. Peluso, G. Di Bernardo, *Stem Cell Rev. Rep.* **2021**, *18*, 23; b) H. Caplan, S. D. Olson, A. Kumar, M. George, K. S. Prabhakara, P. Wenzel, S. Bedi, N. E. Toledano-Furman, F. Triolo, J. Kamhieh-Milz, G. Moll, C. S. Cox, Jr., *Front. Immunol.* **2019**, *10*, 1645; c) T. Zhou, Z. Yuan, J. Weng, D. Pei, X. Du, C. He, P. Lai, *J. Hematol. Oncol.* **2021**, *14*, 24; d) S. L. Waters, L. J. Schumacher, A. J. El Haj, *npj Regen. Med.* **2021**, *6*, 24.
- [4] a) H. Yang, L. N. Gao, Y. An, C. H. Hu, F. Jin, J. Zhou, Y. Jin, F. M. Chen, *Biomaterials* **2013**, *34*, 7033; b) N. H. Nicolay, R. L. Perez, J. Debus, P. E. Huber, *Cancer Lett.* **2015**, *366*, 133; c) L. A. Costa, N. Eiro, M. Fraile, L. O. Gonzalez, J. Saa, P. Garcia-Portabella, B. Vega, J. Schneider, F. J. Vizoso, *Cell. Mol. Life Sci.* **2021**, *78*, 447; d) S. Shi, S. Gronthos, *J. Bone Miner. Res.* **2003**, *18*, 696; e) T. R. Heathman, A. W. Nienow, M. J. McCall, K. Coopman, B. Kara, C. J. Hewitt, *Regener. Med.* **2015**, *10*, 49.
- [5] a) R. A. Marklein, M. W. Klinker, K. A. Drake, H. G. Polikowsky, E. C. Lessey-Morillon, S. R. Bauer, *Cytotherapy* **2019**, *21*, 17; b) M. W. Klinker, R. A. Marklein, J. L. Lo Surdo, C. H. Wei, S. R. Bauer, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E2598.
- [6] a) J. M. Kowal, H. Schmal, U. Halekoh, J. B. Hjelmberg, M. Kassem, *Stem Cells Transl. Med.* **2020**, *9*, 189; b) T. C. von Erlach, S. Bertazzo, M. A. Wozniak, C. M. Horejs, S. A. Maynard, S. Attwood, B. K. Robinson, H. Autefage, C. Kallepitis, A. Del Rio Hernandez, C. S. Chen, S. Goldoni, M. M. Stevens, *Nat. Mater.* **2018**, *17*, 237; c) R. A. Marklein, J. L. Lo Surdo, I. H. Bellay, S. A. Godil, R. K. Puri, S. R. Bauer, *Stem Cells* **2016**, *34*, 935; d) K. A. Kilian, B. Bugarija, B. T. Lahn, M. Mrksich, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4872.
- [7] a) H. Shi, X. Xu, B. Zhang, J. H. Xu, Z. J. Pan, A. H. Gong, X. Zhang, R. Li, Y. X. Sun, Y. M. Yan, F. Mao, H. Qian, W. R. Xu, *Theranostics* **2017**, *7*, 1674; b) C. Hartmann, *Trends Cell Biol.* **2006**, *16*, 151; c) L. Ling, V. Nurcombe, S. M. Cool, *Gene* **2009**, *433*, 19135507; d) M. P. Alfaro, A. Vincent, S. Saraswati, C. A. Thorne, C. C. Hong, E. Lee, P. P. Young, *J. Biol. Chem.* **2010**, *285*, 35645.
- [8] D. W. Liu, X. X. Kou, C. Chen, S. Y. Liu, Y. Liu, W. J. Yu, T. T. Yu, R. L. Yang, R. C. Wang, Y. H. Zhou, S. T. Shi, *Cell Res.* **2018**, *28*, 918.
- [9] a) N. Almasoud, S. Binhamdan, G. Younis, H. Alaskar, A. Alotaibi, M. Manikandan, M. Alfayez, M. Kassem, N. AlMuraikhi, *Sci. Rep.* **2020**, *10*, 16746; b) N. Liu, S. Shi, M. Deng, L. Tang, G. Zhang, N. Liu, B. Ding, W. Liu, Y. Liu, H. Shi, L. Liu, Y. Jin, *J. Bone Miner. Res.* **2011**, *26*, 2082.
- [10] H. B. Low, Z. L. Wong, B. Y. Wu, L. R. Kong, C. W. Png, Y. L. Cho, C. W. Li, F. C. Xiao, X. Xin, H. Yang, J. M. Loo, F. Y. X. Lee, I. B. H. Tan, R. DasGupta, H. M. Shen, H. Schwarz, N. R. J. Gascoigne, B. C. Goh, X. H. Xu, Y. L. Zhang, *Nat. Commun.* **2021**, *12*, 2284.
- [11] a) A. W. James, *Scientifica* **2013**, *2013*, 684736; b) P. Zhao, L. Xiao, J. Peng, Y. Q. Qian, C. C. Huang, *Eur. Rev. Med. Pharmacol.* **2018**, *22*, 3962; c) J. N. Liu, C. Sato, M. Cerletti, A. Wagers, *Curr. Top. Dev. Biol.* **2010**, *92*, 367.
- [12] a) D. J. Wong, H. Liu, T. W. Ridky, D. Cassarino, E. Segal, H. Y. Chang, *Cell Stem Cell* **2008**, *2*, 333; b) S. Gkoutela, F. Castro-Giner, B. M. Szczerba, M. Vetter, J. Landin, R. Scherrer, I. Krol, M. C. Scheidmann, C. Beisel, C. U. Stirnimann, C. Kurzeder, V. Heinzelmann-Schwarz, C. Rochlitz, W. P. Weber, N. Aceto, *Cell* **2019**, *176*, 98.
- [13] O. Levy, R. Kuai, E. M. J. Siren, D. Bhere, Y. Milton, N. Nissar, M. De Biasio, M. Heinelt, B. Reeve, R. Abdi, M. Alturki, M. Fallatah, A. Almalik, A. H. Alhasan, K. Shah, J. M. Karp, *Sci. Adv.* **2020**, *6*, eaba6884.
- [14] a) W. Wruck, N. Graffmann, L. S. Spitzhorn, J. Adjaye, *Front. Cell Dev. Biol.* **2021**, *9*, 717772; b) K. C. O'Connor, *Stem Cells Int.* **2019**, *2019*, 5924878; c) N. Ketterl, G. Brachtel, C. Schuh, K. Bieback, K. Schallmoser, A. Reinisch, D. Strunk, *Stem Cell Res. Ther.* **2015**, *6*, 236.
- [15] R. Muthukrishnan, R. Rohini, 2016 IEEE Int. Conf. on Advances in Computer Applications (ICACA), IEEE, Piscataway, NJ **2016**, p. 18.
- [16] W. Osborne, B. Costello, *Pract. Assess., Res., Eval.* **2004**, *9*, 11.
- [17] a) S. Imboden, X. Liu, B. S. Lee, M. C. Payne, C. J. Hsieh, N. Y. C. Lin, *Sci. Rep.* **2021**, *11*, 6728; b) J. M. Kowal, S. Moller, D. Ali, F. Figeac, T. Barington, H. Schmal, M. Kassem, *Stem Cell Res. Ther.* **2021**, *12*, 265; c) D. Docheva, D. Padula, C. Popov, W. Mutschler, H. Clausen-Schaumann, M. Schieker, *J. Cell. Mol. Med.* **2008**, *12*, 537; d) W. C. Lee, H. Shi, Z. Y. Poon, L. M. Nyan, T. Kaushik, G. V. Shivashankar, J. K. Y. Chan, C. T. Lim, J. Han, K. J. Van Vliet, *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4409; e) M. D. Treiser, E. H. Yang, S. Gordonov, D. M. Cohen, I. P. Androulakis, J. Kohn, C. S. Chen, P. V. Moghe, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 610.
- [18] a) J. Krstic, M. Herrmann, I. Gadjanski, S. Mojsilovic, *Front. Cell. Dev. Biol.* **2017**, *5*, 82; b) G. Dostert, B. Mesure, P. Menu, E. Velot, *Front. Cell. Dev. Biol.* **2017**, *5*, 6; c) R. A. Marklein, J. Lam, M. Guvendiren, K. E. Sung, S. R. Bauer, *Trends Biotechnol.* **2018**, *36*, 105.
- [19] a) M. Bao, J. Xie, A. Piruska, W. T. S. Huck, *Nat. Commun.* **2017**, *8*, 1962; b) S. Khetan, M. Guvendiren, W. R. Legant, D. M. Cohen, C. S. Chen, J. A. Burdick, *Nat. Mater.* **2013**, *12*, 458; c) A. Prasad, E. Alizadeh, *Trends Biotechnol.* **2019**, *37*, 347.

- [20] a) S. R. Neves, P. Tsokas, A. Sarkar, E. A. Grace, P. Rangamani, S. M. Taubenfeld, C. M. Alberini, J. C. Schaff, R. D. Blitzer, I. I. Moraru, R. Iyengar, *Cell* **2008**, *133*, 666; b) K. E. Scott, S. I. Fraley, P. Rangamani, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2021571118.
- [21] N. Jain, K. V. Iyer, A. Kumar, G. V. Shivashankar, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11349.
- [22] J. M. Ramirez, Q. Bai, M. Pequignot, F. Becker, A. Kassambara, A. Bouin, V. Kalatzis, M. Dijon-Grinand, J. De Vos, *Stem Cells Dev.* **2013**, *22*, 1851.
- [23] a) M. Khacho, A. Clark, D. S. Svoboda, J. Azzi, J. G. MacLaurin, C. Meghaizel, H. Sesaki, D. C. Lagace, M. Germain, M. E. Harper, D. S. Park, R. S. Slack, *Cell Stem Cell* **2016**, *19*, 232; b) A. Bahat, A. Gross, *J. Biol. Chem.* **2019**, *294*, 13852; c) R. P. Chakrabarty, N. S. Chandel, *Cell Stem Cell* **2021**, *28*, 394.
- [24] a) P. Muller, A. Langenbach, A. Kaminski, J. Rychly, *PLoS One* **2013**, *8*, 71283; b) A. De Simone, M. N. Evanitsky, L. Hayden, B. D. Cox, J. Wang, V. A. Tornini, J. Ou, A. Chao, K. D. Poss, S. Di Talia, *Nature* **2021**, *590*, 129.
- [25] H. De Belly, A. Stubb, A. Yanagida, C. Labouesse, P. H. Jones, E. K. Paluch, K. J. Chalut, *Cell Stem Cell* **2021**, *28*, 273.
- [26] a) B. Bai, J. He, Y. S. Li, X. M. Wang, H. J. Ai, F. Z. Cui, *Biomed Res. Int.* **2013**, *2013*, 361906; b) X. Dou, X. Wei, G. Liu, S. Wang, Y. Lv, J. Li, Z. Ma, G. Zheng, Y. Wang, M. Hu, W. Yu, D. Zhao, *J. Orthop. Transl.* **2019**, *19*, 81; c) D. J. Xu, Y. Z. Zhao, J. Wang, J. W. He, Y. G. Weng, J. Y. Luo, *BMB Rep.* **2012**, *45*, 247.
- [27] a) H. Mohammadpour, A. A. Pourfathollah, M. N. Zarif, S. M. Hashemi, *Immunopharmacol. Immunotoxicol.* **2016**, *38*, 68; b) K. Takeshita, S. Motoike, M. Kajiya, N. Komatsu, M. Takewaki, K. Ouhara, T. Iwata, K. Takeda, N. Mizuno, T. Fujita, H. Kurihara, *Stem Cell Res. Ther.* **2017**, *8*, 101; c) M. F. A. Cutiongco, B. S. Jensen, P. M. Reynolds, N. Gadegaard, *Nat. Commun.* **2020**, *11*, 1384.
- [28] S. Gronthos, M. Mankani, J. Brahim, P. G. Robey, S. Shi, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 13625.
- [29] a) D. Kim, B. Landmead, S. L. Salzberg, *Nat. Methods* **2015**, *12*, 357; b) B. Langmead, S. L. Salzberg, *Nat. Methods* **2012**, *9*, 357.
- [30] B. Li, C. N. Dewey, *BMC Bioinformatics* **2011**, *12*, 323.
- [31] M. I. Love, W. Huber, S. Anders, *Genome Biol.* **2014**, *15*, 550.