Contents lists available at ScienceDirect

Regenerative Therapy

journal homepage: http://www.elsevier.com/locate/reth

Original Article

The importance of scoring recognition fitness in spheroid morphological analysis for robust label-free quality evaluation

Kazuhide Shirai [a, b], Hirohito Kato [a], Yuta Imai [a], Mayu Shibuta [a], Kei Kanie [a], Ryuji Kato [a, c, *]

[a] *Graduate School of Pharmaceutical Sciences, Nagoya University, Furocho, Chikusa-ku, Nagoya 464-8601, Japan*
[b] *Mathematical Sciences Research Laboratory, Research & Development Division, Nikon Corporation, Yokohama Plant, 471, Nagaodai-cho, Sakae-ku, Yokohama-city, Kanagawa 244-8533, Japan*
[c] *Institute of Nano-Life-Systems, Institute for Innovation for Future Society, Nagoya University, Furocho, Chikusa-ku, Nagoya 464-8601, Japan*

ARTICLE INFO

Article history:
Received 25 December 2019
Received in revised form
6 February 2020
Accepted 20 February 2020

Keywords:
Spheroid
Object recognition
Spheroid morphology
Label-free quality evaluation
Cell manufacturing

ABSTRACT

Because of the growing demand for human cell spheroids as functional cellular components for both drug development and regenerative therapy, the technology to non-invasively evaluate their quality has emerged. Image-based morphology analysis of spheroids enables high-throughput screening of their quality. However, since spheroids are three-dimensional, their images can have poor contrast in their surface area, and therefore the total spheroid recognition by image processing is greatly dependent on human who design the filter-set to fit for their own definition of spheroid outline. As a result, the reproducibility of morphology measurement is critically affected by the performance of filter-set, and its fluctuation can disrupt the subsequent morphology-based analysis. Although the unexpected failure derived from the inconsistency of image processing result is a critical issue for analyzing large image data for quality screening, it has been tackled rarely. To achieve robust analysis performances using morphological features, we investigated the influence of filter-set's reproducibility for various types of spheroid data. We propose a new scoring index, the “recognition fitness deviation (RFD),” as a measure to quantitatively and comprehensively evaluate how reproductively a designed filter-set can work with data variations, such as the variations in replicate samples, in time-course samples, and in different types of cells (a total of six normal or cancer cell types). Our result shows that RFD scoring from 5000 images can automatically rank the best robust filter-set for obtaining the best 6-cell type classification model (94% accuracy). Moreover, the RFD score reflected the differences between the worst and the best classification models for morphologically similar spheroids, 60% and 89% accuracy respectively. In addition to RFD scoring, we found that using the time-course of morphological features can augment the fluctuations in spheroid recognitions leading to robust morphological analysis.

1. Introduction

Spheroids, *in vitro* three-dimensionally cultured cellular aggregates, have been shown to mimic *in vivo* biological functions compared with two-dimensionally cultured cells [1–4]. Therefore, their importance in drug development research has grown. To understand the physiological responses for testing pharmaceutical efficacy and safety, human cell-derived spheroids have been studied as replacements for animal models as a new *in vitro* drug screening platform. Cancer spheroids [5–7], liver spheroids [8], and heart spheroids [9–11] represent some of the cutting-edge cell applications in development. Moreover, based on recent advances in stem cell engineering, stem cell-derived spheroids are expected to be applied clinically [12,13]. In tissue engineering applications, spheroids are used as building blocks to manipulate larger scale tissues or organs [11,14].

One of the most advantageous features of spheroids is the balance of their biological complexity and their scalability. From the

[*] Corresponding author. Graduate School of Pharmaceutical Sciences, Nagoya University, Furocho, Chikusa-ku, Nagoya 464-8601, Japan. FAX: +81 52 747 6813.
E-mail address: kato-r@ps.nagoya-u.ac.jp (R. Kato).
Peer review under responsibility of the Japanese Society for Regenerative Medicine.

https://doi.org/10.1016/j.reth.2020.02.004

aspect of screening, spheroids are highly compatible with high-throughput screening technologies, such as multi-well plate assay systems or high content analysis platforms [15]. From the aspect of manufacturing, spheroids can enable the highest efficiency in large-scale cell source processing, up to the scale of 10^{10} cells, such as in induced pluripotent stem cell manufacturing [16] or mesenchymal stem cell-derived implantable tissues [12]. Despite the growing expectations for such spheroid applications, the technology to control the quality of spheroid production is still limited. Although it is essential to prepare massive numbers of spheroids with controlled quality for any application, spheroid evaluation technology, which can balance three important criteria ("efficiency," "resolution," and "non-invasiveness") is still lacking.

For spheroid evaluation, conventional biochemical assay techniques can feasibly expand their evaluation throughput. However, their evaluation per spheroid is limited to measuring the average value of all spheroid-comprising cells and difficult to discriminate their delicate differences. Conventional molecular biology techniques, such as sequencing or quantitative PCR analysis, can sensitively measure their differences, although still costly for high throughput screening. High content imaging has great potential for obtaining single-cell or intracellular organelle level evaluation data per spheroid. However, the imaging resolution commonly negatively correlates with their throughput. Moreover, most of the fluorescent-staining-based techniques are limited to end-point assays, therefore evaluated spheroids cannot be further used for the leading applications. Non-invasive cell evaluation technologies have been introduced to evaluate *in vitro* three-dimensionally cultured cells including spheroid such as measurements for oxygen gradients [17] and optical coherence tomography [18]. Among these, label-free microscopic image-based analysis is one of the technologies which can balance "efficiency," "resolution," and "non-invasiveness."

Our group has applied label-free image-based morphology analysis for enabling quantitative, high throughput, and non-invasive profiling of cells [19,20], colonies [21,22], and cell aggregates [23]. Marklein et al. reported high content imaging of early morphological signatures of human mesenchymal stem cells [24]. Oja et al. have also reported image-based analysis to detect aging in clinical-grade mesenchymal stromal cell cultures [25]. Maddah et al. have reported the application of a system for automated morphology-based evaluation of induced pluripotent stem cell cultures [26]. Although such label-free image-based analysis works have been growing, studies discussing the robustness of their analysis performance is still scarce. For better image-based analysis, especially for cell manufacturing applications, it is crucial to investigate the robustness of image-based quality evaluations, balancing its accuracy and reproducibility. In this work, we investigated to develop the concept to maximize the "reproducibility" of label-free morphology-based analysis for spheroids.

Generally, the workflow of conventional image-based cell evaluation analysis consists of 3 steps: recognition, measurement, and analysis (Fig. 1, Supplementary information Fig. S1, S2). The very first step, target recognition, is the image processing step, which try to recognize the region of "spheroid area" by the combination of image processing filters for further measures. Although it has a critical impact on all subsequent processes, the recognition of "whole spheroid" has been a very subjective process, rather than an evidence-based process. One of the biggest reasons for the subjectivity is that the three-dimensional spheroids have fine contrast area with their main body, although their outer surface region with loose aggregates makes poor contrast (Supplementary information Fig. S3). By such ambiguous contrast, the definition of the outline of whole spheroid can vary significantly between operators who design the filter-set. In other words, in spheroid images, it is a fact that there exists an "uncertain area" (Supplementary information Fig. S3A) at the outer region of spheroid which reflect the spheroid quality, however their recognition level is highly dependent on operators' decisions. Since operators commonly design their filter-set for label-free images only with limited and representative images, and evaluate their performance only by their feelings, the unexpected variation of "uncertain area" can critically fail the recognition process and disturb the subsequent analysis (Supplementary information Fig. S3B). For example, even if a filter-set was designed to "sharply" recognize spheroids within the first small dataset, their recognition can be "loose" in the second dataset by the existence of new type of "uncertain area" in new data.

To solve this basic issue in morphological analysis, we here investigated the influence of non-robust recognition filter-sets, and propose a "recognition fitness deviation (RFD)" as a new scoring index to objectively rank the most robust recognition filter-set which leads to the best analysis performance. To investigate this concept, we compared the effect of three types of spheroid recognition filter-sets (designated as recipes) and investigated their effects on cell type classification performances only from their label-free images. For this model, phase-contrast microscopic images of spheroids, including cancer cells (A-498, A549, NCI–H23, U-251) and healthy cells (HASMC, NHDF), covering different or similar morphological features, were analyzed. Our RFD scoring, which is designed to reflect the deviation of recognition fitness toward different replicate samples, time points, and cell types, was shown to quantitatively indicate the most robust spheroid recognition recipe, which leads to the best cell type classification model using only spheroid morphology. Moreover, our investigation indicated that time-course morphological feature usage could complement the fluctuations of designed recipes and improve the analysis performance in combination with RFD evaluation.

2. Methods

2.1. Cell culture

Healthy human dermal fibroblast cells (NHDF (Lot No. 01439)), human aortic smooth muscle cells (HASMC (Lot No. 01293)), human adenocarcinoma cells derived from lung cancer (A549 (Lot No. 60150896)), human lung adenocarcinoma cells (NCI–H23 (Lot No. 58078626)), human renal cancer cells (A-498 (Lot No. 58033335)), and human astrocytoma cells (U-251 (Lot No. unidentified)) were used. Cell culture was performed using an appropriate medium according to the culture protocol described in the product information sheet (American Type Culture Collection). The cells were seeded on a 10-cm dish (172958, Thermo Fisher Scientific Inc., Waltham, MA, USA) and cultured. Each medium contained 10% fetal bovine serum (Lot No. 13N059, 172012-500 ML, Nichirei Bioscience, Tokyo, Japan) and 1% penicillin and streptomycin (26253-84, Nacalai Tesque, Kyoto, Japan) was added, and the cells were cultured at 37 °C, under 5% CO_2. Cell suspensions were seeded in a Prime Surface 96-well plate (MS9096U, Sumitomo Bakelite, Tokyo, Japan) at a concentration of 1500 cells/well for spheroid formation.

2.2. Image acquisition

Phase-contrast images (1000 × 1000 pixels) at 4× magnification were captured using the automatic cell culture observation system BioStation CT (Nikon, Tokyo, Japan) at intervals of 6 h for 38 times over approximately 9 days. In this study, we selected time point 24 (= 144 h) to time point 32 (= 192 h), representing a total of 9 time points, as the period where spheroids formed stably for the analysis. For each cell type, 24 spheroids were prepared for each sample

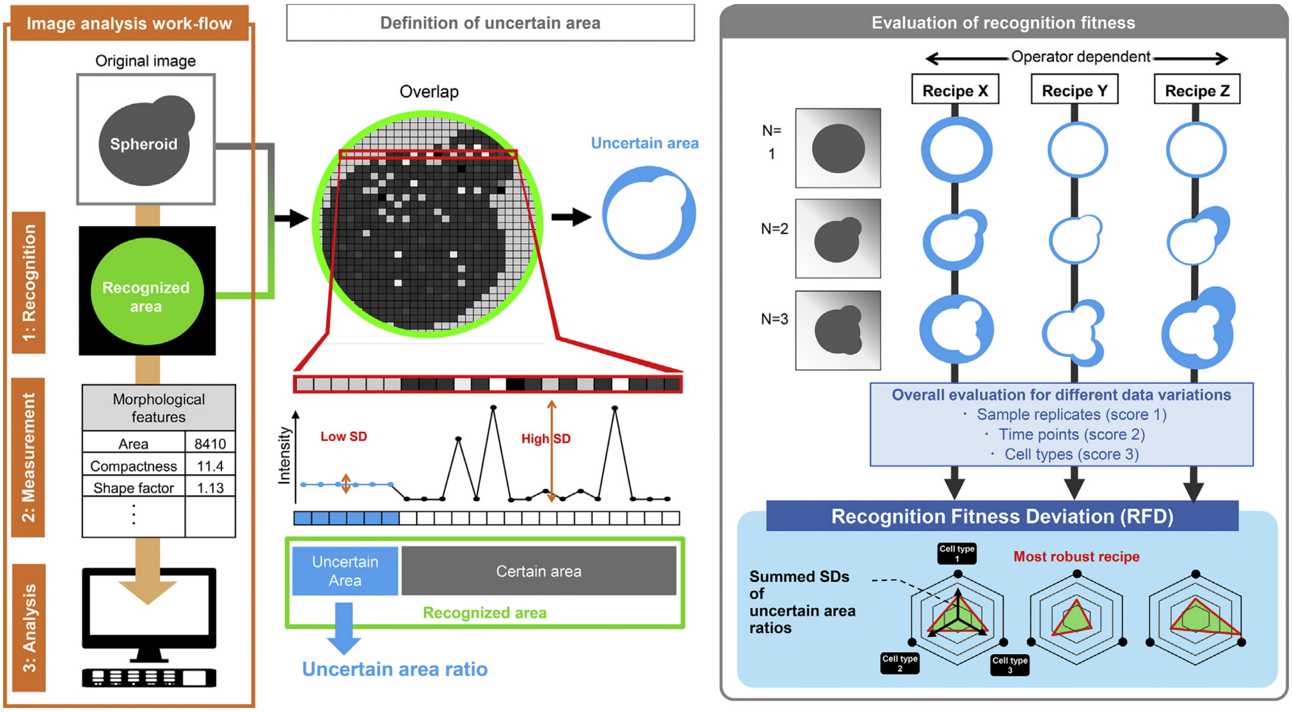


Fig. 1. Schematic illustration of this study. (Left: Orange column) The work-flow in the illustration indicates the conventional image analysis scheme for morphology-based analysis. From the original image, the objective target in the image (spheroids in this study) is recognized by image processing (step 1: Recognition). The recognized area (colored in green) is commonly designed to cover the total spheroid area including their outer borders. Then, from the recognized area, morphological features are measured (step 2: Measurement). Using morphological features as multiple descriptors of the objective target, further analysis (step 3: Analysis) is conducted. (Center column) The uncertain area, the gap between the "recognized area" and the "certain area," is defined as the uncertain area ratio in the image. Because annotation of the true spheroid area is difficult in label-free images, we defined the uncertain area by calculating the "area with low-intensity SD" to measure the recognition fitness in each image. Practically, within the recognition area (green), the uncertain area (light blue) is flagged, and their total ratios were scored as a "uncertain area ratio." It should be noted that the "certain area" only defines the region of main body of spheroid, and its outer border in the whole spheroid is dependent on the recipe. Such certain area ratios are called "recognition fitness" in our study. (Right: Grey column) At present, the fitness and performance of a recipe is highly dependent on operators. In this study, we evaluated such recognition fitness with a more objective scoring criterion, the recognition fitness deviation (RFD). In this concept, the importance of evaluating a recipe by the summary of all the uncertain area ratios in each image, the SD value within the fitness of the recipe for data variations is proposed. By summarizing the SDs of uncertain area ratios with three types of scores, the radar chart can be illustrated to show the robustness of the recipe. The smaller the RFD (the area of radar chart), the more robust the recipe is.

replicates to make total 5328 images (24 spheroids × 37 time points × 6 cell types). The images only from the successfully formed spheroids without noise were selected for analysis (19 for A-498, 21 for A549, 18 for NHDF, 17 for HASMC, 23 for NCI–H23, and 19 for U-251 (total 117)).

2.3. Image processing

CL-Quant software version 3.20 (Nikon, Tokyo, Japan) was used to design filter-sets for recognizing spheroids in images and to measure the recognized area. Multiple combinations of image processing filter-sets were called "recipes." Three different recipes designed with different concepts were compared; Recipe A: using the soft-matching function in CL-Quant, which is the automated machine-learning algorithm based on user-selected areas. Recipe B consisted of: (1) normalization of the background; (2) soft-matching; (3) removal of objects; (4) filling holes. Recipe C included: (1) normalization of the background; (2) thresholding of the intensity; (3) opening and closing to fill holes; (4) flattening of the background; (5) thresholding of the intensity; (6) opening and closing to fill holes; (7) merge recognition areas from step 3 and 6; (8) fill holes. The three recipes were designed by three different operators. Recipe A was designed using 10–20 images from only NHDF. It was intended to capture the spheroid surface area information. Recipe B was modified from recipe A, but was optimized to fit cancer spheroid images (different from the cancer cells used in this study), and was intended to fit sharply to those cancer spheroids, but was first to be applied to the six cell types in this work. Recipe C was designed with 10–20 images of all the six cell types used in this work. However, recipe C was intended to emphasize the subtle differences of the spheroid surface areas, which tend to become characteristically loose for some cell types. From the recognized area using each recipe, a total of 11 morphological features (area, compactness, correlation mean, energy mean, entropy means, homogony mean, inertia means, length: width ratio (ratio of length/width), perimeter, shape factor, std dev (standard deviation of) intensity) were analyzed (Supplementary information Table S1). The feature calculation details are described in previous works [19–23]. Each feature was normalized with standard normalization for further analysis.

2.4. Morphological analysis for comparing the performances of the recipes

For the comparison of the effects from the different morphological features extracted from different recipes, the similar morphological features were analyzed using hierarchical clustering based on Euclidean distance. To further compare the differences of recipes, ridge regression was performed to classify six cell types with leave-one-out cross-validation. In the clustering and classification, the effect of morphological features was compared using only time point 3, or all time points. All analysis was performed using R software (version 3.2).

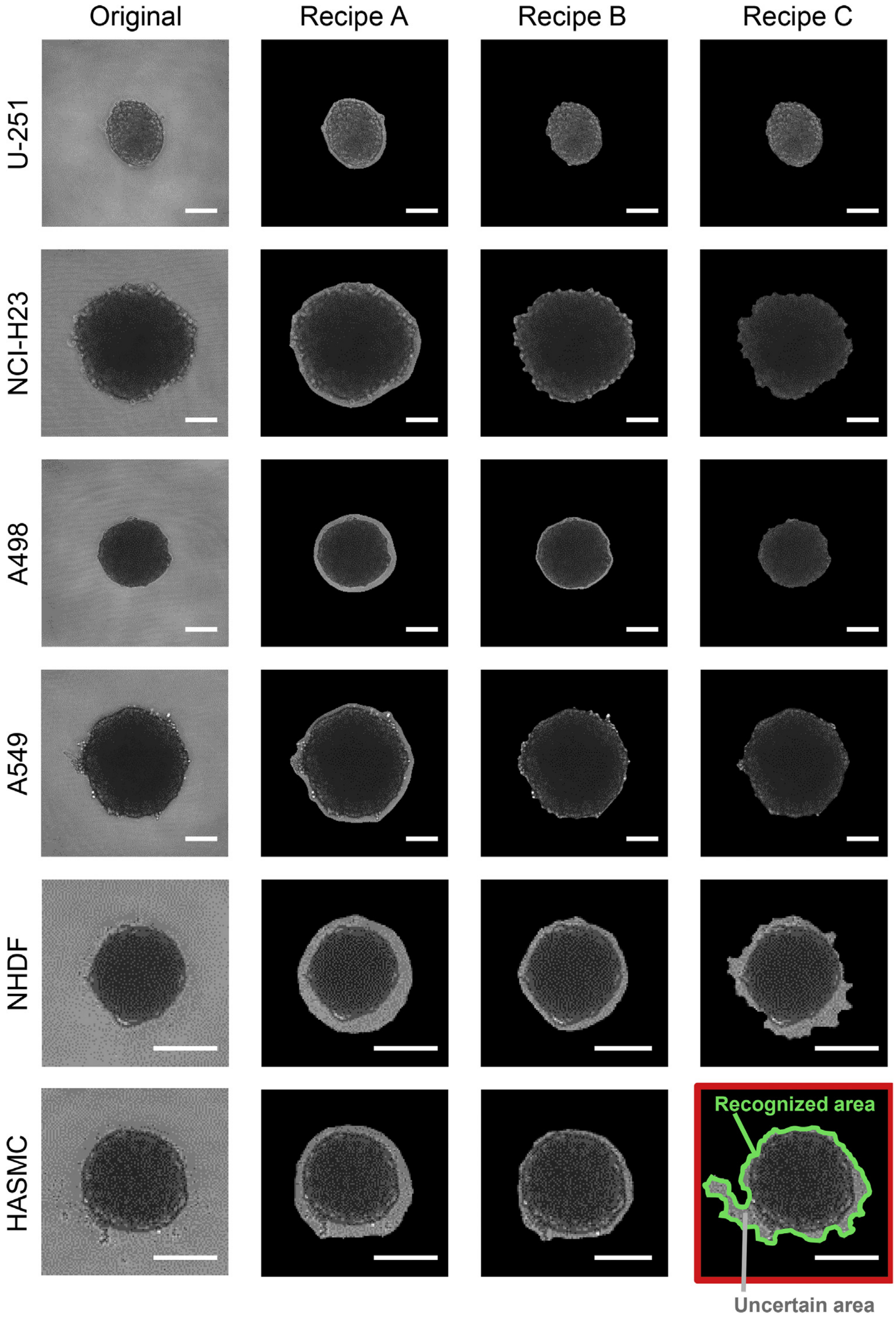


Fig. 2. Representative images of spheroids and their recognition. Using the same initial phase-contrast image (left column), three different types of recipes (recipe A, recipe B, and recipe C) were applied to recognize the same spheroid. The black area is the non-spheroid area defined by each recipe, and in the recognized area, the raw spheroid image (left column) is overlaid to indicate the “uncertain area ratio” visually. The row shows their morphological and recognition differences in six cell types. White bars in U-251, NCI–H23, A498, and A549: 75 μm, in NHDF and HASMC: 150 μm. In the bottom row, the recognized area outline (green line) visualizes the “outline of whole spheroid,” which varies greatly among the recipes. In other words, it indicates that there are conceptual differences of recipes regarding fitting sharply or loosely to capture the spheroid surface morphology.

2.5. Measurement of recognition fitness deviations

To quantitatively evaluate the spheroid recognition fitness in all images, the recognition fitness deviation (RFD) was designed to score the robustness of image processing recipes (Fig. 1). First, for RFD calculations, the “uncertain area ratio” in each image was calculated. In each image, the uncertain area ratio was defined as the ratio of the “uncertain area” (Supplementary information Fig. S1A) in the “recognized area” by the recipe. The “uncertain area” is the area with poor contrast, therefore implementation of outline region can vary between the operators. The “certain area” is the spheroid main body area, where contrast is clearer and operators tend to recognize easily. In our study, we defined the “uncertain area” as a low-intensity value (<55) which repeats within a 5-pixel horizontal window, since in the raw image, the “certain area” of spheroid commonly shows high-intensity standard deviation (SD), and the rest of the image field shows faint intensity differences (Fig. 1A). It is important to note that the “high-intensity SD area,” which we call the “certain area” is the area, which probably includes the true spheroid main body, but is not the whole spheroid. Limited recognition of such area loses the characteristics of spheroids. Using the quantitative definition of uncertain area ratio, we could compare the area where operators differ in implementation in all the images with the same quantitation criteria automatically. When all the uncertain area ratios are calculated for all images using different recipes, the SD values were calculated within different sample replicates (score 1), different time points (score 2), and different cell types (score 3). The sum of scores 1 to 3 reflects the deviation of each recipe's recognition performance. If the deviation is high, the recipe is not working reproductively in some samples. Therefore, we designated these as RFD. As an illustration, sum of score 1 and score 2 is plotted in each hexagon axis per each cell type, and the total area of the radar chart reflects score 3 (Fig. 1).

3. Results

3.1. Diversity of spheroid morphology and fluctuation of spheroid recognition

In this study, six types of cells were selected to mimic the varieties of spheroids (Fig. 2A). Even by seeding the same cell number in a well, their morphologies were found to have diversity with their details. First, size difference is a clear morphological feature. Some spheroids shrink to a smaller size (U-251 and A-498) than other spheroids during the culture. Comparing such sizes in cancer cells (U-251, NCI–H23, A-498, and A549), the difference between normal and cancer cannot be categorized simply with their sizes. Second, the tightness of spheroid aggregation, reflected by the intensity distribution in the spheroid area, is also a characteristic feature. Although there is a slight tendency for normal cell spheroids to appear brighter, it is difficult to classify them as normal or cancer cells or their tissue origins. Therefore, it was clear that spheroid morphological features were more complicated than their differences in morphological characteristics, as in two-dimensionally cultured cells.

To analyze spheroid varieties using image analysis, we compared three recipes (A–C; Fig. 2A). The three recipes were designed by three operators aiming for the same goal, the morphological analysis of spheroids. However, their analytical situations and concepts for designing their recipes were different. They differed not only in their filter-set combinations but also the data which each operator focused upon to develop their recipe. For the design of recipe A, the operator utilized images of only one cell type (NHDF). The operator attempted to recognize the outermost surface of spheroids since NHDFs tend to aggregate loosely, and some cells float to the surface. For the design of recipe B, the operator used images of other cancer cell spheroids, which were not included in the six prepared cell types, and modified recipe A to fit the cell type. Thus, different filters were added in recipe B. For the design of recipe C, the operator utilized images of all six cell types and attempted to create a recipe to recognize various types of cells from scratch. However, in this recipe design, the operator attempted to augment the differences between various spheroids by rendering a recipe that sensitively recognizes the differences of spheroid surface collapse. As a result, their recognition of fitness was found to show diversity.

However, the characteristics of these recipes were only evident when their recognition results for all cell types were paneled for visualization. For example, comparisons for single-cell types, such as U-251 or A-498, did not show clear differences between recipes. By the paneled comparison, recipe A showed overall “fat” recognition, recipe B showed overall “fit” recognition, and recipe C showed fluctuated recognition, which was “invasive” or “disordered,” sensitively reflecting the spheroid surface status. It is essential to realize that such paneled comparison results shown in Fig. 2A are only partial results in the more than 5000 images, showing one timepoint with one image from three replicates. This result strongly indicates that a recipe evaluated only by limited numbers of images does not assure robust performance and can critically disrupt further analysis when the number of images or cell types increases. In other words, the robustness of image processing requires an evaluation from the aspect of its overall performance toward the varieties of data by some quantitative index.

3.2. Evaluation of a recipe's robustness with recognition fitness deviation

To quantitatively and comprehensively evaluate the recipes, we analyzed the “recognition fitness,” which derives from the gap between the “recognized area (defined by the recipe designer's implementation)” and the “certain area (where spheroid main body can be clearly defined)”, designated as “uncertain area” (Fig. 1 and Supplementary information Fig. S1). To objectively score such fitness, we here introduced an algorithm to measure the “uncertain area ratio”. By analyzing large image data covering the variation in the sample replicates, time points, and cell types, we compared the recipe's reproducibility in their recognition fitness (Fig. 3A). In the plots of the uncertain area ratio, we found that even under replicate conditions (17–23 spheroids per each condition), there were outliers indicating unexpected recognition results.

Moreover, the outlier deviation of uncertain area was found in the time course as well (ex. Recipe C for HASMC recognition). When such recognition performances were paneled within all cell types, it was again found that there are specific cell types that show large deviations (ex. NHDF and HASMC recognition by recipe C). By detailed confirmation of each recipe's recognition image, the uncertain area ratio was visually confirmed to reflect the unexpected failures in recognition reproducibility (Fig. 3B and C). In our data, the NHDF and HASMC were the two most difficult spheroids to be robustly recognized.

To further confirm the influence of such fluctuations in recipe performance, we compared the effect on the measured morphological features from their recognized areas (Fig. 3D, and Supplementary information Fig. S3). Even from the same spheroid, the morphological features were found to show significant differences when the deviations of the uncertain area ratios were high. This result indicated that if a recipe were not robust enough, the measured morphological features can contain significant noise.

To visually and quantitatively evaluate the total performances of the recipes, we summarized the SDs of the uncertain area ratio within identical replicate samples (score 1), time points (score 2), and cell types (score 3) to visualize in a radar chart (Fig. 3E). In this visualization, the small and uniform radar area indicates the robustness of a recipe. With this scoring, recipe A can be ranked as the best of the three recipes quantitatively. We further designated these summed SDs of uncertain area ratios as “recognition fitness deviations (RFD).”

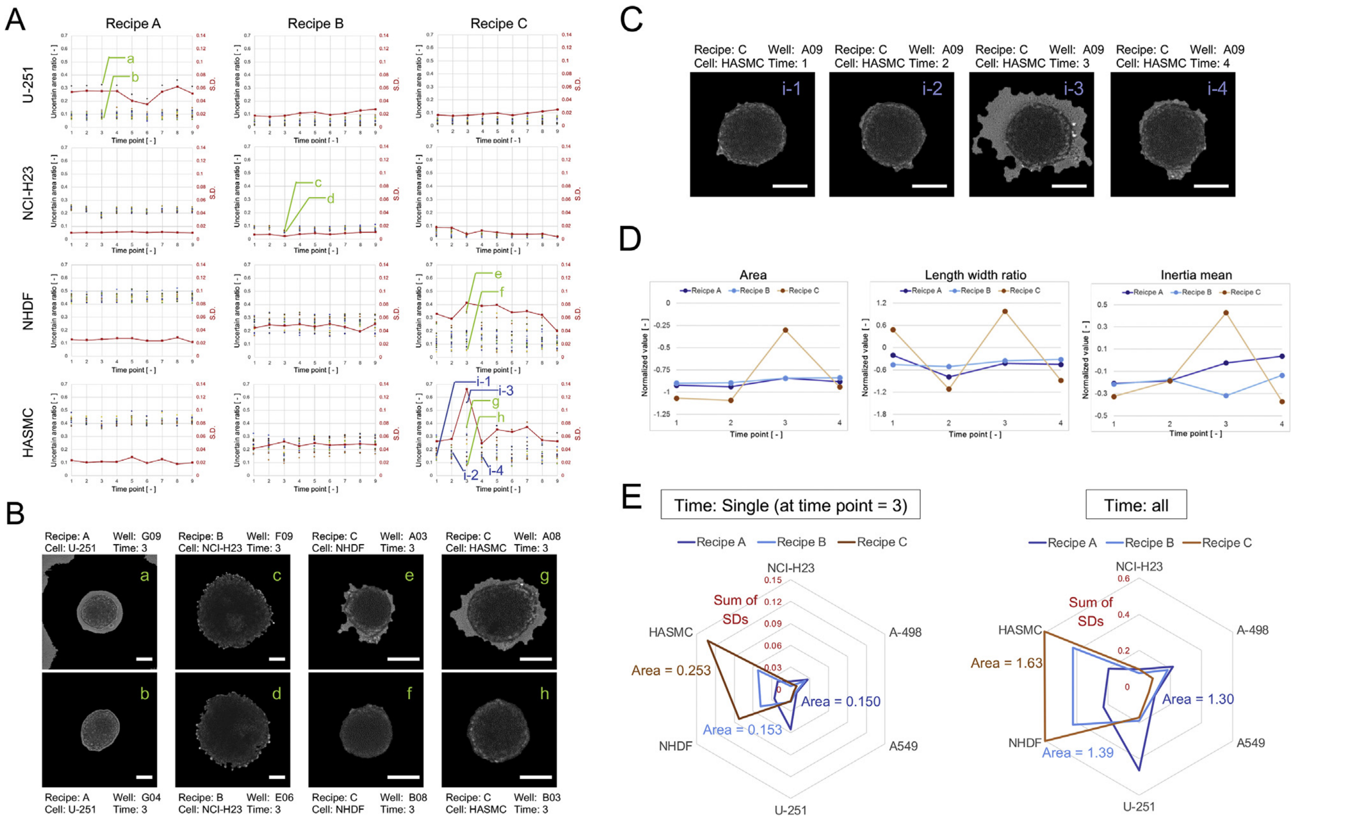


Fig. 3. Evaluation of recipes using recognition fitness deviation (RFD). (A) Variation of uncertain area ratios among varieties of data (varieties within sample replicates, time points, and cell types). In each graph, the X-axis shows the time point (6 h), the first Y-axis (left) shows the uncertain area ratio for each plot in color (each spheroid), and the second Y-axis (right) shows the time-course changes in SD summarizing 17–23 plots (red line). Among the six-cell types evaluated, the red line pattern was similar among A-498, A549, and NCI–H23; NCI–H23 is only shown as a representative. The alphabetically indicated plots, a–h, are representative spheroid examples to indicate different uncertain area ratios in 3B. The alphabetically indicated (i-1)–(i-4) are representative cases of the different uncertain area ratios found in the same spheroid in 3B. (B) Representative images to indicate the uncertain area ratios between spheroids and their recognition areas. (a–h) indicates the alphabetically indicated plots in 3A. U-251 and NCI–H23, and NHDF and HASMC are lined vertically, to compare the differences of recognition fitness between pairs of similar spheroid morphologies. White bars in a–d: 65 μm, in e–h: 130 μm. (C) Representative images are indicating the uncertain area ratios within the time-course in the same spheroid (HASMC). (i-1) –(i-4) indicates the alphabetically indicated plots in 3A. All white bars: 130 μm. (D) The fluctuation of morphological features measured from different recognition fitness within the spheroid (HASMC) captured in 2C. The X-axis indicates the time points (6 h), and the Y-axis indicates the normalized feature values (Area, Length: width ratio, and Inertia mean). (E) The RFD evaluation of the three recipes, as a summary of recognition fitness among all time-points, spheroid image replicates, and cell types. Specifically, fitness among six cell types is indicated with the hexagon axis as a radar chart. If a recipe can be robustly used for a variety of cells, time-courses, and image replicates, the RFD (radar area) becomes smaller.

3.3. Effects of a recipe's recognition in further morphology-based analysis

When a recipe results in poor recognition robustness, the further measured morphological features can poorly express the morphological characteristics of spheroids. As a result, the following analysis using morphological features is affected.

As a model case of this confirmation, we compared the analysis results (step 3 in Fig. 1) based on the three recipe's spheroid recognition and morphological measurement. First, by using the morphological features from a single time point (time point 3), we compared the recipe's effect in hierarchical clustering. In this analysis, we focused on how NHDF and HASMC, the most difficult spheroids to discriminate, would be clustered. The NHDF and HASMC spheroids were mostly clustered in separate clusters with recipe A, whether both cell types were combined in the same cluster in the other two recipes (Fig. 4A). Especially in recipe C, the morphological characteristics indicated by the heatmap became faint, because there appeared a “peaky morphological values” in the total morphological data by the fluctuating recognition.

We further utilized the same morphological features for the classification of six cell types by ridge regression. Recipe A showed better performance both in the (1) six-cell type classifications, and the (2) NHDF/HASMC classification, compared with other recipes. Especially, with the most challenging spheroids to be recognized, the classification of NHDF and HASMC tend to fail in recipe B and recipe C (Fig. 5A). This result indicates that the performance of the morphology-based prediction model can be critically affected by the performance of the utilized image processing recipe because the measured morphological features had less interpretable information. Moreover, recipe A, which showed the lowest RFD score, was found to show the best performance.

Second, since our previous studies showed the importance of using time-course morphological features in morphology-based predictions [21], we investigated the effect of using time-course information on recipe's performance. With hierarchical clustering, the spheroids of NHDF and HASMC were clustered in different clusters in recipe A, although they tend to be mixed under closer trees in the rest of the recipes (Fig. 4B). However, with recipe B, the mis-clustering rate was improved compared with the clustering results using only a single time point (Fig. 4A). With the six-cell type classification, the performances of all three recipes increased with both (six cell types, and NHDF/HASMC classification) compared with the classification model using only a single time point (Fig. 5). Moreover, the classification of NHDF or HASMC with recipe B and recipe C were significantly improved. Consequently, by

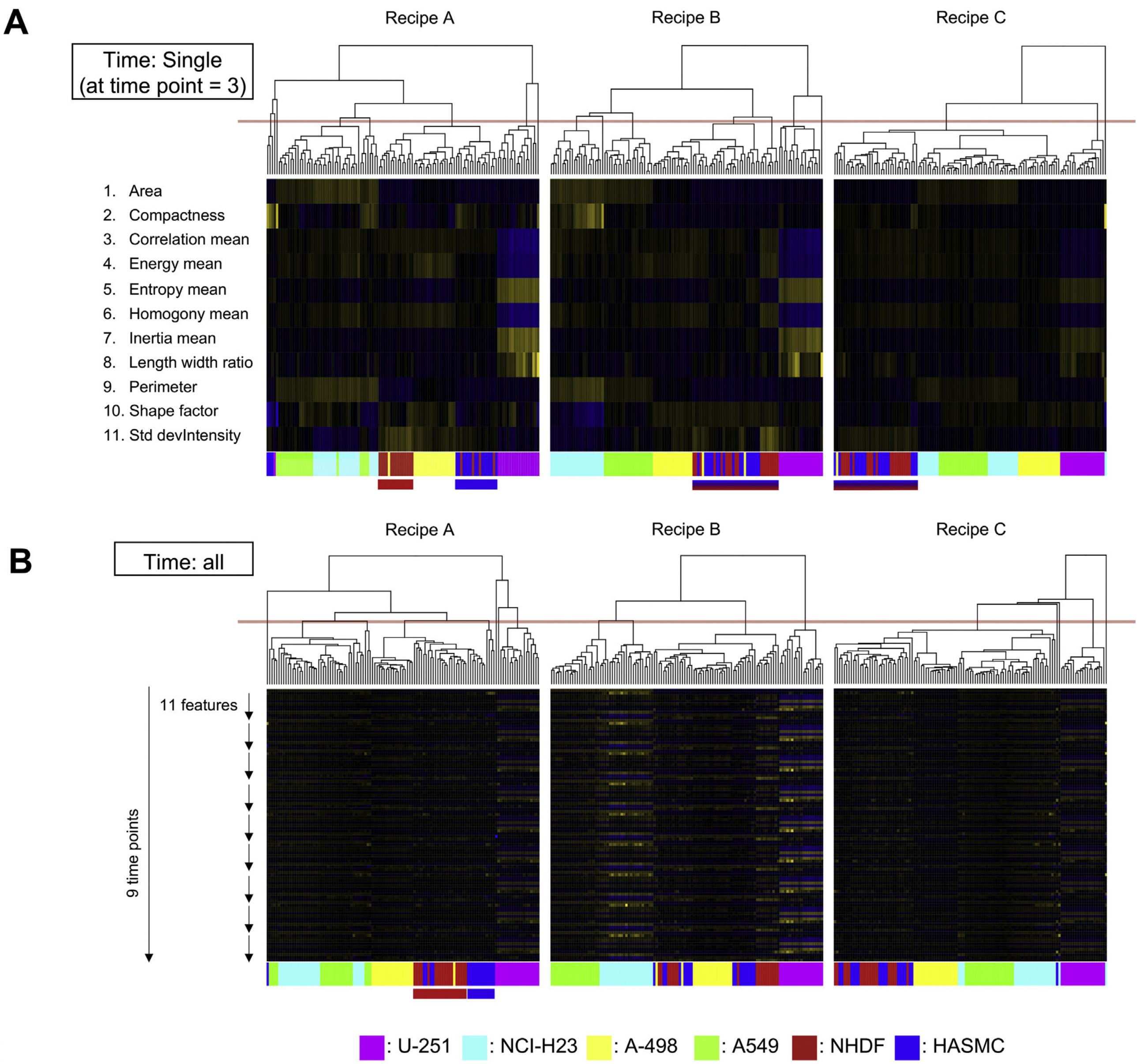


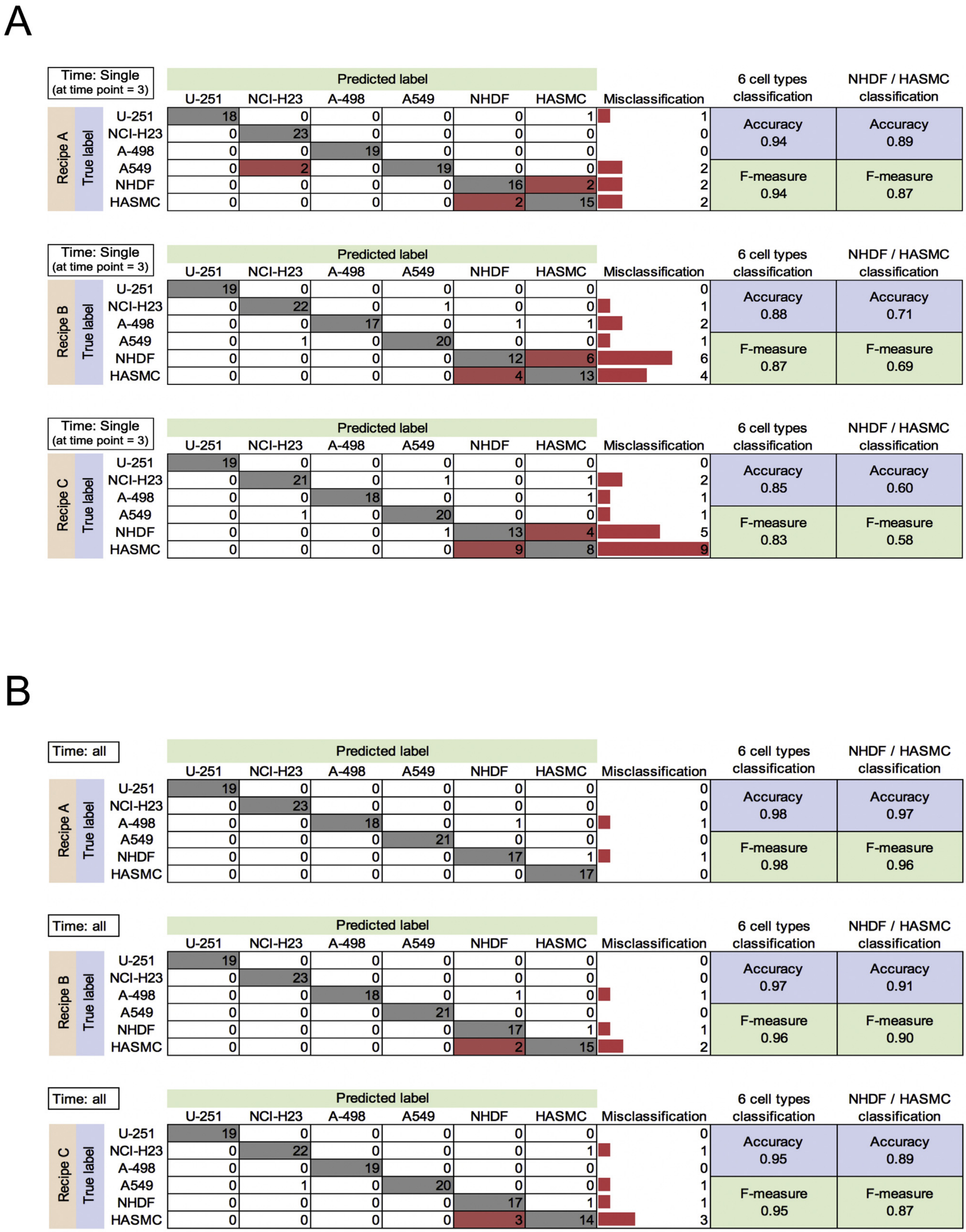
Fig. 4. Comparison of hierarchical clustering results of spheroid morphologies between recipes. (A) Comparison of recipes with morphological features at time point 3. Columns: 17–23 spheroids per six cell types (117 spheroids). Row: 11 morphological features at only time point 3. (B) Comparison of recipes with morphological features at all time points. Columns: 17–23 spheroids per six cell types (117 spheroids). Row: 11 morphological features × nine time points. The heatmap indicates the normalized value for each feature (blue: low, yellow: high). The color label for each spheroid column under the clustering indicates cell types: pink, U-251; light blue, NCI–H23; yellow, A-498; green, A549; red, NHDF, blue, HASMC. The colored bars indicate the cluster of morphologically similar cell types: red, NHDF, blue, HASMC. The red horizontal line at the tree indicates the Euclidian distance = 10.

the use of time-course morphological features, the performance itself and the deviation of performances among recipes were found to be improved. However, it should be noted that even with such performance improvements using time-course data, recipe B and recipe C showed lower performances when compared with recipe A, the lowest RFD scoring recipe.

4. Discussion

In this work, to obtain robust performance in the morphological image-based evaluation of cells, we propose RFD scoring concept as a means of objectively selecting the recipe, instead of the conventional expert experience-based selection, to enhance the reproducibility of spheroid image analysis.

The reproducibility of image processing has rarely been discussed and quantitatively scored, since most of the image processing design was accomplished by manual trial and error until the operator was satisfied. This experience-based image processing design has been the standard in most image processing studies in various fields. To explore evidence-based options to the subjective operator-biased image processing pipeline, we presently propose an objective scoring concept to evaluate image processing recipes (Supplementary information Fig. S1 and S2). Our strategy is simple. Instead of insisting on a recipe based on limited evaluation, our RFD scoring enables the comprehensive and automatic evaluation of the recipe's performance for all the possible variations of the acquired images. Faced with a great variation and volume of images, the RFD scoring can perform the evaluation automatically, instead of



Time: Single (at time point = 3)		Predicted label							6 cell types classification	NHDF / HASMC classification
		U-251	NCI-H23	A-498	A549	NHDF	HASMC	Misclassification		
Recipe A True label	U-251	18	0	0	0	0	1	1	Accuracy 0.94	Accuracy 0.89
	NCI-H23	0	23	0	0	0	0	0		
	A-498	0	0	19	0	0	0	0		
	A549	0	2	0	19	0	0	2	F-measure 0.94	F-measure 0.87
	NHDF	0	0	0	0	16	2	2		
	HASMC	0	0	0	0	2	15	2		

Time: Single (at time point = 3)		Predicted label							6 cell types classification	NHDF / HASMC classification
		U-251	NCI-H23	A-498	A549	NHDF	HASMC	Misclassification		
Recipe B True label	U-251	19	0	0	0	0	0	0	Accuracy 0.88	Accuracy 0.71
	NCI-H23	0	22	0	1	0	0	1		
	A-498	0	0	17	0	1	1	2		
	A549	0	1	0	20	0	0	1	F-measure 0.87	F-measure 0.69
	NHDF	0	0	0	0	12	6	6		
	HASMC	0	0	0	0	4	13	4		

Time: Single (at time point = 3)		Predicted label							6 cell types classification	NHDF / HASMC classification
		U-251	NCI-H23	A-498	A549	NHDF	HASMC	Misclassification		
Recipe C True label	U-251	19	0	0	0	0	0	0	Accuracy 0.85	Accuracy 0.60
	NCI-H23	0	21	0	1	0	1	2		
	A-498	0	0	18	0	0	1	1		
	A549	0	1	0	20	0	0	1	F-measure 0.83	F-measure 0.58
	NHDF	0	0	0	1	13	4	5		
	HASMC	0	0	0	0	9	8	9		

Time: all		Predicted label							6 cell types classification	NHDF / HASMC classification
		U-251	NCI-H23	A-498	A549	NHDF	HASMC	Misclassification		
Recipe A True label	U-251	19	0	0	0	0	0	0	Accuracy 0.98	Accuracy 0.97
	NCI-H23	0	23	0	0	0	0	0		
	A-498	0	0	18	0	1	0	1		
	A549	0	0	0	21	0	0	0	F-measure 0.98	F-measure 0.96
	NHDF	0	0	0	0	17	1	1		
	HASMC	0	0	0	0	0	17	0		

Time: all		Predicted label							6 cell types classification	NHDF / HASMC classification
		U-251	NCI-H23	A-498	A549	NHDF	HASMC	Misclassification		
Recipe B True label	U-251	19	0	0	0	0	0	0	Accuracy 0.97	Accuracy 0.91
	NCI-H23	0	23	0	0	0	0	0		
	A-498	0	0	18	0	1	0	1		
	A549	0	0	0	21	0	0	0	F-measure 0.96	F-measure 0.90
	NHDF	0	0	0	0	17	1	1		
	HASMC	0	0	0	0	2	15	2		

Time: all		Predicted label							6 cell types classification	NHDF / HASMC classification
		U-251	NCI-H23	A-498	A549	NHDF	HASMC	Misclassification		
Recipe C True label	U-251	19	0	0	0	0	0	0	Accuracy 0.95	Accuracy 0.89
	NCI-H23	0	22	0	0	0	1	1		
	A-498	0	0	19	0	0	0	0		
	A549	0	1	0	20	0	0	1	F-measure 0.95	F-measure 0.87
	NHDF	0	0	0	0	17	1	1		
	HASMC	0	0	0	0	3	14	3		

Fig. 5. Comparison of the confusion matrix of cell-type classification performances using morphological features from different recipes and their time points. (A) The classification performances of models using morphological features of time point 3. (B) The classification performances of models using morphological features of all time points. In the matrix, the numbers indicate the counts of spheroids classified by the model. The grey cell indicates the correctly classified, and the red cell indicates the misclassified (≥2). On the right, the total misclassification spheroid number is indicated.

necessitating an image-by-image manual scrutiny by the operator. Therefore, by its nature, RFD scoring can be applied to any type of image analysis of spheroids, including spheroid viability prediction, spheroid metabolic potency assessment, spheroid differentiation analysis, and spheroid morphometry measurement, as the first step of image analysis to compare the custom-made recipes for each data. As an analogy, our RFD score for selecting robust recipe can be interpreted as "Melting temperature value for selecting robust primers in quantitative polymerase chain reaction" (Supplementary information Fig. S1).

Our data revealed the risk and importance of evaluating the "reproducibility" of manually designed spheroid recognition filter-sets, and proposed the RFD scoring to comprehensively and automatically evaluate their performances. To the best of our knowledge, this study is the first to show that a label-free morphological feature can discriminate 6 different cell qualities (the difference of cancer or normal, or differences in cancer cell types). The study is also the first to investigate the reproducibility of spheroid image analysis with 5000 experimentally obtained spheroid images.

Our data clarified that different recipes designed with different concepts show significant variations in spheroid recognition, especially when they are checked in a panel with different data variations. By the objective quantification of such recognition fitness, we found that such variability of recipe performance can occur not only between cell types, but also during their time course, and even within replicated samples. Therefore, to design a robust recipe that can be promise reproducible results in big image data analysis, our data suggest that human-dependent or self-proposed recipe evaluation has a considerable reproducibility risk. Therefore, in this study, we evaluated the cell recognition performances widely throughout the data with a new quantitative scoring index, the RFD.

To investigate the performance of our RFD scoring, we compared three recipes designed differently using different concepts and compared their actual performances not only in the recognition step but also in their morphology-based analysis steps. As a result, we found a good negative correlation of the proposed RFD score and their analysis performances, where a lower RFD indicates a higher recipe reproducible performance. With recipe A and recipe B (modified from recipe A), both recognition performance seemed very similar at a glance (Fig. 2A). However, when RFD were scored in detail and overall, the score indicated higher robustness for recipe A, and recipe A resulted in the best performances in clustering and classification.

In the morphology-based analysis, we examined the effect of time-course data usage in different recipes. As a result, both the results of clustering and classification, the performance of all recipes, including the lower performing recipes (recipe B and recipe C), could be improved. Therefore, it was suggested that time course morphological information that can be obtained from label-free imaging could partially compensate for the fluctuation disorder of spheroid recognitions in less robust recipes. However, it is important to note that even with time-course data usage, the RFD score evaluation showed the best performing recipe, recipe A. In other words, it was clear that with lower RFD score recipe, which shows a lower rate of uncertain area recognition under any condition, works best as a morphology-based analysis model.

In this study, we focused on the issue of the "recognition of object," which is presently designed and evaluated manually by an operator, with spheroid images. However, with the recent progress of deep learning algorithms, biological-image analysis now has new approaches. There are algorithms to detect the object area with high precision by training the object feature through deep learning [27,28]. In such algorithms, the object, the spheroid in our case, can be recognized with higher recognition fitness compared with our three compared recipes. However, although any other algorithms may show higher fitness in some data, our work suggest that their robustness evaluation is more important. Moreover, with deep learnings, more and more volume of annotated spheroid images (which are difficult to obtain) is required to design robust recognition. Apart from the object recognition approach, there are also algorithms, which uses the total image pixel information including all object and background with convolutional feature extraction [29,30]. With such algorithms, there is no need to evaluate our recognition fitness, and their extracted information can be used for further analysis. However, even with such algorithms, our proposing concept of checking the robustness of the algorithm with varieties of data remains essential, because "automatic feature extraction ability" does not promise the robustness of image processing.

Considering the practical application of image-based cell evaluation technology with various types of patients or lot diversities, it should be essential to establish robust image processing to provide robust measurement results for subsequent analysis. Our RFD scoring concept will release the image processing from the present human-oriented decision in image processing design to lead for a more automated cell recognition process that can be optimized with the growth of data. Moreover, by introducing such scoring in image processing, the robustness of the recipe can be optimized using automated machine learning algorithms (Supplementary information Fig. S5). Further studies should evaluate the effect of our RFD scoring in more varied images, including different cell types, image magnifications, and images from different microscopes. We believe our work will contribute to the mechanization and automation of image-based in-process monitoring technology in cell processing.

Declaration of Competing Interest

A collaboration research support from Nikon Corporation was funded to Ryuji Kato. The first author Kazuhide Shirai is the employee of Nikon Corporation, who have been administrated as PhD candidate in the Graduate School of Pharmaceutical Sciences, Nagoya University.

Acknowledgments

This research was partially supported by the Japan Science and Technology Agency (JST) Program for Creating STart-ups as part of the Advanced Research and Technology (START Program) program, and with collaborative funding from Nikon Corp.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.reth.2020.02.004.

References

[1] Pampaloni F, Reynaud EG, Stelzer EHK. The third dimension bridges the gap between cell culture and live tissue. Nat Rev Mol Cell Biol 2007;8:839–45. https://doi.org/10.1038/nrm2236.

[2] Gaebler M, Silvestri A, Haybaeck J, Reichardt P, Lowery CD, Stancato LF, et al. Three-dimensional patient-derived in vitro sarcoma models: promising tools for improving clinical tumor management. Front Oncol 2017;7:1–14. https://doi.org/10.3389/fonc.2017.00203.

[3] Van Den Brand D, Massuger LF, Brock R, Verdurmen WPR. Mimicking tumors: toward more predictive in vitro models for peptide- and protein-conjugated drugs. Bioconjugate Chem 2017;28:846–56. https://doi.org/10.1021/acs.bioconjchem.6b00699.

[4] Hoffmann OI, Ilmberger C, Magosch S, Joka M, Jauch KW, Mayer B. Impact of the spheroid model complexity on drug response. J Biotechnol 2015;205:14–23. https://doi.org/10.1016/j.jbiotec.2015.02.029.

[5] Rodrigues T, Kundu B, Silva-Correia J, Kundu SC, Oliveira JM, Reis RL, et al. Emerging tumor spheroids technologies for 3D in vitro cancer modeling. Pharmacol Ther 2018;184:201–11. https://doi.org/10.1016/j.pharmthera.2017.10.018.

[6] Nunes AS, Barros AS, Costa EC, Moreira AF, Correia IJ. 3D tumor spheroids as in vitro models to mimic in vivo human solid tumors resistance to therapeutic drugs. Biotechnol Bioeng 2019;116:206–26. https://doi.org/10.1002/bit.26845.

[7] Gencoglu MF, Barney LE, Hall CL, Brooks EA, Schwartz AD, Corbett DC, et al. Comparative study of multicellular tumor spheroid formation methods and implications for drug screening. ACS Biomater Sci Eng 2018;4:410–20. https://doi.org/10.1021/acsbiomaterials.7b00069.

[8] Bell CC, Hendriks DFG, Moro SML, Ellis E, Walsh J, Renblom A, et al. Characterization of primary human hepatocyte spheroids as a model system for drug-induced liver injury, liver function and disease. Sci Rep 2016;6:1–13. https://doi.org/10.1038/srep25187.

[9] Ong CS, Fukunishi T, Zhang H, Huang CY, Nashed A, Blazeski A, et al. Biomaterial-free three-dimensional bioprinting of cardiac tissue using human induced pluripotent stem cell derived cardiomyocytes. Sci Rep 2017;7:2–12. https://doi.org/10.1038/s41598-017-05018-4.

[10] Beauchamp P, Moritz W, Kelm JM, Ullrich ND, Agarkova I, Anson BD, et al. Development and characterization of a scaffold-free 3D spheroid model of induced pluripotent stem cell-derived human cardiomyocytes. Tissue Eng C Methods 2015;21:852–61. https://doi.org/10.1089/ten.tec.2014.0376.

[11] Moldovan NI. Progress in scaffold-free bioprinting for cardiovascular medicine. J Cell Mol Med 2018;22:2964–9. https://doi.org/10.1111/jcmm.13598.

[12] Santos JM, Camões SP, Filipe E, Cipriano M, Barcia RN, Filipe M, et al. Three-dimensional spheroid cell culture of umbilical cord tissue-derived mesenchymal stromal cells leads to enhanced paracrine induction of wound healing. Stem Cell Res Ther 2015;6:1–19. https://doi.org/10.1186/s13287-015-0082-5.

[13] Tellez-Gabriel M, Cochonneau D, Cadé M, Jubelin C, Heymann MF, Heymann D. Circulating tumor cell-derived pre-clinical models for personalized medicine. Cancers (Basel) 2019;11:1–16. https://doi.org/10.3390/cancers11010019.

[14] Laschke MW, Menger MD. Spheroids as vascularization units: from angiogenesis research to tissue engineering applications. Biotechnol Adv 2017;35:782–91. https://doi.org/10.1016/j.biotechadv.2017.07.002.

[15] Moriconi C, Palmieri V, Di Santo R, Tornillo G, Papi M, Pilkington G, et al. INSIDIA: a Fiji macro delivering high-throughput and high-content spheroid invasion analysis. Biotechnol J 2017;12:1–7. https://doi.org/10.1002/biot.201700140.

[16] Petrenko Y, Syková E, Kubinová Š. The therapeutic potential of three-dimensional multipotent mesenchymal stromal cell spheroids. Stem Cell Res Ther 2017;8:1–9. https://doi.org/10.1186/s13287-017-0558-6.

[17] Langan LM, Dodd NJF, Owen SF, Purcell WM, Jackson SK, Jha AN. Direct measurements of oxygen gradients in spheroid culture system using electron paramagnetic resonance oximetry. PloS One 2016;11:1–14. https://doi.org/10.1371/journal.pone.0160795.

[18] Hari N, Patel P, Ross J, Hicks K, Vanholsbeeck F. Optical coherence tomography complements confocal microscopy for investigation of multicellular tumour spheroids. Sci Rep 2019;9:1–11. https://doi.org/10.1038/s41598-019-47000-2.

[19] Sasaki H, Takeuchi I, Okada M, Sawada R, Kanie K, Kiyota Y, et al. Label-free morphology-based prediction of multiple differentiation potentials of human mesenchymal stem cells for early evaluation of intact cells. PloS One 2014;9. https://doi.org/10.1371/journal.pone.0093952.

[20] Ishikawa K, Yoshida K, Kanie K, Omori K, Kato R. Morphology-based analysis of myoblasts for prediction of myotube formation. SLAS Discov 2019;24:47–56. https://doi.org/10.1177/2472555218793374.

[21] Yoshida K, Okada M, Nagasaka R, Sasaki H, Okada M, Kanie K, et al. Time-course colony tracking analysis for evaluating induced pluripotent stem cell culture processes. J Biosci Bioeng 2019;128:209–17. https://doi.org/10.1016/j.jbiosc.2019.01.011.

[22] Kato R, Matsumoto M, Sasaki H, Joto R, Okada M, Ikeda Y, et al. Parametric analysis of colony morphology of non-labelled live human pluripotent stem cells for cell quality control. Sci Rep 2016;6:1–12. https://doi.org/10.1038/srep34009.

[23] Shibuta M, Tamura M, Kanie K, Yanagisawa M, Matsui H, Satoh T, et al. Imaging cell picker: a morphology-based automated cell separation system on a photodegradable hydrogel culture platform. J Biosci Bioeng 2018;126:653–60. https://doi.org/10.1016/j.jbiosc.2018.05.004.

[24] Marklein RA, Lo Surdo JL, Bellayr IH, Godil SA, Puri RK, Bauer SR. High content imaging of early morphological signatures predicts long term mineralization capacity of human mesenchymal stem cells upon osteogenic induction. Stem Cell 2016;34:935–47. https://doi.org/10.1002/stem.2322.

[25] Oja S, Komulainen P, Penttilä A, Nystedt J, Korhonen M. Automated image analysis detects aging in clinical-grade mesenchymal stromal cell cultures. Stem Cell Res Ther 2018;9:1–13. https://doi.org/10.1186/s13287-017-0740-x.

[26] Maddah M, Shoukat-Mumtaz U, Nassirpour S, Loewke K. A system for automated, noninvasive, morphology-based evaluation of induced pluripotent stem cell cultures. J Lab Autom 2014;19:454–60. https://doi.org/10.1177/2211068214537258.

[27] Van Valen DA, Kudo T, Lane KM, Macklin DN, Quach NT, DeFelice MM, et al. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. PLoS Comput Biol 2016;12:1–25. https://doi.org/10.1371/journal.pcbi.1005177.

[28] Wan T, Xu S, Sang C, Jin Y, Qin Z. Accurate segmentation of overlapping cells in cervical cytology with deep convolutional neural networks. Neurocomputing 2019;365:157–70. https://doi.org/10.1016/j.neucom.2019.06.086.

[29] Kraus OZ, Ba JL, Frey BJ. Classifying and segmenting microscopy images with deep multiple instance learning. Bioinformatics 2016;32:i52–9. https://doi.org/10.1093/bioinformatics/btw252.

[30] Niioka H, Asatani S, Yoshimura A, Ohigashi H, Tagawa S, Miyake J. Classification of C2C12 cells at differentiation by convolutional neural network of deep learning using phase contrast images. Hum Cell 2018;31:87–93. https://doi.org/10.1007/s13577-017-0191-9.