PSCK-RIFS Operations

Figures S1 and S2 outline the application of the PSCK-RIFS software. A tissue cell preparation of interest is first counted (including viable cell determination; NC in Table S1) and used to initiate serial cultures, typically in triplicate. At an experimenter defined interval (SPLI in Table S1), cultures are counted – including viable cell count – and a constant fraction of the cells or number of the cells is transferred to a new culture dish of the same type (Figure S1). The PSCK software is designed to be versatile for accommodating different formats for the culture time period and/or the number of cells transferred. The total cell count data are transformed into plots of cumulative population doublings (CPD); and the viable cell fraction is used to estimate the dead cell fractions for transient cells (RDT in Table S1) and terminal cells (RDTM in Table S1), which are similarly very abundant during most of the culture period.

Figure S2 depicts the various inputs that are used by the PSCK software to calculate CPD curve simulations for the period defined by the experimental CPD data. Using the cellular model illustrated in Figure S2, the software inputs measured values ("m" superscripted in Figure S2) and discovered factors ("d" superscripted in Figure S2) obtained from the RIFS software. An important aspect of the PSCK-RIFS method is that there is no underlying mathematical algorithm. Instead, simulations are developed purely by a computational counting program. They begin with the number of total tissue cells dictated by the experimental data (i.e., NC). The starting cellular distribution with respect to stem cell fraction, transiently amplifying cell fraction, and terminally differentiated cell fraction can be varied among different idealized patterns representing disrupted tissue structures or cultured primary cell strains (e.g., exponential distribution, uniform distribution). Then the RIFS program feeds known and discovered input factors into PSCK simulations. The PSCK simulation program computes the total cell number produced during each culture interval according to the model and performs randomized splits at the end of each culture interval for the number of passages specified by the investigator. These data are transformed into simulated CPD data plots.

The quality of input factor sets discovered by PSCK-RIFS is assessed quantitatively. Each PSCK-RIFS analysis is performed for 1000 cycles. Each cycle includes RIFS selection of unknown factors, followed by generation of replicate PSCK simulations using the selected unknown factors, followed by comparison of the replicate PSCK CPD simulations to the source experimental replicate CPD data. We developed a quantitative quality score for these comparisons. The quality score is a 2-dimensional root mean square error (RMSE) measure of how close the combined magnitude and variance of the replicate CPD simulations about the experimental mean CPD data approximate the respective features of the experimental replicate CPD data about the experimental mean CPD data. The best quality score is 0.0. Ideally, quality scores ≤ 0.5 are preferred. However, scores ≤ 1.0 are acceptable.

For all studies presented, “n = 5”, refers to 5 independent, 1000-cycle PSCK-RIFS analyses to discover the sets of unknown input factors. For each discovered input factor, values from the 5 independent PSCK-RIFS analyses were averaged to estimate statistical significance and to make comparisons among different culture conditions. The replicate analyses reported have PSCK-RIFS quality scores < 0.5; and all presented analyses have quality scores < 1.0.

Two analyses performed routinely illustrate that the discovered cell kinetics factor sets are discrete. Figure S3A provides an example of a typical history for a 1000-cycle RIFS search for an optimal PSCK factor set. The y-axis gives the quality score as defined above for the ability of factor sets at the end of each search cycle to simulate a triplicate experimental CPD dataset using the PSCK stem cell-based growth simulation program. The x-axis indicates the progression of search iterations from 1 to 1000. The RIFS program keeps track of factor sets that give the best (lowest 2-dimensional RMSE score described above) quality scores and uses this information for subsequent searches (i.e., a machine learning basis). The bold black line indicates this retained and compared minimal quality score factor set. Figure S3B shows the frequency distribution of quality scores (y-axis, number of searches; x-axis quality score) from the search history in Figure S3A. The normal character of this distribution indicates a robust model structure; and it shows that the minimal quality score factor set discovered is a significantly discrete set of factors (arrow).