Cell Stem Cell
Forum

# Separating Stem Cells by Flow Cytometry: Reducing Variability for Solid Tissues

Caroline M. Alexander,[1,*] Joel Puchalski,[1] Kristine S. Klos,[1] Nisha Badders,[1] Laurie Ailles,[2] Carla F. Kim,[3] Peter Dirks,[4] and Matthew J. Smalley[5]
[1]McArdle Laboratory for Cancer Research, 1400 University Ave, Madison, WI 53706, USA
[2]Division of Stem Cell and Developmental Biology, Ontario Cancer Institute, MaRS Centre, Toronto Medical Discovery Tower, 8th floor 8-401, 101 College Street, Toronto, Ontario M5G 1L7, Canada
[3]Children's Hospital Stem Cell Program, Department of Genetics, Harvard Medical School, Harvard Stem Cell Institute, 300 Longwood Avenue, Karp Building, Room 8-216, Boston, MA 02115, USA
[4]Division of Neurosurgery/Program in Developmental and Stem Cell Biology, University of Toronto Hospital for Sick Children, Suite 1503, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada
[5]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK
*Correspondence: alexander@oncology.wisc.edu
DOI 10.1016/j.stem.2009.11.008

**Until there are valid identifiers that visualize stem cells in vivo, we rely upon flow cytometry to enrich for subpopulations with stem cell function. However, data reporting styles for flow cytometric analyses are typically inconsistent, creating challenges in comparing results across publications. In our view, clear reporting guidelines could improve reproducibility of stem cell analyses in solid tissues.**

Presently, the field of cancer stem cell biology (and somatic stem cell biology in general) is awash with reports of two-parameter dot plots that do not resemble one another, despite originating from similar tissues immunophenotyped with the same antibodies. Regardless of the lack of resemblance to original publications, authors will often attribute specific properties to cells that exhibit particular phenotypic traits by flow analysis without confirming their identity or function. Overall, this practice leads to the accumulation of unvalidated conclusions and misinformation on the behavior of stem cells or other cell types (such as differentiated epithelial or nonepithelial cell types).

We encourage the community to make use of the following proposed recommendations for the presentation of stem cell data obtained by flow cytometry. These criteria are not new: broadly stated, standards have been set out by MIBBI (Minimum Information for Biological and Biomedical Investigations), are outlined by Lee et al. (2008), as a consensus of opinion from cytometry professionals, and have even been implemented as the minimum accompanying information for flow cytometric results by some journals. Detailed criteria and techniques have been discussed by Roederer and Herzenberg, especially with respect to setting standards for hematopoietic cell analysis (Herzenberg et al., 2006; Moore and Roederer, 2009; Perfetto et al., 2006; Perfetto et al., 2004; Roederer, 2002a, 2008). Within the broader stem cell community, however, there has been little standardization applied to the separation of cell fractions from solid tissues to date. This oversight is unfortunate because the enzymatic dissociation procedures that are used for generating cell fractions make this analysis even more variable than the analysis of nonadherent cell types. It seems timely to highlight specific practices that may offer the stem cell field improved consistency in reporting across published accounts, given the wildfire adoption of cytometric procedures by laboratories not previously specialized in multichromatic analyses of cell populations.

We propose that a detailed list of experimental details and specific examples be included in submissions that utilize flow cytometric methods (Table 1). We offer our insight as to how the provision of such details will improve consistency across related reports and outline potential pitfalls that might be avoided by following this pattern of experimental reporting.

There are many reasons why presentations of flow data derived from the same tissue type may not look similar. Some of these variables are hard to control for and include differences between cytometers (even the same model) or areas of the world in which the experiments are conducted. These issues can only be truly solved by repeating all the functional characterization in each independent laboratory. Others sources of discrepancy might be attributed to variability between human tumors, to substantial differences between profiles from inbred mouse strains, or to different regimens for proteolytic dissociation of tissues. Specifically, the inclusion of the following information should enable valid cross-comparisons and ensure improved reproducibility, as described below.

## Antibody Binding Conditions Used to Label Cell Populations

Flow cytometry is a quantitative technique when antibody binding is saturating. Individual laboratories should take pains to test "new" antibodies for saturation binding (see Kantor and Roederer (1997) as a source of information on basic experimental guidelines). To describe a staining reaction in enough detail that it can be reproduced, the cell and antibody concentrations utilized should be specified (within the limits of the manufacturer's description; see Figures S1 and S3). Indeed, the antibody clone and specific fluorochrome used often influence the binding reactions in ways that are difficult to rationalize, making the provision of more detailed methodological information important.

## Make and Settings of the Flow Cytometer

For the ability to reproduce functional data from live cell sorts, machine factors

**Table 1. Experimental Details and Specific Examples for Inclusion in Submissions that Utilize Flow Cytometric Methods**

| To Report in Submitted Publications | Purpose | Specific Details to Include | Templated Example |
|---|---|---|---|
| Antibody binding conditions used to label cell populations | Establish consistency of antibody binding immunophenotypes | Antibody clone and fluorochrome used; antibody concentrations and time, temperatures used for labeling | Figure S1 |
| Make and settings of the flow cytometer | Clarify any discrepancy in reporting due to machine factors such as physical pressures exerted, laser calibration, wavelength, and filters used | Make/model of cytometer used; nozzle tip diameter, sheath pressure, and fluid composition; laser power and wavelength used; software package(s) used for sorting and analysis | Figures S1 and S3 |
| Compensation procedures | Achieve consistent elimination of artifacts associated with spectral overlap between fluorochromes | Report use of single stains or fluorescence-minus-one stains; indicate whether compensation was performed prior to separation and/or during post-sort analysis; calculated by operator or with a software-based algorithm | Figures S1, S3, and S5 |
| Display of manual gates applied | Improve equivalence of quantitation across independent experiments | Display the gating hierarchy used; indicate the number and percent of events excluded at each step in the "tree"; gates used to eliminate dead cells, cell doublets, debris, and irrelevant live cells | Figures S1–S5 |
| Display of the raw data | Establish transparency of number of events examined and reveal degree of separation between populations | Raw data plots demonstrating a sufficient event count to establish an adequate "n"; axis ticks that distinguish between log and linear scales should be visible; summary percentages should include a degree of error determined in independent replicates | Figures S1–S5 |
| Validation of results | Evaluation of extent of purification of isolated cells and verify their functional status and degree of enrichment | Reanalysis of fluorescent profile post-sort to establish degree of sort purity; examination of morphology and/or genetic traits to verify the identity and purity of sorted cells; functional assays comparing to similarly handled, unselected cells are needed to determine fold enrichment and degree of activity recovered | Figures S1, S5, and S6 |

that can affect sample recovery, viability, and function, such as nozzle tip diameter, sheath pressure and fluid composition should be reported (see examples in figures in the Supplemental Data). The laser power could be included, if this figure is known to be an important determinant of success. Note that these settings differentially affect various cell types; for example, for cells from the mammary gland, high pressure and low nozzle tip diameter can lead to fewer differentiated cells or basal cells. Specification of the make of the flow cytometer and the name of the software package(s) used during the sort and for any subsequent analysis provides most of the important machine-based parameters needed for background information. There are a great diversity of options for laser wavelengths and emission filters and an expanding repertoire of new fluorochromes (Chattopadhyay et al., 2008); thus, the wavelength of the laser (and possibly the emission filter, if that is not predictable) should be stated, given that these parameters can determine the relative efficiency of fluorochrome signals (see Figures S1 and S3).

## Compensation Procedures

For preserving the quantitative aspect of flow cytometry, fluorescent signals that bleed from one channel to another should be subtracted. For example, if an immunophenotyping reaction includes a bright fluorochrome with an emission spectrum that closely aligns with that of other fluorochromes used in the same sample, a correction factor will be necessary to prevent cell populations from being shifted inappropriately to different quadrants of the dot plot. In other words, cells that are labeled with one bright fluorochrome may read as false-positive expressors for a fluorochrome with a closely overlapping emission spectrum. The correction factors applied to prevent false-positive signals are termed compensation procedures (Roederer, 2002a), and a description of compensation procedures used (single stains for each dye, use of CompBeads [Becton Dickinson], or lot numbers for tandem dyes with variable spectral properties, and whether the correction is applied by the operator as either a machine-based correction, or calculated by software after flow cytometry [Herzenberg et al., 2006; Tung et al., 2004]) should be included in the methods section (see Figures S1 and S3 for examples, or http://www.drmr.com/ for more detailed technical advice). Overall, reproducible patterns and quantitation of polychromatic flow histograms require a consistent setup procedure. This paradigm has been elegantly described by Perfetto et al. and can be presented as a stepwise dissection including system optimization, calibration, and continuous monitoring of the fluorescent signals with respect to sensitivity, accuracy, and precision (Perfetto et al., 2006).

## Display of Manual Gates Applied

Accurate quantitation of cell subpopulations depends as much on accounting for the cells that are left out as the cells that are included on the final data plot. A description of the gating procedure should include a list and display of the sequential gates applied to exclude debris, to select single cells, to assay only live cells, and to exclude irrelevant cells. This pattern is summarized in a gating tree, also known as a population hierarchy (see Figures S1 and S5). For example, PI-positive cells (marking dead cells with permeable membranes) and debris can be "sticky," and thus nonspecifically bind antibody, and are likely to emit autofluorescent signals that contribute spurious, false-positive signals (Figure S2). Cell doublets will bind proportionally more antibody than single cells and can often appear in stem cell fractions, given that stem cell-enriched fractions are often defined as "high" expressors of various cell surface antigens, and so doublets should be gated out. The typical 2D data plot relies on the exclusion of cells not directly relevant to the analysis because they will often express ligands that bind the analytical antibodies as well. Their presence can cloud the view of rare target cells and decrease the purity of a target population. The most common gate applied to epithelial cell populations when attempting to enrich for a rare stem or progenitor pool is described as a Lineage$^+$ gate (Lin$^+$, or "dump channel," based historically on the application of a similar gate during hematopoietic separations) and includes a panel of antibodies labeled with the same flurochrome that bind endothelial and hematopoietic cells (such as CD31 and CD45). However, given that these irrelevant populations are not completely excluded by Lin$^+$ antibodies (either because cells are not homogeneously positive or because the antigens are clipped off cell surfaces during cell preparation), and their number can be high, marking the location of these cells on the final epithelial cell data plot can be important for the accuracy of subsequent interpretation (if for example, endothelial cells overlie a putative stem cell-enriched fraction; see Figure S2). Knowing where any spurious populations from nonsubject lineages lie on the final data plot may convince the reader that any changes in the fractions of interest are, indeed, specific.

In addition to indicating how unwanted events are eliminated, it is equally important to clarify how a positive signal is defined. Thus, for many antigens, an unselected population exists as a continuum of negative-, low-, and high-staining cells, rather than as a collection of obviously discrete populations. Subtraction of background binding can be based on (in order of rigor): (1) a truly negative population, for example genetically null cells (see Figure S4), (2) a nonexpressing population, known to be negative by prior understanding (see Figure S3), or (3) relevant singly stained fluorescently labeled isotype-matched antibody controls and/or fluorescence-minus-one strategies (Roederer, 2002b; Tung et al., 2004) (see Figure S3). Applying one or more of these strategies is particularly important for rare or specialized antigens, which are often visualized by adding a labeled secondary antibody that binds to the antigen-specific primary antibody and are particularly prone to high background binding (Figure S4). Furthermore, if it is clear that unstained cell populations exhibit significant autofluorescence that is detected in (all) analytical channels, the strategy used to exclude this contribution should be described. Many times, staining is described by the subjective terms "high" and "low." Instead, any cutoff used to distinguish relative levels of staining should be delimited quantitatively (for example, "CD49f$^{lo}$ cell fractions were gated as the lower 50% of the population," or "the median fluorescence of the CD49f $^{lo}$ was 10-fold less than that of the CD49f $^{hi}$ cells"; see Figure S3).

After having completed the gating hierarchy, both to negatively and positively select for the population of interest, it should be clear what proportion of cells is shown in the final analytical window. For some cell preparations, the total cell population represented on the final flow data plot may be a relatively minor fraction of the starting cell preparation. Thus, if one cell type is more susceptible to damage than others (especially with respect to the stem/differentiated cell fractions), this trait could lead to large discrepancies between laboratories (for example, if one laboratory displays data that represent 50% of the starting population, whereas another presents 5%). That is, without knowing which cells were lost during the course of the gating procedure, the final reported frequency of the population of interest is impossible to compare across laboratories.

Variations in outcome due to mechanical susceptibilities of isolated cells might be minimized by establishing a reproducible pattern of cell release from solid tissues, which is aided by providing detailed information on mechanical disaggregation, cell-dissociation media, and agitation patterns. Perhaps counterintuitively, tumor cells are often more fragile than their normal counterparts and easy to destroy during preparation. Necrotic tumors containing aneuploid cells may be particularly hard to handle. Devising a method for quality control for the enzymatic release procedure is useful. For example, enzymatic digestion can strip epithelial cell surface antigens, causing the epithelial cells to appear at high frequency in other nonepithelial populations. This effect can be detected by analyzing nonepithelial cell fractions for their expression of epithelial keratins (such as keratin-5 or -8).

An estimate of the total number of events retrieved from a flow cytometer, compared to the cell number added to the analysis tube, will often reveal serious discrepancies that relate to the extensive particles and debris that can be produced during live-tissue processing. The number of events is often assumed to equal the number of cells, but this relationship may not be accurate if/when epithelial or tumor cells are disrupted into numerous debris particles. The presence of scraps of extracellular matrix can also contribute to machine-detected "events" that are actually debris. Thus, in cases where significant debris is present, the percent recovery of live cells may be significantly higher than implied by the values reported in the gating tree.

## Display of the Raw Data

Flow cytometry follows the same rules of reporting as other scientific assays; it requires a sufficient n number (independent sample determinations) to show statistical significance. If the number of independent assays is sufficient, the absolute number of events required can be low (Roederer, 2008). If a single data plot is shown in a figure to illustrate the properties and frequency of an enriched population, these values should be

indicated ± standard deviations, as determine on the basis of multiple independent determinations. Note that axis ticks should always be visible and clear (particularly to distinguish between logarithmic and linear scales), and contour or other density plots are often more visually quantitative than dot plots. Finally, it should be made clear whether methods such as biExponential data transformation are used for visualizing compensated data points that fall below the axis on a logarithmic scale (Herzenberg et al., 2006; Roederer et al., 2004).

## Validation of Results: Evaluating the Purity of Subpopulations

Cell fractions of interest should be characterized after sort-mediated purification (using relevant assays, such as cytospin-immunohistochemistry [see Figure S1] or genetic tests), with the aim of confirming their identity. Note that claims of purity are also supported by reanalysis of sorted cell fractions, but these methods can harbor false negatives if antibodies or antigens and/or epitopes are lost during mechanical separation or because of photo-bleaching of particular fluorochromes (resulting in an underestimate of the sort efficiency). When various cell populations are tested to determine their functional activity, their relative viability should be specified (using a live-cell reporter such as trypan blue). Ideally, enriched fractions should be compared to stained, unseparated populations that have also passed through the sorting apparatus, to control for loss of function due to mechanical shear or other stress such as temperature or nutrient shock (Figure S6).

Among the purest stem cell populations reported to date are the hematopoietic stem cells ($\geq$1 in 3) purified by Morrison and colleagues (Kiel et al., 2005). Isolated fractions that harbor relatively more cells with stem cell activity should be labeled stem cell-enriched (SCE), rather than "stem cells" (see Figure S5), given that this label can be misleading for readers and is often incorrectly summarized in media sound bites. For example, the MRU mammary stem cell fraction, although it contains all the stem cell activity isolated from this tissue, may still be only 5% pure (Shackleton et al., 2006; Stingl et al., 2006; Figure S6).

Experiments that claim to enrich for stem cell activity must be validated via experimental means to show that functional enrichment has, indeed, been accomplished. Thus, the nonpurified cell population should be compared with purified cell fractions, recording the percent recovery of activity and fold-enrichment (per cell) to illustrate how much of the functional activity has been accounted for (Figure S6). It is important to perform functional analyses on all cell subsets, including those claimed to be stem cell deficient. The process of flow sorting itself can compromise stem cell activity by separating the test population from non-stem cell types usually required to support stem cell activity or because of mechanical damage, blocking functional epitopes with cell surface-binding antibodies, or antigenicity of fluorochromes (together, often responsible for 90% loss of activity) (Britt et al., 2009). Importantly, for specific claims about stemness to be made, there should be a functional evaluation of stem cell activity.

## Concluding Thoughts

To summarize, we encourage authors and reviewers to keep the following questions in mind when assessing whether submitted cytometric data is sufficient to support the claims made in a given study: Are the methods described in sufficient detail that the experiment can be reproduced, and do they include procedures for mechanical and enzymatic dissociation, antibody sources and binding reactions, make and settings for the flow cytometer, and any relevant software? For any major findings, are the gating procedures presented, and is the rationale for gate placement clearly defined? Are the relevant controls present that confirm specificity of staining? Is the overall percent cell recovery presented? Is the reproducibility of fractionation indicated? What is the basis for correlating a specific population with an activity or a phenotype? Is the percent purity presented for any given phenotype in a cell fraction? If functional activities are presented, what is the percent activity recovered in purified cell fractions compared to the starting population? Are the cell fractions given accurate names? That is, if a subpopulation is described as stem cell enriched, what is the estimated percentage of purity? What efforts are made to determine whether surrogate stem cell markers are accurate and truly specific to the stem cell-containing fraction?

It is typically expensive and time consuming to set up flow cytometric analyses of animal or human tissues, especially when searching for rare populations. We hope that widespread adoption of reporting guidelines, such as those outlined previously (Lee et al., 2008) and those we propose here for the analysis of cells from solid tissues, will enable the comparison of data generated across laboratories worldwide to yield more accurate conclusions and to reduce the frustration of new investigators in this area.

## SUPPLEMENTAL DATA

Supplemental Data include six figures and can be found with this article online at http://www.cell.com/cell-stem-cell/supplemental/S1934-5909(09)00581-5.

## ACKNOWLEDGMENTS

Many thanks to our expert reviewers, who made this a much more accessible, and hopefully useful, reference article, and Erik Ranheim (University of Wisconsin-Pathology Dept), Sean Morrison (University of Michigan Medical School), Heather LaMarca (Baylor College of Medicine), and Debi Lazzarino (New Jersey Medical School) for comments. Thanks for financial support from DOD Era of Hope Scholars Award (W81XWH-06-1-0491) and the UWCCC (CMA). M.J.S. is supported by Breakthrough Breast Cancer and acknowledges NHS funding to the NIHR Biomedical Research Centre.

## REFERENCES

Britt, K.L., Kendrick, H., Regan, J.L., Molyneux, G., Magnay, F.A., Ashworth, A., and Smalley, M.J. (2009). Breast Cancer Res. *11*, R20.

Chattopadhyay, P.K., Hogerkorp, C.M., and Roederer, M. (2008). Immunology *125*, 441–449.

Herzenberg, L.A., Tung, J., Moore, W.A., and Parks, D.R. (2006). Nat. Immunol. *7*, 681–685.

Kantor, A., and Roederer, M. (1997). In Handbook of Experimental Immunology, Fifth Edition, L.A. Herzenberg, D.M. Weir, L.A. Herzenberg, and C. Blackwell, eds. (Cambridge: Blackwell Science), pp. 49.1–49.13.

Kiel, M.J., Yilmaz, O.H., Iwashita, T., Yilmaz, O.H., Terhorst, C., and Morrison, S.J. (2005). Cell *121*, 1109–1121.

Lee, J.A., Spidlen, J., Boyce, K., Cai, J., Crosbie, N., Dalphin, M., Furlong, J., Gasparetto, M., Goldberg, M., Goralczyk, E.M., et al. (2008). Cytometry A *73*, 926–930.

Moore, J., and Roederer, M. (2009). Cytometry A *75*, 643–649.

Perfetto, S.P., Ambrozak, D., Nguyen, R., Chattopadhyay, P., and Roederer, M. (2006). Nat. Protocols *1*, 1522–1530.

Perfetto, S.P., Chattopadhyay, P.K., and Roederer, M. (2004). Nat. Rev. Immunol. *4*, 648–655.

Roederer, M. (2002a). Curr. Protoc. Cytom. *Chapter 1*, Unit 1.14.

Roederer, M. (2002b). Curr. Protoc. Immunol. *Chapter 5*, Unit 5.8.

Roederer, M. (2008). Cytometry A *73*, 384–385.

Roederer, M., Darzynkiewicz, Z., and Parks, D.R. (2004). Methods Cell Biol. *75*, 241–256.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Nature *439*, 84–88.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Nature *439*, 993–997.

Tung, J.W., Parks, D.R., Moore, W.A., and Herzenberg, L.A. (2004). Clin. Immunol. *110*, 277–283.