Novel Wheat-Germ Agglutinin Barcoding Approach for Diverse Mass Cytometry Samples

Authors: Brendan Cox and Dayton Barker
Advisor: Riley Hannan, PhD

Word Count: 3522
Number of Figures: 4
Number of Tables: 1
Number of Equations: 0
Number of Supplements: 0
Number of References: 12

Novel Wheat-Germ Agglutinin Barcoding Approach for Diverse Mass Cytometry Samples

Dayton Barker[a], Brendan Cox[a,1]

[a] Department of Biomedical Engineering at University of Virginia
[1] Correspondence: dhb8wkz@virginia.edu, +1 (443) 310-2884
[2] Correspondence: bpc3h@virginia.edu, +1 (757) 871-0701

Abstract

Barcoding is a powerful tool enabling the batch analysis of cell populations[1]. The abundant mass channels available in mass cytometry or Cytometry by Time of Flight (CyTOF) support per-sample barcoding, which allows numerous samples to be processed concurrently. The resulting increase in throughput and reduction in experimental variability have made barcoding a widely used tool in multiplexed cytometric studies[1]. Most live cell barcoding approaches, however, are restricted by species, cell lineage, or fixation state. In this Capstone project we demonstrate a live cell barcoding method that mitigates those restrictions, increasing the diversity of samples that can be concurrently measured in a CyTOF experiment. Thiolated wheat germ agglutinin (tWGA) has affinity for most membrane bound cells and is stable enough to be conjugated with metal isotopes. We demonstrate the utility of tWGA conjugates as a live cell barcode before fixation and across cell lineage. Our comparator is a percentage of CyTOF events falling within defined Mahalanobis and barcode separation distance ranges called percent good debarcoded (PGD). We show that this tWGA barcode approach has a PGD of 73% with a z-score of -0.87, which we claim to be comparable to 11 other barcoding methods tested with an average PGD of 81%.

Keywords: mass cytometry, CyTOF, barcoding, wheat-germ agglutinin

Introduction

Mass cytometry has become an indispensable tool in high parameter biomedical research. It allows researchers to detect metal-conjugated antibodies that are bound to antigens of interest on single cells. Its capability can be further expanded through a technique called barcoding, which labels cells originating in different samples such that they can be processed concurrently and later differentiated. (*Figure 1)*.

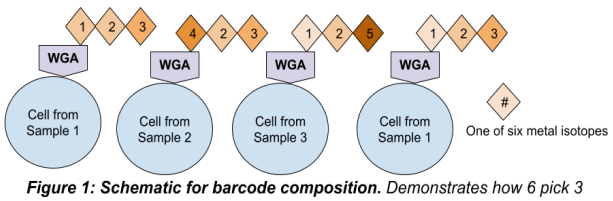

Figure 1: Schematic for barcode composition. *Demonstrates how 6 pick 3 combination of isotopes conjugated to WGA labels cells from different samples*

Traditional barcoding techniques, which can be traced back to the 1970s, use combinations of fluorophores to establish unique barcodes across a set of samples used in flow cytometry[2]. This fluorophore-based method, however, is limited by features inherent to fluorescent measurement including autofluorescence and spectral overlap. There is a more restrictive limit on the dimensionality of data and thus the complexity of a barcode in fluorescent cytometry. Mass cytometry is a much more recent technological advance, introduced in 2009, that ameliorates some of the limitations associated with fluorescence.

CyTOF increases the potential dimensionality of cytometry experiments by labeling samples with metal

isotopes instead of fluorescent markers. Abundant unique mass/charge ratios on metal isotopes mean researchers can multiplex experiments to a greater extent than within the bounds set by spectral overlap. Experiments with 40+ simultaneous parameters are relatively common using CyTOF, the utility of which contributed significantly to its rising popularity. As of 2019, the number of papers published using mass cytometry has more than quadrupled since 2015[3].

Despite its popularity, there are limitations to experimental complexity even within the realm of mass cytometry. Traditional mass barcoding involves the fixation and permeabilization of cells before barcodes, which often target a specific intracellular antigen, can be employed[4]. In experiments with fixation-sensitive epitopes, however, staining must occur before barcoding, which limits the optimization of such a workflow. Live cell barcoding was a solution to that predicament, saving time and material by allowing sample pooling to happen before fixation and permeabilization. The tWGA method proposed in this report is an example of live cell barcoding, which is typically characterized by attachment to the cell surface rather than an intracellular target. An existing and significant limitation, though, is that most antibodies used in live cell barcoding are restricted to a particular cell lineage[5]. Wheat germ agglutinin (WGA), however, is known to have cell surface affinity across species and cell lineage, making it a promising candidate for use in a universal live cell barcode[5,6]. WGA also has a strong history of use in cytometric research. It gained popularity in the 1980s as a reagent for neuronal tracing studies[7] and has since become a useful membrane-labeling tool across fluorescence microscopy and cytometry. As a result, it is widely available in both conjugated or free molecule form. It has also seen significant use in CyTOF specifically as a proxy for cell size and the characterization of bacteria. WGA's frequent use in research stems from its affinity for glycoproteins expressed in cell membranes[8]. To our knowledge, there is no reported membrane-bound cell that does not have surface affinity for WGA. Easily acquired, well-documented, and with the appearance of ubiquitous membrane affinity, WGA is a potential solution for existing limitations in live cell barcoding workflows.

Results

The proposed barcode, given as *WGA 20Plex,* falls within the range given by 11 other barcodes used by labs in both the Department of Biomedical Engineering and School of Medicine at the University of Virginia (*Figure 2)*. We compare the barcodes using a metric called percent good debarcoded (PGD) which is based on two statistical factors: barcode separation distance and Mahalanobis distance (*Figure 3)*. Barcode separation distance is defined as the scaled difference (0-1.0) between the least intense barcode mass and the most intense non-barcode mass within the file. A high barcode separation distance is desirable, as it shows a strong differentiation between barcoded and non-barcoded mass. For calculating PGD, we accept all measurements with barcode separation > 0.1. Mahalanobis distance measures the distance between a point and its population distribution and is essentially used to exclude outliers. We reject anything with a Mahalanobis distance >20 for calculating PGD. PGD can be understood to represent the proportion of debarcoded measurements or "events" that fall within desired constraints. A higher PGD indicates a more successful barcode in that measurements are sufficiently distinguishable from non-barcoded masses and that the

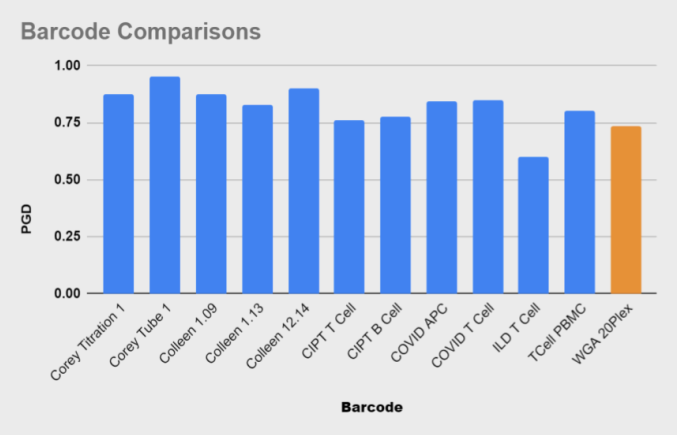

Figure 2. Barcode comparison using PGD metric. The WGA barcode, indicated in orange, competes with existing cell-specific barcodes as shown through percent good debarcoded (PGD).

population of measurements has acceptably low variability.

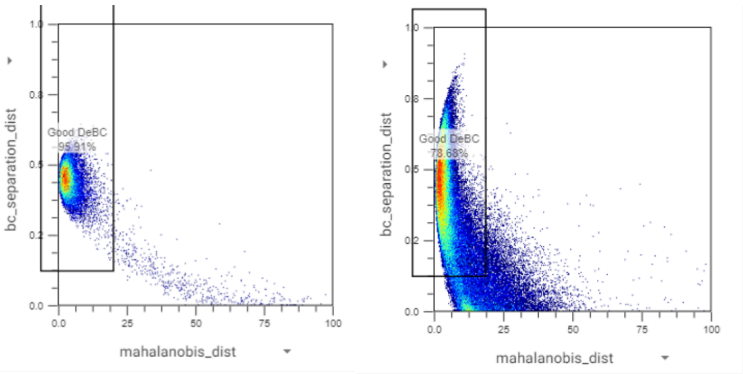

Figure 3. Visualization of PGD calculation. *PGD is* based on gating on Mahalanobis distance threshold (<20, x-axis) and barcode separation distance (>0.1, y-axis). The population of measurements within this gate relative to the quantity outside provides PGD.

The WGA barcode approach developed in this Capstone project had a value of approximately 73%, which was within the range of barcodes currently being used in several University of Virginia labs. With a z-score of -0.87, the WGA method is within one standard deviation of the mean PGD, which we interpret to be a competitive result.

Methodology and Materials

The processing of CyTOF data after barcoding was done using popular cytometric data analysis software OMIQ. Significant time was spent developing and testing the data preparation steps needed to compare the WGA barcode to data derived from other existing barcodes. These preparation steps can be divided into four main subprocesses: normalization, cleanup, debarcoding, and review.

Normalization

Data outputted by the CyTOF machine is loaded into a normalizer program. This program corrects the signal drop off that occurs over the course of a flow cytometry run[9]. Since metal ion buildup causes the cytometry sensor to lose sensitivity over time, a set of calibration beads that are easily pinpointed by the analysis program are spiked into the sample prior to measurement. As a part of the normalization process, the user draws gates over distributions of CyTOF data, sectioning off junk volume outside of the singlet masses. These gates are drawn several times, further refining the selection window until just beads and “good” cells are shown. The program then removes the calibration beads, so that the normalized file contains just the cells in question, corrected for variance within the CyTOF run.

Cleanup

The cleanup process includes a series of gating tasks that isolate singlet cells within a normalized file. The cleanup process begins with scaling the file, making sure that the original user view encompasses all of the data from the CyTOF run. After the file has been scale-checked, gates are created on the following parameters: center, offset, width, residual, bead distance, and live vs. dead cells. When gating, OMIQ is able to differentiate a variety of cell types and pulse shape metrics against time, plotting time on the horizontal axis of a chart and the type/metric on the vertical axis. As you go through the gating process, every step removes events outside the drawn gate, and another gate is drawn over what is left to further refine the selection region until all that is left is singlet cells.

The first four gates are done on gaussian parameters defining the CyTOF pulse, center, offset, width, and residual. Center refers to the mean of the pulse, offset is the distance from the base of the pulse to the zero line, width is the standard deviation, and residual represents how well the data fits the gaussian model[10]. A gaussian model is used to describe CyTOF data due to the bimodal nature of having a varying range of cell expression rather than definitive absences or presences of cells within a pulse[11]. The bead distance metric is used in order to separate doublet cells that may have shown up merged during the course of the flow cytometry run, as they will show up as higher mass single cells in analysis if not removed. Finally, a gate is drawn on live vs. dead cells, in order to fully remove every cell that is not a live singlet (*Figure 4)*.

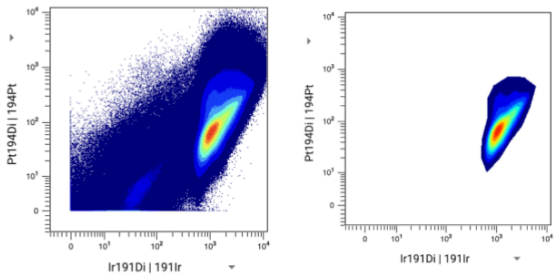

Figure 4. CyTOF data preparation. Before (left) and after (right) cleanup processing to select only singlet masses

Debarcoding

After the file has been cleaned, it is loaded into a debarcoding program to generate individual files for each sample. The debarcoding program, developed by Professor Eli Zunder of UVA Biomedical Engineering, takes the cleaned file as well as a debarcoding key as inputs. The program analyzes CyTOF events and identifies attached metal isotopes, looking for combinations derived from a 6 pick 3 arrangement. It separates them based on the sample number corresponding to the specific combination of metal isotopes. For a 6 pick 3 barcode, a maximum of 21 files will be outputted. This includes one for each of the 20 samples involved and a final file for all events that could not be assigned to a specific barcode. These files can then be loaded back into OMIQ for review.

Review

We consider several metrics in order to determine relative efficacy in comparing existing barcodes to our proposed WGA method. Throughout the cleanup and debarcoding process, counts of total debarcoded cells, split into what can be called “good” and “bad” debarcoded cells, and non-debarcoded cells are registered and displayed. Labeling a barcode as good requires cells to have just three of the six isotope tags, meaning that the debarcoding program will be able to place them into one of the 20 output individual sample files. A bad debarcode is a cell where more or less than 3 isotope tags are present (due to “mixing” as sample isotope tags bleed onto other samples), and these events are ones found outside the barcode separation vs. Mahalanobis distance acceptable range gate. Finally, non-debarcoded cells are those that were missing isotope tags, and so were not grouped by the program and are placed into the 21st “unassigned” file. At this point, sufficient data has been generated and cleaned to calculate PGD and make a comparison between barcoding methods.

Barcode Preparation

Using WGA as an attachment mechanism for barcoding was facilitated by an earlier lab discovery that enabled the use of commercially available thiol-maleimide conjugation kits. 2-iminothiolane (Traut’s Reagent), which has been previously used to conjugate WGA for applications in drug delivery, was used to stably thiolate WGA. This thiolated WGA, termed tWGA, would react at room temperature with polymer-bound metals, supporting the formation of a physical barcode compound. For the purposes of this project, a 20-plex 6 pick 3 palladium barcode is used, although tWGA is theoretically compatible with other heavy metal conjugations.

In the absence of commercial barcode kits, palladium barcode labels could be prepared based on a protocol for the chelation of isothiocyanobenzyl-EDTA and palladium ions[12]. In this case, the palladium heavy metal isotopes are purchased separately, with special attention paid to purity as the efficacy of the barcode will be dependent on pure palladium stock. In this procedure, isothiocyanobenzyl-EDTA serves as a polymer to which palladium ions are attached. To prepare the chelated compound, isothiocyanobenzyl-EDTA and a specific palladium isotope are mixed at a 2:1 molar ratio, snap frozen in liquid nitrogen, and lyophilized overnight before being mixed in a 6 pick 3 combinatorial arrangement and placed into storage[12].

tWGA and the combinatorial palladium chelations can then be mixed, resulting in combinations of three palladium isotopes bound to WGA molecules, as schematized in *Figure 1*. WGA coming from storage will be lyophilized, so the first step in adding the heavy metal barcode labels is reconstitution in an applicable buffer solution (Fluidigm Maxpar Antibody Labeling Kit). Once in solution, a milliliter of 1g/mL WGA solution is added to a 50kDa filtered centrifuge tube and

washed in three separate centrifuge runs using the suspension buffer. The metal isotope is then resuspended in buffer and added to the same tube as the washed WGA. This mixture is gently mixed and then incubated in a water bath at 37 °C for 90 minutes. A series of four more washes and centifugations are done to thoroughly mix the WGA and isotopes. Any unbound isotope or WGA is also washed out during this process. The final volume is tested using a nanodrop spectrometer to validate an acceptable concentration of conjugated WGA.

Barcode Application

Cell samples are taken from storage, thawed, and quickly placed on ice. Cells are spun down and resuspended to a pre-specified dilution volume. The barcode is then added from a prepared master mix as described in the previous section. The samples are thoroughly mixed and incubated for 30 minutes on ice. Each sample is spun down and resuspended in 200 uL of cold staining buffer twice before being combined into a new conical for batch CyTOF processing.

Additional live cell surface staining can occur immediately after application of the WGA barcode. For live cell surface staining it may also be advisable to do a viability stain before fixation. Intracellular staining can then occur after fixation and permeabilization.

Discussion

Future Implications of New Approach

The development of a universal live cell barcode will further increase the power of high throughput single cell analyses. tWGA's extensive compatibility across cell types may open the door for new experiments requiring parallel processing of diverse cell populations. An example of a potential study enabled by the WGA approach would be in investigating the immune response to tumor growth. Traditionally barcoding approaches, such as using a CD45 antibody conjugate to target immune cells, would not allow for a single experiment to capture the interplay between immune cells and potentially relevant tumor cells[5]. A barcode capable of labeling both cell types in one sample might enable researchers to discover mechanisms of communication between these different cell types, observe up or downregulation of particular proteins, or establish a causal relationship between observations while mitigating cross-experiment variability.

Hypothesis and Aims of Design Project

This main hypothesis addressed by this project was whether WGA could be used to barcode cells in a CyTOF workflow. Our proposal divided that question into two key aims: demonstrate efficacy for a WGA barcode and file a provisional patent application. Over the course of the year, we tested the barcode and implemented a rigorous data processing pipeline to show that WGA has an efficacy on par with several existing barcoding procedures. Successfully getting to the PGD metric and compiling data from across several research labs for our comparative analysis was in completion of the first aim. Submission of a provisional patent application went through on Wednesday, May 1, 2024 in completion of the second aim.

Improvements and Future Steps

WGA performed below average for our set of barcode PGDs, but improvements can still be made to optimize its efficacy. A deeper understanding of the mechanism through which WGA has affinity for cell surfaces may help us engineer stronger attachment. WGA's affinity for cell surfaces comes, at least partly, from selective binding to *N*-acetylglucosaminyl (GlcNAc) and *N*-acetylneuraminic acid (sialic acid) residues on glycoconjugates and oligosaccharides found throughout the cell membrane[6]. However, it is not clear whether this is a comprehensive explanation of WGA's affinity for cell surfaces. WGA may have affinity across a broader set of membrane glycoproteins. Additionally, it is likely that different cell types express varying amounts of glycoprotein and might then have varying efficacy under a WGA barcode. Variation in the size of cells may also contribute to varying barcode efficacy. An important future step is thus to continue testing the barcode against a variety of cell types to better characterize where it is effective, as well as improving the preparation and storage process to allow for a longer barcode shelf-life.

Design Constraints

Early design constraints were an important guide in the development of this new barcoding approach. Each design constraint was quantitatively defined for both acceptable and ideal outcomes. First, tWGA needed to efficiently attach to cells across lineage and without permeabilization. This led to a design constraint for attachment efficacy, which we observed as the percent of debarcoded events after CyTOF. Acceptable values fall within the 70-80% range, with ideal results greater than 80%. Attachment efficiency for each of the tested barcodes can be found in *Table 1* under percent debarcoded.

Barcode	Total Live Cells	Total DeBC	Total Live DeBC	Total Good DeBC	%Good DeBC	Total non-DeBC	Total Bad DeBC	%Bad DeBC
Corey Tube 1	155963	155717	155759	148208	95.15%	204	7551	4.85%
Colleen 1.09	7517591	7515381	7515381	6574742	87.48%	2210	940639	12.52%
Colleen 1.13	6411315	6408854	6408854	5302186	82.73%	2461	1106668	17.27%
Colleen 12.14	3266829	3265751	3265751	2943372	90.13%	1078	322379	9.87%
CIPT T Cell	3275642	3179346	3179346	2425302	76.28%	99296	754044	23.72%
CIPT B Cell	2942299	2823357	2823357	2191559	77.62%	118942	631798	22.38%
COVID APC	2498011	2471157	2471157	2085857	84.41%	26854	385300	15.59%
COVID T Cell	2349775	2341501	2341501	1991321	85.04%	8274	350180	14.96%
ILD T Cell	6395489	5577094	5577094	3355324	60.16%	818395	2221770	39.84%
TCell Panel PBMC WGA	77276	77121	77121	61803	80.14%	155	15318	19.86%
WGA 20Plex Mouse	*939381*	*800552*	*800552*	*586519*	*73.26%*	*10041*	*214033*	*26.74%*

Table 1: Barcode Efficacy Review; Data from OMIQ analysis of all samples within each barcode, with final comparison metric being %Good DeBC. Group's barcode is under "WGA 20Plex Mouse"

We developed two additional constraints to define the efficiency and quality of debarcoding: debarcoded events and debarcode quality. The metric for debarcoded events was %debarcoded, for which acceptable values fall within the 80-100% range. In the ideal case, %debarcoded would be >90%. Debarcode quality is given by PGD, which is acceptable above 70%, but ideally above 85%. Another design criteria, barcode longevity, refers to an issue our method is currently having with conjugated isotopes and WGA deteriorating in storage. Barcode efficacy drops after approximately two months in cold storage. Ideally, the longevity design constraint would entail stable conjugations for greater than 5 months. Developing a new storage approach is an ongoing process, but waste can be mitigated by making smaller batches to conserve reagents. The final design specification, cost, is derived from a broader discussion on how to make this procedure competitive with existing methods. The WGA barcoding method needs to be financially on par or cheaper than its counterparts if we expect widespread adoption of the technique. The goal of competitive pricing led to a metric of cost-per-run in the $150-200 range. Reagent costs are heavily dependent on which new isotopes are added to the panel. Commercially available conjugation kits are in the $200+ range and typically contain enough material for 4 reactions. Assuming a 20-plex barcode configuration, reagent cost-per-run comes to approximately $40. This value may be reduced if using reagents and barcodes prepared in-house. Additional costs factored into cost-per-run including CyTOF machine/core use fees, labor, and other miscellaneous expenses.

Assumptions and Limitations

A lack of funding was a significant limiting factor throughout the course of this project. Each isotope used in the barcode was ~$200 for six reactions worth of material, and we had run out by the end of the fall semester. Funding sources were explored to no avail, both within the BME department and School of Medicine at UVA. While a CyTOF run had been completed with the full barcode, for a number of reasons we would have liked to have been able to run additional experiments. As stated before, the shelf life of stored barcodes was below that stated in our design constraints. Without more isotopes to conjugate, our preparation and

storage methods could not be analyzed for changes to increase the longevity of the barcode.

Literature reports that WGA attaches to cell membranes without regard to cell lineage or type, but these claims were not tested in conjunction with our barcode, as we were only able to run it with one sample of mouse cells. We would have liked to experimentally demonstrate efficacy of our barcode across a variety of cell types and species, but were unable to due to monetary constraints.

Conclusion

Barcoding offers significant savings in both time and resources for labs running CyTOF experiments, but current methods remain limited by sample specificity. In this capstone project we demonstrate the efficacy of a live cell barcoding technique using wheat germ agglutinin as an attachment agent, which attaches to all membrane-bound cells regardless of type or lineage. In comparison with 11 other barcodes the WGA 20 Plex method had a 73% PGD value and the average across all methods was 81%. Our value falls within the acceptable range for a barcode method meaning that WGA can effectively be used to barcode live cells of all types.

With further funding the barcode would be refined, as barcode deterioration while in storage may have contributed to a lower PGD value. We would also like to test WGA 20 Plex with other cell types to validate the claim made in literature that WGA attaches ubiquitously across all cells. A provisional patent application has already been submitted for the method and if companies are interested talks will commence on licensing the method out to them, but it will remain free for other labs to use for their CyTOF experiments. Through dissemination of our barcoding method, labs will be able to optimize their workflows and make great time and money savings, while also being able to only work with one barcode as opposed to a different one for each experiment involving a new cell type.

End Matter

Conflict of Interest

The authors declare no conflict of interests.

Acknowledgements

We would like to thank the Sturek Lab for all its support both with funding and guidance throughout the past year, as well as Professor Allen and the rest of the capstone teaching team for their lectures, office hours, advice, and hours spent creating materials to aid us in all areas of the project process.

References

1. Extended live-cell barcoding approach for multiplexed mass cytometry | Scientific Reports. https://www.nature.com/articles/s41598-021-91816-w.

2. Picot, J., Guerin, C. L., Le Van Kim, C. & Boulanger, C. M. Flow cytometry: retrospective, fundamentals and recent instrumentation. *Cytotechnology* 64, 109–130 (2012).

3. Di Zeo-Sánchez, D. E., Sánchez-Núñez, P., Stephens, C. & Lucena, M. I. Characterizing Highly Cited Papers in Mass Cytometry through H-Classics. *Biology* 10, 104 (2021).

4. Gajera, C. R. *et al.* Mass-tag barcoding for multiplexed analysis of human synaptosomes and other anuclear events. *Cytom. Part J. Int. Soc. Anal. Cytol.* 99, 939–945 (2021).

5. Hartmann, F. J., Simonds, E. F. & Bendall, S. C. A Universal Live Cell Barcoding-Platform for Multiplexed Human Single Cell Analysis. *Sci. Rep.* 8, 10770 (2018).

6. Pellegrina, C. D. *et al.* Effects of wheat germ agglutinin on human gastrointestinal epithelium: Insights from an experimental model of immune/epithelial cell interaction. *Toxicol. Appl. Pharmacol.* 237, 146–153 (2009).

7. Brückner, G. & Biesold, D. Wheat germ agglutinin indicates the commencement of neuronal differentiation in the embryonic brain. A comparative carbohydrate histochemical study. *Acta Histochem.* 69, 231–242 (1981).

8. Wheat Germ Agglutinin - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/neuroscience/wheat-germ-agglutinin.

9. Finck, R. *et al.* Normalization of mass cytometry data with bead standards. *Cytom. Part J. Int. Soc. Anal. Cytol.* 83, 483–494 (2013).

10. Bagwell, C. B. *et al.* Automated Data Cleanup for Mass Cytometry. *Cytometry A* 97, 184–198 (2020).

11. Abe, K., Minoura, K., Maeda, Y., Nishikawa, H. & Shimamura, T. Model-based clustering for flow and mass cytometry data with clinical information. *BMC Bioinformatics* 21, 393 (2020).

12. Zunder, E. R. *et al.* Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. *Nat. Protoc.* 10, 316–333 (2015).