CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/115821/

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Smith, Paul J. 2016. Cytometric routes to single cell transcriptomics. Cytometry Part A 89 (5) , pp. 424-426. 10.1002/cyto.a.22869

Publishers page: http://dx.doi.org/10.1002/cyto.a.22869

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.

Commentary on: Samadder P, Weng N, Doetschman T, Heimark RL, Galbraith DW. Flow cytometry and single nucleus sorting for Cre-based analysis of changes in transcriptional states. Cytometry A 2016;89:430-42.

By
Paul J Smith*; Emeritus Professor of Cancer Biology, Cancer & Genetics Institute, School of Medicine, Cardiff University, UK, CF14 4XN,
smithpj2@cf.ac.uk

Title: Cytometric routes to single cell transcriptomics

A touchstone for cytometry and the article of Samadder et al. (in this issue, page XXX) is an appreciation of both the nature and consequences of heterogeneity while grappling with the stochastic attributes of biology. Although flow cytometry is not a single-cell analysis in its strictest sense, an enduring attraction is its ability to identify and indeed isolate the needles in the biological haystacks that are complex cellular systems. It offers the most accessible visualization of population heterogeneity that allows for a disassembly of complexity (1), while providing insights into the dynamics operating within cellular micro-communities (2). A recurring challenge is how to avoid the masking of biological complexity and subtlety when conventional transcriptomics employs bulk tissues or even mixtures of sub-populations often defined by specific protein expression. It is apparent that a revolution in cellular measurement technology is under way and cytometry is well-placed to participate in this ongoing revolution (3).

Cytometry in its various incarnations frequently requires well-designed "link" methodologies to enable its platform technology to access advances in a range of disciplines not least in oncology. Consider how multi-parameter flow cytometry has previously reached into histopathology through the processing of paraffin-embedded tissues from human tumours or how high-resolution flow cytometry imaging can be applied to critical cellular subsets such as circulating tumour cells. Further the creation of tractable model systems – most obvious being Genetically-Engineered Mouse (GEM) models – calls upon cytometry to deliver readily deployable link methodologies.

The paper by Smadder et al.,(in this issue, page XXX) does not seek to introduce a radical new approach, rather the authors describe and validate a pragmatic step forward in allowing cytometry to extract information from GEM models at the single nuclei level. The motivation is the challenge of detecting early indicators of different disease states within a dominating background and potentially behaviorally dynamic stromal background. The approach provides a key methodological link that permits the collection of nuclei from tissues for a new horizon in transcriptomics.

Specifically, the authors (page XXX) describe how they have undertaken to adapt fluorescence-activated nuclear sorting (FANS) for use in GEM models. The

importance of isolation of fluorescently tagged nuclei in cells in any linear methodology for single cell transcriptomics has been noted previously (4). The methodology borrows from experience in the flow cytometric methods for the analysis of nuclear-targeted GFP studies in plants and more recent advances in single cell genomic analysis including RNA-sequencing of single nuclei (5). The aspirations relate to the generation of driver cell lines as a resource and also wider options for the targeted localization of fluorescent proteins in well defined model systems to explore organ development and pathology.

The advantages and indeed the labor involved in the establishment and preparation of GEM models mammalian development are widely appreciated. EGFP was the first fluorescent protein expressed in transgenic mice (6,7) and remains in widespread use today (4). Together with its spectral variants and high performance monomeric red fluorescent proteins these genetically pliable fluorescent tags play beautifully into the flow cytometry platform for the identification, quantification and isolation of subpopulations of interest. The authors created transgenic mouse lines expressing chimeric histone 2B-GFP protein under the control of a constitutively-active, actin-derived promoter, separated by a Floxed-STOP sequence. Cre recombinase, within the F1 progeny, acts to excise the STOP sequence. The resulting transcriptional activation is readily identified in multiple tissues by GFP-positive nuclei conveniently prepared and sorted from various tissues. Critically, the importance of single cell resolution within the degrees of heterogeneity for sorted populations has in part been addressed by the previous finding that dissection of regulatory events at the single-cell level by the pooling of 10 nuclei can obscure the innate variability (5).

FANS purified nuclei could be used for nuclear proteomics and/or chromatin immuno-precipitation procedures (4). An intriguing possibility is to build upon the traditional use of flow cytometry in the field of stem cell research for the isolation of cells and enrichment of cell populations and for transcriptomics based lineage tracing without necessarily relying upon a biomarker identification of cells of interest. Here a combination of inducible recombinases, fluorescent reporter constructs, and live-cell imaging together with activator control over inducible Cre could generate clonal populations in situ for lineage tracing (8).

The limitations on transcriptomics are continually relaxing for subpopulation analysis. An earlier generic approach was to use an RNA tagging-isolation method such as thiouracil (TU) tagging in GEM systems. The GEM model expresses an enzyme that tags the nascent RNA strand with TU while context control is achieved by cell type-specific inducible promoter control of the transgene. This provides a sensitive method to isolate transcripts from cell type-specific subpopulations. A related approach is Translating Ribosome Affinity Purification (TRAP) in which EGFP-tagged ribosomes allow enable the isolation of mRNA undergoing translation but appears to run a greater risk of contamination (9). A more recent method is the microscopy based Transcriptome In Vivo Analysis (TIVA) whereby an activated photo-cleavable tag in a single cell anneals to mRNA that can be subsequently affinity purified or indeed combined with other isolation methods. (9)

However, in this post-microarray/next-generation sequencing era, RNAseq technology (10) is now making single cell sequencing achievable with a precision of measurement for the levels of transcripts and their isoforms that other methods cannot address. In RNAseq, sometimes referred to as whole transcriptome shotgun sequencing, RNA is converted to a library of cDNA fragments with adaptors, that can then be sequenced in a high-throughput manner to obtain short sequences from either one or both ends to produce genomic transcription maps. These maps reflect the transcript sequences and levels of gene expression revealing new dimensions of heterogeneity.

As always technology fulfils unmet need - in the development of automated instrumentation for the high-throughput isolation of single cells for further processing. For example, the development of a microfluidics platform for amplification of nucleic materials to support large-scale whole transcriptome studies to survey heterogeneity (11). The impact of the Smadder et al. study (page XXX) will be the provision of a readily accessible methodology for delivering nuclei of defined origins and temporal states to the precision tool of RNAseq. Using a less complex cell preparation approach, they have demonstrated the ability to profile DNA content distributions of isolated nuclei. Then to use minimal numbers of sorted nuclei as the sources of polyadenylated RNA for RT-PCR to determine how the nuclei qualify for the presence or absence of transcripts that are known to be diagnostic for specific pancreatic cell types. The approach is supported by a previous demonstration that RNA-seq can be performed using single sorted nuclei when prepared in this manner. The simplified cell preparation and isolation approach is adaptable to multiple models for which Cre driver lines have been established.

Biology occurs in context, in neighborhoods, in dynamic environments with extrinsic influences and under programmed destinies that are subverted in the case of oncogenesis. Not surprisingly heterogeneity and cross-talk in populations of cells has prompted the development of innovative tagging methods for *in situ* transcriptome profiling (9). An attraction of single cell or even type specific *in situ* genetic transcriptome profiling is the potential to explore microenvironment influences on defined cell lineages.

It is now common to predict the downstream data processing, analysis and cross-platform data integration challenges presented by emerging technologies (12). The caveat is that we are long way from understanding how dynamical systems can be used to describe cell fate transitions although they require a multidisciplinary approach for understanding how perturbations and have unexpected outcomes (3). However, in understanding the nature of heterogeneity, much can be borrowed from earlier conceptual models and indeed robust definitions. Here, classical cytometry has an important role in dissecting heterogeneity for model construction. For example, time-lapse live-cell microscopy can provide data about short-term fluctuations in promoter activity, while flow cytometry can reveal the longer term changes as cell states navigate their epigenetic landscapes. In 1957, C.H.

Waddington provided a visualization of the "epigenetic landscape" with the metaphor of a cell moving over a surface as a ball, attempting to find a path while encountering complex contours representing different states, revealing the plasticity of a cell to realize a sought state. The mathematical study of dynamical systems inspired later work, including that of Sui Huang and co-workers who were able to describe how cells can reach stable states even with minimal gene interactions but also achieve a weakly stable state that could be readily perturbed. Inevitably one appreciates how such weakly stable states can have downstream consequences and undesirable outcomes in cancer such as progression to therapeutic resistance or metastatic spread (13). Huang has further contributed to our joint vocabulary by reaching through the epigenetic noise to aid our description of non-genetic heterogeneity in dynamic cellular systems. For example, micro-heterogeneity is manifest in flow cytometry as the spread of single Gaussian-like curve in the typical log-scale presentation of data distribution. On the other hand, macro-heterogeneity due to the presence of a variety of discrete cell types or of cells in obviously distinct states is familiar us all as multi-modal distributions. Flow cytometry is familiar with the latter while advances in transcriptomics is a disassembling the nature and consequences of micro-heterogeneity - drawing us away from the comfortable position that apparently uniform cell populations consist of identical cells (1).

The Samadder et al. study (page XXX) again aligns flow cytometry with the ambitions of transcriptomics and is a practical contribution to the ongoing efforts to bring insights to the heterogeneity seen in normal tissues and those subpopulations under stress - particularly during oncogenesis. These efforts will no doubt reveal how selective pressures can define the ability of a tumour to exercise options for adaptation whether behavioral, epigenetic or driven by genomic instability.

LITERATURE CITED

1. Huang S. Non-genetic heterogeneity of cells in development: more than just noise. Development 2009;136:3853-62.
2. Smith PJ, Falconer RA, Errington RJ. Micro-community cytometry: sensing changes in cell health and glycoconjugate expression by imaging and flow cytometry. J Microsc 2013;251:113-22.
3. Trapnell C. Defining cell types and states with single-cell genomics. Genome Res 2015;25:1491-8.
4. Miller RL. Transgenic mice: beyond the knockout. Am J Physiol Renal Physiol 2011;300:F291-300.
5. Grindberg RV, Yee-Greenbaum JL, McConnell MJ, Novotny M, O'Shaughnessy AL, Lambert GM, Arauzo-Bravo MJ, Lee J, Fishman M, Robbins GE and others. RNA-sequencing from single nuclei. Proc Natl Acad Sci U S A 2013;110:19802-7.
6. Mao X, Fujiwara Y, Chapdelaine A, Yang H, Orkin SH. Activation of EGFP expression by Cre-mediated excision in a new ROSA26 reporter mouse strain. Blood 2001;97:324-6.

7. Okabe M, Ikawa M, Kominami K, Nakanishi T, Nishimune Y. 'Green mice' as a source of ubiquitous green cells. FEBS Lett 1997;407:313-9.
8. Kretzschmar K, Watt FM. Lineage tracing. Cell 2012;148:33-45.
9. Guldner IH, Zhang S. A journey to uncharted territory: new technical frontiers in studying tumor-stromal cell interactions. Integr Biol (Camb) 2015;7:153-61.
10. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 2009;10:57-63.
11. Streets AM, Zhang X, Cao C, Pang Y, Wu X, Xiong L, Yang L, Fu Y, Zhao L, Tang F and others. Microfluidic single-cell whole-transcriptome sequencing. Proc Natl Acad Sci U S A 2014;111:7048-53.
12. Khan IA, Lupi M, Campbell L, Chappell SC, Brown MR, Wiltshire M, Smith PJ, Ubezio P, Errington RJ. Interoperability of time series cytometric data: a cross platform approach for modeling tumor heterogeneity. Cytometry A 2011;79:214-26.
13. Huang S. Genetic and non-genetic instability in tumor progression: link between the fitness landscape and the epigenetic landscape of cancer cells. Cancer Metastasis Rev 2013;32:423-48.