Cataloguing individual proteins from a single cell

Jagannath Swaminathan, Cofounder of Erisyon Inc. – jagannath@erisyon.com

- **Erisyon introduces fluorosequencing, a single molecule protein sequencing technology.**
- **Identifying rare cells and mapping tissue heterogeneity is useful for biomedical research and in advancing precision medicine.**
- **Unlike genomics and transcriptomics, capturing information of proteins from single cells has been a major technical challenge. This is due to limited analytical sensitivity of current technologies and quantitative protein losses occurring during routine sample preparation process.**
- **We propose to use the single molecule detection sensitivity of fluorosequencing in combination with a novel solid phase peptide capture method to identify 1000s of proteins from a single cell.**

The single cell analysis market is valued at $1.6B in 2016 and projected to $5.2B in 2018[1]. While identifying rare cells and monitoring differentiation stages of tissues and organs by sequencing the genomes and/or the transcriptomes of single cells is routinely performed and drives the market, protein level information from these individual cells lag behind[1]. Capturing the information on the localization, abundances and modification states of proteins from single cells is important as proteins are the primary determinants of the state and functions of a cell. The approximately 50-500 pg of protein from a common mammalian cell's proteome (~20,000 types of genes are known to translate to proteins) is extremely complex with (a) large dynamic range of protein abundances (estimated from 10^2 – 10^8 copies/cell[2]), (b) a diverse heterogeneity of its isoforms and modification states (estimated to be 100,000 proteo-forms) and (c) a large range of turnover rates of proteins (estimated 10-1000h in primary cell[3]). The current proteomic technologies such as mass-spectrometry or antibody-based methods are underpowered to handle these requirements for detailed single cell protein analysis. Fluorosequencing, a **parallelized single molecule peptide sequencing technology**[4,5], could be a platform technology for single cell proteomics, owing to its intrinsic features of single molecule sensitive detection and absolute protein quantification[6].

Current technical challenges in single cell proteomics

While RNA sequencing methods have identified transcriptomic activities of single cells[7], it is a poor proxy for a cellular proteome[8]. Additionally, the multitude of post-translational modifications – its sites and abundances- are invisible from the transcriptomic data. Thus a number of attempts have been made to detect and measure proteins in a single cell using antibodies[9] or mass spectrometry[10] or a combination of the two, as in CyTOF[11]. However, they lack scalability and sensitivity, thus limiting the number of proteins identified to much fewer than a hundred and in a biased manner.

Apart from detection sensitivity, the other major problem in single cell proteomics is the invariable protein losses during the demanding sample handling steps. Efforts to streamline and automate the separation process for peptide mass spectrometry has enabled increased identification to a few hundred proteins, a far cry from the estimated 15,000 proteins present in the cells[10]. Large scale identification of proteins from single cells derived from a limited amount of tissues or tumor biopsy has not been feasible till date.

Case study for red blood cell differentiation

It is well known that mature red blood cells are non-nucleated, devoid of ribosomes and have a reduced proteome (estimated to approximately 1000 proteins) and dominated by hemoglobin. Understanding the dynamic nature of RBC's biology and its differentiation pathway from progenitor reticulocytes play an important role in hematology and transfusion medicine. However, these states and changes cannot be measured by the transcriptomic analyses, leaving single cell proteomics as the only direct method for the single cell analysis. Additionally, single cell proteomics obviates a technical problem faced in estimating the proteins and protein-protein interactions from bulk-cell measurements. With 90% of the proteome in a RBC made up of Hemoglobin, a 10% contamination of other cells in the sample would produce an incorrect interpretation. Understanding low abundant phosphorylation events that drives cytoskeleton and metabolic changes indicates a highly sensitive and a quantitative method for single cell proteomics.

Identification of proteins from single cells by Fluorosequencing

The features of single molecule sensitivity, digital quantification and the ability to identify sites of modifications, such as phosphorylation makes fluorosequencing the ideal method for single cell proteomic analysis. Techniques of limited dilution will be used to isolate single cells in well plates. Acoustic-sonication will be used to lyse the cell and release its contents to the surrounding buffer. In combination with endo-protease such as gluC, beads coated with Pyridinecarboxaldehyde (PCA) would be introduced at a neutral pH buffer. Erisyon is developing the PCA mediated

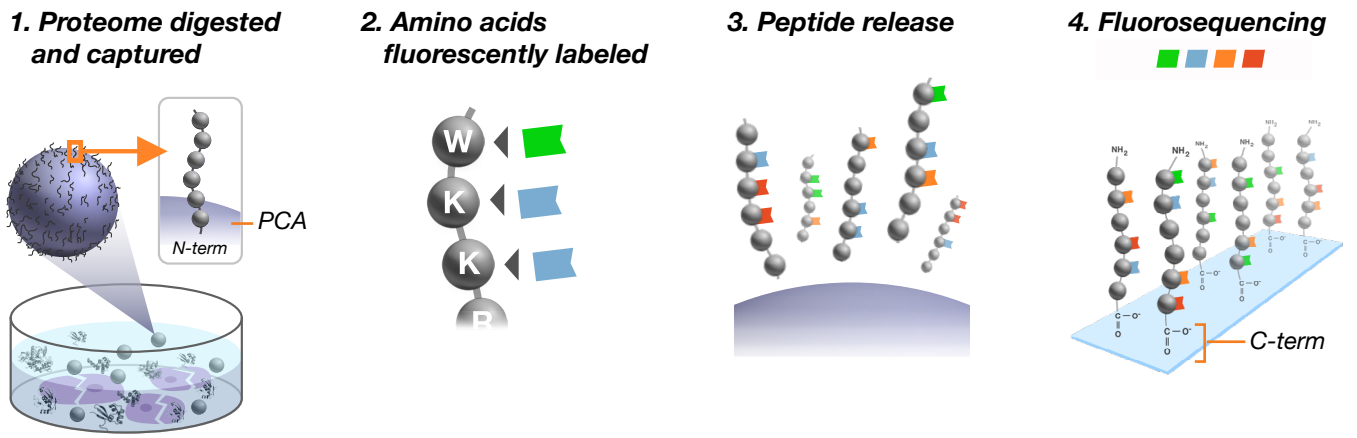

Fig 1. Proposed route to catalogue all the proteins from a single cell. *– Cells are diluted and isolated into micro-wells, ensuring at most a single cell/well. (1) Each distinct cell is broken by ultrasonication and proteins digested by stable endo-proteases such as gluC. The digested peptides are captured covalently via its N-terminal on solid phase resin, coated with the PCA reagent. (2) The functional amino acid side chains (lysines, cysteines etc) are labeled with fluorophore on the solid support with washing out the excess fluorophores, salts or other reagents and (3) released from the bead using a hydrazine mediated cleavage reaction. (4) The labeled peptides are immobilized via it's C-termini on a glass slide and sequenced to identify all the source proteins present in a single cell.*

peptide capture solution for the upstream sample preparation procedure and described in a recent publication[13]. Briefly the peptide's amino terminus reacts with the PCA reagent forming a covalent five-membered ring structure. The side chains of lysines are unaffected by the reagent[12]. The covalent nature of the binding enables stringent washing off the salts, detergents and other chemicals present in the sample. Additionally, the innovation makes it possible to automate the labeling and clean-up procedures for fluorosequencing application. A scar-less release of the peptides from the bead is a major inventive step for the upstream sample handling procedure. The figure describes the proposed procedure for single cell proteomics by fluorosequencing. Simulations performed under the current state of fluorosequencing technology, indicates that about 4000 cytosolic proteins can be identified from a single cell. This assay platform can also be used to fingerprint (the pattern of fluorosequences) the extracellular proteins in order to study cell differentiation, cell-typing etc.

Conclusions

While single cell genomics and transcriptomics have matured and is applied for identifying rare cells to learning organ differentiation, single cell proteomics hasn't been so advanced. Erisyon's technology of single molecule protein sequencing and bead capture method potentially overcomes the limitation of detecting and working with low abundant protein molecules from a single cell.

References

1. Strzelecka, P. M., Ranzoni, A. M. & Cvejic, A. Dissecting human disease with single-cell omics: application in model systems and in the clinic. *Dis. Model. Mech.* 11, dmm036525 (2018).
2. Zubarev, R. A. The challenge of the proteome dynamic range and its implications for in-depth proteomics. *PROTEOMICS* 13, 723–726 (2013).
3. Mathieson, T. et al. Systematic analysis of protein turnover in primary cells. *Nat. Commun.* 9, 689 (2018).
4. Swaminathan, J., Boulgakov, A. A. & Marcotte, E. M. A theoretical justification for single molecule peptide sequencing. *PLoS Comput. Biol.* 11, e1004080 (2015).
5. Swaminathan, J. et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* (2019).
6. Yates, J. R. Innovations in Proteomics: The Drive to Single Cells. *J. Proteome Res.* 17, 2563–2564 (2018).
7. Tang, F., Lao, K. & Surani, M. A. Development and applications of single-cell transcriptome analysis. *Nat. Methods* 8, S6-11 (2011).
8. Liu, Y., Beyer, A. & Aebersold, R. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell* 165, 535–550 (2016).
9. Hughes, A. J. et al. Single-cell western blotting. *Nat. Methods* 11, 749–755 (2014).
10. Specht, H. & Slavov, N. Transformative Opportunities for Single-Cell Proteomics. *J. Proteome Res.* 17, 2565–2571 (2018).
11. Bendall, S. C. et al. Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science* 332, 687–696 (2011).
12. MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* 11, 326–331 (2015).
13. Howard, Cecil J., et al. "Solid-Phase Peptide Capture and Release for Bulk and Single-Molecule Proteomics." bioRxiv (2020).