Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Genetics & Development

Advances in stem cell proteomics

Nade Abazova[1,2,3] and Jeroen Krijgsveld[1,2]

Stem cells are at the basis of organismal development, characterized by their potential to differentiate towards specific lineages upon receiving proper signals. To understand the molecular principles underlying gain and loss of pluripotency, proteomics plays an increasingly important role owing to technical developments in mass spectrometry and implementation of innovative biochemical approaches. Here we review how quantitative proteomics has been used to investigate protein expression, localization, interaction and modification in stem cells both *in vitro* and *in vivo*, thereby complementing other omics approaches to study fundamental properties of stem cell plasticity.

Addresses
[1] German Cancer Research Center (DKFZ), Heidelberg, Germany
[2] Excellence Cluster CellNetworks, Heidelberg University, Heidelberg, Germany
[3] European Molecular Biology Laboratory (EMBL), Heidelberg, Germany

Corresponding author: Krijgsveld, Jeroen (j.krijgsveld@dkfz.de)

Current Opinion in Genetics & Development 2017, **46**:149–155

This review comes from a themed issue on **Cell reprogramming**

Edited by **Jianlong Wang** and **Miguel Esteban**

http://dx.doi.org/10.1016/j.gde.2017.07.007

Introduction

Stem cell biology is currently one of the most dynamic fields in life science. This is explained by the fact that stem cells are at the basis of organismal development, where intricate mechanisms underlying pluripotency maintenance and cellular specialization are core components of biological complexity emergence [1]. Additionally, the high developmental potential of stem cells and the manifold properties they share with cancer cells hold great promise for a wide range of clinical applications, including drug screens, regenerative medicine and cancer research [2,3].

The vast complexity of stem cell biology demands multi-scale, multi-disciplinary research efforts, including genomic, epigenetic and functional studies. Proteomics is making an increasing impact in the field, focusing on proteins as the main functional entities driving cellular processes. This has benefited from technical advances in mass spectrometry for protein identification and quantification (Box 1) as well as from innovative biochemical approaches to target specific sub-proteomes. Combined with tailor-made bioinformatic tools, this now provides an integrated and accessible set of technologies to characterize cellular proteomes at unprecedented depth and detail. Here we review recent advances in stem cell proteomics, highlighting how quantitative studies of protein expression, localization, interaction and modification have contributed to increase our understanding of stem cell identity, cell fate transitions, and chromatin regulation (Figure 1).

Proteomic profiling of cellular plasticity

Proteomics of stem cells *in vitro*

Mass-spectrometry-based proteomics has been applied to a wide range of developmental processes including spermatogenesis [4], lineage specification [5] and neural differentiation [6], among many others in a body of literature that exceeds the scope of this review. Clearly the number and diversity of proteomic applications in the stem cell field are the result of continuous progress in introducing various *in vitro* cell culture systems that recapitulate specific developmental processes that are not easily accessible *in vivo*. A prominent example is the introduction of iPS cells to study fundamental aspects of gain and loss of pluripotency, in which proteomics is taking an increasing share [7]. For instance, the groups of Nagy and Heck identified two waves in proteome resetting in the early and late phase of reprogramming of secondary mouse embryonic fibroblasts (MEFs) to iPS cells [8], mirroring findings from our lab in a similar system [9]. The proteome data covered extensive cellular programs such as energy metabolism, chromatin regulation and cell cycle, reflecting a profound but well-orchestrated transition in cell identity. Other studies have focused on characterizing the molecular basis of pluripotency by comparing proteomes of mouse embryonic stem cells (mESCs) in ground state and primed state (epiblast stem cells, EpiSC) of pluripotency, identifying global differences in glycolysis [10] and in many proteins involved in chromatin regulation [10,11], reflecting the notion that pluripotency is largely regulated epigenetically [12].

Proteomics of stem cells *in vivo*

Although *in vitro* systems are very powerful to create (generally) homogeneous populations under well-controlled conditions, the ultimate goal is to understand cellular plasticity in the complex environment of the stem cell niche *in vivo*. Challenged by the low number of cells that can typically be obtained from animal models,

Box 1 Mass spectrometric technologies for protein identification and quantification

Protein identification

- **MSMS**: Tandem mass-spectrometry, indicating the 2-step process in the mass spectrometer by which a peptide is first detected to determine its mass, and then fragmented to generate a fragmentation pattern leading to its identification.
- **LC–MSMS**: Liquid chromatography coupled to tandem mass-spectrometry. Chromatographic separation of complex peptide mixtures leads to a larger number peptides that can be identified, and thus to greater proteome sampling depth.
- **Shotgun proteomics**: Experimental strategy for global proteome profiling aiming to identify all proteins in a sample, typically employing LC–MSMS. Also referred to as 'discovery proteomics'.
- **Targeted proteomics**: Experimental strategy, typically employing LC–MSMS, aimed to detect and quantify a pre-defined set of peptides in a sample, or usually across a series of samples.
- **Phospho-proteomics**: Strategy aimed to globally characterize proteins carrying a phosphate as a post-translational modification. Critically involves enrichment of phospho-peptides to compensate for their low abundance.

Protein quantification

- **SILAC**: Stable isotope labeling with amino acids in cell culture. Proteins that are differentially labeled with heavy and light stable isotopes can be distinguished by mass. Their mass spectrometric signal can be relatively quantified to generate protein ratios between samples.
- **Di-methyl labeling**: The principle is as for SILAC, however the stable isotope is introduced chemically after protein isolation rather than during cell culture. Thus, this method can be applied to any sample.
- **iTRAQ**: Isobaric tag for relative and absolute quantitation. Chemical stable isotope labeling strategy that allows up to 8-plex sample multiplexing.
- **TMT**: Tandem mass tag. Isobaric labeling strategy that is conceptually similar to iTRAQ, allowing multiplexing of samples up to 10-plex.
- **Label-free quantification**: Protein quantification not using stable isotopes, based on the use of mass spectrometric features such as peak intensity, volume or peptide counts.

increased sensitivity of current mass spectrometers have allowed the combination of FACS sorting with proteomic analysis to study the underpinnings of stem cell properties *in vivo*. For instance, a combined transcriptome and proteome analysis of quiescent and cycling intestinal crypt cells identified SMOC2 as a signature protein of intestinal stem cells [13]. Focusing on the hematopoietic system, we have shown that multipotent stem/progenitor cells are distinguished from myeloid precursors by the differential expression of hundreds of proteins in metabolism and cytoskeletal organization, but also by a select group of proteins involved in interferon signaling specifically expressed in the stem cell population [14]. Furthermore, refined sorting of hematopoietic stem cell (HSC) and multipotent progenitor (MPP) populations, which are both multipotent but differ in their self-renewal potential, allowed the identification of several differentially expressed proteins involved in transcriptional and cell cycle regulation [15]. Beyond the analysis of cells in non-perturbed conditions, we investigated proteomic and transcriptomic changes during HSC activation in a mouse model inducing acute inflammatory response [16•]. This revealed the emergence of a set of platelet-specific proteins shown to represent stem-like megakaryocyte-committed progenitors (SL-MkPs) in the HSC, but not the MPP compartment. Strikingly, since this was only observed at the proteome, but not the transcriptome level, this indicates a post-transcriptional mechanism elicited to replenish the platelet-pool depleted during inflammation [16•].

Cell surface proteins

Proteins specifically expressed at the outer surface of cells enjoy a long-standing interest since they can serve as markers to visualize specific cell types, or to purify them from a heterogeneous population [17]. Although classically this has been approached by the biochemical purification of plasma membrane proteins [18,19], more recent applications utilize chemical approaches to biotinylate lysines [17] or glycan-moieties [20] in cell-surface exposed proteins. These methods have proven very powerful to identify lineage-specific markers in embryonic (ES), epiblast (EpiSC), trophoblast (TS), and extraembryonic endoderm (XEN) stem cells [21], to monitor changes in cell-surface proteins during neural differentiation [20], to identify selectable markers for iPS cells [22] and iPSC-derived hepatocyte-like cells [23], and to immunophenotype cells in large panels by positive or negative selection [24]. Interestingly, benefiting from deep proteome coverage in global protein expression data, population-specific surface markers have been identified even without their specific enrichment [25].

Protein interactome

Cellular plasticity is not only regulated by changing protein expression, but also by modulating protein–protein, protein–DNA and protein–RNA interactions. In particular, transcriptional regulation during gain or loss of pluripotency is typically dependent on highly coordinated processes involving formation of protein complexes on chromatin, while post-transcriptional regulation relies on dynamic RNA–protein interactions. Proteomics has been crucial to characterize these interaction networks (or interactomes) in an unbiased manner, thereby defining pluripotency-regulating circuits that involve dozens or hundreds of proteins [26,27].

Protein–protein interactions

In the context of stem cell biology, the majority of protein–protein interaction studies in ESCs were centered around the core pluripotency transcription factors, OCT4 [28–30], SOX2 [31,32] and NANOG [33,34]. In general, affinity purification was used to pull down the protein of interest along with its interaction partners, which were then identified and quantified via mass spectrometry. Using this method, several novel regulators of pluripotency have been recently identified, such