Bioinformatics in stem cell characterization

Syed M Shah[1], N Saini[2], S Ashraf[3], M S Chauhan[4]

[1-4]*Animal Biotechnology Center National Dairy Research Institute, Karnal*-132001, *India*
syedakhyderi14@gmail.com

Abstract

Speaking in realistic sense a stem cell is endowed with potentials only, which are options rather than specific gene-expression patterns, thereby making its identification vague. Their characterization thus necessitates an overall genomic and proteomic analysis, coupled with mathematical modeling providing for the need for a wide range of bioinformatics tools to efficiently and effectively process large amounts of data being generated. A large number of computational tools and bioinformatic methods are thus required to deal with the experimental riches of this complex and multivariate data and the subsequent transition from data collection to information or knowledge in order to arrive at a universal definition of stemness.

Keywords-Bioinformatics, stem cell characterization,stemness

1. Introduction

A cell is regarded as a stem cell if it fulfills essential characteristics of self renewal, capability for multilineage differentiation and *in vivo* functional reconstitution of a given tissue. Cells derived from many different sources fulfill these criteria and are thus, considered stem cells. Based on their source of origin, these could be adult stem cells, embryonic stem cells, embryonic germ cells or placental stem cells. These cells have different potency and therefore accordingly classified as totipotent, pleuripotent, multipotent or unipotent. The development of organs during embryogenesis depends on these cells and, in the adult, frequent cell loss is compensated for by their activity. Stem cells are therefore indispensable for the integrity of complex and long-lived organisms. The exact definition of stem cells and the ability to isolate them are, therefore, the matters of supreme importance. However, despite the efforts of many investigators who strive to determine their nature, a definitive stem-cell 'portrait' is lacking (Zipori, 2004).

1.1 The stem-cell definition enigma

Stem cells are usually defined in retrospect by functional assays that test cellular properties. These assays are long term and elaborate — for example, in the hematopoietic system, the prototype model to define the general biological properties of mammalian SCs, several markers are used to define a population, which is then isolated and injected into animals that are examined several months later. The end point indicates that, within the injected population, only some cells show stem-cell properties while many of the cells within the defined population do not conform to a stem-cell functional phenotype, despite the fact that they share the same markers used for the purification process. Therefore, defining stem cells according to the properties of the markers they express excludes other cells that possess obvious stem-cell functional phenotypes and might be as important or as abundant. The solution to determining the molecular configurations that dictate a stem-cell state should, therefore, come from an overall genomic and proteomic analysis, coupled with mathematical modeling.

1.2 Characterization of stem cells

The most common strategies employed for characterization of stem cells include morphology, transcription and surface markers (receptors), karyotyping, pluripotency and differentiation. Despite of all these strategies various possible confounding variables involved in stem cell characterization include source of the stem cells, whether embryonic, fetal or adult, systematic, whether embryonic, liver, heart or brain, organismal, whether human, mouse dog, chimp or farm animals, epigenetic and genetic background and cultural, depending on substrate, media, growth factors and *in vivo* or *in vitro* conditions. The generalizations are further hampered due to differences in stem cell behavior and morphology and differences between highly evolved cell lines and their wild type counterparts (Robert, 2004). This necessitates the need for wide range of bioinformatics tools to efficiently and effectively process large amounts of data being generated. A large number of computational tools have been developed to deal with the experimental riches of this complex and multivariate data and transition from data collection to information or knowledge (Kapetanovic *et al*, 2004), in

order to arrive at a universal definition of stemness. These include the bioinformatic methods like clustering, gene shaving, artificial neural networks, boosting, bagging, fuzzy logic and so on.

1.3 Bioinformatic tools for stem cell receptor assessment

It is suggested that the state of large sets of molecules such as proteins, nucleic acids and lipids, as well as the position of these molecules within the cells, like the specific patterns of chromatin organization form a stem-cell signature. Many of these traits might be modified quantitatively or via post translation. Therefore, patterns of gene expression should be combined with analysis of molecules, their modifications and intracellular localizations for the right identification of stem cells. The stem-cell signature should therefore be determined by systems-biology tools that can identify patterns, rather than by the analysis of individual genes or even multiple gene-product behaviours. The resulting transcriptome and proteome should then be compared, placing special emphasis on the analysis of post-transcriptional and posttranslational modifications, and on changes in the trafficking and localization of molecules within cells. Mathematical modelling of the vast amount of data generated from such studies should unravel the constellation of cellular characteristics that determine the stem state (Zipori, 2004). The various bioinformatics tools which play a role include ORF finder for detection of open reading frames, Pfam, Prodom, SMART, Prosite and eMatrix for analysis of protein motifs to identify sequence motifs characteristic of certain protein families, Tmpred for transmembrane helices examination and analysis of signal peptides by SignalP. SignalIP is the best method among binary approaches which employs neural networks and predicts N-terminal secretion signals cleaved by Signal Peptidase-I and their associated cleavage sites (Choo & Ranganathan, 2009). A recent modification of SignalIP is the implementation of a hidden Markov Model (HMM) which seeks to differentiate uncleaved signal anchors from cleaved signal peptides (Flower et al, 2010). The sequencing of the differentially expressed genes and the sequence comparison, using BLAST algorithm, to several databases viz., SwissProt, GenBank protein and nucleotide collections, expressed sequence tags (ESTs), murine and human EST contigs and SCDb (stem cell database) itself (a measure of internal redundancy) helps in further characterization for the definition of stemness (Robert, 2000). The multiple sequence alignment tools which align three or more sequences to identify regions of conservation play a significantly important role in assessment of stem cell receptors. These include CLUSTAL, WebPRANK, T-Coffee and MUSCLE. Needle and LAlign are pairwise sequence alignment tools to identify regions where the sequences are conserved and conversely the regions where the sequence is not conserved. Various functional genomic tools which help to examine and explore data generated from the gene expression experiments include Gene Expression Atlas, EFO tools and EBI R-workbench. The NCBI's Digital Differential Display (DDD) tool is used for comparing EST profiles in order to identify genes with significantly different expression levels. PSORT is another bioinformatic tool which predicts the sub cellular location of gene products and proteins. The development of algorithms for clustering, analysis of the data generated by techniques like microarray, real-time PCR, FACS, immunophenotyping, spectrophotometry, and a host of other molecular biology techniques used in stem cell research provide the working horses for the identification of stem cell signatures.

2. Conclusion

The essence of the mature and relatively stable phenotype is that the cell harbours a set of molecular tools that makes it specialized for its tasks. In the stem-cell state, which is characterized by instability, this contention does not exist. A stem cell has no function and is endowed with potentials only, which are options rather than specific gene-expression patterns, making the identification of stemness undefined. To this adds the relative uncommonness of stem cells (0.0001% to 5%) within a tissue, and the diverse variables involved in stem cell characterization. This generates a wide range of data which could only be processed, classified and clustered using bioinformatic tools to arrive at a universal definition of stemness.

3. References

1• Flower D, Macdonald I, Ramakrishan K, Davies M and Doytchinova I (2010) Immunomee research. 6, 2-16.

2• Choo K, Tan T and Ranganathan S(2009) BMC Bioinformatics 10, S2.

3• Zipori D(2004) Nature Reviews Genetics 5, 873-78.

4• Robert JS (2004) Model systems in stem cell biology, BioEssays 26, 1005–1012.

5• Kapetanovic I, Rosenfeld S and Izmirlian G (2004) New York Academy of Sciences. 1020,10-21.