Lateral Thinking of Emerging Technology

Molecular Markers for Characterization of Stem Cell Lineages

JYOTI SRIVASTAVA, SANJAY PREMI, DEEPALI PATHAK, SUDHIR KUMAR and SHER ALI*
Molecular Genetics Laboratory, National Institute of Immunology, Aruna Asaf Ali Marg, New Delhi-110067

(Received on 2 February 2006; Accepted on 6 June 2006)

Stem cell research offers immense potential to address several key issues related to human health. Different cell lineages of the stem cells are expected to have varying transcriptomes and proteomes. Thus, no two cell types are identical with respect to their gene expression profiles. If the expression profile of the genes is established in different cell types, the same can be used as reliable marker for characterization of cell lineages. Present article deals with the analysis of this issue and assessment of tools that can be used as marker systems for characterization of different cell lineages. The rationale is related to the experimental approach based on the fact that a higher eukaryotic chromosome is endowed with coding and non-coding sequences. From the pool of repeat sequences, the short tandem repeat (STR) sequences (minisatellites) may be used as decoy to access the coding ones by conducting minisatellite associated sequence amplification (MASA) using cDNA from a given cell type representing a cell lineage. Cloning, sequencing and expression analyses of MASA amplicons would uncover gene(s) specific to a particular cell lineage. Since expression profile of each cell type is envisaged to differ, different cell lineages would give rise to different expression profiles. Following this approach, a large number of genes expressing exclusively in a given cell type may be accessed. Working on the hypothesis, we have accessed a total of 43 structural, regulatory and functional genes. The added advantage of this novel approach is that genes are accessed even without screening the cDNA library. Prospect of this approach in the context of stem cells research is discussed.

Key Words: Molecular marker, Short tandem repeats (STR), Minisatellite Associated Sequence Amplification (MASA), Real Time PCR, Transcriptomes and Proteomes

Introduction

Stem cells are now used to address a number of problems related to human health. However, little is known about the status of transcriptomes and proteomes of stem cells. The decisive, instructive, and permissive signals that decide the fate of self renewal or differentiation of these cells are provided by growth factors in the microenvironment or "stem cell niche" [1]. Identification of these growth factors and their respective inputs are critical to understand the developmental and physiological inscription of stem cell-mediated tissue generation, turnover and repair. Furthermore, extending such knowledge to control the differentiation, expansion and lineage specific expression profile of stem cells holds promise for applications in regenerative medicine and biopharmaceuticals.

Broadly, stem cells have been categorized as embryonic stem cells (ESCs) and adult stem cells (ASCs). Accurate maintenance and homeostasis of mammalian tissues depend on the ASCs [2], in contrast to ESCs, having multipotency to give rise to all the tissue types of the body [3]. Adult stem cells are found in a variety of tissues, including skin, muscle, intestine and the hematopoietic system [4, 5]. Within these tissues, stem cells are often integrated in, and protected by niches that provide a unique environment for their specialized functions [6, 7]. In recent years, plasticity of adult stem cells and their potential use for medical therapeutics have been scrutinized. The embryonic cells have the ability to give rise to a variety of cell lineages maintaining unlimited symmetrical self-renewal and clonality [8], providing a potential source of differentiated cells for therapeutic purposes. Human embryonic stem cells (huES) were first isolated and successfully propagated in 1998 [9]. To date 71 independent huES cell lines has been developed worldwide [9, 10] of which, 11 are currently used for research purposes. However, little is known about the mechanism of differentiation and factors regulating their development. Much of the information on ES cells has been derived from the studies on mouse. These include culture conditions promoting differentiation of mouse ES cells into heart, blood, muscle, blood vessels, brain and insulin producing islet cells [11, 12]. Owing to its importance in research and clinical applications, the first step should be the identification of a set of ES cell-specific genes that may be used as molecular markers for their characterization. That would also facilitate to identify unique regulatory pathways of these candidate gene(s).

A large number of mega-, mini- and microsatellite sequences (Short tandem repeats motifs) have been characterized from a number of species and several of these have been used as genetic markers [13] for genome

***** *Author for correspondence – E-mail: sheralib5@hotmail.com, alisher@nii.res.in; Phone: 01126703753; Fax: 01126162125*

analysis [14]. Some of these are evolutionarily conserved (shared across the species) [15] whereas others remain unique (species specific) to a given genome [16]. Repeat sequences are also known to shrink and expand fuelling the process of copy number alteration providing an opportunity to assess their possible roles in reproductive isolation and speciation [17, 18].

Though little is known about the differentiation of stem cells and lineage specific molecular markers, MASA coupled with Real Time PCR may be used to monitor the expression of a number of genes. Such genes if not already mapped in a given species may also be localized using fluorescence in situ hybridization (FISH).

Differentiation Potential of Stem Cells

As described earlier, stem cells are uniquely capable of giving rise to multiple types of differentiated progeny and maintaining tissues. Since the term 'pluripotent' is used to describe stem cells that can form all the three embryonic germ layers; mesoderm, endoderm and ectoderm which are the source of the entire cell types of the body. Ectoderm forms skin cells of epidermis, neurons of brains and pigment cells, mesoderm forms cardiac muscle, skeletal muscle cells, kidney tubule cells, red blood cells and gut smooth muscle, while the fate of pancreatic cells, thyroid cells and lung cells depend upon the inner layer, endoderm (Table 1).

Since it is difficult to delineate distinct features specific to stem cells, focusing on expression profile of individual cell type with its multipotency and self renewal [19] may prove to be very informative. Embryonic stem cells are derived from the so-called 'inner cell mass' of blastocyst developed in culture within 5 days of fertilization of the oocyte [20]. ESCs can form all the somatic tissues, but not the 'extra embryonic' ones necessary for the development of placenta and membranes, and thus cannot form a complete individual. ESCs are immortal having high levels of telomerase expression without undergoing senescence [21]. High potential of proliferation, euploid karyotypes, epigenetic status and high level expression of Oct4 are other important characteristics of the embryonic stem cells [9, 22]. The summary of the different approaches used to understand the elaborate process of differentiation of ESCs is given in recent reviews [9, 10].

Table 1. Development of different tissues from the embryonic stem cells

Embryonic Germ Layer	Differentiated Tissue
Endoderm	Thymus
	Thyroid , parathyroid glands
	Larynx , trachea , lung
	Urinary bladder , vagina , urethra
	Gastrointestinal (Gi) organs (liver, pancreas)
	Lining of the Gi tract
	Lining of the respiratory tract
Mesoderm	Bone marrow (blood)
	Adrenal cortex
	Lymphatic tissue
	Skeletal , smooth and cardiac muscle
	Connective tissue (including bone, cartilage)
	Urogenital system, Primordial germ cells
	Heart and blood vessels(vascular system)
Ectoderm	Skin
	Neural tissue(neuroectoderm)
	Adrenal medulla
	Pituitary gland
	Connective tissue of the head and face
	Eyes, ears

ASCs differentiate to give rise to major specialized type of cells and have been shown to exhibit an unexpected versatility. Studies have shown adult stem cells originating in one germ layer form a variety of other derivatives of the same germ layer (e.g. bone marrow-to-muscle within the mesodermal lineage) and this may trans-differentiate into the derivatives of other germ layers [23, 24, 25]. ASCs are also dedicated to reproduction and somatic tissue homeostasis respectively [26]. ESC genes have been found to show higher complexity with significantly more number of expressing genes than ASCs [26].

Potential Applications of the Stem Cells

With the advancement of therapeutic cloning, the use of different kinds of stem cells has increased several folds. There are over 4,000 registered genetic diseases for which, in principle, different types of stem cells are required to provide the cure. Stem cells can be used for different purposes [27], but an eminent use of ESCs is seen in transplantation medicine for cell replacement therapies. Examples of diseases where a person may be benefited from ESC include diabetes, Parkinson's disease, Stroke, Arthritis, Multiple Sclerosis, Heart failure and Spinal cord. It is known that ESCs are capable of generating neural, cardiac, skeletal muscle, pancreas and liver cells in teratocarcinomas in immuno-deficient mice *in vivo* as well as in tissue culture cells [12, 28]. In addition, adult stem cells may also be used either exclusively or in combination with treatments to achieve significant "healthcare benefits" to the patients suffering from Brain Tumors, Ovarian Cancer, Solid Tumors, Multiple Myeloma, Breast Cancer, Non-Hodgkin's Lymphoma, Multiple Sclerosis, Systemic Lupus, Rheumatoid Arthritis, Anaemia, Stroke Blindness and Immunodeficiency [23, 24, 29, 30].

Molecular Basis of Pluripotency

Few signaling pathways are hypothesized to regulate the capacity of ESCs for continued self renewal and differentiation. The most important one is the inhibition

of BMP-4 via the action of FGF-2 and Noggin which is thought to be crucial for the maintenance of pluripotency and self renewal in stem cells [10, 31]. The other pathways are phosphorylation of SMAD2/3 and stabilization of β-catenin via the PI3K/Akt pathway [10]. The signaling pathways like Wnt, Notch, Sonic hedgehog (Shh) and intrinsic factors such as chromatin remodeling factors and cell cycle regulators have also been reported to be involved in stem cell renewal [32, 33].

Several transcription factors have been found to play a critical role in self-renewal of stem cells whose expression profile is used to characterize embryonic cells. One such factor, Oct3/4, belonging to the POU family of transcriptional regulators, is expressed in pluripotent cell populations both *in vivo* and *in vitro* [34, 35]. Several target genes for Oct 3/4 have been identified including Utf-1 [36], Rex-1 [37], PDGFaR [38], Otx-2 [39], Lefty-1 [39] and Nanog [40, 41] but their roles in embryonic cell pluripotency and self-renewal are still unclear. Other transcriptional regulators such as Sox-2 [42] act in coordination with Oct 3/4. The molecular mechanisms through which Oct 3/4 and various co-factors regulate ES cell pluripotency and self-renewal have been covered in recent report [43]. Future research in this area is envisaged to enable the identification of new genes responsible for pluripotency.

Markers for Stem Cells

Usually, surface markers are used to characterize stem cells, including various glycolipids and glycoproteins which were originally identified on human embryonic carcinoma cells or in human preimplantation embryos, such as SSEA-4, TRA-1-60 and TRA-1-81 [44]. ESCs also express surface antigens initially described in other stem cell populations such as AC133, *c-kit* (CD117), flt3 (CD135) and CD9 [45, 46]. Although several laboratories have reported the expression of these markers in ESCs, their quantitation in different cell lines has only recently been reported [45, 46]. Studies evaluating long-term stability of ESCs have indicated that expression of characteristic surface markers is maintained in ES cell lines following prolonged cultures [46, 47] as well as those maintained in a serum-free culture medium containing a combination of growth factors, transforming growth factor-beta 1 (TGF-β1 basic fibroblast growth factor (bFGF) and/or leukemia inhibitory factor (LIF)[47]. These findings indicate that hES cells maintained over long-term culture in different conditions retain expression of characteristic pluripotent stem cell markers.

Commitment of Hematopoietic Stem Cells and Multilineage Gene Expression

Multipotent hematopoietic progenitor cells appear to be primed for co-expression of a number of genes, characteristics of different lineages [48]. Lineage commitment in multipotential cells has been regarded as an immutable stochastic process but may be modified by extrinsic signaling. The mechanisms, by which lineage-restricted patterns of transcription factor expression is established, are of particular relevance to our understanding their role in normal development [49]. This suggests that multiple lineage-restricted transcription factors are expressed prior to lineage commitment. The regulated accessibility of genes appears to be a critical feature in the stem cell state. The so-called 'multilineage priming' model suggests that stem cells express low levels of a large number of genes that are highly expressed in its differentially committed progeny [48]. A large number of genes have also been found to express in differentiating HSCs, whereas committed precursors express only the proportion of the genes that are related to their lineages [50]. In addition, hematopoietic stem cells express a relatively large number of non-hematopoiesis-associated genes, in particular neuronal genes. Although most of these neuronal genes have not been linked to the induction of neuronal differentiation, their neuronal exclusiveness remains elusive since it illustrates the wide variety of genes expressed in the hematopoietic stem cells [51]. A similar situation seems to hold true for embryonic stem cells also. Embryonic stem cells with a tendency to acquire neuronal identity, express several neuronal lineage-specific genes under normal culture conditions [51]. It appears that differentiation is accompanied by a successive restriction in the repertoire of genes that can be expressed. Thus, there is a relationship amongst differentiation potential, chromatin accessibility and the multipotency at the transcriptional level. Chromatin modifications maintain its status through DNA replication and mitosis. In particular, methylation of the lysine residues in histone 3 appears to play a role in epigenetic memory. These histone modifications recruit polycomb or trithorax group protein complexes to maintain the silencing or activation status of genes [52]. Since stem cells self renew without losing any of their features, they should inherit the same epigenetic status in order to maintain their genetic characteristics. From this viewpoint, it is not unexpected that polycomb group members, such as Bmi1, play a role in the maintenance of several stem cell populations, their self renewal [53] and stem cell memory [54]. A common code for the stem cell-specific expression of factors involved in the epigenetic modifications and/or transcription silencing or activation has not been described. This may not be very surprising as all cells including differentiated ones need enough epigenetic memory to maintain their inherited transcription profile in order to prevent cell death or acquiring an altered cell fate.

Lineage Specific Molecular Marker for Stem Cells

The development of tools for global gene expression analysis allowed characterization of the transcriptomes and proteomes of different stem cell populations. The study conducted on transcription profiles of embryonic, neural and hematopoietic stem cells generated a great deal of enthusiasm [55, 56]. However, this could not provide unique transcriptional profile of a stem cell lineage which could be used as specific marker. Instead, it was observed that some stem cells were enriched with genes not exclusive to these cells, indicating that the combinations of many genes rather than a few individual ones are responsible to provide the stem cell its characteristics. Surprisingly, two common expression profiles for different stem cells showed minimal overlap. Other studies also failed to uncover a unique genetic fingerprint and it was suggested that the limited overlap was due to somewhat different experimental approaches [57, 58]. Even though differences in the selected cells, microarrays, culture methods and computational analyses might have affected the number of overlapping genes for different stem cell types. It appears unlikely that these differences could account for the limited overlap observed from different studies. If specific set of genes contributing to the characteristic features of stem cells exist, the comparison of the expression profiles of widely different stem cell populations should have revealed a common genetic signature [55, 56, 58, 59, 60, 61, 62]. However, analysis of the different stem cell transcriptomes suggests that a single common genetic program does not control the properties of different stem cell and hence different expression profile is expected in different lineages and populations of stem cells.

Minisatellites as Markers for Analysis of Gene Expression

A typical eukaryotic genome contains a sizable part of repetitive DNA sequences. These sequences may be divided into those that are tandemly arrayed and the ones interspersed in the genome. Minisatellites are usually defined as the repetition in tandem of a short (6- to 100-bp) motif spanning 0.5 kb to several kilo bases [63]. Some of the minisatellite consensus sequences, when used as primers in PCR, generate a number of bands which are useful for the development of genetic markers. This technique is referred to as Minisatellite Associated Sequence Amplification (MASA) [14, 16]. The size and high degree of heterozygosity of microsatellites and minisatellites have made them one of the most important DNA based markers for genetic analysis. For instance, the hypervariable minisatellite 33.15, originating from the autosomal myoglobin gene [13] has been widely used to detect genetic polymorphism and genome analysis in humans and several non-human species [14, 64]. A list of other oligos based on minisatellite regions with their nucleotide sequence is given in Table 2. These minisatellites have been found to be highly useful for the analysis of human and non-human genomes.

MASA conducted using cDNA as template, uncovers a number of amplicons each representing a specific mRNA transcript tagged with minisatellite. This suggests that the minisatellites can be an integral part of the transcribing genes. This has been demonstrated in one of our recent studies using a 33.15 consensus repeat as primer for MASA and cDNA from various tissues of buffalo. In the process, 25 amplicons of six different sizes representing several functional, regulatory and structural genes were uncovered [14]. Of these amplicons, the 846/847 bp fragment having homology with Adenylate kinase gene showed gonad specific changes in its nucleotide at six different positions. Another 487 bp fragment showed 100% homology with human Y-chromosome specific sequence. Based on this, we speculate that MASA and Real Time PCR can be used to characterize the transcribing sequences from the stem cells. This in turn would allow to access lineage specific expression of the transcribing genes.

Our Hypothesis: MASA Coupled with Real Time PCR may be Used for Molecular Mining of Transcriptomes of the Stem Cell Lineages

Despite several studies, no common genetic program for the expression profiles of stem cells lineages is available. We hypothesize that MASA coupled with Real Time PCR may be used to uncover the expression profile of genes in different stem cell lineages which will provide detailed information about the transcriptomes of that particular lineage as shown in schematic representation (Fig. 1).

We have already standardized the procedures using buffalo as a model system. RNA was isolated from

Table 2. Synthetic oligos based on consensus repeat motifs used for DNA profiling

Mini-satellites	Number of nucleo-tides	Nucleotide sequence
OAT 33.15	16	5' CACCTCTCCACCTGCC 3'
OAT 15.1	15	5' ATAGATAGATAGATA3'
OAT 15.2	15	5' ACAGACAGACAGACA3'
OAT 18.1	19	5' ATAGATAGATAGATAGATA 3'
OAT 18.2	19	5' ACAGACAGACAGACAGACA 3'
OAT 33.6	22	5' CCTCCAGCCCTCCTCCAGCCCT 3'
OAT 24.1	23	5' ATAGATAGATAGATAGATAGATA 3'
OAT 24.2	23	5' ACAGACAGACAGACAGACAGACA 3'
OAT 24.3	18	5' TTAGGGTTAGGGTTAGGG 3'
OAT 15	15	5' $(TGG)_5$ 3'
OAT 18	18	5' $(TGG)_6$ 3'
OAT 21	21	5' $(TGG)_7$ 3'
OAT 20Y	20	5' TTCCATTCCATTCCATTCCA 3'
Poly CA	16	5' CACACACACACACACA 3'
OMS 1	16	5' GGATGGATGGATGGAT 3'

different tissues of the animal and converted in to cDNA following standard protocols. The cDNA from testis, ovary, spleen, kidney, lung heart and liver was used as template for MASA with a 16 mer oligo based on consensus of 33.15 repeat loci (5'CACCTCTCCACCTGCC3') and 25 amplicons of six different sizes (1263, 846/847, 602, 576, 487 and 324 base pairs) were uncovered [14]. These amplicons represented 43 transcribing genes across the species encompassing several functional, regulatory and structural ones. Some of these fragments showed tissue specific nucleotide changes and expression. For instance, the 846/847 bp fragment representing the Adenylate kinase gene showed six gonad specific (ovary and testis) nucleotide changes at six positions. Some of the fragments represented very important genes e.g. Secreted modular calcium binding protein-1 (SMOC-1) and Leucine rich repeat neuronal 6A (LRRN6A). This evoked our interest to design primers for Real Time PCR and assess the transcription level of these genes. Though SMOC-1 gene transcripts were uncovered from liver only, Real Time PCR showed its maximum (160 folds) expression in liver (Fig. 2). Similarly, Real Time PCR showed maximum expression of AKL, LRRN6A and TCR-? in the testis and T-cell receptor like gene in spleen. Detailed work is reported in our recent publication [14]. It is clear that MASA coupled with Real Time PCR has immense potential to uncover accurate expression of a large number of genes tagged with minisatellites and can be used to design tissue and species specific molecular markers. Following this approach, expression profiles of the genes may be developed using cDNA from different lineages of stem cells. In the process, gene(s) showing highest level of expression in a specific lineage may intern be used as molecular marker. In principle, transcriptomes of any cell lineage can be accessed using MASA approach. Thus, current MASA approach would compliment other existing approaches adding a newer dimension to the stem cell biology and therapeutic cloning.

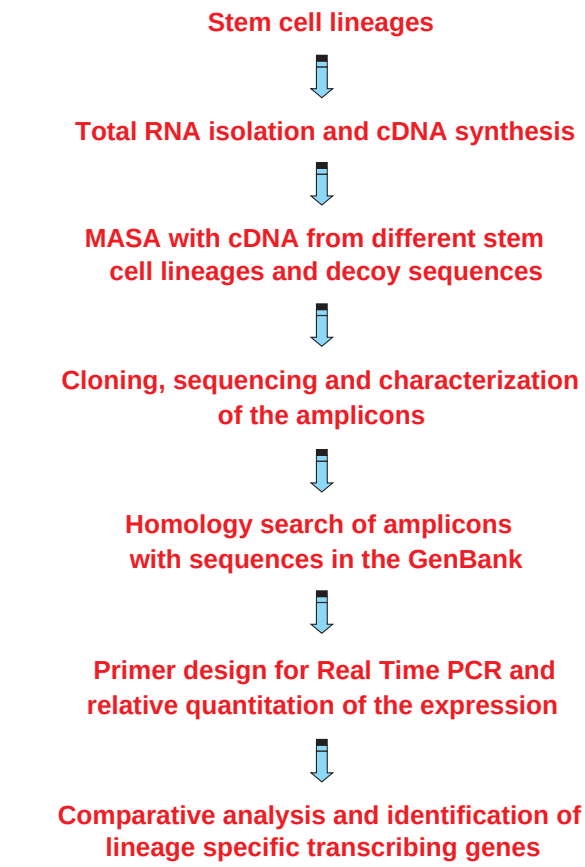
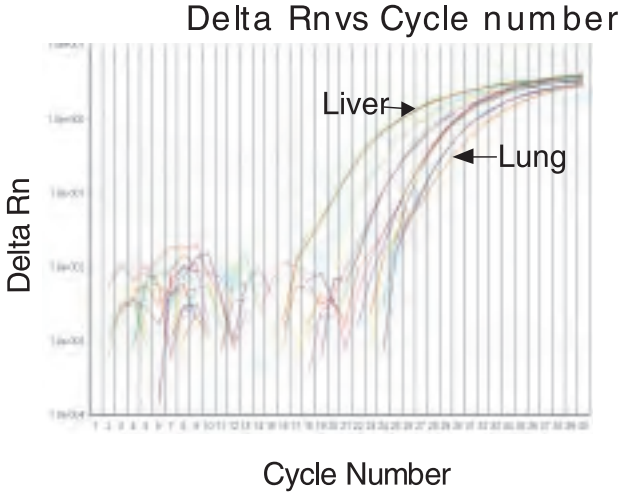


Fig. 1: *Schematic illustration of assessment of transcribing genes from different stem cell lineages without screening the cDNA library employing Minisatellite associated sequence amplification (MASA) coupled with real Time PCR*



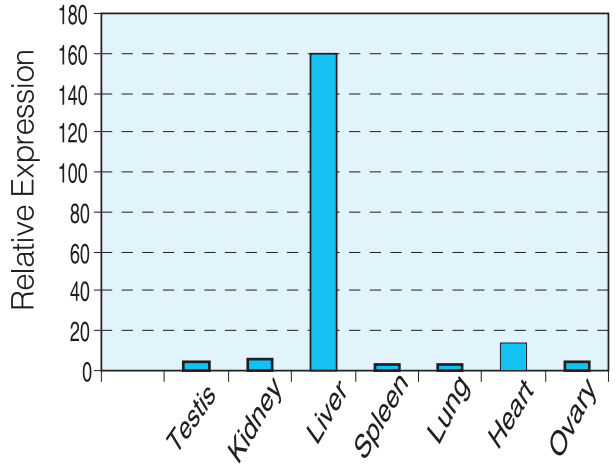


Fig. 2: *Real Time PCR amplification plot showing relative expression of Secreted Modular Calcium Binding Protein (SMOC-1) in different tissues of buffalo* ***(a)*** *and its highest expression (160 folds) in liver compared to that in lung taken as an endogenous control* ***(b)*** *(Taken from Srivastava et al. 2006,* DNA Cell Biol. ***25(1),*** *31-48).*

Acknowledgments

This work was supported by a DBT grant no. BT/PR2225/Med/13/077/2000 and DST grant no. SP/SO/DO3/99) to SA and a core grant from the Department of Biotechnology, Govt. of India to National Institute of Immunology, New Delhi. We thank Shri Khem Singh Negi for technical assistance. Equipment donation from the Alexander Von Humboldt Foundation, Bonn, Germany, is gratefully acknowledged.

References

1. R Schofield *Blood Cells* **4** (1978) 4-7
2. C Niemann *Bio Essays* **28** (2005) 1-5
3. H Darr and N Benvenisty *Handb. Exp. Pharmacol.* **174** (2006) 1-19
4. T Reya and H Clevers *Nature* **434** (2005) 843-850
5. IL Weissman, DJ Anderson and F Gage *Annu Rev Cell Dev Biol* **17** (2001) 387-403
6. E Fuchs, T Tumbar and G Guasch *Cell* **116** (2004) 769-778
7. FM Watt and BL Hogan *Science* **287** (2000) 1427-1430
8. R McKay *Nature* **406** (2000) 361-364
9. LM Hoffmann and MK Carpenter *Nature Biotechnology* **23(6)** (2005) 699-708
10. LA Hyslop, L Armstrong, M Stojkovic and M Lako *Exp Rev Mol Med* **7(19)** (2005) 1-21
11. GM Keller *Curr Opin Cell Biol* **7** (1995) 862-869
12. N Lumelsky, O Blondel, P Laeng, I Velasco, R Ravin and R McKay *Science* **292** (2001) 1389-1394
13. AJ Jeffreys, V Wilson and SL Thein *Nature* **314(6006)** (1985) 67-73
14. J Srivastava, S Premi, D Pathak, Z Ahsan, M Tiwari, LC Garg and S Ali *DNA Cell Biol* **25(1)** (2006) 31-48
15. D Tautz *Nucl Acids Res* **17** (1989) 6463-6471
16. S Ali, AA Azfer, A Bashamboo, PK Mathur, PK Malik, VB Mathur, AK Raha and S Ansari *Gene* **228** (1999) 33-42
17. Y Nakamura, M Leppert, P O'connell, R Wolff, T Holm, M Culver, C Martin, E Fujimoto, M Hoff, E Kumlin et al. *Science* **235** (1987) 1616-1622
18. MV John, I Parwez , VSS Mylavarapu, S Mehta, N Marwah and S Ali ***Gene*** **172** (1996) 191-197
19. D Zipori (2005) *Stem Cells* **23(6)** (1995) 719-726
20. JA Thomson, J Itskovitz-Eldor, SS Shapiro, MA Waknitz, JJ Swiergiel, VS Marshall and JM Jones *Science* **282** (1998) 1145-1147
21. GT Werd and C Mummery *Human Reproduction* **19(4)** (2003) 672-682
22. A Vats, RC Bielby, NS Tolley, R Nerem and JM Palak *Lancet* **366** (2005) 592-602
23. MF Pittenger, AM MacKay, SC Beck, RK Jaiswal, R Douglas, JD Mosca, MA Moorman, DW Simonetti, S Craig and DR Marshak *Science* **284** (1999) 143-147
24. DL Clarke, CB Johansson, J Wilbebirtz, B Veress, E Nilsson, H Karlstrom, U Lendahl and J Frisen *Science* **288** (2000) 1559-1561
25. DJ Anderson, FH Gage and IL Weissman *Nat Med* **7** (2001) 393-395
26. R Xi, D Kirilly and T Xie *Curr Opin Genet Devel* **15** (2005) 381-387
27. G Keller and HR Snodgrass *Nat Med* **5** (1999) 151-152
28. GS Schierle, O Hansson, M Leist, P Nicotera, H Widner, and P Brundin *Nat Med* **5** (1999) 97-100
29. P Stiff, B Chen, W Franklin, D Oldenberg, E Hsi, R Bayer, E Shpall, J Douville, R Mandalam, D Malhotra, T Muller, RD Armstrong, A Smith *Blood* **95(6)** (2000) 2169-2174
30. RJ Chan and MC Yoder *Curr Neurovascular Res* **1(3)** (2004) 197-206
31. Ying Qi-Long, J Nichols, I Chambers and A Smith *Cell* **115** (2003) 281-292
32. R Nusse and HE Varmus *Cell* **31** (1982) 99-109
33. KM Cadigan and R Nusse *Genes Dev* **11** (1997) 3286-3305
34. K Okamoto, H Okazawa, A Okuda, M Sakai, M Muramatsu and H Hamada *Cell* **60** (1990) 461-472
35. MH Rosner, MA Vigano, K Ozato, PM Timmons, F Poirier, PW Rigby and LM Staudt *Nature* **345** (1990) 686-692
36. M Nishimoto, A Fukushima, A Okuda and M Muramatsu *Mol Cell Biol* **19** (1999) 5453-5465
37. E Ben-Shushan, J Thompson, L Gudas and Y Bergman *Mol Cell Biol* **18** (1998) 1866-1878
38. HJ Kraft, S Mosselman, HA Smits, P Hohenstein, E Piek, Q Chen, K Artzt and EJ Van-Zoelen *J Biol Chem* **271** (1996) 12873-12878
39. H Niwa, J Miyazaki and AG Smith *Nat Genet* **24** (2000) 372-376
40. K Mitsui, Y Tokuzawa, H Itoh, K Segawa, M Murakami, K Takahashi, M Maruyama, M Maeda and S Yamanaka *Cell* **113(5)** (2003) 631-642
41. I Chambers, D Colby, M Robertson, J Nichols, S Lee, S Tweedie and A Smith *Cell* **113(5)** (2003) 643-655
42. H Yuan, N Corbin, C Basilico and L Daileiy *Genes Dev* **9** (1995) 2635-2645
43. H Niwa *Cell Struct Funct* **26** (2001) 137-148
44. PW Andrews, I Damjanov, D Simon, GS Banting, C Carlin, NC Dracopoli and J Fogh *Lab Invest* **50(2)** (1984) 147-62
45. MK Carpenter, ES Rosler, GJ Fisk, R Brandenberger, X Ares, T Miura, M Lucero and MS Rao *Dev Dyn* **229(2)** (2004) 243-58
46. DS Kaufman, ET Hanson, RL Lewis, R Auerbach and JA Thomson *Proc Natl Acad Sci USA* **98** (2001) 10716-10721
47. M Amit, MK Carpenter, MS Inokuma, CP Chiu, CP Harris, MA Waknitz, J Itskovitz-Eldor and JA Thomson *Dev Biol.* **227(2) (**2000) 271-278
48. MA Cross and T Enver *Curr Opin Genet Dev* **7(5)** 609-613
49. MA Cross, CM Heyworth, AM Murrell, EO Bockamp, TM Dexter and AR Green *Oncogene* **9(10)** ((1994)) 3013-3016
50. K Akashi, X He, J Chen, H Iwasaki, C Niu, B Steenhard, J Zhang, J Haug and L Li *Blood* **101** (2003) 383-389
51. V Tropepe, S Hitoshi, C Sirard, TW Mak, J Rossant, D van der Kooy *Neuron* **30** (2001) 65-78
52. L Ringrose and R Paro *Annu Rev Genet* **38** (2004) 413-443
53. IK Park, D Qian, M Kiel, MW Becker, M Pihalja, Weissman SJ Morrison and MF Clarke *Nature* **423** (2003) 302-305
54. A Iwama, H Oguro, M Negishi, Y Kato, Y Morita, H Tsukui, H Ema, T Kamijo, Y Katoh-Fukui, H Koseki, M van Lohuizen and H Nakauchi *Immunity* **21** (2004) 843-851
55. NB Ivanova, JT Dimos, C Schaniel, JA Hackney, KA Moore and IR Lemischka *Science* **298** (2002) 601-604
56. M Ramalho-Santos, S Yoon, Y Matsuzaki, RC Mulligan and DA Melton *Science* **298** (2002) 597-600
57. AV Evsikov and D Solter *Science* **302** (2003) 393

58. NO Fortunel, HH Otu, HH Ng, J Chen, X Mu, T Chevassut, X Li, M Joseph, C Bailey, JA Hatzfeld, A Hatzfeld, F Usta, VB Vega, PM Long, TA Libermann and B Lim *Science* **302** (2003) 393
59. N Sato, IM Sanjuan, M Heke, M Uchida, F Naef and AH Brivanlou *Dev Biol* **260** (2003) 404-413
60. JM Sperger, X Chen, JS Draper, JE Antosiewicz, CH Chon, SB Jones, JD Brooks, PW Andrews, PO Brown and JA Thomson *Proc Natl Acad Sci USA* **100** (2003) 13350-13355
61. B Bhattacharya, T Miura, R Brandenberger, J Mejido, Y Luo, AX Yang, BH Joshi, I Ginis, RS Thies, M Amit, I Lyons, BG Condie, J Itskovitz-Eldor, MS Rao and RK Puri *Blood* **103** (2004) 2956-2964
62. I Ginis, Y Luo, T Miura, S Thies, R Brandenberger, S Gerecht-Nir, M Amit, A Hoke, MK Carpenter, J Itskovitz-Eldor and MS Rao *Dev Biol* **269** (2004) 360-380
63. AR Wyman and R White *Proc Natl Acad Sci* **77** (1980) 6754-6758
64. JN Weitzel, JM Hows, AJ Jeffreys, GL Min and JM Goldman *Br J Haematol* **70(1)** (1988) 91-97
65. ES Rosler, GJ Fisk, X Ares, J Irving, T Miura, MS Rao and MK Carpenter *Dev Dyn* **229(2)** (2004) 259-274