



**ANALYSIS OF SIGNAL NETWORKS IN HEMATOPOIETIC STEM CELLS
DETERMINING THE STEM CELL FATE**

Swati Sharma¹, Gurudutta U. Gangenahalli*² and Upma Singh³

¹PhD Scholar, Department of Applied Chemistry, School of Vocational Studies & Applied Sciences, Gautam Buddha University, Yamuna Expressway, Greater Noida, Gautam Buddh Nagar, Uttar Pradesh (India)- 201308.

²Scientist 'G' FNABS, Fellow RS Medicine (UK), FRSB (UK), FRSC (UK); Head, Division of Stem Cell & Gene Therapy Research, Institute of Nuclear Medicine & Allied Sciences (INMAS), Defence Research & Development Organisation (DRDO), Lucknow Road, Timarpur, Delhi -110054, India.

³Assistant Professor, Department of Applied Chemistry, School of Vocational Studies & Applied Sciences, Gautam Buddha University, Yamuna Expressway, Greater Noida, Gautam Buddh Nagar, Uttar Pradesh (India)- 201308.

***Corresponding Author: Dr. Gurudutta U. Gangenahalli**

Scientist 'G' FNABS, Fellow RS Medicine (UK), FRSB (UK), FRSC (UK); Head, Division of Stem Cell & Gene Therapy Research, Institute of Nuclear Medicine & Allied Sciences (INMAS), Defence Research & Development Organisation (DRDO), Lucknow Road, Timarpur, Delhi - 110054, India.

Article Received on 21/11/2017

Article Revised on 11/12/2017

Article Accepted on 31/12/2017

ABSTRACT

Hematopoietic stem cells (HSCs) are the unique cells having the core property of self-renewal/ proliferation and differentiation which make them capable of giving rise to all blood cell lineages. Proliferation and migration of HSCs in a regulated manner maintain healthy homeostasis by the interplay between differentiation, self-renewal/ proliferation and dormancy. Here, we identified and described functional molecular signatures eliciting different biochemical responses in molecular signaling of CD34⁺ HSCs in response to a cocktail of Interleukin-3 (IL-3), Fms-Like Tyrosine Kinase-3 Ligand (FLT-3), and Stem Cell Factor (SCF), using a network-based approach. The bioinformatic analysis of sum total of 22,283 gene transcripts (extracted from microarray data GSE3003) resulted into the characterization of genes (in two subgroups: up-regulated & down-regulated) eliciting stem cell responses: proliferation, differentiation and self-renewal. As *in-vitro* proof of concept, nine significant genes, RBM3, DHX32, PSMD9, SAR1A, SMYD3, ST6GAL1, CKAP4, CTR9 are selected for further analysis, having a key role in molecular signaling. Taken together, our results showed the identification of some novel regulatory genes through computational means and their functional importance in biological pathways.

KEYWORDS: Hematopoietic stem cells (HSCs), self-renewal, proliferation.

INTRODUCTION

Hematopoietic stem cells (HSCs) are functionally defined, a rare population of progenitor cells that have the property of self-renewal and differentiation. HSCs are long known to respond to a wide array of extracellular regulatory signals of complex nature and differentiate into the erythroid, lymphoid and myeloid lineages.^[1] The commitment of HSCs towards a particular lineage cell type is governed by a wide range of factors comprising of various cytokines and growth factors.

Surprisingly, despite the ability to self-renew and an insufficient understanding of HSC ontogeny, it is difficult to expand HSCs *in-vitro*. This has limited its application in human Hematopoietic stem cell (hHSC) transplantation either for the treatment of ionizing radiation exposure or in the treatment of cancer or genetic diseases of the blood or blood-forming system. However, cytokine-mediated therapies (a specific combination of cytokines and hematopoietic growth

factors) are increasingly adopted for treatment of malignancies. Here, in this study, we identified and described functional molecular signatures eliciting different biochemical responses in molecular signaling of CD34⁺ HSCs in response to a cocktail of IL-3, FLT-3 and SCF using a network-based approach.

Recent studies have made attempts for *in-vitro* expansion of HSCs with several combinations of stem cell stimulators, such as IL-1/3/6/11, SCF, TPO, FLT-3, GM-CSF, G-CSF, M-CSF and erythropoietin.^[2] However, the synergistic action of specific combination (IL-3, FLT-3 & SCF), exploited in this study, showed clinical evidence of their effects in the treatment of bone marrow transplant and radiation injury. Oswald et al. (2006) reported the significant qualitative changes in the gene-expression profile of HSCs expanded in a Collagen I Matrix, in the presence of IL-3, FLT-3 and SCF.^[3] However, the responses elicited by overlapping groups of genes and signaling pathways in this background remain unknown. Together, these findings enable us to

analyze the gene expression data (microarray data) from NCBI-GEO accession no. GSE3003 and to identify the novel regulatory genes, serving as biomarkers and elucidating their role in molecular networks.

MATERIAL AND METHODS

Microarray data analysis

The microarray data obtained from GEO series accession number GSE3003 has been analysed using computational tools such as Search Tool for the Retrieval of Interacting Genes (STRING), GeneMANIA, Protein Analysis Evolutionary Relationships (PANTHER) etc. and referred in the current study. Also, based on the analysis performed on the sum of 22,283 gene transcripts, protein-protein interaction (PPI) networks were constructed for the genes found responsible for proliferation, differentiation and self-renewal.

The data extracted from GSE3003 series was normalized to eliminate the effects of non-biological variations using mean log-centering method assuming that mean \log_2 (expression ratio) should be equal to 0 for the gene-set. The normalization factor was calculated as:

$$N_{mlc} = \frac{\sum_{k=1}^{N_{gene-set}} \log_2 \left(\frac{R_k}{G_k} \right)}{N_{gene-set}}$$

The intensities are then rescaled such that $G'_k = G_k \times (2^{N_{mlc}})$ and $R'_k = R_k$. The normalized expression ratio becomes:

$$T'_k = \frac{R'_k}{G'_k} = \frac{R_k}{G_k \times (2^{N_{mlc}})} = \frac{T_k}{2^{N_{mlc}}}$$

Which is equivalent to:

$$\log_2(T'_k) = \log_2(T_k) - \log_2(2^{N_{mlc}}) = \log_2(T_k) - N_{mlc}$$

This adjusts the ratio such that the mean \log_2 (expression ratio) for the gene-set is equal to 0.

Source: An introduction to Microarray data analysis.^[4]

Clustering of genes

Later, the normalized data was used for creating an expression matrix to enlist the Differentially Expressed Genes (DEGs) based on their fold change (FC) and p-values. The threshold of 10 FC and $p \leq 0.05$ was selected to identify the highly expressed genes. These genes were analyzed and clustered based on the biochemical function, i.e., proliferation, differentiation and self-renewal potential.

The resulting subgroups of genes (up-regulated and down-regulated) were clustered, combining STRING and literature mining approaches, which led to the construction of PPI networks for genes governing proliferation and self-renewal.

RESULTS

Identification of novel regulatory genes

Total of 22,283 gene candidates was found expressed in the data analysis, which was then normalized using log-centering method based on their fold change and p-values. The initial screening and segregation resulted into 10,710 up-regulated and 11,573 down-regulated genes.

We selected a threshold of 10 fold change to identify the highly expressed genes (1802 up-regulated and 1618 down-regulated) that may play a crucial role in signal transduction pathways (**unpublished data**). To analyze them further, we clustered both sets of selected genes (up- & down-regulated) based on their biochemical properties (proliferation, differentiation and self-renewal) using computational tools and data mining approaches. However, we employed the network-based analysis particularly for exploring the genes responsible for proliferation and self-renewal.

PPI Network analysis

The highly expressed genes (1802 up-regulated and 1618 down-regulated) which were selected for analysis were clustered based on their biological properties of proliferation and self-renewal (see supplementary files). The protein-protein interaction (PPI) networks were also constructed in each case using STRING software tool, which guided us to delineate the role, played by them in clustered group of genes and comprehend the associations among them.^[5]

The up-regulated and down-regulated genes eliciting proliferation and self-renewal were retrieved using STRING database with evidence view display option, where each interaction edge is coloured differently and represented the different sources of the interaction data (figure 1 & 2).

The interaction network is visualized as a graph with the protein molecule/ gene corresponds to node and edge represents the biological relationships between the nodes. Most of the proteins are concentrated at the center of the network while few are arranged loosely at the periphery. The multiple lines with which some of the interactors are connected to one another represents that these interactions were derived from more than one source of information. Analysis of these interaction networks combining the literature mining approach enables us to select 9 notable genes/ proteomic biomarkers (highlighted in boxes in figure- 1 & 2).

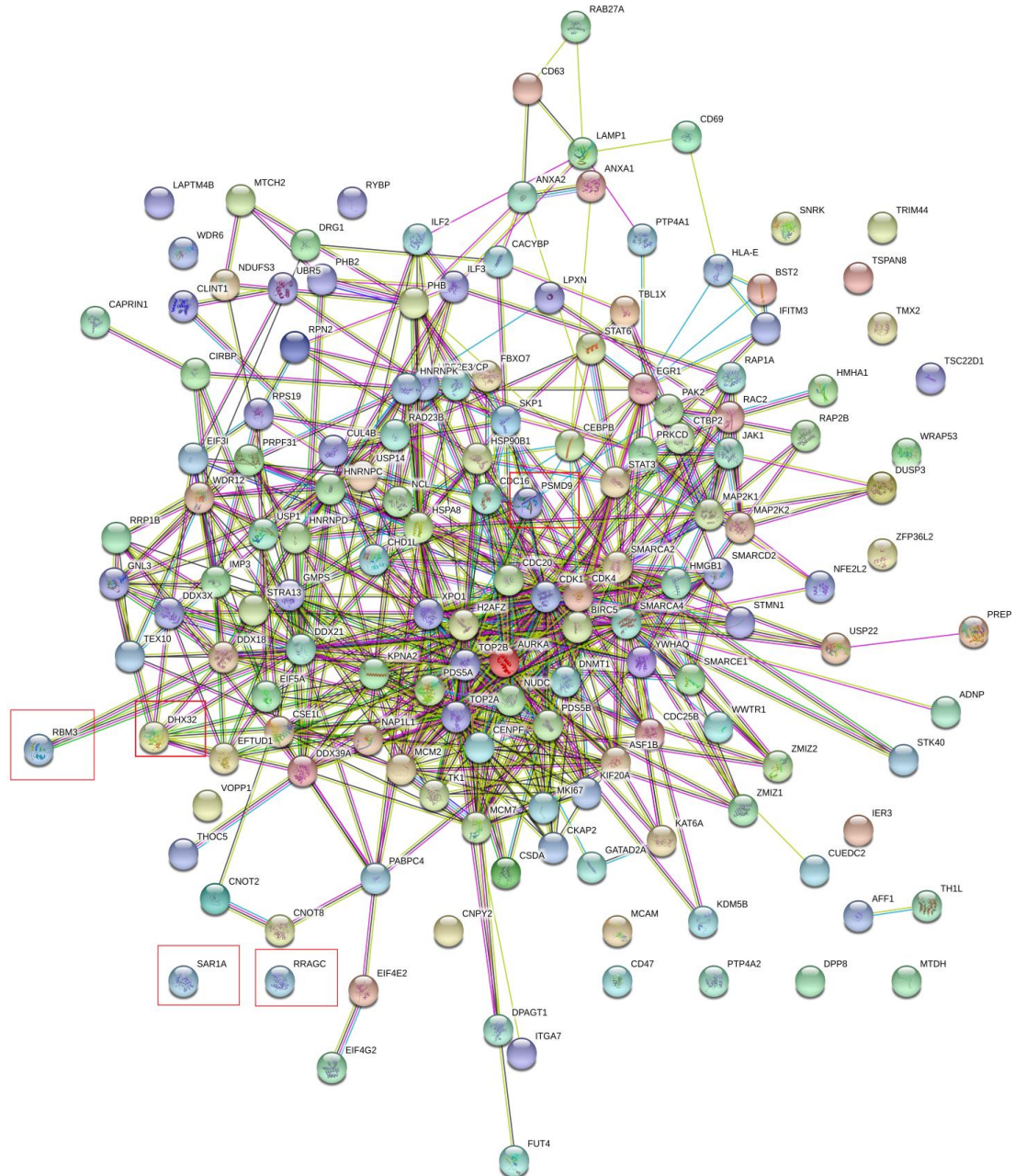


Figure-1: Highly expressed up-regulated genes eliciting proliferation and self-renewal.

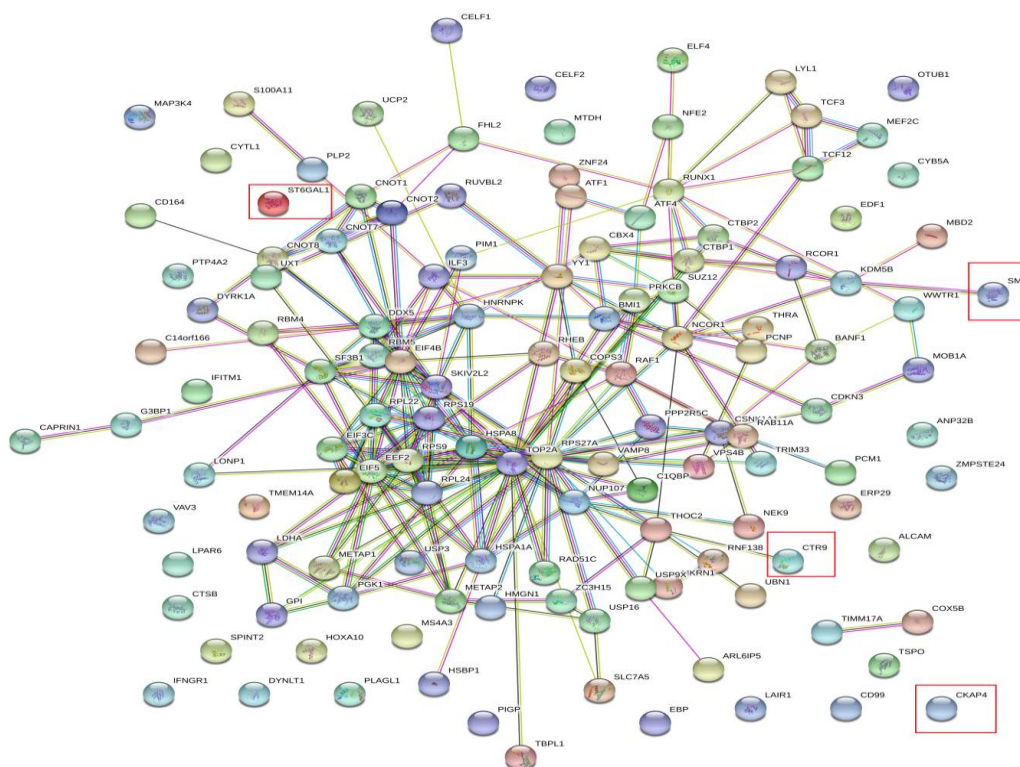


Figure-2: Highly expressed down-regulated genes eliciting proliferation and self-renewal.

Together, these findings indicated about the nine regulatory genes/ proteins which may serve as a biomarker, through several sources such as databases of physical interactions and databases of curated biological pathway knowledge. The targeted sets of genes/ proteins are up-regulated proliferation (RBM3, DHX32, PSMD9, RRAGC, SAR1A) and down-regulated proliferation (SMYD3, ST6GAL1, CKAP4) and self-renewal (CTR9).

Subsequently, these nine proteins were treated as an individual biomarker and their potential interacting partners also identified using the GENEMANIA software tool in order to analyze their functional organization (as shown in Table-1).^[6]

Table-1: Functional biomarkers of targeted genes and their interacting partners.

Gene Symbol	Interacting partners
Up-regulated proliferation	
RBM3	HNRNPA2B1, HNRNPK, ABI1, HNRNPA1, PSMA7, CHAF1A, SLC25A5, TOMM20, UXT, ABCB7, SLMO2, NACA, CDK2, PLP2, MSN, EEF1B2, NONO, RPS20, NAP1L1
DHX32	DHX33, TMEM231, DHX35, DHX34, NSUN7, DHRS7B, DHX9, DHX40, SAV1, DHX30, DHX29, EIF2B4, C10orf118, DHX38
PSMD9	PSMC3, PDX1, PSMC6, ACO2, ANAPC5, NSF, TCF3, PSMB1, PSMB6, TMOD3, PSMD10, PSMC5, RHOA, SLC25A3, UBE2I, PSMB5, ERP29, PSMD11
RRAGC	RRAGA, ATP2B2, ATP2B4, NOL8, COQ4, SUPT16H, RUVBL2, RUVBL1, PSMC2, WDR41, SGPP1, RRAGD, HPS1, PHLDA3, SNX6
SAR1A	TPD52, ATF6, TGFB2, SEC23A, TGFB3, SMAD7, TGFB1, DCTN1, TGFB2R2, ARF4, ANXA7, PRDX4, DERL1, PSMA5, GDE1, MAGED1
Down-regulated proliferation	
SMYD3	HELZ, TGDS, FAR1, MEST, PBX2, TRIT1, MAGED1, EGLN1, HDAC1, CCNB1IP1, MAGED2, VKORC1, INSL6, EED
ST6GAL1	CD22, SEL1L3, ST3GAL6, CLINT1, BTG2, CXCR5, MS4A1, GCDH, ST6GALNAC5, ST3GAL4, ST3GAL2, DCAF11
CKAP4	ADRA1D, HDLBP, SEC61A1, KDELR2, RPN2, P4HB, MANF, HAMP, PSME4, TMED2, HSP90B1, FAM98A, GOLGA3, YIPF2, L1TD1
Down-regulated self-renewal	
CTR9	CDC73, WDR61, PAF1, RTF1, CXCR5, CNTNAP2, BLK, FGG, SSRP1, MYH14, RNGTT, SUPT16H, CSNK2B, TMEM41B, PPM1B

The numbers of interactions represented in each case of gene or protein in particular module are associated with text-mining, experimental and database evidence.

Functional classification of targeted biomarkers

The selected nine biomarkers were and their potential interacting partners (as mentioned in Table-1) were also subjected to another software tool- PANTHER to infer proteins based on their biological functions & Gene Ontology (GO) using scientific evidence & evolutionary relationships.^[7] The combined group of up-regulated genes (RBM3, DHX32, PSMD9, RRAGC & SAR1A) and down-regulated (SMYD3, ST6GAL1, CKAP4 & CTR9) along with their interacting partners were fed into software (PANTHER) separately to explore their protein families based on their function, ontology terms and biological pathways (as shown in Figure-3 & 4).

GO molecular function analysis for the up-regulated group of genes (shown in Figure-3) with their interacting partners revealed the following GO terms overrepresented: 1.0% antioxidant category (GO:0016209), 28.8% binding (GO:0005488), 30.8% catalytic activity (GO:0003824), 1.9% receptor activity (GO:0004872), 4.8% structural molecule activity (GO:0005198), 1.9% translation regulator activity (GO:00045182) and 3.8% transporter activity (GO:0005215). Proteins belonging to biological process category, include the GO terms with: 8.7% biological regulation (GO:0065007), 15.4% cellular component organization or biogenesis (GO:0071840), 56.7% cellular process (GO:0009987), 9.6% developmental process (GO:0032502), 9.6% localization (GO:0051179), 54.8% metabolic process (GO:0008152), 3.8% multicellular organismal process (GO:0032501), 1.0% reproduction (GO:0000003) and 9.6% response to stimulus (GO:0050896).

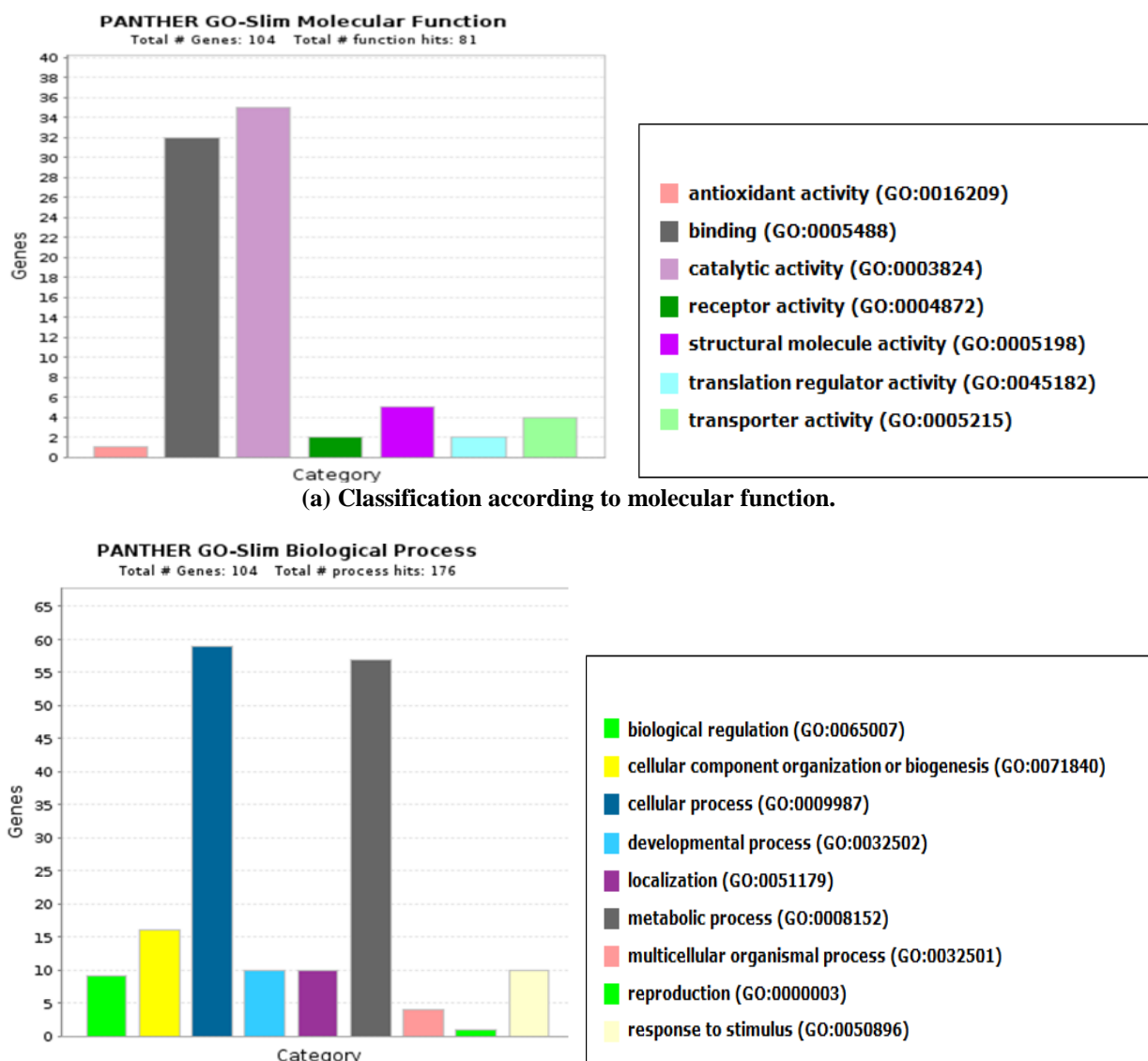
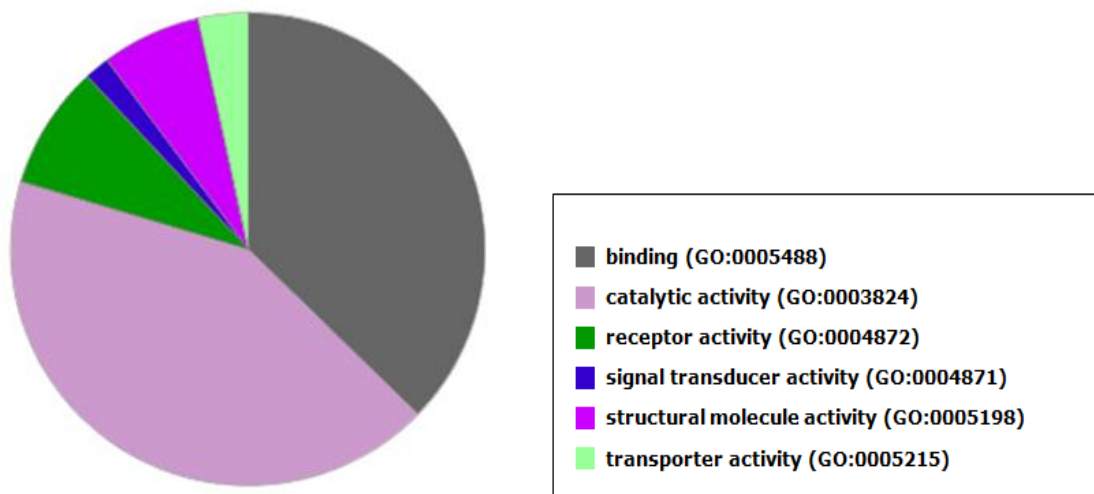


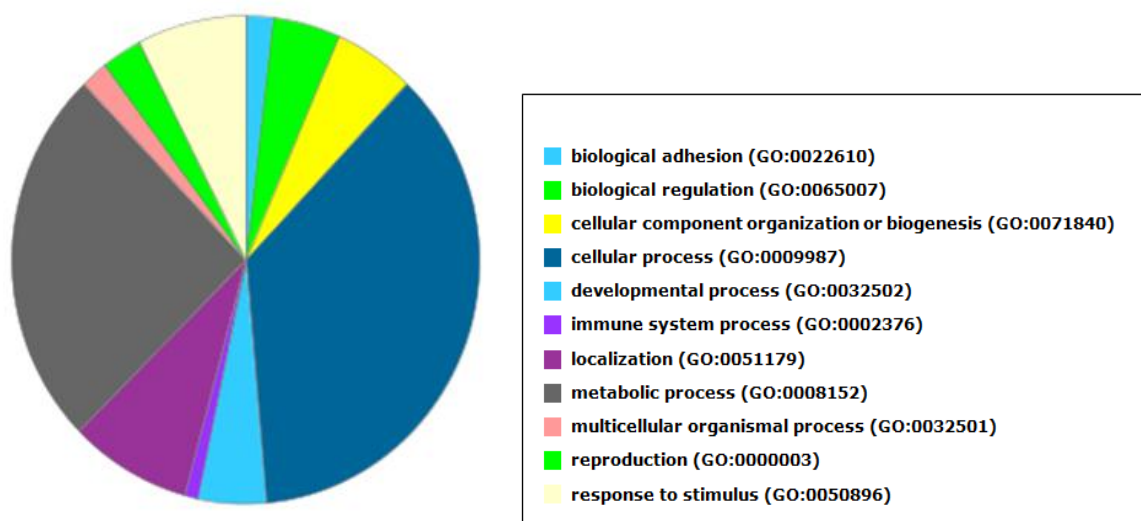
Figure-3: Functional classification of up-regulated genes/proteins based on their (a) molecular function (b) biological process.

Similarly, the *down-regulated group of genes* (as shown in Figure-4) with their interacting partners illustrated that, in the molecular function category, GO terms were represented: 26.2% binding (GO:0005488), 29.8% catalytic activity (GO: 0003824), 6.0% receptor activity (GO: 0004872), 1.2% signal transducer activity (GO:0004871), 4.8% structural molecule activity (GO: 0005198) and 2.4% transporter activity (GO: 0005215). In the category of biological process, these terms were

represented: 6.0% biological regulation (GO: 0065007), 7.1% cellular component organization or biogenesis (GO: 0071840), 46.4% cellular process (GO: 0009987), 6.0% developmental process (GO: 0032502), 1.2% immune system process (GO: 0002376), 10.7% localization (GO: 0051179), 32.1% metabolic process (GO: 0008152), 2.4% multicellular organismal process (GO: 0032501), 3.6% reproduction (GO: 0000003) and 9.5% response to stimulus (GO: 0050896).



(a) Classification according to molecular function.



(b) Classification according to biological process.

Figure-4: Functional classification of down-regulated genes/proteins based on their (a) molecular function (b) biological process.

DISCUSSION

The overall work describes the regulatory proteins and molecular signatures eliciting different biochemical responses in molecular signaling of CD34⁺ HSCs in response to a cocktail of IL-3, FLT-3 and SCF using a network-based approach with the data extracted from GSE3003.^[3] The data analysis and clustering revealed the significant gene candidates those were found responsible for proliferation, differentiation and self-renewal. From the data analysis, up-regulated genes

found responsible for proliferation is 144, differentiation is 98 and self-renewal is 10. Similarly, 111 genes were found eliciting proliferation, 58 differentiation and 28 were accounted for self-renewal in the down regulated genes list.

Further, PPI network for up-regulated and down-regulated gene responsible for proliferation and self-renewal were constructed which indicated the interactions among the genes and resulted in the

selection of nine potential biomarkers for further analysis. Further, the GeneMania software tool provided the potential interacting partner for each of the nine selected biomarkers, which were further explored for their molecular significance using another software tool-PANTHER.^[7]

To explore the molecular function of the up- & down-regulated gene and to gain insight into the biological processes to which these might involve, the selected genes/ proteins along with the interacting partners were analyzed using PANTHER (shown in fig. 3 & 4). The results indicated the involvement of these selected genes into various cellular activities under the molecular function and biological responses establishing the facts in support of the finding of this study. For instance, very few protein like PRDX4 showing antioxidant activity are associated with the free radicals, generated due to increased oxidative activity, which is one of the main actions of reactive oxygen species (ROS) signaling during control of cellular growth.^[9] PRDX4, an active member of antioxidant peroxiredoxin (PRDX) protein family, resides in the endoplasmic reticulum, from where it regulates oxidative stress by reducing hydrogen peroxide (H₂O₂) levels.^[10] It has been known that many growth factors during their action facilitate an increase in the production of H₂O₂ and thereby free radicals, which stimulates growth factor signaling by inhibiting tyrosine phosphatases and tumor suppressor phosphatase. The latter inhibits the hydrolysis of PtdIns3,4,5P3 (PIP3), which functions in cell migration, proliferation, and survival.^[8,10] Further, a large number of genes/ proteins like SMYD3, SMAD2, TGFB1, TOMM20, PSMC5, DHX32, PDX1, RRAGB, HAMP, ZMYND15, PBX2, GCDH, ST6GALNAC2, etc. exhibiting binding and catalytic activity signifies their central role in the transduction process responsible for triggering many of the other signaling pathways. The binding & catalytic activity possessed by protein could be selective, non-covalent, often stoichiometric, interaction with ligand receptor or at a specific site on another protein involving catalytic enzymes.^[11] For instance, SMAD signaling pathway mediates the action of the transforming growth factor β (TGF- β) superfamily, which controls transcription through the Smad transcription factors and regulates many cellular functions such as proliferation, apoptosis, extracellular matrix formation and angiogenesis.^[12] However, relatively lesser number of proteins (1.9% in up-regulated and 6.0% in down-regulated) depicting receptor activity revealed very fewer chances of autocrine signaling among the identified subsets. Then, the few other proteins identified such as TOMM20, ATP2B4, SLMO2, SEC61A1, etc. acts as a carrier that facilitates the directed movement of other proteins within a cell, or between cells.

Moreover, while exploring the proteins that regulated the biological process and molecular function of stem cells, we could find 8.7% among the up-regulated proteins (like ATP2B4, ATP2B2, RRAGB, etc.) and that of 6.0%

from the down-regulated gene sets (like HAMP, GCDH, P4HB, etc.) regulating biological processes. Evident to their involvement in the regulation of biological process, a study conducted by Laplante et al. 2012, confirmed an oncogenic activation of mTOR signaling induces several processes required for cancer cell growth, survival, and proliferation.^[13] It was discovered by two groups independently that amino acid-dependent activation of mTORC1 requires the Rag GTPases.^[14,15] However, four Rag proteins, RagA to RagD found in mammals forms an obligate heterodimer consisting of RagA or RagB with RagC or RagD. The two members of this heterodimer appear to have divergent nucleotide loading states, such that binding of RagA/B to GTP and RagC/D to GDP and vice versa is done. Amino acids promote loading of RagA/B with GTP through an unknown mechanism, which allows the interaction of heterodimer and raptor component of mTORC1.^[14] Thus resulting in the translocation of mTORC1 from an inadequately characterized cytoplasmic location to the lysosomal surface, where docking of Rag GTPases on a multi-subunit complex called 'Ragulator' is observed.^[15] Hence, both Rag GTPase & Ragulator are essential for the activation of mTORC1 by amino acids. Additionally, majority of the proteins in this particular category were found to elicit cellular and metabolic activities either through cell-cell signaling (SMAD2, SAV1, SMAD7, ST3GAL2 etc.) or signal transduction (RRAGA, SAR1A, CDK2, BTG2 etc.) or through a protein metabolic process (TGFB3, PSMD9, PSMC5, EIF2B4, ST6GALNAC5, SMYD3, VKORC1 etc.).

Among numerous strategies to regulate proteins varying from modulating gene expression to post-translational modifications, we identified few members such as DCTN1, PSMD11, SAR1A, HNRNPK, CCDC171, OLFM2, etc. subjected to the signal-induced regulation of localization through the mode of protein transport. Some of these proteins function as determinants of localization context while others perform the same functions in multiple compartments. One common use of this mode of regulation is in the sequestration of an enzyme, such as a kinase, in one compartment to preclude interactions with substrates located in another compartment. Another common use is when a protein has similar roles in multiple compartments with differing needs, such as in the mobilization of DNA repair proteins between specific genome-containing compartments (nucleus, mitochondria, chloroplast) and the cytoplasm.^[16]

CONCLUSION

The significance HSCs in tissue regenerative therapies has been widely known through various clinical evidences. Further, attempts are ongoing for elucidating the ways of reprogramming of the stem cell functional properties, which would be helpful in improving the overall efficiency of transplants. The stem cell niche and microenvironment poses several limitations due to their complexity. However, exploring the synergistic effects

of the different combination of cytokines and growth factors will substantially provide insights and new patterns of their relative stimulation in stem cells. The current study has identified and characterized the novel regulatory genes (in combination such as CKAP4 & MANF; CTR9 & CNTNAP2 etc.) that have a significant role in HSC proliferation and self-renewal.

ACKNOWLEDGEMENT

The help of Dr. Yogesh Kumar Verma for providing technical comments and editing of the manuscript is acknowledged.

CONFLICT OF INTEREST

The author declares no conflict of interest exists.

REFERENCES

- Han T, Yang C, Chang K, Zhang D, Imam FB, Rana TM. Identification of novel genes and networks governing hematopoietic stem cell development. *EMBO Reports*, 2016; 17(12): 1814-1828. doi:10.15252/embr.201642395
- Domen J, Weissman IL. Self-renewal, differentiation or death: regulation and manipulation of hematopoietic stem cell fate. *Mol Med Today*, 1999; 5(5): 201-208; [https://doi.org/10.1016/S1357-4310\(99\)01464-1](https://doi.org/10.1016/S1357-4310(99)01464-1)
- Oswald J, Steudel S, Salchert et al. Gene-Expression Profiling of CD34 Hematopoietic Cells Expanded in a Collagen I Matrix. *Stem Cells*, 2006; 24: 494-500. doi: 10.1634/stemcells.2005-0276
- Babu MM. An introduction to Microarray data analysis. *Computational Genomics* (Ed: R. Grant), Horizon Press, U.K., 2004; 225-249.
- Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Research*, 2015; 43(Database issue): D447-D452. doi:10.1093/nar/gku1003
- Warde FD, Donaldson SL, Comes O, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research*; 38(Web Server issue), 2010; W214-W220. doi:10.1093/nar/gkq537
- Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K et al. PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.*, 2003; 13: 2129-2141.
- Berridge MJ. *Cell Signalling Pathways*, Cell Signalling Biology Module 2. Portland Press Limited, 2014.
- Kiessling LL, Gestwicki JE, Strong LE. Synthetic Multivalent Ligands as Probes of Signal Transduction. *Angewandte Chemie (International ed in English)*, 2006; 45(15): 2348-2368. doi:10.1002/anie.200502794
- Palande KK, Beekman R, van der Meeren LE, Beverloo HB, Valk PJM, Touw IP. The Antioxidant Protein Peroxiredoxin 4 Is Epigenetically Down Regulated in Acute Promyelocytic Leukemia. Bunting KD, ed. *PLoS ONE*, 2011; 6(1): e16340. doi:10.1371/journal.pone.0016340
- Ellenrieder V, Fernandez Zapico ME, Urrutia R. TGF β -mediated signaling and transcriptional regulation in pancreatic development and cancer. *Current opinion in gastroenterology*, 2001; 17(5): 434-440.
- Laplante M, Sabatini DM. mTOR signaling in growth control and disease. *Cell*, 2012; 13; 149(2): 274-93. doi: 10.1016/j.cell.2012.03.017
- Kim E, Goraksha-Hicks P, Li L, Neufeld TP, Guan K-L. Regulation of TORC1 by Rag GTPases in nutrient response. *Nature cell biology*, 2008; 10(8): 935-945. doi:10.1038/ncb1753
- Sancak Y, Peterson TR, Shaul YD, et al. The Rag GTPases bind raptor and mediate amino acid signaling to mTORC1. *Science (New York, NY)*, 2008; 320(5882): 1496-1501. doi:10.1126/science.1157535
- Sancak Y, Bar-Peled L, Zoncu R, Markhard AL, Nada S, Sabatini DM. Ragulator-Rag complex targets mTORC1 to the lysosomal surface and is necessary for its activation by amino acids. *Cell*, 2010; 141(2): 290-303. doi:10.1016/j.cell.2010.02.024
- Bauer NC, Doetsch PW, Corbett AH. Regulating Protein Localization. *Traffic*, 2015; 16: 1039-1061. doi:10.1111/tra.12310