Defence Life Science Journal, Vol. 2, No. 3, July 2017, pp. 354-357, DOI : 10.14429/dlsj.2.11673
© 2017, DESIDOC

RESEARCH PAPER

A Hypothetical Tool for Integrated Omics, Drugs and Clinical Stem Cell Data Analysis

Yogesh Kumar Verma and Gangenahalli Ugraiah Gurudutta*

Division of Stem Cell & Gene Therapy Research, Institute of Nuclear Medicine & Allied Sciences, Delhi - 110 054, India
**E-mail: gugdutta@rediffmail.com*

ABSTRACT

Understanding the molecular basis of stem cell (SC) function is essential to regenerate damaged tissue due to various types of injuries and pathologies. However, there are bottlenecks that prevent the use of Stem cell in the treatment of wide range of diseases. Complete information regarding the control of gene expression in Stem cell is necessary to understand the regulation of Stem cell fates such as self-renewal, differentiation, migration/ homing and apoptosis. However, in the absence of multi-functional, single window tool for the analysis of various omics and drugs data related to Stem cell, the knowledge consolidation is missing to fully harness its potential. A tool which can integrate various databases, analysis software and gives a probable solution for Stem cell mediated regeneration of tissues using available therapeutic options would likely to help in reducing morbidity and mortality associated with various pathologies. This subsumes omics based repurposing of drugs for Stem cell guided tissue regeneration, which will be a boon for clinicians to treat various debilitating diseases/ malfunctions.

Keywords: Integrated system analysis; Omics-drugs-disease-clinical databases; Tissue engineering; Radiation injury

NOMENCLATURE

pp	Patient/population specific solution
NMR	Nuclear magnetic resonance
PET	Positron emission tomography
CT	Computer tomography
MS	Mass spectroscopy
GC	Gas chromatography
LC	Liquid chromatography
CE	Capillary electrophoresis
iTRAQ	Isobaric tags for relative and absolute quantitation
ICAT	Isotope-coded affinity tag
2DE	2 Dimensional gel electrophoresis

1. PERSPECTIVE

Stem cell is a clonal, self-renewing entity that is multipotent and thus can generate several differentiated cell type (describing cellular attributes only)[1]. Increasing understanding of molecular mechanisms governing their potential can help in better understanding and remedy for health problems such as neurodegenerative diseases, heart ailments, cancers and radiation injury etc. Omics analysis of SC has allowed distinguishing the major features that define a cell as SC and, identification of its stemness, differentiation and tissue based niche regulators. Omics data include quantification of mRNA transcripts levels (transcriptome), protein abundance (proteome), metabolic fluxes (fluxome), intracellular and extracellular metabolites concentration (metabolome) and information on protein-protein and protein-DNA interactions (interactome). To efficiently extract relevant biological insight from vast amount of data generated, appropriate and goal-dependent tools are necessary[2]. In order to extrapolate the information stored in omics, drugs and disease database, a new tool is required which could integrate all the available and evolving information and suggests us an optimal solution to a problem (Fig. 1). This would help us in translating the ever-increasing biological information about SC from bench-to-bedside.

At present there are many tools viz. DvD[4], MeSH[5], connectivity map[6], Metacore[7], etc., which could integrate few of the databases and could query the databases on the basis of user input. Even a database (StemBase)[8] has been created, specifically, to facilitate the discovery of gene's function relevant to SC control and differentiation. Nevertheless, these tools lack capability to integrate all the databases to generate a possible solution. The basic idea is to query the available repositories for the identification of SC specific biomarkers under various milieu and novel drug targets for their fates regulation. This will facilitate repurposing of known chemical/ drug like compounds/ clinically used drugs, for which most of the *in vitro* data, *in vivo* data and clinical data is available. The predicted candidate molecules from this tool may find applications in genome based clinical trials, population studies, outcome research and in health care system[9].

1.1 Proposed Tool Capability

This system will be able to retrieve available data, query it, and generate signature profiles and interaction network from

Received : 27 April 2017, Revised : 23 May 2017
Accepted : 01 June 2017, Online published : 02 August 2017

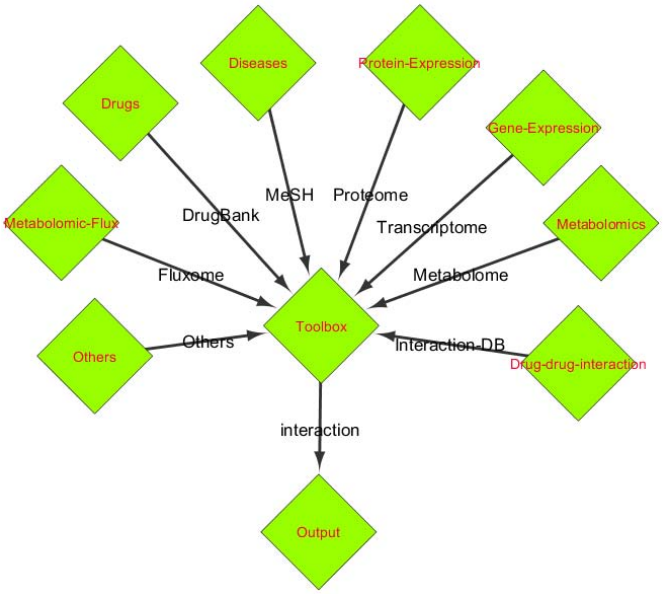


Figure 1. Boxes (nodes) indicating the data available regarding drug-drug interaction, metabolomics, gene expression, protein expression, diseases, drugs and, metabolomic flux etc. Directed arrows (edges) representing their respective databases for the information pool. The tool/ toolbox would integrate information parsed from Fluxome, Drugbank, MeSH (NCBI), Proteome, Transcriptome, Metabolome and other databases to generate an optimal output. The output (prediction) could be used for clinical correlation and new targets identification in lab settings, which might regulate SC fates. This image was generated in Cytoscape[3].

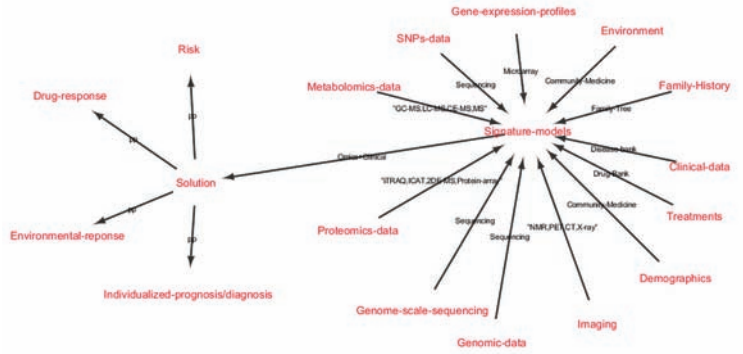


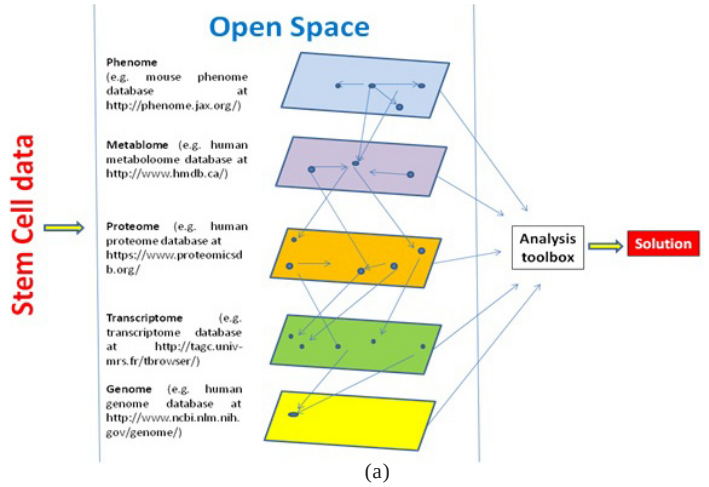
Figure 2. A patient/ population data could act as an input for generating signature profile, which could be compared with the data available in repositories such as gene-expression profiles, SNPs data, metabolomics data, proteomics data, sequencing data and imaging data. This data could be correlated with clinical data and treatments available using SCs as a source. This system analysis may also include family history and demographics. The optimal solution will help in diagnosis/ prognosis of an individual, drug response, environmental response and the associated risks. The image was generated in Cytoscape. The directed edges indicating the techniques/ procedure from which this data was generated.

the databases for a given instance. For example, if a person with disease/ dysfunction/ injury approaches to a clinic, this tool should be able to give prognosis (survival analysis) on the basis of genomics/ proteomics biomarkers spotted on a biosensor chips. The tool could also predict the cause of malfunction {(mutations from Genomics and Single Nucleotide Polymorphisms (SNPs)} and the protein interactions (Proteome and Interactome) being affected (Fig. 2). For this, individual's clinical data during the course of ageing (from childhood to old age) could be stored temporally in a repository as *in vogue* practice of stem cell banking for futuristic applications. This information will be vital in case of emergent situation, without wasting time on lengthy diagnostic procedures. However on the flip side, commercialisation of this repository by private firms will make it exclusive for those who could afford this. The identification of signature model deduced from this information would likely to give cues to the underlying molecular mechanism. These pathways may be tweaked in SC using available/ repurposed drugs for regenerating the affected tissue. It would also include analysis of new metabolites generated due to these pathologies (Metabolomics), the survival chances (Disease bank) and the drugs available (Drug bank) for this patient. In addition, descendents of the patients could be counselled for their chances of inheriting similar phenotype and precautions to be taken for mitigating it. The predicted signature would also find utility in *in vitro* to guide a tissue/organs generation for tissue engineering and increased transplantation efficiency using repurposed-drug-primed SC. The big data analysis based tissue engineering will be a major turn in the conceptual approach to reconstruction of tissues and organs. In addition, advances in SC biology using multidata analysis would make it possible to engineer tissues which were otherwise not amenable to reconstruction.

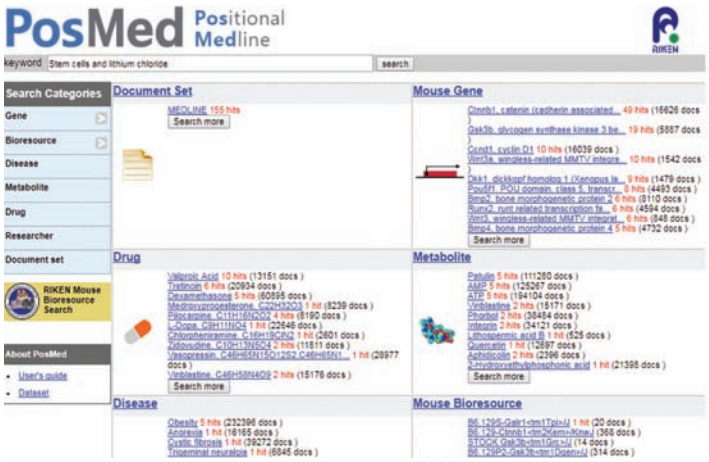
In totality, hitherto, this type of tool is unavailable however endeavour to achieve this has been initiated in the form of Omic space[10]. It is a coordinate based integration and analysis of genomic phenome interactions. It is available on a public server at http://omicspace.riken.jp (known as PosMed or Positional Medicine). It ranks genes, metabolites, diseases and drugs based on the statistical significance of associations which are connected through biological databases such as MEDLINE, OMIM, pathway, coexpressions, molecular interactions and ontology terms (Fig. 3(a), 3(b)).

Nevertheless, it is not capable of providing a possible solution; it only suggests a list of interactions which are needed to be mined manually. In the absence of availability of proposed tool, individual databases and an array of analysis tools may be employed to achieve desired output, for example, to get genes, disease, drugs, metabolites and protein interaction (Table 1). The list of databases and analysis tools would keep on increasing along with a new tool employed for a new analysis. And results obtained would be very difficult to understand and integrate from various tools; this warrants need of a single tool which could parse data from multiple databases.

A central database exchanging information with the proposed tool might also result in better understanding of immune compatibility and enhanced methods for matching patient to donor. Nonetheless, this would need to comply with institutional review board (IRB) protocol and may include creation of stem cell standards like minimum information about stem cell experiments (MIASCE)[12] similar to minimum information about a microarray experiment (MIAME)



(a)



(b)

Figure 3. (a) Omic space is a multiple partially orthogonal domain, where various sorts of biological networks, experimental findings and model-based predictions are represented and integrated through various databases such as phenome, metabolome, proteome, transcriptome and genome. Tool for this type of analysis are being developed (PosMed is one such tool). The SC data could be feed into such a system, this would likely to generate a solution to a problem at hand. and (b) LiCl is well known GSK3β inhibitor used clinically for bipolar disorders. LiCl is also shown to have effect on SC mobilization and granulopoiesis[11]. To analyze the effect of LiCl on SC and parse data available in this regard, a quick search at ProsMed server generated information about genes being regulated by LiCl in SC, in addition to, metabolites and diseases associated with LiCl exposure to SC. The drugs showing action similar to LiCl may be considered as potential target for repurposing whereas the disease association suggests side effects of LiCl treatment.

guidelines[13]. It is also likely that stem cells therapeutics could benefit from the development of informatics tools that combine existing databases and data sources and compliment these with best practices for collaboration. Following multipronged strategy, all stakeholders are needed to be engaged and ensure funding organisations and regulatory authorities are supportive of this proposal. In addition, stem cell researchers must collaborate to overcome hurdles and strive towards development of a common tool available to public.

In summary, a tool for integrated omics, drugs and clinical SC data analysis will increase the potential of SC to infinity and hitherto untreatable diseases. However, extrapolation of *in silico* prediction to humans should not be considered without supporting pre-clinical/ clinical data and prior regulatory authority approval, which will otherwise undermine the beneficial uses of this tool.

Table 1. A simple omics analysis involving 12 steps referring to individual database lacking most of the information, which proposed tool would parse from a central database

Steps	Omics analysis	Databases
1	Microarray of biological sample (such as single cell microarray)	Minimum information about microarray experiment (MIAME) (http://www.ncbi.nlm.nih.gov/geo/)
2	Pathway analysis	Pathway miner (http://www.biorag.org/pathway.html)
3	Protein-protein interaction analysis	Interactome (http://interactome.dfci.harvard.edu/)
4	Correlation analysis	ExAtlas server (https://lgsun.irp.nia.nih.gov/exatlas/)
5	Gene Ontology and commonality	Cytoscape (www.cytoscape.org/)
6	Trancription factor	AMADEUS (acgt.cs.tau.ac.il/amadeus/)
7	Pathway-gene interaction	DAVID (https://david.ncifcrf.gov/)
9	Gene-disease association	Diseasome (https://datahub.io/dataset/fu-berlin-diseasome)
10	Gene-drug association	DGI-db (dgidb.genome.wustl.edu/)
11	Genetic association and tissue expression	Reactome (www.reactome.org/)
12	Gene-metabolite analysis	Biocyc (biocyc.org/)

REFERENCES

1. Lanza, R.; Gearhart, J.; Hogan, B.; Melton, D.; Pedersen, R.; Thomson, J.; Thomas, E.D. & West, M. Essentials of stem cell biology. Elsevier Academic Press, 2006.
2. Choi, S. Introduction to system biology. Totowa, New Jersey, Humana Press., 2007.
3. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B. & Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.*, 2003, **13**(11), 2498-04. doi: 10.1101/gr.1239303
4. Pacini, C.; Iorio, F.; Gonçalves, E.; Iskar, M.; Klabunde, T.; Bork, P. & Saez-Rodriguez, J. DvD: An R/Cytoscape

pipeline for drug repurposing using public repositories of gene expression data. *Bioinformatics*, 2013, **29**(1),132-4. doi: 10.1093/bioinformatics/bts656

5. http://www.ncbi.nlm.nih.gov/mesh (Accessed on 12 July 2017).
6. Lamb, J.; Crawford, E.D.; Peck, D.; Modell, J.W.; Blat, I.C.; Wrobel, M.J.; Lerner, J.; Brunet, J.P.; Subramanian, A.; Ross, K.N.; Reich, M.; Hieronymus, H.; Wei, G.; Armstrong, S.A.; Haggarty, S.J.; Clemons, P.A.; Wei, R.; Carr, S.A.; Lander, E.S. & Golub, T.R. The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease. *Science*, 2006, **313**(5795), 1929-35. doi: 10.1126/science.1132939
7. Shmelkov, E.; Tang, Z.; Aifantis, I. & Statnikov, A. Assessing quality and completeness of human transcriptional regulatory pathways on a genome-wide scale. *Biol. Direct*, 2011, **6**, 15-27. doi: 10.1186/1745-6150-6-15
8. Perez-Iratxeta, C.; Palidwor, G.; Porter, C.J.; Sanche, N.A.; Huska, M.R.; Suomela, B.P.; Muro, E.M.; Krzyzanowski, P.M.; Hughes, E.; Campbell, P.A.; Rudnicki, M.A. & Andrade, M.A. Study of stem cell function using microarray experiments. *FEBS Lett.*, 2005, **579**, 1795-01. doi: 10.1016/j.febslet.2005.02.020
9. Willard, H.F. & Ginsburg, G.S. Genomic and personalized medicine. Volumes I & II. Academic Press, 2009.
10. Toyoda, T. & Wada, A. Omic space: Coordinate-based integration and analysis of genomic phenomic interactions. *Bioinformatics*, 2004, **20**, 1759-65. doi: 10.1093/bioinformatics/bth165
11. Focosi, D.; Azzara, A.; Kast, R.E.; Carulli, G. & Petrini, M. Lithium and hematology: Established and proposed uses. *J. Leukoc. Biol.*, 2009, **85**, 20-8. doi: 10.1189/jlb.0608388
12. Litterman, N.K. & Ekins, S. Databases and collaboration require standards for human stem cell research. *Drug Discov. Today*, 2015, **20**, 247-54. doi:10.1016/j.drudis.2014.10.006
13. Brazma, A.; Hingamp, P.; Quackenbush, J.; Sherlock, G.; Spellman, P.; Stoeckert, C.; Aach, J.; Ansorge, W.; Ball, C.A.; Causton, H.C.; Gaasterland, T.; Glenisson, P.; Holstege, F.C.; Kim, I.F.; Markowitz, V.; Matese, J.C.; Parkinson, H.; Robinson, A.; Sarkans, U.; Schulze-Kremer, S.; Stewart, J.; Taylor, R.; Vilo, J. & Vingron, M. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nature Genet.*, 2001, **29**, 365-71. doi: 10.1038/ng1201-365

CONFLICT OF INTEREST

The authors declare no competing financial interests.

ACKNOWLEDGEMENTS

We thank Dr. AK Singh, Director, INMAS, for his continuous support. This work was funded by Defence Research Development Organisation (DRDO), India.

CONTRIBUTORS

Dr Yogesh Kumar Verma, completed his MSc and PhD in Biomedical Science from Delhi University. Presently working as Scientist 'D' in Institute of Nuclear Medicine and Allied Sciences, Delhi. He has 21 publications and 03 patents (filed) to his credit. He is presently working in the area of stem cell research, microencapsulation, tissue engineering and omics data mining and analysis for decreasing various types of injuries. Contribution in the current study: Conception, writing and data analysis of the manuscript.

Dr Gurudutta Gangenahalli, presently an Additional Director and Head of the Division of Stem Cell Research at Institute of Nuclear Medicine and Allied Sciences, Delhi. He worked in the area of therapeutic potential human stem cell fate response signaling, such as apoptosis, adherence, osteogenesis, differentiation, homing, by using genetic-engineering of human stem cell genes (of CD34, BCL-2, CXCR4, PU.1, SCFr/c-Kit, APC, OSx, Wnt etc) and by high-throughput gene-expression analysis. He is also involved in developing the human stem cell shielding formulations and NMR stem cell tracking methods. Contribution in the current study: Research guidance and editing of the manuscript.