Gene interaction studies in cellular reprogramming of adult stem cells to embyronic like stem cells

Nishanthi Gangadaran[1] & Jannet Vennila James[2]*

[1]Department of Bioinformatics, School of Biotechnology & Health Sciences, Karunya University, Coimbatore – 641114, South India; [2]Department of Bioinformatics, School of Biotechnology, & Health Sciences Karunya University, Coimbatore – 641114, South India; Jannet Vennila James - Email: jannet_r@karunya.edu; Phone: 9487846612; *Corresponding author

Abstract:
The sophisticated process of reprogramming of adult stem cells to embryonic-like stem cells, known as cellular reprogramming, involves the risk of generation of tumorigenic cells due to the complexity involved. Oct4 protein is the inevitable element for inducing pluripotency along with Sox2 and Nanog proteins. In this study, the set of genes interacting with Oct4, Sox2 and Nanog were analyzed and categorized based on their molecular function. Later, the domains of translated products of 46 transcription factors interacting with Oct4, Sox2 and Nanog were identified, clustered them based on the nature of the domain and multiple sequence alignment was performed to find any functionally important consensus regions in the sequence. The key finding of this study is the 13 member cluster of homeo domain transcription factors exhibited some consensus in their sequence.

Keywords: Adult stem cells, Embryonic-like stem cells, Cellular reprogramming, tumorigenic Oct4, Pluripotency, Sox2, Nanog, yamanaka factors.

Background:
Cellular reprogramming is the process of conversion of adult stem cells to embryonic-like stem cells. Four factors are involved in this process of reprogramming say the yamanaka factors-Oct4, Sox2, Klf4 and c-Myc **[1]**. Oct4 is an indispensable component for this conversion. But the success of reprogramming depends on several factors such as amount of factors used and the time period of exposure. Over exposure may result in dangerous results such as generation of tumorigenic cells **[1]**. Later it was discovered how long an adult cells need to be exposed to reprogramming factors before they get converted to an embryonic-like state **[2]**. They defined the sequence of events that occur during reprogramming **[3, 4]**. This finding informs that the duration of exposure on the cells is important for the programming to be effective. High expression of Oct4 and Klf4 combined with lower expression of c-Myc and Sox2 produced Induced Pluripotent Stem (iPS) cells **[2]**. In order to maintain the pluripotency, glycolysis is decoupled from oxidative phosphorylation (OxPhos-which includes Krebs cycle

and ETC) **[5, 6]**. Now, Oct4 being an indispensable element in reprogramming, understanding its functions is of great importance **[6]**. Oct4 forms a hetero dimer with Sox2 and the complex binds to the DNA. Sox2 binds to the consensus motif CATTGTT, and Oct4 binds to ATTTGCAT. Oct4 –Sox2 complex binds to thousands of regulatory sites in the ESC genome and positively regulate Nanog **[7]**. It was found that, Oct4 bind to a factor called PSBP (pluripotential cell-specific Sox element-binding protein), which in turn binds to Sox2 and control Nanog expression. At the same time both Oct4 and Sox2 binds to DNA **[8, 9]**. Oct4 interacts with thousands of other proteins which are of importance for maintaining the pluripotent state of the cell **[10]**. Several factors act as positive or negative regulators for Oct4 **[11]**.

Methodology:
There are thousands of genes with which Oct4, Sox2 and Nanog interacts for inducing pluripotency. Out of these thousands of interactions, there are 404 genes interacting with Oct4 and Sox2.

BIOINFORMATION

According to Boyer *et al*, **[12],** 353 of these genes were co-occupied by Oct4, Sox2 and Nanog. These 353 genes are of importance because all the three factors play major role in inducing pluripotency **[12].** These 353 genes were classified based on their interaction with Nanog and later based on the location of interaction. Some of the genes interact with Oct4 and Sox2 using the same site whereas others interact with different locations. The functions of these genes were identified by database searching. Using a tool GOEAST (Gene Ontology Enrichment Analysis Software Toolkit) **[13]**, genes were clustered based on their molecular functions. Out of the several categories, those genes showing transcription regulation activity were separated and the domains of their translated products were identified from Uniprot. The interaction networks of the transcription regulation genes were studied using a tool STRING **[14]**.

Genes were further classified based on their domain nature. Multiple sequence alignment was performed using MultAlin on clusters of domains to identify the occurrence of any functionally important consensus regions in the sequence. Later, the influence of Oct4, Sox2 and Nanog on these genes were determined and drugs were identified that can induce the same effect, by using a tool X2K (Expression 2 Kinases) **[15]**. Attempts were also made to find phytochemical analogs for those drugs, which may serve as an alternative factor for Oct4.

Results and Discussion:
Classification
353 genes co-occupied by Oct4, Sox2 and Nanog and reported by Boyer *et al*, **[12]** were used for analysis. These genes were then classified in to two categories based on their type of interaction with Oct4 and Sox2, after retrieving them from Uniprot database **[16].** Among 353 genes, 187 genes exhibited interaction with Oct4 and Sox2 at different location and 166 genes at same location. In real time, Oct4 and Sox2 form a hetero dimer before interacting with DNA for inducing pluripotency. This implies that, Oct4 and Sox2 cannot interact with genes on same location at the same time. This make the set of same location interaction less important compared to different location interactions. 29 out of above mentioned 187 genes and 20 out of 166 genes were found not to interact with Nanog.

Clustering and interaction network analysis
These genes were then clustered based on their molecular function using a tool GOEAST Transcription factors are essential for the regulation of gene expression and hence are important in any network **[7, 17]**. From the cluster, 46 out of 353 genes were involved in transcription regulation. The interaction networks of these 46 transcription regulation genes were studied STRING.

Domain wise clustering and multiple sequence alignment
These 46 were then clustered based on the nature of the domain of their translated products. Different types of domains starting from homeo domains to coiled coils, zinc fingers; trans-membrane helices etc. were involved. These clusters were then used to perform multiple sequence alignment using the tool MultAlin **[18]** for the identification of consensus regions in the sequence. Identification of functionally important consensus regions is of great significance because these could be the region

with which Oct4, Sox2 and Nanog are interacting for inducing pluripotency. By finding alternative factors that can bind to those regions and induce the same effect as that of the trio elements will help in deriving new factors for inducing pluripotency. A domain wise classification created several clusters. Out of the different clusters, the cluster of homeo domain transcription factors having 13 members exhibited some consensus in their sequence **(Figure 1).** The amino acids tryptophan, phenylalanine, asparagine and arginine are conserved throughout in homeo domain containing proteins. This consensus was found to be functionally important because, while analyzing the functions of genes, those having this consensus region in their sequence exhibited sequence-specific DNA binding transcription factor activity whereas the others didn't. This enunciates the importance of this consensus region in the sequence.
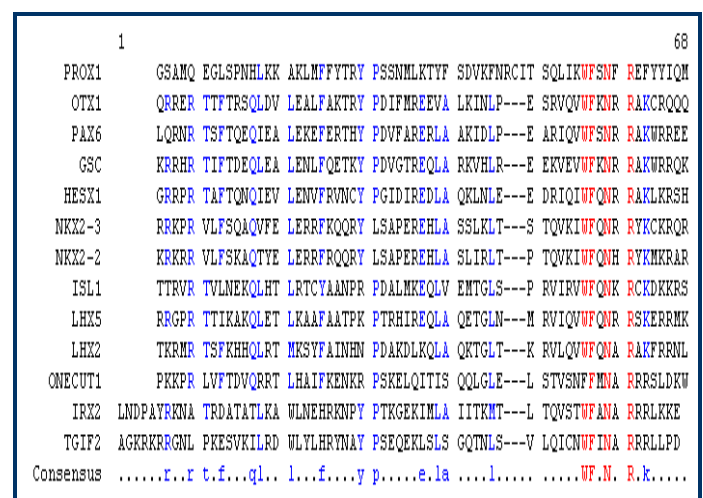

Figure 1: Multiple sequence alignment of homeo domains Transcription factors

Gene regulation
The influences of Oct4, Sox2 and Nanog on the genes were identified. Out of the 46 transcription factors, 26 were down regulated by Oct4, Sox2 and Nanog action and 14 were up regulated. And out of the 13 homeo domain containing transcription factors, 8 of them were down regulated. This implies that majority of the homeo domain transcription factors favors differentiation and hence their down regulation **Table 1 (see supplementary material)** is favored by Oct4, Sox2 and Nanog for maintaining the cell in pluripotent state.

Drug interaction and phytochemical analogs
Later using a tool X2K– Expressions 2 Kinases, drugs that can induce the same effect as that of Oct4, Sox2 and Nanog on these genes were identified. Attempts were made to identify phytochemical analogs of the drugs. Most of them belonged to the family of Boraginaceae, compositae, Leguminosae and Euphorbiaceae.

Conclusion:
The above study has helped in identifying alternate factor for inducing pluripotency for a subclass of genes. In the above analysis, a sub class of genes, say 13 homeo domain containing transcription factors that are involved in plurpotency network out of the 353 genes interacting with all the three factors could be replaced by some phytochemical products. Identification of

BIOINFORMATION

some consensus region for all the genes involved in pluripotency and determining the effect of Oct4, Sox2 and Nanog on those regions, and identifying an alternate factor that can induce the same effect as that of the trio elements will help us in replacing these tumor inducing components with safer methods for inducing pluripotency. This emphasizes the fact that identifying functionally important consensus regions in all the genes involved in pluripotency can help us find an alternate factor for Oct4 for inducing pluripotency without the risk of tumor formation. Finding an alternate factor for inducing pluripotency without the risk of tumor formation can revolutionize reprogramming and widen the scope of its use in saving human lives. Phytochemical analogs identification should be encouraged because this can bring down the cost of reprogramming and equip its wide spread implementation.

References:
[1]   Zhu S *et al. Cell Stem Cell.* 2010 **7**: 651 [PMID: 21112560]
[2]   Carey BW *et al. Cell Stem Cell.* 2011 **9**: 588 [PMID: 22136932]
[3]   http://www.sumobrain.com.
[4]   Stadtfeld M *et al. Cell Stem Cell.* 2008 **2**: 230 [PMID: 18371448]
[5]   Shyh-Chang N *et al. EMBO J.* 2011 **30**: 4851 [PMID: 22166995]
[6]   Babaie Y *et al. Stem Cells.* 2007 **25**: 500 [PMID: 17068183]
[7]   Ding J *et al. Cell Res.* 2012 **22**: 155 [PMID: 22083510]
[8]   Boiani M & Schöler RH, *Nat Rev Mol Cell Bio.* 2005 **6**: 872 [PMID: 16227977]
[9]   http:// www.qiagen.com
[10] http://www.medicalnewstoday.com.
[11] Saxe JP *et al. PLoS One.* 2009 **4**: e4467 [PMID: 19221599]
[12] Boyer LA *et al. Cell.* 2005 **122:** 947 [PMID: 16153702]
[13] http://omicslab.genetics.ac.cn/GOEAST
[14] http://string-db.org.
[15] http://www.maayanlab.net/X2K/index.html
[16] http://www.uniprot.org
[17] Nishiyama A *et al. Cell Stem Cell.* 2009 **5**: 420 [PMID: 19796622]
[18] http://multalin.toulouse.inra.fr/multalin

Supplementary material:

Table 1: Gene Regulated by Oct4, Sox2 and Nanog

Up Regulated Genes		Down Regulated Genes					Unknown
lcmt	ankrd1	foxd3	dmrt1	sox5	nkx2-2	**otx1**	muc1
hesx1	zic2	otx1	Rorb	yaf2	Nfix	**isl1**	nr4a3
zic1	hmg20a	zic4	lhx2	myf5	**lhx2**		af10
sox2	Jup	hand2	nkx2-3	Spic	**nkx2-3**		fli1
Rest	commd7	isl1	pax6	tbx5	**pax6**		tgif2
irx2	ankrd15	neurog	Gsc	dach1	**Gsc**		**lhx5**
Msc	**Gsc**	hand1	onecut1	zic5	**onecut1**		**prox1**
enpp2	**onecut1**	nr2e1	foxb1	gli3	**nkx2-2**		**tgif2**

Note: Genes shown in red color indicates homeo domain containing genes