Life Science Informatics Publications
Research Journal of Life Sciences, Bioinformatics, Pharmaceutical and Chemical Sciences
Journal Home page http://www.rjlbpcs.com/
RJLBPCS
ISSN 2454-6348

Original Research Article **DOI: 10.26479/2019.0502.21**

THE CASE OF EFFICIENT ANIMAL MODELS FOR REGENERATIVE MEDICINE: A COMPUTATIONAL INVESTIGATION

Dhanya V. Menon[1,2*], Darshan Patel[1], Prashanth Nagaraja [3], Chaitanya G. Joshi [4]

1. P. D. Patel Institute of Applied Sciences, Charusat University, Anand, Gujarat.
2. School of Regenerative Medicine, Manipal Academy of Higher Education, Bangalore.
3. Department of Biotechnology, Reva University, Bangalore.
4. Department of Animal Biotechnology, College of Veterinary Sciences and Animal Husbandry, Anand, Gujarat.

ABSTRACT: Induced pluripotent stem cells have revolutionized stem cell research. There are many limitations in using human patients and also human embryonic stem cells (ESCs) for stem cell research. Selecting the most appropriate model for stem cell research is essential for the advancements in regenerative medicine. Validation of the iPS cell therapies requires efficient animal models, supporting faster and effective clinical trials. Transcription factor networks play a major role in reprogramming terminally differentiated cells into iPSCs. This study aims to understand the relatedness between large animal models like pig and dog with that of human, based on four major reprogramming factors- Oct4, Sox2, Klf4 and cMyc (OSKM). OSKM interactomes were constructed to comprehend protein interactions, and closely interacting genes were used for functional enrichment analysis for understanding the functional characters. Phylogenetic analysis of each factor in the selected species was performed, which led to more insights into the evolutionary relationships amongst them. Further functional studies elucidating reprogramming pathways will enable the identification of efficient animal models for applications in stem cell therapy and regenerative medicine.

KEYWORDS: Induced pluripotent stem cells (iPSC); domestic animal iPSCs; pluripotency pathways; interactome; gene enrichment; phylogenetic analysis.

Corresponding Author: Dhanya V. Menon*

P. D. Patel Institute of Applied Sciences, Charusat University, Anand, Gujarat.

ABBREVIATIONS

iPSC, Induced pluripotent stem cells; ESC, Embryonic stem cells; somatic cell nuclear transfer, SCNT; STRING, Search Tool for the Retrieval of Interacting Genes; GO, Gene Ontology; Oct4,octamer-binding transcription factor 4;Sox2, Sex determining region Y box-2; Klf4, Kruppel like factor; bp, base pair; MSA, multiple sequence alignment; db, database; ICM, inner cell mass; LIF, leukemia inhibiting factor; bFGF, basic fibroblast growth factor; PPI, protein-protein interaction; MEGA, Molecular Evolutionary Genetic Analysis; OSKM, Oct4, Sox2, Klf4, cMyc.

1. INTRODUCTION

Many years of path-breaking research in domestic animals led to the development of "Dolly" the sheep, by somatic cell nuclear transfer [1]. This raised many questions, and an attempt to answer a few of these led to the exciting discovery of induced pluripotent stem cells by Takahashi and Yamanaka [2].iPSCs offer an efficient alternative to ESCs- as they can be produced from any cell source and can be differentiated into any desired cell type. Their self-renewing potential makes them ideal candidates for biomedical research. Both the efficiency and limitation of iPSCs need to be investigated thoroughly by using animal models. Selecting the most appropriate model for stem cell research is essential for any potential advancement in regenerative medicine. The mouse is a popular animal model in stem cell biology and regenerative medicine but species differences have to be carefully examined, before extrapolating the findings in mice to that in higher animals. This statement holds true, especially in diseases which do not occur naturally in the mouse. The mouse has a shorter life expectancy which makes it an unsuitable model in aging-related studies. Primate models are the closest to humans but there are many limitations to their use in research settings, mainly due to difficulties in their care and maintenance. Dog and pig have been widely used as large animal models, due to physiological, anatomical and size similarities to humans [3]. These models have been proved to be successful in several diseases like cardiovascular illnesses, diabetes, ophthalmic diseases, spinal cord injuries and some cancers. Pre-clinical porcine models have been designed for retinitis pigmentosa, Huntington`s and Alzheimer`s diseases, and cystic fibrosis. iPSCs have been derived from dog[4,5] and pig [6]. Long term follow-up of stem cell therapy is possible in these models, unlike that in mouse. They can be effectively used for preclinical disease modelling, biomarker development, and therapeutic cloning. Also, large animal iPSCs have many biotechnological applications like biopharming, disease resistance, animal improvement and germplasm conservation [7]. Canine iPSCs has been differentiated into endothelial cells, used for treating immune deficient murine models of myocardial infarction and hindlimb ischemia [8]. Porcine iPSCs have been differentiated into rod photoreceptors and integrated into a damaged swine retina [9]. iPS cells are reprogrammed from somatic cells upon overexpression of four vital reprogramming factors, OSKM, generally known as Yamanaka factors (2) as illustrated in figure 1.

Figure 1: **OSKM factors leading a differentiated fibroblast cell on the road to pluripotency**

Oct4, along with Sox2, Klf4 and Nanog forms the "core" reprogramming factors, inducing and maintaining the pluripotency. Reprogramming is an intricate process which involves many regulatory proteins along with OSKM factors, forming a network to enable the process. This protein network plays an active role in the survival and maintenance of iPSCs in all species. An understanding of these network factors (interactome) helps in elucidating the pluripotency and self-renewal pathways in domestic animals like pig and dog. Optimizing the media and growth conditions is essential for ensuring good quality stem cells. This can be done with an understanding of the pathways involved in maintaining the pluripotency and self-renewal. We performed the interactome analysis and gene enrichment studies of OSKM factors, focussing on two large animal models dog and pig, in comparison with that of mouse and human. Many common interacting partners present in OSKM interactomes were shared targets of OSKM factors which infer their roles in pluripotency and self-renewal. Subsequent phylogenetic analysis by Mega6(10) gave a clear understanding of the evolutionary relationship among the species based on the OSKM reprogramming factors. Interestingly, the phylogenetic studies gave more understanding of the relatedness of large animal models such as dog and pig to that of the human.

2. MATERIALS AND METHODS

Interactome analysis

The relatedness between OSKM proteins and their interacting partners were studied using STRING analysis with a medium confidence score of 0.4. The interactomes involve physical and functional relationships, which are derived from experimental evidence, co -expression, neighbourhood, text mining (mentioned in pub med) and databases. Molecular actions in the network edges were selected where the line shape indicated the predicted mode of action. Different colours of interaction lines showed different molecular actions; green line represents activation, blue line for binding, yellow line for transcriptional regulation, red line for inhibition and purple line for post-translational modification.

Functional enrichment studies

The interactome data files from STRING database(version 10.5) were uploaded to TOPFUN(11) database for plotting and visualizing results of gene annotations based on three major levels; biological processes, molecular functions and cell components of the individual organisms. Significant enrichment was considered when the false discovery rate was ≤.05.

Sequence retrieval and similarity search

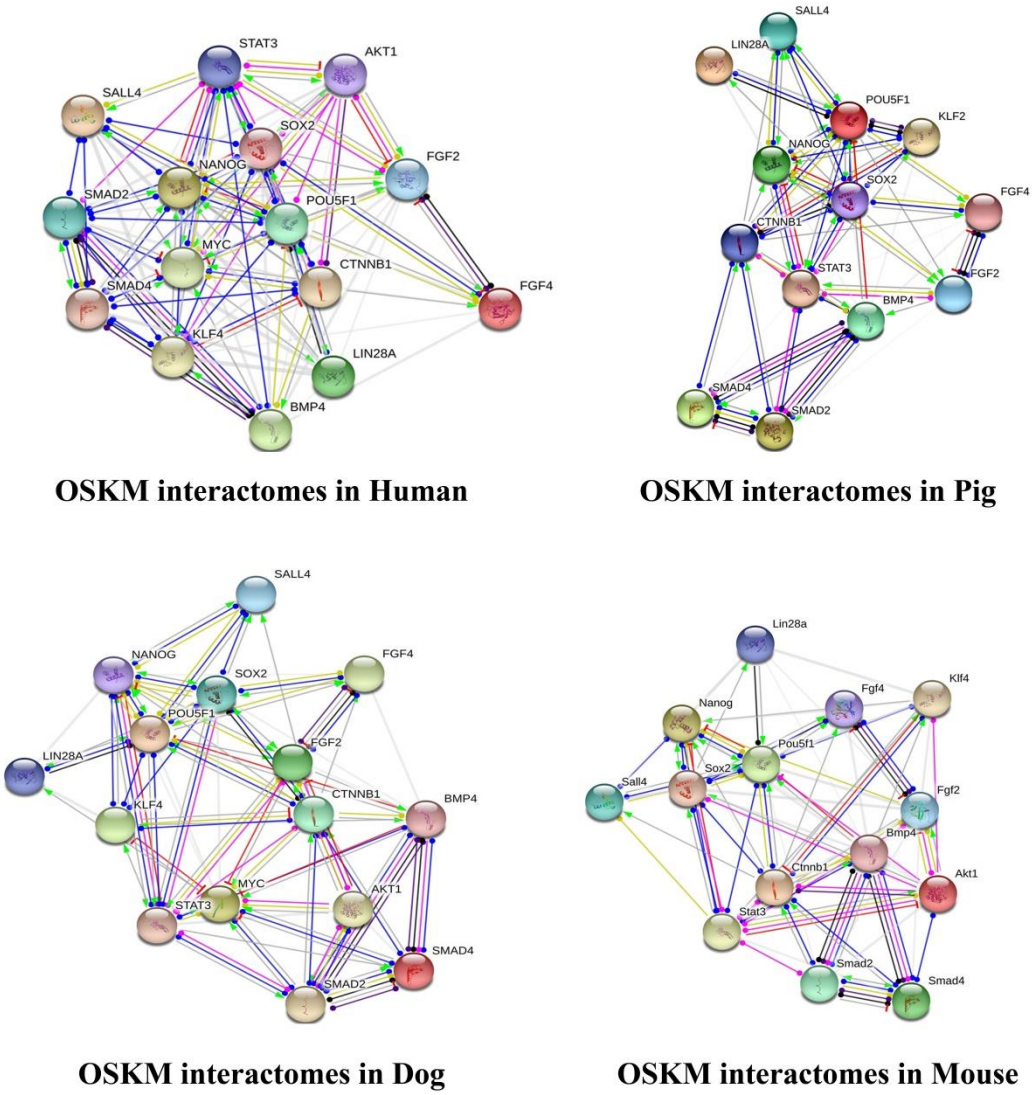
Annotated protein sequences of Oct4, Sox2, Klf4 and cMyc of the selected organisms were retrieved from the NCBI Genbank database (http://www.ncbi.nlm.nih.gov/). The protein sequences from selected species were aligned by BLASTP to predict sequences with a high homology to the template sequences. Chimpanzee and rat were used as internal controls to that of human and mouse. Sequences with an E-value < 10^{-20} and more than 80% alignment coverage were selected for further analysis. These templates were used to predict evolutionary relationship and protein structure prediction.

Phylogenetic tree construction

The selected OSKM template sequences of six species were aligned by the multiple sequence alignment with the program MUSCLE in MEGA6, with default parameters. Best fit model of amino acid substitution for each factor was selected and the phylogenetic tree was constructed by MEGA 6 using the Neighbour-Joining method, based on p- value and substitution matrix with a bootstrap value of 1000.

3. RESULTS AND DISCUSSION

Protein-protein interaction network was built by selecting OSKM proteins in the selected species. A positive correlation was observed between OSKM and other transcription factors as shown in figure 2. This showed that these TF`s interact with each other and participate in the same biological pathways.



OSKM interactomes in Human **OSKM interactomes in Pig**

OSKM interactomes in Dog **OSKM interactomes in Mouse**

Figure 2: Interactome analysis of OSKM transcription factors in A- Human; B-Pig; C-Dog; D-Mouse

A tight protein network and subsequent enrichment were observed in the OSKM interactomes, pointing to the role of individual proteins in stem cell maintenance and development. These proteins constitute protein complexes and reveal their roles in reprogramming mechanisms as enhancers and activators. Human and mouse showed different interacting partners. Most of the identified proteins were involved in many processes especially in stem cell population maintenance, stem cell development, stem cell differentiation, cell fate determination, chromatin remodelling and epigenetic regulation. In tandem with expectations, terms like stem cell population maintenance, positive regulation of transcription by RNA polymerase II, stem cell differentiation, cell fate commitment appeared in biological process as shown in table 1. Transcription regulatory region, DNA binding regulatory region, nucleic acid binding, double-stranded DNA binding, DNA-binding transcription factor activity showed the highest significance in molecular functions as shown in table 2. Among cell components, nuclear transcription factor complex, chromatin, RNA polymerase II

transcription factor complex were represented as shown in table 3. These data indicate that interacting proteins sharing gene annotations are involved in various aspects of stem cell biology and that their coordinated action is necessary for efficient TF binding, as well as enhancing or modulating functions. Many of these proteins are known to be involved in the pluripotency maintenance, repression of which might lead to differentiation. Lineage commitment is also caused by the regulation of a few proteins.

Table 1: Enriched biological processes as defined by gene ontology

Biological Process	False discovery rate	pValue
stem cell population maintenance	3.32E-19	1.34E-22
maintenance of cell number	4.10E-19	1.65E-22
somatic stem cell population maintenance	5.82E-18	2.35E-21
positive regulation of transcription by RNA polymerase II	1.90E-13	7.64E-17
positive regulation of macromolecule biosynthetic process	1.16E-12	4.69E-16
embryonic morphogenesis	2.92E-12	1.18E-15
positive regulation of gene expression	2.94E-12	1.19E-15
positive regulation of cellular biosynthetic process	3.76E-12	1.52E-15
response to growth factor	4.44E-12	1.79E-15
positive regulation of nitrogen compound metabolic process	4.50E-12	1.81E-15
positive regulation of biosynthetic process	5.05E-12	2.04E-15
stem cell differentiation	7.77E-12	3.13E-15
positive regulation of nucleic acid-templated transcription	2.04E-11	8.22E-15
positive regulation of transcription, DNA-templated	2.04E-11	8.22E-15
positive regulation of RNA biosynthetic process	2.40E-11	9.69E-15
embryo development	3.48E-11	1.40E-14
positive regulation of RNA metabolic process	3.65E-11	1.47E-14
positive regulation of nucleo base-containing compound metabolic process	2.57E-10	1.04E-13
cell fate commitment	3.26E-10	1.31E-13
cell fate commitment involved in formation of primary germ layer	3.44E-10	1.39E-13

regulation of cell development	4.37E-10	1.76E-13
regulation of transcription by RNA polymerase II	4.80E-10	1.94E-13
tissue morphogenesis	1.46E-09	5.89E-13
gastrulation	1.65E-09	6.67E-13

Table 2: Enriched molecular functions as defined by gene ontology

Molecular functions	False discovery rate	pValue
transcription regulatory region DNA binding	8.51E-09	9.99E-11
regulatory region nucleic acid binding	8.51E-09	1.07E-10
transcriptional activator activity, RNA polymerase II proximal promoter sequence-specific DNA binding	2.84E-08	5.36E-10
double-stranded DNA binding	9.47E-08	2.38E-09
DNA-binding transcription factor activity	1.13E-07	4.31E-09
transcription factor binding	1.13E-07	4.65E-09
transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific DNA binding	1.13E-07	4.99E-09
transcription factor activity, RNA polymerase II proximal promoter sequence-specific DNA binding	1.16E-07	5.82E-09
transcription regulatory region sequence-specific DNA binding	3.55E-07	2.01E-08
sequence-specific double-stranded DNA binding	4.22E-07	2.81E-08
sequence-specific DNA binding	4.22E-07	2.92E-08
chromatin binding	8.02E-07	6.06E-08
I-SMAD binding	8.45E-07	6.91E-08
miRNA binding	3.22E-06	2.84E-07
proximal promoter sequence-specific DNA binding	3.90E-06	3.86E-07
RNA polymerase II transcription factor activity, sequence-specific DNA binding	3.90E-06	3.92E-07
R-SMAD binding	7.87E-06	8.42E-07
RNA polymerase II transcription factor binding	1.74E-05	1.97E-06
protein dimerization activity	2.07E-05	2.48E-06
protein heterodimerization activity	2.51E-05	3.16E-06

transcription factor activity, transcription factor binding	3.48E-05	4.63E-06
transcription factor activity, protein binding	3.48E-05	4.81E-06
RNA polymerase II proximal promoter sequence-specific DNA binding	5.91E-05	8.55E-06
protein-containing complex binding	7.78E-05	1.20E-05
phosphatase binding	7.78E-05	1.22E-05
repressing transcription factor binding	8.55E-05	1.40E-05
transforming growth factor beta receptor, cytoplasmic mediator activity	1.59E-04	2.70E-05
SMAD binding	1.68E-04	2.97E-05
chromatin DNA binding	2.19E-04	3.99E-05

Table 3: Cellular components as defined by gene ontology

Cell components	False discovery rate	pValue
nuclear transcription factor complex	4.63E-10	4.45E-12
transcription factor complex	1.06E-08	1.02E-10
chromatin	9.81E-08	9.43E-10
nuclear chromatin	2.29E-07	2.20E-09
nuclear chromosome part	6.90E-06	6.64E-08
chromosomal part	6.93E-06	6.66E-08
nuclear chromosome	1.11E-05	1.06E-07
chromosome	1.68E-05	1.62E-07
RNA polymerase II transcription factor complex	1.18E-04	1.13E-06
nuclear euchromatin	1.64E-04	1.58E-06
activin responsive factor complex	1.80E-04	1.73E-06
euchromatin	3.98E-04	3.83E-06
SMAD protein complex	1.68E-03	1.61E-05

Based on the interesting differences in the OSKM interactome studies, phylogenetic analysis was performed to understand the sequence similarity of human and mouse reprogramming factors to that of other domestic animals as represented in figures 3-6.

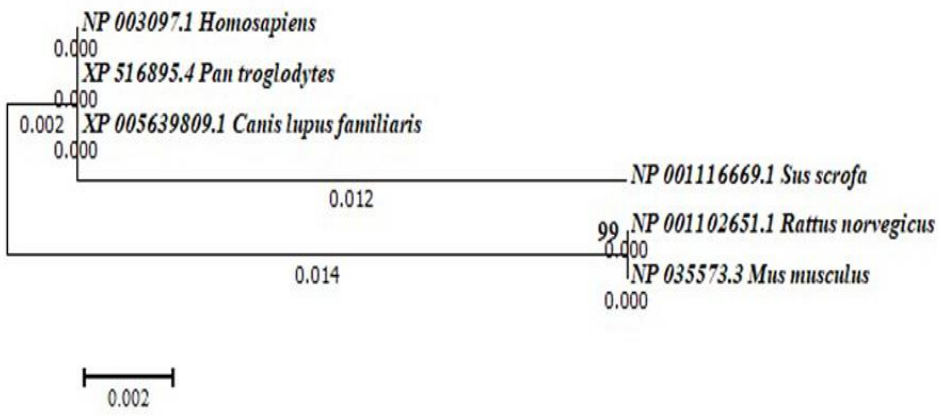


Figure 3: Phylogenetic tree of Oct4 protein in six species

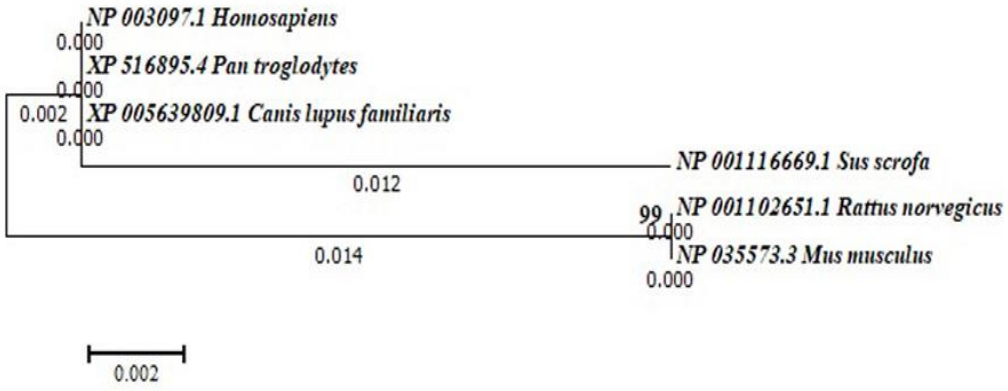


Figure 4: Phylogenetic tree of Sox2 protein in six species

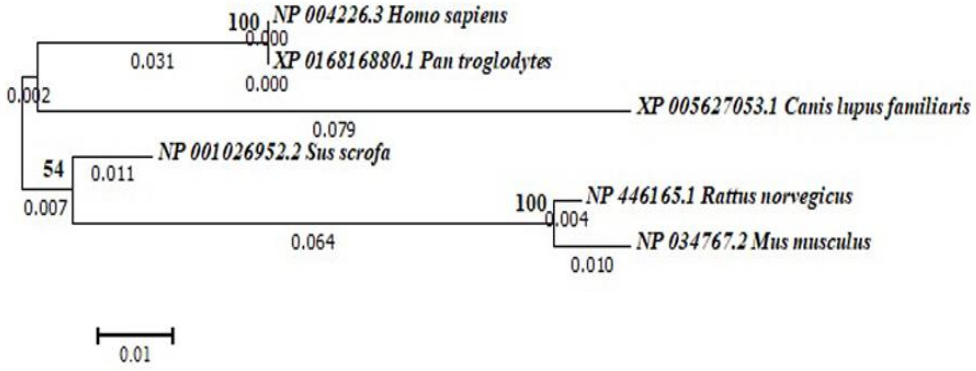


Figure 5: Phylogenetic tree of Klf4 protein in six species

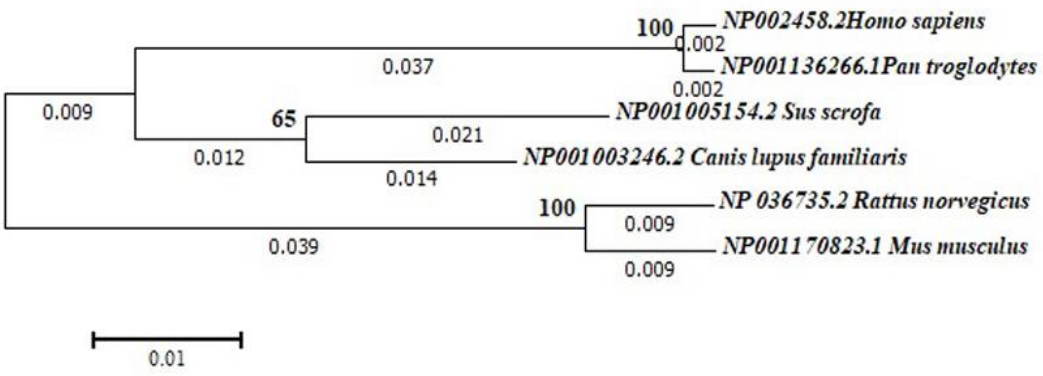


Figure 6: Phylogenetic tree of cMYC protein in six species

The homology of human OSKM with that of other species is given in table 4.

Table 4: Homology of human OSKM with that of other species

Species	Protein	Accession No.	Identity (%)	Protein	Accession No.	Identity (%)	Protein	Accession No.	Identity (%)	Protein	Accession No.	Identity (%)
Homo sapiens	OCT4	NP002692.2	100%	SOX2	NP003097.1	100%	KLF4	NP004226.3	100%	CMYC	NP002458.2	100%
Pan troglodytes		XP528230.1	96%		XP516895.4	98%		XP016816880.1	93%		NP001136266.1	99%
Canis lupus familiaris		XP38830.1	91%		XP005639809.1	98%		XP005627053.1	79%		NP001003246.2	92%
Sus scrofa		NP001106531.1	95%		NP001116669.1	98%		NP001026952.2	94%		NP001005154.2	93%
Rattus Norwegicus		NP001009178.1	83%		NP001102651.1	98%		NP446165.1	90%		NP036735.2	89%
Mus Musculus		NP038661.2	83%		NP035573.3	98%		NP034767.2	90%		NP001170823.1	91%

Apart from chimpanzee which showed 96% identity, pig and dog showed high sequence similarity for the Oct4 protein with the human while mouse and rat showed 83% similarity. The identity of OSKM protein sequences of each species with that of human is given in table 4. Sox2 was the most conserved among all species. Human, chimpanzee and dog were grouped together followed by mouse and rat. This was similar to the previous analysis[7,12] reports. Klf4 and cMyc sequences showed higher sequence similarity among animals. The phylogenetic relatedness between the animals can also influence the success of stem cell transplantations in appropriate models[13]. We observed a tight protein network around the Oct4 protein in the interactome analysis. Among the OSKM factors, many studies had repeatedly shown that Oct4 is the master regulator for efficient generation of iPS cells[14].Oct4 mutants exhibited less reprogramming ability though they could maintain the self-renewal of ESCs [15].Sox2 is co-localized with Oct4 and interacts with other

transcription factors to mediate pluripotency[16]. The interactions of Oct4 and Sox2 were evident in our interactome studies as well. While Oct4, Sox2 and Klf4 had many shared targets, c-Myc interactions were highly distinct from the other three pluripotency factors, suggesting that they have different downstream targets. Interactomes of all species showed Sall4 as an important component. Sall4 has been found to be essential in pluripotency maintenance in both mouse and human ESCs, knockdown of which can cause the downregulation of Oct4 and Sox2 [17]. Various interactome pathways are involved in the pluripotency maintenance of ESCs and iPSCs such as TGFβ, BMP/SMAD, MEK/ERK, FGF pathways, LIF/ Signal Transducer and Activator 3 (STAT3). There are two pluripotency states- naïve, which is ICM-like and prime, which is epiblast-like [18]. Mouse ESCs express the naive state and human ESCs express the prime state. LIF activates the JAK-STAT pathway which is involved in maintaining the naïve state. bFGF activates the PI3K/AKT pathway which maintains the prime state. Accordingly, two growth factors, leukemia inhibitory factor (LIF) and basic fibroblast growth factor (bFGF) are added to the culture media for pluripotency maintenance of domestic animal iPSCs Though there are many reports on iPSC derivation from various cell sources and types in mice and humans, lesser reports are available in iPSC derivation in domestic animals. Understanding key partners and their roles in pluripotency helps in developing new approaches for iPSC derivation and characterization in domestic animals. This can be further confirmed by evidence-based studies like targeted functional genomics or protein-protein interaction experiments. Using specific reprogramming factors and adding the most appropriate cytokines along with the accurate designing of culture media conditions can help to overcome the issues related to stem cell models in higher animals. Our article suggests that reprogramming pathways in higher animals like dog and pig is more similar to human than to mice. This is further validated by the similarity search and phylogenetic analysis. OSKM reprogramming factors of these animals shows higher similarity to human OSKM, which supports their use as better animal models for stem cell therapy.

4. CONCLUSION

Mouse is a popular animal model in stem cell biology and regenerative medicine but species differences have to be carefully examined before extrapolating the findings in mice to that in higher animals. Development of large animal models is necessary to fill the gap in pluripotency pathway interpretations and to assist the transfer of iPSCs-based therapies from mice to the arena of biotechnology and regenerative medicine. Based on the results obtained from our studies, pig and dog could prove more effective as animal models. More detailed functional studies can lead to the choice of appropriate models for stem cell therapy and also, as disease models.

ACKNOWLEDGEMENT

This work was supported by Department of Science and Technology, Government of India under Women Scientists Scheme (WOS-A/LS-1096/2014).

CONFLICT OF INTEREST

The authors declare no conflict of interests.

REFERENCES

1. Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KHS. Viable offspring derived from fetal and adult mammalian cells. Nature. 1997;385:810.
2. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. Cell. 2006;126(4):663–76.
3. Mao J, Zhang Q, Deng W, Wang H, Liu K, Fu H, et al. Epigenetic Modifiers Facilitate Induction and Pluripotency of Porcine iPSCs. Stem Cell Reports. 2017;8(1):11–20.
4. Menon DV, Patel D, Joshi CG, Kumar A. The road less travelled: the efficacy of Canine Pluripotent Stem Cells. Experimental Cell Research.2019.
5. Shimada H, Nakada A, Hashimoto Y, Shigeno K, Shionoya Y, Nakamura T. Generation of canine induced pluripotent stem cells by retroviral transduction and chemical inhibitors. Molecular Reproduction and Development.2009;77(1):2–2.
6. Ezashi T, Telugu BPVL, Alexenko AP, Sachdev S, Sinha S, Roberts RM. Derivation of induced pluripotent stem cells from pig somatic cells. Proc Natl Acad Sci USA. 2009;106(27):10993–8.
7. Ogorevc J, Orehek S, Dovč P. Cellular reprogramming in farm animals: an overview of iPSC generation in the mammalian farm animal species. Journal of Animal Science and Biotechnology. 2016;7(1).
8. Lee AS, Xu D, Plews JR, Nguyen PK, Nag D, Lyons JK, et al. Preclinical Derivation and Imaging of Autologously Transplanted Canine Induced Pluripotent Stem Cells. Journal of Biological Chemistry. 2011;286(37):32697–704.
9. Zhou L, Wang W, Liu Y, Fernandez de Castro J, Ezashi T, Telugu BPVL, et al. Differentiation of induced pluripotent stem cells of swine into rod photoreceptors and their integration into the retina. Stem Cells. 2011;29(6):972–80.
10. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.
11. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Research. 2009:W305–11.
12. Ma Y, Zhang X, Ma H, Ren Y, Sun Y, Wang Q, et al. Bioinformatic analysis of the four transcription factors used to induce pluripotent stem cells. Cytotechnology. 2014;66(6):967–78.
13. Cibelli J, Emborg ME, Prockop DJ, Roberts M, Schatten G, Rao M, et al. Strategies for Improving Animal Models for Regenerative Medicine. Cell Stem Cell. 2013;12(3):271–4.
14. Hammachi F, Morrison GM, Sharov AA, Livigni A, Narayan S, Papapetrou EP, et al. Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency. Cell Rep. 2012;1(2):99–109.

15. Oka M, Moriyama T, Asally M, Kawakami K, Yoneda Y. Differential Role for Transcription Factor Oct4 Nucleocytoplasmic Dynamics in Somatic Cell Reprogramming and Self-renewal of Embryonic Stem Cells. Journal of Biological Chemistry. 2013;288(21):15085–97.
16. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, et al. Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell.;168(3):442-459.e20.
17. Yang F, Yao Y, Jiang Y, Lu L, Ma Y, Dai W. Sumoylation is important for stability, subcellular localization, and transcriptional activity of SALL4, an essential stem cell transcription factor. J Biol Chem. 2012;287(46):38600–8.
18. Nichols J, Smith A. Naive and Primed Pluripotent States. Cell Stem Cell. 2009;4(6):487–92.