Article

Reconstruction, visualization and explorative analysis of human pluripotency network

Priyanka Narad[1], **Kailash C.Upadhyaya**[2], **Anup Som**[3]
[1]Amity Institute of Biotechnology, Amity University, Uttar Pradesh, India
[2]Amity Institute of Molecular Biology & Genomics, Amity University, Uttar Pradesh, India
[3]Centre of Bioinformatics, Institute of Interdisciplinary Studies, University of Allahabad, Allahabad, India
Email: som.anup@gmail.com

Received 8 May 2017; *Accepted 15 June 2017*; *Published 1 September 2017*

Abstract
Identification of genes/proteins involved in pluripotency and their inter-relationships is important for understanding the induction/loss and maintenance of pluripotency. With the availability of large volume of data on interaction/regulation of pluripotency scattered across a large number of biological databases and hundreds of scientific journals, it is required a systematic integration of data which will create a complete view of pluripotency network. Describing and interpreting such a network of interaction and regulation (i.e., stimulation and inhibition) links are essential tasks of computational biology, an important first step in systems-level understanding of the underlying mechanisms of pluripotency. To address this, we have assembled a network of 166 molecular interactions, stimulations and inhibitions, based on a collection of research data from 147 publications, involving 122 human genes/proteins, all in a standard electronic format, enabling analyses by readily available software such as Cytoscape and its Apps (formerly called "Plugins"). The network includes the core circuit of OCT4 (POU5F1), SOX2 and NANOG, its periphery (such as STAT3, KLF4, UTF1, ZIC3, and c-MYC), connections to upstream signaling pathways (such as ACTIVIN, WNT, FGF, and BMP), and epigenetic regulators (such as L1TD1, LSD1 and PRC2). We describe the general properties of the network and compare it with other literature-based networks. Gene Ontology (GO) analysis is being performed to find out the over-represented GO terms in the network. We use several expression datasets to condense the network to a set of network links that identify the key players (genes/proteins) and the pathways involved in transition from one state of pluripotency to other state (i.e., native to primed state, primed to non-pluripotent state and pluripotent to non-pluripotent state).

Keywords human pluripotency network; network layout; network measures; gene enrichment analysis; gene expression data; embryonic stem cell; naive and primed states.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Embryonic stem cells (ESC) are important due to the following two properties: One is the self-renewal, which is the ability to go through numerous cycles of cell division while maintaining the undifferentiated

state (Thomson et al., 1998; Jain et al., 2011). The other one is pluripotency, which is the capacity to differentiate into specialized cell types (Reubinof et al., 2000). The ability to give rise to any mature cell type makes the pluripotent stem cells as a potentially ideal cell candidate for basic science as well as for regenerative medicine (De Los Angeles et al., 2015). However successful application of pluripotent stem cells is impeded due to lack of complete knowledge of underlining mechanisms of pluripotency. The molecular mechanism behind this process is a complex interplay between the genes/proteins that includes transcription factors (TFs) with epigenetic regulators and signaling pathways. These mechanisms involve protein–protein interactions, which participate in key biological processes in cells. They are supplemented by DNA–protein interactions describing gene regulation by the control of transcription (called gene-regulatory network) (Kininmonth et al., 2012).

With the availability of large volume of data on interaction/regulation of pluripotency scattered across a large number of biological databases and hundreds of scientific journals, it is required a systematic integration of data which will create a complete view of pluripotency network and allow researchers to know the up to date knowledge on pluripotency. Thus, in this paper, we propose to assemble a new map of the interaction/regulation network of human pluripotency that incorporates genes and proteins, and is based on manual literature curation. Describing and interpreting such a network of interaction and regulation (i.e., stimulation and inhibition) links are essential tasks of computational biology, an important first step in systems-level understanding of the underlying mechanisms of pluripotency.

Furthermore, current high throughput genomics technologies generate large quantities of high dimensional data. The challenge for integrative bioinformatics is to capture, integrate, and analyze these data in a consistent way to provide new and deeper insights into complex biological systems. For example, large amounts of differential gene expression data on pluripotency are becoming available, describing a variety of experimental conditions such as gene knockout and knockdown, and the effect of interventions. Combining differential gene expression data with the network of interaction/regulation should be used to understand the role of epigenetic, signaling, and transcriptional mechanisms in pluripotency, and screen for putative mechanistic relationships revealed by the differential gene expression data. Thus, we assembled an electronic network (interaction/regulation) depicting the cellular and molecular mechanisms encompassing pluripotency, called hPluriNet. Nodes and edges are main components of the network. Nodes represent genes/proteins and edges represent stimulation, inhibition, or interaction links. Continuing and updating the network are important; therefore we will continue addition of nodes/edges to the network. In this paper, we describe how the network is assembled, its layout and properties, annotation of associated biological information, and its usefulness in data analysis.

Recently, pluripotency networks have been constructed based on manual literature curation as well as in-silico approach. Particularly, the *PluriNetWork* (Som et al., 2010) which is the largest literature-based network currently available, describing pluripotency in mouse, a network of 274 genes/proteins and 574 molecular interactions, based on a collection of research data from 177 publications. Another literature-based pluripotency network for mouse was assembled by Xu et al. that contains 134 genes/proteins and 220 interactions (Xu et al., 2010). Other types of pluripotency network are based on machine learning, using high-throughput interaction and gene expression data as input. In particular, Muller et al. developed the PluriNet (Muller et al., 2008), an undirected network describing stem cell regulation in human. The PluriNet contained 299 genes/proteins. In another effort, Newman & Cooper (2010) generated the PluriUp, a network describes pluripotency in human based on their AutoSOME machine learning (clustering) approach. The PluriUp is a cluster of 3421 genes upregulated in pluripotent stem cells which indicates that the PluriUp was inflated by the genes those are not directly involved in the induction, maintenance or loss of pluripotency.

To overcome this problem, Som et al. (2012) derived a human pluripotency network, containing 136 genes/proteins and 196 interactions by considering mouse literature-based network (i.e., *PluriNetWork*) as a model network. However, there is no such literature-based network on human pluripotency that could be used as a model network for studying early development, and scientists are expected to benefit from the easy access to a human network of pluripotency players and mechanisms, which will help them, make sense of high-throughput data.

2 Material and Methods

2.1 Network assembly

We created the *hPluriNet* by manually adding the nodes (genes/proteins) and the edges (activations, interactions and inhibitions) that are reported in the literature showing direct mechanisms involved in the induction and maintenance of pluripotency in human model system (Fig. 1). The source file (i.e., Cytoscape file) of Fig. 1 is given in the supplementary data (S4). The list of curated literatures, nodes and edges information retrieved from the literatures, and type of edges (i.e., activations, inhibitions and interactions) are described in Supplementary data (S1). We restricted the criteria of inclusion of nodes and the edges: nodes and links added must be directly involved in induction and maintenance of pluripotency and in human model system. This inclusion criterion ensures the quality of the network and also prevents it from unnecessary expansion. From June 2013 we performed a fortnight search of any new publication reporting any relevance to pluripotency maintenance in human and relevant data were incorporated to expand the network till date, resulting in a model human pluripotency network of 166 edges and 122 nodes all in a standard electronic format in the Cytoscape editor (Shannon et al., 2003). Beside the network assembling, various useful biological information have been annotated in the network such as gene/protein name, Ensamble ID, UniGene ID, type of interaction, literature reference, and pubmed ID (Fig. 2a & 2b).

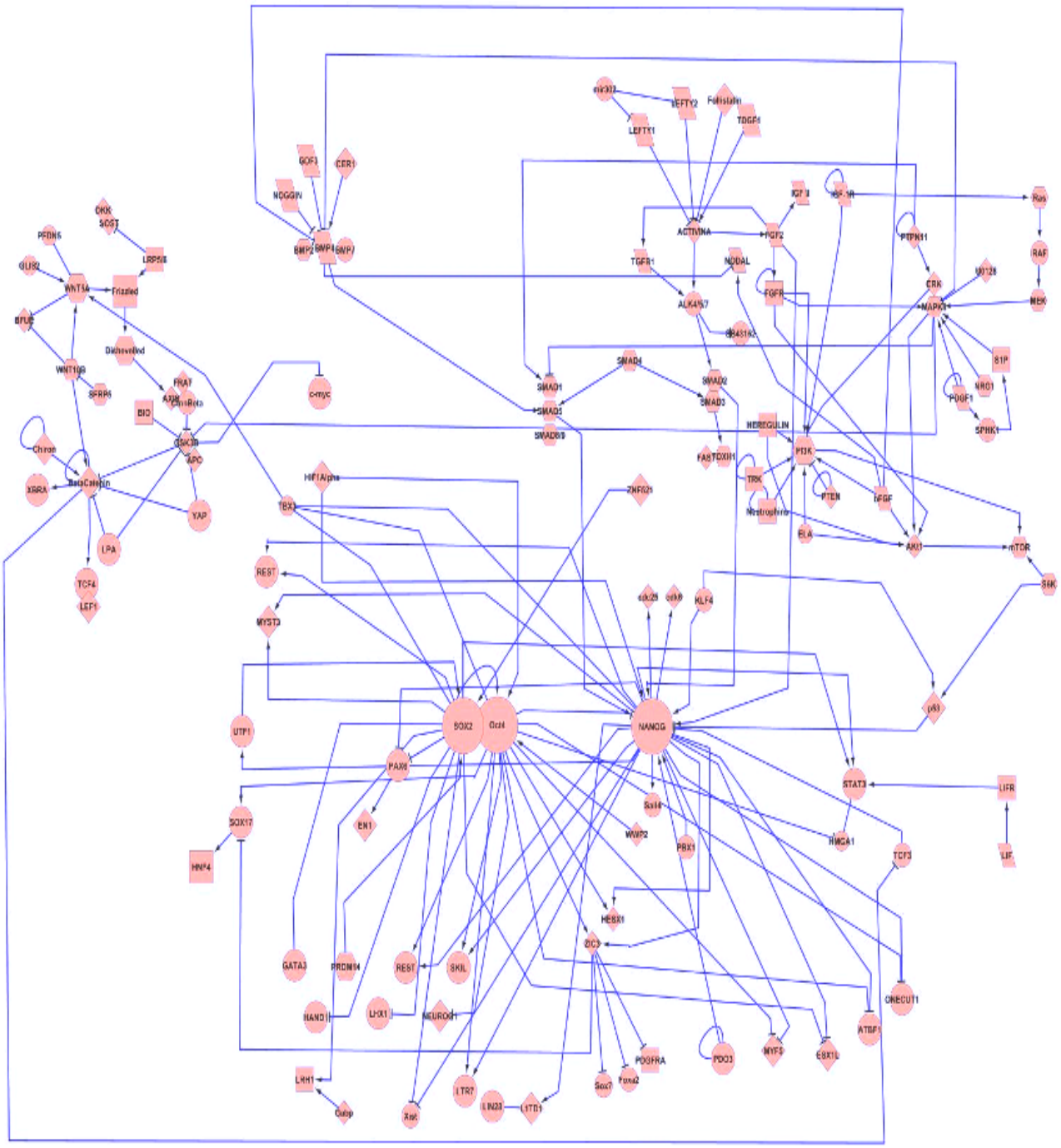

Fig. 1 Electronic Representation of *hPluriNet*. The network consists of 122 nodes and 166 edges. The overall network looks like an electronic circuit clearly depicting the nodes (Genes/proteins) and the edges (activation/inhibition/interaction). The network includes the core circuit of Oct4 (Pou5f1), Sox2 and Nanog, its periphery (such as Stat3, Klf4, Utf1, Zic3, and c-Myc), connections to upstream signaling pathways (such as Activin, WNT, FGF, and BMP), and epigenetic regulators (such as L1TD1, LSD1 and PRC2). Different node representations are given in the network: genes/enzymes (Diamond), transcription factors (Ellipse), signaling components (Hexagon), growth factors (Parallelogram), and receptors (Rectangle). Cytoscape file of this Figure is given in the supplementary data (S4).

2.2 Network layout and functionality

One of the most important aspects of any biological network visualization is layout style. There are several popular layout styles used in biological networks visualization such as circular, hierarchical, force-directed, and simulated annealing. We present our network in a circuit-like representation that will allow easy human interpretation of large and complex biological networks. The network layout was produced by manually adding nodes and edges in an electronic circuit representation using the Cytoscape editor. The layout includes nodes and edges where the nodes represent genes or gene products (i.e. proteins). Three types of mechanisms were considered for the edges: (i) activation is denoted by an arrow, (i) inhibition is denoted by a T-bar, and (iii) interaction is denoted by a straight line. The network consists of five components – diamond shape represents enzymes, ellipse shape represents transcription factors (TFs), hexagon shape represents signaling component, parallelogram shape represents growth factors, and rectangle shape represents receptors. Each gene/protein is defined by a set of attributes described in Fig. 2a. A number of different attributes from different databases have been annotated in the *hPluriNet*. These attributes include Ensembl ID, symbol of gene, full name and type of molecule. The edge attributes are described in Fig. 2b. It includes start node, end node, type of interaction, and the reference of the publication from which the information is extracted.

Data Panel

ID	node.label	Full Name	TYPE	ENSEMBL ID
node7_101	FOXH1	Forkhead box protein H	signalling component	ENSG00000160973
node25_...	GATA3	GATA binding protein 3	transcription factor	ENSG00000107485
node29_...	ESX1L	X chromosome-linked human homeobox gene	protein	ENSG00000123576
node0_096	XBRA	Chordin	transcription factor	ENSG00000090539
node25	CER1	Cerberus 1, DAN family BMP antagonist	signalling component	ENSG00000147869
node6	WWP2	WW domain containing E3 ubiquitin protein ligase 2	enzyme/ protein	ENSG00000198373
node1_356	RAF	Raf Family Kinases	transcription factor	ENSG00000078061

Fig. 2A Node attributes of the selected edges in the *hPluriNet*. It represents the node annotation panel in Cytoscape. The attributes included for annotation; node label which denotes the symbol of the gene, full name, ensemble gene ID, and type of molecule.

Data Panel

ID	START N...	END NODE	TYPE	REFERENCE
node3_9...	PTPN11	Smad1/5/8	Activation	Protein post-translational modifications and regulation of pluripotency in human stem cells
node16_...	GSK3B	c-myc	Inhibition	Stephen Dalton, Signaling networks in human pluripotent stem cells, Current Opinion in Cell Biology, Volume 25, Issue 2, April 2013, ...
node1 (Di...	OCT4/SO...	HAND1	Inhibition	Boyer et al. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell . 2005 September 23; 122(6): 947-956. doi...
node1_9...	SMAD4	SMAD2	Activation	Angelique Scherch et al. Distinguishing between Mouse and Human Pluripotent Stem Cell Regulation: The Best laid plans of Mice an...
node17_...	BetaCate...	XBRA	Activation	Singh et al. Signalling Network Crosstalk in Human Pluripotent Cells: A Smad2/3 regulated switch that controls balance between self r...
node21 (...	ActivinA	Alk4/5/7	Activation	Ludovic Vallier*, Morgan Alexander and Roger A. Pedersen.Activin/Nodal and FGF pathways cooperate to maintain pluripotency of hum...
node9 (Di...	ZIC3	PDGFRA	Inhibition	Mol Biol Cell. 2007 Apr;18(4):1348-58. Epub 2007 Jan 31. Zic3 is required for maintenance of pluripotency in embryonic stem cells.
node1_0...	Sphk1	S1P	Activation	Avery et al. Stem Cells & Development.15:729-740 (2006).
node0_4...	BIO	GSK3B	Inhibition	Avery et al. Stem Cells & Development.15:729-740 (2006).
node12 (...	LRP5/6	Frizzled	Activation	MASUKO KATOH and MASARU KATOH.CER1 is a common target of WNT and NODAL signaling pathways in human embryonic stem ...
node2 (Di...	NANOG	ONECUT1	Inhibition	Boyer et al. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell . 2005 September 23; 122(6): 947-956. doi...
node0 (Di...	GATA3	OCT4/SO...	Activation	Montserrat et al., Reprogramming of Human Fibroblasts to Pluripotency with Lineage Specifiers, Cell Stem Cell(2013), http://dx.doi.org...
node23_...	bFGF	BMP4	Inhibtion	Xu el al. basic FGF & Supression of BMP signalling sustain undifferentiated proliferation of Human ES Cells.
node23_...	bFGF	NODAL	Activation	Saunders et al. Nanog function in pluripotency and reprogramming. Stem Cells 2013;31:1227-1236.

Node Attribute Browser | Edge Attribute Browser | Network Attribute Browser | VistaClara

Fig. 2B Edge attributes of the selected edges in the *hPluriNet*. It represents the edge annotation panel in Cytoscape. The attributes annotated are the start node, end node, type of interaction such as activation, inhibition or interaction, and reference of the paper from which the edge has been derived.

2.3 Differential gene expression data

Combining differential gene expression data with the network of interaction/regulation is useful for identifying the pathways are involved in transition from one state of pluripotency to other state (i.e., Naive to Primed, Primed to Non-pluripotent, and Pluripotent to Non-pluripotent). We selected and preprocessed several gene expression datasets to condense the network to a set of network links to understand and highlight the most important start-ups and shut-downs in the process of maintenance of pluripotency in human. Expression data were retrieved from Gene Expression Omnibus (Edgar et al., 2002) and pre-processed using the RMA algorithm (Wu and Irizarry, 2004). These data were then integrated into the network and condensed using ExprEssence app (formerly called "plugin"), which searches for differentially altered links in a given network using multiple set of expression data (Warsow et al. 2010). For the purpose of identifying putative mechanistic relationships and validating existing knowledge for Naive and Primed state in hESC, we selected three datasets namely GSE59430 (Theunissen et al. 2014), GSE21222 (Hanna et al., 2010) and GSE46872 (Gafni et al., 2013). For the non-pluripotent state the datasets namely GSE7177 (Marteyn et al., 2011), GSE7178 (Marteyn et al., 2011), and GSE22392 (Chin et al., 2010) were selected. The detailed description of the datasets with the sample and replicates is provided in S2 (supplementary data). For the purpose of studying the developmental timeline, we selected the gene expression dataset GSE29397 (Vassena et al., 2011). This sample dataset consists of expression profiles of developmental changes starting from as early as the oocyte stage to the ESC stage provided in S3 (supplementary data).

3 Results and Discussion

3.1 Network topology and its measures

Network measure offers a quantifiable description of a network that characterizes various biological systems (Zhang, 2012). Here we measured the network properties that allow us to characterize the network topology and compare *hPluriNet* with other biological networks. Topological properties of the *hPluriNet* were analyzed using Network Analyzer, which is a Cytoscape app developed by Assenov et al. (2008). The network parameters were computed by treating *hPluriNet* as an undirected network. The *hPluriNet* consists of 122 nodes and 166 edges having the degree exponent (γ) of 1.36, which states that the *hPluriNet* is a scale-free network (Barabási and Albert, 1999; Hakes et al., 2008). The other network measures are: network diameter is 10, characteristic path length is 4.29, clustering coefficient is 0.123, and connected component is 9. We compared the physical properties of the *hPluriNet* with other biological networks. Particularly, the PluriNetWork is a similar kind of literature curated network that describes pluripotency in mouse. The degree exponent of the PluriNetWork is 1.14, network diameter is 10, characteristic path length is 3.25, and clustering coefficient is 0.256, and connected component is 8. The above results show that both the networks are similar in nature. However, interestingly, the variation of the parameters between the networks clearly indicates that pluripotency is species-specific that was established by the experimental evidences (Rao, 2004; Ginis et al., 2004; Weinberger et al., 2016). Further we compared the parameters of *hPluriNet* with other two literature curated networks namely; protein-protein interaction (PPI) network of Curcumin (Gan et al., 2015), PPI network of human T lymphocyte (Orr et al., 2012). The rationale behind considering non-pluripotency network is to find out whether *hPluriNet* has any specific network features or characteristics. The computed parameters are given in Table 1 that reveals that the *hPluriNet* is a scale-free, biological and species specific network.

Table 1 Comparison of network properties of *hPluriNet* to other biological networks namely mouse PluriNetWork, curcumin protein-protein interaction (PPI) network and human T lymphocytes PPI network.

Network Properties	*hPluriNet*	PluriNetWork	Protein interaction network of Curcumin	Protein interaction network of human T lymphocytes
Node Degree exponent	-1.361	-1.14	-1.39	-0.383
Clustering Coefficient	0.123	0.256	0.641	0.5
Connected Components	9	8	1	7
Network Diameter	10	10	11	7
Network Centralization	0.231	0.418	0.165	0.299
Characteristic Path length	4.299	3.250	4.394	2.39
Network Heterogeneity	1.379	2.352	0.995	0.981

3.2 Gene ontology analysis

We performed gene enrichment analysis of the genes involved in *hPluriNet* to find the gene ontology (GO) terms over-represented in the biological process (BP), molecular function (MF) and cellular compartment (CC). The GO analysis was conducted for all 122 genes. We used BINGO (Maere et al., 2005), a cytoscape app for gene ontology analysis. For the reference annotations we used "GO slim generic" gene ontology annotation, which is a set of high-level GO terms (Ashburner et al., 2000). Fig. 3 shows the GO representation of BPs, MFs and CCs of the genes over-represented in *hPluriNet*. As observed from the diagram, terms like "transcription regulation activity", "transcription factor activity", "chromatin binding", "embryonic development", "cell growth", "reproduction", "cell proliferation", and "protein modification process" are significantly over-represented. Over-representation of the above GO terms is expected and it can be taken as evidence that the proteins/genes of the network perform molecular functions and biological processes related to pluripotency.

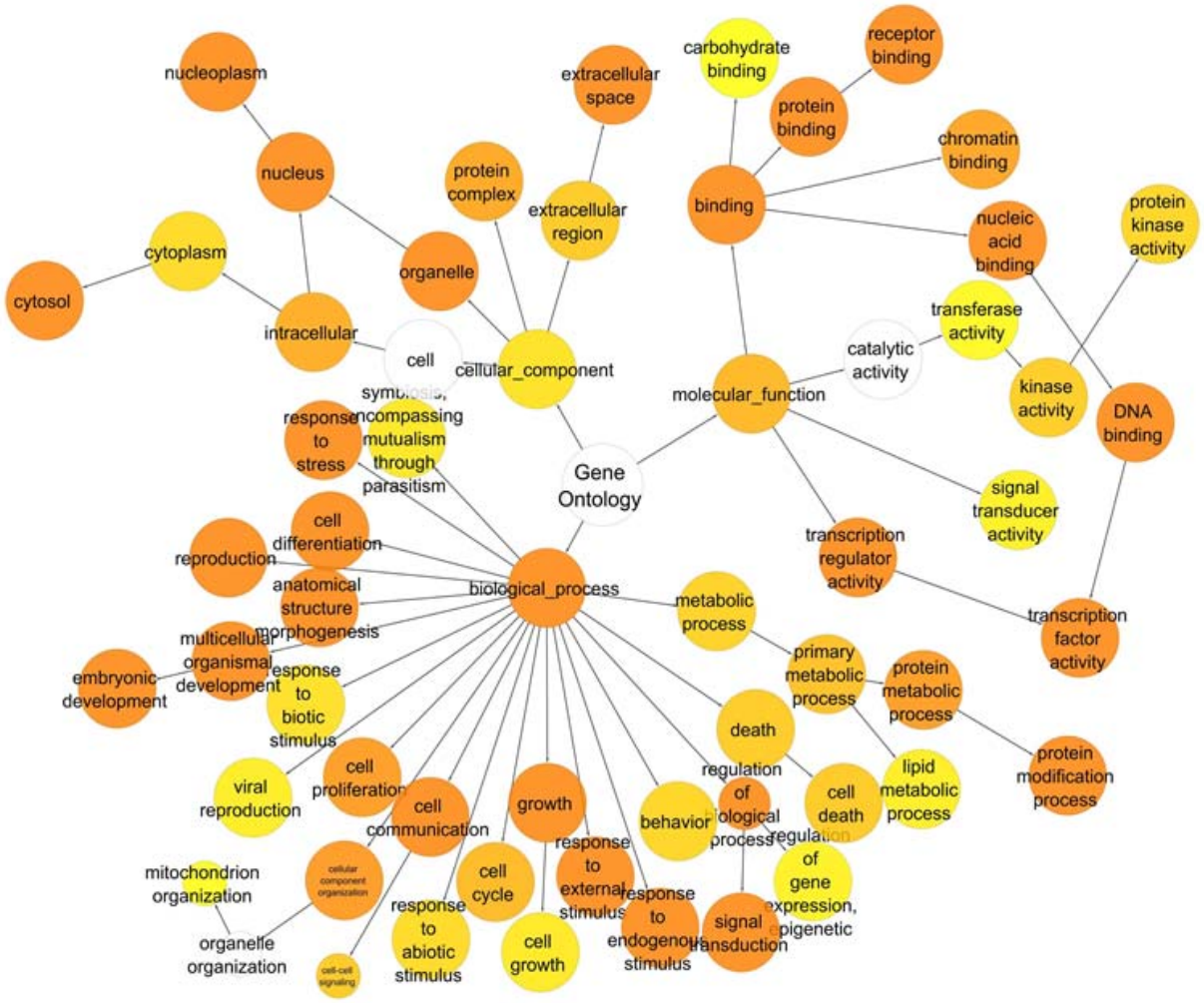

Fig. 3 Enrichment analysis of the *hPluriNet* genes at a significance level p = 0.05, using GO terms from GO Slim. The Enrichment analysis (through BiNGO graph) visualizes the GO categories that were found significantly over-represented in the context of the GO hierarchy. According to BINGO documentation, the size (area) of the nodes is proportional to the number of genes in our gene set which are annotated to that node. The color of the node represents the (corrected) p-value. White nodes are not significantly over-represented, the other ones are with a color scale ranging from yellow (p-value = significance level, here 0.05) to dark orange (p-value = 5 orders of magnitude smaller than significance level, here 0.0000005). The color saturates at dark orange for p-values which are more than 5 orders of magnitude smaller than the chosen significance level.

3.3 Visualization and explorative analysis of pluripotency-related data

In this section, we discuss the applications of *hPluriNet* using microarray data pertaining to hESC as described in the methodology section. We use ExprEssence app for the network condensation to identify the most differentially altered links when transition takes place from one state of pluripotency to the other. In the condensed network, top 10% differential altered links were considered for the analysis purpose.

3.3.1 Transition from naive to primed state

Stem cell states are dynamic and classified into native or ground state and primed state based on their growth characteristics and potential to give rise to all somatic lineages (Weinberger et al., 2016). In this analysis, we investigated to find out the genes/proteins and the pathways those are most crucial for the transition from native state to primed state. We integrated gene expression data in the *hPluriNet* and then condensed the

network using ExprEssence app. We observed that within the core pluripotency circuit, shut-down of activation/inhibition links around NANOG such as activation of important transcription factors REST and UTF1, and inhibition of lineage associated genes such as ATBF1 and MYF5 as the transition occurs from the native to the primed state (Fig.4a). Another interesting finding is that SOX2 is under-expressed in the native and primed states; hence it may not be imperative in the maintenance of pluripotency. We have further studied the role of SOX2 and has been discussed in the next paragraph. The TFs such as KLF4 and UTF1 have been reported to be instrumental in the maintenance of pluripotency in human (Nishimoto et al., 2001). Both these TFs show an upregulation in the native state, whereas the transition towards the primed state occurs, the expression intensity decreases. Important start-ups are centered on BMP4. BMP4 has a repressive effect on the maintenance of pluripotency in human (Sun et al., 2006); hence these start-ups indicate the loss of pluripotency when development takes place from native state to the primed state (Fig. 4a). CER1 is an important for activation of BMP4 signaling pathway (Katoh and Katoh, 2006). CER1 is significantly downregulated in the native state and is upregulated in the primed state that causes the activation of BMP4 in the primed state. Role of signaling pathways in hESC is well known and highlighted in the condensed network such as bFGF, mTOR, ERK, and RAS signaling are activated and over-expressed in both the states (Chen et al., 2012).

Maintenance of pluripotency in human embryonic stem cell is thought to be regulated by several key TFs including OCT4, NANOG, and SOX2 (Boyer et al. 2005). Although the functions of OCT4 and NANOG in human ESCs are fully characterized (Hyslop et al., 2005; Zaehres et al., 2005), but the role of SOX2 in human pluripotency is not well defined (Fong et al. 2008). To provide some elucidation towards this, we observed that SOX2 is not involved in transition from the native state to the primed state when top 10% pathways were considered. We further modulated the network and observed top 20% pathways and found that SOX2 is not associated in transition from native to primed state. According to Weinberger et al. (2016), the native state is observed at the pre-implantation blastocyst state (E6-E7), whereas the primed state is observed at post implantation embryo stage (E9-E12). We performed expression analysis across the developmental timeline form 2-cell to blastocyst stages for analyzing the expression of SOX2 in the native and primed state and found that SOX2 is upregulated at the 6-cell stage and continues to be upregulated till the morula stage. As it enters the pre-implantation blastocyst stage (native), SOX2 is downregulated. As it moves from pre to post implantation stage (primed), upregulation of SOX2 occurs. This observation led to an important inference that SOX2 is an important player but it is not an essential as are OCT4 and NANOG in maintaining pluripotency in human embryonic stem cell.

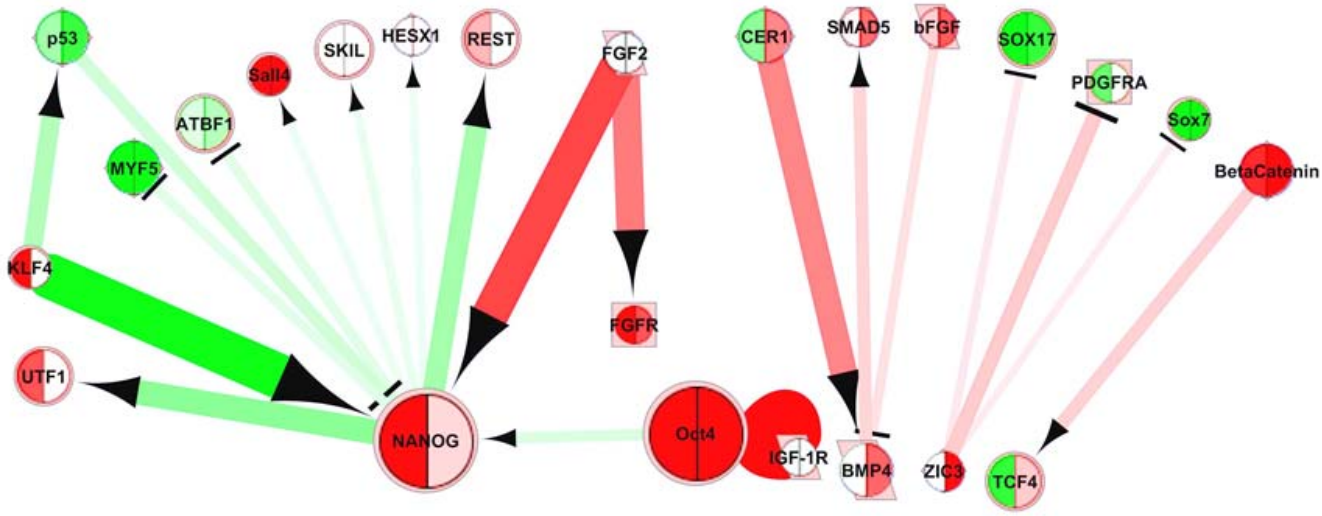

Fig. 4 A Network condensation through ExprEssence (Naive vs Primed). *hPluriNet* condensed comparing microarray data from human native and primed pluripotent states. The top 10% startups (red) and the top 10% shutdowns (green) are presented. See the text for details.

3.3.2 Transition from primed to non-pluripotent

Transition from primed state to non-pluripotent state was studied by condensing the network. We observed major shut-downs around OCT4 and NANOG (Fig. 4b), which highlights their role in the maintenance of pluripotency in human. In the developmental progress towards the non-pluripotent state differentiation starts that causes shut-down of major pluripotency players and activation of WNT, BMP4, and AKT1 pathways (Fig. 4b). These results were supported by several studies (Sun et al., 2006; Katoh and Katoh, 2007; Chen et al., 2012; Davidson et al., 2012). In addition to these, we also observed that the proteins ZIC3 and L1TD1 were upregulated in the primed state and downregulated in the non-pluripotent state indicate that ZIC3 and L1TD1 are indispensable players for the maintenance of pluripotency in human (Fig. 4B).These observations have also been supported by the experimental studies (Lim et al., 2007; Wong et al. 2011). Among the signaling pathways, we observed the downregulation of PI3K by the inhibition of PTEN, indicating a strong role in the maintenance of pluripotency. This has been advocated the several studies (Armstrong et al., 2006; Alva et al., 2011; Chen et al., 2012). We also observed strong start-up links centric to WNT5A resulting in activation of FRIZZLED causing an upregulation of both pathways indicating their role in differentiation (Davidson et al., 2007; 2012). Start-up of mTOR via AKT1 by bFGF and MAPK signaling is also an important inference of the transition from primed to non-pluripotent state established their roles in differentiation process (Easley et al., 2010; Zhou and Wang, 2012).

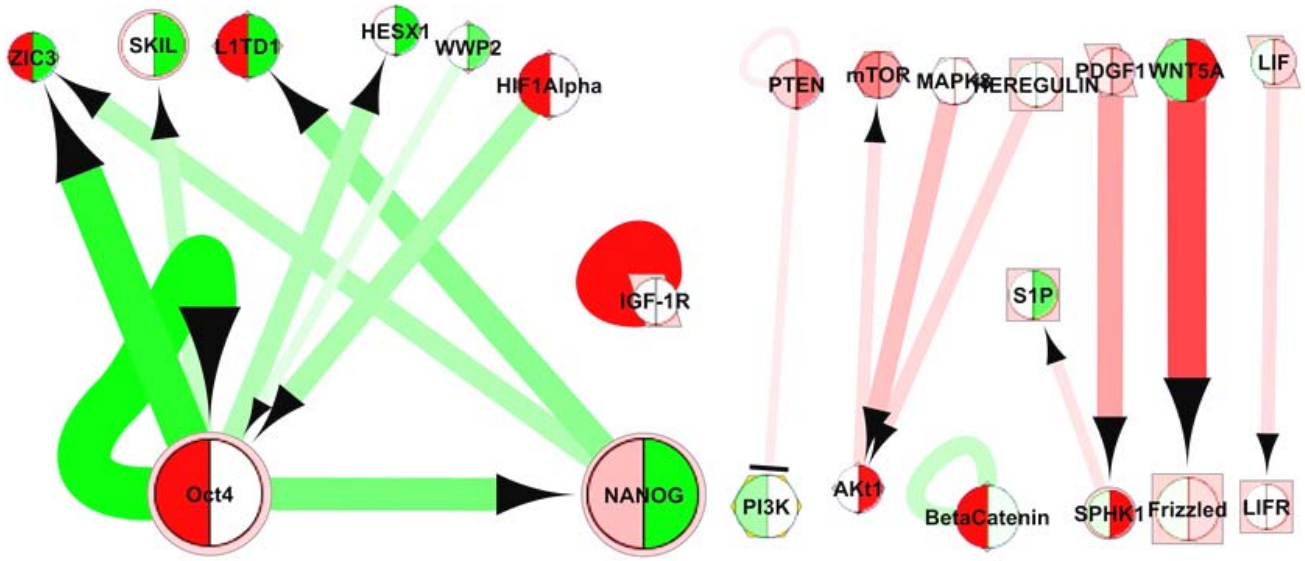

Fig. 4B Network condensation through ExprEssence (Primed vs. Non-pluripotent). *hPluriNet* condensed comparing microarray data from primed and non-pluripotent states. The top 10% startups (red) and the top 10% shutdowns (green) are presented.

3.3.3 Transition from pluripotent to non-pluripotent state

We condensed the network to determine the most differentially altered pathways those are critically involved in transition from pluripotent to non-pluripotent state. The results include major shut downs centric to OCT4 and NANOG, further validating that the two TFs are essential for induction and maintenance of pluripotency in human (Babaie et al., 2007). We also observed the downregulation of proteins including TFs such as KLF4, UTF1, ZIC3, and epigenetic regulator L1TD1, which further validates our hypothesis that they are important for induction and maintenance of pluripotency as the transition to non-pluripotent state resulted in decrease of their expression (Fig. 4c). Major start-ups are also observed around WNT signaling which is in congruency with our primed vs non-pluripotent results, further validating our observation that WNT signaling pathway promotes differentiation (Davidson et al., 2007, 2012). The MAPK signaling activates AKT1 signaling when the transition from pluripotent to non-pluripotent state occurs. mTOR signaling activated S6K which leading to upregulation of p53 in the non-pluripotent state: a critical mediator of human embryonic stem cells survival and differentiation process (Easley et al., 2010; Qin et al., 2007). The top 10% start-up and shut-down links involved in transition from pluripotent to non-pluripotent state are depicted in Fig.4c.

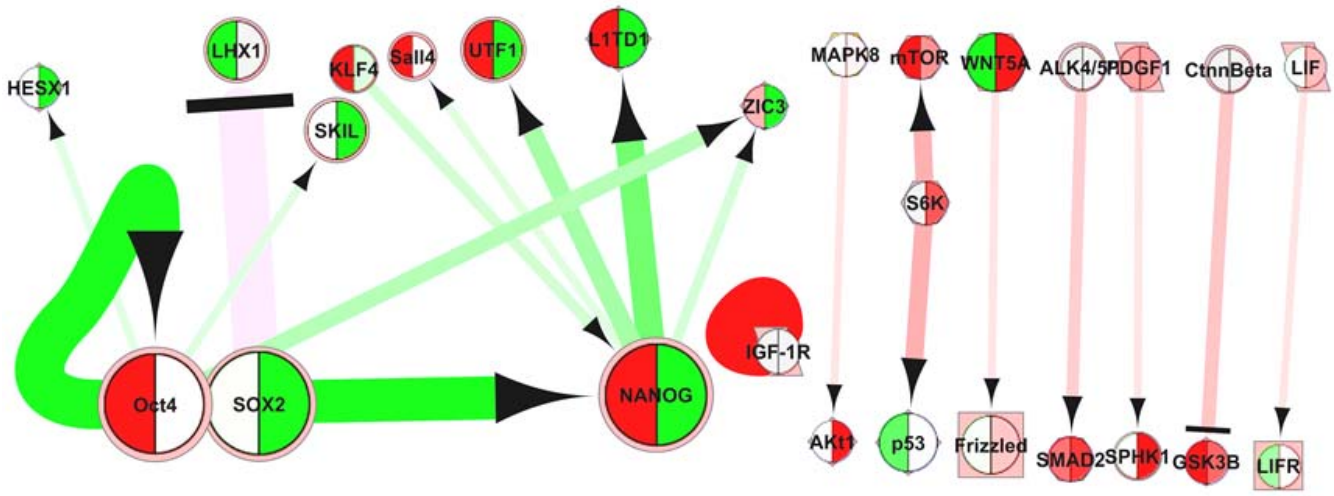

Fig. 4C Network condensation through ExprEssence (Pluripotent vs. Non-pluripotent). *hPluriNet* condensed comparing microarray data from pluripotent and non-pluripotent states. The top 10% startups (red) and the top 10% shutdowns (green) are presented.

3.4 Expression profile across developmental timeline

Gene expression profiling of human pre-implantation was followed from oocyte to blastocyst stages, and compared with ESCs. Most of the major pluripotency TFs (OCT4, SOX2, NANOG, and KLF4) show very low level of expression up to 4-cell stage of development (Fig. 5a). When development progresses further to the six-cell stage, the core pluripotency TFs, NANOG, SOX2 and KLF4 are over-expressed, however, the expression of OCT4 continues to remain low (Fig. 5b). At the 8-cell stage, the OCT4 gets activated and thus at this stage the entire core circuitry becomes completely operational. This observation implies that the pluripotency program begins at the 8-cell stage (Fig. 5b). L1TD1 is also an important start up link along with OCT4 which may be considered as hallmark of the onset of pluripotency in human (Närvä et al., 2012). The signaling components involved in pluripotency such as MAPK, RAS, PI3K, and FGF get activated at the 8-cell stage which apparently helps in induction and/or maintenance of pluripotency (Fig. 5c). Also, expression of a group of TFs (MYST3, REST, GATA3 and PRDM14) associated with the core circuit increases at the morula stage. An important observation emerges that SOX2 expression peaks at the morula stage whereas its expression decreases at the blastocyst stage (Fig.5d).

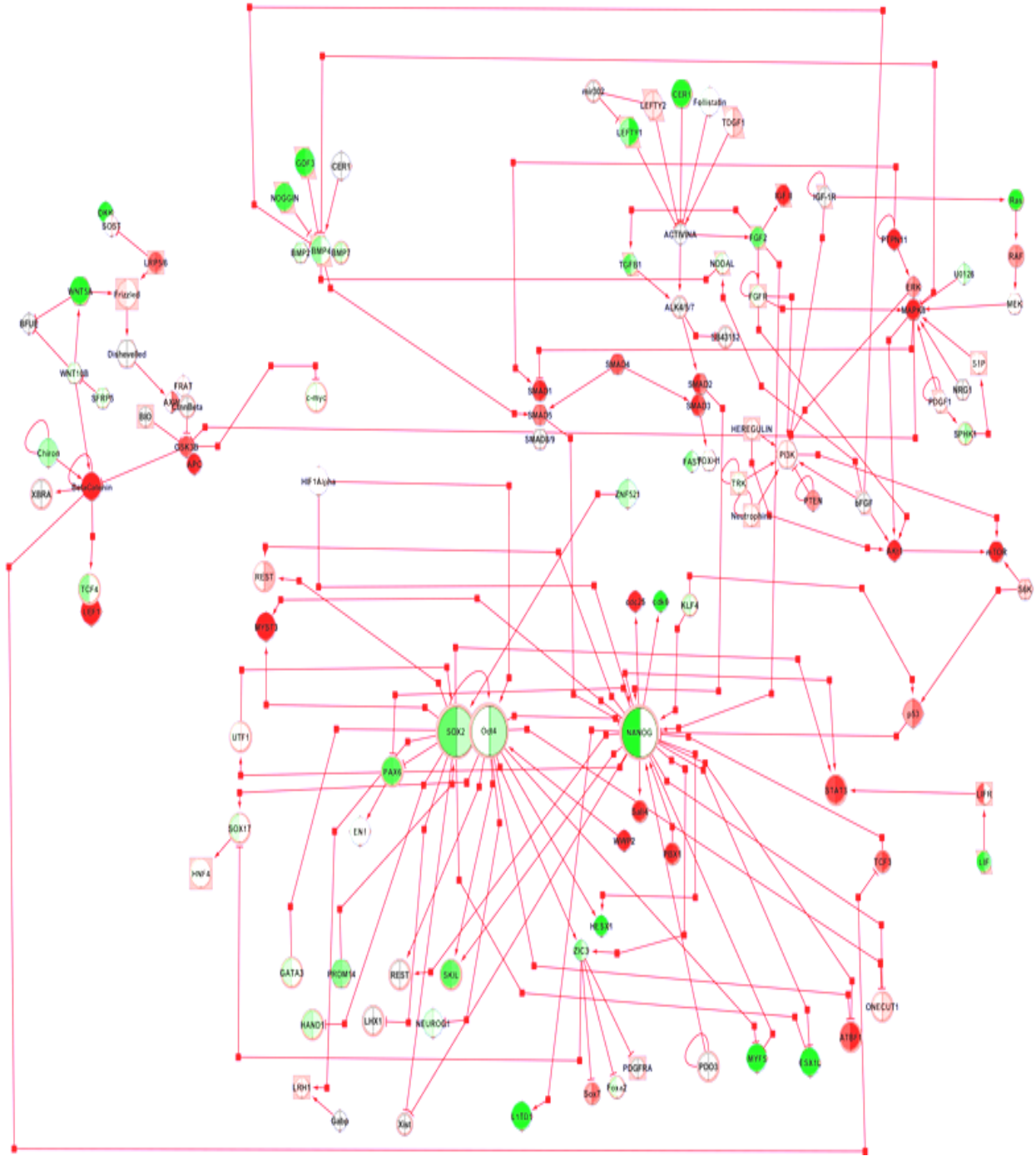

Fig. 5a Expression Profiling across developmental timeline (2-cell vs. 4-cell stage). *hPluriNet* with differential gene expression data showing similar expression patterns in 2-cell and 4-cell stages.

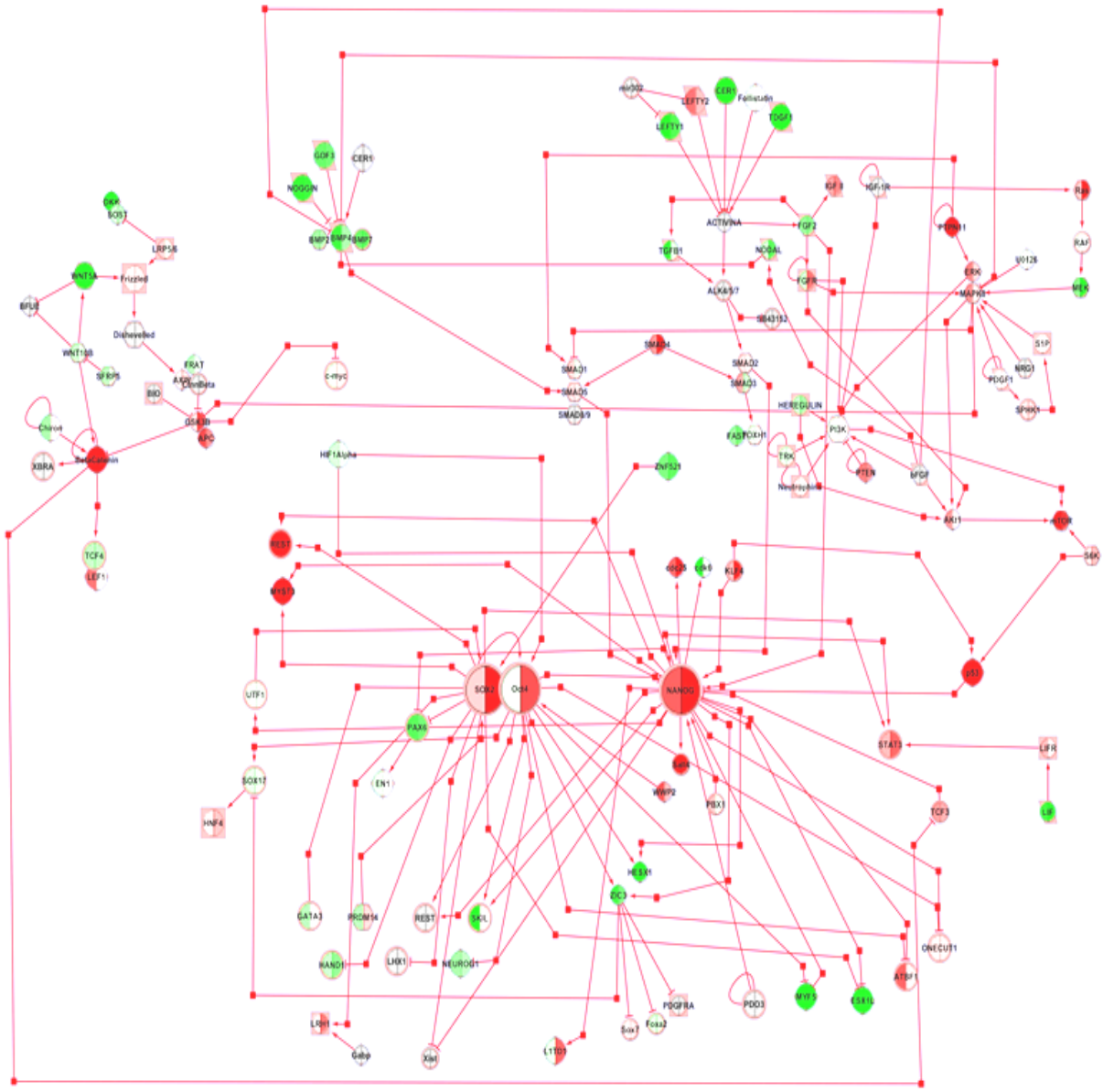

Fig. 5b Expression Profiling across developmental timeline (6-cell vs. 8-cell stage). *hPluriNet* with differential gene expression data shown the transcriptional activation beings at the 6-cell to 8-cell stages.

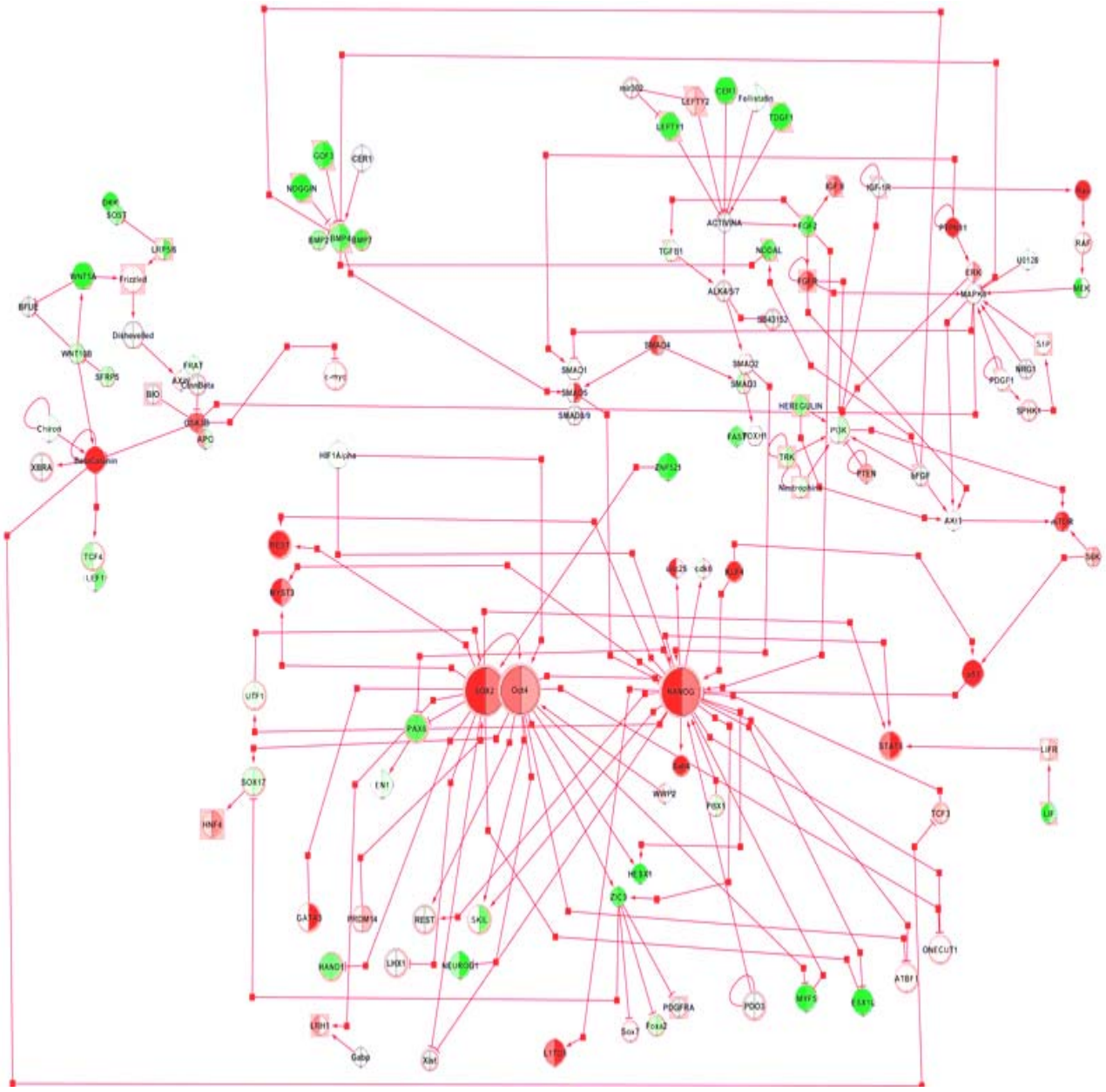

Fig. 5c Expression Profiling across developmental timeline (8-cell vs. Morula stage). *hPluriNet* with differential gene expression data established that the TFs activated the core regulatory circuitry as it progressed to the morula stage.

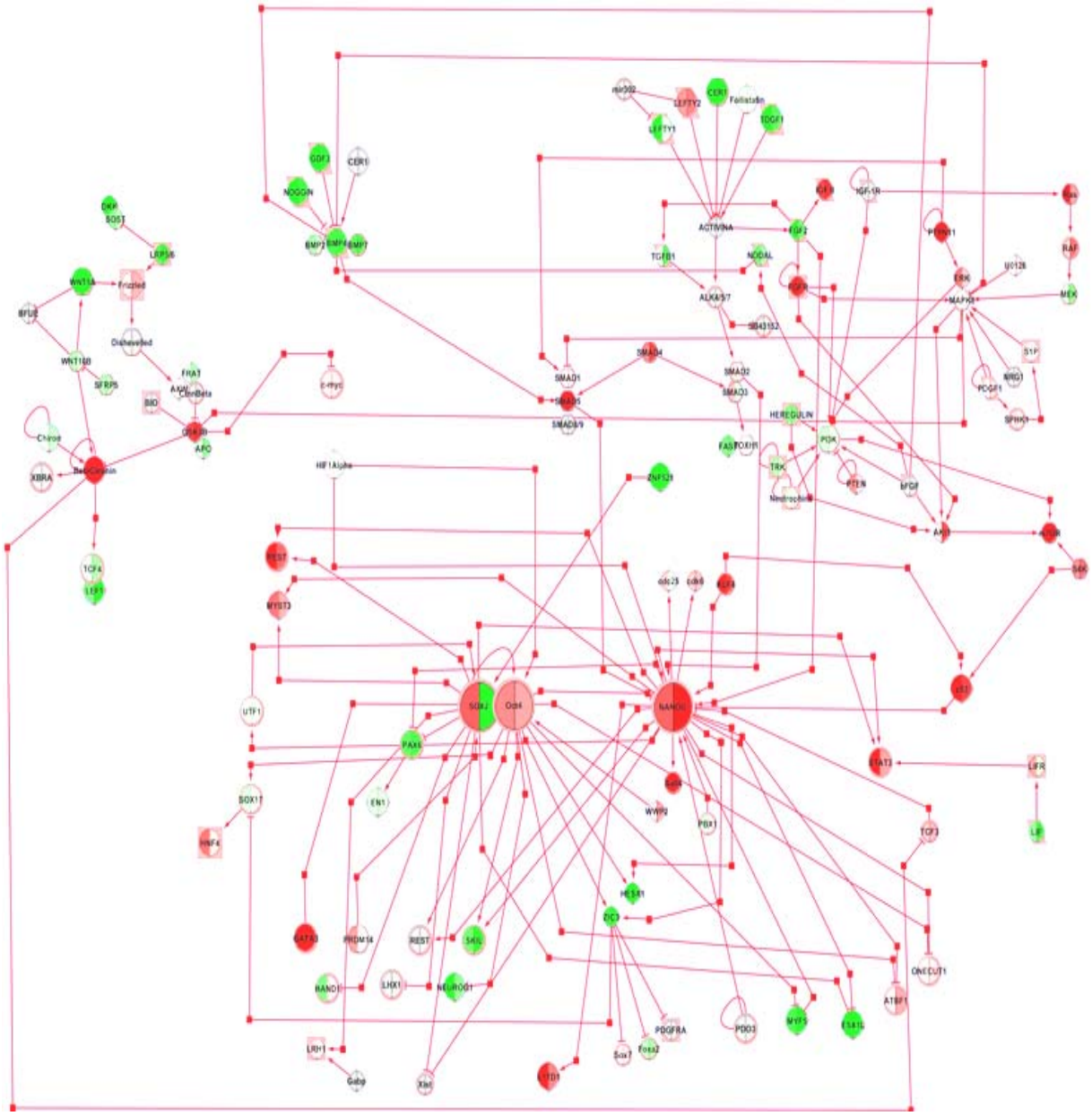

Fig. 5d Expression Profiling across developmental timeline (Morula vs. Blastocyst stage). *hPluriNet* with differential gene expression data shown the genes/proteins up/down-regulation patterns from morula stage to blastocyst stage.

4 Conclusions

In conclusion, we have assembled an up-to-date extended picture of the underlining mechanisms of pluripotency in human which would be useful as a model system for studying early development and for the generation of specific cell types for use in regenerative medicine and drug screening. Assembling an electronic circuitry of the interaction/regulation network of genes/proteins involved in human pluripotency and annotation of associated biological information provide a formal way to combine different types of biological data into a single conceptual framework. Furthermore visualization and explorative analysis of pluripotency related differential gene expression data identifies the key players and the pathways involved in transitions from native to prime state, primed to non-pluripotent state, and pluripotent to non-pluripotent state.

In future, we are interested in adding of small molecules and microRNAs to the *hPluriNet* wherever these mechanisms have shown to be of importance. Furthermore, the comparison between human and mouse pluripotency networks should be performed, which will be useful to identify species-specific pathways evolution, and afford a deeper understanding of the evolution of pluripotency.

Acknowledgment

AS gratefully acknowledge University Grants Commission (UGC), India for providing financial support. We thank Dr. Vinod Verma of University of Allahabad for useful discussions.

References

Alva JA, et al. 2011. Phosphatase and tensin homolog regulates the pluripotent state and lineage fate choice in human embryonic stem cells. Stem Cells, 29(12): 1952-1962

Armstrong L, et al. 2006. The role of PI3K/AKT,MAPK/ERK and NFKBeta signalling in the maintenance of human embryonic stem cell pluripotency and viability. Human Molecular Genetics 15(11): 1894-1913

Ashburner M, et al. 2000. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics, 25(1): 25-29

Assenov Y, et al., 2008. Computing topological parameters of biological networks. Bioinformatics, 24(2): 282-284

Babaie Y, et al. 2007. Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. Stem Cells, 25(2): 500-510

Barabasi AL, Albert R. 1999. Emergence of scaling in random networks. Science, 286(5439), 509-512

Boyer LA, et al. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell, 122(6): 947-956

Chen YG, Li Z, Wang XF. 2012. Where PI3K/AKT meets SMADs: The crosstalk determines Human Emmbryonic stem cell fate. Cell Stem Cell, 10(3): 231-232

Chin MH, et al. 2010. Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. Cell Stem Cell, 7(2): 263-269

Davidson KC, et al. 2012. Wnt/β-catenin signaling promotes differentiation, not self-renewal, of human embryonic stem cells and is repressed by Oct4. Proceedings of the National Academy of Sciences, 109(12): 4485-4490

Davidson KC, et al. 2007. Wnt3a regulates survival, expansion, and maintenance of neural progenitors derived from human embryonic stem cells. Molecular and Cellular Neuroscience, 36(3): 408-415

De Los Angeles,A, et al. 2015. Hallmarks of pluripotency. Nature, 525(7570): 469-478

Easley IV, et al. 2010. mTOR-mediated activation of p70 S6K induces differentiation of pluripotent human embryonic stem cells. Cellular Reprogramming (Formerly" Cloning and Stem Cells"), 12(3): 263-273

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research, 30(1): 207-210

Fong H, Hohenstein KA, Donovan PJ. 2008. Regulation of self-renewal and pluripotency by Sox2 in human embryonic stem cells. Stem Cells, 26(8): 1931-1938

Gafni O, et al. 2013. Derivation of novel human ground state naive pluripotent stem cells. Nature, 504 (7479): 282-286

Gan Y, et al. 2015. Prediction of the anti-inflammatory mechanisms of curcumin by module-based protein interaction network analysis. Acta Pharmaceutica Sinica B, 5(6): 590-595

Ginis I, et al., 2004. Differences between human and mouse embryonic stem cells. Developmental Biology, 269(2): 360-380

Hakes L, et al. 2008. Protein-protein interaction networks and biology—what's the connection? Nature Biotechnology, 26(1): 69-72

Hanna J, et al. 2010. Human embryonic stem cells with biological and epigenetic characteristics similar to those of mouse ESCs. Proceedings of National Academy of Sciences USA, 107(20): 9222-9227

Hyslop L, et al., 2005. Downregulation of NANOG induces differentiation of human embryonic stem cells to extraembryonic lineages. Stem Cells, 23(8): 1035-1043

Jain S, Bhooshan SV, Naik PK. 2011. Mathematical modeling deciphering balance between cell survival and cell death using insulin. Network Biology, 1(1): 46-58

Katoh M, Katoh M. 2006. Cer1 is a common target of Wnt and nodal signaling pathways in human embryonic stem cells. International Journal of Molecular Medicine, 17(5): 795-799

Kininmonth S, van Oppen M, Castine S. 2012. The small genetic world of Seriatopora hystrix. Network Biology, 2(1): 1-15

Lim LS, et al. 2007. Zic3 is required for maintenance of pluripotency in embryonic stem cells. Molecular Biology of the Cell, 18(4): 1348-1358

Maere S, Heymans K, Kuiper M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics, 21(16): 3448-3449

Marteyn A, et al. 2011. Mutant human embryonic stem cells reveal neurite and synapse formation defects in type 1 myotonic dystrophy. Cell Stem Cell, 8(4): 434-444

Muller FJ, et al. 2008. Regulatory networks define phenotypic classes of human stem cell lines. Nature, 455(7211): 401-405

Närvä E, et al. 2012. RNA Binding Protein L1TD1 Interacts with LIN28 via RNA and is Required for Human Embryonic Stem Cell Self Renewal and Cancer Cell Proliferation. Stem Cells, 30(3): 452-460

Newman AM, Cooper JB. 2010. AutoSOME: a clustering method for identifying gene expression modules without prior knowledge of cluster number. BMC Bioinformatics, 11(1): 117

Nishimoto M, et al. 2001. Structural analyses of the UTF1 gene encoding a transcriptional coactivator expressed in pluripotent embryonic stem cells. Biochemical and Biophysical Research Communications, 285(4): 945-953

Orr SJ, et al. 2012. Proteomic and protein interaction network analysis of human T lymphocytes during cell cycle entry. Molecular Systems Biology, 8(1): 573

Qin, H., et al. 2007. Regulation of apoptosis and differentiation by p53 in human embryonic stem cells. Journal of Biological Chemistry, 282(8): 5842–5852

Rao M. 2004. Conserved and divergent paths that regulate self-renewal in mouse and human embryonic stem

cells. Developmental Biology, 275(2): 269-286

Reubinoff BE, et al. 2000. Embryonic stem cell lines from human blastocysts: somatic differentiation in vitro. Nature Biotechnology, 18(4): 399-404

Shannon P, et al. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11): 2498-2504

Som A, et al. 2010. The PluriNetWork: an electronic representation of the network underlying pluripotency in mouse, and its applications. PLoS One, 5(12): e15165

Som A, et al. 2012. Derivation of an interaction/regulation network describing pluripotency in human. Gene 502(2): 99-107

Sun J, et al. 2006. BMP4 activation and secretion are negatively regulated by an intracellular gremlin-BMP4 interaction. Journal of Biological Chemistry, 281(39): 29349-29356

Theunissen TW, et al. 2014. Systematic identification of culture conditions for induction and maintenance of naive human pluripotency. Cell Stem Cell, 15(4): 471-487

Thomson JA, et al. 1998. Embryonic stem cell lines derived from human blastocysts. Science, 282(5391): 1145-1147

Vassena R, et al. 2011. Waves of early transcriptional activation and pluripotency program initiation during human preimplantation development. Development, 138(17): 3699-3709

Warsow G, et al. 2010. ExprEssence-revealing the essence of differential experimental data in the context of an interaction/regulation net-work. BMC Systems Biology ,4(1): 164

Weinberger L, et al. 2016. Dynamic stem cell states: naive to primed pluripotency in rodents and humans. Nature Reviews Molecular Cell Biology, 17(3): 155-169

Wong RC, et al. 2011. L1TD1 is a marker for undifferentiated human embryonic stem cells. PLoS One, 6(4): e19355

Wu Z, Irizarry RA. 2004. Preprocessing of oligonucleotide array data. Nature Biotechnology, 22(6): 656-658

Xu H, et al. 2010. Toward a complete in silico, multi-layered embryonic stem cell regulatory network. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2(6): 708-733

Ye S, Liu D, Ying QL. 2014. Signaling pathways in induced naive pluripotency. Current Opinion in Genetics and Development, 28: 10-15

Zaehres H, et al. 2005. High-efficiency RNA interference in human embryonic stem cells. Stem Cells, 23(3): 299-305

Zhang WJ. 2012. Several mathematical methods for identifying crucial nodes in networks. Network Biology, 2(4): 121-126

Zhou J, Li D, Wang F. 2012. Assessing the function of mTOR in human embryonic stem cells. Methods in Molecular Biology, 821: 361-372