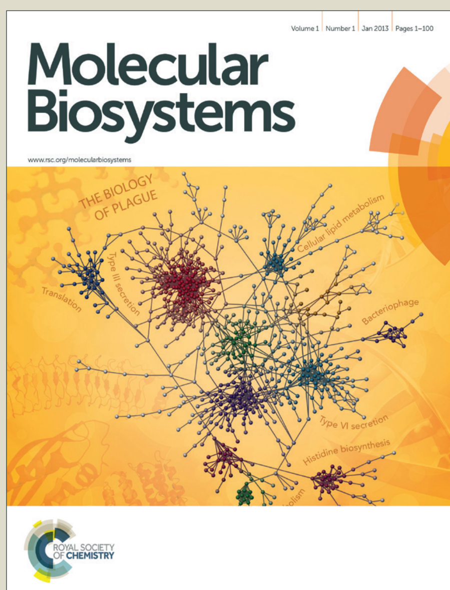
Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

www.rsc.org/molecularbiosystems

Journal Name

RSCPublishing

ARTICLE

Construction and analysis of a protein-protein interaction network related to self-renewal of mouse spermatogonial stem cells

Wenhai Xie,[a] Jin Sun,[a] and Ji Wu,[a,b]

Spermatogonial stem cells (SSCs) are responsible for sustained spermatogenesis throughout the reproductive life of the male. Extensive studies of SSCs have identified dozens of genes that play important roles in sustaining or controlling the pool of SSCs in the mammalian testis. However, there is still limited knowledge on whether or how these key genes interact with each other during SSC self-renewal. Here, we constructed a protein-protein interaction (PPI) network for SSC self-renewal based on interactions between 23 genes essential for SSC self-renewal, which were obtained from a text mining system, and the interacting partners of the 23 key genes, which were differentially expressed in SSCs. The SSC self-renewal PPI network consisted of 246 nodes connected by 844 edges. Topological analyses of the PPI network were conducted to identify genes essential for maintenance of SSC self-renewal. The subnetwork of the SSC self-renewal network suggested that the 23 key genes involved in SSC self-renewal were connected together through another 94 genes. Clustering of the whole network and subnetwork of SSC self-renewal revealed several densely connected regions, implying significant molecular interaction modules essential for SSC self-renewal. Notably, we found the 23 genes responsible for SSC self-renewal by forming a continuous PPI network centered on *Pou5f1*. Our study indicates that it is feasible to explore important proteins and regulatory pathways in biological activities by combining a PPI database with the high-throughput data of gene expression profiles.

Introduction

In mammalian testes, spermatogenesis is a classic adult stem cell-dependent process sustained by self-renewal and differentiation of spermatogonial stem cells (SSCs).[1] During spermatogenesis, SSCs give rise to daughter cells through mitotic cell division. The majority of the daughter cells differentiate and undergo meiosis and cellular morphogenesis to produce mature sperm.[2] Meanwhile, a small proportion of the daughter cells is functionally similar to SSCs maintaining the pool of SSCs through self-renewal. The number of SSCs estimated by transplantation is only approximately 0.002% of all germ cells in the testes.[3] Over the past few decades, many classic studies of SSCs have established the foundation to understand the molecular mechanisms underlying SSC self-renewal using techniques such as spermatogonial transplantation and long-term SSC culture.[4] Recently, high-throughput technologies such as microarray and transcriptome sequencing have been adopted for genome-wide expression studies of SSCs, revealing dozens of genes responsible for SSC self-renewal and several key signaling pathways involved in SSC self-renewal.[5-7]

SSCs reside on the basement membrane of seminiferous tubules and require extrinsic chemokines to maintain a specific niche for self-renewal. Chemokine receptors *Cxcr4* and *Csf1r* have been found to transduce signals from the extracellular matrix or sertoli cells into SSCs.[8, 9] In addition, several cell surface molecules have been identified as SSC markers. In rodents, *Thy1*, *Cd9*, *Cdh1*, *Gfra1*, and *Ret*, which are mainly cell surface markers, have been used to isolate SSCs from the testis.[10-14] In humans, G protein-coupled receptor 125 is a marker that can be used to isolate spermatogonial progenitor cells.[15, 16] By establishing a long-term SSC culture system, some signaling factors have been found to promote the self-renewal of SSCs, such as glial cell line-derived neurotrophic factor (GDNF), fibroblast growth factor 2 (FGF2), and short-type PB-cadherin.[17-19] GDNF activates the phosphoinositide 3-kinase-Akt pathway that maintains self-renewal of mouse SSCs.[20-22] FGF2 mediates SSC self-renewal via up-regulation of *Etv5* and Bcl6b through MAP2K1 activation.[18] Additionally, *Akt* has been identified as a critical regulator of SSC self-renewal.[21] Several transcription factors have been found to be required for maintenance of SSCs, such as *Pou5f1*, *Plzf*, and *Sall4*.[23-26] The expression of transcription factors *Pou3f1*, *Id4*, *Bcl6b*, *Egr2*, and *Egr3* is up-regulated by GDNF in SSC cultures in vitro.[5, 27, 28] Furthermore, RNA-binding proteins *Piwi2* and *Nonas2* are essential for maintenance of SSCs.[29, 30]

To date, although a number of genes have been identified in sustaining or controlling the pool of SSCs in the mammalian

Molecular BioSystems Accepted Manuscript

testis, there is still limited knowledge regarding the interactions among these key genes in SSCs.

With the continuous increase of information in protein-protein interaction (PPI) databases, topological analyses of PPI networks have been performed to explore important proteins and regulatory pathways in biological activities. In systems biology, the arrangement of molecular networks from gene expression data based on known interactions permits understanding of basic biological mechanisms.[31] A germline regulatory network, which was established by integration of transcriptome, interactome and phenotype information provide clues for male infertility[32]. Network modeling and topological analysis can provide additional information on the particular properties of genes and proteins involved in biological activities that require multiple genes and proteins to function together. Thus, the combination of PPI analysis and gene expression can provide a better understanding of the mechanisms responsible for SSC self-renewal (Figure 1).

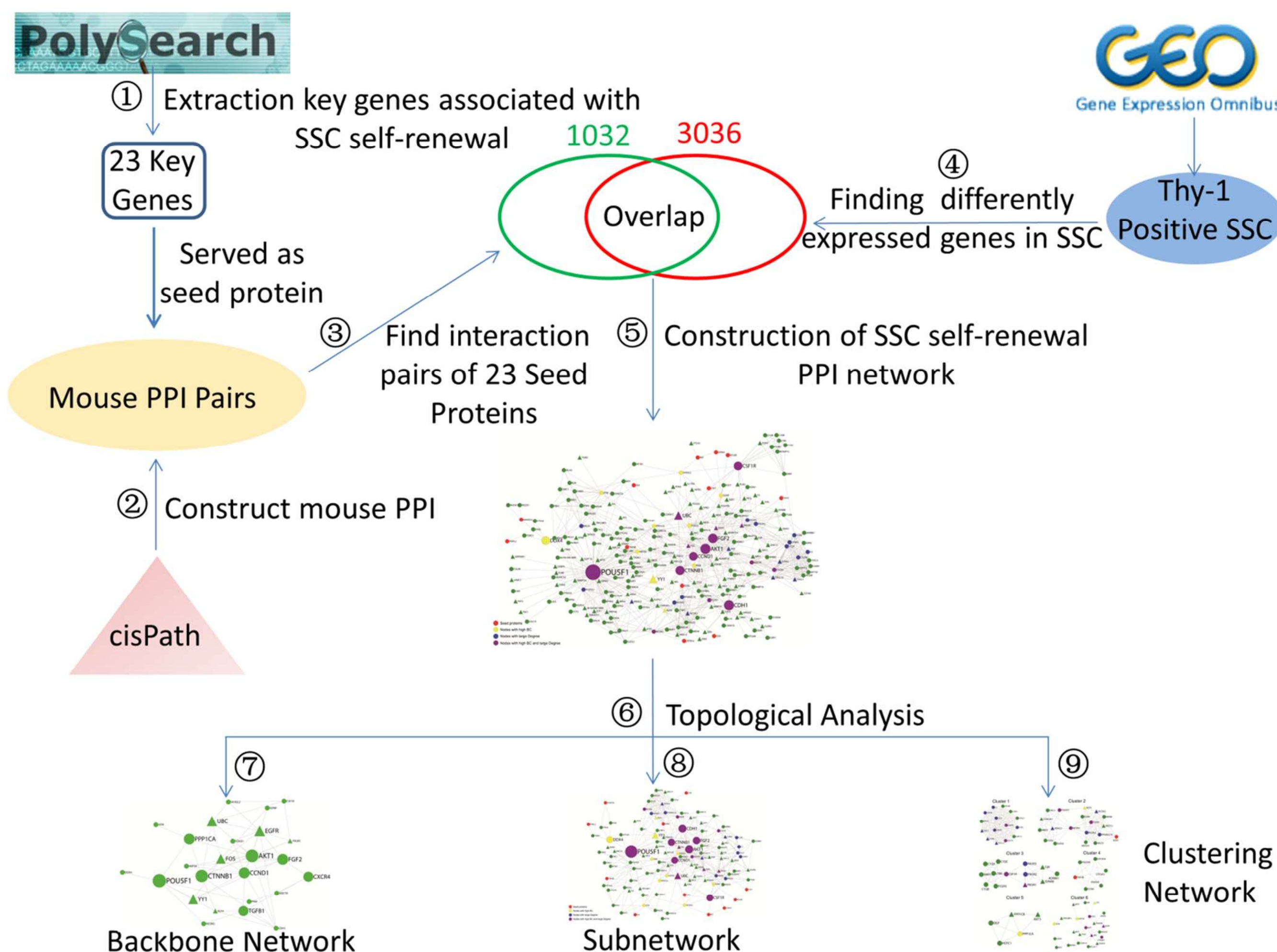


Fig. 1 Flow diagram of this study. First, key genes associated with SSC self-renewal were extracted by PolySearch. Using the 23 key genes related to SSC self-renewal as seed proteins, we found 1032 (green oval circle) interacting partners from the mouse PPI, and then determined the overlapping proteins between interacting partners of the seed proteins and SSC overexpressed genes (3036 genes in the red oval circle). Construction of the PPI network was based on the seed proteins and overlapping genes. Topological analysis of the PPI network revealed key node proteins and interaction modules essential for SSC self-renewal.

Materials and Methods

1. Extraction of key genes associated with SSC self-renewal

We identified candidate genes associated with SSC self-renewal using a text mining system, PolySearch, which produces a list of concepts relevant to the user's query.[33] The query type used in this study was 'Text-word-Gene/Protein' and the query keyword was 'spermatogonial stem cell'. The PolySearch system returned 59 genes/proteins from 116 articles after removing repeat cites. PolySearch result was found in Table S1 (ESI†). After manual confirmation, we obtained a list of 23 key genes that play important roles in maintenance of SSC self-renewal (Table 1).

2. Combining several well-known PPI databases to construct a mouse PPI

The R package of cisPath was used to generate PPI pairs for the mouse from PINA,[34] iRefIndes,[35] and STRING 9.1.[36] The combined PPI included 17,991 proteins and 341,219 interactions. Using the genes listed in Table 1 as seed proteins,

Molecular BioSystems Accepted Manuscript

 This journal is © The Royal Society of Chemistry 2012

we compiled a PERL script to extract their interacting partners from the combined PPI. These seed proteins directly interacted with 1,032 proteins. (Table S2 (ESI†)).

3. SSC differentially expressed genes obtained from published microarray data

The gene expression data of SSCs were retrieved from Wu *et al*.[37, 38] and JM *et al*.[9] Data from GSM356344, GSM356345, GSM356346, GSM761164, GSM761165, GSM761166, and GSM761167 were used for gene expression profiles of SSCs. Testis somatic cell data from GSM468625, GSM468627, and GSM468628 were used as the control to obtain differentially expressed gene list. These gene expression profiles of SSCs were all extracted from enriched Thy-1-positive SSCs. The method of microarray data analysis was described as follows. Microarray analysis was conducted in the R software environment for statistic computing. The Affy package was used to process the raw data of SSCs and testis somatic cells[39]. Firstly, the raw data of three samples were loaded into the same one AffyBatch class. After detection the quality of the data, we detected and normalized the expression of probes using rma function which was precoded with background correction and probe level normalization procedures in Affy package. Secondly, we respectively compared the expression profiles between the two SSC datasets with testis somatic cell dataset using linear models and empirical bayes methods methods[40]. Moderated t-statistic was used to verify the different expressed genes between SSC and testic somatic cell. The Benjamini and Hochberg's method was used to control the false discovery rate. We used the adjusted p-values to control the false discovery rate. FDR ratio was set as 5%. Genes that at least two-folds changed with adjusted p-values below 0.05 were selected as SSC differentially expressed. In order to reduce the bias brought by two different microarray datasets, we combined the two SSC differentially expressed gene lists together as SSC differentially expressed gene list (Table S3 (ESI†)).

4. Construction and analysis of the PPI network for SSC self-renewal

The constructed PPI network for SSC self-renewal consisted of the 23 key genes and their direct PPI neighbors differentially expressed in SSCs. We compiled a PERL script to identify interactions between these genes. In the theory of the network, the connectivity degree (k) and betweenness centrality (BC) value of nodes are two fundamental parameters that are usually adopted to evaluate the nodes in a network.[41] The degree (k), defined as the number of interacting partner proteins, is the most basic characteristic of a node. The BC value is the fraction of the number of shortest paths that pass through each node in a network, which measures how often the node is located on the shortest paths between other nodes.[42] The shortest path is the path containing the least number of vertices between two vertices in a network. A node with a higher BC value indicates that it has more influence over the information flow in the network. Therefore, BC values are generally a useful indicator to detect bottlenecks in a network. Closeness centrality (CC) is the inverse of the average length of the shortest paths to/from all other nodes in the graph and measures how close a node is to other nodes.[42] The node with the highest CC is usually the topological center of the network. In this study, Cytoscape 3.0.2[43] was used to calculate the properties of nodes and perform measure under default parameters.

Table 1 Genes with essential roles in SSC self-renewal.

Symbol	Function	References
Cell surface markers		
Thy1	surface marker of SSCs used in isolation	10
Cd9	Cell surface marker on mouse and rat SSCs	11
Cdh1	Marker for undifferentiated spermatogonia	13
Gpr125	Human SSC marker	15
Gfra1	GDNF receptor	14
Ret	Downstream target of GDNF mediated pathway	17
Transcription factors		
Plzf	Transcription repressor required to regulate self-renewal and maintenance of SSCs pool	24, 23
Id4	Transcription repressor expression is up-regulated by GDNF	27
POU5f1	Required for SSCs maintenance in culture and for colonization activity following cell transplantation.	26
Pou3f1	an Important Intrinsic Regulator of GDNF-Induced Survival and Self-Renewal of SSCs	28
TAF4b	A transcriptional regulator required for the regulation of SSCs proliferation.	44
Egr2	GDNF up-regulated Transcription factors	5
RNA-binding protein		
Piwi2	RNA-binding protein required for SSCs self-renewal	29
Nanos2	Required for maintenance of SSC and is downstream of GDNF mediated signaling pathway	30
Niche related		
CXCR4	Receptor of chemokine required for maintenance of SSC	8
Csf1r	Receptor of extrinsic stimulator	9
Extrinsic growth factors		
GDNF	Growth factors required for SSCs self-renewal	22
FGF2	Growth factors required for SSCs self-renewal	18
Csf1	an extrinsic stimulator of mouse SSCs self-renewal	9
Others genes related to SSC self-renewal		
Foxo1	Required in maintenance of mouse SSCs and initiation of spermatogenesis	45
Atm	ATM-deficient result to SSCs defection through DNA damage-induced cell-cycle arrest	46
Bcl6b	GDNF and FGF2 up-regulated gene	5,18
Etv5	GDNF and FGF2 up-regulated gene	5,18

Molecular BioSystems Accepted Manuscript

Journal Name RSCPublishing

ARTICLE

5. Creation of the backbone network of the PPI network for SSC self-renewal

In the theory of the graph, proteins with high BC values are usually thought to be bottlenecks controlling the information flow in the transportation network. We set the critical node with a high BC value at 10% of the total nodes in the network. The proteins with a higher BC value and the links between them will make up a backbone network.[42] Thus, we extracted the proteins with the top 10% of BC values and the links between them from the PPI network for SSC self-renewal to create a backbone network.

6. Construction of a subnetwork consisting of all shortest paths between the 23 key genes

Pesca2.0,[47] a plug-in for Cytoscape, was used to calculate the shortest paths between the 23 key genes. The subnetwork consisted of all nodes in these shortest paths. The subnetwork indicates the possible minimal number of connections among the key 23 genes responsible for the self-renewal of SSCs.

7. Topological analysis to identify densely connected regions in the PPI network

MCODE,[48] another plug-in for Cytoscape, was used to cluster the whole network to identify densely connected regions. We selected clusters consisting of more than five node proteins as candidate clusters. After clustering the PPI network for SSC self-renewal, function annotation of the nodes located in the clusters was performed by DAVID.[49]

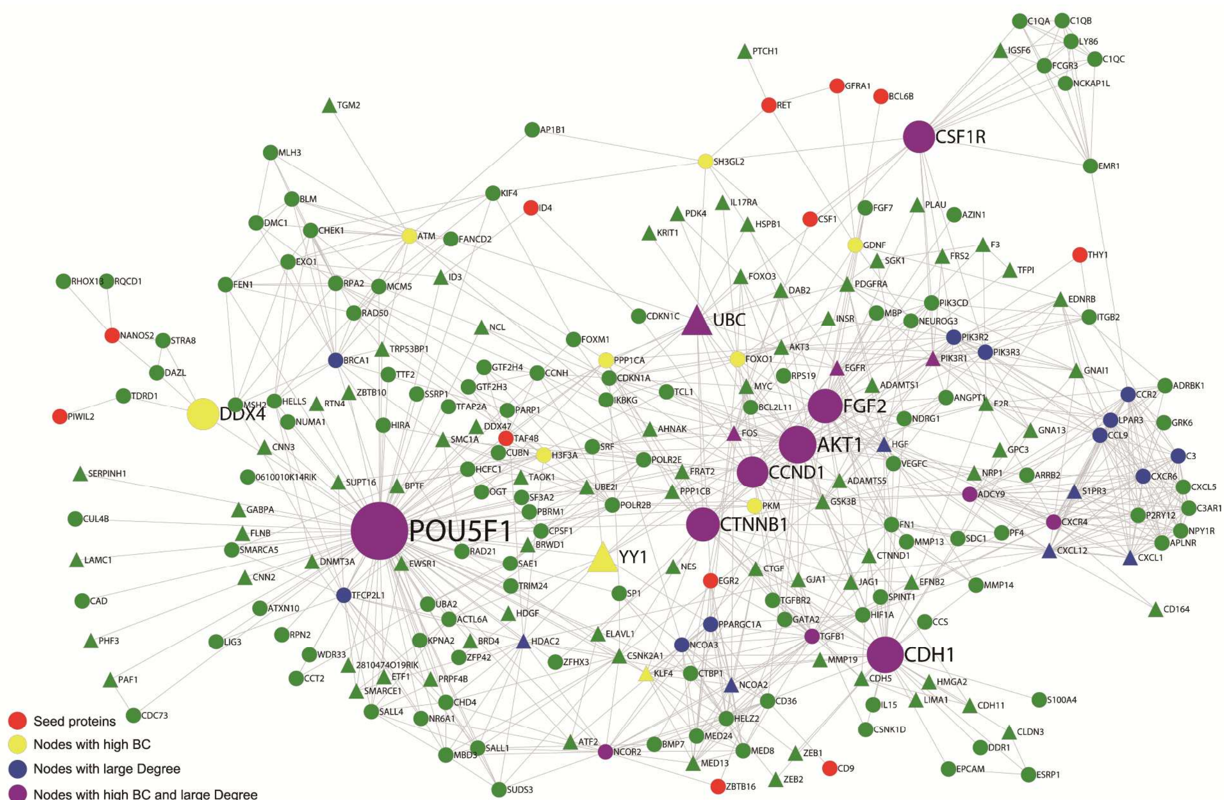


Fig. 2 Overview of the constructed PPI network for SSC self-renewal. The whole network consisted of 246 nodes and 844 edges. Key nodes in the PPI network are highlighted in different colors. The size of the nodes corresponds to their BC values. POU5F1 was located at the center of the subnetwork. Nodes with triangle shape are down-regulated genes in SSC.

Results

1. A PPI network for SSC self-renewal

The network was derived from 23 key genes that interacted with SSC differentially expressed genes and consisted of 246 nodes connected by 844 edges (Fig. 2). The backbone network of the SSC self-renewal PPI network consisted of 25 nodes that were connected via 58 edges (Fig. 3). We compared the measurement parameters of the network between the SSC self-renewal PPI network and backbone network (Table 2). The largest degree in the whole network was 88, while the average degree was 6.862. Thus, most of the nodes in this network were

Molecular BioSystems Accepted Manuscript

low connected, which is the classical character of a PPI network.[50]

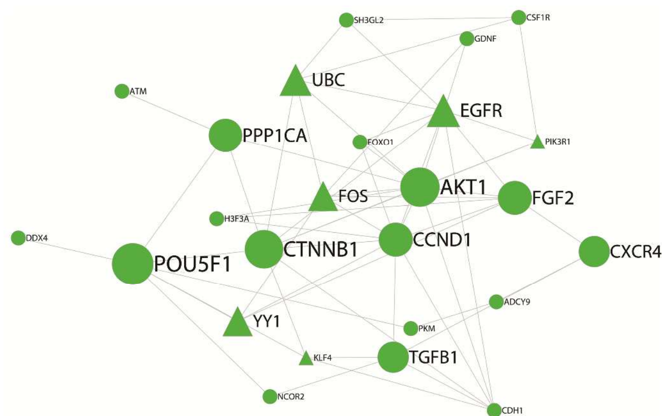


Fig. 3 Backbone of the SSCs self-renewal PPI network. The backbone network consists of 25 high BC value nodes and 58 edges. The size of the nodes correspond to their BC values.

Table 2 The general network measurements for networks

Parameter	Whole network	Backbone network	Subnetwork
Number of nodes	246	25	117
Average degree	6.862	4.64	7.453
Diameter	8	5	4
Mean shortest path length	3.227	2.297	2.914

2. Key nodes in the PPI network for SSC self-renewal

Of the 246 total nodes, 25 nodes had a high BC value (Table S4 (ESI†)), 35 nodes had a large degree, and 15 nodes had both a high BC value and high degree (Table S5 (ESI†)). To distinguish the different roles of these key nodes in the network, they were highlighted with different colors and sizes. The size of the nodes corresponded to their BC values (Fig. 2). Based on nodes with a high BC value and large degree, the top three nodes in the two lists were in the same order. The three nodes were POU5F1, AKT1, CDH1 and FGF2. POU5F1 was not only a hub protein with the largest degree but also a bottleneck node with the largest BC value. We also found that POU5F1 had the largest CC value, indicating that POU5F1 was located at the center of the network.

3. Backbone network of the PPI network for SSC self-renewal

The backbone network consisted of 25 high BC value nodes and 58 edges. The size of these nodes corresponded to their BC value (Fig. 3). AKT1 was located at the center of the backbone network with the highest CC value and largest degree. POU5f1 has the largest BC value, so it controls the information flow in the backbone network. POU5f1 has seven neighbors: CTNNB1, DDX4, NCOR2, PKM, PPP1CA, KLF4 and YY1. KLF4 and YY1 were down-regulated in SSC.

4. Subnetwork consisting of all shortest paths between the 23 key genes

The subnetwork consisted of 117 nodes and 436 edges. We found that AKT1 was located at the center of the subnetwork and had the highest CC value and largest degree (Fig. 4). POU5f1 had the highest BC value and second largest degree, which did not coincide with POU5f1 as the center of the PPI network for SSC self-renewal. The top 25 nodes with a high BC value in the subnetwork are listed in Table S6 (ESI†).

5. Densely connected proteins in the constructed PPI network for SSC self-renewal

Based on topology, we found six densely connected regions in the whole SSC self-renewal network (Fig. 5). The Gene Ontology (GO) term of cell surface receptor-linked signal transduction was overrepresented in cluster 1 and 3 nodes, indicating that these nodes may be involved in signal transduction to sustain the self-renewal of SSCs. Seed protein CXCR4, an important factor in maintenance of the SSC niche, was found in cluster 1, which coincides well with the functions of *Cxcr4* in signal transduction to maintain the SSC microenvironment.[8] Seed protein CSF1R located at cluster 3 has been found to enhance SSC self-renewal.[9] CXCL1 and CXCL12 are extrinsic chemokines essential to form SSC self-renewal nich. Nodes in clusters 1 and 3 may act as key molecules composing a special microenvironmental niche on the basement membrane of seminiferous tubules, which is required for SSC self-renewal. Functionally annotated nodes in cluster 6 enriched cell proliferation, cell cycle and transcription regulation related GO terms. For example, seed protein ATM and nodes CHECK1, PPPO1CA, and HCFC1 are cell cycle-related genes. *Atm* deficiency results in a defect of SSCs through DNA damage-induced cell cycle arrest.[46] The proliferation gene *Mcm5* was also located in cluster 6. Transcription regulated terms were overrepresented in Cluster 2 nodes. Trancription factor *Pouf51, Klf4, Sall1* and several nucleus molecules located in cluster 2. The cluster 2 described a detailed transcription regulation network. Other clusters contained genes that play key roles in SSC self-renewal. For example, gonad-specific *Taf4b* as a component of TFIID lists in cluster 4 is a transcriptional regulator required for regulation of SSC proliferation.[44] Cell growth factors FGF2 mediated signal pathway was displayed Cluster 6, which promotes SSC growth in vitro.

Discussion

By combining key genes involved in self-renewal of SSCs and the unique gene expression profiles of SSCs, we obtained a continuous PPI network consisting of 246 genes and 844 interactions. Based on topological analyses, we evaluated important proteins in the SSC self-renewal PPI network. Notably, POU5F1 was located at the center of the network. *Pou5f1* is an important transcription factor in maintaining the stemness of many stem cell types such as embryonic stem cells [51] and SSCs.[26] In addition, *Pou5f1* is a necessary transcription factor in induced pluripotent stem cells.[52] Berg *et al.*[51] has also proposed a PPI network centered on POU5f1, which controls pluripotent cell identity.

Clustering the SSC self-renewal network to find densely connected genes might help to delineate the underlying mechanisms of SSC self-renewal. Nodes in clusters 1 and 3 were involved in signal transduction to sustain the self-renewal of SSCs. We can directly perceive interactions of *Csf1r* and *Cxcr4* through senses. Nodes in cluster 6 suggested that the signaling pathways of FGF2 promote SSC self-renewal.

Sall4 plays a critical role in maintenance of the pluripotency and self-renewal of embryonic stem cells,[53, 54] and can activate the expression of *Pou5f1* in embryonic stem cells. *Pou5f1* interacts with Nanog to form a transcriptional core network essential for the maintenance of “stemness” of embryonic stem

Molecular BioSystems Accepted Manuscript

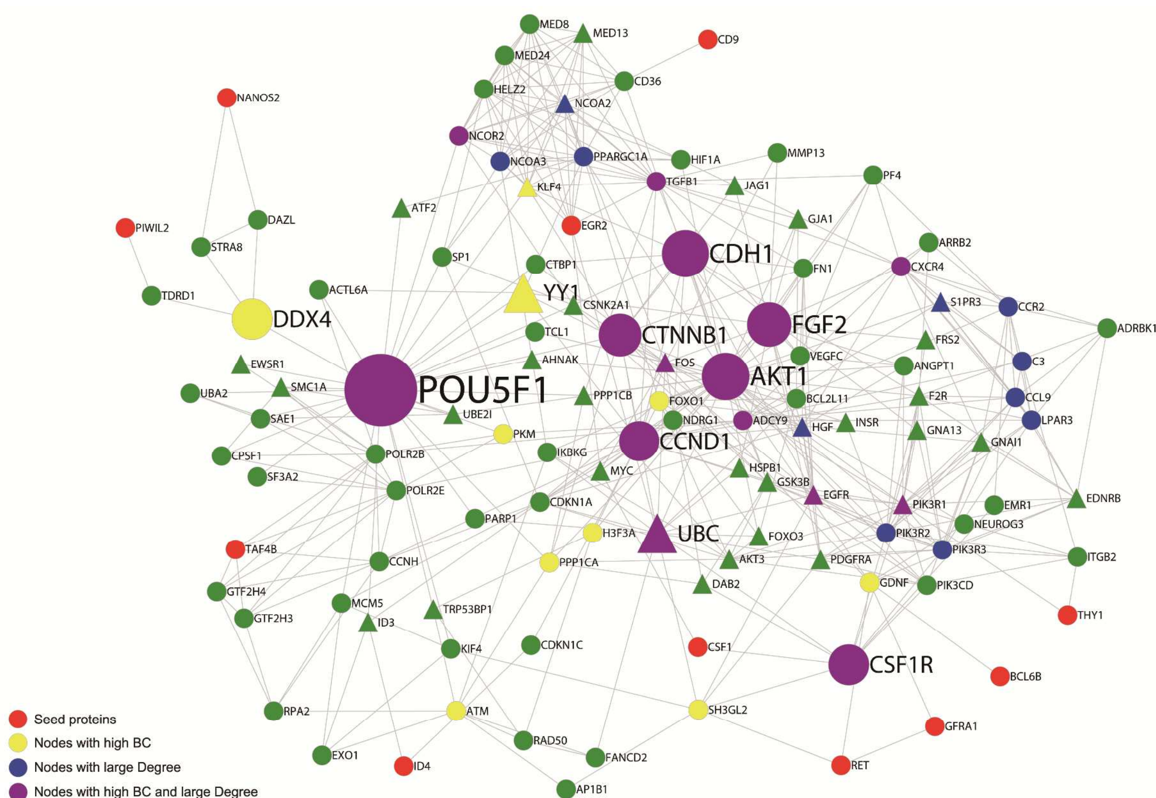


Fig. 4 The subnetwork consisting of all shortest paths between the 23 key genes. The seed proteins are connected by all shortest paths in the SSCs self-renewal PPI network. The subnetwork consists of 79nodes and 209 edges. The size of nodes corresponds to their BC values. AKT1 locats at the center of the subnetwork.

cells.[55, 56] In the PPI network for SSC self-renewal, SALL4 connected with POU5F1, SALL1, CHD4, MBD3, and TFCP2L1. Nodes in cluster 2 suggest that a transcriptional core network consisting of *POU5f1* and *Egr2* may be exist in SSCs. *Ncor2* connected two independent clusters centered on *Pou5f1* and *Egr2*. Another famous transcription factor Klf4, down-regulated in SSC was found in cluster 2. The interactions maybe sugesst some way how Klf4 was down-regulated in SSC. However, Hobbs *et al*.[25] proposed that *Plzf* antagonizes the function of *Sall4* by displacing *Sall4* from cognate chromatin to induce *Sall1* expression in SSCs. *Mbd3* and *Hdac2* are involved in epigenetic regulation. *Mbd3* is an effector for 5hmC in embryonic stem cells.[57] *Tfcp2l1* is at the intersection of leukemia inhibitory factor- and two small molecule inhibitor-mediated self-renewal pathways and plays a critical role in embryonic stem cell identity.[58] More experiments are needed to verify the regulatory mechanisms among *Sall4*, *Pou5f1*, and *Plzf* in SSC self-renewal.

The subnetwork was formed by the shortest paths between the 23 key genes in SSC self-renewal. Most of the top 25 nodes with a high BC value in the subnetwork were the same as those in the SSC self-renewal network. Only seven proteins were not in the list of top 25 nodes in the SSC self-renewal network (Table S6 (ESI†)). Unlike POU5F1 at the center of the whole network, we found that AKT1 was located at the center of the subnetwork. As mentioned above, AKT1 was at the intersection of two different growth factor-mediated signaling pathways, which may explain why AKT1 was located at the center of the subnetwork.

By clustering the subnetwork of the SSC self-renewal network, we found two clusters that was similar to the combination of clusters 1, 2 and 4 in the whole SSC self-renewal network (Fig. 6). CXCR4 interacted with TGFB1 that is a known growth factor involved in cellular proliferation, growth, and development. The two clusters in subnetwork could be bridged together by ADCY9 and CXCR4. As nodes in the two cluster of the subnetwork were involved in cell surface receptor-linked signal transduction and transcriptional regulation, we speculated that extracellular signals transduce into the cell nucleus via *Cxcr4* in SSC self-renewal. Via POLR2E or POLR2B, ADCY9 connected with TAF4B, a gonad-specific transcriptional regulator required for regulation of SSC proliferation.[44] *Polr2e* and *Polr2b* encode subunits of RNA polymerase II. Therefore, *Taf4b* may interact with RNA polymerase II to mediate the precise expression of genes essential for SSC self-renewal. *Adcy9* is a membrane-bound enzyme that catalyzes the formation of cyclic AMP from ATP.[59] Cyclic AMP is often considered an endogenous suppressor that inhibits cell proliferation and promotes cell differentiation. Thus, *Adcy9* may act as a gate controlling SSC proliferation or differentiation.

Molecular BioSystems Accepted Manuscript

 This journal is © The Royal Society of Chemistry 2012

Journal Name RSCPublishing

ARTICLE

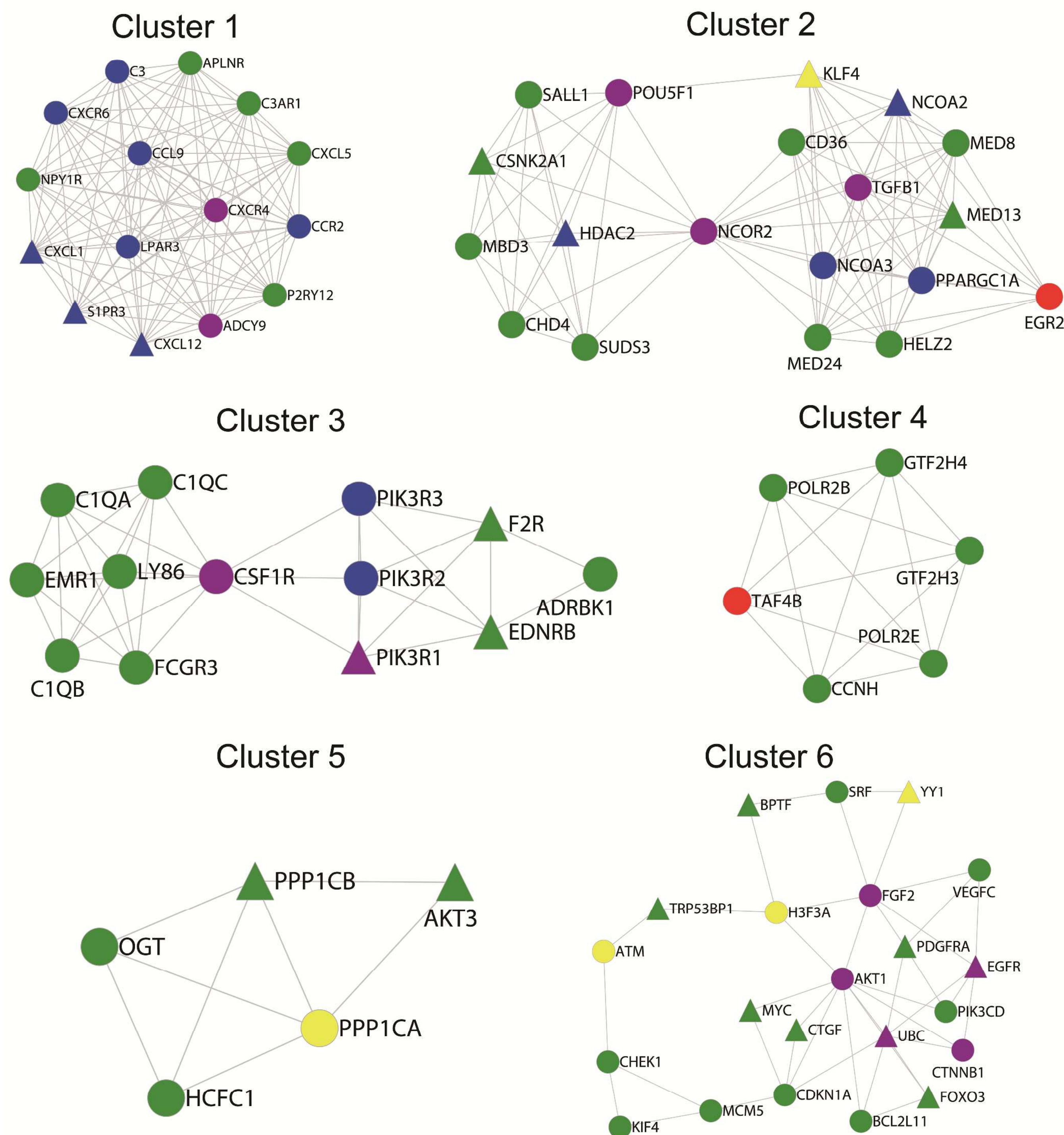


Fig. 5 Clusters of the whole PPI network for SSC self-renewal. After clustering the whole PPI network for SSC self-renewal, seven closely connected regions were directly extracted from the SSC self-renewal PPI network.

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 2013

Journal Name RSCPublishing

ARTICLE

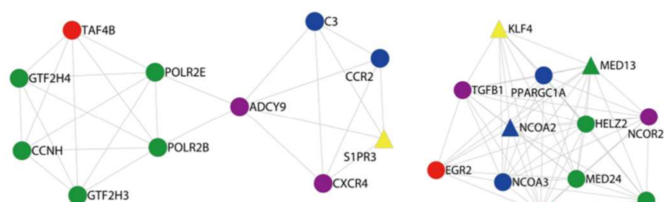


Fig. 6 Cluster from the subnetwork of the SSC self-renewal PPI network. The two clusters consisted of three independent modules that were similar to the combination of clusters 1, 2, and 4 from the whole PPI network for SSC self-renewal.

As the heterogeneous of SSC, there is no criterion on defining true SSC. SSC identified by single marker maybe not exactly represent the true SSC. Not only Thy1 but also *Gfra1 and Id4* positive SSC comprise A_s, A_{pr} and A_{al} spermatogonia with slight differences in gene expression.[60, 61] Thus, if there are datasets of true SSC transcriptome, the prediction on mechanisms of SSC self-renewal will be more accuracy. Considering the heterogeneous of SSC, we selected the same one marker purified SSC as data sources. In this manuscript, the constructed PPI network still can provide some clues for mechanisms of SSC self-renewal.

Conclusion

It is feasible to analyze interactions between dozens of key genes by combining information in PPI databases with the high-throughput data of gene expression profiles. We adopted a systems biology method to construct a continuous PPI network that mediates SSC self-renewal. PPI analysis and gene expression studies may provide a better understanding of the mechanisms responsible for SSC self-renewal.

Acknowledgements

This work was supported by National Basic Research Program of China (2013CB967401 and 2010CB945001), the National Nature Science Foundation of China (81370675, 81200472 and 81121001), and Shanghai Jiao Tong University Medicine-Engineering Fund (YG2013ZD04).

Notes

[a] Key Laboratory for the Genetics of Developmental & Neuropsychiatric Disorders (Ministry of Education), Bio-X Institutes, Shanghai Jiao Tong University, Shanghai 200240, China.
Email: jiwu@sjtu.edu.cn. Tel: 86-21-34207263.
[b] Key Laboratory of Fertility Preservation and Maintenance of Ministry of Education, Ningxia Medical University, Yinchuan 750000, China

Electronic Supplementary Information (ESI) available: Table S1 shows the result of PolySearch including returned 'genes/protein' and corresponding Pubmed ID of cited ariticles. Table S2 shows the list of 23 seed proteins interacting partners. Table S3 shows the list of SSC diferently expressed genes. Table S4 shows the list of high BC nodes and their CC values in SSCs self-renewal PPI network. Table S5 shows the list of large degree nodes and their CC values in SSCs self-renewal PPI network. Table S6 shows the list of high BC value nodes and their CC values in subnetwork.
[details of any supplementary information available should be included in supplementary data.doc].

References

1. M. I. G. a. J. A. V. Flavia Regina Oliveira de Barros, *ISBN: 978-953-51-0096-6*, 2012, 303-304.
2. J. M. Oatley and R. L. Brinster, *Annual Review of Cell and Developmental Biology*, 2008, **24**, 263-286.
3. M. Nagano, M. R. Avarbock and R. L. Brinster, *Biology of Reproduction*, 1999, **60**, 1429-1436.
4. M. Kanatsu-Shinohara, N. Ogonuki, K. Inoue, H. Miki, A. Ogura, S. Toyokuni and T. Shinohara, *Biology of Reproduction*, 2003, **69**, 612-616.
5. J. M. Oatley, M. R. Avarbock, A. I. Telaranta, D. T. Fearon and R. L. Brinster, *Proceedings of the National Academy of Sciences*, 2006, **103**, 9524-9529.
6. K. von Kopylow, C. Kirchhoff, D. Jezek, W. Schulze, C. Feig, M. Primig, V. Steinkraus and A.-N. Spiess, *Human Reproduction*, 2010, **25**, 1104-1112.
7. S. R. Singh, O. Burnicka-Turek, C. Chauhan and S. X. Hou, *Journal of Cellular and Molecular Medicine*, 2011, **15**, 468-483.
8. Q.-E. Yang, D. Kim, A. Kaucher, M. J. Oatley and J. M. Oatley, *Journal of Cell Science*, 2013, **126**, 1009-1020.
9. J. M. Oatley, M. J. Oatley, M. R. Avarbock, J. W. Tobias and R. L. Brinster, *Development*, 2009, **136**, 1191-1199.
10. H. Kubota, M. R. Avarbock and R. L. Brinster, *Proceedings of the National Academy of Sciences*, 2003, **100**, 6487-6492.
11. M. Kanatsu-Shinohara, S. Toyokuni and T. Shinohara, *Biology of Reproduction*, 2004, **70**, 70-75.
12. B.-Y. Ryu, K. E. Orwig, H. Kubota, M. R. Avarbock and R. L. Brinster, *Developmental Biology*, 2004, **274**, 158-170.
13. M. Tokuda, Y. Kadokawa, H. Kurahashi and T. Marunouchi, *Biology of Reproduction*, 2007, **76**, 130-141.
14. A. Buageaw, M. Sukhwani, A. Ben-Yehudah, J. Ehmcke, V. Y. Rawe, C. Pholpramool, K. E. Orwig and S. Schlatt, *Biology of Reproduction*, 2005, **73**, 1011-1016.
15. M. Seandel, D. James, S. V. Shmelkov, I. Falciatori, J. Kim, S. Chavala, D. S. Scherr, F. Zhang, R. Torres, N. W. Gale, G. D.

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry 2013

Yancopoulos, A. Murphy, D. M. Valenzuela, R. M. Hobbs, P. P. Pandolfi and S. Rafii, *Nature*, 2007, **449**, 346-350.
16. Z. He, M. Kokkinaki, J. Jiang, I. Dobrinski and M. Dym, *Biology of Reproduction*, 2010, **82**, 363-372.
17. C. K. Naughton, S. Jain, A. M. Strickland, A. Gupta and J. Milbrandt, *Biology of Reproduction*, 2006, **74**, 314-321.
18. K. Ishii, M. Kanatsu-Shinohara, S. Toyokuni and T. Shinohara, *Development*, 2012, **139**, 1734-1743.
19. J. Wu, Y. Zhang, G. G. Tian, K. Zou, C. M. Lee, Q. Yu and Z. Yuan, *Cellular Signalling*, 2008, **20**, 1052-1060.
20. L. Braydich-Stolle, N. Kostereva, M. Dym and M.-C. Hofmann, *Developmental Biology*, 2007, **304**, 34-45.
21. J. Lee, M. Kanatsu-Shinohara, K. Inoue, N. Ogonuki, H. Miki, S. Toyokuni, T. Kimura, T. Nakano, A. Ogura and T. Shinohara, *Development*, 2007, **134**, 1853-1859.
22. J. M. Oatley, M. R. Avarbock and R. L. Brinster, *Journal of Biological Chemistry*, 2007, **282**, 25842-25851.
23. F. W. Buaas, A. L. Kirsh, M. Sharma, D. J. McLean, J. L. Morris, M. D. Griswold, D. G. de Rooij and R. E. Braun, *Nat Genet*, 2004, **36**, 647-652.
24. J. A. Costoya, R. M. Hobbs, M. Barna, G. Cattoretti, K. Manova, M. Sukhwani, K. E. Orwig, D. J. Wolgemuth and P. P. Pandolfi, *Nat Genet*, 2004, **36**, 653-659.
25. Robin M. Hobbs, S. Fagoonee, A. Papa, K. Webster, F. Altruda, R. Nishinakamura, L. Chai and Pier P. Pandolfi, *Cell stem cell*, 2012, **10**, 284-298.
26. C. T. Dann, A. L. Alvarado, L. A. Molyneux, B. S. Denard, D. L. Garbers and M. H. Porteus, *STEM CELLS*, 2008, **26**, 2928-2937.
27. M. J. Oatley, A. V. Kaucher, K. E. Racicot and J. M. Oatley, *Biology of Reproduction*, 2011, **85**, 347-356.
28. X. Wu, J. M. Oatley, M. J. Oatley, A. V. Kaucher, M. R. Avarbock and R. L. Brinster, *Biology of Reproduction*, 2010, **82**, 1103-1111.
29. Y. Unhavaithaya, Y. Hao, E. Beyret, H. Yin, S. Kuramochi-Miyagawa, T. Nakano and H. Lin, *Journal of Biological Chemistry*, 2009, **284**, 6507-6519.
30. A. Sada, A. Suzuki, H. Suzuki and Y. Saga, *Science*, 2009, **325**, 1394-1398.
31. P. R. Hiesinger and B. A. Hassan, *Cell*, 2005, **123**, 1173-1174.
32. F. Chalmel, A. Lardenois, B. Evrard, R. Mathieu, C. Feig, P. Demougin, A. Gattiker, W. Schulze, B. Jégou, C. Kirchhoff and M. Primig, *Human Reproduction*, 2012, **27**, 3233-3248.
33. D. Cheng, C. Knox, N. Young, P. Stothard, S. Damaraju and D. S. Wishart, *Nucleic Acids Research*, 2008, **36**, W399-W405.
34. M. J. Cowley, M. Pinese, K. S. Kassahn, N. Waddell, J. V. Pearson, S. M. Grimmond, A. V. Biankin, S. Hautaniemi and J. Wu, *Nucleic Acids Research*, 2012, **40**, D862-D865.
35. S. Razick, G. Magklaras and I. Donaldson, *BMC Bioinformatics*, 2008, **9**, 405.
36. A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering and L. J. Jensen, *Nucleic Acids Research*, 2013, **41**, D808-D815.
37. X. Wu, S. M. Goodyear, J. W. Tobias, M. R. Avarbock and R. L. Brinster, *Biology of Reproduction*, 2011, **85**, 1114-1123.
38. X. Wu, J. A. Schmidt, M. R. Avarbock, J. W. Tobias, C. A. Carlson, T. F. Kolon, J. P. Ginsberg and R. L. Brinster, *Proceedings of the National Academy of Sciences*, 2009.
39. L. Gautier, L. Cope, B. M. Bolstad and R. A. Irizarry, *Bioinformatics*, 2004, **20**, 307-315.
40. G. K. Smyth, *Stat Appl Genet Mol Biol*, 2004, **3**, 12.
41. K. Raman, *Automated Experimentation*, 2010, **2**, 2.
42. J. Ran, H. Li, J. Fu, L. Liu, Y. Xing, X. Li, H. Shen, Y. Chen, X. Jiang, Y. Li and H. Li, *BMC Systems Biology*, 2013, **7**, 32.
43. R. Saito, M. E. Smoot, K. Ono, J. Ruscheinski, P.-L. Wang, S. Lotia, A. R. Pico, G. D. Bader and T. Ideker, *Nat Meth*, 2012, **9**, 1069-1076.
44. A. E. Falender, R. N. Freiman, K. G. Geles, K. C. Lo, K. Hwang, D. J. Lamb, P. L. Morris, R. Tjian and J. S. Richards, *Genes & Development*, 2005, **19**, 794-803.
45. M. J. Goertz, Z. Wu, T. D. Gallardo, F. K. Hamra and D. H. Castrillon, *The Journal of Clinical Investigation*, 2011, **121**, 3456-3466.
46. K. Takubo, M. Ohmura, M. Azuma, G. Nagamatsu, W. Yamada, F. Arai, A. Hirao and T. Suda, *Cell stem cell*, 2008, **2**, 170-182.
47. S. Giovanni and L. Carlo, *University of Verona, Center for BioMedical Computing, Italy*.
48. G. Bader and C. Hogue, *BMC Bioinformatics*, 2003, **4**, 2.
49. G. Dennis, B. Sherman, D. Hosack, J. Yang, W. Gao, H. Lane and R. Lempicki, *Genome Biol*, 2003, **4**, R60.
50. G. Lima-Mendez and J. van Helden, *Molecular BioSystems*, 2009, **5**, 1482-1493.
51. D. L. C. van den Berg, T. Snoek, N. P. Mullin, A. Yates, K. Bezstarosti, J. Demmers, I. Chambers and R. A. Poot, *Cell stem cell*, 2010, **6**, 369-381.
52. J. B. Kim, V. Sebastiano, G. Wu, M. J. Araúzo-Bravo, P. Sasse, L. Gentile, K. Ko, D. Ruau, M. Ehrich, D. van den Boom, J. Meyer, K. Hübner, C. Bernemann, C. Ortmeier, M. Zenke, B. K. Fleischmann, H. Zaehres and H. R. Schöler, *Cell*, 2009, **136**, 411-419.
53. U. Elling, C. Klasen, T. Eisenberger, K. Anlag and M. Treier, *Proceedings of the National Academy of Sciences*, 2006, **103**, 16319-16324.
54. J. Yang, L. Chai, T. C. Fowles, Z. Alipio, D. Xu, L. M. Fink, D. C. Ward and Y. Ma, *Proceedings of the National Academy of Sciences*, 2008.
55. J. Zhang, W.-L. Tam, G. Q. Tong, Q. Wu, H.-Y. Chan, B.-S. Soh, Y. Lou, J. Yang, Y. Ma, L. Chai, H.-H. Ng, T. Lufkin, P. Robson and B. Lim, *Nat Cell Biol*, 2006, **8**, 1114-1123.
56. Q. Zhou, H. Chipperfield, D. A. Melton and W. H. Wong, *Proceedings of the National Academy of Sciences*, 2007, **104**, 16438-16443.
57. O. Yildirim, R. Li, J.-H. Hung, Poshen B. Chen, X. Dong, L.-S. Ee, Z. Weng, Oliver J. Rando and Thomas G. Fazzio, *Cell*, 2011, **147**, 1498-1510.
58. S. Ye, P. Li, C. Tong and Q. L. Ying, *The EMBO Journal*, 2013, **32**, 2548-2560.
59. B. M. Hacker, J. E. Tomlinson, G. A. Wayman, R. Sultana, G. Chan, E. Villacres, C. Disteche and D. R. Storm, *Genomics*, 1998, **50**, 97-104.

Molecular BioSystems Accepted Manuscript

60. T. Nakagawa, M. Sharma, Y.-i. Nabeshima, R. E. Braun and S. Yoshida, *Science*, 2010, **328**, 62-67.
61. F. Chan, M. J. Oatley, A. V. Kaucher, Q.-E. Yang, C. J. Bieberich, C. S. Shashikant and J. M. Oatley, *Genes & Development*, 2014, **28**, 1351-1362.

Molecular BioSystems Accepted Manuscript

 This journal is © The Royal Society of Chemistry 2012