

StemChecker: a web-based tool to discover and explore stemness signatures in gene sets

José P. Pinto^{1,*}, Ravi K. Kalathur^{1,2}, Daniel V. Oliveira¹, Tânia Barata¹, Rui S.R. Machado¹, Susana Machado¹, Ivette Pacheco-Leyva¹, Isabel Duarte¹ and Matthias E. Futschik^{1,3,*}

¹Systems Biology and Bioinformatics Laboratory (SysBioLab), University of Algarve, Faro, Algarve, 8005-139, Portugal, ²Experimental and Clinical Cell Therapy Institute, Spinal Cord and Tissue Regeneration Center Salzburg, Paracelsus Medizinische Privatuniversität, Salzburg, A-5020, Austria and ³Centre of Marine Sciences (CCMAR), University of Algarve, Faro, Algarve, 8005-139, Portugal

Received March 06, 2015; Revised April 30, 2015; Accepted May 10, 2015

ABSTRACT

Stem cells present unique regenerative abilities, offering great potential for treatment of prevalent pathologies such as diabetes, neurodegenerative and heart diseases. Various research groups dedicated significant effort to identify sets of genes—so-called *stemness signatures*—considered essential to define stem cells. However, their usage has been hindered by the lack of comprehensive resources and easy-to-use tools. For this we developed StemChecker, a novel *stemness* analysis tool, based on the curation of nearly fifty published *stemness signatures* defined by gene expression, RNAi screens, Transcription Factor (TF) binding sites, literature reviews and computational approaches. StemChecker allows researchers to explore the presence of *stemness signatures* in user-defined gene sets, without carrying-out lengthy literature curation or data processing. To assist in exploring underlying regulatory mechanisms, we collected over 80 target gene sets of TFs associated with pluri- or multipotency. StemChecker presents an intuitive graphical display, as well as detailed statistical results in table format, which helps revealing transcriptionally regulatory programs, indicating the putative involvement of stemness-associated processes in diseases like cancer. Overall, StemChecker substantially expands the available repertoire of online tools, designed to assist the stem cell biology, developmental biology, regenerative medicine and human disease research community. StemChecker is freely accessible at <http://stemchecker.sysbiolab.eu>.

INTRODUCTION

Stem cells have been the focus of intense biomedical research in recent years. Their self-renewal ability, together with their potential to differentiate into other cell types represent particularly attractive features, not only for clinical applications in the field of regenerative medicine, but also for the study of fundamental processes like embryology and the development of complex multi-cellular organs (1). Moreover, their reported involvement in neurodegenerative diseases (2,3), diabetes (4) and cancer (5,6) have raised considerable interest in stem cell biology across a wide range of biomedical research fields. Despite the great progress achieved in stem cell biology in the last decade, many of the stem cells' features still await full clarification. Of key importance is the elucidation of the genetic program that underlies the core properties of stem cells, i.e. the capacity for self-renewal and for generation of differentiated progeny—or in short, their *stemness*. Although it is possible nowadays to generate stem cells *in vitro*, it is still not fully clear how *stemness* is established and maintained (7). To gain a better understanding of this issue, various research groups applied a diverse range of approaches to identify the set of genes—so-called *stemness signatures*—that are considered necessary, or at least associated with the defining characteristics of stem cells (8–12). Although no 'universal' *stemness signature* has been established, the proposed signatures have proven to be invaluable resources for the study of stem cell generation, maintenance and differentiation (10–11,13). Furthermore, they have been found to be highly informative indicators for the study of diseases such as cancer (14). Despite its undeniable potential, the usage of *stemness signatures* has been hampered by the lack of a comprehensive resource, as well as easy-to-use analysis tools. Most published signatures are hidden as supporting materials or scattered across multiple repositories, limiting a broader access and wider usage. Additionally, these signatures can be substantially distinct, particularly the ones derived by dif-

*To whom correspondence should be addressed. Tel: +351 289 800 100 (Ext7292); Fax: +351 289 818 419; Email: jppinto@ualg.pt
Correspondence may also be addressed to Matthias Futschik. Tel: +351 289 244 403; Fax: +351 289 818 419; Email: mfutschik@ualg.pt

ferent experimental approaches, making the analysis and comparison between signature gene sets a non-trivial task (9).

For this reason, we have developed **StemChecker** (freely accessible at <http://stemchecker.sysbiolab.eu>). This web server is based on the most extensive and up-to-date curation of published *stemness signatures*. It enables researchers to easily explore the presence of *stemness signatures* in their gene lists, without the burden of having to carry out lengthy literature curation, data processing and statistical analysis. Additionally, to assist in exploring the underlying regulatory mechanisms, we collected and integrated a large number of target gene sets of transcription factors (TFs) associated with stem cell identity and pluripotency.

DATA COLLECTION AND DATA SET CURATION

To obtain a comprehensive collection of *stemness signatures*, we carried out an extensive review of published studies, in which human or mouse *stemness signatures* were reported. Additionally, we surveyed the literature for studies that described human or murine gene sets associated with stem cell identity and maintenance. To further expand the underlying data sets, we queried publicly accessible resources for genes annotated as being related to stem cells and pluripotency. Finally, we retrieved the results from published ChIP-chip and ChIP-Seq studies, where known stem cell-related transcription factors have been investigated, both in human and in mouse stem cells. In total, we collected and curated 132 *stemness signatures* and transcription factor target gene sets.

Stemness signatures

The *stemness signatures* present in **StemChecker** are classified into five major categories, depending on their source:

1. *Transcription factor target genes* containing only the positively regulated genes that are targeted by key transcriptional regulators such as OCT4 (POU5F1), NANOG and SOX2 for embryonic stem cells (ESC) (15).
2. *Expression profiles* containing 34 sets of up-regulated genes in nine stem cell types: ESC, Hematopoietic Stem Cells (HSC), Mesenchymal Stem Cells (MSC), Embryonal Carcinoma (EC), Mammary Stem Cells (MaSC), Neural Stem Cells (NSC), Intestinal Stem Cells (ISC), induced Pluripotent Stem Cells (iPSC) and Spermatogonial Stem Cells (SSC).
3. *RNAi screens* including 5 sets from genome wide RNAi screening experiments for genes essential for self-renewal (16–20).
4. *Literature curation* including gene sets extracted from publicly accessible resources such as Reactome, KEGG, PluriNetWork and HSC-Explorer that were based on independent curation of published studies (21,22).
5. *Computationally derived* genes sets collected from two resources: PluriNet (based on computational network analysis) and GeneCards database (based on text-mining).

Even though the different data sources resulted in divergent gene sets, many pairs of *stemness signatures* show

highly significant overlap (Supporting Figures S1 and S2). This finding suggests that the data integration in **StemChecker** can help to identify subsets of genes whose association with *stemness* is supported by multiple independent evidence.

Transcription factor gene sets

Transcription factor gene sets encompass target genes from 46 human and mouse TFs that are known to play an important role in stem cell differentiation and maintenance. A total of 11331 regulatory interactions for human and 166286 for mouse are contemplated in these data sets, providing the user with a straightforward and powerful way of finding potential transcription regulators, active in stem cells, for their genes of interest.

Further details on the data sets can be obtained at the ‘Browse Sets’ tab in **StemChecker** and in supporting Table S1.

IMPLEMENTATION

StemChecker was created using a combination of JavaScript and JavaServer Faces (JSF) 2.1, a Java-based framework for the development of user interfaces. JSF’s functions were extended through the PrimeFaces library. Curated data sets are stored in a MySQL database. Communication between Java and MySQL components was implemented through the Hibernate library. Radar charts are generated using Chart.js (<http://www.chartjs.org>), an open source JavaScript library for creating charts. The sunburst chart is produced employing D3.js (<http://d3js.org/>), a JavaScript library for manipulating documents based on data, and code obtained from the Sequences sunburst chart (<http://bl.ocks.org/kerryrodden/7090426>).

STEMCHECKER USAGE

With **StemChecker** we have created a user-friendly online application that can be used both by experienced bioinformaticians as well as researchers with little expertise in computer science.

Workflow

We developed a workflow for **StemChecker** based on four simple steps: *Input–Select–Search–Interactive Analysis* (Figure 1).

In the first step the user inputs the identifiers for their gene(s) of interest. **StemChecker** currently accepts official gene symbols, EntrezGene ids, or a mix of both.

The second step consists in selecting the data sets that should be included in the search, i.e., the user can either restrict the search to the *stemness signatures* that best suit their research interest, or alternatively, search all available data sets. The default searching options select all human data sets, but the user can alternatively search only mouse only data sets, or both organisms (in which case, only EntrezGene ids can be used as query).

In the third step the user initiates the search and analysis by pressing the *Submit* button.



Figure 1. StemChecker Workflow. (1) The user inputs a gene set and (2) selects the *stemness signature* data sets that should be searched, the ordering options for the checkerboards and the organism. (3) The query is submitted, and (4) the results are presented in three tabs: the *Overview/Stats* tab gives a general view of the results through a combination of radar charts and tables with the statistical significance of enrichment in *Stem Cell types*, *Stemness Signatures* and *Transcription Factors* (each in its individual sub-tab); the *Stemness Match* tab contains a checkerboard that shows the presence of the queried genes in each *stemness signature*; finally the *Transcription Factor Match* tab contains a similar checkerboard showing the presence of individual genes in each TF data set.

The final step of the workflow allows the user to interactively inspect the results which are displayed in three pages, each in its own tab: (i) *Overview/Stats* showing the overlap of the input genes with **StemChecker**'s data sets and their statistical significance; (ii) *Stemness Signature Match* displaying a checkerboard-like table, which indicates the occurrence of each individual query gene in the different *stemness signatures* and (iii) *Transcription Factor Match* showing, which of the input genes are targets of transcription factors based on the curated data sets.

The *Overview/Stats* tab shows a summary of the successful searches performed, informing the user about the number of valid input genes, invalid ones and the number of *Stemness Signatures* and TF data sets queried. Below the summary, individual results for Stem Cell types, *Stemness Signatures* and Transcription Factors are presented in three separate tabs, each one displaying one radar chart and one table with detailed statistics. The radar charts are an intuitive and fast way of visualizing the overlap between the user's input genes and the data sets in **StemChecker**. By default, the values plotted represent the significance of the overlap ($-\log_{10} P$ -value). The *Switch* button plots the fractional overlap, which can range from 0 to 100%, with each grid representing a 10% interval. The exact values are shown when the pointer is placed over the generated plot. The main purpose of these plots is to provide an overview of the global results and to give the user an instant indication whether their input genes display any tendency toward a particular *stemness signature*, transcription factor or stem cell type.

Tables below the charts display the underlying statistical details i.e. individual results for particular stem cell types, *stemness signatures* and transcription factors. For each data set, its total size, the number of overlapping genes, the statistical significance of the overlap (P -value calculated by the Hypergeometric test) and the Bonferroni adjusted P -value are shown.

All the images and statistical results from **StemChecker**'s analysis are downloadable via the *Download* button below the respective image/table.

The *Stemness Match* and *Transcription Factor Match* tabs, both display a checkerboard-like table, indicating the presence of each valid input gene in each of the selected *stemness* or transcription factor data sets. Briefly, each row represents one query gene, and each column one selected data set. Each blue rectangle in this table indicates the presence of the input gene in the corresponding data set. By default, the checkerboards are ordered so that genes (rows) with more matches will be displayed on top, and gene sets (columns) with the most genes matched will be located to the left. Both these settings can be changed in the Checkerboard Options from the Analysis tab. Additionally, the user can fully customize the checkerboards to display only the most relevant data pertaining to its particular research interests by hiding the rows and columns that show superfluous information, followed by the download of the personalized image in PDF format.

This presentation is very intuitive, and it provides the user with a quick way of visualizing the data sets that share the most genes with the input gene list. It also indicates to the user if a subset of the input genes is found in common *stemness signatures* or is under the control of a common TF. This

indication helps to identify the genes that were responsible for detected associations with *stemness signatures* or TF regulons.

Case Studies

To illustrate the utility of **StemChecker** for the detection of *stemness signatures* and regulatory programs in user-defined gene lists, we used **StemChecker** for the analysis of two different cases: (i) a set of prognostic genes for the clinical outcome of Pancreatic Ductal Adenocarcinoma (PDAC) and (ii) up-regulated genes during *in vitro* differentiation of murine ESCs.

Pancreatic ductal adenocarcinoma. PDAC is one of the deadliest solid cancers and it is characterized by an extremely low survival rate, and high resistance to chemotherapy. In a recent meta-analysis of expression profiles from 466 PDAC patients, a set of 225 genes with prognostic value for the patient's survival time could be derived (23). From this set, we extracted 180 genes, whose up-regulation was correlated with a short survival time, and used these genes as input for **StemChecker**. Remarkably, we found that 59 genes have been included in the collected *stemness signatures* for human ESCs. This is a highly significant enrichment (adjusted P -value = 3.99×10^{-8}) of ESC-associated genes (Figure 2A, Supporting Table S2). In contrast, no significant over-representation of *stemness signatures* was found for prognostic genes, whose down-regulation correlated with shorter survival time (Supporting Table S3). A general activation of ESC-associated genes in PDAC was also indicated by the evaluation of up-regulated genes in matching pairs of tumor and adjacent non-tumor tissues as detected in an independent study (24) (Supporting Figure S3; Supporting Table S4). Although closer examination is certainly warranted, these observations suggest that the activation of *stemness*-related mechanisms might play a role in PDAC—a hypothesis which has been put forward previously (25).

In vitro differentiation of murine ESCs. To exemplify how **StemChecker** can help to reveal underlying genetic regulatory programs, we analyzed a time-series experiment for the differentiation of murine ESCs (8). In this experiment, differentiation of murine ESCs was induced by the 'hanging drop' method, which led to the formation of embryoid bodies (EBs). Here, we compared the enrichment of up-regulated genes among TF target genes at different time points (Supporting Table S5). For day zero, expectably, we observed a highly significant overrepresentation of targets of the pluripotency master regulators Oct4, Sox2 and Nanog (Figure 2B). This is contrasted by results for up-regulated genes in EBs at day 3 after induction of differentiation (Figure 2C, Supporting Table S6). For this time point, we detected that up-regulated genes tended to be target genes of components of the polycomb complexes (Suz12, Ezh2, Ring1B) in ESCs. This finding corroborates previous observations that polycomb complexes function as repressors of genes with a role in embryonic development (26) and illustrates the utility of **StemChecker** to elucidate transcriptional regulation in stem cell biology.

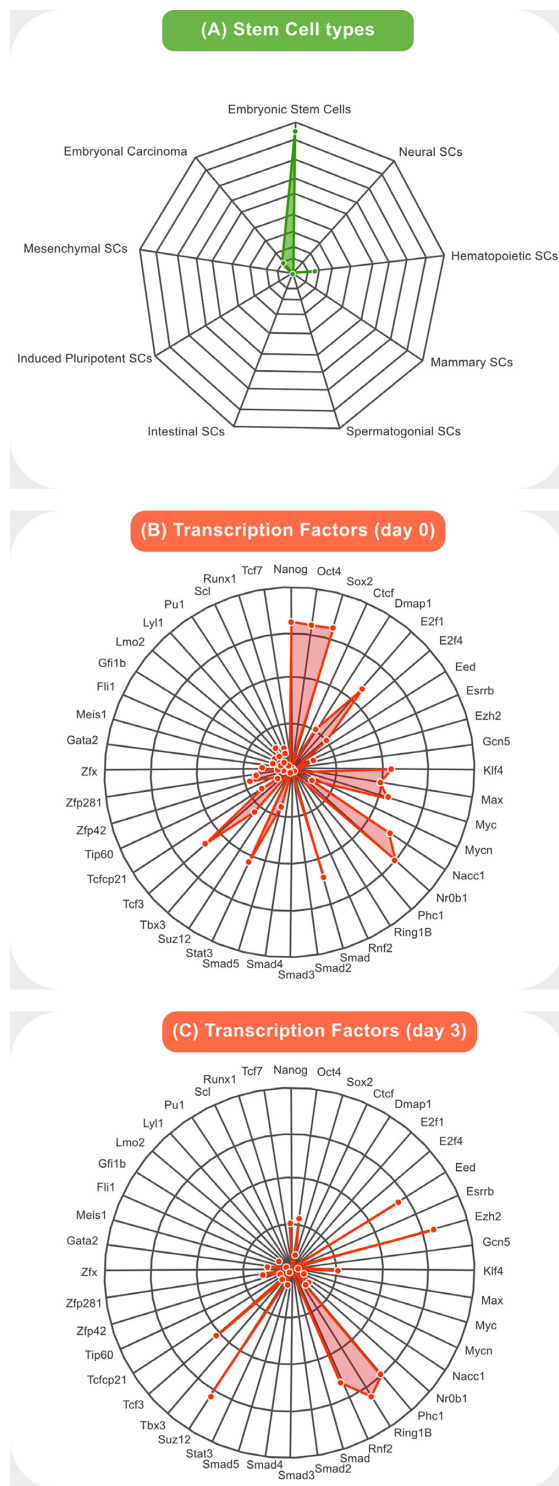


Figure 2. StemChecker Case Studies. (A) Analysis of a set of 180 genes up-regulated in pancreatic ductal adenocarcinoma patients with poor survival prognosis. This radar chart displays a highly significant enrichment for genes associated with Embryonic Stem Cell Type. (B) Analysis of a set of up-regulated genes in a study of *in vitro* differentiation of mouse embryonic stem cells. In time point zero, the chart shows that this gene set (i.e. up-regulated in undifferentiated ESC) is significantly enriched in targets of Nanog, Sox2 and Oct4 (the pluripotency master regulators) (C) At day 3 of the induced *in vitro* differentiation, up-regulated genes are mostly enriched in targets from Suz12, Ezh2 and Ring1B (components of polycomb complexes) in ESC.

FUTURE DIRECTIONS

StemChecker aims to be an efficient and user-friendly web server with substantial impact on the rapidly expanding field of stem cell biology and related areas. As such, we are continuously improving its data sources, through the addition of new published *stemness* and transcription factor signatures, all properly curated prior to their inclusion online.

Additionally, new ways of displaying data and analysis tools have been the major focus of our development team, in order to provide the user with intuitive, clear and concise ways of exploring the data sets and the analyses' results. Some new functionalities are scheduled to be added in the near future, particularly the implementation of quantitative *stemness measures* (currently under development), and the possibility of simultaneous analysis of several gene lists for comparative studies. Another feature that will be implemented is the option to mask genes associated with cell cycle and proliferation from analysis. Such masking can help to clarify whether significant enrichment was detected mainly due to a high content of proliferative genes in the input list - a scenario, which might arise, for instance, for genes lists associated with fast growing tumors.

Finally, we invite all **StemChecker** users to submit their feedback and suggestions of new studies to include, so that we can continue to improve this web resource, tailoring it to the specific needs of the stem cell biology community.

CONCLUSION

Stem cells' ability to self-renew and differentiate into virtually all other cell types has gathered much interest by the research community, which saw them as a powerful tool for the understanding of fundamental processes, while picking up their potential for many biomedical applications. These studies eventually led to the search for the *stemness gene signatures* responsible for these cells' properties, generating a wealth of data, which most of the times is not easily accessible to the research community.

StemChecker was developed to address this limitation. It is a unique online tool customized to meet the needs of the stem cell research community. It functions as a 'first stop' resource where researchers can rapidly check whether their genes of interest can be associated with *stemness*; which known pluri- or multipotency-associated transcription factors regulate them and which stem cell types express them.

Its functionalities proceed from a comprehensive set of 49—and counting—manually curated *stemness signatures*, and over 80 published transcription factor target gene sets associated with *stemness*. **StemChecker** combines and integrates these data, providing an easy-to-use interface, which displays the overall results in an intuitive and meaningful way, while still providing a detailed individual gene analysis, together with the statistical significance of the overlap.

StemChecker substantially expands the available repertoire of stem cell-related online analysis tools, offering the research community the ability to quickly check and explore the *stemness genetic signatures* contained in their gene sets, which greatly complements existing tools for integrative stem cell biology such as ESCAPE (27) or StemCellNet (28).

SUPPORTING DATA

Supporting Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Ramiro Magno for his priceless technical help, Franklim Barata for designing the **StemChecker** logo, Joaquín Lamia for stimulating discussions, the referees for their detailed and constructive criticism, and the valuable help from our experimental colleagues, which have tested **StemChecker** and requested many of its current features.

FUNDING

Portuguese Fundação para a Ciência e Tecnologia (FCT) [SFRH/BPD/96890/2013 to J.P.P., SFRH/BPD/70718/2010 to R.K.K., PTDC/BIA/GEN/116519/2010 and FCT Investigator Grant IF/00881/2013 to M.E.F.]; Programa Doutoral ProRegeM—Mecanismos de Doença e Medicina Regenerativa [PD/00117/2012 to R.R.M.]. Funding for open access charge: Portuguese Fundação para a Ciência e a Tecnologia [PTDC/BIA-GEN/116519/2010].
Conflict of interest statement. None declared.

REFERENCES

- Murry,C.E. and Keller,G. (2008) Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. *Cell*, **132**, 661–680.
- Lunn,J.S., Sakowski,S.A. and Feldman,E.L. (2014) Concise review: stem cell therapies for amyotrophic lateral sclerosis: recent advances and prospects for the future. *Stem Cells*, **32**, 1099–1109.
- Kalathur,R.K., Hernandez-Prieto,M.A. and Futschik,M.E. (2012) Huntington's disease and its therapeutic target genes: a global functional profile based on the HD Research Crossroads database. *BMC Neurol.*, **12**, 47.
- Dave,S. (2014) Mesenchymal stem cells derived in vitro transdifferentiated insulin-producing cells: A new approach to treat type 1 diabetes. *Adv. Biomed. Res.*, **3**, 266.
- Nguyen,L.V., Vanner,R., Dirks,P. and Eaves,C.J. (2012) Cancer stem cells: an evolving concept. *Nat. Rev. Cancer*, **12**, 133–143.
- Glinsky,G.V. (2008) 'Stemness' genomics law governs clinical behavior of human cancer: implications for decision making in disease management. *J. Clin. Oncol.*, **26**, 2846–2853.
- Cyranoski,D. (2014) Stem cells: the black box of reprogramming. *Nature*, **516**, 162–164.
- Gaspar,J.A., Doss,M.X., Winkler,J., Wagh,V., Hescheler,J., Kolde,R., Vilo,J., Schulz,H. and Sachinidis,A. (2012) Gene expression signatures defining fundamental biological processes in pluripotent, early, and late differentiated embryonic stem cells. *Stem Cells Dev.*, **21**, 2471–2484.
- Fortunel,N.O., Otu,H.H., Ng,H.H., Chen,J., Mu,X., Chevassut,T., Li,X., Joseph,M., Bailey,C., Hatzfeld,J.A. *et al.* (2003) Comment on 'Stemness': transcriptional profiling of embryonic and adult stem cells' and 'a stem cell molecular signature'. *Science*, **302**, 393.
- Ivanova,N.B., Dimos,J.T., Schaniel,C., Hackney,J.A., Moore,K.A. and Lemischka,I.R. (2002) A stem cell molecular signature. *Science*, **298**, 601–604.
- Ramalho-Santos,M., Yoon,S., Matsuzaki,Y., Mulligan,R.C. and Melton,D.A. (2002) 'Stemness': transcriptional profiling of embryonic and adult stem cells. *Science*, **298**, 597–600.
- Sperger,J.M., Chen,X., Draper,J.S., Antosiewicz,J.E., Chon,C.H., Jones,S.B., Brooks,J.D., Andrews,P.W., Brown,P.O. and Thomson,J.A. (2003) Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 13350–13355.
- Cai,J., Weiss,M.L. and Rao,M.S. (2004) In search of 'stemness'. *Exp. Hematol.*, **32**, 585–598.
- Shats,I., Gatz,M.L., Chang,J.T., Mori,S., Wang,J., Rich,J. and Nevins,J.R. (2011) Using a stem cell-based signature to guide therapeutic selection in cancer. *Cancer Res.*, **71**, 1772–1780.
- Boyer,L.A., Lee,T.I., Cole,M.F., Johnstone,S.E., Levine,S.S., Zuckerman,J.P., Guenther,M.G., Kumar,R.M., Murray,H.L., Jenner,R.G. *et al.* (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, **122**, 947–956.
- Chia,N.Y., Chan,Y.S., Feng,B., Lu,X., Orlov,Y.L., Moreau,D., Kumar,P., Yang,L., Jiang,J., Lau,M.S. *et al.* (2010) A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*, **468**, 316–320.
- Ding,L., Paszkowski-Rogacz,M., Nitzsche,A., Slabicki,M.M., Heninger,A.K., de Vries,I., Kittler,R., Junqueira,M., Shevchenko,A., Schulz,H. *et al.* (2009) A genome-scale RNAi screen for Oct4 modulators defines a role of the Paf1 complex for embryonic stem cell identity. *Cell Stem Cell*, **4**, 403–415.
- Wang,C.H., Ma,N., Lin,Y.T., Wu,C.C., Hsiao,M., Lu,F.L., Yu,C.C., Chen,S.Y. and Lu,J. (2012) A shRNA functional screen reveals Nme6 and Nme7 are crucial for embryonic stem cell renewal. *Stem Cells*, **30**, 2199–2211.
- Hu,G., Kim,J., Xu,Q., Leng,Y., Orkin,S.H. and Elledge,S.J. (2009) A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes Dev.*, **23**, 837–848.
- Yang,S.-H., Kalkan,T., Morrisroe,C., Smith,A. and Sharrocks,A.D. (2012) A genome-wide RNAi screen reveals MAP kinase phosphatases as key ERK pathway regulators during embryonic stem cell differentiation. *PLoS Genet.*, **8**, e1003112.
- Som,A., Harder,C., Greber,B., Siatkowski,M., Paudel,Y., Warsow,G., Cap,C., Scholer,H. and Fuellen,G. (2010) The PluriNetWork: an electronic representation of the network underlying pluripotency in mouse, and its applications. *PLoS One*, **5**, e15165.
- Montrone,C., Kokkaliaris,K.D., Dirk Loeffler,D., Lechner,M., Kastenmüller,G., Schroeder,T. and Ruepp,A. (2013) HSC-explorer: a curated database for hematopoietic stem cells. *PLoS one*, **8**, e70348.
- Haider,S., Wang,J., Nagano,A., Desai,A., Arumugam,P., Dumartin,L., Fitzgibbon,J., Hagemann,T., Marshall,J.F., Kocher,H.M. *et al.* (2014) A multi-gene signature predicts outcome in patients with pancreatic ductal adenocarcinoma. *Genome Med.*, **6**, 105.
- Zhang,G., Schetter,A., He,P., Funamizu,N., Gaedcke,J., Ghadimi,B.M., Ried,T., Hassan,R., Yfantis,H.G., Lee,D.H. *et al.* (2012) DPEP1 inhibits tumor cell invasiveness, enhances chemosensitivity and predicts clinical outcome in pancreatic ductal adenocarcinoma. *PLoS One*, **7**, e31507.
- Balic,A., Dorado,J., Alonso-Gómez,M. and Heeschen,C. (2012) Stem cells as the root of pancreatic ductal adenocarcinoma. *Exp. Cell Res.*, **318**, 691–704.
- Boyer,L.A., Plath,K., Zeitlinger,J., Brambrink,T., Medeiros,L.A., Lee,T.I., Levine,S.S., Wernig,M., Tajonar,A., Ray,M.K. *et al.* (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, **441**, 349–353.
- Xu,H., Baroukh,C., Dannenfeller,R., Chen,E.Y., Tan,C.M., Kou,Y., Kim,Y.E., Lemischka,I.R. and Ma'ayan,A. (2013) ESCAPE: database for integrating high-content published data collected from human and mouse embryonic stem cells. *Database*, bat045.
- Pinto,J.P., Reddy Kalathur,R.K., Machado,R.S., Xavier,J.M., Braganca,J. and Futschik,M.E. (2014) StemCellNet: an interactive platform for network-oriented investigations in stem cell biology. *Nucleic Acids Res.*, **42**, W154–W160.