

From RNAi Screens to Molecular Function in Embryonic Stem Cells

Li Ding · Ina Poser · Maciej Paszkowski-Rogacz · Frank Buchholz

Published online: 28 April 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The ability of embryonic stem (ES) cells to generate any of the around 220 cell types of the adult body has fascinated scientists ever since their discovery. The capacity to re-program fully differentiated cells into induced pluripotent stem (iPS) cells has further stimulated the interest in ES cell research. Fueled by this interest, intense research has provided new insights into the biology of ES cells in the recent past. The development of large-scale and high throughput RNAi technologies has made it possible to sample the role of every gene in maintaining ES cell identity. Here, we review the RNAi screens performed in ES cells to date and discuss the challenges associated with these large-scale experiments. Furthermore, we provide a perspective on how to streamline the molecular characterization following the initial phenotypic description utilizing bacterial artificial chromosome (BAC) transgenesis.

Keywords RNA interference · siRNA · shRNA · esiRNA · Genome-wide screen · Bacterial artificial chromosome · TransgeneOmics

Introduction

Embryonic stem cells (ESCs) are isolated from the inner cell mass (ICM) of mouse blastocysts [1]. Under appropriate conditions, ES cells can be kept undifferentiated for many passages, while maintaining their competence to differentiate into cells of all three germ layers [2]. These characteristics make ES cells invaluable for studying the molecular mechanisms of pluripotency and lineage differentiation processes for regenerative therapies of a variety of human diseases [3]. With the advance of cellular reprogramming techniques that include somatic cell nuclear transfer (SCNT) and induced pluripotent stem (iPS) cells, the investigation of pluripotency becomes even more important. iPS cells are generated from somatic cells by the forced expression of a few embryonic transcription factors [4]. Like ES cells, iPS cells express pluripotency markers including Oct4, Nanog and SSEA1, possess the potential to differentiate into cells of all three germ layers, and prominently contribute to chimeras [5]. iPS technology holds enormous promise as it raises the prospect for the generation of patient-specific and disease-specific stem cells, which one day may prove useful in custom-tailored cell therapies [reviewed in 6].

Cell fate choices in ES cells are regulated by a complex orchestration of multiple pathways. Details regarding the genetic and epigenetic regulations that control these remarkable features of pluripotent cells have started to emerge, and they include a molecular circuitry controlled by major transcription factors, epigenetic modifications, and post-transcriptional gene regulation [reviewed in 7]. Investigations of the interplay of transcription factors, epigenetic regulators, and miRNAs have substantially improved our understanding of ES cell biology. However, recent discoveries of new pluripotency factors and path-

L. Ding · M. Paszkowski-Rogacz · F. Buchholz (✉)
University Hospital Carl Gustav Carus and Medical Faculty,
University of Technology Dresden,
Fetscherstr. 74,
01307 Dresden, Germany
e-mail: buchholz@mpi-cbg.de

L. Ding · I. Poser · M. Paszkowski-Rogacz · F. Buchholz
Max-Planck-Institute for Molecular Cell Biology and Genetics,
Pfötenhauerstrasse 108,
01307 Dresden, Germany

ways suggest that our understanding of the full pluripotency network is still far from being complete [8–11].

Both gain-of-function and loss-of-function assays have been employed to identify genes that play a role in ES cell biology. In mammalian cells, gain-of-function screens usually utilize libraries of pooled cDNA clones, which require iterative rounds of selection to yield functional gene products that are identified by clone sequencing. Genome-wide gain-of-function screens have been successfully performed in ESCs, and discovered novel pluripotency genes including *Nanog*, *Akt1* and *RhoJ* [2, 12]. In contrast to the pooled cDNA libraries, recent advances of individual arrayed cDNA libraries allow screens to be performed with higher efficiency, complexity and sensitivity. Consequently, several novel regulators of pluripotency including *Timp2*, *Hig2*, and *Mki67ip* have recently been identified from an arrayed cDNA library [10].

Another powerful strategy, that has been employed extensively to investigate ES cell biology, is loss-of-function experiments. Gene targeting via homologous recombination [13] has been a powerful tool to generate knock-out ES cell lines to investigate functional consequences upon deletion for a number of genes [14–16]. However, the generation of targeting constructs and the selection and validation of clones is time-consuming and cost intensive, limiting the throughput for this approach.

RNA Interference Resources

The discovery of post-transcriptional gene silencing via RNA interference (RNAi) allows for rapid and efficient disruption of gene functions in cells, and has become one of the most popular methods for loss-of-function studies. In an RNAi reaction, the cellular RNase III enzyme Dicer cleaves the dsRNA into 21–25 bp small interfering RNAs (siRNAs). These siRNAs are then incorporated into the RNA-induced silencing complex (RISC), which subsequently interacts with the homologous mRNA in a sequence-specific manner, resulting in the degradation of target mRNA and hence decreased production of the corresponding protein [17].

In mammalian cells, the RNAi effect can be achieved by direct, transient introduction of siRNAs into the cytoplasm of cells via transfection or electroporation. siRNAs can be generated by chemical synthesis [17], or through enzymatic digestion of long dsRNA employing an endoribonuclease (esiRNAs) [18]. Alternatively, the silencing triggers can be expressed in cells as short hairpin RNAs (shRNAs), where they are processed to siRNAs by the cellular RNAi machinery [19].

The short nature of siRNAs allows for a flexible design of sequences to silence the target transcript. However, the silencing potency varies greatly between all possible

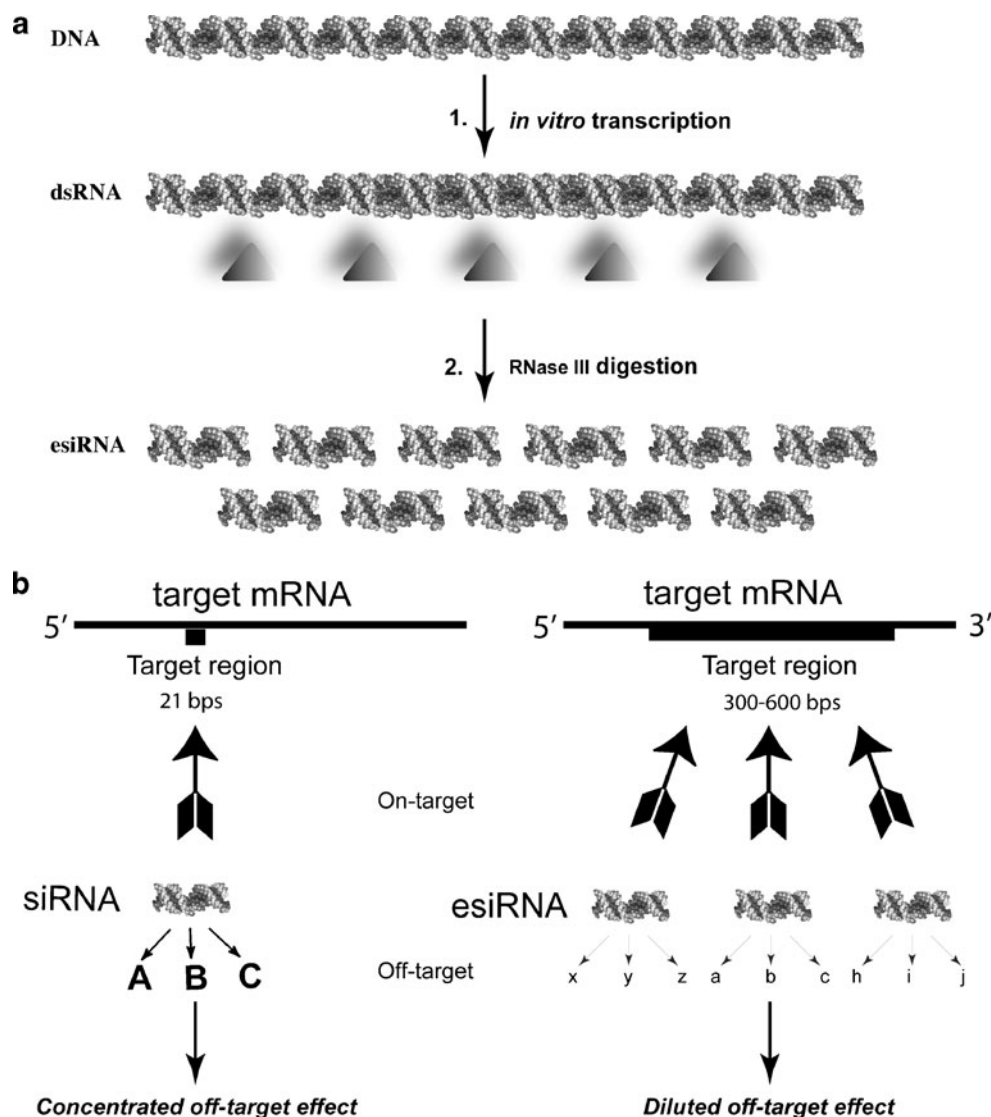
silencing triggers and each siRNA or shRNA molecule has a different ability to silence its target transcript. To ensure a high knockdown efficacy, algorithms have been developed to design potent silencing triggers that are based on the dynamic and sequence-specific properties of the silencing molecules [20, 21]. Although the use of these algorithms have improved the overall silencing efficacy, typically different silencing triggers have to be experimentally tested to obtain optimal knock-downs on the target mRNA.

An evenly important desirable property of an RNAi reagent is high target specificity. It is now well established that siRNAs and shRNAs are silencing other transcripts beside their intended targets. These so-called off-target effects arise mostly from partial homologies of the employed silencing trigger to other expressed transcripts [22]. Consequently, it is possible that an observed phenotype is not due to the knock-down of the intended mRNA, but to the down-regulation of an off-targeted transcript. Hence, off-target effects are a central challenge in the RNAi field, especially for large-scale screening experiments.

Some improvements to increase the specificity of siRNAs and shRNAs have recently been made, for instance by bioinformatically scanning siRNA sequences for their potential to hybridize to 3'-UTR sequences [23]. However, an algorithm that dependably predicts off-target signatures of silencing triggers is presently not available. A simple approach to reduce off-target effects is pooling of diverse siRNAs that target different regions of the same transcript [24, 25]. The rationale for the increased specificity via pooling is that the silencing capacities of individual siRNAs for the intended mRNA are cumulated, whereas off-target effects are diluted out. However, many different siRNAs have to be pooled to obtain the most specific results making pooling with chemically synthesized siRNAs a costly effort.

Different groups have developed an alternative approach to generate complex pools of silencing triggers for RNAi in mammalian cells [18, 26]. This technology is based on the enzymatic digestion of long dsRNA by an RNase III enzyme, generating heterogeneous pools of siRNA duplexes covering a broad region (normally 300–600 bps) of the target mRNA (Fig. 1a). The resulting endoribonuclease-prepared siRNAs (esiRNAs) have proven to knock down transcripts efficiently in mammalian tissue culture cells [27], developing embryos [28] and adult mice [29]. Their pool diversity is a distinct advantage of esiRNAs over siRNAs or shRNAs, and explains their high degree of target specificity (Fig. 1b), which ultimately translates into a lower false positive rate [25, 30]. These distinct features make esiRNA an attractive source for RNAi experiments, especially for medium- to large-scale screens.

Fig. 1 a Scheme for the production of endoribonuclease prepared siRNA (esiRNA). Important synthesis steps are indicated. The RNase III to cleave the dsRNA into esiRNA is shown as *triangles* **b** Comparison of mRNA targeting with esiRNA and siRNA. The regions that are targeted with the two silencing triggers are shown as a *thick line* beneath the mRNAs. The on-target effects are illustrated by *bold arrows*. Note the additive on-target effects of the complex pool of esiRNA. Off-target effects are symbolized by *letters* and the size of the font represents the prominence of the off-target effect



Given its versatility and ease of use, RNAi has become the method of choice to probe gene function in mammalian cells in general [19] and in ES cells in particular [31]. Importantly, the completion of sequenced genomes for a variety of different organisms, including human and mouse, has made it possible to construct genome-scale libraries of RNAi reagents, which can be used to perform high-throughput RNAi screens. Consequently, genome-wide RNAi resources for siRNAs, esiRNAs and shRNAs have been generated and employed for functional screens in a variety of different cells [32].

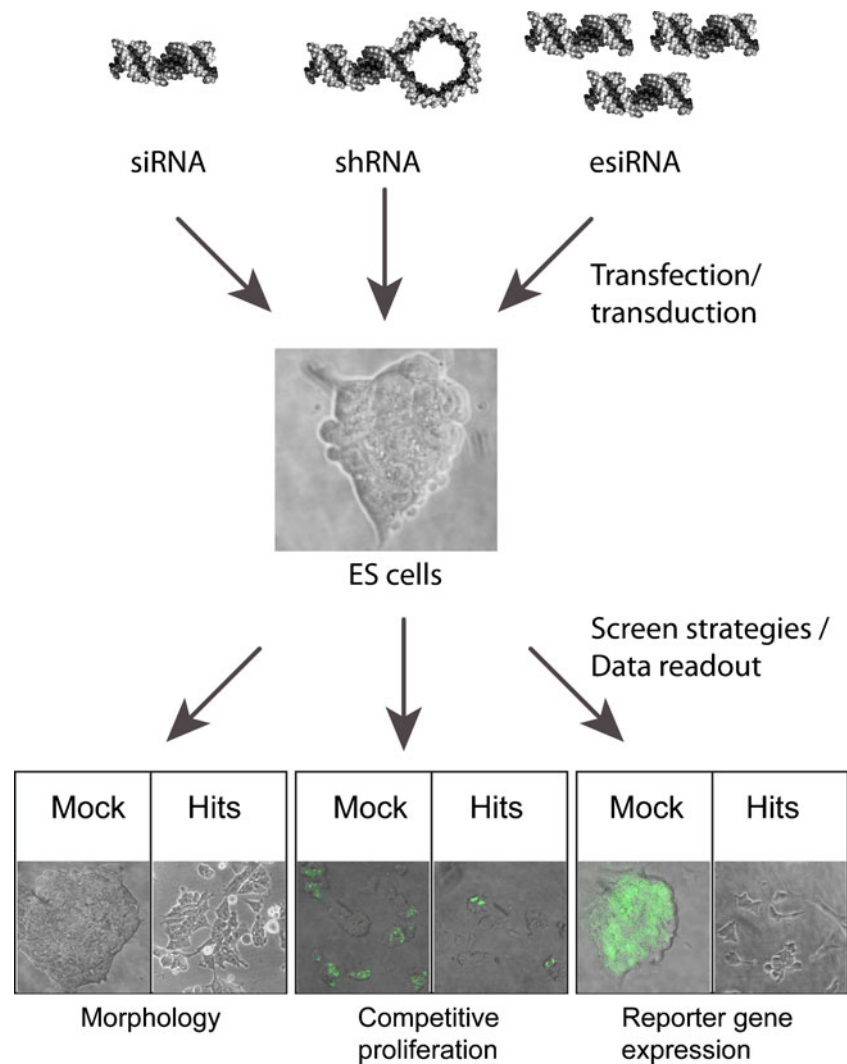
RNAi Screens in ES Cells

The potential of ES cells in research and clinical applications requires a comprehensive understanding of the pluripotency network and pathways that regulate cell fate

decisions. In the past few years, a number of RNAi screens using shRNA, siRNA or esiRNA libraries have successfully been performed in mouse and human ES cells [8, 9, 33–36]. The screens were performed employing different assays to delineate genes with roles in ES cells, including cell proliferation, cell morphology and reporter assays (Fig. 2). Results of these screens have greatly extended our knowledge on ES cell biology and have led to the identification of novel components required to maintain ES cell identity. Importantly, these screens have also provided new insights into understanding the interplay of multiple pathways of transcriptional factors, epigenetic regulators and cellular signals.

Pluripotent ES cells proliferate at high rates in tissue culture, with mouse ES cell doubling times of around 12 h [37]. Upon differentiation these rates decrease significantly and therefore the proliferation rate can be used to distinguish pluripotent ES cells from cells that have exited

Fig. 2 Summary of RNAi screens in ES cells. Silencing trigger delivery methods as well as screen strategies and assay read-outs are illustrated



the pluripotent state. Based on this fact, Ivanova et al. developed a fluorescence-based competition assay of cell proliferation by mixing a fixed ratio of shRNA transduced ES cells (GFP positive) with wild type ES cells (GFP negative; [34]). The ratio of GFP positive versus GFP negative cells was then monitored over time and knockdowns that led to a decrease of GFP positive cells were studied in detail. Using this approach ten genes were identified that are required to maintain ES cell identity, including known pluripotency genes (Nanog, Oct4, Sox2), and novel factors (Tbx3, Esrrb, Tc11, and Dppa4). Importantly, this study established a second, distinct pathway beside the already known Oct4, Nanog and Sox2 pathway that is necessary for ES cell self-renewal [34].

ES cells have a distinct morphology that distinguishes them from other cell types. They are small with little cytoplasm and form tightly packed colonies with round or polygonal borders. Upon differentiation, the cells typically

get bigger and flatten out. Furthermore, they often lose the tightly packed appearance, which leads to an extension of the colony. These features were used by Fazio et al. to screen a panel of about 1000 esiRNAs targeting structural and regulatory chromatin proteins [33]. Knockdown-induced differentiation was then scored for defects in viability and alterations in cell or colony morphology. 68 genes were identified that resulted in altered viability or morphology and the Tip60-p400 complex was selected for in-depth characterization because several components of this complex scored in the screen and exhibited similar differentiation phenotypes. Upon depletion of Tip60-p400, ES cells exhibited defects in three distinguishing features: Alkaline Phosphatase (AP) activity, EB formation and teratoma formation. Expression profiling revealed a significant overlap between the sets of genes misregulated in Tip60-p400 knockdown and Nanog knockdown suggesting that Tip60-p400 functions in the same pathway as Nanog to repress developmental genes.

Several genes are predominantly expressed in ES cells and are turned off upon differentiation of the cells. These genes include, for instance, the pluripotency transcription factors Oct4/Pou5f1 and Nanog. Expression of these genes can therefore be used as a reporter for pluripotency [38]. Based on this fact, GFP reporter cell lines have been established, where GFP is driven by promoters of pluripotency genes such as Oct4/Pou5f1 [39] and Nanog [40]. Consequently, these cells can be used to monitor the pluripotency status simply by surveying GFP expression levels [41]. Three different laboratories used this assay and employed Oct4/Pou5f1-GFP cell lines in combination with siRNA [8, 11] or esiRNA [9] libraries. The usage of reporter cell lines allowed for a rapid, automated readout and therefore all three screens were carried out as genome-wide RNAi screens. Two of the screens were performed in mouse ES cells [8, 9]. Although using different silencing triggers, both screens used the same reporter cell line and a similar strategy to monitor Oct4 expression level by FACS. The screen by Ding et al. nominated 296 genes as primary hits (Z -score >2), and validated 16 genes out of the top 30 primary hits (Z -score >4) using independent esiRNAs [9]. The screen by Hu et al. identified 148 genes that scored significantly (Z score >2) as primary hits [8]. They performed data filtering, and nominated 104 genes as high-confidence candidates, including genes that either scored by more than one siRNA from the primary screen, or genes that are highly expressed in ES cells or embryonic tissues based on published expression profiles. Although at first glance, the overlap of the hit lists from these two screens appears small, a second look reveals interesting correlations between the two data sets (Fig. 3). Out of the 16 validated knockdowns from the esiRNA screen [9], seven (44%) were also present in the 104 genes nominated by Hu et al. Moreover, different components of an additional four (25%) well-known protein complexes,

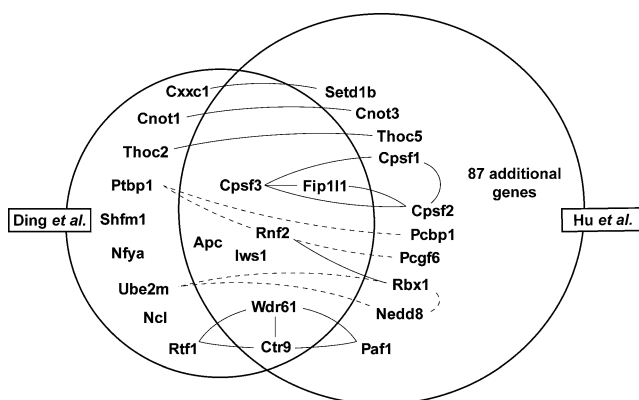


Fig. 3 Comparative analysis of validated genes required for maintenance of ES cell identity from Ding et al. [9] and Hu et al. [8]. A Venn diagram depicting the overlap, proteins that form complexes (*solid lines*) and proteins that interact (*dotted lines*; String 8.3) is shown

namely the Ccr4-Not-, Thoc-, Paf1- and Set1-complexes, were present among the 16 validated hits, indicating that these complexes play a role in maintaining pluripotency. Furthermore, String analyses [42] identified interaction partners of Ptbp1 and Ube2m to be required for ESC identity in the list of Hu et al., suggesting potential links of these proteins to maintain pluripotency. Hence, the overlap of genes identified in both screens is high when stringent validation criteria are taken into consideration. Interestingly, in both screens Oct4 and Sox2 were detected, whereas Nanog was not identified, indicating that transient knockdown of Nanog did not reduce Oct4 levels significantly during the assay period.

Hu et al. selected the genes Cnot3 and Trim28 for detailed investigations. They found that Cnot3 and Trim28 are highly expressed in ES cells and embryonic tissues and downregulated during ES cell differentiation. Chromatin IP experiments revealed that Cnot3 and Trim28 co-occupy many putative gene promoters with c-Myc and Zfx, but not other pluripotency-associated transcription factors. The transcriptional targets of this module are enriched for genes involved in cell cycle, cell death, and cancer. Importantly, in their analysis only Cnot3 but not other components of the Ccr4-Not complex were found to be required to maintain ES cell identity, leading to the proposal that Cnot3 action is independent of the other Cnot complex components to maintain pluripotency. However, Ding et al. identified and validated Cnot1 as a gene required to maintain ES cell identity challenging the view that Cnot3 is the only Cnot complex component required to maintain pluripotency.

Ding et al. selected the Paf1 complex (Paf1C) for a detailed analysis. All Paf1C components were required to maintain ES cell identity and rapid differentiation upon Paf1C depletion was observed, activating ectodermal, mesodermal and trophodermal markers. Interestingly, knockdown of Paf1C components did not result in activation of endodermal markers, indicating lineage restricted differentiation upon Paf1C depletion. The mechanism of this restricted differentiation remains to be investigated. Furthermore, the study showed that the Paf1C is enriched at promoters of pluripotency genes and that depletion of Paf1C results in a rapid loss of H3K4me3 at these promoters, suggesting a link between Paf1C and the H3K4me3 methyltransferase Set1C/COMPASS. This link was indeed substantiated subsequently by the finding that Paf1C interacts directly with Bre1 in H2B ubiquitylation, which stimulates Set1C dependent H3K4 di- and trimethylation [43].

Lastly, the first genome-wide RNAi screen in human ES cells (hESCs) was recently published [11]. Reminiscent of two of the mouse ES cell screens, this screen was also performed in Oct4/Pou5f1-GFP hESCs, which express GFP under the control of Oct4/Pou5f1 regulatory elements.

Based on the quantification of GFP fluorescence, 566 primary hits were nominated in this study. Gene ontology (GO) analysis of the primary hits showed an enrichment for transcription factors and translation factors, including the INO80 chromatin remodeling complex, the mediator complex, the TAF complex, the eukaryotic initiation factor complex, and genes that are implicated in splicing. Substantial differences were detected when comparing the hit list of the human ES cell screen with the four mouse RNAi screens, suggesting that hESCs may possess divergent molecular mechanisms to mouse ES cells to maintain pluripotency. Whether this is indeed the case, or whether the differences rather reflect different culture conditions will have to be investigated in the future. Differences between hES cells and mES cells have been known ever since hES cells were first established [44]. For examples, mES cells grow in three-dimensional, tightly packed colonies with a population doubling time of approximately 12 h and require LIF and BMP4 growth factor signaling for their continued self-renewal. In contrast, hES cells form flattened two-dimensional colonies and are maintained in bFGF and ActivinA/TGF- β to prevent differentiation. Human ES cells proliferate slowly, with a population doubling time averaging 36 h. Epigenetically, human and mouse ESCs display a different pattern of X chromosome inactivation and promoter occupancy by pluripotency transcription factors [45].

Reports on the derivation of murine epiblast stem cells (EpiSCs) recently provided a new perspective on the nature of human ES cells [46, 47]. EpiSCs are derived from postimplantation murine epiblast embryos under culture conditions similar to hES cell culture conditions. EpiSCs express transcription factors known to regulate pluripotency, maintain their genomic integrity, and robustly differentiate into the major somatic cell types as well as primordial germ cells. The EpiSC lines are distinct from mouse ES cells in their epigenetic state and the signals controlling their differentiation. In contrast, EpiSCs demonstrate striking similarities to hESCs with regard to their molecular properties, growth factor requirements, colony morphology, X-inactivation status, and culture dynamics [46, 47]. These findings suggest that hES cells correspond to cells of a postimplantation/pregastrulation stage of mammalian development, whereas mES cells correspond to preimplantation stages [46]. Hence, this difference may also explain the lack of overlap of genes required to maintain ES cell identity between the mouse- and human ESC RNAi screens. It will therefore be interesting to investigate the genes required to maintain EpiSCs and to compare these genes to the list obtained in the screen with hES cells.

Independent of the reasons for the differences in the ES cell RNAi screens, the results show that the

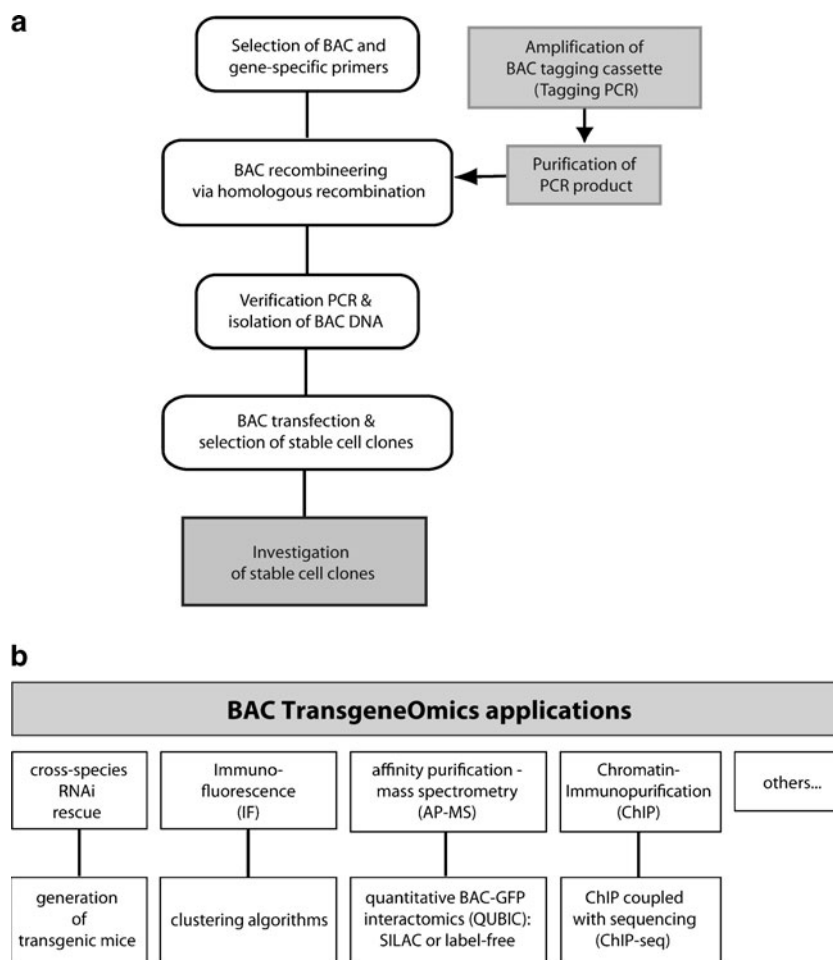
regulation of pluripotency is complex and that many factors contribute to maintain ES cell identity. The large number of hits identified in the RNAi screens poses a new challenge to the understanding of ES cell biology; the detailed dissection of the molecular mechanisms underlying the knockdown phenotypes with current techniques will take a considerable amount of time. Hence, strategies to streamline investigations of identified genes in RNAi screens are desperately needed.

TransgeneOmics

One powerful method to rapidly investigate gene function combines RNAi with bacterial artificial chromosome (BAC) recombination technology and is called BAC TransgeneOmics [48–50]. In TransgeneOmics a BAC carrying a third allele of the gene of interest is introduced and stably integrated into the genome of the host cell. To facilitate selection and further biochemical analysis, the BAC containing the gene of interest is modified via homologous recombination in *E. coli* employing bacteriophage enzymes [51]. Typically, a tag that contains a fluorescence marker (GFP) for localization studies in addition to other features (TEV, S-peptide, PreScission, selection marker) useful for downstream applications e.g. affinity purification is fused during the recombination process. Importantly, BAC recombineering is highly efficient and can be carried out in high throughput (Fig. 4a) [49].

First, a suitable BAC containing the gene of interest and suitable primers are identified using the software BACfinder (www.mitocheck.org/cgi-bin/BACfinder) [52]. The BAC is then typically ordered from a resource center such as BACPAC CHORI, California, USA (<http://bacpac.chori.org/>), which normally provide the plasmid as a bacterial stab culture. The bacteria harboring the BAC are then transduced with plasmids allowing expression of phage recombination systems, such as the λ red, or ET cloning system [53]. The advantage of using homologous recombination in *E. coli* for BAC modification is, that it alleviates the limitations of the use of restriction enzymes. Because the integration site is defined through homologous regions, which are stretches of DNA shared by the DNA molecules that recombine, a wide range of DNA modifications at any chosen position and unlimited by fragment size is easily possible [53]. Because homologous arms of only 35 to 60 nucleotides are sufficient for successful recombination [51], they can be generated through oligonucleotide synthesis and subsequently attached to the DNA fragment via polymerase chain reaction (PCR). This results in a linear PCR fragment, which

Fig. 4 a BAC transgeneOmics pipeline. Main steps are indicated as followed: First, a suitable BAC clone is selected, followed by modification using a gene-specific BAC tagging cassette that is inserted via homologous recombination (BAC recombineering). Finally, after transfection, stable cell pools are selected and investigated, **b** Overview of biochemical and molecular approaches based on BAC TransgeneOmics



exhibits gene-specific homology arms on both ends. To increase the number of positive homologous recombination events, an antibiotic-resistant gene (e.g. neomycin-kanamycin) is typically included in the PCR fragment. Therefore, bacteria carrying the inserted cassette are easily identified through drug selection.

Correct insertion of the tagging construct is verified by checking clones by a PCR reaction that amplifies the region around the insertion site. Subsequently, the purified BAC DNA is transfected into cultured mammalian cells, typically through lipofection. After selection of the marker carried on the modified BAC, cell pools stably expressing the tagged transgene are used for downstream analyses (see below). Because BACs are much larger (up to 250 kb and more) than conventional cDNA constructs and harbor the gene of interest in its genomic context, they also frequently contain the endogenous regulatory sequences that control the expression levels of the gene. This allows for expression of the tagged protein at near physiological levels, ensuring cell type-specific processing, regulation, and splicing [49].

Streamlined Molecular Characterization of RNAi Phenotypes through BAC Transgenesis

BAC technology offers a variety of different biochemical and molecular biological approaches for further analysis of candidate genes (Fig. 4b). This versatility makes the “BAC TransgeneOmics” approach an ideal tool for studying ESC differentiation involving different genes being expressed at specific time points. Additionally, recombineering techniques allow for the precise manipulation of BAC transgenes by introducing specific mutations to analyze functional consequences of mutated alleles [54]. The most straightforward TransgeneOmics applications are described below.

Cross-Species RNAi Rescue

As discussed above, off-target effects are frequently observed in RNAi experiments in mammalian cells [55]. Therefore, validation of RNAi phenotypes is essential to avoid reporting false positive results. The ultimate test of the specificity of an RNAi experiment is a rescue

experiment by expression of the target gene in a form refractory to the silencing trigger, thus reverting the RNAi-induced phenotype. This can be achieved by silent point mutations introduced into cDNA constructs encoding the target gene, such that the siRNA is no longer complementary to the modified region [56]. However, this approach is limited by the availability of a full-length cDNA and often requires time consuming cloning to obtain the desired construct. Furthermore, it is very difficult to achieve endogenous expression levels employing cDNA constructs and mis-expression of genes can cause phenotypes by themselves.

To overcome these hurdles, a rescue approach by introducing BAC encoded orthologous genes from a closely related species, e.g. by transfecting human genes into mouse cells or vice versa has been established [48]. This method achieves two things: first, the transgene is expressed at physiological levels, because the gene is surrounded by its own regulatory elements and second, the different DNA sequence of the cross species gene frequently allows to specifically silence the endogenous transcript. The cross-species mRNAs are refractory to the silencing trigger, therefore, abolishing the need for introducing point mutations into the candidate transcript. For example, the role of the Paf1C in the maintenance of mouse ES cell identity was validated with this approach [9]. For this purpose, human BACs encoding *Ctr9*, *Rtf1*, and *Leo1* were engineered by BAC TransgeneOmics [49] and stably transfected into the mouse ES cells to render these cell lines resistant to the corresponding esiRNAs, hence, manifesting an essential role of the Paf1C for ESC identity [9].

Protein Localization

Determination of protein localization within cells is a vital and frequently used method to characterize a protein of interest including assumptions on its potential role in protein pathways and networks. Antibodies are typically used to investigate the localization of proteins. However, their generation is cost-intensive and time-consuming and the optimal staining protocol for each antibody is different and has to be optimized. Furthermore, antibody staining does not allow for a spatio-temporal investigation of protein localization. The generation of BAC-tagged GFP fusion proteins offers a versatile approach that overcomes many of these limitations. First, cell lines expressing BAC-tagged GFP fusion proteins can be rapidly and cost-effectively generated. Importantly, the fusion protein is typically expressed near physiological levels, avoiding localization artifacts often observed when GFP-fusion proteins are over-expressed from cDNA constructs. Second, a single, well-defined antibody against GFP can be used employing the same staining protocol for all tagged proteins (Fig. 5).

Third, following the GFP signal in live cells can provide valuable spatio-temporal information of protein localization such as protein behavior during the cell cycle or during cell differentiation.

Proteomics

To interpret RNAi based genotype-phenotype relationships, the identification of protein interactions is often informative. Again antibodies are typically used to isolate proteins that associate to the protein of interest, which are subsequently identified via mass spectrometry [57, 58]. However, similar restrictions as for the protein localization also apply for the use of antibodies to investigate protein-protein interactions. The use of generated BAC-transgenic cell lines alleviates many of these limitations and allows for a streamlined analysis of protein-protein interactions of large numbers of proteins. For HeLa cells the power of this approach has recently been demonstrated [49, 52] and is also applicable in ES cells. For example, immunoprecipitations followed by mass spectrometry analyses of Paf1C components identified *Wdr61* as a novel component of the Paf1C in ES cells that, like other Paf1C components, is required to maintain ES cell identity [9]. Because of the scalability of this approach, it appears feasible that with TransgeneOmics a first genome-wide ES cell proteome interaction network could be built within the next few years.

Chromatin Immunoprecipitation

To determine the interaction sites for DNA binding proteins, another application, chromatin immunoprecipitation (ChIP), is highly suitable for the “BAC TransgeneOmics” approach. Again, to avoid the reliance on specific antibodies, GFP-tagged BAC transgenes can be used in ChIP employing a standardized assay, eliminating the need to establish the best protocol for a given antibody. Importantly, large-scale ChIP applications, such as ChIP-chip and ChIP-seq are also feasible with the BAC TransgeneOmics approach. Employing this technology it was demonstrated that the Paf1C component *Ctr9* is enriched at certain promoters in ES cells, particularly at promoters of pluripotency genes [9]. Scaling of this approach promises to deliver high quality data of genome binding sites for many DNA binding proteins, which should help to decipher chromosome organization and transcription factor networks in ES cells.

Perspective

Through the development of RNAi technologies in ES cells we have witnessed a dramatic extension of our knowledge about these cells. Different assays to identify genes with roles in ES cell biology have been employed in a variety of

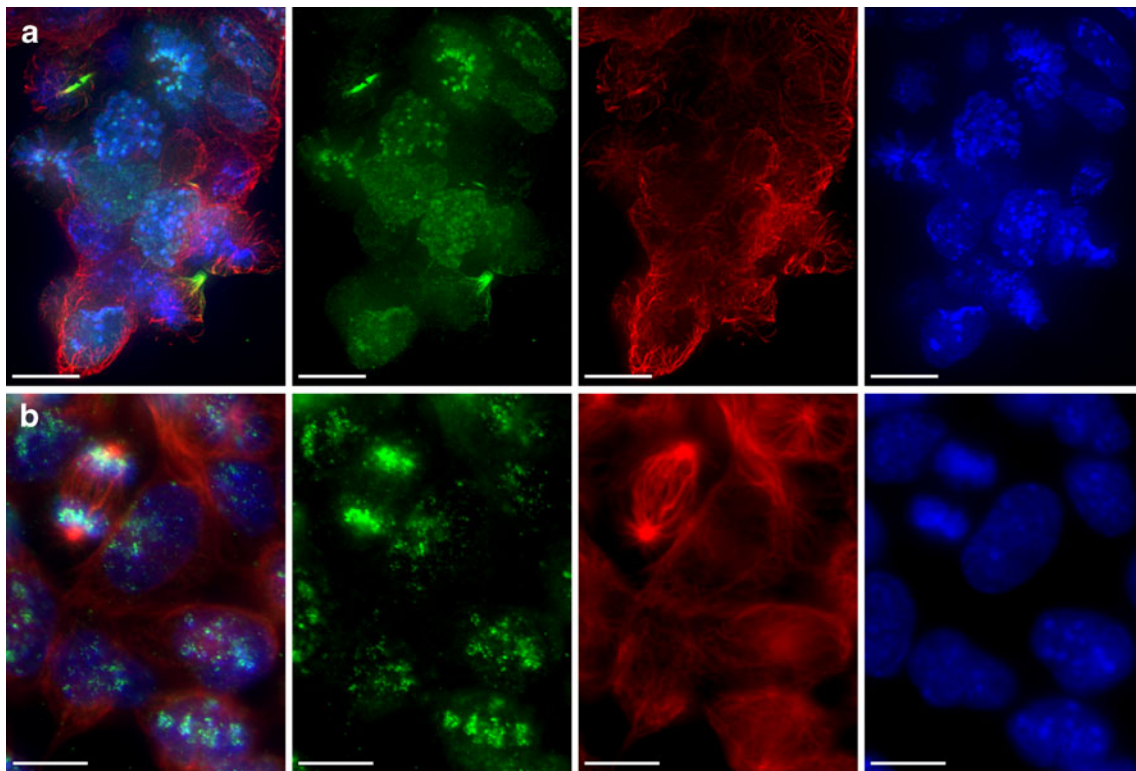


Fig. 5 Immunostaining of mouse ES cells stably expressing GFP-tagged transgenes. **a** AURKB, **b** Ki67. *Green*: GFP-tagged transgene; *Red*: alpha-tubulin; *Blue*: DAPI (scale bar: 10 μ m)

screens and we now have first genome-wide data sets of genes required to maintain ES cell identity. These experiments have also allowed a first glimpse of the complexity to maintain a balance between pluripotency and the capacity for rapid differentiation in these cells. However, these results have also shown us that we are only beginning to understand ES cell biology at a systems level and that many more assays are required to obtain a comprehensive list of the genes that play a role to maintaining ES cell identity. A focus to identify novel factors has so far relied on Oct4/Pou5f1 as a reporter for pluripotency. It will be interesting to employ other pluripotency reporter cell lines (e.g. Nanog, Rex1 etc.) in equivalent assays to search for additional factors that have an impact on ES cells. In a slight modification, reporter cell lines for lineage markers should also prove valuable to search for genes that are implicated in the differentiation into certain lineages. In this scenario, undifferentiated ES cells would not be fluorescent and the number of differentiated cells for a particular lineage (e.g. Cdx1-GFP for trophoderm, or brachyury-GFP for mesoderm etc.) could be scored after each knockdown by counting the percent of GFP positive cells. This study would not only provide novel insight into ES cell biology but could also improve protocols to differentiate ES cells selectively into defined lineages.

The identification of genes implicated in ES cell identity is a prerequisite to understand ES cells at a systems level. However, the mere identification and phenotypic description upon depletion will not be sufficient to understand the function of each gene mechanistically. The TransgeneOmics approach is suitable to obtain a deeper mechanistic insight into the molecular reasons underlying the depletion phenotypes. It is scalable and possible to perform in high-throughput, promising to rapidly deliver the necessary information to build meaningful mathematical models of pluripotency. Hence, the generation of a repository of BAC-tagged ES cells should be a useful resource for the scientific community. Importantly, this resource could be the starting point to perform the described experiments not only in ES cells but also in any other cell type of the body, because animals could be generated from these cells [49], promising to further accelerate gene function studies in the future.

Acknowledgments We thank Nicolas Berger and Anja Nitzsche for providing images of BAC transgenic ES cells and other members of the Buchholz laboratory for discussions. This work was supported by the BMBF grants GO-Bio (0315105) and NGFN-plus DiGTOP (01GS1102), the DFG grant SPP1356 (BU1400/3-1) and the EU FP7 grant SyBoSS (HEALTH-F4-2010-242129).

Conflicts of interest The authors declare no potential conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Evans, M. J., & Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature*, *292*(5819), 154–156.
- Chambers, I., Colby, D., Robertson, M., et al. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, *113*(5), 643–655.
- Keller, G. (2005). Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes & Development*, *19*(10), 1129–1155.
- Takahashi, K., Tanabe, K., Ohnuki, M., et al. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, *131*(5), 861–872.
- Okita, K., Ichisaka, T., & Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. *Nature*, *448*(7151), 313–317.
- Yoshida, Y., & Yamanaka, S. (2010). Recent stem cell advances: induced pluripotent stem cells for disease modeling and stem cell-based regeneration. *Circulation*, *122*(1), 80–87.
- Niwa, H. (2007). How is pluripotency determined and maintained? *Development*, *134*(4), 635–646.
- Hu, G., Kim, J., Xu, Q., Leng, Y., Orkin, S. H., & Elledge, S. J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. *Genes & Development*, *23*(7), 837–848.
- Ding, L., Paszkowski-Rogacz, M., Nitzsche, A., 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*(5), 403–415.
- Abujarour, R., Efe, J., & Ding, S. (2010). Genome-wide gain-of-function screen identifies novel regulators of pluripotency. *Stem Cells*, *28*(9), 1487–1497.
- Chia, N. Y., Chan, Y. S., Feng, B., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature*, *468*(7321), 316–320.
- Pritsker, M., Ford, N. R., Jenq, H. T., & Lemischka, I. R. (2006). Genomewide gain-of-function genetic screen identifies functionally active genes in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(18), 6946–6951.
- te Riele, H., Maandag, E. R., & Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(11), 5128–5132.
- Tong, C., Li, P., Wu, N. L., Yan, Y., & Ying, Q. L. (2010). Production of p53 gene knockout rats by homologous recombination in embryonic stem cells. *Nature*, *467*(7312), 211–213.
- Amano, H., Itakura, K., Maruyama, M., Ichisaka, T., Nakagawa, M., & Yamanaka, S. (2006). Identification and targeted disruption of the mouse gene encoding ESG1 (PH34/ECAT2/DPPA5). *BMC Developmental Biology*, *6*, 11.
- Lubitz, S., Glaser, S., Schaft, J., Stewart, A. F., & Anastassiadis, K. (2007). Increased apoptosis and skewed differentiation in mouse embryonic stem cells lacking the histone methyltransferase Mll2. *Molecular Biology of the Cell*, *18*(6), 2356–2366.
- Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., & Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature*, *411*(6836), 494–498.
- Yang, D., Buchholz, F., Huang, Z., et al. (2002). Short RNA duplexes produced by hydrolysis with Escherichia coli RNase III mediate effective RNA interference in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(15), 9942–9947.
- Hannon, G. J. (2002). RNA interference. *Nature*, *418*(6894), 244–251.
- Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W. S., & Khvorov, A. (2004). Rational siRNA design for RNA interference. *Nature Biotechnology*, *22*(3), 326–330.
- Henschel, A., Buchholz, F., & Habermann, B. (2004). DEQOR: a web-based tool for the design and quality control of siRNAs. *Nucleic Acids Research*, *32*(Web Server issue), W113–W120.
- Jackson, A. L., Burchard, J., Schelter, J., et al. (2006). Widespread siRNA “off-target” transcript silencing mediated by seed region sequence complementarity. *RNA*, *12*(7), 1179–1187.
- Jackson, A. L., & Linsley, P. S. (2010). Recognizing and avoiding siRNA off-target effects for target identification and therapeutic application. *Nature Reviews. Drug Discovery*, *9*(1), 57–67.
- Parsons, B. D., Schindler, A., Evans, D. H., & Foley, E. (2009). A direct phenotypic comparison of siRNA pools and multiple individual duplexes in a functional assay. *PLoS ONE*, *4*(12), e8471.
- Kittler, R., Surendranath, V., Heninger, A. K., et al. (2007). Genome-wide resources of endoribonuclease-prepared short interfering RNAs for specific loss-of-function studies. *Nature Methods*, *4*(4), 337–344.
- Myers, J. W., Jones, J. T., Meyer, T., & Ferrell, J. E., Jr. (2003). Recombinant Dicer efficiently converts large dsRNAs into siRNAs suitable for gene silencing. *Nature Biotechnology*, *21*(3), 324–328.
- Kittler, R., Putz, G., Pelletier, L., et al. (2004). An endoribonuclease-prepared siRNA screen in human cells identifies genes essential for cell division. *Nature*, *432*(7020), 1036–1040.
- Calegari, F., Marzesco, A. M., Kittler, R., Buchholz, F., & Huttner, W. B. (2004). Tissue-specific RNA interference in post-implantation mouse embryos using directional electroporation and whole embryo culture. *Differentiation*, *72*(2–3), 92–102.
- Xuan, B., Qian, Z., Hong, J., & Huang, W. (2006). EsiRNAs inhibit Hepatitis B virus replication in mice model more efficiently than synthesized siRNAs. *Virus Research*, *118*(1–2), 150–155.
- Myers, J. W., Chi, J. T., Gong, D., Schaner, M. E., Brown, P. O., & Ferrell, J. E. (2006). Minimizing off-target effects by using diced siRNAs for RNA interference. *Journal of RNAi and Gene Silencing*, *2*(2), 181–194.
- Ding, L., & Buchholz, F. (2006). RNAi in embryonic stem cells. *Stem Cell Reviews*, *2*(1), 11–18.
- Mohr, S., Bakal, C., & Perrimon, N. (2010). Genomic screening with RNAi: results and challenges. *Annual Review of Biochemistry*, *79*, 37–64.
- Fazio, T. G., Huff, J. T., & Panning, B. (2008). An RNAi screen of chromatin proteins identifies Tip60-p400 as a regulator of embryonic stem cell identity. *Cell*, *134*(1), 162–174.
- Ivanova, N., Dobrin, R., Lu, R., et al. (2006). Dissecting self-renewal in stem cells with RNA interference. *Nature*, *442*(7102), 533–538.
- Jian, R., Cheng, X., Jiang, J., Deng, S., Hu, F., & Zhang, J. (2007). A cDNA-based random RNA interference library for

- functional genetic screens in embryonic stem cells. *Stem Cells*, 25(8), 1904–1912.
36. Zhang, J. Z., Gao, W., Yang, H. B., Zhang, B., Zhu, Z. Y., & Xue, Y. F. (2006). Screening for genes essential for mouse embryonic stem cell self-renewal using a subtractive RNA interference library. *Stem Cells*, 24(12), 2661–2668.
 37. Udy, G. B., Parkes, B. D., & Wells, D. N. (1997). ES cell cycle rates affect gene targeting frequencies. *Experimental Cell Research*, 231(2), 296–301.
 38. Habermann, F. A., Wuensch, A., Sinowatz, F., & Wolf, E. (2007). Reporter genes for embryogenesis research in livestock species. *Theriogenology*, 68(Suppl 1), S116–S124.
 39. Ying, Q. L., Nichols, J., Evans, E. P., & Smith, A. G. (2002). Changing potency by spontaneous fusion. *Nature*, 416(6880), 545–548.
 40. Silva, J., Nichols, J., Theunissen, T. W., et al. (2009). Nanog is the gateway to the pluripotent ground state. *Cell*, 138(4), 722–737.
 41. Ying, Q. L., Nichols, J., Chambers, I., & Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell*, 115(3), 281–292.
 42. Szklarczyk, D., Franceschini, A., Kuhn, M., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Research*, 39(Database issue), D561–D568.
 43. Kim, J., Guermah, M., McGinty, R. K., et al. (2009). RAD6-Mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. *Cell*, 137(3), 459–471.
 44. Thomson, J. A., Itskovitz-Eldor, J., Shapiro, S. S., et al. (1998). Embryonic stem cell lines derived from human blastocysts. *Science*, 282(5391), 1145–1147.
 45. Buecker, C., & Geijsen, N. (2010). Different flavors of pluripotency, molecular mechanisms, and practical implications. *Cell Stem Cell*, 7(5), 559–564.
 46. Tesar, P. J., Chenoweth, J. G., Brook, F. A., et al. (2007). New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature*, 448(7150), 196–199.
 47. Brons, I. G., Smithers, L. E., Trotter, M. W., et al. (2007). Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature*, 448(7150), 191–195.
 48. Kittler, R., Pelletier, L., Ma, C., et al. (2005). RNA interference rescue by bacterial artificial chromosome transgenesis in mammalian tissue culture cells. *Proceedings of the National Academy of Sciences of the United States of America*, 102(7), 2396–2401.
 49. Poser, I., Sarov, M., Hutchins, J. R., et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature Methods*, 5(5), 409–415.
 50. Sarov, M., Schneider, S., Pozniakovski, A., et al. (2006). A recombineering pipeline for functional genomics applied to *Caenorhabditis elegans*. *Nature Methods*, 3(10), 839–844.
 51. Zhang, Y., Buchholz, F., Muylers, J. P., & Stewart, A. F. (1998). A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genetics*, 20(2), 123–128.
 52. Hutchins, J. R., Toyoda, Y., Hegemann, B., et al. (2010). Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science*, 328(5978), 593–599.
 53. Muylers, J. P., Zhang, Y., & Stewart, A. F. (2001). Techniques: recombinogenic engineering—new options for cloning and manipulating DNA. *Trends in Biochemical Sciences*, 26(5), 325–331.
 54. Bird, A. W., & Hyman, A. A. (2008). Building a spindle of the correct length in human cells requires the interaction between TPX2 and Aurora A. *The Journal of Cell Biology*, 182(2), 289–300.
 55. Jackson, A. L., & Linsley, P. S. (2004). Noise amidst the silence: off-target effects of siRNAs? *Trends in Genetics*, 20(11), 521–524.
 56. Lassus, P., Rodriguez, J., & Lazebnik, Y. (2002). Confirming specificity of RNAi in mammalian cells. *Science Signaling*, 2002(147), p113.
 57. Gingras, A. C., Gstaiger, M., Raught, B., & Aebersold, R. (2007). Analysis of protein complexes using mass spectrometry. *Nature Reviews. Molecular Cell Biology*, 8(8), 645–654.
 58. Kocher, T., & Superti-Furga, G. (2007). Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nature Methods*, 4(10), 807–815.