

Development 140, 4110-4115 (2013) doi:10.1242/dev.094615
© 2013. Published by The Company of Biologists Ltd

Genome-wide shRNA screening to identify factors mediating Gata6 repression in mouse embryonic stem cells

Sarah Cooper* and Neil Brockdorff*

SUMMARY

The use of whole-genome pooled shRNA libraries in loss-of-function screening in tissue culture models provides an effective means to identify novel factors acting in pathways of interest. Embryonic stem cells (ESCs) offer a unique opportunity to study processes involved in stem cell pluripotency and differentiation. Here, we report a genome-wide shRNA screen in ESCs to identify novel components involved in repression of the *Gata6* locus, using a cell viability-based screen, which offers the benefits of stable shRNA integration and a robust and simple protocol for hit identification. Candidate factors identified were enriched for transcription factors and included known Polycomb proteins and other chromatin-modifying factors. We identified the protein Bcor, which is known to associate in complexes with the Polycomb protein Ring1B, and verified its importance in *Gata6* repression in ESCs. Potential further applications of such a screening strategy could allow the identification of factors important for regulation of gene expression and pluripotency.

KEY WORDS: Embryonic stem cells, Polycomb, shRNA screen, Mouse

INTRODUCTION

Embryonic stem cells (ESCs) are pluripotent, retaining the capacity to differentiate into different cell types representing the three germ layers of the developing embryo, and can be maintained indefinitely by culturing in the presence of defined factors (Blair et al., 2011). As such, ESCs provide an important *in vitro* model for understanding processes in early development. Studies on pluripotency are also important in terms of the potential medical or therapeutic applications of ESCs, for example through reprogramming of somatic cells (Hussein and Nagy, 2012).

The transcriptional circuitry of ESCs has been extensively studied through the application of high-throughput genomic methods (Boyer et al., 2005), and also using small interfering RNA (siRNA) library-based screening approaches (Ding et al., 2009; Hu et al., 2009). Among the factors implicated in the maintenance of pluripotency are the Polycomb group (PcG) proteins that repress expression of several thousand key developmental genes in ESCs (Ku et al., 2008). The two major polycomb complexes, PRC1 and PRC2, enzymatically modify chromatin, catalysing mono-ubiquitylation of histone H2A (H2AK119u1) and methylation of histone H3 (H3K27me3), respectively (Simon and Kingston, 2009). Interestingly, in ESCs, PcG target loci exhibit a unique bivalent chromatin state in which H3 acetylation and H3K4 methylation, modifications normally associated with gene activity, co-exist with PcG-mediated repressive modifications. It is thought that this represents poising of loci, which encode factors important for specifying differentiated cell lineages (Azuara et al., 2006; Bernstein et al., 2006). It remains unclear which factors recruit PcG complexes

to target loci and what is the basis for establishment of a bivalent chromatin configuration.

Genome-wide loss-of-function RNAi screens provide a powerful means to analyse functionally the genome in an unbiased manner. Such screens have proved highly successful in *Caenorhabditis elegans*, *Drosophila melanogaster* and mammalian tissue culture models. Mammalian screens have largely focused on the field of cancer biology and have been used to identify genes involved in cell proliferation, survival, invasion and migration, and responses to drugs (reviewed by Simpson et al., 2012).

In mammalian cells, RNAi-based screening can be achieved by direct, transient introduction of siRNAs or through expression of short hairpin RNAs (shRNAs) in the cell, which are subsequently processed by the RNAi machinery. A significant drawback in the former case is the cost and infrastructure requirements for robotic screening of cells in a plate-based format. In addition, transfection efficiencies of certain cell types can be very low, and siRNAs only provide a transient loss of function that might not be sufficient for certain screening requirements. These shortcomings are overcome, at least in part, by using shRNA libraries. Whole-genome shRNA libraries can be screened in pool format, significantly reducing costs and simplifying infrastructure requirements (Hu and Luo, 2012), and the use of lentiviral vectors allows screening in most cell types, including primary cells (Root et al., 2006). Moreover, stable integration of shRNA constructs into the genome allows long-term knockdown.

Our aim was to identify novel factors involved in repression at a bivalent gene promoter in ESCs. We therefore generated a reporter cell line in which a neomycin-resistance marker is integrated into an endogenous PcG-repressed target locus, the *Gata6* gene. We screened using a whole-genome lentiviral shRNA library to identify factors involved in *Gata6* repression. Validations carried out on selected candidates defined the protein Bcor as being important for PcG-mediated repression at the *Gata6* locus. We discuss potential applications of this screening strategy, both in the identification of transcriptional regulatory factors and more widely in defining molecular mechanisms in differentiation and development.

Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK.

* Author for correspondence (sarah.cooper@bioch.ox.ac.uk; neil.brockdorff@bioch.ox.ac.uk)

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

Accepted 22 July 2013

MATERIALS AND METHODS

ESC culture

ESC lines E14TG2A, Eed4 conditional knockout (cKO) (Ura et al., 2008) and ES-ERT2 *Ring1A*^{-/-} (Endoh et al., 2008) were cultured as described previously (Nesterova et al., 2008). To induce conditional knockout of Eed, Eed4 cKO were treated for 8 days with 1 µg/ml doxycycline. Conditional deletion of *Ring1B* (*Rnf2* – Mouse Genome Informatics) from ES-ERT2 *Ring1A*^{-/-} cells was carried out by the addition of 800 nM 4-hydroxytamoxifen for 8 days. ESCs were differentiated using all-trans retinoic acid (Sigma) as described (Rougeulle et al., 2004).

Reporter construction

The Gata6 reporter line was created by homologous recombination in E14TG2A cells as described in supplementary material Fig. S1 using standard techniques (Caparros et al., 2002).

Stable shRNA knockdown in ESCs

Production of shRNA plasmids and lentivirus packaging were performed as previously described (Tavares et al., 2012). The sequences of the complementary shRNA oligonucleotides are shown in supplementary material Table S1. ESCs (7.5×10^5 per well of a 6-well dish) were plated with 450 µl of lentivirus and polybrene (final concentration 8 µg/ml). Selection (1.75 µg/ml puromycin) was applied 48 hours after transduction. After 7 days, colonies were harvested for analysis or individually picked and expanded for analysis. In the colony-counting experiment, G418 selection was applied 72 hours after transduction at 0, 100, 250 or 400 µg/ml concentration. After 7 days the number of surviving colonies were counted.

Expression analysis

RNA was extracted from cell pellets, then DNase treated and cDNA was synthesised using standard methods. Expression of genes (see supplementary material Table S1 for primer sequences) was analysed by quantitative PCR. Relative gene expression analysis is normalised to the expression of the housekeeping gene *Idh2* (known to remain unchanged upon differentiation).

Genome-wide shRNA screen

Sigma MISSION pooled mouse Lentiplex shRNA library (SHPM01) comes pre-packaged in lentivirus and divided into ten pools each containing 25 µl of packaged library. For each pool, we used 2.5 µl of virus which gave approximately ten times coverage, and <0.1 multiplicity of infection (MOI).

We trypsinised 3.5×10^7 cells and plated 2.5×10^6 per 90 mm dish (12–20 dishes) in 5 ml of ES media (Nesterova et al., 2008). Polybrene was added in 1 ml media to give a final concentration of 8 µg/ml. Lentivirus packaged library (2.5 µl) was diluted in media and distributed between the dishes to give a total volume of 7 ml per dish. Additionally 2.5×10^6 cells in a 90 mm dish were also infected with a homemade scrambled virus of an identical titre. Cells were incubated for 24 hours, and then each 90 mm dish was split into three 145 mm dishes (1:9 split), allowing enough area for single colonies to form. Controls for transduction efficiency are necessary to determine the coverage of the screen, so one-hundredth of the cells from a single 90 mm dish were grown under puromycin selection on a 90 mm dish. This was also carried out for the scrambled control.

Twenty-four hours after transferring the cells to 145 mm dishes, puromycin (1.75 µg/ml) and G418 (400 µg/ml) were added to both the library plates and the scrambled control. Surviving colonies were picked 7–12 days later from the library plates into 96-well plates. The day when colonies were picked was determined by when the colonies on the scrambled control plates were all dead.

DNA was extracted from the 96-well plate. Nested PCR was carried out using standard 25 µl volume Taq PCR, annealing temperature 58°C and extension time 1 minute. The product of PCR1 (1 µl) was used as the template for PCR2. The final 313-bp product (5 µl) (including the shRNA hairpin) was treated with Exonuclease 1 (1.2 U) and shrimp alkaline phosphatase (0.66 U) for 30 minutes at 37°C and inactivated at 80°C for 10 minutes and sent for sequencing, with nested PCR2 forward primer.

Gene functional analysis

Gene functional analysis was carried out using DAVID gene functional classification tool and DAVID functional annotation tool (Huang et al., 2007; Huang et al., 2009).

Chromatin immunoprecipitation (ChIP)

Cells stably expressing either a scrambled control or Bcor sh1 hairpin were expanded and harvested for ChIP as previously described (Blackledge et al., 2010), using the following commercial antibodies (5 µg/IP): anti-histone H3K27me3 (Diagenode CS-069-100) and anti-histone H3 (Abcam ab1791).

RESULTS AND DISCUSSION

Principle of the screen

We devised the screening strategy shown in Fig. 1A. To establish a reporter cell line, we integrated a neomycin reporter gene, conferring resistance to the drug G418, into an endogenous bivalent locus by homologous recombination, maintaining genomic context and chromatin states. Expression analysis of selected bivalent loci in PRC2 (*Eed*^{-/-}) and PRC1 (*Ring1B*^{-/-}, *Ring1A*^{-/-}) null ESCs

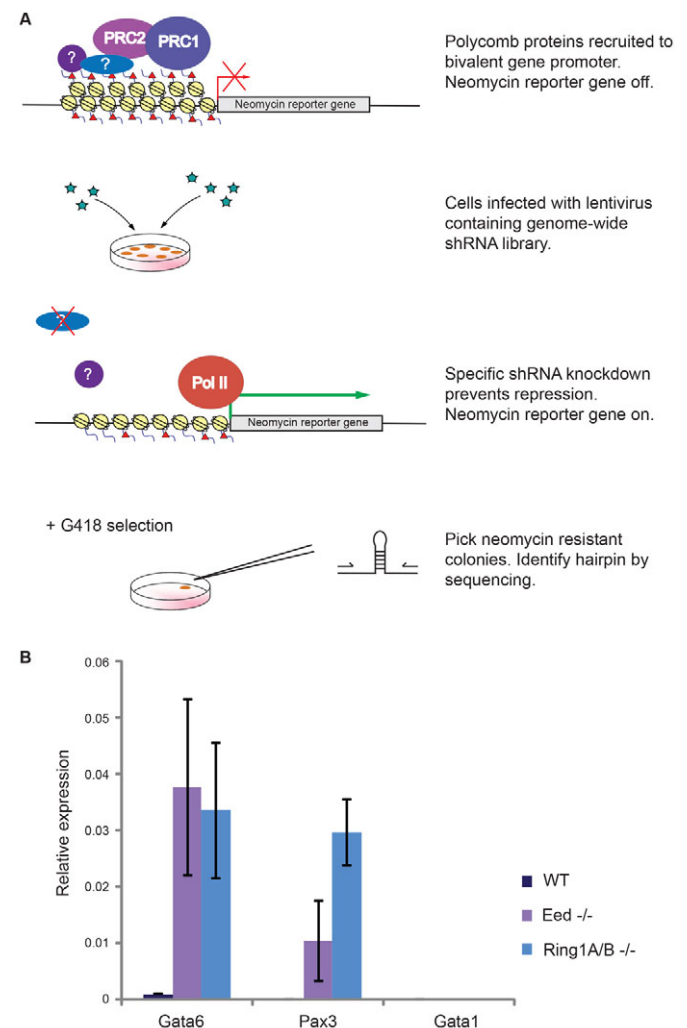


Fig. 1. Design of a genome-wide shRNA screen in ESCs. (A) The principle of the genome-wide shRNA screen. See text for further details. (B) Gene expression analysis of *Gata6*, *Pax3* and *Gata1* using qRT-PCR in WT, *Eed* conditional knockout and *Ring1A*, *Ring1B* double conditional knockout ESCs. *Gata6* and *Pax3* are both PcG target genes in ESCs. *Gata1* is repressed in ESCs but is not a PcG target. Error bars represent 95% confidence intervals of three technical replicates.

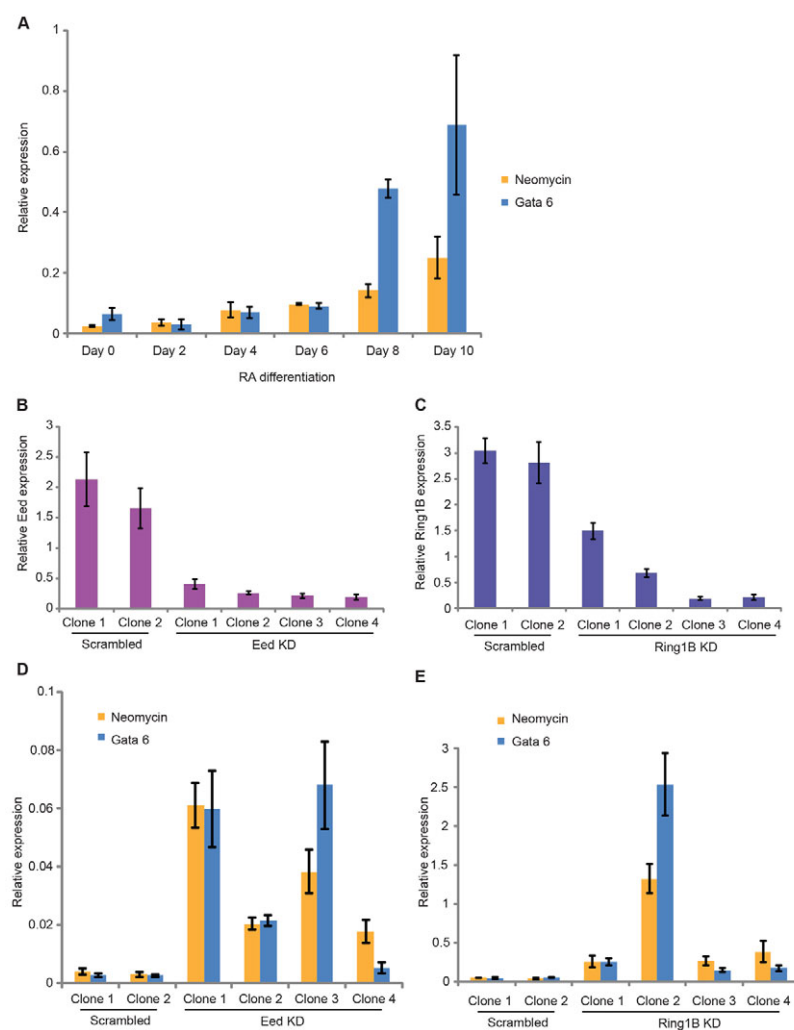


Fig. 2. Validation of the reporter cell line. (A) qRT-PCR analysis during retinoic acid differentiation of ESCs showing neomycin and *Gata6* expression. (B) *Eed* expression analysis of clones containing scrambled or shRNA targeting *Eed*. (C) *Ring1B* expression analysis of clones containing scrambled or shRNA targeting *Ring1B*. (D) Neomycin and *Gata6* expression of the clones shown in B. (E) Neomycin and *Gata6* expression of the clones shown in C. Error bars represent 95% confidence intervals of three technical replicates. KD, knockdown.

determined that *Gata6* and *Pax3* repression is highly dependent on the presence of PcG proteins (Fig. 1B). We elected to use the *Gata6* locus as a reporter as it was strongly upregulated in both PRC1 and PRC2 mutant ESCs, with expression levels >25-fold higher than in the wild-type (WT) control.

Our screening strategy enabled the use of a pooled shRNA library, taking advantage of unique features of ESCs. Upon transduction of the cells, at an MOI of 1 (one virus particle/one shRNA per cell), a single gene will be targeted for knockdown within each cell. If this gene plays a role in repression of the neomycin reporter gene, de-repression of the neomycin selectable marker will occur and the cell will be resistant to G418. As ESCs grow as colonies representing clones of an individual cell, this allows a single colony containing a single shRNA to be picked and identified by PCR and Sanger sequencing. This strategy avoids expanding the cells and therefore removes the problem of over-representation of a particular gene due to a knockdown conferring a selective advantage. In addition, in a viability-based screen such as this, only the positive 'hits' in the screen will survive and this reduces the amount of work required to screen the whole genome.

We used the pooled MISSION mouse TRC shRNA library that is available commercially, packaged into lentivirus (Sigma MISSION Lentiplex library SHPM01) (Moffat et al., 2006). The library is cloned into the lentiviral vector pLKO.1, which contains a

puromycin selectable marker with a PGK promoter for stable integration, and the RNA Pol III promoter U6 to express the shRNA. The shRNAs consist of a pre-miRNA-like 21-nucleotide double-stranded stem, with a six-nucleotide loop, which is then processed into functional siRNAs in the cells (Root et al., 2006). The library covers at least 15,000 mouse genes, with an average of five hairpins targeting each gene.

Validation of the *Gata6* reporter cell line

A promoterless neomycin reporter gene was targeted to the start codon of one allele of the endogenous *Gata6* gene. The neomycin gene contained a stop codon with a polyA signal, which prevented transcription of the remaining *Gata6* exons. A puromycin selectable marker used to monitor integration was subsequently removed by transient Cre recombinase expression to leave a single LoxP site downstream of the neomycin gene (supplementary material Fig. S1).

We monitored expression of the reporter allele specifically using primers spanning *Gata6* exon1 and the neomycin reporter gene (supplementary material Fig. S1). Similarly to the endogenous *Gata6* gene, the neomycin reporter gene was repressed in ESCs, had chromatin modifications indicative of PcG-mediated repression (Fig. 4D) and followed an identical pattern of upregulation upon retinoic acid differentiation (Fig. 2A). This indicates that insertion of the neomycin cassette did not interfere either with PcG

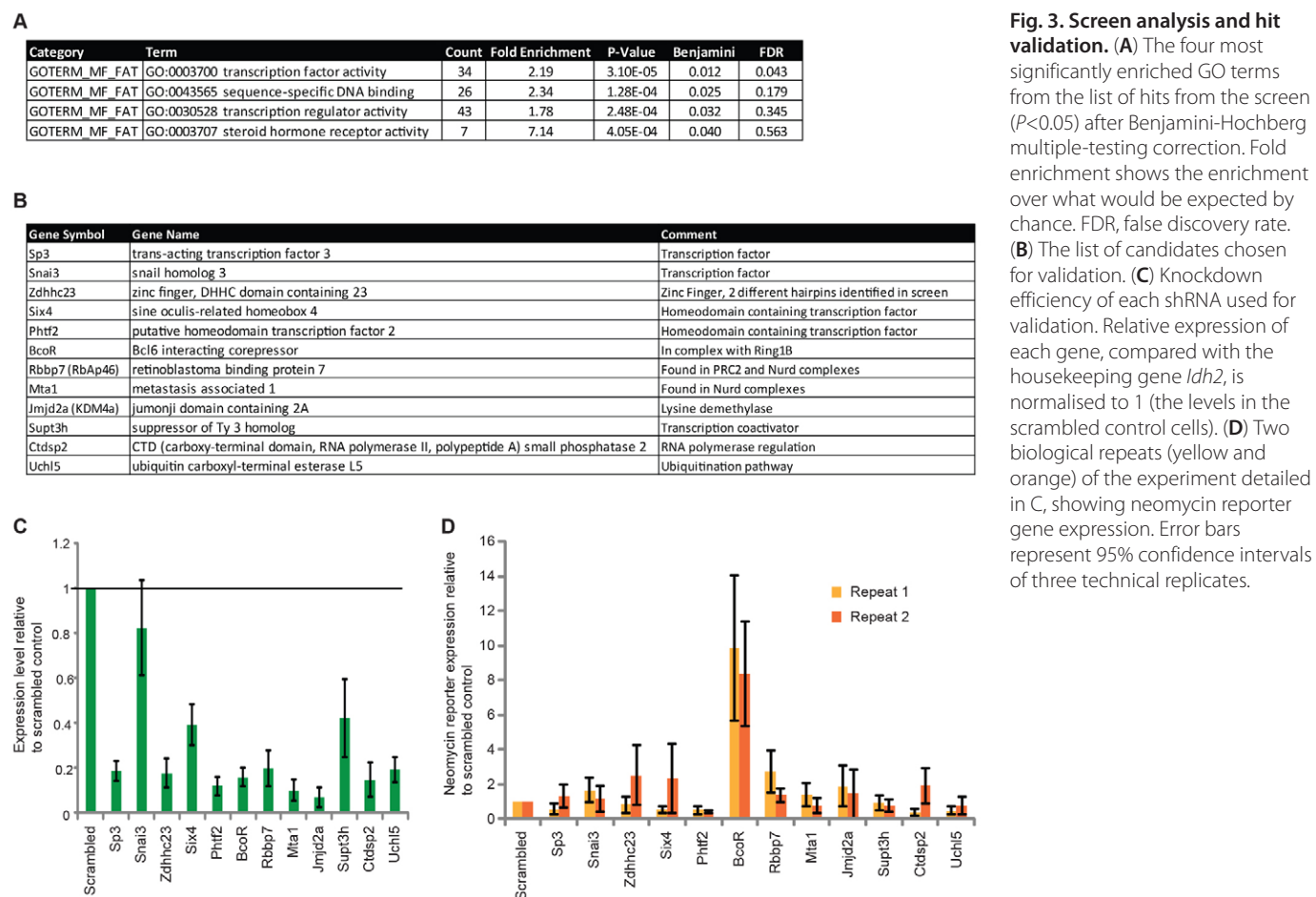


Fig. 3. Screen analysis and hit validation. (A) The four most significantly enriched GO terms from the list of hits from the screen ($P < 0.05$) after Benjamini-Hochberg multiple-testing correction. Fold enrichment shows the enrichment over what would be expected by chance. FDR, false discovery rate. (B) The list of candidates chosen for validation. (C) Knockdown efficiency of each shRNA used for validation. Relative expression of each gene, compared with the housekeeping gene *Idh2*, is normalised to 1 (the levels in the scrambled control cells). (D) Two biological repeats (yellow and orange) of the experiment detailed in C, showing neomycin reporter gene expression. Error bars represent 95% confidence intervals of three technical replicates.

recruitment to the locus or with the function of *cis*-regulatory elements.

To validate the reporter cell line, we established stable knockdown lines using shRNAs directed against Eed and Ring1B, core components of PRC2 and PRC1, respectively. Four independent clones that showed significant knockdown levels (Fig. 2B,C) were analysed for expression of alleles with the integrated neomycin resistance gene and the intact *Gata6* gene. De-repression of both alleles occurred both in Eed and Ring1B knockdown cell lines relative to scrambled shRNA controls (Fig. 2D,E). However, the degree of upregulation of *Gata6* was dependent on the individual knockdown clone analysed (Fig. 2D,E), probably attributable to the heterogeneity that exists within ESC populations (Chambers et al., 2007) (see also below).

Screening in ESCs

Having validated the reporter cell line, we proceeded to carry out a full genome-wide shRNA screen. Pilot studies were performed in order to optimise the number of cells and MOI, along with the timings and concentration of puromycin, for stable integration of the shRNA hairpin, and G418, which acts as the reporter. An important consideration is that the ESCs used in this screen were grown in serum plus leukaemia-inhibitor factor and, therefore, are not a fully homogeneous population (Chambers et al., 2007). The level of selection (concentration of the G418) therefore needed to be carefully modulated to take into account variations of reporter expression within the population of ESCs.

The conditions of the screen were empirically determined such that a negative scrambled shRNA control, when transduced into

cells, gave no surviving colonies upon G418 selection. The optimised protocol, outlined in supplementary material Fig. S2, was then used to screen the pLKO.1 Lentiplex shRNA library, comprising ~80,000 independent hairpins. Because lentiviral transduction of ESCs is around five to ten times less efficient than with other cell types, practical considerations limited us to screening the library at approximately tenfold coverage.

The shRNA library is divided into ten pools, each containing ~8000 hairpins, and each pool was screened individually. Results from sequencing the hairpins from 480 individual colonies are shown in supplementary material Table S2, and analysis of the list using the gene functional classification tool (DAVID) (Huang et al., 2007; Huang et al., 2009) in supplementary material Table S3. Functional enrichment analysis was also carried out using the 480 hits from the screen (supplementary material Table S4) and the gene ontology (GO) terms significantly enriched, after a multiple testing correction, are shown in Fig. 3A. Encouragingly, the most significantly enriched term is transcription factor activity (GO:0003700), consistent with *a priori* predictions of factors likely to have a role in *Gata6* repression.

Validation of putative regulators

Twelve genes, including transcription factors, homeodomain factors, components of PcG and Nurd complexes, and chromatin and ubiquitylation factors, were selected for further validation (Fig. 3B). An identical sequence to the hairpin recovered from the screen was used to knockdown independently the gene of interest in stable, pooled ESC lines. Figure 3C shows that, with the exception of *Snai3*, a good level of knockdown was achieved in all cases. We

assayed the expression of the reporter gene in two biological repeats (Fig. 3D). One of the candidates, *Bcor*, gave a robust and significant upregulation of neomycin expression in the pooled population, but only limited effects were observed with the other knockdowns. In light of this, we further analysed individual stable knockdown clones for two candidates identified in the screen: *Rbbp7*, a component of both the PRC2 and Nurd complex, and *Mta1*, a component of the Nurd complex (supplementary material Fig. S3). In both cases, we observed upregulation of the reporter and endogenous *Gata6* but, similarly to *Eed* and *Ring1B* knockdowns (Fig. 2B,C), there were substantial differences in the degree of *Gata6* upregulation between individual clones, and this was not directly correlated with the knockdown efficiency. Although the reasons for this variation are not certain, one possibility is that it reflects stochastic variation of the differentiation state and subsequent priming of bivalent genes within a population of ESCs. This stochastic variation of expression might aid the identification of novel factors if in certain cells only a smaller knockdown efficiency is required to reach higher reporter expression.

Bcor-mediated repression of *Gata6* in ESCs

We went on to investigate the role of *Bcor* in *Gata6* repression. *Bcor* is a co-repressor that interacts with the transcription factor *Bcl6*, and was of particular interest because it has been shown to be associated with a non-canonical PRC1 complex containing *Ring1b*, *Nspc1* (*Pcgl1* – Mouse Genome Informatics), *Rybp* and *Kdm2b* (Gearhart et al., 2006; Sánchez et al., 2007; Lagarou et al., 2008; Gao et al., 2012; Tavares et al., 2012). To validate further the role of *Bcor* in *Gata6* repression, we carried out an independent knockdown experiment using a second hairpin to deplete *Bcor* in the reporter cell line (Fig. 4A). Analysis of expression levels of the *Gata6* allele with the integrated neomycin resistance cassette and the intact *Gata6* allele demonstrated de-repression relative to scrambled control to a similar level to that seen with the original shRNA hairpin (Fig. 4B).

We verified upregulation of neomycin expression by counting the number of neomycin-resistant colonies after transduction of *Bcor* shRNA and G418 selection. By utilising a variety of G418 concentrations, we observed that colonies persisted at 400 µg/ml G418 only upon *Bcor* knockdown (supplementary material Fig. S4). This demonstrates the efficiency of the screening approach for sensitive detection of relatively small changes in neomycin resistance gene expression.

To analyse whether the observed upregulation is mediated by loss of PcG repression, we performed chromatin immunoprecipitation for H3K27me3 in the *Bcor* knockdown cells. This analysis showed a specific loss of the PcG-linked histone modification at both the neomycin reporter gene and across the endogenous *Gata6* locus upon *Bcor* knockdown (Fig. 4C,D). Reduced H3K27me3 has also been observed at selected PcG targets following knockdown of *Kdm2b* (Farcas et al., 2012), a different component of the *Bcor* complex. *Kdm2b*, as well as having a role as an H3K36me2 demethylase (He et al., 2008), also contains a CXXC domain, which can bind to unmethylated CpG dinucleotides. This protein complex links *Bcor* and PcG proteins to G/C sequences (Farcas et al., 2012; He et al., 2013; Wu et al., 2013) and suggests a potential mechanism for PcG recruitment to selected CpG islands in mouse ESCs.

Conclusions

Here, we describe a novel approach to performing genome-wide loss-of-function screens in mouse ESCs, using a commercially available pooled shRNA library and analysing hits by sequencing

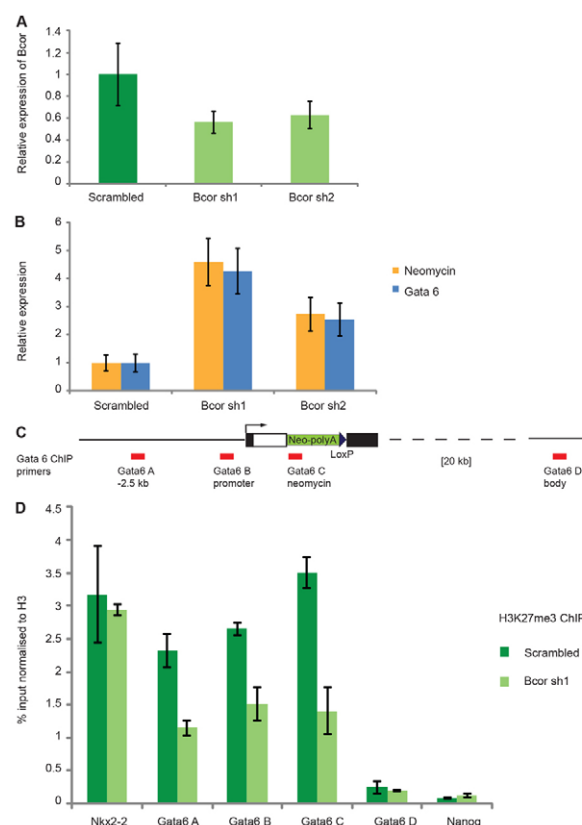


Fig. 4. *Bcor* facilitates Polycomb repression of the *Gata6* locus.

(A) *Bcor* expression after stable knockdown of *Bcor* using two different hairpin sequences or a scrambled shRNA control. Values are shown relative to the scrambled control. (B) Neomycin reporter gene expression and endogenous *Gata6* expression in the same knockdown lines shown in A. (C) Schematic of *Gata6* locus showing ChIP primers Gata6 A–D (supplementary material Table S1). (D) ChIP analysis using H3K27me3 antibody in cells expressing either a scrambled or *Bcor*-specific shRNA. *Nkx2-2* is a positive control and *Nanog* is a negative control. Error bars show 95% confidence interval of three biological repeats.

individual colonies that survive after applying selection. Our screen hits were significantly enriched for transcription factors, and we identified a potential PcG recruitment factor, *Bcor*. It is important to note that the screen does not differentiate between specific regulators of *Gata6* expression and more general regulators, such as PcG repressors. A secondary screen to identify general PcG recruiters could use alternative bivalent gene reporters to find factors that appear in both screens. Similarly, it will be interesting to determine whether there are lineage-specific factors necessary for PcG repression. Given that neomycin selection can be optimised for a broad range of expression levels, a similar strategy could be applied to different reporter loci in ESCs to investigate factors involved in pluripotency, lineage differentiation and epigenetics.

Acknowledgements

We would like to thank members of the lab for discussion and advice, and Andrew Bassett for critical reading of the manuscript. We thank Hiroshi Koide for the *Eed4* cells and Haruhiko Koseki for the ES-ERT2 *Ring1A*^{−/−} cells.

Funding

This work was funded by the Wellcome Trust UK [grant number WT081385]. Deposited in PMC for immediate release.

Competing interests statement

The authors declare no competing financial interests.

Author contributions

S.C. and N.B. designed experiments, S.C. performed experiments and S.C. and N.B. wrote the manuscript.

Supplementary material

Supplementary material available online at

<http://dev.biologists.org/lookup/suppl/doi:10.1242/dev.094615/-/DC1>

References

- Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M. et al. (2006). Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8**, 532-538.
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K. et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315-326.
- Blackledge, N. P., Zhou, J. C., Tolstorukov, M. Y., Farcas, A. M., Park, P. J. and Klose, J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell* **38**, 179-190.
- Blair, K., Wray, J. and Smith, A. (2011). The liberation of embryonic stem cells. *PLoS Genet.* **7**, e1002019.
- Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, 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.
- Caparros, M. L., Alexiou, M., Webster, Z. and Brockdorff, N. (2002). Functional analysis of the highly conserved exon IV of XIST RNA. *Cytogenet. Genome Res.* **99**, 99-105.
- Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L. and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230-1234.
- 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.
- Endoh, M., Endo, T. A., Endoh, T., Fujimura, Y., Ohara, O., Toyoda, T., Otte, A. P., Okano, M., Brockdorff, N., Vidal, M. et al. (2008). Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. *Development* **135**, 1513-1524.
- Farcas, A. M., Blackledge, N. P., Sudbery, I., Long, H. K., McGouran, J. F., Rose, N. R., Lee, S., Sims, D., Cerase, A., Sheahan, T. W. et al. (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. *Elife* **1**, e00205.
- Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y. and Reinberg, D. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* **45**, 344-356.
- Gearhart, M. D., Corcoran, C. M., Wamstad, J. A. and Bardwell, V. J. (2006). Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. *Mol. Cell Biol.* **26**, 6880-6889.
- He, J., Kallin, E. M., Tsukada, Y. and Zhang, Y. (2008). The H3K36 demethylase Jhdmlb/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). *Nat. Struct. Mol. Biol.* **15**, 1169-1175.
- He, J., Shen, L., Wan, M., Taranova, O., Wu, H. and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. *Nat. Cell Biol.* **15**, 373-384.
- Hu, G. and Luo, J. (2012). A primer on using pooled shRNA libraries for functional genomic screens. *Acta Biochim. Biophys. Sin. (Shanghai)* **44**, 103-112.
- 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.
- Huang, D., Sherman, B. T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M. W., Lane, H. C. et al. (2007). DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* **35** Web Server issue, W169-W175.
- Huang, D., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44-57.
- Hussein, S. M. and Nagy, A. A. (2012). Progress made in the reprogramming field: new factors, new strategies and a new outlook. *Curr. Opin. Genet. Dev.* **22**, 435-443.
- Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., Presser, A., Nusbaum, C., Xie, X., Chi, A. S. et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* **4**, e1000242.
- Lagarou, A., Mohd-Sarip, A., Moshkin, Y. M., Chalkley, G. E., Bezstarosti, K., Demmers, J. A. and Verrijzer, C. P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev.* **22**, 2799-2810.
- Moffat, J., Grueneberg, D. A., Yang, X., Kim, S. Y., Kloepper, A. M., Hinkle, G., Piqani, B., Eisenhaure, T. M., Luo, B., Grenier, J. K. et al. (2006). A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell* **124**, 1283-1298.
- Nesterova, T. B., Popova, B. C., Cobb, B. S., Norton, S., Senner, C. E., Tang, Y. A., Spruce, T., Rodriguez, T. A., Sado, T., Merkenschlager, M. et al. (2008). Dicer regulates Xist promoter methylation in ES cells indirectly through transcriptional control of Dnmt3a. *Epigenetics Chromatin* **1**, 2.
- Root, D. E., Hacohen, N., Hahn, W. C., Lander, E. S. and Sabatini, D. M. (2006). Genome-scale loss-of-function screening with a lentiviral RNAi library. *Nat. Methods* **3**, 715-719.
- Rougeulle, C., Chaumeil, J., Sarma, K., Allis, C. D., Reinberg, D., Avner, P. and Heard, E. (2004). Differential histone H3 Lys-9 and Lys-27 methylation profiles on the X chromosome. *Mol. Cell Biol.* **24**, 5475-5484.
- Sánchez, C., Sánchez, I., Demmers, J. A., Rodríguez, P., Strouboulis, J. and Vidal, M. (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdmlb histone demethylase and the Bcl6 interacting corepressor. *Mol. Cell Proteomics* **6**, 820-834.
- Simon, J. A. and Kingston, R. E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **10**, 697-708.
- Simpson, K. J., Davis, G. M. and Boag, P. R. (2012). Comparative high-throughput RNAi screening methodologies in C. elegans and mammalian cells. *New Biotechnol.* **29**, 459-470.
- Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H. et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678.
- Ura, H., Usuda, M., Kinoshita, K., Sun, C., Mori, K., Akagi, T., Matsuda, T., Koide, H. and Yokota, T. (2008). STAT3 and Oct-3/4 control histone modification through induction of Eed in embryonic stem cells. *J. Biol. Chem.* **283**, 9713-9723.
- Wu, X., Johansen, J. V. and Helin, K. (2013). Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol. Cell* **49**, 1134-1146.

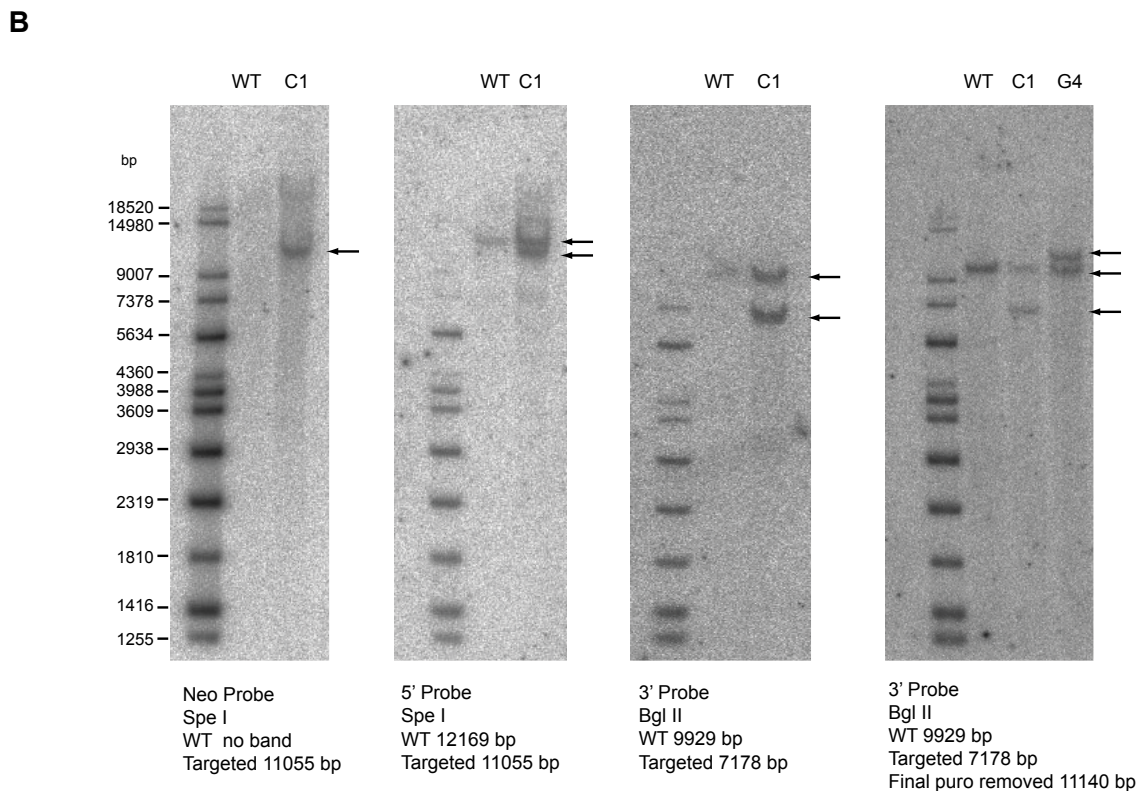
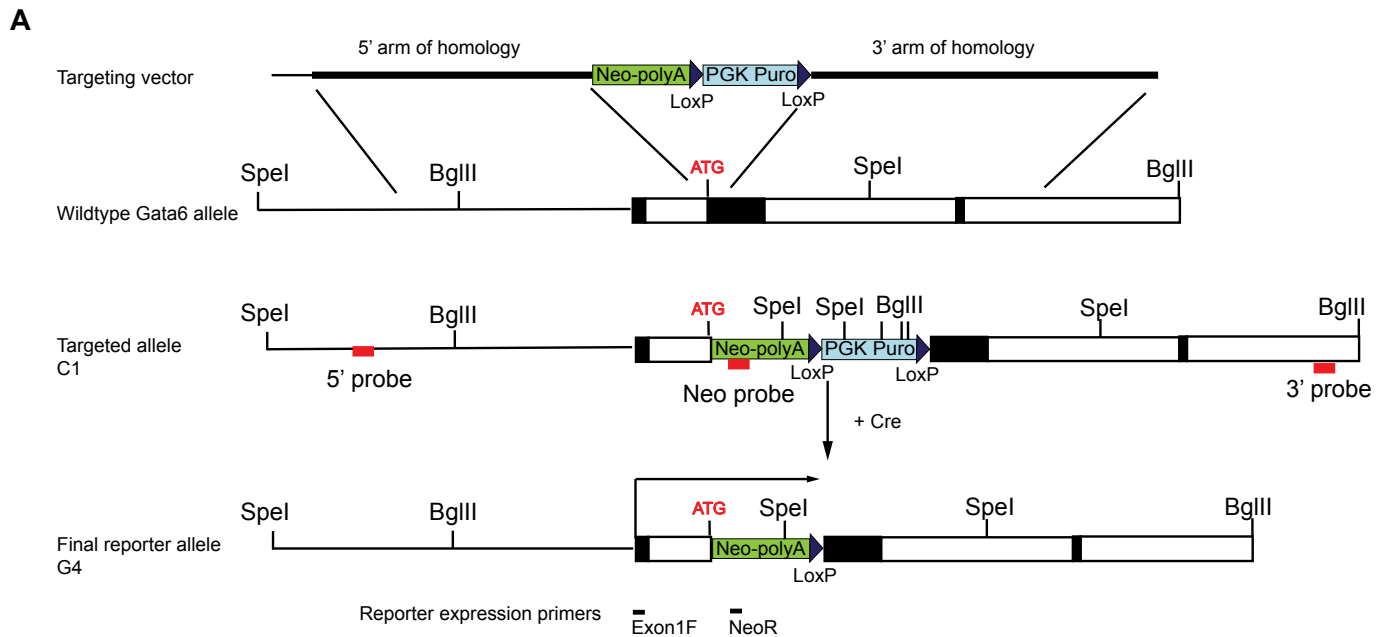


Fig. S1. Targeting of the neomycin reporter gene to the endogenous *Gata6* locus. (A) A schematic of the strategy for targeting a promoterless neomycin with a polyA signal at the start codon of endogenous *Gata6* gene. The first targeting step introduces a floxed puromycin selection cassette, which is then removed using a transient expression of Cre to create the final reporter cell line. Positions of the 5', 3' and Neo Southern probes are shown as are the locations of *Bgl*II and *Spe*I restriction sites, which are used for the Southern blot strategies. The positions of the qRT-PCR primers, to identify the reporter specific transcript, are indicated. (B) Phosphorimager scans of Southern blots to show gene targeting. Probes, restriction sites and expected sizes for each allele are detailed. WT, E14 genomic DNA; C1, first targeted line including the floxed puromycin cassette; G4, final targeted line after removal of puromycin cassette.

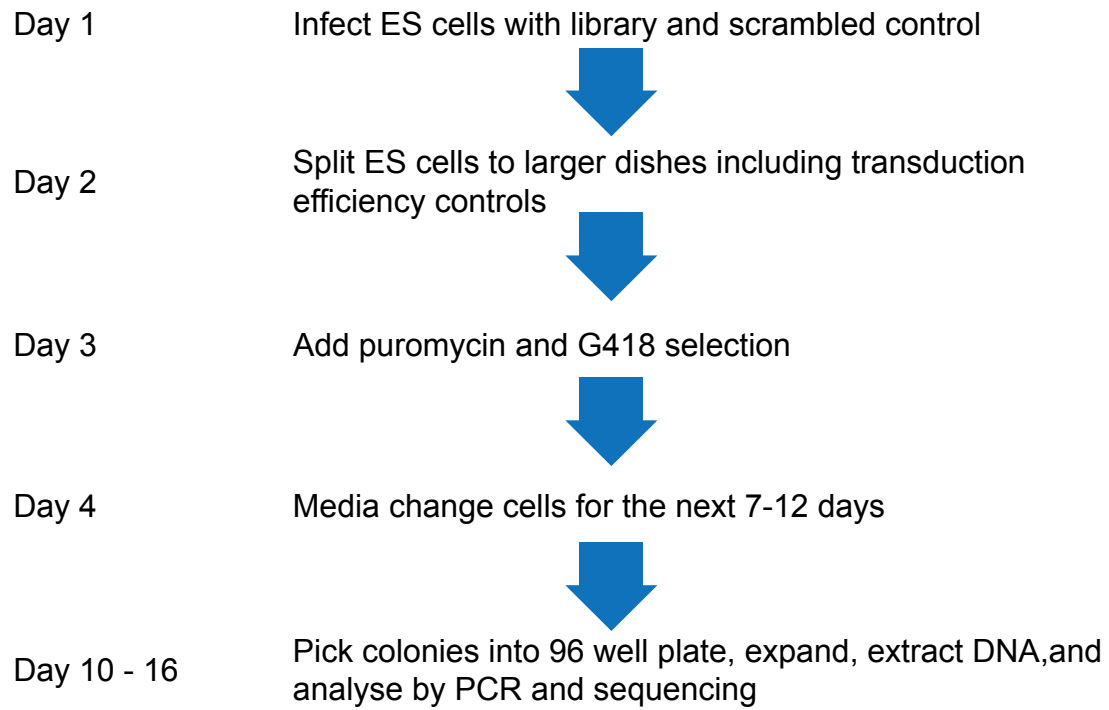


Fig. S2. Schematic overview of the screening process.

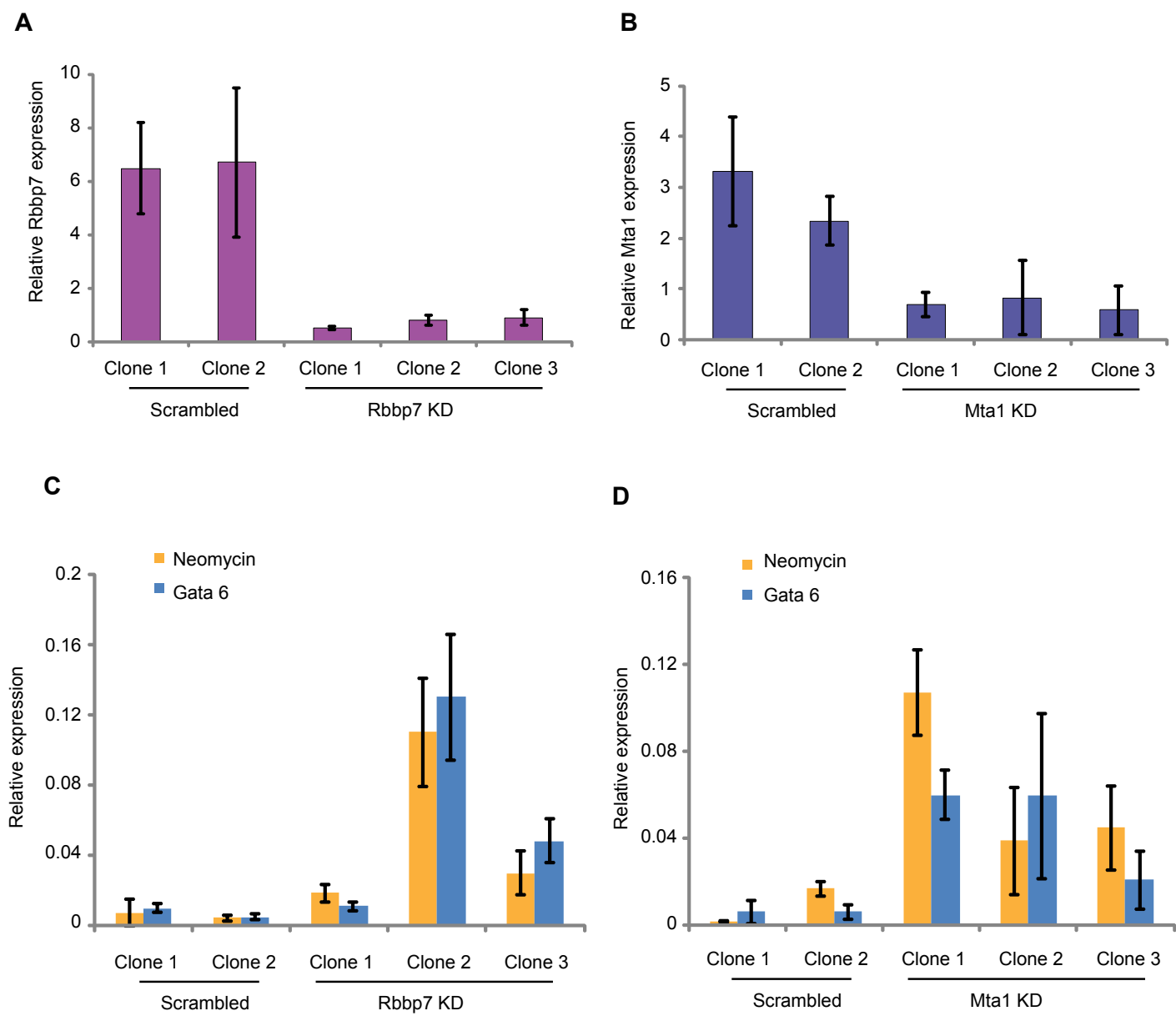


Fig. S3. Validation of *Rbbp7* and *Mta1* hits. (A) *Rbbp7* expression analysis of clones containing scrambled or shRNA targeting *Rbbp7*. (B) *Mta1* expression analysis of clones containing scrambled or shRNA targeting *Mta1*. (C) Neomycin and *Gata6* expression of clones shown in A. (D) Neomycin and *Gata6* expression of clones shown in B. Error bars represent 95% confidence intervals of three technical replicates.

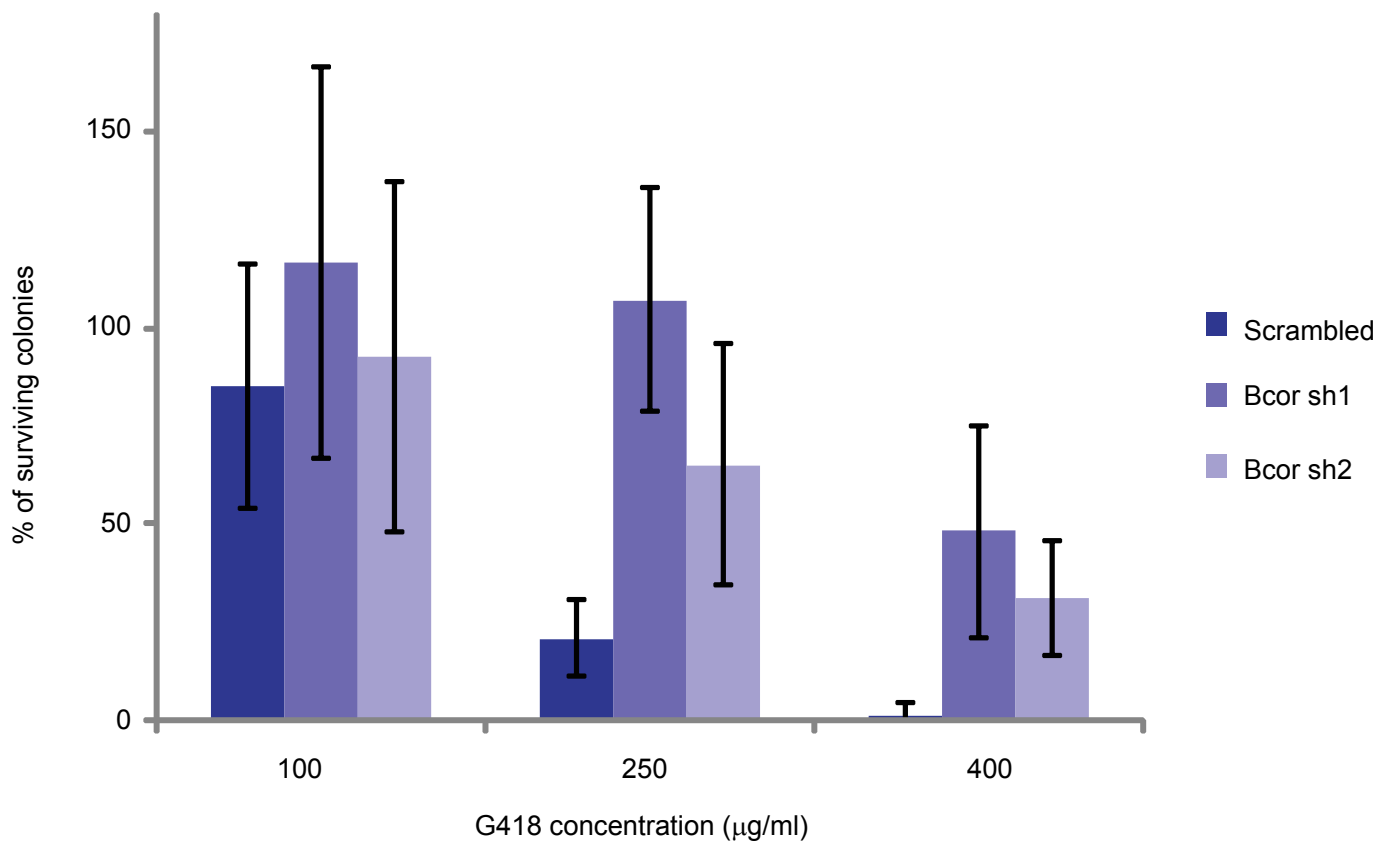


Fig. S4. Colony-counting experiment after Bcor knockdown. Number of colonies surviving at three concentrations of G418 (100, 250 and 400 µg/ml) shown as a percentage of number of colonies surviving without G418, in cells stably expressing either a scrambled control shRNA or a Bcor knockdown shRNA (Bcor sh1 and Bcor sh2). Error bars show 95% confidence intervals of three biological replicates.

Table S1. Primer and oligonucleotide sequences.

[Download Table S1](#)

Table S2. Full results of the screen. List of genes (symbols and full names) together with the sequence of the hairpin identified in the screen. The library comes as ten pools for ease of screening. The list of genes is arranged in pool numbers and screening was carried out using each pool individually. The shRNA hairpins for each gene are within a single pool, but there are no specific categories of genes within each pool.

[Download Table S2](#)

Table S3. Gene functional classification of screen hits. List of screen hits analysed by DAVID using the gene functional classification tool.

[Download Table S3](#)

Table S4. Functional annotation of screen hits. List of screen hits analysed by DAVID using the functional annotation tools. Category: the database from which the term is found. Term: specific name that is enriched. Count: number of genes included in that term and as a percentage of total. %: percentage of hits from the screen which fall into that term. *P* value. Genes: list of genes included in the term. List total: Total number of screen hits represented in that category. Pop hits: Number of genes in that category. Pop total: total number of genes. Fold enrichment: enrichment over what would be expected by chance. Bonferroni and Benjamini-Hochberg: *P* value adjusted for multiple testing correction. FDR: false discovery rate.

[Download Table S4](#)