An Embryonic Stem Cell-Based System for Rapid Analysis of Transcriptional Enhancers

Citation

Published Version
doi:10.1002/dvg.20820

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:8715742

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
An Embryonic Stem Cell Based System for Rapid Analysis of Transcriptional Enhancers

Kaloyan M. Tsanov\textsuperscript{1,2}, Yuichi Nishi\textsuperscript{1}, Kevin A. Peterson\textsuperscript{1}, Jing Liu\textsuperscript{1,3}, Manfred Baetscher\textsuperscript{2,3} and Andrew P. McMahon\textsuperscript{1,2,3*}

\textsuperscript{1}Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts

\textsuperscript{2}Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts

\textsuperscript{3}Harvard Stem Cell Institute, Cambridge, Massachusetts

*Correspondence: Andrew P. McMahon, Department of Stem Cell and Regenerative Biology, SF358A, Harvard University, 7 Divinity Avenue, Cambridge MA, 02138; Phone: (617) 496-3757; Fax: (617) 496-3763; E-mail: mcmahon@mcb.harvard.edu

Running Head: ESC based system for rapid enhancer analysis

Keywords: embryonic stem cells; gene targeting; enhancer elements; transcription; gene regulatory networks

Grant Sponsor: NIH, Grant number: R37 NS033642 (to A.P.M.)
Abstract

With the growing use of genome-wide screens for cis-regulatory elements, there is a pressing need for platforms that enable fast and cost-effective experimental validation of identified hits in relevant developmental and tissue contexts. Here, we describe a murine embryonic stem cell (ESC) based system that facilitates rapid analysis of putative transcriptional enhancers. Candidate enhancers are targeted with high efficiency to a defined genomic locus via recombinase-mediated cassette exchange (RMCE). Targeted ESCs are subsequently differentiated in vitro into desired cell types, where enhancer activity is monitored by reporter gene expression. As a proof of principle, we analyzed a previously characterized, Sonic hedgehog (Shh)-dependent, V3 interneuron progenitor (pV3)-specific enhancer for the Nkx2.2 gene, and observed highly specific enhancer activity. Given the broad potential of ESCs to generate a spectrum of cell types, this system can serve as an effective platform for the characterization of gene regulatory networks controlling cell fate specification and cell function.
Enhancers are cis-regulatory DNA elements that modulate transcription independent of their position and orientation with respect to the transcriptional start site (Banerji et al., 1981). A given enhancer contains binding sites for different transcription factors; the specific combination of DNA-protein interactions determines whether the enhancer is active or silent, ensuring appropriate spatial and temporal activity of target genes. As key regulators of gene expression, enhancers play critical roles in development, disease, and evolution (Ben-Tabou de-Leon and Davidson, 2007; Visel et al., 2009b). Consequently, a systematic comprehensive decoding of their regulatory actions is an important goal, and especially so for the understanding of how cell fates are determined on differentiation of stem cells.

The advent of powerful genomic technologies has allowed for the unprecedented genome-wide prediction of enhancer elements (Rada-Iglesias et al., 2011; Visel et al., 2009a). However, to rapidly validate and characterize the vast number of genomic regions with regulatory potential remains a considerable challenge, particularly in mammals. In a typical enhancer assay, a putative enhancer element is placed upstream of a minimal promoter driving a reporter gene whose expression is analyzed in an appropriate experimental system (Loots, 2008). There are significant drawbacks associated with many of the existing methodologies. For example, transient transfections of reporter constructs are easy to perform but assess the elements in multiple copies and in a non-chromosomal context. At the other end of the spectrum, mouse transgenesis offers a more natural in vivo context, yet it is expensive, time-consuming, and laborious. This is further aggravated by the fact that the copy number and integration sites of reporter constructs cannot be controlled, thus necessitating the analysis of multiple lines per element to control for position effects at the site of integration and copy number differences that
confound analysis. Given these and other constraints, it is critical to develop faster, scalable, and more cost-effective experimental systems that replicate normal regulatory mechanisms.

Embryonic stem cells (ESCs) are well suited for this purpose, since they are amenable to controlled genetic manipulations (Capecchi, 1989; Turan et al., 2011) and can be differentiated into a variety of cell lineages closely replicating the normal programs of cell fate specification (Keller, 2005). The general utility of ESC based approaches for analysis of cis-regulatory elements has been shown (Xian et al., 2005; Zhang et al., 2008). However, a high-throughput application requires the development of an efficient, reproducible and rapid strategy that enables position-independent activity of potentially any enhancer of interest. To these ends, we developed a novel system that enables reproducible targeting of candidate enhancers through recombinase-mediated cassette exchange (RMCE) at the Gt(ROSA)26Sor locus (hereafter referred to as the Rosa26 locus) (Turan et al., 2011; Zambrowicz et al., 1997). The Rosa26 locus is a transcriptionally permissive environment and eliminates potential position effects associated with random integration (Zambrowicz et al., 1997). Targeted ESCs can be assayed directly or differentiated into various cell types to assess tissue-specific enhancer activity. As a proof of principle, we generated neuralized embryoid bodies (EBs) from ESCs and analyzed Sonic hedgehog (Shh)-mediated activation of the Nkx2.2 gene in V3 interneuron progenitors (pV3) (Briscoe et al., 1999). We selected a pV3-specific enhancer of Nkx2.2 that has previously been characterized in transgenic mice and shown to be dependent upon a single Gli transcription factor binding site (Lei et al., 2006; Vokes et al., 2007). Characterization of this enhancer in vitro recapitulates the in vivo analysis, validating our enhancer analysis platform. In addition, we demonstrate how the system can be further optimized to improve its capability for high-throughput applications.
To allow for high-throughput analysis of putative enhancers, we took advantage of the high efficiency of RMCE and selected the Rosa26 locus as a target site, as its activity in most cell types indicates a chromatin configuration that is predicted to permit normal, position-independent regulation of test enhancers (Zambrowicz et al., 1997). To facilitate RMCE, we targeted a FLPo recombinase expression cassette, flanked by F3 and FRT recognition sites, to the Rosa26 locus (cell line Rosa26 (FLPo)) (Fig. 1a). This configuration allows for high recombination efficiency, prevents internal recombination, and provides an endogenous source of recombinase activity that is conveniently self-terminated upon recombination (Raymond and Soriano, 2007; Seibler and Bode, 1997; Seibler et al., 2005). To prevent random integration and achieve stringent selection of correctly targeted clones, we used a neomycin resistance gene (Neo) that lacks a promoter and an ATG translation start codon (Fig. 1b). Only upon successful recombination, a splice acceptor and ATG codon engineered in the Rosa26 (FLPo) line allow for Neo expression from the endogenous Rosa26 promoter (Beard et al., 2006). Indeed, we observed 98% selection efficiency (48/49 recombinant colonies), in line with similar targeting strategies (Seibler et al., 2005; Wang et al., 2007) (Fig. 1c and data not shown). A correctly integrated enhancer-reporter cassette was flanked by single copies of the full-length chicken β-globin insulator to block local influences on enhancer activity (Burgess-Beusse et al., 2002; Chung et al., 1997). Together, all these features enable a highly efficient analysis, where the need to screen dozens of colonies for correct integration and assess multiple clones per construct due to position effects is eliminated.

Once we optimized the parameters for consistent targeting results, we explored a number of protocol modifications to improve the scalability of the approach (Table 1). On average, at least one recombinant colony per million electroporated cells was obtained consistently under all
conditions tested. As a result, we have adapted our system to a 6-well-plate format, allowing for moderate-throughput applications. Additional optimization, such as testing the use of even fewer cells per electroporation or lipofection-based delivery methods, may improve on targeting efficiency with regard to the number of input cells.

Next, we validated our platform for enhancer analysis (schematized in Fig. 2). A 420-bp enhancer located 1.7 kb upstream of the *Nkx2.2* coding sequence recapitulates the pV3-specific *Nkx2.2* expression pattern *in vivo* (Lei et al., 2006). This activity is Shh-dependent; mutation of a unique Gli factor binding site—Gli proteins are the transcriptional effectors of the Shh pathway—abolishes enhancer activity. Rosa26 (FLPo) cells were targeted with this element and subjected to a 5-day EB differentiation protocol that utilizes the neuralizing activity of retinoic acid (RA), in conjunction with a small-molecule agonist of the Shh pathway (SAG), to generate a mix of Shh-dependent neural progenitors that includes *Nkx2.2*+, pV3 cells (Chen et al., 2002; Wichterle et al., 2002). As expected, nearly all *Nkx2.2*+ cells co-expressed the β-gal reporter (91.1 ± 1.7%) (Fig. 3i-l, q; data from two independent clones). We did detect a population of β-gal single-positive cells (20.6 ± 3.0% of all β-gal+ cells), consistent with *in vivo* observations (Lei et al., 2006). This likely reflects the absence of sequences that refine *Nkx2.2* expression after Shh activation, notably those that suppress activity in the floor plate. Importantly, enhancer activity was Shh-dependent. First, omission of the Shh agonist abolished reporter along with *Nkx2.2* expression (2.4 ± 1.1% β-gal+ cells) (Fig. 3e-h, r). Further, a mutation of the critical Gli binding site similarly abrogated reporter expression (1.8 ± 0.6% β-gal+ cells) (Fig. 3m-p, q, r). An enhancer-less reporter showed minimal levels of reporter activation (1.6 ± 0.4% β-gal+ cells) despite appropriate derivation of V3 progenitors (25.9 ± 4.2% *Nkx2.2*+ cells), indicating that promoter background is negligible and does not influence the analysis (Fig. 3a-d, q, r).
Collectively, these data demonstrate that transcriptional activity of the enhancer element can be reliably analyzed in this *in vitro* system.

We also developed an exchange vector with a nuclear *lacZ::GFP* fusion reporter to simplify scoring of nuclear signals. Using this modified construct, we obtained essentially identical results (91.2 ± 1.5% *Nkx2.2*+ cells co-expressed reporter for the wild-type enhancer; 15.5 ± 2.3% reporter cells were single-positive) (Supp. Fig. 1). Besides facilitating scoring of immunostained samples, this reporter also enables potential FACS analysis with the GFP marker as an additional means of quantitatively assessing enhancer activity.

In conclusion, we have developed a novel murine ESC based system for fast, scalable, and cost-effective analysis of putative transcriptional enhancers. Our methodology can be applied to other neural and non-neural enhancers, given the broad array of cell types that can be generated from ESCs (Keller, 2005), and potentially to human ESCs, since protocols for their genetic manipulation and directed differentiation are available (Hockemeyer and Jaenisch, 2010; Irion *et al.*, 2008). The same general strategy can also be implemented for the analysis of other types of *cis*-regulatory elements, such as silencers or insulators. Thus, our system provides a versatile tool that could find broad application in large-scale studies of the transcriptional gene regulation in ESCs or diverse developmental and tissue contexts.

**Methods**

**Plasmid construction**

To construct the *Rosa26* targeting vector (pRosa-17), a PGK-Neo-3xpA cassette was replaced with ATG-F3-PGK-Puro-pA and PGK-FLPo-pA cassettes in the pBigT vector (Srinivas *et al.*, 2006).
The former fragment was PCR amplified from the pPGKpuro vector (Tucker et al., 1996) (ATG and F3 sequences were included in the primer tails), and the latter obtained by excision from the pPGKFLPobpA vector (Raymond and Soriano, 2007). Next, an FRT-1xIns fragment was PCR amplified from the PHSP68lacZ2XINS vector (Vokes et al., 2007) (FRT sequence was included in the primer tail), inserted downstream of FLPo-pA to replace a pA sequence in the pBigT backbone, and the entire ATG-to-1xIns fragment cloned into pRosaPAS (Mao et al., 2005).

To construct the exchange vector (pXCHG3), an F3-ΔATGNeo-pA-3xpA cassette was PCR amplified from pBigT (Srinivas et al., 2001) (ATG was excluded and F3 sequence included in the primer tails), and an FRT site generated by annealing of pre-synthesized oligonucleotides (IDT). The two fragments were ligated and inserted downstream of a pA sequence cloned from pPGKFLPobpA (Raymond and Soriano, 2007) in pBlueScript (Stratagene). A 1xIns-hsp68-lacZ-pA cassette (Vokes et al., 2007) was cloned immediately upstream of the FRT site. F3 and FRT sequences used in all vectors were respectively: GAAGTTCCCTATTCCGAAGTTCCTATTCTT-CAAATAGTATAGGAACTTC and GAAGTTCCCTATACCTTCTAGAGAATAGGAACTTC-CAATAGTAGGAACCTTC. Nkx2.2 enhancer variants were subsequently inserted at a unique KpnI site upstream of the hsp68 promoter. The Nkx2.2-420bp and Nkx2.2-GliIM constructs were generated as previously described (Lei et al., 2006).

To generate the modified exchange vector (pXCHG-IHZG), 1xIns and hsp68 sequences were PCR amplified from PHSP68lacZ2XINS (Vokes et al., 2007), and cloned into the pBlueScript-KS(+) vector (Stratagene) to generate pBS-Ins-Hsp68. Next, the lacZ coding sequence was PCR amplified from PHSP68lacZ2XINS (Vokes et al., 2007), and a 3X-NLS-GFP-pA fragment PCR amplified from pCIG (Megason and McMahon, 2002). The two
fragments were cloned into pBS-Ins-Hsp68 to generate pBS-1xIns-hsp68-lacZ-3XNLS-GFP-pA (pIHZG). An attR1-attR2 Gateway® selection cassette (Invitrogen) was inserted immediately upstream of hsp68, and the entire 1xIns-to-pA fragment used to replace 1xIns-hsp68-lacZ-pA in pXCHG3. The Nkx2.2-420bp enhancer was cloned into the pENTR1A vector (Invitrogen) and inserted in pXCHG-IHZG via Gateway® LR recombination, as per manufacturer’s protocol (Invitrogen).

For the optimization experiments, a pCAGGS-FLPo expression vector was constructed by cloning the FLPo coding sequence (Raymond and Soriano, 2007) downstream of the CAGGS promoter of pCIG (Megason and McMahon, 2002).

**Gene targeting**

For generation of the Rosa26 (FLPo) line, 20 μg targeting vector was linearized with SwaI, phenol/chloroform purified, and added to 10^7 V6.5 cells (Eggan et al., 2001) in 800 μl cold PBS. Cells were electroporated at 230 V and 500 μF using Gene Pulser (BioRad), and plated onto 10-cm dishes with puromycin-resistant MEFs. 3 μg/μl puromycin (Sigma) was added on the following day, medium was changed daily, and colonies were picked after 6 days of selection, as described previously (Hogan et al., 1994). Picked clones were expanded in 96-well plates with MEF feeders. Properly targeted clones were identified by junction PCR, expanded, and stocked. Primer sequences for genotyping were as follows: for 5’ end, CCGCCTAAAGAAGAGGCTGT-GCTTTGG (Rosa05) and CAAGGAAACCCTGGACTTCTGC (Rosa15); for 3’ end, CTGGGCTGCTGGTTGATGACCCTGC (Rosa02) and GGGCAATCTGGGAAGGTTCCTTA-AGAA (Rosa11).
For RMCE, 20 μg circular exchange vector was added to 6 x 10^6 Rosa26 (FLPo) cells in 800 μl cold PBS. Cells were electroporated at 240 V and 500 μF using Gene Pulser (BioRad), and plated onto 10-cm dishes with neomycin-resistant MEFs. For 6-well-plate format, the same conditions were used, except that 4 x 10^6 cells were electroporated and 1.5 x 10^6 cells plated per well in 6-well plates. 300 μg/ml G418 (Geneticin®, Invitrogen) was added on the following day and medium was changed daily. Colonies were picked 6 days after the onset of selection, expanded in 48-well plates with MEF feeders, and cultured in complete medium containing 200 μg/ml G418. Appropriate recombination events were identified by junction PCR, targeted cells were expanded, and frozen stocks prepared. Primer sequences for genotyping were as follows: for 5’end, CTCTGAGTTGTTATCAGTAAGGGAGC (Xchg01) and GATTGTCTGTTGTGC-CCAGTCATAG (Xchg11); for 3’ end, GTCGCTACCATTACCAGTTGGTCT (Xchg02) and CCCAGATGACTTATCCTCCATTT (Xchg12). PCR conditions were as follows: 94ºC for 2 min; 35 cycles of 94ºC for 10 s, 55ºC for 30 s, 68ºC for 1 min (5’ end) or 2 min (3’ end); 68ºC for 5 min. The Taq PCR Core Kit was used following manufacturer’s protocol (Qiagen).

For testing different targeting conditions, a given parameter was modified and the effect of the modification on the yield of recombinant colonies determined as indicated in Table 1. When exogenous FLPo was provided, this was achieved by co-electroporation with 20 μg pCAGGS-FLPo. Each condition was repeated in at least three replicates and data on colony yield are presented as mean ± SEM, with the exceptions noted in Table 1.

**ESC culture and directed differentiation**

All ESC lines were maintained using standard procedures, in complete ESC medium containing 15% fetal bovine serum (FBS; Hyclone) and 1,000 units/ml recombinant leukemia inhibitory
factor (LIF; Chemicon) (Hogan et al., 1994). Directed differentiation followed published procedures (Wichterle et al., 2002). Two days after EB formation, medium was replaced and supplemented with 500 nM retinoic acid (RA; Sigma) and, where applicable, 1 μM of the Shh-agonist, SAG (Alexis Biochemicals). EBs were cultured for additional 3 days to induce neural progenitor stages, at which point they were harvested, fixed, sectioned and processed for immunostaining, as described previously (Wichterle et al., 2002).

**Immunocytochemistry and quantification**

Primary antibodies were as follows: rabbit anti-β-galactosidase (1:40,000, Cappel), and mouse anti-Nkx2.2 (1:20, Developmental Studies Hybridoma Bank). Secondary antibodies were appropriately-conjugated IgG (H+L) Alexa Fluor® 488 or 568 (1: 500, Invitrogen). Stained sections were imaged on a LSM510/710 META confocal microscope (Zeiss). To quantify co-expression, β-gal and Nkx2.2 immunoreactivity was quantified by individual cell counts, and the overlap determined relative to the number of Nkx2.2+ cells. To calculate the proportion of β-gal single-positive cells, the latter were counted and represented as a percentage of all β-gal+ cells. To calculate the total proportion of β-gal+ or Nkx2.2+ cells, the respective cells were counted and represented as a percentage of Hoechst-positive cells. In all cases, regions corresponding to fields of view (40X magnification) of at least five different EBs (three for total counts and no-SAG controls) were scored for each sample. Data are presented as mean ± SEM. Differences between samples were compared by one-way ANOVA, where statistical significance is defined as P < 0.05.
Acknowledgements

We thank the staff at the Harvard Stem Cell Institute Genome Modification Facility for assistance with the generation of the Rosa26 (FLPo) line, A. Kobayashi, J. McMahon, L. Chen and S. Ohba for technical advice, B. Allen for feedback on the draft, and all members of the McMahon laboratory for invaluable help and discussions. K.M.T. was supported by the Harvard College Research Program and the Pechet Family Fund for Undergraduate Research. K.A.P. was supported by an NRSA postdoctoral fellowship (F32 GM087939). Work in A.P.M.’s laboratory was supported by a grant from the NIH (R37 NS033642).

References


Figure Legends

Figure 1. Engineering the Rosa26 locus for RMCE analysis of enhancer activity. (a) Schematic of the strategy to generate the Rosa26 (FLPo) allele. ‘X’ marks the insertion point within the Rosa26 locus (genome coordinates chr6:113,026,031 (mm9)). (b) Schematic of the RMCE used to target putative enhancers. Gray arrowheads denote primers used for PCR screening for positive clones. (c) Representative PCR results from a targeting experiment. The highlighted bands indicate correct integration at the 5’ (left) and 3’ (right) ends. 1-7, individual clones; NC, negative control (Rosa (FLPo) DNA).

Abbreviations: SA, splice acceptor; ATG, translation initiation codon; F3/FRT, recombination sites; PGK, phosphoglycerokinase promoter; Puro, puromycin resistance gene; pA, polyadenylation/transcription stop signal; FLPo, codon-optimized FLP recombinase; Ins, chicken β-globin insulator; ΔATG-Neo, promoter/ATG-less neomycin resistance gene; Enh, putative enhancer; hsp68, heat shock protein 68 minimal promoter; lacZ, E. coli β-galactosidase gene.
Figure 2. Flowchart of the experimental analysis of enhancer activity in neuralized EBs.

Putative enhancers (enh) are coupled to the hsp68 minimal promoter driving the expression of a nuclear lacZ::GFP reporter (lacZnGFP), and introduced at a defined locus (Rosa26) in ESCs via RMCE. Following highly efficient selection for neomycin resistance, successful recombinants are expanded and differentiated in vitro to desired cell types, which results in the expression of specific markers (red and orange). Neuralized EB sections are immunostained and analyzed quantitatively to determine overlap between the expression of the reporter gene (green) and an endogenous gene of interest (red). For abbreviations, see Fig. 1. *This time is given for the differentiation protocol used in this study.

Figure 3. Shh-pathway-dependent activation of an Nkx2.2 enhancer in ESC-derived neural progenitors. (a-d) Neural progenitors harboring an empty exchange vector. Activity of the Nkx2.2 enhancer (Nkx2.2-420bp) in the absence (e-h), or presence of SAG (i-l); arrowheads in inset (k) point to Nkx2.2β-gal+ cells. (m-p) SAG-treated neural progenitors harboring an Nkx2.2 enhancer with a mutated Gli binding site (Nkx2.2-GliM). Magnification: 40X. (q) Quantification of β-gal/Nkx2.2 co-expression. **, P < 0.01. (r) Quantification of Nkx2.2 (purple) and β-gal (blue) expression.
Figure 1
Figure 2
Figure 3
Tables

Table 1. Summary of tested protocol modifications.

<table>
<thead>
<tr>
<th>Protocol modification</th>
<th>Number of recombinant colonies (per $10^6$ electroporated cells)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of electroporated cells ($x10^6$)</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>1.0 ± 0.1</td>
</tr>
<tr>
<td>4.8</td>
<td>1.4 ± 0.3</td>
</tr>
<tr>
<td>6.4</td>
<td>2.9 ± 0.1</td>
</tr>
<tr>
<td>Amount of electroporated DNA ($\mu$g/$10^6$ cells)</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>2.5</td>
</tr>
<tr>
<td>8.3</td>
<td>1.7</td>
</tr>
<tr>
<td>12.5</td>
<td>2.9</td>
</tr>
<tr>
<td>Exogenous FLPo</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2.1 ± 0.5</td>
</tr>
<tr>
<td>Yes</td>
<td>2.8 ± 0.7</td>
</tr>
<tr>
<td>Seeding cell density ($x10^5$/cm$^2$)</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>1.4 ± 0.3</td>
</tr>
<tr>
<td>0.6</td>
<td>1.6 ± 0.1</td>
</tr>
<tr>
<td>1.5</td>
<td>2.9 ± 1.7</td>
</tr>
<tr>
<td>5.0</td>
<td>1.2 ± 0.2</td>
</tr>
</tbody>
</table>

*Average of at least three replicates.
Supplementary Figure Legend

Supplementary Figure 1. Nkx2.2-enhancer-driven expression of a nuclear reporter. Activity of the wild-type Nkx2.2 enhancer (Nkx2.2-420bp) in the absence (a-d) or presence (e-h) of SAG; arrowheads in inset (g) point to Nkx2.2β-gal c cells. Magnification: 40X.
Supplementary Figure 1