

## Video Article

# Prediction and Validation of Gene Regulatory Elements Activated During Retinoic Acid Induced Embryonic Stem Cell Differentiation

Zoltan Simandi<sup>\*1</sup>, Attila Horvath<sup>\*2</sup>, Peter Gergely Nagy<sup>1</sup>, Laszlo Nagy<sup>1,2,3</sup><sup>1</sup>Sanford-Burnham-Prebys Medical Discovery Institute at Lake Nona<sup>2</sup>Department of Biochemistry and Molecular Biology, Research Center for Molecular Medicine, University of Debrecen, Medical and Health Science Center<sup>3</sup>MTA-DE "Lendulet" Immunogenomics Research Group, University of Debrecen<sup>\*</sup>These authors contributed equallyCorrespondence to: Laszlo Nagy at [lnagy@sbpdiscovery.org](mailto:lnagy@sbpdiscovery.org)URL: <http://www.jove.com/video/53978>DOI: [doi:10.3791/53978](https://doi.org/10.3791/53978)

Keywords: Transcription, enhancer, gene expression, retinoic acid, Retinoic Acid Receptor, embryonic stem cell, ChIP-seq, eRNA, luciferase assay

Date Published: 6/1/2016

Citation: Simandi, Z., Horvath, A., Nagy, P.G., Nagy, L. Prediction and Validation of Gene Regulatory Elements Activated During Retinoic Acid Induced Embryonic Stem Cell Differentiation. *J. Vis. Exp.* (), e53978, doi:10.3791/53978 (2016).

## Abstract

Embryonic development is a multistep process involving activation and repression of many genes. Enhancer elements in the genome are known to contribute to tissue and cell-type specific regulation of gene expression during the cellular differentiation. Thus, their identification and further investigation is important in order to understand how cell fate is determined. Integration of gene expression data (e.g. microarray or RNA-seq) and results of chromatin immunoprecipitation (ChIP)-based genome-wide studies (ChIP-seq) allows large-scale identification of these regulatory regions. However, functional validation of cell-type specific enhancers requires further *in vitro* and *in vivo* experimental procedures. Here we describe how active enhancers can be identified and validated experimentally. This protocol provides a step-by-step workflow that includes: 1) identification of regulatory regions by ChIP-seq data analysis, 2) cloning and experimental validation of putative regulatory potential of the identified genomic sequences in a reporter assay, and 3) determination of enhancer activity *in vivo* by measuring enhancer RNA transcript level. The presented protocol is detailed enough to help anyone to set up this workflow in the lab. Importantly, the protocol can be easily adapted to and used in any cellular model system.

## Video Link

The video component of this article can be found at <http://www.jove.com/video/53978/>

## Introduction

Development of a multicellular organism requires precisely regulated expression of thousands of genes across developing tissues. Regulation of gene expression is accomplished in large part by enhancers. Enhancers are short non-coding DNA elements that can be bound with transcription factors (TFs) and act from a distance to activate transcription of a target gene<sup>1</sup>. Enhancers are generally *cis*-acting and most frequently found just upstream of the transcription start site (TSS), but recent studies also described examples where enhancers were found much further upstream, on the 3' of the gene or even within the introns and exons<sup>2</sup>.

There are hundreds of thousands of potential enhancers in the vertebrate genomes<sup>1</sup>. Recent methods based on chromatin immunoprecipitation (ChIP) provide high-throughput data of the whole genome that can be used for enhancer analysis<sup>3-9</sup>. Though data obtained by ChIP-seq experiments greatly increases the likelihood to identify cell and tissue-specific enhancers, it is important to keep in mind that detected binding sites do not necessarily identify direct DNA binding and/or functional enhancers. Thus, further functional analysis of newly identified enhancers is indispensable. In this work, we present a basic three-step process of putative active enhancer identification and validation. This includes: 1) selection of putative transcription factor binding sites by bioinformatics analysis of ChIP-seq data, 2) cloning and validation of these regulatory sequences in reporter constructs, and 3) measurement of enhancer RNA (eRNA).

Exposure of embryonic stem (ES) cells to retinoic acid (RA) is frequently used to promote neural differentiation of the pluripotent cells<sup>10</sup>. RA exerts its effects by binding to RA receptors (RAR $\alpha$ ,  $\beta$ ,  $\gamma$ ) and retinoid X receptors (RXR $\alpha$ ,  $\beta$ ,  $\gamma$ ). RARs and RXRs in a form of heterodimer bind to DNA motifs called RA-response elements, that is typically arranged as direct repeats of AGGTCA sequence (called as half site) and regulate transcription. Ligand-treatment experiments allowed the identification of several retinoic acid regulated genes in ES cells<sup>11,12</sup>. However, enhancer elements for many of these genes has not been described yet. To demonstrate how the here-described workflow can be used for enhancer identification and validation we show step-by-step the selection and characterization of two retinoic acid-dependent enhancers in embryonic stem cells.

## Protocol

### 1. Enhancer Selection Based on Chip-seq Analysis

- Download the RXR ChIP-seq raw data fastq file (mm\_ES\_RXR\_24h\_ATRA.fastq.gz) from <http://ngsdebftp.med.unideb.hu/bioinformatics/>
- Download and extract the required BWA index file for the alignment (in our case: Mus\_musculus\_UCSC\_mm10).(ftp://igenome:G3nom3s4u@usdd-ftp.illumina.com/Mus\_musculus/UCSC/mm10/Mus\_musculus\_UCSC\_mm10.tar.gz  
NOTE: Visit to [https://github.com/ahorvath/Bioinformatics\\_scripts](https://github.com/ahorvath/Bioinformatics_scripts) for more information regarding the steps of bioinformatics analysis and to download the scripts used below.
- Align the example fastq file to mm10 genome (use the script: perform\_alignment.sh). This will create a folder with a .bam file and statistics of the alignment. Aligned RXR ChIP-seq data (mm\_ES\_RXR\_24h\_ATRA.bam), and the corresponding index file (.bai) are available here: <http://ngsdebftp.med.unideb.hu/bioinformatics/>  
NOTE: BWA is a software package that aligns relatively short sequences (e.g. ChIP-seq results) to a sequence database, such as the mouse reference genome (e.g. mm10)<sup>18</sup>.
- Run the script (callpeaks.sh) for peak calling and *de novo* motif analysis. Use the .bam file as the input. The script is based on Homer findPeaks<sup>14</sup>.  
NOTE: The output files give us information about the total number of peaks, enrichment of the motifs, the percentage of the background, and the target sequences with motifs, etc. (Figure 1). Typically several motifs are listed in order of their significance.
- Remap the #1 ranked motif (NR half `AGGTCA`) (use the script: remap\_motif.sh)  
NOTE: As the result, the script generates a .bed file that will show which ChIP-peaks are covering genomic regions that contain the canonical binding site of the transcription factor of interest.
- Download and visualize results of the aligned RXR ChIP-seq reads (mm\_ES\_RXR\_24h\_ATRA.bam (also download the .bai)) and the AGGTCA motif occurrences (mm\_ES\_RXR\_24h\_ATRA\_homerpeaks\_motif1\_mm10s\_200\_remaped\_mbed.bed) by Integrative Genomics Viewer (IGV)<sup>15</sup> to identify putative RAR/RXR binding sites (Figure 2 and Figure 3).  
NOTE: Integration of various ChIP-seq data will help to more precisely predict good candidate enhancers<sup>16,17</sup>. Recent studies revealed that active enhancers are typically enriched for P300 and correlate with H3K4me1 and H3K27ac, and are located in open chromatin regions, thereby displaying DNase-I hypersensitivity is also recommended<sup>18-21</sup>.
- Use the "Define a region of interest" in IGV<sup>15</sup> to obtain 200 - 400 bp sequence of the selected genomic region and use these sequences for subsequent primer designs.

### 2. Reporter Assay

NOTE: The luciferin/luciferase system is used as a very sensitive reporter assay for transcriptional regulation. Depending on the enhancer activity luciferase enzyme is produced that will catalyze the oxidation of luciferin to oxyluciferin resulting in bioluminescence, which can be detected. As a first step, identified putative enhancer sequences should be subcloned into a reporter vector (e.g. TK-Luciferase<sup>22</sup>, pGL3 or NanoLuc).

#### 1. Primer Design

- Design PCR primers for amplification of 200 - 400 bp of putative enhancer regions by Primer3 (available here: <http://bioinfo.ut.ee/primer3/>)<sup>23</sup>
- Copy-paste the genomic sequence from IGV that includes the putative binding site into Primer3 (see step 1.7). Set product size range to 250 - 300 bp in Primer3.
- Validate the PCR primers using the UCSC Genome Browser's *In silico* PCR tool (<https://genome.ucsc.edu/cgi-bin/hgPcr>)<sup>24</sup>  
NOTE: This protocol in a later step uses HindIII and BamHI restriction enzymes for cloning. Check whether the PCR amplified sequence as predicted by the UCSC *In-silico* PCR tool will contain AAGCTT or GGATCC restriction sites (e.g. <http://nc2.neb.com/NEBcutter2/>).
- Obtain the required primers from a commercial vendor.

#### 2. PCR Amplification of Enhancer Sequence from Genomic DNA

- Purify template genomic DNA from primary cells (e.g. primary embryonic fibroblasts). Use commercial kit for gDNA isolation and follow the manufacturer's instructions.  
NOTE: High quality of gDNA is essential for successful PCR amplification. Appropriate BAC clones can be used if PCR is not easy from gDNA.
- Prepare 50 µl PCR reaction containing 100 ng gDNA, 5 µl 10x buffer, 3 µl MgSO<sub>4</sub> (25 mM), 5 µl dNTP (2 mM each), 1.5 µl Fwd primer (10 µM), 1.5 µl Rev primer (10 µM), 1 µl DNA polymerase  
NOTE: Use high-fidelity DNA polymerase with a low error rate for the PCR reactions.
- Set the PCR cycle (2 min 95 °C, (20 sec 95 °C, 10 sec 58 - 65 °C, 18 sec 70 °C) x25 repeat, 2 min 70 °C, forever 4 °C). Set annealing temperature with the consideration of the primer's predicted T<sub>m</sub> values.
- Purify the PCR product with a commercial PCR purification kit and follow the manufacturer's instructions.
- Introduce the restriction sites for cloning by repeating the PCR with primers as used above, but adding overhangs on their 5' ends (e.g. Fwd: 5'-ATATAAGCTTxxxxxxxxxxxxxxxxxxx-3' (HindIII), Rev: 5'-TATAGGATCCxxxxxxxxxxxxxxxxxxx-3' (BamHI)).
- Run 5 µl of the PCR product on agarose gel to check for unspecific PCR products.  
NOTE: If more bands are detected, run the entire PCR product and purify the appropriate band by a commercial gel extraction kit.

#### 3. Restriction

NOTE: To clone insert upstream of TK promoter<sup>22</sup> the purified PCR product and the vector (e.g. TK-Luc) should be digested with HindIII and BamHI.

1. Prepare a 50  $\mu$ l reaction mixture containing 1  $\mu$ g purified PCR product, 1  $\mu$ l HindIII (20 U/ $\mu$ l), 1  $\mu$ l BamHI (20 U/ $\mu$ l) and 5  $\mu$ l 10x buffer.
2. Prepare a 250  $\mu$ l reaction mixture containing 5 $\mu$ g vector (TK Luc, or alternative reporter vector), 5  $\mu$ l HindIII (20 U/ $\mu$ l), 5  $\mu$ l BamHI (20 U/ $\mu$ l), 5  $\mu$ l thermosensitive Alkaline Phosphatase (1 U/ $\mu$ l) and 25  $\mu$ l 10x buffer.  
NOTE: AP treatment of the vector greatly decreases the self-ligation of single-digested vector and thus the background.
3. Incubate the reaction mixtures for 4 hr at 37 °C, then purify the digested PCR products and the vector with a commercial PCR purification kit.

#### 4. Ligation

1. Set up the reaction for ligation in PCR tubes including 2  $\mu$ l 10x T4 DNA ligase buffer, 10 ng TK-Luc vector, 50 ng insert (digested PCR product), 1  $\mu$ l T4 DNA ligase (400 U/ $\mu$ l) and nuclease-free water up to 20  $\mu$ l.  
NOTE: Prepare a negative control where the reaction does not contain insert.
2. Gently mix the reaction by pipetting up and down, spin down briefly the PCR tubes using mini-centrifuge and then incubate at room temperature for 10 min.
3. Heat inactivate at 65 °C for 10 min then chill on ice and use 5  $\mu$ l of the product for transformation into 50  $\mu$ l competent cells (e.g. DH5 $\alpha$ ).
4. Use standard heat shock procedure to transform bacteria, pick up colonies and isolate plasmid DNA with a commercial kit. Validate all constructs by sequencing prior to the next steps.

#### 5. Transfection of ES Cells

1. Use 48-well plates. Add 200  $\mu$ l 0.1% gelatin solution/well 30 min prior to cell plating.
2. Plate embryonic stem (ES) cells a day before transfection. Use feeder-free ES cells and plate at a density of  $3 \times 10^4$  cells per well in 250  $\mu$ l ES media/well.
3. Prepare the plasmid mixes (each mix is calculated for 8 wells, 4 untreated and 4 retinoic acid-treated). *Mix 1*: 1,250 ng TK-Luc Empty (negative control), 950 ng  $\beta$ -Galactosidase coding vector ( $\beta$ Gal), *Mix 2*: 1,250 ng TK-Luc Hoxa1 enhancer (positive control), 950 ng  $\beta$ Gal, *Mix 3*: 1,250 ng TK-Luc PRMT8 enhancer (region of interest), 950 ng  $\beta$ Gal.  
NOTE: ES cells express the desired transcription factors (RAR/RXR). Leave wells untransfected to determine base values later during the luciferase and  $\beta$ -galactosidase measurements.
4. Add transfection quality reduced serum media to each plasmid mix up to 106  $\mu$ l total volume (enough for 8 wells).
5. Add 4.5  $\mu$ l ES quality transfection reagent then mix carefully by pipetting 15 times up and down. Incubate the transfection mix for 10 min at room temperature.
6. Add 13  $\mu$ l of the transfection mix to the cells per well, mix thoroughly by pipetting. Place the cells back into the incubator (37 °C) overnight before ligand treatment.
7. Remove media carefully by aspiration from cells and add 250  $\mu$ l fresh media/well containing 0.01 - 1  $\mu$ M all-trans retinoic acid (RA) as final concentration or DMSO as vehicle. Incubate cells for 24 - 48 hr.  
NOTE: Retinoic acid is light sensitive.

#### 6. Preparation of Cell Lysates

1. Dilute 5x Lysis Buffer (1.25 ml 0.5 M Tris pH = 7.8, 1 ml 1 M dithiothreitol (DTT) (in H<sub>2</sub>O), 10 ml 0.1 M EDTA pH = 8.0, 50 ml glycerol, 5 ml Triton X-100, sterile water up to 100 ml) to 1x using sterile water, add 20  $\mu$ l 1 M DTT/10 ml and equilibrate to room temperature before use. Prepare 10 ml for a 48-well plate on the day of the assay.
2. Remove media carefully by aspiration from cells. Rinse cells once with 1x PBS. Add 200  $\mu$ l 1x Lysis Buffer/well directly to the cells.
3. Shake for 2 hr at room temperature (chemical lysis). Freeze the cell lysates on the plate at -80 °C (mechanical lysis). Lysates can be stored at -80 °C for several days.
4. Thaw the lysates. After complete thawing, transfer cell lysates to a 96-well plate using an electronic multichannel pipette. Transfer 80  $\mu$ l of the cell lysate to 96-well clear plate for  $\beta$ -galactosidase assay and 40  $\mu$ l of the cell lysate to a white plate for luciferase measurement. Avoid forming any bubbles.

#### 7. $\beta$ -galactosidase Assay

NOTE: Beta galactosidase or alternative second reporters should be used as an internal control to account for non-specific experimental variations.

1. Prepare  $\beta$ -galactosidase substrate buffer: 80.0 g Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O, 3.8 g NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O, 30.0 g KCl, 1.0 g MgSO<sub>4</sub>·7H<sub>2</sub>O, make up to 1,000 ml with sterile water. Adjust pH to 7.4. Filter Sterilize (0.2 micron) and store at 4 °C
2. Mix 1 ml  $\beta$ -galactosidase substrate buffer with 4 mg ONPG (o-nitrophenyl- $\beta$ -D-galactosidase). Add 3.5  $\mu$ l of 2-Mercaptoethanol ( $\beta$ ME) per 1 ml of buffer just before use. Add 100  $\mu$ l  $\beta$ -galactosidase substrate buffer to 80  $\mu$ l cell lysate.
3. Incubate the reactions at 37 °C until a faint yellow color has developed. Read absorbance at OD 420 nm on a plate reader and export data.

NOTE: The time varies depending on the transfection efficiency, usually no longer than 30 min.

#### 8. Luciferase Assay

1. Prepare 50 ml luciferin substrate reagent: 15.1 mg D-luciferin salt, 82.7 mg ATP salt, 185 mg MgSO<sub>4</sub>·7H<sub>2</sub>O, add to 50 ml polypropylene tube, then add 1.5 ml 1 M HEPES buffer, 48.5 ml ATP free water, vortex, make sure all compounds dissolve completely. Keep 5 - 10 ml aliquots at -80 °C.
2. Warm 5 ml luciferin substrate reagent to room temperature before starting. Pipette 100  $\mu$ l substrate reagent to each well of the white plate containing the 40  $\mu$ l of the cell lysate and measure the signal immediately using a luminescent counter machine.
3. Obtain activities from at least three independent experiments in triplicate transfections.
4. Calculate first the base normalized luciferase activity and base normalized  $\beta$ -galactosidase values for each well by subtracting the average values measured for non-transfected cells from each value measured. Then calculate the ratio of luciferase/ $\beta$ -galactosidase activity.

### 3. Characterization of Enhancer RNA

NOTE: A more direct indicator of enhancer activity has emerged from recent genome-wide studies that identified many short non-coding RNAs, ranging in size from 50 to 2,000 nt, which are transcribed from enhancers, and are termed enhancer RNAs (eRNAs)<sup>16,25,26</sup> (Figure 4). eRNA induction highly correlates with the induction of adjacent exon-coding genes. Thus, signal-dependent enhancer activity can be quantified *in vivo* by comparing eRNA production between various conditions by RT-qPCR.

#### 1. Guidelines for Primer Design for eRNA detection by RT-qPCR

1. Select genomic regions for eRNA measurements that are at least 1.5 - 2 kb away from annotated transcription start sites.  
NOTE: Importantly, high levels of mRNA transcription across intragenic enhancers prevent accurate quantification of eRNA in the sense orientation, antisense eRNAs at intragenic enhancers are detectable and are similar in level to eRNAs at extragenic enhancers.
2. As a general rule, use regions 200 - 1,000 bp from the center of the transcription factor binding site of interest for primer design either on the sense or anti-sense strand (Figure 4 and Figure 6).  
NOTE: If GRO-seq, DNase-seq, H3K4me1, H3K4me3 or H3K27ac ChIP-seq data are available from the desired cell types and conditions it can be used for more accurate primer design. eRNAs are transcribed from putative enhancer regions characterized by high levels of H3K4me1 and me3. Expression of eRNAs positively correlates with the enrichment of activated enhancer histone mark H3K27ac.
3. Design PCR primers for fluorescent dye-based RT-qPCR measurement of eRNAs by Primer3 (<http://bioinfo.ut.ee/primer3/>)<sup>23</sup> Generate anti-sense sequence (reverse complement of the sense strand) for primer design as per: [http://www.bioinformatics.org/sms/rev\\_comp.html](http://www.bioinformatics.org/sms/rev_comp.html). Insert 200 - 300 bp of sense or anti-sense sequence for primer design.

#### 2. Measurement of eRNA Transcription

1. Isolate total RNA from treated cells with commercial acidic phenol/chloroform-based extraction reagent according to the manufacturer's recommendations.
2. Use DNaseI to treat isolated RNA prior to using them in reverse transcription reaction. Inactivate DNase according to the manufacturer's recommendation prior to reverse transcription.
3. Measure RNA concentration after DNase I treatment and use 1,000 ng per RT reaction. Use high quality reverse transcriptase.  
NOTE: As large number of eRNAs may not be polyadenylated, reverse transcription requires the use of random primers.
4. Detect reverse transcribed eRNA by RT-qPCR using standard procedure. Measure a non-treatment-dependent mRNA for normalization (e.g. 36B4, Ppia, Gapdh, Actb).

### Representative Results

We used a pan-specific RXR antibody in order to identify genome-wide which RA-regulated genes have receptor enrichment in their close proximity. Bioinformatics analysis of RXR ChIP-seq data obtained from ES cells treated with retinoic acid revealed the enrichment of the nuclear receptor half site (AGGTCA) under the RXR occupied sites (Figure 1). Using a bioinformatics algorithm we mapped back the motif search result for the half site to the RXR ChIP-seq data (Figure 2). This analysis helped us to accurately identify those ChIP peaks which were overlapping with canonical nuclear receptor binding sites. Visualization of these sites in IGV indicated enrichment of these transcription factors in the close proximity of *Hoxa1*, a previously characterized RAR/RXR target (Figure 2 and Figure 3). We also identified a novel RA target gene, namely *PRMT8*<sup>27</sup>. This latter region contains a direct repeat with no spacer nucleotides between the two half sites (AGGTCAAGGTCA)(Figure 3). To functionally validate that the element identified for *Hoxa1* can indeed bind RAR/RXR and regulated by RA we constructed TK-luciferase reporter vectors that contained ~300bp genomic region, including the identified elements. We also constructed vectors without the response element (Figure 5). ES cells were transfected and luciferase activity was measured in the absence and presence of retinoic acid (Figure 5). Constructs containing the retinoic acid response element (RARE) did show increased luciferase signal intensity upon RA treatment, while the construct without the RARE was not inducible.

We also studied the retinoic acid-dependent activity of the *Hoxa1* enhancer. To ascertain if the sense and anti-sense eRNA expression of the *Hoxa1* RAR/RXR-bound enhancer correlate with the mRNA expression of the gene, we also measured the level of *Hoxa1* mRNA (Figure 6). These results confirmed that the enhancer activity is induced by short-term RA treatment (3 hr), thus confirming that the enhancer is likely involved in RA-dependent enhancer regulation.

### Homer *de novo* Motif Results

[Known Motif Enrichment Results](#)  
[Gene Ontology Enrichment Results](#)

If Homer is having trouble matching a motif to a known motif, try copy/pasting the matrix file into [STAMP](#)

More information on motif finding results: [HOMER | Description of Results | Tips](#)

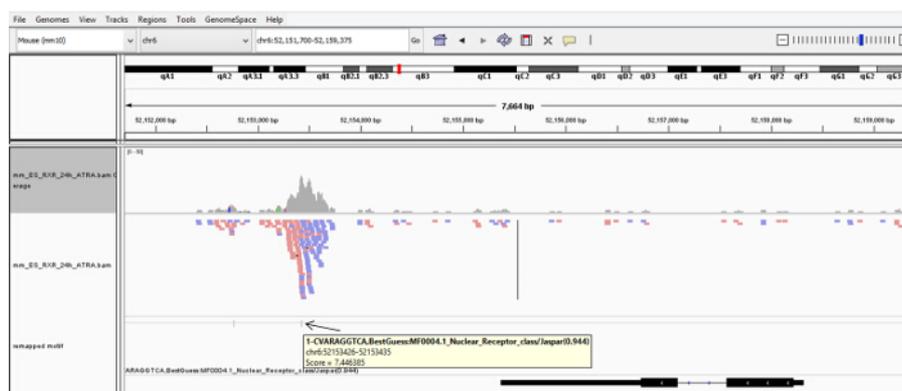
Total target sequences = 7463

Total background sequences = 42345

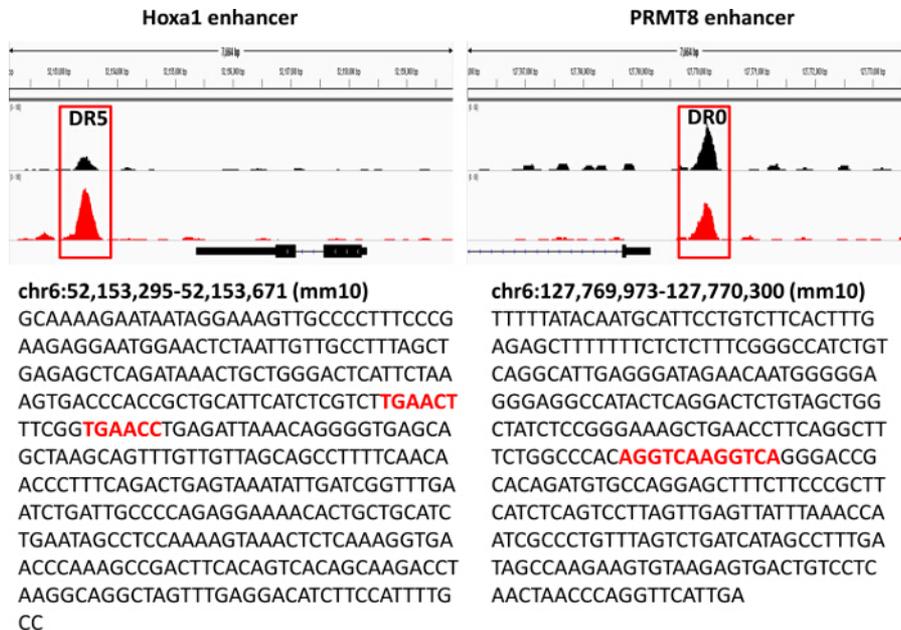
\* - possible false positive

Rank	Motif	P-value	log P-value	% of Targets	% of Background	STD/Bg (STD)	Best Match/Details	Motif File
1		1e-1958	-4.532e+03	77.66%	25.11%	61.1bp (94.6bp)	MF0004.1_Nuclear_Receptor_class/Jasper <a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>	<a href="#">motif file</a> <a href="#">/matr</a>
2		1e-774	-1.782e+03	13.57%	0.99%	70.3bp (92.4bp)	RARg(NR)/ES-RARg-ChIP-Seq(GSE30538)/Homer <a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>	<a href="#">motif file</a> <a href="#">/matr</a>
3		1e-394	-9.078e+02	19.04%	5.14%	64.2bp (97.1bp)	RARg(NR)/ES-RARg-ChIP-Seq(GSE30538)/Homer <a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>	<a href="#">motif file</a> <a href="#">/matr</a>
4		1e-220	-5.295e+02	12.33%	3.53%	73.6bp (93.7bp)	OCT4-SOX2-TCF-NANOG(POU/Homeobox/HMG)mES-ChIP-Seq/Homer <a href="#">More Information</a>   <a href="#">Similar Motifs Found</a>	<a href="#">motif file</a> <a href="#">/matr</a>

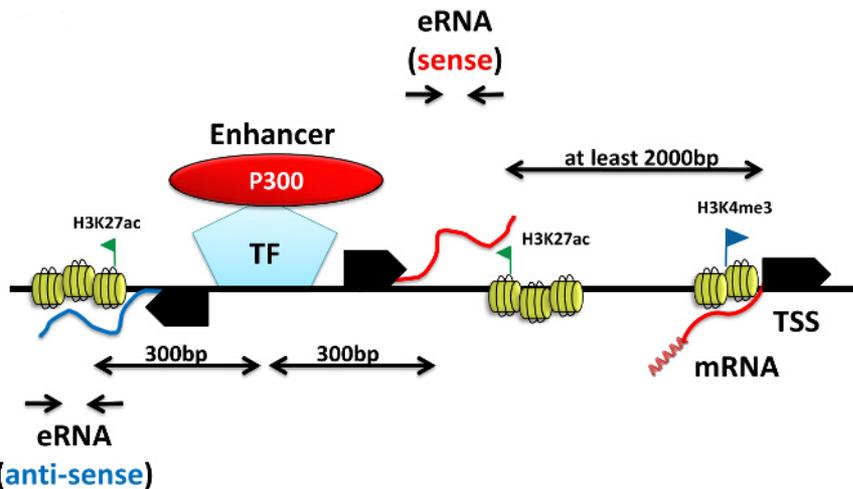
**Figure 1: Homer *De Novo* Motif Results.** An output HTML (homerResults.html) generated by Homer for the example RXR ChIP-seq is shown. Sequence logos corresponding to top enriched transcription factor motifs identified by *de novo* motif discovery at RXR-bound loci. 7463 RXR occupied genomic regions were used for the motif search. 77.66% of these regions contains AGGTCA-like motif (ranked as #1). For detailed description visit: (<http://homer.salk.edu/homer/motif/>) [Please click here to view a larger version of this figure.](#)



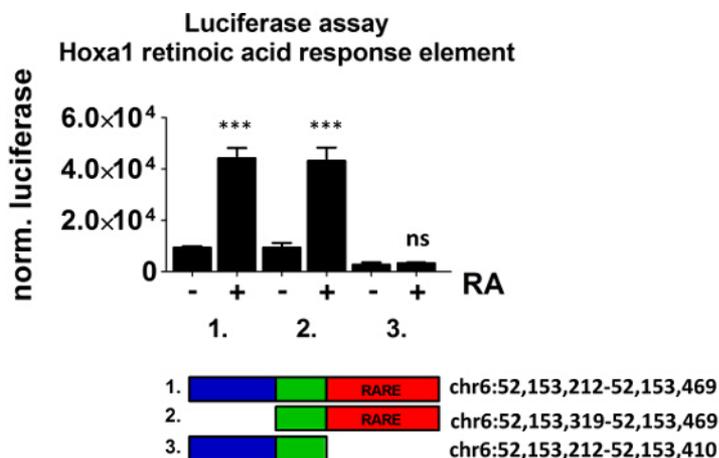
**Figure 2: Finding Genomic Regions Containing Motif of Interest.** Visualization of the aligned RXR ChIP-seq reads (mm\_ES\_RXR\_24h\_ATRA.bam and the AGGTCA motif occurrences (mm\_ES\_RXR\_24h\_ATRA\_homerpeaks\_motif1\_mm10s\_200\_remaped\_mbed.bed) by Integrative Genomics Viewer (IGV). Alignments are colored by read strand (reads on + strand: red, - strand: blue). Genome used for the alignment: mm10. [Please click here to view a larger version of this figure.](#)



**Figure 3: Genomic Loci of Hoxa1 and PRMT8 Enhancers.** IGV snapshot of RXR ChIP-Seq signals obtained from untreated and RA-treated (24 hr, 1  $\mu$ M) embryonic stem cells<sup>27</sup> are shown. Identified binding sequences are colored red. [Please click here to view a larger version of this figure.](#)

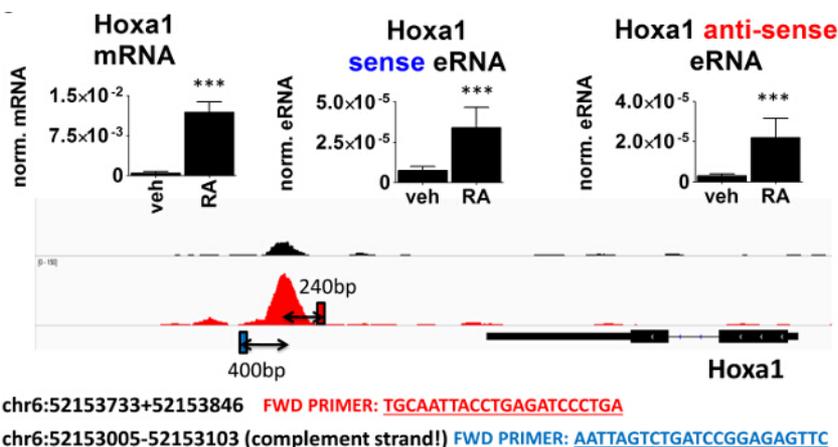


**Figure 4: Experimental Design for Testing eRNA Coding Sequences.** Enhancers chosen for enhancer RNA (eRNA) measurement should be located at least 1.5 - 2 kb away relative to the transcription start site (TSS) of the closest gene. ChIP-seq data of the transcription factor (TF) of interest and the motif analysis can be used to identify direct binding sites genome-wide. As binding of transcriptional coactivator (e.g. P300) often marks distal enhancers, regions enriched for both the TF and P300 are good enhancer candidates. Expression of eRNAs is positively correlate with the enrichment of activated enhancer histone mark H3K27ac. Regions 200 - 1,000 bp away in sense or anti-sense direction from the center of the transcription factor binding of is recommended for eRNA primer design. [Please click here to view a larger version of this figure.](#)



CAGCTGGAAGCGTCTCTCTTCCCAACAAAAAGAACTAAGAGGGAAGAGGAAAGTGTTTAAGTTTCTA  
GTAAACATTGTTTAGCAAAGAATAATAGGAAAGTTGCCCTTCCCGAAGAGGAATGGAACCTAATTG  
TTGCCCTTAGCTGAGAGCTCAGATAAACTGCTGGGACTCATTCTAAAGTGACCCACCGTGCATTATCTC  
GTCTTGAACCTTCGGTGAACCTGAGATTAACAGGGGTGAGCAGCTAA

**Figure 5: Luciferase Activity of the Indicated Reporter Constructs in ES Cells.** Indicated genomic regions of the Hoxa1 enhancer were cloned into TK-Luc vector and transfected into embryonic stem cells. Construct 1 and 2 contained the retinoic acid response element (RARE, underlined sequence). Cells were treated with RA (24 hr, 1 μM). βGal was used as an internal control for normalization. Bars represent mean normalized values from four biological replicates. ± s.d. \*\*\* P < 0.005 [Please click here to view a larger version of this figure.](#)



CAATTGGTCTGTGCCACAttttcacaagattagctctagtactcaacgagattcgttttctgagatgaaccgGAACCTCCGGATCAGACTAATTaacatttt  
taaaggattcagtgccacttatttacttgaacactgagtgatgggcagaaaaactcgaggattggcagaaaaactcgaggagcgcagtgaggagcagctggaaagcctct  
cttccccacaaaagaactaagagggagaggaaggttttaagtttctagtaaacatttttagcaaaaagataataggaaagtggcccttcccgaagaggaatggaactc  
aattgttccttagctgagagctcagataaactctgggactcattctaaagtaccaccctgctcattcatctctgttaacttccgtgaaacctgagattaaacaggggtgagca  
gctaagcagtttggtagcagcttttcaacaacccttcagactgagtaaatattgatcggttgactgattgccccagaggaaacactgctgcatctgaaatagctccaaaag  
taaaactcacaaggtgaacccaagccgacttcacagtcacagcaagacctaaggcaggctagttaggagacattccatttggcagcttctgatgatgaagatggagaaata  
aaaaataaaaacaaaacaaattttatATCAATTTCGAGGCCCAA

**Figure 6: Retinoic Acid Induced Enhancer RNA Transcription.** Hoxa1 mRNA and eRNA levels as measured by RT-qPCR. Total RNA was isolated from ES cells treated for 3 hr with retinoic acid (RA) or DMSO as vehicle (veh). Forward primers for the sense and anti-sense eRNA detection and the PCR amplicons are shown. Distance of regions detected by eRNA RT-qPCR from the center of the RXR binding site is indicated. Values were normalized to the average of 36B4 mRNA. Data represent mean ± s.e.m. \*\*\* P < 0.005 [Please click here to view a larger version of this figure.](#)

## Discussion

In recent years, advances in sequencing technology have allowed large-scale predictions of enhancers in many cell types and tissues<sup>7-9</sup>. The workflow described above allows one to perform primary characterization of candidate enhancers chosen based on ChIP-seq data. The detailed steps and notes will help anyone to set up a routine enhancer validation in the lab.

The most critical step in the luciferase reporter assay is the transfection efficiency. It is recommended to include a GFP-construct in order to estimate the transfection efficiency in every experiment. Having fresh and good quality of plasmids can be critical and can significantly improve the transfection. The presented protocol can be adapted to any cellular model system. It is highly recommended to use the appropriate cell type for the characterization of tissue-specific enhancers. Thus, luciferase reporter assay requires that the cells chosen for the experiment should be

efficiently transfectable. Alternatively, easily transfectable cells should be used where the transcription factor of interest is (over)expressed. This can be particularly important in case of some tissue-specific enhancers.

In contrast, enhancer RNA can be easily used in any cell without such limitations to study enhancer activity in different conditions *in vivo*. Key parameters for efficient eRNA detection are listed above. Isolation of good quality, non-degraded RNA is essential. DNase treatment is very important in order to avoid the detection of genomic DNA contamination instead of the eRNA. As RNA is used for subsequent reverse transcription reaction to get cDNA, DNase should be appropriately inactivated. It is advisable to design primers for the detection of both sense and anti-sense eRNAs.

Carrying out the above described enhancer trap experiment (transfection of a construct containing an enhancer element and a luciferase-based reporter) can provide strong functional evidence for enhancer activity. However, the ultimate proof, which is rarely presented in research articles, would be genetic elimination of the designated enhancer and thus determination of its role *in vivo* or at least in a cell<sup>28</sup>.

The above-described workflow does not give us information which gene is regulated by the validated enhancer. Further research could use 3C (Chromosome Conformation Capture) experiment that show whether the implicated enhancer is interacting with the promoter of the gene of interest.

## Disclosures

The authors have nothing to disclose.

## Acknowledgements

The authors would like to acknowledge Dr. Bence Daniel, Matt Peloquin, Dr. Endre Barta, Dr. Balint L Balint and members of the Nagy laboratory for discussions and comments on the manuscript. L.N is supported by grants from the Hungarian Scientific Research Fund (OTKA K100196 and K111941) and co-financed by the European Social Fund and the European Regional Development Fund and Hungarian Brain Research Program - Grant No. KTIA\_13\_NAP-A-I/9.

## References

1. Wamstad, J. A., Wang, X., Demuren, O. O., & Boyer, L. A. Distal enhancers: new insights into heart development and disease. *Trends in cell biology*. **24**, 294-302 (2014).
2. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. Enhancers: five essential questions. *Nat Rev Genet*. **14**, 288-295 (2013).
3. Mardis, E. R. ChIP-seq: welcome to the new frontier. *Nature methods*. **4**, 613-614 (2007).
4. Muller, F., & Tora, L. Chromatin and DNA sequences in defining promoters for transcription initiation. *Biochim Biophys Acta*. **1839**, 118-128 (2013).
5. Ounzain, S., & Pedrazzini, T. The promise of enhancer-associated long noncoding RNAs in cardiac regeneration. *Trends Cardiovasc Med*. (2015).
6. Lam, M. T., Li, W., Rosenfeld, M. G., & Glass, C. K. Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci*. **39**, 170-182 (2014).
7. Dickel, D. E. *et al.* Function-based identification of mammalian enhancers using site-specific integration. *Nature methods* **11**, 566-571 (2014).
8. Blum, R., Vethanatham, V., Bowman, C., Rudnicki, M., & Dynlacht, B. D. Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. *Genes Dev*. **26**, 2763-2779 (2012).
9. Vermunt, M. W. *et al.* Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell reports* **9**, 767-779 (2014).
10. Bain, G., Kitchens, D., Yao, M., Huettner, J. E., & Gottlieb, D. I. Embryonic stem cells express neuronal properties *in vitro*. *Dev Biol*. **168**, 342-357 (1995).
11. Mahony, S. *et al.* Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol* **12**, R2 (2011).
12. Simandi, Z., Balint, B. L., Poliska, S., Ruhl, R., & Nagy, L. Activation of retinoic acid receptor signaling coordinates lineage commitment of spontaneously differentiating mouse embryonic stem cells in embryoid bodies. *FEBS Lett*. **584**, 3123-3130 (2010).
13. Li, H., & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. **25**, 1754-1760 (2009).
14. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).
15. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-26 (2011).
16. Daniel, B. *et al.* The active enhancer network operated by liganded RXR supports angiogenic activity in macrophages. *Genes Dev* **28**, 1562-1577 (2013).
17. Zhu, Y. *et al.* Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res* **41**, 10032-10043 (2013).
18. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858 (2009).
19. Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-318 (2007).
20. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283 (2010).
21. Bonn, S. *et al.* Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* **44**, 148-156 (2012).
22. Hollenberg, S. M., Giguere, V., Segui, P., & Evans, R. M. Colocalization of DNA-binding and transcriptional activation functions in the human glucocorticoid receptor. *Cell*. **49**, 39-46 (1987).
23. Untergasser, A. *et al.* Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**, e115 (2012).
24. Rhead, B. *et al.* The UCSC Genome Browser database: update 2010. *Nucleic Acids Res* **38**, D613-619 (2010).

25. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187 (2010).
26. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390-394 (2011).
27. Simandi, Z. *et al.* PRMT1 and PRMT8 regulate retinoic acid-dependent neuronal differentiation with implications to neuropathology. *Stem Cells* **33**, 726-741 (2014).
28. Zhou, H. Y. *et al.* A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev* **28**, 2699-2711 (2014).