Gene: X 2 (2019) 100008

Contents lists available at ScienceDirect

Gene: X

journal homepage: www.journals.elsevier.com/gene-x

Development of a high efficient promoter finding method based on transient transfection

Yao Lu[a,1], Qilong Li[a,1], Kexin Zheng[b], Chenghao Fu[a], Chunying Jiang[a], Dayu Zhou[b], Chao Xia[a], Shiliang Ma[a,*]

[a] College of Bioscience and Biotechnology, Shenyang Agricultural University, Shenyang, Liaoning, PR China
[b] College of Food Science and Technology, Shenyang Agricultural University, Shenyang, Liaoning, PR China

ARTICLE INFO

Keywords:
Promoter trap
Mouse breast cancer
Cancer-specific promoter
Gene expression regulation
Bioinformatics
Gene finding

ABSTRACT

In metazoan genome, the mechanism of gene expression regulation between transcriptional regulatory elements and their target gene is spatiotemporal. Active promoters possess many specific chromosomal features, such as hypersensitive to DNaseI and enrichment of specific histone modifications. In this article, we proposed a novel method which possesses a high efficiency to find promoters *in vitro*. A promoter-trap library was constructed with totally 706 random mouse genomic DNA fragment clones, and 260 promoter-active fragments of the library were screened by transient transfection into 4T1 cells. To demonstrate the accuracy of this promoter finding method, 13 fragments with promoter activities were randomly selected for published DNase-seq and ChIP-seq data analysis, downstream transcripts prediction and expression confirmation. qRT-PCR results showed that six predicted transcription units were successfully amplified in different mouse tissues/cells or in reconstituted mouse mammary tumors. Our results indicate that this promoter finding method can successfully detect the promoter-active fragments and their downstream transcripts.

1. Introduction

In metazoan genome, the mechanism of gene expression regulation between transcriptional regulatory elements and their target genes is spatiotemporal (Long et al., 2016). As one of the *cis*-regulatory elements, promoters are considered to be the key to the initiation of transcription (Lenhard et al., 2012). Many studies have shown that under the different genomic context, promoters perform different transcriptional activities, which can influence the transcriptional efficiency and mRNA expression levels of their target genes at a certain extent (Stark, 2014; Arensbergen et al., 2014; Trinklein et al., 2004). Furthermore, as an important biological, functional and regulatory diversity element, promoter can both proximally and distally regulate its target gene. Currently, there are 82,853 candidate promoters have been found in the mouse genome and this number is still increasing (Yue et al., 2014).

Active promoters possess many specific chromosomal features, including DNaseI hypersensitive sites (DHS) and the enrichment of specific histone modifications, such as mono- and tri-methylation of histone 3 (H3) Lys^4 (H3K4), *etc*. (Valen and Sandelin, 2011; Li et al., 2007) Due to the different features they contain, promoters in metazoan genome can be detected by using different technologies with several biological perspectives, broadly dividing into RNA-based and epigenomic approaches. For the former, active promoters are investigated by technologies such as DNase I hypersensitive sites sequencing (DNase-seq), RNA sequencing (RNA-seq), cap analysis of gene expression (CAGE), and global run-on sequencing (GRO-seq) (Core et al., 2014; Consortium et al., 2014; Carninci et al., 2005). And the latter involves

Abbreviations: ATAC-seq, Assay for transposase-accessible chromatin using sequencing; CAGE, cap analysis of gene expression; ChIP-seq, Chromatin immunoprecipitation followed by massively parallel DNA sequencing; CMV, Cytomegalovirus; Ct, threshold; DHS, DNaseI hypersensitive sites; DNase-seq, DNase I hypersensitive sites sequencing; dNTPs, deoxy-ribonucleoside triphosphate; EF1a1, eukaryotic translation elongation factor 1 alpha 1; eRNAs, enhancer RNAs; FBS, fetal bovine serum; GRO-seq, global run-on sequencing; H3K4me3, histone H3 lysine 4 trimethylation; Itpr2, inositol 1, 4, 5-triphosphate receptor 2; LSINCT5, long stress-induced non-coding transcript 5; MCS, multiple cloning site; MPRA, Massively parallel reporter assays; mSEAP, mouse synthetic secreted embryonic alkaline phosphatase; PBS, phosphate buffered solution; pNPP, p-nitropheny-phosate; qRT-PCR, quantitative RT-PCR; RNA-seq, RNA sequencing; SD, standard deviation; STARR-seq, Self-transcribing active regulatory region sequencing; TFs, transcription factors; tpk1, thiamine pyrophosphokinase; TSS, transcription start sites

* Corresponding author.
E-mail address: msl@syau.edu.cn (S. Ma).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.gene.2019.100008
Received 30 August 2018; Received in revised form 23 January 2019; Accepted 8 February 2019
Available online 12 February 2019

in detecting characteristic chromatin modifications especially H3K4me3 positivity and H3K4me1 depletion (Wang et al., 2008; Barski et al., 2016; Creyghton et al., 2011; Radaiglesias et al., 2009).

In this article, we proposed a novel approach to find genome sequences function as promoters *in vitro*. The mouse promoter-trap library was successfully constructed with the insertion of the mouse breast cancer cell 4T1 randomly digested genomic DNA fragments into the promoter-less reporter vector. And 260 fragments with promoter activities were selected by detecting the mouse synthetic secreted embryonic alkaline phosphatase (mSEAP) (Kain, 1997) Furthermore, 13 of all fragments with promoter activities were randomly selected for sequencing and then mapped on mouse genome (mm9). The accuracy of our promoter finding approach was verified by analyzing the published data of DNase-seq and ChIP-seq. Furthermore, downstream transcription units of the 13 selected fragments were bioinformatically predicted and experimentally tested.

2. Materials and methods

2.1. Ethics statement

This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Research Council. The protocol was approved by the Committee on the Ethics of Animal Experiments of Shenyang Agricultural University (Permit Number: 2015-029). All surgeries were performed under sodium pentobarbital anesthesia, and every effort was made to minimize suffering.

2.2. Materials

BALB/c mice were obtained from the Changsheng Animal Resources Center (Benxi, Liaoning, China). Mouse breast cancer cell lines 4T1 and 4T07 were purchased from Cell Bank of Chinese Academy of Sciences (Shanghai, China). Fluorescent protein reporter gene vector pEGFP-N1 was stored in our laboratory. Transfection reagent Lipofectamine 2000 (cat. no. 11668027) was purchased from Invitrogen (Carlsbad, CA, USA). The mSEAP substrate, pNPP (p-nitropheny-phosate) (cat. no. N9389), was purchased from Sigma-Aldrich (St. Louis, MO, USA). GT115 competent cells (cat. no. gt115-11) and pCpGfree-basic (named as Pro-trap) vector were purchased from InvivoGen (San Diego, CA, USA); pMD™18-T Vector Cloning Kit (cat. no. 6011) was purchased from TaKaRa (Otsu, Japan). E.Z.N.A. ® Gel Extraction Kit (cat. no. D2500-02) and E.Z.N.A. ® Plasmid Mini Kit I (cat. no. D6943-02) were purchased from Omega Bio-tek (Norcross, GA, USA). Primer pairs were synthesized by Dingguo Biotechnology Company (Beijing, China). KAPA SYBR® FAST Universal 2× qPCR Master Mix (cat. no. 07959389001) was purchased from KAPA Biosystems (Boston, MA, USA). RevertAid First Strand cDNA Synthesis Kit (cat. no. K1622) was purchased from Thermo Scientific (Carlsbad, CA, USA).

2.3. Cell lines and cell culture

Mouse breast cancer cell line 4T1 and 4T07 were cultured in RPMI 1640 (Gibco, Grand Island, NY, USA, cat. no. 11875093), supplemented with 10% fetal bovine serum (FBS) (TBD, Tianjin, China, cat. no. TBD21HY) and 100 U/mL penicillin plus 100 μg/mL streptomycin. The cells were maintained in an incubator (Thermo Scientific) of 5% CO_2 atmosphere at 37 °C. Cells were passaged or harvested when cell confluence reached 90%.

2.4. Promoter-trap library construction

2.4.1. Genomic DNA extraction and fragmentation

When 4T1 cells grown to confluence of 90%, cells were washed in PBS (phosphate buffered solution) buffer and trypsinized, washed and re-suspended in PBS. Cell genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany, cat. no. 69506) according to manufacturer's instructions. RNase A was used for purification of DNA extraction. The genomic DNA was assessed by 1.0% agarose gel electrophoresis. For the generation of random genomic DNA fragments, the purified genomic DNA was double enzyme digested with *Hin*dIII (NEB, Ipswich, MA, USA, cat. no. R0104S) and *Bam*HI (NEB, cat. no. R0136S). Furthermore, the genomic DNA after digestion was purified by using E.Z.N.A. Gel Extraction Kit (Omega Bio-tek). Purified DNA fragments after double enzyme digestion were reserved for promoter-trap library construction.

2.4.2. Promoter-trap vector preparation

The backbone of our promoter-trap vector was from a promoter-less vector that contained the mSEAP gene as a reporter gene downstream the multiple cloning sites. Here this vector was used as the negative control vector designated as Pro-trap. For the construction of the positive control vector, the CMV (Cytomegalovirus) promoter (Boshart et al., 1985; Foecking and Hofstetter, 1986) was PCR amplified from pEGFP-N1 vector with the primers named as pCMV-F and pCMV-R (S1 Table). The PCR mixture contained 1 μL pEGFP-N1 template DNA, 2.5 μL 10× Taq Easy Buffer, 2 μL (2.5 mM) dNTPs (deoxy-ribonucleoside triphosphate), 0.5 μL (5 U/μL) Taq polymerase, 1 μL (10 mM) forward primer, 1 μL (10 mM) reverse primer and 17 μL H_2O. PCR conditions were as follows: 94 °C pre-denaturation 4 min, followed by 94 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min, 40 cycles; the final extension step was at 72 °C for 10 min. PCR products were visualized on a 1.0% agarose gel and then purified by E.Z.N.A. Gel Extraction Kit (Omega Bio-tek). The target fragments were then ligated into the pMD 18-T Vector (TaKaRa). After sequencing confirmation, the positive control vector was constructed by inserting the amplified CMV promoter from the pEGFP-N1 into the *Bam*HI-*Hin*dIII sites of Pro-trap vector, designated as Pro-CMV-trap.

2.4.3. Promoter-trap library construction

Both the Pro-trap and the randomly digested genomic DNA fragments from 4T1 cells were purified after restriction enzyme digestion with *Bam*HI/*Hin*dIII. The purified restriction enzyme digested big fragment from Pro-trap and the purified restriction enzyme digested random genomic DNA fragments were ligated and named as Pro-DNA-trap. The Pro-DNA-trap was transformed into *E. coli* GT115 cells. Positive clones were selected from LB plated with Zeocin, and subsequently inoculated into Zeocin containing liquid LB and cultured at 37 °C with shaking at 180 rpm for 12 h. The plasmids were purified with E.Z.N.A. ® Plasmid Mini Kit I (Omega Bio-tek) and quantified on a 1.0% agarose gel. Each clone obtained in this way was named as P-4T1-number (the number corresponding to its clone number), and the entire collection of clones constituted the promoter library. All plasmids were stored at −20 °C.

2.5. 4T1 cells transfection

4T1 cells were maintained at the same conditions as described above. For transfection with the obtained promoter-trap library clones above, 4T1 cells were plated into a 24-well cell culture plate. 5×10^5 4T1 cells were dispersed uniformly in each well. The promoter positive control vector Pro-CMV-trap was mixed with liposome Lipofectamine 2000 (Invitrogen) to identify the optimal transfection time following the manufacturer's instructions. Cell culture supernatants were collected from each transfected wells at time points of 6 h, 12 h, 24 h, 36 h and 48 h. Then, mSEAP activity analysis followed the protocols as described (Berger et al., 1988) and the absorbance was measured at 405 nm with a plate reader.

After determining the transfection time with the highest mSEAP activity, all vectors from the promoter-trap library plasmids as well as the positive control vector Pro-CMV-trap or the negative control vector

Pro-trap were transiently transfected into 4T1 cells as described above. Cell culture supernatants were collected from each transfected wells at the optimal transfection time. The mSEAP activity was assayed as previously. The pRL-CMV vector (Promega, Madison, WI, USA, cat. no. E2261) was included to normalize transfection efficiency in each transfection experiment. Cells were lysed by using the Steady-Glo® Reagent (Promega), following the protocol recommendation. The luciferase activity was measured as described in Standard Protocol of the Steady-Glo® Luciferase Assay System (Promega, cat. no. E2510) by using the luminometer (Wallac 1420 Victor 2, EG&G Wallac).

2.6. Screening active promoters obtained from promoter-trap library

After measured and calculated the activities of the reporter gene mSEAP, positive clones with promoter activities were selected by comparing the relative reporter gene mSEAP activities with the negative control vector Pro-trap. Thirteen fragments with promoter activity were randomly selected for bidirectional sequence identification by Genewiz Biotechnology Co. Ltd (Suzhou, Jiangsu, China).

2.7. Bioinformatics analysis in promoter identification and transcripts prediction

The published DNase-seq and ChIP-seq data from ENCODE database (S5 Table) were analyzed and visualized with the WashU EpiGenome Browser V46.1 (http://epigenomegateway.wustl.edu/browser/) (Brozovic et al., 2017). The mouse genomic resource database UCSC (University of California Santa Cruz of BLAST-LiFor) was used to determine the localization of the genomic position of the selected fragments. The 100,000 base pair sequences downstream each corresponding promoter fragment were chosen from UCSC. In addition, the FGENESH (Solovyev et al., 2006) (www.softberry.com) program was used to predict transcription units of these 100,000 bp sequences and further validated with NCBI blast program (Sayers et al., 2010).

2.8. Predicted transcription units analysis in mouse breast cancer cell line 4T1 by RT–PCR

4T1 cells were maintained at the same conditions as described above. Total RNA was extracted using TRIzol™ Reagent (Invitrogen, cat. no. 15596026) and treated with RNase-free DNase I (Invitrogen, cat. no. AM1907) according to the manufacturer's instructions. mRNAs were reverse transcribed followed the protocols of the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) using oligo$(dT)_{18}$ primer and random hexamers primer as first strand primers, respectively. The obtained cDNA was treated with DNase-free RNase A. Then, cDNA was used as template in subsequent PCR reactions. PCR was carried out under the following conditions: 94 °C pre-denaturation for 3 min, followed by 94 °C for 30 s, 60 °C for 30 s and 72 °C for 1 min, 40 cycles; the final extension step was at 72 °C for 10 min. All RT-PCR primers were listed in S1 Table.

2.9. Predicted transcription units analysis in 4T1 homografted BALB/c mice by qRT–PCR

2.9.1. Mouse homograft model

6–8 week old female BALB/c mice were maintained at the University of Animal Facility under specific pathogen-free conditions. Mice were randomly divided into different experimental groups, *i.e.* tumor-free group (n = 5) and transplant tumor group (n = 5). For transplantations, a total of 2×10^5 cells were resuspended in 100 µL PBS and injected subcutaneously into the right 4th mammary gland fat pad of the mouse using a 30-gauge needle. Mice were palpated for tumor formation. Tumor growth was monitored every 3 days by measuring the tumor width (W) and length (L) with a caliper. Tumor volume was estimated according to the formula Volume (mm^3) = $L \times W^2 \times 0.4$. Mice were sacrificed at 42 days after transplantation. Tumors and other tissues (livers, lungs and spleens) were then excised and weighed for RNA extracting.

2.9.2. qRT–PCR analysis

Total RNAs from BALB/c mice tumors, tissues (livers, lungs and spleens) and cells (4T1 and 4T07) were extracted by using TRIzol Reagent (Invitrogen) as described above.

For qRT-PCR (quantitative RT-PCR) analysis (Bustin et al., 2009), the total RNAs obtained above were first treated with 10 U DNase I (Invitrogen) to eliminate potential genomic DNA contamination. The first-strand cDNA was synthesized respectively from these total RNAs followed the protocols of the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) by using oligo $(dT)_{18}$ primer as first strand primer. Mouse eukaryotic translation elongation factor 1 alpha 1(EF1a1) gene (accession No. NC_000075.6) was used as an internal control. qRT-PCR was performed using specific primers (S1 Table) to amplify products between 250 and 400 bp, respectively. qRT-PCR amplification mixtures (20 µL) contained 100 ng template cDNA, 0.5 µM of each primer and KAPA SYBR FAST qPCR master mix buffer (10 µL) (KAPA Biosystems, USA). Reactions were run on the LightCycler480 II (Roche, Switzerland). The cycling conditions consisted of 4 min polymerase activation at 94 °C and 40 cycles at 94 °C for 30 s, 58 °C for 30 s and 72 °C for 30 s. Experiments were performed on three biological replicates for samples from each tissue or cell. Melting curves were analyzed to confirm the specificity of the reactions. Ct (threshold cycles) values were calculated by the $2^{-\Delta\Delta Ct}$ method. Ct values from three technical replicates were averaged and normalized with the Ct values of the internal control EF1a1, and then the standard deviations and errors were calculated.

2.10. Statistical analysis

Student's *t*-test was used to identify significantly different genes between paired health and metastatic samples. Values were considered statistically significant at $P < 0.05$.

3. Results

3.1. Promoter-trap library construction

The process of the promoter-trap library construction was shown as Fig. 1. For the construction of promoter-trap library, four vectors were prepared and shown as Fig. 2. The promoter-trap library was successfully constructed by inserting randomly digested 4T1 cells genomic DNA into the vector Pro-trap. Each clone from the promoter-trap library above was confirmed by using agarose gel electrophoresis, forming a library that was composed of 706 clones.

3.2. Promoter activities determination

For identifying the optimal transfection time, the highest mSEAP activity was observed after transfection of Pro-CMV-trap into 4T1 cells. The result indicated that 36 h was optimal transfection time with the highest activity of mSEAP (Fig. 3A). Then, each single clone from the promoter-trap library, named as P-4T1-number (the number corresponding to the clone number), was transiently transfected 4T1 cells for determining the fragments with promoters activities. The results of each clone that tested by mSEAP were shown in S2 Fig. Promoter activities were positively correlated with the protein expression of mSEAP in breast cancer cell line 4T1. As a result, 66.14% fragments possessed promoter activities, which between the negative control (Pro-trap) and the positive control (Pro-CMV-trap), 3.60% fragments can increase the expression of mSEAP compared with the positive control, which was determined as highly active promoters (Fig. 3B). And 30.26% promoters were inferior or equal to the negative control (Fig. 3C).

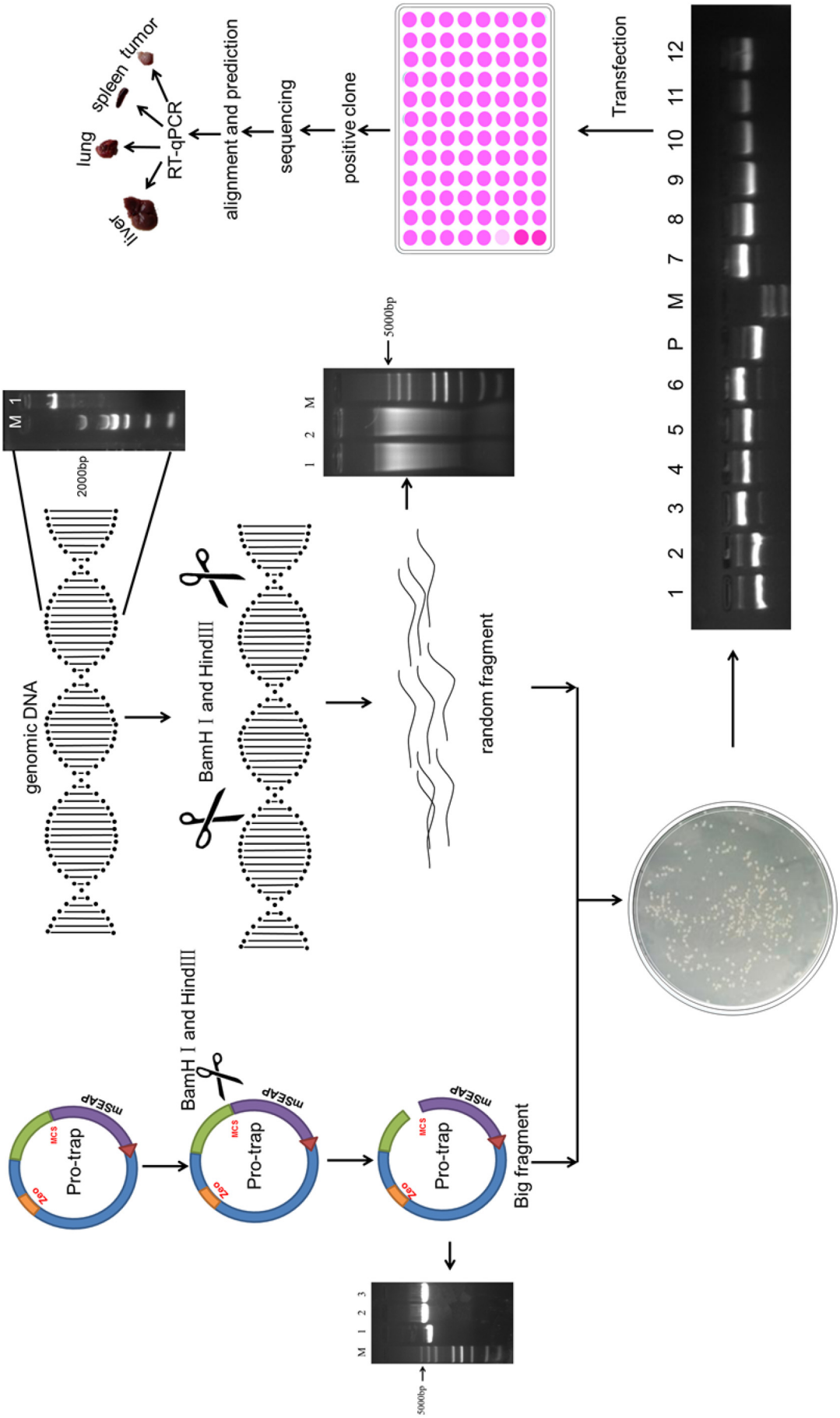

Fig. 1. Flow chart of gene finding with promoter-trap technology. The main components of the promoter-trap vector (Pro-trap) include a bacteria selection marker Zeo, a eukaryotic reporter gene mSEAP and a multiple cloning site (MCS) directly upstream of the reporter gene. Both the Pro-trap and the 4T1 cells genomic DNA were restriction enzyme digested with *Bam*HI/*Hind*III, and the purified enzyme digested big fragment from the pro-trap and the random fragments from the mouse 4T1 genome were ligated and transformed into *E. coli* GT115 cells. After transformation, each single clone was picked up and amplified. Recombinant plasmids were gel purified and quantified before reporter activity assay with a 96-well plate. Positive clones were determined after subtraction of reporter activities from the negative control. Furthermore, downstream transcription units of the 13 selected fragments were bioinformatically predicted and amplified by qRT-PCR in different tissues/cells.

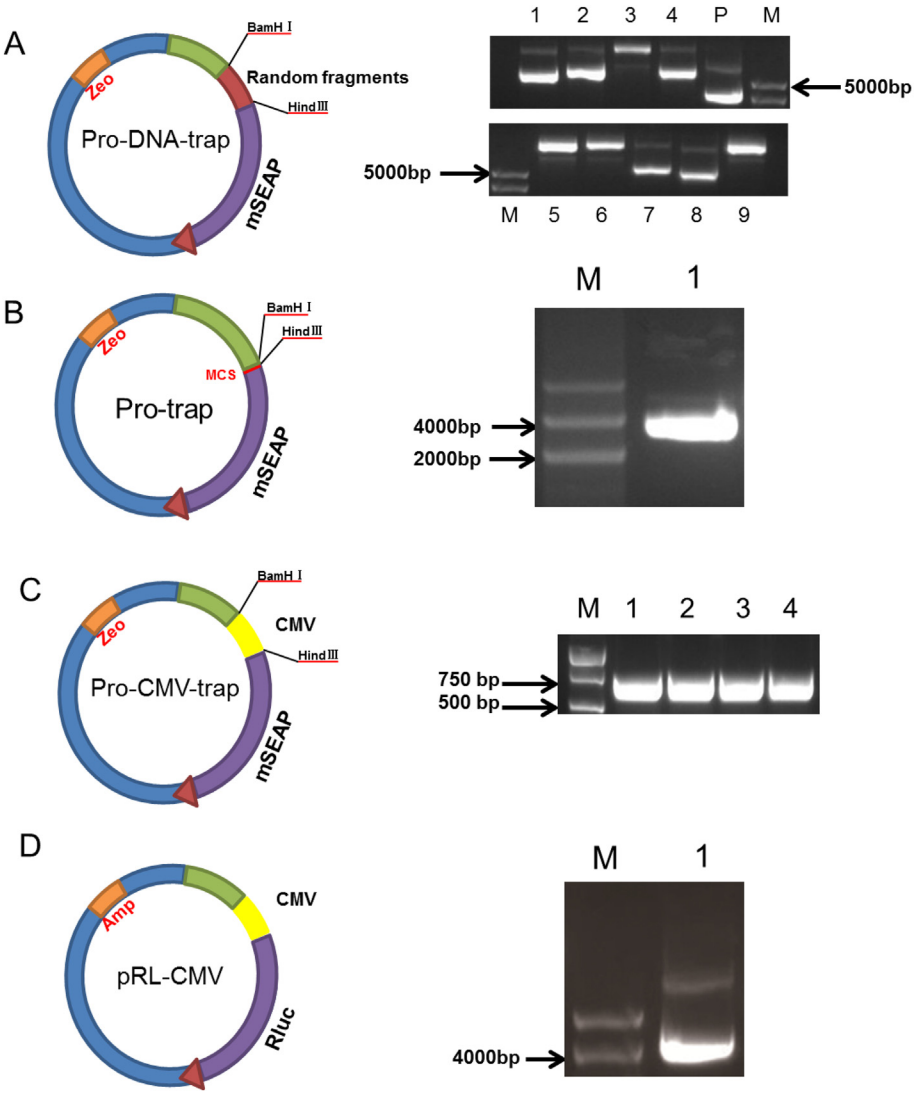

Fig. 2. Diagram of the construction of promoter trap library. (A) Pro-DNA-trap was a reporter vector driven by random 4T1 cells genomic DNA fragments. It was used for insertion of the random 4T1 cells genomic DNA fragments obtained by *Hin*dIII and *Bam*HI double restriction enzyme digestion. Agarose gel map demonstrated the varieties of the inserted random DNA fragments. Lane 1–9: Recombinant vector containing trapped fragments with different base pairs. P: Pro-trap vector. M: Marker. (B) The Pro-trap was a promoter-less vector used as the negative control, and the agarose gel map showed the Pro-trap vector. Lane 1: Double digested Pro-trap. M: Marker. (C) The Pro-CMV-trap was a vector with the insertion of a CMV promoter into the *Bam*HI/*Hin*dIII sites as the positive control, and the agarose gel map showed the fragment size of the CMV promoter. Lane 1–4: the fragments of the CMV promoter. M: Marker. (D) The pRL-CMV was a vector with a CMV promoter for normalizing transfection efficiency. Lane 1: the pRL-CMV. M: Marker.

3.3. Promoter localization confirmation

Thirteen positive clones with different promoter activities were randomly selected from the promoter-trap library and the promoter activities of each clone were shown in Fig. 4A. Twelve of these 13 clones were highly active promoters, and the other one was active promoter. These 13 clones above were sequenced and then mapped to mouse genome mm9. The results showed that the lengths of the 13 trapped fragments ranged from 294 bp to 3784 bp (S3 Seq), which the localizations were randomly distributed in different mouse chromosomes (Fig. 4B). In general, 7 of 13 (53.8%) trapped fragments with promoter activities were located in nongene regions, which mainly consist of repetitive elements. Five of 13 (38.4%) trapped fragments with promoter activities were located in gene intron regions, and 1 of 13(7.6%) trapped fragments with promoter activities was located in gene exon region (Fig. 4C and S4 Table).

3.4. The accuracy verification by analyzing the published data in mouse normal cell lines and tissues

To further analyze the degree of open chromatin of the 13 randomly selected trapped fragments, we extracted 6 DNase-seq datasets from ENCODE database (S5 Table) to identify the DNase I hypersensitivity. By analyzing the published DNase-seq data, the results showed that the trapped fragments displayed various degrees of chromatin accessibility in different cell lines (B cells, ES-E14 and Erythroblast) and tissues (Liver, Lung and Spleen). The trapped fragment, P-4T1-273, displayed a structure with obvious chromatin accessibility in erythroblasts and livers. The trapped fragment, P-4T1-629, also displayed an accessible chromosome structure in erythroblasts. The trapped fragment, P-4T1-28, displayed a heterochromatin state in erythroblasts, lungs and spleens. The trapped fragment, P-4T1-146, displayed a heterochromatin state in B cells, ES-E14 cells, erythroblasts and lungs. The trapped fragment, P-4T1-516, displayed a heterochromatin state in all selected cell lines and tissues (Fig. 5A and S6 Fig).

Comparing with the published models of chromatin state from ENCODE project in different cell lines and tissues, the ratio of heterochromatin and euchromatin in trapped promoter regions was also different. In B cells, ES-E14 cells and spleens, 11 of 13 fragments (84.61%) displayed euchromatin state. In erythroblasts and lungs, 10 of 13 fragments (76.92%) displayed euchromatin state. In livers, 12 of 13 fragments (92.30%) displayed euchromatin state (Fig. 5B). This indicated that the fragments with promoter activities in 4T1 cells trapped by our method also displayed as accessible chromatins in all selected cell lines and tissues.

To further verify the accessibility of our approach to trap promoters, we analyzed the H3K4me3 modification in trapped fragments from published ENCODE database (S5 Table). The results showed that P-4T1-328 was strongly modified with H3K4me3 in megakaryocytes, while obviously decreased in erythroblasts. This indicated that P-4T1-328

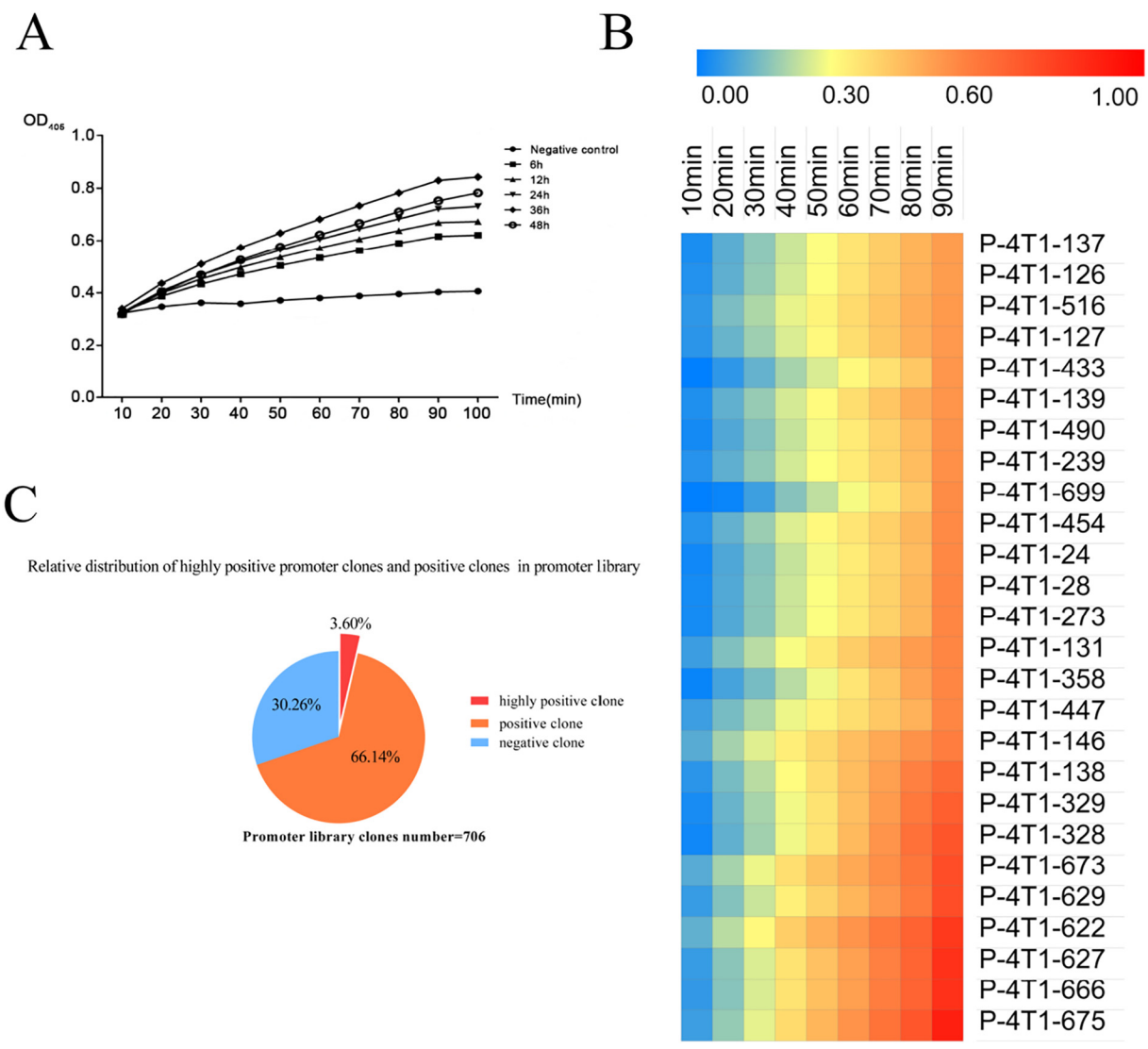

Fig. 3. Screening of inserted fragments with reporter gene activity. (A) After transfection of 4T1 cells with either negative control (pro-trap) or positive control (CMV-trap), the optimal transfection time point was shown as 36 h. The time course of the mSEAP activities was indicated. The figure was plotted by the mean value of the three replicates. ● stands for the negative control, ■ stands for the cell culture supernatants collected at 6 h, ▲ stands for the cell culture supernatants collected at 12 h, ▼ stands for the cell culture supernatants collected at 24 h, ◆ stands for the cell culture supernatants collected at 36 h, ⊖ stands for the cell culture supernatants collected at 48 h. (B) The heat map of SEAP activities of positive clones that possess promoter activities higher than positive control corresponding to its 96-well plate. The figure was displayed by the mean value of the three replicates. (C) The relative distribution of 706 clones with different activities. Negative clones represent the promoter activities lower than that of the negative control; positive clones represent the promoter activities between that of the positive and negative control; highly positive clones represent the promoter activities higher than that of the positive control. Each clone was compared with negative and positive control on its corresponding 96-well plate.

may play a key role during erythro-megakaryocytic development. P-4T1-516 was modified with H3K4me3 in all cell lines except Mel cells. Although P-4T1-28 displayed a heterochromatin state in erythroblasts, it was obviously modified with H3K4me3. In this study, the regions of all 13 selected trapped fragments were obviously modified with H3K4me3 in megakaryocytes. The results of the H3K4me3 modification of these 13 fragments in different cell lines also verified the accessibility of our approach. The peak scores and signals of H3K4me3 from ChIP-seq data were respectively shown in Fig. 5C and S7 Fig.

3.5. Downstream putative transcripts prediction and predicted transcription units confirmation by RT-PCR

The results of the predicted transcription units that located in 100,000 bp downstream of 13 trapped promoters were shown in S8 and S9 Figs. These predicted transcription units were then amplified used Gene-specific primers (S1 Table). Predicted transcription units were named as G-4T1-number (the number corresponding to its clone number). The results of RT-PCR showed that six predicted transcription units were successfully amplified, including *G-4T1-24*, *G-4T1-328*, *G-4T1-329*, *G-4T1-447*, *G-4T1-490* and *G-4T1-516*, while the others were failed to amplify by using oligo $(dT)_{18}$-primed cDNA in 4T1 cell line. Then the fragments that failed to amplify before were subsequently amplified by using random hexamer primers. *G-4T1-28* was successfully amplified by using random hexamer primers. *G-4T1-28* and *G-4T1-544* were located in the same strand with its known genes, respectively, and the rest fragments were located in different strands of its target genes. *G-4T1-28* was located in the second intron of *Mus musculus inositol 1, 4, 5-triphosphate receptor 2* (*Itpr2*), while the exon and splice of its predicted transcription unit are different from the known gene. However, only the fragment of the known gene was successfully amplified in 4T1 cells cDNA by RT-PCR, and failed to amplify the predicted exons and

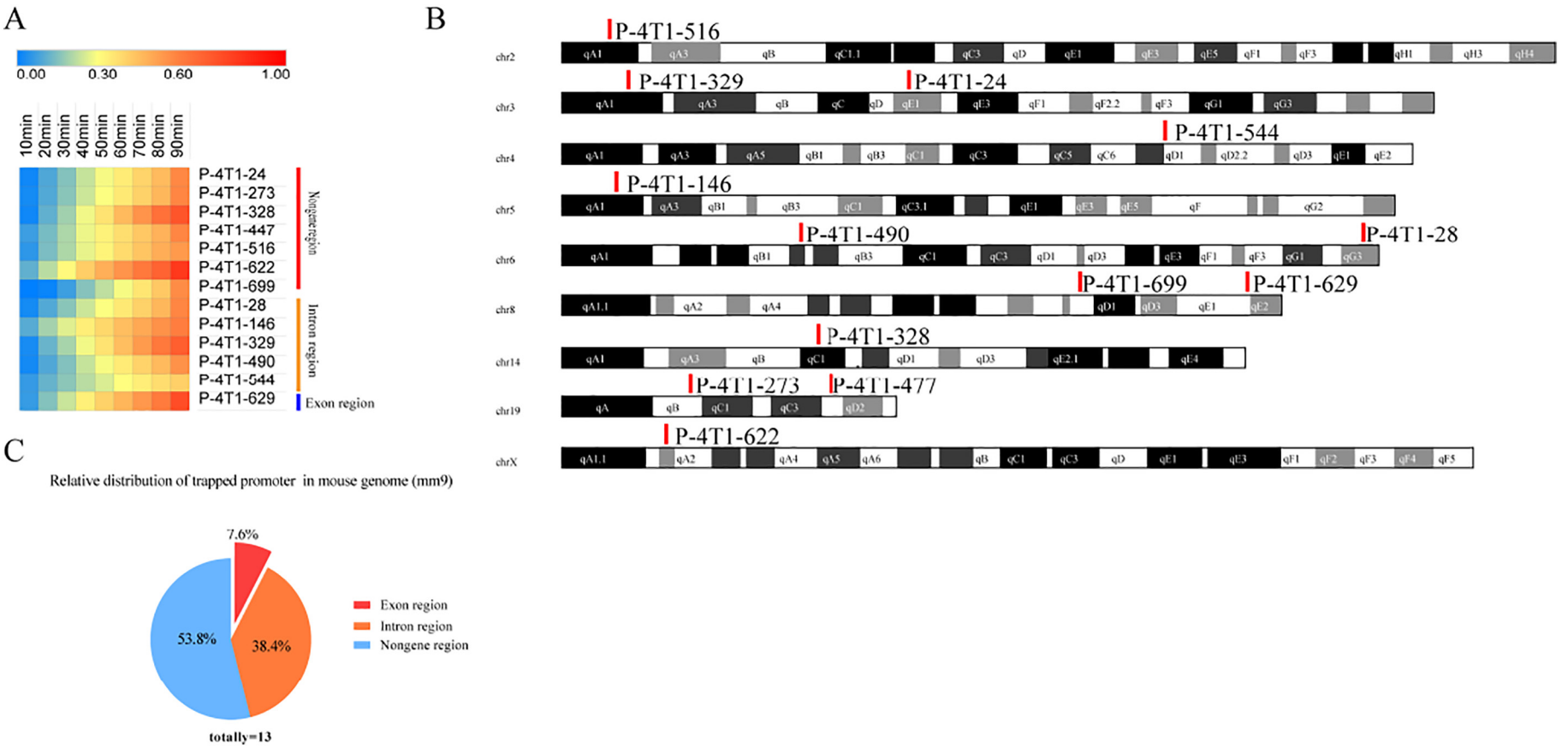

Fig. 4. Genomic localization of the 13 selected fragments with promoter activities. (A) The heat map of SEAP activities of the 13 randomly selected clones. The SEAP activities of each clone were subtracted from the negative control in the corresponding 96-well plate. The figure was displayed by the mean value of the three replicates and sorted according the location of the selected clones on mouse genome. The red line on the right represented the trapped fragments (including P-4T1-24, P-4T1-273, P-4T1-328, P-4T1-447, P-4T1-516, P-4T1-622, P-4T1-699) located in the nongene regions; the orange line on the right represented the trapped fragments (including P-4T1-28, P-4T1-146, P-4T1-329, P-4T1-490, P-4T1-544) located in the intron regions; the blue line on the right represented the trapped fragments (P-4T1-629) located in the exon region. (B) The 13 trapped fragments were mapped on mouse genome (mm9). The red line pointed the location of each trapped fragment on mouse genome. (C) Proportion of the location of the 13 trapped fragments. In total 13 trapped fragments, 7.6% was located in the gene exon region, 53.8% was located in the nongene regions and 38.4% was located in the gene intron regions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

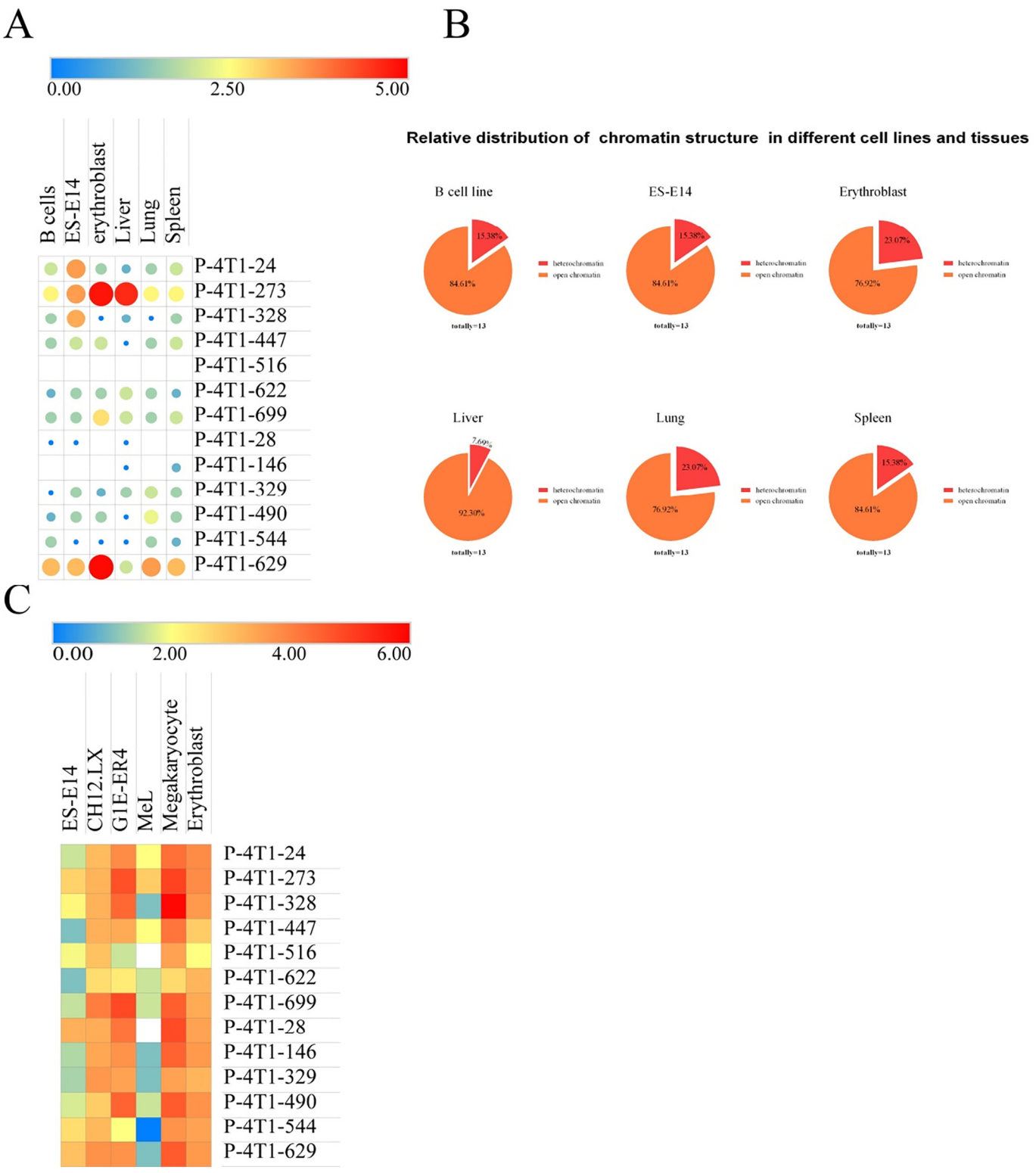

Fig. 5. DNase-seq and H3K4me3 ChIP-seq analysis detected trapped promoter activities in normal cell lines and tissues. (A) The heat map of the DNase I hypersensitivity of the 13 trapped fragments. All data were converted to log2 scale. The maximum red circle represented the value of 5 and the minimum blue circle represented the value of 0. The white block represented the trapped fragment displayed as heterochromatin in its corresponding cell lines and tissues. (B) The relative distribution of chromatin structure of the 13 trapped fragments in different cell lines and tissues. Red region represented the location of its trapped fragment on mouse genome. (C) The heat map of the H3K4me3 modification in the 13 trapped fragments. All data were converted to log2 scale. Red represented the value of 6, blue represented the value of 0 and gray represented the unmodified region. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

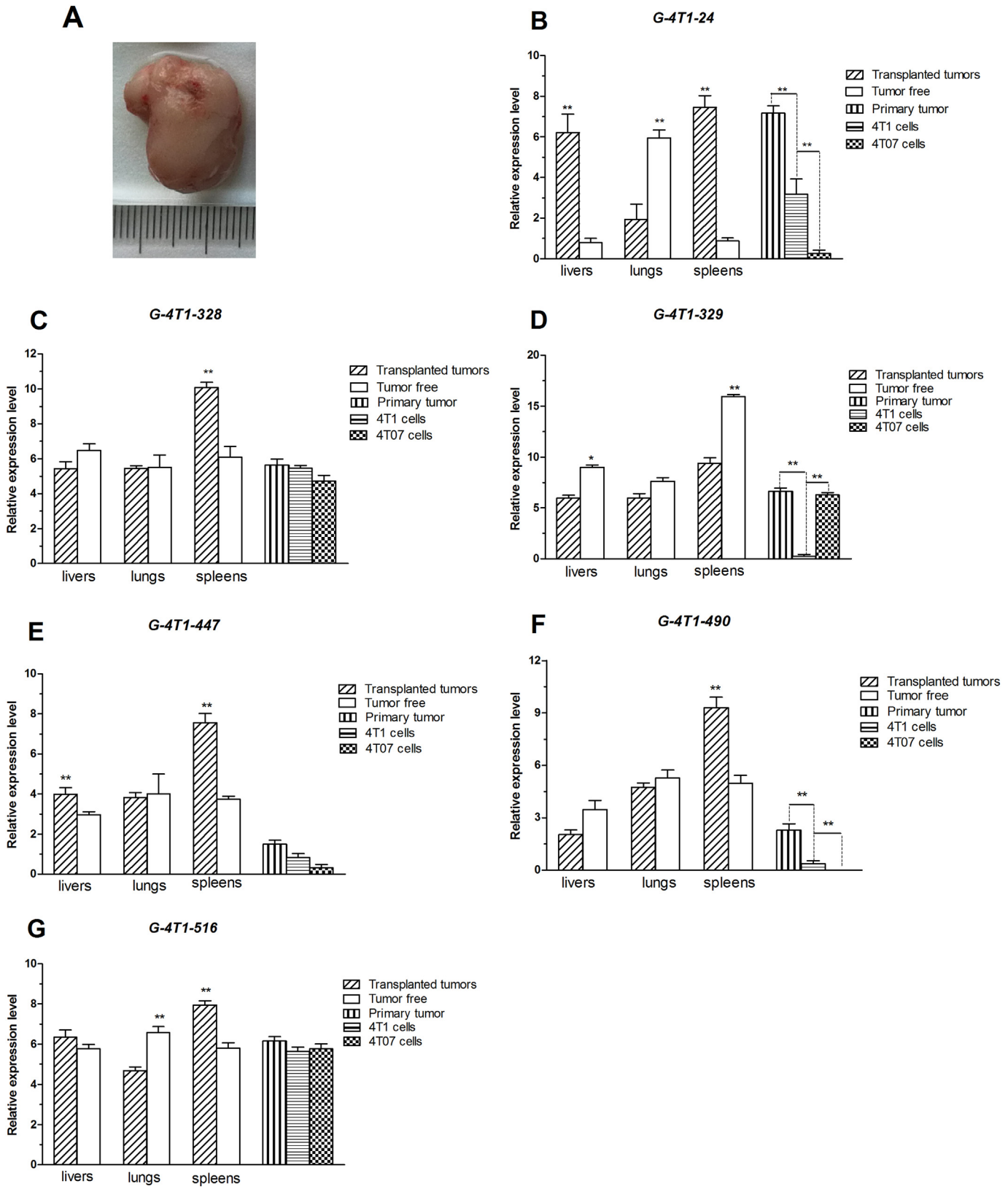

Fig. 6. Tissue specific expression of predicted genes. (A) The representative image of tumors from tumor-bearing mice at 42 days after transplantation of 4T1 cells. Tumor volume was estimated according to the formula Volume (mm^3) = L × W2 × 0.4. (B–G) qRT-PCR was carried out with primer pairs listed in S1 Table. The level of each gene was relative to that of EF1a1. Experiments were performed on three biological replicates for samples from each tissue. Bars represented the means ± standard deviation (SD) of 3 independent replicates. *P value < 0.05, **P value < 0.01. ▥ stands for the gene expression level in 4T1 cell cognate reconstituted mouse tumors, ▤ stands for the gene expression level in mouse tumor cells 4T1, ⬚ stands for the gene expression level in mouse tumor cells 4T07, ▨ stands for the gene expression level in BALB/c mouse livers, lungs, spleens after transplanted tumors, ▭ stands for the gene expression level in tumor free BALB/c mouse.

splice. *G-4T1-329* was located in the second intron of the sense strand of *predicted gene 10745*, which is a transcriptional sequence with 455 bp downstream, and amplified in 4T1 cells cDNA by RT-PCR successfully. *G-4T1-490* was located in the third intron of the sense strand of *thiamine pyrophosphokinase* (*tpk1*) gene, and a 272 bp transcriptional fragment was successfully amplified in 4T1 DNA by RT-PCR. G-4T1-146 and G-4T1-544 were failed to amplify the transcriptional sequences downstream.

3.6. Identified transcription units analysis in mouse samples by qRT-PCR

The mouse homograft model was established with 4T1 cells. Tumors grew up to 300 mm^3 when mice were sacrificed at 42 days after transplantation (Fig. 6a). Metastases were observed in lungs, but not in livers and spleens. The possibility of the metastasis after tumor formation was related to the growth time of the tumor *in situ*. In the mouse homograft model we established, it was highly metastatic in lungs, followed by livers, while spleens and kidneys were relatively rare.

To provide the transcription units expression profiles among different mouse tissues/cells experimentally, 6 transcription units (*G-4T1-24*, *G-4T1-329*, *G-4T1-328*, *G-4T1-447*, *G-4T1-490* and *G-4T1-516*) were quantified by qRT-PCR. The detection of these 6 selected transcription units in 4T1 cells and 4T07 cells, 4T1 homografted BALB/c mouse tumor tissues, livers, lungs and spleens were shown in Fig. 6, respectively.

The expression profile of transcription unit *G-4T1-24* was shown in Fig. 6B. It was apparent that the expression levels of transcription unit *G-4T1-24* in *in vitro* cultured 4T1 cells was much higher than that in 4T07 cells. In the *in vivo* 4T1 cell homografted mouse, the expression levels of transcription unit *G-4T1-24* in tumor tissues were apparently higher than that in the *in vitro* cultured 4T1 cells. In addition, the expression levels of transcription unit *G-4T1-24* in the *in vivo* 4T1 cell homografted mouse livers and spleens were significantly higher than that in the corresponding tumor free mouse livers and spleens. However, the expression levels of transcription unit *G-4T1-24* in the *in vivo* 4T1 cell homografted mouse lungs were lower than that in the tumor free mouse lungs.

For the transcription unit *G-4T1-328*, the expression differences were shown in Fig. 6C. Obvious higher levels of its expression in the *in vivo* 4T1 cell homografted mouse spleens than its expression in the tumor free mouse spleens were observed. However, no significant differences were observed when refer to its expressions in either the *in vitro* cultured 4T1 cells in comparison with the *in vitro* cultured 4T07 cells or *in vitro* cultured 4T1 cells in comparison with its *in vivo* 4T1 cell homografted mouse tumor tissues. Moreover, no significant differences of *G-4T1-328* expression levels were observed between the *in vivo* 4T1 cell homografted mouse lungs or livers and the cognate tumor free mouse lungs or livers.

For the transcription unit *G-4T1-329*, the expression differences were shown in Fig. 6D. The gene expression levels in the *in vitro* cultured 4T1 cells were much lower than that in the *in vitro* cultured 4T07 cells. However, the expression of *G-4T1-329* in the *in vivo* 4T1 cell homografted mouse tumor tissues rose to significant levels in comparison with its *in vitro* expression levels. In the *in vivo* 4T1 cell homografted mouse, the expression levels of *G-4T1-329* in livers, lungs and spleens were generally lower than in the corresponding tumor free mouse tissues, with the biggest expression difference was observed between the *in vivo* 4T1 cell homografted mouse spleens and the tumor free mouse spleens.

Fig. 6E showed the expression levels of the *G-4T1-447* transcription unit. Its expression levels in the *in vitro* cultured 4T1 cells were significantly higher than that in the *in vitro* cultured breast cancer 4T07 cells. In the *in vivo* transcription unit expression analysis, the expression levels of the *G-4T1-447* in mouse lungs and spleens from the 4T1 cell homografted mouse were significantly higher than that in the tumor free mouse. However, no significant expression level differences were observed when it came to the comparison of 4T1 cell homografted mouse livers with the cognate tumor free mouse livers.

For the transcription unit *G-4T1-490*, the expression differences were shown in Fig. 6F. The *in vitro* expression was almost undetectable in either the cultured 4T1 cells or the cultured 4T07 cells. In the *in vivo* 4T1 cell homografted mouse, the expression levels of transcription unit *G-4T1-490* in tumor tissue were higher than that in the *in vitro* cultured 4T1 cells. In the *in vivo* 4T1 cell homografted mouse, the expression levels of *G-4T1-490* in the livers and lungs were lower than those in the corresponding tumor free tissues, with the exception that the expression of transcription unit *G-4T1-490* in the *in vivo* 4T1 cell homografted mouse spleens exhibited higher expression levels than those in the tumor free mouse spleens.

For the transcription unit *G-4T1-516*, the expression differences were shown in Fig. 6G. Its *in vitro* expression levels were almost the same in both the cultured 4T1 cells and the cultured 4T07 cells. In the *in vivo* 4T1 cell homografted mouse, the expression levels of transcription unit *G-4T1-516* in tumor tissues were almost the same as in the *in vitro* cultured 4T1 cells. In the *in vivo* 4T1 cell homografted mouse, the expression levels of *G-4T1-516* in the livers and spleens were marginally higher than those in the corresponding tumor free tissues, with the exception that the expression levels of *G-4T1-516* in the *in vivo* 4T1 cell homografted mouse lungs were lower than those in the tumor free mouse lungs.

4. Discussion

Currently, the transcriptional regulation processes of gene expression can be uncovered from different perspectives by using several technologies, including PEAT (Ni et al., 2010), 5′RACE (Frohman et al., 1988), CAGE (Shiraki et al., 2003), *etc*. Chromatin immunoprecipitation followed by massively parallel DNA sequencing (ChIP-seq) investigates the interaction between DNA and proteins to illustrate the activity of promoters and enhancers through profiling chromatin modifications like H3K4me1, H3K9me3 and H3K27ac *etc*. (Sharifizarchi et al., 2017; Luke et al., 2016) CAGE maps precisely transcription start sites (TSS) through capture the 5′ end of RNAPII transcript (Haberle and Lenhard, 2016). However, it also has some limitations like it can only detect the total mature RNA or some specific RNAs such as polyadenylated RNA. And some processed coding and non-coding RNAs may captured by CAGE tags because of their cap-like structures (Haberle and Lenhard, 2016). GRO-seq detects nascent transcripts even including unstable transcripts, such as enhancer RNAs (eRNAs) (Hah et al., 2013; Core et al., 2008). Therefore, it overcomes the limitations of CAGE while requires strictly for sample preparation (Murakawa et al., 2016). Assay for transposase-accessible chromatin using sequencing (ATAC-seq) determines the chromatin accessibility by integrating Tn5 transposase into active regulatory regions (Buenrostro et al., 2013). Self-transcribing active regulatory region sequencing (STARR-seq) identifies the regulatory elements by detecting reporter activity of putative regulatory sequences (Arnold et al., 2013). However, one of the limitations is it determines the activity of reporter gene *in vitro*, thereby the real genome context is not taken into account. Massively parallel reporter assays (MPRA) (Patwardhan et al., 2009) identifies the regulatory active sequences from synthesized DNA fragments for analyzing regulatory elements and uncovering the rules of proximal promoter regulation. However, the limitations of this method also exist, such as the fragments synthesized by this method are not entirely the same as genome fragments under the natural context. Furthermore, with the extension of the amplification time and the increase of the fragment length, mutations in the synthesized fragments produced by DNA polymerase will increase quickly. Therefore, the number of cycles in PCR procedure must be reduced and the amplified fragments should limited within a certain length, which may deviate from the length of promoters or enhancers under the natural genome context (Inoue et al., 2017). Recently, with the development of three-dimensional

technology, transcriptional regulation is deeply investigated at spatial chromatin structure. Based on 3C (Miele and Dekker, 2008) and high-throughput sequencing, the Hi-C technology (Whalen et al., 2016; Hakim and Misteli, 2012) was invented to investigate the connection between promoters and distant enhancers at three-dimensional chromatin structures. Thanks to these technologies, the mystery of potential promoters or enhancers and their regulatory mechanisms are partly unveiled. In addition, some projects such as ENCODE (Yue et al., 2014; Hariharan, 2012) and FANTOM (Hoon et al., 2015) have greatly promoted the comprehension of the human and mouse transcriptome including transcription factors (TFs) binding, the transcriptional regulatory interaction network, as well as the epigenetic modifications. In this study, by using promoter finding method based transient transfection, only a single pair restriction enzyme digestion screening round, a random mouse promoter-trap library containing 706 clones was obtained, and 66.14% genomic DNA fragments with promoter activities was successfully obtained *in vitro*.

In metazoan genome, nucleosome density will be reduced and display hypersensitive to DNase I when a region is preparing for transcription (Kim and Shiekhattar, 2015). By utilizing this feature, DNase-seq can be used for investigating the chromatin accessibility to identify promoters (Natarajan et al., 2012). In this study, the published DNase-seq data of different normal cells and tissues was used to verify the feasibility of our approach. In the 13 trapped fragments we randomly selected, 75% of the fragments displayed euchromatin state in the selected cell lines and tissues. Notably, P-4T1-516 displayed heterochromatin state in all selected cell lines and tissues. Comparing with the high promoter activity in 4T1 cells, it suggested that the fragment of P-4T1-516 may possess a cancer-gained promoter activity in mouse breast cancer. With the increasing number of transcriptional regulation studies, the chromatin modifications have been proven as a distinguish signature for discriminating promoters, such as the histone markers H3K4me3, which usually resides active gene promoters (Benoit et al., 2011). Published ChIP-seq data of H3K4me3 in different normal cells and tissues was further used to verify the feasibility of our approach. All 13 trapped fragments we randomly selected were modified with H3K4me3 in all selected cell lines and tissues except P-4T1-28 and P-4T1-516, which were only not displayed in Mel cells. These results further indicated that our approach in this article is feasible. Epigenetic modifications of promoter regions were differed in a tissue-specific manner. The result of WashU suggested that P-4T1-328 displayed a high histone modification degree of H3K4me3 in megakaryocyte cell lines (max score = 59). More than 50% of regions of trapped promoters score were over 20 points in megakaryocyte cell lines.

Another picture is that many of our promoter-trap identified transcription units were localized in the non-coding region of mouse genome. This kind of non-coding region could serve as miRNA templates, demonstrating the involvement of the non-coding region in forming transcriptional regulatory networks in cells (Fang et al., 2012). JM Silva proposed that the *long stress-induced non-coding transcript 5*(*LSINCT5*) can affect cellular proliferation through its over expression in breast and ovarian cancers (Silva et al., 1900). Non-coding RNAs have been divided into housekeeping noncoding RNAs and regulatory noncoding RNAs (You et al., 2014). Housekeeping non-coding RNAs are usually expressed constitutively, while regulatory noncoding RNAs are tissue specific. Our qRT-PCR results showed that some identified transcripts that signature with tissue-specific expression properties may contain the potential of regulatory non-coding RNAs.

Although we have improved the promoter-trap strategy based on transient transfection, the potential drawbacks still exist. First, the random genome DNA fragments only can be obtained after digested with the specific restriction enzyme sites. Second, there exists a probability that the fragments with transcription activities containing the specific restriction sites were cut off, which result in failing to detect the promoter activities. Last, it still needs to detect the activities of trapped fragments to further identify whether it harbors promoter activities or not after constructed the promoter-trap library.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2019.100008.

Declaration of interest statement

The authors declared that they have no conflicts of interest to this work. We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

Funding

This work was supported by the “Financial Support for Selected Researchers Back from Abroad (2012)” from Liaoning Province (http://www.lninfo.gov.cn/) (Grant No. 88030312004).

References

Arensbergen, V., Joris, Steensel, V., Bussemaker, Harmen, J., 2014. In Search of the Determinants of Enhancer–promoter Interaction Specificity. Mouton, pp. 695–702.
Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., Stark, A., 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339 (6123), 1074–1077.
Barski, A., Cuddapah, S.K., Roh, T., Schones, D., Wang, Z., Wei, G., et al., 2016. High-resolution profiling of histone methylations in the human genome. Cell 129 (4), 823–837.
Benoit, G., Paul, D., Lin, H.H.S., Harry, A., Kyoko, H.H., Axel, I., et al., 2011. H3 lysine 4 is acetylated at active gene promoters and is regulated by H3 lysine 4 methylation. PLoS Genet. 7 (3), e1001354.
Berger, J., Hauber, J., Hauber, R., Geiger, R., Cullen, B.R., 1988. Secreted placental alkaline phosphatase: a powerful new quantitative indicator of gene expression in eukaryotic cells. Gene 66 (1), 1.
Boshart, M., Weber, F., Jahn, G., Dorsch-Häsler, K., Fleckenstein, B., Schaffner, W., 1985. A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. Cell 41 (2), 521.
Brozovic, M., Dantec, C., Dardaillon, J., Dauga, D., Faure, E., Gineste, M., et al., 2017. ANISEED 2017: extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. Nucleic Acids Res.(Database issue), gkx1108.
Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J., 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods 10 (12), 1213.
Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., et al., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 55 (4), 611.
Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., et al., 2005. The transcriptional landscape of the mammalian genome. Science 309 (5740), 1559–1563.
Consortium, F., Pmi, T.R., CLST, Forrest, A.R., Kawaji, H., Rehli, M., et al., 2014. A promoter-level mammalian expression atlas. Nature 507 (7493), 462–470.
Core, L.J., Waterfall, J.J., Lis, J.T., 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322 (5909), 1845–1848.
Core, L.J., Martins, A.L., Danko, C.G., Waters, C.T., Siepel, A., Lis, J.T., 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. Nat. Genet. 46 (12), 1311–1320.
Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., et al., 2011. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. 107 (50), 21931–21936.
Fang, W., Wang, X., Bracht, J.R., Nowacki, M., Landweber, L.F., 2012. Piwi-interacting RNAs protect DNA against loss during Oxytricha genome rearrangement. Cell 151 (6), 1243–1255.
Foecking, M.K., Hofstetter, H., 1986. Powerful and versatile enhancer-promoter unit for mammalian expression vectors. Gene 45 (1), 101–105.
Frohman, M.A., Dush, M.K., Martin, G.R., 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc. Natl. Acad. Sci. U. S. A. 85 (23), 8998.
Haberle, V., Lenhard, B., 2016. Promoter architectures and developmental gene regulation. Semin. Cell Dev. Biol. 57, 11–23.
Hah, N., Murakami, S., Nagari, A., Danko, C.G., Kraus, W.L., 2013. Enhancer transcripts mark active estrogen receptor binding sites. Genome Res. 23 (8), 1210–1223.
Hakim, O., Misteli, T., 2012. SnapShot: chromosome confirmation capture. Cell 148 (5), 1068.e1.
Hariharan, M., 2012. Architecture of the Human Regulatory Network Derived From ENCODE Data.
Hoon, M.D., Shin, J.W., Carninci, P., 2015. Paradigm shifts in genomics through the FANTOM projects. Mamm. Genome 26 (9–10), 1–12.
Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., Mcmanus, M.T., et al.,

2017. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. Genome Res. 27 (1), 38–52.
Kain, S.R., 1997. Use of secreted alkaline phosphatase as a reporter of gene expression in mammalian cells. Methods Mol. Biol. 63 (63), 49–60.
Kim, T.K., Shiekhattar, R., 2015. Architectural and functional commonalities between enhancers and promoters. Cell 162 (5), 948–959.
Lenhard, B., Sandelin, A., Carninci, P., 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nat. Rev. Genet. 13 (4), 233–245.
Li, B., Carey, M., Workman, J.L., 2007. The role of chromatin during transcription. Cell 128 (4), 707–719.
Long, H.K., Prescott, S.L., Wysocka, J., 2016. Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell 167 (5), 1170.
Luke, I., Lexie, P., Wu, H., Lucia, D., Harald, O., Alex, S., et al., 2016. Wiz binds active promoters and CTCF-binding sites and is required for normal behaviour in the mouse. elife 5.
Miele, A., Dekker, J., 2008. Mapping Cis - and Trans - Chromatin Interaction Networks Using Chromosome Conformation Capture (3C). Humana Press, pp. 105–121.
Murakawa, Y., Yoshihara, M., Kawaji, H., Nishikawa, M., Zayed, H., Suzuki, H., et al., 2016. Enhanced identification of transcriptional enhancers provides mechanistic insights into diseases. Trends Genet. 32 (2), 76–88.
Natarajan, A., Yardımcı, G.G., Sheffield, N.C., Crawford, G.E., Ohler, U., 2012. Predicting cell-type-specific gene expression from regions of open chromatin. Genome Res. 22 (9), 1711–1722.
Ni, T., Corcoran, D.L., Rach, E.A., Song, S., Spana, E.P., Gao, Y., et al., 2010. A paired-end sequencing strategy to map the complex landscape of transcription initiation. Nat. Methods 7 (7), 521.
Patwardhan, R.P., Lee, C., Litvin, O., Young, D.L., Pe'Er, D., Shendure, J., 2009. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. Nat. Biotechnol. 27 (12), 1173.
Radaiglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., Wysocka, J., 2009. A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470 (7333), 279–283.
Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V. (Eds.), 2010. Database Resources of the National Center for Biotechnology Information. Haptics Symposium.
Sharifizarchi, A., Gerovska, D., Adachi, K., Totonchi, M., Pezeshk, H., Taft, R.J., et al., 2017. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. BMC Genomics 18 (1), 964.
Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., et al., 2003. Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc. Natl. Acad. Sci. U. S. A. 100 (26), 15776–15781.
Silva, J.M., Boczek, N.J., Berres, M.W., Ma, X., Smith, D.I., 1900. LSINCT5 is over expressed in breast and ovarian cancer and affects cellular proliferation. RNA Biol. 8 (3), 496–505.
Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D., 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7 (Suppl. 1(1)), S10.1.
Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. 15 (4), 272.
Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P., Myers, R.M., 2004. An abundance of bidirectional promoters in the human genome. Genome Res. 14 (1), 62–66.
Valen, E., Sandelin, A., 2011. Genomic and chromatin signals underlying transcription start-site selection. Trends Genet. 27 (11), 475–485.
Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., et al., 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. 40 (7), 897–903.
Whalen, S., Truty, R.M., Pollard, K.S., 2016. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat. Genet. 48 (5).
You, J., Zhang, Y., Liu, B., Li, Y., Fang, N., Zu, L., et al., 2014. MicroRNA-449a inhibits cell growth in lung cancer and regulates long noncoding RNA nuclear enriched abundant transcript 1. Indian J. Cancer 51 (Suppl. 3(7)), e77.
Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., et al., 2014. A comparative encyclopedia of DNA elements in the mouse genome. Nature 515 (7527), 355–364.