Circularis™ Mini-monomer™ Promoter Discovery and Sequence Selection Technology

Introduction

Transcription is the first step of information transfer in cells as represented by the classical "Central Dogma" of Watson and Crick. As a result, it represents an obvious point for controlling the production of the ultimate object of that information flow—proteins. Transcription control sequences include promoters, terminators, and RNA stability-control sequences, and are categorically referred to as regulatory elements. Traditional sequencing methods for finding regulatory sequences include Whole Genome Sequencing (WGS) and RNA-Seq. While these methods provide sufficient data to give insight into the actual proteins expressed, these methods provide no insight into the regulatory control sequences. Another common approach is the analysis of suspected promoter regions 1kb immediately upstream of individual genes being studied. While many transcription factors bind to this region, their specific binding motif is not known, making the optimization of a promoter region virtually a guessing game. A more recent method of analyzing how transcription factors interact with DNA to regulate gene expression is chromatin immuno-precipitation (ChIP-Seq). ChIP-Seq provides epigenetic information that can complement gene expression data, but it is limited by antibody affinity and specificity as well as by an exceptionally low number of available antibodies (167 ChIP-Seq antibody targets in the ENCODE[i] ChIP-Seq Experiment Matrix). Circularis has developed alternative methods[ii] that allow global selection, analysis and evolution of all types of regulatory elements.

Using one of our methods, all active promoters in an organisms' genome can be found, analyzed and then evolved for superior function. Using another, related method, we can discover and optimize promoter sequences. Our method is naïve—it requires no foreknowledge of an organism's transcription control sequences. It can be applied to the

	WGS	RNA-seq	ChIP-seq	1Kb "Region"	In-silico Design	Circularis
Genomic Coverage	++++	+	+	+	+++	+++
Regulatory Coverage	-	-	++	++	++	++++
Binding Motif	-	-	++	++	+	+++
# Elements Measured	-	-	+	+	++	++++
Required Resources	Very High	Very High	High	Highest	Very High	Lowest
Validation Time	Very High	Very High	High	Highest	High	Lowest
Total Costs	Very High	Very High	Very High	Highest	Very High	Lowest

Table 1. Methods of discovering regulatory elements.

genomic DNA of organisms from any kingdom. It can also be applied to entirely synthetic sequences to form artificial transcription control sequences. One important aspect of our system is that any of the transcription control sequences that we develop are not be bound by the Supreme Court's Myriad decision, and as such, are patentable.

Materials and methods

While there is a wide range of applications for our technology, this paper will focus on prokaryotic *in vivo* and *in vitro* proof of concept experiments that demonstrate how our technology is used to discover a known promoter. The overall flowchart of our technology is shown in Figure 1.

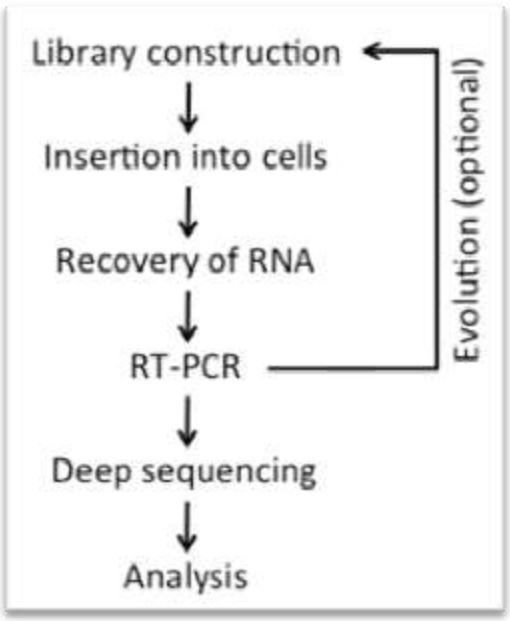

Figure 1. Flowchart of Circularis Mini-Monomer method.

Plasmid Construction

To test our ability to find functional promoters in a library of genomic DNA, two constructs were made in the Circularis Mini-monomer plasmid—pCirc T7, containing a functional, 18 nucleotide T7 promoter sequence and, as a control, pCirc 7T, containing a non-functional T7-related sequence—the T7 promoter sequence written backwards (this is not the reverse complement, which would place a functional T7 promoter in the opposite orientation, but the functional sequence written 3' to 5'). For the evolution of a known promoter, a library of randomized T7 promoters containing at each of 17 out of 18 positions, 75% of the correct nucleotide and 25% spread among the 3 incorrect nucleotides, was cloned into the same plasmid forming pCirc T7lib. The library contained approximately 500,000 unique members. Each of these individual plasmids or the plasmid library was purified from E. coli Top10 cells using a Zyppy™ commercial plasmid preparation kit (Zymoresearch).

In-vitro Transcriptions

In-vitro transcriptions of 20µl were performed using the standard conditions provided with the T7 RNA polymerase (NEB) using 0.4µg of supercoiled plasmid DNA and without optional ribonuclease inhibitor. Transcriptions were incubated for 1 to 4 hours at 37°C, then terminated by addition of an equal volume of 25:24:1 phenol:chloroform:iso-amyl alcohol, followed by vortexing and removal of the aqueous phase into a new tube. One tenth volume of 3M NaAc (pH5.2) and 2.5 volumes of 99.5% ethanol were added, the samples were incubated at -20°C for 15 minutes, then centrifuged at room temperature for 10 minutes to precipitate the synthesized RNA. The liquid phase was removed, the pellet was briefly air dried, then re-suspended in 20µl of sterile water.

In-vivo Transcriptions

Both pCirc T7 and pCirc 7T were transformed into chemically competent T7 expressing cells (New England BioLabs) and grown overnight in LB with 1% glucose and 100µg/mL carbenicillin. Using the stationary phase cell culture for each construct, new cultures in LB

(100µg/mL carbenicillin) at an OD_{600} of 0.1 were started. The OD_{600} was continually monitored for each culture, and once the cell cultures reached an OD_{600} reading of 0.6, IPTG was added and cells were allowed to grow for 2 hours.

Two hours following IPTG addition, the OD_{600} readings were again measured. Cells with the functional T7 promoter grew to an OD_{600} of 0.7 while cells with the nonfunctional 7T promoter grew to an OD_{600} of 2.1. Additional LB was added to the 7T cultures to bring the OD_{600} to 0.7. The functional promoter culture (pCirc T7) was then serially diluted with nonfunctional promoter culture (pCirc 7T) up to 1 in 10^5.

RNA from each of the dilution series was then isolated using the RNeasy Plus Minikit (Qiagen). An RT PCR was then performed using dilution series and results were run on an agarose gel.

Reverse transcription-polymerase chain reaction (RT-PCR)

Complementary DNA (cDNA) was synthesized using 1µl of the in vitro synthesized RNA in reverse transcription reactions of 20µl using the standard conditions provided with the SuperScript™ III reverse transcriptase (LifeTechnologies) used and performed without the optional ribonuclease inhibitor. Polymerase chain reactions (PCR) of 25µl were performed using 1µl of the cDNA using the standard conditions provided with the Pfusion™ DNA polymerase (NEB) and primers appropriate to amplify the desired sequences and adding the appropriate sequences for Illumina deep sequencing. Sequencing results available in subsequent publication(s).

Results

Transcription reactions with and without the addition of T7 RNA polymerase were performed using the pCirc T7 and pCirc 7T plasmids. After extraction, reverse transcription and PCR for 20 cycles, 5µl of each reaction was loaded onto a 1.5% agarose gel. The result is shown in Figure 2A and 2C, for *in vitro* and *in vivo* work, respectively. The expected PCR product is only recovered when the pCirc T7 transcript is incubated with T7 RNA polymerase. No correctly sized band is seen in the pCirc 7T lane. The appearance of a faint, somewhat larger PCR artefact band in the lanes with this plasmid, can

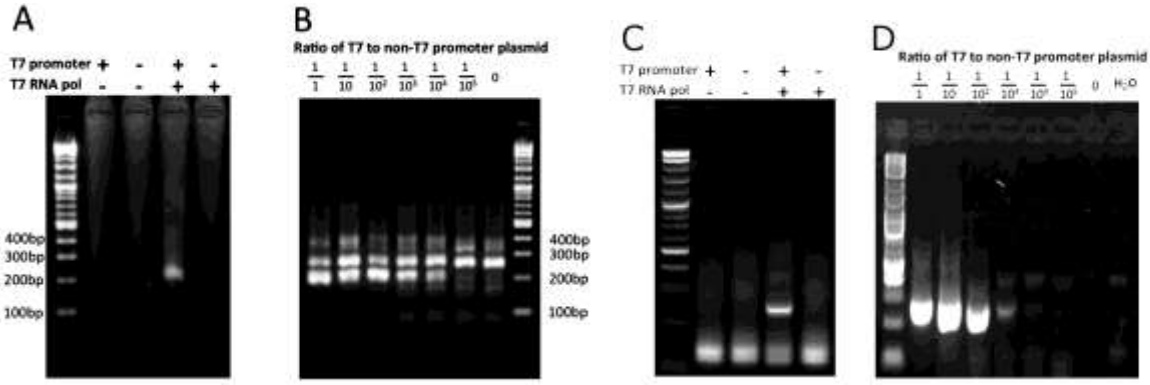

Figure 2. A-B *in vitro* results, C-D *in vivo* results. A-C, Recovery of promoter sequence as a PCR product from pCirc T7 transcripts. The T7 promoter positive lanes contain pCirc T7 plasmid and the T7 promoter, negative lanes contain pCirc 7T plasmid. The expected band is 221 base pairs. B-D, Recovery of promoter sequence as a PCR product from a dilution series of pCirc T7 plasmid in a constant amount of pCirc 7T plasmid. The ratio above the lanes represents the ratio of pCirc T7 plasmid to pCirc 7T plasmid. The "0" lane contains only the pCirc 7T plasmid. The *in vivo* results in D include a water only lane to determine the presence of an artefact band in the PCR tube.

be attributed to remnant DNA from the polymerase supplier, and is addressed below.

To determine the limit of our ability to recover a rare promoter sequence in a library of genomic fragments, a 10 fold dilution series of pCirc T7 was performed with a constant amount of pCirc 7T. After transcription, extraction, and RT-PCR with 35 cycles of PCR amplification, 5µl of each reaction was loaded on a 1.5% agarose gel. The results of this experiment are shown in Figure 2B and 2D, for *in vitro* and *in vivo* work, respectively. The correct band is seen at a dilution of 1 pCirc T7 plasmid in 10^4 pCirc 7T plasmids. The artefact band seen faintly in the pCirc 7T lane in Figure 2A is now substantial in Figure 2B. An examination of the sequences of the plasmid has putatively identified an alternative primer binding site, which reduced the prominence of the band in the *in vivo* data. The appearance of this artefact band in the water lane of the *in vivo* data suggests that there is remnant DNA in the tube from the supplier.

Conclusion

We have demonstrated that we can recover rare promoter sequences from a population containing large numbers of non-promoter sequences. Applying this same technology, we are able to find functional promoters within a library of genomic DNA cloned into our pCirc Mini-monomer system. Once the Mini-Monomer is stable and proven to be transcriptionally active in the host cells, the process of discovery takes less than 3 weeks, enabling us to discover new promoters at speeds and numbers that are cost-prohibitive using current technologies.

Our next steps with the T7 promoter library will include mutating the consensus sequence to generate a variant library of approximately 500,000 novel promoter sequences, measuring the activity of individual promoter sequences, and confirming protein production.

---

[i] ENCODE: Encyclopedia of DNA Elements. https://www.encodeproject.org/

[ii] Methods covered by two provisional PCT patents:
Compositions and methods for identifying polynucleotides of interest
Methods for RNA promoter identification