A dual sgRNA mediated tiling-deletion based genetic screen to identify regulatory DNA sequence in mammalian cells

Authors:

Yarui Diao[1], Rongxin Fang[1], Bin Li[1] and Bing Ren[1, 2, *]

Affiliations:

[1] Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA.

[2] Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, and Moores Cancer Center, University of California at San Diego, La Jolla, CA 92093, USA.

[*] E-mail: biren@ucsd.edu

Abstract:

Recently, CRISPR/Cas9 genome editing has been employed to introduce short insertions and deletions and identify transcriptional regulatory elements in a high-throughput fashion. However, scalability of these methods is limited by the number of targeted perturbations that can be introduced. Further, existing approaches neglected the issue of *cis*- and *trans*-regulation. Here we report a highly scalable method, *Cis* Regulatory Element Scan by Tiling-deletion and sequencing (CREST-seq), for unbiased discovery and functional assessment of cis regulatory sequences in the genome with improved coverage and sensitivity comparing to existing single sgRNA screen approach.

Introduction:

Developmental programs of multi-cellular organisms are controlled by *cis*-regulatory sequences including enhancers, promoters, and insulators[1-31-31-31-31-31-31-31-3]. Millions of candidate *cis*-regulatory elements have been annotated in the human genome based on histone modification, transcriptional factor binding, and DNase I hypersensitivity[4-9]. These putative regulatory sequences harbor a disproportionally large number of sequence variants associated with diverse human traits and diseases, supporting the hypothesis that non-coding sequence variants contribute to common traits and diseases via transcriptional regulation[10-12]. However, research on the role of these putative functional elements in disease pathogenesis has been hindered by a dearth of direct evidence for their biological function in the native genomic context.

High throughput CRISPR/Cas9-mediated mutagenesis using single guide RNAs (sgRNAs) have been developed to functionally characterize *cis*-regulatory elements in mammalian cells [13-18]. Currently, the utility of this approach for *de novo* discovery of *cis*-regulatory elements is limited by several factors: (1) Not all sequences are suitable for CRISPR/Cas9 mediated genome editing due to the lack of protospacer adjacent motifs (PAMs) that are required for targeting and DNA cutting by CRISPR/Cas9 [19-21]; (2) CRISPR/Cas9 mediated genome editing with single sgRNAs tends to cause point mutations or short insertion or deletions [14], which may not be sufficient to disrupt the function of a *cis*-regulatory element; (3) an unrealistic number of sgRNAs would be required to interrogate the human genome; (4) it has been challenging to distinguish the *cis*- and *trans*-regulatory effects. To overcome these limitations, we developed CREST-seq, short for *Cis* Regulatory Elements Scan by Tiling-deletion and Sequencing, which enables efficient discovery and functional characterization of *cis*-regulatory elements by introducing massively parallel, kilobase-long deletions to the genome. The utility of CREST-seq in human embryonic stem cells and shows that CREST-seq outperforms the previous CRISPR/Cas9 screening studies to study *cis*-regulatory elements.

Reagents:

1) DMEM/HIGH GLUCOSE (Hyclone)
2) Phosphate Buffered Saline (Hyclone)
3) FBS (Hyclone, USA)
4) penicillin-streptomycin (Hyclone)
5) PolyJet In Vitro DNA Transfection Reagent (Signagen)
6) Opti-MEM(Gibco, Life Technologies)
7) 0.25% Trypsin-EDTA(1x) (Gibco, Life Technologies)
8) Puromycin (Merck Millipore)
9) Blasticidin (10 ug/mL)(InvivoGen)
10) UltraPure TM Distilled Water(InvivoGen)
11) MicroElute Genomic DNA Kit (Omega)
12) Phusion® High-Fidelity PCR Kit (New England Biolab, MA)
13) dNTP mixture (Takara)
14) AmPure XP Beads (Beckman Coulter)
15) Magnetic Plate (Life Technologies)
16) Ethanol (molecular biology grade)
17) Primer set for primary PCR (marker gene variable region).
18) Indexing primer set for indexing PCR
19) Gibson Assembly® Master Mix (NEB)
20) T4 DNA ligase (400,000 units/ml, NEB)
21) Esp3I(BsmBI, ThermoFisher)

Equipment:

1) Manual Pipettes
2) Pipette tips
3) Tissue culture facilities and equipment
4) Light microscope
5) Haemocytometer
6) 6-well plate, 10cm and 15cm petridish for cell culture
7) 1.5 mL microfuge tubes.
8) 200 ul PCR tubes.
9) nanodrop
10) centrifuge
11) thermomixer
12) Thermocycler
13) Vortex
14) Gel electrophoresis equipment
15) Access to an Illumina sequencer

Procedure:

A. Before getting your hands wet…

1. Oligo library Design (~2 weeks to receive the order):

Design paired sgRNA library with the following format and order from Agilent.

5’AAAGGACGAAACACC(sgRNA#1)GTTTAGAGACGNNNNNNNNCGTCTCACCTT(sgRNA#2)GT
TTTAGAGCTAGAAATAGCAAG-3’

*N denotes random nucleotide.

2. Order gBlock from IDT (~ 1 to 2 weeks)

Order the following sequence containing tracRNA and mouse U6 promoter sequence:

5’-

tacgcaCGTCTCAGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAGTCCG
TTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTCTCGAGTACTAGGATCCATTAG
GCGGCCGCGTGGATAACCGTATTACCGCCATGCATTAGTTATTAGATCTAGAGATCCGACGCG
CCATCTCTAGGCCCGCGCCGGCCCCCTCGCACAGACTTGTGGGAGAAGCTCGGCTACTCCCCT
GCCCCGGTTAATTTGCATATAATATTTCCTAGTAACTATAGAGGCTTAATGTGCGATAAAAGA
CAGATAATCTGTTCTTTTTAATACTAGCTACATTTTACATGATAGGCTTGGATTTCTATAACTT
CGTATAGCATACATTATACGAAGTTATAAACAGCACAAAAGGAAACTCACCCTAACTGTAAA
GTAATTGTGTGTTTTGAGACTATAAGTATCCCTTGGAGAACCACCTTGGAGACGtgcgta-3’

3. Order the PCR primers to amplify oligo library and gBlock DNA.

Fwd_oligo: 5’-TGTGGAAAGGACGAAACACCG-3’

Rev_oligo: 5’-GACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC-3’

Fwd_gBlock: 5’-TACGCACGTCTCAGTTTAAGAGCTATGCTGGAAACAG-3’

Rev_gBlock: 5’-TACGCACGTCTCCAAGGTGGTTCTCCAAGGGATACT-3’

B. Plasmid DNA cloning and amplification

Day 1 (receiving oligo library and gBlock):

1. PCR amplification of oligo library (with less than 20 cycles) amd gBlock DNA.

2. After PCR, purify library DNA and gBlock DNA with (Agencourt AMPure XP beads)/sample ratio 3:1 and 1.8:1, respectively.
3. Linearize lentiCRISPRv2 plasmid with BsmBI followed by gel purification.
4. Perform Gibson Assembly per manufacturer's instructions. Use up to 500ng of total DNA per 10ul reaction and 5:1 insert vs. vector (mole ratio). Incubate 2 hours.
5. Remove protein with PHENOL:CHLOROFORM:ISOAMYL ALCOHOL (25:24:1) followed by DNA precipitate with 70% Ethanol. Glycogen could be used to facility the precipitation. Dilute Gilbson Assembly reaction if the volume is small. Samples can be stored at -80 overnight.

Day 2

1. Dissolve precipitated DNA with TE.
2. Amplify DNA library following the protocol described by Sanjana et al [22], and calculate the coverage of DNA library.

Day 3

1. Maxi prep of plasmid DNA.
2. Digest plasmid library with BsmBI, followed by gel purification.
3. Digest gBlock DNA, followed by Agencourt AMPure XP beads purification.
4. Overnight ligation at 16 degree using high concentration (400,000u/ml) T4 DNA ligase per NEB's instruction.

Day 4.

1. Purify ligated DNA with PHENOL:CHLOROFORM:ISOAMYL ALCOHOL(25:24:1) followed by 70% Ethanol precipitation. Incubate DNA at -20 for 2 hours (or overnight).
2. Amplify DNA library following the protocol described by Sanjana et al [22], and calculate the coverage of DNA library.
3. Order PCR primers with the following sequence:

Fwd_1: TCCCTACACGACGCTCTTCCGATCTNNNNNGGAAAGGACGAAACACCG
Fwd_2: TCCCTACACGACGCTCTTCCGATCTNNNNNHGGAAAGGACGAAACACCG
Fwd_3: TCCCTACACGACGCTCTTCCGATCTNNNNNHHGGAAAGGACGAAACACCG
Fwd_4: TCCCTACACGACGCTCTTCCGATCTNNNNNHHYGGAAAGGACGAAACACCG

Rev_1: GGAGTTCAGACGTGTGCTCTTCCGATCNNNNNTGCTATTTCTAGCTCTAAAAC
Rev_2: GGAGTTCAGACGTGTGCTCTTCCGATCNNNNNVTGCTATTTCTAGCTCTAAAAC
Rev_3: GGAGTTCAGACGTGTGCTCTTCCGATCNNNNNVMTGCTATTTCTAGCTCTAAAAC
Rev_4: GGAGTTCAGACGTGTGCTCTTCCGATCNNNNNVMAATGCTATTTCTAGCTCTAAAAC
Fwd_TruSeq: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGA

Rev_TruSeq:CAAGCAGAAGACGGCATACGAGAT(index)GTGACTGGAGTTCAGACGTGTGCTCTTCCG

Day 5.

1. Maxi prep of plasmid with Endotoxin free kit.
2. QC library quality and coverage:
   a. Mix Fwd_1/2/3/4 at 1:1:1:1 as Fwd_mix, and mix Rev_1/2/3/4 at 1:1:1:1 as Rev_mix
   b. PCR amplicaiton of plasmid DNA with Fwd_mix and Rev_mix, <10 cycles.
   c. Purify PCR DNA as template for 2nd PCR reaction to add TruSeq adaptor sequence.
   d. PCR with Fwd_TruSeq and Rev_TruSeq in less than 10 cycles.
3. Purify DNA for deep sequencing in HiSeq, paired end 100 cycles. Map the reads to the library design for QC purpose.

C. Lentiviral prep, low MOI infection, screening and collection of cell populations with desired phenotype.

The following steps are performed as described by by Sanjana et al [22] and the workflow is illustrated in Figure 1. The major steps are:

1. The CREST-seq lentiviral library was prepared as previously described [23]. Briefly, 5ug of lentiCRISPR plasmid library was co-transfected with 4 ug PsPAX2 and 1 ug pMD2.G (Addgene #12260 and #12259) into a 10-cm dish of HEK293T cells in DMEM (Life Technologies) containing 10% FBS (Life Technologies) by PolyJet transfection reagents (Signagen, Cat# SL100688). Scale up the number of HEK293T cells and transfection depending on the yield and library size. Growth medium was replaced 6 hours after transfection.
2. Collect culture media at 12 hour, 24 hour and 36 hour post transfection.
3. Concentrate lentviral by ultra-centrifugation and titrate leniviral using the same type cells for screen.

D. CREST-seq, or tiling deletion screen, to identify regulatory DNA elements.

Day 0.

Infection the reporter cells with lentiviral library at MOI<0.3.

Day 2.

48 hours post viral infection, replace fresh media containing 500ng/ml Puromycin.

Day 3-14

Regular cell culture, changing media and spit the cells when needed.

Day 15

1. Dissociate the cells for FACS sorting. The testing samples, negative control samples, and the non-sorted control samples should be collected with at least three replicates.
2. Purify genomic DNA.
3. Two-step PCR amplification of sgRNA pairs from genomic DNA. Follow the protocol described earlier (Day 5, **Plasmid DNA cloning and amplification**).
4. Purify PCR product and subject to high throughput sequencing.
5. Data analysis. We used MAGeCK for data processing and enrichment analysis.

Figure

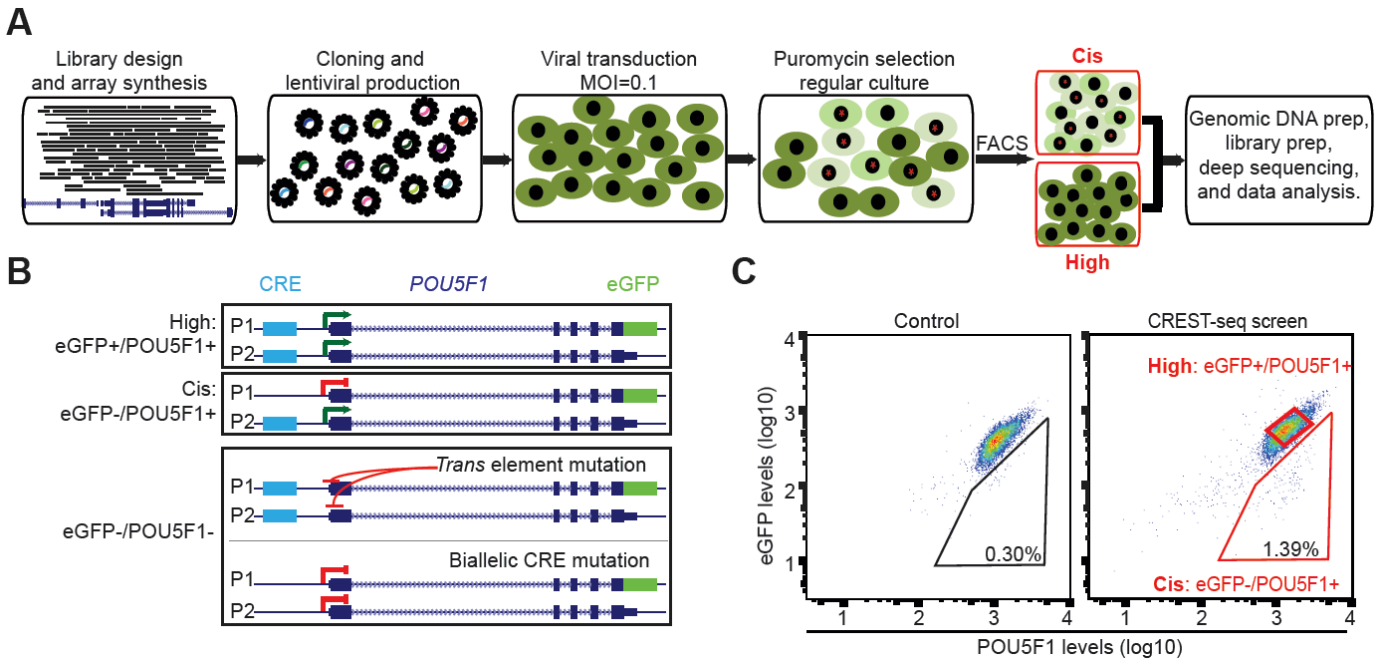

Figure 1. CREST-seq experimental design and application to the *POU5F1* locus in hESC.

(A) Workflow of CREST-seq. A total of 11,570 sgRNA pairs were designed to excise genomic sequences with an average deletion size of ~2kb in the 2Mbp *POU5F1* locus. The library of oligonucleotides corresponding to the sgRNA pairs was cloned into the lentiCRISPRv2 vector, and packaged into a lentiviral library that was in turn transduced into the H1 POU5F1-eGFP cells with MOI=0.1. 24 hours after infection, the cells were treated with Puromycin for 7 days, cultured in regular media for another 7 days. Prior to FACS, the cells were dissociated into single cells, and stained with antibodies specifically recognizing POU5F1 (PE) or eGFP (APC), respectively. The eGFP-/POU5F1+ ("Cis") and eGFP+/POU5F1+ ("High") cells were collected separately by FACS. The integrated copies of sgRNA pairs were amplified by PCR from genomic DNA followed by deep sequencing.

(**B**) Schematic illustration of allelic deletions and the expected effects on allelic POU5F1 expression. The eGFP-tagging allele is designated as P1 and the wild-type allele as P2. In wide-type cells eGFP POU5F1 protein levels are linearly correlated (eGFP+/POU5F1+). Mono-allelic disruption of a POU5F1 CRE on the P1 allele would lead to reduced eGFP expression while POU5F1 protein levels remain relatively unchanged (eGFP-/POU5F1+). Biallelic disruption of a POU5F1 CRE would lead to eGFP-/POU5F1-.

(C) FACS analysis of H1 POU5F1-eGFP cells transduced with control lentivirus expressing Cas9 but not sgRNA (left) or the CREST-seq lentiviral library (right) 14 days post transduction.

Reference:

1 Gorkin, D. U., Leung, D. & Ren, B. The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell* **14**, 762-775, doi:10.1016/j.stem.2014.05.017
S1934-5909(14)00204-5 [pii] (2014).

2 Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13-25, doi:10.1016/j.cell.2014.02.009
S0092-8674(14)00201-3 [pii] (2014).

3 Lee, T. I. & Young, R. A. Transcriptional regulation and its misregulation in disease. *Cell* **152**, 1237-1251, doi:10.1016/j.cell.2013.02.014
S0092-8674(13)00203-1 [pii] (2013).

4 Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91-100, doi:10.1038/nature11245 (2012).

5 Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116-120, doi:10.1038/nature11243 (2012).

6 Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134-1148, doi:10.1016/j.cell.2013.04.022
S0092-8674(13)00464-9 [pii] (2013).

7 Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

8 Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

9 Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43-49, doi:10.1038/nature09906 (2011).

10 Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75-82, doi:10.1038/nature11232 (2012).

11 Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337-343, doi:10.1038/nature13835 (2015).

12 Gjoneska, E. *et al.* Conserved epigenomic signals in mice and humans reveal immune basis of Alzheimer's disease. *Nature* **518**, 365-369, doi:10.1038/nature14252 (2015).

13 Canver, M. C. *et al.* BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-197, doi:10.1038/nature15521 (2015).

14 Korkmaz, G. *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* **34**, 192-198, doi:10.1038/nbt.3450 (2016).

15 Rajagopal, N. *et al.* High-throughput mapping of regulatory DNA. *Nat Biotechnol* **34**, 167-174, doi:10.1038/nbt.3468 (2016).

16 Diao, Y. *et al.* A new class of temporarily phenotypic enhancers identified by CRISPR/Cas9-mediated genetic screening. *Genome Res* **26**, 397-405, doi:10.1101/gr.197152.115 (2016).

17 Sanjana, N. E. *et al.* High-resolution interrogation of functional elements in the noncoding genome. *Science* **353**, 1545-1549, doi:10.1126/science.aaf7613 (2016).

18 Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science*, doi:10.1126/science.aag2445 (2016).

19 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).

20 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740, doi:10.1099/mic.0.023960-0 (2009).
21 Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67, doi:10.1038/nature13011 (2014).
22 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).
23 Meng, Z. *et al.* Berbamine inhibits the growth of liver cancer cells and cancer-initiating cells by targeting Ca(2)(+)/calmodulin-dependent protein kinase II. *Mol Cancer Ther* **12**, 2067-2077, doi:10.1158/1535-7163.MCT-13-0314 (2013).