A saturating mutagenesis CRISPR-Cas9–mediated functional genomic screen identifies cis- and trans-regulatory elements of Oct4 in murine ESCs

Matthew C. Canver1,‡, Pratibha Tripathi2,3,4, Michael J. Bullen2,3,4, Moshe Olshansky5,4, Yogesh Kumar2,3, Lee H. Wong6, Stephen J. Turner4,7, Samuel Lessard8,9, Luca Pinello10, Stuart H. Orkin1,11,* and Partha Pratim Das2,3,‡

From the 1Division of Hematology/Oncology, Boston Children’s Hospital and Department of Pediatric Oncology, Dana–Farber Cancer Institute (DFCI), Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts, USA, the 2Department of Anatomy and Developmental Biology, Monash University, Clayton, Victoria, Australia, the 3Development and Stem Cells Program, the 4Department of Microbiology, and the 5Department of Biochemistry and Molecular Biology, Cancer Program, Monash Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia, the 6Computational Biology and Bioinformatics, Baker Heart and Diabetes Institute, Melbourne, Victoria, Australia, the 7Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria, Australia, the 8Research Center, Montreal Heart Institute, Montréal, Quebec, Canada, the 9Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal, Quebec, Canada, and the 10Molecular Pathology and Cancer Center, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA, and the 11Howard Hughes Medical Institute, Boston, Massachusetts, USA

Edited by John M. Denu

Regulatory elements (REs) consist of enhancers and promoters that occupy a significant portion of the noncoding genome and control gene expression programs either in cis or in trans. Putative REs have been identified largely based on their regulatory features (co-occupancy of ESC-specific transcription factors, enhancer histone marks, and DNase hypersensitivity) in mouse embryonic stem cells (mESCs). However, less has been established regarding their regulatory functions in their native context. We deployed cis- and trans-regulatory elements scanning through saturating mutagenesis (ctSCAN-SMS) to target elements within the ~12–kb cis-region (cis-REs; CREs) of the Oct4 gene locus, as well as genome-wide 2,613 high-confidence trans-REs (TREs), in mESCs. ctSCAN-SMS identified 10 CREs and 12 TREs as novel candidate REs of the Oct4 gene in mESCs. Furthermore, deletions of these candidate REs confirmed that the majority of the REs are functionally active, and CREs are more active than TREs in controlling Oct4 gene expression. A subset of active CREs and TREs physically interact with the Oct4 promoter to varying degrees; specifically, a greater number of active CREs, compared with active TREs, physically interact with the Oct4 promoter. Moreover, comparative genomics analysis reveals that a greater number of active CREs than active TREs are evolutionarily conserved between mice and primates, including humans. Taken together, our study demonstrates the reliability and robustness of ctSCAN-SMS screening to identify critical REs and investigate their roles in the regulation of transcriptional output of a target gene (in this case Oct4) in their native context.

Large-scale genomic studies reveal that ~80% of the human genome may be involved in gene regulation, whereas only ~2% of the genome encodes proteins (1). The functional noncoding genome can be broadly divided into regulatory elements (REs) and regions that encode noncoding RNAs (1, 2). Furthermore, REs can be subdivided into cis-REs (CREs) and trans-REs (TREs), based on their position relative to their target gene(s). CREs are present proximally or distally relative to their target gene on the same chromosome, whereas TREs are located distally relative to their target gene on different chromosomes (3, 4). Putative REs have been identified using various methods, including transcription factor binding, enhancer histone marks, DNA accessibility (open chromatin regions), enhancer-promoter interactions, and gene expression (1, 5–10). REs are usually enriched with sequence variants that are associated with diverse human traits and diseases (11–13). In addition, REs play crucial roles in evolutionary turnover and divergence (14–17).

Initial efforts systematically evaluated RE functions using reporter assays on a massive scale (18, 19); however, such approaches fail to interrogate RE functions within their native genomic contexts. Advances in CRISPR-Cas9–mediated genome editing technology (20, 21) have transformed analysis of protein-coding genes (22, 23) as well as REs in situ in chromatin. A few high-throughput CRISPR-Cas9–mediated functional genetic screens were performed to characterize CREs in mammalian cells (24–29). Prior screens to identify functional CREs were focused on either targeting putative CREs of gene(s) of interest (gene-centric) or targeting putative CREs bound by selected TFs (TF-centric). Nevertheless, identification of functional TREs presents a challenge that has attracted less attention.

Here, we deployed genome-wide cis- and trans-regulatory elements scanning through saturating mutagenesis and sequencing (ctSCAN-SMS) in mouse embryonic stem cells (mESCs) to

This article contains supporting information.
‡ These authors contributed equally to this work.
* For correspondence: Stuart H. Orkin, stuart_orkin@dfci.harvard.edu; Partha Pratim Das, partha.das@monash.edu.
Present address for Samuel Lessard: Rare Blood Disorders, Sanofi, Waltham, Massachusetts, USA.
Editors' Pick: ctSCAN-SMS identifies regulatory elements of Oct4

identify critical CREs and TREs of the Pou5f1/Oct4 gene (a master pluripotency regulator of mESCs). We uncovered new functionally active CREs and TREs and how they regulate Pou5f1/Oct4 gene expression in mESCs.

Results

Design of a saturating CRISPR-Cas9 pooled library for ctSCAN-SMS

In mESCs, several putative REs, including 8,563 enhancers (ENs) and 231 superenhancers (SEs) have been identified based on co-occupancy of ESC-specific TFs (OCT4, NANOG, SOX2, KLF4, and ESRRB), mediators (MED1), enhancer histone marks (H3K4me1 and H3K27ac), and DNase I hypersensitivity (30). SEs contain multiple ENs; SEs are also more densely co-occupied with TFs, enhancer histone marks, and chromatin regulators, as compared with ENs, and associate with greater transcriptional output (30). We undertook a high-throughput CRISPR-Cas9-mediated genome-editing approach to comprehensively target putative REs in mESCs. First, we generated a genome-wide map of open chromatin regions using ATAC-Seq in mESCs, as ATAC-Seq identifies most EN REs (10). ATAC-Seq peaks were then overlaid within all putative ENs (8,563) and SEs (231) to designate “high-confidence” REs (2,613) (Fig. 1a, Fig. S1a, and Table S1). As these REs are distributed genome-wide and on different chromosomes relative to the Oct4 gene locus (in trans), we termed these REs TREs. All possible single guide RNAs (sgRNAs) (20 nt) were designed (within the TREs for tilling upstream of the Streptococcus pyogenes Cas9 NGG-protopsacer-adjacent motif (PAM) sequences) to target the high-confidence 2,613 TREs (Fig. 1a, Fig. S1a, and Tables S1 and S2). This analysis yielded 70,480 sgRNAs with a median gap of 5 bp between adjacent genomic cleavages (Fig. 1, c and d). Likewise, 1,827 sgRNAs were designed at the surrounding −12-kb (−10 to +2 kb of the TSS of the Oct4) region of the mouse Oct4 gene locus to dissect the CREs of Oct4 (Fig. 1, b and c). In addition, the library included 2,000 nontargeting (NT) sgRNAs as negative controls, 119 sgRNAs targeting GFP (of the Oct4-GFP reporter that used for the screen), and 150 sgRNAs targeting coding sequence of mESC-TFs as positive controls (Fig. 1c). In total, the RE CRISPR-Cas9 pooled library contained 74,576 sgRNAs (Fig. 1c). These sgRNAs were synthesized, pooled, and cloned into a lentiviral vector for deep sequencing. The deep sequencing of the pooled library confirmed the presence of >95% sgRNAs that target TREs and >99% sgRNAs that target CREs and control sgRNAs (Fig. S1b–f and Table S2).

Candidate CREs and TREs of the Oct4 gene identified by ctSCAN-SMS

The pooled library was transduced into an Oct4-GFP reporter mESC line, which constitutively expresses Cas9 (31–33). The Oct4-GFP reporter was used as a “readout” for the screen to measure the reduction in GFP levels upon perturbation of any targeted RE regions by their corresponding sgRNAs. Lentiviral transduction of the pooled library was performed at low multiplicity (MOI) to ensure that each cell contained predominantly one sgRNA (Fig. S2a). After puromycin drug selection (because sgRNA constructs carry the puromycin drug resistance gene), “GFP-low” cells were sorted using FACS (Fig. S2a and b). As a control, cells were collected before FACS (the “pre-sort” sample). Genomic DNA was isolated from both “GFP-low” and “pre-sort” cell populations, and next-generation sequencing was employed to enumerate the sgRNAs in each population (Fig. S2a). The screen was performed in triplicate.

We calculated an “enrichment score” of each sgRNA by comparing its frequency (presence) in GFP-low over pre-sort cells. Enrichment scores were determined based on the two best replicates (Table S2). The highest and lowest enrichment scores obtained from GFP-targeting sgRNAs (mean log2FC 4.87, p < 0.0001) and NT sgRNAs (mean log2FC 0.44, p < 0.0001) indicated that the screen was technically successful (Fig. 2a and 3a). We ranked all sgRNAs based on their enrichment scores (Table S2) and analyzed their off-target scores (ranging between 0 and 100) (34) (Figs. S3a and S4a and Table S2). A higher off-target score signifies fewer off-targets of an sgRNA. We found that the majority of the evaluated sgRNAs (87.6% sgRNAs for CREs and 84.5% sgRNAs for TREs) had off-target scores >10 (Table S2); these sgRNAs are statistically significant based on the hidden Markov model (HMM) analysis (Fig. S3a).

To identify candidate CREs of the Oct4, we considered all HMM sgRNAs with off-target scores >10 and mapped them within the −12-kb surrounding region (−10 to +2 kb of TSS) of the Oct4 gene locus. This yielded 16 candidate CREs (numbered 1–16), based on the mean enrichment score (mean log2FC) of sgRNAs per CRE. Each of the candidate CREs had a mean log2FC >0.5 (range of mean log2FC 0.93–3.95), with p < 0.0001, which was higher than the mean enrichment score of NT sgRNAs (mean log2FC 0.44, with p < 0.0001) (Fig. 2a and b). Among 16 CREs, CREs 10 and 12 have been recognized previously as distal and proximal enhancers, respectively (31); CREs 13–16 were present within the promoter region of Oct4 (±2 kb of TSS) (31). The remaining 10 CREs were newly identified candidate CREs of the Oct4 gene (Fig. 2a and b). Of note, CRE 12 displayed a relatively lower mean enrichment score (log2FC 0.77, p < 0.0001) compared with other candidate CREs, because the original Oct4-GFP reporter lacks a portion of CRE 12 (33). Nonetheless, we included CRE 12 for further validation.

To classify the candidate TREs, we applied a statistically relevant HMM analysis to the sgRNA enrichment scores (35), which initially identified 263 candidate TREs. Furthermore, we applied stringent criteria to select candidate TREs for validation, as follows: (i) TREs must have HMM sgRNAs with off-target scores >10 (Fig. 3a and Fig. S4a); (ii) TREs must possess at least four sgRNAs with mean log2FC >0.5 (range of mean log2FC 0.86–2.52), with p < 0.0001 (Fig. 3b); (iii) TREs must co-occupy with ESC-TFs (OCT4, NANOG, and SOX2) and enhancer histone marks (H3K27ac and H3K4me1) (Fig. 3c and Fig. S4b) and (iv) they must contain “dynamic” open chromatin regions (i.e. “open” chromatin regions present in the undifferentiated state (0 h) that gradually become “closed” with differentiation (24 and 96 h) of mESCs) (Fig. 3d and Fig. S4c). Based on these criteria, we selected 12 candidate TREs of the Oct4 gene (Fig. 3a and b).
Dissection of functionally active CREs and TREs of the Oct4 gene

We selected a total of 33 REs, including 20 CREs (16 candidate CREs and 4 control CREs (CREs with no significant enrichment scores of sgRNAs)) and 13 TREs (12 candidate TREs and one control TRE (with no significant enrichment scores of sgRNAs)) of the Oct4 gene locus for validation using WT mESCs. “Paired” sgRNAs (5’ and 3’ sgRNAs) tagged with mCherry were used to target both of the flanking ends of each selected candidate RE to create a deletion. Following transfection of paired sgRNAs-mCherry into the WT mESCs, mCherry-positive cells were sorted to measure the endogenous Oct4 mRNA expression and to confirm the genomic deletions of CREs/TREs (Figs. 2f and 3e, Figs. S3d and S4e, and Table S3). We observed a significant reduction in Oct4 expression to different extents upon deletions of CREs 1, 3, 5, 7, 10, 12, and 13–16 (of 16 candidate CREs) (Fig. 2f). Deletions of newly identified CREs 1, 3, 5, and 7 showed a greater reduction in Oct4 expression than the other deletions.
EDITORS' PICK: ctSCAN-SMS identifies regulatory elements of Oct4

List of candidate CREs of the Oct4 gene

| Candidate CREs | Genomic locations of RESs (mm8) | Length (bp) | No. of sgRNAs |
|---------------|---------------------------------|-------------|---------------|
| CRE_1         | chr17:35,632,944-35,633,217     | 273         | 51            |
| CRE_2         | chr17:35,634,773-35,635,108     | 335         | 40            |
| CRE_3         | chr17:35,635,706-35,636,370     | 664         | 98            |
| CRE_4         | chr17:35,636,568-35,636,884     | 316         | 35            |
| CRE_5         | chr17:35,637,687-35,637,998     | 311         | 53            |
| CRE_6         | chr17:35,638,316-35,638,683     | 347         | 45            |
| CRE_7         | chr17:35,639,981-35,639,479     | 496         | 73            |
| CRE_8         | chr17:35,639,516-35,640,048     | 532         | 60            |
| CRE_9         | chr17:35,640,274-35,640,811     | 537         | 56            |
| CRE_10        | chr17:35,640,811-35,641,342     | 531         | 97            |
| CRE_11        | chr17:35,641,411-35,641,715     | 304         | 50            |
| CRE_12        | chr17:35,641,862-35,642,698     | 834         | 187           |
| CRE_13        | chr17:35,642,726-35,643,342     | 616         | 140           |
| CRE_14        | chr17:35,643,343-35,644,057     | 714         | 113           |
| CRE_15        | chr17:35,644,097-35,644,345     | 248         | 35            |
| CRE_16        | chr17:35,644,380-35,644,880     | 500         | 70            |

Control CRE_1 (C1) chr17:35,631,818-35,631,951 133 0
Control CRE_2 (C2) chr17:35,632,289-35,632,557 268 0
Control CRE_3 (C3) chr17:35,633,280-35,633,418 133 0
Control CRE_4 (C4) chr17:35,633,801-35,633,934 133 0

sgRNAs targeting GFP (positive controls) NA NA 118
Non-targeting sgRNAs (negative controls) NA NA 1984
expression, compared with deletions of known distal and proximal enhancers (CREs 10 and 12) of Oct4 (Fig. 2f). However, “regulatory features,” such as co-occupancy of ESC-TFs (Oct4, NANOG, and SOX2–ONs), enhancer histone marks (H3K27ac and H3K4me1), and dynamic open chromatin regions (ATAC-Seq peaks at 0 h compared with 24, 96 h), were more prominent at CREs 7, 10, and 12 compared with CREs 1, 3, and 5 (Fig. 2c–e and Fig. S3b and c). Moreover, deletions of CREs 13–16 (present at the promoter region of Oct4) showed a significant reduction in Oct4 expression, as expected (Fig. 2f). Nonetheless, only CREs 13 and 14 showed substantial co-occupancy of ONS, H3K27ac, and dynamic open chromatin regions, as compared with other CREs present at the Oct4 promoter (Fig. 2c–e and Fig. S3b and c). In contrast, deletions of control CREs (C1–C4) led to no significant changes in Oct4 expression (Fig. 2f). Additionally, control CREs exhibited low-level co-occupancy of ONS, H3K27ac, H3K4me1, and dynamic open chromatin regions (Fig. 2d and e) and Fig. S3b and c). These data confirm the existence of multiple “active” CREs (active REs depict REs that reduce the Oct4 expression upon their deletions), including newly identified active CREs of the Oct4. Yet some active CREs fail to display recognized regulatory features (i.e., without of any co-occupancy of TFs, enhancer histone marks, and open chromatin regions), as described recently (26, 28), suggesting that a subset of functionally active REs may lack typical regulatory features.

Deletions of eight TREs (TREs 1, 2, 3, 4, 7, 9, 11, and 12) of 12 candidates showed a significant reduction in Oct4 expression to various extents (Fig. 3e). Nonetheless, all 12 TREs were co-occupied with ONS, H3K27ac, H3K4me1 enhancer marks and exhibited dynamic open chromatin regions (Fig. 3c and d) and Fig. S4b and c), as all of the candidate TREs (TREs 1–12) were short-listed for validation based on their regulatory features. Conversely, control TREs lack typical regulatory features (Fig. 3c and d), and deletion did not affect Oct4 expression (Fig. 3e). Moreover, neighboring genes of most of the validated active TREs were lowly expressed in mESCs (Fig. S4d), indicating that these genes may not be critical for mESC state maintenance. Importantly, we observed that the reduction of Oct4 expression was greater for the majority of active CREs than active TREs upon their deletions (Figs. 2f and 3e). This suggests that CREs generally contribute more than TREs to control overall Oct4 gene expression. Taken together, these data demonstrate the reliability and robustness of the ctSCAN-SMS screen to identify new candidate REs of the Oct4 in a high-throughput manner. Moreover, we confirm that a subset of these newly identified candidate CREs and TREs are functionally active REs of the Oct4 gene, and CREs are more active compared with TREs.

Cis- and trans-regulation of the Oct4 gene expression

REs (particularly enhancers) physically interact with gene promoters and control transcription (36). Chromosome conformation capture–based methods—4C, Hi-C, capture Hi-C, ChiA-PET, and HiChiP—have been utilized to identify physical contacts between promoters and REs (enhancers) (4). To interrogate the potential mechanisms by which candidate CREs and TREs regulate Oct4 gene expression, we examined interactions between REs (CREs and TREs) and the Oct4 promoter using published 4C-Seq data. These data were generated to study intrachromosomal and interchromosomal interactions between REs and the Oct4 promoter at a genome-wide scale (37). We used Oct4-234 (a region at ~1.5 kb upstream of the TSS of Oct4) as a “viewpoint,” as described previously (Fig. S5a) (37). Next, “contact frequencies” were calculated between the viewpoint and CREs (using a 1-kb resolution window, surrounding the 30-kb region of the Oct4 gene locus) (Fig. 4a and c), as well as between the viewpoint and TREs (using a 50-kb resolution window, surrounding each of the TREs) (Fig. 4b). This analysis revealed ranges of contact frequencies between functionally validated active CREs/TREs and the Oct4 promoter (Figs. 2f, 3e, and 4a and b). For example, functionally validated active CREs, including CREs 3, 5, and 7 (newly identified CREs), CREs 10 and 12 (previously known as distal and proximal enhancers of Oct4), and CREs 13–16 (residing at the promoter region of Oct4) demonstrated intrachromosomal interactions with the Oct4 promoter (Fig. 4a and c). Established enhancers, such as CREs 10 and 12, showed higher contact frequencies compared with newly identified CREs 3, 5, and 7 (Fig. 4a and c). However, active CRE 1 and control CREs (C1–C4) did not show significant interactions with the Oct4 promoter (Fig. 4a and c). In contrast, among functionally validated active TREs 1, 2, 3, 4, 7, 9, 11, and 12, only a minority (TREs 2, 3, 4, and 11) displayed higher interchromosomal interactions (compared with control TRE) with the Oct4 promoter (Fig. 4b).

Likewise, a high-resolution Micro-C (micrococcal nuclease–based Hi-C assay that captures genome-wide 3D chromatin organization/contact frequencies at single-nucleosome resolution (~100–200 bp)) from mESCs (38), displayed ranges of intrachromosomal interactions between active CREs and the Oct4 promoter (Fig. S5, b and c). Further, these data evaluated all other intrachromosomal interactions between any two genomic loci around the ~20-kb region of the Oct4 gene locus at 200 bp resolution (Fig. S5b). In addition, interchromosomal interactions were also observed between a few active TREs.
EDITORS' PICK: ctSCAN-SMS identifies regulatory elements of Oct4

List of candidate TRES of the Oct4 gene

Based on HMM based enrichment scores of sgRNAs.
All the sgRNAs must have off-target score >10.
TRES must possess at least 4 sgRNAs with mean enrichment scores (mean log2FC) >0.5, P-values <0.0001.
TRES should co-occur with ESC-TFs (OCT4, NANOG, SOX2), enhancer histone marks (H3K27ac, H3K4me1) and dynamic open chromatin regions.

| Candidate TRES | Genomic locations | Length (bp) | No. of sgRNAs |
|---------------|------------------|-------------|---------------|
| TRE_1         | chr: 14:59,989,540-59,989,621 | 81          | 20            |
| TRE_2         | chr: 4:71,807,347-71,807,416 | 69          | 12            |
| TRE_3         | chr: 9:74,996,403-74,996,575 | 172         | 21            |
| TRE_4         | chr: 6:9,978,361-13,978,560 | 199         | 20            |
| TRE_5         | chr: 4:13,023,503-13,023,629 | 126         | 11            |
| TRE_6         | chr: 19:53,534,741-53,534,849 | 108         | 19            |
| TRE_7         | chr: X:130,208,023-130,208,140 | 117         | 14            |
| TRE_8         | chr: 12:42,828,460-42,828,512 | 52          | 5             |
| TRE_9         | chr: 18:82,062,626-82,062,657 | 31          | 6             |
| TRE_10        | chr: 11:83,347,393-83,347,468 | 75          | 11            |
| TRE_11        | chr: 1:7,722,856-7,723,033 | 177         | 10            |
| TRE_12        | chr: 10:7,365,884-7,366,087 | 203         | 25            |
| TRE_control   | chr: 1:120,957,901-120,958,247 | 346         | 0             |
| sgRNAs_GFP (positive controls) | NA | NA | 118 |
| sgRNAs_non-targeting (negative controls) | NA | NA | 1984 |

Relative intensity

Relative intensity

Relative OCT4 mRNA expression
(TREs 4 and 11) and the Oct4 promoter (Fig. S5d). Both 4C-Seq and Micro-C data showed greater contact frequencies of active CREs compared with active TREs with the Oct4 promoter, and a greater number of active CREs compared with active TREs physically interacted with the Oct4 promoter at a 1-kb resolution window. Bottom triangle, heat map of normalized contact frequencies between 1-kb bins represented with the color codes (the highest is 1 with red; the lowest is 0 with white). In the top part, the black line (within the gray region) represents the normalized median contact frequencies between a locus and the viewpoint. The gray region exhibits the 20th to 80th percentile of the normalized contact frequencies.

Conserved functionally active CREs and TREs of the Oct4 gene

Recent studies demonstrate that the majority of species-specific REs/ENs evolved de novo from ancestral DNA regulatory sequences (16, 39). Also, evidence implies that loss or gain of REs (called RE turnover) takes place during evolution (17). To understand the importance of validated active CREs and TREs of the mouse Oct4 gene in evolutionary turnover, we analyzed their regulatory sequence conservation. Active CREs 3 and 5, CREs 10 and 12 (known distal and proximal ENs), and CREs 13–16 (present within the promoter) demonstrated significant conservation between mice and primates, including humans (Fig. 5a), whereas active CREs 1 and 7 did not show appreciable sequence conservation between mice and primates (Fig. 5a). Compared with active CREs, only a small subset of active TREs (TREs 3 and 4) showed significant sequence conservation between mice and primates (Fig. 5b and c). These observations suggest that active CREs are more critical than active TREs for Oct4 expression in mice and primates (including humans) during evolution.

**Figure 3. Identification and validation of active TREs of the Oct4 gene.** a, dot plot analysis displays the enrichment score of each sgRNAs by comparing its frequency in the GFP-low cells with that in the pre-sort cells at the selected candidate 12 TREs. b, list of identified candidate TREs of the Oct4 gene. c, binding dynamics of OCT4, NANOG, SOX2, H3K27ac, and H3K4me1 represented within the TREs of Oct4. d, ATAC-Seq demonstrates the changes in open chromatin regions/ENs from undifferentiated to differentiated mESC state (0, 8, 24, and 96 h) within the TREs. e, quantitative RT-PCR data illustrate the mRNA expression changes of endogenous Oct4 upon deletion of individual TREs of Oct4. Oct4 mRNA levels were normalized to GAPDH. Data are represented as mean ± S.E. (error bars) (n = 3); p values were calculated using analysis of variance. *, p < 0.05; **, p < 0.01; ***, p < 0.001; ns, nonsignificant.

**Figure 4. Physical interactions between CREs, TREs, and the Oct4 promoter in Oct4 gene regulation.** a, 4C-Seq data represent normalized interaction frequencies between CREs and the Oct4 promoter. The interaction/contact frequencies between CREs and the Oct4 promoter were measured at a 1-kb resolution window. b, interaction/contact frequencies between TREs and the Oct4 promoter quantified at a 50-kb resolution window. c, contact profile of CREs and the Oct4 promoter at a 1-kb resolution window. Bottom triangle, heat map of normalized contact frequencies between 1-kb bins represented with the color codes (the highest is 1 with red; the lowest is 0 with white). In the top part, the black line (within the gray region) represents the normalized median contact frequencies between a locus and the viewpoint. The gray region exhibits the 20th to 80th percentile of the normalized contact frequencies.
Next, we analyzed regulatory sequence conservation of previously identified high-confidence CREs (−449, −571, and −694) of human OCT4. These CREs are located distally between ~450 and 700 kb upstream of the human OCT4 TSS and physically interact with the OCT4 gene (28). We found that these human CREs are also evolutionarily conserved at the upstream regions of the mouse Oct4 locus (Fig. S6).

Taken together, our comprehensive comparative genomic analysis shows that a larger subset of active CREs compared with active TREs of OCT4 are well-conserved between mice and primates, including humans. These observations support the existence of both conserved and nonconserved REs of Oct4 among mice and primates (including humans), which is consistent with “RE turnover/divergence” (gain and loss of REs) at the OCT4 locus during evolution that may be critical for positive selection, as proposed earlier (17, 39).

**Discussion**

A handful of modest scale CRISPR-Cas9–mediated functional screens (using hundreds to thousands of sgRNAs) have been performed to target specific noncoding CREs of a gene(s) of interest (24–26, 29, 40). These screens successfully identified functional CREs of the target gene(s). In the context of identification of CREs of the OCT4 gene, a previous CRISPR-Cas9–mediated screen was performed to target 174 candidate CREs of OCT4 within its 1-MB topological associated domain in human ESCs; it revealed four temporary CREs and two known proximal CREs. The temporary CREs showed “transient
enhancer regulatory activity in OCT4 gene expression (27). However, the functional relevance of these temporary CREs is uncertain in human OCT4 gene regulation. Furthermore, another CRISPR-Cas9–mediated screen was performed by the same group using a different strategy, called CREST-Seq. This method employed 11,570 paired sgRNAs to introduce deletions to target 2 Mb surrounding the OCT4 locus in human ESCs, which created 2-kb deletions on average with an overlap of 1.9 kb between two adjacent deletions. This screen identified a total of 45 CREs, of which 17 CREs (with regulatory features) reside at the promoters of “unrelated” genes (intrachromosomally) that act as typical enhancers of the OCT4 gene (28). Our study employed ctSCAN-SMS—an unbiased, high-resolution, high-throughput screening approach using 1.827 sgRNAs to target CREs and 70,480 sgRNAs to target TREs of the mouse Oct4 gene (Fig. 1).

Previous CRISPR-Cas9 screens were focused on CREs of the target gene(s). In contrast, our screen was designed to identify both CREs and TREs. Indeed, we discovered 16 CREs (including 10 novel CREs) and 12 novel TREs of the murine Oct4 gene, as potential candidate REs (Figs. 2a and 3a). Furthermore, deletion studies confirmed that the majority of these CREs (10 of 16) and TREs (8 of 12) are functionally active in controlling the Oct4 expression. Overall, CREs are more functionally active than TREs (Figs. 2f and 3e). In addition, we showed that a subset of these active CREs and TREs physically interacts with the Oct4 promoter to different extents through intra- and interchromosomal interactions, respectively (Fig. 4 and Fig. S5). Notably, a greater number of active CREs, compared with active TREs, physically interact with the Oct4 promoter. Moreover, the interactions between active CREs (compared with active TREs) and the Oct4 promoter are more prominent (Fig. 4 and Fig. S5). Nonetheless, “enhancer activities” of several active CREs/TREs are not directly correlated to their physical contact frequencies with the Oct4 promoter (Figs. 2f, 3e, and 4 and Fig. S5). Interestingly, we found that several active CREs (CREs 1, 3, and 5) lack typical regulatory features (Figs. 2d–f and 6); this supports earlier studies that found that unmarked REs (UREs) without typical regulatory features may play critical roles in transcriptional output (26, 28). Moreover, comparative genomics analysis revealed that numerous active CREs and active TREs (a greater number of active CREs than active TREs) of Oct4 are evolutionarily conserved between mice and primates (including humans) (Fig. 5). However, we also observed divergence of Oct4 REs among mice and primates (including humans), which may account for the vital roles of “RE turnover” during evolution (3, 39).

Several studies demonstrate that “multiple REs” act either in a cooperative or competitive fashion to control transcriptional output (39). Our study identified multiple active REs of Oct4 and revealed a spectrum of regulatory activities of “individual” CREs and TREs in Oct4 gene expression (Figs. 2f and 3e). Further systematic studies will be required to elucidate how multiple REs function combinatorially to control Oct4 transcriptional output. In conclusion, we have demonstrated the utility of ctSCAN-SMS as an approach to identify functional REs of a gene locus and dissect their regulatory contributions to the transcriptional output of a target gene within its normal chromosomal context.

Experimental procedures

mESCs

mESCs were cultured in mouse ESC medium that contains Dulbecco’s modified Eagle’s medium (Thermo Fisher Scientific) supplemented with 15% fetal calf serum (Merck Millipore), 0.1 mM β-mercaptoethanol (Sigma–Aldrich), 2 mM l-glutamine (Thermo Fisher Scientific), 0.1 mM nonessential amino acid (Thermo Fisher Scientific), 1% of nucleoside mix (Merck Millipore), 1000 units/ml recombinant leukemia inhibitory factor (LIF/ESGRO) (Merck Millipore), and 50 units/ml penicillin/
**Editors’ Pick: ctSCAN-SMS identifies regulatory elements of Oct4**

strepotinycin (Thermo Fisher Scientific), as described previously (32, 41). mESCs were cultured at 37 °C, 5% CO₂.

**Mouse RE CRISPR-Cas9 pooled library design for the ctSCAN-SMS**

For this study, we selected a list of putative 8,563 EN and 231 SE REs from the mESCs, as described previously (30). First, we generated ATAC-Seq (10) data from WT mESCs and mapped within all of the putative EN and SE REs to identify open chromatin regions as well as high-confidence REs. Next, ±100 bp (200 bp) around the center of the ATAC-Seq peaks were obtained from the high-confidence REs. In total, we identified 2,613 REs for targeting. All possible sgRNAs (20 nt) were designed upstream of the S. pyogenes Cas9 NGG-PAM sequences at these defined REs (2,613), which created 70,480 sgRNAs with a median gap of 5 bp between adjacent genomic cleavages. Because these EN and SE REs were distributed in trans of the Oct4 gene locus, we called these REs trans-REs (TREs). Likewise, 1,827 sgRNAs were designed prior to all possible NGG-PAM sequences at the adjacent 12-kb (−10 to +2 kb of the TSS of Oct4) region of the mouse Oct4 gene locus to systematically dissect the REs of Oct4. As these REs reside adjacent to the Oct4 gene locus, they are called cis-REs (CREs). We also included 2,000 NT sgRNAs as negative controls and 119 sgRNAs targeting GFP of the Oct4-GFP reporter and 150 sgRNAs targeting coding sequence of mESC-TFs as positive controls. Altogether, the RE CRISPR-Cas9 pooled library contained a total of 74,576 sgRNAs.

**RE CRISPR-Cas9 pooled library construction for the ctSCAN-SMS**

All of the sgRNA oligonucleotides of the library were synthesized as described previously (32) using a B3 synthesizer (CustomArray, Inc.), pooled together, PCR-amplified and cloned into Esp3I-digested plentiGuide-Puro (Addgene plasmid ID: 52963) lentiviral vector, using a Gibson assembly master mix (New England Biolabs). Gibson assembly products were transformed into electrocompetent cells (E. cloni, Lucigen) and plated on 245 × 245-mm square LB-agar plates to obtain the sufficient number of bacterial colonies at a ~50× library coverage. Bacterial colonies were collected from the plates, genomic DNA was isolated, and plasmid libraries were prepared for high-throughput sequencing to confirm the representation of individual sgRNAs in the RE CRISPR-Cas9 pooled library.

**Lentiviral library production**

HEK293T cells were seeded onto 15-cm dishes ~24 h prior to transfection. Cells were transfected at 80% confluence in 16 ml of medium with 8.75 μg of VSVG, 16.25 μg of pSPAX2, and 25 μg of the RE CRISPR-Cas9 pooled lentiviral plasmids, using 150 μg of linear polyethyleneimine (Sigma–Aldrich). Medium was changed with fresh medium, 16–24 h after transfection. Lentiviral supernatant was collected at 48 and 72 h post-transfection and subsequently concentrated by ultracentrifugation (24,000 rpm, 4 °C, 2 h) (Beckman Coulter SW32).

**CRISPR-Cas9–mediated ctSCAN-SMS in mESCs**

Oct4-GFP reporter mESCs with stably expressed Cas9 were transduced with an RE CRISPR-Cas9 pooled lentiviral library at low MOI to avoid more than one lentiviral integration per cell. Test transductions were performed to estimate the viral titration and transduction rate. Briefly, 300,000 Oct4-GFP + Cas9 mESCs were plated per well of a 12-well plate. After 24 h, different amounts (1, 2, 4, 6, and 8 μl) of the lentiviral library were added to the cells. 10 μg/ml blasticidin (InvivoGen) and 1 μg/ml puromycin (Sigma–Aldrich) were added 24 h after the transduction to select for lentiviral library integrants (puromycin-resistant) in cells with Cas9 (blasticidin-resistant). Cells were selected for the next 3–4 days. The same number of cells were seeded as a control but not infected with lentiviral library and not treated with blasticidin and puromycin. The number of blasticidin- and puromycin-resistant cells and control cells were counted to calculate the viral titer and transduction rate (to achieve 30%).

For the actual screen, we seeded ~112 million (~75,000 sgRNAs in the pooled library, with 500× coverage, for a 30% transduction rate) Oct4-GFP + Cas9 mESCs in the same format (i.e. 300,000 cells/well of the 12-well plate) for each independent screening replicate. The lentiviral library was added to each well of a 12-well plate to achieve a 30% transduction rate with low MOI (MOI 0.1) to make sure each infected cell obtained one viral particle. 24 h post-transduction, fresh mESC medium was added to the cells with 10 μg/ml blasticidin (Invivogen) and 1 μg/ml puromycin (Sigma–Aldrich) and selected for 4 days. These selected cells were used to sort the “GFP-low” cells. The “pre-sort” cells were collected before sorting and used as a control. Genomic DNA was isolated from both the “GFP-low” and “pre-sort” cell populations, and libraries were prepared for deep sequencing to enumerate the presence of sgRNAs in these cell populations as described previously (32). The screening was performed in biological triplicates.

**Deletion of CREs and TREs of Oct4 in mESCs**

Paired sgRNAs (5’ and 3’ sgRNAs) were designed to target both the ends of each selected candidate RE to create a deletion. Both the sgRNAs were cloned into lentiGuide-plasmid that carries Cas9 and mCherry, using the Golden Gate cloning approach as described previously (32). 1,000,000 WT (J1) mESCs were transfected with 1 μg of each 5’ and 3’ sgRNA-Cas9-mCherry plasmids using Lipofectamine 2000 (Thermo Fisher Scientific). After 24 h of transfection, mCherry-positive cells were sorted; at least 50,000–100,000 mCherry-positive cells were collected either to isolate the total RNA for measuring the endogenous Oct4 mRNA expression levels by quantitative RT-PCR or to perform the genotyping PCR for checking deletions of targeted REs.

**Flow cytometry**

Cells were dissociated using trypsin and washed with 1× PBS, followed by sorting. (i) During the CRISPR-Cas9 screening, Oct4-GFP reporter mESCs were sorted based on GFP-low intensity after the transduction with the RE CRISPR-Cas9 pooled library; (ii) to quantify the Oct4 mRNA expressions...
upon deletions of CREs and TREs, their targeting sgRNAs-mCherry–positive cells were sorted.

**RNA isolation and RT-qPCR**

DNA-free total RNA was isolated from mESCs using the RNeasy Mini Kit (Qiagen), and cDNA was prepared using the iScript cDNA synthesis kit (Bio-Rad). RT-qPCR was performed using iQ SYBR Green Supermix (Bio-Rad) on the Bio-Rad iCycler RT-PCR detection system.

**ctSCAN-SMS screen data analysis**

sgRNA sequences present in the GFP-low and pre-sort pools were enumerated. Enrichment was determined by the log₂ transformation of the median number of occurrences of a particular sgRNA in the GFP-low pool divided by the median number of occurrences of the same sgRNA in the pre-sort pool across the best two biological screen replicates.

**ATAC-Seq experiment and data analysis**

ATAC-Seq was performed according to the previously described protocol (10), with some modifications. Briefly, each ATAC-Seq experiment and data analysis across the best two biological screen replicates.

**Micro-C data analysis**

Raw Micro-C data from mESCs was downloaded from GEO (GSE130275) (38). Samples were processed with a standard juicer pipeline (45), using the mm9 mouse genome. Reads from all 37 samples were merged together (after removal of duplicates). This resulted in more than 4 billion contacts, allowing the creation of a Hi-C file at a very high (200-bp) resolution. For CRE contacts, the chr17 intrachromosomal contact map was created using an in-house tool to produce normalized observed/expected counts. These counts were used for the creation of a heat map/matrix and bar plot. For each CRE (including Oct4 TSS), all 200-bp bins overlapping the loci were used. For TRE contacts, a whole-genome 1-kb resolution contact map was created and balanced using our in-house C++ code (this is the GW_SCALE normalization available in juicer_tools_1.22.01.jar). These counts were used for the creation of the bar plot. In this case, the Oct4 promoter was defined as ±3 kb of TSS, and TREs were used 200 kb bin centered at the particular TRE. The heat map was generated using a heatmap2 R function from the gplots package, and bar plots were generated using a bar plot R function.

**Data availability**

All of the high-throughput sequencing data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number GSE140911. ChiP-Seq and RNA-Seq data used are from GSE113335 (32) and GSE43231 (41).

**Acknowledgments**

We thank the Monash University FACS core facility. We thank Xiaofeng Wang for Illumina HiSeq2500 high-throughput sequencing at Harvard Medical School as well as the Genewiz high-throughput sequencing facility, China.

**Author contributions**

M. C. C., P. T., Y. K., L. P., and P. P. D. data curation; M. C. C., Y. K., and L. P. software; M. C. C., M. O., Y. K., L. P., and P. P. D. formal analysis; M. C. C., P. T., M. J. B., L. H. W., S. H. O., and P. P. D. investigation; M. C. C., M. O., Y. K., and L. P. visualization; M. C. C., S. H. O., and P. P. D. writing-original draft; M. C. C., P. T., S. H. O., and P. P. D. writing-review and editing; P. T. and M. J. B. validation; P. T., M. J. B., L. P., and P. P. D. methodology; S. J. T., S. L., and L. P. resources; S. J. T., S. H. O., and P. P. D. project administration; S. H. O. and P. P. D. conceptualization; S. H. O. and P. P. D. supervision; S. H. O. and P. P. D. funding acquisition.

**Funding and additional information**

This work was supported by National Health and Medical Research Council (NHMRC) of Australia Grants APP1159461 and APP1182804 (to P. P. D.) and National Human Genome Research Institute (NHGRI) Career 12 M. C. Canver, P. Tripathi, M. J. Bullen, M. Olshansky, Y. Kumar, L. H. Wong, S. J. Turner, S. Lessard, L. Pinello, S. H. Orkin, and P. P. Das, manuscript in preparation.
Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: RE, regulatory element; CRE, cis-RE; TRE, trans-RE; TF, transcription factor; ctSCAN-SMS, cis- and trans-regulatory elements scanning through saturating mutagenesis and sequencing; ESC, embryonic stem cell; mESC, mouse ESC; EN, enhancer; SE, superenhancer; sRNA, single guide RNA; TSS, Transcription start site; MOI, multiplicity of infection; FC, -fold change; H3K27ac, histone H3 Lys-27 acetylation; H3K4me1, histone H3 Lys-4 monomethylation; HMM, hidden Markov model; GAPDH, glyceraldehyde-3-phosphate dehydrogenase.

References

1. ENCODE Project Consortium, (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74
2. Cech, T. R., and Steitz, J. A. (2014) The noncoding RNA revolution—trashing old rules to forge new cells. Cell 157, 77–94 CrossRef Medline
3. Miele, A., and Dekker, J. (2008) Long-range chromosomal interactions and gene regulation. Mol. Biol. 40, 1046–1057 CrossRef Medline
4. Elkoon, R., and Agami, R. (2017) Characterization of noncoding regulatory DNA in the human genome. Nat. Biotechnol. 35, 732–746 CrossRef Medline
5. Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shokry, M., Wright, C., Chen, F., Afzaal, V., Ren, B., Rubin, E. M., and Pennacchio, L. A. (2009) ChIP-seq accurately predicts tissue-specific enhancer localization of common disease-associated variation in regulatory DNA. Science 326, 1097–1102 CrossRef Medline
6. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., and Greenleaf, W. J. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. 30, 271–277 CrossRef Medline
7. Siepel, A., and Arbiza, L. (2014) Cis-regulatory elements and human evolution. Curr. Opin. Genet. Dev. 29, 81–89 CrossRef Medline
8. Melnikov, A., Muragan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Frei, S., Gnirike, A., Callan, C. G., Kinney, J. B., Kellis, M., Lander, E. S., and Mikkelsen, T. S. (2012) Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. 30, 265–270 CrossRef Medline
9. Patwardhan, R. P., Hiatt, J. B., Witten, D. M., Kim, M. J., Smith, R. P., May, D., Lee, C., Andrie, J. M., Lee, S.-I., Cooper, G. M., Abitz, N., Pennacchio, L. A., and Shendure, J. (2012) Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. 30, 255–259 CrossRef Medline
10. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013) RNA-guided human genome engineering via Cas9. Science 339, 823–826 CrossRef Medline
11. Mojica, F. M., Diez-Villaseñor, C., Garcia-Martinez, J., and Almendros, C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (Reading) 155, 733–740 CrossRef Medline
12. Shalem, O., Sanjan, N. E., and Zhang, F. (2015) High-throughput functional genomics using CRISPR-Cas9. Nature 16, 299–311 CrossRef Medline
13. Wang, T., Wei, J. S., Sabatini, D. M., and Lander, E. S. (2014) Genetic screens in human cells using the CRISPR-Cas9 system. Science 343, 80–84 CrossRef Medline
14. Canver, M. C., Smith, E. C., Sher, F., Pinello, L., Sanjan, N. E., Shalem, O., Chen, D. D., Schupp, P. G., Vinjamur, D. S., Garcia, S. P., Luc, S., Kurita, R., Nakamura, Y., Fujiwara, Y., Maeda, T., et al. (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. Nature 527, 192–197 CrossRef Medline
15. Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Miyake, K., Zwart, W., Elkon, R., and Agami, R. (2016) Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat. Biotechnol. 34, 192–198 CrossRef Medline
16. Rajagopal, N., Srinivasan, S., Kooshesh, K., Guo, Y., Edwards, M. D., Banejee, B., Syed, T., Emons, B. J. M., Gifford, D. K., and Sherwood, R. I. (2016) High-throughput mapping of regulatory DNA. Nat. Biotechnol. 34, 167–174 CrossRef Medline
17. Diao, Y., Li, B., Meng, Z., Jung, I., Lee, A. Y., Dixon, J., Maliskova, L., Guan, K.-L., Shen, Y., and Ren, B. (2016) A new class of temporarily phenotypic enhancers identified by CRISPR-Cas9-mediated genetic screening. Genome Res. 26, 397–405 CrossRef Medline
18. Diao, Y., Fang, R., Li, B., Meng, Z., Yu, J., Qiu, Y., Lin, K. C., Huang, H., Liu, T., Marina, R. J., Jung, I., Shen, Y., Guan, K.-L., and Ren, B. (2017) A tiling-
deletion-based genetic screen for cis-regulatory element identification in mammalian cells. Nat. Methods 14, 629–635 CrossRef Medline
29. Sanjana, N. E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A., and Zhang, F. (2016) High-resolution interrogation of functional elements in the noncoding genome. Science 353, 1545–1549 CrossRef Medline
30. Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., and Young, R. A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 CrossRef Medline
31. Yeom, Y. I., Fuhrmann, G., Ovitt, C. E., Ohbo, K., Gross, M., Hübner, K., and Schöler, H. R. (1996) Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. Development 122, 881–894 Medline
32. Seruggia, D., Tripathi, P., Canver, M. C., LeBlanc, L., Di Giammarino, D. C., Bullen, M. J., Nefzger, C. M., Sun, Y. B. Y., Farouni, R., Polo, J. M., Pinello, L., Apostolou, E., Kim, J., Orkin, S. H., et al. (2019) TAF5L and TAF6L maintain self-renewal of embryonic stem cells via the MYC regulatory network. Mol. Cell 74, 1148–1163.e7 CrossRef Medline
33. Szabó, P. E., Hübscher, K., Schöler, H., and Mann, J. R. (2002) Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. Mech. Dev. 115, 157–160 CrossRef Medline
34. Hsieh, T.-H. S., Cattoglio, C., Slobodyanyuk, E., Hansen, A. S., Rando, O. J., Hsieh, M. H., Rahl, P. B., Lee, T. I., and Young, R. A. (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. Cell Syst 3, 76–88 CrossRef Medline