THE INTERNATIONAL JOURNAL OF DEVELOPMENTAL BIOLOGY
www.intjdevbiol.com

doi: 10.1387/ijdb.160108jb

SUPPLEMENTARY MATERIAL

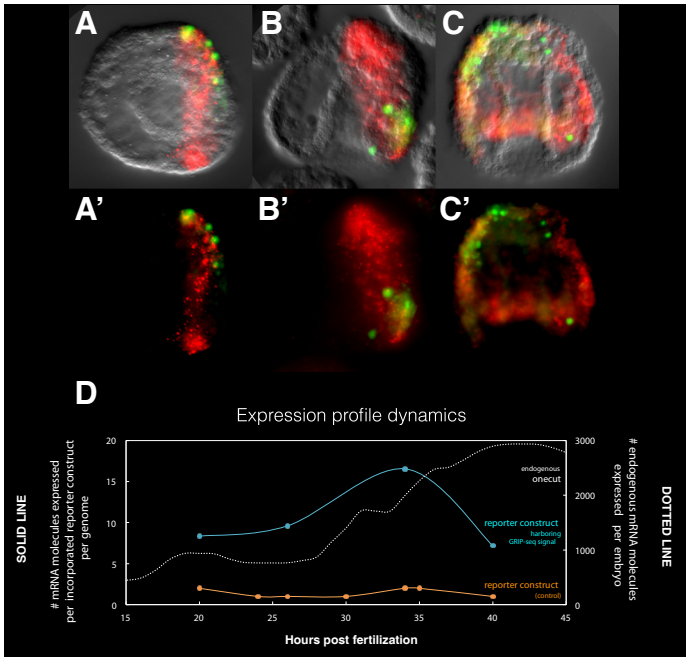
corresponding to:

Genome-wide identification of enhancer elements

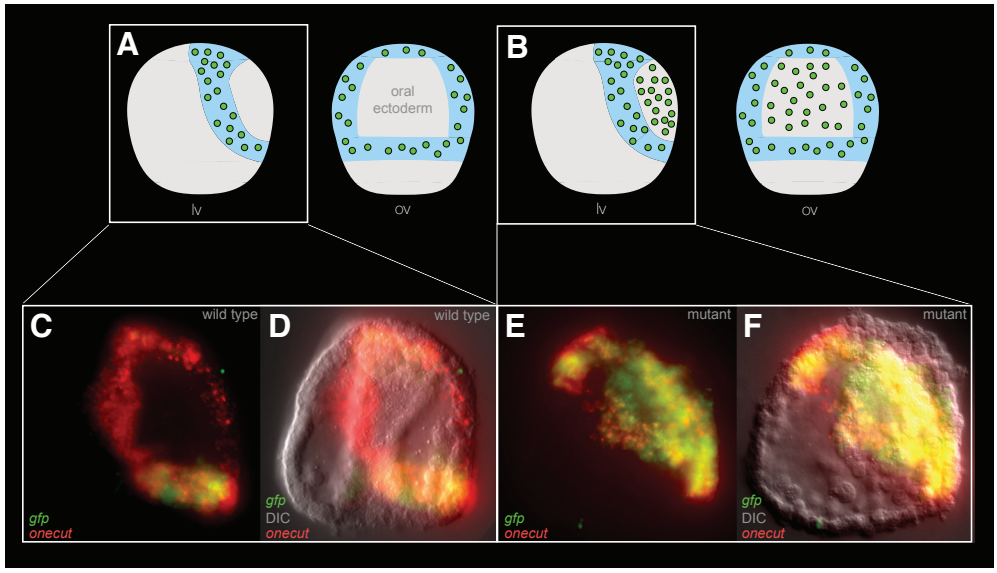
SARAH TULIN*, JULIUS C. BARSI*, CARLO BOCCONCELLI* and JOEL SMITH

***Address correspondence to:** Julius C. Barsi. Division of Biology and Biological Engineering, California Institute of Technology, 1200 East California Boulevard, Mail Code 156-29, Pasadena, CA 91125, USA. Tel: +54-929-4469-0388. E-mail: barsi@caltech.edu - Web: http://www.its.caltech.edu/~barsi/

Full text for this paper is available at: http://dx.doi.org/10.1387/ijdb.160108jb



Suppl. Fig. S1. Qualitative and quantitative expression profile of the *onecut* GFP-reporter from Fig. 3. (A-C') *Double fluorescent RNA* in situ *hybridization of a 35 hpf transgenic embryo with endogenous* onecut *expression in red and exogenous* gfp *expression in green.* **(A-C)** *Double fluorescent images overlaid onto a DIC image to orient the signal relative to the embryo.* **(A'-C')** *The fluorescent signal without DIC.* **(A,A')** *Apical views of the embryo* **(B,B')**. *Lateral views of the embryo.* **(C,C')** *Oral views of the embryo.* **(A-C')** *Amended from (Barsi, Davidson, 2016).* **(D)** *The quantitative expression profile of the* onecut *intronic CRE GRP reporter construct described in Fig. 3 was determined using qPCR with primers against GFP (blue line), RFP (Ctrl, orange line), and endogenous* onecut *(white line). The expression of the reporter construct recapitulates the profile observed for endogenous* onecut *across all developmental stages assessed. This qualitative* **(A-C)** *and quantitative* **(D)** *information further supports the conclusions derived from the* in situs *of Fig. 3.*



Suppl. Fig. S2. *In situ* experimental assay showing ectopic expression in mutated distal *onecut* CRE. (A) *Schematic embryo showing GFP expression (green dots) overlapping endogenous* onecut *expression in the ciliated band (shaded light blue).* **(B)** *Schematic embryo showing ectopic GFP expression (green dots) spreading throughout the oral face, beyond the endogenous* onecut *expression when Gsc sites in the distal reporter are mutated.* **(C-F)** *Genuine data on which the above schematics are based, amended from a previously published Fig. (Barsi, Davidson, 2016).* **(C,D)** *Double fluorescent RNA* in situ *hybridization of 35 hpf transgenic embryos carrying the reporter construct described in Fig. 5F.* **(D)** *Fluorescent signal from (C) superimposed onto a DIC microphotograph of the embryo.* **(E,F)** *Double fluorescent RNA* in situ *hybridization of 35 hpf transgenic embryos carrying the same reporter construct, but with Gsc sites mutated.* **(F)** *Fluorescent signal from (E) superimposed onto a DIC microphotograph of the embryo. Abbreviations: lv, lateral view; ov, oral view.*

ChIA-PET using Pol II (Goh et al. 2012)	GRIP-seq
Fix with both EGS and formaldehyde, 45', 20'	Fix with only formaldehyde, 10'
Cell lysis	Cell lysis
Nuclear lysis	No nuclear lysis
sonicate chromatin, Branson sonicator, 9'	sonicate chromatin, Covaris, 130sec
centrifuge	centrifuge
Pre-clear chromatin with protein G beads	no pre-clearing
Pol II antibody added to beads, o/n at 4° C	antibody not pre-bound to beads
Antibody coated beads added to sheared chromatin, o/n at 4° C	Pol II antibody added to beads and chromatin, o/n at 4° C
repair ends "on bead"	repair ends with NEB kit "on bead"
ligate biotinylated half-linkers "on bead"	no tags
Phosphorylate half-linker-ligated beads	N/A
circularize, o/n 16° C	ligate ends with NEB kit "on bead", 16hr, 16° C
reverse crosslinks, ProtK, 2hr, 50° C	reverse crosslinks, ProtK, 16hr, 65° C
Phenol chloroform extract and precipitate	no phenol chloroform extraction
release PET-DNA molecules with addition on SAM and MmeI enzyme, 2 hrs 37° C	N/A
addition of Streptavidin beads	N/A
N/A	linear DNA removed using plasmid safe DNase
N/A	circularized DNA molecules amplified using Rolling Circle Amplification with phiX DNA polymerase
Ligate 454 Adapters, o/n at 16° C	make library using Nextera kit
Nick translate ChIA-PET DNA, with DNA pol I o/n at RT	N/A
16-20 cycles of PCR amplification	N/A
precipitation of DNA	N/A
gel extract 233bp band, perform "gel crush"	Pippin prep to gel extract 450-550bp

Suppl. Fig. S3. A comparison of ChIA-PET and GRIP-seq method steps. *A table comparing the major steps in ChIA-PET and GRIP-seq methods.*

Supplemental 4: README

This is a guide to using the files in the database.

There are files from three samples:
1. GRIP-seq experimental no. 1, ChIP'd with anti-Pol(II), sequences named N702-
2. GRIP-seq experimental no. 2, ChIP'd with anti-Pol(II), sequences named N705-
3. GRIP-seq control, ChIP'd with IgG alone, sequences named N703-

For each sample there are 5 files:

1. raw read files, zipped fastqc files, all R1 reads in one file
2. raw read files, zipped fastqc files, all matching R2 files in a separate file
3. Bowtie-mapped reads, converted to a .bam file
4. an index of the mapped reads, .bai file
5. a read count file, 25bp window, .tdf file

To get from the raw reads to the mapped reads, follow the following steps

1. Filter the reads for quality using the instructions in Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. Genome Biol. 2011

> quality_passed.R1.fastqc file
> quality_passed.R2.fastqc file

2. Using Bowtie2, map the quality-passed paired R1 and R2 fastqc files onto the S. Purpuratus genome, version 3.1, which can be downloaded from:
http://www.spbase.org/SpBase/download/

2a. Before using Bowtie2 you have to make an index of the genome you are aligning to. This gives the genome coordinates an index so the reads can then have the same index. The command is below, it creates a set of 6 files.

> bowtie2-build genome_name.fasta index_prefix_name
> for example: bowtie2-build Sea_Urchin_genome.fasta Sea_Urchin_ref_index_bowtie

2b. Run Bowtie2 on each sample using the following command

> bowtie2 --phred33 --fr -I 100 -X 500 -p 8 --seed 123 -q -t -x
Sea_Urchin_ref_index_bowtie -1

3. After Bowtie alignments are made, the resulting sam files need to be
converted to bam files. They can then be loaded into a program like IGV for visualization

3a. this first command converts sam files to bam files using samtools

> samtools view -bS #filenamehere#.sam > #filenamehere#.bam

3b. this second command filters the hits for read that map well (mapq score > 10)

> samtools view -b -h -q 10 #filenamehere#.bam > #filenamehere#_filtered.bam

3c. this next command sorts the bam file

> samtools sort #filenamehere#_filtered.bam #filenamehere#_filtered_sorted

3d. this last command indexes the sorted bam file, giving it coordinates, no output name needed. It will create the .bai file

> samtools index #filenamehere#_filtered_sorted.bam

4. To make a read count file for visualization, one option is to use tools from a program like IGV from the Broad center: http://www.broadinstitute.org/igv/

4a. First, create a .genome file, using version 3.1 of the sea urchin genome and
the sea urchin transcriptome gene models: http://www.spbase.org/SpBase/download/
within the file labeled: gff3 files for 3.1 assembly (Build7)

4b. Load the .genome file. Load the sorted.bam file for all GRIP-seq samples

4c. In IGV, go to Tools > Run igvtools. Run "Count" with a window size of 25, or
adjust to a different window size. A .tdf file will be created. Load these count
file alongside the sorted.bam file.

5. To view reads as pairs, once the read files have been opened you can right click on the left hand panel of a read file and select "View as pairs". Other helpful sorting options are under "sort alignments by": (1) "sort by insert size" to see the read pairs that map the furthest apart, and (2) "sort by start location" to line up pairs based on the mapping position of the left most pair.

Peak Calling

Peak Calling was performed on the GRIP-seq files using MACS software, version 1.4.2 20120305, through the Galaxy Cistrome interface.

The parameters used were:
custom genome size= 814000000
p-value= .001
Keep duplicate tags at the exact same location?= Keep ALL
Use Model= TRUE
small fold enrichment for model building= 10
large fold= 30

Suppl. Fig. S4. A README file for computational methods. *This describes the files in the publicly available database and gives the commands used to generate each file. It also contains instructions for viewing the files in the Broad Institute's IGV viewer, including instructions for how to view reads "as pairs," as discussed in the Results section.*