Oct4-dependent lincRNA Expression in Human Embryonic Stem Cells

Lauren Gulland, Sarah Kinnings, John Whitaker, and Wei Wang

Department of Chemistry and Biochemistry, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0378

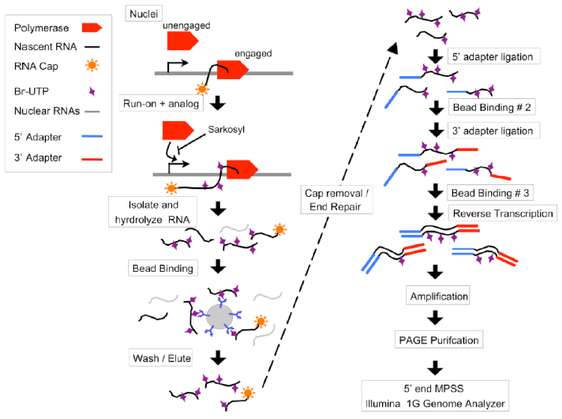
Overview

The purpose of our investigation was to determine whether Oct4 has a role in lincRNA expression in human embryonic stem cells. To test this, we used GRO-seq data from two different samples: wild type and knock down. The wild type sample represented normal human embryonic stem cells, while in the knockdown, Oct4, a well-known transcription factor, was knocked down. Using the GRO-seq data from the experimental lab, we began our process to test Oct4's effect on lincRNA expression in the genome.

An Introduction to GRO-Seq

Genome-wide Run-On Sequencing, more commonly referred to as GRO-seq, is a process that "maps the distribution of short transcripts generated by transcriptionally engaged RNA polymerases that are able to transcribe (run-on) a short distance and incorporate an affinity tag into the nascent RNAs. Sequencing of these RNAs and aligning to the genome provide a density and orientation map on mRNA-encoding genes of the transcriptionally competent Pol II. These Pol II include those on the body of genes that are caught in the process of transcriptional elongation, as well as those that accumulate promoter-proximal paused Pol II". In other words, GRO-seq allows us to detemine which regions of the genome are being actively transcribed at a particular point in time.

A diagram of the GRO-seq process.



Core *et al*. (2008) *Science*, 322, 1845-1848

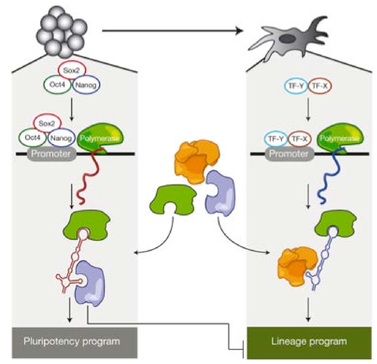
The statistics resulting from the read mapping performed using Bowtie and Perl, seen to the right

	Total GRO-seq reads	Total reads aligning uniquely to genome	Reads that failed to align	Reads that didn't align uniquely	Total uniquely aligned monoclonal reads
H1 wildtype	56,899,820	14,350,306 (25.22%)	21,383,985 (37.58%)	21,165,529 (37.20%)	4,531,016
H1 Oct4 KD	51,353,655	11,853,346 (23.08%)	14,930,542 (29.07%)	24,569,767 (47.84%)	3,207,630

lincRNAs and Oct4

Thousands of large intergenic non-coding RNAs, or lincRNAs, have been found in mammalian genomes, although we are not quite sure what their function is or why they are important. However, some tests have been done on the knockdown of lincRNAs in embryonic stem cells, and when performed, the outcome is similar to when major embryonic stem cell state regulators and transcription factors are knocked down.

Oct4 is the known transcription factor of the POU family and is majorly involved in the self-renewal of embryonic stem cells. Oct4 is tightly regulated, as any disruption may cause the unwanted differentiation of the embryonic stem cells.



TFs bind to promoters on lincRNAs

lincRNAs transcribed

lincRNAs bind to chromatin regulatory proteins

Identification of Oct4 Enhancer Target lincRNAs

Guttman, M. et al. (2011) *Nature* **477**, 295-300

Bowtie and Perl – matching and sorting

The GRO-seq reads received from the experimental lab were mapped to the their respective places on the human genome using a program called Bowtie. However, not all of the reads aligned to the genome. Utilizing the programming language Perl, only the uniquely aligned reads were extracted, and, of those, only the monoclonal reads were extracted. Although we started out with more than 50 million reads for each of the two samples (wild type and knock down), we ended up with only 4.5 million when all of the unsuitable reads were removed. These filtered reads were then the ones that used in our lincRNA analysis (see the table on the left).

Mapping data to genome
- Runs data through Bowtie
- Sorts useable reads from not unique or unmapped reads

Sorting and Filtering
- Formats and sorts the Bowtie output
- Deletes repeats

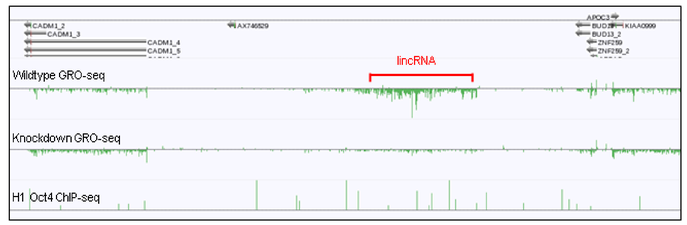
Matching to lincRNAs
- Filters through reads and matches to lincRNAs
- Prints which strand each read is on and totals them

A simple flowchart of the logic behind the execution of the Perl programs and when Bowtie was used.

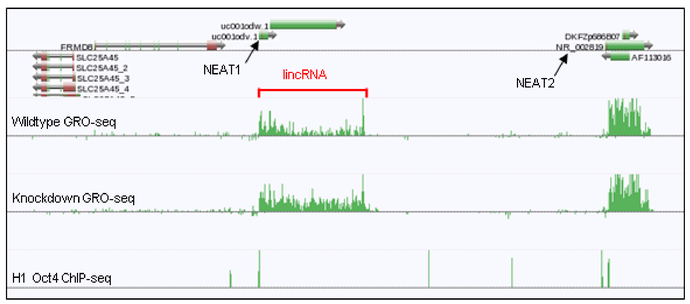
Results and Identification of Oct4 Regulated lincRNAs

When Oct4 is knocked down, the lincRNA's GRO-seq may increase or decrease. If it turns out that the lincRNA was being repressed by Oct4, then its GRO-seq levels after knockdown may spike as the stem cell differentiates. However, if the Oct4 was inducing transcription of the lincRNA, its wild type GRO-seq will be noticeably greater.

On the other hand, there are also lincRNAs that are not regulated by Oct4, and we will not see a change in their GRO-seq from wild type to knock down.



When Oct4 is knocked down, some lincRNA's GRO-seqs are significantly reduced. The wildtype lincRNA reads also overlap with Oct4 ChIP-seq peaks, confirming that Oct4 is regulating this lincRNA.



An example of lincRNA repression by Oct4: when knocked down, the lincRNA is over-expressed.

Conclusions

- Oct4, an embryonic stem cell state regulator, plays a crucial role in transcription regulation of some lincRNAs, while it is not involved at all with others.
- Oct4 can act as an enhancer or a repressor of lincRNA transcription.

Acknowledgements

Professor Wei Wang's laboratory at UCSD
Dr. Elizabeth Komives, Program Supervisor
Mr. and Mrs. Stuart Gulland

Citations

Min, Irene M., Joshua J. Waterfall, and Leighton J. Core. "Regulating RNA Polymerase Pausing and Transcription Elongation in Embryonic Stem Cells." *Regulating RNA Polymerase Pausing and Transcription Elongation in Embryonic Stem Cells. Cold Spring Harbor Laboratory Press, 18 May 2011. Web. 26 July 2012. <http://genesdev.cshlp.org/content/25/7/742.full.pdf html?sid=4a91e285-0299-46a2-851a-538e63344dca>.*