Yale University

MBB
266 Whitney Avenue
PO Box 208114
New Haven, CT 06520-8114

Telephone:
203 432 6105
360 838 7861 (fax)
mark@gersteinlab.org
GersteinLab.org

September 7, 2017

Dear Editor,

Please find enclosed our manuscript entitled "Comprehensive survey of LINE-1 transcriptional activity in human cell-lines, healthy somatic tissue and tumors", which we hope will be considered for publication in Nature Biotechnology. We hope that this new manuscript regarding the development of a method to gauge LINE-1 transcriptional activity in RNA-seq experiments and the landscape of LINE-1 in a vast number of human samples will be considered for publication.

Transposable elements are one of the most important sources of variation in the mammal genome but, until this moment, we couldn't assess their activity comprehensively. Due to their repetitive nature, we have been only able to assess their activity in a small number of samples at a time. One way of solving this problem is by leveraging the multitude of available RNA-Seq experiments. The transcriptome from different tissues and diseases can help us uncover the biology of LINE-1. We present a new method to gauge LINE-1 autonomous transcription based on regular RNA-seq. Using this uniform, unbiased and accurate method, we characterized thousands of RNA-seq experiments and describe the landscape of LINE-1 activity in normal healthy tissue. We also present a theory that suggests that LINE-1 activity is entangled with genome instability in tumors by creating double strand breaks and small insertions and deletions. Thus, we not only present a totally new solution to tackle the long-standing problem but also perform the largest and most comprehensive assessment of LINE-1 activity in human cells. Our major findings are:

1. LINE-1 creates small insertions and deletions in tumoral genomes.
2. LINE-1 is also active in healthy somatic tissue.
3. Cell turnover is correlated with the transcriptional activity of LINE-1s.
4. The majority of reads overlapping LINE-1 elements derive from pervasive transcription.

We are submitting our paper as a companion to the other manuscripts written in the ENCODE Consortium's 3rd phase, i.e., ENCODE3. Our paper does not depend heavily on the other ENCODE3 manuscripts. Nevertheless, it is very important because it is the only study that uses ENCODE data to analyze the transcriptional activity of Transposable Elements. We also would like to point out the importance of the ENCODE project as it supplies the freely accessible datasets for analysis of transposable elements activity in human cell lines.

We appreciate you taking the time to review and respond to our manuscript. Please address all correspondence concerning this manuscript to pi@gersteinlab.org. We attach below a list of our referees

Alysson Muotri (muotri@ucsd.edu , UC San Diego)
Peter Park (peter_park@harvard.edu , Harvard)
Cedric Feschotte (cf458@cornell.edu , Cornell University)
Precott L. Deininger (pdeinin@tulane.edu , Tulane University)

Yours sincerely,

Mark Gerstein
Albert L. Williams Professor
of Biomedical Informatics