MASSACHUSETTS GENERAL HOSPITAL
HARVARD MEDICAL SCHOOL

Martin Aryee
Assistant Professor, Department of Pathology, Massachusetts General Hospital & Harvard Medical School
Assistant Professor in the Department of Biostatistics, Harvard T.H. Chan School of Public Health
Associate Member, Broad Institute of Harvard and MIT

149 13th Street, 6th floor, Room 6016
Charlestown, Massachusetts 02129
Telephone: 617-726-5695
Email: aryee.martin@mgh.harvard.edu

Dear Editors,

Please find attached a manuscript entitled **diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data**, which we are submitting for your consideration.

As recent ground-breaking findings (Flavahan *et al.* 2016—*Nature*; Hnisz *et al.* 2016—*Science*) have implicated the role of three-dimensional structural changes in DNA leading to carcinogenesis, our group has been interested in understanding variability in the spatial arrangement of the genome and its impact on cellular phenotypes. In particular, we were interested in comparing different cell states to identify putative regions of differential topologies and characterize the transcriptomic and epigenomic corollaries associated with three-dimensional changes. To our knowledge, no previous software environment provided both rigorous statistical testing of differential DNA loop data and the necessary annotation function to assess the transcriptional and epigenetic correlates of variation in three-dimensional changes in DNA.

Thus, we present a set of computational methods and results using our novel R/Bioconductor software *diffloop* to facilitate analyses of DNA loop structures. We believe the following features of our software and analysis will be of interest to your readership:

- Statistically rigorous quality control and association testing to identify differential topological features.
- Facile integration of CNV, histone modification, transcription factor binding, RNA-Seq, chromatin accessibility, and DNA methylation data with topological data.
- Biological annotation of chromatin interactions.
- Rapid visualization of statistically significant differential loops and regions.
- Integration into the R/Bioconductor environment to allow interfacing with other workflows

In the manuscript, we perform a comparison of differential POL2 loops between the MCF-7 and K562 cancer cell lines, and link these DNA loops to differences in epigenetic and transcriptional

state. While this simple comparison enabled us to display the functional versatility of the package, we note that *diffloop* has been designed to handle more complex experimental design.

We anticipate that *diffloop* will be of considerable interest to the *Cell Systems* community and that our software meets the compelling need to identify differential DNA loops with statistical rigor. As such, we are eager for this software/manuscript to be considered as a "Tool." All code used to produce our results is readily available allowing other users to expand on our analyses and integrate this package into their existing workflows. All pertinent information and analyses are linked in the repository noted in the manuscript. The *diffloop* analysis pipeline can be executed on any modern computer that supports the R programming language (version 3.3.0 and newer) and should take less than 1 hour for users familiar with the R/Bioconductor ecosystem to execute.

Again, we anticipate that *diffloop* will be of considerable interest to the *Cell Systems* community. We eagerly await the feedback of the editorial board as well as any peer-reviewers.

Sincerely,

Martin Aryee