Yale University

MB&B
260/266 Whitney Avenue
PO Box 208114
New Haven, CT 06520-8114

Tel: 203 432 6105
Email: Mark.Gerstein@yale.edu

September 9, 2014

Dear Editor,

we are transferring our manuscript entitled "**Analysis of structural variation breakpoints from 1,092 humans reveals details of mutation mechanisms**" for consideration for publication in Nature Communications. In this manuscript we derived confident set of breakpoints for 8,943 structural variants detected in 1,092 human individuals sequenced by the 1000 Genomes Project. We further performed a comprehensive analysis of the breakpoints including comparison in the context of genome functional and epigenetic contents. For this purpose we used data generated by the ENCODE and by the NIH Roadmap Epigenomics Mapping Consortium. Our study provides insights into the mutational mechanism leading to formation of structural variations. In particular, we ~~hypothesize~~ that structural variations mediated by non-allelic homologous recombination could originate in early embryonic and germ cells without replication and then are passed on through the germline. Furthermore, for structural variants generated by template switching during replication we suggest existence of particular spatial and temporal configurations for DNA to generate a variant.

Our analysis is of importance not only for fundamental understanding of mutations in the genome but also for clinical use – in particular, for understanding genetic alteration associated with cancer, senescence, and aging. We, therefore, believe that our manuscript will be of great interest to a broad audience and, especially, to the readers of Nature Communications.

We have recently noticed the "Call for data analysis papers". We feel that out manuscript fits the announcement. We are kindly asking you to consider the manuscript for this call.

The manuscript has been evaluated by anonymous reviewers, and our understanding is that, you'll find the reviews as part of the transferred documents. In our opinion the reviewers questioned novelty of the study, confounding/technical factors in the analysis and asked for additional evidence/analyses. After reading the reports we could see why reviewers think this way and, we'd like to provide few comments in this regard. We are sorry that it was not entirely clear that our work is not simple re-analysis of the old data/breakpoint set, rather derivation of a new set with novel analyses. We can definitely clarify ~~unclear moments~~ and re-emphasize novel analysis. Similarly, we can address concerns about confounding factors, provide additional evidence and perform additional analyses requested by the reviewers. Finally, in several occasions, we believe, ~~that~~ the reviewers were confused or mis-interpreted the conclusions/inputs and we'd be glad to provide responses and additional explanations to such points. For example, ~~the~~ reviewer #2 states "In general the authors avoid difficult regions including limiting the dataset to variants greater than 1kb, …", while ~~apparently~~ roughly half of our data set are CNVs smaller that 1 kbp (Fig. 1C lower panel). Similarly, ~~the~~ reviewer #3 states "The study of Koren et al only reports replication timing in 100 kb intervals, …" and makes conclusion that our analysis of micro-insertions and replication timing for ranges of up to 1 kbps is meaningless. In fact, Koren data are of about 1 kbp resolution, and this allowed us to conduct and report novel analysis of association of micro-insertions with replication timing. In this brief note we are not providing point-by-point response to all the critiques but would be happy to do so if requested.

Yours sincerely,

Mark Gerstein
Albert L. Williams Professor of
Biomedical Informatics