RESPONSE LETTER

Referee 1

-- Ref1.1 – General comments --

Reviewer Comment	The authors have addressed all my concerns.
Author Response	We appreciate referee 1's comments.

Referee 2

-- Ref2.1 – General comments --

Reviewer Comment	The authors are right in stating that there is no other reference that implements a noncoding mutation burden analysis. The only one I know is Weinhold et al., (2014), and I also agree with the authors that a simple binomial test as applied in that reference is not good enough to correctly compute the mutation burden in noncoding regions. What I wonder is if the change from a binomial to a beta-binomial distribution is a good enough solution. Unfortunately, the controls provided in the new version don't seem to be enough to prove that, see comments below. Also, we have tried to run the software and we found many problems and unsatisfactory results in the only case we managed to run it (described below). Overall I agree with the authors that it would be an important contribution to describe and provide a method that does the noncoding mutation burden analysis correctly. I am not convinced that LARVA does it well, at least in its current version, based on our test on running the software (see below) and on the description in the manuscript.
Author Response	We thank the reviewer for these comments. We have addressed the reviewer's concerns over LARVA's false positive and false negative rate by testing LARVA against simulated variant datasets, indicating that LARVA does control both false positives and false negatives rigorously. Furthermore, we have addressed the software issues raised by the reviewer. We address each of these in a point-by-point format below.
Excerpt From Revised Manuscript	

-- Ref2.2 – False positive and false negative rate --

Reviewer Comment	AUTHOR'S RESPONSE We emphasize our contribution in the following listed points.

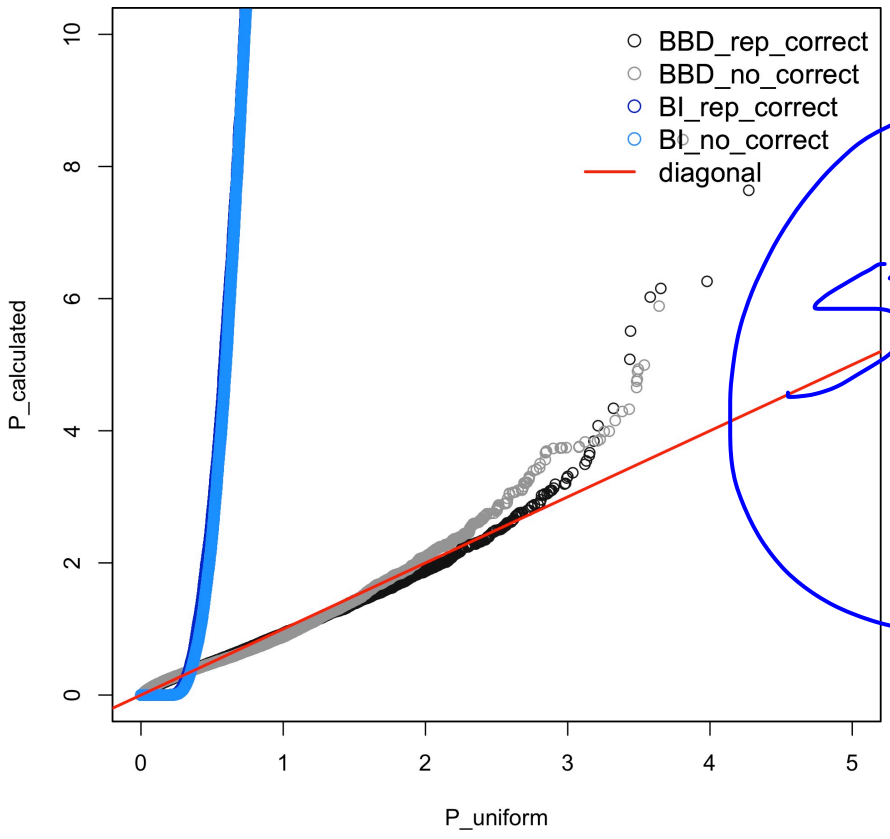
	1. We are among the first to implement the somatic burden test with overdispersion control, which is specifically designed for noncoding somatic variant analysis. MY NEW COMMENT I agree with that. It is important not only to be among the first but more importantly to make sure that the test is correct, give a good control of false positives and false negatives, and provide a code that users can run. AUTHOR'S RESPONSE 2. We release a convenient annotation resource for the whole community by gathering all the noncoding regulatory regions from more than 122 experiments from the ENCODE project. Notably, this data has never been collected in one place before, which will greatly facilitate subsequent research. 3. Our released noncoding regulatory element corpus provides a natural and meaningful solution about how to pool biologically relevant regions to perform the mutation burden test. We do not have to rely on the bin procedure, which is a relatively ad-hoc method. 4. Once highly mutated regions are detected in a certain cancer type, users can immediately understand the functions of these regions. MY NEW COMMENT I agree with authors that 2, 3 and 4 are useful additional resources provided with the code of LARVA, however the first and more important think is that authors convince that LARVA is able to detect noncoding recurrently mutated drivers, which I understand from the description of the paper it is the main aim, with an acceptable rate of false positives and false negatives. This is not clear in this version of the software and manuscript.
Author Response	We thank the reviewer for agreeing with our contribution. We would like to preface our response with an important caveat: due to our limited understanding of true noncoding cancer drivers, it is extremely difficult to accurately gauge false positive and false negative rates for noncoding cancer driver discovery methods at this time. Nevertheless, we have expanded our false positive and false negative rate analysis by testing LARVA on simulated variant datasets, demonstrating that **[tbd]**
Excerpt From Revised Manuscript	

-- Ref2.3 – P-values for all genes --

Reviewer Comment	Finding only 7 significantly mutated coding genes analyzing 5032 tumors is a surprising low number. I agree that 6759 significantly mutated genes with the binomial test is a not an acceptable number of genes, surely full

	of false positives. It would be useful if authors provide a supplementary table with the obtained pvalue per gene, not only for the 7 genes claimed as highly mutated by LARVA.
Author Response	We thank the reviewer for this comment. We have added a new table to our supplement that lists the p-values for all tested gene regions.
Excerpt From Revised Manuscript	

-- Ref2.4 – QQ plots --

Reviewer Comment	QQ plots should be in - log10 scale to be able to see in detail the most important part of the plot, which correspond to the significant regions. With the QQ plot provided it is not clear if the distribution of pvalues is correct. Authors could use this code for example: http://www.broadinstitute.org/files/shared/diabetes/scandinavs/qqplot.R
Author Response	We thank the reviewer for this comment. We have updated the QQ plots in our manuscript in accordance with these suggestions. As seen from figure, Binomial tests with/without replication timing correction severely deviate the theoretical P values distribution, but beta-binomial results follows the diagonal line except for several true signal points. 

Excerpt From Revised Manuscript	We have replaced the figure in Text S1.

-- Ref2.5 – Software errors I --

Reviewer Comment	Since I wasn't convinced myself of the validity of the method by reading the new version of the manuscript I thought the best would be to run the software ourself. We decided to run LARVA on a pancancer dataset retrieved from tumorportal (http://www.tumorportal.org/load/data/per_ttype_mafs/PanCan.maf). Unfortunately we were not able to get any results as the program halted the execution raising errors. We first tried to analyze the coding regions of the pancancer dataset. The program kept running for more than 100 hours (> 4 days) and eventually halted raising an R error. Error in if (any(mu <= 0) | any(mu >= 1)) stop(paste("mu must be between 0 and 1 ", : missing value where TRUE/FALSE needed Calls: pval_varying_length -> pBB Execution halted
Author Response	We thank the reviewer for bringing this to our attention. We have addressed the long running time by profiling our code, and optimizing the computations in portions of the code where the running time did not scale well with the size of the input. We have released revised code along with our revised manuscript. Furthermore, we have migrated our R codebase into C++, giving us more direct control over the source code. Our new code is not prone to the error the reviewer encountered.
Excerpt From Revised Manuscript	

-- Ref2.6 – Software Errors II --

Reviewer Comment	We next tried to run LARVA with a dataset of 505 tumor whole-genomes across 14 cancer types as reported in Fredriksson et al., 2014 in promoters and ultra-sensitive regions. For both promoters and ultra-sensitive regions we used the annotations present in the folder data/annotations/ of LARVA. The program didn't run successfully on promoters and raised an error after approx. 12 hours. Following is the trace of the error: Error in d$p.bbd.cor[d$p.bbd.cor <= 0] = rep(d$p.tiny, sum(d$p.bbd.cor <= : replacement has length zero Execution halted
Author Response	We thank the reviewer for bring this to our attention. We have determined that this error can occur in rare boundary conditions

	in our R code. We have migrated our R codebase into C++, and now have more direct control over the functioning of our code. Our new code handles these conditions properly.
Excerpt From Revised Manuscript	

-- Ref2.7 – Software P-value Output --

Reviewer Comment	We finally managed to run LARVA with this dataset in ultra-sensitive regions. In this case the program performed the analysis quickly. However, when we check the files with the results we found cases, with the exception of 'p.bbd.cor.adj', where the pvalues were greater than 1. How can this be possible? Following are the maximum values of each pvalue type: p.bbd 3.488000 p.binomial 16.425000 p.bbd.cor 3.286000 p.binomial.cor 20.596000 p.bbd.adj 1.148000 p.bbd.cor.adj 0.357000 p.binomial.adj 14.031000 p.binomial.cor.adj 17.464000
Author Response	We thank the reviewer for bringing this to our attention. We failed to mention in our software documentation that this numerical output is, in fact, the –log10-transformed p-value. This has been rectified in the revised version's documentation.
Excerpt From Revised Manuscript	

THUS THESE ARE CORRECT

-- Ref2.8 – Software P-value QQ Plots --

Reviewer Comment	After filtering the results for regions that overlapped genes or pseudogenes and for regions without mutations, we did QQplots as follow: we discarded pvalues > 1 (considering them wrong) and we plot on the y axis the -log10 of the sorted observed pvalues and on the x axis the -log10 of a uniform distribution of expected pvalues between 0 and 1. The QQplots were generated by using the code provided here:http://www.broadinstitute.org/files/shared/diabetes/scandinavs/qqplot.R) The resulting plots showed that the both the 'pbb' pvalues distributions (p.ddb and p.ddb.cor, top row of the figure) are deflated respect to a perfect correlation between observed and expected pvalues (red diagonal line) and thus the methods are finding less significant genes that what expected by the null model. On the other hand the binomial method (bottom row of the figure) is somehow inflated respect to the red diagonal. While the binomial method is likely to find a number of false positive candidates, the method proposed by the authors is likely to miss many true positive candidates.

Author Response	**[tbd]**
Excerpt From Revised Manuscript	