Referee #2 (Remarks to the Author):

Blue written by WM, after extensive discussion with PDM

Pre-Amble
This manuscript describes a bioinformatics resource that would cater functional annotations generated by the ENCODE projects to the needs of cancer genomics. The authors incorporated external cancer resources such as TCGA into the resource. The resource includes cancer-normal pairings of cell lines for the analysis; prediction of local mutation rate at the megabase-scale based on epigenomic features; a set of high confidence "compact" annotations of putative regulatory elements; extended gene annotations that include protein coding regions and cis-regulatory elements; analysis of "key regulators" based on expression data; and regulatory networks highlighting cancer-specific rewiring events. The manuscript does not report new datasets in addition to the ENCODE release and offers limited conceptual novelty. However, there is a possibility that the resource would be very popular among cancer genomics researchers.

Referee 2 provided helpful feedback on our statistical approach and description thereof. This feedback was an opportunity for us to critically reflect on our statistical choices and improve the clarity of our writing. In some cases, we maintained our original approach but with new justification of appropriateness, and in other cases we have modified our approach based on analyses we performed in response to Referee 2's comments. In the revised manuscript, we have taken care to implement or address each point raised by the reviewer. We feel that this has strengthened our paper, and we thank the reviewer for his or her attention and expertise.

In reading Referee 2's comments, we realized that we also need to clarify that we are part of the ENCODE data release. This means that our manuscript would be the first publication of many of the ENCODE data sets analyzed herein, listed in table R2.1.

2.1 The negative binomial regression (Gamma-Poisson mixture model) was introduced in Nik-Zainal et al. Nature 2016 and Marticorena et al., Cell 2017. Why was not this available method applied, and what is the benefit for the procedure used by the authors? Also, does Gamma-Poisson model fits data for most cancers well or is it just an approximation? One can use non-conjugate priors but this is probably beyond the scope of this work.

Response: There may be a point of confusion here. When constructing the BMR in the original manuscript, we did in fact use negative binomial regression (Gamma-Poisson), because of its suitability for handling overdispersed count data. The Gamma-Poisson model fit most cancer types well, with an R^2 >0.7 for 23/32 [MADE UP] cancers studied (Figure R2.1, below). The non-conjugate priors suggestion was an interesting suggestion we followed up with when preparing the revised manuscript [NOT YET!], using the method of Statistician et al, but it was found to afford comparable model fit, so we stuck with the simpler Gamma-Poisson model.
The referee's language here suggests that he or she believes we used some method other than the Gamma-Poisson model. Perhaps our PCA figures led the referee to believe that we used PCA to construct our BMR. In fact, our use of PCA in the original manuscript was meant only for demonstration purposes to show that informational content of the various epigenetic features. We have revised the manuscript to make this point more clear.

2.2 It seems that the Poisson model was not rejected for cancers with very low mutation counts (liquid tumors). Is this a power issue rather than the property of the mutation process?

Although it could certainly be the case that tumors with low mutation counts have a different underlying subclonal structure and mutational processes, this is not our intended implication. In the original manuscript we had assumed that this was a power issue, and favored the simpler Poisson model for prediction purposes where there was insufficient data to fit the parameters for a Gamma-Poisson predictive model. The referee's question inspired us to test this assumption. We downsampled [NOT YET!] highly mutated tumors to resemble the counts of liquid tumors and found that, indeed, low counts will not reject the Poisson model for any tumor

type. This leads us to believe that all tumor types have some degree of overdispersion, but that this is only detectable once a tumor reaches > 1000 mutations [MADE UP].

2.3 The approach with principle components used for the BMR estimation does not seem to work well. Starting with the second PC most components have roughly the same prediction power. One possibility is that higher principle components do not capture the additional signal and reflect noise in the data, and the correlation with mutation rate is due to an overfit of the NB regression (it is unclear whether it was analyzed with cross-validation). Another possibility is that the signal is spread over many components. In the latter case, this is not an optimal method choice.

*WM running out of time. Ref2 is right about PCA's weirdness here. Our cross-validation is a nice attempt to address ref's concerns about overfitting but it is not good enough. Need to cross-validate that the additional PC's matter more. Ref misunderstands point of PCs. They are for demonstration only, not as a pre-processing step for NB.

The referee's comments here were particularly helpful. Initially, we had assumed that the reason for the long tail of informative PC's was because .

2.4 I do not agree with the power analysis presented to support the idea of compact annotations. I understand that this is a toy analysis neglecting specific properties of mutation rate known for regulatory regions and also sequence context dependence of mutation rate. The larger issue is that the analysis assumes that ALL functional sites are within the compact annotation. In that case, power indeed would decrease with length. However, in case some of the functional sites are outside the compact annotation power would not decrease and is even likely to increase with the inclusion of additional sequence. Is there a justification for all functional sites to reside within compact annotations? Can this issue be explored? Some statistical tests incorporate weighting schemes.

*WM running out of time. Our response here is not good enough. Jing's upward sloping than downward sloping plot only AFFIRMS refs objections, does not say how we are ROBUST to their objection or intelligently handle it.
* The way to show robustness is to show the effects of uncertainty. How does compact gene fare power-wise on expectation if we lob off nucleotides with a 50% depletion of functional impact per nucleotide compared to core, with a 10% depletion of functional impact per nucleotide compare to core, with a 90% depletion …
* Let's use ref2's weighting scheme! Define annotation-credence scores as probabilities of functional affect for nucleteotide in and near an element. Count a 50%-uncertain-functional-effects nucleotide for 0.5 points in the numerator if mutation but only 0.5 points in the denominator. We need to do this analysis. Why not hit all of ref2's points?

2.5 Some of the QQ-plots in supplementary figures look problematic. Also, for some tumors with low count statistics QQ-plots are expected to always be deflated, so the interpretation of QQ-plots may be non-trivial.

[someone help]

2.6 The idea of extended genes and the use of multiple information sources to construct them is a strength of the paper. However, it is unclear whether the analysis takes into account complexities of the mutation model in regulatory regions. The influence of tri- or even penta-nucleotide context can be significant. Next, TF binding and nucleosome occupancy is known to interfere with the activity of DNA repair system. It would be great to see a formal analysis about how extended genes increase power of cancer driver discovery.

These were also very helpful suggestions, for which we thank the reviewer. We have now revised our approach to take into account pentanucleotide context [NOT YET!]. There is no proven method to account for TF binding on DNA repair, making it outside the scope of our paper, but we have now included a discussion about how this confound limits the interpretation of some of our results [NOT YET.] We appreciate the referee's comments that our extended gene concept and its power advantage is an important concept in our paper, and we now give it a formal treatment in the revised manuscript. We conclude that a doubling the length of a test region by merging

two equivalent segments maximally doubles the power [MADE UP] in the setting where both elements are subject to the same strength and direction of selection, although the effect trails off in certain conditions.

2.minor.1 I would not use the term “burden test”. This usage is slightly confusing because this term is commonly used in human genetics where it refers to a case-control test.

2.mior.2 Similarly, it is unclear what is meant by “deleterious SNVs” as the term is commonly used in human genetics in reference to germline variants under negative selection.