**Response to Reviewer #1**

**One of the major difficulties of this sort of work is genuinely showing a significant and demonstrable improvement over existing techniques and knowledge. The analyses undertaken are useful and provide reassurance that the methods are effective but do not provide a huge advance over the pre-existing studies. The authors are clearly already aware of this and state in the discussion that the major gains are likely to be made by the incorporation of epigenetic data. I feel that this should be demonstrated before the paper could be considered for publication in this journal. It is not immediately obvious that simply adding additional independent epigenetic features is sufficient to uncover interesting biology (there maybe complex dependencies between mutational signatures and epigenetic features). An empirical study that demonstrates that this simple model really applies in this context would be desirable to justify the extrapolation.**

An empirical study that demonstrates that this simple model really applies in this context would be desirable to justify the extrapolation. While we believe that incorporating epigenetic data, and other types of data, into mutation signatures is an important and exciting direction for future work, we also believe that a proper investigation of this will be a major undertaking and view it as a project in itself that is outside the scope of our current paper. We also believe that, even without this additional work, the combination of novel models, software and visualization tools that are included here already represent a substantive and important advance - one that will be of considerable interest to both methodological, and applied researchers, and perhaps, as reviewer 2 suggests, will dramatically change the way that mutational signatures are identified in practice.

**Response to Reviewer #3**

**If I had one request, I’d ask whether or not the authors have found any examples of mutation signatures that are clear in the full model, but really lose something important in the independence model. Such examples are not hard to find with sequence logos. They write on p.5 that they “likely do exist in practice”, but give no example.**

One very clear example is the Pol epsilon mutation signature (identified as “Signature 10” in the previous literature, Alexandrov et al., Nature, 2013), which puts strong probability masses on just two patterns (TpCpT > TpApT and TpCpG > TpTpG), and a slightly weaker mass to TpTpT > TpGpT (see the Figure R1 below). Since the features of this signature are not independent (the frequencies of the immediate 3’ base (T and G) are influenced by the substitution pattern (C>A and C>T, respectively)), our model has to use two signatures to represent this signature (Figure R2).

We have discussed in the manuscript that the strand specificity shown in TpCpT > TpApT pattern (see the left panel in Figure R2) is somewhat consistent with the previous research (Shinbrot et al., Genome Research, 2014). However, TpCpG > TpTpG did not show such strand specificity (see the right panel in Figure R2). Considering this biological difference, we suspect there may be a possibility that the “Signature 10” in fact consists of two distinct signatures. However, we do not have enough evidence to prove this at present, and we would like to explore this using the more abundant data in the future.



**Figure R1.** The pol epsilon signature (Signature 10) derived in the Alexandrov et al., Nature, 2013. The barplots are divided by 6 substitution pattern. In each division, 16 bars show joint probabilities of substitution pattern and 5' and 3' bases (ApNpA, ApNpC,

ApNpG, ApNpT, …, TpNpT).



**Figure R2.** The signatures corresponding to Pol epsilon obtained from the Colon-Rectum cancer data using our approach.