## <ID>REF1.1 – Positive comments on the resource releases

<TYPE>$$$NoveltyPos

<ASSIGN>

<PLAN>&&&AgreeFix

<STATUS>%%%100DONE

|  |  |
| --- | --- |
| RefereeComment | This manuscript describes how the ENCODE project data could be utilized to derive insights for cancer genome analysis. It has several examples to illustrate this point, e.g., how to better estimate background mutation rate in a cancer genome, how to modify gene annotation for finding mutation-enriched regions (e.g., by bundling enhancer regions to target genes using Hi-C/ChIA-PET), and describing the changes in regulatory networks in cancer.Obviously, the ENCODE project involves a great deal of planning and a lot of experimental work by many groups, and the overall aim of re-highlighting the ENCODE as a resource to cancer research seems worthwhile in general, perhaps even in a high-profile journal. |
| AuthorResponse | We thank the referee for this positive feedback. |

##

## <ID>REF1.2 – BMR: comparison with existing literature

<TYPE>$$$BMR,$$$Text

<ASSIGN>@@@JZ,@@@WM,@@@PDM

<PLAN>&&&OOS

<STATUS>%%%95DONE

|  |  |
| --- | --- |
| RefereeComment | Just to take the first application as an example, the problem of estimating background somatic mutation rate accurately in order to better identify cancer drivers has been studied extensively in the literature. One paper, “Mutational heterogeneity in cancer and the search for new cancer-associated genes” (Nature 2013), is cited in the current manuscript, but there are many others. For instance, Weinhold et al, 2014 (Genome-wide analysis of noncoding regulatory mutations in cancer, Nat Genetics), Araya et al, 2015 (Identification of significantly mutated regions across cancer types highlights a rich landscape of functional molecular alterations, Nat Genetics), and similar non-coding mutation identification papers all include steps to account for epigenetic features in their background rate calculation. |
| AuthorResponse |  |
| Excerpt 1.2-A (in main text) | Wait for main text |

## <ID>REF1.3 – BMR: Match

<TYPE>$$$BMR,$$$Text

<ASSIGN>@@@JZ,@@@WM

<PLAN>&&&DisagreeFix

<STATUS>%%%50DONE

|  |  |
| --- | --- |
| RefereeComment | Most large-scale cancer genome sequencing papers also have models at various levels sophistication, most of them including the issue of proper tissue-type matching. “matched” cell lines are better than unmatched or addition of more epigenetic features results in some improvement is almost trivial at this point. Which marks contribute to this is also not new. |
| AuthorResponse |  |
| Excerpt 1.3-A (main manuscript) |  |
| Excerpt 1.3-B (cross validation in supplement) | l for each cancer type and listed the performance as below. |

## <ID>REF1.4 – BMR: cell of origin features vs. many features

<TYPE>$$$BMR,$$$Calc

<ASSIGN>@@@JZ,@@@JL

<PLAN>&&&DisagreeFix,&&&More

<STATUS>%%%70DONE

|  |  |
| --- | --- |
| RefereeComment | Importantly, Polak et al, 2015 (Cell-of-origin chromatin organization shapes the mutational landscape of cancer, Nature) in fact show that cell-of-origin chromatin features are much stronger determinants of cancer mutations profiles than chromatin feature of matched cancer cell lines, and that cell type origin can be predicted from the mutational profile. |
| AuthorResponse |  |
| Excerpt 1.4-A(added to disc. sect) |  |

## <ID>REF1.5 – BMR: Tissues vs. Cell lines

<TYPE>$$$BMR,$$$Calc

<ASSIGN>@@@JZ,@@@JL

<PLAN>&&&DisagreeFix,&&&More