Dear Dr. Bahcall,

We would like to submit our manuscript entitled “ENCODEC: A large scale integrative resource from ENCODE for cancer research” for consideration for publication in *Nature*. This paper serves as a companion encyclopedia to the ENCODE project, with a specific focus on **C**ancer research. For these reasons, we have named this companion encyclopedia “ENCODEC”.

The ENCODE top-tier cell-lines are the deepest, most richly annotated human genomic samples in the world. In this companion encyclopedia, we hope to accomplish the following:

1. Make unique genomic information from ENCODE more accessible to the cancer research community.
2. Showcase how its data sets can be mined for cancer insights
3. Demonstrate how ENCODE cell lines can be used to more powerfully analyze clinical cancer genomic cohorts.

Around 70 percent of the ENCODE cell lines are cancerous, and top-tier ENCODE cell lines are enriched with signals from hundreds of experimental assays. We first organized these cell lines into tumor-normal pairings across five major cancer types. Through large-scale data integration, we created our ENCODEC resource, including 1) uniformly processed signal tracks; 2) accurate non-coding annotations; 3) precise regulatory region and gene linkages; 4) TF and RBP networks.

The integration of novel assays, such as STARR-seq, Hi-C, RAMPAGE, and eCLIP, means that the accuracy of the resource we release here extends far beyond any previous effort. Using this newly created resource, we were able to carefully explore the cancer mutational landscape, and prioritize key positions in the genome that are associated with tumorigenesis. Specifically, we:

• synthesized signals from hundreds of assays to form joint estimates of background mutation rate with unprecedented accuracy. We associated a comprehensive set of regulatory elements with coding regions to evaluate overall mutation burden, and discovered novel candidate cancer driver genes with significant prognostic value. We also prioritized key regulatory regions with higher than expected mutations.

• collected hundreds of ChIP-seq experiments to set up cell-type specific TF networks. We defined the notion of *rewiring index*, in which TF regulatory logic units change in relation to chromatin state, and in response to many mutations. We get a clear sense of what types of TF network changes occur in oncogenesis, and how these changes relate to known events in oncogenesis.

• built up generalized TF and RBP networks to identify key regulators that drive differential gene expression in tumor cells in 15 cancer types. For discovered key regulators such as MYC, we demonstrated how our network information could help reveal how MYC interacts with other regulators in multiple cancer types.

Altogether, we prioritized key regulators, high-impact regulatory regions, and deleterious SNVs that impact these regions. We accomplished this using a step-wise scheme, that considered regulatory power, hierarchies in networks, regulatory changes, mutational burden, and local context effects. We further validated our prioritizations using several small-scale assays. In particular, we confirmed that MYC and SUB1 up-regulate their targets in breast and lung cancers through knockdown experiments. We also validated enhancer activities through luciferase assays, and demonstrated the effect of selected SNVs in these regions by introducing mutation into wild-type sequences. Finally, we consolidated our annotations, pipelines, analysis results, and experimental validations into the comprehensive ENCODEC resource.

This work constitutes a powerful resource that fully exploits the richness of deep sequencing functional assays in top-tier ENCODE cell lines. We have linked our ENCODEC annotations to the main ENCODE encyclopedia resource for easy retrieval. Our work marks a path towards improved cancer genome interpretation through incorporation of high dimensional functional assays at large-scale. We believe this would be of great interest to the broad readership of Nature.

We would like to suggest the following investigators as potential referees:

* Matthieu Lupien
* Han Liang, MD Anderson
* Josh Stuart
* Ewan Birney (EBI)

Thank you for your consideration.

Sincerely yours,

XXX