Hi Orli,

Let me summarize some of the key points from our ENCODE and Cancer paper.

We integratde and processed a variety of assays from the ENCODE project to make a companion ***ENCODE*** encyclopedia as a useful resource for the **C**ancer community (***ENCODEC***).

Around 70 percent of the ENCODE cell lines are cancerous, with the top-tier ones enriched with hundreds of functional assays. Hence, we first organized these cell lines into tumor-normal pairings for five major cancer types and after large-scale data integration we created our ENCODEC resource, which provides: 1) uniformly processed signal tracks; 2) accurate non-coding annotations; 3) precise regulatory region and gene linkage; 4) TF and RBP regulatory networks. The accuracy of the resource we released here is far any previous effort before we integrated a serial of novel assays, such as Hi-C, STARR-seq, RAMPAGE, and eCLIP.

With such resource, we could better dissect the cancer mutational landscape and prioritize key positions on the genome that are associated with tumorigenesis. Specifically,

* We could synthase signals from hundreds of experimental assays to jointly estimate background mutation rate at an unprecedented accuracy. We reconcile comprehensive regulatory elements with coding regions to jointly pick up high-mutated genes from distributed mutational signals. As a result, we discovered novel candidate cancer drivers with significant prognostic value. We could use this scheme to prioritize key regulatory regions in the genome.
* We collected hundreds of ChIP-seq experiments to set up cell-type specific TF networks. We defined the notion rewiring index, in which TF regulatory logic units change in relation to chromatic 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.
* We reconciled the cell-type specific TF and RBP networks applied in 15 cancer types to identify key regulators that drives tumor-to-normal differential expression. For discovered key regulators, such as MYC, we demonstrated how our network information can help to reveal how it works with other regulators in multiple cancer types.

Altogether, we validated our prioritized key regulators, high-impact regulatory regions, and deleterious SNVs through small-scale validations. Through siRNA knockdown experiment, we confirmed the role of MYC and SUB1 to up regulate their targets in breast and lung cancers. We validated the enhancer activities through luciferase assays and demonstrated effect of our selected SNV in these regions by introducing mutations into wild type sequences.

We believe that ENOCDEC is a powerful resource that connects the high quality single non-coding annotation nicely, and we have integrated it into the overall ENCODE encyclopedia resource. It also shed light on demonstrating by integrating high dimensional function assays, we could better understand the cancer genome.