1. I think the supplement does a good job explaining the tradeoffs you used to select your pairs
2. It was not clear where the community would go to find them, and whether they would understand the relationship between them and other data/analyses at the portal, including the ENCODE encyclopedia.

Comments in the paper:

* I would say cell types; there were plenty of primary cells in ENCODE 2, including NHEK, HMEC.
* Cell types; some are explants, some are primary cells, to me those are very different than cell lines.
* At least 7 of the factors in Fig 3 don’t bind DNA in a sequence specific manner; are the results from ChIP in the relevant cell type(s), or predicted from motif analysis?
* Good news if you have a model for AML; are they linked to CML?
* Again, are you finding RBPs, or sequence motifs for RBP families?
* Fig 4A RBPs includes TAF15 and GTF2F1; these are part of the Pol II initiation complex, they associate with DNA prior to initiation of transcription, and facilitate initiation.
* Doesn’t the ENCODE encyclopedia have these things? Are your calculations the same or different? Shouldn’t this be explained either way?
* Consistent with oncogenesis model? Reproducible? Consistent with neighboring SNV?
* Every locus lies in a TAD; do mean the candidate enhancer and predicted target promoter are in the same TAD?
* FWIW, there are both sequence specific TFs and cofactors bound at this location, in Fig 5D, nicely marked by DHS and flanking H3K27ac; one known function for GATA3 is a key factor in mammary epithelia

Other comments

* It was unclear what the relationship is, if any, between this work and Shirley Liu’s.
* in some cases I thought multiple comparisons were interesting
* PLEASE don’t refer to everything in ENCODE as cell lines

Analysis:

* I’d be curious about how well the tumor cells cluster with ES cells, as opposed to the mature cell type they appear to be derived from.
* I am most concerned about the A549/IMR-90 pairing, I am next most concerned about the K562/GM12878