Comments Response:

1. *It was not clear where the community would go to find them (the annotations like promoters, enhancers, and others), and whether they would understand the relationship between them and other data/analyses at the portal, including the ENCODE encyclopedia.*

   Response: We will certainly make this clearer. We're planning to add a section into the paper explaining how we merge our annotations with those in the encyclopedia. We're currently communicating with Zhiping on this, and we'll send you more details on how this would be done shortly.

2. *I would say cell types; there were plenty of primary cells in ENCODE 2, including NHEK, HMEC.*

   Response: You raise a good point. We'll use the term "cell type" more often, and we'll clarify that, although we're focusing on top-tier cell lines (for which a lot of data are available), ENCODE also includes many other cell types.

3. *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?*

   Response: Our results are based on is from ChIP-seq data, and we agree that this should be made clearer in the figure caption and the main text. In addition to focusing on TFSS, we also report the rewiring status of cofactors and chromatin remodellers.

4. *Good news if you have a model for AML; are they linked to CML?*

   Response: Indeed, we realized before that the pairing is incomplete. K562 is from a CML patient. In the mutational analysis section, we have to use CLL data because there is not very much whole-genome sequencing data available for CML. Given the insufficient data for CML, we have to refer to AML patient information for the survival analysis. It has been reported that a deletion of IKZF1 is a hallmark of ALL types of BCR-ABL fusions, but not of CML. A deletion of IKZF1 has also been identified as an acquired lesion when CML is transformed to ALL. We will likely de-emphasize this in next version of our manuscript.

5. *Again, are you finding RBPs, or sequence motifs for RBP families?*

   Response: We are finding RBPs, and *not* the motifs. The binding data is obtained from eCLIP experiments.

6. *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.*

   Response: In our classification, both TAF15 and GTF2F1 were classified as TF+RBP. We test such factors with peaks from both eCLIP and ChIP-seq data. If there is a significant association of

the binding score with target T/N expression change in eCLIP, we put it into the RBP Heatmap in Figure 4. Similarly, if the ChIP-seq peaks help to explain T/N expression change, we put it into the TF in Figure 4.

7. *Doesn't the ENCODE encyclopedia have these enhancers, promoters? Are your calculations the same or different? Shouldn't this be explained either way?*

   Response: We are carefully merging our annotation with the main encyclopedia. As explained earlier, we are closely working with Zhiping to cross reference our resources. Also we mentioned in previous sections that these annotations are results from integration from many assays, like STARR-seq, ChIP-seq, DNase-seq, RNA-seq, Hi-C, ChIA-pet, and RAMPAGE data. We would link everything to encyclopedia. If there is overlap, we label it. Otherwise we could create new labels.

8. *Consistent with oncogenesis model? Reproducible? Consistent with neighboring SNV?*

   Response: There is no clear conclusion offered regarding the role(s) of the *SYCP2* gene. As is, we just left it at the statement "aberrant expression of this protein may contribute to genetic instability during HPV-associated cancer development". We did not discuss this in further detail.

9. *Every locus lies in a TAD; do you mean the candidate enhancer and predicted target promoter are in the same TAD?*

   Response: We just show that this is quite an active region with considerable interactions within this TAD. The enhancer-target prediction results were obtained using our tool and combination of ChIA-PET and Hi-C data.

10. *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*

    Response: Thanks for the note. We can certainly look more closely into this. 7 out of 8 regions we validated were successful with strong regulatory evidences, and this shows robustness of our prioritization scheme.

11. *It was unclear what the relationship is, if any, between this work and Shirley Liu's.*

    Response: We apologize for the potential confusion here – perhaps we should have mentioned that Shirley is a co-author of this paper, and she has contributed to these analyses.

12. *In some cases, I thought multiple comparisons were interesting*

    Response: This is a good point. We are revising the manuscript to incorporate the notion of the normal by merging multiple related normal cell types. For example, for MCF-7, we could merge

the MCF10A and HMEC ChIP-seq to construct the "normal" regulatory network.

13. *PLEASE don't refer to everything in ENCODE as cell lines*

    Response: We see your point – where appropriate, we will use the term "cell type".

14. *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.*

    Response: This is a very good point. We are now carrying out this calculation.

15. *I am most concerned about the A549/IMR-90 pairing, I am next most concerned about the K562/GM12878*

    Response: We are concerned about the lung pairing, and we have a detailed section on pairing in the supplementary text. For the blood cancer, we are trying to use CD34+ progenitor cell as an appropriate comparison when applicable. We have tried to fix this with a merged normal, but we admit that the pairing issue is problematic.

16. *For HeLa, one could compare it to any primary epithelial cell type, or perhaps a few epithelial cell types. Of course, it is thought that HPV integration in the HeLa genome was a critical event in transformation (PMID: 3028716, PMID: 23925245), so viral TF, protein coding genes, and regulatory DNA may play a role in addition to host genes.*

    Response: This is a good point. In fact, we have previously attempted to pair HeLa-S3, but we could not easily find a healthy proxy for it. We will continue looking into the best cell type for doing this comparison.