1.

Since the FANTOM5 Atlas is the most comprehensive collection of transcribed enhancers across different primary cells and tissues, I would like to see a comparison of the model predictions in human to the enhancer dataset of the FANTOM5 Atlas dataset taking into account cell-type/tissue specificity. In a similar fashion, what is the overlap with the integrative ENCODE annotation proposed by Hoffman et al. NAR 2013. Assuming that the size of training datasets is the only limiting factor for achieving high discrimination performance, what is the minimum number of samples that guarantees good performance in the deployed method?

2.

Page 9: “Similarly, we did genome wide prediction of regulatory regions in ENCODE top tier human cell lines, including H1-hESC, GM12878, K562, HepG2 and MCF-7 (all available through our website)”.

Following my previous comment, I would like to see the comparison analysis with CAGE-defined enhancers and promoters for some cell-specific cases, comparison with the integrative ENCODE annotation proposed by Hoffman for all top-tier cell-lines as well as comparison with other studies (see previous papers) that validated the regulatory activity of different segments in K562, HepG2 or H1-hESC cell-lines