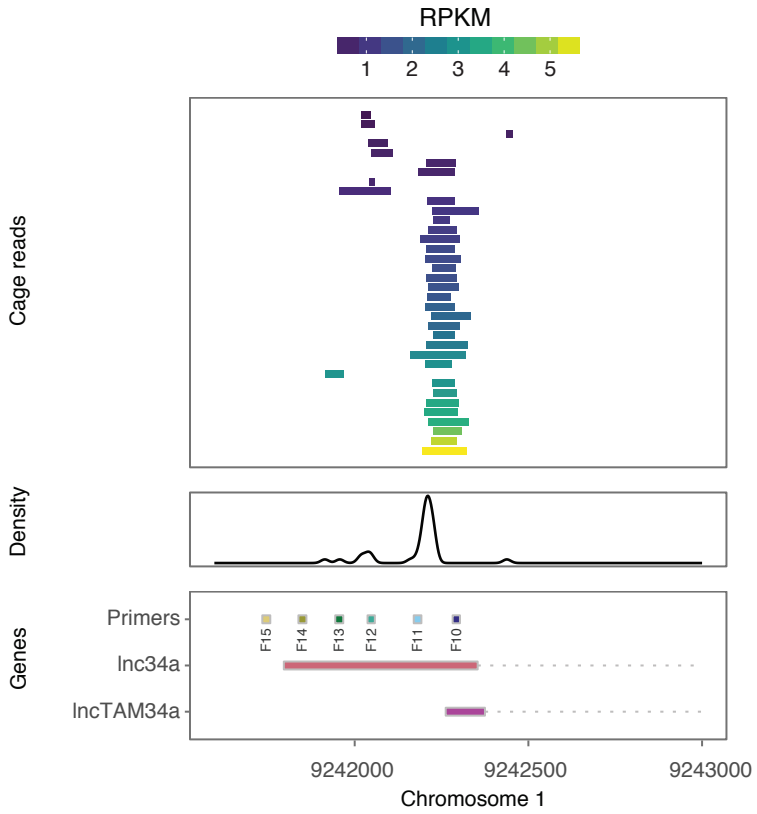

Supplementary Figure 6: All available CAGE data from the ENCODE project for 36 cell lines was downloaded from the UCSC genome browser for genome version hg19. Of these, 28 cell lines had CAGE transcription start sites mapping to the plus strand of chromosome 1 and in regions corresponding to 200 base pairs upstream of the Lnc34a start site (9241796 - 200) and 200 base pairs upstream of the GENCODE annotated lncTAM34a start site (9242263 + 200). These cell lines included: HFDPC, H1-hESC, HMEpC, HAoEC, HPIEpC, HSaVEC, GM12878, hMSC-BM, HUVEC, AG04450, hMSC-UC, IMR90, NHDF, SK-N-SH_RA, BJ, HOB, HPC-PL, HAoAF, NHEK, HVMF, HWP, MCF-7, HepG2, hMSC-AT, NHEM.f_M2, SkMC, NHEM_M2, and HCH. In total 74 samples were included. 17 samples were polyA-, 47 samples were polyA+, and 10 samples were total RNA. In addition, 34 samples were whole cell, 15 enriched for the cytosolic fraction, 10 enriched for the nucleolus, and 15 enriched for the nucleus. All CAGE reads were plotted and the RPKM of the individual reads was used to colour each read to indicate their relative abundance (top panel). In addition, a density plot (middle panel) shows the distribution of the CAGE reads in the specified interval and the transcription start regions for Lnc34a and lncTAM34a as well as primer positions from the primer walk assay (bottom panel).

An unannotated transcript, Lnc34a, arising from the antisense orientation of the miR34a locus and with a transcription start site >250 bp upstream of the annotated lncTAM34as start site, has been previously reported in a study examining colorectal cancer (Wang et al. 2016). Among the findings in Wang et al. the authors discover that Lnc34a negatively regulates miR34a expression via recruitment of DNMT3a, PHB2, and HDAC1 to the miR34a promoter. Although the Lnc34a and lncTAM34a transcripts share some sequence similarity, we believe them to be separate RNAs transcripts. Furthermore, Lnc34a may be highly context dependent and potentially only expressed at biologically significant levels in colon cancer stem cells, or other stem-like cells, in agreement with the conclusions drawn in the paper.

Several lines of evidence point to the fact that lncTAM34a and Lnc34a are not the same transcript and, in addition, that Lnc34a expression may be confined to a highly specific subset of colorectal cancer stem cells (CCSC). First, we were unable to detect transcription upstream of the 5' start site that was defined in the primer walk assay (Fig. 1E and Supplementary Fig. 1B) although the reported Lnc34a start site is >250 base pairs upstream of the F12 primer used in this assay. This could simply be due to the fact that Lnc34a is not expressed in HEK293t cells in which the assay was performed. To further investigate the existence of transcription start sites in the antisense orientation of the miR34a locus, we interrogated CAGE data from 28 cell lines.

The results show a high density of CAGE tags aligning to the region corresponding to the annotated lncTAM34a start site. Several additional peaks, albeit with a much lower average expression, aligns slightly upstream of the annotated lncTAM34a start site, one of which, corresponds to the upstream start site detected in our primer walk analysis (Figure 1e). Despite this, we find no CAGE tags aligning at or upstream of the transcription start site of the Lnc34a transcript. This potentially indicates that Lnc34a is tightly regulated and specifically expressed in the CCSC context, as suggested by the authors. An alternative intrepretation could be that Lnc34a expression is present in a subset of the examined cell lines although at levels too low to be detected. Finally, Lnc34a may not be 5'-capped precluding its detection by CAGE.