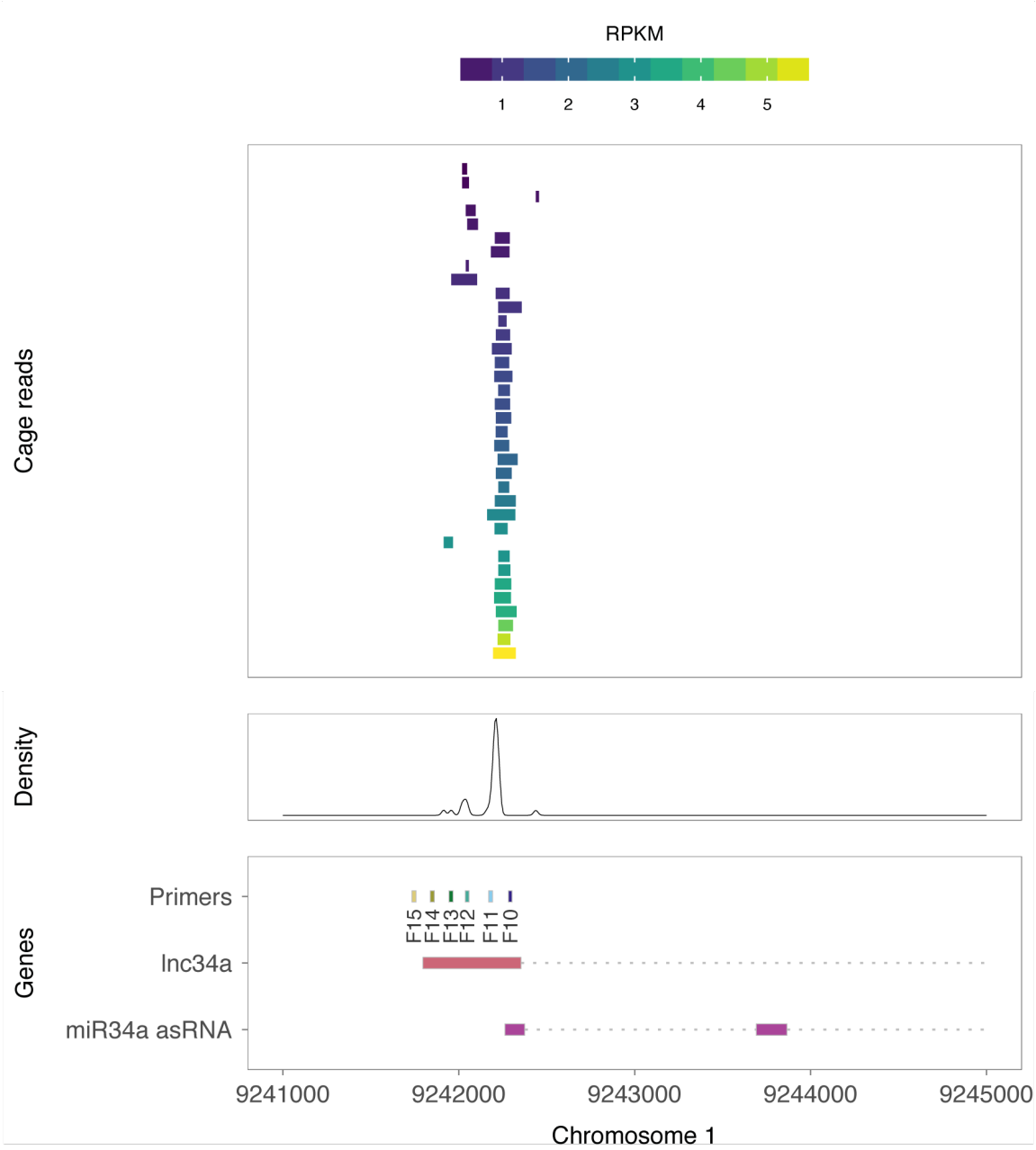
Supplementary Document 1

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 miR34a asRNAs 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 miR34a asRNA 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 miR34a asRNA 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. 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.



CAGE analysis at the miR34a locus: 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 miR34a asRNA 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, 15 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 miR34a asRNA as well as primer positions from the primer walk assay (bottom panel).

The results show a high density of CAGE tags aligning to the region corresponding to the annotated miR34a asRNA start site. Several additional peaks, albeit with a much lower average expression, aligns slightly upstream of the annotated miR34a asRNA 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 the transcription start site 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 claimed by the authors. An alternative intrepretation could be that lnc34a expression is present in a subset of the examined cell lines although at too low levels to be detected. Finally, lnc34a may not be 5'-capped precluding its detection by CAGE.

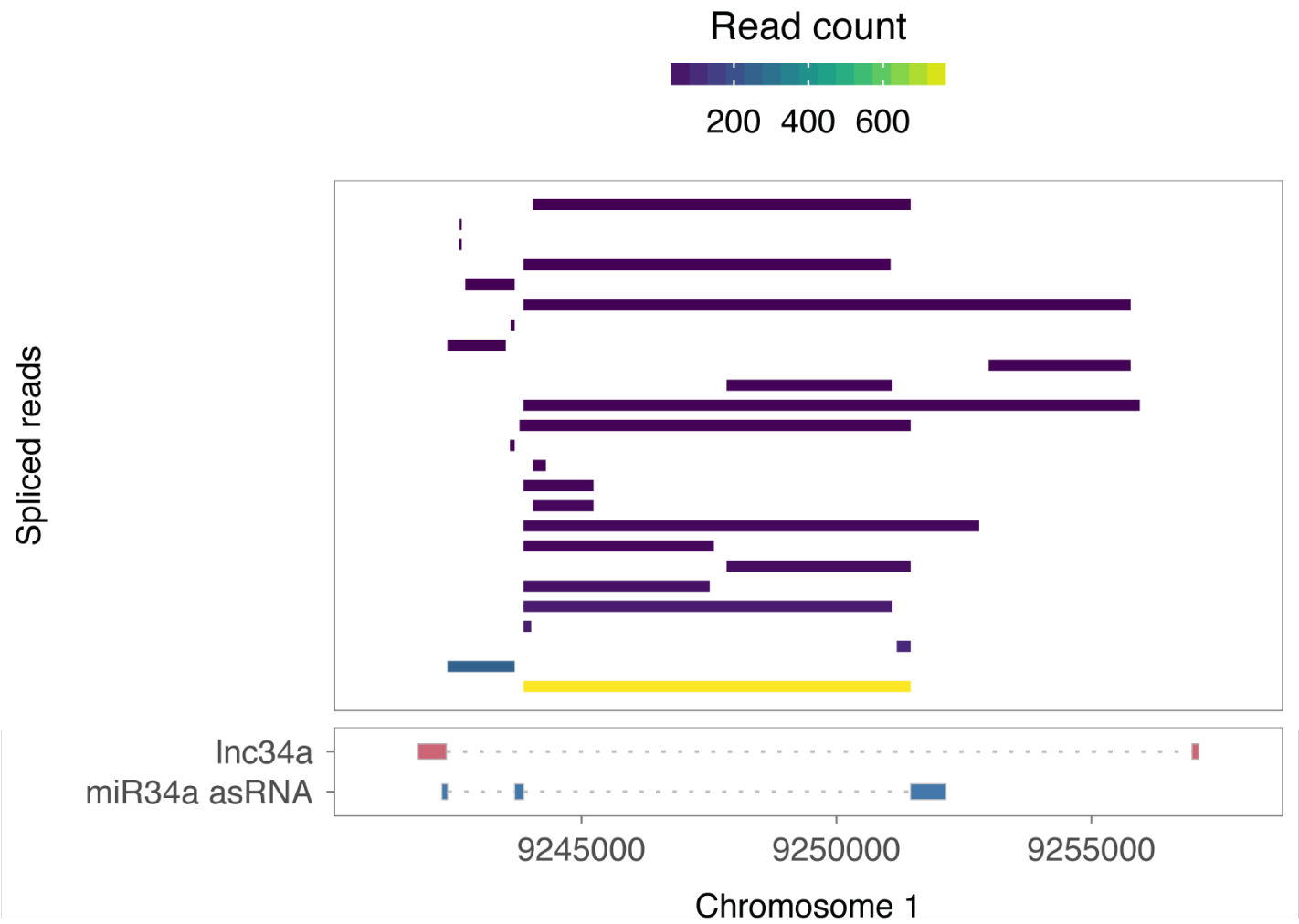
In order to lnc34a expression in a manner that is not dependant on 5'-capping, we proceeded to examine spliced RNA sequencing reads from 36 cell lines, taking advantage of the fact that lnc34a has an exon which is not present in any miR34a asRNA isoforms.

These results indicate that, although splice junctions corresponding to multiple isoforms detected by PCR cloning of the miR34a asRNA were detected, the data give no support for the presence of the splice junction between the first and second exon of lnc34a. In summary, these results indicate that lnc34a expression may be confined to colon cancer stem cells as evidence for its expression in the examined cell lines is not forthcomming.

In addition, we note several comments in the public review that was published in conjunction with the work by Wang et al. The authors mention, and provide data, indicating that lnc34a expression is not changed upon ectopic expression of TP53. In contrast, miR34a asRNA is strongly regulated by TP53 as the evidence shows in this and others (Léveillé 2015, Rashi-Elkeles 2014, Hünten 2015, Ashouri 2016, Kim 2017) findings.

Furthermore, Wang et al. also mention in the public review that lnc34a has a low expression level in HCT116 cells although we detect robust expression of miR34a asRNA in this cell type (**Figure 1b**).

In summary, these results indicate that lnc34a expression is not present in the cell types examined where there exists strong evidence for the presence miR34a asRNA. For these reasons, we believe miR34a asRNA and lnc34a to be individual antisense RNA transcripts.



Splice junction analysis at the miR34a locus: All available whole cell (i.e. non-fractionated) spliced read data originating from the Cold Spring Harbor Lab in the ENCODE project for 38 cell lines was downloaded from the UCSC genome browser. Of these cell lines, 36 had spliced reads mapping to the plus strand of chromosome 1 and in the region between the lnc34a start (9241796) and transcription termination (9257102) site (note that miR34a asRNA resides totally within this region). Splice junctions from the following cell lines were included in the final figure: A549, Ag04450, Bj, CD20, CD34 mobilized, Gm12878, H1hesc, Haoaf, Haoec, Hch, Helas3, Hepg2, Hfdpc, Hmec, Hmepc, Hmscat, Hmscbm, Hmscuc, Hob, Hpcpl, Hpiepc, Hsavec, Hsmm, Huvec, Hvmf, Hwp, Imr90, Mcf7, Monocd14, Nhdf, Nhek, Nhemfm2, Nhemm2, Nhlf, Skmc, and Sknsh. All splice junctions were included in the figure and coloured according to the number of reads corresponding to each (top panel). In cases where the exact same read was detected multiple times the read count was summed and represented as one read in the figure. miR34a asRNA and lnc34a transcripts are represented for reference (bottom panel).