Background and significance:

The paper by Adey *et al.*, illustrates the potential of current sequencing methods for understanding cancer origins as well as some of the limitations. The ~100 base sequence reads produced by the Illumina technology (the most widely used and cheapest sequencing technology available today) are typically used just to determine alterations of individual bases. Here the authors applied a variety of methods to obtain long range information that allowed them to determine nearly whole chromosome level of continuity. The result is a fairly comprehensive look at the genetic makeup of the HeLa cancer cell line, including the results of the HPV-18 viral integration.

The HeLa cell line, as we have mentioned before in class, was the first permanent cell line ever established for human cells. As the authors note it is extremely widely used, with the number of PubMed abstracts referring to it is now up past 78,000. However, Henrietta Lacks, the person from whom the tumor cells were taken, never gave her permission to use the cells in research. The case was made famous in a book by Rebecca Skloot called "The Immortal life of Henrietta Lacks." This is the first paper that obtained permission from the surviving family members for the use of the HeLa cell line in research.

As DNA sequencing becomes cheaper and more powerful, it will become increasingly feasible to obtain comprehensive genomic views of individuals' cancer cells. Tests are already being employed in clinic that examine the sequences of ~200 cancer "driver" genes, and more tests are rapidly entering clinical practice. However, the functional significance of the sequence variants discovered often remains unclear, and as shown in this paper, understanding just what is represented at the genetic level can be challenging. As you read the paper, use your growing knowledge of the molecular basis of cancer to see if you can figure out why these cells represent such an aggressive cancer cell line.

Notes:

The HPV genes E6 and E7 function analogously to the SV40 large T antigen and adenovirus: E6 binds TP53 and E7 binds RB1.

Don't worry too much about the exact methods used to obtain haplotype information. They are clever, but with any luck advancing technology will make them unnecessary.

The ENCODE project (ENCyclopedia Of DNA Elements) has been a concerted effort to identify the functional portions of the human genome, delineating not only the protein-coding genes more precisely, but also finding regulatory elements and non-coding RNAs. HeLa cells, because of their ease of use, were used extensively in the project.

Questions:

1. There are many versions of HeLa cells. Which cell line did they use? Why?

2. Why do the authors refer to this genome as “haplotype-resolved”? Give a reason why haplotype resolution might be useful.
3. What do the authors suggest is the result of the long-range interaction between the HPV integration site and *myc*?