ASSEMBLY VERSUS ALIGNMENT

After sequencing fragments of DNA to obtain reads, most genomic pipelines follow one of two steps. The reads can be de novo assembled to construct longer stretches called contigs from scratch, with overlapping sequences on the ends dictating which read pieces belong next to each other (left). Alternatively, reads can also be aligned to a reference genome to identify small genetic variations (right). Where de novo assembly can be thought of as assembling a puzzle without the use of the picture on the box, alignment is the equivalent of piecing together a puzzle by looking at that picture. However, because a singular reference genome fails to capture all of the genetic diversity across humans, some sections of DNA might not be able to align to the reference genome well.