Bowtie2 and Hisat2 are both still actively developed. Hisat2 uses some code from Bowtie2 but makes use of a different underlying index (first it was a hierarchical FM index, now its a hierarchical graph FM index) that is more cache friendly than what Bowtie2 uses (a relatively standard FM index). Both are great tools. If you are doing spliced alignment (i.e. aligning RNA-seq reads to the genome) you should use Hisat, as Bowtie doesn't support spliced alignment. If you aren't doing spliced alignment (e.g. DNA-seq reads or some such, or mapping RNA-seq reads directly to the transcriptone rather than the genome), either tool should do a good job. I've not seen them compared directly for that purpose, but I've seen anecdotal evidence that Bowtie2 may be a bit more sensitive (at least with less parameter tweaking) for unspliced alignments.

A downside to Bowtie2 is that the aligner was designed with the intent of aligning DNA reads to a reference genome and does not allow for the addition of a transcript annotation file to aid in the mapping of RNA reads.

BWA is generally slower than Bowtie2 with similar sensitivity and both tools can perform gapped alignment for the identification of indels and can effectively map paired-end reads. However, BWA is a bit more accurate and provides information on which alignments are trustworthy. Small numbers of bad alignments can result in many false variant calls, so accuracy is paramount, and is the basis for choosing BWA.

BWA-MEM also has the ability to report multiple alignments per read in the form of secondary and supplementary alignments (on by default). Like Bowtie2, a downside to the BWA-MEM algorithm is that it was originally created to align DNA reads to a reference genome and therefore does not have the ability to import a transcript file to identify splice junctions in RNA.

***HISAT2***

HISAT2 also deals with repetitive sequences differently than other aligners by combining repeat sequences in the reference genome into one sequence. This reduces the number of alignments reported by only reporting one alignment for a read aligning to these regions rather than one per repetitive element. As HISAT2 is designed to align both genomic and RNA reads, the ability to input a file of known splice sites allows for better mapping of exons to places in the genome.

STAR allows for the genome to be pre-built once and the resulting indexes to be loaded in for each individual alignment, which drastically decreases the runtime. STAR provides the option of inputting an annotation file of known splice junctions to aid in the process.

Bwa (Burrows–Wheeler-Alignment) was developed for mapping short DNA sequences against a reference genome and was extended for RNA-Seq data analysis. For indexing, the algorithm constructs a suffix array and Burrows–Wheeler-Transformation (BWT), and subsequently matches the sequences using a backward search [[11](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7084517/#B11-ijms-21-01720)]. STAR (Spliced Transcripts Alignment to a Reference) is a specialized tool for RNA-Seq reads that uses a seed-extension search based on compressed suffix arrays [[12](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7084517/#B12-ijms-21-01720)] and can detect splice-junctions. HISAT2 (Hierarchical Indexing for Spliced Alignment of Transcripts 2) is also a splice-aware aligner using a graph-based alignment approach (graph Ferragina Manzini index) that can align DNA and RNA sequences

STAR has a higher tolerance for more soft-clipped and mismatched bases compared to HISAT2, which leads to a higher mapping rate for STAR and more unmapped reads for HISAT2. STAR showed the highest fraction of mapped reads for both accessions among all compared mapping tools

**1.** **Please list the considerations you have to pick an aligner for your experiment (Make sure you mentioned the different consideration when your sample is genomic DNA vs RNA).**

The experiment is trying to find how mutation of SOD1 is related to  Amyotrophic Lateral Sclerosis(ALS) in humans and figure out the transcriptional and functional changes done by mutant SOD1 in human motor neurons. The experiment is using RNA Seq technology to identify differentially expressed genes by mutation of SOD1 compared to control. To choose an aligner for this experiment, I had several considerations like whether the alignment is for DNA or RNA seq analysis, how much time it takes to align and how efficient it works compared to other aligners. If the alignment is for DNA then bowtie2 or BWA works well because they were built for DNA seq analysis and as a result they are not splicing aware, though BWA was extended for use in RNA seq but it is not much efficient like RNA seq aligners as it is not splicing aware. If the sample is genomic RNA, then STAR and HISAT2 aligners would be the highest priority because they are splicing aware, which is very important while working with RNAs because we have to be aware of the introns present in the genome, though HISAT2 supports both mapping of DNA and RNA seq. I would choose STAR and HISAT2 to compare how these two aligners perform (both the aligners build an index first and align the reads with the index, but HISAT2 uses Graph FM index based on BWT graphs whereas STAR uses a Suffix Array index) and then use BWA to compare with the other two aligners and see how efficient it is.

**3. Briefly describe the differences between the alignment results, and select an alignment tool to use if you were the author of this study. Justify your pick.**

I will first start with STAR. I ran STAR twice, by providing with the annotation and without annotation. For both run, I got almost same results, with annotation file it was 99.96% mapped and 99.96% properly paired, and without annotation file it was 99.81% mapped and 99.81% properly paired. Also there was slightly less number in total reads for annotated one than basic which might be due to the fact that without annotation file the aligner might have failed for a perfect match rather two partial matches and it counted as two reads so it might be of that.

Now for HISAT2, I also ran it twice like STAR, with annotation and without annotation. Same as Star, both hisat2 with and without annotation file gave almost same results, with annotation file it was 97.53% mapped and 96.29% properly paired, and without annotation file it was 97.52% mapped and 96.31% properly paired. So from this we can understand that hisat2 works well with annotation file provided by small percentage compared to not providing the annotation. But if compared between hisat2 and Star, Star aligner is much more superior in case of alignment results. Star has the ability of "soft-clipping ends of highly mismatched reads" (1) which I feel helps Star in higher mapping rate than hisat2. Another point is worth mentioning that if you compare total reads between hisat2 and Star with annotations, and hisat2 and Star without annotations separately, hisat2 usually have less total reads which might be due to the fact that hisat2 combines repeated sequences into a single sequence which decreases the number of reads for aligning as it counts it as one single read rather counting multiple times for a single read.

Now analyzing results for BWA, it has pretty good mapped reads 99.96% (same as STAR with annotation) but decreased properly paired reads 86.28% compared to the other two. The decrease in properly paired reads may be due to the fact that BWA is splicing unaware. Also flagstat results showed higher number of total reads compared to Star and hisat2. This is because each aligner handles reads differently, like BWA reports each and every match and flags the best one, so considering the bam files contain all these multiple reads, samtools does not figure out and count the flagged one rather counts every maps as reads in bam file, so assuming that same fastq file was provided, the total count number in BWA is high due to this reason. Also I observed some supplementary reads in the BWA bam file which might be related to chimeric alignments. Also, BWA takes more time than both STAR and hisat2.

So comparing all the results above, I would choose Star aligner based on its efficiency for the research.