*What I'd like is a one-page survey of the different transcriptome methodologies, which could include labmethods and alignment tools if you think it's relevant (e.g., Erange supports using bowtie and blatinstead of Eland for read mapping).*

*I hope that's not too general? Basically anything you think is significant is good, like the statistical assumptions behind their approaches and anything relating to the need to develop more sophisticated probabilistic models, e.g., to improve the identification and quantification of low copy-number RNAs and discriminate between RNAs from related genes.*

One of the biggest challenges for next-gen sequencing is informatics, and that ranges from primary analysis to secondary alignment and assembly. While many programs exist for mapping short reads, there's still a paucity of good tools for emerging applications such as transcriptome analysis.

There are no exceptions to the rule that speed and sensitivity are mutually exclusive. While some programs are faster, others are more accurate. The choice of program will depend on the specific research goals, as no single algorithm can outperform the others in all aspects.

For example, Roche's Genome Sequencer FLX comes with an aligner called Newble which performs well on Roche data researchers must often explore other tools for large amounts of data and shorter reads. Eland is the default software that comes with the Illumina Genome Analyzer and is known to run fast for short read data, including tag-based assays. Although able to align longer reads, Eland only uses the first 32 base pairs of a read in alignment. Embedded Perl scripts are required to align shorter reads and/or mate pairs efficiently. Eland also misses short indels and can align with at most 2 mismatches, whereas MAQ can handle more.

The MAQ algorithm is better for paired end reads and longer reads, and does both mapping and variant calling. As an exhaustive aligner, it is useful when accuracy is more important than speed, such as for SNp or indel detection. Although users like the rich feature set of MAQ, which includes the ability to do gapped alignment for paired end reads, it is much slower than Eland and may not be preferred when computing resources are limited.

Researchers often discover that their goals require the use of a variety of aligners, especially if their interests involve finding small insertions and deletions. This may involve building a patchwork including Eland, Maq, and SOAP, an aligner optimized for exhaustive, whole genome alignment of short reads. An recent alternative to SOAP for whole genome alignment using Illumina is Novoalign, which does gapped alignment, provides quality-based scoring (like Maq), can handle multiple alignment and paired end reads. Reports are that it’s faster than SOAP but a careful analysis is yet to be conducted. SHRiMP does Smith-Waterman so has a gain in accuracy over other software, but is slower as a consequence. It is a good choice for small genomes and could be improved by making use of paired end information and calculating mapping quality.

Despite the existence of useful tools such as those described above, the fact remains that the alignment of transcriptome data often requires the use of multiple different tools, in essence, a `pipeline’ of tools whose components may vary depending upon the research goals. The necessary tools must be compiled from various resources and combined by the researcher. In addition, alignments produced by different alignment tools provide different results, so combining different alignments may be necessary to provide more reliable estimates of genomic aberrations (young). How to combine information from multiple aligners is an open research question.