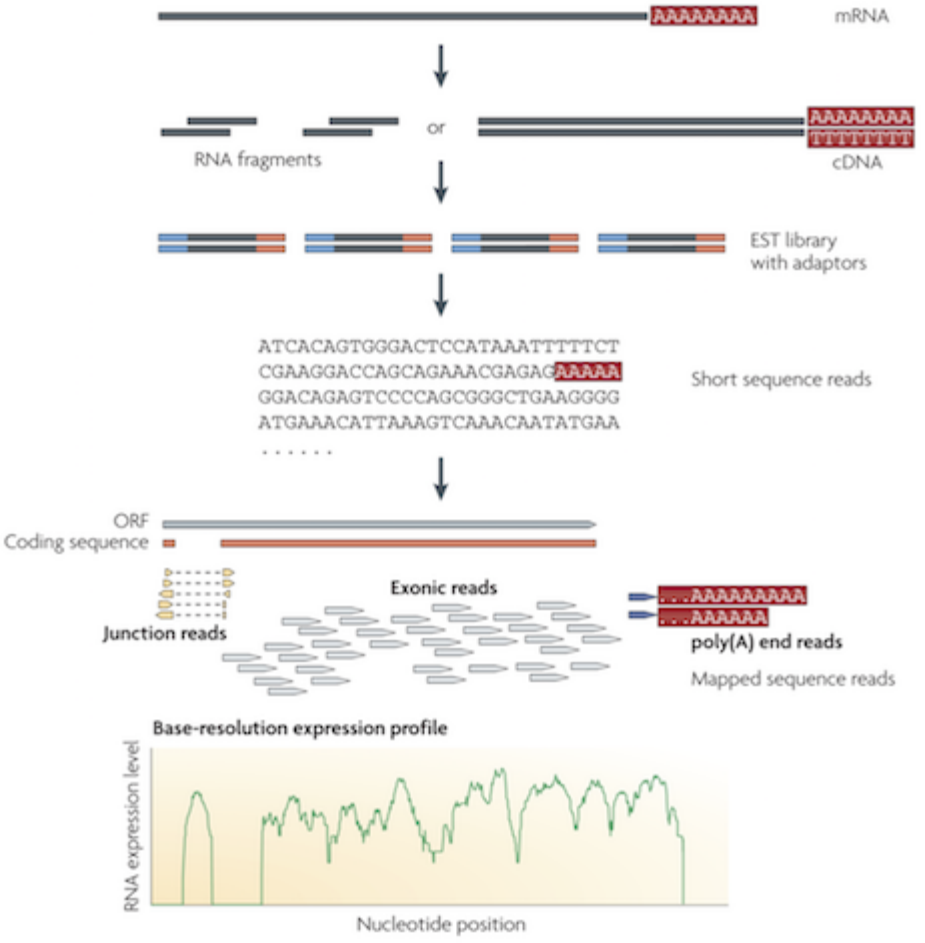
RNASeq analyses

RNASeq procedure

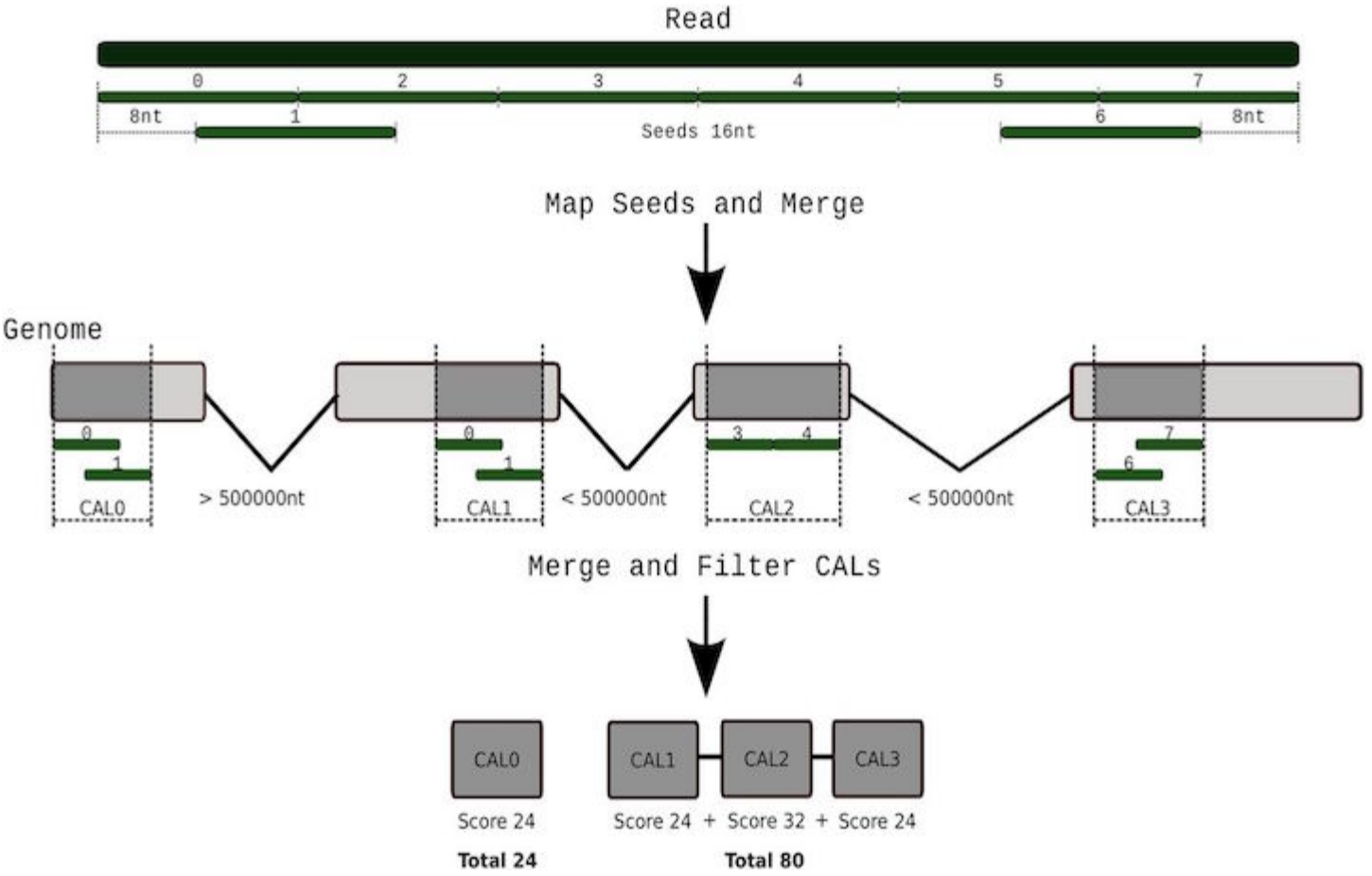
Wang et al. Nat Rev Genetics. 2009. doi:10.1038/nrg2484



Multiple approaches

1. Genome sequenced, align RNAseq reads to genome
2. de novo Assembly of mRNA into transcripts
3. Quantify gene expression from reads aligned to genome or transcripts

Reads to Genome mapping



Tarraga et al 2017. DNA Research.10.1093/dnares/dsv039

Reads to Genome mapping

Challenges: mRNA is spliced, genome contains introns

Splice-aware short read aligners. Speed and accuracy tradeoffs

- Tophat + Bowtie
- HISAT/HISAT2
- GMAP/GSNAP
- STAR

Quantify expression

- Count reads overlapping exons
- Table of total read counts per gene
- Normalize counts for gene length and sequencing library depth
- Gene expression then is FPKM - Fragments per Kilobase per Millions of reads
- Tools: htseq-count, stringtie
- BEDtools
- R tools with iRanges

Evaluating expression differences

Statistical tools for evaluating gene expression differences

- Ballgown bioconductor package
- DESeq bioconductor package
- edgeR bioconductor package

Alternative approach for Quantifying

Compare reads to **Transcripts** instead of Genome

- Kalisto and Sailfish are common tools
- Bray et al 2016 "Near-optimal probabilistic RNA-seq quantification" doi:10.1038/nbt.3519
- Patro et al 2014 "Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms" doi:10.1038/nbt.2862

Denovo assembly

Trinity Assembler for RNASeq

```
$ module load trinity-rnaseq
$ module switch perl/5.22.0
$ Trinity --seqType fq --left reads_1.fq --right reads_2.fq --CPU 8 --max_m
```

ORF identification

TransDecoder

- Finds Open Reading Frames in mRNA transcripts

```
$ module load transdecoder
$ TransDecoder.LongOrfs -t target_transcripts.fasta
```

RNAseq read mapping

Using HISAT2

```
# srun --ntasks 8 --pty bash -l
$ mkdir rnaseq; cd rnaseq
$ cp -r /bigdata/gen220/shared/projects/RNAseq ./
$ module load hisat2
$ cd genome
$ ls -l
$ hisat2-build yeast_genome.fasta yeast
$ cd ..
```

RNAseq read mapping

```
$ hisat2  -x genome/yeast -1 fastq/yeast_RNASeq_1.fq -2 fastq/yeast_RNASeq_
  -S RNASeq_aln.sam  -p 16
$ module load samtools
$ samtools view -b RNASeq_aln.sam > RNASeq_aln.bam
$ samtools sort RNASeq_aln.bam > RNASeq_aln.sort.bam
$ samtools index RNASeq_aln.sort.bam
$ samtools flagstat RNASeq_aln.sort.bam
```

Process BAM files for other tools

- give to htseq-count to get the read depth
- process with stringtie

```
$ module load stringtie
GTF=genome/genes.gff
$ stringtie -G $GTF -b stringtie_out -e -o stringtie.gtf -A stringtie.gene_
```