What is the difference between these cells?

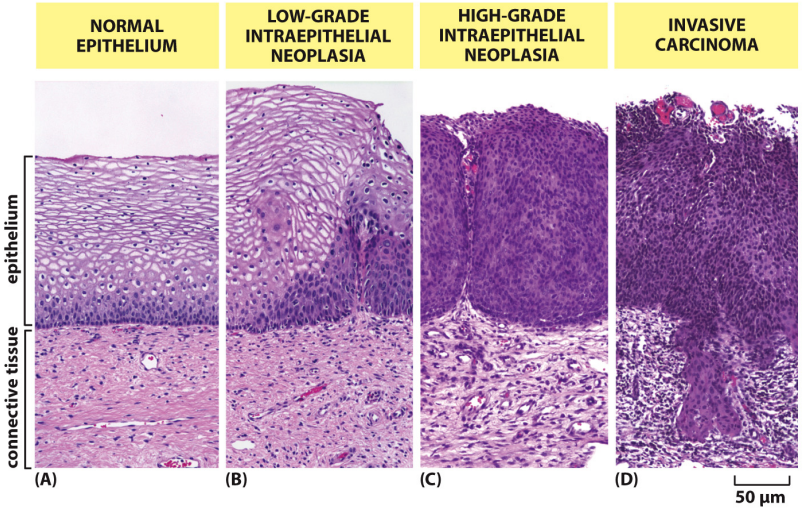


Figure 20-8 Molecular Biology of the Cell 6e (© Garland Science 2015)

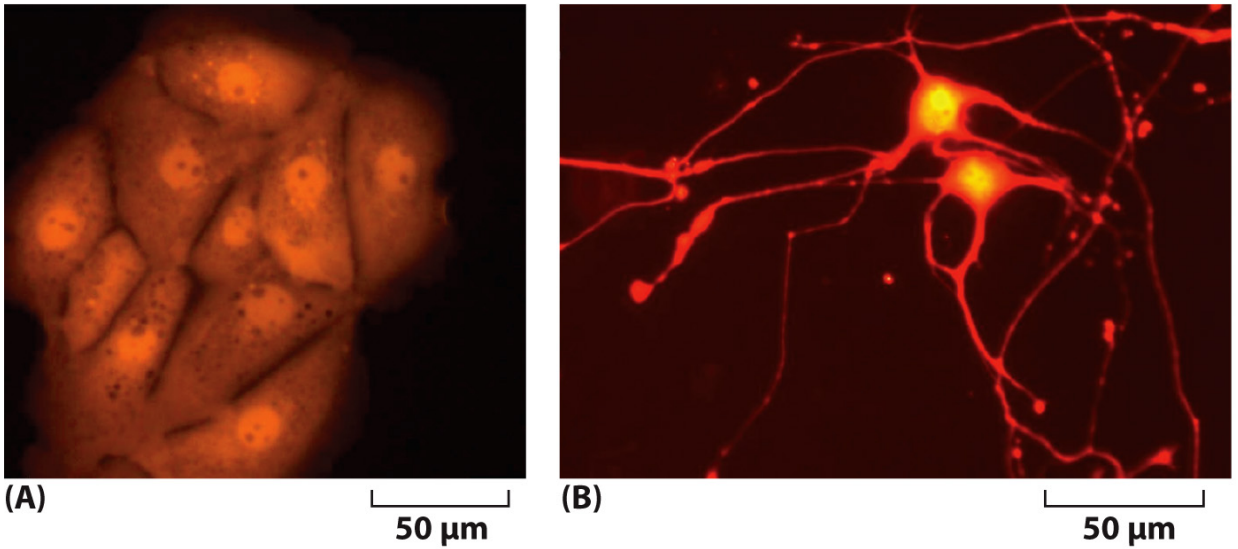


Figure 7-34 Molecular Biology of the Cell 6e (© Garland Science 2015)

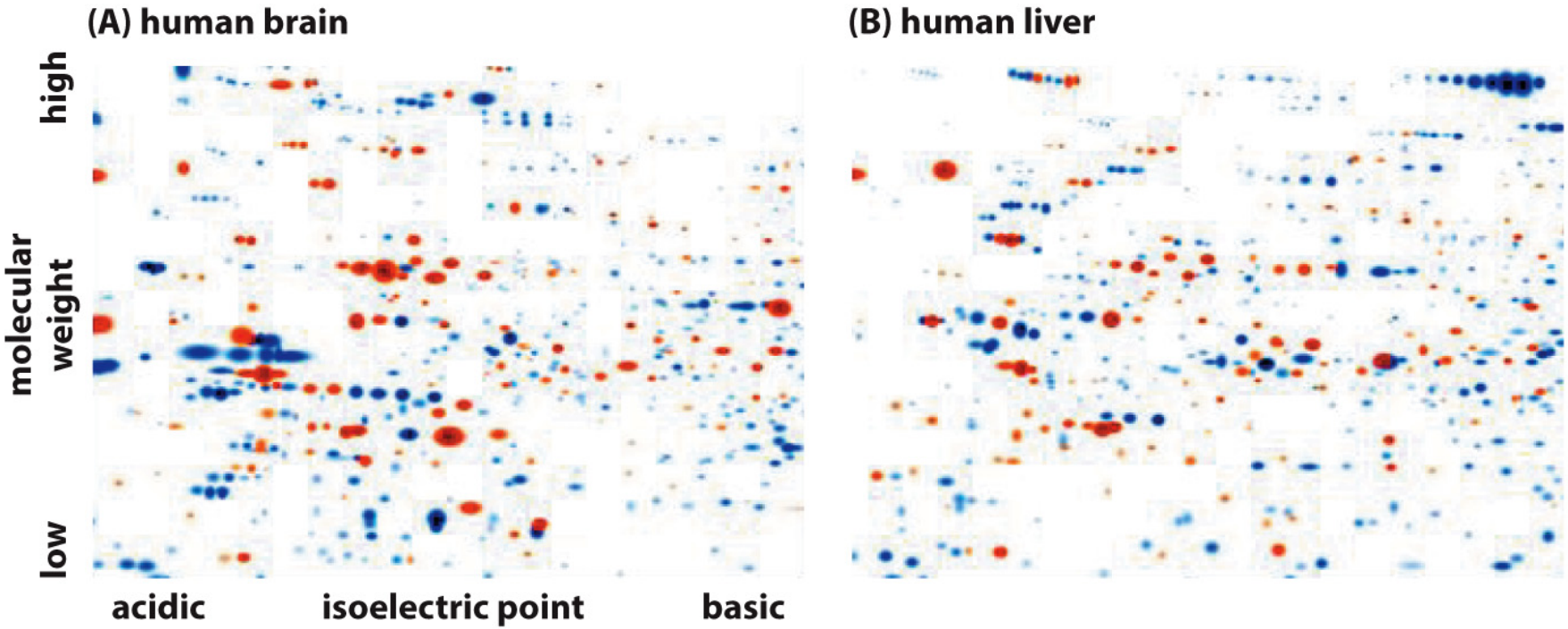


Figure 7-4 Molecular Biology of the Cell 6e (© Garland Science 2015)

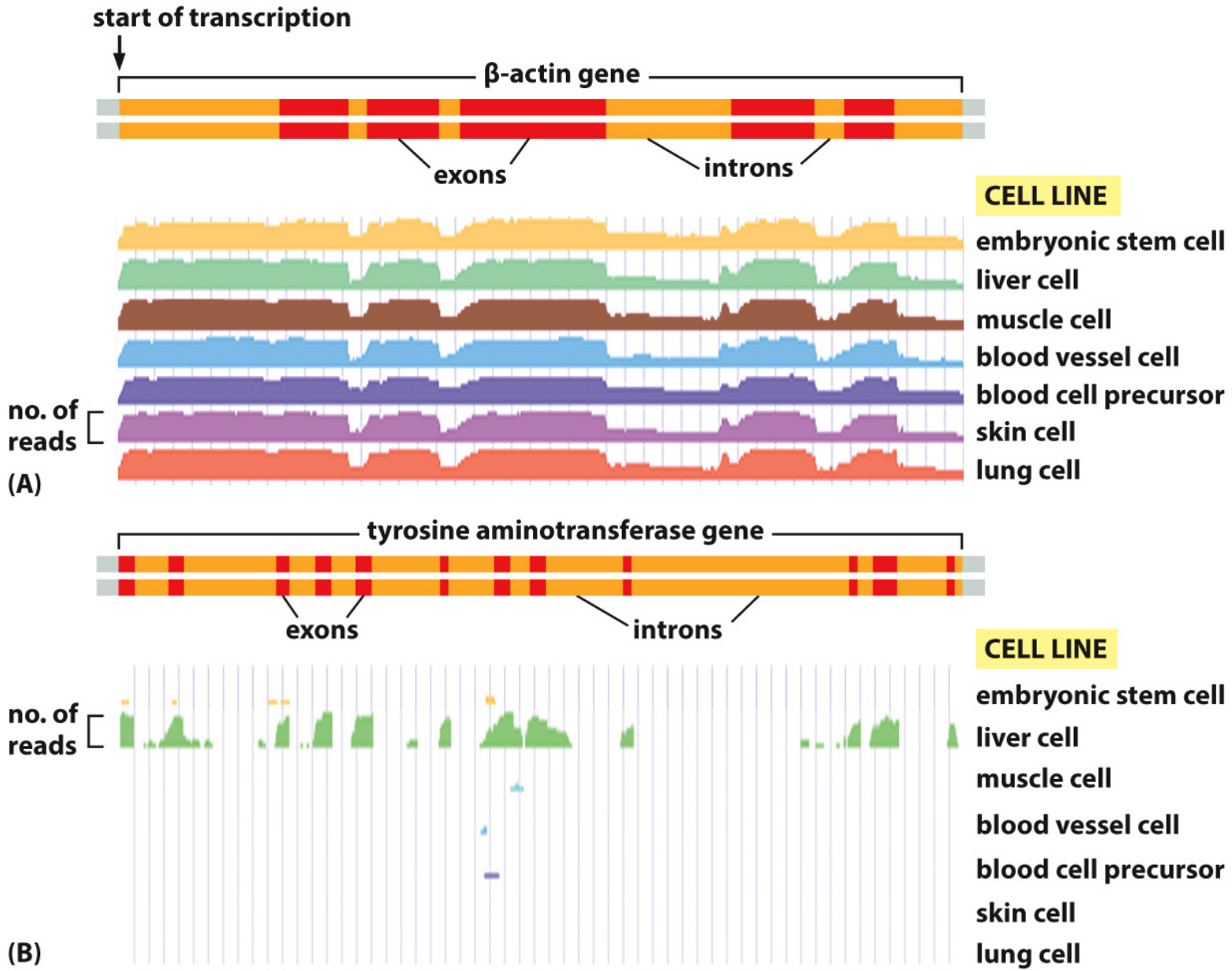


Figure 7-3 Molecular Biology of the Cell 6e (© Garland Science 2015)

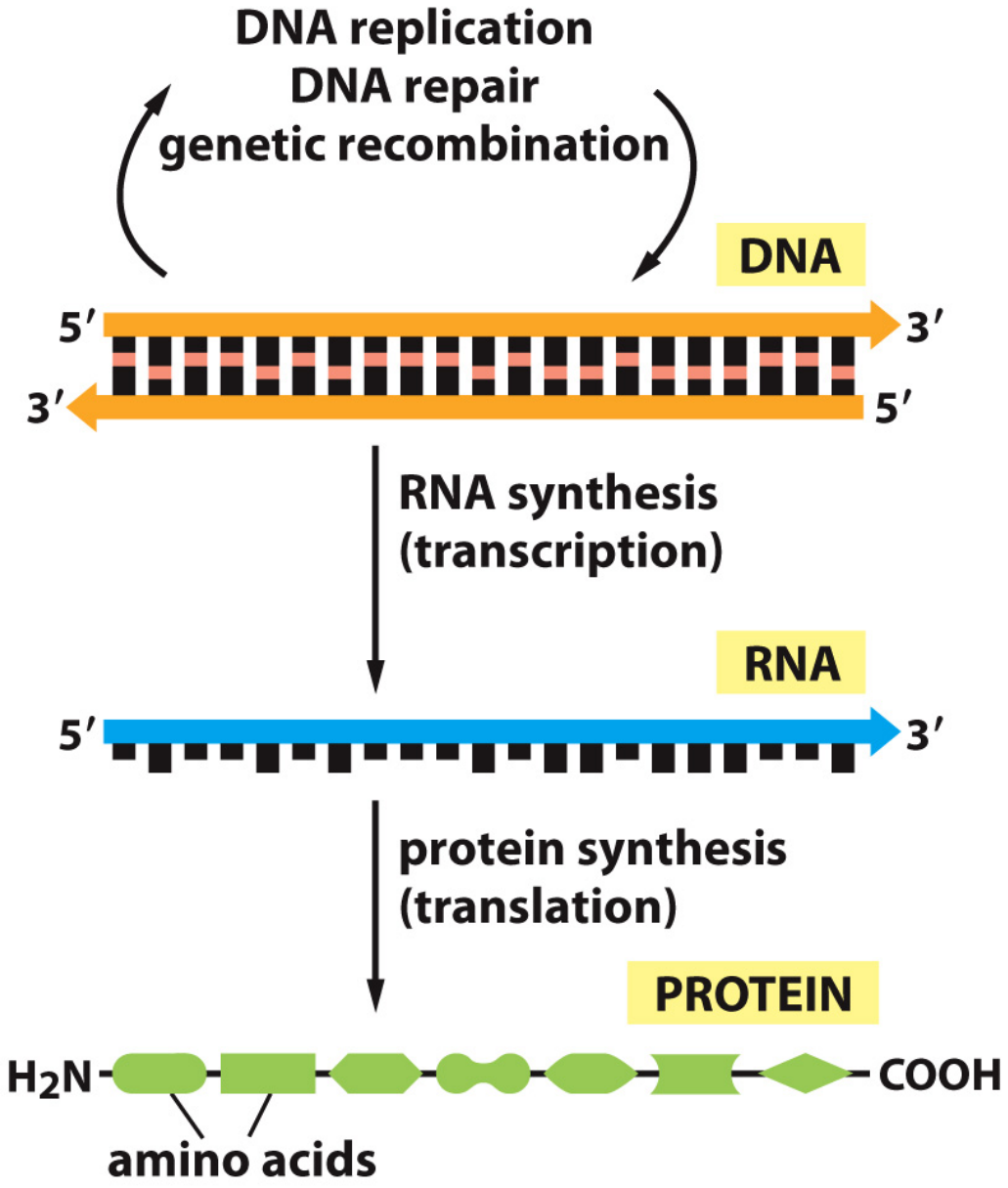


Figure 6-1 Molecular Biology of the Cell 6e (© Garland Science 2015)

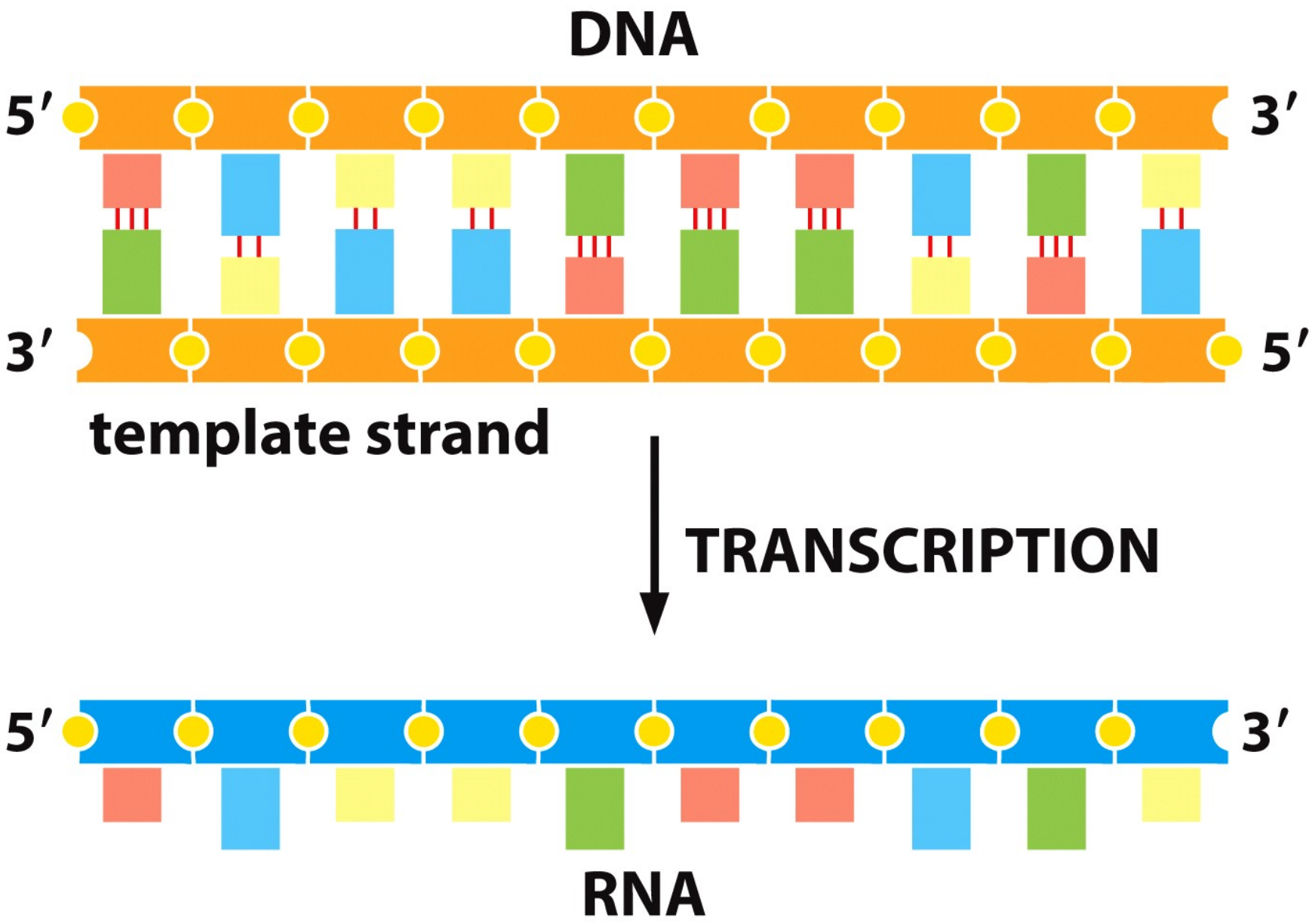


Figure 6-8 Molecular Biology of the Cell 6e (© Garland Science 2015)

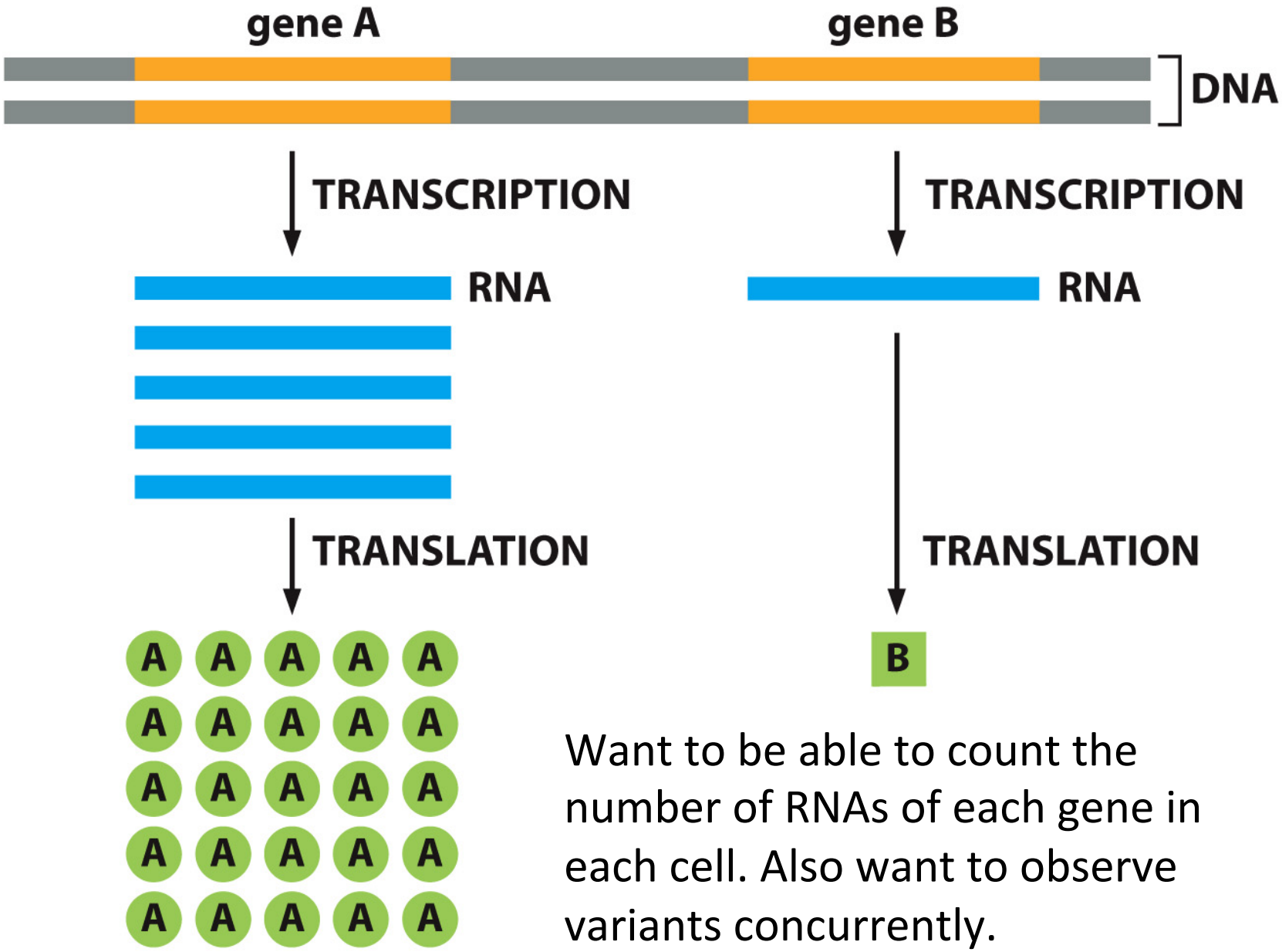


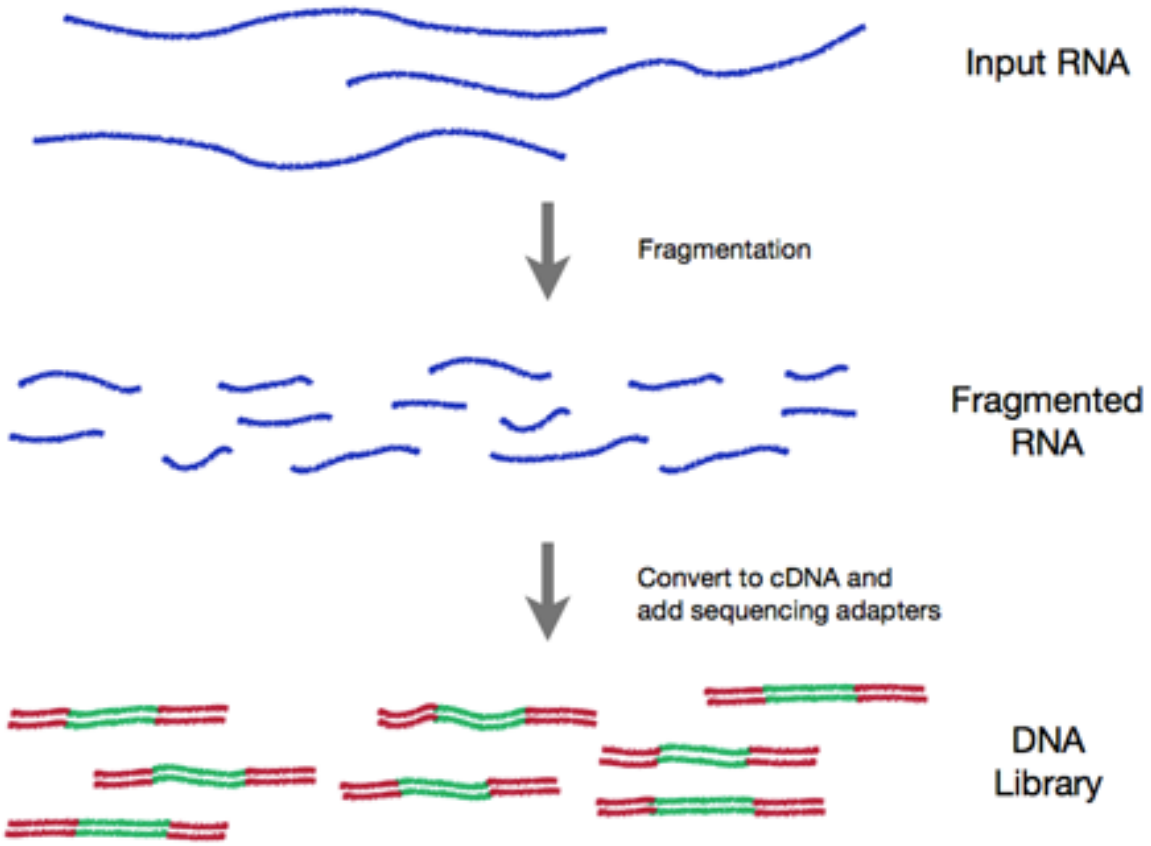
Figure 6-3 Molecular Biology of the Cell 6e (© Garland Science 2015)

Want to be able to count the number of RNAs of each gene in each cell. Also want to observe variants concurrently.

How do we measure RNA transcription?

RNA-Seq

- Sequence the RNAs found in a population of cells
  - We'll talk about single-cell RNA-seq later



RNA-seqlopedia, http://rnaseq.uoregon.edu/

Target Enrichment

- What is most common RNA in a cell?
  - Ribosomal RNA
  - Up to 80% of the amount of RNA in a cell is rRNA
- How would you eliminate or minimize rRNA in your cDNA library?

Methods of enriching for mRNAs

1. Selection of target RNAs via hybridization.
2. Removal of non-target RNAs via hybridization.
3. Copy-number normalization via duplex-specific nuclease digestion.
4. Target enrichment via size-selection

Methods of enriching for mRNAs

1. Selection of target RNAs via hybridization.
2. Removal of non-target RNAs via hybridization.
3. Copy-number normalization via duplex-specific nuclease digestion.
4. Target enrichment via size-selection

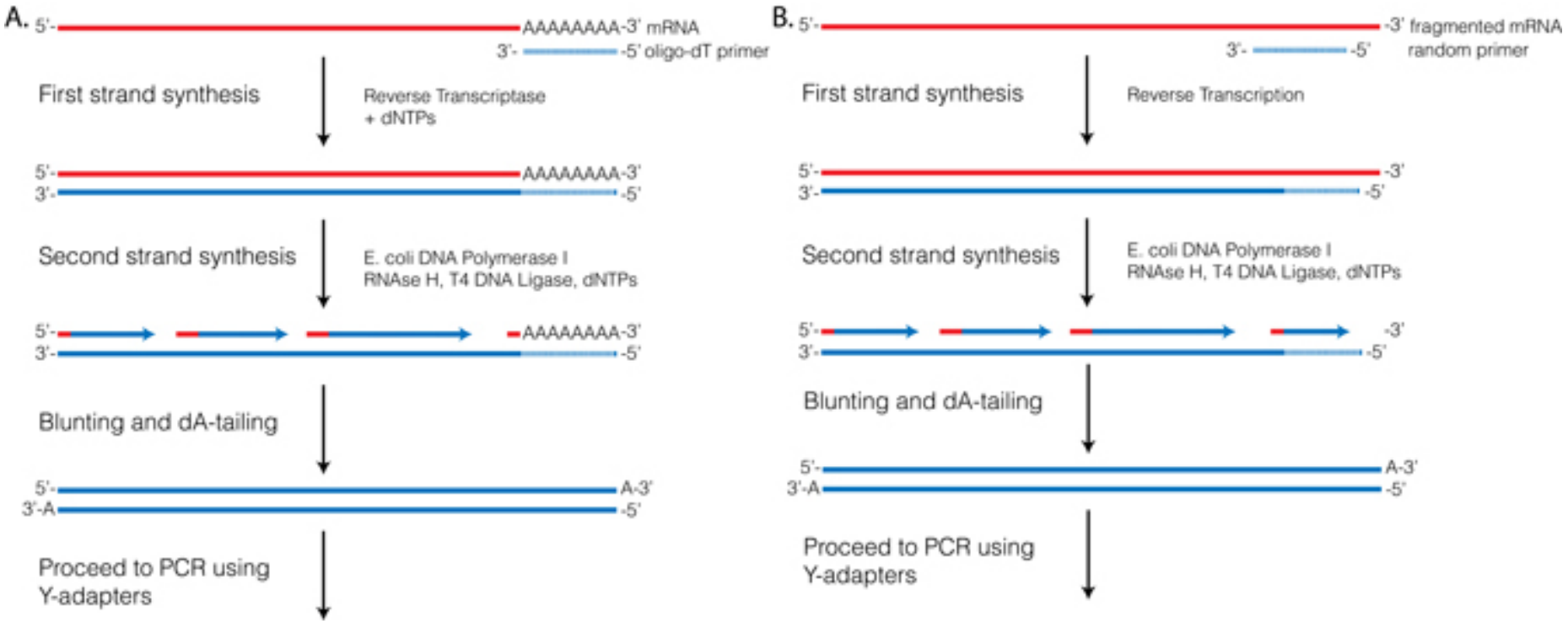
What kinds of RNA are going to be maintained or lost in each of these techniques?

1. Selection by hybridization to oligo-dTs
2. rRNA depletion by binding to oligos complementary to rRNA

Fragment the RNA

Early protocols fragmented the cDNA however, RNA is easier to fragment without bias, so more recent protocols are switching to that.

- Metal Ions (usually Mg^{2+}, Zn^{2+}, Mn^{2+})
- Heat
- Sonication
- Enzymatically
  - Often through the use of transposons



RNA-seqlopedia, http://rnaseq.uoregon.edu/#library-prep-first-strand-synthesis

Given the way that cDNA is made from RNA, which end of the mRNA are you least likely to see in your final sequencing data? Why? What data does this mean you might be missing? How might you try to target these sequences?

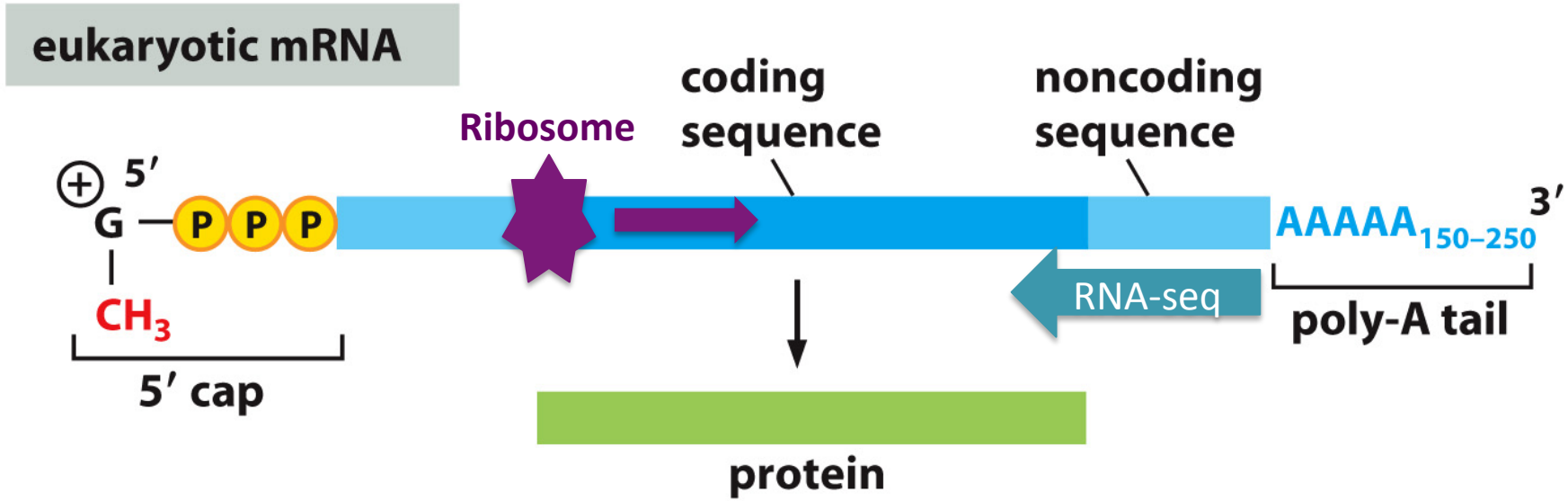


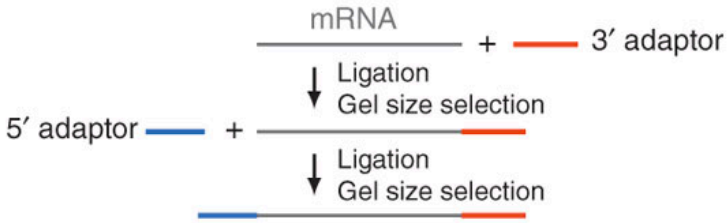
Figure 6-21a Molecular Biology of the Cell 6e (© Garland Science 2015)

Depending on where the primers bind, might not get to the 5' end. 5'UTR were less well represented in early RNA-seq data. New techniques for creating less biased data created.

a

RNA ligation[29]

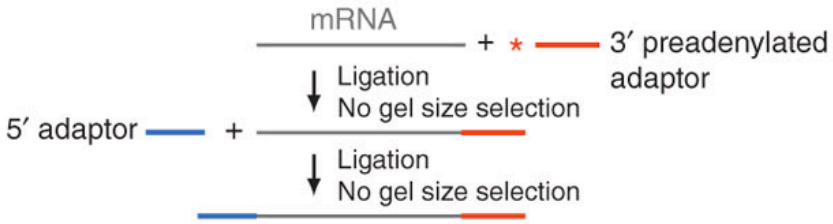
3′ and 5′ adaptors ligated sequentially to RNA with cleanup



Illumina RNA ligation

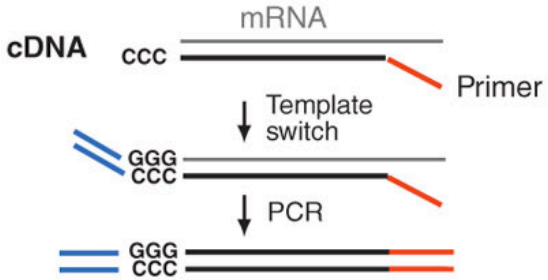
3′ preadenylated adaptors and 5′ adaptors ligated sequentially to RNA without cleanup

(S. Luo and G. Schroth, personal communication)



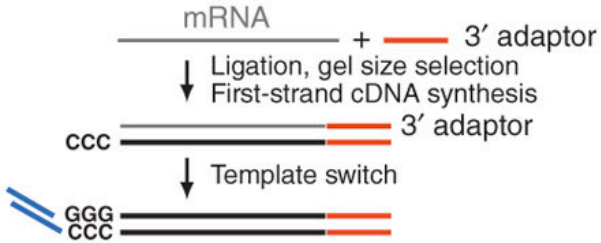
SMART[30]

Nontemplate ‘C’s on 5′ end of cDNA



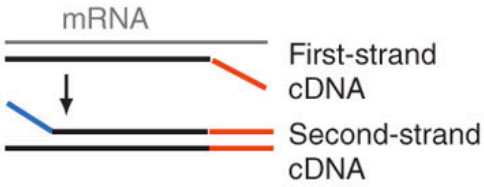
SMART–RNA ligation (hybrid)

Adaptor ligated on 3′ end of RNA and nontemplate ‘C’s on 5′ end of cDNA; template switching, PCR

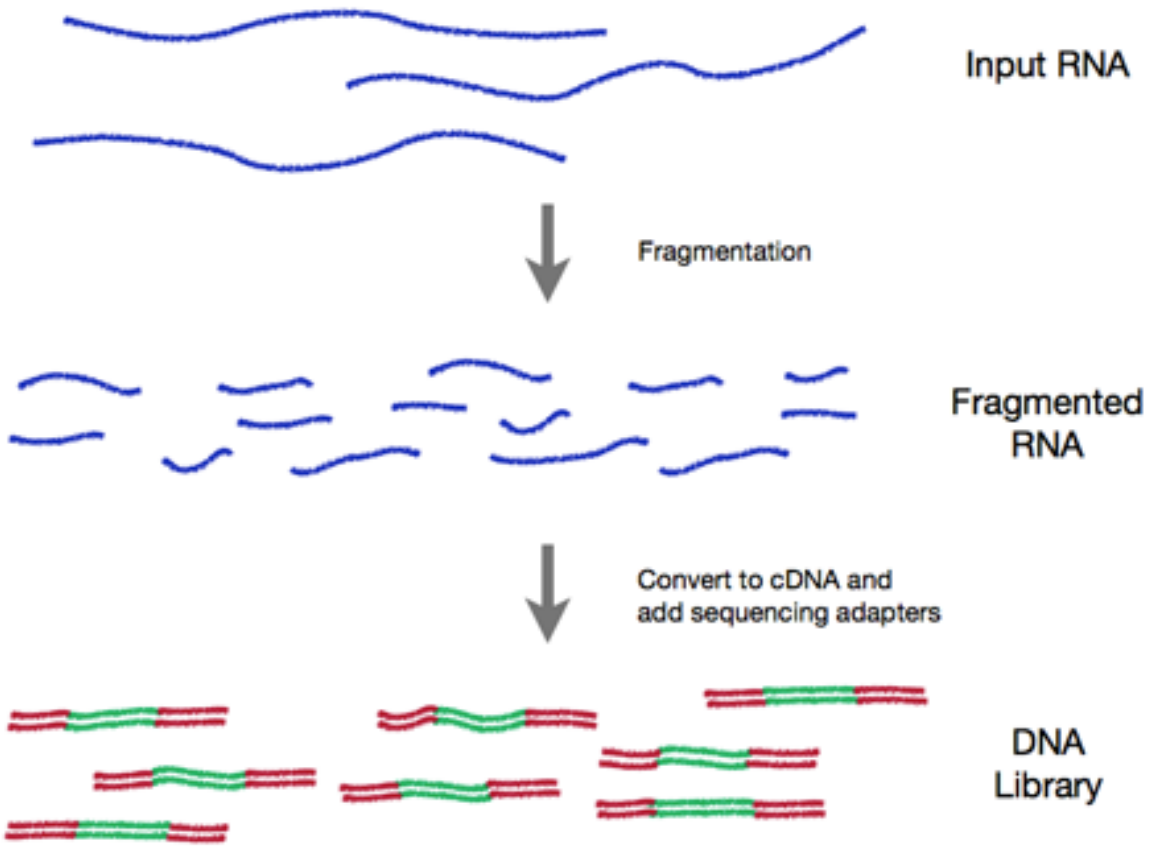


NNSR priming[31]

First- and second-strand cDNA synthesis with adaptors on ends of the primers

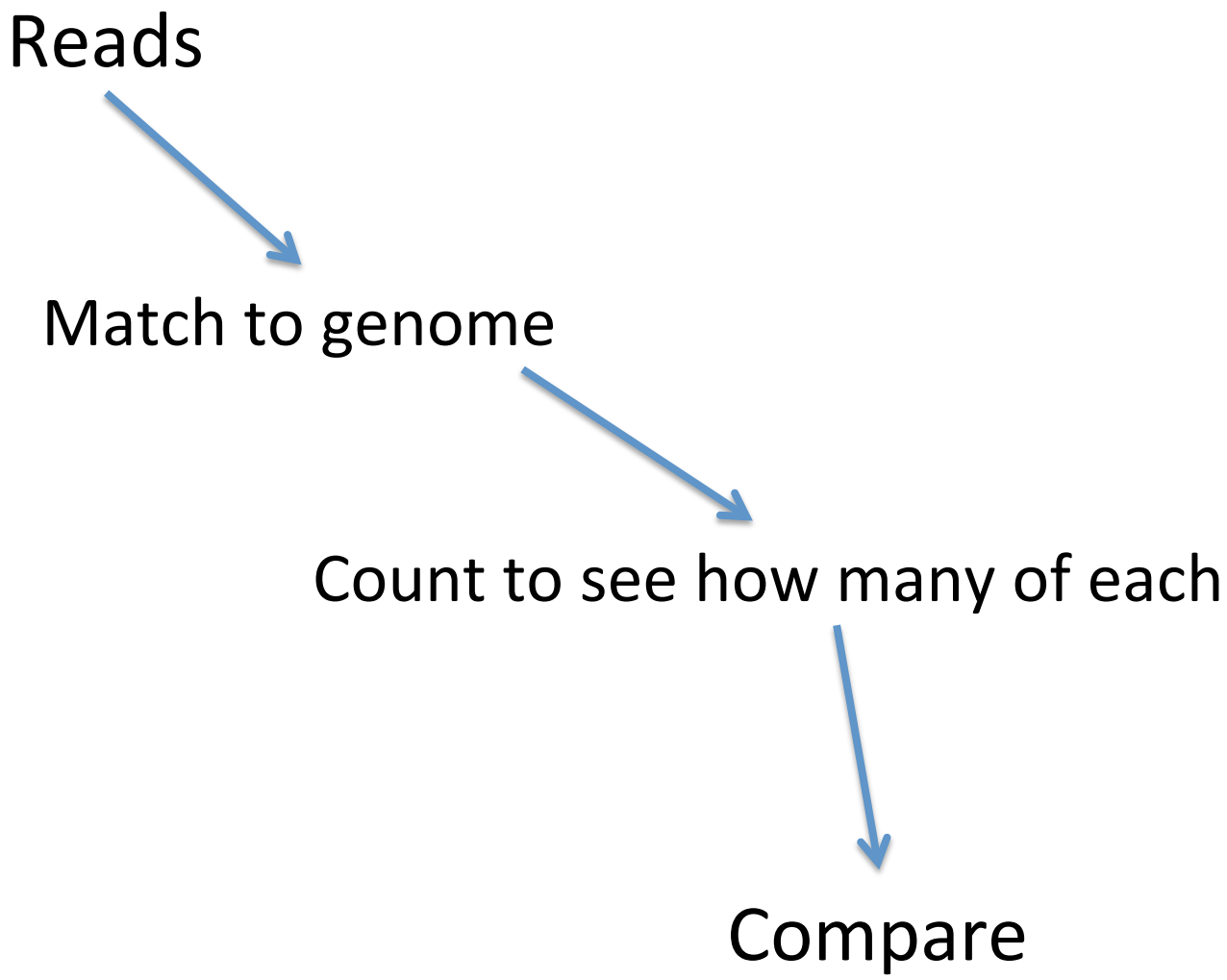


Levin, JZ, et al.; *Nature Methods* 7, 709–715 (2010) doi:10.1038/nmeth.1491

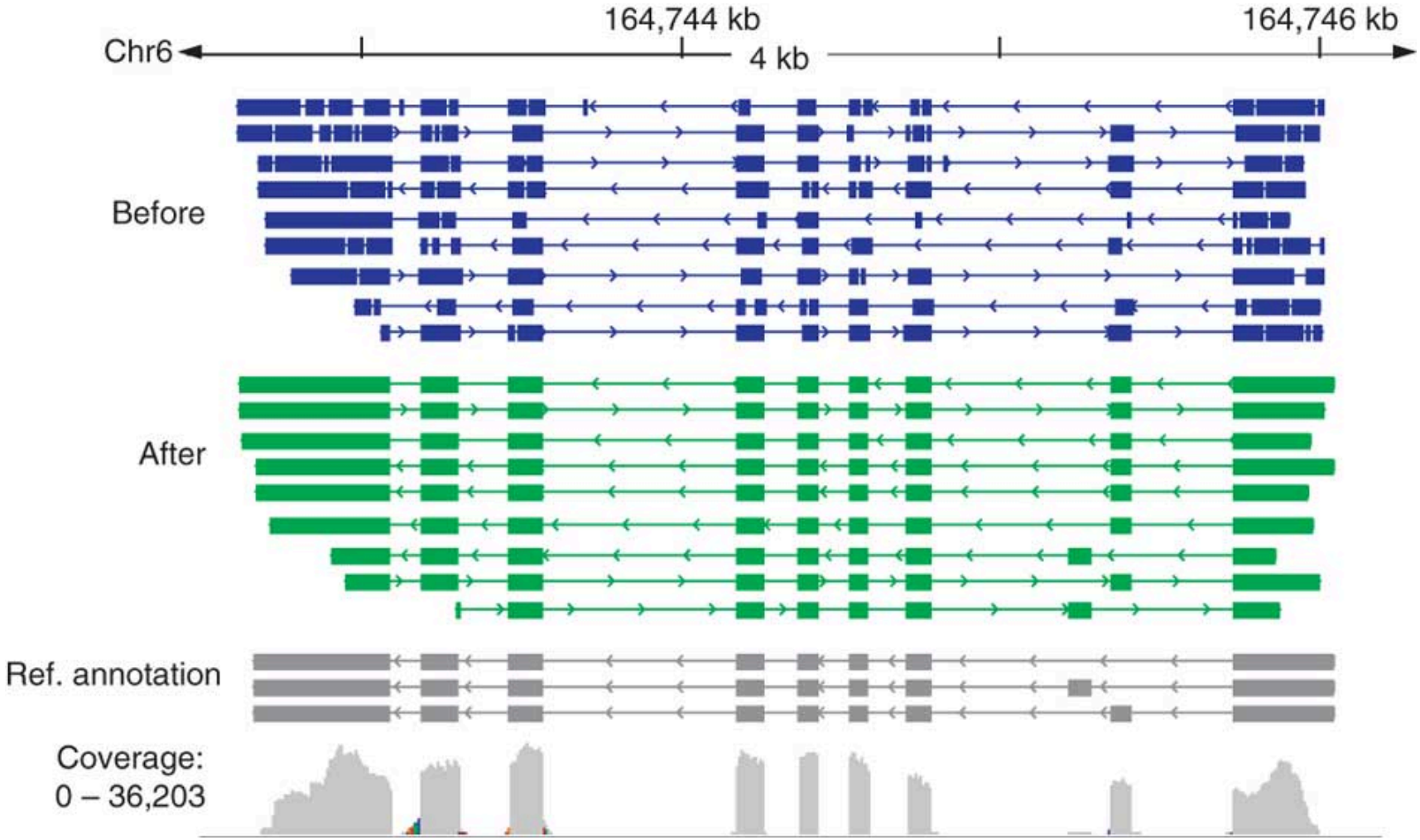


RNA-seqlopedia, http://rnaseq.uoregon.edu/

Sequence


Reads
Match to genome
Count to see how many of each
Compare

Data from RNA Seq



Koren, S. et al.; *Nature Biotechnology* 30, 693–700 (2012) doi:10.1038/nbt.2280

RNA-Seq vs Genome sequence

In a genome, each base has the same probability of being seen.

L length of the read, ***G*** length of genome

Is this true in RNA-seq? Why or why not? How will this change your library size calculations?

RNA-Seq vs Genome sequence

In RNA-seq the probability of seeing a given base will depend on the expression level as well as the length of each mRNA and the length of the entire transcriptome and number of the reads.

- High copy number RNAs dominate the reads
- To observe low copy number (low expression) RNAs, you need a larger library of reads (more coverage)
- Long RNAs will have more fragments in the sequencing population

What do we do with the reads?

Given what we said we wanted to know, what do we do next?

Transcriptome vs Exome

Transcriptome

- List of all the known mRNA sequences in a specific cell type
  - Includes each known splice variant separately

Exome

- All exons in an organism
  - Often determined from a set of transcriptomes
  - Often used for genetic disorders and cancer analysis

Have a reference genome

- Align to genome or transcriptome
- Can observe quantitative differences
- Can observe splicing differences
- Depending on reference used, can observe new genes, gene fusions, and splicing pattern

Why one type of reference versus the other? Why would transricptome be best in some cases?

No reference genome

- Generate *de novo* transcriptome
  - Use similar tools as genome
  - End up with many small contings
  - Put in some order
  - May not resemble genome in structure, but contains a list of many of the genes
- Can observe quantitative differences
- Can observe slicing differences

What would a transcriptome contig with 2 splicing isoforms look like (exon skipping)?

Summary