

# Mapping - part 2

Mark Stenglein, MIP 280A4

The command-line environment is hard at first, but you can do it!

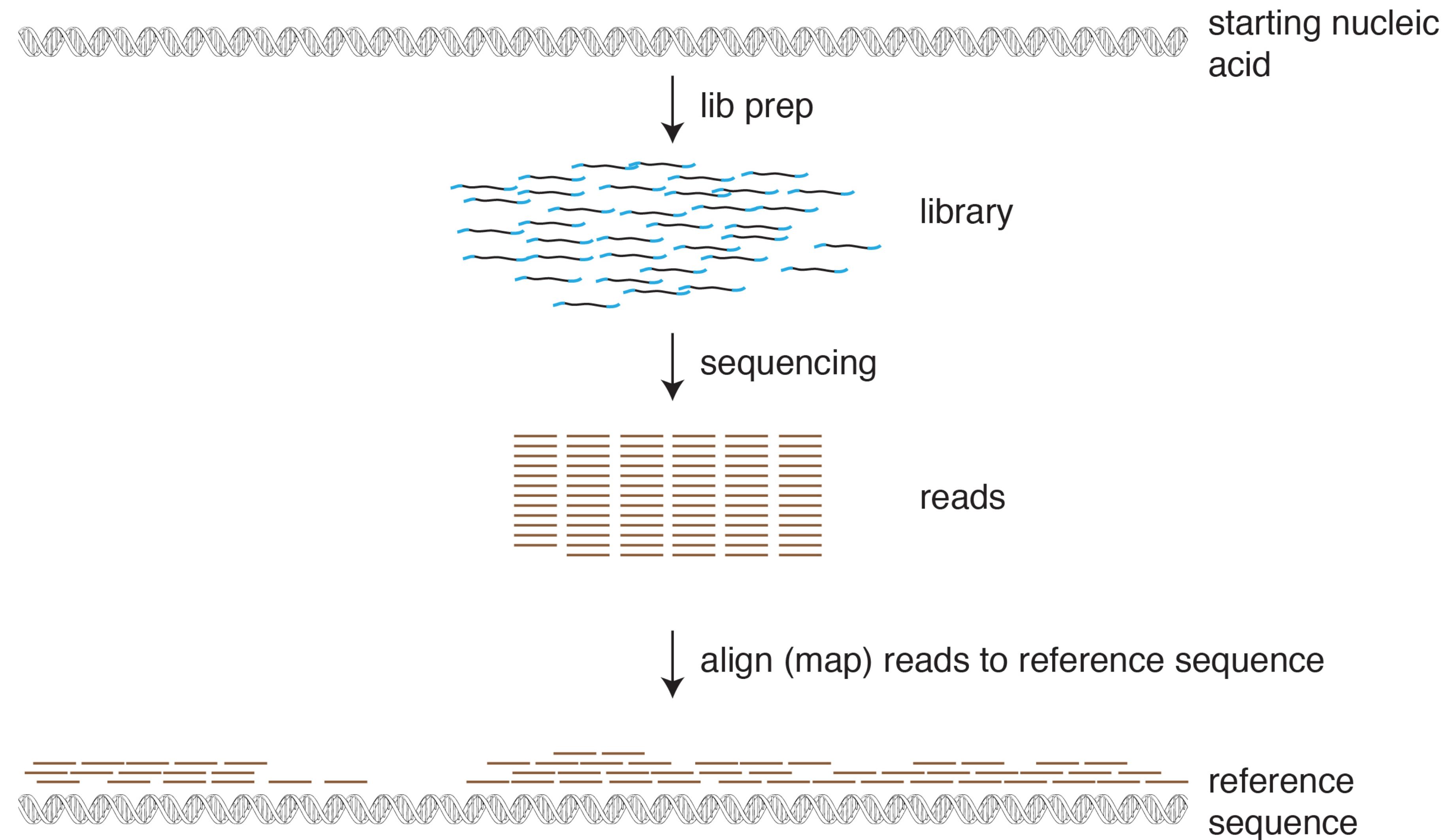
The screenshot shows a news article from Education Week. The title is "'Growth Mindset' Linked to Higher Test Scores, Student Well-Being in Global Study". Below the title is a photo of a woman, Sarah D. Sparks, and the text "By Sarah D. Sparks — April 09, 2021 4 min read". The main image is a graphic comparing fixed and growth mindsets. On the left, a blue head silhouette is split vertically; the left half contains a red brain network with a padlock, labeled "Fixed Mindset". On the right, a blue head silhouette is split vertically; the right half shows a person climbing a ladder to a flower pot with a gold coin, labeled "Growth Mindset".

<https://culturelligence.com/to-grow-and-understand-your-mindset/>



<https://www.sproutbrite.com/products/the-power-of-yet-classroom-banner>

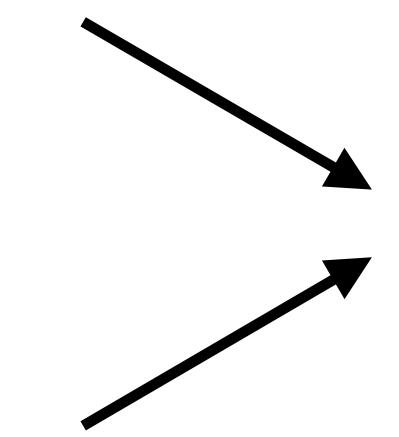
**Mapping** is the process by which sequencing reads are aligned to the region of a genome from which they derive.



## Mapping inputs

Reads  
(Or existing sequences)

Reference sequence(s)

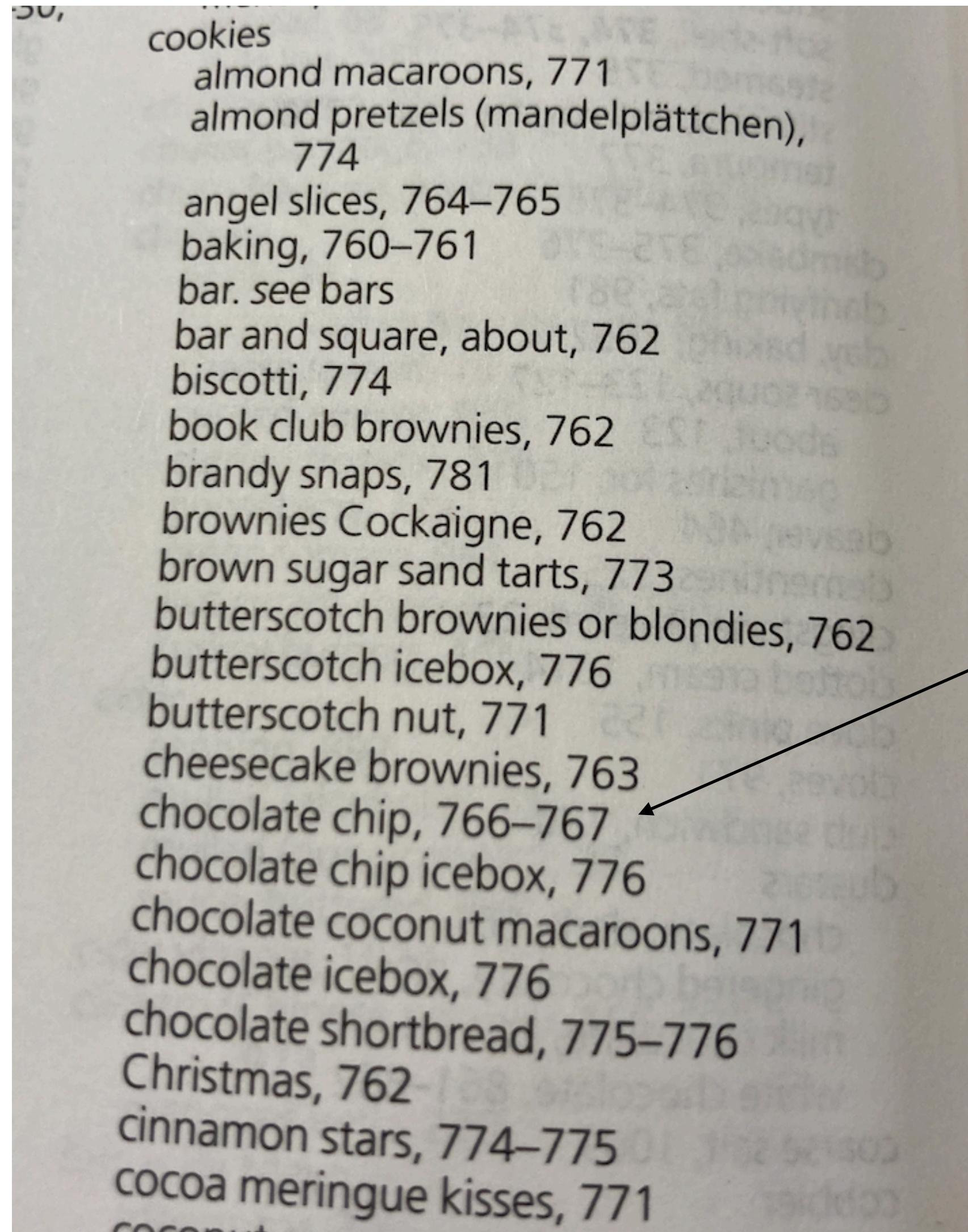


## Mapping

## Mapping output

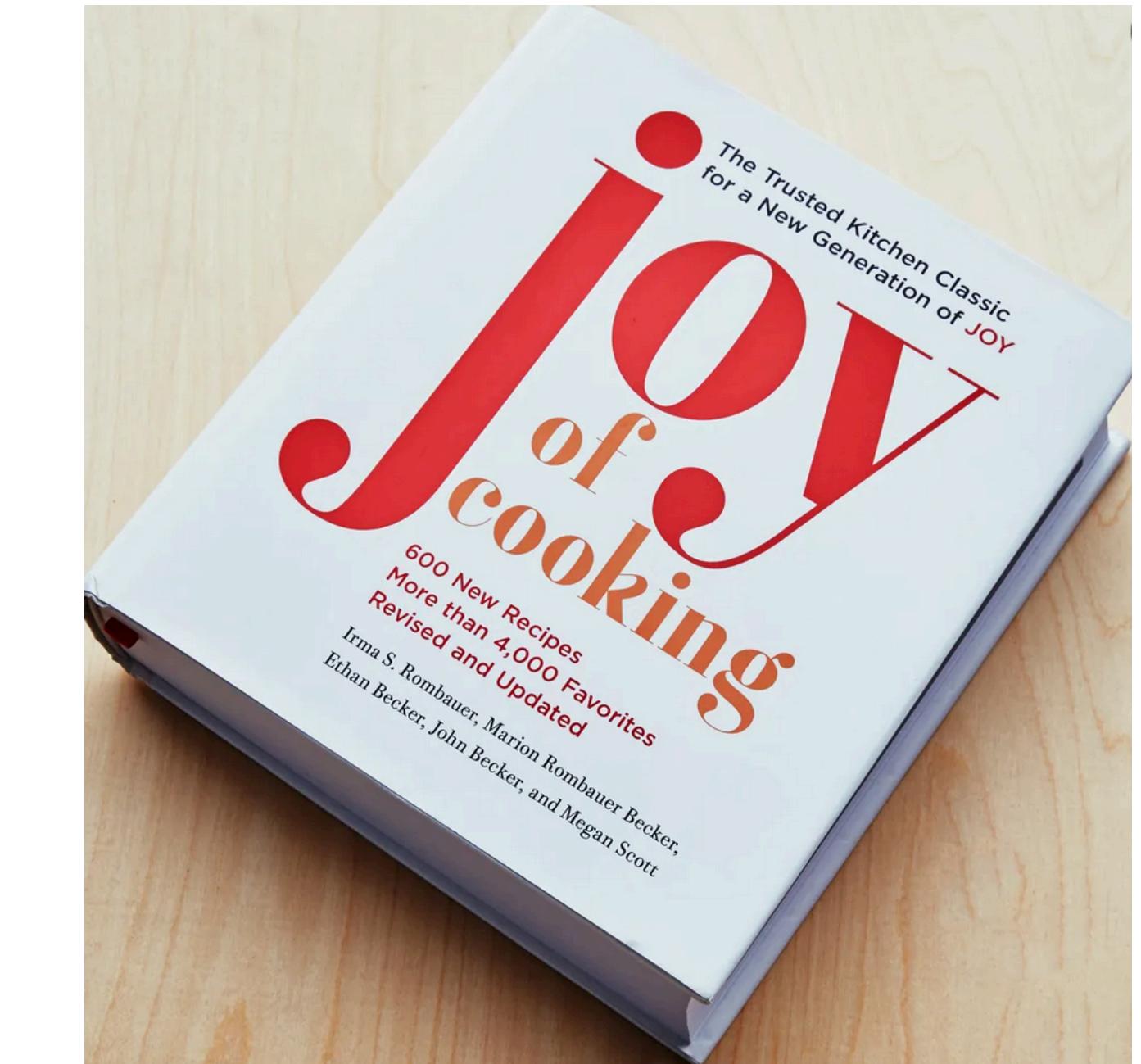
Does each read map?  
Where on the ref. seq. does it map?  
*How well* does it map?

Mapping tools map fast because they **pre-index** reference sequences

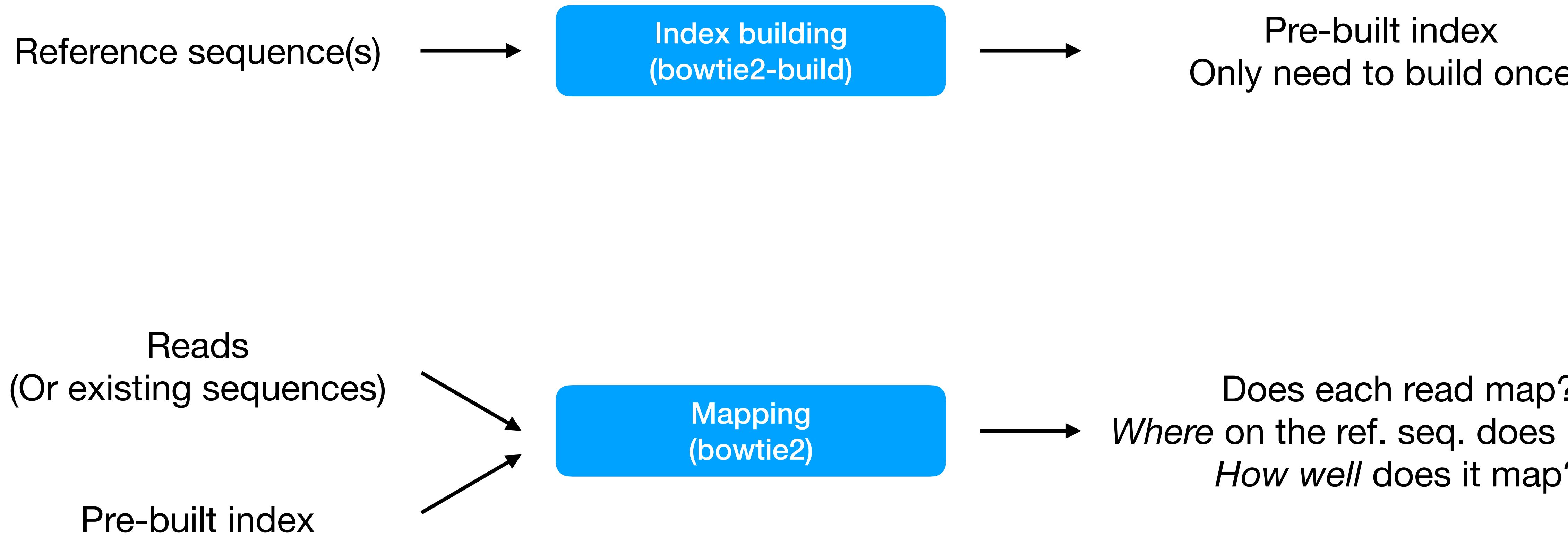


Indexes help you find things faster

chocolate chip  
cookies on page  
766



BLAST databases are another example of pre-built indexes



The output of mapping software is SAM format files

Header lines, start with @, provide info about the reference sequences and the mapping software used

After header lines, one line per mapped read, with 11 columns separated by tab

SAM column	Info
1	The read's name
3	The mapped-to reference sequence name
4	Position in the reference sequence where the read maps
5	Mapping quality score
6	Whether there are mismatches to the reference sequence
10	The read's sequence
11	The read's basecall quality scores