CMSC423

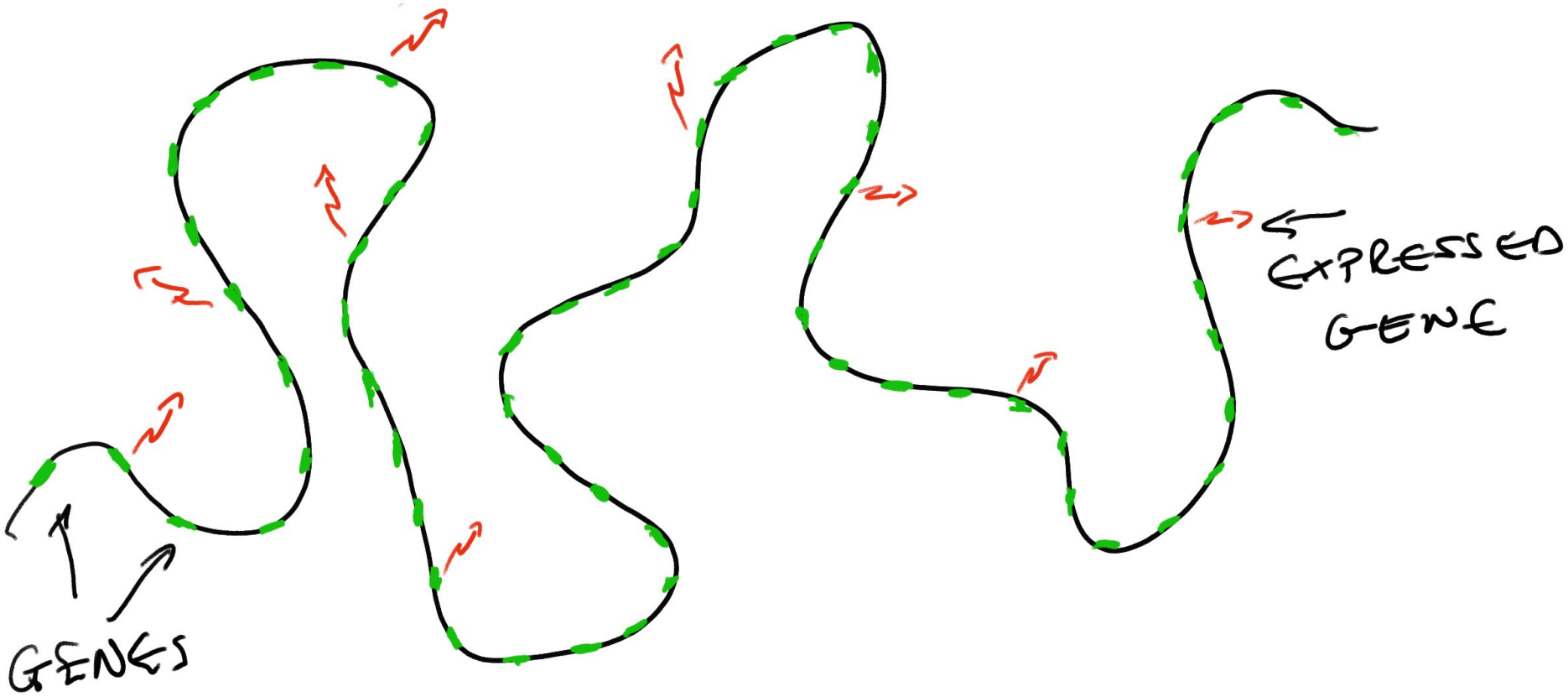
Chapter 2 – Introduction to motif finding

Recap

- Chapter 1 – look for "interesting" regions in a genome (regions where some patterns are frequent)
- Z+KMP – look for a specific pattern in a genome
- Here:
  - we don't know what we're looking for
  - signal spread out throughout the genome

This week

- Find something that's common to many pieces of DNA



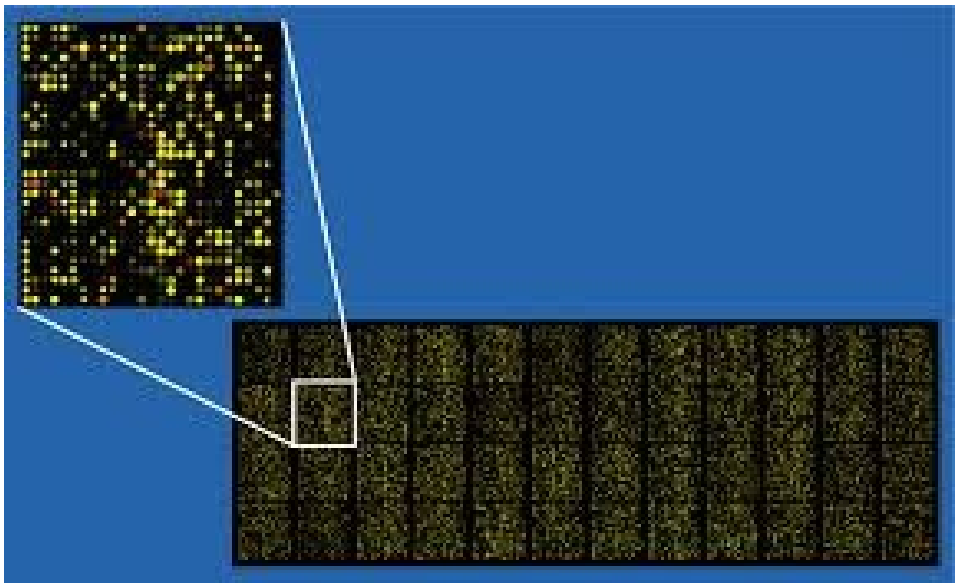


image from Wikipedia

Stop and Think

- Can you use what you've learned in class so far to solve this problem?

Why earlier approaches don't work

- We don't know the motif (KMP and Z don't work)
- Motifs are too inexact (frequency doesn't work)

The solution

- Randomized search:
  - take a random string from each upstream region
  - check the score of the profile
  - repeat until we find highest scoring profile

- What we need:
  - define the profile
  - define the score
  - come up with a search strategy

Next: better defining the problem