

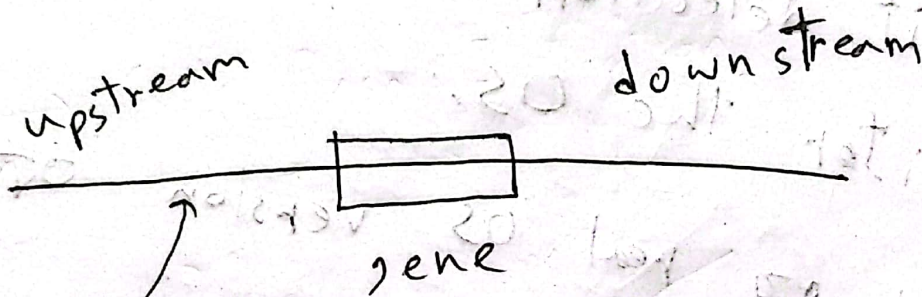
BioInfo

31/1/24

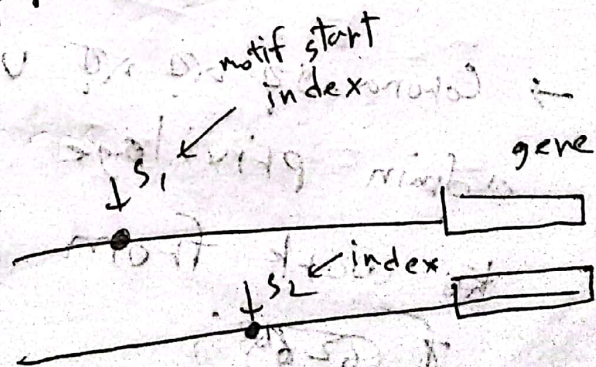
$(15, 4)$ Motif \rightarrow 15 length pattern
with at most 4 mismatches in
each pattern.

\rightarrow Biological problem এ এটি আর
ঠাঠা জালা যায না।

* ଏହା cell ଏ, ଏହା group of
gene high-rate ଏ express କରୁ,
ଏହା cell to cell functionality
vary କରୁ, ଏହା whole genome
ଏ cell ଏ same.

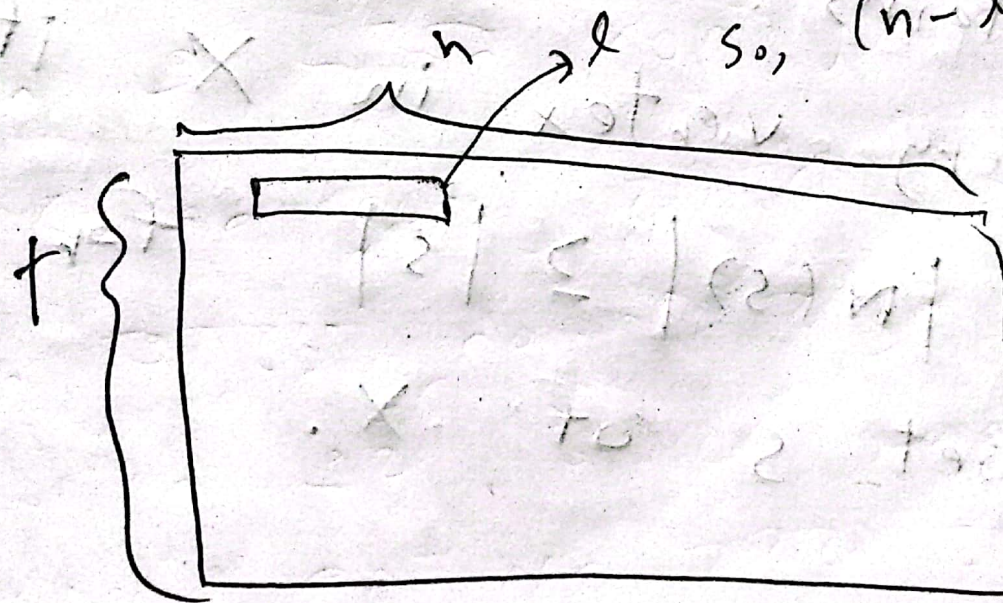


Upstream ଏହା କରୁ sequence
ଏହା ଏ enzyme ଏହା attach
କରୁ, ଏହା ~~expression~~ 3' ~~gene~~
gene ଏ expression rate କରୁ



BioInfo

- Pairwise Comparison for mismatch
~~two~~ than = real motif and ~~for~~ present
sequence comparison.



So, $(n-l+1)$ possible motifs in 1 row.

So, search space

$$(n-l+1)^t$$

motif logo

Height in motif logo = 2 - entropy



because
we have 4
characters. So
2 bits.

Row for score
median string problem

5/2/24

BioInfo

Greedy Motif Search

$$(n-k+1) \times t \times (n-k+1)$$

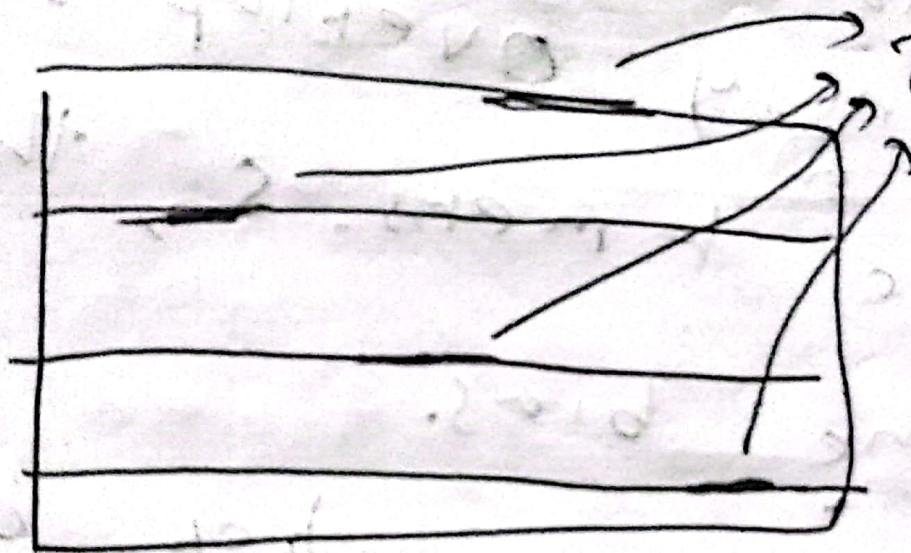
$$= t(n-k+1)^2$$

Randomized motif search

$t(n-k+1)$ on each iteration

Bio Info

Assumption:



True motif

True motif
not over

particular
atgc
random
uniformly
distributed

According to assumption, for ~~parts~~ parts not motif, profile will be almost:

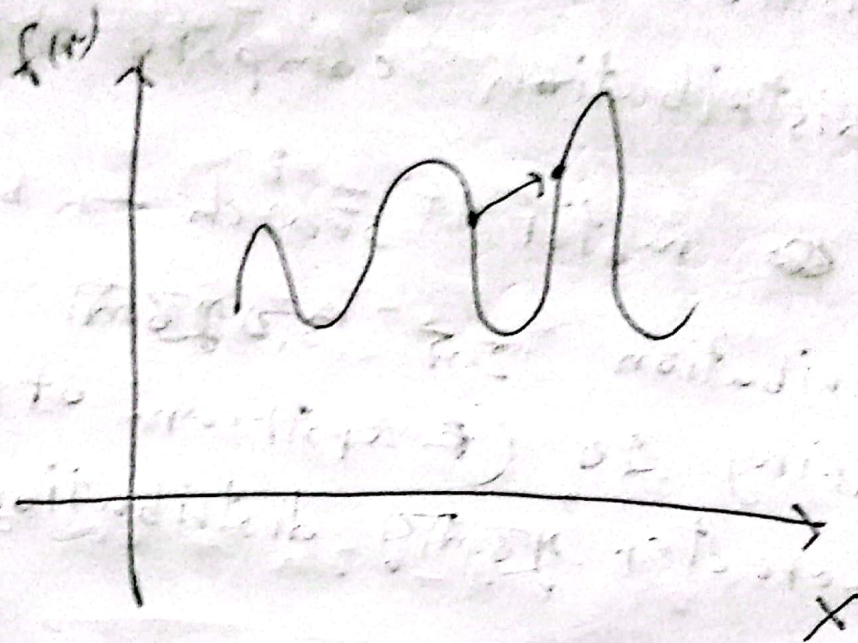
A	.25	.25	.25	.25
C	.25	.25	.25	.25
G	.25	.25	.25	.25
T	.25	.25	.25	.25

It will not be good, scores will always remain same for all position.

• But randomly choosing initial k-mers, generally will not give this. Because some part of the motifs may overlap some part of these k-mers. So, there will be some bias.

• ~~some~~ some number of motifs
chose k-mers as 3500

କିନ୍ତୁ ନାହିଁ, ତାହାଙ୍କ ମଧ୍ୟ ଡାହାଣାଡ଼ା
algorithm ଅନୁସାରେ ଆମର,



Problem: ଏହି algo ଠିକ୍ ଏବଂ
iteration ଏ ~~target~~ assumed
motif ଏ drastic change ଥିବ
ନାହିଁ, ଏବଂ search space ଏ
ଅନ୍ତରାଳ ଠିକ୍ move ନାହିଁ, ଏବଂ
converge ନା ଥିବ critical point
ଥାଏ ନାହିଁ ~~ନାହିଁ~~ ଠିକ୍ ନାହିଁ
ନାହିଁ,

solution \rightarrow Gibbs Sampling

KL-divergence \rightarrow ଏକ background distribution ଏବଂ ଆମ ଆବୃତ୍ତି distribution compare କରୁ.

for motif search \rightarrow background distribution ର ଆବୃତ୍ତି input string ଏବଂ (upstream of all genes) character frequency distribution.

\rightarrow ଯେଉଁ letter ଥିବା ଆଉ overall, ଯେ motif ଏବଂ ଥିବା ଥାଏ, ଏହି system ଦ୍ଵାରା ପଢ଼ା ଯାଏ.