Map out ~10 CCS
to see what/where certain regions are

Final   HMM

start as equal

2-B01
3-C01
4-D01

5'adapter → 5'barcode

RNA

3' barcode

start as equal

RNA
RC

3' barcode
RC

5'barcode RC → 5'adapter RC
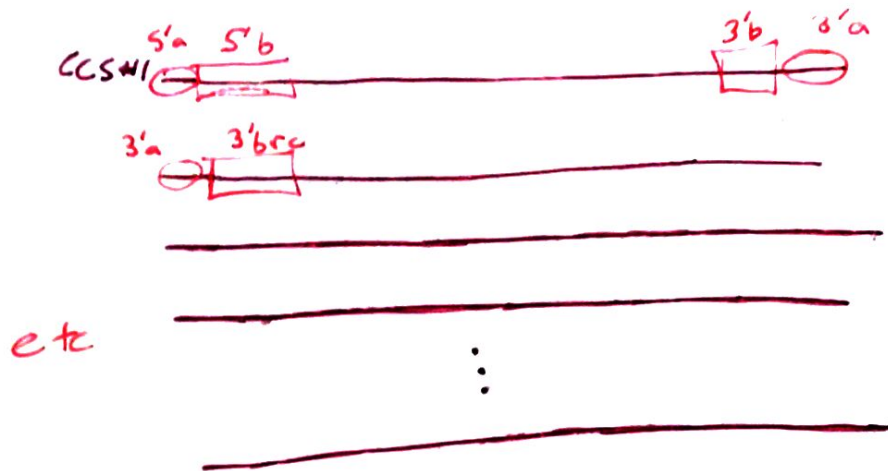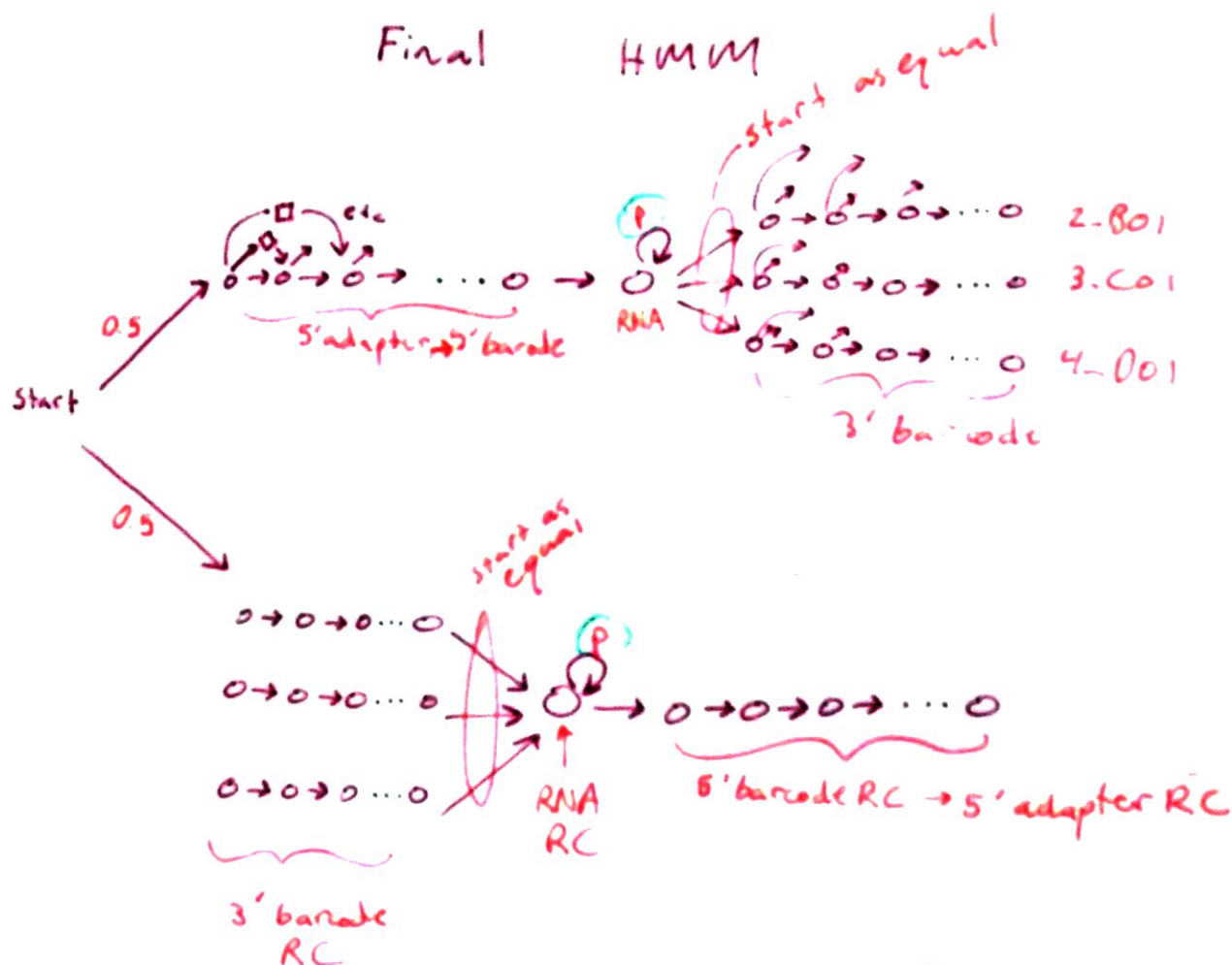
(P): geometric distr of length
↳ not sure how to impliment

The longer the CCS, the
more likely   RNA   self-cycles.
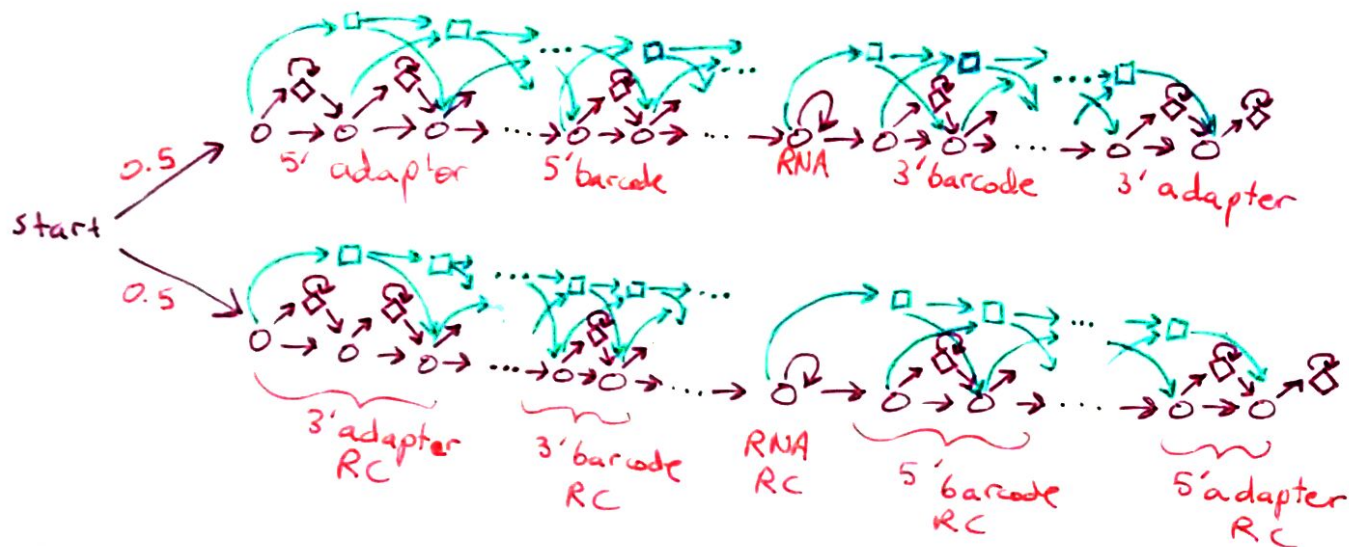
Let ℓ = avg len of our CCSs

$$\tilde{p} = e^{\ell}$$

$$p = \text{scaled } \tilde{p} = \frac{e^{\hat{p}}}{1+e^{\hat{p}}}$$

totally arbitrary idea of mine

prob of self-cycle

e^x

length of ccs

Phase One HMM

I HMM for each cell line:

Emission probs

→ "RNA", insert sections:
$$A = .25$$
$$C = .25$$
$$G = .25$$
$$T = .25$$

→ All others: based on reference seq:

ex say 5' adapter is   A C G C T A T

Probs:

A: .05/3   A: .05/3
C: .05/3   C: .05/3
G: 0.95   G: .05/3
T: .05/3   T: 0.95

Transition probs

match → insert : 0.05

match → delete: 0.05

match → match : 0.90

insert → insert : 0.5

insert → match: 0.5

delete → delete: 0.5

delete → match: 0.5

RNA → RNA : p

RNA → match(3' barcode): $1 - p - 0.05$

RNA → delete: 0.05

p:

$$\tilde{p} = \frac{e^l}{\bar{p}} \quad (l: \text{avg len of ccs})$$

$$p = \frac{e^{\tilde{p}}}{1 + e^{\tilde{p}}}$$