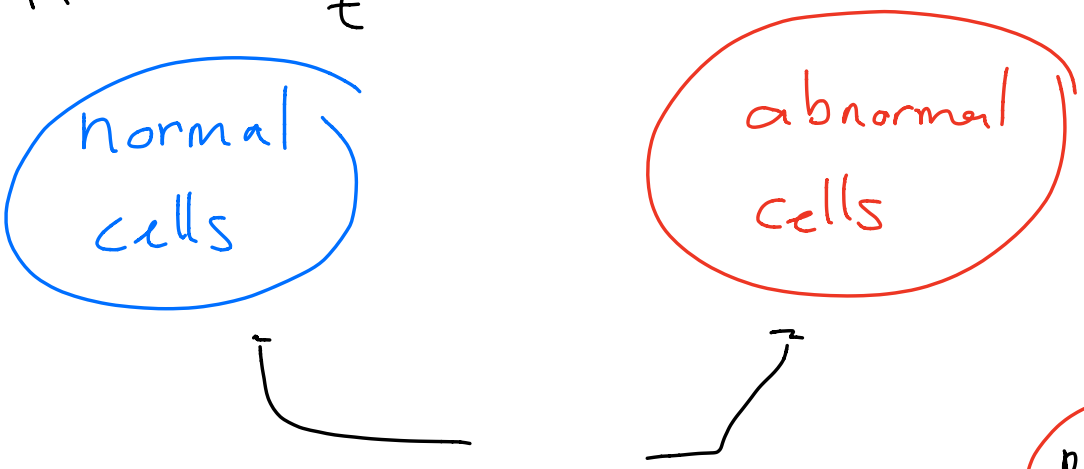
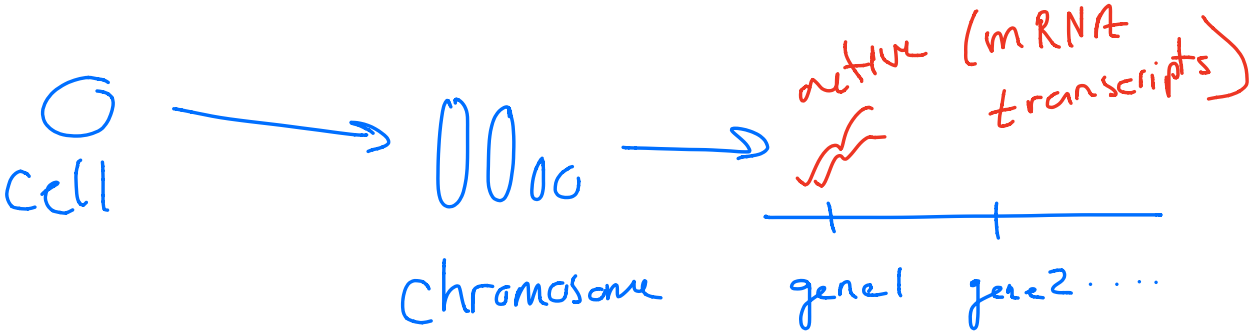
RNA-Seq



- behave differently
- Genetic Mechanism?
- Differences in gene expression



High throughput tells us

(i) which genes are active

(ii) how much they are transcribed

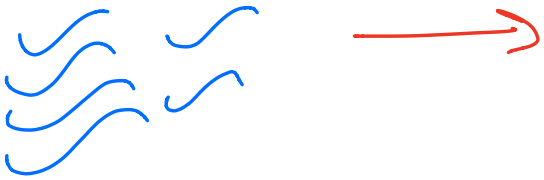
Task: Use RNA-seq to measure gene expression in (non) normal cells for comparison

Steps to RNA-Seq

1. Prepare library
2. Sequence
3. Data Analysis

Prepping RNA-seq Library:

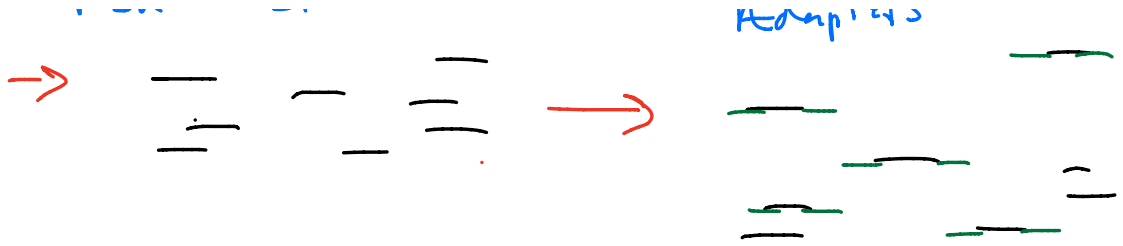
Step 1: Isolate RNA



Step 2: Break into fragments

Step 3: Convert to Double Stand DNA

Step 4: Add Seq. w/ Line



* Doesn't work all the time

Step 5: PCR Amplification

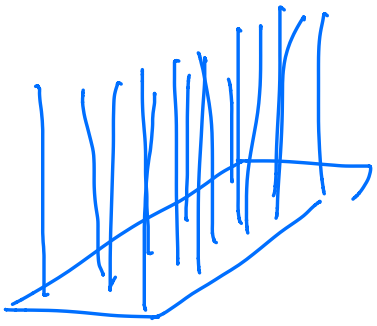
→ Amplifying segments w/ adapters

→ Step 6 QC

Sequence the Library

G
C
A
G
C
A
⋮

Goal Sequence this strand



flow cell

Works like elevator reading each "level" producing a set of slices

Issues: · Reading each cell correctly
· Low diversity (clustering on the flow grid)

Alignment:

1. Start with genome OOoo gattacata...
2. Split into small fragments

   gattac attaca ttacat ...
3. Create index and hashing
4. Split read from RNA-seq
5. Match segments in (4) to (2).

Why breaking up segements?

- Fuzzy matching / hashing

Count Reads per Gene

After counting reads per gene

$X =$

Gene id.	Sample1	Sample2	...
KIBG	24	30	
ASM	13	5771	
⋮	⋮	⋮	

Rmk: Will need to normalize this mat. based on VST.

Analysis: PCA plots, DESeq2