

# Project 2: influenza & deep sequencing

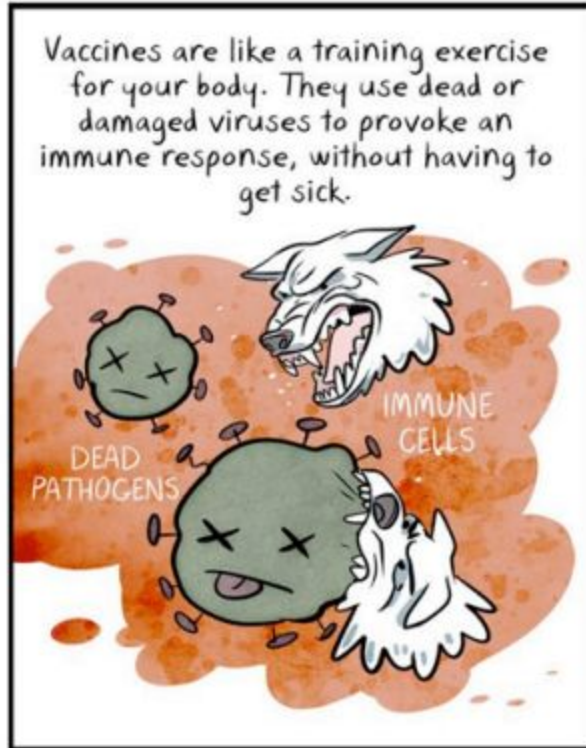
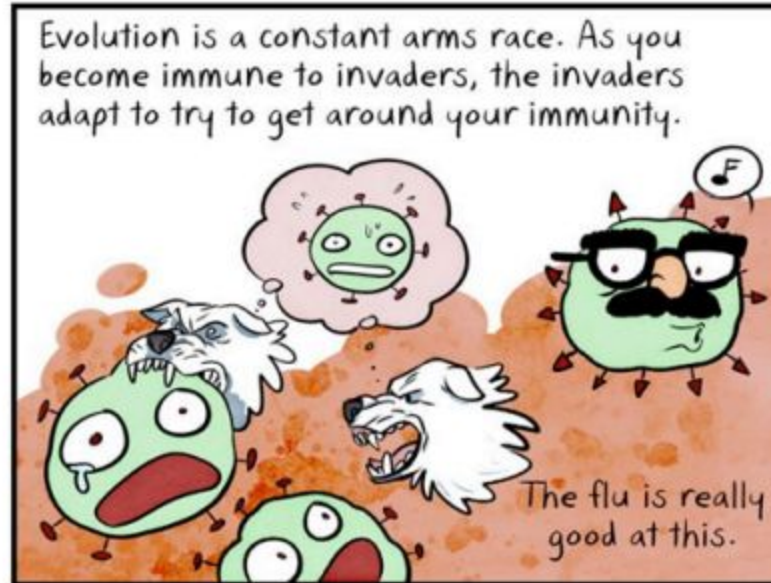
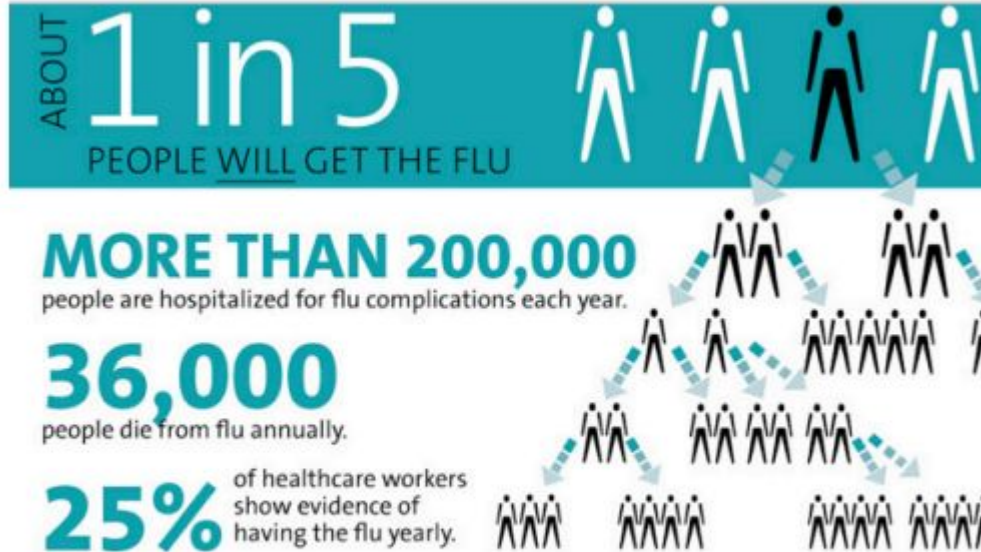
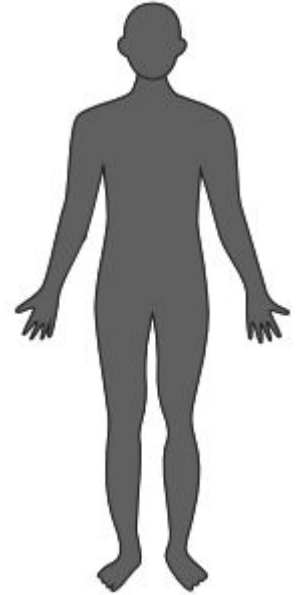


image credit: Maki Naro

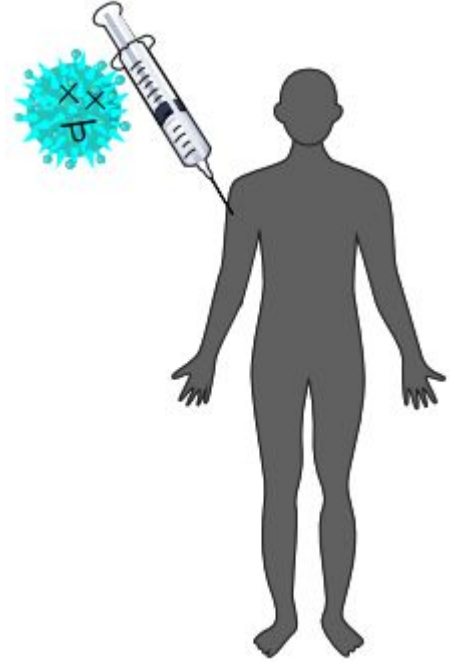
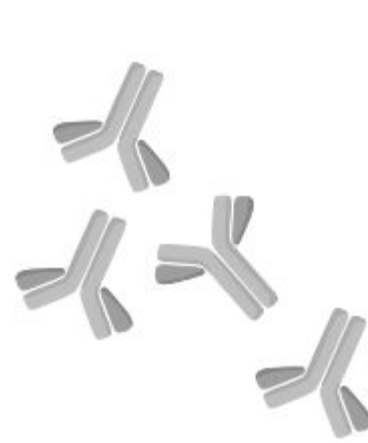
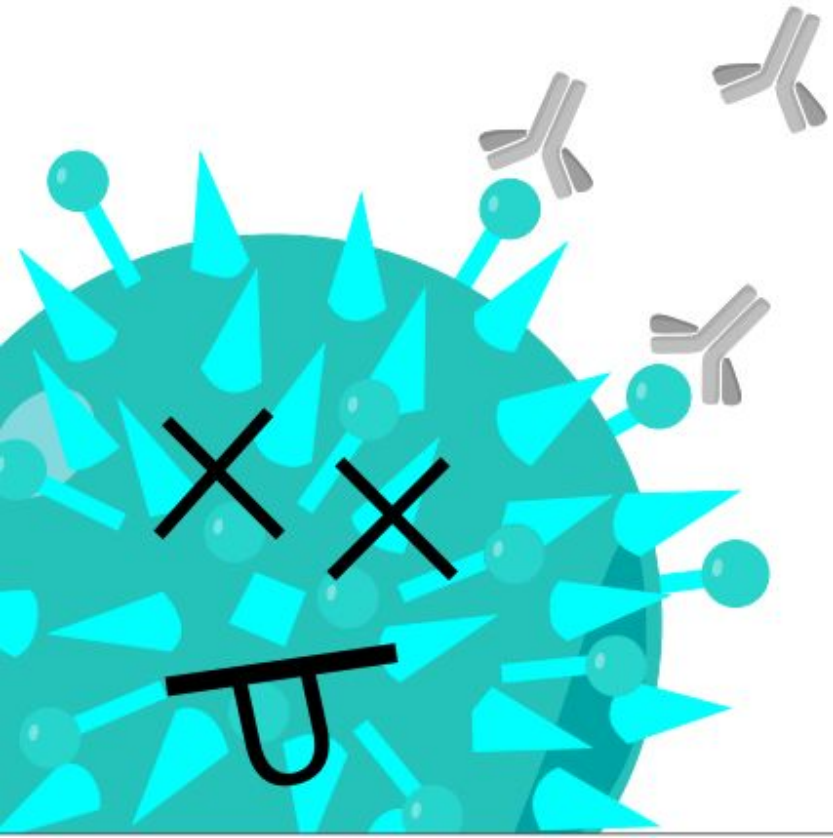


# bio/ Influenza virus

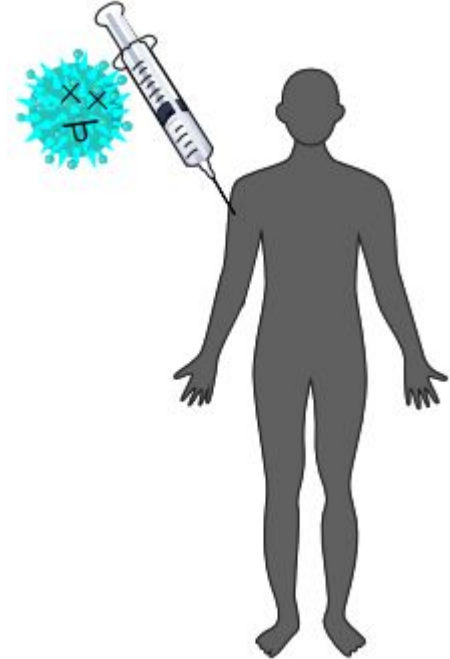
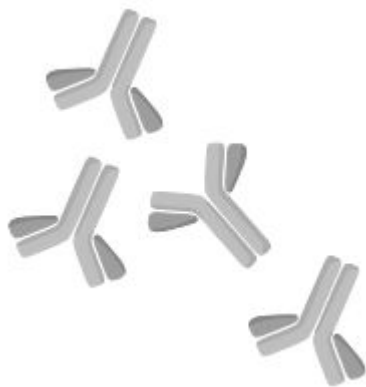
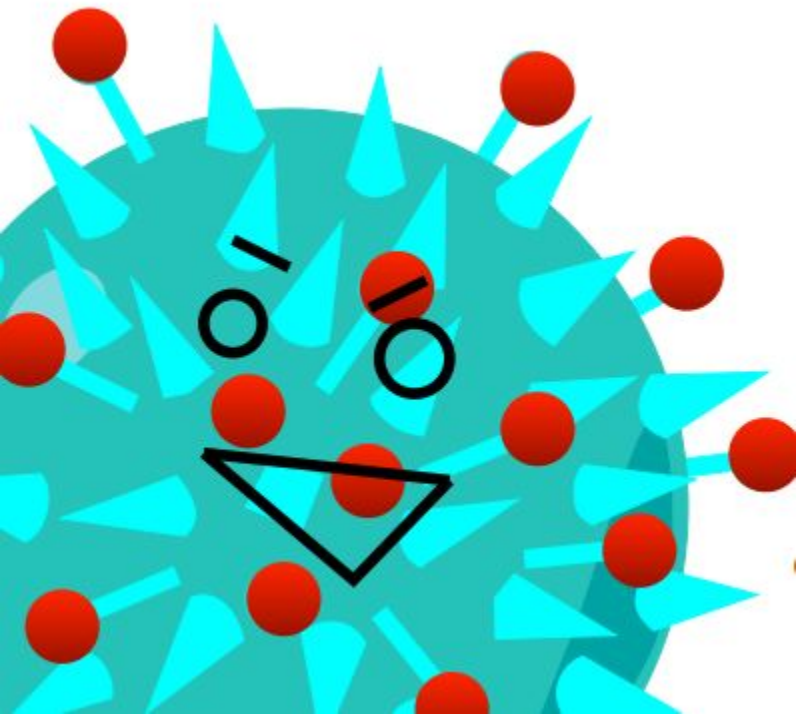


THE CDC SAYS THAT MOST HEALTHY ADULTS CAN INFECTION ANOTHER PERSON  
**ONE DAY BEFORE** SYMPTOMS EXIST AND FOR **A WEEK AFTER** THE ILLNESS BEGINS.

# bio/ Influenza vaccination



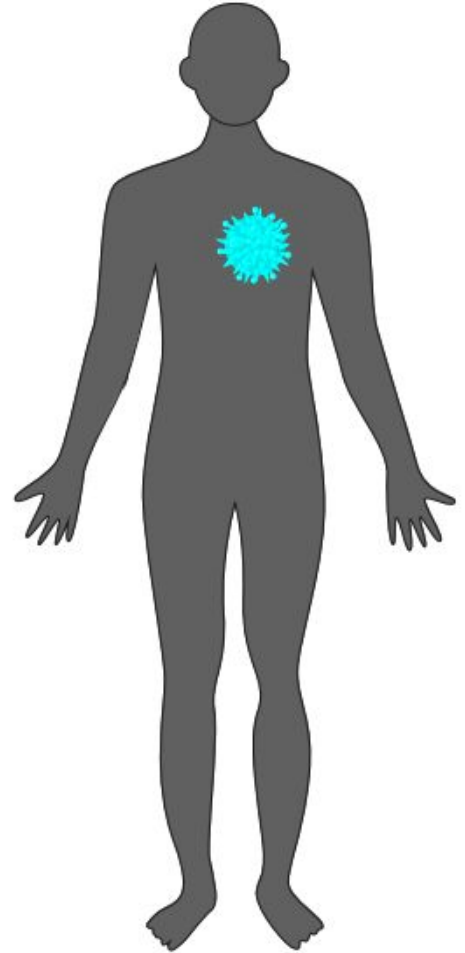
Antibodies to influenza target the viral surface protein hemagglutinin (H)



Epitope mutation prevents antibodies from binding

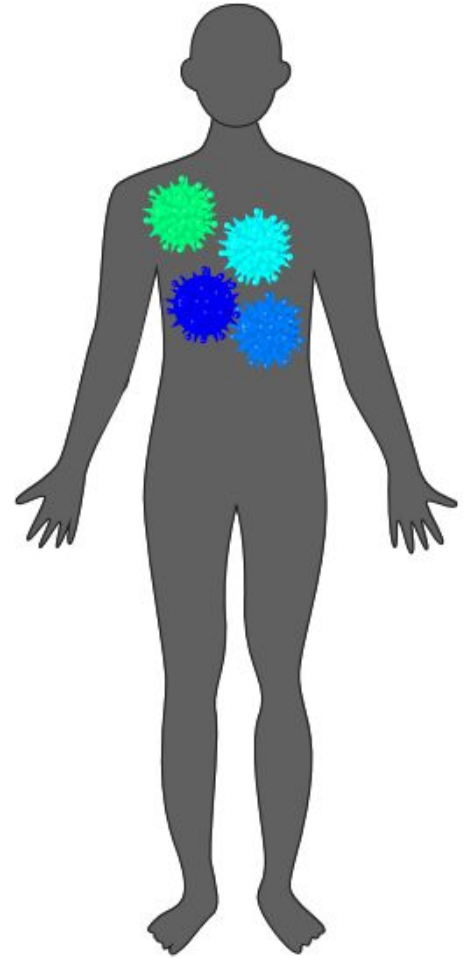
## bio/ Influenza mutation rate

- influenza mutates at a rate of one mutation per genome per replication



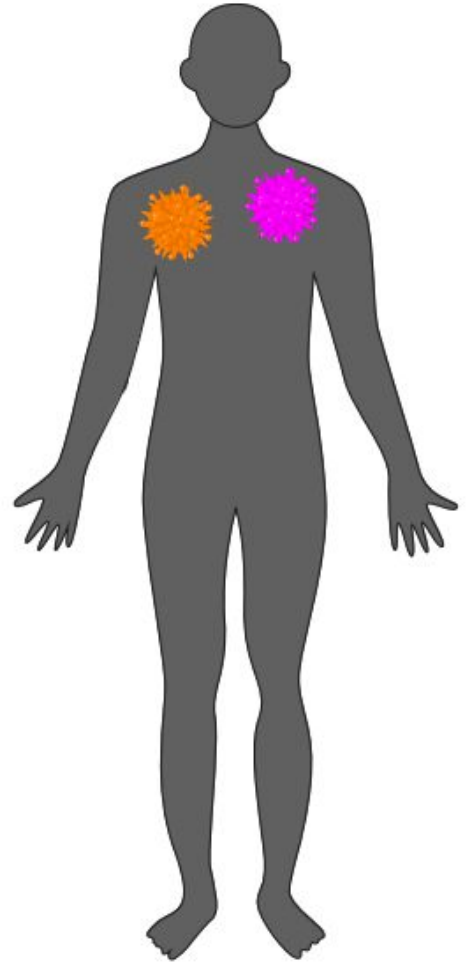
## bio/ Influenza mutation rate

- influenza mutates at a rate of one mutation per genome per replication
- exists as quasispecies, even within a single host



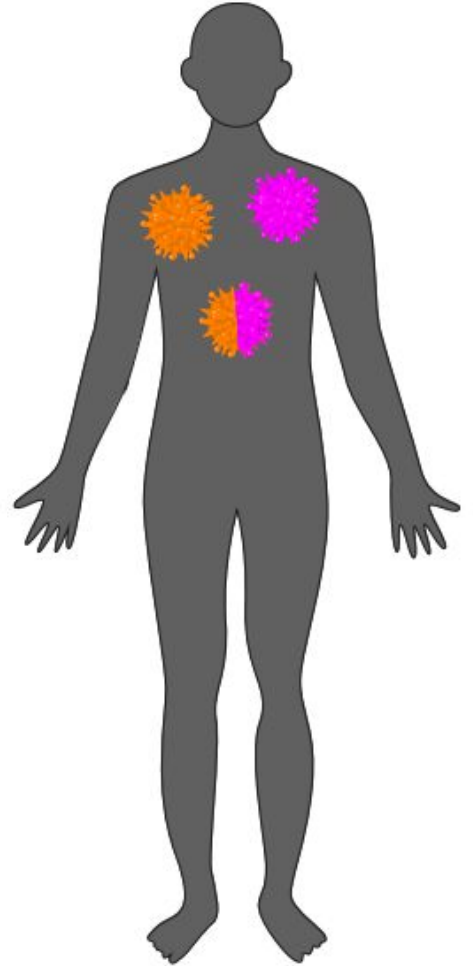
## bio/ Influenza mutation rate

- influenza mutates at a rate of one mutation per genome per replication
- exists as quasispecies, even within a single host
- if more than one strain of flu infects the same host,



## bio/ Influenza mutation rate

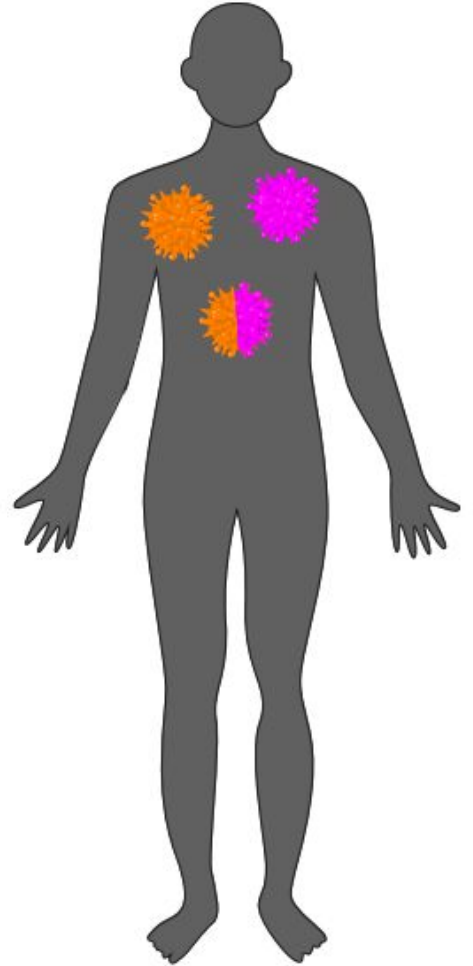
- influenza mutates at a rate of one mutation per genome per replication
- exists as quasispecies, even within a single host
- if more than one strain of flu infects the same host, they can recombine





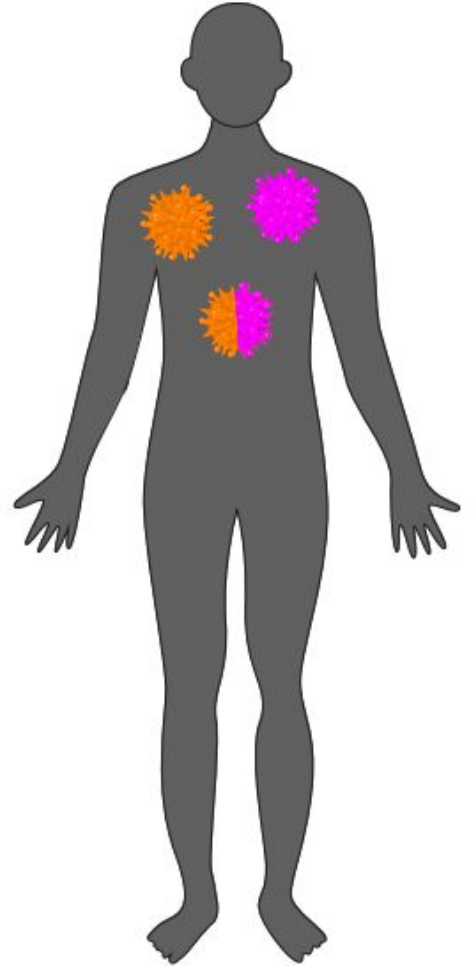
## bio/ Influenza mutation rate

- Antigenic drift: is the result of new mutations that are introduced into currently circulating strains over time



## bio/ Influenza mutation rate

- Antigenic drift: is the result of new mutations that are introduced into currently circulating strains over time
- Antigenic shift occurs when genetic segments are mixed with novel components, sometimes between species.



## bio/ Influenza mutation rate


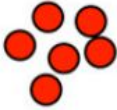

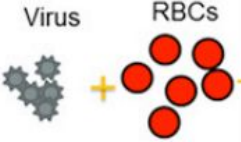


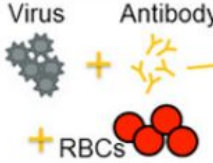
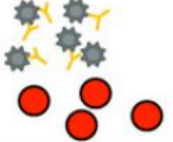

- Antigenic drift is attributable to influenza's low fidelity RNA polymerase that lacks a function for error proof-reading.
- With each replication cycle, polymerase errors create de novo mutations, increasing the genetic diversity of the virus.



RNA Replicase structure [PDB: 3PHU](#).



# bio/ Hemagglutination inhibition assay

Components	Interaction	Microtiter Results
RBCs 		No Reaction 
Virus + RBCs 		Hemagglutination 
Virus + Antibody + RBCs 		Hemagglutination Inhibition 

Use antibodies that specifically bind to each known flu strain.

Test unknown sample against each set of antibodies separately.

# bio/ Hemagglutination inhibition assay

test virus + Ab to strain A

test virus + Ab to strain B

test virus + Ab to strain C

test virus

no virus

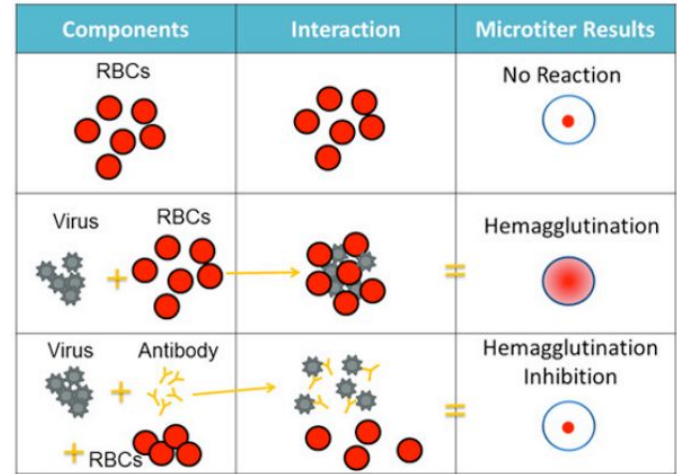


Use antibodies that specifically bind to each known flu strain.

Test unknown sample against each set of antibodies separately.

# bio/ Hemagglutination inhibition assay

- HI can only detect viruses that match the ones used to make the antibodies
- Conventional sequencing can identify new strains, but will miss rare variants
- Next generation 'deep sequencing' can study all variants in mixed population

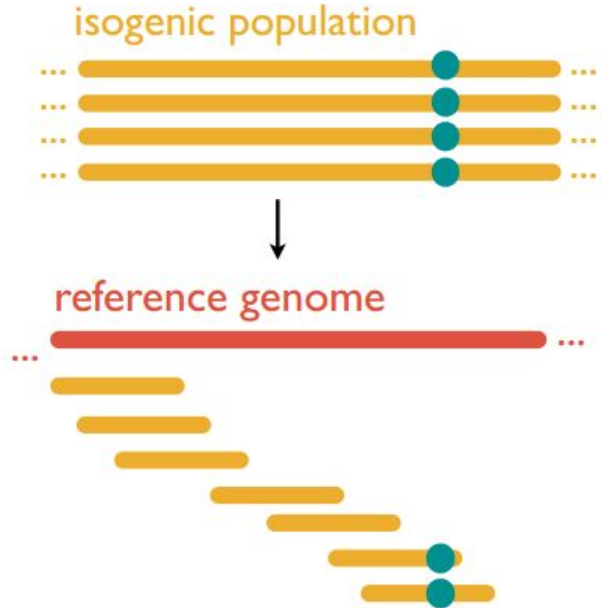


# /informatics Isogenic vs Mixed Populations

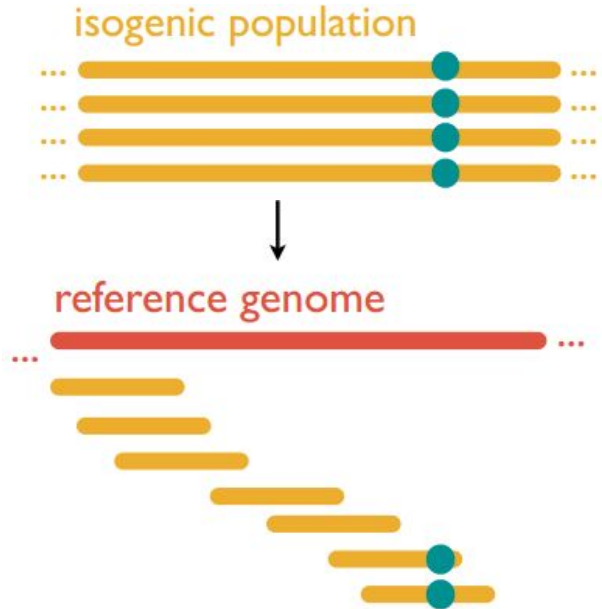




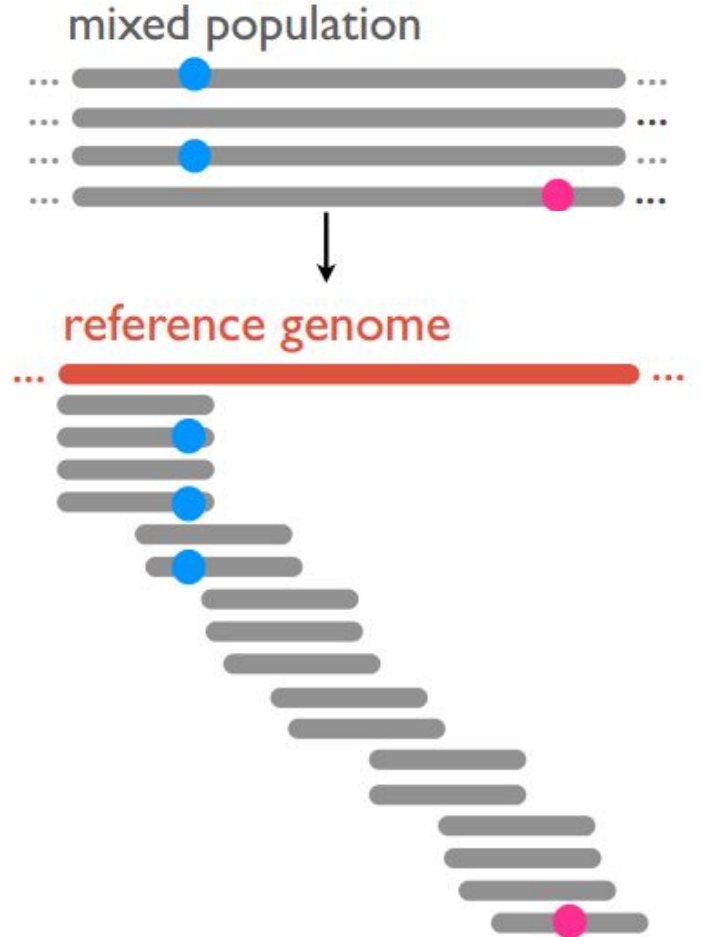
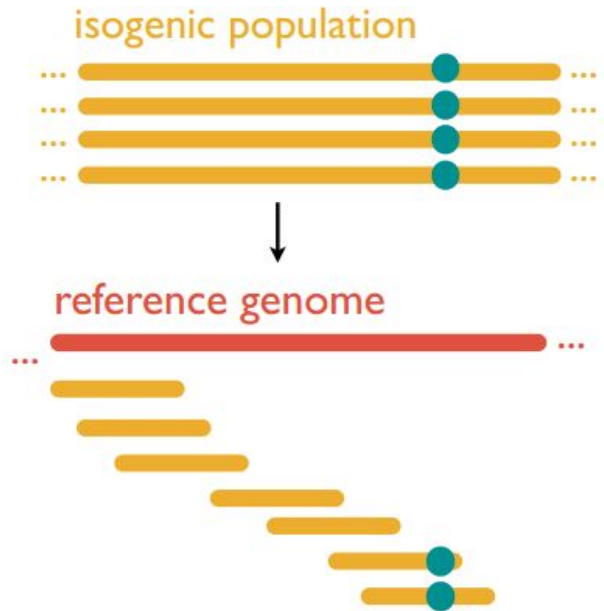
# /informatics Isogenic vs Mixed Populations



# /informatics Isogenic vs Mixed Populations



# /informatics Isogenic vs Mixed Populations



# /informatics Deep Sequencing

In sequencing runs with an identical number of clusters and the same read length, which target will have the highest coverage?

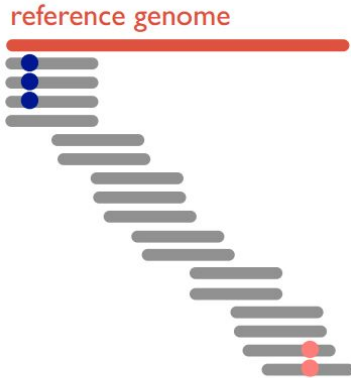
- A) the human genome
- B) a tumor cell
- C) the gene for insulin
- D) the human Y chromosome
- E) the *E. coli* genome

# /informatics Isogenic vs Mixed Populations



# /informatics Isogenic vs Mixed Populations

Which alignment  
shows the correct deep  
sequencing results?



A



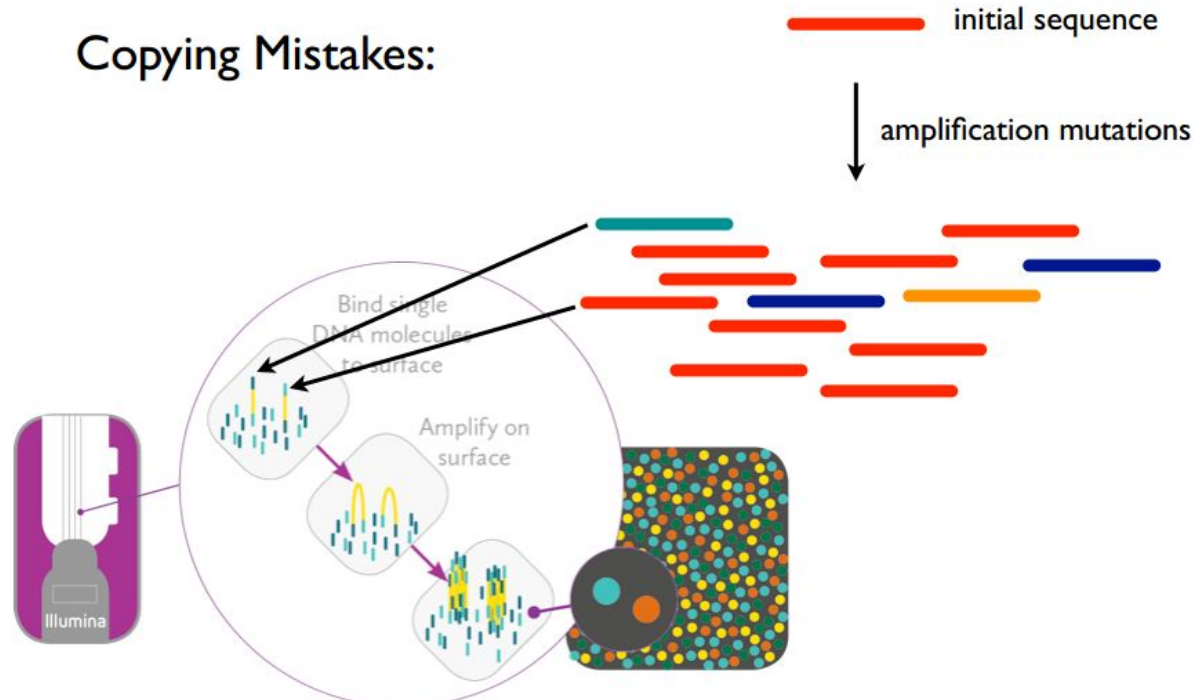
B



C

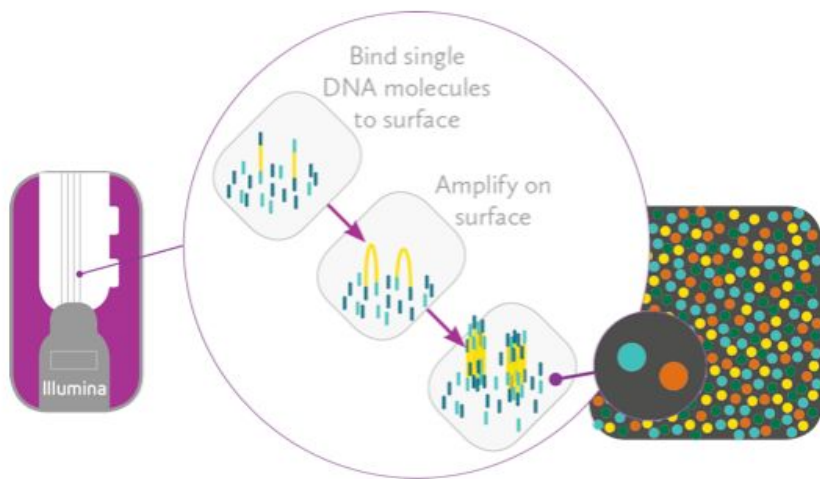
# /informatics Sources of sequencing error

“Upstream” Errors: occur during amplification of DNA prior to sequencing

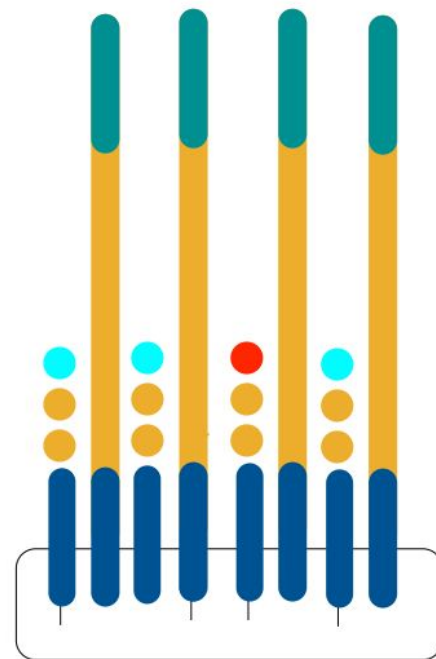


# /informatics Sources of sequencing error

“During” errors: occur in cluster generation or sequencing by synthesis



Incorrect base

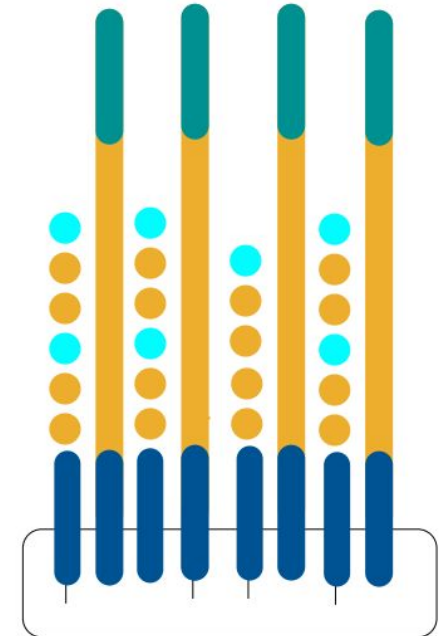
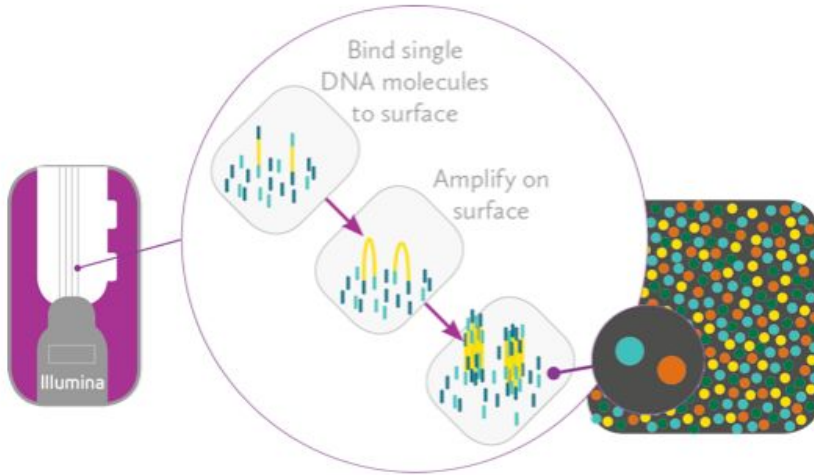




# /informatics Sources of sequencing error

“During” errors: occur in cluster generation or sequencing by synthesis

Missing base



## bio/ Experimental control

- Minimize amplification
- High-Fidelity enzymes
- Biological replicates
- Technical replicates
- Cross-platform replicates

## bio/ Experimental control

- Biological replicates: Analysis of multiple biological samples under the same conditions from the same host.
- Technical replicates: Repeat analysis of the exact same sample with the same technique.
- Cross-platform replicates: Repeat analysis of the exact same sample with the different platforms.