

RNA-seq - Wprowadzenie

Podstawowe założenia

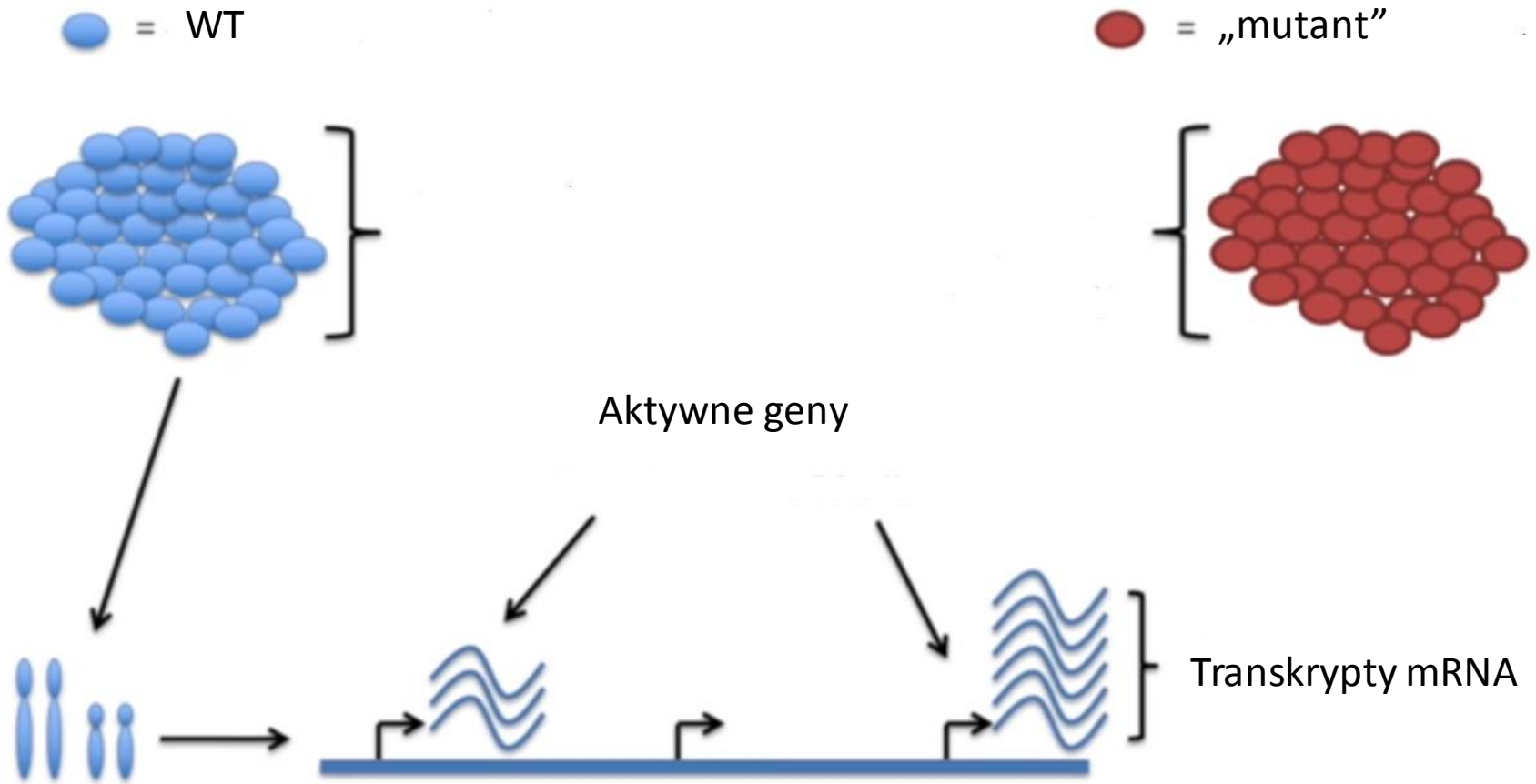


Komórki „mutanty” zachowują się inaczej od komórek WT

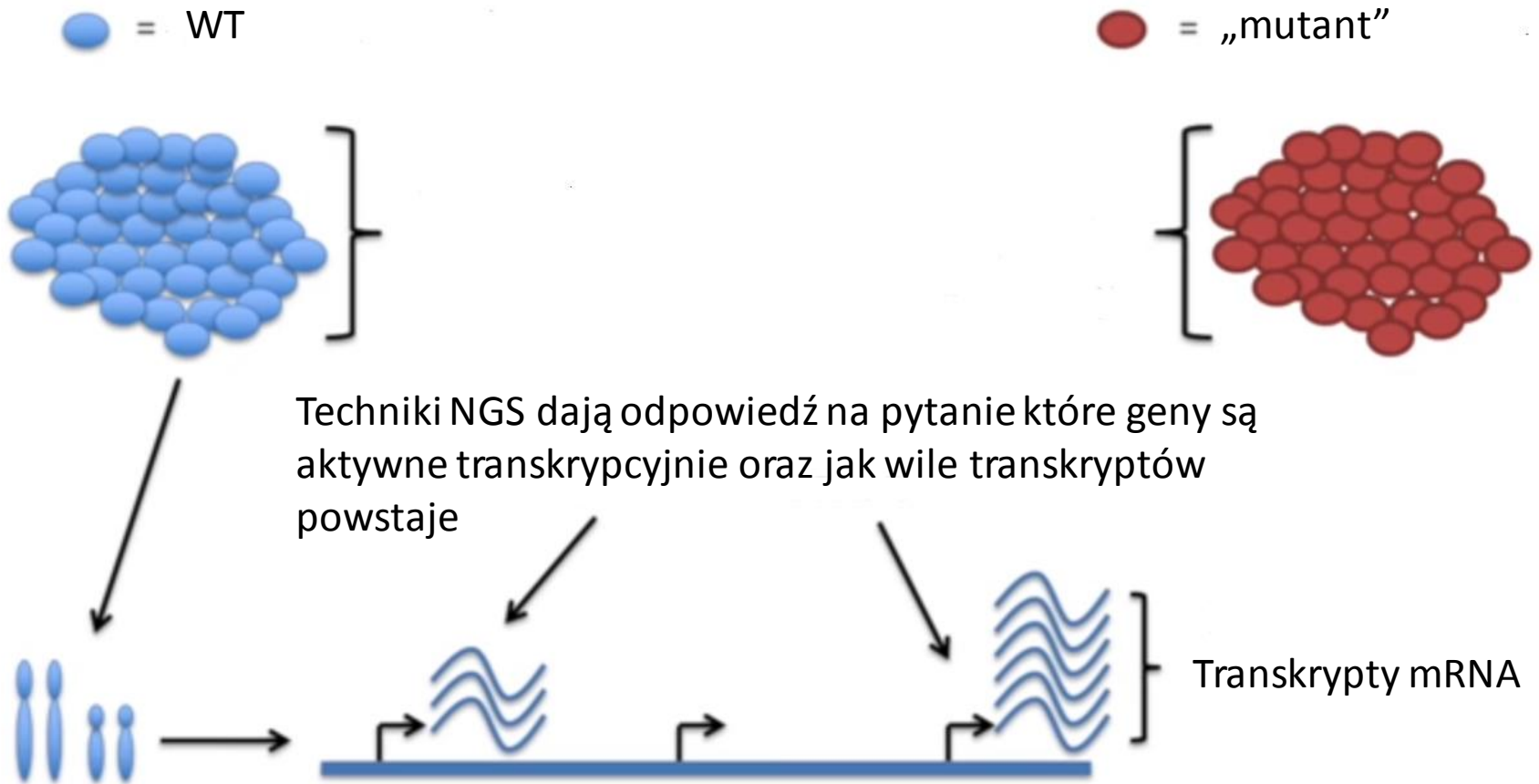
Chcemy poznać molekularny mechanizm tych różnic

Jak możemy to osiągnąć?

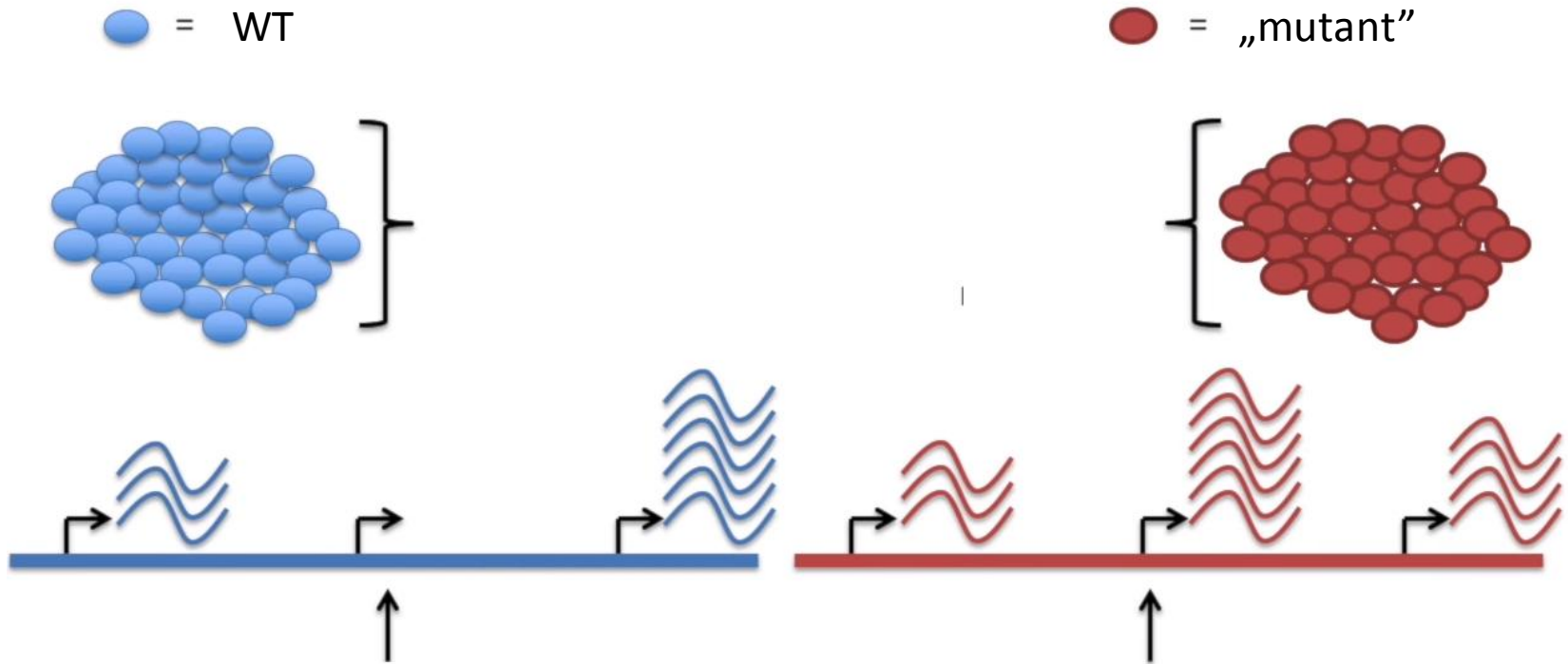
Podstawowe założenia



Podstawowe założenia



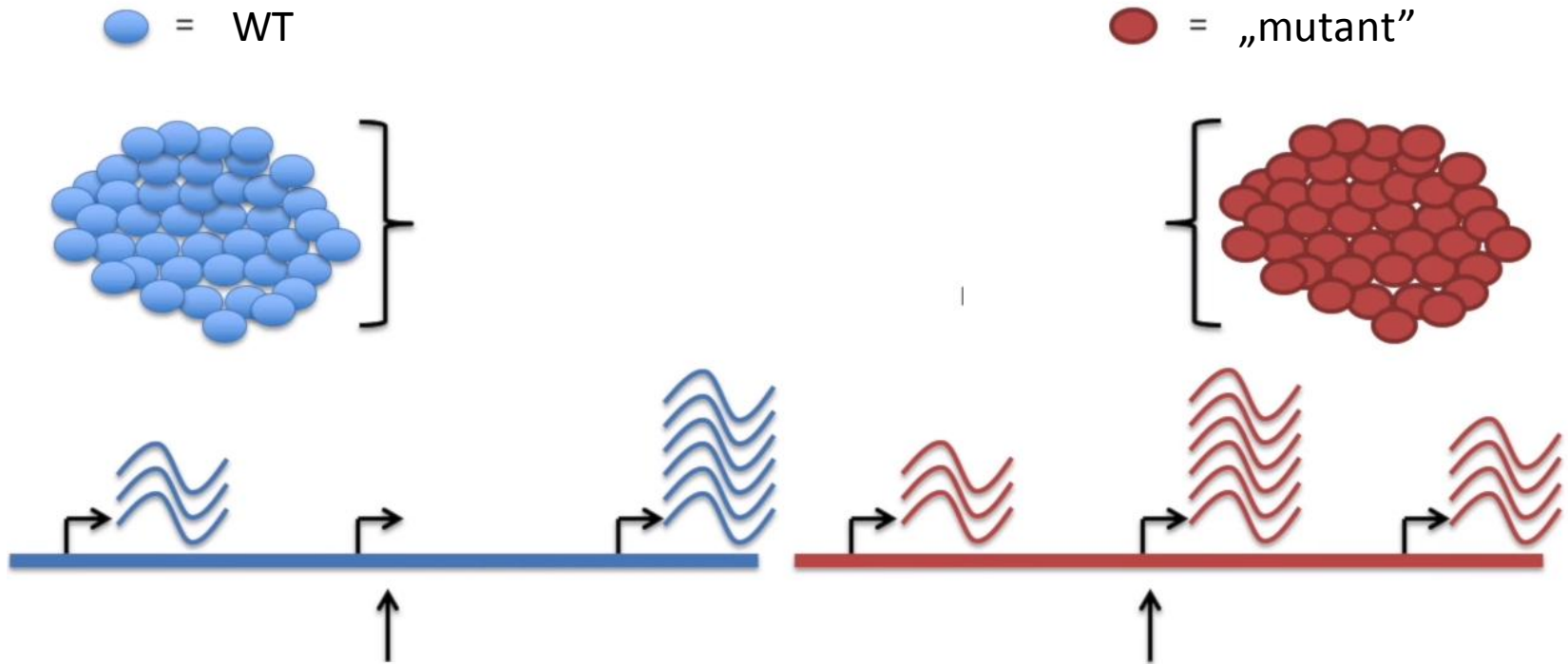
Podstawowe założenia



Możemy wykorzystać RNAseq do
zmierzenia ekspresji w komórkach WT...

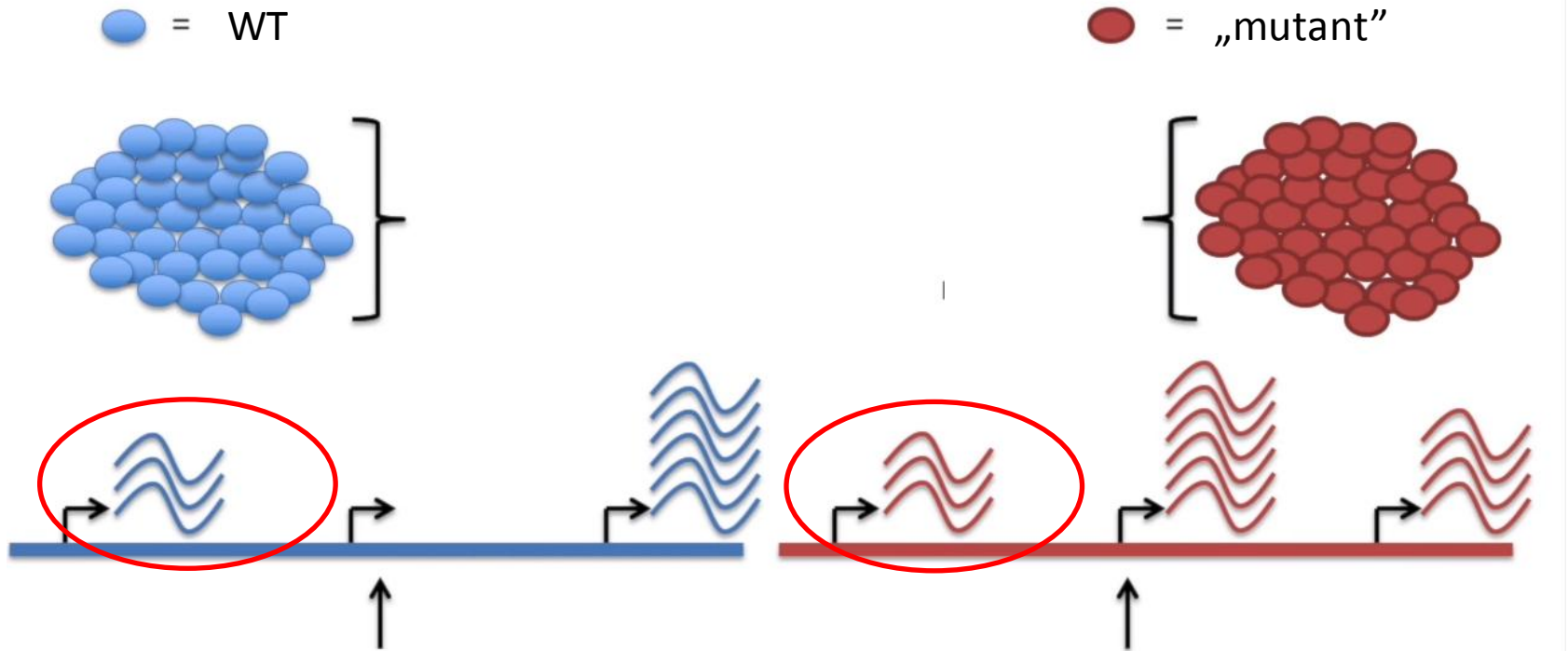
... oraz w komórkach zmutowanych...

Podstawowe założenia

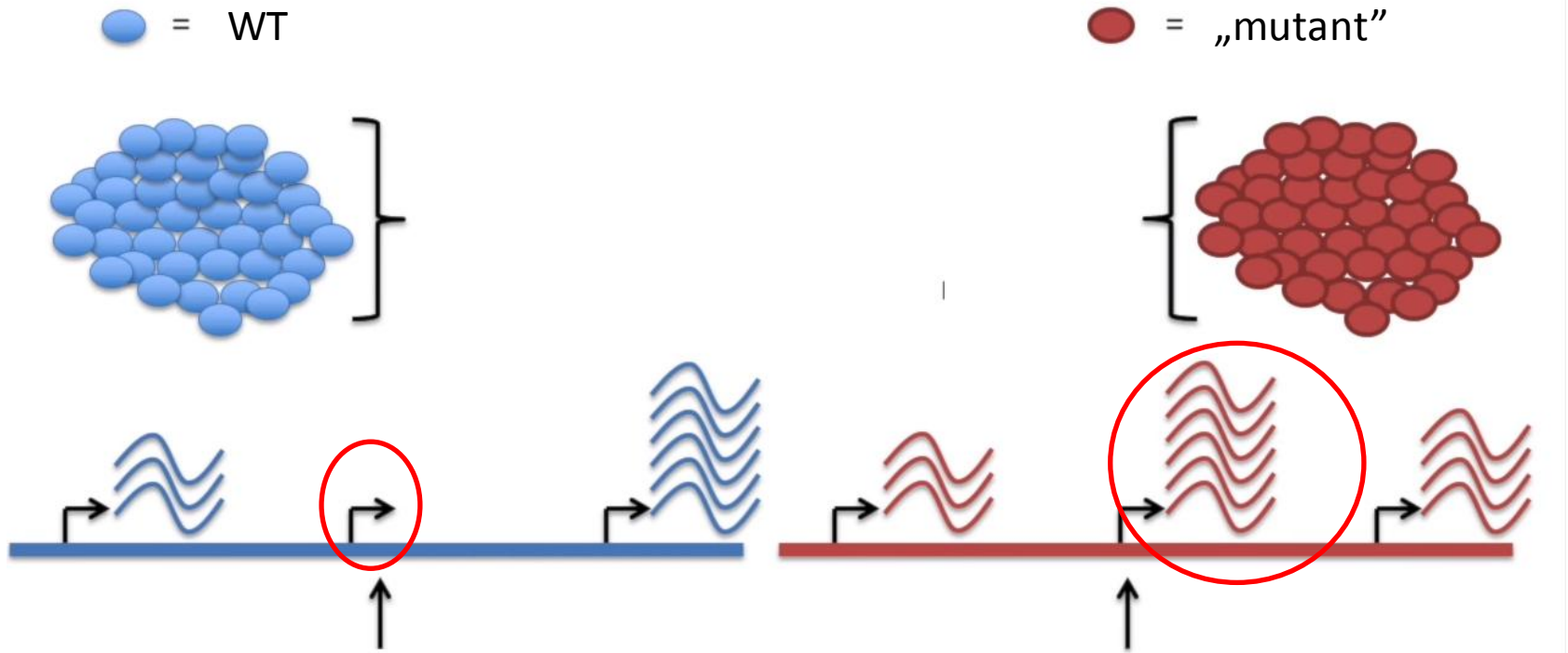


Następnie możemy porównać te dwa typy komórek i wywnioskować które geny mają zmienioną ekspresję

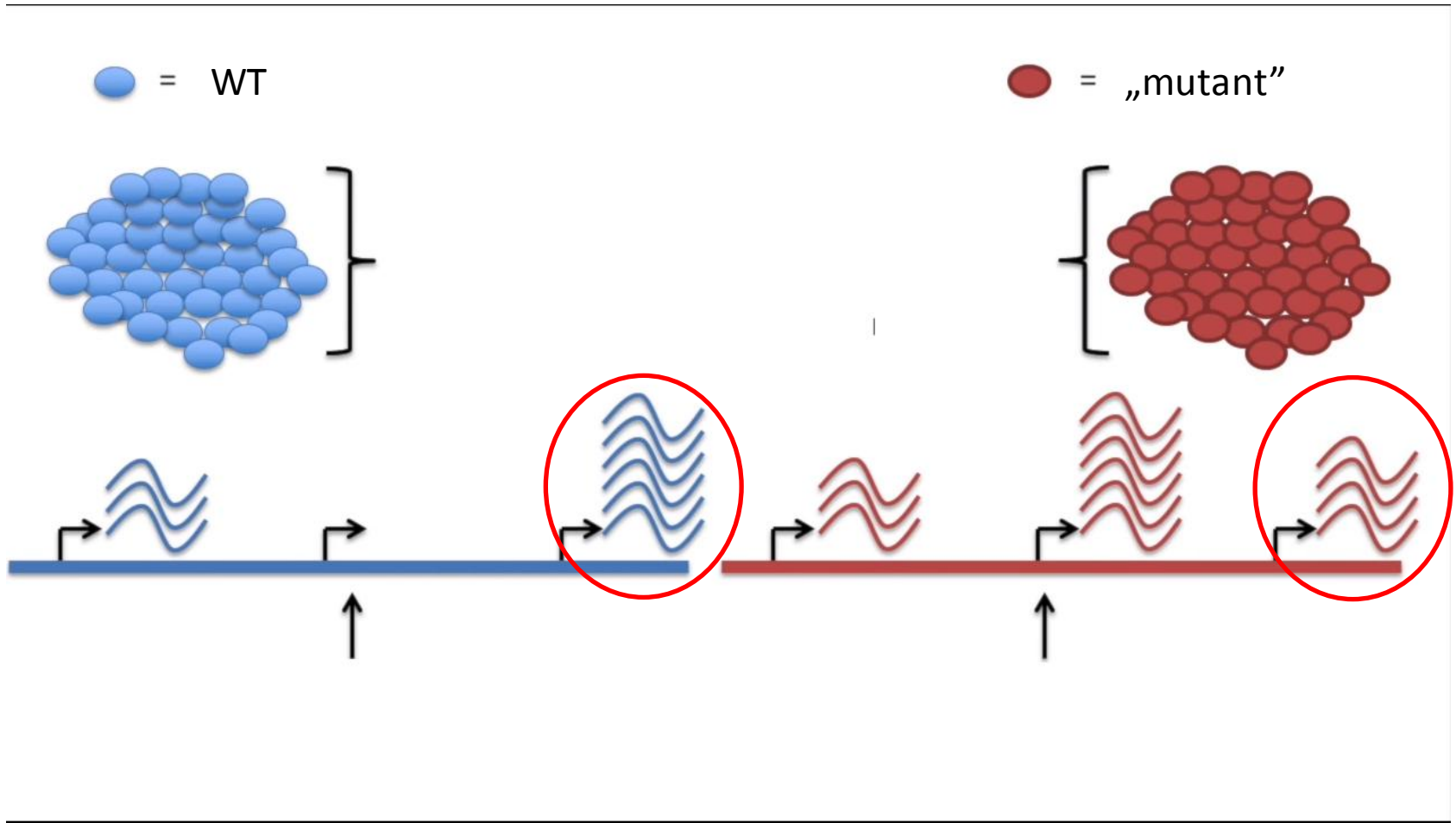
Podstawowe założenia



Podstawowe założenia



Podstawowe założenia



Podstawowe etapy RNAseq

1. Przygotowanie biblioteki
2. Sekwencjonowanie
3. Analiza danych

Przygotowanie biblioteki

Istnieje wiele protokołów i kilka urządzeń (technik) sekwencjonowania, ale obecnie najszerzej stosowanym protokołem i urządzeniem (technikom) jest ta zaproponowana przez Illuminę

Przygotowanie biblioteki

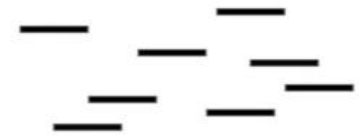
Krok 1: Izolacja RNA



Krok 2: Fragmentacja RNA



Krok 3: Synteza cDNA



Transkrypty mogą mieć długość kilku tysięcy par zasad, dlatego muszą być pofragmentowane do odcinków długości 200-300 pz, bo aparaty do sekwencjonowania mogą sekwencjonować tylko fragmenty takiej długości

Przygotowanie biblioteki

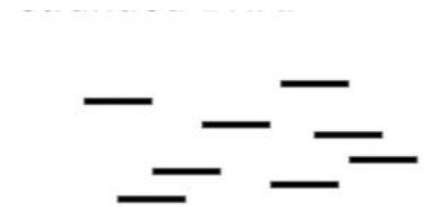
Krok 1: Izolacja RNA



Krok 2: Fragmentacja RNA



Krok 3: Synteza cDNA



DNA jest bardziej stabilne niż RNA i łatwiej go amplifikować oraz modyfikować

Przygotowanie biblioteki

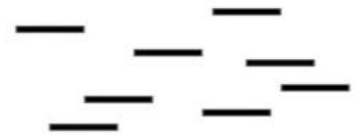
Krok 1: Izolacja RNA



Krok 2: Fragmentacja RNA



Krok 3: Synteza cDNA



Krok 4: Ligacja adaptorów



1. Adaptory umożliwiają urządzeniom sekwencyjnym rozpoznanie fragmentów

Przygotowanie biblioteki

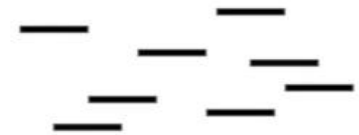
Krok 1: Izolacja RNA



Krok 2: Fragmentacja RNA



Krok 3: Synteza cDNA



Krok 4: Ligacja adaptorów



2. Adaptory umożliwiają sekwencjonowanie fragmentów z dwóch różnych próbek w jednym czasie (obniża koszty)

Przygotowanie biblioteki

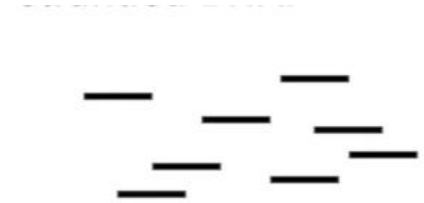
Krok 1: Izolacja RNA



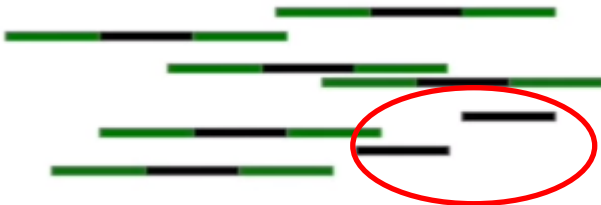
Krok 2: Fragmentacja RNA



Krok 3: Synteza cDNA



Krok 4: Ligacja adaptorów



Przygotowanie biblioteki

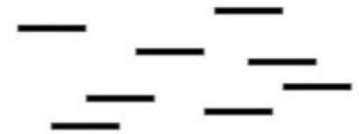
Krok 1: Izolacja RNA



Krok 2: Fragmentacja RNA



Krok 3: Synteza cDNA



Krok 4: Ligacja adaptorów



Krok 5: Amplifikacja (PCR)



Tylko fragmenty z adaptorami ulegają amplifikacji

Przygotowanie biblioteki

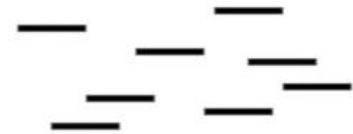
Krok 1: Izolacja RNA



Krok 2: Fragmentacja RNA



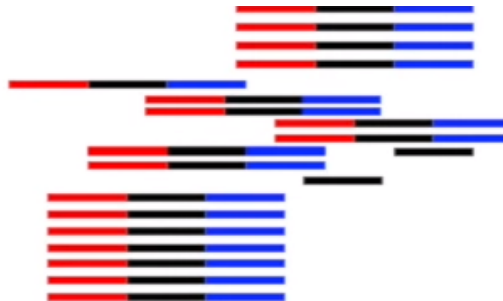
Krok 3: Synteza cDNA



Krok 4: Ligacja adaptorów



Krok 5: Amplifikacja (PCR)



Krok 6: QC

1. Weryfikacja stężenia biblioteki
2. Weryfikacja długości fragmentów w bibliotece

Sekwencjonowanie biblioteki

Jak działa NGS

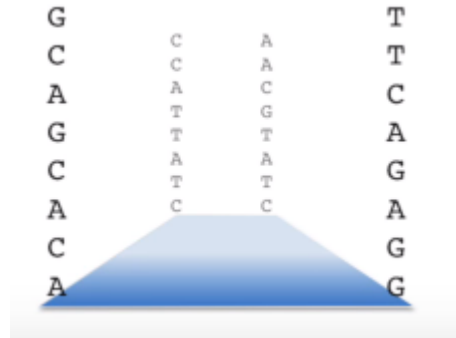
G
C
A
G
C
A
C
A



Wyobraźmy sobie fragment
DNA, który chcemy
sekwencjonować
W sekwenatorze ten fragment
jest umieszczany pionowo

Właściwie to jest 400 milionów fragmentów
w kwadracie (grid)

Ten kwadrat nosi nazwę „flow cell”



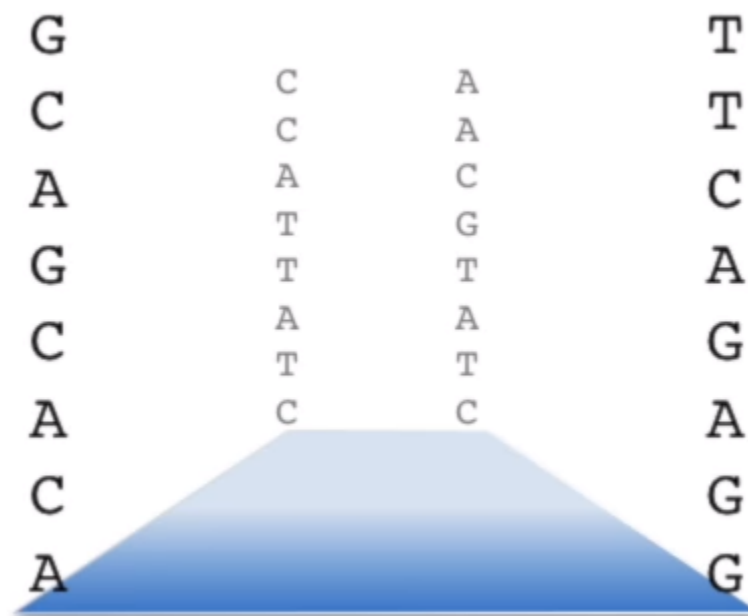
Urządzenie posiada fluorescencyjne sondy, każdy kolor sondy odpowiada jednemu nukleotydowi

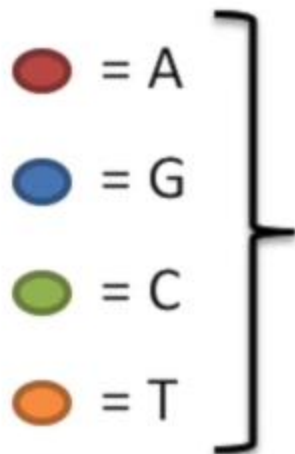
● = A

● = G

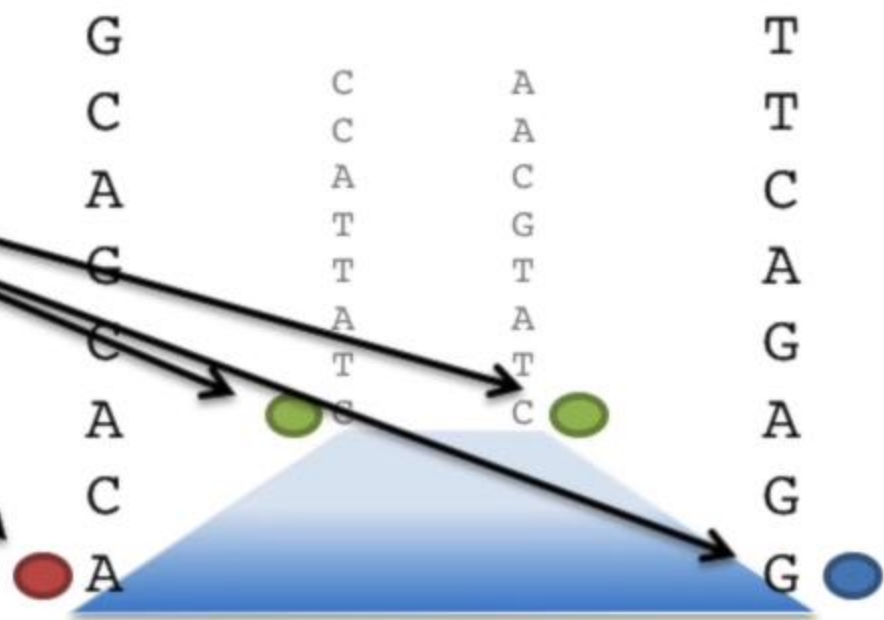
● = C

● = T

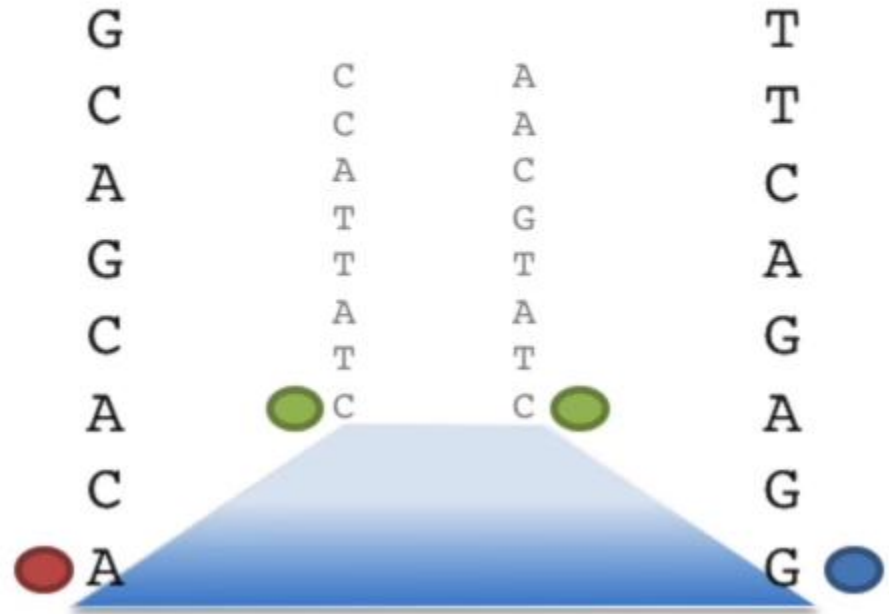
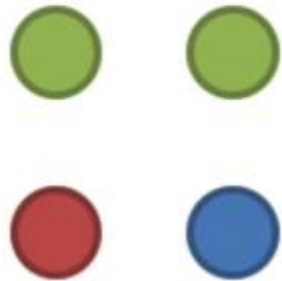




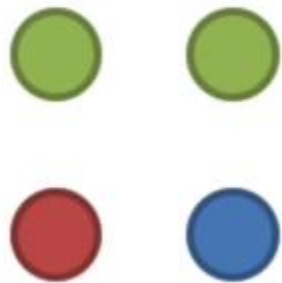
Sądy przyłączają się do pierwszego nukleotydu w każdej sekwencji



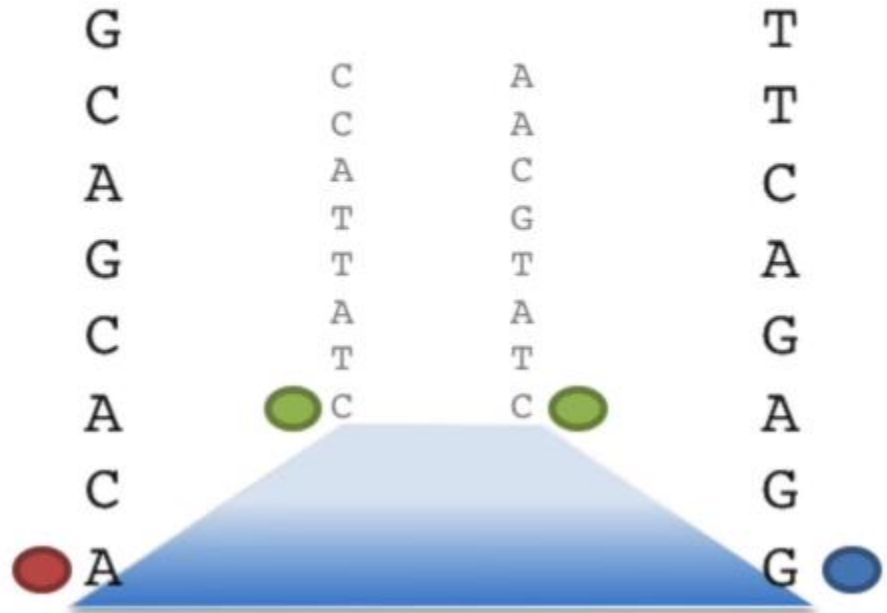
Urządzenia robi zdjęcie,
które wygląda mniej
więcej tak (zdjęcie
robione jest „z góry”)



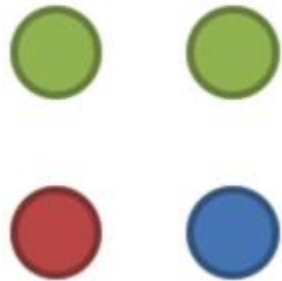
Urządzenia robi zdjęcie, które wygląda mniej więcej tak (zdjęcie robione jest „z góry”)



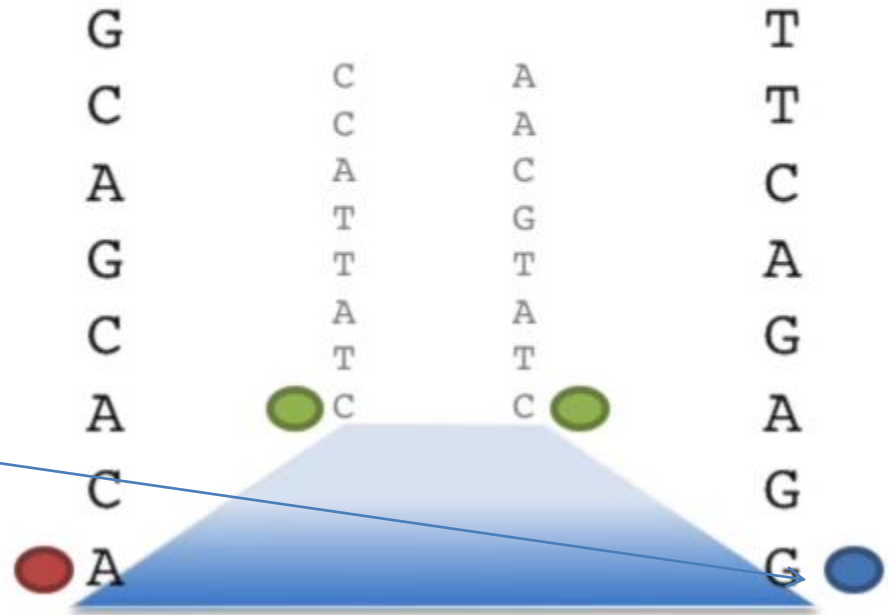
Analiza zdjęcia mówi maszynie, że pierwszy nukleotyd w lewym rogu ma kolor „A”



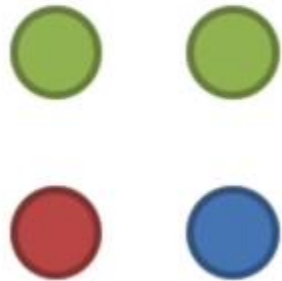
Urządzenia robi zdjęcie,
które wygląda mniej
więcej tak (zdjęcie
robione jest „z góry”)



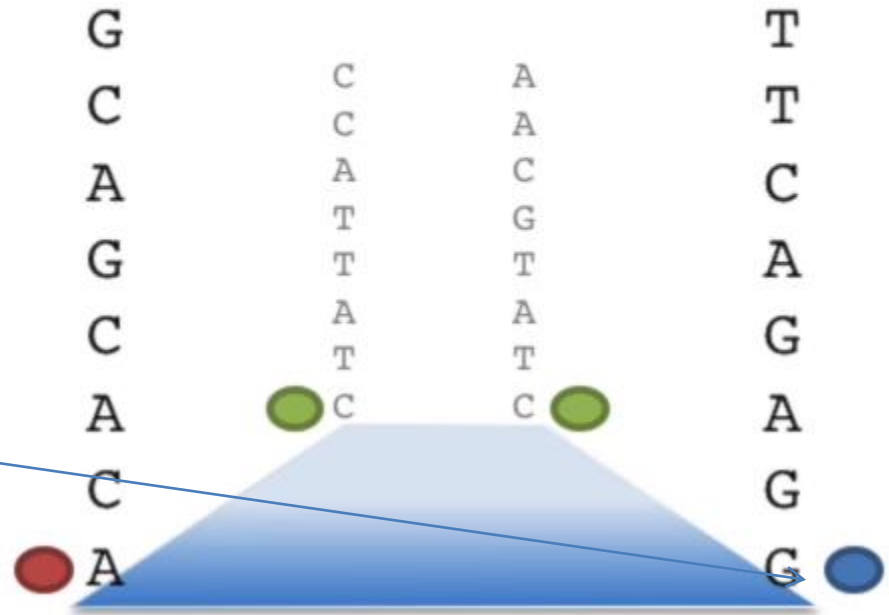
To jest „G” itd.



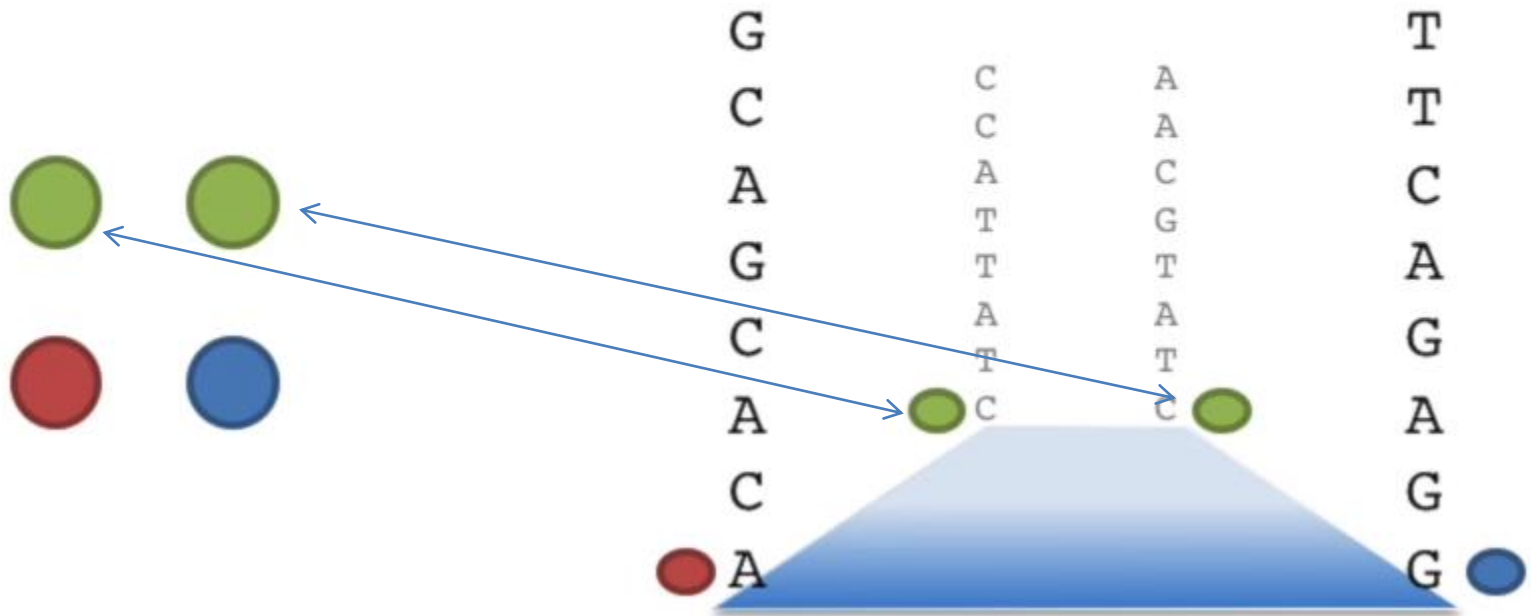
Urządzenia robi zdjęcie,
które wygląda mniej
więcej tak (zdjęcie
robione jest „z góry”)



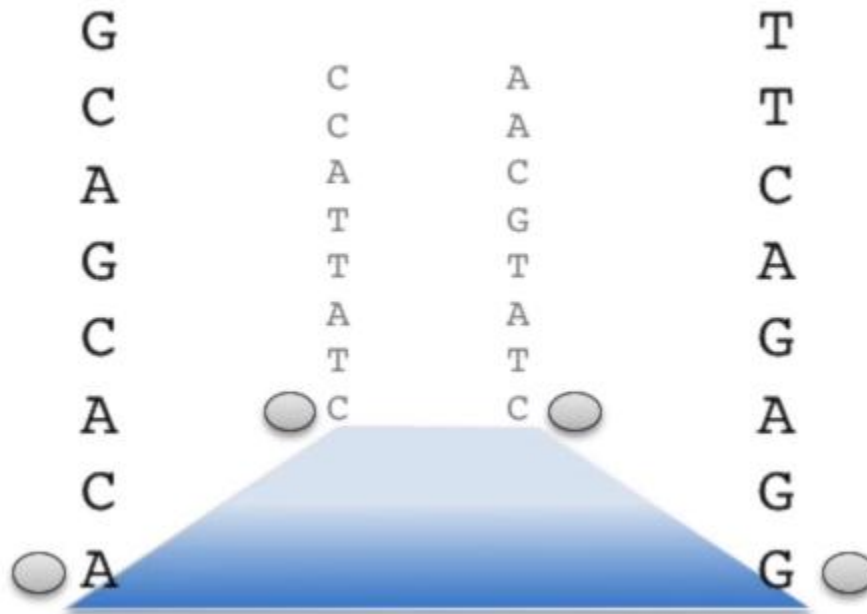
To jest „G”



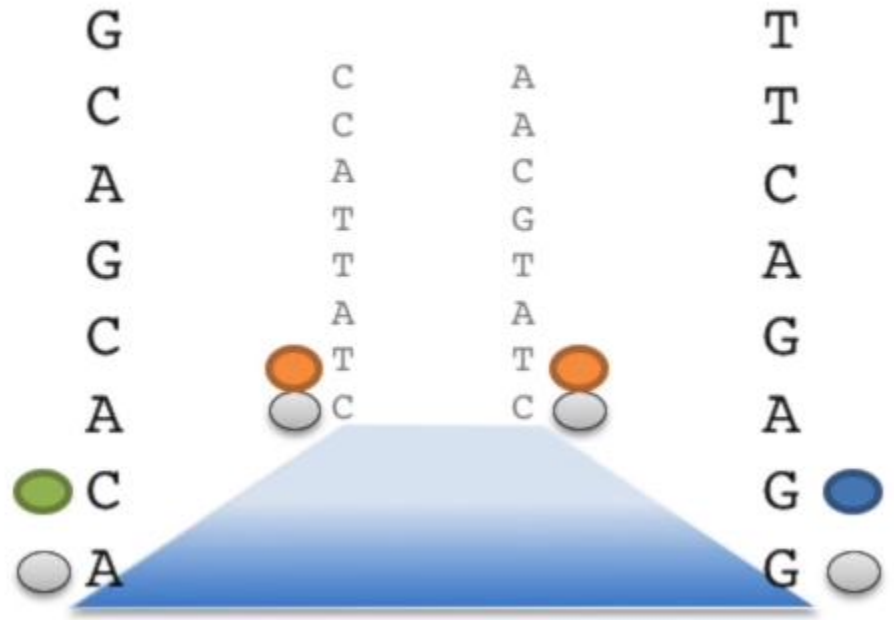
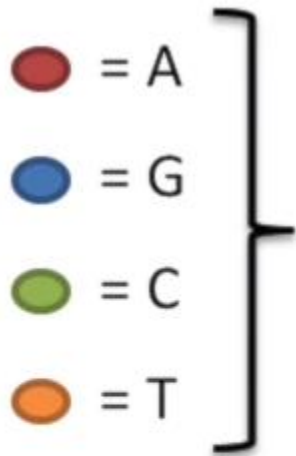
Urządzenia robi zdjęcie,
które wygląda mniej
więcej tak (zdjęcie
robione jest „z góry”)



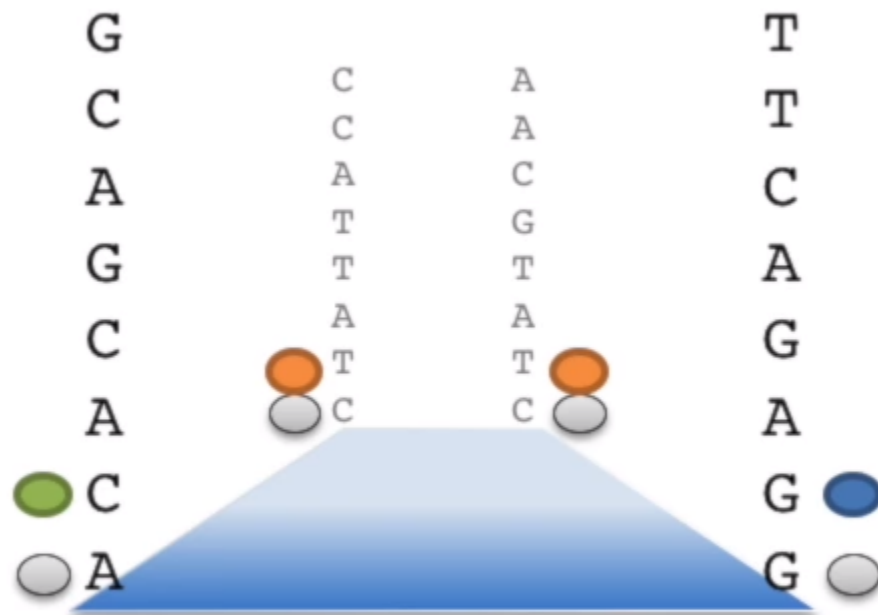
Maszyna wymywa barwnik fluorescencyjny z sondy



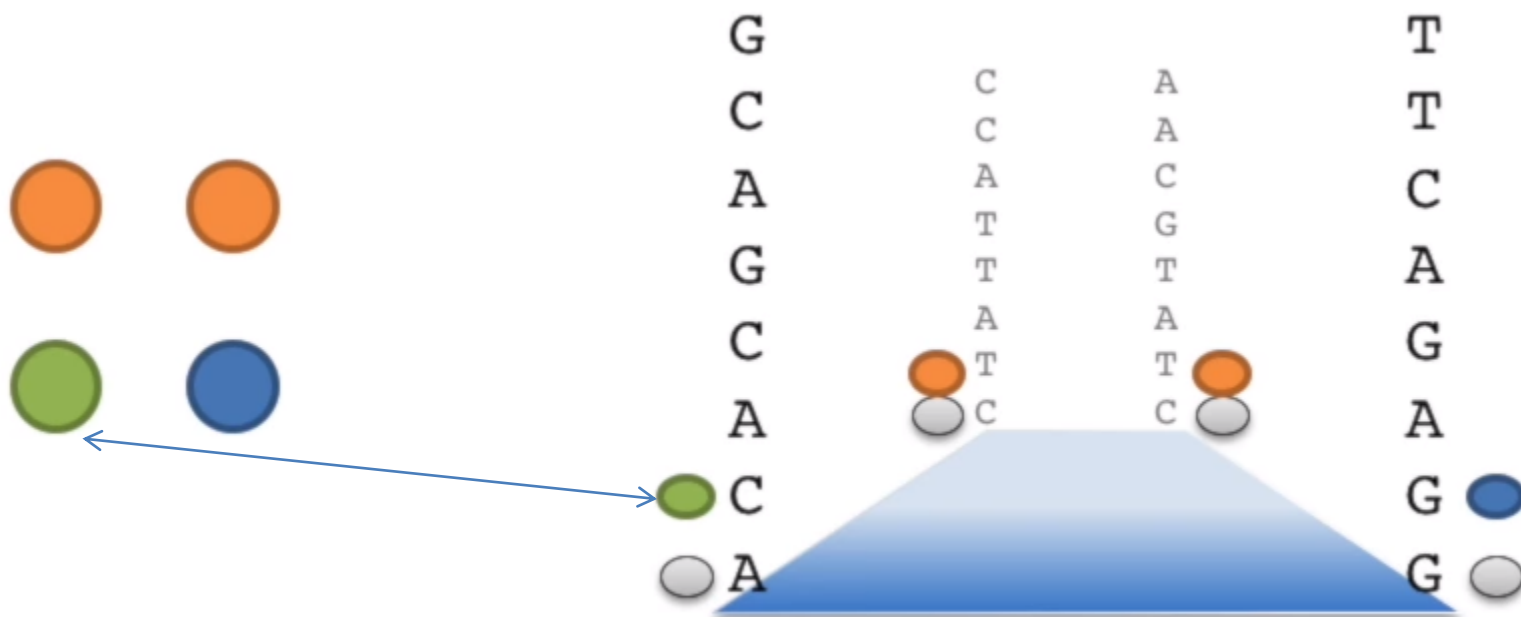
Następnie kolejna partia sąd jest wiązana do kolejnego nukleotydu



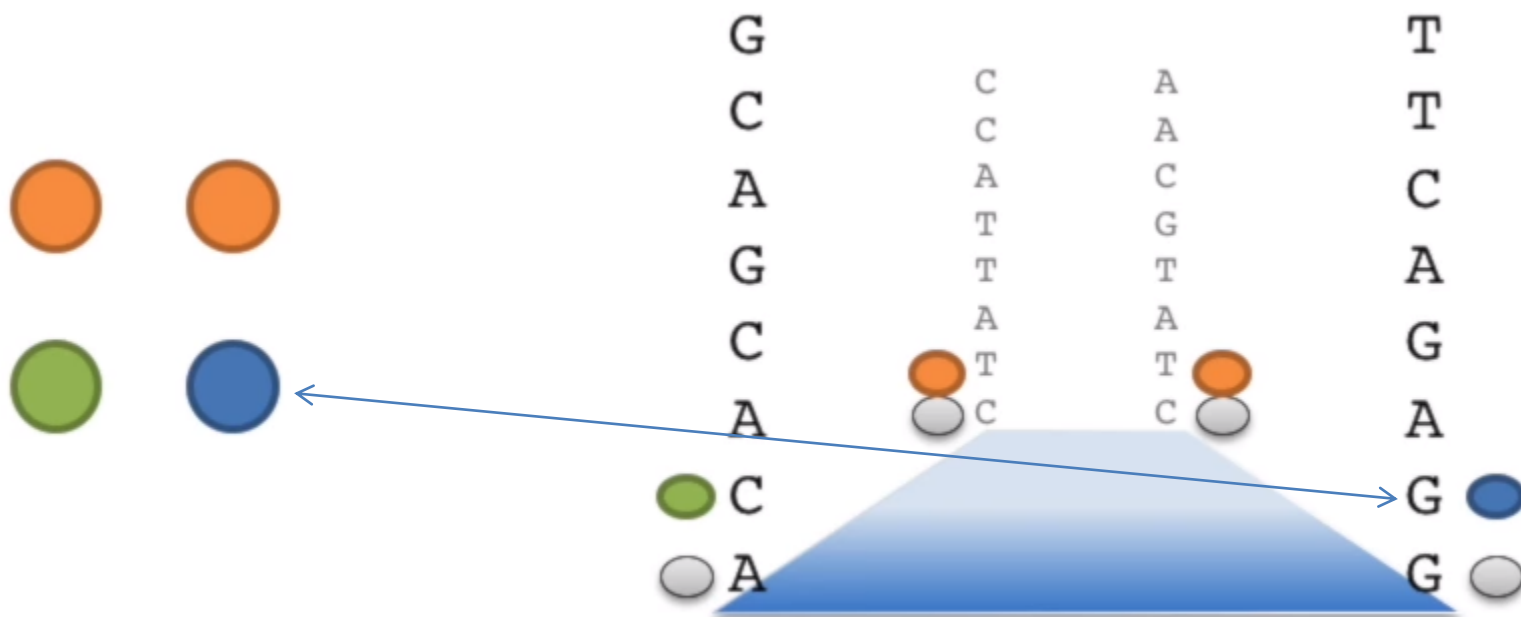
Maszyna robi kolejne zdjęcie i określa typ nukleotydu w każdym fragmencie



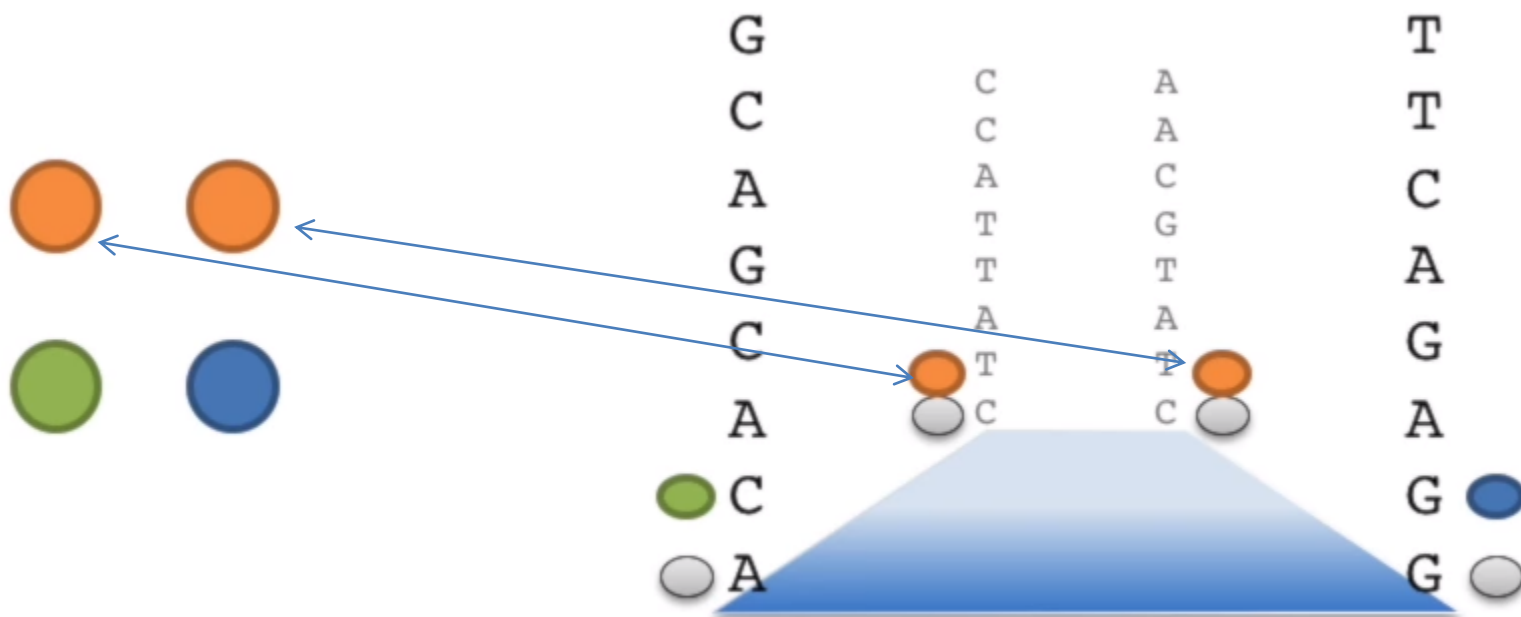
Maszyna robi kolejne zdjęcie i określa typ nukleotydu w każdym fragmencie



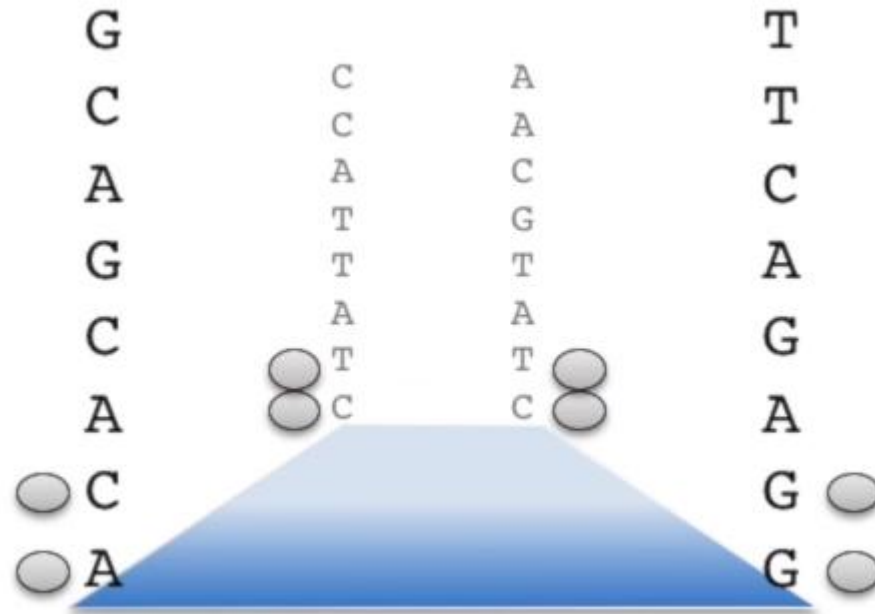
Maszyna robi kolejne zdjęcie i określa typ nukleotydu w każdym fragmencie



Maszyna robi kolejne zdjęcie i określa typ nukleotydu w każdym fragmencie

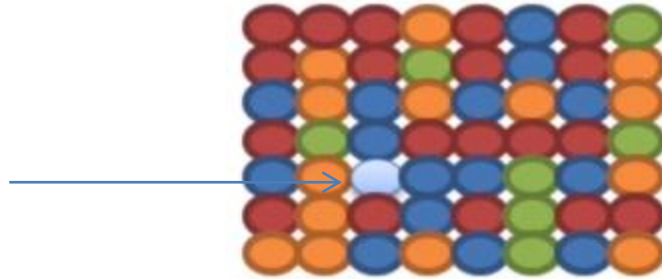


Maszyna wmywa barwnik fluorescencyjny z sondy, proces powtarza się



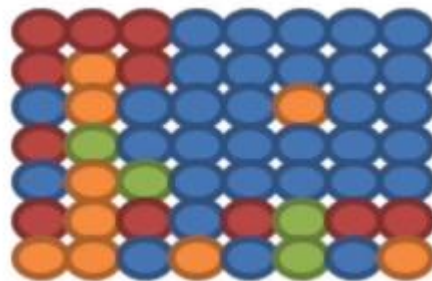
W przypadku 400 milionów fragmentów DNA analizowanych jednocześnie obraz jest bardzo „gęsty”

Zdarza się, że sonda nie świeci tak intensywnie jak powinna



QS (Quality Scores) który jest wynikiem odczytu razem z nukleotydem jest liczbową reprezentacją pewności, że maszyna prawidłowo odczytała nukleotyd w tym miejscu sekwencjonowanego fragmentu (każdy nukleotyd w każdym fragmencie ma swój QS

Taki efekt nazywamy „low diversity”. W takich obszarach obraz staje się rozmyty i identyfikacja poszczególnych (pojedynczych) fragmentów staje się trudniejsza



Duża liczba fragmentów z tym samym kolorem w jednym fragmencie „flow cell” może skutkować niską wartością QS

Efekt „low diversity” jest szczególnie problematyczny w przypadku sekwencjonowania kilku pierwszych nukleotydów, kiedy maszyna ustala w których punktach kwadratu fragmenty DNA są umiejscowione

Surowe dane - FASTQ

```
@NS500177:196:HFTTTAFXX:1:11101:10916:1458 2:N:0:CGCGGCTG
ACACGACGATGAGGTGACAGTCACGGAGGATAAGATCAATGCCCTCATTAAGCAGCCGGTGTA
+
AAAAAEEEEEEEEEEEE//AEEEEEEEEEEEEEEEE/EE/<<EE/AEEEAEE//EEEEEEEEAEA<
```

Każdy odczyt składa się z 4 linijek
Każda linijka zawiera inne informacje

FASTQ

1. @ unikalne ID sekwencji
2. Sekwencja nukleotydów
3. + ?
4. QS dla każdego odczytanego nukleotydu

- 1 standardowy eksperyment sekwencjonowania z 400 milionami fragmentów DNA wygeneruje 1,6 biliarda linijek danych

Analiza surowych danych

- Filtracja danych
- Przyrównanie/złożenie (alignment) odczytów wysokiej jakości do genomu
- Zliczenie liczby odczytów (reads) przypadających dla każdego genu

Filtracja danych

- Odczyty o niskiej wartości określenia sekwencji (niskie prawdopodobieństwo)
- Artefakty związane z chemią wykorzystywaną w sekwencjonowaniu

