

# SEKVENCIRANJE NOVE GENERACIJE

---

Funkcijska genomika

2013/2014

doc.dr. Tadeja Režen

# Sekvenciranje nove generacije

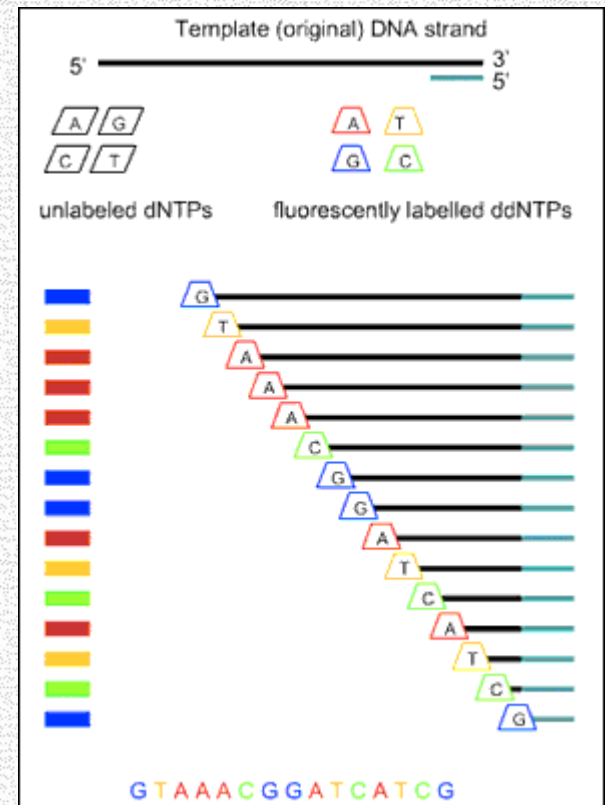
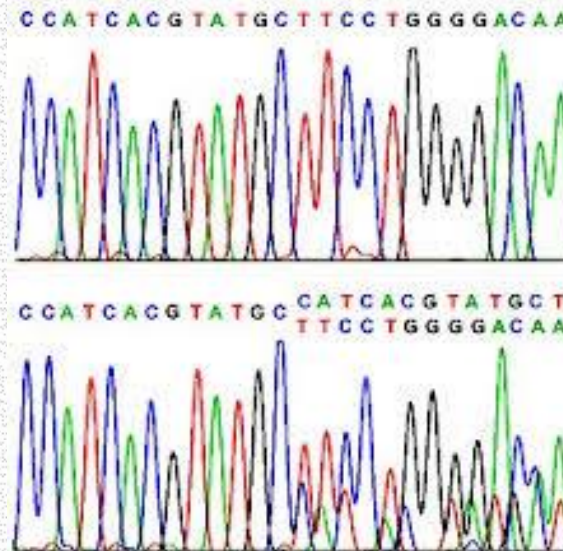
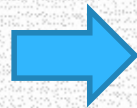
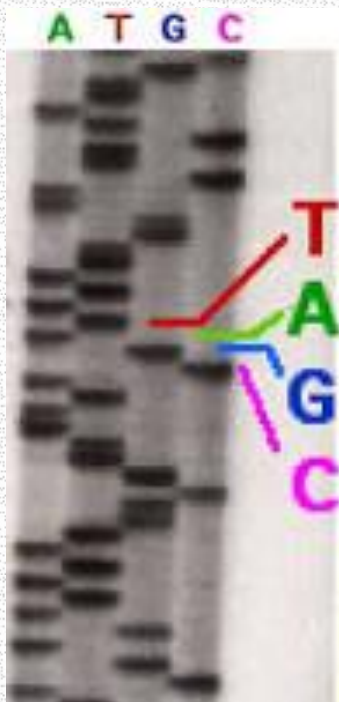
- 1) Uvod – tehnologija in aplikacije
- 2) Priprava vzorcev in sekvenciranje
- 3) Analiza podatkov
- 4) Demonstracija: GS Junior
- 5) Razprava

## NAMEN VAJE

V sklopu vaje se boste spoznali s teoretičnimi osnovami sekvenciranja nove generacije. V praktičnem delu se boste spoznali s pripravo vzorcev in analizo podatkov na napravi za določanje zaporedja GS Junior.

# Sekvenciranje prve generacije

Sangerjeva metoda – sekvenciranje z zaustavitvijo verige z dideoksinukleotidom, Frederick Sanger, 1977





# Sekvenciranje naslednje generacije

- Danes lahko sekvenciramo v eni analizi celoten človeški genom ali genom katerega koli drugega živega bitja. Lahko določimo zaporedje vseh prepisov RNA, ki so v celici. V nekaj dneh na eno bazo natančno.
- Več fragmentom naenkrat paralelno določamo zaporedje, vendar so ti med seboj prostorsko ločeni: na 20  $\mu\text{m}$  kroglicah v pikotitrski ploščici (454 sistem), 1  $\mu\text{m}$  klonskih kroglicah (SOLiD), na klonskih kolonijah pritrjenih na podlago (HiSeq).

Poznamo drugo in tretjo generacijo:

- Druga generacija s pomočjo PCR pomnoži DNA fragmente za sekvenciranje. Pomnoževanje je nujno, da je signal fluorescence dovolj velik, da ga zazna CCD kamera.
- Tretja generacija sekvenciranja ne potrebuje pomnoževanje vzorca, ampak prebere zaporedje ene izvorne DNA molekule.

# Naslednja generacija sekvenciranja

Poznamo sekvenciranje:

- 1) S sintezo na kroglicah na pikotitrski ploščici na sistemu 454 (GS FLX Titanium XL+, GS FLX Titanium XLR70 in GS Junior) podjetja Roche
- 2) S sintezo na pretočni celici (flow cell) na napravi HiSeq 2500q, HiSeq X Ten, NextSeq500 in MiSeq podjetja Illumina
- 3) Z ligacijo na kroglicah na napravi SOLiD 5500xl podjetja Life Technologies (Applied Biosystems); SOLiD – Sequencing by Oligonucleotide Ligation and Detection
- 4) S sintezo na kroglicah na Ion čipu na napravi Ion Torrent (Life Technologies) ; <http://www.youtube.com/watch?v=MxkYa9XCvBQ>

# Tretja generacija sekvenciranja

## **Heliscope (tSMS, angl. »true single molecule sequencing«)**

- Priprava DNA knjižnice (fragmentacija DNA in dodajanje poliA repov)
- Hibridizacija DNA fragmentov na poliT oligonukleotide, ki so pritrjeni na pretočno celico in ti s sintezo sekvencirajo paralelno. Sekvencira do 55 b.

## **SMRT (angl. »single molecule real time sequencer«)**

- Priprava DNA knjižnice s fragmentacijo.
- Čip ima več tisoč jamic, kjer je pritrjena DNA polimeraza in le ta prepisuje DNA fragment s fluorescenčno označenimi nukleotidi. Sekvencira daljše od 1000 baz.

## **RNAP (angl. » single molecule real time sequencer«)**

- RNA polimeraza je pritrjena na polistirensko kroglico in DNA fragment je pritrjen na drugo kroglico, naprava meri razdaljo med kroglicama

## **»Nanopore« naprava za določanje zaporedja**

- DNA molekula potuje skozi alfa-hemolizin poro s ciklodekstrinom, kjer nanj vezan nukleotid spremeni ionski tok skozi poro.



# Aplikacije

## DNA:

- a) sekvenciranje genoma *de novo*: sestavljanje novih genomov,
- b) resekvciranje genoma: ponovno sestavljanje genomov, detekcija novih in pogostih različic, strukturne različice,
- c) tarčno (re)sekvenciranje , sekvenciranje amplikonov – genotipizacija,
- d) sekvenciranje celotnega eksoma,
- e) ChiP-Seq (kromatinska imunoprecipitacija): vezavna mesta TF,
- f) metilirana DNA - Met-Seq: epigentika in struktura kromatina.

## RNA:

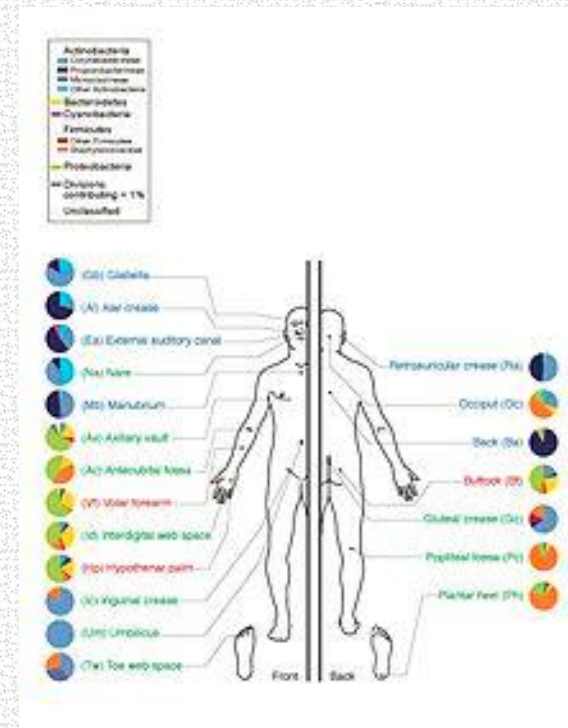
- a) sekvenciranje celotnega transkriptoma (RNA-Seq) in sestava *de novo*,
- b) določevanje diferencialno izraženih genov,
- c) sekvenciranje kratkih RNA: miRNA in druge nekodirajoče RNA.

# Aplikacije

## Metagenomi:

- genetski material kompleksnih vzorcev iz okolja
- identifikacija vrst mikroorganizmov – biodiverzитета
- Pridobimo zaporedja vseh genov vseh mikroorganizmov v vzorcu.

- Mikrobiom človeka
- Biogorivo
- Ekosistemi – ekologija, varstvo okolja
- Biotehnologija





# Primerjava naprav za določevanje zaporedja

	<b>GS Junior</b>	<b>GS FLX Titanium / 1K</b>	<b>Illumina HiSeq</b>	<b>Illumina GAIIe</b>	<b>SOLiD 4</b>	<b>SOLiD PI</b>
<b>Read length (bp)</b>	400	400 / 800	2x100	2x100 – 2x150	2x50 - 2x75	2x50
<b>Accuracy</b>	>99.9%	>99.9%	98%	98%	99.95+%	99.95+%
<b># reads/run</b>	100 000	1 000 000	550M	200M	640M - 2.4 B	640 M
<b>Run time</b>	10 h	10 / 16 h	8 d	10 d	16 d – 24 d	16 d
<b>Yield/run</b>	40 Mb	400 / 1,000 Mb	200 Gb	20 Gb	64-300 Gb	50 Gb
<b>Throughput (Gb/d)</b>	0.08	0.8 – 1.0	25	2	4-23	7



# Izbora metode

Izbira metode sekvenciranja je odvisna od cilja raziskave!

Cilj in namen raziskave vplivata na odločitve o:

- vrsti vzorca (genomska DNA, eksomi, transkriptom, amplikoni,...),
- globini sekvenciranja (10x – 300x),
- Pripravi vzorcev za sekvenciranje: sekvenciranje samo z ene strani (single-end), z obeh strani (paired-end), z ujemajočimi konci (mate-pair),...
- Izbira načina sekvenciranja: kratki – dolgi fragmenti (SNP, insercije/delecije, procesiranje DNA, miRNA,...)

Izbira reagentov za Illumina

<http://www.illumina.com/prebuilt/kit-selector/sequencing-sample-prep-kit-selector.html?iframe>



# Globina sekvenciranja

Globina pomeni kolikokrat bo določen fragment DNA prebran med sekvenciranjem.

Večja globina = večja statistična moč/manjši del DNA sekvenciran

Odvisna od vrste študije in vrste vzorca (velikost genoma):

- SNP v človeškem genomu (pogost, redek) : 10 – 30 x
- SNP v zarodni liniji (podedovan) : 50 x
- SNP v somatski celici : 300x
- transkriptom : odvisno od ravni izražanja genov

Lander/Watermanova enačba:

C – globina

G – dolžina haploidnega genoma

L – dolžina branja

N - število branj

$$C = \frac{L * N}{G}$$

# Globina pokritja

HiSeq Output Calculations	
TruSeq v3 Reagents (one flow cell)	
Clusters/mm <sup>2</sup> (800K @85%PF) %PF may vary based on library	680,000
Area of a lane (mm <sup>2</sup> )	273.6
Reads/lane	186,048,000
Genome or region size (in bases)	3,000,000,000
Coverage	30
Total number of cycles (e.g. 200 for 2x100)	200
<b>Total output required (in bases)</b>	<b>90,000,000,000</b>
<b>Output/lane (bases/lane)</b>	<b>37,209,600,000</b>
<b>Number of lanes</b>	<b>2.42</b>
<b>Number of samples/lane</b>	<b>0.41</b>

# Globina vs dolžina

Človeški genom je 3.000.000.000 baz (3 Gb) dolg.

Genom ima okoli 180.000 exonov, kar predstavlja 1% genoma.

Človeški eksom je 30.000.000 baz (30 Mb) dolg.

V povprečju sekvenciranje eksonov odkrije 12.000 različic (okoli 10 % je novih), sekvenciranje genoma pa okoli 5 milijonov različic (le nekaj % je novih).

Zato je sekvenciranje eksonov cenejše in lahko tudi dovolj informativno za iskanje mutacij v genih, ki sodelujejo v razvoju bolezni in omogoča natančnejše diagnoze pacientov.



# Priprava vzorcev in sekvenciranje

Koraki skupni vsem aplikacijam druge generacije so:

- 1) priprava DNA knjižnice,
- 2) pomnoževanje s PCR,
- 3) čiščenje in priprava za sekvenciranje,
- 4) sekvenciranje,
- 5) analiza podatkov.

Vzorci:

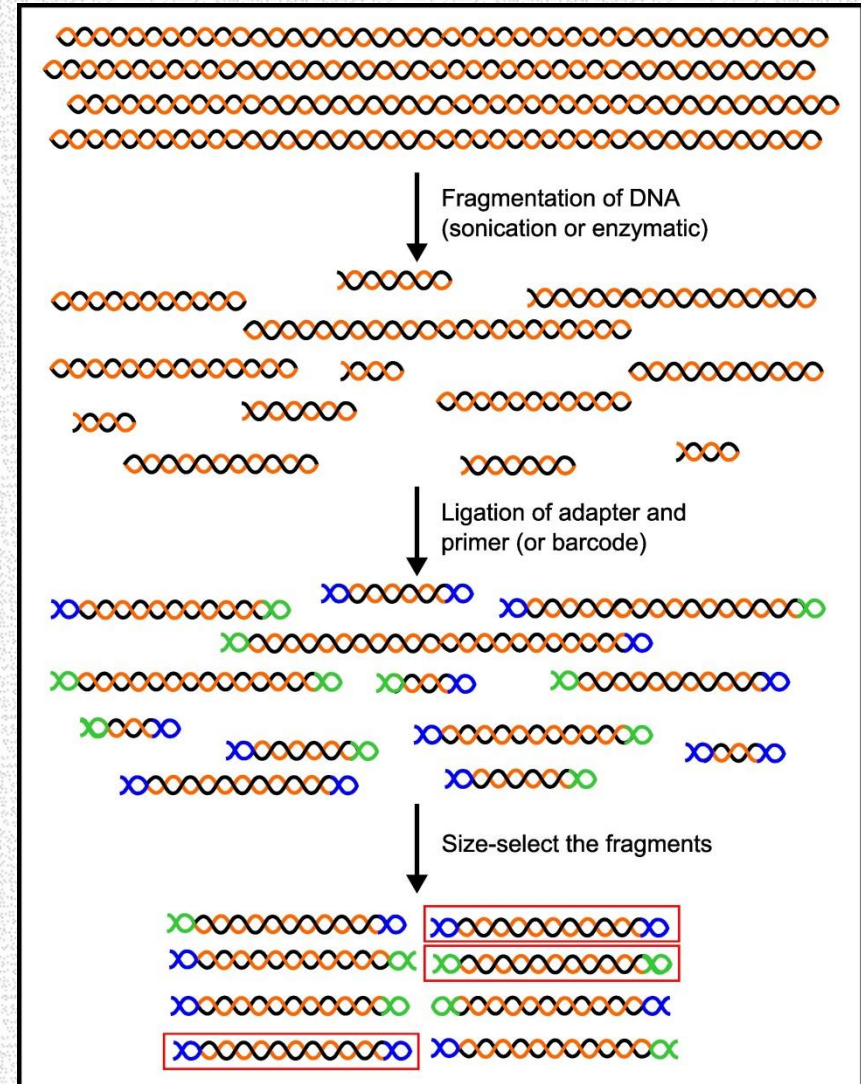
- DNA: genomska, plazmidna, amplikoni
- RNA: mRNA, miRNA, rRNA

Kvaliteta je zelo pomembna!!! (260/280, 260/230, stopnja razgrajenosti)  
Izbira metode izolacije ima močan učinek na kvaliteto in kvantiteto izolata!

# 1) Priprava DNA knjižnice

Intaktna genomska DNA (do 5  $\mu\text{g}$ )

- 1) Fragmentacija
- 2) Poravnava visečih koncev
- 3) Lepljenje adapterjev
- 4) Čiščenje fragmentov
- 5) Kratek PCR
- 6) Izolacija fragmentov določene dolžine in čiščenje
- 7) Analiza kvalitete DNA knjižnice



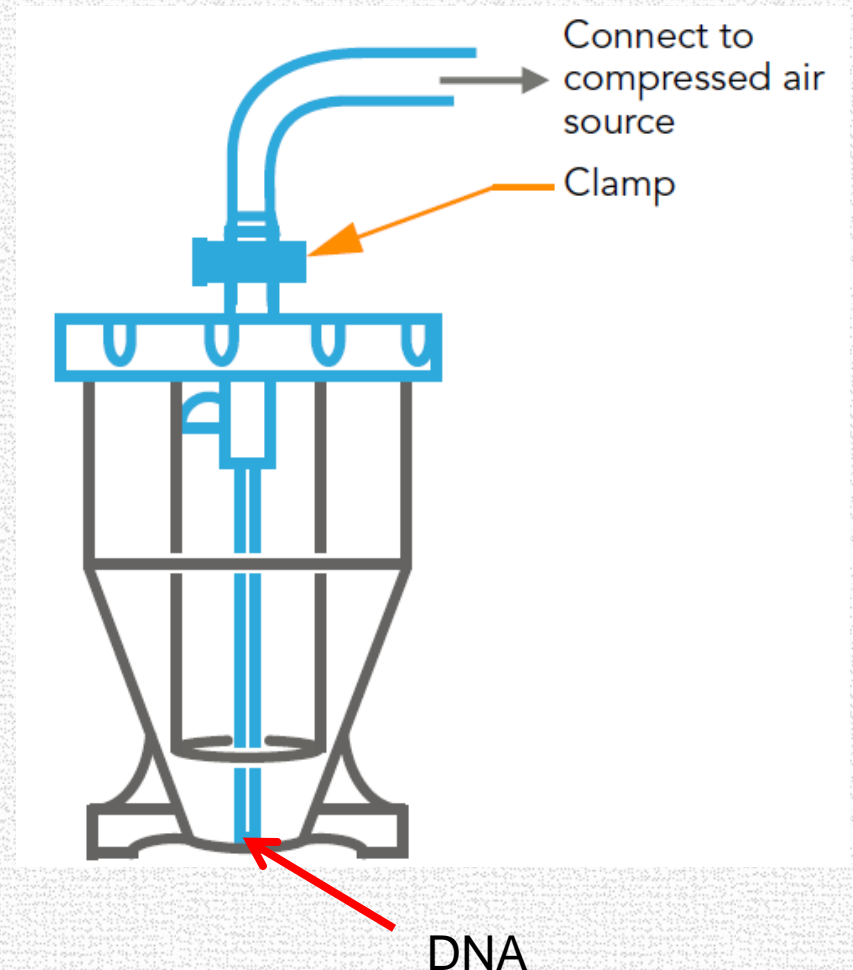
# 1) Priprava DNA knjižnice

## 1) Fragmentacija DNA

Z nebulizacijo ali z ultrazvokom („sonication“) naključno raztrgamo DNA na različno dolge fragmente.

Nebulizacija: DNA raztopino večkrat potisnemo skozi ozko odprtino s pomočjo stisnjenega zraka ali dušika.

Koliko odvisno od izbrane tehnologije sekvenciranja:  
Illumina < 800 bp, SOLiD od 100 do 250 bp,...

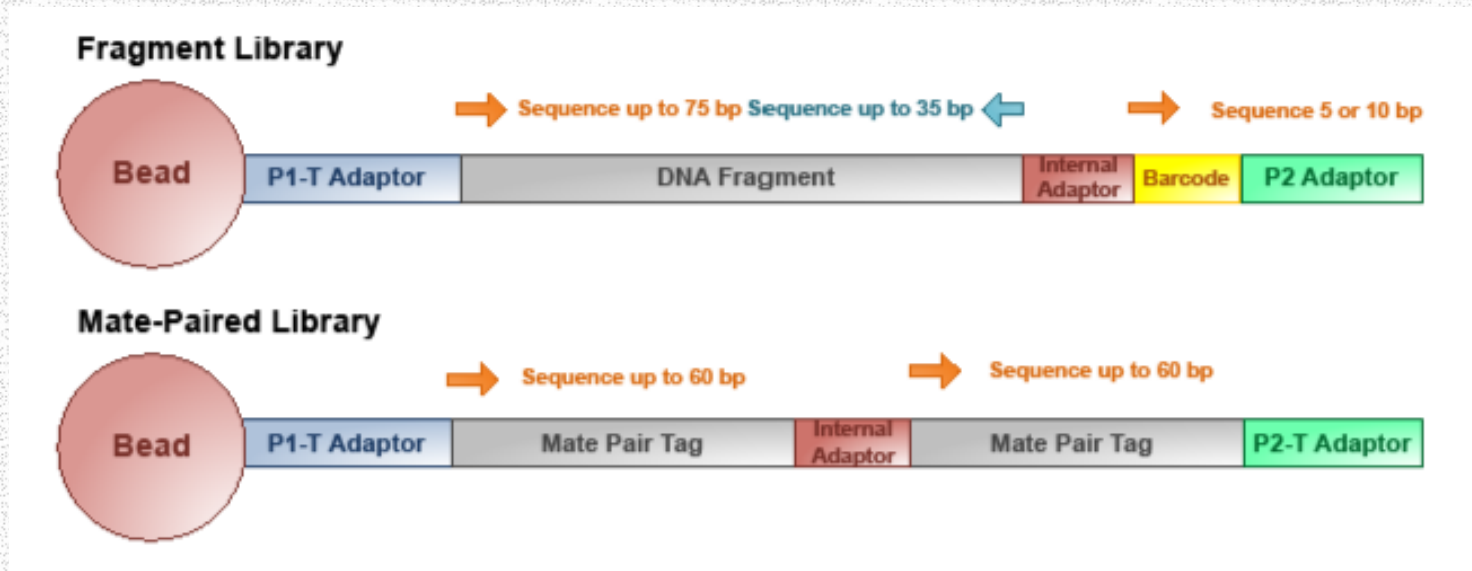




# 1) Priprava DNA knjižnice

## 2 – 3) Poravnava visečih koncev in lepljenje adapterjev:

1. DNA je raztrgana in ima zato 5'-in 3'- viseče konce, ki se poravnajo v ravne konce in doda se dA rep.
2. Lepljenje dveh adapterjev z ligazo, vsakega na svoj konec:  
454 sistem: adapter A in B, SOLiD : adapter P1 in P2.



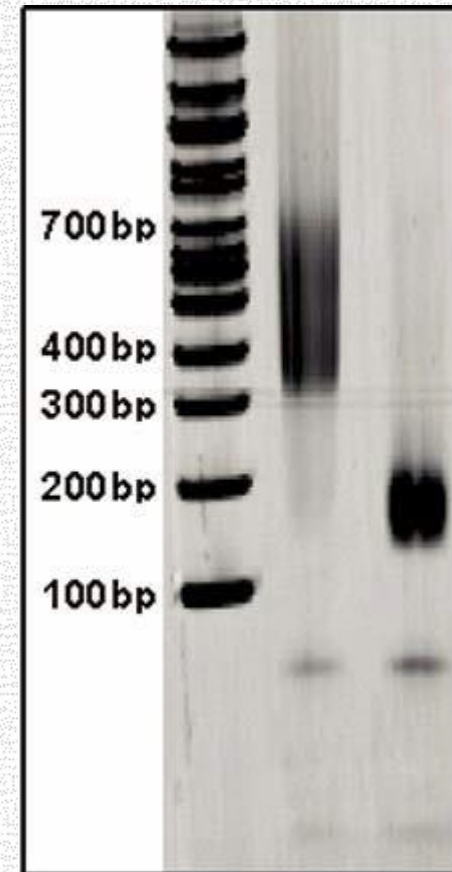
# 1) Priprava DNA knjižnice

4 – 6) Čiščenje in izolacija DNA fragmentov določene dolžine

Kratek PCR, da namnožimo fragmente z različnima adapterjema.

Med čiščenjem z magnetnimi kroglicami

Agaroznem gelu



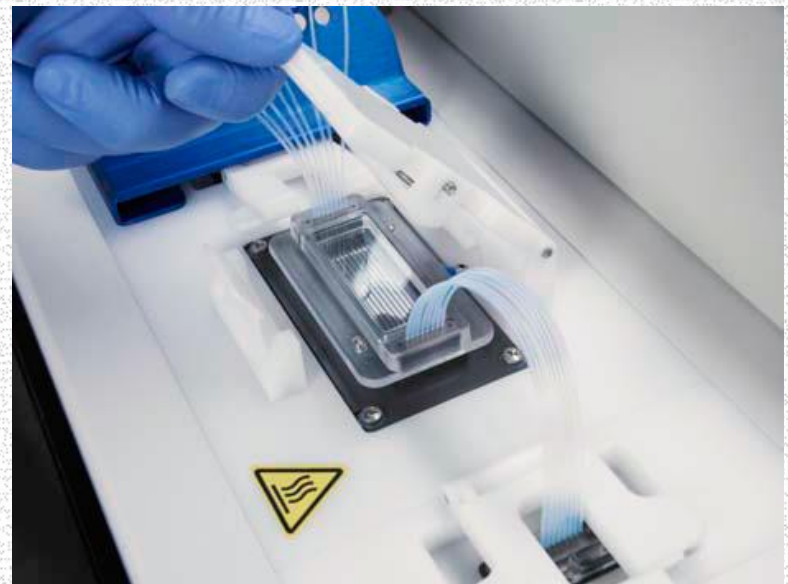
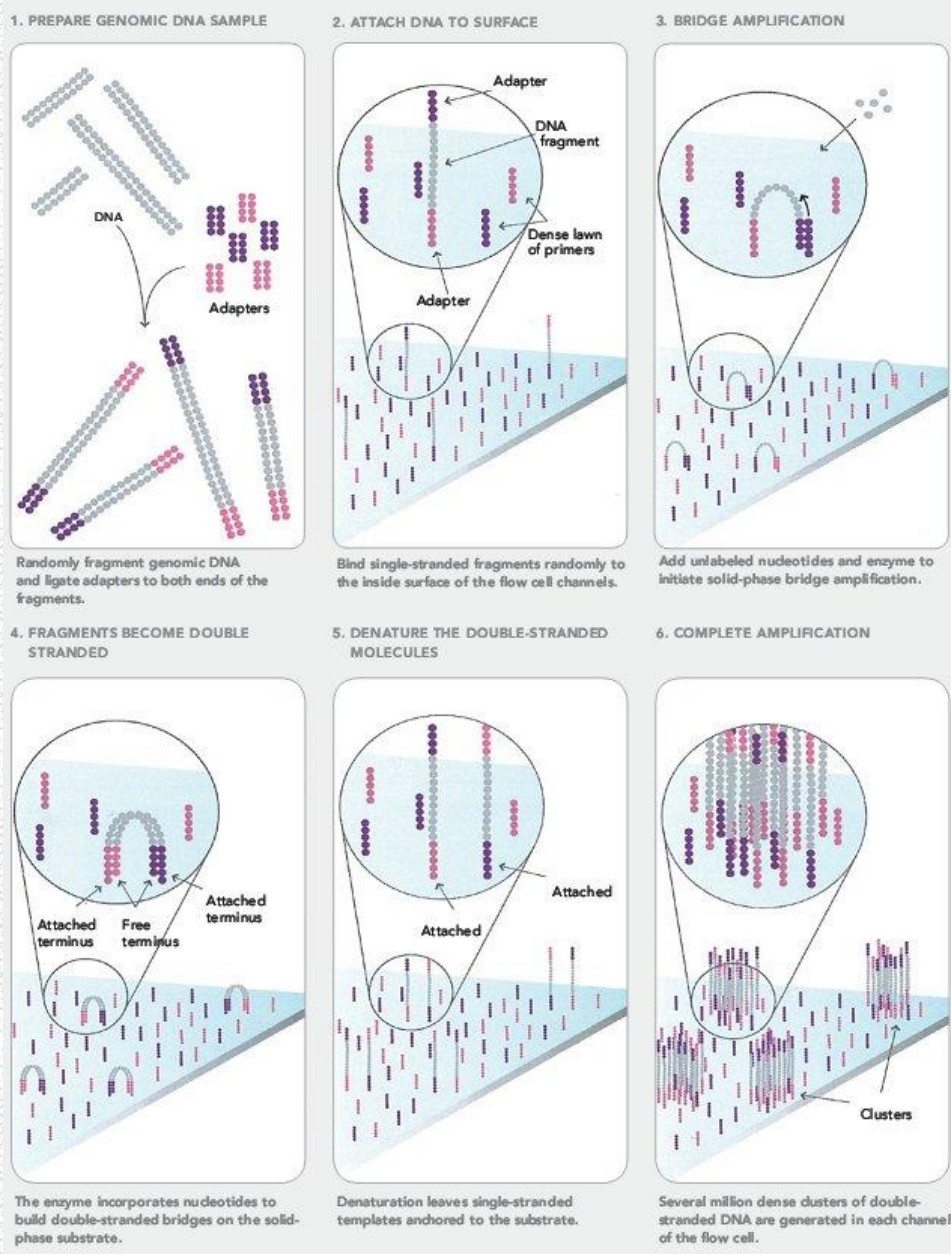
## 2) Pomnoževanje DNA knjižnice

Koraki pri pomnoževanju knjižnice DNA fragmentov:

1. DNA fragmenti se preko adapterjev vežejo na:
  - kroglice (SOLiD, Roche, Ion Torrent), po en fragment na kroglico,
  - pretočno celico (flow cell) (Illumina).
2. PCR:
  - Emulzijski na kroglicah (Roche, SOLiD, Ion Torrent)
  - Na nosilcu preko mostu (Illumina)
3. Čiščenje produktov
4. Priprava za sekvenciranje







# Sekvenciranje – detekcija fluorescence s CCD kamero

## ❑ 454 sistem (Roche):

Pomnoženi DNA fragmenti so vezani na kroglice (eden na kroglico in mešanico teh nanesimo na pikotitrsko ploščico (picotiter plate). Ploščica je iz zlepljenih optičnih vlaken. Sledi določevanje zaporedja na kroglicah s sintezo DNA verige.

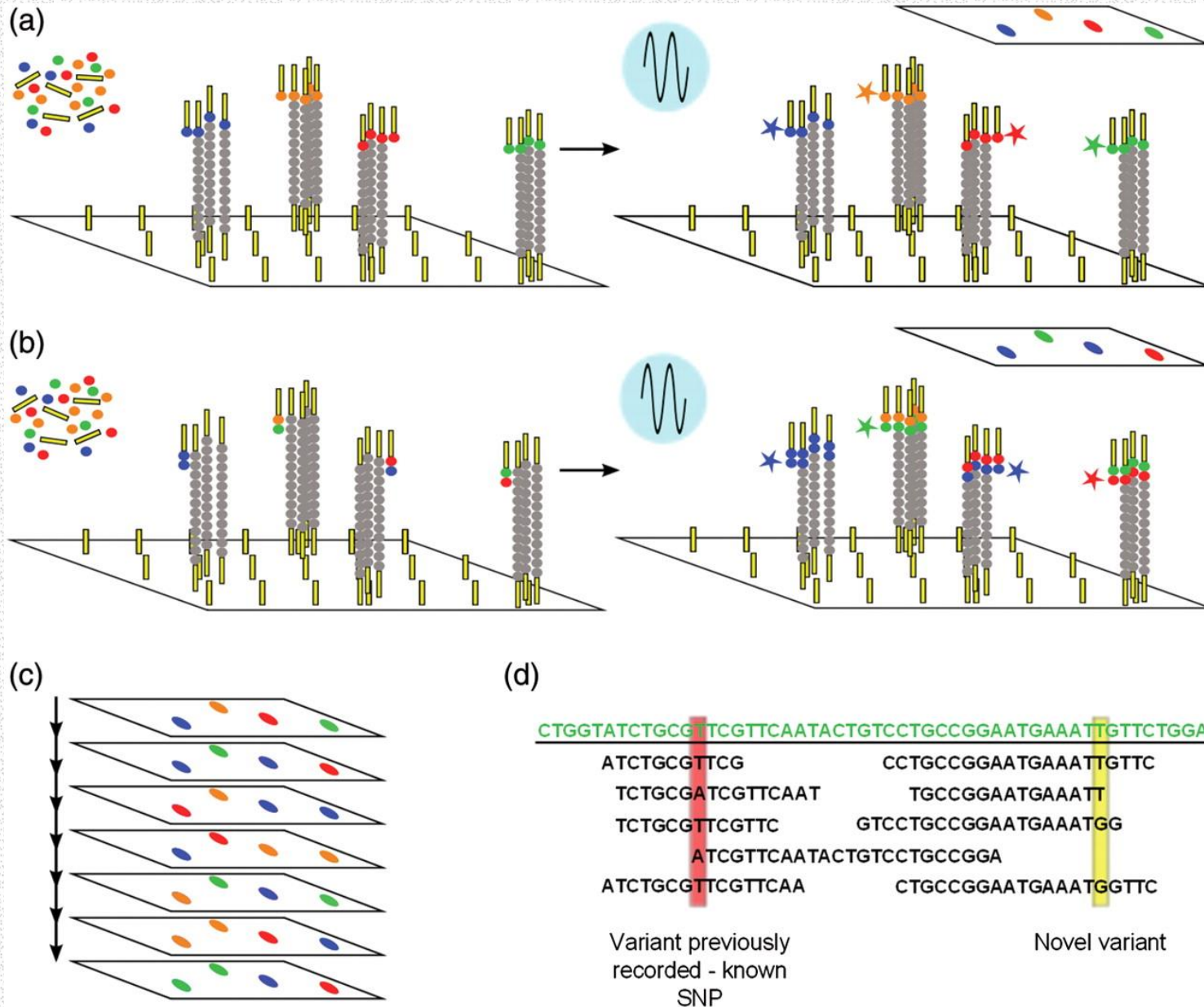
## ❑ SOLiD:

Pomnoženi DNA fragmenti se vežejo na stekleni nosilec. Sledi sekvenciranje z ligacijo oligonukleotida, detekcijo, odstranitvijo produkta in lepljenjem novega oligonukleotida in ponavljanje ciklov.

## ❑ Illumina:

DNA fragmenti se pomnožijo na pretočni celici in na tej poteka tudi sekvenciranje teh fragmentov preko sinteze.








# Sekvenciranje RNA

RNA:

- Kvaliteta:
  - čistost:  $260/230 = 2,0-2,2$ ;  $260/280 = 2,$
  - RIN: od 7 do 10 (nad 8), vzorci naj varirajo znotraj 1 in 1,5 razreda.
- Količina: 25 ng mRNA ali 200 ng RNA brez rRNA ali do 4 ug totalne RNA.
  
- Odstranitev DNA
- Odstranitev rRNA
- Obogatitev mRNA

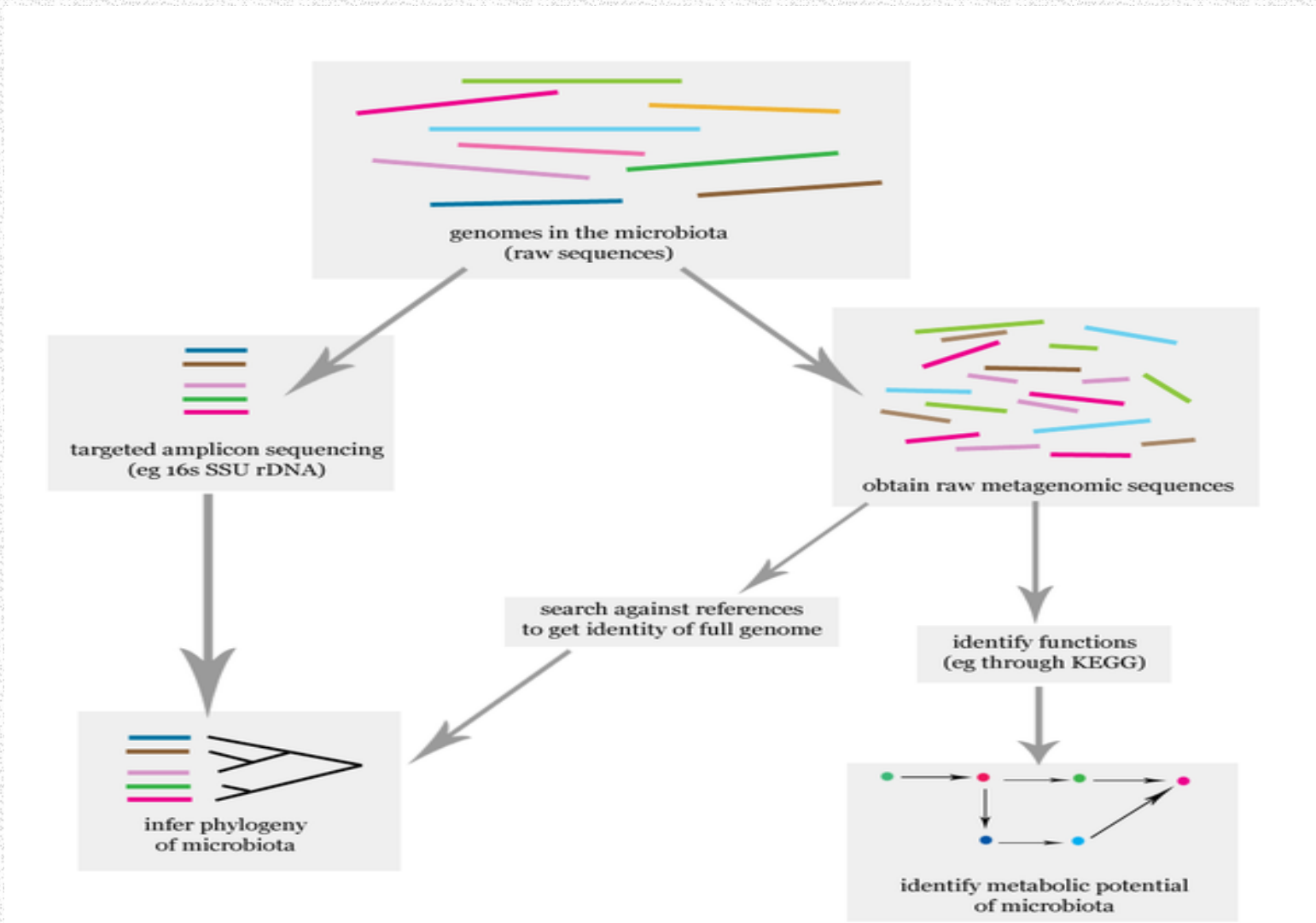
# RNA-Seq –priprava knjižnice

Koraki so:

1. Če začneš s totalno RNA, sledi najprej selekcija RNA s poli A repi preko magnetnih kroglic (Illumina).
  2. Če začneš z RNA brez rRNA ali že izolirano mRNA, ni selekcije.
  3. Fragmentacija RNA,
  4. Prepis v dvoverižno cDNA,
  5. Na obe strani se doda A,
  6. Lepljenje adapterjev,
  7. PCR,
  8. Kontrola knjižnice (velikost fragmentov, čistost in količina).
- 

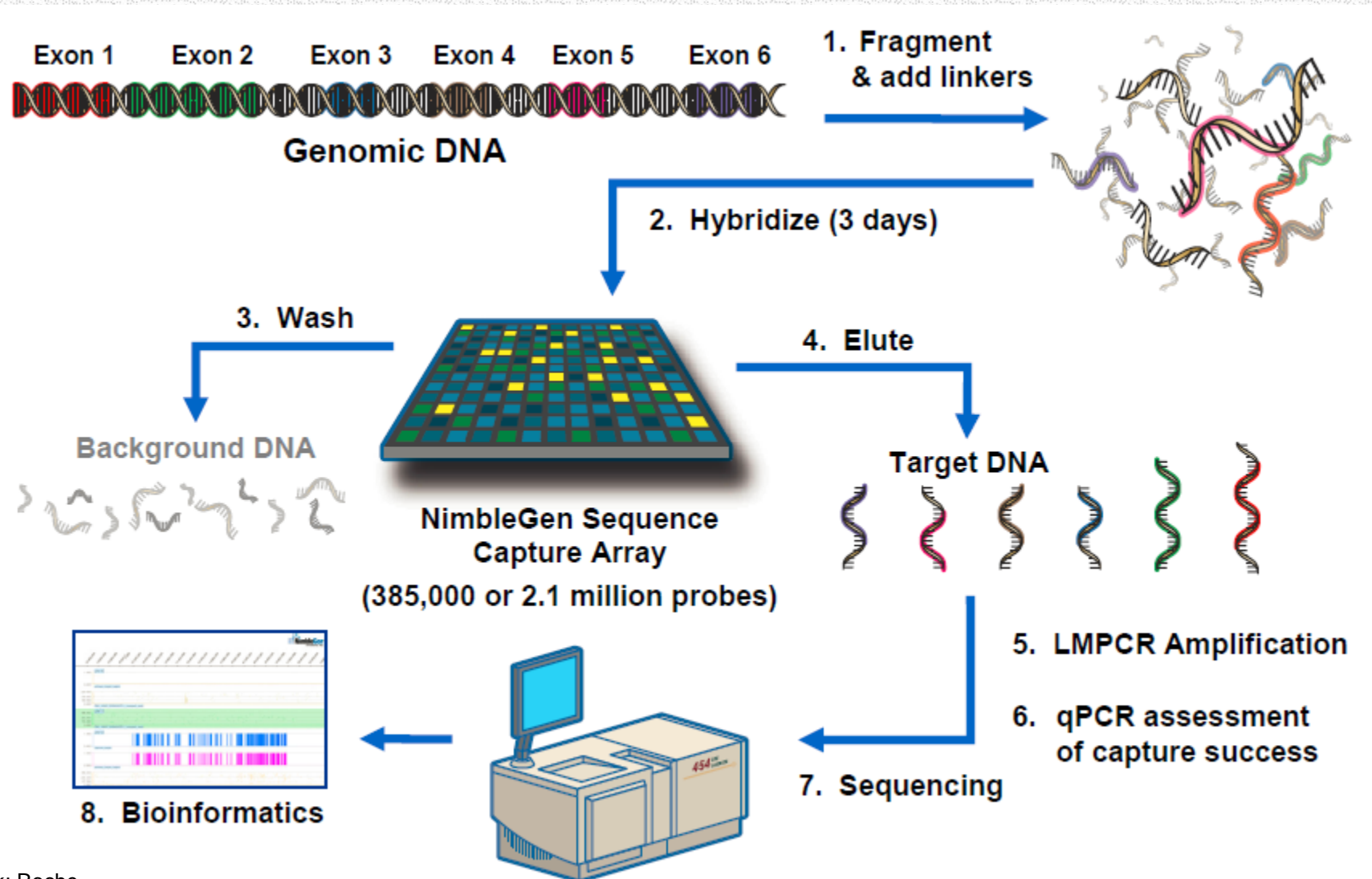
Sekvenciranje

# Mikrobiom





# Tarčno resekveniranje - eksom



# Drowned in next generation sequencing data

HELP!



<http://biocomicals.blogspot.com>

AG



# Analiza podatkov

## ❑ Količina podatkov – shranjevanje in analiza (terabiti)

Roche – 500 milijonov baz na analizo

Illumina – milijardo baz na analizo

Primer: SOLiD – ena analiza = 90 do 200 gigabaz podatkov –  
vzporejanje z referenčnim genomom: 4 h na 250 serverjih

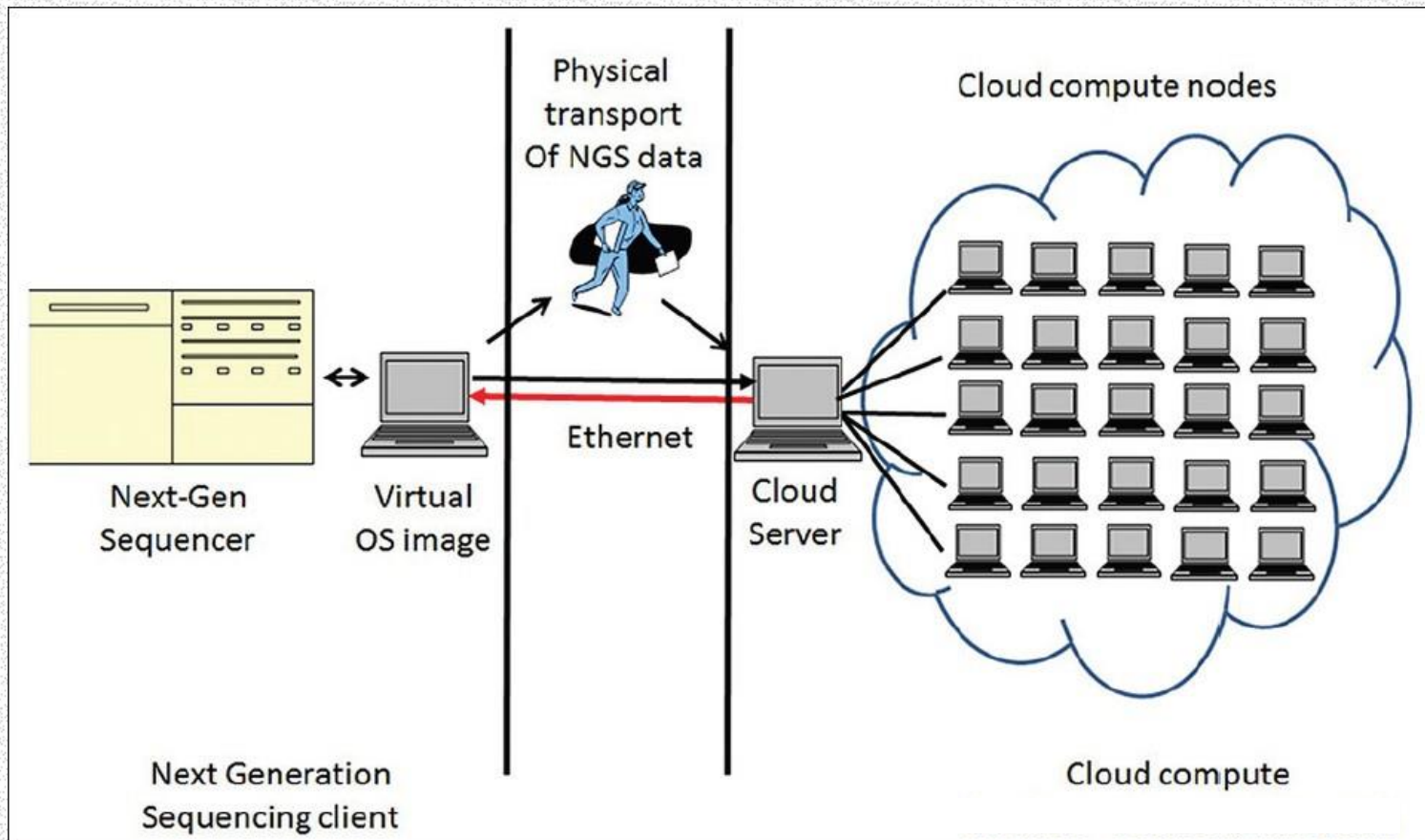
## ❑ Ogromna izbira programov in metod (preko 200)

Komercialni – enostavni za uporabo, podpora, dragi, ni izbire

Prosto dostopni - zahtevajo več bioinformatičnega znanja (UNIX programiranje), ni podpore, zastoj, ni omejitev v analizi

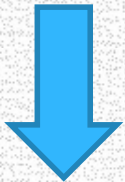


# Analiza podatkov - oblaki



# Vzporejanje odčitkov zaporedij

Genomska DNA



Fragmenti DNA



Vzporejanje na referenčni genom



**TGACATGTCGATCGATGCTGGATCGA**  
**GACATG GATCGA GATCGA**  
**GTCGAT GCTGGA**  
**GACATG CGATGC GGATCG**

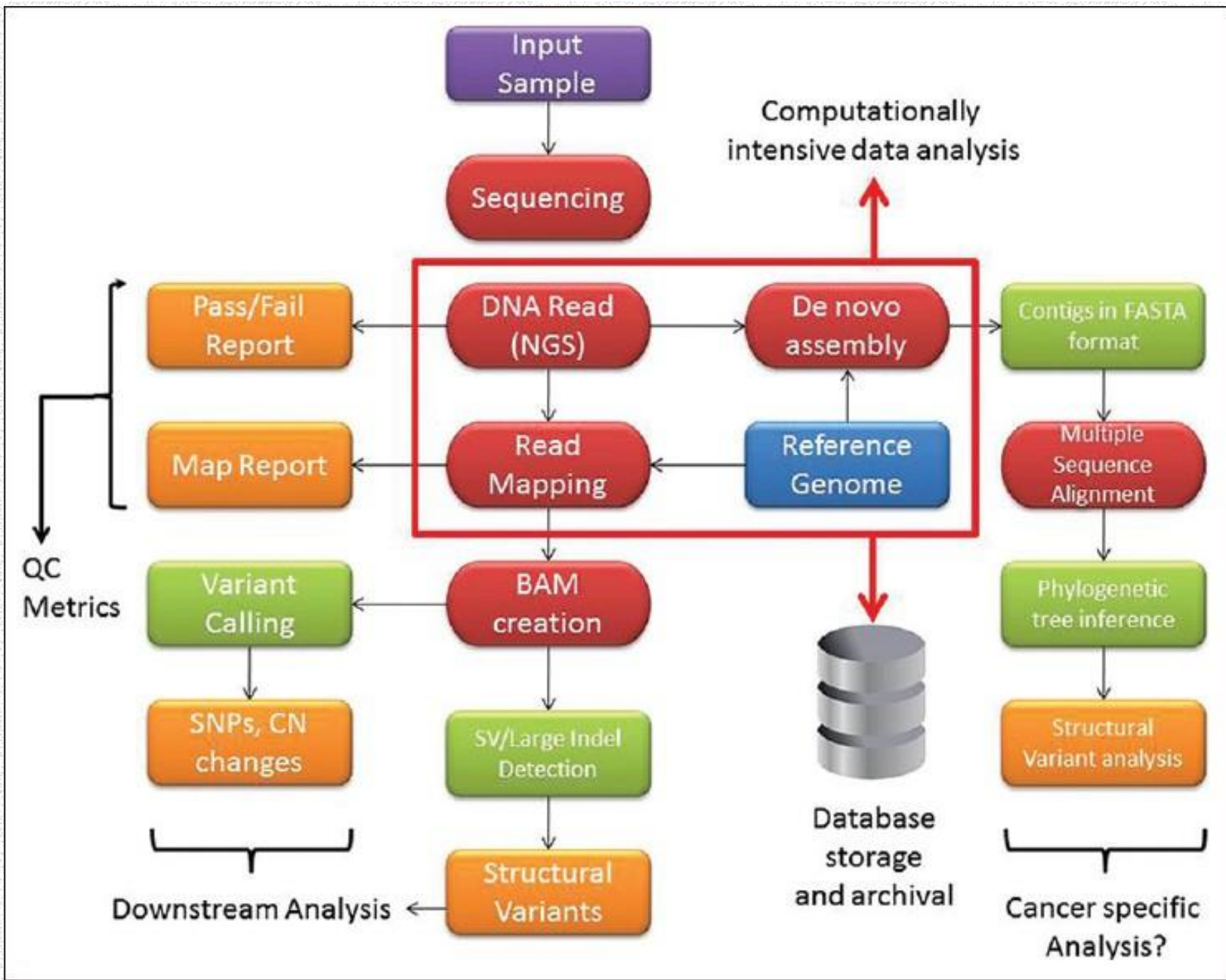
# Koraki v analizi podatkov (iskanje različic)

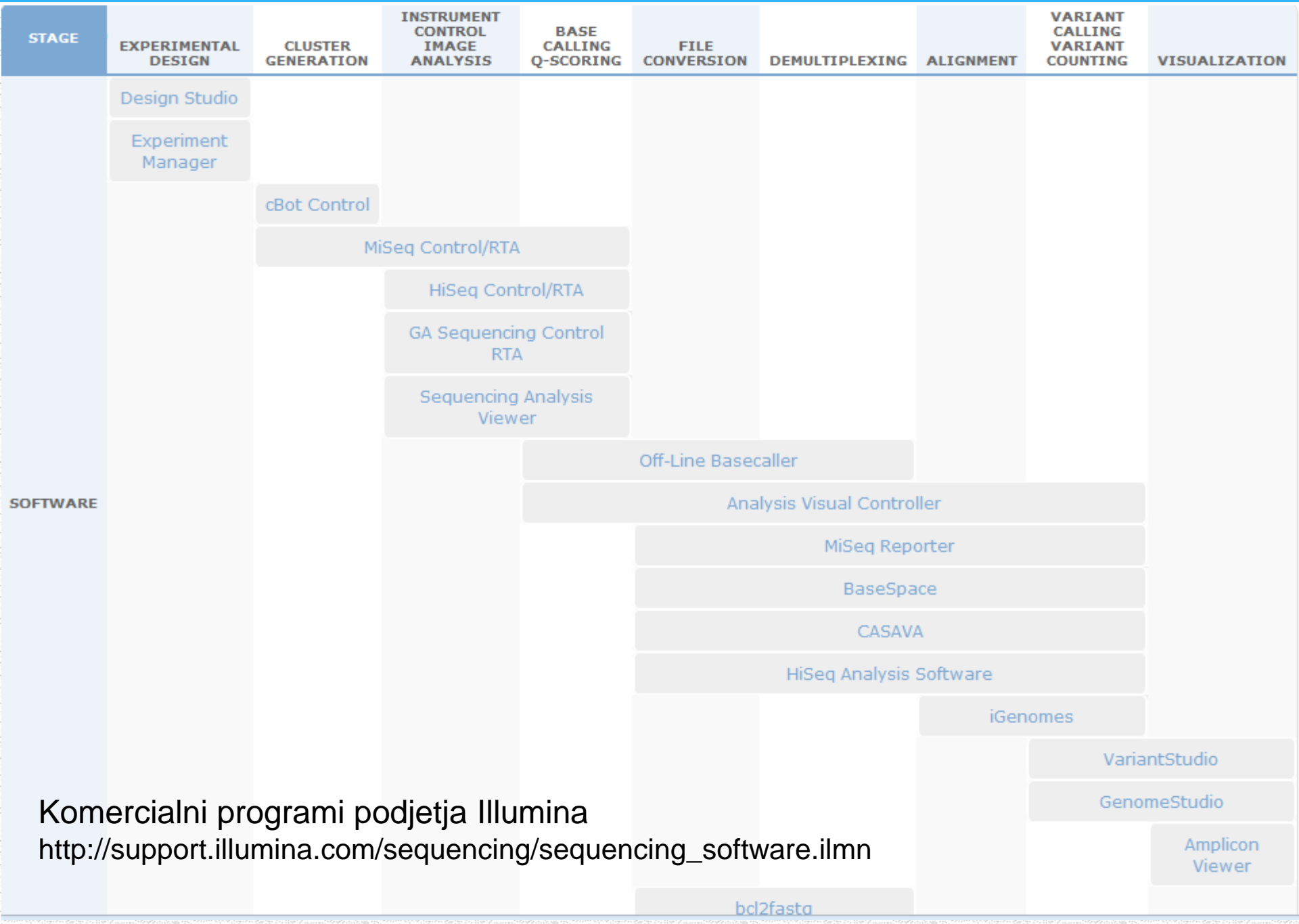
1. Procesiranje slike in generiranje odčitkov zaporedja (reads)
2. Ocena kvalitete odčitkov zaporedja
3. Vzporeditev odčitkov zaporedja na referenčni genom
4. Identifikacija različic (variant)
5. Anotacija različic
6. Vizualizacija

Pabinger S. *et al.*, A survey of tools for variant analysis of next-generation genome sequencing data, 2013, Brief Bioinform, 21.

<https://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course>







Komercialni programi podjetja Illumina

[http://support.illumina.com/sequencing/sequencing\\_software.ilmn](http://support.illumina.com/sequencing/sequencing_software.ilmn)

# 1) Ocena kvalitete

- Ocena kvalitete odčitkov zaporedij. Ali smo dobili dovolj odčitkov zaporedij v analizi? Če ne, zavrnamo celotno analizo.
- Potrebno je odstraniti odčitke, ki ne ustrezajo kriterijem. So verjetno artefakti nastali med pripravo ali sekvenciranjem vzorca.
- Potrebno je skrajšati oziroma odrezati dele odčitkov, kjer branje ni več zanesljivo ali vsebuje zaporedje adapterja in drugih pomožnih zaporedij.
- Vsaka platforma ima svoje kriterije in programe, ki to omogočajo. Obstajajo tudi neodvisni programi.

## 2) Vzporeditev odčitkov zaporedij

- ❑ Problem: odkriti vir odčitka zaporedja – „paired-end“ metoda je pomembna sekvenciranje genoma ali eksonov.
- ❑ Problem vzporeditve odčitkov zaporedij, ki se ne vzporejajo dobro zaradi na primer insercije/delecije, ki je ni v referenčnem genomu; ponavljajočih se zaporedij,...Taki odčitki so izločeni iz nadaljnjih analiz – izgubljena informacija.
- ❑ Pomembna izbira referenčnega zaporedja:  
UCSCS (University of Santa Cruz) – ENCODE – GRC (Genome Reference Consortium)
- ❑ Na voljo je vrsta programov vsak s svojim pristopom k poravnavi odčitkov.  
Bowtie, BWA, SOAP, MAQ, mrFAST, Novoalign, SSAHA2, Stampy,...



# Identifikacija, anotacija in vizualizacija različic

- Ali identificirajo zarodne različice  
CRISP; GATTK, SAMtools, SNVer, VarScan 2,..
- Ali identificirajo somatske različice  
GATK; SAMtools, SomaticSniper, VarScan 2,...
- Identifikacija CNV – razlika v številu kopij  
CNVnator, RDXplorer, CONTRA, ExomeCNV,...
- Identifikacija SV – strukturnih različic  
BreakDancer, Breakpointer, CLEVER,....
  
- Anotacija različic:  
ANNOVAR, AnniTool, NGS-SNP, VARIANT, VEP,.....
  
- Vizualizacija različic:  
Na spletnih straneh: Ensembl, USCS, VEGA,  
Artemis, Integrative Genomics, Viewer,...

# Sestava zaporedij *de novo*

Sestava transkriptoma *de novo*

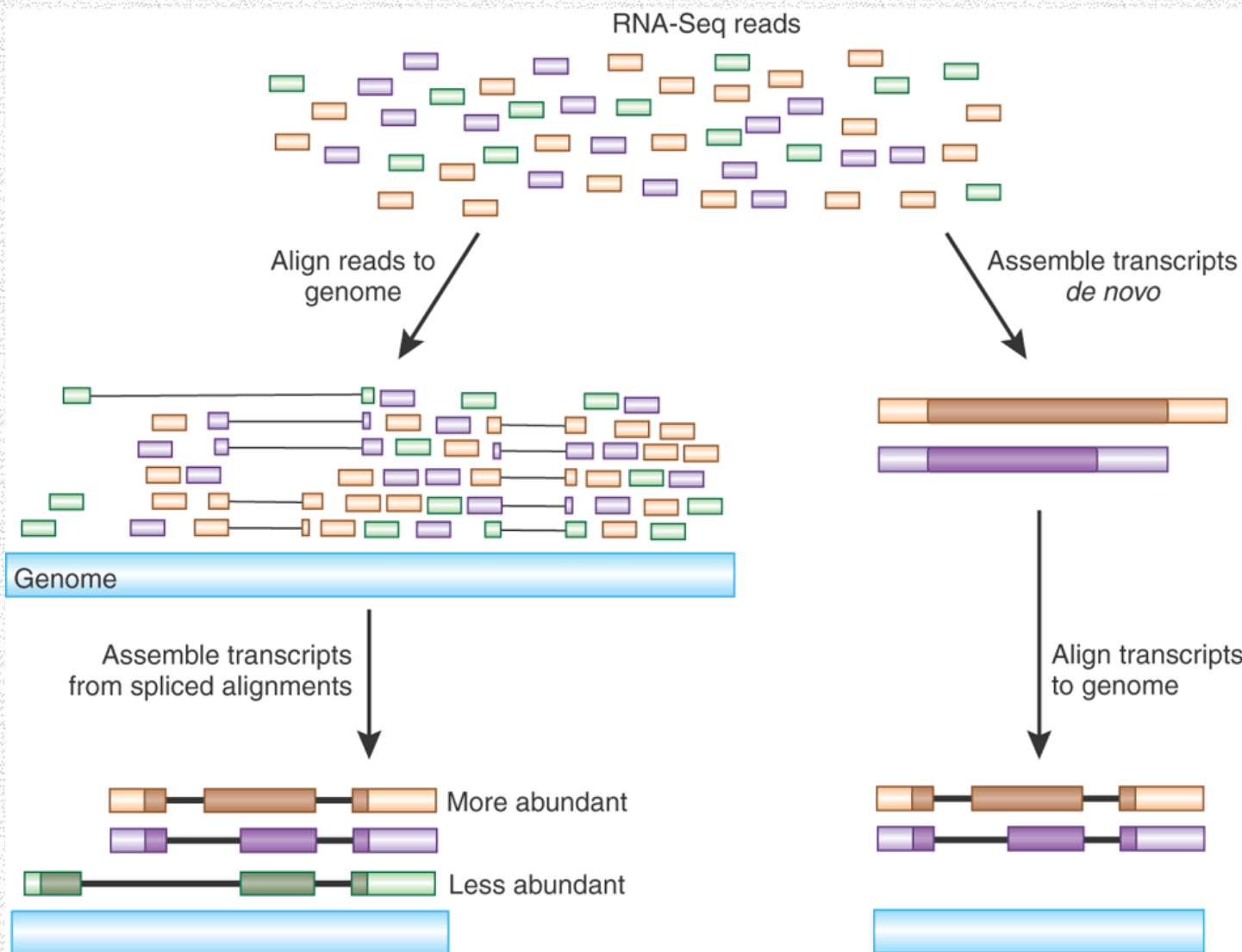
SOAPdenovo-Trans, Trinity, Cufflinks, Velvet/Oases,...

Sestava genoma *de novo* (velik ali mali genom, kratki ali dolgi fragmenti)

Velvet, EULER, ABySS, SOPAdenovo, MIRA3, CABOG,..  
(reads – contigs - scaffolds)

Anotacija genov

# Analiza pri sekvenciranju RNA





# GS Junior

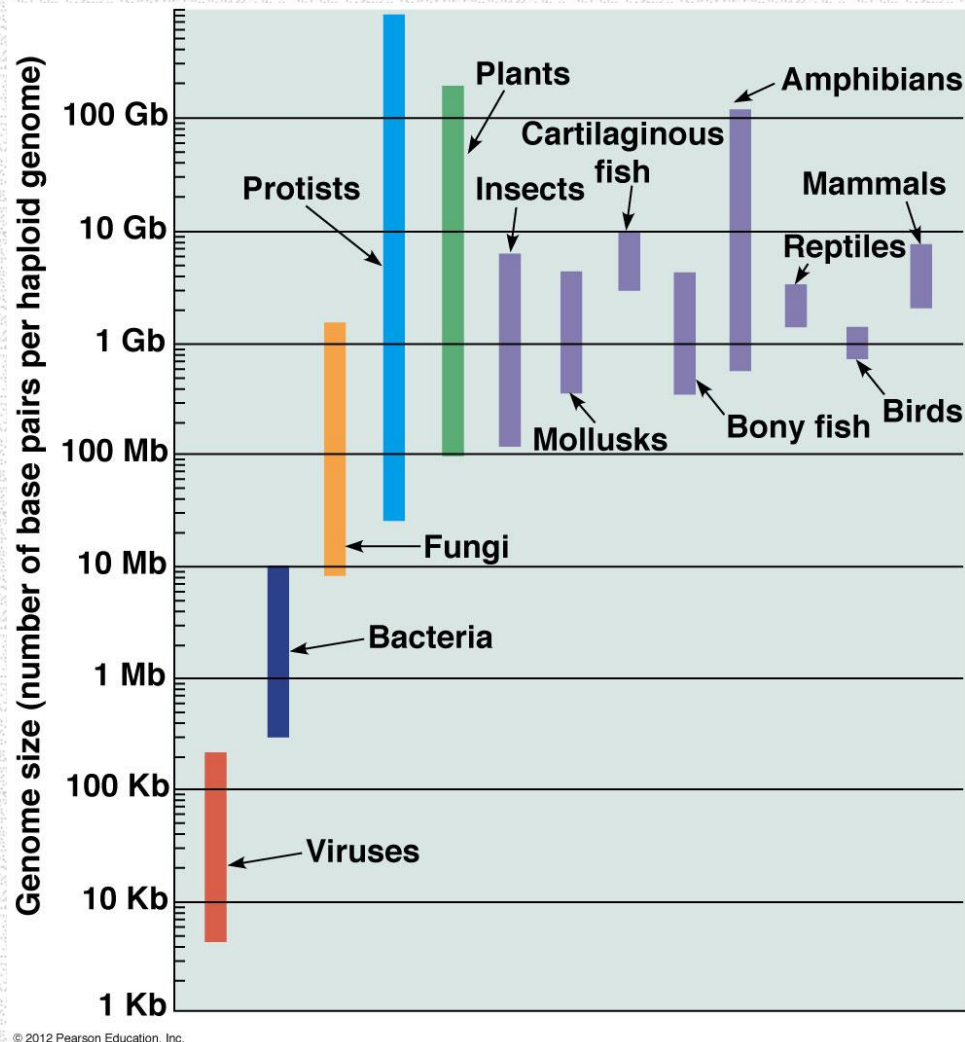


<http://vimeo.com/24229714>

# GS Junior

Pikotitrna plošča	1
Dolžina branja	~400 bp
Učinkovitost	- 85% vseh prebranih baz je iz produktov dolgih >400 bp - 45% vsej prebranih baz je iz produktov dolgih >500 bp
Vseh prebranih baz	~35 Mb
Fragmentov na analizo	~100,000 „shot gun“, 70,000 amplikon
Natančnost	Q20 dolžine produktov je 400 baz (99% natančnost pri 400. bazi in višja za baze pred njo)
Čas analize	10 ur
Vzorci	genomska DNA, PCR produkti (amplikoni) ali cDNA,...
Hkratno sekvenciranje	do 153 vzorcev

# Velikost genomov





# GS Junior

Lahko sekvenciramo:

- celotno genomsko DNA (virusi, bakterije, glive),
- dele daljših genomskih DNA (živali, rastline) z namenom resekveciranja,
- PCR produkte (amplikone), BAC in cDNA oziroma transkriptom.

Aplikacije so:

- Tarčno resekveciranje: genotipizacija, detekcija redkih različic, sekvenciranje posameznih genov ali področij genoma.
- Sekveciranje genoma: sekvenciramo SNPs, insercije, delecije, genomske regije, čista genomska DNA bakterij, rastlin, živali, gliv, virusov, BAC ali katerekoli drugega vira DNA.
- Metagenomika: diverziteta mikrobnih združb.
- Sekveciranje transkriptoma: novi transkripti in isoforme.

# Metode priprave DNA knjižnice

Pristopi k pripravi knjižnice za sekvenciranje so:

- ❑ »**Shot gun**«: dolge odseke DNA fragmentiramo v naključno dolge fragmente, jim nalepimo posebne adapterje in jim preberemo zaporedje; poseben program jih sestavi v pravo zaporedje.
- ❑ **Amplikoni**: na PCR produkt ligiramo adapterje ali pa uporabimo dolge oligonukleotide, ki imajo poleg specifičnega zaporedja tudi adapter.
- ❑ »**Paired end**«: genomsko DNA fragmentiramo v 20, 8 ali 3 kb dolge fragmente, na konce nalepimo adapterje in jih zlepimo v krog. Krožna DNA je fragmentirana in fragmenti z adapterji so izolirani in sekvencirani.
- ❑ **cDNA knjižnica**: RNA se fragmentira in nato prepíše v cDNA s pomočjo naključnih začetnih oligonukleotidov, doda se druga veriga, adapterji in sledi sekvenciranje.

## Izbira metode - pokritost

S katero metodo bomo sekvencirali je odvisno od tega kaj želimo detektirati. Primer je resekveciranje določenega gena in iskanje SNP.

- ❑ Poznati moramo ali je sprememba zaporedja redka ali pogosta oziroma ali je somatskega izvora ali podedovana kajti to vpliva na pokritje/globino sekvenciranje, ki jo moramo imeti v analizi, da jo bomo lahko identificirali.
- ❑ Primer je sprememba zaporedja, ki je prisotna samo v 1% celic, kar pomeni, da jo bomo na 5000 odčitkov prebrali samo 50x. To pomeni, da lahko samo 14 vzorcev hkrati analiziramo.
- ❑ Naslednji primer je detekcija heterozigota, kjer potrebujemo 40x pokritje, da bomo imeli 99,9-odstotno verjetnost, da ga bomo detektirali.
- ❑ V eni analizi dobimo 70.000 odčitkov, kar pomeni, da z 40x pokritjem lahko preberemo 1750 amplikonov dolgih 400 bp hkrati.
- ❑ Spoznali se bomo s tarčnim resekveciranjem gena z amplikoni.



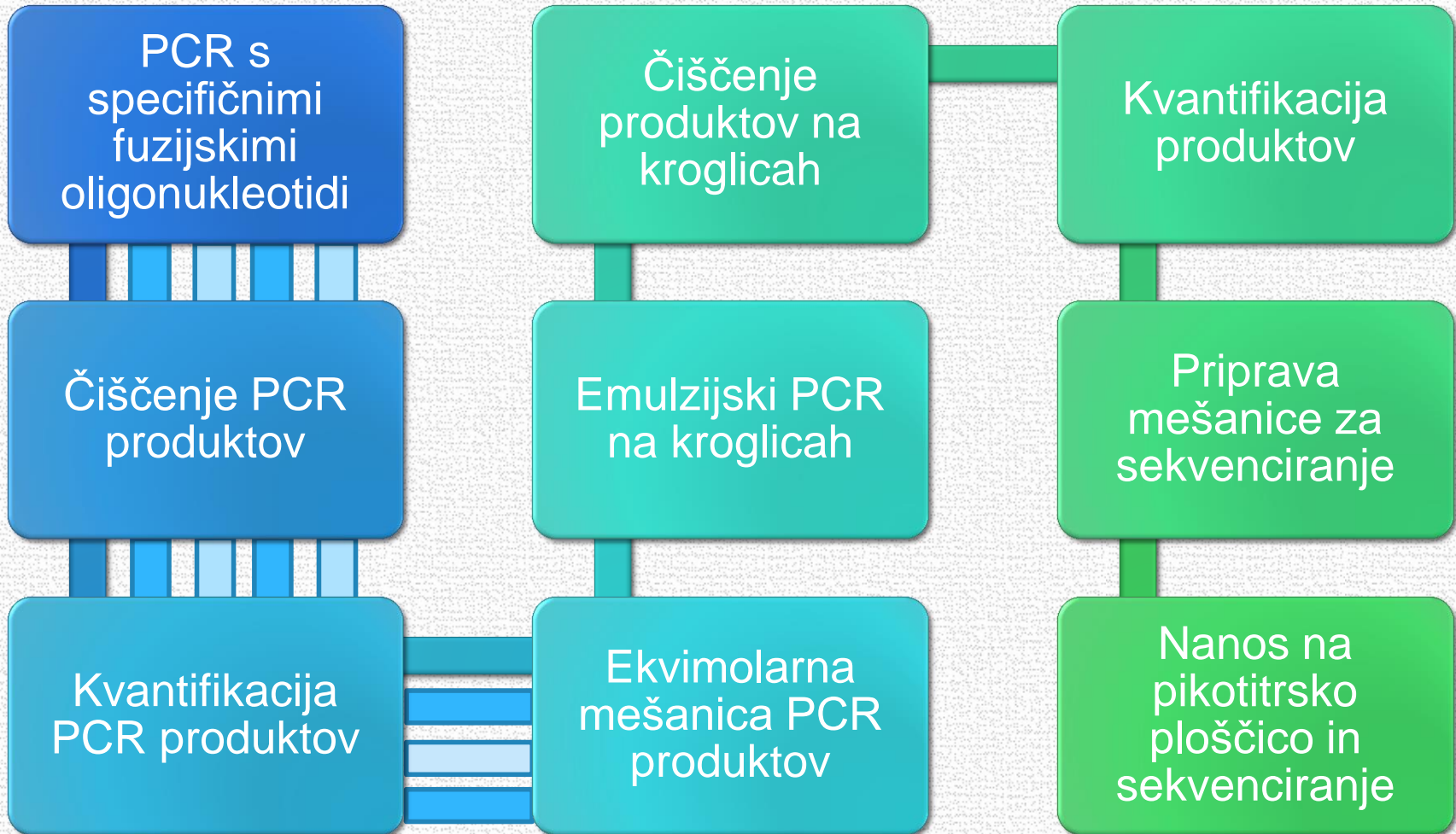
## Izbor metode

Prvi korak je izbor metode priprave PCR produktov in načina sekvenciranja.

- Odločiti se morate ali boste dodajali adapter in MID med pomnoževanjem produkta in sta ti zaporedji del začetnega oligonukleotida ali pa boste prilepili adapter, KEY in MID na PCR produkte po pomnoževanju.
- Pomembno je tudi vedeti ali želite sekvencirati kratke (pod 800 bp) ali dolge PCR produkte (nad 1500 bp) ali želite sekvenco prebrati z obeh verig ali samo z ene.

Primer: kratki PCR produkti (od 400 do 500 bp), kjer želimo prebrati DNA z obeh strani in kjer boste dodali adapter in MID med pomnoževanjem z vgnezenim PCR.

# Priprava vzorcev (2-3 dni)



# Priprava PCR produktov s specifičnimi fuzijskimi začetnimi oligonukleotidi

PCR reakcija s „**FastStart High Fidelity**“ PCR sistemom:

- Vsebuje „FastStart“ DNA polimerazo in termostabilen protein za kontrolno branje.

Velikost PCR produktov mora biti med 200 in 600 bp oziroma morajo biti do 150 bp različno dolgi.

5 ng genomske DNA



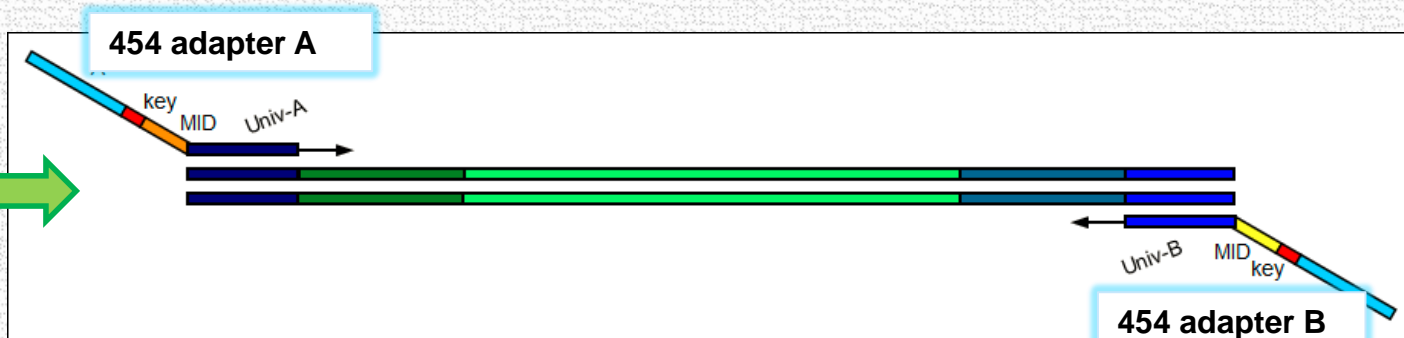
1. PCR



100-1000x



2. Vgnezdeni PCR





# Fuzijski začetni oligonukleotidi

**454 ADAPTER – KEY – MID(multiplex identifier) – specifično zaporedje**

Smerni začetni oligonukleotid (Primer A-Key):

5'-CGTATCGCCTCCCTCGCGCCA-TCAG-MID-{specifično zaporedje}-3'

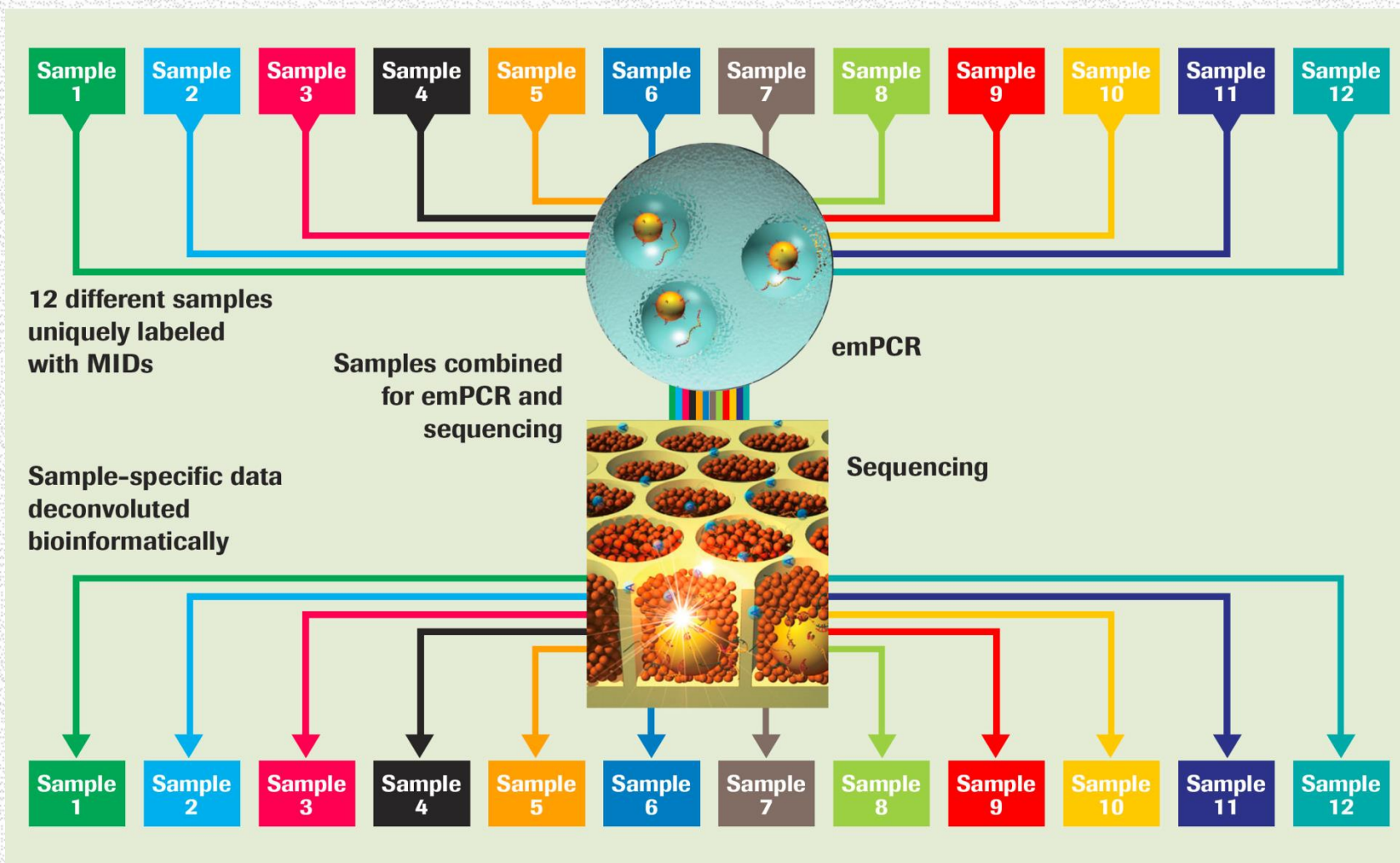
Protismerni začetni oligonukleotid (Primer B-Key):

5'-CTATGCGCCTTGCCAGCCCGC-TCAG-MID-{specifično zaporedje}-3'

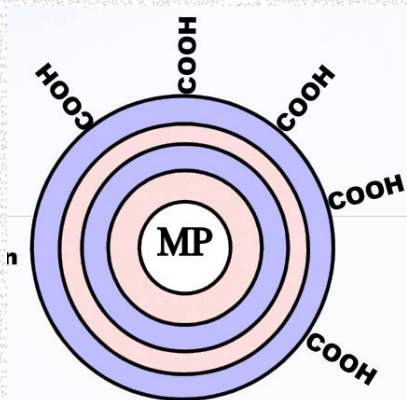
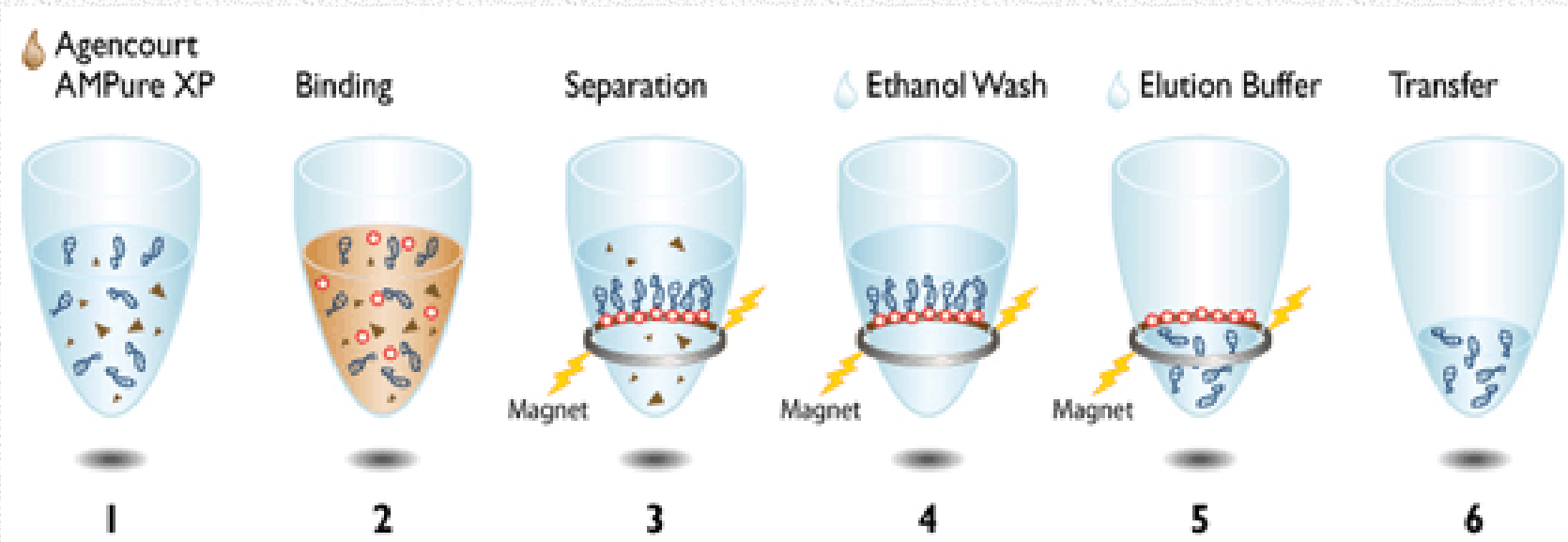
ID	MID Sequence
MID1	ACGAGTGCGT
MID2	ACGCTCGACA
MID3	AGACGCACTC
MID4	AGCACTGTAG
MID5	ATCAGACACG
MID6	ATATCGCGAG
MID7	CGTGTCTCTA

ID	MID Sequence
MID8	CTCGCGTGTC
MID9	TAGTATCAGC
MID10	TCTCTATGCG
MID11	TGATACGTCT
MID12	TACTGAGCTA
MID13	CATAGTAGTG
MID14	CGAGAGATAC

# Hkratno sekvenciranje vzorcev

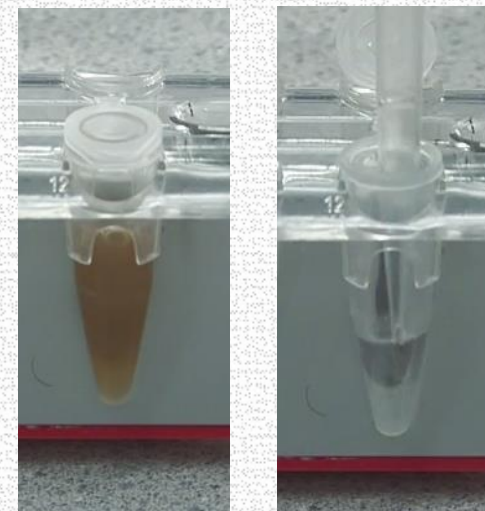


# Čiščenje PCR produktov



En PCR produkt = ena PCR reakcija = eno čiščenje

Čiščenje odstrani vse fragmente pod 200 bp.





# Kvantifikacija PCR produktov

- S PicoGreen reagentom na Nanodrop ND-3000

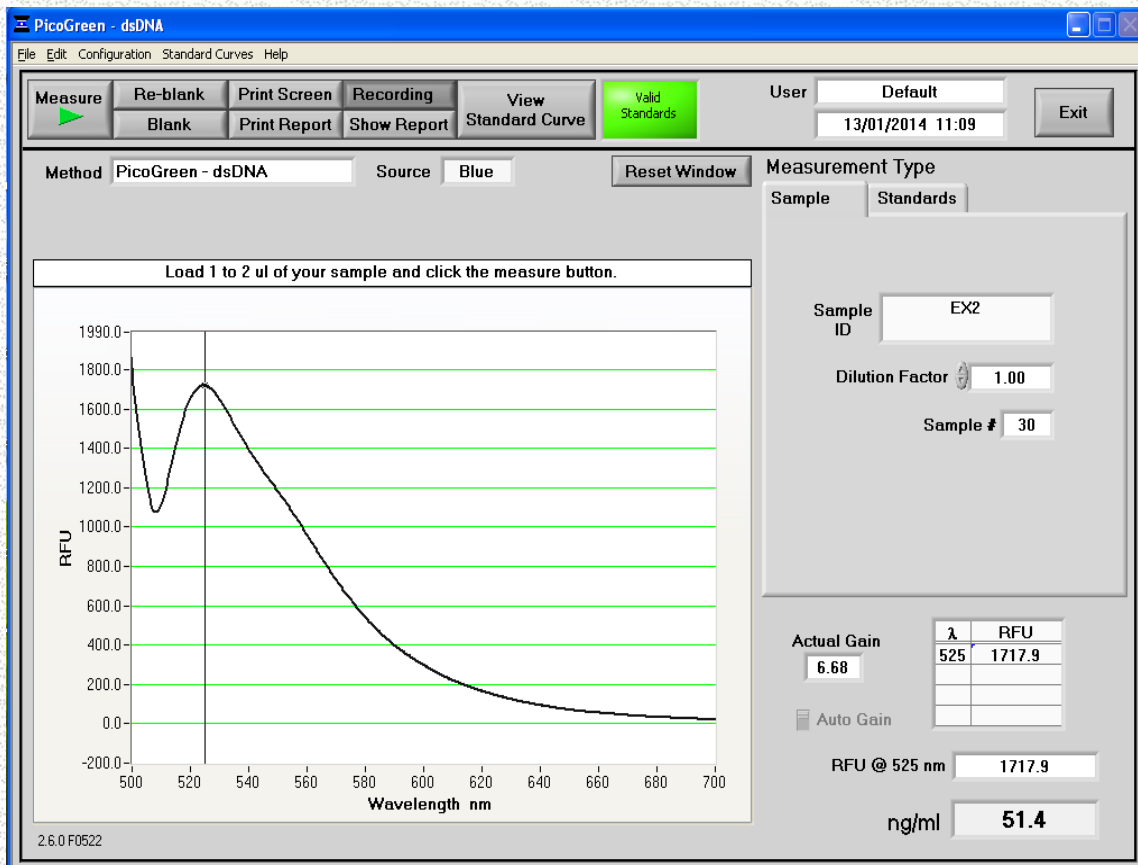
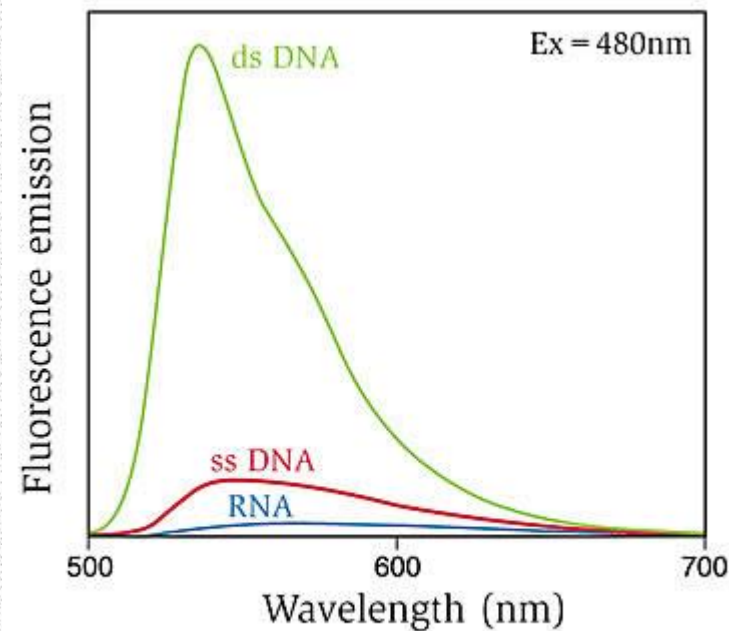


Figure 2 - dsDNA specificity



Fluorescence enhancement of the PicoGreen® Quantitation Reagent upon binding dsDNA, ssDNA and RNA.

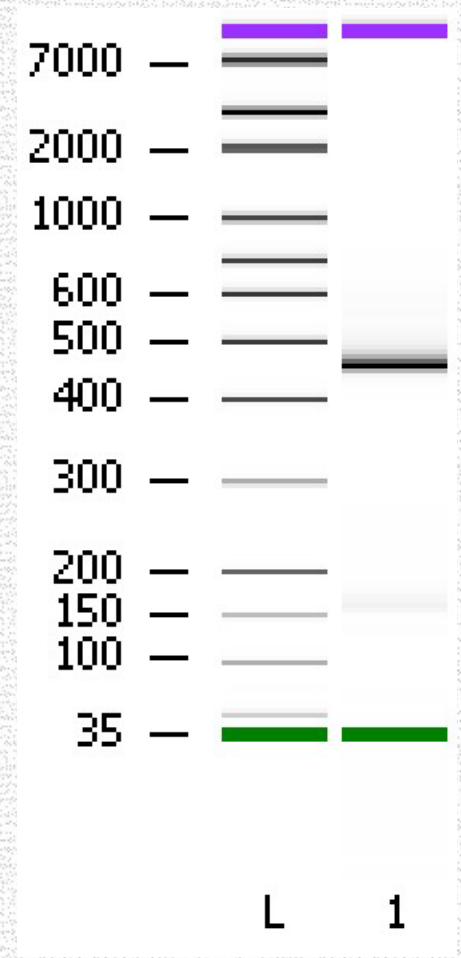
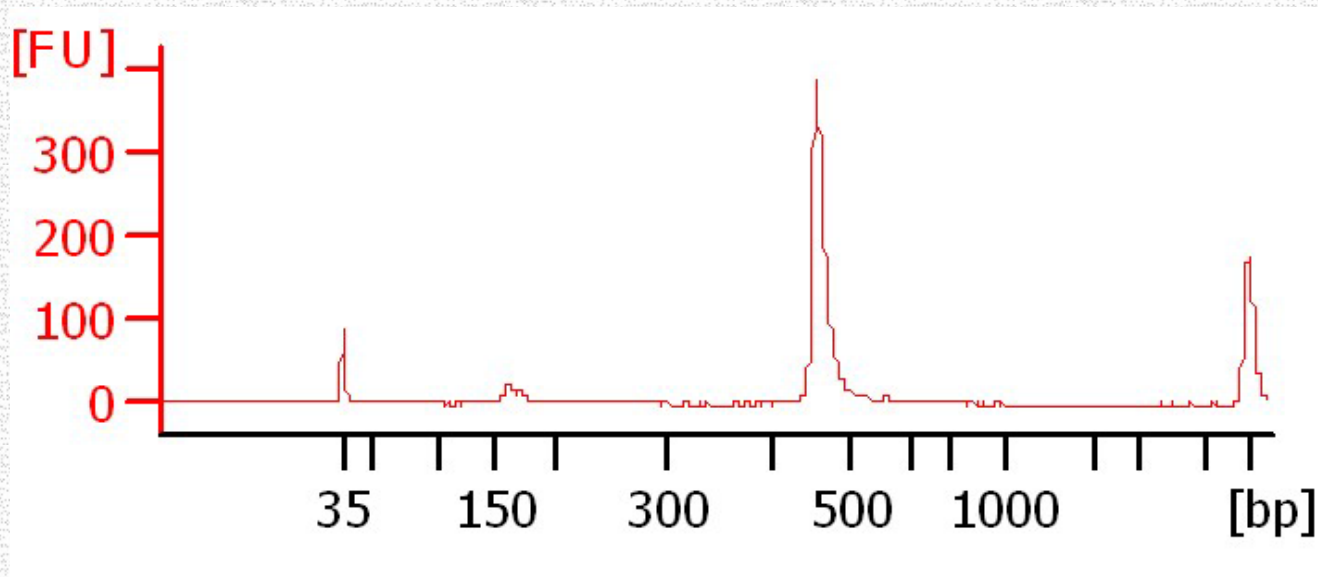
# Priprava mešanice PCR produktov

- Izračunate koncentracijo PCR produktov v molekul/ $\mu$ l.

$$\frac{\text{Molekul}}{\mu\text{l}} = \frac{(\text{koncentracija vzorca } \left[\frac{\text{ng}}{\mu\text{l}}\right] * 6,022 * 10^{23})}{(656,6 * 10^9 * \text{dolžina amplikona } [\text{bp}])}$$

- Pripravite mešanico vseh PCR produktov, ki jih boste hkrati sekvencirali :  $1 \times 10^9$  molekul/ $\mu$ l v TE pufri.
- Preverite čistost/dolžino produktov na s High Sensitivity DNA chip na Agilent Bioanalizator 2001.

# Kvaliteta PCR produkto



Region table for sample 1 : sample 1

From [bp]	To [bp]	Corr. Area	% of Total	Average Size [bp]	Size distribution in CV [%]	Conc. [pg/μl]	Molarity [pmol/l]
200	1,000	557.9	93	464	2.4	414.74	1,354.4

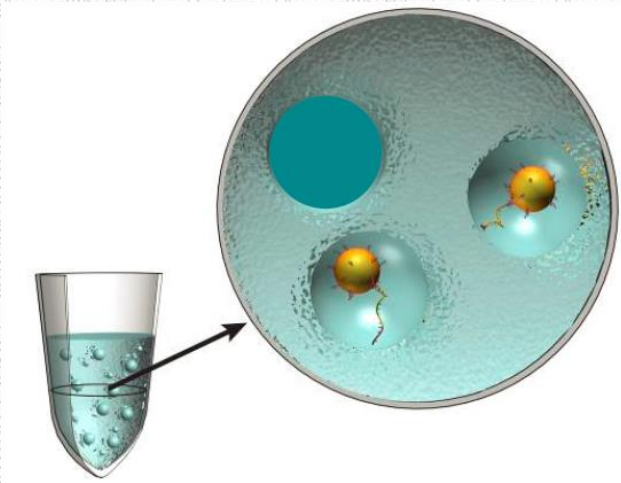


# PCR na kroglicah v emulziji voda-olje

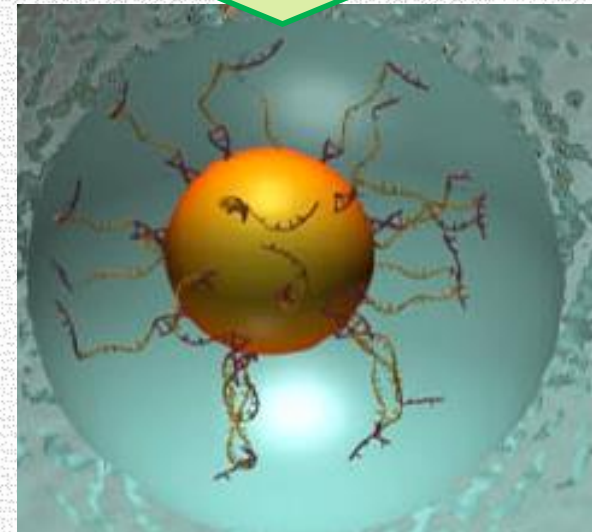
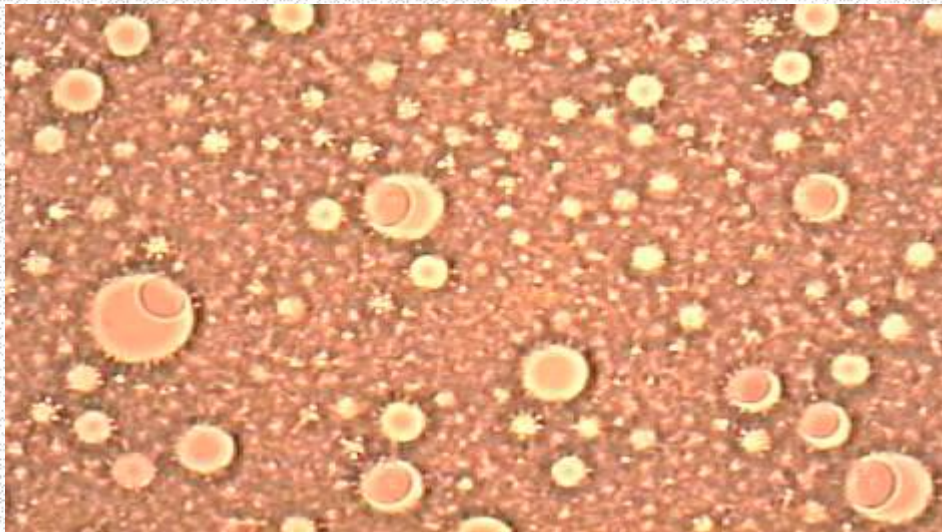
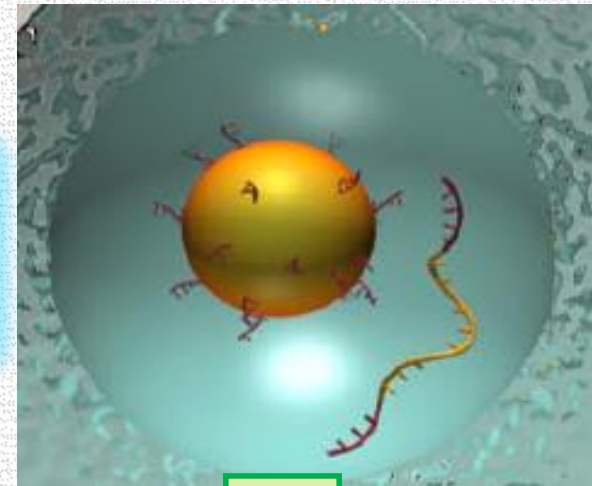
- 1) Pripravimo dve encimski mešanici (A in B), dve mešanici kroglic (A in B) in dve redčitvi knjižnice PCR produktov.
- 2) PCR produkte redčite in dodajte toliko v mešanico, da boste dobili razmerje dve molekuli na kroglico.
- 3) Vse skupaj zmešate in napravite emulzijo s posebno napravo.



# Mikroreaktorji v emulziji ( $10^6/\text{ml}$ )

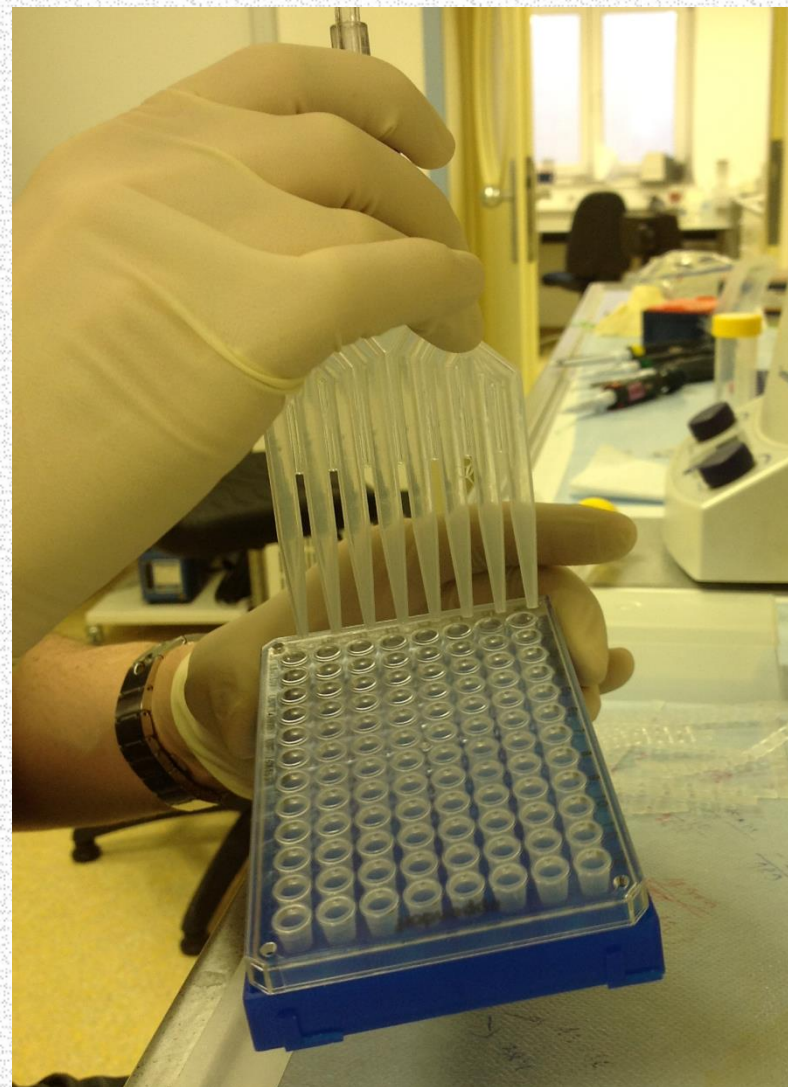
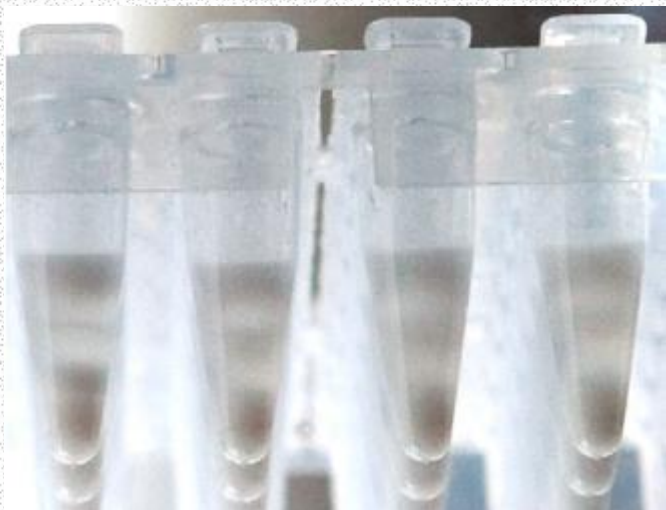
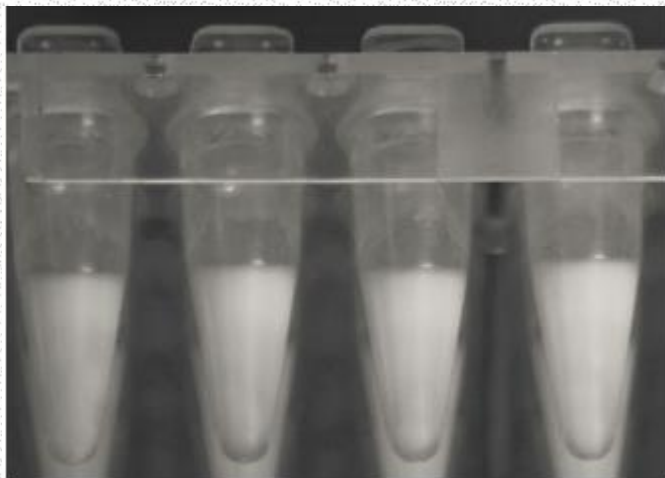


ena kroglica  
=  
en DNA fragment



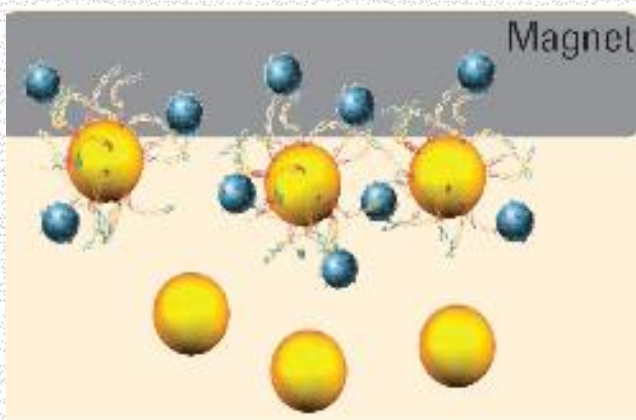


# Izolacija kroglic s pomnoženimi fragmenti





# Izolacija kroglic s pomnoženimi fragmenti



Odstranitev ostankov PCR reagentov



Obogatitev kroglic z DNA fragmenti

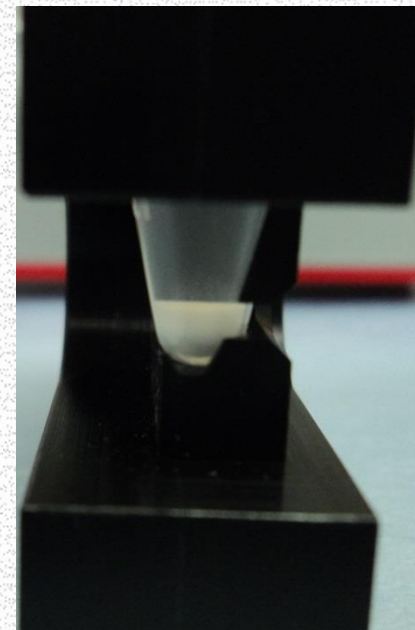
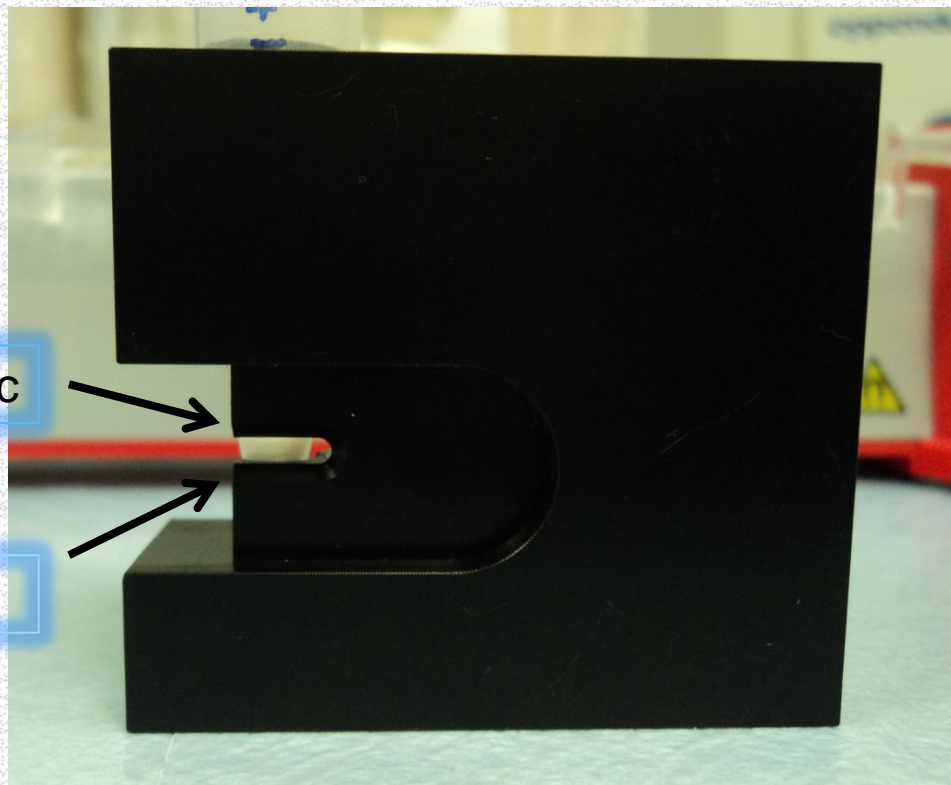


Prileganje oligonukleotidov za sekvenciranje



Kvantifikacija kroglic

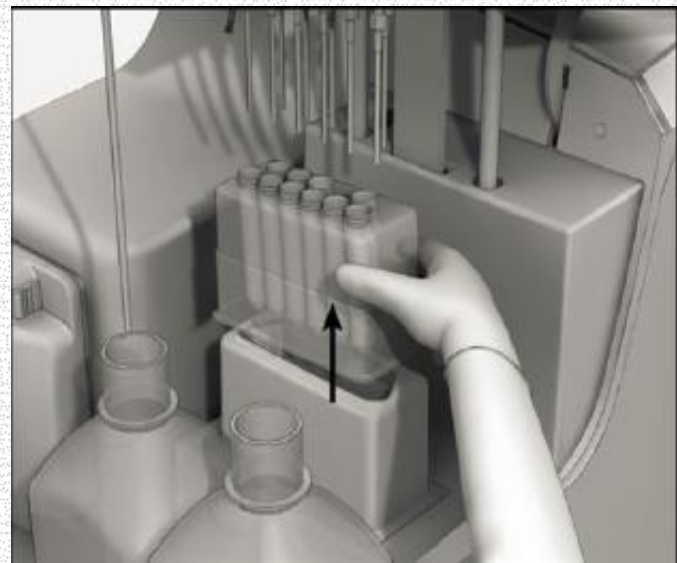
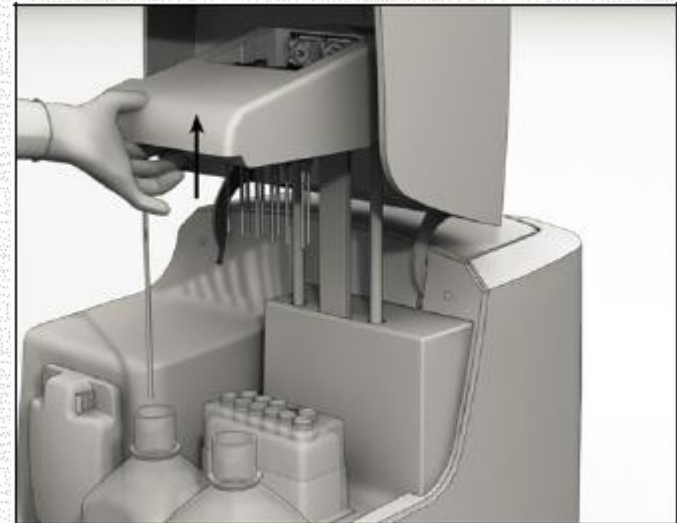
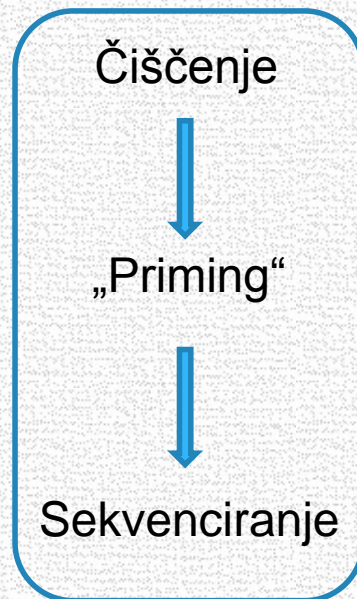
# Kvantifikacija kroglic



Za sekvenciranje potrebujete 500.000 kroglic.

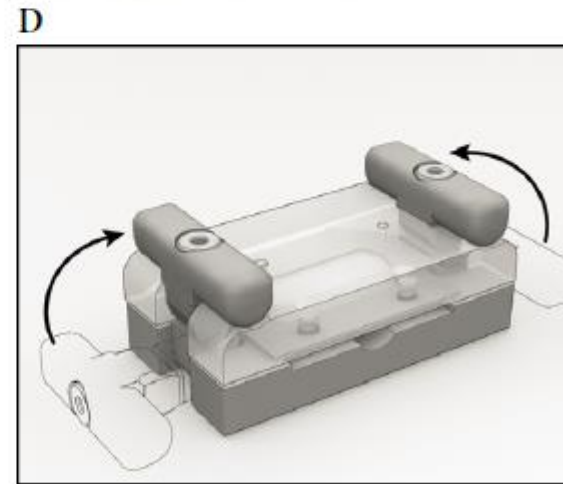
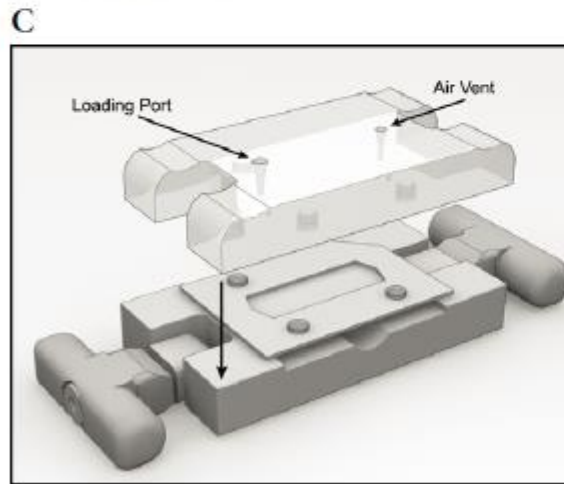
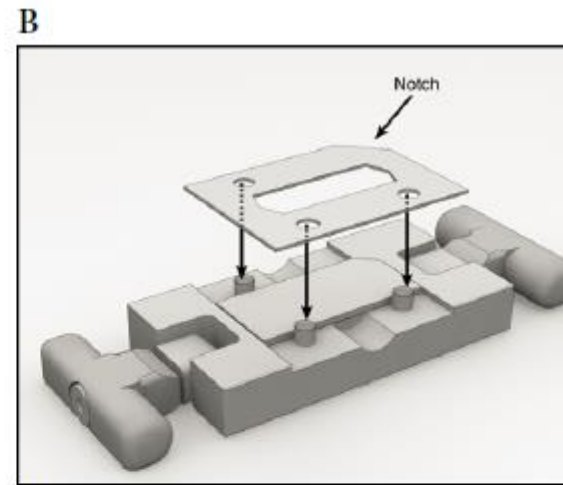
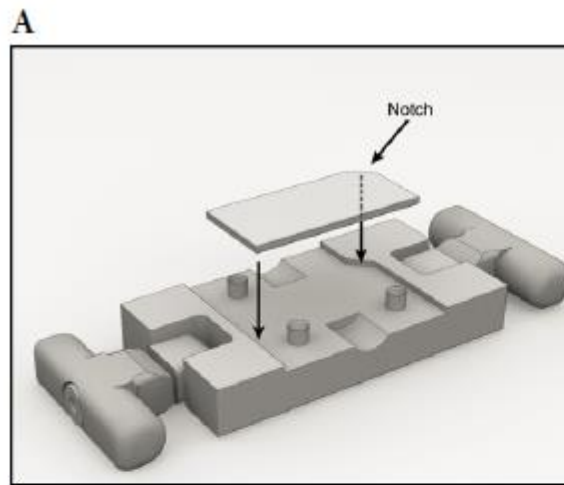
- ✓ Če jih je več (do 2 milijona oziroma do zgornje meje), se preostanek kroglic spravi.
- ✓ Če je manj kot 500.000, jih lahko spravite.
- ✓ Če jih je več kot 2 milijona, zavržite!

# Priprava GS Junior naprave





# Sestava naprave za nanos kroglic



# Priprava pikotitrne ploščice

**Plast kroglic**

**Vrsta kroglic**

Plast 1 kroglice z encimi (pre)

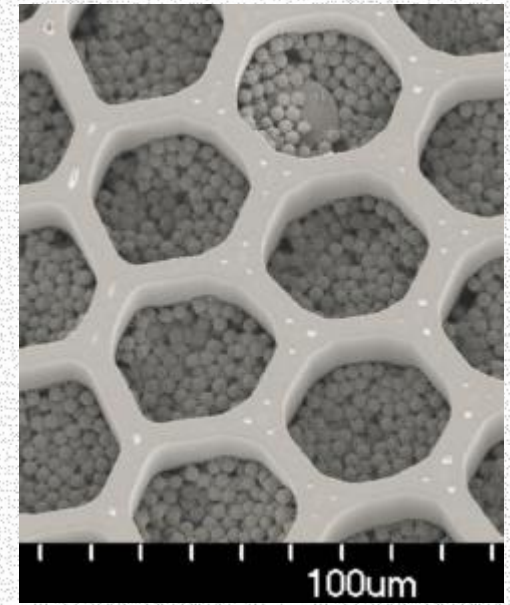
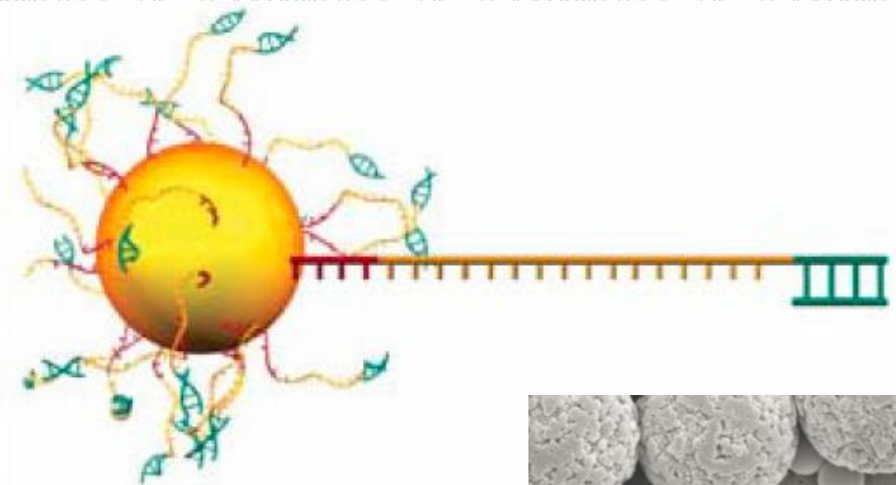
Plast 2 pakirne kroglice in kroglice z DNA

Plast 3 kroglice z encimi (post)

Plast 4 Kroglice z ppiazo

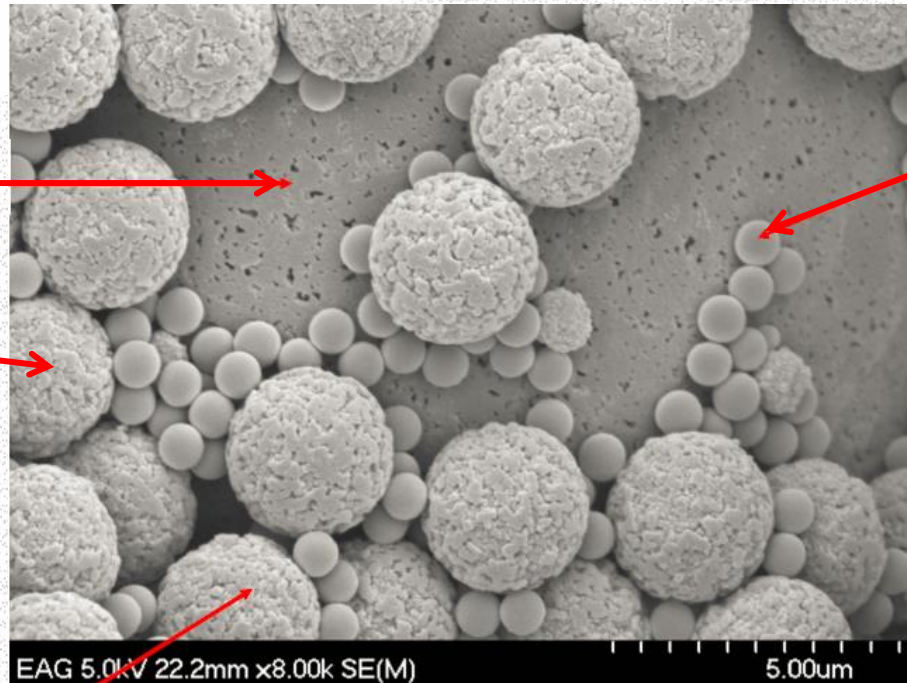


# Pikotitrna ploščica



Kroglica z DNA  
(20 µm)

Kroglica z encimi  
(2,8 µm)



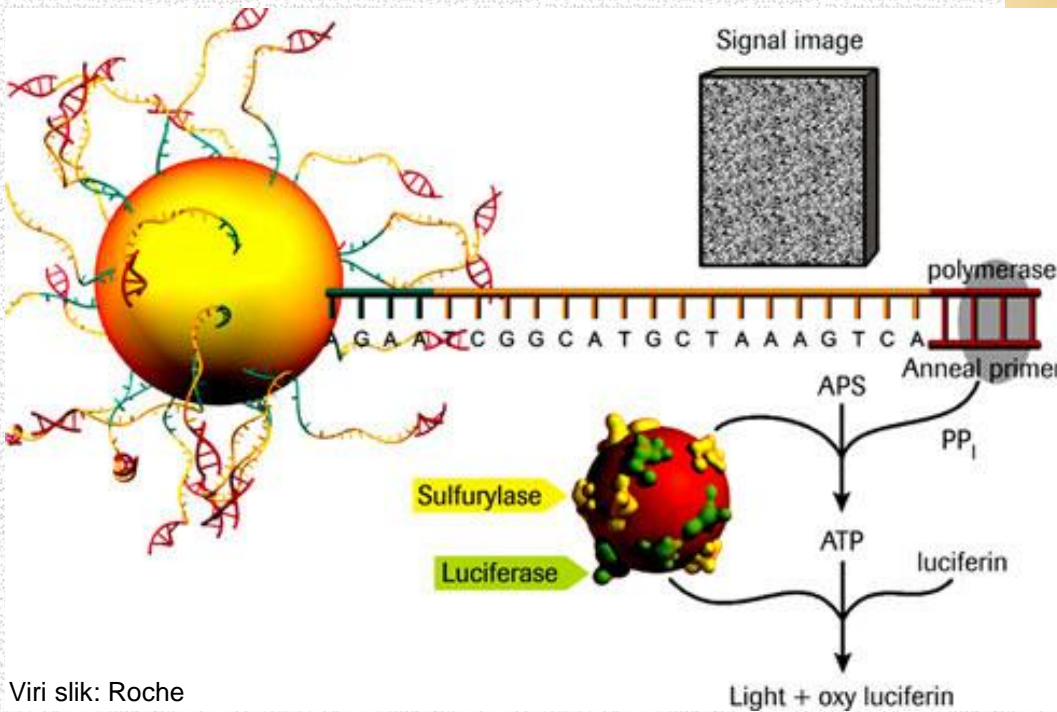
Pakirna kroglica  
(0,8 µm)



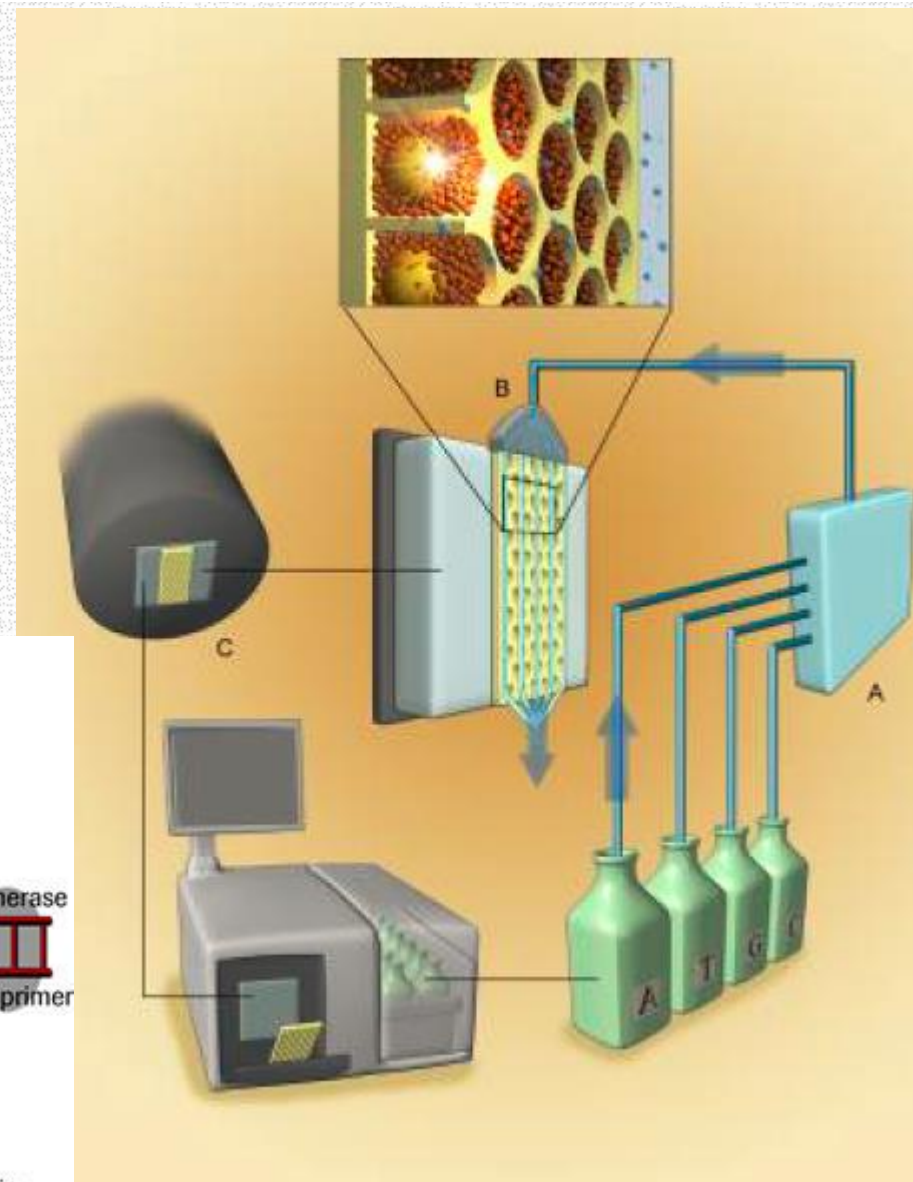
# Vrata kamere, kjer je pikotitrska ploščica



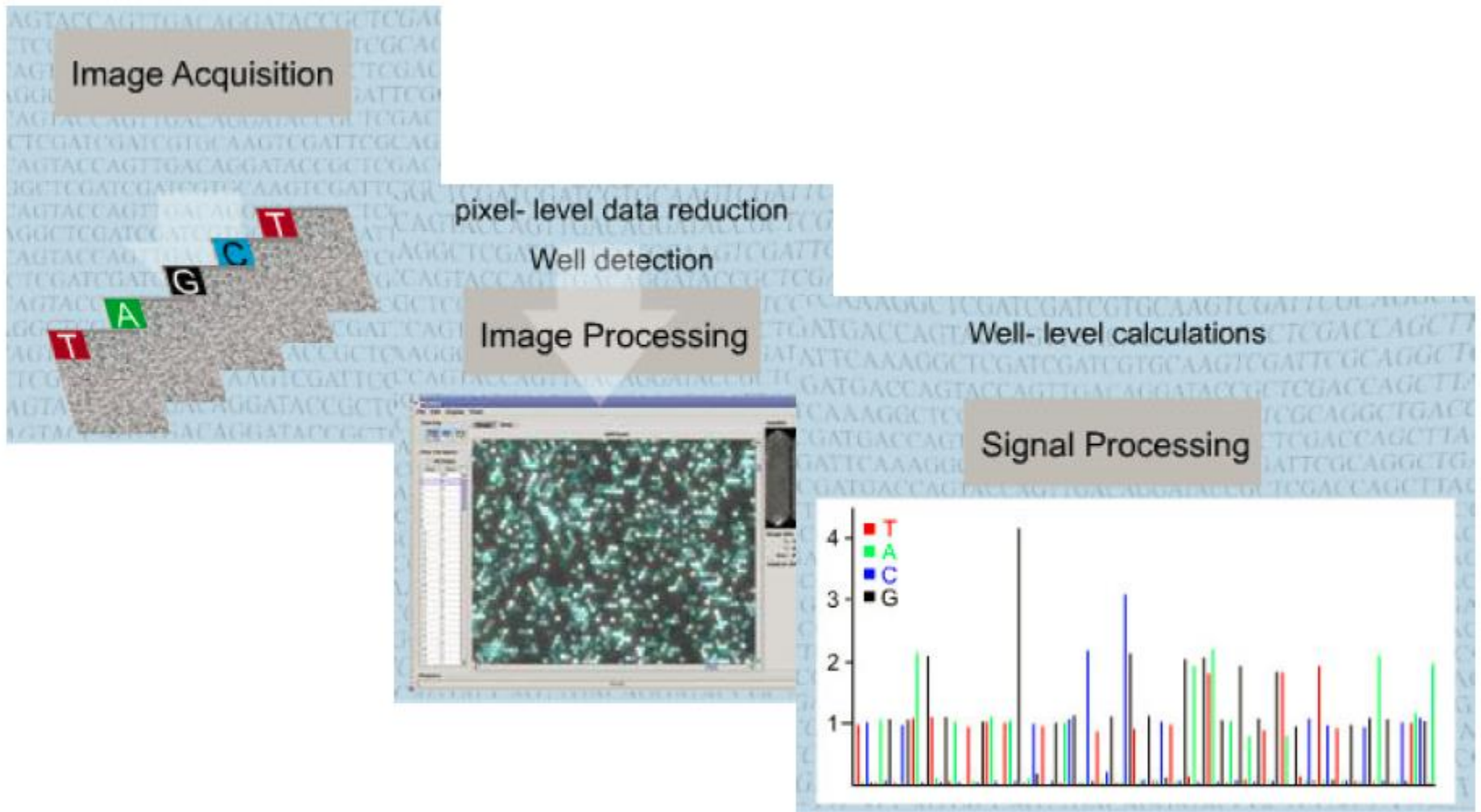
# Sekvenciranje s sintezo



Viri slik: Roche

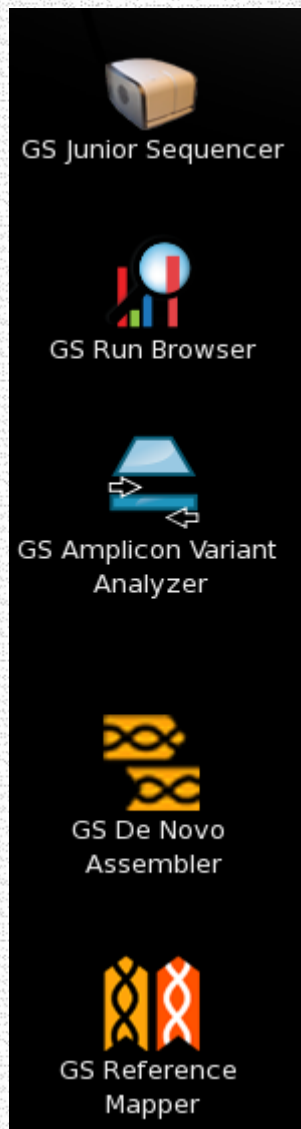


# Analiza podatkov



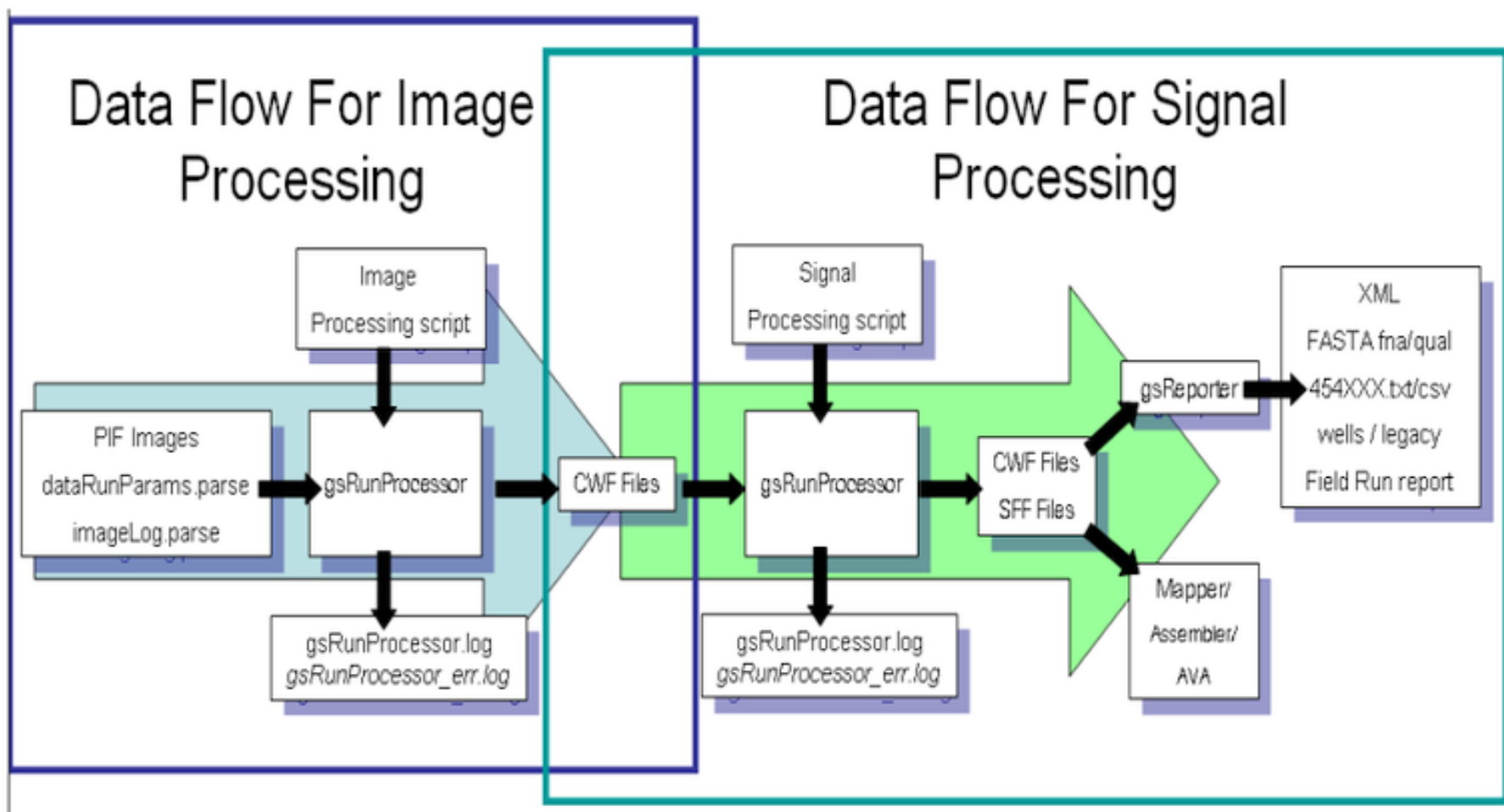


# Programi



Program	Funkcija
<b>GS Junior Sequencer</b>	Upravljanje z napravo, pridobivanje slik
<b>GS Run Processor</b>	Procesiranje slik in procesiranje signala (.cwf in .sff)
<b>GS Run Browser</b>	Vizualizacija rezultatov, upravljanje programov
<b>GS Reporter</b>	Izvoz rezultatov v drugih formatih
<b>GS De Novo Assembler</b>	Sestavi nove genome.
<b>GS Reference Mapper</b>	Vzporedi zaporedja z referenčnim genomom in poišče SNP, insercije, delecije in strukturne spremembe
<b>GS Amplicon Variant Analyzer (AVA)</b>	Vzporedi PCR produkte na izbrano referenčno zaporedje, detektira in kvantificira znane in nove različice

# Analiza podatkov



# Procesiranje signala

## GS Run Processor

- Procesira slike: normalizacija in subtrakcija ozadja, detekcija aktivnih jamic in ustvari .cwf datoteke (composite wells format).
- Procesira signal in naredi .sff (standard flowgram format).
- SFF datoteke potrebujejo nadaljnji programi.
- CWF spremeni GS Reporter v druge formate (FASTA,...).

## Procesiranje signala:

1. Ostrani se signal iz jamic, kjer je zaporedje prekratko (short), sta dva ali več fragmentov (mixed), ni signala (empty).
2. Odreže se dele zaporedja, kjer je kvaliteta preslaba in odreže se dodana zaporedja adapterjev, MID,..
3. Normalizira se za učinkovitost inkorporacije nukleotidov, motnje signala med jamicami, napake pri dovajanju reagentov za sintezo, popravek ozadja,...).
4. Prebere se zaporedje.



GS Run Processor data set: D\_2013\_03\_27\_17\_05\_14\_JR08110684\_fullProcessingAmplicons

Overview Wells Signals Reads Control DNA Filters Adaptor Match

Well Categories

Status

- No Key
- Passed Filter (Library)
- Failed (Library)
- Passed Filter (Control DNA)
- Failed (Control DNA)

Flows

Show All

1	SUB
2	SUB
3	SUB
4	SUB
5	apyrase00
6	apyrase08
7	apyrase10
8	apyrase12
9	SUB
10	apyrase00
11	apyrase08
12	apyrase10
13	apyrase12
14	SUB
15	apyrase00
16	apyrase08
17	apyrase10
18	apyrase12
19	SUB
20	apyrase00
21	apyrase08
22	apyrase10
23	apyrase12
24	SUB
25	apyrase00
26	apyrase08
27	apyrase10
28	apyrase12
29	SUB
30	apyrase00
31	apyrase08
32	apyrase10
33	apyrase12
34	SUB
35	PPI
36	SUB
37	SUB
38	T
39	A
40	C

Exit

Open

About

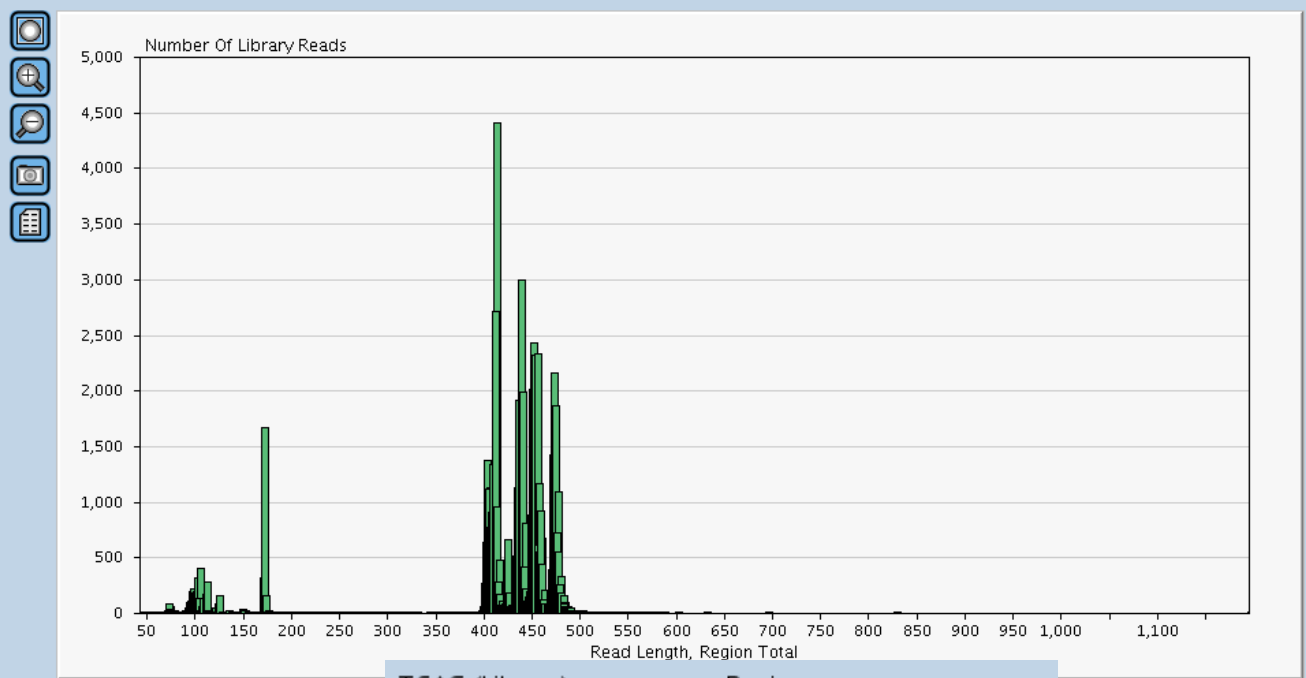
Help

GS Run Browser

GS Run Processor data set: D\_2013\_03\_27\_17\_05\_14\_JR08110684\_fullProcessingAmplicons

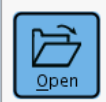
- Overview
- Wells
- Signals
- Reads
- Control DNA
- Filters
- Adaptor Match

- Read Attributes
- Read Length
  - Base Contribution
  - Base Quality
- Well Categories
- TCAG (Library)
  - CATG (Control-Type I)
  - ATGC (Control-Type II)



TCAG (Library)	Region	Total
Raw Wells	236,249	236,249
Key Pass Wells	212,598	212,598
Passed Filter Wells	72,895	72,895
Total Bases	29,921,526	29,921,526
Average Read Length	410.47	410.47
Length Std Deviation	93.83	
Longest Read Length	1,195	1,195
Shortest Read Length	43	43
Median Read Length	437.0	437.0
Modal Read Length	411.0	411.0

>30%  
>70.000







**GS Amplicon Variant Analyzer**

Project Name: Training\_AVA\_project  
 Location: /home/adminrig/Desktop/Training\_AVA\_project

Overview  Project  Computations  **Variants**  Global Align Consensus Align Flowgrams

Alignment Read Type

Consensus

Individual

Show values

Combined

Forward + reverse

All three

Show denominators

Filter values

Min = 0.00 %

Max = 100.00 %

Apply min/max % to

Forward or reverse

Forward and reverse

Available data

Combined also

Variant status

All

Compact table

No Variants To Load

Variants

Reference	Variant	Max	Anja	Ela	Emila	Gregor	Jaroslav	Krystyna	Marjan	Monika	Rok	Tanja
CYP5 1	8903:T/-	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
CYP5 1	11787:G/T	100.00	100.00	0.00	0.00	0.00	-	0.00	0.00	0.00	▼ 61.11	0.00
CYP5 1	23809:C/T	▼ 75.51	0.00	0.00	0.00	0.00	0.00	▼ 75.51	0.00	0.00	0.00	0.00
CYP5 1	27481:G/C	100.00	100.00	0.00	35.96	0.00	56.18	41.56	0.00	28.24	0.00	0.00
CYP5 1	27607:T/C	100.00	100.00	0.00	35.96	0.00	56.18	41.56	0.00	▼ 41.18	0.00	0.00

Variants:

Samples:

Meet filter:

Exit

New

Open

Save

Back

About

Help

# GS Amplicon Variant Analyzer (AVA)

Reference	Variant	Max	Sample_1	Sample_2	Sample_3	Sample_4
Exon-3	229:T/G	46.38	46.38	0.00	0.00	-
Exon-4	98:A/C	0.59	0.14	0.59	-	0.35
Exon-8	91:C/A	49.53	0.00	49.53	0.00	0.00

20% > heterozigot > 80%

← homozigot →

Analiza najdenih različic:

1. Izvozite zaporedje v FASTA formatu.
2. Blast zaporedja v Genome Browser.
3. Preverite ali je različica v podatkovnih zbirkah: spletna stran Ensembl.
4. Kako različica vpliva na aminokislinsko sestavo in delovanje proteina uporabite programa SIFT in PolyPhen.
5. Statistična analiza.

# Razprava

1. Kako potrdimo rezultate (nov SNP) pridobljene s sekvenciranjem nove generacije?
2. Kakšne probleme imamo pri iskanju novih vrst in kako bi jih rešili?



# Literatura

Next-generation sequencing for dummies:

[http://users.ugent.be/~avierstr/nextgen/Next\\_generation\\_sequencing\\_web.pdf](http://users.ugent.be/~avierstr/nextgen/Next_generation_sequencing_web.pdf)

<http://seqanswers.com/> - forum z odgovori

Chandra Shekhar Pareek *et al.*, Sequencing technologies and genome sequencing, 2011, J Appl Genetics, 52, 413-435.

Metzker M.L. *et al.*, Sequencing technologies – the next generation, 2010, 11: 31-46.

<http://www.454.com/>

<http://www.illumina.com/applications/sequencing.ilmn>

<http://www.lifetechnologies.com/us/en/home/life-science/sequencing.html>