



Applicazioni del Sequenziamento di Nuova Generazione (NGS) nel contesto One Health

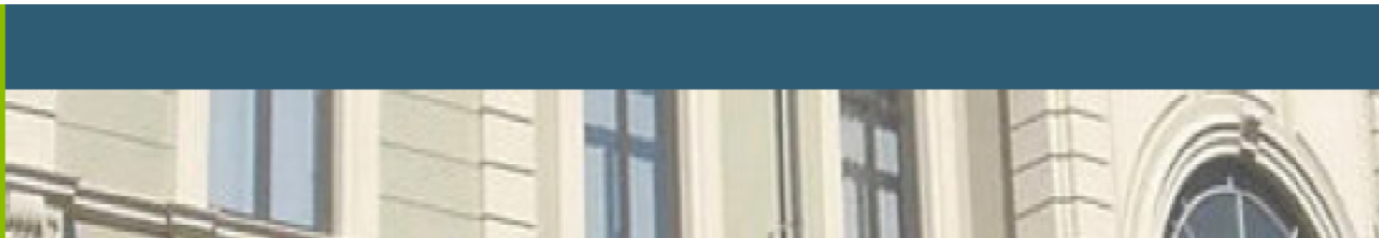
NGS: tutti i segreti del laboratorio

Francesco Cerutti

*S.S. Genetica e Immunobiochimica
Istituto Zooprofilattico Sperimentale del Piemonte,
Liguria e Valle d'Aosta*

Angelo Romano

*S.C. Controllo Alimenti e Igiene delle Produzioni
Istituto Zooprofilattico Sperimentale del Piemonte,
Liguria e Valle d'Aosta*



GLOSSARIO MINIMO DI SOPRAVVIVENZA

Next Generation Sequencing

Metodo di sequenziamento parallelo ad alta produttività, che produce migliaia o milioni di sequenze contemporaneamente.

Deep Sequencing

Tecniche di analisi della sequenza nucleotidica in grado di aumentare: portata, complessità, sensibilità e precisione dei risultati.

Adattatori

Brevi oligonucleotidi, fusi alle estremità del DNA da sequenziare. A seconda della tecnologia utilizzata gli adattatori possono essere accoppiati ad oligonucleotidi complementari su biglie o possono essere legati a flow cell e differiscono tra le piattaforme di sequenziamento.

Library

Insieme di frammenti di DNA con adattatori legati ad ogni estremità.

Bridge PCR

Amplificazione *in situ* di una molecola di DNA specifico su un supporto solido.

Paired-End Sequencing.

Sequenza di entrambe le estremità dello stesso frammento di DNA target e tenere traccia dei dati appaiati.

PCR in emulsione

Amplificazione di library basato sull'utilizzo di biglie. Un singolo frammento viene attaccato alla superficie di una biglia che si trova in un'emulsione contenente i reagenti di amplificazione, che avviene in parallelo in milioni di biglie con milioni di frammenti a singolo filamento.

Read

Dati ottenuti dal sequenziamento, nella forma di una stringa di basi nucleotidiche.

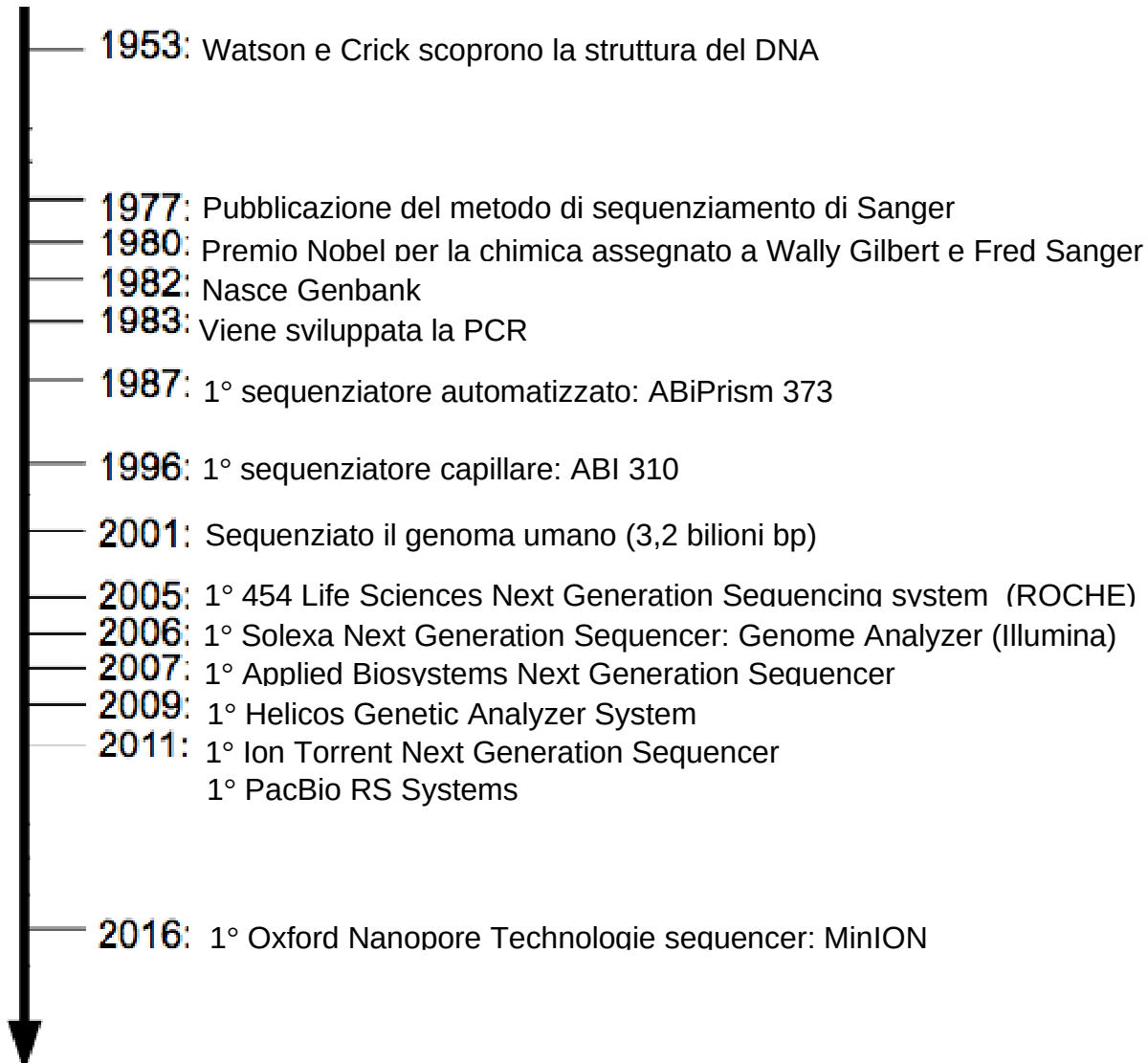
Assembly

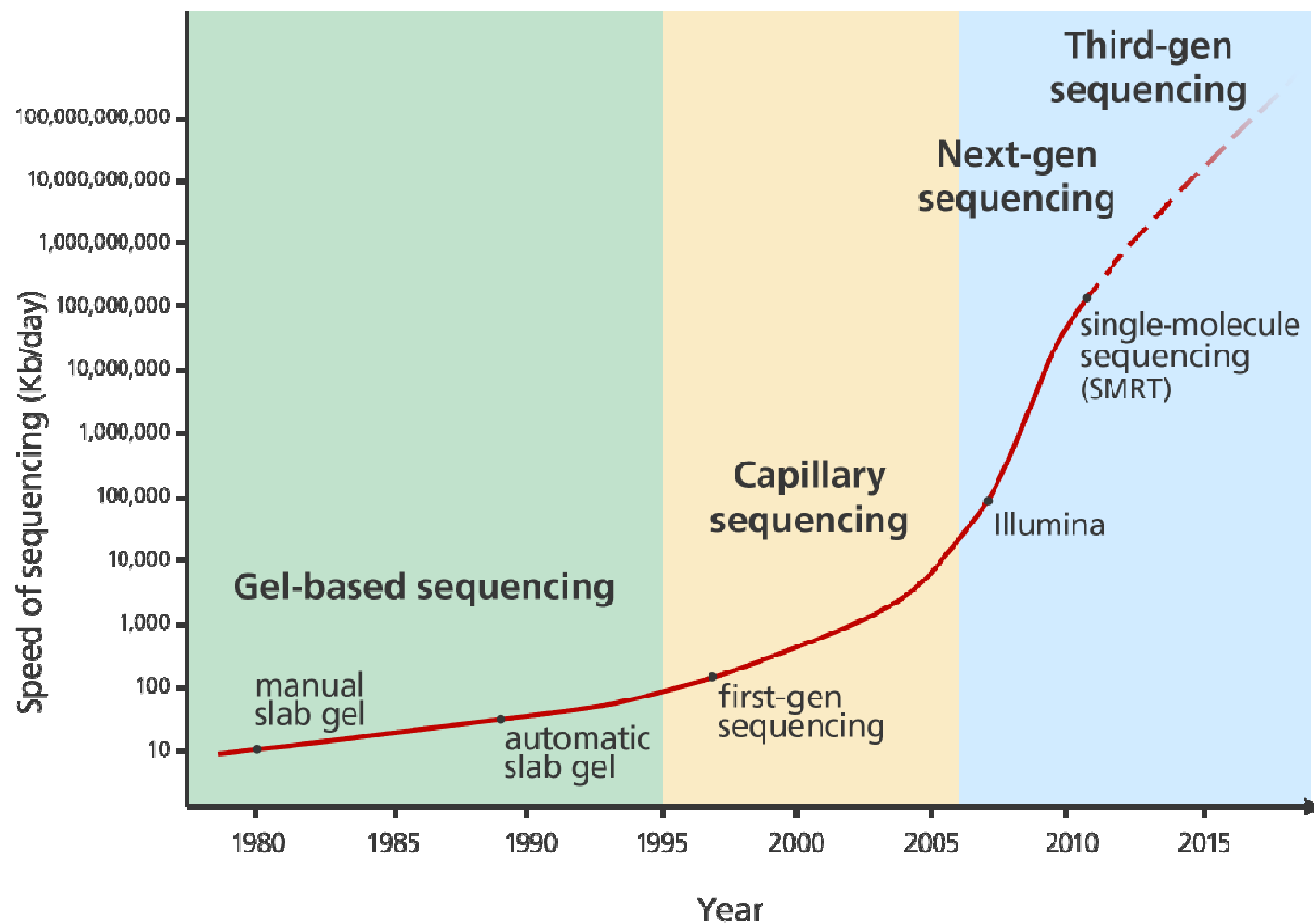
Un processo computazionale per trovare sovrapposizioni di stringhe identiche di lettere tra un insieme di frammenti di sequenza, per formare sequenze più lunghe.

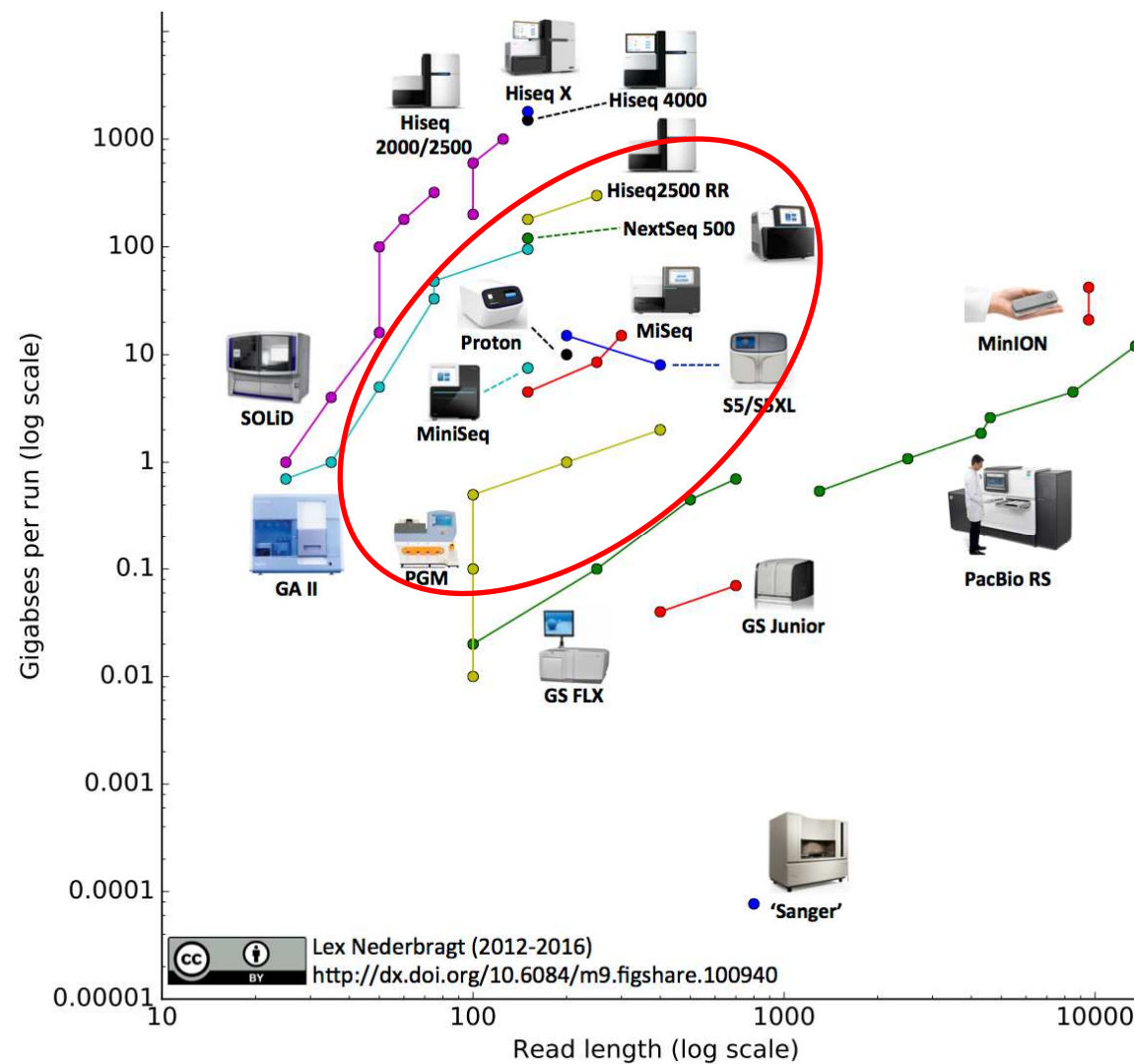
Sequenziamento di terza generazione

Metodo di sequenziamento che funziona leggendo la sequenza nucleotidica a livello di singola molecola.





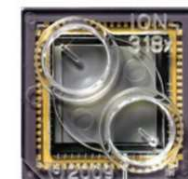






ION PGM

200 o 400 bp read
5.5 milioni di read (differenti chip)
4/7 h di corsa



ION Proton

Fino a 200 bp read
60-80 milioni di read
2/4 h di corsa



Ion S5

200 o 400 bp read
3-80 milioni di read (differenti chip)
2/8 h di corsa



MiniSeq

Fino a 2x150 bp read
14/50 milioni di read
7/24 ore di corsa



MiSeq

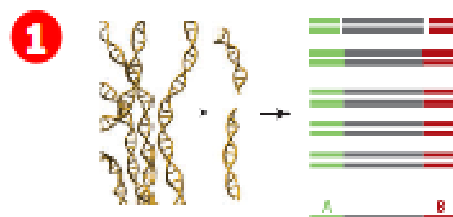
Fino a 2x300 bp read
1/25 milioni di read
5/55 ore di corsa



HiSeq

Fino a 2x 125 bp read
2,5/5 bilioni di read
1/3,5 giorni di corsa

- 1 Library preparation
- 2 Clonal amplification
- 3 Cyclic array sequencing

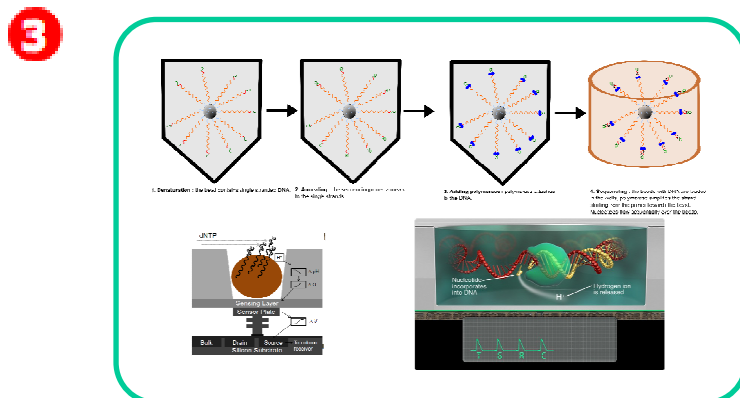


Frammentazione DNA
e preparazione Library

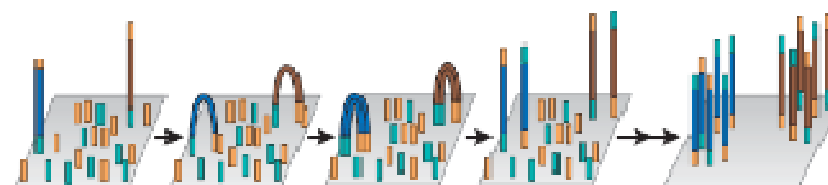
PCR in emulsione



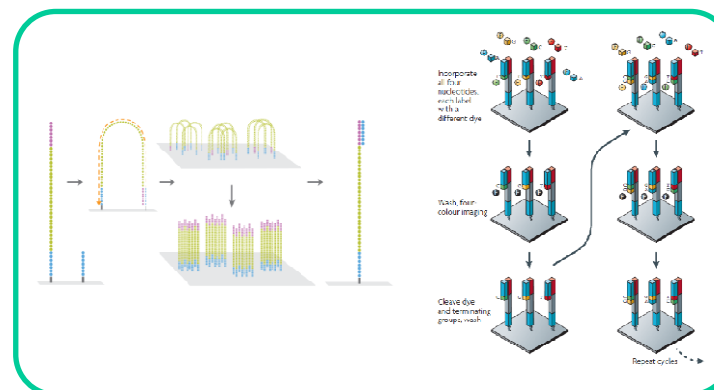
Semiconductor sequencing



Bridge PCR



Sequencing by synthesis





Tecnologia Illumina

Sample Prep

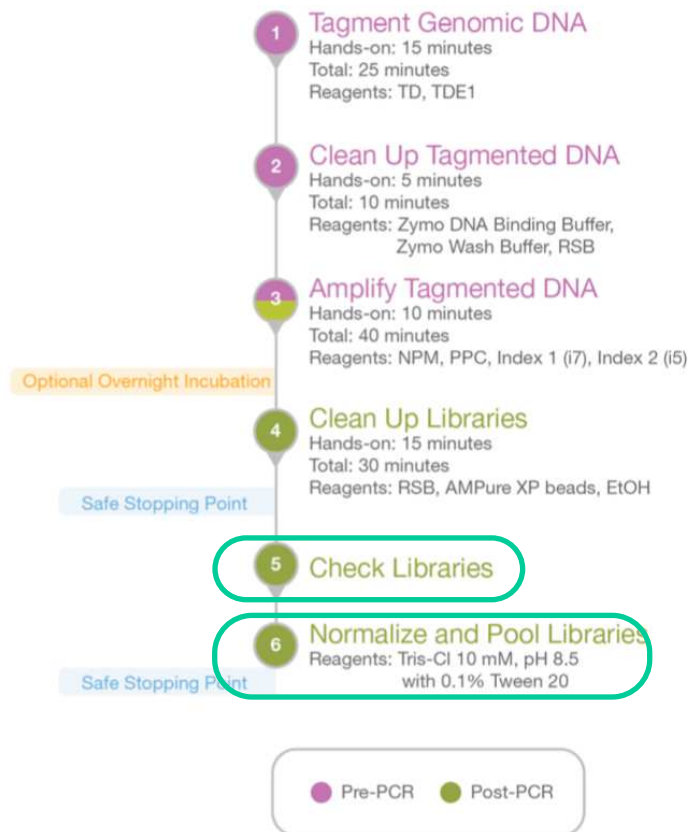
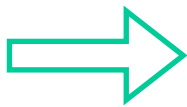
Cluster Generation

Sequencing

Data Analysis

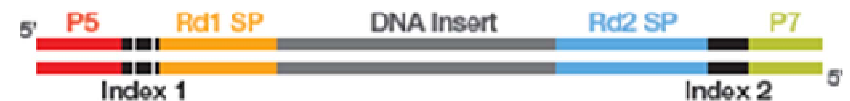


Sample Prep

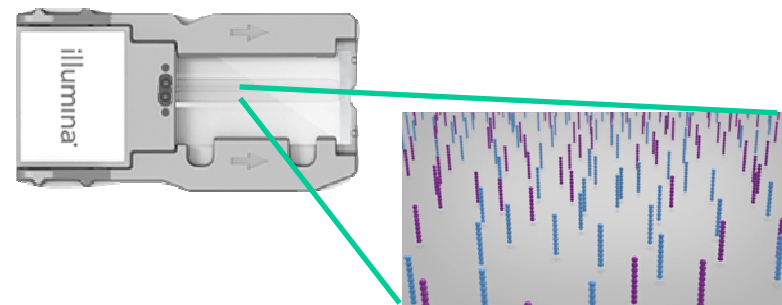
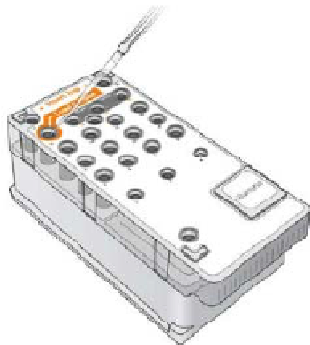


Ha lo scopo di produrre frammenti di dsDNA che contengano ad entrambe le estremità gli

ADATTATORI

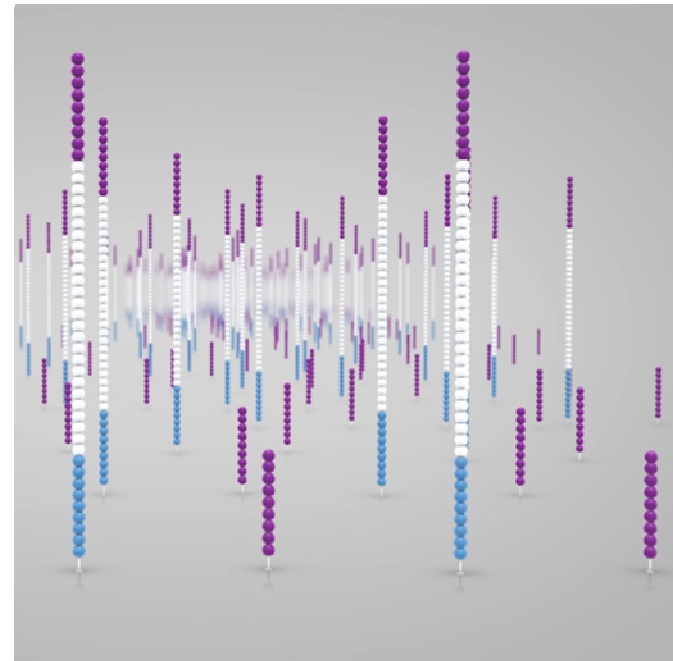


Cluster Generation

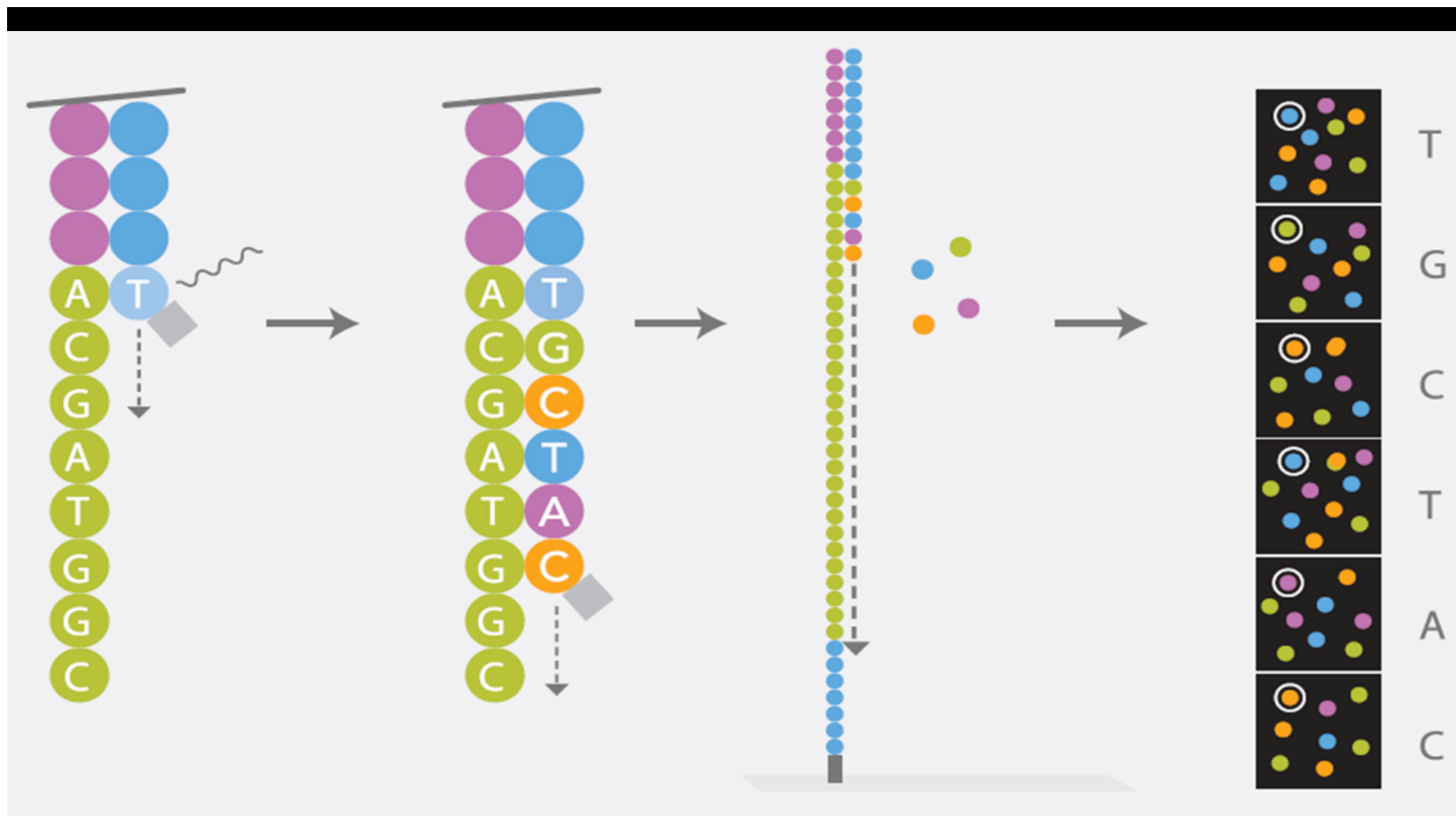


Cluster Generation

- Ibridazione library DNA e estensione
- Denaturazione dsDNA e clivaggio template
- Bridge PCR con formazione di dsDNA bridge
- Denaturazione dsDNA bridge
- Ripetizione bridge PCR
- Denaturazione dsDNA bridge
- Clivaggio della sequenza reverse
- Bloccate estremità 3'
- Ibridazione dei primer di sequenza



Sequencing





Sequencing

illumina





AB3730



800



MiSeq



1



Sequenziamento di terza generazione



Sequenziamento di terza generazione

Limiti delle precedenti “generazioni”

Metodo Sanger

- ⌘ Sequenziamento di un solo frammento
- ⌘ Dipendente da amplificazione (ed eventuali errori correlati)
- ⌘ Elevato costo per base

Metodi di seconda generazione

- ⌘ Altissimo numero di *read*, ma di poche centinaia di bp
- ⌘ Parzialmente dipendente da amplificazione
- ⌘ Poca risoluzione nelle sequenze ripetute
- ⌘ Difficoltà ad assemblare *de novo* i genomi completi

Sequenziamento di terza generazione

Cos'è il sequenziamento di terza generazione

Single - molecule Sequencing

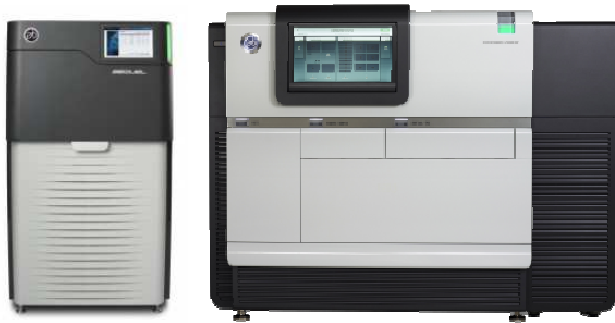
*Sequenze di >10,000 bp
Alcune anche di ~100,000 bp*

*Sequenziamento di genomi virali (2-
200 Kbp) e batterici (0.5-10 Mbp)*

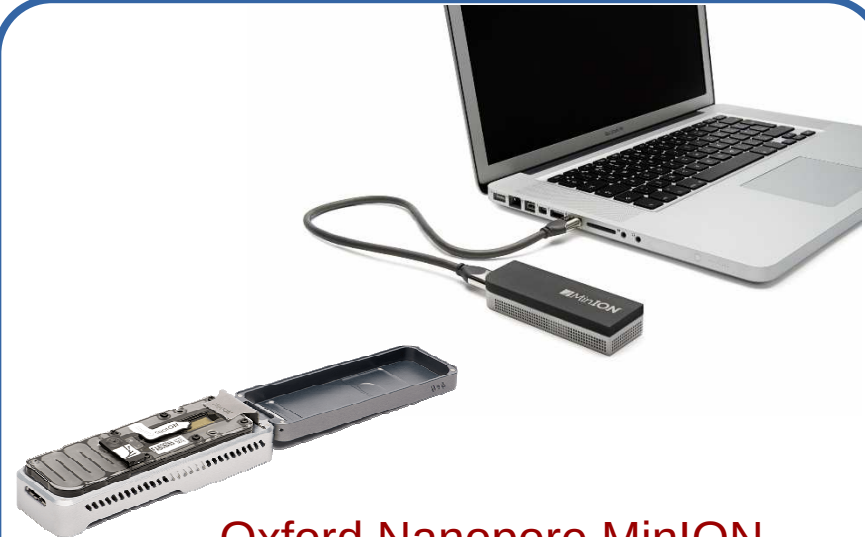
DNA di ottima qualità

Sequenziamento di terza generazione

Strumenti in commercio



Pacific Bioscience Sequel e RS II



Oxford Nanopore MinION



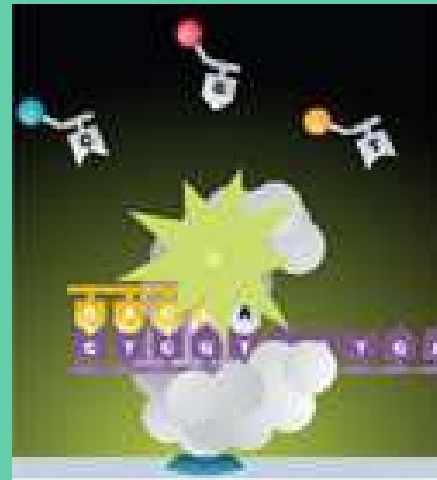
Illumina Moleculo
Protocollo per HiSeq
2000 e HiSeq 2500

Sequenziamento di terza generazione

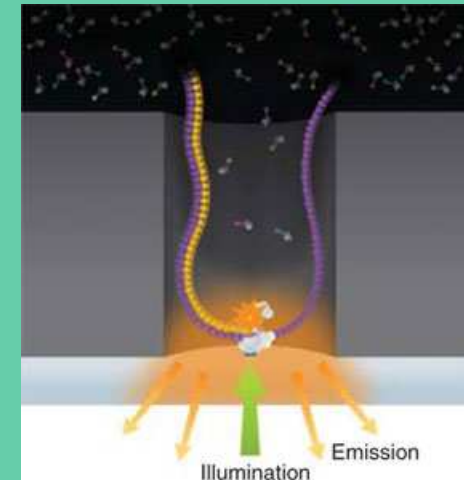
La tecnologia PacBio in breve



DNA Polimerasi
immobilizzata in un
pozzetto (zero-mode
waveguides)



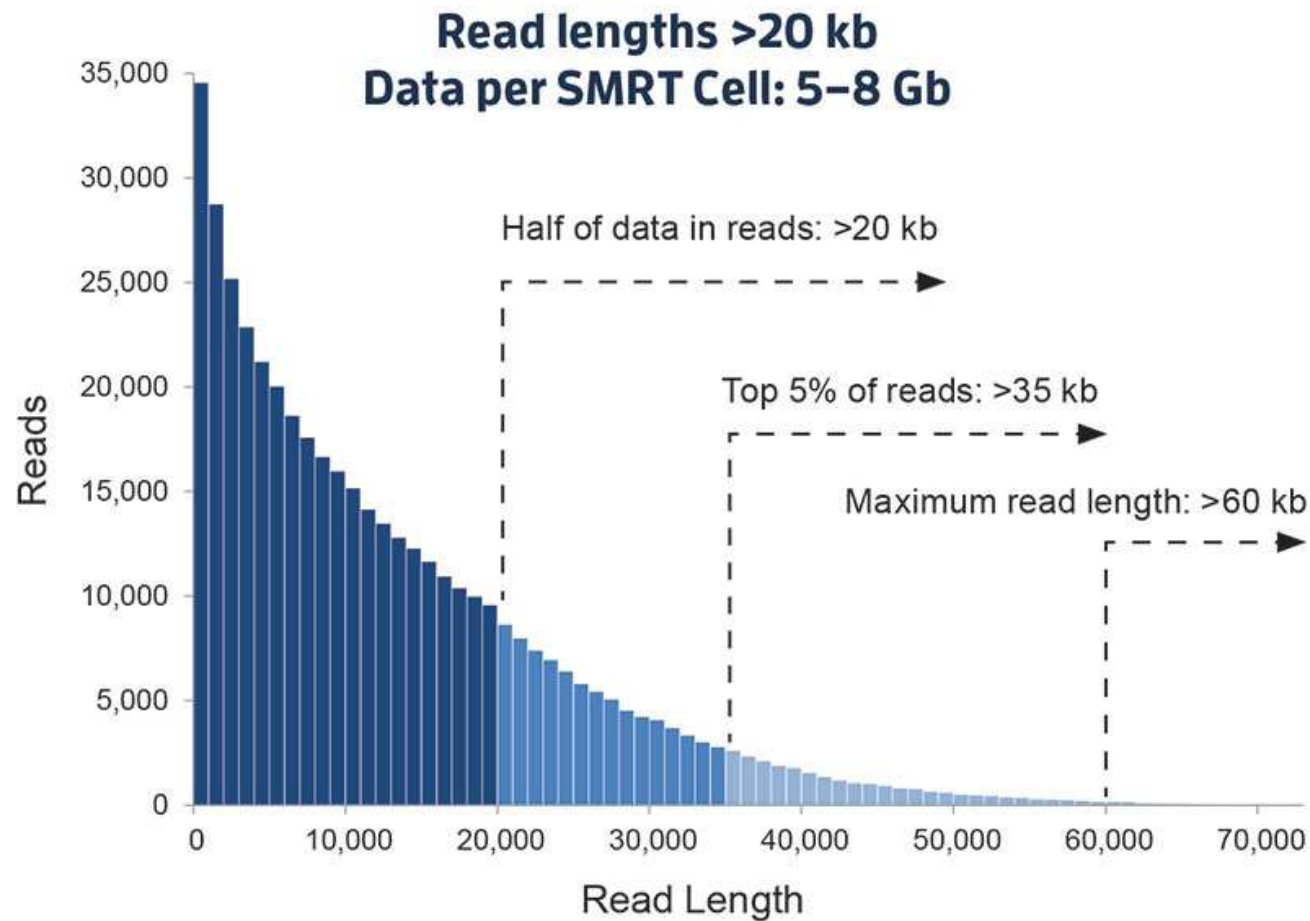
Incorporazione
nucleotidi marcati
durante la sintesi del
filamento



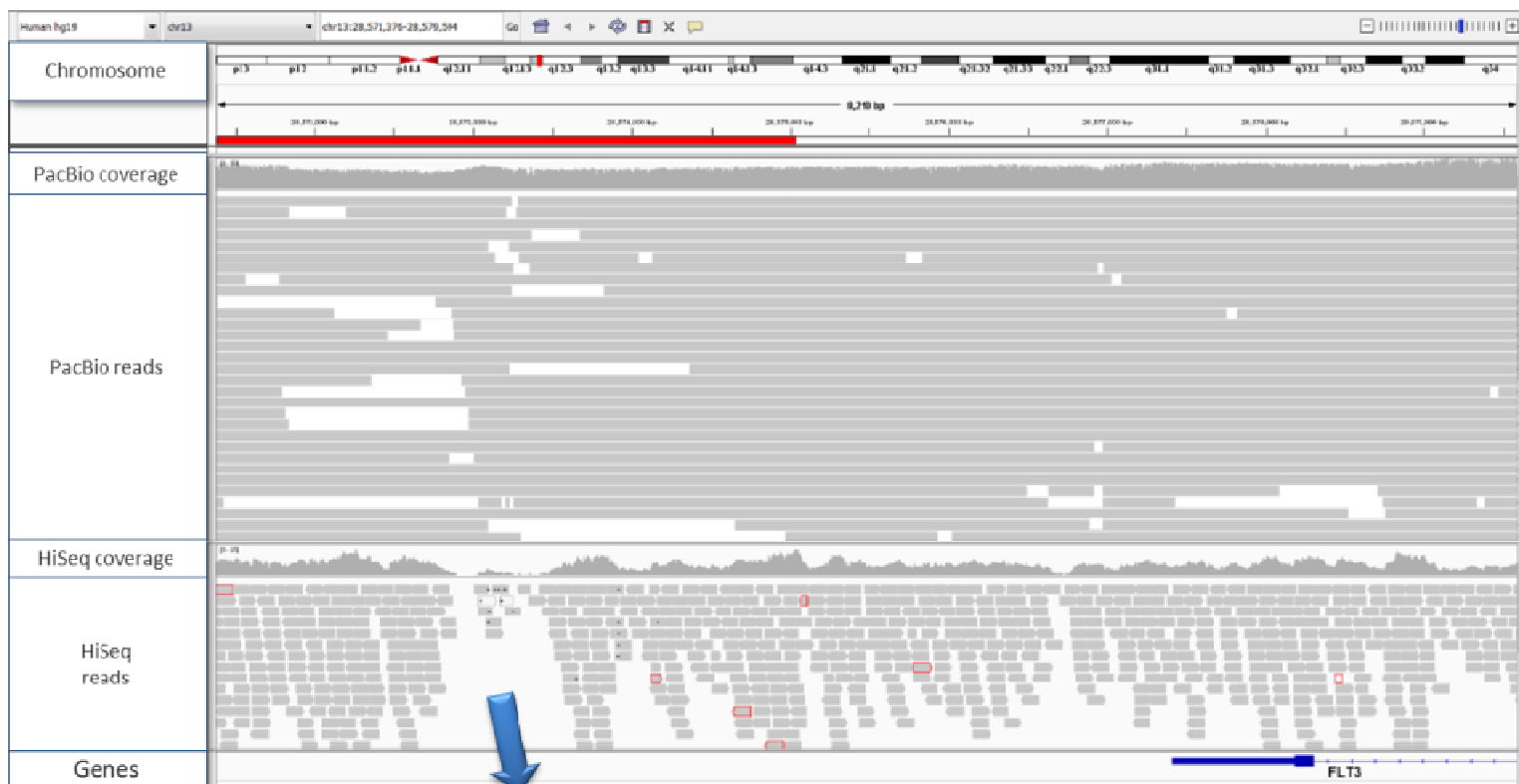
Rilevazione di segnale
luminoso
esclusivamente al
fondo del pozzetto

Single Molecule, Real-Time (SMRT)
Sequencing technology

Sequenziamento di terza generazione



Sequenziamento di terza generazione



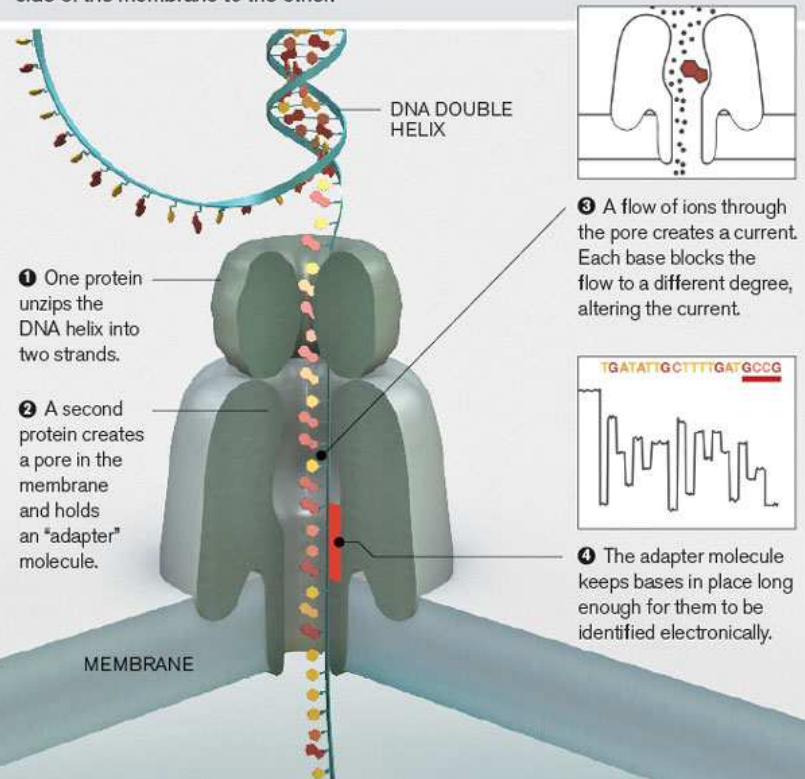
TATATATAAATATA*GTTATATATACATAAATATATGTTATATATTTATATATTTATATACATATAAATATATATTTATATACATATATAAAATATA
TATATATTTATATATAAATATATGTTATATATATTTATATACATATAAATATATATTTATATACATATAAATATATTTATATACATATAAATATATGTTAT
ACATATAAATATATATTTATATACATATAAATATATTTATATACATATAAATATATGTTATATACATATATATATGTTATATACATATAAATATATATGTTAT
TACATATAAATATATATGTTATATACATATAAATATATATAAATTTATATACATATATATTTATATTTATATGTTATATACATATATTTATATTTATATGT
ATATACATATATATATTTATATTTATATGTTATATACATATATATATGTTATATAAATATATATTTATATGTTATATAATATACATACATATATATATTTTT

94% AT-rich

Sequenziamento di terza generazione

La tecnologia Oxford Nanopore in breve

DNA can be sequenced by threading it through a microscopic pore in a membrane. Bases are identified by the way they affect ions flowing through the pore from one side of the membrane to the other.



...rata e la molecola a singolo
...on un rilevatore di corrente.
...che viene rilevato e tradotto in
...elotidica



Sequenziamento di terza generazione

La tecnologia Oxford Nanopore in breve

PacBio Sequel

- & Lunghezza read: 10-20Kb
- & ~ 365,000 read per SMRT cell
- & 85% accuratezza, ma correggibile fino al 99%
- & Utile per assemblaggio di genomi e per identificazione di geni
- & Meno utile per RNA-Seq

Oxford Nanopore MinION

- & Lunghezza read: ~10Kb
- & Output: 5-10Gb
- & 85% accuratezza, ma correggibile fino al 98-99%
- & Utile per scaffold di genomi batterici e identificazione di patogeni
- & Usato durante l'outbreak di Ebola

Sequenziamento di terza generazione

La tecnologia Oxford Nanopore in breve

PacBio Sequel

- ⌘ Lunghezza read: 10-20Kb
- ⌘ ~ 365,000 read per SMRT cell
- ⌘ 85% accuratezza, ma correggibile fino al 99%
- ⌘ Utile per assemblaggio di genomi e per identificazione di geni
- ⌘ Meno utile per RNA-Seq

Oxford Nanopore MinION

- ⌘ Lunghezza read: ~10Kb
- ⌘ Output: 5-10Gb
- ⌘ 85% accuratezza, ma correggibile fino al 98-99%
- ⌘ Utile per scaffold di genomi batterici e identificazione di patogeni
- ⌘ Usato durante l'outbreak di Ebola

- ⌘ Sequenziatore USB che sta in una mano per 1000\$
- ⌘ Richiede un PC/laptop con porta USB 3.0



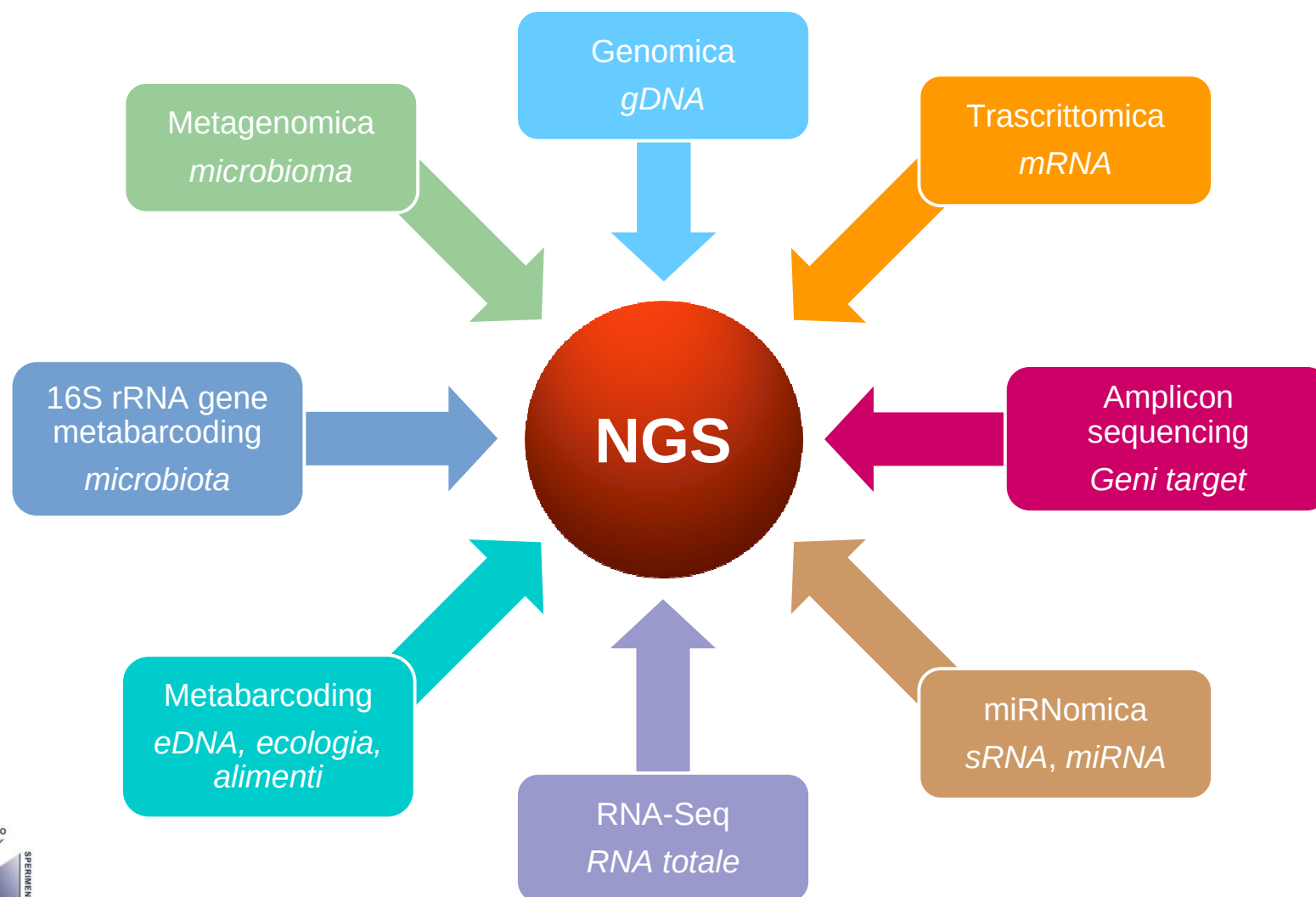
Sequenziamento di terza generazione

Pro

- Read più lunghe
- Copertura delle regioni ripetute
- Minore copertura richiesta

Contro

- Alto tasso di errore
- Costi elevati
- Resa minore in numero di basi





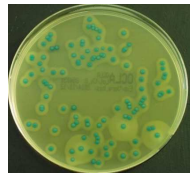
***“Whole Genome Sequencing Is The
Biggest Thing To
Happen To Food Microbiology Since
Pasteur Showed Us How To Culture
Pathogens...”***



*~Dr. Jorgen Schlundt
Exec Director and Founder
The Global Microbial Identifier*



Caratterizzazione isolati batterici



Caratterizzazione «tradizionale»

Utilizzo di differenti metodi per ciascun ceppo:

- Sierotipizzazione
- Ricerca fattori virulenza
- Antibiotico resistenza
- Metodi di tipizzazione biomolecolare (PFGE, spa Type)

Next-generation sequencing:

- «one lab method» non organismo specifico
- possibilità di analisi differenti



Mix di tecniche

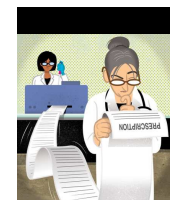
Serotyping, ATB Res, PCR, PFGE, MLST, MLVA, SNPs



Whole Genome Sequencing

- Analisi di sequenza per fini differenti
- Comparabilità a ritroso





Data Analysis



Data Analysis

Risorse online “Web Based”

BaseSpace **SEQUENCE HUB** DASHBOARD PREP RUNS PROJECTS **APPS** PUBLIC DATA

Applications

 16S Metagenomics Illumina, Inc.	 Amplicon DS Illumina, Inc.	 Bacterial Analysis Pipeline GoSeqit ApS	 BWA Aligner BaseSpace Labs
 BWA Enrichment Illumina, Inc.	 Cancer Variant Caller Samsung SDS	 ChIPSeq BaseSpace Labs	 CrossMap BaseSpace Labs
 Cufflinks Assembly & DE Illumina, Inc.	 DeepCheck8-HBV ABL S.A.	 DeepCheck8-HCV ABL S.A.	 DeepCheck8-HIV ABL S.A.
 DESeq2 BaseSpace Labs	 E. coli Serotyping GoSeqit	 Annotator EDGC Annotator GENE DIAGNOSTICS Genome Center	 eGB Elastic Genome Browser Grand Life Sciences
 Enrichment	 FASTA Upload	 FASTQ Toolkit	 FastQC

Ion Reporter

Home **Samples** Analyses Workflows Admin

Dashboard Notifications

Quick links to get started

 **Samples**
Samples represent a collection of data (sequence reads) from one or more sequencing runs.

[Define sample](#) [View samples](#)

 **Workflows**
Workflows are a set of analysis components that have been put together to automate the analysis of your data.

[Create workflow](#) [View workflows](#)

 **Analyses**
Analyses represent workflows that have been executed on a set of samples.

[Launch analysis](#) [View analyses](#)

Data Analysis

Center for Genomic Epidemiology

Home Services Datasets User Home

Overview of Services

Workflows

[Bacterial Analysis Batch Upload Pipeline](#) (Works)

Phenotyping

[ResFinder](#) (Works)
[ShimodoriFinder](#) (Works)
[VirulenceFinder](#) (Works)
[Restriction-ModificationFinder](#) (Works)

Typing

[SeqSero](#) (Works)
[SeqSeroFinder](#) (Works)
[PAST](#) (in development)
[VirusFinder](#) (in development)
[SeqType](#) (Works)
[MLST](#) (Works)
[pMLST](#) (Works)
[PhamFinder](#) (Works)
[KmerFinder](#) (Works)
[SpacedFinder](#) (Works)
[MashTyper](#) (This service is not implemented on the new server)
[TaxonomyFinder](#) (This program is in development)
[TAPR](#) (This service is not implemented on the new server)

Phylogeny

[spTree](#) (Works)
[NDSite](#) (Works)
[CAPI-Muscle](#) (Works)
[TreeViewer](#) (Works)

Other

[Assembler](#) (Works)
[RNAUploader](#) (in development)
[PanFunFrog](#) (Works)
[MashFinder](#) (Works)
[MLFinder](#) (Works)
[SPFinder](#) (Works)
[HostFinder](#) (in development)
[GenomicClassifyer](#) (Not associated with CGE)
[VirusFCM](#) (Not associated with CGE)

Centri di ricerca

Sviluppo tool di analisi genomica dedicati

Risorsa disponibile precompilata

<https://usegalaxy.org>

App ★ Bookmarks Portale Intranet Posta Aziendale SELF Query Reportin... SOIPA MLVA eBURST V3 CGE Server - WORK... AdventNet Manage... Put

Galaxy Analyze Data Workflow Shared Data Visualization Help User

Tools

search tools

[Get Data](#)
[Send Data](#)
[Lift-Over](#)
[Text Manipulation](#)
[Convert Formats](#)
[Filter and Sort](#)
[Join, Subtract and Group](#)
[NGS: QC and manipulation](#)
[NGS: Mapping](#)
[NGS: RNA-seq](#)
[NGS: SAMtools](#)
[NGS: BAM Tools](#)
[NGS: Picard](#)
[NGS: VCF Manipulation](#)
[Extract Features](#)
[Fetch Sequences](#)
[Fetch Alignments](#)
[Get Genomic Scores](#)
[Operate on Genomic Intervals](#)
[Statistics](#)
[Graph/Display Data](#)
[Phenotype Association](#)

Galaxy is an open source, web-based platform for data intensive biomedical research. If you are new to Galaxy [start here](#) or consult our [help resources](#).

Tweets

Galaxy Project @galaxyproject 10h
 In 2 weeks: Advanced RNA-Seq DGE with Galaxy. 25 Nov @VLSI Carlton, VIC, Australia. Still Space & FREE. [bit.ly/1Nb7jdm](#) #usegalaxy [Show Summary](#)

Galaxy Project @galaxyproject 11h
 Ever wondered who writes #usegalaxy? These people do: [github.com/galaxyproject/...](#) [pic.twitter.com/VIPc84T1F6](#) [Show Photo](#)

Galaxy Project @galaxyproject 16h
 Online IUC Contribution Fest - Metagenomics Tools and Workflows [Show Photo](#)

Tweet to @galaxyproject

Server Galaxy

spaTyper 1.0

ResFinder 2.1

[illegible]

spa Typing				
spa Type	Repeats	Contig	Position	Orientation
t2953	11-12-21-17-34-24-34-22-25-25	NODE_66_length_676549_cov_71.170280	325676-325953	plus

MLST 1.8

```

MLST Results
PRONACC-004-Saureus_s15
Sequence Type: ST-581
Organism: Staphylococcus aureus
MLST Profile: saureus

*****
GENE      % IDENTITY  HSP Length  Allele Length  GAPS  BEST MATCH
*****
ARCC      100.00      456         456            0      ARCC-6
AROE      100.00      456         456            0      AROE-57
GLPF      100.00      465         465            0      GLPF-45
GMK       100.00      417         417            0      GMK-18
PTA       100.00      474         474            0      PTA-10
TPI       100.00      402         402            0      TPI-14
YQIL      100.00      516         516            0      YQIL-52

```

CSI Phylogeny 1.4 (Call SNPs & Infer Phylogeny)

PathogenFinder 1.1

Center for Genomic Epidemiology

Home

Services

Instructions

Output

The input organism was predicted as human pathogen

Probability of being a human pathogen (%)

9.91

Input proteome coverage (%)

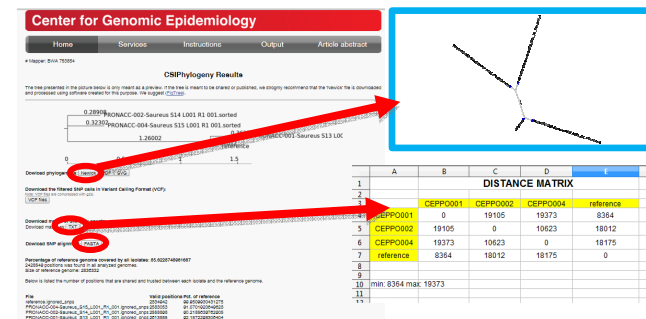
35.05

Matched Pathogenic Families

565

Matched Not Pathogenic Families

0



Data Analysis

Risorse online "Web Based"

*Pipeline analisi
genomiche dedicate*

*Comunità di utenti molto
attiva e sviluppatori*

Analisi bioinformatica

Software commerciali con
GUI

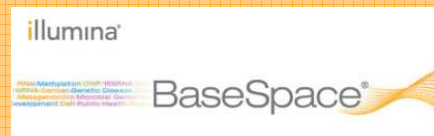


Che
approccio?

Software free sviluppati in
ambiente linux



Piattaforme web (gratuite e a pagamento)

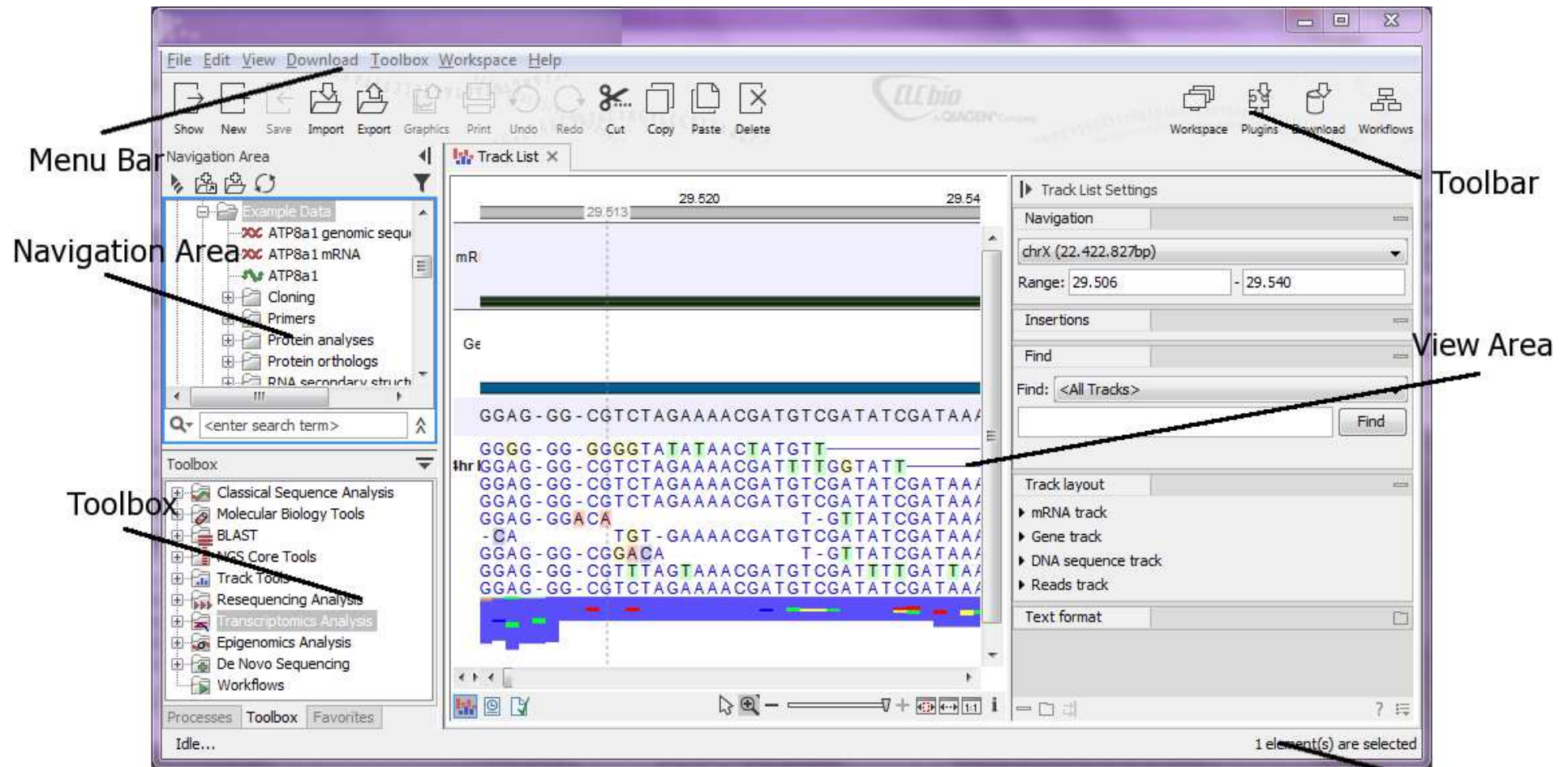


Analisi bioinformatica



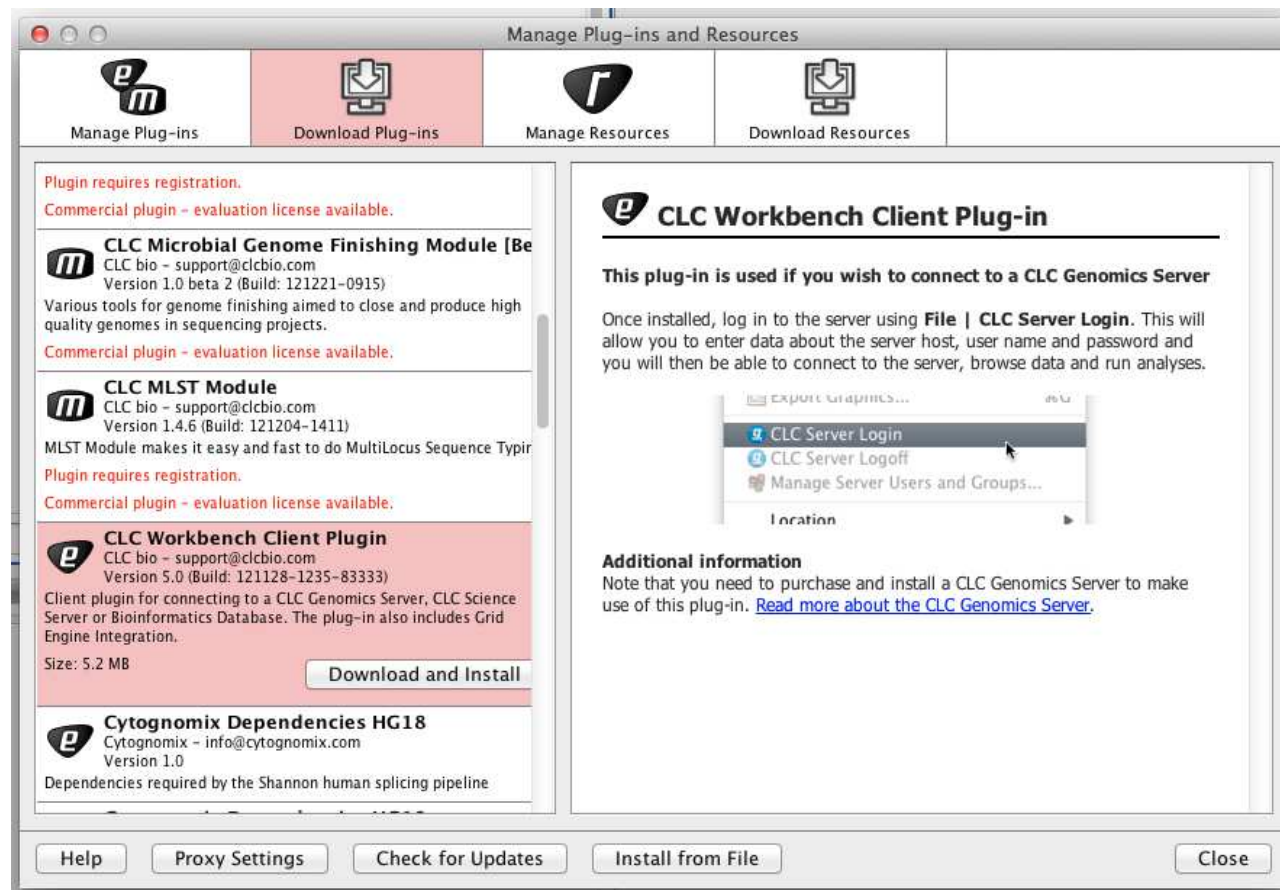
CLC Genomics Workbench

Analisi bioinformatica



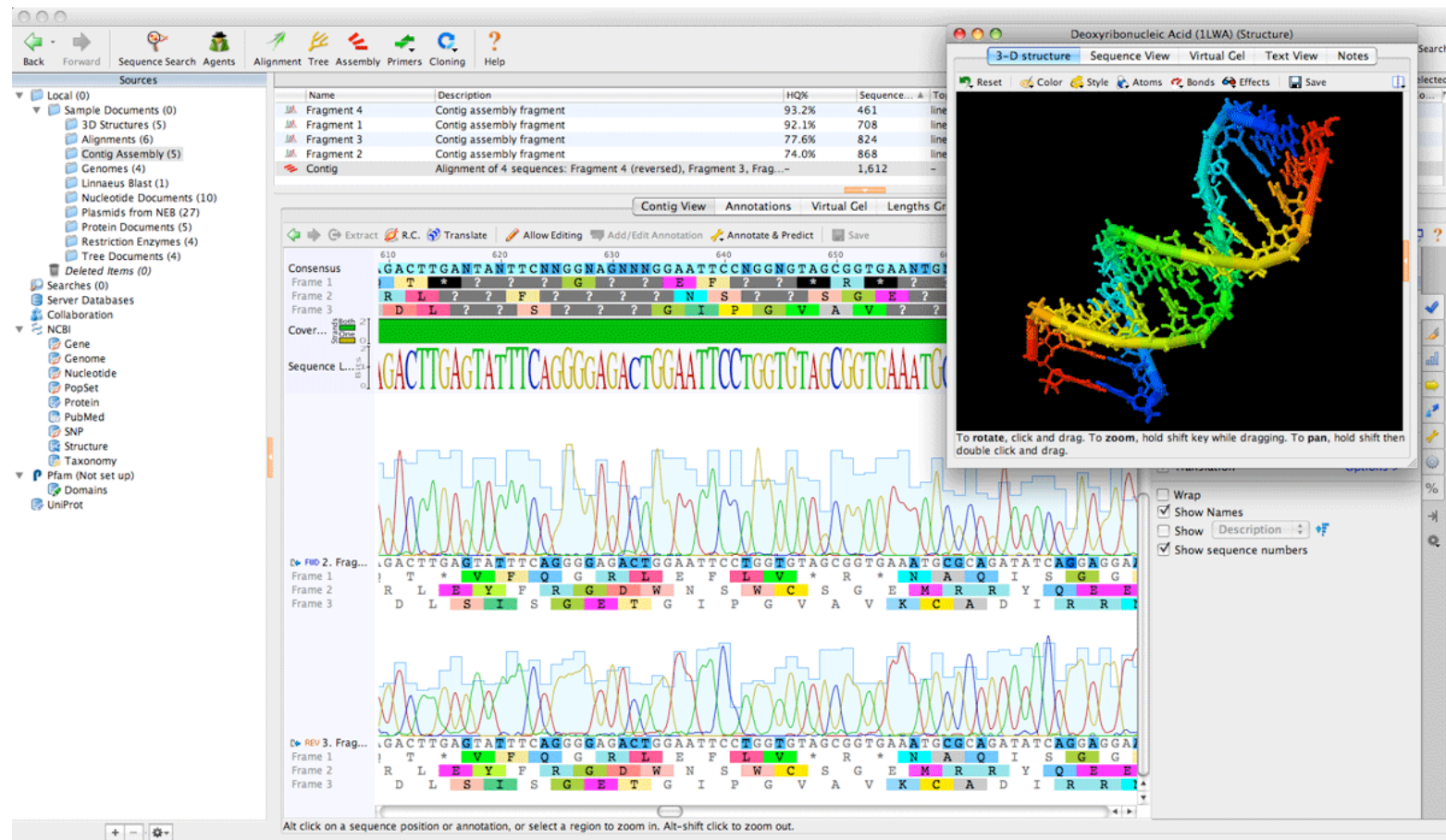
CLC Genomics Workbench

Analisi bioinformatica



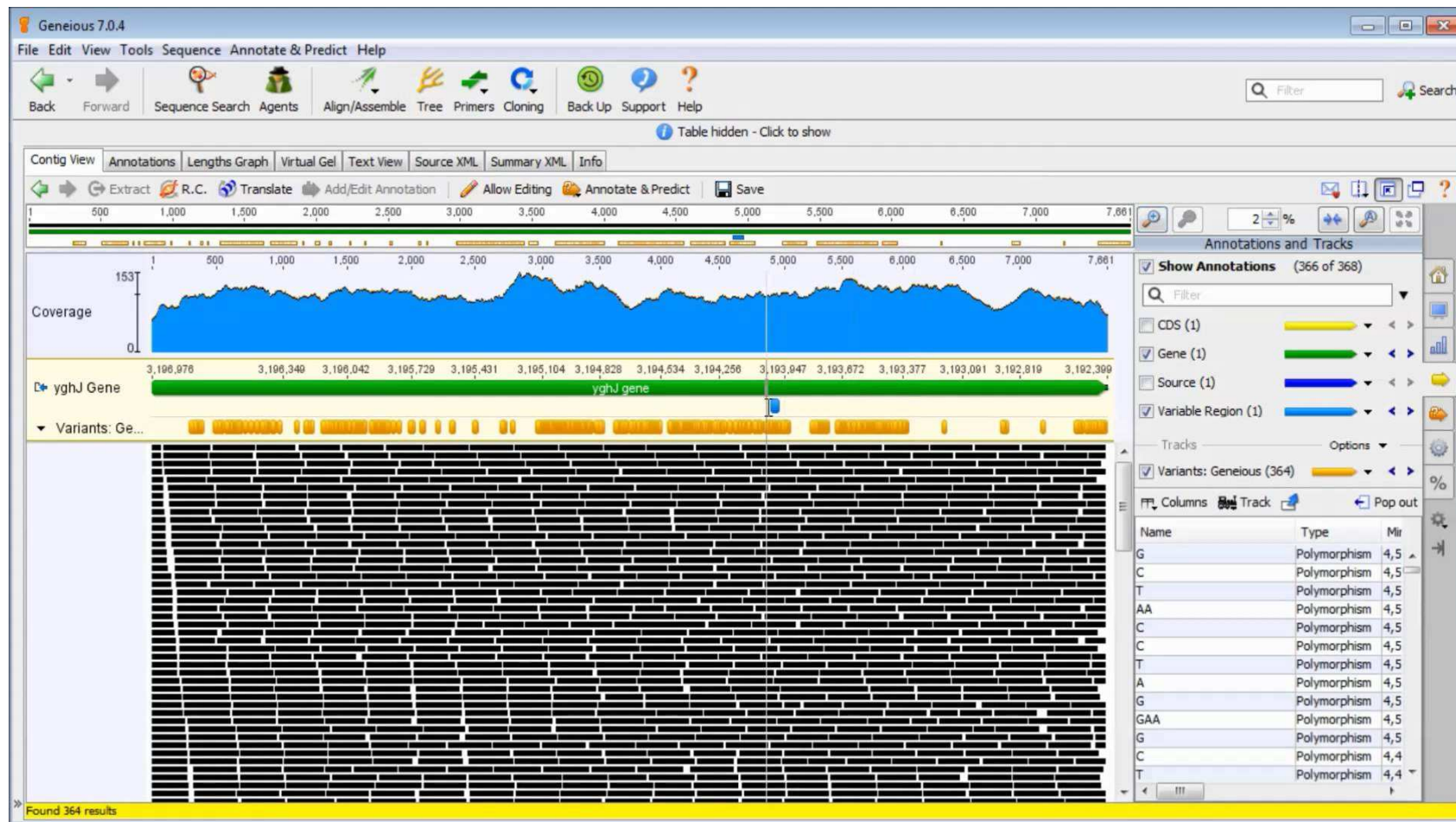
CLC Genomics Workbench

Analisi bioinformatica



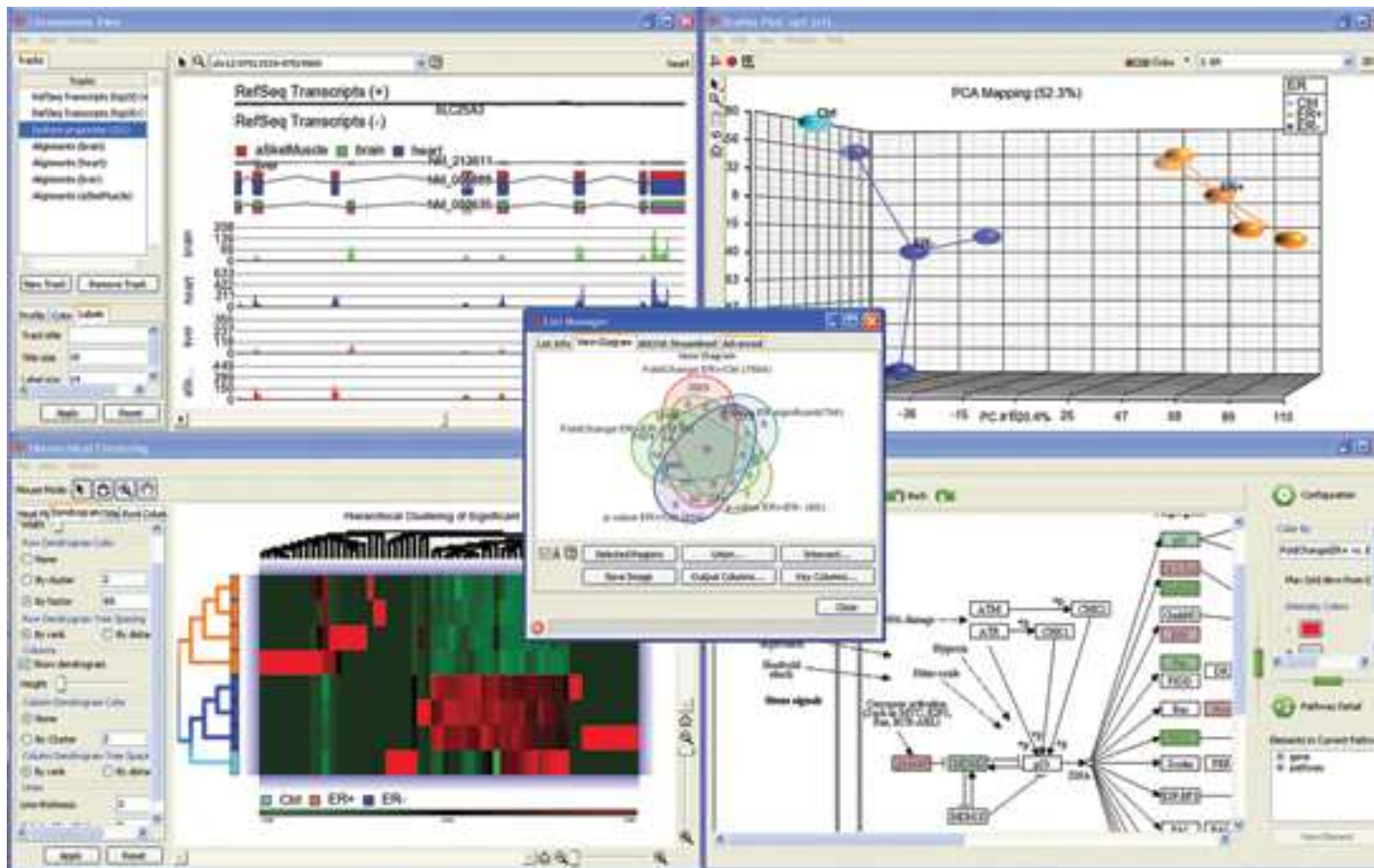
Geneious

Analisi bioinformatica



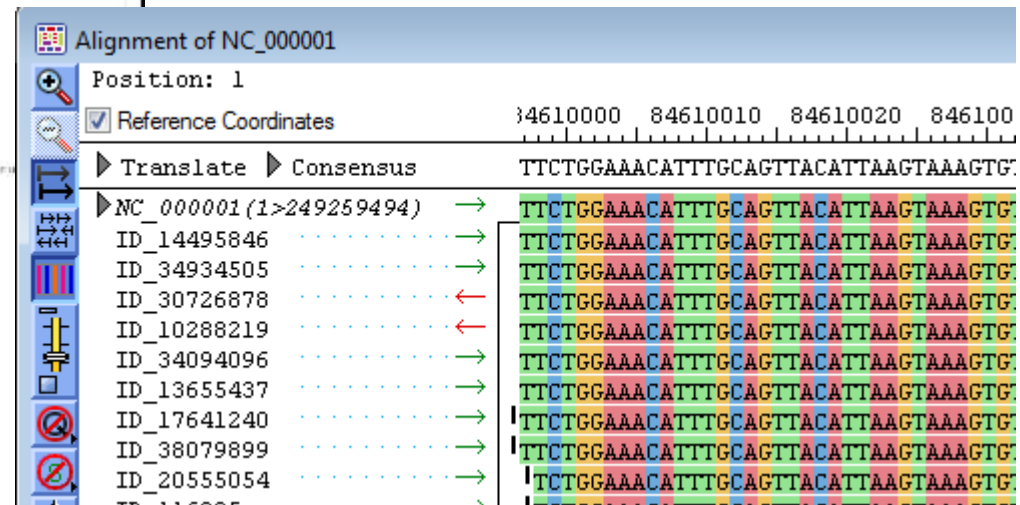
Geneious

Analisi bioinformatica



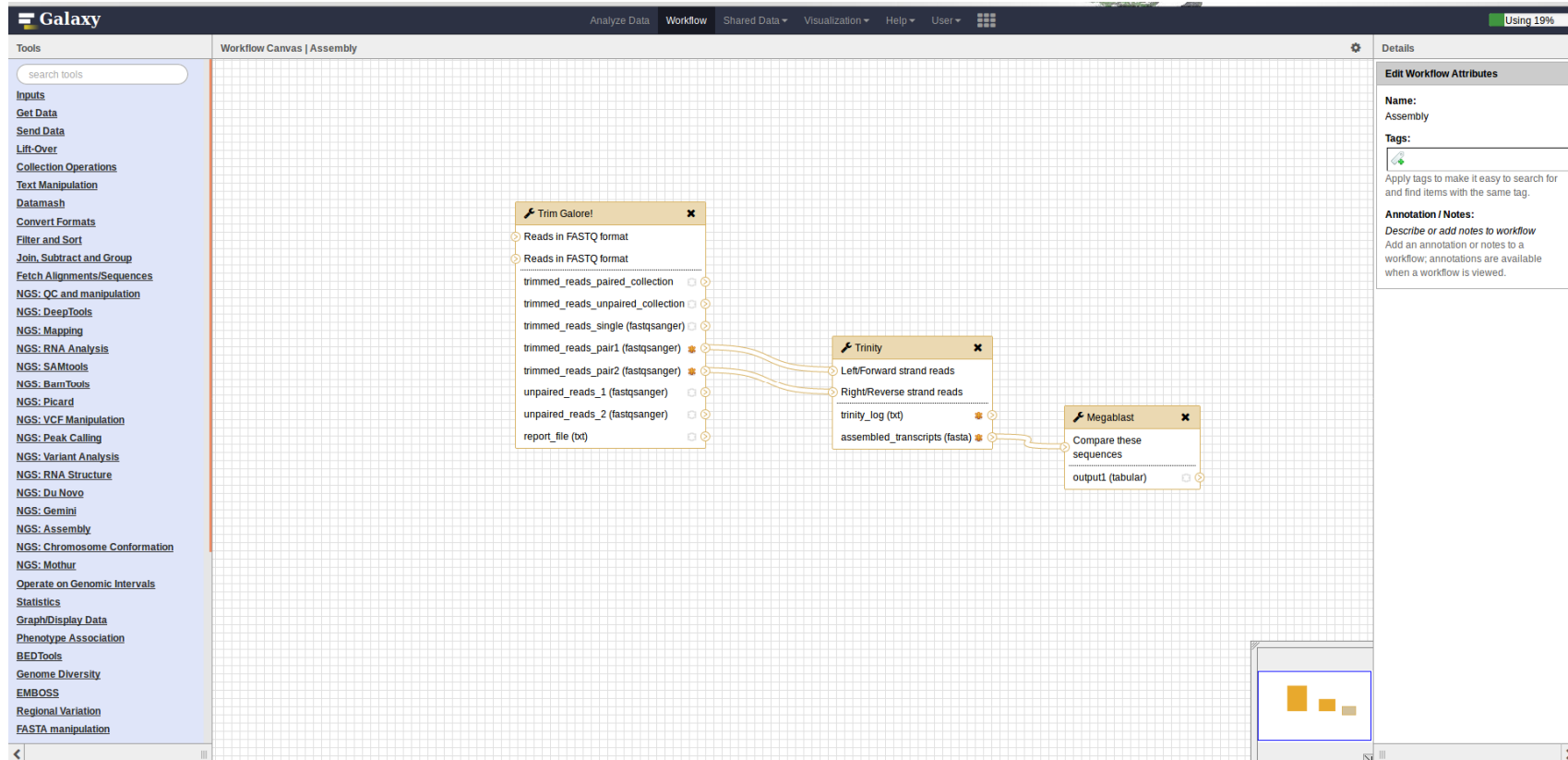
Partek Genomics Suite

Analisi bioinformatica



SeqMan Pro

Analisi bioinformatica



Galaxy

Analisi bioinformatica

BaseSpace **SEQUENCE HUB** ▼

DASHBOARD PREP RUNS PROJECTS **APPS** PUBLIC DATA

Search Apps ? IZS Torino ▼

Applications

Rescaf
BaseSpace Labs

SPAdes Genome Assembler
Algorithmic Biology Lab

SRA Import
BaseSpace Labs

SRA Submission
BaseSpace Labs

String Graph Assembler
BaseSpace Labs

USC BigWig Generator
Ramsingh Lab Keck-Hematology

Velvet de novo Assembly
BaseSpace Labs

Search Apps

Categories

ChIP-Seq (5)	De Novo Assembly (7)
Differential Expression (15)	Gene Fusion Detection (7)
HIPAA (11)	Metagenomics (12)
Methyl-Seq (6)	Proteomics (10)
Quality (9)	Resequencing (19)
RNA-Seq (21)	Small RNA (7)
Synthetic Long Reads (3)	Targeted Sequencing (29)
Tumor Normal (9)	Variant Analysis (25)

contact us

BaseSpace (Illumina)

Analisi bioinformatica

Qualità

- ↗ FastQC
- ↗ Cutadapt
- ↗ Trimmomatic
- ↗ ...

De-novo Assembly

- ↗ Spades
- ↗ Abyss
- ↗ Trinity RNAseq
- ↗ MIRA
- ↗ Megahit
- ↗ Velvet
- ↗ ...

Reference mapping

- ↗ BWA
- ↗ Bowtie
- ↗ BMap
- ↗ ...

Metabarcoding

- ↗ QIIME
- ↗ Mothur
- ↗ R/Bioconductor

Software per linux

Analisi bioinformatica

Software commerciali con GUI

- ↳ Interfaccia grafica
- ↳ Facili da usare
- ↳ Workflow completo
- ↳ Costo elevato
- ↳ Non tutti i tool disponibili

Piattaforme web

- ↳ Interfaccia grafica
- ↳ Facili da usare
- ↳ Workflow completo
- ↳ Gratuiti (Galaxy) o poco costosi (BaseSpace)
- ↳ Non tutti i tool disponibili
- ↳ Lunghi tempi di attesa per tempi di elaborazione

Software free sviluppati in ambiente linux

- ↳ Molti tool disponibili e gratuiti
- ↳ Massima personalizzazione dell'analisi
- ↳ Non facili da usare
- ↳ Personale formato

Analisi bioinformatica

