



Genomica computazionale
Università di Trieste, Anno 2025/2026
Fabrizio Mafessoni

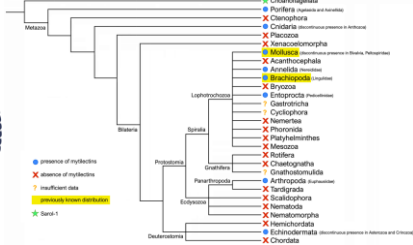
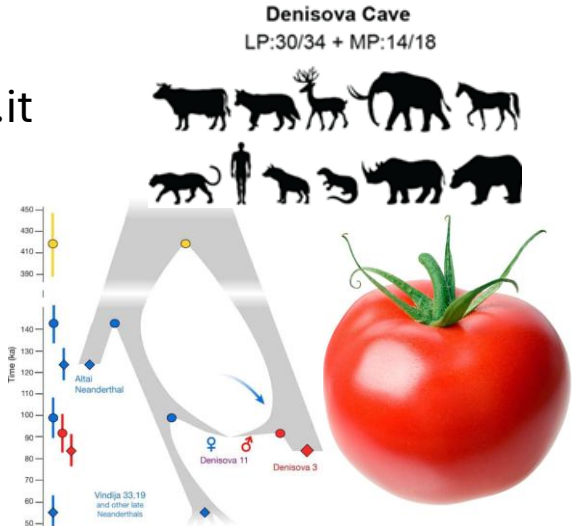
whoami



Fabrizio Mafessoni
fabrizio.mafessoni@units.it



Daniela Eugenia Nerelli
Tutor e dottoranda all'Università



Syllabus

Il corso introduce ai concetti fondamentali e all'applicazione pratica dei metodi computazionali per lo studio dei genomi. Questi strumenti vengono applicati e approfonditi in diversi contesti, come l'assemblamento di genomi, l'epigenomica, la caratterizzazione funzionale di mutazioni e di zone del genoma, la genomica delle popolazioni, l'analisi del DNA antico e la metagenomica. IN DETTAGLIO: Ruolo della genomica computazionale, attraverso un excursus dei principali progressi in biologia e genetica. Sequenze di DNA e formati fasta e fastq. Visualizzazione fasta e fastq. Utilizzo della linea di comando per la pratica e gestione e manipolazione di dati genomici. Assemblaggio di genomi. Valutazione di qualità dei genomi. Annotazione di genomi. Mappatura e formato bam. La trasformazione di Burrow-Wheeler. BWA and Bowtie. Competitive mapping. Primers di scripting in R, shell e statistica. Distribuzioni e maximum likelihood. Genotipizzazione e i formati vcf e bed. Analisi di files vcf (short reads). Mappatura di long reads e genotipizzazione di varianti strutturali. Progetto di genotipizzazione: implementazione di un semplice programma di genotipizzazione. Allineamento di genomi (Progressive Cactus) e basi di filogenetica. Individuazione di zone conservate e di interesse funzionale nel genoma. PhastCons, GERP, PhyloP. Metodi basati su Machine Learning. Teoria degli Hidden Markov Models, loro implementazione ed applicazioni nell'individuazione di zone di interesse funzionale nel genoma (PhastCons), in epigenomica e regolazione dell'espressione genica (peak calling, ChromHMM ed epigenomica), nell'identificazione di famiglie geniche e domini funzionali (Profile Hidden Markov Models e HMMer) e genomica di popolazione (PSMC e altri metodi basati sul Sequentially Markovian Coalescent). Paleogenomica e DNA antico. Metagenomica. Metodi per lo studio di comunità microbiche ed eucariotiche. Metodi basati su k-mers (Kraken). Machine learning in genomica.

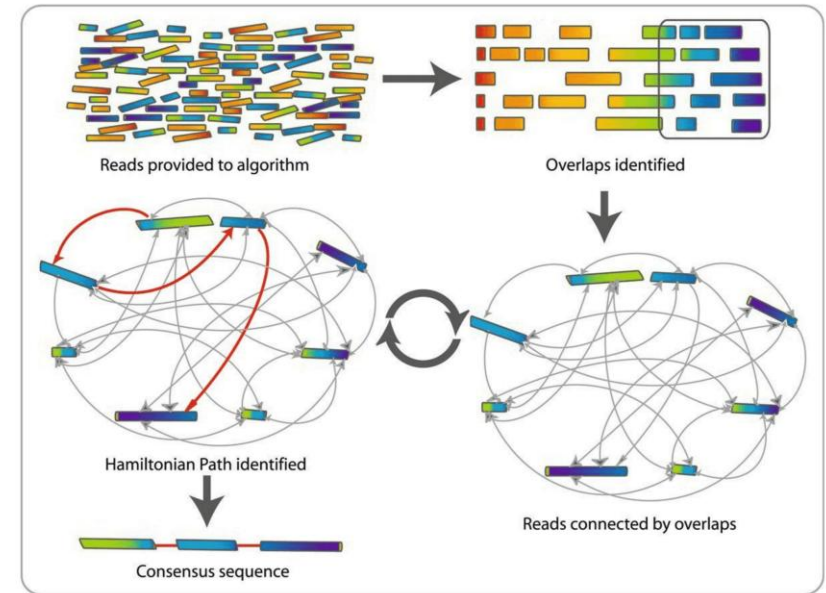
Ambiti biologici

Genomica

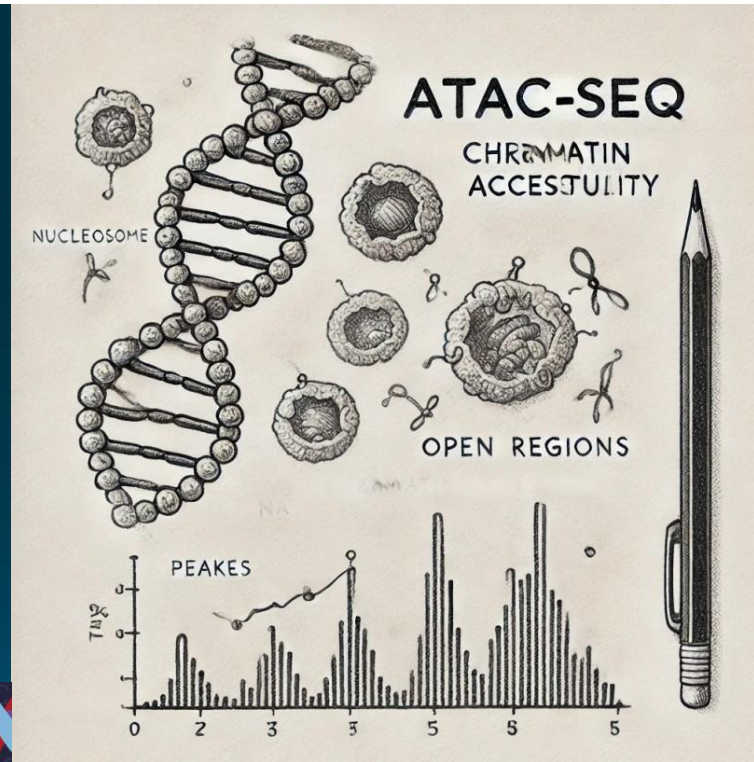
Epigenomica

Metodi

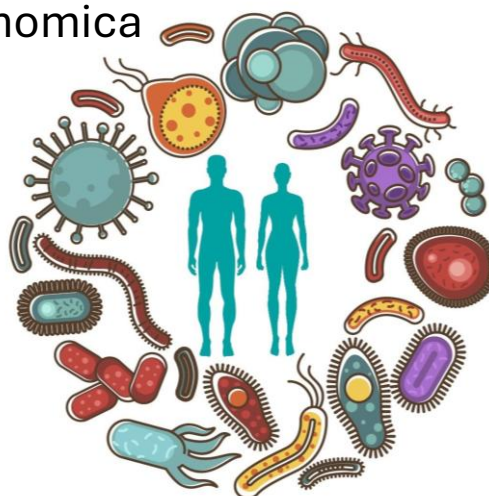
Algoritmi e statistica



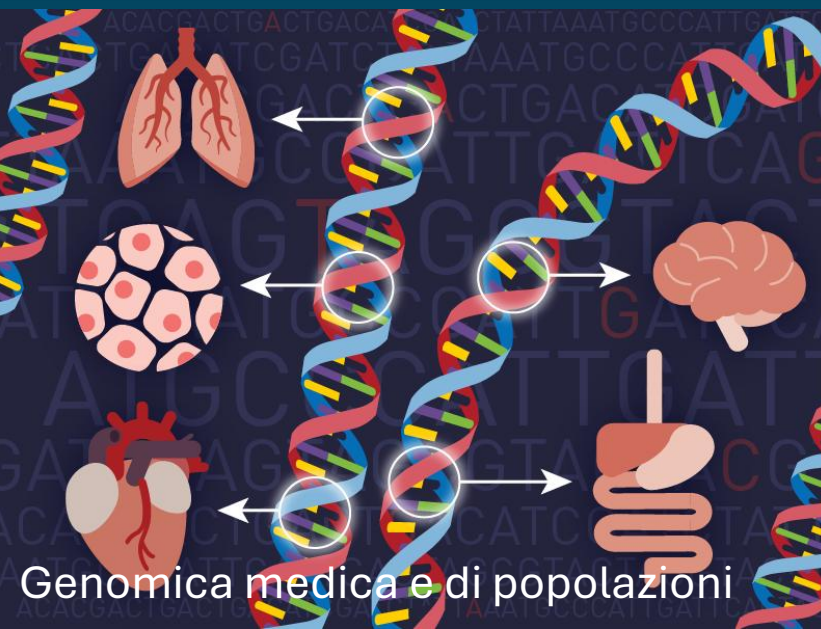
Terminale e programmi bioinformatici



Metagenomica

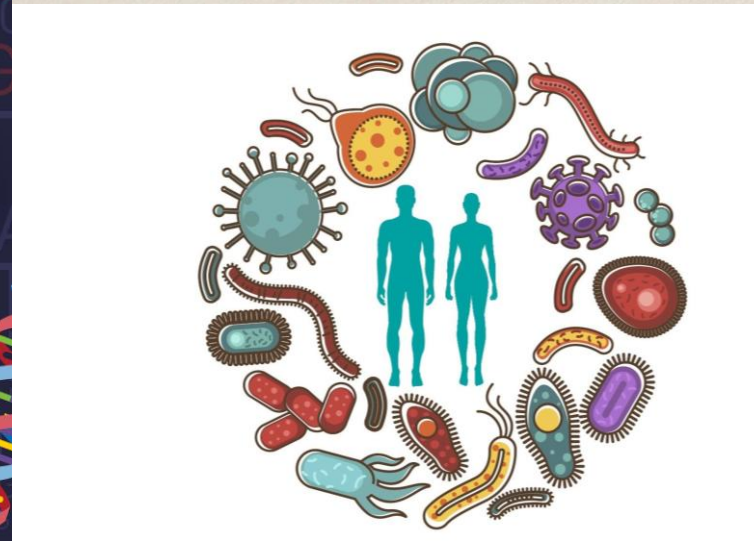
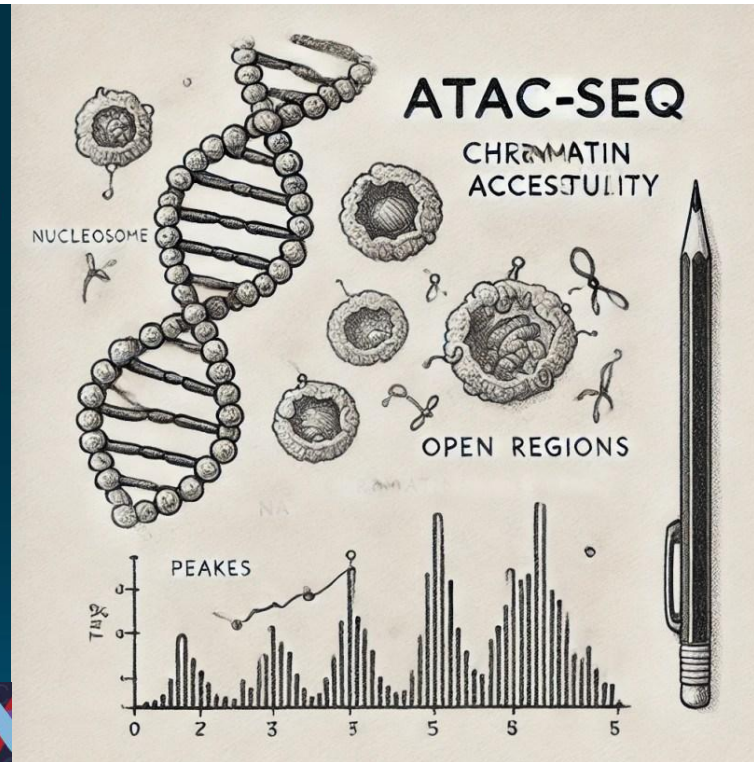
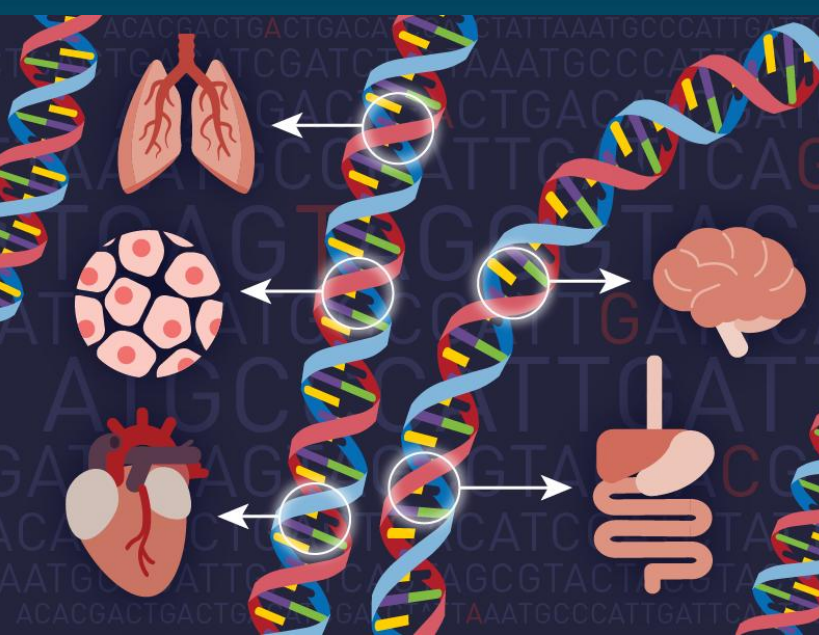


ASSEMBLAGGIO, ANNOTAZIONE E MAPPATURA

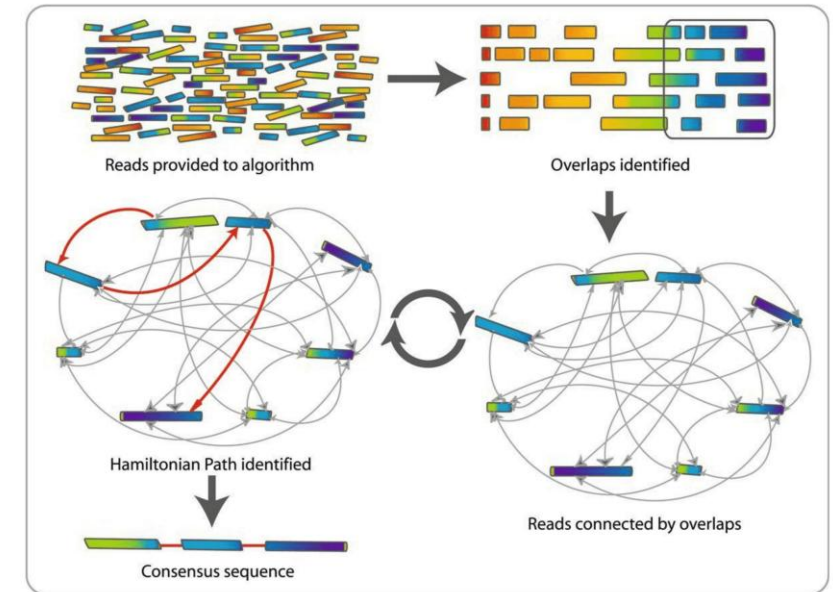


Genomica medica e di popolazioni

Teoria (24 ore)



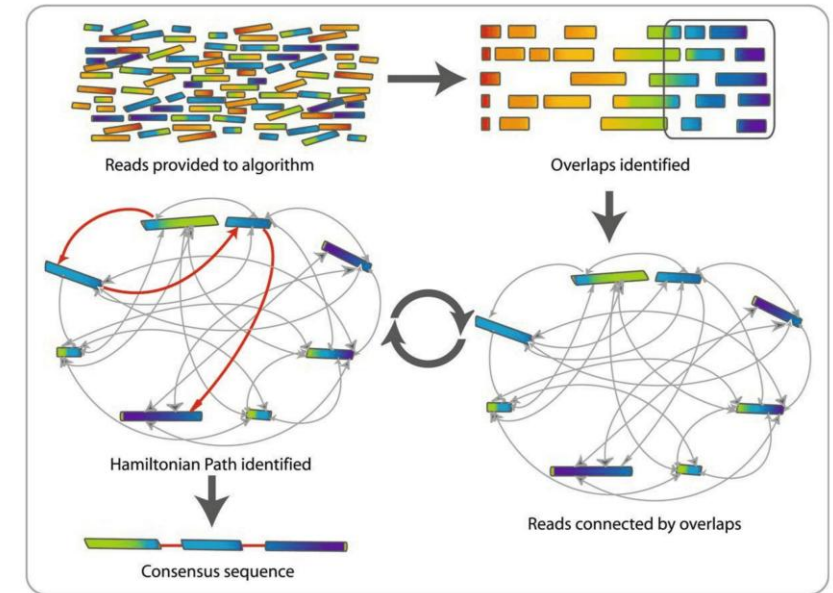
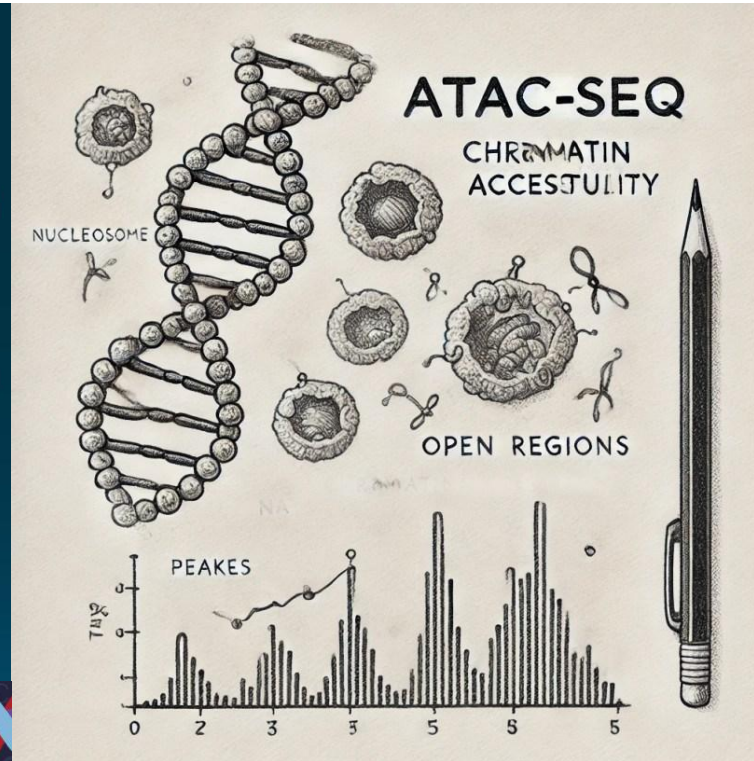
Algoritmi e statistica



Terminale e programmi bioinformatici

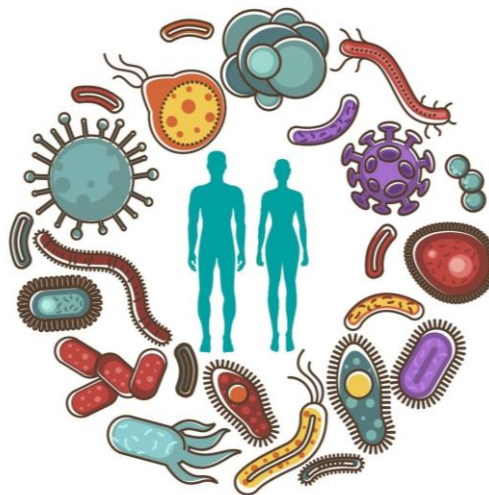


Teoria (24 ore)



Abbiamo tempo, chiedete!

Matematica e algoritmi complessi saranno tenuti al minimo (ma studiate bene le poche basi iniziali per godervi meglio il resto!)



• Perché il terminal/linea di comando?

• Potente e versatile:

- La maggior parte delle applicazioni di genomica computazionali sono solo su riga di comando
- Possibilità di eseguire programmi in remoto senza interfaccia grafica e su server, più rapidamente ed efficientemente

• Flessibile:

- Potete scrivere i vostri programmi e creare pipeline customizzate

• Riproducibile:

- Il codice si può salvare meglio che una sequenza di punta e clicca col mouse

• Più semplice:

- Ve lo prometto!



Terminale e programmi bioinformatici



Pratica (36 ore)

- Perché il terminal/linea di comando?

- Potente e versatile:

- La maggior parte delle applicazioni di genomica computazionali sono solo su riga di comando
 - Possibilità di eseguire programmi in remoto senza interfaccia grafica e su server, più rapidamente ed efficientemente

- Flessibile:

- Potete scrivere i vostri programmi e creare pipeline customizzate

- Riproducibile:

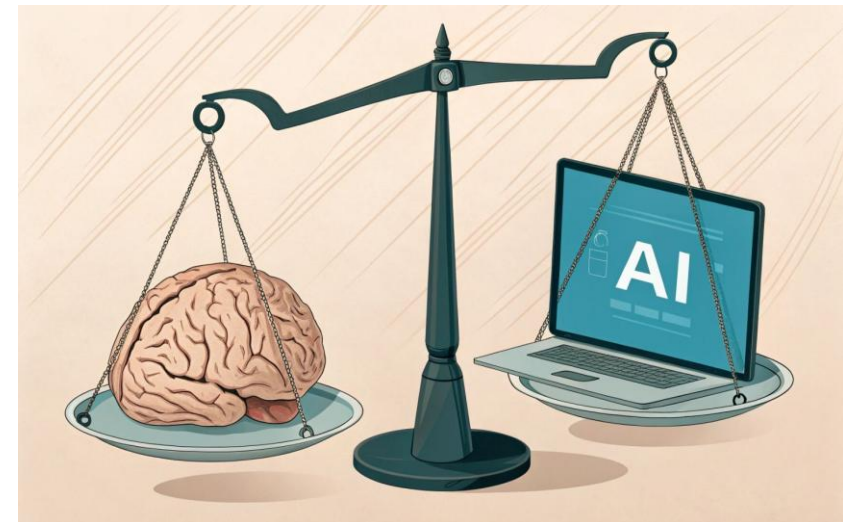
- Il codice si può salvare meglio che una sequenza di punta e clicca col mouse

- Più semplice:

- Ve lo prometto!



- Perché il terminal/linea di comando?
 - Potente e versatile:
 - La maggior parte delle applicazioni di genomica computazionali sono solo su riga di comando
 - Possibilità di eseguire programmi in remoto senza interfaccia grafica e su server, più rapidamente ed efficientemente
 - Flessibile:
 - Potete scrivere i vostri programmi e creare pipeline customizzate
 - Riproducibile:
 - Il codice si può salvare meglio che una sequenza di punta e clicca col mouse
 - Più semplice:
 - Ve lo prometto!
- Per tutto il corso i comandi sul terminal/linea di comando saranno in Courier New
- Chatgpt va bene. Ma non perdetevi in comandi astrusi!
- Per alleggerire le 4 ore del martedì avremo sempre una parte teorica ed una pratica.
- Portate i laptop! Seguite e chiedete a me o Daniela!

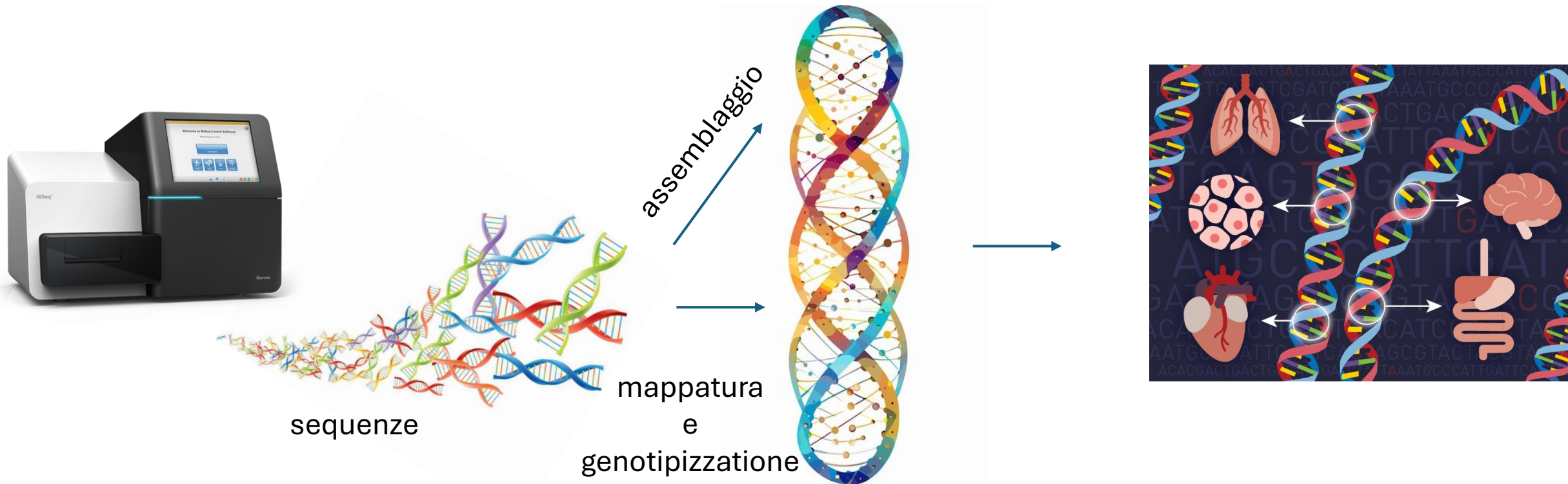


Scopo ultimo del corso: saper partire da sequenze di DNA fino ad assemblare interi genomi e caratterizzarli funzionalmente

Dal sequenziatore..

..al genoma..

..alla funzione

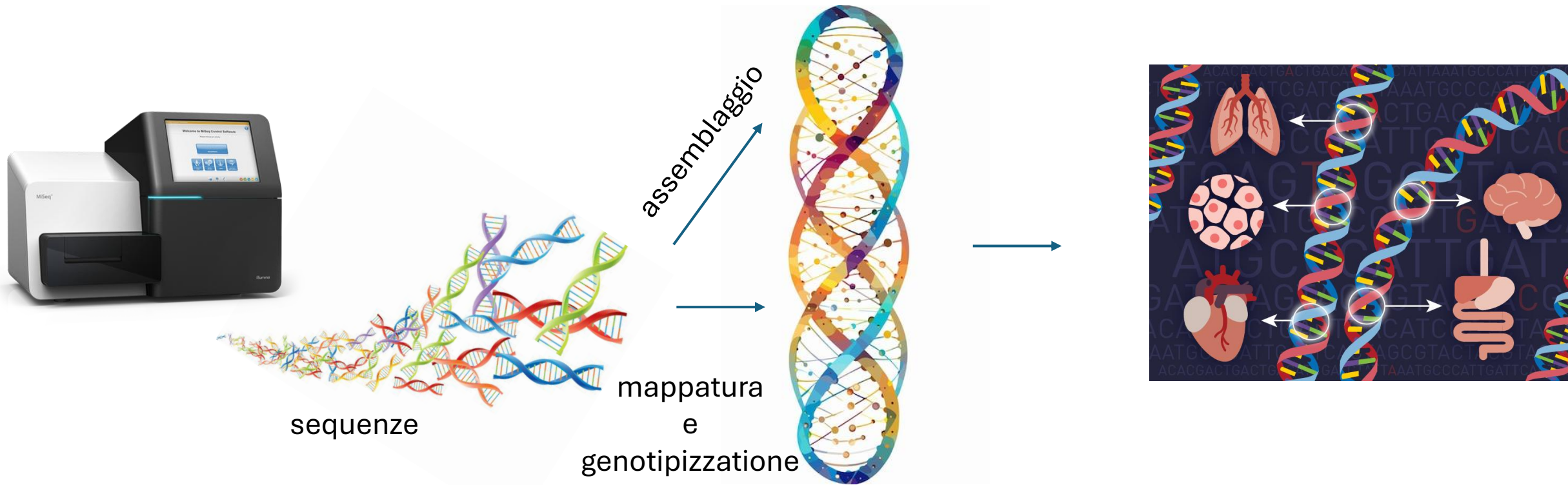


Scopo ultimo del corso: saper partire da sequenze di DNA fino ad assemblare interi genomi e caratterizzarli funzionalmente

Dal sequenziatore..

..al genoma..

..alla funzione



Formati di file bioinformatici

.fastq

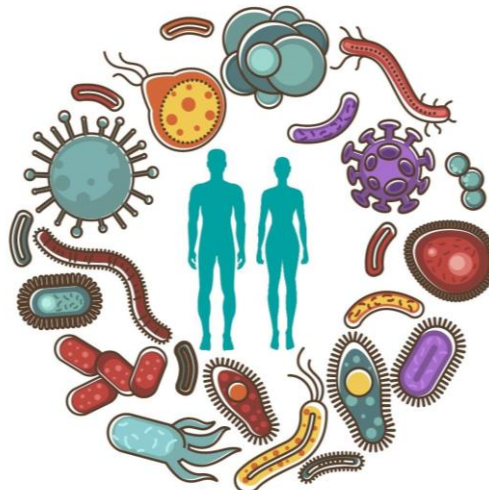
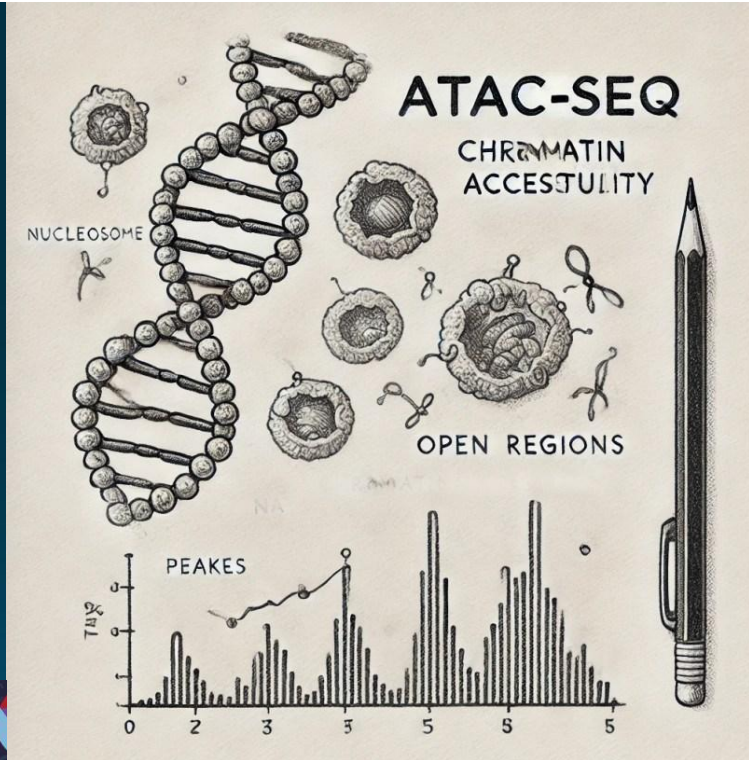
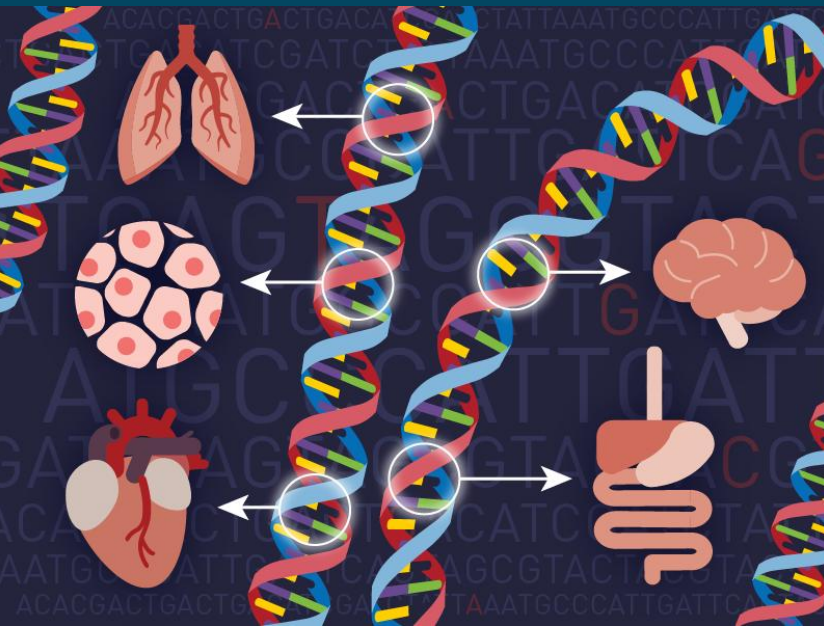
.fasta, .bam, .vcf

.bed, .gff

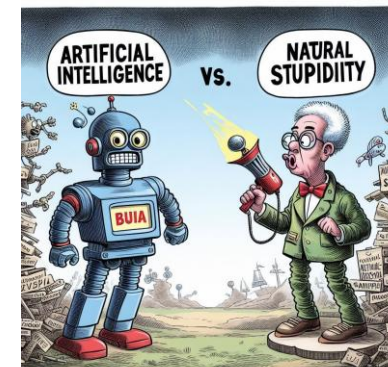
Note sul progetto finale



ASSEMBLAGGIO, ANNOTAZIONE E MAPPATURA

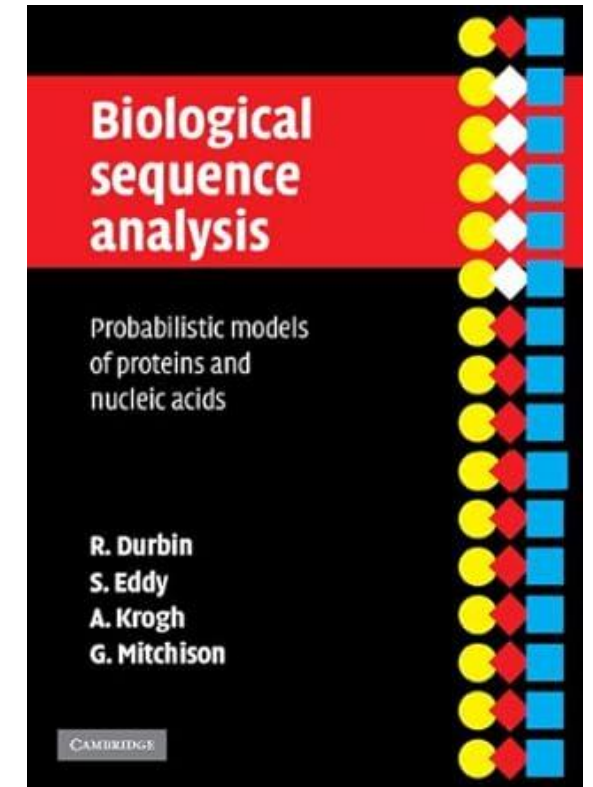
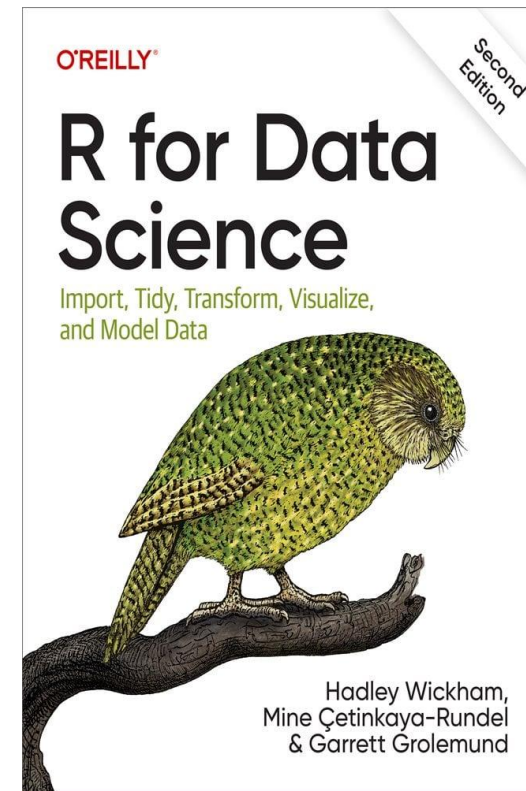
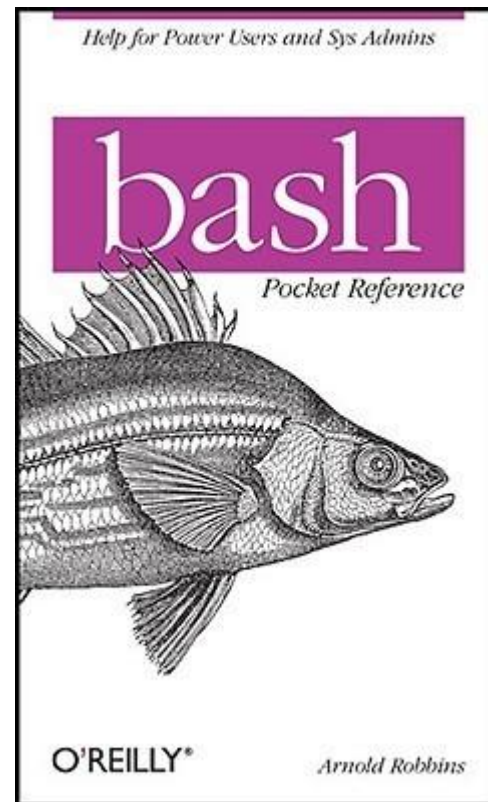
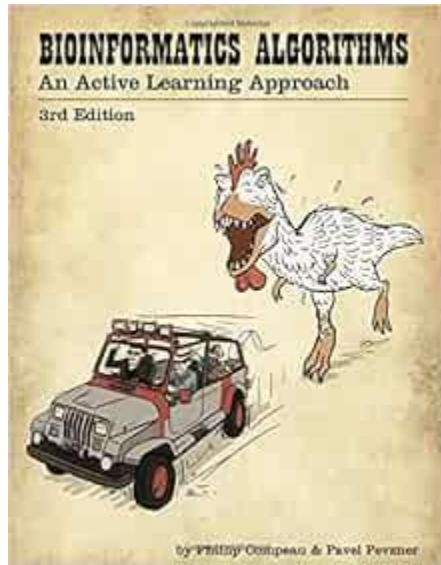


- Vari argomenti disponibili ma iniziate a pensare durante il corso ad un argomento di interesse
- 4 ore di base più consultazioni/esercitazioni aggiuntive
- Breve pipeline/analisi o implementazione di un piccolo algoritmo/programma
- Va bene Chatgpt ma senza esagerare..



Libri di testo

Materiale e capitoli verranno forniti a lezione





Introduzione alla Genomica computazionale

Milestones | 10 February 2021

Milestones in Genomic Sequencing

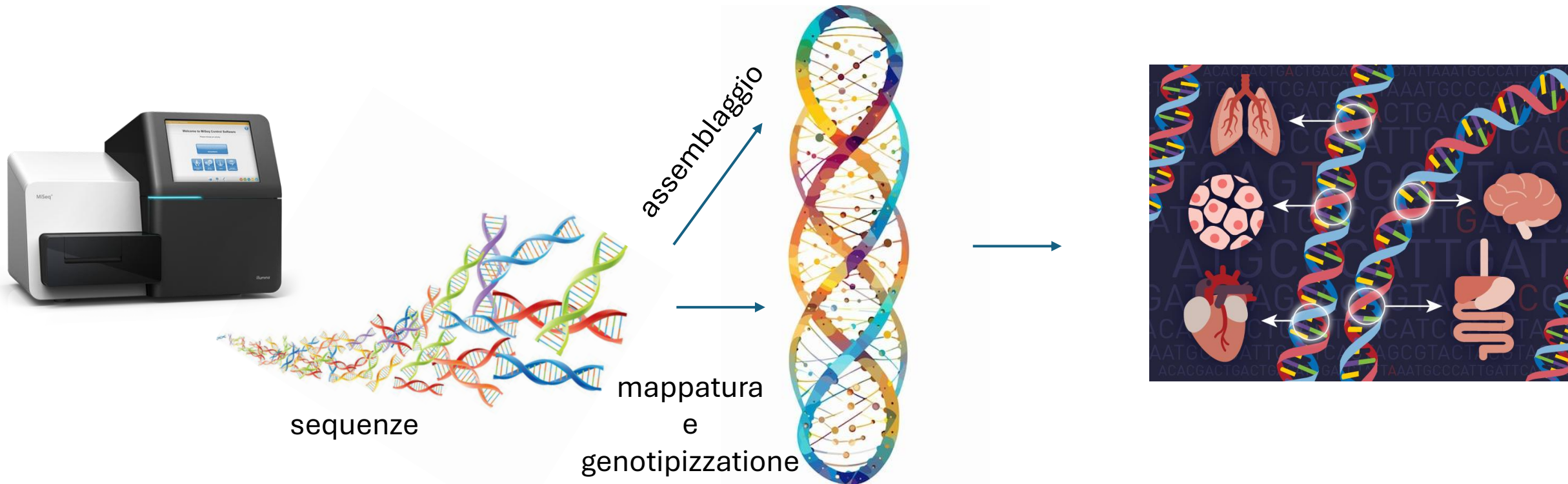


Le “sequenze di DNA”/”sequencing reads”/”reads” sono l’unità di partenza per le analisi di genomica computazionale

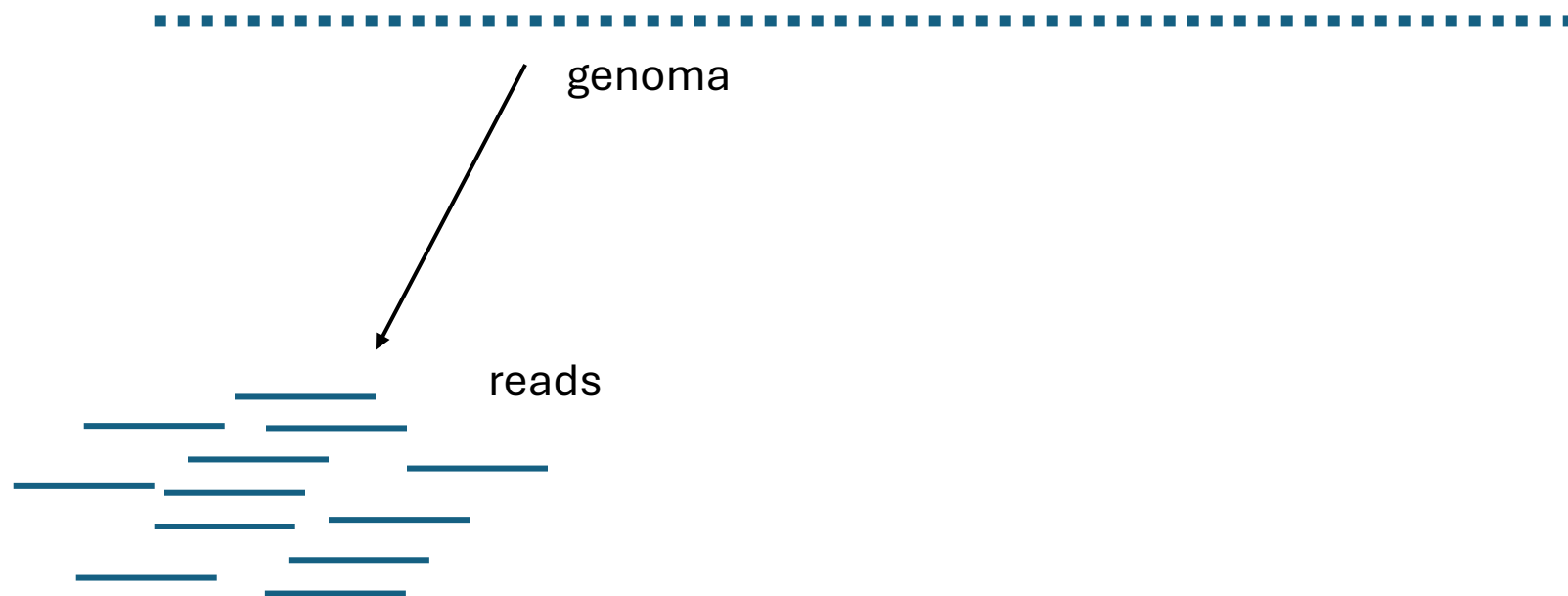
Dal sequenziatore..

..al genoma..

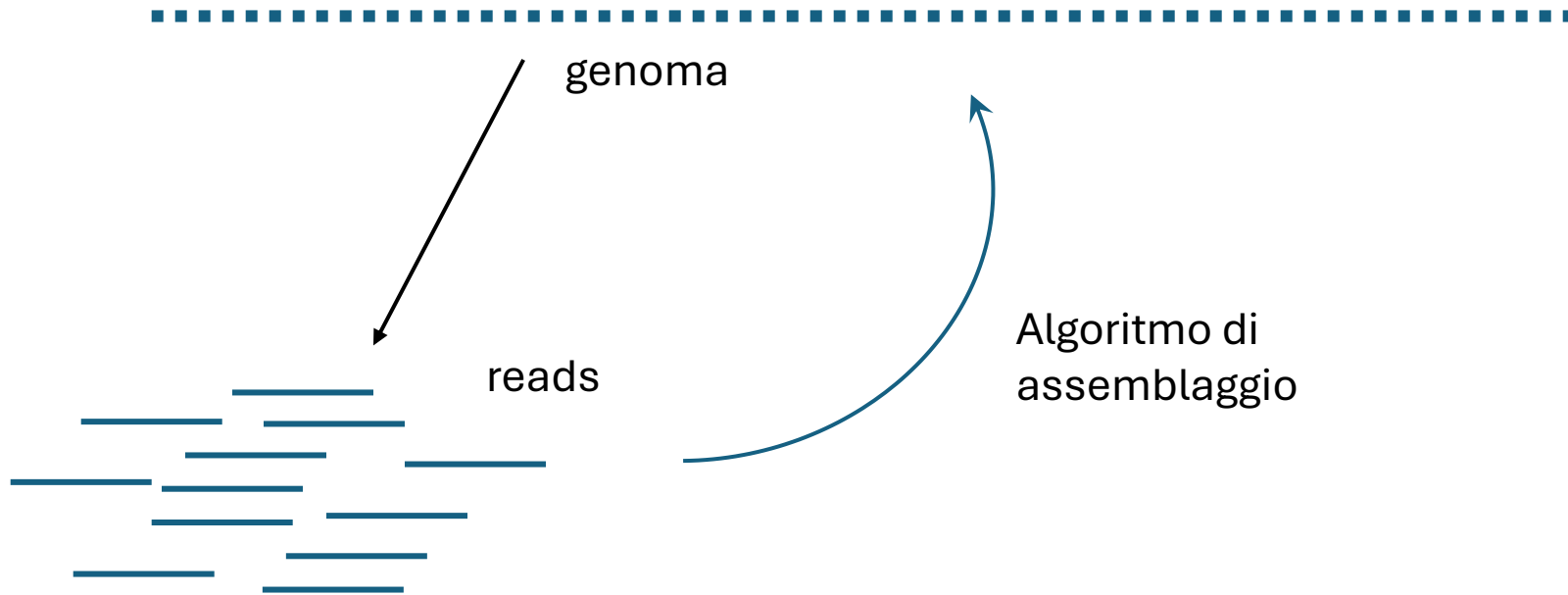
..alla funzione



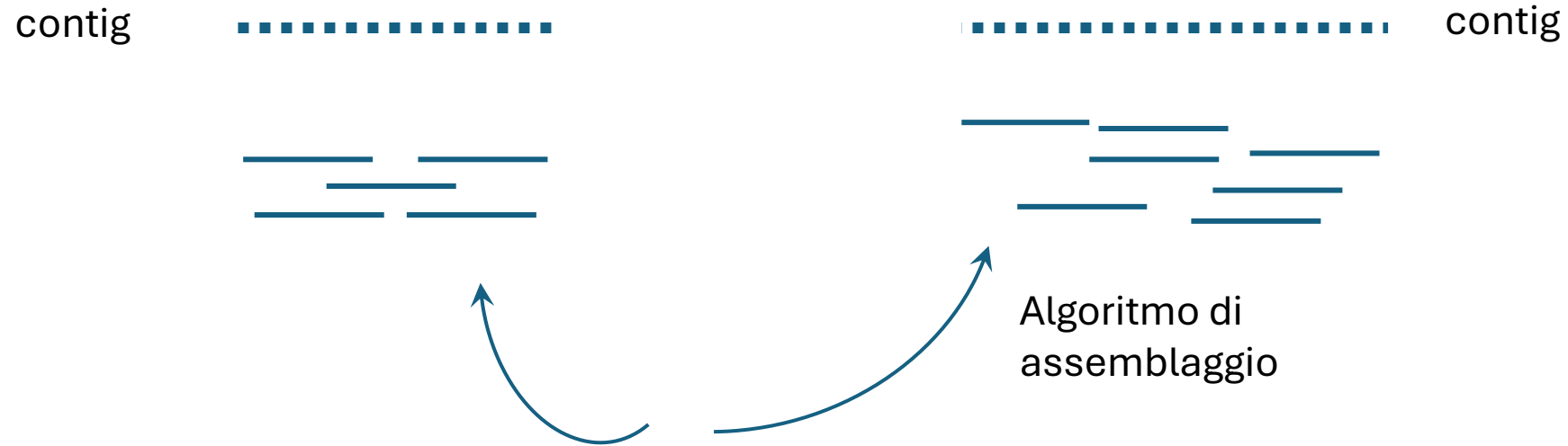
L'assemblaggio di genomi



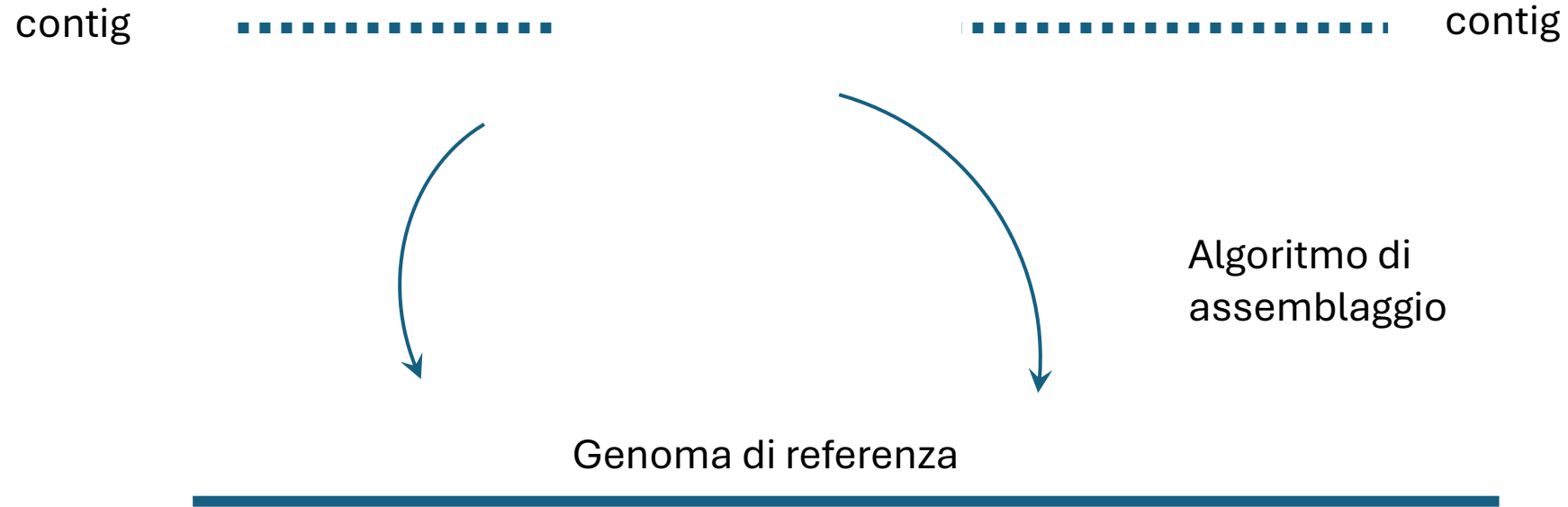
L'assemblaggio di genomi



L'assemblaggio di genomi

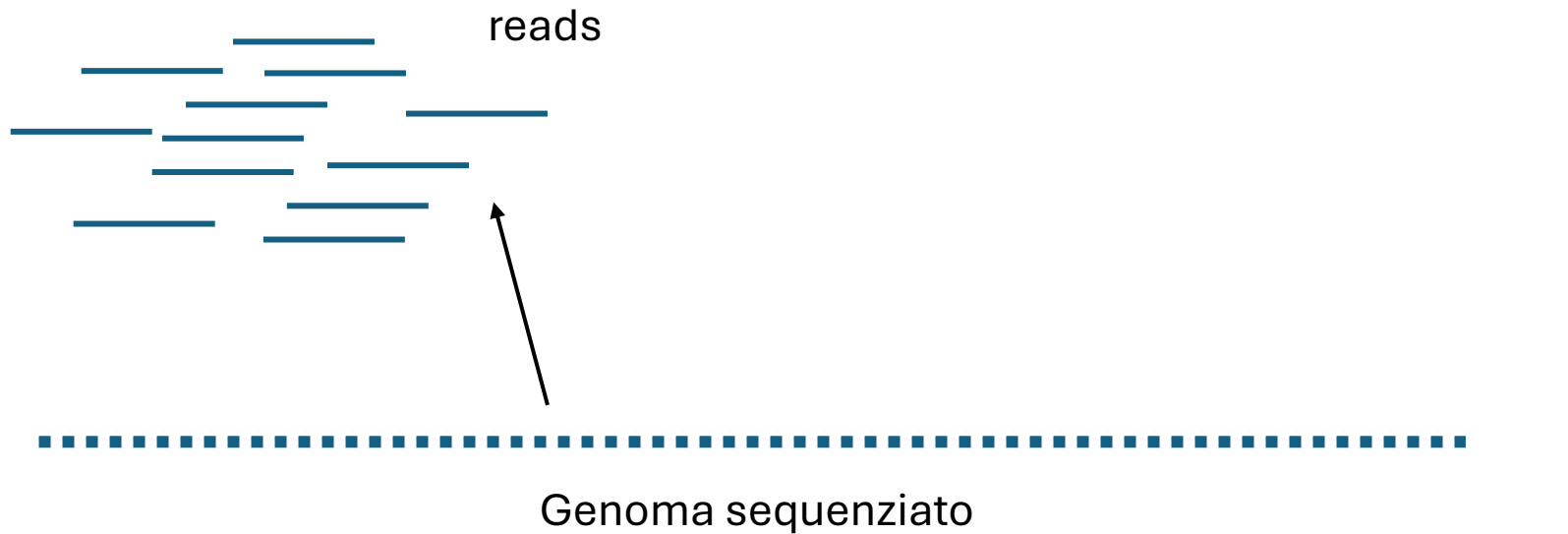


L'assemblaggio di genomi

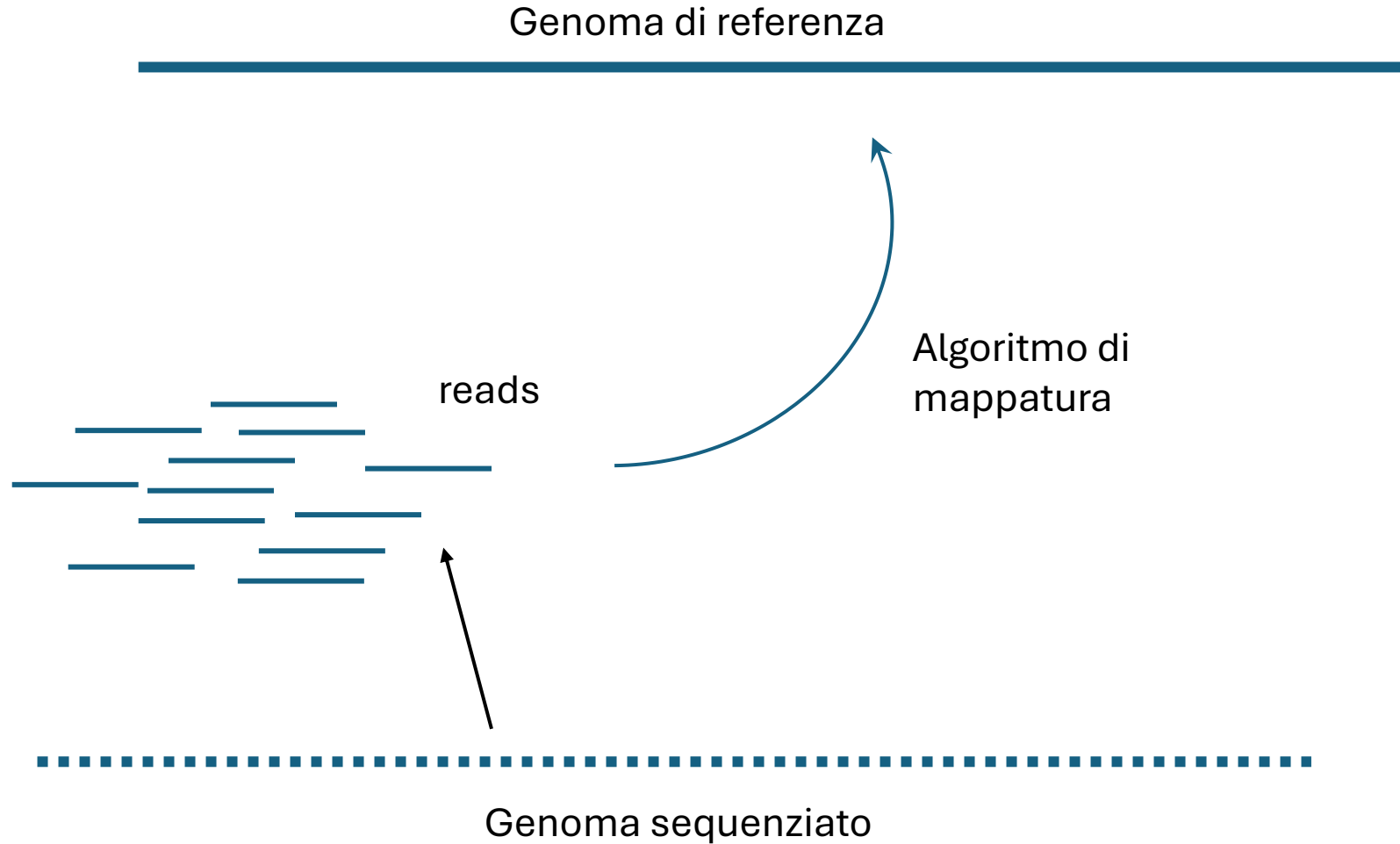


La mappatura delle sequenze

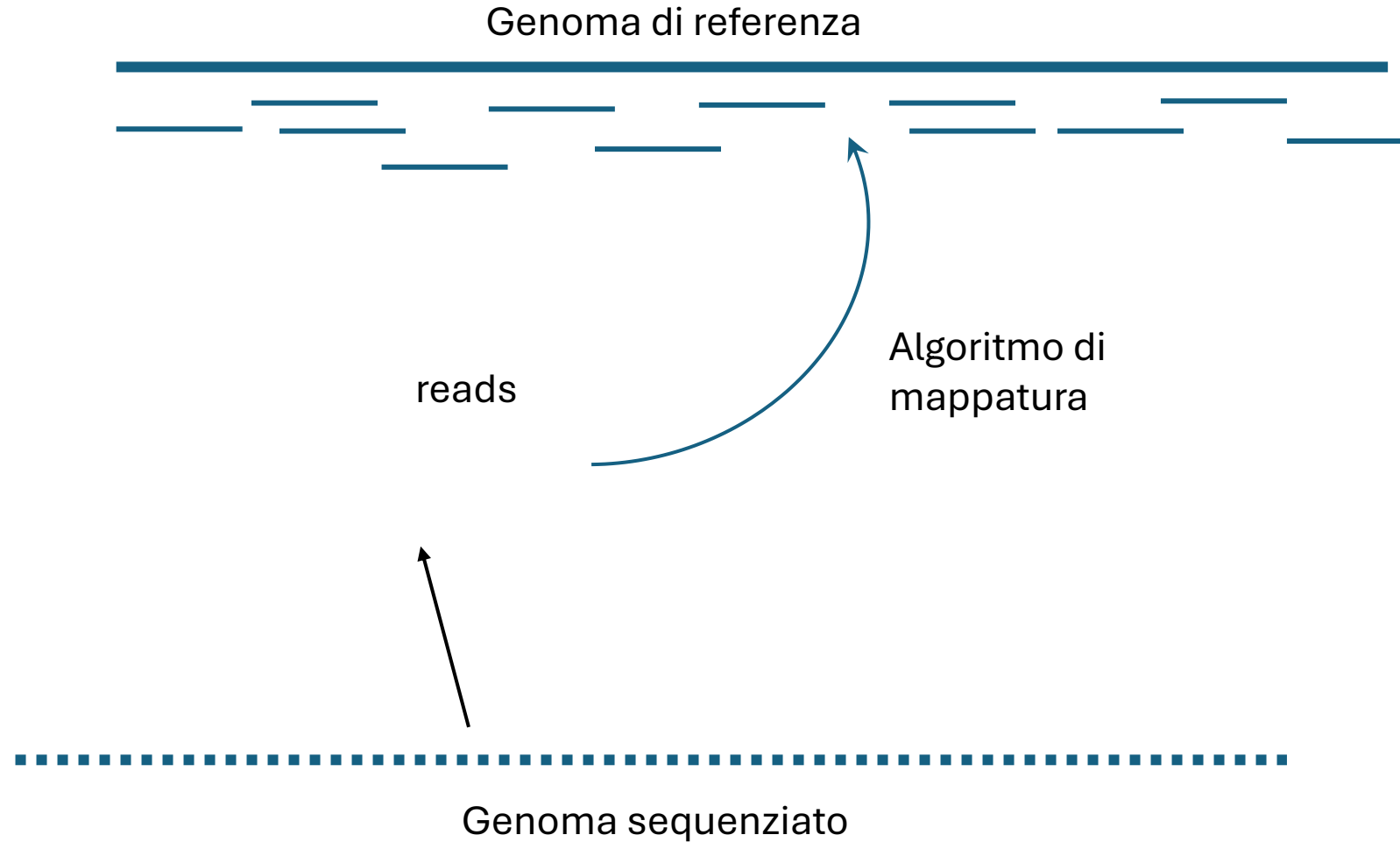
Genoma di referenza



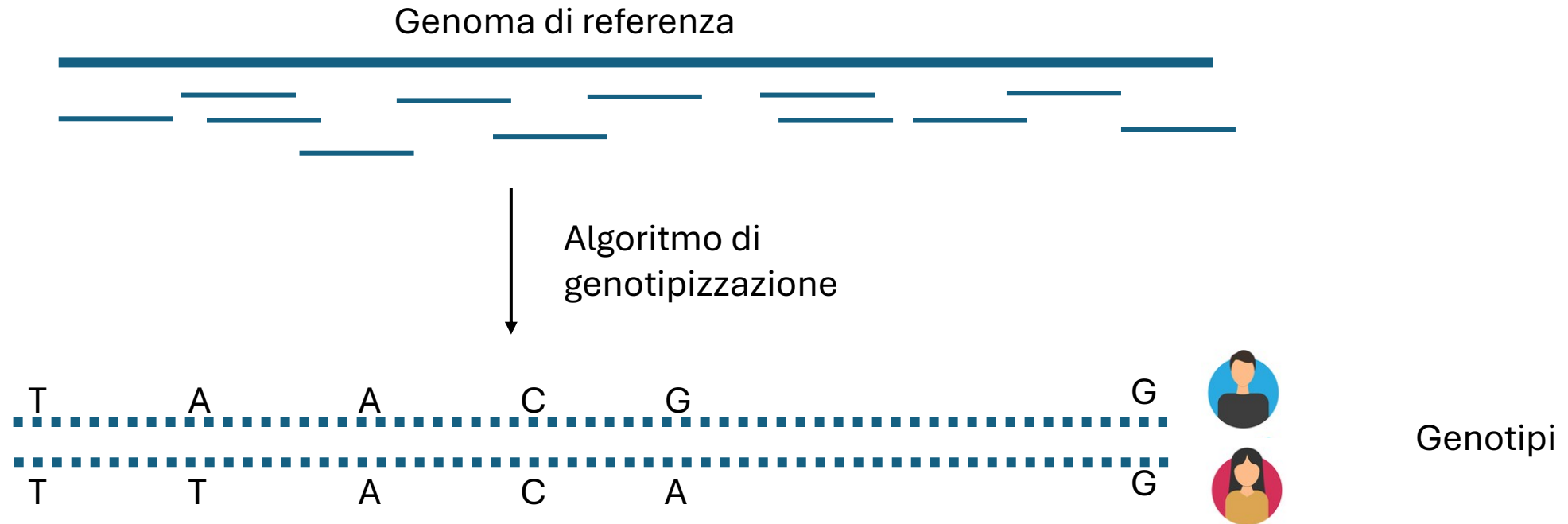
La mappatura delle sequenze



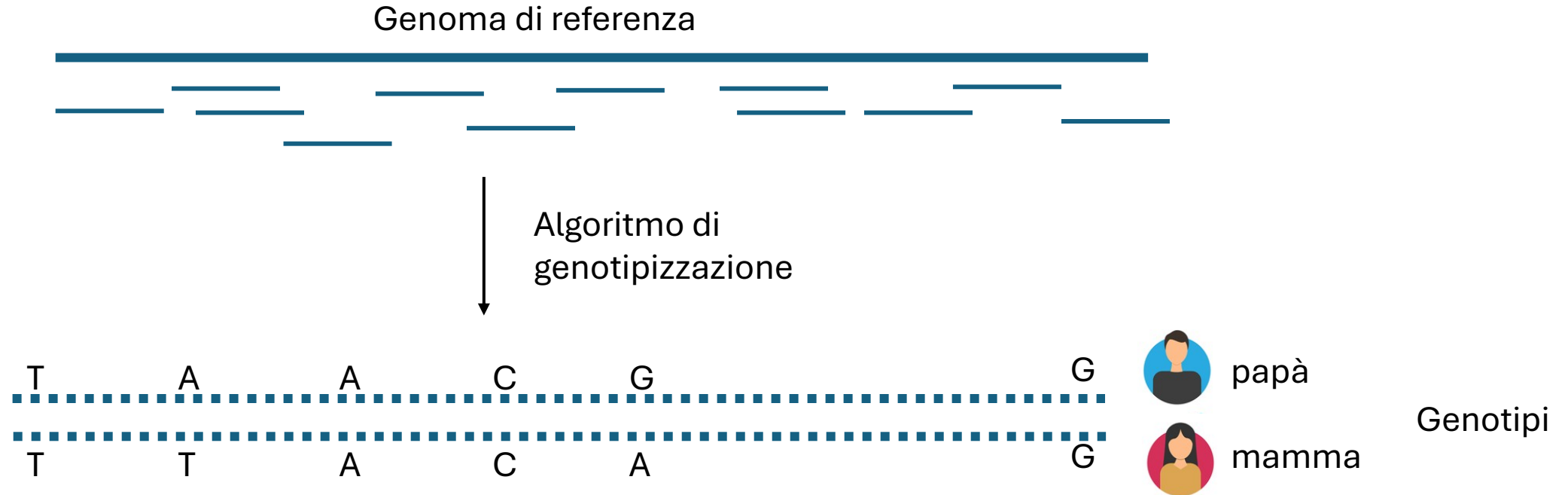
La mappatura delle sequenze



La genotipizzazione



La genotipizzazione

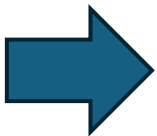


Prima parte del corso:

Scopo: utilizzo delle **sequenze** ottenute dal sequenziatore per operazioni comuni di genomica.

Utilizzo della command line UNIX per utilizzare programmi bioinformatici e manipolare sequenze.

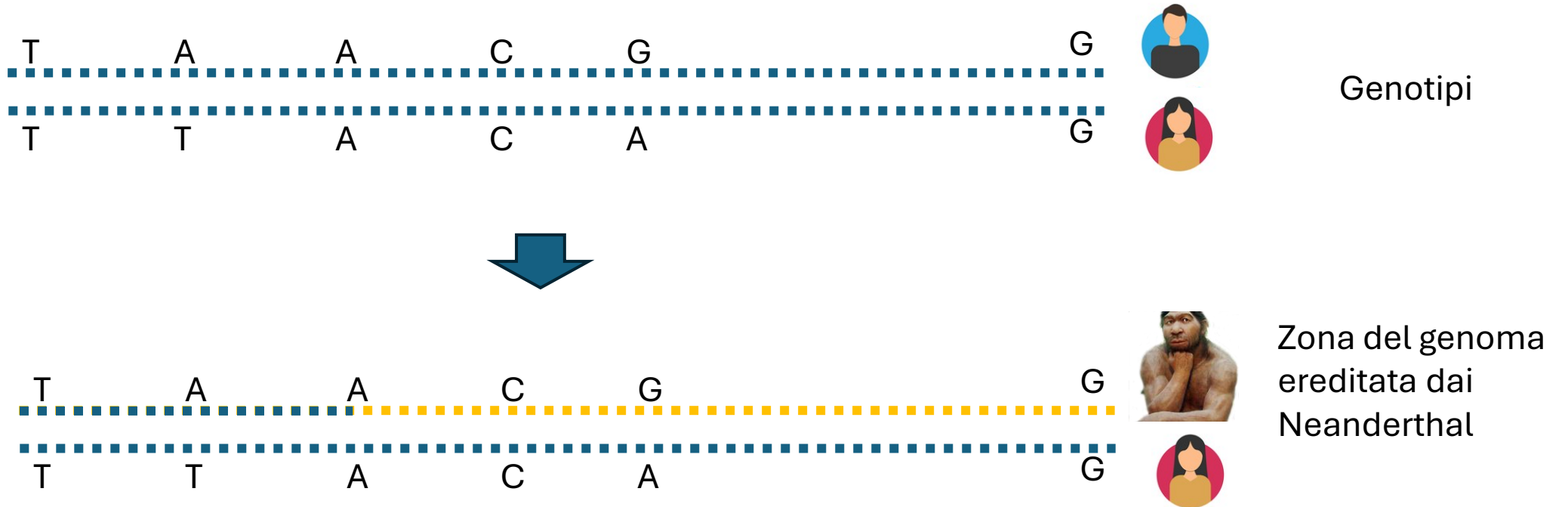
Comprensione degli algoritmi.



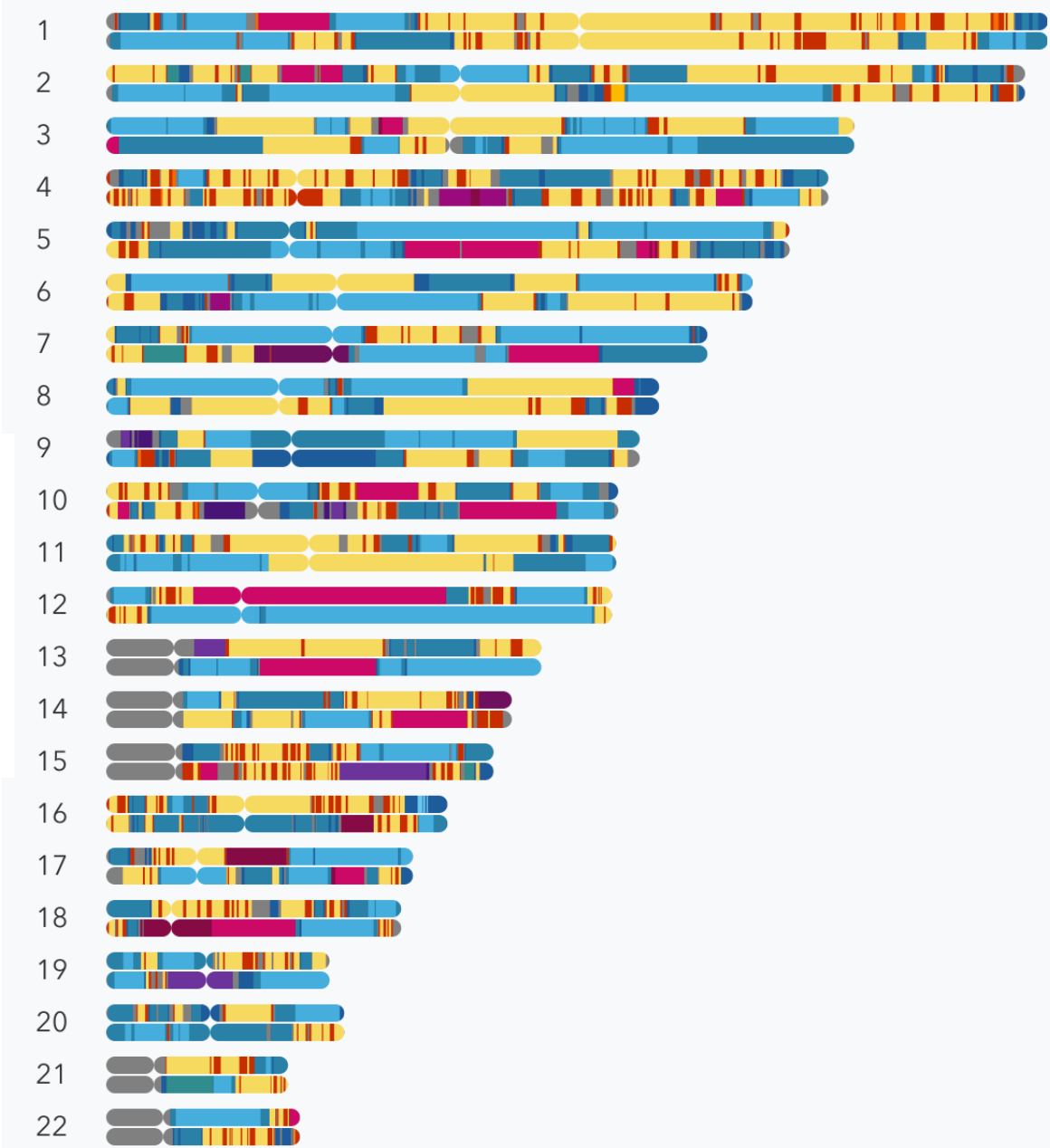
Caratterizzazione funzionale ed evolutiva del genoma



Caratterizzazione evolutiva del genoma

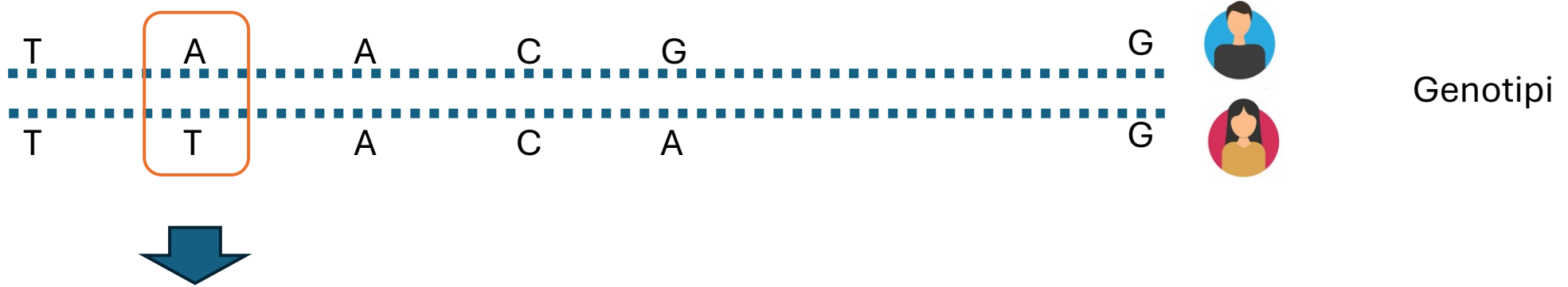


Chromosome-painting



European	48.3%
Southern European	43.3%
Iberian	25.3%
Ashkenazi	0.5%
Northwestern European	< 0.1%
East Asian & Native American	38.7%
Native American	30.7%
East Asian	0.1%
Southeast Asian	< 0.1%
Sub-Saharan African	7.1%
West African	5.3%
Central & South African	0.6%
East African	0.3%

Caratterizzazione funzionale del genoma



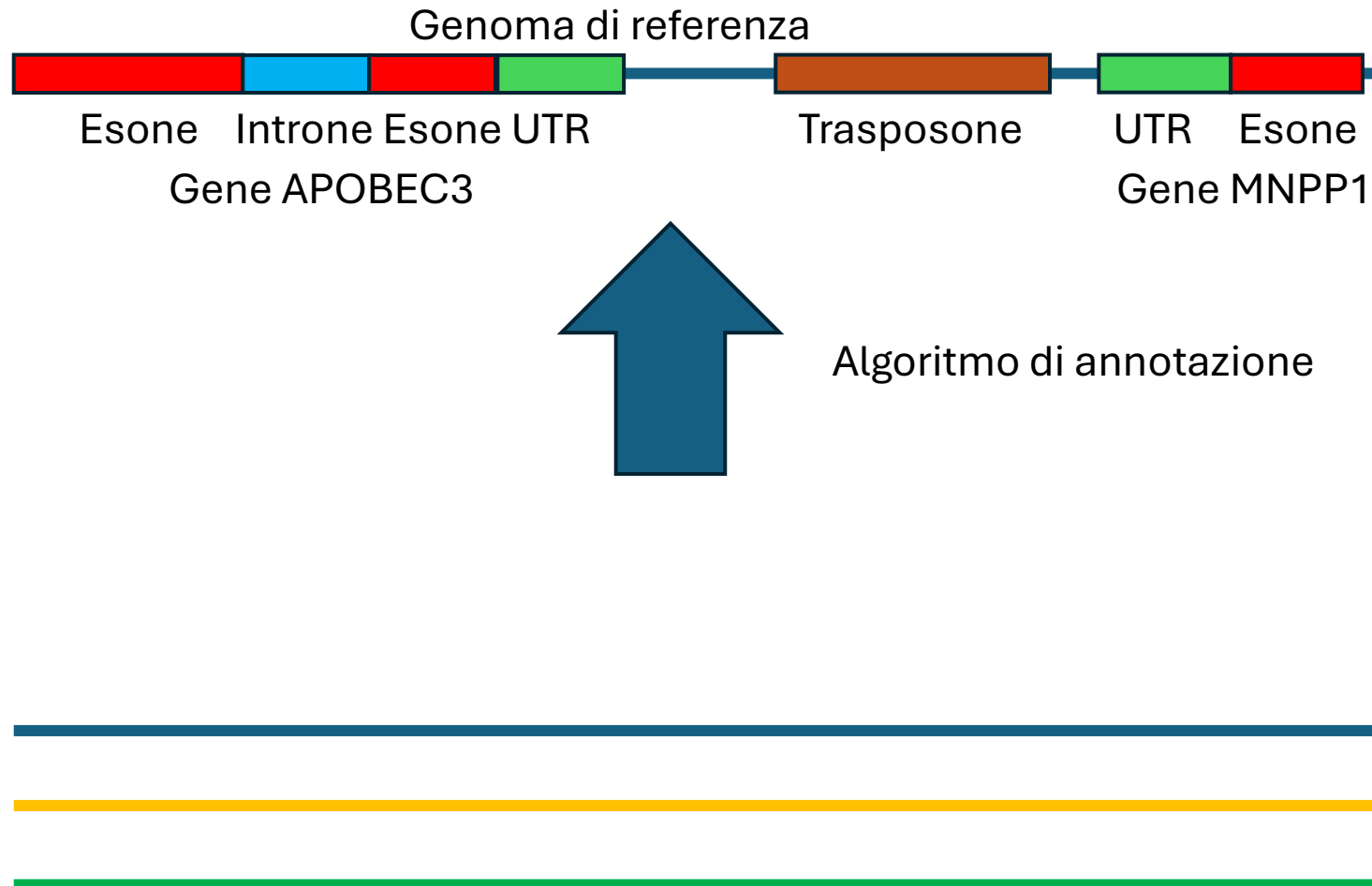
Questa mutazione incrementa il rischio di tumore alla prostata



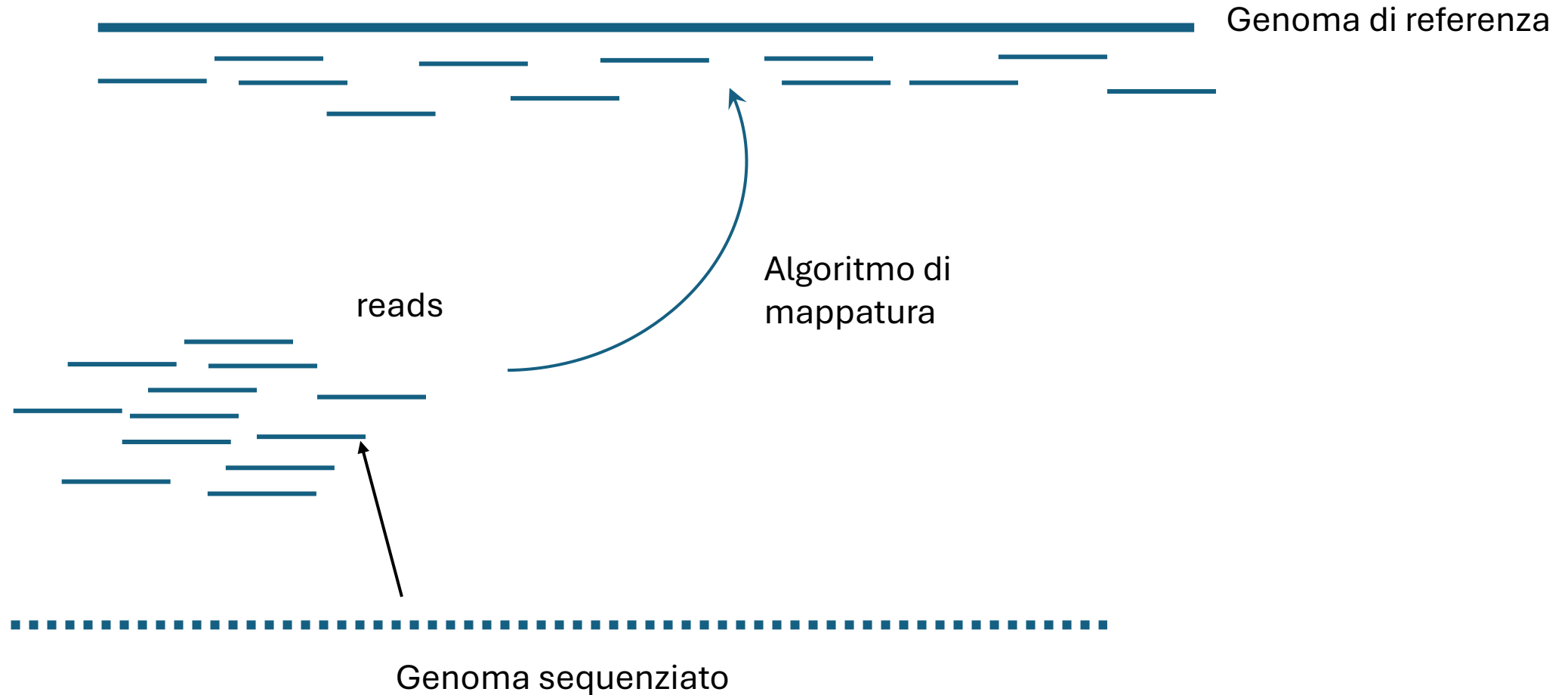
Caratterizzazione funzionale del genoma: L'annotazione

Genoma di referenza

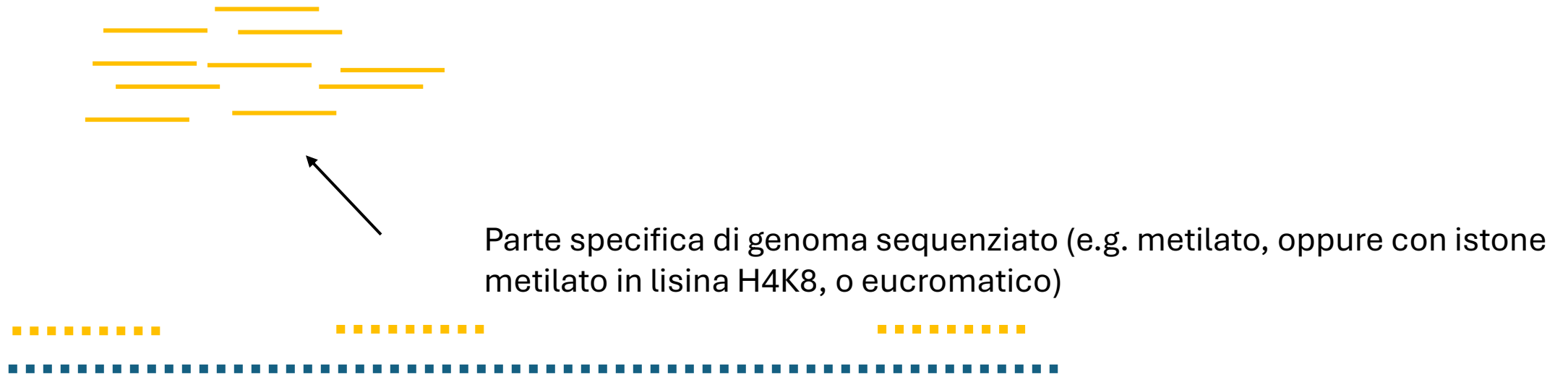
Caratterizzazione funzionale del genoma: L'annotazione



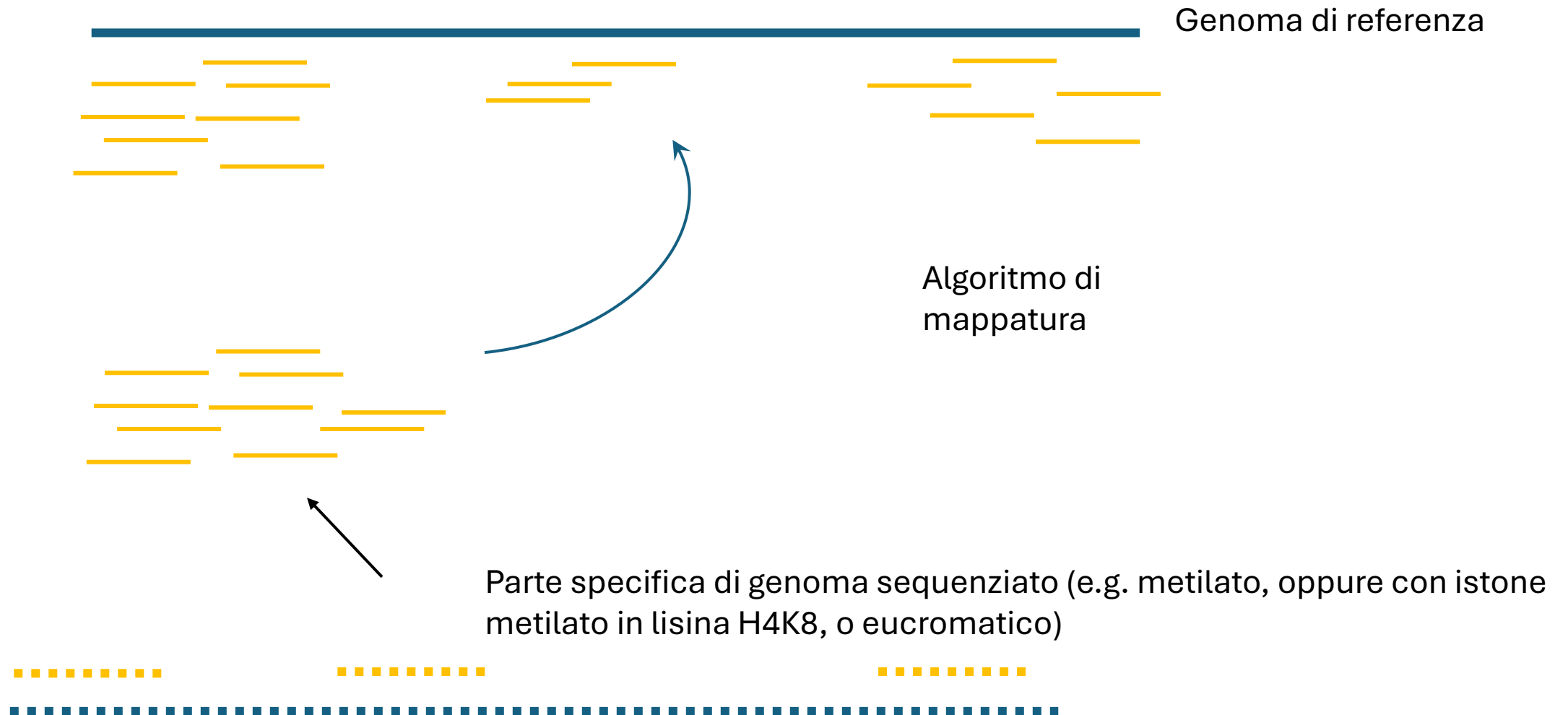
Mappatura delle sequenze e caratterizzazione funzionale del genoma



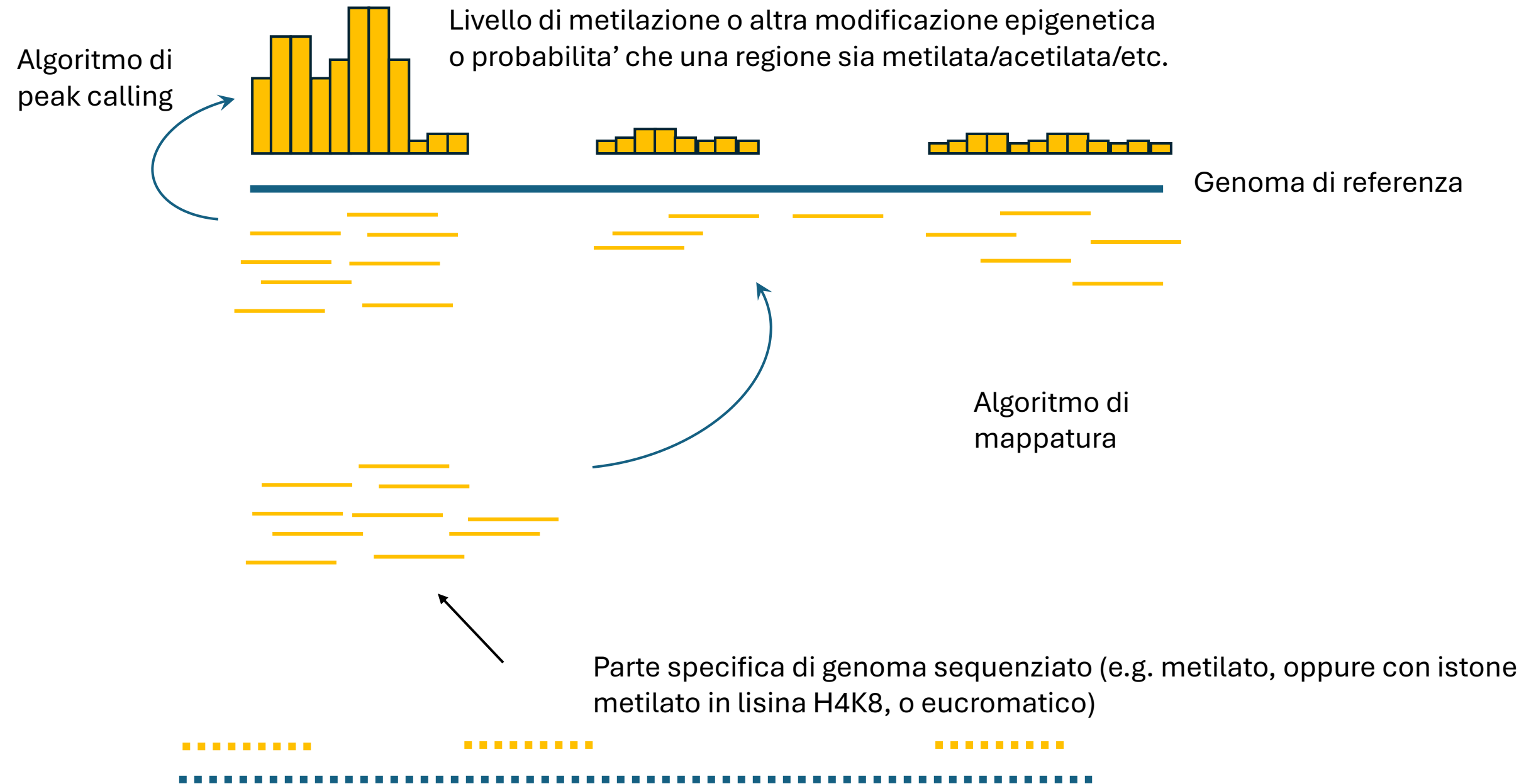
Peak calling per epigenomica



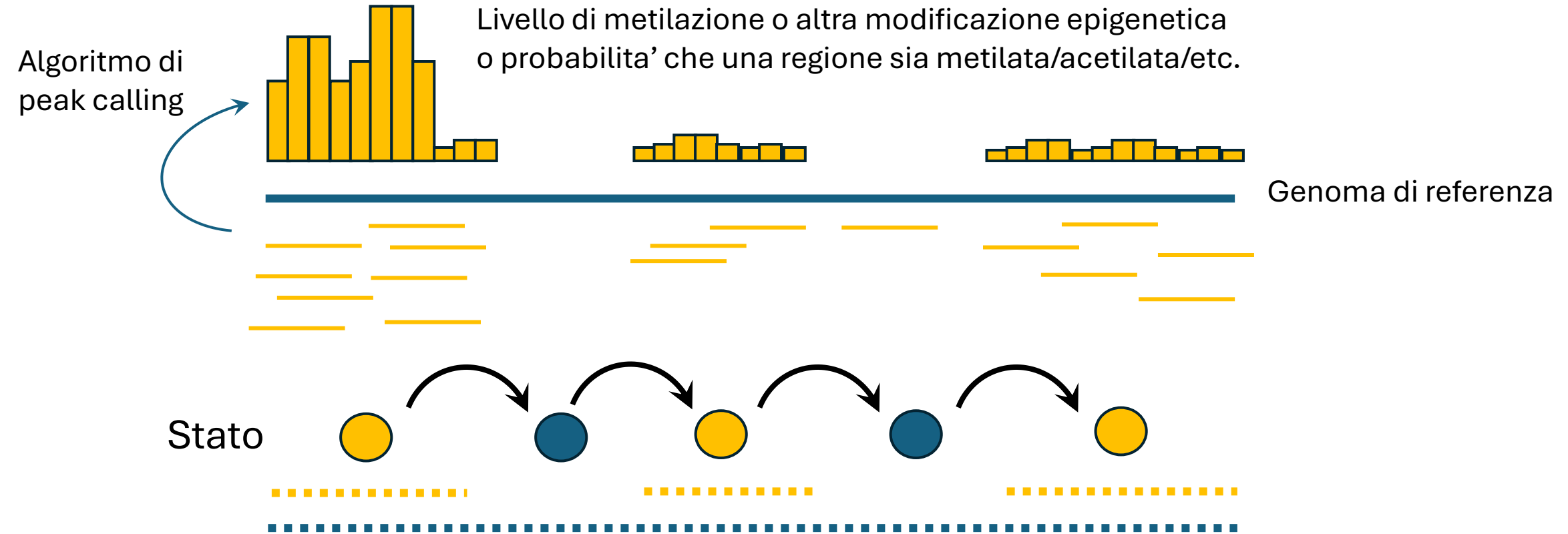
Peak calling per epigenomica



Peak calling per epigenomica



Peak calling per epigenomica

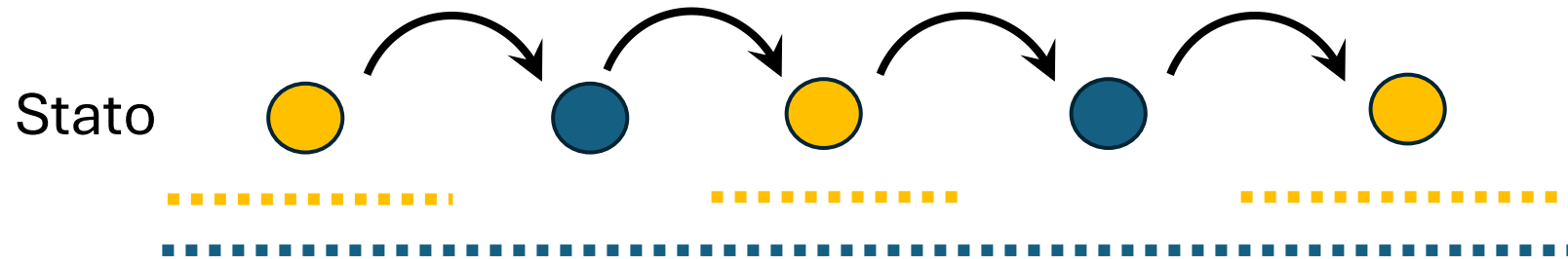


Esempio:

- Regione non-metilata
- Regione metilata

Hidden-Markov Models

Classe di algoritmi probabilistici utilizzatissimi in genomica in quanto rappresentano la naturale struttura del genoma in una serie di elementi che si susseguono linearmente



Esempio:

Regione non-metilata

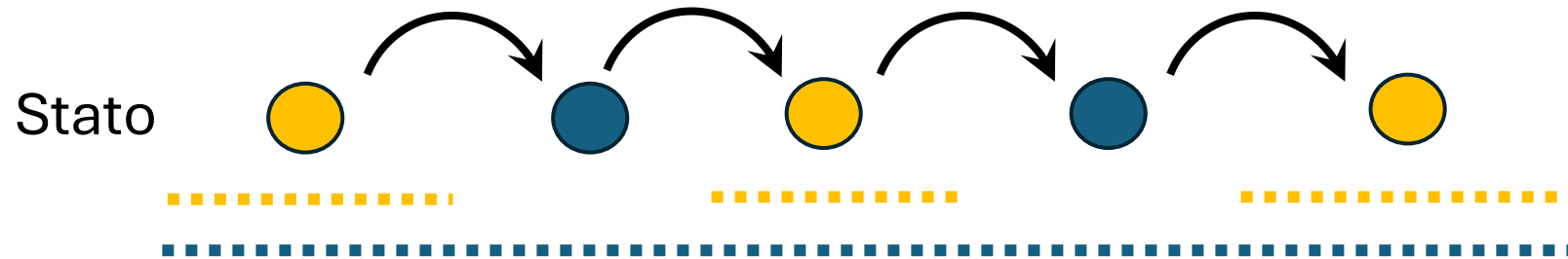


Regione metilata



Hidden-Markov Models

Classe di algoritmi **probabilistici** utilizzatissimi in genomica in quanto rappresentano la naturale struttura del genoma in una serie di elementi che si susseguono linearmente



Esempio:

Regione non-metilata



Regione metilata



Programma

- Operazioni di genomica computazionale di base su sequenze



- Assemblaggio
- Annotazione
- Mappatura e filtraggio delle reads



- Genotipizzazione, chiamata delle varianti strutturali e genomica di popolazione di base

- Algoritmi genomici avanzati



- Hidden Markov Model
 - Peak calling e suddivisione del genoma in contesi epigenomici (CHROMHMM)
 - Chromosome-painting
 - Identificazioni di zone funzionali del genoma sulla base della conservazione filogenetica
- Metagenomica



Classi di algoritmi

Categoria	Definizione	Esempi	Pro e cons
 Probabilistici	La casualità(randomness) fa parte dell'equilibrio.	Modelli binomiali per la genotipizzazione, MCMC, Hidden-Markov come CHROMHMM, PhastCons, etc.	C'è un po' di probabilità. Facilmente interpretabili
 Deterministici	Stesso input -> stesso output	Burrow-Wheeler Transform per la mappatura; De Bruijn Graphs per l'assemblaggio	Esatti e prevedibili.
 Euristici	Scorciatoie per problemi difficili	Filtraggio di varianti	Approssimati. Semplici ma sporchi.
 Machine-learning/AI	Data driven; eterogenei	DeepVariant; AlphaFold.	Potenti. Necessitano moltissimi dati. Poco interpretabili

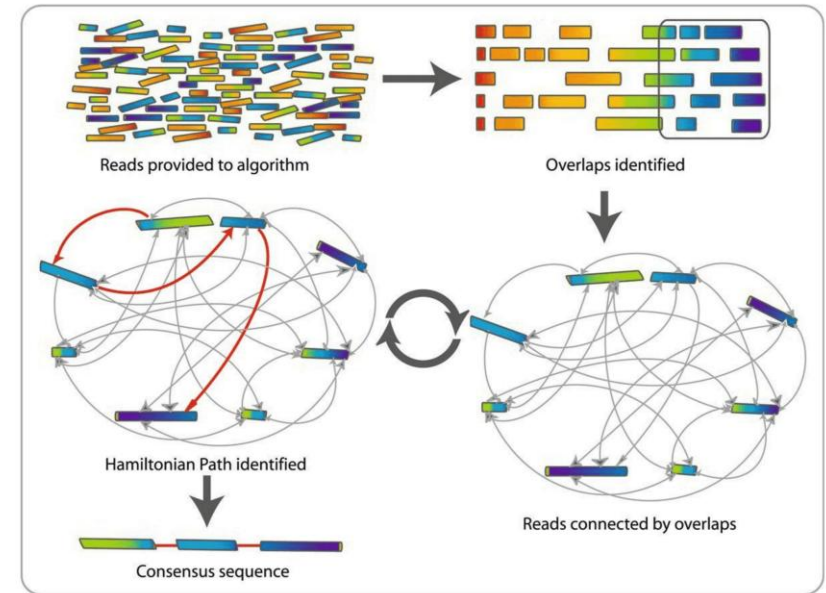
Ambiti biologici

Genomica

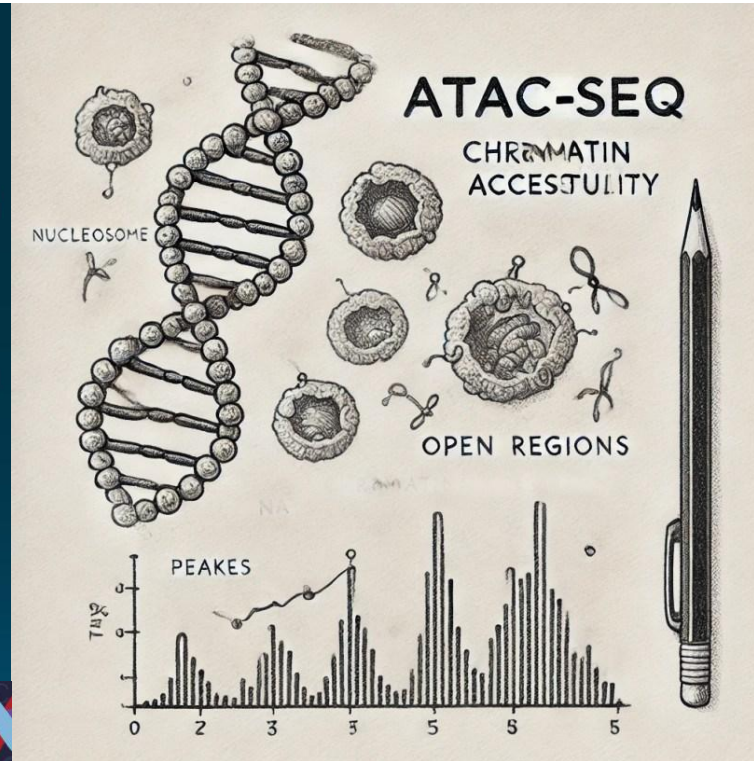
Epigenomica

Metodi

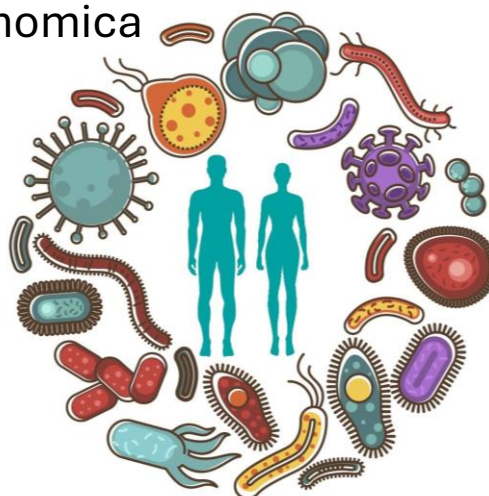
Algoritmi e statistica



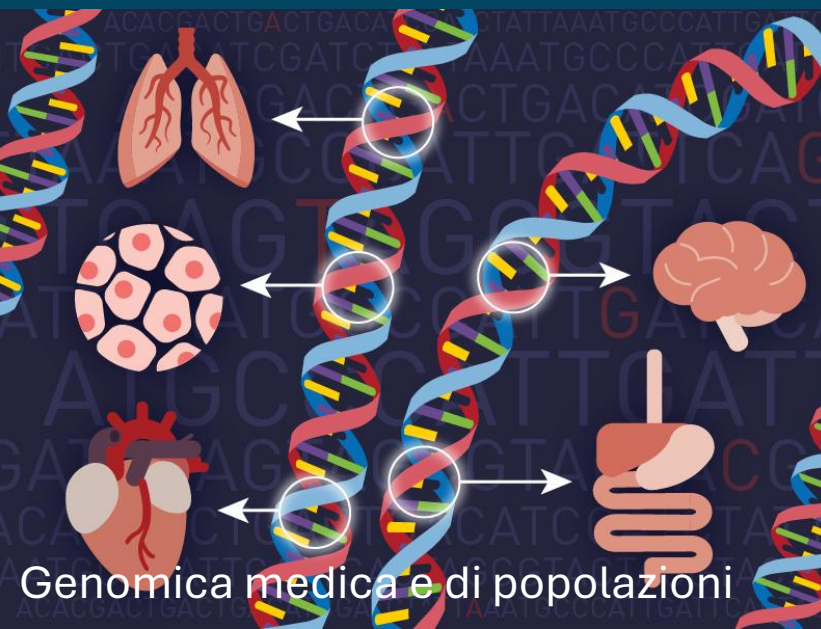
Terminale e programmi bioinformatici



Metagenomica



ASSEMBLAGGIO, ANNOTAZIONE E MAPPATURA



Genomica medica e di popolazioni

Glossario

- Assemblaggio/assembly: processo con il quale delle sequenze di DNA vengono utilizzate per inferire *de novo* la sequenza del genoma di provenienza
- Genoma di riferimento: la sequenza lineare di un genoma, potenzialmente utilizzata come punto di partenza per analisi bioinformatiche o popolazionistiche successive
- Annotazione: descrizione di un genoma che assegna caratteristiche funzionali (tipicamente geni e altre strutture) alle sue varie regioni
- Mappatura/mapping: processo con il quale delle sequenze di DNA vengono assegnate a delle coordinate rispetto ad un genoma di riferimento
- Genotipizzazione: processo con il quale vengono inferiti il genotipo di un individuo (quindi lo stato, e.g. eterozigote, omozigote, quale nucleotida) di un individuo

Preparazione dei laptop
alle esercitazioni di
Genomica
Computazionale



ubuntu



- Per tutto il corso i comandi sul terminal/linea di comando saranno in Courier New
- Useremo solo programmi e dati open e free-ware.
- Se qualcuno ha problemi di dispositivo fateci sapere!
- Per alleggerire le 4 ore del martedì avremo sempre una parte teorica ed una pratica. Il giovedì alterneremo, ma assumete che il laptop possa sempre essere utilizzato.
- Chatgpt va bene. Ma non perdetevi in comandi astrusi! E poca matematica ma chiave.

Installazione del terminal/linea di comando



Mac OS X



LINUX



Soluzione 1:

Andate direttamente su Store, oppure al seguente link

<https://apps.microsoft.com/detail/9pdxgncfsczv?hl=en-us&gl=US>

Microsoft | Store | Home | Apps | Games | About

Search apps, games, and more

 **Ubuntu**
Canonical Group Limited
4.4 ★ | 683 ratings | [Developer tools](#)

Install a complete Ubuntu terminal environment in minutes with Windows Subsystem for Linux (WSL). Develop cross-platform applications, improve you...

[Download](#) [View in Store](#)

 **EVERYONE**

Screenshots

Discover more >

 **Ubuntu 22.04.5 LTS**
4.2 ★ | [Developer tools](#) | Free

Windows 11 Soluzione 2:

Aprirete la powershell come amministratori

Digitate

```
wsl --install
```

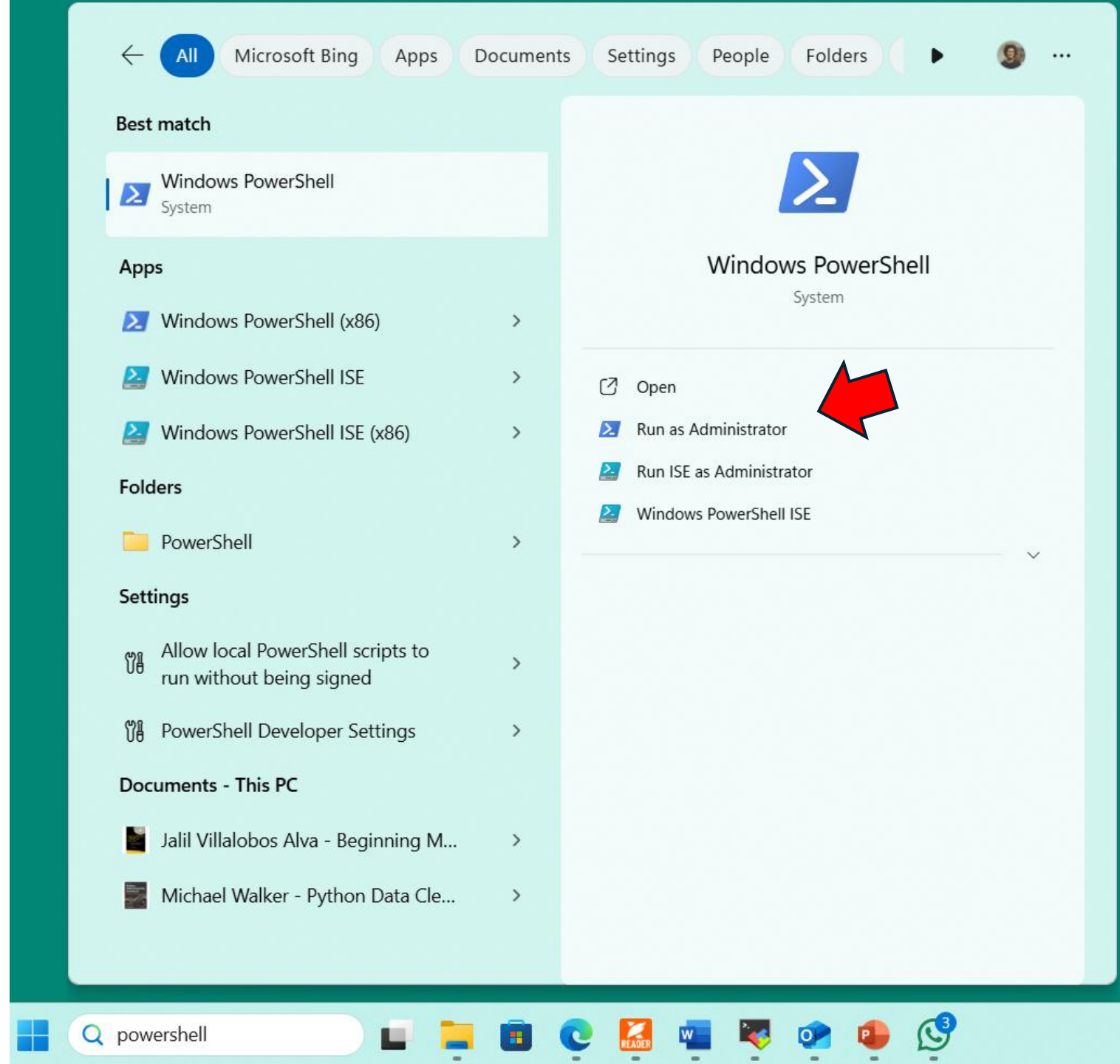


YouTube · Tricknology

How To Install Ubuntu in Windows 11 Using Subsystem For Linux

In this video tutorial, I will show you guys how to install Ubuntu in Windows 11 using Windows Subsystem for Linux. Enable WSL 1:...

111.9K views · 1 Sep 2021



Soluzione 3:

Premete la combinazione di tasti **Win+R**, e scrivete `optionalfeatures`

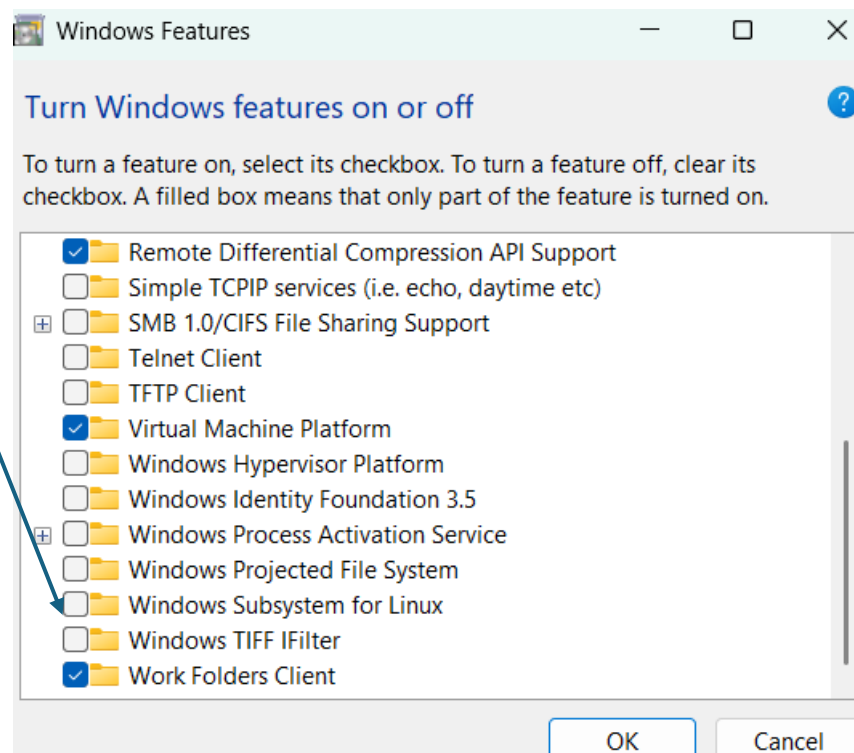
Selezionate

Windows Subsystem for Linux
Virtual Machine Platform

Riavviate il computer.

Usate le soluzioni suggerite per Windows 11, ad esempio, tramite Powershell

```
wsl -install
```



Ora avrete un computer che di fatto
contiene un secondo sistema
operativo, Ubuntu!

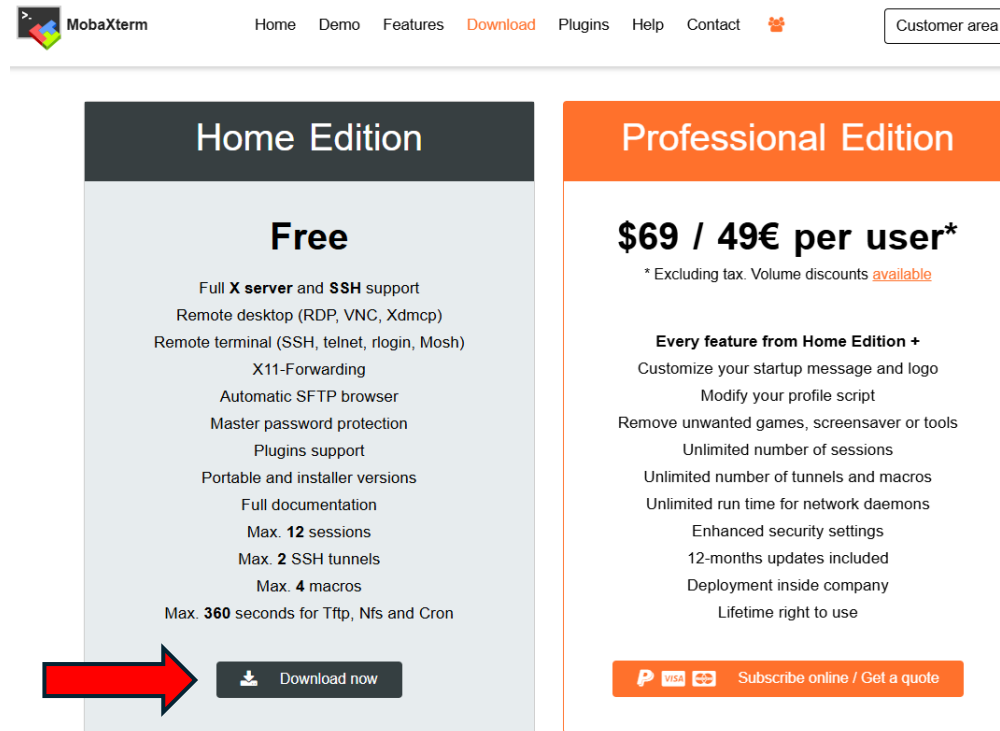
Se qualcuno ha versioni piu antiche di Windows fatemi sapere!



ubuntu

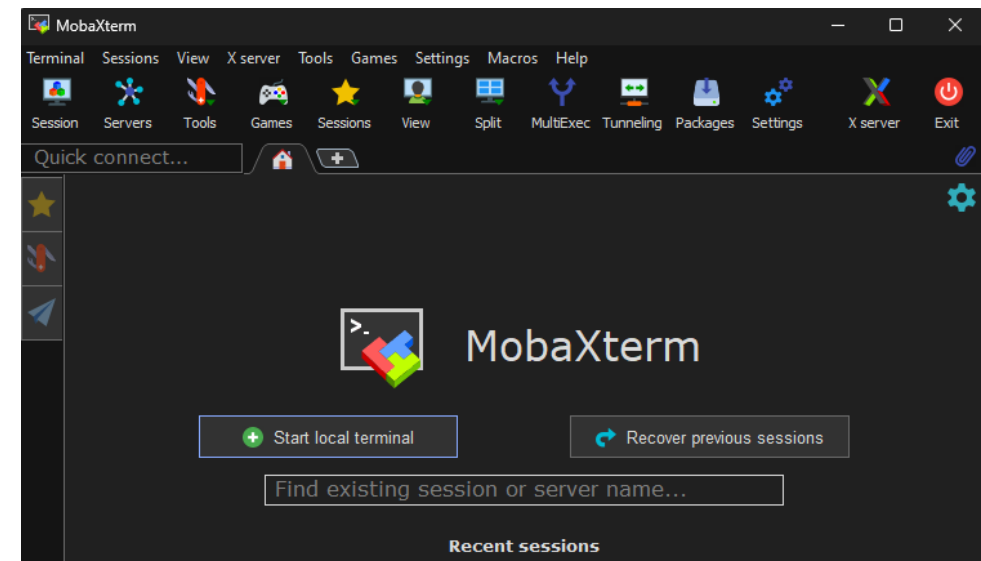
Installazione di MobaXterm (consigliato)

- Andate sul sito <https://mobaxterm.mobatek.net/download.html>



The screenshot shows the MobaXterm website with two main sections: Home Edition and Professional Edition. The Home Edition is labeled 'Free' and lists features like Full X server and SSH support, Remote desktop (RDP, VNC, Xdmcp), Remote terminal (SSH, telnet, rlogin, Mosh), X11-Forwarding, Automatic SFTP browser, Master password protection, Plugins support, Portable and installer versions, Full documentation, Max. 12 sessions, Max. 2 SSH tunnels, Max. 4 macros, and Max. 360 seconds for Tftp, Nfs and Cron. A red arrow points to the 'Download now' button. The Professional Edition is labeled '\$69 / 49€ per user*' and lists features like Every feature from Home Edition +, Customize your startup message and logo, Modify your profile script, Remove unwanted games, screensaver or tools, Unlimited number of sessions, Unlimited number of tunnels and macros, Unlimited run time for network daemons, Enhanced security settings, 12-months updates included, Deployment inside company, and Lifetime right to use. A 'Subscribe online / Get a quote' button is at the bottom.

MobaXterm aggiunge delle funzionalità utili per la grafica e per connettervi a server. Una volta installato cercate una icona simile per avviarlo:



Download MobaXterm Home Edition (current version):

 MobaXterm Home Edition v25.2 (Portable edition)



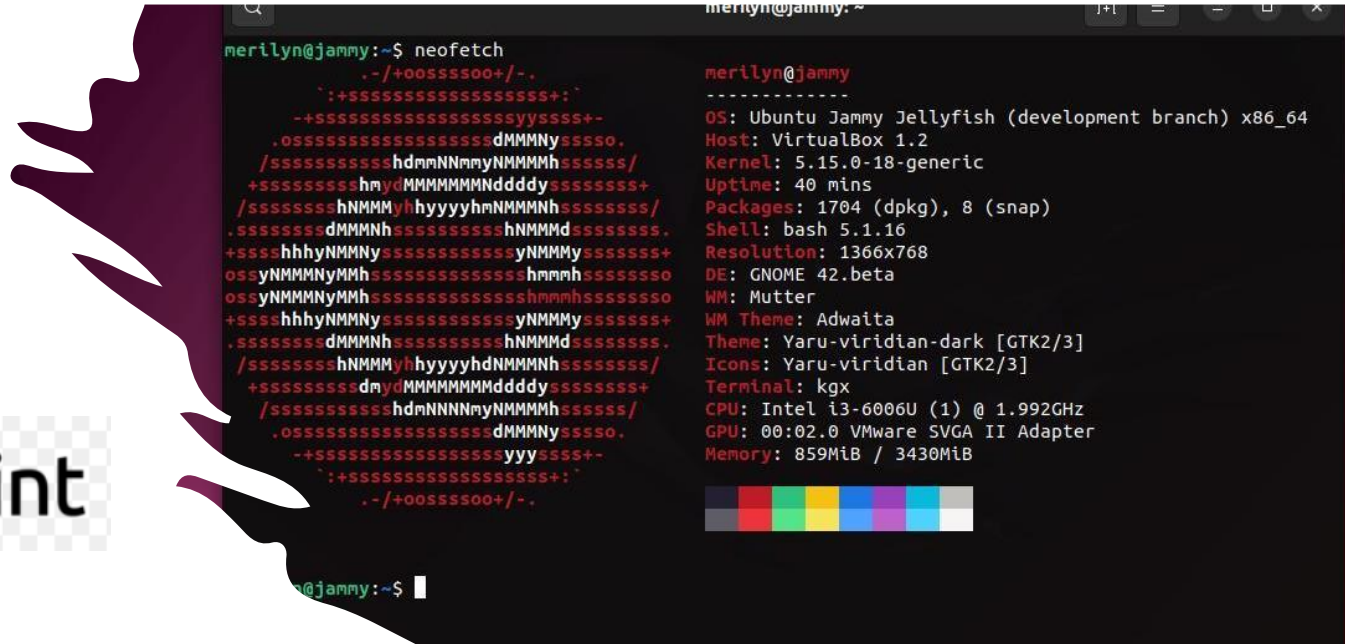
 MobaXterm Home Edition v25.2 (Installer edition)

Preparazione per utenti



LINUX

- Se avete Ubuntu/Debian semplicemente installate un terminale a vostra scelta
- `sudo apt install tilix`
- `sudo apt install terminator`
- Se avete altri sistemi (davvero???) fatemi sapere





Avete già installato **Terminal.app** di default. Per aprirlo:

- ⌘ (Command) + Spazio e poi cercate **Terminal**
- Oppure da Applicazioni, poi Utilities, poi Terminal

Se volete piu comfort (ma non una grande differenza) potete installare **iTerm2**.

NOTA: la maggior parte dei comandi sarà identica tra utenti Mac e gli altri, ma piccole differenze emergeranno per alcuni comandi. Specialmente durante l'installazione.

Per installare i programmi, vi consiglio di usare Homebrew, che vi permette di installare programmi più semplicemente.

Per scaricarlo, andare sul Terminale e poi:

```
/bin/bash -c "$(curl -fsSL
https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"
echo 'eval "$(/opt/homebrew/bin/brew shellenv)"' >> ~/.zprofile
eval "$(/opt/homebrew/bin/brew shellenv)"
brew --version
brew doctor
```

Potete poi installare i vari comandi con:

```
brew install wget curl git tree nano vim gnuplot
```

```
brew install coreutils gnu-sed grep #questo renderà i vostri comandi più simili a quelli degli altri utenti
```



Aprite Ubuntu/Terminal/MobaXterm

```
echo "Bravissimo/a $( whoami ) , stai usando la riga di comando !"
```





Installazione
programmi



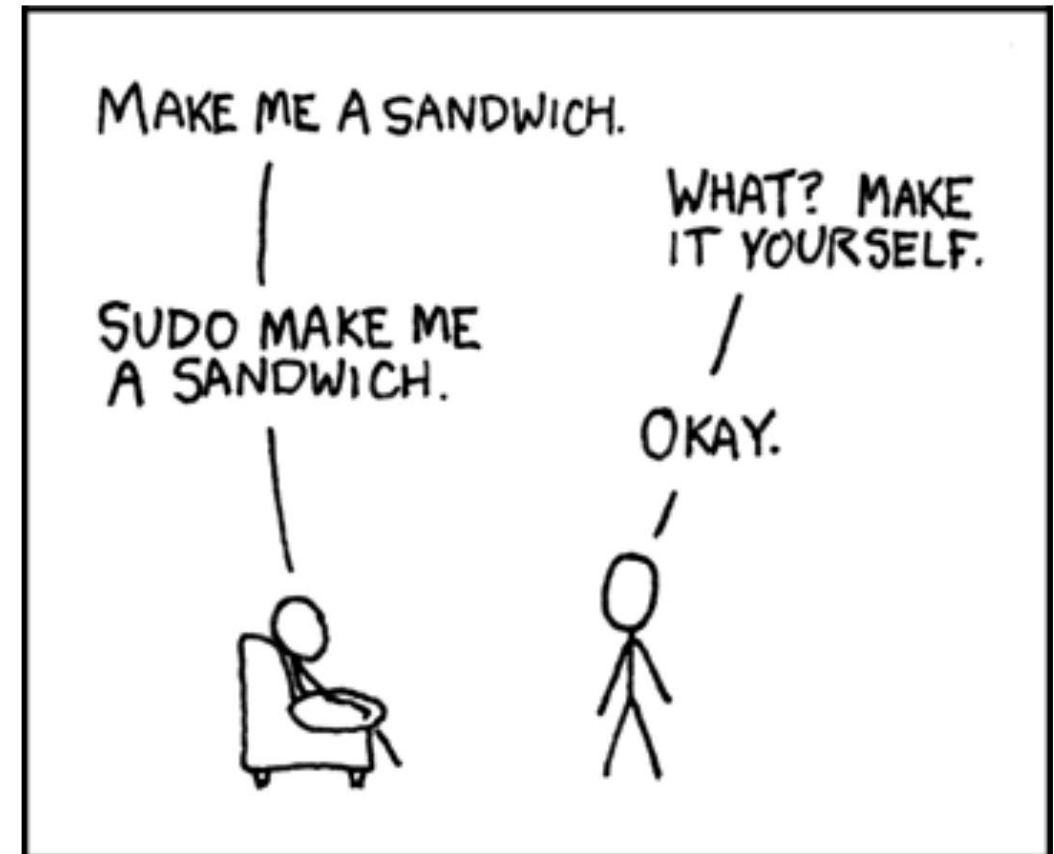
I hate programming
I hate programming
I hate programming
It works!
I love programming

- `sudo` + comando vi permette di eseguire qualsiasi comando come amministratore. Ma dovete ricordare la vostra password!
- `apt` è il comando di base per installare programmi su ubuntu

```
sudo apt update && sudo apt -y upgrade
sudo apt -y install build-essential
sudo apt -y install git curl wget ca-
certificates nano
sudo apt -y install vim less unzip tar
sudo apt -y install grep gawk sed coreutils
diffutils
sudo apt -y install samtools bcftools
bedtools tabix
sudo apt -y install fastqc bwa
sudo apt -y install bowtie2
sudo apt -y install r-base r-base-dev
```



LINUX



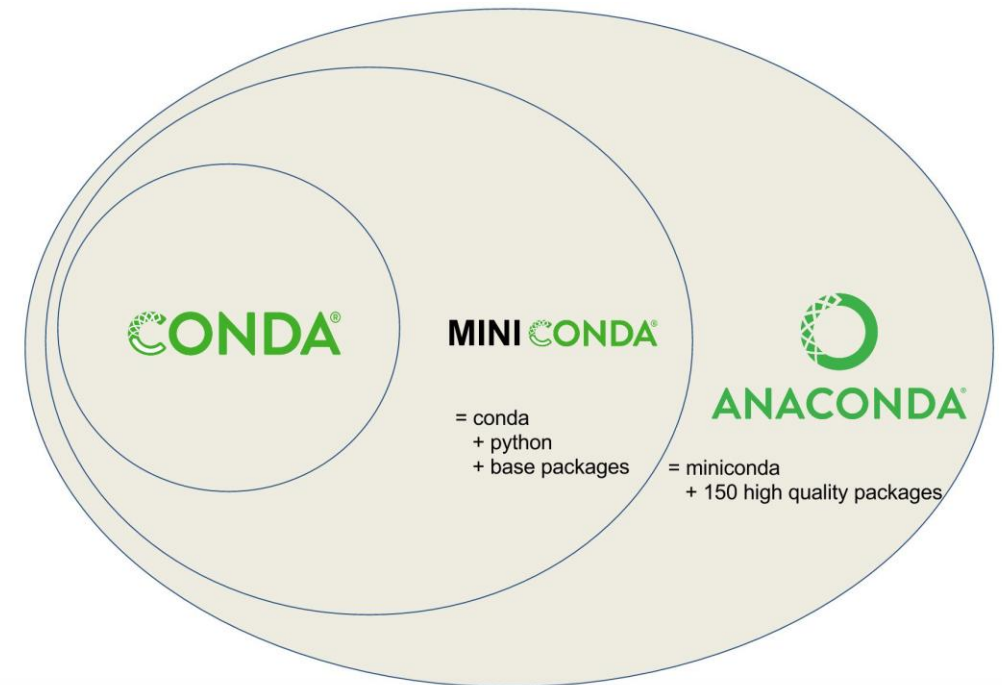


Macintosh

```
brew install samtools bcftools bedtools htslib bwa bowtie2 fastqc  
brew install --cask r
```

Per tutti (Mac, Windows, Ubuntu, Linux vari e stranezze varie)

- Conda è un sistema di «manager» di pacchetti, che vi permette di installare facilmente vari programmi e creare «ambienti» dentro il quale lavorare.
- Una volta installato, stessa procedura per ogni piattaforma.



Installazione di miniconda su mac

- `curl -fL -o Miniconda.sh
https://repo.anaconda.com/miniconda/Miniconda3-
latest-MacOSX-arm64.sh`
- `# Intel Macs (x86_64):`
- `# curl -fL -o Miniconda.sh
https://repo.anaconda.com/miniconda/Miniconda3-
latest-MacOSX-x86_64.sh`
- `bash Miniconda.sh -b`
- `~/miniconda3/bin/conda init zsh`
- `exec zsh`



Installazione di miniconda su linux/windows

- `sudo apt update && sudo apt -y install curl bzip2`
- `curl -fL -o Miniconda.sh https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh`
- `bash Miniconda.sh -b`
- `~/miniconda3/bin/conda init bash`
- `exec bash`



Installazione programmi usando miniconda (per tutti)

- Configurazione dei canali (una volta “per sempre”)

```
conda config --set channel_priority strict
```

```
conda config --add channels conda-forge
```

```
conda config --add channels bioconda
```

- Creazione di un ambiente per il corso

```
conda create -n GenomicaComputazionale -y samtools  
bedtools htlib bcftools bwa bowtie2
```

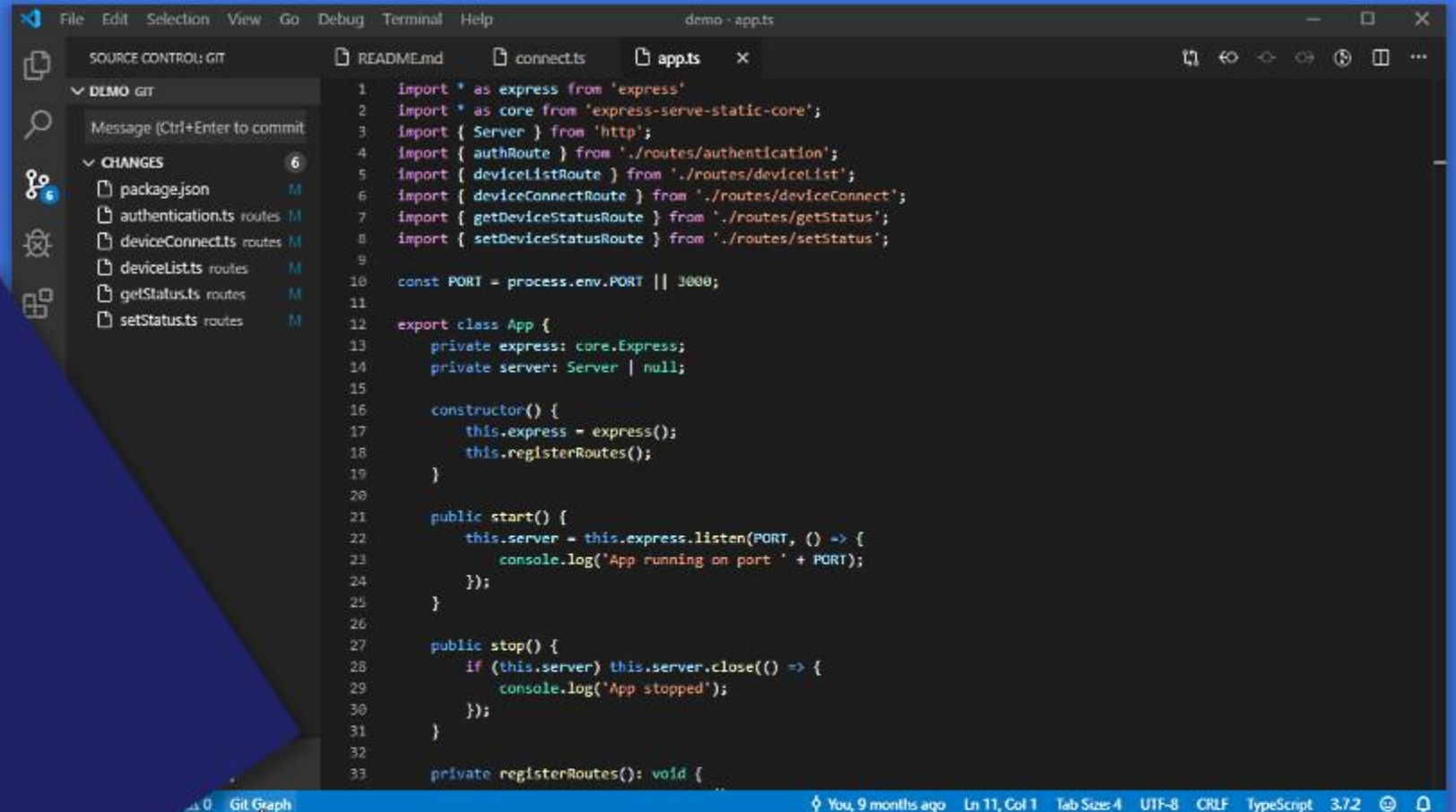
```
conda activate GenomicaComputazionale
```

```
conda install -n genomics101 -y r-base r-essentials
```

```
conda deactivate
```



E ora l'editor (<https://code.visualstudio.com/>)



Visual Studio Code

Se avete mac o ubuntu e vi siete gia' innamorati della linea di comando

- Mac

- `brew install --cask visual-studio-code`

- Ubuntu

- `sudo apt -y install wget gpg`
 - `wget -qO- https://packages.microsoft.com/keys/microsoft.asc | gpg -dearmor > packages.microsoft.gpg`
 - `sudo install -o root -g root -m 644 packages.microsoft.gpg /etc/apt/trusted.gpg.d/`
 - `sudo sh -c 'echo "deb [arch=amd64,arm64,armhf] https://packages.microsoft.com/repos/code stable main" > /etc/apt/sources.list.d/vscode.list'`
 - `sudo apt update`
 - `sudo apt -y install code`

Ma se preferite potete usare qualsiasi altro (Kate, gedit, Eclipse, etc.)

Menzione d'onore per vim:

```
(base) fabrizio@asusfabri:~$  
(base) fabrizio@asusfabri:~$ vim youarebeautiful.sh
```

Premete *i* e scrivete

```
#!/bin/bash  
YOURNAME=$( whoami )  
echo "You are beautiful, in every single way, and today $( date ) dear ${YOURNAME} you wrote your first program"  
  
~  
~  
-- INSERT --
```

Poi uscite dall'INSERT mode con ESC e scrivete **:wq**

```
(base) fabrizio@asusfabri:~$ chmod +x youarebeautiful.sh  
(base) fabrizio@asusfabri:~$ ./youarebeautiful.sh
```

Complimenti, il vostro computer è pronto! E' ora un piccolo laboratorio per fare analisi e ricerca di genomica computazionale!



Potenziati problemi

- **Le parentesi graffe {}** sulla tastiera italiana si digitano con Shift+AltGr+[e Shift+AltGr+]. Se avete problemi su mobaXterm, usate

```
echo '"\C-o": "{ "' >> ~/.inputrc
```

```
echo '"\C-p": "}" "' >> ~/.inputrc
```

e così potrete inserirle con Ctrl+o e Ctrl+p

- **Se installando Ubuntu su Windows ottenete l'errore**

Error: 0x80370114 Non è stato possibile avviare l'operazione perché una funzionalità richiesta non è installata.

o simili, seguite la Soluzione 3 delle slides (come per Windows 10) e poi proseguite sull'installazione: dovete prima abilitare alcune funzioni per abilitare le Virtual Machines.

Potenziati problemi

- **Copia e incolla non funziona**

- ◆ **Windows**

- 1. MobaXterm**

- **Copia** → selezionare col mouse (va in automatico negli appunti).
 - **Incolla** → `Shift+Insert` oppure tasto destro → *Paste*.
 - Si può però settare per `Ctrl+C` e `Ctrl+V`

- 2. WSL (Linux su Windows)**

- Se aperto in Windows Terminal → usare le scorciatoie sopra (`Ctrl+Shift+C/V`).
 - Se aperto in vecchio CMD/PowerShell → serve QuickEdit Mode.

- ◆ **macOS (Terminale o iTerm2)**

- **Copia** → `Cmd+C`
 - **Incolla** → `Cmd+V`

- ◆ **Linux (Ubuntu, Debian, ecc.)**

- **Copia** → `Ctrl+Shift+C`
 - **Incolla** → `Ctrl+Shift+V`
 - (NB: `Ctrl+C` da solo interrompe il programma, non copia!)

Potenziati problemi

- **Siete finiti in root (/) e non avete username**

Può accadere se durante l'installazione di WSL non scegliete di creare un utente

In questo caso l'utente entra direttamente come **root** (non ha un account non privilegiato).

Non è un problema, ma non è consigliato: meglio avere un account “normale” e usare sudo solo quando serve.

👉 Soluzione: creare un nuovo utente e renderlo predefinito. Ad esempio in Ubuntu/WSL:

dentro la sessione root

```
adduser studente
```

```
usermod -aG sudo studente
```

Poi, da **PowerShell** in Windows:

```
wsl -d Ubuntu -u studente
```

Oppure rendere permanente:

```
ubuntu config --default-user studente
```

Potenziati problemi

- **Avete scelto uno username indesiderato (esempio wsl-install) e/o avete dimenticato la password**

Se siete su Windows: L'account esiste, ma senza password non si può usare sudo (permessi da amministratore).

Bisogna **resettare la password** entrando come root. Avviare WSL come root da Powershell:

```
wsl -d Ubuntu -u root
```

Dentro, cambiare la password dell'utente:

```
passwd wsl-install
```

→ inserire nuova password (due volte).

Riavviare WSL normalmente:

```
wsl -d Ubuntu -u wsl-install
```

Se su mac, resettate la password con:

```
sudo passwd wsl-install
```



WE ARE HERE TO HELP!

IMPORTANTE

Se avete problemi persistenti (o siete su un tablet con ssh) possiamo darvi temporaneamente accesso sul server del dipartimento. Avvertiteci!