



# BIOINFORMATICA

## II

*ALLINEAMENTO TRASCRITTOMA*

# Argomenti



**01** TRASCRITTOMA E  
RNA-SEQ

**02** ALLINEAMENTO

**03** SOFTWARE DI  
ALLINEAMENTO





01

# TRASCRIPTOMA E RNA-SEQ



# TRASCRIPTOMA



La **trascrizione** è il processo mediante il quale le informazioni genetiche contenute nel DNA sono copiate in molecole di RNA

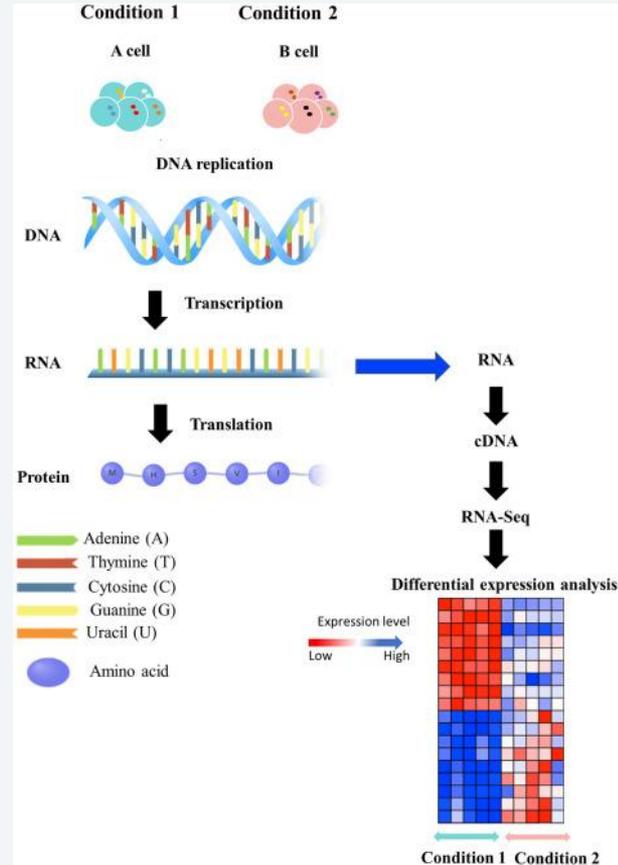
Le molecole di RNA, chiamate **trascritti**, svolgono un ruolo cruciale nella sintesi delle proteine e in molti altri processi biologici

**Il trascrittoma rappresenta l'insieme di tutti i trascritti presenti in una cellula o in un organismo in un momento stabilito**

Comprendere il trascrittoma è fondamentale per comprendere come vengano regolati i geni e come le cellule rispondano a cambiamenti ambientali o a malattie



# TRASCRITTOMA



# RNA-SEQ: Definizione



**RNA-seq (RNA sequencing)** è una tecnica avanzata di sequenziamento utilizzata per analizzare in modo globale e quantitativo l'RNA presente in un campione biologico. È una delle tecnologie chiave per studiare l'espressione genica e le dinamiche trascrizionali in diversi tipi di cellule, tessuti o condizioni

Il livello di trascrizione varia da gene a gene e dipende dal contesto biologico (come il tipo di cellula), lo stato di sviluppo o le condizioni ambientali



# RNA-SEQ: Procedura



## **Isolamento dell'RNA**

*Si estrae l'RNA totale da un campione biologico, come cellule, tessuti o organismi*

*In molti casi, si arricchisce il campione per specifici tipi di RNA (ad esempio, solo mRNA rimuovendo rRNA, o si analizzano RNA non codificanti)*

## **Preparazione del cDNA**

*Poiché le tecnologie di sequenziamento leggono il DNA, l'RNA è convertito in cDNA (DNA complementare) attraverso la trascrizione inversa*

## **Frammentazione e costruzione della libreria:**

*L'RNA o il cDNA è frammentato in sequenze più brevi*

*Si aggiungono adattatori specifici alle estremità dei frammenti, necessari per il sequenziamento*



# RNA-SEQ: Procedura



## **Sequenziamento:**

*I frammenti di cDNA sono sequenziati utilizzando piattaforme come Illumina, PacBio o Oxford Nanopore.*

*Ogni frammento sequenziato genera una "read", cioè una sequenza di nucleotidi letta dalla macchina*

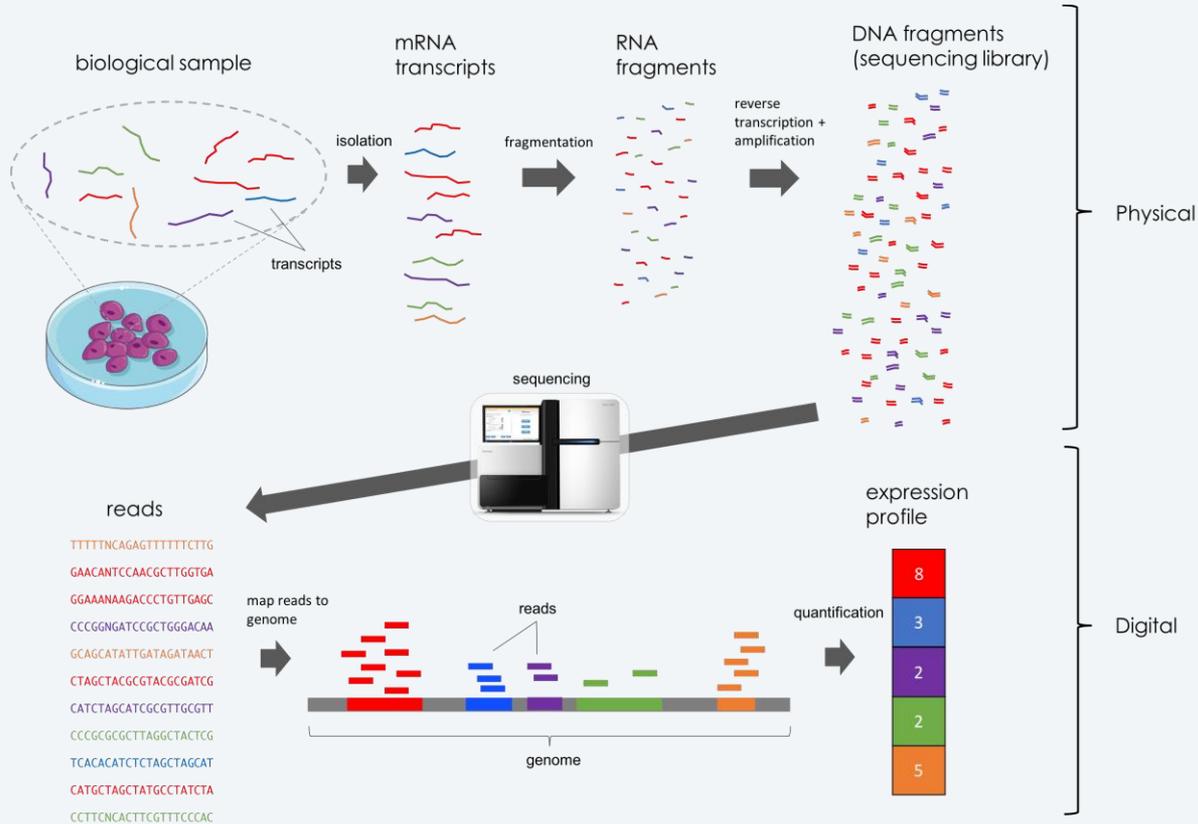
## **Allineamento e analisi dei dati:**

*Le read sono allineate a un genoma di riferimento (se disponibile) o assemblate de novo*

*L'analisi fornisce informazioni quantitative (quanti trascritti di ogni gene sono presenti) e qualitative (quali varianti di trascritti sono presenti, identificazione di nuovi RNA)*



# RNA-SEQ: Procedura



# RNA-SEQ: Possibili problemi



**Bassa qualità del RNA:** La degradazione dell'RNA o la contaminazione da DNA possono influenzare negativamente la qualità dei dati. È quindi fondamentale valutare la qualità dell'RNA, ad esempio attraverso il valore RIN (RNA Integrity Number), prima di procedere con la preparazione della libreria

**Bias nella preparazione della libreria:** La conversione di RNA in cDNA e l'amplificazione possono introdurre rumore (*bias*), favorendo alcune sequenze rispetto ad altre

*Ad esempio, l'efficienza della trascrittasi inversa può variare a seconda della sequenza e della struttura dell'RNA*

**Normalizzazione:** Poiché il numero totale di reads può variare da un campione all'altro, è necessario normalizzare i dati tra le diverse librerie per rendere comparabili i livelli di espressione



# RNA-SEQ: Possibili problemi



**Mappatura dei reads:** La mappatura dei reads al genoma di riferimento può presentare sfide, in particolare in regioni del genoma con duplicazioni, ripetizioni o alta omologia tra i geni

**Espressione di geni a basso livello:** I geni espressi a livelli molto bassi possono essere difficili da rilevare e analizzare, in quanto la copertura potrebbe essere insufficiente per fornire una stima affidabile dell'espressione

**Isoforme e splicing alternativo:** L'RNA-seq può rilevare diverse isoforme di un gene, ma determinare l'espressione di singole isoforme può essere complesso a causa della sovrapposizione di reads tra diverse varianti di isoforme



# RNA-SEQ: Possibili problemi



**Interferenza da trascritti non codificanti:** L'RNA-seq può rilevare non solo mRNA, ma anche una varietà di RNA non codificanti, che potrebbero complicare l'analisi se non si è interessati a questi trascritti.

**Variazione tecnica e biologica:** Come menzionato in precedenza, è cruciale distinguere la variazione dovuta a differenze reali nell'espressione genica da quella causata da rumore tecnico o variazione biologica casuale.

**Risoluzione temporale:** Gli esperimenti di RNA-seq rappresentano uno "snapshot" dell'espressione genica in un dato momento. La dinamica dell'espressione genica nel tempo potrebbe richiedere esperimenti di sequenziamento ripetuti, multipli e complessi



# RNA-SEQ: Possibili analisi



## ❖ Misurare l'espressione genica

*Determina quali geni sono attivi (trascritti in RNA) e quantifica il loro livello di espressione*

## ❖ Identificare nuove varianti di trascritti

*RNA-seq può rilevare splicing alternativo, nuovi trascritti o isoforme geniche*

## ❖ Caratterizzare RNA non codificanti

*Identifica e quantifica RNA che non codificano per proteine, come miRNA, lncRNA, e altri RNA regolatori*

## ❖ Analizzare cambiamenti in condizioni diverse

*Confronta l'espressione genica tra campioni, ad esempio tra tessuti sani e malati, o prima e dopo un trattamento*

## ❖ Indagare eventi di editing dell'RNA

*Può rilevare modifiche a livello delle sequenze di RNA rispetto al DNA originale*



# RNA-SEQ: Possibili analisi



## ❖ Mappatura delle read ai geni

Le read sequenziate sono allineate (mappate, *mapping*) al genoma di riferimento per identificare da quale gene provengono. Più un gene è espresso, più copie del suo mRNA saranno presenti nella cellula e, di conseguenza, più read corrispondenti verranno rilevate durante il sequenziamento

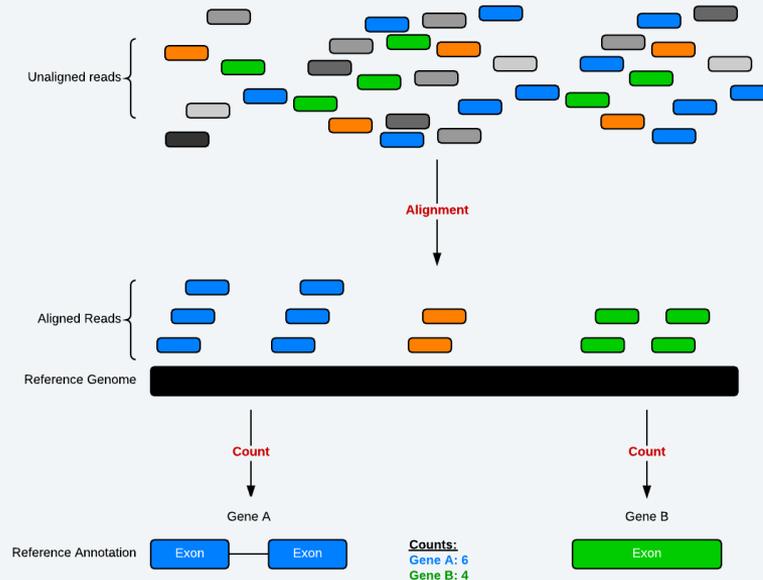
## ❖ Proporzionalità con il livello di espressione

Il numero di read che si allineano a un gene riflette il livello di espressione del gene, perché è proporzionale alla quantità di mRNA prodotto. Ad esempio, se un gene è molto attivo, produrrà più mRNA e quindi genererà più read mappate; se un gene è poco o per nulla attivo, il numero di read sarà molto basso o nullo



# RNA-SEQ: Comportamento reads

Le reads tendono ad agglomerarsi in aree di interesse comune per motivi intrinseci legati alla biologia del campione, al protocollo sperimentale, e al metodo di sequenziamento



# RNA-SEQ: Comportamento reads



I motivi del raggruppamento sono:

## 1. Esprimibilità differenziale dei geni

Non tutti i geni sono trascritti in quantità uguali. Alcuni geni sono più espressi e producono molti trascritti, mentre altri sono meno espressi o completamente silenziati.

Le reads si accumulano nei tratti di geni che sono attivamente trascritti nel campione analizzato.

Ad esempio, un campione di tessuto epatico mostrerà un forte agglomerato di reads attorno ai geni coinvolti nel metabolismo, come ALB (albumina).

## 2. Struttura dei trascritti

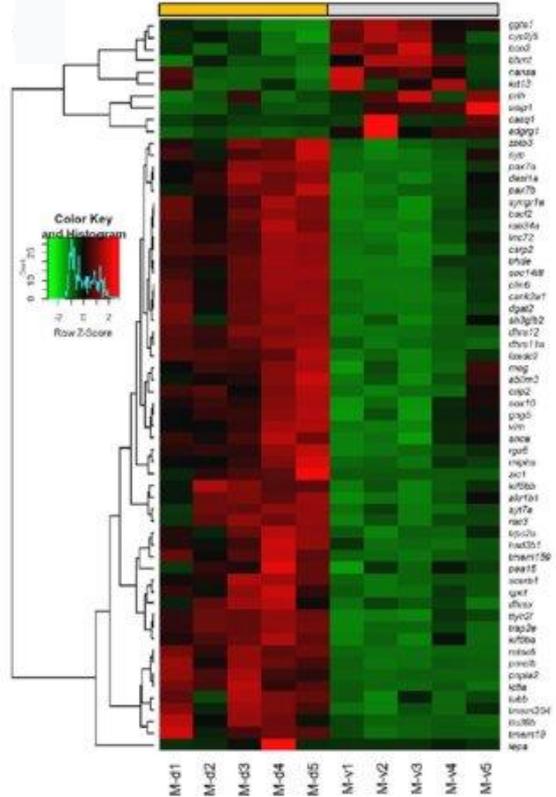
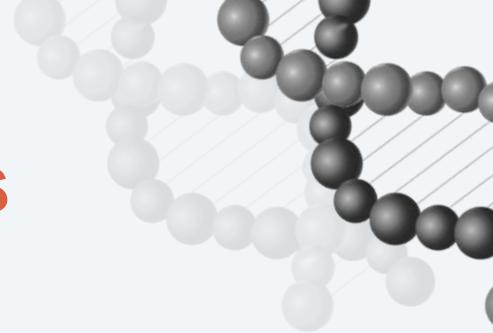
Le reads derivano dalle regioni esone di un trascritto, poiché gli introni vengono rimossi durante lo splicing.

Di conseguenza, le reads si agglomerano nei tratti corrispondenti agli esoni del gene.

Nei casi di splicing alternativo, le reads possono essere distribuite in modo non uniforme, riflettendo le isoforme prevalenti.



# RNA-SEQ: Comportamento reads

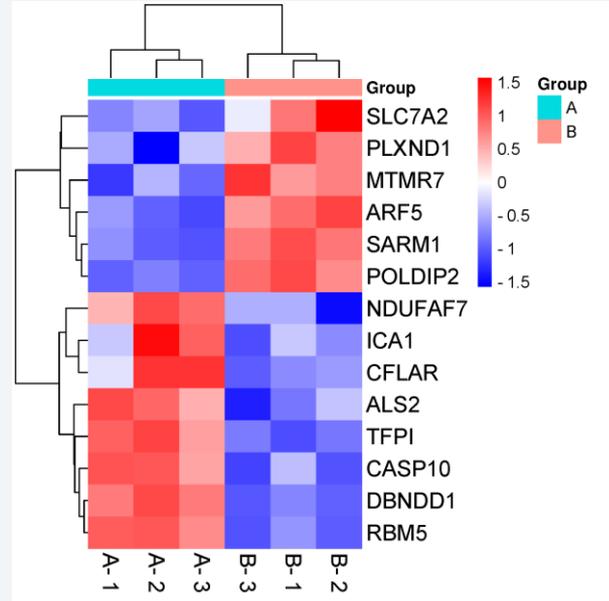


*T. dubois* Maswa  
dorsal/ventral

VERDE: Up-regolated  
ROSSO: Down-regolated

GENI

CAMPIONI



# RNA-SEQ: Comportamento reads



## 3. Caratteristiche del protocollo RNA-Seq

Errori in sequenziamento (Bias di copertura) o sequenziamento mirato (in alcuni esperimenti (es., RNA-Seq mirato), si utilizzano sonde per arricchire trascritti di interesse, creando agglomerati di reads in regioni specifiche)

## 4. Sequenze ripetitive o regioni ad alta affinità

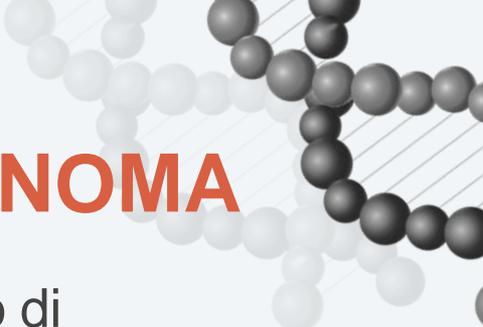
Alcune regioni del genoma, come quelle altamente ripetitive, possono attirare reads in modo spurio o non uniforme a causa di artefatti del mapping o di problemi di unicità delle sequenze. Esempio: Trascritti derivanti da geni delle famiglie ribosomiali o elementi trasponibili possono mostrare falsi agglomerati.

## 5. Interesse biologico specifico

L'accumulo di reads in certe regioni può rappresentare processi biologici importanti: RNA instabile o degradato o Regolazione post-trascrizionale.



# ALLINEAMENTO RNA-seq SU GENOMA

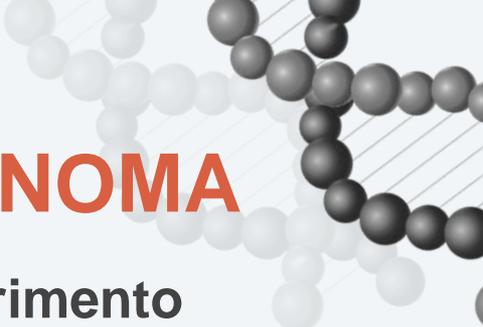


Il mapping di dati RNA-seq sul genoma è il processo di allineamento delle sequenze di RNA ottenute mediante sequenziamento ad alto rendimento (**RNA-Seq**) sul genoma di riferimento

Questo passaggio è cruciale per identificare quali parti del genoma sono trascritte e per quantificare l'espressione genica



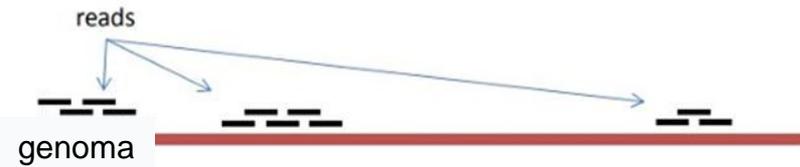
# ALLINEAMENTO RNA-seq SU GENOMA



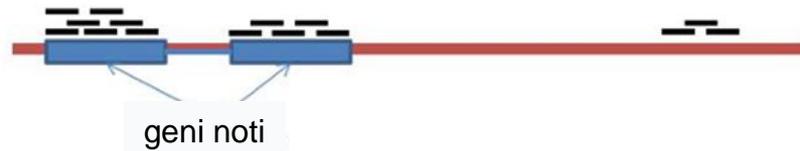
**L'analisi dei trascritti mappati sul genoma di riferimento** consente di rispondere a domande chiave nella ricerca biologica, come:

- ❖ Quanti trascritti sono prodotti da un gene specifico?
  - ❖ Come cambia l'espressione genica in diverse condizioni o in risposta a trattamenti?
  - ❖ Quali varianti di splicing sono presenti in un gene?
  - ❖ Quali geni sono coinvolti in una specifica via metabolica o processo biologico
- 

# ALLINEAMENTO RNA-seq SU GENOMA (ORGANISMI MODELLO)



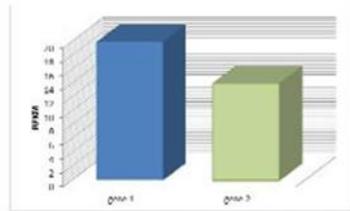
Allineamento su un  
genoma di riferimento



Assegnamento delle read  
ai geni annotati



Rilevazione di eventuali  
geni "nuovi" non annotati



Quantificazione dell'espressione  
e analisi statistica



02

**ALLINEAMENTO**



# ALLINEAMENTO

L'allineamento delle RNA-seq permette di trasformare dati grezzi di sequenziamento in informazioni biologiche preziose, come il livello di espressione genica, la presenza di varianti di splicing, e l'identificazione di nuovi trascritti

# ALLINEAMENTO: Strategie





02

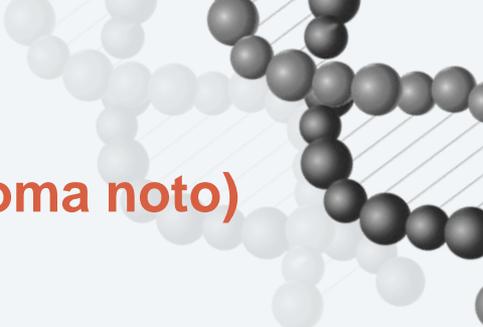
# SOFTWARE DI ALLINEAMENTO



# SOFTWARE ALLINEAMENTO

Programma	Tecnica di Indicizzazione	Tecnica di Allineamento	Specificità
Bowtie2	FM-index (Burrows-Wheeler Transform)	Allineamento completo con tolleranza agli errori	Ottimizzato per reads corte, tollera errori e variazioni
BWA	FM-index (Burrows-Wheeler Transform)	BWA-MEM: Allineamento semi-globale con tolleranza agli errori	Adatto a reads lunghe, gestione di allineamenti secondari
HISAT2	FM-index + Indice Grafo (Graph-based index)	Allineamento tramite indice basato su grafo, gestisce lo splicing	Ottimizzato per RNA-seq, gestisce giunzioni di splicing
Salmon	K-mer index (Pseudo-allineamento)	Pseudo-allineamento senza allineamento esplicito	Quantifica direttamente l'espressione, corregge bias, molto veloce

# SOFTWARE ALLINEAMENTO (genoma noto)



## Preparazione dei dati grezzi

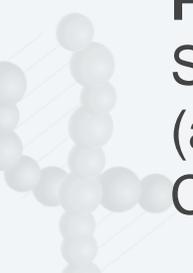
### Sequenziamento e qualità delle read:

Dopo il sequenziamento, si ottengono milioni di read (frammenti di RNA convertiti in cDNA e poi sequenziati).

Si utilizza un software come **FastQC** per controllare la qualità delle read (lunghezza, punteggi di qualità, contaminazione da adattatori, ecc.)

### Filtraggio e trimming:

Si eliminano basi o read di bassa qualità e si rimuovono contaminanti (ad esempio, adattatori) utilizzando strumenti come **Trimmomatic**, Cutadapt o **FASTQ**.



# SOFTWARE ALLINEAMENTO (genoma noto)

## Allineamento delle read al genoma di riferimento

Se si dispone di un genoma di riferimento (ad esempio, il genoma umano o di un organismo modello), si seguono questi passaggi:

### a) Preparazione del genoma di riferimento

Scaricare il genoma di riferimento e l'annotazione genica (formato GTF o GFF) da database come **Ensembl**, **NCBI** o **UCSC**.

Indicizzare il genoma con strumenti come **HISAT2**, **STAR**, o **Bowtie2**. L'indicizzazione velocizza l'allineamento creando una struttura che permette un confronto rapido tra le read e il genoma

# SOFTWARE ALLINEAMENTO: (genoma noto)

## b) Allineamento delle read

Utilizzare un software di allineamento come:

**HISAT2/BOWTIE2**: Ottimizzato per RNA-seq, gestisce splicing e read accoppiate (paired-end).

**STAR**: Efficiente e adatto a dataset di grandi dimensioni.

**TopHat**: Una scelta più tradizionale ma ancora usata in alcuni flussi di lavoro

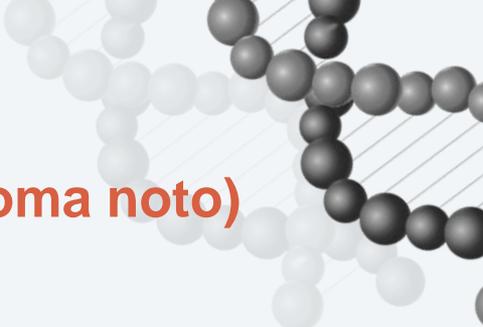
### Durante l'allineamento:

Le read sono mappate sul genoma.

Gli introni sono riconosciuti e saltati per identificare gli esoni (*splicing-aware mapping*)

Si generano file di output in formato SAM o BAM (compressed SAM), che contengono informazioni sulle posizioni delle read all'interno del genoma.

# SOFTWARE ALLINEAMENTO (genoma noto)



## c) Valutazione dell'allineamento

Controllare la percentuale di read mappate per valutare la qualità dell'allineamento.

Utilizzare strumenti come **Samtools** per gestire i file BAM



# SOFTWARE ALLINEAMENTO (genoma non noto)

Se non esiste un genoma di riferimento, si procede con l'**assemblaggio de novo** per ricostruire i trascritti.

## a) Assemblaggio delle read

Utilizzare strumenti come **Trinity**, **SPAdes** o **Velvet** per assemblare le read in contigui (segmenti di trascritti ricostruiti).

L'assemblaggio combina read sovrapposte per formare trascritti completi

## b) Clustering

Durante l'assemblaggio *de novo* (es. con Trinity), sono prodotti molti trascritti o isoforme per ciascun gene, spesso in quantità ridondante o imprecisa.

**Corset** raggruppa queste isoforme in cluster a livello genico, semplificando l'analisi e riducendo l'ambiguità nell'interpretazione biologica

# SOFTWARE ALLINEAMENTO: (genoma non noto)

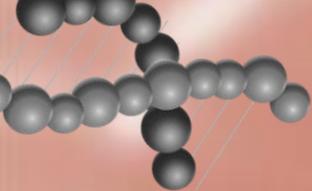
## c) Annotazione dei trascritti

Si utilizzano database come **BLAST**, **InterProScan** o **eggNOG** per identificare i trascritti assemblati confrontandoli con proteine o RNA noti.

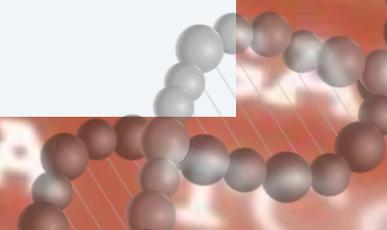
Classificare i trascritti come codificanti (mRNA) o non codificanti (lncRNA, miRNA).

## d) Quantificazione

Una volta assemblati i trascritti, si utilizzano strumenti come **Salmon** o **Kallisto** per stimare l'abbondanza di ogni trascritto nel dataset RNA-seq.



# BIOINFORMATICA



# HISAT2

**HISAT2 è un programma bioinformatico progettato per l'allineamento di sequenze di letture di RNA o DNA a un genoma di riferimento**

Supporta la gestione di giunzioni di splicing, essenziali per dati RNA-Seq, ed è in grado di trattare grandi dataset in modo efficiente

# HISAT2: Algoritmo



**HISAT2** è un algoritmo che utilizza un approccio unico basato su un sistema di indicizzazione gerarchica per gestire l'ampia dimensione del genoma dei mammiferi e le complessità associate allo splicing

## **Passo 1 - Indicizzazione gerarchica:**

- HISAT2 suddivide il genoma di riferimento in molte piccole parti o "bin".
  - Per ogni bin, è costruita la struttura di indicizzazione di FM combinata con BWT
- 

# HISAT2: Algoritmo



## **Passo 2 - Indicizzazione FM globali e locali:**

- L'indicizzazione principale del genoma è effettuata tramite il "global FM-index".
- Per gestire lo splicing, HISAT2 utilizza anche ciò che chiama "local FM-index" per le regioni che circondano i noti siti di splicing, consentendo allineamenti efficienti attraverso questi siti

## **Passo 3 - Gestione dello splicing:**

- HISAT2 può allineare efficacemente reads che coprono siti di splicing
  - La combinazione di indici globali e locali di tipo FM-index permette a HISAT2 di gestire efficacemente reads che coprono regioni esoniche e introniche, identificando correttamente le giunzioni tra esoni
- 

# HISAT2: Algoritmo



## **Passo 4 - Allineamento di reads:**

- HISAT2 inizia cercando corrispondenze esatte o quasi esatte per segmenti dei reads usando l'FM-index.
- Una volta identificato un potenziale allineamento, HISAT2 prova a estendere l'allineamento per tutto il read, tenendo conto delle potenziali discrepanze come mismatch, inserzioni o cancellazioni.

## **Passo 5 - Scelta degli allineamenti multipli:**

Se una read può essere mappata in più posizioni nel genoma, HISAT2 può segnalare il read come "multi-mapping" o scegliere una posizione basata su vari criteri.



# HISAT2: Algoritmo



## **Passo 6 - Generazione di output:**

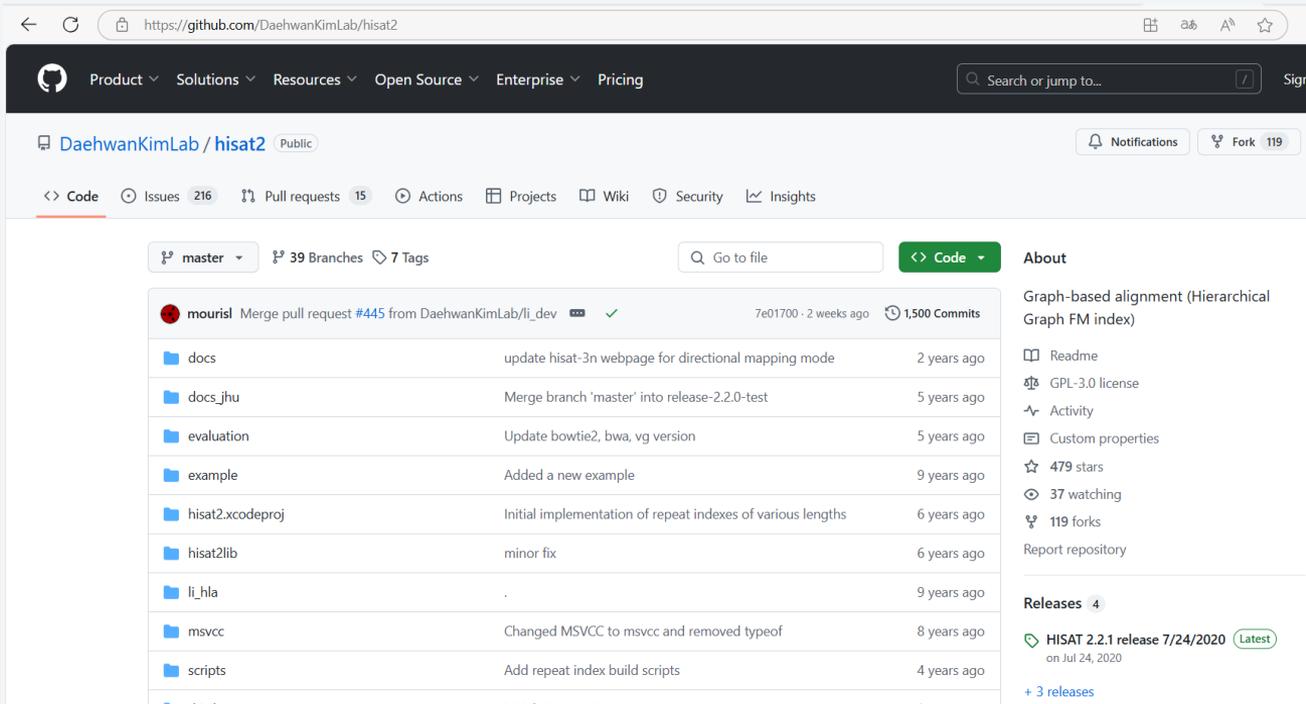
HISAT2 produce un file SAM/BAM come output, che può essere ulteriormente elaborato o utilizzato per altre analisi.

Un vantaggio di HISAT2 è la sua efficienza in termini di memoria e velocità. L'uso di una struttura di indicizzazione gerarchica consente a HISAT2 di gestire genomi di grandi dimensioni, come quello umano, con una quantità relativamente piccola di memoria.



# HISAT2: Sito

<https://github.com/DaehwanKimLab/hisat2>



Navigation: Product, Solutions, Resources, Open Source, Enterprise, Pricing. Search: Search or jump to... Sign in

Repository: DaehwanKimLab / hisat2 (Public). Notifications, Fork 119

Navigation: Code, Issues 216, Pull requests 15, Actions, Projects, Wiki, Security, Insights

Branches: master (39 Branches, 7 Tags). Go to file. Code

File/Folder	Description	Last Commit
docs	update hisat-3n webpage for directional mapping mode	2 years ago
docs_jhu	Merge branch 'master' into release-2.2.0-test	5 years ago
evaluation	Update bowtie2, bwa, vg version	5 years ago
example	Added a new example	9 years ago
hisat2.xcodeproj	Initial implementation of repeat indexes of various lengths	6 years ago
hisat2lib	minor fix	6 years ago
li_hla	.	9 years ago
msvcc	Changed MSVCC to msvcc and removed typeof	8 years ago
scripts	Add repeat index build scripts	4 years ago

**About**  
Graph-based alignment (Hierarchical Graph FM index)

- Readme
- GPL-3.0 license
- Activity
- Custom properties
- 479 stars
- 37 watching
- 119 forks
- Report repository

**Releases 4**

- HISAT 2.2.1 release 7/24/2020 (Latest) on Jul 24, 2020
- + 3 releases

# HISAT2: Installazione

È sufficiente eseguire i seguenti comandi su Linux:

```
git clone https://github.com/DaehwanKimLab/hisat2.git  
cd hisat2  
make
```

# HISAT2: Esempio



Passo 1 - Prima di effettuare l'allineamento, bisogna creare un indice del genoma. Questo passaggio viene fatto una sola volta per ogni genoma di riferimento

Comando:

```
hisat2-build reference_genome.fasta reference_index
```

Si genera un set di file indice (reference\_index.\*), utilizzati in seguito per l'allineamento



# HISAT2: Esempio

## Passo 2a - Filtrare Giunzioni di Splicing

HISAT2 consente di ottenere informazioni sulle giunzioni di splicing utilizzando l'opzione `--novel-splicesite-outfile`:

```
hisat2 -x reference_index -1 reads_1.fastq -2 reads_2.fastq -S output.sam --  
novel-splicesite-outfile splicesites.txt
```

# HISAT2: Esempio



## Passo 3 - Conversione in BAM

Per analisi successive, il file SAM può essere convertito in BAM con samtools:

```
samtools view -bS output.sam > output.bam
```

```
samtools sort output.bam -o output_sorted.bam
```



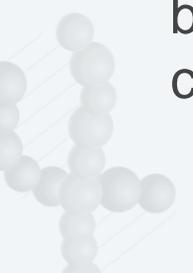
# HISAT2: Esempio



## Output e Analisi

File SAM/BAM: Contiene l'allineamento delle letture al genoma.  
Statistiche di allineamento: HISAT2 stampa un riassunto degli allineamenti al termine del processo (es. percentuale di letture allineate).

HISAT2 è comunemente utilizzato come primo passo per pipeline bioinformatiche, seguito da analisi di espressione genica (ad esempio con featureCounts o StringTie) o analisi delle varianti (con GATK).



# SALMON

Salmon è un programma utilizzato principalmente per analisi di dati biologici, specialmente nella genomica e trascrittomica. È progettato per quantificare l'espressione genica (RNA-Seq) rapidamente e con alta precisione



# SALMON

# SALMON: Algoritmo



**Salmon** è un algoritmo di **quasi-allineamento** (e **quantificazione** probabilistica) che utilizza modelli statistici per stimare l'abbondanza relativa di trascritti a partire da reads di sequenziamento

## **Passo 1 -Preindicizzazione:**

Prima dell'analisi, Salmon costruisce un indice a partire da una sequenza di riferimento (ad esempio, trascrittoma) usando una **tabella hash di k-mer** (vedere seed-extend)

Questa fase permette di identificare rapidamente dove un frammento di sequenza può mappare nel trascrittoma



# SALMON: Algoritmo



## **Passo 2 - Quasi-Allineamento:**

Le letture RNA-Seq sono mappate ai trascritti di riferimento in modo approssimativo, senza effettuare un allineamento completo. Questo è possibile grazie alla corrispondenza con i k-mer pre-indicizzati. In questa fase, l'algoritmo tiene conto di variazioni e di errori di sequenziamento

## **Passo 3 – Correzione bias**

Salmon corregge le stime per *bias* tecnici (come il GC bias o bias introdotti dalla preparazione della libreria), migliorando la precisione



# SALMON: Algoritmo



## **Passo 4 - Quantificazione Probabilistica:**

Utilizza un modello probabilistico per assegnare le reads ai trascritti. Poiché una reads può mappare su più trascritti (ad esempio, isoforme), Salmon calcola la probabilità che una lettura derivi da ciascun trascritto.

Il risultato è una stima dell'abbondanza relativa per ogni trascritto.

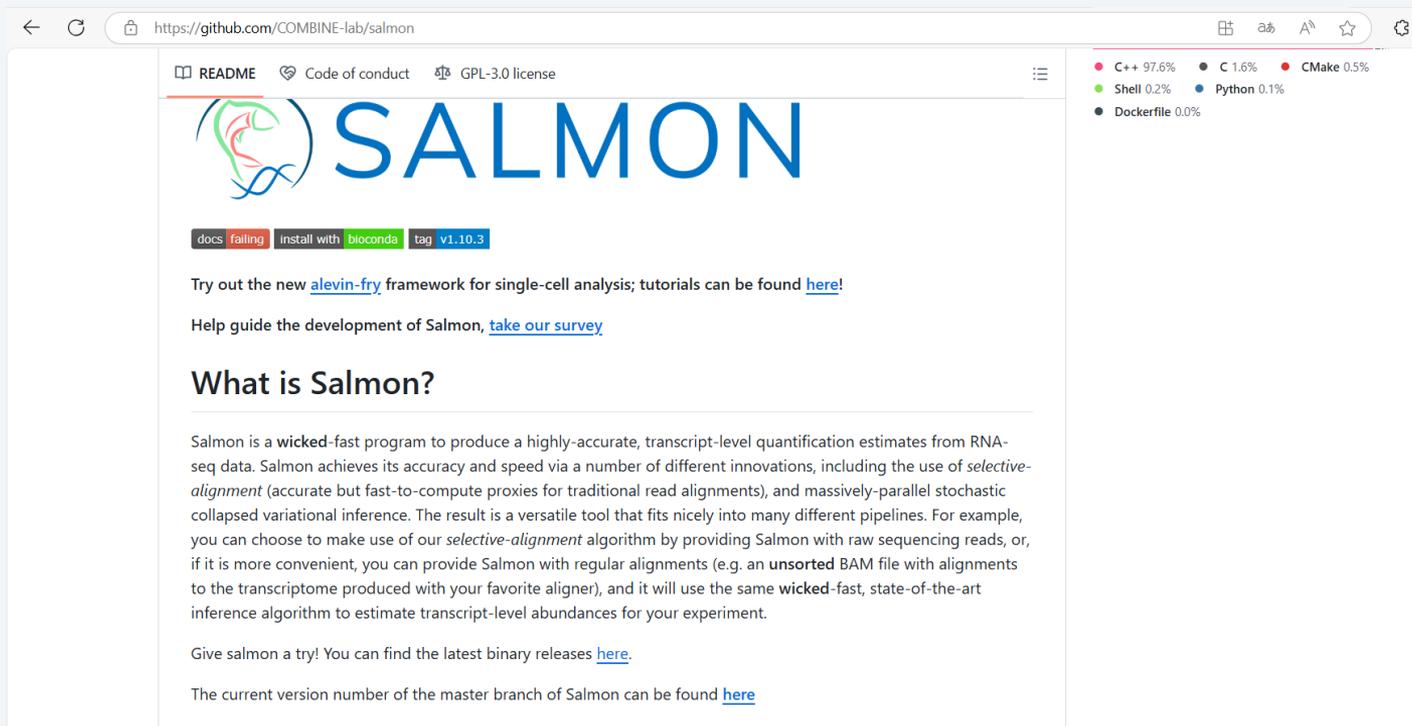
## **Passo 5 – Risultati:**

L'output include l'abbondanza stimata dei trascritti, espressa in termini di TPM (Transcripts Per Million) o altre metriche



# SALMON: Sito

<https://github.com/COMBINE-lab/salmon>



The screenshot shows the GitHub repository page for Salmon. At the top, there are navigation links for README, Code of conduct, and GPL-3.0 license. The main heading is "SALMON" with a logo featuring a DNA double helix and a salmon. Below the heading, there are buttons for "docs", "failing", "install with bioconda", and "tag v1.10.3". A promotional message says "Try out the new [alevin-fry](#) framework for single-cell analysis; tutorials can be found [here](#)!". Below that, it says "Help guide the development of Salmon, [take our survey](#)". The section "What is Salmon?" follows, with a detailed paragraph explaining the tool's accuracy and speed. At the bottom, it says "Give salmon a try! You can find the latest binary releases [here](#)." and "The current version number of the master branch of Salmon can be found [here](#)". On the right side, there is a language distribution chart.

Language	Percentage
C++	97.6%
C	1.6%
CMake	0.5%
Shell	0.2%
Python	0.1%
Dockerfile	0.0%

# SALMON: Installazione

È sufficiente eseguire il comando su Linux:

<https://github.com/COMBINE-lab/salmon>



# SALMON: Esempio



Si vuole quantificare l'espressione genica in un campione usando Salmon.

## 1. Preparazione dell'indice

**salmon index -t transcripts.fasta -i transcript\_index**

transcripts.fasta è il file FASTA contenente i trascritti di riferimento, e transcript\_index è la directory dove sarà memorizzato l'indice



# SALMON: Esempio

## 2. Quantificazione

```
salmon quant -i transcript_index -l A -r reads.fastq -o output_dir
```

- i specifica l'indice generato.
- l A rileva automaticamente il tipo di libreria (i tipi di file immessi)
- r è il file FASTQ con le letture RNA-Seq
- o indica la directory di output

### Altre librerie

- A (Automatic): Questa opzione consente a Salmon di determinare automaticamente il tipo di libreria dai dati forniti
- O (Opposite): Specifica che la libreria è "opposite-stranded", cioè il filamento complementare rispetto alla direzione della trascrizione è rappresentato nelle reads
- U (Unstranded): Indica che la libreria è non direzionale: le letture non hanno un orientamento specifico rispetto alla direzione di trascrizione.

# SALMON: Esempio

## 3. Risultati

Nell'output (output\_dir), si ottiene un file chiamato quant.sf con informazioni come:

<Nome del trascritto, Lunghezza effettiva del trascritto, TPM, conteggi stimati>

Esempio del contenuto di quant.sf:

Name	Length	EffectiveLength	TPM	NumReads
ENST00000456328.2	1657	1427	10.5	150
ENST00000515242.2	632	402	5.8	80

# Grazie!

**Domande?**

[franco.liberati@unitus.it](mailto:franco.liberati@unitus.it)

[deb.scienceontheweb.com](http://deb.scienceontheweb.com)

