



# UNIVERSITÀ DI TRENTO

Dipartimento di Ingegneria e Scienza dell'Informazione

Corso di Laurea in  
Informatica

ELABORATO FINALE

## ANALYSIS OF RNA-SEQ TRANSCRIPTOMIC DATA FROM TOTAL AND POLYSOMAL mRNA FRACTIONS FROM AN EPITHELIAL CANCER CELL LINE

Supervisore

.....

Laureando

Giacomo Fantoni

Anno accademico 2020/2021

# Ringraziamenti

*...thanks to...*

# Indice

<b>Sommario</b>	<b>2</b>
<b>1 Introduzione</b>	<b>2</b>
1.1 TransSNPs . . . . .	2
1.2 Sbilanciamento allelico . . . . .	2
1.2.1 Profilamento polisomico . . . . .	2
<b>2 Linee cellulari e condizioni</b>	<b>2</b>
2.1 HCT116 . . . . .	2
2.1.1 DHX30 . . . . .	3
2.1.2 PCBP2 . . . . .	3
2.2 Trattamenti . . . . .	3
2.2.1 DMSO . . . . .	3
2.2.2 Nutlin . . . . .	3
<b>3 Processamento dei dati</b>	<b>3</b>
3.1 Dati disponibili . . . . .	3
3.1.1 Sequenze biologiche . . . . .	3
3.1.2 Genoma di riferimento . . . . .	3
3.1.3 Variant call . . . . .	3
3.1.4 Struttura dei geni . . . . .	3
3.2 Troncatura e allinamento . . . . .	3
3.2.1 Troncatura . . . . .	3
3.2.2 Allineamento . . . . .	3
3.2.3 Ordinamento . . . . .	3
3.2.4 Indicizzazione . . . . .	3
3.3 Deduplicazione, riallinamento e recalibrazione . . . . .	3
3.3.1 Deduplicazione . . . . .	3
3.3.2 Riallineamento e recalibrazione . . . . .	3
3.4 Ottenere le varianti alleliche . . . . .	4
3.5 Ottenere i dati delle frazioni alleliche . . . . .	4
3.5.1 Filtrare le frazioni alleliche . . . . .	4
3.6 Ottenere gli SNP nel 3'-UTR . . . . .	4
<b>4 Analisi dei dati</b>	<b>4</b>
4.1 Conta degli SNP trovati con ASEQ . . . . .	4
4.2 Considerazioni sulla recalibrazione . . . . .	4
4.3 Ottenere i dati per gli SNP di interesse . . . . .	4
4.4 Analisi degli sbilanciamenti di frazione allelica . . . . .	4
4.5 Conclusioni . . . . .	4
<b>Bibliografia</b>	<b>4</b>

<b>A</b>	<b>Titolo primo allegato</b>	<b>6</b>
A.1	Titolo . . . . .	6
A.1.1	Sottotitolo . . . . .	6
<b>B</b>	<b>Titolo secondo allegato</b>	<b>7</b>
B.1	Titolo . . . . .	7
B.1.1	Sottotitolo . . . . .	7

# Sommario

Sommario è un breve riassunto del lavoro svolto dove si descrive l'obiettivo, l'oggetto della tesi, le metodologie e le tecniche usate, i dati elaborati e la spiegazione delle conclusioni alle quali siete arrivati.

Il sommario dell'elaborato consiste al massimo di 3 pagine e deve contenere le seguenti informazioni:

- contesto e motivazioni
- breve riassunto del problema affrontato
- tecniche utilizzate e/o sviluppate
- risultati raggiunti, sottolineando il contributo personale del laureando/a

## 1 Introduzione

Questo capitolo è volto a descrivere i processi biologici considerati durante il progetto. Cito principalmente dal draft paper sui transSNPS

### 1.1 TransSNPs

Definizione di SNP e loro impatto. Descrizione degli SNP considerati in questo esperimento.

### 1.2 Sbilanciamento allelico

Definizione di sbilanciamento allelico e perchè viene considerato.

#### 1.2.1 Profilamento polisomico

Come si identifica lo sbilanciamento allelico.

## 2 Linee cellulari e condizioni

Descrizione delle linee cellulari e dei materiali utilizzati.

### 2.1 HCT116

Descrizione della linea cellulare e motivazione del suo utilizzo.

#### 2.1.1 DHX30

Funzione di DHX30, cosa ne è stato fatto nei campioni e motivazioni dietro alla scelta.

#### 2.1.2 PCBP2

Funzione di PCBP2, cosa ne è stato fatto nei campioni e motivazioni dietro alla scelta.

## **2.2 Trattamenti**

### **2.2.1 DMSO**

Descrizione del trattamento e motivazioni.

### **2.2.2 Nutlin**

Descrizione del trattamento e motivazioni.

# **3 Processamento dei dati**

## **3.1 Dati disponibili**

### **3.1.1 Sequenze biologiche**

Descrizione dei fastq.

### **3.1.2 Genoma di riferimento**

Descrizione del genoma di riferimento.

### **3.1.3 Variant call**

Descrizione dei vcf.

### **3.1.4 Struttura dei geni**

Descrizione del gtf.

## **3.2 Troncatura e allinamento**

Descrizione del processo e perchè viene fatto.

### **3.2.1 Troncatura**

Trimmomatic, cosa fa come è stato usato.

### **3.2.2 Allineamento**

STAR, cosa fa come è stato usato.

### **3.2.3 Ordinamento**

SAMTOOLS SORT cosa fa come è stato usato.

### **3.2.4 Indicizzazione**

SAMTOOLS index cosa fa come è stato usato.

## **3.3 Deduplicazione, riallineamento e recalibrazione**

Descrizione del processo e perchè viene fatto

### **3.3.1 Deduplicazione**

Come sopra.

### **3.3.2 Riallineamento e recalibrazione**

Come sopra.

## **3.4 Ottenere le varianti alleliche**

Intersezione tra VCF e GTF.

## **3.5 Ottenere i dati delle frazioni alleliche**

ASEQ cosa fa come viene usato.

### **3.5.1 Filtrare le frazioni alleliche**

Condizioni di filtraggio per i risultati di ASEQ.

## **3.6 Ottenere gli SNP nel 3'-UTR**

Filtraggio del gtf e intersezione con i VCF

# **4 Analisi dei dati**

## **4.1 Conta degli SNP trovati con ASEQ**

Discussione dei risultati di ASEQ.

## **4.2 Considerazioni sulla recalibrazione**

Discussione dei risultati di ASEQ prima e dopo la recalibrazione

## **4.3 Ottenere i dati per gli SNP di interesse**

Discussione degli SNP con i dati necessari per lo studio e scelta degli SNP di interesse.

## **4.4 Analisi degli sbilanciamenti di frazione allelica**

Analisi finali.

## **4.5 Conclusioni**

# Bibliografia



# Allegato A    Titolo primo allegato

## A.1    Titolo

### A.1.1    Sottotitolo

# Allegato B    Titolo secondo allegato

## B.1    Titolo

### B.1.1    Sottotitolo