

II PROVETTA – Fondamenti di Biologia Computazionale 2012-2013

17 maggio 2013

1. Distanza di genotipo:

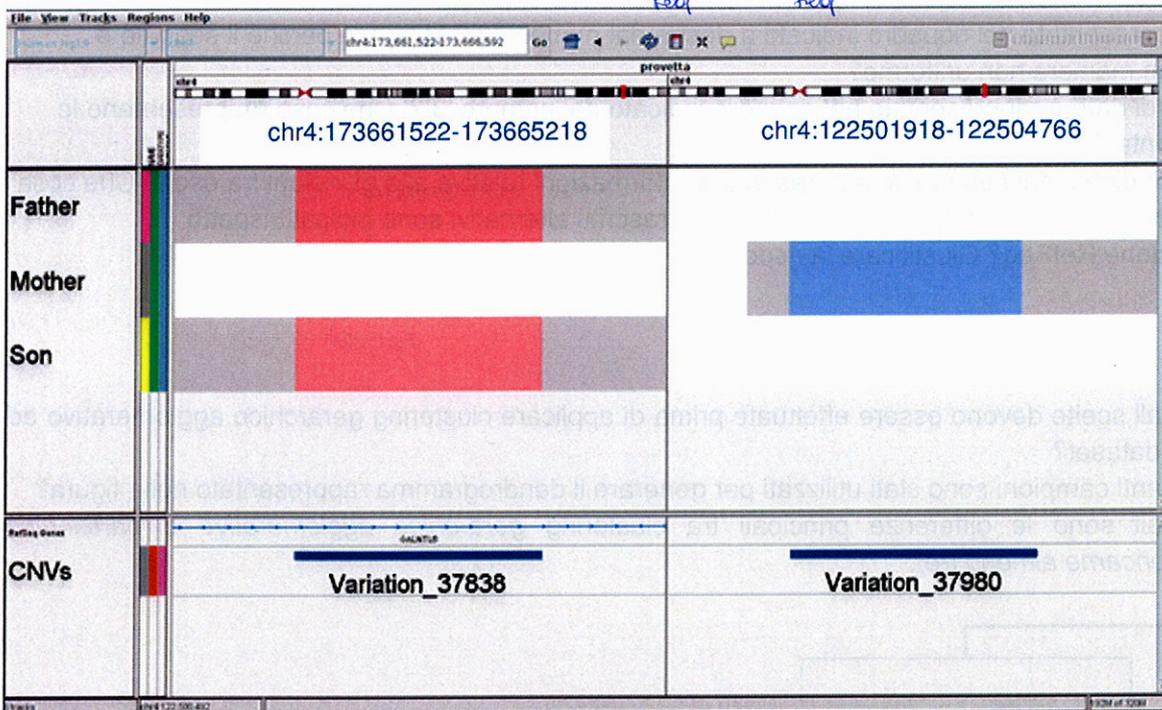
- a. Definire la formula per calcolare la distanza di genotipo; ✓
- b. Calcolare e commentare le distanze di genotipo utilizzando (a.) tra i campioni S1, S2, S3, ed S4 appartenenti ad una stessa popolazione per la quale sono note le chiamate di genotipo di 30 SNPs:

S1=(BB,BB,BB,AB,AB,AB,AB,AA,BB,AB,AA,AB,AB,AA,BB,AA,BB,AB,AA,BB,BB,AB,BB,AB,AB,AA,AB,AB,AB,AB)
 S2=(AA,AB,AB,AA,AA,na,AA,AB,AB,AA,BB,AB,AA,AB,AB,AB,BB,AB,AB,AB,AB,BB,BB,AB,AB,AB,AB,BB,AB,AA)
 S3=(BB,AB,BB,AA,BB,BB,AB,AA,AB,BB,AA,BB,AA,AB,AB,BB,AA,AB,na,AA,AB,BB,BB,AA,AA,BB,AB,AA,AB,AB)
 S4=(BB,BB,AA,BB,AA,AA,AA,AB,AA,AB,AB,AB,AA,BB,AB,AB,AA,AB,BB,AB,AA,BB,BB,AA,AB,AA,BB,AA,BB,BB)

- c. Sulla base delle distanze, si reputa che gli SNPs inclusi nel test siano stati opportunamente selezionati?

2. Copy Number Variants (CNVs)

Considerare la seguente schermata del visualizzatore IGV. Contiene i dati segmentati di un trio HapMap in due zone genomiche contenenti due CNV denotati con identificatori Variation_37838 e Variation_37980, che sono due Loss ed hanno rispettivamente frequenze di 21% e 10% nella popolazione HapMap considerata.

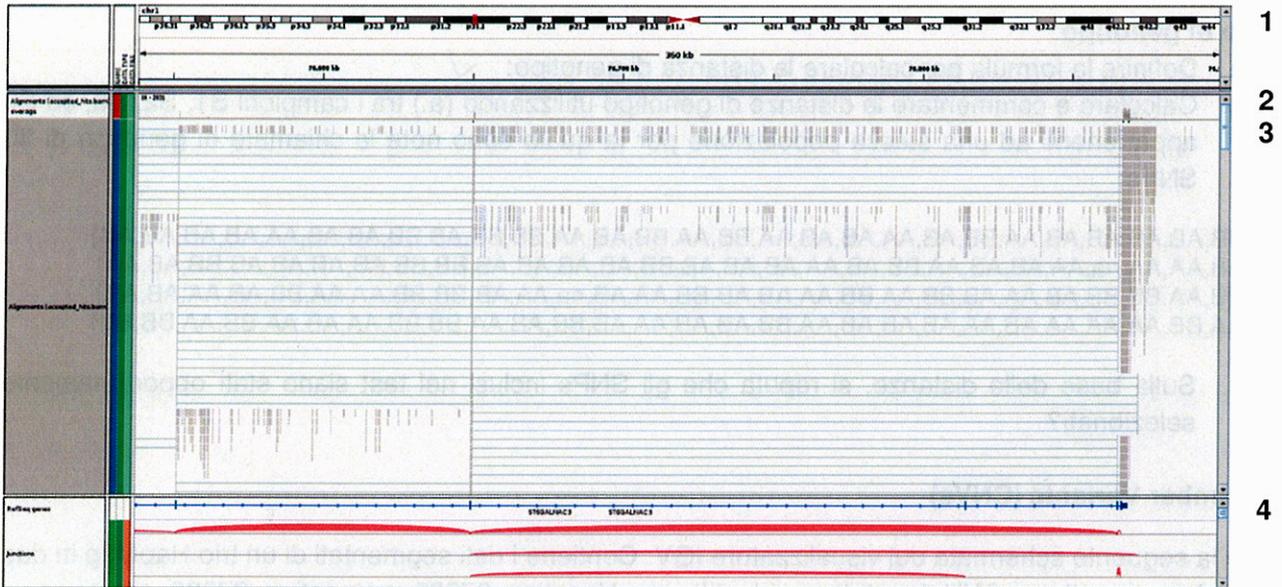


- a. Per ciascun CNV commentare i dati come visualizzati da IGV ed interpretarli sulla base delle informazioni fornite;
- b. Calcolare le dimensioni di ciascun CNV;
- c. Sarebbe possibile costruire un test di distanza di genotipo utilizzando i CNVs al posto degli SNPs? Giustificare la risposta.



3. Visualizzazione Dati Generati da Sequenziatore:

5



- Che tipo di dati sono visualizzati in figura?
- Cosa è rappresentato nel riquadro indicato a destra dal numero '2'? Motivare perché il segnale è distribuito in maniera non uniforme?
- Quale tipo di dato è rappresentato nel riquadro indicato dal numero '3'? Che cosa rappresentano le linee orizzontali?
- Il riquadro indicato dal numero '4' rappresenta le informazioni relative alle giunzioni fra esoni. Che cosa rappresenta lo spessore degli archi rossi? Quanti trascritti alternativi sono indicati rispetto all'annotazione RefSeq? Giustificare la risposta.

4. Clustering:

20

- Quali scelte devono essere effettuate prima di applicare clustering gerarchico agglomerativo ad un dataset?
- Quanti campioni sono stati utilizzati per generare il dendrogramma rappresentato nella figura?
- Quali sono le differenze principali tra clustering *gerarchico agglomerativo* e *partitioning* (elencarne almeno tre).

