

Tipologia di domande ed istruzioni generali

-Domande a risposta multipla: un'unica risposta possibile, nessuna penalizzazione in caso di errore (+1 per risposta corretta)

-Domande vero falso e domande con due possibili alternative: +1 in caso di risposta corretta, -0,5 in caso di risposta errata. Nel caso di due possibili alternative, CANCELLARE la risposta errata o CERCHIARE la risposta corretta

-Domande a completamento: nessuna penalizzazione in caso di errore (+1 per risposta corretta)

- 1) Per "coverage" si intende la quantità di reads ottenute da un sequenziamento/il numero di volte che in media ogni posizione di un genoma è letto durante il sequenziamento
- 2) Le cosiddette "colonie" sono cluster clonali di sequenze/librerie di sequenziamento costruite sfruttando l'attività di trasposoni
- 3) La cosiddetta bridge amplification è un processo tipico del sequenziamento _____.
Nel caso in cui vengano sequenziate entrambe le estremità di un frammento si parla di sequenziamento _____
- 4) La metodica SMRT è utilizzata dai sequenziatori della linea Oxford Nanopore/Pacific BioSciences e permette di ottenere output di decine di miliardi di reads per corsa/reads molto lunghe se comparate ad altri approcci NGS. Questa metodica è stata sviluppata più recentemente rispetto al sequenziamento Sanger VERO FALSO
- 5) In linea di principio è desiderabile poter partire dalla quantità maggiore possibile di materiale genetico (acidi nucleici estratti) per ottenere una migliore rappresentazione della complessità del campione in una libreria di campionamento VERO FALSO
- 6) In linea di massima, in un esperimento di RNA-sequencing è importante allocare un maggior numero di risorse nell'aumentare:
 - a) Il numero di repliche tecniche
 - b) Il numero di repliche biologiche
 - c) La copertura di sequenziamento
 - d) La quantità di RNA estratto
- 7) La procedura di trimming permette di rimuovere.
 - a) Basi a bassa qualità
 - b) Adattatori di sequenziamento
 - c) Nucleotidi ambigui
 - d) Tutte le precedenti
- 8) Gli RPKM permettono una efficiente comparazione dei livelli di espressione sia between samples che within samples VERO FALSO

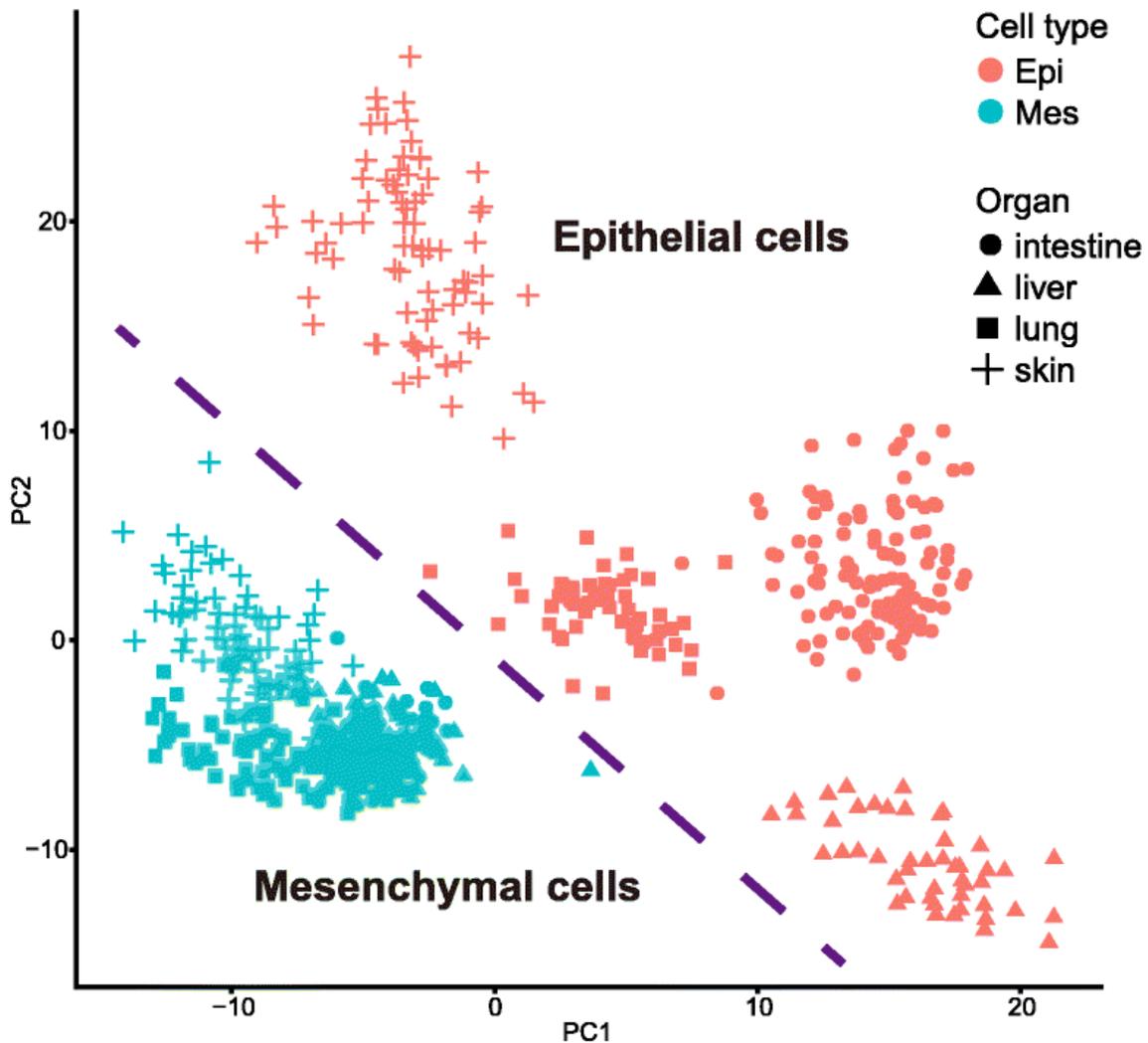
- 9) Per valutare se un trascrittoma è completo è sufficiente valutare la lunghezza media dei contig e l'N50 VERO FALSO
- 10) L'annotazione di geni nei genomi solitamente avviene tramite il rilevamento di similarità con proteine già descritte utilizzando _____, tramite la predizione de novo oppure tramite l'allineamento con dati di RNA-seq/dati di ChIP-seq
- 11) Una upstream regulator analysis permette di prevedere quali siano i possibili effetti di un pattern di espressione genica a livello cellulare e tissutale VERO FALSO
- 12) I network di regolazione costruiti da IPA sono costituiti da tre livelli:
- a) _____
- b) _____
- c) _____
- 13) Lo z-score è una misura statistica che indica la significatività dell'alterazione di un pathway biologico VERO FALSO
- 14) Il genoma del dromedario ha una dimensione di circa 2 Gb. Se volessi risequenziarlo utilizzando una strategia Illumina single end con reads di lunghezza 100bp, ottenendo una copertura 100X, di quante reads avrei approssimativamente bisogno?
- a) 200mila
- b) 2 milioni
- c) 20 milioni
- d) 2 miliardi
- 15) Il fatto che un organismo abbia un elevato numero di cromosomi rappresenta/non rappresenta un problema per il suo sequenziamento.
- 16) Il sequenziamento di genomi tramite piattaforme Oxford Nanopore in situazioni ottimali presuppone:
- a) Di preservare i campioni in RNAlater fino al momento dell'estrazione del DNA genomico
- b) La valutazione della presenza di due bande ben evidenti su gel d'agarosio denaturante
- c) L'estrazione di DNA ad alto peso molecolare
- d) La frammentazione del DNA genomico estratto tramite sonicazione o acoustic shearing
- 17) Se in un esperimento di espressione genica la soglia di significatività statistica è stata posta ad un p-value corretto 1×10^{-3} ad abbinata ad una soglia di fold change $> |2|$ indicate quale tra i seguenti geni è rilevato come differenzialmente espresso
- a) Gene A: FC = -3,5, FDR p-value = 0,03
- b) Gene B: FC = -5,1, FDR p-value = 1×10^{-5}
- c) Gene C: FC = 2,1, FDR p-value = 0,02
- d) Gene D: FC = 1,6, DFR p-value = 0,00005

18) Uno dei vantaggi principali degli approcci metagenomici rispetto a quelli microbiologici classici è quello di poter ottenere una migliore risoluzione delle comunità. Ciò è principalmente dovuto a:

- a) Al fatto che la maggior parte dei microorganismi non sono in grado di crescere in coltura
- b) Al fatto che non siano necessarie repliche biologiche
- c) Al tempo minore richiesto per l'analisi
- d) Al fatto che praticamente tutte le specie microbiche sono ormai state sequenziate e sono quindi riconoscibili con certezza tramite analisi di sequenza

19) Il traguardo del “\$1000 genome” è stato raggiunto grazie all'avvento sul mercato dei sequenziatori PacBio VERO FALSO

Osservate la seguente immagine, che riporta dati derivanti da un esperimento di single-cell RNA-seq condotto sulle cellule epiteliali e mesenchimali di 4 tessuti.



20) Questo grafico rappresenta una _____

21) Per quanto riguarda le cellule epiteliali, la dimensione che differenzia in modo significativo le cellule polmonari da quelle dell'intestino è la _____

22) Complessivamente, le cellule mesenchimali dei vari tessuti mostrano un maggior grado di diversità nei profili di espressione rispetto alle cellule epiteliali VERO FALSO

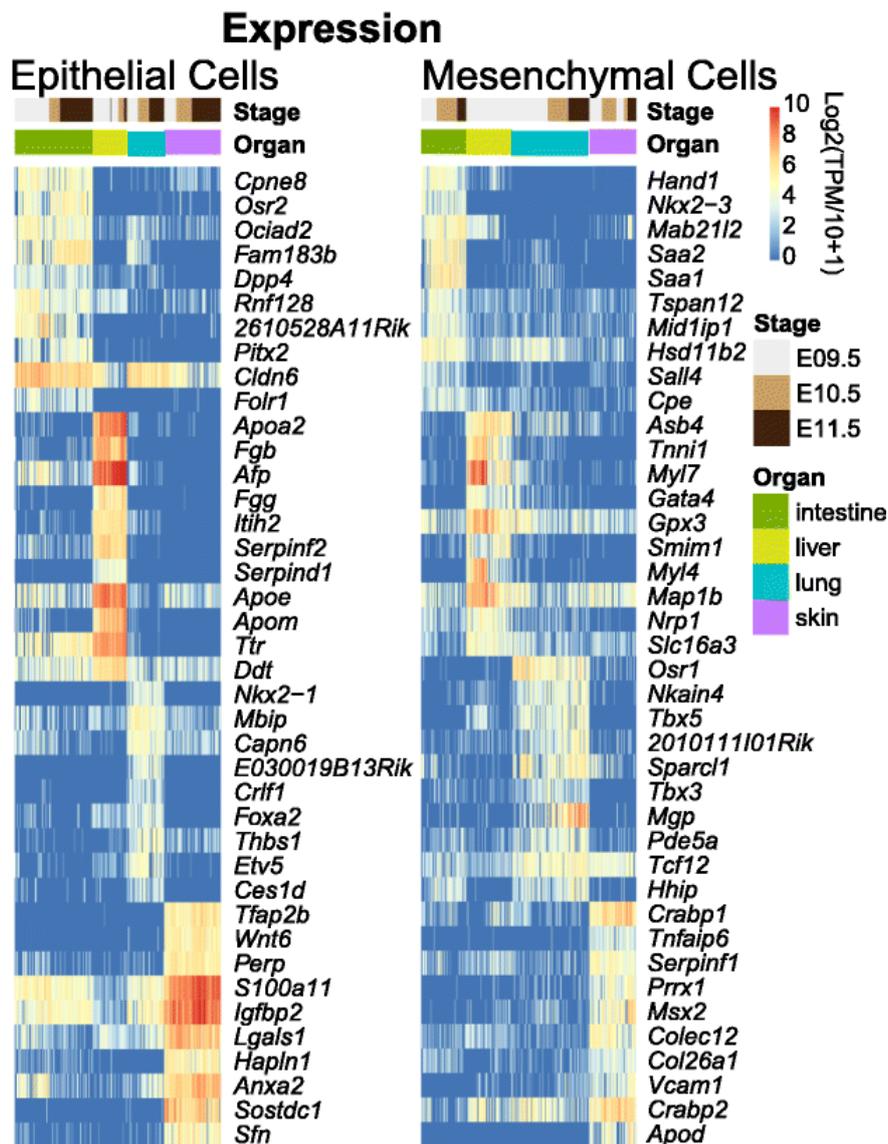
23) Da questo grafico è possibile:

-ricavare i livelli di espressione di singoli geni VERO FALSO

-dire che non ci sono geni differenzialmente espressi nel confronto tra le cellule mesenchimali dei 4 tessuti VERO FALSO

-dire che è possibile distinguere cellule mesenchimali ed epiteliali derivate da uno stesso tessuto sulla base dei profili di espressione VERO FALSO

-per quanto riguarda le cellule epiteliali, quelle che mostrano un profilo di espressione più distante in un confronto a coppie sono quelle di pelle e _____



Questo secondo grafico è una heat map che riporta un dettaglio dei geni differenzialmente espressi ricavati dall'analisi degli stessi dati di espressione dai quali è stato ricavato il grafico precedente.

24) I colori delle celle dipendono dal livello di espressione/fold change, che sono stati _____

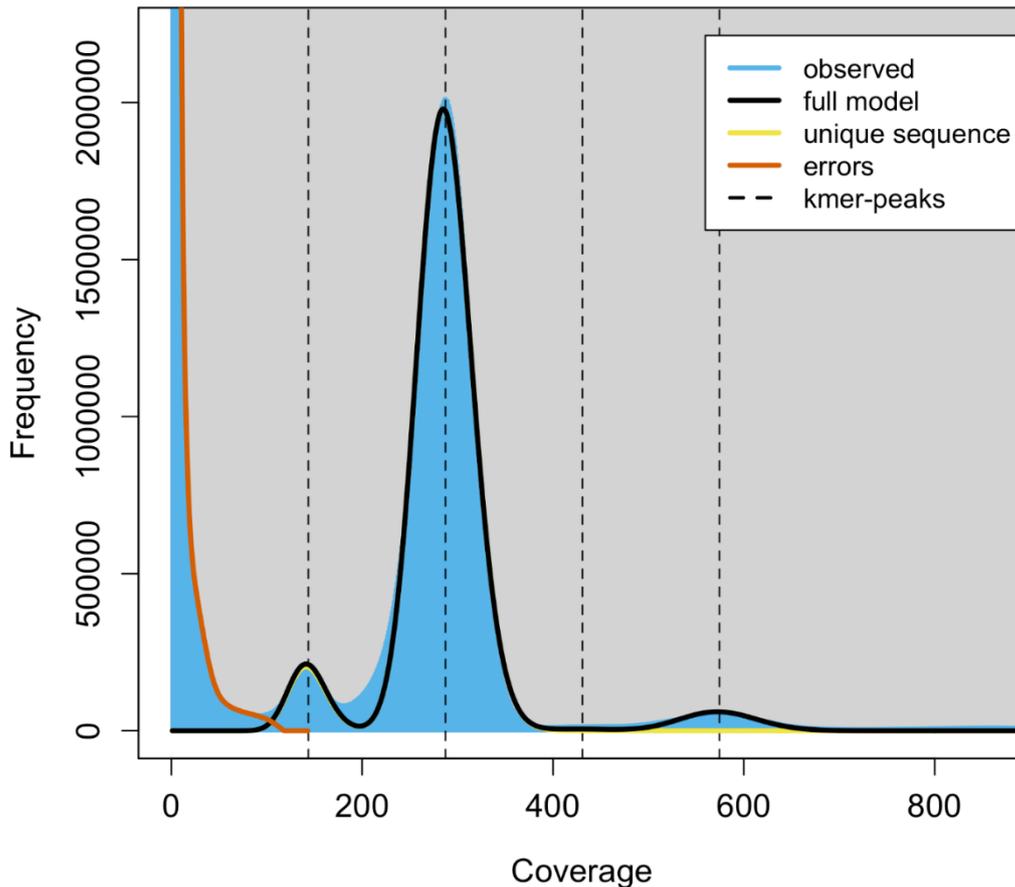
25) Un marker di espressione specifico delle cellule epiteliali del fegato potrebbe essere _____, mentre S100a11 è un marker specifico delle cellule della pelle/dell'intestino

26) Anche se ciò non viene esplicitamente mostrato nel grafico, è ragionevole supporre che i geni siano stati ordinati sull'asse verticale in base a:

- a) Livello di espressione
- b) Fold change
- c) Clustering gerarchico
- d) Z-score

27) La trasformazione in scala logaritmica che è stata adoperata è stata fatta per migliorare l'informatività della figura/per migliorare l'analisi statistica dei dati

Il seguente grafico rappresenta la frequenza dei k-meri osservati nel sequenziamento di un genoma



28) Il grafico indica che il genoma è stato sequenziato con una copertura di circa 300X/600X e che il genoma è aploide/diploide

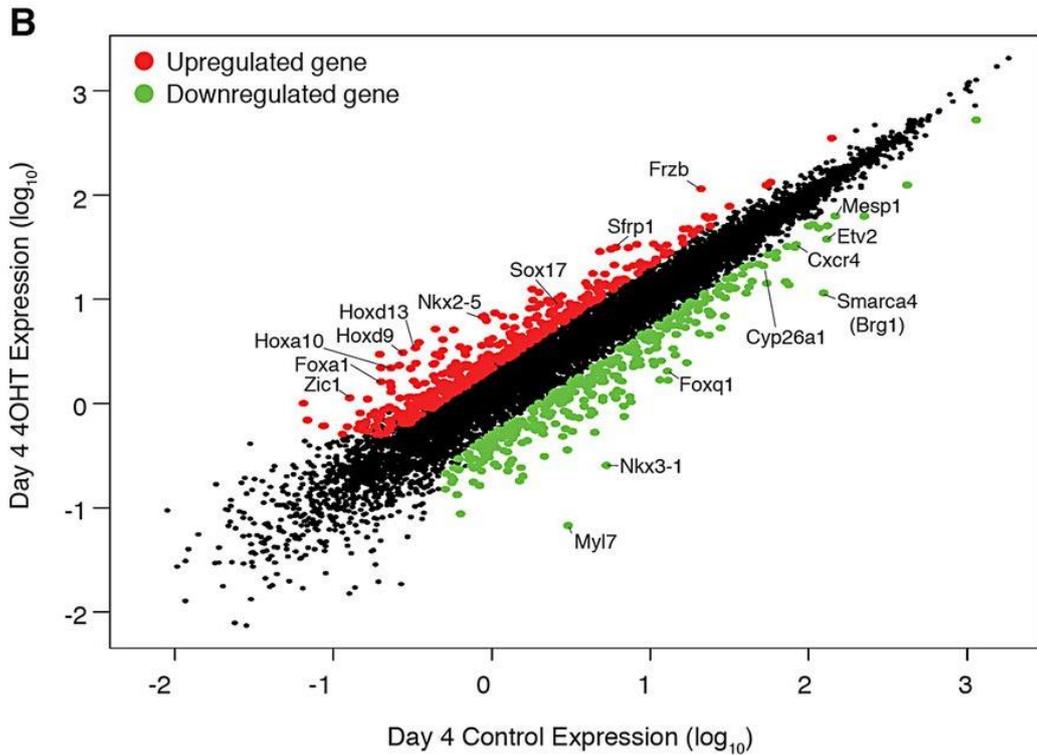
29) Il genoma in questione mostra un alto tasso di eterozigosi VERO FALSO

30) Il picco mostrato all'estrema sinistra del grafico indica:

- a) Duplicazioni
- b) Elementi trasponibili
- c) Mutazioni de novo
- d) Errori di sequenziamento

31) Il picco individuato a circa 600X indica regioni genomiche _____

- 32) L'RNA 16S è uno dei marker più importanti per quanto riguarda gli studi di metagenomica e filogenesi molecolare in quanto si tratta di una sequenza evolutivamente estremamente conservata/altamente variabile. Le analisi sfruttano la presenza di 9 regioni _____ che possono essere amplificate, sequenziate e confrontate con un database di riferimento.
- 33) La procedura di "binning" permette di assegnare le sequenze ad una determinata OTU, identificata come "specie" nel momento in cui la loro percentuale di identità supera il 90%/97%
- 34) Se ci fosse la necessità di indagare la presenza di SNP associate a malattie genetiche tramite un risequenziamento genomico, sarebbe opportuno impostare la procedura di mapping con un parametro similarity fraction = 1. VERO FALSO
- 35) Un test ipergeometrico analizza:
- I geni differenzialmente espressi e le relative annotazioni
 - I geni differenzialmente espressi ed i relativi valori di fold change
 - Gli assemblaggi trascrittomici ed i risultati di BUSCO
 - Gli assemblaggi genomici ed i valori di N50
- 36) Indicate quale tra i seguenti è un valore di N50 migliore
- 1×10^{-5}
 - 10000
 - 0
 - 100
- 37) Una curva di rarefazione può essere creata per:
- Valutare il tasso di eterozigosi di un genoma
 - Valutare la copertura di sequenziamento di un genoma
 - Valutare la completezza di un trascrittoma
 - Valutare il grado di campionamento di una comunità microbica
- 38) Un gene caratterizzato da TPM =10 è sempre maggiormente espresso di un gene caratterizzato da TPM =5/è espresso a livelli più alti di un gene caratterizzato da TPM =5 solo se quest'ultimo è più corto
- 39) La metodica di preparazione di librerie per RNA-seq 3'-tag permette di abbassare i costi rispetto ad un approccio classico perché:
- Richiede molte meno reads per campione
 - Permette una migliore copertura dell'intera lunghezza di un trascritto
 - Non richiede un genoma o un trascrittoma di riferimento
 - Non richiede che le librerie vengano fornite di barcode se sequenziate sulla medesima lane
- 40) Un'analisi ChIP-seq individua come siti di legame di un fattore di trascrizione di interesse quelli dove si ottiene una copertura di reads più alta/più bassa rispetto ad un campione di controllo. Questa metodica richiede che la proteina venga crosslinkata con il DNA genomico/l'mRNA ed utilizza un passaggio di _____ per isolare le porzioni di acido nucleico legati dalla proteina di interesse.



41) Dal grafico sopra posso dire che

- Myl7 è downregolato VERO FALSO
- Frzb è espresso a livelli più alti di Sfrp1 nel campione di controllo VERO FALSO
- Myl7 sembra essere il gene con il fold change di entità maggiore di tutto l'esperimento, in valore assoluto VERO FALSO
- I geni Hoxa10, Hoxd9 e Hoxd13, se riportati in un volcano plot, sarebbero localizzati nel quadrante in alto a sinistra del grafico VERO FALSO