

Soluzioni del Capitolo 11

11.1 a. 5; b. 3; c. 8; d. 6; e. 2; f. 7; g. 1; h. 4; i. 10; j. 9.

11.2 Quando si esaminano marcatori di DNA anonimi, si osserva direttamente la sequenza di DNA di un individuo. I termini dominante e recessivo sono solitamente usati solo quando si discute del fenotipo di un organismo, quindi in un certo senso questa domanda è priva di significato. Inoltre, la maggior parte di questi loci di DNA anonimi non si trova nei geni e quindi non hanno alcun effetto sul fenotipo dell'organismo. Tuttavia, i genetisti spesso affermano che i marcatori del DNA sono ereditati in modo co-dominante per indicare che entrambi gli alleli possono essere visti nella sequenza del DNA e che il genotipo di un eterozigote dipende in ugual misura da entrambi gli alleli. Questo uso della parola co-dominante ha senso se si pensa al fenotipo come la sequenza di DNA al locus marcatore (SNP) o alla dimensione di una banda di un prodotto PCR su un gel (SSR).

11.3 È più probabile che gli SNP nelle regioni codificanti le proteine abbiano un effetto deleterio sull'organismo, poiché è più probabile che influiscano sulla funzione delle proteine, quindi si prevede che siano relativamente rari. Ancora più importante, la maggior parte degli SNP si trova nel DNA non codificante semplicemente perché il DNA non codificante negli esseri umani costituisce una proporzione molto maggiore del genoma (~98%) rispetto al DNA codificante (~2%).

11.4 a. Il genoma di un gamete è grande circa 3×10^9 bp. Questo numero moltiplicato per 1×10^{-8} sostituzioni di basi per bp per gamete produce ~30 sostituzioni di basi nel genoma di un gamete. Poiché gli individui sono formati da due gameti (un uovo e uno spermatozoo), il genoma di ciascuno dovrebbe contenere circa 60 nuove mutazioni per sostituzione di base, che non sono presenti nei genomi di nessuno dei genitori.

b. Queste nuove mutazioni devono essersi verificate nelle linee germinali dei genitori. Potrebbero essersi verificate durante le divisioni mitotiche che aumentano il numero di cellule della linea germinale (cioè negli spermatogoni o negli oogoni), o durante la meiosi (negli spermatociti o negli ovociti). Ogni divisione cellulare offre l'opportunità di accumulare mutazioni.

c. Più cicli di divisione cellulare avvengono prima della produzione di spermatozoi o uova, più mutazioni si troveranno nel gamete. Poiché per la produzione di spermatozoi sono necessari più cicli di divisione cellulare rispetto alle uova, lo spermatozoo dovrebbe contenere più mutazioni per sostituzione di base rispetto alle uova. (Poiché ogni gamete è il risultato di una singola meiosi, la maggior parte di queste divisioni avviene nelle cellule spermatogoniali e oogoniali.) Il tasso medio di sostituzioni di basi nei loci SNP, menzionato nel problema (1×10^{-8} sostituzioni di basi per bp per gamete), è in realtà la risultante di valori molto diversi per spermatozoi e uova: il tasso medio di nuove sostituzioni di basi dello spermatozoo è in realtà notevolmente superiore a quello dell'uovo. Pertanto, la maggior parte delle nuove mutazioni che si trova nel genoma, ma non nei genomi di nessuno dei genitori, deve essere stata ottenuta dallo spermatozoo.

Ricerche recenti suggeriscono che questa conclusione potrebbe spiegare i risultati secondo cui il tasso di autismo e di problemi comportamentali come il disturbo da deficit di attenzione è più alto tra i figli di padri più vecchi rispetto ai padri più giovani. Più vecchio è il padre, più cicli di divisione cellulare sono avvenuti prima che venissero prodotti gli spermatozoi. (Nei testicoli, le cellule spermatogoniali si dividono continuamente durante l'età adulta.) Pertanto, più il padre è anziano, più nuove mutazioni si

produrranno negli spermatozoi. (Un buon argomento per i maschi più giovani per congelare e conservare i loro spermatozoi). Al contrario, le femmine umane nascono con un corredo già completo di oociti (o quasi), quindi l'età materna non è un fattore nel numero di sostituzioni di basi che si accumulano nelle uova. (Naturalmente, le donne anziane sono soggette a un aumento dei tassi di non disgiunzione cromosomica che porta, per esempio, alla sindrome di Down tra i loro figli.)

- 11.5** Le regioni in cui Watson e Venter condividono gli stessi SNP implicano che questi due individui devono aver condiviso un antenato comune recente (entro diverse generazioni) da cui hanno ereditato questi SNP. L'intersezione di tali regioni di SNP condivisi con regioni in cui i due non condividono SNP è il risultato della ricombinazione. Cioè, un particolare cromosoma del loro recente antenato comune iniziò a ricombinare con altri cromosomi omologhi di altre discendenze che portavano a Watson o a Venter. Questi cromosomi omologhi avevano SNP diversi. Pertanto, i cromosomi odierni in Watson e Venter sono patchwork di parti ottenute dal comune antenato recente e parti ottenute da altri antenati che non erano comuni ai due uomini; la ricombinazione nel corso delle generazioni ha rimescolato in continuazione queste diverse parti.

Si deve capire che qualsiasi confronto tra i genomi di due persone oggi viventi mostrerà pattern in cui si alternano SNP simili e non simili. Più le persone sono strettamente imparentate, più lunghe saranno le regioni in cui gli SNP coincidono. Per le persone più distanti, come per esempio gli Africani rispetto agli Asiatici o agli Europei, le somiglianze saranno più difficili da trovare, ma saranno comunque presenti perché l'umanità come specie è piuttosto giovane.

- 11.6 a.** 3 miliardi – 100 milioni = 2,9 miliardi. Il punto centrale qui è che la maggior parte dei nucleotidi nel genoma umano è identica in tutte le persone.
- b.** Gli studi condotti sinora su migliaia di genomi umani avranno già rilevato la maggior parte degli SNP in cui entrambi gli alleli sono relativamente comuni nelle popolazioni umane (~15 milioni), ma molti altri alleli SNP rari che si troveranno solo in una o pochissime persone (per esempio, mutazioni che si sono verificate molto di recente) non sono stati ancora identificati. Per individuare tutte le mutazioni rare in tutti i genomi umani

sulla terra, si dovrebbero sequenziare i genomi di tutti gli esseri umani di oggi.

È interessante considerare perché la popolazione umana ha così tanti SNP rari e non così tanti SNP comuni. Si ricordi che l'accumulo di mutazioni dipende dalla divisione cellulare, quindi ogni nuovo individuo può avere decine di mutazioni non presenti nei genomi dei suoi genitori (vedi Problema 4). La popolazione totale degli esseri umani sulla terra era storicamente molto piccola, ma a partire dalla rivoluzione industriale, nel 1750 circa, si verificò una tremenda esplosione demografica. Molte nuove mutazioni sarebbero state introdotte nella specie durante questo breve periodo di tempo e queste mutazioni sono rare perché solo poche persone le condividono (la persona che per prima ha ereditato una mutazione rara non può avere avuto molti discendenti). Al contrario, le mutazioni che danno origine a SNP comuni devono essersi verificate molte generazioni fa, ma esistono relativamente poche di tali mutazioni perché la popolazione prima del 1750 era relativamente piccola.

c. La possibilità che una qualsiasi coppia di basi sia biallelica nel milione o più di esseri umani esaminati fino a oggi è $(1 \times 10^8)/(3 \times 10^9) \approx 0,33 \times 10^{-1} = 3,3 \times 10^{-2}$ o circa 1/30. La possibilità che un'altra mutazione avvenga in questa stessa coppia di basi sarebbe $(3,3 \times 10^{-2})^2 = 1,09 \times 10^{-3}$ o circa 1/917. Questo numero $(1,09 \times 10^{-3})$ per $3 \times 10^9 = 3,27 \times 10^6$, o circa 3 270 000 posizioni in cui 3 diversi alleli potrebbero essere trovati in qualsiasi posizione nucleotidica nei genomi di tutte le persone esaminate fino a oggi. Per 4 alleli, questo sarebbe $(3,3 \times 10^{-2})^3 \times 3 \times 10^9 \approx 10^8 \times 10^3$ o circa 108 000 posizioni. Si può vedere che per la grande maggioranza dei 100 milioni di loci SNP caratterizzati fino a oggi, esistono solo due alleli alternativi in questo campione di oltre 1 milione di persone.

I calcoli di cui sopra sono solo approssimazioni. Presumono che tutte le possibili mutazioni che cambiano una coppia di nucleotidi in un'altra siano ugualmente probabili, ma in realtà le transizioni sono circa il doppio delle trasversioni perché i meccanismi che generano spontaneamente le transizioni sono prevalenti. I calcoli per 3 o 4 diversi alleli, inoltre, non tengono conto del fatto che se nella popolazione esistono già 2 o 3 alleli, solo un sottoinsieme di possibili mutazioni cambierebbe la coppia di basi in un nuovo allele. (Per esempio, se gli alleli A, C e G fossero già pre-

senti nella popolazione, un allele preesistente potrebbe cambiare solo in T per rendere tetra-allelico il locus SNP.) Questi fattori tenderebbero a rendere i loci tri- e tetra-allelici ancora meno comuni nei genomi umani di quanto indicherebbero i calcoli precedenti.

D'altra parte, questo problema considera solo il milione di persone i cui genomi sono stati caratterizzati fino a oggi; la popolazione umana è attualmente più di 7 miliardi. Restano da scoprire molti alleli SNP, ma è probabile che quasi tutti siano estremamente rari.

11.7 a. Gli SSR sono polimorfismi in cui gli alleli differiscono per il numero di ripetizioni in tandem di una sequenza semplice lunga meno di circa 10 bp.

b. Gli SSR sono probabilmente generati da un fenomeno di stallo della DNA polimerasi (appaiamento errato dei filamenti di DNA) in corrispondenza di sequenze ripetute durante la replicazione del DNA. Questo meccanismo è stato illustrato nella Figura 6.11.

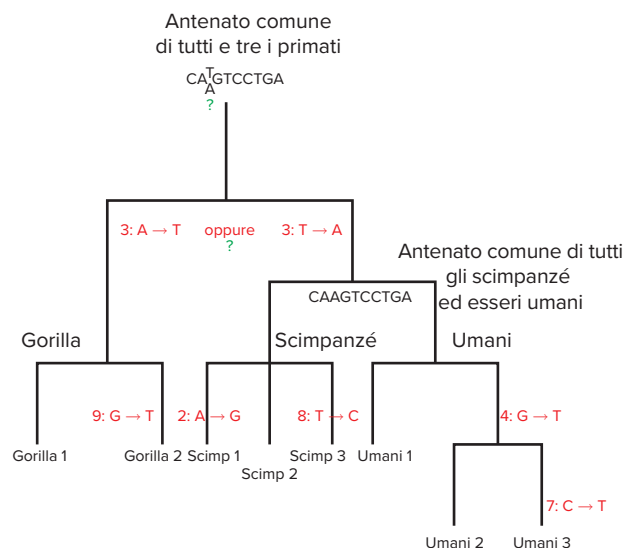
c. I CNV sono ripetizioni di unità superiori a 10 bp. È improbabile che lo stallo della DNA polimerasi generi questo tipo di cambiamenti. Al contrario, si pensa che il crossing-over ineguale tra le ripetizioni sia responsabile della maggior parte dei cambiamenti nel numero di ripetizioni CNV (vedi Figura 11.5). In contrasto con le unità ripetute brevi degli SSR, le unità ripetute dei CNV sono sufficientemente lunghe da potersi appaiare in modo errato nella profase della meiosi I, favorendo il meccanismo di crossing-over ineguale.

d. Gli SNP potrebbero potenzialmente verificarsi in uno qualsiasi dei 3 miliardi di bp nel genoma umano. Gli SSR potrebbero verificarsi solo in posizioni che già contengono diverse ripetizioni di una piccola sequenza di DNA. Pertanto, ci sono molti meno loci SSR rispetto ai loci SNP.

11.8 La figura che segue illustra le risposte alle parti (a), (b) e (c). In tutte le notazioni rosse, la lettera a sinistra della freccia è il nucleotide ancestrale e quella a destra della freccia è quello derivato. La sequenza nell'antenato comune di tutte e tre le specie è in alto. Per la posizione n. 3, non si può sapere se la sequenza ancestrale fosse A o T. Si sa che la mutazione che causa questo polimorfismo è nata dopo che la linea dei gorilla si è separata da quella che porta a scimpanzé/uomo.

Si noti che i dati indicano una notevole sottostruttura della popolazione umana. La mutazio-

ne $G \rightarrow T$ in posizione n. 4 si è verificata in un momento, e poi tra i discendenti che hanno ereditato quella mutazione si è verificata una mutazione in posizione n. 7.



11.9 a. Le coppie di basi omozigoti hanno la stessa identità in ogni lettura (per esempio, il 100% delle letture potrebbe essere una A in una determinata posizione). Al contrario, le coppie di basi eterozigoti hanno due identità diverse in reads diverse (per esempio, in una certa posizione, il 50% delle letture potrebbe essere A e il 50% G).

b. Le sequenze in eterozigosi indicano che i due genitori dell'animale sono dal punto di vista genetico strettamente imparentati, con genomi molto simili.

c. La perdita della diversità genetica potrebbe aver contribuito all'estinzione dei mammut lanosi. Come si vedrà nel Capitolo 24, questa perdita di diversità genetica potrebbe essere stata causata da un collo di bottiglia della popolazione, un evento in cui la popolazione è diventata molto piccola. Una popolazione con una bassa diversità genetica è una popolazione che non può affrontare con successo i cambiamenti ambientali. Per esempio, tutti i mammut rimasti sull'isola di Wrangel potrebbero essere stati suscettibili a un nuovo virus o parassita, causando la perdita dell'intera popolazione.

11.10 a. La sequenza del genoma umano mostra la sequenza dell'allele normale della PKU. Si vuole sapere se la sindrome PKU in questo paziente è causata o meno da una mutazione nel gene della fenilalanina idrossilasi. Si sospetta che ci possa essere una tale mutazione in questo particolare

esone, quindi si sequenzierà il prodotto della PCR. Se c'è una mutazione in questo esone da 1 kb, si vuole sapere esattamente di cosa si tratta, come influisce sull'enzima e forse qualcosa sulla storia di questa mutazione nelle popolazioni umane. Per esempio, se si confronta la sequenza in molti pazienti e si tiene traccia della loro provenienza, è possibile avere un'idea dell'origine di questa mutazione nel tempo storico e nella regione geografica. Se non si trova alcuna mutazione in questo esone da 1 kb che modifichi la sequenza amminoacidica dell'enzima, potrebbe esserci allora una mutazione in un esone diverso. Altre alternative includono la possibilità di una mutazione in una sequenza non codificante (che interessa la trascrizione o lo splicing) o una mutazione in un altro gene che potrebbe causare gli stessi sintomi.

b. Un genoma umano aploide contiene 3×10^9 bp. Pertanto $(3 \times 10^9 \text{ bp/genoma aploide}) \times (6,6 \times 10^2 \text{ g/mole}) \times (\text{mole}/6,02 \times 10^{23} \text{ bp}) = 3,3 \times 10^{-12} \text{ g/genoma aploide}$. In altre parole, un genoma aploide pesa $3,3 \times 10^{-12} \text{ g}$, cioè 3,3 picogrammi. Ciascun genoma aploide conterrà un solo gene della fenilalanina idrossilasi da utilizzare come stampo per la reazione PCR. Si avvia la reazione PCR con 1 ng ($1 \times 10^{-9} \text{ g}$) di DNA umano. Pertanto $(1 \times 10^{-9} \text{ g di DNA}) \times (1 \text{ genoma aploide}/3,3 \times 10^{-12} \text{ g}) \times (1 \text{ molecola stampo}/1 \text{ genoma aploide}) = 0,3 \times 10^3 \text{ molecole di stampo} = 300 \text{ molecole di stampo in } 1 \text{ ng di DNA}$.

c. Si inizia la PCR con 300 molecole di stampo. Se la PCR viene eseguita per 25 cicli, questo numero di molecole raddoppia esponenzialmente 25 volte. Pertanto, ci si ritroverà con $300 \text{ molecole} \times 2^{25} \approx 10^{10}$, cioè circa 10 miliardi di molecole. Questo risultato spiega il potere della PCR: si è iniziato con solo 300 molecole di stampo e si finisce con 10 miliardi di copie della regione che si sta amplificando. In pratica, le rese non sono così elevate perché non tutte le potenziali molecole stampo vengono amplificate a ogni ciclo. Tuttavia, l'amplificazione è ancora molto significativa. Il prodotto PCR è lungo 1 kb, quindi $(10^{10} \text{ molecole di prodotto PCR}) \times (10^3 \text{ bp/molecola di prodotto PCR}) \times (\text{mole}/6,02 \times 10^{23} \text{ bp}) \times (6,6 \times 10^2 \text{ g/mole}) = 1,1 \times 10^{-8} \text{ g} = 110 \text{ ng}$. Si è iniziato con 1 ng dell'intero genoma e alla fine si ottengono 110 ng di una regione di 1 kb del genoma dopo la PCR!

d. A ogni ciclo, il numero di molecole di DNA a doppio filamento presenti nella PCR raddoppia. Se lasciamo n = numero di molecole di dsDNA

alla fine della reazione, t = numero di molecole stampo all'inizio delle reazioni e c = numero di cicli: $n = t \times 2^c$.

11.11 I due primer devono fiancheggiare la sequenza target e dovrebbero essere orientati in modo che i loro orientamenti da 5' a 3' disegnati come frecce, siano rivolti l'uno verso l'altro (con le estremità 3' verso il centro). Questo requisito riflette il fatto che la DNA polimerasi aggiunge nucleotidi in sequenza alle estremità 3' dei primer. È anche necessario che la DNA polimerasi estenda il primer nella direzione della sequenza che deve essere amplificata. Solo il set b soddisfa questi criteri.

11.12 a. Se il DNA umano fosse una sequenza casuale di quantità uguali di A, C, G e T (non è esattamente così), allora la possibilità che uno dei primer si appai in una regione casuale del DNA sarebbe $(1/4)^{18}$, cioè 1 possibilità su 7×10^{10} . In altre parole, qualsiasi particolare sequenza di 18 basi si verificherebbe a caso solo una volta ogni 70 miliardi di nucleotidi. Poiché il genoma umano è lungo 3 miliardi di coppie di nucleotidi, è molto improbabile che anche uno solo dei primer si rinaturi in qualsiasi altra parte tranne che nel gene CFTR. La probabilità è ancora molto più bassa che entrambi i primer si riappaino ad altri tratti casuali di DNA che sono abbastanza vicini tra loro da consentire la formazione di un prodotto PCR. (Questa probabilità è difficile da calcolare esattamente a causa della variazione della possibile distanza tra i due primer.)

b. Il limite inferiore della lunghezza dei primer è governato da diversi fattori. Il più importante è che i primer devono appaiare al DNA genomico. I legami idrogeno normalmente necessari per consentire al DNA di formare una doppia elica stabile sono quelli forniti da 15 coppie di nucleotidi. Un secondo fattore è che i primer devono essere sufficientemente lunghi in modo che la PCR amplifichi specificamente solo il DNA bersaglio. Usando la stessa logica che abbiamo usato nella parte (a), la probabilità che un primer di 16 nucleotidi si appai casualmente al DNA è $(1/4)^{16} = 1$ su 4 miliardi. Poiché ci sono solo 3 miliardi di nucleotidi nel genoma aploide umano, un primer da 16 bp probabilmente si ibriderà con una sola sequenza nel genoma umano. Un particolare primer da 15 bp (al contrario di un primer da 16 bp) sarà presente ogni $(1/4)^{15}$ basi, ovvero circa 1 volta ogni miliardo di bp. Ciò significa che è probabile che la

sequenza di un primer da 15 bp ricorra 3 volte nel genoma umano. La possibilità che due di questi primer si appaiano a DNA casuale abbastanza vicini tra loro per permettere l'amplificazione della PCR è estremamente bassa, come discusso nella parte (a). Tuttavia, poiché è probabile che un primer di almeno 16 bp si ibridi a una sequenza genomica unica, la maggior parte dei ricercatori ha individuato 16 bp come limite inferiore.

È interessante notare (sebbene non richiesto dal problema), che anche primer PCR molto lunghi (più di 40 nucleotidi) possono creare alcuni problemi. In particolare, i primer più lunghi potrebbero ibridarsi con sequenze di DNA genomico a cui non corrispondono perfettamente. (I cattivi appaiamenti interni sono tollerati e l'ibridazione può verificarsi fintanto che ci sono abbastanza nucleotidi appaiati nelle basi intorno, in particolare all'estremità 3' del primer.) Pertanto, primer molto lunghi potrebbero amplificare regioni del genoma diverse dal bersaglio. I primer lunghi potrebbero anche formare strutture a forcina che ne diminuirebbero l'efficienza.

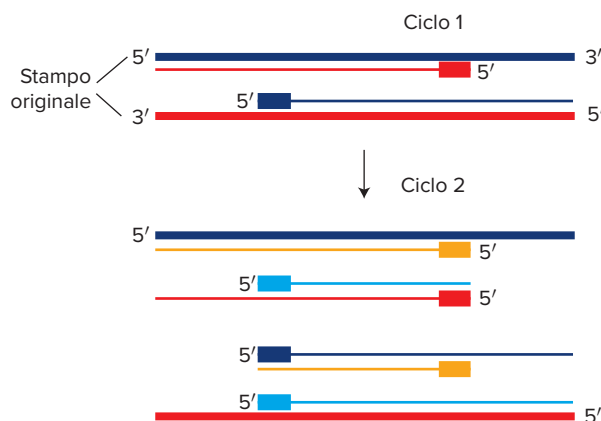
c. Sarebbe più probabile ottenere un prodotto PCR se la mancata corrispondenza fosse all'estremità 5'. L'estremità 3' del primer è l'estremità cui la DNA polimerasi aggiunge nucleotidi nuovi alla catena. Pertanto, i disallineamenti all'estremità 3' ostacolerebbero la DNA polimerasi nell'aggiungere di nuovi nucleotidi, ma questo non è vero per l'altra estremità del primer. (Si noti che questa domanda è importante perché significa che è possibile aggiungere nucleotidi non correlati allo stampo alle estremità 5' dei primer PCR. In questo modo, per esempio, è possibile aggiungere un sito di riconoscimento di un enzima di restrizione alle estremità del prodotto PCR per facilitare il clonaggio.)

- 11.13** In primo luogo, si genera un cDNA dagli mRNA purificati dalle cellule retiniche. Se la lunghezza totale dei 16 esoni del gene di interesse non supera 1 kb, o giù di lì, si può provare ad amplificare per PCR i cDNA completi usando coppie di primer in cui un primer è complementare alla coda poli-A e l'altro è complementare, sull'altro filamento all'inizio previsto dell'esone 1 (che di solito è di circa 25-35 bp a valle del promotore [TATA-box]). Se il gene ha molti prodotti di splicing alternativo, purché abbiano in comune le sequenze dell'esone 1 che sono incluse nel primer PCR, verranno amplificati prodotti PCR diversi

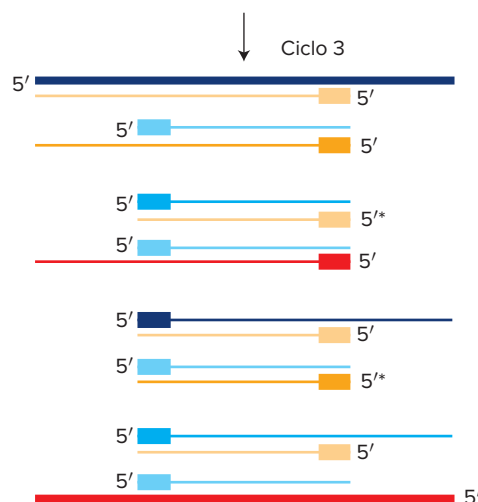
nella reazione, ciascuno corrispondente a un diverso prodotto di splicing del trascritto primario. Il sequenziamento di ogni prodotto rivelerà le giunzioni esone/introne.

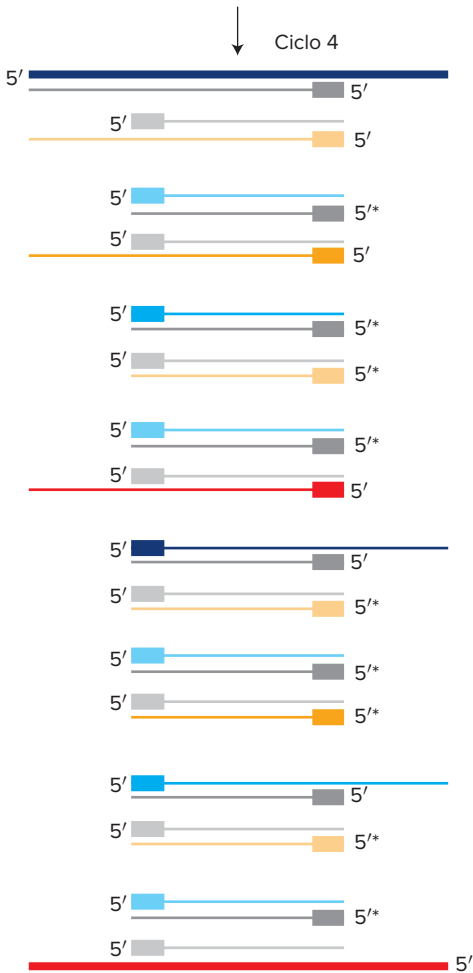
Se i cDNA sono troppo lunghi per essere amplificati in un'unica reazione, si dovranno amplificare dei cDNA parziali con diverse coppie di primer.

- 11.14 a.** Dopo i primi due cicli di PCR, nessuna delle molecole di DNA a doppio filamento inizia e termina anche con un primer:



- b. Perché a ogni ciclo di PCR, l'estremità 5' di ogni filamento di DNA appena sintetizzato inizia con un primer; con ogni ciclo di PCR vengono create sempre più molecole a doppio filamento che iniziano e finiscono con i primer. Come si può vedere nella figura che segue, 2 di queste molecole (indicate da *) sono prodotte al ciclo 3, ma già al ciclo 4 esistono 9 di queste molecole. Se si analizzano i prodotti del ciclo 4, si può prevedere che al ciclo 5 saranno 24 le molecole di DNA a doppio filamento che iniziano e finiscono con un primer.





11.15 a. Entrambi i primer avrebbero bisogno della sequenza 5' NNNNNGAATTC 3' aggiunta alle loro estremità 5' (N = qualsiasi base). Per esempio, uno dei primer potrebbe essere:

5' NNNNNGAATTCGCTACTTCGCGTATTCCA 3'

La sequenza di N può essere più lunga, ma non può essere meno di 6 basi. Il Problema 11.12, parte (c), spiega perché queste sequenze dovrebbero essere aggiunte alle estremità 5' dei primer.

b. La regione del genoma umano non può contenere un sito *EcoRI*; in tal caso, infatti, se digeriti con *EcoRI* per creare estremità appiccicose, i prodotti di amplificazione verrebbero tagliati in due parti e quindi non potrebbero essere clonati come unità nel vettore.

c. La stessa sequenza della parte (a) dovrebbe essere aggiunta a un primer e la sequenza 5' NNNNNNGGATCC 3' dovrebbe essere aggiunta all'altro primer. Se si digerisce il prodotto della PCR sia con *EcoRI* sia con *BamHI*, sarà possibile integrare gli inserti in un plasmide in una direzione particolare, a condizione che anche il pla-

smide venga tagliato nel suo poly-linker con entrambi gli enzimi. La direzione dell'inserto in un plasmide a volte può essere importante. Per esempio, se il plasmide verrà utilizzato per esprimere una proteina nei batteri, un inserto di cDNA deve avere un orientamento specifico rispetto al promotore.

11.16 a. Se si riscontra un doppio picco e questo rappresenta un errore, gran parte del prodotto PCR deve presentare lo stesso errore. Questo scenario potrebbe verificarsi solo se l'errore si è verificato all'inizio dell'amplificazione PCR ed è stato quindi copiato per molti cicli PCR. Se l'errore fosse avvenuto più tardi, solo una piccola parte delle molecole amplificate lo avrebbe contenuto. Probabilmente queste poche molecole non sarebbero rilevabili tra la proporzione molto più ampia di quelle senza l'errore.

b. È più probabile che si verifichi un errore in uno degli ultimi cicli di PCR, perché ci sono più molecole presenti e quindi più possibilità di errore. Fortunatamente, più tardi si verificano questi errori, meno è probabile che siano visibili perché solo una piccola parte del prodotto PCR avrebbe questo errore, come spiegato nella parte (a).

c. Fondamentalmente, si potrebbe determinare il genotipo della persona più volte con reazioni PCR indipendenti. Questi errori sono così rari che è quasi impossibile che lo stesso errore si verifichi in più reazioni PCR indipendenti. Non si possono eseguire facilmente più PCR indipendenti nella genotipizzazione pre-impianto, perché si dovrebbero usare più cellule singole dallo stesso embrione di circa 8 cellule.

d. Questa funzione di esonucleasi permetterebbe alla DNA polimerasi di *P. furiosa* di riparare molti degli errori commessi nell'incorporazione delle basi e quindi alla fine si troverebbero meno errori di quanti se ne rileverebbero con l'enzima di *T. aquaticus*, quindi la reazione PCR sarebbe più accurata e affidabile.

11.17 Ricordiamo che le persone sono diploidi e quindi hanno due copie dei loci autosomici. Potrebbero essere omozigoti per una qualsiasi delle tre sequenze nella tabella, oppure potrebbero essere eterozigoti per due delle tre sequenze. Se si stava eseguendo il sequenziamento Sanger di prodotti PCR, per gli eterozigoti si vedrebbero picchi doppi in cui le due sequenze in un eterozigote sono diverse.

a. Le tre possibili risposte sono quindi:

CAAGTCCTGA/CAATTCCTGA (1/2) +
CAATTCTGA/CAATTCTGA (3/3)

CAAGTCCTGA/CAAGTCCTGA (1/1) +
CAATTCCTGA/CAATTCCTGA (2/3)

CAAGTCCTGA/CAATTCCTGA (1/2) +
CAATTCCTGA/CAATTCCTGA (2/3)

Si noti che i dati di sequenza degli eterozigoti 1/2 o 2/3 rivelano inequivocabilmente entrambe le sequenze aploidi da 10 nt perché c'è solo un doppio picco, cioè un solo nucleotide variabile. Poiché non ci sono doppi picchi nelle sequenze degli omozigoti (1/1 o 3/3), i tracciati di sequenza (ferrogrammi) per questi individui rivelano entrambe le sequenze aploidi (che sono ovviamente identiche) nei loro genomi.

b. Se una persona era (1/3), ci sono due differenze e quindi due doppi picchi. Una possibile soluzione è:

CAAGTCCTGA/CAATTCCTGA (1/3)

Tuttavia, ferrogramma è compatibile anche con:

CAAGTCCTGA/CAATTCCTGA (4/2)

dove 4 è un nuovo allele che non è né 1, 2 o 3. Non è possibile distinguere queste due possibilità usando questa metodologia. Non si potrebbe concludere che una persona il cui tracciato è quello prodotto da (1/3) sia in effetti (1/3), perché potrebbe essere (4/2).

Si noti che questo problema non avrebbe senso se si esaminassero i genotipi delle persone mediante sequenziamento del DNA a singola molecola (come descritto in Figura 11.22) perché ogni sequenza proverrebbe da un singolo cromosoma, non da una miscela dei due cromosomi omologhi, come è il caso del materiale amplificato per PCR.

- 11.18** La malattia di Huntington è una malattia letale dominante a esordio tardivo. La malattia è associata all'espansione di una ripetizione trinucleotidica CAG nel gene *HD*. Come indicato nel testo, gli individui normali hanno fino a 35 copie della tripletta. Gli individui con regioni ripetute contenenti più di 35 ripetizioni sono suscettibili alla malattia di Huntington; tutti gli alleli della malattia con 42 o più ripetizioni sono completamente penetranti. In generale, un numero maggiore di ripetizioni è correlato con un'età più giovane di insorgenza della malattia (vedi Figura 11.11).
- a.** Gli individui A, B, C ed E presentano tutti una banda PCR (un allele) che è molto più grande (lunga tra 270 e 380 nucleotidi) rispetto al secondo allele (lunga tra 200 e 220 nucleotidi). Non è possibile correlare queste dimensioni di banda

con i numeri di ripetizione del trinucleotide perché non conosciamo la posizione (relativa alle ripetizioni) delle sequenze a cui si rinaturano i primer PCR. In ogni caso, queste bande molto più lunghe devono avere più copie della ripetizione rispetto alle bande più piccole. La regione ripetuta nell'individuo B è la più lunga; quindi è probabile che questa persona abbia l'insorgenza più precoce della malattia.

b. Entrambe le bande PCR degli individui D e F sono più piccole, quindi sembra che abbiano ricevuto alleli *HD*⁺ da entrambi i genitori. Pertanto, non manifesteranno la malattia.

c. Questo problema può essere risolto solo se si conosce il numero di ripetizioni in un allele. Se assumiamo che l'allele *HD*⁺ più lungo mostrato nella figura (la banda più piccola nell'individuo B) contenga il numero massimo di ripetizioni per un allele non patogeno (35 ripetizioni; vedi Figura 11.11), allora le 220 bp di questo Il prodotto PCR dovrebbe includere $35 \times 3 = 105$ bp di ripetizioni CAG, più 70 bp tra l'estremità 5' di un primer PCR e la ripetizione CAG più vicina. Pertanto, $220 - 175 = 45$ bp rimarranno dalla fine delle ripetizioni CAG all'estremità 5' del secondo primer PCR. Si noti che la distanza tra i primer e l'inizio e la fine delle ripetizioni SSR è la stessa per tutti gli alleli.

11.19 a. Nel paziente il cui grafico è mostrato in alto, il numero di ripetizioni CAG nell'allele *HD*⁺ è di circa 15; le cellule somatiche del paziente sul fondo hanno circa 20 ripetizioni CAG nell'allele *HD*⁺.

b. Questi risultati dicono molto sui meccanismi che danno origine agli alleli mutanti *HD*. Innanzitutto, il numero di ripetizioni varia tra i diversi spermatozoi dello stesso individuo, quindi i processi di espansione hanno luogo nelle cellule della linea germinale durante la spermatogenesi. In secondo luogo, sembra che maggiore è il numero originale di ripetizioni in un allele *HD*, più è probabile che il numero di ripetizioni negli spermatozoi vari e maggiore è il grado di variazione potenziale. Nel grafico in alto, quasi tutti gli spermatozoi con l'allele mutante *HD* hanno più di 62 ripetizioni, alcuni spermatozoi hanno più di 120 ripetizioni e la distribuzione è piuttosto ampia. La distribuzione delle dimensioni degli spermatozoi mutanti nel grafico in basso è più stretta. Inoltre, si noti il fatto che vi è poca o nessuna variazione nella dimensione negli spermatozoi contenenti un allele *HD*⁺. Terzo, è interessante che gli spermatozoi per lo più sembrano accumulare più ripetizioni CAG piuttosto che perdere

ripetizioni CAG: pochi spermatozoi hanno un numero minore di ripetizioni CAG rispetto ai geni *HD* delle cellule somatiche.

c. C'è una certa probabilità che il numero di ripetizioni CAG possa espandersi durante la spermatogenesi, anche partendo da un numero inferiore di ripetizioni CAG. Un uomo con un allele *HD* nella parte alta del range di normalità (~30 ripetizioni CAG) o nell'area grigia tra 36 e 41 ripetizioni (che potrebbe non mostrare alcun sintomo), può produrre spermatozoi con più di 42 ripetizioni. Si noti che questi dati predicono che il padre non affetto (ma non la madre non affetta, perché questa particolare espansione ripetuta si verifica durante la spermatogenesi) di un paziente con malattia di Huntington dovrebbe avere un allele compreso tra 30 e 42 ripetizioni. È interessante notare che nella sindrome dell'X fragile (un'altra malattia da espansione dei trinucleotidi), l'espansione si verifica solitamente durante la meiosi nella madre (vedi Figura 6.10b).

d. Ci si aspetterebbe che ciascuno dei due campioni di cellule del sangue mostri una distribuzione molto stretta dei numeri di ripetizione CAG attorno ai numeri rilevati negli alleli *HD* normali e mutanti (15 e 62 per il paziente in alto; 20 e 48 per il paziente in basso). Dovrebbe esserci una variazione molto piccola nel numero di ripetizioni tra cellule diverse del sangue della stessa persona, poiché l'espansione nel numero di ripetizioni sembra essere limitata alla linea germinale e non si verifica nelle cellule somatiche.

11.20 Le prove furono utilizzate per dimostrare che il camioncino dell'imputato si trovava in un luogo specifico nel deserto dell'Arizona dove cresceva il particolare albero di Palo Verde. Presumibilmente questa posizione era molto vicina al luogo in cui è stato scoperto il cadavere.

11.21 a. In generale, gli SNP sono biallelici; hanno solo 2 alleli nelle popolazioni umane (vedi il Problema 11.6). Gli SSR utilizzati nel CODIS hanno molti alleli diversi che variano nel numero di unità ripetitive. Questo significa che la possibilità che due persone qualsiasi condividano lo stesso insieme di alleli per 14 loci SNP è molto più alta della possibilità che condividano lo stesso insieme di alleli per 14 loci SSR.

Nel Capitolo 23 si parlerà dell'equilibrio di Hardy-Weinberg, che permette di calcolare la probabilità che due persone scelte casualmente condividano gli stessi alleli di un qualsiasi polimorfismo, purché si conosca la frequenza di quell'allele tra tutti i alleli nella popolazione studiata. Na-

turalmente, minore è la frequenza con cui gli alleli nel soggetto del test si trovano nella popolazione, minore è la probabilità che un individuo casuale della popolazione abbia esattamente gli stessi alleli. Poiché i loci CODIS non sono in linkage tra loro (e quindi vengono ereditati in modo indipendente), è possibile moltiplicare fra loro le basse probabilità di una corrispondenza degli alleli per ciascun locus, ottenendo una probabilità estremamente bassa di una corrispondenza con gli alleli di tutti i 14 loci CODIS. Questa logica è alla base del fingerprinting del DNA.

b. I cani e altri animali domestici presentano elevata consanguineità. Ciò significa che è molto più probabile che due cani qualsiasi condividano gli stessi alleli di qualsiasi polimorfismo, che non due umani qualsiasi; inoltre, è molto più probabile che i cani siano omozigoti a un certo locus. Si dovrebbero quindi esaminare molte più sequenze di DNA per trovarne una che mostri variazione tra due cani qualsiasi.

11.22 a. C: Zar; E: Zarina; ABDH: Figlie; G: Figlio (Tsarevitch); F: Servitore non imparentato. Non esiste un modo completamente sistematico per analizzare questi dati, quindi alla fine si dovranno selezionare le alternative più probabili e vedere quali si adattano. Il punto di partenza più semplice è con i maschi (C e G), uno dei quali potrebbe essere lo zar Nicola II e l'altro potrebbe essere lo Tsarevitch Alexei. Poiché G ha due alleli per D3S1358, nessuno dei quali si trova in A e B (che devono essere sorelle perché condividono tutti gli alleli a tutti i loci tranne uno), G non trasmette alcun allele a queste figlie e quindi G potrebbe essere lo Zarevitch ma non lo Zar. Una volta fatta questa ipotesi, il resto dell'analisi diventa semplice. Un modo alternativo per iniziare è cercare l'individuo non imparentato, che deve essere F perché ha almeno un allele di ogni locus che non si trova in nessuna delle altre persone.

b. B (12,12); C (11,12); E (9,11)

c. No, nessuna delle figlie ha alleli identici in tutti i loci.

d. Le dimensioni degli scheletri potrebbero riflettere l'età delle figlie. Forse esistono impronte dentali che potrebbero corrispondere ai denti di uno o più scheletri.

e. Tutte le figlie sono presenti negli scheletri, quindi Anastasia doveva essere uno di essi.

f. $1/8 = 12,5\%$. (La Zarina e la nonna di Filippo dovrebbero condividere $1/2$ dei loro alleli perché sono sorelle. C'è una probabilità su due per la madre del principe Filippo di ereditare un allele co-

mune dalla nonna di Filippo e una probabilità su due per Filippo di ereditare un allele comune di sua madre: $1/2 \times 1/2 \times 1/2 = 1/8$.)

g. $1/16 = 6,25\%$. Un altro fattore di $1/2$ per lo Zarevitch di ereditare un allele comune dalla Zarina.

h. La zarina era la prozia del principe Filippo.

11.23 a. In questo problema, non si prende in considerazione il singolo spermatozoo, ma lo sperma, che contiene milioni di spermatozoi diversi. Se l'uomo è eterozigote per due alleli di un locus SSR, circa metà dei suoi spermatozoi conterranno un allele e l'altra metà l'altro allele.

b. Il locus viola è sul cromosoma X. Si noti che c'è solo una copia allelica nel campione di liquido seminale e una copia nei tamponi orali degli individui 1 e 4. Queste due persone devono essere maschi. Gli individui 2 e 3 hanno due copie del locus viola, quindi 2 copie del cromosoma X; queste due persone sono femmine.

c. Il locus verde è sul cromosoma Y. C'è una copia nel liquido seminale e nei genomi somatici dei maschi. Gli individui 2 e 3 non hanno bande verdi perché, essendo femmine, non hanno un cromosoma Y.

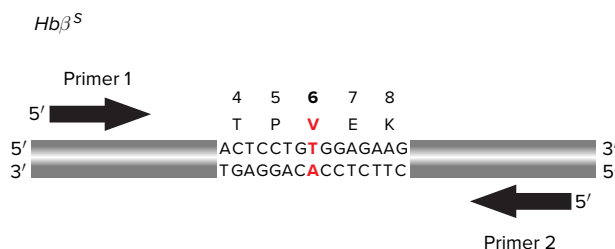
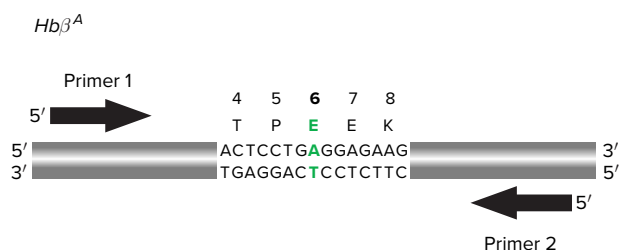
d. Se un individuo fosse lo stupratore tutte le bande di tutti i colori dovrebbero corrispondere al campione di sperma.

e. Da notare che l'individuo 2 corrisponde allo stupratore per 5 alleli su 10 (uno nero, due rossi e due arancioni). È probabile che l'individuo 2 (una donna come stabilito nella parte b) sia la sorella dello stupratore.

f. Le bande sono lunghe 200 e 212 bp. Ciascun primer per la PCR è lungo 20 bp; La PCR richiede due primer. L'unità di ripetizione è 4 bp. Pertanto, l'allele da 200 bp avrebbe $(200 - 40) = 160/4 = 40$ ripetizioni. L'allele 212 avrebbe 3 ripetizioni in più, cioè 43.

11.24 a. La tecnica di ibridazione su microarray non è molto sensibile, quindi è necessario amplificare la regione di interesse per generare una sonda marcata in fluorescenza in quantità sufficiente per vedere eventuali segnali.

b. Questa domanda essenzialmente chiede di ridisegnare la Figura 11.9, riprodotta di seguito.



c. A 80°C , il legame idrogeno tra i due filamenti di DNA (un filamento sta sul chip di silicio e l'altro è la sonda fluorescente) viene inibito dal calore, quindi le sonde non possono legarsi al DNA sul chip.

d. La temperatura di incubazione influisce sull'accuratezza della rinaturazione tra sequenze complementari. Se la temperatura è sufficientemente bassa, una mancata corrispondenza non influirà sulla capacità dei due filamenti di DNA di rinaturare e rimanere uniti.

e. Gli embrioni 2 e 3 sono omozigoti per l'allele A (AA); gli embrioni 1, 5 e 7 sono eterozigoti AS e gli individui 4 e 6 sono omozigoti SS. Per evitare che il bambino abbia l'anemia falciforme, si devono scegliere gli omozigoti AA (embrioni 2 e 3). Se fossero disponibili solo pochi embrioni, si potrebbero anche impiantare gli eterozigoti AS (che di solito sono fenotipicamente normali, tranne che in circostanze eccezionali, come una lunga esposizione a basse quantità di ossigeno). Non si dovrebbero scegliere gli omozigoti SS, poiché alla nascita avrebbero l'anemia falciforme.

Nei primi tempi della fecondazione in vitro, i medici impiantarono diversi embrioni (a volte più di 5) nell'utero della futura gestante. Poiché gli effetti negativi delle nascite multiple sono diventati sempre più evidenti, il numero di embrioni impiantati è stato ridotto; oggi, nella maggior parte dei casi di genotipizzazione preimpianto, non vengono scelti più di due o tre embrioni.

11.25 I due ASO compatibili per l'allele *Hb β^A* sono:

5' CTGACTCCTGAGGAGAAGTCG 3'
o 3' GACTGAGGACTCCTCTTCAGC 5'

I due ASO compatibili per l'allele *Hb β^S* sono:

5' CTGACTCCTGTGGAGAAGTCG 3'
o 3' GACTGAGGACACCTCTTCAGC 5'

11.26 a. Un paziente il cui allele *HD* ha uno dei due particolari alleli SNP solitamente associati all'*HD*⁻ verrà sottoposto a iniezione di ASO che includono questo particolare allele SNP. Gli ASO sono quindi complementari solo all'*mRNA HD*⁻

(e non all'mRNA dell'allele normale HD^+). Poiché parte dell'mRNA HD^- formerà un ibrido DNA-RNA, l'mRNA HD^- sarà degradato dalla RNasi H.

b. I pazienti dovrebbero avere uno dei due particolari alleli SNP solitamente associati all'allele della malattia. La PCR e il sequenziamento di ciascun locus SNP sarebbero il primo passo per determinare se sia presente uno dei due particolari alleli SNP. Se uno di questi lo è, i ricercatori dovrebbero quindi assicurarsi che l'allele SNP a cui si lega l'ASO sia all'interno dell'allele della malattia e non all'interno del normale allele HD^+ . Un modo per stabilire questo fatto è creare una genoteca dalle cellule somatiche del paziente, iniziando con frammenti di DNA genomico abbastanza grandi da contenere sia il locus SNP in questione sia la regione di ripetizione CAG del gene HD. Esistono molti metodi per identificare i plasmidi ricombinanti che contengono le parti rilevanti del gene HD. Una volta identificati, i singoli inserti di DNA genomico verrebbero sequenziati per determinare se la regione di ripetizione CAG espansa e l'allele SNP complementare all'ASO sono presenti nella stessa copia del gene.

c. Solo i pazienti che hanno almeno uno dei due particolari alleli SNP complementari all'ASO all'interno di un esone dell'allele della malattia, e non l'allele normale, sarebbero candidati per questo farmaco. Gli scienziati potrebbero essere in grado di progettare ASO specifici per il paziente. Determinando la sequenza del DNA genomico degli alleli HD^+ e HD^- di un paziente specifico, i ricercatori potrebbero essere in grado di trovare un locus SNP eterozigote in un esone HD che potrebbe essere distinto dagli ASO.

11.27 a. Nella Figura 11.15b, è necessario un solo primer perché la stessa sequenza adattatore è stata aggiunta a entrambe le estremità del frammento prima della PCR. Il primer da usare per la fase di amplificazione della PCR è complementare quindi a entrambe le estremità 3' del frammento sulle due eliche opposte.

b. È necessario tagliare il DNA con un enzima di restrizione che produce estremità appiccicose; questa scelta rende molto semplice aggiungere l'adattatore alle due estremità. (L'adattatore va realizzato con un'estremità netta e una coesiva, come mostrato nella figura.) Non è rilevante se l'estremità coesiva ha l'estremità sporgente al 5' o al 3'. È opportuno tagliare il DNA con un enzima che ha un sito di riconoscimento di 4 bp. Il motivo qui è principalmente tecnico. La PCR può ampli-

ficare solo regioni di DNA inferiori a circa 10 kb. Pertanto, se si utilizza un enzima che riconosce 6 bp o 8 bp, gran parte, o anche la maggior parte, dei frammenti di restrizione del DNA genomico sarebbe troppo grande per l'amplificazione PCR.

11.28 a. Padre: 1-AC 2-GG 3-AG 4-GT 5-CC 6-CG 7-AG 8-GT; Madre: 1-AA 2-CG 3-GG 4-AT 5-CC 6-CC 7-AG 8-TT

b. Il locus SNP 4, con gli alleli A, G e T.

c. Si conosce l'intera sequenza del genoma umano. Gli ASO sul microarray sono stati sintetizzati basandosi su questa conoscenza.

d. Se gli 8 loci sono ciascuno a una distanza di circa 10 Mb, SNP1 e SNP8 sarebbero distanti circa 70 Mb. $70/191 = 36\%$ del cromosoma 4.

11.29 a. Un uistiti chimerico potrebbe avere 4 diversi alleli di un singolo locus SSR se i suoi genitori fossero eterozigoti senza alleli in comune e se quell'uistiti e il suo gemello dizigote ereditassero da ciascun genitore tutti alleli diversi per quel locus.

b. Il maschio potrebbe aver generato tutti e 4 i cuccioli perché ognuno di loro ha almeno uno dei suoi alleli a tutti e quattro i loci SSR mostrati.

c. Cucciolo 2 e Cucciolo 3 sono gemelli dizigoti chimerici. Possiamo affermare che sono chimerici perché ogni bambino ha più di 2 alleli a tre dei loci SSR. I dati sono coerenti con il fatto che questi due animali sono una coppia di gemelli dizigoti chimerici perché condividono tutti i loro loci SSR. Da notare che sebbene i cuccioli 2 e 3 abbiano le stesse bande, le intensità relative di queste bande non sono le stesse nei due animali. Il motivo è che i due animali non hanno le stesse proporzioni di cellule chimeriche che hanno avuto origine dai due zigoti.

d. Se la madre o il padre dei gemelli fossero chimerici, la loro linea germinale potrebbe essere chimerica e potrebbero produrre gameti con fino a quattro diversi alleli di un singolo SSR. Poiché Cucciolo 2 e Cucciolo 3 ereditano ciascuno solo 1 allele per ciascun locus SSR dai loro genitori, il numero massimo di alleli rilevabili nei due cuccioli è due da ciascun genitore. Pertanto, i profili del DNA dei gemelli dizigoti non forniscono informazioni sufficienti sul fatto che uno dei loro genitori fosse o meno chimerico.

e. Se la madre e suo fratello fossero gemelli dizigoti, alcune delle sue cellule germinali potrebbero provenire da suo fratello e alcune delle sue uova potrebbero quindi contenere il suo genoma.

11.30 a. Condividi esattamente metà del tuo DNA con ciascuno dei tuoi genitori e ciascuno dei tuoi figli. Il motivo è che ciascuno dei tuoi genitori ti ha

dato una delle sue copie di ciascuno dei loro cromosomi (ricombinati) nei loro gameti. Allo stesso modo, hai dato a ciascuno dei tuoi figli uno dei tuoi due cromosomi (ricombinati).

Per tutti gli altri tuoi parenti, la frazione di DNA che condividi sarà in media quella mostrata nella Figura B, che però è solo una valutazione approssimativa. Le ragioni sono che ogni gamete contiene una combinazione casuale di cromosomi ereditati per via materna e per via paterna (definiti dai centromeri) e anche che i cromosomi ricombinano. Anche se i cromosomi non hanno ricombinato, l'assortimento indipendente delle nostre 23 paia di cromosomi (uno materno e uno paterno ereditato) nei gameti significa che esistono $2^{23} = 8\,388\,608$ gameti diversi, con diverse combinazioni dei 23 cromosomi.

Nella Figura A del box Uno sguardo in avanti si possono osservare le conseguenze dell'assortimento indipendente dei cromosomi e della ricombinazione sulla correlazione genetica con alcuni dei tuoi parenti. Sebbene tu e tuo fratello abbiate ereditato esattamente il 50% del DNA di ciascuno dei genitori, potresti non condividere esattamente il 50% del tuo DNA con tuo fratello. Il motivo è che ciascuno dei gameti dei genitori conteneva una miscela casuale dei loro genomi (il loro DNA ereditato da madre e padre). A un estremo, è teoricamente possibile che tu e tuo fratello non abbiate alcuna sequenza di DNA in comune: ognuno di voi potrebbe aver ereditato una copia diversa del genoma dei due genitori per ogni segmento di ogni cromosoma. All'altro estremo, tu e tuo fratello potreste in teoria aver ereditato il 100% dello stesso DNA: avreste potuto ereditare la stessa copia di ciascuna delle due copie del genoma dei vostri genitori per ogni segmento di ogni cromosoma (cioè, sia i gameti maschili sia quelli femminili che sono stati usati per produrre ognuno di voi era identico). Poiché il numero di tipi di gameti diversi è così grande, ciascuno di questi due estremi è altamente improbabile. Mediamente, i fratelli condividono circa il 50% del loro DNA. Un ragionamento simile spiega perché condividi non esattamente il 25%, ma in media il 25% del tuo DNA con zie, zii, nonni e così via.

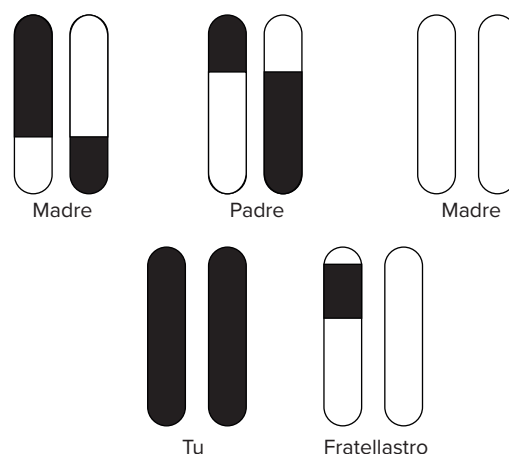
b. Nella Figura B si può osservare che ci si aspetta che parenti diversi condividano le stesse frazioni di DNA: i tuoi fratelli e i tuoi figli condividono il 50% del tuo DNA, le tue zie, zii, nipoti e nipoti, nonni e nipoti condividono tutti il 25% di il tuo DNA.

I genetisti potrebbero utilizzare l'età del sospettato e i vari parenti per distinguere tra queste di-

verse relazioni. Inoltre, alcune di queste relazioni determinano modelli distinti di segmenti di DNA condivisi. Per esempio, sul lato destro della Figura A si osserva che gli alleli che un individuo condivide con i suoi genitori sono presenti solo su uno dei due omologhi dei suoi genitori, ma gli alleli che un individuo condivide con i suoi fratelli possono essere presenti su entrambi gli omologhi.

11.31 a. Due fratelli con madri diverse ma con lo stesso padre condividerebbero, in media, il 25% del loro DNA. Non condividerebbero il DNA delle loro madri, ma in media, metà del DNA del padre comune.

b. È possibile disegnare un numero infinito di diagrammi diversi; uno di questi è il seguente:



11.32 a. L'allele della malattia è dominante: la condizione è rara e si evidenzia un modello verticale di ereditarietà (i bambini affetti hanno un padre affetto). Inoltre, l'allele della malattia è autosomico; non può essere legato all'X perché c'è una trasmissione da padre a figlio. Si noti inoltre che l'allele non può essere legato all'X per un altro motivo: sia i maschi sia le femmine hanno due copie degli SNP (supponendo che questi SNP siano sintenici con il gene della malattia).

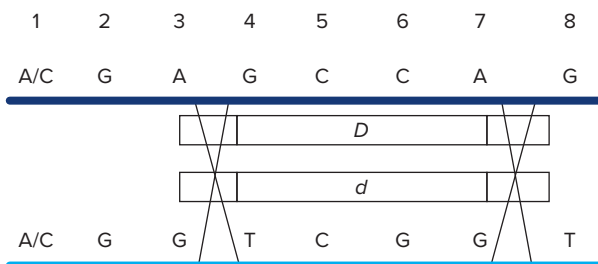
b. Figlio affetto: 1-C 2-G 3-A 4-G 5-C 6-C 7-A 8-T
Figlia affetta: 1-A 2-G 3-A 4-G 5-C 6-C 7-A 8-G
Figlio affetto: 1-C 2-G 3-A 4-T 5-C 6-G 7-G 8-T
Figlia non affetta: 1-LA 2-G 3-G 4-T 5-C 6-G 7-G 8-T

c. SNP2 e SNP5 non forniscono informazioni sul linkage al gene della malattia perché il padre è un omozigote per entrambi gli SNP. (Si ricordi che si può seguire il linkage solo in un doppio eterozigote.)

d. SNP1 non è associato perché metà dei bambini affetti e metà dei bambini non affetti hanno l'allele A di SNP1, mentre l'altra metà dei bambini

affetti e l'altra metà dei bambini non affetti hanno l'allele C. $R = NR$, quindi $RF = 50\%$.

e. ed **f.** Si osservi la figura che segue. Il blu scuro e l'azzurro indicano le due copie del cromosoma 4 nel padre. Il cromosoma blu scuro ha l'allele *D* che causa la malattia dominante. Il cromosoma blu chiaro ha il normale allele recessivo *d*. Un evento di ricombinazione nella linea germinale del padre che ha dato origine al figlio affetto si è verificato tra SNP7 e SNP8. L'evento di ricombinazione nel padre che dà origine al figlio sano deve essere compreso tra SNP3 e SNP4. Il gene della malattia deve trovarsi tra questi due siti di crossing-over. Il gene della malattia, quindi, deve essere compreso tra SNP3 e SNP8; deve essere a sinistra di SNP 4 e a destra di SNP7. Le caselle attraversate dalle X che indicano i crossing-over mostrano le regioni di incertezza. I crossing-over devono avvenire all'interno dei box, ma non sappiamo esattamente dove.



11.33 a. Sì, esistono prove di linkage. Si sono verificati due incroci informativi: II-1 × II-2 e III-7 × III-8. Gli accoppiamenti I-1 × I-2 e III-1 × III-2 non sono informativi, perché in nessuno dei due casi un genitore è eterozigote sia per l'allele della malattia sia per il locus SNP.

Assegniamo *D* come allele della malattia dominante e *d* come allele normale recessivo. L'individuo I-1 ha introdotto l'allele *D* nella famiglia. Poiché I-1 è omozigote per l'allele *C* dello SNP, deve aver trasmesso un cromosoma *D* con *C* a suo figlio II-1. Il cromosoma II-1 ereditato da sua madre (I-2) doveva essere *d* con *T*. Cinque figli di II-1 (III-2, III-3, III-4, III-5 e III-6) hanno ereditato da lui alleli in configurazione parentale (o *D-C* o *d-T*). Solo uno dei suoi figli (III-7) ha ereditato una configurazione ricombinante (*D-T*). Tutti e cinque i bambini dell'altro accoppiamento informativo (III-7 × III-8) hanno ereditato una configurazione parentale degli alleli (*D-T* o *d-C*) dal padre (III-7).

(È fondamentale notare qui che le configurazioni parentali per i cromosomi ereditati da II-1 sono *D-C* e *d-T*, mentre le configurazioni parentali per i cromosomi ereditati da III-7 sono *D-T* e

d-C. La ragione ovviamente è che III-7 ha ereditato un cromosoma paterno ricombinante, ma questo ora diventa una configurazione genitoriale per i cromosomi III-7 trasmessi ai suoi figli.)

I due incroci informativi insieme hanno prodotto 11 individui di progenie, di cui solo 1 (III-7) è ricombinante. La distanza genetica stimata è quindi $1/11 \times 100 = 9,1$ unità della mappa.

b. Il punteggio Lod massimo può essere calcolato utilizzando l'equazione mostrata nel Box Strumenti della Genetica intitolato "La Statistica del Lod Score":

Rapporto di verosimiglianza (assumendo 10 Parentali e 1 Ricombinante e quindi un $FR = 0,091$) = $[(1 - (1/11))^{10} \times (1/11)^1] / (1/2)^{11} = 71,7$

Questo valore significa che è 71,7 volte più probabile che lo SNP e l'allele della malattia siano linked (con un $FR = 0,091$), rispetto al fatto che non lo siano. Il punteggio Lod per un rapporto di verosimiglianza di $71,7 = \log_{10}(71,7) = 1,86$. La convenzione tra la maggior parte dei genetisti è che per considerare i dati sufficienti a dimostrare il linkage, il punteggio Lod deve essere almeno 3,0, il che significa che il rapporto di verosimiglianza deve essere almeno 1000. Pertanto, i dati in questo pedigree non sono sufficienti per dedurre il linkage.

11.34 a. Gli accoppiamenti W, Y e Z non sono informativi perché nessuno dei due genitori è un doppio eterozigote. L'accoppiamento X è informativo perché entrambi i genitori sono doppi eterozigoti e perché tutte le possibili permutazioni dei gameti possono essere distinte come tipi parentali o ricombinanti.

b. È probabile che sia A sia B siano SSR piuttosto che SNP. La maggior parte degli SNP sono biallelici nelle popolazioni umane [vedi Problema 11.6 parte (c)], ma i dati mostrano 4 diversi alleli di A e 4 alleli di B. Un allele A è etichettato A5, il che implica che nelle popolazioni potrebbero esistere 5 o più alleli.

11.35 a. È chiaro che l'allele *C* è comune a quasi tutti gli individui che hanno la malattia di Huntington. Pertanto, è probabile che G8 sia sul cromosoma 4 e sia strettamente associato all'*HD*.

b. Sorprendentemente, se non si fanno supposizioni, non si può dire con certezza che le persone nel pedigree siano state formate da gameti parentali o ricombinanti rispetto al gene *HD* e al marcatore *G8*. Il motivo è che non esistono casi in cui i bambini, i genitori e i nonni siano tutti genotipizzati. Quindi non si conosce la fase a meno che non si assuma un linkage molto stretto.

c. Guardando nel pedigree prima le persone affette genotipizzate per *G8*, si osserva che tutte hanno l'allele C. Il presupposto più semplice è quindi che tutti abbiano ricevuto la configurazione parentale *HD C* che era presente in I-1. Dei discendenti non affetti di I-1 che potrebbero aver ereditato l'allele *HD* dal genitore affetto, nessuno ha l'allele C di *G8* a eccezione di VI-5. È possibile che questa donna possa aver ricevuto C dal padre sano (in altre parole, V-3 è sicuramente AC e anche V-4 potrebbe essere AC). Pertanto, la lettura più semplice di questo pedigree è che nessuno degli individui deve aver ricevuto dal genitore affetto un gamete ricombinante (rispetto al gene *HD* e al marcatore *G8*).

d. Senza ricombinanti, la distanza sarebbe 0 unità mappa. Ma espresso in modo più accurato, in un modo tenga conto della qualità dei dati, si dovrebbe dire che la distanza della mappa è inferiore a $1/47 \times 100 = 2,1$ cM.

e. Il calcolo del Lod score con il metodo delineato nel riquadro Strumenti della Genetica intitolato “La Statistica del Lod Score”, e l'esistenza di 47 individui di progenie di accoppiamenti informativi (tutti parentali):

$$\text{Rapporto di verosimiglianza} = [(1-0)^{47} \times 0^0] / (1/2)^{47} = 1,4 \times 10^{14}$$

$$\text{Lod} = \log_{10}(1,4 \times 10^{14}) = 14,15$$

Cioè, è 14 ordini di grandezza più probabile che *HD* e *G8* siano associati che non associati. Questo calcolo del Lod score è una stima approssimativa perché non stiamo tenendo conto del fatto che alcune delle nostre ipotesi potrebbero essere errate, ovvero alcune delle persone con *HD C* potrebbero essere ricombinanti. Tuttavia, il punteggio di Lod che calcoliamo è così alto che un calcolo corretto (più complesso) che tenesse conto delle fasi sconosciute lo modificherebbe solo di poco. Pertanto, i dati in questo albero genealogico forniscono prove conclusive per il linkage fra *HD* e *G8*.

11.36 a. È possibile che lo SNP sia strettamente legato alla mutazione che causa la malattia, ma non è la mutazione stessa. Semplicemente non si sono esaminati abbastanza bambini per rilevare alcuni rari ricombinanti.

b. Si potrebbe: (1) determinare se il locus SNP è raro verificando se si trova nei database degli SNP noti; (2) determinare se lo SNP è all'interno di un gene e, in tal caso, è probabile che influisca sulla funzione del gene (cioè, lo SNP è non anonimo?) Probabilmente la sostituzione della coppia di ba-

si si traduce in una mutazione di senso che colpisce un amminoacido conservato o in una mutazione non senso. Forse la sostituzione interrompe una giunzione di splicing; (3) controllare altre famiglie non imparentate, alcune con la malattia e altre senza. Forse la stessa base G è sempre correlata alla malattia. Forse si scoprirebbe che in tutte le famiglie affette, il gene in cui si verifica questa sostituzione ha costantemente un SNP non anonimo, anche se il polimorfismo nella famiglia non è il cambiamento T/G notato nel problema. Lo stesso gene potrebbe ospitare mutazioni diverse in famiglie affette diverse, in una situazione di eterogeneità allelica.

[Il Capitolo 20 illustrerà un altro metodo che potrebbe aiutare a determinare se un particolare gene o un cambiamento della coppia di basi causa una condizione genetica. L'idea è di manipolare il genoma di un animale modello (il topo o la *Drosophila melanogaster*) per eliminare la funzione del gene o per introdurre lo stesso cambiamento di base nel gene. Se gli animali sviluppano un fenotipo aberrante simile alla malattia umana, un tale risultato costituirebbe una forte evidenza di causa-effetto. Una nuova tecnica chiamata “CRISPR/Cas9” discussa nel Capitolo 20 è altamente efficiente nel creare tali “mutazioni mirate” nel genoma del topo.]

11.37 a. *D* è l'allele dominante che causa la malattia; *d* è l'allele normale; *m1*, *m2* e *m3* sono gli alleli SSR. (Stiamo usando *D* e *d* invece dei normali simboli genetici umani *D*⁺ e *D* per sottolineare che l'allele della malattia è dominante.) In ogni caso, è necessario prima determinare la fase degli alleli del gene della malattia e degli alleli SSR; in altre parole, è necessario determinare quali alleli del gene della malattia (*D* o *d*) e quali alleli SSR (*m1* o *m2*) siano sullo stesso cromosoma nel genitore maschio. In secondo luogo, data la FR del 10% tra i due loci e la conoscenza dell'allele SSR che il bambino ha ereditato dal genitore maschio (la condizione), è necessario utilizzare la probabilità condizionata per determinare la possibilità che il bambino abbia anche ereditato *D*.

Nel pedigree a sinistra, i cromosomi del genitore maschio sono *D m2/d m1*. Si producono i seguenti gameti alle frequenze indicate tra parentesi: *D m2* (0,45); *dm1* (0,45); *D m1* (0,05); *d m2* (0,05). La figlia A ha ereditato *m1* da suo padre; ciò significa che ha ereditato *d m1* o *D m1*. Il rapporto 45:5 (o rapporto 9:1) dei gameti *d m1*: *D m1* significa esiste la probabilità del 10% che il bambino A abbia ereditato l'allele *D* (malato). Il bam-

bino B ha ereditato $m2$ da suo padre, il che significa che ha ereditato $D m2$ o $d m2$. Il rapporto 9:1 dei gameti $D m2$: $d m2$ significa che c'è una probabilità del 90% che il bambino B sia malato (allele D ereditato).

Nel pedigree centrale, la figlia C ha ereditato l'allele $m2$ da suo padre. La probabilità che la bambina C sia affetta non è influenzata dal fatto che il primo figlio sia affetto, quindi il calcolo è lo stesso del bambino B; cioè, anche la probabilità che la bambina C abbia la malattia è del 90%.

Nel pedigree a destra, la fase degli alleli nel genitore maschio non è nota. Tuttavia, possiamo utilizzare l'informazione che il fratello della bambina D è affetto per determinare la probabilità di una fase o dell'altra. Il fratello della bambina D ha ereditato $D m1$ da suo padre. Ciò significa che la probabilità è del 90% che il genotipo del padre sia $D m1/d m2$ e la probabilità del 10% che il suo genotipo sia $d m1/D m2$. La bambina D ha ereditato $m2$ da suo padre. La probabilità che la bambina D abbia ereditato $m2 D$ da suo padre è la somma ponderata delle probabilità di ereditare questo cromosoma a partire da ciascuno dei possibili genotipi del padre di cui sopra. Se il genotipo del padre è $D m1/d m2$, usando la stessa logica che abbiamo usato sopra per i bambini A-C, ci sono 9 gameti $d m2$ prodotti per ogni gamete $D m2$, quindi la probabilità relativa di $D m2$ è 10%. Se il genotipo del padre è $d m1/D m2$, la probabilità relativa di $D m2$ è del 90%. Pertanto, la probabilità che la bambina D abbia la malattia (cioè che abbia ereditato $D m2$) è: $(0,9)(0,1) + (0,1)(0,9) = 0,09 + 0,09 = 0,18$, o 18%.

b. Un genetista umano non utilizza volentieri un marker SSR a 10 unità mappa (cM) di distanza per diagnosticare un gene, perché il 10% delle volte la presenza in un feto dell'allele SSR legato all'allele della malattia porterebbe a una diagnosi errata. Il 10% delle volte in cui è presente nel feto l'altro allele SSR (di solito sullo stesso cromosoma dell'allele recessivo, non patogeno), il feto risulterebbe invece affetto. Pertanto, il 10% dei negativi sarebbero falsi negativi e il 10% dei positivi sarebbero falsi positivi.

11.38 a. Il primo figlio affetto mostra quale allele marcatore è più probabilmente legato all'allele della malattia in ciascun genitore. Diciamo "molto probabile" perché esiste una piccola possibilità che uno degli alleli $CFTR$ del primo figlio sia il prodotto di un crossing-over, tra il marker SSR e la mutazione che causa la malattia nel gene $CFTR$, verificatasi nelle linee germinali di uno

dei genitori. Il bambino affetto ha ricevuto l'allele SSR grande dal padre e l'allele SSR piccolo dalla madre. Ciò significa che l'allele SSR grande del padre e quello piccolo della madre quasi certamente contrassegnano, in ciascun genitore, le mutazioni responsabili della malattia.

Il feto ha ereditato l'allele piccolo del padre e quello piccolo della madre. Pertanto, il feto è quasi certamente un portatore e la possibilità che il feto sia affetto (od omozigote per un allele $CFTR$ normale) è molto vicina allo 0%. Anche supponendo che i loci SSR siano allineati con le mutazioni che causano la malattia allo stesso modo nei genitori e nel primo figlio, esiste una piccola possibilità che un crossing-over abbia separato il marker SSR e la mutazione causale in uno dei geni $CFTR$ del feto. Tuttavia, molto probabilmente gli alleli SSR sia nel primo figlio, sia nel feto sono associati agli alleli della malattia allo stesso modo in cui lo sono in entrambi i genitori.

b. Il feto diventerà quasi sicuramente una donna portatrice. Poiché questo è un carattere recessivo raro, le probabilità che qualcuno dei suoi figli abbia la fibrosi cistica sono basse. Non solo dovrebbero ricevere l'allele mutante $CFTR$ da questa donna (probabilità = 0,5), ma dovrebbero anche ricevere un allele mutante $CFTR$ dal padre. Il problema afferma che il 3% della popolazione è portatore di un gene $CFTR$ mutante e supponiamo inoltre che quest'uomo non sia affetto da fibrosi cistica. Pertanto, l'incrocio che potrebbe produrre un bambino con fibrosi cistica è $CF CF^+ \times CF CF^+$. Sappiamo già che la donna è portatrice (probabilità = 1), e la probabilità che un dato uomo sia $CF CF^+$ è 0,03. La probabilità che ogni bambino prodotto da questo accoppiamento sia $CF CF$ e abbia la malattia è 1/4. Pertanto, la probabilità che un figlio di questa donna (e un uomo scelto a caso dalla popolazione) abbia la fibrosi cistica è $1 \times 0,25 \times 0,03 = 0,0075 = 0,75\%$.

c. In contrasto con la situazione descritta nel Problema 11.37, la possibilità di falsi positivi o falsi negativi nello scenario qui descritto è molto bassa, quindi un tale test sarebbe abbastanza accurato (sebbene non accurato al 100%). Ancora più importante, questo metodo indiretto aggira il problema che esistono più di 1500 mutazioni diverse del gene $CFTR$ che potrebbero causare la fibrosi cistica; i medici quindi non saprebbero mai quali mutazioni testare direttamente. Ma seguendo il marcatore SSR, i medici sarebbero in grado di fare previsioni accurate sulla situazione della malattia del feto con un'elevata (ma non to-

tale) sicurezza, anche se non sapessero quale specifico allele mutante di CFTR fosse coinvolto. Si noti che in futuro, quando il sequenziamento dell'intero genoma diventerà più di routine, si potranno testare tutte le mutazioni che causano malattie nel gene CFTR del feto, quindi questo metodo indiretto non sarebbe necessario.

11.39 a. Il farmaco dovrebbe essere efficace nell'eterozigote composto G551D, anche se probabilmente meno che in un omozigote, perché nell'eterozigote ci sarà la metà delle proteine che possono essere rese funzionali dal farmaco.

b. I pazienti giovani non hanno accumulato tanto muco nei polmoni quanto quelli più anziani e potrebbe essere più facile per il farmaco raggiungere il CFTR nella membrana. Inoltre, i pazienti più anziani probabilmente hanno già qualche danno polmonare irreversibile. Il farmaco non è realmente progettato per riparare i tessuti danneggiati, ma per prevenire che si verifichino danni.

c. I ricercatori hanno inserito i composti nel mezzo di crescita delle cellule omozigoti per il mutante G551D e quindi hanno esaminato al microscopio le cellule che hanno le ciglia oscillanti. Questo è stato infatti l'approccio che ha portato allo sviluppo dell'ivacaftor. Ciò non può funzionare nel caso di pazienti che non hanno la funzione della proteina CFTR (perché sono omozigoti o transeterozigoti per alleli nulli). Il motivo è che questa tecnica dipende dalla presenza nella membrana di alcune proteine mutanti che potrebbero essere rese funzionali dopo il legame del farmaco. Se non è presente tale proteina, questo approccio non consente la formazione di canali ionici del cloro funzionali.

d. Ivacaftor da solo è inefficace su $\Delta F508$ perché anche con il farmaco, la proteina mutante non viene inserita nella membrana cellulare. Lumacaftor è necessario affinché alcune delle proteine mutanti $\Delta F508$ possano piegarsi abbastanza bene da essere inserite nella membrana cellulare. Apparentemente, quelle molecole $\Delta F508$ che vengono inserite nella membrana hanno una funzione di canale inefficiente. I due farmaci in combinazione sono efficaci perché lumacaftor consente alle molecole $\Delta F508$ di inserirsi nella membrana; una volta posizionate, ivacaftor consente loro di funzionare in modo più efficiente come canali ionici del cloruro.

11.40 a. Gli adattatori vengono aggiunti alle estremità dei frammenti di DNA genomico per due motivi: (I) i frammenti hanno tutti le stesse estremità 5' in modo che possano attaccarsi tutti alla superfi-

cie della cella a flusso mediante appaiamento di basi con oligonucleotidi di sequenza complementare, che sono pre-ancorati alla cella di flusso; (II) I frammenti hanno tutti le stesse estremità 3', in modo che la sintesi dei filamenti complementari di tutti i frammenti possa essere innescata da oligonucleotidi con la stessa sequenza di basi.

b. Cluster di frammenti di DNA identici vengono generati in modo che la fluorescenza emessa dai deossiribonucleotidi marcati aggiunti a ogni ciclo di sintesi sia rilevabile. A ogni cluster, vengono aggiunti circa 1000 deossiribonucleotidi con lo stesso marcatore fluorescente (dello stesso colore) a circa 1000 frammenti identici a singolo filamento. Se invece dei cluster fossero disponibili solo singoli frammenti di DNA genomico come stampo, i segnali generati a ciascun ciclo non sarebbero rilevabili.

c. È necessario rimuovere il tag fluorescente alla fine di ogni ciclo in modo da poter rilevare il colore del tag fluorescente del nucleotide incorporato al ciclo successivo. Si ricordi che questo metodo alla fine produrrà un filmato per ogni frammento dello stampo che mostra emissioni di colore successivi che indicano la sequenza del filamento appena sintetizzato. Non sarebbe possibile ottenere un film del genere se il tag fluorescente aggiunto a ogni round non fosse rimosso.

d. Il gruppo bloccante assicura che venga aggiunto un solo nucleotide a ogni ciclo in modo che il suo colore possa essere registrato prima che vengano aggiunti ulteriori nucleotidi. Se il gruppo bloccante non viene rimosso prima del ciclo successivo, non è possibile aggiungere il nucleotide successivo alla catena in crescita al ciclo successivo e quindi non è possibile sequenziare il frammento genomico oltre la prima base.

e. Per sequenziare il secondo filamento (inverso): (I) denaturare i DNA a doppio filamento formati sequenziando il primo filamento (in avanti) e lavare via il filamento appena sintetizzato: il filamento stampo originale rimane; (II) tagliare i filamenti opposti (quelli collegati agli oligo viola nella Figura 11.22) dagli oligo e lavarli via; (III) sequenziare come prima.

11.41 a. Il ricercatore dovrebbe cercare possibili casi di transeterozigosi; cioè, analizzare i geni per scoprire se entrambi gli alleli sono mutanti ma in posizioni diverse. In teoria, il ricercatore esaminerebbe uno di questi geni identificati per verificare se è probabile che le due mutazioni si traducano in una perdita di funzione, come le mutazioni non senso o frameshift. Se una o entram-

be le mutazioni sono di senso, sarebbe interessante determinare se alterano amminoacidi altamente conservati.

b. Sfortunatamente, è difficile sapere cosa cercare nell'intero genoma al di fuori delle regioni codificanti. Forse il paziente ha una mutazione in una regione intergenica o in un introne, che colpisce l'espressione di un gene bersaglio, ma come potrebbe lo scienziato riconoscerla tra i tanti polimorfismi che esistono nell'intero genoma del paziente? Un possibile approccio è focalizzare la ricerca su nucleotidi particolarmente ben conservati, o quelli in noti motivi di legame per determinati fattori di trascrizione. Il Capitolo 18 descrive un grande progetto chiamato ENCODE (l'Enciclopedia degli elementi del DNA) intrapreso da un consorzio di molti laboratori che sta tentando di identificare tutti gli elementi al di fuori delle regioni codificanti che potrebbero influenzare l'espressione dei geni. Se si cercano mutazioni che potrebbero produrre fenotipi mutanti, gli elementi identificati dal progetto ENCODE sarebbero i primi candidati.

Il Problema 11.42 discute un approccio alternativo in cui gli scienziati potrebbero provare a determinare se una condizione genetica determina cambiamenti quantitativi nell'espressione di un particolare gene nei tessuti di un paziente.

11.42 a. Utilizzando questo tipo di sequenziamento degli mRNA (il metodo è anche chiamato RNA-Seq), i ricercatori potrebbero tentare di trovare mRNA i cui livelli cambiano drasticamente nel paziente rispetto ai controlli non affetti. Questo approccio potrebbe essere molto utile se non si trovano mutazioni nelle regioni codificanti dei geni, perché le mutazioni potrebbero influenzare la quantità di espressione di un gene. Gli scienziati potrebbero anche essere in grado di rilevare i cambiamenti nei pattern di splicing del trascritto di un gene. Ma anche questa tecnica presenta problemi. Come scegliere quale tessuto esaminare? Se non si sceglie quello giusto, potrebbero non trovarsi cambiamenti nei livelli di espressione genica. (A volte il fenotipo della malattia suggerisce in quale tessuto guardare.) Inoltre, le informazioni ottenute con questo approccio non sarebbero informative a meno che si osservassero notevoli cambiamenti nelle quantità di uno specifico mRNA.

b. Se sono stati osservati cambiamenti rilevanti nella dimensione o nella quantità di una proteina in un paziente, allora il gene che codifica per quella proteina potrebbe essere responsabile della malattia. Successivamente si dovrebbe essere

in grado di rilevare cambiamenti nelle regioni regolative che causano un'espressione aberrante del gene o mutazioni non senso o frameshift nella regione codificante. Probabilmente non si otterrebbe alcuna informazione da questo metodo se la malattia fosse causata da una mutazione di senso nella regione codificante perché la maggior parte delle volte né la quantità né la dimensione della proteina sarebbero cambiate. Tuttavia, esistono casi eccezionali in cui una mutazione di senso potrebbe alterare la configurazione tridimensionale della proteina rendendola instabile, per cui la quantità della proteina diminuirebbe.

Si noti che un problema con la tecnica di Western blotting è che si può analizzare solo una proteina alla volta ed è necessario un anticorpo specifico per esaminare la proteina. Quindi, questa tecnica si usa solo se un particolare gene è un probabile candidato come gene della malattia.

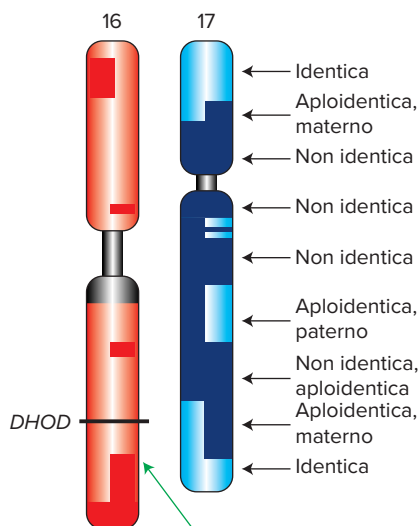
Un approccio alternativo chiamato spettrometria di massa consentirebbe di esaminare i livelli di tutte le proteine che costituiscono il proteoma di un determinato tipo di cellula. In questo caso, un estratto proteico viene digerito con una proteasi e le masse dei frammenti risultanti vengono quindi misurate mediante spettrometria di massa. Queste dimensioni vengono confrontate con un catalogo delle dimensioni dei frammenti proteolitici di tutti i geni previsti dal genoma annotato. Questa tecnica può essere potente, ma è difficile ottenere dati quantitativi accurati che riflettano i livelli di proteine.

11.43 a. La sindrome di Miller potrebbe essere dovuta a una mutazione dominante se: (I) il carattere non è completamente penetrante, per cui entrambi i genitori avevano l'allele mutante ma non esprimevano il carattere; (II) si è verificata una nuova mutazione nella linea germinale di uno dei genitori, così che molti oociti (e poi uova) o spermatozoi (e poi spermatozoi) potrebbero avere la stessa mutazione. La mutazione non può essersi verificata in uno stadio tardivo della linea germinale (cioè durante la meiosi negli oociti o negli spermatozoi primari) perché più di un bambino è affetto. La parte (b) esplora questo nuovo scenario di mutazione in modo più dettagliato.

b. Nella linea germinale si verificano in effetti nuove mutazioni. Affinché un'ampia parte dei gameti prodotti da un genitore portasse la nuova mutazione, la mutazione avrebbe dovuto verificarsi relativamente presto nella proliferazione mitotica di oogoni o spermatogoni. Più è precoce lo stadio di proliferazione della linea germinale,

meno cellule sono presenti e quindi minori sono le probabilità che si verifichi la mutazione. Questo fatto spiega perché le nuove mutazioni che si verificano nella linea germinale e che causano la mutazione di una elevata percentuale di gameti sono estremamente rare. I genetisti umani devono comunque prendere in considerazione tali possibilità, anche se non sono molto probabili.

c. Nella figura che segue, in qualsiasi punto si osservi un cambiamento del colore da una tonalità scura a una chiara, o viceversa, deve essersi verificato un evento di ricombinazione durante la meiosi in uno dei genitori che ha prodotto il gamete che ha dato vita a uno dei figli. La freccia verde nella figura mostra l'evento di ricombinazione che si è verificato più vicino al gene *DHOD*. Questo evento di ricombinazione è visto perché *DHOD* si trova in una regione identica del braccio lungo del cromosoma 16 (dove il figlio affetto e la figlia affetta condividono gli stessi alleli della madre e gli stessi alleli del padre), ma più vicino al telomero si nota una regione aploidica materna in cui i due pazienti condividono gli stessi alleli della madre ma hanno alleli diversi dal padre. Pertanto, nel sito della freccia verde deve essersi verificato un evento di ricombinazione durante la meiosi I in uno spermatozoo primario del padre. Dal modo in cui i dati vengono visualizzati non sappiamo se questo evento di ricombinazione sia avvenuto durante la meiosi che ha portato allo spermatozoo che ha prodotto il figlio o la figlia.



d. I dati rivelano circa 2 volte più crossing-over del previsto, ma i numeri effettivi sono ancora sorprendentemente vicini alle aspettative. La figura mostra 8 crossing-over sul cromosoma 16 e

10 sul cromosoma 17 nei 4 gameti che sono stati usati per creare i due bambini affetti. I due cromosomi assommano insieme a circa 170 Mb e quindi sono lunghi 170 cM. 1 cM equivale a un crossing-over su 100 gameti, quindi ci si aspettano circa 170 crossing-over su 100 gameti per i cromosomi 16 e 17. In 4 gameti, sarebbero previsti circa 7 crossover, ma i dati mostrano poco più del doppio di questo numero (18).

e. Si dovrebbero semplicemente confrontare i nucleotidi dell'intero genoma dei genitori con quelli dei bambini e cercare nuove alterazioni. Quindi si divide il numero di alterazioni nei due bambini per 4 volte la dimensione del genoma aploide (perché ci sono 4 genomi aploidi nei due bambini).

È interessante notare che in molti casi (a seconda della disposizione effettiva degli alleli), si può effettivamente determinare in quale genitore si è verificata una particolare mutazione. I dati di questo studio hanno rivelato che i due bambini di questa famiglia hanno ricevuto ciascuno in media 30 nuove mutazioni che si sono verificate nella linea germinale della madre e circa 50 che si sono verificate nella linea germinale del padre. Ci si aspetterebbe che il numero di nuove mutazioni paterne aumenti nella progenie dei padri più anziani ma non delle madri, perché la linea germinale maschile viene continuamente propagata attraverso le divisioni mitotiche e più divisioni avvengono, maggiore è la probabilità che si verifichino mutazioni (vedi Problema 11.4).

11.44 a. e b. Le conclusioni sono sintetizzate nella tabella che segue. Tutti i loci autosomici dovrebbero avere lo stesso numero di read, il 90% delle quali dovrebbe provenire dal genoma della madre e il 10% dal genoma del feto. Se la madre e il feto fossero omozigoti per lo stesso allele, dovrebbe esserci un solo tipo di read che rappresenta quell'allele. Se esistono due o tre tipi di read (due o tre alleli), le proporzioni relative dipendono dal fatto che gli alleli siano o meno condivisi dalla madre e dal feto. Per un locus sul cromosoma X, ci si aspetterebbe o la stessa situazione dei loci autosomici (lo stesso numero di read; 90% materne, 10% fetali) se il feto fosse una femmina XX, oppure il 5% di letture in meno (di cui il 95% sarebbe materno e il 5% fetale) se il feto fosse un maschio XY con solo 1 cromosoma X. Per un locus sul cromosoma Y, il numero di letture sarebbe solo il 5% del numero di letture per un locus autosomico. (Ovviamente, questi dati sono stati molto semplificati rispetto ai risultati effettivi che i ricercatori otterrebbero da un tale test.)

Locus	Sequenze	Numero di read	Genoma materno	Genoma del feto	Autosomico X-linked Y-linked
1	α : TCTTTGGTAAACGCAAG	1000	$\alpha \alpha$	$\alpha \alpha$	Autosomico
2	α : GTACCGGAGGCAGCCTC β : GTACCGCGGCAGCCTC	500 500	$\alpha \beta$	$\alpha \beta$	Autosomico
3	α : AGCCATTGCGGATCCGA β : AGCTATTGCGGATCCGA	950 50	$\alpha \alpha$	$\alpha \beta$	Autosomico
4	α : GGGGCCTTATGATAAGG	50	--	$\alpha -$	Y-linked
5	α : CAGTTCCTGGAGTTGTA β : CAGTTCATGGAGTTGTA	550 450	$\alpha \beta$	$\alpha \alpha$	Autosomico
6	α : GCAGCCCGTGTCTTAA β : GCAGCCCGTGTCTCAA	500 450	$\alpha \beta$	$\alpha -$	X-linked
7	α : CACTCAGTCTACGGAC β : CACTCGGTCTACGGAC γ : CACTCAGTCTAAGGAC	500 450 50	$\alpha \beta$	$\alpha \gamma$	Autosomico

c. Il feto è chiaramente maschio perché uno dei loci (4) è legato al cromosoma Y.

d. L'ottavo locus con 1500 read suggerisce che esistono copie aggiuntive di questa regione nel DNA. Molto probabilmente, anche l'ottavo locus è autosomico e sia la madre sia il feto hanno tre copie del locus invece delle due normali. Una possibilità è che il locus si trovi sul cromosoma 21 e sia la madre sia il feto siano trisomici per questo cromosoma (cioè entrambi affetti dalla sindrome di Down). In alternativa, il locus potrebbe trovarsi su qualsiasi autosoma e sia la madre sia il feto potrebbero essere eterozigoti per la duplicazione.

11.45 a. I ricercatori non avrebbero cercato solo varianti omozigoti. Per quanto ne sappiamo, i genitori di Nic non sono parenti stretti, quindi è del tutto possibile che Nic sia un eterozigote composto (cioè, che Nic abbia ereditato due diverse mutazioni nello stesso gene, una da ciascun genitore non affetto).

b. (I) La mutazione che causa la malattia di Nic potrebbe essere stata una nuova mutazione dominante che si è verificata nella linea germinale di uno dei suoi genitori (quindi non affetti). (II) È anche possibile che la condizione sia dominante ma la penetranza sia bassa, quindi uno dei genitori di Nic potrebbe aver portato la mutazione ma non esserne stato colpito. Fortunatamente per Nic, le ipotesi originali del medico erano corrette. Sarebbe stato molto più difficile trovare la mutazione della malattia se si fossero dovute considerare queste altre ipotesi.

c. Si potrebbe raggiungere questo obiettivo molto semplicemente usando la PCR. Si progettano primer che amplificano la regione del genoma contenente la mutazione causale. Quindi si amplifica per PCR questa regione dai genomi dei due genitori di Nic (per esempio, da un tampone della loro

mucosa orale). Se questa mutazione non è stata trovata in nessuno dei due genomi, deve trattarsi di una mutazione nuova. Se si trovasse la mutazione in eterozigosi nel genoma della madre, ma non in quello del padre, questo approccio PCR fornirebbe invece la prova che la mutazione causale è legata all'X e che il carattere è recessivo.

11.46 a. Il problema afferma che il primo esone di questo gene contiene il codone ATG iniziale. Pertanto, la sequenza del prodotto proteico che può essere dedotta dalla sequenza nucleotidica del primo esone è: N Met-Cys-Gly-Ala.... C.

b. Poiché è stato trovato un solo tipo di sequenza, l'individuo 2 deve essere un omozigote per l'allele 5' CAACGCTTAGGATGTGGAGCCT 3'. Questo risultato è sorprendente se questo allele è raro nella popolazione e suggerirebbe che i genitori di questa persona possano essere stati strettamente imparentati. Il risultato è meno sorprendente se questo allele è più diffuso nella popolazione. (Si vedrà nel seguito che questo allele è un allele di malattia e quindi dovrebbe essere raro nella maggior parte delle popolazioni.)

c. La malattia è causata da un allele dominante, perché gli eterozigoti come l'individuo 3, con un allele normale (trovato nella sequenza di riferimento) e un allele mutato, mostrano il fenotipo mutato.

d. Sì, i dati mostrano l'esistenza di eterogeneità allelica. Tra questi quattro pazienti, diversi alleli mutanti possono causare la malattia.

e. Nessuno di questi individui è un eterozigote composto (chiamato anche transeterozigote). (Un eterozigote composto presenta due alleli mutanti diversi, ma nessuno di questi individui lo fa.)

f. I dati non forniscono prove di eterogeneità del locus nella sindrome di Brugada. Si può spiegare la malattia in tutti gli individui colpiti con l'esistenza di mutazioni in un singolo gene. Eterogeneità del locus vuol dire che mutazioni in geni diversi potrebbero causare la sindrome. I dati non escludono questa possibilità (perché potrebbero esistere altri pazienti che hanno la malattia per mutazioni in un altro gene), ma l'ipotesi di eterogeneità del locus non è necessaria per spiegare i dati forniti.

g. L'individuo 4 dovrebbe essere normale (cioè non affetto dalla sindrome di Brugada). Un allele è l'allele della sequenza di riferimento. L'altro allele presenta una mutazione silente (che cambia un codone TGC in un codone TGT, entrambi per l'amminoacido Cys).

h. e i. La mutazione di senso dell'individuo 1 nell'allele inferiore cambia GGC (Ala) in GAC (Asp).

Questo allele sarebbe classificato come SNP perché viene modificata solo una coppia di nucleotidi. La mutazione provoca una perdita di funzione, per quanto non sia possibile affermarlo solo in base a questa mutazione (ovvero, è necessario guardare i risultati per altri pazienti; vedi oltre). Se si tratta di una mutazione con perdita di funzione con effetti dominanti, il motivo deve essere l'aploinsufficienza. Questo allele è probabilmente ipomorfo con una minore attività del prodotto proteico perché è una perdita di funzione, ma è probabile che venga prodotta qualche proteina (poiché tutti gli amminoacidi sono presenti tranne uno che è cambiato). Si potrebbe forse classificarlo come una mutazione nulla anziché ipomorfa se la mutazione di senso risultasse in una proteina senza funzione.

L'individuo 2 è omozigote per una mutazione non senso nel secondo codone. Ancora una volta, questo è un SNP perché è interessata solo una coppia di basi. Questo allele è nullo, con perdita di funzione, perché non viene prodotta alcuna proteina. Il fatto che l'individuo 2 sopravviva significa che il gene non può essere essenziale. Sarebbe interessante per uno scienziato esaminare attentamente l'elettrocardiogramma di questa persona. Potrebbe essere possibile che la funzione cardiaca in questo individuo sia colpita più che in altri pazienti con sindrome di Brugada perché lui/lei è omozigote senza alcuna funzione genica.

L'individuo 3 è un eterozigote per una mutazione frameshift in cui un nucleotide viene perso dal frame di lettura. Questa mutazione, classificata come DIP, non come SNP, è una mutazione nulla perché l'allele non può produrre alcuna proteina normale. Il gene è aploinsufficiente, e questo spiega perché l'eterozigote mostra il fenotipo mutato. (Questo fatto spiega anche perché è probabile che la mutazione nell'individuo 1 sia a perdita di funzione; cioè, l'eterozigosi per almeno una mutazione a perdita di funzione produce la malattia.)

Nell'individuo 4, l'allele che non corrisponde alla sequenza di riferimento ha una mutazione silente (uno SNP anonimo) come discusso nella parte (g).

j. Sì, la funzione del gene è aploinsufficiente, perché gli individui 2 e 3, che sono eterozigoti per gli alleli nulli del gene (hanno un allele di tipo selvatico e un allele nullo), mostrano i sintomi patologici della sindrome di Brugada.

11.47 a. HPRT1 è l'ipoxantina guanina fosforibosiltransferasi 1. (Questo è il nome dell'enzima codificato dal gene HPRT1.)

b. Il gene *HPRT1* è sul cromosoma X.

c. Una malattia è la sindrome di Lesch-Nyhan: (I) i sintomi principali sono ritardo mentale, paralisi cerebrale, movimenti spastici, calcoli urinari e comportamenti autodistruttivi; (II) il carattere è recessivo; (III) la sindrome è causata da alleli a perdita di funzione; (IV) la sindrome mostra eterogeneità allelica; (V) la sindrome non mostra eterogeneità di locus: l'unica causa è la mutazione di *HPRT1*.

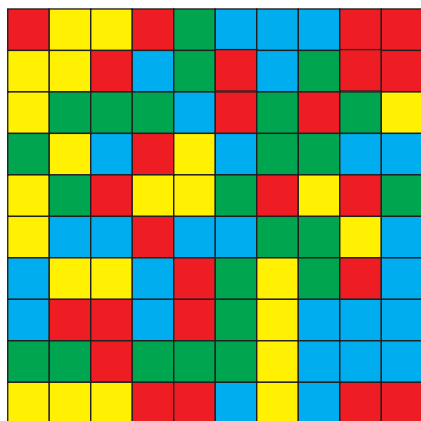
L'altra malattia è la sindrome di Kelley-Seegmiller: (I) i sintomi principali sono calcoli renali e gotta; (II) il carattere è recessivo; (III) la sindrome è causata da alleli a perdita di funzione; (IV) la sindrome mostra eterogeneità allelica; (V) la sindrome non mostra eterogeneità di locus: l'unica causa è la mutazione di *HPRT1*.

Sorge spontanea la domanda perché gli alleli a perdita di funzione nello stesso gene causano due sindromi diverse. La risposta è che la sindrome di Lesch-Nyhan coinvolge alleli di *HPRT1* nulli o fortemente ipomorfi, mentre la sindrome di Kelley-Seegmiller è causata da alleli di *HPRT1* che influenzano la produzione o la funzione dell'enzima HPRT1 in modo meno drastico. Entrambe le condizioni provocano la sovrapproduzione di acido urico tossico (a livelli più elevati nei pazienti con Lesch-Nyhan rispetto a quelli con Kelley-Seegmiller). Almeno alcuni sintomi associati a queste sindromi possono essere alleviati in una certa misura da farmaci come l'allopurinolo, che riducono i livelli di acido urico nel sangue, ma non incidono sulla causa delle malattie.

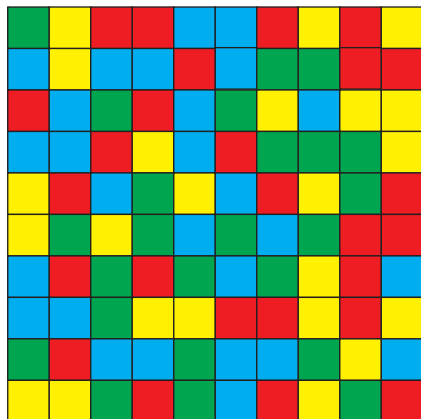
11.48 Una persona potrebbe avere sintomi di una malattia a causa di una mutazione somatica de novo, che non ha ereditato da nessuno dei genitori. La mutazione somatica potrebbe generare un allele dominante per una malattia, oppure una persona che eredita un allele recessivo di una malattia da un genitore potrebbe poi contrarre la malattia a causa di una nuova mutazione somatica nella copia normale del gene che ha ereditato dall'altro genitore. Indipendentemente dal fatto che la mutazione somatica sia dominante o recessiva, il clonaggio posizionale sarebbe impossibile perché nessun altro individuo nel pedigree avrebbe la malattia né l'allele mutante. La mutazione non verrebbe rilevata dal sequenziamento del genoma se le cellule utilizzate come fonte del DNA per l'analisi non la contenessero.

11.49 La macchina di sequenziamento “vede” i colori di ciascun cluster a ogni ciclo. Il colore di un cluster è determinato da quale dei quattro nucleotidi contrassegnati dalla marcatura fluorescente (un fluorocromo diverso per ogni nucleotide) è stato incorporato nel filamento di DNA sintetizzato. La sequenza di colori che appare in ogni cluster indica la sequenza del filamento di DNA che viene prodotto. I dati per 3 cicli di sintesi a 100 cluster assomiglierebbero a questi:

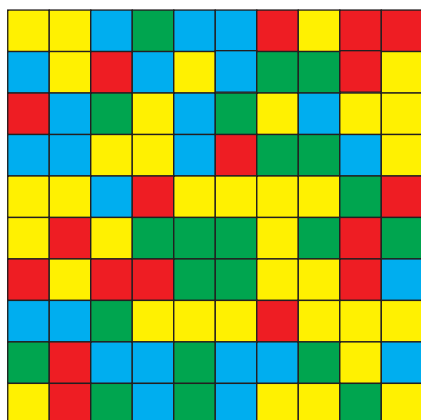
Ciclo 1



Ciclo 2



Ciclo 3



Ogni quadrato della griglia rappresenta il segnale di un singolo cluster. La sequenza di lettura del primo quadretto in alto a sinistra, per esempio, è rosso-verde-giallo (o TAG utilizzando i colori della Figura 11.22), la sequenza di lettura del quadretto in alto a destra è rosso-giallo-rosso (o TGT) e così via.

11.50 a. Le mutazioni puntiformi che generano codoni di stop (mutazioni non senso) producono proteine tronche.

b. Le mutazioni di stop di solito generano alleli amorfi. Se non si sono trovati mai alleli amorfi di un gene in nessun individuo, molto probabilmente significa che un tale allele è raro (presente in meno di 1/60 706 persone). La rarità dell'allele potrebbe indicare che l'omozigosi per gli alleli amorfi del gene provoca una malattia grave, o forse anche che l'eterozigosi per un allele amorfo e un allele normale causa la malattia (il gene è aploinsufficiente).

c. I dati di sequenziamento ExAC forniscono un elenco di geni umani che sono probabilmente essenziali, geni per i quali è probabile che la perdita di funzione causi malattie. Quando si sequenzia il genoma o l'esoma di un individuo per trovare SNP rari non anonimi, gli scienziati possono utilizzare questo elenco per decidere quali alleli SNP rari nel genoma di un individuo potrebbero essere alleli della malattia.