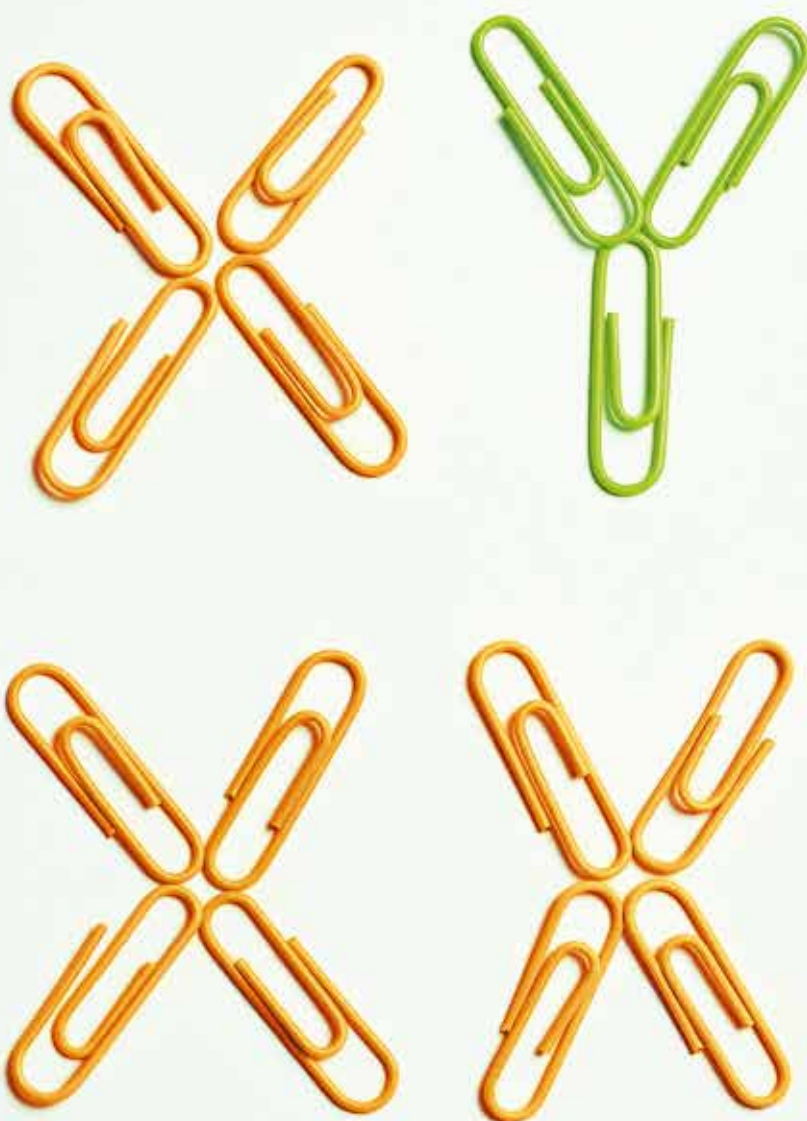


Tom Strachan Andrew Read

Genetica molecolare umana

Seconda edizione Zanichelli condotta sulla quinta edizione inglese

A cura di Rossella Tupler



Tom Strachan Andrew Read

Genetica molecolare umana

Seconda edizione Zanichelli condotta sulla quinta edizione inglese

A cura di Rossella Tupler

Se vuoi accedere alle risorse online riservate

1. Vai su **my.zanichelli.it**
2. Clicca su *Registrati*.
3. Scegli *Studente*.
4. Segui i passaggi richiesti per la registrazione.
5. Riceverai un'email: clicca sul link per completare la registrazione.
6. Cerca il tuo codice di attivazione stampato in verticale sul bollino argentato in questa pagina.
7. Inseriscilo nella tua area personale su **my.zanichelli.it**

Se sei già registrato, per accedere ai contenuti riservati di altri volumi ti serve solo il relativo codice di attivazione.

Titolo originale: *Human Molecular Genetics*, 5th Edition
Copyright © 2019 by Taylor and Francis Group, LLC
CRC Press is an imprint of Taylor&Francis Group, an Informa business
Authorised translation from the English language edition published by CRC Press,
a member of the Taylor & Francis Group LLC.
All rights reserved.

Traduzione: Valentina Salsi (2, 3, 4, 5, 8, 10, 14, 19), Rossella Tupler (4, 5, 11, 12, 15-18),
Antonio Vallarola (1, 6, 7, 9, 13, 20-22).
Revisione: Rossella Tupler

© 2021 Zanichelli editore S.p.A., via Imerio 34, 40126 Bologna [52030]
www.zanichelli.it

I diritti di elaborazione in qualsiasi forma o opera, di memorizzazione anche digitale su supporti di qualsiasi tipo (inclusi magnetici e ottici), di riproduzione e di adattamento totale o parziale con qualsiasi mezzo (compresi i microfilm e le copie fotostatiche), i diritti di noleggio, di prestito e di traduzione sono riservati per tutti i paesi. L'acquisto della presente copia dell'opera non implica il trasferimento dei suddetti diritti né li esaurisce.

Le fotocopie per uso personale (cioè privato e individuale, con esclusione quindi di strumenti di uso collettivo) possono essere effettuate, nei limiti del 15% di ciascun volume, dietro pagamento alla S.I.A.E del compenso previsto dall'art. 68, commi 4 e 5, della legge 22 aprile 1941 n. 633. Tali fotocopie possono essere effettuate negli esercizi commerciali convenzionati S.I.A.E. o con altre modalità indicate da S.I.A.E. Per le riproduzioni ad uso non personale (ad esempio: professionale, economico, commerciale, strumenti di studio collettivi, come dispense e simili) l'editore potrà concedere a pagamento l'autorizzazione a riprodurre un numero di pagine non superiore al 15% delle pagine del presente volume.

Le richieste vanno inoltrate a:
Centro Licenze e Autorizzazioni per le Riproduzioni Editoriali (CLEARedi),
Corso di Porta Romana 108, 20122 Milano
e-mail: autorizzazioni@clearedi.org e sito web: www.clearedi.org

L'autorizzazione non è concessa per un limitato numero di opere di carattere didattico riprodotte nell'elenco che si trova all'indirizzo
www.zanichelli.it/chi-siamo/fotocopie-e-permessi

L'editore, per quanto di propria spettanza, considera rare le opere fuori del proprio catalogo editoriale. La loro fotocopia per i soli esemplari esistenti nelle biblioteche è consentita, oltre il limite del 15%, non essendo concorrenziale all'opera. Non possono considerarsi rare le opere di cui esiste, nel catalogo dell'editore, una successiva edizione, le opere presenti in cataloghi di altri editori o le opere antologiche. Nei contratti di cessione è esclusa, per biblioteche, istituti di istruzione, musei ed archivi, la facoltà di cui all'art. 71-ter legge diritto d'autore. Per permessi di riproduzione, anche digitali, diversi dalle fotocopie rivolgersi a: ufficiocontratti@zanichelli.it

Coordinamento editoriale: Marika De Acetis

Realizzazione editoriale e indice analitico: Stilgraf, Bologna

Copertina:

– *Progetto grafico:* Falcinelli & Co., Roma

– *Immagine di copertina:* © pkujahe/Sinitar Photo/iStockphoto

Prima edizione Zanichelli: maggio 2012

Seconda edizione Zanichelli: gennaio 2021

Ristampa: **prima tiratura**

5 4 3 2 1 2021 2022 2023 2024 2025

Realizzare un libro è un'operazione complessa, che richiede numerosi controlli:
sul testo, sulle immagini e sulle relazioni che si stabiliscono tra essi.

L'esperienza suggerisce che è praticamente impossibile pubblicare un libro
privo di errori. Saremo quindi grati ai lettori che vorranno segnalarceli.

Per segnalazioni o suggerimenti relativi a questo libro scrivere al seguente indirizzo:

Zanichelli editore S.p.A.
Via Imerio 34
40126 Bologna
fax 051293322
e-mail: linea_universitaria@zanichelli.it
sito web: www.zanichelli.it

Prima di effettuare una segnalazione è possibile verificare se questa sia già stata inviata in precedenza,
identificando il libro interessato all'interno del nostro catalogo online per l'Università.

Per comunicazioni di tipo commerciale: universita@zanichelli.it

Stampa:

per conto di Zanichelli editore S.p.A.
Via Imerio 34, 40126 Bologna

Indice

Prefazione XIV

PARTE

1

CONCETTI FONDAMENTALI SU DNA, CROMOSOMI, CELLULE, SVILUPPO ED EREDITARIETÀ

1

Principi di base della struttura degli acidi nucleici e dell'espressione genica

- Il dogma centrale della biologia molecolare 4

1.1 La composizione degli acidi nucleici e dei polipeptidi 4

- Gli acidi nucleici e i polipeptidi sono sequenze lineari di semplici unità 4
- Il ruolo dei legami chimici nella stabilità e nella funzione delle macromolecole 8

1.2 L'appaiamento delle basi nel DNA e nell'RNA, la doppia elica e la replicazione del DNA 9

- Le basi metilate nel DNA 10
- L'accoppiamento alternativo delle basi e la conformazione del DNA 11
- La complementarietà e la notazione delle sequenze 12
- La replicazione semiconservativa 12
- La natura semidiscontinua della replicazione del DNA 13
- Le diverse DNA polimerasi dei mammiferi 14
- La struttura dell'RNA e i genomi a RNA 15

1.3 La trascrizione dell'RNA e l'espressione genica 17

- Il processo trascrizionale di base 18
- Le classi di RNA polimerasi nelle cellule eucariote 19
- L'RNA polimerasi II e la trascrizione dei geni nel nucleo che codificano per le proteine 19
- Diversi gruppi di geni per gli RNA sono trascritti dalle tre RNA polimerasi eucariote 20

1.4 La maturazione dell'RNA 21

- Lo splicing dell'RNA rimuove le regioni non necessarie dal trascritto primario 21
- Sono aggiunti nucleotidi specifici alle estremità della maggior parte dei trascritti prodotti dall'RNA polimerasi e dei precursori degli RNA transfer 23
- Gli RNA non codificanti sono prodotti dopo una serie di eventi di taglio e modifiche chimiche dei singoli nucleotidi 25

1.5 La traduzione, le modifiche post-traduzionali e la struttura delle proteine 27

- Gli RNA messenger sono decodificati per sintetizzare polipeptidi specifici 28
- Il codice genetico è degenerato e non è completamente universale 29
- Il processamento post-traduzionale: le modifiche chimiche degli amminoacidi e il taglio dei polipeptidi 30
- La complessa relazione tra sequenza amminoacidica e struttura proteica 34

FOCUS 1.1 Le principali classi di proteine coinvolte nella replicazione del DNA 14

FOCUS 1.2 L'estrema variabilità dei genomi virali 16

RIASSUNTO 37

SUGGERIMENTI DI LETTURA 37

2

Nozioni fondamentali su cellule e cromosomi

2.1 La struttura, la varietà cellulare e l'evoluzione della cellula 40

- I procarioti e gli eucarioti rappresentano una separazione fondamentale delle forme di vita cellulari 40
- La suddivisione del genoma in compartimenti cellulari distinti negli eucarioti 42
- La straordinaria diversità delle cellule nel corpo 43
- L'evoluzione delle cellule eucariote e l'origine di mitocondri, nucleo e citoscheletro 46
- Semplici mutazioni geniche possono aver dato origine al fondamentale sviluppo di organismi pluricellulari 48

2.2 Il DNA e il numero di copie di cromosomi durante il ciclo cellulare 49

- La ploidia può variare nelle cellule di un singolo individuo 49
- Differenze nella ploidia e nel contenuto di DNA durante il ciclo cellulare 50

2.3 La divisione cellulare e la trasmissione del DNA alle cellule figlie 52

- La mitosi è la modalità normale di divisione cellulare 52
- La meiosi è una forma specializzata di divisione cellulare che produce spermatozoi e cellule uovo 53
- La mitosi e la meiosi hanno importanti somiglianze e differenze 56
- La replicazione e la segregazione del DNA mitocondriale 57

2.4 La struttura e la funzione dei cromosomi 57

- Il DNA cromosomico è compattato attraverso un ripiegamento che inizia con il legame delle proteine istoniche per formare i nucleosomi 58
- Eucromatina, eterocromatina e la variabilità nella compattazione della cromatina in interfase 59
- Ogni cromosoma ha il suo territorio specifico nel nucleo interfascico 60
- I centromeri svolgono un ruolo centrale nei movimenti dei cromosomi ma si sono evoluti in modi molto diversi in diversi organismi 60
- Esistono poche evidenze riguardo a sequenze consenso conservate nelle origini di replicazione degli eucarioti complessi 62
- I telomeri hanno strutture specializzate per preservare le estremità dei cromosomi lineari 63

FOCUS 2.1 L'organizzazione intracellulare delle cellule animali 42

FOCUS 2.2 Le divisioni cellulari asimmetriche 46

FOCUS 2.3 Le componenti del fuso mitotico 61

RIASSUNTO 66

SUGGERIMENTI DI LETTURA 66

3

Principi basilari sulle interazioni tra cellule e sulla biologia del sistema immunitario

3.1 I principi della segnalazione cellulare 69

- Le molecole di segnalazione si legano a specifici recettori sulle cellule riceventi per indurre un'alterazione del comportamento cellulare 70
- Alcune molecole segnale legano recettori intracellulari che attivano direttamente i geni bersaglio 72
- La via di segnalazione attraverso i recettori della superficie cellulare spesso coinvolge cascate di chinasi 73
- Le piccole molecole intermedie della segnalazione intracellulare nella trasduzione del segnale 74
- La segnalazione sinaptica non richiede l'attivazione dei fattori di trascrizione 75

3.2 La proliferazione cellulare e la morte cellulare programmata 76

- La maggior parte delle cellule negli animali adulti non si divide, ma alcuni tessuti e cellule si rinnovano rapidamente 76
- I mitogeni promuovono la proliferazione cellulare superando i meccanismi che frenano la progressione del ciclo cellulare in G1 77
- I limiti della proliferazione cellulare e il concetto di senescenza cellulare 79
- Un grande numero delle nostre cellule è programmato in modo naturale per morire 79
- L'importanza della morte cellulare programmata 80
- L'apoptosi è effettuata dalle caspasi in risposta a segnali di morte o a prolungato stress cellulare 82

3.3 L'adesione cellulare e la formazione dei tessuti 83

- I diversi tipi di giunzioni cellulari possono regolare il contatto tra le cellule 84
- La matrice extracellulare regola il comportamento cellulare e serve da impalcatura per sostenere i tessuti 85
- I tipi cellulari specializzati sono organizzati in tessuti 87

3.4 La biologia del sistema immunitario 89

- L'origine delle cellule del sistema immunitario e loro caratteristiche 90
- Il sistema immunitario innato: combattere i patogeni utilizzando barriere e una risposta rapida basata su una modalità generale di riconoscimento dei patogeni 94
- Il sistema immunitario adattativo attiva risposte immunitarie specifiche potenziate dalle cellule della memoria 98
- L'immunità umorale dipende dalle attività di anticorpi solubili 99
- Nell'immunità mediata da cellule, le cellule T riconoscono cellule che contengono frammenti di proteine estranee 100

FOCUS 3.1 La struttura dei fattori di trascrizione 71

FOCUS 3.2 Il complesso maggiore di istocompatibilità e la struttura e funzione delle proteine MHC 102

RIASSUNTO 105

SUGGERIMENTI DI LETTURA 105

4

Le caratteristiche delle prime fasi dello sviluppo, del differenziamento cellulare e delle cellule staminali

4.1 I tipi di cellule e il differenziamento cellulare nelle prime fasi dello sviluppo dei mammiferi 107

- Una panoramica dello sviluppo dei mammiferi 108
- Il controllo epigenetico dello sviluppo e la necessità d'interrompere la simmetria 109
- Fecondazione: l'inizio di una nuova vita 110
- Dallo zigote umano allo stadio di blastocisti e allo sviluppo di tre tipi di linee cellulari 111
- L'impianto 114
- La gastrulazione e la formazione dei tre foglietti germinativi embrionali 115
- Le decisioni sul destino cellulare sono generate da segnali dai tessuti circostanti o attraverso divisioni cellulari asimmetriche 118
- Le basi molecolari della specificazione cellulare primitiva e modelli alternativi di come i differenti tipi cellulari si sviluppano 120
- Lo sviluppo delle cellule germinali e la determinazione del sesso nei mammiferi 122

4.2 Le cellule staminali e il differenziamento cellulare 124

- Perché le cellule staminali sono importanti per la ricerca scientifica e medica? 125
- La divisione delle cellule staminali e l'equilibrio tra la proliferazione delle cellule staminali e il differenziamento 126
- Le cellule staminali tissutali sono importanti per il rinnovamento tissutale e sono mantenute in microambienti specializzati 127
- La manipolazione delle cellule ottenute da embrioni nei primi stadi dello sviluppo per generare linee immortali di cellule staminali pluripotenti 128
- Generare linee di cellule staminali pluripotenti attraverso la riprogrammazione epigenetica dei genomi delle cellule differenziate 133
- Dirigere artificialmente il differenziamento e progettare il transdifferenziamento delle cellule umane 135

FOCUS 4.1 Le membrane extraembrionali e la placenta nei mammiferi 108

FOCUS 4.2 La compattazione degli embrioni dei mammiferi: polarità delle prime cellule e basi per lo sviluppo di due tipi di cellule 113

FOCUS 4.3 Formazione dei gemelli nella specie umana 117

RIASSUNTO 136

SUGGERIMENTI DI LETTURA 137

5

I modelli di trasmissione ereditaria

5.1 L'eredità monogenica e multifattoriale 140

5.2 I modelli di alberi genealogici mendeliani 142

- L'identificazione della modalità di trasmissione 145
- Le complicazioni degli alberi genealogici rispetto ai tratti mendeliani semplici 146

5.3 Il mosaicismo e le nuove mutazioni 149

- Le anomalie genetiche possono essere costituzionali o presentarsi come un mosaico 149
- Le nuove mutazioni sono spesso presenti in forma di mosaico 151
- Come rilevare il mosaicismo 152

5.4 I caratteri non mendeliani 153

- Agli inizi del XX secolo ci fu una controversia tra i fautori dell'eredità mendeliana e di quella quantitativa 153
- La teoria poligenica spiega come i caratteri quantitativi possono essere geneticamente determinati 153
- L'ereditabilità misura la proporzione della variabilità totale di un carattere che dipende da differenze genetiche 155
- Il modello dell'effetto soglia estende la teoria poligenica ai caratteri dicotomici 156

FOCUS 5.1 Riassunto dei principali tipi di ereditarietà 143

RIASSUNTO 158

SUGGERIMENTI DI LETTURA 159

PARTE 2

LA COMPRESIONE DEI GENOMI

6

Le tecnologie chiave per studiare il DNA: amplificazione, ibridazione e sequenziamento

6.1 Il clonaggio del DNA nelle cellule batteriche 164

- Trasformazione delle cellule batteriche con DNA ricombinante 164
- L'amplificazione 165
- I vettori 165
- La separazione fisica dei cloni 165
- La produzione del DNA ricombinante 166
- Un esempio di clonaggio standard di DNA nelle cellule batteriche usando un plasmide e una cellula ospite modificata geneticamente 168
- Il principio delle librerie di DNA, le applicazioni e i limiti del clonaggio del DNA 169

- Il clonaggio e l'espressione del DNA nelle cellule batteriche come mezzo per produrre grandi quantità della proteina d'interesse 170

- La necessità di proteine di fusione e tag di affinità 171

- La tecnica del *phage display* 171

6.2 L'amplificazione del DNA mediante replicazione *in vitro* 173

- La PCR: caratteristiche di base e quantificazione 173

- PCR quantitativa e real time PCR 175

- Vantaggi e svantaggi della PCR 175

- L'amplificazione isoterma è un'alternativa alla PCR per amplificare sequenze di DNA *in vitro* 176

- Metodi non selettivi di amplificazione del DNA per ottenere materiale sufficiente da campioni con una quantità limitata di acido nucleico 176

6.3 Ibridazione degli acidi nucleici: principi e usi 177

- Gli eteroduplex artificiali 177

- Due usi diversi dell'ibridazione degli acidi nucleici 178

- I saggi d'ibridazione: usare acidi nucleici noti o oligonucleotidi per trovare sequenze correlate in una popolazione di acidi nucleici 179

- Le due classi di saggi d'ibridazione 182

- Saggi d'ibridazione standard con sonde per visualizzare acidi nucleici immobilizzati 184

- L'ibridazione con microarray: ibridazione in parallelo su larga scala di acidi nucleici marcati con sonde non marcate 185

- L'ibridazione degli acidi nucleici per purificare selettivamente le sequenze d'interesse 186

6.4 I principi del sequenziamento del DNA e il sequenziamento di Sanger con dideossinucleotidi 187

- I concetti di base del sequenziamento di Sanger con dideossinucleotidi 187

6.5 Il sequenziamento massivo parallelo del DNA (sequenziamento di nuova generazione) 190

- Il sequenziamento massivo in parallelo di DNA amplificato: una panoramica 193

- La costruzione di librerie di DNA 194

- L'amplificazione di frammenti di DNA separati 194

- Le reazioni di sequenziamento del DNA 195

- Il sequenziamento massivo parallelo di DNA amplificato: le piattaforme di sequenziamento più usate 197

- Il sequenziamento massivo in parallelo di DNA non amplificato 201

FOCUS 6.1 Le endonucleasi di restrizione, da protettori dei batteri a strumenti per la genetica 167

FOCUS 6.2 La marcatura degli acidi nucleici 180

FOCUS 6.3 Elettroforesi su gel ed elettroforesi capillare per separare gli acidi nucleici in base alle dimensioni 189

FOCUS 6.4 Il pirosequenziamento 198

RIASSUNTO 203

SUGGERIMENTI DI LETTURA 204

7

L'analisi della struttura e dell'espressione dei geni e dei genomi

7.1 L'analisi della struttura del genoma e i progetti genoma 206

- Le mappe strutturali sono necessarie per il sequenziamento del genoma completo 207

- L'ordine lineare dei cloni del DNA genomico in un contiguo di sequenze deve rispecchiare le posizioni subcromosomiche originali 207
- Il Progetto Genoma Umano è stato uno sforzo internazionale e il primo grande progetto della biologia 208
- Le prime mappe genetiche umane avevano una risoluzione bassa e furono costruite principalmente con marcatori di DNA anonimi 209
- La creazione di mappe ad alta densità di marcatori del DNA e di contigui di cloni per ciascun cromosoma umano 210
- La corsa per ottenere una bozza della sequenza del genoma umano 213
- I progetti genoma per diversi organismi modello 216
- Le potenti banche dati genomiche e i browser genetici aiutano a immagazzinare e analizzare i dati genomici 216
- I metodi bioinformatici per predire i geni, le loro funzioni e l'annotazione dei geni 218
- Mettere in ordine le sequenze nei genomi complessi 220

7.2 Le analisi di base dell'espressione genica 222

- La quantificazione dell'espressione di geni individuali utilizzando la PCR quantitativa real time 222
- La mappatura dell'espressione ad alta risoluzione mediante ibridazione *in situ* e immunocitochimica 223

7.3 L'analisi in parallelo a elevata processività dell'espressione genica 226

- I microarray di DNA e di oligonucleotidi permettono di determinare il profilo trascrizionale rapidamente e globalmente 226
- L'approccio moderno allo studio dei profili di espressione genica globale sfrutta sempre più il sequenziamento per quantificare i trascritti 229
- L'analisi dell'espressione globale delle proteine rilevata mediante la spettrometria di massa 231

7.4 Le analisi genomiche su singola cellula 235

- La comprensione delle variazioni tra cellule: le innumerevoli applicazioni della genomica a singola cellula in biologia 236
- Le prospettive di avanzamento della ricerca medica usando la genomica a singola cellula 238
- La tecnologia dei saggi di sequenziamento basati sul DNA in singole cellule 239

FOCUS 7.1 Clonaggio di grandi frammenti di DNA con cromosomi artificiali di lievito e batterici 212

FOCUS 7.2 Le principali tappe nella mappatura e nel sequenziamento del genoma umano 214

FOCUS 7.3 Di chi è il genoma umano sequenziato? 215

FOCUS 7.4 La ricerca di omologie di sequenza 219

FOCUS 7.5 I metodi di rilevamento usati nella PCR quantitativa real time 224

FOCUS 7.6 La produzione di anticorpi 225

FOCUS 7.7 L'analisi dei dati di espressione ottenuti con i microarray 227

FOCUS 7.8 La spettrometria di massa nella proteomica 234

RIASSUNTO 243

SUGGERIMENTI DI LETTURA 244

8

Principi fondamentali della manipolazione genetica nelle cellule di mammifero

Una panoramica dei meccanismi di editing genomico, silenziamento genico e transgenesi della linea germinale 246

8.1 Il trasferimento artificiale di materiale genetico all'interno delle cellule di mammifero 248

- Le colture cellulari: colture di cellule primarie, metodi di coltura e conservazione delle cellule 249
- L'utilità e l'ottenimento di linee cellulari immortalizzate 250
- Metodi non virali per il trasferimento del materiale genetico nelle cellule di mammifero 251

8.2 I principi dell'espressione dei transgeni nelle cellule di mammifero 259

- I promotori inducibili permettono di controllare l'espressione dei transgeni 260
- I metodi per l'espressione stabile e transiente in cellule di mammifero 261

8.3 I meccanismi di editing genomico basati sulla ricombinazione omologa 262

- L'utilizzo della ricombinazione omologa per ottenere editing genomico: le strategie generali e la necessità di trovare sistemi di selezione 263

8.4 L'utilizzo di endonucleasi programmabili sito-specifiche per ottenere editing genomico 266

- L'utilizzo dei meccanismi di riparazione delle rotture del doppio filamento del DNA 266
- Le strategie di editing genomico che prevedono l'utilizzo di endonucleasi sito-specifiche contenenti un dominio di taglio per il DNA e un dominio modulare di legame al DNA 267
- Strategie di editing genomico che prevedono l'utilizzo di endonucleasi guidate da RNA nel sistema CRISPR/Cas 269

8.5 Il silenziamento genico 271

- La tecnologia antisense come mezzo per silenziare l'espressione di uno specifico gene 271
- L'interferenza a RNA è il metodo principale per valutare la funzione di un gene nelle cellule di mammifero in coltura 272

8.6 I meccanismi di transgenesi della linea germinale e gli animali transgenici 274

- Il trasferimento genico nei gameti e nei precursori della linea germinale 274
- La microiniezione nel pronucleo: un metodo consolidato per ottenere organismi transgenici 275
- La manipolazione genetica di cellule embrionali pluripotenti per trasmettere sequenze modificate alla linea germinale 276
- I metodi di manipolazione genica utilizzati per creare animali transgenici attraverso alleli nulli o specifiche mutazioni puntiformi 277
- Il trasferimento nucleare come metodo per produrre animali domestici geneticamente modificati 278

FOCUS 8.1 Come ottenere linee linfoblastoidi trasformate da EBV 251

FOCUS 8.2 L'interferenza a RNA come meccanismo naturale di difesa cellulare 272

FOCUS 8.3 Inattivazione genica condizionale in animali transgenici 278

RIASSUNTO 280

SUGGERIMENTI DI LETTURA 282

9

La scoperta dell'architettura e del funzionamento del genoma umano

9.1 Una panoramica del genoma umano 283

- Il genoma mitocondriale ricorda un genoma batterico ridotto all'essenziale e il suo numero di copie varia in modo molto significativo tra le cellule 284

- L'autonomia limitata del genoma mitocondriale e l'utilizzo di un codice genetico diverso 286
- Il trasferimento recente e continuo di sequenze di mtDNA nel genoma nucleare 287
- Il genoma nucleare umano è relativamente povero di geni e ha rilevanti quantità di DNA eterocromatico 288
- La composizione variabile in basi del genoma nucleare e il significato delle isole CpG 290
- Il genoma umano contiene circa 20000 geni che codificano proteine, ma non è possibile stabilire il loro numero totale esatto 291
- Esistono numerosi RNA umani non codificanti, ma identificare e perfino definire i geni degli RNA funzionali non è immediato 292
- Il lungo viaggio per finire la sequenza del genoma nucleare 295
- Una breve panoramica di alcune risorse informatiche utilizzate per studiare la sequenza del genoma umano e i prodotti genici 297

9.2 L'organizzazione e la distribuzione dei geni nel genoma umano 300

- I geni umani mostrano una grande variabilità nelle dimensioni e nell'organizzazione interna 300
- La densità genica variabile e le unità di trascrizione sovrapposte nel genoma nucleare 302
- Le diverse origini dei geni duplicati nel genoma umano 302
- I diversi modi con cui sono organizzate le famiglie di geni codificanti proteine 306
- L'organizzazione e la distribuzione delle famiglie di geni codificanti RNA 308

9.3 L'eterocromatina e le unità ripetute di trasposoni 310

- L'eterocromatina costitutiva è in gran parte definita da lunghe unità di DNA ripetute in tandem 310
- Le ripetizioni derivate da trasposoni costituiscono la maggior parte del genoma umano e si sono sviluppate in gran parte attraverso la retrotrasposizione 312
- La doppia natura dei trasposoni: amici e nemici 314

9.4 Primi passi per capire come funziona il nostro genoma 315

- Il progetto ENCODE: il primo tentativo sistematico di catalogare elementi funzionali del DNA nel genoma umano 316
- L'approccio alle funzioni dei geni umani attraverso la manipolazione genetica delle cellule in coltura e l'estrapolazione da organismi modello 320
- Gli sforzi internazionali per produrre mappe di proteomi umani e interattomi delle proteine umane 323

FOCUS 9.1 Le isole di CpG nei vertebrati 290

FOCUS 9.2 Gli pseudogeni, i retrogeni e la copiatura diretta dagli RNA di geni codificanti proteine 305

FOCUS 9.3 La CHIP-Seq, un metodo per definire i siti di legame delle proteine associate al DNA 319

FOCUS 9.4 L'identificazione delle interazioni tra proteine mediante screening con doppio ibrido del lievito 324

RIASSUNTO 326

SUGGERIMENTI DI LETTURA 326

10

La regolazione genica e l'epigenoma

10.1 Accessibilità e conformazione della cromatina 330

- Il posizionamento dei nucleosomi e i complessi di rimodellamento della cromatina 332

10.2 Gli istoni e le altre proteine che legano il DNA 333

- I fattori trascrizionali e altre proteine che legano il DNA 335

10.3 La regolazione mediata dalla metilazione del DNA e dagli RNA non codificanti 336

- Il significato delle sequenze CpG 338
- I ruoli dell'RNA nella regolazione dell'espressione genica 340
- Le interazioni tra i meccanismi epigenetici 342

10.4 L'inattivazione dell'X: un cambiamento epigenetico trasmissibile da una cellula alla cellula figlia, ma non dai genitori ai figli 342

- A livello dei *loci* soggetti a imprinting, l'espressione dipende dall'origine parentale 346
- La memoria epigenetica transgenerazionale: un argomento poco conosciuto e controverso 350

10.5 La creazione del trascritto: promotori ed enhancer 351

- La trascrizione richiede l'assemblaggio sul promotore di un complesso di preinizio 351
- Gli enhancer 353
- L'allungamento del trascritto 355

10.6 La regolazione post-trascrizionale 357

- Lo splicing alternativo permette a un trascritto primario di codificare molte isoforme di una proteina 357
- L'editing dell'RNA può cambiare la sequenza dell'mRNA dopo la trascrizione 358
- La regolazione della traduzione 359
- La scoperta di molti piccoli RNA che regolano l'espressione genica ha determinato un cambiamento di paradigma nella biologia cellulare 359

FOCUS 10.1 La tecnica *Chromosome Conformation Capture* 332

FOCUS 10.2 Nomenclatura delle modifiche istoniche 334

FOCUS 10.3 Lo studio della metilazione del DNA 338

RIASSUNTO 362

SUGGERIMENTI DI LETTURA 363

PARTE

3

VARIAZIONI GENETICHE TRA INDIVIDUI E SPECIE

11

Una panoramica delle variazioni genetiche umane

11.1 Le origini delle variazioni delle sequenze del DNA 368

- Le variazioni genetiche derivanti da errori endogeni nella funzione del DNA e dei cromosomi 369
- Varie sorgenti endogene ed esogene possono causare danno al DNA alterando la sua struttura chimica 370

11.2 La riparazione del DNA 372

- La riparazione del danno del DNA o della sequenza alterata su un singolo filamento 373
- La riparazione delle lesioni del DNA che colpiscono entrambi i filamenti 375
- Il danno al DNA non rilevato, la tolleranza del danno al DNA e la sintesi translesione 377

11.3 La genomica delle popolazioni e la scala delle variazioni genetiche umane 378

- Le varianti del DNA, i polimorfismi e la genomica delle popolazioni 378
- Le variazioni su piccola scala: le varianti a singolo nucleotide e le piccole inserzioni e delezioni 380
- Le variazioni strutturali: inversioni, traslocazioni, inserzioni e delezioni di ampie dimensioni, e varianti del numero di copie 382
- Il sequenziamento dell'intero genoma permette la misurazione diretta dei tassi di mutazione della linea germinale e l'analisi completa delle variazioni del DNA nella linea germinale 383
- I cambiamenti post-zigotici del DNA: le mutazioni *de novo*, gli alti tassi di mutazioni somatiche e le variazioni strutturali eccezionali 387

11.4 Le variazioni genetiche funzionali e le variazioni nelle proteine 389

- Le variazioni nelle sequenze delle proteine 389
- La maggior parte delle variazioni genetiche non ha effetti sul fenotipo, ma una piccola frazione è dannosa ed è soggetta a selezione purificante 390
- La selezione positiva sostiene i cambiamenti adattativi promuovendo la frequenza delle varianti del DNA vantaggiose 391
- La diversità delle proteine generata dalla duplicazione genica e dal processamento alternativo di un gene 393

11.5 Le eccezionali variazioni genetiche nel sistema immunitario adattativo 393

- I meccanismi somatici che permettono la produzione specifica d'immunoglobuline a livello di cellule e di recettori dei linfociti T 395
- Le funzioni delle proteine del sistema maggiore di istocompatibilità (MHC) e il concetto di limitazione delle proteine MHC 398
- Qual è la base dell'eccezionale polimorfismo delle proteine MHC? 399
- L'importanza del sistema HLA in medicina 400

FOCUS 11.1 Per varie ragioni lo schema delle variazioni a singolo nucleotide nel genoma umano non è casuale 381

FOCUS 11.2 Effetti delle differenze sessuali e dell'età parentale sui tassi di mutazione nelle linee germinali 384

FOCUS 11.3 Una forte selezione positiva recente può indurre uno *sweep* selettivo con cancellazione locale delle variazioni genetiche 392

FOCUS 11.4 I geni HLA, gli alleli e gli aplotipi 401

RIASSUNTO 402

SUGGERIMENTI DI LETTURA 403

12**La genetica delle popolazioni umane****12.1 Le frequenze alleliche e le frequenze dei genotipi: la relazione di Hardy-Weinberg** 405

- Un esperimento mentale: scegliere i geni da un insieme di geni 405
- La predizione delle frequenze dei portatori di condizioni mendeliane 407

12.2 Le frequenze degli aplotipi e il *linkage disequilibrium* 408**12.3 Il cambiamento delle frequenze alleliche** 412

- La stima dei tassi di mutazione 412

- Le varianti neutre e la deriva genetica 413
- La selezione 415
- La manipolazione delle frequenze genetiche: l'eugenetica 418

12.4 La struttura della popolazione e l'endogamia 419

- L'identificazione della discendenza 420
- La consanguineità e il coefficiente di relazione 421

FOCUS 12.1 Il controllo della distribuzione di Hardy-Weinberg 406

FOCUS 12.2 Le misure del *linkage disequilibrium* 409

FOCUS 12.3 L'equilibrio tra mutazione e selezione 417

FOCUS 12.4 La selezione contro una condizione autosomica recessiva 419

RIASSUNTO 424

SUGGERIMENTI DI LETTURA 425

13**La genomica comparata e l'evoluzione del genoma****13.1 La genomica comparata** 428

- L'identificazione di sequenze di DNA funzionalmente importanti che sono state conservate mediante selezione negativa (purificante) 428
- L'identificazione delle basi dei recenti adattamenti evolutivi specifici del lignaggio 429
- Diversi programmi informatici consentono gli allineamenti automatizzati delle sequenze del genoma 430
- L'utilizzo di confronti tra specie per convalidare o rifiutare la predizione di geni e per identificare nuovi geni 432
- L'uso della genomica comparata per stimare la porzione di genoma sottoposta a selezione negativa 433
- La genomica comparata come metodo globale per identificare elementi di DNA non codificante conservati 435
- Risorse online per la genomica comparata 438

13.2 La duplicazione genica, le differenze tra le specie nel numero dei geni e i vantaggi evolutivi degli esoni 440

- I paradossi del valore C e del valore G: la complessità non è semplicemente correlata con le dimensioni del genoma o con il numero dei geni 440
- Nei lignaggi dei vertebrati, dopo la divisione dai tunicati, sono avvenuti due o tre eventi importanti di duplicazione dell'intero genoma 441
- La duplicazione genica: un'importante via dell'evoluzione verso lo sviluppo di una complessità funzionale 442
- La duplicazione genica e la divergenza: l'esempio della famiglia genica delle globine 444
- L'importanza evolutiva dei cambiamenti nel numero di copie di geni di uno specifico lignaggio 445
- La duplicazione e il mescolamento degli esoni determinano una maggiore complessità genica 447

13.3 L'evoluzione dei cromosomi dei mammiferi 448

- Durante l'evoluzione del genoma dei mammiferi sono avvenuti importanti riarrangiamenti cromosomici 448
- Nei cromosomi sessuali eteromorfi, quello più piccolo è presente in uno solo dei due sessi, possiede pochi geni e di solito non presenta ricombinazione 450
- Le regioni pseudoautosomiche non sono molto conservate e sono evolutivamente instabili 451

- I cromosomi del sesso nei mammiferi si sono evoluti in seguito alla comparsa di un *locus* che determina il sesso su un autosoma, facendolo divergere dal cromosoma omologo corrispondente 452
- Nella maggior parte dei casi, i geni abbondantemente espressi nei testicoli e localizzati sul cromosoma Y sono mantenuti dalla conversione genica intracromosomica 454
- L'inattivazione del cromosoma X si è sviluppata in risposta alla perdita di geni dal cromosoma Y 456

13.4 L'evoluzione delle sequenze regolatrici e l'origine delle sequenze funzionali dai trasposoni 456

- Il DNA non codificante dei metazoi complessi è dominato da sequenze derivate da elementi trasponibili ed è instabile dal punto di vista evolutivo 456
- La divergenza delle sequenze regolatrici come modo per spiegare la divergenza morfologica 458
- Nuove sequenze funzionali, tra cui esoni, elementi regolatori, lncRNA e a volte perfino geni, hanno spesso origine da elementi trasponibili 460

13.5 La filogenetica e il nostro posto nell'albero della vita 462

- La filogenetica molecolare utilizza allineamenti di sequenze per costruire alberi evolutivi 463
- Gli alberi evolutivi possono essere costruiti in modi diversi e la loro affidabilità è testata con metodi statistici 463

FOCUS 13.1 La valutazione della pressione selettiva sulle sequenze di DNA codificante utilizzando i rapporti *dN/dS* 430

FOCUS 13.2 Una controversia post-ENCODE: il DNA spazzatura (*junk*) e la proporzione del genoma che è funzionalmente importante 436

FOCUS 13.3 Le classi dei Chordata 440

FOCUS 13.4 La maggior parte dei nuovi geni deriva dalla copia di geni vecchi o di segmenti genici funzionali, ma alcuni sono originati *de novo* dal DNA non codificante 461

RIASSUNTO 465

SUGGERIMENTI DI LETTURA 466

14

L'evoluzione umana

14.1 Le origini della specie umana 469

- Il posizionamento degli esseri umani all'interno della classe dei mammiferi 470
- Le testimonianze fossili e archeologiche delle origini umane 471
- L'analisi del DNA dai fossili: lo studio del DNA antico 474
- Le sequenze di DNA degli antichi ominini 475

14.2 La storia dell'evoluzione umana attraverso le sequenze genomiche 476

- Le origini e l'unicità della specie umana: gli approfondimenti genetici 476
- Gli esseri umani hanno poca diversità genetica rispetto alla maggior parte degli altri mammiferi 477
- L'estesa condivisione di varianti comuni all'interno della popolazione umana e il modello Out of Africa 480
- Le prove dell'introgresione arcaica 483

14.3 La deduzione della storia di maschi e femmine sulla base del DNA mitocondriale e del cromosoma Y 485

- Gli alberi del DNA mitocondriale e dell'MSY hanno le loro radici in Africa 486
- Altre informazioni dal confronto tra il mtDNA e l'MSY 488

14.4 Le conseguenze della nostra storia evolutiva per la salute umana 489

- La selezione negativa ubiquitaria agisce sulla variabilità genomica umana 490
- La selezione positiva di varianti vantaggiose 491
- L'importanza delle malattie infettive nell'evoluzione umana 493
- La genetica evolutiva delle malattie infiammatorie 496

FOCUS 14.1 Il linguaggio, la genetica e la storia 474

FOCUS 14.2 La dimensione effettiva di una popolazione 478

RIASSUNTO 496

SUGGERIMENTI DI LETTURA 497

PARTE

4

MALATTIE GENETICHE UMANE

15

Le anomalie cromosomiche e le varianti strutturali

15.1 L'analisi dei cromosomi umani 501

- L'analisi del cariotipo al microscopio 502
- L'ibridazione *in situ* fluorescente è una sintesi delle tecniche di microscopia e di biologia molecolare 505
- L'ibridazione genomica comparativa permette d'identificare gli sbilanciamenti in tutto il genoma 506

15.2 Le grandi anomalie cromosomiche 508

- Le anomalie numeriche includono le poliploidie e le aneuploidie 509
- Le conseguenze cliniche delle anomalie cromosomiche di numero 509
- Diverse anomalie cromosomiche costituzionali sono il risultato di errata riparazione del danno, inappropriata ricombinazione o errori nella duplicazione 511
- Le conseguenze delle anomalie cromosomiche strutturali 512

15.3 Le varianti strutturali, le microdelezioni e le microduplicazioni 516

- Le varianti strutturali ricorrenti e non ricorrenti 517
- L'invasione del filamento può portare a conversione genica 521
- Conclusioni 522

FOCUS 15.1 La nomenclatura delle bande cromosomiche 504

RIASSUNTO 523

SUGGERIMENTI DI LETTURA 523

16

Le basi molecolari delle malattie: collegare i fenotipi ai genotipi

16.1 La perdita di funzione 526

- La delezione o la rottura di un gene normalmente porterebbero a una perdita di funzione 526

- La perdita di funzione potrebbe essere dovuta alla delezione o a un'alterazione del promotore 527
- La rimozione dell'accesso a un enhancer può causare una perdita di funzione tessuto-specifica 528
- Lo splicing aberrante causa frequentemente la perdita di funzione 529
- I cambiamenti che colpiscono il codone d'inizio AUG influenzeranno la traduzione 531
- Le inserzioni e le delezioni spesso causano scivolamento delle cornici di lettura 532
- I codoni di stop prematuri di solito agiscono come mutazioni nulle 533
- I cambiamenti missenso possono o non possono influenzare la funzione della proteina 534
- Qualche volta una condizione recessiva dovuta a perdita di funzione dipende da un modesto livello residuo della funzione 536
- Gli effetti dominanti negativi si verificano in una persona eterozigote quando il prodotto del gene mutato interferisce con la funzione del prodotto normale 537

16.2 L'acquisto di funzione 538

- Di quando in quando un cambiamento missenso porta un enzima a catalizzare una nuova reazione 538
- Un prodotto genico può acquisire una funzione attraverso vari meccanismi 539
- Alcune mutazioni producono un RNA con nuove proprietà tossiche 540
- Alcune mutazioni producono una proteina mutante con nuove proprietà tossiche 541
- I complessi meccanismi patogenetici delle mutazioni dinamiche 542

16.3 Le mutazioni dinamiche: le espansioni instabili di elementi ripetuti 543

- Le mutazioni dinamiche possono avvenire in molti geni e malattie differenti 543
- Alcuni geni mostrano l'espansione patogena di tratti di polialanine 545

16.4 I meccanismi patologici molecolari nelle malattie mitocondriali 546

- I difetti dei geni mitocondriali mostrano un'eredità matrilineare e possibile eteroplasmia 546
- Le malattie causate dalle mutazioni del mtDNA sono estremamente variabili 547

16.5 Le correlazioni tra genotipo e fenotipo 549

- Le mutazioni che determinano perdita o acquisto di funzione nello stesso gene causeranno differenti fenotipi 549
- I cambiamenti che determinano perdita di funzione generano malattie sia dominanti sia recessive 552
- Si cercano spesso le correlazioni tra genotipo e fenotipo, ma si trovano raramente 554
- Lo sviluppo di banche dati dei fenotipi su vasta scala sarà uno strumento importante per il futuro della patologia molecolare 558

FOCUS 16.1 La predizione degli effetti dello splicing 530

FOCUS 16.2 Predizioni *in silico* degli effetti dei cambiamenti missenso 535

RIASSUNTO 558

SUGGERIMENTI DI LETTURA 559

17

La mappatura e l'identificazione di geni che causano malattie monogeniche

17.1 Il clonaggio posizionale inizialmente cerca di identificare i geni responsabili di una malattia mappandoli in una precisa posizione cromosomica 562

- I ricombinanti sono identificati genotipizzando i genitori e la progenie per coppie di *loci* 562
- Il riconoscimento dei ricombinanti negli alberi genealogici umani non è sempre immediato 564
- La mappatura delle malattie umane fa affidamento su gruppi di marcatori genetici rappresentativi del genoma 565
- Una volta che si è mappato il *locus* per una malattia, è necessario molto altro lavoro per identificare la reale variante causativa 568

17.2 La condivisione degli aplotipi e l'autozigosità 569

- L'autozigosità è l'omozigosità per sequenze identiche per discendenza 569

17.3 Il sequenziamento dell'intero esoma e dell'intero genoma permette di affrontare senza ipotesi preconcepite e pregiudizi l'identificazione della causa di una condizione monogenica 572

- Il sequenziamento dell'esoma si basa su kit commerciali per la cattura degli esoni 573
- L'identificazione delle varianti vere nei dati del sequenziamento è ben lontana dall'essere banale 573

17.4 Le strategie basate sul sequenziamento degli esomi per identificare i geni associati a malattie 575

- L'identificazione del gene mutato nella sindrome di Miller, una condizione autosomica recessiva 576
- L'identificazione del gene mutato nella sindrome di Schinzel-Giedion, una condizione sporadica 577
- L'identificazione del gene mutato nella sindrome di Kabuki, una condizione dominante eterogenea 578
- L'identificazione di geni in condizioni molto eterogenee si basa sul ritrovamento di cambiamenti *de novo* 579

17.5 Come si conferma che il gene candidato è quello corretto 579

- Gli insiemi di dati esistenti possono fornire una grande quantità d'informazioni su un gene candidato 580
- Le varianti rilevanti potrebbero essere create e studiate in colture cellulari o nell'animale *in toto* 581

FOCUS 17.1 Calcolo dei lod score (Z) 566

FOCUS 17.2 Il calcolo bayesiano della soglia di significatività per un lod score a due punti 567

FOCUS 17.3 L'uso di *zebrafish* per l'analisi funzionale delle varianti associate a ciliopatie 582

RIASSUNTO 583

SUGGERIMENTI DI LETTURA 584

18

Le malattie complesse: identificare i fattori di suscettibilità e comprenderne la patogenesi

Introduzione 585

18.1 Lo studio delle malattie complesse: gli approcci epidemiologici 586

- Il rischio relativo (λ) è una misura del raggruppamento familiare 586
- Gli studi sui gemelli soffrono di numerose limitazioni 587
- Gli studi sulle adozioni rappresentano lo standard di riferimento per districare i fattori genetici da quelli ambientali 588
- L'analisi di segregazione tenta di usare i dati epidemiologici per comprendere l'architettura genetica dei caratteri multifattoriali 589

18.2 Lo studio delle malattie complesse usando il linkage 590

- Le coppie di fratelli affetti forniscono il materiale principale per l'analisi di linkage tra coppie di parenti 590
- I deludenti risultati degli studi di linkage hanno favorito l'interesse verso gli studi di associazione 591

18.3 Lo studio delle malattie complesse mediante l'associazione 592

- L'era moderna degli studi di associazione su scala genomica è iniziata nel 2007 con il Consorzio caso-controllo del Wellcome Trust 593
- Gli attuali studi di associazione su scala genomica usano la costruzione della fase, l'inferenza e la metanalisi 598
- Dalla statistica alla biologia 601

18.4 Le limitazioni degli studi di associazione su scala genomica 601

18.5 Che cosa abbiamo imparato dalla genetica dei caratteri complessi? 604

- L'identificazione delle varianti causative 605
- Andare avanti 607

FOCUS 18.1 Linkage e associazione 592

FOCUS 18.2 Misure del rischio 596

RIASSUNTO 608

SUGGERIMENTI DI LETTURA 609

19

Genetica e genomica del cancro

Introduzione 611

19.1 Gli oncogeni 613

- Le mutazioni driver spesso attivano gli oncogeni 614
- I microRNA possono agire come oncogeni 619
- L'attivazione di un oncogene risulta tumorigenica solamente in circostanze particolari 619

19.2 I geni oncosoppressori 620

- Lo studio del retinoblastoma ha fornito un paradigma per la comprensione dei geni oncosoppressori 620
- I microRNA spesso si comportano come i geni oncosoppressori 624
- I geni oncosoppressori sono spesso silenziati a livello epigenetico attraverso la metilazione del DNA 624

19.3 Gli oncogeni e gli oncosoppressori più importanti agiscono prevalentemente nella regolazione dei punti di controllo del ciclo cellulare e nel mantenimento dell'integrità del genoma 625

- pRb: un regolatore cruciale della progressione attraverso la fase G_1 626
- *CDKN2A*: un singolo gene codificante due proteine regolatrici 626
- p53: il guardiano del genoma 627
- I difetti nei meccanismi di riparazione del DNA sono una causa importante di instabilità genomica 628

19.4 Una visione del cancro su scala genomica 629

- Le analisi condotte su piattaforme multiple descrivono le variazioni a vari livelli: cromosomico, di sequenza del DNA, epigenetico, di RNA e proteico 629
- I dati genomici permettono una nuova classificazione dei tumori 632
- Pensare in termini di vie molecolari piuttosto che di singole mutazioni o geni riduce la complessità e suggerisce collegamenti con la lista di caratteristiche di Hanahan e Weinberg 633
- Gli approcci *genome-wide* possono identificare vie biochimiche inattese 634
- Alcuni tumori mostrano cambiamenti genetici coordinati su larga scala 635
- L'evoluzione di un tumore può essere desunta dalle analisi comparative o seguita direttamente attraverso studi su singola cellula 636

19.5 Applicazioni delle nostre nuove conoscenze sul cancro 638

- Terapie antitumorali mirate 638

FOCUS 19.1 La regolazione dei livelli di β -catenina nei tumori del colon-retto 623

RIASSUNTO 643

SUGGERIMENTI DI LETTURA 644

PARTE

5

GENETICA MOLECOLARE UMANA APPLICATA

20

I test genetici in ambito sanitario e legale

20.1 Che cosa analizzare e perché 648

20.2 Le analisi per la ricerca di una variante genetica specifica 649

- Gli SNP chip consentono la genotipizzazione massiva in parallelo degli SNP distribuiti nel genoma 652

20.3 I test clinici diagnostici 653

- Il sequenziamento è il metodo d'elezione per testare un campione per varianti su piccola scala 653
- Filtro dell'elenco di varianti 655
- I metodi non basati sul sequenziamento sono utilizzati per rispondere a domande specifiche 657
- I test devono essere sempre rigorosamente valutati prima di essere introdotti nei servizi clinici 658

20.4 Lo screening di popolazione 659

- Quantificare la performance di un test 659

- Lo screening per le condizioni genetiche può essere effettuato in diversi periodi della vita 660
- Lo screening per la suscettibilità a condizioni multifattoriali 665
- I risultati incidentali rappresentano una sorta di screening opportunistico 667

20.5 La farmacogenetica e la medicina personalizzata 668

- Molte differenze genetiche influenzano il metabolismo dei farmaci 669
- La variabilità genetica del bersaglio può avere effetto sulla farmacodinamica di una sostanza 672
- La warfarina è un banco di prova per l'uso di dati farmacogenetici nella medicina personalizzata 673
- La farmacogenetica e la farmacogenomica tendono a essere più di interesse accademico che pratico, ma questo potrebbe cambiare 674

20.6 L'analisi forense del DNA: identificazione individuale e di parentela 675

- L'analisi del profilo del DNA usa gli elementi brevi ripetuti in tandem amplificati mediante PCR 676
- I problemi giudiziari 680
- I problemi etici 680
- Test di paternità e di parentela 682

FOCUS 20.1 I chip di SNP 653

FOCUS 20.2 Definire le prestazioni di un test diagnostico 660

FOCUS 20.3 La curva caratteristica di funzionamento del ricevitore (ROC) 661

RIASSUNTO 682

SUGGERIMENTI DI LETTURA 683

21

Organismi modello e modelli di malattie

21.1 Una panoramica sugli organismi modello 685

- Gli organismi modello unicellulari aiutano a comprendere la biologia cellulare di base e i microbi patogeni 686
- Alcuni modelli di invertebrati offrono screening genetici ad alto rendimento e informazioni utili sulla funzione genica 687
- Vari modelli di pesci, rane e uccelli offrono vie accessibili per studiare lo sviluppo dei vertebrati 688
- I mammiferi sono i modelli animali più rilevanti, ma sono penalizzati da limitazioni pratiche e preoccupazioni etiche 689

21.2 Modelli cellulari di malattie 691

- I modelli di malattie basati sull'uso di colture cellulari tradizionali 691
- Le colture di cellule staminali pluripotenti permettono di estendere la gamma di modelli cellulari di malattie umane 693
- La generazione di modelli di malattie *in vitro* mediante l'uso di colture di organoidi basate sulle cellule staminali 695

21.3 Come si generano i modelli animali di malattie genetiche 696

- La generazione di modelli di malattie dovuti a cambiamenti del DNA che provocano un'acquisizione di funzione dannosa 697
- La creazione di modelli di malattie causate da mutazioni che provocano perdita di funzione 700

- Molti modelli animali di malattie genetiche umane sono stati identificati mediante lo screening del fenotipo in programmi di mutagenesi casuale su scala genomica 701
- La generazione di nuovi fenotipi mutanti murini e modelli di malattie attraverso progetti di mutagenesi su larga scala e ad alto rendimento 702

21.4 Quanto sono utili i modelli animali di malattie genetiche? 703

- I vantaggi dei modelli di malattie genetiche manipolabili geneticamente 704
- I roditori forniscono i modelli più ampiamente usati di malattie genetiche nei mammiferi 705
- I limiti dei modelli di roditori e di altri mammiferi e il motivo per cui creare modelli animali per alcuni disturbi in cui sono implicati singoli geni è stato impegnativo 707
- Le aspettative e le difficoltà della generazione di modelli di malattie nei primati 708
- Osservazioni conclusive 710

FOCUS 21.1 Le caratteristiche dei due principali modelli animali invertebrati 688

FOCUS 21.2 Le caratteristiche di alcuni diffusi modelli animali vertebrati 690

FOCUS 21.3 Tre modi diversi per generare modelli animali di sindrome di Down 698

FOCUS 21.4 I background genetici murini e i geni modificatori 706

FOCUS 21.5 Perché può essere difficile replicare i fenotipi umani nei topi 709

RIASSUNTO 711

SUGGERIMENTI DI LETTURA 711

22

Approcci genetici alla cura delle malattie

22.1 Una panoramica del trattamento della malattia genetica e del trattamento genetico della malattia 714

- Tre diversi approcci generali nel trattamento dei disturbi genetici 714
- Il trattamento genetico della malattia può essere condotto a molti livelli diversi 716

22.2 Il trattamento di malattie con proteine ingegnerizzate per via genetica 717

- Le proteine ricombinanti terapeutiche prodotte mediante ingegneria genetica 717
- Gli anticorpi geneticamente modificati con un potenziale terapeutico aumentato 717

22.3 I principi di base della terapia genica e gli interventi terapeutici basati su RNA (*RNA therapeutics*) 721

- Due strategie ad ampio raggio nella terapia genica somatica 721
- Il problema della consegna: progettare strategie ottimali e sicure per trasportare i costrutti genici nelle cellule dei pazienti 722
- La terapia genica è talvolta eseguita *in vivo*, ma la terapia genica *ex vivo* offre vantaggi significativi 724
- I sistemi non virali per il trasporto di costrutti genici terapeutici: sicurezza a spese dell'efficienza 724

- Il trasporto virale di costrutti genici terapeutici: efficienza relativamente alta ma problemi di sicurezza 726

22.4 La sovraespressione dei geni per il trattamento di malattie ereditarie recessive 728

- I molti successi della terapia *ex vivo* con aumentata espressione mirata alle cellule staminali ematopoietiche 729
- La terapia genica *in vivo*: approcci, barriere e recenti successi 731

22.5 Gli approcci terapeutici mediante l'uso di RNA, le prospettive terapeutiche dell'editing genetico e gli approcci genetici per prevenire le malattie 733

- La terapia mediante RNA: silenziamento genico con l'interferenza a RNA e la modulazione dello splicing dell'RNA 733
- Le prospettive dell'editing terapeutico del genoma usando le nucleasi programmabili 734
- La terapia sostitutiva mitocondriale: un caso specializzato di prevenzione della trasmissione

di malattie genetiche attraverso la modifica della linea germinale 736

FOCUS 22.1 La produzione di anticorpi monoclonali terapeutici completamente umani 719

FOCUS 22.2 Dal trapianto di midollo osseo alla terapia genica *ex vivo* 725

FOCUS 22.3 I due maggiori problemi di sicurezza nei trial di terapia genica 730

FOCUS 22.4 Lo studio del potenziale terapeutico per la terapia dell'HIV mediante trapianto di cellule staminali e l'editing genetico 737

RIASSUNTO 739

SUGGERIMENTI DI LETTURA 740

Glossario 741

Indice analitico 760

BANDEGGIO DEI CROMOSOMI UMANI 784

Prefazione

Ci sono stati enormi cambiamenti da quando, nel 2011, è stata pubblicata la quarta edizione inglese di *Genetica molecolare umana*, motivo per cui abbiamo completamente riscritto e riorganizzato il testo per questa edizione. Sono pochi i capitoli che hanno mantenuto la stessa identità, ma abbiamo cercato in tutto il libro di raggiungere lo stesso obiettivo di allora: fornire un inquadramento generale dei principi piuttosto che una lista di informazioni (per le quali è sicuramente più proficuo attingere alle fonti online), in modo da costruire un ponte tra il libro, che contiene i fondamenti della disciplina, e i risultati delle ricerche pubblicate negli articoli scientifici, oltre a comunicare il nostro entusiasmo per un'area della ricerca in così rapida evoluzione.

Lo sviluppo che da solo ha registrato i maggiori cambiamenti dal 2011 è il sequenziamento massivo del DNA, che si è espanso in ogni area della genetica umana. Di conseguenza, abbiamo deciso di trattare in modo più ampio e aggiornato le tecnologie di sequenziamento massivo in parallelo, compreso il nuovo, eccitante, settore degli studi di genomica su singola cellula. Per certi versi, questa rivoluzione nel sequenziamento ha reso le cose più semplici. Molte tecniche trattate nell'edizione precedente del libro sono state ampiamente o totalmente superate dal sequenziamento. Chi leggerà questo testo noterà, comunque, che abbiamo scelto di continuare a spiegare i cariotipi. Abbiamo deciso che è corretto illustrare le analisi dei cariotipi quando hanno un importante valore formativo: spesso è più facile capire cosa sta succedendo guardando un cariotipo piuttosto che dei dati di sequenza, nonostante molti laboratori oggi preferiscano usare il sequenziamento e non il microscopio per questi scopi.

Nella prefazione della precedente edizione avevamo scritto che «possiamo ragionevolmente aspettarci che il sequenziamento del genoma di un elevato numero di organismi e di persone sarà completato prima della prossima edizione di questo libro», e questa aspettativa è stata ampiamente soddisfatta. La genetica umana è ora solidamente entrata nel mondo dei Big Data e delle grandi collaborazioni internazionali, e ciò si riflette nella nostra trattazione.

Durante la revisione e la scrittura dei capitoli, siamo stati particolarmente grati a Mark Jobling e alla sua squadra, che hanno scritto l'eccellente *Human Evolutionary Genetics* (seconda edizione, Garland Science, 2013), quando hanno accettato di contribuire alla stesura di un capitolo sull'evoluzione umana. L'analisi del DNA contemporaneo e di quello antico hanno fatto passi da gigante in pochi anni e stanno svelando affascinanti informazioni sulle nostre origini e sulla nostra storia, ma nessuno di noi si sentiva in grado di scrivere con sufficiente autorità su un argomento così importante.

Le altre integrazioni e revisioni principali sono:

- una riscrittura radicale delle prime fasi dello sviluppo embrionale nei mammiferi e della trattazione sulle cellule staminali, con una spiegazione dettagliata delle origini del differenziamento cellulare e un esteso approfondimento sulle cellule staminali pluripotenti, sulle cellule staminali tissutali e sulla riprogrammazione cellulare;

- un capitolo specifico sulla manipolazione delle cellule dei mammiferi, che riunisce alcuni contenuti sparsi in diversi capitoli dell'edizione precedente e che delinea l'evoluzione dell'editing del genoma a partire dall'uso della semplice ricombinazione omologa fino all'enfasi moderna sulle nucleasi programmabili;
- un capitolo che si snoda tra l'architettura del genoma umano, il progetto ENCODE e altre nuove iniziative per comprendere le funzioni del nostro genoma;
- un capitolo che fornisce una visione unificante della regolazione genica e dell'epigenetica;
- un capitolo che fornisce una panoramica delle varianti genetiche, dalle loro origini nella sequenza del DNA, ai meccanismi di riparazione del DNA, alle classi di varianti, alla genomica di popolazione, fino alle varianti genetiche funzionali;
- un nuovo capitolo sulla genetica umana di popolazione, un argomento che riteniamo non fosse adeguatamente approfondito nell'edizione precedente;
- un capitolo specifico sulla patologia molecolare, che mette insieme ed espande alcuni materiali già presenti;
- un'ampia trattazione dei successi e dei limiti degli studi di associazione sull'intero genoma (GWAS, *Genome-Wide Association Studies*) nell'identificare i fattori di suscettibilità per le condizioni complesse comuni. Mentre l'era dei GWAS apre la strada agli approcci basati sul sequenziamento su larga scala, ci sembra il momento appropriato per farne un'analisi critica;
- una nuova trattazione della diagnostica del DNA che riflette i cambiamenti più importanti introdotti con l'uso di routine del sequenziamento dell'intero esoma e dell'intero genoma:
- una revisione della genetica del cancro e della genomica che riflette gli sviluppi delle analisi multipiattaforma, l'uso delle biopsie liquide e lo sviluppo di trattamenti mirati;
- un nuovo capitolo che mette insieme gli organismi modello e la modellizzazione delle malattie, compreso il nuovo settore, in rapida crescita, della modellizzazione cellulare tra cui gli organoidi, modelli biologici resi possibili dagli avanzamenti della ricerca di base nella biologia dello sviluppo.

Oltre a questi argomenti specifici, ogni pagina del libro è stata rivista e aggiornata per fornire una panoramica della genetica molecolare umana aggiornata allo stato attuale della ricerca.

Questo libro è stato possibile solo grazie al lavoro di Joanna Koster e della squadra di Taylor & Francis da lei diretta, che ha convertito i nostri testi e i nostri bozzetti nel prodotto finito, tra cui Paul Bennet, Jordan Wearing, Matt McClements, Ruth Maxwell, Becky Hainz-Baxter, e probabilmente altri che hanno lavorato di volta in volta sul progetto. Come sempre, siamo molto grati alle nostre mogli, Meryl e Gilly, per il loro supporto durante tutta la lunga gestazione di questo libro.

TOM STRACHAN
ANDREW P. READ

La mappatura e l'identificazione di geni che causano malattie monogeniche

Di per sé, la sequenza del genoma non fornisce informazioni sull'eventuale tipo di fenotipo che una variante potrebbe causare, o su quale variante potrebbe essere responsabile di una malattia. I ricercatori devono fare questo collegamento e, per farlo, attualmente usano principalmente due strategie. L'approccio più usato è il *sequenziamento dell'esoma*, ossia della totalità degli esoni codificanti tutte le proteine, o dell'intero genoma dei pazienti. Questo approccio può generare una lista di circa 20 000 varianti nell'esoma o di 4 milioni nell'intero genoma. Poi, in qualche modo, la lista dev'essere ristretta fino a trovare la variante causativa. Questo processo è descritto nei ►Paragrafi 17.3-17.5.

In passato, prima dell'avvento del sequenziamento di nuova generazione, trovare le varianti era molto più difficile. La strategia allora era definire, cercando di avvicinarsi il più possibile, la localizzazione cromosomica della variante causativa sconosciuta, in modo da minimizzare la quantità di sequenziamenti di Sanger necessari per identificare le varianti. Questo processo è conosciuto come **clonaggio posizionale** (*positional cloning*). Talvolta è possibile applicare questo metodo, cercando un paziente che ha la malattia in questione ed è anche portatore di una delezione o di un riarrangiamento cromosomico. Questa associazione può essere assolutamente casuale, una pura coincidenza, oppure l'anomalia cromosomica può effettivamente influenzare l'espressione del gene causativo in quel particolare paziente. In genere questa evenienza può risultare particolarmente plausibile se ci si trova di fronte a una condizione trasmessa come tratto autosomico dominante comparso *de novo* in un paziente che è contemporaneamente portatore di un'anomalia cromosomica *de novo*. Queste osservazioni suggeriscono che molto probabilmente il gene responsabile della condizione si potrebbe trovare nella regione coinvolta nell'anomalia cromosomica. Il passo successivo è cercare se sono presenti delle varianti in quella posizione in altri pazienti che hanno la stessa malattia, in assenza però di un'anomalia cromosomica visibile.

Alcuni casi fortunati furono davvero importanti agli albori delle ricerche per l'identificazione dei geni con il clonaggio posizionale, ma furono davvero casi eccezionali. Molto più frequentemente, la posizione si dovette trovare attraverso l'*analisi di linkage*, usando un gruppo di famiglie in cui più persone erano affette dalla malattia. Questo tipo di studi oggi è meno rilevante per la ricerca genetica, non perché sia obsoleto dal punto di vista tecnico, ma perché quasi tutte le malattie monogeniche in cui si potevano trovare famiglie adatte per queste ricerche sono già state studiate e il gene causativo è stato identificato. Esistono numerose famiglie con più persone colpite da condizioni non mendeliane, come il diabete o la schizofrenia, che sono ancora oggi studiate, ma con un approccio differente, che sarà descritto nel ►Capitolo 18. L'analisi di linkage è frequentemente usata anche come approccio nella *mappatura* (in gergo tecnico detta anche *mappaggio*) per *autozigosità*, descritta nel ►Paragrafo 17.2. Nel ►Paragrafo 17.1 descriveremo i principi generali dell'analisi di linkage e del clonaggio posizionale, perché l'analisi di linkage è un aspetto importante del ragionamento genetico che chi impara oggi la genetica dovrebbe capire, almeno nei termini generali. Inoltre, come spiegato da Ott e colleghi (2015) (*vedi PMID 25824869; Suggestimenti di lettura*), il linkage ha ancora un

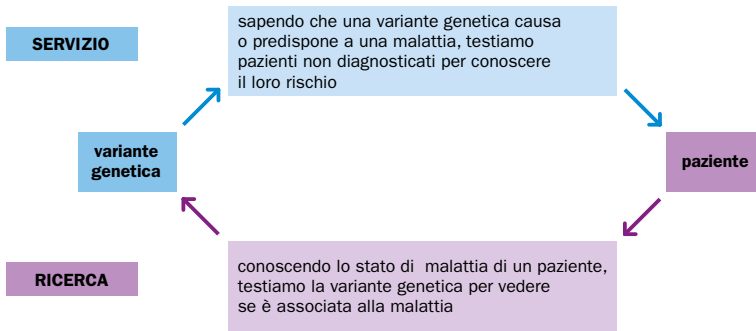


Figura 17.1 L'uso dei test genetici per la clinica e per la ricerca. La figura riassume la differenza che si incontra quando il test esamina una sola variante. Nell'era del sequenziamento dell'esoma, la distinzione è diventata meno chiara, facendo sorgere importanti problemi per il consenso e la generazione di un referto.

posto nel mondo del sequenziamento dell'intero genoma umano, perché può essere uno strumento potente per creare delle priorità tra i genomi da sequenziare e per ridurre la lista delle varianti che devono essere studiate.

Questo capitolo tratta la ricerca per identificare le varianti genetiche sottostanti le condizioni monogeniche. In passato quest'attività era chiaramente distinta dalle analisi genetiche condotte a scopo diagnostico o a scopo predittivo, come riassunto nella **Figura 17.1**. Adesso che il sequenziamento dell'esoma è usato non solo nel campo della ricerca, ma anche nel campo della diagnostica clinica, c'è una sostanziale sovrapposizione tra i due settori. In entrambi i casi bisogna essere chiari sul consenso che il paziente o il soggetto della ricerca hanno dato in relazione a ciò che si può studiare, in particolare devono essere chiari i termini su ciò che sarà restituito nel referto o come risultato. Questi temi saranno considerati nel ►Capitolo 20, quando tratteremo i test e lo screening per le varianti genetiche.

17.1 Il clonaggio posizionale inizialmente cerca di identificare i geni responsabili di una malattia mappandoli in una precisa posizione cromosomica

Durante gli anni '80 e '90 del Novecento, i geni responsabili per la maggior parte delle malattie monogeniche più frequenti furono identificati attraverso il clonaggio posizionale. La **Figura 17.2** mostra il principio alla base di questa tecnica. Nell'analisi di linkage, si testa un pannello (*panel*) di varianti note, dette *marcatori genetici*, distribuite in tutto il genoma. L'obiettivo è trovare un marcatore genetico che, in modo affidabile, possa essere usato per tracciare la variante sconosciuta che causa la malattia in una collezione di alberi genealogici. Se gli eventi di ricombinazione non separano mai, o raramente, la malattia dal marcatore, possiamo supporre che la variante che causa la malattia e il marcatore si trovino vicini nella stessa posizione cromosomica.

La ricombinazione è un evento normale in tutte le divisioni meiotiche. Durante la profase della I divisione meiotica, le coppie dei cromosomi omologhi in sinapsi e i singoli cromatidi si scambiano dei segmenti con il crossing over (►Figure 2.14-2.16). In prima approssimazione, i crossing over sono distribuiti in maniera casuale nel genoma, sebbene, come mostrato nel ►Capitolo 12 (►Figura 12.5), a livello molecolare siano concentrati in 30 000 hotspot.

■ I ricombinanti sono identificati genotipizzando i genitori e la progenie per coppie di loci

La ricombinazione di solito si osserva analizzando la progenie piuttosto che i gameti, sebbene alcuni ricercatori abbiano tipizzato i singoli spermatozoi mediante PCR ultrasensibile. Ne consegue che nella mappatura genetica umana sia normale parlare di una *persona* come un **ricombinante** o un **non ricombinante**. In questo caso è sottinteso che in realtà noi stiamo parlando di uno dei gameti parentali che hanno generato quella persona. Se esiste una qualsiasi ambiguità a proposito di

Figura 17.2 L'identificazione dei geni associati a malattia attraverso il clonaggio posizionale. Il metodo dipende dalla capacità di raccogliere un numero sufficiente di famiglie con molti casi che possano permettere di condurre con successo l'analisi di linkage. L'eterogeneità dei loci (geni differenti che causano la condizione in famiglie differenti) o schemi di ereditarietà irregolare rappresentano l'ostacolo maggiore all'ottenimento di un buon risultato. Il gene causativo è identificato attraverso la dimostrazione della presenza di mutazioni in quel gene in un pannello di individui ammalati non collegati tra loro.



quale genitore è il gamete coinvolto, questo deve essere specificato. La proporzione di gameti che sono ricombinanti per due loci è la **frazione di ricombinazione** tra i due loci.

Nella **Figura 17.3** l'individuo II₁ è eterozigote per i due loci, A e B. Il suo genotipo è A₁A₂B₁B₂. I suoi alleli A₁ e B₁ derivano dalla madre e A₂ e B₂ dal padre. Qualsiasi spermatozoo che porta la combinazione A₁B₁ o A₂B₂ è un non ricombinante per quei due loci, mentre qualsiasi spermatozoo che porta la combinazione A₁B₂ o A₂B₁ è un ricombinante. Come si può osservare, dei suoi sette figli, due sono stati generati da spermatozoi ricombinanti e cinque da spermatozoi non ricombinanti. La frazione di ricombinazione è 0,28.

La frazione di ricombinazione è una misura della distanza genetica tra i due loci

Se due loci sono su cromosomi differenti, si distribuiscono in modo indipendente nei gameti. In media, un 50% dei gameti è ricombinante e un 50% non è ricombinante per ciascuna coppia di loci che si trovano su cromosomi differenti. La frazione di ricombinazione attesa è 0,5. Quando due loci sono **sintenici**, cioè si trovano sullo stesso cromosoma, essi si trasmettono sempre insieme, a meno che non siano separati durante la ricombinazione. Solo un crossing over avvenuto nella regione tra i due loci li può separare. Se due loci si trovano molto vicini sul cromosoma, i ricombinanti sono molto pochi, mentre i loci che si trovano molto distanti tra loro ricombinano liberamente, considerato che i crossing over avvengono su ogni braccio di cromosoma. Ne consegue che la frazione di ricombinazione tra due loci è la misura della loro distanza sul cromosoma. Questa è la **distanza genetica**. È misurata in *centimorgan* (cM, la lettera maiuscola M è in onore del pioniere della genetica T.H. Morgan) ed è distinta dalla distanza fisica che si misura in kilobasi (kb) o megabasi (Mb) di DNA. Le coppie di loci che ricombinano l'1% delle volte si dice che distano 1 cM. La distanza fisica e la distanza genetica non corrispondono precisamente. L'ordine dei loci su una mappa fisica e su una mappa genetica dovrebbe essere lo stesso, ma ci sono *punti caldi di ricombinazione* (*hot-spot*), in cui una piccola distanza fisica si trasforma in un'ampia distanza genetica e viceversa.

La frazione di ricombinazione non eccede mai il valore di 0,5. Se due loci sono molto lontani sul cromosoma ci possono essere due o più crossing over che li separano, ma, come mostra la **Figura 17.4**, non più del 50% dei gameti prodotti sarà ricombinante per quei due loci. Se ci sono 10 loci, A, B, C, ..., J in quest'ordine lungo

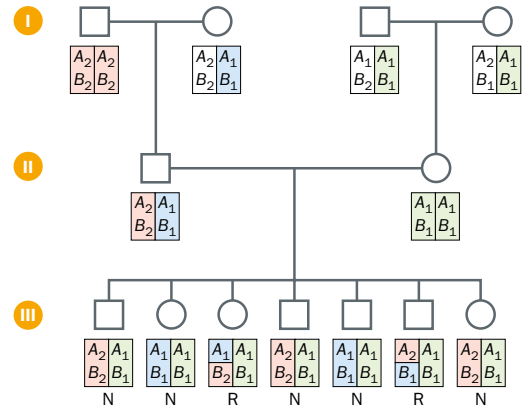


Figura 17.3 I ricombinanti e i non ricombinanti.

In questa famiglia ci sono due loci (A e B) in cui segregano, rispettivamente, gli alleli A₁ e A₂ e B₁ e B₂. I rettangoli colorati indicano le combinazioni degli alleli che possono essere seguite attraverso le diverse generazioni nell'albero genealogico della famiglia. Nella generazione III, possiamo distinguere le persone che hanno ricevuto dal loro padre, II₁, lo spermatozoo non ricombinante (N; A₁B₁ o A₂B₂) o ricombinante (R; A₁B₂ o A₂B₁). La loro madre, II₂, è omozigote in questi due loci e quindi non possiamo identificare gli individui che, nella generazione III, si sono sviluppati da oociti ricombinanti o non ricombinanti.

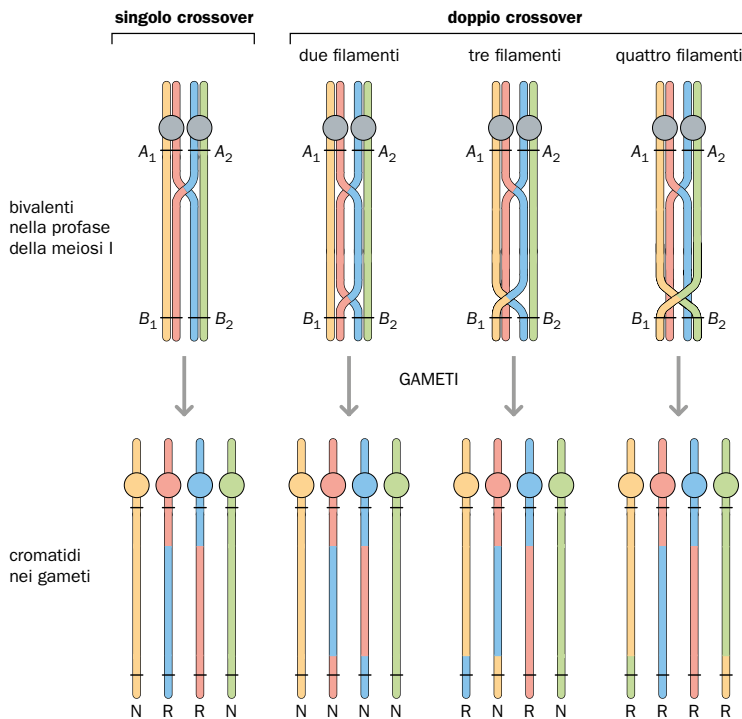


Figura 17.4 I crossing over singoli e doppi.

La figura mostra una coppia di cromosomi omologhi, uno con gli alleli A₁ e B₁, l'altro con gli alleli A₂ e B₂, in due loci. Ciascun cromosoma è composto da due cromatidi fratelli. Solo due dei quattro cromatidi sono coinvolti in uno dei qualsiasi crossing over. I cromatidi nei gameti marcati con N portano una combinazione degli alleli parentali (A₁B₁ o A₂B₂). I cromatidi nei gameti marcati con R portano una combinazione ricombinante degli alleli parentali (A₁B₂ o A₂B₁). Da notare che non ricombinanti o ricombinanti sono definiti solo in relazione a questi due loci. Per esempio, nel risultato del doppio crossing over a tre filamenti, il secondo cromatidio da sinistra è stato coinvolto nel crossing over, ma non è ricombinante per i loci A e B, perché ha gli alleli A₁ e B₁, che rappresentano una combinazione parentale. Un singolo crossing over genera due cromatidi ricombinanti e due cromatidi non ricombinanti (50% di ricombinanti). I tre tipi di doppio crossing over avvengono in percentuale variabile, pertanto come risultato medio un doppio crossing over darà il 50% di ricombinanti.

il cromosoma e ciascuna coppia è a 10 cM di distanza, i *loci* *A* e *J* continueranno a mostrare non più del 50% di ricombinazione. Se prepariamo una mappa genetica che mostra i 10 *loci*, la distanza complessiva tra *A* e *J* dovrebbe essere di 100 cM, ma se di fatto stabilissimo la distanza tra *A* e *J* direttamente nell'albero genealogico, questa sarebbe di 50 cM. Ne consegue che le distanze genetiche non sono additive, ma sono in relazione tra loro sulla base di una **funzione di mappa**. I lettori interessati all'argomento possono approfondire con il libro di Ott e colleghi (*vedi Suggestioni di lettura*).

■ Il riconoscimento dei ricombinanti negli alberi genealogici umani non è sempre immediato

È facile riconoscere che l'individuo II₁ nella ►Figura 17.3 ha due figli ricombinanti e cinque figli non ricombinanti, ma non è sempre così semplice. Solo gli alberi genealogici con una struttura "ideale", cioè con tre o più generazioni, e di cui sono disponibili per tutti dei campioni appropriati e le informazioni cliniche, permettono un'interpretazione diretta. A ogni modo, i ricercatori che tentano di mappare nuovi geni associati a malattie rare possono solo studiare le famiglie così come le trovano. Consideriamo le tre famiglie nella Figura 17.5. In questi casi supponiamo che i due *loci* siano quello che determina una malattia rara autosomica dominante e un marcatore microsatellite che ha sei alleli *A*₁-*A*₆.

- Nella famiglia A l'interpretazione è immediata e segue la logica della ►Figura 17.3.
- Nella famiglia B la donna doppio eterozigote II₁ è di **fase ignota**. Tra i suoi sei figli, cinque possono essere non ricombinanti e uno ricombinante, oppure cinque possono essere ricombinanti e uno non ricombinante. In questa famiglia non possiamo identificare i ricombinanti in modo certo, anche se la prima possibilità sembra molto più probabile della seconda.
- Nella famiglia C si aggiungono altre informazioni. Sono stati identificati altri parenti portatori dell'allele *A*₁ ricevuto dal loro padre, II₃, ma non possiamo essere sicuri che l'allele *A*₁ sia identico per discendenza e sia lo stesso allele *A*₁ della loro zia II₁ (►Figura 12.10 per la distinzione tra gli alleli identici per stato e quelli identici per discendenza).

Sono necessari dei metodi appropriati per estrarre informazioni utili per l'analisi di linkage da famiglie come queste, in cui l'identificazione dei ricombinanti o dei non ricombinanti è ambigua o incompleta. Per affrontare questo problema, si possono usare *punteggi (lod score)* generati con l'analisi al computer, come descritto più avanti.

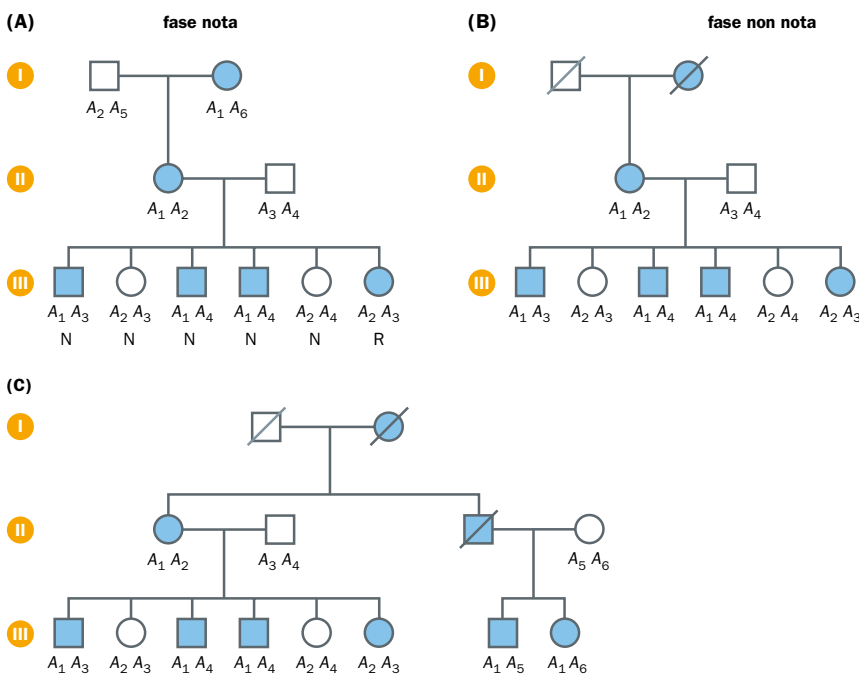


Figura 17.5 Il riconoscimento dei ricombinanti.

Le tre versioni di una famiglia con una malattia autosomica dominante tipizzata con un marcatore microsatellite con alleli *A*₁-*A*₆. **(A)** Tutte le meiosi sono in fase nota. Possiamo identificare in modo non ambiguo i soggetti III₁-III₅ come non ricombinanti (N) e III₆ come ricombinante (R). **(B)** La stessa famiglia, ma con fase non nota. La madre II₁ potrebbe aver ereditato con la malattia sia l'allele *A*₁, sia l'allele *A*₂; quindi la sua fase non è nota. III₁-III₅ possono essere non ricombinanti e III₆ ricombinante oppure III₁-III₅ possono essere ricombinanti e III₆ non ricombinante. **(C)** La stessa famiglia dopo aver rintracciato altri parenti. Anche III₇ e III₈ hanno ereditato l'allele *A*₁ con la malattia dal padre, ma non possiamo essere sicuri che l'allele *A*₁ del padre sia identico per discendenza a quello di sua sorella II₁. Potrebbero esserci due copie dell'allele *A*₁ tra i quattro alleli marcatori dei nonni. La probabilità che questo avvenga dipende dalla frequenza dell'allele *A*₁. Quindi, nonostante l'albero genealogico ricostruito contenga informazioni aggiuntive sul linkage se comparato all'albero genealogico B, l'estrazione di queste informazioni è difficile.

■ La mappatura delle malattie umane fa affidamento su gruppi di marcatori genetici rappresentativi del genoma

La mappatura genetica nella specie umana si basa in linea di massima sugli stessi principi della mappatura nel moscerino *Drosophila* o in qualsiasi altro organismo diploide che si riproduce sessualmente. Tuttavia ci sono due differenze pratiche. Innanzitutto, come abbiamo appena visto, dobbiamo fare affidamento su alberi genealogici la cui struttura non ideale rende spesso difficile l'interpretazione, invece che basarci su esperimenti di inbreeding chiari, puliti. E, in seconda istanza, mentre in *Drosophila* possiamo mappare una variante del colore degli occhi verso una variante della forma delle ali, negli esseri umani non possiamo mappare una malattia verso altre malattie. Le malattie monogeniche umane sono rare e le famiglie in cui due di queste malattie segregano sono doppiamente rare. Anche se ne potessimo trovare una, sarebbe probabilmente troppo piccola per riuscire a generare un numero sufficiente di dati.

Per mappare una malattia umana, o un *locus* che determina qualsiasi altro fenotipo poco comune, abbiamo bisogno di raccogliere famiglie in cui segrega il carattere d'interesse e poi trovare qualche altro carattere mendeliano che segreghi nelle famiglie che si vogliono studiare, in modo che un soggetto possa essere individuato come ricombinante o non ricombinante. Un **marcatore genetico** adatto allo scopo deve avere le seguenti caratteristiche:

- deve mostrare uno schema chiaro di ereditarietà mendeliana, preferibilmente codominante, cosicché il genotipo possa sempre essere dedotto dal fenotipo;
- deve essere ottenuto facilmente e a basso costo usando materiale prontamente disponibile (per esempio un prelievo di saliva, piuttosto che una biopsia del cervello);
- il *locus* che determina il carattere dovrebbe essere altamente polimorfico, in modo tale che sia elevata la probabilità che una persona selezionata casualmente sia eterozigote;
- dovrebbe appartenere a un gruppo di centinaia di marcatori dello stesso tipo, situati in posizioni note sui cromosomi e distribuiti in tutto il genoma.

I primi tentativi di mappatura negli umani usarono le varianti delle proteine come gli isotipi tissutali o i gruppi sanguigni. Un lavoro fondamentale di Botstein e colleghi nel 1980 (vedi PMID 6247908; *Suggerimenti di lettura*) faceva emergere il potenziale delle varianti del DNA per la mappatura dei geni umani. Le varianti originali descritte nel lavoro erano i *polimorfismi dei frammenti di restrizione* (RFLP, *Restriction Fragment Length Polymorphism*). Si tratta di variazioni nelle dimensioni dei frammenti prodotti dalla digestione del DNA di una persona usando un enzima di restrizione, conseguenti a varianti geniche che inseriscono o aboliscono dei siti di restrizione, ► Figura 7.4A. Dalla metà degli anni '90 del Novecento in poi, per l'analisi di linkage nella specie umana, sono stati usati i microsatelliti amplificati per PCR in multiplex, usando primer marcati con fluorofori, in modo che possano essere analizzati mediante sequenziatori automatici con elettroforesi capillare. In tempi più recenti, per genotipizzare ciascun membro di una famiglia sono stati usati array di *polimorfismi a singolo nucleotide* (SNP) che permettono di analizzare 500 000 SNP in una singola operazione. Si potrebbero anche usare i dati di sequenziamento dell'intero genoma.

I dati grezzi sono interpretati da programmi informatici che generano un lod score Una volta che abbiamo raccolto le famiglie in cui una malattia mendeliana segrega e le abbiamo genotipizzate con il marcatore adatto, come facciamo a sapere se abbiamo trovato un linkage? Per rispondere, si devono valutare due aspetti:

- come possiamo ricavare la frazione di ricombinazione?
- che test statistico dovremmo usare per vedere se la frazione di ricombinazione è significativamente differente da 0,5, il valore atteso sull'ipotesi nulla dell'assenza di linkage?

Nell'albero genealogico mostrato nella **Figura 17.5C**, non è possibile identificare in modo non ambiguo i ricombinanti e contarli. È comunque possibile calcolare la probabilità complessiva che i *loci* siano in linkage (frazione di ricombinazione pari a zero, $\theta = 0$) o che non siano in linkage (frazione di ricombinazione pari a 0,5). Il rapporto tra questi due valori fornisce la *probabilità di linkage (odds)* e il logaritmo della probabilità è il **lod score**. I lod score sono rappresentati dal simbolo **Z**.

Focus 17.1

Calcolo dei lod score (Z)

Si calcola la verosimiglianza complessiva di un pedigree sulla base di due ipotesi:

- posto che due *loci* siano davvero in linkage, con frazione di ricombinazione θ , la probabilità di una meiosi di essere ricombinante è θ e la probabilità di essere non ricombinante è $1 - \theta$;
- se i due *loci* sono, di fatto, non in linkage, la probabilità di una meiosi di essere ricombinante o non ricombinante è 0,5.

Per la famiglia nella ► Figura 17.5A ci sono 5 non ricombinanti ($1 - \theta$) e un ricombinante θ .

- La probabilità complessiva che siano in linkage è $(1 - \theta)^5 \times \theta$.
- La probabilità complessiva che non siano in linkage è $(0,5)^6$.
- Il rapporto di verosimiglianza è $(1 - \theta)^5 \times \theta / (0,5)^6$.
- Il lod score è il logaritmo in base 10 di questo rapporto della verosimiglianza.

θ	0	0,1	0,2	0,3	0,4	0,5
Z	$-\infty$	0,577	0,623	0,509	0,299	0

Per la famiglia nella ► Figura 17.5B la fase nella madre, II₁, è ignota. Se lei ha ereditato l'allele A₁ con la malattia, ci sono cinque non

ricombinanti e un ricombinante. Se lei ha ereditato l'allele A₂ con la malattia, ci sono cinque ricombinanti e un non ricombinante.

Il rapporto di verosimiglianza complessivo è

$$\frac{1}{2} \left[(1 - \theta)^5 \cdot \frac{\theta}{(0,5)^6} \right] + \frac{1}{2} \left[\theta^5 \cdot \frac{1 - \theta}{(0,5)^6} \right]$$

Questo permette di considerare entrambe le possibili fasi con una uguale probabilità a priori.

θ	0	0,1	0,2	0,3	0,4	0,5
Z	$-\infty$	0,276	0,323	0,222	0,076	0

Per la famiglia nella ► Figura 17.5C, per calcolare la probabilità che III₇ e III₈ siano ricombinanti o non ricombinanti, dobbiamo considerare le probabilità per ciascun genotipo possibile di I₁, I₂ e II₃, pesata per la probabilità per quel genotipo. Per I₁ e I₂, le probabilità del genotipo osservato dipendono sia dalle frequenze geniche sia dai genotipi osservati in II₁, III₇ e III₈. Per II₃ le probabilità del genotipo sono calcolate sulla base delle semplici regole mendeliane. Tutto questo è molto complicato, quasi impossibile da calcolare a mano. Per qualsiasi famiglia, tranne che per le più semplici, i non masochisti si affidano ai programmi informatici appositamente progettati.

Nel 1955 Newton Morton dimostrò che i lod score sono il parametro statistico più efficace per valutare il linkage nei pedigree e derivò da varie strutture standard di alberi genealogici delle formule per generare il lod score in funzione della frazione di ricombinazione θ . Il Focus 17.1 mostra come si applica questa analisi ad alberi genealogici con strutture semplici come quelle delle ► Figure 17.5A, B. Tranne che in casi così semplici, l'analisi di linkage nella specie umana dipende interamente da programmi computazionali; questi applicano degli algoritmi in grado di analizzare le ramificazioni degli alberi genealogici in termini di probabilità dei genotipi, avendo a disposizione i dati sul pedigree e una tabella delle frequenze geniche. Ott e colleghi (2015) (vedi PMID 25824869; *Suggerimenti di lettura*), descrivono i vari programmi di linkage che possono essere utilizzati e forniscono gli indirizzi Web attraverso cui è possibile accedere.

Poiché è una funzione della frazione di ricombinazione, il lod score si calcola per una gamma di valori di θ . I risultati possono essere rappresentati graficamente per ottenere le curve di lod score in relazione alla frazione di ricombinazione. La frazione di ricombinazione più probabile è quella in cui il lod score è più alto. Se non ci sono ricombinanti, il lod score sarà massimo per $\theta = 0$. Se ci sono dei ricombinanti, Z avrà un picco alla frazione di ricombinazione più probabile ($0,167 = 1/6$ per la famiglia A nella ► Figura 17.5, ma sarà più difficile da predire per la famiglia B senza calcoli più dettagliati).

In un gruppo di famiglie, la probabilità complessiva di linkage è data dal prodotto delle probabilità per ogni singola famiglia. I lod score, essendo logaritmi, possono essere sommati attraverso le famiglie. Quindi il risultato dell'analisi di linkage è una tabella di lod score corrispondenti alle varie frazioni di ricombinazione, con gli score sommati attraverso un'intera collezione di famiglie e con righe separate per una serie intera di marcatori. I lod score positivi forniscono prove in favore dell'esistenza di linkage, i lod score negativi costituiscono prove contro l'esistenza di linkage. Da notare che solo le frazioni di ricombinazione tra 0 e 0,5 hanno qualche significato e che tutti i lod score sono zero per $\theta = 0,5$, perché in questo caso misurano il rapporto tra due probabilità identiche e il $\log_{10}(1) = 0$.

I lod score di +3 e -2 sono i criteri per assegnare il linkage e per escluderlo (per un singolo test) La seconda domanda che abbiamo posto riguarda la *soglia della significatività* statistica. A questo punto la risposta è, a prima vista, sorprendente: $Z = 3$ è la soglia per accettare il linkage, con una probabilità di errore di tipo 1 del 5% (rigettando erroneamente l'ipotesi nulla). In molte analisi statistiche, $p < 0,05$ è usato come il valore soglia della significatività statistica, ma $Z = 3,0$ corrisponde alla probabilità di 1000 a 1 (1000:1) odds [$\log_{10}(1000) = 3,0$]. La scelta di una soglia così stringente si basa sull'improbabilità intrinseca che due *loci*, presi a caso,

che dimostrino perché mutazioni in quel gene dovrebbero produrre il fenotipo in questione. Questo secondo stadio è sempre importante, indipendentemente da come un gene candidato è stato identificato, ma è particolarmente importante per i geni che sono stati identificati attraverso la ricerca di varianti *de novo* in pazienti con malattie che non presentano caratteristiche cliniche distintive. Le linee di prova possono essere acquisite usando insieme di dati esistenti e generando nuovi dati.

■ **Gli insiemi di dati esistenti possono fornire una grande quantità d'informazioni su un gene candidato**

Non serve nulla di più di un computer abilitato all'accesso a Internet per poter consultare una notevole quantità d'informazioni che possono aiutare a giudicare se un gene è un candidato promettente per la malattia che si sta studiando. Gli aspetti su cui si può sperare di trovare informazioni includono:

- se il gene è già stato implicato in malattie o fenotipi;
- se sono state descritte varianti strutturali comprendenti il gene e, in caso positivo, se erano state trovate in soggetti sani o in persone con qualche condizione patologica;
- quali varianti sono state riportate per quel gene e con quale frequenza;
- in quale tessuto del corpo e in quale periodo durante lo sviluppo il gene è espresso;
- quali isoforme di splicing sono state descritte per quel gene;
- se ci sono informazioni sulla probabile funzione biochimica del prodotto genico;
- se sono note le altre proteine con cui il prodotto genico interagisce;
- se esistono informazioni sui meccanismi di controllo dell'espressione del gene;
- quali sono gli ortologhi e paraloghi del gene conosciuti;
- qual è l'estensione della conservazione della sequenza tra gli ortologhi e i paraloghi, con particolare attenzione alla conservazione degli amminoacidi coinvolti nelle varianti candidate per la malattia;
- l'eventuale disponibilità di dati sugli effetti di mutazioni, knockdown o iper-espressione degli ortologhi in altre specie.

La presenza di un precedente, la rarità e la conservazione sono probabilmente gli elementi iniziali più utili per identificare le varianti che causano malattie mendeliane.

- La banca dati ClinVar (www.ncbi.nlm.nih.gov/clinvar) raccoglie informazioni su possibili varianti patogenetiche. I dati sono raccolti dai singoli laboratori o sono estratti da altre banche dati. Nel luglio 2018 conteneva 676 579 risultati. Nella ClinVar è possibile cercare una specifica variante, oppure tutte le varianti segnalate per un singolo gene, o, ancora, fare una ricerca per varianti in una data posizione genomica o per varianti associate a una malattia o a un fenotipo. Così come sono elencate le varianti e qualsiasi fenotipo dichiarato associato, ClinVar include una valutazione della probabile patogenicità e i link a pubblicazioni rilevanti e a molte altre caratteristiche disponibili nella banca dati. Quando la stessa variante è stata riportata da più di un soggetto, ClinVar mette le due informazioni fianco a fianco, in modo che gli utilizzatori possano controllare che le informazioni siano concordanti o rivederle per interpretazioni alternative.
- Il Genome Aggregation Database (gnomad.broadinstitute.org), il successore dell'ampiamente usato ExAC Database, può essere impiegato per controllare la frequenza di una variante patogenetica putativa nei controlli sani. GnomAD contiene così tanti esomi (nell'ottobre 2017, 126 216 esomi più 15 136 sequenze *whole genome*) e così tante varianti rare che sarebbe irrealistico pensare che una variante patogenetica candidata sia completamente assente nella banca dati. La rarità dovrebbe essere definita in base alla frequenza stimata della malattia. Una variante candidata per una condizione dominante, se presente in GnomAD, dovrebbe essere presente a una frequenza molto inferiore rispetto alla frequenza della malattia nella popolazione. Una variante candidata per una condizione recessiva dovrebbe essere presente a una frequenza molto inferiore rispetto alla frequenza dei portatori, come stimato dalla relazione di Hardy-Weinberg (con un margine per l'endogamia in una rara condizione recessiva; ►Paragrafo 12.4).
- In passato si sarebbero controllate poche centinaia d'individui sani della popolazione generale per essere sicuri che la variante non fosse un comune polimorfismo non patogenetico. Un calcolo presentato nel ►Paragrafo 5.3 può essere usato per stabilire un limite superiore alla frequenza di una variante nella popolazione, se un pannello di controlli è stato esaminato senza trovare nessuna

persona sana portatrice della variante. In ogni caso, a meno che uno non stia studiando una popolazione molto inusuale, questo approccio ora è stato completamente superato dall'uso delle banche dati ExAC e GnomAD. Il limite principale di queste risorse immensamente potenti è che i soggetti inseriti non sono necessariamente privi di malattie. Per esempio, le banche dati includono dati generati da coorti di studi sulla schizofrenia. Ci si potrebbe chiedere realisticamente che cosa significhi un soggetto "sano". Nessuno di noi è perfetto. Il risultato di GnomAD può comparire in una schermata sola che riguarda le malattie pediatriche gravi, perché molti di quei casi sono stati esclusi dalla banca dati, ma l'elenco potrebbe richiedere un filtraggio se si stanno cercando informazioni su malattie a esordio tardivo o condizioni meno severe.

- La conservazione è la base dei programmi di PolyPhen e SIFT che sono usati per stabilire la patogenicità di varianti missenso, così com'è alla base di un certo numero di altri programmi (► Focus 16.2). Programmi alternativi per lo studio a livello dei nucleotidi includono PhyloP, disponibile come una delle funzioni visualizzabili nel browser UCSC, e GERP o GERP++ (*Genomic Evolutionary Rate Profiling*). CADD (*Combined Annotation-Dependent Depletion*) e VEP (*Variant Effect Predictor*, disponibile attraverso Ensembl) integrano analisi di molti tipi di data.

I risultati ottenuti dall'uso di questi strumenti dovrebbero essere trattati unicamente come indicativi. Un motivo per cui le predizioni basate sulla conservazione danno falsi negativi, marcando una variante dannosa come benigna, è la **deviazione patogenica compensata** (*compensated pathogenic deviation*). La **Figura 17.14** mostra degli esempi. Gli studi funzionali potrebbero confermare che una certa variante amminoacidica è patogenetica nel contesto della proteina umana. Tuttavia, in certi animali quella stessa variante è normale e rappresenta l'amminoacido wild type nella proteina ortologa. A causa di questo, i programmi come PolyPhen e SIFT la classificano come benigna, ma è benigna solo nel contesto di proteine che contengono particolari cambiamenti amminoacidici in altre posizioni. Le catene polipeptidiche si ripiegano attraverso interazioni tra differenti residui amminoacidici e spesso funzionano attraverso quelle interazioni, così, in principio, non c'è nulla di sorprendente in questo. Lo studio di Jordan e colleghi (*vedi* PMID 26123021; *Suggerimenti di lettura*) ha stimato che almeno il 3% delle sostituzioni amminoacidiche umane sono soggette a questo effetto in qualche altra specie. La ► Figura 12.9 mostrava un fenomeno simile in un RNA non codificante.

Le informazioni circa l'espressione e le interazioni di un prodotto genico possono essere di grande aiuto. La cosa più ovvia è che un gene sia espresso nelle cellule, nei tessuti e negli organi colpiti dalla malattia. Quindi, nello studio di Vissers sulla disabilità intellettiva, descritto prima, due geni sono stati scartati, nonostante fosse stato confermato che si trattava di due mutazioni *de novo*, perché non erano espressi nel sistema nervoso centrale. Inoltre avevano funzioni note che facevano ritenere improbabile che potessero avere un ruolo nella disabilità intellettiva. Anche se le specifiche funzioni di un prodotto genico non sono ben descritte, le informazioni sulle interazioni genetiche possono essere informative. Le sue interazioni possono porre il gene in reti o in vie molecolari che danno qualche idea sul tipo di fenotipo che la variante potrebbe produrre.

■ Le varianti rilevanti potrebbero essere create e studiate in colture cellulari o nell'animale *in toto*

Oltre a questi impieghi dei dati esistenti, si possono rendere necessari degli esperimenti *ad hoc* per esplorare gli effetti di una variante. Come descritto nel ► Paragrafo 16.1, si possono usare degli esperimenti di trasfezione transiente per controllare l'effetto di una variante sulla regione del promotore oppure un saggio di splicing con un minigene può testare l'effetto sullo splicing (► Figura 16.3). Nei casi in cui la proteina ha un ruolo biochimico osservabile nelle cellule, il funzionamento della proteina wild type o della variante può essere confrontato in cellule in coltura.

	BBS4	N165	H366		RPGRIP1L	P189	F193	R937	R961
<i>H. sapiens</i>		N	H	<i>H. sapiens</i>		P	F	R	R
<i>O. cuniculus</i>		H	R	<i>E. caballus</i>		L	L	L	T
<i>O. princeps</i>		H	R	<i>L. africana</i>		L	L	L	T
<i>E. telfairi</i>		H	R	<i>P. capensis</i>		L	L	L	T
<i>T. nigroviridis</i>		H	T	<i>T. manatus</i>		L	L	L	T

Figura 17.14 La deviazione patogenica compensata.

I saggi funzionali confermarono che due sostituzioni amminoacidiche, p.N165H nella proteina BBS4 e p.R937L nella proteina RPGRIP1L, sono patogenetiche anche se l'amminoacido sostituito corrisponde all'amminoacido wild type normale nelle altre specie. In quegli animali sostituzioni amminoacidiche in altre porzioni della proteina compensano l'effetto patogenetico: R366 o T366 in BBS4; L189, L193 o T961 in RPGRIP1L. I programmi come PolyPhen e SIFT erroneamente classificano p.N165H e p.R937L come benigne. (Adattata da Jordan DM *et al.* [2015] *Nature* 524: 225-229; PMID 26123021. Con l'autorizzazione di Springer Nature. Copyright © 2015.)

CRISPR/Cas o altre tecnologie per l'editing dei geni (►Paragrafo 8.4) permettono di creare specifiche varianti in una cellula o si possono usare cellule staminali pluripotenti indotte da un paziente con la variante e differenziarle, se necessario, nell'appropriato tipo cellulare.

Se queste strategie basate sulle cellule non forniscono una risposta chiara, può essere necessario ricreare la variante in un organismo modello e osservare il suo effetto. Vi sono vari organismi modello adatti a questo scopo. Questi includono il topo, *zebrafish* e *Drosophila*. Come esempio, nel **Focus 17.3** si mostra come *zebra-*

Focus 17.3

L'uso di *zebrafish* per l'analisi funzionale delle varianti associate a ciliopatie

Le ciliopatie sono malattie causate alle disfunzioni delle ciglia. Le ciglia sono presenti in quasi tutte le cellule dei mammiferi e hanno una varietà di funzioni, non solo per il movimento dei fluidi extracellulari, ma anche nei meccanismi di segnalazione (rassegna di Gerdes e colleghi, vedi PMID 19345185; *Suggerimenti di lettura*). Nel funzionamento delle ciglia possono essere coinvolte fino a 1000 proteine e ci sono molte ciliopatie mendeliane. Una di queste è la *sindrome di Bardet-Biedl* (BBS; OMIM #209900). La BBS è una malattia autosomica recessiva caratterizzata da retinite pigmentosa, obesità, disfunzione renale, polidattilia, disordini del comportamento e ipogonadismo. Nei pazienti con BBS è stata osservata la perdita di funzione in omozigosi di uno qualsiasi dei 20 geni coinvolti. Ci sono state delle controversie sul fatto se la BBS sia triallelica, con varianti che contribuiscono alla patogenicità in più di un *locus*, attraverso un carico complessivo della ridotta funzione ciliare, o se sia una semplice malattia recessiva e le varianti aggiuntive trovate in altri *loci* secondari siano soltanto incidentali e irrilevanti. Da qui è nato l'interesse per lo sviluppo di test funzionali per studiare le varianti patogenetiche.

Di seguito descriviamo un sistema che usa lo *zebrafish* sviluppato dal gruppo di Nicholas Katsanis della Duke University (Figura A). Il test si sviluppa in tre stadi.

- Innanzitutto si usa un oligonucleotide morfolino antisense (►Paragrafo 8.5) per ridurre l'espressione del gene ortologo in un embrione di *zebrafish* wild type di 1-4 cellule. Il test si basa sul fatto

che il knockdown possa causare difetti di gastrulazione che possono includere assi corporei accorciati, somiti più lunghi e notocorde ampie e difettose. Questi fenotipi sono indicativi di anomalie della segnalazione della *polarità cellulare planare* (PCP, *Planar Cell Polarity*) che molto probabilmente è alla base di molti fenotipi clinici osservati nei pazienti BBS. Ammesso che il knockdown produca questi effetti, è possibile procedere al passaggio successivo.

- Successivamente, negli embrioni wild type sono coiniettati gli oligonucleotidi morfolino e l'mRNA wild type del gene umano d'interesse. Ammesso che questo trattamento riporti al fenotipo normale, è possibile procedere al passaggio successivo.
- Infine, negli embrioni wild type sono coiniettati gli oligonucleotidi di morfolino e l'mRNA in cui è presente la variante in studio. Se con questo trattamento il fenotipo si mantiene normale la variante è considerata non patogenetica, ma se il difetto nella gastrulazione rimane, significa che la variante non ha la funzione dell'mRNA wild type.

In alcuni casi l'iniezione del solo mRNA mutante, senza l'oligonucleotide morfolino, causa rilevanti anomalie. Questo suggerisce un effetto dominante. Nei casi in cui l'effetto può essere indotto dalla coiniezione dell'mRNA wild type, si considera che la variante abbia un effetto dominante negativo; se l'mRNA wild type non riesce a compensare, la variante ha un effetto di acquisizione di funzione patogenetica.

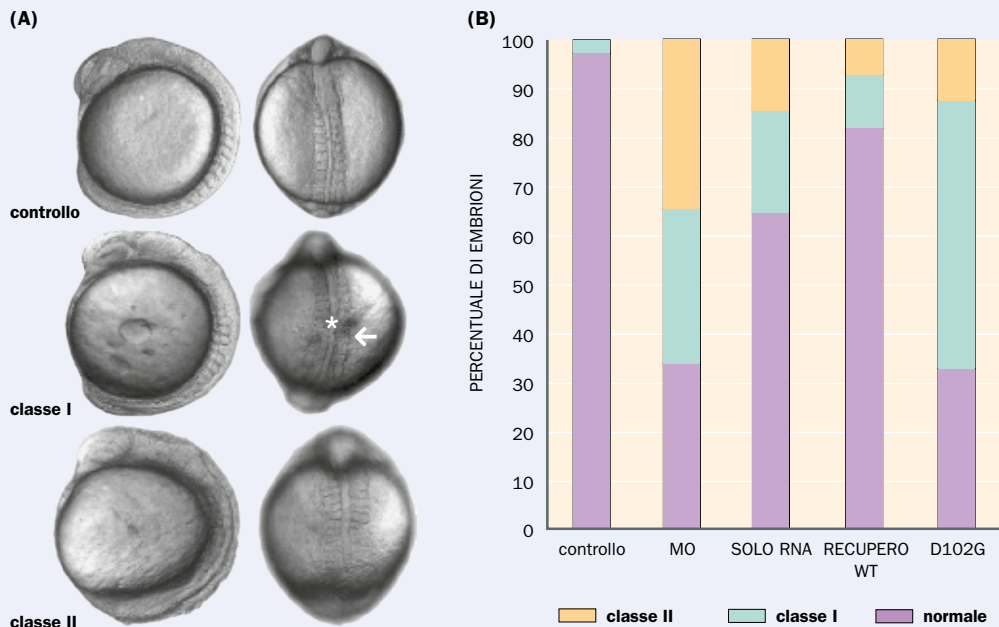


Figura A Un test funzionale per studiare le varianti trovate nei pazienti con la sindrome di Bardet-Biedl. (A) Immagini di embrioni di *zebrafish* con 9 somiti che mostrano uno sviluppo normale e anomalie di classe I e classe II. L'asterisco mostra una notocorda deformata, la freccia segna i somiti allungati. (B) L'iniezione di un oligonucleotide morfolino (MO) che blocca *BBS4* causa embrioni

anormali con grande frequenza (seconda barra). L'effetto può essere ampiamente recuperato coiniettando l'mRNA umano wild type (WT) *BBS4* (quarta barra), ma non con l'mRNA portatore della mutazione p.D102G (quinta barra). L'RNA p.D102G quindi non possiede la funzione dell'mRNA wild type. (Cortesia di Erica E Davies, Duke University.)

fish sia stato usato per investigare se le varianti trovate nei pazienti con la *sindrome di Bardet-Biedl* fossero patogenetiche. I dati mostrati nella ►Figura 17.14 sono stati ottenuti usando questo sistema.

Uno degli organismi modello più ampiamente usati è il topo. È più facile che i topi mostrino un fenotipo direttamente correlato alla corrispondente mutazione umana. Tuttavia, nonostante le numerose similarità tra topi e umani, non ci si deve aspettare una corrispondenza perfetta tra i fenotipi mutanti degli ortologi nelle due specie. Non è raro che una mutazione patogenetica umana riprodotta nel topo non determini alcun effetto fenotipico o che sia letale. Per esempio, in un progetto di ricerca furono introdotte nel topo mutazioni knockout in 37 geni per i quali l'OMIM riportava fenotipi umani recessivi. Diciassette knockout murini risultarono letali. I genetisti del topo sono ben consapevoli che i fenotipi dipendono molto dal background genetico. Una certa variante è spesso espressa diversamente nei diversi ceppi di topo di laboratorio (►Focus 21.4). Le massicce differenze nel background tra topi ed esseri umani hanno quindi effetti molto più grandi. Negli esseri umani gli effetti delle differenze nel background genetico si mostrano come penetranza ed espressività variabile.

La generazione sistematica e la fenotipizzazione di un mutante nullo per ogni gene murino è in corso grazie all'International Mouse Phenotyping Consortium (IMPC; www.mousephenotype.org). I modelli knockout possono avere effetti fenotipici più estremi delle mutazioni puntiformi che causano perdita di funzione, in parte perché queste ultime non sempre causano una perdita completa di funzione. Così i modelli knockout nel topo sono utili indicatori per la totalità dei possibili effetti fenotipici dovuti alla perdita di funzione, almeno nei topi, mentre solo un sottogruppo può essere valutato nei topi omozigoti per mutazioni puntiformi. Per le varianti che determinano l'acquisto di funzione, generalmente è necessario costruire un topo mutante portatore dell'esatta mutazione corrispondente. In altri tempi questo era un compito lungo e impegnativo; alcuni dei nuovi strumenti per l'editing del genoma, in particolare la tecnologia di editing CRISPR/Cas (►Paragrafo 8.4) ha reso questo compito molto più semplice.

RIASSUNTO

- La mappatura genetica è usata per identificare la posizione cromosomica della variante responsabile di un fenotipo monogenico e generare una lista ristretta di geni candidati.
- Nella mappatura tradizionale basata sul lod score, si studia, in una raccolta di famiglie, la co-segregazione del fenotipo rilevante con gli alleli dei marcatori genetici. I principali marcatori usati sono i microsatelliti, brevi sequenze ripetute in tandem (STRP, *Short Tandem Repeat Polymorphism*) o polimorfismi a singolo nucleotide (SNP, *Single Nucleotide Polymorphism*). Il criterio statistico per la significatività è il lod score.
- Oggi la mappatura tradizionale basata sul lod score è usata raramente come primo approccio per identificare un gene, perché quasi tutti i fenotipi mendeliani per cui sono disponibili ampie famiglie sono già stati mappati. Questo tipo di strategia può ancora aiutare a ridurre la lunga lista di varianti identificate dal sequenziamento dell'esoma o del genoma.
- La mappatura tradizionale basata sul lod score ha avuto molto successo nel mappare le condizioni monogeniche, ma la mappatura di *loci* che conferiscono suscettibilità a comuni malattie complesse richiede, per l'analisi, approcci differenti che sono descritti nel ►Capitolo 18.
- La mappatura per autozigosità può essere usata per malattie recessive. Gli SNP sono usati per identificare segmenti cromosomici che sono in omozigosi e identici per discendenza in una serie di persone affette da una o più famiglie di consanguinei.
- Il sequenziamento dell'intero genoma o dell'intero esoma è ampiamente usato per la ricerca di geni mutati in individui con condizioni sporadiche o molto rare.
- Le condizioni sporadiche sono spesso il risultato di mutazioni dominanti insorte *de novo* e queste possono essere trovate nel sequenziamento dell'esoma dell'individuo ammalato e di entrambi i suoi genitori.
- I dati grezzi prodotti dagli strumenti del sequenziamento di nuova generazione richiedono un ampio e dettagliato processamento bioinformatico, prima per generare una lista di varianti veramente presenti nel campione analizzato e poi per filtrare la lunga lista di tali varianti per identificare una lista ristretta dei candidati per la condizione in studio.
- Le varianti identificate dal sequenziamento in individui affetti devono essere valutate per la loro presunta patogenicità e per la loro rilevanza sulla genesi del fenotipo.
- Per le varianti di sequenze codificanti, i loro presunti effetti sulla funzione genica possono essere predetti da vari programmi computazionali, per esempio PolyPhen-2 e SIFT a livello delle proteine e VEP o programmi di predizione dei siti di splicing a livello del DNA.
- Una variante candidata non dovrebbe essere presente nei soggetti sani con una frequenza troppo alta. Per una condizione recessiva questa dovrebbe avere la frequenza dei portatori predetti; per una condizione dominante dovrebbe essere, idealmente, zero, se la malattia ha un'alta penetranza, ma in ogni caso sostanzialmente molto più bassa della frequenza della condizione nella popolazione.
- L'attinenza di un gene candidato per un fenotipo può essere stabilita usando le conoscenze sulla sua funzione biochimica, sui tempi e sulla localizzazione della sua espressione e sulle osservazioni fenotipiche ottenute analizzando persone o animali che hanno mutazioni in geni correlati.
- Una valutazione definitiva della patogenicità richiede studi funzionali in estratti cellulari, cellule intatte o modelli animali.

SUGGERIMENTI DI LETTURA

Linkage e mappatura

- Botstein D *et al.* (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* **32**:314–331; PMID 6247908.
- Ott J (1999) *Analysis of Human Genetic Linkage*, 3rd edn. Johns Hopkins University Press.
- Ott J *et al.* (2015) Genetic linkage analysis in the age of whole-genome sequencing. *Nat Rev Genet* **16**:275–284; PMID 25824869.

Mappatura per autozigosità

- Houwen RHJ *et al.* (1994) Genome screening by searching for shared segments: mapping a gene for benign recurrent intrahepatic cholestasis. *Nat Genet* **8**:380–386; PMID 7894490.
- Varon R *et al.* (1998) Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* **93**:467–476; PMID 9590180.

Identificazione dei geni con il sequenziamento dell'esoma

- DePristo MA *et al.* (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**:491–498; PMID 21478889.
- Gilissen C *et al.* (2014) Genome sequencing identifies major causes of severe intellectual disability. *Nature* **511**:344–347; PMID 24896178.
- Hoischen A *et al.* (2010) *De novo* mutations of *SETBP1* cause Schinzel-Giedion syndrome. *Nat Genet* **42**:483–485; PMID 20436468.
- Lelieveld SH *et al.* (2015) Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum Mutat* **36**:815–822; PMID 25973577.
- Ng SB *et al.* (2010) Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**:30–35; PMID 19915526.
- Ng SB *et al.* (2010) Exome sequencing identifies *MLL2* mutations as a cause of Kabuki syndrome. *Nat Genet* **42**:790–793; PMID 20711175.

- Visser LELM *et al.* (2016) Genetic studies in intellectual disability and related disorders. *Nat Rev Genet* **17**:9–18; PMID 26503795.

Valutazione della conservazione e patogenicità di una variante

- Adzhubei IA *et al.* (2010) A method and server for predicting damaging missense mutations. *Nat Methods* **7**:248–249; PMID 20354512. (PolyPhen-2; genetics.bwh.harvard.edu/pph2)
- Cooper GM *et al.* (2005) Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res* **15**:901–913; PMID 15965027. (GERP)
- Gerdes JM *et al.* (2009) The vertebrate primary cilium in development, homeostasis, and disease. *Cell* **137**:39–45; PMID 19345185.
- Jordan DM *et al.* (2015) Identification of *cis*-suppression of human disease mutations by comparative genomics. *Nature* **524**:225–229; PMID 26123021.
- Kircher M *et al.* (2014) A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet* **46**:310–315; PMID 24487276. (CADD)
- Kumar P *et al.* (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* **4**:1073–1082; PMID 19561590. (SIFT; sift.jcvi.org/)
- Siepel A *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**:1034–1050; PMID 16024819. (PhyloP)

Test funzionali di patogenicità

- White JK *et al.* (2013) Genome-wide generation and systematic phenotyping of knockout mice reveals new roles for many genes. *Cell* **154**:452–464; PMID 23870131.
- Zaghloul NA *et al.* (2010) Functional analyses of variants reveal a significant role for dominant negative and common alleles in oligogenic Bardet-Biedl syndrome. *Proc Natl Acad Sci USA* **107**:10602–10607; PMID 20498079.

Tom Strachan, Andrew Read

Genetica molecolare umana

Seconda edizione Zanichelli condotta sulla quinta edizione inglese

A cura di Rossella Tupler

I progressi nell'ambito della genetica molecolare umana sono stati moltissimi nell'ultimo decennio, ma non ci sono dubbi su quale sia lo sviluppo più coinvolto nei cambiamenti: il sequenziamento massivo del DNA.

Le applicazioni delle tecnologie di sequenziamento di nuova generazione (NGS, *Next Generation Sequencing*), che si sono espanse in ogni area della disciplina, sono trattate qui in modo ampio e aggiornato, includendo anche il campo nuovo ed entusiasmante degli studi di genomica su singola cellula e le innovazioni nelle tecniche per la manipolazione dei geni.

Le tecnologie meno recenti sono state sintetizzate e raggruppate, ma non escluse dalla trattazione; per quanto quasi del tutto superate nei laboratori, sono didatticamente molto utili, come nel caso dello studio dei cariotipi: il microscopio infatti aiuta a capire cosa avviene nel nucleo di una cellula in modo più chiaro rispetto all'analisi informatica. Fin dalla prima edizione, lo scopo di *Genetica molecolare umana* è proprio fornire i fondamenti della disciplina, non stilare una lista di informazioni, per le quali è più proficuo attingere direttamente alle fonti online.

Anche la genomica funzionale, le cellule staminali e i metodi per creare modelli di malattia trovano in questa edizione largo spazio, così come maggior attenzione è riservata all'ereditarietà, alle varianti del DNA – inserite nel contesto delle popolazioni – e al ruolo dell'epigenetica nel controllo dell'espressione genica. Alle cellule staminali, in particolare, è dedicata una spiegazione dettagliata delle origini del differenziamento cellulare e un esteso approfondimento sulle cellule staminali pluripotenti, sulle cellule staminali tissutali e sulla riprogrammazione cellulare. Inoltre, gli studi di associazione *genome-wide* (GWAS), che aprono la strada agli approcci basati sul sequenziamento su larga scala, sono maturi per un'analisi critica, che ne percorre i successi e i limiti.

Infine, i grandi progressi compiuti in pochi anni nell'analisi del DNA contemporaneo e di quello antico, e le nuove informazioni sulle origini e sulla storia dell'essere umano che se ne ricavano, sono oggetto di un nuovo capitolo sulla genetica evolutiva, «La genomica comparata e l'evoluzione del genoma».

Tom Strachan è professore emerito di Genetica molecolare umana presso la Newcastle University, UK. Ha scritto, con Judit Goodship e Patrick Chinnery, *Genetica & Genomica nelle scienze mediche*, tradotto da Zanichelli nel 2016.

Andrew Read è professore emerito di Genetica umana presso la University of Manchester, UK.

Nel 2007 Tom Strachan e Andrew Read hanno ricevuto lo European Society of Human Genetics Education Award.

Le risorse multimediali



online.universita.zanichelli.it/strachanze

A questo indirizzo sono disponibili le risorse multimediali di complemento al libro. Per accedere alle risorse protette è necessario registrarsi su **my.zanichelli.it** inserendo il codice di attivazione personale contenuto nel libro.

Libro con ebook



Chi acquista il libro può scaricare gratuitamente l'**ebook**, seguendo le istruzioni presenti nel sito. L'ebook si legge con l'applicazione *Booktab Z*, che si scarica gratis da App Store (sistemi operativi Apple) o da Google Play (sistemi operativi Android).

STRACHAN*GENETICA MOL UMANA 2ED LUM

ISBN 978-88-08-52030-2



9 788808 520302

2 3 4 5 6 7 8 9 0 (60H)