

Indice generale

Prefazione

XIII

PARTE

A

Nozioni fondamentali

CAPITOLO 1

La biologia essenziale

3

- 1.1 Genomi, genomica e avvento della Bioinformatica 3
- 1.2 Genoma dei procarioti 5
 - 1.2.1 Struttura e dimensioni 5
 - 1.2.2 Proprietà composizionali 6
 - 1.2.3 RNA non codificanti 7
 - 1.2.4 Elementi mobili e sequenze ripetute 8
- 1.3 Genoma degli eucarioti 8
 - 1.3.1 Struttura e dimensioni 9
 - 1.3.2 Proprietà composizionali 12
 - BOX 1.1 Isocore nei genomi di vertebrati* 13
 - 1.3.3 Corredo genico 14
 - BOX 1.2 Definizione operativa di gene eucariotico* 16
 - 1.3.4 Genoma umano 17
- 1.4 Genomi degli organelli: mitocondri e cloroplasti 21
 - 1.4.1 Mitocondri 22
 - 1.4.2 Cloroplasti 24
- 1.5 Genomi virali 24

CAPITOLO 2

L'informatica essenziale

26

- 2.1 Architettura dei calcolatori 26
- 2.2 Sistemi operativi 27
- 2.3 Algoritmi e programmi 29
- 2.4 Linguaggi di programmazione 31
- 2.5 Server e web server 33
- 2.6 Basi di dati 33
 - 2.6.1 Definizioni 33
 - 2.6.2 DB relazionali 33
 - 2.6.3 Progettazione di un database 34

2.6.4 Accesso ai dati	35
2.6.5 Qualità di un database	35
BOX 2.1 Operatori booleani	36

CAPITOLO 3

La statistica essenziale	37
3.1 Dal campione alla stima dei parametri	37
3.1.1 Campionamento	37
3.1.2 Distribuzioni e stima	37
3.1.3 Stima non-parametrica attraverso metodi di ricampionamento	39
<i>Metodo bootstrap</i>	40
<i>Metodo jackknife</i>	40
3.2 Inferenza statistica	40
3.2.1 Sistema d'ipotesi	40
3.2.2 Errore α e β	41
3.2.3 Livello di significatività: il <i>p-value</i>	42
3.2.4 Test parametrici	42
3.2.5 Test non-parametrici	42
BOX 3.1 Distribuzione Gaussiana	42
<i>Test di permutazione</i>	43
3.3 Statistica multivariata e metodi di raggruppamento	44
BOX 3.2 Vicinanza	44
3.3.1 Metodi gerarchici	45
3.3.2 Metodi non-gerarchici	45
3.3.3 Validità dei gruppi	46
BOX 3.3 Alcuni software per l'analisi di dati	47

PARTE

B

Allineamenti

CAPITOLO 4

L'evoluzione biologica	51
4.1 Evoluzione molecolare	51
4.2 Meccanismi molecolari alla base dei processi evolutivi	52
4.3 Geni ortologi e paraloghi	55

CAPITOLO 5

Allineamenti tra sequenze	57
5.1 Confrontare sequenze biologiche	57
5.2 Allineamento globale di coppie di sequenze	59
5.3 Programmazione dinamica	60

5.4	Matrici di sostituzione	63
5.5	Allineamento locale di coppie di sequenze	67
5.6	Ricerche per similarità in banca dati	68
5.7	BLAST	70
5.7.1	Parametri di input e output	73
5.7.2	Significatività degli allineamenti di sequenze	73
5.7.3	Interpretazione dei risultati	75
5.8	Allineamento di sequenze a genomi	76
5.9	Allineamento multiplo di sequenze	79

CAPITOLO 6

	Alberi filogenetici	83
6.1	Cosa sono gli alberi filogenetici	83
6.2	Metodi per la costruzione di alberi filogenetici	85
6.2.1	Metodi basati su <i>clustering</i> : UPGMA e <i>neighbor-joining</i>	85
	UPGMA	85
	<i>Neighbor-joining</i>	86
6.2.2	Massima parsimonia	87
6.2.3	Metodi di massima verosimiglianza	90
6.3	Test statistici per valutare il grado di affidabilità delle ricostruzioni filogenetiche	91

PARTE

C

Genomi

CAPITOLO 7

	Piattaforme di sequenziamento degli acidi nucleici	95
7.1	Sequenziamento degli acidi nucleici, dal Sanger al NGS	97
7.2	Tecniche di sequenziamento NGS	98
7.2.1	Sequenziatori NGS di seconda generazione	99
	<i>Roche 454</i>	99
	<i>ThermoFisher SOLiD</i>	101
	<i>ThermoFisher Ion Torrent/Proton</i>	101
	<i>Illumina</i>	103
7.2.2	Sequenziatori NGS di terza generazione	104
	<i>Pacific Bioscience</i>	105
	<i>Oxford Nanopore</i>	106
7.3	Formato dei dati genomici	107
7.4	<i>Paired-end vs single-end</i>	111
	<i>BOX 7.1 Formato SAM/BAM per l'allineamento di sequenze su un genoma</i>	111
7.5	Valutazione della qualità delle read	112

CAPITOLO 8

Ricostruzione e annotazione di genomi	113
8.1 Come si ricostruisce la sequenza di un genoma?	113
8.1.1 Algoritmi di assemblaggio <i>Overlap-Layout-Consensus</i> (OLC)	114
8.1.2 Algoritmi di assemblaggio per sequenziamenti di nuova generazione	118
8.1.3 Valutazione della qualità di un assemblaggio	119
8.2 Annotazione del genoma	121
8.2.1 Genomi procariotici	121
8.2.2 Annotazione di genomi eucariotici	122
8.2.3 Predizione di geni <i>ab initio</i>	123
8.2.4 Modelli Markoviani	124
BOX 8.1 Banca dati RefSeq	124
8.2.5 Annotazione basata su sequenziamento NGS dell'RNA	126
BOX 8.2 Formato dei file per annotazioni genomiche	126
8.3 Elementi regolatori nel genoma	127
8.4 Genomica di popolazioni	129
BOX 8.3 SNP array	130
8.4.1 Identificazione delle varianti genomiche nella popolazione	131
BOX 8.4 Formato VCF	131
BOX 8.5 Teorema di Bayes	132
8.4.2 Associazione tra presenza di variante e fenotipo clinico	133
8.4.3 Identificazione di varianti rare	133
8.5 Identificazione di grandi riarrangiamenti e variazioni strutturali	133
8.6 Banche dati genomiche	134
8.7 Metagenomica e DNA <i>barcoding</i>	135

CAPITOLO 9

Applicazioni della genomica	138
9.1 Diagnosi e predisposizioni	138
BOX 9.1 Conseguenze delle mutazioni	139
9.1.1 Varianti somatiche	140
9.2 Analisi prenatali	141
9.3 Test di paternità e genomica forense	142

PARTE

D

Trascrittomi

CAPITOLO 10

L'analisi del trascrittoma	147
10.1 Annotazione di geni e trascritti alternativi	150
10.2 Approcci sperimentali per l'analisi del trascrittoma	153

10.2.1 DNA microarray	153
10.2.2 Tecniche basate sul sequenziamento: RNA-Seq	155
10.3 Analisi di dati RNA-Seq	158
10.4 Analisi dei livelli di espressione genica	160
10.4.1 <i>Gene clustering</i>	160
10.4.2 Identificazione di geni differenzialmente espressi	160
10.5 Funzione dei geni differenzialmente espressi	162
<i>BOX 10.1 Correzione per la molteplicità dei test</i>	163
<i>BOX 10.2 Gene Ontology</i>	164
10.6 RNA editing	164

CAPITOLO 11

La struttura dell'RNA	166
11.1 Componenti della struttura secondaria	166
11.2 Metodi per la predizione della struttura secondaria	167
11.2.1 Ricerca su larga scala di motivi di RNA funzionali	173

PARTE

E

Proteomi

CAPITOLO 12

Proteine e proteomi	177
12.1 Annotazione funzionale di proteine	177
12.2 Interazioni proteiche	179
12.3 Banche dati	179
12.3.1 Sequenze proteiche: UniProt	179
12.3.2 Motivi funzionali proteici: PROSITE ed ELM	181
12.3.3 Strutture proteiche: PDB e PDBe	182
12.3.4 Interazioni proteiche: IntAct, MINT e STRING	183
12.3.5 Altre banche dati	184

CAPITOLO 13

Metodi bioinformatici per l'analisi di sequenze proteiche	186
13.1 Descrittori di elementi funzionali nelle sequenze proteiche	186
13.1.1 Espressioni regolari	187
<i>Definizione di espressione regolare</i>	187
<i>Associazione tra espressione regolare e motivo funzionale</i>	188
<i>Ricerca di un'espressione regolare in una sequenza</i>	189
<i>Valutazione dell'affidabilità dell'associazione funzionale</i>	190
<i>BOX 13.1 Misure della performance di un algoritmo predittivo</i>	191

13.1.2	Profili di sequenza	191
	<i>Definizione di un profilo di sequenza</i>	191
	<i>Ricerca di una PSSM in una sequenza</i>	195
13.1.3	Modelli nascosti di Markov (<i>Hidden Markov Models</i>)	196
	<i>Catene di Markov</i>	196
	<i>Modelli Markoviani nascosti</i>	197
	<i>HMM per descrivere una famiglia proteica o un motivo</i>	198
	<i>Costruzione di un motivo</i>	198

CAPITOLO 14

Strutture proteiche 201

14.1	Informazione strutturale	201
14.1.1	Livelli di descrizione delle strutture proteiche	201
14.1.2	Metodi sperimentali per la determinazione della struttura	202
	<i>Diffrazione ai raggi X</i>	202
	<i>Spettroscopia a risonanza magnetica nucleare</i>	203
	<i>Validazione delle strutture</i>	204
14.2	Predizione della struttura delle proteine	205
14.2.1	Predizione della struttura secondaria delle proteine	206
	BOX 14.1 Metodi di apprendimento automatico	207
	<i>Reti neurali artificiali</i>	207
	<i>Metodi per la predizione della struttura secondaria</i>	210
14.2.3	Predizione della struttura terziaria delle proteine	210
	<i>Modeling comparativo o homology modeling</i>	210
	BOX 14.2 Root-Mean-Square Deviation	212
	<i>Threading</i>	213
	<i>Metodi ab initio</i>	214
14.3	Metodi per l'analisi delle strutture proteiche	215
14.3.1	Ordine e disordine	215
14.3.2	Confronti di strutture	215
14.3.3	Metodi strutturali per l'annotazione funzionale	217

CAPITOLO 15

Interazioni proteiche 218

15.1	Metodi per l'analisi delle interazioni proteiche	218
15.1.1	Inferenza per omologia	218
15.1.2	Mutazioni correlate	219
15.1.3	Profili filogenetici	220
15.1.4	Metodo della stele di Rosetta per le interazioni proteiche	221
15.1.5	Similarità di alberi filogenetici	222
15.1.6	Prossimità genomica	222
15.1.7	Espressione correlata	223
15.1.8	<i>Docking</i>	223
15.2	Interazione tra proteine e ligandi	224
15.2.1	Descrizione e confronti di ligandi	224

15.2.2	Progettazione razionale di farmaci	226
15.2.3	Banche dati di ligandi	226
15.3	Analisi delle interazioni	227
	<i>BOX 15.3 Grafi</i>	228

PARTE

F

Grandi progetti e prospettive future

CAPITOLO 16

	La Bioinformatica: tra presente e futuro	233
16.1	Grandi progetti	233
16.1.1	Progetto 1000 Genomi	233
16.1.2	Progetti ENCODE ed epigenomici	234
16.1.3	Progetto GTEx	234
16.1.4	Progetto <i>The Cancer Genome Atlas</i> (TCGA) e suoi sviluppi	236
16.1.5	Progetto <i>Human Microbiome</i>	238
16.2	Competizioni scientifiche	240
16.2.1	CASP-CAPRI	241
16.2.2	CAFA	242
16.2.3	DREAM – <i>Dialog for Reverse Engineering Assessment and Methods</i>	242
16.3	Grandi risorse pubbliche: Ensembl e UCSC	243
16.4	Sfide dell'era omica	245
16.4.1	Gestione dei dati e infrastrutture di ricerca internazionali per la Bioinformatica	245
16.4.2	Integrazione dei dati e biologia dei sistemi	246
16.4.3	Comprensione dei risultati	247
16.5	Altre applicazioni	247
16.5.1	Ambito agro-alimentare	248
16.5.2	Antropologia molecolare e studio delle popolazioni	248
16.5.3	Tossicologia ambientale	249
	Indice analitico	251