

Programma del corso integrato di Bioinformatica con laboratorio

(moduli didattici BIO/11 e ING-INF/05)

Anno Accademico 2014 – 2015

Lezioni dettate dai Proff. Elena Amendola e Luigi Cerulo

Elementi di programmazione in ambiente R

- Nozioni di base sul linguaggio R
- concetto di algoritmo
- variabili e strutture dati
- logica booleana e selezione dati
- istruzione condizionale IF
- istruzioni iterative WHILE e FOR
- concetto di funzione e libreria di funzioni
- struttura di un programma
- input/output da file (formati csv, tsv)
- Introduzione alla libreria R Bioconductor
- Introduzione alla libreria R BioString (formati fasta)
- elementi di output grafico (istruzione plot)

Elementi di genomica

- Metodi per il sequenziamento massivo di DNA ed RNA.
- Banche dati di acidi Nucleici (EMBL, NCBI, DDBJ)
- Banche dati di proteine (PDB, UniProt)
- Il Progetto Genoma Umano
- Organizzazione del genoma
- DNA ripetitivo a basso numero di copie
- Geni codificanti proteine: identificazione ed organizzazione
- Splicing alternativo
- Meccanismi responsabili per l'origine di nuovi geni
- Pseudogeni
- Evoluzione di funzioni geniche
- Geni di RNA non codificanti: piccoli RNA nucleari, miRNA, piRNA, siRNA, lunghi RNA regolatori
- Classi di DNA altamente ripetitivo: eterocromatina e trasposoni
- Organizzazione e funzione delle sequenze regolative
- Variabilità del genoma umano
- Metodi di analisi dell'espressione genica
- Il progetto ENCODE

Bioinformatica

- Database biologici
 - Uso dei navigatori genomici (UCSC Genome, Ensemble, BioMart)
 - Elaborazione in R delle tabelle di annotazione UCSC, Ensemble, BioMart
- Analisi di sequenze
 - Algoritmi di allineamento (Needleman–Wunsch, Smith–Waterman)
 - Matrici di scoring (PAM, BLOSUM)

- Algoritmi per analisi di segnali (matrici PWM)
- Algoritmi di predizione regioni codificanti (Genescan)
- Algoritmi per l'assembling di dati DNA-seq
- Algoritmi per il mapping di dati RNA-seq
- Genomica funzionale
 - Algoritmi di Clustering
 - Analisi di espressione (dati di microarray ed RNA-seq)
 - Gene set enrichment (GO, GSEA, IPA)

Seminari previsti

- Tecnologie di Next Generation Sequencing
- Modelli statistici per l'analisi dei dati RNA-seq
- Metodi e strumenti di genomica comparata

Testi consigliati

(uno ed uno solo per ciascun ambito):

1. Marketa Zvelebil, Jeremy Baum
"Understanding Bioinformatics" (2007, Garland Science, ISBN: 978-0815340249)
2. Richard C. Deonier, Simon Tavaré, Michael S. Waterman
"Computational Genome Analysis: An Introduction" (2005, Springer, ISBN: 978-0387987859)
3. Tom Strachan, Andrew Read
"Genetica Umana Molecolare" (2012, Zanichelli 1° edizione italiana condotta sulla 4° edizione americana)
4. Selezione di recenti pubblicazioni scientifiche