



UNIVERSITÀ DI PISA

ANALISI GENETICHE E GENOMICHE

STEFANO LANDI

Anno accademico

2020/21

CdS

BIOLOGIA MOLECOLARE E
CELLULARE

Codice

176EE

CFU

6

Moduli	Settore/i	Tipo	Ore	Docente/i
ANALISI GENETICHE E GENOMICHE	BIO/18	LEZIONI	48	STEFANO LANDI ROBERTO MARANGONI

Obiettivi di apprendimento

Conoscenze

Il corso di Analisi Genetiche e Genomiche si pone l'obiettivo di formare laureati con una specifica competenza nelle tematiche e nelle metodologie caratterizzanti la Genetica, la Genomica e la Trascrittomica nelle loro applicazioni per la ricerca di base e applicata.

Modalità di verifica delle conoscenze

La verifica delle conoscenze avviene con esame scritto eventualmente (opzionale) seguito da verifica orale (con spostamento della valutazione di + o - 1 punto)

Capacità

Al termine del corso lo studente sarà in grado di imparare un metodo per allineare sequenze derivanti dai sequenziatori di nuova generazione (Next Generation Sequencing) e disegnare alcuni test molecolari di analisi del DNA, nonché di contestualizzare l'utilizzo e l'applicazione di una determinata metodica genomica, genetica e di biologia molecolare, con particolare riferimento al campo della biomedicina. Lo studente apprenderà a usare l'ambiente R e in particolare la suite Bioconductor per l'elaborazione di pipeline di analisi su dati genome-wide.

Modalità di verifica delle capacità

La verifica delle capacità acquisite avverrà in sede di esame tramite domande aperte o esercizi mirati

Comportamenti

Al termine del corso lo studente avrà le basi necessarie (da approfondire) per progettare uno studio di indagine genetica e molecolare su malattie mendeliane e complesse.

Modalità di verifica dei comportamenti

I comportamenti acquisiti verranno verificati in sede di esame.

Prerequisiti (conoscenze iniziali)

Le conoscenze della genetica di base, biologia molecolare e alcune nozioni di statistica sono essenziali. E' utile, ma non indispensabile, la conoscenza dell'ambiente R.

Indicazioni metodologiche

Le lezioni sono tutte di tipo frontale.

Programma (contenuti dell'insegnamento)

Questo programma e' provvisorio alla data del 16/09/2020.

Presentazione del corso. Libro di testo consigliato. Definizione di genetica. Definizione di Genomica. Il genoma mitocondriale.

I principi del sequenziamento di nuova generazione (next generation sequencing. Prof. Landi).

Modalità pratiche con utilizzo del proprio PC per analisi dati NGS con la suite Bioconductor in ambiente R (3 CFU a cura del Prof. Marangoni).



UNIVERSITÀ DI PISA

Prof. Landi:

Approccio generico per studiare sequenze di DNA: ibridazione molecolare e clonaggio del DNA. Tecnica del clonaggio in un sistema cellulare. La PCR. Principio di funzionamento. Descrizione dei singoli reagenti necessari. Visualizzazione del prodotto di PCR. Elettroforesi su gel d'agarosio. Problema delle contaminazioni. Marcatura dei prodotti di PCR: saggi diretti e indiretti. Aggiunta di code al 5'. Applicazioni della PCR. Il disegno dei primers. La Tm dei primers. Hairpin, self-dimers e Dimer. Primers design tools. Dimostrazioni pratiche.

Varianti comuni della PCR. Multiplex PCR. Vantaggi e applicazioni. Nested PCR. Applicazioni. Semi-nested PCR. La PCR asimmetrica. Il metodo diretto del sequenziamento di DNA (reazione di Sanger). PCR con primers allele-specifici (ASO e ARMS-PCR). La PCR inversa. PCR hotstart. Touch-down PCR. Long range PCR. Whole-genome amplification (WGA). DOP-PCR. Adaptor-ligation PCR. Inter-Alu PCR. Phi29 DNA-polymerase (MDA). Primers mutagenetici. Un esempio di mutagenesi sito-specifica con primers mutagenetici: il "QuickChange Lightning Site-Directed Mutagenesis".

La quantificazione dell'espressione genica. Il Northern Blot. Valutazione dell'integrità dell'RNA e calcolo della purezza e concentrazione. RT-PCR. Real Time PCR. SYBER e EVA-GREEN. Calcolo della efficienza di reazione. Valutazione della specificità di reazione. Molecular Beacons e Sonde Taqman. Disegno di primers e sonde per real-time PCR. Utilizzo delle giunzioni esone-esone per il posizionamento dei primers in qPCR. Calcolo della quantità di espressione con metodo qPCR. Utilizzo di geNorm per la selezione dei geni di riferimento. Quantificazione assoluta e relativa (metodo del Delta-Ct e quantity transformation; metodo del Delta-Delta-Ct approssimato).

I numeri notevoli dell'RNA. Micro-array per analisi di trascrittomiche. Principio di funzionamento. Microarray spottati a cDNA, a oligonucleotidi pre-sintetizzati, o a oligonucleotidi sintetizzati in situ. Sintesi di sonde nell'array Affymetrix (metodo di stampa fotolitografico). Tipi di sonde: a cDNA, 65-mers e 25-mers. Metodo competitivo a due colori. Metodi a singolo colore. Analisi dei dati. Estrazione degli spot. Eliminazione degli spot con SNR inferiore a 3. Sottrazione del Background. Normalizzazione. Selezione dei geni differenzialmente espressi. Clustering analysis.

Applicazioni in ricerca e in diagnostica (Mamma Print).

Nanostring: principio di funzionamento e applicazioni in ricerca e in diagnostica

Biopsie tissutali; biopsie liquide e relative applicazioni; Droplet Digital PCR: aspetti tecnici e applicazioni; analisi a single cell level: utilizzo di strumentazioni come FACS sorting e Laser Capture Microdissection. Analisi a single cell level: RNA scope, tecnologia e applicazioni.

Organizzazione del genoma umano. Categorie di RNA non codificanti: rRNA, tRNA, snRNA, snoRNA, RNA TERT, RNA Y, RNasi MRP, miRNA, siRNA, piRNA.

Sequenze altamente ripetute: retrotrasposoni LINE1 e SINE con particolare riferimento alle sequenze Alu. Sequenze intersperse: retrotrasposoni LTR. Trasposoni a DNA e meccanismo di trasposizione. Conseguenze mutageniche della retrotrasposizione. Pseudogeni processati. DNA satellite. DNA satellite come marker di instabilità citogenetica. Minisatelliti della famiglia telomerica. Minisatelliti della famiglia ipervariabile. Metodi di indagine dei minisatelliti. DNA fingerprinting in genetica forense. Microsatelliti. Evoluzione delle sequenze ripetute in tandem (mini e micro satelliti). Slippage-misalignment.

Micro-satellite instability. Metodi di indagine dei microsatelliti. DNA profiling in genetica forense. Micro-inserzioni. Micro-delezioni.

SNPs. Strategie di studio degli SNPs: distinzione tra tecnologie di discovery e tecnologie di genotyping. Strategia di scansione o discovery: metodi conformazionali ed enzimatici. Metodi conformazionali: spiegazione di SSCP, DGGE, dHPLC, HRM tra i metodi conformazionali. CCM e Sanger sequencing tra i metodi enzimatici. Metodi di genotipizzazione: DOT BLOT; RFLP (SOUTHERN BLOT E PCR); ASO PCR; OLA PCR; PIREQUENZIAMENTO; MALDI-TOF; TAQ MAN ALLELIC DISCRIMINATION ASSAY.

Genotipizzazione tramite array: APEX; BEAD ARRAY ILLUMIN (GOLDENGATE ASSAY E INFINIUM ASSAY); GENE CHIP AFFYMETRIX. CNV (duplicazioni; duplicazioni segmentali; delezioni e inversioni); pseudogeni non processati; cenni sulla whole genome duplication; esempi di duplicazioni obbligatorie: i geni delle globine; meccanismi di formazione alla base delle CNV: non-allelic homologous recombination, ligazione non omologa delle estremità, cromotripsis; implicazioni fisiologiche delle CNV: esempi di polimorfismi in geni implicati nel metabolismo dei farmaci; concetti di farmacogenetica e farmacogenomica; implicazioni patologiche delle CNV: esempi di CNV rilevate in disturbi neurologici; metodiche di analisi delle CNV: TaqMan assay, Digital Droplet PCR, MLPA, QF-PCR.

Comparative Genomic Hybridization (Classical CGH). BAC array CGH, tiling BAC array CGH. SNP array per CGH. Le inversioni segmentali. Un polimorfismo di inversione in 17q21.31

Dal fenotipo al gene. I tratti mendeliani. Identificazione della regione candidata tramite analisi di linkage. Meiosi informative e non informative. Misurare il "contenuto di informazione di un marcatore genetico" (PIC) tramite calcolo della eterozigosità. Esempi di alberi genealogici dove capire la "fase gametica" e la prole "ricombinante" e "parentale". Calcolo del LOD SCORE (mappatura a due punti). Da un singolo marcatore a più marcatori. Il multipoint LOD score. La prima mappa con densità elevate dei marcatori umani (CHLC, CEPH). Il multipoint LOD score. La mappatura per i tratti recessivi sfruttando l'autozigosità. Segmenti cromosomici con "Identity by Descent" (IBD) diversi dai segmenti cromosomici con "Identity by State" (IBS). Esempio di calcolo di LOD score in figli di secondi cugini. Esempi dalla letteratura di mappature per autozigosità. Possibilità di restringere la regione candidata sfruttando l'ancestro comune (esempio della Nijmegen Breakage Syndrome).

Clonaggio posizionale, ordinamento dei contigui di cloni e predizione dei loci genici. L'era post-genomica: uso delle banche dati per l'identificazione dei geni nella regione candidata. Geni della regione candidate da sottoporre a screening di mutazione: prioritizzazione sulla base del pattern di espressione, della funzione, delle informazioni provenienti da geni ortologhi e paraloghi. L'importanza dei pazienti con anomalie cromosomiche per identificare un gene-malattia. Mutation screening: sequenziamento degli esoni dei geni prioritizzati nei probandi affetti. Analisi parallela del cDNA. Esempi della fibrosi cistica e dell'emofilia grave (fattore VIII). Verifiche all'ipotesi che le mutazioni riscontrate in un probando siano effettivamente causative della patologia: la corrispondenza del modello genetico con la segregazione della mutazione; mutazioni o polimorfismi? Sequenziamento di probandi di famiglie diverse: tre possibili situazioni (stessa mutazione, stesso gene diversa mutazione, nessuna mutazione). Inferire l'effetto funzionale della mutazione (predizione in silico). Identificazione dei geni-malattie con strategie indipendenti dalla posizione: conoscenza del prodotto proteico e utilizzo di modelli animali. Esempi di patologie il cui gene è stato rinvenuto tramite clonaggio posizionale: Distrofia muscolare di Duchenne e Fibrosi Cistica.

Genetica e genomica del cancro. Tratti distintivi di una cellula perchè possa essere definita tumorale. Il cancro come un processo multi-stadio. Il cancro come malattia genetica. Oncogeni e oncosoppressori: esempi di mutazioni a carico di queste categorie geniche. Eterogeneità genetica: intratumorale, intermetastatica, intrametastatica, interpaziente. Differenza tra mutazioni passeggero e mutazioni drivers nel cancro e pathways maggiormente coinvolti. Numero e tipologie di mutazioni più frequentemente riscontrate in varie tipologie tumorali. Terapie: dalle convenzionali alle personalizzate. L'esempio del test KRAS. Immunoterapia e definizione di neoantigeni in funzione del carico mutazionale tipico di una cellula tumorale. Visione genomica del cancro. Il Cancer Genome Project.

Storia del Progetto Genoma Umano; Sequenziamento gerarchico; Sequenziamento shotgun; Tecnologia NGS: workflow generale; Metodologie di sequenziamento utilizzate in NGS: sequencing by synthesis; pyrosequencing; proton sequencing; sequencing by ligation; Descrizione della



UNIVERSITÀ DI PISA

tecnologia NGS con piattaforma Illumina. Descrizione della tecnologia NGS con piattaforma Ion Torrent. Descrizione della tecnologia NGS con piattaforma 454 Roche. Descrizione della tecnologia NGS con piattaforma Solid. Descrizione della tecnologia NGS con piattaforma Helicos. Tecniche di sequenziamento di terza generazione: SMRT (Pacific Bioscience) e Oxford Nanopore. Analisi bioinformatica dei dati NGS: base calling; allineamento; variant calling; filtering e annotation. Vantaggi e Svantaggi dell'NGS. Applicazioni dell'NGS in genomica, trascrittomica, epigenomica e metagenomica. Exome sequencing e whole genome sequencing: workflow operativo, analisi bioinformatica ed esempi in letteratura con particolare riferimento alle malattie genetiche rare. RNA sequencing: workflow operativo, analisi bioinformatica e relative applicazioni dell'RNA-seq nell'analisi di espressione genica differenziale, identificazioni di varianti di splicing, identificazioni di trascritti di fusione e analisi allele specifica. Esempi in letteratura. NGS e cancro: pannelli di geni utilizzati in diagnostica. Metagenomica: workflow operativo ed esempi pratici.

prof. Marangoni: i 3 crediti sono focalizzati sulle metodiche bioinformatiche di analisi ed elaborazione dei dati di NGS. Lo strumento fondamentale è rappresentato dalla suite Bioconductor che permette di implementare la quasi totalità delle pipeline normalmente impiegate per l'analisi di dati molecolari. I principali punti che verranno coperti sono:

- Introduzione all'ambiente R (elementi fondamentali del linguaggio, tipi di variabili, costrutti)
- L'ambiente "tidy R": caratteristiche e usabilità per le applicazioni bioinformatiche
- La suite Bioconductor: installazione e caratteristiche fondamentali
- Acquisizione di dati NGS
- Un esempio di workflow di mappatura e genotipizzazione
- Esempi di workflow per analisi di dati NGS genome-wide

Bibliografia e materiale didattico

1) Libri di testo consigliati:

"Genetica molecolare umana", by Tom Strachan & Andrew P. Read (Zanichelli)
Powepoint forniti dai docenti

Modalità d'esame

Prova d'esame orale, preceduta da 20 minuti dedicati allo svolgimento di esercizi scritti.

Altri riferimenti web

cercare su e-learning e Moodle

Ultimo aggiornamento 16/09/2020 10:41