Immagine che contiene testo, schermata, numero, linea

Descrizione generata automaticamenteNascita dei genome browsers

Questi sono nati per diverse ragioni, tra cui: l'annotazione dei dati genomici è stata fondamentale per attribuire un significato e una funzione alle sequenze genetiche. Senza un'annotazione adeguata, una sequenza genomica da sola non avrebbe avuto un senso chiaro o comprensibile.

Per affrontare questa sfida, uno dei pionieri e più grandi bioinformatici nella storia della genomica, Jim Kent, ha sviluppato strumenti e metodi per visualizzare graficamente il genoma. Questo ha portato alla creazione di browser genomici, che hanno richiesto anni di lavoro per essere sviluppati. Grazie a questi browser genomici, come Ensembl, UCSC e NCBI, è possibile esplorare le regioni cromosomiche, studiare le regioni di regolazione e confrontare le regioni genomiche tra diverse specie.

È importante sottolineare che il genoma umano assemblato è composto da 3.096.649.726 paia di basi (detto anche "golden path length"), che corrispondono a oltre 20.000 geni codificanti proteine, quasi 24.000 geni non codificanti e oltre 15.000 pseudogeni, oltre a più di 200.000 trascritti. Queste informazioni sono enormi in termini di dimensioni e complessità, e la rappresentazione grafica fornita dai genome browser ha reso più accessibili e comprensibili tali dati*.*

*Che informazioni ci danno questi genome browser?*

Ci permettono di:

* Esplorare regioni cromosomiche;
* Esplorare regioni di regolazione fiancheggianti un gene;
* Effettuare ricerche di elementi (per parola chiave o similarità di sequenza) su scala dell’intero genoma;
* Investigare l’organizzazione del genoma;
* Comparare l’architettura del genoma in organismi differenti (comparative genomics).

Si distinguono: **annotazioni base** e **annotazioni avanzate**.

**ANNOTAZIONI DI BASE**

I genome browser, come Ensembl, UCSC e NCBI, forniscono informazioni dettagliate sui geni, compresi gli introni, gli esoni e le regioni non tradotte 5' e 3' UTR. Questi browser consentono di visualizzare la struttura dei geni, mostrando l'organizzazione degli esoni e degli introni lungo la sequenza genomica. I genome browser forniscono informazioni sui trascritti, ovvero le sequenze di RNA generate dai geni attraverso il processo di trascrizione. È possibile esplorare i diversi trascritti associati a un gene specifico, comprese le loro varianti alternative, e visualizzare le sequenze codificanti e non codificanti presenti all'interno dei trascritti.

I browser genomici sono anche una risorsa preziosa per l'identificazione e lo studio dei non coding RNA, che sono molecole di RNA che non vengono tradotte in proteine ma svolgono importanti funzioni regolatorie all'interno della cellula. È possibile accedere alle informazioni sui non coding RNA tramite i genome browser e studiare la loro posizione e interazione con i geni codificanti. Un'altra caratteristica dei genome browser è la possibilità di esplorare i pseudogeni, che sono sequenze simili a geni funzionanti ma che hanno perso la loro capacità di produrre una proteina funzionale.

**ANNOTAZIONI AVANZATE**

I genome browser, come Ensembl, UCSC e NCBI, offrono anche informazioni citogenetiche dettagliate. Questi browser consentono la visualizzazione del cromosoma in bande e sottobande, che forniscono un'indicazione visiva della struttura del cromosoma e delle sue caratteristiche distintive.

Tali browser permettono di identificare polimorfismi, inserzioni e delezioni all'interno della sequenza genomica. Queste variazioni genetiche possono essere evidenziate e visualizzate tramite i browser, consentendo agli utenti di analizzare le differenze tra individui o tra specie.

Sono in grado di individuare sequenze ripetute all'interno del genoma umano, le quali possono essere importanti per la comprensione della struttura genomica, delle regioni regolatrici e delle regioni coinvolte nella ricombinazione genetica. I browser genomici permettono di identificare e studiare queste sequenze ripetute in modo dettagliato.

I dati di espressione sono un'altra caratteristica fornita dai genome browser. Questi dati consentono di esplorare l'espressione dei geni in diversi tessuti o condizioni biologiche. Attraverso i browser, è possibile visualizzare i livelli di espressione dei geni e ottenere informazioni sulle regioni del genoma coinvolte nell'espressione genica differenziale.

Immagine che contiene testo, schermata, Carattere

Descrizione generata automaticamente

Ognuno dei tre ha caratteristiche diverse:

* NCBI allinea mRNA di RefSeq, mRNA di Genebank utilizzando Megablast;
* ENSEMBL allinea tutte le proteine umane note di SP/Trembl utilizzando un suo algoritmo,
* UCSC allinea mRNA di RefSeq e Genbank e dalle ultime release SP/Trembl con BLAST.

Si differenziano per il modo in cui prendono i dati e li analizzano.

**ENSEMBL**

Il progetto Ensembl è stato avviato nel 1999 con l'obiettivo principale di studiare organismi modello. Nel corso degli anni, il progetto si è evoluto e ora analizza i genomi di diverse specie. Questo offre agli utenti numerose risorse per i processi di allineamento genomica.

Ensembl fornisce anche risorse specifiche per una vasta gamma di organismi, tra cui metazoi, piante, funghi, protisti e batteri. Ciò significa che i ricercatori possono accedere a dati genomici e strumenti di analisi per una varietà di specie, consentendo di studiare la diversità e l'evoluzione genomica tra diversi organismi. Tutti i genomi depositati in Ensembl seguono una pipeline predefinita, in modo che i dati derivino da metodiche simili, garantendo uno standard qualitativo uniformemente elevato. Ogni genoma viene periodicamente aggiornato, soprattutto quando viene rilasciato un nuovo genoma di riferimento. Ad esempio, la release più recente del genoma umano è la GRCh38.p13. Con ogni nuova release, possono comparire o scomparire alcune annotazioni a seguito di nuove evidenze sperimentali o correzioni di predizioni in silico. Inoltre, in alcuni casi più rari, l'assemblaggio stesso del genoma può essere migliorato, permettendo una rappresentazione più accurata della sequenza genomica.

**SISTEMA DI ANNOTAZIONE**

Nel sistema di annotazione di un genome browser, le informazioni relative alla posizione di geni, trascritti, varianti ed elementi di regolazione vengono gestite in modo strutturato e funzionale. Ogni entità come gene, mRNA, proteina e esone ha un ID interno univoco che permette il riconoscimento e il collegamento ad altre schede correlate. Questi ID facilitano una ricerca rapida attraverso il motore di ricerca interno del browser. L'annotazione strutturale fornisce coordinate rispetto al cromosoma, mentre quella funzionale fornisce nomi e altre informazioni sui geni. Il genome browser permette quindi di navigare, esplorare e collegare in modo efficiente tutte queste informazioni all'interno del contesto genomico.

**ESPLORAZIONE DI UN GENOMA – PRIMI PASSI**

Nel genome browser, è possibile accedere a un chromosome summary cliccando su un cromosoma nel "view karyotype".

Alcuni chiarimenti importanti riguardo alle informazioni fornite nel genome browser includono:

* La dimensione dell'assemblaggio del genoma può apparire più grande di quanto ci si aspetti (3,2 Gb) a causa dell'inclusione di regioni che corrispondono ad aplotipi alternativi derivanti da molti individui.
* Una misura più accurata (sebbene non perfetta) della dimensione effettiva del genoma è data dalla "golden path length", che rappresenta la somma delle lunghezze di tutti i segmenti di genoma non ridondanti. Tuttavia, è una misurazione ideale.
* Sono disponibili anche statistiche sulle "alternative sequence" che rappresentano gli aplotipi alternativi che si differenziano dal genoma di riferimento.
* Il numero di geni annotati può variare leggermente con ogni release, poiché vengono scoperti nuovi geni e vengono corretti gli errori.

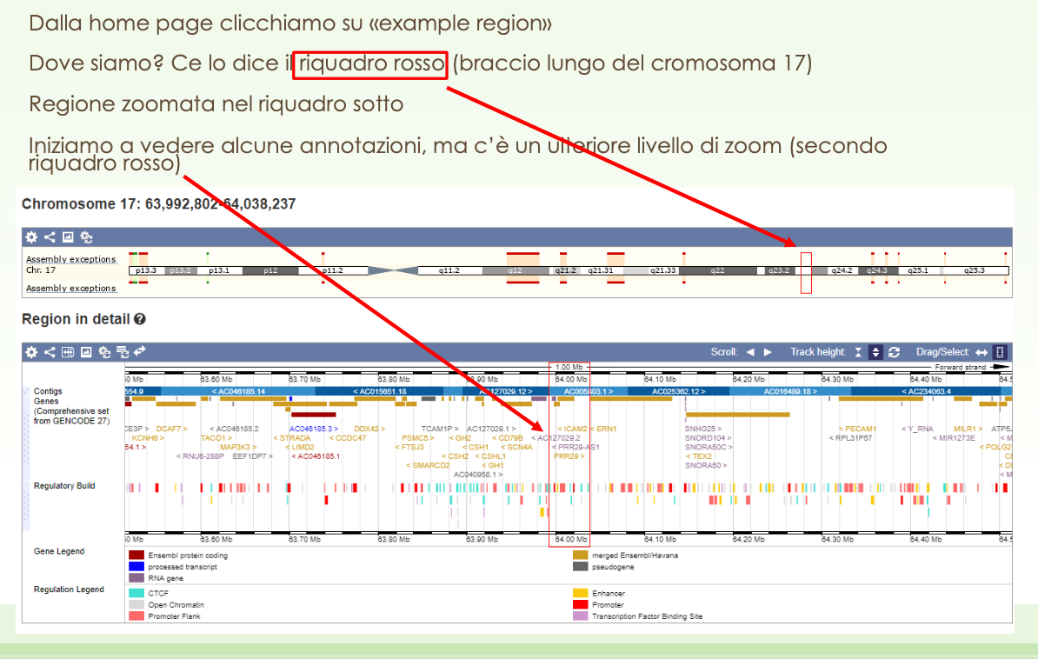
Inoltre, nella home page del genome browser è possibile trovare ulteriori informazioni sul genoma tramite la sezione "more about this genebuild", che fornisce una tabella riassuntiva dell'intero genoma

Immagine che contiene testo, schermata, numero

Descrizione generata automaticamente

**CHROMOSOME SUMMARY**

* Visione bandeggi
* Densità genica
* Densità geni non codificanti
* Densità pseudogeni
* Composizione GC
* Densità di varianti geniche
* Dimensione del cromosoma
* Numero di geni codificanti e non
* Numero di pseudogeni

Immagine che contiene testo, schermata, Sito Web

Descrizione generata automaticamenteIl rettangolino indica che quello che stiamo cercando è in quella posizione. E in seguito è possibile sapere che cosa vi sia in quella regione.

Ogni genome broswer ha una legenda e i colori sono personalizzabili.

*.*

TIPI DI ANNOTAZIONE CHE POSSO MOSTRARE

Immagine che contiene testo, numero, software, Carattere

Descrizione generata automaticamenteOltre a geni e trascritti...

• Varianti geniche (SNP e indel derivate da vari progetti di resequencing);

• Regulatory build: promotori, enhancer, ecc.

• Elementi per genomica comparata, come livello di conservazione in vari organismi di elementi codificanti o di regolazione;

• Ognuno di questi elementi è contrassegnato da un ID ed è collegato alla sua pagina;

SCHEDA RELATIVA AD UN TRASCRITTO

Immagine che contiene testo, schermata, software, Pagina Web

Descrizione generata automaticamente

Il UCSC Genome Browser offre una visione dettagliata dei siti di splicing, nonché delle dimensioni degli esoni e degli introni. Fornisce collegamenti incrociati a varianti geniche correlate, schede degli esoni individuali, geni e al browser stesso con una visualizzazione centrata sul trascritto.

Attraverso il collegamento alla "transcript table", è possibile ottenere una panoramica di tutti i trascritti prodotti da un gene.

Sono disponibili anche link diretti alle schede esterne che contengono le sequenze di riferimento (RefSeq) degli mRNA e delle proteine depositate in UniProt.

È importante prestare attenzione alle "Flags" che caratterizzano ciascun mRNA, in particolare il TLS (transcript support level), che indica su una scala da 1 a 5 quanto il trascritto sia supportato.

Un TLS-1 indica un alto grado di affidabilità, mentre è consigliabile evitare di consultare trascritti con TSL-5. Infatti, i trascritti che non codificano proteine sono generalmente associati a un TSL-5.

Immagine che contiene testo, schermata, software, numero

Descrizione generata automaticamenteIl «protein summary» può essere utile per identificare in modo grafico la posizione dei domini conservati che caratterizzano ciascuna proteina. Un dominio conservato solitamente identifica una regione strutturalmente conservata e che di conseguenza è legata ad una particolare funzione.

ESPLORARE LA VARIABILITA’ GENETICA

Ogni gene è associato a polimorfismi noti, che possono essere visualizzati nella "variant table". Questi polimorfismi possono essere SNP (single nucleotide polymorphisms), delezioni o inserzioni.

A seconda della loro localizzazione (5' UTR, 3' UTR o sequenza codificante), gli SNP possono essere sinonimi o non sinonimi. Queste varianti possono essere associate o meno a specifici fenotipi o patologie, come indicato nella colonna "clinical significance".

Esiste anche una scheda dedicata a ciascun SNP che fornisce una serie di informazioni correlate.

I link presenti sotto "explore this variant" ci aiutano a comprendere meglio alcuni aspetti.

La sezione "population genetics" ci consente di valutare la frequenza di una determinata mutazione nella popolazione generale o in un gruppo etnico specifico. Il link alle "citations" ci porta a riferimenti bibliografici che trattano i singoli polimorfismi.

La sezione "genes and regulation" indica la correlazione con gli effetti sull'espressione genica e localizza l'SNP nei trascritti di mRNA e proteine. Questi dettagli possono essere esplorati ulteriormente cliccando su "genomic context".

GENETICA DI POPOLAZIONE – FREQUENZE ALLELICHE E GENOTIPICHE

Vengono forniti dati suddivisi per gruppi etnici e singole popolazioni sotto forma di semplici diagrammi a torta o istogrammi. Sono dati derivati da progetti di risequenziamento genomico su larga scala (HapMap, 1000 Genome Project, GnomAd). Sono indicate frequenze e numeri assoluti delle osservazioni.

I dati possono esportati e possono essere poi studiati manualmente.

*La prof mostra come cercare in maniera pratica un gene su Ensembl.*

**UCSC GENOME BROSWER**

Il UCSC Genome Browser presenta le stesse informazioni di Ensembl, ma con annotazioni e visualizzazioni grafiche differenti.

Ospita 46 genomi, oltre al genoma umano, e consente di visualizzare tracce di interesse utilizzando le opzioni di visualizzazione come hide/show e diverse modalità di impacchettamento (pack/full/dense, ecc.).

* Le tracce sono suddivise in categorie come annotazioni di geni e proteine, link a dati di fenotipi e letteratura, dati di espressione, elementi di regolazione e strumenti di genomica comparata.
* È possibile utilizzare vari strumenti come la ricerca di omologia, il confronto tra genomi e la PCR virtuale.
* Il Gene Sorter fornisce dati di espressione, omologia e altre informazioni su gruppi di geni correlabili in modi diversi.
* Blat consente di mappare rapidamente una sequenza di interesse nel genoma e fare allineamenti con diverse specie presenti nel browser.

Il portale di ENCODE è una sezione dedicata all'Encyclopedia of DNA Elements, un consorzio internazionale che mira a creare un elenco degli elementi funzionali del genoma umano a tutti i livelli, compresi i geni codificanti, non codificanti e regolatori. Vengono utilizzati diversi saggi, come test di ipersensibilità del DNA, metilazione del DNA, immunoprecipitazione della cromatina (ChIP) di proteine che interagiscono con il DNA, tra cui istoni modificati e fattori di trascrizione, seguiti da sequenziamento (ChIP-Seq).

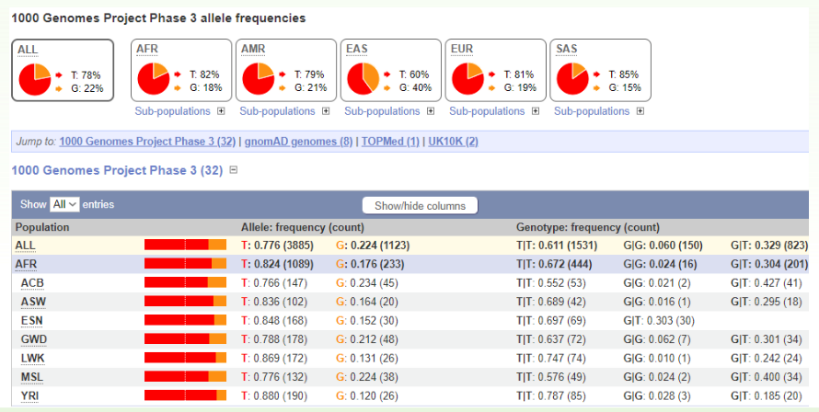
* UCSC Genome Browser fornisce anche informazioni sugli SNP (Single Nucleotide Polymorphism) con una visualizzazione completa e dettagliata.
* BLAT è uno strumento incluso nel UCSC Genome Browser che consente di mappare rapidamente una sequenza genomica desiderata e fare allineamenti con specie diverse presenti nel browser.

Immagine che contiene testo, software, Pagina Web, Sito Web

Descrizione generata automaticamente