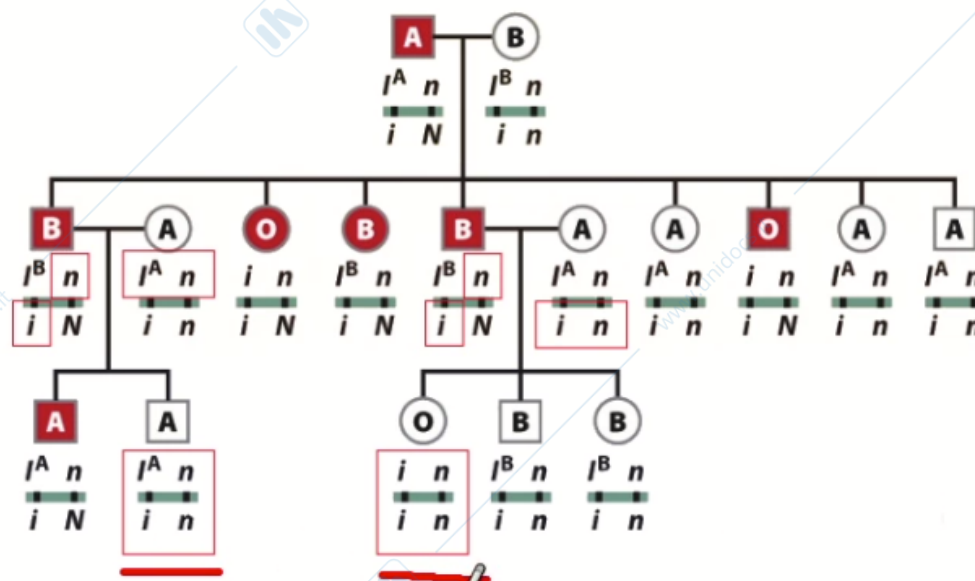


Mappatura del genoma umano

L'incrocio a due e tre punti possono essere identificati come strumenti base, classici per misurare le distanze genetiche, distanze in funzione della ricombinazione tra loci sul genoma e quindi procedere piano piano ad una mappatura dell'intero genoma → questo però è un occhio che proviene dalla senza dubbio importantissima ricerca che è stata fatta su *Drosophila*, nell'uomo linkage e a due e tre punti non si possono fare perché richiedono incroci controllati ed in più richiedono, visto che andiamo a stimare categorie che numericamente abbracciano un range molto ampio, la possibilità di studiare una prole molto numerosa. Tuttavia possono utilizzare altri approcci → si possono utilizzare i dati del pedigree, dal punto di vista teorico non è molto distante da un incrocio a due punti, si tratta di una stima della ricombinazione che invece di fare con un incrocio controllato (a due punti) si fa ragionando sul pedigree, oltre a ciò si possono utilizzare tecniche di citogenetica e applicazione di associazione genomica, che probabilmente sono il futuro (o almeno il presente) della mappatura → cose introdotte negli ultimi 20, che tramite un notevole sforzo di analisi/tecnologico permette di ottenere dati ricchi.

Iniziando da un'applicazione di linkage su pedigree e studiamo una delle prime dimostrazioni documentate di associazione nell'uomo → quella tra il locus che determina la sindrome unghia-rotula (condizione genetica dominante autosomica → nel pedigree identificata con N) e quello che determina i gruppi sanguigni (A,B, i → tre alleli A e B codominanti e dominanti su i).

Ricombinazione su pedigree (unghia-rotula)



Tracciando i cromosomi lungo la famiglia possiamo notare che non ci sono ricombinazioni (i cromosomi si ereditano così come sono) eccetto per **due individui, che devono aver ereditato un cromosoma ricombinato**. Pertanto la **frequenza di ricombinazione** è $2/13$, ovvero **0,15**. **NOTA**: numeri piccoli.

Cercheremo di constatare che la sindrome unghia-rotula e i gruppi sanguigni del sistema A,B,0 si ereditano in maniera associata.

Quadratino rosso → individuo malato

Quadrato bianco → individuo normale

Gli individui malati, soprattutto nella prima generazione, hanno gruppi B o O, già a colpo d'occhio il pedigree ci suggerisce che ci sia un'associazione tra questi due caratteri. Quello che bisogna fare è ricostruire i cromosomi lungo tutto il pedigree, E DUNQUE ricostruire quindi i genotipi di unghia-rotula, ricostruire i genotipi del gruppo A,B,O ed anche le loro relazioni sui cromosomi.

Iniziamo con il gruppo sanguigno, dai nonni → entrambi possono essere o eterozigoti (iA ed iB) oppure omozigoti (AA) (BB), quello che si osserva però è che nei figli ci sono anche degli individui con gruppo sanguigno O, pertanto per avere gruppo sanguigno zero devono avere ereditato un allele i da ciascuno dei nonni → per cui i nonni sono entrambi eterozigoti, trovato il loro genotipo banalmente si trovano tutti gli altri.

Per quanto riguarda il gene per la sindrome unghia rotula, trattandosi di una condizione molto rara praticamente non si osserva mai l'omozigote dominante, gli individui malati avranno il genotipo eterozigote Nn ed i non affetti il genotipo omozigote nn .

A questo punto bisogna posizionare questi geni sui cromosomi, identificare i cromosomi parentali → il nonno è $iA Nn$ e la nonna $iBnn$, però ci sono due modi per sistemare questi geni sui cromosomi → posso avere i due recessivi (i ed n) sullo stesso cromosoma o su due cromosomi diversi (in trans), in questo caso ci aiuta un colpo d'occhio sul pedigree → gli individui con gruppo A sono quasi sempre sani, mentre quelli con gruppo B o gruppo zero sono quasi sempre malati, pertanto l'allele A starà preferenzialmente con n , mentre l'allele i starà con l'allele malato → pertanto assegniamo al nonno la condizione in trans.

Per la nonna c'è un solo modo di disporre gli alleli.

Dunque si possono tracciare i cromosomi lungo il pedigree, tutto torna, vengono trasmessi cromosomi intatti, senza combinazione → soltanto due individui non tornano (vedi slide)

- Il primo prende un cromosoma dalla madre, che glielo tramanda intatto, l'altro cromosoma lo deve prendere dal padre, il figlio però sul cromosoma presenta i , che stavano originariamente su cromosomi separati, il padre non aveva un cromosoma i da passare → nel padre è avvenuta ricombinazione tra i suoi due cromosomi, si è formato un nuovo cromosoma per fusione di i e d in i , e questo nuovo cromosoma è stato trasmesso al figlio → è avvenuta ricombinazione.
- Lo stesso è avvenuto per il cromosoma posseduto dal secondo nipote.
- Per cui in questo pedigree osserviamo 13 eventi, 13 figli, 13 formazioni di gameti e su ciò osserviamo 2 ricombinazioni → abbiamo pertanto stimato (si tratta di numeri molto piccoli, nel caso di *Drosophila* non verrebbero presi in considerazione, ma nell'uomo, magari aggiungendo altri pedigree della stessa condizione, si può arrivare a dare una stima della frequenza di ricombinazione, che in questo caso è $2/13 \rightarrow 0,15$. Se facessimo finta di procedere solo con questo pedigree di questa famiglia (in realtà bisogna studiare altri pedigree della stessa condizione per avere un numero maggiore di individui da considerare) avremmo stimato che il gene che determina la condizione unghia-rotula e quello che determina il gruppo sanguigno sono a 15 centimorgan l'uno dall'altro.
Si possono dunque stimare le distanze di mappa partendo dal pedigree.

- Un altro modo di lavorare che è importante per l'uomo proprio perché le alternative sono scarse e poi perché è un sistema ad alto costo ed alta tecnologia per cui è stato sviluppato prima per gli organismi per cui c'era più interesse (e dunque l'uomo) anche se oggi è una tecnologia che viene applicata a tanti altri organismi, in ogni caso gli studi di ASSOCIAZIONE GENOMEWIDE, non si basano sull'ereditarietà che si osserva all'interno di una famiglia per uno o pochi caratteri, ma mirano ad osservare in un colpo solo/analisi sola, le relazioni tra migliaia di caratteri (per studi

moderni umani possiamo arrivare ad 1 o più milioni di caratteri tutti genotipizzati in uno stesso individuo in uno stesso momento) e una condizione di interesse → non viene studiata una famiglia, ma individui diversi presi dalla popolazione. Si riesce dunque a sopperire con grandi numeri ed un approccio statistico al dato che si poteva avere, più preciso, anche su piccoli numeri, andando a mappare una famiglia, però in una famiglia non si possono mappare centinaia di migliaia di loci in contemporanea, si tratta di un approccio diverso, che non si basa su caratteri morfologici ma su caratteri molecolari.

- Cerchiamo di capire cosa si intende per associazioni genomewide → diciamo che in una popolazione ci sono numerose varianti, numerosi cromosomi omologhi e se ne trovano differenziate, perché quasi ogni individuo avrà il proprio pattern di distribuzione allelica, sul cromosoma ci sono migliaia di geni ed ogni individuo avrà una distribuzione di dominanti, recessivi, alleli diversi di questi geni. Quando una mutazione di interesse, una mutazione che determina quindi una condizione patologica che ci interessa avviene, al momento in cui avviene, avviene su un singolo cromosoma (poi si potrà anche diffondere nella popolazione, ma quando avviene, avviene su uno specifico cromosoma, specifica collocazione sul cromosoma ed ha insieme a sé gli alleli per gli altri geni che casualmente si trovavano sul cromosoma da dove è nata la condizione), questo nell'individuo in cui è avvenuta la mutazione. Nelle generazioni successive, quel cromosoma ad ogni generazione, subisce meiosi, fa crossing over e si mescola con altri cromosomi che casualmente erano finiti sullo stesso individuo → il fatto che avvengano crossing over fa sì che il cromosoma primo che aveva subito la condizione e gli altri cromosomi omologhi si scambino pezzi, ora, più è lontano un frammento di questo cromosoma dal locus di interesse che stiamo cercando di mappare, più è facile che questo venga ci sia un crossing over e che questo venga soppiantato, sostituito da altri pezzi provenienti da altri cromosomi, mentre per sostituire loci molto molto vicini al locus di interesse è necessario un evento abbastanza raro → il crossing over deve cadere esattamente tra il locus adiacente e quello di interesse. Quello che ci si aspetta quindi è che l'associazione tra un qualunque marcatore lungo il cromosoma ed il marcatore di interesse (e cioè il gene che stiamo mappando) si perda abbastanza velocemente, nel giro di poche generazioni, se il marcatore è lontano, ma che permanga per molte generazioni se il marcatore è vicino. Dunque noi andiamo proprio a cercare questi marcatori associati al gene di interesse, perché se ne vede l'associazione, probabilmente saranno fisicamente vicini.

Associazioni genomewide

- $A^1 B^1 C^1 D^1 E^1 GO^M F^1 G^1 H^1 I^1$
- $A^2 B^2 C^2 D^2 E^2 GO^S F^2 G^2 H^2 I^2$
- $A^3 B^3 C^3 D^3 E^3 GO^S F^3 G^3 H^3 I^3$
- $A^4 B^4 C^4 D^4 E^4 GO^S F^4 G^4 H^4 I^4$
- $A^5 B^5 C^5 D^5 E^5 GO^S F^5 G^5 H^5 I^5$
- $A^6 B^6 C^6 D^6 E^6 GO^S F^6 G^6 H^6 I^6$
- $A^7 B^7 C^7 D^7 E^7 GO^S F^7 G^7 H^7 I^7$

A^6	B^2	C^3	D^7	E^1	GO^M	F^1	G^6	H^4	I^7
A^7	B^2	C^4	D^3	E^1	GO^M	F^1	G^5	H^7	I^4
A^2	B^6	C^2	D^7	E^7	GO^S	F^1	G^4	H^6	I^5
A^6	B^3	C^6	D^5	E^1	GO^S	F^3	G^6	H^5	I^3

Si osserva un'associazione fra il carattere (GO^M) e due specifici marcatori (E^1 e F^1). GO^M è localizzato nelle loro vicinanze.

In realtà, lungo un cromosoma, in questi studi, si fa in modo di avere migliaia di geni di marcatori lungo ciascun cromosoma. Quello che si vede nell'immagine è che il cromosoma che ha subito la mutazione è il cromosoma 1 (Gene GO, che nel resto della popolazione è presente come variante sana, in questo cromosoma 1 è presente come variante malata, quello che voglio fare è localizzare la posizione di questo gene, sano o malato che sia, e mapparli).

Quindi, genotipizzo, in tutta la popolazione il gene di interesse (divido in gruppi in cui è presente la variante malata e sana) e tutti gli altri marcatori che posso visualizzare sul genoma (in realtà centinaia di migliaia), poi questi cromosomi si mescolano e data la posizione del gene che voglio mappare, la probabilità dopo un certo numero di generazioni, di trovare il gene ancora associato a marcatori distanti è minima, i geni si mischiano e appaiono in maniera casuale, eccetto per i marcatori ai loci E ed F, che sono molto vicini al locus che sto tentando di mappare. Se non è avvenuto alcun crossing over in grado di separare E ed F dal gene, troverò che il gene di interesse forma malata è sempre associato all'allele 1 del locus F e all'allele 1 del gene E, che erano quelli che lo accompagnavano quando era accaduta la prima mutazione → quindi si procede determinando i genotipi di un alto numero di individui per un altissimo numero di marcatori, prendendo sia individui che hanno la condizione sia individui che hanno la condizione sana per il locus che voglio mappare, e poi vado a vedere con quali altri marcatori, con quali altri loci il gene di interesse è associato, se lo trovo associato ad uno specifico allele del locus E o del locus F, dirò che il gene di interesse si trova fisicamente vicino ai geni E ed F.

L'idea intermedia, il passaggio logico è che se l'allele malato del gene di interesse è associato agli alleli 1 di E ed F → vuol dire che mai un crossing over gli ha separati, la spiegazione più plausibile a ciò è che siano talmente vicini che la probabilità di un crossing over che li separi è molto molto bassa.

Ma cosa sono questi centinaia di migliaia di marcatori che posso studiare lungo i cromosomi? Non sono marcatori fenotipici, si conoscono molti geni che hanno eredità mendeliana nell'uomo (le fossette, le lentiggini, l'attaccatura dei capelli e la forma del lobo delle orecchie, sono molti caratteri biochimici, molte malattie genetiche, ma sempre una manciata di caratteri).

Per arrivare a numeri più elevati (centinaia di caratteri) bisogna andare nella categoria di caratteri molecolari, che non osservo nel fenotipo dell'individuo, ma che analizzo sulla sua sequenza di DNA, la loro natura fisica è una differenza nella sequenza genetica nei diversi individui. Non si guardano differenze a caso, ma differenze che tecnicamente posso genotipizzare (stabilire il genotipo) che possono andare ad indagare con qualche tecnologia che non mi riobblighi a rifequenziare tutto il genoma, si utilizzano delle caratteristiche del genoma che posso individuare anche senza ottenere la sequenza completa del genoma di ogni individuo.

Questi marcatori molecolari hanno anche un altro vantaggio → sono in qualche modo standardizzabili, se si pensa a caratteri morfologici o biochimici (colore del fiore, funzionalità dell'enzima che cambia) → sono caratteri specifici, che richiedono conoscenze diverse ogni volta che si cambia specie studiata (si pensi al colore del fiore di pisello come esempio di dominanza e nella bocca di leone come esempio di dominanza incompleta), che richiedono tecniche specifiche. Questo, insieme al fatto che guardare insieme caratteri morfologici, fisiologici, biochimici, richiede esperti diversi, laboratori diversi e macchinari diversi fa sì che studi di questo tipo non siano espandibili nell'ottica di HIGH THROUGHPUT (un sistema settato per cui qualunque campione metta all'inizio io ottengo risultati e risultati sui grandi numeri alla fine, senza bisogno di metterci troppo le mani).

I marcatori molecolari sono ottimi da questo punto di vista, essenzialmente si va a vedere è una sequenza di DNA, che questo sia umano, di pianta, di drosophila, alla fine, dal punto di vista della tecnologia è essenzialmente la stessa cosa (anche se useremo marcatori differenti etc) e quindi questa possibilità ha reso l'utilizzo dei marcatori molecolari molto potente (vengono introdotti negli anni 80, oggi l'introduzione di nuovi metodi ha permesso un ampliamento dei marcatori molecolari e dunque facilitato la creazione di mappe genetiche).

Alcuni esempi:

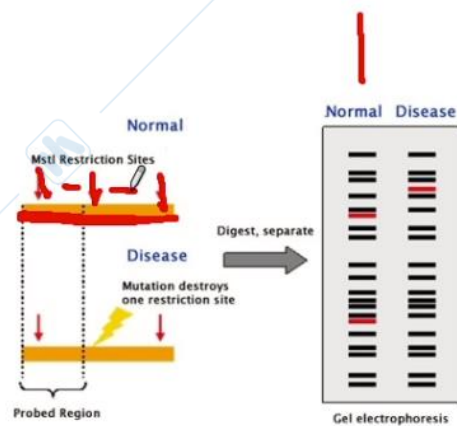
I primi marcatori usati sono stati gli RFLP → restriction fragment length polymorphism (polimorfismi della lunghezza dei frammenti di restrizione).

→ Gli enzimi di restrizione sono enzimi che tagliano il DNA riconoscendo una specifica lunghezza di basi (che può avere una lunghezza di 4,6,8 basi), ogni volta che c'è una sequenza target per l'enzima di restrizione (se c'è il DNA è esposto all'enzima) allora l'enzima taglia.

Dunque posso digerire un DNA di interesse e ottenere tutta una serie di frammenti che hanno lunghezze diverse.

Il cromosoma normale nella slide presenta 3 siti di restrizione che mi determinano la formazione di due frammenti di DNA, diciamo che una mutazione nella sequenza di DNA impatta esattamente sul sito di riconoscimento dell'enzima che sto usando e di fatto cambia il sito di restrizione di una base → questo non viene più riconosciuto dall'enzima, di fatto scompare questo sito di restrizione.

Se digerisco questo cromosoma mutato avrà solo due siti di restrizione e dunque un solo frammento di lunghezza pari alla somma degli altri due. Questo è un marcatore, posso visualizzare le bande, posso determinarne la lunghezza e posso determinare il genotipo di entrambi in funzione della presenza o assenza di questi siti di restrizione. Su un genoma si possono visualizzare moltissimi siti di restrizione diversi, posso di fatto cercare differenze nel bandeggio tra organismi normali e organismi che portano la malattia.



RFLP – restriction fragment length polymorphism, ovvero polimorfismi nei frammenti di restrizione.



Un altro tipo di marcatori che è stato introdotto successivamente e che richiede un macchinario che è un sequenziatore di DNA capillare, qui i pezzetti di genoma che posso visualizzare e studiare come microsatelliti sono sequenze che hanno delle brevi ripetizioni ripetute in tandem.

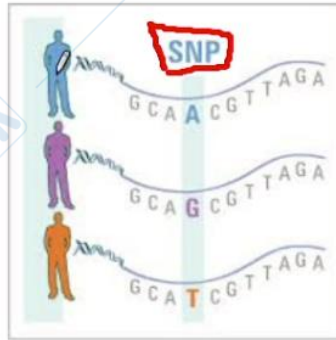
L'allele 1 ha, ad esempio, una sua sequenza a sinistra e poi 10-20 ripetizioni CACACACACA... queste zone tendono ad essere molto variabili per l'aumento o perdite di unità della zona ripetuta, quindi l'allele 1 sarà formato, diciamo, da 10 ripetizioni di CA, l'allele 2 da 12 ripetizioni di CA, l'allele 3 da 14 ripetizioni CA. Quando vado a farli migrare per lunghezza (ogni picco corrisponde ad una lunghezza) ottengo dei genotipi che posso scrivere nella forma rappresentata.

L'individuo 1 ha due copie dell'allele che un certo numero di ripetizioni

L'individuo 2 ha due copie dell'allele che ha un certo numero di ripetizioni, maggiori dell'allele 1 (sempre omozigote)

L'individuo 4 presenta 2 bande perché presenta 2 alleli, uno ha lo stesso numero di ripetizioni dell'individuo 1 e l'altro lo stesso numero di ripetizioni dell'individuo 2.

Di fatto si possono leggere in termini di genotipi, nel caso dell'individuo 4 si tratta di genotipi codominanti perché riesco a leggerne 2.



SNPs – single nucleotide polymorphism, ovvero polimorfismi a livello di singolo nucleotide.



Marcatore associato → da uno studio genomewide emerge uno o più marcatori associati al tratto di interesse (c'è uno specifico allele del marcatore associato ad uno specifico allele del tratto di interesse, mai separati dal crossing over e quindi sono con tutta probabilità vicini sul cromosoma). Associazione implica vicinanza fisica, non implica un coinvolgimento funzionale, il marcatore che posso aver trovato associato, essere vicini non significa essere implicati funzionalmente, in genere il marcatore è del tutto estraneo al tratto di interesse, questo perché è localizzato nella stessa regione ma non nello stesso gene, normalmente si trovano in zone non codificanti nel genoma, non hanno un effetto funzionale, non sono loro la causa del fenotipo, del malfunzionamento (Dunque se il marcatore in questione che posso aver trovato associato è una data mutazione di un nucleotide, uno SNP associato, non è questo a determinare il carattere) sono però vicine → il che è un risultato importante, ho visualizzato la regione in cui si trova il gene di interesse, posso dunque andare a focalizzare il mio studio soltanto su quella regione → molti geni di interesse anche per patologie sono stati isolati con questa tecnica → si trova un marcatore associato che mi identifica una regione genomica, vado a vedere che geni ci sono in quella regione (1, 2, o 3 geni), posso quindi studiare 1, 2 o 3 geni, come si esprimono, se portano sequenze diverse negli individui sani e malati e confermare quale sia quello effettivamente coinvolto come causa, quello che funzionalmente determina il fenotipo patologico.

Dunque marcatore associato significa semplicemente che si trova nelle immediate vicinanze, non significa che quella mutazione sia coinvolta nel fenotipo mutato.

Abbiamo inoltre altri sistemi di mappatura che vanno sotto il nome di MAPPA FISICA.

Per mappa fisica si intende una mappa direttamente associabile alla morfologia del cromosoma, che a seconda della finezza e precisione della mappa, può essere a diversi livelli → potrebbe corrispondere grossolanamente al sapere se il marcatore si trova sul braccio lungo o sul braccio corto del cromosoma o sapere se si trova in un certo cromosoma, ma potrebbe anche far riferimento alla posizione del marcatore rispetto ad una banda visualizzabile al microscopio (in generale i cromosomi di drosophila si prestano molto bene a questo tipo di studio, ma anche i cromosomi umani, in realtà, se trattati nel modo giusto, sviluppano un bandeggio, un alternarsi di bande chiare e colorate che possono essere utilizzate per creare un disegno, una mappa fisica del

cromosoma). Una mappa ancora più fine coincide con la sequenza completa del genoma, se io ho quindi l'intera sequenza del genoma e so dove sono due geni posso semplicemente mapparli con una distanza espressa in numero di paia di basi.

NB→ Tutti questi metodi hanno un riferimento fisico, ogni punto può essere fisicamente associato ad un punto del cromosoma.

Mapa fisica

Si basa su informazioni che fanno riferimento a **quantità fisiche**:

- lunghezza (es. in paia di basi)
- posizione rispetto a siti cromosomici visibili (braccio corto)
- posizione rispetto ad una banda visualizzabile al microscopio



Fig. 13-26: A detailed match of the entire set of polytene chromosomes.

Un metodo per posizionare i geni sui cromosomi è rappresentato dall'**ibridazione delle cellule somatiche**, tecnica che è stata molto usata, si usa per rispondere a domande quali "su quale cromosoma sta un particolare gene?" e non solo. Un limite della tecnica è che può visualizzare soltanto geni la cui presenza possa essere determinata a livello molecolare, perché si lavora su colture cellulari, devo poter visualizzare la presenza di un gene con un test biochimico, se il carattere che mi interessa è la capacità di produrre un enzima (ad esempio) lo posso usare, se è "i capelli scuri" non lo posso usare, perché da una cellula non posso vedere il colore dei capelli, magari posso vedere un pigmento prodotto, ma non il colore dei capelli.

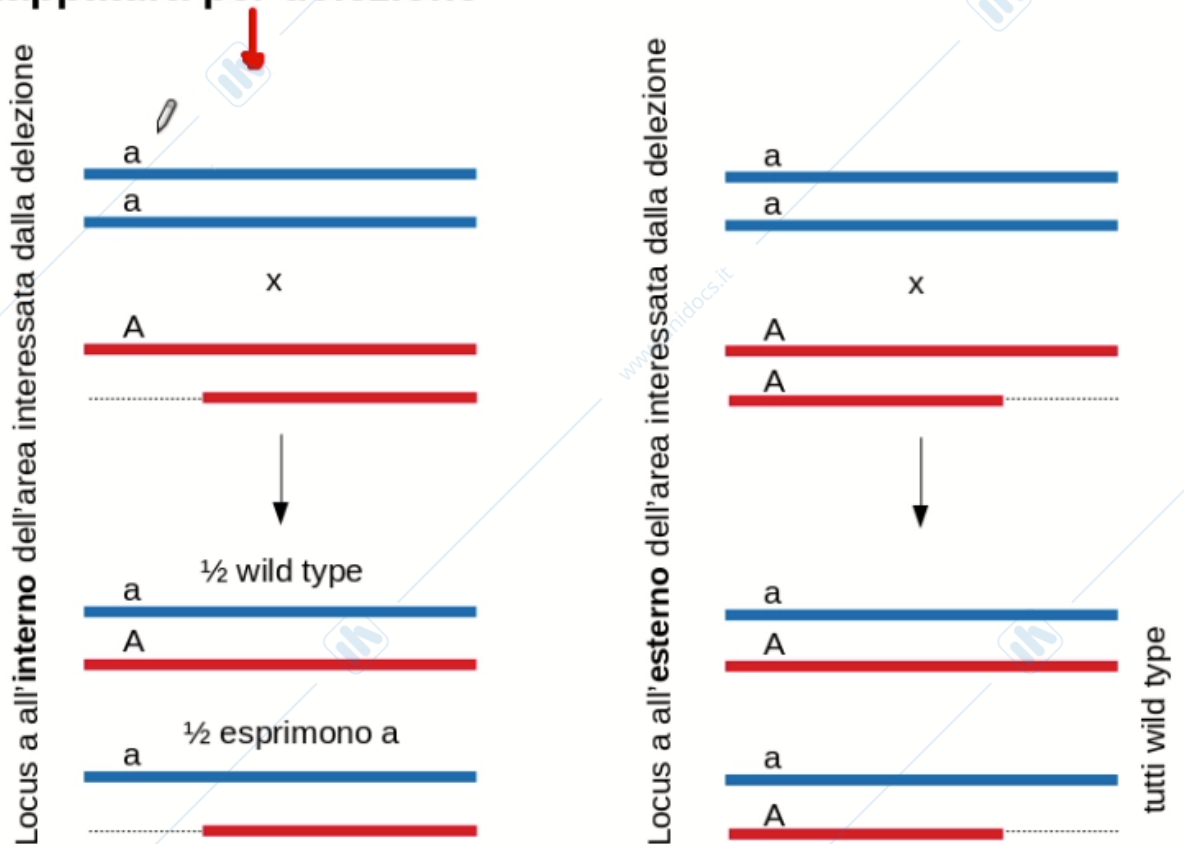
Si sfrutta una caratteristica delle cellule in coltura, che possono essere isolate, vive, da un organismo e che si riproducono per alcune generazioni e poi muoiono (tuttavia esistono linee cellulari chiamate immortalizzate, che derivano da cellule tumorali, che si riproducono per un tempo indeterminato e che sono pertanto ottimi strumenti da laboratorio)→ si fanno fondere due cellule diverse (una di uomo ed una di topo) utilizzando un reagente, Polietilen glicol, che altera le membrane cellulari→ queste alla fine si fondono. Le cellule, una volta fuse, per le prime fasi sono instabili a livello cromosomico→ perdono cioè numerosi cromosomi→ dopo un po' si stabilizzano, in particolare le cellule ibride conservano i cromosomi di topo e tendono a perdere i cromosomi umani (ne rimangono pochi e sono sempre gli stessi in ogni linea). Avrò quindi una linea che ha solo il cromosoma 1 umano, una linea che ha solo i cromosomi 3 e 4, un'altra che ha solo i cromosomi 5 e 7. In ciascuna di queste linee posso fare la mia misura biochimica e vedere quale produce l'enzima di cui sto cercando il gene responsabile, troverò che poche linee cellulari rispondono a questa ricerca e vedendo quale cromosoma hanno in comune, posso comprendere che su questo sarà localizzato il gene di interesse.

Questa tecnica può essere utilizzata anche per cercare la zona di cromosoma utilizzando linee cellulari che hanno un dato cromosoma con una specifica delezione, quindi utilizzando il suddetto ragionamento posso capire in quale zona del cromosoma sia localizzato il gene.

Mappatura per delezione → è un'altra tecnica che è stata ampiamente utilizzata che è stata ampiamente utilizzata, normalmente si conoscono linee delle famiglie in cui sono presenti delle delezioni cromosomiche, che sono state ben caratterizzate a livello citologico, si conosce quindi la mappa in termini di bandeggio dei cromosomi di questi individui (questa tecnica è stata molto usata per drosophila) e in base al bandeggio si vede quale pezzetto di cromosoma è stato perso.

→ si tratta di un ottimo strumento per localizzare geni altri, di cui non si conosce la localizzazione, ad esempio si può fare un incrocio tra un individuo omozigote per una condizione recessiva da mappare con cromosomi senza delezioni, con un ceppo wild type per la condizione da mappare (non presenta la mutazione dunque) e che porta una delezione in forma eterozigote → SE IL GENE DI INTERESSE È NELL'AREA CHE corrisponde alla delezione si avrà metà dei figli wild type e metà mutanti, se il gene è fuori dall'area deleta si avranno tutti figli wild type (genotipo diverso, ma stesso fenotipo → tutti i figli con fenotipo dominante) → è un sistema di mappatura fisica perché posso dire se è all'interno dell'area deleta, conosco dove si trova nel cromosoma in base al bandeggio.

Mappatura per delezione



Un'altra tecnica interessante in questi termini che si può usare per localizzare su un genoma qualunque gene o qualunque marcatore di cui io conosco la sequenza nucleotidica → è l'ibridazione in situ fluorescente (FISH) → la tecnica si usa moltissimo se già uno lavora su sequenze, non si può usare per

mappare il gene per il colore del fiore, ma si può usare per mappare un marcatore molecolare di cui io conosco la sequenza, il prerequisito è conoscere di DNA che uno vuol mappare, il frammento.

Quello che si fa è produrre un piccolo frammento di DNA che porta la sequenza di interesse e marcarlo con delle molecole fluorescenti → questo prende il nome di probe (sonda), quest'ultima molecola fluorescente, avendo una specifica sequenza, ha una forte tendenza ad ibridizzarsi nel genoma, non in una regione a caso, ma nella posizione in cui nel genoma si trova la sequenza complementare (le sequenze reverse complement hanno forte tendenza ad appaiarsi) per cui in un genoma in cui si trova un mix di 3 miliardi di paia di basi, questi probe tenderanno a muoversi in soluzione fino ad andare a trovare il punto in cui c'è la sequenza complementare, poi a questo punto posso visualizzare il kariogramma, i cromosomi meiotici metafasici al microscopio visualizzando i cromosomi → irraggiando con una lunghezza d'onda adatta che eccita le molecole fluorescenti che erano attaccati al probe, il probe spiccherà come pallini colorati (si possono usare colori diversi) → NON C'è ALTISSIMA RISOLUZIONE, MA È UNO STRUMENTO ESTREMAMENTE POTENTE, BASTA DIFATTI CONOSCERE la sequenza di DNA che si può mappare dove si trova sul genoma e vederlo in termini fisici come pallino sovrainpresso.

→ Il sequenziamento completo del genoma umano, ovvero il PROGETTO GENOMA UMANO, progetto di dimensioni estreme (paragonabile solo all'esplorazione spaziale in termini di lavoro, costi, ma anche impatto sulla società).

La tecnologia di sequenziamento con cui è iniziato il progetto era disponibile già dagli anni '70, ma la dimensione del genoma umano (3,3 miliardi di paia di basi di genoma aploide) era esageratamente più grande di quella che mai era stata sequenziata. Negli anni 80 si inizia a parlarne, nell'85 si decide di avviare il progetto che di fatto inizia nel 1990 con una stima di 15 anni di lavoro, verrà dichiarato completato nel 2003. Il progetto è guidato da Francis Collins, finanziato in larga parte dagli Stati Uniti da DOE (department of energy) e da NIH (National Institute Of Health) per un totale di oltre 5 miliardi di dollari → si parla di centinaia di gruppi di ricerca che parallelamente lavorano allo stesso progetto, molte nazioni hanno dato il loro contributo (anche se il lavoro è stato portato avanti principalmente da Stati Uniti ed Inghilterra) e questo era in un consorzio pubblico, dunque organizzato dallo Stato, gestito da centri di ricerca statali ed università, che si rispecchiava in una visione di Scienza pubblica, il progetto del consorzio pubblico va dal '90 al 2003, il competitore si presenta nel '98 → uno scienziato Craig Venter fonda una piccola ditta (Celera Genomics) e con un buon investimento (non paragonabile all'investimento pubblico del progetto genoma umano) si lancia nell'impresa e lancia un progetto parallelo, sfidando il consorzio pubblico ad arrivare prima di loro al risultato, con una frazione molto minore dei costi e del tempo a disposizione. Il progetto genoma si dichiara concluso pochi anni dopo, l'annuncio è del 2000, ma i risultati, una "working draft" del genoma, vengono pubblicati nel 2001 (11 anni dopo l'inizio). In questi pochi sono successe molte cose → una competizione molto spiccata tra consorzio pubblico e celera genomics, che però viene plasticamente riconciliata in qualche modo con una foto comune con il presidente dell'epoca (Clinton) e due pubblicazioni parallele sulle due riviste più importanti al mondo (Scienze e Nature → celera genomics). Di fatto il genoma proposto deriva da un mix di dati provenienti dall'uno e dall'altro fronte.

Cosa era stato sequenziato nel 2001? Una draft, la bozza migliore viene pubblicata nel 2003 (anche se viene rimaneggiata anche dopo) → non è stato sequenziato tutto il genoma, ma la parte eucromatica, perché contiene geni, si esprime, e perché è più gestibile per il sequenziamento, non sono state sequenziate, anche perché di interesse limitato, le zone ad esempio del centromero, dei telomeri, che sono stabilmente in forma di eterocromatina e dunque anche più difficili da maneggiare, comunque si parla di una parte sostanziale del genoma o del totale del genoma che ci interessa → con un'accuratezza del 99,9% (proprio come lo voleva Francis Collins, "facciamo il genoma, ma facciamolo con un'accuratezza del 99,9%").

Ancora oggi si studia il genoma, i fronti di approfondimento sono l'interpretazione funzionale (come funziona, quali sono i geni, come si attivano, come si controllano, quali sono le relazioni) e le variabilità tra genomi di individui diversi (quello che è stato sequenziato è un genoma, il realtà il mix di un limitato numero di individui) ci interessa sapere quali sono le variazioni tra gruppi separati da molte generazioni e quindi anche per tracciare una storia dell'umanità.

Venter → capo, CEO, di un'industria privata che deve mantenersi e deve fare soldi → punta a scoprire geni di interesse medico che possono interessare le compagnie farmaceutiche, brevettarli e vendere i brevetti alle aziende farmaceutiche, mira dunque a produrre un capitale di ritorno. Propone una tecnica diversa, mai stata utilizzata fino a quel momento per genomi di una certa complessità → shearing/shot gun → frammentare il genoma tutto a caso e rimontarlo in base alle similarità di sequenza, Venter arriva vicino a ricomporlo. Venter, che aveva esperienza (perché aveva lavorato in un centro di sequenziamento pubblico) con Celera Genomics si dedica essenzialmente all'uomo.

Collins → scienziato del sistema pubblico, accademico, punta sulla collaborazione (progetto inclusivo, internazionale), l'unico scopo è quello di essere la scoperta del secolo e di essere donata pubblicamente alla comunità scientifica e a tutta l'umanità. Parte con tecniche ben rodiate nel passato, prima una mappa genetica del genoma poi una mappa fisica, poi la rottura del genoma in tanti pezzi, sequenziare ciascun pezzo, poi utilizzare mappa fisica e mappa genomica per rimettere i pezzi insieme. Collins comincia con dei piccoli genomi (drosophila) per vedere se il sistema adottato funziona.

→ Clinton risolve la situazione perché passa i geni umani come non brevettabili.

Due diversi progetti, due diverse filosofie per una delle imprese fondamentali del secolo passato, ha cambiato il modo di lavorare per chi fa biochimica, genetica, biotecnologie, biologia molecolare, ed è stato uno spartiacque sociale, tutti guardavano al progetto genoma umano come a quella esperienza scientifica che ci avrebbe fatto fare un salto di qualità in tutti i campi. È stato sicuramente un grande passo avanti.