

# Variabilità del genoma umano

---

# Variabilità del genoma umano

- Come visto nei genomi virali, anche il genoma umano presenta variazioni di sequenza nel DNA di un individuo rispetto ad un altro.
- Così come osservando due persone si può dire che sono diverse, anche i genomi di due persone presentano differenze che, con gli appropriati strumenti, possono essere evidenziate.
- Se il genoma fosse un libro, il libro di ogni persona conterrebbe gli stessi paragrafi e capitoli, disposti nello stesso ordine. Ogni libro racconterebbe la stessa storia.
- Ma un libro potrebbe contenere un errore a pagina 303 che manca in un altro, o un altro libro potrebbe usare un'ortografia britannica a pagina 135 - "*colour*" – e l'altro usare l'ortografia americana - "*color*".

La genetica umana e medica fa uso di piccole variazioni neutre, i polimorfismi del DNA, e cerca gli errori, le mutazioni geniche, per svolgere le proprie analisi.

# Polimorfismi genetici

- Il **99.9%** del DNA degli individui presenti nella popolazione è identico
- Solo i gemelli monozigoti possiedono lo stesso DNA genomico



- Queste differenze di composizione nel DNA degli individui è sufficiente per dar luogo alla eterogeneità della popolazione.
- Queste differenze nel DNA tra individui della medesima popolazione sono detti **polimorfismi genetici**
- I polimorfismi generalmente non impattano o solo in maniera marginale il fenotipo individuale

# Polimorfismi vs. mutazioni geniche

## VARIABILITÀ GENETICA

```
graph TD; A[VARIABILITÀ GENETICA] --> B[Il polimorfismo è una variazione nel DNA che può avere un effetto fenotipico minimo o nullo; la sua frequenza nella popolazione è MAGGIORE dell' 1%]; A --> C[Una mutazione è un cambio del genotipo che produce variabilità modificando il fenotipo; la sua frequenza nella popolazione è INFERIORE all'1%]; B --> D[FENOTIPO NORMALE]; C --> E[FENOTIPO MALATTIA];
```

Il **polimorfismo** è una variazione nel DNA che può avere un effetto **fenotipico minimo o nullo**; la sua frequenza nella popolazione è **MAGGIORE dell' 1%**



**FENOTIPO NORMALE**

Una **mutazione** è un cambio del genotipo che produce variabilità **modificando il fenotipo**; la sua frequenza nella popolazione è **INFERIORE all'1%**.

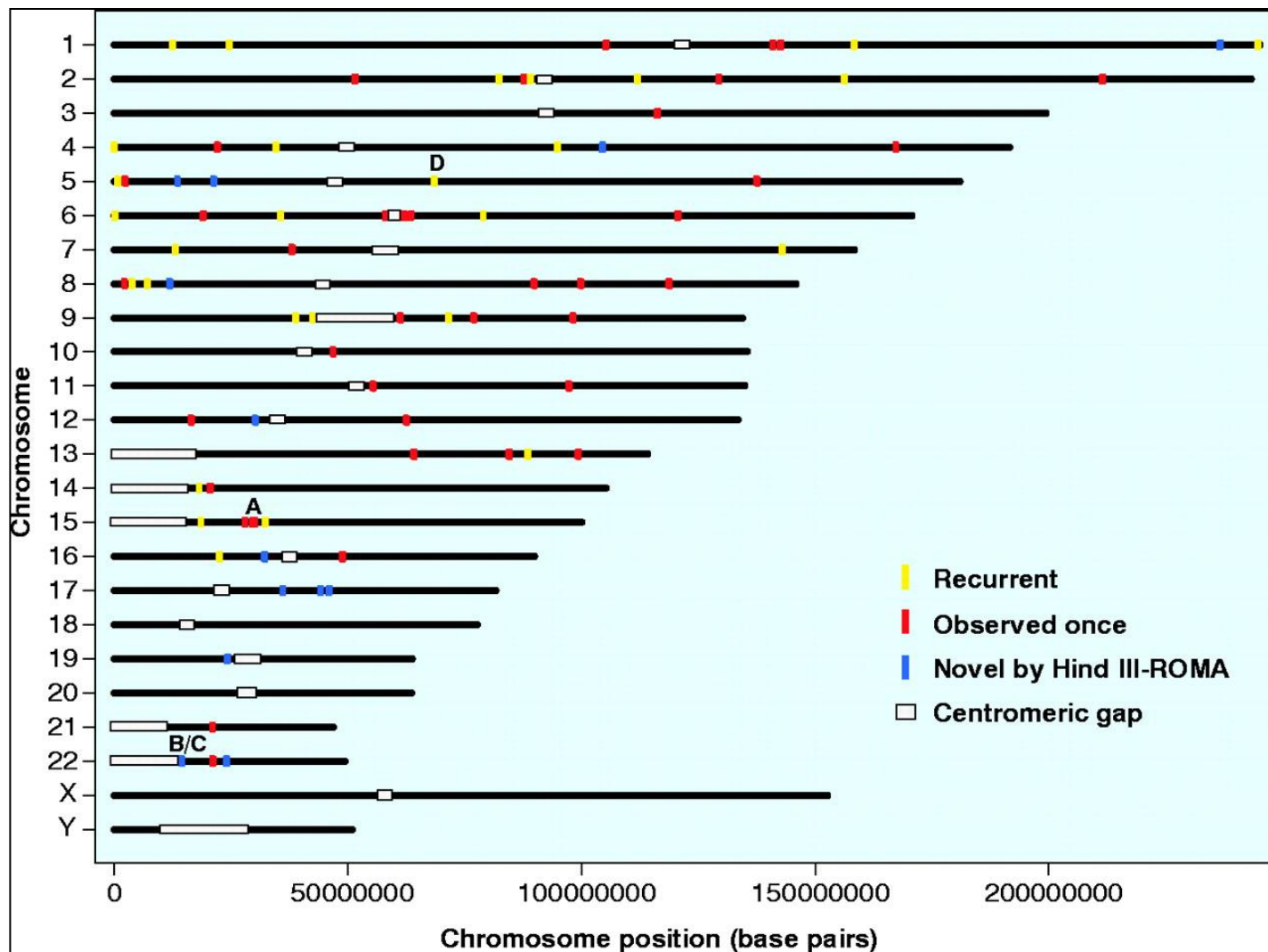


**FENOTIPO MALATTIA**

# Copy Number Polymorphisms (CNPs) in the Human Genome

- Abbiamo visto che molte differenze genetiche tra umani e altri organismi sono il risultato di rimodellamento di grandi regioni di *sintenia*: regioni cromosomiche contigue sul cromosoma di una specie, ma separate o su diversi cromosomi in altre specie.
- Differenze sono anche dovute ad ampie duplicazioni o delezioni di regioni cromosomiche
- È perciò ragionevole aspettarsi che differenze nel numero di copie in segmenti di DNA possano essere una fonte di variazione genetica tra gli umani

# CNPs in Human Genome



*From the analysis of  
20 individuals*

# Single Nucleotide Polymorphisms (SNPs)

- All'estremo opposto rispetto alle variazioni cromosomiche, troviamo le variazioni nucleotidiche, cioè cambiamenti in singoli nucleotidi: in una specifica posizione nel DNA di un individuo potrebbe essere presente una C, in un altro individuo, nella stessa posizione potrebbe essere presente una T
- In genere queste variazioni non sono dannose e si definiscono polimorfismi



Trattandosi di un singolo nucleotide: single nucleotide polymorphism o SNP

# Tipologie comuni di polimorfismi del DNA

## 1. Cambio di un nucleotide (Single Nucleotide Polymorphism, SNP)

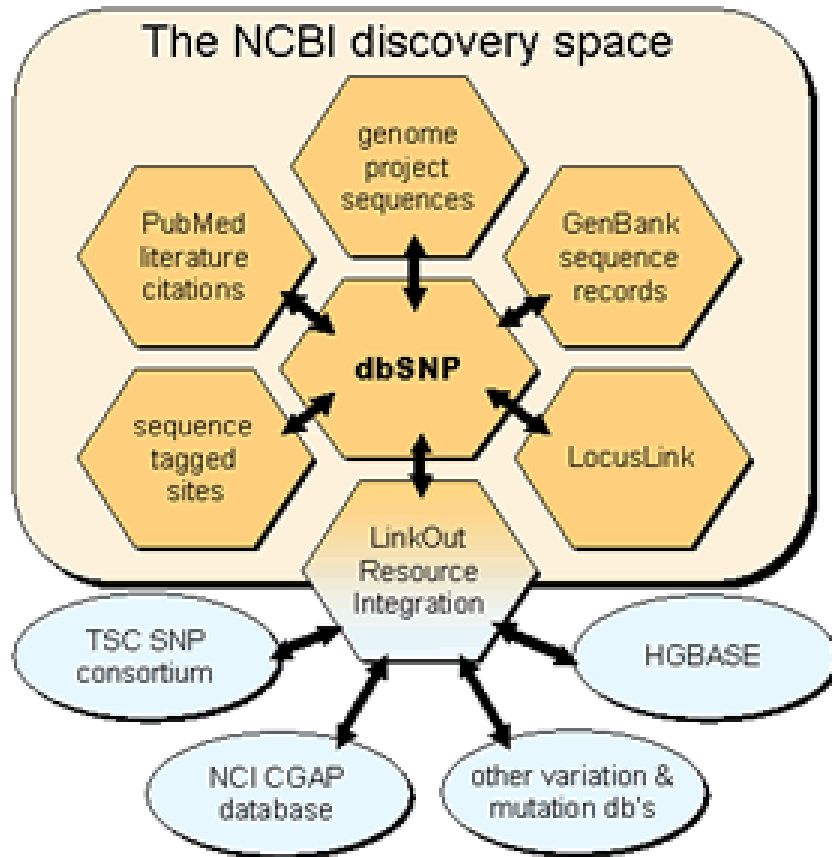
... AAC **C**A ACG CCG CGA GAT ...  
... AAC **A**T ACG CAG CGA GAT ...

## 2. Sequenze ripetute in TANDEM in numero variabile (Short Tandem Repeat, STR)

... AAC **C**A **C**A **C**A **C**A **C**A **C**A **C**A **C**A GTT ... (CA) 7 volte  
... AAC **C**A **C**A **C**A GTT ... (CA) 3 volte



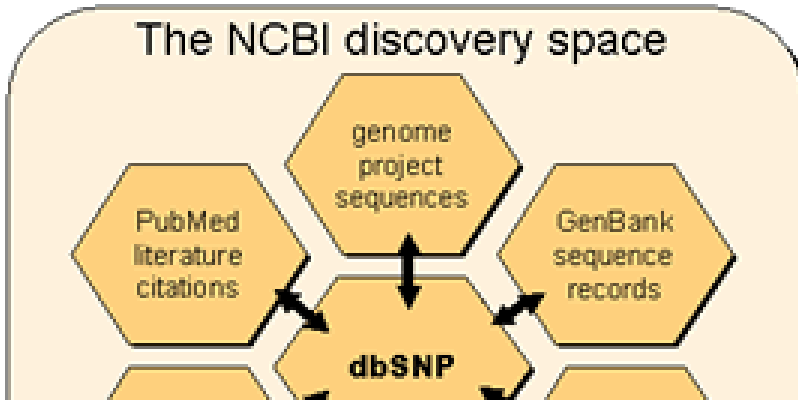
# SNPs: Single Nucleotide Polymorphisms



public  
SNP  
database  
([dbSNP](https://www.ncbi.nlm.nih.gov/snp/)).

<https://www.ncbi.nlm.nih.gov/snp/>

# SNPs: Single Nucleotide Polymorphisms

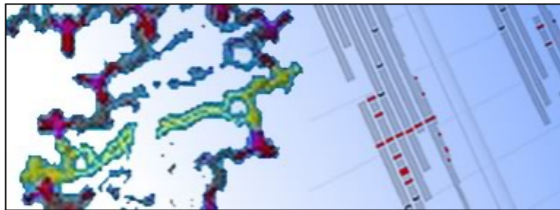


public  
SNP  
database  
([dbSNP](https://www.ncbi.nlm.nih.gov/snp/)).

<https://www.ncbi.nlm.nih.gov/snp/>

NCBI Resources  How To  [Sign in to NCBI](#)

dbSNP    [Advanced](#) [Help](#)



## dbSNP

dbSNP contains human single nucleotide variations, microsatellites, and small-scale insertions and deletions along with publication, population frequency, molecular consequence, and genomic and RefSeq mapping information for both common variations and clinical mutations.

### Getting Started

[dbSNP 20th Anniversary](#)

[Overview of dbSNP](#)

[About Reference SNP \(rs\)](#)

[Factsheet](#)

[Entrez Updates \(February 5, 2020\)](#)

### Submission

[How to Submit](#)

[Hold Until Published \(HUP\) Policies](#)

[Submission Search](#)

### Access Data

[Variation Services API](#)

[FTP Download](#)

[Tutorials on GitHub](#)

# SNPs: NCBI dbSNP

È possibile riconoscere tutti gli SNPs associati a ciascun gene o regioni intergeniche. Qui ad esempio *CDH4*

NCBI Resources How To Sign in to NCBI  
Variation Viewer Homo sapiens: GRCh38.p12 (GCF\_000001405.38) Chr 20 (NC\_000020.11): 60,770,692 - 62,422,351  
New to Variation Viewer? Read our quick overview! X  
Region: CDH4 Gene NM\_001794.5 Transcript  
NC\_000020.11  
Genes, NCBI Homo sapiens Annotation Release 109.20200228  
CDH4  
CDH4  
CDH1

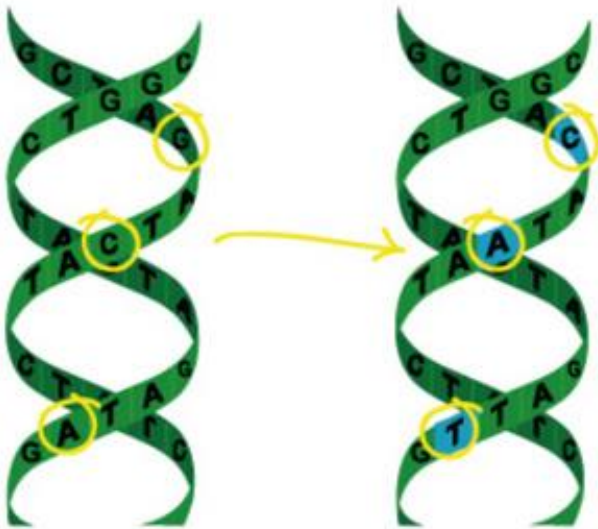
Variant ID	Location	Variant type	Gene	Molecular consequences	Most severe clinical significance	1000G MAF
rs951612025	60,770,693	single nucleotide variant	LOC105372699 and 1 more	intron variant	Not-Provided	
rs1297612170	60,770,696	single nucleotide variant	LOC105372699 and 1 more	intron variant	Not-Provided	
rs148944055	60,770,698	single nucleotide variant	LOC105372699 and 1 more	intron variant	Not-Provided	T = 0.00499201
rs1245225583	60,770,758	single nucleotide variant	LOC105372699 and 1 more	intron variant	Not-Provided	
rs534902289	60,770,763	single nucleotide variant	LOC105372699 and 2 more	intron variant	Not-Provided	A = 0.000199681
rs553149525	60,770,765	single nucleotide variant	LOC105372699 and 1 more	intron variant	Not-Provided	C = 0.000599042
rs6027811	60,770,788	single nucleotide variant	LOC105372699 and 1 more	intron variant	Not-Provided	A = 0.496406
rs1398227139	60,770,790	single nucleotide variant	LOC105372699 and 1 more	intron variant	Not-Provided	



<http://www.ncbi.nlm.nih.gov/SNP/index.html>

<https://www.ncbi.nlm.nih.gov/variation/view/>

# SNPs: quanti sono ?



Sono ad oggi stati identificati oltre 150 milioni di polimorfismi nella popolazione umana, ovvero 1 polimorfismo ogni 20 nucleotidi circa

<http://www.ncbi.nlm.nih.gov/SNP/index.html>

Sherry ST, et al. "dbSNP: the NCBI database of genetic variation." *Nucleic Acids Research*. 2001; 29:308-311.

Chromosome	Base pairs	Variations	Frequency
1	248.956.422	12.151.146	20
2	242.193.529	12.945.965	19
3	198.295.559	10.638.715	19
4	190.214.555	10.165.685	19
5	181.538.259	9.519.995	19
6	170.805.979	9.130.476	19
7	159.345.973	8.613.298	18
8	145.138.636	8.221.520	18
9	138.394.717	6.590.811	21
10	133.797.422	7.223.944	19
11	135.086.622	7.535.370	18
12	133.275.309	7.228.129	18
13	114.364.328	5.082.574	23
14	107.043.718	4.865.950	22
15	101.991.189	4.515.076	23
16	90.338.345	5.101.702	18
17	83.257.441	4.614.972	18
18	80.373.285	4.035.966	20
19	58.617.616	3.858.269	15
20	64.444.167	3.439.621	19
21	46.709.983	2.049.697	23
22	50.818.468	2.135.311	24
X	156.040.895	5.753.881	27
Y	57.227.415	211.643	270
mtDNA	16.569	929	18
<b>total</b>	<b>3.088.286.401</b>	<b>155.630.645</b>	<b>20</b>

# Short Tandem Repeats (STRs) o Microsatelliti

Microsatelliti sono corti segmenti di DNA caratterizzati dalla presenza di sequenze ripetute, quali CACACACA.

In ciascun allele, la unità ripetuta (CA) può presentare un numero variabile di ripetizioni

- Gli *STR (Short Tandem Repeats) o microsatelliti* presentano perciò alleli multipli: ad esempio

**A<sub>1</sub>:** ... TACAGC CA CA CA CA GCACTT ...

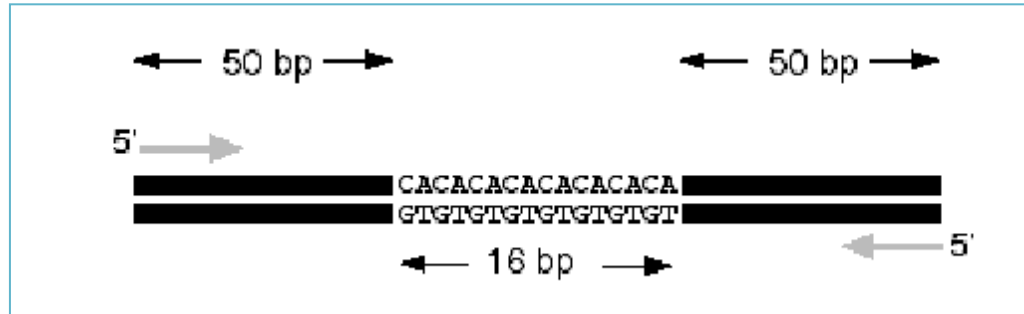
**A<sub>2</sub>:** ... TACAGC CA CA CA CA CA GCACTT ...

**A<sub>3</sub>:** ... TACAGC CA CA CA CA CA CA GCACTT ...

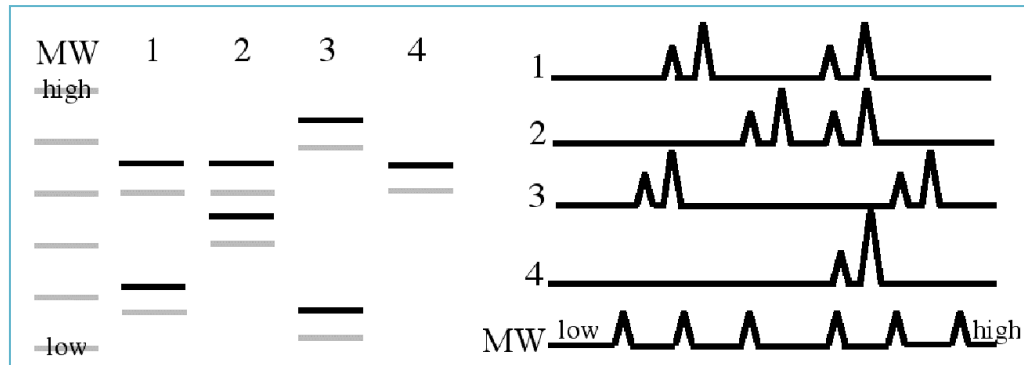
**A<sub>4</sub>:** ... TACAGC CA CA CA CA CA CA CA GCACTT ...

# Short Tandem Repeats (STRs) or Microsatellites

I microsatelliti vengono rilevati analisi PCR



I prodotti della PCR sono separati mediante elettroforesi su gel capillare



È possibile valutare la dimensione dei prodotti PCR e quindi quante volte il dinucleotide "CA" è stato ripetuto per ciascun allele.

*Spesso si vedono bande minori in aggiunta alle bande maggiori, chiamati bande «balbuzienti»*

# Differenze tra SNP e STR: Frequenze alleliche

- Gli *SNP* (*Single Nucleotide Polymorphisms*) presentano generalmente solo due alleli nella popolazione: ad esempio

**A1:** ... AAC **CA** ACG ...      **Frequenza Allelica = 40%**

**A2:** ... AAC **TA** ACG ...      **Frequenza Allelica = 60%**

- Gli *STR* (*Short Tandem Repeats*) o *microsatelliti* presentano alleli multipli: ad esempio

**A1:** ... AAC **CA CA CA** GTT ...      **FA = 30%**

**A2:** ... AAC **CA CA CA CA** GTT ...      **FA = 30%**

**A3:** ... AAC **CA CA CA CA CA** GTT ...      **FA = 25%**

**A4:** ... AAC **CA CA CA CA CA CA** GTT ...      **FA = 15%**

# Differenze tra SNP e STR: Frequenze genotipiche

## ALLELI

## GENOTIPI

### ■ SNPs

A1: 40%

A2: 60%

	A1	A2
A1	0,16	0,24
A2	0,24	0,36

52% omozigoti

48% eterozigoti

### ■ Microsatelliti

A1: 30%

A2: 30%

A3: 25%

A4: 15%

	A1	A2	A3	A4
A1	9%	9%	8%	5%
A2	9%	9%	8%	5%
A3	8%	8%	6%	4%
A4	5%	5%	4%	2%

26% omozigoti

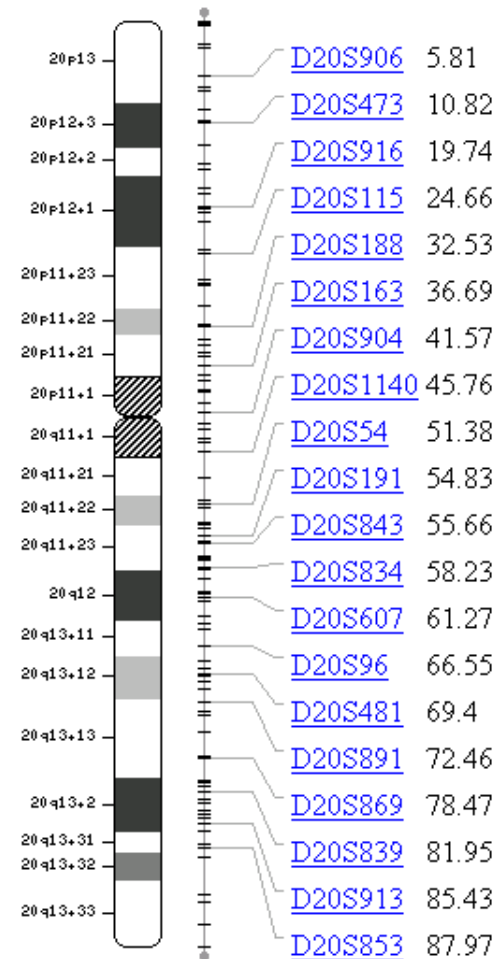
74% eterozigoti

*Per il loro multi-allelismo, i microsatelliti hanno alto tasso di eterozigosi nella popolazione*



# STR sono marcatori genetici

- Gli STR sono altamente polimorfici per la presenza di multipli alleli
- Hanno una posizione nota nel genoma
- Sono facilmente tipizzabili e stabili
- Questi marcatori polimorfici sono stati usati per ottenere la prima mappa genetica del genoma umano

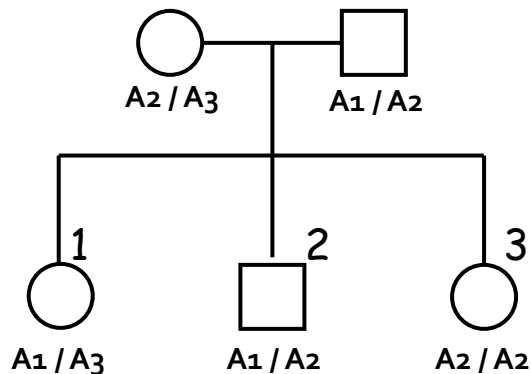


# Analisi familiare

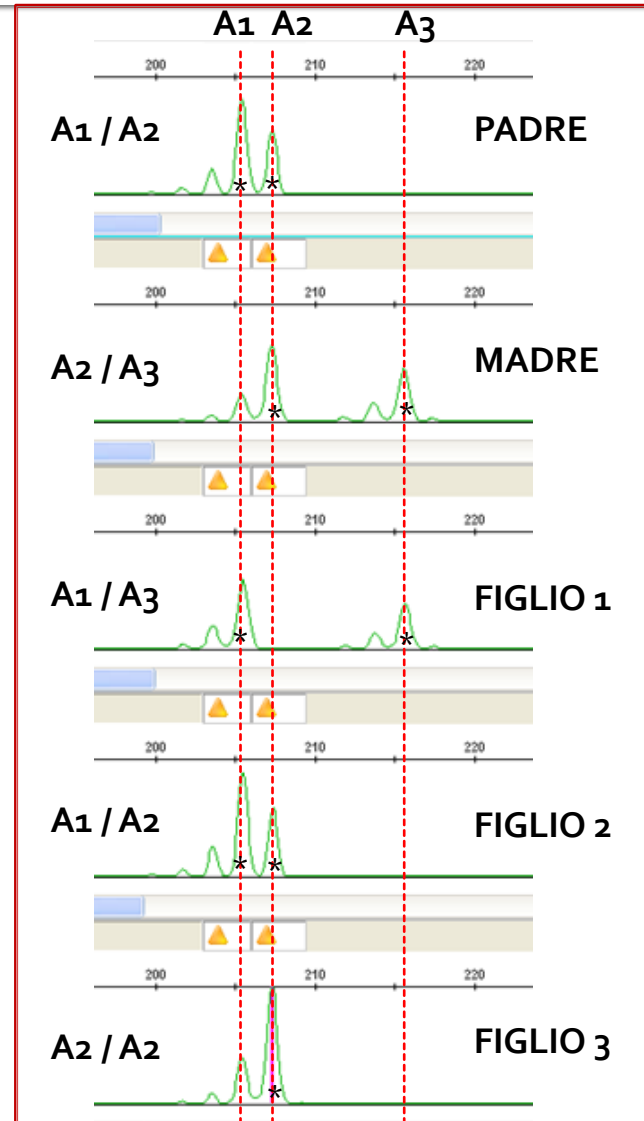
Sebbene nella popolazione i microsatelliti presentino multipli alleli,

Di questo pool allelico, ogni individuo ne presenta ovviamente non più di due

Impiegati per analisi di pedigree familiari



- Analisi di paternità
- Analisi di parentela



# Polimorfismi del DNA impiegati nella tipizzazione di reperti in indagini criminali

Le analisi forensi del DNA fanno uso di polimorfismi del DNA per la tipizzazione del DNA e vengono utilizzate per:

- identificare possibili stupratori e altri criminali;

Nell'esempio a fianco, sono stati utilizzati due marcatori genetici: uno che rivela le bande alleliche in alto, l'altro bande alleliche in basso.

1. (VICTIM) DNA della vittima stessa;
2. (EVIDENCE n. 1) DNA dallo sperma rimosso dalla vagina della vittima;
3. (EVIDENCE n. 2) DNA della macchia di seme lasciato sui vestiti della vittima;
4. (SUSPECT # 1) DNA del sospetto 1;
5. (SUSPECT # 2) DNA del sospetto 2;
6. (MARKER) un insieme di frammenti di DNA di lunghezza nota.
7. (CONTROL) DNA di una persona precedentemente testata per assicurarsi che le sonde funzionino correttamente.

In base a questo test, il sospetto n. 2 può essere chiaramente escluso.

Ma il sospetto n. 1 è colpevole?

- La probabilità della combinazione allelica osservata nel genotipo dipende dalle frequenze alleliche nella popolazione

Ad esempio:

Frequenze Alleliche: A1= 25%; A2= 20%

Frequenza genotipo =0.20%

Frequenze Alleliche: B1= 10%; B2= 40%

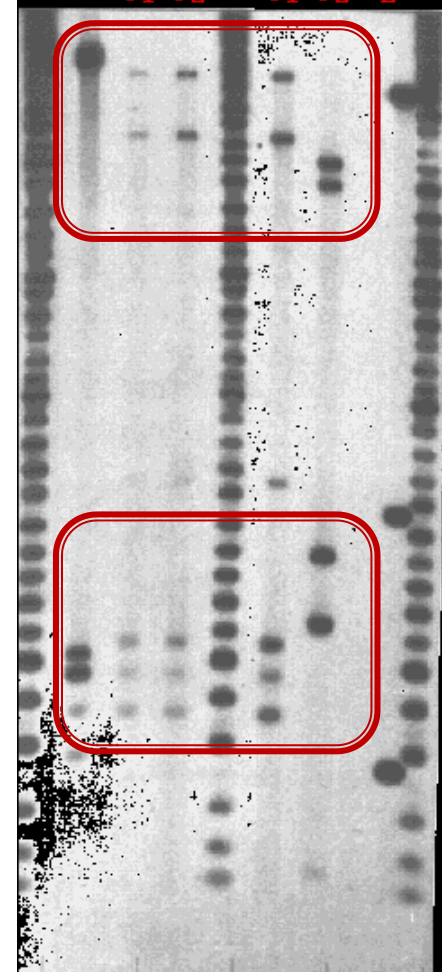
- La probabilità di abbinamento casuale: si ottiene moltiplicando

Frequenza Genotipo Vittima	0.20% × 0.20%
	0.0004%

Frequenza Genotipo Sospetto 1
-------------------------------

1 possibilità su 2,500

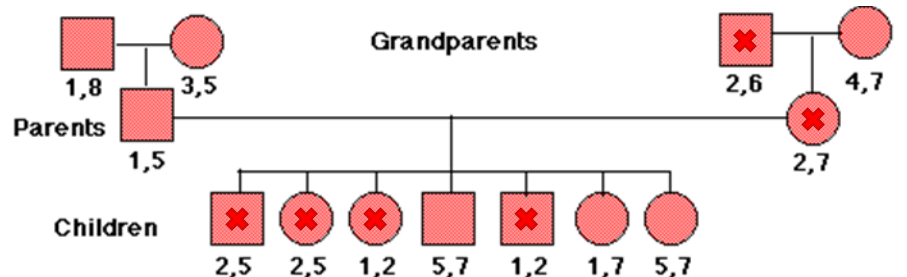
M  
A  
R  
K  
E  
R  
  
V  
I  
C  
T  
I  
M  
  
E  
V  
I  
D  
E  
N  
C  
E  
#1  
#2  
  
M  
A  
R  
K  
E  
R  
  
S  
U  
S  
P  
E  
C  
T  
#1  
#2  
  
S  
U  
S  
P  
E  
C  
T  
  
C  
O  
N  
T  
R  
O  
L



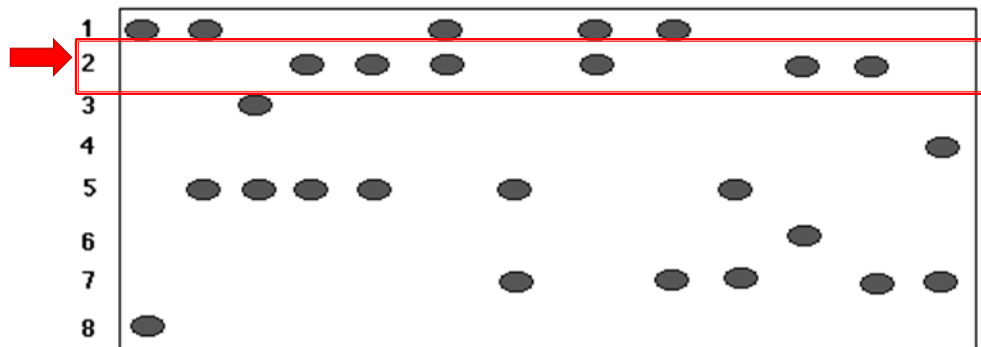
# Un polimorfismo del DNA può funzionare come surrogato di un locus malattia

Esistono malattie genetiche per le quali non è stato ancora scoperto il gene malattia. Fino a quando il gene non è noto, non è possibile effettuare test diretti.

Un "marcatore" genetico può servire da sostituto del gene stesso della malattia.



Esiste un marcatore con 8 diversi alleli nella popolazione: tutti coloro che hanno ereditato l'allele-2 hanno anche un certo disordine ereditario e nessuno a cui manca l'allele-2 ha il disordine >>> deduciamo che il gene per la malattia è strettamente legato a questo allele



× inherited disorder

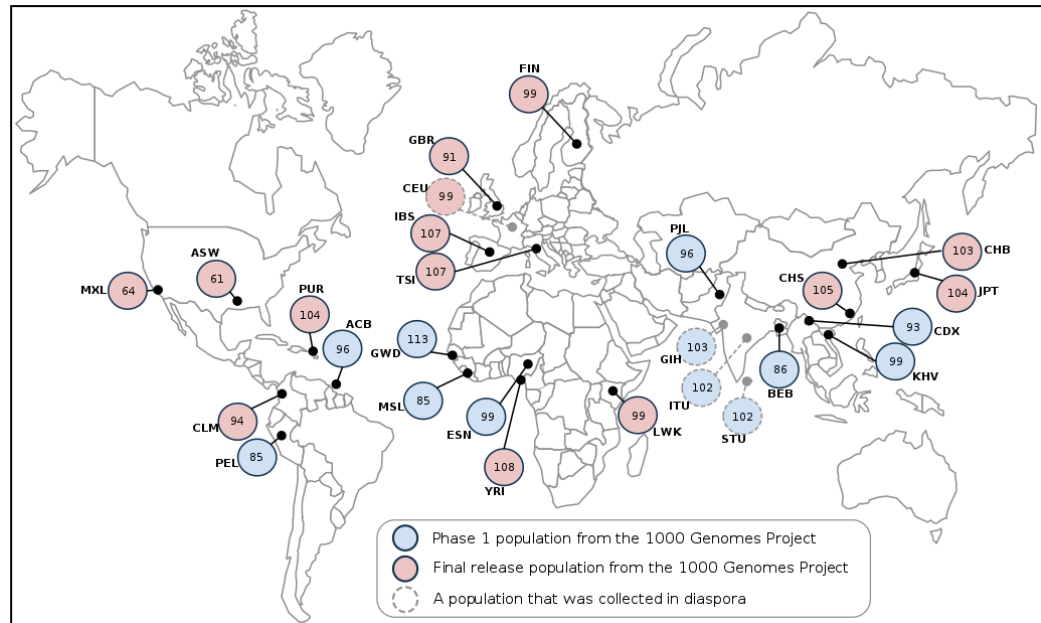
>> queste informazioni possono essere utilizzate per riconoscere la trasmissione della malattia in questa famiglia

>> queste informazioni possono essere utilizzate per identificare il gene della malattia, che deve essere localizzato a breve distanza genetica dal marcatore (nessun *crossing-over*)

# 1000 Genomes Project

- 1000 Genomes Project è uno studio internazionale con l'obiettivo principale di creare un catalogo completo e dettagliato delle variazioni genetiche presenti nel genoma umano proveniente da diverse popolazioni e aree geografiche.
- Scoprire > 95% delle varianti (ad esempio SNP, CNV, indels)
- Definire le frequenze degli alleli
- Identificare eventuali fenomeni di «linkage disequilibrium» nelle diverse popolazioni mondiali

- Utilizzabile per studi di associazione relativi a variazioni genetiche nelle malattie dell'uomo.

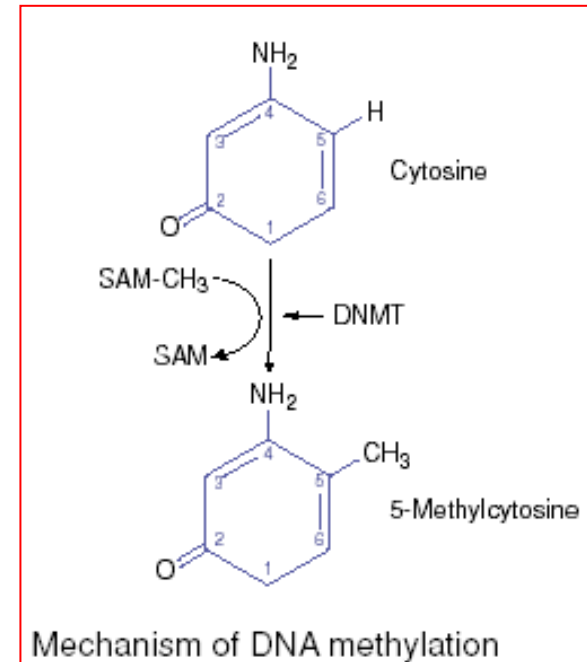


Origine dei genomi sequenziati



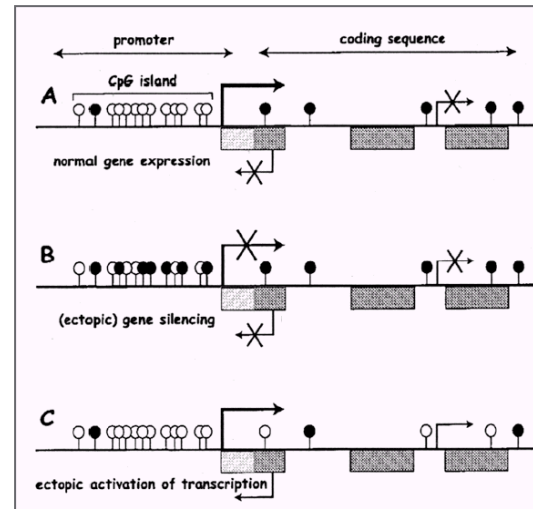
# DNA Methylation: epigenetica del genoma umano

- Nelle cellule di mammifero, le citosine presenti nel dinucleotide CpG sono generalmente metilate tramite l'aggiunta di un gruppo metile (-CH<sub>3</sub>) da parte di un enzima (DNA metiltransferasi, DNMT) una volta incorporati nel DNA
- Eccezione sono le isole CpG, regioni dai 500 ai 2000 nucleotidi arricchiti a lungo in dinucleotidi CpG che si trovano in circa il 60% dei promotori genici.
- Le isole CpG dei promotori genici di solito non sono metilate

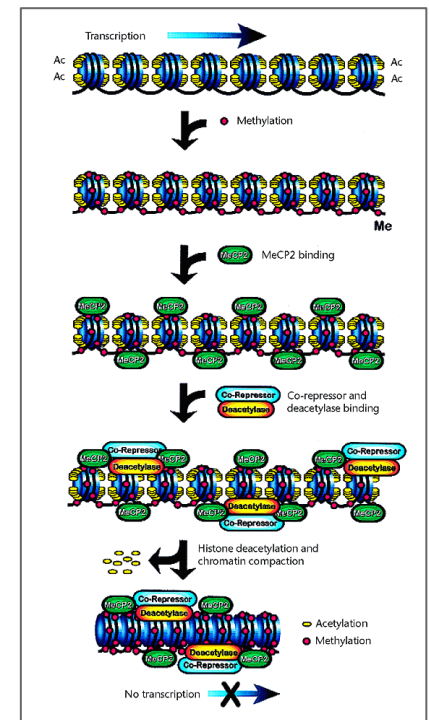


# DNA Methylation: repressione della trascrizione

- La Metilazione del DNA dei promotori genici contenenti *isole CpG* induce la repressione della trascrizione genica



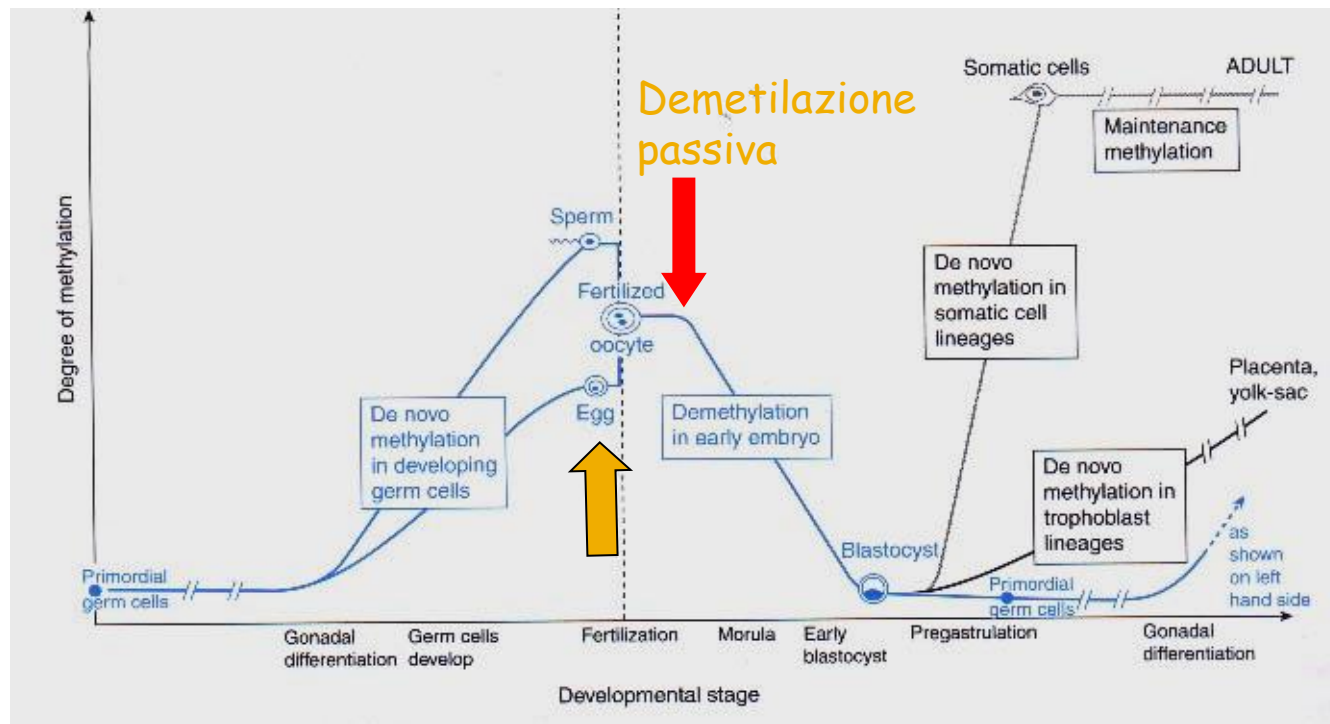
- Il DNA metilato recluta iston deacetilasi (HDAC) e reprime la trascrizione genica





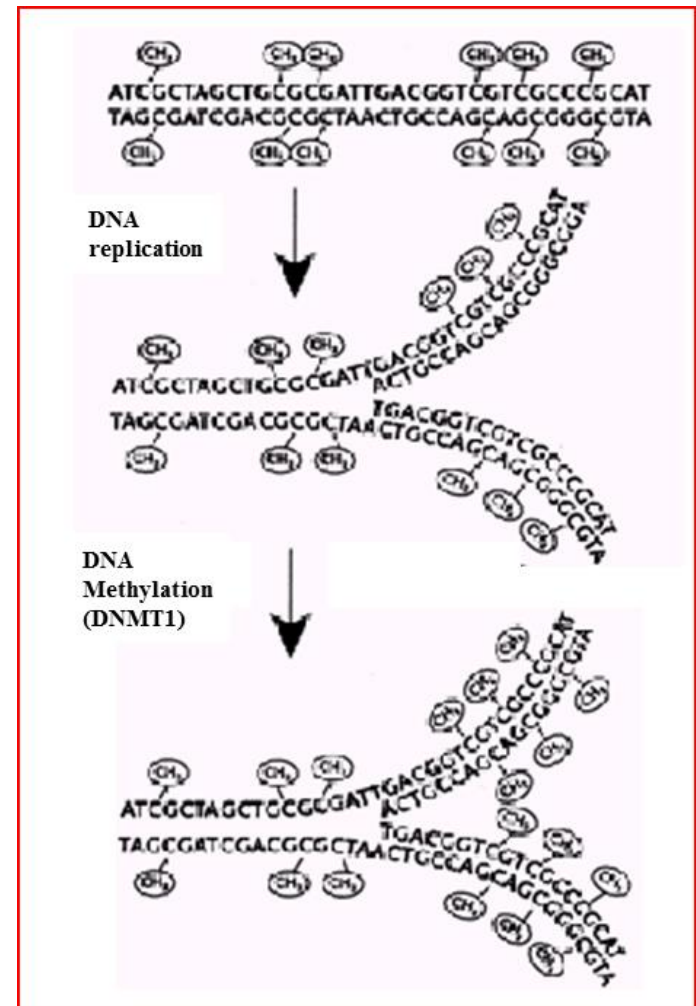
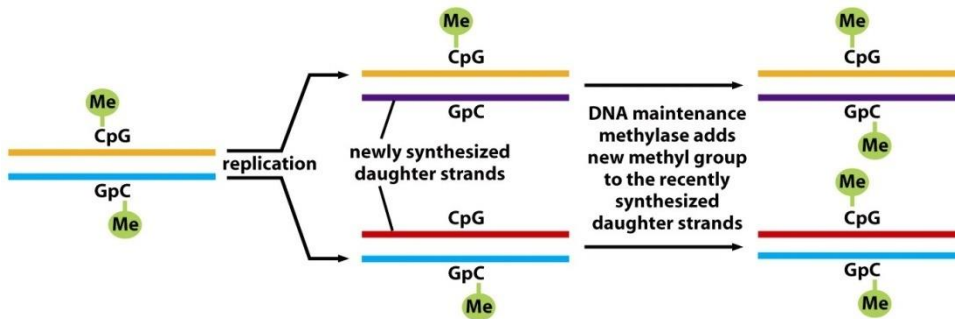
# DNA Methylation prime fasi dello sviluppo

Metilazione del DNA nello sviluppo dei mammiferi



# DNA Methylation: mantenimento e trasmissione

- Una volta fissato, come viene trasmesso/ereditato stabilmente di cellula in cellula lo stato di metilazione del DNA?
- È possibile per la natura simmetrica delle sequenze modificate (CpG)
- La preferenza della DNA metiltransferasi di "mantenimento" DNMT1 per il DNA emimetilato



# DNA Methylation: importanza

- Che rilevanza ha il mantenimento dello stato di metilazione nel genoma umano?
  - Silenziamento elementi trasponibili: necessario al mantenimento stabilità del genoma
  - Semplificazione funzionale del genoma: apertura delle sole regioni dei promotori genici per un corretto inizio della trascrizione e chiusura delle altre regioni.
  - Coordinamento con i sistemi di riparo del DNA per la correzione dello strand giusto in caso di errori nell'incorporazione nucleotidica alla duplicazione del DNA

# Genoma umano: in sintesi

- Come sia costituito il genoma umano: quali siano i numeri caratterizzanti
- Quali siano gli elementi codificanti, quali no
- Quali le inconsistenze ancora presenti
  
- Quali siano i database pubblici che ci consentono di acquisire informazioni su specifici geni o regioni cromosomiche

- Abbiamo imparato cosa siano i polimorfismi
- Che tipo di polimorfismi esistono nel genoma umano
  
- Come possano essere utilmente impiegati per lo svolgimento di analisi genetiche di parentela o in ambito forense o per attività di ricerca

- Abbiamo imparato cosa sia la metilazione del DNA
  
- Quale il suo significato biologico e funzionale sia nella regolazione dell'espressione genica che del mantenimento della stabilità genomica

***FINE***