

- 
- **STUDI DI ASSOCIAZIONE E METODICHE INNOVATIVE E**
 - **GENOME WIDE SCANNING (GWAS)**



MENDEL

1865
Leggi
ereditarietà



Alfred Hershey e Martha Chase

1952
L'informazione
genetica è
contenuta in
DNA

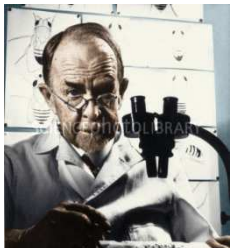


Frederick Sanger

1975
Sequenziamento
del DNA

1927
Applicazione
delle leggi di
Mendel

1953
Modello del
DNA



Thomas Hunt Morgan



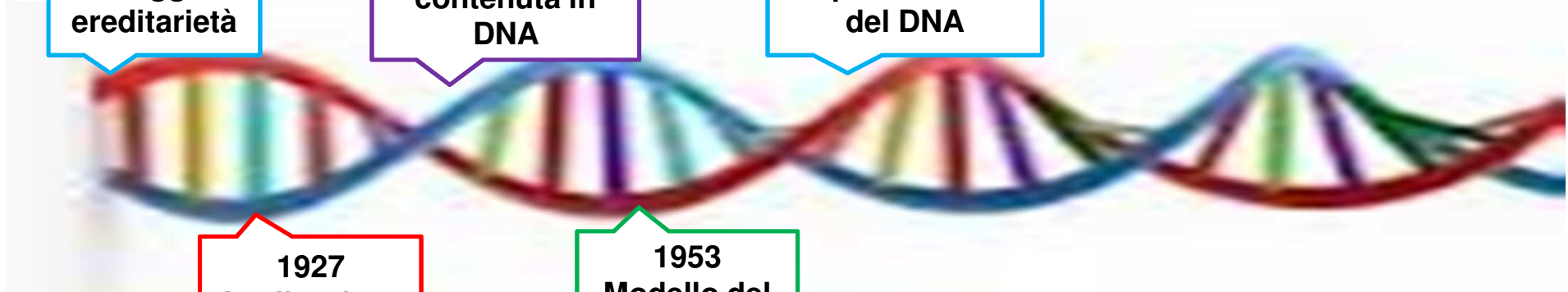
Hermann Joseph Muller



Alfred Henry Sturtevant



Watson & Crick





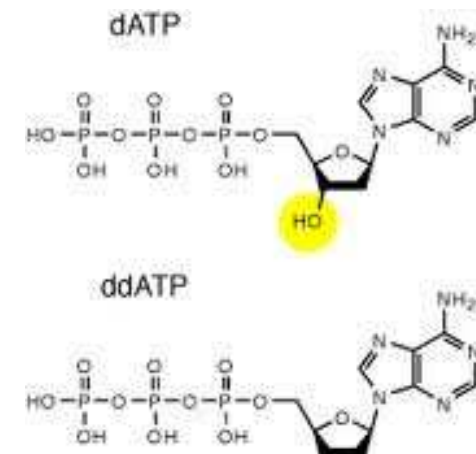
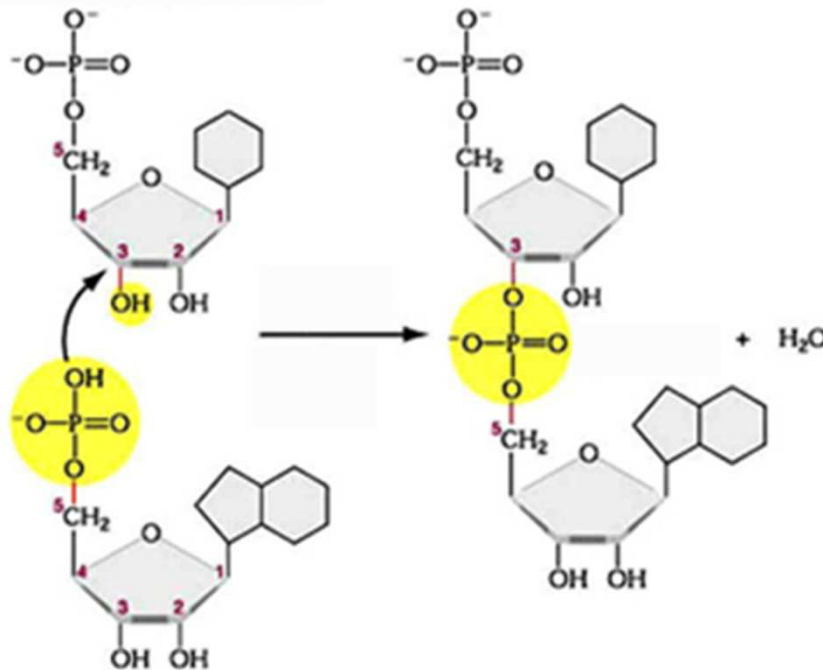
METODICHE DI ANALISI

- SEQUENZIAMENTO SECONDO SANGER
- PROGETTO GENOMA UMANA
- SEQUENZIAMENTO NGS (NEXT GENERATION SEQUENCING)

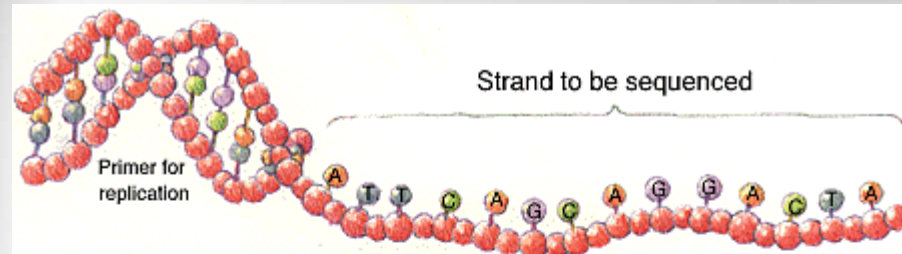
SEQUENZIAMENTO SANGER

Utilizzo:

- primer
- DNA polimerasi
- deossinucleotidi
- nucleotidi modificati
(dideossitrifosfato, ddNTPs)
per interrompere la reazione di sintesi

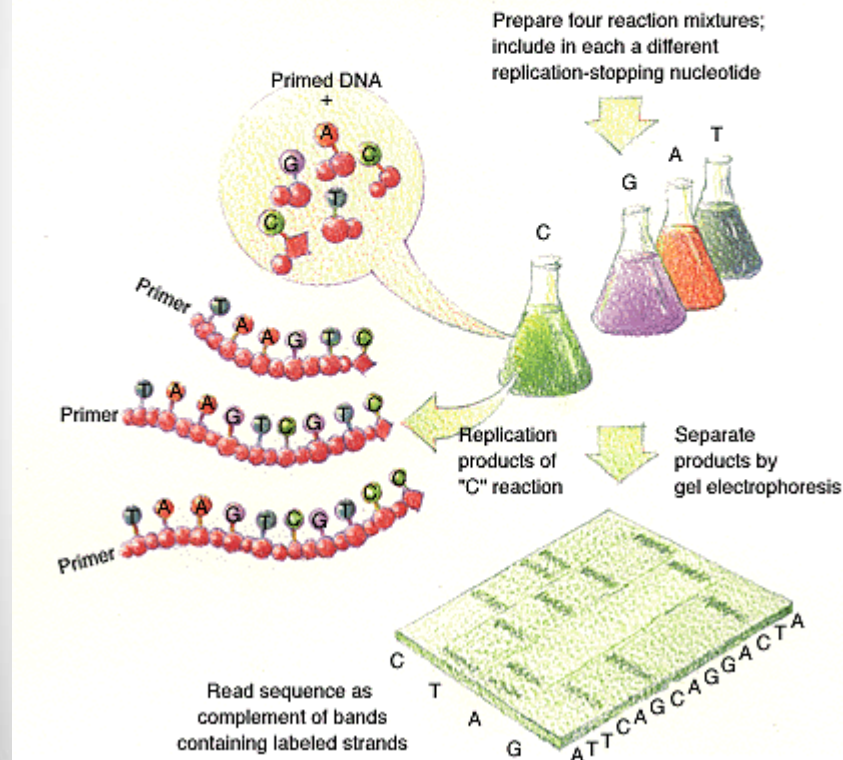


SEQUENZIAMENTO SANGER

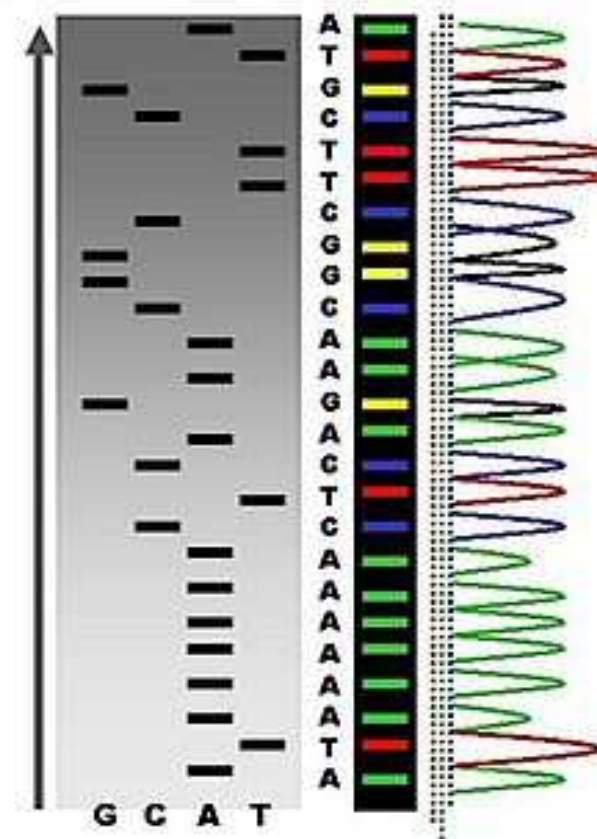


Utilizzo:

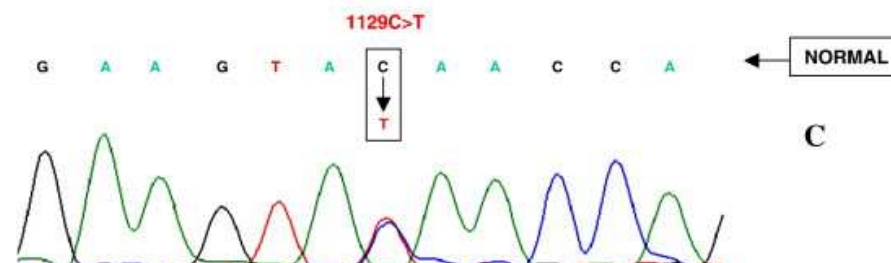
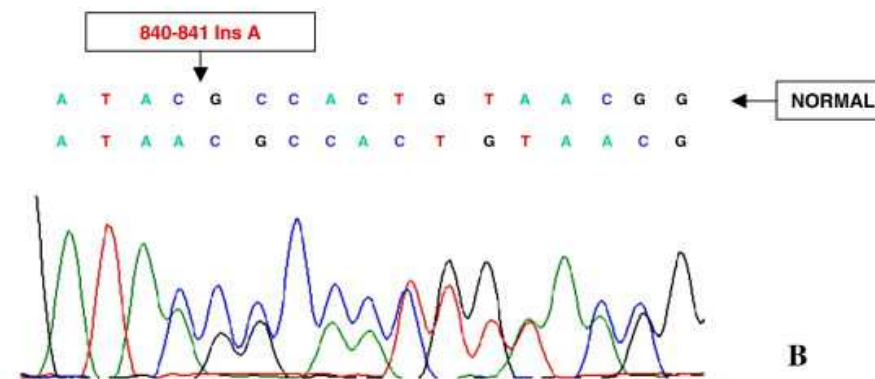
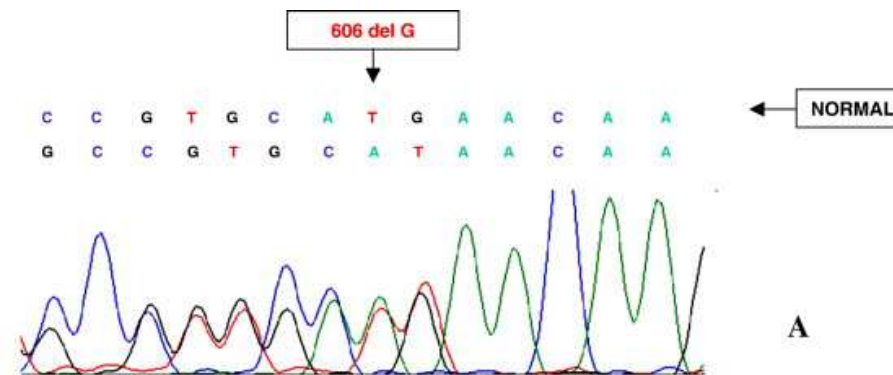
- primer
- DNA polimerasi
- deossinucleotidi
- nucleotidi modificati
(dideossitrifosfato, ddNTPs)
per interrompere la reazione di sintesi



SEQUENZIAMENTO SANGER



Esempi di cromatogrammi





MENDEL

1865
Leggi
ereditarietà



Alfred Hershey e Martha Chase

1952
L'informazio
ne genetica
è contenuta
in DNA



Frederick Sanger

1975
Sequenziamento
del DNA



1927
Applicazione
delle leggi di
Mendel



Thomas Hunt
Morgan



Hermann Joseph
Muller



Alfred Henry
Sturtevant

1953
Modello del
DNA



Watson & Crick

1990 - 2001
progetto
genoma
umano



Francis
Collins



Craig
Venter



PROGETTO GENOMA UMANO

SCOPI

- La disponibilità della sequenza rende più semplice l'identificazione dei geni responsabili delle malattie mendeliane
- Sequenza completa di tutti i geni
- Possibilità di determinare la struttura esoni introni
- Rivelare le regioni di controllo non codificanti
- Identificare polimorfismi

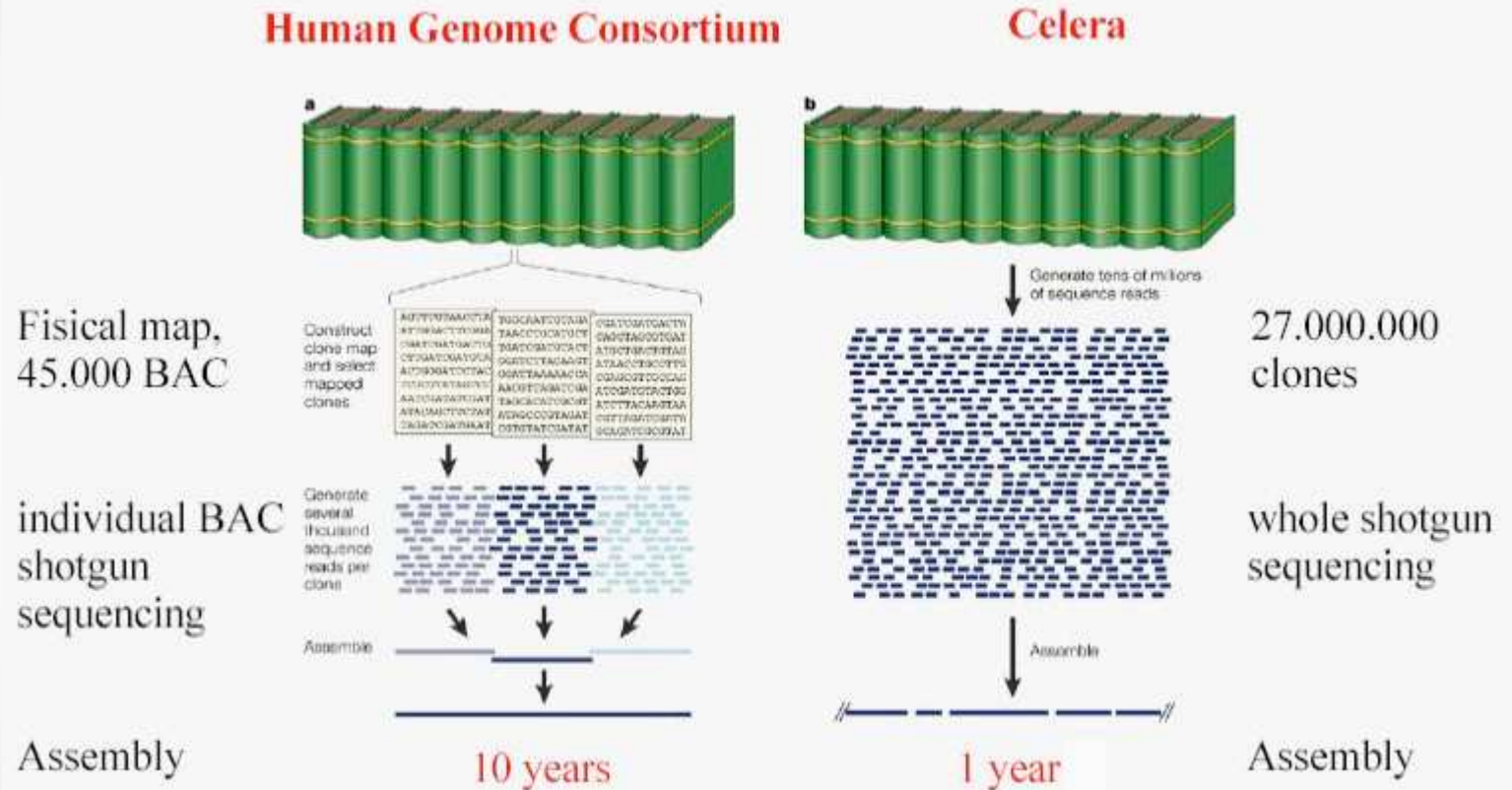


PROGETTO GENOMA UMANO

LE TAPPE

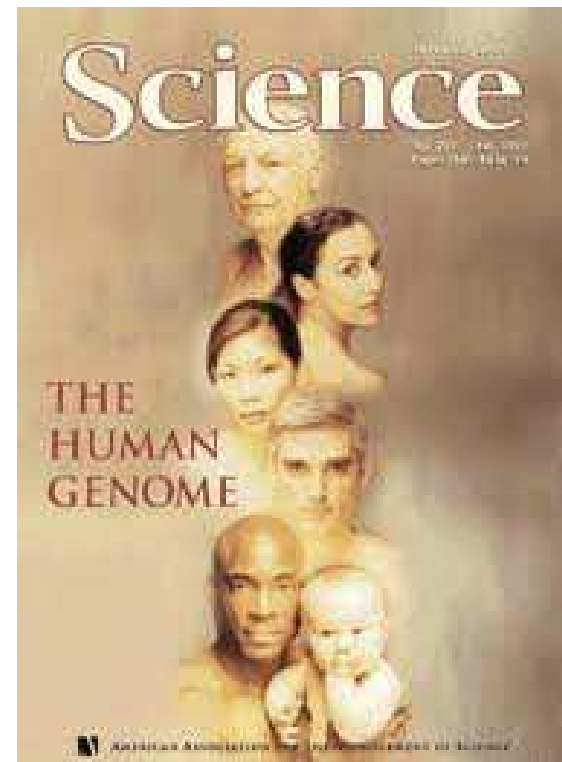
- 1990 Negli Stati Uniti nasce ufficialmente lo Human Genome Project (HGP), sotto la guida di James Watson. Negli anni successivi Regno Unito, Giappone, Francia, Germania, Cina si uniscono al progetto formando un consorzio pubblico internazionale. In Italia il progetto genoma nasce nel 1987 ma si interrompe nel 1995.
- 1992 Craig Venter lascia l'NIH e il progetto pubblico. Fonderà una compagnia privata, la Celera Genomics, portando avanti un progetto genoma parallelo.
- 1993 Francis Collins e John Sulston diventano direttori rispettivamente del National Human Genome Research Center negli USA e del Sanger Center in Inghilterra, i 2 principali centri coinvolti nel HGP.
- 1999 (Dicembre) Pubblicata su Nature la sequenza completa del cromosoma 22.
- 2000 (Maggio) pubblicata su Nature la sequenza completa del cromosoma 21.
- 2000 (Giugno) Francis Collins e Craig Venter annunciano congiuntamente di aver completato la "bozza" del genoma Umano.

PROGETTO GENOMA UMANO STRATEGIE



PROGETTO GENOMA UMANO

- 2001 La bozza completa del genoma umano (che gli inglesi chiamano working draft) è pubblicata su Nature (quella del consorzio pubblico) e su Science (quella della Celera).





PROGETTO GENOMA UMANO IN NUMERI

20 CENTRI DI SEQUENZIAMENTO IN 6 PAESI:
USA, UK, GIAPPONE, FRANCIA, GERMANIA, CINA

DURATA: 13 ANNI

COSTO TOTALE 2,7 MILIARDI DI DOLLARI

Il Genoma Umano in numeri

- **23 paia di cromosomi**
- **3.000.000.000 paia di basi**
- **30.000-40.000 geni**
- **Non si conosce la funzione di circa la meta' dei geni scoperti**
- **Meno del 2% del genoma codifica per proteine**
- **Piu' del 50% del genoma e' costituito da sequenze ripetute con funzione sconosciuta**



MENDEL

1865
Leggi
ereditarietà



Alfred Hershey e Martha Chase

1952
L'informazio
ne genetica
è contenuta
in DNA



Frederick Sanger

1975
Sequenziamento
del DNA

2005 – oggi
NEXT
GENERATION
SEQUENCING

1927
Applicazione
delle leggi di
Mendel



Thomas Hunt
Morgan



Hermann Joseph
Muller



Alfred Henry
Sturtevant

1953
Modello del
DNA



Watson & Crick

1990 - 2001
progetto
genoma
umano



Francis
Collins



Craig
Venter



SEQUENZIAMENTO DI NUOVA GENERAZIONE NEXT GENERATION SEQUENCING

Vantaggi delle piattaforme di nuova generazione

- Rivoluzionaria diminuzione del costo e del tempo per generare dati di sequenza (lavorano in multi-parallelo)
 - 10 Giorni per sequenziare il genoma di un individuo
- Richiesta meno robotica nelle fasi precedenti al caricamento sul sequenziatore
- Eccezionale risoluzione per molti tipi di esperimenti (es. analisi di espressione, sequenziamento di DNA immunoprecipitato e di micro RNA, analisi di medie/grandi inserzioni-delezioni nei genomi....)

NEXT GENERATION SEQUENCING

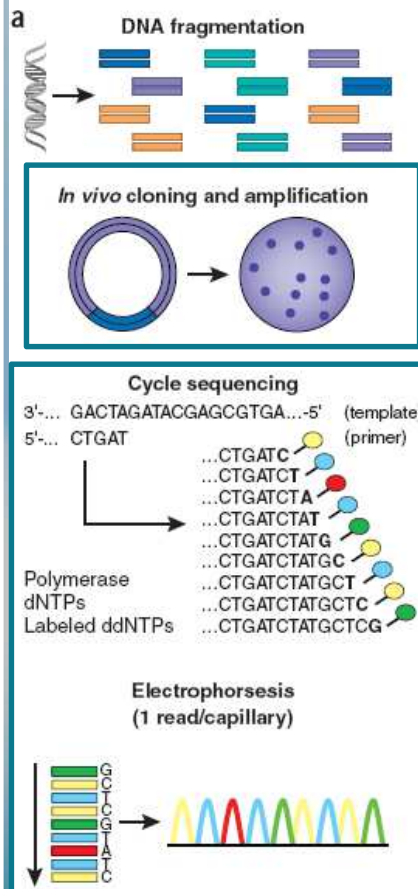


ion torrent
by life technologies™

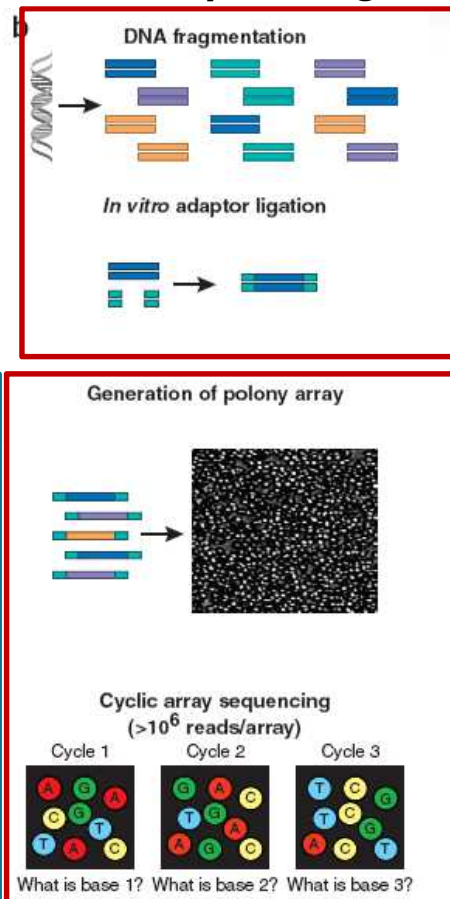


SANGER vs NEXT GENERATION SEQUENCING

Sanger sequencing



NGS sequencing



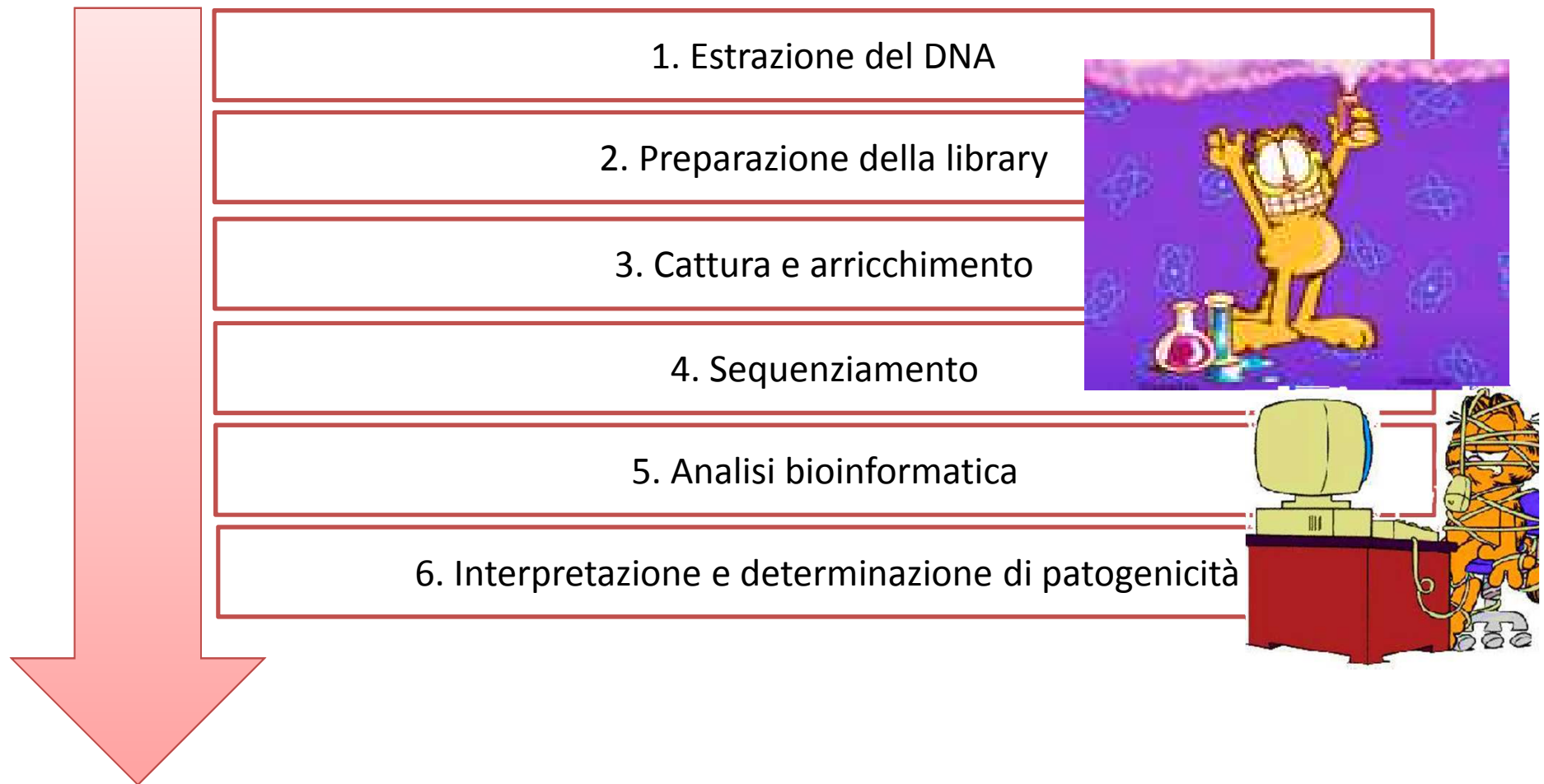
Advantages of NGS:
Construction of a sequencing library
clonal amplification to generate sequencing features

No in vivo cloning,
transformation, colony picking...

Array-based sequencing

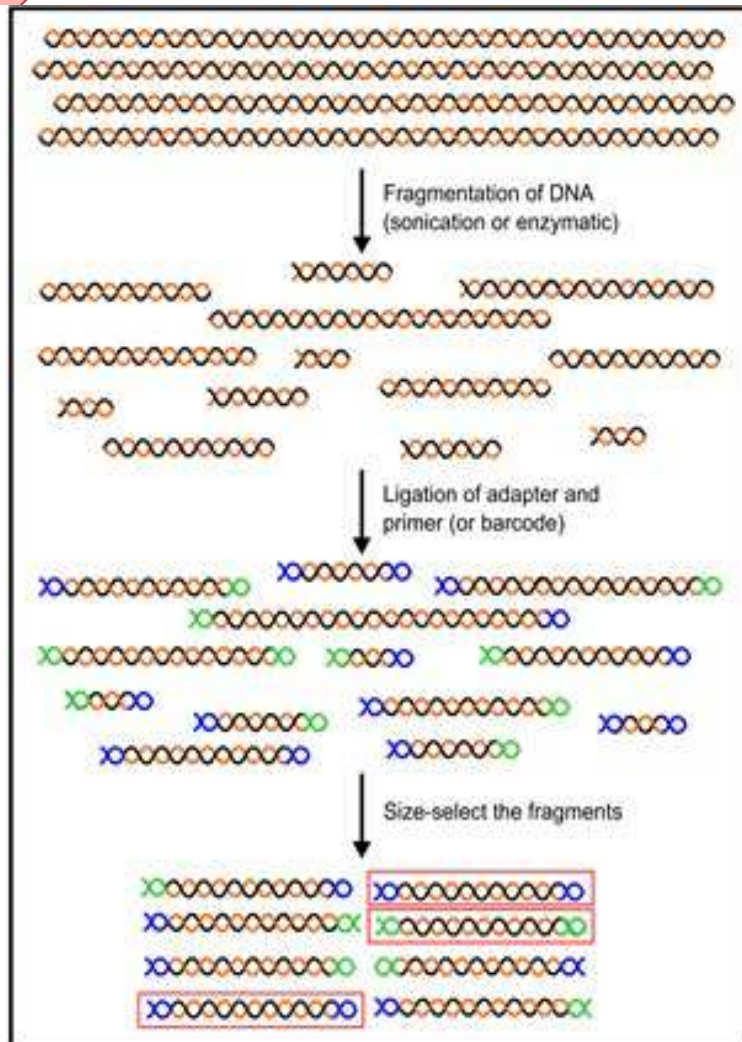
Higher degree of parallelism
than capillary-based sequencing

WORKFLOW del SEQUENZIAMENTO



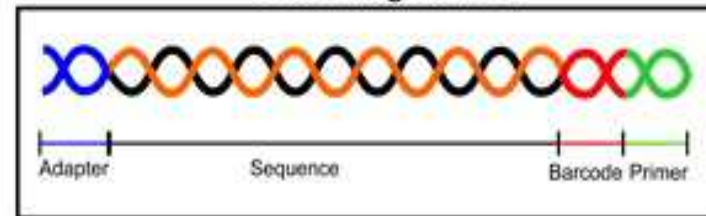
1. Estrazione del DNA

2. Preparazione della library



Library preparation

Good fragments :

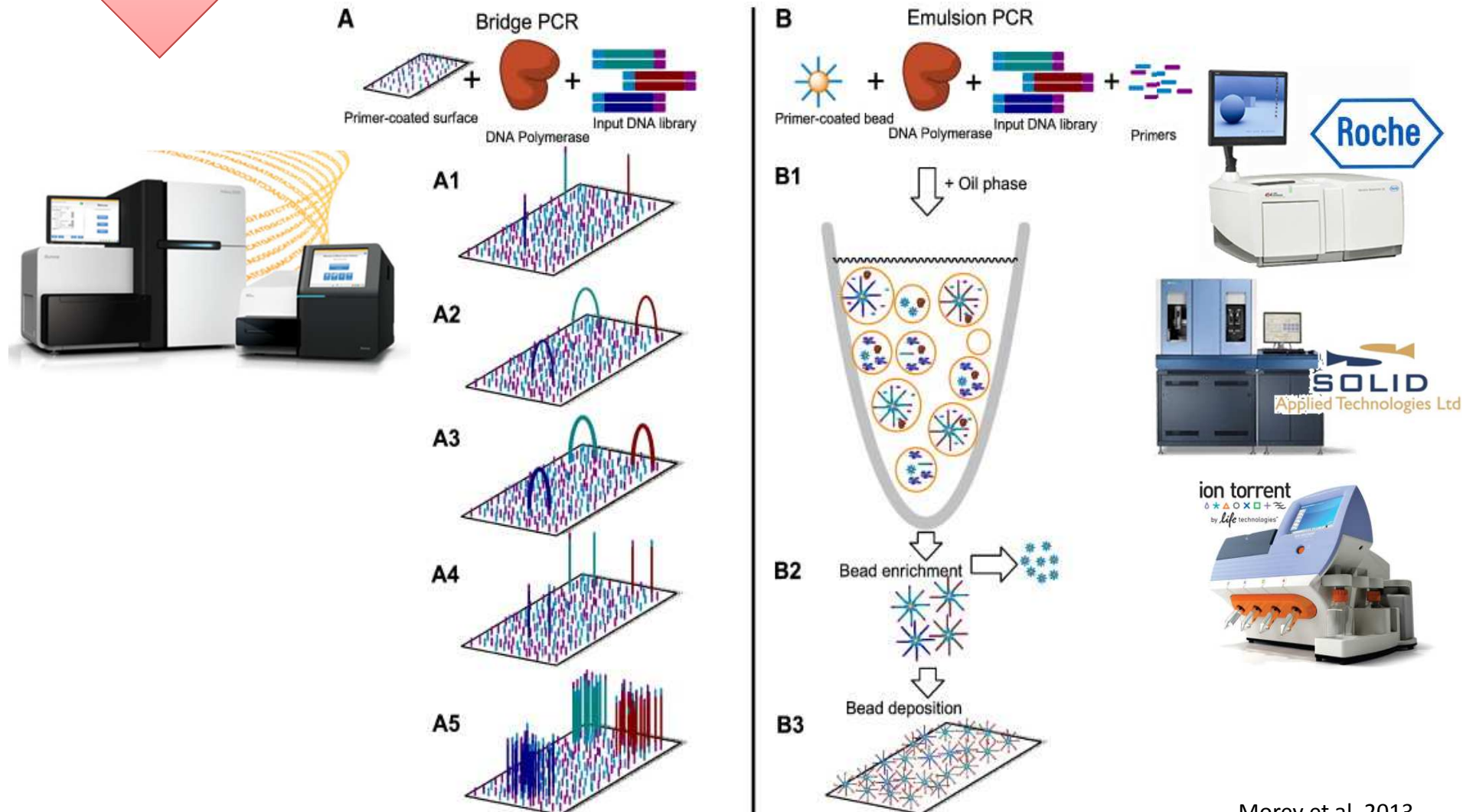


1. Estrazione del DNA

2. Preparazione della library

3. Cattura e arricchimento

Clonal array generation

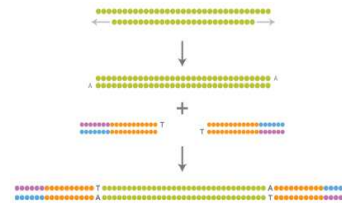


Morey et al. 2013

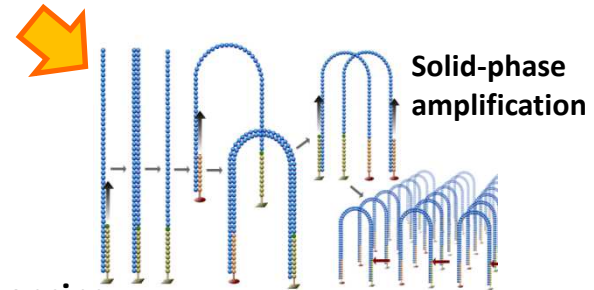
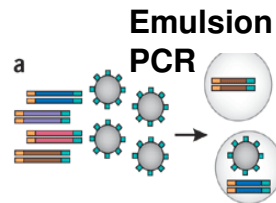
WORKFLOW del SEQUENZIAMENTO

- 1 Library preparation
- 2 Clonal amplification
- 3 Cyclic array sequencing

1 Library preparation

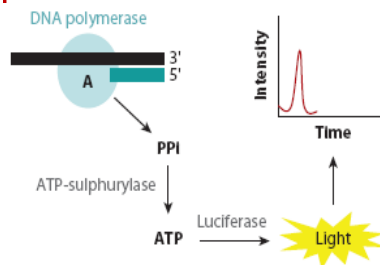


2 Clonal amplification



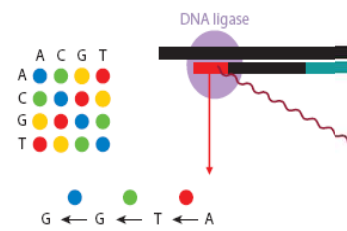
3 Cyclic array sequencing

pyrosequencing



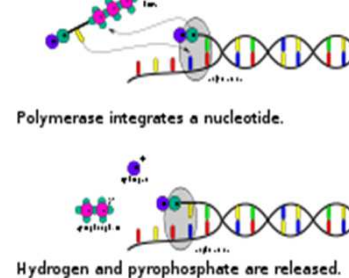
454

Sequencing by ligation



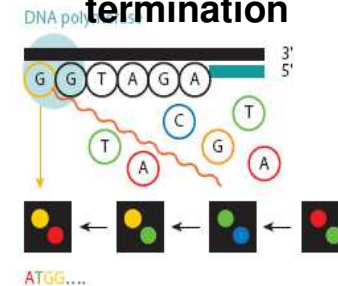
SOLID

semiconductor sequencing



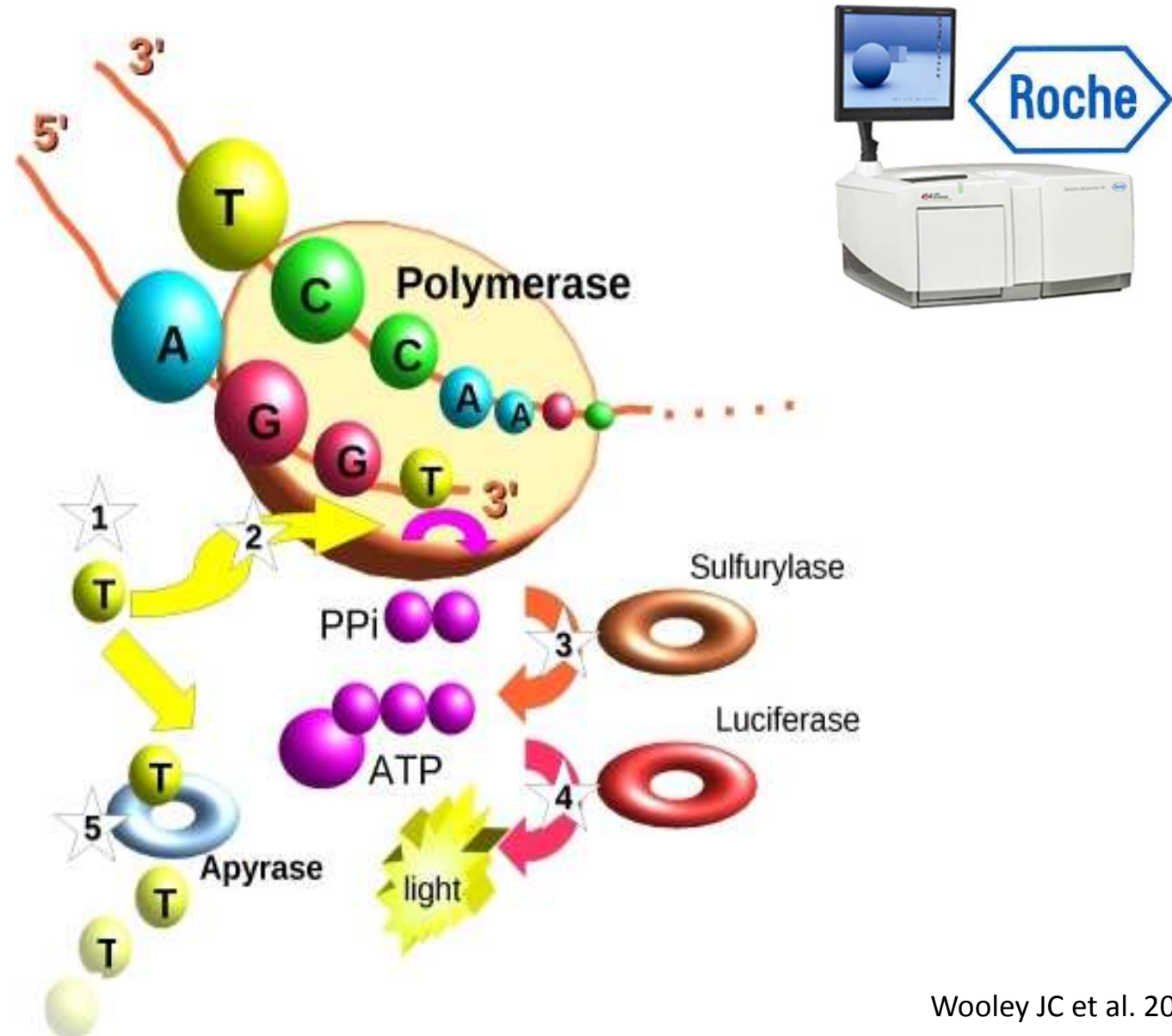
ION TORRENT

Sequencing by reversible termination



ILLUMINA

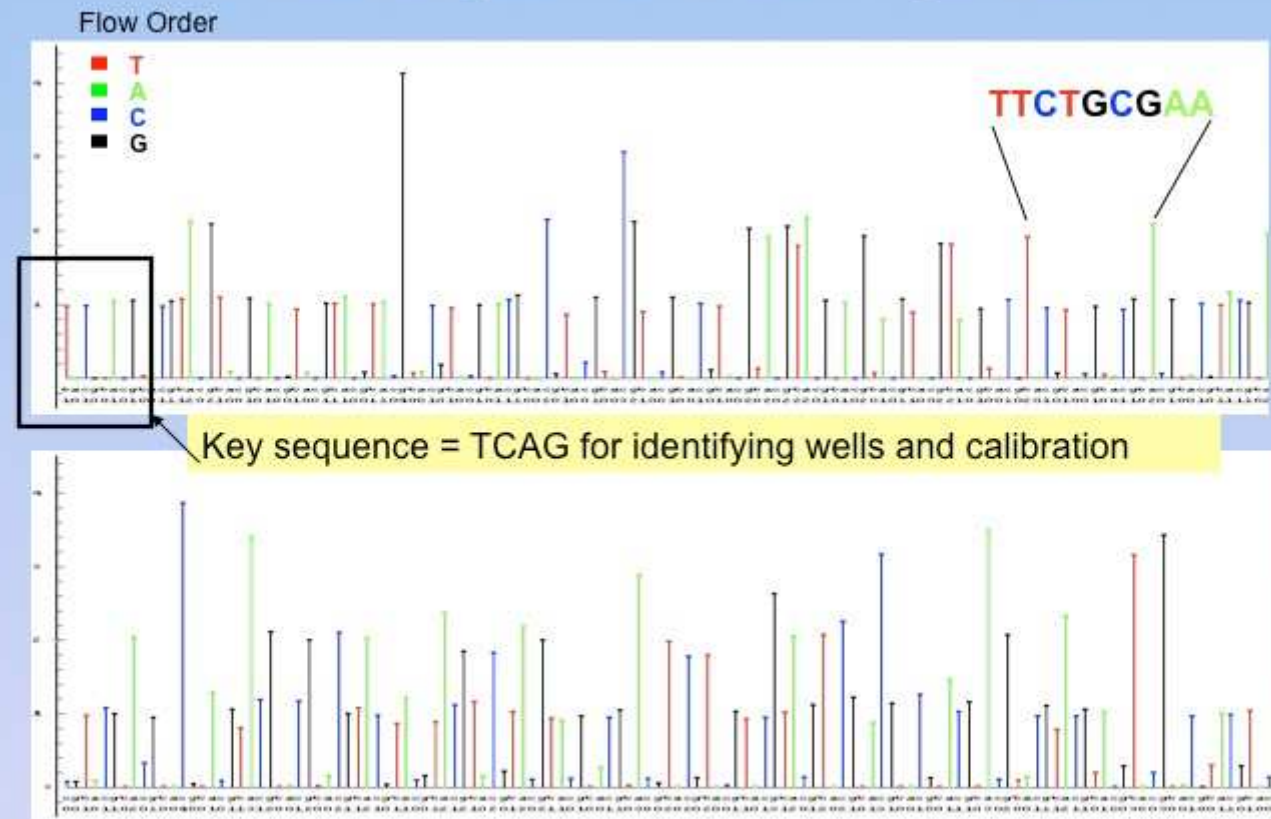
4. Sequenziamento: Pyrosequenziamento



3

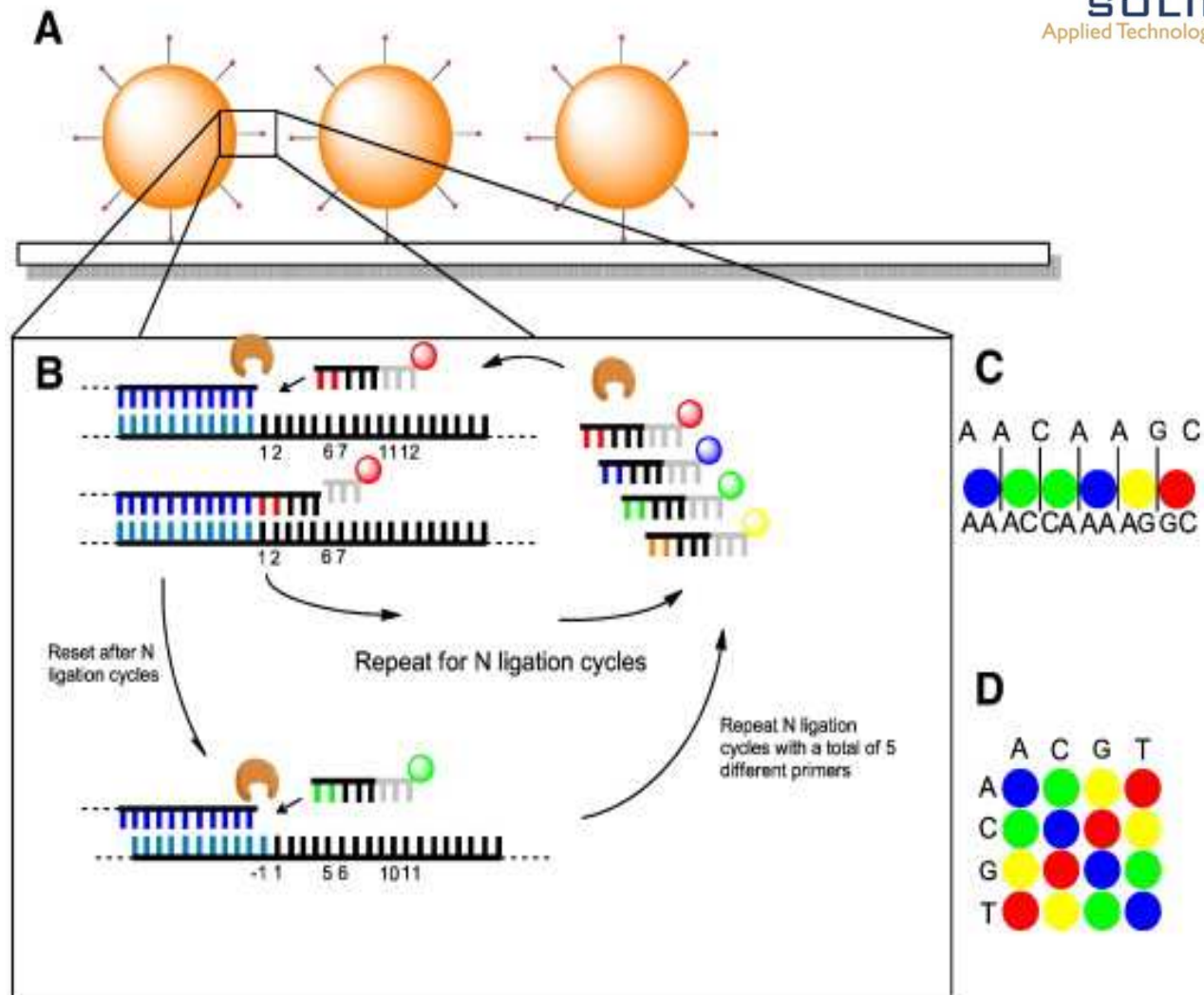
SEQUENCING pyrosequencing

Example of a Flowgram



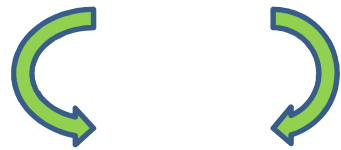
4. Sequenziamento: sequencing by ligation

Ligase Sequencing (SOLiD)

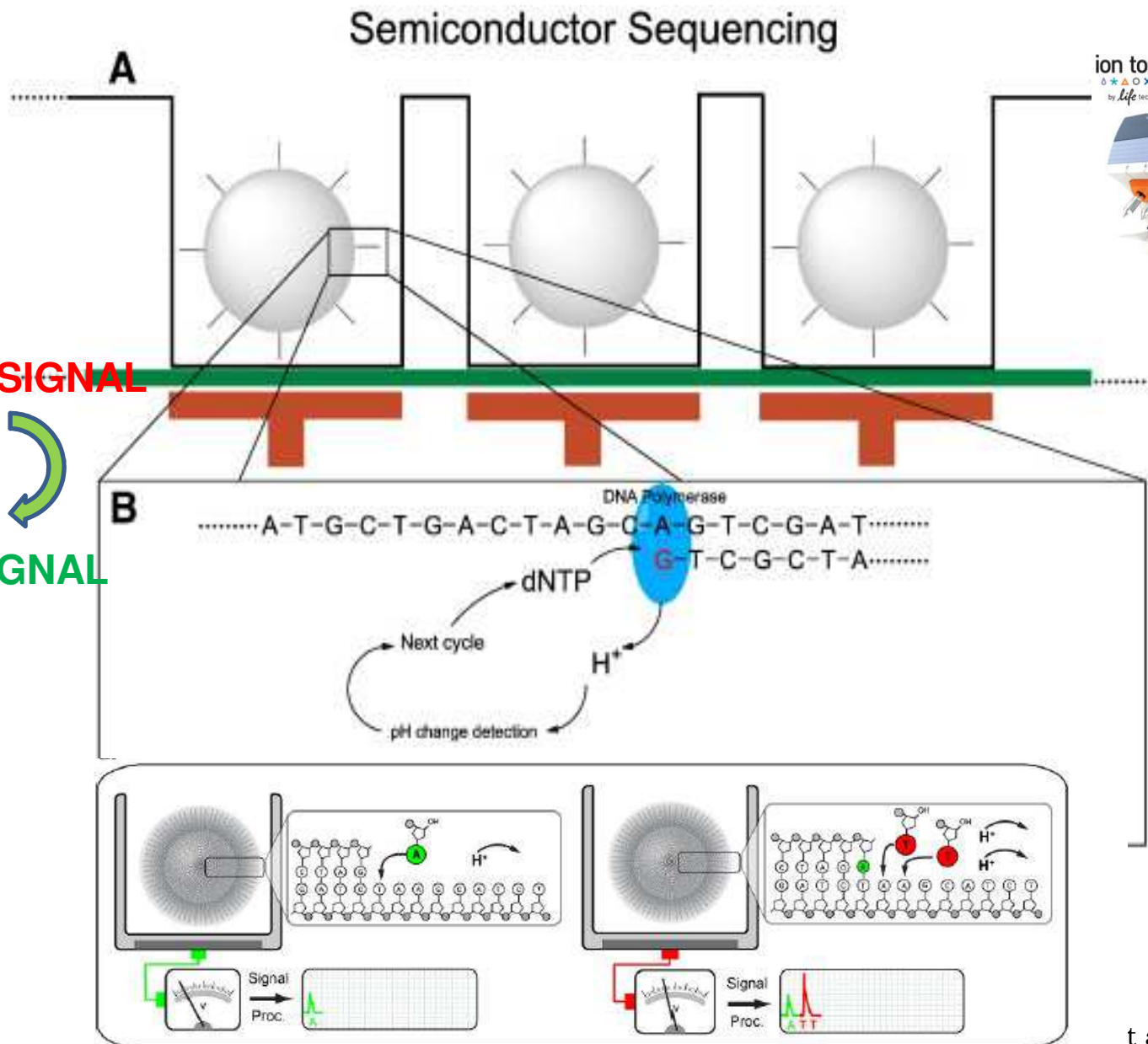


4. Sequencing: semiconductor sequencing

CHEMICAL SIGNAL

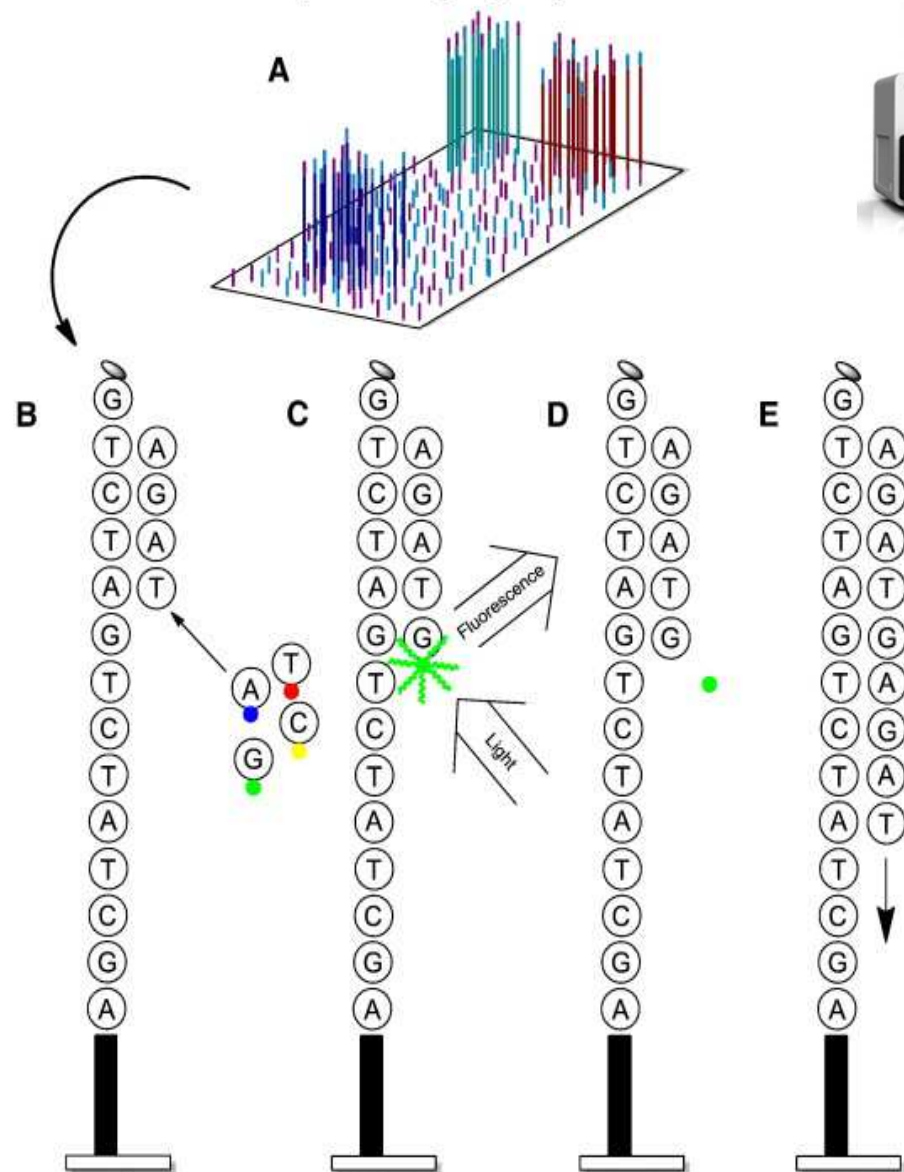


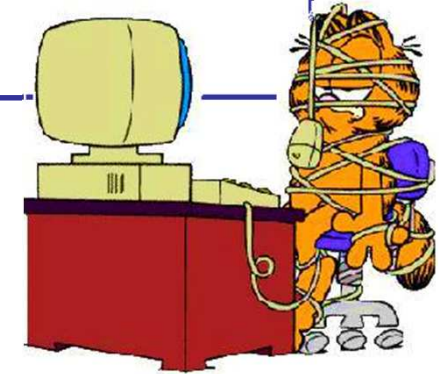
DIGITAL SIGNAL



4. Sequencing: sequencing by synthesis

Sequencing by Synthesis





DATA ANALYSIS

FASTQ file format

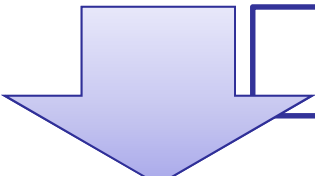
.....

```
@FCC0H3RACXX:8:1101:2380:2249#ACTGTTCC/1
TTACCGCGGCTGCTGGCACGGAGTTAGCCGGGGCTTCTTCTCCGGCTACCGTCATTATC
+
bbbeeeeeggggggiiiiiihiighiiiiiiigggeeedeedacccccccacacdddc
```

«@» Precede l'id univoco della sequenza

«+» Separa la sequenza nucleotidica dalla qualità

b = 98 | **e** = 101 | **g** = 103 | **i** = 105 | **h** = 104 | **c** = 99 | **a** = 97 | **d** = 100



5. Bionformatics analysis

QUALITY CONTROL



REFERENCE SEQUENCE



- IMAGE ACQUISITION



- DATA FILTERING



- ALIGNMENT



- VARIANT CALLING



- ANNOTATION

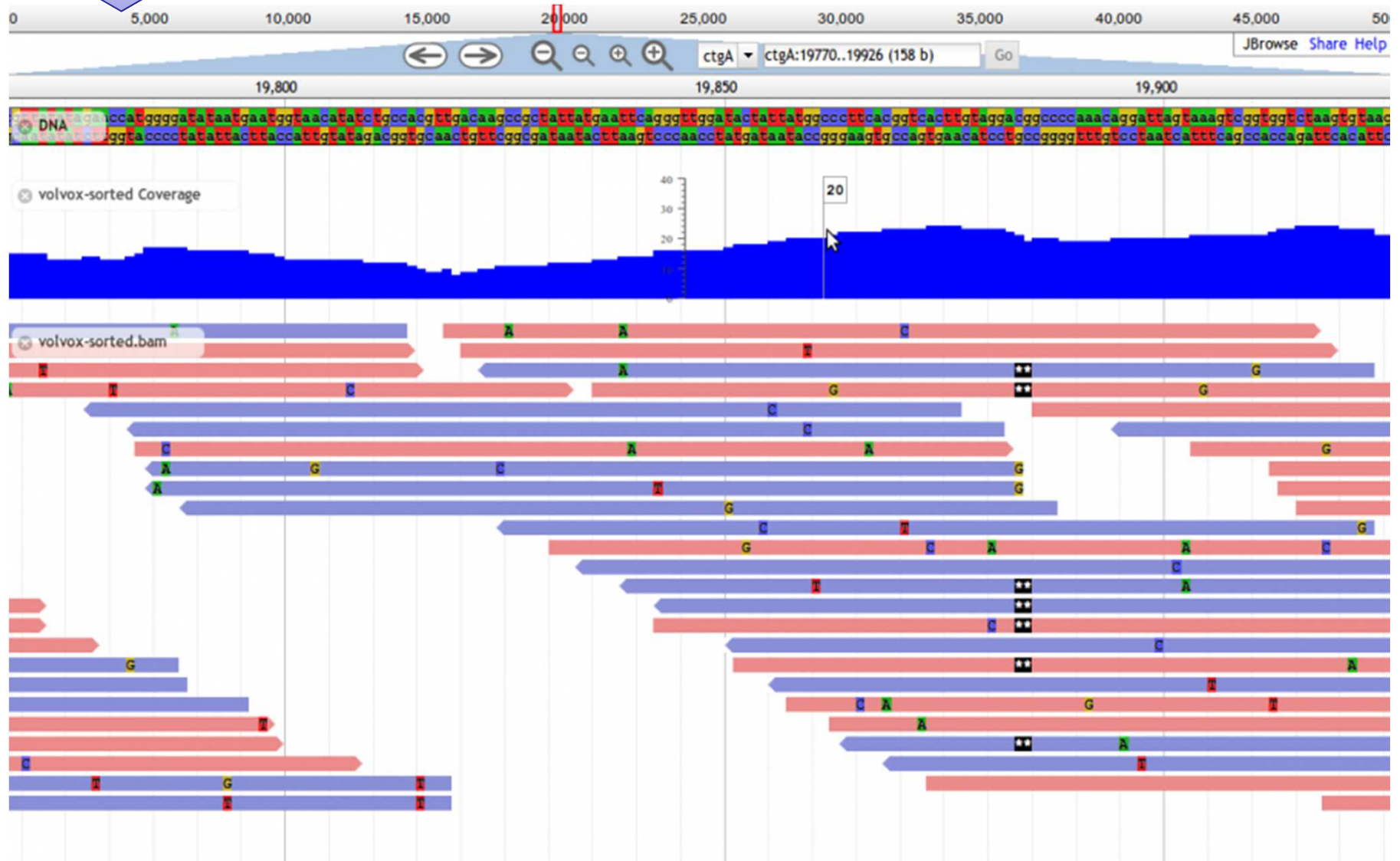


- TAQ COUNTING



- DATA REPORTING

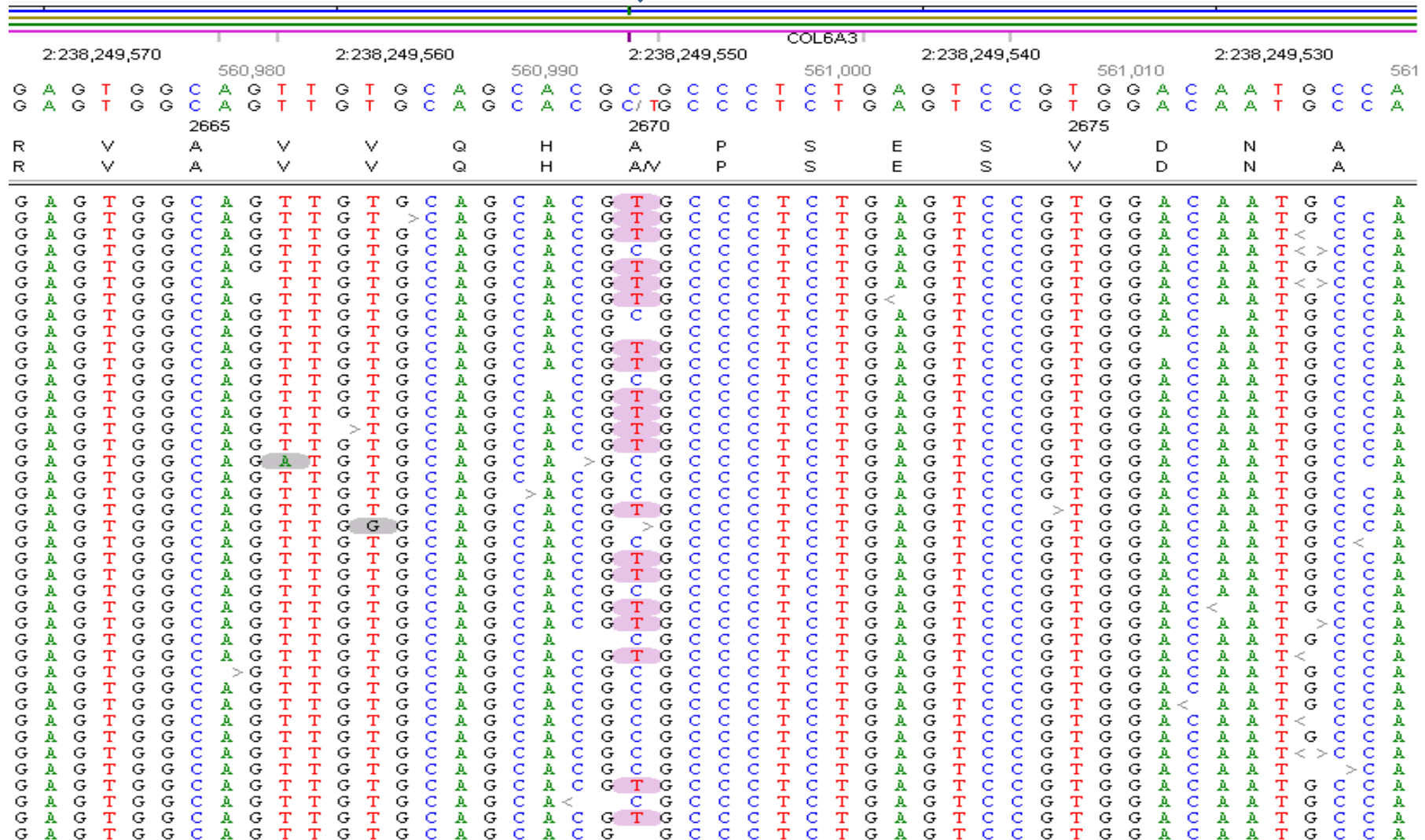
5. Bionformatics analysis



[illegible]

AFFETTO ETEROZIGOTE DI MUTAZIONE DOMINANTE MIOPATIA DI BETHLEM

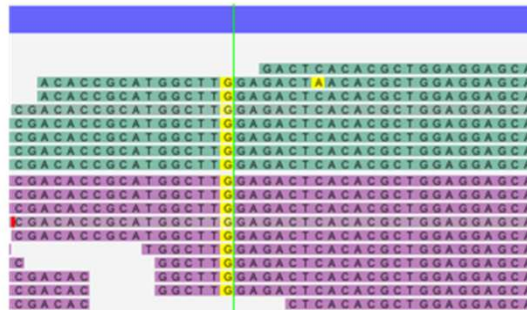
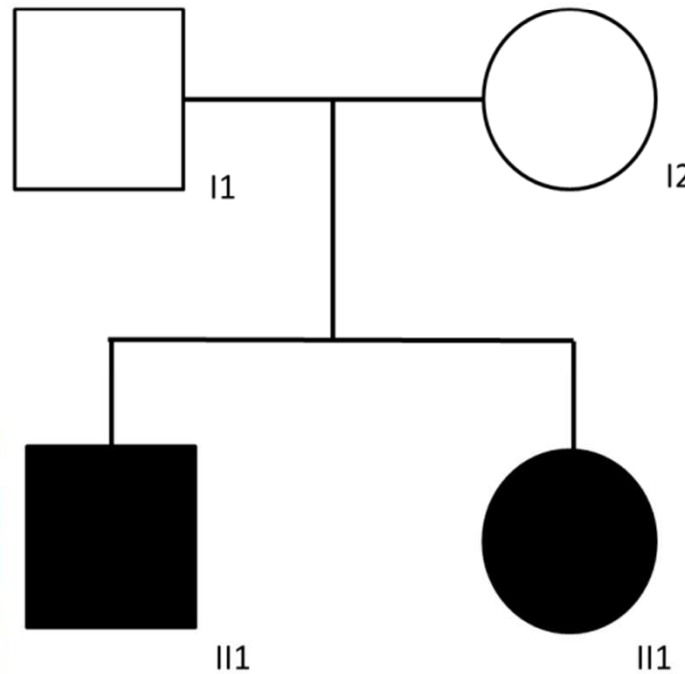
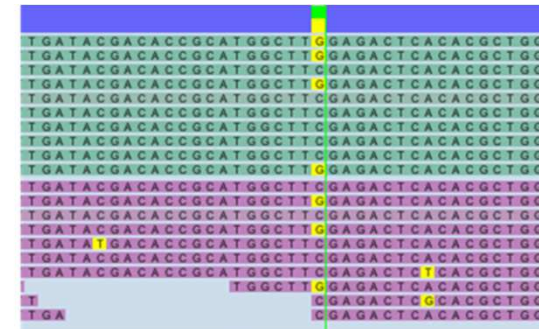
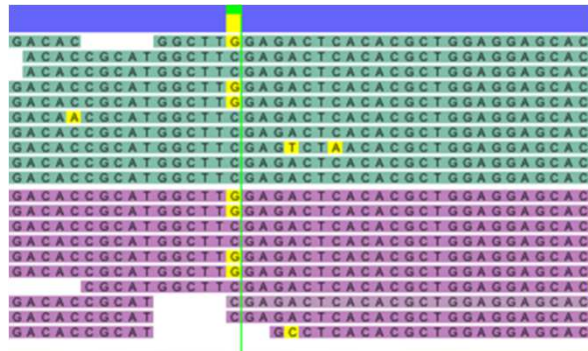
1747/11 COL6A3 c.8009C>T p.A2670V (het)



DEL GENE DYSTROFINA 2619/11 DMD c.5697_5698insA

[illegible]

ANALISI DEL PEDIGREE E SEGREGAZIONE FAMILIARE





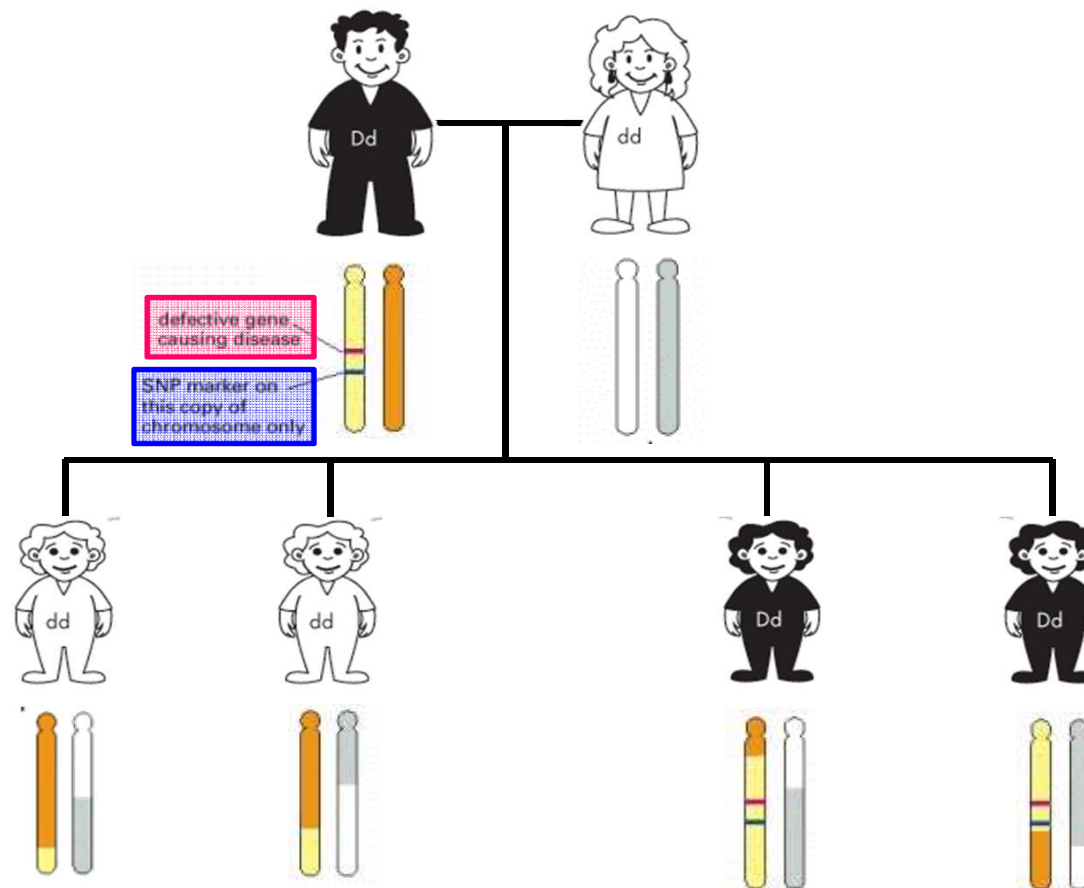
APPLICAZIONI DI NEXT GENERATION SEQUENCING

1. ricerca di mutazioni in geni noti: screening in patologie caratterizzate da eterogenicità genetica e fenotipica
 2. identificazione di nuovi geni-malattia
-
1. Studi di associazione (GWAS)

L'ANALISI DI LINKAGE

permette di determinare la posizione cromosomica di un **locus** responsabile di una determinata **malattia**/carattere genetico rispetto a **marcatori polimorfici molecolari** la cui localizzazione è nota

Marcatore genetico: Locus genico che identifica univocamente una regione cromosomica





Marcatore genetico: Locus genico che identifica univocamente una regione cromosomica

Caratteristiche di un marcatore:

- **Polimorfico** (min 2 alleli alternativi)
- **Facile da identificare e stabile** tra le generazioni
- **Segrega in modo mendeliano**

TIPI DI MARCATORE:

- RFLP (*restriction fragment length polymorphism*)
- SSLP (*simple sequence length polymorphism*)
- **SNP** (*single nucleotide polymorphism*)



Marcatore genetico: Locus genico che identifica univocamente una regione cromosomica

Caratteristiche di un marcatore:

- **Polimorfico (min 2 alleli alternativi)**
- **Facile da identificare e stabile tra le generazioni**
- **Segrega in modo mendeliano**

TIPI DI MARCATORE:

- **RFLP (*restriction fragment length polymorphism*)**

Variant 1
EcoRI does not cut

GCC**GCATTC**TA
CGG**CGTAAG**AT

Variant 2
EcoRI does cut

GCC**GAATTC**TA
CGG**CTTAAG**AT



Uncut



} Cut

1

2-1

2

Phenotype



Marcatore genetico: Locus genico che identifica univocamente una regione cromosomica

Caratteristiche di un marcatore:

- **Polimorfico (min 2 alleli alternativi)**
- **Facile da identificare e stabile tra le generazioni**
- **Segrega in modo mendeliano**

TIPI DI MARCATORE:

- **SNP (*single nucleotide polymorphism*)**
 - >1% popolazione
 - Snps in genoma: 10 milioni
 - Breve distanza 1 ogni 290 bp
 - Progetto 1000 genomi (esempio di SNPs)
 - SNPs usati in GWAS studies



GWAS

(genome wide association study)

- Studi prospettici su popolazione di larga scala
- Utilizzano come marcatori SNP
- Studi ipotesi-indipendenti
- Correlazione SNP-patologia

SNPs associati a patologia



Ruolo causale diretto nella sua origine

Linkage disequilibrium con variante patogenetica

Esempi:

- ☐ Decode genetics (90% della popolazione)
- ☐ UK BIOBANK (500,000 individui)
- ☐ Australian project (2,000,000)



GWAS

(genome wide association study)

Validità clinica:

- marcatore specifico ($a/a+c$) 75%
- marcatore sensibile ($d/b+d$) 80%
- Alto valore predittivo + ($a/a+b$) 11%
- O Alto valore predittivo – ($d/c+d$) 99%

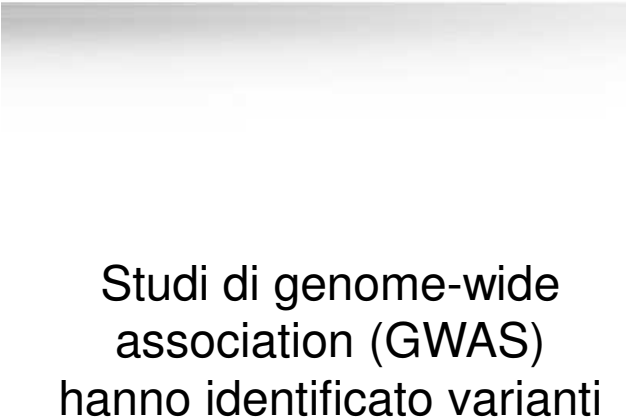
	malati	sani
+	119 (a)	926 (b)
-	39 (c)	3704 (d)

- Analisi di variabili
 - genetiche (num e freq)
 - epidemiologiche (rischio malattia e rischio dei singoli genotipi)
- porta a calcolo di AUC indicatore di validità clinica, indica se il test riconosce nella popolazione quelli che si ammaleranno

0,5(nessuna capacità discriminante) $<AUC<1$ (individuazione perfetta)

Test valido con $AUC \geq 0.8$

A vertical illustration of a DNA double helix structure. The two strands are represented by thick black ribbons that spiral around each other. They are connected by horizontal silver-colored bars representing the base pairs. The structure is set against a light blue background that has a subtle gradient and a faint, curved white line on the right side.



genetiche associate a
patologie complesse



Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer

Michailidou K et al.

Nature Genetics 47, 373–380 (2015) doi:10.1038/ng.3242

Best variant	Locus	Position ^b	Alleles ^c	Genes within ± 2 kb
rs12405132	1q21.1	145,644,984	C/T	LOC10028814,NBPF10,RNF115
rs12048493	1q21.2	149,927,034	A/C	—
rs72755295	1q43	242,034,263	A/G	EXO1
rs6796502	3p21.3	46,866,866	G/A	—
rs13162653	5p15.1	16,187,528	G/T	—
rs2012709	5p13.3	32,567,732	C/T	—
rs7707921	5q14	81,538,046	A/T	ATG10
rs9257408	6p22.1	28,926,220	G/C	—
rs4593472	7q32.3	130,667,121	C/T	FLJ43663
rs13365225	8p11.23	36,858,483	A/G	—
rs13267382	8q23.3	117,209,548	G/A	LINC00536
rs11627032	14q32.12	93,104,072	T/C	RIN3
chr17:29230520:D	17q11.2	29,230,520	GGT/G	ATAD5
rs745570	17q25.3	77,781,725	A/G	—
rs6507583	18q12.3	42,399,590	A/G	SETBP1



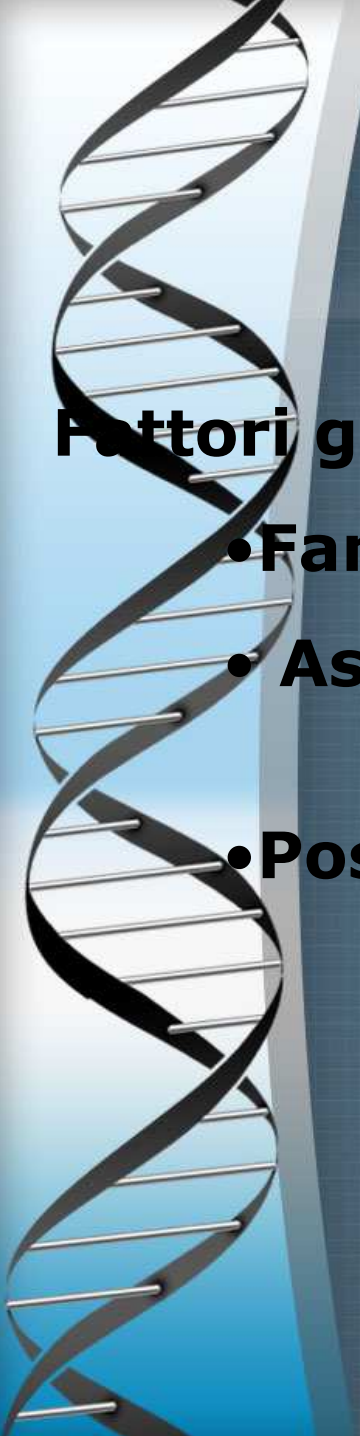
ESEMPI DI ANALISI GWAS IN MALATTEI COMPLESSE

- DIABETE
- ALZHEIMER



Malattie ad Eredità complessa: **Diabete mellito insulino-dipendente**

- Frequenza 1:2500 a 5 anni e 1:300 a 18 anni in Nord-America
- Distruzione su base autoimmune delle cellule β del pancreas
- Mancata produzione di Insulina
- Componente genetica e ambientale
- 13 loci di suscettibilità. HLA DR3 + DQ β_1 *0201 o DR4 + DQ β_1 *0302 (aa neutro)
- DQ β_1 *602 (Asp) protettivo
- Infezioni virali, precoce esposizione all'albumina bovina (mimetismo molecolare con antigeni cellule β)
- Anticorpi anti-cellule β



Storia naturale

Fattori genetici

- **Familiarità della malattia**
- **Associazione con antigeni di istocompatibilità (HLA-DR3, DR4 e DQ2)**
- **Possibili geni candidati (Tabella 3)**

Tabella 3

LOCUS	CROMOSOMA	GENE CANDIDATO
-------	-----------	----------------

IDDM 1

6p21

HLA

11p15

Insulin VTNR

15q26

11q13

6q25

18q21

2q31

6q27

3q21-q25

10p13-q11

14q24-q31

2q33

CTLA-4

2q34-q35

6q21

7p13



Fattori ambientali

Fattori ambientali:

- **Infezioni virali (Echovirus, Coxsackie A e B)**
- **Alimenti (proteine del latte vaccino, glutine, vitamina D, altri)**
- **Tossine**

I fattori ambientali favoriscono l'instaurarsi dei meccanismi autoimmunitari che conducono alla distruzione delle beta-cellule



Meccanismi autoimmunitari

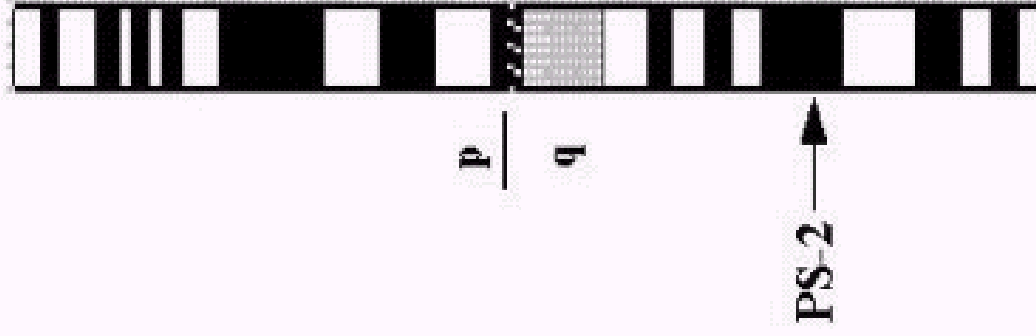
- ✓ **Autoanticorpi circolanti diretti contro le cellule insulari (ICA)**
- ✓ **Gli ICA ad alto titolo (> 20 unità JDF) predicono un rischio del 40-60% nei successivi 5-7 anni**
- ✓ **Gli ICA e gli altri auto-anticorpi sono presenti nel periodo di “prediabete”**
- ✓ **Quando sono presenti anticorpi multipli, la capacità predittiva aumenta**



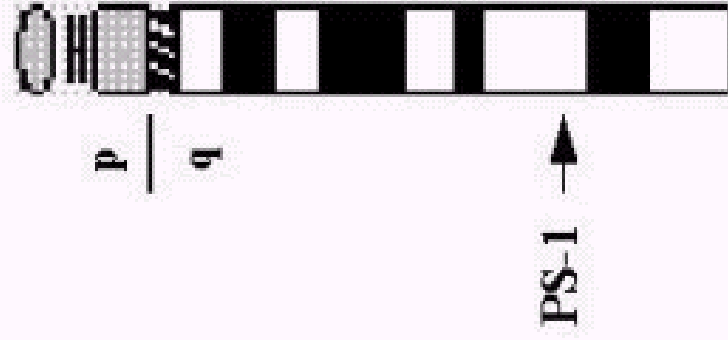
Malattie ad Eredità complessa: **Malattia di Alzheimer**

- 75% dei casi sporadici
- 10% Autosomica dominante, altri casi origine multifattoriale
- Accumulo nel cervello di:
 - proteina β -amiloide (placche amiloidi extracellulari)
 - proteina Tau (fusi neurofibrillari intraneuronali)
- 4 geni responsabili della forma familiare
- Individui con Sindrome di Down sviluppano la malattia precocemente

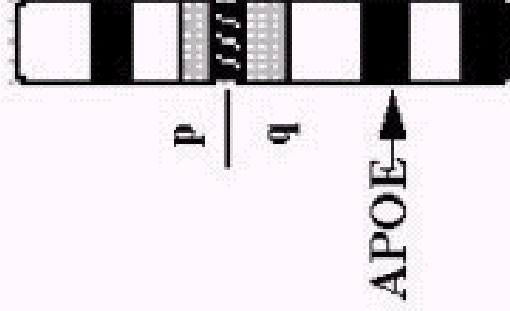
Chromosome 1



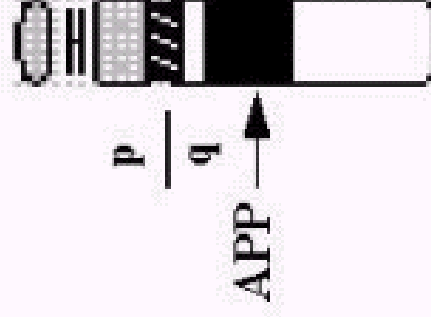
Chromosome 14



Chromosome 19

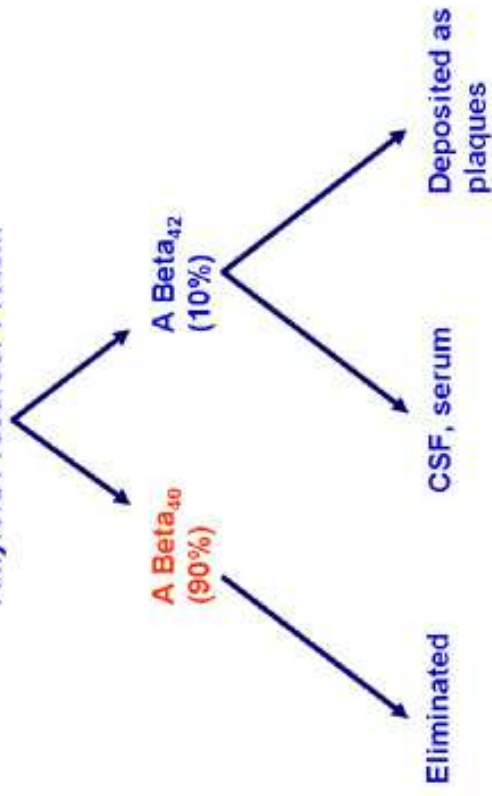


Chromosome 21



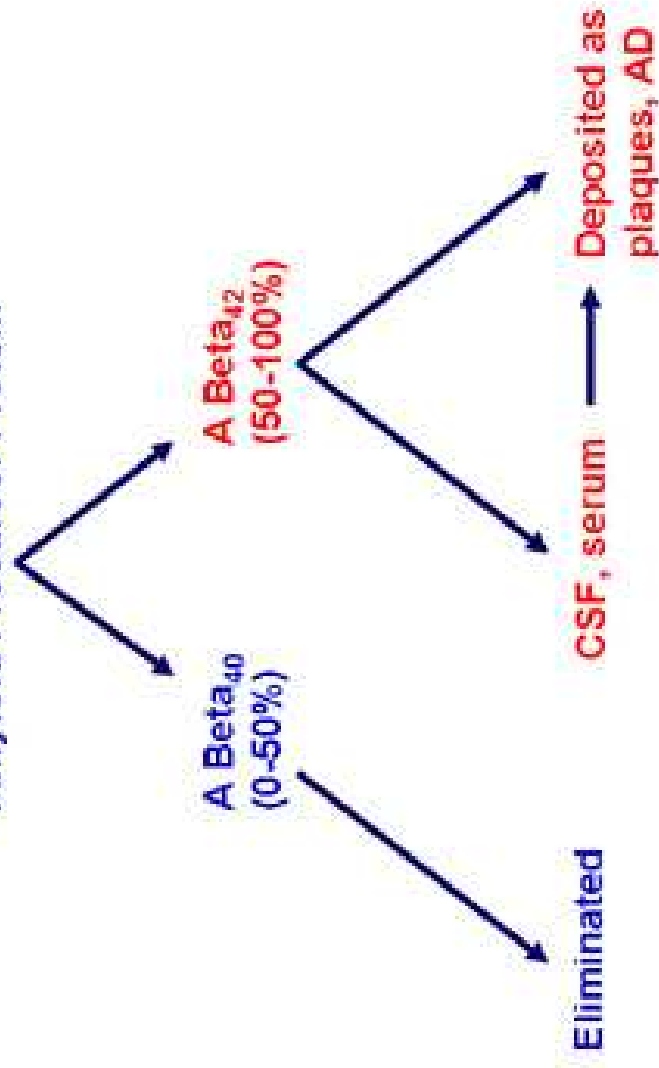
Normal Amyloid Precursor Metabolism

Amyloid Precursor Protein

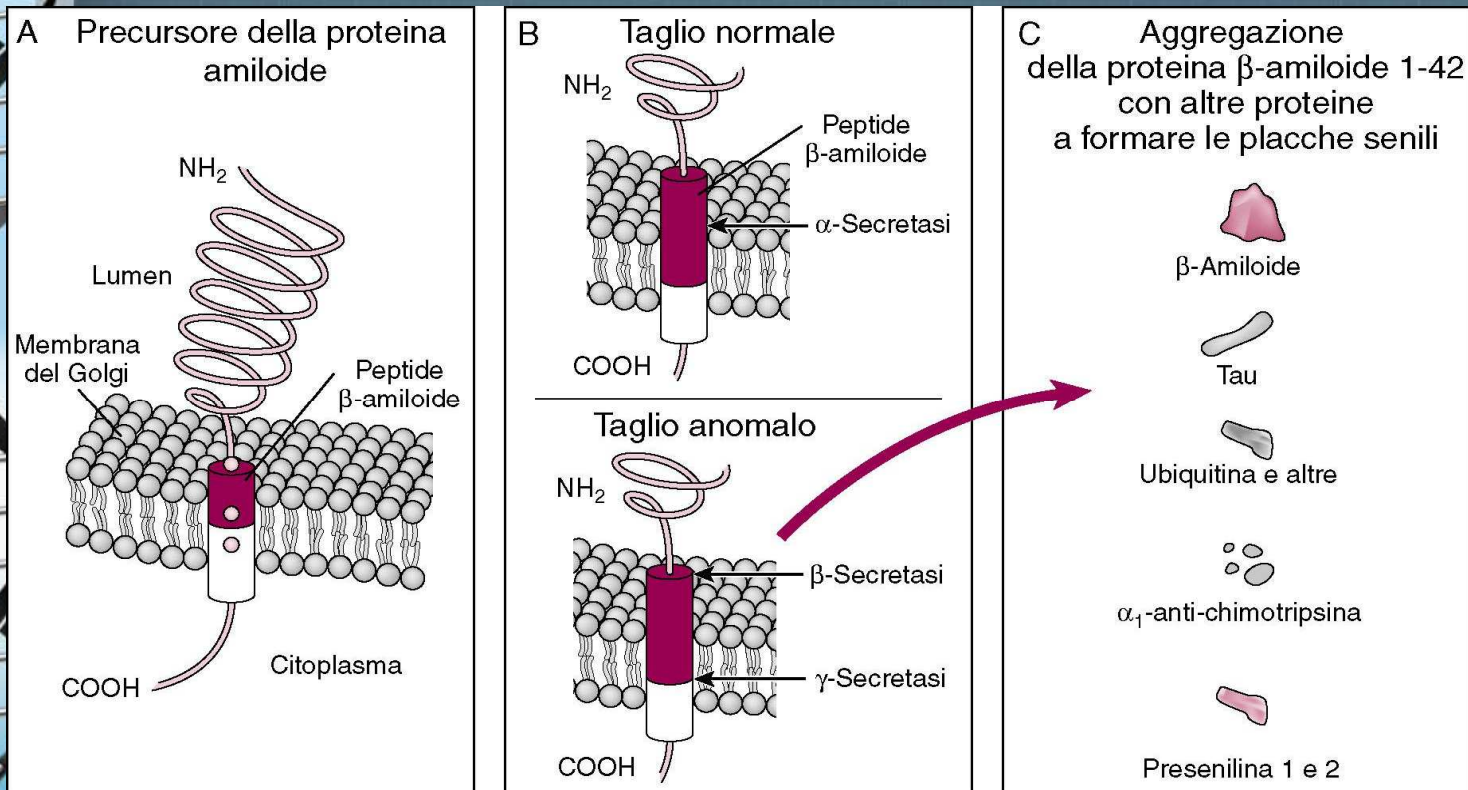


Abnormal Amyloid Precursor Metabolism

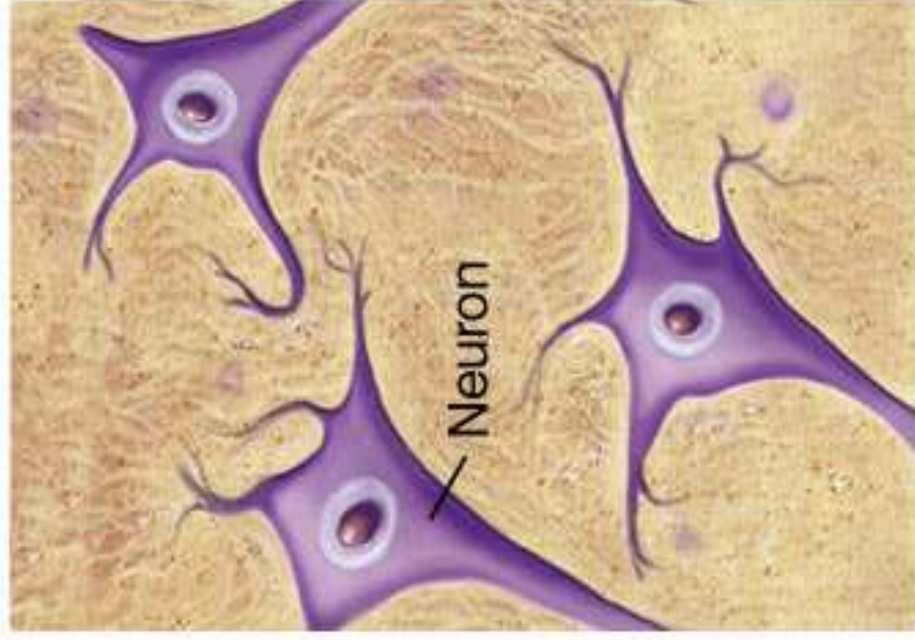
Amyloid Precursor Protein



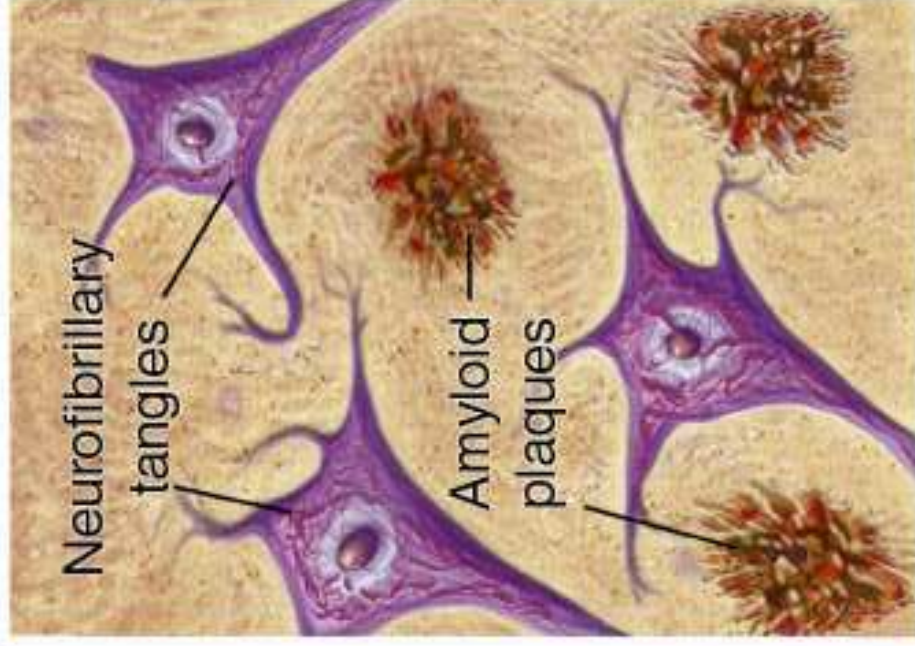
Formazione della proteina amiloide

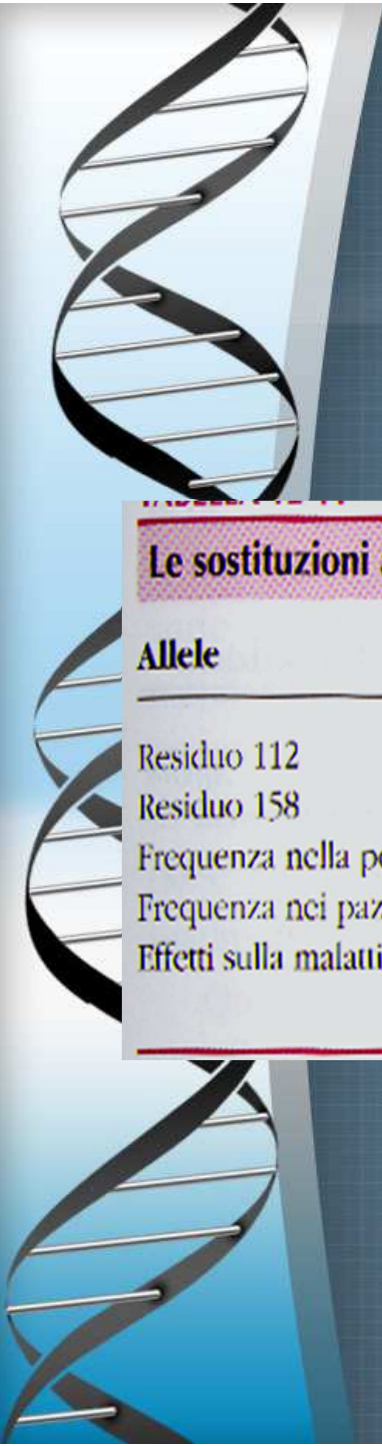


Normal



Alzheimer's



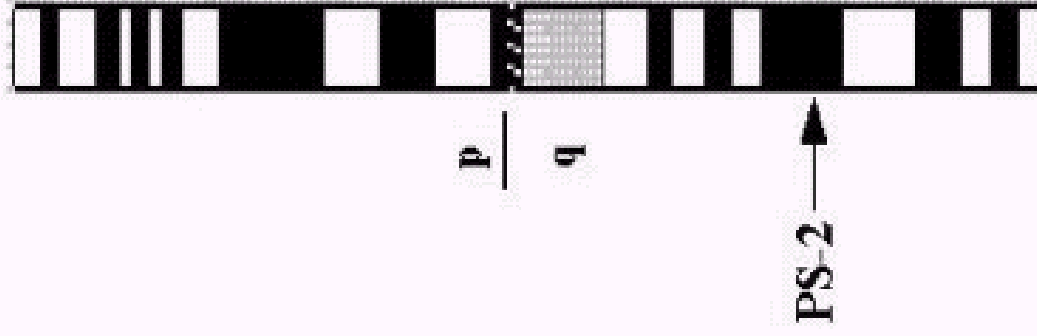


Il gene APOE e la malattia di Alzheimer

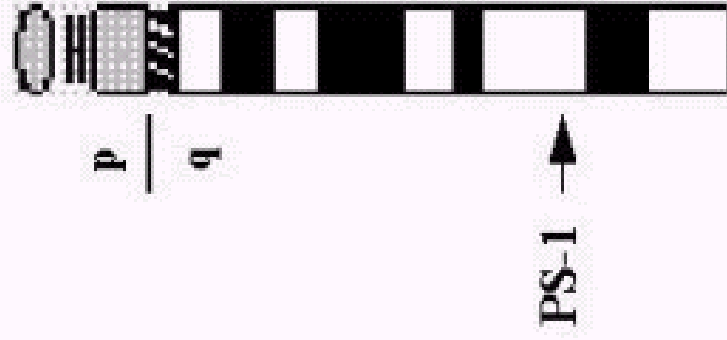
Le sostituzioni amminoacidiche che rappresentano i tre alleli APOE più comuni

Allele	$\epsilon 2$	$\epsilon 3$	$\epsilon 4$
Residuo 112	Cys	Cys	Arg
Residuo 158	Cys	Arg	Arg
Frequenza nella popolazione caucasica	10%	75%	15%
Frequenza nei pazienti con malattia di Alzheimer	2%	58%	40%
Effetti sulla malattia di Alzheimer	Protettivo	Non nota	30-50% di rischio genetico di AD

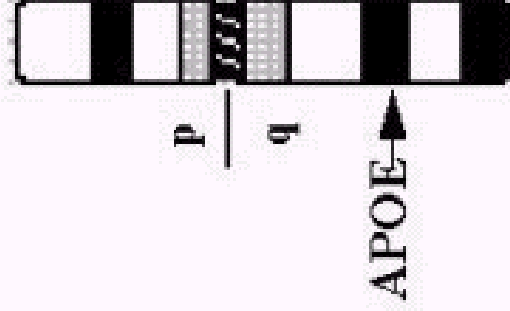
Chromosome 1



Chromosome 14



Chromosome 19



Chromosome 21

