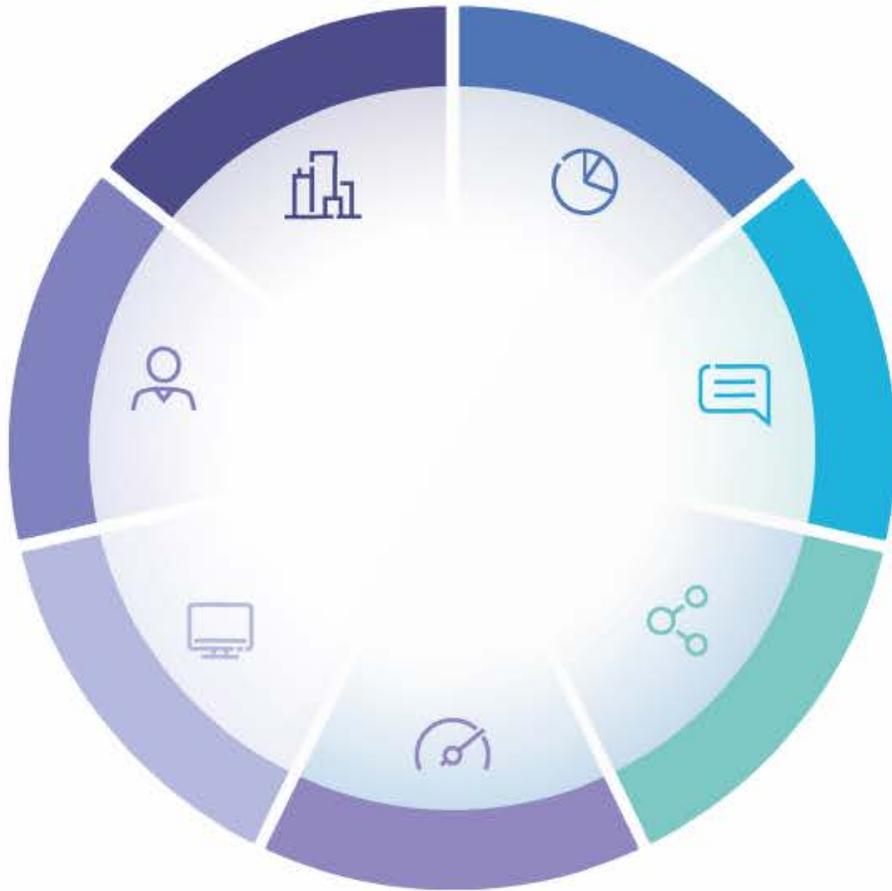


Il Progetto Genoma Umano

Presentazione a cura di Assunta Croce, PhD

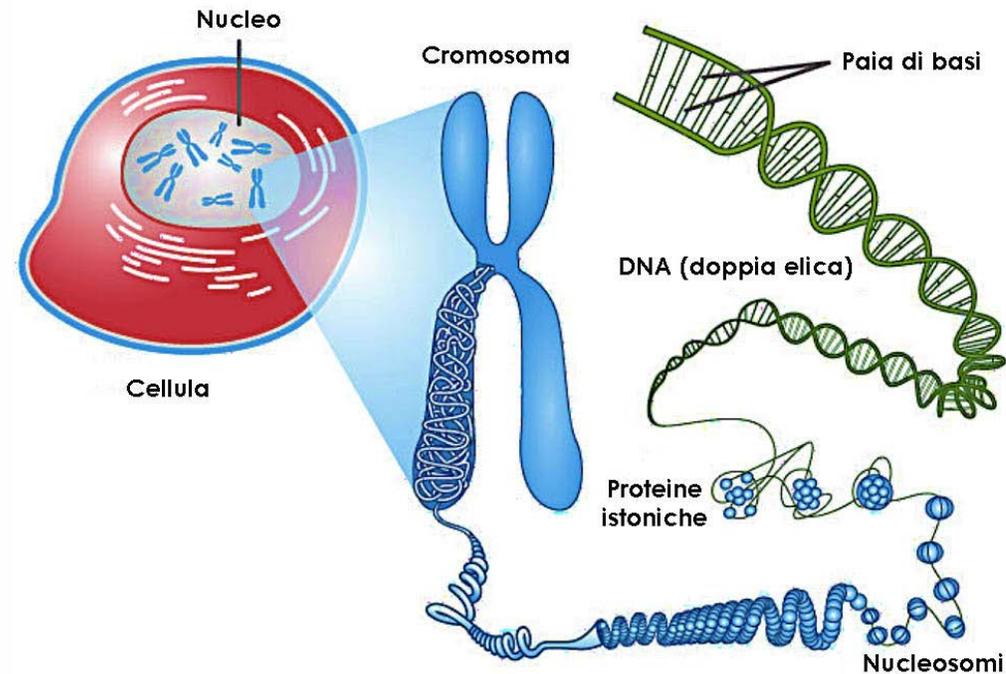


Indice

1. IL PROGETTO, UNA SFIDA SCIENTIFICA
2. COME SI SEQUENZIA IL DNA?
3. L'IMPATTO DEL PROGETTO GENOMA UMANO
4. MEDICINA PERSONALIZZATA

Che cosa è il Genoma?

- Insieme di tutte le informazioni genetiche contenute nel DNA di una cellula di organismo vivente
- Il manuale di istruzioni che indica alla cellula e all'organismo COME realizzare i processi alla base della vita



Il Progetto Genoma Umano, un'esplorazione di noi stessi



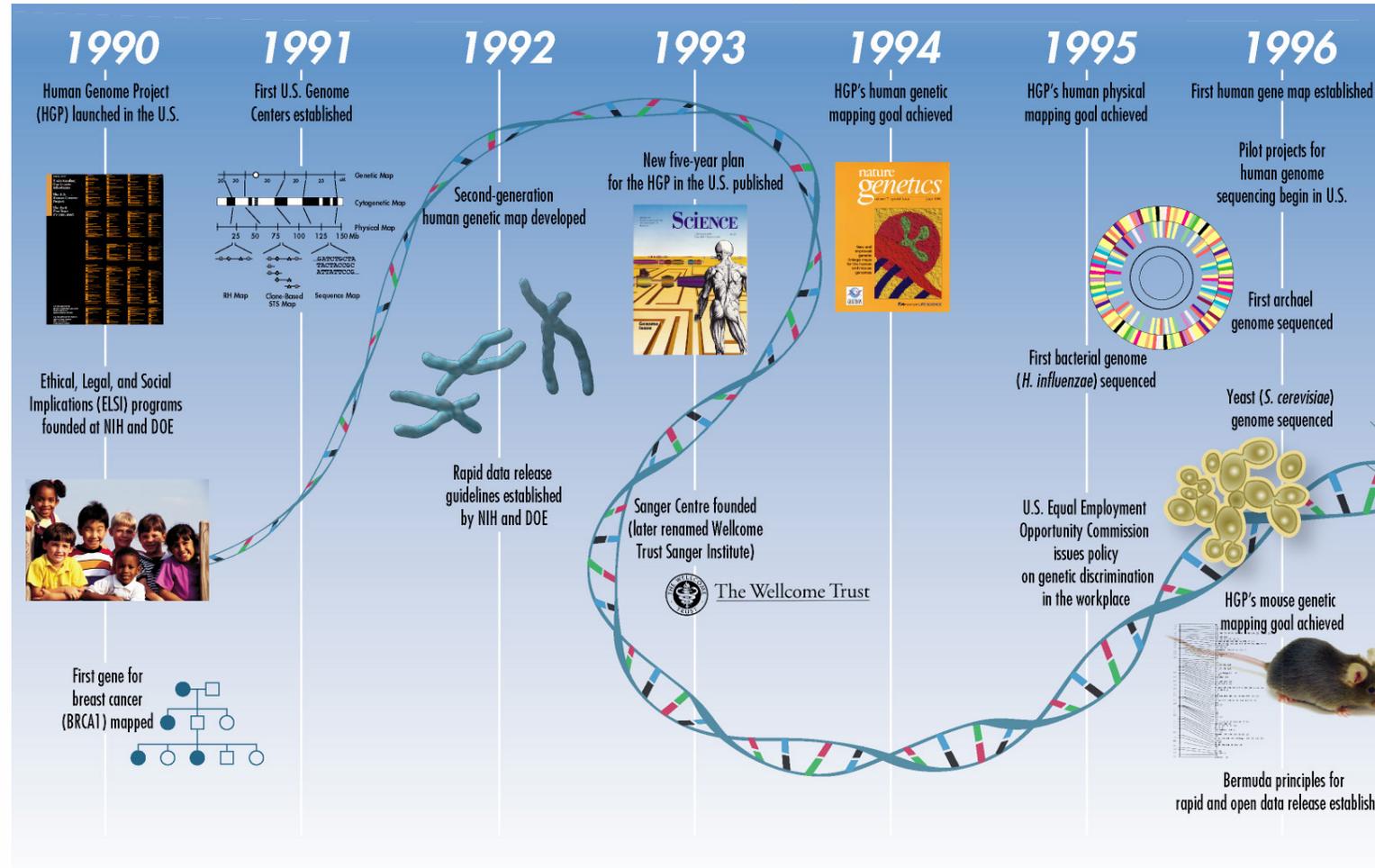
Credit & Copyright: Conselice et al., Hubble Heritage Team (STScI/AURA), NASA

Obiettivi del progetto

- **Sequenziare il genoma umano**, cioè scoprire l'esatta **successione** dei nucleotidi presenti nel DNA **rappresentativo** di un essere umano (diversi volontari) in 15 anni
- Definire una **mappa fisica e genetica** del genoma umano
- Sequenziare e mappare **5 organismi modello** (incluso topo)

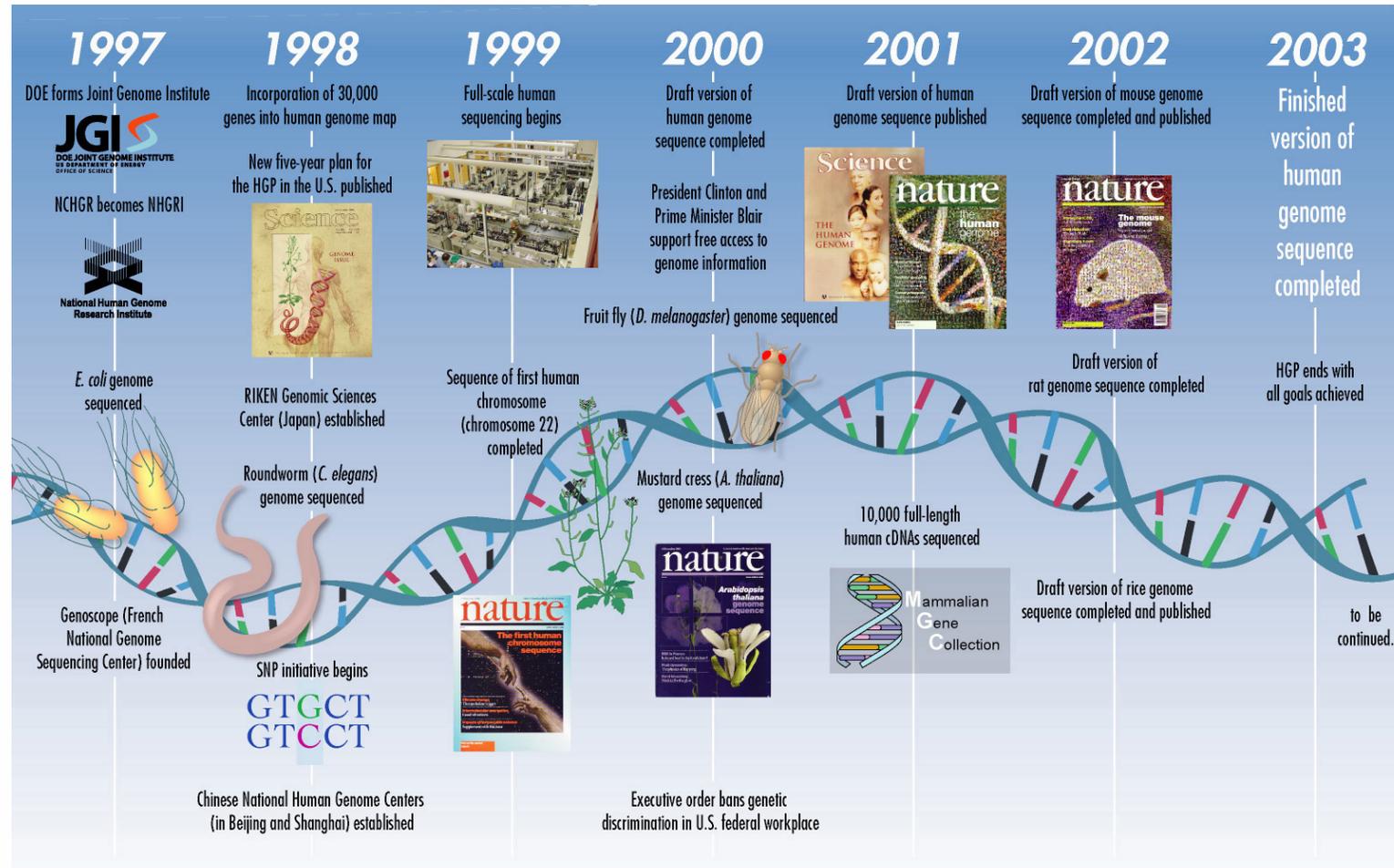
Stabiliti nel 1988 da un US Academy of Sciences e poi ripresi dall'ente americano che lanciò il progetto (Dipartimento dell'Energia)

Timeline



Credit: https://www.mun.ca/biology/scarr/Human_Genome_Project_timeline.html

Timeline



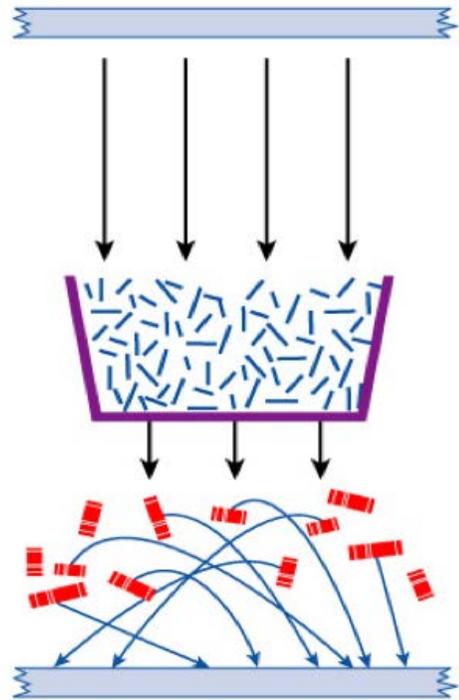
Credit: https://www.mun.ca/biology/scarr/Human_Genome_Project_timeline.html

L'annuncio nel 2001



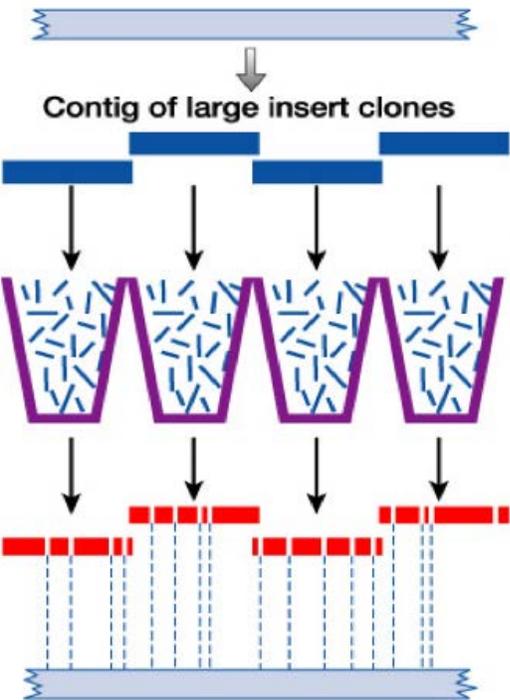
Due approcci diversi

Whole-genome shotgun
(CELERA Genomics)



Genome
Random fragmentation
Sequencing and assembly
Anchoring
Genome assembly

Hierarchical shotgun
(Conorzio Pubblico)



Modified from Human Molecular Genetics, Garland Science 2004

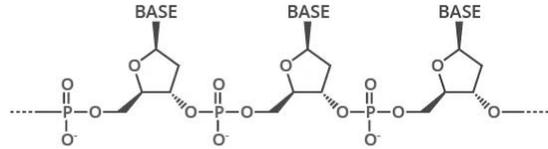
I numeri del progetto

- Costo del progetto: 3 miliardi di dollari
- Completato in 13 anni (2 anni in anticipo)
- 20 Istituzioni di 6 Paesi diversi: UK, Francia, Germania, Giappone e Cina

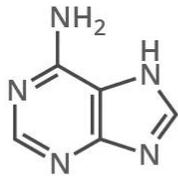
Come si sequenzia il DNA?

La struttura del DNA

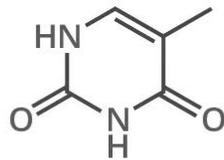
THE SUGAR PHOSPHATE 'BACKBONE'



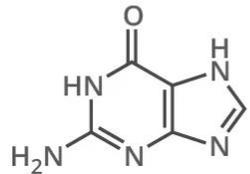
A ADENINE



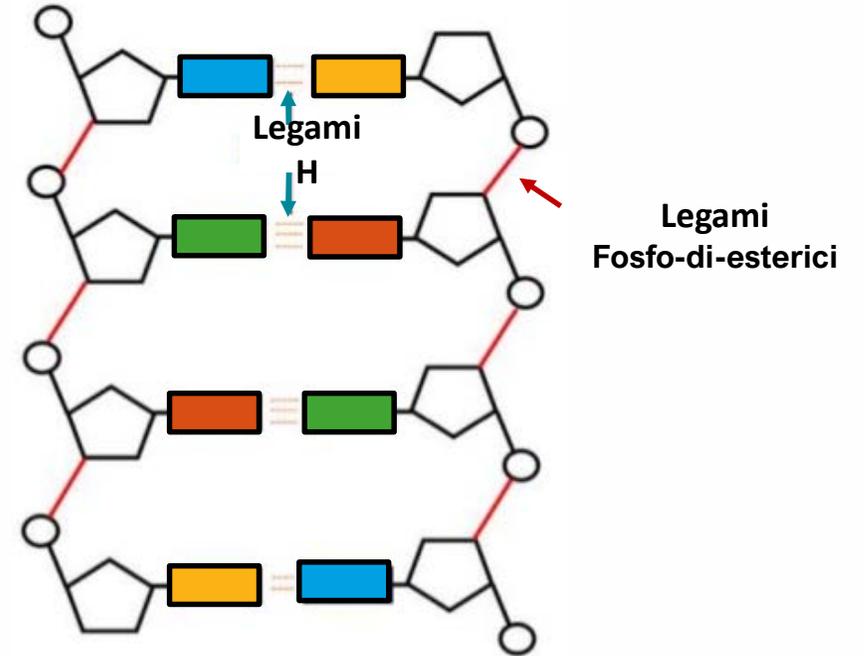
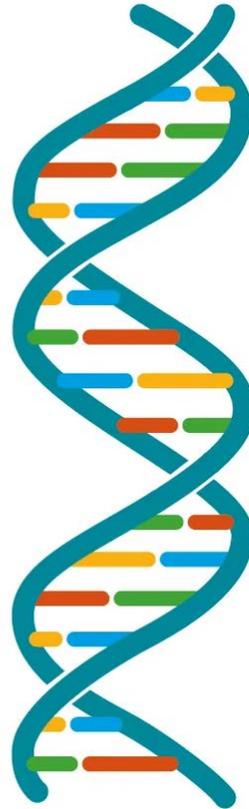
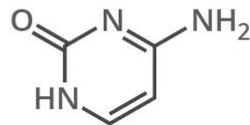
T THYMINE



G GUANINE

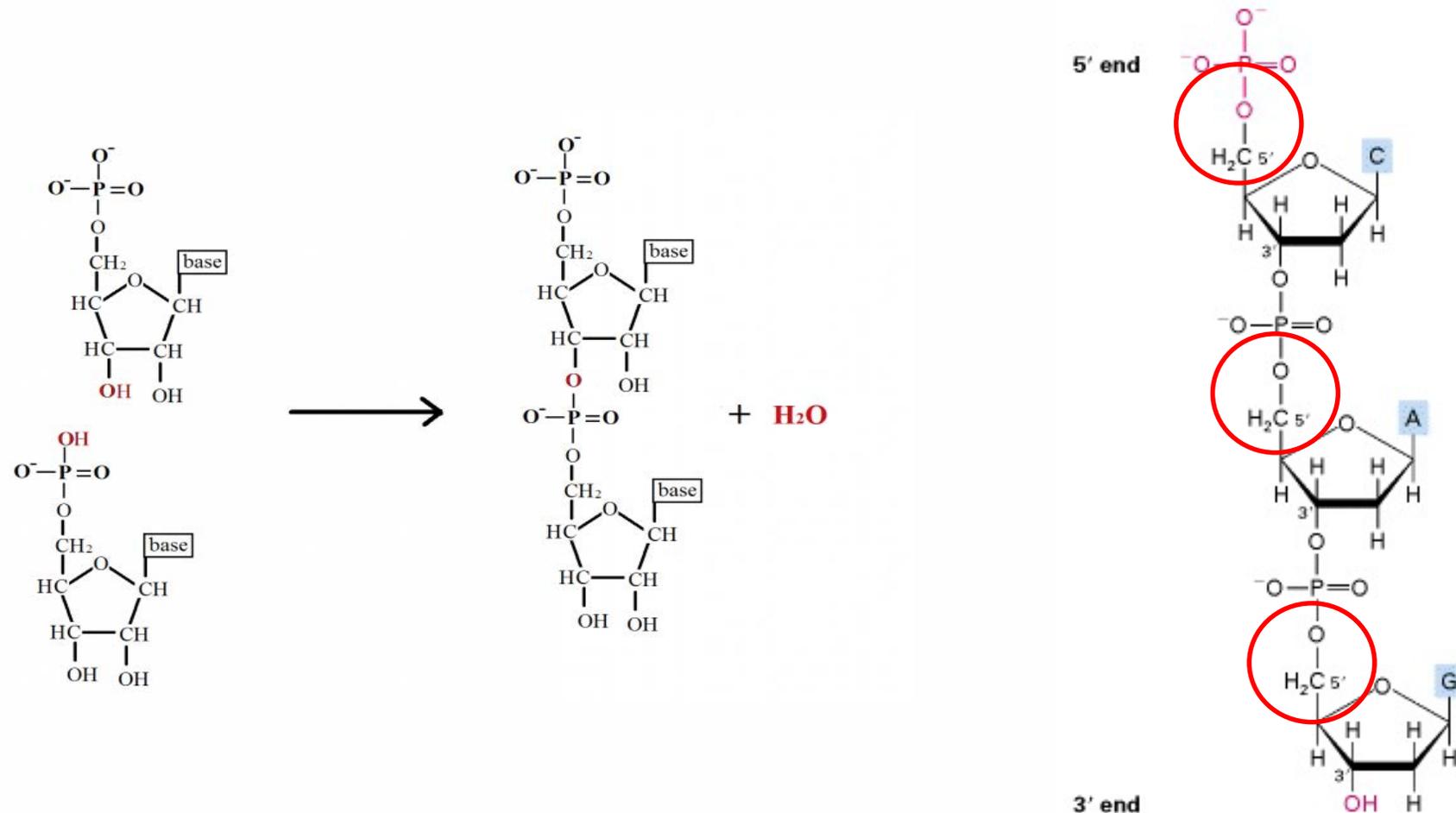


C CYTOSINE



Modified from www.compoundchem.com and from bioknowledgy.weebly.com

I legami fosfo-di-esterici



Metodo Sanger

- Metodo enzimatico (poiché sfrutta gli enzimi della replicazione del DNA)
- Sviluppato da Frederick Sanger nel 1977
- Basato sulla capacità di mimare *in vitro* la replicazione del DNA e sulla possibilità di interrompere la sintesi e identificare l'ultimo nucleotide inserito

Una scoperta da Nobel

Nature Vol. 265 February 24 1977

687

articles

Nucleotide sequence of bacteriophage Φ X174 DNA

F. Sanger, G. M. Air*, B. G. Barrell, N. L. Brown†, A. R. Coulson, J. C. Fiddes, C. A. Hutchison III‡, P. M. Slocombe§ & M. Smith*

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

A DNA sequence for the genome of bacteriophage Φ X174 of approximately 5,375 nucleotides has been determined using the rapid and simple 'plus and minus' method. The sequence identifies many of the features responsible for the production of the proteins of the nine known genes of the organism, including initiation and termination sites for the proteins and RNAs. Two pairs of genes are coded by the same region of DNA using different reading frames.

THE genome of bacteriophage Φ X174 is a single-stranded, circular DNA of approximately 5,400 nucleotides coding for nine known proteins. The order of these genes, as determined by genetic techniques¹⁻⁴, is *A-B-C-D-E-J-F-G-H*. Genes *F*, *G* and *H* code for structural proteins of the virus capsid, and gene *H* (as defined by sequence work) codes for a small basic protein

strand DNA of Φ X has the same sequence as the mRNA and, in certain conditions, will bind ribosomes so that a protected fragment can be isolated and sequenced. Only one major site was found. By comparison with the amino acid sequence data it was found that this ribosome binding site sequence coded for the initiation of the gene *G* protein¹⁵ (positions 2,362-2,413).

At this stage sequencing techniques using primed synthesis with DNA polymerase were being developed¹⁶ and Schott¹⁷ synthesised a decanucleotide with a sequence complementary to part of the ribosome binding site. This was used to prime into the intergenic region between the *F* and *G* genes, using DNA polymerase and ³²P-labelled triphosphates¹⁸. The ribo-substitution technique¹⁹ facilitated the sequence determination of the labelled DNA produced. This decanucleotide-primed system was also used to develop the plus and minus method⁷. Suitable synthetic primers are, however, difficult to prepare and as

The Nobel Prize in Chemistry 1980



Photo from the Nobel Foundation archive.

Paul Berg

Prize share: 1/2



Photo from the Nobel Foundation archive.

Walter Gilbert

Prize share: 1/4



Photo from the Nobel Foundation archive.

Frederick Sanger

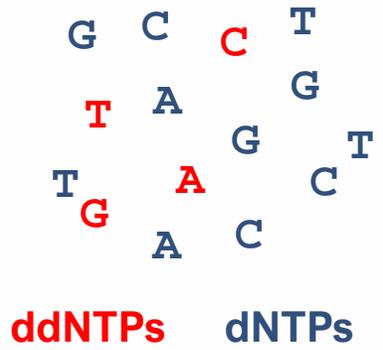
Prize share: 1/4

The Nobel Prize in Chemistry 1980 was divided, one half awarded to Paul Berg "for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA", the other half jointly to Walter Gilbert and Frederick Sanger "for their contributions concerning the determination of base sequences in nucleic acids."

Step 1 – Allestimento reazione

5' - CCTACGATGTTACGACTATACATGGCAT - 3'

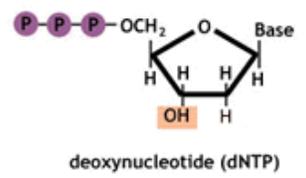
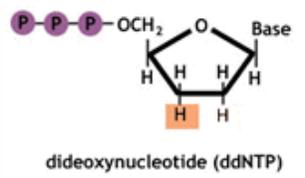
3' - GGATGCTACAATGCTGATATGTACCGTA - 5'



5' - CCTACG - 3'
5' - CCTACG - 3'
5' - CCTACG - 3'
5' - CCTACG - 3'

primer

DNA polimerasi



Step 2 – Amplificazione

5' - CCTACGATGTTACGACTATACATGGCAT - 3'

3' - GGATGCTACAATGCTGATATGTACCGTA - 5'



Denaturazione

5' - CCTACGATGTTACGACTATACATGGCAT - 3'

3' - GGATGCTACAATGCTGATATGTACCGTA - 5'

Step 2 – Amplificazione

5' - CCTACGATGTTACGACTATACATGGCAT - 3'

3' - GGATGCTACAATGCTGATATGTACCGTA - 5'



5' - CCTACGATGTTACGACTATACATGGCAT - 3'

3' - GGATGCTACAATGCTGATATGTACCGTA - 5'

5' - CCTACG - 3'

5' - CCTACG - 3'

primer

Step 2 – Amplificazione

5' - CCTACGATGTTACGACTATACATGGCAT - 3'

3' - GGATGCTACAATGCTGATATGTACCGTA - 5'



5' - CCTACGATGTTACGACTATACATGGCAT - 3'

3' - GGATGCTACAATGCTGATATGTACCGTA - 5'

5' - CCTACG - 3' → → →

DNA
polimerasi

I ddNTPs interrompono la reazione

5' - CCTACGATGTTACGACTATACATGGCAT - 3'

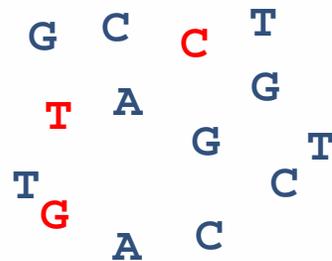
3' - GGATGCTACAATGCTGATATGTACCGTA - 5'



5' - CCTACGATGTTACGACTATACATGGCAT - 3'

3' - GGATGCTACAATGCTGATATGTACCGTA - 5'

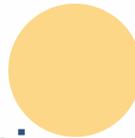
5' - CCTACGATGTT



ddNTPs

dNTPs

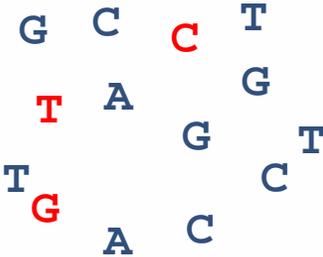
DNA
polimerasi



Quattro reazioni per quattro provette

Provetta 1

5' - CCTACGATGTTACGACTATACATGGCAT - 3'
3' - GGATGCTACAATGCTGATATGTACCGTA - 5'



ddNTPs dNTPs

5' - CCTACG - 3'
5' - CCTACG - 3'
5' - CCTACG - 3'

primer

DNA
polimerasi



Quattro reazioni per quattro provette

Provetta 1

5' - CCTACGATGTTACGACTATACATGGCAT - 3'

3' - GGATGCTACAATGCTGATATGTACCGTA - 5'



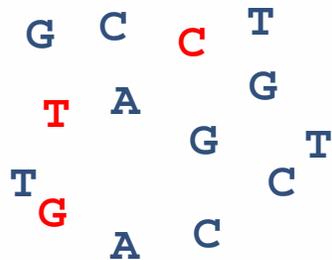
3' - GGATGCTACAATGCTGATATGTACCGTA - 5'

5' - CCTACGATGTTA

5' - CCTACGATGTTACGA

5' - CCTACGA

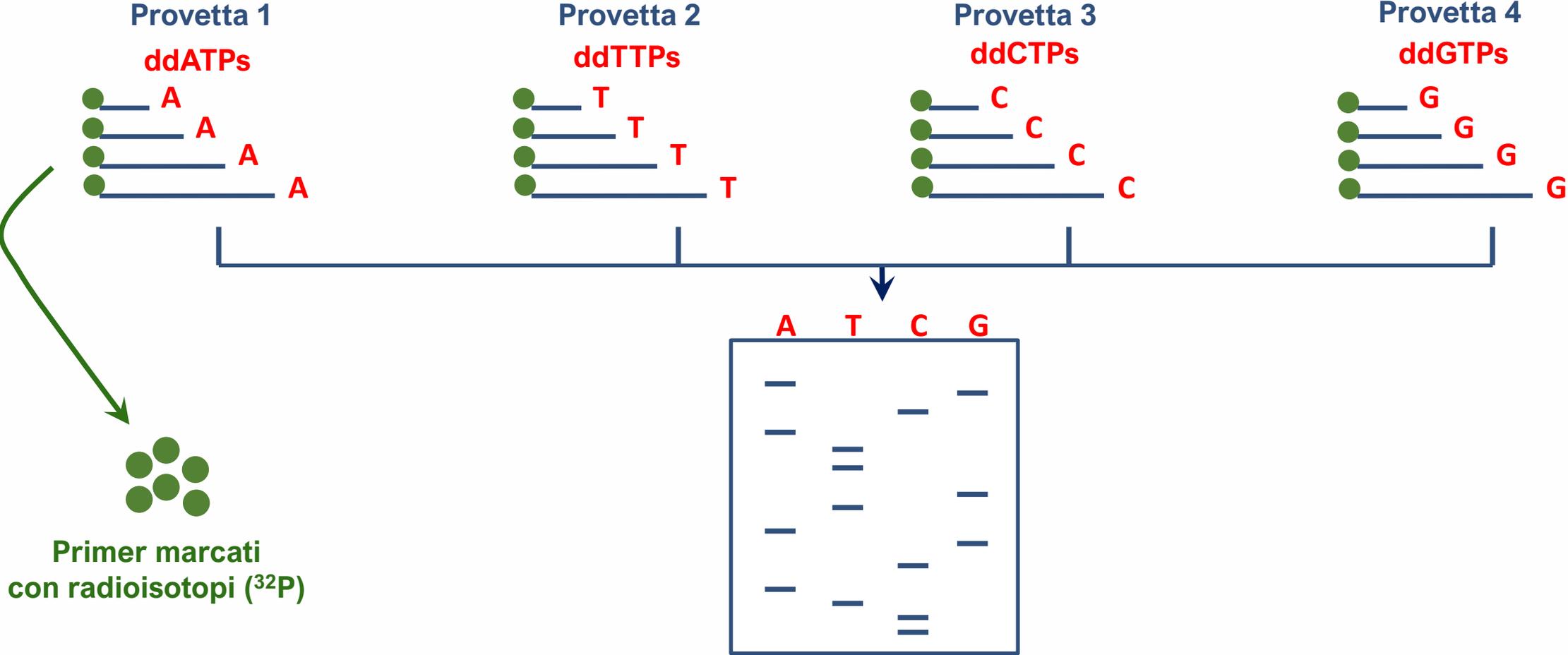
5' - CCTACGATGTTACGACTA



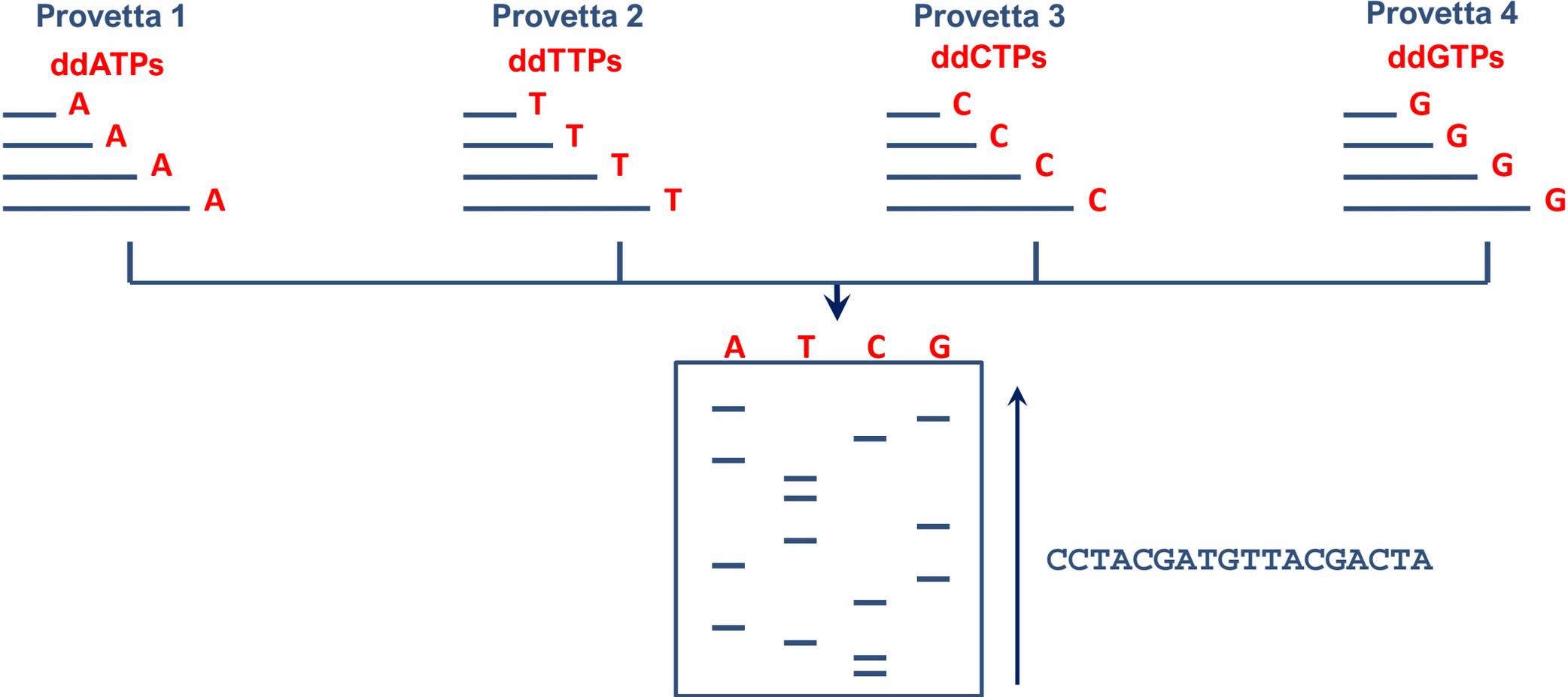
ddNTPs

dNTPs

Quattro reazioni per quattro provette

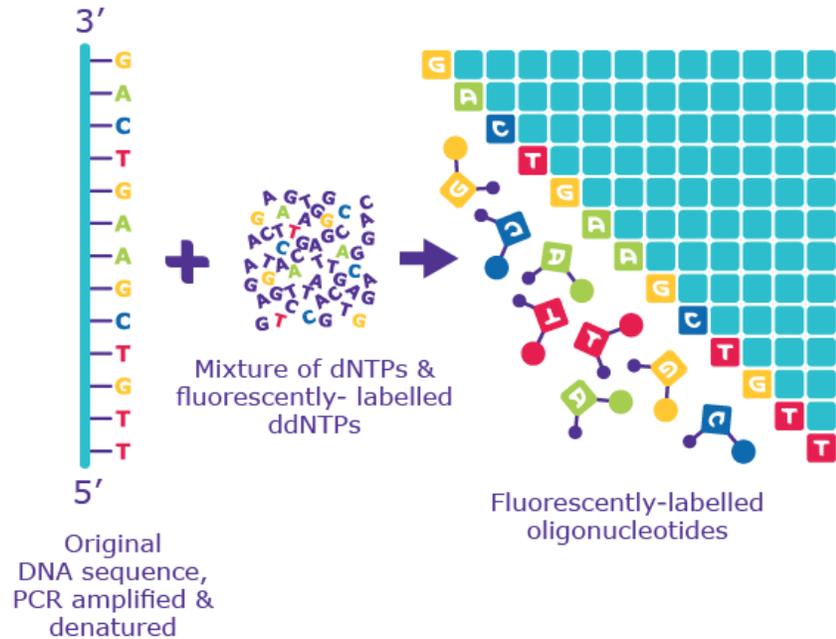


Quattro reazioni per quattro provette

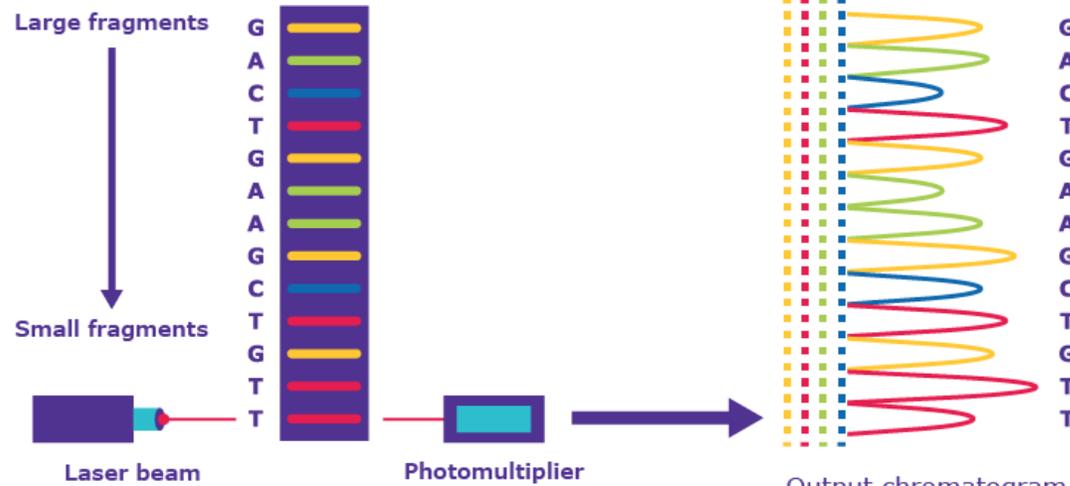


Sequenziamento automatico

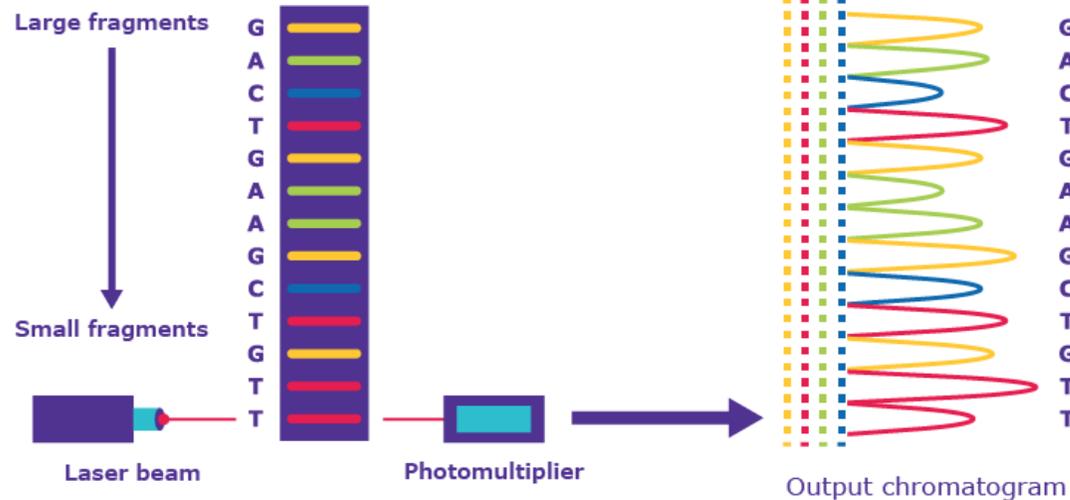
1 PCR with fluorescent, chain-terminating ddNTPs



2 Size separation by capillary gel electrophoresis

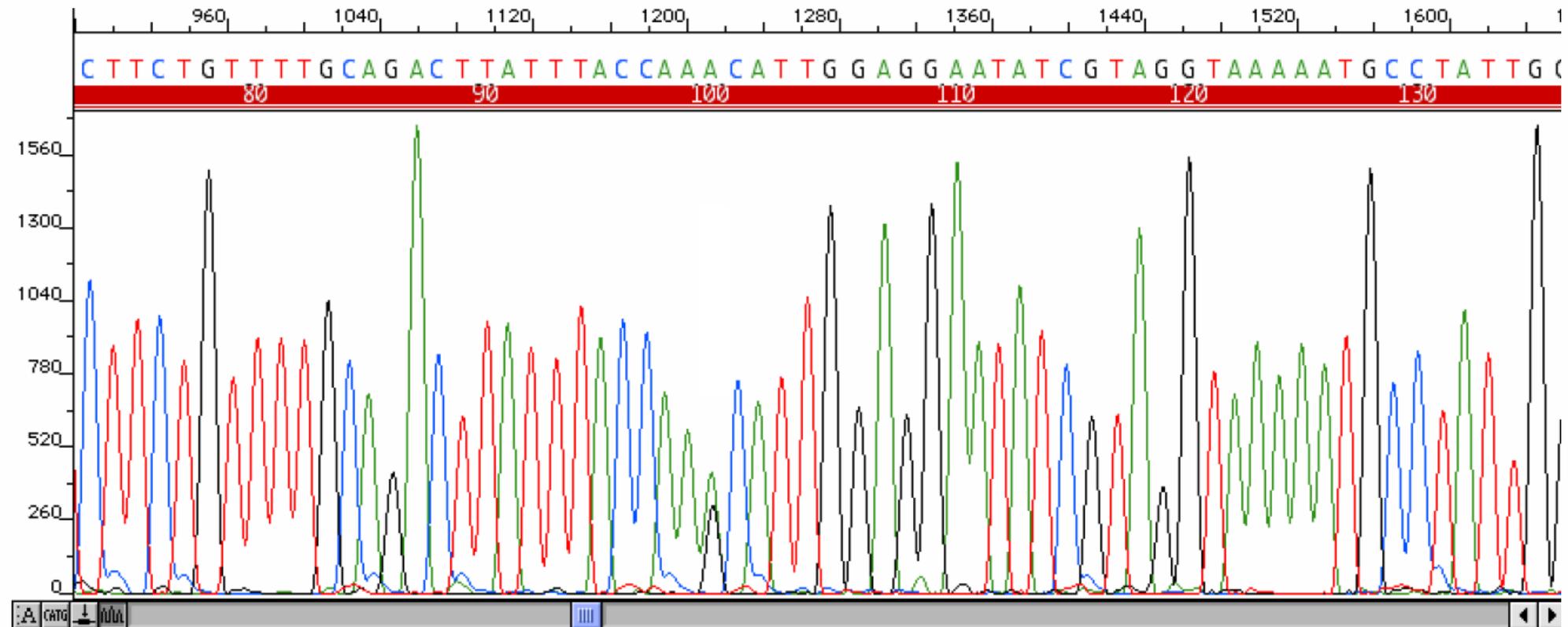


3 Laser excitation & detection by sequencing machine



Credit: Sigma Aldrich

Elettroferogramma

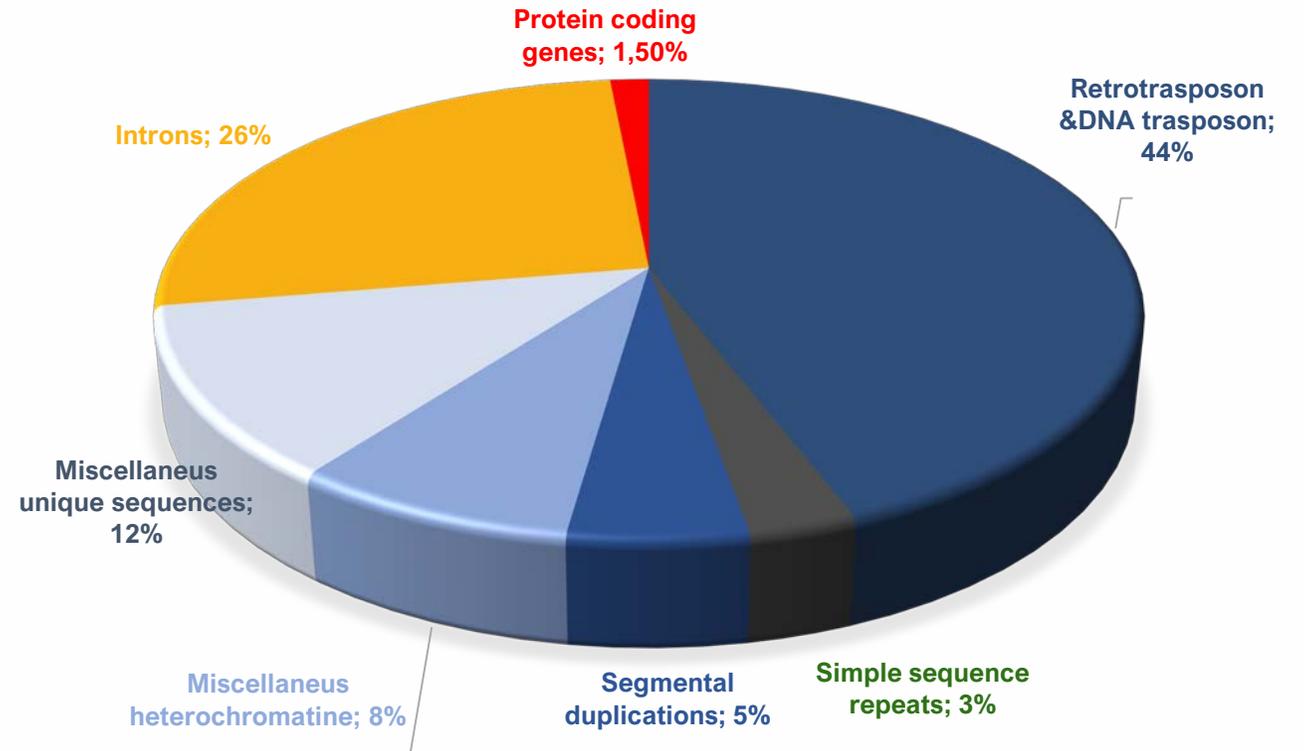


Credit: S. Volorio, Sequencing Unit Cogentech

L'impatto del Progetto Genoma Umano

Che cosa è stato scoperto?

- Il genoma umano è lungo circa 3.2 miliardi di basi
- Possediamo tanto DNA da andare e tornare dal Sole per 41 volte!
- Possediamo «solo» 21.000 geni



Adapted from T. R. Gregory Nat Rev Genet. 9:699-708, 2005

Sviluppo di piattaforme tecnologiche

1. Strutture informatiche per «**depositare**» le sequenze
2. Sviluppo di **algoritmi/software** per analizzare i dati
3. Nascita di **nuove discipline** (genomica comparativa, farmacogenomica, bioinformatica)



- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts into NCBI databases



Download

Transfer NCBI data to your computer



Learn

Find help documents, attend a class or watch a tutorial



Develop

Use NCBI APIs and code libraries to build applications



Analyze

Identify an NCBI tool for your data analysis task



Research

Explore NCBI research and collaborative projects



Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

Find SRA datasets in the cloud using BigQuery Taxonomy Analysis tables!

27 Apr 2020

Now that the Sequence Read Archive (SRA) is publicly available on the cloud

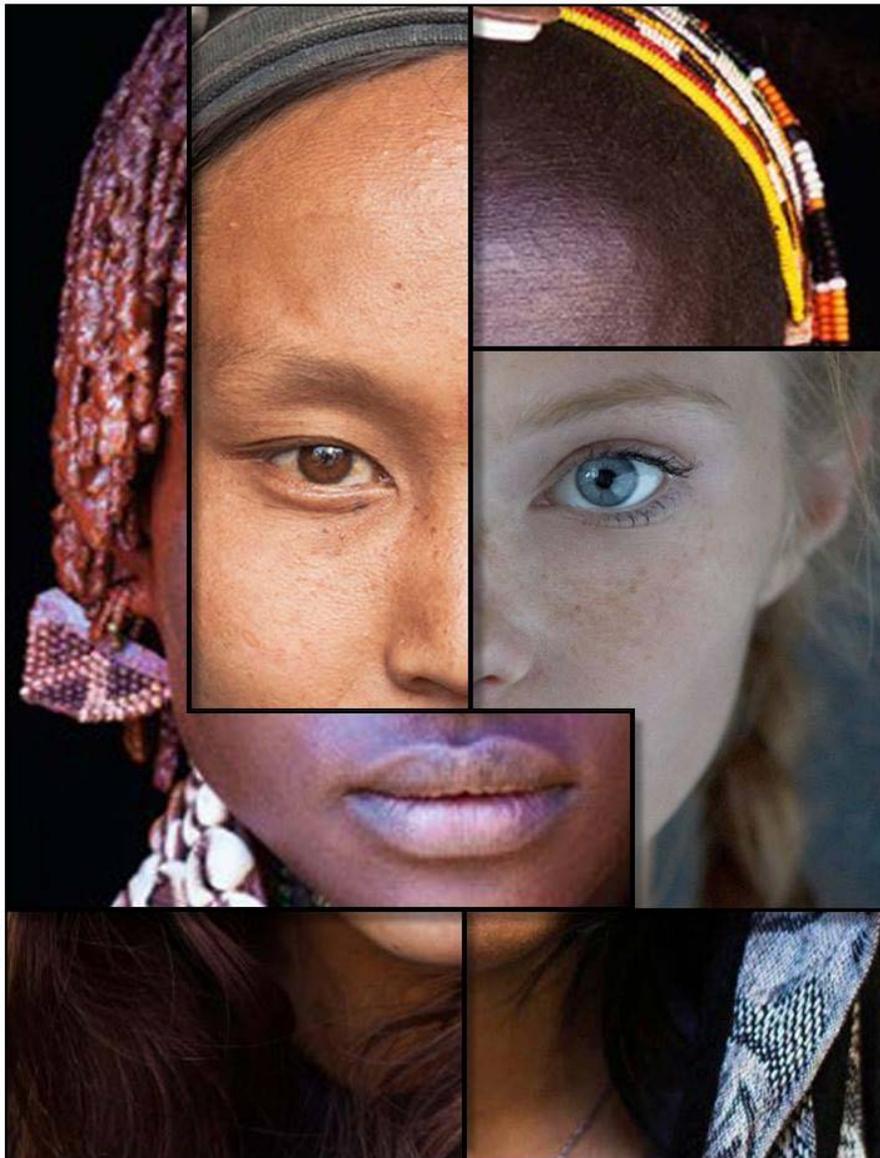
New feature added to Primer-BLAST to better design primers for expression assays

24 Apr 2020

Così lontani, così vicini



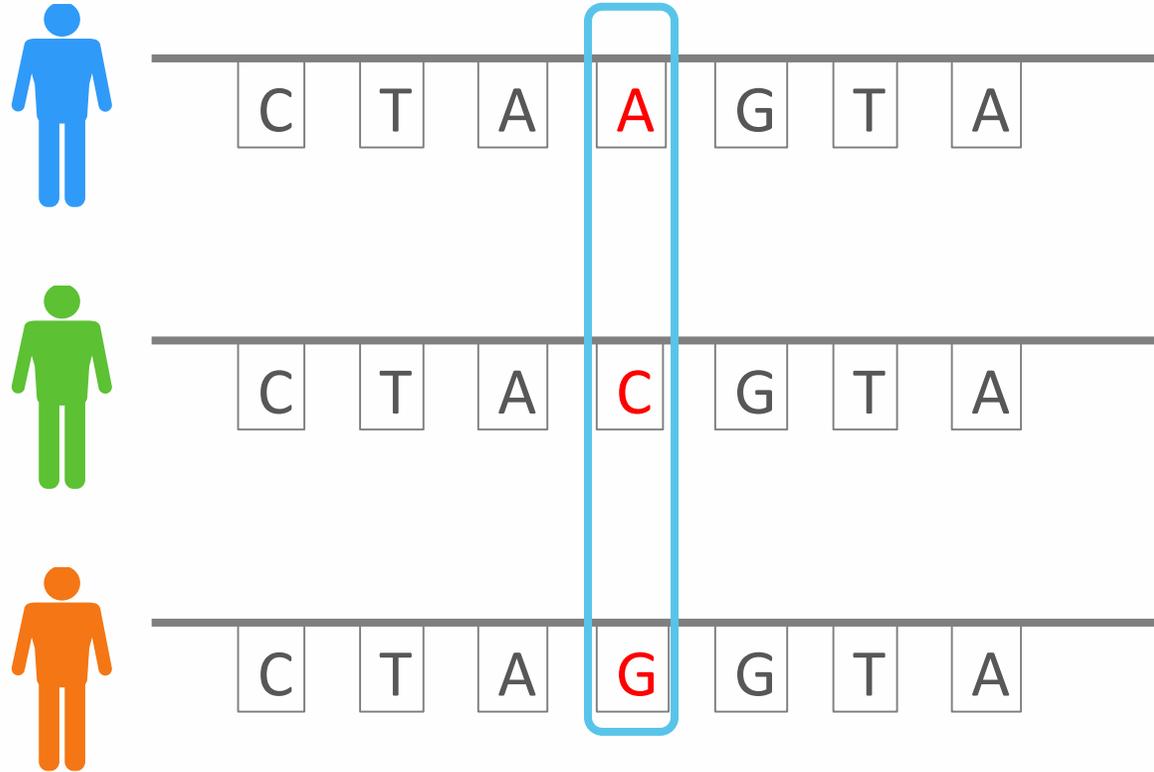
- Sequenziato genoma di scimpanzé nel 2005
- Genoma dello scimpanzé e dell'uomo IDENTICI per il 96%
- 29% delle proteine sono identiche



Nello **0.3%** del nostro genoma è racchiusa la **diversità umana**

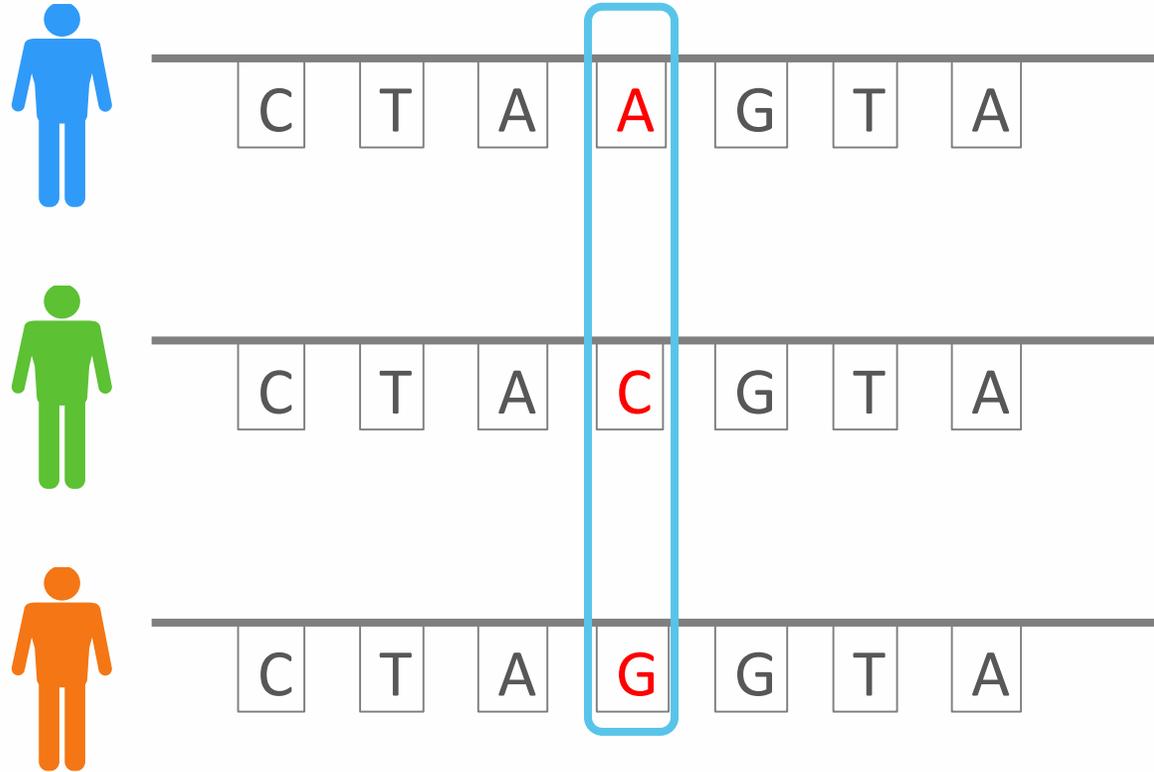
Credit: <https://www.facebook.com/ManifestodellaDiversitaUmana/>

Variazioni nel genoma



Single Nucleotide Polymorphism (SNP)

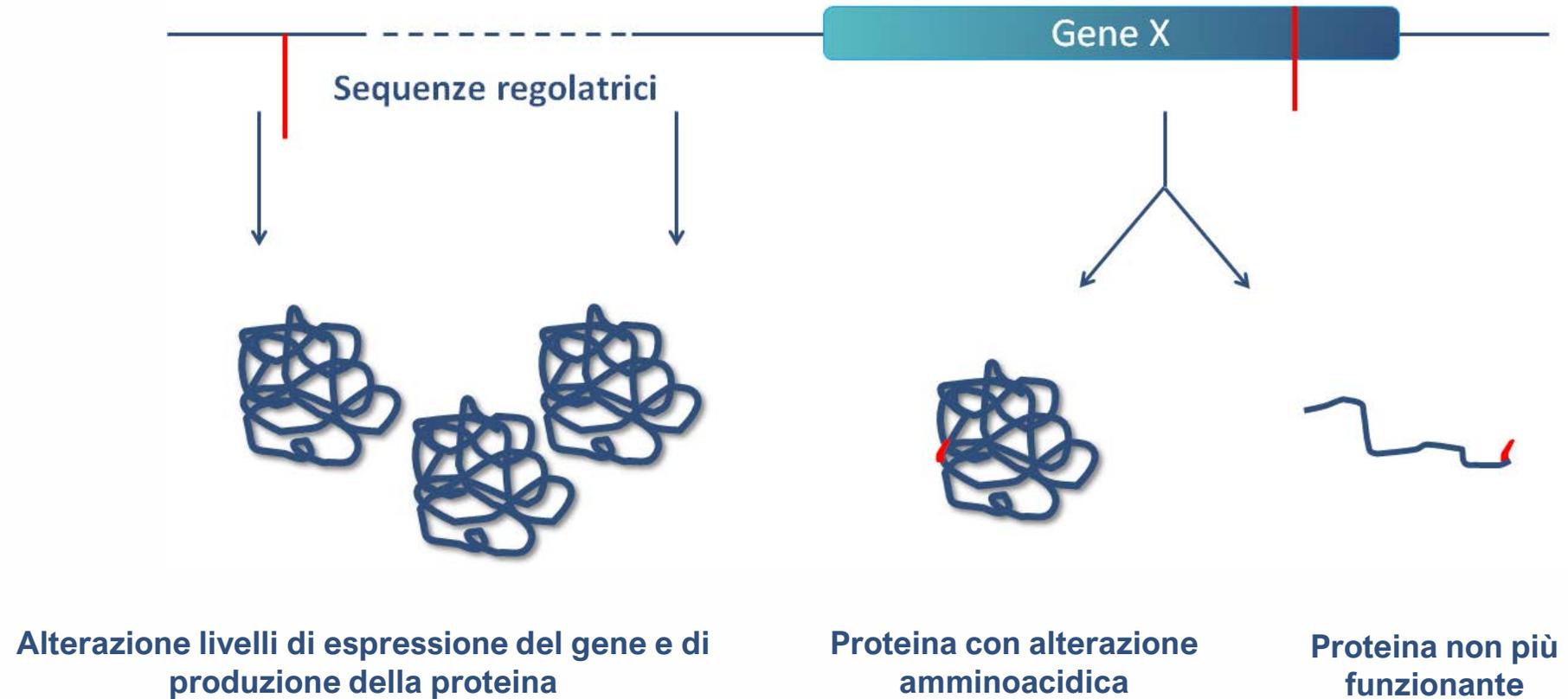
Variazioni nel genoma



- 1 SNP ogni 300 paia di basi
- Circa 10 milioni SNPs

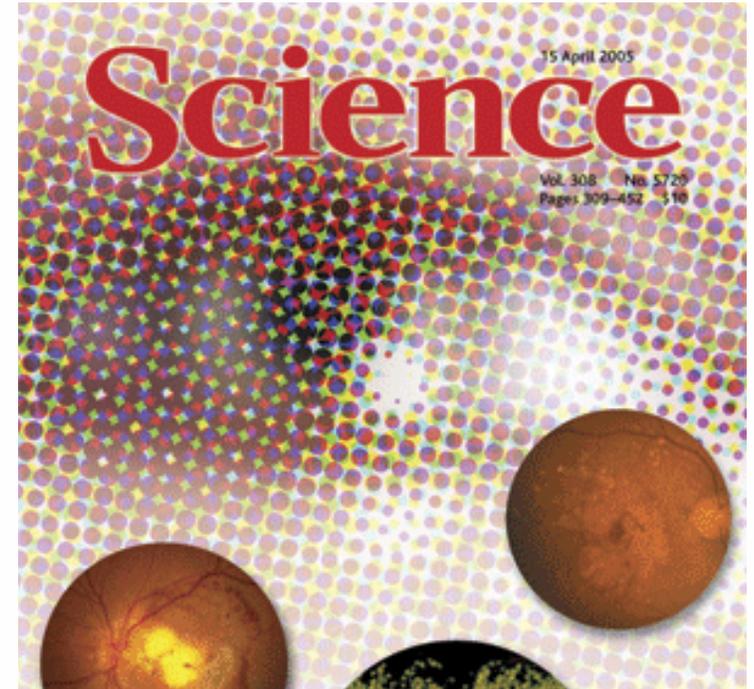
Single Nucleotide Polymorphism (SNP)

Che significato biologico hanno?



Genome wide association studies

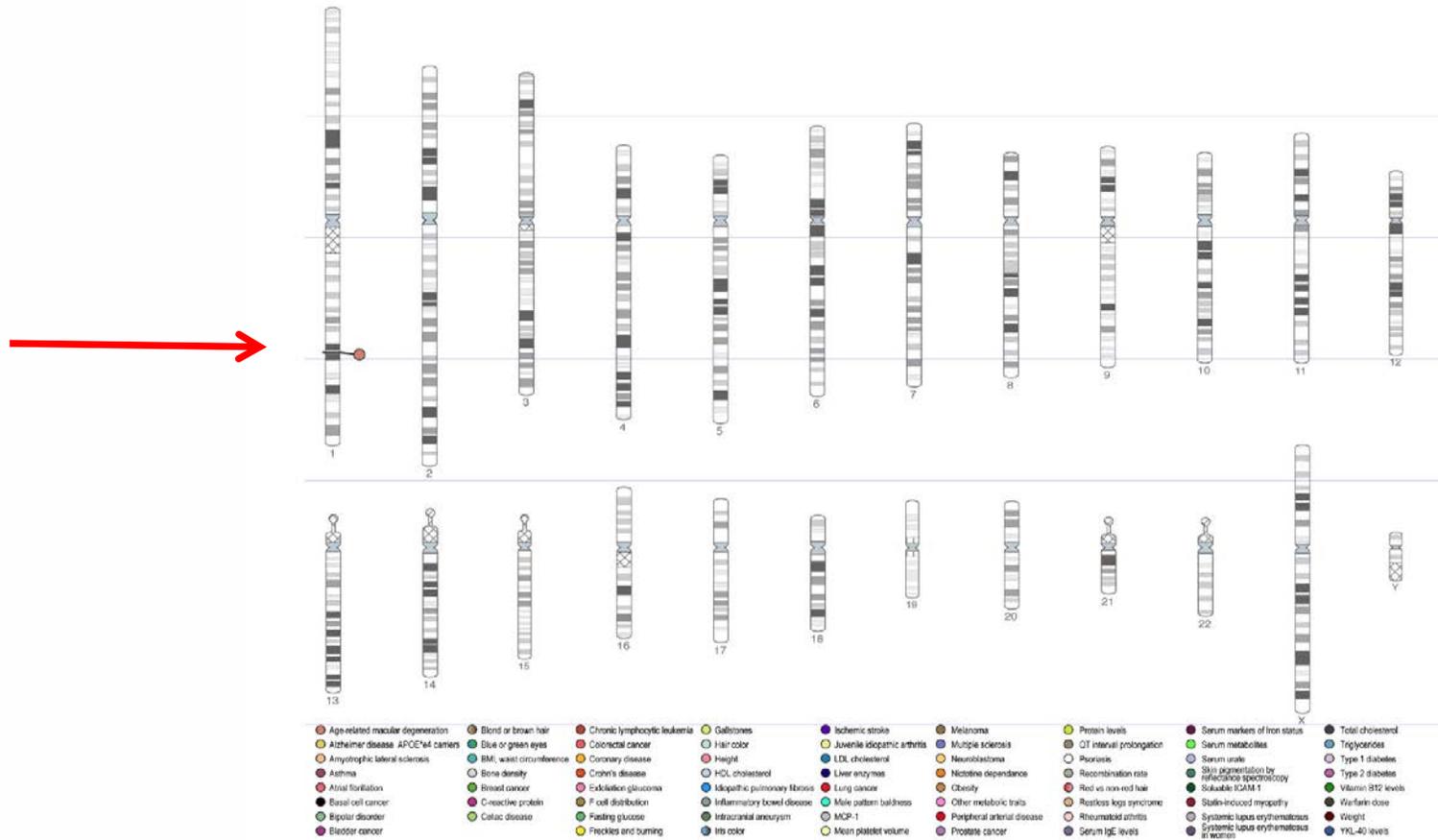
- Analisi su larga scala di genomi di individui diversi alla ricerca di **varianti geniche associate allo sviluppo di una certa patologia (GWAS)**
- **Consortio HapMap** identifica una variante genica che predispone a un tipo di degenerazione maculare (2005)



Complement Factor H Polymorphism and Age-Related Macular Degeneration

Albert O. Edwards,^{1*} Robert Ritter III,¹ Kenneth J. Abel,²
Alisa Manning,³ Carolien Panhuysen,^{3,6} Lindsay A. Farrer^{3,4,5,6,7}

Genome wide association studies (2005)



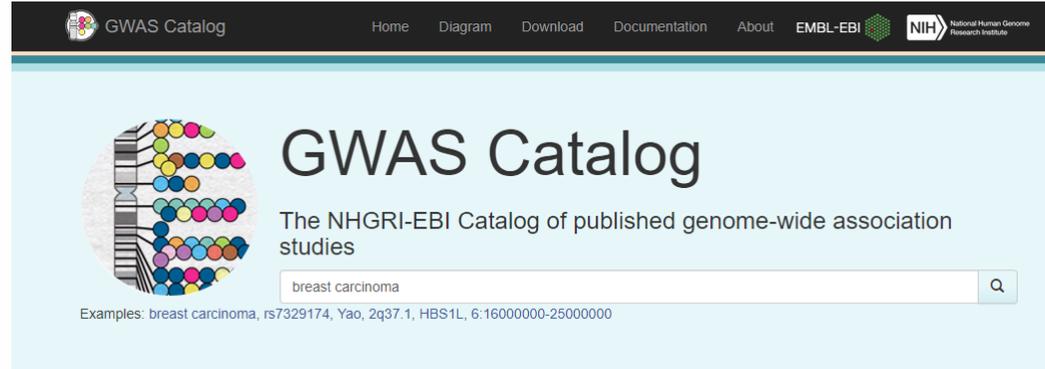
Credit: genome.gov

Genome wide association studies (2012)



Credit: <https://www.ebi.ac.uk/gwas/diagram>

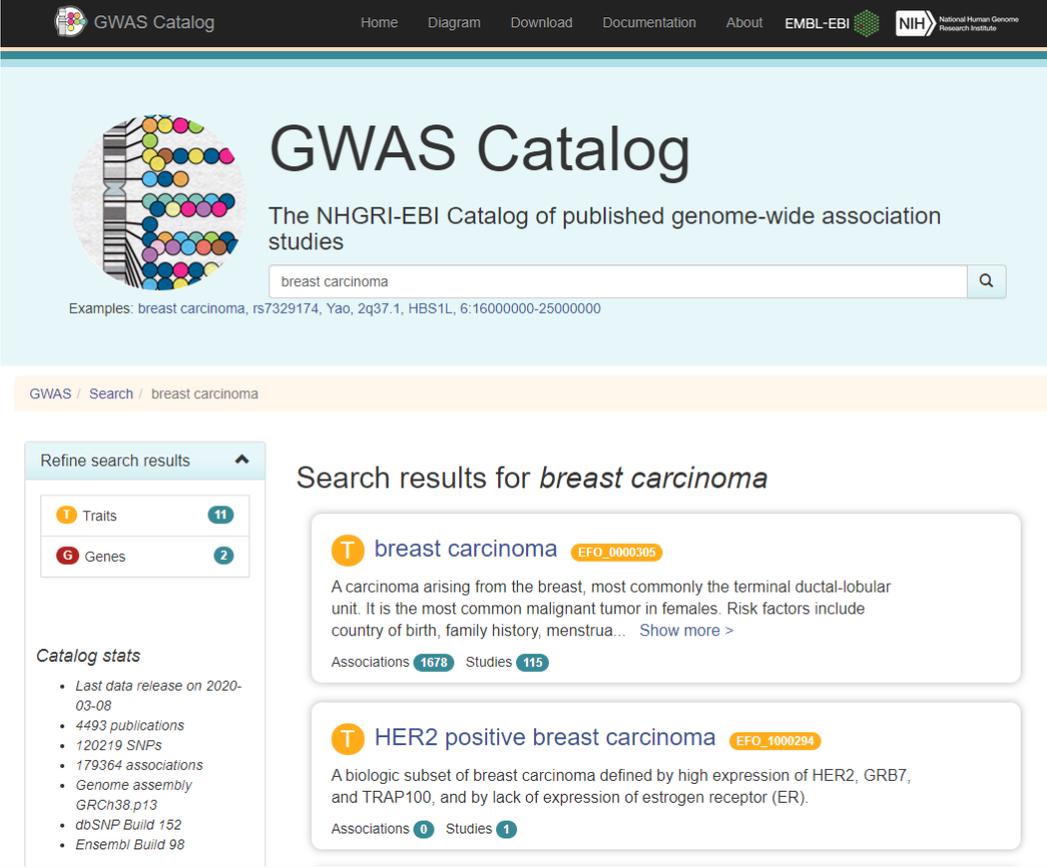
Un catalogo sempre aggiornato



The screenshot shows the GWAS Catalog website homepage. At the top, there is a navigation bar with the following links: Home, Diagram, Download, Documentation, About, EMBL-EBI, and NIH National Human Genome Research Institute. The main content area features a circular graphic on the left composed of colorful dots representing genetic markers. To the right of this graphic, the text reads "GWAS Catalog" in a large font, followed by "The NHGRI-EBI Catalog of published genome-wide association studies" in a smaller font. Below this text is a search bar containing the text "breast carcinoma" and a search icon. At the bottom of the search bar, there are examples: "Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000".

<https://www.ebi.ac.uk/gwas/>

Un catalogo sempre aggiornato



The screenshot displays the GWAS Catalog website interface. At the top, there is a navigation bar with links for Home, Diagram, Download, Documentation, About, EMBL-EBI, and NIH. The main header features the GWAS Catalog logo and the text "The NHGRI-EBI Catalog of published genome-wide association studies". A search bar contains the query "breast carcinoma" and a search button. Below the search bar, there are examples of search terms: "Examples: breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000".

Below the search bar, the breadcrumb trail reads "GWAS / Search / breast carcinoma".

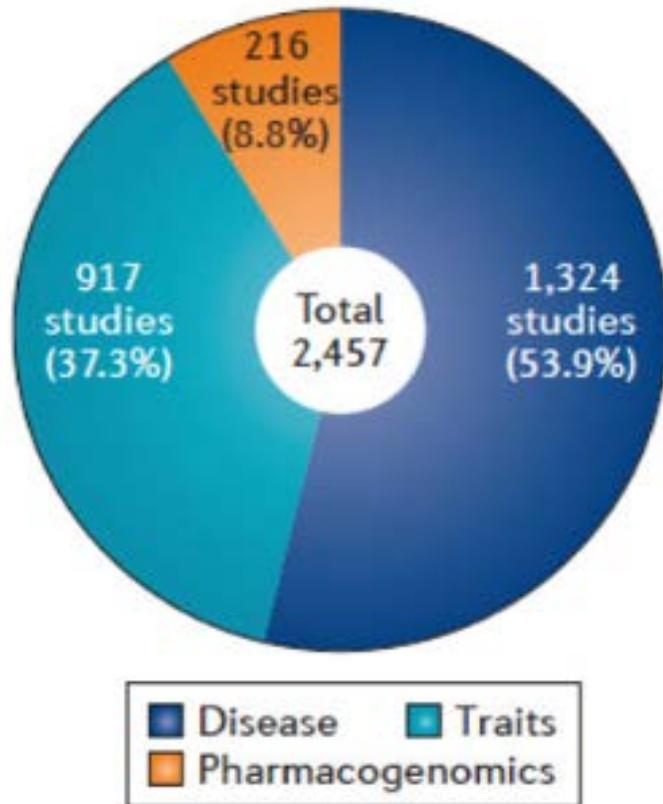
The search results are displayed in a list format. The first result is "breast carcinoma" (EFO_0000305). The description states: "A carcinoma arising from the breast, most commonly the terminal ductal-lobular unit. It is the most common malignant tumor in females. Risk factors include country of birth, family history, menstrua... Show more >". The statistics for this result are "Associations 1678" and "Studies 115".

The second result is "HER2 positive breast carcinoma" (EFO_1000294). The description states: "A biologic subset of breast carcinoma defined by high expression of HER2, GRB7, and TRAP100, and by lack of expression of estrogen receptor (ER).". The statistics for this result are "Associations 0" and "Studies 1".

On the left side of the search results, there is a "Refine search results" panel. It shows two filters: "Traits" with 11 results and "Genes" with 2 results. Below this panel, there is a "Catalog stats" section with the following information:

- Last data release on 2020-03-08
- 4493 publications
- 120219 SNPs
- 179364 associations
- Genome assembly GRCh38.p13
- dbSNP Build 152
- Ensembl Build 98

<https://www.ebi.ac.uk/gwas/search?query=breast%20carcinoma/>



La maggior parte degli studi Genome-wide in 3 categorie

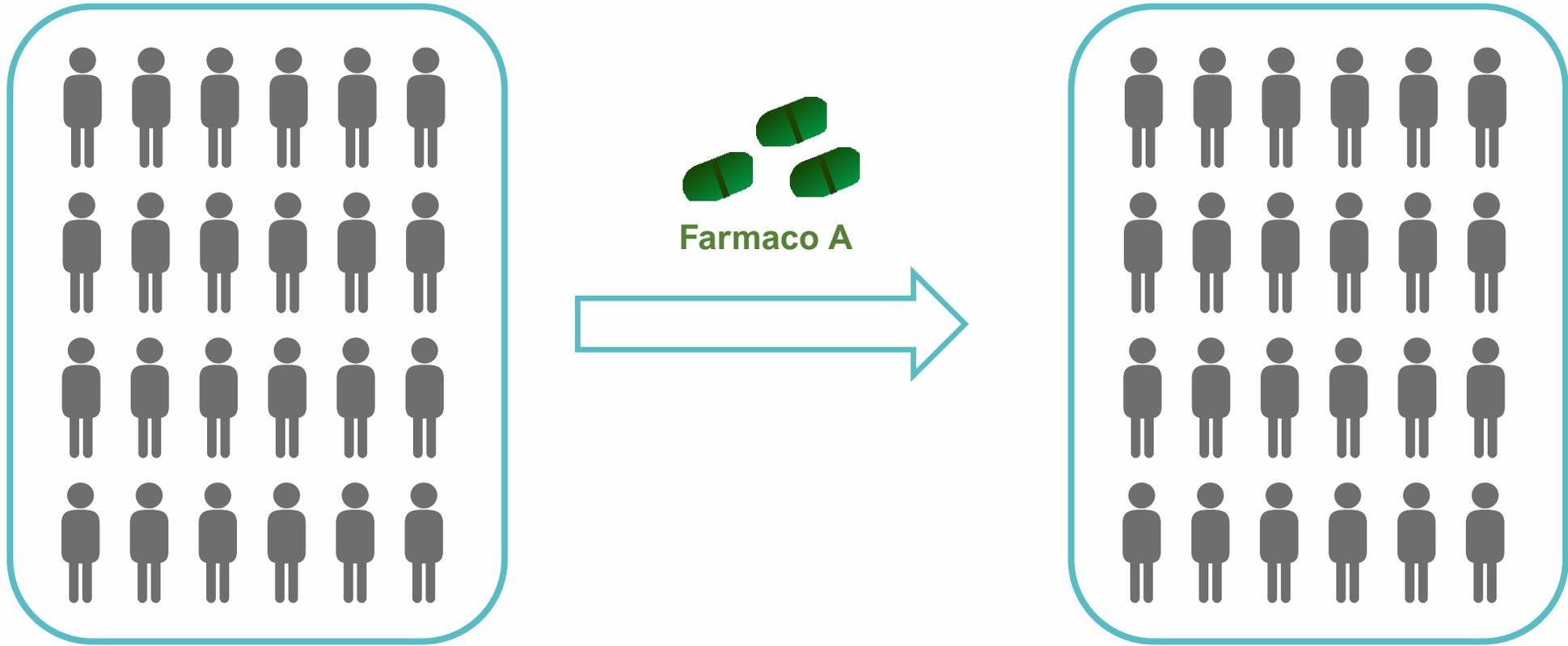
From: Giacomini et al., Nat Rev Drug Discov. 2016 16(1): 1

Medicina Personalizzata

Un cambio di paradigma

- Passaggio dal “**one size fits all**” (lo stesso trattamento per tutti) a “**personalized medicine**”
- La genomica consente di operare classificazioni molecolari e di scegliere il trattamento più efficace alla dose terapeutica più indicata per quello specifico paziente, per una specifica malattia caratterizzata dal punto di vista molecolare

Lo stesso farmaco per tutti



Farmaco A efficace nel 20% della popolazione
80% dei pazienti non avrà beneficio dal trattamento

Scoprire il profilo molecolare dei pazienti



Somministrare il Farmaco A solo ai pazienti che hanno il profilo molecolare adeguato e che risponderanno alla terapia

Un farmaco per ognuno



Tattamento sulla base del profilo molecolare del paziente

Un esempio: HER-2 e tumore al seno

1987

ARTICLES

Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene

DJ Slamon, GM Clark, SG Wong, WJ Levin, A Ullrich, WL McGuire

+ See all authors and affiliations

Science 09 Jan 1987:
Vol. 235, Issue 4785, pp. 177-182
DOI: 10.1126/science.3798106

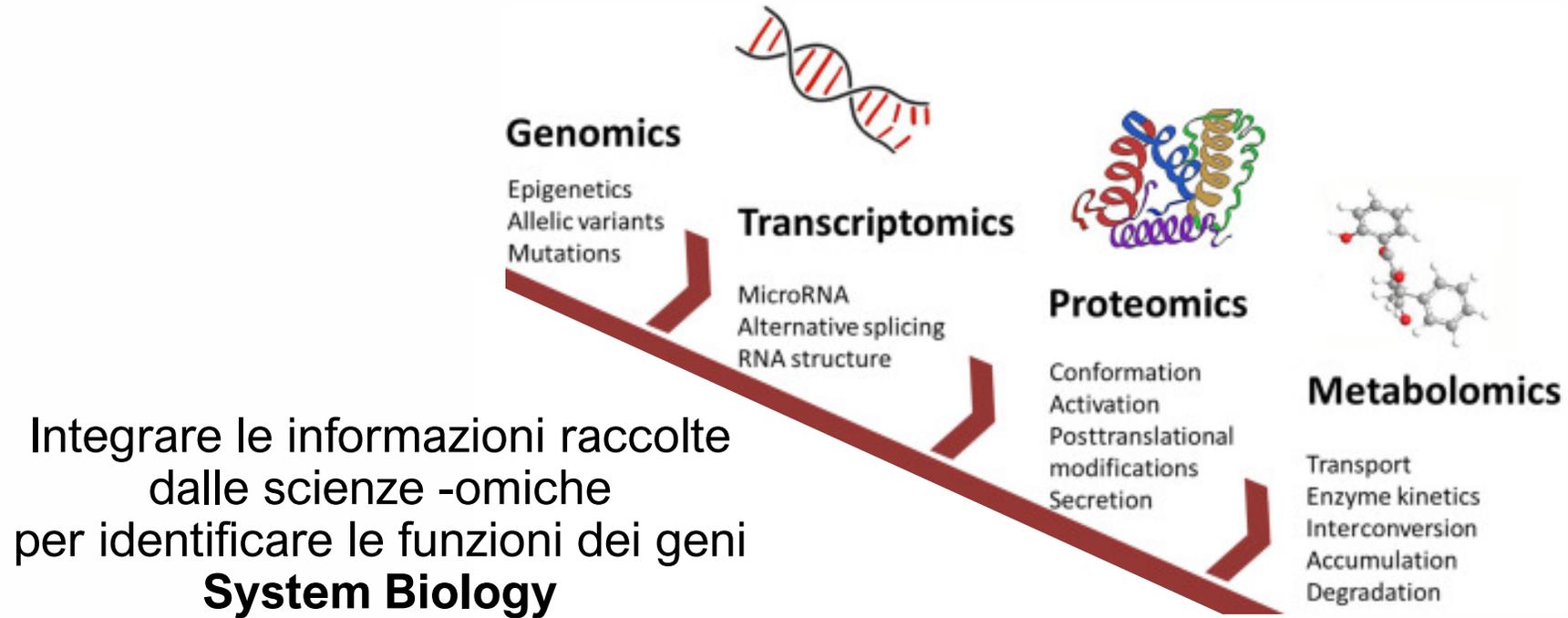
2006

**Food and Drug Administration approva il primo
farmaco molecolare
anticorpo monoclonale contro HER-2**

L' impatto

- Miglioramento dei tassi di sopravvivenza di **oltre il 30%** (tumori stadio 1-3)
- Test per positività a HER-2 entrato **di routine nella diagnosi molecolare** dei tumori alla mammella
- Il farmaco molecolare utilizzato anche **in altri tipi di cancro** (quelli con over-espressione di HER-2)

La sfida che ci attende



**consulta tutte le risorse didattiche su
www.fondazione mediasorin.it**