Marco Seri
Direttore Scientifico
IRCCS Azienda Ospedaliero-Universitaria di Bologna

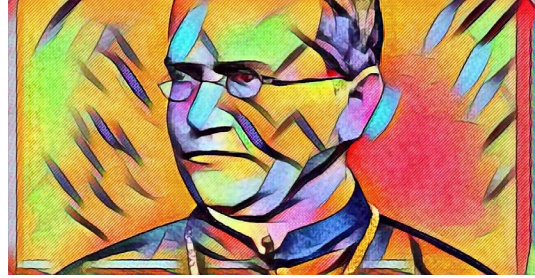
marco.seri@aosp.bo.it

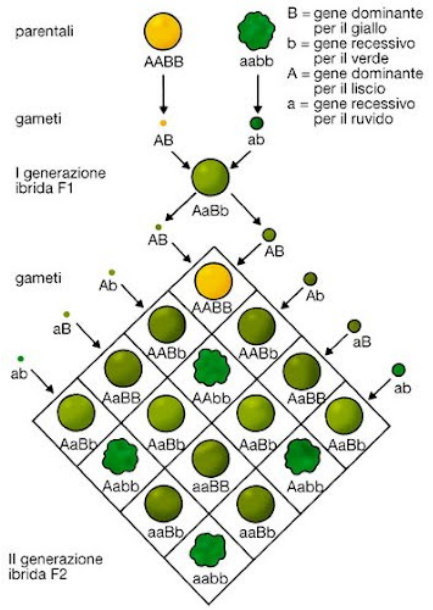
- La rivoluzione dell’era Genomica

- L’effetto del progetto genoma umano a livello clinico: l’approccio allo studio delle malattie genetiche rare tramite esoma

- L’utilizzo del genoma nei casi ancora non diagnosticati dopo esoma

La Genetica Medica 200 anni dopo Mendel

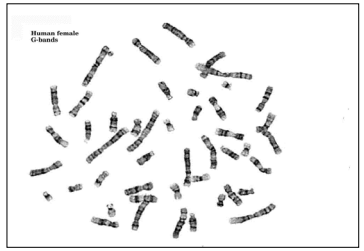


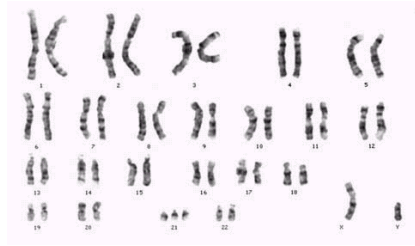


Gregor Johann *Mendel* (Hynčice, 20 luglio 1822 – Brno, 6 gennaio 1884)

A very brief history of medical genetics

1956 description of the correct chromosome number in humans

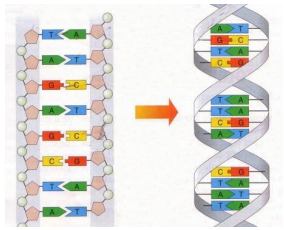




1959 discovery of a chromosome change associated with a clinical disorder (Down s.)

1902 concept of inborn errors of metabolism (alkaptonuria)

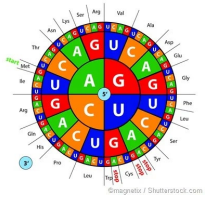
1953 structure of DNA



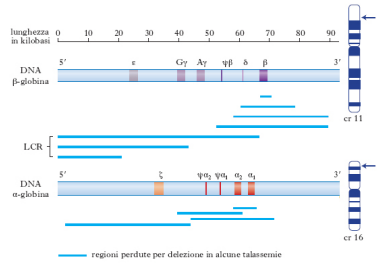
1957 single amino acid difference identified in the “sickle” hemoglobin protein

	Thr	Pro	Glu	Glu	beta^A chain
	...A C T	C C T	G **A** G	G A G...	beta^A gene
Codon #	4	5	6	7	
	...A C T	C C T	G **T** G	G A G...	beta^S gene
	Thr	Pro	**Val**	Glu	beta^S chain

1966 “cracking” of the genetic code



1977 first human genes to be cloned: chorionic somatomammotropin, α- and β-globin



La rivoluzione del Progetto Genoma Umano

- 1985 – Proposto
- 1986 - 89 - Discusso, dibattuto e pianificato
- Oct. 1, 1990 – Data ufficiale di inizio progetto
- Sept. 30, 2005 – Data presunta di completamento del progetto
- ma…….


PRESIDENT OF THE UNITED

Victor A. McKusick (1921-2008)
initiator and orchestrator of the transition between
Medical Genetics and Genetic Medicine

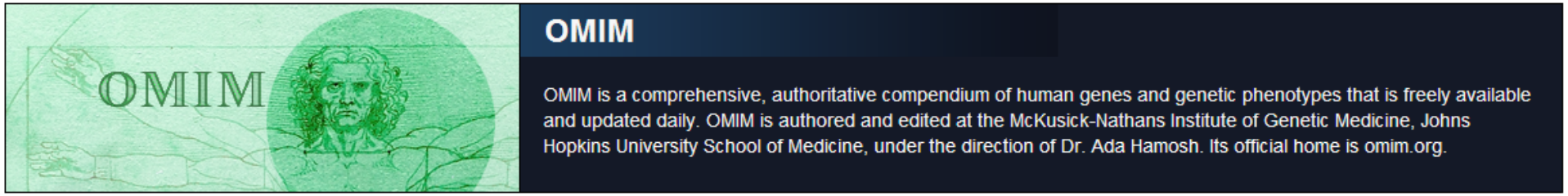
Japan Prize 2008

MENDELIAN INHERITANCE IN MAN (MIM)

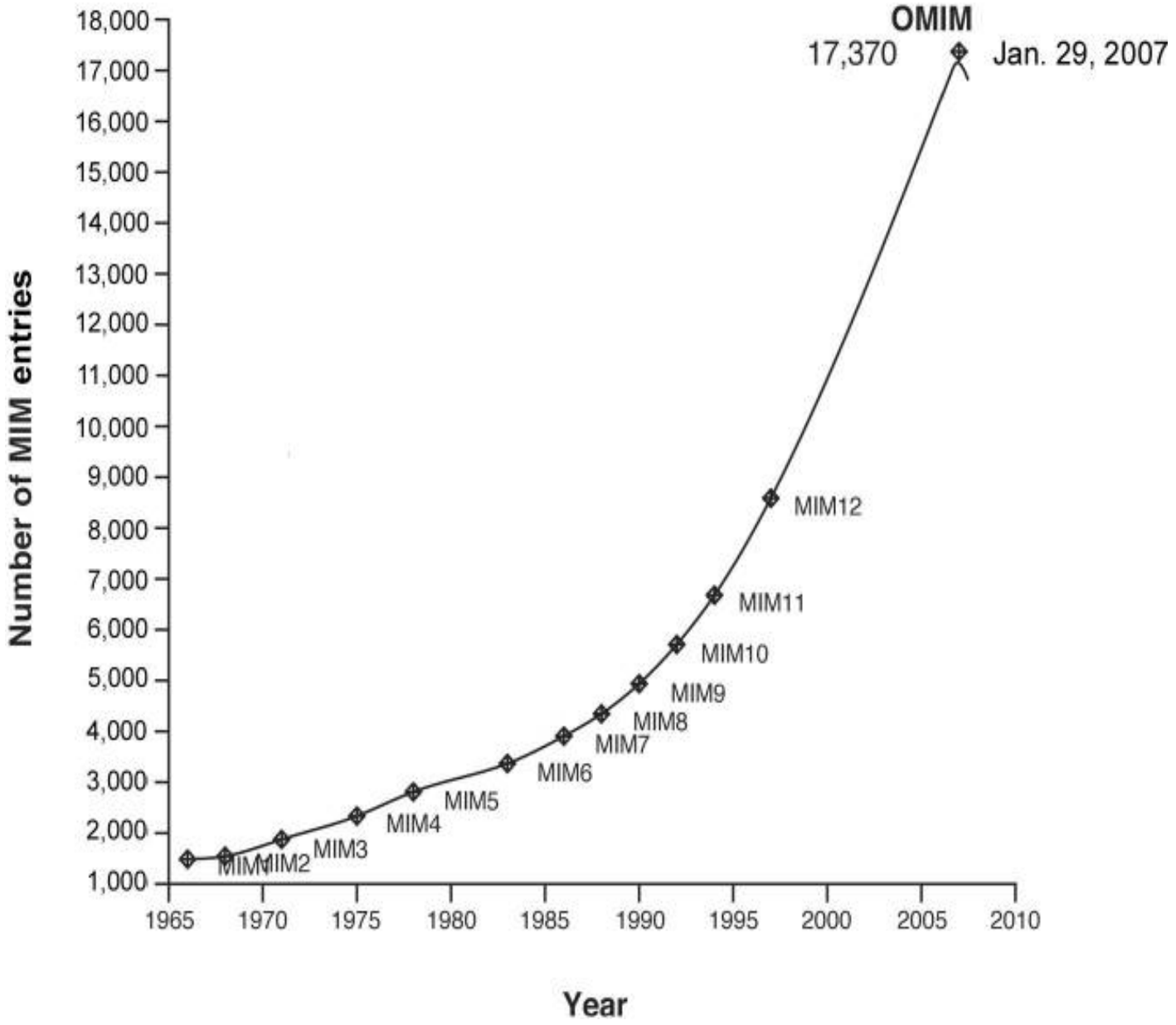
1^ edizione 1966: 1500 voci – oggi: >26.000

Dal 1998 solo in formato elettronico: On-lineMIM

http://www.ncbi.nlm.nih.gov/omim



Un incremento esponenziale di informazioni



OMIM Entry Statistics

Number of Entries in OMIM (Updated May 8th, 2024) :

MIM Number Prefix	Autosomal	X Linked	Y Linked	Mitochondrial	Totals
Gene description *	16,407	770	51	37	17,265
Gene and phenotype, combined +	21	0	0	0	21
Phenotype description, molecular basis known #	6,420	387	5	34	6,846
Phenotype description or locus, molecular basis unknown %	1,388	109	4	0	1,501
Other, mainly phenotypes with suspected mendelian basis	1,634	100	3	0	1,737
Totals	25,870	1,366	63	71	27,370

OMIM Gene Map Statistics

OMIM Morbid Map Scorecard (Updated May 8th, 2024) :

Total number of phenotypes* for which the molecular basis is known	7,530
Total number of genes with phenotype-causing mutation	4,910

* Phenotypes include (1) single-gene mendelian disorders and traits; (2) susceptibilities to cancer and complex disease (e.g., BRCA1 and familial breast-ovarian cancer susceptibility, 113705.0001, and CFH and macular degeneration, 134370.0008); (3) variations that lead to abnormal but benign laboratory test values ("nondiseases") and blood groups (e.g., lactate dehydrogenase B deficiency, 150100.0001 and ABO blood group system, 110300.0001); and (4) select somatic cell genetic disease (e.g., GNAS and McCune-Albright syndrome, 139320.0008 and IDH1 and glioblastoma multiforme, 147700.0001.)

Distribution of Phenotypes across Genes (Updated May 8th, 2024) :

Number of genes with 1 phenotype	3,439
Number of genes with 2 phenotypes	893
Number of genes with 3 phenotypes	325
Number of genes with 4+ phenotypes	253

Dissected OMIM Morbid Map Scorecard (Updated May 8th, 2024) :

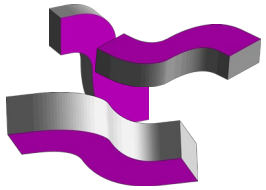
Class of phenotype	Phenotype	Gene *
Single gene disorders and traits	6,468	4,545
Susceptibility to complex disease or infection	680	503
"Nondiseases"	151	118
Somatic cell genetic disease	238	131

*Some genes may be counted more than once because mutations in a gene may cause more than one phenotype and the phenotypes may be of different classes (e.g., activating somatic BRAF mutation underlying cancer, 164757.0001. and germline BRAF mutation in Noonan syndrome, 164757.0022.)

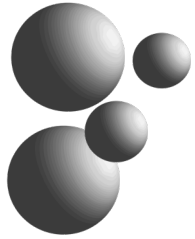
Genetic Medicine Paradigm

DNA

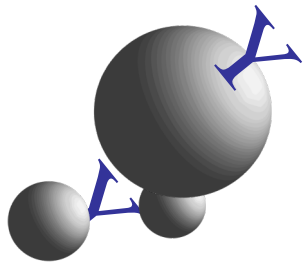
RNA



Proteins



Modified Proteins



Biological Function

Transcription

Translation

Post-Translation Modification

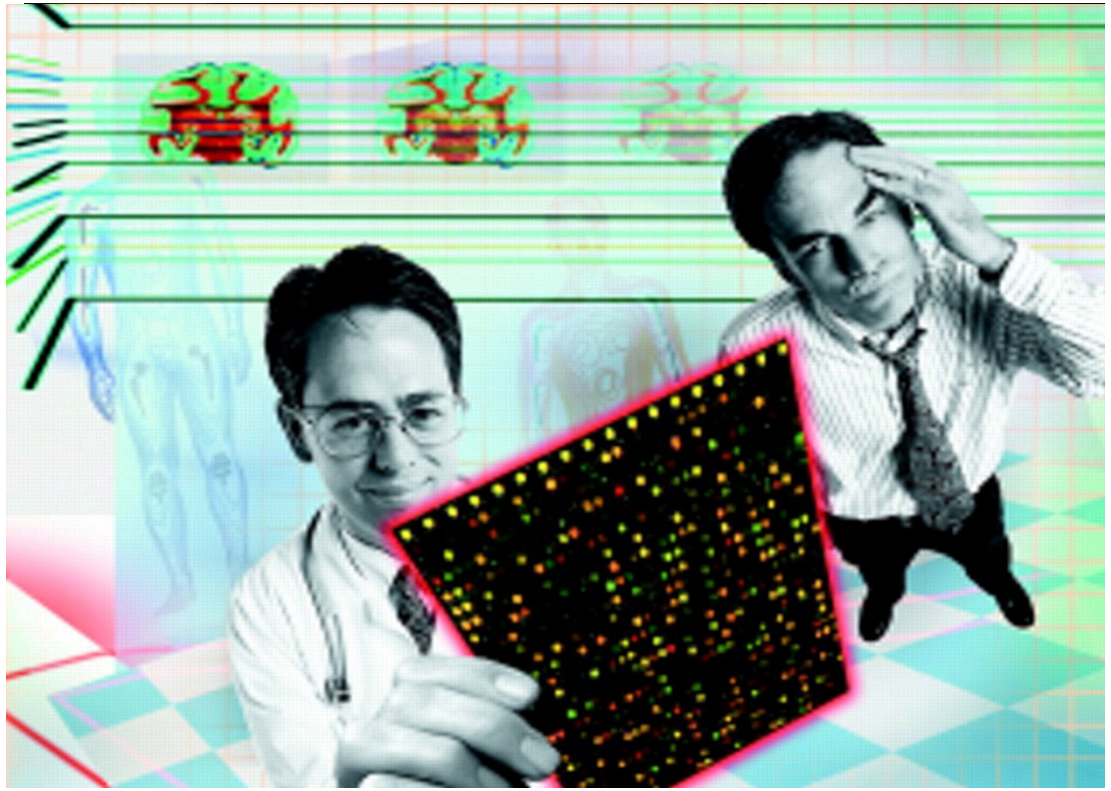
23.000

> 1,000,000 Proteins

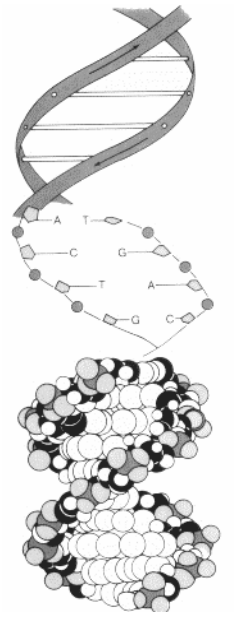
Siamo entrati nell'era post-genomica

- ***Finora le tecnologie disponibili permettevano l'analisi di singoli geni***

- ***Oggi possiamo analizzare il funzionamento di migliaia di geni***

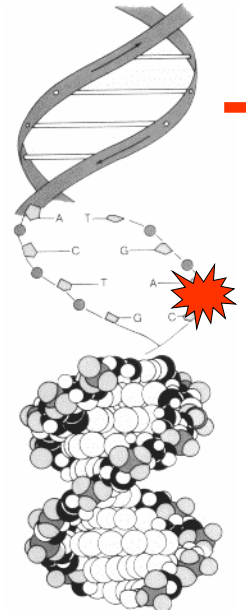


Geni e malattie: una nuova dimensione



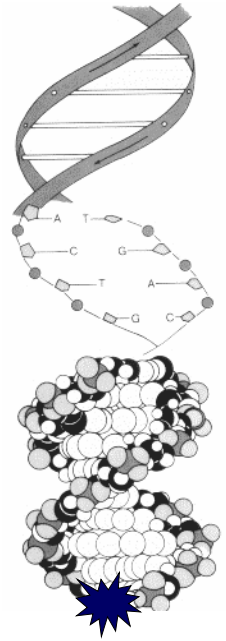
normale

Sano



variazione 1:
mutazione

Malattia
monofattoriale



Sano, ma
con una proteina
che funziona di più o
di meno: suscettibilità

Variazione 2: polimorfismo

- La rivoluzione dell’era Genomica

- L’effetto del progetto genoma umano a livello clinico: l’approccio allo studio delle malattie genetiche rare tramite esoma

- L’utilizzo del genoma nei casi ancora non diagnosticati dopo esoma

Sequenziamento personalizzato

nature Vol 452|17 April 2008|doi:10.1038/nature06884

LETTERS

The complete genome of an individual by massively parallel DNA sequencing

HUMAN GENETICS

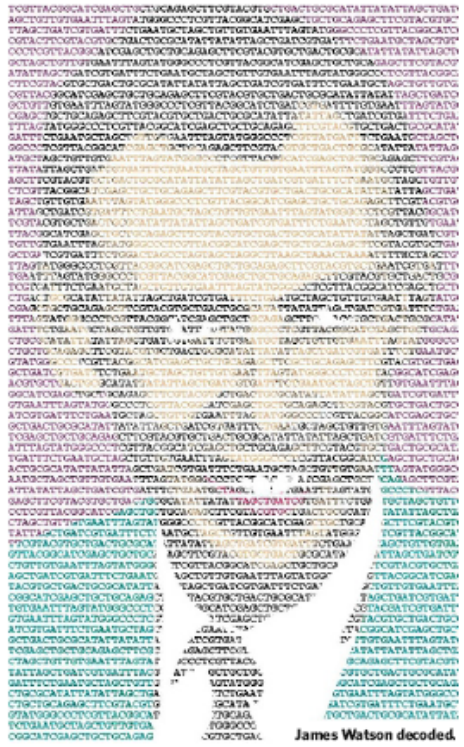
Dr Watson's base pairs

Maynard V. Olson

The application of new technology to sequence the genome of an individual yields few biological insights. Nonetheless, the feat heralds an era of 'personal genomics' based on cheap sequencing.

This issue of *Nature* contains a paper that is, in a curious way, a sequel to one published 55 years ago — the description by James Watson and Francis Crick[1] of the double-helical structure of DNA. At the information-carrying core of this beautiful structure, with its far-reaching implications for biology and medicine, are the base pairs that Watson discovered by fitting together cardboard cut-outs of the bases adenine, thymine, guanine and cytosine. Now, on page 872, Wheeler *et al.*[2] describe the use of massively parallel DNA sequencing to determine the order of the base pairs in Watson's own genome. This achievement is a technical *tour de force* that points towards routine use of whole-genome sequencing as a research tool in human genetics. Given the choice of James Watson as an identified research subject, the paper is also a conspicuous effort to publicize the arrival of the era of personal genomics and the willingness of a famous geneticist to put his genome sequence in the public domain.

Technically, the paper's interest stems from its reliance on a DNA-sequencing platform that differs greatly from the one used during the first great era of genome sequencing, which culminated in the Human Genome Project (HGP). In the HGP platform, each kilobase-pair fragment of genomic DNA was captured



James Watson decoded.

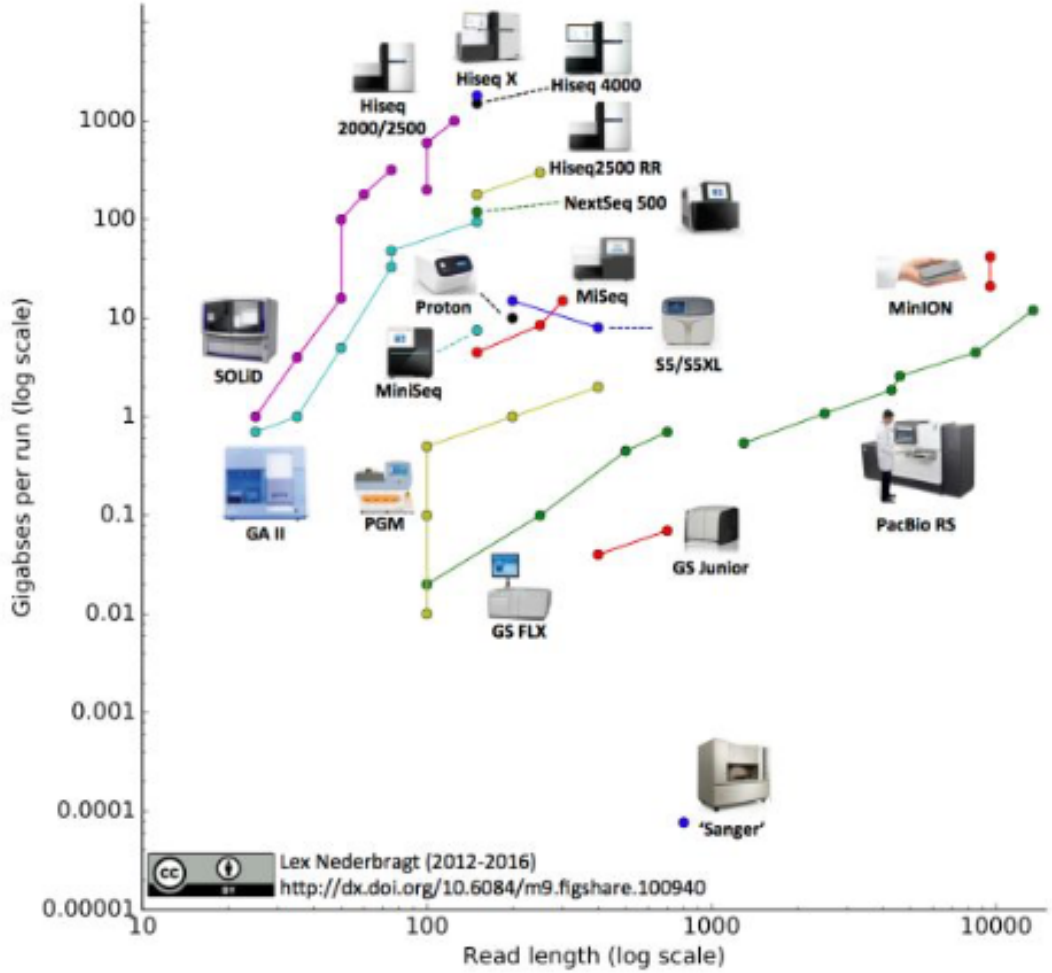
PHOTO BETTMANN/CORBIS

efficiency of the new methods lies in massive parallelization of the biochemical and measurement steps. The instruments used by Wheeler *et al.* are marketed by 454 Life Sciences, a component of Roche Diagnostics, which joined forces with the Human Genome Sequencing Center at Baylor College of Medicine in Houston, Texas, to sequence Watson's genome.

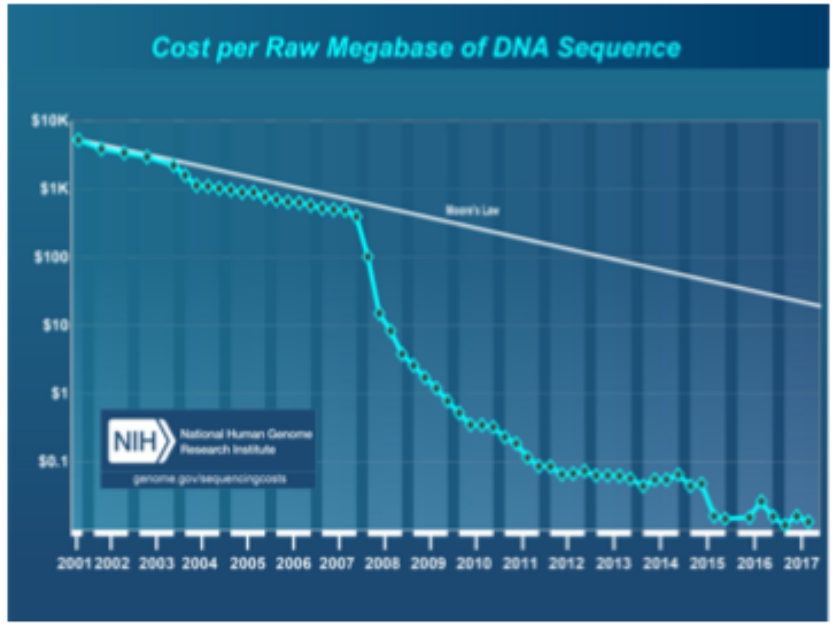
The 454 instruments achieve massive parallelization in two different ways[5]. In an initial step, single DNA molecules are attached to synthetic beads and then amplified enzymatically. During amplification, the beads are trapped in tiny water droplets within a water–oil emulsion; hence, more than 100,000 samples can be processed in parallel in a single test tube. In a later step, during which optical measurements are used to collect the actual sequencing data, each bead is confined to a picolitre-scale well etched into the end of a glass fibre within a fibre-optic bundle. Although costs have not yet dropped to the much-ballyhooed target of US$1,000 per genome[6], they are now low enough to make the era of personal genomics a reality rather than a distant dream.

What can we expect to learn from the sequences of individual genomes? The main lesson from the analyses by Wheeler *et al.* is that it will be extremely difficult to extract medically, or even

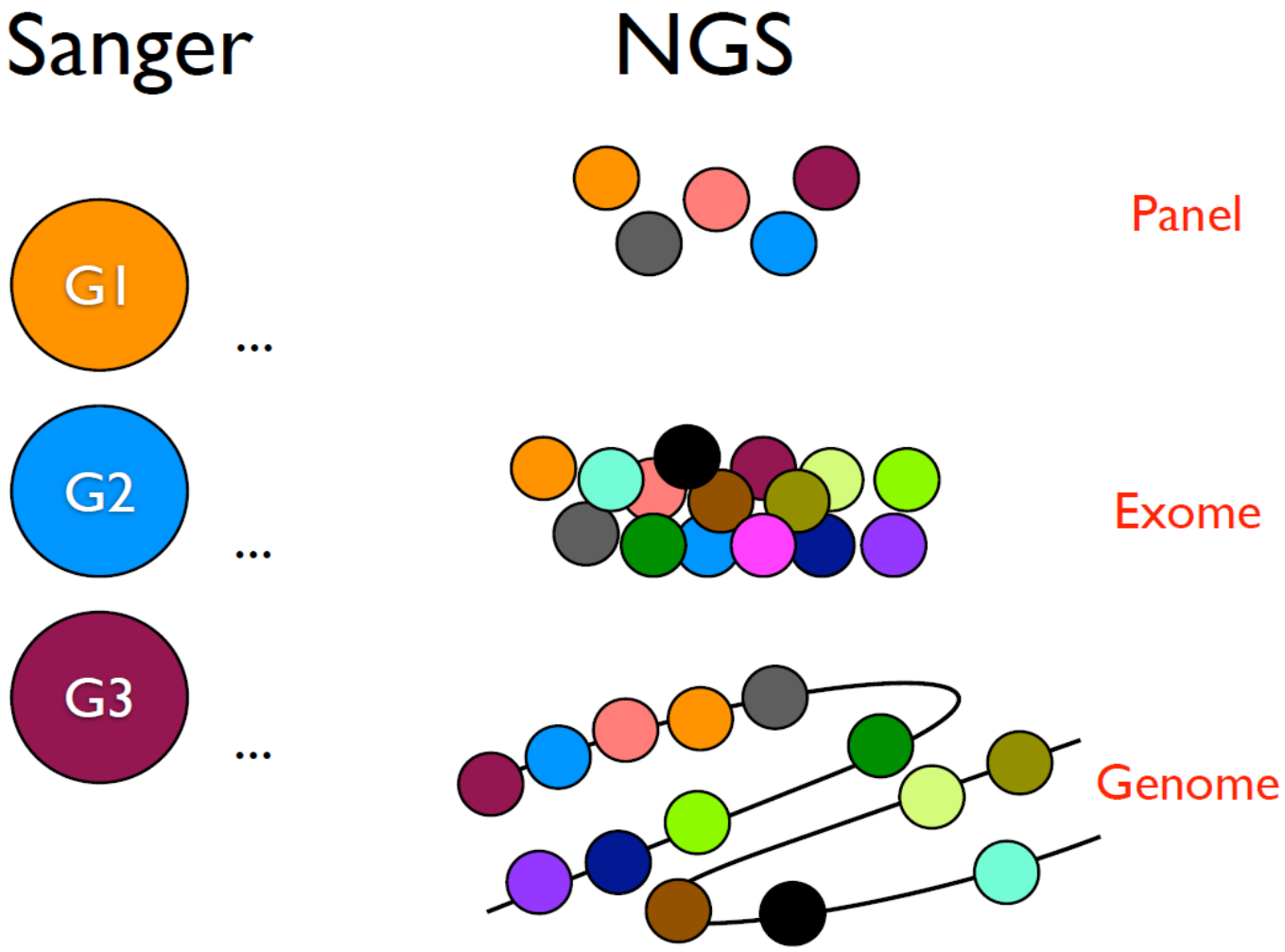
The explosion of DNA sequencing capacity



10M-fold increase in the sequencing capacity (Gigabases per run) from Sanger to today DNA sequencers



Le prospettive cliniche del Progetto Genoma Umano



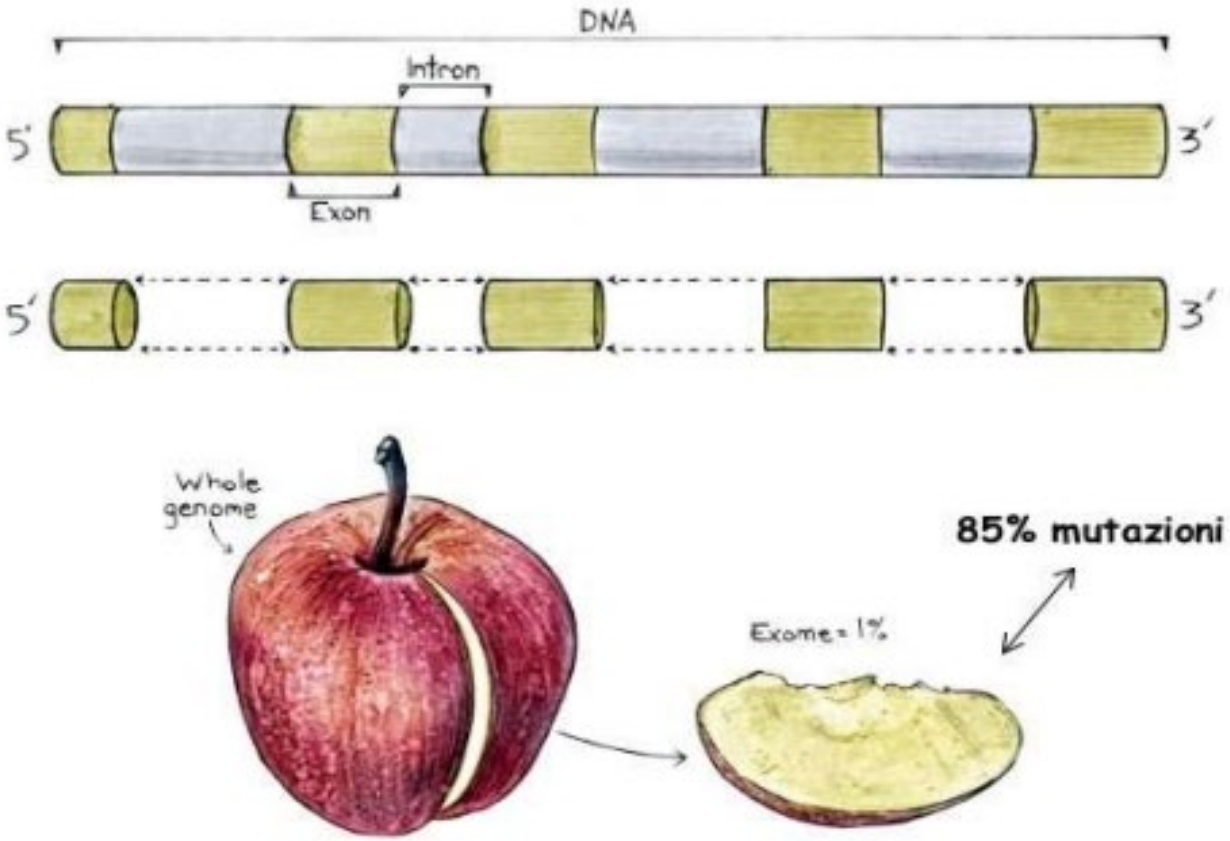
Le malattie rare

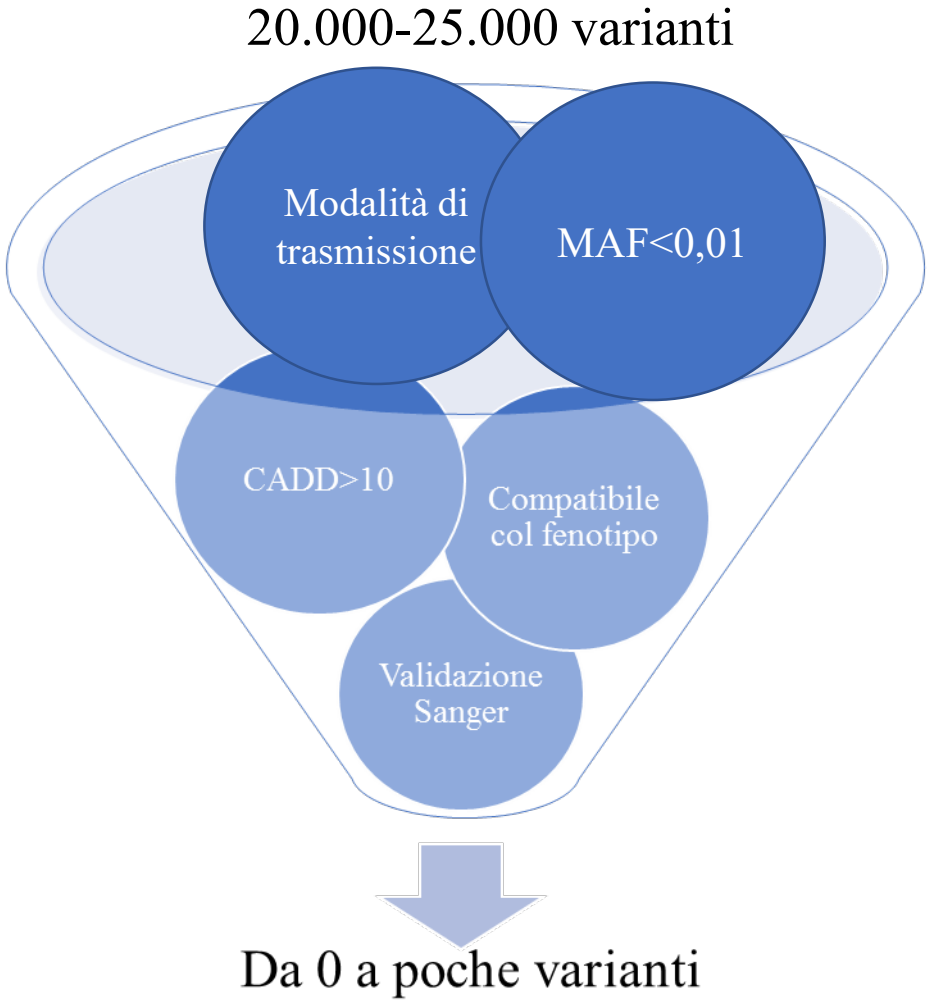
- Frequenza < 1:2.000
- Più di 7.000 descritte
- Nel loro complesso colpiscono il 6-8% della popolazione
- In Italia ne sono affette 2 milioni di persone, con 19.000 nuovi casi l'anno
- La maggior parte ha origine genetica

- ULTRARARE < 1:2.000.000
- Potrebbero esserne affette pochissime o una persona al mondo
- Conoscenze estremamente limitate
- Necessità di condivisione dei dati per la loro descrizione e definizione molecolare
- L'NGS ha rivoluzionato l'approccio diagnostico alle malattie rare

Whole Exome Sequencing (WES)

Tecnica NGS in grado di analizzare virtualmente tutte le porzioni codificanti (esoni) di un genoma umano

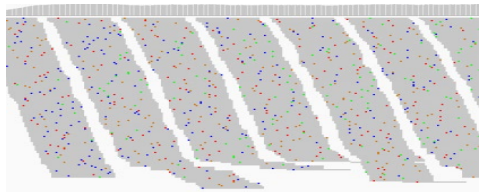




SEQUENCING

100 bp PE reads

ALIGNMENT



PROCESSING

Local realignment

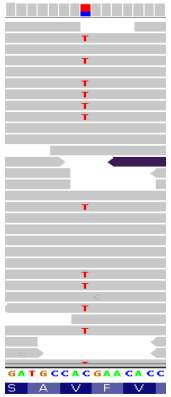
Duplicate marking

Base quality recalibration

Analysis-ready reads

PRIORITIZATION

CALLING



FILTERING

- dbSNP135
- 1KG project
- ESP project
- 500 ctrl chrs

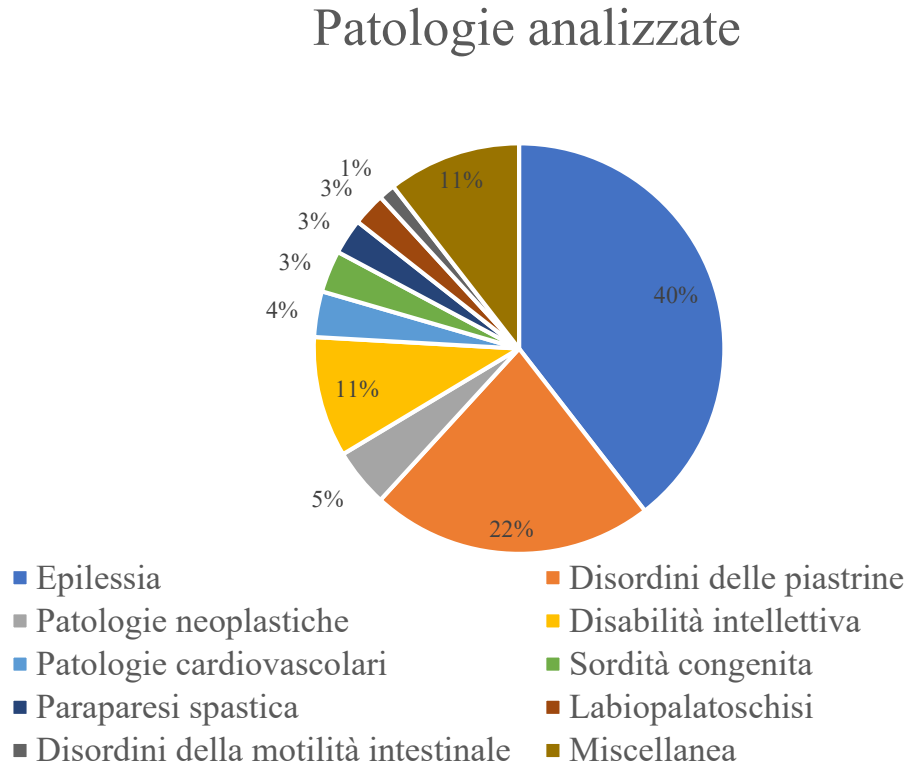
- non-coding
- synonymous

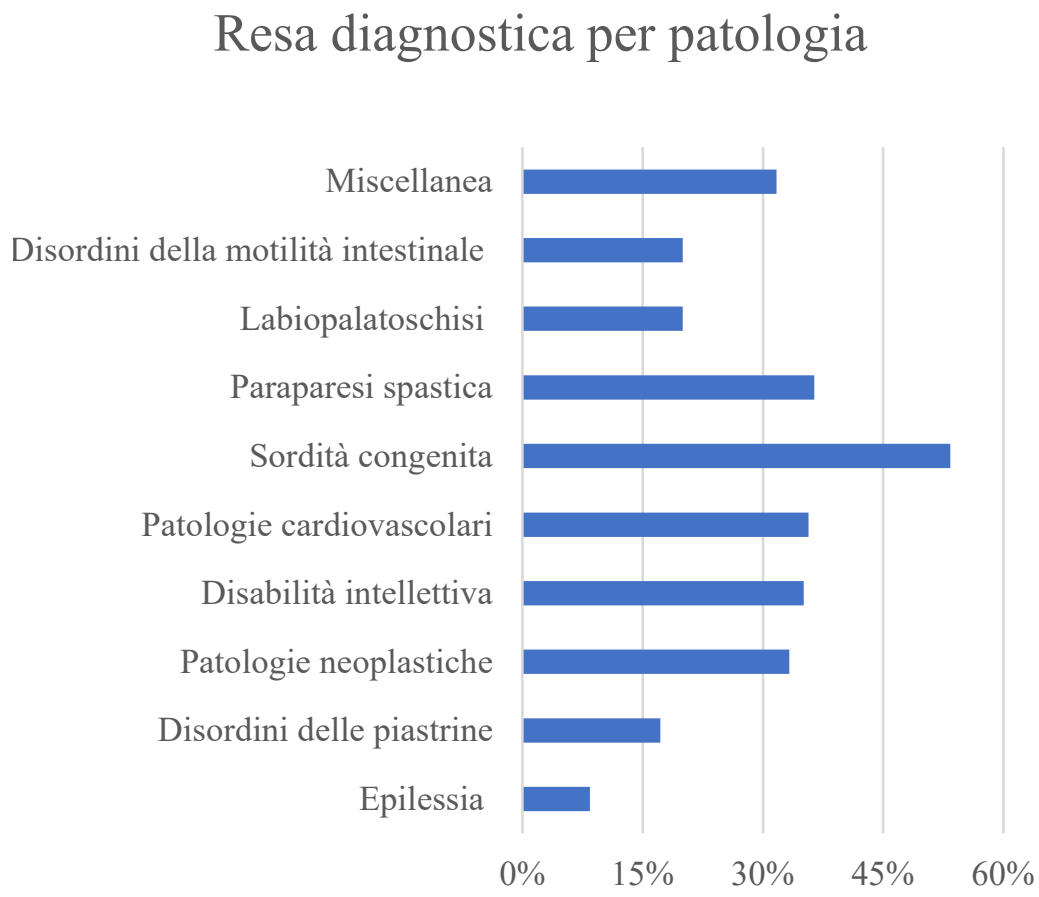
- RS<0
- spurious positions or genes

CANDIDATE VARIANTS

Resa diagnostica

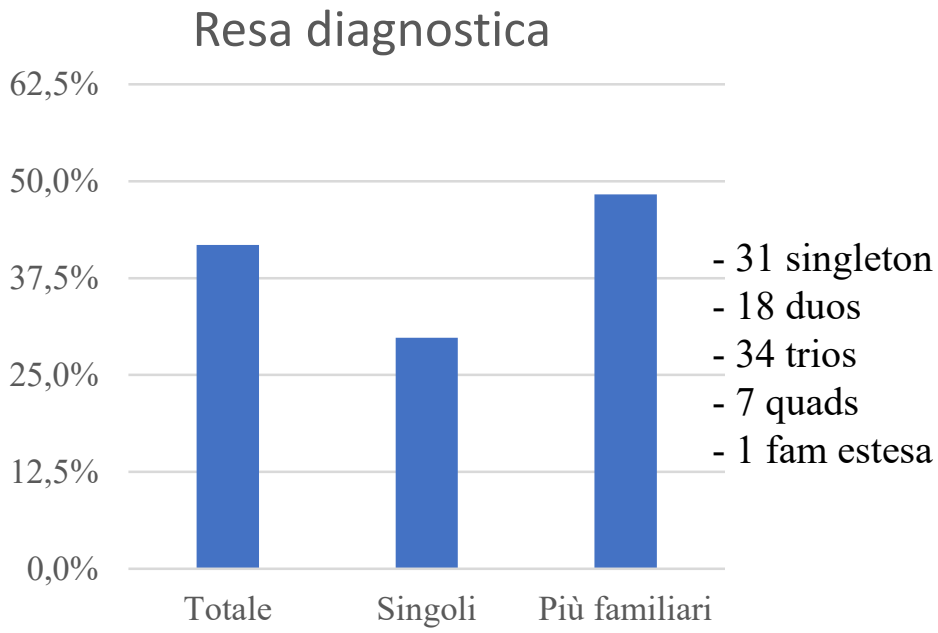
- Resa complessiva del 21,1%
- In caso di analisi di più familiari 29,9%
- Se si escludono le coorti più numerose reclutate per progetti di ricerca 35,4%

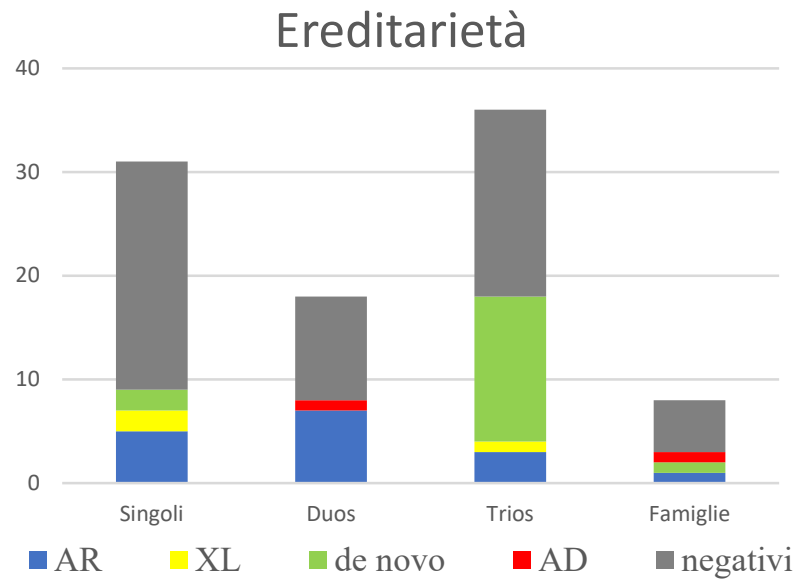




Pazienti seguiti clinicamente presso il nostro centro

- 91 probandi per un totale di 206 individui
- Tutti avevano eseguito almeno un'altra indagine molecolare 'di primo livello' (prevalentemente CGH-array, sequenziamento Sanger, pannelli NGS…)
- Resa diagnostica 41,8% (38/91 casi), in caso di analisi di più familiari la resa raggiunge il 48,3% (29/60 famiglie)

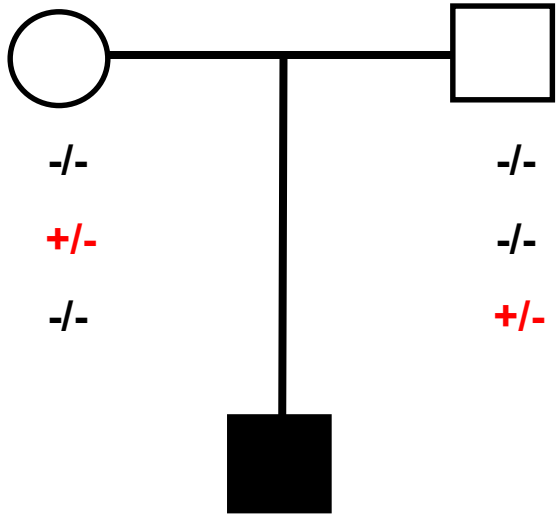




Genotype-Phenotype Complexity

In 3/38 pazienti con diagnosi (7,9%) sono state identificate due patologie genetiche

Nocturnal Frontal Lobe Epilepsy and Congenital Deafness



KCNT1 Ala934Thr	-/-	-/-
TMPRSS3 Pro277Leu	+/-	-/-
TMPRSS3 H70fs	-/-	+/-

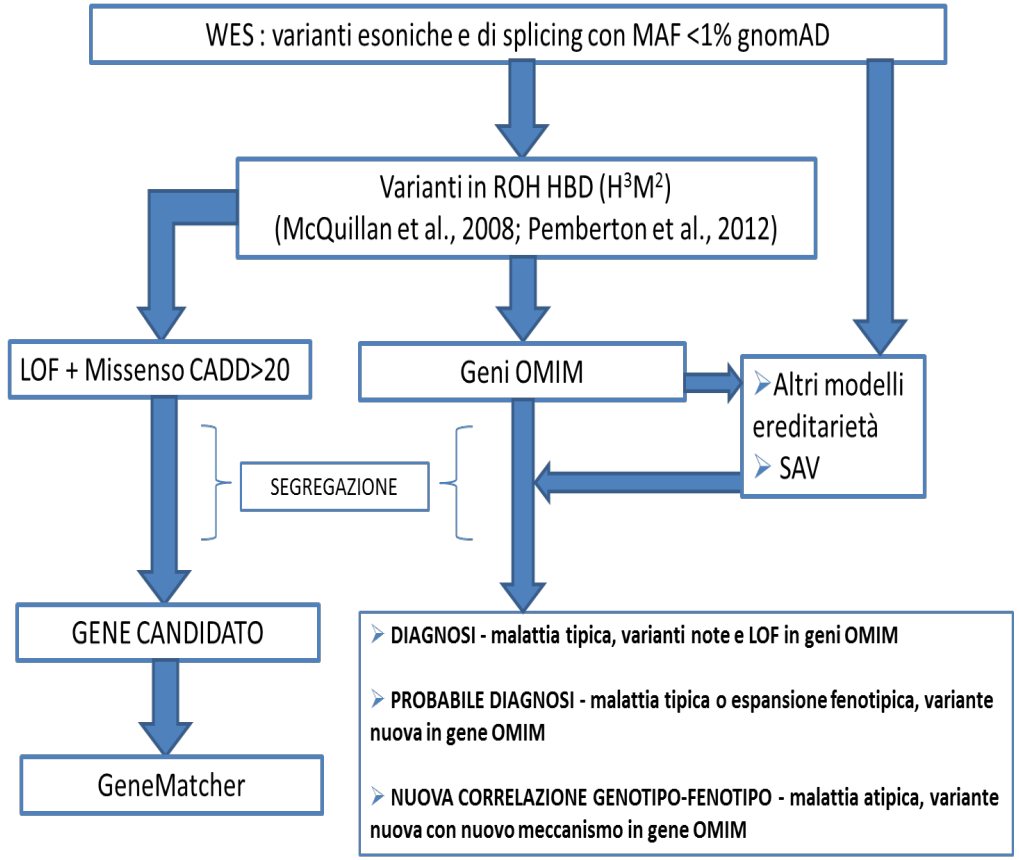
KCNT1 Ala934Thr	+/-
TMPRSS3 Pro277Leu	+/-
TMPRSS3 H70fs	+/-

KCNT1 (MIM 608167) *TMPRSS3 (MIM 605511)*

Famiglie consanguinee

48 probandi provengono da famiglie consanguinee (40 con genitori cugini di 1° grado)

- Resa diagnostica del 43,6%
- La maggior parte delle famiglie sono state analizzate come singleton

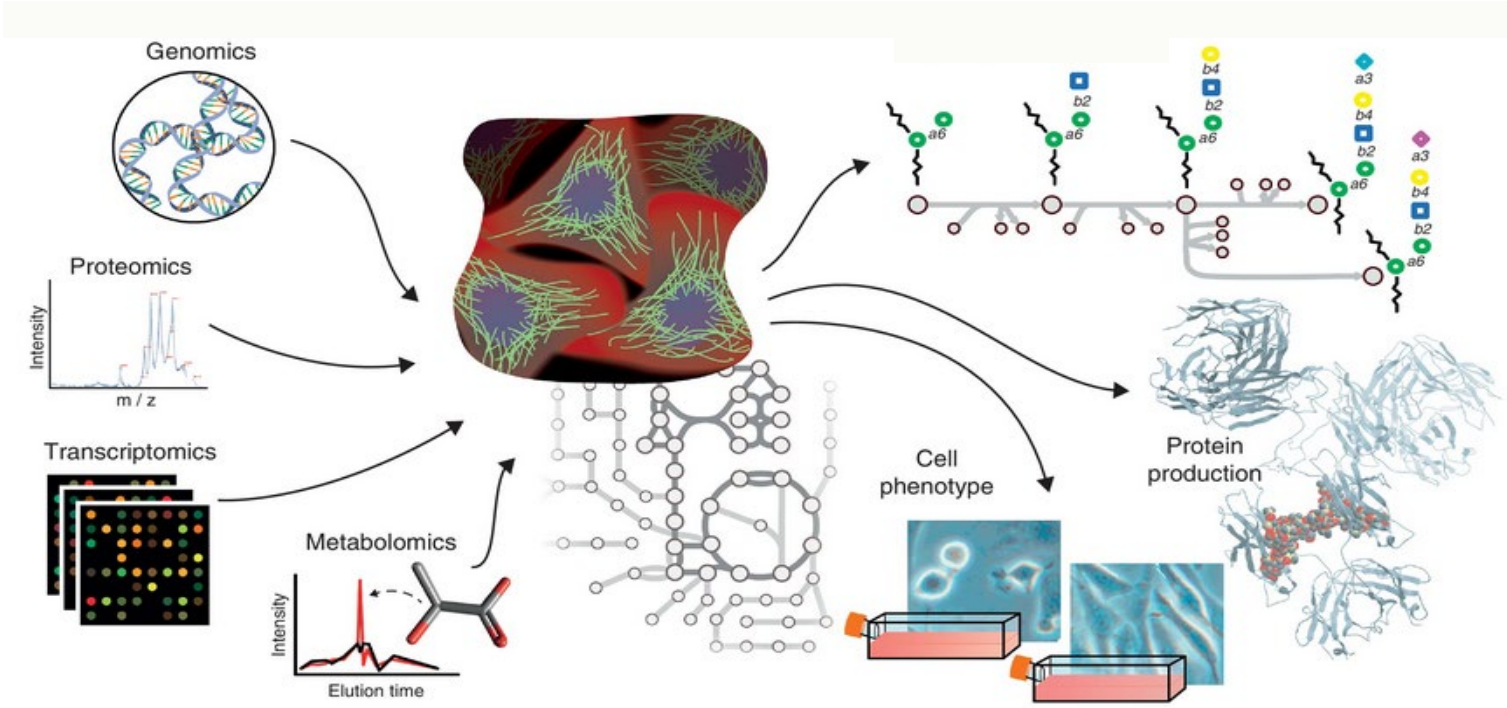


Novel disease-gene correlations

Gene	Disease	Journal	Year
ANKRD26	Thrombocytopenia 2	American Journal of Human Genetics	2011
CACNA2D2	Cerebellar atrophy with seizures and variable developmental delay	PLOS One	2013
PAK3	X-linked syndromic developmental disorder	Human Molecular Genetics	2014
RAD21	chronic intestinal pseudo-obstruction	Gastroenterology	2015
NOTCH3	recessive early-onset arteriopathy and cavitating leukoencephalopathy	EMBO Mol Med	2015
SOS2	Noonan Syndrome	Human Mutation	2015
PRIMA1	nocturnal frontal lobe epilepsy	Annals of Clinical and Translational Neurology	2016
ALDH18A1	Spastic paraplegia 9A, autosomal dominant	Brain	2016
ATAD3A	Harel-Yoon syndrome	American Journal of Human Genetics	2016
ANKRD26	acute myeloid leukemia	Journal of Haematology and Oncology	2017
KIAA1109	Alkuraya-Kucinskas syndrome	American Journal of Human Genetics	2018
TRAPPC2L	Encephalopathy, progressive, early-onset, with episodic rhabdomyolysis	Journal of Medical Genetics	2018
CDC42	Takenouchi-Kosaki syndrome	American Journal of Human Genetics	2018
MYOF	thyroid cancer	Cancer Genetics and Epigenetics	2018
SOX4	Intellectual developmental disorder with speech delay and dysmorphic facies	American Journal of Human Genetics	2019
SMPD4	Neurodevelopmental disorder with microcephaly, arthrogryposis, and structural brain anomalies	American Journal of Human Genetics	2019
NKAP	Intellectual developmental disorder, X-linked syndromic, Hackman-Di Donato type	American Journal of Human Genetics	2019
STARD7	Epilepsy, familial adult myoclonic, 2	Nature Communications	2019
MARCH6	Epilepsy, familial adult myoclonic, 3	Nature Communications	2019
SSBP1	Optic atrophy 13 with retinal and foveal abnormalities	Journal of Clinical Investigations	2020
CCDC32	Cardiofacioneurodevelopmental syndrome	Human Molecular Genetics	2020
SLC12A2	Delpire-McNeill Syndrome	Brain	2020
MAPK1	Noonan syndrome 13	American Journal of Human Genetics	2020
AP1G1	Usmani-Riazuddin syndrome autosomal dominant and recessive	American Journal of Human Genetics	2021
LIG3	Mitochondrial DNA depletion syndrome 20 (MNGIE type)	Brain	2021
SPRED2	Noonan syndrome 14	American Journal of Human Genetics	2021
PIK3C2B	focal epilepsy	Brain	2022
ZMYND8	autosomal dominant neurodevelopmental disorder with cardiac malformations	Genetics in Medicine	2022
NOTCH1	CNS Immune Activation and Microangiopathy	Annals of Neurology	2022
SRSF1	Neurodevelopmental disorder with dysmorphic facies and behavioral abnormalities	American Journal of Human Genetics	2023
CELSR3			

Studi funzionali sono necessari per interpretare le mutazioni identificate

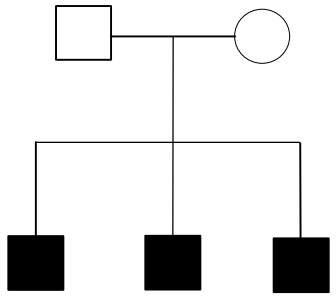
Mutazione identificata >>> come provoca la malattia (il fenotipo)?



ABBIAMO BISOGNO DI STUDI FUNZIONALI IN MODELLI CELLULARI O IN MODELLI ANIMALI

Nuovi geni malattia

LIG3



CIPO, problemi autonomici, leucoencefalopatia.

TYMP, ACTG2 neg.

WES: due varianti in eterozigosi composta in tutti i fratelli

PROVE DI PATOGENICITA'

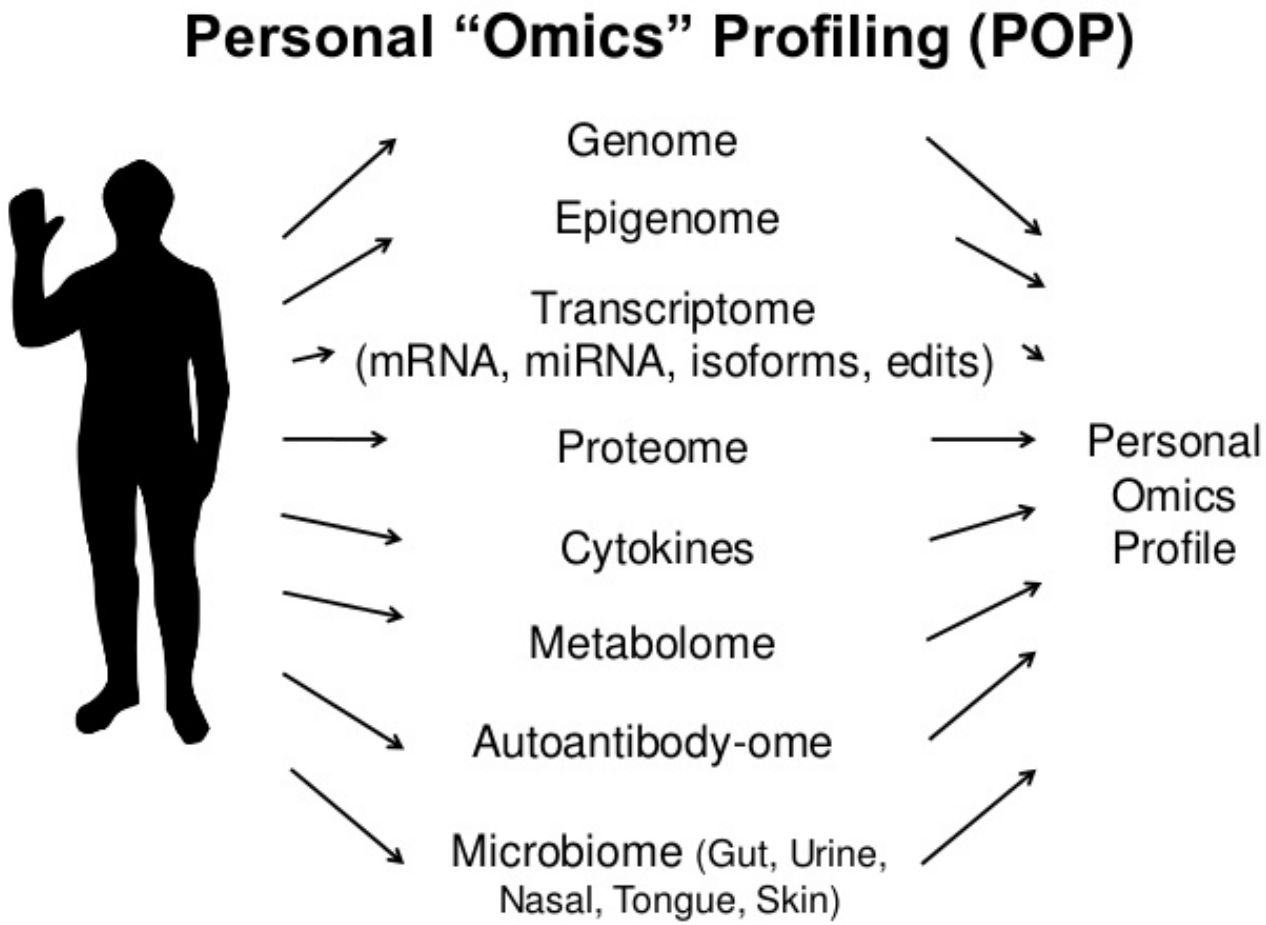
- *LIG3 codifica per la ligasi III, attiva sia nel nucleo sia nel mitocondrio, dove rappresenta l'unica ligasi coinvolta nel riparo del mtDNA, in associazione con la polimerasi γ.*
- Zebrafish *knockout* mostrano alterazione dell'encefalo e della peristalsi intestinale.
- Il fenotipo viene corretto dall'inserimento di *LIG3* wt umano.
- La transfezione con *LIG3* contenente le nostre mutazioni non era in grado di ricostituire il fenotipo normale: mutazioni LOF.
- Studi *in vitro* hanno dimostrato anomalie della catena respiratoria e un aumento delle specie reattive dell'ossigeno.

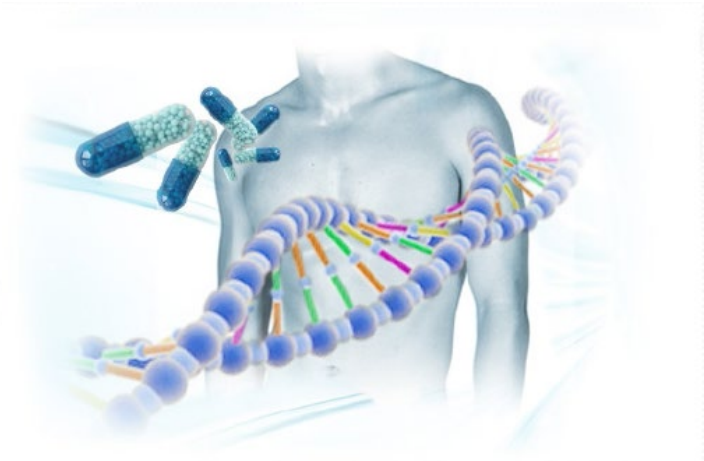
→ Utilizzo di Glutammina per la sua azione antiossidante

- La rivoluzione dell’era Genomica
- L’effetto del progetto genoma umano a livello clinico: l’approccio allo studio delle malattie genetiche rare tramite esoma
- L’utilizzo del genoma nei casi ancora non diagnosticati dopo esoma

E dall'esoma in poi…

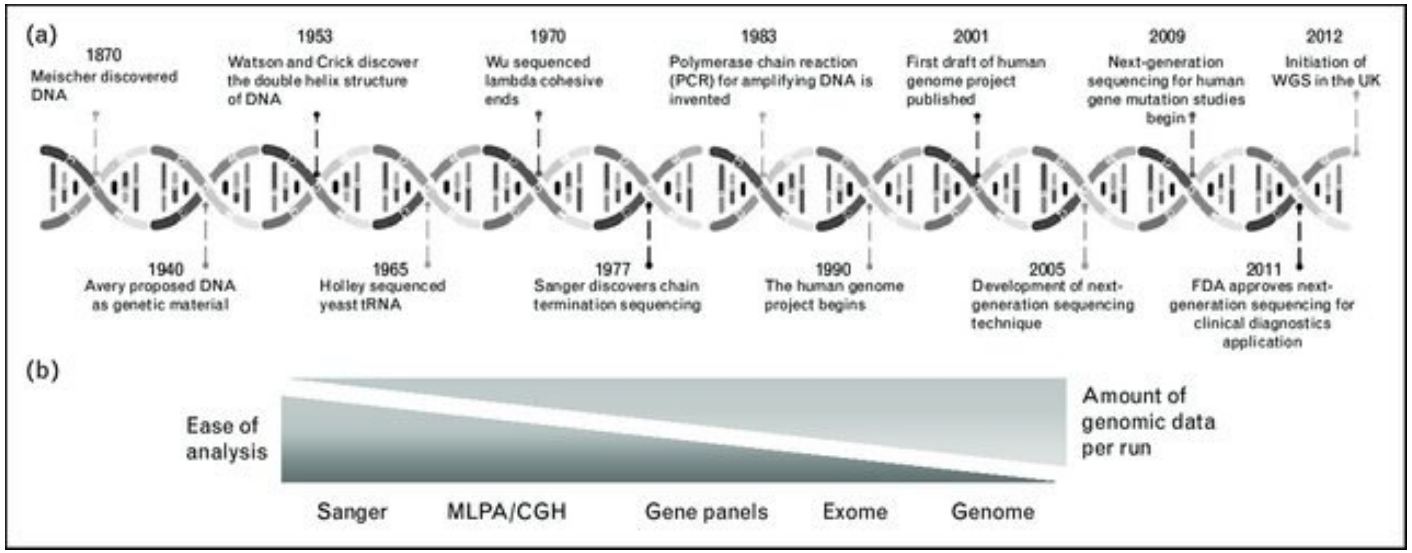
- Pazienti negativi all'analisi dell'esoma>>>> **analisi del genoma** (costi/analisi dei dati ancora complessa)

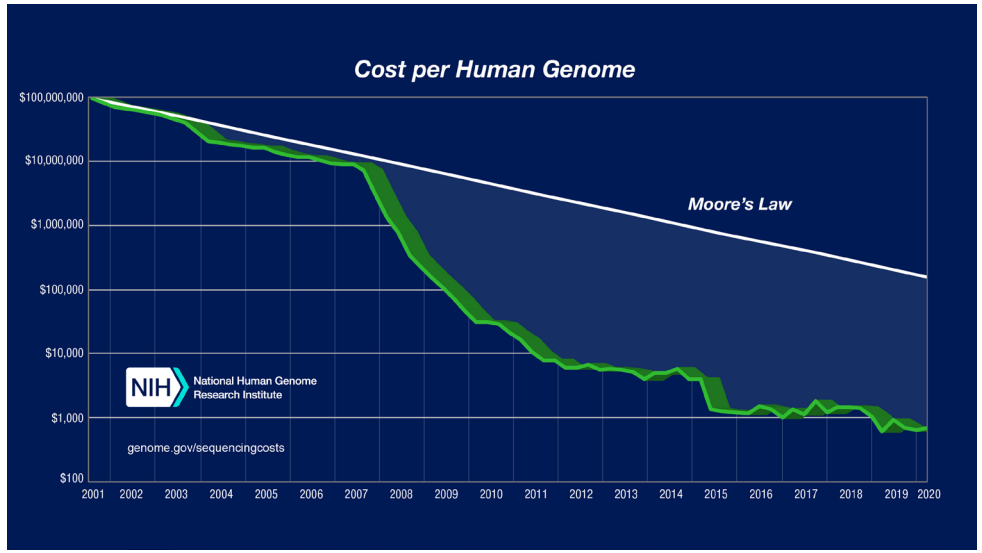




WHOLE GENOME SEQUENCING

- Metodica di sequenziamento in grado di determinare il 98% del genoma umano
  → regioni codificanti e non codificanti (introni e regioni regolatorie intergeniche).

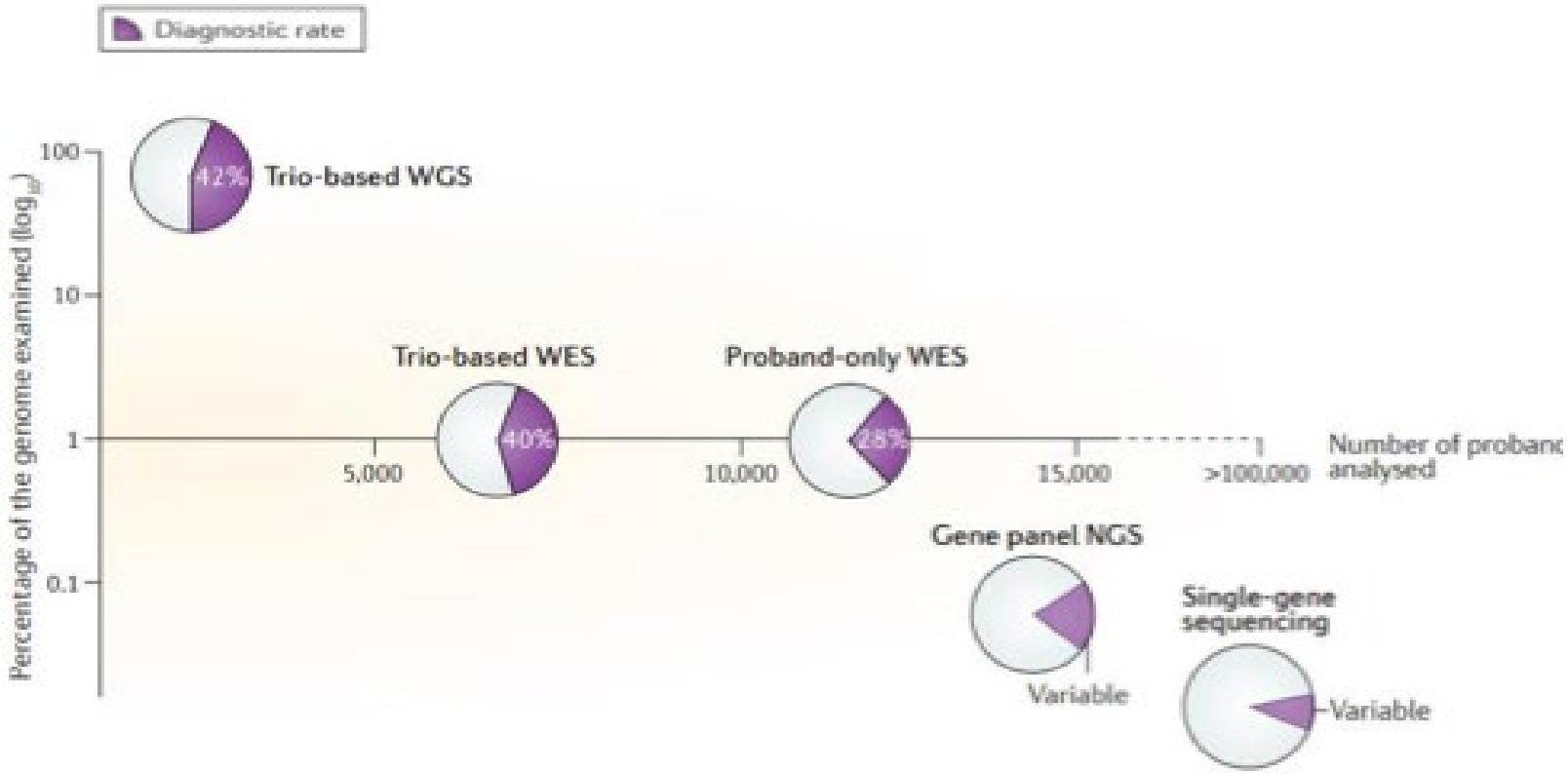




L'avvento e la diffusione delle piattaforme commerciali di NGS ne hanno permesso una resa sempre più efficiente: in parallelo, si è assistito ad un progressivo abbattimento dei costi.

WHOLE GENOME SEQUENCING

Diagnostic rate reported in literature for different NGS strategies

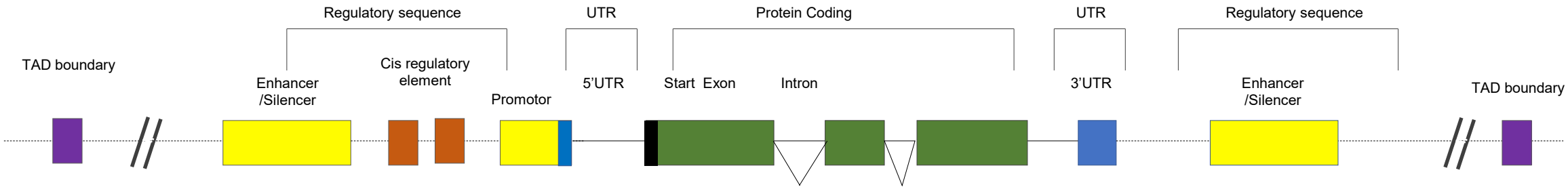


Wright et al., *Nat Rev Genet* , 2018

Move toward non-coding DNA

A large portion of non-coding DNA is functional and regulatory elements tightly control gene and protein expression.

- Key elements in the non-coding genome such as promoters, silencer and enhancers ensure that genes are turned on or off at the right moment
- Cis-regulatory elements regulate gene transcription through the binding of transcription factors (TFs)



Mechanisms of which disrupt non-coding elements cause severe disease:

- Splicing, transcription and translation alteration
- RNA processing and stability
- chromatin interactions

However, interpretation of non-coding regions remains of uncertain significance and more efforts are required to enable a consistent and precise interpretation.

Main project: Genomic feature integration for improved variant interpretation in Whole Genome Sequencing (WGS)

Background:

WGS can uncover all type of genetic variation in coding and non-coding DNA in unbiased way.

Well-established

Single Nucleotide Variants

Small Insertion or deletion

Partially studied

Short tandem repeats (STRs)/Repeat expansions (RE)

Copy number Variants (Structural Variant unbalanced)

Balanced structural variants (SVs) – e.g. inversions

Challenges with WGS data

Millions of variants are identified in a typical genome :

- ✓ ~ 4,000,000 of small variants per samples
- ✓ ~ 20,000 Structural Variants for sample
- ✓ ~ 300,000 Short tandem repeats for sample

Challenges with WGS data

Millions of variants are identified in a typical genome :

Questions

1. How do we handle the huge amount of data?

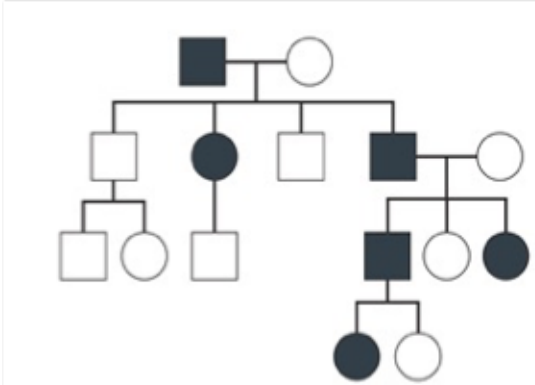
2. How do we facilitate data prioritization for clinical evaluation?

Prioritization: useful steps

Population Frequency

Variant Impact

Inheritance



Gene-Disease association

ClinVar

OMIM
Human Genetics Knowledge for the World

human phenotype ontology

Quality information: coverage

Pathogenicity score

CADD

M-CAP Scores

SIFT score

pLI and Missense Z-score

Limitations of WGS

WGS potentially can identify each type of variation in genomes however some questions remain challenging:

- WGS produces a lot of data and a lot of filtration is required
- Short reads WGS might identify SVs but no single callers reach a high level of precision and recall for all types and all sizes of variants
- Interpretation of non-coding variants remains challenging, but a lot of effort is begin made to increase the understanding

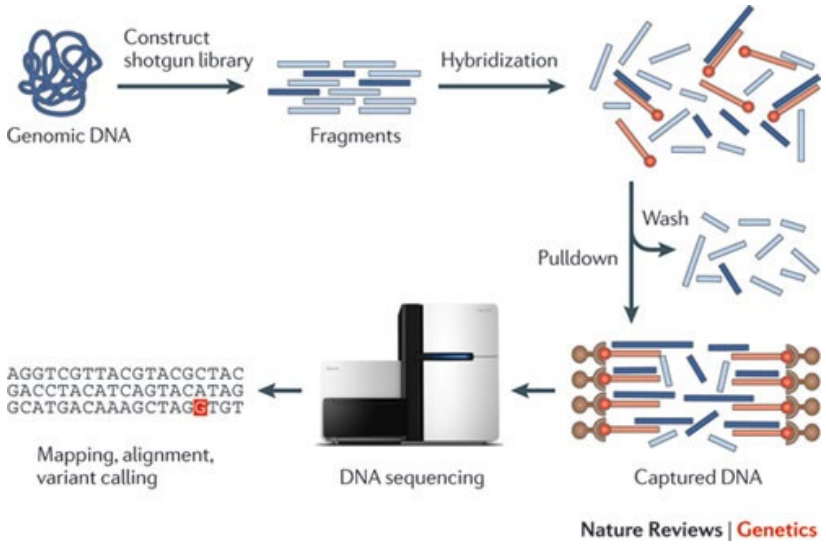
CASI RISOLTI – VARIANTI CANDIDATE - CODING

Family ID	Sample	HGVS	Conseguence	AF GnomAD	CADD	ACMG	OMIM	GEN
1	18963_S11	***CREBBP***(NM_004380.3):c.5602C>T (p.Arg1868Trp)	missense	na	27.7	Pathogenic	Rubinstein-Taybi syndrome 1 AD	denovo 0/1
2	18836_S1	***MT-ATP8*** m.8535A>G (p.Lys57*)	mit				deficit del complesso V mitocondriale (ATP sintasi) tipo 2, MT	96% di eteroplasmia su sangue periferico
7	17303_S11	***HECW2***(NM_001348768.2) :c.3988C>T (p.Arg1330Trp)	missense	na	27.8	Pathogenic	Neurodevelopmental disorder with hypotonia, seizures, and absent language, AD	denovo 0/1
8	22900_S13	***AFF4***(NM_014423.4):c.772C>T	missense	na	26.7	Pathogenic	CHOPS syndrome, AD	denovo 0/1
10	25428_S16	***TRIT1***(ENST00000316891.10):c.448C>T (p.Arg150Ter)	non-sense	na	35	Likely Pathogenic	Combined oxidative phosphorylation deficiency 35, AR	0/1
	25427_S12	TRIT1(ENST00000316891.10):c.682T>C (p.Trp228Arg)	missense	0.00000657	31	Uncertain Significance	Combined oxidative phosphorylation deficiency 35, AR	0/1
11	ISNB_10448_S17	***CYFIP2***(NM_001037333.3) :c.2631del (p.Asp877GlufsTer57)	frameshift	na		Likely Pathogenic	Developmental and epileptic encephalopathy, AD	denovo 0/1

CASI RISOLTI – VARIANTI CANDIDATE – NON CODING

Family ID	Sample	HGVS	Conseguence	AF GnomAD	CADD	ACMG	OMIM	GEN	ncER	ReMM	FATHMM
5	18319_S5	CHD7(NM_017780.4):c.5210+1235A>G	non-coding	na	14.75	Likely Benign	CHARGE syndrome, AD	denovo 0/1	92.99	0.042	0.1424
4	17100_S14	SLC25A12(ENST00000422440.7):2:171808202_171808320	Dels	na in gnomadSV		Uncertain_significance	Developmental and epileptic encephalopathy,AD	inherited 0/1			
		SLC25A12(ENST00000422440.7):c.1446+27T>A	non-coding	na	6	Likely Benign		denovo 0/1	98.51	0.867	0.2951
9	30469A_S14	LMBR1(ENST00000353442.10):c.424-5999T>G	non-coding	na	5	Likely Benign	TIBIA, HYPOPLASIA OR APLASIA OF, WITH POLYDACTYLY AD	inherited 0/1	88.52	0.708	0.1118

Nuove prospettive diagnostiche e terapeutiche



GENOMES ON PRESCRIPTION

The first clinical uses of whole-genome sequencing show just how challenging it can be.

BY BRENDAN MAHER

The first thing Debbie Jorde noticed about her newborn daughter was that her arms were bent at unnatural angles. She had other problems, too: a cleft palate, eight fingers, eight toes and no lower eyelids. She would eventually be diagnosed with Miller syndrome, a disease so rare that doctors have long assumed that each case arises through spontaneous mutation, rather than being passed down through families. Doctors told Jorde that her chances of having a second child with the syndrome were less than one in a million.

They were wrong. Jorde's son, born three years after his sister, had the same features. Lynn Jorde, Debbie's current husband and a geneticist at the University of Utah in Salt Lake City, still cringes when Debbie recounts what the doctors had told her. "The right answer for that situation is that there have been so few cases that we really can't predict the risk," he says.

Thanks to next-generation genome sequencing, Debbie and her children now know the family's genetic risks. Lynn and his collaborators had been talking about sequencing the genomes of an entire 'nuclear' family affected by a genetic disease, both to identify the mutation responsible and to investigate how genes are inherited in unprecedented detail. Debbie, her former husband and her now-adult children, Heather and Logan Madsen, were happy to be take part, and in 2009 became the first family in the world to have their genomes fully sequenced[1].

Over the course of six months, the research team cross-compared the whopping amount of DNA data from the four genomes. With the help of a parallel sequencing effort that included others

Grazie per l'attenzione!