Lezione 3: Genoma umano come esempio di genoma eucariote

Schema della lezione

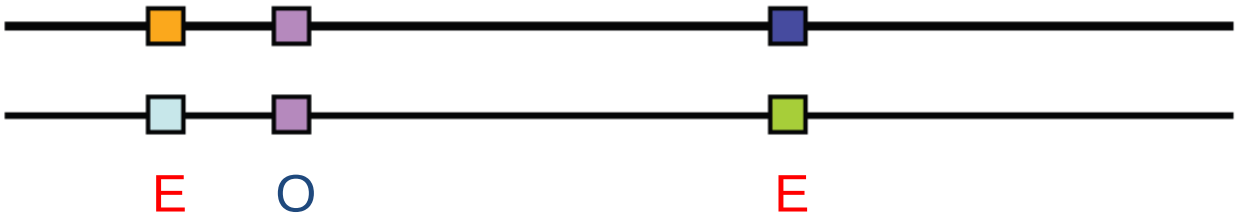
- Sommario degli elementi contenuti in un genoma eucariote
- Variabilità: dove si trova e come si definisce
- I grandi progetti internazionali di caratterizzazione della variabilità dei genomi umani
- Alcuni esempi di regioni codificanti il proteoma: geni di lunghezze molto diverse e famiglie geniche (DNA moderatamente ripetitivo)

Genomi eucarioti

- Regioni codificanti proteine
- Regioni codificanti RNA strutturali (es. rRNA)
- DNA moderatamente ripetitivo
    - Famiglie geniche come actina, globine
    - Geni per RNA ribosomali (ripetuti in centinaia di copie)
    - tRNA (50 siti ognuno con 10/100 copie nell'uomo)
    - Geni per istoni in molte specie
- Elementi **ripetuti** a funzione non nota (es. SINE (tra cui le Alu nei primati) e LINE) e **altamente ripetuti** (es. mini e microsatelliti o STR), Telomeri (corte unità ripetitive TTAGGG nell'uomo)
- Regioni intergeniche non ripetute
- Regioni funzionali ancora da definire con esattezza (ENCODE http://www.nature.com/encode/#/threads)

I geni negli individui

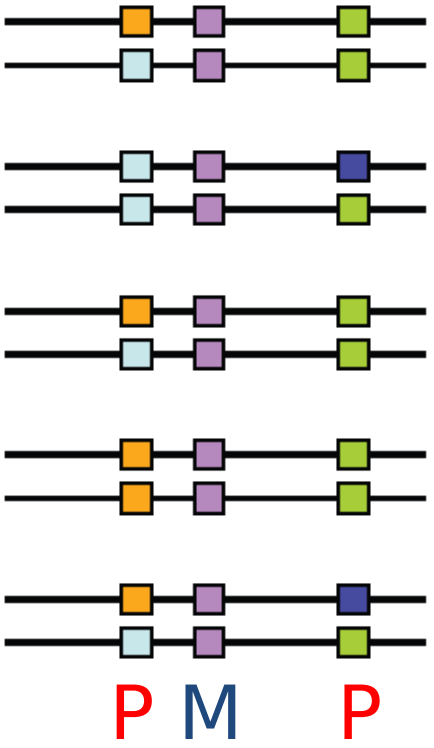
- Omozigote: lo stesso allele ad un locus in organismi diploidi (o poliplodi)



- Eterozigote: Alleli diversi ad un locus in organismi diploidi (o poliplodi)

I geni nelle popolazioni

- Polimorfico: presenza di >1 tipo genico distinto nella popolazione (variante, forma alternativa)

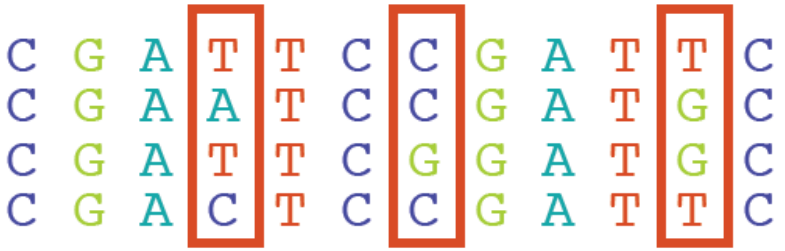


- Monomorfico: nella popolazione esiste un solo tipo genico

I geni nelle popolazioni

- Polimorfico: presenza di >1 tipo genico distinto nella popolazione (variante, forma alternativa)
- Monomorfico: nella popolazione esiste un solo tipo genico

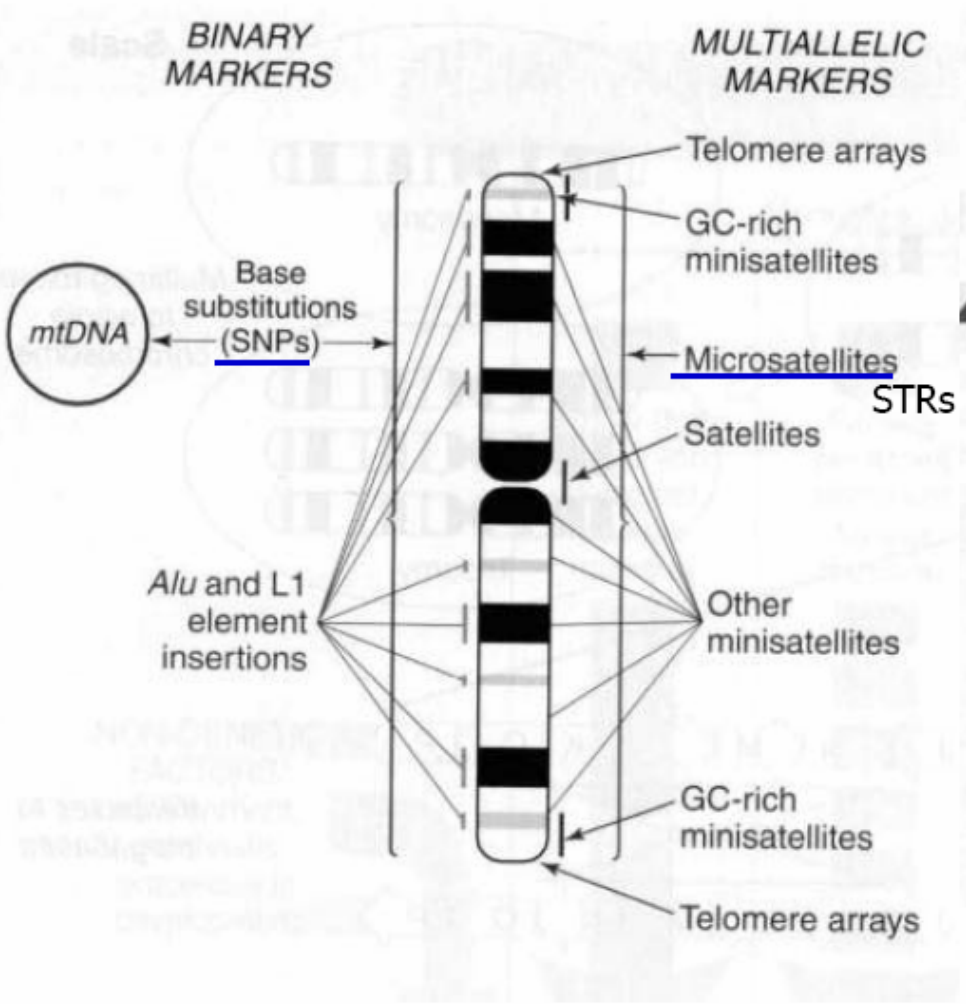
Il polimorfismo può essere inteso su scala genetica (esempio: single nucleotide polymorphism, SNP)

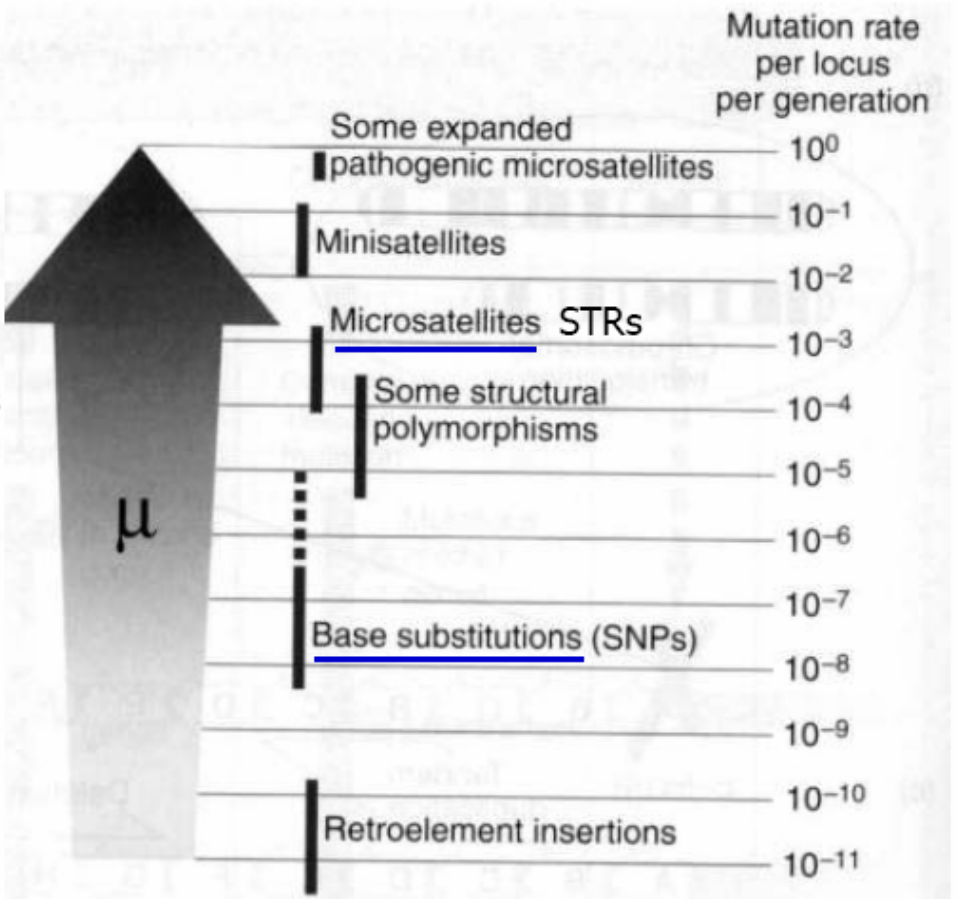


O su scala popolazionistica (colori, tipi di giaguari in Sud America)

Distribuzione nel genoma dei diversi tipi di marcatori molecolari polimorfici

Tassi di mutazione (frequenza)





Jobling et al. 2003

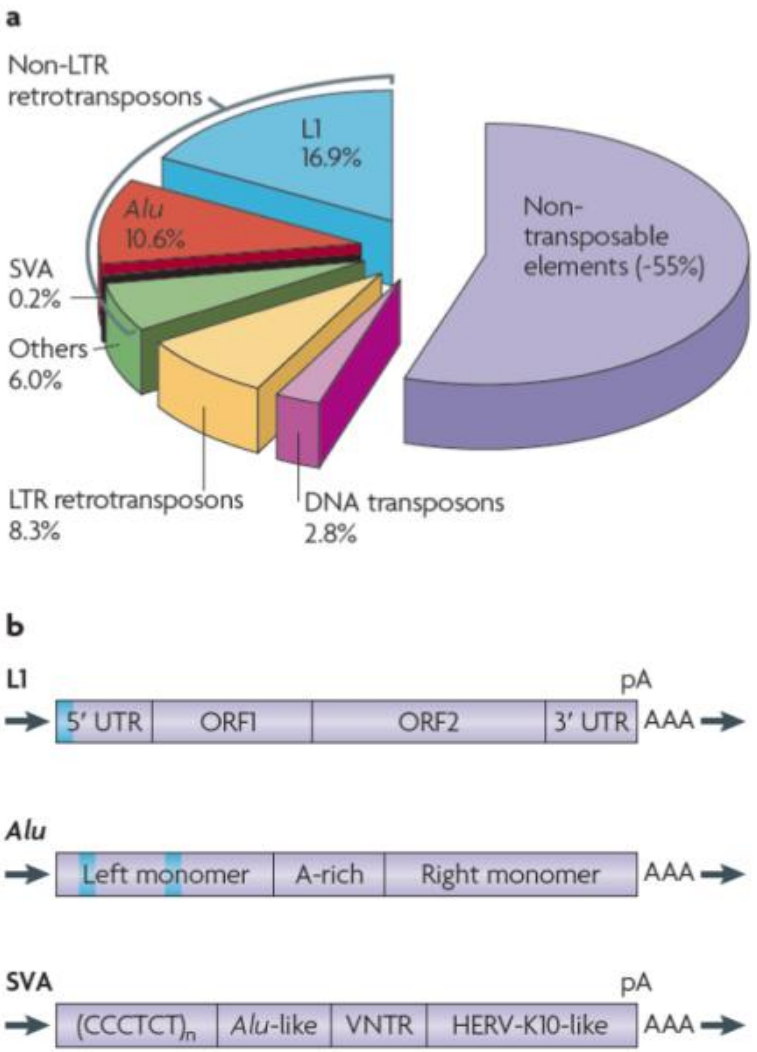
Review

Nature Reviews Genetics **10**, 691-703 (October 2009) | doi:10.1038/nrg26

The impact of retrotransposons on human genome evolution

Richard Cordaux & Mark A. Batzer

Their ability to move within genomes gives transposable elements an intrinsic propensity to affect genome evolution. Non-long terminal repeat (LTR) retrotransposons — including LINE-1, *Alu* and SVA elements — have proliferated over the past 80 million years of primate evolution and now account for approximately one-third of the human genome. In this Review, we focus on this major class of elements and discuss the many ways that they affect the human genome: from generating insertion mutations and genomic instability to altering gene expression and contributing to genetic innovation. Increasingly detailed analyses of human and other primate genomes are revealing the scale and complexity of the past and current contributions of non-LTR retrotransposons to genomic change in the human lineage.

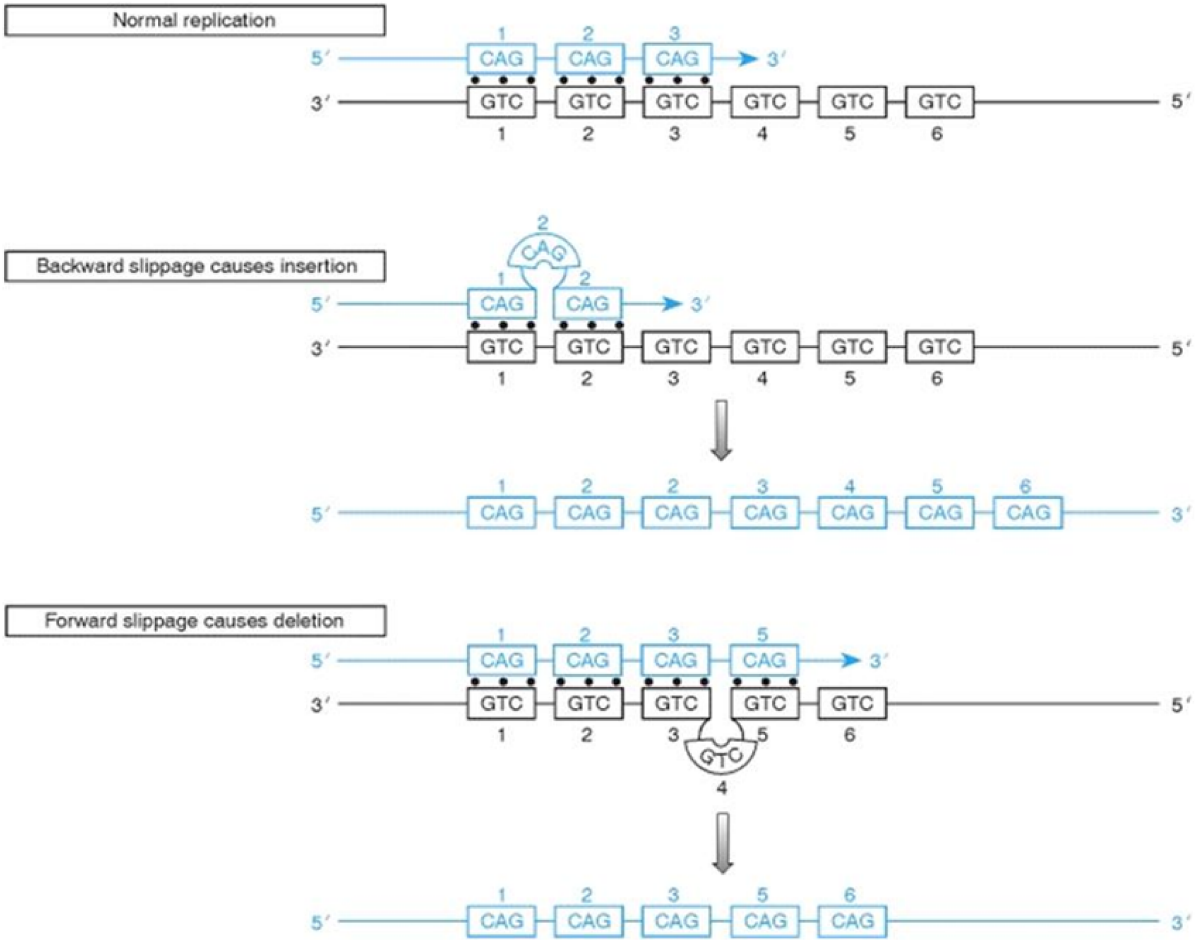


The transposable element content of the human genome

About 45% of the human genome can currently be recognized as being derived from transposable elements, the vast majority of which are non-LTR retrotransposons such as L1, *Alu* and SVA elements. L1, LINE-1; LTR, long terminal repeat.

I microsatelliti (short tandem repeats, STRs)

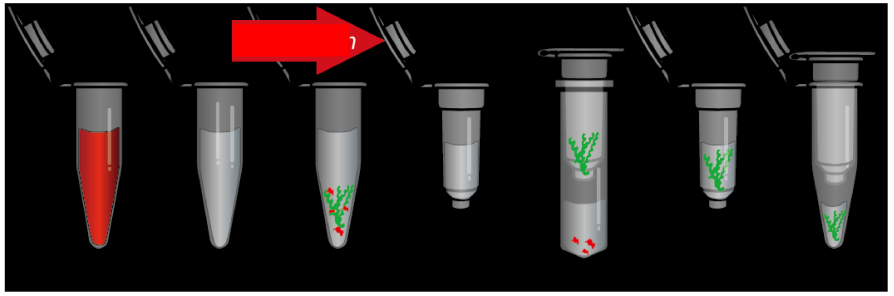
Sequenze ripetitive di 2-5 paia di basi in tandem



From Strachan and Read, Human Molecular Genetics 2

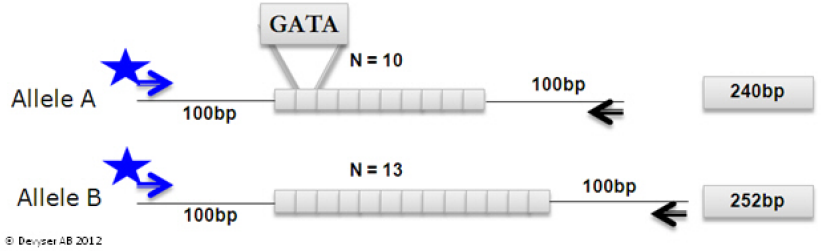
Come ottenere un profilo individuale STR (DNA fingerprinting)

Raccolta del campione biologico



Estrazione del DNA

Amplificazione delle regioni di interesse (primer marcati fluorescenti)



Lettura dei frammenti con un sequenziatore automatico a capillare

THE FBI FEDERAL BUREAU OF INVESTIGATION

CONTACT US | ABOUT US | MOST WANTED | NEWS | STATS

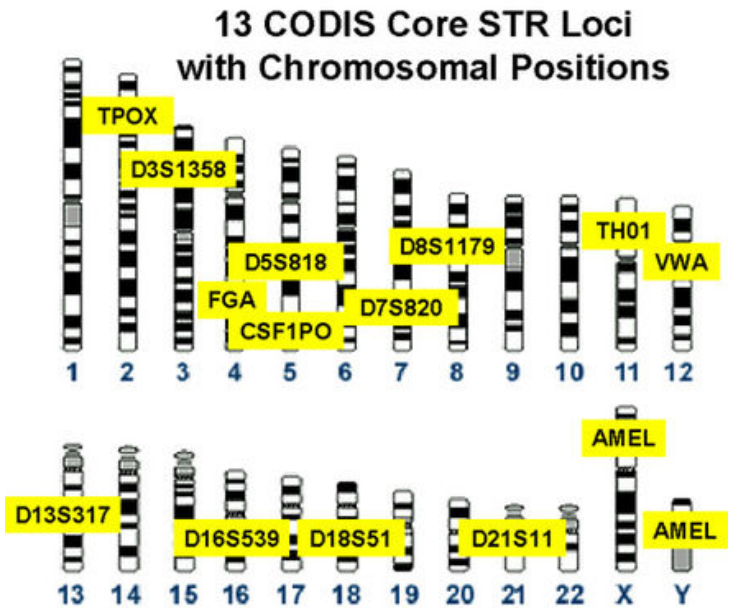
Laboratory Services

Home • About Us • Laboratory Services • Biometric Analysis

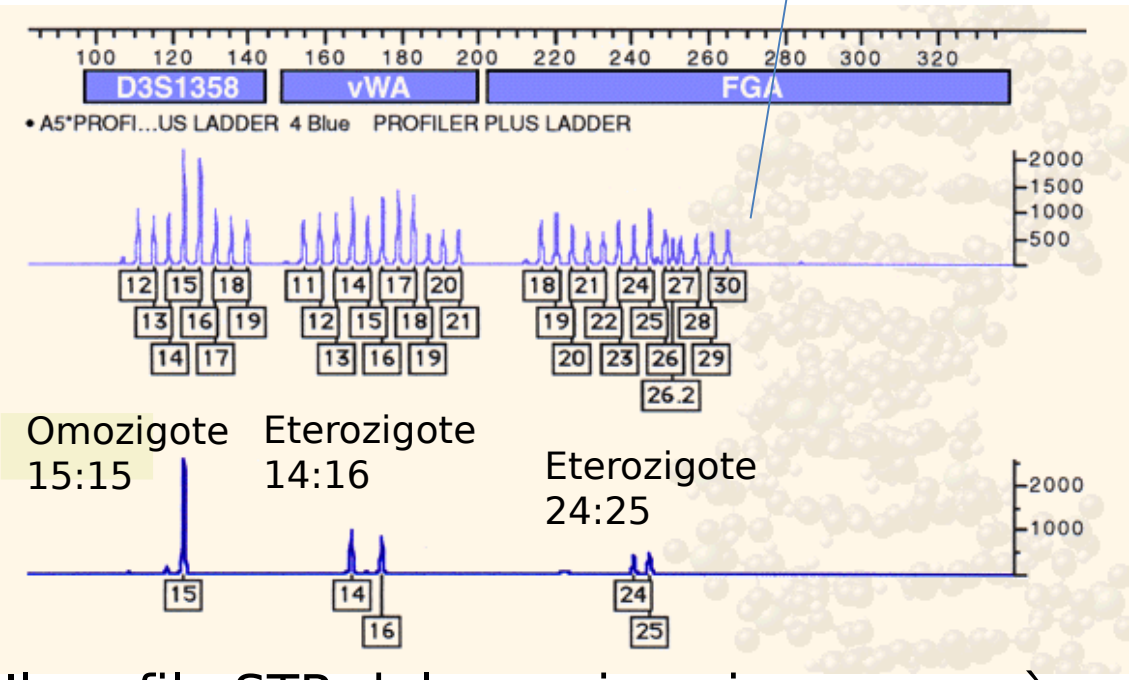
Combined DNA Index System (CODIS)

Mission

The CODIS Unit manages the Combined DNA Index System (CODIS) and the National DNA Index System (NDIS) and is responsible for developing, providing, and supporting the CODIS Program to federal, state, and local crime laboratories in the United States and selected international law enforcement crime laboratories to foster the exchange and comparison of forensic DNA evidence from violent crime investigations. The CODIS Unit also provides administrative management and support to the FBI for various advisory boards, Department of Justice (DOJ) grant programs, and legislation regarding DNA.

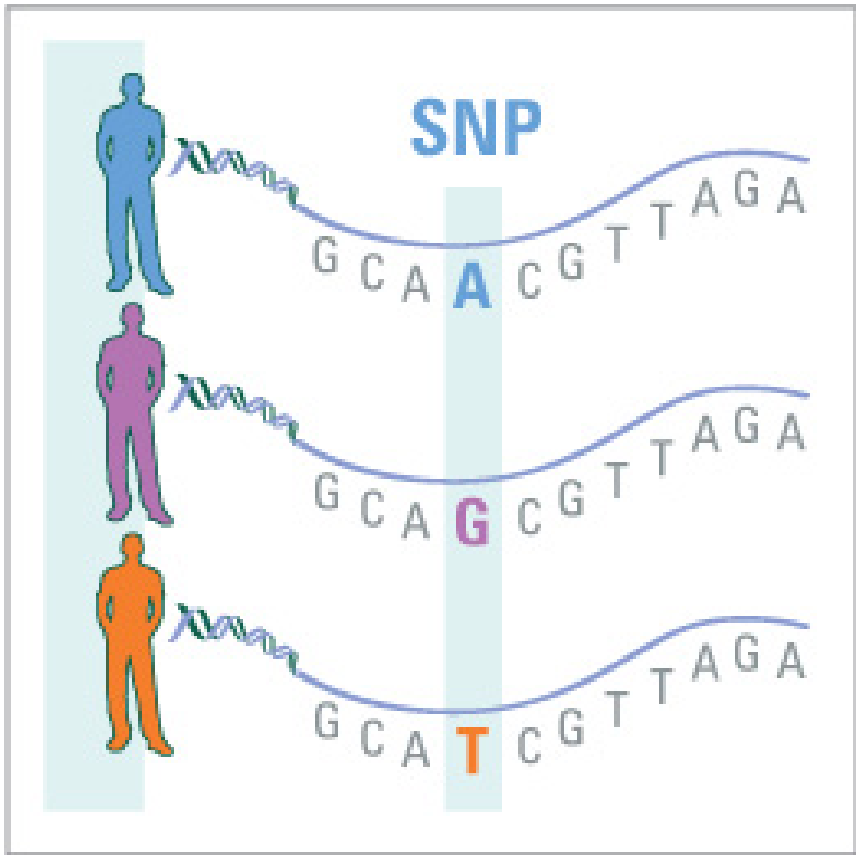


Ladder:
una specie di "legenda"
Include tutti gli alleli conosciuti (ripetizioni) ad ogni locus in analisi



Il profilo STR del campione in esme può essere definito confrontando gli alleli con

Single nucleotide polymorphism (SNP)

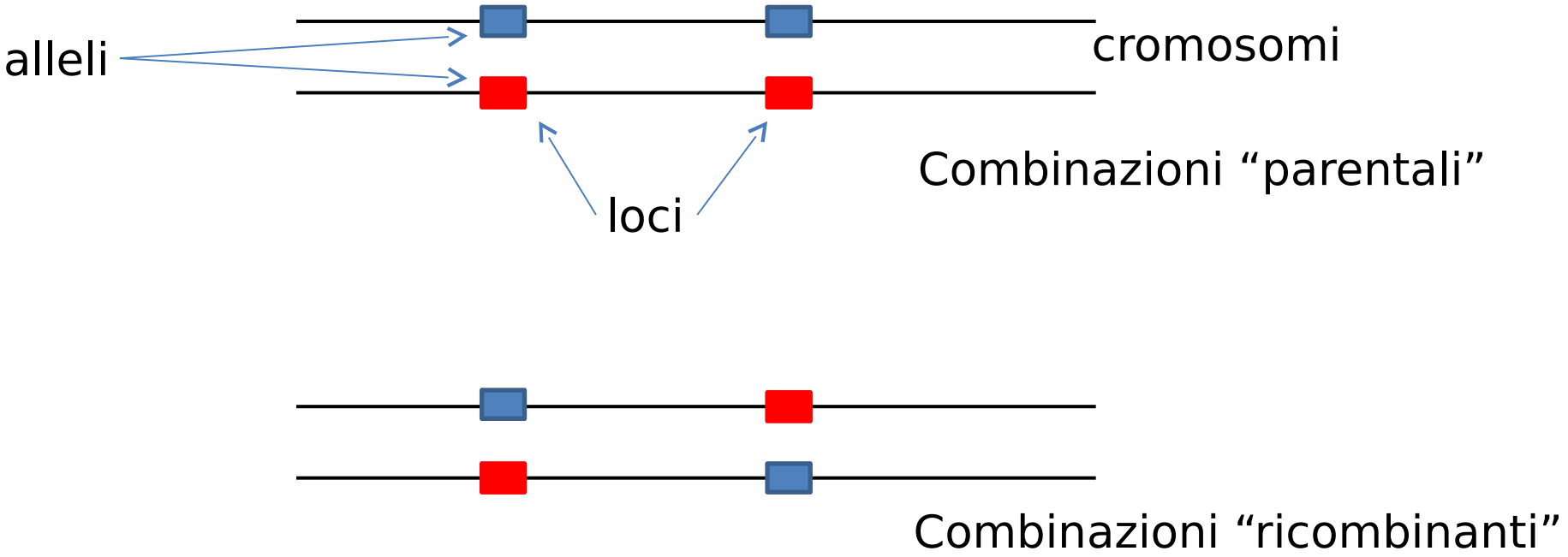


La sequenza genomica di ogni coppia di individui della nostra specie è uguale (in media) per il 99.9%.

Circa 1 in 1000 “lettere” del DNA umano può variare in forma di SNP.

Linkage disequilibrium (LD)

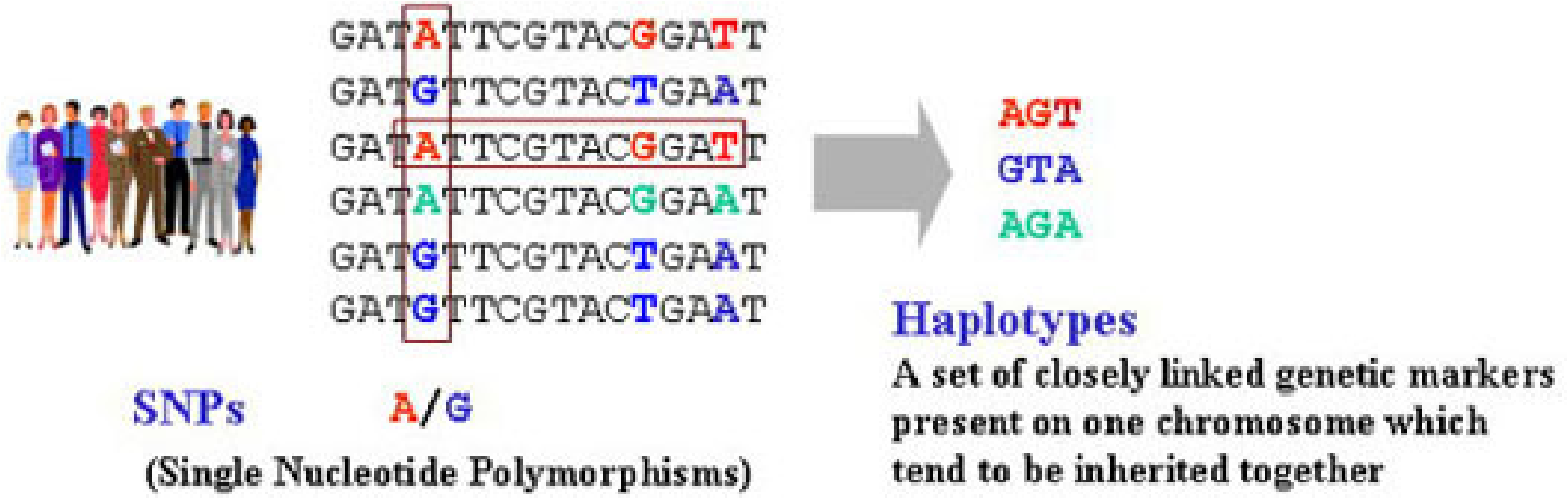
Associazione non casuale tra ALLELI a loci diversi in una popolazione dovuta alla tendenza ad essere ereditati insieme.



Se c’è LD ci sono “troppo poche” copie di ricombinanti (significativamente meno dell’atteso)

Il LD comporta la formazione di aplotipi

Aplotipo: set di alleli (o marcatori genetici) strettamente legati presenti su un cromosoma e che tendono ad essere ereditati insieme



*Siccome la **A** in prima base è sempre insieme alla **G** in seconda base, posso tipizzare una sola delle due e saprò cosa c'è nell'altra posizione*

TABLE 10.1:
LARGE-SCALE PROJECTS PROVIDING INFORMATION ON HUMAN GENETIC VARIATION

Project	Launch date	Primary aims	Sample size	Populations	Genetic analyses	Cell lines	Data release	Website (Key reference) Further information
HGDP	1991	collection of isolated population samples	1064	51, worldwide	chosen by investigator	yes	on publication	http://www.cephb.fr/en/hgdp/diversity.php/ [Cann HM et al. (2002) *Science* 296, 261.] Box 10.2
HapMap	2002	haplotype map for medical genetics	1184	11, Africa, Europe, South and East Asia	SNP genotyping	yes	full public release	http://hapmap.ncbi.nlm.nih.gov/index.html.en [The International HapMap Project (2003) *Nature* 426, 789.] Box 3.6
Genographic	2005	elucidate migration history	~500,000	many, worldwide; including public participation	Y-SNP and Y-STR genotyping, mtDNA HVSI sequencing	no	on publication	https://genographic.nationalgeographic.com/genographic/lan/en/index.html [Wells, Deep Ancestry: Inside the Genographic Project (2006) National Geographic Books]
1000 Genomes	2008	discover variants at ≥1% frequency for medical genetics	2500	27, Africa, Europe, South Asia, East Asia, Americas	whole-genome sequencing	yes	full public release	http://www.1000genomes.org/ [The 1000 Genomes Project Consortium (2010) *Nature* 467, 1061.] Box 3.2

Jobling et al. Human Evolutionary genetics 2nd ed. 2014

2002- 2009

http://hapmap.ncbi.nlm.nih.gov/

Grande progetto internazionale con lo scopo di caratterizzare il pattern (modello)di variabilità genetica e di LD in campioni di individui di popolazioni provenienti da diverse aree geografiche

Caratteristiche principali:

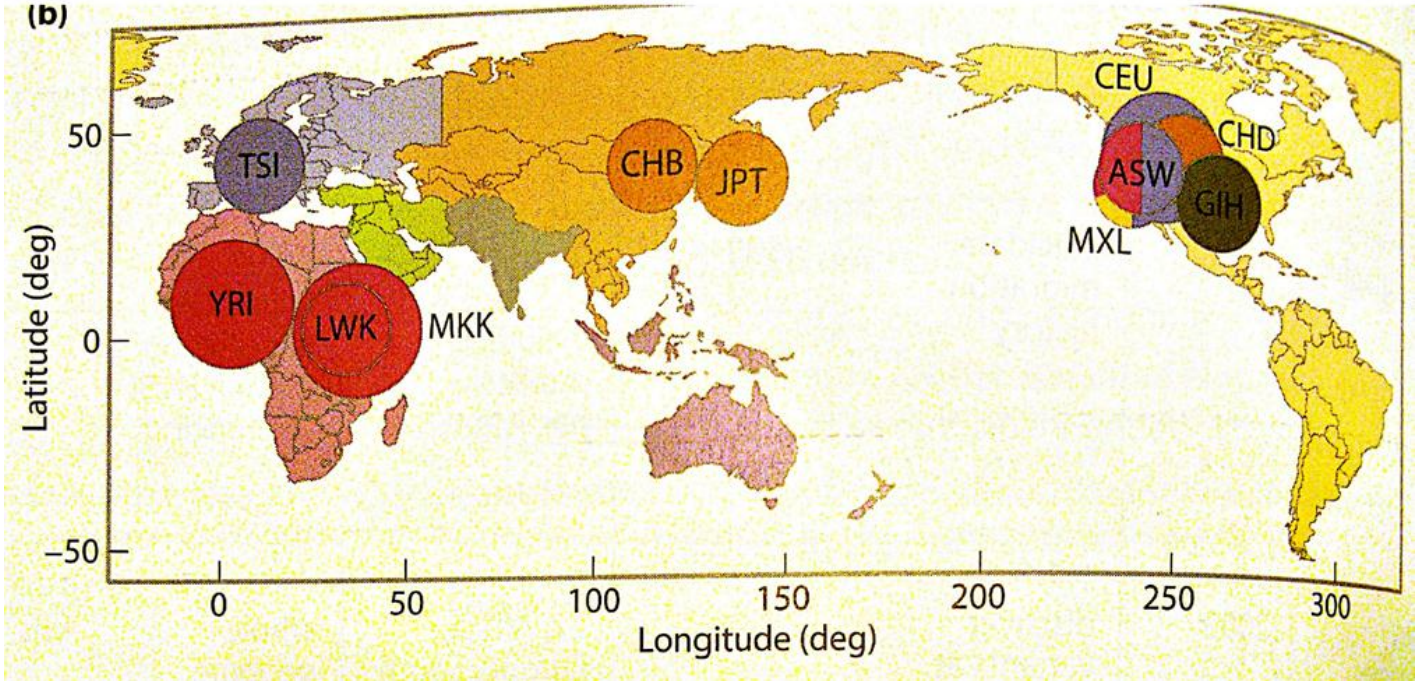
- Tipizzazione di SNP, non sequenziamento di regioni
- Popolazioni di etnia definita, più alcune mescolate

Applicazioni:

- Study design di associazioni geni-malattie

Campioni popolazionistici inclusi nel progetto HapMap.

ASW	African ancestry in Southwest USA
CEU	Utah residents with Northern and Western European ancestry from the CEPH collection
CHB	Han Chinese in Beijing, China
CHD	Chinese in Metropolitan Denver, Colorado
GIH	Gujarati Indians in Houston, Texas
JPT	Japanese in Tokyo, Japan
LWK	Luhya in Webuye, Kenya
MXL	Mexican ancestry in Los Angeles, California
MKK	Maasai in Kinyawa, Kenya
TSI	Toscani in Italia
YRI	Yoruba in Ibadan, Nigeria



L'area dei cerchi è proporzionale alla dimensione del campione.
I colori indicano l'origine etnica dei campioni
Notare la scarsa rappresentazione di popolazioni indigene di alcuni continenti come l'Australia e il nord America

Jobling et al. Human Evolutionary genetics 2nd ed. 2014

TABLE 1:
DETAILS OF POPULATION SAMPLES USED IN HAPMAP

Standard abbreviation	Origin	Sample size and composition
Phase I and II		
YRI	Yorubans from Ibadan, Nigeria	90 (30 parent–child trios)
CEU	Utah (US) residents of N and W European ancestry	90 (30 parent–child trios)
CHB	Han Chinese from Beijing, China	45 unrelated individuals
JPT	Japanese from Tokyo, Japan	44 unrelated individuals
Phase III		
ASW	individuals of African ancestry from Southwest USA	90 (11 parent–child trios, 24 parent–child duos + 9 unrelated individuals)
CHD	Chinese from Metropolitan Denver, Colorado, USA	100 unrelated individuals
GIH	Gujarati Indians from Houston, Texas, USA	100 unrelated individuals
LWK	Luhya from Webuye, Kenya	100 unrelated individuals
MEX	individuals of Mexican ancestry from Los Angeles, California, USA	90 (30 parent–child trios)
MKK	Maasai from Kinyawa, Kenya	180 (30 parent–child trios + 90 unrelated individuals
TSI	Tuscans from Italy	100 unrelated individuals

Jobling et al. Human Evolutionary genetics 2nd ed. 2014

Box 3.6: The nuts and bolts of HapMap

HapMap proceeded in three phases:

- *Phase I* aimed to genotype at least one common (MAF ≥0.05) SNP per 5 kb in each of 269 samples (see **Table 1** and **Figure 1**), and published data on 1 million SNPs;[22] genotyping was carried out by nine genome centers using six different technologies.
- *Phase II* increased the typed SNPs on the same samples to 3.1 million,[23] including SNPs with MAF <0.05. It largely employed one genotyping technology, and estimated per-genotype accuracy at ≥99.5%.
- *Phase III* included additional samples from a more diverse set of populations (see Table 1 and Figure 1), and has typed 1.3–1.5 million SNPs using two standardized methods.

All the HapMap samples are available for purchase as DNA aliquots and as immortalized lymphoblastoid cell lines (**Section 4.1**). This has made them an invaluable standardized set of samples: the Phase I/II samples have been studied not only for SNP variation, but also for CNV and gene expression, and have undergone genome sequencing using next-generation methods as part of the 1000 Genomes Project (see Box 3.2). Note that because none of the samples was collected as a representative of a particular larger population (for example YRI to represent Yorubans, or Nigerians, or Africans), HapMap recommends that specific local identifiers be used to describe them.

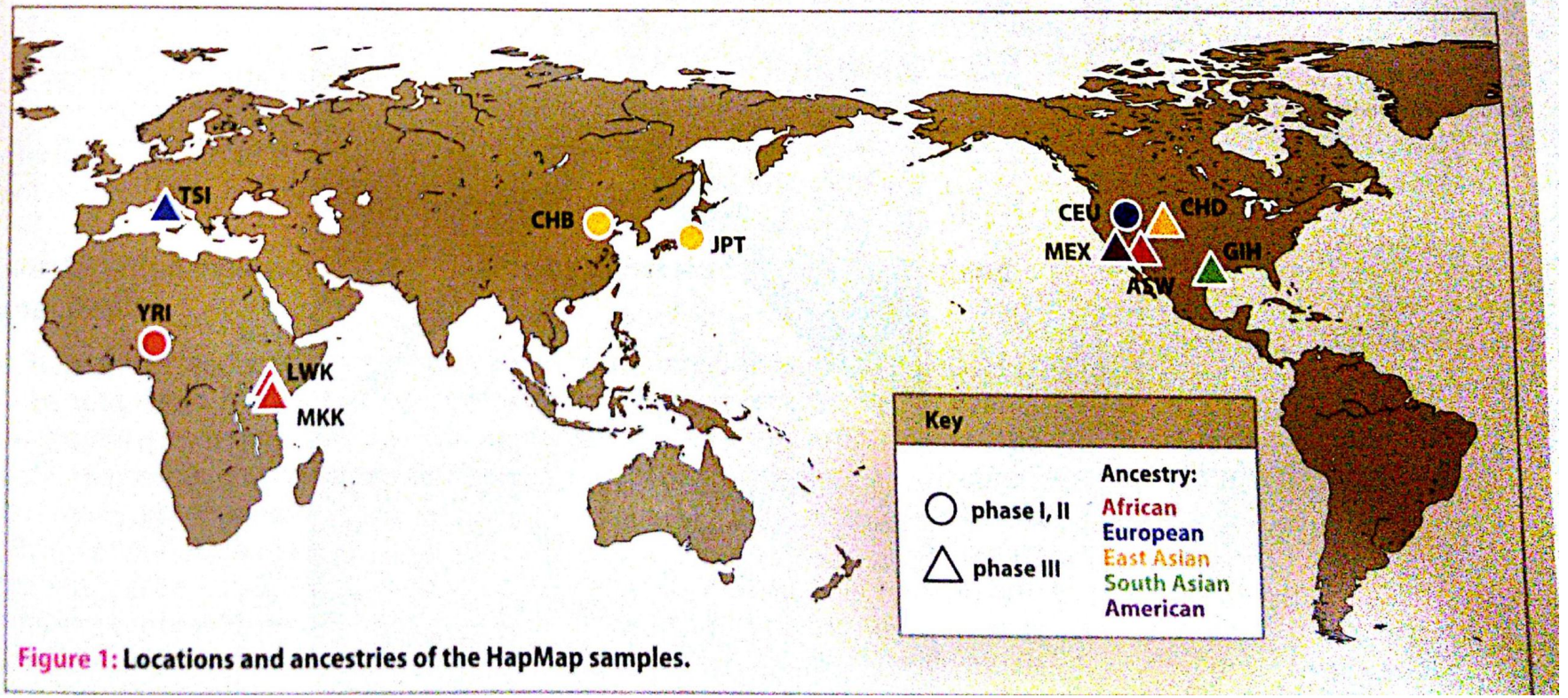
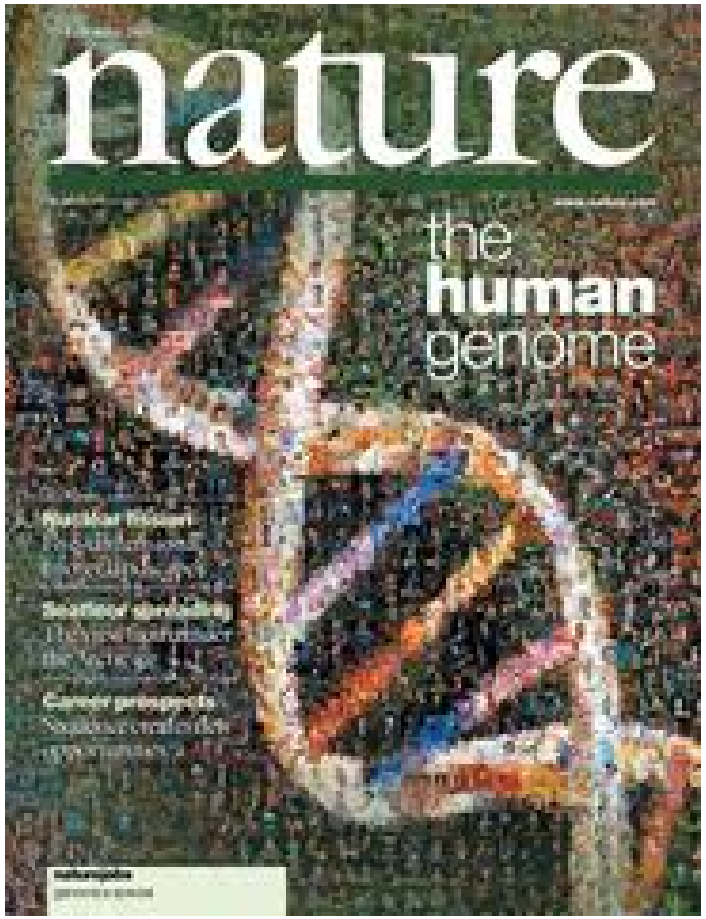
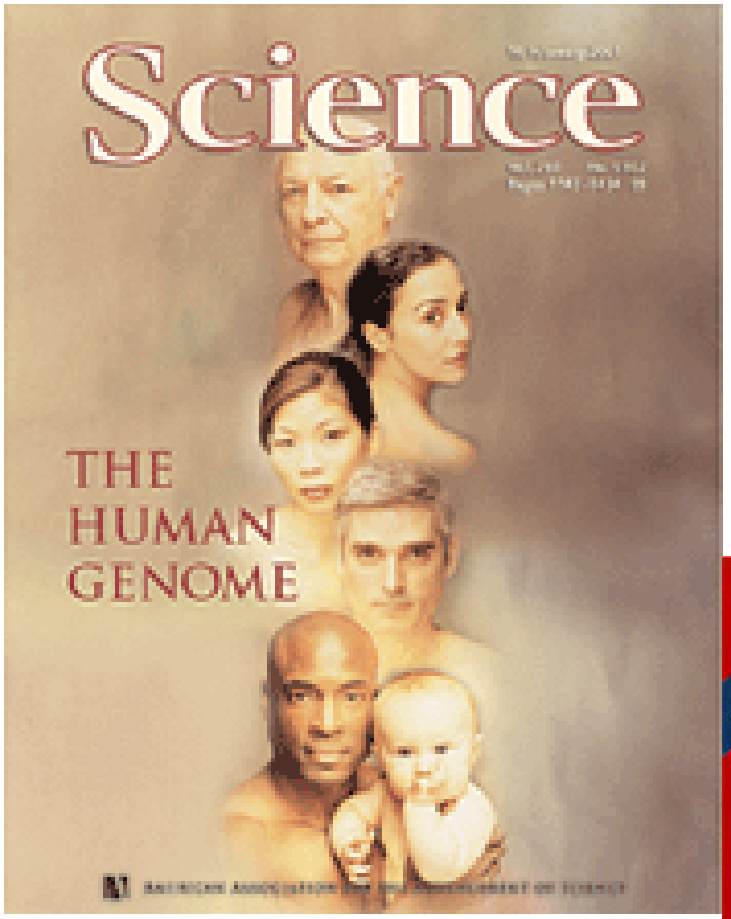


Figure 1: Locations and ancestries of the HapMap samples.

Febbraio 2001: primo annuncio completamento del sequenziamento del genoma umano



Consorzio Pubblico



Celera Genomics

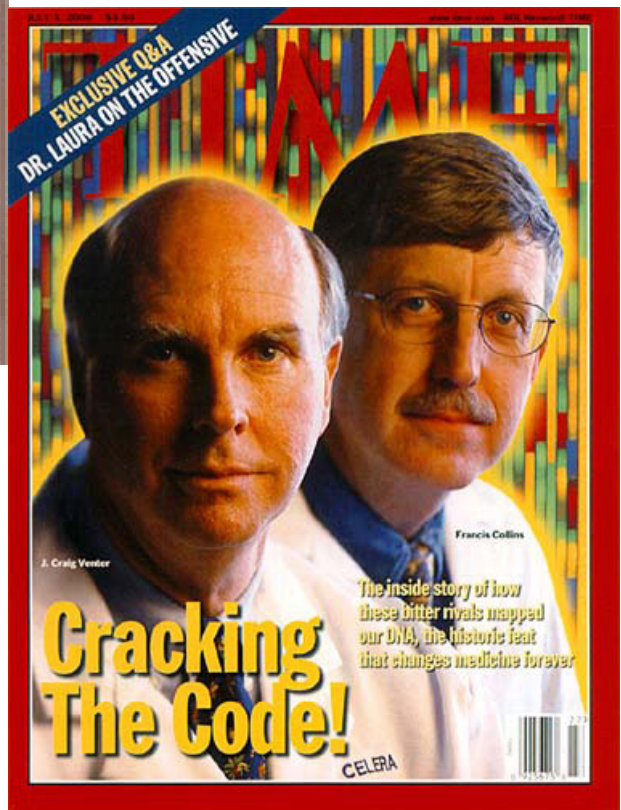


TABLE 10.1:
LARGE-SCALE PROJECTS PROVIDING INFORMATION ON HUMAN GENETIC VARIATION

Project	Launch date	Primary aims	Sample size	Populations	Genetic analyses	Cell lines	Data release	Website (Key reference) Further information
HGDP	1991	collection of isolated population samples	1064	51, worldwide	chosen by investigator	yes	on publication	http://www.cephb.fr/en/hgdp/diversity.php/ [Cann HM et al. (2002) *Science* 296, 261.] Box 10.2
HapMap	2002	haplotype map for medical genetics	1184	11, Africa, Europe, South and East Asia	SNP genotyping	yes	full public release	http://hapmap.ncbi.nlm.nih.gov/index.html.en [The International HapMap Project (2003) *Nature* 426, 789.] Box 3.6
Genographic	2005	elucidate migration history	~500,000	many, worldwide; including public participation	Y-SNP and Y-STR genotyping, mtDNA HVSI sequencing	no	on publication	https://genographic.nationalgeographic.com/genographic/lan/en/index.html [Wells, Deep Ancestry: Inside the Genographic Project (2006) National Geographic Books]
1000 Genomes	2008	discover variants at ≥1% frequency for medical genetics	2500	27, Africa, Europe, South Asia, East Asia, Americas	whole-genome sequencing	yes	full public release	http://www.1000genomes.org/ [The 1000 Genomes Project Consortium (2010) *Nature* 467, 1061.] Box 3.2

Jobling et al. Human Evolutionary genetics 2nd ed. 2014

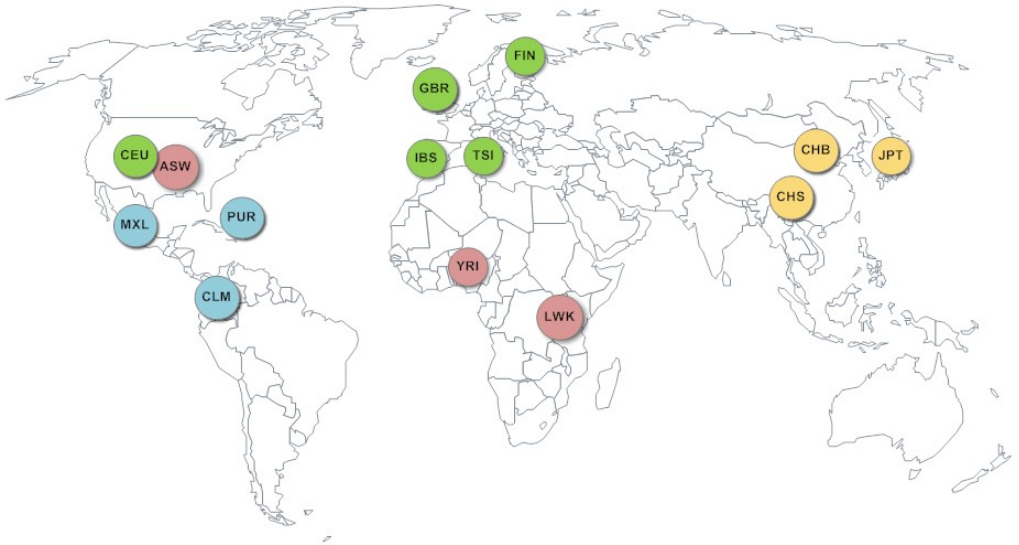
Il progetto 1000 genomi umani

Inizia nel **2008** e ha come scopo la costruzione di un catalogo dettagliato della variabilità genetica umana.

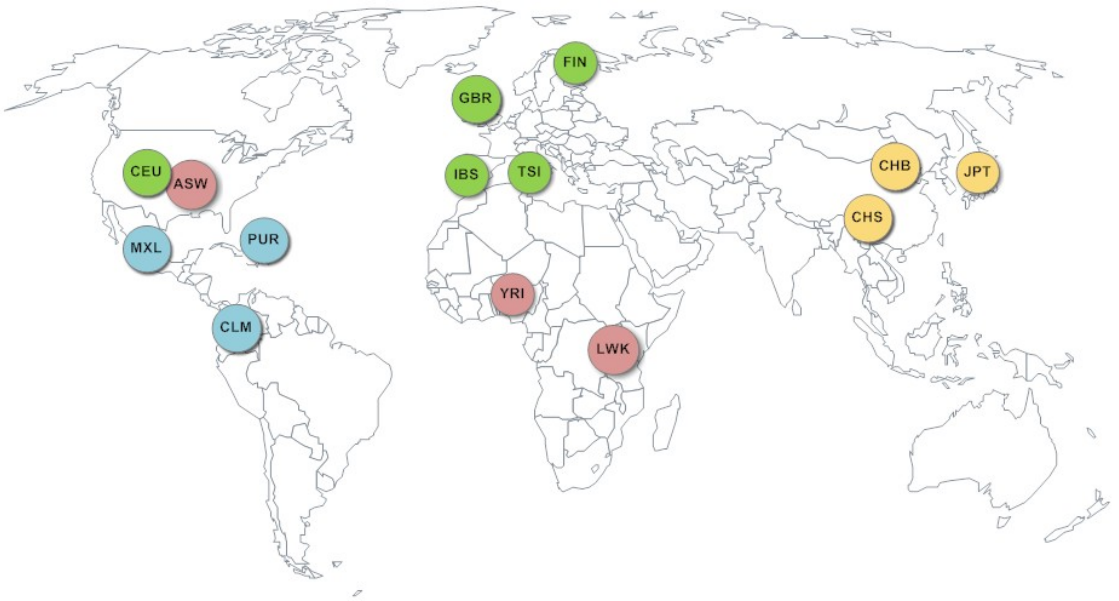
-**almeno 1000 genomi** di donatori anonimi da diverse popolazioni mondiali in 3 anni, usando metodi di sequenziamento massivo di nuova generazione, meno costoso e più veloce.

Hinxton, Cambridge, UK

-Il progetto pilota termina nel **2010**.
-Nel **2012** è stato annunciato il



Population Code	Population Description	Super Population Code
CHB	Han Chinese in Bejing, China	EAS
JPT	Japanese in Tokyo, Japan	EAS
CHS	Southern Han Chinese	EAS
CDX	Chinese Dai in Xishuangbanna, China	EAS
KHV	Kinh in Ho Chi Minh City, Vietnam	EAS
CEU	Utah Residents (CEPH) with Northern and Western European Ancestry	EUR
TSI	Toscani in Italia	EUR
FIN	Finnish in Finland	EUR
GBR	British in England and Scotland	EUR
IBS	Iberian Population in Spain	EUR
YRI	Yoruba in Ibadan, Nigeria	AFR
LWK	Luhya in Webuye, Kenya	AFR
GWD	Gambian in Western Divisions in the Gambia	AFR
MSL	Mende in Sierra Leone	AFR
ESN	Esan in Nigeria	AFR
ASW	Americans of African Ancestry in SW USA	AFR
ACB	African Caribbeans in Barbados	AFR
MXL	Mexican Ancestry from Los Angeles USA	AMR
PUR	Puerto Ricans from Puerto Rico	AMR
CLM	Colombians from Medellin, Colombia	AMR
PEL	Peruvians from Lima, Peru	AMR
GIH	Gujarati Indian from Houston, Texas	SAS
PJL	Punjabi from Lahore, Pakistan	SAS
BEB	Bengali from Bangladesh	SAS
STU	Sri Lankan Tamil from the UK	SAS
ITU	Indian Telugu from the UK	SAS



Fasi del progetto

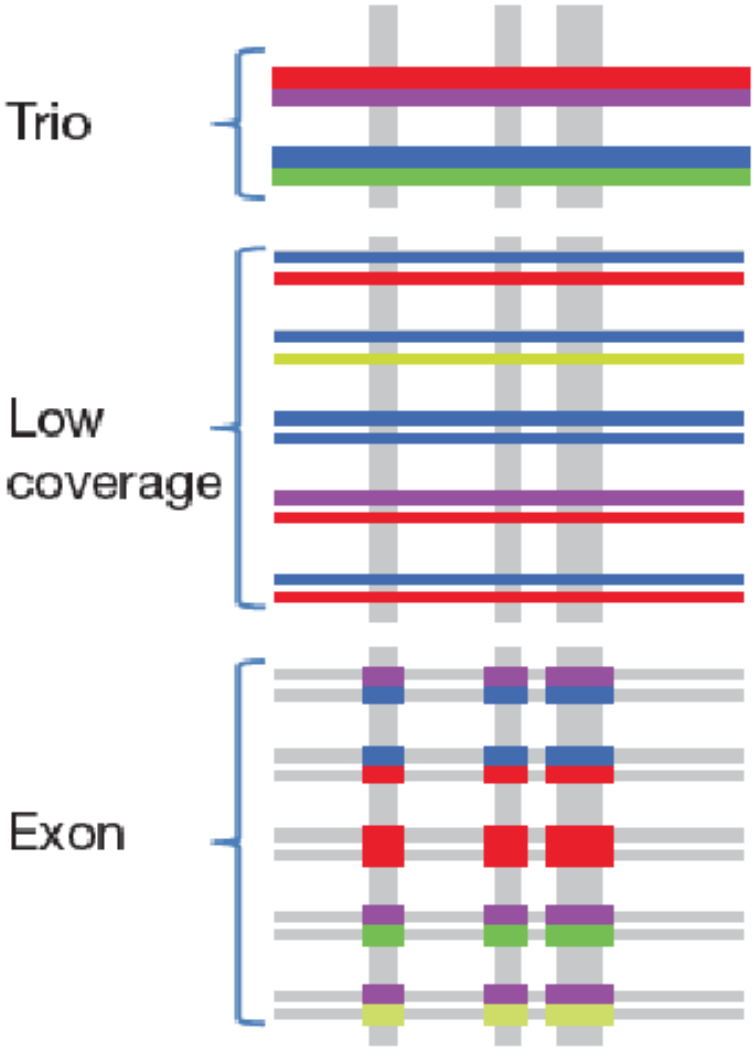
- **Pilot:** The 1000 Genomes project ran a pilot study between 2008 and 2010
- **Phase 1:** The initial round of exome and low coverage sequencing of 1000 individuals
- **Phase 2:** Expanded sequencing of 1700 individuals and method improvement
- **Phase 3:** Sequencing of 2500 individuals and a new variation catalogue
- **SAM/BAM:** Sequence Alignment/Map Format, an alignment format
- **VCF:** Variant Call Format, a variant format

Differenze principali con HapMap

Genomi, non SNP > anche le varianti molto rare vengono tipizzate (lo scopo è caratterizzare tutto ciò che è ≥ 0.1%)

Il sequenziamento di genomi interi fornisce informazioni anche su tutto il resto della variabilità (esempio regioni in cui sono presenti numeri di copie multiple di alcuni geni/sequenze, CNV o copy number variants)

Pilot project: Tre datasets



Due famiglie di origine africana ed europea, ognuna con due genitori e una figlia

A-C-T-G-C-A-C
A-G-G-A-A-T-C

Individual haploid genomes

Phased by transmission

20-60 X

59 YRI, 60 CEU, 30 CHB, 30 JPT non imparentati

A-.-T-G-C-A-C
A-.-G-G-A-T-C

Common haplotypes

Statistical phasing

2-4 X

. . T G . A .
. . G A . T .

Exon variants

Unphased

50 X

8140 esoni di 906 geni random (1.4 Mb tot) in 697 individui di 7 popolazioni (YRI, LWK, CEU, TSI, CHB, JPT, CHD)

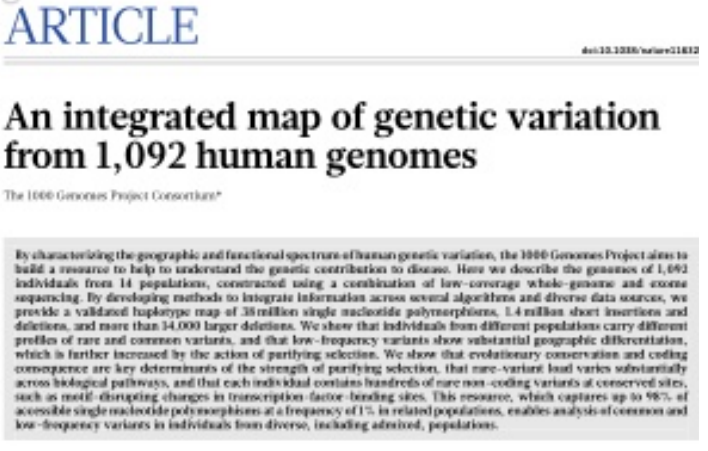
1000 Genomes

A Deep Catalog of Human Genetic Variation

Home | About | Data | Analysis | Participants | Contact | Browser | Wiki | FTP search

1000 Genomes Project Design and Progress

- Pilot data collected in 2008; paper published October 2010 in Nature
  - Companions in Science and Genome Research
  - Other companions later
- Full project data collection and analysis underway
  - Phase 1 results published Nov 1st 2012
  - Phase 2 / Phase 3 being completed
- Sequencing completion - early 2013
  - Analysis completion in 2013-2014



ARTICLE

doi:10.1038/nature11632

An integrated map of genetic variation from 1,092 human genomes

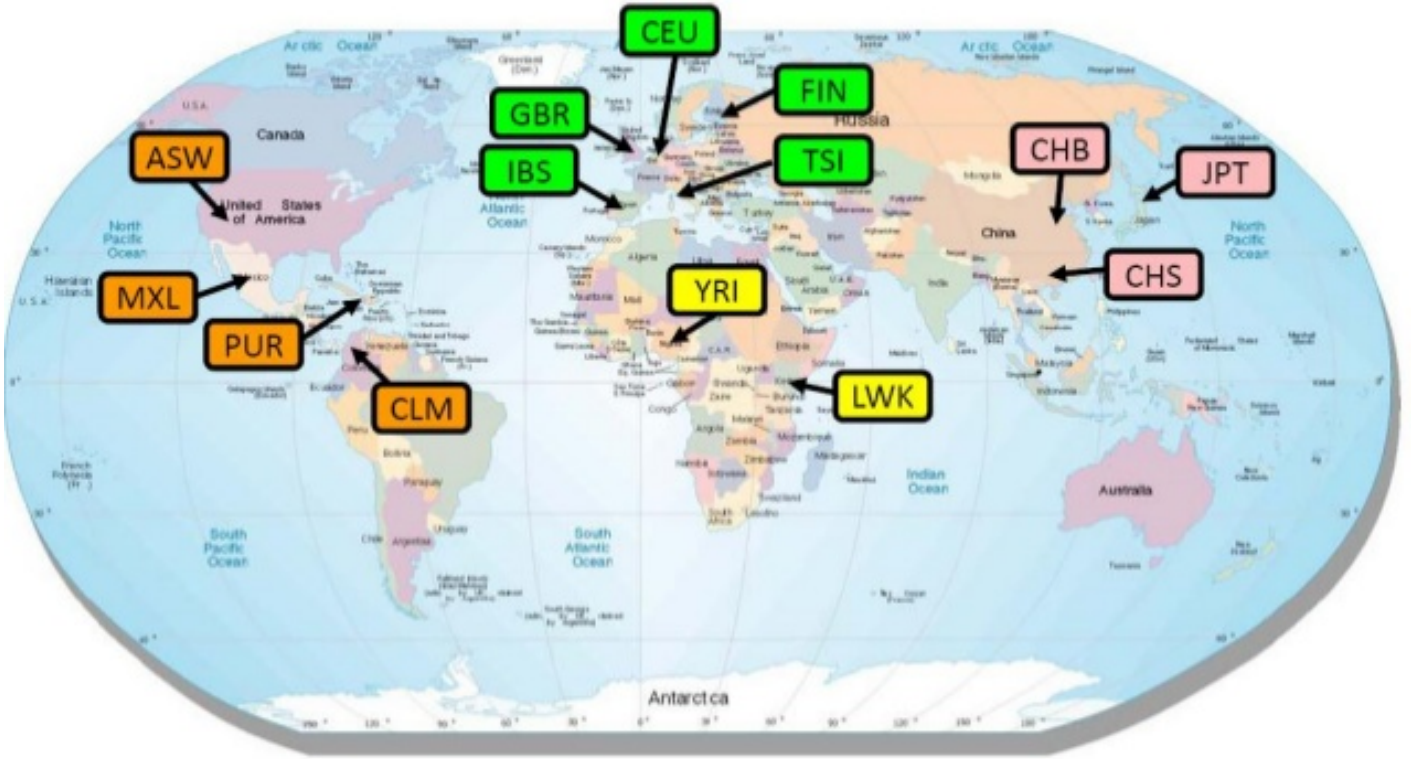
The 1000 Genomes Project Consortium*

By characterizing the geographic and functional spectrum of human genetic variation, the 1000 Genomes Project aims to build a resource to help to understand the genetic contribution to disease. Here we describe the genomes of 1,092 individuals from 14 populations, constructed using a combination of low-coverage whole-genome and exome sequencing. By developing methods to integrate information across several algorithms and diverse data sources, we provide a validated haplotype map of 38 million single nucleotide polymorphisms, 1.4 million short insertions and deletions, and more than 14,000 larger deletions. We show that individuals from different populations carry different profiles of rare and common variants, and that low-frequency variants show substantial geographic differentiation, which is further increased by the action of purifying selection. We show that evolutionary conservation and coding consequence are key determinants of the strength of purifying selection, that rare-variant load varies substantially across biological pathways, and that each individual contains hundreds of rare non-coding variants at conserved sites, such as motif-disrupting changes in transcription-factor-binding sites. This resource, which captures up to 98% of accessible single nucleotide polymorphisms at a frequency of 1% in related populations, enables analysis of common and low-frequency variants in individuals from diverse, including admixed, populations.

1000 Genomes

A Deep Catalog of Human Genetic Variation

1000G Phase I populations



HGSC

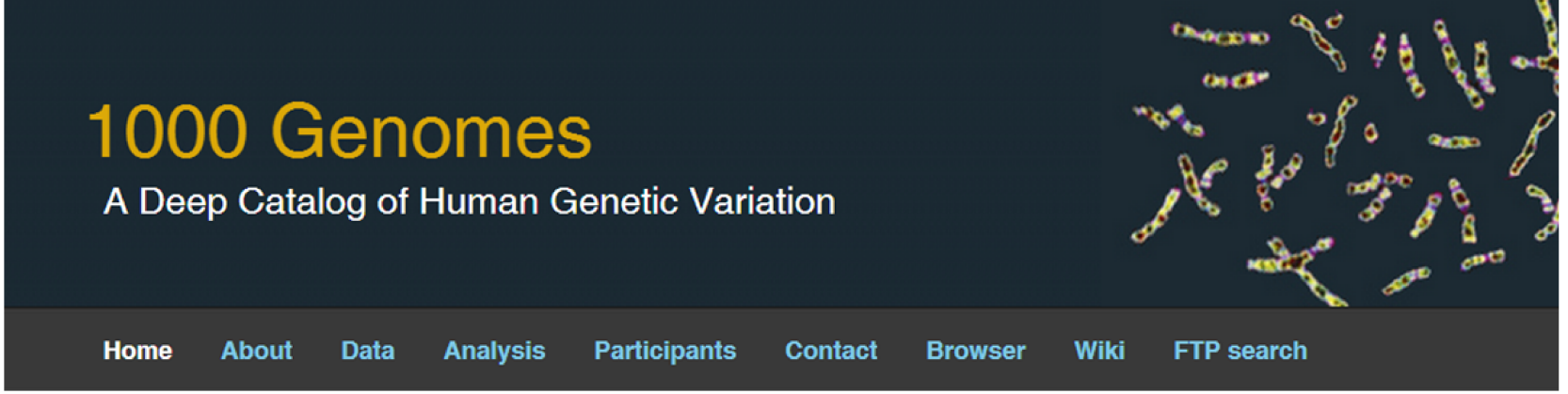


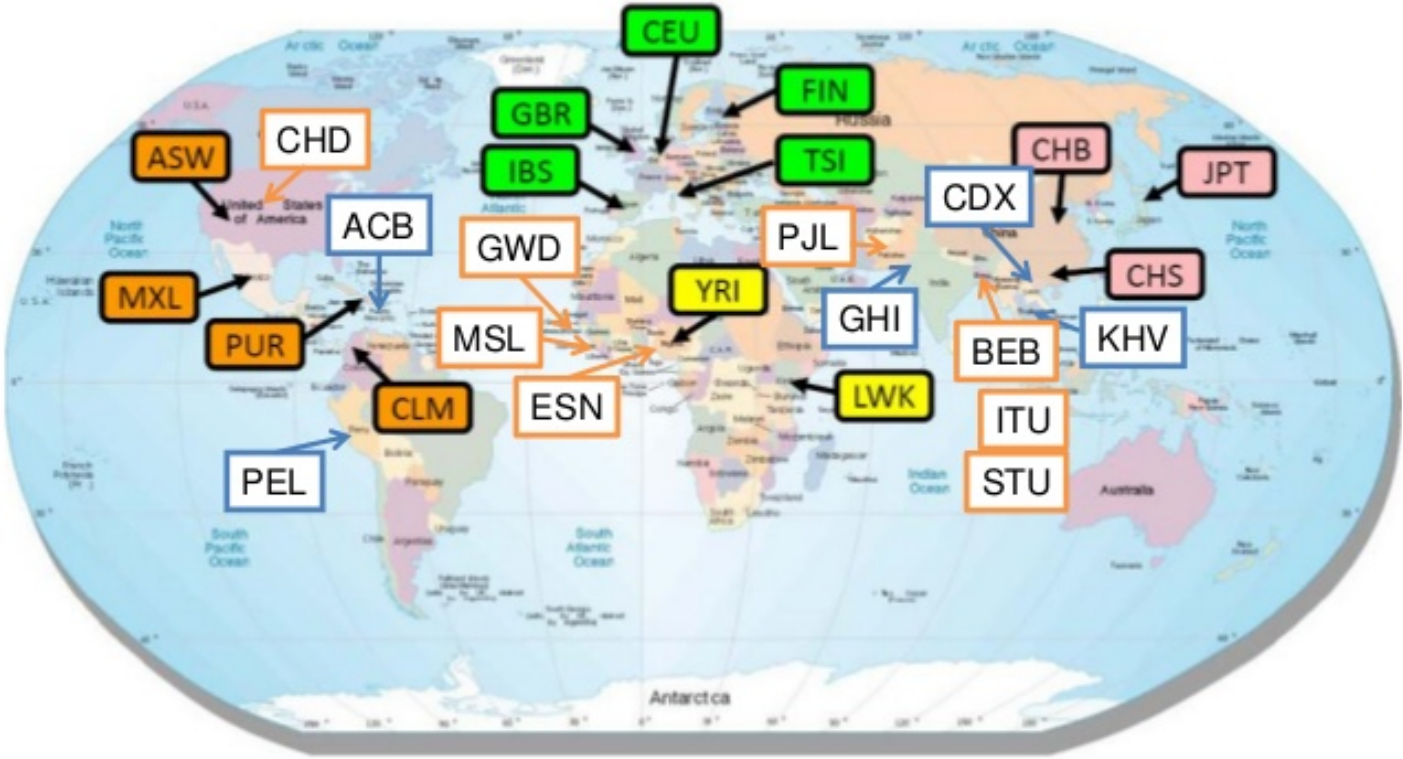
Table 1 | Summary of 1000 Genomes Project phase I data

	Autosomes	Chromosome X
Samples	1,092	1,092
Total raw bases (Gb)	19,049	804
Mean mapped depth (×)	5.1	3.9
SNPs		
No. sites overall	36.7 M	1.3 M
Novelty rate†	58%	77%
No. synonymous/non-synonymous/nonsense	NA	4.7/6.5/0.097 K
Average no. SNPs per sample	3.60 M	105 K
Indels		
No. sites overall	1.38 M	59 K
Novelty rate†	62%	73%
No. inframe/frameshift	NA	19/14
Average no. indels per sample	344 K	13 K
Genotyped large deletions		
No. sites overall	13.8 K	432
Novelty rate†	54%	54%
Average no. variants per sample	717	26

1000 Genomes Project Design and Progress

- Pilot data collected in 2008; paper published October 2010 in Nature
  - Companions in Science and Genome Research
  - Other companions later
- Full project data collection and analysis underway
  - Phase 1 results published Nov 1st 2012
  - Phase 2 / Phase 3 being completed
- Sequencing completion - early 2013
  - Analysis completion in 2013-2014

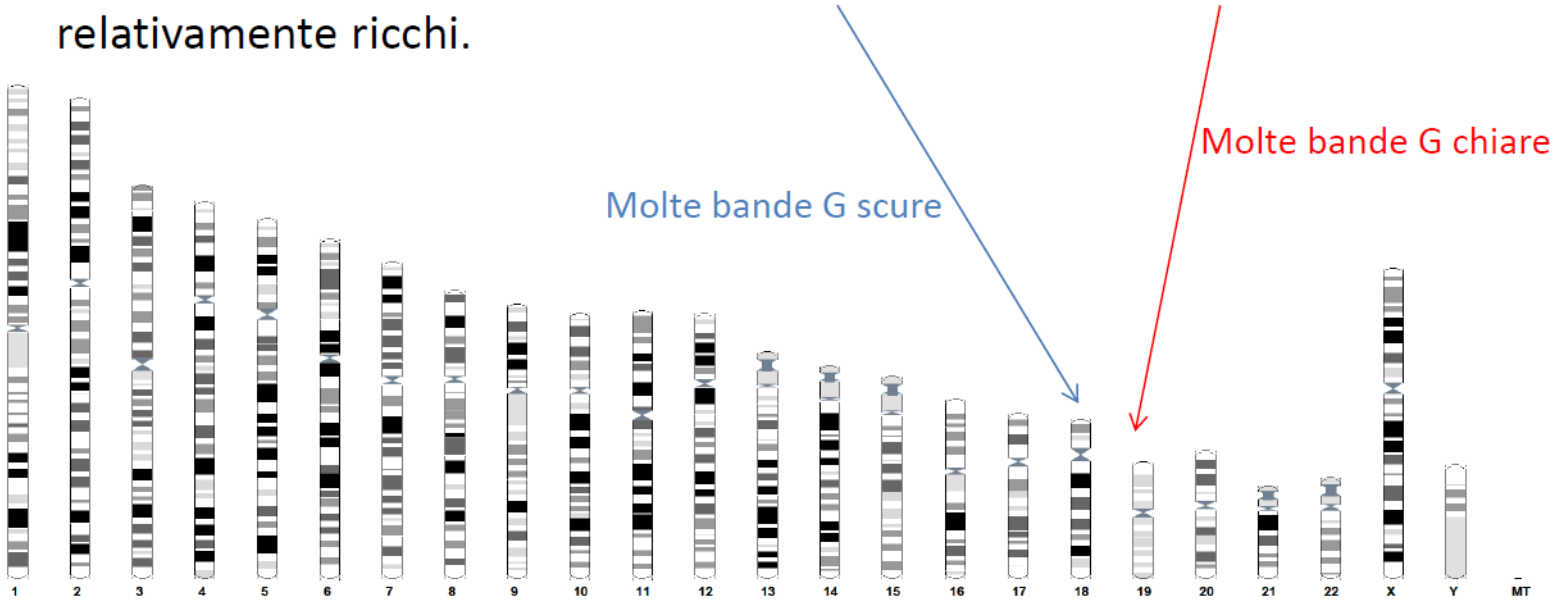
1000G Phase 2/3 populations



Alcuni esempi di geni codificanti

Regioni codificanti il proteoma

- Genoma umano: ~ 20.000 geni codificanti proteine. Alcune regioni povere: regioni subtelomeriche, cromosoma 18 e X. Viceversa il 19 e il 22 sono relativamente ricchi.



http://www.ensembl.org/Homo_sapiens/Location/Genome

Regioni codificanti il proteoma

- Esoni e introni (con eccezioni): lunghezza media di un esone: 200 bp

- La grande varietà di dimensione dei geni dipende soprattutto dagli introni: gene per l'insulina 1.7 kb, LDL receptor 5.45 kb, dystrophin gene 2400 kb

- Esistono regioni codificanti su entrambi i filamenti

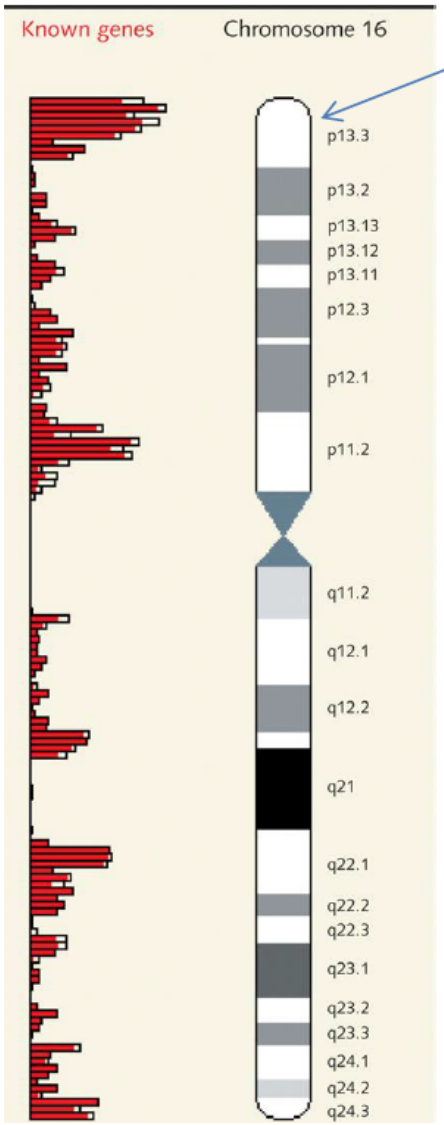
http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000254647;r=11:2159779-2161341
http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000198947;r=X:31097677-33339441

http://www.ensembl.org/

REF SEQ: http://www.ncbi.nlm.nih.gov/nuccore/NG_012232.1

Un esempio specifico: la famiglia delle globine

Il locus **α-globin** sul cromosoma 16



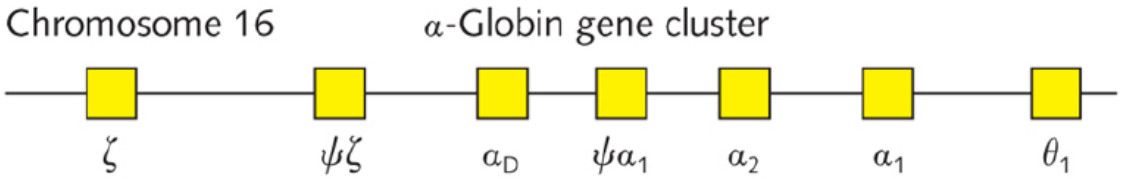
p13.3

NCBI map viewer: regione genomica p13.3 cromosoma 16
Regione 100k-200k

NCBI map viewer > human > cromosome 16

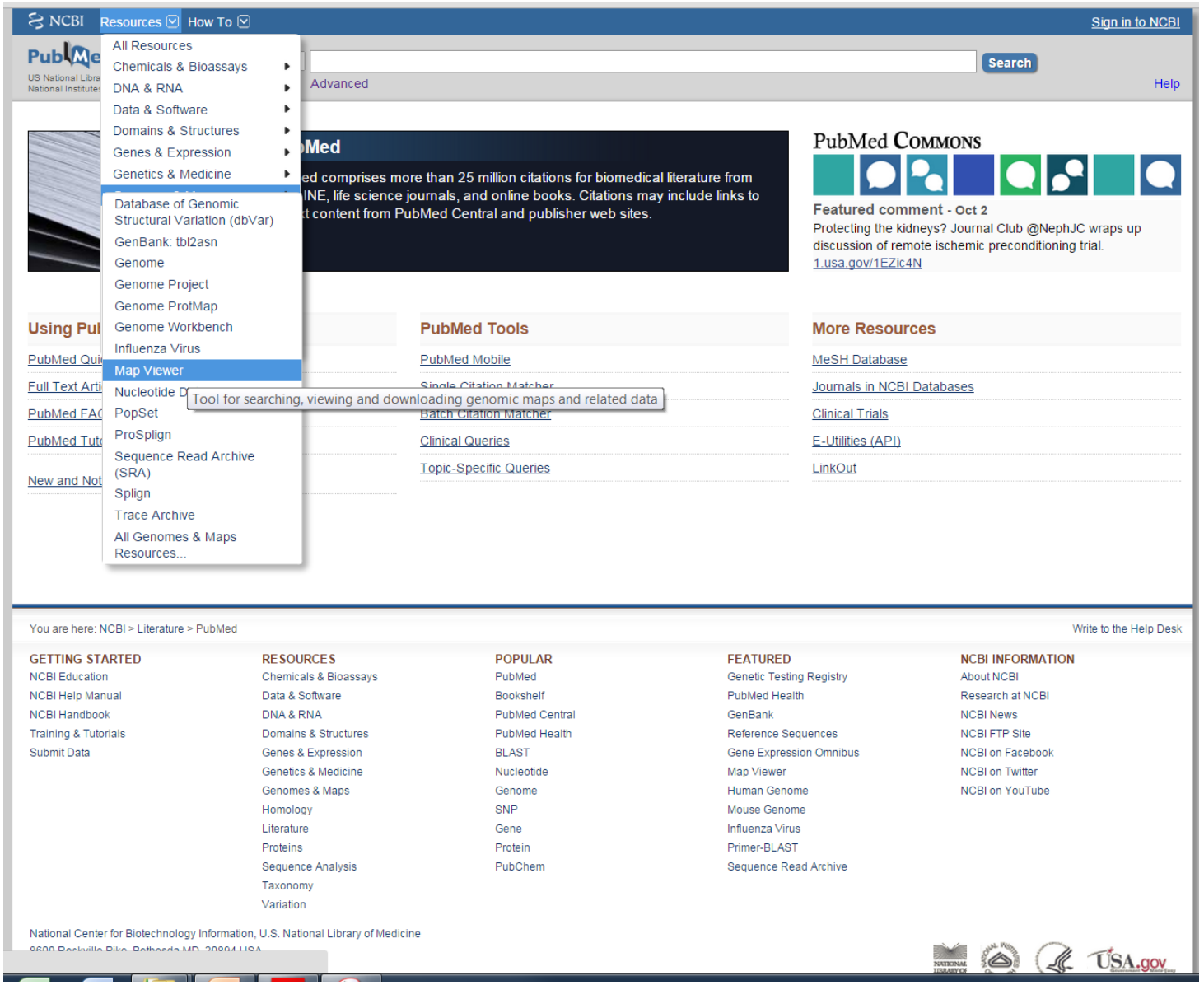
The α-globin locus:

Cinque geni espressi, ζ, α_D, α_2, α_1, e θ_1,
e due pseudogeni, $\psi\zeta$ *e* ψa_1

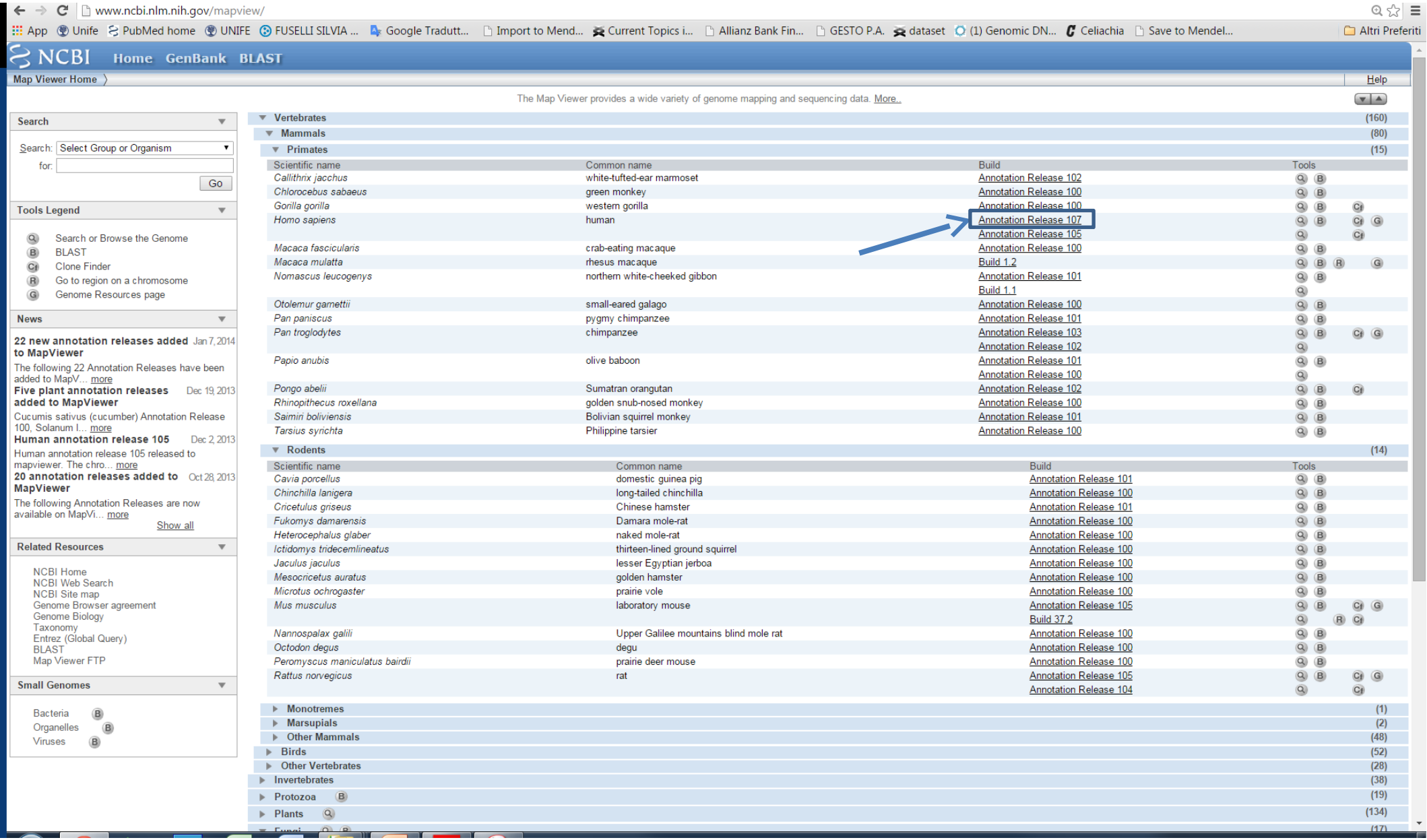


Lesk, Introduction to genomics 2nd ed

NCBI MapViewer



NCBI MapViewer > Homo sapiens annotation Release 107



www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606&build=107.0

App | Unife | PubMed home | UNIFE | FUSELLI SILVIA ... | Google Tradutt... | Import to Mend... | Current Topics i... | Allianz Bank Fin... | GESTO P.A. | dataset | (1) Genomic DN... | Celiachia | Save to Mendel... | Altri Preferiti

NCBI

NCBI Map Viewer

PubMed | Nucleotide | Protein | Genome | Gene | Structure | PopSet | Taxonomy | Help

Search for [ ] on chromosome(s) [ ] assembly [All] Find Advanced Search

Map Viewer

- Map Viewer Home
- Map Viewer Help
- Human Maps Help
- FTP

NCBI Resources

- Assembly
- CCDS
- Gene
- Genome
- RefSeq

Organism Data in GenBank

- EST
- Genomic
- mRNA
- Protein
- SRA
- WGS

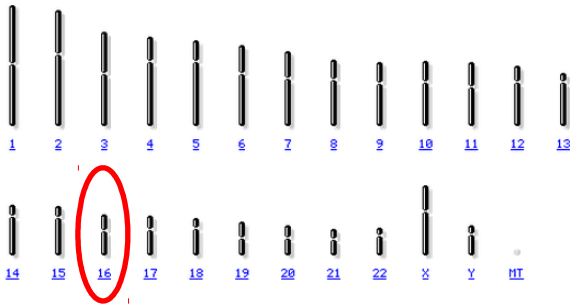
Related Resources

- GRC
- HGNC
- OMIM

Homo sapiens (human) genome view

BLAST search the human genome

Annotation Release 107 statistics Switch to previous build



Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo; Homo sapiens

March 2015, Annotation Release 107

This full annotation run includes the following assembly(ies):

- GRCh38.p2 (accession GCF_000001405.28)
- CHM1_1.1 (accession GCF_000306695.2)

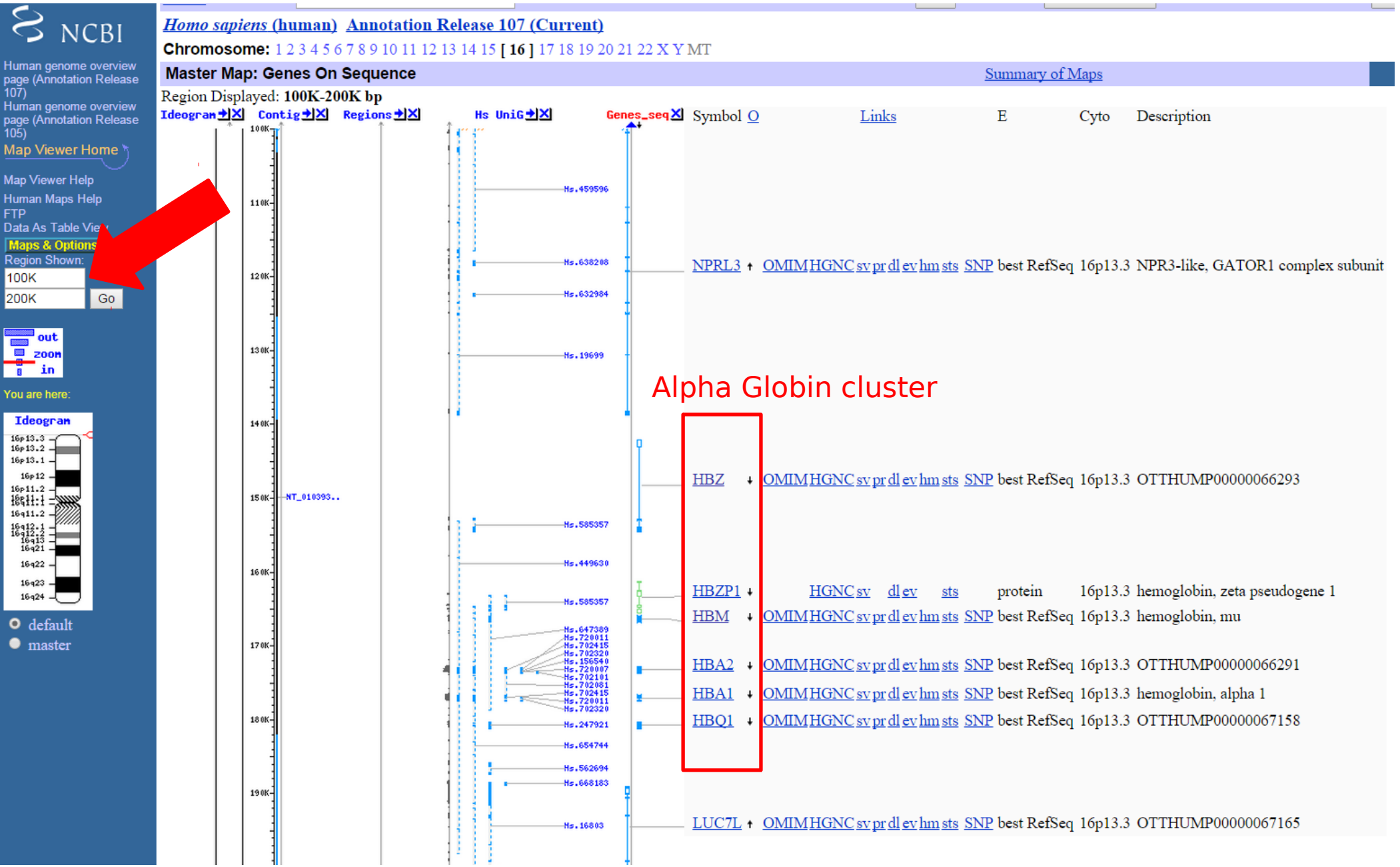
The NCBI Map Viewer provides graphical displays of features on the genome assembly. Map features that can be seen along the sequence include annotated genes and transcripts, Gnomon-predicted gene and transcript models, aligned transcript and genomic sequences, RefSeq scaffolds (the 'Contig' map), the assembly tiling path (the 'Component' map), and more. For some species, additional non-sequence maps such as Genetic maps, Radiation hybrid maps, and others may be available.

Available Documentation:

- Species Maps Help
- General Map Viewer Help
- NCBI Handbook: Map Viewer chapter
- NCBI Handbook: Map Viewer Exercises
- NCBI Annotation Process
- Gnomon

BI MapViewer > Homo sapiens annotation Release 107 > chromosome 16 >

ion shown 100K-200K



NCBI MapViewer > Homo sapiens annotation Release 107 > chromosome 11 >

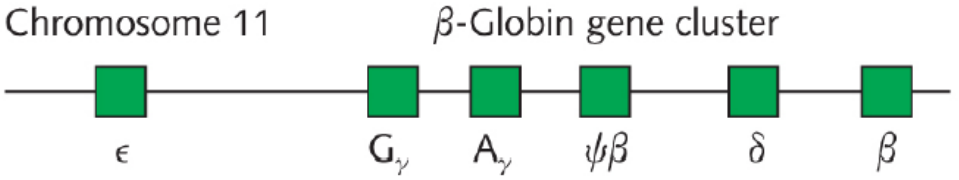
Region shown 5.2M-5.3M

Il lucus **β-globin** sul cromosoma 11

NCBI map viewer: regione genomica p15 cromosoma 11

The β-globin locus:

Quattro geni espressi e uno pseudogene



Lesk, Introduction to genomics 2nd ed

Homo sapiens (human) Annotation Release 107 (Current)

BLAST human sequences

Chromosome: 1 2 3 4 5 6 7 8 9 10 [**11**] 12 13 14 15 16 17 18 19 20 21 22 X Y MT

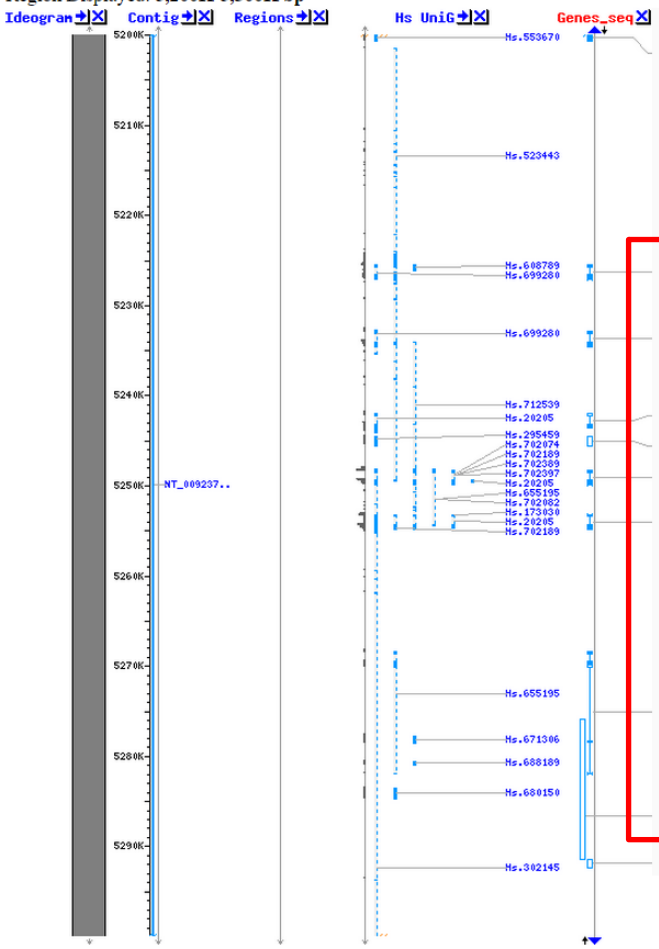
Master Map: Genes On Sequence

Summary of Maps

Maps & Options

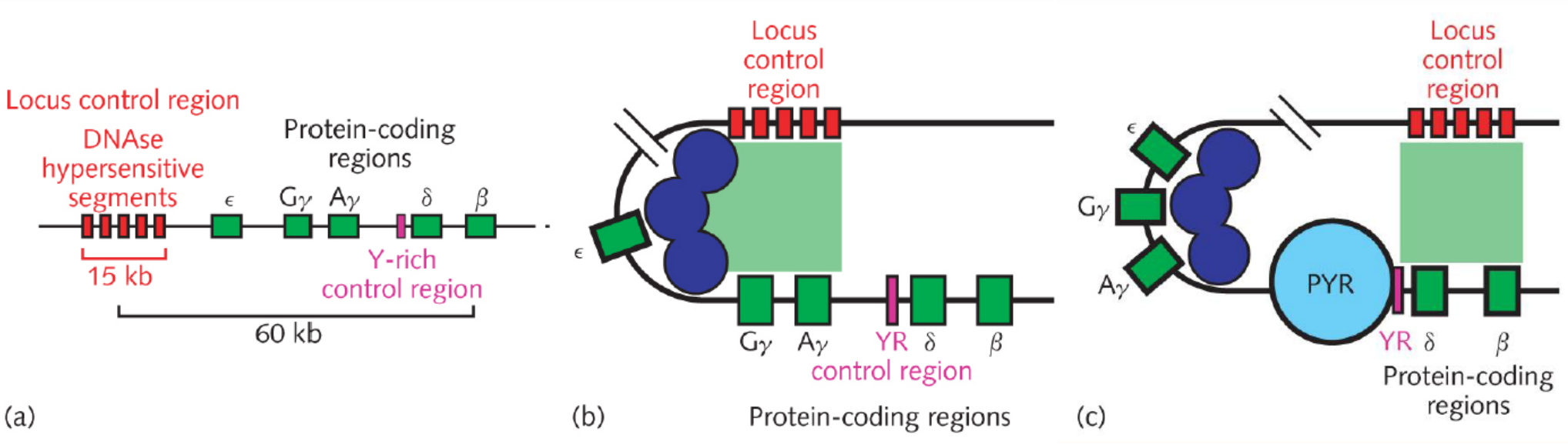
Region Displayed: **5,200K-5,300K bp**

Download/View Sequence/Evidence



Symbol	O	Links	E	Cyto	Description
OR51V1	↑	HGNC sv pr dl ev hm sts SNP	best RefSeq	11p15.4	olfactory receptor, family 51, subfamily V, member 1
HBB	↑	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	11p15.5	hemoglobin, beta
HBD	↑	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	11p15.5	OTTHUMP00000069639
HBBP1	↑	HGNC sv dl ev sts SNP	best RefSeq	11p15.5	hemoglobin, beta pseudogene 1
BGLT3	↑	OMIM HGNC sv dl ev SNP	best RefSeq	11p15.4	beta globin locus transcript 3 (non-protein coding)
HBG1	↑	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	11p15.5	hemoglobin, gamma A
HBG2	↑	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	11p15.5	hemoglobin, gamma G
HBE1	↑	OMIM HGNC sv pr dl ev hm sts SNP	best RefSeq	11p15.5	hemoglobin, epsilon 1
LCRB	↑	OMIM sv dl ev	best RefSeq	11p15.5	locus control region, beta
OR51AB1P	↑	HGNC sv dl ev sts	best RefSeq	11p15.4	olfactory receptor, family 51, subfamily AB, member 1 pseudogene

Controllo dell'espressione del cluster β: locus control region



β-Globin region: geni codificanti e Locus Control Region, che consiste di cinque DNAse hypersensitive segments.

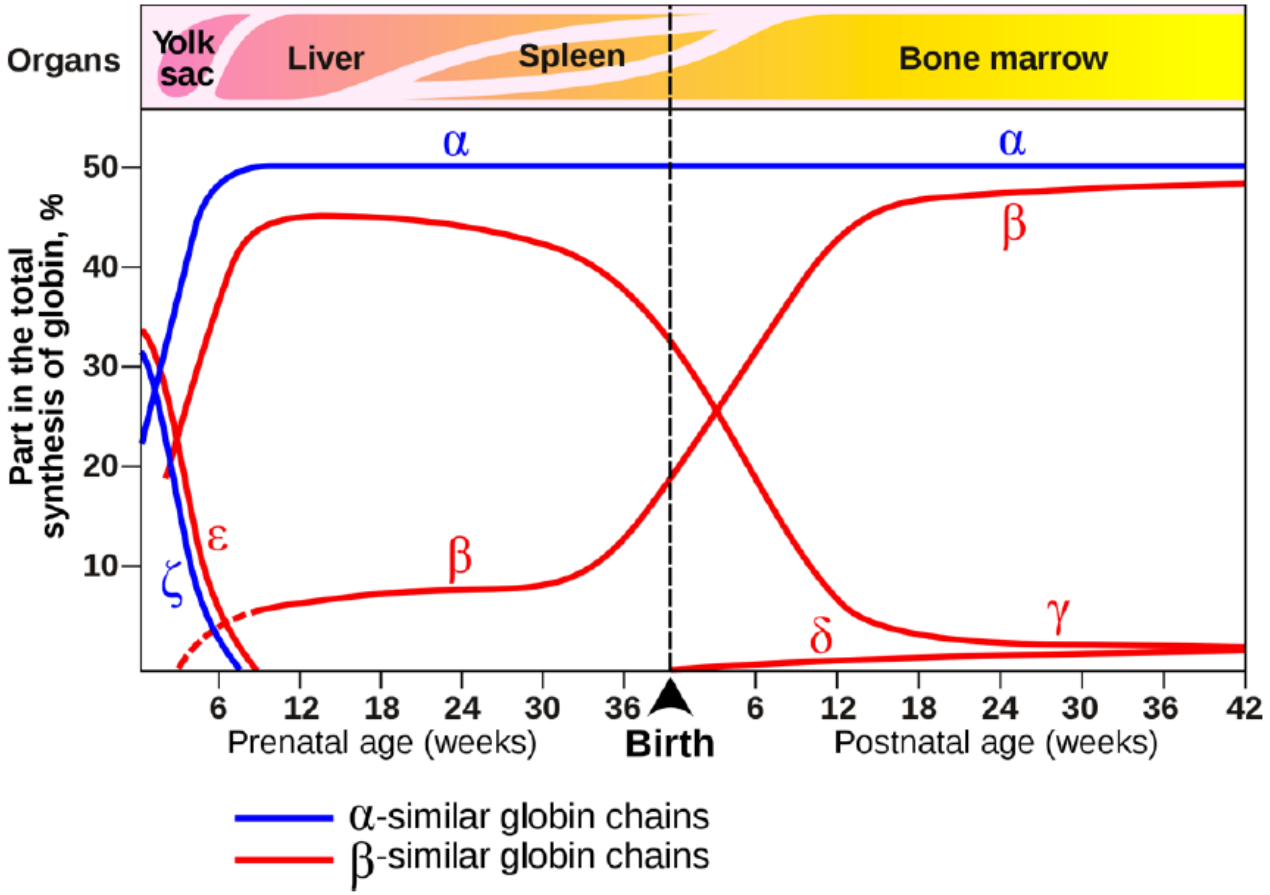
Feto: interazione tra il locus control region e i geni Gγ and Aγ genes, mediata da proteine. Cerchi blu: complessi di rimodellamento della cromatina.

Cerchio azzurro: complesso proteico PYR che lega la regione ricca di pirimidine (YR; Y = pirimidine) al 5' del gene. Questo legame riconfigura il sistema: PYR blocca il controllo "fetale", menter la locus control region promuove l'espressione del gene β.

Lesk, Introduction to genomics 2nd ed

L'ordine dei geni sul cromosoma ha un significato: la trascrizione segue rigidamente lo sviluppo dell'organismo.

Embrione (fino a 6 settimane dopo il concepimento): tetramero $\zeta_2\varepsilon_2$. A seguire, fino a ceirca 8 settimane dopo la nascita la specie predominante diventa l'emoglobina fetale, $\alpha 2\gamma 2$, segue poi l'emoglobina adulta, principalemente $\alpha 2\ \beta 2$.



L'emoglobina nell' adulto e di tipo HbA 96% α2β2, HbA2 α2δ2 3% e HbF α2γ2 1%

Numerose duplicazioni e divergenze

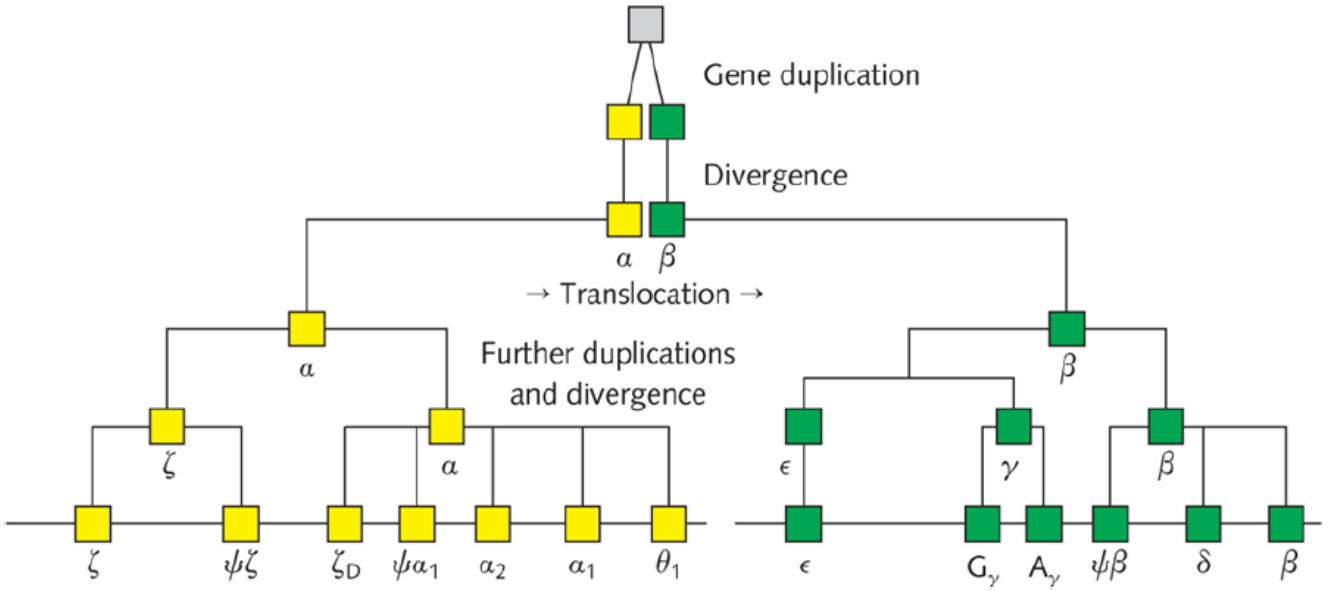


Figure 1.11 Haemoglobin genes and pseudogenes are distributed on their chromosomes in a way that appears to reflect their evolution via duplication and divergence. That is, adjacent genes are similar in sequence. The evolutionary tree can be drawn without any intersecting lines.

Paraloghi: geni che si sono originati da un antenato comune in seguito a duplicazione genica.

Lesk, Introduction to genomics 2nd ed