Xenómica para o mundo contemporáneo

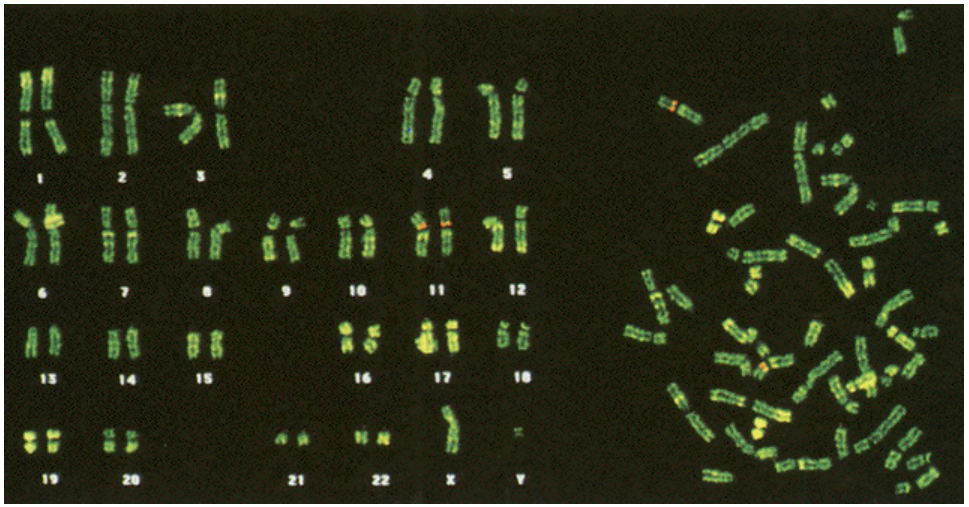
Angel Carracedo

Instituto de Medicina Legal e Centro Nacional de Xenotipado-Universidade de Santiago

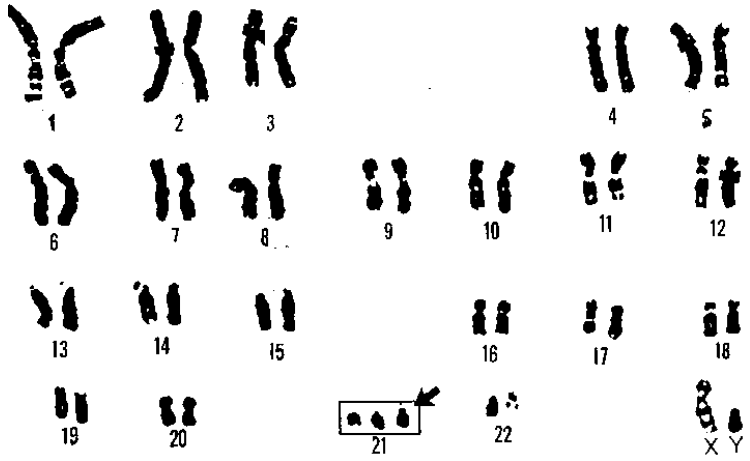
Fundación Galega de Medicina Xenómica

Climántica febreiro 2010

EL CARIOTIPO HUMANO



23 pares de cromosomas: 1956

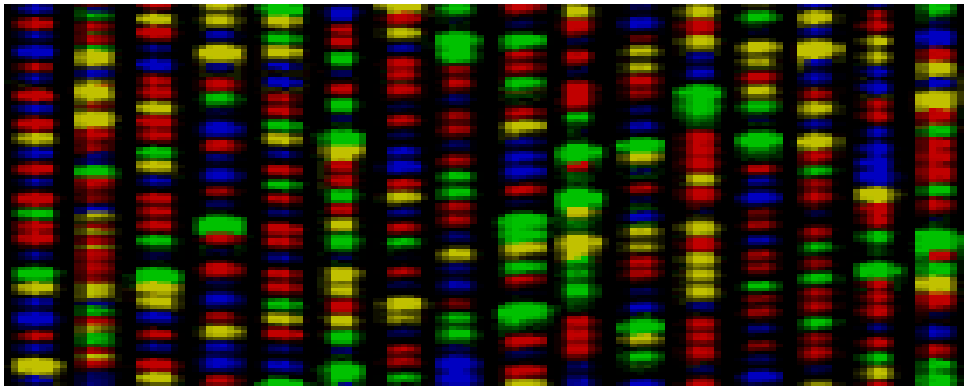


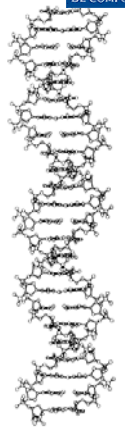
J. Lejeune (1960)

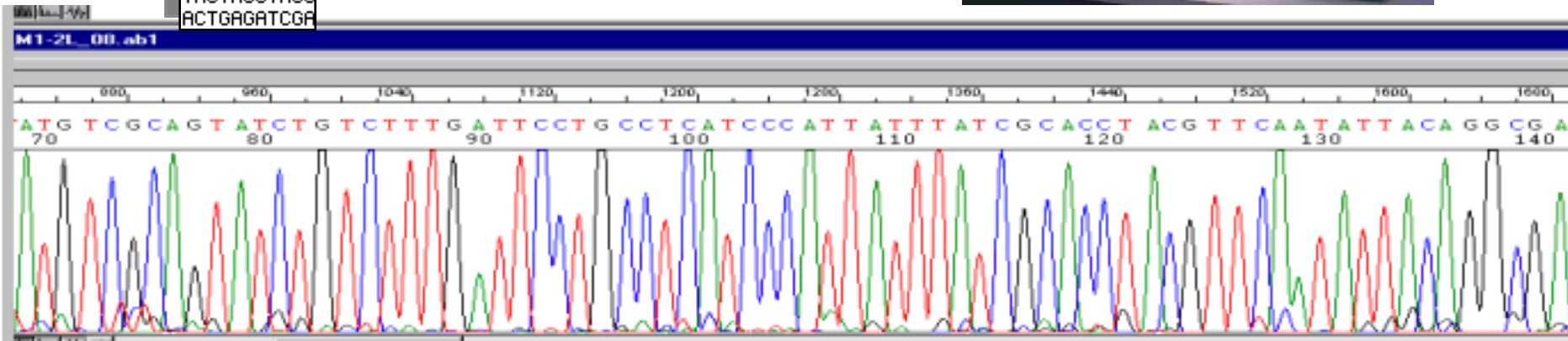
Fundación Pública Galega de Medicina Xenómica

1985-1995

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

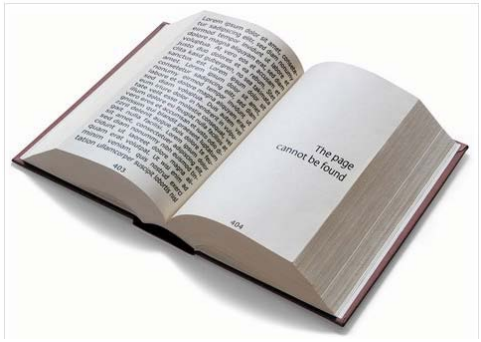






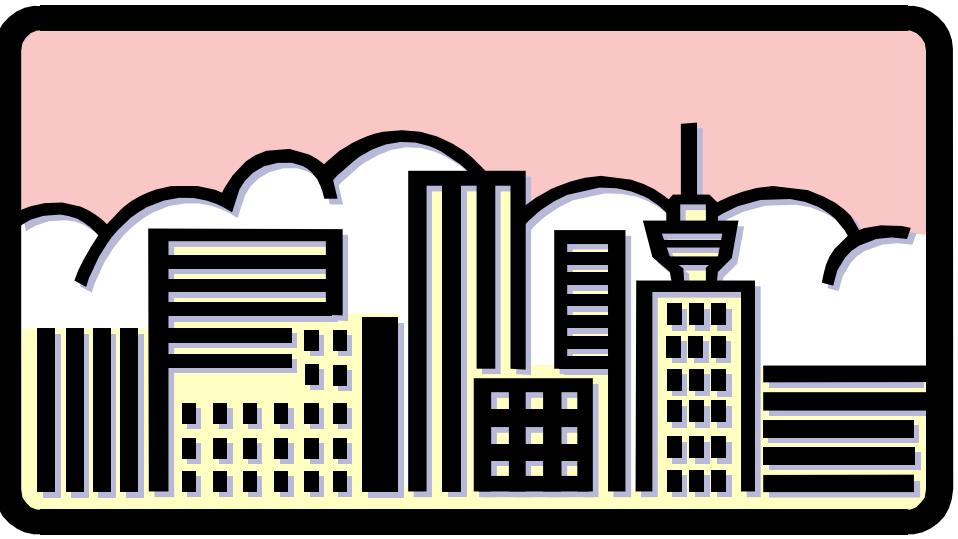
El genoma es...

- El conjunto completo de genes de un organismo, donde se guarda toda la información genetica.



3.200.000.000 de pares de bases (A, T, C, G)

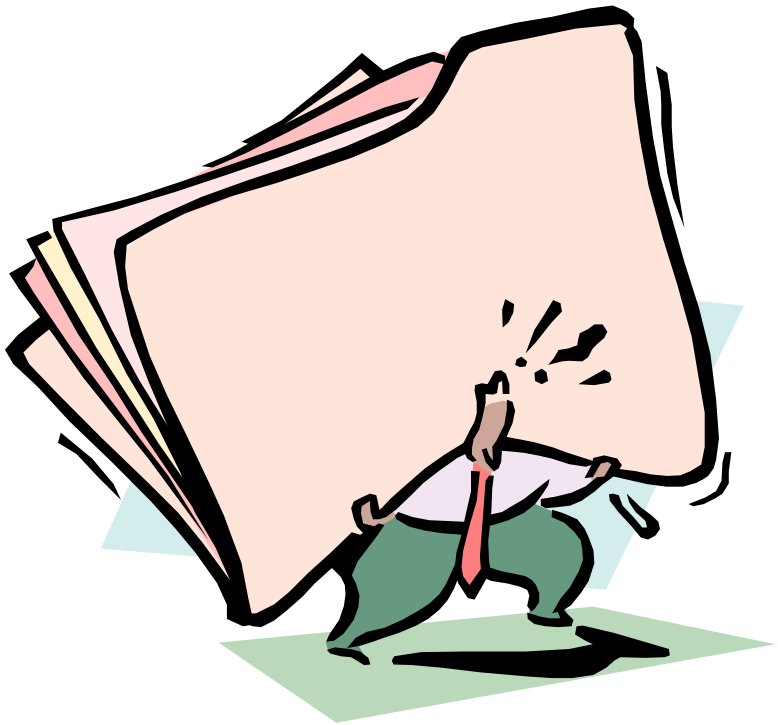
Human Genome Project

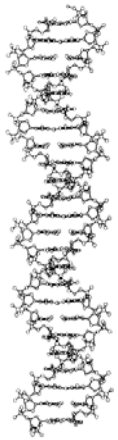


GENOMA HUMANO

3.000.000.000 de pares de bases (A, T, C, G)

Alrededor de 30.000 genes



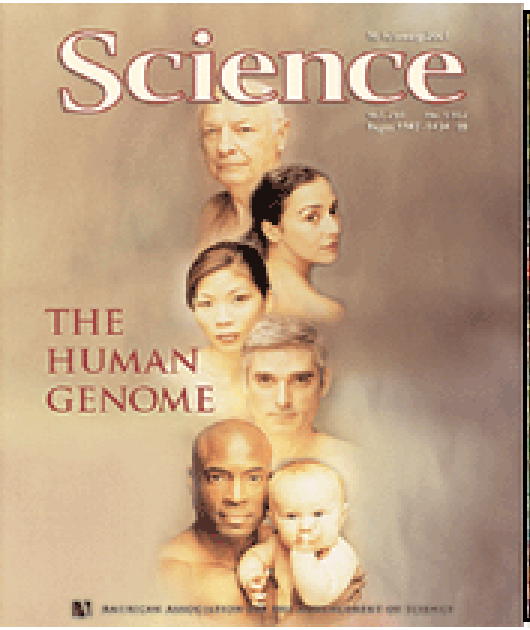


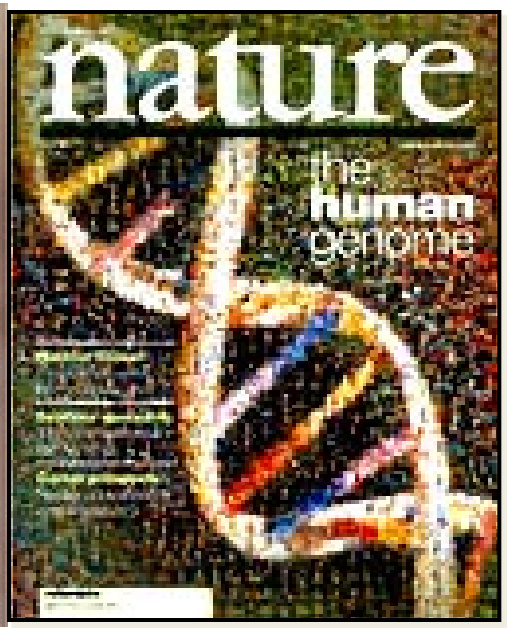
ATACCTGCGTCGGATGCTGCGATTGCTGACCAACATCGTGACAGTTAG
ACAAACGATTGACTGTTAGGATTGACCACCAATTACGATGACGTTGG…

CeGen
USC

Fundación Pública Galega de Medicina Xenómica

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA





EMBL Outstation
European Bioinformatics Institute

OMIM
Online Mendelian Inheritance in Man
Johns Hopkins University

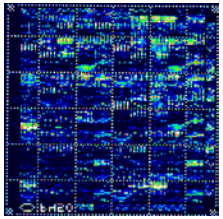
GeneTests

HGMD™
The Human Gene Mutation Database
Cardiff

PDB™
PROTEIN DATA BANK

SWISS-PROT

GeneCards™



¿Qué sabemos?

*Que tenemos unos 3 mil millones de pares de bases (A,C,T, G)

*Que un gen tiene de media 3000 bases pero varía mucho su tamaño

*Que el número de genes total que tenemos es alrededor de 26,000

*Que los seres humanos somos un 99,9% idénticos

*Que sólo el 2% del genoma codifica para proteínas

*Que la mayoría del ADN es no codificante

*Que sobre un 50% del ADN son frases que se repiten (mucho más que ninguna otra especie)

¿Qué no sabemos?

*El número exacto de los genes

*La función de la mayoría

*La manera como interaccionan

*La mayoría de los genes implicados en enfermedades comunes

*Muchas de las funciones del ADN no codificante

*La coordinación exacta de la expresión génica, síntesis de proteínas y eventos postraslacionales.

*El proteoma (Contenido total de proteínas, su función e interrelaciones)

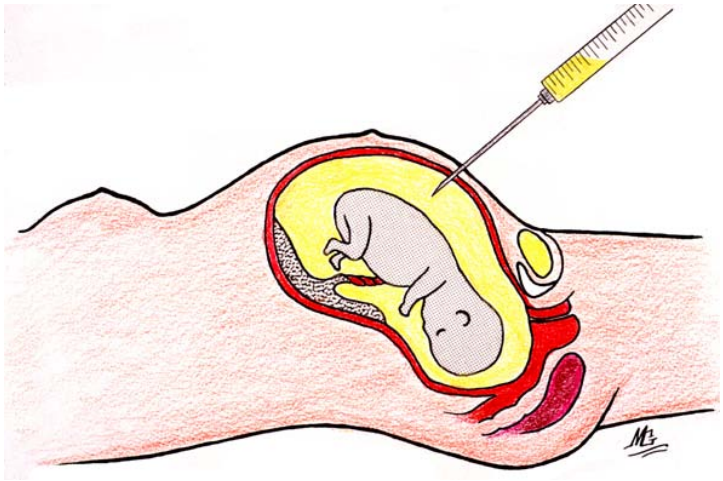
Genética médica: Fundación Galega de Medicina Xenómica

UTILIDAD DE LOS ANÁLISIS GENÉTICOS

*DIAGNÓSTICO

*PRONÓSTICO

*TRATAMIENTO



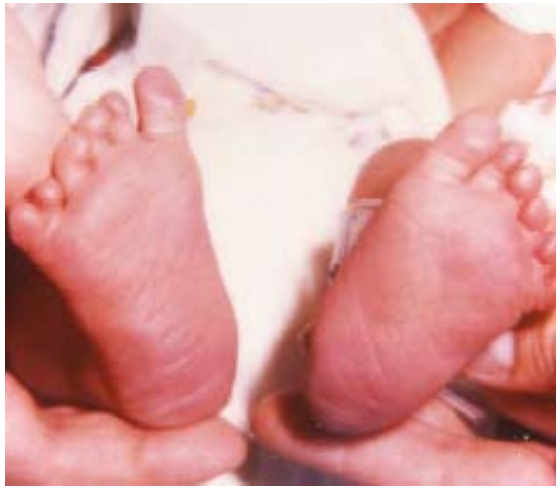
Consejo genético

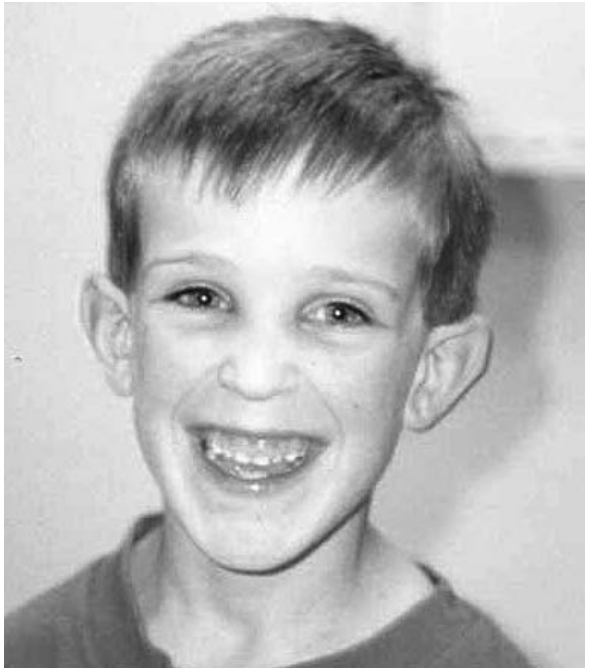
Diagnóstico prenatal

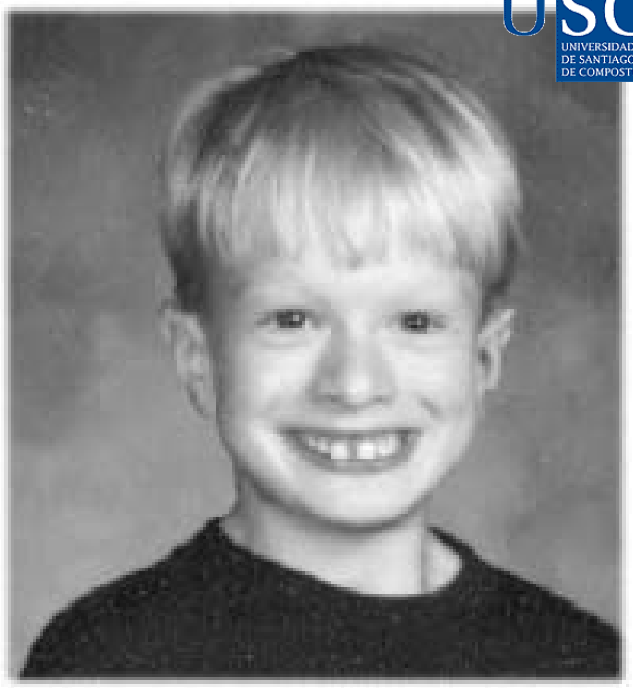
Diagnóstico preimplantacional

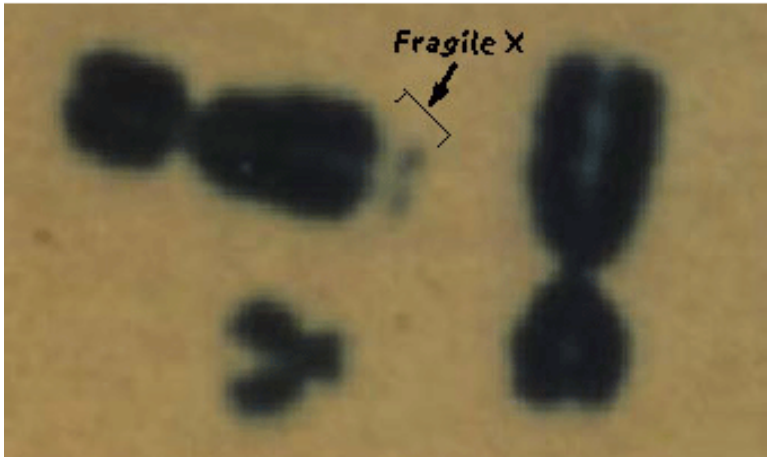
Fundación Pública Galega de Medicina Xenómica

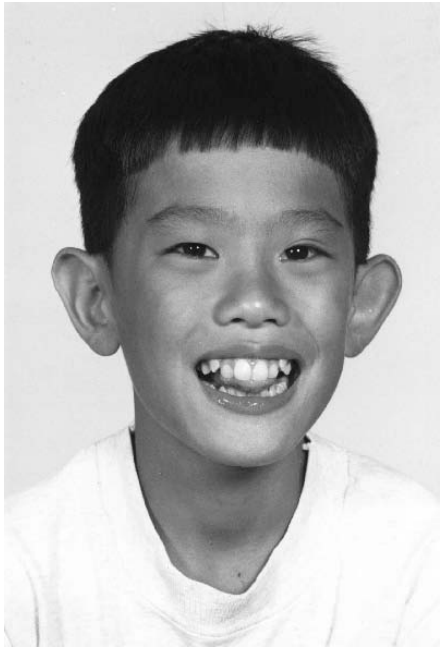
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

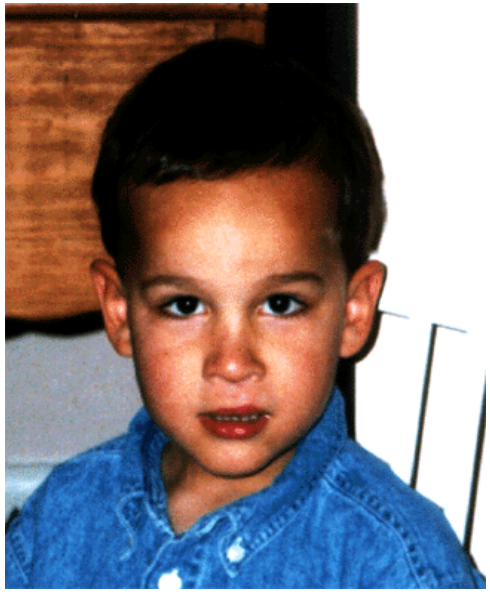


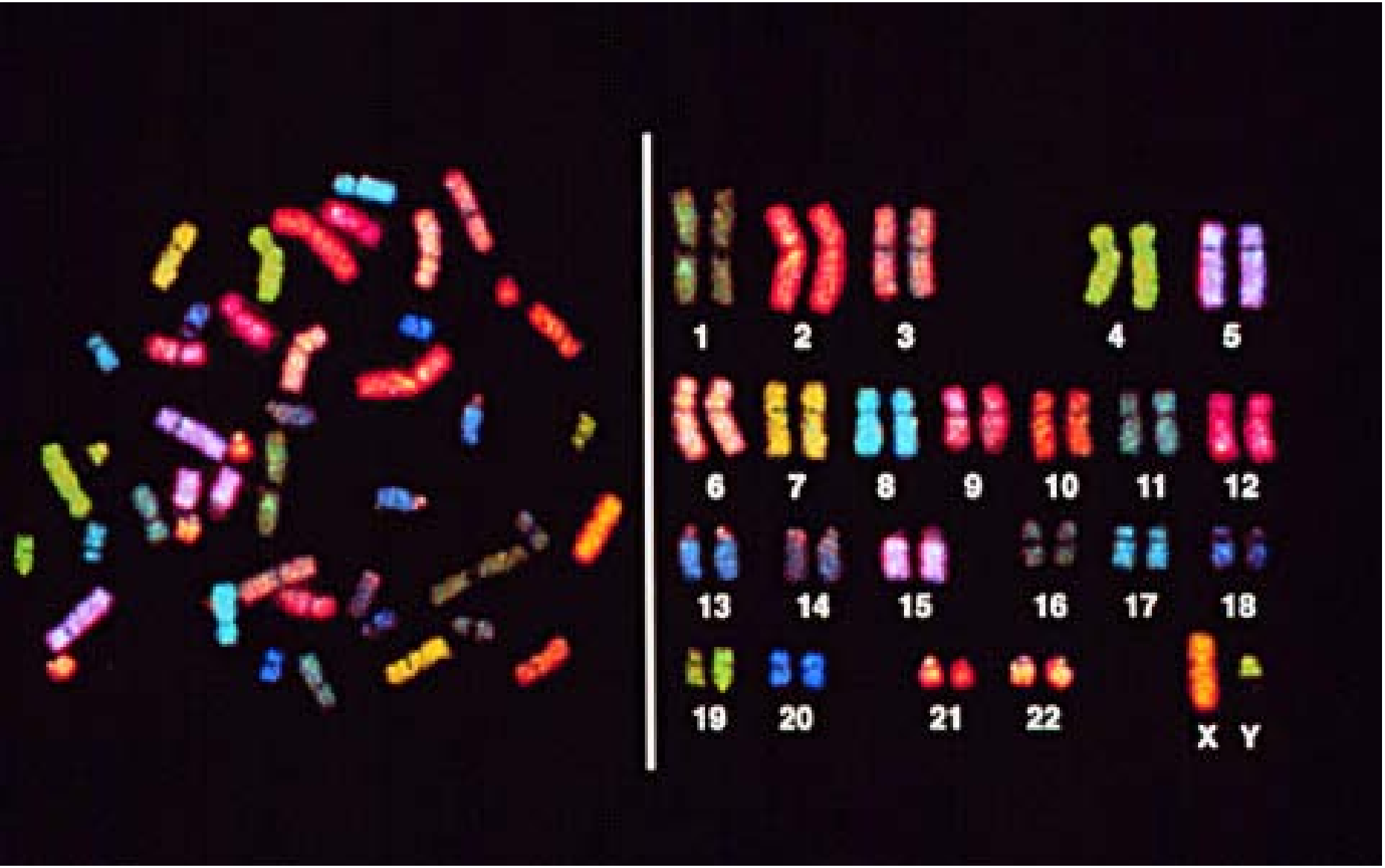




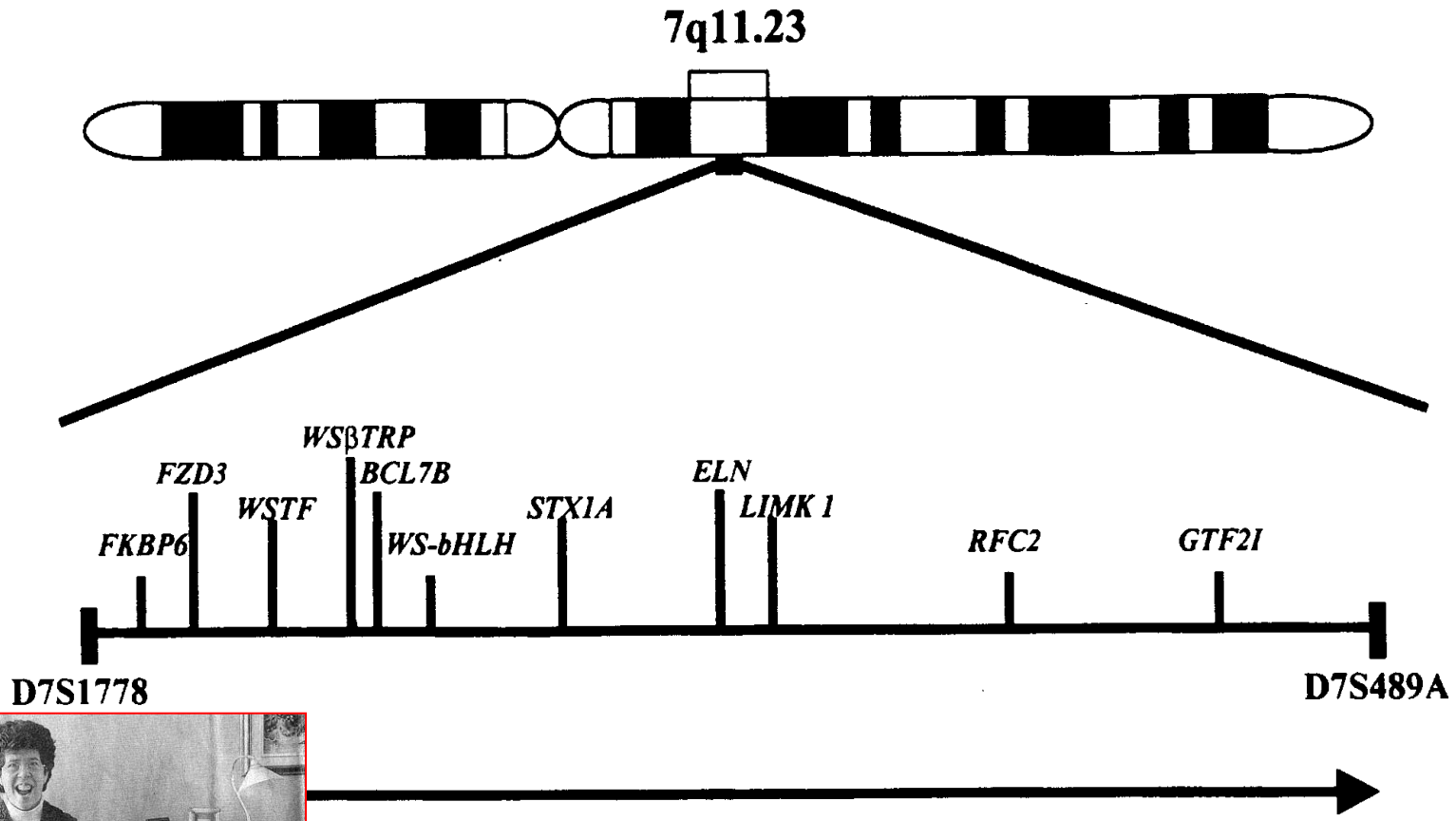




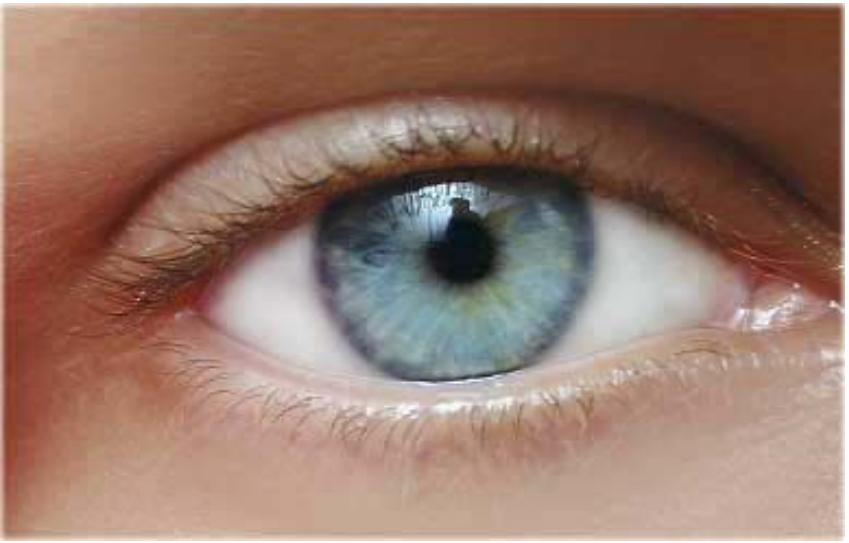


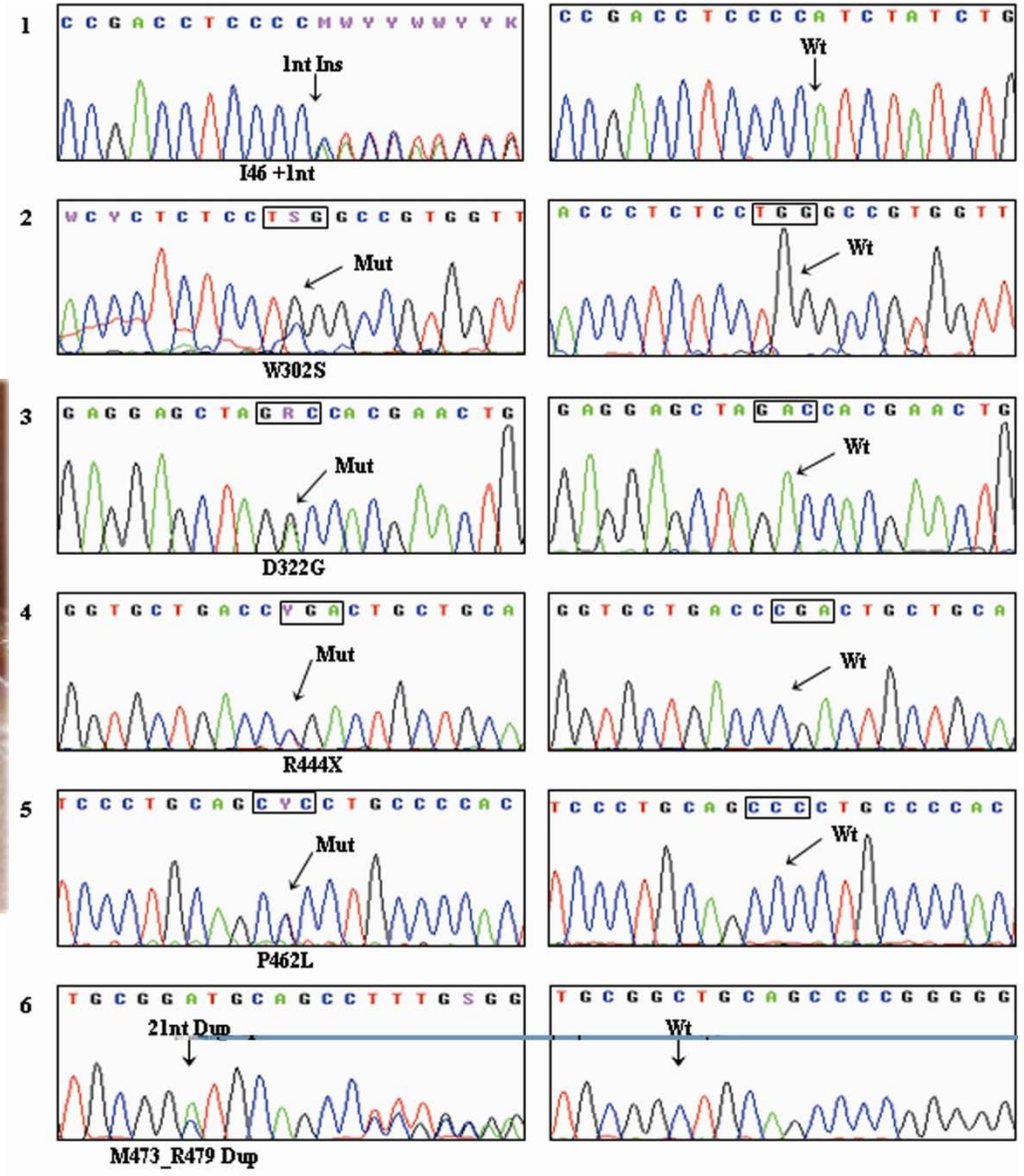

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y

Williams Deletion Region

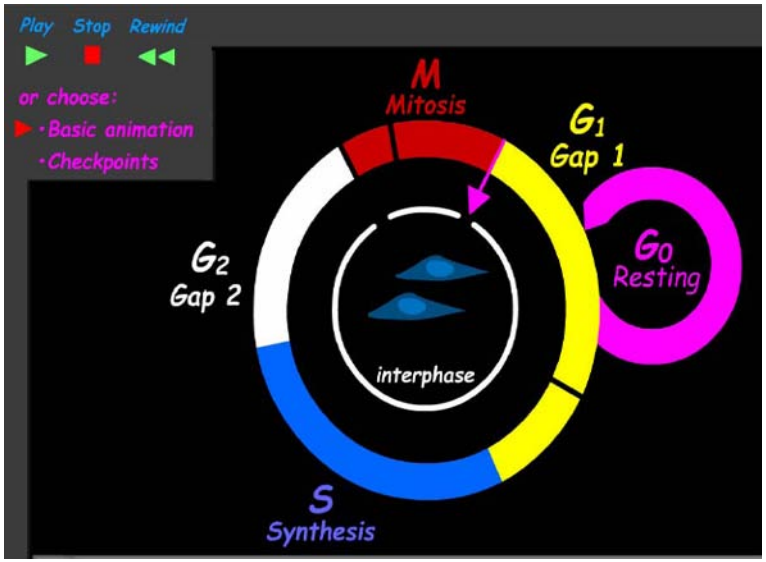


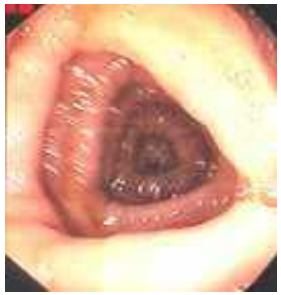
Retinitis pigmentosa

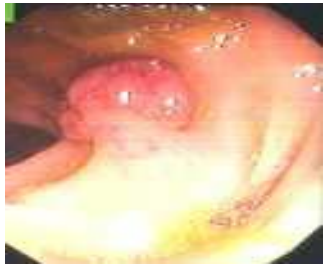


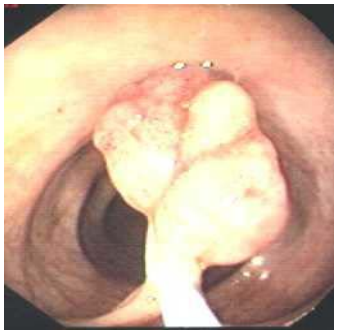


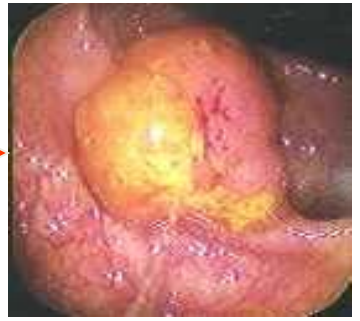
Genetics of sudden death in young people









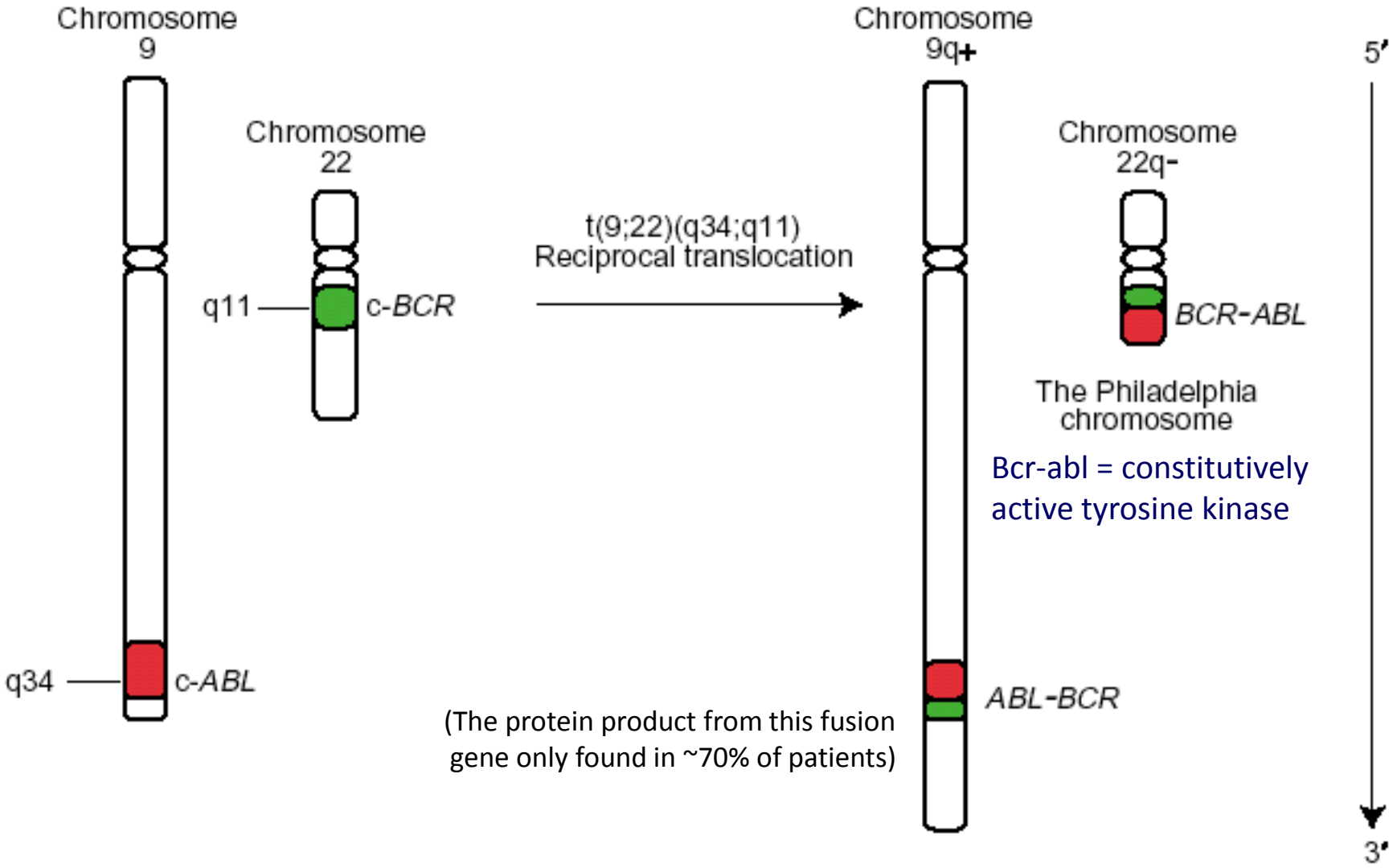


Colon normal

Adenoma

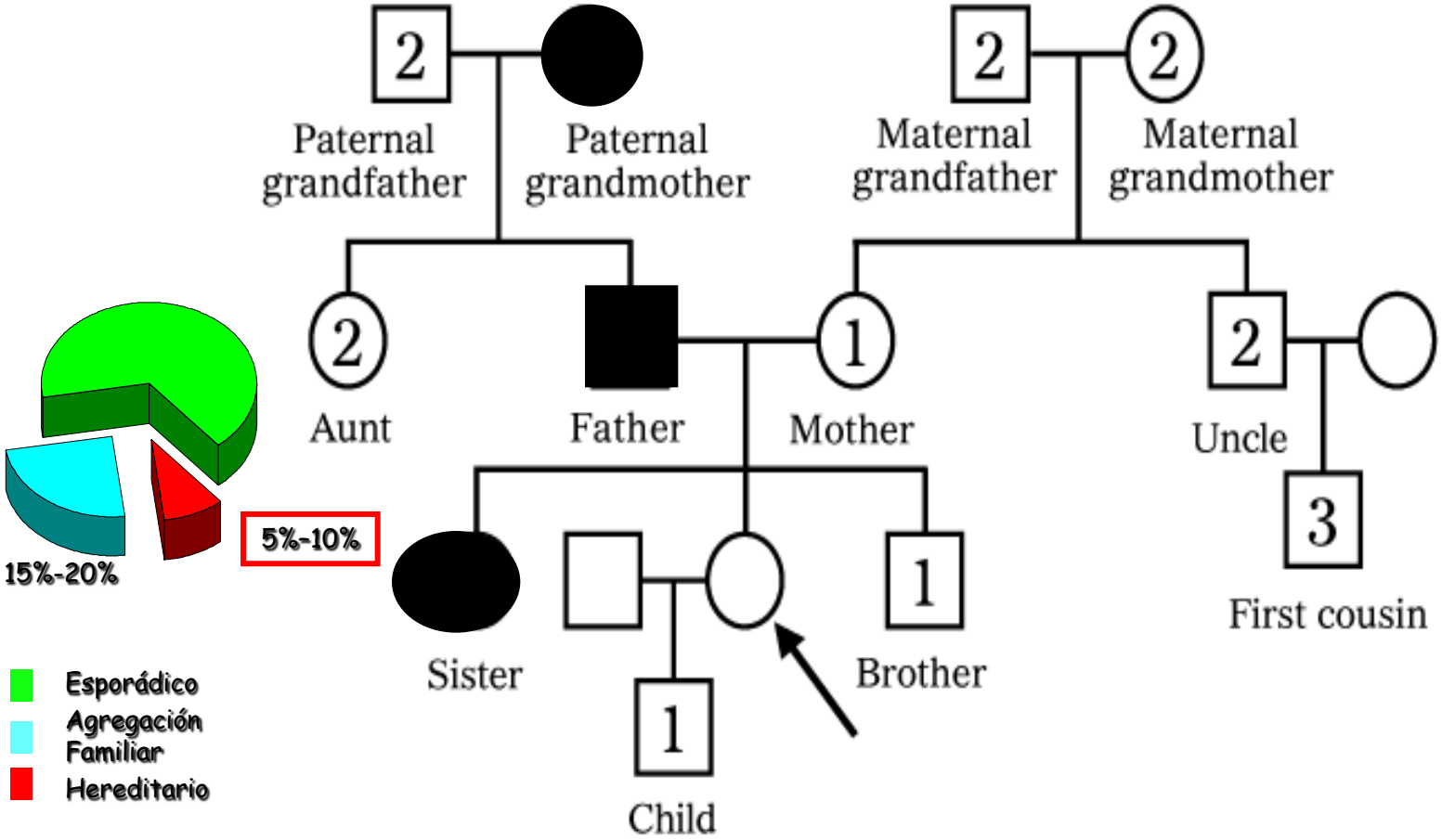
Adenoma avanzado

Cancer

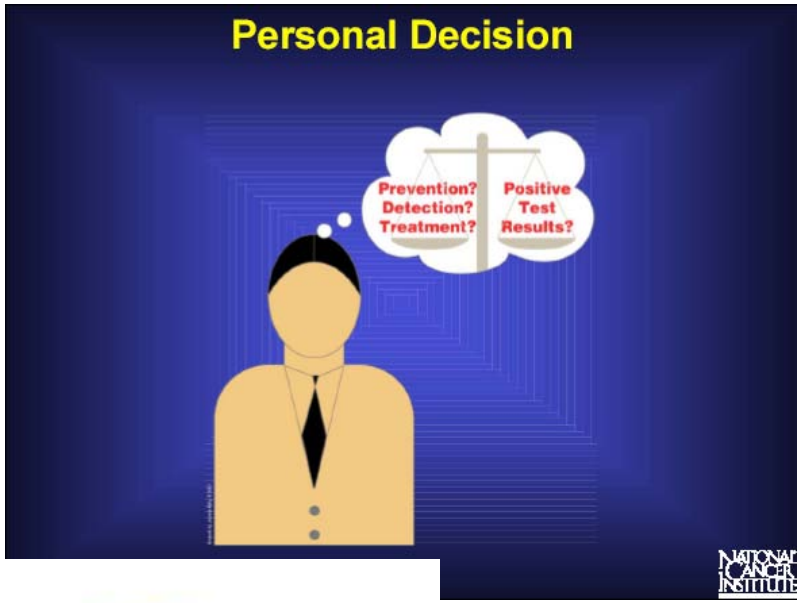


Chronic myeloid leukaemia (CML) is characterised by the t(9;22)(q34;q11) reciprocal translocation

CÁNCER HEREDITARIO



Consejo Genético

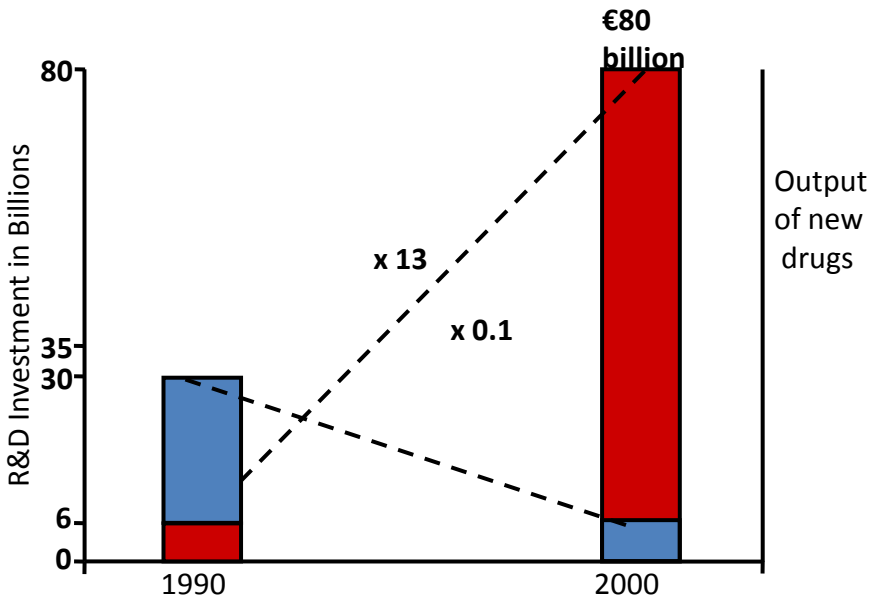


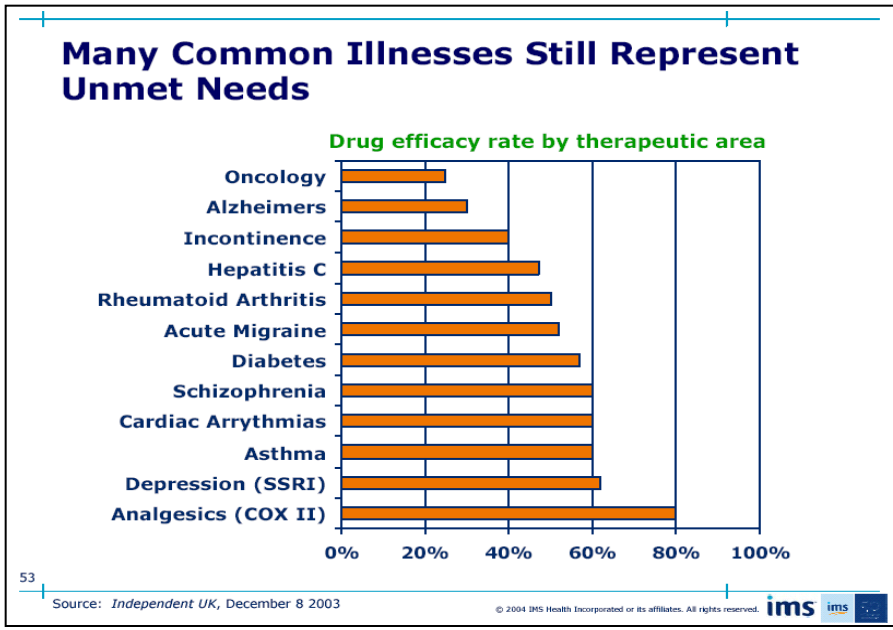
THE HUMAN VARIOME PROJECT

Our knowledge of the etiopathogenesis of the disease is limited

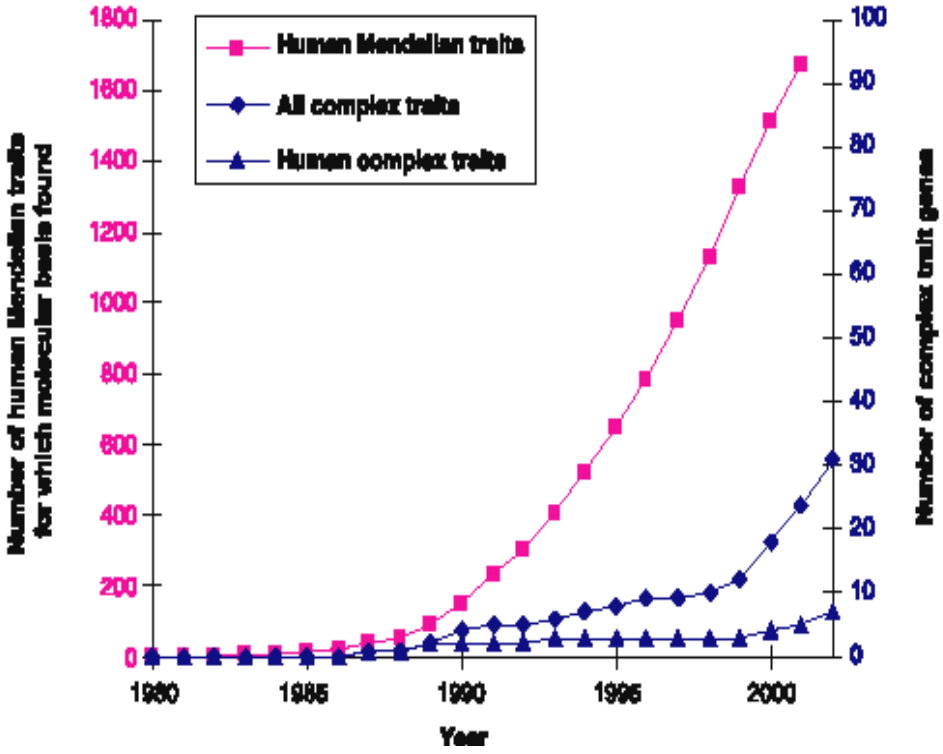
Our classification of diseases is mainly based in signs and symptoms

Limited therapeutic efficacy in many TAs



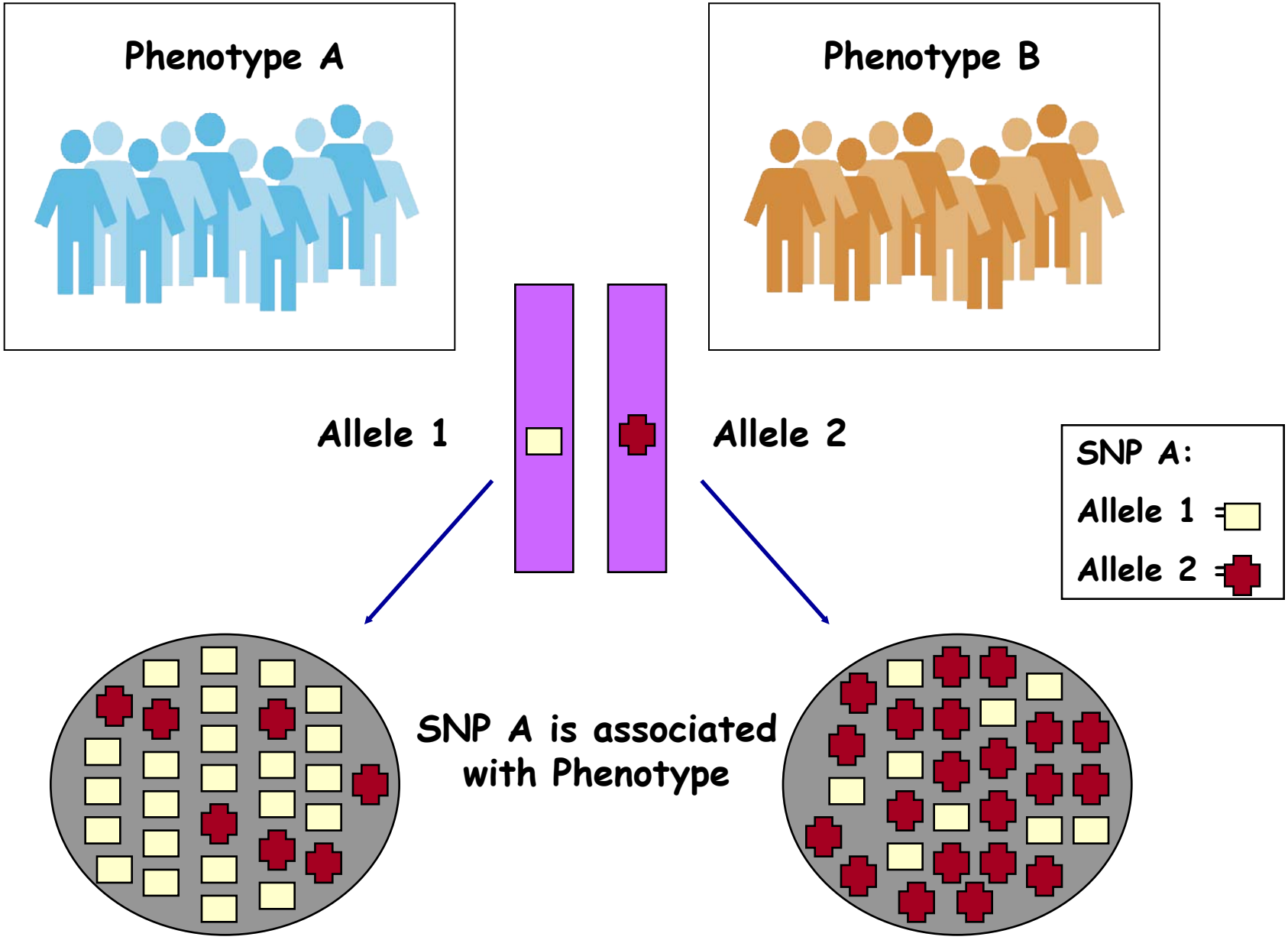


Using genetics to find genes that underlie complex traits (such as cancer) is a potential useful tool for a better understanding of the disease and pharmacogenetics and pharmacogenomics

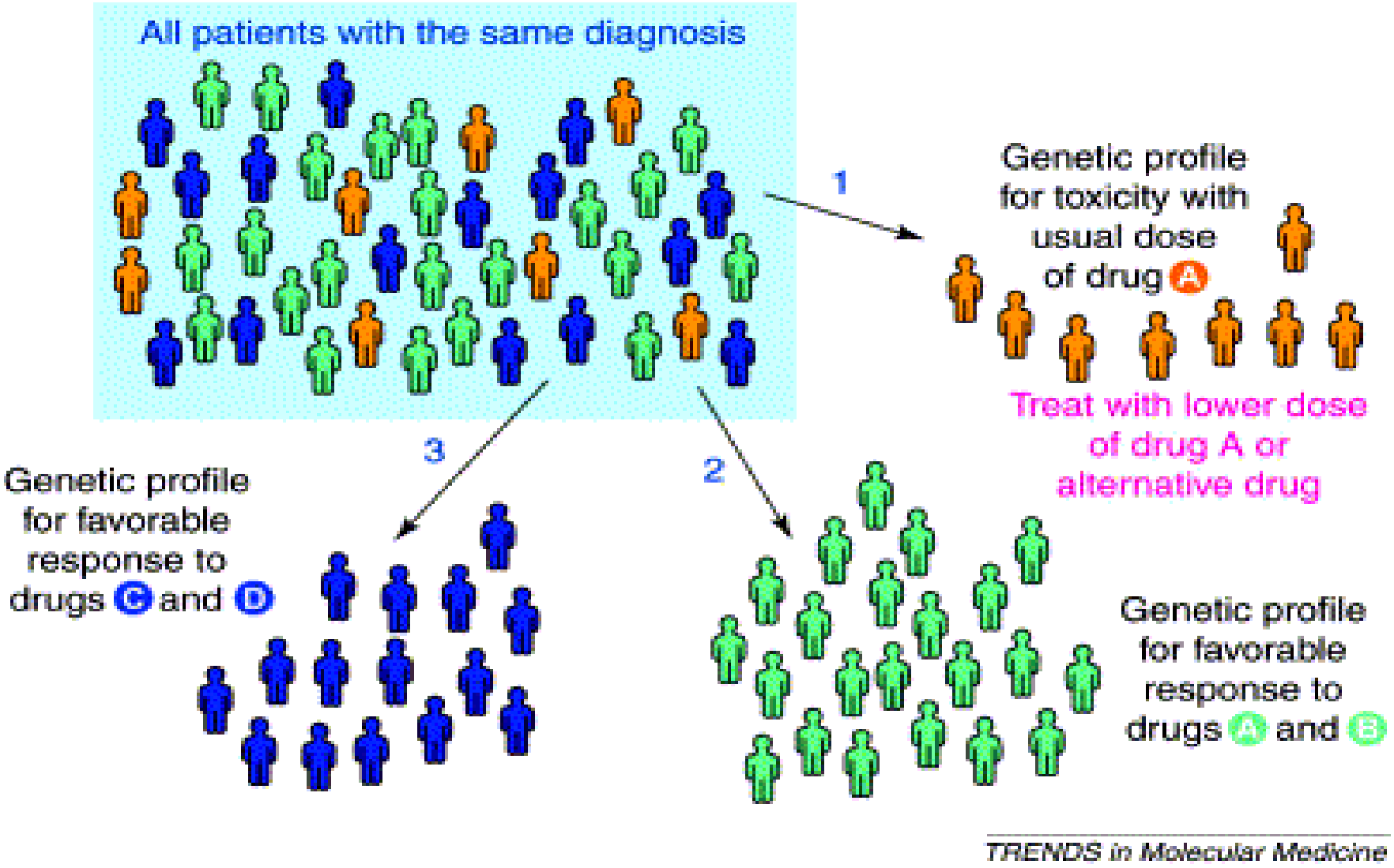


But still the advance in knowledge ir genes involved in complex traits is limited !

Human Genetic Association Study Design



Genotyping means patient classification



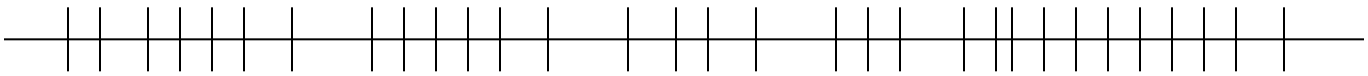
SNP: SINGLE NUCLEOTIDE POLYMORPHISM

ATCGGCGTACCTGATTCCGAATCCGTATCG
ATCGGCGTACCTGAATCCGAATCCGTATCG

Distribution: 1 every 250bp

>10 million SNPs in the human genome (around 6M validated)->11M in dbDNP

Mutation rate: 10^{-7}

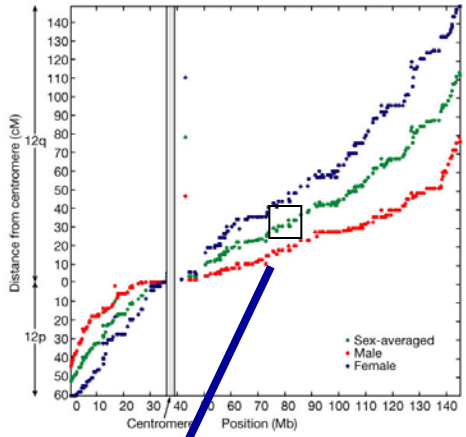


Human DNA

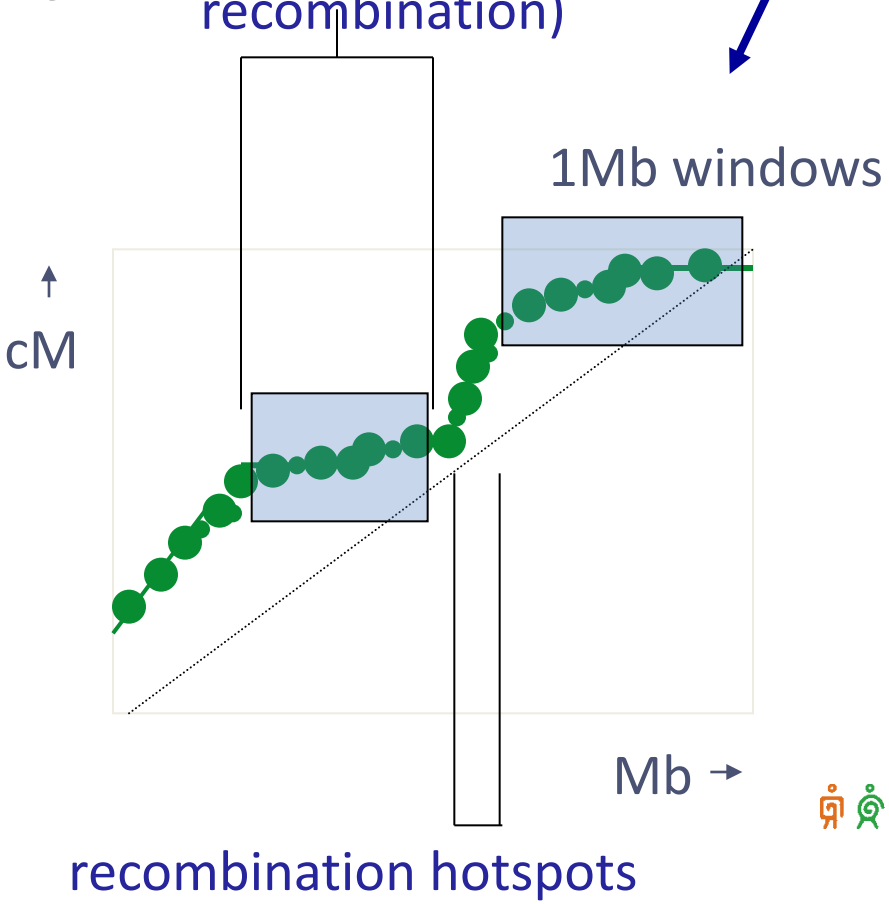
A SNP MAP

Characteristics of SNP Variation

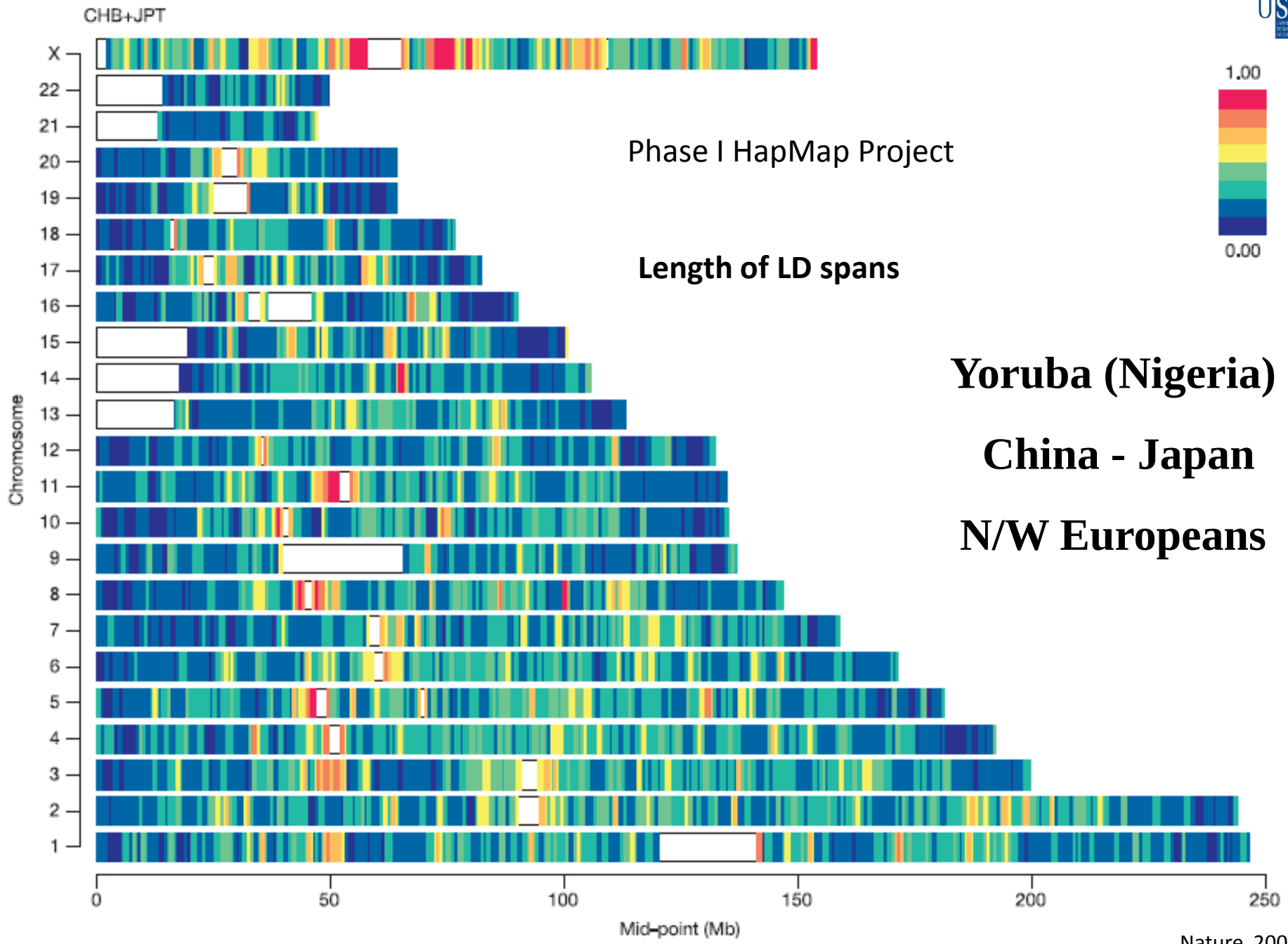
- Clustering is observed on all the autosomes: Haplotype blocks: Blocks with little evidence of recombination
- Some clusters appear functional : MHC on chromosome 6 (with extensive replication)



LD blocks (little or no recombination)



Gabriel et al. Science, 296,2002


CHB+JPT
X
22
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
Chromosome
0
50
100
150
200
250
Mid-point (Mb)
1.00
0.00
Phase I HapMap Project
Length of LD spans
Yoruba (Nigeria)
China - Japan
N/W Europeans

Nature, 2005

Spanish National Genotyping Center GeGen

Scientific International Committee

Ethical International Committeel

Coordination

NODE 1 Barcelona (CRG)

SNPlex / Illumina

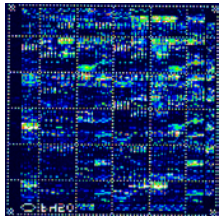
NODE 2 Santiago de Compostela (USC)

Sequenom / SNPlex/

Affymetrix

NODE 3 Madrid (CNIO)

Illumina




CeGen
TECAN
eedom evo
TECAN
te-mo

Spanish National Genotyping Center

Pre-genotyping

DNA extraction and quantification
Project design

SNP selection based on:
Function prediction
- Previous LD
- Known disease association

Genotyping

Genotyping platforms:

Designs

Genotyping

Data checking

Post-genotyping

- Data download
- Format changes
- Data analysis
- Customer support

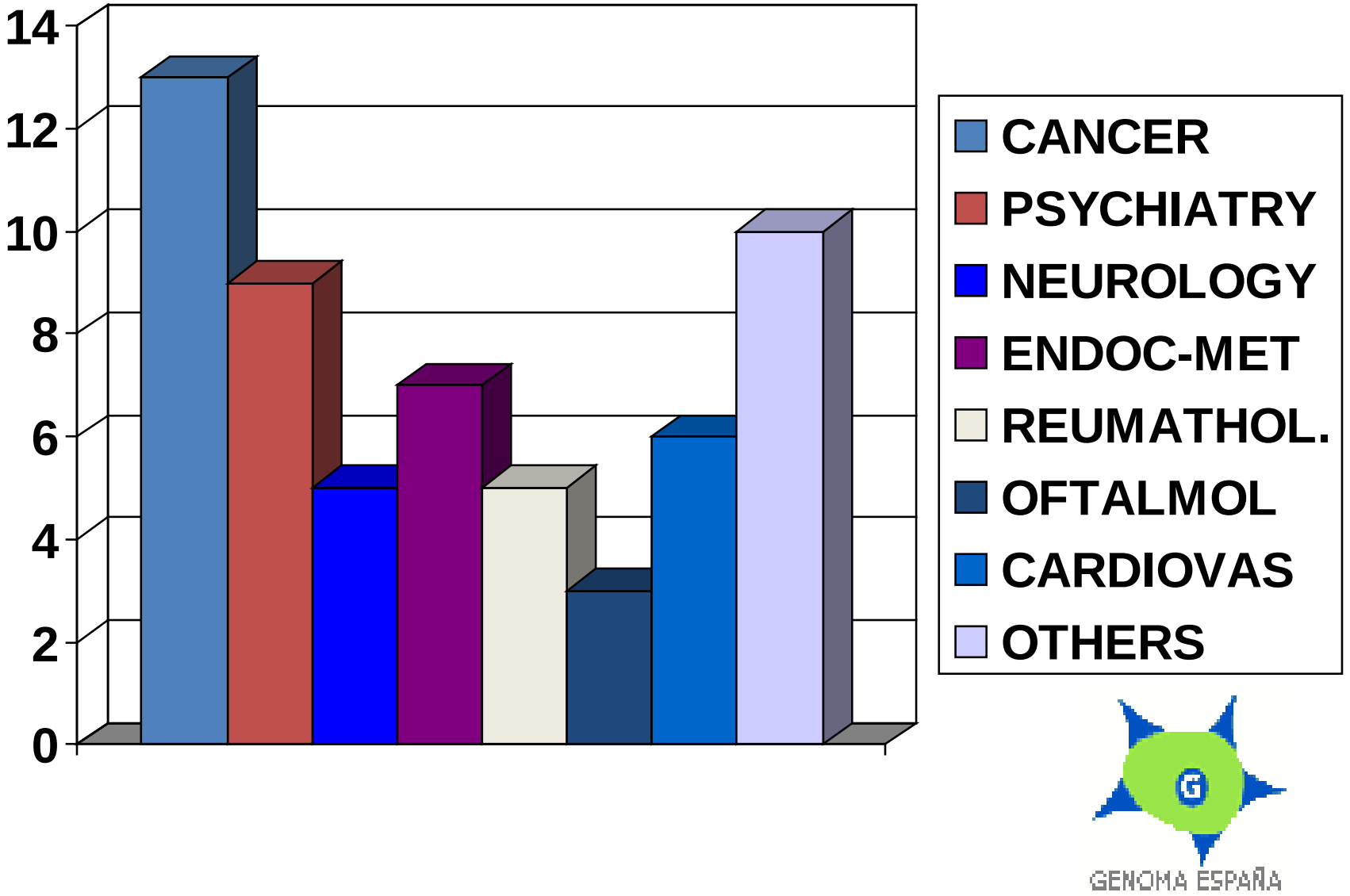
Researchers

Biólogos (Genética)
Médicos
Técnicos

Tecnólogos
(Biólogos-
Biotecnólogos)
Técnicos
Bioinformáticos

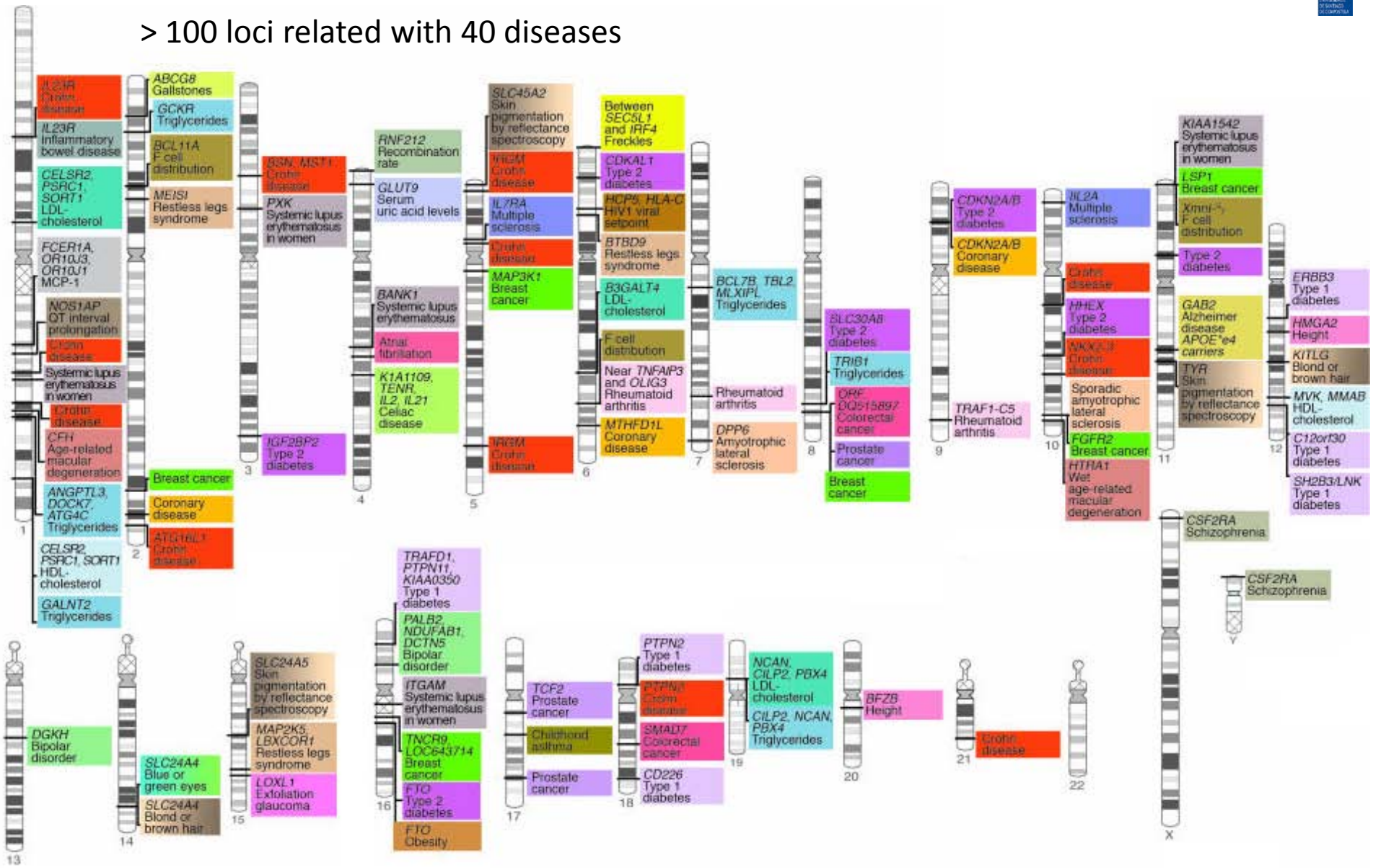
Biólogos (Genética)
Matemáticos
Bioinformáticos
Médicos

I+D: Biológos, Químicos, Físicos (Nanotecnólogos), Bioinformáticos


14
12
10
8
6
4
2
0
CANCER
PSYCHIATRY
NEUROLOGY
ENDOC-MET
REUMATHOL.
OFTALMOL
CARDIOVAS
OTHERS

GWAs preliminary results

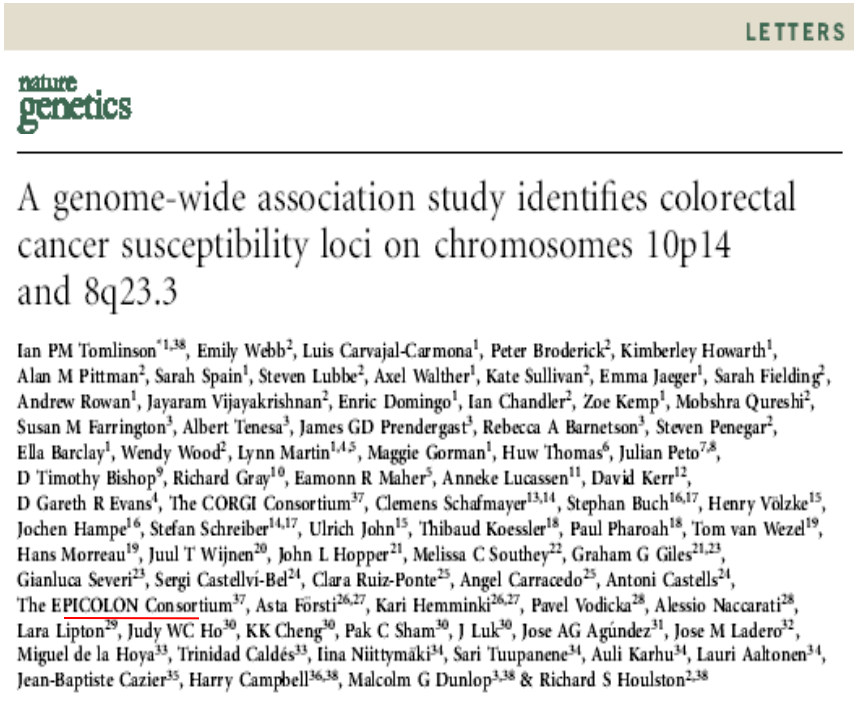
> 100 loci related with 40 diseases



J. Clin. Invest. **118**:1590–1605 (2008)

RESULTS GWAS CRC
Tomlinson et al. 2008 Nature Genetics

Locus	gen	MAF	OR alelico	P-val
8q24	Ninguno/MYC/ POU5F1p1	0.48	1.35 (1.20-1.53)	$1.3*10^{-14}$
18q21	SMAD7	0.47	0.84 (0.75-0.94)	$1.0*10^{-12}$
15q	GREM1	0.19	1.17 (1.06-1.30)	$4.4*10^{-14}$
8q23	EIF3H	0.07	1.27 (1.20-1.34)	$3.3*10^{-18}$
10p	Ninguno	0.33	0.87 (0.83-0.91)	$2.5*10^{-12}$
11q23	POU52AF1		1.11 (1.08-1.15)	$7.7*10^{-28}$



LETTERS

nature genetics

A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3

Ian PM Tomlinson[1,38], Emily Webb[2], Luis Carvajal-Carmona[1], Peter Broderick[2], Kimberley Howarth[1], Alan M Pittman[2], Sarah Spain[1], Steven Lubbe[2], Axel Walther[1], Kate Sullivan[2], Emma Jaeger[1], Sarah Fielding[2], Andrew Rowan[1], Jayaram Vijayakrishnan[2], Enric Domingo[1], Ian Chandler[2], Zoe Kemp[1], Mobshra Qureshi[2], Susan M Farrington[3], Albert Tenesa[3], James GD Prendergast[3], Rebecca A Barnetson[3], Steven Penegar[2], Ella Barclay[1], Wendy Wood[2], Lynn Martin[1,4,5], Maggie Gorman[1], Huw Thomas[6], Julian Peto[7,8], D Timothy Bishop[9], Richard Gray[10], Eamonn R Maher[5], Anneke Lucassen[11], David Kerr[12], D Gareth R Evans[4], The CORGI Consortium[37], Clemens Schafmayer[13,14], Stephan Buch[16,17], Henry Völzke[15], Jochen Hampe[16], Stefan Schreiber[14,17], Ulrich John[15], Thibaud Koessler[18], Paul Pharoah[18], Tom van Wezel[19], Hans Morreau[19], Juul T Wijnen[20], John L Hopper[21], Melissa C Southey[22], Graham G Giles[21,23], Gianluca Severi[23], Sergi Castellví-Bel[24], Clara Ruiz-Ponte[25], Angel Carracedo[25], Antoni Castells[24], The EPICOLON Consortium[37], Asta Försti[26,27], Kari Hemminki[26,27], Pavel Vodicka[28], Alessio Naccarati[28], Lara Lipton[29], Judy WC Ho[30], KK Cheng[30], Pak C Sham[30], J Luk[30], Jose AG Agúndez[31], Jose M Ladero[32], Miguel de la Hoya[33], Trinidad Caldés[33], Iina Niittymäki[34], Sari Tuupanene[34], Auli Karhu[34], Lauri Aaltonen[34], Jean-Baptiste Cazier[35], Harry Campbell[36,38], Malcolm G Dunlop[3,38] & Richard S Houlston[2,38]

Common variants conferring risk of schizophrenia

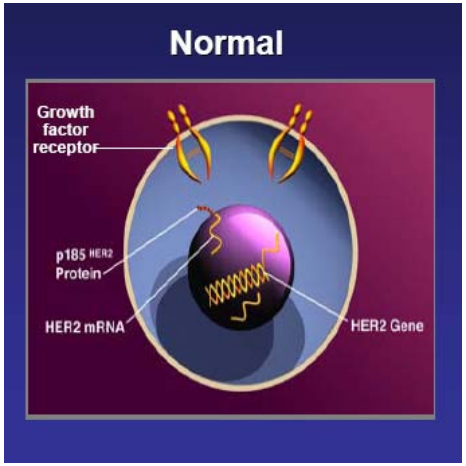
Nature 1 de julio

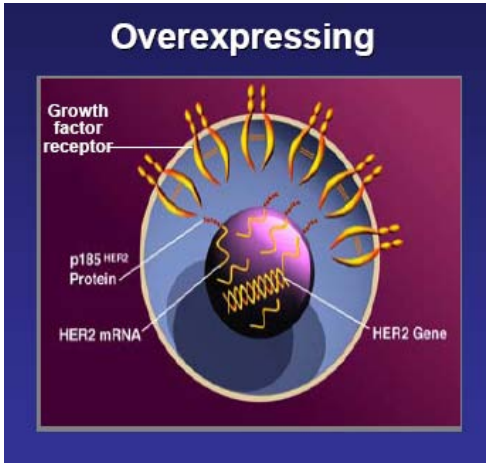
Genome-wide significant association of seven markers with schizophrenia

ne/ SNP[allele]	Frequency	SGENE-plus* (2,663 cases; 13,498 controls)		Follow-up (4,999 cases; 15,555 controls)		SGENE-plus + follow-up (7,662 cases; 29,053 controls)		SGENE-plus + follow-up + ISC + MGS (12,951 cases; 34,594 controls)		Region/ neighbouring gene
		OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	
rs6913660[C]†☆	0.85	1.22 (1.10, 1.36)	0.00023	1.11 (1.04, 1.19)	0.0021	1.14 (1.08, 1.21)	4.7×10^{-6}	1.15 (1.10, 1.21)	1.1×10^{-9}	MHC/ *HIST1H2BJ*
rs13219354[T]‡☆	0.90	1.25 (1.11, 1.42)	0.00043	1.19 (1.08, 1.30)	0.00022	1.21 (1.12, 1.30)	4.4×10^{-7}	1.20 (1.14, 1.27)	1.3×10^{-10}	MHC/ *PRSS16*
rs6932590[T]§☆	0.78	1.15 (1.05, 1.26)	0.0024	1.17 (1.10, 1.25)	4.9×10^{-7}	1.17 (1.11, 1.23)	4.4×10^{-9}	1.16 (1.11, 1.21)	1.4×10^{-12}	MHC/ *PRSS16*
rs13211507[T]‖☆	0.92	1.24 (1.08, 1.42)	0.0027	1.27 (1.15, 1.40)	3.1×10^{-6}	1.26 (1.16, 1.36)	3.1×10^{-8}	1.24 (1.16, 1.32)	8.3×10^{-11}	MHC/ *PGBD1*
rs3131296[G]¶☆	0.87	1.21 (1.08, 1.36)	0.0011	1.20 (1.11, 1.30)	5.3×10^{-6}	1.21 (1.13, 1.29)	2.1×10^{-8}	1.19 (1.13, 1.25)	2.3×10^{-10}	MHC/ *NOTCH4*
rs12807809[T]	0.83	1.19 (1.08, 1.32)	0.00045	1.13 (1.06, 1.21)	0.00022	1.15 (1.09, 1.22)	5.0×10^{-7}	1.15 (1.10, 1.20)	2.4×10^{-9}	*NRGN*
rs9960767[C]#☆	0.056	1.30 (1.11, 1.51)	0.0011	1.20 (1.08, 1.33)	0.00044	1.23 (1.13, 1.34)	2.2×10^{-6}	1.23 (1.15, 1.32)	4.1×10^{-9}	*TCF4*

4999 cases and 15,555 controls from Denmark (Aarhus), Denmark (Copenhagen), Germany (Bonn), Germany (Munich), Hungary, the Netherlands, Norway, Russia, Sweden Finland; Spain (Santiago) and Spain (Valencia))

Farmacogenética

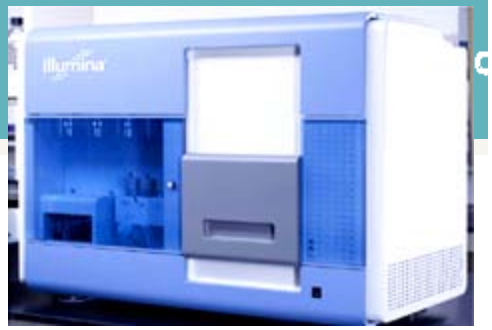




- ***Safety***
  - **• TPMT (6-MP, azathioprine)**
  - **• UGT1A1 (irinotecan)**
  - **• CYP2C9/VKORC1 (warfarin)**
  - **• CYP2D6 (Strattera)**
  - **• HLAB*5701 (Abacavir)**
- ***Efficacy***
  - **• EGFR status (Erbitux, Tarceva)**
  - **• Her2/neu status (Herceptin)**
  - **• Philadelphia chromosome ~ Bcr-abl (Gleevec)**
  - **• C-kit (Gleevec)**
  - **• K ras mutation (Cetuximab)**

Evolución del coste de secuenciación

- Human Genome Project:
  - Sanger sequencing
  - 13 years
  - > 2800 scientists
  - 2.7 billion $
  - (> 98%) finished in 2003
- Genome of James Watson:
  - GS-FLX
  - 4.5 months
  - 27 scientists
  - < $1.5 million $
  - finished in 2007: Nature (2008) 452:872.



2008: JCVI. 50 genomas a $300.000 cada uno

2008: Solexa. $150.000

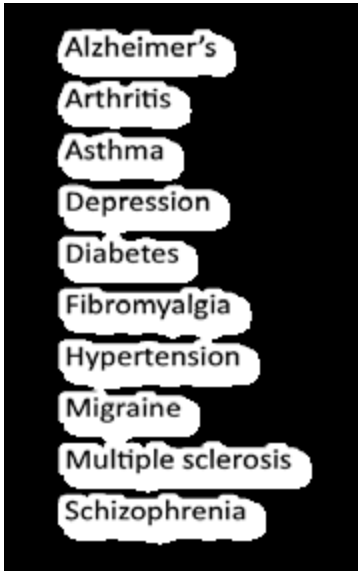
2012: BioNanomatrix y Complete Genomics. $1000

gEUVADIS Promotors
– ES: Xavier Estivill
– UK: Leena Peltonen and Richard Durbin
– NL: Joris Veltman and Han Brunner
– CH: Stylianos Antonarakis
Members
– FR: Mark Lathrop and Arnold Munich
– DE: Stefan Schreiber and Thomas Meitinger
– SE: Ann-Christine Syvänen
– UK: Ewan Birney and Peter Donnelly
– ES: Roderic Guigó, Cedric Notredame and Angel Carracedo
– NL: Gert Jan van Ommen
– AU: Kurt Zatloukal
– European Commission: Manuel Hallen and Jacques Remacle

gEUVADIS

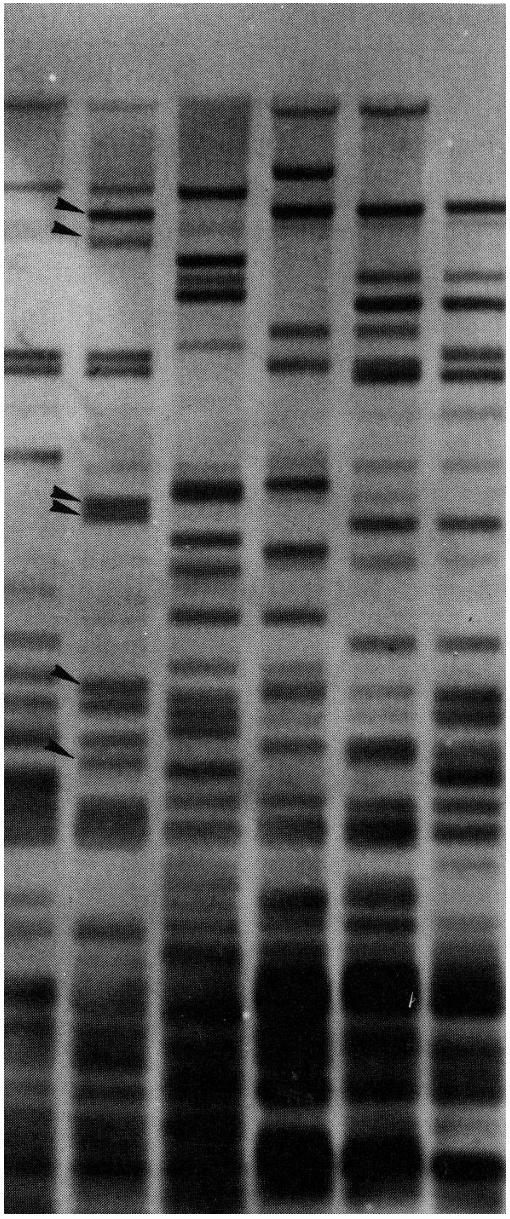
- **Phase I (2009-2011):**
  - **Sequence 1,000 genomes for each of 10 common disorders**
  - **Research teams as core of the analysis of each disorder**
  - **Research funds for each disorder could come from several sources**

- **Phase II (2012-2014):**
  - **Sequence 1,000 genomes for each of 50 disorders**
  - **Deep sequencing (single molecule sequencing)**
  - **Digital molecular analysis of DNA and RNA**
  - **Phenotype Capture**



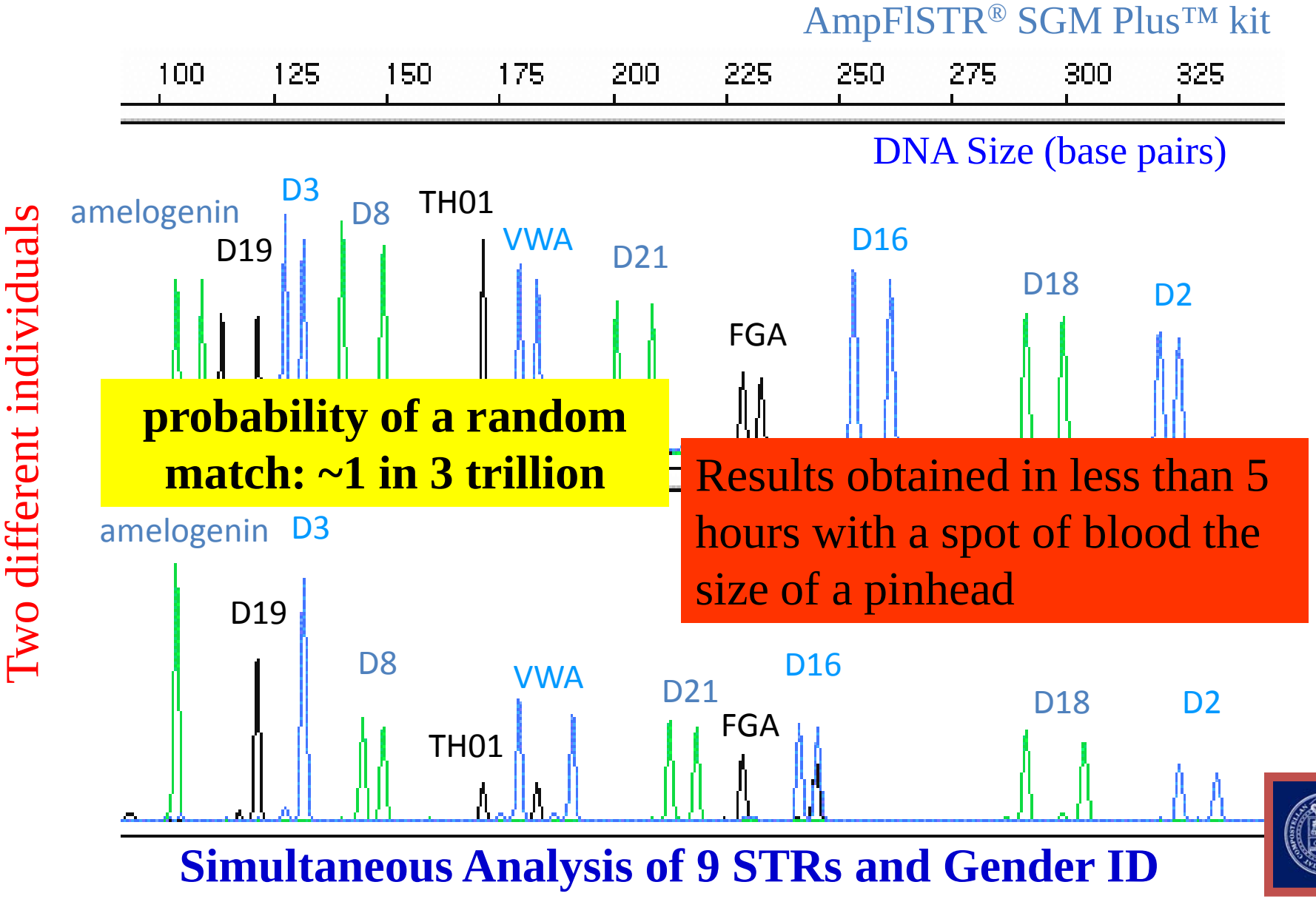
Genética de poblaciones y Genética forense

Instituto de Investigación Sanitaria
SANTIAGO DE COMPOSTELA

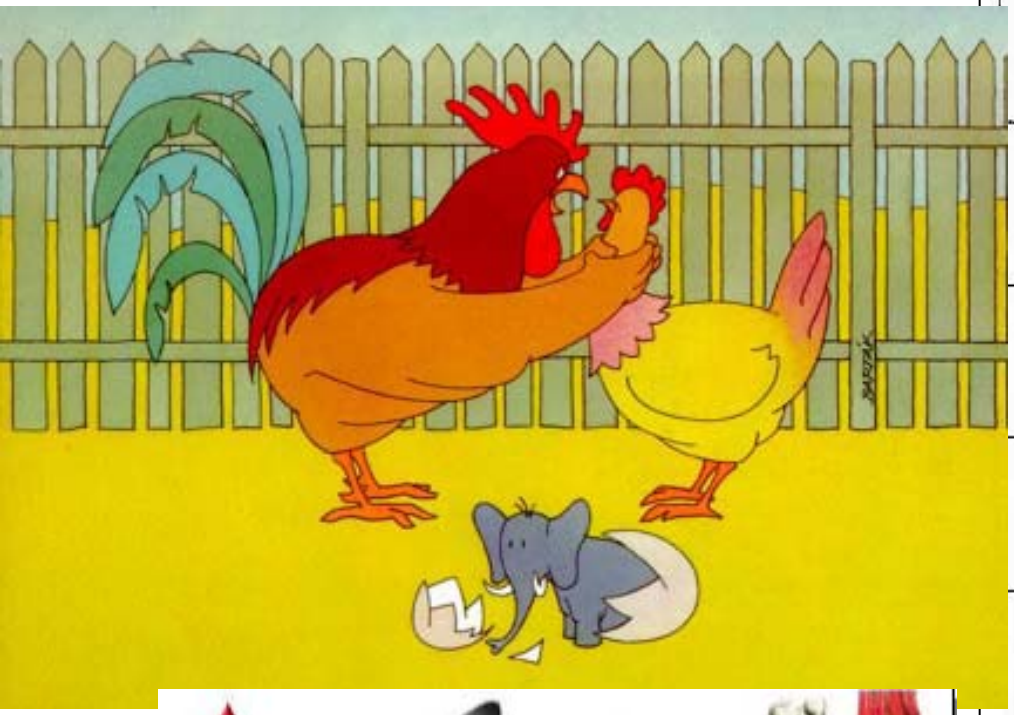


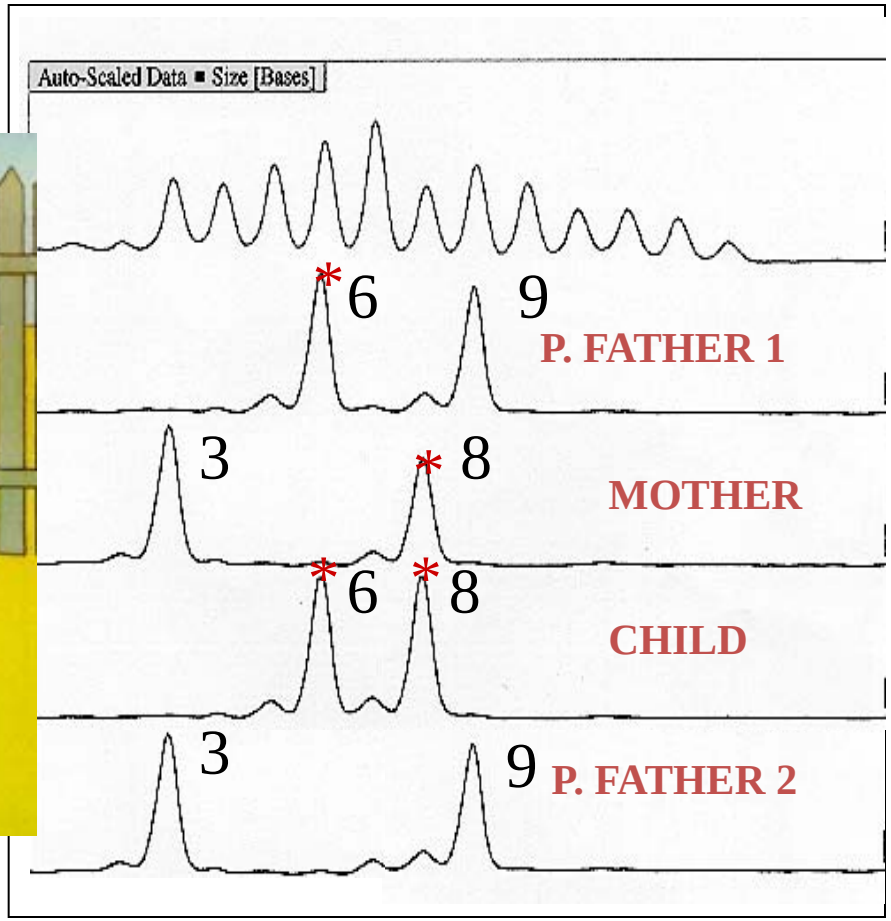
Alec Jeffreys DNA fingerprint 1985

Human Identity Testing with Multiplex STRs

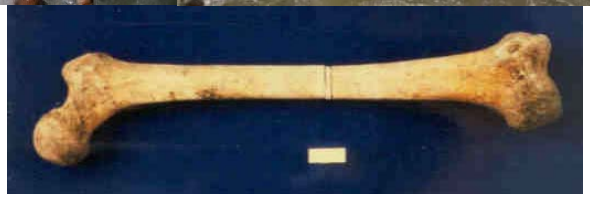


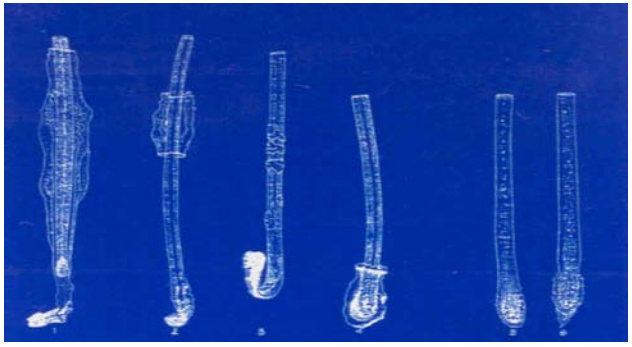
PATERNITY TESTING

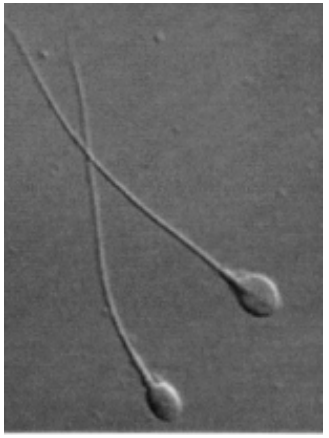


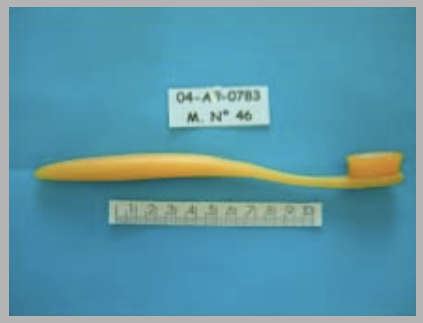



MBEIROS

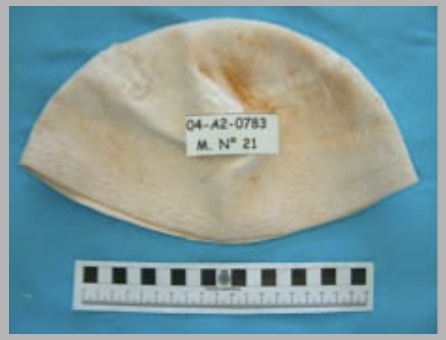






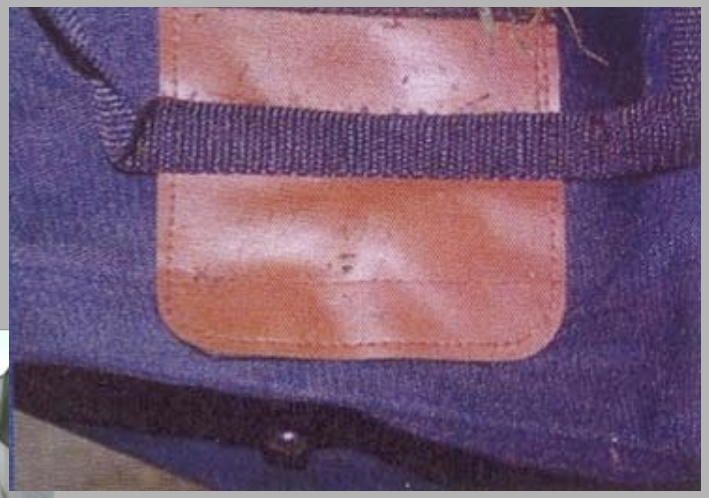


11-M was a 'closed' ancestry analysis



personal items from bomb assembly site in Liganés

un-detonated explosives in a holdall found at El Pozo station



Calculation results

Executing the query with 3 default populations and the 34 SNPs of the individual to classify :
GTCCCCCTAGAACTCCAACTGGGGTTCCAACCAACCAAGGCCAACCTTACAAGGTTCGAAGGTTACAC

The **-log(LIKELIHOOD)** (lower is best) and PERCENTILE (percent of population samples with lower likelihoods than individual submitted).

Gal-Dani	41.109466	2.50%
Moz-Som	66.935324	0.00%
Chi-Taiw	65.629561	0.00%

This person was European

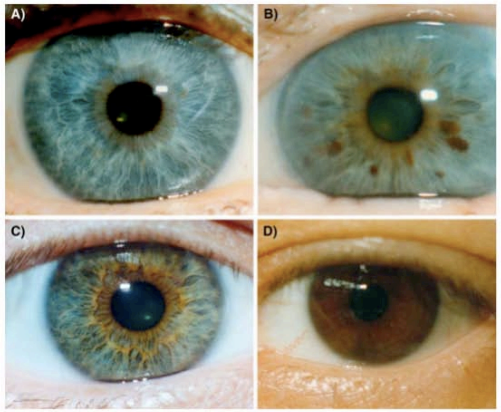
training set:	-log likelihood	percentile	exp	¨times more likely to be:	verbal predicate
Gal-Dani	41.109466	2.50%	1.40083E-18	EUR not AFR	
Moz-Som	66.935324	0.00%	8.5184E-30	1.64448E+11	*164 billion times more likely to be European than African*
Gal-Dani	41.109466	2.50%	1.40083E-18	EUR not ASN	
Chi-Taiw	65.629561	0.00%	3.14372E-29	44559668806	*44 billion times more likely to be European than Asian*
Moz-Som	66.935324	0.00%	8.5184E-30	AFR not ASN	
Chi-Taiw	65.629561	0.00%	3.14372E-29	0.270965709	

GT	CC	CC	CT	AG	AA	CT	CC	AA	CT	GG	GG	TT	CC	AA	CC	AA	CC	AA	GG	CC	AA	CC	TT	AC	TT	GG	TT	AA	CG	TT	GG	AC	AC

... and had blue eyes: rs12913832 = GG

GG

Blue



AG

Green (Unknown)

BEY1

Green Hazel

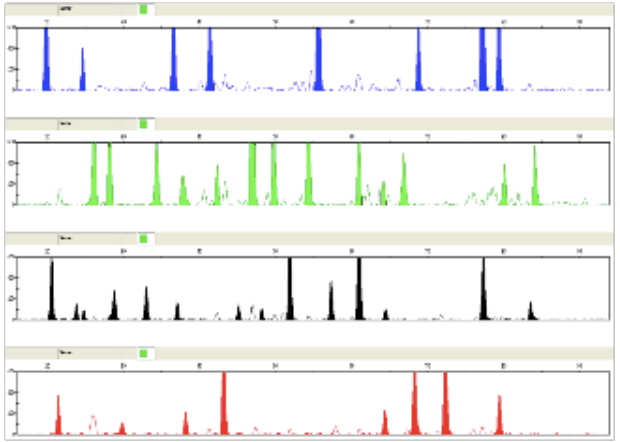
AG

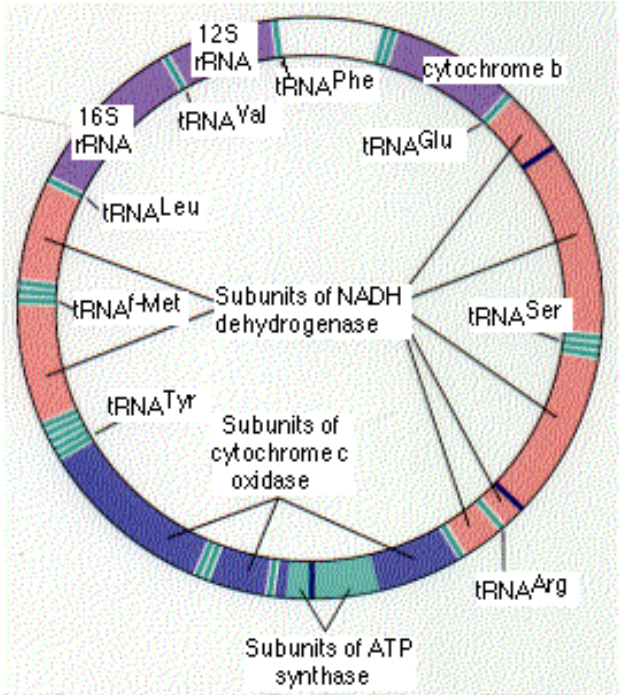
AG / (AA)

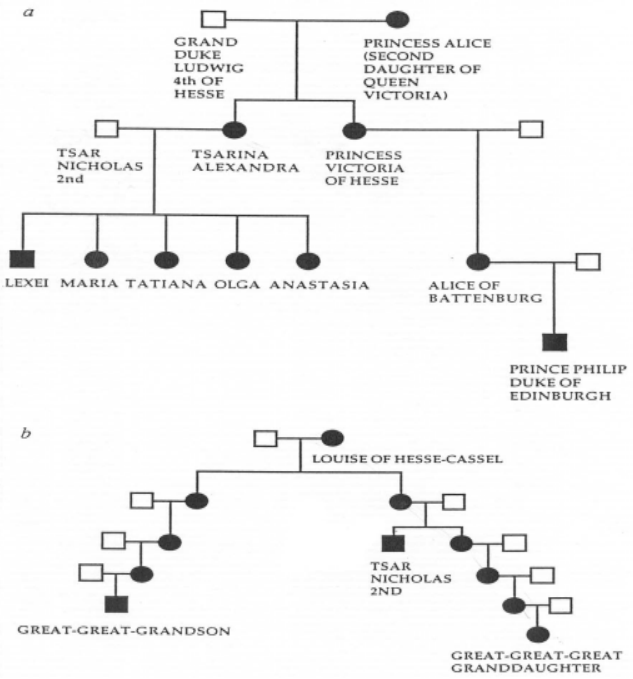
Brown

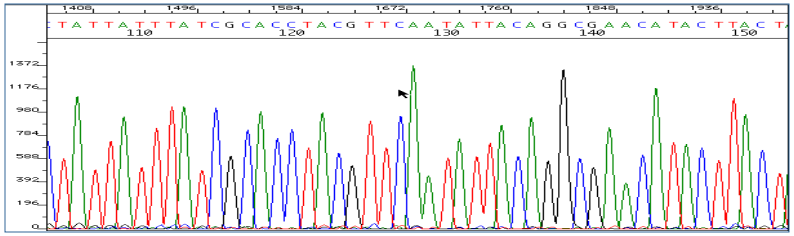
BEY2

34plex









Análisis de los restos de la familia Romanov

Gill et al. Nature Genetics, 1994

MUESTRAS ANALIZADAS:

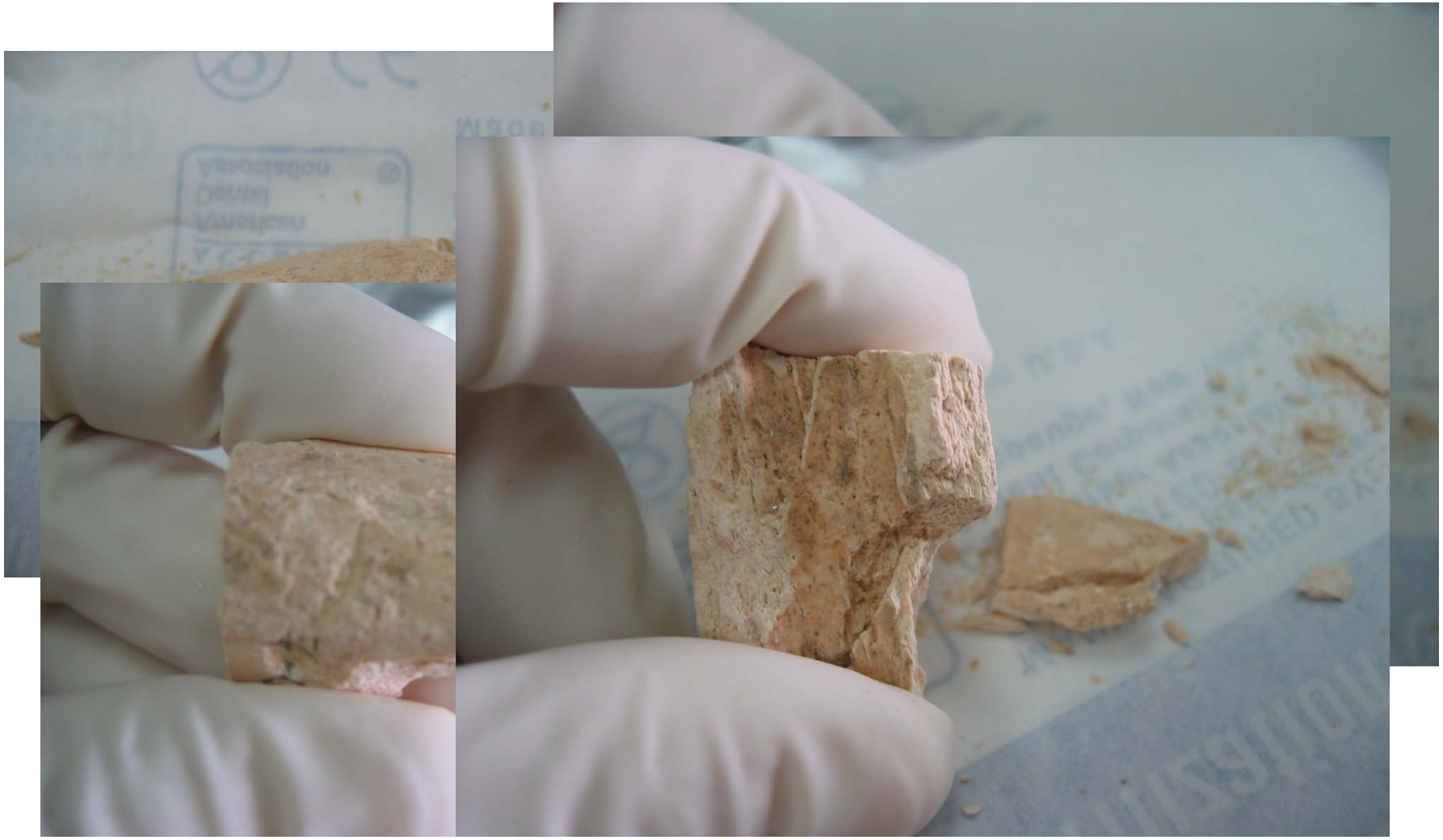
- Corazón
- Cabellos de María Antonieta, madre de Luis XVII
- Cabellos de Juana Gabriela y María Josefa, hijas de María Teresa de Austria, abuela del infante

EL "CORAZÓN DE PIEDRA" PERTENECE AL DEL HIJO DE MARÍA ANTONIETA

NADIE HA DEMOSTRADO QUE SEA EL HIJO DE LUIS XVI….

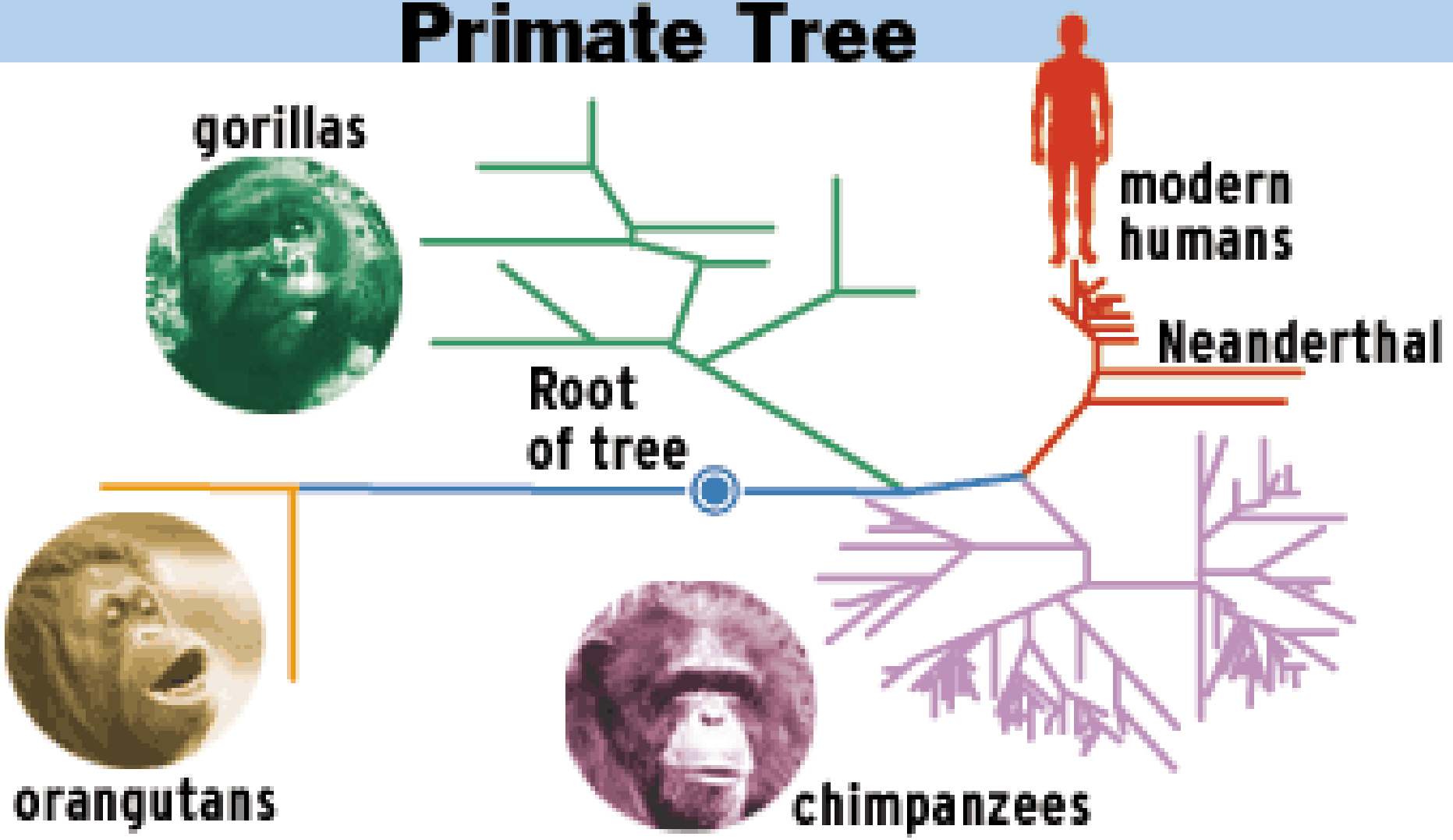
MUESTRA 3

¿CRISTÓBAL COLÓN?



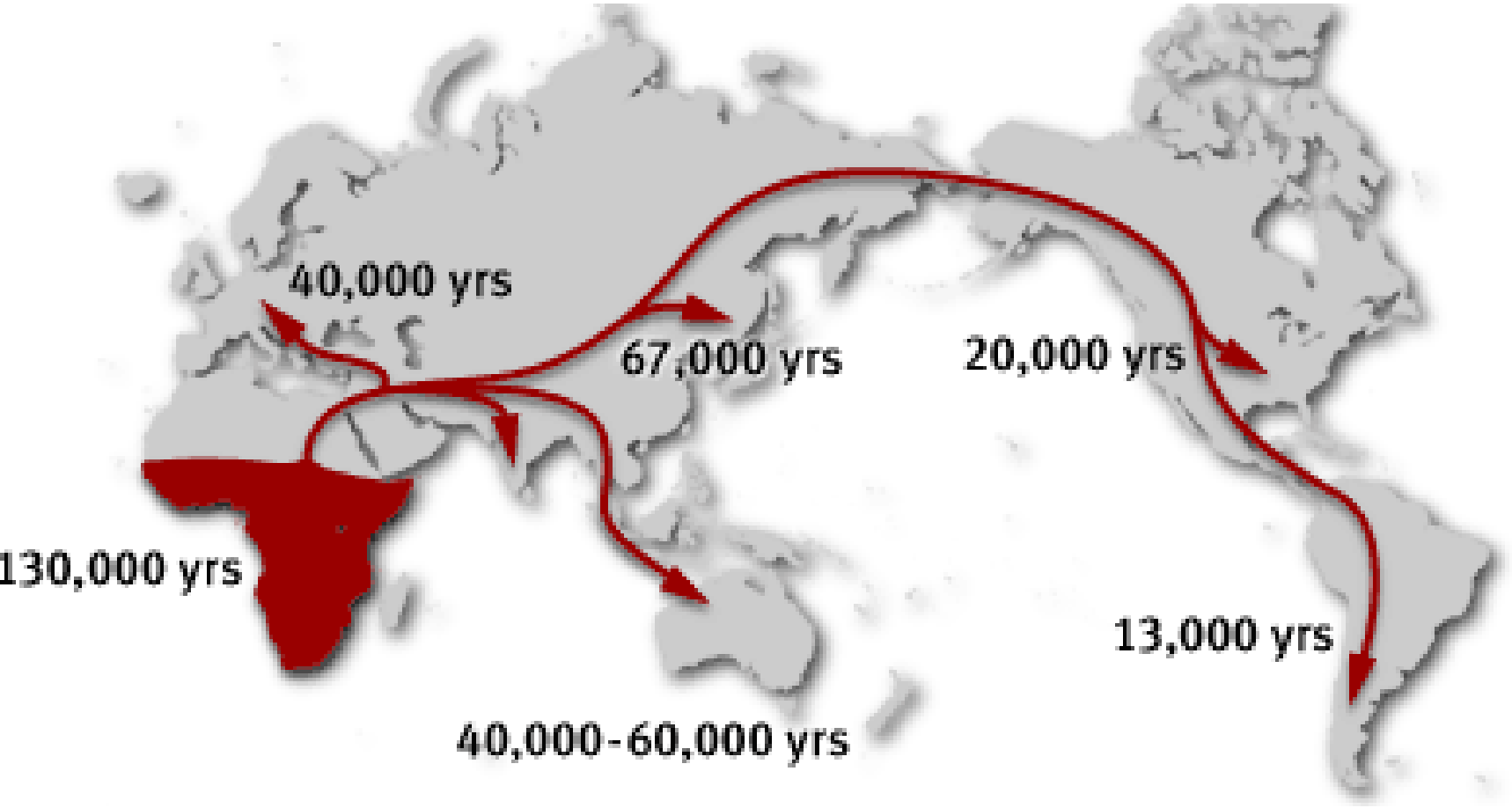
IML.- Proceso para extraer y amplificar muestras críticas

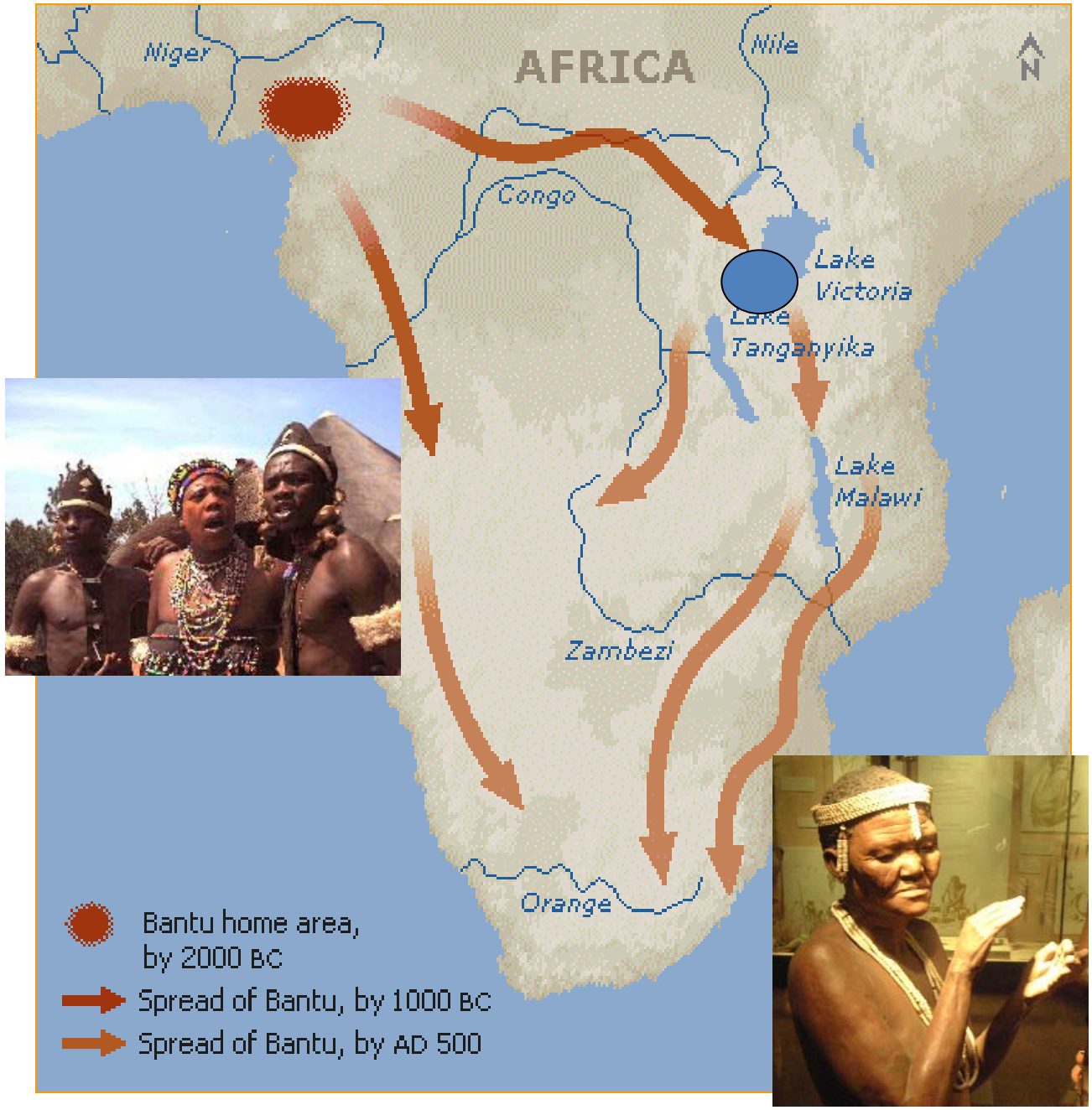
Primate Tree



Humano-neanderthal
Humano-humano
Humano-chimpancé
% de pares
Número de diferencias
0
5
10
15
20
25
30
35
40
45
50
55
60
65

“Out of Africa”





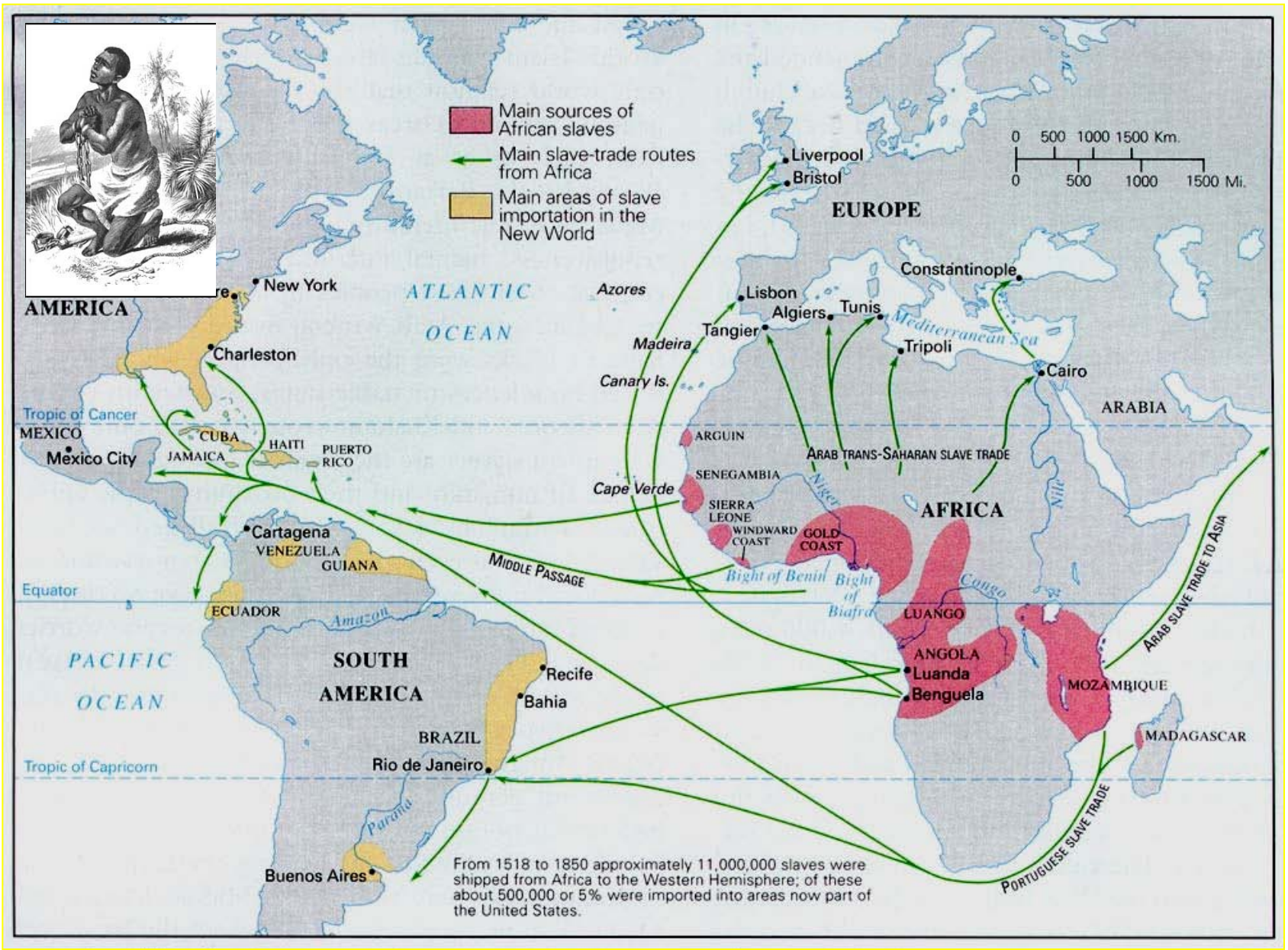
Bantu farmers adopted invented a type of grain that grows in winter as opposite to the Mediterrenean type of agriculture.

Later on they developped new technologies which would give them clear advantages during the expansion:

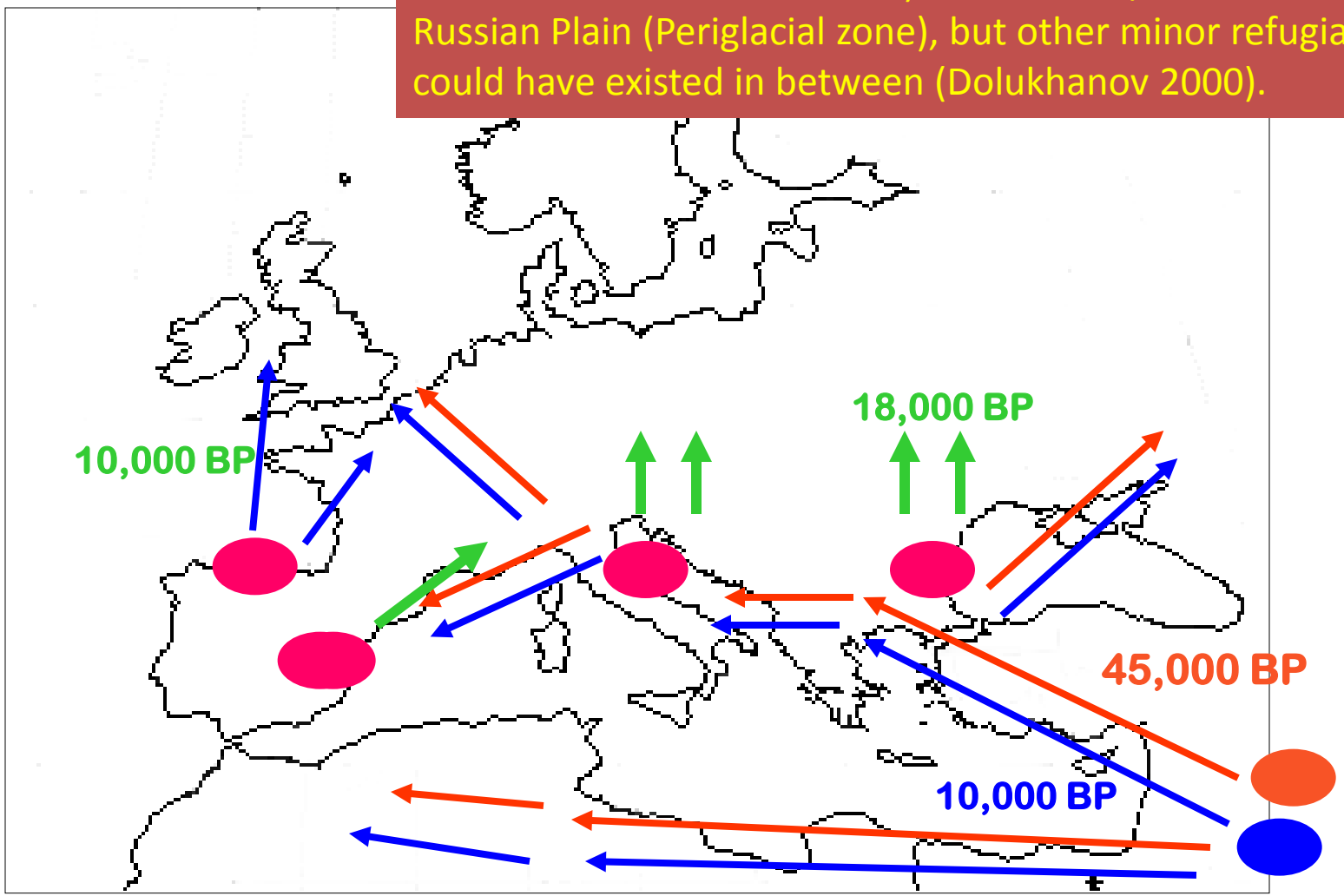
.- iron making

.- growing of domesticated bananas and plantains (north-eastern corner of equatorial rain forest)

mtDNA signature of SLAVE TRADE



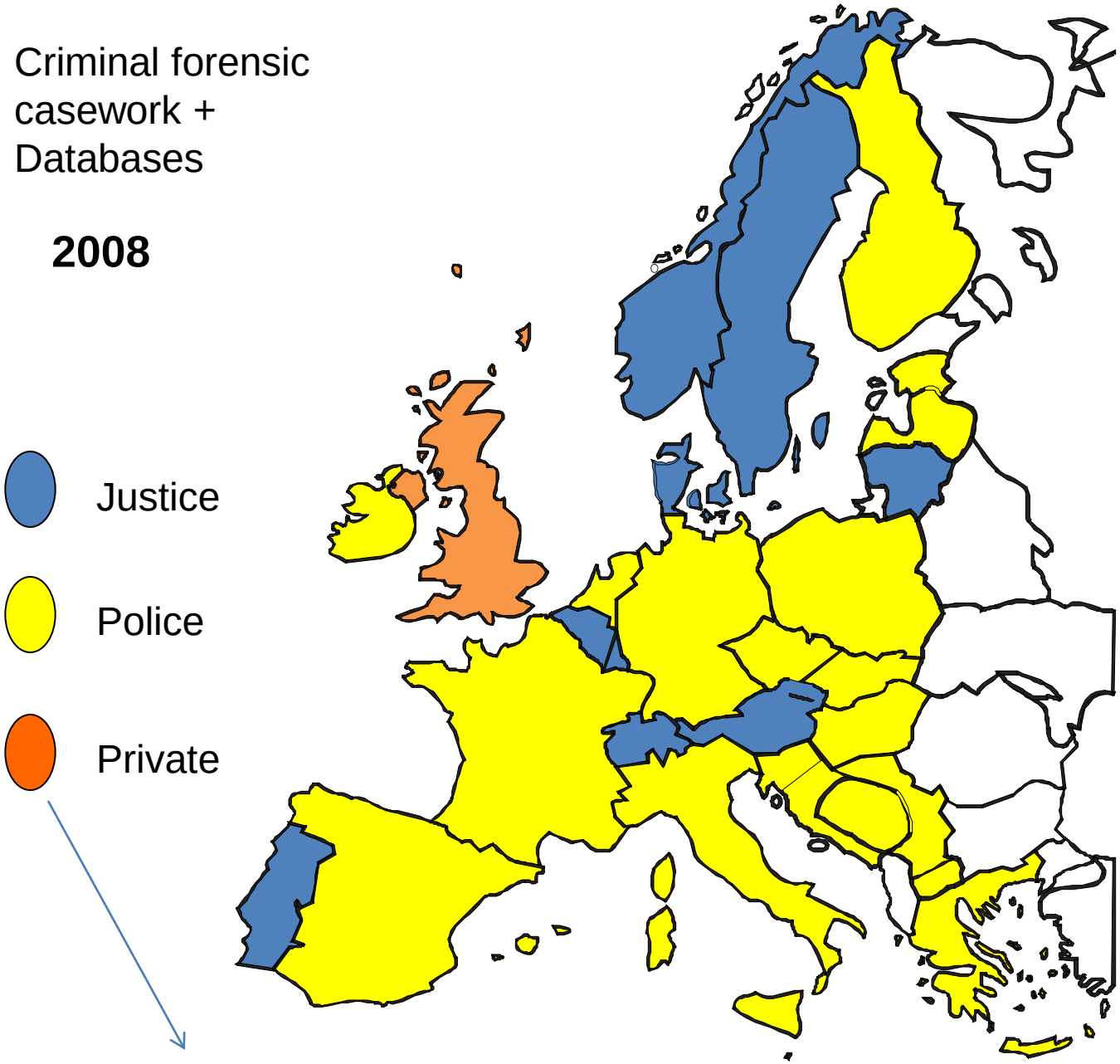
Northern Europe was gradually recolonized from refugia after the Last Glacial Maximum (LGM), ~20,000 years ago (Housley et al. 1997). The two major refugia were in southwestern France/Cantabria (Atlantic and western Mediterranean zone) and Ukraine/Central Russian Plain (Periglacial zone), but other minor refugia could have existed in between (Dolukhanov 2000).



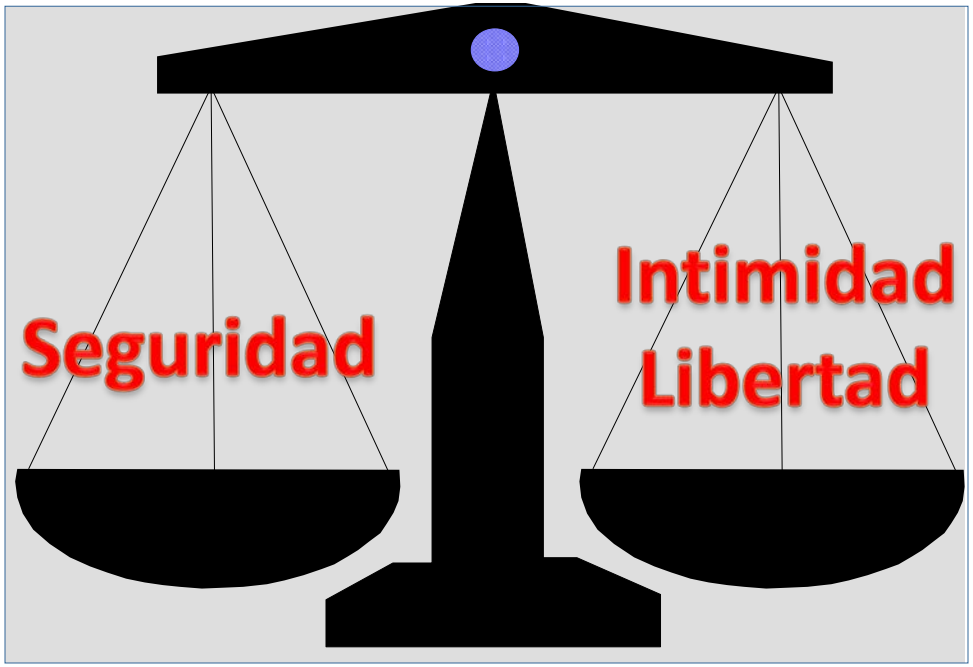

T&S
Hay secretos simples de revelar...
Test ADN de Paternidad
Disponible en Farmacias Ahumada
www.paternidad.cl
800 800 007

Criminal forensic casework + Databases

2008



Economic pressure in favor of general databases



El cambio de valores en los últimos años (¿Por qué valoramos distinto seguridad y libertad?)

Las bases de datos son necesarias pero la libertad individual también

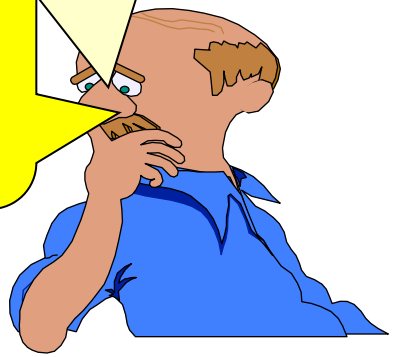
Una pendiente deslizante peligrosa que recuerda el mito de la caverna de Platón-Un mundo feliz o GATTACA

Determinismo genético versus determinismo cultural. ¿Otra vez el debate?

¿Hai diferencias de CI entre as razas?
¿Esto é xenético ou ambiental?

Seguramente una pregunta
mellor sería...

¿Qué mide exactamente o CI?
¿E qué significa unha raza?



50% investigación
50% a la aplicación práctica de la genética

70% son licenciados o doctores: Biología (42), Farmacia (8), Medicina (7), Biotecnología (3), Matemáticas (2), Bioinformática (2), Física (1), Química (1), Veterinaria (1), Filosofía-Derecho (2)
30% técnicos

12 nacionalidades- Inglés es esencial- Movilidad