



Análisis genético de formas no mendelianas de enfermedades comunes.

Jordi Pérez-Tur

Unitat de Genètica Molecular

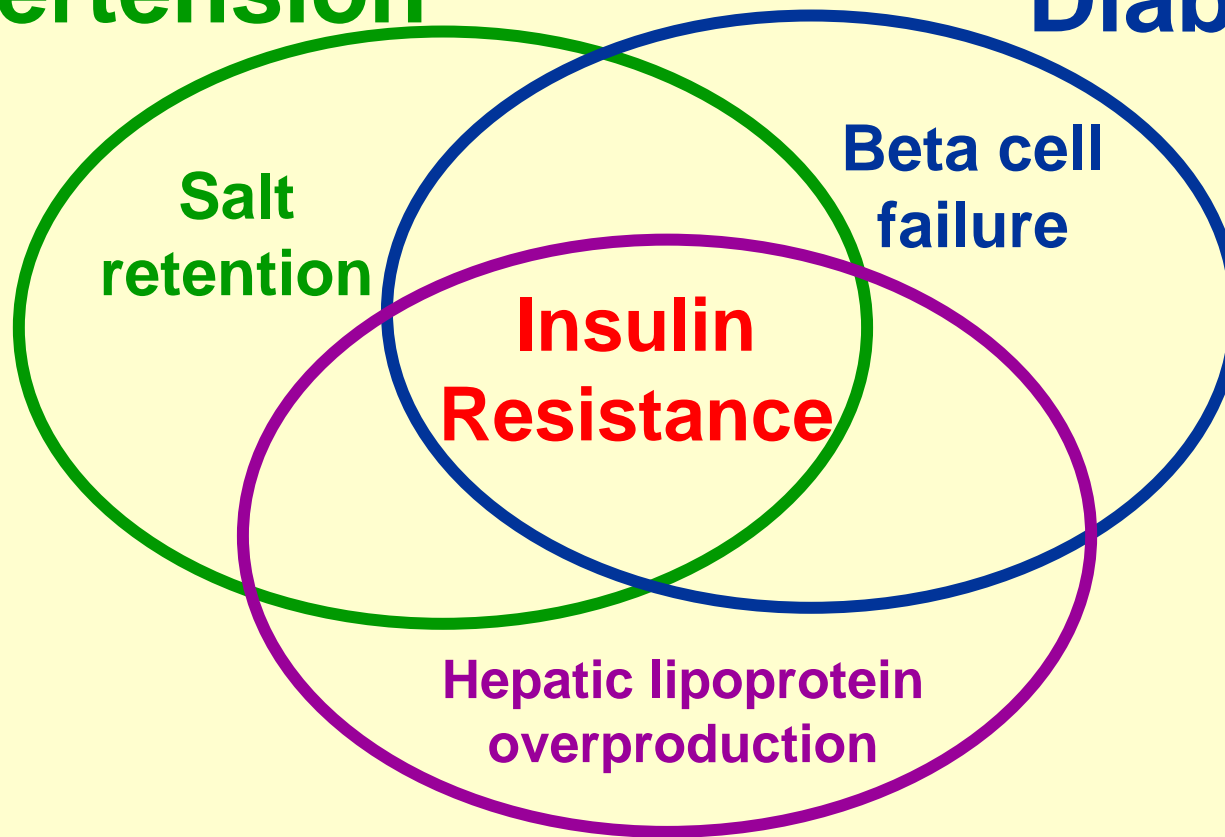
CIBERNED

Institut de Biomedicina de València-CSIC

The Human Insulin Resistance Syndromes (Metabolic Syndrome X)

Hypertension

Diabetes



Dyslipidaemia

Predominant causes of coronary heart disease

Herramientas para la identificación de genes (enf. Complejas)

- Secuencias de genomas
 - Anotación
 - Búsqueda por homología
 - Cartografiado comparativo
 - Recursos basados en SNP
 - “*Data/text mining*”
- LD y mapas de haplotipos
- QTL/Marcadores indirectos (*surrogate markers*)
- Plataformas tecnológicas

La genética molecular se (y nos) dirige hacia la
medicina personalizada...

la práctica de la medicina que incluye la
individualidad genética humana

Variación común en enfermedades complejas.

Estudios caso-control simples. Comparando frecuencias de variantes genéticas comunes se identifica factores de riesgo. Permiten identificar alelos de riesgo y alelos protectores.

Fenotipo	Gene	Variante
Úlcera péptica	ABO	B
IDDM*	HLA	DR3,4
Alzheimer dementia	APOE	E4
Trombosis venas profundas	F5	Leiden
Falciparum malaria*	HBBE	S
SIDA*	CCR5	32
Cáncer Colorectal	APC	3920A
NIDDM	PPAR	12A

Análisis genético de enfermedades complejas.

1. ¿Es un carácter genético? Epidemiología.
2. Métodos clásicos (microsatélites):
 1. Análisis de parejas de parientes afectados (*Affected relative-pair analysis*)
 2. Basados en desequilibrio de transmisión (TDT)
 3. Análisis de parejas de hermanos (*sib pairs*)
3. Métodos recientes/en desarrollo:
 1. Basados en LD
 2. Fundamentalmente: GWA
4. Otros métodos: Marcadores indirectos.
5. Epigenética, nuevos polimorfismos (CNV)

2. Métodos clásicos: Uso de microsatélites.

2.1. Affected-relative pair analysis.

2.2. TDT/STDT

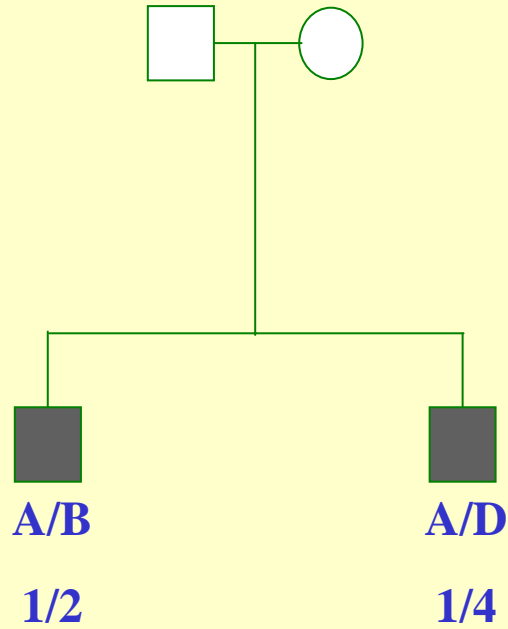
2.3. Asociación/Gen candidato (no microsatélites):

2.3.1. Utilizan poblaciones no emparentadas

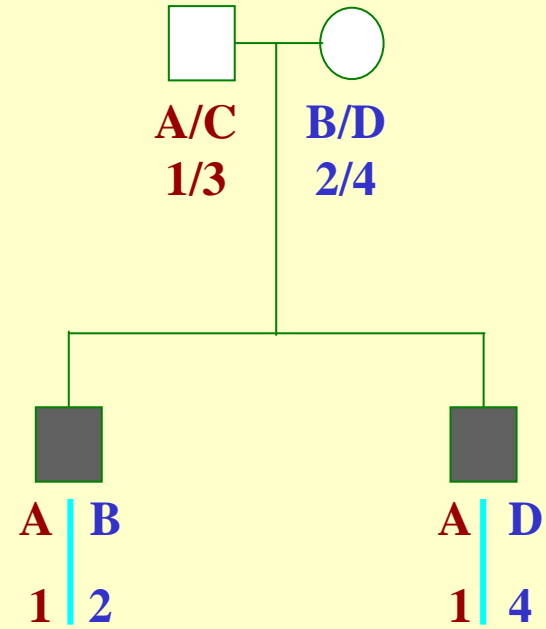
2.3.2. Combinan desequilibrio de ligamiento con gen candidato

2.3.3. Variantes alélicas en genes candidatos pueden estar estadísticamente más representadas en la una población que en la otra.

2.3. Análisis de tipo *sib-pair*

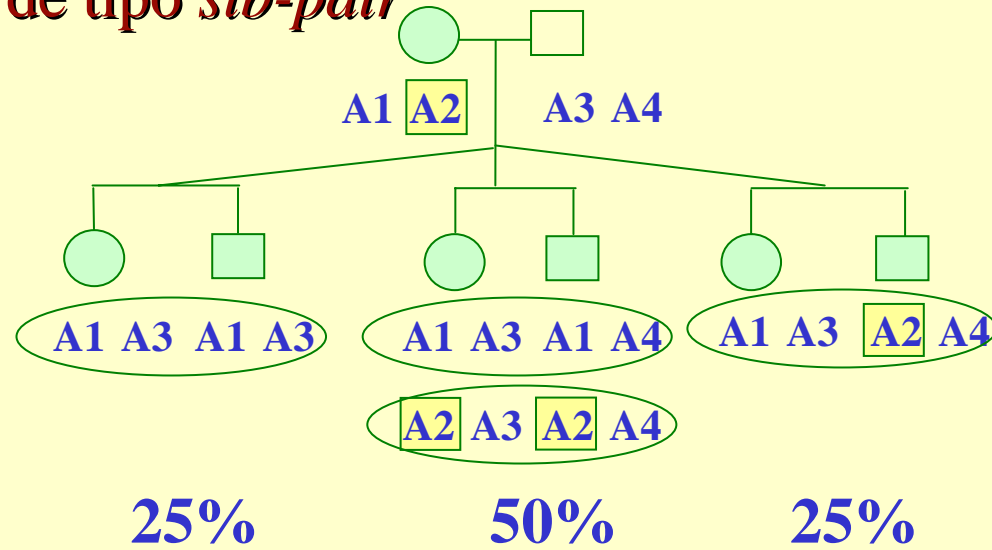


$Ig \dagger qwlfrv \# sru \# Hwdgr$
 $+IE V,$



$Ig \dagger qwlfrv \# sru \# R uljhq$
 $+IE G,$

2.3. Análisis de tipo *sib-pair*



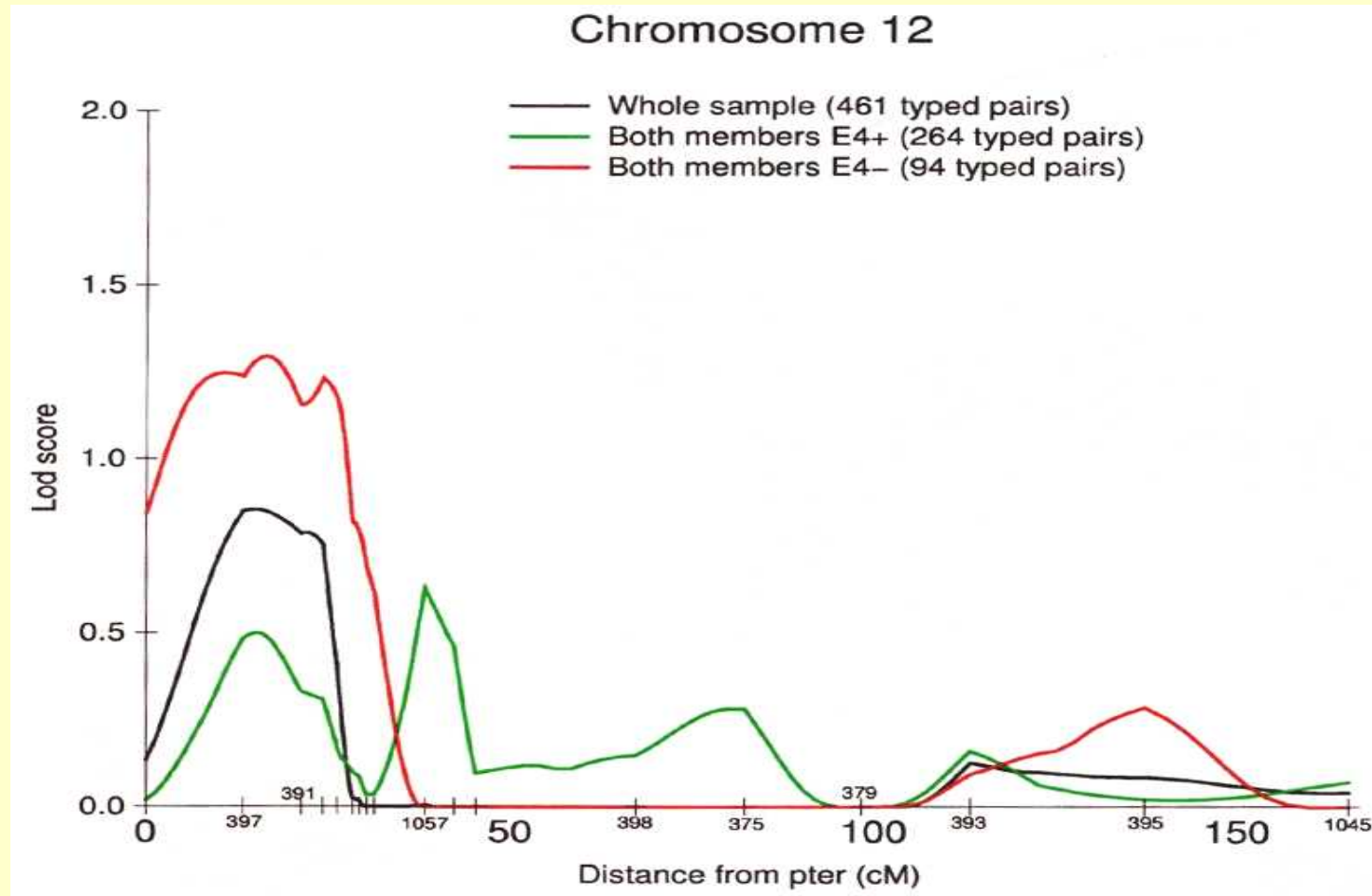
Búsqueda Genómica

Estadío 1°
 Búsqueda preliminar

Estadío 2°
 Limpieza

Estadío 3°
 Refinado

2.3. Análisis de tipo *sib-pair*



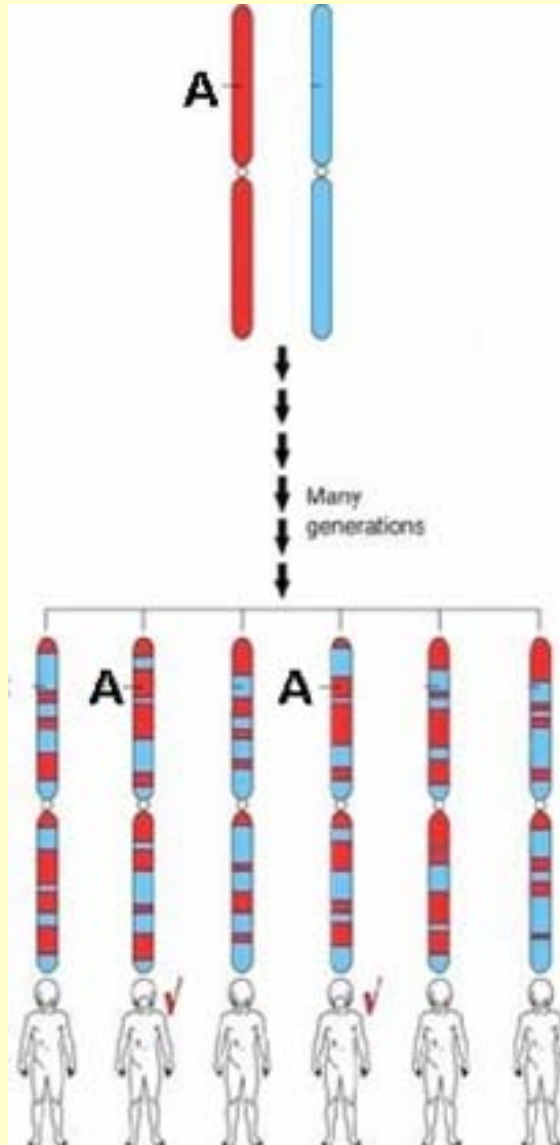
2. Cartografiado por LD.

Siempre que una mutación tenga un origen simple, se podrá identificar por su asociación con los marcadores cercanos.

Definiciones.

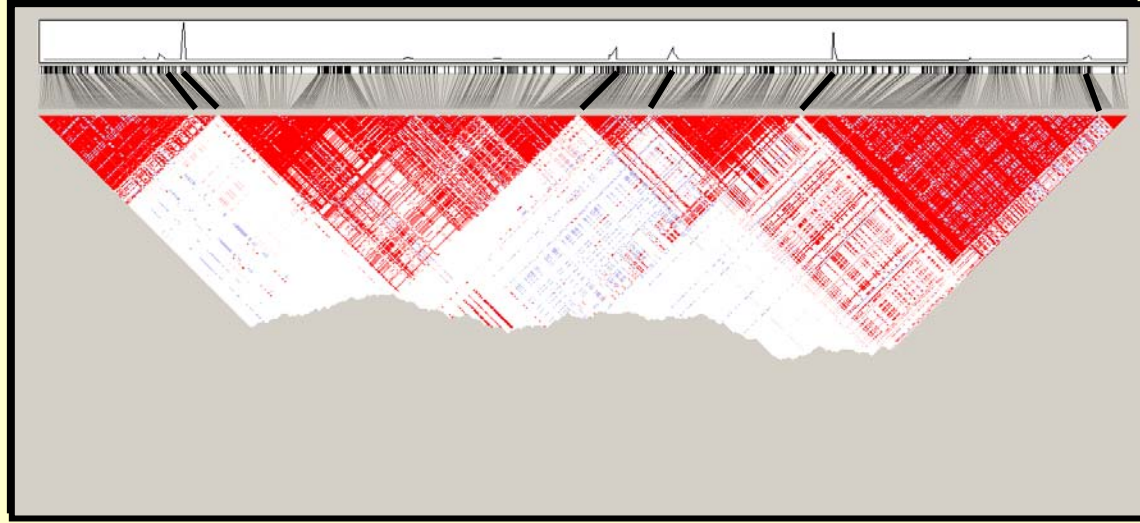
- *LD (desequilibrio de ligamiento)*: Es la medida de la desviación respecto de una asociación al azar para dos alelos en 2 SNPs distintos. Asume no recombinación y se mide por D' , r^2 y LOD.
- *Haplotipos en fase*: Distribución estimada de alelos de SNPs.
- *Tag SNPs*: Conjunto mínimo de SNPs para identificar un haplotipo. Entre dos SNPs, si $r^2 = 1$ resulta redundante genotipar los 2 puesto que uno fina el otro

Origen del desequilibrio de ligamiento.

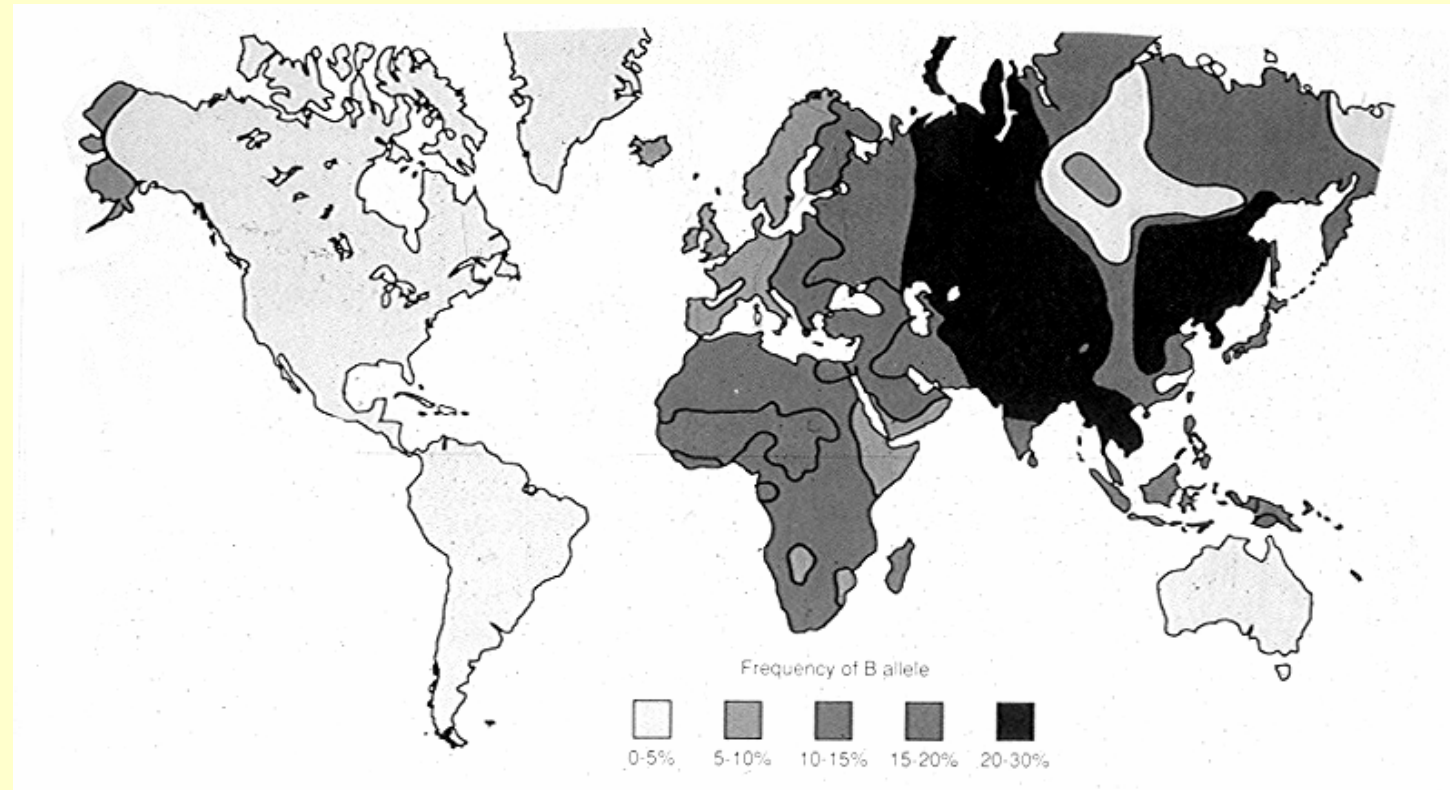


En una población determinada 55% puede tener una versión de un haplotipo, 30% puede tener otra, el 8% puede poseer una tercera y el resto dispondrá de haplotipos aún menos frecuentes.

LD es el resultado de los *puntos calientes* de entrecruzamiento.



Distribución mundial del alelo I^B (ABO)



Genética No Mendeliana: El proyecto HapMap



Poblaciones Incluidas

- Yoruba (Ibadan, Nigeria)
30 tríos (padres/hijo)
- Chinos Han (Beijing, China)
45 individuos no relacionados
- Japoneses (Tokio, Japón)
45 individuos no relacionados
- CEPH (residentes en Utah de ascendencia del Noroeste de Europa)
30 tríos padres/hijo

Proyecto HapMap.

- Un recurso gratuito y fácilmente accesible para aumentar la potencia de estudios de asociación con caracteres biomédicos.
- Genotipado de alta densidad de SNPs proporciona información sobre estructura de LD del genoma y permite obtener información técnica sobre el uso de cada SNP.
- Toda la información es accesible de manera gratuita a través de la web y facilita el diseño de estudios.

Proyecto HapMap: Desarrollo

FASE I: Completa.

1.000.000 SNPs genotipadas en las 270 muestras.

ENCODE también realizado.

FASE II: Completa.

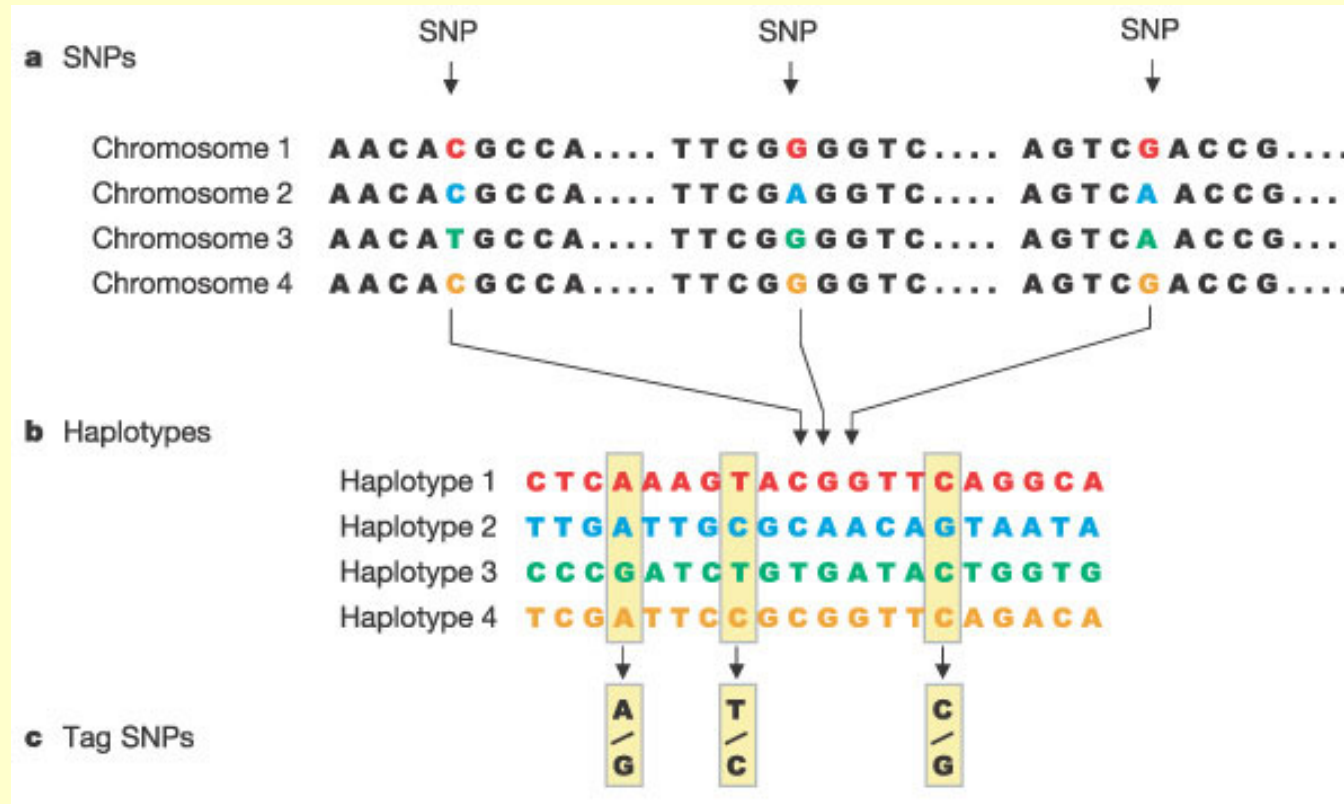
3.500.000 SNPs genotipados en las 270 muestras.

Proyecto HapMap: Desarrollo

Objetivos:

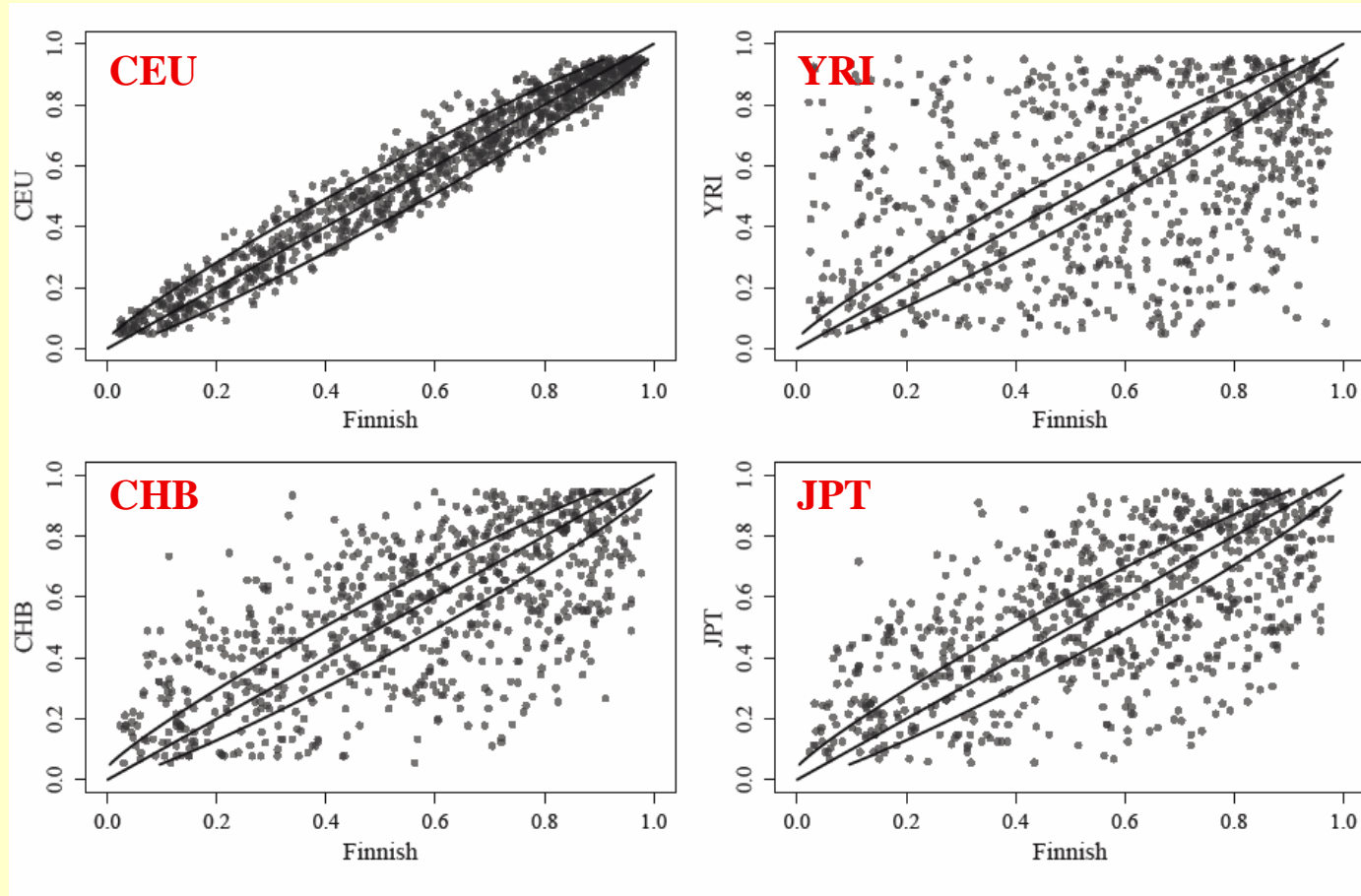
- Proporcionar una herramienta que permita la identificación de haplotipos (determinación de regiones a estudiar) en los que existan variantes genéticas que influyeran:
 - El riesgo frente a una enfermedad.
 - El *tipo* de progresión de la misma.
 - La respuesta individual frente al tratamiento.
 - Permitir un diagnóstico precoz (=intervención precoz) de las enfermedades.
- Definir los perfiles de variación genética a lo largo del genoma humano.
- Servir de guía para seleccionar los tagSNP.

Genética No Mendeliana: El proyecto HapMap



El objetivo del Proyecto HapMap es identificar los *Tag SNPs* a fin de genotipar entre 300.000 y 600.000 SNPs obteniendo la misma información que si se genotiparan los 10.000.000 de SNPs que se estima que pueden existir.

Frecuencias alélicas: HapMap vs FUSION.



La genética molecular se (y nos) dirige hacia la
medicina personalizada...

la práctica de la medicina que incluye la
individualidad genética humana

El proyecto HapMap es esencial en la definición de
esta individualidad incluyendo aquella frente a la
susceptibilidad por la enfermedad y
la respuesta a tratamientos

Utilidad del LD en estudios de asociación.

Si soy una **variante causal** lo que resulta importante para la detección de la asociación es lo bien que correlaciono con el resto de SNPs o haplotipos analizados.

Aún queda mucho trabajo por hacer: Demostrar causalidad de los SNPs asociados

Even trained statisticians often fail to appreciate the extent to which statistics are vitiated by the unrecorded assumptions of their interpreters. . . It is easy to prove that the wearing of tall hats and the carrying of umbrellas enlarges the chest, prolongs life, and confers comparative immunity from disease. . . . A university degree, a daily bath, the owning of thirty pairs of trousers, a knowledge of Wagner's music, a pew in church, anything, in short, that implies more means and better nurture . . . can be statistically palmed off as a magic-spell conferring all sorts of privileges. . . . The mathematician whose correlations would fill a Newton with admiration, may, in collecting and accepting data and drawing conclusions from them, fall into quite crude errors by just such popular oversights as I have been describing.

**George Bernard Shaw, Preface,
The Doctors Dilemma (1906)**

Diseño de estudios de GWA.

Representa una oportunidad sin precedentes para identificar variantes que predisponen frente a las enfermedades.

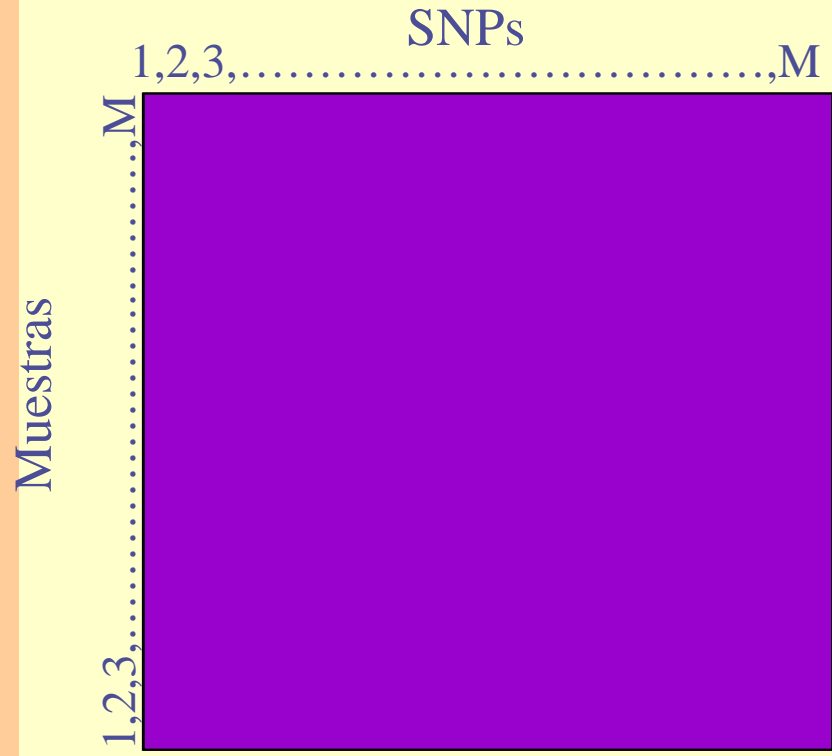
Facilitado por HapMap y la reducción en los costes de secuenciación.

Genotipado de al menos 100.000 SNPs en unas 100s-1000s muestras.
Diseño experimental crítico!!!

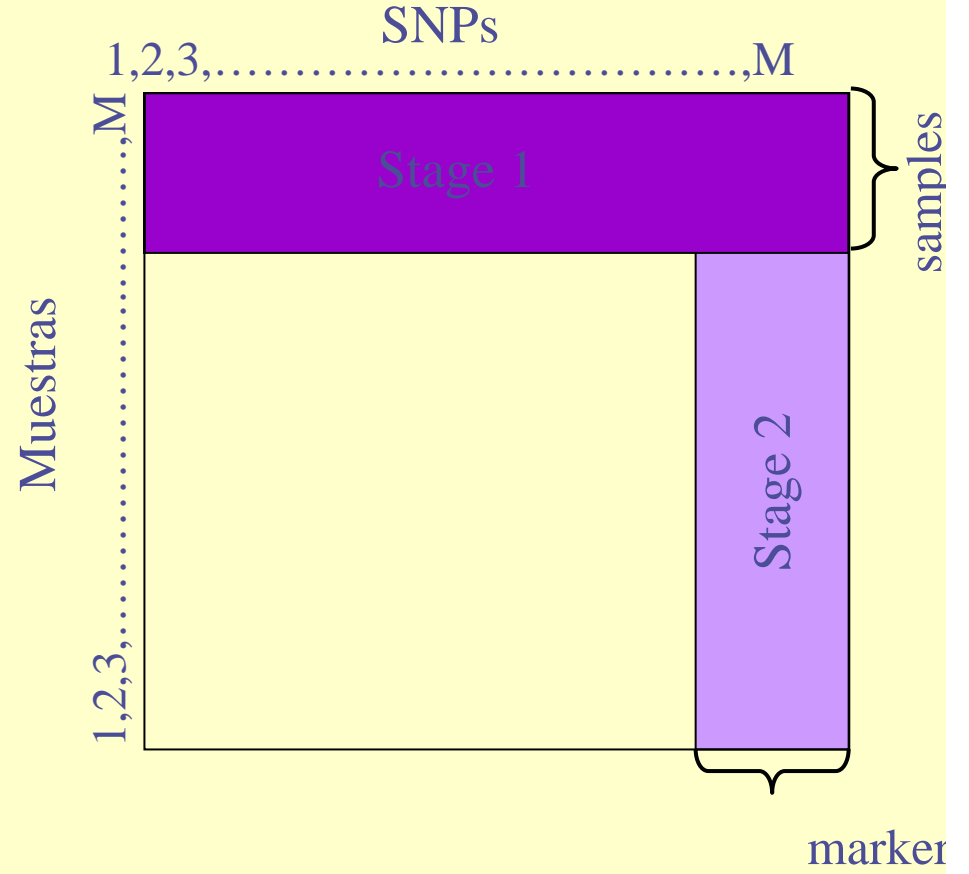
Para estudios a muy gran escala interesa llevar a cabo un estudio en dos etapas.

Comparativa

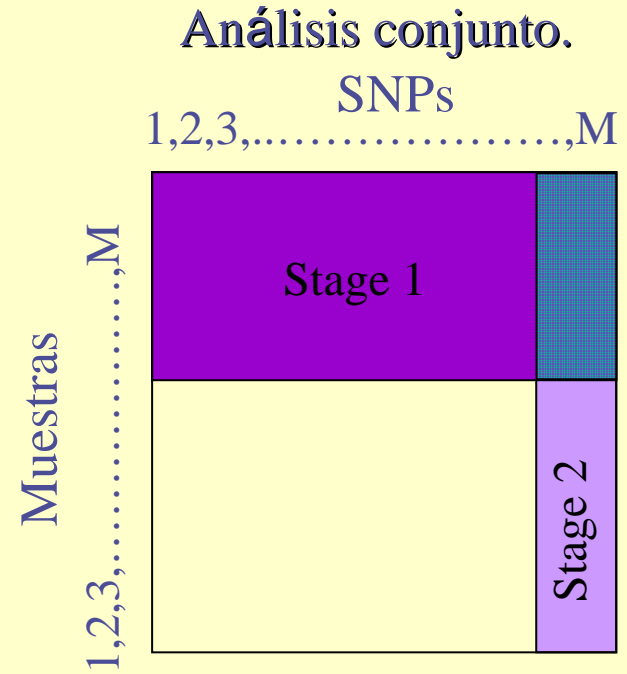
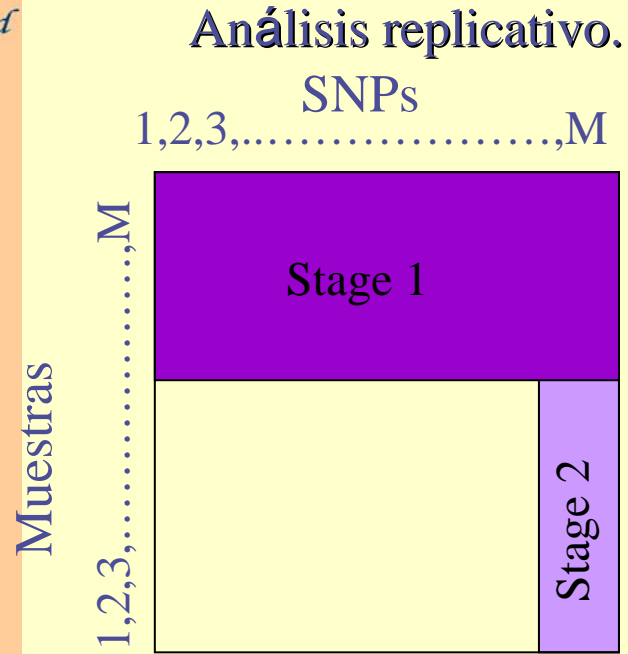
Una etapa



Dos etapas

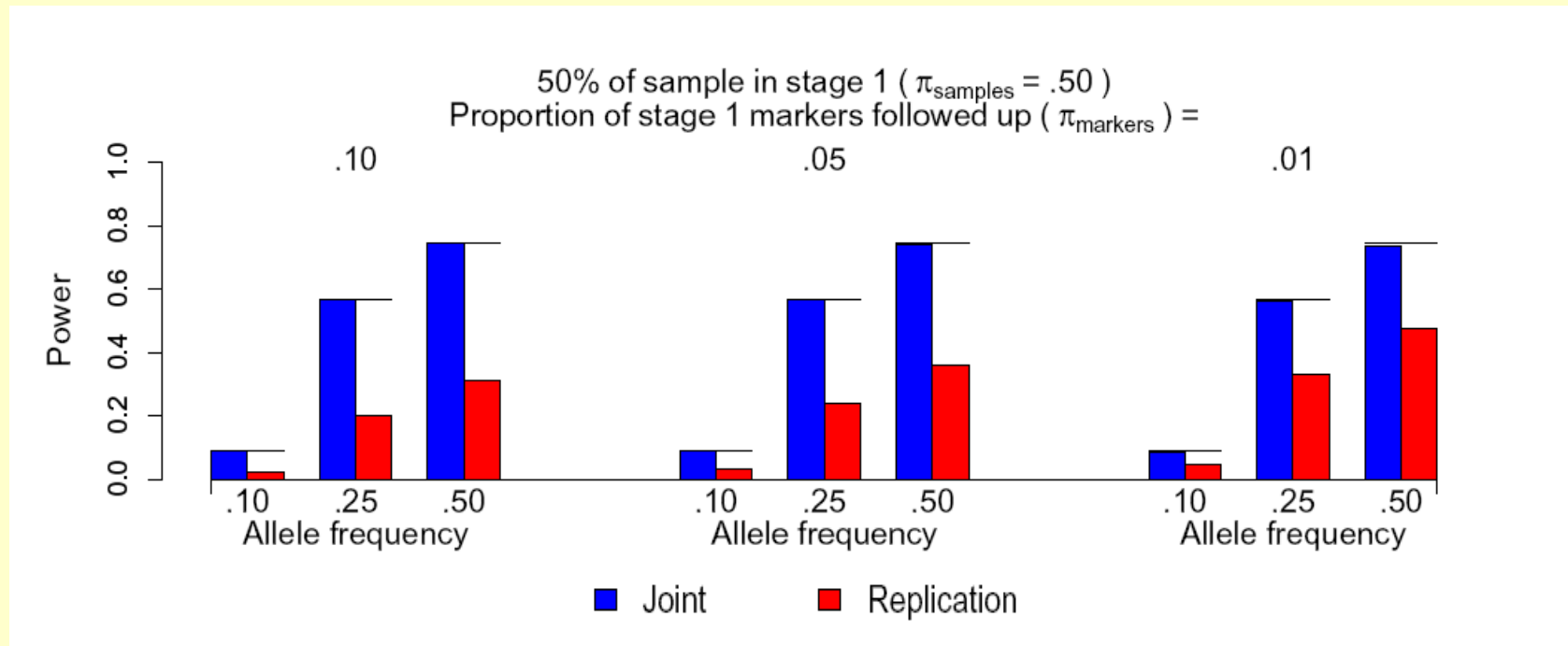


Diseño de estudios de GWA.



Diseño de estudios de GWA.

El análisis conjunto es más potente que el replicativo.



Factores que influyen en el coste y eficiencia de los GWAs

Fracción de muestras obtenidas y genotipadas en el primer sector.

Fracción de SNPs que se genotipan en el estadio 2.

Relación de coste por genotipo entre Estadio 1 y Estadio 2

Para un GWA en 2 estadíos:

Es eficiente (Análisis conjunto mejor que replicación)

Si el coste de genotipar en los dos estadíos es similar, se pueden genotipar un 30% de la muestras en el primer estadío (signif a 0,05 y 250K de densidad)

Si cambian las condiciones (densidad de marcadores, cambio en condiciones de trabajo, menor requerimiento de significación...) hay que aumentar el número de polimorfismos.

Aplicación del HapMap:

Diseño de estudios para:

Evaluar SNPs.

Identificación/selección de tag-SNPs.

Evaluación de densidad.

Mejorar la asociación genética.

Comparación de estudios múltiples (mismos SNPs)

Conexión con características genómicas.

Integración con otra información funcional y de expresión.

Otros: LOH, selección, *admixture*.

¿Cómo seleccionar los SNPs?

¿Cuál es la hipótesis genética?

¿Qué variantes quieren testar frente a la enfermedad?

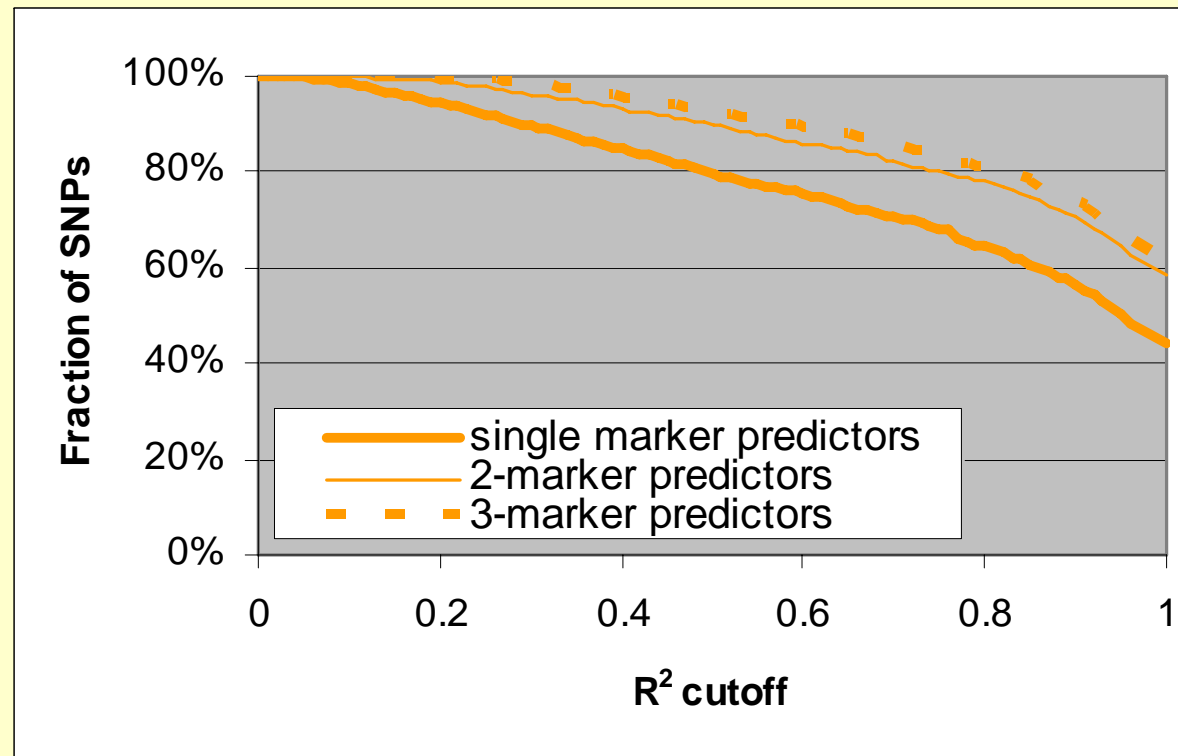
En función de la anotación funcional (p.ej. Codif.)

En función de la frecuencia del alelo menor.

Aquellos previamente implicados en asociaciones.

La página web de HapMap permite obtener toda la información en formatos adecuados.

Los haplotipos aumentan la cobertura.



Tecnología.

1. Prehistoria: RFLP/ARMS
2. TaqMan.
3. dHPLC.
4. SNPLex
5. Sequenom
6. (Affymetrix)
7. Illumina
8. ????

Identificación de genes (1.980-2.002)

Nº caracteres mendelianos.

Nº de caracteres complejos.

