



Fundamentos de genética

Objetivos de aprendizaje

- Revisar los tipos de variación genética humana
- Reconocer los objetivos y hallazgos del Proyecto 1000 Genomas
- Distinguir entre rasgos mendelianos y complejos
- Comprender la base de los Estudios de asociación del genoma completo (GWAS)
- Comprender la estimación de puntuaciones poligénicas

Describir

- Variación genética humana
- Genomas de referencia basados en poblaciones: Proyecto 1000 Genomas
- Rasgos complejos
- Estudios de asociación del genoma completo (GWAS)
- Puntuaciones poligénicas

Variación genética humana

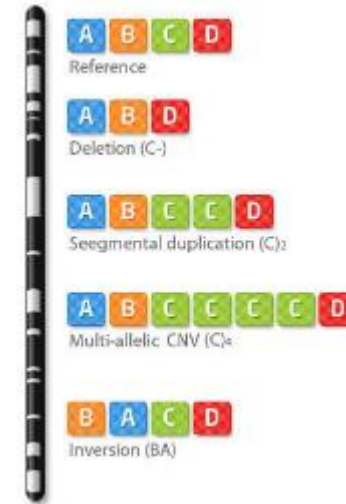
Variaciones genéticas

Número de cromosomas

Estructural
(> 1 kb)

Deleción, duplicación segmentaria, translocación e inversión

Variante del número de copias (CNV)

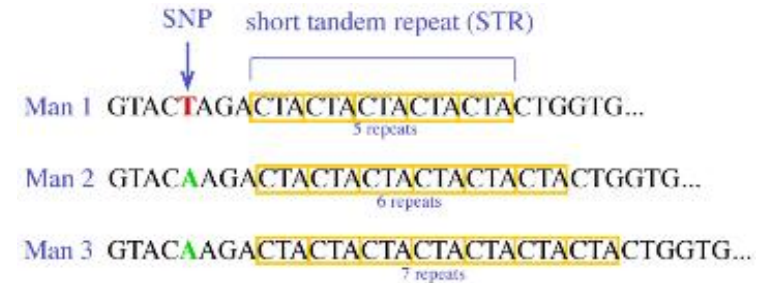


Repeticiones en tándem (por ejemplo, microsatélites [STR])

Secuencia
(<1 kb)

Inserciones y deleciones (Indeles)

Polimorfismo de un solo nucleótido (SNP)



CTTTATAGAGCATAAGCAGCAGCGCAACACCCTTAT mutation
CTTTATAGAGCATA--AGCAGCAGCGCAACACCCTTAT reference

Variación detectada en un genoma humano típico

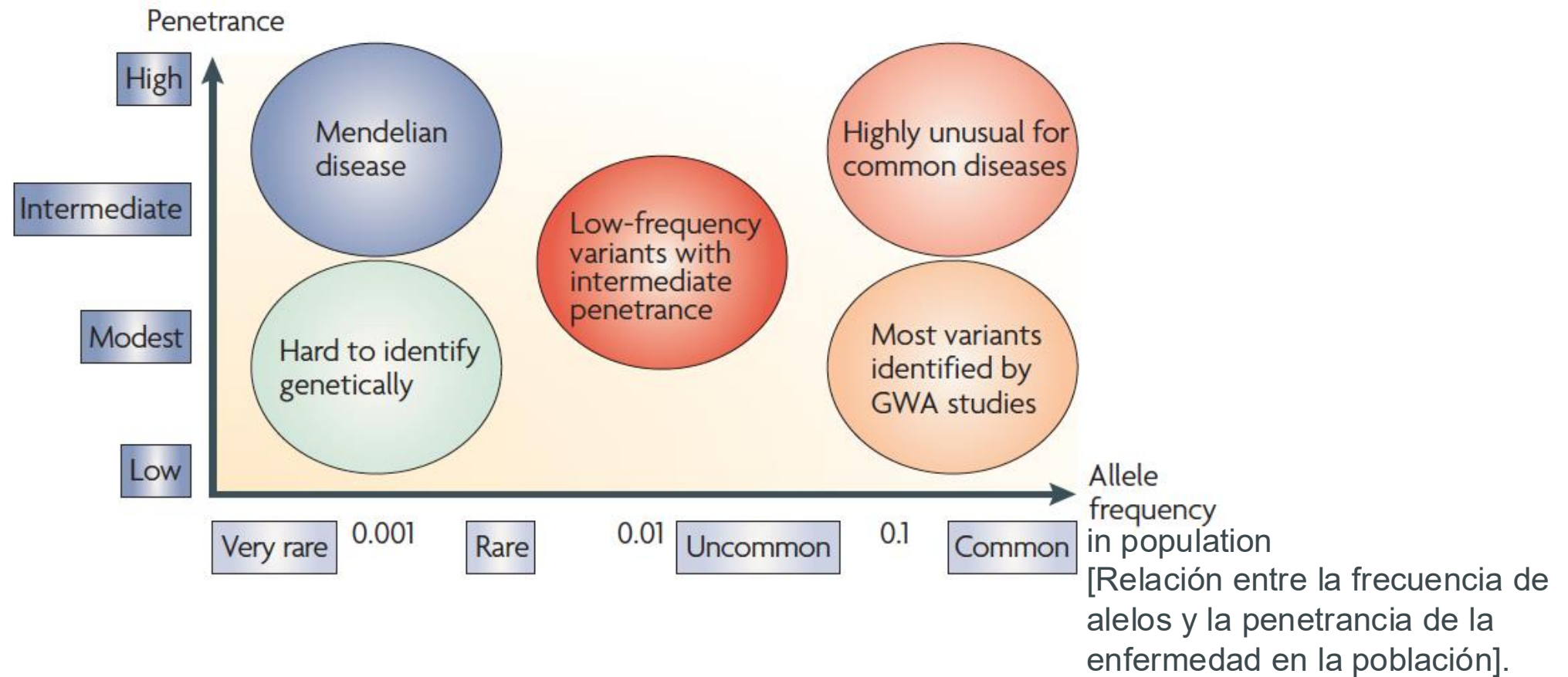
Sitios de variantes autosómicas medianas
por genoma

	África	EAS	EUR
Muestras	661	504	503
SNP	4,31 M	3,55 M	3,53 M
raro (<0,5 %)	18 k	15 k	8 k
Indeles	625 k	546 k	546 k
Grandes deleciones	1,1 k	940	939
Nonsynon	12,2 k	10,2 k	10,2 k
Semifallo	139	144	116
Intrón	2,06 M	1,68 M	1,68 M
UTR	37,2 k	30,0 k	30,0 k
Promotor	102 k	81,6 k	82,2 k
Pérdida de función	182	153	149
HGMD	20	16	18
Semifallo	0	1	2
ClinVar	28	24	29
Semifallo	0	0	1

M, millón; k, mil; HGMD, base de datos de mutaciones genéticas humanas: mutaciones de enfermedades; ClinVar: variantes con significación clínica.

El Consorcio del Proyecto 1000 Genomas. Auton *et al.* A global reference for human genetic variation [Una referencia mundial sobre la variación genética humana]. Nature 2015;526:68–74.

Relationship between allele frequency and disease penetrance



Genomas de referencia basados en
poblaciones: Proyecto 1000 Genomas



El objetivo del Proyecto 1000 Genomas fue encontrar la mayoría de las variantes genéticas con MAF > 1 % en las poblaciones estudiadas.



<http://www.internationalgenome.org/home>

Consorcio del Proyecto 1000 Genomas (Auton et al) Nature 2015; 526: 68–74.

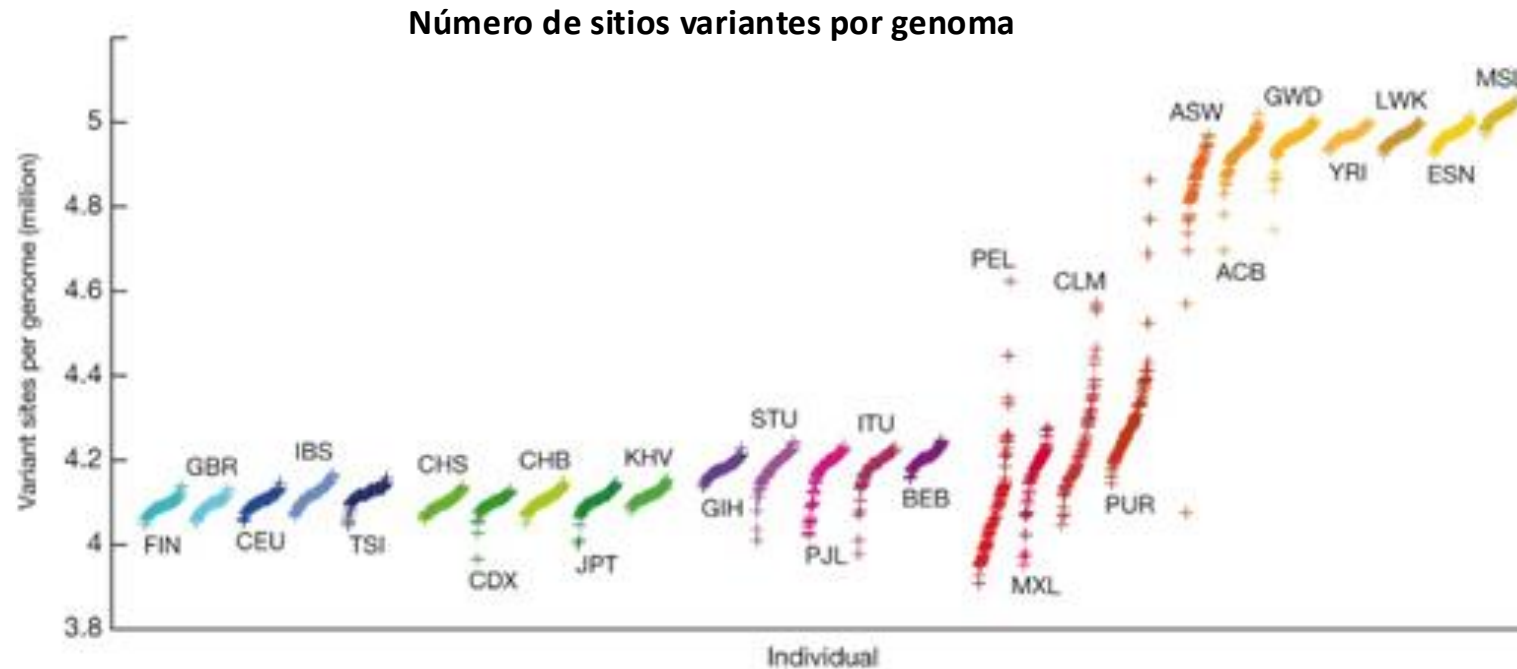
Población		N
Ascendencia africana en el suroeste de EE. UU.	ASW	62
Caribe africano en Barbados	ACB	120
Bengalí en Bangladesh	BEB	144
Británicos de Inglaterra y Escocia	GBR	100
Dai chino en Xishuangbanna, China	CDX	102
Residentes de CEPH y Utah con ascendencia del norte y oeste de Europa, EE. UU.	CEU	186
Colombiano en Medellín, Colombia	CLM	136
Esan en Nigeria	ESN	173
Finlandés en Finlandia	FIN	103
Gambiano en la División Oeste – Mandinka	GWD	179
Indios gujarati en Houston, Texas, EE. UU.	GIH	109
Chinos han en Pekín, China	CHB	120
Han del Sur de China	CHS	163
Poblaciones ibéricas en España	IBS	157
Telugu indio en el Reino Unido	ITU	118
Japoneses en Tokio, Japón	JPT	120
Kinh en la ciudad de Ho Chi Minh, Vietnam	KHV	124
Luhya en Webuye, Kenia	LWK	120
Mende en Sierra Leona	MSL	128
Ascendencia mexicana en Los Ángeles, CA, EE. UU.	MXL	71
Peruano en Lima Perú	PEL	122
Puertorriqueño en Puerto Rico	PUR	139
Punjabi en Lahore, Pakistán	PJL	158
Tamil de Sri Lanka en el Reino Unido	STU	128
Toscana en Italia	TSI	114
Yoruba en Ibadán, Nigeria	YRI	120



<http://www.internationalgenome.org/home>

Consortio del Proyecto 1000 Genomas (Auton et al) Nature 2015; 526: 68–74.

El conjunto de datos final contiene datos de secuencia de 2504 individuos de 26 poblaciones;
> 88 M variantes; 84,7 M SNP.



Rasgos complejos

Rasgos genéticos humanos

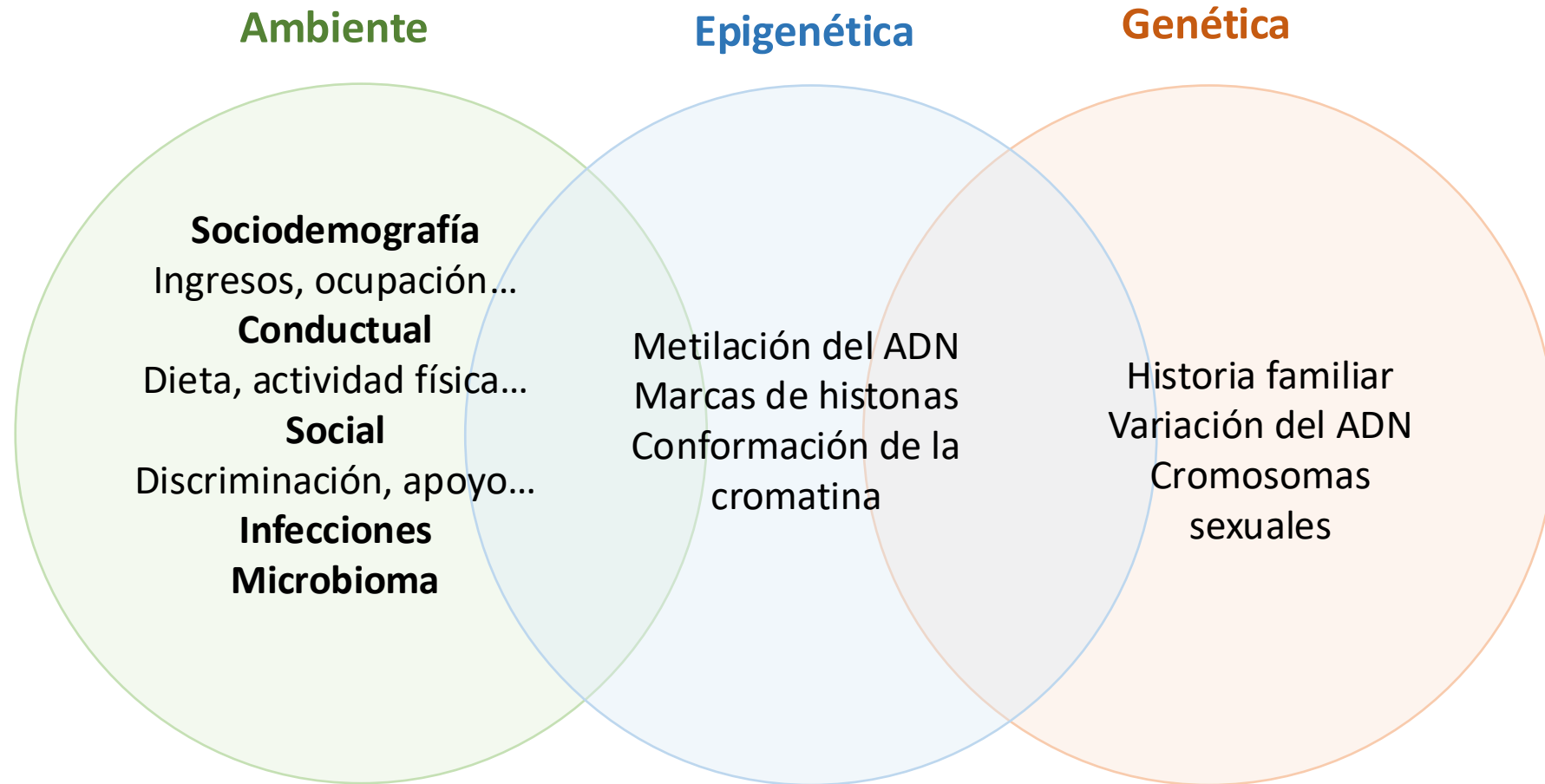
Trastorno mendeliano

- Raro
- Monogénico o poligénico
- Alta penetración o tamaño del efecto
- Alelos raros en la población con alto riesgo para el individuo
- Patrón de herencia: autosómica dominante, recesiva o ligada al cromosoma X
- Influencias ambientales mínimas
- A menudo se producen perturbaciones sustanciales en la función de genes y proteínas

Trastorno complejo

- Común
- Poligénico
- Baja penetración o tamaño del efecto
- Alelos raros y comunes en la población con riesgo o tamaño del efecto pequeño a moderado
- Patrón de herencia complejo
- Influencias ambientales
- Variaciones comunes menores, principalmente cambios regulatorios

Enfermedades y rasgos complejos



Por ejemplo, altura, color de ojos, presión arterial alta, enfermedades cardiovasculares, diabetes, obesidad, cáncer, enfermedades psiquiátricas, Alzheimer, asma, artritis, enfermedades autoinmunes...

Estudios de asociación del
genoma completo (GWAS)



GWAS - Genome-wide Association Studies

NHGRI FACT SHEETS

genome.gov

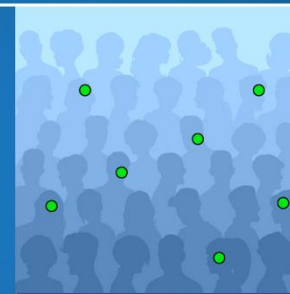
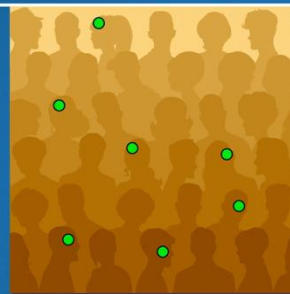
Individuals with disease

Individuals without disease



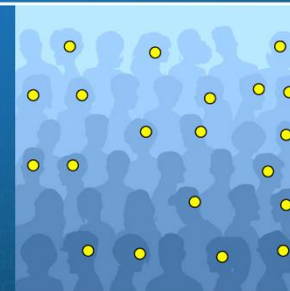
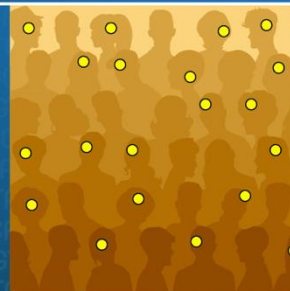
Using a CHIP can genotype
500,000 - 5 Million SNPs

SNP 1



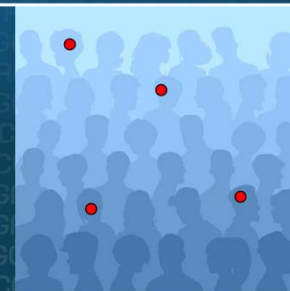
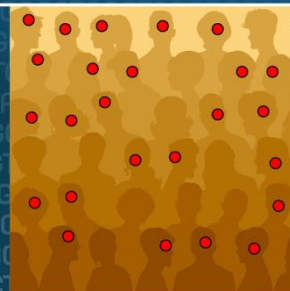
SNP 1
No association
to disease

SNP 2



SNP 2
No association
to disease

SNP 3



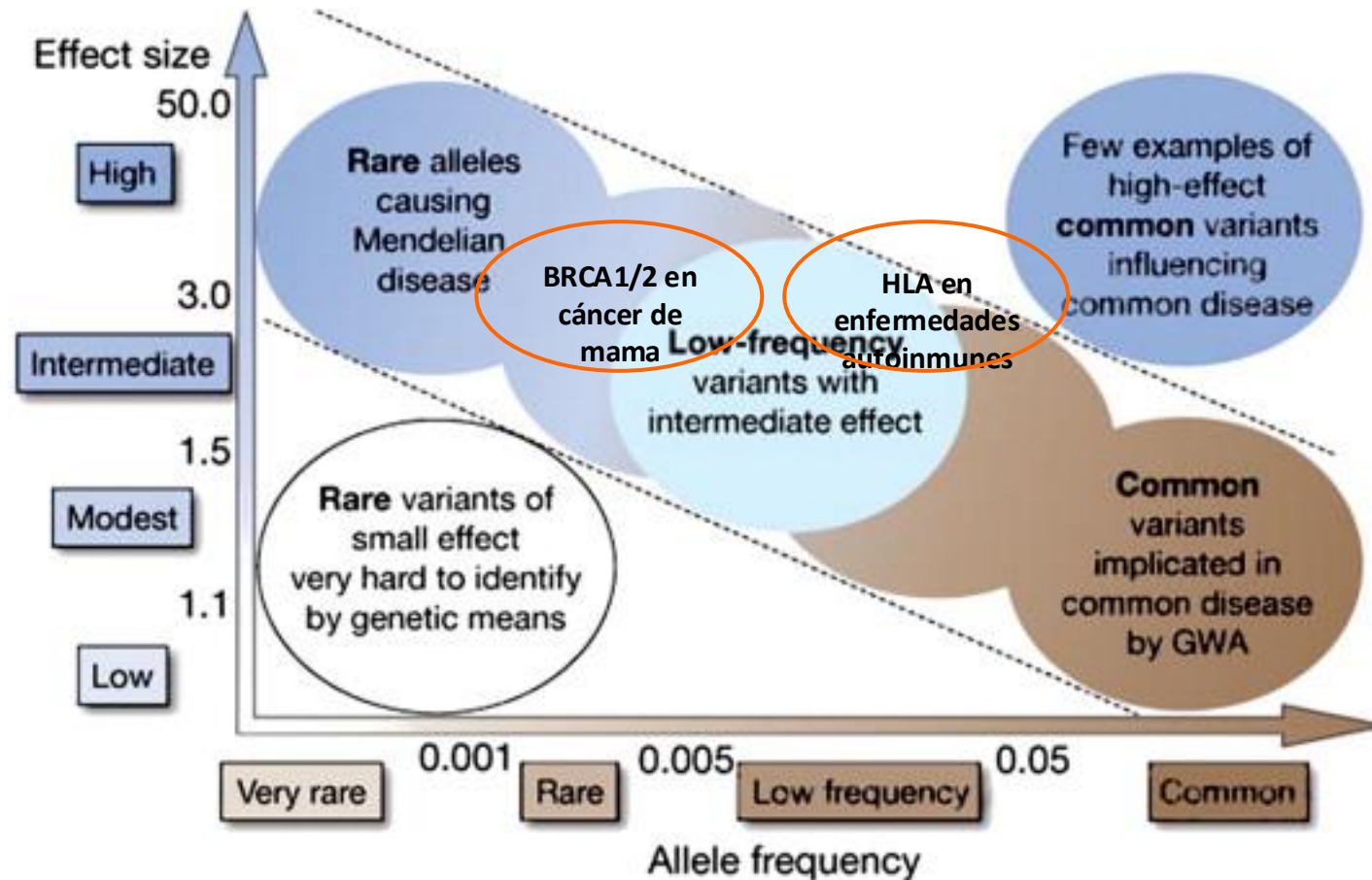
SNP 3
Associated
to disease



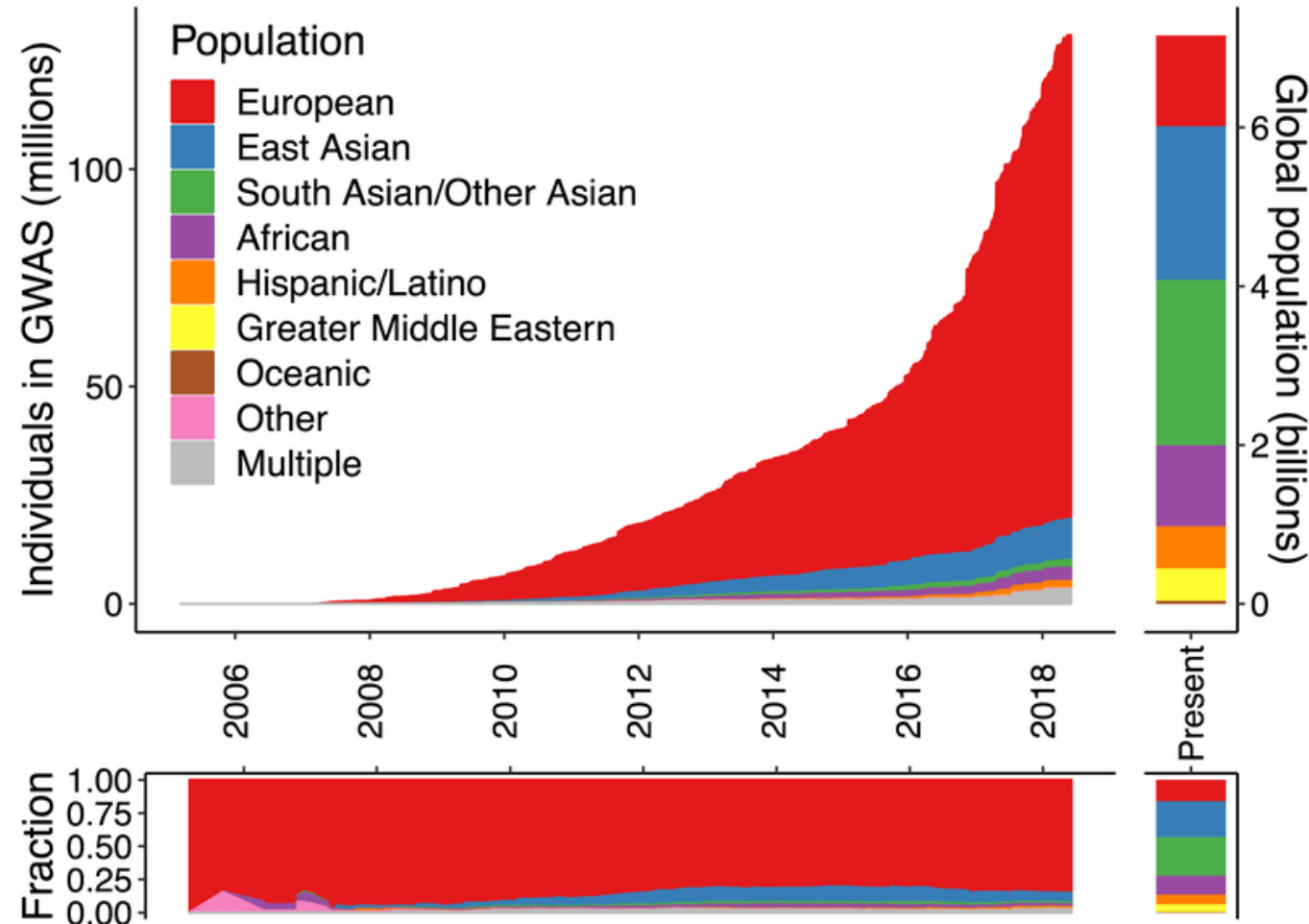
NIH

National Human Genome
Research Institute

Viabilidad de identificar variantes genéticas mediante la frecuencia de alelos de riesgo y la fuerza del efecto genético (razón de momios)



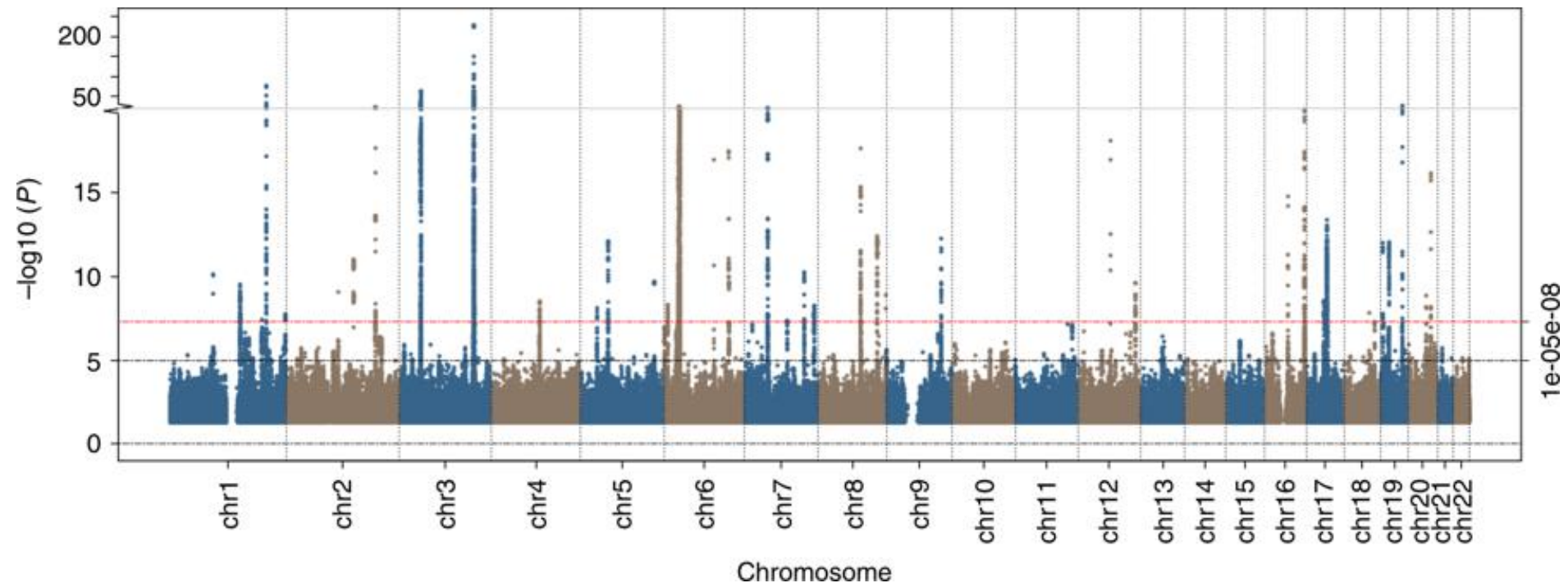
Ancestry of GWAS participants over time compared to the global population [Ascendencia de los participantes de GWAS a lo largo del tiempo en comparación con la población mundial].



Martin *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities [El uso clínico de las puntuaciones de riesgo poligénico actuales puede exacerbar las disparidades en la salud.]. Nat Genet. 2019;51:584-591.

Interpretación de los resultados

Manhattan plot [Gráfico de Manhattan]



Manhattan plot of genome-wide association analysis of self-reported ulcers in UK Biobank [Gráfico de Manhattan del análisis de asociación genómica de úlceras autoinformadas en el Biobanco del Reino Unido] (<https://www.ukbiobank.ac.uk/>)*.

* Estudio poblacional prospectivo de > 500.000 individuos con datos fenotípicos y genéticos extensos y accesibles.

Table 2 | Novel ancestry-specific non-HLA associated regions.

Most significant SNP(s)	Chr.	Position (b37)	Gene region*	Region rank	Ref. allele	RAF case	RAF control	P value	OR (95% CI)	Regional stepwise P value	dbSNP function†
EA Tier 1											
rs1132200 ^{d,P}	3q13	119150836	TMEM39A ¹ -TIMMDC1	29	A	0.138	0.159	1.37×10^{-7}	0.83 (0.77-0.89)		missense
rs1131265 ^d	3q13	119222456	TMEM39A ¹ -TIMMDC1	29	C	0.161	0.186	1.42×10^{-9}	0.81 (0.76-0.87)	5.96×10^{-9}	coding-synon
rs1534154	3q13	119311030	TMEM39A ¹ -TIMMDC1	29	G	0.177	0.165	2.60×10^{-4}	1.11 (1.05-1.18)	1.35×10^{-3}	
rs3733345	4p16	954247	DGKQ	33	G	0.444	0.466	5.84×10^{-8}	0.89 (0.85-0.93)	5.84×10^{-8}	untranslated-3
rs4690229 ⁱ	4p16	970724	DGKQ	33	T	0.486	0.462	1.62×10^{-8}	1.13 (1.09-1.19)		
rs10498722 ^d	6p22	25186512	LRRC16A	12	A	0.096	0.080	2.87×10^{-10}	1.30 (1.20-1.41)	2.87×10^{-10}	
rs35789010 ⁱ	6p22	25514179	LRRC16A	12	A	0.089	0.072	4.59×10^{-19}	1.46 (1.35-1.59)		intron
rs4712969	6p22	25764192	SLC17A4	8	T	0.119	0.093	1.83×10^{-22}	1.42 (1.32-1.52)	1.83×10^{-22}	intron
rs36014129 ⁱ	6p22	25884519	SLC17A4	8	A	0.101	0.079	1.21×10^{-24}	1.50 (1.39-1.62)		
rs2327832	6q23	137973068	OLIG3-LOC100130476‡	6	C	0.239	0.212	1.76×10^{-13}	1.22 (1.15-1.28)	2.38×10^{-8}	
rs17779870	6q23	138156425	OLIG3-LOC100130476‡	6	C	0.132	0.154	5.35×10^{-7}	0.85 (0.80-0.91)	1.80×10^{-5}	intron
rs5029939	6q23	138195723	TNFAIP3‡	6	C	0.061	0.033	2.39×10^{-29}	1.81 (1.63-2.01)	7.21×10^{-22}	intron
rs2230926 ^P	6q23	138196066	TNFAIP3‡	6	C	0.061	0.033	2.79×10^{-29}	1.81 (1.63-2.01)		missense
rs77000060 ⁱ	6q23	138237989	TNFAIP3‡	6	T	0.055	0.030	1.84×10^{-29}	1.89 (1.69-2.11)		
rs73137125	7q11	74018950	GTF2IRD1-GTF2I [§]	15	G	0.219	0.230	3.27×10^{-3}	0.92 (0.88-0.97)	1.08×10^{-5}	
rs73366469	7q11	74033600	GTF2IRD1-GTF2I [§]	15	C	0.126	0.098	2.68×10^{-13}	1.29 (1.21-1.38)	1.11×10^{-15}	
rs2955587	8p23	8098079	FAM86B3P	25	C	0.468	0.442	7.91×10^{-10}	1.15 (1.10-1.20)	7.91×10^{-10}	intron
rs2980512 ⁱ	8p23	8140901	FAM86B3P	25	C	0.497	0.467	3.54×10^{-10}	1.15 (1.10-1.20)		
rs1966115	8q21	79556891	PKIA-ZC2HC1A	23	A	0.296	0.256	1.43×10^{-7}	1.14 (1.09-1.20)	4.11×10^{-11}	
rs12114284	8q21	79558441	PKIA-ZC2HC1A	23	A	0.291	0.256	2.75×10^{-5}	1.11 (1.06-1.17)	1.93×10^{-10}	
rs930297	17q25	73404537	GRB2	38	G	0.106	0.116	1.43×10^{-7}	0.83 (0.77-0.89)	4.91×10^{-8}	
rs1463485 ^d	17q25	73851791	GRB2	38	G	0.223	0.197	7.34×10^{-5}	1.14 (1.07-1.21)	2.27×10^{-5}	near-gene-5

No se han identificado previamente regiones nuevas mediante asociaciones de SNP con valores $P < 5 \times 10^{-8}$ y están resaltadas en gris. Las regiones que son las primeras asociaciones observadas en una ascendencia particular se indican con un superíndice¹ en la región del gen.

i: SNP imputado.

d o r: Modelo dominante o recesivo; si no se indica, se utilizó el modelo aditivo.

P: Asociación publicada: este SNP se ha identificado como causal o como el SNP más significativo en la región del gen.

*Se denomina así por los genes que delimitan la región de asociación, a menos que la bibliografía implique fuertemente a un gen específico.

†Efecto funcional previsto de dbSNP.

‡La región OLIG3-LOC100130476 alcanza significancia de nivel 1 incluso después de ajustar la señal TNFAIP3, una región SLE establecida.

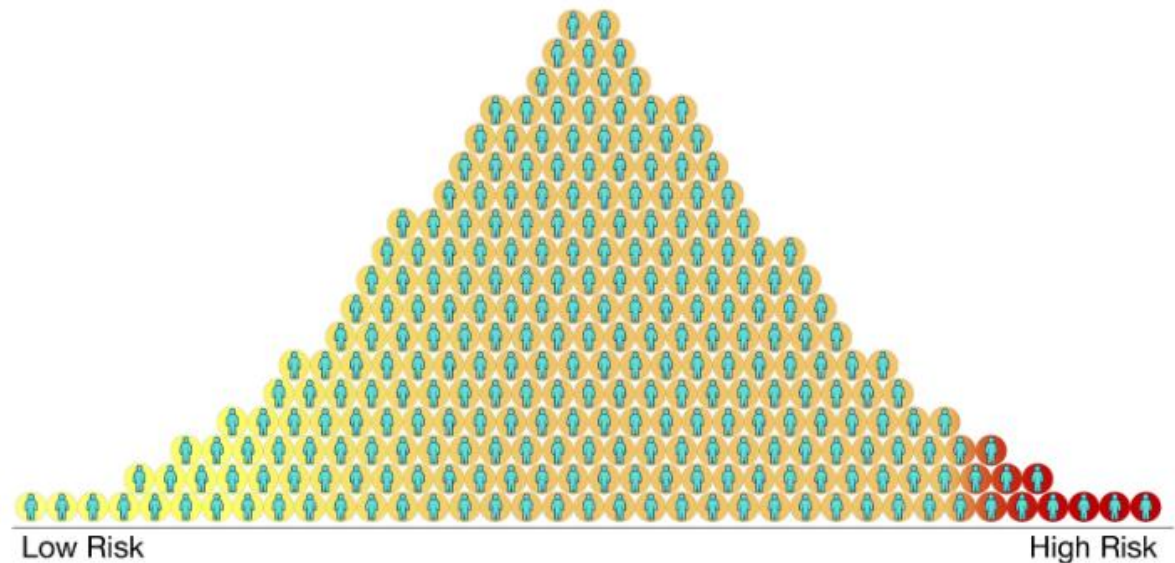
§Validamos que la región MGAT5 es distinta del gen de la lactasa (LCT) en el cromosoma 2, ajustando el punto máximo en el gen LCT (rs55634455i, $P=4,77 \times 10^{-4}$, OR=0,87). Después de este ajuste, los SNP superiores seguían siendo significativos y se vieron mínimamente afectados por el ajuste (rs2460382d: $P=5,17 \times 10^{-3}$, OR=0,91; rs10496726: $P=6,65 \times 10^{-5}$, OR=0,86).

Las probabilidades de desarrollar lupus son 1,8 veces mayores para aquellos con el alelo T.

Puntuaciones poligénicas (riesgo)

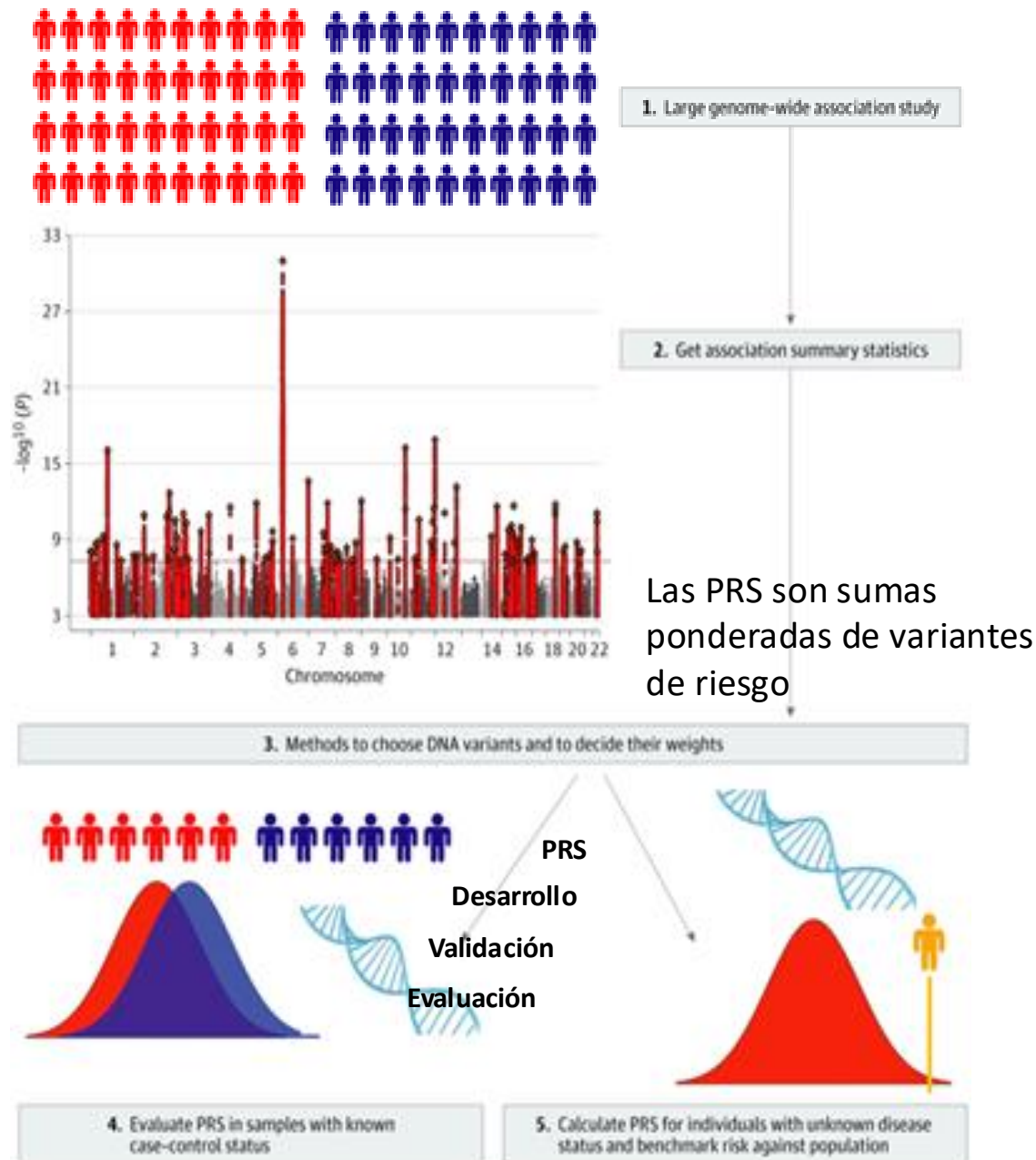
Puntuación de riesgo poligénico (PRS)

La puntuación de riesgo poligénico (PRS) o puntuación poligénica (PGS) es una estimación numérica basada en la presencia de múltiples variantes de riesgo genómico y sus estimaciones de riesgo asociadas (razón de momios, riesgos relativos, etc.) para predecir el riesgo de que una persona desarrolle una enfermedad.



- **Una puntuación de riesgo poligénico sólo puede explicar el riesgo relativo de una enfermedad, no el riesgo absoluto.**
- **Las puntuaciones de riesgo poligénico son probabilidades, no certezas.**

Construcción de una puntuación poligénica



Para predecir el resultado de interés, la PRS se puede combinar con variables de riesgo no genéticas (por ejemplo, edad, sexo, ascendencia y variables clínicas).

Wand *et al.* Improving reporting standards for polygenic scores in risk prediction studies [Mejorar los estándares de informes para puntuaciones poligénicas en estudios de predicción de riesgos]. Nature 2021.

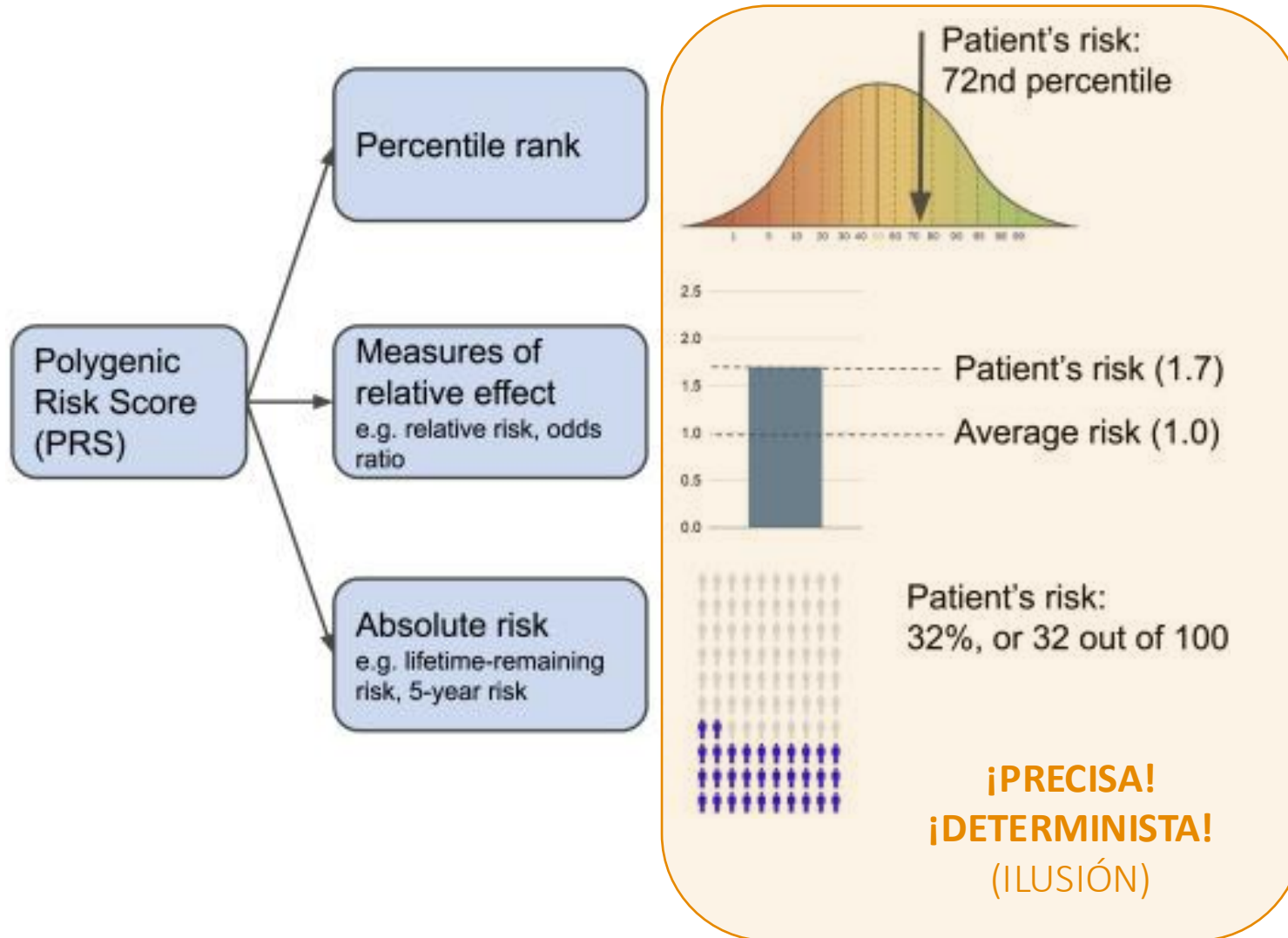
Wray *et al.* From Basic Science to Clinical Application of Polygenic Risk Scores [De la ciencia básica a la aplicación clínica de las puntuaciones de riesgo poligénico]. JAMA Psychiatry 2021.

Beneficios potenciales de las puntuaciones poligénicas

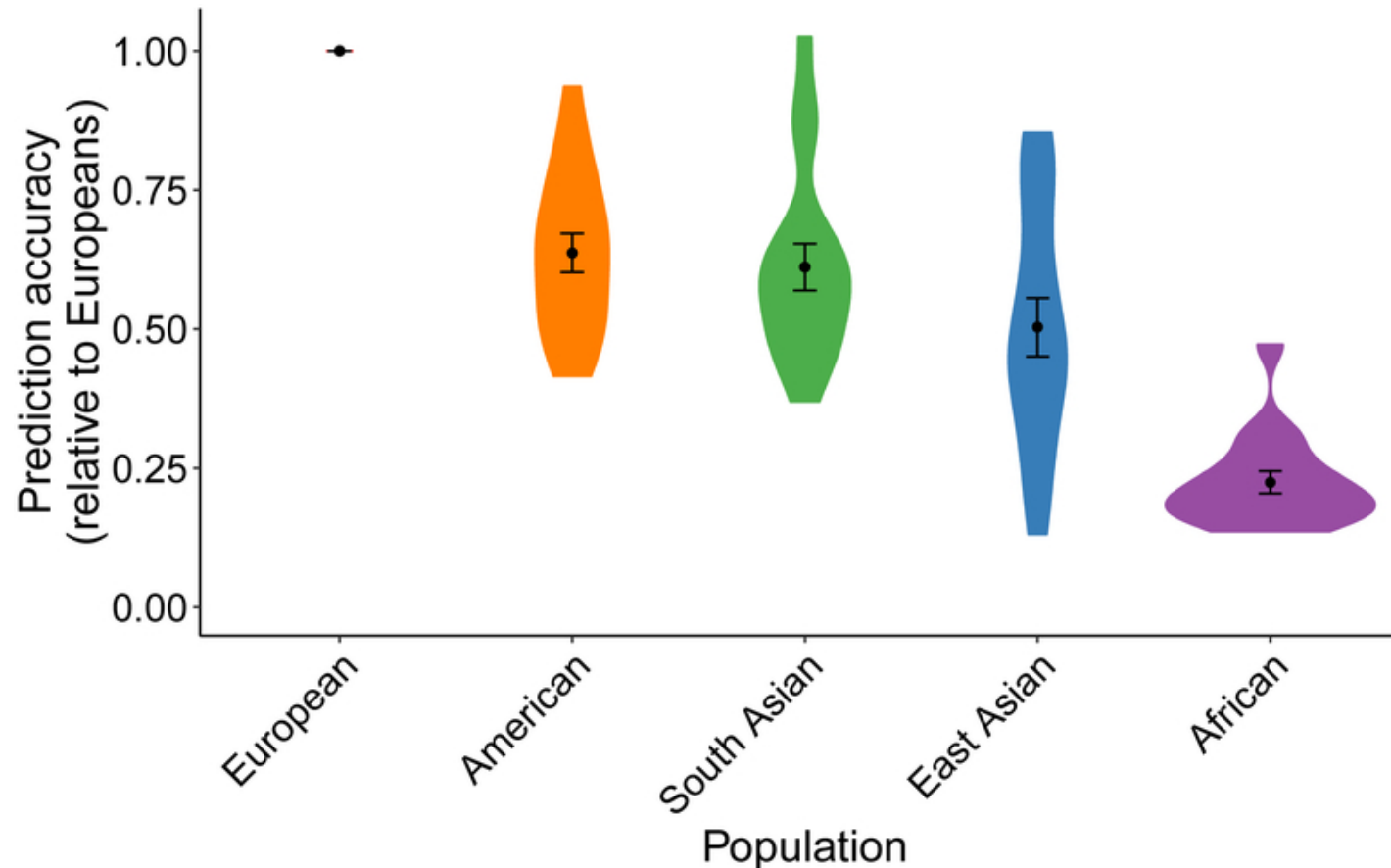
- Predicción del riesgo de enfermedad
- Mejorar los diagnósticos
- Retrasar la progresión y recurrencia de la enfermedad
- Mejorar la detección poblacional
- Fomentar conductas que reduzcan el riesgo
- Mejorar el costo y eficiencia de los ensayos

Grupo de trabajo sobre puntuación de riesgo poligénico de la Alianza Internacional de Enfermedades Comunes. Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps [Uso responsable de las puntuaciones de riesgo poligénico en la clínica: beneficios potenciales, riesgos y brechas]. Nat Med 2021.

Tres opciones para representar la información de riesgo de una PRS



Precisión de la predicción de la PRS en diferentes ascendencias

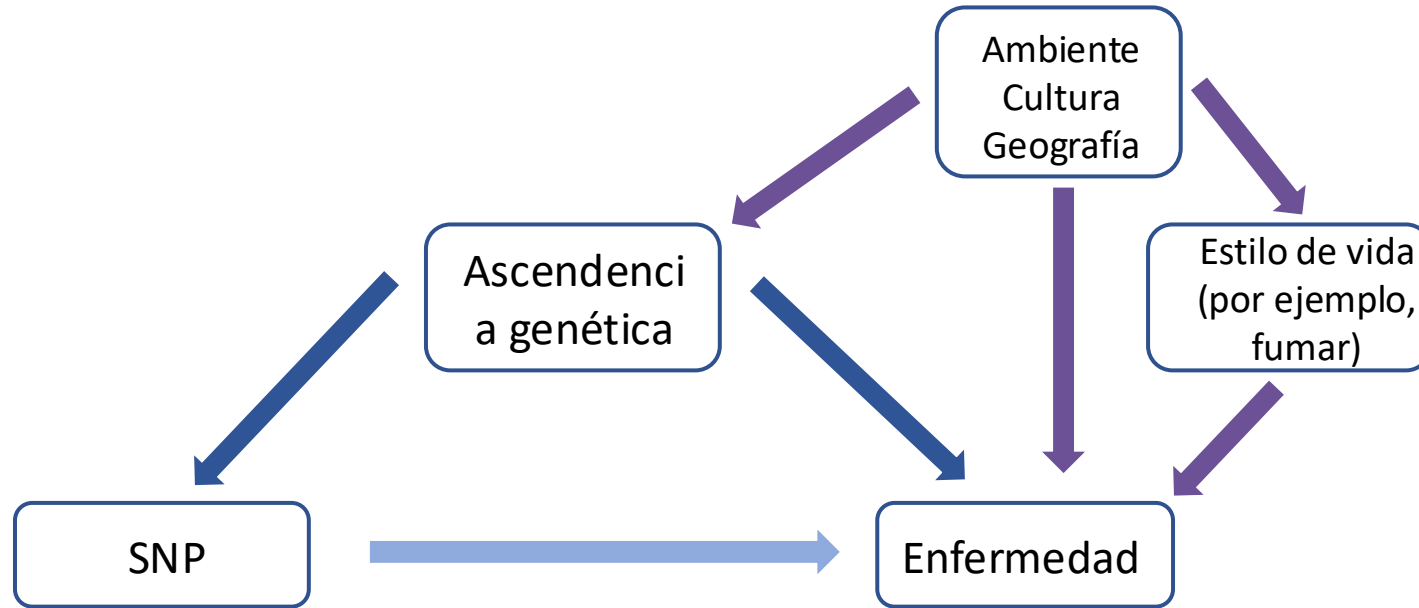


Precisión de la predicción relativa a individuos de ascendencia europea en 17 rasgos cuantitativos y 5 poblaciones continentales en UKBB

➔ Las PRS capturan una pequeña fracción de la varianza genética de un rasgo y tienen una precisión predictiva variable según la población

Martin *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities [El uso clínico de las puntuaciones de riesgo poligénico actuales puede exacerbar las disparidades en la salud.]. Nat Genet 2019;51:584-591.

Puntuaciones de riesgo de enfermedad



El desarrollo de modelos de riesgo es complejo y requiere conjuntos de datos diversos.

➔ Es necesario incluir el “medio ambiente” a medida que pasamos de modelos de riesgo genético a modelos de riesgo integrados para la salud de precisión.

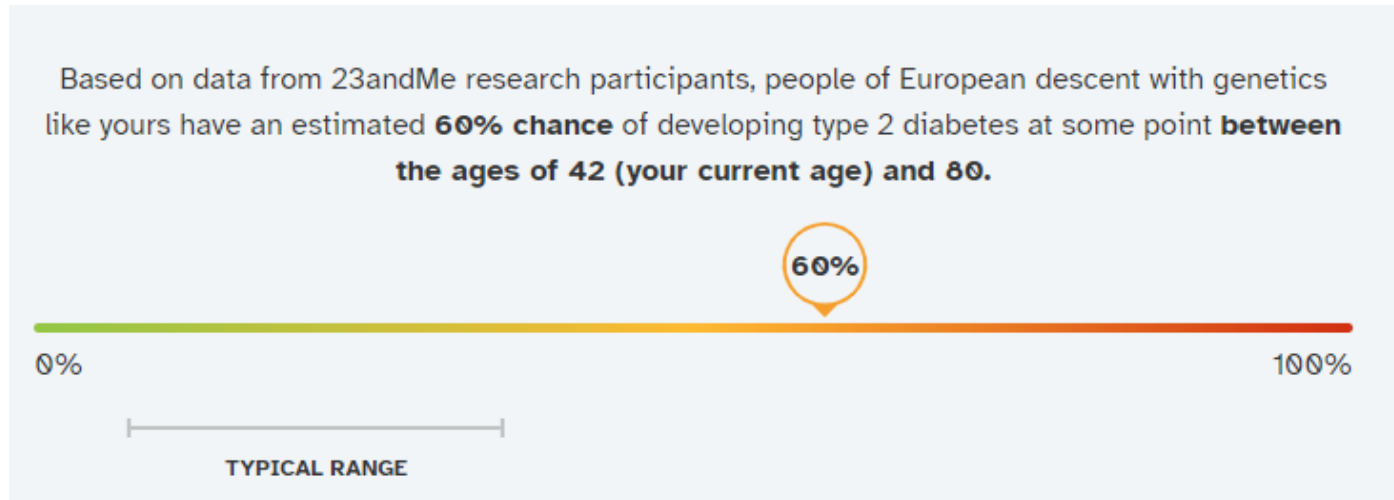
Riesgos de las puntuaciones poligénicas

- Resultados “incorrectos” o “falsos positivos”
- Sesgos de representación (por ejemplo, demografía)
- Fallos en la transmisión de incertidumbre (falso determinismo genético)
- Comunicación de resultados
- Efectos psicosociales (por ejemplo, ansiedad y depresión)
- Discriminación y aumento de las disparidades en materia de salud
- Usos inapropiados (por ejemplo, selección de embriones)

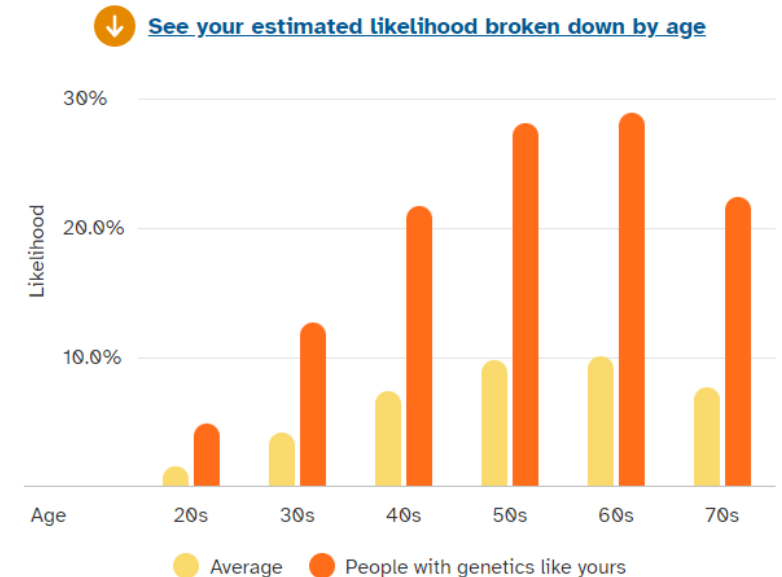
¿Preguntas?



Caso de uso: Diabetes tipo 2 (DT2)



- Pruebas directas al consumidor: 23andMe
- Prueba PRS para la diabetes:
 - 1166 centros distintos
 - El modelo se basa en datos de 720.000 participantes de investigación de ascendencia europea
 - Informe limitado por la FDA, por lo que solo puede proporcionar sugerencias de bienestar



The chances of developing type 2 diabetes are greatest during midlife. This graph shows the likelihood of developing type 2 diabetes at different ages for people with genetics like yours, compared to the average likelihood. These estimates are based on data from 23andMe research participants of European descent.

Otros factores: edad, antecedentes familiares, origen étnico, prediabetes, diabetes gestacional, exceso de peso, inactividad física, dieta y tabaquismo

Diabetes tipo 2 – 23andMe

- La puntuación de riesgo se basa en datos de investigación de 23andMe basados en autoinformes combinados con datos de los CDC sobre la probabilidad de desarrollar la enfermedad con ascendencia europea.
- Se creó un modelo de regresión logística utilizando la diabetes tipo 2 como respuesta, las variantes genéticas como predictores y la edad, el sexo, los primeros cinco componentes genéticos principales y la fecha de genotipificación como covariables.

https://permalinks.23andme.com/pdf/23-12_predictivemodel_methodology_02oct2015.pdf
https://permalinks.23andme.com/pdf/23_19-Type2Diabetes_March2019.pdf

Performance across ethnicities

We evaluated model performance for people of European, African, Hispanic/Latino, East Asian, and South Asian descent. This analysis included data from 8,000 or more research participants of each of these ethnicities. The predictive power of the model (AUC) varies across ethnicities, possibly due to factors like limitations in the amount of data available from each ethnicity.

ETHNICITY	AUC VALUE
European	0.645
South Asian	0.611
Hispanic/Latino	0.625
East Asian	0.621
African	0.578

The "Area Under the receiver operating characteristic Curve" (AUC or AUROC) measures how well a statistical model predicts whether or not people have a trait. AUC values usually range from 0.5 to 1, where higher numbers mean the model has more predictive power. The predictive power of the type 2 diabetes model is limited by the fact that type 2 diabetes depends not only on the genetic factors included in this model, but also on non-genetic factors and likely on additional genetic factors that are not yet known.

Controversia

Ventajas

- Incorpora la variación genética en la etiología de la enfermedad
- Crear una puntuación de riesgo para un rasgo o enfermedad poligénica compleja que puede no tener un solo gen para analizar
- Puede haber intervenciones tempranas o específicas
- Útil para la estratificación de riesgos

Desventajas

- Predictor débil de enfermedad
- Determina los niveles de falsos positivos o falsos negativos dentro de las puntuaciones de riesgo
- Intervenciones costosas o peligrosas
- Estrés del consumidor por los resultados
- Variación en las puntuaciones de riesgo utilizando diferentes métodos

¿Preguntas?

ramosp@musc.edu