Usages et enjeux en génétique humaine:

Prédiction de risques pour les maladies multifactorielles

Anne-Louise Leutenegger

Inserm U1141, NeuroDiderot-GenMedStroke

anne-louise.leutenegger@inserm.fr

Instituts thématiques

Inserm

Institut national de la santé et de la recherche médicale

Un siècle de Fisher – Septembre 2019

Université de Paris

Prédiction de risques pour les maladies multifactorielles

- Pour quoi faire ?
  - »Identifier les personnes les plus à risque pour leur proposer un suivi personnalisé
    - Modifications du mode de vie
    - Thérapies préventives (préemptives)

- Comment ?
  - »Calcul d'un score à partir des données génétiques d'une personne
    - Marqueurs génétiques (SNPs) utilisés dans les GWAS
  - »Score de risque polygénique
    - « Polygenic risk score » (PRS)
    - « Genome-wide polygenic risk score » (GPS)

Prédiction de risques pour les maladies multifactorielles

Prediction of individual genetic risk to disease from genome-wide association studies

Naomi R. Wray,[1,4] Michael E. Goddard,[2,3] and Peter M. Visscher[1]

Genome Research 17:1520–1528 ©2007

Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores

Bjarni J. Vilhjálmsson,[1,2,3,4,*] Jian Yang,[5,6] Hilary K. Finucane,[1,2,3,7] Alexander Gusev,[1,2,3] ... Shaun Purcell,[3,13,17,18] Daniel Chasman,[22,24] Benjamin Neale,[3,8] Michael Goddard,[25,26] Peter M. Visscher,[5,6] Peter Kraft,[1,2,3,27] Nick Patterson,[3] and Alkes L. Price[1,2,3,27,*]

The American Journal of Human Genetics *97*, 576–592, October 1, 2015

Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations

Amit V. Khera[1,2,3,4,5], Mark Chaffin[4,5], Krishna G. Aragam[1,2,3,4], Mary E. Haas[4], Carolina Roselli[4], Seung Hoan Choi[4], Pradeep Natarajan[2,3,4], Eric S. Lander[4], Steven A. Lubitz[2,3,4], Patrick T. Ellinor[2,3,4] and Sekar Kathiresan[1,2,3,4*]

NATURE GENETICS | VOL 50 | SEPTEMBER 2018 | 1219–1224 |

23andMe®

Depuis 2007

Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes

Nasim Mavaddat,[1,*] Kyriaki Michailidou,[1,2] Joe Dennis,[1] Michael Lush,[1] Laura Fachal,[3] Andrew Lee,[1] ... Nilanjan Chatterjee,[6,208,209] Peter Kraft,[68,109] Montserrat García-Closas,[6] Jacques Simard,[62] and Douglas F. Easton[1,3]

The American Journal of Human Genetics *104*, 21–34, January 3, 2019

PREDILIFE prédire pour prévenir

communiqué de presse institutCurie

Paris, France, 06 Mai 2019 – PrediLife (Euronext Growth Paris : ALPRE), société développant des solutions innovantes de prédiction de risque de pathologies pour une médecine personnalisée, annonce sa collaboration avec l'Institut Curie dans le cadre du lancement de la nouvelle génération de MammoRisk®, un test de prédiction de risque de cancer du sein qui intègre maintenant un score polygénique issu d'un test salivaire.

Score de risque polygénique (PRS)

- Calcul du PRS d'une personne
  - »Score pour un marqueur génétique (SNP): un poids multiplié par le nombre d'allèles à risque portés par l'individu (0, 1 ou 2)
    - Poids: force de l'association (effet du SNP) mesurée dans les études pangénomiques (GWAS)
  - »PRS: la somme des scores de chaque marqueur

$$PRS = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \ldots + \beta_n x_n$$

where β_k is the per-allele log odds ratio (OR) for breast cancer associated with SNP k, x_k is the allele dosage for SNP k, and n is the total number of SNPs included in the PRS.

Mavaddat et al, AJHG, 2019

Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations

Amit V. Khera[1,2,3,4,5], Mark Chaffin[4,5], Krishna G. Aragam[1,2,3,4], Mary E. Haas[4], Carolina Roselli[4], Seung Hoan Choi[4], Pradeep Natarajan[2,3,4], Eric S. Lander[4], Steven A. Lubitz[2,3,4], Patrick T. Ellinor[2,3,4] and Sekar Kathiresan[1,2,3,4*]

NATURE GENETICS | VOL 50 | SEPTEMBER 2018 | 1219–1224 |

- Cinq maladies multifactorielles étudiées
- Plusieurs algorithmes de calcul du PRS
  - »GWAS: choix des SNPs (~6 millions) et estimation de leur effet
- Cohorte UK Biobank: 500 000 personnes 40-69 ans
- Au total, ~20 % de la population étudiée avait un risque multiplié par 3 de développer au moins l'une des cinq maladies
  - » insuffisance coronarienne (8%), fibrillation auriculaire (6,1%), diabète de type 2 (3,5%), maladies inflammatoires chroniques de l'intestin (MICI) (3,2%), cancer du sein (1,5%)
- *Le moment est venu d'envisager l'inclusion du PRS dans la pratique clinique*

Réactions dans la communauté génétique

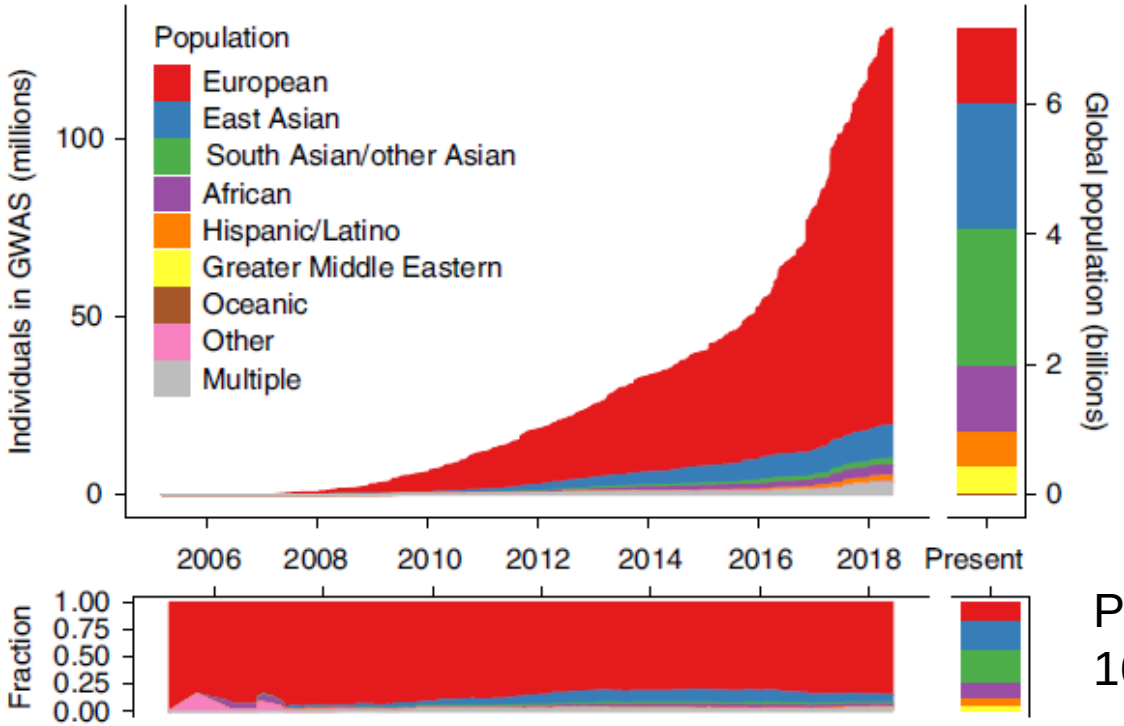
- Des réactions positives « great leap forward »

- Des critiques éthiques
  - »Qualité inégales des prédictions selon les origines des personnes
  - »Son utilisation abusive, par exemple pour prédire « l'intelligence »
  - »Communication des PRS aux personnes, information complexe

- Des critiques méthodologiques …
  - »Algorithme de calcul (choix SNPs, calcul des poids)
  - »Prédiction statistique sans compréhension de la biologie: SNPs associés et non pas les facteurs génétiques causaux

… mais rarement sur les hypothèses sous-jacentes du modèle polygénique

Manque de représentativité des populations étudiées en GWAS



GWAS:
79% EUR

Population mondiale:
16% EUR

Fig. 1 | Ancestry of GWAS participants over time, as compared with the global population. Cumulative data, as reported by the GWAS catalog[76]. Individuals whose ancestry is 'not reported' are not shown.

- GWAS biaisés vers les Européens

Martin et al, Nature Genetics, 2019

PRS pour des individus d'origine non-Européenne

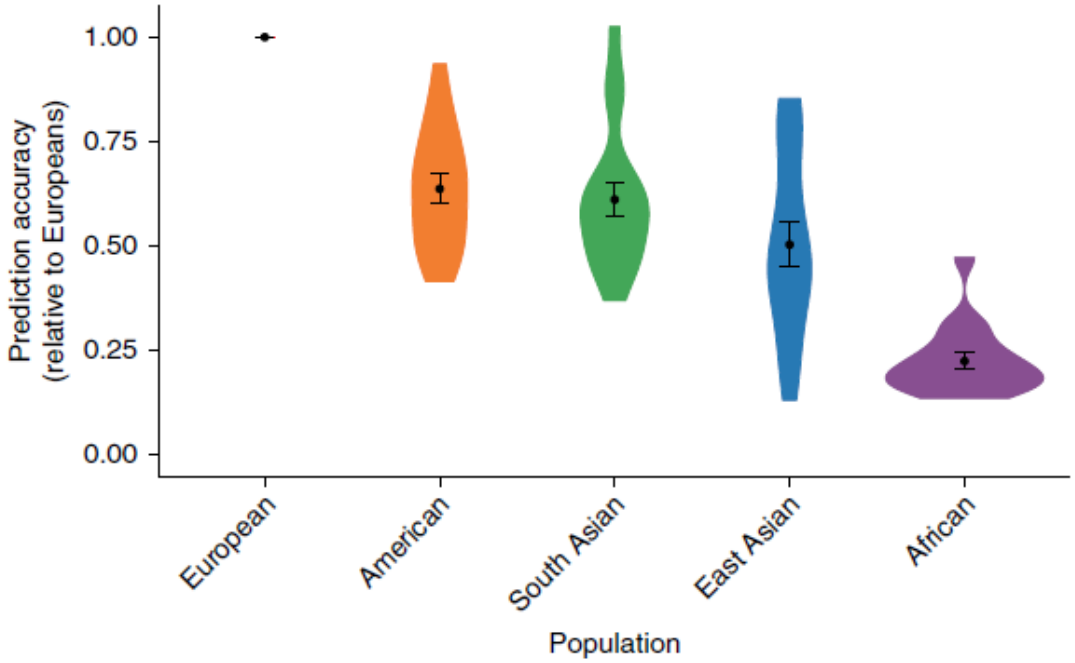


Fig. 3 | Prediction accuracy relative to European-ancestry individuals across 17 quantitative traits and 5 continental populations in the UKBB. All

- Sélection des SNPs et poids viennent des GWAS
- Qualité de prédiction décroit avec la divergence génétique entre populations

Martin et al, Nature Genetics, 2019

Prédire « l'intelligence » : niveau d'éducation

Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals

James J. Lee[1,58], Robbee Wedow[2,3,4,58], Aysu Okbay[5,6,58]*, Edward Kong[7], Omeed Maghzian[7], Patrick Turley[16,17,59], Peter M. Visscher[10,54,59]*, Daniel J. Benjamin[9,47,56,59]* and David Cesarini[47,51,57,59]

NATURE GENETICS | VOL 50 | AUGUST 2018 | 1112–1121 |

The new genetics of intelligence

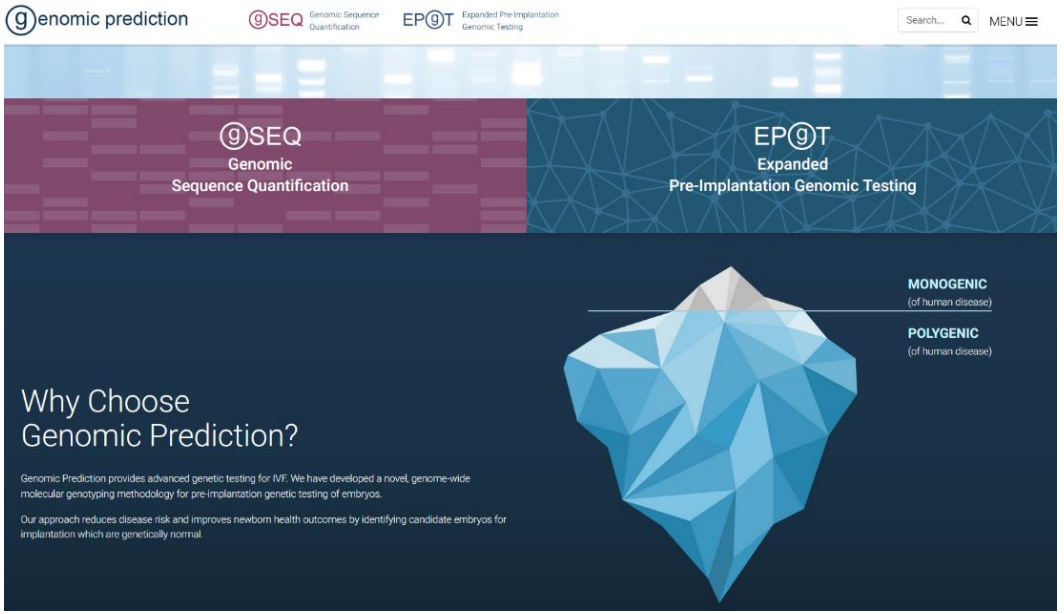
Robert Plomin[1] and Sophie von Stumm[2]

NATURE REVIEWS | GENETICS
VOLUME 19 | MARCH 2018 | 149

Depuis Septembre 2018, PGT-P (Polygenic Disorders):

- Type 1 and Type 2 Diabetes
- Coronary Artery Disease, Heart Attack Risk, Hypercholesterolemia, Hypertension
- Breast Cancer, Testicular Cancer, Prostate Cancer, Malignant Melanoma, Basal Cell Carcinoma
- Intellectual Disability
- Idiopathic Short Stature

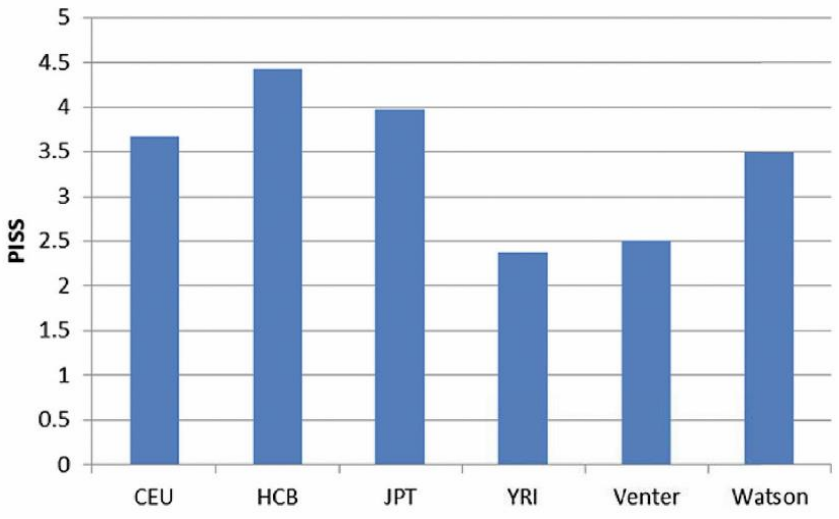
The Genomic Prediction team comprises scientists at the forefront of polygenic trait prediction in human adults. Now, they are directing their proven methodology towards the human embryo.



Réaction des sociétés savantes ?

Prédire « l'intelligence » : niveau d'éducation

- Neil Risch, discours présidentiel ASHG 2015
  - »Polygenic IQ SNP Score (PISS) avec 7 SNPs les plus associés dans 4 populations: Africains les plus faibles
  - »Ajouter un prix Nobel (James Watson) et un chercheur pionnier du séquençage du génome humain (Craig Venter): très moyens !



Prédire « l'intelligence » : niveau d'éducation

- Neil Risch, discours présidentiel ASHG 2015
  »PISS

- 2019 ASHG/ESHG session: « Potential Policy Implications of Genetic Research into Educational Attainment »

Prédire « l'intelligence » : niveau d'éducation

Recent genome-wide association studies (Lee et al., Nature Genetics, 2018) have shown that polygenic scores composed of common variants can explain ~13% of the variance in educational attainment, and this number is set to increase as studies grow. While there has been considerable discussion in policy circles and the media about the ethics of gene editing, the implications of genetic prediction for behavioral and cognitive traits have received less attention. There is an urgent need to consider whether and how the findings from these studies should be applied in public policy. For example, should polygenic scores for educational attainment be used as a component of school admissions processes, or to identify children likely to need particular educational help? What threshold of precision would be necessary for useful deployment? Should we allow antenatal embryo selection based on these scores in IVF clinics? Should insurers be allowed to use polygenic scores in determining policy pricing? What are the potential unintended consequences of such policies, and what does the public think about them? How does knowledge about the heritability of educational attainment affect political notions of equality? How can we stop this debate being hijacked by extremist groups? In this session, genetics and policy experts will present the latest research on these topics and hold a panel discussion.

Information aux personnes

- Que dire à une personne avec PRS élevé mais pas d'autres facteurs de risque? Quelle intervention proposée?
  Et inversement?

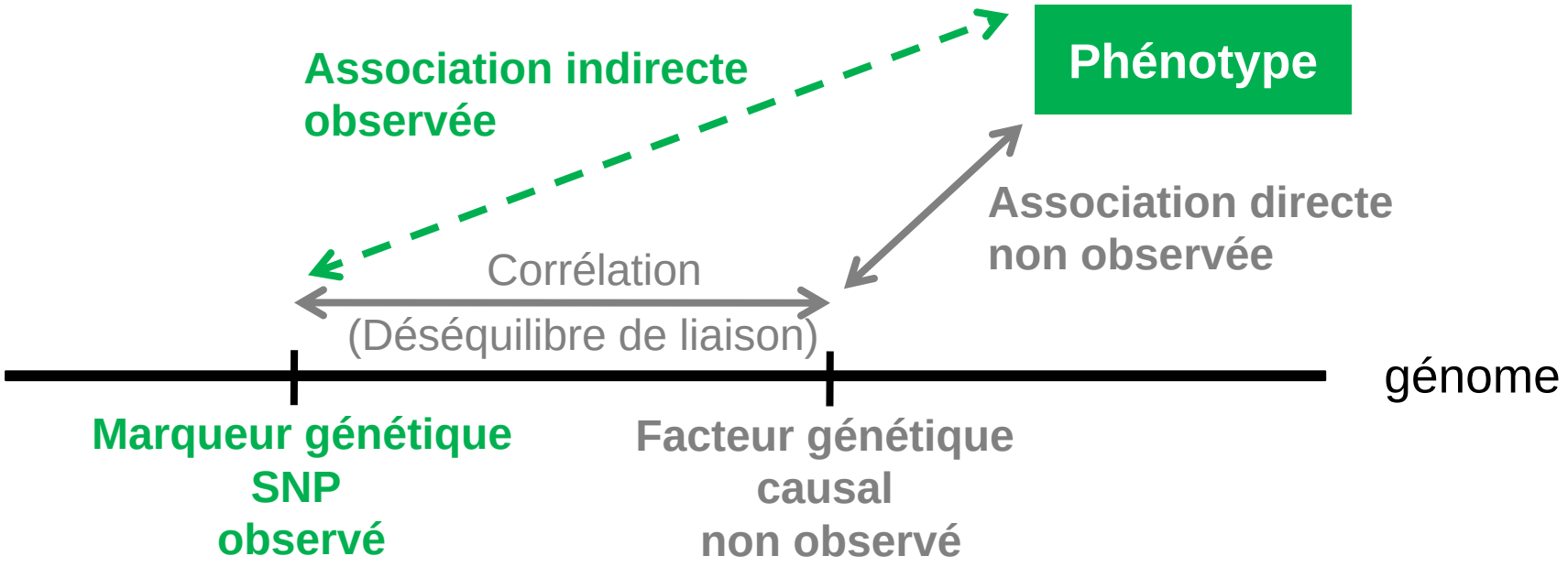
- Est-ce que PRS sera suffisant pour motiver une personne à changer de mode de vie ou à prendre un médicament sans signe de maladie?

Algorithme de calcul

- Tendance à inclure tous les SNPs, mais certains avec un poids proche de zéro
  - »Que veut alors dire allèle à risque?
  - »Ces SNPs changent-ils vraiment les prédictions d'une personne?
- Fausse impression qu'un PRS sur des milliers de SNPs est plus solide, qu'on a une meilleure connaissance des bases génétiques de la maladie

- Algorithmes propriétaires des entreprises privées: secret industriel
  - »23andMe: fourni à la FDA en 2017, site web white paper Avril 2019

SNPs ≠ facteurs génétiques causaux directement impliqués dans la maladie



Etude d'association pangénomique
« Genome-wide association study » (GWAS)

Hypothèses sous jacentes au PRS: modèle polygénique

- Nombreux facteurs génétiques avec un effet faible
- Effet additif des allèles aux facteurs génétiques
  - »Pas de dominance
- Effet additif des facteurs génétiques
  - »Pas d'interaction GxG
- Pas d'interaction entre les facteurs génétiques et l'environnement (GxE)

Rarement remises en cause

Modèle polygénique

$$PRS = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \ldots + \beta_n x_n$$

where β_k is the per-allele log odds ratio (OR) for breast cancer associated with SNP k, x_k is the allele dosage for SNP k, and n is the total number of SNPs included in the PRS. Previous analyses found no evidence for statistically significant interactions between SNPs[19,20] and little evidence for departures from a log-additive model for individual SNPs. Assuming this is true in general, the PRS summarizes efficiently the combined effects of SNPs on disease risk.

Mavaddat et al, AJHG, 2019

- Peut-on considérer qu'il est valide juste parce qu'on n'a pas pu le rejeter sur la base des SNPs?

Maladie cœliaque

maladie auto-immune induite par l'ingestion de gluten

- Association avec hétérodimère HLA-DQ2 (=DQA05 et DQB02), connue depuis 1989
- GWAS: SNP rs2187668 (Van Heel et al, 2007)
  - »SNP en fort déséquilibre de liaison avec DQA05-DQB02 en cis, mais pas en trans
- Conséquences de l'utilisation du SNP:
  - »Mauvaise mesure de l'association avec HLA-DQ2 (poids PRS)
  - »Mauvaise classification des individus en terme de risque
  - »Interaction entre gènes DQA et DQB non détectées

Adaptée de F Clerget-Darpoux

Conclusion

- Etudes d'association pangénomique ont mis en évidence de nombreux SNPs
  mais sans comprendre leur fonctionnement exact

- Modèle polygénique

- N'a-t-on pas besoin de mieux comprendre la relation génotype-phénotype pour pouvoir prédire les risques de maladies multifactorielles?

« All models are wrong but some are useful.

….

However, the approximate nature of the model must always be borne in mind »

(Georges Box, 1987)

Extra slides

Khera et al 2018

Table 1 | GPS derivation and testing for five common, complex diseases

Disease	Discovery GWAS (*n*)	Prevalence in validation dataset	Prevalence in testing dataset	Polymorphisms in GPS	Tuning parameter	AUC (95% CI) in validation dataset	AUC (95% CI) in testing dataset
CAD	60,801 cases; 123,504 controls[16]	3,963/120,280 (3.4%)	8,676/288,978 (3.0%)	6,630,150	LDPred ($\rho = 0.001$)	0.81 (0.80–0.81)	0.81 (0.81–0.81)
Atrial fibrillation	17,931 cases; 115,142 controls[30]	2,024/120,280 (1.7%)	4,576/288,978 (1.6%)	6,730,541	LDPred ($\rho = 0.003$)	0.77 (0.76–0.78)	0.77 (0.76–0.77)
Type 2 diabetes	26,676 cases; 132,532 controls[31]	2,785/120,280 (2.4%)	5,853/288,978 (2.0%)	6,917,436	LDPred ($\rho = 0.01$)	0.72 (0.72–0.73)	0.73 (0.72–0.73)
Inflammatory bowel disease	12,882 cases; 21,770 controls[32]	1,360/120,280 (1.1%)	3,102/288,978 (1.1%)	6,907,112	LDPred ($\rho = 0.1$)	0.63 (0.62–0.65)	0.63 (0.62–0.64)
Breast cancer	122,977 cases; 105,974 controls[33]	2,576/63,347 (4.1%)	6,586/157,895 (4.2%)	5,218	Pruning and thresholding ($r^2 < 0.2$; $P < 5 \times 10^{-4}$)	0.68 (0.67–0.69)	0.69 (0.68–0.69)

AUC was determined using a logistic regression model adjusted for age, sex, genotyping array, and the first four principal components of ancestry. The breast cancer analysis was restricted to female participants. For the LDPred algorithm, the tuning parameter ρ reflects the proportion of polymorphisms assumed to be causal for the disease. For the pruning and thresholding strategy, r^2 reflects the degree of independence from other variants in the linkage disequilibrium reference panel, and *P* reflects the *P* value noted for a given variant in the discovery GWAS. CI, confidence interval.

Janssens and Joyner, 2019

Table 1. Comparison between PRSs based on genome-wide significant SNPs and 30 alternatives based on up to 7.2 million SNPs.[a]

	Number of SNPs included		AUC			Number of PRS (out of 30) with	
	Only genome-wide significant	PRS with highest AUC	Only genome-wide significant	PRS with highest AUC	ΔAUC	ΔAUC < 0	ΔAUC < 0.01
Coronary artery disease	74	6 629 369 (ρ = 0.1%)	0.791	0.806	0.015	2	27
Atrial fibrillation	55	6 705 798 (ρ = 0.3%)	0.766	0.773	0.007	21	30
Type 2 diabetes	72	6 893 037 (ρ = 1%)	0.700	0.725	0.025	7	25
Inflammatory bowel disease	288	6 882 324 (ρ = 10%)	0.614	0.633	0.019	19	23
Breast cancer	572	5158	0.677	0.685	0.008	19	30

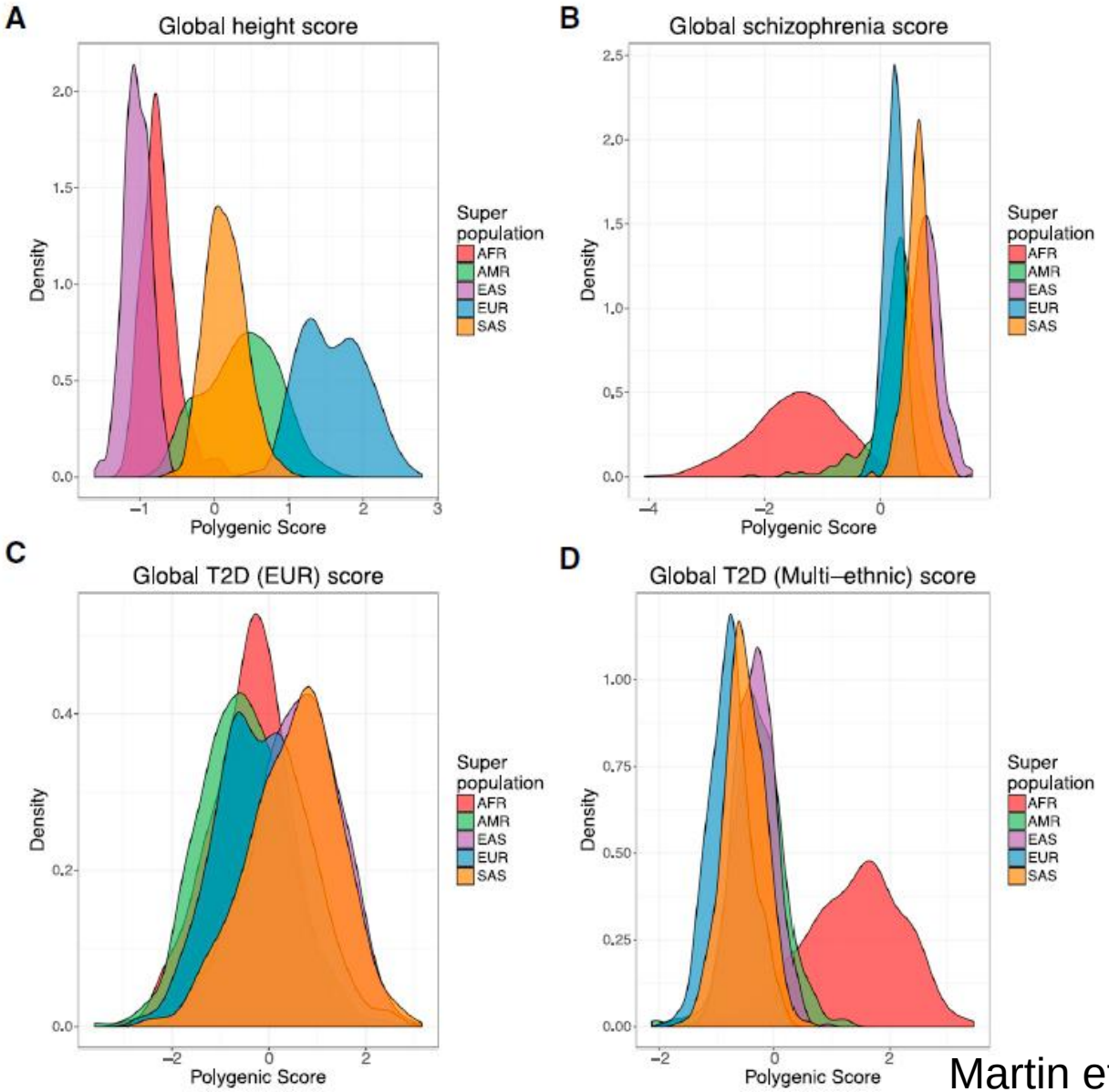
[a] Table is based on Supplementary Tables 1–5 from Khera et al. *(4)*. ρ, percentage of SNPs expected to have nonzero effects.

Martin et al 2019

Supplementary Table 7. Numbers of individuals with a given ancestry in UK Biobank. Target individuals describes the numbers of individuals used in **Figure 3**.

Super population	Total	Unrelated	Target individuals	Neale Lab GWAS
EUR	447206	370407	5000	350326
SAS	9950	9015	9015	0
AFR	9288	8503	8503	0
AMR	4724	4329	4329	1
EAS	2421	2306	2306	0
other	14788	12252	N/A	10867
TOTAL	488377	406812	29153	361194

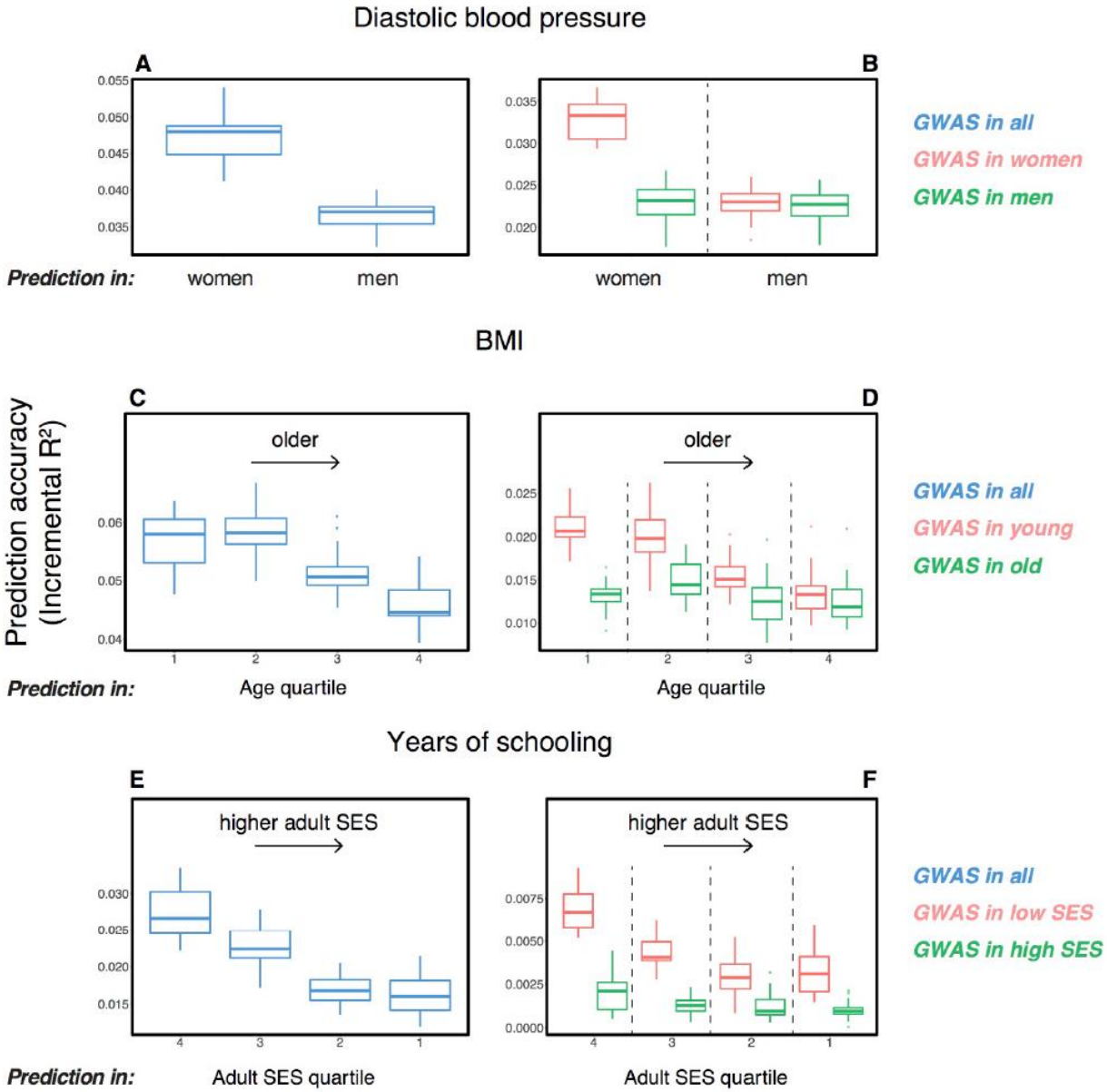
PRS dans différentes populations



Martin et al, AJHG, 2017

Mostafavi et al, biorxiv, 2019

Même au sein d'une même population



Choix des marqueurs et des poids?

- Choix des marqueurs
  - »SNPs significatifs des GWAS
  - »Génome entier (Vilhjalmsson et al 2015, LDPred)
- Choix des poids
  - »Basés sur les résultats des GWAS, SNP individuel
  - »Estimés conjointement sur tous les SNPs sélectionnés (Privé et al 2019; bigstatsr)

Genomic Prediction

PGT-P (Polygenic Disorders):

- Type 1 and Type 2 Diabetes
- Coronary Artery Disease, Heart Attack Risk, Hypercholesterolemia, Hypertension
- Breast Cancer, Testicular Cancer, Prostate Cancer, Malignant Melanoma, Basal Cell Carcinoma
- Intellectual Disability
- Idiopathic Short Stature

What is PGT-P?

- PGT-P stands for Preimplantation Genetic Testing for Polygenic disorders.
- Polygenic disorders are caused by variants in multiple genes.

Is PGT-P a diagnostic test?

- No. PGT-P is for assessment of risk, NOT a diagnostic test, embryos cannot be diagnosed. Furthermore, there is a risk of discarding embryos that may have led to healthy outcomes with regards to the polygenic disease(s) tested through this methodology. Conversely, disease risk is not eliminated in embryos designated as “normal risk” and embryos chosen for transfer may ultimately develop into individuals with the condition(s) that was screened for using PGT-P.

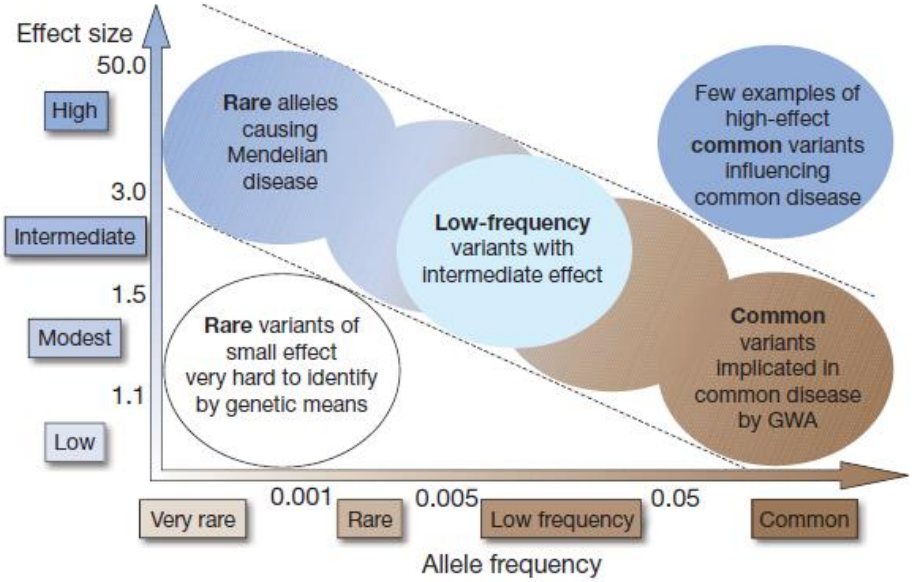
Does a high risk for PGT-P translate into “Not Suitable for Transfer?” If not, what is it useful for?

- A recommendation for or against transfer is not provided for polygenic conditions. We believe reproductive decisions surrounding PGT-P results should be determined by the patient and the physician, with the benefit of a Genetic Counselor.
- Identifying embryos that are at high risk may help some families prioritize which embryos they would choose for a transfer cycle, and in which sequence of preference, similar to how embryo morphology is being used currently. Furthermore, knowing the profile of the implanted embryo may help inform medical management in the future.
- Independent third party estimates as to the risk reduction may be found here.
  - » Screening human embryos for polygenic traits has limited utility, Karavani et al, 2019
  - » Given currently available polygenic predictors and typical IVF yields, the average gain due to selection would be≈2.5cmif selecting for height, and ≈2.5IQ (intelligence quotient) points if selecting for cognitive function.

ASHG/ESHG Building Bridges: Potential Policy Implications of Genetic Research into Educational Attainment

Recent genome-wide association studies (Lee et al., Nature Genetics, 2018) have shown that polygenic scores composed of common variants can explain ~13% of the variance in educational attainment, and this number is set to increase as studies grow. While there has been considerable discussion in policy circles and the media about the ethics of gene editing, the implications of genetic prediction for behavioral and cognitive traits have received less attention. There is an urgent need to consider whether and how the findings from these studies should be applied in public policy. For example, should polygenic scores for educational attainment be used as a component of school admissions processes, or to identify children likely to need particular educational help? What threshold of precision would be necessary for useful deployment? Should we allow antenatal embryo selection based on these scores in IVF clinics? Should insurers be allowed to use polygenic scores in determining policy pricing? What are the potential unintended consequences of such policies, and what does the public think about them? How does knowledge about the heritability of educational attainment affect political notions of equality? How can we stop this debate being hijacked by extremist groups? In this session, genetics and policy experts will present the latest research on these topics and hold a panel discussion.

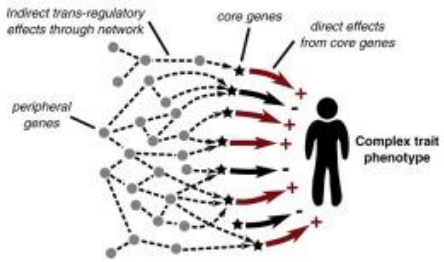
1:00 PM **Introduction.** A.I. Young. Univ Oxford, UK.
1:10 PM **Polygenic prediction of education: U.S. attitudes regarding when and how it should be deployed.** D. Conley. Princeton Univ.
1:20 PM **Genetics of education and egalitarian policy goals.** K.P. Harden. Univ Texas Austin.
1:30 PM **Educational reform, ability, and labor market screening.** A. Okbay. Vrije Univ Amsterdam, Netherlands.
1:40 PM **Lost in translation: Bias and methodological challenges in policy applications of polygenic educational scores.** M. Mills. Univ Oxford, UK.
1:50 PM **Panel discussion.**



Fom Manolio et al, Nature 2009

Question: Why is the architecture of complex traits dominated by huge numbers of small effect variants?

Approach: We built a quantitative phenotype model based on core gene expression



$$\mathrm{Var}(Y_i) = \underbrace{\sum_{j=1}^{M} \gamma_j^2 V_{j,\mathrm{cis}}}_{M \text{ core terms}} + \underbrace{\sum_{j=1}^{M} \gamma_j^2 V_{j,\mathrm{trans}}}_{M \text{ trans terms}} + \underbrace{\sum_{j=1}^{M} \sum_{k=1}^{j-1} 2\gamma_j \gamma_k C_{j,k}}_{M^2-M \text{ covariance terms}}$$

Conclusion: Most of the trait heritability is explained by many small *trans*-regulatory effects from peripheral genes

Trans Effects on Gene Expression Can Drive Omnigenic Inheritance

Xuanyao Liu,[1,*] Yang I. Li,[1,2,*] and Jonathan K. Pritchard[3,4,*]

Cell *177*, 1022–1034, May 2, 2019