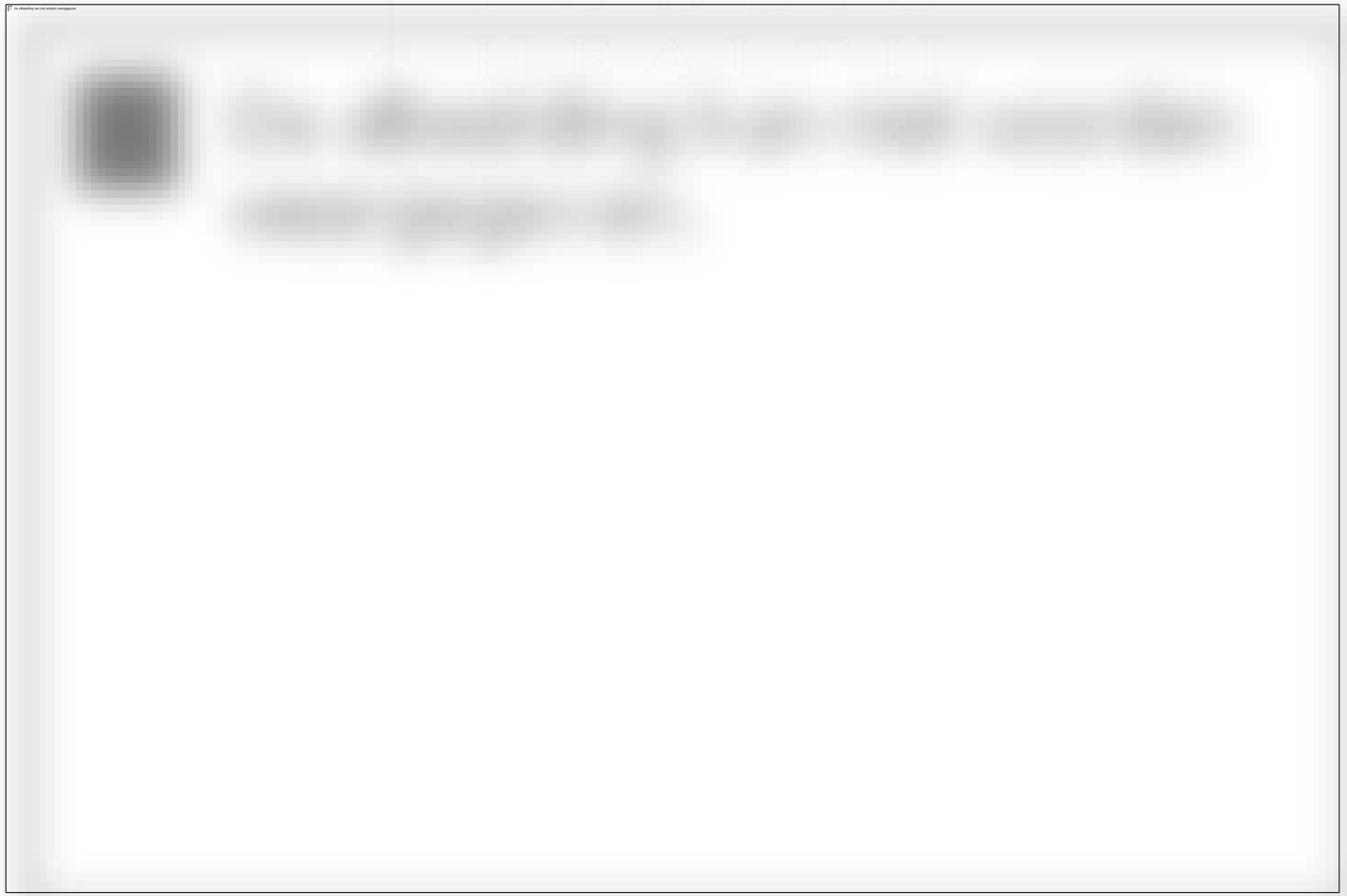
Polygenic Risk Scores: Going Where?

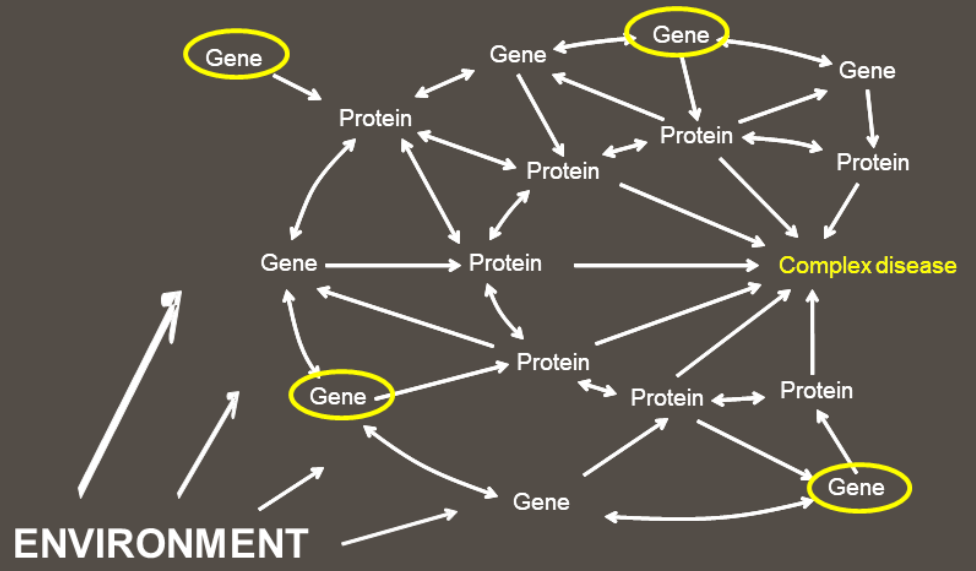


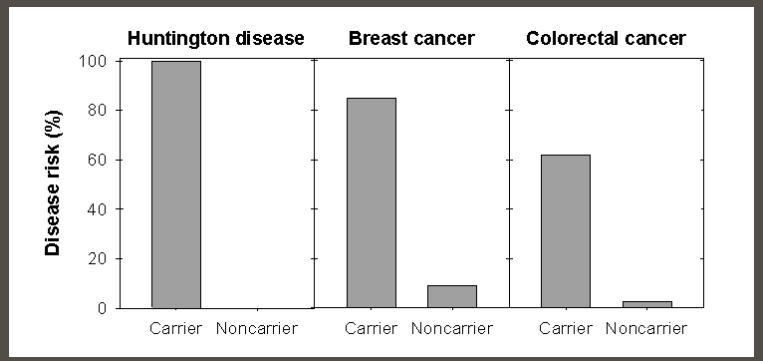
A. Cecile J.W. Janssens, PhD
Research professor of epidemiology
Rollins School of Public Health, Emory University
@cecilejanssens

Monogenic

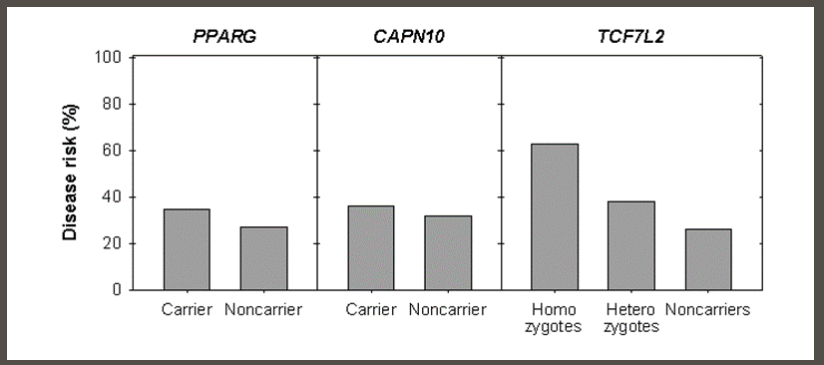
Gene → Something → Disease

Complex/multifactorial





Mutations

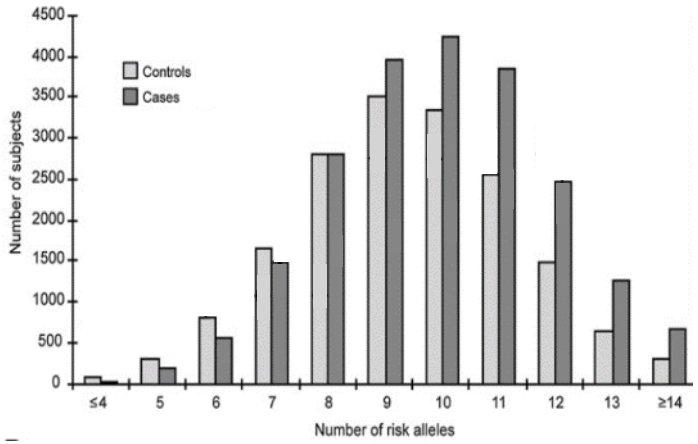


Variations
(single nucleotide polymorphisms; SNPs)

Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42 103 individuals

Malcolm G Dunlop,[1] Albert Tenesa,[2] Susan M Farrington,[1] Stephane Ballereau,[1] [redacted] Steven Penegar,[24] Harry Campbell,[25] Ian Tomlinson,[22] Richard S Houlston[24]

SNP	Location	Estimate	SE	OR
Study populations with SNP genotype data for all 10 risk loci (n=39 266)				
rs10411210	19q13	0.12	0.02	1.13
rs9929218	16q22	0.11	0.02	1.11
rs6983267	8q24	0.17	0.01	1.19
rs4779584	15q23	0.13	0.02	1.14
rs4939827	18q21	0.19	0.01	1.21
rs3802842	11q23	0.13	0.02	1.14
rs10795668	10p14	0.11	0.02	1.12
rs16892766	8q23	0.20	0.03	1.23
rs961253	20p12	0.10	0.02	1.11
rs4444235	14q22	0.09	0.01	1.09



10 SNPs

Each 0,1,2 risk alleles

Total 20 risk alleles

Unweighted score

Range : 0 – 20

Weighted score

Range: 0 – 2.7

(namely 2*0.12 + 2*0.11 + …. + 2*0.09 = 2.7)

Dunlop et al. *Gut* 2013

First mentions of genetic information, susceptibility for common diseases, not yet polygenic models

1998

The new genetics

The new genetics in clinical practice

John Bell

1999

1998 ASHG PRESIDENTIAL ADDRESS

Making Genomic Medicine a Reality

Arthur L. Beaudet

1999

SPECIAL ARTICLE SHATTUCK LECTURE

Medical and Societal Consequences of the Human Genome Project

Francis S. Collins, M.D., Ph.D.

Early skeptical views

2000

The New England Journal of Medicine

WILL GENETICS REVOLUTIONIZE MEDICINE?

NEIL A. HOLTZMAN, M.D., M.P.H.
THERESA M. MARTEAU, PH.D.

2001

Misconceptions about the use of genetic tests in populations

Paolo Vineis, Paul Schulte, Anthony J McMichael

1999

When can a risk factor be used as a worthwhile screening test?

N J Wald, A K Hackshaw, C D Frost

Summary points

To be a worthwhile screening test, a risk factor must be strongly associated with a disorder

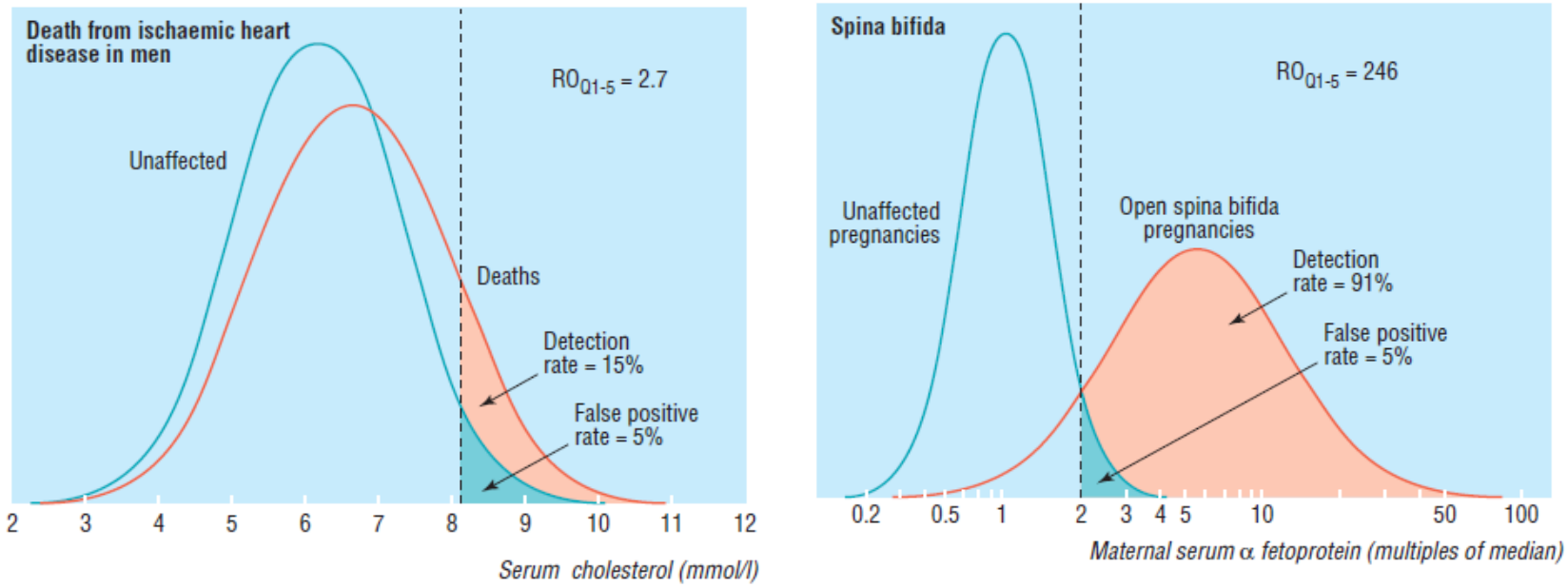


Fig 4 Distribution of maternal serum α fetoprotein in pregnancies affected and unaffected by open spina bifida (derived from Wald et al[2]) and distribution of serum cholesterol in men who did and did not die of ischaemic heart disease (derived from Wald et al[1])

2002

article

Polygenic susceptibility to breast cancer and implications for prevention

Paul D.P. Pharoah[1,2], Antonis Antoniou[3], Martin Bobrow[4], Ron L. Zimmern[2], Douglas F. Easton[3] & Bruce A.J. Ponder[1]

Published online: 4 March 2002, DOI: 10.1038/ng853

- First mention of risk distributions
- Fitted on cancer data from relatives of BC patients
- No mention of individual variants or how to build polygenic risk models

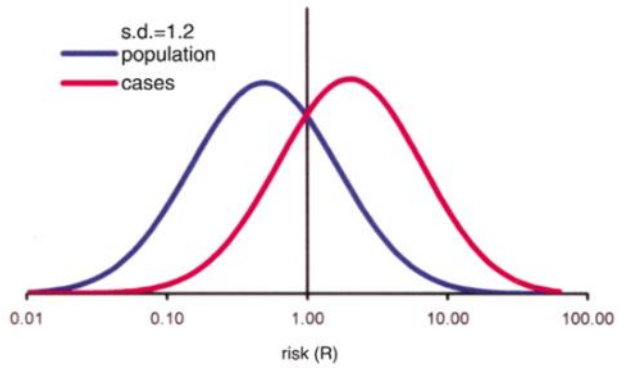
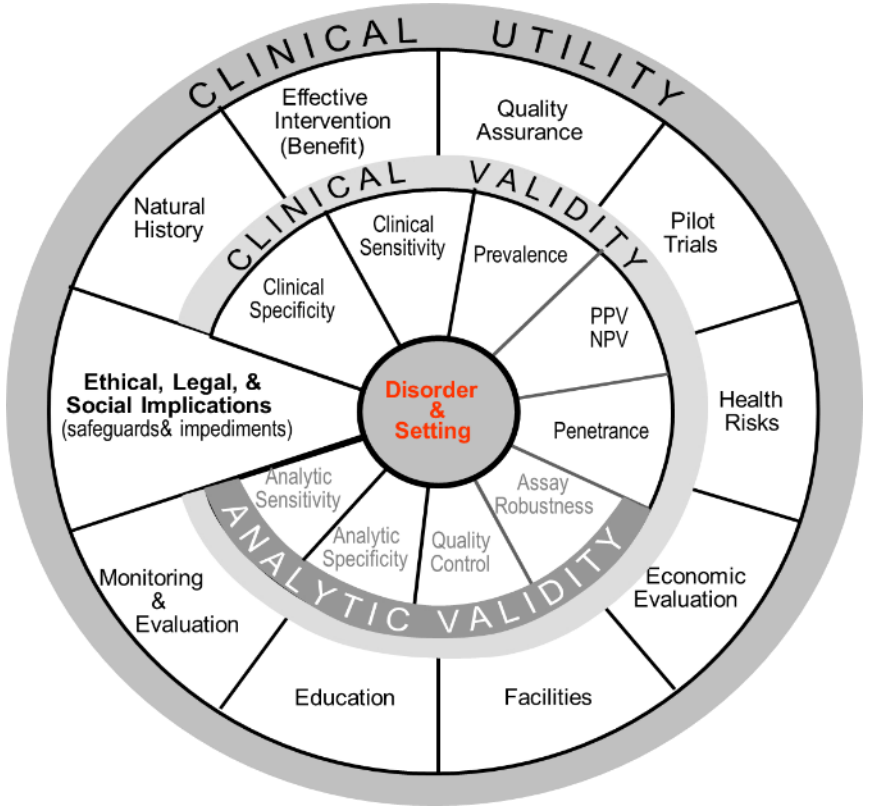


Fig. 1 Distribution of breast cancer risk in the population and in individual cases. Risks are shown on a log scale; the arithmetical average risk for the entire population has been set at 1.0 (see Methods). The risk distribution in individuals who will develop breast cancer (cases) is shifted to the right. The standard deviation describes the spread of risk between high and low values within the population, and thus the potential to discriminate different levels in different individuals.

2003

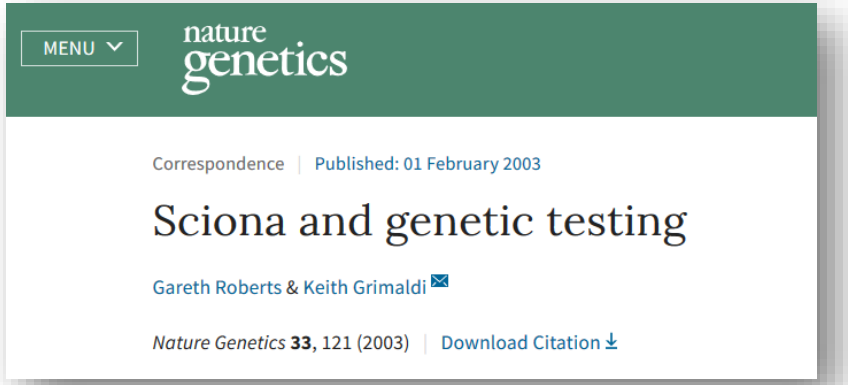
ACCE model: evaluating genetic tests

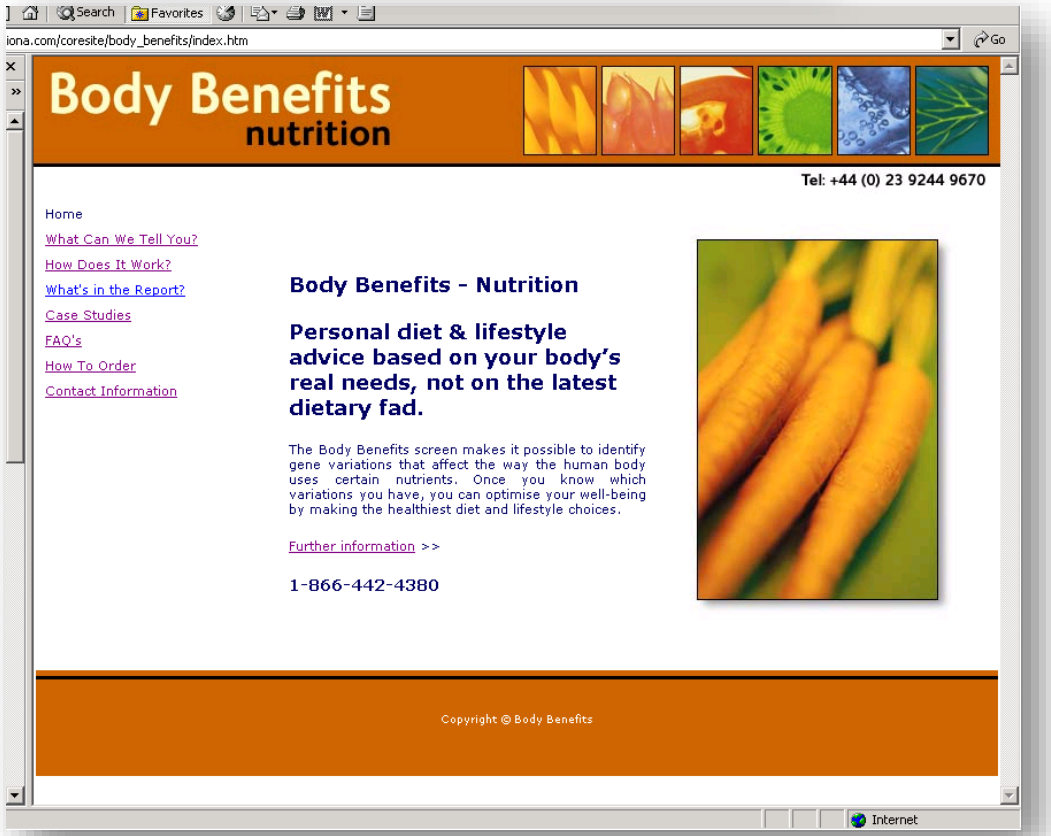
- Comprehensive framework
- Key: Disorder & Setting: What is predicted in whom, for what purpose?
- Assessment changes if setting changes (different population or purpose)



Haddow & Palomaki, Human Genome Epidemiology, 2003

2003


MENU
nature genetics
Correspondence | Published: 01 February 2003
Sciona and genetic testing
Gareth Roberts & Keith Grimaldi
Nature Genetics 33, 121 (2003) | Download Citation


Search
Favorites
iona.com/coresite/body_benefits/index.htm
Go
Body Benefits
nutrition
Tel: +44 (0) 23 9244 9670
Home
What Can We Tell You?
How Does It Work?
What's in the Report?
Case Studies
FAQ's
How To Order
Contact Information
Body Benefits - Nutrition
Personal diet & lifestyle advice based on your body's real needs, not on the latest dietary fad.
The Body Benefits screen makes it possible to identify gene variations that affect the way the human body uses certain nutrients. Once you know which variations you have, you can optimise your well-being by making the healthiest diet and lifestyle choices.
Further information >>
1-866-442-4380
Copyright © Body Benefits
Internet

2003

Am. J. Hum. Genet. 72:636–649, 2003

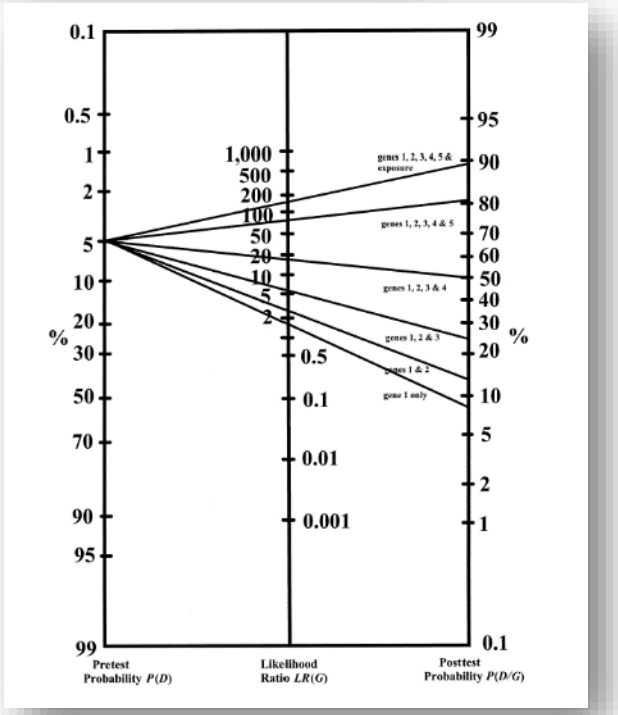
Improving the Prediction of Complex Diseases by Testing for Multiple Disease-Susceptibility Genes

Quanhe Yang,[1] Muin J. Khoury,[2] Lorenzo Botto,[1] J. M. Friedman,[4] and W. Dana Flanders[3]

[1]National Center on Birth Defects and Developmental Disabilities and [2]Office of Genomics and Disease Prevention, Centers for Disease Control and Prevention, and [3]Department of Epidemiology, School of Public Health, Emory University, Atlanta; and [4]Department of Medical Genetics, University of British Columbia, Vancouver

- First study to show how multiple genes can be combined to predict risk, using regression analysis
- Focused on posterior risk for carriers of one or more multiple risk alleles

- (very strong per-allele effects by today's standards (RR 1.5-3.5))



2004

Am. J. Hum. Genet. 74:585–588, 2004

Revisiting the Clinical Validity of Multiplex Genetic Testing in Complex Diseases

To the Editor:

The usefulness of genetic testing to identify high-risk patients for common multifactorial diseases is subject to debate. Optimism about the public health opportunities is counterbalanced with skepticism, since genetic factors appear to play a role in only a minority of patients with complex diseases, the number of genes involved is large, and their penetrance is incomplete (Holtzman and Marteau 2000; Vineis et al. 2001).

A. CECILE J. W. JANSSENS,[1] M. CAROLINA PARDO,[2] EWOUT W. STEYERBERG,[1] AND CORNELIA M. VAN DUIJN[2]

Am. J. Hum. Genet. 74:588–589, 2004

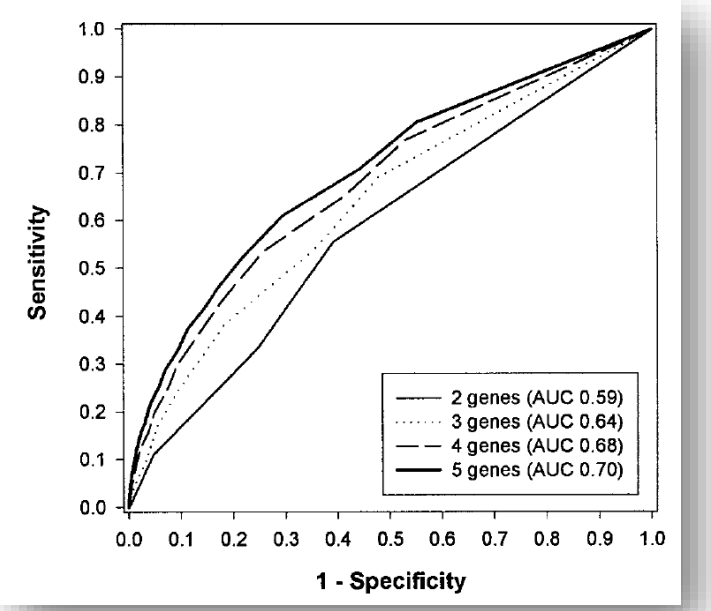
Revisiting the Clinical Validity of Multiplex Genetic Testing in Complex Diseases: Reply to Janssens et al.

To the Editor:

We appreciate the comments by Janssens and her associates (2004 [in this issue]) regarding our study on the use of likelihood ratios to improve the prediction of complex diseases by testing for multiple-susceptibility genes (Yang et al. 2003). As Janssens et al. correctly point out, our study considers only the predicted probability of disease for subjects who have all positive testing results, and this is likely to be an infrequent occurrence. We think that the suggestion made by Janssens et

QUANHE YANG,[1] MUIN J. KHOURY,[2] LORENZO BOTTO,[1] J. M. FRIEDMAN,[4] AND W. DANA FLANDERS[3]

- Evaluation of test performance should include <u>all people</u>, also noncarriers of risk alleles
- Proposed using Area under the Receiver Operating Curve (AUC)



2005

Sciona™ Optimal health through genetics™

GENOVATIONS™
Predictive Genomics for Personalized Medicine

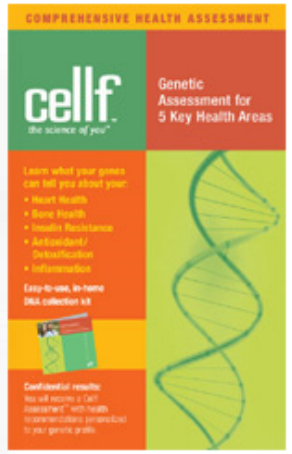
genotrim™

salugen™
DNA Customized Nutrition

INTEGRATIVE GENOMICS

SURACELL
Personal Genetic Health™

genelex



CELLF COMPREHENSIVE

A combined analysis of nineteen genes that may play an important role in how your body manages bone health, heart health, antioxidant and detoxification function, insulin sensitivity and inflammation. Perfect for those who want a complete snapshot of their overall health profile and recommendations for achieving optimal health without a specific disease focus.

Price: **$252.00**

Item	QTY
CELLF COMPREHENSIVE	1

CHECKOUT

http://www.genotrim.nl/

Stop believing the myths about weight loss and GenoTrim™

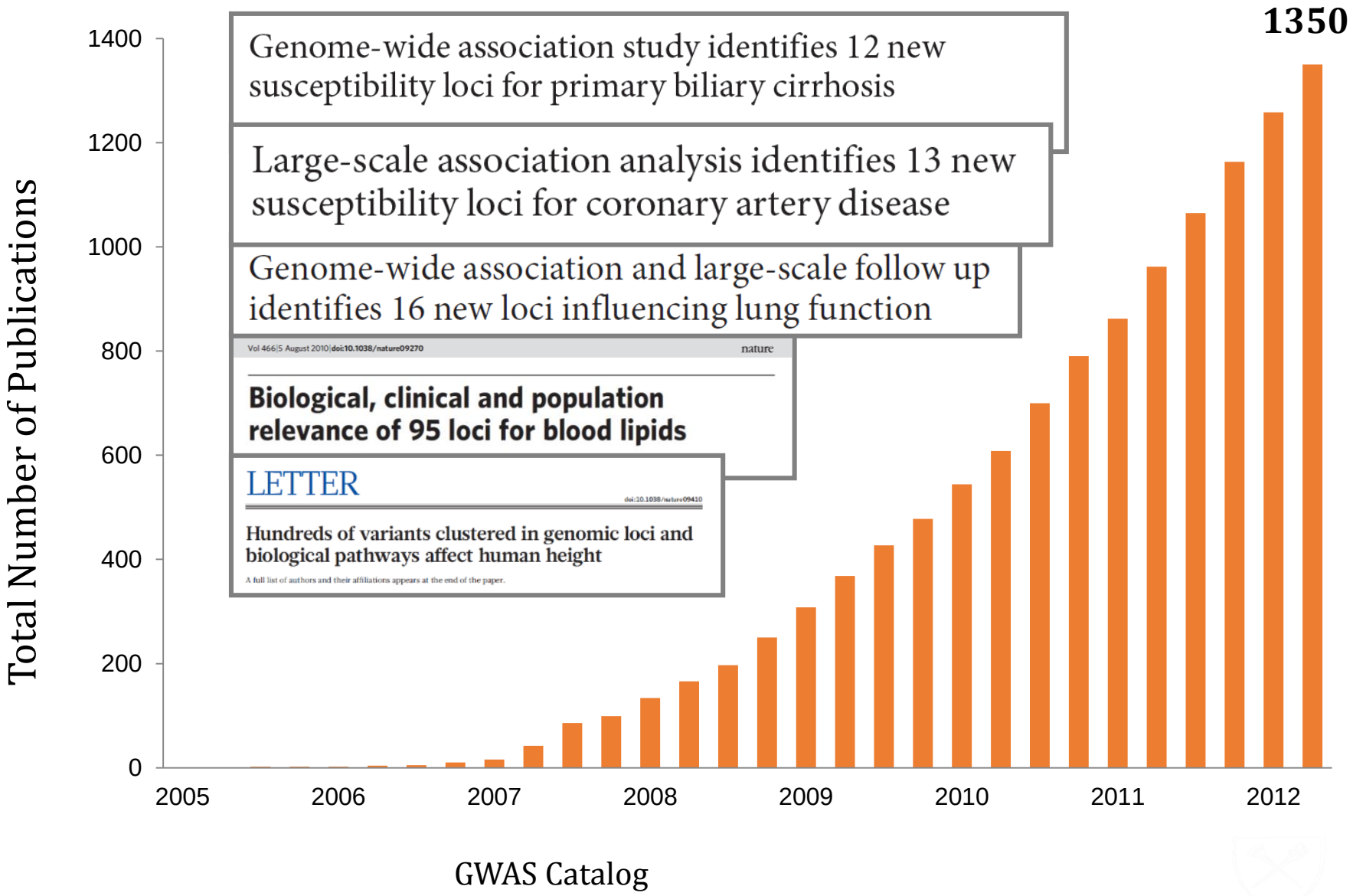
Introducing the first DNA customized solution for weight!*

Personalized nutritional and lifestyle recommendations from the genetic age.

Advice that lasts a lifetime because your genes are not a fad.

EMORY

GWAS Discoveries



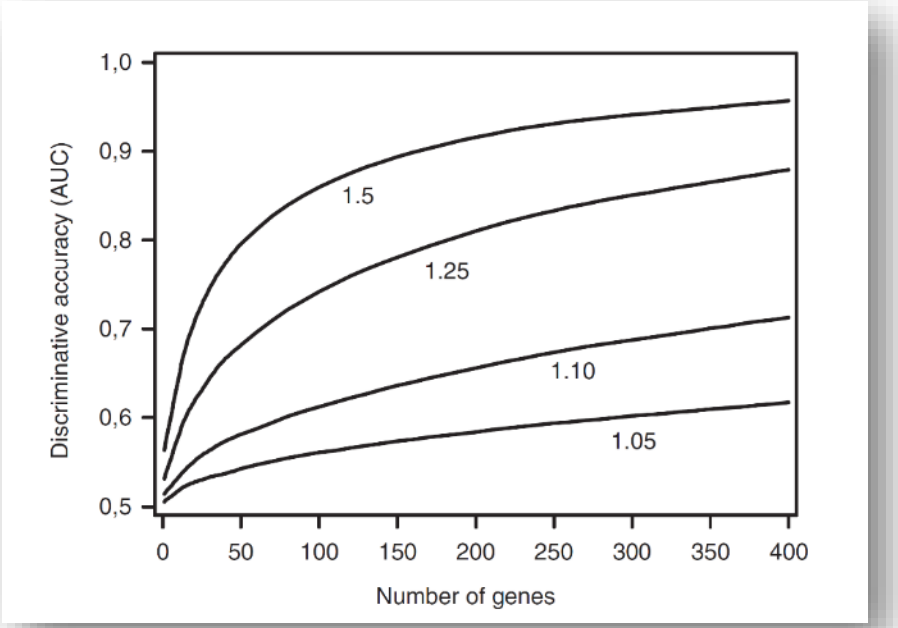
2006

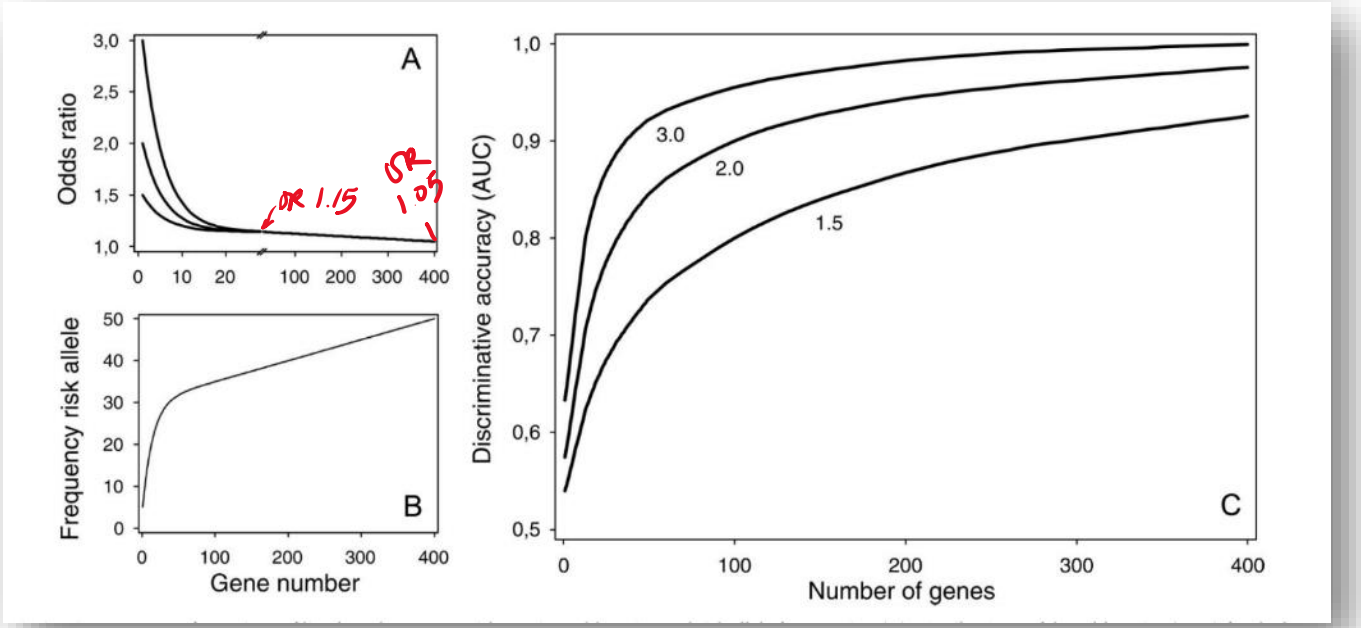
July 2006 · Vol. 8 · No. 7

article

Predictive testing for complex diseases using multiple genes: Fact or fiction?

A. Cecile J. W. Janssens, PhD[1], Yurii S. Aulchenko, PhD[2], Stefano Elefante, PhD[2], Gerard J. J. M. Borsboom, MSc[1], Ewout W. Steyerberg, PhD[1], and Cornelia M. van Duijn, PhD[2]





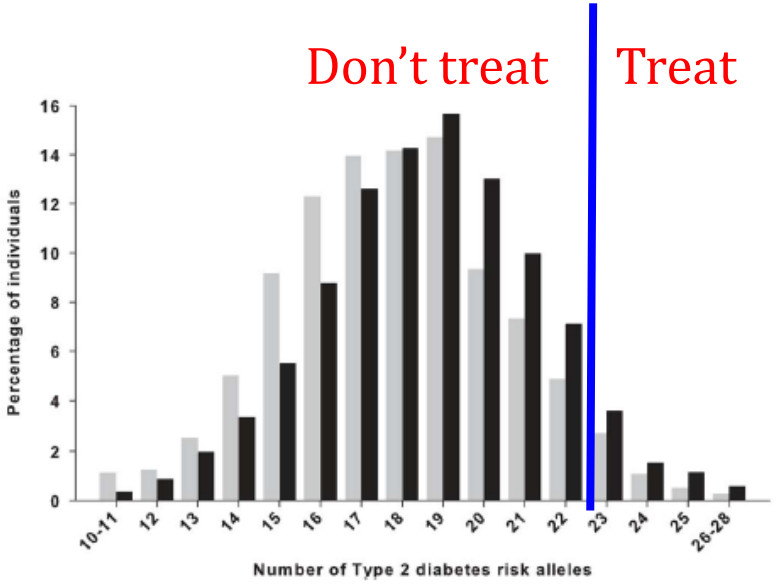
Higher AUC requires a few variants, not too rare, with stronger effects (say, per allele OR > 1.5)

How to get high AUC: common variants with strong effects

Type 2 diabetes AUC = 0.60				Hypertriglyceridemia AUC = 0.80	
TCF7L2	1.36	SLC30A8	1.10	APOA5 19WW	7.36
KCNJ11	1.25	TSPAN8	1.09	APOA5 -1131CC	5.57
CDKN2A/2B	1.21	CDC123	1.10	APOE non-e3	2.14
PPARG	1.21	WFS1	1.07	GCKR TT	2.11
ADAM30	1.15	TCF2	1.07	TRIB1 AA	2.02
CDNK2A/2B	1.13	ADAMTS9	1.05	TBL2 CC	2.81
IGF2BP2	1.12	HHEX-IDE	1.02	GALNT2 GG	2.10
FTO	1.11	THADA	1.04		
CDKAL1	1.11	JAZF1	1.00		

Lango et al *Diabetes* 2008; Wang et al. *Hum Mol Genet* 2008

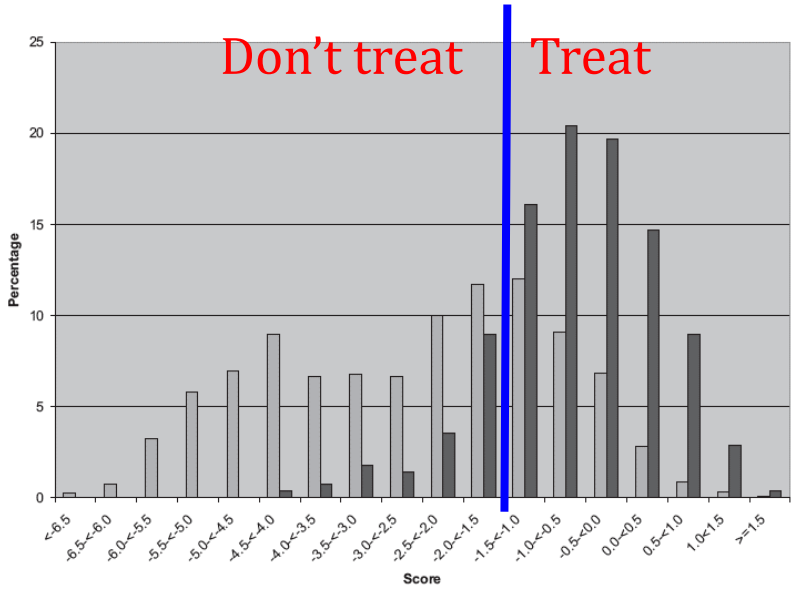
Type 2 diabetes



Lango et al *Diabetes* 2008

AUC = 0.60

AMD

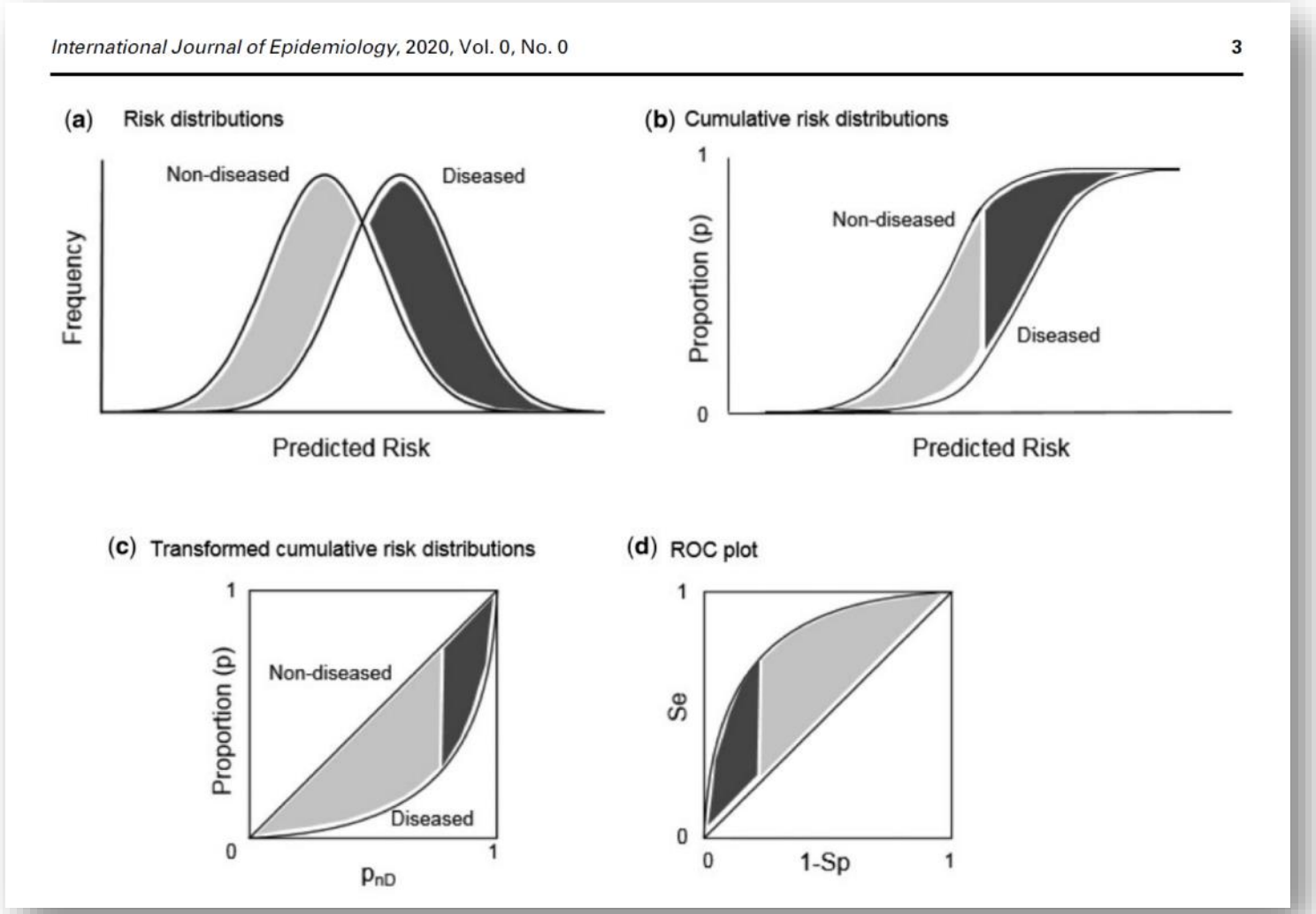


Seddon et al. *IOVS* 2009

AUC = 0.76

AUC = degree of separation between risk distributions of affected and unaffected individuals—nothing more, nothing less

From risk distributions to ROC/AUC: transforming axes



Janssens & Martens, *Int J Epidemiol 2020*

Intended use: Increasing efficiency of healthcare

2008

Polygenes, Risk Prediction, and Targeted Prevention of Breast Cancer

Paul D.P. Pharoah, Ph.D., Antonis C. Antoniou, Ph.D., Douglas F. Easton, Ph.D., and Bruce A.J. Ponder, F.R.S.

Table 2. Absolute Risks of Breast Cancer According to Percentile of Population.*

Percentile of Population	Relative Risk	Lifetime Risk†	10-Yr Risk at 50 Yr of Age†	Age at Which 10-Yr Risk ≥2.3%
		%		yr
5	0.63	6.1	1.5	NA‡
10	0.69	6.7	1.6	NA‡
20	0.77	7.4	1.8	NA‡
40	0.90	8.6	2.1	53
60	1.03	9.7	2.4	49
80	1.20	11.0	2.7	45
90	1.35	12.0	3.0	43
95	1.49	14.0	3.4	41

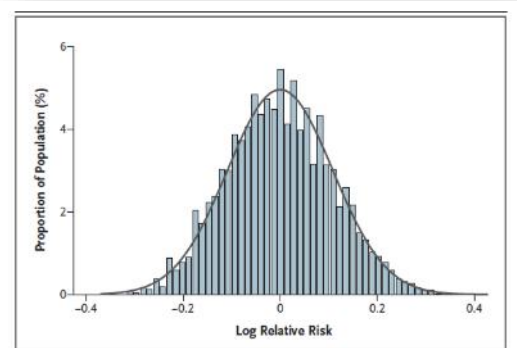
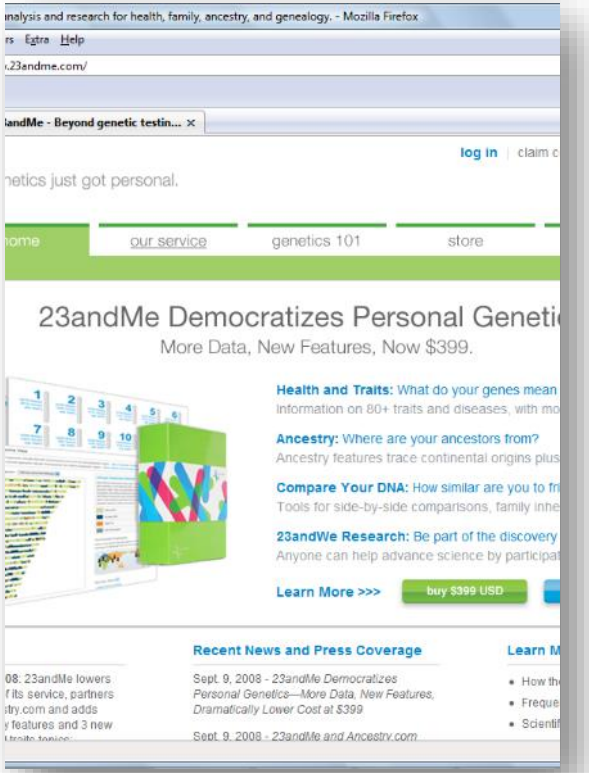
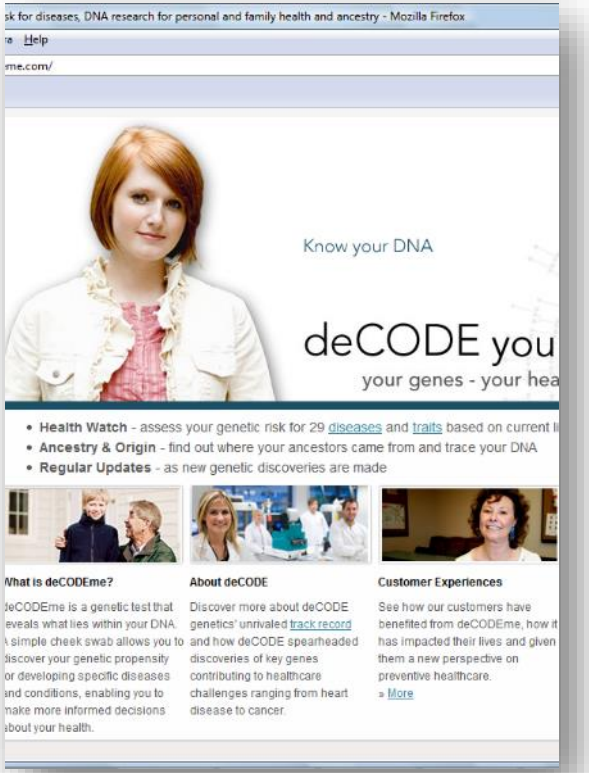


Figure 1. Distribution of Genetic Risk in the Population.
The log relative risk scale of −0.4 to 0.4 is equivalent to 0.4 to 2.5 on the relative risk scale.

2008





From The Sunday Times

September 7, 2008

Rival genetic tests leave buyers confused

Firms that offer to predict your risk of disease give worryingly varied results

Nic Fleming

TIMES RECOMMENDS

2018

Evaluation of polygenic risk scores for ovarian cancer risk prediction in a prospective cohort study

Xin Yang,[1] Goska Leslie,[1] Aleksandra Gentry-Maharaj,[2] Andy Ryan,[2] Maria Intermaggio,[3] Andrew Lee,[1] Jatinderpal K Kalsi,[2] Jonathan Tyrer,[4] Faiza Gaba,[5] Ranjit Manchanda,[2,5,6] Paul D P Pharoah,[1,4] Simon A Gayther,[7,8] Susan J Ramus,[3,9] Ian Jacobs,[2,10,11] Usha Menon,[2] Antonis C Antoniou[1]

J Med Genet 2018;**55**:546–554.

Use of schizophrenia and bipolar disorder polygenic risk scores to identify psychotic disorders

BJPsych The British Journal of Psychiatry (2018) 213, 535-541. doi: 10.1192/bjp.2018.89

Maria Stella Calafato, Johan H. Thygesen, Siri Ranlund, Eirini Zartaloudi, Wiepke Cahn, Benedicto Crespo-Facorro, Álvaro Díez-Revuelta, Marta Di Forti, Genetic Risk and Outcome of Psychosis (GROUP) consortium*, Mei-Hua Hall, Conrad Iyegbe, Assen Jablensky, Rene Kahn, Luba Kalaydjieva

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

The American Journal of Human Genetics *104*, 1–14, January 3, 2019 1

Nasim Mavaddat,[1,*] Kyriaki Michailidou,[1,2] Joe Dennis,[1] Michael Lush,[1] Laura Fachal,[3] Andrew Lee,[1] Jonathan P. Tyrer,[3] Ting-Huei Chen,[4] Qin Wang,[1] Manjeet K. Bolla,[1] Xin Yang,[1] Muriel A. Adank,[5] Thomas Ahearn,[6] Kristiina Aittomäki,[7] Jamie Allen,[1] Irene L. Andrulis,[8,9] Hoda Anton-Culver,[10]

Predictive modeling of schizophrenia from genomic data: Comparison of polygenic risk score with kernel support vector machines approach

Am J Med Genet. 2018;1–6.

Timothy Vivian-Griffiths[1] | Emily Baker[1] | Karl M. Schmidt[2] | Matthew Bracher-Smith[1] | James Walters[1] | Andreas Artemiou[2] | Peter Holmans[1] | Michael C. O'Donovan[1] | Michael J. Owen[1] | Andrew Pocklington[1] | Valentina Escott-Price[1]

nature genetics LETTERS https://doi.org/10.1038/s41588-018-0183-z

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*]

ARTICLES https://doi.org/10.1038/s41588-018-0147-3 nature genetics

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], Meghan Zacher[8], Tuan Anh Nguyen-Viet[9], Peter Bowers[7], Julia Sidorenko[10,11], Richard Karlsson Linnér[5,6,12], Mark Alan Fontana[9,13], Tushar Kundu[9], Chanwook Lee[7], Hui Li[7], Ruoxi Li[9], Rebecca Rover[9]

JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY VOL. 72, NO. 16, 2018

Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults

Implications for Primary Prevention

Michael Inouye, PhD,[a,b,c,i,n,o] Gad Abraham, PhD,[a,b,c,i,n,o] Christopher P. Nelson, PhD,[f] Angela M. Wood, PhD,[c] Michael J. Sweeting, PhD,[c] Frank Dudbridge, PhD,[c,g] Florence Y. Lai, MPhil,[f] Stephen Kaptoge, PhD,[c,h] Marta Brozynska, PhD,[a,b,c] Tingting Wang, PhD,[a,b,c] Shu Ye, MD, PhD,[f] Thomas R. Webb, PhD,[f] Martin K. Rutter, MD,[j,l] Ioanna Tzoulaki, PhD,[k,l] Riyaz S. Patel, MD,[m,n] Ruth J.F. Loos, PhD,[o] Bernard Keavney, MD,[p,q] Harry Hemingway, MD,[r] John Thompson, PhD,[g] Hugh Watkins, MD, PhD,[s,t] Panos Deloukas, PhD,[u] Emanuele Di Angelantonio, MD, PhD,[c,h] Adam S. Butterworth, PhD,[c,h] John Danesh, DPhil,[c,h,v] Nilesh J. Samani, MD,[f,v] for the UK Biobank CardioMetabolic Consortium CHD Working Group

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

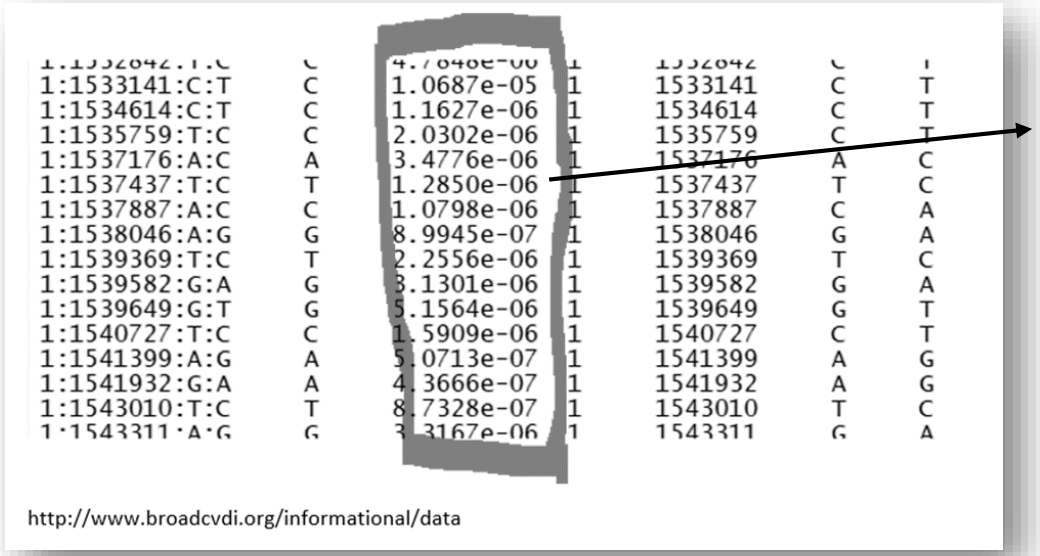
	Number of SNPs included		AUC			Number of PRS (out of 30) with	
	Only genome-wide significant	PRS with highest AUC	Only genome-wide significant	PRS with highest AUC	ΔAUC	ΔAUC < 0	ΔAUC < 0.01
Coronary artery disease	74	6 629 369 (ρ = 0.1%)	0.791	0.806	0.015	2	27
Atrial fibrillation	55	6 705 798 (ρ = 0.3%)	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.

Janssens & Joyner, *Clin Chem* 2019

Polygenic risk scores using weights that can't be observed?

Khera et al.
Nat Genet 2018



1:1533141:C:T	C	1.0687e-05	1	1533141	C	T
1:1534614:C:T	C	1.1627e-06	1	1534614	C	T
1:1535759:T:C	C	2.0302e-06	1	1535759	C	T
1:1537176:A:C	A	3.4776e-06	1	1537176	A	C
1:1537437:T:C	T	1.2850e-06	1	1537437	T	C
1:1537887:A:C	C	1.0798e-06	1	1537887	C	A
1:1538046:A:G	G	8.9945e-07	1	1538046	G	A
1:1539369:T:C	T	2.2556e-06	1	1539369	T	C
1:1539582:G:A	G	3.1301e-06	1	1539582	G	A
1:1539649:G:T	G	5.1564e-06	1	1539649	G	T
1:1540727:T:C	C	1.5909e-06	1	1540727	C	T
1:1541399:A:G	A	5.0713e-07	1	1541399	A	G
1:1541932:G:A	A	4.3666e-07	1	1541932	A	G
1:1543010:T:C	T	8.7328e-07	1	1543010	T	C
1:1543311:A:G	G	3.3167e-06	1	1543311	G	A

http://www.broadcvdi.org/informational/data

= 0.000001285

Per allele OR: 1.000001285

Most SNPs had weights lower than 0.00001

The SMALLEST per allele odds ratio that can be CALCULATED from UK Biobank SNP data

	Number of alleles in		
	Patients	Controls	
Risk allele	250,001	250,000	500,001
Non-risk allele	249,999	250,000	499,999
	500,000	500,000	1,000,000

Per allele OR 1.000008000032

Cold Spring Harbor Laboratory

bioRxiv

THE PREPRINT SERVER FOR BIOLOGY

1 comment

Software as a Service for the Genomic Prediction of Complex Diseases

Alessandro Bolli, Paolo Di Domenico, Giordano Bottà

doi: https://doi.org/10.1101/763722

PPV = # Cases top 3% / # People top 3%

=

PRS panel	SNPs in PRS	AUC (95% CI)	PPV (3%)	Cases in top 3%
Khera full	6630150	0.805 (0.798–0.812)	12.[illegible]	1031
Khera 1%	66300	0.798 (0.792–0.805)	11.[illegible]	945
Khera 0.1%	6630	0.794 (0.788–0.801)	10.[illegible]	909
Khera 74	74	0.789 (0.784-0.797)	10.[illegible]	804

0.81 0.80 0.79 0.79

PRS + age + sex

EMORY

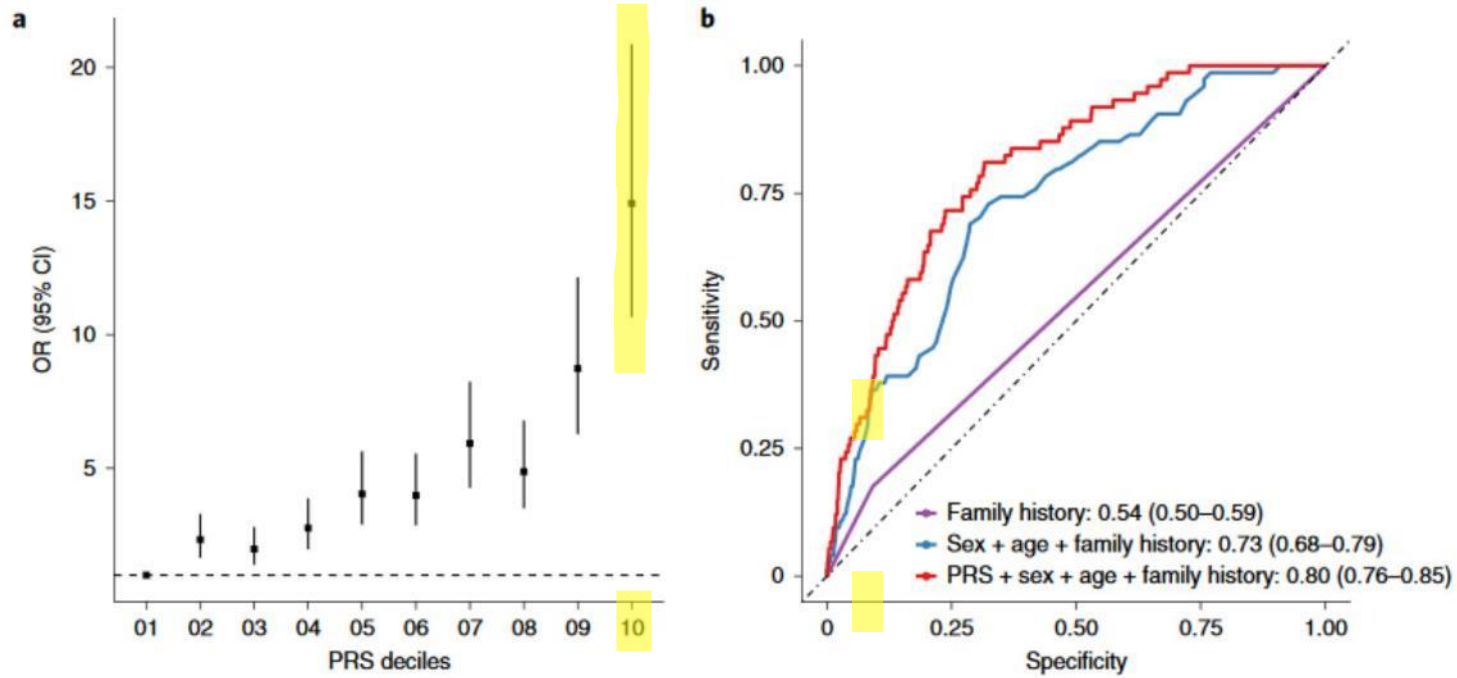
nature genetics

ARTICLES

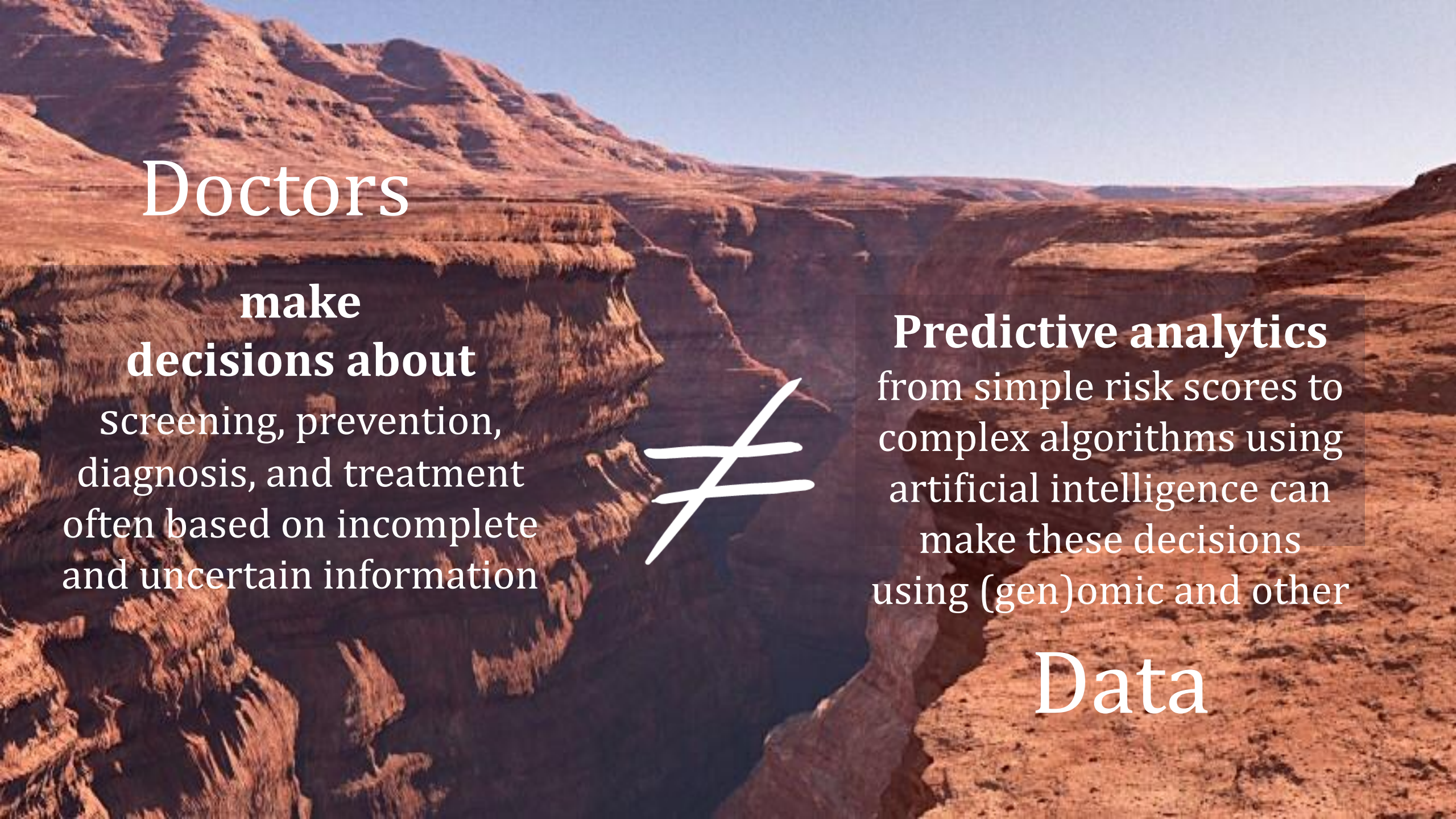
https://doi.org/10.1038/s41588-019-0556-y

Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression

Jamie E. Craig[1,40], Xikun Han[2,3,40*], Ayub Qassim[1,40], Mark Hassall[1], Jessica N. Cooke Bailey[4],



Moving forward


Doctors
make
decisions about
Screening, prevention, diagnosis, and treatment often based on incomplete and uncertain information
≠
Predictive analytics
from simple risk scores to complex algorithms using artificial intelligence can make these decisions using (gen)omic and other
Data

Essential questions for prediction research: intended use (Why?)

In whom?

How?

When?

What?

0 — *time* → — t

(Clinical) risk factors
and/or
Polygenic risk score

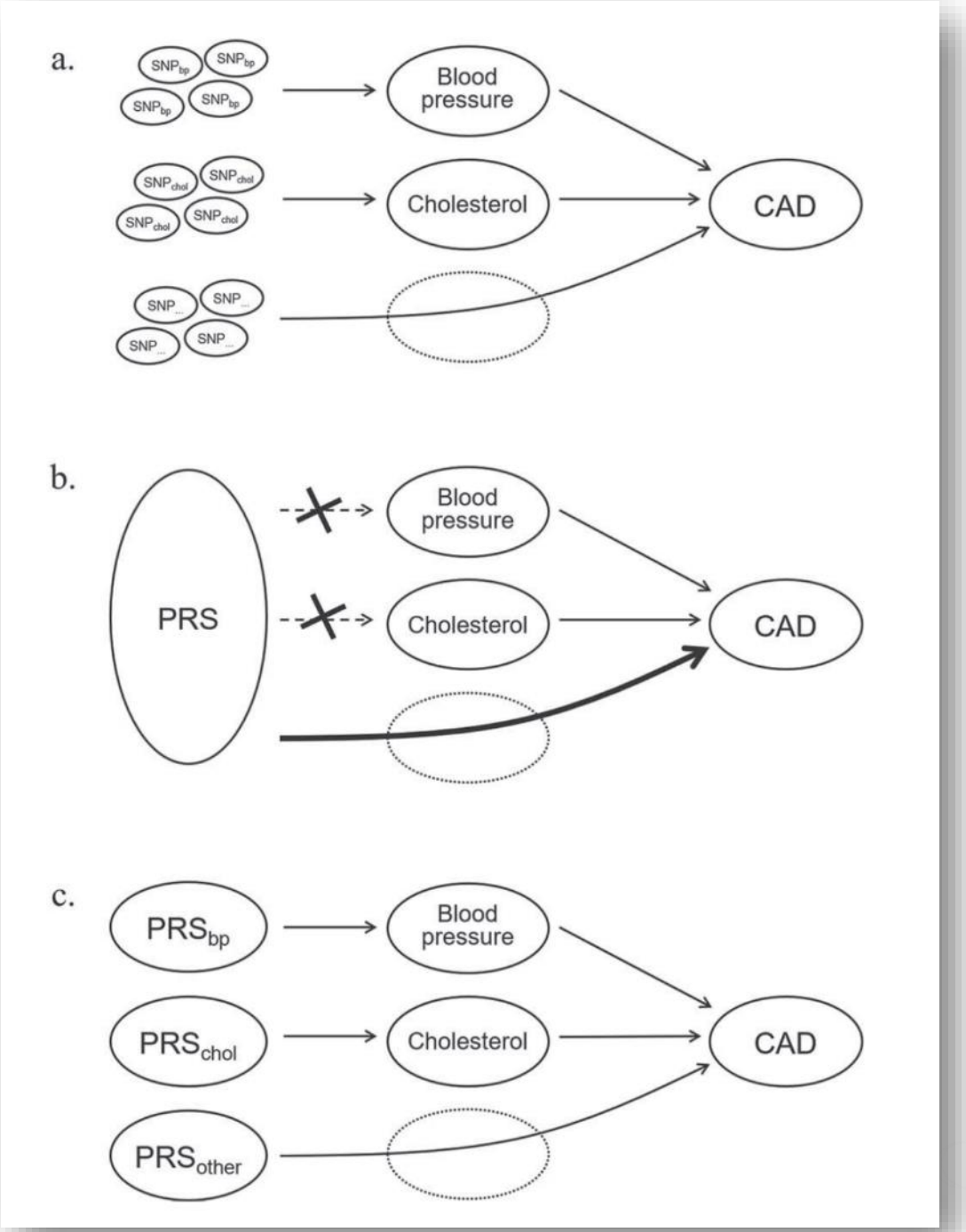
Disease
Recurrence
Treatment response
Side effects
…

Make a prediction model

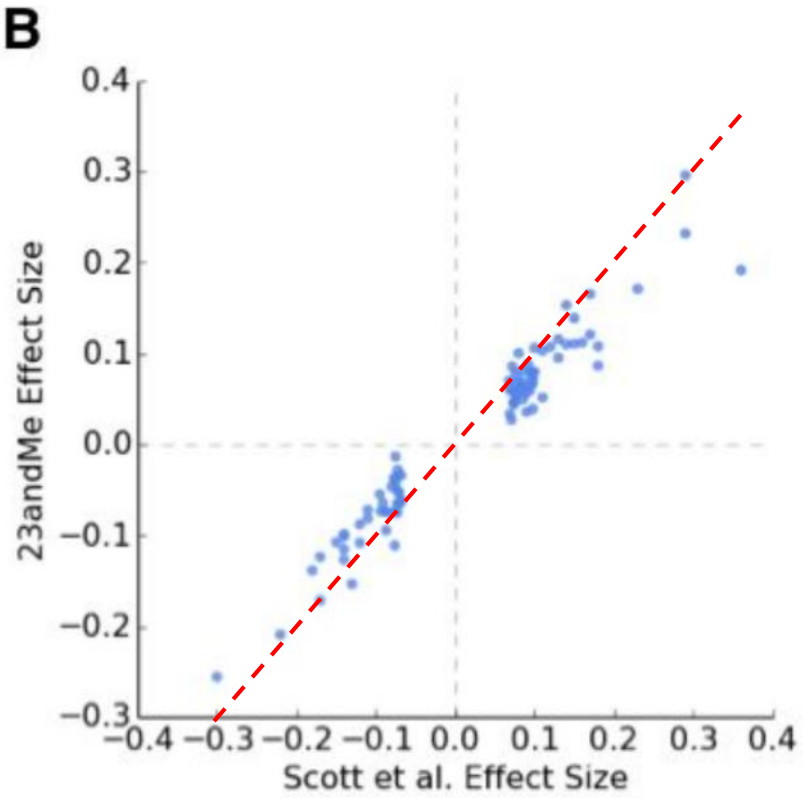
Don't just put variables in a model in data you have available

Validity of PRS

1. Independent effects?
2. Estimation of weights: how?
3. Millions of SNPs: really?



Accuracy of SNP weights



Scott et al:
n ~ 150,000

23andMe
n ~ 1,500,000

Source: 23andMe 2019

- Even with large n, weights may differ between samples
- Calibration of PRS in intended population crucial but often forgotten

Prediction of atrial fibrillation

METHODS

Participants

We examined the association between AF genetic risk and incident AF in 5 prospective studies. Briefly, these studies were the MDCS (Malmö Diet and Cancer Study),[20] MESA (Multi-Ethnic Study of Atherosclerosis),[21] PREVEND (Prevention of Renal and Vascular Endstage Disease),[22] PROSPER (Prospective Study of Pravastatin in the Elderly at Risk),[23] and BioVU (Vanderbilt University Deidentified DNA Biobank).[24] We also examined the association between AF genetic risk and stroke in MGH-GASROS (Massachusetts General Hospital Genes Associated

up. Models were adjusted for variables included in a previously validated composite risk score for 5-year AF risk prediction (CHARGE [Cohorts for Heart and Aging Research in Genomic Epidemiology]-AF risk score).[9] The composite CHARGE-AF risk score included age, height, weight, systolic and diastolic blood pressures, smoking status, antihypertensive medication use, diabetes status, heart failure status, myocardial infarction status, electrocardiographic evidence of left ventricular hypertrophy, and PR interval. Electrocardiographic variables that were not available were omitted from the scores on a study-by-study

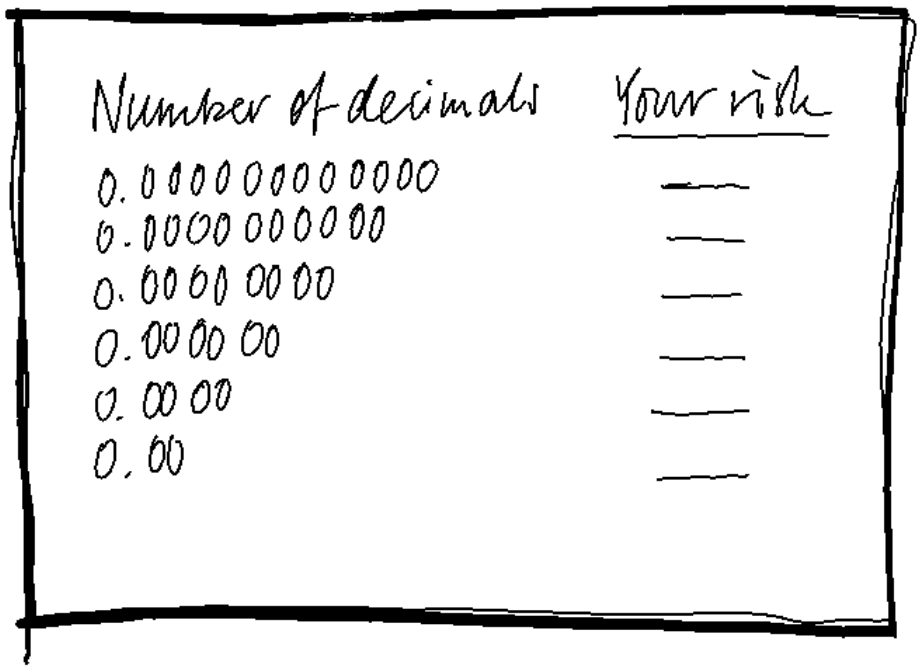
Lubitz et al. *Circulation* 2017

AUC

	Clinical risk factors	+ PRS of 719 SNPs
MDCS	0.75	0.76
MESA	0.80	0.80
PREVEND	0.76	0.80
PROSPER	0.62	0.63
BioVU	0.67	0.67

When do variants contribute to risk?

- Replication of SNP in independent data?
- Statistical significance?
- When its weight is high enough to change risks:

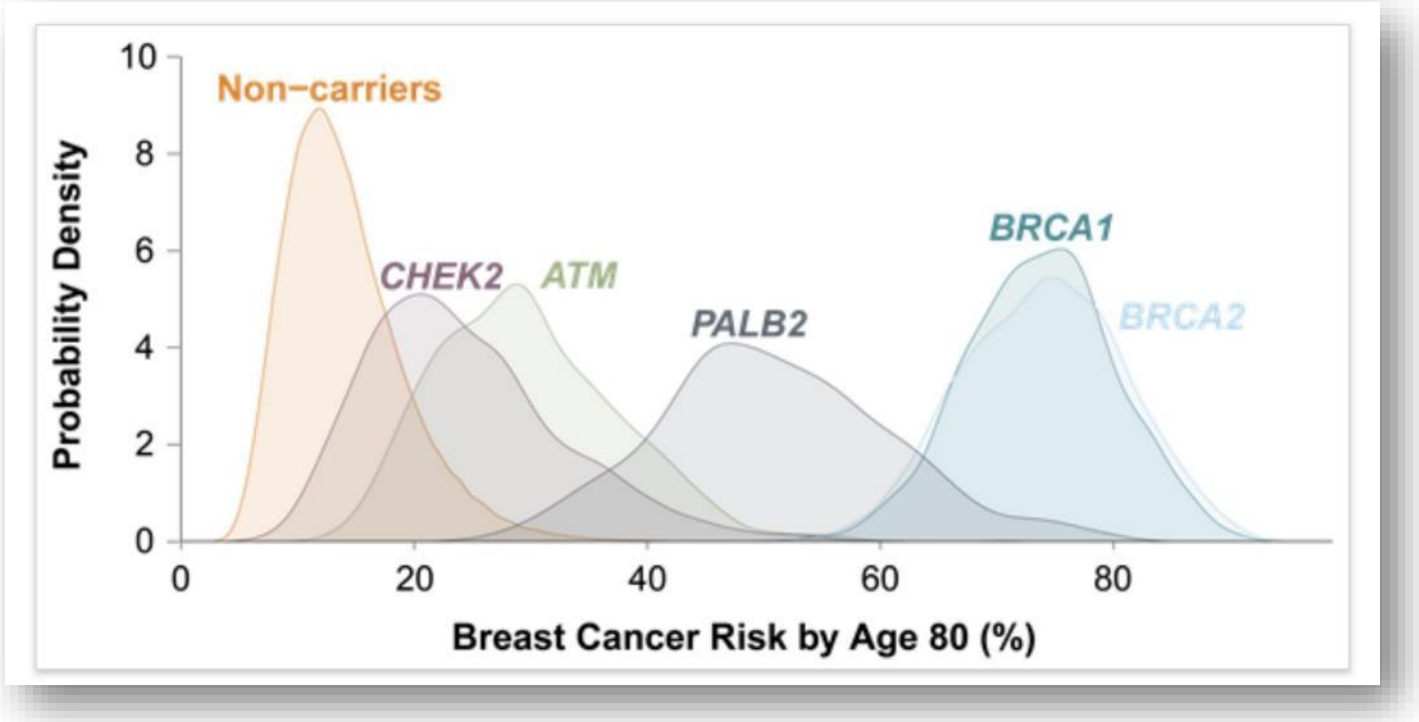


Better prediction studies: focus on intended use

- Large enough cohort that is representative (enough) for the population in which application of the PRS is foreseen
  - No small or case-control studies. No entire biobank datasets that include 18-90-year-olds if your disease of interest has a more limited range for age of onset (which disease hasn't?)
- Clinically relevant follow-up time
  - Both too short and too long can be irrelevant.
- Adequate consideration of non-genetic predictors
  - Compare with current and (reasonable) alternative opportunities for prediction
- Appropriate modeling
  - There is tons of guidance on prediction modeling. Follow conventions or justify why you don't.
- Relevant evaluation
  - Include at least calibration, discrimination, and recalibration (the latter only if the currently used risk model has treatment thresholds)
  - Make fair and informative comparisons with current and alternative prediction opportunities
  - Demonstrate what PRS adds. Show models with and without PRS.
  - Report weights/coefficients for all predictors
- Responsible interpretation, ideally based on external validation
  - Statistical significance ≠ clinical or public health relevance. Prediction is about relevance.
  - No overinterpretation of small effects and small improvements.
  - Realistic and honest implications for healthcare.

Highest AUC is not the goal, clinical utility is; small ΔAUC may be statistically significant, but not change medical decisions or stratification of risk groups

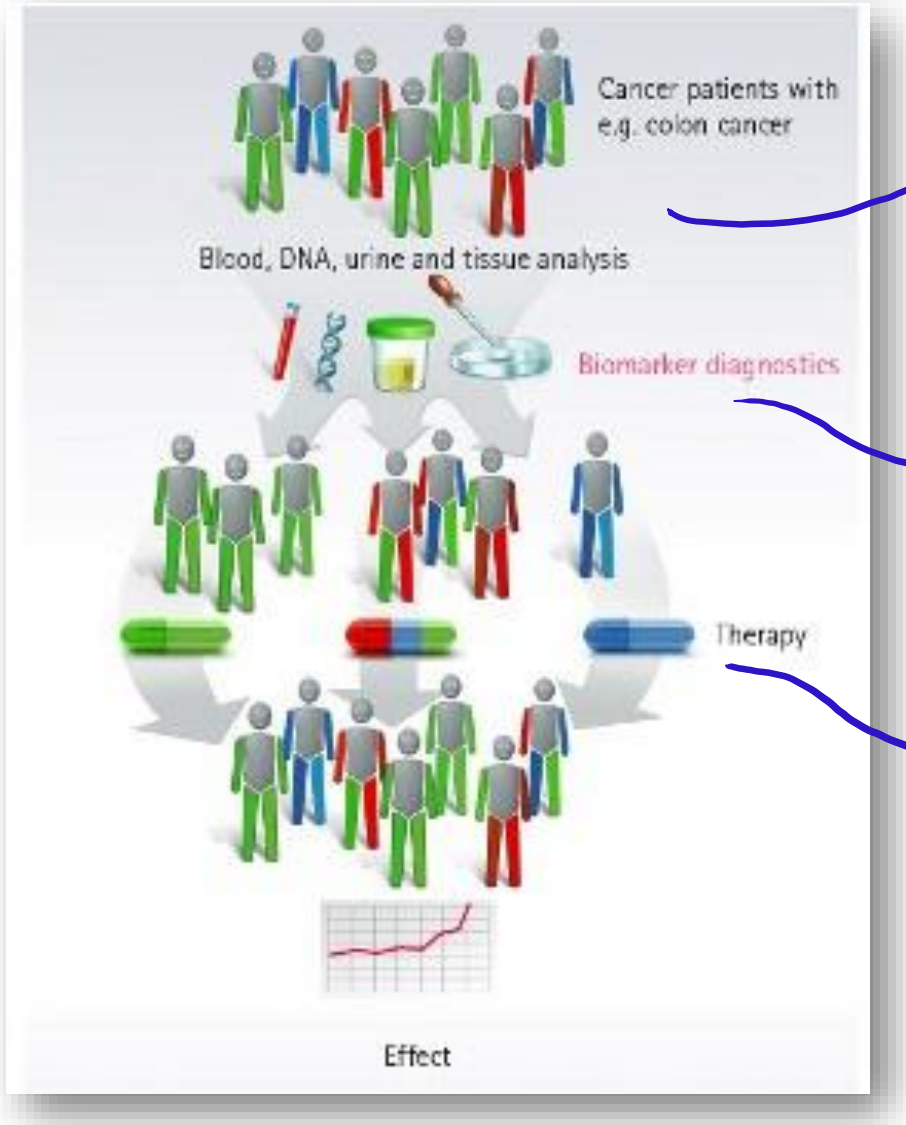
Combining mutations and PRS



Nice, but:

- Do we know how these risks are calculated?
- Are they validated? Calibrated? How confident are we about accurate risks in the tails?
- Do people value this level of precision? What is the utility?

Personalized medicine: When? And when not?



1. Variability in 'outcome', with some outcomes worth avoiding, e.g.,
   Treatment: works in most, not/less in some
   Adverse reactions: none in most, severe in some

2. Predictability of variability
   Treatment effects: in whom does it not work?
   Adverse reactions: who has adverse reactions?

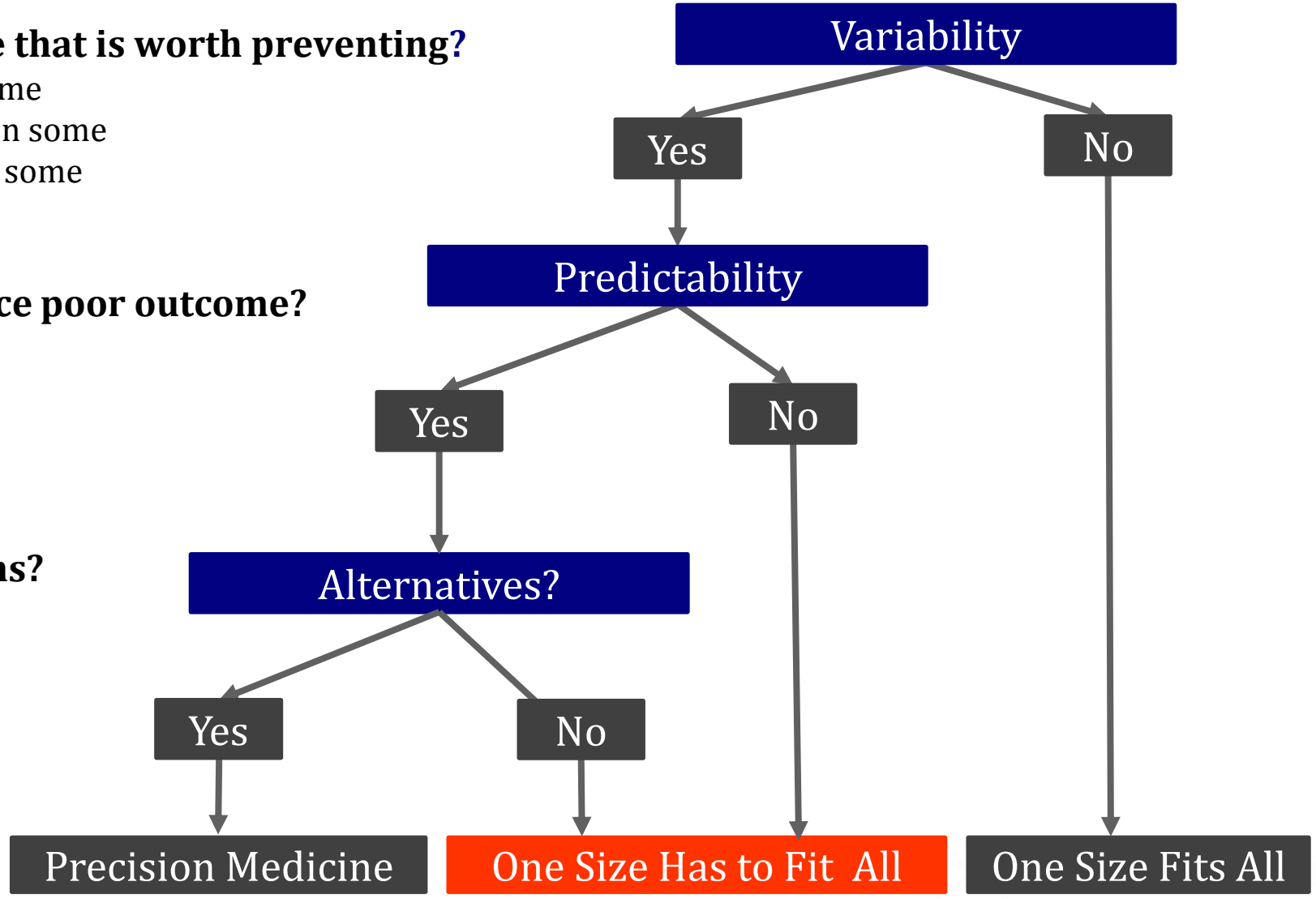
3. Incentive to alternative strategy
   If expected smaller treatment effect:
   If expected adverse reactions:
   Alternative treatment available?
   Is withholding treatment an option?

Personalized Medicine?

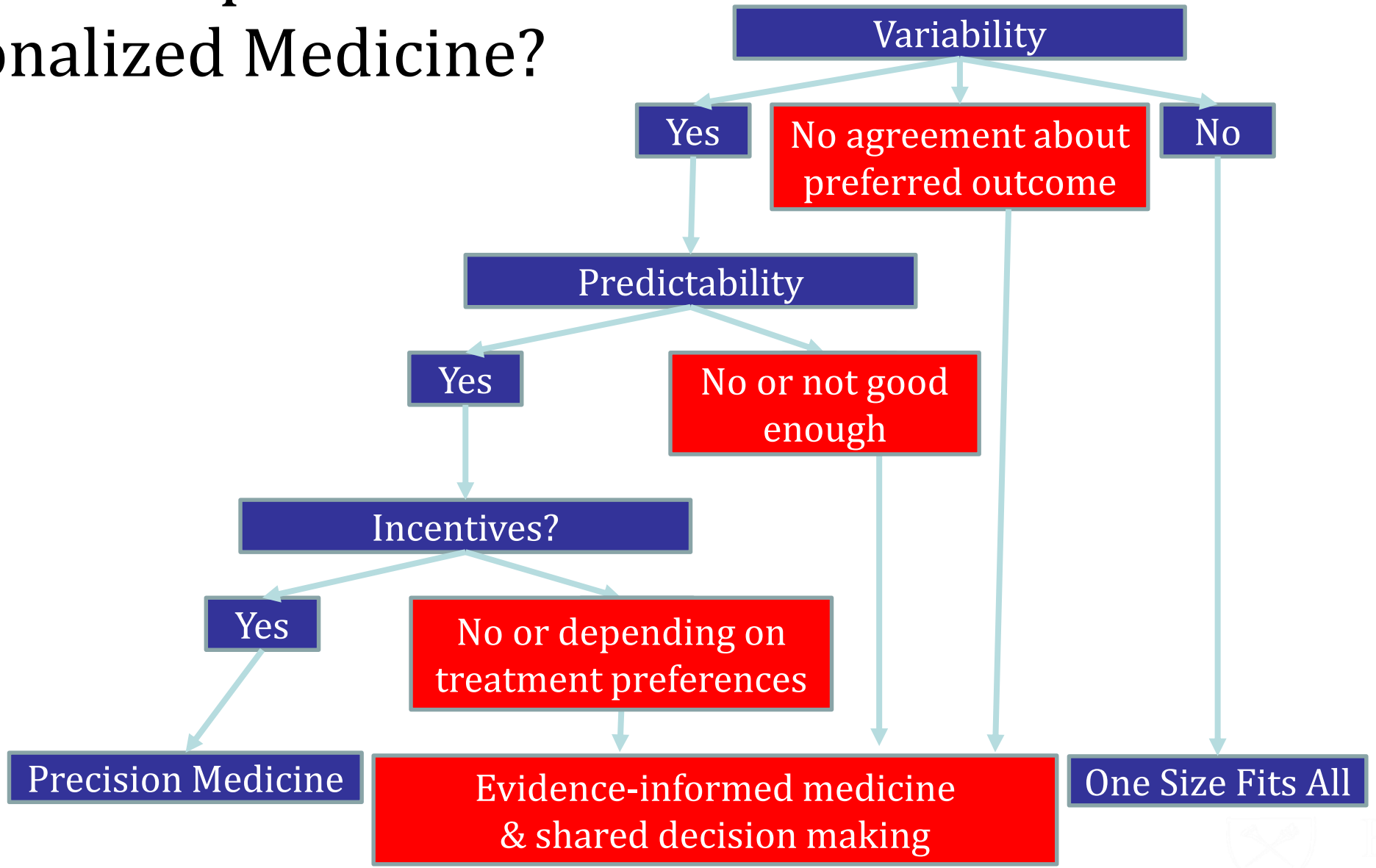
1. **Is there variability in the outcome that is worth preventing?**
   Treatment: works in most, not/less in some
   Adverse reactions: none in most, severe in some
   Prognosis: good for most, unfavorable in some

2. **Can we predict who will experience poor outcome?**
   In whom will treatment not work?
   Who will experience adverse reactions?
   Who has poor prognosis?

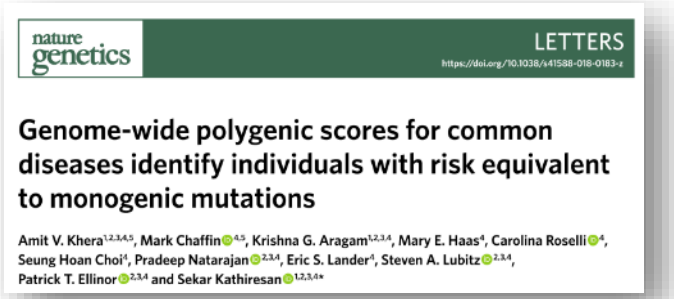
3. **Are there alternative interventions?**
   Are alternative treatments available?
   Is withholding treatment an option?

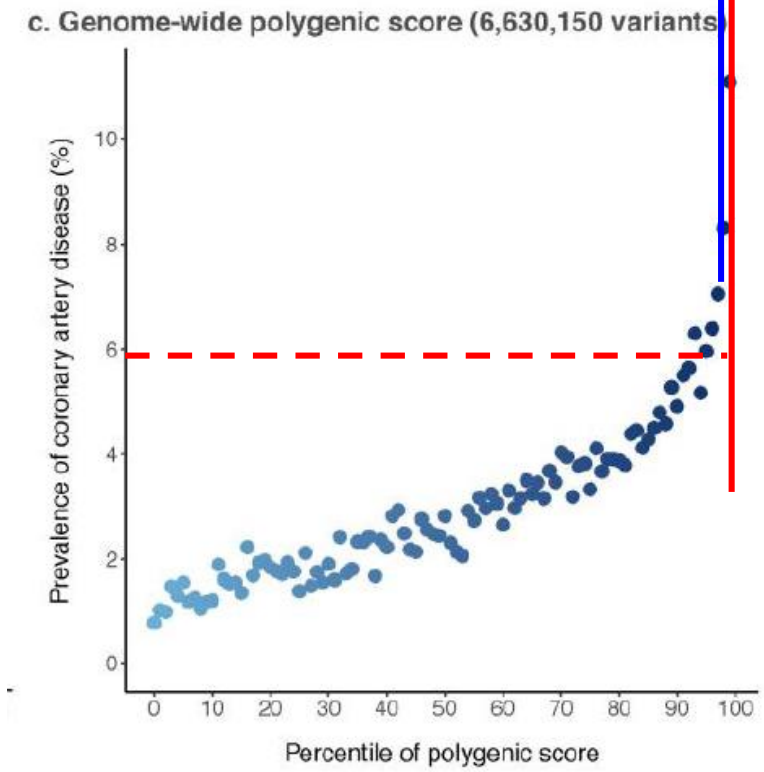


How data improves personalized Medicine?



Don’t overpromise





Confidence interval:

- Predicting for observations in the study sample
- Uncertainty due to random error

‘Prediction interval’:

- Predicting for new observations, outside the study sample
- Uncertainty due to random + nonrandom error

See for discussion on PI: https://twitter.com/cecilejanssens

Making prediction more accurate means improving quality of data and model

Data x Model → Prediction

DNA sequencing

Increasingly cheaper and more accurate

(Easy)

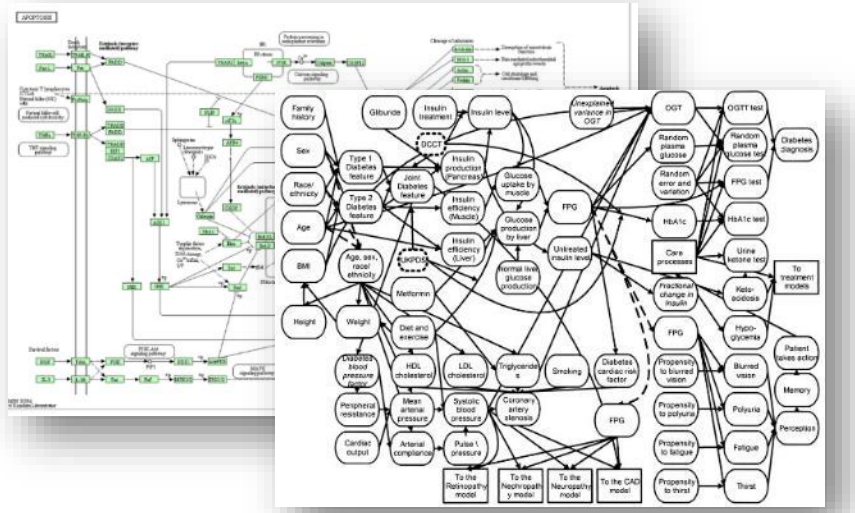
Non-genetic data

What to measure, when, and how?

(Hard)

Modeling disease pathways beyond PRS

$$score = \beta_1 * snp_1 + \beta_2 * snp_2 + \cdots \beta_n * snp_n$$



(Very hard)