Genomic analysis in participants with ancestries outside of Europe: opportunities and challenges

Sekar Kathiresan, MD
@skathire

March 25, 2018

Massachusetts General Hospital

Harvard Medical School

Two men and cardiovascular disease

1874-1965
90 years

1932-1984
52 years

1874-1965
90 years

1932-1984
52 years

PROTECTION

Are there mutations that confer 'resistance' to disease?

RISK

Why do some get disease at an early age?

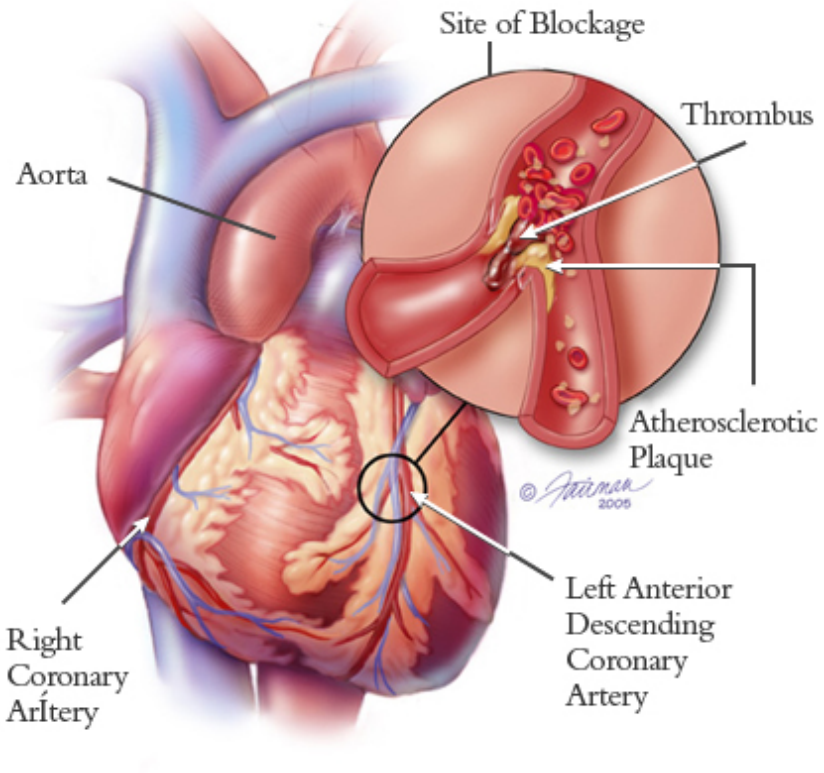
Two questions

Can we find mutations that protect against disease & develop medicines that mimic them?

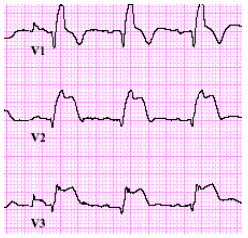
Can we identify those at high genetic risk & offer preventive intervention?

Myocardial infarction (MI)

- Leading cause of death worldwide
- Heritable



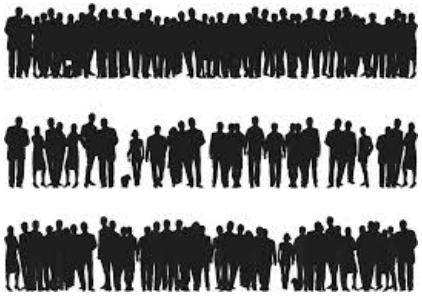
Symptoms

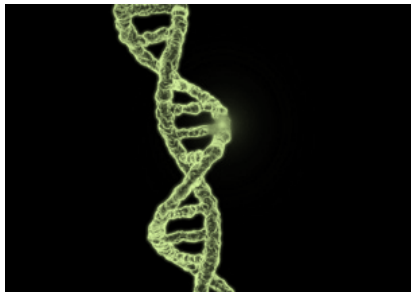


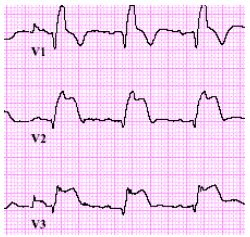
EKG change

Elevation in cardiac biomarkers

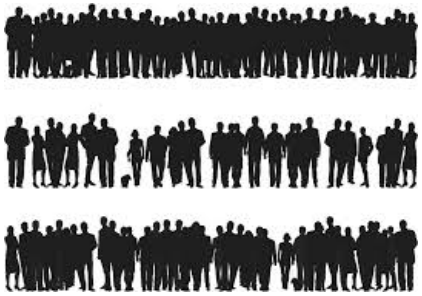
Roger, *Circulation* 2012

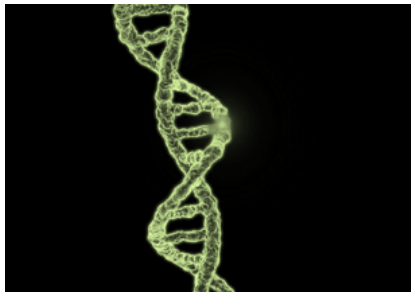


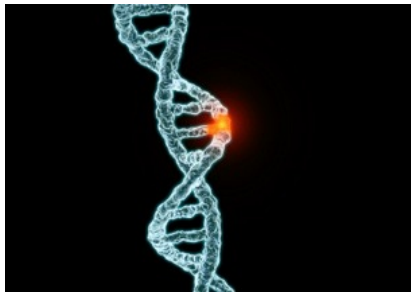


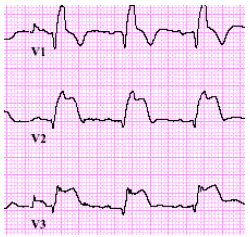


Average risk of disease



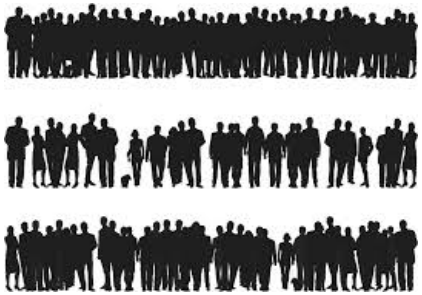




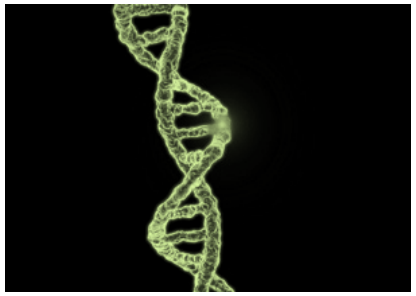


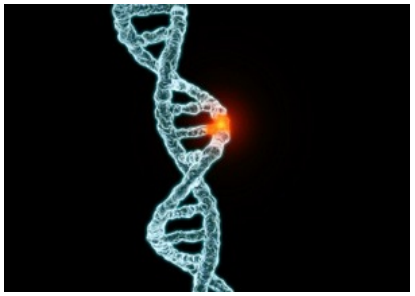
Average risk of disease

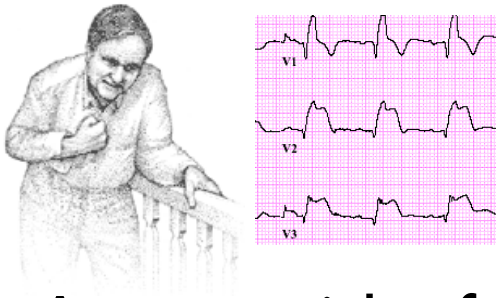
Lower risk



Develop medicine that mimics mutation





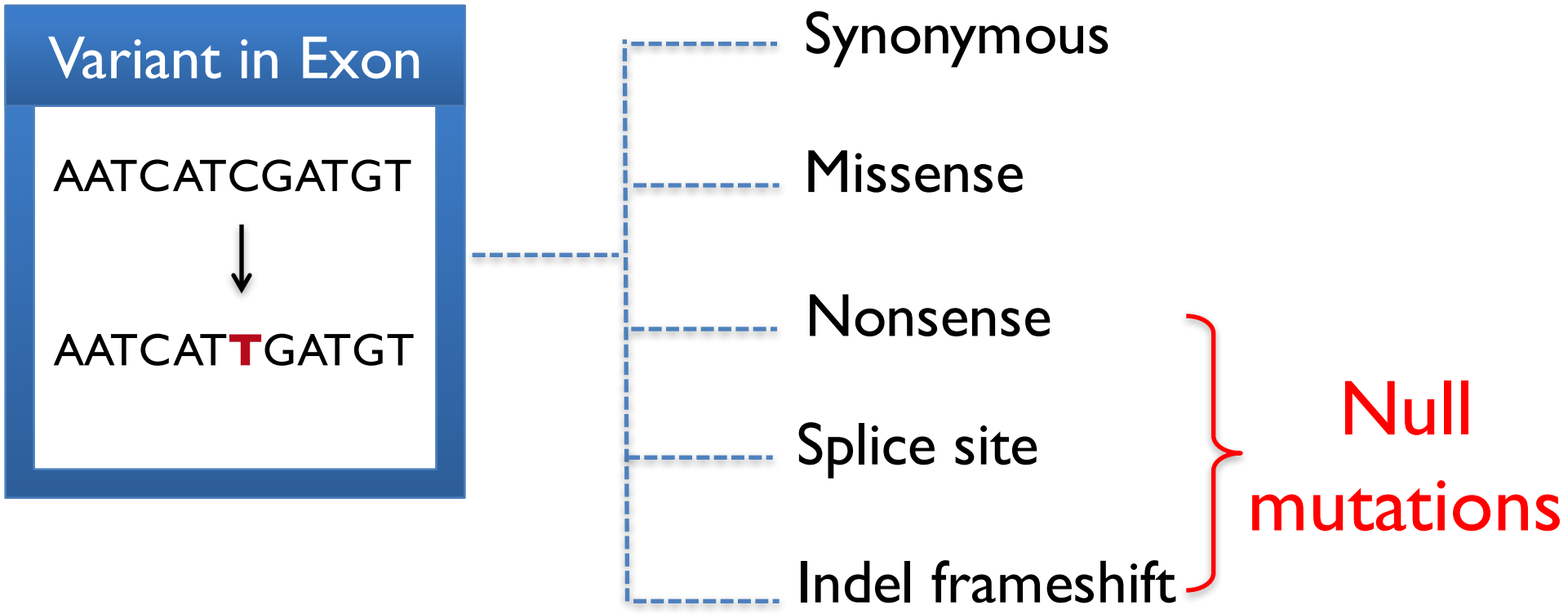


Average risk of disease

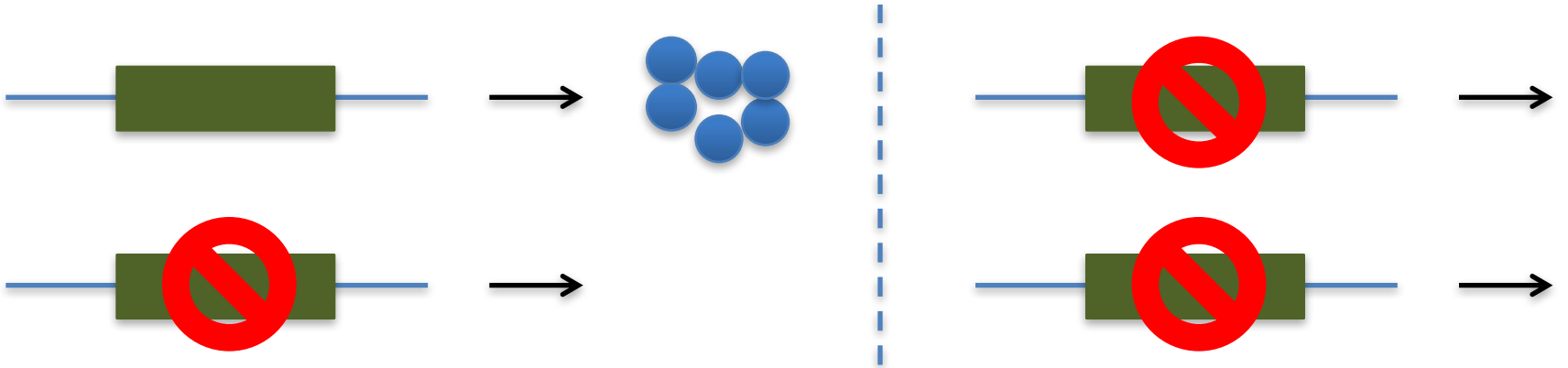
Lower risk

Lower risk

NULL mutations: provide clear direction of effect



Particularly useful for therapeutic target selection are null mutations that reduce risk



Heterozygous null

Homozygous null
'Knockout'

Efficacy

Safety

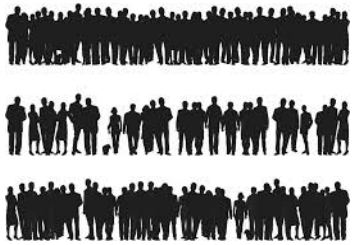
Best example to date: PCSK9

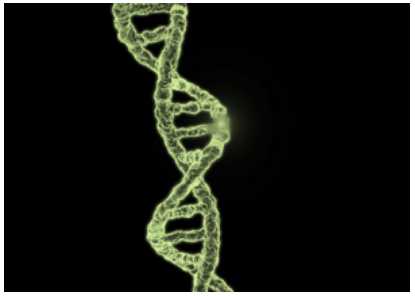
Mutations in *PCSK9* cause autosomal dominant hypercholesterolemia

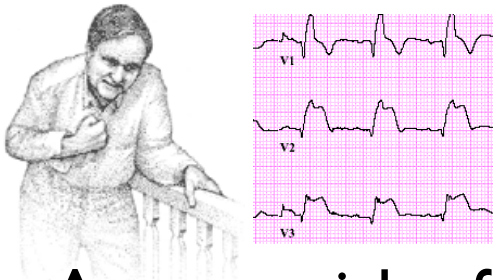
Marianne Abifadel[1,2], Mathilde Varret[1], Jean-Pierre Rabès[1,3], Delphine Allard[1], Khadija Ouguerram[4], Martine Devillers[1], Corinne Cruaud[5], Suzanne Benjannet[6], Louise Wickham[6], Danièle Erlich[1], Aurélie Derré[1], Ludovic Villéger[1], Michel Farnier[7], Isabel Beucler[8], Eric Bruckert[9], Jean Chambaz[10], Bernard Chanu[11], Jean-Michel Lecerf[12], Gerald Luc[12], Philippe Moulin[13], Jean Weissenbach[5], Annick Prat[6], Michel Krempf[4], Claudine Junien[1,3], Nabil G Seidah[6] & Catherine Boileau[1,3]

Abifadel, *Nature Genetics* (2003)

PCSK9: null mutation lower LDL, MI





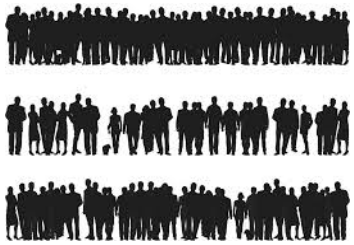


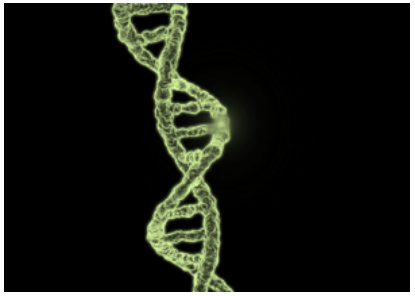
Average risk of heart attack

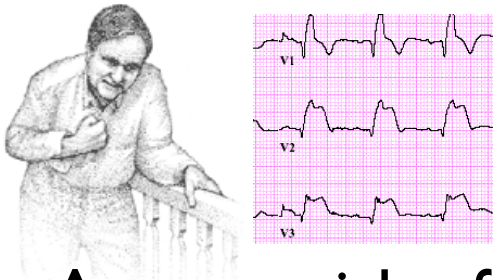
 **1 in 40 blacks**
Null mutation (het) in PCSK9 gene
Lower LDL cholesterol **80% lower risk heart attack**

Abifadel, *Nat Genet* (2003)
Cohen, *N Engl J Med* (2006)

PCSK0 null mutation at 1:40 frequency: found exclusively in blacks







Average risk of heart attack

1 in 40 blacks

Null mutation (het) in PCSK9 gene

Lower LDL cholesterol

80% lower risk heart attack

Null mutation found *only* in those of African ancestry

Abifadel, *Nat Genet* (2003)
Cohen, *N Engl J Med* (2006)

African-American woman completely deficient in PCSK9 (healthy)

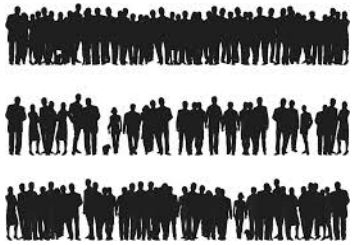
ARTICLE

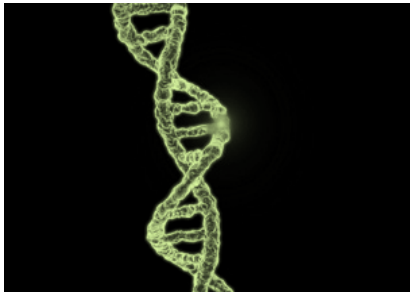
Molecular Characterization of Loss-of-Function Mutations in *PCSK9* and Identification of a Compound Heterozygote

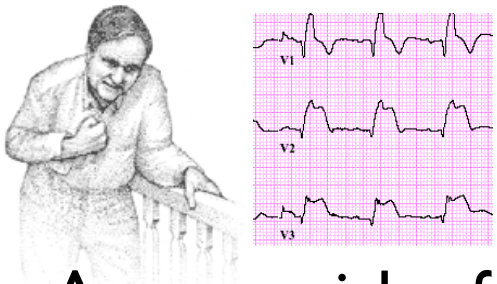
Zhenze Zhao,* Yetsa Tuakli-Wosornu,* Thomas A. Lagace, Lisa Kinch, Nicholas V. Grishin, Jay D. Horton, Jonathan C. Cohen, and Helen H. Hobbs

Elevated levels of circulating low-density lipoprotein cholesterol (LDL-C) play a central role in the development of atherosclerosis. Mutations in proprotein convertase subtilisin/kexin type 9 (PCSK9) that are associated with lower plasma levels of LDL-C confer protection from coronary heart disease. Here, we show that four severe loss-of-function mutations prevent the secretion of PCSK9 by disrupting synthesis or trafficking of the protein. In contrast to recombinant wild-type PCSK9, which was secreted from cells into the medium within 2 hours, the severe loss-of-function mutations in *PCSK9* largely abolished PCSK9 secretion. This finding predicted that circulating levels of PCSK9 would be lower in individuals with the loss-of-function mutations. Immunoprecipitation and immunoblotting of plasma for PCSK9 provided direct evidence that the serine protease is present in the circulation and identified the first known individual who has no immunodetectable circulating PCSK9. This healthy, fertile college graduate, who was a compound heterozygote for two inactivating mutations in *PCSK9,* had a strikingly low plasma level of LDL-C (14 mg/dL). The very low plasma level of LDL-C and apparent good health of this individual demonstrate that PCSK9 plays a major role in determining plasma levels of LDL-C and provides an attractive target for LDL-lowering therapy.

PCSK9 inhibitor: mimics mutation







Average risk of heart attack

 **1 in 50 people**	 **Alirocumab Evolocumab**
Null mutation (het) in PCSK9 gene	**Block PCSK9**
Lower LDL cholesterol **80% lower risk heart attack**	**Lower LDL cholesterol** **15-20% lower risk CV events, death**

Are there more
protective, null mutation
examples for MI?

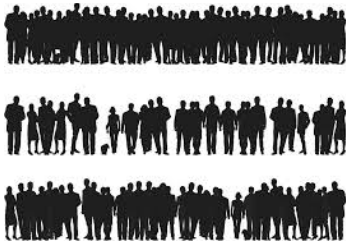
In Single Gene, a Path to Fight Heart Attacks

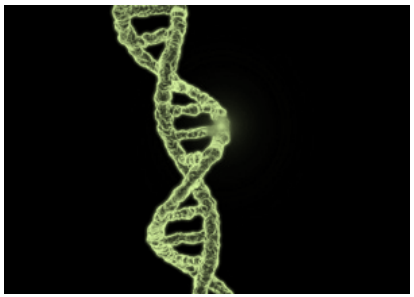
By **GINA KOLATA** JUNE 18, 2014

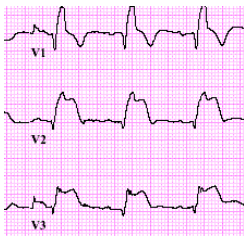
Two major studies by leading research groups published on Wednesday independently identified mutations in a single gene that protect against heart attacks by keeping levels of triglycerides — a kind of fat in the blood — very low for a lifetime.

Pollin et al., *Science* (2008)

Apolipoprotein C3 (APOC3)







1 in 20 Amish
1 in 150 US/Europe

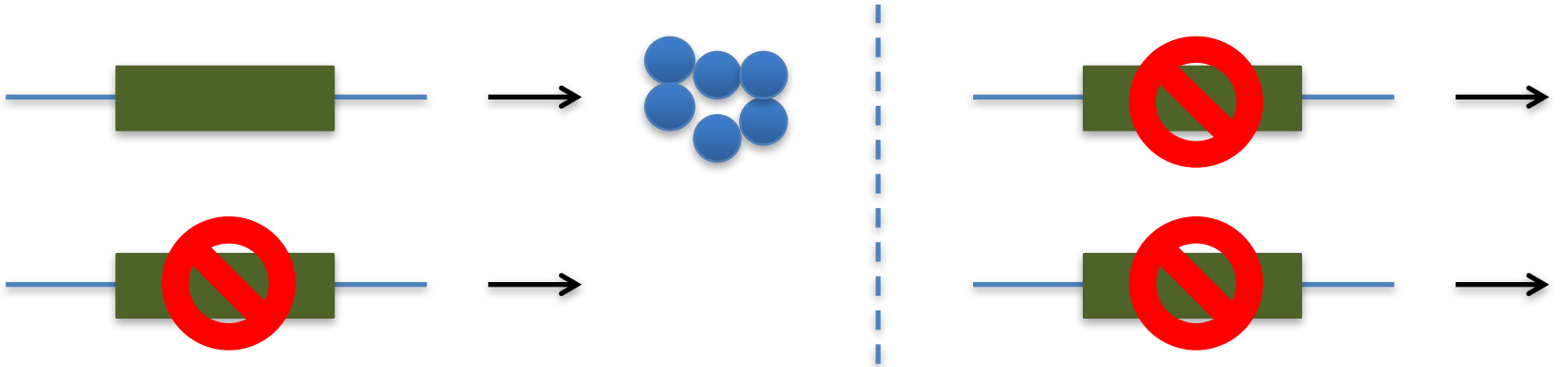
Null mutation in APOC3 gene

Lower triglycerides

40% lower risk MI

Pollin et al., *Science* (2008)
Crosby et al., *N Engl J Med* (2014)
Jorgensen et al., *N Engl J Med* (2014)

Can we identify a 'human knockout' for APOC3 and if so, healthy?



Heterozygous null

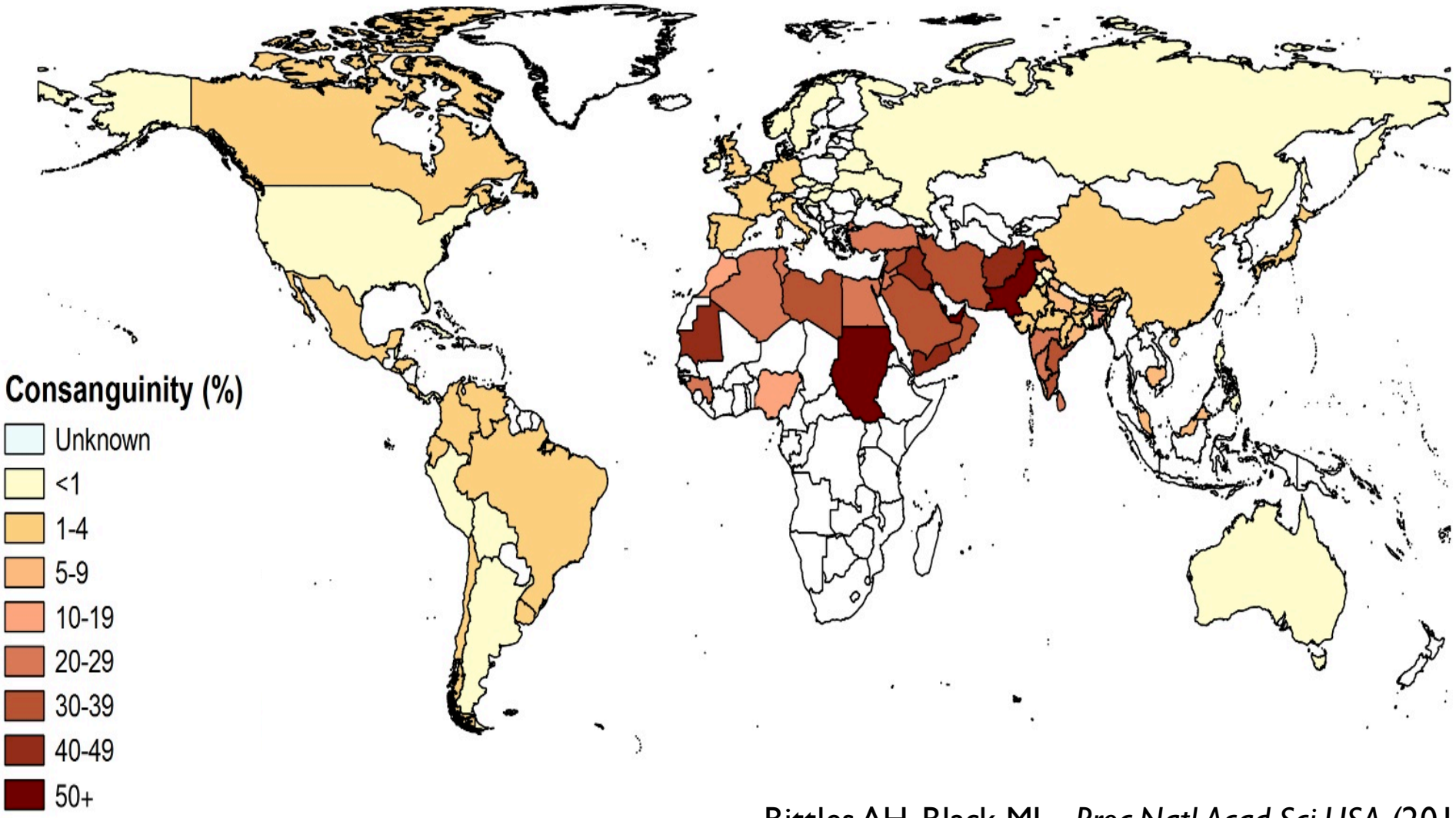
Homozygous null

'Knockout'

✔

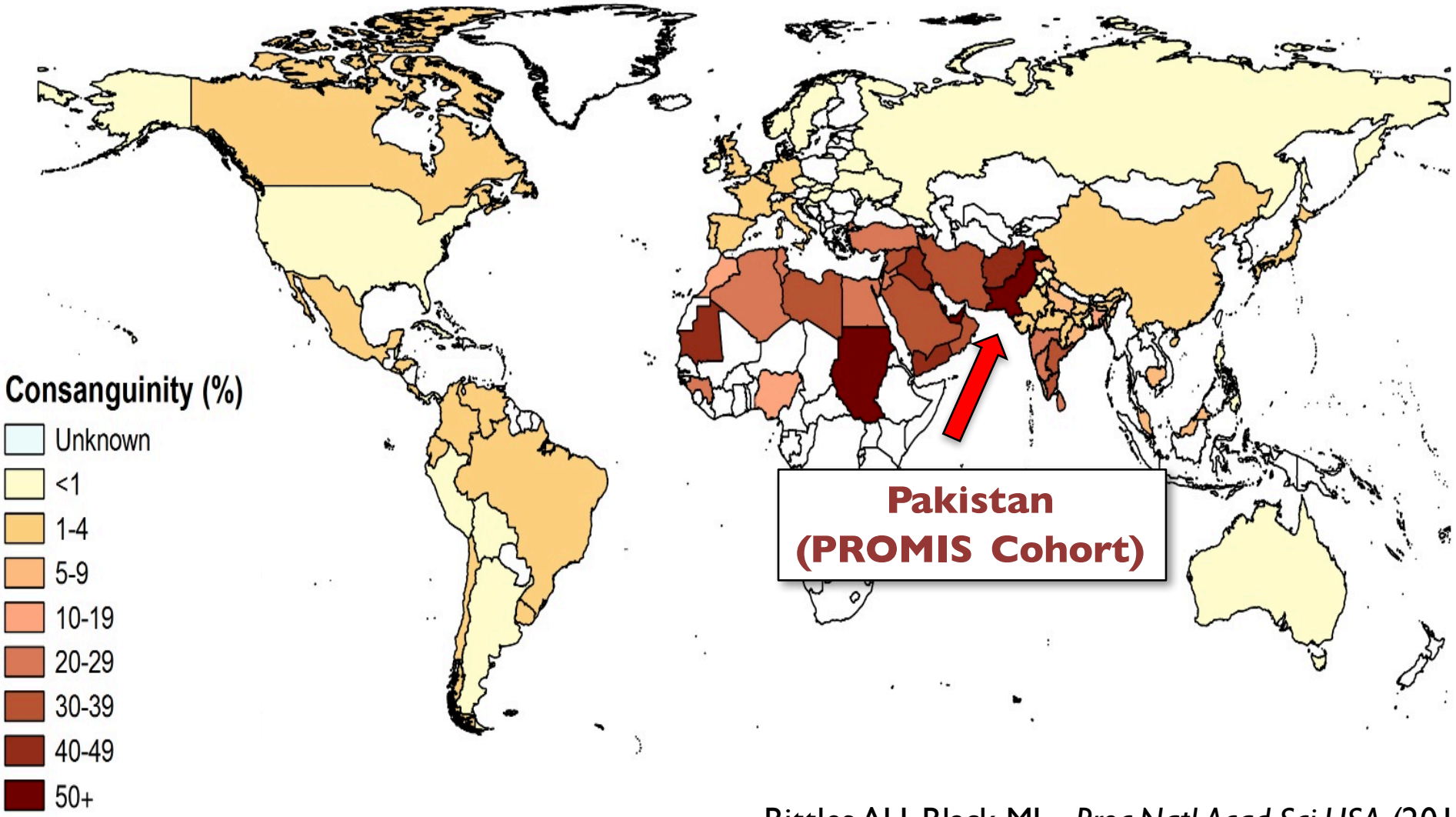
?

When parents are closely related (consanguinity), children more likely to be homozygous



Bittles AH, Black ML. *Proc Natl Acad Sci USA* (2010)

Global map of consanguinity

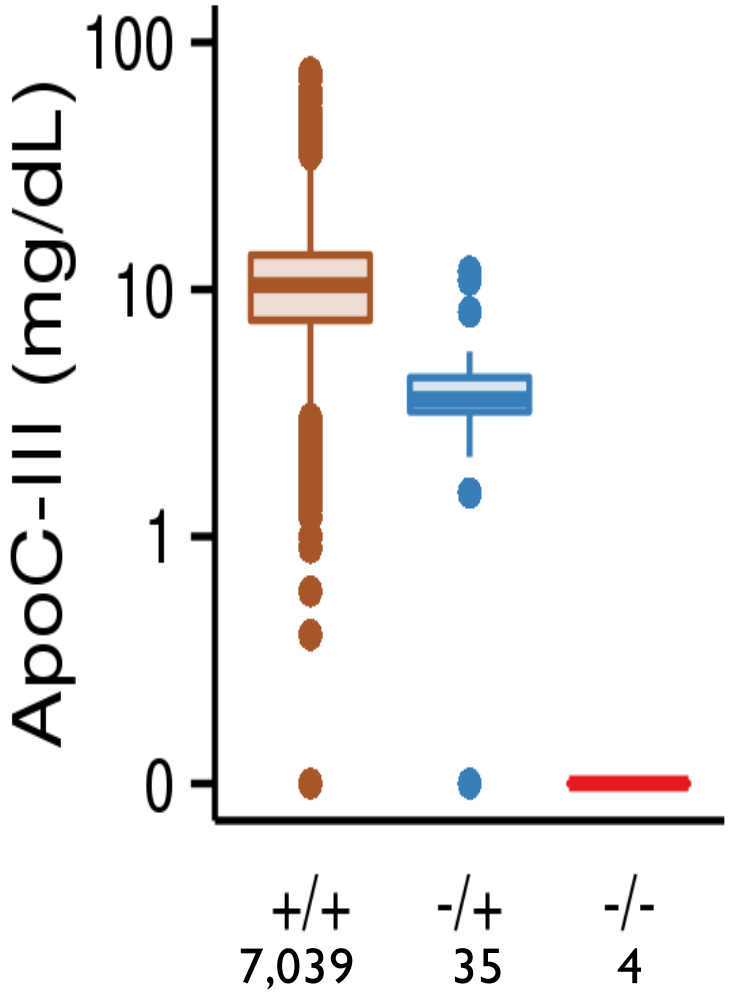


Bittles AH, Black ML. *Proc Natl Acad Sci USA* (2010)

In Pakistan, we have identified world's first *APOC3* homozygous nulls

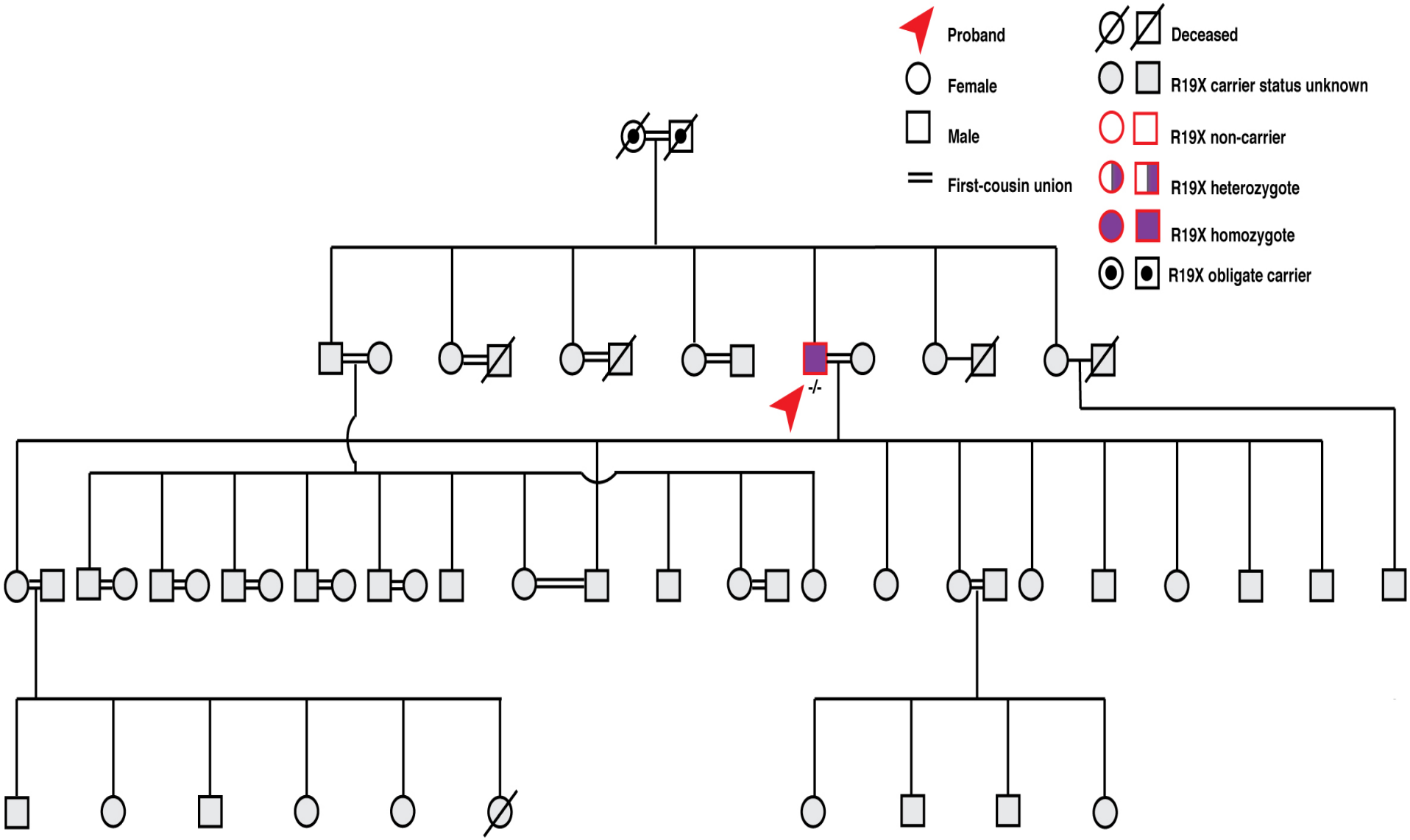
Danish Saleheen Pradeep Natarajan

John Danesh

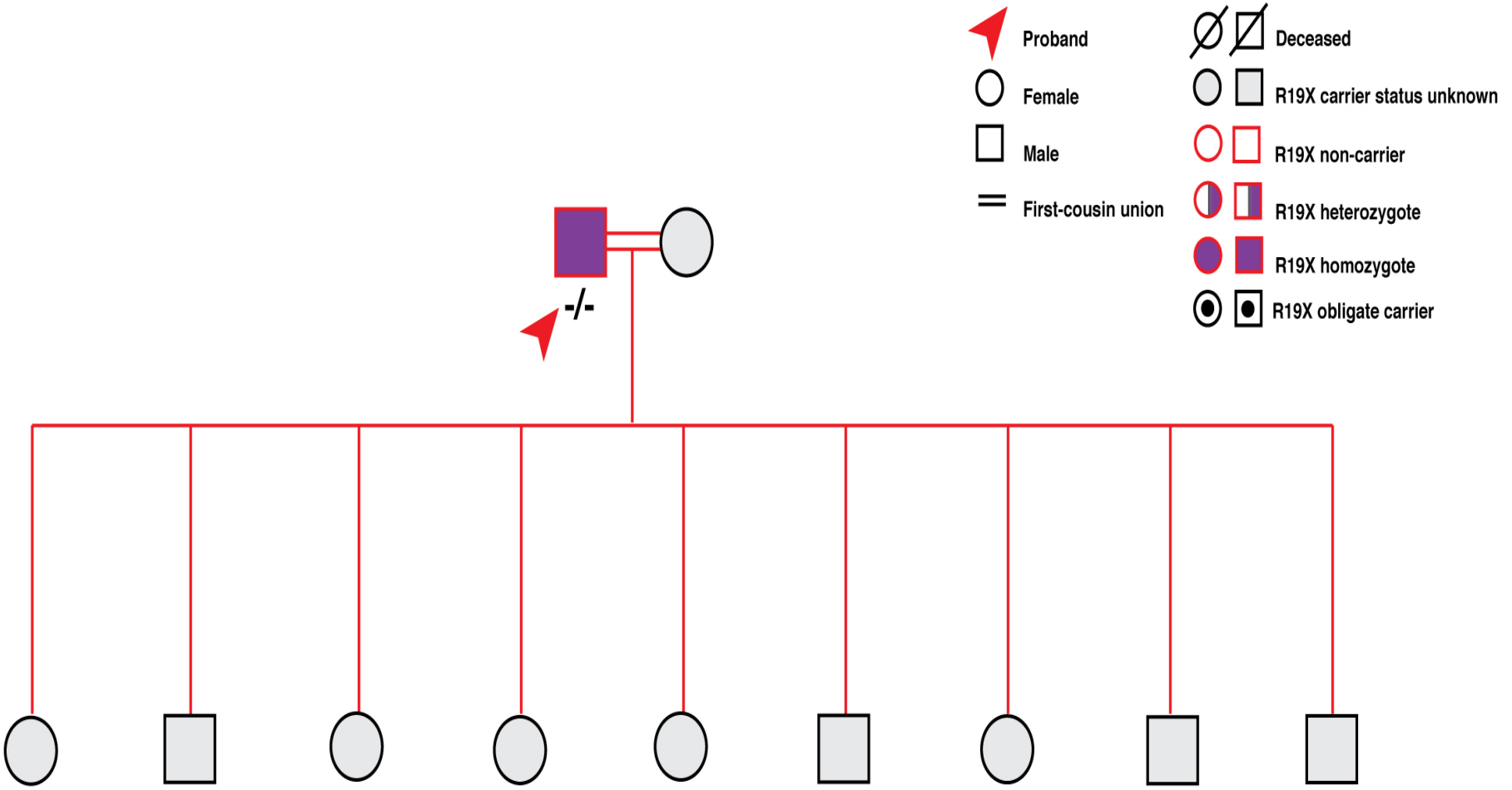


Saleheen*, Natarajan* et al., *Nature* (2017)

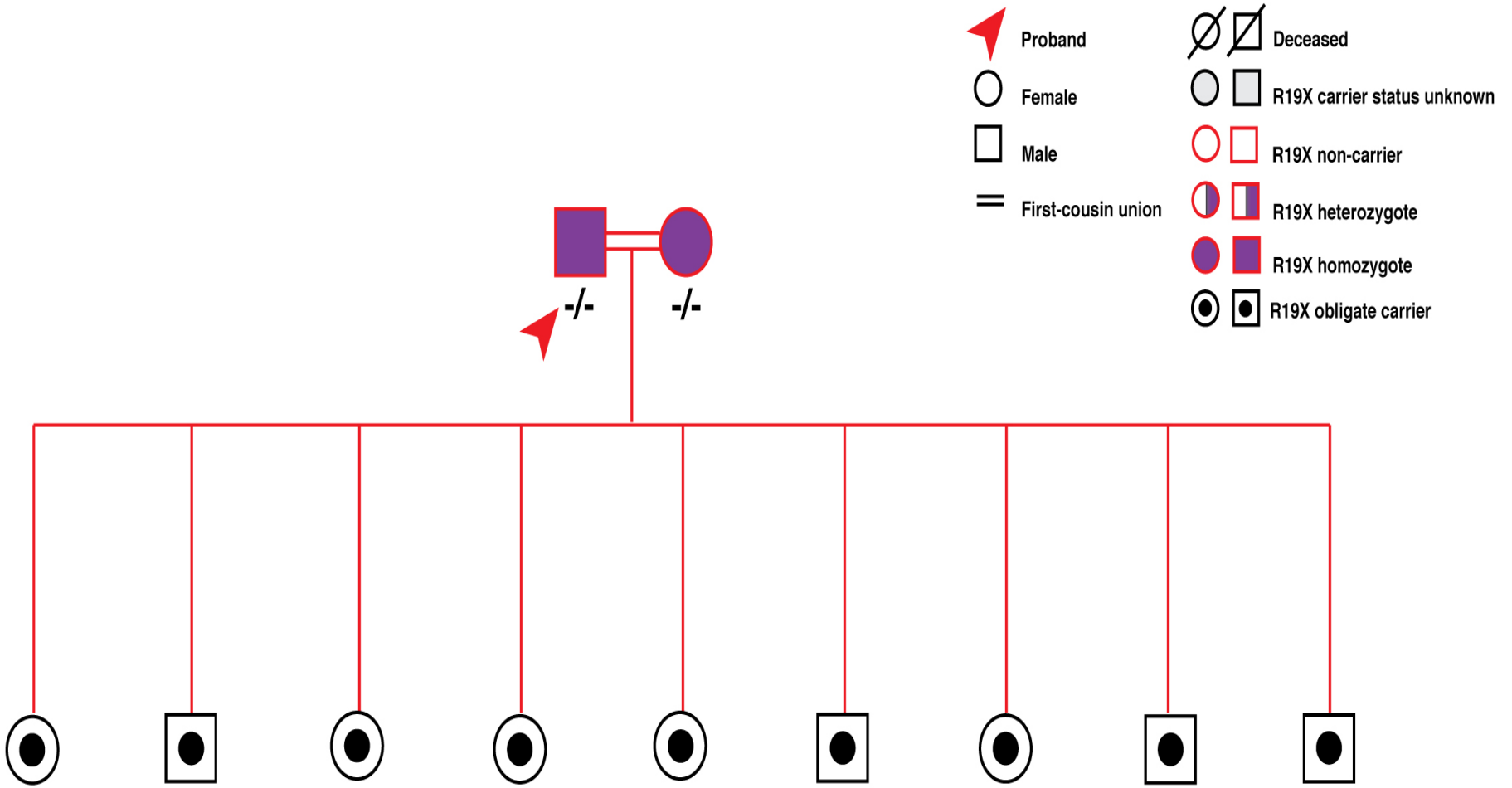
Family where 15 unions are between first cousins



We recruited this nuclear family for hypothesis-driven phenotyping



Surprise: both father and mother are completely defective for *APOC3* (homozygous null)



All nine children are obligate homozygote null

Why are APOC3 mutation carriers protected from heart attack?

People WITHOUT mutation

People WITH mutation

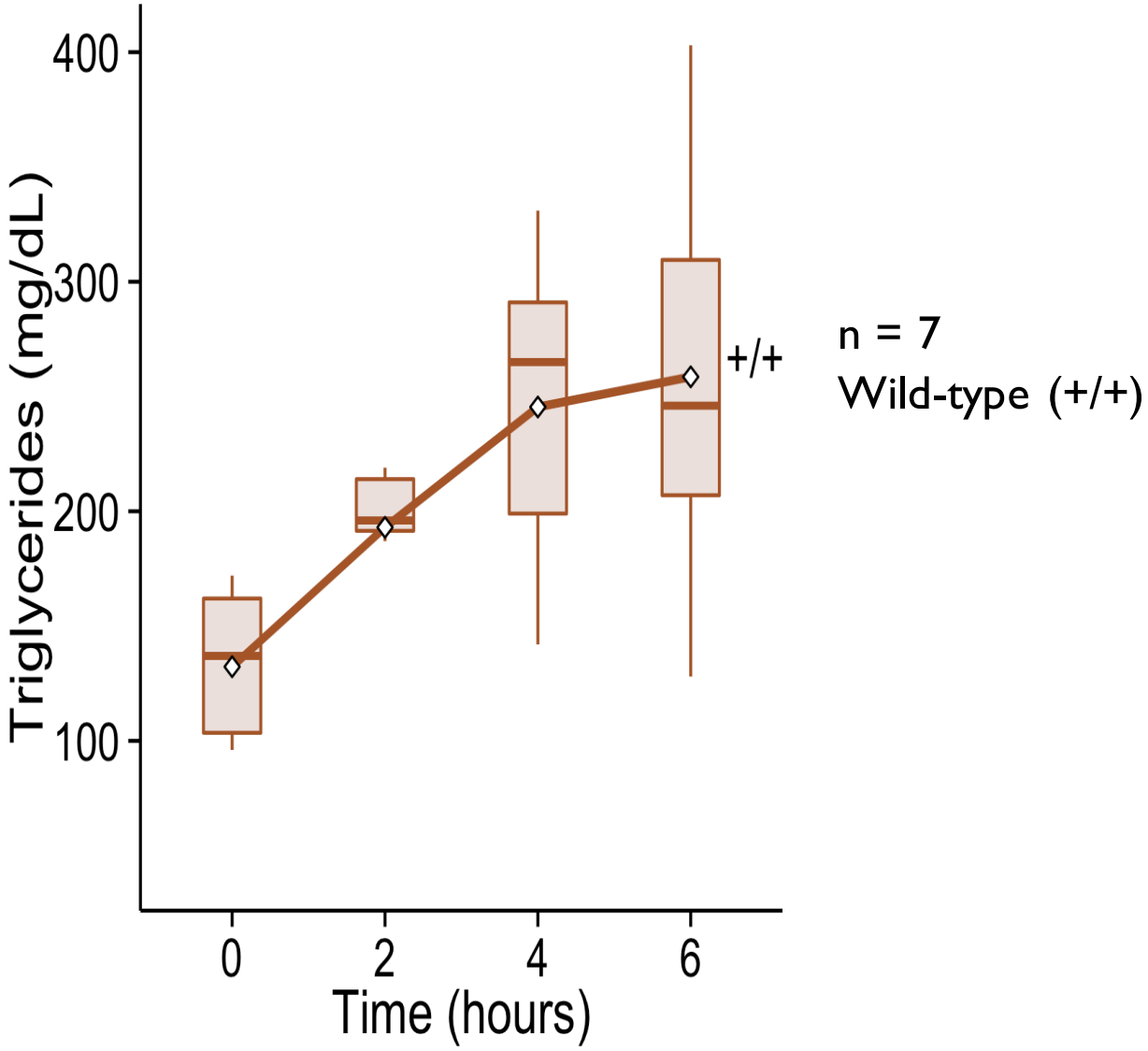
Oral fat challenge test in mutation non-carriers and carriers

People WITHOUT mutation

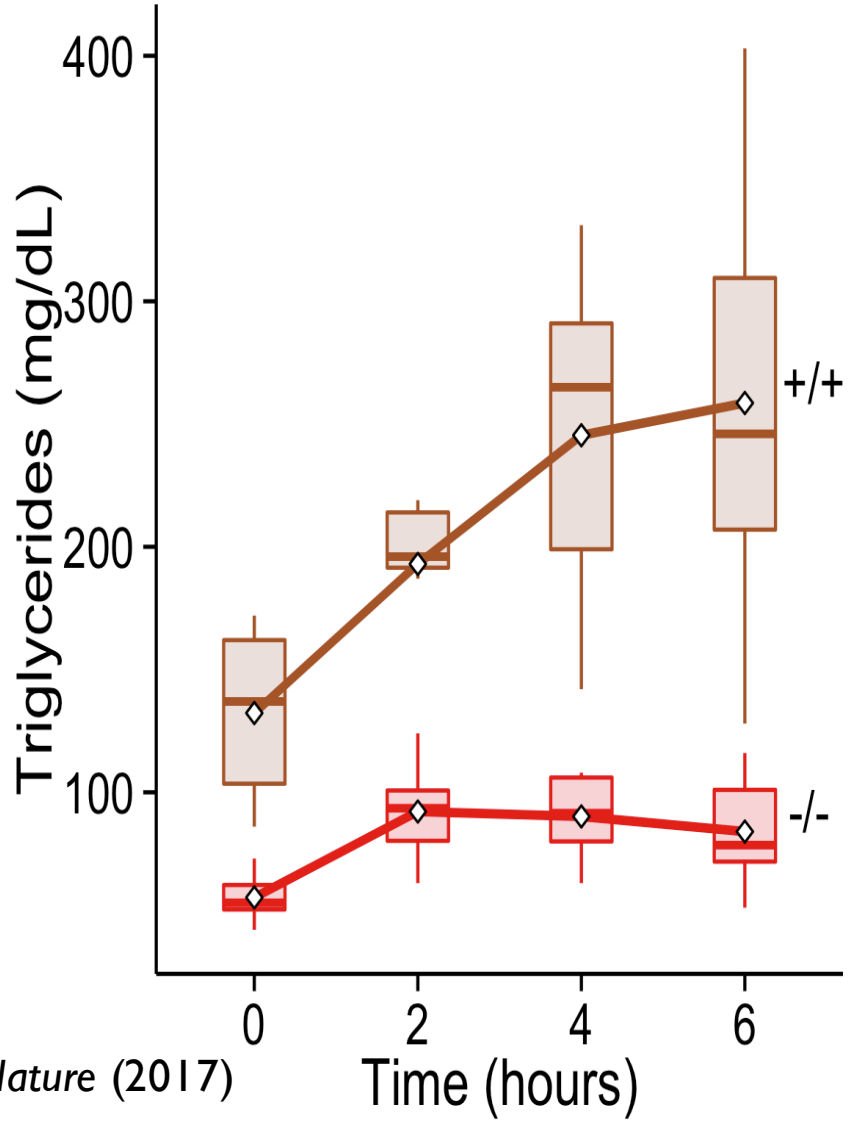
Dietary fat challenge

People WITH mutation

After a fat challenge in all of us, blood triglycerides rises



Complete absence of *APOC3* leads to much lower blood TG after meal

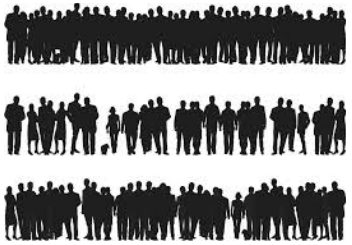


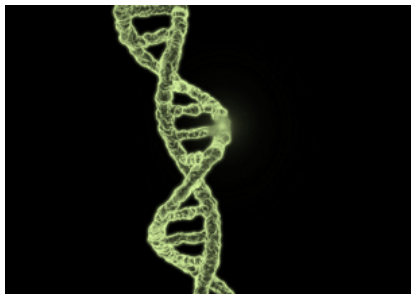
n = 7
Wild-type (+/+)

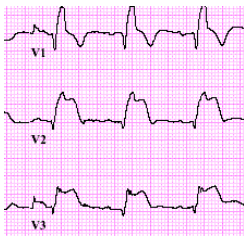
n = 6
Homozygous null (-/-)

Saleheen*, Natarajan* et al., *Nature* (2017)

Medicines now being developed to mimic APOC3 null mutation

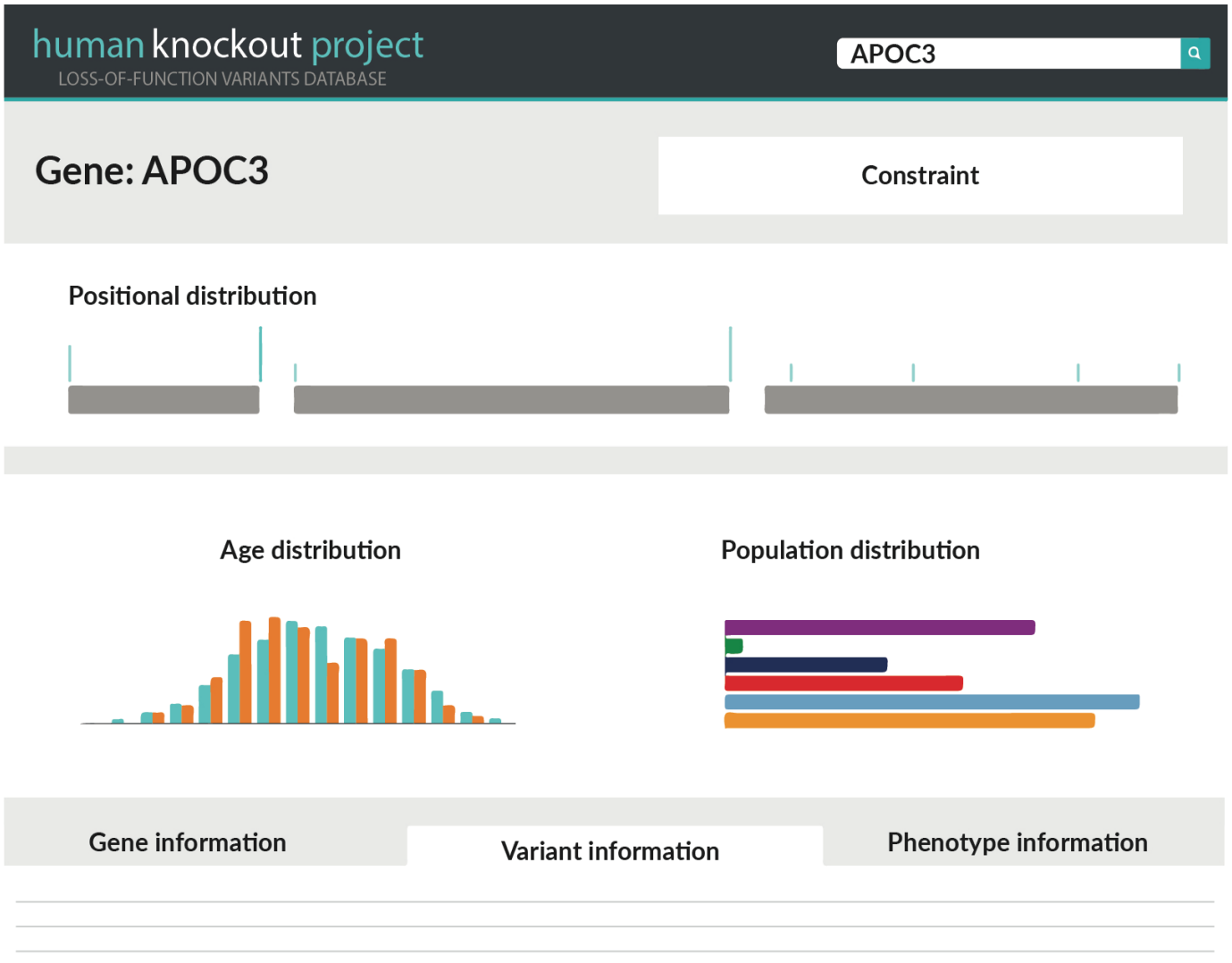






 **1 in 20 Amish** **1 in 150 US/Europe**		**Antisense oligonucleotide** (Ionis & Novartis)
Null mutation in APOC3 gene		**Block APOC3**
Lower triglycerides **40% lower risk MI**		**✓ Lower triglycerides** **? MI risk**

Imagine if:
a public resource of human knockouts



Daniel MacArthur

doi:10.1038/nature22034

Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity

Danish Saleheen[1,2]*, Pradeep Natarajan[3,4]*, Irina M. Armean[4,5], Wei Zhao[1], Asif Rasheed[2], Sumeet A. Khetarpal[6], Hong-Hee Won[7], Konrad J. Karczewski[4,5], Anne H. O'Donnell-Luria[4,5,8], Kaitlin E. Samocha[4,5], Benjamin Weisburd[4,5], Namrata Gupta[4], Mozzam Zaidi[2], Maria Samuel[2], Atif Imran[2], Shahid Abbas[9], Faisal Majeed[2], Madiha Ishaq[2], Saba Akhtar[2], Kevin Trindade[6], Megan Mucksavage[6], Nadeem Qamar[10], Khan Shah Zaman[10], Zia Yaqoob[10], Tahir Saghir[10], Syed Nadeem Hasan Rizvi[10], Anis Memon[10], Nadeem Hayyat Mallick[11], Mohammad Ishaq[12], Syed Zahed Rasheed[12], Fazal-ur-Rehman Memon[13], Khalid Mahmood[14], Naveeduddin Ahmed[15], Ron Do[16,17], Ronald M. Krauss[18], Daniel G. MacArthur[4,5], Stacey Gabriel[4], Eric S. Lander[4], Mark J. Daly[4,5], Philippe Frossard[2]§, John Danesh[19,20]§, Daniel J. Rader[6,21]§ & Sekar Kathiresan[3,4]§

Science

REPORTS

Cite as: V. M. Narasimhan *et al.*, *Science* 10.1126/science.aac8624 (2016).

Health and population effects of rare gene knockouts in adult humans with related parents

Vagheesh M. Narasimhan,[1] Karen A. Hunt,[2]* Dan Mason,[3]* Christopher L. Baker,[4]* Konrad J. Karczewski,[5,6]* Michael R. Barnes,[7] Anthony H. Barnett,[8] Chris Bates,[9] Srikanth Bellary,[10] Nicholas A. Bockett,[2] Kristina Giorda,[11] Christopher J. Griffiths,[2] Harry Hemingway,[12,13] Zhilong Jia,[7] M. Ann Kelly,[14] Hajrah A. Khawaja,[7] Monkol Lek,[5,6] Shane McCarthy,[1] Rosie McEachan,[3] Anne O'Donnell-Luria,[5,6] Kenneth Paigen,[4] Constantinos A. Parisinos,[2] Eamonn Sheridan,[3] Laura Southgate,[2] Louise Tee,[14] Mark Thomas,[1] Yali Xue,[1] Michael Schnall-Levin,[11] Petko M. Petkov,[4] Chris Tyler-Smith,[1] Eamonn R. Maher,[15,16] Richard C. Trembath,[2,17] Daniel G. MacArthur,[5,6] John Wright,[3] Richard Durbin,[1]†‡ David A. van Heel[2]†‡

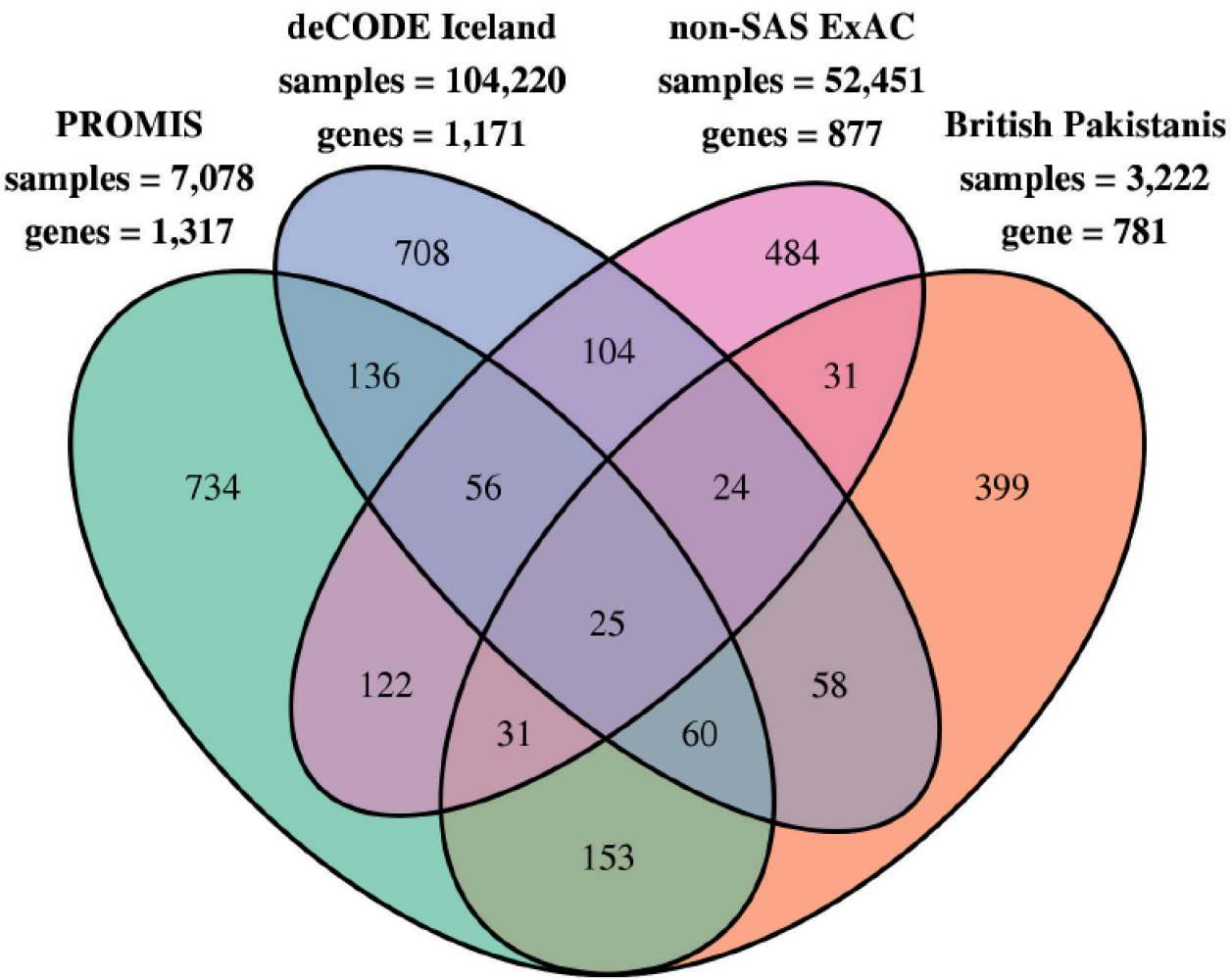
Identification of a large set of rare complete human knockouts

Patrick Sulem[1,6], Hannes Helgason[1,2,6], Asmundur Oddson[1], Hreinn Stefansson[1], Sigurjon A Gudjonsson[1], Florian Zink[1], Eirikur Hjartarson[1], Gunnar Th Sigurdsson[1], Adalbjorg Jonasdottir[1], Aslaug Jonasdottir[1], Asgeir Sigurdsson[1], Olafur Th Magnusson[1], Augustine Kong[1,2], Agnar Helgason[1,3], Hilma Holm[1,4], Unnur Thorsteinsdottir[1,5], Gisli Masson[1], Daniel F Gudbjartsson[1,2] & Kari Stefansson[1,5]

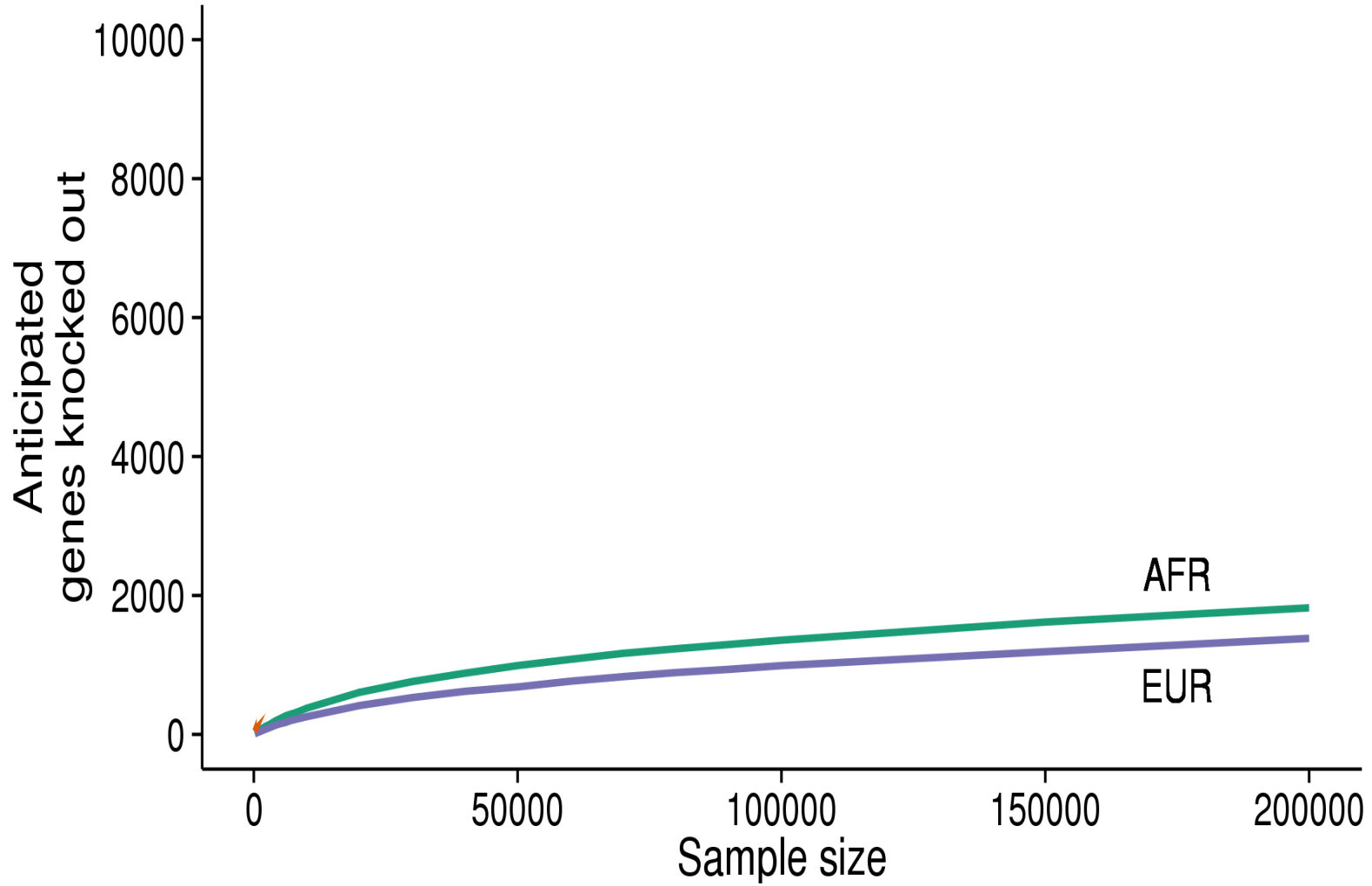
Analysis of protein-coding genetic variation in 60,706 humans

Monkol Lek[1,2,3,4], Konrad J. Karczewski[1,2*], Eric V. Minikel[1,2,5*], Kaitlin E. Samocha[1,2,5,6*], Eric Banks[2], Timothy Fennell[2], Anne H. O'Donnell-Luria[1,2,7], James S. Ware[2,8,9,10,11], Andrew J. Hill[1,2,12], Beryl B. Cummings[1,2,5], Taru Tukiainen[1,2], Daniel P. Birnbaum[2], Jack A. Kosmicki[1,2,6,13], Laramie E. Duncan[1,2,6], Karol Estrada[1,2], Fengmei Zhao[1,2], James Zou[2], Emma Pierce-Hoffman[1,2], Joanne Berghout[14,15], David N. Cooper[16], Nicole Deflaux[17], Mark DePristo[18], Ron Do[19,20,21,22], Jason Flannick[2,23], Menachem Fromer[1,6,19,20,24], Laura Gauthier[18], Jackie Goldstein[1,2,6], Namrata Gupta[2], Daniel Howrigan[1,2,6], Adam Kiezun[18], Mitja I. Kurki[2,25], Ami Levy Moonshine[18], Pradeep Natarajan[2,26,27,28], Lorena Orozco[29], Gina M. Peloso[2,27,28], Ryan Poplin[18], Manuel A. Rivas[2], Valentin Ruano-Rubio[18], Samuel A. Rose[6], Douglas M. Ruderfer[19,20,24], Khalid Shakir[18], Peter D. Stenson[16], Christine Stevens[2], Brett P. Thomas[1,2], Grace Tiao[18], Maria T. Tusie-Luna[30], Ben Weisburd[2], Hong-Hee Won[31], Dongmei Yu[6,25,27,32], David M. Altshuler[2,33], Diego Ardissino[34], Michael Boehnke[35], John Danesh[36], Stacey Donnelly[2], Roberto Elosua[37], Jose C. Florez[2,26,27], Stacey B. Gabriel[2], Gad Getz[18,26,38], Stephen J. Glatt[39,40,41], Christina M. Hultman[42], Sekar Kathiresan[2,26,27,28], Markku Laakso[43], Steven McCarroll[6,8], Mark I. McCarthy[44,45,46], Dermot McGovern[47], Ruth McPherson[48], Benjamin M. Neale[1,2,6], Aarno Palotie[1,2,5,49], Shaun M. Purcell[19,20,24], Danish Saleheen[50,51,52], Jeremiah M. Scharf[2,6,25,27,32], Pamela Sklar[19,20,24,53,54], Patrick F. Sullivan[55,56], Jaakko Tuomilehto[57], Ming T. Tsuang[58], Hugh C. Watkins[44,59], James G. Wilson[60], Mark J. Daly[1,2,6], Daniel G. MacArthur[1,2] & Exome Aggregation Consortium†

Thousands genes where KO human exists

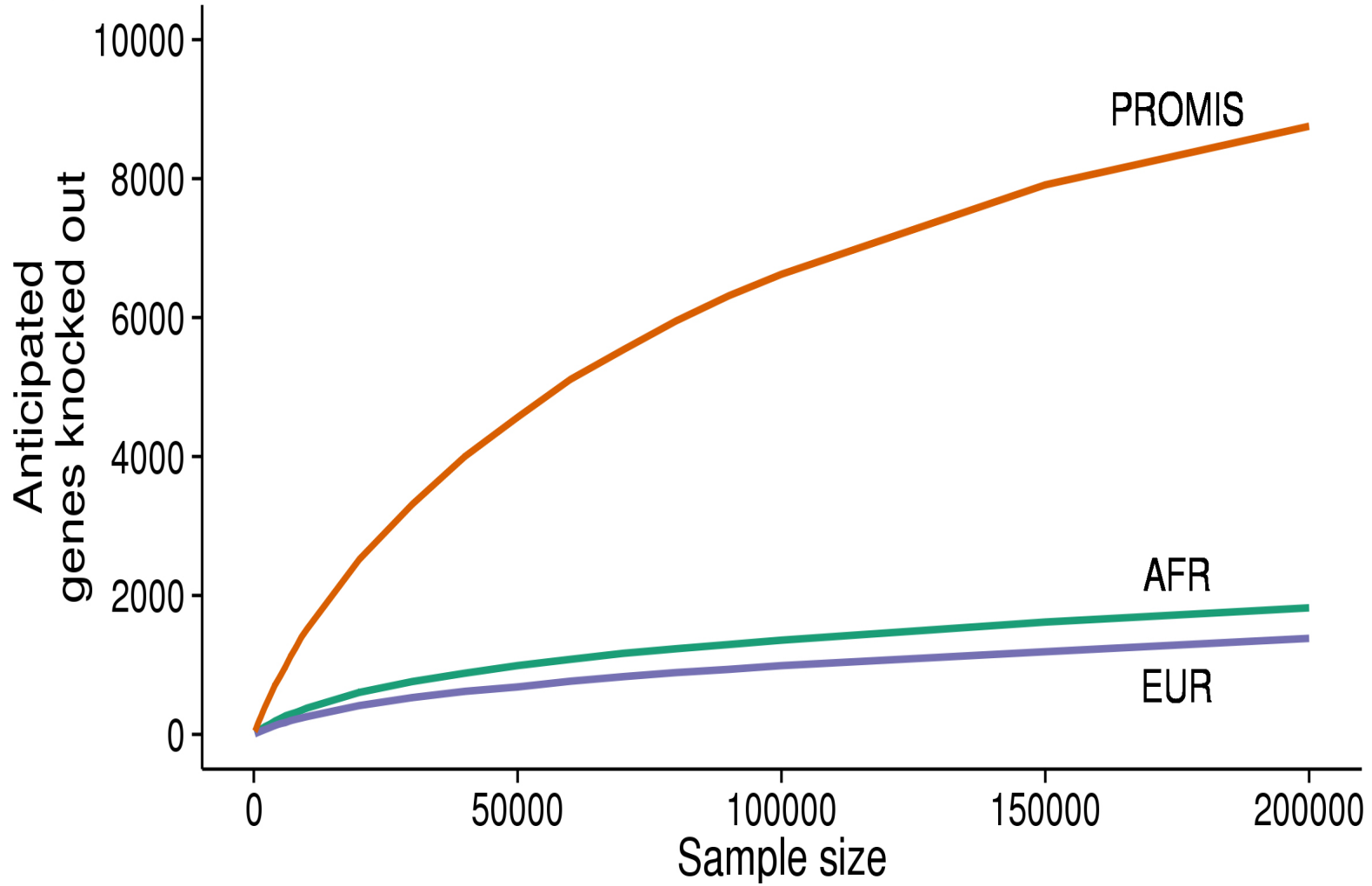


How many more null genes can we expect?



$$P(A_iA_i) = (1 - F) * cmaf_i^2 + F * cmaf_i$$

Up to ~8,000 genes homozygous null after sequencing 200,000 Pakistanis

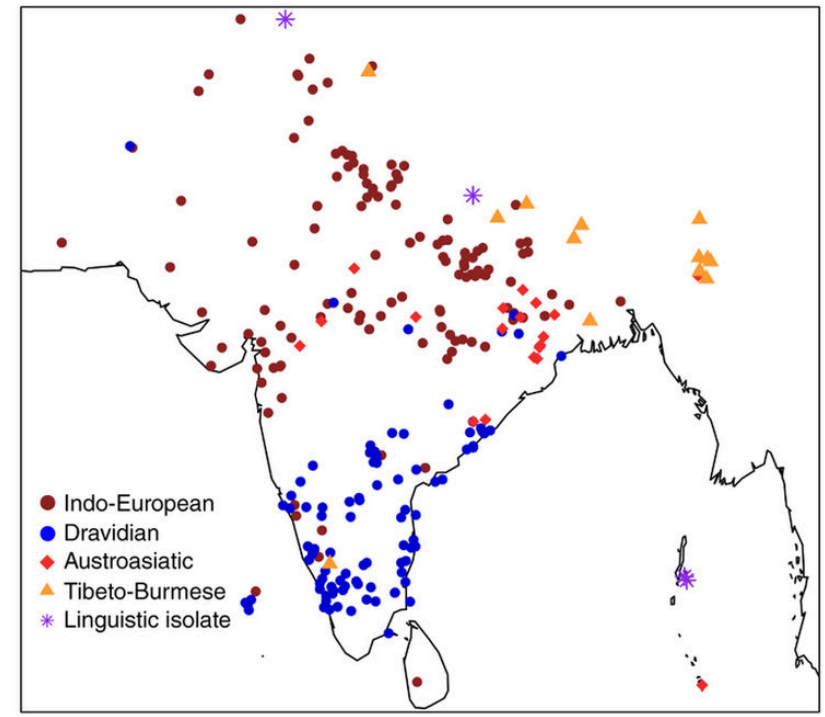


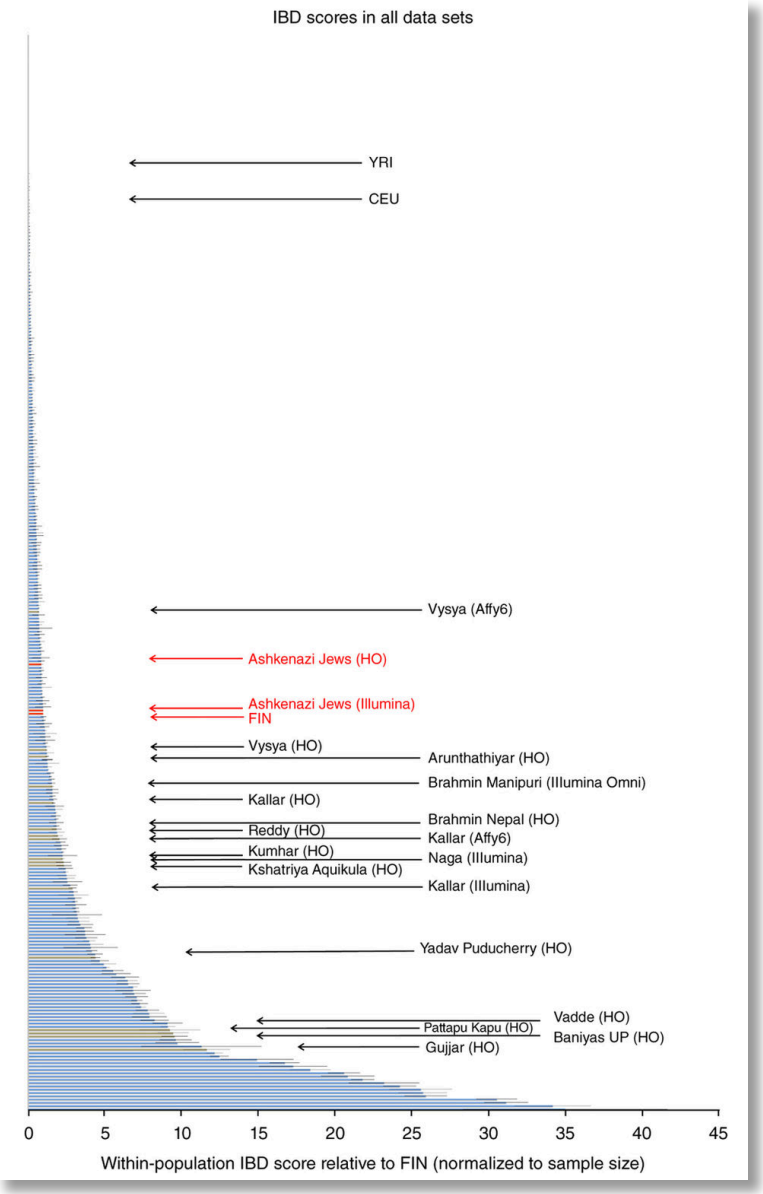
$$P(A_iA_i) = (1 - F) * cmaf_i^2 + F * cmaf_i$$

In India, 14 groups at >1M size with founder events more extreme than Finns, Azhkenazi Jews

The promise of discovering population-specific disease-associated genes in South Asia

Nathan Nakatsuka[1,2], Priya Moorjani[3,4], Niraj Rai[5,15], Biswanath Sarkar[6], Arti Tandon[1,4], Nick Patterson[4], Gandham SriLakshmi Bhavani[7], Katta Mohan Girisha[7], Mohammed S Mustak[8], Sudha Srinivasan[9], Amit Kaushik[10], Saadi Abdul Vahab[11], Sujatha M Jagadeesh[12], Kapaettu Satyamoorthy[11], Lalji Singh[13], David Reich[1,4,14,16] & Kumarasamy Thangaraj[5,16]





Two questions

Can we find mutations that protect against disease & develop medicines that mimic them?

Can we identify those at high genetic risk & offer preventive intervention?

Why do some suffer MI at a very early age?

Lifestyle

Polygenic

Monogenic

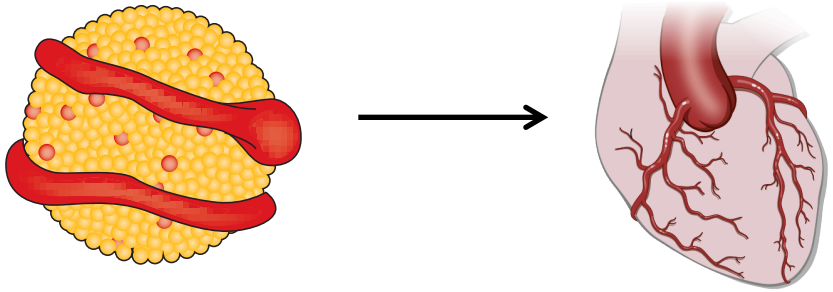
MI at age < 55

Age onset at MI

Traditional approach:
Genetic prediction focuses on rare, monogenic mutations

Traditional approach:
Genetic prediction focuses on rare, monogenic mutations

Familial hypercholesterolemia

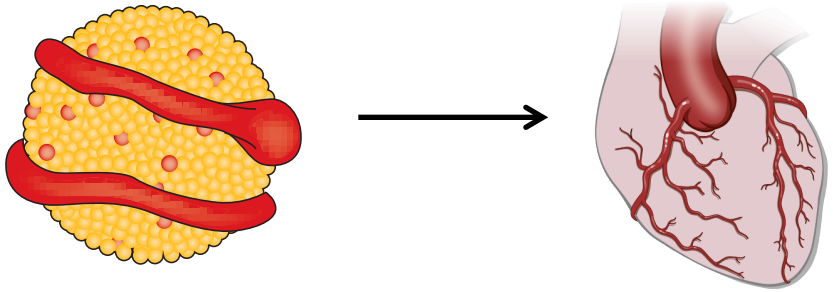


↑ **Cholesterol**

Heart attack
3x increased risk

Traditional approach:
Genetic prediction focuses on rare, monogenic mutations

Familial hypercholesterolemia



↑ Cholesterol

Heart attack
3x increased risk

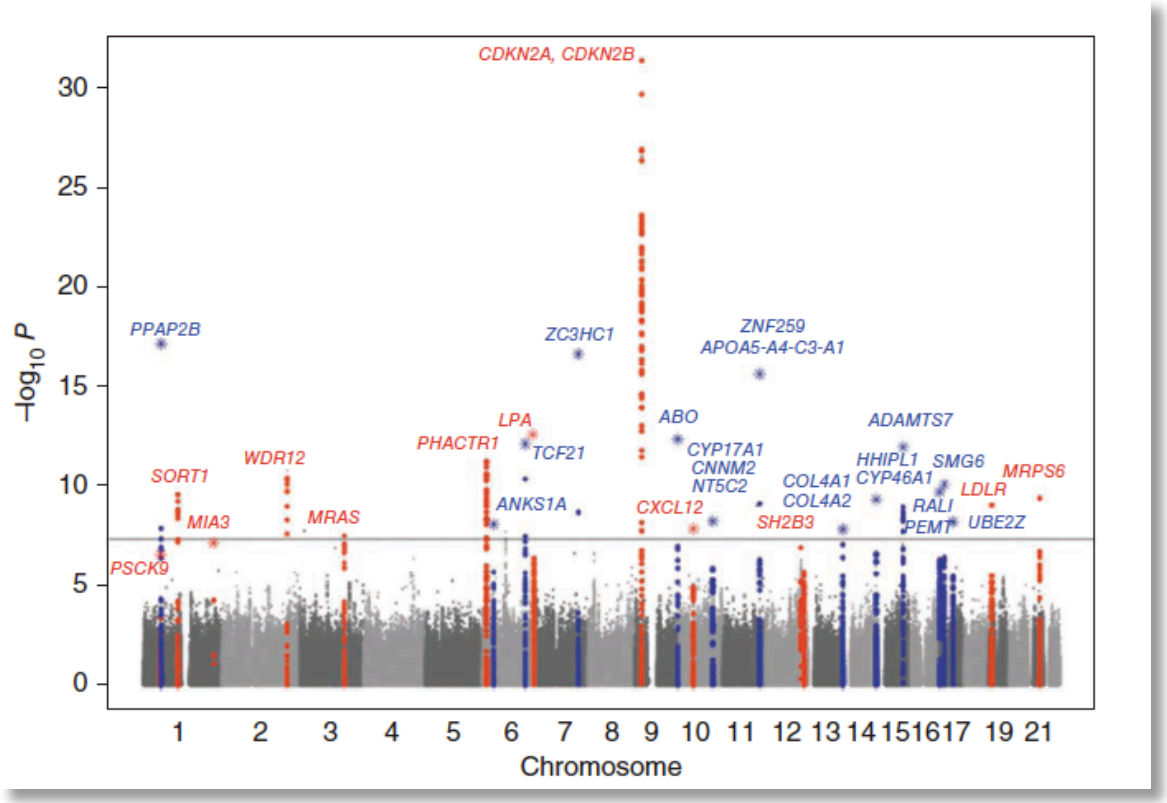
0.4% of the population

Question: Can we identify additional patients with a polygenic risk model?

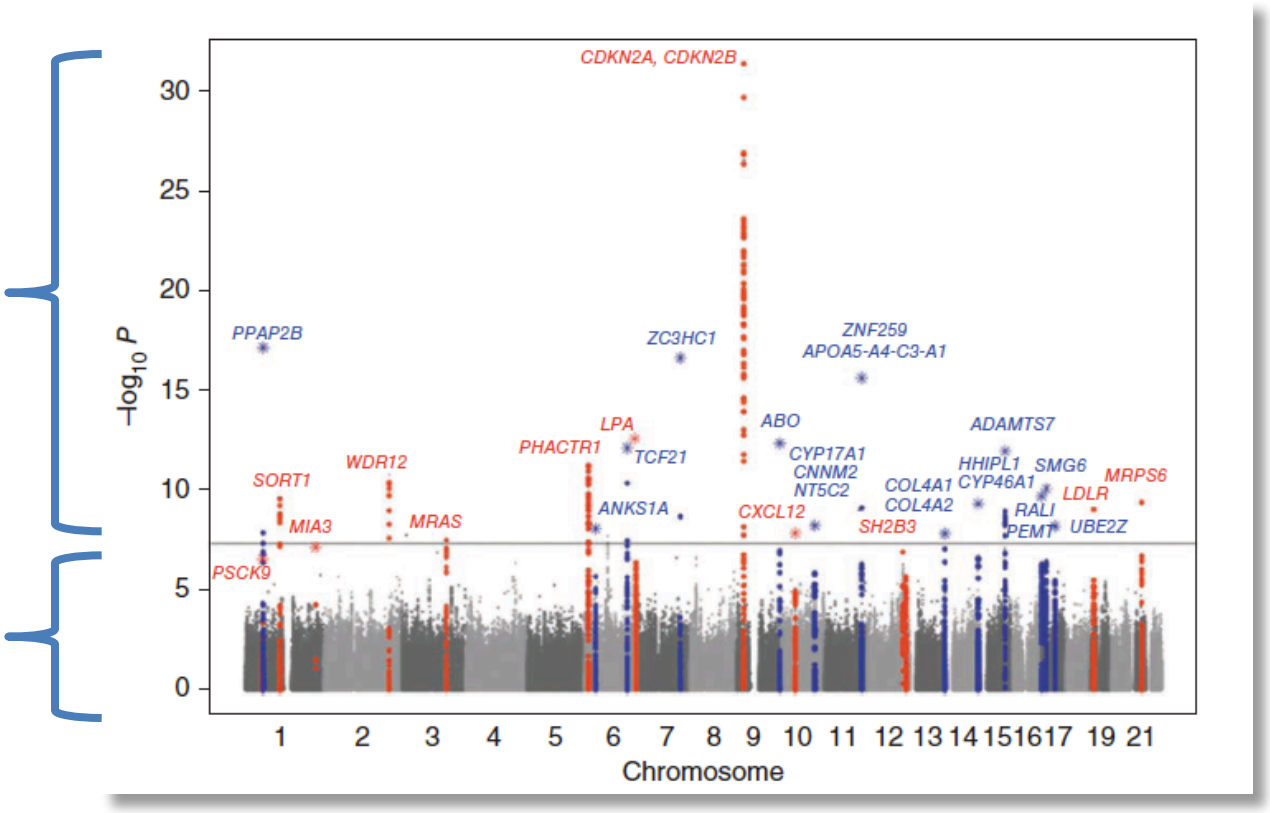
Concept: polygenic risk scores

'N' polymorphisms → 0, 1, or 2 copies of the risk allele → Score ranging from 0 to 2N for each person

Kathiresan, *N Engl J Med* (2008)
Ripatti, *Lancet* (2010)
Khera, *N Engl J Med* (2016)



Polygenic risk scores: move from top SNPs to a genome-wide set of 6.6M for prediction

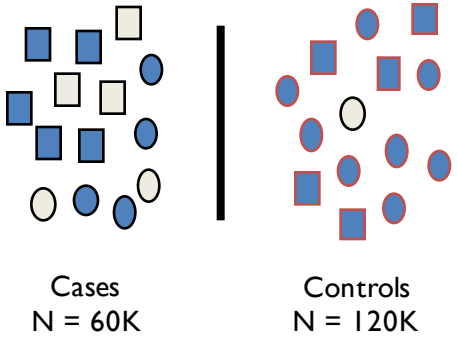


Khera*, Chaffin*, bioRxiv 2017

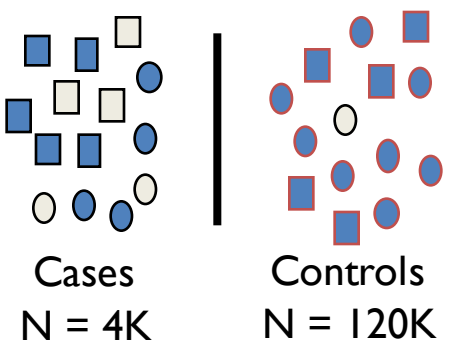
Amit V. Khera

Hypothesis: a polygenic score including a genome-wide set of SNPs can identify individuals with risk equivalent to a monogenic mutation

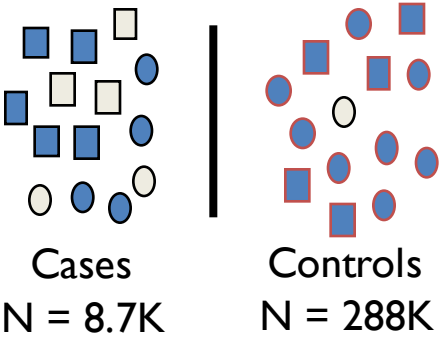
Step 1
6.6 million variants from genome-wide association study



Step 2
Testing
Dataset: 125K



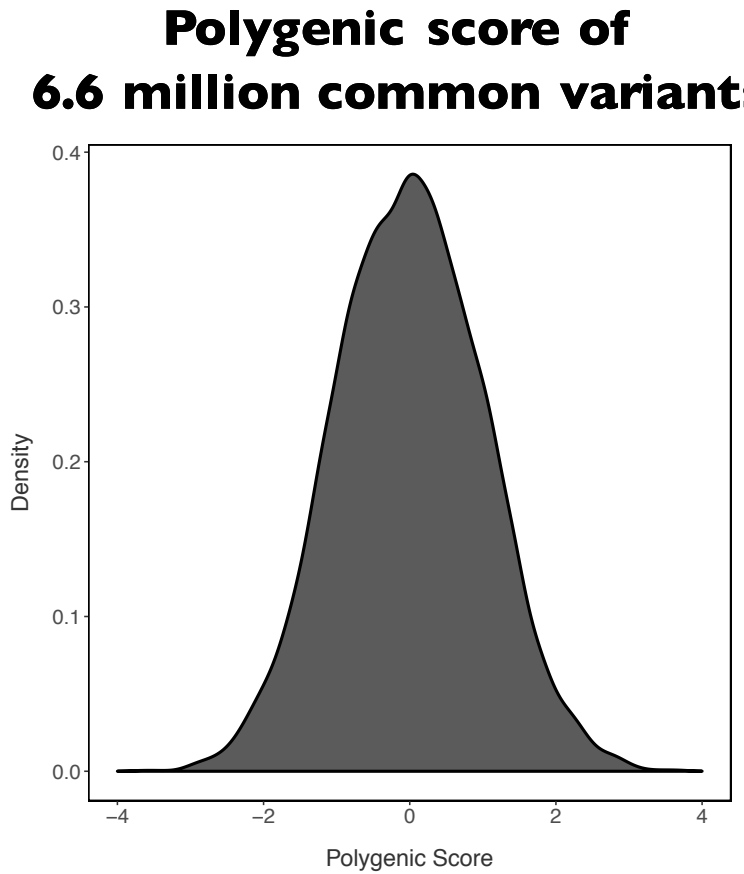
Step 3
Validation
Dataset: 300K



Genotypes: from arrays + imputation

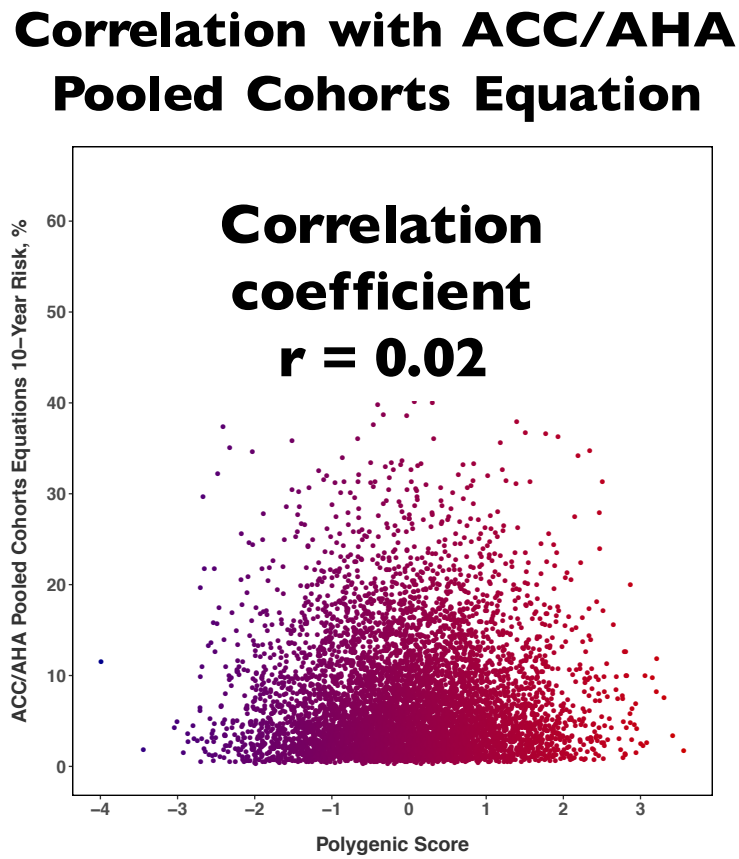
Khera*, Chaffin*, *bioRxiv* (2017)

A new quantitative metric of genetic liability to heart attack



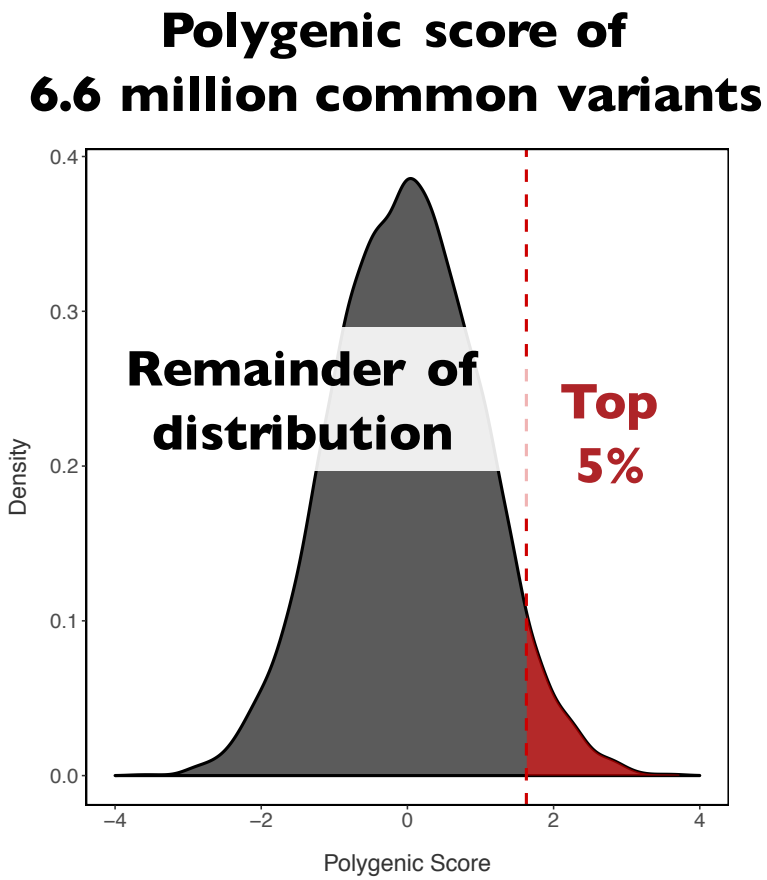
Khera*, Chaffin*, *bioRxiv* (2017)

Genome-wide polygenic score: little correlation with currently measured MI risk factors



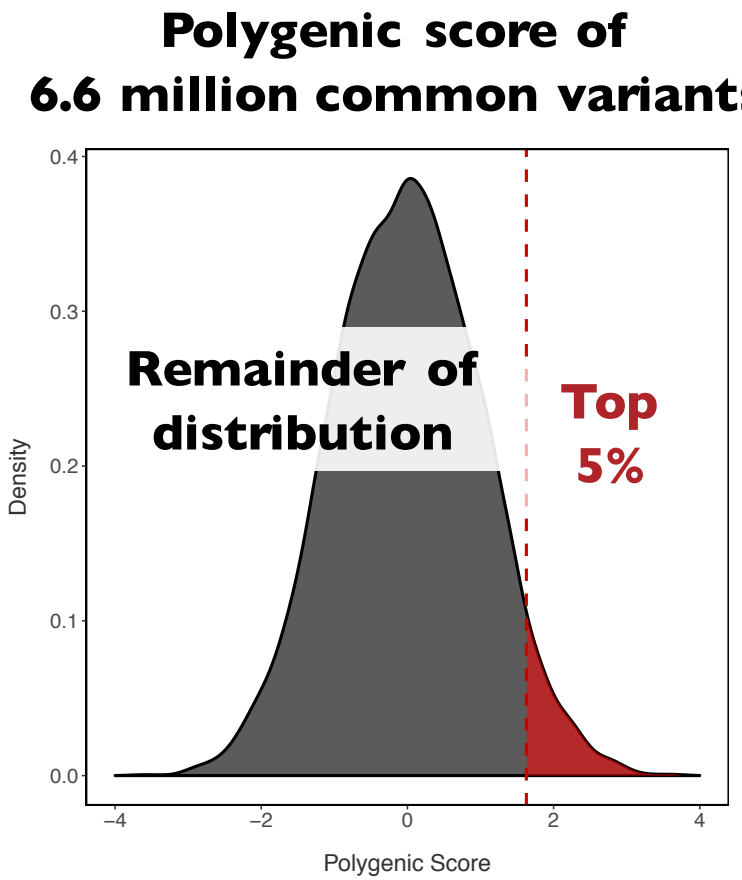
Khera*, Chaffin*, *bioRxiv* (2017)

What if we label top 5% tail of distribution as ‘carriers’ and remainder as ‘non-carriers’?



Khera*, Chaffin*, *bioRxiv* (2017)

1:20 participants, odds ratio 3.3 for MI when compared with all others



High polygenic score definition	Odds ratio
Top 5%	**3.3**
Top 1%	4.7

Khera*, Chaffin*, *bioRxiv* (2017)

Monogenic, polygenic contributions to early MI

	Monogenic	Polygenic
Prevalence in population	0.4%	5%
Odd ratio for MI	3.2	3.3
Mode of detection	↑ LDL cholesterol	**Currently UNAWARE**
Mechanism of risk	apoB lipoproteins	'Gamish'

Monogenic, polygenic contributions to early MI

	Monogenic	Polygenic
Prevalence in population	0.4%	5%
Odd ratio for MI	3.2	3.3
Mode of detection	↑ LDL cholesterol	**Currently UNAWARE**
Mechanism of risk	apoB lipoproteins	'Gamish'
Intervention	Lifestyle Medications	**?**

Is polygenic risk for MI modifiable?

Lifestyle

↓48%

Khera, *N Engl J Med* (2016)

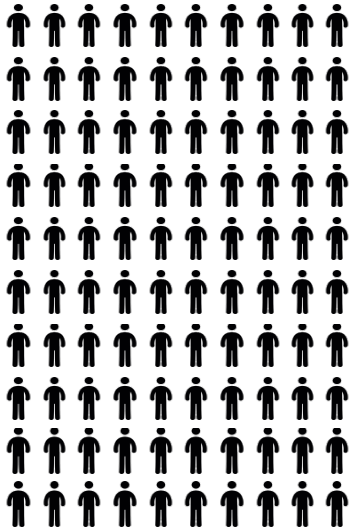
Medicines

↓44%

Mega*, Stitziel*, *Lancet* (2015)
Natarajan, *Circulation* (2017)

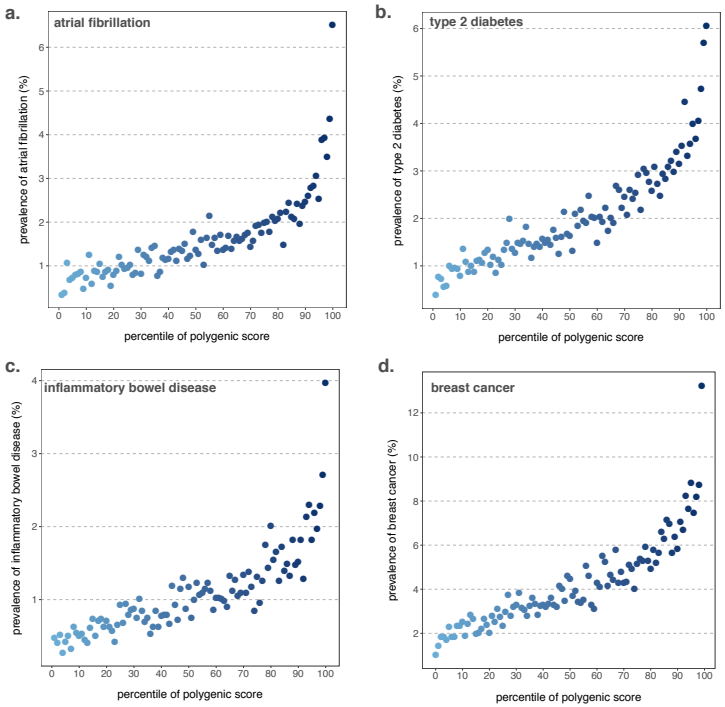
Summary:
Monogenic vs polygenic contribution to early MI

100 patients with early MI



	↑ Risk
Monogenic risk	**3-4 fold**
Polygenic risk	**3-4 fold**
Monogenic & polygenic	**6-8 fold**

Approach works for other common diseases . . . including those without monogenic risk factors



	% of population at >3-fold risk
Heart attack	**5.0%**
Atrial fibrillation	**6.2%**
Diabetes	**3.6%**
Inflammatory Bowel	**3.0%**
Breast cancer	**2.1%**

Potential for impact on clinical practice

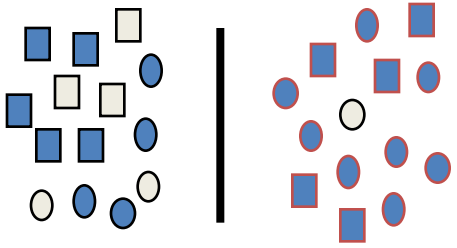
Why much better prediction now?

- Larger genome-wide association studies, more precise effect estimate for each variant
- Better computational methods to create genome-wide polygenic scores
- Larger cohorts to test and validate genome-wide polygenic scores (e.g., UK Biobank, 500K participants with GWAS data)

Will the MI polygenic score derived, validated in whites generalize to other ethnicities?

European ancestry

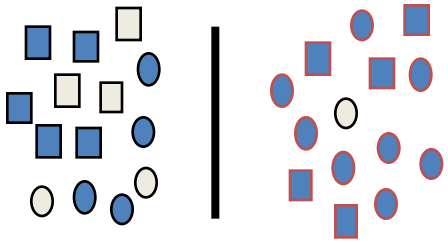
Step 1
6.6 million variants from genome-wide association study



European ancestry

Step 2
Testing
Dataset: 125K

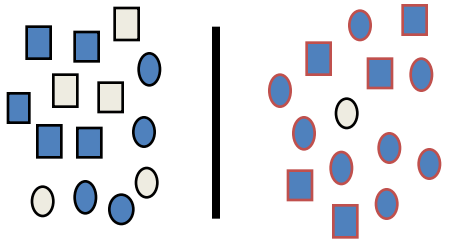
biobank uk



European ancestry

Step 3
Validation
Dataset: 300K

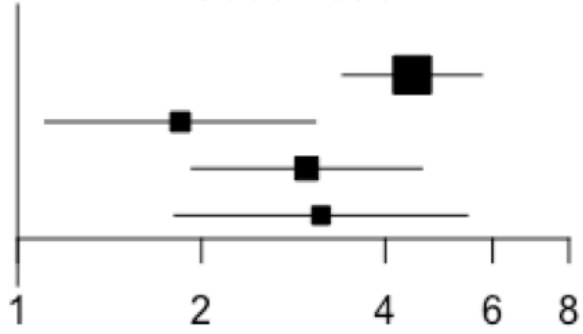
biobank uk



Khera*, Chaffin*, *bioRxiv* (2017)

Impact of high MI polygenic score by race

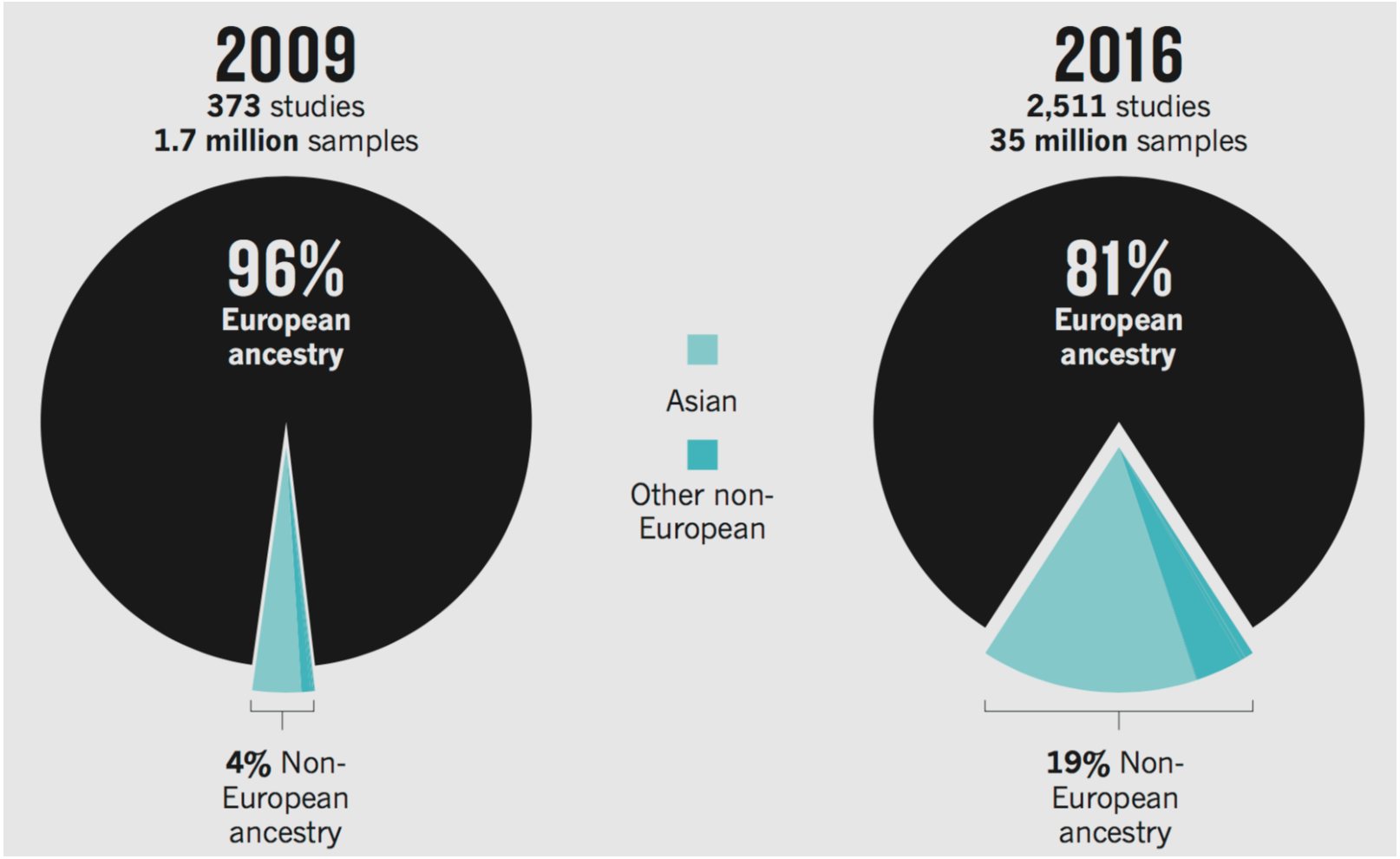
Race	N (%) Cases	N (%) Controls	Odds Ratio	95% CI	P-Value
White	(329 / 1537) 21%	(78 / 1544) 5%	4.43	[3.41; 5.75]	<0.0001
Black	(26 / 336) 8%	(49 / 962) 5%	1.85	[1.11; 3.07]	0.018
Asian	(44 / 328) 13%	(48 / 961) 5%	2.97	[1.93; 4.59]	<0.0001
Hispanic	(27 / 168) 16%	(38 / 751) 5%	3.14	[1.81; 5.45]	<0.0001



South Asians, UK Biobank (414 cases / 7203 controls): OR 3.06

Important to extend this approach to additional ancestral backgrounds—including undertaking or expanding GWAS in non-European ethnic groups

Major challenge: <20% published GWAS studies in ancestries outside Europe



Popejoy, Fullerton *Nature* (2016)

Summary: multi-ethnic studies

Opportunity

Gene variants providing unique biologic insights distributed <u>all over world</u>

Challenge

Polygenic scores: those of non-European ancestry could be left behind