

A misty landscape at sunrise or sunset. A river flows from the left towards the center, with a small wooden bridge or gate crossing it. The sky is a mix of blue and orange, and the ground is covered in grass and trees, some of which are shrouded in mist.

The Genome of the Netherlands Project



UMC Utrecht
Center for Molecular Medicine

Paul de Bakker

“A large number of Mendelian factors”

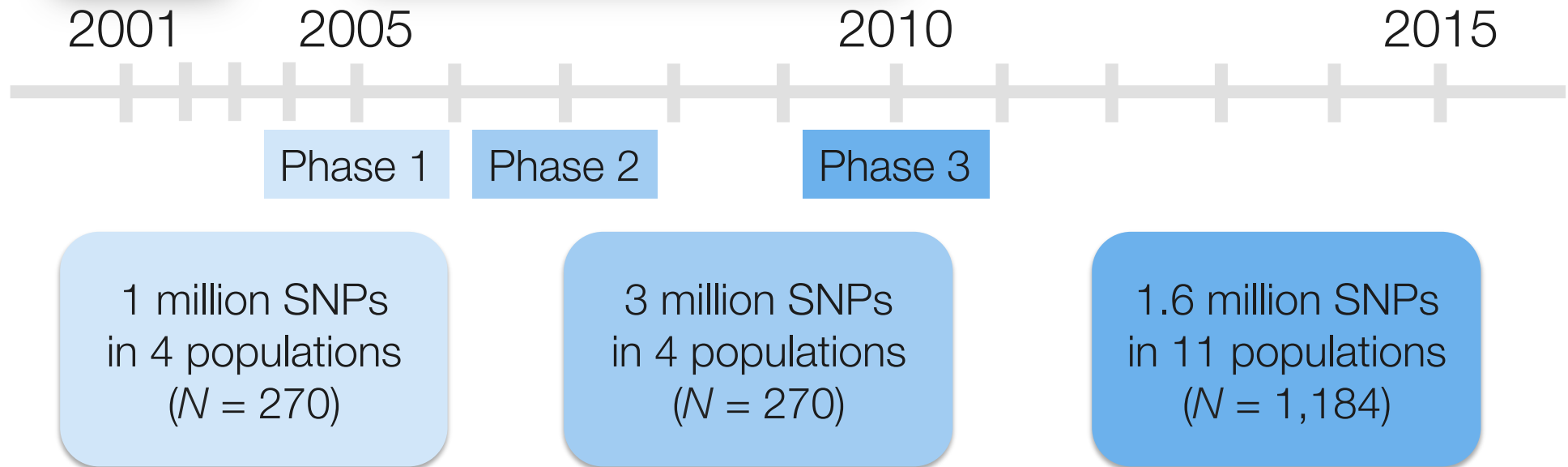
It is not sufficient to ascribe this last residue to the effects of environment. Numerous investigations by GALTON and PEARSON have shown that all measurable environment has much less effect on such measurements as stature. Further, the facts collected by GALTON respecting identical twins show that in this case, where the essential nature is the same, the variance is far less. The simplest hypothesis, and the one which we shall examine, is that such features as stature are determined by a large number of Mendelian factors, and that the large variance among children of the same parents is due to the segregation of those factors in respect to which the parents are heterozygous. Upon this hypothesis we will attempt to determine how much more of the variance, in different measurable features, beyond that which is indicated by the fraternal correlation, is due to innate and heritable factors.

Ronald Fisher. The correlation between relatives on the supposition of Mendelian inheritance. 1918

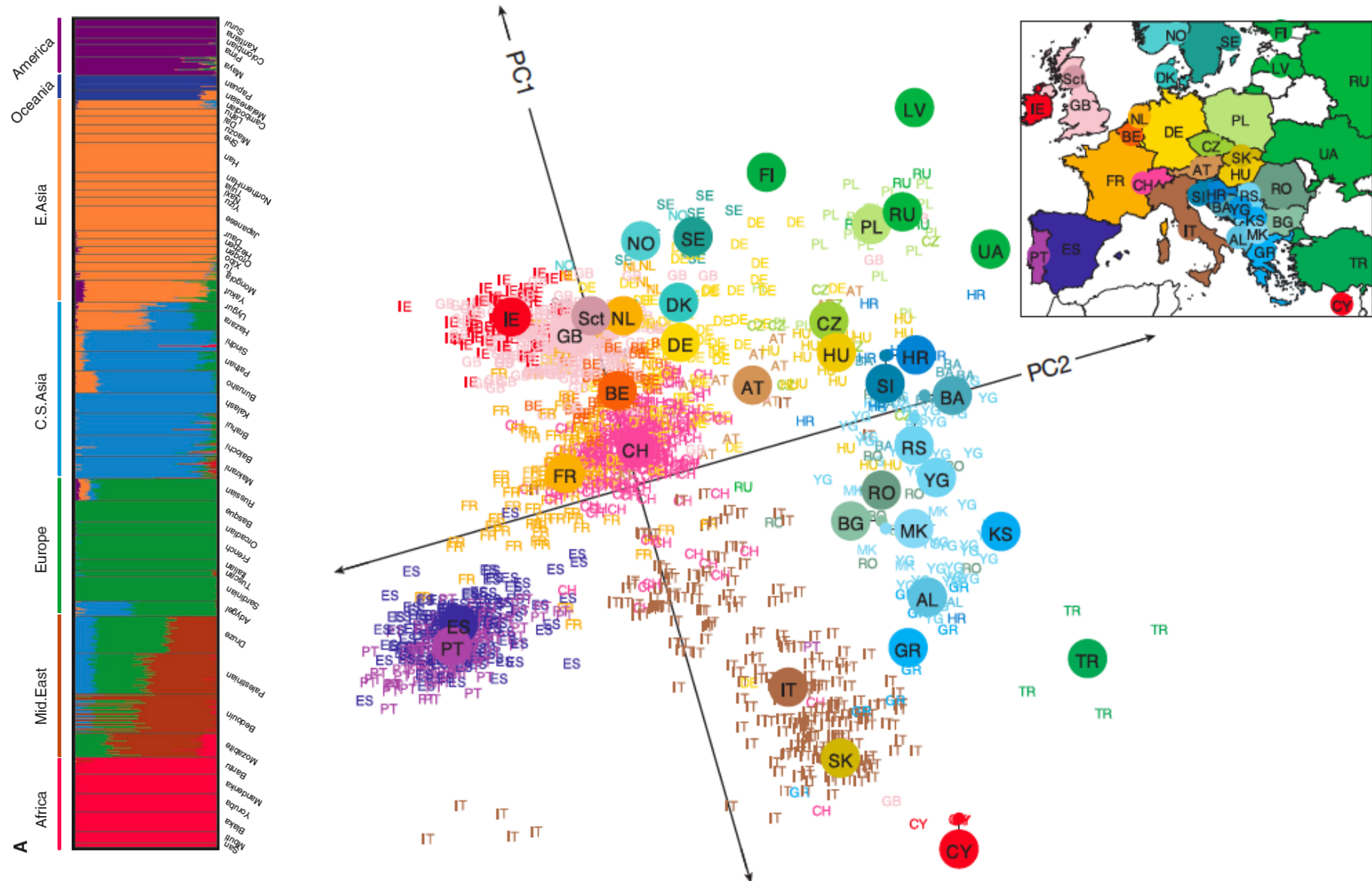
Common variation



Characterization of DNA polymorphism



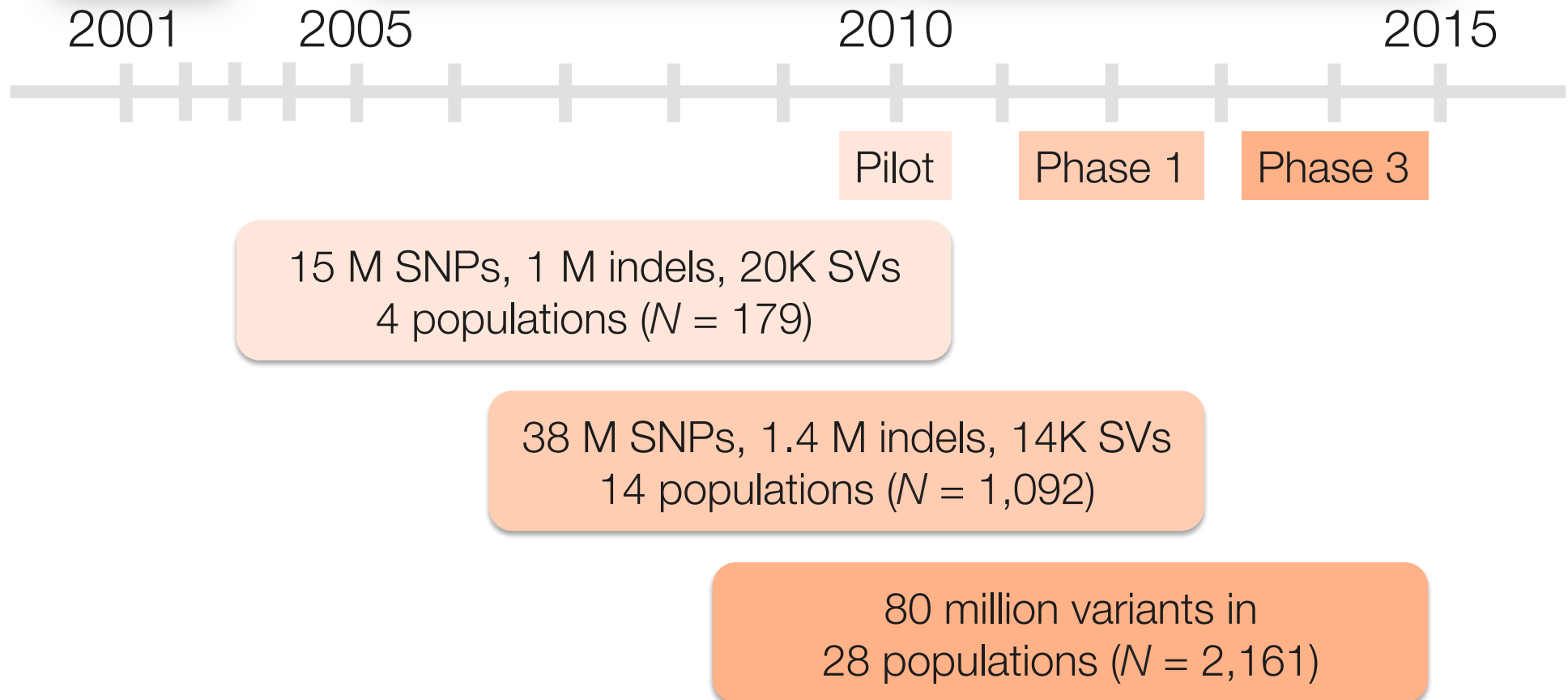
Differentiation across populations



Li et al. Science. 2008

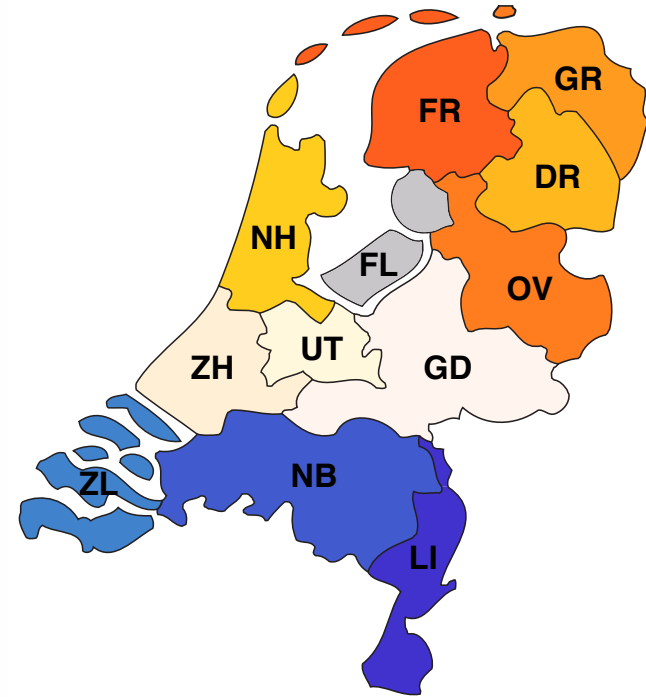
Novembre et al. Nature 2008

Characterization of DNA polymorphism



nature
genetics

VOLUME 46 NUMBER 8 AUGUST 2014
www.nature.com/naturegenetics



Whole-genome sequence variation, population structure and demographic history of the Dutch population

The Genome of the Netherlands Consortium*

Dutch genomes
Cattle genomes
Marmoset genome

Go • NL
GENOME of the NETHERLANDS

The Genome of the Netherlands

231 trios

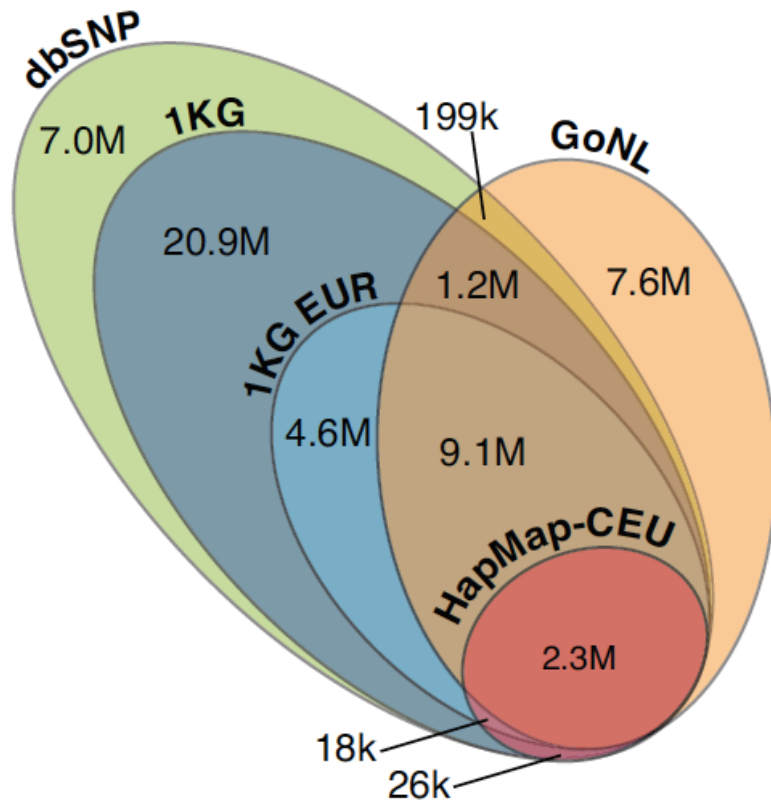
11 quartets
(monozygotic twins)

8 quartets
(dizygotic twins)

Whole-genome sequencing
at medium depth (~13x)



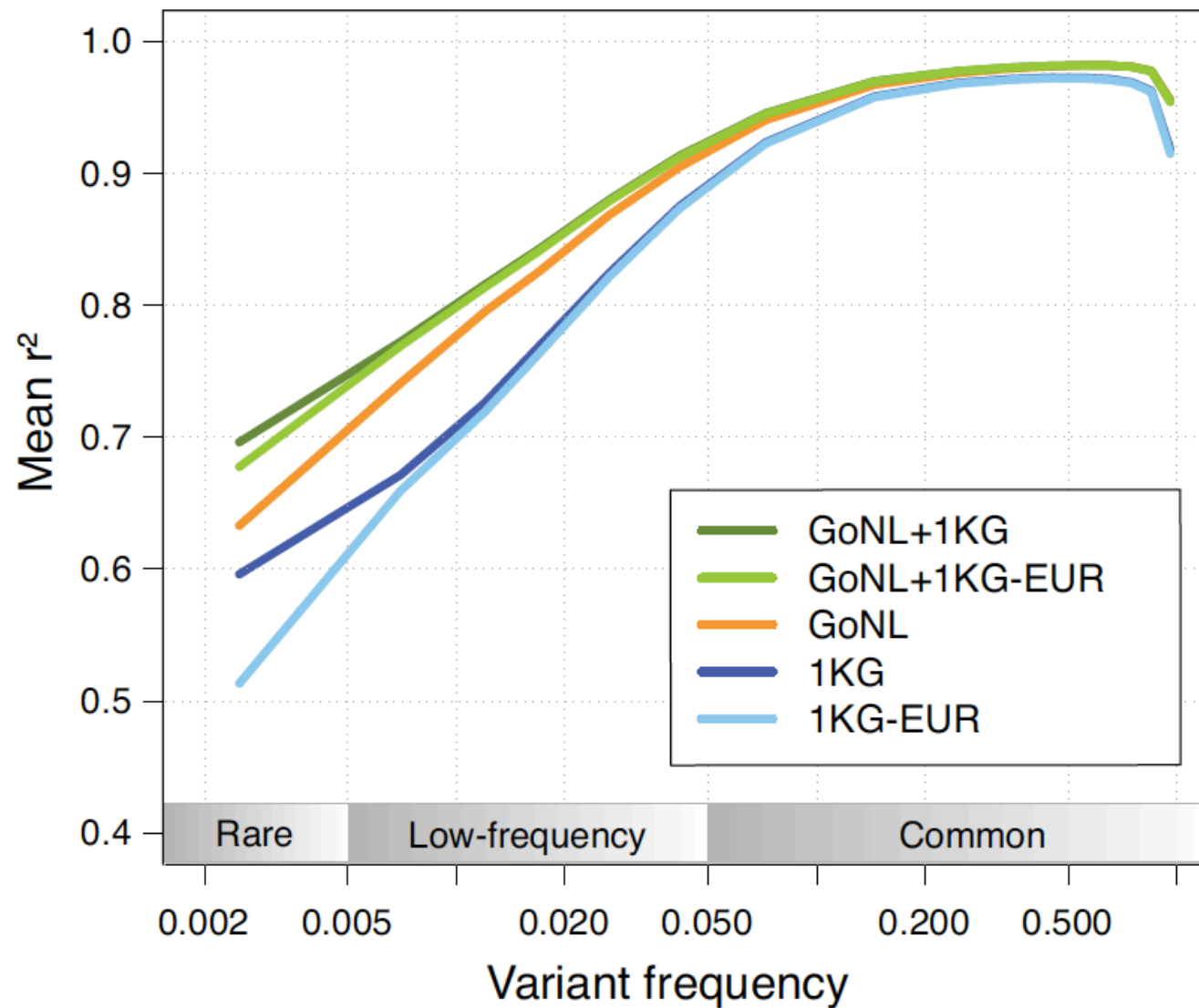
Variation discovery (SNVs)

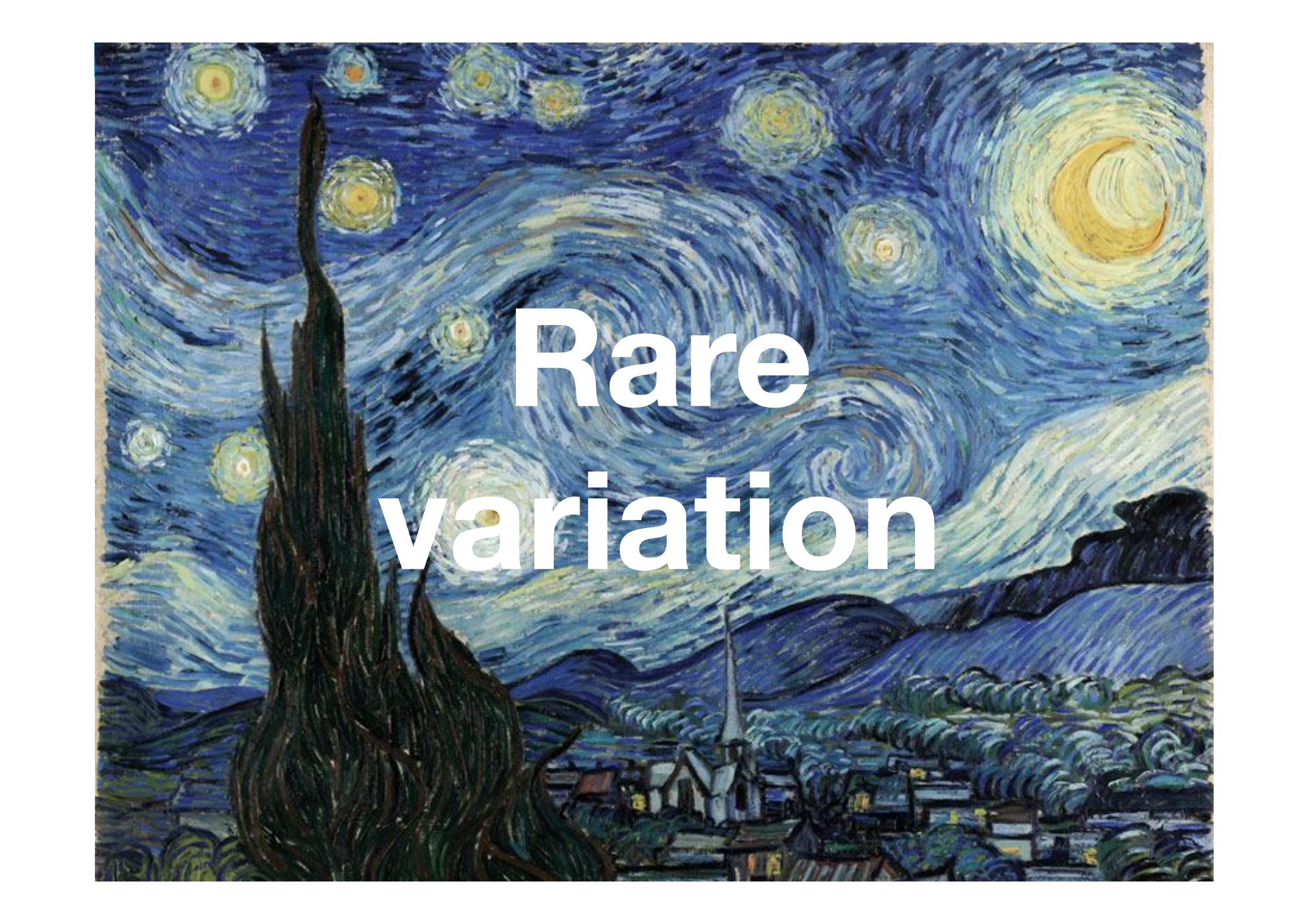


- 20.4 million SNVs
- Includes essentially all HapMap (>98%) and most 1KG-EUR (>71%) SNPs
- Only 39% of rare SNVs (but no singletons) also seen in 1KG



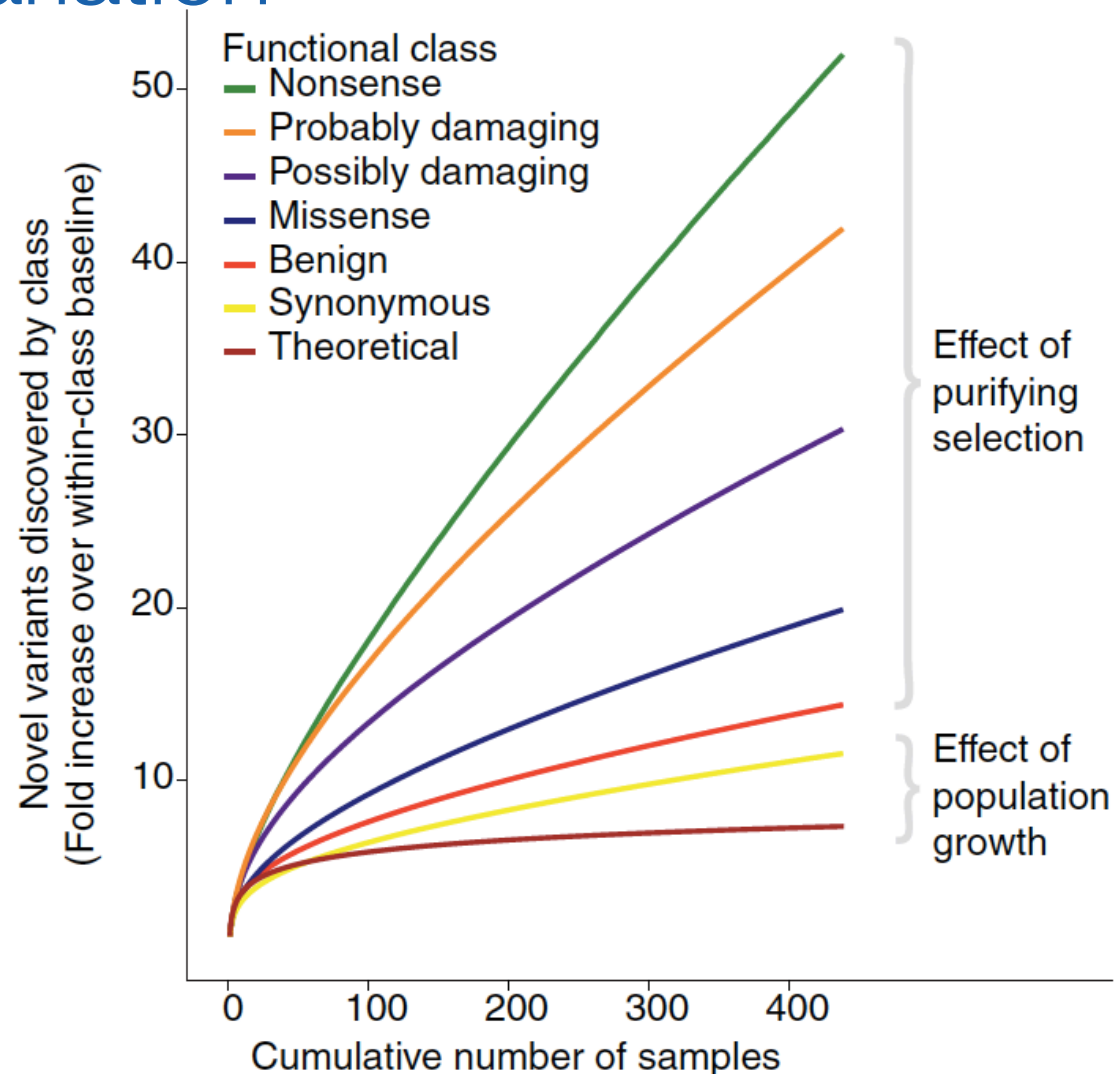
Better imputation accuracy with GoNL+1KG



The background of the slide is a reproduction of the painting 'The Starry Night' by Vincent van Gogh. It depicts a night scene with a dark, swirling sky filled with many bright, glowing stars and a large, luminous crescent moon. In the foreground, a dark, jagged cypress tree stands on the left. Below the sky, a small village with a prominent church spire is visible, nestled among rolling hills. The overall style is characterized by visible, expressive brushstrokes and a rich, vibrant color palette dominated by blues and yellows.

Rare variation

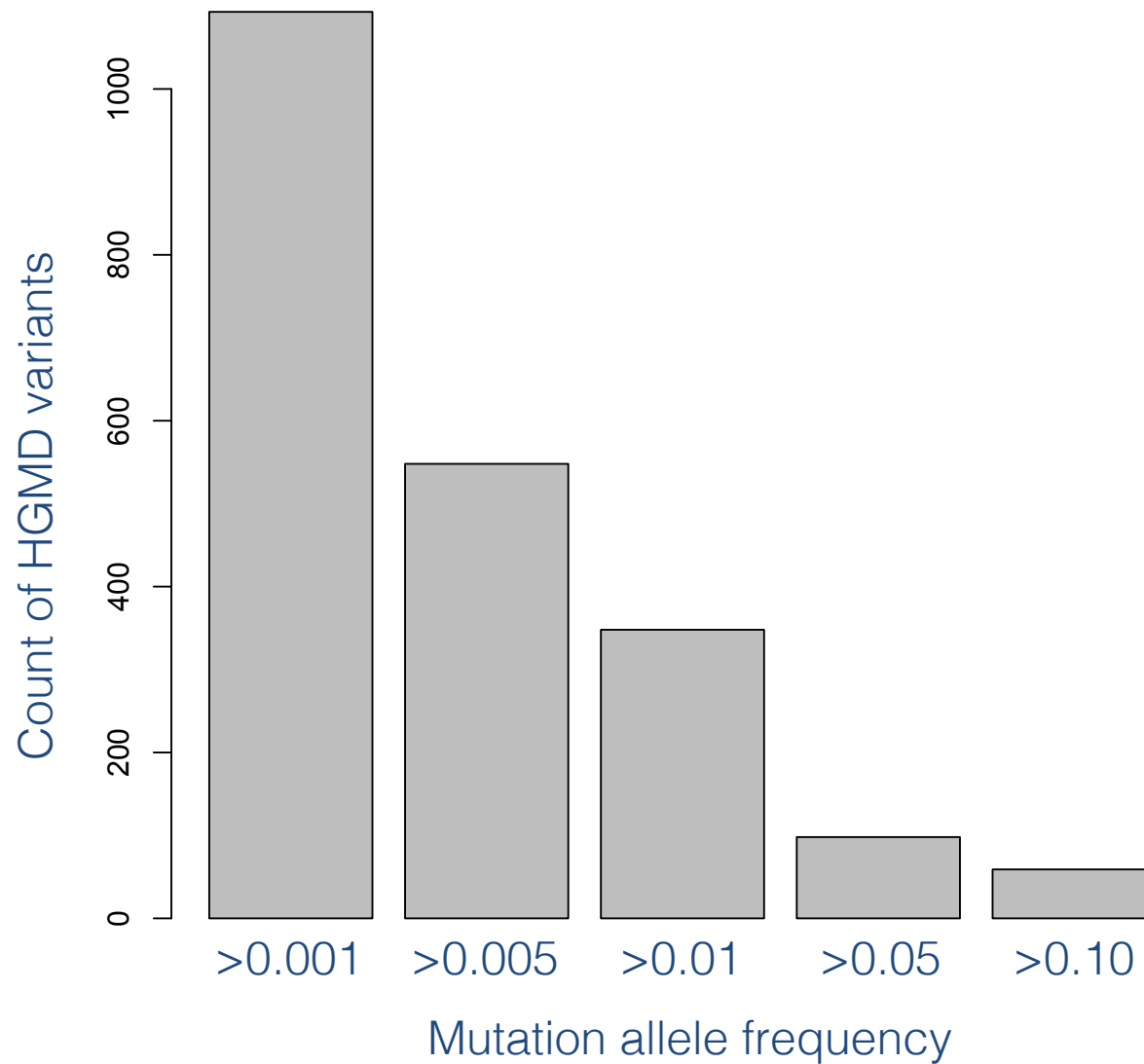
Exome sequencing reveals abundance of rare variation



Kiezun et al. Nature Genetics 2012



HGMD mutations in GoNL

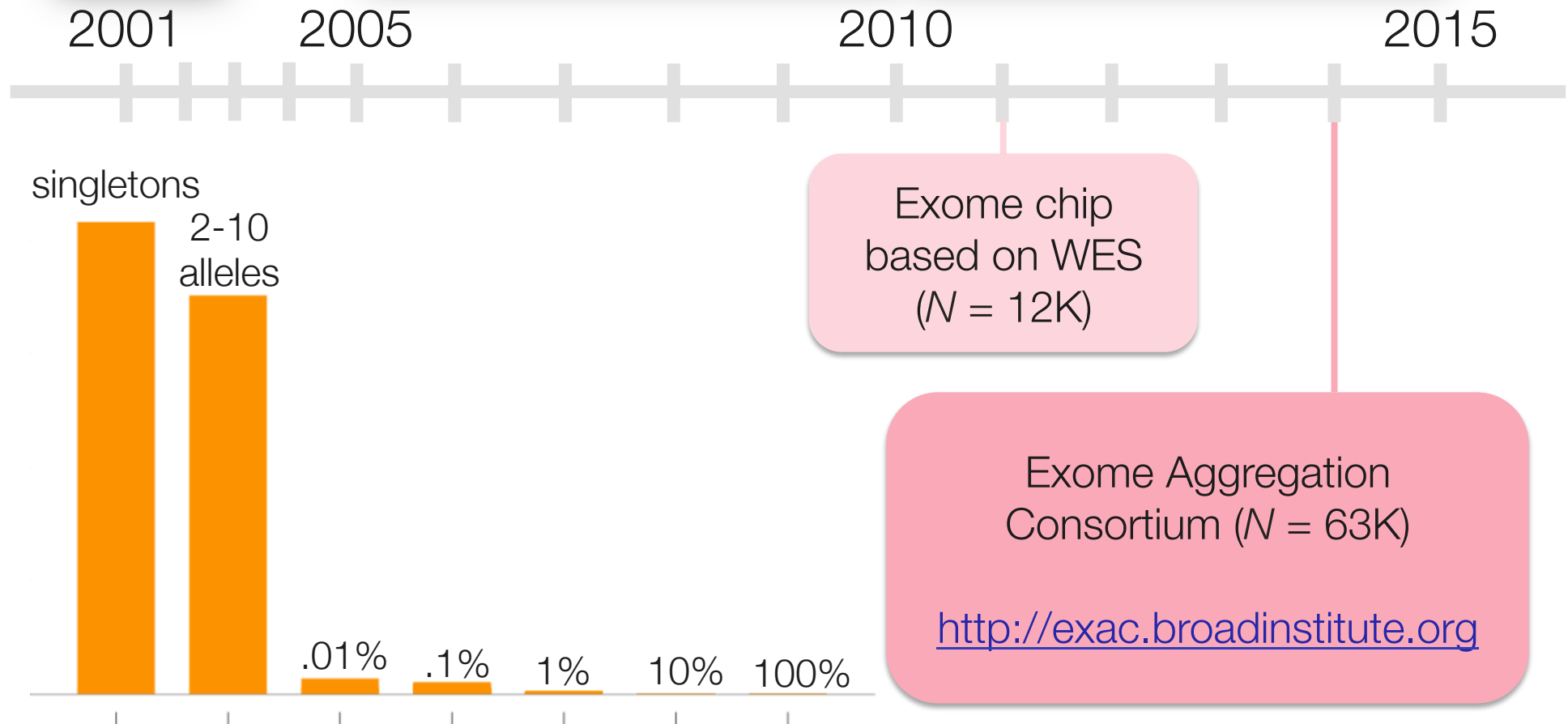


“Disease-causing” mutations

Table 2 HGMD disease-causing mutations in GoNL samples

Chr.	Position	Gene	Mutation allele	Reference allele	Disease in HGMD	Disease prevalence	Inheritance pattern	Affected individuals ^c	Phenotypic manifestations ^a	OMIM ID	Mutation allele frequency in GoNL ^d	Mutation allele frequency 1000G CEU
4	6,302,519	WFS1	A	G	Wolfram syndrome	0.0002% ^a	AR	257	Hyperglycemia, vision and hearing loss	604928, 222300	0.728	0.759
13	52,515,354	ATP7B	G	A	Wilson disease	0.003% ^a	AR	167	Liver disease, neuropsychiatric problems	277900	0.574	0.582
16	3,304,463	MEFV	T	C	Familial Mediterranean fever	0.10% in Mediterranean populations; rarer elsewhere ^a	AR	36	Recurrent fevers, inflammation of the abdomen, chest, joints	249100, 134610	0.277	0.224
11	6,415,463	SMPD1	A	G	Niemann-Pick disease	0.0004% ^a	AR	37	Nervous system deterioration, failure to thrive, fatal in infancy or early childhood (type A)	257200, 607616, 257220, 607625	0.230	0.230
20	61,463,522	COL9A3	A	C	Pseudoachondroplasia	0.003% ^a	AD	177	Short stature, joint pain	177170	0.197	0.200
10	13,340,236	PHYH	A	G	Refsum disease	Unknown, current estimate 0.0001% ^a	AR	18	Anosmia, progressive blindness, deafness, hand/feet bone abnormalities, arrhythmia	266500	0.188	0.153
15	52,643,564	MYO5A	A	G	Griscelli syndrome	<0.0001% ^b	AR	10	Albinism (all types), intellectual disability (type 1), recurrent infection (type 2)	214450, 607624, 609227	0.159	0.141
19	36,339,247	NPHS1	T	C	Congenital nephrotic syndrome (Finnish type)	0.01% in Finland; rarer elsewhere ^b	AR	2	Proteinuria, rapid progression to renal failure	256300	0.082	0.082 (0.110) ^e
14	94,847,262	SERPINA1	A	T	α 1 antitrypsin deficiency	0.02–0.06% ^a	AR	2	Lung disease, liver disease	613490	0.039	0.053

Characterization of DNA polymorphism



Words of caution

- There is an abundance of rare variants
- There will always be “interesting” variants in your sample
- Assuming incomplete penetrance, we cannot rely on variant filters
- We must compare variants in cases and controls (general population)

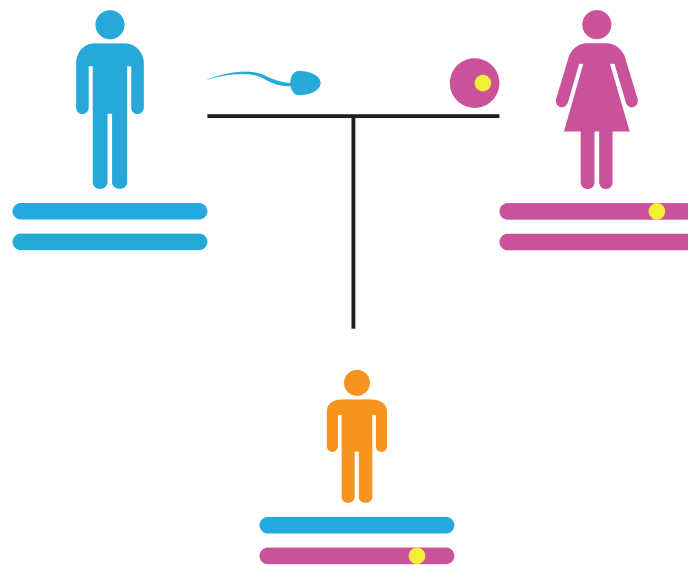




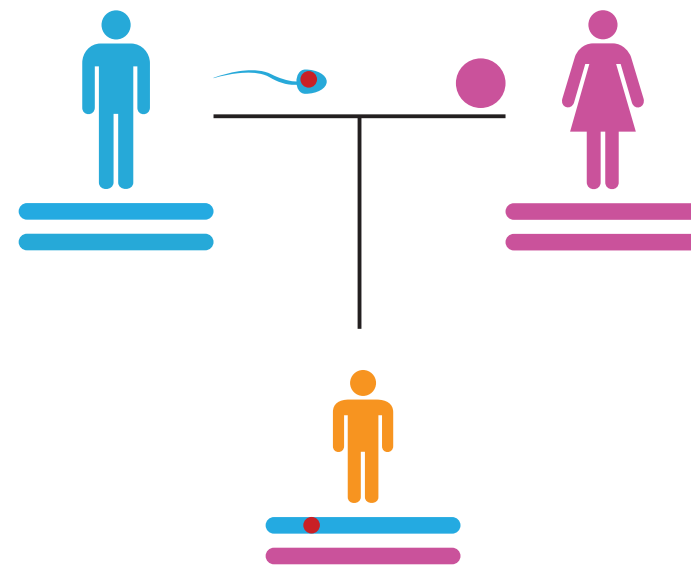
De novo
variation

Inherited vs *de novo* mutations

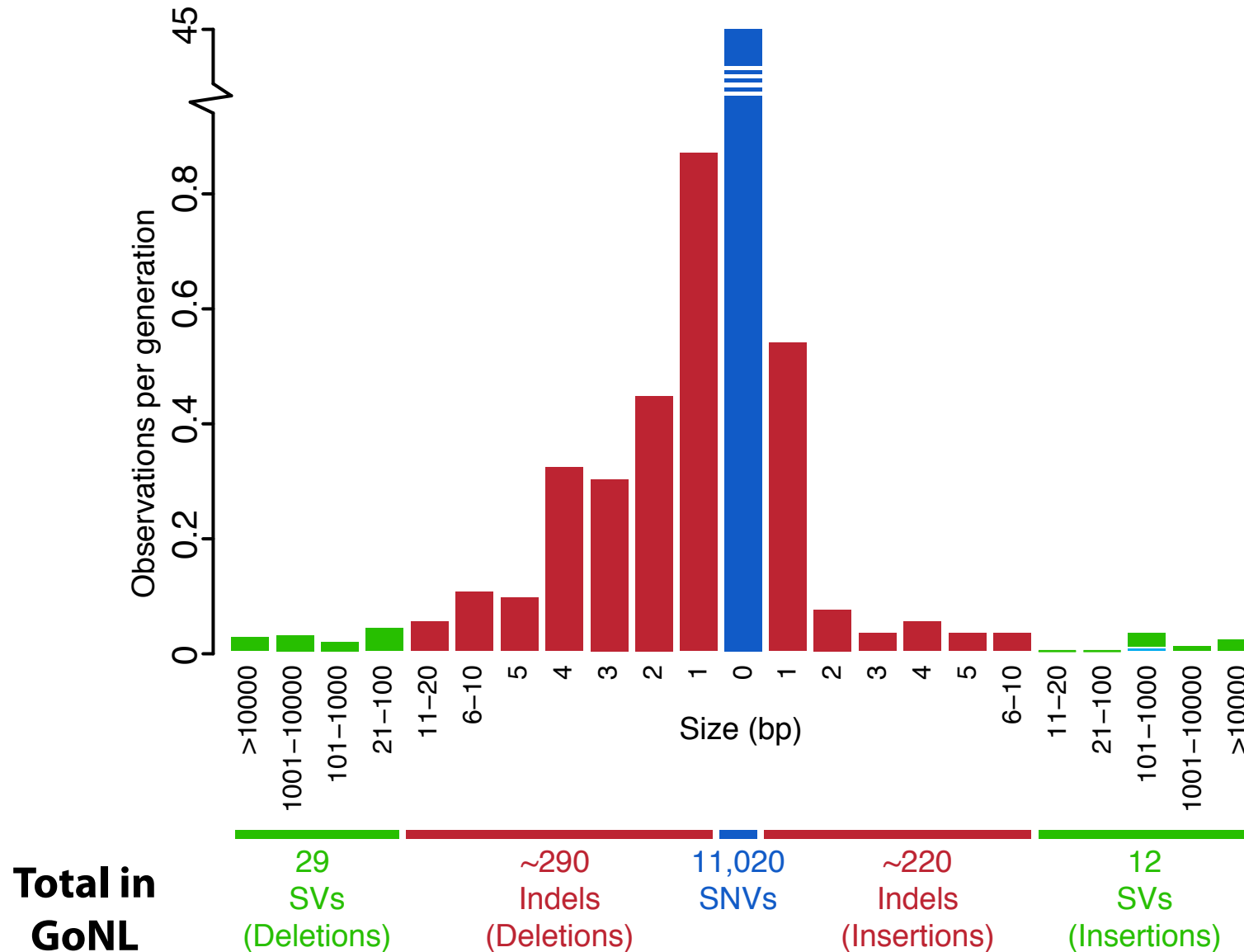
Inherited



De novo

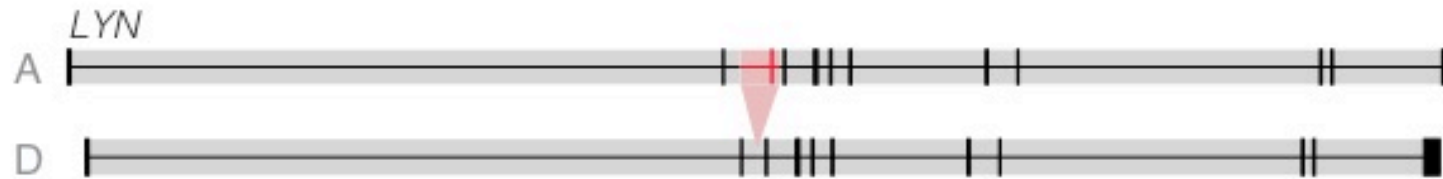


Spectrum of *de novo* events in GoNL

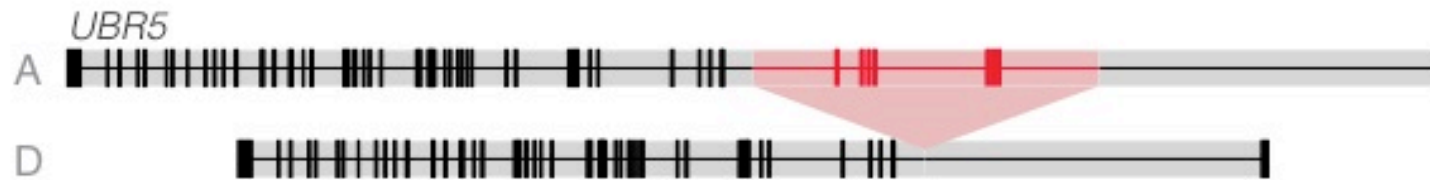


Examples of *de novo* SVs in GoNL

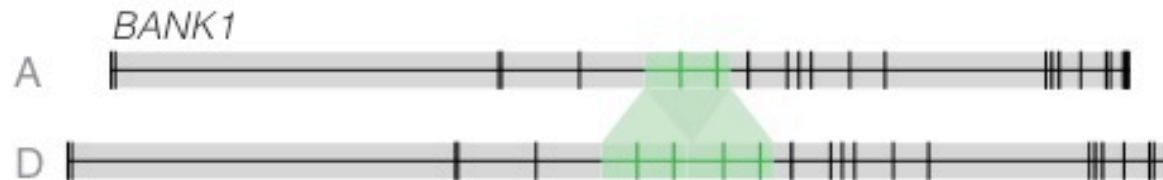
B



C



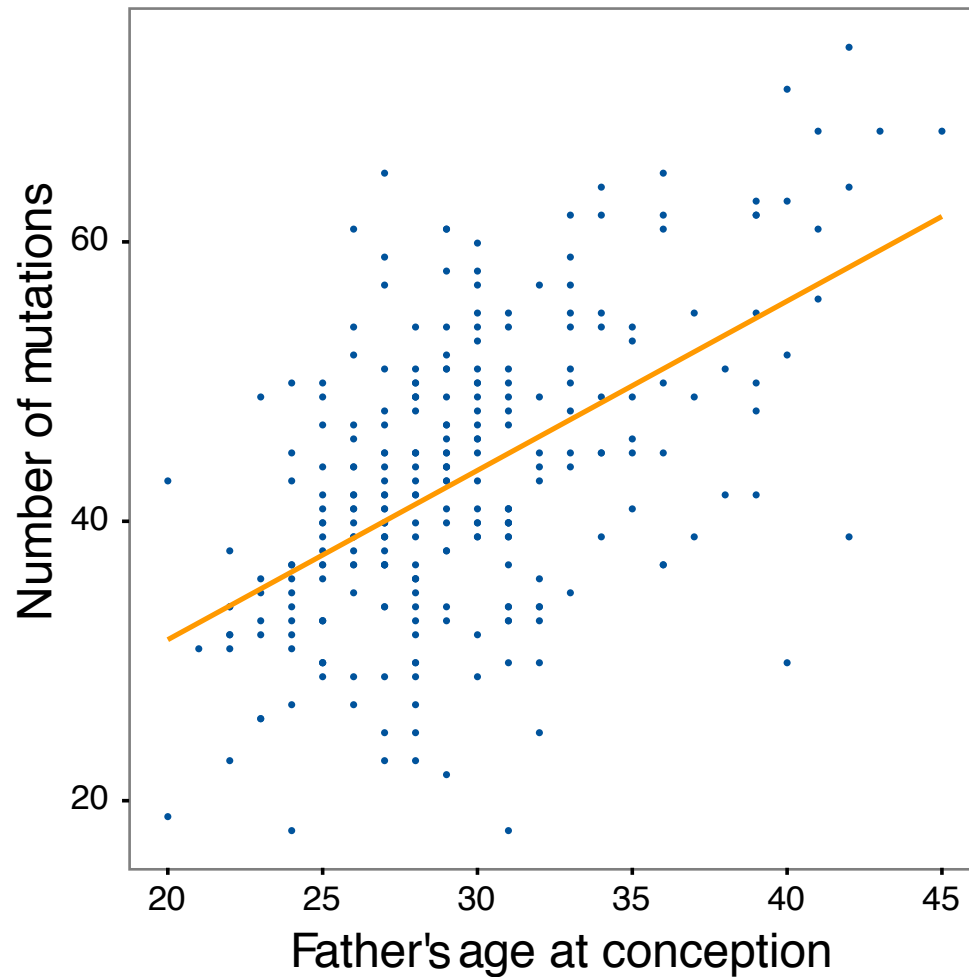
D



E



De novo mutation and age



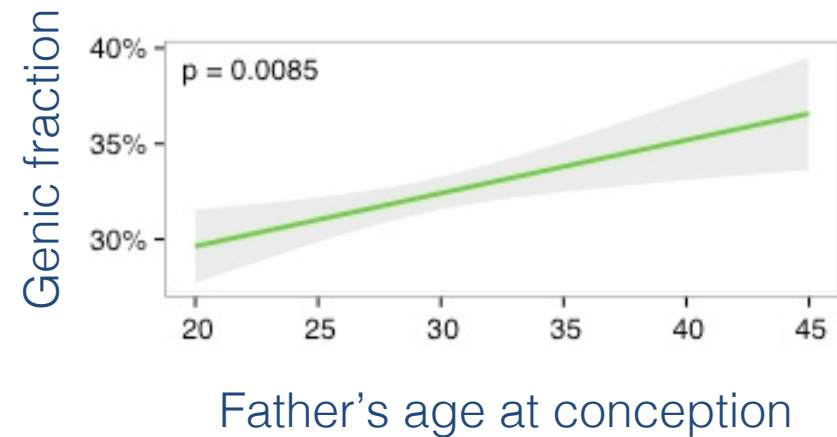
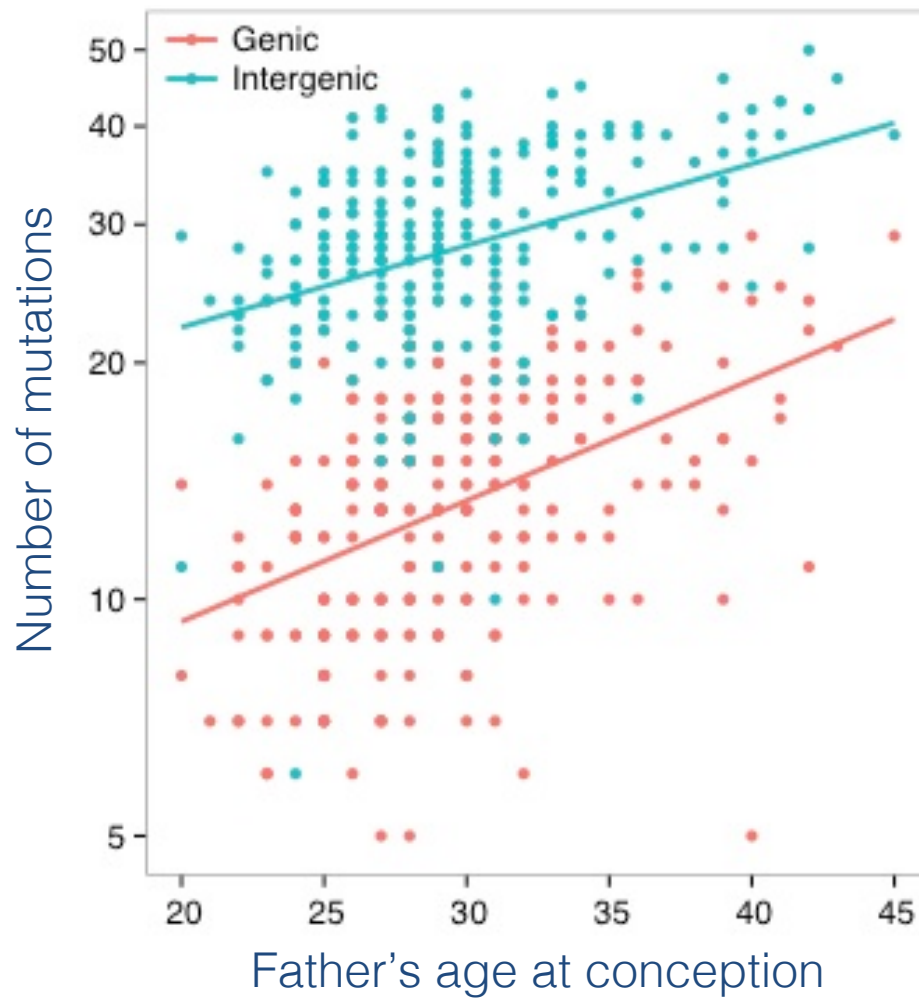
SNVs (N = 2,613)
76% paternal

Indels (N = 56)
64% paternal

SVs (N = 29)
69% paternal



More genic mutations with age



Summary

- Whole-genome sequencing enable comprehensive characterization genetic variation in global populations
- Challenges interpreting “disease” variants, especially rare variants
- Novel insights into *de novo* mutational processes
- Imputation from largest reference panels improve power for GWAS



Acknowledgements

The Genome of the Netherlands Consortium:

Laurent C Francioli^{1,40}, Androniki Menelaou^{1,40}, Sara L Pulit^{1,40}, Freerk van Dijk^{2,3,40}, Pier Francesco Palamara⁴, Clara C Elbers¹, Pieter B T Neerincx^{2,3}, Kai Ye^{5,6}, Victor Guryev⁷, Wigard P Kloosterman¹, Patrick Deelen^{2,3}, Abdel Abdellaoui⁸, Elisabeth M van Leeuwen⁹, Mannis van Oven¹⁰, Martijn Vermaat^{11,12}, Mingkun Li¹³, Jeroen F J Laros^{11,12}, Lennart C Karssen⁹, Alexandros Kanterakis^{2,3}, Najaf Amin⁹, Jouke Jan Hottenga⁸, Eric-Wubbo Lameijer⁶, Mathijs Kattenberg⁸, Martijn Dijkstra^{2,3}, Heorhiy Byelas^{2,3}, Jessica van Setten¹, Barbera D C van Schaik¹⁴, Jan Bot¹⁵, Isaac J Nijman¹, Ivo Renkens¹, Tobias Marschall¹⁶, Alexander Schönhuth¹⁶, Jayne Y Hehir-Kwa^{17,18}, Robert E Handsaker^{19,20}, Paz Polak^{19,21}, Mashaal Sohail^{19,21}, Dana Vuzman^{19,21}, Fereydoun Hormozdiari²², David van Enckevort¹², Hailiang Mei¹², Vyacheslav Koval²³, Matthijs H Moed⁶, K Joeri van der Velde^{2,3}, Fernando Rivadeneira^{9,23}, Karol Estrada^{19,23,24}, Carolina Medina-Gomez²³, Aaron Isaacs⁹, Steven A McCarroll^{19,20}, Marian Beekman⁶, Anton J M de Craen⁶, H Eka D Suchiman⁶, Albert Hofman⁹, Ben Oostra²⁵, André G Uitterlinden^{9,23}, Gonneke Willemsen⁸, LifeLines Cohort Study²⁶, Mathieu Platteel², Jan H Veldink²⁷, Leonard H van den Berg²⁷, Steven J Pitts²⁸, Shobha Potluri²⁸, Purnima Sundar²⁸, David R Cox^{28,39}, Shamil R Sunyaev^{19,21}, Johan T den Dunnen^{11,29}, Mark Stoneking¹³, Peter de Knijff³⁰, Manfred Kayser¹⁰, Qibin Li³¹, Yingrui Li³¹, Yuanping Du³¹, Ruoyan Chen³¹, Hongzhi Cao³¹, Ning Li³², Sujie Cao³², Jun Wang^{31,33,34}, Jasper A Bovenberg³⁵, Itsik Pe'er^{4,36}, P Eline Slagboom⁶, Cornelia M van Duijn⁹, Dorret I Boomsma⁸, Gert-Jan B van Ommen³⁷, Paul I W de Bakker^{1,38,41}, Morris A Swertz^{2,3,41} & Cisca Wijmenga^{2,3,41}

