

Effect of global migration on the European population genotype as illustrated by the genome of the Netherlands: Effects of behavior, migration patterns and geography on genotypes and phenotypes

Dorret Boomsma

Netherlands Twin Register, Vrije Universiteit Amsterdam

# Outline

- Population stratification (or population structure) : the presence of a systematic difference in allele frequencies between (sub)populations, possibly due to e.g. different ancestry
- Characterizing genomic variation across populations (based on SNP & whole-genome sequencing)
- Genetic variation and migration
- Relation with phenotypes: biomarkers, height, schizophrenia, education, MDD

# Population stratification

---

Ways to study population stratification:

- ▶ Principal Component Analysis
- ▶ (= Multidimensional Scaling in Plink)

Applying PCA to human genetic variation has helped in:

- ▶ Correcting for stratification
- ▶ Detecting population substructure
- ▶ Making inferences about human history

Ways to study genetic differences between populations:

- ▶  $F_{st}$ : measure of genetic variability within and between populations

# Principal Component Analysis (PCA)

---

- ▶ PCA is a statistical method for exploring datasets with a large number of measures (e.g. genetic variants) by reducing the dimensions to the few principal components (PCs).
- ▶ The first PC is the mathematical combination of measures that accounts for the largest amount of variability in the data. The second PC (uncorrelated with the first) accounts for the second largest amount of variability, etc.
- ▶ PC Plots provide guidance for interpretation of findings and deciding which PCs to use for correction for population stratification.

# Perspective “world”

---

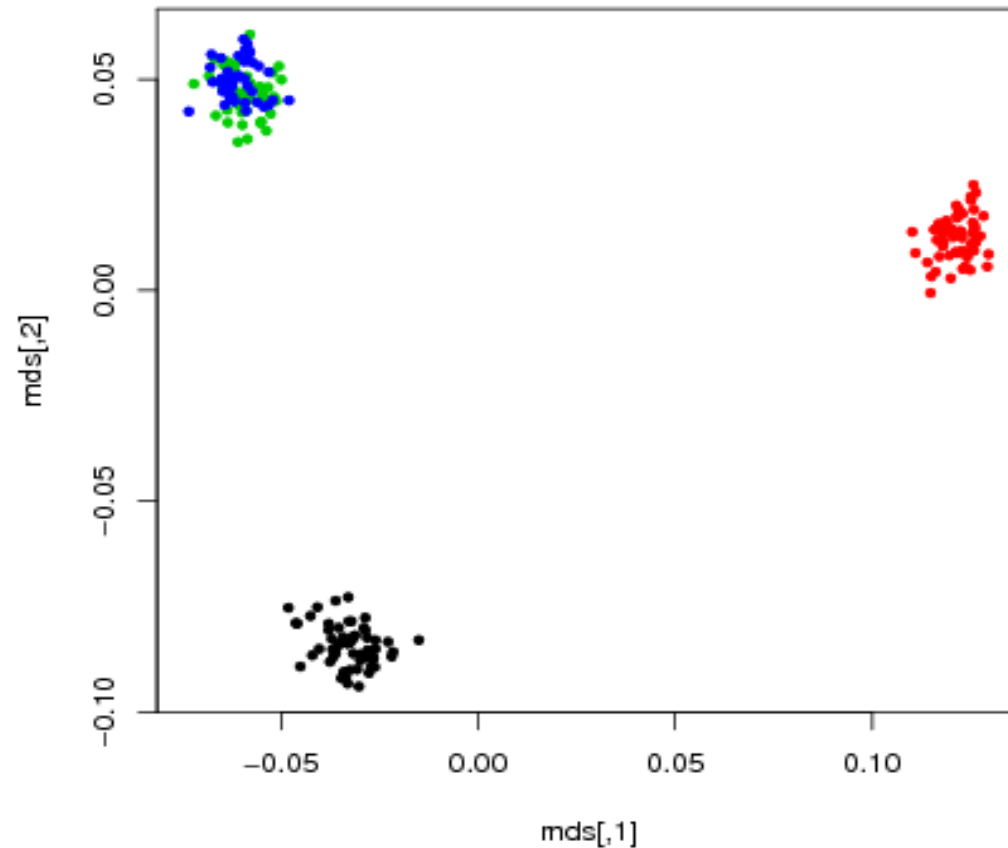
CEPH/European

Yoruba

Han Chinese

Japanese

(Multidimensional  
Scaling (MDS))

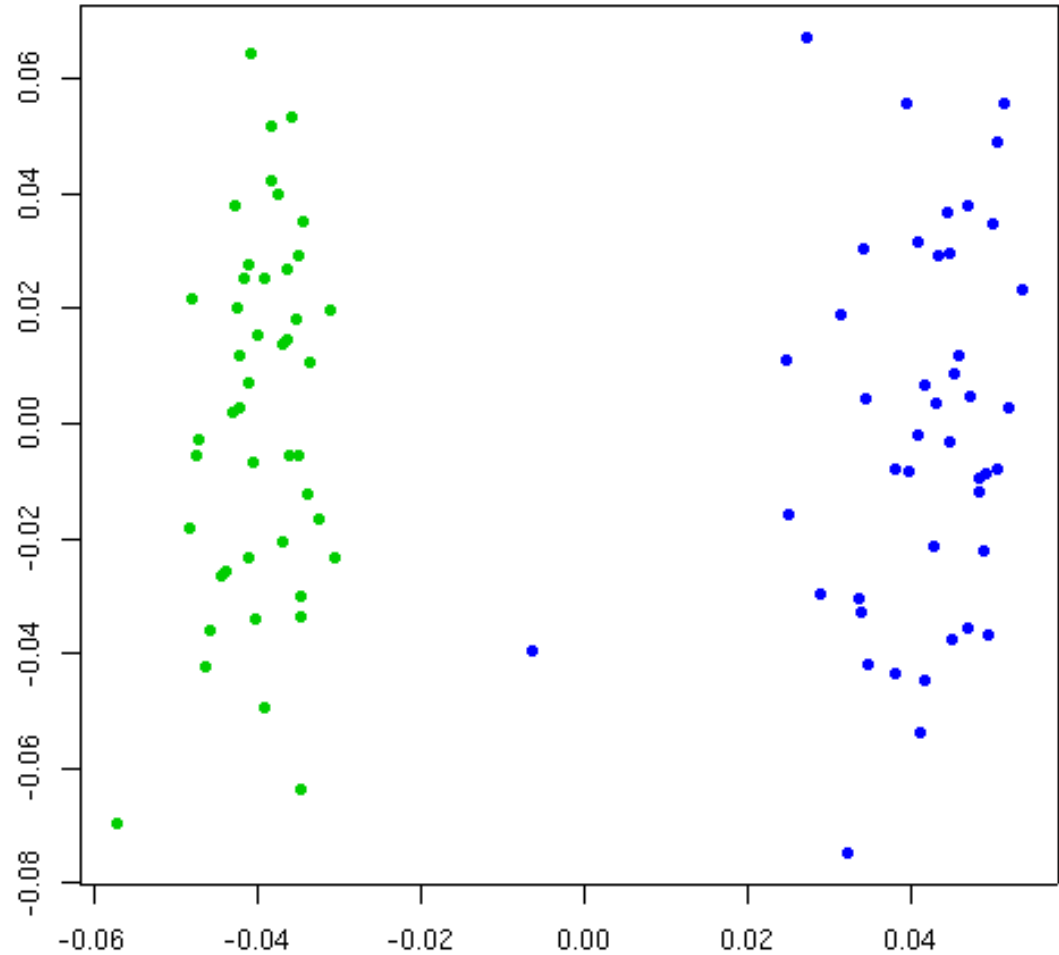
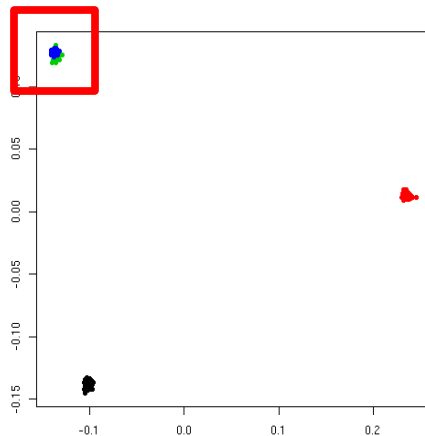


CEPH: 10 French families / 48 Mormon families from Utah / 1 Pennsylvania Amish family / 2 Venezuelan families

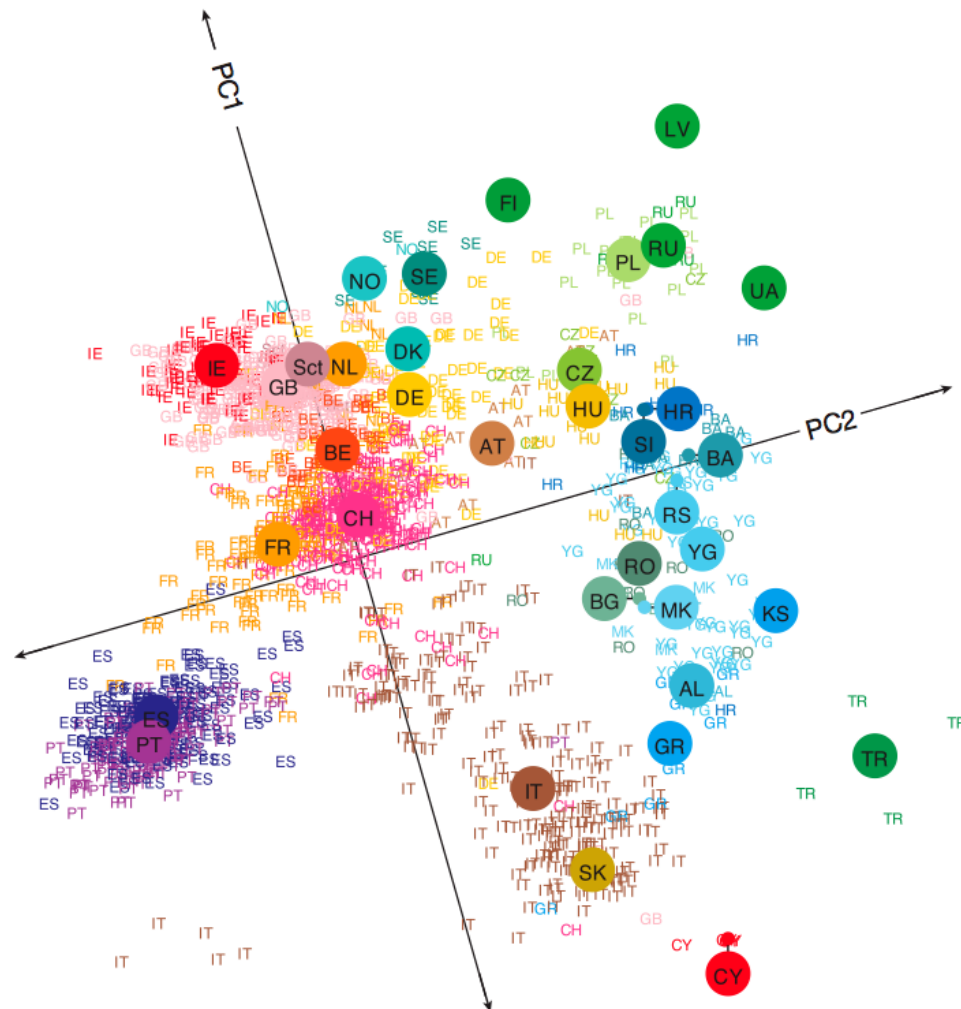


# Perspective “China / Japan”

Han Chinese  
Japanese



# Perspective "Europe"



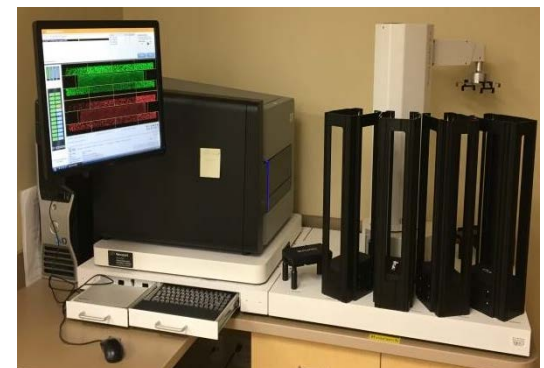
# Twin-family populations genotyped on GSA

Cohort	Country of origin	Sample				
		<i>N</i>	Female (%) <sup>a</sup>	Composition	Unrelated individuals	Tissue
Avera Twin Register	USA	602	66.4	MZ, DZ, parents, sibs	238	Buccal
NTR	Netherlands	1135	55.4	MZ, DZ, parents, sibs	6139	Blood
NTR	Netherlands	9091		MZ, DZ, parents, sibs		Buccal
Australian	Australia	1922	100	MODZT	1448	Blood
Nigerian	Nigeria	100	100	MOSDZ	96	Saliva
Total		12,850			7921	

GSA = Global Screening Array; NTR = Netherland's Twin Register; *N* = number of samples; MZ = monozygotic twins; DZ = dizygotic twins; MODZT = mothers of DZ twins; MOSDZ = mothers of opposite-sex dizygotic twins; sibs = siblings; parents = parents of twins.

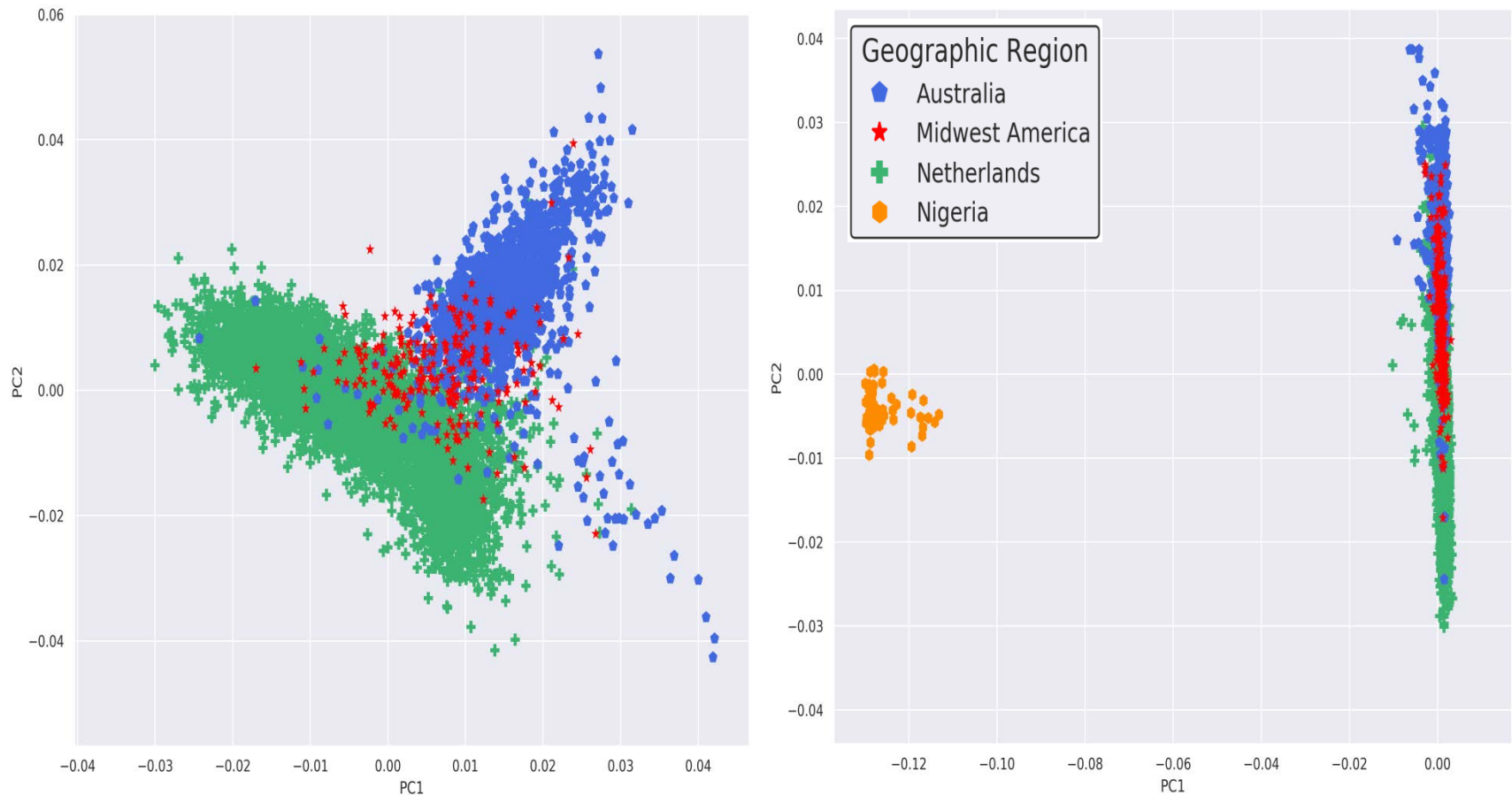
<sup>a</sup>Percentage of female samples is reported from the unrelated set.

Beck, J. J., Hottenga, J. J., Mbarek, H., Davies, G. E. (2019). Genetic Similarity Assessment of Twin-Family Populations by Custom-Designed Genotyping Array. *Twin Res Hum Genet*, 1-10. doi:10.1017/thg.2019.41



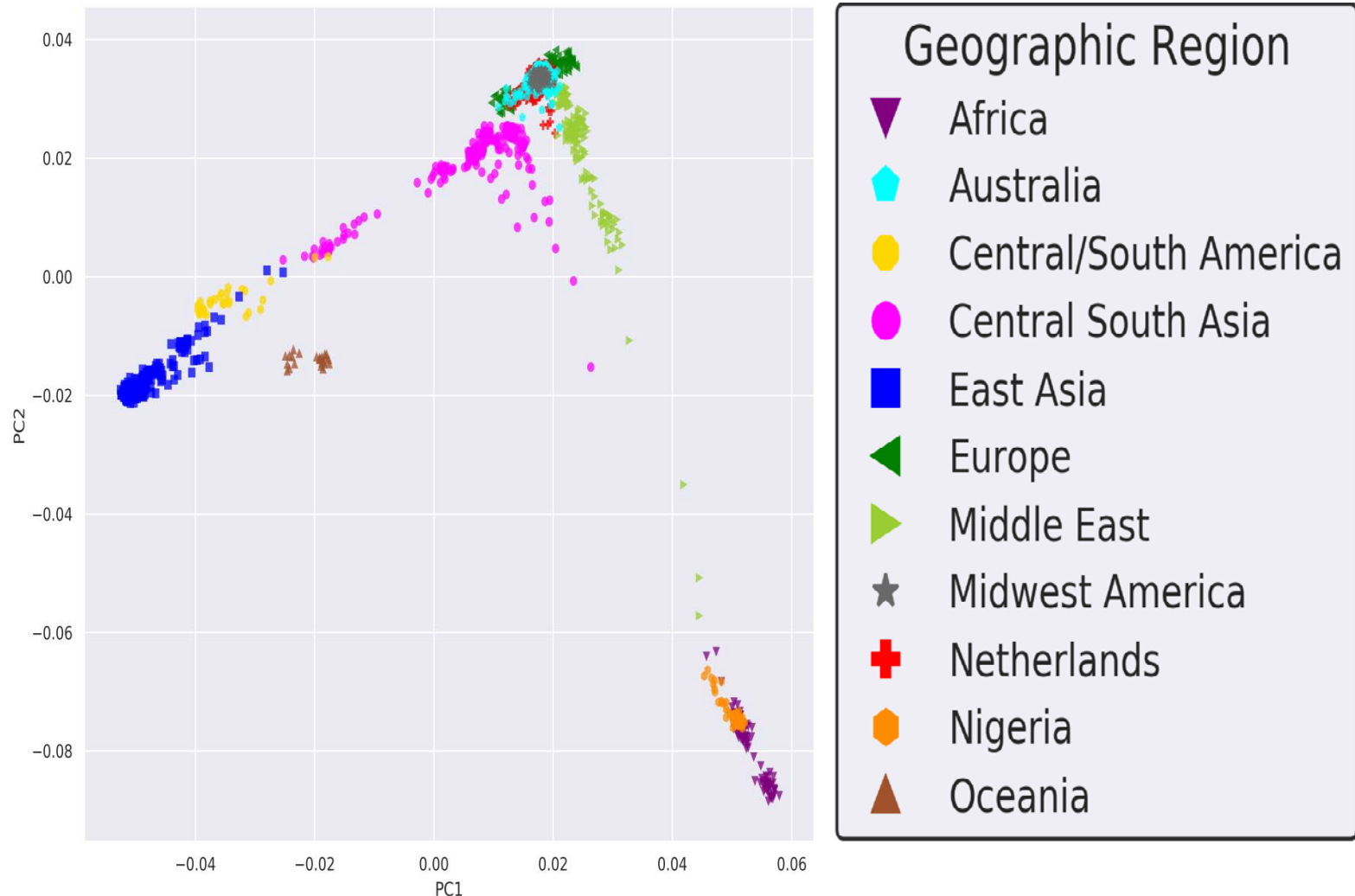


# Top two PCs of 109,702 GSA SNPs reveal genetic similarity of Australian, Dutch, and Midwestern American Populations



Beck, J. J et al. (2019). Genetic Similarity Assessment of Twin-Family Populations by Custom-Designed Genotyping Array. *Twin Res Hum Genet*, 1-10. doi:10.1017/thg.2019.41

# Projection of study population PCs on to worldwide Human Genome Diversity Project (HGDP) populations (54,820 SNPs)



Beck, J. J et al. (2019). Genetic Similarity Assessment of Twin-Family Populations by Custom-Designed Genotyping Array. Twin Res Hum Genet, 1-10. doi:10.1017/thg.2019.41

# Twin-family population differentiation as estimated by $F_{ST}$

---

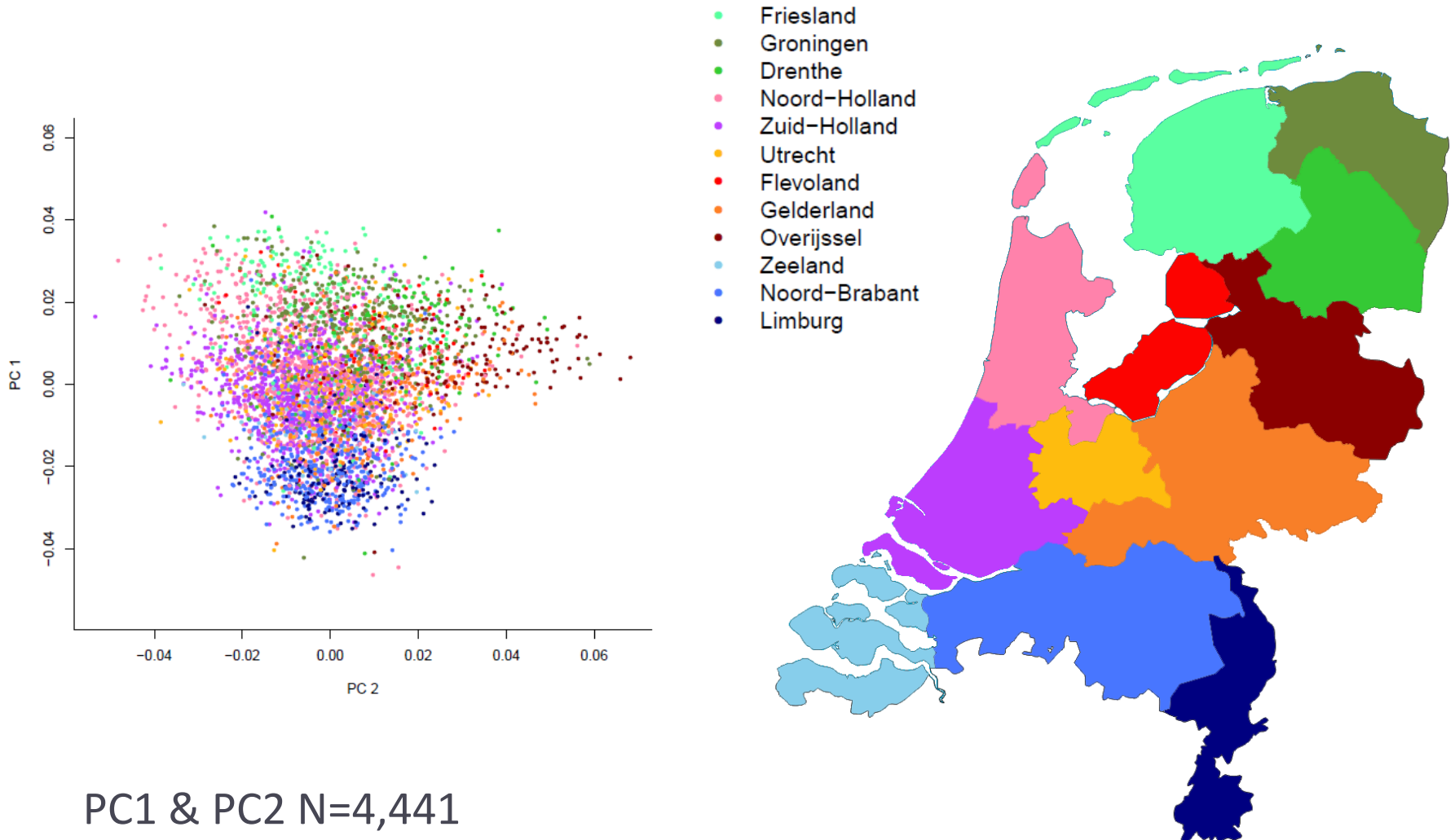
Comparison		Weir and Cockerham $F_{ST}$	
Population 1	Population 2	Est	Std. error
Midwest US	Dutch	0.00017	7.40e-06
Midwest US	Australian	0.00019	8.34e-06
Midwest US	Nigerian	0.14362	0.00100
Dutch	Australian	0.00045	8.73e-06
Dutch	Nigerian	0.14382	0.00101
Australian	Nigerian	0.14347	0.00102

The number of SNPs used for each comparison was 564,020.  $F_{ST}$  as estimated by *popstats* software. Block size for the jackknife estimator was default 5 Mb.

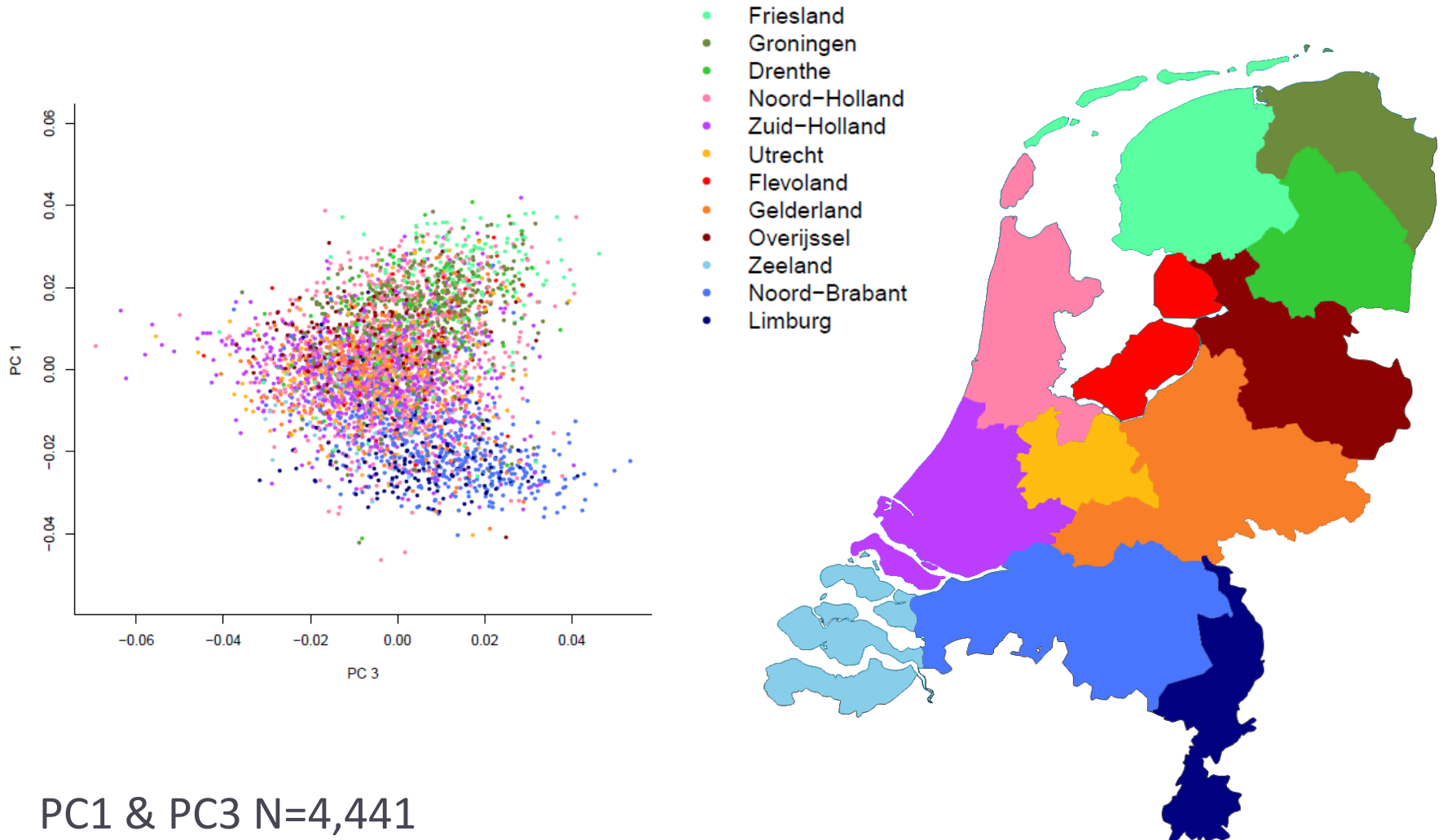
---



# Perspective “Netherlands”

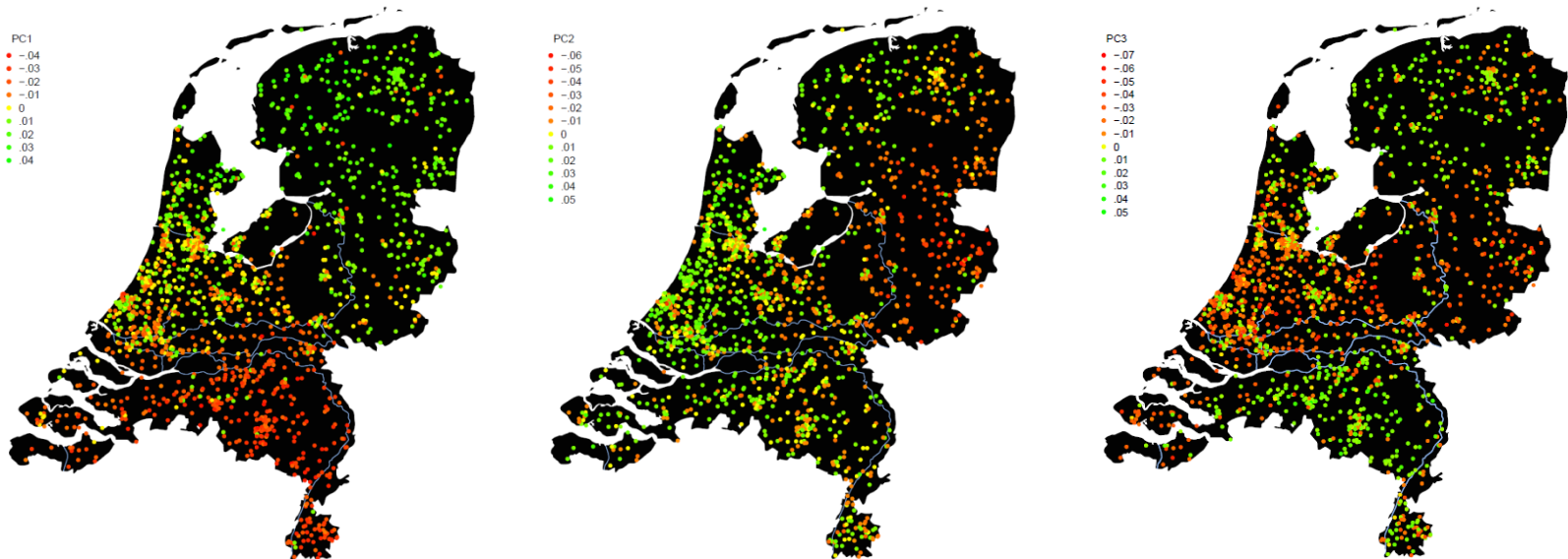


# Perspective “Netherlands”



# PC1, PC2, and PC3

- ▶ The color of the dots represent the mean PC value per postal code (based on current living address of the 4,441 subjects).



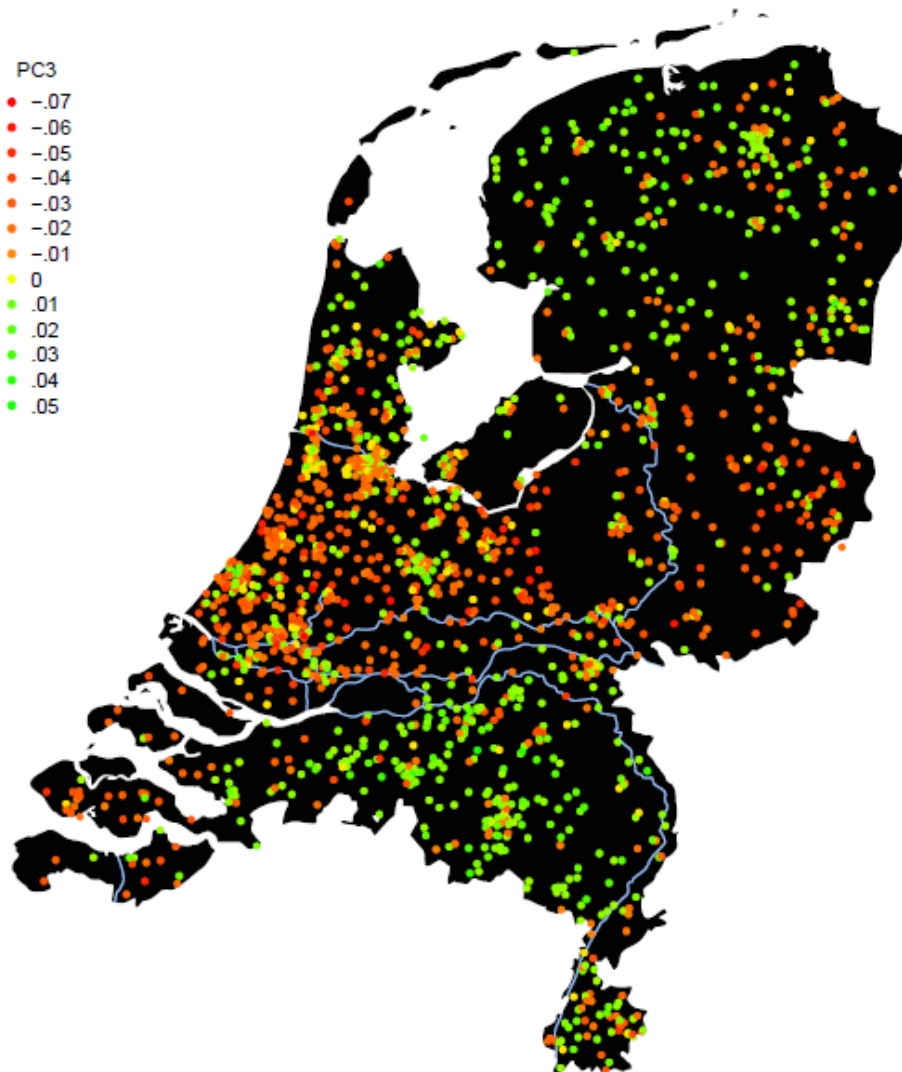
# Height and PC 1

---

- ▶ Northern Dutch are taller on average than the Dutch from the Southern parts of the Netherlands.
- ▶ In our sample, height does not correlate very high with the North-South gradient of the current living address:
  - ▶ males:  $r = .036$ ,  $p = .232$ ; females:  $r = .050$ ,  $p = .020$
- ▶ Height however correlates higher with the North-South PC:
  - ▶ males:  $r = .142$ ,  $p < .001$ ; females:  $r = .153$ ,  $p < .001$
- ▶ This shows that the PC is indeed a better measure for ancestral origin than current geographical location.

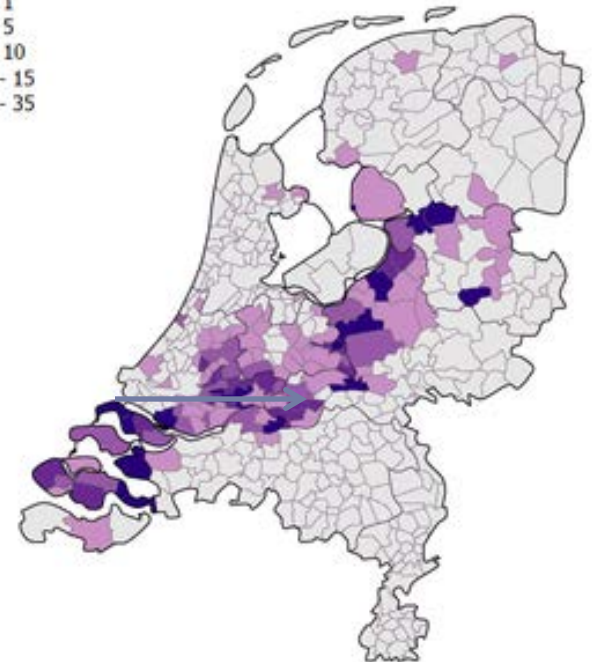
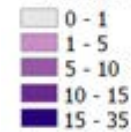


# PC3 (N=4,441)



Votes for conservative  
Christian party (SGP)

Percentage van het aantal kiesgerechtigden





# Netherlands Twin Register

# NTR recruitment: twin families



First projects (1983):  
cardiovascular risk. Families  
approached via Amsterdam City  
Council.

Large scale city council  
approach (1990): addresses of  
all adolescent twins in NL.

1988: new born twins via  
commercial partner

Ongoing: new born & adult  
twins and their family members

# Focus

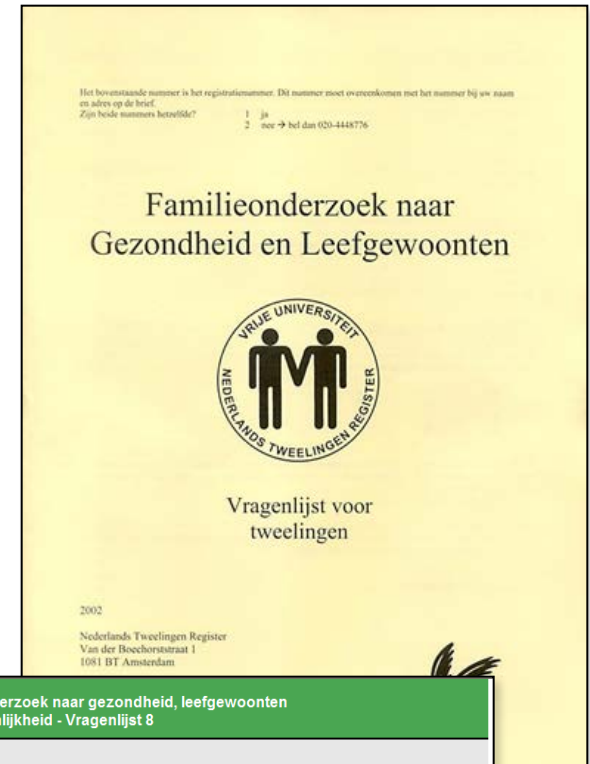
- Heritability, genetic & cultural transmission
- Gene finding / genetic association studies (GWAS 'revolution')
- Omics (epigenome, transcriptomics, metabolomics)
- Epidemiological projects / Causality ('MR')
- 1 in 40 people is a twin -> are they different?



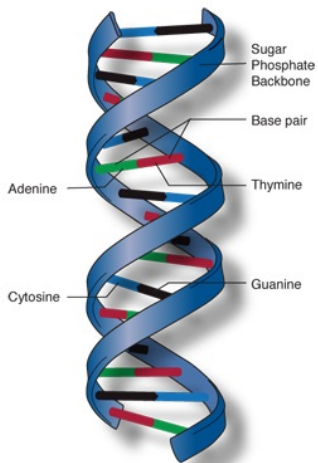
**NTR (April 2019): Number of participants by role and age group**

	0-11 years	12-17 years	18+	Total
<b>Multiples</b>	<b>27474</b>	<b>20949</b>	<b>73734</b>	<b>122,157</b>
<b>Siblings</b>	<b>628</b>	<b>2338</b>	<b>11938</b>	<b>14,904</b>
<b>Parents</b>	<b>0</b>	<b>0</b>	<b>114278</b>	<b>114,278</b>
<b>Spouses</b>	<b>0</b>	<b>0</b>	<b>2616</b>	<b>2,616</b>
<b>Offspring</b>	<b>280</b>	<b>28</b>	<b>778</b>	<b>1,086</b>
<b>Other</b>	<b>1</b>	<b>4</b>	<b>683</b>	<b>688</b>
<b>Total</b>	<b>28,383</b>	<b>23,319</b>	<b>204,027</b>	<b>255,729</b>

# Survey, biobank & non-survey studies







# NTR-BIOBANK



DNA, RNA, blood,  
urine, feces, etc.

## INSTRUCTIE

### Wat zit er in de enveloppe?

U heeft 5 buisjes ontvangen: een wat  
dikker buis met 16 wattenstaafjes en  
4 dunne buisjes met alleen een beetje  
vloeistof.

- 1 staafje voor de binnenkant van  
de linkerwang.
- 1 staafje voor de binnenkant van  
de rechterwang.
- U wrijft per wattenstaafje ongeveer

niet meer nodig. Deze kunt u dus  
weggoien.

Wilt u de vier dunne buisjes met de  
mondultstrijkes vervolgens direct in  
de bijgesloten antwoordenveloppe  
met de gewone post aan ons opsturen?  
Let op: als het vriest, de enveloppe  
niet in een buitenbrievenbus gooien.

Nadat het erfelijk materiaal is geanalyseerd,  
zullen eventuele restanten  
voorlopig opgeslagen worden voor  
mogelijk aanvullende bepalingen ten  
behoeve van het onderzoek naar leef-  
gewoonten, gedrag en gezondheid.  
Wanneer u dit niet wilt, stelt u ons  
dan daarvan s.v.p. op de hoogte. In  
dat geval zullen wij het overgebleven  
materiaal vernietigen.

Indien u nog vragen heeft, kunt u  
contact opnemen met ons secretari-  
aat. Onze gegevens staan op de achter-  
zijde van deze brochure.



### Wat te doen met het materiaal?

De mondultstrijkes moeten 4 keer  
worden afgenomen, verspreid over  
twee dagen (2 keer per dag; bijv. voor  
de ochtend- en avondmaaltijden).  
Niet van te voren de mond spoelen,  
tanden poetsen of eten!

10-20 seconden zorgvuldig en met  
enige druk (het hoeft niet hard; het  
mag geen pijn doen).

Na het wrijven deponert u het staaf-  
je, met het watje naar beneden, in de  
vloeistof in één van de dunne buisjes.  
Doe alle 4 staafjes in ditzelfde dunne  
buisje. De andere dunne buisjes zijn  
voor de volgende mondultstrijkes.  
Wilt u het deksel van het buisje  
goed dichtdraaien? (Niet te  
hard om barsten van de  
buis te voorkomen).  
Als u na twee dagen alle  
vier de afnames heeft ge-  
daan, heeft u de dikke buis

Per afname gebruikt u 4 wattenstaaf-  
jes:

- 1 staafje voor de binnenkant van  
de bovenlip en het tandvlees van  
de bovenkaak.
- 1 staafje voor de binnenkant van  
de onderlip en het tandvlees van  
de onderkaak.



	Total N	YNTR survey	ANTR survey
YNTR survey	147360	147360	24072
ANTR survey	52039	24072	52039
DNA (blood or buccal)	29138	15984	19397
Genome-wide SNPs	26146	14511	17704
DNA methylation whole blood 450K	3025	355	2942
DNA methylation buccal 450K	22	22	21
DNA methylation buccal 850K	1526	1526	58
DNA sequencing whole blood	349	9	330
RNA expression U219 array	3370	284	3231
RNA sequencing	1606	256	1561
RNA seq LPS stimulated samples	390	13	380
Metabolomics in serum or plasma	6485	621	6311
Metabolomics in urine	1564	1564	156
Microbiome	568	187	561
Hair sample	222	7	222

## Lipids

- Total cholesterol
- LDL & HDL cholesterol
- Triglycerides

## Glucose metabolism

- Glucose
- Insulin
- HbA1c

## Inflammation

- C-reactive protein ( CRP)
- tumor necrosis factor (TNFa)
- Interleukin-6 (IL6 and IL6R)

## Other

- Liver enzymes:

Gamma glutamyl transferase (GGT)  
alanine aminotransferase (ALT)  
aspartate aminotransferase (AST)

- Cotinine
- Fibrinogen
- Creatinine

# Adult NTR Biobank (~10.000)

## Haematology profile

### WHITE BLOOD CELLS

- Count
- Differentiation (neutrophils, monocytes, lymphocytes, eosinophils, basophils)

### RED BLOOD CELLS

- Count
- Hemoglobine
- mean corpuscular volume (MCV)
- Hematocrite
- mean corpuscular hemoglobin (MCH)
- corpuscular hemoglobin concentration (MCHC)
- red cell distribution width (RDW)

### PLATELETS

- Count
- mean platelet volume (MPV)

Gut  
microbiome

## Metabolomics

Lipidomics / Biocrates

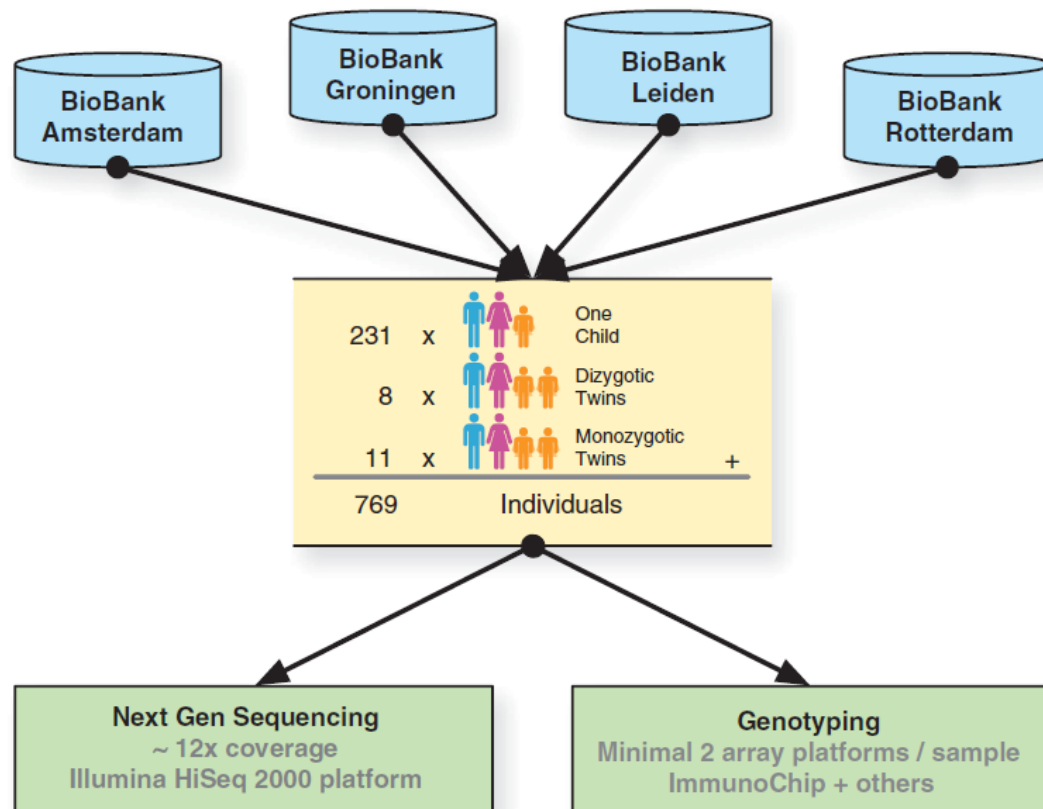
<sup>1</sup>H-NMR/ Brainshake Nightingale

LCLs, RNA expression, DNA (sequence / epigenetics) /Telomeres



# Genome of the Netherlands (GoNL)

- ▶ Goal: Capture whole genome-sequence variation in the Netherlands
- ▶ Trios and quartets from 11 of the 12 Dutch provinces were sampled without ascertaining on phenotype or disease.



---

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

Genome of the Netherlands Consortium

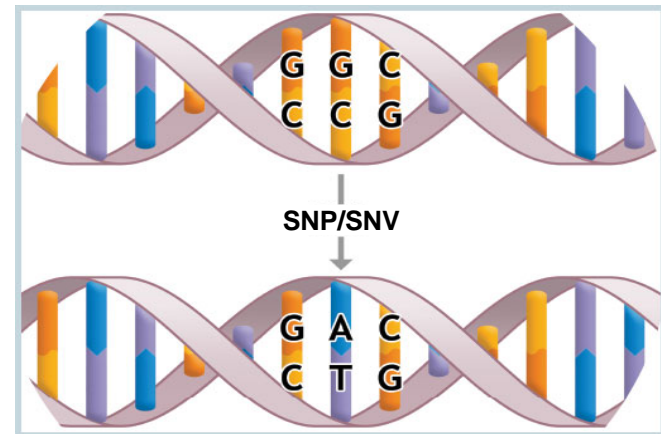
- Single Nucleotide Variants (SNVs)
- Structural Variants
- Loss of Function (LoF) variants
- De novo* mutations
- Population structure and demographic inference

# Single Nucleotide Variants (SNVs)

## dbSNP definition:

dbSNP variation class <sup>a, b</sup>	Rules for assigning allele classes
Single Nucleotide Variations (SNVs) <sup>a</sup>	<p>Strictly defined as single base substitutions involving A, T, C, or G.</p> <p>Formerly called “SNP”. Name changed to “SNV” to emphasize that the dbSNP database contains both rare and polymorphic variants.</p>

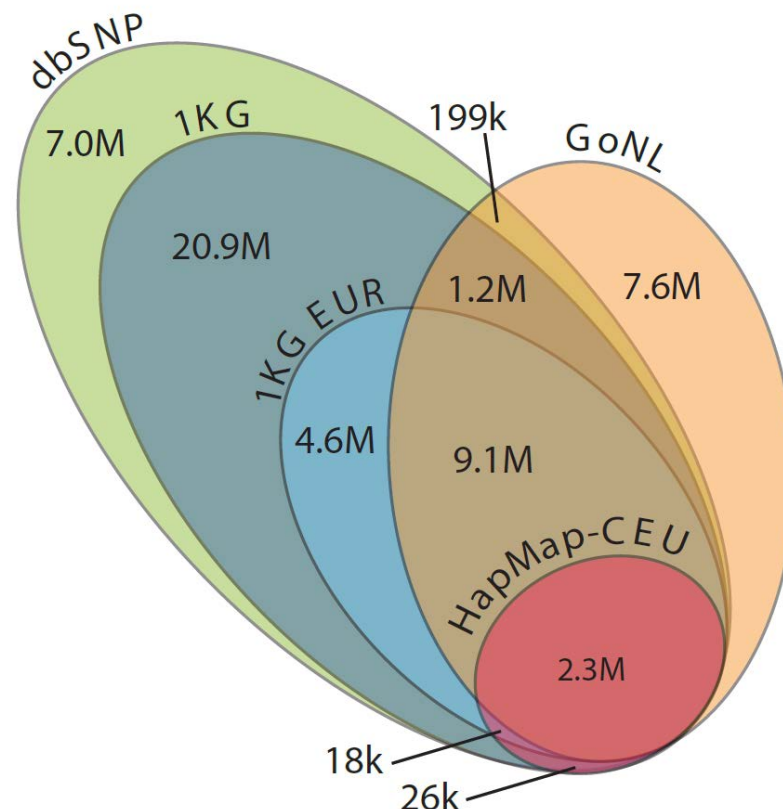
So basically, SNVs are common and rare SNPs.





# Single Nucleotide Variants (SNVs)

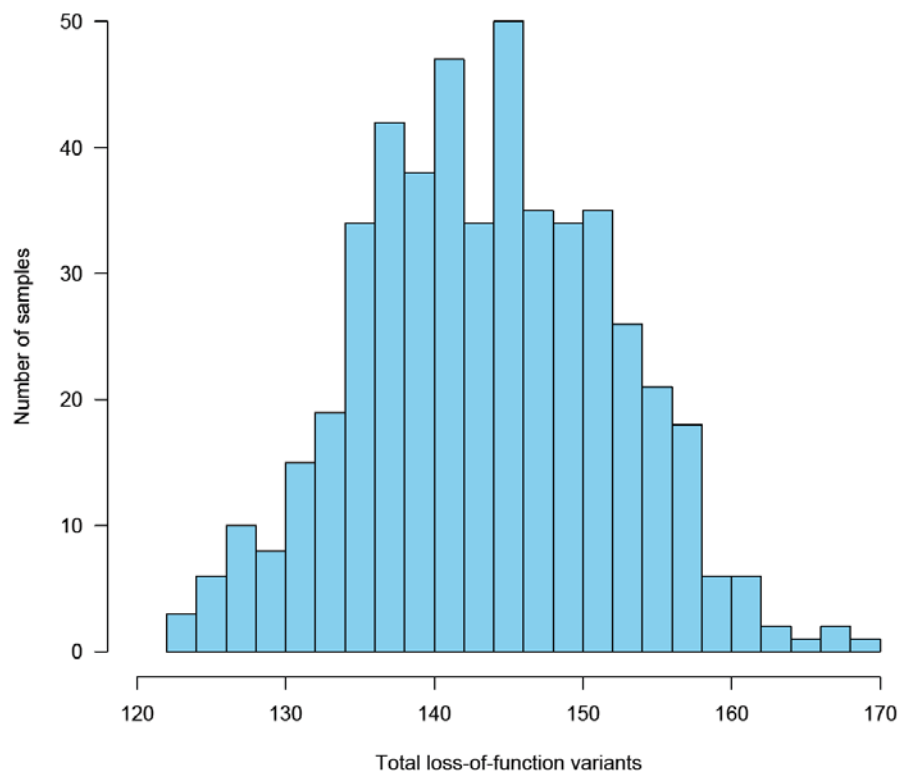
- ▶ 98.2% of sites present in the European sample of HapMap Phase 2 (CEU) were observed in GoNL
- ▶ 71.1% of sites in the European subset of the 1000 Genomes Project Phase 1 (1KG-EUR) were present in GoNL
- ▶ 16.5% of 2 million singletons in 1KG-EUR were also observed in GoNL



# Loss-of-function (LoF) variants

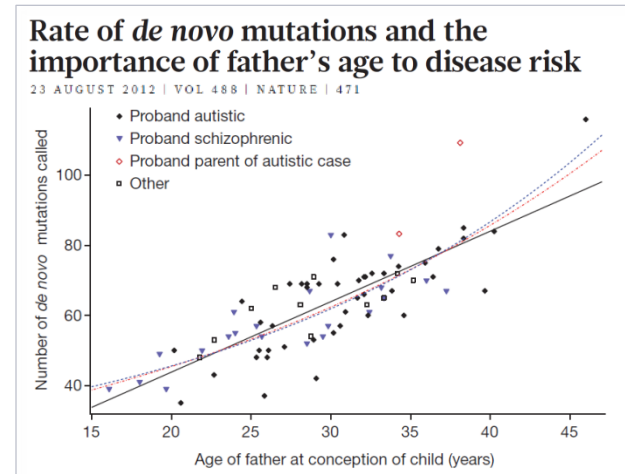
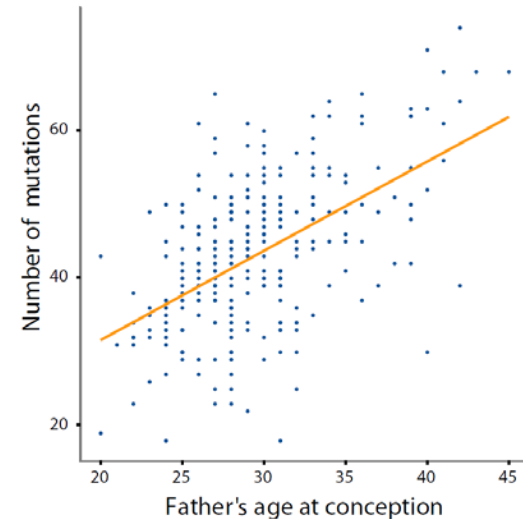
---

- ▶ Each individual carries on average 144 LoF variants:
  - ▶ 60 SNVs
  - ▶ 69 indels
  - ▶ 15 large deletions
- ▶ The majority of these LoF mutations are common, suggesting that these variants are not subject to strong selective pressure, and thus are likely phenotypically benign.

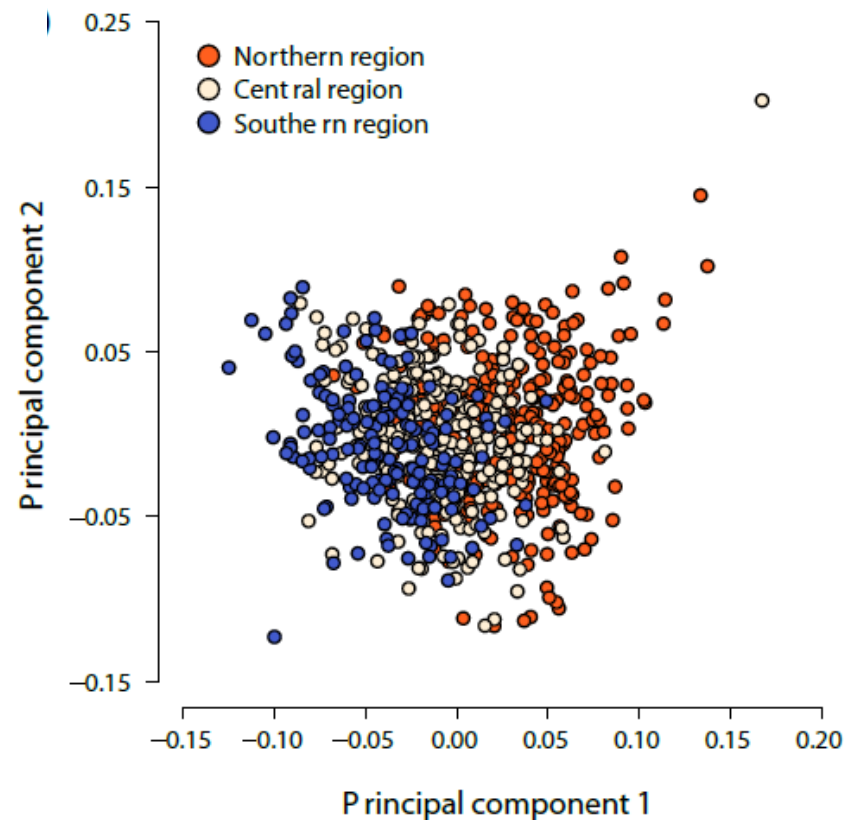
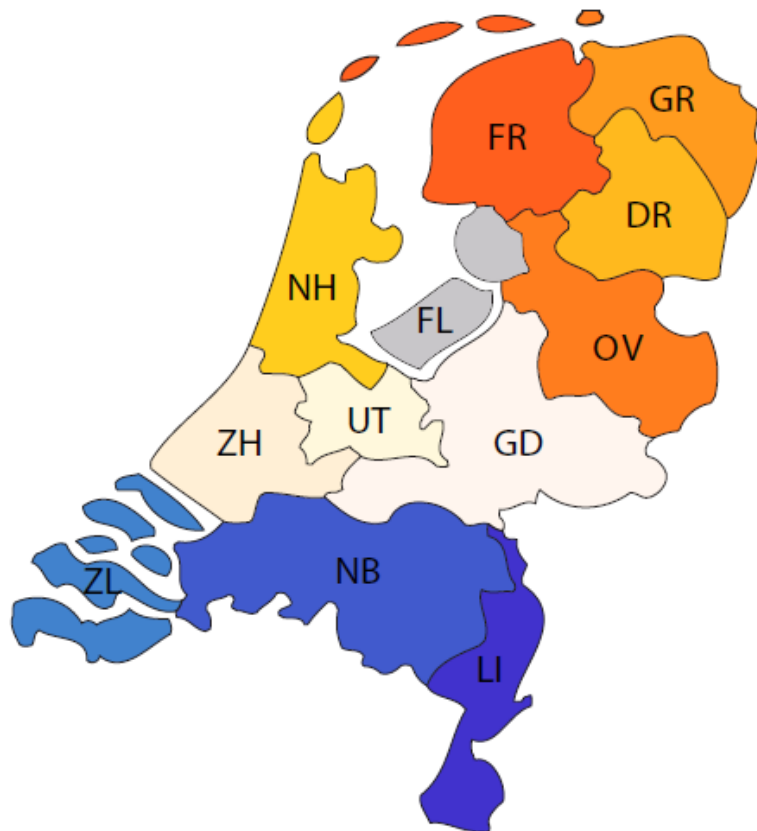


# De novo mutations

- ▶ 11,020 high confidence *de novo* SNVs (18-74 per offspring) were detected.
- ▶ There was a significant positive correlation ( $R^2 = 0.47$ ,  $p < 2.2 \times 10^{-16}$ ) between father's age at conception and number of mutations in the offspring.

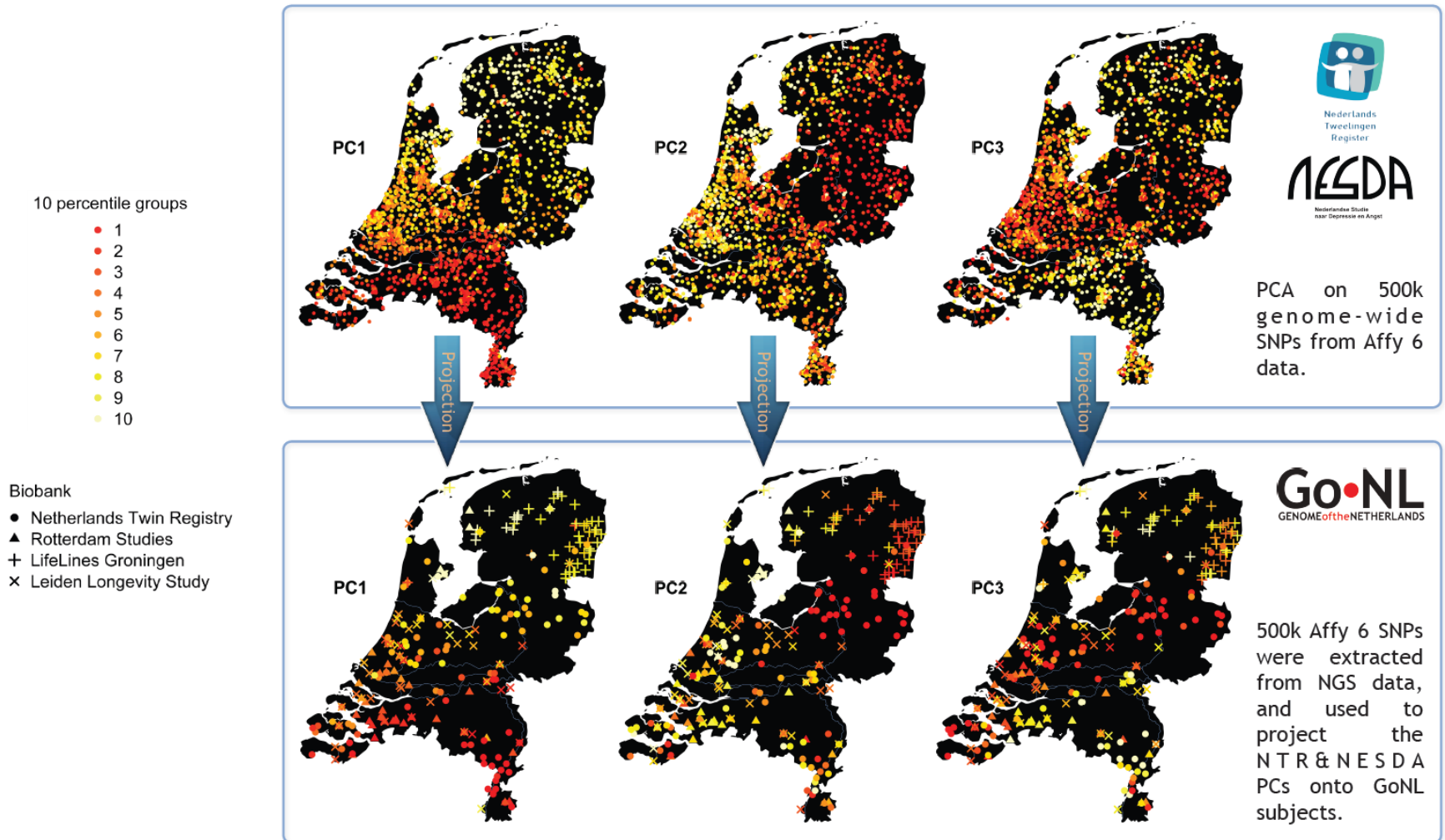


# Population structure in GoNL

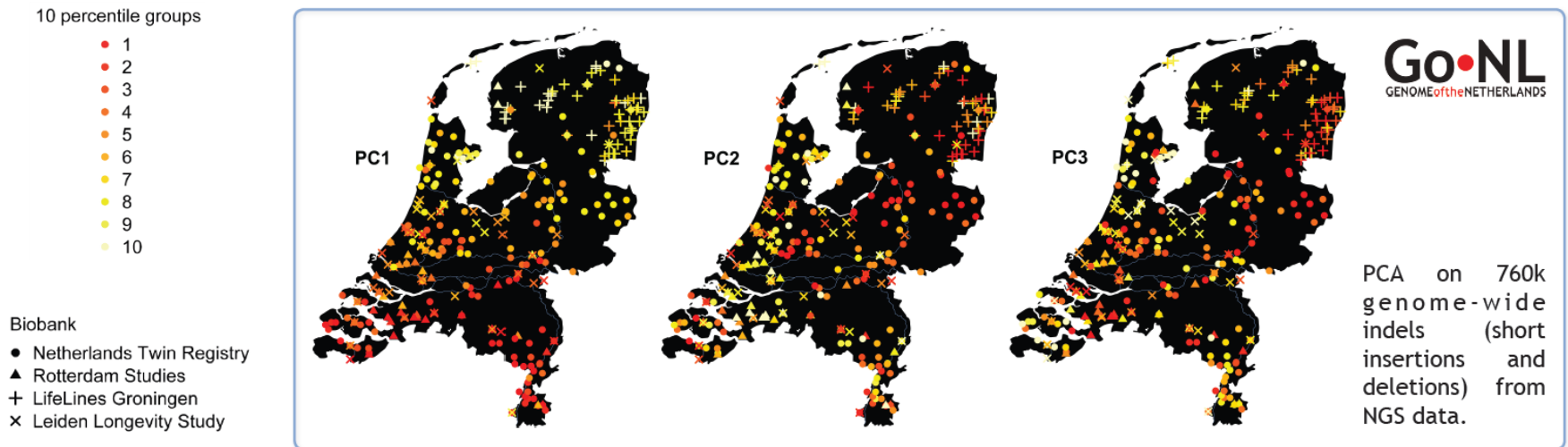




# Ancestry-informative PCs replicated in next-generation sequencing dataset GoNL



# Ancestry-informative PCs replicated in next-generation sequencing dataset GoNL



## Genonderzoek //

**Tussen Nederlanders** in het zuiden en in het noorden bestaan ook genetische **verschillen** // Dit is nu voor het eerst **systematisch** vastgelegd // In de Randstad komen noord en zuid **samen**

nrc.next

# De Nederlander bestaat niet

LUCAS BROUWERS  
REDACTEUR WETENSCHAP

AMSTERDAM. Dat Limburgers van Friezen verschillen weet iedereen. Maar hoe? Genetici van de Vrije Universiteit in Amsterdam vonden dat de verspreiding van genvarianten die onze oogkleur, stofwisseling, hersenen, lichaamslengte en ons immuunsysteem beïnvloeden per windstreek verschilt. Daarmee hebben ze voor het eerst systematisch de subtiele genetische verschillen tussen Nederlanders blootgelegd. Ze beschreven hun resultaten eind vorige week in *European Journal of Human Genetics*.

De grootste genetische verschillen bestaan tussen het noorden en zuiden van Nederland, ontdekten de genetici. De Randstad ligt als een smeltkroes daartussen: hier komen noord en zuid elkaar tegen en worden hun kenmerken de genvarianten

door elkaar gehusseld.

Het gen met de scherpste noord-zuidverdeling was *HERC2*, dat mede bepaalt of iemand bruine of blauwe ogen heeft. Noorderlingen hebben vaker de blauwe variant van dit gen, in het zuiden komen juist bruine ogen vaker voor. De genetici opperen dat de blauwe genvariant door seksuele selectie in frequentie toenam: mannen en vrouwen met een opvallende oogkleur waren aantrekkelijker voor het andere geslacht.

Onder de overige genen met een asymmetrische verspreiding over Nederland waren opvallend veel genen die in de hersenen actief zijn. Hebben de genetici hier een biologische basis gevonden voor die stereotiepe verschillen in temperament tussen noord en zuid, oost en west? „Dat is moeilijk om te zeggen”, zegt geneticus Abdel Abdellaoui, die het onderzoek uitvoerde, aan de telefoon. „Maar het is aannemelijk dat er in het verleden selectie op gedrag plaatsvond. Misschien speelden deze genvarianten daarbij een rol.”

Net als in de rest van Europa bleek de genetische variatie in het zuiden van Nederland groter dan in het noorden. Een gevolg van de *out of Africa*-migratie, denken de onderzoekers, waar de genetische variatie het grootst is. Maar het betekent niet dat zich ook daadwerkelijk een zuid-noordmigratie binnen onze landsgrenzen voltrok. Abdellaoui: „Het kan ook dat Zuid-Europeanen vaker naar Zuid-Nederland migreerden en Noord-Europeanen vaker naar Noord-Nederland. Denk aan de Romeinen, die niet verder kwamen dan de grote rivieren.”

Naast de verschillen tussen noord en

zuid was ook het noordoosten duidelijk van de rest van Nederland te onderscheiden. De derde genetisch afwijkende regio is een strook die door het midden van het land slingert. „Ik dacht even aan de Bijbelgordel toen ik het kaartje zag”, zegt Abdellaoui.

De genetici brachten delen van het genoom in kaart van meer dan 7.000 Nederlanders uit het Nederlands Tweelingen Register en de Nederlandse Studie naar Depressie en Angst (de onderzoekers checkten wel of depressieven afweken, maar niet).



DE FRIEZIN DOUTZEN KROES EN DE LIMBURGER MARK VAN BOMMEL. FOTO ANP

PC1 correlates with height (in the North people are taller)

PC1 correlates strongly with hair & eye color

We did not observe associations with biomarkers and PC1-PC3

# Association Between Population Density and Genetic Risk for Schizophrenia

Lucía Colodro-Conde, PhD; Baptiste Couvy-Duchesne, PhD; John B. Whitfield, PhD; Fabian Streit, PhD; Scott Gordon, PhD; Kathryn E. Kemper, PhD; Loic Yengo, PhD; Zhili Zheng, MD, PhD; Maciej Trzaskowski, PhD; Eveline L. de Zeeuw, PhD; Michel G. Nivard, PhD; Marjolijn Das, PhD; Rachel E. Neale, PhD; Stuart MacGregor, PhD; Catherine M. Olsen, PhD; David C. Whiteman, MBBS, PhD; Dorret I. Boomsma, PhD; Jian Yang, PhD; Marcella Rietschel, PhD; John J. McGrath, MD, PhD; Sarah E. Medland, PhD; Nicholas G. Martin, PhD

European Journal of Human Genetics (2013), 1–9  
© 2013 Macmillan Publishers Limited All rights reserved 1018-4813/13  
www.nature.com/ejhg



## ARTICLE

### Population structure, migration, and diversifying selection in the Netherlands

Abdel Abdellaoui<sup>\*,1</sup>, Jouke-Jan Hottenga<sup>1</sup>, Peter de Knijff<sup>2</sup>, Michel G Nivard<sup>1</sup>, Xiangjun Xiao<sup>3</sup>, Paul Andrew Brooks<sup>4</sup>, Erik A Ehli<sup>5</sup>, Yueshan Hu<sup>5</sup>, Gareth E Davies<sup>5</sup>, James J Hudziak<sup>6</sup>, Patrick F Sullivan<sup>7</sup>, Toos van Beijsterveldt<sup>1</sup>, Gonneke Willemsen<sup>1</sup>, Eco J de Geus<sup>1</sup>, Brenda WJH Penninx<sup>8</sup> and Dorret I. Boomsma<sup>8</sup>

Behav Genet  
DOI 10.1007/s10519-013-9610-1

## ORIGINAL RESEARCH

### Association Between Autozygosity and Major Depression: Stratification Due to Religious Assortment

Abdel Abdellaoui · Jouke-Jan Hottenga · Xiangjun Xiao · Paul Scheet · Erik A. Ehli · Gareth E. Davies · James J. Hudziak · Dirk J. A. Smit · Meike Bartels · Gonneke Willemsen · Andrew Brooks · Patrick F. Sullivan · Johannes H. Smit · Eco J. de Geus · Brenda W. J. H. Penninx · Dorret I. Boomsma

OPEN ACCESS Freely available online



### Educational Attainment Influences Genome-Wide Runs of Homozygosity through Migration and Assortative Mating

Abdel Abdellaoui, Jouke-Jan Hottenga, Gonneke Willemsen, Meike Bartels, Toos van Beijsterveldt, Erik A. Ehli, Gareth E. Davies, Andrew Brooks, Patrick F. Sullivan, Brenda W.J.H. Penninx, Eco J. de Geus, Dorret I. Boomsma

