B B M R I • N L

Population structure in the Netherlands: short indels and deletions from NGS

Go•NL GENOMEofthe NETHERLANDS

Abdel Abdellaoui, Victor Guryev, Laurent Francioli, Jayne Y. Hehir-Kwa, Wigard Kloosterman, Tobias Marschall, Alexander Schoenhuth, Eric-Wubbo Lameijer, Slavik Koval, Fereydoun Hormozdiari, Joep de Ligt, Najaf Amin, Freerk van Dijk, Lennart Karssen, Hailiang Mei, Evan E. Eichler, Gert-Jan van Ommen, Paul de Bakker, Cisca Wijmenga, Cock van Duijn, Eline Slagboom, Dorret I. Boomsma, Kai Ye on behalf of GoNL Consortium

Background

While microarray data have contributed much to population genetics, the higher resolution of whole-genome sequence data is expected to yield new insights about population stratification, population history, the prevalence of selection pressures, and the identification of functional variants under selection. Ancestry differences in the Netherlands show clear geographic distributions, as previously mapped using principal component analysis (PCA) on microarray SNPs. Ancestry-informative PCs can reveal the consequences of population history, and can also be used to detect selection pressures and traces of migration.

Methods

The first 10 PCs were computed on 490 unrelated Dutch individuals using LD-pruned common indels and deletions. Ancestry-informative PCs were identified by correlating PCs with geography and three previously identified ancestry informative SNP PCs (500k Affy 6 SNPs). F_{st}'s were computed for each ancestry-informative PC by comparing the top 150 with the bottom 150 subjects for each PC according to Weir & Cockerham on three sets of common variants: 8,685,291 SNPs, 838,979 short indels (<20 bp), and 17,140 deletions (20bp-10kb).

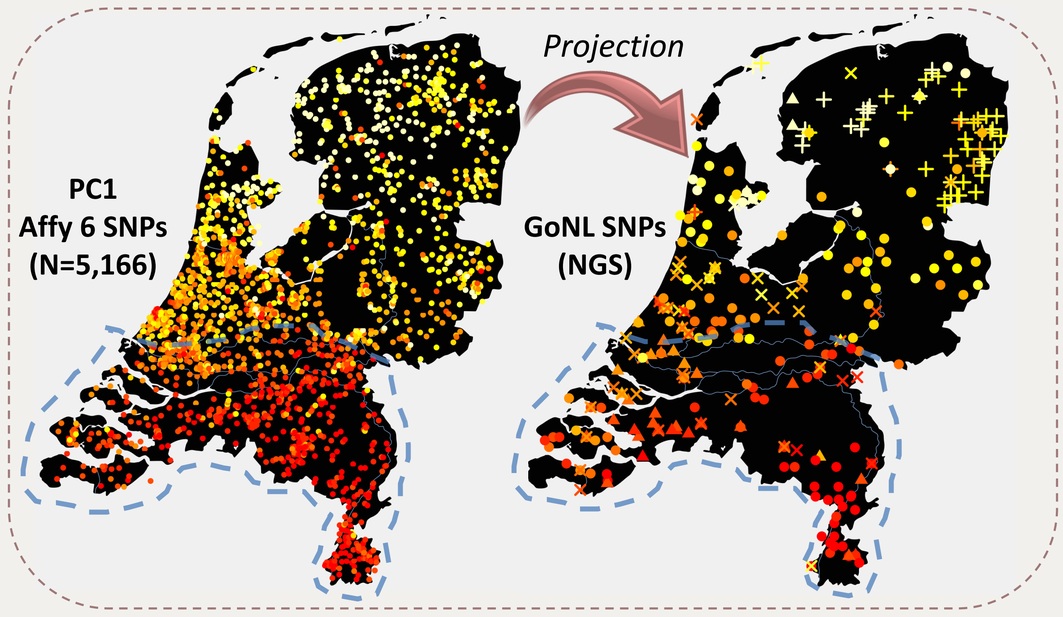
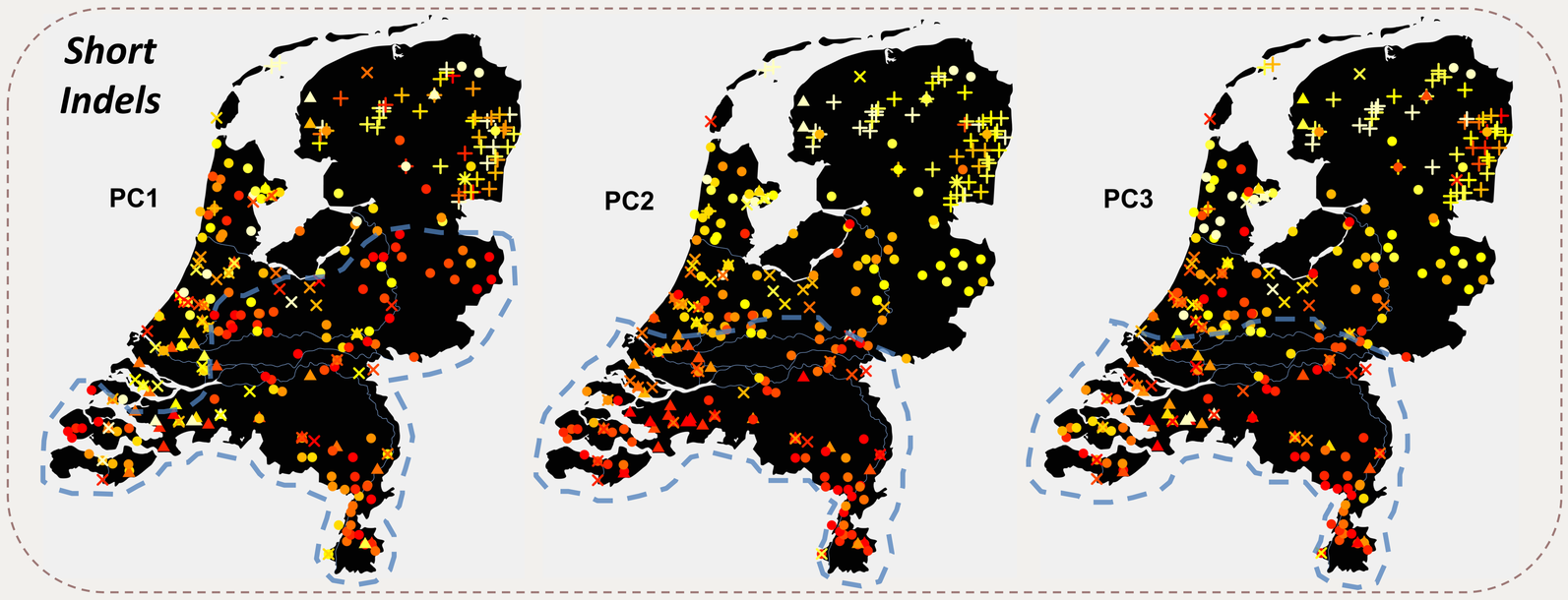


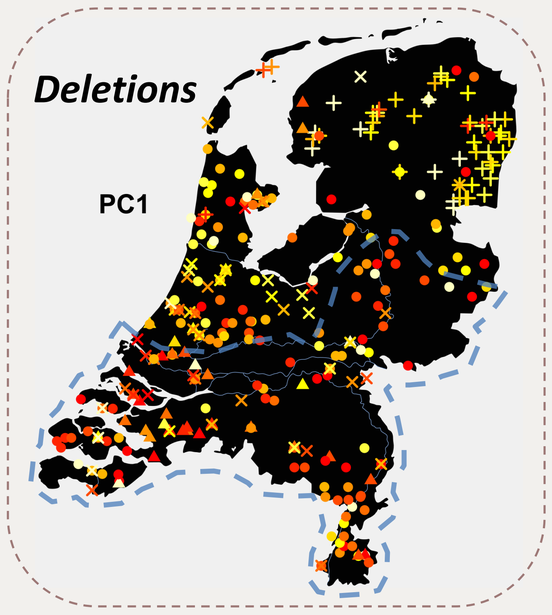
Table: Pearson correlations of PCs with genome-wide homozygosity (F), height, and the European North-South PC from 1000 Genomes

PCs	F	Height	Eur ↕PC
PC1 from 500k Affy 6 SNPs:			
Affy 6 PC1	***r* = .25**	***r* = .15**	***r* = .66**
GoNL PC1	***r* = .36**	***r* = .12**	***r* = .78**
Short indels:			
PC1	***r* = .26**	*r* = .05	***r* = .17**
PC2	***r* = .18**	***r* = .11**	***r* = .55**
PC3	***r* = .20**	***r* = .18**	***r* = .30**
Deletions:			
PC1	***r* = .10**	*r* = .02	***r* = .26**

Bold: p < .05; Red: p < .001

Previous research showed that the correlation between the North-South cline and genome-wide homozygosity (F) is likely due to a European serial founder effect. The Dutch North-South cline shows decreasing homozygosity and increasing LD in the North, and correlates highly with the European North-South cline (.66), where the same trend is observed. The correlation between the Dutch North-South cline and height is also in the same direction in Europe.





Results

Indels showed six, and deletions showed five putative ancestry-informative PCs with significant correlations with geography and/or ancestry-informative SNP PCs ($p < 5 \times 10^{-4}$). For both indels and deletions, the PC explaining most variation (PC1) shows a different North-South distribution than PC1 from Affy 6 SNPs, does not correlate with height, and shows a weaker correlation with the European North-South cline.

Genome-wide, F_{st}'s did not differ significantly between coding & non-coding, or between genic & non-genic regions.

Conclusions

Indels and deletions from NGS capture ancestry differences not observed with micro-array SNP data. Several North-South distributions are observed that vary in their correlations with genome-wide homozygosity, height, and the European North-South cline. Diversifying selection pressures are not visible on a genome-wide level. Significant diversifying selection pressures have been observed previously through F_{st} outlier analyses in micro-array data from this population. Similar analyses on NGS data are on the way.

Contact information: a.abdellaoui@vu.nl

VRIJE UNIVERSITEIT AMSTERDAM