THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels

Citation for published version:
Genome Netherlands Consortium, van Leeuwen, EM, Karssen, LC, Deelen, J, Isaacs, A, Medina-Gomez, C, Mbarek, H, Kanterakis, A, Trompet, S, Postmus, I, Verweij, N, van Enckevort, DJ, Huffman, JE, White, CC, Feitosa, MF, Bartz, TM, Manichaikul, A, Joshi, PK, Peloso, GM, Deelen, P, van Dijk, F, Willemsen, G, de Geus, EJ, Milaneschi, Y, Penninx, BWJH, Francioli, LC, Menelaou, A, Pulit, SL, Rivadeneira, F, Hofman, A, Oostra, BA, Franco, OH, Leach, IM, Beekman, M, de Craen, AJM, Uh, H-W, Trochet, H, Hocking, LJ, Porteous, DJ, Sattar, N, Packard, CJ, Buckley, BM, Brody, JA, Bis, JC, Campbell, H, Wilson, JF, Hayward, C, Polasek, O, Vitart, V, Rudan, I & Wright, AF 2015, 'Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels', *Nature Communications*, vol. 6, 6065. https://doi.org/10.1038/ncomms7065

Digital Object Identifier (DOI):
10.1038/ncomms7065

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Nature Communications

Publisher Rights Statement:
This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 25. Apr. 2024

nature COMMUNICATIONS

ARTICLE

Received 9 Jul 2014 | Accepted 9 Dec 2014 | Published 9 Mar 2015

DOI: 10.1038/ncomms7065

OPEN

Genome of the Netherlands population-specific imputations identify an *ABCA6* variant associated with cholesterol levels

Elisabeth M. van Leeuwen[1], Lennart C. Karssen[1], Joris Deelen[2], Aaron Isaacs[1], Carolina Medina-Gomez[3], Hamdi Mbarek[4], Alexandros Kanterakis[5], Stella Trompet[6], Iris Postmus[7], Niek Verweij[8], David J. van Enckevort[9], Jennifer E. Huffman[10], Charles C. White[11], Mary F. Feitosa[12], Traci M. Bartz[13], Ani Manichaikul[14], Peter K. Joshi[15], Gina M. Peloso[16], Patrick Deelen[5], Freerk van Dijk[5], Gonneke Willemsen[4], Eco J. de Geus[4], Yuri Milaneschi[17], Brenda W.J.H. Penninx[17], Laurent C. Francioli[18], Androniki Menelaou[18], Sara L. Pulit[18], Fernando Rivadeneira[3], Albert Hofman[1], Ben A. Oostra[19], Oscar H. Franco[1], Irene Mateo Leach[8], Marian Beekman[2], Anton J.M. de Craen[7], Hae-Won Uh[20], Holly Trochet[10], Lynne J. Hocking[21], David J. Porteous[22], Naveed Sattar[23], Chris J. Packard[24], Brendan M. Buckley[25], Jennifer A. Brody[26], Joshua C. Bis[26], Jerome I. Rotter[27], Josyf C. Mychaleckyj[14], Harry Campbell[15], Qing Duan[28], Leslie A. Lange[28], James F. Wilson[15], Caroline Hayward[10], Ozren Polasek[29], Veronique Vitart[10], Igor Rudan[15], Alan F. Wright[10], Stephen S. Rich[14], Bruce M. Psaty[30], Ingrid B. Borecki[31], Patricia M. Kearney[25], David J. Stott[24], L. Adrienne Cupples[11,32], The Genome of the Netherlands Consortium*, J. Wouter Jukema[6], Pim van der Harst[8], Eric J. Sijbrands[33], Jouke-Jan Hottenga[4], Andre G. Uitterlinden[3], Morris A. Swertz[5], Gert-Jan B. van Ommen[34], Paul I.W. de Bakker[18,35], P. Eline Slagboom[2], Dorret I. Boomsma[36], Cisca Wijmenga[37] & Cornelia M. van Duijn[1]

Variants associated with blood lipid levels may be population-specific. To identify low-frequency variants associated with this phenotype, population-specific reference panels may be used. Here we impute nine large Dutch biobanks (~35,000 samples) with the population-specific reference panel created by the Genome of the Netherlands Project and perform association testing with blood lipid levels. We report the discovery of five novel associations at four loci (P value $<6.61\times10^{-4}$), including a rare missense variant in *ABCA6* (rs77542162, p.Cys1359Arg, frequency 0.034), which is predicted to be deleterious. The frequency of this *ABCA6* variant is 3.65-fold increased in the Dutch and its effect ($\beta_{LDL\text{-}C}=0.135$, $\beta_{TC}=0.140$) is estimated to be very similar to those observed for single variants in well-known lipid genes, such as *LDLR*.

[1] Department of Epidemiology, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands. [2] Department of Molecular Epidemiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. [3] Department of Epidemiology and Internal Medicine, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands. [4] Department of Biological Psychology, VU University Amsterdam and EMGO+ Institute for Health and Care Research, Amsterdam 1081BT, The Netherlands. [5] Department of Genetics, Genomics Coordination Center, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands. [6] Department of Cardiology, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. [7] Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. [8] Department of Cardiology, University Medical Center Groningen, University of Groningen, Groningen 9700 RB, The Netherlands. [9] BioAssist, Netherlands Bioinformatics Center, Nijmegen 6500 HB, The Netherlands. [10] MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Edinburgh EH4 2XU, UK. [11] Department of Biostatistics, Boston U School of Public Health, Boston, Massachusetts 02118, USA. [12] Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63108, USA. [13] Department of Biostatistics and Medicine, University of Washington, Seattle, Washington 98101, USA. [14] Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia 22908, USA. [15] Centre for Population Health Sciences, University of Edinburgh, Edinburgh, Scotland EH8 9AG, UK. [16] Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02176, USA. [17] Department of Psychiatry, VU University Medical Center Amsterdam/GGZinGeest, EMGO+ Institute for Health and Care Research, Neuroscience Campus Amsterdam, Amsterdam 1081HL, The Netherlands. [18] Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht 3584 CG, The Netherlands. [19] Department of Clinical Genetics, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands. [20] Department of Genetical Statistics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. [21] Division of Applied Health Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK. [22] Centre for Genomic and Experimental Medicine, MRC IGMM, University of Edinburgh, Edinburgh EH4 2XU, UK. [23] BHF Glasgow Cardiovascular Research Centre, Faculty of Medicine, University of Glasgow, Glasgow G12 8QQ, UK. [24] Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow G12 8QQ, UK. [25] Department of Pharmacology and Therapeutics, University College Cork, Cork, Ireland. [26] Department of Medicine, University of Washington, Seattle, Washington 98101, USA. [27] Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, California 90502, USA. [28] Department of Genetics, University of North Carolina, Chapel Hill, North Carolina NC 27599, USA. [29] Department of Public Health, Faculty of Medicine, University of Split, Split 21000, Croatia. [30] Department of Medicine and Epidemiology, University of Washington, Seattle, Washington 98101, USA. [31] Department of Genetics and Biostatistics, Washington University School of Medicine, St Louis, Missouri 63108, USA. [32] Framingham Heart Study, Framingham, Massachusetts 01702, USA. [33] Department of Internal Medicine, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands. [34] Department of Human Genetics, Leiden University Medical Center, Leiden P.O. Box 9600, 2300 RC, The Netherlands. [35] Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht 3584 CG, The Netherlands. [36] Department of Biological Psychology, VU University Amsterdam, Amsterdam 1081BT, The Netherlands. [37] Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen 9700 RB, The Netherlands. Correspondence and requests for materials should be addressed to C.M.v.D. (email: c.vanduijn@erasmusmc.nl).

*A full list of consortium members appears at the end of the paper.

© 2015 Macmillan Publishers Limited. All rights reserved.

Genome-wide association studies (GWAS) have identified a large number of loci associated with blood lipid levels and analysis suggest there are additional susceptibility loci that have not yet been discovered[1–3]. Despite the fact that rare functional variants are known to play a major role in lipid metabolism[1–3], there has been limited success in finding such variants in population-based studies using next-generation sequencing. Even if the effect of these variants is expected to be larger than that of common variants, the sample size needed to detect these rare or low-frequency variants increases dramatically with variant rarity. As the frequency of rare variants may increase in certain populations because of drift and founder effects[4], the power of searches for rare functional variants may improve by the use of reference sets specific to distinct populations. Such references allow for better quality imputation of rare variants especially those with increased frequency in the population of interest[3,5,6]. Previous studies have successfully detected rare variants by imputation into larger sets of individuals in isolated populations followed by association testing to detect variants associated with the trait of interest[7–9].

Here we describe an imputation-based GWAS for circulating lipid levels using a custom-built reference panel for the Dutch population (Genome of the Netherlands, GoNL, http://www.nlgenome.nl/), in which the whole genomes of 250 parent–offspring trios were sequenced at $\sim 13\times$ coverage[5,6]. Owing to the trio design, the phasing quality of the reference panel was better than that of the 1000 Genomes (1-kG) Phase 1 panel. In this study we show that using this population-specific reference panel we were able to identify five novel associations at four loci.

Results

Nine large Dutch epidemiological cohorts (comprising 36,000 samples in total) were imputed with the GoNL reference panel ($\sim$19.5 million single-nucleotide polymorphisms (SNPs)) on an identical protocol[6,10]. All cohorts conducted association analysis on the imputed variants assuming an additive genetic effect on high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), total cholesterol (TC) and triglyceride (TG) levels (Methods, Supplementary Methods and Supplementary Table 1), and the results were meta-analysed. We used conditional analysis implemented in GCTA[11] to identify variants associated independently with lipid levels.

Both rare (minor allele frequency (MAF) <0.01), low ($0.01<\text{MAF}<0.05$) and common variants ($\text{MAF}>0.05$) were associated with HDL-C ($N=60$ variants), LDL-C ($N=142$ variants), TC ($N=134$ variants) and TG ($N=16$ variants) in both known and novel loci (Methods, Supplementary Tables 2–5 and Supplementary Fig. 1). In Fig. 1 we compare the allele frequencies that reach genome-wide significance in the GCTA analysis (P value $<5\times10^{-8}$) to those reported in refs 1,2 (Fig. 1). The majority of the known HDL-C (31 of 45, 68.9%), LDL-C (24 of 34, 70.6%), TC (33 of 48, 68.6%) and TG (13 of 30, 43.3%) loci described in ref. 1 replicated at a P value $<3.18\times10^{-4}$ (Bonferroni correction based on 157 variants;

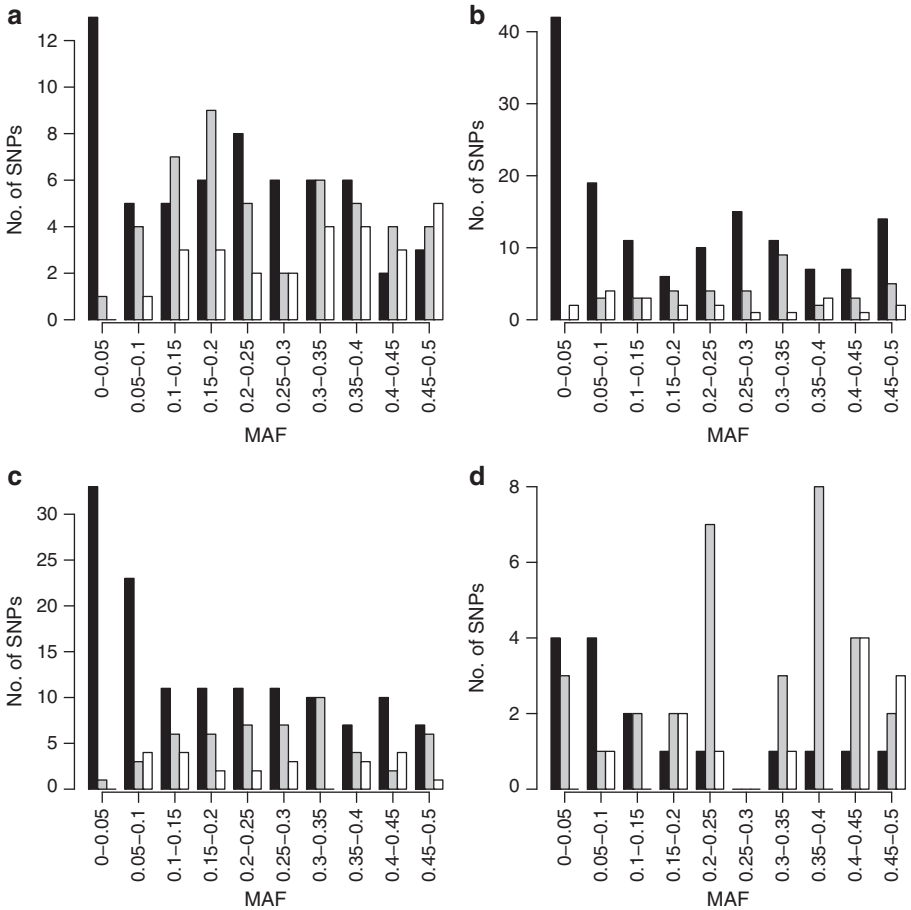

Figure 1 | Identified variants for plasma lipid levels. Distribution of the variants identified by conditional analysis implemented by GCTA to be independently associated with the lipid traits (**a**) HDL-C (60 variants), (**b**) LDL-C (142 variants), (**c**) TC (134 variants) and (**d**) TG (16 variants)) over MAF bins after meta-analysis of discovery cohorts (black). The histograms also include loci identified in ref. 1 (grey) and ref. 2 (white).

© 2015 Macmillan Publishers Limited. All rights reserved.

Table 1 | Summary descriptions for the variants associated with HDL-C, LDL-C, TC or TG.

SNP	Chr	Position	EA	NEA	Gene	MAF_{GoNL}	$MAF_{1\text{-}kG}$	$MAF_{GoNL}/MAF_{1\text{-}kG}$ (*P* value for two population proportions)
rs4752801	11	47,907,641	G	A	Close to the *NUP160*	0.347	0.338	1.027 (0.258)
rs149580368	17	41,874,745	A	C	Between *C17orf105* and *MPP3*	0.029	0.015	1.923 (<0.0001)
rs77542162	17	67,081,278	G	A	*ABCA6*	0.030	0.008	3.647 (<0.0001)
rs144984216	19	20,479,901	T	C	*ZNF826P*	0.028	0.011	2.555 (<0.0001)
rs117162033	19	8,627,569	T	C	*MYO1F*	0.007	0.007	0.957 (<0.0001)

EA, effect allele; GoNL, Genome of the Netherlands; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; MAF_{GoNL} and $MAF_{1\,kG}$, the minor allele frequency of the effect allele in the GoNL reference panel and in the 1-kG reference panel (Phase 1 integrated release v3, April 2012, all ancestries), respectively; NEA, non-effect allele; SNP, single-nucleotide polymorphism; TC, total cholesterol; TG, triglyceride.

Table 2 | Results for the variants associated with HDL-C, LDL-C, TC or TG.

Trait	SNP	Discovery phase						Replication phase						Combined discovery and replication			
		N	MAF	Rsq	β	s.e.$_\beta$	*P* value	*N*	MAF	Rsq	β	s.e.$_\beta$	*P* value	MAF	β	s.e.$_\beta$	*P* value
HDL-C	rs4752801	33,613	0.355	0.992	−0.023	0.003	1.62E−12	31,422	0.362	0.985	−0.012	0.003	5.63E−05	0.359	−0.017	0.002	8.39E−15
HDL-C	rs149580368	36,000	0.036	0.674	−0.075	0.010	4.23E−14	21,281	0.023	0.621	−0.079	0.014	5.90E−09	0.031	−0.077	0.008	1.53E−21
LDL-C	rs77542162	35,624	0.034	0.734	0.135	0.023	6.67E−09	21,969	0.026	0.773	0.125	0.031	4.35E−05	0.031	0.131	0.019	1.33E−12
TC	rs77542162	36,109	0.034	0.731	0.140	0.025	1.29E−08	29,196	0.027	0.785	0.095	0.028	6.61E−04	0.031	0.120	0.019	7.31E−11
TC	rs144984216	31,622	0.046	0.573	−0.140	0.024	7.88E−09	24,913	0.025	0.632	−0.056	0.036	1.22E−01	0.039	−0.114	0.020	1.58E−08
TG	rs117162033	26,122	0.016	0.511	−0.143	0.025	8.02E−09	10,296	0.021	0.573	−0.133	0.030	7.98E−06	0.018	−0.139	0.019	3.10E−13

HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; MAF, minor allele frequency; SNP, single-nucleotide polymorphism; TC, total cholesterol; TG, triglyceride. MAF, the weighted average of minor allele frequency for the effect allele across all studies in the discovery phase, replication phase or combined, respectively. *N*, sample size after QC. Rsq, the mean imputation quality of all cohorts. β is the effect of the effect allele in mmol l^{-1}.

Methods, Supplementary Figs 2 and 3 and Supplementary Tables 6–7). We also confirmed several of the HDL-C (6 of 27, 22.2%), LDL-C (7 of 21, 33.3%), TC (4 of 23, 17.4%) and TG (1 of 12, 8.3%) loci described in ref. 2 at a *P* value $<6.02\times10^{-4}$ (Bonferroni correction based on 83 variants) despite a sample size of ~20% of the other studies.

To identify novel loci associated with blood lipid levels, we selected from the list of variants identified by GCTA, those variants located more than 1 Mb away from previously identified loci. This resulted in six novel associations at five loci (Methods, Tables 1 and 2 and Supplementary Table 8). The five loci are not in linkage disequilibrium (LD) with previously described GWAS loci (Methods and Supplementary Table 9). Conditional analysis in the discovery cohorts showed that these new variants were independent from previously identified loci (Supplementary Table 10 and Supplementary Fig. 4). Of the five loci, three (rs149580368, rs77542162 and rs144984216) have an increased frequency in GoNL compared with 1-kG (Phase 1 integrated release v3, April 2012, all ancestries; Table 1), suggesting that there may have been genetic drift in the Dutch population for these loci[4]. Yet, as each of these loci has a MAF>0.005, we assumed that these alleles also segregate in other populations of European descent[4], such as those of the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. Therefore, we set out replication in independent samples from the CHARGE cohorts using the 1-kG reference panel (Phase 1 integrated release v3, April 2012, all ancestries). We were able to replicate five out of the six variants using the Bonferroni-corrected *P* value threshold of 8.33×10^{-3} (Table 2, Methods and Supplementary Table 11).

Of the replicated variants, rs77542162 is the most interesting variant. This missense variant is associated with both LDL-C and TC (Supplementary Figs 5 and 6) and is located on chromosome 17 within the *ABCA6* gene (ATP-binding cassette, subfamily A (ABC1), member 6). The frequency of this variant is 1.31-fold higher in the discovery cohorts than in the replication cohorts and even 3.65-fold higher in the GoNL population than in the 1-kG population. This missense variant changes the amino acid cysteine into arginine at position 1359 (Cys1359Arg) and is predicted to be damaging for the structure and function of the protein by Polyphen2 (ref. 12), MutationTaster[13] and LRT[14]. The effect size of rs77542162 ($\beta_{LDL\text{-}C}=0.135$ and $\beta_{TC}=0.140$) is very similar to those observed for other single variants in well-known lipid genes, such as *LDLR* and *CETP*, as reported in ref. 1. The membrane-associated protein encoded by this gene is a member of the superfamily of ATP-binding cassette (ABC) transporters that transport various molecules across extra- and intracellular membranes. This protein is a member of the ABC1 subfamily, which is the only major ABC subfamily found exclusively in multicellular eukaryotes. *ABCA6* is clustered with four other ABC1 family members on chromosome 17q24 and appears to play a role in macrophage lipid homeostasis.

One other replicated variant, rs149580368, is also enriched with a 1.92-fold increase in frequency in the Dutch population compared with the 1-kG population. This intergenic variant (Supplementary Fig. 7), without a significant *cis*-eQTL effect, is located between the protein-coding genes *C17orf105* (chromosome 17 open reading frame 105) and *MPP3* (membrane protein, palmitoylated 3). Two replicated variants have similar frequencies in the GoNL and 1-kG reference sets: rs4752801 (Supplementary Fig. 8), an new intergenic variant with a high frequency (MAF = 0.355) that is located in a region previously identified[1], and rs117162033 (Supplementary Fig. 9), an intronic variant in the myosin F (*MYO1F*)-coding gene. *C17orf15, MPP3* and *MYO1F* have no known impact on lipid levels. As the imputation quality of rs117162033 is lower than the other variants, we validated the imputation of this variant using the same approach as published in ref. 15. We compared in a random sample of 65 participants of the GoNL reference panel their sequence and best-guess GoNL-imputed genotypes and found that the concordance was 100% (all participants were correctly imputed). The association between TG and the intronic variant in the *MYO1F* gene is remarkable because of the low frequency of the variant. This confirms the conclusions as published before about the GoNL reference panel, that the trio-based phasing contributed significantly to the imputation quality of rare variants[5].

© 2015 Macmillan Publishers Limited. All rights reserved.

In this current study, the GoNL reference panel was used for imputations of the discovery cohorts and the 1-kG reference panel for the imputation of the replication cohorts. Although it would be interesting to impute with a combined reference panel of both the GoNL data, the 1-kG data and other sequence data, this effort is ongoing.

This study shows that the imputation of a population-specific reference panel into large epidemiological cohorts can reveal both low-frequency and rare variants associated with blood lipid levels using classical association testing approaches. The three variants with increased frequency in the Dutch population as compared with the 1-kG population include a rare, predicted to be deleterious missense variant in *ABCA6,* which has increased frequency 3.65 times larger in the Dutch population. The effect of this variant is comparable to that of variants in the *LDLR* gene, a gene for which several population-based screening programmes have been initiated. Our findings suggest that next-generation-sequencing effort may yield clinically relevant findings. Our paper further shows that next-generation-sequencing efforts in *specific homogeneous* populations as the Dutch may yield clinically relevant findings *worldwide*.

Methods

Study descriptions. The descriptions of the including cohorts can be found in the Supplementary Methods. A written informed consent was obtained from all study participants for all cohorts and local ethical committees at participating institutions approved individual study protocols.

Study samples and phenotypes. A summary of the details of both the discovery and replication cohorts participating in this study can be found in Supplementary Tables 1 and 12.

Only samples of Dutch ancestry were used in the discovery cohorts; the samples in the replication cohorts are from various ancestries (see Supplementary Table 12). In all studies, except MESA Whites, all individuals who used lipid-lowering medication at the time the lipid levels were measured, were excluded. In MESA Whites, the total cholesterol values for individuals on lipid-lowering medication were divided by 0.8. In all studies except for LLS and PREVEND, the subjects were fasting when the lipid levels were measured. In LLS all samples were non-fasted and in PREVEND 2.99% were non-fasted. The LDL-C levels were measured within the ERF, Croatia-Korcula, Croatia-Split, Croatia-Vis, FamHS and Lifelines cohorts, within the other cohorts the Friedewald equation was used to calculate the LDL-C levels[16].

The lipid measurements were adjusted for sex, age and age^2 in all cohorts. Various methods were used to account for family relationships: in ERF grammar-gamma, GenABEL version 1.7.6 (refs 17,18) was used; in the Croatia-Korcula, Croatia-Split, Croatia-Vis and Generation Scotland cohorts mmscore (GenABEL)[17] was used; and in LLS, qt-assoc was used. In CHS the clinic was used as extra covariate, in Lifelines PC1 and PC2, in FamHS the field centre, the genotyping array (Illumina 550 k, 610 k and 1 M), PC5 only for TC and PC1 only for LDL, in FHS the cohort (offspring and third generation) and PCs, in MESA Whites 2 PCs and study site, in NTR-NESDA PCs and chip effect, in ORCADES the genotyping array and PC1, PC2 and PC3, in PROSPER-Dutch only PC1 and in both PROSPER-Scottish and PROSPER-Irish PC1-PC4.

Genotyping and imputations. Detailed information about genotyping and imputations per cohort can be found in the Supplementary Methods. In summary, all cohorts were genotyped using commercially available Affymetrix or Illumina genotyping arrays, or custom Perlegen arrays. Quality control was performed independently for each study. To facilitate meta-analysis, each replication cohort performed genotype imputation using IMPUTE[19] or Minimac[20] with reference to the GoNL project data for the discovery cohorts and with reference to the 1-kG project data for the replication cohorts.

GWAS in all discovery cohorts. All nine discovery cohorts ran separate the genome-wide association study for each of the four traits: HDL-C, LDL-C, TC and TG. Supplementary Table 13 shows the genomic control factor λ per trait per cohort and Supplementary Figs 10–13 show the λ per MAF bin per trait per cohort. We therefore used only the SNPs with a $R^2 > 0.3$, $R^2 < 1.1$ and expected minor allele count (expMAC $= 2 \times$ MAF $\times R^2 \cdot$ sample size) > 10. Most inflations are observed within the ERF study, especially in the lowest-frequency variants, which is probably caused by the family structure in this cohort.

Meta-analysis of discovery cohorts. The association results of all studies were combined and the s.e.-based weights were calculated using METAL[21]. This tool also applies genomic control by automatically correcting the test statistics to account for small amounts of population stratification or unaccounted relatedness. METAL also allows for heterogeneity. We used the following filters: $0.3 < R^2 < 1.1$ and expMAC > 10.

After meta-analyses of all available variants, we excluded the variants that are not present in at least six of the nine cohorts. We also excluded all variants that are labelled as being in the inaccessible genome, since the quality of those SNPs cannot be guaranteed[22]. The remaining variants per trait, see Supplementary Table 14, were used to create Manhattan plots and QQ plots, see Supplementary Figs 14 and 15. The meta-analysis resulted in 1,905 SNPs with a P value less than 5×10^{-8} for HDL-C, 2,626 SNPs for LDL-C, 3,133 SNPs for TC and 1,310 for TG.

Confirmation of known loci. Previously, Teslovich *et al.*[1] and Willer *et al.*[2] identified 157 loci associated with one of more of the lipids. Teslovich *et al.*[1] identified 47, 37, 52 and 32 loci to be associated with HDL-C, LDL-C, TC and TG, respectively. The positions of these loci were reported on human genome build 36; we therefore lifted these positions over to human genome build 37 and checked the association results after the meta-analysis of all discovery cohorts. The effect size of these loci was reported in mg dl^{-1}, whereas in this study we use mmol l^{-1}. We therefore multiplied the effect size for the loci associated with TG with 0.0259 and the other loci with 0.011. Supplementary Fig. 2 and Supplementary Table 6 show the comparison per trait of our meta-analysis of all discovery cohorts with the results of the meta-analysis in ref. 1. We did the same for the loci identified in ref. 2, see Supplementary Fig. 3 and Supplementary Table 7. The effect size of these loci could not be compared with our results, since trait residuals within each study participating in the meta-analysis in ref. 2 were adjusted for sex and age^2 and subsequently quantile normalized. Their GWAS was performed with the inverse normal transformed trait values.

Selection of independent variants. In order to select only associated variants that were independent of previous findings, we used the GCTA tool[11]. This tool performs a stepwise selection procedure to select multiple associated SNPs by a conditional and joint analysis approach using summary-level statistics from a meta-analysis and LD corrections between SNPs estimated from the GoNL reference panel, release 4. This analysis revealed 60 independent variants associated with HDL-C, 142 independent variants associated with LDL-C, 134 independent variants associated with TC and 16 independent variants associated with TG. By using this approach, we were able to identify additional independent variants in known loci. Figure 1 shows that we identified both common and rare variants and more rare variants compared with refs 1,2. There is an overlap between the genome-wide significant SNPs of the different traits, and also between the independent SNPs of the different traits, as shown in Supplementary Fig. 1.

Identification of potential novel variants. To identify potential novel variants, we first excluded all variants within 1 Mb of a known loci from refs 1,2. Since the number of loci associated with the four traits differ, we end up with 7,946,245 SNPs for HDL-C, 8,014,693 SNPs for LDL-C, 7,923,530 SNPs for TC and 7,468,790 SNPs for TG. For all traits we do find some genome-wide significant loci, see Supplementary Figs 16 and 17. We used the GCTA tool to select only those variants that are independently associated with the lipid trait. This analysis revealed two novel independent variants associated with HDL-C, one novel independent variant associated with LDL-C, two novel independent variants associated with TC and one novel independent variants associated with TG, see Supplementary Table 8 and Supplementary Fig. 18. We used PLINK to test whether these six variants are in LD with the known loci from refs 1,2. None of the six variants are in LD with known loci associated with the same trait on the same chromosome ($R^2 < 0.14$).

Replication of potential novel variants. The six potential novel loci were replicated in 11 cohorts: CHS, Croatia-Korcula, Croatia-Split, Croatia-Vis, FamHS, FHS, Generation Scotland, MESA Whites, ORCADES, PROSPER-Scottish and PROSPER-Irish. The association results of all cohorts were combined and the s.e.-based weights were calculated using METAL[21]. The Bonferroni correction for multiple testing was 8.33×10^{-3}. This resulted in the significant replication of five out of the six variants, see Supplementary Fig. 19 and Supplementary Table 11.

Conditional analysis. Within the discovery cohorts we performed a conditional analysis to see whether the novel variants are independent of the known loci from refs 1,2. Supplementary Table 10 shows the results within these cohorts with and without adjusting for the known loci for the trait in question, if available in the GoNL reference panel. Since the unadjusted and adjusted results are similar, we conclude that the newly identified variants are independent of the known loci.

References

1. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466,** 707–713 (2010).

© 2015 Macmillan Publishers Limited. All rights reserved.

2. Global Lipids Genetics Consortium *et al.* Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45,** 1274–1283 (2013).
3. Willer, C. J. & Mohlke, K. L. Finding genes and variants for lipid levels after genome-wide association analysis. *Curr. Opin. Lipidol.* **23,** 98–103 (2012).
4. Pardo, L. M., MacKay, I., Oostra, B., van Duijn, C. M. & Aulchenko, Y. S. The effect of genetic drift in a young genetically isolated population. *Ann. Hum. Genet.* **69,** 288–295 (2005).
5. Genome of the Netherlands Consortium. Whole-genome sequence variation, population structure and demographic history of the Dutch population. *Nat. Genet.* **46,** 818–825 (2014).
6. Boomsma, D. I. *et al.* The Genome of the Netherlands: design, and project goals. *Eur. J. Hum. Genet.* **22,** 221–227 (2014).
7. Jonsson, T. *et al.* A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. *Nature* **488,** 96–99 (2012).
8. Holm, H. *et al.* A rare variant in MYH6 is associated with high risk of sick sinus syndrome. *Nat. Genet.* **43,** 316–320 (2011).
9. Styrkarsdottir, U. *et al.* Nonsense mutation in the LGR4 gene is associated with several human diseases and other traits. *Nature* **497,** 517–520 (2013).
10. Deelen, P. *et al.* Improved imputation quality of low-frequency and rare variants in European samples using the 'Genome of The Netherlands'. *Eur. J. Hum. Genet.* **22,** 1321–1326 (2014).
11. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).
12. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat. Methods* **7,** 248–249 (2010).
13. Schwarz, J. M., Rödelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7,** 575–576 (2010).
14. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19,** 1553–1561 (2009).
15. Scott, L. J. *et al.* A genome-wide association study of type 2 diabetes in finns detects multiple susceptibility variants. *Science* **316,** 1341–1345 (2007).
16. Warnick, G. R., Knopp, R. H., Fitzpatrick, V. & Branson, L. Estimating low-density lipoprotein cholesterol by the Friedewald equation is adequate for classifying patients on the basis of nationally recommended cutpoints. *Clin. Chem.* **36,** 15–19 (1990).
17. Aulchenko, Y. S., Ripke, S., Isaacs, A. & van Duijn, C. M. GenABEL: an R library for genome-wide association analysis. *Bioinformatics* **23,** 1294–1296 (2007).
18. Svishcheva, G. R., Axenovich, T. I., Belonogova, N. M., van Duijn, C. M. & Aulchenko, Y. S. Rapid variance components-based method for whole-genome association analysis. *Nat. Genet.* **44,** 1166–1170 (2012).
19. Howie, B. N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5,** e1000529 (2009).
20. Howie, B. *et al.* Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44,** 955–959 (2012).
21. Willer, C. J., Li, Y., Gonçalo, R. & Abecasis, G. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26,** 2190–2191 (2010).
22. Browning, S. R. & Browning, B. L. Haplotype phasing: existing methods and new developments. *Nat. Rev. Genet.* **12,** 703–714 (2011).

Acknowledgements

We especially thank all volunteers who participated in our study. This study made use of data generated by the 'Genome of the Netherlands' project, which is funded by the Netherlands Organization for Scientific Research (grant no. 184021007). The data were made available as a Rainbow Project of BBMRI-NL. Samples were contributed by LifeLines (http://lifelines.nl/lifelines-research/general), the Leiden Longevity Study (http://www.healthy-ageing.nl; http://www.langleven.net), the Netherlands Twin Registry (NTR: http://www.tweelingenregister.org), the Rotterdam studies (http://www.erasmus-epidemiology.nl/rotterdamstudy) and the Genetic Research in Isolated Populations programme (http://www.epib.nl/research/geneticepi/research.html#gip). The sequencing was carried out in collaboration with the Beijing Institute for Genomics (BGI).

We would like to thank all the members of the CHARGE Lipids Working Group for their contribution in this project (a full list of consortium members is provided in Supplementary Note 1).

Cardiovascular Health Study: This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, HHSN268200960009C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL105756 and HL103612 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the National Institute on Aging (NIA). A full list of CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm.

The CROATIA cohorts would like to acknowledge the invaluable contributions of the recruitment teams in Vis, Korcula and Split (including those from the Institute of Anthropological Research in Zagreb and the Croatian Centre for Global Health at the University of Split), the administrative teams in Croatia and Edinburgh and the people of Vis, Korcula and Split. SNP genotyping was performed at the Wellcome Trust Clinical Research Facility in Edinburgh for CROATIA-Vis, by Helmholtz Zentrum München, GmbH, Neuherberg, Germany for CROATIA-Korcula and by AROS Applied Biotechnology, Aarhus, Denmark for CROATIA-Split. They would also like to thank Jared O'Connell for performing the pre-phasing for all cohorts before imputation.

The ERF study as a part of EuroSPAN (European Special Populations Research Network) was supported by European Commission FP-6 STRP grant number 018947 (LSHG-CT-2006-01947) and also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013)/grant agreement HEALTH-F4-2007-201413 by the European Commission under the programme 'Quality of Life and Management of the Living Resources' of 5th Framework Programme (no. QLG2-CT-2002-01254). High-throughput analysis of the ERF data was supported by joint grant from the Netherlands Organisation for Scientific Research and the Russian Foundation for Basic Research (NWO-RFBR 047.017.043). This research was financially supported by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007). Statistical analyses for the ERF study were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003 PI: Posthuma) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam. We are grateful to all study participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, J. Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection.

The FamHS is funded by a NHLBI grant 5R01HL08770003, and NIDDK grants 5R01DK06833603 and 5R01DK07568102.

The Framingham Heart Study SHARe Project for GWAS scan was supported by the NHLBI Framingham Heart Study (Contract No. N01-HC-25195) and its contract with Affymetrix Inc for genotyping services (Contract No. N02-HL-6-4278). DNA isolation and biochemistry were partly supported by NHLBI HL-54776. A portion of this research utilized the Linux Cluster for Genetic Analysis (LinGA-II) funded by the Robert Dawson Evans Endowment of the Department of Medicine at the Boston University School of Medicine and Boston Medical Center. We are grateful to Han Chen for conducting the 1000G imputation.

The Family Heart Study was supported by the by grants R01-HL-087700 and R01-HL-088215 from the National Heart, Lung, and Blood Institute (NHLBI).

Generation Scotland is a collaboration between the University Medical Schools and National Health Service in Aberdeen, Dundee, Edinburgh and Glasgow (UK). We would like to acknowledge the invaluable contributions of the families who took part in the Generation Scotland: Scottish Family Health Study, the general practitioners and Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes academic researchers, IT staff, laboratory technicians, statisticians and research managers. SNP genotyping was performed at the Wellcome Trust Clinical Research Facility in Edinburgh. GS:SFHS is funded by the Scottish Executive Health Department, Chief Scientist Office, grant number CZD/16/6. SNP genotyping was funded by the Medical Research Council, United Kingdom.

We wish to acknowledge the services of the LifeLines Cohort Study, the contributing research centres delivering data to LifeLines and all the study participants.

MESA Whites and the MESA SHARe project are conducted and supported by contracts N01-HC-95159 through N01-HC-95169 and RR-024156 from the NHLBI. Funding for MESA SHARe genotyping was provided by NHLBI Contract N02.HL.6.4278. MESA Family is conducted and supported in collaboration with MESA investigators; support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071252, R01HL071258 and R01HL071259. We thank the participants of the MESA study, the Coordinating Center, MESA investigators and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org.

Netherland Twin Register (NTR) and Netherlands Study of Depression and Anxiety (NESDA): Funding was obtained from the Netherlands Organization for Scientific Research (NWO) and MagW/ZonMW grants Middelgroot-911-09-032, Spinozapremie 56-464-14192, Geestkracht programme of the Netherlands Organization for Health Research and Development (Zon-MW, grant number 10-000-1002), Center for Medical Systems Biology (CSMB, NWO Genomics), NBIC/BioAssist/RK(2008.024), Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL, 184.021.007), VU University's Institute for Health and Care Research (EMGO+) and Neuroscience Campus Amsterdam (NCA); the European Science Foundation (ESF, EU/QLRT-2001-01254), the European Community's Seventh Framework Program (FP7/2007-2013), ENGAGE (HEALTH-F4-2007-201413); the European Science Council (ERC Advanced, 230374); and the European Research Council (ERC-284167). Part of the genotyping and analyses were funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health, Rutgers University Cell and DNA Repository (NIMH U24 MH068457-06), the Avera Institute, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH R01 HD042157-01A1, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995).

PREVEND genetics is supported by the Dutch Kidney Foundation (Grant E033), the EU project grant GENECURE (FP-6 LSHM CT 2006 037697), the National Institutes of Health (grant 2R01LM010098), The Netherlands Organisation for Health Research and Development (NWO-Groot grant 175.010.2007.006, NWO VENI grant 916.761.70, ZonMw grant 90.700.441) and the Dutch Inter University Cardiology Institute Netherlands (ICIN).

© 2015 Macmillan Publishers Limited. All rights reserved.

The PROSPER study was supported by an investigator-initiated grant obtained from Bristol-Myers Squibb. J.W.J is an Established Clinical Investigator of the Netherlands Heart Foundation (grant 2001 D 032). Genotyping was supported by the seventh framework programme of the European commission (grant 223004) and by the Netherlands Genomics Initiative (Netherlands Consortium for Healthy Aging grant 050-060-810).

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII) and the Municipality of Rotterdam. We are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

The generation and management of GWAS genotype data for the Rotterdam Study is supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012). This study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) project no. 050-060-810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database.

Author contributions

E.M.v.L. organized the study and designed the study with substantial input of L.C.K., A.I., P.I.W.d.B. and C.M.v.D. E.M.v.L. drafted the manuscript with substantial input of L.A.C., A.Me, B.M.P., C.W., G.M.P., J.F.W., J.E.H., L.C.F., L.C.K., J.D., P.E.S., D.I.B., J.E.H., H.M., P.M.K., P.I.W.d.B., S.L.P., S.T., C.M.v.D. and G.-J.B.v.O. All authors had the opportunity to comment on the manuscript. Data collection, GWAS and statistical analysis were performed by T.M.B., J.A.B., J.C.B., B.M.P. (CHS); J.E.H., C.H., O.P., V.V., I.R., A.F.W. (CROATIA); E.M.v.L., B.A.O., C.M.v.D. (ERF); C.C.W., L.A.C. (FHS), M.F.F., I.B.B. (FamHS); J.E.H., H.T., L.J.H., D.J.P. (Generation Scotland); G.M.P., Q.D., L.A.L. (JHS); A.Ma., J.I.R., J.C.M., S.S.R. (MESA); A.K., P.D., F.v.D., M.A.S., C.W. (Lifelines); J.D., M.B., A.J.M.C., H.-W.U., P.E.S. (LLS); H.M., G.W., E.J.d.G., Y.M., B.W.J.H.P., J.-J.H., D.I.B. (NTR-NESDA); N.V., I.M.L., P.v.H. (PREVEND); S.T., I.P., N.S., C.J.P., B.M.B., P.M.K., D.J.S., J.W.J. (PROSPER); P.K.J., H.C., J.F.W. (ORCADES); E.M.v.L., C.M.-G., F.R., A.H., O.H.F., E.J.S., A.G.U., C.M.v.D. (Rotterdam Study). D.J.v.E. recruited cohorts. Creation of the GoNL reference panel was carried out by L.C.F., A.Me., S.L.P. and P.D. Design of the GoNL project was made by C.W., M.A.S., C.M.v.D., D.I.B., P.E.S., G.-J.B.O., P.I.W.d.B. E.M.v.L. performed the meta-analysis. Biological association of loci and bioinformatics were carried out by E.M.v.L. and C.M.v.D.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: B.M.P. served on the DSMB of a clinical trial funded by the manufacturer (Zoll Lifecor) and on the Yale Open Data Access Project funded by Johnson & Johnson. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: van Leeuwen, E. M. *et al.* Genome of the Netherlands population-specific imputations identify a ABCA6 variant associated with cholesterol levels. *Nat. Commun.* 6:6065 doi: 10.1038/ncomms7065 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

The Genome of the Netherlands Consortium

Pieter B.T. Neerincx[5,37], Clara C. Elbers[18], Pier Francesco Palamara[38], Itsik Pe'er[38,39], Abdel Abdellaoui[4], Wigard P. Kloosterman[18], Mannis van Oven[40], Martijn Vermaat[41], Mingkun Li[42], Jeroen F.J. Laros[41], Mark Stoneking[42], Peter de Knijff[43], Manfred Kayser[40], Jan H. Veldink[44], Leonard H. van den Berg[44], Heorhiy Byelas[5,37], Johan T. den Dunnen[41], Martijn Dijkstra[5,37], Najaf Amin[1], K. Joeri van der Velde[5,37], Jessica van Setten[18], Mathijs Kattenberg[4], Barbera D.C. van Schaik[45], Jan Bot[46], Isaäc J. Nijman[18], Hailiang Mei[9], Vyacheslav Koval[33], Kai Ye[2,47], Eric-Wubbo Lameijer[2], Matthijs H. Moed[2], Jayne Y. Hehir-Kwa[48], Robert E. Handsaker[49,50], Shamil R. Sunyaev[49,51], Mashaal Sohail[49,51], Fereydoun Hormozdiari[52], Tobias Marschall[53], Alexander Schönhuth[53], Victor Guryev[54], H. Eka D. Suchiman[2], Bruce H. Wolffenbuttel[55], Mathieu Platteel[37], Steven J. Pitts[56], Shobha Potluri[56], David R. Cox[56,‡], Qibin Li[57], Yingrui Li[57], Yuanping Du[57], Ruoyan Chen[57], Hongzhi Cao[57], Ning Li[58], Sujie Cao[58], Jun Wang[57,59,60], Jasper A. Bovenberg[61]

[38]Department of Computer Science, Columbia University, New York, NY 10027-7003, USA. [39]Department of Systems Biology, Columbia University, New York, NY 10032, USA. [40]Department of Forensic Molecular Biology, Erasmus Medical Center, Rotterdam 3000 CA, The Netherlands. [41]Leiden Genome Technology Center, Department of Human Genetics, Leiden University Medical Center, Leiden 2300 RC, The Netherlands. [42]Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig, 4103, Germany. [43]Forensic Laboratory for DNA Research, Department of Human Genetics, Leiden University Medical Center, Leiden, 2300 RC, The Netherlands. [44]Department of Neurology, University Medical Center Utrecht, Utrecht, 3584 CG, The Netherlands. [45]Bioinformatics Laboratory, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Medical Center, Amsterdam 1090GE, The Netherlands. [46]SURFsara, Science Park, Amsterdam 1098 XG, The Netherlands. [47]The Genome Institute, Washington University, St. Louis, MO 98101, USA. [48]Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen 6500 HB, The Netherlands. [49]Broad Institute of Harvard and MIT, Cambridge, MA 2142, USA. [50]Department of Genetics, Harvard Medical School, Boston, MA 2115, USA. [51]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 2115, USA. [52]Department of Genome Sciences, University of Washington, Seattle, WA 98101, USA. [53]Centrum voor Wiskunde en Informatica, Life Sciences Group, Amsterdam 1098 XG, The Netherlands. [54]European Research Institute for the Biology of Ageing, University Medical Center Groningen, University of Groningen, Groningen 9700 RB, The Netherlands. [55]Department of Endocrinology, University Medical Center Groningen, Groningen 9700 RB, The Netherlands. [56]Rinat-Pfizer Inc, South San Francisco, CA 10017, USA. [57]BGI-Shenzhen, Shenzhen 518083, China. [58]BGI-Europe, Copenhagen DK-1870, Denmark. [59]Department of Biology, University of Copenhagen, Copenhagen 2100, Denmark. [60]The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Copenhagen 2100, Denmark. [61]Legal Pathways Institute for Health and Bio Law, Aerdenhout, The Netherlands. ‡Deceased.

© 2015 Macmillan Publishers Limited. All rights reserved.