

A Polynesian-specific copy number variant encompassing the MICA gene associates with gout

Ke Wang^{1,2}, Murray Cadzow², Matt Bixley², Megan P. Leask³, Marilyn E. Merriman², Qiangzhen Yang¹, Zhiqiang Li^{1,4}, Riku Takei^{2,3}, Amanda Phipps-Green², Tanya J. Major², Ruth Topless², Nicola Dalbeth⁵, Frances King⁶, Rinki Murphy⁵, Lisa K. Stamp⁷, Janak de Zoysa⁵, Zhuo Wang¹, Yongyong Shi^{1,4,*} and Tony R. Merriman^{2,3,*}

¹Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Jiao Tong University, Shanghai 200030, People's Republic of China

²Department of Biochemistry, University of Otago, Dunedin 9054, New Zealand

³Division of Clinical Immunology and Rheumatology, University of Alabama at Birmingham, Birmingham, AL 35233, USA

⁴Biomedical Sciences Institute of Qingdao University (Qingdao Branch of SJTU Bio-X Institutes), Qingdao University, Qingdao 266003, China

⁵Department of Medicine, University of Auckland, Auckland 1023, New Zealand

⁶Ngati Porou Hauora Charitable Trust, Te Puia Springs, New Zealand

⁷Department of Medicine, University of Otago, Christchurch 8013, New Zealand

*To whom correspondence should be addressed. Tel.: 205-427-5853; Email: Tony.merriman@otago.ac.nz and shiyongyong@gmail.com

Abstract

Gout is of particularly high prevalence in the Māori and Pacific (Polynesian) populations of Aotearoa New Zealand (NZ). Here, we investigated the contribution of common population-specific copy number variation (CNV) to gout in the Aotearoa NZ Polynesian population. Microarray-generated genome-wide genotype data from Aotearoa NZ Polynesian individuals with ($n = 1196$) and without ($n = 1249$) gout were analyzed. Comparator population groups were 552 individuals of European ancestry and 1962 of Han Chinese ancestry. Levels of circulating major histocompatibility complex (MHC) class I polypeptide-related sequence A (MICA) were measured by enzyme-linked immunosorbent assay. Fifty-four CNV regions (CNVRs) appearing in at least 10 individuals were detected, of which seven common ($>2\%$) CNVRs were specific to or amplified in Polynesian people. A burden test of these seven revealed associations of insertion/deletion with gout (odds ratio (OR) 95% confidence interval [CI] = 1.80 [1.01; 3.22], $P = 0.046$). Individually testing of the seven CNVRs for association with gout revealed nominal association of CNVR1 with gout in Western Polynesian (Chr6: 31.36–31.45 Mb, OR = 1.72 [1.03; 2.92], $P = 0.04$), CNVR6 in the meta-analyzed Polynesian sample sets (Chr1: 196.75–196.92 Mb, OR = 1.86 [1.16; 3.00], $P = 0.01$) and CNVR9 in Western Polynesian (Chr1: 189.35–189.54 Mb, OR = 2.75 [1.15; 7.13], $P = 0.03$). Analysis of European gout genetic association data demonstrated a signal of association at the CNVR1 locus that was an expression quantitative trait locus for MICA. The most common CNVR (CNVR1) includes deletion of the MICA gene, encoding an immunomodulatory protein. Expression of MICA was reduced in the serum of individuals with the deletion. In summary, we provide evidence for the association of CNVR1 containing MICA with gout in Polynesian people, implicating class I MHC-mediated antigen presentation in gout.

Introduction

Gout is a common complex metabolic and inflammatory disease caused by hyperuricemia and immune response to the deposition of monosodium urate (MSU) crystals in and around body tissues, particularly joints (1). The clinical features consist of intermittent gout flares, chronic gouty arthritis, tophi and increased prevalence of comorbidities that include hypertension, type 2 diabetes mellitus, chronic kidney disease, dyslipidemia and cardiovascular disease (1–5). The pathogenesis of gout requires progression through several phases: from hyperuricemia to the deposition of MSU crystals to subsequent immune response to these crystals (6). In some individuals, the crystals cause a toll-like receptor-mediated formation and activation of the NLRP3 inflammasome in

monocytes and production of interleukin-1 β that drives the gout flare.

Genome-wide association studies (GWAS) that focus on single-nucleotide variants (SNVs) have provided considerable insight into the molecular pathogenesis of hyperuricemia (7–9), although GWAS in gout have been more limited (10). However, common SNVs explain only a relatively small fraction of the heritability of gout, with genome-wide heritability estimates of around 0.3 (11). Structural variants are defined as genomic rearrangements affecting >50 bp of sequence, and they impact a greater proportion of the genome than SNVs. Therefore, they have a greater likelihood of having an impact on molecular function and phenotype. Structural variants are firmly implicated in common disease

Received: January 16, 2022. Revised: April 1, 2022. Accepted: April 19, 2022

© The Author(s) 2022. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

including schizophrenia (12), Alzheimer's disease (13), intellectual disabilities (14), autism (15,16), Crohn's disease (17,18) and rheumatic diseases (19–21). In gout and hyperuricemia, single small studies have reported that copy number variations (CNVs) of *ABCF1*, *IL17REL* and *FCGR3A* are associated with the risk of gout in the Chinese population (22), and CNV upstream of *SLC2A9* associates with urate levels in Europeans (23), suggesting a role for genomic structural variation in the pathogenesis of gout.

The prevalence of gout among Māori (the indigenous people of Aotearoa New Zealand) and New Zealand Pacific adults is 8% and 14%, respectively, compared to 4% for non-Māori and non-Pacific people (24). While structural inequities contribute to increased prevalence and poorer outcomes (25), research on population-specific genetic variants associated with gout is important to provide insights into the disparities between different populations, especially in populations currently under-served with respect to participation in genetic studies. It is becoming clear that genetic variation that is specific to Pacific populations does exist and that such variation contributes to the pathogenesis of gout (26–28) and other metabolic diseases (29). Our primary hypothesis was that there are Polynesian-specific (or amplified) CNV regions (CNVRs) associated with gout and which may contribute to the increased prevalence of gout in Polynesian populations. Therefore, we aimed to identify Polynesian-specific CNVs contributing to gout using genome-wide genotype data in a total of 1196 people with and 1249 people without gout.

Results

Discovery and characteristics of the CNVs

Aotearoa NZ Polynesian and European participants that had CNV count ≤ 312 (85th percentile) were used for CNV identification and analysis. Among 4353 CNV calls detected after the quality control filtering, 2607 (1427 duplications and 1180 deletions) were found in people with gout and 1746 (899 duplications and 847 deletions) in people without gout. The distribution of CNVs along each of the chromosomes is shown in Figure 1A. The identified CNVs ranged in size from 5007 to 144 150 872 bp (144 Mb), with an average length of 385.3 kb (Fig. 2). There were three CNVs > 10 Mb that were present in two European individuals (Supplementary Material, Fig. S1)—duplications at Chr8: 2.09–146.25 Mb, Chr15: 31.53–43.38 Mb and Chr15: 45.57–102.40 Mb (the same individual had both Chr15 duplications), representing apparent trisomy of chromosome 8 and chromosome 15, respectively. We did not further investigate these individuals as they did not present with significant health issues and the observed data may be a result of artefact and/or low-level somatic mosaicism. Supplementary Material, Table S1 lists all 54 CNVRs that were present in ≥ 10 individuals, including details of each copy number type that PennCNV reported.

Representative examples of LRR and BAF plots are shown in Supplementary Material, Figure S2 for the 20 most common CNVRs. Comparing length of all types of CNVs between gout and non-gout revealed no significant differences with respect to the average counts and length of CNV (Supplementary Material, Fig. S3).

Testing Polynesian-specific and -amplified CNVR for association with gout

Specifically, we examined the seven CNVRs of $> 2\%$ prevalence in Polynesian that were either specific to Polynesian [CNVR1 (Chr6: 31.36–31.45 Mb), CNVR2 (Chr6: 29.92–29.94 Mb), CNVR6 (Chr1: 196.75–196.92 Mb), CNVR8 (Chr13: 111.17–111.33 Mb), CNVR9 (Chr1: 189.35–189.54 Mb), CNVR15 (Chr14: 62.63–62.77 Mb)] or were amplified in Polynesian [$< 1\%$ in European and Chinese; CNVR7 (Chr2: 110.86–110.98 Mb)]. A test for association with gout was performed by comparing individuals with a CNV in any of the seven reported regions (pooled) to those who were diploid for all seven CNVRs. These seven CNVRs were associated with gout in Polynesian using Model 2 [odds ratio (OR) = 1.80, $P = 0.046$] (Table 1), with the strongest association in the Western Polynesian sample set (OR = 2.43, $P = 2.1E-06$).

The seven CNVRs were then individually tested for association with gout in the Polynesian sample set. Three provided evidence for association with gout at a nominal level of significance in the fully adjusted model: CNVR1 (OR = 1.72, $P = 0.04$, Western Polynesian), CNVR6 (OR = 1.86, $P = 0.01$, meta-analyzed Polynesian) and CNVR9 (OR = 2.75, $P = 0.03$, Western Polynesian) (Table 2). Linear regression testing for association with serum urate level in controls provided no evidence for association of any of the three CNVRs, even in the unadjusted models (Supplementary Material, Table S2). Association with gout but not serum urate level is consistent with a possible role of the deletions in the progression from hyperuricemia to gout.

Further validation of CNVR1, CNVR6 and CNVR9

Examination of the CNVR1, CNVR6 and CNVR9 genomic regions in the Database of Genomic Variants (DGV) revealed that these regions are established CNV sites with multiple reported losses and gains (Supplementary Material, Table S3). There are 6, 9 and 37 DGV Gold Standard Variants that overlap with CNVR1, CNVR9 and CNVR6, respectively (Supplementary Material, Table S3).

A total of 91 Polynesian participants with both microarray and 30X whole genome sequencing (WGS) data were used to validate CNVR1, CNVR6 and CNVR9. Read depth of these individuals was plotted from WGS data. The average consistency of the copy number calls in microarray and WGS data was 94% (Supplementary Material, Table S4, 96%, 88% and 97%, respectively), indicating that our CNV calls were reliable. Representative examples of WGS read-depth plots are shown in Supplementary Material, Figure S4.

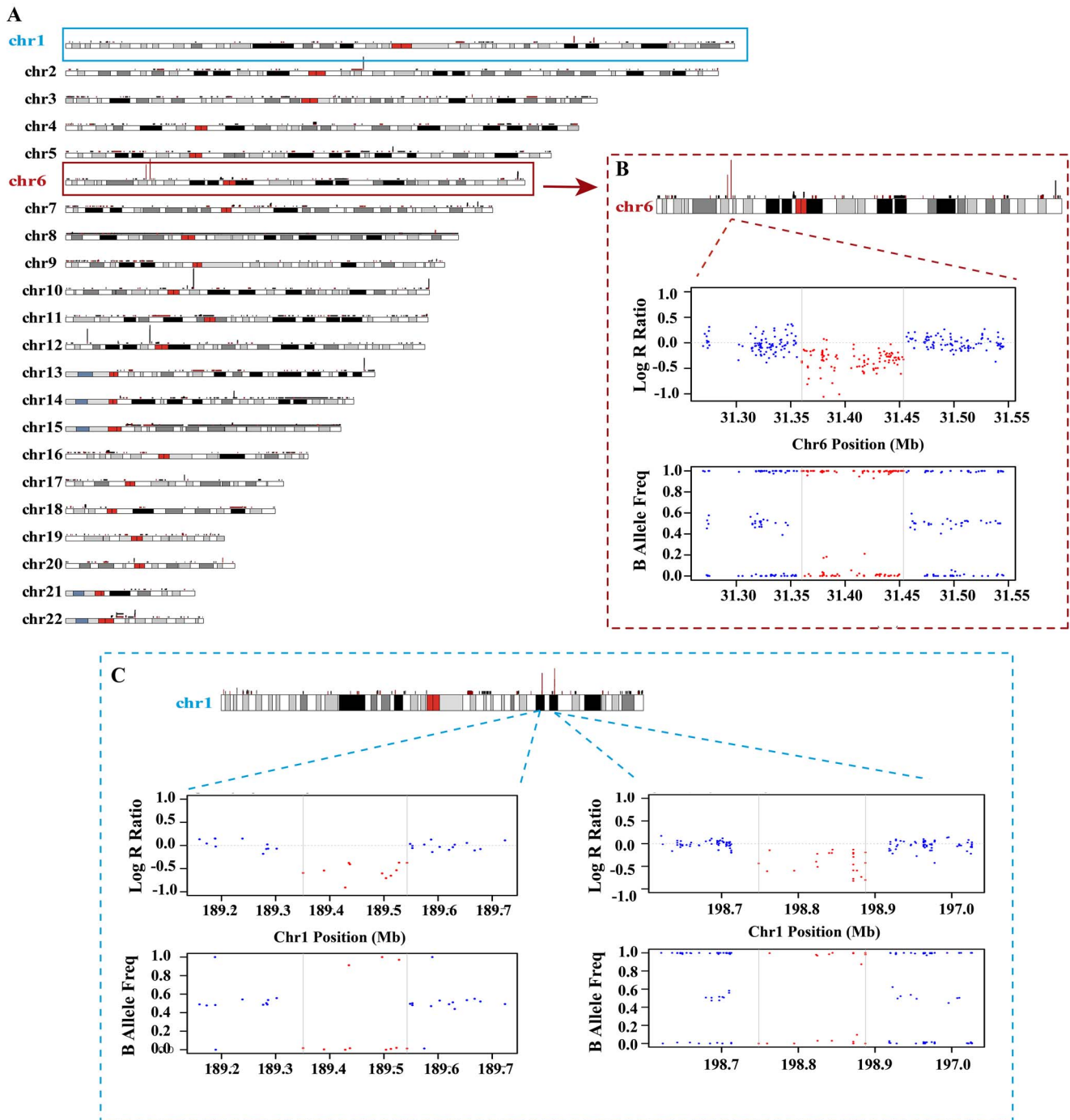


Figure 1. (A) Genomic distribution of CNVs. Red lines show copy number loss and black lines show copy number gain. (B, C) Example of deletions in CNV regions.

Association analysis of loci encompassing CNVR1, CNVR6 and CNVR9 with gout in the UK Biobank

We used gout GWAS SNV data in a European dataset to test the CNVR1, CNVR6 and CNVR9 loci for association with gout (UK Biobank; 7131 cases and 325 239 controls). [Supplementary Material, Figure S5](#) presents locus zoom plots for each locus. There was a signal of association in the CNVR1 region, but not CNVR6 or CNVR9. (The lead SNV at CNVR6 with $P < 1 \times 10^{-4}$ (rs143765601) was

uncommon with no gout-associated SNVs in linkage disequilibrium and therefore was considered unreliable.) At CNVR1, there appeared to be two signals of genetic association—one signal was marked by top associated variant rs3016018 and the second marked by rs9265955. We tested each for association with expression of genes within CNVR1 using the Gene and Tissue Expression (GTEx) resource ([Supplementary Material, Fig. S6](#)). Rs3016018 is associated with the expression of multiple protein-coding genes outside CNVR1 (e.g. HLA-S,

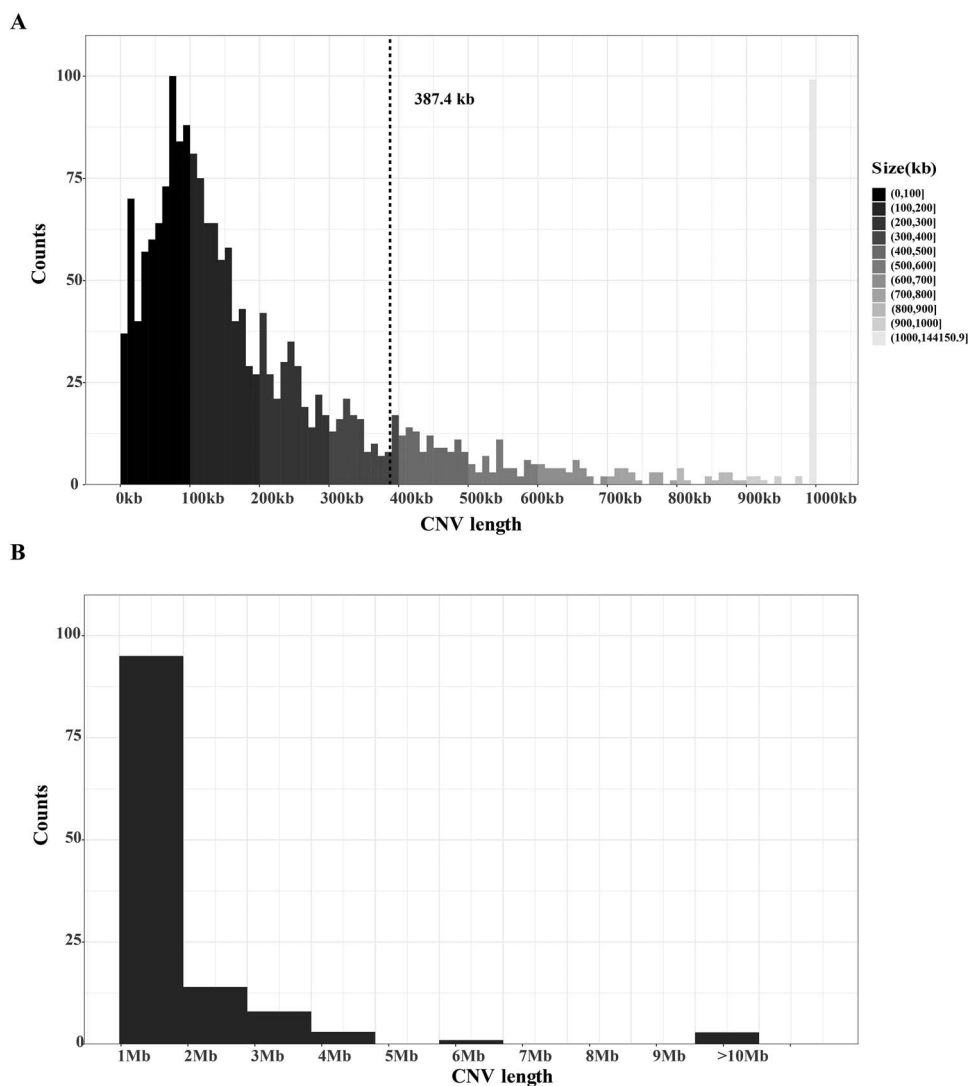


Figure 2. The size distribution of CNVs. **(A)** Size (in kb) compared to the number of CNVs. The CNVs were grouped into 10 kb windows. For the purposes of visualization, large CNVs (>1 Mb) were classified to 1 Mb. The dotted line represents the mean CNV size. **(B)** Distribution of large CNVs (>1 Mb).

Table 1. Association with gout of the burden of seven common specific or amplified CNVRs in the Aotearoa NZ Polynesian sample sets

Ethnicity	Gout		Control		Unadjusted		Adjusted (Model 1)		Adjusted (Model 2)	
	With CNV	Without CNV	With CNV	Without CNV	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P
EP	255 (38.2%)	413 (61.8%)	221 (34.7%)	416 (65.3%)	1.39 (1.12–1.73)	0.003	1.3 (1.01–1.67)	0.039	1.21 (0.94–1.58)	0.14
WP	224 (57.3%)	167 (42.7%)	123 (46.1%)	144 (53.9%)	2.44 (1.83–3.25)	1.3E-09	2.77 (1.96–3.93)	1.0E-08	2.43 (1.69–3.52)	2.1E-06
EPWP	23 (56.1%)	18 (43.9%)	22 (40%)	33 (60%)	2.34 (1.08–5.16)	0.032	3.18 (1.33–8.06)	0.011	2.48 (0.79–8.13)	0.1231
OP-meta	502 (45.6%)	598 (54.4%)	366 (38.2%)	593 (61.8%)	1.91 (2.22–3.01)	0.005 P_{het}	2.12 (1.13–3.95)	0.019 P_{het}	1.80 (1.01–3.22)	0.046 P_{het}
						< 0.01, $I^2 = 80%$		< 0.01, $I^2 = 85%$		< 0.01, $I^2 = 80%$

EP, Eastern Polynesian; EPWP, Mixed Eastern–Western Polynesian; WP, Western Polynesian; OP-meta, WP, EP and EPWP were combined by meta-analysis with random effects; CI, confidence intervals; OR, odds ratio; Model 1, adjusted by age and sex; Model 2, Model 1 plus adjustment by batches and PCs 1–10. P-values < 0.05 are shown in bold.

PSORS1C3, NOTCHC4, C4B); however, rs9265955, which also associated with the expression of multiple genes outside of CNVR1, is also associated with the expression of MHC class I polypeptide-related sequence A (MICA), the single protein-coding gene within CNVR1 ($P = 3.6 \times 10^{-5}$). The gout risk (C) allele of rs9265955 increased the expression of MICA. This evidence from the UK Biobank supports the MICA locus as involved in the etiology of gout.

Physical connectedness within CNVR1

We used both GeneHancer (30) (visualized in the UCSC genome browser), which infers genomic connectedness based on multiple genomic data sources, and HiC physical connectedness data (31) to demonstrate genomic interaction between MICA and long non-coding regulatory RNAs HCP5 and LINC01149 that are both contained within CNVR1 (Fig. 3). This is consistent with a role of these long non-coding RNAs

Table 2. Logistic regression association analysis with gout

CNVR	Ethnicity	Cases				Controls				Unadjusted			Adjusted (Model 1)			Adjusted (Model 2)			Gene list
		CN=0	CN=1	CN=3	CN=0	CN=1	CN=3	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	
CNVR1	WP	3 (0.8%)	60 (15.4%)	1 (0.3%)	1 (0.4%)	38 (14.2%)	0 (0%)	1.60 (1.05-2.45)	0.03	1.84 (1.12-3.07)	0.02	1.72 (1.03-2.92)	0.04	$P_{het} = 0.13, I^2 = 52%$	HCG26, HCPFS,				
	EP	11 (1.7%)	104 (15.6%)	6 (0.9%)	4 (0.6%)	88 (13.8%)	9 (1.4%)	1.36 (1.02-1.81)	0.04	1.16 (0.84-1.61)	0.37	1.10 (0.79-1.54)	0.57	$P_{het1} = 0.05, I^2 = 68%$	MICA				
	EPWP	0 (0%)	9 (22.0%)	0 (0%)	0 (0%)	3 (5.5%)	0 (0%)	5.66 (1.57-26.74)	0.01	6.06 (1.53-31.69)	0.02	6.33 (1.24-43.16)	0.04	$P_{het2} = 0.07, I^2 = 62%$					
CNVR2	OP-meta	14 (1.3%)	173 (15.7%)	7 (0.6%)	5 (0.5%)	129 (13.5%)	9 (0.9%)	1.63 (1.07-2.48)	0.02	1.73 (0.94-3.17)	0.08	1.56 (0.87-2.78)	0.14						
	WP	12 (3.1%)	62 (15.9%)	0 (0%)	8 (3.0%)	40 (15.0%)	0 (0%)	1.50 (1.02-2.23)	0.04	1.45 (0.91-2.34)	0.12	1.24 (0.76-2.03)	0.40	$P_{het} = 0.20, I^2 = 38%$	HLA-G, HLA-				
	EP	3 (0.5%)	56 (8.4%)	6 (0.9%)	3 (0.5%)	65 (10.2%)	6 (0.9%)	0.95 (0.67-1.35)	0.77	0.98 (0.65-1.48)	0.93	0.91 (0.6-1.38)	0.64	$P_{het1} = 0.47, I^2 = 0%$	H, HLA-j				
CNVR6	EPWP	0 (0%)	6 (14.6%)	0 (0%)	2 (3.6%)	9 (16.4%)	0 (0%)	0.83 (0.27-2.38)	0.74	1.1 (0.31-3.62)	0.88	0.52 (0.11-2.25)	0.39						
	OP-meta	15 (1.4%)	124 (11.3%)	6 (0.6%)	13 (1.4%)	114 (11.9%)	6 (0.6%)	1.14 (0.88-1.47)	0.31	1.16 (0.86-1.56)	0.34	1.01 (0.74-1.37)	0.96						
	WP	18 (4.6%)	24 (6.1%)	0 (0%)	11 (4.1%)	12 (4.5%)	0 (0%)	1.75 (1.04-3.01)	0.04	1.71 (0.93-3.23)	0.09	1.71 (0.91-3.31)	0.11	$P_{het} = 0.45, I^2 = 0%$	CFHR1, CFHR2,				
CNVR7	EP	13 (2.0%)	12 (1.8%)	0 (0%)	4 (0.6%)	11 (1.7%)	0 (0%)	1.84 (0.97-3.6)	0.07	1.92 (0.93-4.1)	0.08	1.92 (0.91-4.15)	0.09	$P_{het1} = 0.50, I^2 = 0%$	CFHR3, CFHR4				
	EPWP	3 (7.3%)	3 (7.3%)	0 (0%)	0 (0%)	2 (3.6%)	0 (0%)	5.29 (1.15-37.34)	0.05	5.34 (1.02-41.76)	0.07	3.21 (0.51-27.77)	0.23	$P_{het2} = 0.83, I^2 = 0%$					
	OP-meta	34 (3.1%)	39 (3.6%)	0 (0%)	15 (1.6%)	25 (2.6%)	0 (0%)	1.90 (1.28-2.83)	2.0E-03	1.93 (1.22-3.05)	5.0E-03	1.86 (1.16-3.00)	0.01						
CNVR8	WP	0 (0%)	0 (0%)	44 (11.3%)	0 (0%)	0 (0%)	25 (9.4%)	1.69 (1.02-2.85)	0.05	2.19 (1.19-4.14)	0.01	1.76 (0.95-3.37)	0.08	$P_{het} = 0.67, I^2 = 0%$	LINC00116,				
	EP	0 (0%)	3 (0.4%)	3 (7.3%)	0 (0%)	1 (0.2%)	8 (1.3%)	1.83 (0.81-4.38)	0.16	1.45 (0.58-3.92)	0.44	1.20 (0.47-3.31)	0.71	$P_{het1} = 0.62, I^2 = 0%$	MALL, NPHP1				
	EPWP	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	1 (1.8%)	4.98 (0.61-102.41)	0.17	4.67 (0.47-112.51)	0.23	2.00 (0.17-53.25)	0.61	$P_{het2} = 0.80, I^2 = 0%$					
CNVR9	OP-meta	0 (0%)	3 (0.3%)	59 (5.4%)	0 (0%)	1 (0.1%)	34 (3.5%)	1.79 (1.17-2.75)	8.0E-03	2.01 (1.21-3.33)	7.0E-03	1.58 (0.94-2.66)	0.083						
	EP	0 (0%)	0 (0%)	35 (9.0%)	0 (0%)	0 (0%)	19 (7.1%)	1.75 (1-3.17)	0.06	2.27 (1.16-4.57)	0.02	1.84 (0.94-3.71)	0.08	$P_{het} = 0.29, I^2 = 18%$	CARKD, CARS2,				
	EPWP	0 (0%)	2 (0.3%)	17 (2.5%)	0 (0%)	0 (0%)	19 (3.0%)	1.09 (0.57-2.08)	0.80	0.98 (0.46-2.1)	0.97	0.86 (0.4-1.89)	0.71	$P_{het1} = 0.18, I^2 = 42%$	RAB20				
CNVR15	OP-meta	0 (0%)	2 (0.2%)	53 (4.8%)	0 (0%)	0 (0%)	43 (4.5%)	1.35 (0.89-2.06)	0.16	1.48 (0.90-2.42)	0.12	1.25 (0.75-2.07)	0.39	$P_{het2} = 0.18, I^2 = 42%$					
	WP	0 (0%)	28 (7.2%)	0 (0%)	1 (0.4%)	9 (3.4%)	0 (0%)	1.28 (0.73-2.24)	9.0E-03	3.00 (1.29-7.55)	0.01	2.75 (1.15-7.13)	0.03	$P_{het} = 0.25, I^2 = 27%$	None				
	EP	2 (0.3%)	25 (3.7%)	1 (0.1%)	0 (0%)	24 (3.8%)	0 (0%)	2.68 (1.33-5.86)	0.39	1.25 (0.67-2.36)	0.48	1.24 (0.65-2.37)	0.51	$P_{het1} = 0.22, I^2 = 34%$					
CNVR15	EPWP	0 (0%)	2 (4.9%)	0 (0%)	0 (0%)	1 (1.8%)	0 (0%)	3.24 (0.3-70.95)	0.34	4.40 (0.36-106.27)	0.26	2.08 (0.09-84.61)	0.67	$P_{het2} = 0.37, I^2 = 0%$					
	OP-meta	2 (0.2%)	55 (5%)	1 (0.1%)	0 (0%)	34 (3.5%)	0 (0%)	1.71 (1.11-2.64)	0.02	1.74 (1.06-2.88)	0.03	1.63 (0.97-2.73)	0.07						
	WP	0 (0%)	0 (0%)	9 (2.3%)	0 (0%)	0 (0%)	6 (2.3%)	1.38 (0.49-4.16)	0.54	0.84 (0.27-2.8)	0.77	0.78 (0.24-2.7)	0.69	$P_{het} = 0.60, I^2 = 0%$	None				
CNVR15	EP	0 (0%)	0 (0%)	21 (3.1%)	0 (0%)	0 (0%)	12 (1.9%)	1.93 (0.96-4.07)	0.07	2.19 (0.98-5.1)	0.06	2.15 (0.95-5.09)	0.07	$P_{het1} = 0.18, I^2 = 44%$					
	EPWP	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (5.5%)	N.A.	-	N.A.	-	N.A.	-	$P_{het2} = 0.17, I^2 = 46%$					
	OP-meta	0 (0%)	0 (0%)	30 (2.7%)	0 (0%)	0 (0%)	21 (2.2%)	1.73 (0.96-3.13)	0.07	1.59 (0.82-3.09)	0.18	1.54 (0.78-3.05)	0.22						

P-values < 0.05 are shown in bold; CN, copy number (only reported genotype were shown); EP, Eastern Polynesian; EPWP, Mixed Eastern-Western Polynesian; WP, Western Polynesian; OP-meta, WP, EP and EPWP were combined by meta-analysis with random effects for CNVR1 and fixed effects for the other CNVRs; CI, confidence intervals; OR, odds ratio; N.A., not available; %, the frequency of case/control; Model 1, adjusted by age and sex; Model 2, Model 1 plus adjustment by batches and PCs 1-10; P_{het} , P_{het1} , P_{het2} , heterogeneity of P for unadjusted, Model 1 and Model 2, respectively, in the meta-analysis.

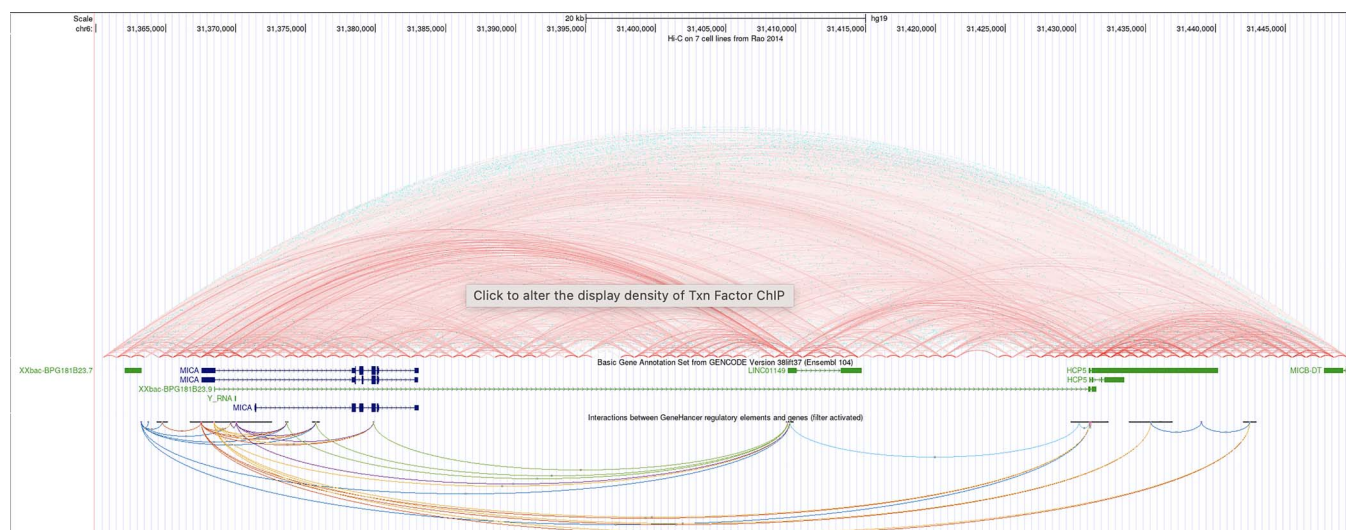


Figure 3. Genomic connectedness at CNVR1. Schematic of UCSC genome tracks at the CNVR1 deletion encompassing the *MICA*::*HCP5* locus. **(top)** Hi-C genome track for GM12878 cells at CNVR1. Hi-C fragments connect *MICA* and *HCP5* plus *LINC01149*. **(bottom)** GeneHancer connections for *HCP5* (yellow) and *MICA* (red) map to predicted enhancers within CNVR1.

(lncRNAs) in regulating expression of *MICA* within CNVR1.

Correlation of CNVR1 genotype with serum *MICA* levels

CNVR1 encompassed one protein-coding gene, *MICA*. It is a cell-surface receptor for natural killer (NK) T-cells and has a soluble form. Levels of soluble *MICA* were measured by enzyme-linked immunosorbent assay (ELISA) in 13 individuals with CN=0, 49 with CN=1 and 43 with CN=2, 95% of whom were men. People without the deletion had a statistically significantly increased mean level of circulating *MICA* protein (Fig 4; $P=0.03$) with those of CN=0 genotype exhibiting the lowest level (55.7 pg/ml), those with CN=1 an intermediate level (172.9 pg/ml) and those with CN=2 the highest level (219.9 pg/ml).

Discussion

We assessed the role of common large (>5 kb) CNV in gout in the Aotearoa NZ Māori and Pacific (Polynesian) populations and identified three common Polynesian-specific CNV regions [Chr6: 31.36–31.45 Mb (CNVR1); Chr1: 196.75–196.92 Mb (CNVR6); Chr1: 189.35–189.54 Mb (CNVR9)] with evidence for association with gout. Additional genetic support for *MICA* as a candidate causal gene in CNVR1 was obtained from European gout genetic association data and from the GTEx genetic control of gene expression resource. However equivalent data were not obtained for either CNVR6 or CNVR9.

Given the limited power of the Polynesian gout sample set, and the unavailability of a replication cohort, our strategy focused on reducing the statistical impact of multiple testing. Hence, we used an approach that would allow us to identify population-specific CNV rather than a genome-wide approach. The European and Han Chinese sample sets were included to serve this purpose. Future studies can focus on the impact of a wider

number of common CNV on gout in these populations, leveraging considerably larger sample sizes.

The locus with the strongest evidence for a role in gout was CNVR1, which is dominated by deletion. Genetic (SNV) associations identified by GWAS for immune-mediated disease have been reported at this locus, including HIV infection (32) and autoimmune diseases such as Grave's disease (33) and systemic lupus erythematosus (34). CNVs that overlap CNVR1 have been reported to be associated with nasopharyngeal carcinoma predisposition (35,36), type 1 diabetes (37) and schizophrenia or bipolar disorder (38) and, now, gout. Recurrent hemizygous deletion at the locus associates with idiopathic pulmonary hemosiderosis, a rare disease without an identified trigger (39). We also provided additional evidence of gout risk in Europeans and gene expression data from GTEx for a role of *MICA* within CNVR1 in the etiology of gout. The CNV encompasses one protein-coding gene (*MICA*), which is a strong candidate causal gene. Levels of the *MICA* protein are reduced in people with the deletion (Fig 4). *MICA* encodes a genotoxic stress-induced protein and serves as a ligand for Natural Killer Group 2 member (NKG2D). NKG2D is an immune receptor on the surface of NK cells (40), which binds to immune cells expressing *MICA* and triggering lysis in these target cells. NK cells can also indirectly activate monocytes and neutrophils through direct interaction or cytokine secretion (41,42). In gout, reduced levels of *MICA* might result in accumulation of macrophages and an increased immune response to MSU crystals. Urate induces *MICA* expression via TAK1 promoting NK cell killing and immunosuppression (43–45). Reduction of urate in an animal model via inhibition of xanthine oxidoreductase activity or gene knockout ablates the genotoxic stress-induced expression of *MICA*. Additionally, genetic variants that associate with *MICA* serum levels also associate with inflammatory

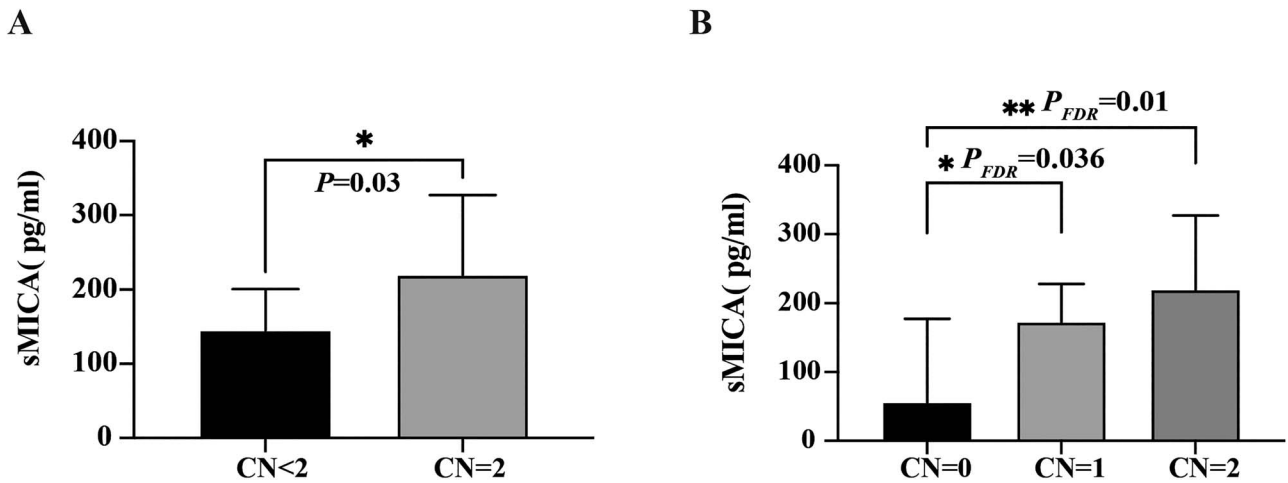


Figure 4. Evaluation of serum levels of soluble MICA. **(A)** Comparison of soluble MICA levels according to the copy number with/without deletions, P -value was from one-tailed test. **(B)** Comparisons of soluble MICA levels according to the copy number, P -value was from Kruskal–Wallis tests and corrected by false discovery rate. * $P \leq 0.05$, ** $P \leq 0.01$.

phenotypes giant cell arteritis, ankylosing spondylitis and hepatitis C and also with DNA methylation in the HLA region (46–49). In addition to MICA, three non-protein-coding genes map within CNVR1. HCP5 is a lncRNA gene primarily expressed in the immune system and its expression is modulated in chronic kidney disease (50), allergic rhinitis (51) and multiple cancers (52). HCP5 could be an example of an immune-priming lncRNA (53) that perhaps modulates expression of MICA—we were able to show that within CNVR1 there is physical interaction between the HCP5 and MICA genes via the LINC01149 locus (Fig. 3). Also mapping within the deletion are pseudogene HCG26 and non-coding Y RNA ENSG00000199332.1—the functional impact of each is poorly understood. On the basis of current understanding, we speculate that the CNVR1 deletion may increase the risk of gout via perturbed immune activation and surveillance pathways downstream of MICA and control of its expression by regulatory RNAs in the locus.

Despite CNVR1 being prevalent in Eastern Polynesian (15.8% in non-gout), it was not associated with gout in this ancestral group for reasons that are, as yet, unclear. However, we have observed a dichotomy in association with gout between Eastern and Western Polynesian at the ABCG2 locus, despite the gout-associated rs2231142 risk allele being prevalent in both ancestral groups (54). Similarly, the ABCG2 rs10011796 risk allele was associated with tophi in Western Polynesian but not in Eastern Polynesian people (55). The dichotomy may relate to other interacting variants that also differ in frequency between the Eastern and Western Polynesian ancestral groups. Our results underline that not only pan-ancestry but also sub-regional population differences should be taken into account when conducting biomedical genetic research in the Māori and Pacific populations of Aotearoa NZ.

In conclusion, we identified Polynesian-specific CNVs. Three CNVRs at Chr6: 31.36–31.45 Mb, Chr1: 196.75–196.92 Mb and Chr1: 189.35–189.54 Mb showed evidence

for association with gout in Polynesian. We identified MICA as a strong candidate causal gene and demonstrated reduced circulating MICA levels in people with deletion at CNVR1. Our study provides new insights into the pathogenic mechanism underlying gout in people of Māori and Pacific ancestry living in Aotearoa NZ.

Materials and Methods

Participants

The Aotearoa NZ Polynesian sample set (individuals of Aotearoa NZ and Cook Island Māori, Samoan, Tongan, Niuean and Tokelauan ancestry) comprised 1196 participants with and 1249 without gout (Table 3). All people with gout met the 1977 American Rheumatism Association preliminary gout classification criteria (56). The sample set included 171 Māori with and 98 Māori without gout from the rohe (area) of the Ngāti Porou iwi (tribe) of the Tairāwhiti region of Aotearoa NZ. The Aotearoa NZ European sample set (552 participants without gout) and the Han Chinese sample set (57) (1962 participants without gout) were included for comparison purposes to facilitate identification of Polynesian-specific CNV. All ethnicity was self-reported.

The Polynesian sample set was divided into three groups (Eastern Polynesian, Western Polynesian and Mixed Eastern–Western Polynesian). The Eastern Polynesian sample set composed of Aotearoa NZ Māori, Cook Island Māori, French Polynesian and other Polynesian (excluding Pukapuka) (732 gout, 795 without gout). The Western Polynesian sample set composed of Samoa, Tonga, Tuvalu, Niue, Tokelau, Pukapuka and other Polynesian (420 gout, 385 without gout). The Mixed Eastern–Western Polynesian group composed 44 people with and 69 without gout of mixed Eastern and Western Polynesian ethnicity. Serum urate measurements were performed at the Southern Community Laboratories (www.sclabs.co.nz) for all Aotearoa NZ samples.

Table 3. Age, sex and serum urate details of Polynesian sample sets

Sample set	Classification	Male			Female		
		Number (%)	Age ^a (SD)	Urate mmol/l ^b (SD)	Number (%)	Age ^a (SD)	Urate mmol/l ^b (SD)
EP	Gout	561(76.6)	56.4 (12.0)	0.40 (0.13)	171 (23.4)	60.7 (12.1)	0.36 (0.14)
	Non-gout	331 (41.6)	45.7 (15.9)	0.38 (0.11)	464 (58.4)	46.7 (15.0)	0.32 (0.05)
WP	Gout	364(86.7)	48.0 (12.3)	0.44 (0.13)	56(13.3)	54.3 (13.5)	0.38 (0.17)
	Non-gout	200 (51.9)	35.5 (14.5)	0.41 (0.10)	185 (48.1)	38.4 (16.6)	0.32 (0.09)
EPWP	Gout	36(81.8)	42.9 (12.9)	0.46 (0.12)	8(18.2)	48.1 (11.0)	0.45 (0.17)
	Non-gout	34 (49.3)	35.1 (17.5)	0.43 (0.09)	35 (50.7)	38.0 (14.5)	0.33 (0.09)

EP, Eastern Polynesian; EPWP, Mixed Eastern–Western Polynesian; WP, Western Polynesian. ^aAge at diagnosis for gout cases, at recruitment for controls.

^bSerum urate levels at recruitment. Data presented for individuals for whom sex information was available.

Generation of exome array genotype and WGS Data

A total of 3937 Aotearoa NZ Polynesian and European individuals were genotyped in two batches (batch 1 with 291 and batch 2 with 3646 individuals) on the Illumina Infinium CoreExome-24 platform. (2997 used in this study were a subset.) From the data generated, log R ratios (LRRs) and B allele frequencies (BAFs) were extracted from the autosomes and used to identify CNV. Samples with a genotyping call rate of <98% and SNVs with a call rate of <95% were not utilized in the subsequent analysis (58). Principal component (PC) analysis was applied to the LRR values, and the first six PCs were visualized in scatter plots to evaluate the possibility of batch effects. Differential effects were not identified when samples were classified according to batch (Supplementary Material, Fig. S7) or ancestral group (Supplementary Material, Fig. S8). The Han Chinese controls of 1962 individuals were genotyped using the Affymetrix axion genome-wide Chinese Han Beijing array. Samples with a genotyping call rate of <95% and SNVs with a call rate of <95% were excluded. A total of 1848 Han Chinese samples were used in the subsequent analyses. Ninety-one Aotearoa NZ Polynesian individuals were sequenced to high coverage on a HiSeqX machine using TruSeq Nano libraries to allow validation of identified CNVRs.

CNV identification and assignment of genotype

Supplementary Material, Figure S9 illustrates the workflow. All genomic locations were derived from NCBI GRCh37/UCSC hg19 coordinates. LRR and BAF values were used to determine copy number. As a normalized measure of total signal intensity, LRR values of 0 represent two copies with lower values in identifying areas of loss and higher values areas of gain. The BAF is a measure of the allelic intensity ratio, with values ranging from 0 to 1. Areas of homozygosity (e.g. deletion) have BAF of 0 or 1, normal diploid regions have BAF of 0, 0.5 or 1 and allelic imbalance areas (e.g. duplication) show intermediate values. PennCNV (v1.0.5) (59) was applied to identify CNV (CN ≠ 2), including homozygous deletion, heterozygous deletion and duplication. This

tool optimized a trained hidden Markov model transition file and a guanine–cytosine (GC) content file that was created for all the SNVs to segment the normalized and transformed probe intensities (LRRs and BAFs) and incorporate population SNV frequencies of the B allele (PFB). In this study, the PFB values for all SNVs of each group were computed from the relevant non-gout sample sets via their intensity files and the ‘compile_pfb.pl’ script in PennCNV. Then, CNV calls were generated using the ‘detect_cnv.pl’ script, along with the customized parameter files (PFB files and GC content files) and intensity files.

CNV quality control and definition of CNVR

Any CNVs that were supported by <10 SNPs or were <5 kb in length and were low confidence (confidence score <40) were removed (Supplementary Material, Fig. S9). Additional subject quality control was performed to exclude subjects: samples with LRR standard deviation >0.28, a waviness factor (the amount of dispersion in signal intensity) (60) >0.05 or <−0.05, or CNV count >312 (85th percentile, see Supplementary Material, Fig. S10) were not utilized in downstream analyses (61). Subsequently, all CNV calls were manually confirmed by visualizing LRR and BAF plots. When overlapping CNVs were identified, they were merged into unique CNV regions via CNVRuler (62), taking the outmost boundaries of the union of those CNVs (Supplementary Material, Table S1).

Statistical analysis

All statistical analysis was performed using R (version 4.0.3). To test for statistically significant differences in the distribution of copy number between cases and controls, regression analyses were implemented within each Polynesian population group. Two models were used: adjustment with age and sex (Model 1); and Model 1 plus adjustment by genotyping batches and PCs 1–10 derived from the genome-wide genotype data as previously described (29) (Model 2). Combination of the three Polynesian subsets was done by meta-analysis with fixed-effects model (except for CNVR1 and burden test). Kruskal–Wallis test and Mann–Whitney test were used to compare serum MICA levels between groups. A *P* < 0.05 threshold indicated nominal evidence for association.

Validation of CNV calls in WGS and publicly available datasets

Among the Polynesian cohort, 91 individuals with both microarray and WGS data were used to confirm the presence of CNV calls. Each chromosome was divided into 1 kb bins, and read depth of coverage in each bin was calculated by Samtools (63). The read depth of the bins was plotted to confirm the deletions and duplications. The DGV (<http://dgv.tcag.ca/dgv/app/home>) that includes annotated CNVs was also used for confirmation of CNV.

Gout GWAS in UK Biobank

A total of 332 370 European individuals from the UK Biobank (64) were included under the approval number 12611. The UK Biobank has ethical approval from the North West Multi-Centre Research Ethics Committee (11/NW/0382) and obtained written informed consent from all participants prior to the study. Gout cases were defined as individuals with any of self-reported gout, urate-lowering therapy use (allopurinol or sulfipyrazone) with no diagnosis of leukemia or lymphoma (ICD codes C81 and C96), primary or secondary diagnosis of gout (ICD code M10) (11). Exclusion criteria were sex chromosome and self-reported sex mismatch, genotype QC failure, relatedness (KING coefficient >0.177) and primary or secondary kidney disease. The GWAS was done using a total of 27 287 012 variants imputed from 845 487 genotyped variants. A logistic regression model was produced for each variant adjusting for age, sex and the first 40 genetic PCs using Plink version 1.9.6.10 (65).

Enzyme-linked immunosorbent assay

A Human MICA ELISA Kit (Invitrogen, Thermo Fisher Scientific, Waltham, Maryland, USA) was used to detect soluble MICA (following the manufacturer's instructions) in serum that had been stored at -80°C . Absorbance was measured at 450 nm. A standard curve of the logarithmic relationship between concentration and absorbance was used to calculate the concentration of soluble MICA in serum samples. All people homozygous for the deletion (CN=0) were selected for measuring of MICA serum levels. For each of CN=1 and CN=2, we began by randomly selecting 30 gout and 30 non-gout individuals (i.e. 60 for each genotype). However, after accounting for unavailability of serum or insufficient serum, the eventual numbers assayed were 49 for CN=1 and 43 for CN=2. Association data across the CNVR1, CNVR6 and CNVR9 were specifically queried from these GWAS data.

Author contributions

T.R.M., Y.S. and K.W. designed and supervised the project. K.W., M.C., M.B., M.P.L., M.E.M., Q.Y., Z.L., R.T., A.P.-G., T.J.M., R.T., N.D., F.K., R.M., L.K.S., J.d.Z. and Z.W. collected and contributed to analysis of data. K.W. and T.R.M. wrote the manuscript with input from co-authors. All authors contributed to and have approved the final manuscript.

Acknowledgements

The authors thank the many participants who so generously donated their genetic sample and other information for the study. The authors would also like to thank Ria Akuhata, Jordyn Allan, Nancy Aupouri, Jill Drake, Carol Ford, Roddi Laurence, Christopher Franklin, Meaghan House and Gabrielle Sexton for recruitment; Jennie Harré Hindmarsh (formerly of Ngāti Porou Hauora) for her contribution and NeSI (New Zealand eScience Infrastructure) for provision of compute resource.

Supplementary Material

Supplementary Material is available at HMG online.

Conflict of Interest statement. None of the authors declare any conflict of interest related to the research presented in this manuscript.

Funding

New Zealand Health Research Council Program (grants 08/075, 11/1075, 14/527); Natural Science Foundation of China (U1804284, 32070679, 81871051, 82071500, 81701321, 81871055); National Key R&D Program of China (2019YFA0905400, 2017YFC0908105, 2016YFC1306903); Shanghai Science and Technology Committee (17490712 200); Shanghai Municipal Science and Technology Major Project (2017SHZDZX01, 18DZ2260200); Shanghai Natural Science Funding (16ZR1449700); National Program for Support of Top-Notch Young Professionals; Shanghai Key Laboratory of Psychotic Disorders (13dz2260500); Shanghai Youth Top-notch Talent Support Program (19MC911100); Shanghai Municipal Health Commission (ZK2015B01, 201540114); Taishan Scholar Program of Shandong Province (tsqn201812153); Natural Science Foundation of Shandong Province (ZR2019YQ14).

References

1. Dalbeth, N., Gosling, A.L., Gaffo, A. and Abhishek, A. (2021) Gout. *Lancet*, **397**, 1843–1855.
2. Terkeltaub, R.A. (2003) Gout. *N. Engl. J. Med.*, **349**, 1647–1655.
3. Annemans, L., Spaepen, E., Gaskin, M., Bonnemaire, M., Malier, V., Gilbert, T. and Nuki, G. (2008) Gout in the UK and Germany: prevalence, comorbidities and management in general practice 2000–2005. *Ann. Rheum. Dis.*, **67**, 960–966.
4. Feig, D.I., Kang, D.-H. and Johnson, R.J. (2008) Uric acid and cardiovascular risk. *N. Engl. J. Med.*, **359**, 1811–1821.
5. Richette, P., Clerson, P., Périsin, L., Flipo, R.-M. and Bardin, T. (2015) Revisiting comorbidities in gout: a cluster analysis. *Ann. Rheum. Dis.*, **74**, 142–147.
6. Dalbeth, N., Merriman, T.R. and Stamp, L.K. (2016) Gout. *Lancet*, **388**, 2039–2052.
7. Tin, A., Marten, J., Halperin Kuhns, V.L., Li, Y., Wuttke, M., Kirsten, H., Sieber, K.B., Qiu, C., Gorski, M., Yu, Z. et al. (2019) Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. *Nat. Genet.*, **51**, 1459–1474.

8. Nakatochi, M., Kanai, M., Nakayama, A., Hishida, A., Kawamura, Y., Ichihara, S., Akiyama, M., Ikezaki, H., Furusyo, N., Shimizu, S. et al. (2019) Genome-wide meta-analysis identifies multiple novel loci associated with serum uric acid levels in Japanese individuals. *Comm. Biol.*, **2**, 115.
9. Boocock, J., Leask, M., Okada, Y., Asian Genetic Epidemiology Network Consortium, Matsuo, H., Kawamura, Y., Shi, Y., Li, C., Mount, D.B., Mandal, A.K. et al. (2020) Genomic dissection of 43 serum urate-associated loci provides multiple insights into molecular mechanisms of urate control. *Hum. Mol. Genet.*, **29**, 923–943.
10. Major, T.J., Dalbeth, N., Stahl, E.A. and Merriman, T.R. (2018) An update on the genetics of hyperuricaemia and gout. *Nat. Rev. Rheumatol.*, **14**, 341–353.
11. Cadzow, M., Merriman, T.R. and Dalbeth, N. (2017) Performance of gout definitions for genetic epidemiological studies: analysis of UK Biobank. *Arthritis Res. Ther.*, **19**, 181.
12. Walsh, T., McClellan, J.M., McCarthy, S.E., Addington, A.M., Pierce, S.B., Cooper, G.M., Nord, A.S., Kusenda, M., Malhotra, D., Bhandari, A. et al. (2008) Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science*, **320**, 539–543.
13. Rovelet-Lecrux, A., Hannequin, D., Raux, G., Meur, N.L., Laquerrière, A., Vital, A., Dumanchin, C., Feuillet, S., Brice, A., Vercelletto, M. et al. (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. *Nat. Genet.*, **38**, 24–26.
14. Froyen, G., Corbett, M., Vandewalle, J., Jarvela, I., Lawrence, O., Meldrum, C., Bauters, M., Govaerts, K., Vandeleur, L., Van Esch, H. et al. (2008) Submicroscopic duplications of the hydroxysteroid dehydrogenase HSD17B10 and the E3 ubiquitin ligase HUWE1 are associated with mental retardation. *Am. J. Hum. Genet.*, **82**, 432–443.
15. Pinto, D., Pagnamenta, A.T., Klei, L., Anney, R., Merico, D., Regan, R., Conroy, J., Magalhaes, T.R., Correia, C., Abrahams, B.S. et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*, **466**, 368–372.
16. Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J. et al. (2007) Strong association of de novo copy number mutations with autism. *Science*, **316**, 445–449.
17. Fellermann, K., Stange, D.E., Schaeffeler, E., Schmalzl, H., Wehkamp, J., Bevins, C.L., Reinisch, W., Teml, A., Schwab, M., Lichter, P. et al. (2006) A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. *Am. J. Hum. Genet.*, **79**, 439–448.
18. Bentley, R.W., Pearson, J., Gearry, R.B., Barclay, M.L., McKinney, C., Merriman, T.R. and Roberts, R.L. (2010) Association of higher DEFB4 genomic copy number with Crohn's disease. *Am. J. Gastroenterol.*, **105**, 354–359.
19. McKinney, C., Merriman, M.E., Chapman, P.T., Gow, P.J., Harrison, A.A., Highton, J., Jones, P.B., McLean, L., O'Donnell, J.L., Pokorny, V. et al. (2008) Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. *Ann. Rheum. Dis.*, **67**, 409–413.
20. McKinney, C., Fanciulli, M., Merriman, M.E., Phipps-Green, A., Alizadeh, B.Z., Koeleman, B.P., Dalbeth, N., Gow, P.J., Harrison, A.A., Highton, J. et al. (2010) Association of variation in Fc gamma receptor 3B gene copy number with rheumatoid arthritis in Caucasian samples. *Ann. Rheum. Dis.*, **69**, 1711–1716.
21. McKinney, C., Broen, J.C., Vonk, M.C., Beretta, L., Hesselstrand, R., Hunzelmann, N., Riemekasten, G., Scorza, R., Simeon, C.P., Fonollosa, V. et al. (2012) Evidence that deletion at FCGR3B is a risk factor for systemic sclerosis. *Genes Immun.*, **13**, 458–460.
22. Dong, Z., Li, Y., Zhou, J., Jiang, S., Wang, Y., Chen, Y., Zhao, D., Yang, C., Qian, Q., Ma, Y. et al. (2017) Copy number variants of ABCF1, IL17REL, and FCGR3A are associated with the risk of gout. *Protein Cell*, **8**, 467–470.
23. Scharpf, R.B., Mireles, L., Yang, Q., Köttgen, A., Ruczinski, I., Susztak, K., Halper-Stromberg, E., Tin, A., Cristiano, S., Chakravarti, A. et al. (2014) Copy number polymorphisms near SLC2A9 are associated with serum uric acid concentrations. *BMC Genet.*, **15**, 81.
24. Dalbeth, N., Dowell, T., Gerard, C., Gow, P., Jackson, G., Shuker, C. and Te Karu, L. (2018) Gout in Aotearoa New Zealand: the equity crisis continues in plain sight. *N. Z. Med. J.*, **131**, 8–12.
25. Guillen, A.G., Te Karu, L., Singh, J.A. and Dalbeth, N. (2020) Gender and ethnic inequities in gout burden and management. *Rheum. Dis. Clin. N. Am.*, **46**, 693–703.
26. Klück, V., van Deuren, R.C., Cavalli, G., Shaikat, A., Arts, P., Cleophas, M.C., Crisan, T.O., Tausche, A.K., Riches, P., Dalbeth, N. et al. (2020) Rare genetic variants in interleukin-37 link this anti-inflammatory cytokine to the pathogenesis and treatment of gout. *Ann. Rheum. Dis.*, **79**, 536–544.
27. Tanner, C., Boocock, J., Stahl, E.A., Dobbyn, A., Mandal, A.K., Cadzow, M., Phipps-Green, A.J., Topless, R.K., Hindmarsh, J.H., Stamp, L.K. et al. (2017) Population-specific resequencing associates the ATP-binding cassette subfamily C member 4 gene with gout in New Zealand Māori and Pacific Men. *Arthritis Rheumatol.*, **69**, 1461–1469.
28. Ji, A., Shaikat, A., Takei, R., Bixley, M., Cadzow, M., Topless, R.K., Major, T.J., Phipps-Green, A., Merriman, M.E., Hindmarsh, J.H. et al. (2021) Aotearoa New Zealand Māori and Pacific population-specific gout risk variants: CLNK is a separate risk gene at the SLC2A9 locus. *J. Rheumatol.*, **48**, 1736–1744.
29. Krishnan, M., Major, T.J., Topless, R.K., Dewes, O., Yu, L., Thompson, J.M.D., McCowan, L., de Zoysa, J., Stamp, L.K., Dalbeth, N. et al. (2018) Discordant association of the CREBRF rs373863828 A allele with increased BMI and protection from type 2 diabetes in Māori and Pacific (Polynesian) people living in Aotearoa/New Zealand. *Diabetologia*, **61**, 1603–1613.
30. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Stein, T.I., Rosen, N., Kohn, A., Twik, M., Safran, M. et al. (2017, 2017) GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*, bax028.
31. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D. and Lander, E.S. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin loop organization. *Cell*, **7**, 1665–1680.
32. van Manen, D., Kootstra, N.A., Boeser-Nunnink, B., Handulle, M.A. and van't Wout, A.B. and Schuitemaker, H. (2009) Association of HLA-C and HCP5 gene regions with the clinical course of HIV-1 infection. *AIDS*, **23**, 19–28.
33. Lane, L.C., Kus, A., Bednarczuk, T., Bossowski, A., Daroszewski, J., Jurecka-Lubieniecka, B., Cordell, H.J., Pearce, S.H.S., Cheetham, T. and Mitchell, A.L. (2020) An intronic HCP5 variant is associated with age of onset and susceptibility to Graves disease in UK and Polish cohorts. *J. Clin. Endocrinol. Metab.*, **105**, e3277–e3284.
34. Ciccacci, C., Perricone, C., Ceccarelli, F., Rufini, S., Di Fusco, D., Alessandri, C., Spinelli, F.R., Cipriano, E., Novelli, G., Valesini, G. et al. (2014) A multilocus genetic study in a cohort of Italian SLE patients confirms the association with STAT4 gene

- and describes a new association with HCP5 gene. *PLoS One*, **9**, e111991.
35. Tse, K.P., Su, W.H., Yang, M.L., Cheng, H.Y., Tsang, N.M., Chang, K.P., Hao, S.P., Yao Shugart, Y. and Chang, Y.S. (2011) A gender-specific association of CNV at 6p21.3 with NPC susceptibility. *Hum. Mol. Genet.*, **20**, 2889–2896.
 36. Low, J.S.Y., Chin, Y.M., Mushiroda, T., Kubo, M., Govindasamy, G.K., Pua, K.C., Yap, Y.Y., Yap, L.F., Subramaniam, S.K., Ong, C.A. et al. (2016) A genome wide study of copy number variation associated with nasopharyngeal carcinoma in Malaysian Chinese identifies CNVs at 11q14.3 and 6p21.3 as candidate loci. *PLoS One*, **11**, e0145774.
 37. Cooper, N.J., Shtir, C.J., Smyth, D.J., Guo, H., Swafford, A.D., Zanda, M., Hurles, M.E., Walker, N.M., Plagnol, V., Cooper, J.D. et al. (2015) Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes. *Hum. Mol. Genet.*, **24**, 1774–1790.
 38. Vega-Sevey, J.G., Martínez-Magaña, J.J., Genis-Mendoza, A.D., Escamilla, M., Lanzagorta, N., Tovilla-Zarate, C.A. and Nicolini, H. (2020) Copy number variants in siblings of Mexican origin concordant for schizophrenia or bipolar disorder. *Psychiatry Res.*, **291**, 113018.
 39. Vazquez-Garcia, G., Mentch, F., Chang, X., Zuppa, A., Goldfarb, S.B. and Hakonarson, H. (2018) A recurrent hemizygous deletion at the MHC region impacting MICA, LINC01149, HCP5, HCG26 associates with idiopathic pulmonary hemosiderosis. *Am. J. Resp. Crit. Care Med.*, **197**, A6322.
 40. Bauer, S. (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science*, **285**, 727–729.
 41. Malhotra, A. and Shanker, A. (2011) NK cells: immune cross-talk and therapeutic implications. *Immunotherapy*, **3**, 1143–1166.
 42. Costantini, C. and Cassatella, M.A. (2011) The defensive alliance between neutrophils and NK cells as a novel arm of innate immunity. *J. Leukoc. Biol.*, **89**, 221–233.
 43. Wang, J., Liu, K., Xiao, T., Liu, P., Prinz, R.A. and Xu, X. (2022) Uric acid accumulation in DNA-damaged tumor cells induces NKG2D ligand expression and antitumor immunity by activating TGF- β -activated kinase 1. *OncolImmunology*, **11**, 2016159.
 44. Xu, X., Rao, G. and Li, Y. (2016) Xanthine oxidoreductase is required for genotoxic stress-induced NKG2D ligand expression and gemcitabine-mediated antitumor activity. *Oncotarget*, **7**, 59220–59235.
 45. Xu, X., Rao, G.S., Groh, V., Spies, T., Gattuso, P., Kaufman, H.L., Plate, J. and Prinz, R.A. (2011) Major histocompatibility complex class I-related chain A/B (MICA/B) expression in tumor tissue and serum of pancreatic cancer: role of uric acid accumulation in gemcitabine-induced MICA/B expression. *BMC Cancer*, **11**, 194.
 46. Ahsan, M., Ek, W.E., Rask-Andersen, M., Karlsson, T., Lind-Thomsen, A., Enroth, S., Gyllenstein, U. and Johansson, Å. (2017) The relative contribution of DNA methylation and genetic variants on protein biomarkers for human diseases. *PLoS Genet.*, **13**, e1007005.
 47. Sun, W., Kechris, K., Jacobson, S., Drummond, M.B., Hawkins, G.A., Yang, J., Chen, T.H., Quibrera, P.M., Anderson, W., Barr, R.G. et al. (2016) Common genetic polymorphisms influence blood biomarker measurements in COPD. *PLoS Genet.*, **12**, e1006011.
 48. Australo-Anglo-American Spondyloarthritis Consortium (TASC), Reveille, J.D., Sims, A.M., Danoy, P., Evans, D.M., Leo, P., Pointon, J.J., Jin, R., Zhou, X., Bradbury, L.A. et al. (2010) Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. *Nat. Genet.*, **42**, 123–127.
 49. Sakaue, S., Kania, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M. et al. (2021) A cross-population atlas of genetic associations for 220 human phenotypes. *Nat. Genet.*, **53**, 1415–1424.
 50. Li, N., Cui, Y., Yin, M. and Liu, F. (2019) Screening potential prognostic biomarkers of long non-coding RNAs for predicting the risk of chronic kidney disease. *Braz. J. Med. Biol. Res.*, **52**, e8333.
 51. Yang, C., Shangguan, C., Cai, C., Xu, J. and Qian, X. (2022) LncRNA HCP5 participates in the Tregs functions in allergic rhinitis and rives airway mucosal inflammatory response in the nasal epithelial cells. *Inflammation*, Published online 5 Feb 2022.
 52. Zou, Y. and Chen, B. (2021) Long non-coding RNA HCP5 in cancer. *Clin. Chim. Acta*, **512**, 33–39.
 53. Fanucchi, S., Fok, E.T., Dalla, E., Shibayama, Y., Börner, K., Chang, E.Y., Stoychev, S., Imakaev, M., Grimm, D., Wang, K.C. et al. (2019) Immune genes are primed for robust transcription by proximal long noncoding RNAs located in nuclear compartments. *Nat. Genet.*, **51**, 138–150.
 54. Phipps-Green, A.J., Hollis-Moffatt, J.E., Dalbeth, N., Merriman, M.E., Topless, R., Gow, P.J., Harrison, A.A., Highton, J., Jones, P.B.B., Stamp, L.K. et al. (2010) A strong role for the ABCG2 gene in susceptibility to gout in New Zealand Pacific Island and Caucasian, but not Māori, case and control sample sets. *Hum. Mol. Genet.*, **19**, 4813–4819.
 55. He, W., Phipps-Green, A., Stamp, L.K., Merriman, T.R. and Dalbeth, N. (2017) Population-specific association between ABCG2 variants and tophaceous disease in people with gout. *Arthritis. Res. Ther.*, **19**, 43.
 56. Wallace, S.L., Robinson, H., Masi, A.T., Decker, J.L., McCarty, D.J. and Yü, T.F. (1977) Preliminary criteria for the classification of the acute arthritis of primary gout. *Arthritis Rheum.*, **20**, 895–900.
 57. Li, C., Li, Z., Liu, S., Wang, C., Han, L., Cui, L., Zhou, J., Zou, H., Liu, Z., Chen, J. et al. (2015) Genome-wide association analysis identifies three new risk loci for gout arthritis in Han Chinese. *Nat. Commun.*, **6**, 7041.
 58. Szatkiewicz, J.P., Neale, B.M., O'Dushlaine, C., Fromer, M., Goldstein, J.I., Moran, J.L., Chambert, K., Kähler, A., Magnusson, P.K.E., Hultman, C.M. et al. (2013) Detecting large copy number variants using exome genotyping arrays in a large Swedish schizophrenia sample. *Mol. Psychiatry*, **18**, 1178–1184.
 59. Wang, K., Li, M., Hadley, D., Liu, R., Glessner, J., Grant, S.F.A., Hakonarson, H. and Bucan, M. (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.*, **17**, 1665–1674.
 60. Diskin, S.J., Li, M., Hou, C., Yang, S., Glessner, J., Hakonarson, H., Bucan, M., Maris, J.M. and Wang, K. (2008) Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucl. Acids Res.*, **36**, e126.
 61. Glessner, J.T., Bick, A.G., Ito, K., Homsy, J., Rodriguez-Murillo, L., Fromer, M., Mazaika, E., Vardarajan, B., Italia, M., Leipzig, J. et al. (2014) Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ. Res.*, **115**, 884–896.
 62. Kim, J.H., Hu, H.J., Yim, S.H., Bae, J.S., Kim, S.Y. and Chung, Y.J. (2012) CNVRuler: a copy number variation-based case-control association analysis tool. *Bioinformatics*, **28**, 1790–1792.

63. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
64. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliot, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J. et al. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature*, **562**, 203–209.
65. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.L., Daly, M.J. and Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage studies. *Am. J. Hum. Genet.*, **81**, 559–575.