Genetic Association and Transferability for Urinary Albumin-Creatinine Ratio as a Marker of Kidney Function in four Sub-Saharan African Populations and non-continental Individuals of African Ancestry

Jean-Tristan Brandenburg (PhD)[1,2], Wenlong Carl Chen (PhD)[1,2,3], Palwende Romuald Boua (PhD)[1,4], Melanie Ann Govender (MSc)[1,5], Godfred Agongo (PhD)[6,7], Lisa K. Micklesfield (PhD)[8], Hermann Sorgho (PhD)[4], Stephen Tollman (MD, PhD)[9], Gershim Asiki (MD, PhD)[10,11], Felistas Mashinya (PhD)[12], Scott Hazelhurst (PhD)[1,13], Andrew P Morris (PhD)[14], June Fabian (MD)[15*], Michèle Ramsay (PhD)[1,5*] as members of AWI-Gen and the H3Africa Consortium

[1] Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[2] Strengthening Oncology Services Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[3] National Cancer Registry, National Health Laboratory Service, Johannesburg, South Africa
[4] Clinical Research Unit of Nanoro, Institut de Recherche en Sciences de la Santé, Burkina Faso
[5] Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[6] Navrongo Health Research Centre, Navrongo, Ghana
[7] Department of Biochemistry and Forensic Sciences, School of Chemical and Biochemical Sciences, C. K. Tedam University of Technology and Applied Sciences, Navrongo, Ghana
[8] SAMRC/Wits Developmental Pathways for Health Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[9] Medical Research Council/Wits University Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[10] African Population and Health Research Center, Nairobi, Kenya.
[11] Department of Women's and Children's Health, Karolinska Institute, Stockholm, Sweden
[12] Department of Pathology and Medical Sciences, School of Health Care Sciences, Faculty of Health Sciences, University of Limpopo, Polokwane, South Africa
[13] School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa
[14] Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, The University of Manchester, Manchester, UK
[15] Wits Donald Gordon Medical Centre, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

*These authors contributed equally to this work and share senior authorship

Corresponding authors: Dr Jean-Tristan Brandenburg and Prof Michèle Ramsay, Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of the Witwatersrand, 9 Jubilee Road, Parktown, Johannesburg, South Africa

Tel: +27 11 717 6635
E-mail: jean-tristan.brandenburg@wits.ac.za and michele.ramsay@wits.ac.za

Support: National Institutes of Health (NIH) H3Africa AWI-Gen Collaborative Centre (U54HG006938).

Running title: GWAS for UACR in African-ancestry populations

Keywords (6)

Chronic kidney disease (CKD), genome-wide association study (GWAS), urinary albumin-creatinine ratio (UACR), African diversity, polygenic risk score

Abstract (251 words)

Background: Genome-wide association studies (GWAS) have predominantly focused on populations of European and Asian ancestry, limiting our understanding of genetic factors influencing kidney function in Sub-Saharan African (SSA) populations. This study presents the largest GWAS for urinary albumin-to-creatinine ratio (UACR) in SSA individuals, including 8,970 participants living in different African regions and an additional 9,705 non-resident individuals of African ancestry from the UK Biobank and African American cohorts.

Methods: Urine biomarkers and genotype data were obtained from two SSA cohorts (AWI-Gen and ARK), and two non-resident African-ancestry studies (UK Biobank and CKD-Gen Consortium). Association testing and meta-analyses were conducted, with subsequent fine-mapping, conditional analyses, and replication studies. Polygenic scores (PGS) were assessed for transferability across populations.

Results: Two genome-wide significant ($P<5x10^{-8}$) UACR-associated loci were identified, one in the *BMP6 region* on chromosome 6, in the meta-analysis of resident African individuals, and another in the *HBB* region on chromosome 11 in the meta-analysis of non-resident SSA individuals, as well as the combined meta-analysis of all studies. Replication of previous significant results confirmed associations in known UACR-associated regions, including *THB53*, *GATM*, and *ARL15*. PGS estimated using previous studies from European ancestry, African ancestry, and multi-ancestry cohorts exhibited limited transferability of PGS across populations, with less than 1% of observed variance explained.

Conclusion: This study contributes novel insights into the genetic architecture of kidney function in SSA populations, emphasizing the need for conducting genetic research in diverse

cohorts. The identified loci provide a foundation for future investigations into the genetic susceptibility to chronic kidney disease in underrepresented African populations.

Introduction

Chronic kidney disease (CKD) is a leading risk factor for years of life lost and premature mortality, with a 41.5% relative increase in mortality worldwide from 1990 to 2017 (GBD Chronic Kidney Disease Collaboration et al., 2020; Kovesdy, 2022). The estimated global prevalence of CKD is 9.1% and while predicted to be higher in Sub-Saharan Africa (SSA), the true prevalence and associated risk factors remain understudied (Kaze et al., 2018; GBD Chronic Kidney Disease Collaboration et al., 2020). The Africa Wits-INDEPTH partnership for Genomic Studies (AWI-Gen) cohort, which included participants from four SSA countries in West, East, and Southern Africa, reported overall CKD prevalence as 10.7% (95% confidence interval [CI]: 9.9-11.7), with notable geographic regional differences. The most important risk factors for CKD in SSA were Older age, female sex, diabetes, hypertension, and human immunodeficiency virus (HIV) (George et al., 2019).

Over the past decade, genome-wide association studies (GWAS) have identified numerous genetic loci associated with CKD and commonly used measures of kidney function, namely urine albumin-creatinine ratio [UACR], estimated glomerular filtration rate [eGFR], and serum creatinine (Böger et al., 2011; Pattaro et al., 2012, 2016; Teumer et al., 2016, 2019; Hellwege et al., 2019; Tin and Köttgen, 2020). Majority of the GWAS for kidney function have examined associations with eGFR, while UACR, as a measure for albuminuria, has been investigated less often (Mahajan et al., 2016; Pattaro et al., 2016; Gorski et al., 2017; Haas et al., 2018; Teumer et al., 2019; Wuttke et al., 2019; Zanetti et al., 2019). In a recent GWAS in 564,257 individuals of multi-ancestry origins 68 associated risk loci for UACR were identified and a priority list of genes where compiled to explore as targets for the treatment of albuminuria (Teumer et al., 2019).

While the majority of kidney disease-associated risk loci have been identified in studies on participants of European and East Asian ancestry, and the African diaspora (Lee et al., 2018), few have focused on participants living in SSA (Böger et al., 2011; Pattaro et al., 2012; Lin et al., 2019; Morris et al., 2019). Recently, in a study of genetic associations of eGFR in a Ugandan population-based cohort, the association between eGFR and the *GATM* locus was replicated (Fatumo et al., 2020).

Replication and transferability of GWAS signals across populations of different ancestries, and specifically with African ancestry populations, tend to be poor despite regional replication often identifying shared associated genomic regions (Pattaro et al., 2012). This may be due to differences in linkage disequilibrium (LD) with the causal variant, allele frequency differences between the populations, underlying population structure, and variabilities in environmental exposures. African populations, with their great genetic diversity and deep evolutionary roots, represent an opportunity for genetic discovery to identify and fine-map disease-associated risk variants (Gomez et al., 2014; Pereira et al., 2021).

Polygenic scores (PGS) are used to quantify and stratify populations according to genetic risk. A PGS based on 63 eGFR-associated alleles showed significant association with kidney disease-related phenotypes, such as chronic kidney failure and hypertensive kidney disease in the Million Veteran Study (US) on 192,868 white and non-Hispanic individuals (Hellwege et al., 2019). A PGS based on 64 urine albumin-to-creatinine ratio (UACR) associated alleles was significantly associated with CKD (Teumer et al., 2019). Further analysis revealed positive associations of the PGS with an increased risk of HT and diabetes. However, PGS often translate poorly across different ancestries (Martin et al., 2017; Kamiza et al., 2022; Kachuri et al., 2023). Since most published GWAS for kidney disease and kidney function markers are based on European ancestry populations, the predictive accuracy of models developed from these studies is expected to be significantly diminished for African populations (Adam et al., 2022; Choudhury et al., 2022; Kamiza et al., 2023; Majara et al., 2023).

In this study, we present a GWAS for UACR conducted within resident Sub-Saharan African individuals. This population cross-sectional study includes a cohort of 8,970 individuals from four SSA countries from the AWI-Gen study (Ali et al., 2018), the African Research on Kidney Disease (ARK) study (Kalyesubula et al., 2020), with 9,705 individuals of African-ancestry from the UK Biobank (UKB) and African American participants from the CDK-Gen Consortium (Teumer et al., 2019). The primary objectives are to: (1) identify genetic loci associated with UACR as a marker of kidney function in individuals from SSA and of African ancestry; (2) explore the replication of findings identified in previous GWAS; (3) perform analysis and comparison of PGS derived from non-African and multi-ancestry population studies and evaluate their transferability to African populations.

Participants and Methods

Study participants

Africa Wits-INDEPTH Partnership for Genomic Research (AWI-Gen)

The study participants are a subset of the population cross-sectional AWI-Gen study (Ramsay et al., 2016; Ali et al., 2018). Adults primarily between the ages of 40 and 80 years were recruited from six SSA study sites in West Africa (Nanoro in Burkina Faso and Navrongo in Ghana), East Africa (Nairobi in Kenya) and South Africa (Bushbuckridge - hereinafter referred to as Agincourt, Dikgale, and Soweto) of self-identified black ethnicity. Data collection was described in detail previously (Ali et al., 2018; George et al., 2019). Detailed demographic data, health-related questionnaire data, and anthropometric measurements were collected. Peripheral blood samples and urine samples were collected for biomarker assays (the relevant assays are described below). DNA was extracted from peripheral blood-derived buffy coat samples and used for genotyping. Urine microalbumin was measured using a colorimetric method on the Cobas© 6000/c501 analyzer, and urine creatinine was measured by the modified Jaffe method(Craik et al., 2023). This study was approved by the Human Research Ethics Committee (Medical), University of the Witwatersrand, South Africa (M121029, M170880) and the ethics committees of all participating institutions. All participants provided written informed consent following community engagement and individual consenting processes.

African Research on Kidney Disease (ARK)

The African Research Kidney Disease (ARK) study is a well characterised population-based cohort study of 2021 adults (20–80 years) of self-identified black ethnicity from Agincourt, (Mpumalanga, South Africa) with demographic data, health-related questionnaire data, and anthropometric measurements collected at enrolment (Fabian et al., 2022). Blood and urine were collected for biomarker assays (the relevant assays are described below). DNA was extracted from buffy coat samples and used for genotyping. Urine microalbumin was measured using a colorimetric method on the Cobas© 6000/c501 analyzer, and urine creatinine was measured by the modified Jaffe method(Craik et al., 2023). *This* study was approved by the Human Research Ethics Committee (Medical), University of the Witwatersrand, South Africa (M160939). All participants provided written informed consent following community engagement and individual consenting processes. The geographical

area of recruitment overlaps with the Agincourt sub-cohort of AWI-Gen but there is no overlap in participants

UK Biobank (UKB)

Individuals of self-reported Caribbean and African ancestry from the UKB were identified for this study. Of this subset of UKB individuals, those with both genotyping and UACR data were retained for the analysis. UACR was derived using urinary levels of albumin and creatinine. In the UKB, albumin was measured using the immuno-turbidimetric analysis method (Randox Biosciences, UK) while creatinine was measured using the enzymatic analysis method (Beckman Coulter, UK) (Casanova et al., 2019).

Phenotype generation and **harmonization**

UACR was calculated for AWI-Gen and ARK studies using urinary levels of albumin and creatinine as previously described (George et al., 2019; Fabian et al., 2022). Participants with missing values for albumin and creatinine were excluded from this study. We applied filtering criteria similar to those employed by the CKD-Gen consortium (Köttgen and Pattaro, 2020). In cases where the values for urine albumin and urine creatinine fell outside the upper and lower limits of detection, the values were replaced with the respective upper and lower limits: for urine creatinine, the range was 3 to 400 mmol/L and for urine albumin, the range was 3.75 to 475 mg/L for AWI-Gen and ARK. For the UKB dataset, the upper limit was 6.7 mg/L for urine albumin. Albuminuria was defined as UACR >3.0mg/mmol.

Genotyping

AWI-Gen and ARK

Genomic DNA was genotyped using the H3Africa custom genotyping array. The H3Africa custom array was designed as an African-common-variant-enriched GWAS array (https://www.h3abionet.org/h3africa-chip) (Illumina) with ~2.3 million single nucleotide polymorphisms (SNPs).

UK-Biobank

Genotyping was performed by Affymetrix on two closely related purpose-designed arrays. ~50,000 participants were genotyped using the UK BiLEVE Axiom array (Resource 149600) and the remaining ~450,000 were genotyped using the UK Biobank Axiom array (Resource 149601). The dataset is a combination of results from both arrays. A total of 805,426 markers

were released in the genotype data. We extracted individuals with self-reported (Data-Field in dataset 21000) African Ancestry split between African (UKB-African) and Caribbean origins (UKB-Caribbean) from the raw dataset (Casanova et al., 2019).

Quality control

For each dataset, AWI-Gen (Choudhury et al., 2022), ARK and UKB, the following pre-imputation quality control (QC) steps were applied: removal of non-autosomal and mitochondrial SNPs; SNPs with genotype missingness greater than 0.05; minor allele frequency (MAF) less than 0.01; and Hardy-Weinberg equilibrium (HWE) P-value less than 0.0001. Individuals were excluded if they had more than 5% overall genotype missingness; heterozygosity lower than 0.150 and higher than 0.343; and discordant genotype and phenotype sex information. We used the GWAS QC workflow of the H3Africa Consortium Pan-African Bioinformatics Network to perform data QC (H3ABioNet H3AGWAS) (https://github.com/h3abionet/h3agwas) (Baichoo et al., 2018; Brandenburg et al., 2022a). In our final QC step, we identified and excluded outliers, admixed and related individuals using PCASmart, a feature of the EIGENSOFT software (Price et al., 2006), Admixture software (Alexander et al., 2009) using AGV (Gurdasani et al., 2015) and 1000 Genomes Project data (Auton et al., 2015) and PLINK (Version 1.9) (Purcell et al., 2007). More detail on filter parameters for each software can be found in **Supplementary Table 1**.

Imputation

Genotype imputation was performed on each dataset separately (AWI-Gen, ARK, and UKB) using the Sanger Imputation Server with the African Genome Resources reference panel (https://www.sanger.ac.uk/tool/sanger-imputation-service/). EAGLE2 was used for pre-phasing and the PBWT algorithm was used for imputation (Loh et al., 2016). After imputation, poorly imputed SNPs with info scores less than 0.3 and with a HWE P-value less than $1x10^{-04}$ were removed. The genomic positions were mapped to GRCh37p11.

Phenotype transformation for association testing

For AWI-Gen, ARK, and UKB datasets, UACR was transformed on the logarithm scale. Linear regression of variables was performed with covariates in R (Version 3.6): ln(UACR) ~age + sex + genetic principal components (PCs) 1-5. Residuals were extracted and transformed using Rank-Based Inverse Normal Transformation to ensure the normal distribution of residuals (Casanova et al., 2019). PCs were calculated using a sub-set of LD

pruned pre-imputed SNPs in PLINK (Version 1.9) (Purcell et al., 2007). The sub-set was derived by LD pruning using PLINK (Version 1.9) (Purcell et al., 2007) with an LD (r^2) threshold of 0.2 with windows of 50 kb and 10 kb for step size.

Association testing

Mixed model association testing was performed with imputed genotype probabilities using GEMMA (Version 0.98.1) (Zhou and Stephens, 2012). GEMMA uses a relatedness matrix to account for genetic structure and relatedness between individuals. The relatedness matrix was built with a sub-set of pre-imputed SNPs described above.

Mixed model association testing was performed independently on each dataset. A total of nine datasets were tested. The datasets were defined as follows: six datasets for AWI-Gen: AWI-Agincourt, AWI- Dikgale, AWI-Nanoro, AWI-Nairobi, AWI-Navrongo and AWI-Soweto; one dataset for ARK: ARK-Agincourt; and two datasets for UK Biobank: UKB-Caribbean and UKB-African. For each dataset, Quantile-to-quantile plots (QQ-plots) were generated, and inflation factors were calculated using SNPs with MAF>0.01 to verify that the association signals were not inflated due to unaccounted population sub-structure. The genome-wide significance level for novel discovery was considered at $P<5x10^{-08}$.

CKD-Gen

We used previously published meta-analysis summary statistics from the CKD-Gen Consortium. The CKD-Gen Consortium datasets consist of three meta-analysis summary statistics: 1) CKD-Gen European ancestry individuals (CKD-Gen-EA); 2) CKD-Gen African American ancestry individuals (CKD-Gen-AA); and 3) CKD-Gen Multi-ancestry individuals (CKD-Gen-MA) which include individuals from CKD-Gen-EA and CKD-Gen-AA (Teumer et al., 2019). The CKD-Gen Consortium meta-analysis summary statistics were retrieved from http://ckdgen.imbi.uni-freiburg.de/.

Briefly, CKD-Gen-AA is a meta-analysis based on 7 studies with African American participants. For each study, genotyping was performed using genome-wide arrays followed by application of study-specific quality filters prior to phasing, imputation, and association analysis software (description can be found in Supplementary Table 1 and 2 from (Teumer et al., 2019)). Meta-analysis was performed using fixed effects inverse-variance weighted meta-analysis of the study-specific GWAS result files with imputation quality (IQ) score > 0.6 and MAC > 10, effective sample size ≥ 100, and a beta < 10, using METAL (for more details see (Teumer et al., 2019)).

Meta-analysis

Three fixed-effect meta-analyses were conducted using the METASOFT software (Han and Eskin, 2011). The first meta-analysis ($Meta_{SSA}$) used the GWAS summary statistics generated from individual-level data from resident SSA populations. This included AWI-Agincourt, AWI- Dikgale, AWI-Nanoro, AWI-Nairobi, AWI-Navrongo, AWI-Soweto and ARK-Agincourt. The second meta-analysis ($Meta_{NONRES}$) included data from individuals of African ancestry who are not residing in SSA. We used the GWAS summary statistics generated from individual-level data from the UK-Biobank (UKB-African and UKB-Caribbean) and CKD-Gen African American sub-set (CKD-Gen-AA). The third meta-analysis ($Meta_{ALL}$) consisted of a meta-analysis that pooled the summary statistics of all studies from AWI-Agincourt, AWI- Dikgale, AWI-Nanoro, AWI-Nairobi, AWI-Navrongo, AWI-Soweto, ARK-Agincourt, UKB-African, UKB-Caribbean and CKD-Gen-AA. **Figure 1** outlines the meta-analysis workflow. The genome-wide significance level for novel discovery was considered at $P<5x10^{-08}$.

Post association analysis

Plotting

QQ-plots and Manhattan plots were generated using the FastMan library (Paria et al., 2022) (available at https://github.com/kaustubhad/fastman) and the Hudson library (available at https://github.com/anastasia-lucas/hudson). These visualizations were created using SNPs with a MAF threshold of 0.01 or more. For regional plots, we utilized the standalone version of the LocusZoom software (Pruim et al., 2010).

Genetic LD reference

For the estimation of the LD reference panel for conditional and joint (COJO) analysis, clumping, and fine-mapping, three LD reference panels were constructed using genotype data from the appropriate datasets. For resident SSA dataset comparisons, the LD reference panel (LD_{SSA}) was constructed using AWI-Gen and ARK individual-level genotype data. For non-resident SSA dataset comparisons, the LD reference panel (LD_{NONRES}) was constructed using UKB individual-level genotype data. For the combined datasets comparison, the LD reference panel (LD_{ALL}) was constructed using AWI-Gen, ARK, and UKB individual-level genotype data.

Fine-mapping and lead SNPs

For each locus with a lead SNP with a P-value below $5x10^{-08}$, fine-mapping was conducted using the H3ABioNet H3AGWAS pipeline and implementing a stepwise model selection procedure through GCTA (Yang et al., 2011, 2012; Brandenburg et al., 2022a) to identify independently associated SNPs. Subsequently, we utilized the FINEMAP software (Version 1.4) (Benner et al., 2016), considering one causal variant, to define the credible set with 99% confidence using a stochastic approach (Benner et al., 2016).

Conditional analyses (GCTA)

Conditional analyses used the GCTA software implemented within the H3AGWAS pipeline, with summary statistics obtained from the meta-analyses as input. In these analyses, the lead SNPs identified in each meta-analysis were conditioned upon lead SNPs found in previously published studies. Changes in the p-value, both increasing or decreasing significance, of the lead SNP, confirmed a relationship between the two SNPs.

Replication of previous findings

Replication was performed according to the following criteria: 1) Exact replication: if any genome-wide significant lead SNPs found in CKD-Gen-EA and CKD-Gen-MA reached statistical significance ($p<0.05$) in $Meta_{SSA}$, $Meta_{NONRES}$ or $Meta_{ALL}$ after Bonferroni correction (A total of 60 independent lead SNPs were identified in the CKD-Gen datasets, of which 55 lead SNPs were from CKD-Gen-EA and 57 lead SNPs were from CKD-Gen-MA) and that the lead SNPs have same direction of effect. 2) LD Window replication: for a given genome-wide significant SNP found in the CKD-Gen datasets, SNPs had been extracted from $Meta_{SSA}$, $Meta_{NONRES}$ and $Meta_{ALL}$ that are in LD with said CKD-Gen lead SNP. LD was pruned using the clump procedure in PLINK (Version 1.9) (r2 = 0.1, windows size 1000 kb, $P_1 = 5x10^{-08}$, $P_2 = 0.1$). The lowest p-value(s) from SNPs within the given LD window were extracted and this LD window was considered statistically significant if the p-value was less than $5x10^{-04}$ in both datasets. Additionally, the direction of effect between the CKD-Gen and Meta-datasets ($Meta_{SSA}$, $Meta_{NONRES}$ and $Meta_{ALL}$) must be consistent. Conditional analyses were performed between the genome-wide significant SNP(s) in CKD-Gen and lead SNP in our meta-analyses to confirm the replication.

For replication, the findings from $Meta_{SSA}$ were compared to CKD-Gen-MA and CKD-Gen-EA, and the findings from $Meta_{NONRES}$ and $Meta_{ALL}$ were only compared to CKD-Gen-EA to

avoid sample overlaps within the CKD-Gen datasets (as CKD-Gen-AA is contained within CKD-Gen-MA).

Annotation and Expression Quantitative Trait Locus (eQTL) analysis

Functional annotation of genome-wide significant SNPs found in $Meta_{SSA}$, $Meta_{NONRES}$ and/or $Meta_{ALL}$ was done using the ANNOVAR software (Wang et al., 2010).
eQTL analysis was performed using the database of cis-eQTLs in both glomerular and tubulointerstitial tissues, derived from participants in the Nephrotic Syndrome Study Network (NEPTUNE) using SNPs with false discovery rate (FDR) < 0.05 (Han et al. 2023), a window of 100,000 bp had been defined around a genome-wide significant locus to identify close eQTLs.

Polygenic Scores

PGS were computed for each dataset independently (AWI-Agincourt, AWI- Dikgale, AWI-Nanoro, AWI-Nairobi, AWI-Navrongo, AWI-Soweto ARK-Agincourt, UKB-African and UKB-Caribbean). The effect sizes from 3 previous studies were used: CKD-Gen-AA, CKD-Gen-MA, and CKD-Gen-EA. PRS-CS (Ge et al., 2019), software that estimates posterior SNP effect sizes by implementing continuous shrinkage (CS) priors, was used to calculate the PGS. As external LD references are required for this analysis, the African LD data derived from the 1000 Genomes Project was leveraged for this purpose (accessible at https://github.com/getian107/PRScs). The PGS values were regressed against the residualized UACR value in a linear regression model that adjusted for age, sex, and the first five principal components to assess the performance of PGS.

Results

Study participants and phenotype data

Genomic and phenotypic data were accessible for 7,959 individuals in the AWI-Gen datasets, 1,011 individuals in the ARK dataset, and 2,916 individuals in the UK-Biobank dataset with 1,205 individuals and 1,711 individuals in UKB-African and UKB-Caribbean respectively (**Supplementary Figure S1**). CKD-Gen AA was a meta-analysis of 7 studies including 6,795 individuals in total. Overall, there was a higher prevalence of albuminuria (17.9%; median UACR 1.01mg/mmol) among individuals from the UKB with African and Caribbean ancestry compared to individuals residing in SSA, where notable regional differences were

observed. The highest prevalence of albuminuria occurred in AWI-Agincourt, South Africa (14.1%; median UACR 0.59mg/mmol) while the lowest prevalence occurred in AWI-Nanoro, West Africa (prevalence of albuminuria 4.5%, median UACR 0.35 mg/mmol) (**Table 1**).

Meta-analysis

Meta-analyses were conducted to investigate the genetics of UACR in resident Sub-Saharan African datasets ($Meta_{SSA}$) (**Figure 2a**), non-resident Sub-Saharan African datasets ($Meta_{NONRES}$) (**Figure 2b**) and all African ancestry datasets ($Meta_{ALL}$) (**Figure 2c**).

No genomic inflation was observed for the individual-dataset association testing performed on the 9 datasets. All genomic inflation factors (lambda) were below 1.1. This was visually confirmed on the dataset specific QQ-plots and Manhattan plots (**Supplementary Figure S2 and Figure S3(a-i)**). Dataset-specific significant findings are reported in **Supplementary Table 2** and **Supplementary Figure S4(a-c)**.

One genome-wide significant locus with the lead SNP rs9505286 ($p=4.3.10^{-08}$) was identified in $Meta_{SSA}$ on chromosome 6. (Figure 3a). One genome wide significant locus with the lead SNP rs73404549 was identified on chromosome 11 in $Meta_{NONRES}$ ($p=5.6.10^{-11}$) and $Meta_{ALL}$ ($p=7.7\ 10^{-13}$) (**Table 2, Figure 3b and 3c)**)

SNP rs9505286 (chr6:7820353) is located in the intronic regions of *BMP6*. Two SNPs were identified in the 95% credible set using FINEMAP (**Figure 3a, Supplementary Table 4**). An eQTL in the region was found to be associated with the expression of several genes notably *BLOC1S5 and TXNDC5* (**Table 2; Supplementary Figure S5; Supplementary Table 3, Supplementary Table 5**).

SNP rs73404549 (chr11:5320654) is located near the *HBE1*, *OR51B4*, and *HBB* genes. This signal is primarily driven by results from West African ancestry datasets in the $Meta_{NONRES}$ and $Meta_{ALL}$ (**Supplementary Figure S6**). Notably, this SNP is monomorphic in the Southern African and East African datasets. Furthermore, rs73404549 is in LD with rs334 ($r^2=0.52$; 72,422 bp apart), the SNP that defines the sickle cell mutation (HbS). SNP rs334 was also significant in $Meta_{ALL}$ ($P_{ALL}=8.55x10^{-9}$).

In the window of 100,000 bp around rs73404549, 14 SNPs are significant eQTLs in : glomerular and tubulointerstitial tissues in the region were found to be associated with the expression of the *HBG*, *HBG2*, *NM_000184*, *TRIM21*, *OR51B5* and *TRIM68* genes (**Supplementary** Table 5). Seven and two SNPs were identified in the 95 % credible set

using FINEMAP in $Meta_{NONRES}$ and $Meta_{ALL}$ results respectively (**Supplementary Table 3; Supplementary Figure S6**).

Replication of previous findings

Replication analysis confirmed associations in three previously identified regions in *THBS3*, *SPATA5L1/GATM*, and *ARL15* (**Supplementary Table 3**).

In the *THBS3* region, the $Meta_{ALL}$ meta-analysis rs370545 was the lead SNP in our dataset, with a P-value of $1x10^{-04}$. However, a conditional analysis using rs2974937 (lead SNP in CKD-Gen-EA) resulted in a decrease in significance level ($P_{conditional_analysis}$=0.85). This suggests that the association in the *THBS3* region was driven by rs2974937 in $Meta_{ALL}$ even though it was not the lead SNP in this region (**Supplementary Table 4; Supplementary Figure 7**).

In the *ARL15* region, a statistically significant association signal was observed in $Meta_{SSA}$ (rs1664781, P=$1.8x10^{-04}$). Conditional analysis using rs1694068 (lead SNP in CDK-Gen-EA) revealed a reduction in P-value for rs1664781 ($P_{conditional_analysis}$=0.87), suggesting that rs1694068 and rs1664781 are in LD thus confirming the association in this region (**Supplementary Table 4, Supplementary Figure 8**).

In the *SPATA5L1/GATM* region, the $Meta_{ALL}$ meta-analysis identified rs1694067 as the lead SNP in this region with a P-value of $7.0x10^{-05}$. Furthermore, the lead SNP rs1153847 identified in CKD-Gen-EA, rs1153847, was present in our dataset, and its association was replicated ($P_{Bonferoni_adjusted}$=0.04). For the window-based replication, a conditional analysis using rs2467858 (genome-wide significant SNP in CKD-Gen-EA), a reduced P-value was observed ($P_{conditional_analysis}$=0.87) confirming rs1694067 and rs2467858 are in LD and replicated the CKD-Gen signal. (**Supplementary Table 4; Supplementary Figure 9**).

Polygenic score analyses

The Variance explained by the PGS for UACR residuals was between 0% and 0.82%. PGS constructed using the betas from CKD-Gen-EA and CKD-Gen-MA performed better for the non-SSA resident datasets, particularly in the UKB-African, showing the best predictivity (%variance: 0.82, p=$1x10^{-04}$) and statistically significant correlation between the PGS and the UACR residual (**Figure 4, Supplementary Table S6**).

Using the PGS constructed from CKD-Gen-MA, ARK-Agincourt (% variance: 0.61, P=0.01) and AWI-Agincourt (%variance: 0.58, P=0.002) demonstrated better predictivity in SSA populations. PGS constructed from CKD-Gen-AA did not improve the variance explained.

Variance explained was lower using PGS constructed from CKD-GEN-AA than CKD-Gen-MA or CKD-Gen-EA.

Discussion

This study is the first GWAS for UACR conducted in Sub-Saharan African populations. Two genomic regions were identified to be associated with UACR among 8,970 participants from West, East, and Southern Africa and among 9,705 non-resident African-ancestry participants from the UK Biobank and CKD-Gen Consortium.

In the first locus, the SNP rs9505286 reached genome-wide significance in Sub-Saharan African individuals ($Meta_{SSA}$). rs9505286 is located in the intronic region of *BMP6* and eQTLs found in a 100,000 bp window where shown to affect the expression of *TXNDC5*. Both these genes have been linked with kidney function. *BMP6* has been shown to be a potential major regulator of kidney fibrosis, the pathological hallmark of irreversible loss of kidney function in CKD (Jenkins and Fraser, 2011). Genes from the BMP gene family play diverse roles in embryonic kidney development, regulating essential aspects of both ureteric bud and nephron development (Oxburgh et al., 2011). eQTLs around lead SNPs show potential interaction with *TXNDC5*, a gene that promotes renal fibrosis by promoting TGF-β signalling in kidney fibroblasts (Chen et al., 2021; Hung et al., 2022). Variability of kidney function, confounding factors and allele frequency differences between datasets may explain why the rs9505286 signal was not replicated in $Meta_{ALL}$ or $Meta_{NONRES}$ (Marigorta et al., 2018).

In the second locus, the SNP rs73404549 was found to be statistically significant in non-resident individuals with African Ancestry ($Meta_{NONRES}$) and overall ($Meta_{ALL}$), but not in Sub-Saharan African individuals ($Meta_{SSA}$). This can be explained by the fact that the variant allele of rs73404549 is extremely rare or absent in East and South African populations. SNP rs73404549 was found to be in linkage disequilibrium with rs334, the sickle cell mutation (HbS) in the *HBB* gene. Notably, rs334 had been associated with various kidney traits, including UACR and CKD in African, African American and US Hispanic/Latino populations (Naik et al., 2014; Gurdasani et al., 2019; Fatumo et al., 2020). The HbS mutation has been linked to malaria resistance among heterozygotes, with differences in allele frequency attributed to variations in selection pressures between Bantu-speaking populations in West and South/East Africa (Gurdasani et al., 2019; Choudhury et al., 2020).

In addition to the *HBB* region, our GWAS revealed transferability of three previously identified signals. Of the 60 UACR-associated loci identified in European and Multi-Ancestry studies, only three were replicated, including variants in *GATM*. This region was also associated with eGFR in a Ugandan population (Fatumo et al., 2020). Furthermore, we replicated the association with *ARL15* in the region of chromosome 1. *ARL15* is a regulator of Mg2+ transport thereby promoting the complex N-glycosylation of cyclin M proteins (CNNM 1-4) and could play a role in the pathogenesis of hypertension mediated via altered tubular handling of magnesium in the kidney (Zolotarov et al., 2021).

The transferability of PGS developed using the effect sizes quantified in three previous association studies in European ancestry, African ancestry and multi-ancestry populations showed limited predictability, explaining less than 1% of the variability in UACR. PGS in resident African populations (AWI-Gen and ARK) explained between 0.58% and 0.60% of the variance of UACR compared to UKB-African, where best prediction was observed (0.80%). The poor predictability of UACR using summary statistics derived from African Americans was likely due to the small sample size of the discovery dataset. Unfortunately, there have been few studies on PGS approaches to compare findings with, and the genetic heritability of UACR is relatively low, estimated at 4.3% (Teumer et al., 2019).

The limited transferability of PGS and previous GWAS signals across ancestral groups could be due to differences in genetic architecture and/or pleiotropic effects. Different demographic histories and genetic selection pressures between European and African populations could modify the ability to replicate previous GWAS results due to differences in allele frequencies between non-African and African populations, with generally lower LDs in African genomes. Environmental factors and variability in the prevalence and aetiology of kidney-function-related risk factors such as diabetes and hypertension (Fatumo et al., 2020) could also influence the genetic architecture of kidney disease in Africans populations (Limou et al., 2014; Teumer et al., 2019; Brandenburg et al., 2022b).

Selection pressures have increased the frequencies *APOL1* kidney risk variants and HbS due to their protective properties in areas of Africa where trypanosomiasis and malaria are endemic. This may have contributed to shaping genetic susceptibility to kidney disease in African individuals. In our study, the *APOL1* gene region did not exhibit significant associations with UACR. The indel rs71785313 was not imputed using the African Sanger reference for imputation, and a specific study had previously been published to describe *APOL1* variant distribution in the AWI-Gen dataset using other imputation panels, but the

locus did not reach genomic significance ($5e^{-8}$) for association with eGFR and uACR (Brandenburg et al., 2022b).

While the burden of CKD in SSA is high, it is noteworthy that no prior GWAS on UACR has been conducted on the continent. Despite its uniqueness, our study is limited by its relatively modest sample size, which impacts statistical power to detect small-effect associations reaching genome-wide significance thresholds. Kidney function markers were measured at a single time point, and spot urine albumin and creatine levels are sensitive to incident infections and other environmental factors that could affect the prevalence of albuminuria.

It is important to note that our study populations are mainly treatment naïve in relation to kidney disease and other cardiometabolic conditions, which may be an advantage in detecting genetic associations (Pereira et al., 2021). Other studies, based on lipid-associated loci, attributed non-transferability of associated loci to pleiotropic effects, gene-environment interactions, and also to variability in allele frequencies and LD patterns (Kuchenbaecker et al., 2019; Choudhury et al., 2022), as we hypothesize for UACR.

In conclusion, this study describes genetic associations with UACR in a unique SSA cohort and non-resident individuals with African ancestry. CKD in African populations remains understudied but from available data, hypertension, rather than diabetes is the most commonly associated risk factor and in some regions, up to 60% of people with CKD do not have an associated “traditional” risk factor common to high-income settings, suggesting alternate underlying molecular pathways or aetiologies for CKD (Kalyesubula et al., 2018; Nakanga et al., 2019; Muiru et al., 2020). Our study identified two novel SNPs associated with UACR in populations of African ancestry. We further replicated three known UACR-associated loci. Regional genetic diversity due to different selection pressures appear to play a role in the genetic aetiology of CKD across the African continent. These factors likely contribute to the limited transferability of previous association signals and the poor transfer of polygenic scores developed in non-African populations to African populations. Larger genomic studies are necessary to better understand the genetic architecture of kidney function and chronic kidney disease across African populations.

Disclosure statement

The authors have nothing to disclose.

Ethics Statement

This study received approval from the Human Research Ethics Committee (Medical), University of the Witwatersrand/South Africa (M121029; M170880; M160939), the approval of the Centre Muraz Institutional Ethics Committee/Burkina Faso (015-2014/CE- CM) and the approval of the National Ethics Committee For Health Research/Burkina Faso (2014-08-096), the Ghana Health Service Ethics Review Committee (ID No: GHS-ERC:05/05/2015), the AMREF Health Ethics and Scientific Review Committee in Kenya (approval no. P114/2014) and the Navrongo Institutional Review Board (ID No: NHRCIRB178). All the participants signed an Informed Consent Form before any study procedure was performed.

Conflict of Interest

None

Acknowledgement

We acknowledge the sterling contributions of our field workers, phlebotomists, laboratory scientists, administrators, data personnel, and all other staff who contributed to the AWI-Gen and ARK data and sample collection, processing, storage, and shipping, and the participants who generously agreed to be part of the studies. The UK Biobank kindly shared data under the UK Biobank Resource Application Number 63215 and approved submission of the paper.

Funding

This study was funded by the National Institutes of Health (NIH) through the H3Africa AWI-Gen project (NIH grant number U54HG006938)). AWI-Gen is supported by the National Human Genome Research Institute (NHGRI), Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD), and Office of the Director (OD) at the National Institutes of Health. PB is funded by the National Research Foundation/The World Academy of Sciences “African Renaissance Doctoral Fellowship” (Grant no. 100004). JTB is supported by grants from the National Human Genome Research Institute (U54HG006938) as part of the H3A Consortium (AWI-Gen); and the Science For African Foundation - REACCT-CAN Grant (Del-22-008). MR is the South African Research Chair in Genomics and Bioinformatics of African populations hosted by the University of the Witwatersrand (SARChI), funded by the South African Department of Science and Innovation, and administered by the National Research Foundation. WCC is supported by the Research Networks for Health Innovations in sub-Saharan Africa Funding Initiative of the German Federal Ministry of Education and Research (RHISSA) Grant and the Science For African Foundation - REACCT-CAN Grant (Del-22-008). The ARK study was jointly funded by the South African MRC, MRC UK (via the Newton Fund), and GSK Africa Non-Communicable Disease Open Lab (via a supporting grant; project number 074). The views expressed in this manuscript do not necessarily reflect the views of the funders.

Author Contributions

JTB and MR designed the study, JTB performed the analyses and wrote the first draft. WCC, MR, AM, SH and JF made significant contributions to guiding the analyses and interpreting the results, as well as writing and reviewing the manuscript. All authors read, edited and approved the manuscript.

Data availability statement

All data that support the findings of this study are available from the corresponding authors on reasonable request. The AWI-Gen data set is available from the European Genome-phenome Archive (EGA) database (https://ega-archive.org/), accession number EGAS00001002482 (phenotype dataset: EGAD00001006425; genotype dataset: EGAD00010001996). The availability of these datasets is subject to controlled access through the Data and Biospecimen Access Committee of the H3Africa Consortium. ARK data is available in WIReDSpace repository on request (https://wiredspace.wits.ac.za/), https://doi.org/10.54223/uniwitwatersrand-10539-33016 35

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication). The processed data generated in this study are provided in Supplementary Material. Permission was obtained to access the genotype and phenotype dataset for UKBB (research project number: 63215).

Publicly available databases include Teumer et al. in CKD-GEN consortium website: https://ckdgen.imbi.uni-freiburg.de/.

Tables

Table 1: Study participants and phenotype data. Participant characteristics for each AWI-Gen study site, ARK-Agincourt and UKB-African and UKB-Caribbean, with phenotype distributions of UACR (median) and covariables used in the study.

Dataset	Sample size	Country of residence	[1]Age (years)	Males (%)	[2]UACR	[3]Albuminuria (%)
Non-resident in [4]SSA	**9,705**					
[5]CKD-Gen-AA	6,795	USA				
[6]UKB-African	1,205	UK	51.8	48.5	1.10 (1.70)	20.0
[7]UKB-Caribbean	1,711	UK	52.8	59.6	0.93 (1.44)	16.4
[8]UKB-All	2,916	UK	52.4	55.0	1.01 (1.55)	17.9
Resident in SSA	**8,970**					
[9]AWI-Nanoro	1,702	Burkina Faso	49.6	50.8	0.35 (0.44)	4.9
[10]AWI-Navrongo	1,548	Ghana	51.1	54.4	0.41(0.48)	6.7
[11]AWI-Nairobi	1,481	Kenya	48.8	54.0	0.63 (0.67)	11.4
[12]AWI-Agincourt	1,545	South Africa	54.3	59.8	0.59 (0.99)	14.1
[13]AWI-Dikgale	917	South Africa	52.1	68.3	0.63 (0.51)	10.4

[14]AWI-Soweto	766	South Africa	49.5	100.0	0.36 (0.57)	11.4
AWI-All	7,959		50.9	51.0	0.48 (0.51)	9.5
[15]ARK-Agincourt	1,011	South Africa	38.8	58.0	0.57 (0.94)	13.4
Total	**17,664**		**50.3**	**52.6**	**0.59 (0.85)**	**11.9**

[1]data reported as the mean.
[2]UACR: urine albumin: creatinine ratio: mg/mmol; reported as median (interquartile range).
[3]Albuminuria: UACR>3.0mg/mmol.
[4]SSA: Sub-Saharan Africa.
[5]CKD-Gen Consortium: African American Ancestry individuals. CKD-GEN-AA data are summary statistics of meta-analysis from Teumer et al. downloadable at the CKD-GEN consortium website (http://ckdgen.imbi.uni-freiburg.de/), information relative to samples had been reported in relevant papers.
[6]UKB Biobank: Individuals of self-reported African Ancestry; [7]self-reported Caribbean Ancestry; [8]self-reported African and Caribbean Ancestry.
[9]Africa Wits -INDEPTH partnership for Genomic Studies (AWI): individuals of African Ancestry from [9,10]West Africa; [11]East Africa; [12-14]South Africa.
[15]African Research on Kidney Disease (ARK) Study: South Africa.

Table 2: Lead genome-wide significantly associated SNPs for sub-Saharan African population meta-analysis ($Meta_{SSA}$), non-resident African ancestry population meta-analysis ($Meta_{NONRES}$) and the combined African ancestry population meta-analysis ($Meta_{ALL}$).

						$Meta_{SSA}$		**$Meta_{NONRES}$**		**$Meta_{ALL}$**	
Chr:bp	**rsID**	**Genes**	**eQTL (genes)**	**EA**	**NEA**	**P (beta, se)**	**EAF**	**P (beta, se)**	**EAF**	**P (beta, se)**	**EAF**
6:7820353	rs9505286	intronic: BMP6	*BLOC1S5, TXNDC5, ENSG00000223342. DSP-AS1*	T	C	4.3e-08 (-0.19,0.03)	0.95	0.17 (-0.05,0.07)	0.96	7.2e-07 (-0.13,0.03)	0.96
11:5320654	rs73404549	intergenic : HBE1, OR51B4	*HBG, HBG2, NM_000184, TRIM21, OR51B5, TRIM68*	T	C	3.2e-03 (0.17,0.06)	0.05	**5.6e-11 (0.21,0.06)**	0.05	**7.7e-13 (0.2,0.04)**	0.05

Chr: bp: chromosome-coordinate (hg19)
rsID : SNPs identifiant
Genes : genes annotation using ANNOVAR
eQTL : gene's transcript had been found significant associated with positions around (100kb)
EA : Effect allele
NEA : non effect allele
P (beta, se) : result of association : P / p-value, beta : effect, se : standard error
EAF : Effect allele frequency

Figure Legends

Figure 1: Study design showing data sources, the analysis strategy and post-GWAS analysis approach.

Figure 2: Manhattan plot - GWAS of UACR in the (a) $Meta_{SSA}$ (b) $Meta_{NONRES}$ (c) $Meta_{ALL}$ datasets. Lead genome-wide significant SNPs ($P<5x10^{-08}$) and gene annotations are highlighted.

Figure 3: Regional plot using LocusZoom of genome wide significant SNPs found in meta-analyses, (a) rs9505286 from the result of $Meta_{SSA}$, (b) rs9966824 from the result $Meta_{NONRES}$ (c) rs9966824 from the result $Meta_{ALL.}$

Figure 4. Percent variance explained between PGS and residual phenotypes computed using age, sex and 5 PCs. Key: The negative relationship between PGS and the phenotype in the result of the linear model, * P<0.05, ** P<0.01 and ** P<0.001. Details in Supplementary Table 6

References

Adam, Y., Sadeeq, S., Kumuthini, J., Ajayi, O., Wells, G., Solomon, R., et al. (2022). Polygenic Risk Score in African populations: progress and challenges. *F1000Res* 11, 175. doi: 10.12688/f1000research.76218.2.

Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655–1664. doi: 10.1101/gr.094052.109.

Ali, S. A., Soo, C., Agongo, G., Alberts, M., Amenga-Etego, L., Boua, R. P., et al. (2018). Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Global Health Action* 11, 1507133. doi: 10.1080/16549716.2018.1507133.

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. doi: 10.1038/nature15393.

Baichoo, S., Souilmi, Y., Panji, S., Botha, G., Meintjes, A., Hazelhurst, S., et al. (2018). Developing reproducible bioinformatics analysis workflows for heterogeneous computing environments to support African genomics. *BMC Bioinformatics* 19, 457. doi: 10.1186/s12859-018-2446-1.

Benner, C., Spencer, C. C. A., Havulinna, A. S., Salomaa, V., Ripatti, S., and Pirinen, M. (2016). FINEMAP: efficient variable selection using summary data from genome-

wide association studies. *Bioinformatics (Oxford, England)* 32, 1493–1501. doi: 10.1093/bioinformatics/btw018.

Böger, C. A., Chen, M.-H., Tin, A., Olden, M., Köttgen, A., de Boer, I. H., et al. (2011). CUBN is a gene locus for albuminuria. *Journal of the American Society of Nephrology□: JASN* 22, 555–570. doi: 10.1681/ASN.2010060598.

Brandenburg, J.-T., Clark, L., Botha, G., Panji, S., Baichoo, S., Fields, C., et al. (2022a). H3AGWAS: a portable workflow for genome wide association studies. *BMC Bioinformatics* 23, 498. doi: 10.1186/s12859-022-05034-w.

Brandenburg, J.-T., Govender, M. A., Winkler, C. A., Boua, P. R., Agongo, G., Fabian, J., et al. (2022b). Apolipoprotein L1 High-Risk Genotypes and Albuminuria in Sub-Saharan African Populations. *Clin J Am Soc Nephrol* 17, 798–808. doi: 10.2215/CJN.14321121.

Casanova, F., Tyrrell, J., Beaumont, R. N., Ji, Y., Jones, S. E., Hattersley, A. T., et al. (2019). A genome-wide association study implicates multiple mechanisms influencing raised urinary albumin–creatinine ratio. *Human Molecular Genetics* 28, 4197–4207. doi: 10.1093/hmg/ddz243.

Chen, Y.-T., Jhao, P.-Y., Hung, C.-T., Wu, Y.-F., Lin, S.-J., Chiang, W.-C., et al. (2021). Endoplasmic reticulum protein TXNDC5 promotes renal fibrosis by enforcing TGF-β signaling in kidney fibroblasts. *J Clin Invest* 131, 143645. doi: 10.1172/JCI143645.

Choudhury, A., Aron, S., Botigué, L. R., Sengupta, D., Botha, G., Bensellak, T., et al. (2020). High-depth African genomes inform human migration and health. *Nature* 586, 741–748. doi: 10.1038/s41586-020-2859-7.

Choudhury, A., Brandenburg, J.-T., Chikowore, T., Sengupta, D., Boua, P. R., Crowther, N. J., et al. (2022). Meta-analysis of sub-Saharan African studies provides insights into genetic architecture of lipid traits. *Nature Communications* 13, 2578. doi: 10.1038/s41467-022-30098-w.

Craik, A., Gondwe, M., Mayindi, N., Chipungu, S., Khoza, B., G□mez-Oliv□, X., et al. (2023). Forgotten but not gone in rural South Africa: Urinary schistosomiasis and implications for chronic kidney disease screening in endemic countries [version 3;

peer review: 2 approved]. *Wellcome Open Research* 8. doi: 10.12688/wellcomeopenres.18650.3.

Fabian, J., Gondwe, M., Mayindi, N., Chipungu, S., Khoza, B., Gaylard, P., et al. (2022). Chronic kidney disease (CKD) and associated risk in rural South Africa: a population-based cohort study [version 2; peer review: 2 approved]. *Wellcome Open Research* 7. doi: 10.12688/wellcomeopenres.18016.2.

Fatumo, S., Chikowore, T., Kalyesubula, R., Nsubuga, R., Asiki, G., Nashiru, O., et al. (2020). Discovery and fine-mapping of kidney function loci in first genome-wide association study in Africans. doi: 10.1101/2020.06.09.142463.

GBD Chronic Kidney Disease Collaboration, B., Purcell, C. A., Levey, A. S., Smith, M., Abdoli, A., Abebe, M., et al. (2020). Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet (London, England)* 395, 709–733. doi: 10.1016/S0140-6736(20)30045-3.

Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A., and Smoller, J. W. (2019). Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature Communications* 10, 1776. doi: 10.1038/s41467-019-09718-5.

George, J. A., Brandenburg, J.-T., Fabian, J., Crowther, N. J., Agongo, G., Alberts, M., et al. (2019). Kidney damage and associated risk factors in rural and urban sub-Saharan Africa (AWI-Gen): a cross-sectional population study. *The Lancet Global Health* 7, e1632–e1643. doi: 10.1016/S2214-109X(19)30443-7.

Gomez, F., Hirbo, J., and Tishkoff, S. A. (2014). Genetic variation and adaptation in Africa: implications for human evolution and disease. *Cold Spring Harbor perspectives in biology* 6, a008524–a008524. doi: 10.1101/cshperspect.a008524.

Gorski, M., van der Most, P. J., Teumer, A., Chu, A. Y., Li, M., Mijatovic, V., et al. (2017). 1000 Genomes-based meta-analysis identifies 10 novel loci for kidney function. *Scientific reports* 7, 45040. doi: 10.1038/srep45040.

Gurdasani, D., Carstensen, T., Fatumo, S., Chen, G., Franklin, C. S., Prado-Martinez, J., et al. (2019). Uganda Genome Resource Enables Insights into Population History and

Genomic Discovery in Africa. *Cell* 179, 984-1002.e36. doi: 10.1016/j.cell.2019.10.004.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., et al. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature* 517, 327–332. doi: 10.1038/nature13997.

Haas, M. E., Aragam, K. G., Emdin, C. A., Bick, A. G., Hemani, G., Davey Smith, G., et al. (2018). Genetic Association of Albuminuria with Cardiometabolic Disease and Blood Pressure. *American Journal of Human Genetics* 103, 461–473. doi: 10.1016/j.ajhg.2018.08.004.

Han, B., and Eskin, E. (2011). Random-effects model aimed at discovering associations in meta-analysis of genome-wide association studies. *American journal of human genetics* 88, 586–598. doi: 10.1016/j.ajhg.2011.04.014.

Han, S. K., McNulty, M. T., Benway, C. J., Wen, P., Greenberg, A., Onuchic-Whitford, A. C., et al. (2023). Mapping genomic regulation of kidney disease and traits through high-resolution and interpretable eQTLs. *Nature Communications* 14, 2229. doi: 10.1038/s41467-023-37691-7.

Hellwege, J. N., Velez Edwards, D. R., Giri, A., Qiu, C., Park, J., Torstenson, E. S., et al. (2019). Mapping eGFR loci to the renal transcriptome and phenome in the VA Million Veteran Program. *Nature Communications* 10, 3842. doi: 10.1038/s41467-019-11704-w.

Hung, C.-T., Tsai, Y.-W., Wu, Y.-S., Yeh, C.-F., and Yang, K.-C. (2022). The novel role of ER protein TXNDC5 in the pathogenesis of organ fibrosis: mechanistic insights and therapeutic implications. *Journal of Biomedical Science* 29, 63. doi: 10.1186/s12929-022-00850-x.

Jenkins, R. H., and Fraser, D. J. (2011). BMP-6 emerges as a potential major regulator of fibrosis in the kidney. *Am J Pathol* 178, 964–965. doi: 10.1016/j.ajpath.2010.12.010.

Kachuri, L., Chatterjee, N., Hirbo, J., Schaid, D. J., Martin, I., Kullo, I. J., et al. (2023). Principles and methods for transferring polygenic risk scores across global populations. *Nature Reviews Genetics.* doi: 10.1038/s41576-023-00637-2.

Kalyesubula, R., Fabian, J., Nakanga, W., Newton, R., Ssebunnya, B., Prynn, J., et al. (2020). How to estimate glomerular filtration rate in sub-Saharan Africa: design and methods of the African Research into Kidney Diseases (ARK) study. *BMC Nephrology* 21, 20. doi: 10.1186/s12882-020-1688-0.

Kalyesubula, R., Hau, J. P., Asiki, G., Ssebunya, B., Kusemererwa, S., Seeley, J., et al. (2018). Impaired renal function in a rural Ugandan population cohort. *Wellcome Open Res* 3, 149. doi: 10.12688/wellcomeopenres.14863.3.

Kamiza, A. B., Toure, S. M., Vujkovic, M., Machipisa, T., Soremekun, O. S., Kintu, C., et al. (2022). Transferability of genetic risk scores in African populations. *Nature Medicine* 28, 1163–1166. doi: 10.1038/s41591-022-01835-x.

Kamiza, A. B., Touré, S. M., Zhou, F., Soremekun, O., Cissé, C., Wélé, M., et al. (2023). Multi-trait discovery and fine-mapping of lipid loci in 125,000 individuals of African ancestry. *Nature Communications* 14, 5403. doi: 10.1038/s41467-023-41271-0.

Kaze, A. D., Ilori, T., Jaar, B. G., and Echouffo-Tcheugui, J. B. (2018). Burden of chronic kidney disease on the African continent: a systematic review and meta-analysis. *BMC Nephrology* 19, 125. doi: 10.1186/s12882-018-0930-5.

Köttgen, A., and Pattaro, C. (2020). The CKDGen Consortium: ten years of insights into the genetic basis of kidney function. *Kidney International* 97, 236–242. doi: 10.1016/j.kint.2019.10.027.

Kovesdy, C. P. (2022). Epidemiology of chronic kidney disease: an update 2022. *Kidney Int Suppl (2011)* 12, 7–11. doi: 10.1016/j.kisu.2021.11.003.

Kuchenbaecker, K., Telkar, N., Reiker, T., Walters, R. G., Lin, K., Eriksson, A., et al. (2019). The transferability of lipid loci across African, Asian and European cohorts. *Nature Communications* 10. doi: 10.1038/s41467-019-12026-7.

Lee, J., Lee, Y., Park, B., Won, S., Han, J. S., and Heo, N. J. (2018). Genome-wide association analysis identifies multiple loci associated with kidney disease-related traits in Korean populations. *PLOS ONE* 13, e0194044.

Limou, S., Nelson, G. W., Kopp, J. B., and Winkler, C. A. (2014). APOL1 kidney risk alleles: population genetics and disease associations. *Advances in chronic kidney disease* 21, 426–433. doi: 10.1053/j.ackd.2014.06.005.

Lin, B. M., Nadkarni, G. N., Tao, R., Graff, M., Fornage, M., Buyske, S., et al. (2019). Genetics of Chronic Kidney Disease Stages Across Ancestries: The PAGE Study. *Frontiers in genetics* 10, 494. doi: 10.3389/fgene.2019.00494.

Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., A Reshef, Y., K Finucane, H., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature Genetics* 48, 1443–1448. doi: 10.1038/ng.3679.

Mahajan, A., Rodan, A. R., Le, T. H., Gaulton, K. J., Haessler, J., Stilp, A. M., et al. (2016). Trans-ethnic Fine Mapping Highlights Kidney-Function Genes Linked to Salt Sensitivity. *American journal of human genetics* 99, 636–646. doi: 10.1016/j.ajhg.2016.07.012.

Majara, L., Kalungi, A., Koen, N., Tsuo, K., Wang, Y., Gupta, R., et al. (2023). Low and differential polygenic score generalizability among African populations due largely to genetic diversity. *Human Genetics and Genomics Advances* 4, 100184. doi: 10.1016/j.xhgg.2023.100184.

Marigorta, U. M., Rodríguez, J. A., Gibson, G., and Navarro, A. (2018). Replicability and Prediction: Lessons and Challenges from GWAS. *Trends Genet* 34, 504–517. doi: 10.1016/j.tig.2018.03.005.

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., et al. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet* 100, 635–649. doi: 10.1016/j.ajhg.2017.03.004.

Morris, A. P., Le, T. H., Wu, H., Akbarov, A., van der Most, P. J., Hemani, G., et al. (2019). Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nature communications* 10, 29. doi: 10.1038/s41467-018-07867-7.

Muiru, A. N., Charlebois, E. D., Balzer, L. B., Kwarisiima, D., Elly, A., Black, D., et al. (2020). The epidemiology of chronic kidney disease (CKD) in rural East Africa: A population-based study. *PLoS One* 15, e0229649. doi: 10.1371/journal.pone.0229649.

Naik, R. P., Derebail, V. K., Grams, M. E., Franceschini, N., Auer, P. L., Peloso, G. M., et al. (2014). Association of Sickle Cell Trait With Chronic Kidney Disease and Albuminuria in African Americans. *JAMA* 312, 2115. doi: 10.1001/jama.2014.15063.

Nakanga, W. P., Prynn, J. E., Banda, L., Kalyesubula, R., Tomlinson, L. A., Nyirenda, M., et al. (2019). Prevalence of impaired renal function among rural and urban populations: findings of a cross-sectional study in Malawi. *Wellcome Open Res* 4, 92. doi: 10.12688/wellcomeopenres.15255.1.

Oxburgh, L., Brown, A. C., Fetting, J., and Hill, B. (2011). BMP signaling in the nephron progenitor niche. *Pediatr Nephrol* 26, 1491–1497. doi: 10.1007/s00467-011-1819-8.

Paria, S. S., Rahman, S. R., and Adhikari, K. (2022). fastman: A fast algorithm for visualizing GWAS results using Manhattan and Q-Q plots. *bioRxiv*. doi: 10.1101/2022.04.19.488738.

Pattaro, C., Köttgen, A., Teumer, A., Garnaas, M., Böger, C. A., Fuchsberger, C., et al. (2012). Genome-wide association and functional follow-up reveals new loci for kidney function. *PLoS genetics* 8, e1002584–e1002584. doi: 10.1371/journal.pgen.1002584.

Pattaro, C., Teumer, A., Gorski, M., Chu, A. Y., Li, M., Mijatovic, V., et al. (2016). Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nature communications* 7, 10023. doi: 10.1038/ncomms10023.

Pereira, L., Mutesa, L., Tindana, P., and Ramsay, M. (2021). African genetic diversity and adaptation inform a precision medicine agenda. *Nat Rev Genet* 22, 284–306. doi: 10.1038/s41576-020-00306-8.

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38, 904–909. doi: 10.1038/ng1847.

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., et al. (2010). LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 26, 2336–2337. doi: 10.1093/bioinformatics/btq419.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* 81, 559–575. doi: 10.1086/519795.

Ramsay, M., Crowther, N., Tambo, E., Agongo, G., Baloyi, V., Dikotope, S., et al. (2016). H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Global health, epidemiology and genomics* 1, e20–e20. doi: 10.1017/gheg.2016.17.

Teumer, A., Li, Y., Ghasemi, S., Prins, B. P., Wuttke, M., Hermle, T., et al. (2019). Genome-wide association meta-analyses and fine-mapping elucidate pathways influencing albuminuria. *Nature Communications* 10, 4130. doi: 10.1038/s41467-019-11576-0.

Teumer, A., Tin, A., Sorice, R., Gorski, M., Yeo, N. C., Chu, A. Y., et al. (2016). Genome-wide Association Studies Identify Genetic Loci Associated With Albuminuria in Diabetes. *Diabetes* 65, 803–817. doi: 10.2337/db15-1313.

Tin, A., and Köttgen, A. (2020). Genome-Wide Association Studies of CKD and Related Traits. *Clinical Journal of the American Society of Nephrology*, CJN.00020120. doi: 10.2215/CJN.00020120.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research* 38, e164–e164. doi: 10.1093/nar/gkq603.

Wuttke, M., Li, Y., Li, M., Sieber, K. B., Feitosa, M. F., Gorski, M., et al. (2019). A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nature Genetics* 51, 957–972. doi: 10.1038/s41588-019-0407-x.

Yang, J., Ferreira, T., Morris, A. P., Medland, S. E., Madden, P. A. F., Heath, A. C., et al. (2012). Conditional and joint multiple-SNP analysis of GWAS summary statistics

identifies additional variants influencing complex traits. *Nature Genetics* 44, 369–375. doi: 10.1038/ng.2213.

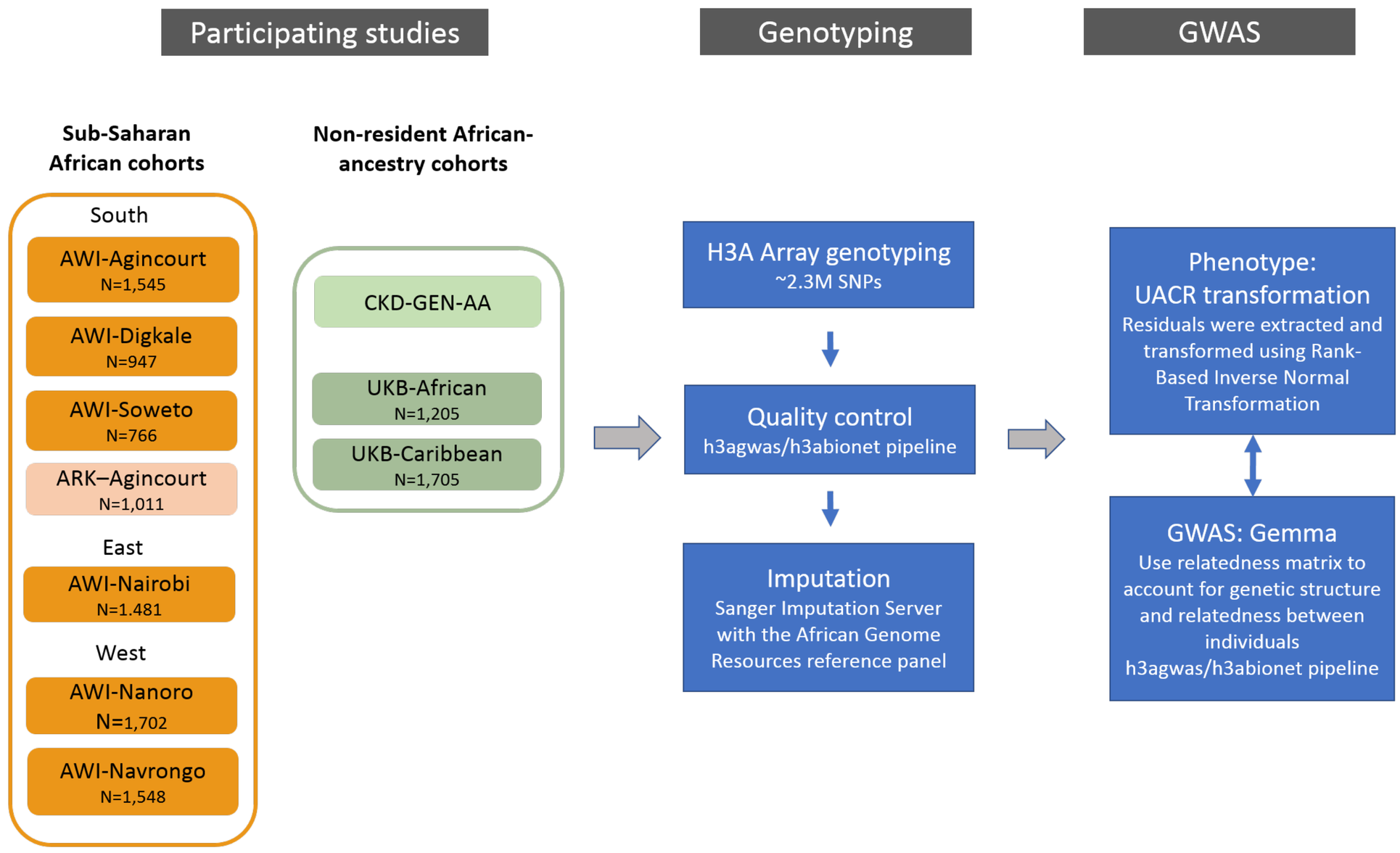
Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011.

Zanetti, D., Rao, A., Gustafsson, S., Assimes, T. L., Montgomery, S. B., and Ingelsson, E. (2019). Identification of 22 novel loci associated with urinary biomarkers of albumin, sodium, and potassium excretion. *Kidney International* 95, 1197–1208. doi: 10.1016/J.KINT.2018.12.017.

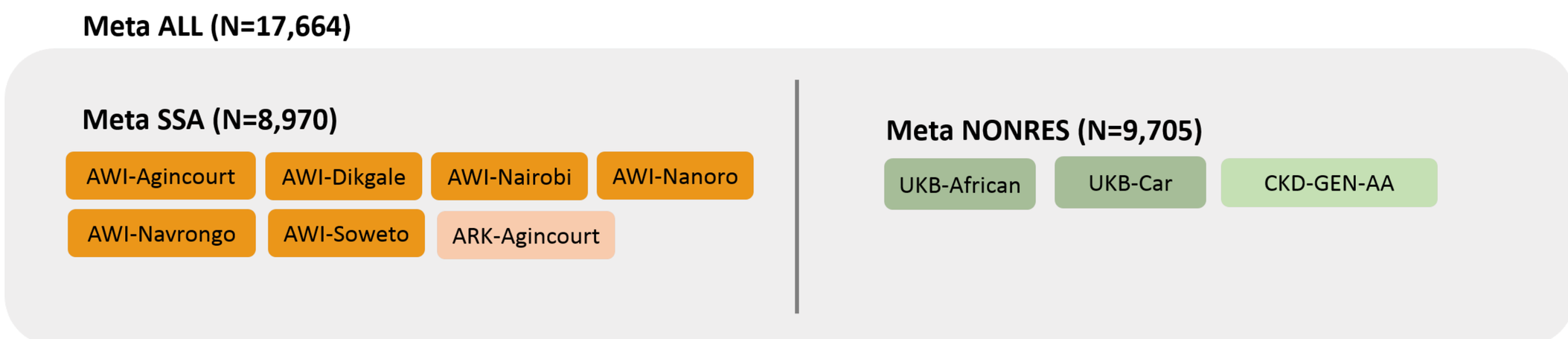
Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature genetics* 44, 821–824. doi: 10.1038/ng.2310.

Zolotarov, Y., Ma, C., González-Recio, I., Hardy, S., Franken, G. A. C., Uetani, N., et al. (2021). ARL15 modulates magnesium homeostasis through N-glycosylation of CNNMs. *Cell Mol Life Sci* 78, 5427–5445. doi: 10.1007/s00018-021-03832-8.

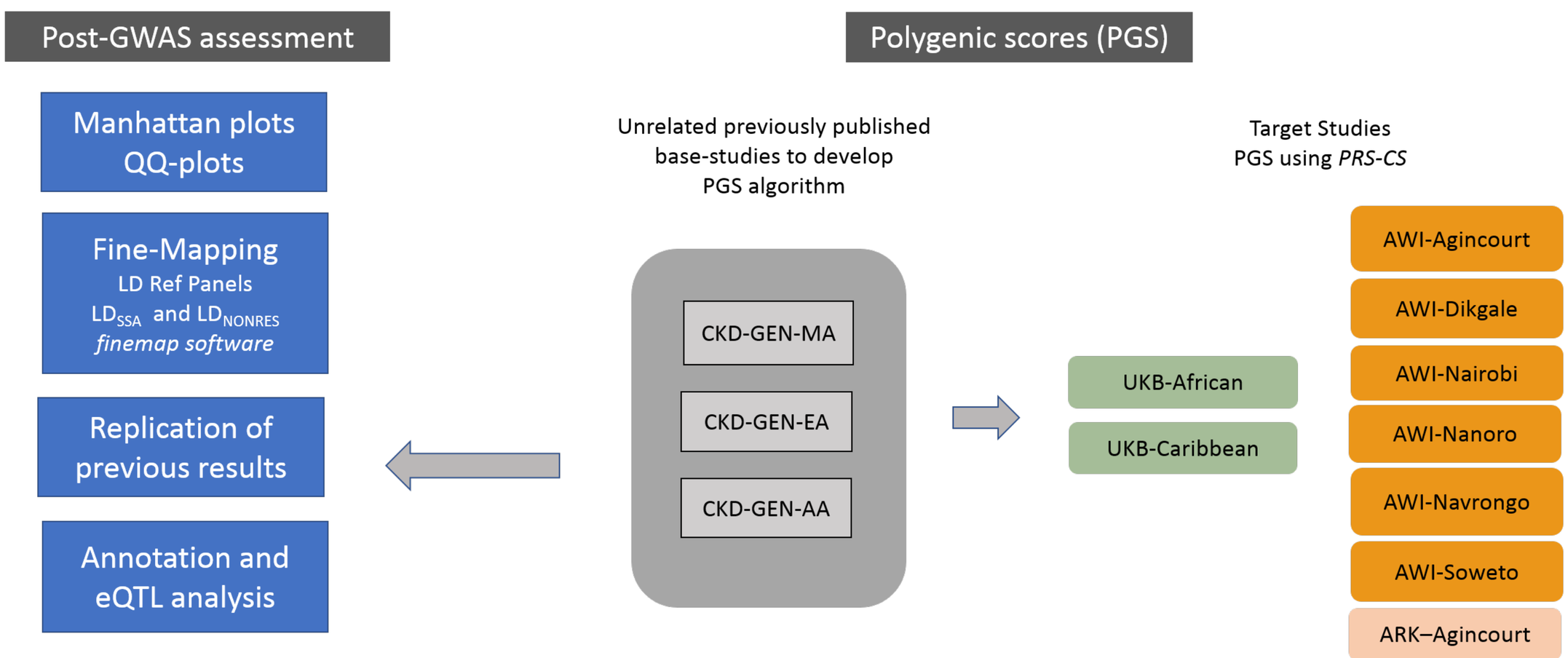
Studies – Genotyping - GWAS

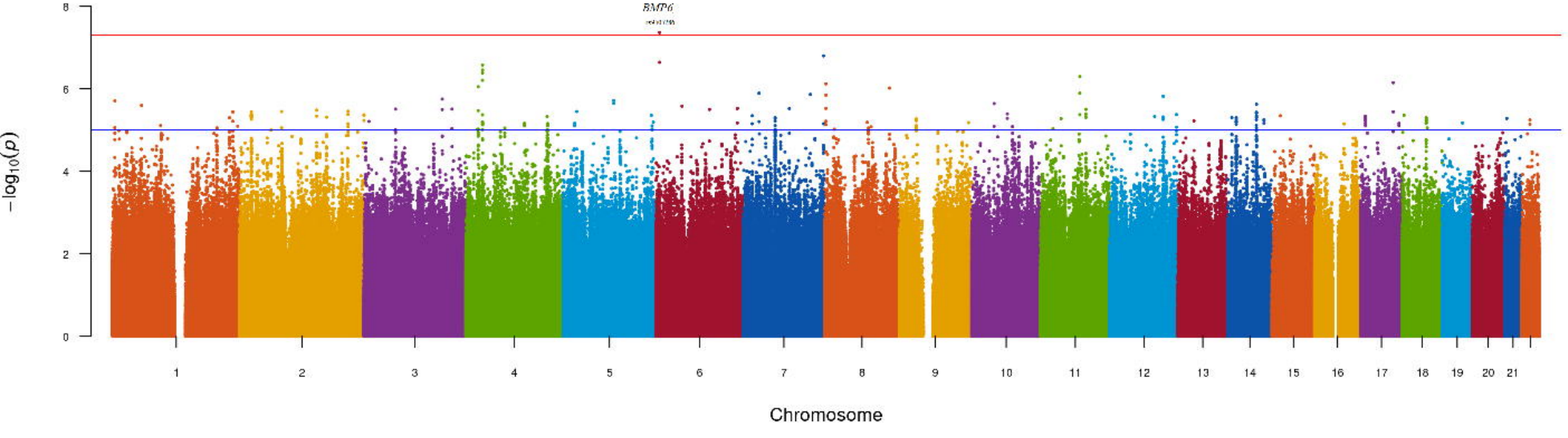


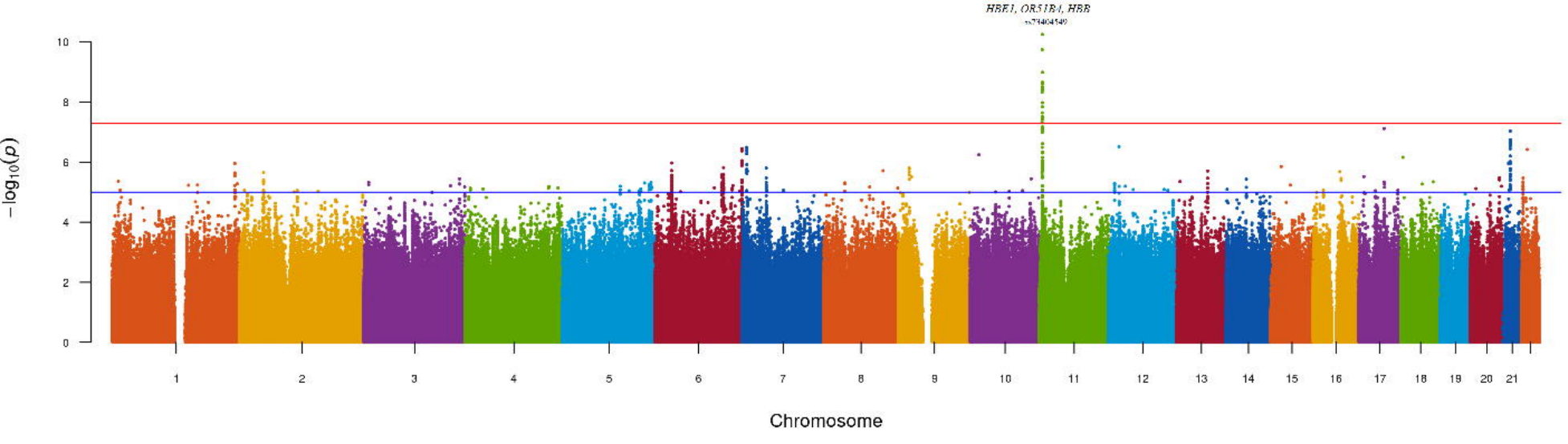
Meta-Analysis Strategy

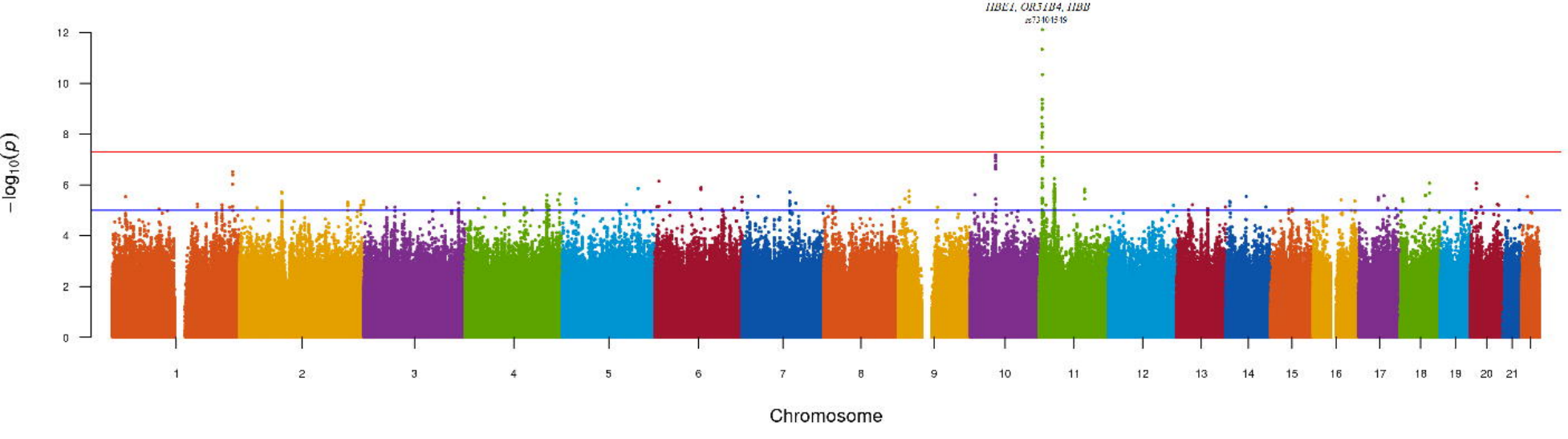


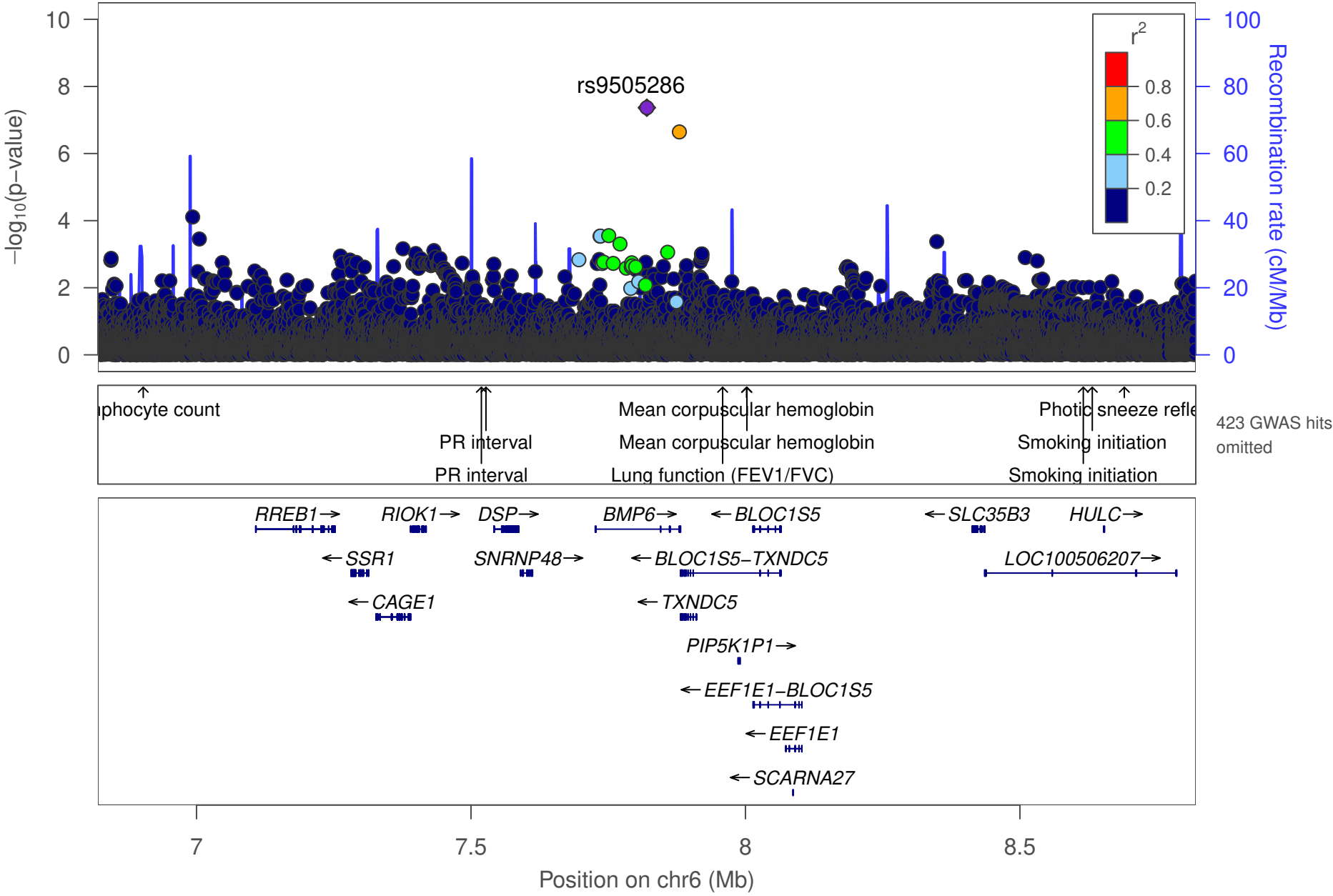
Post-GWAS Analysis and Polygenic Scores

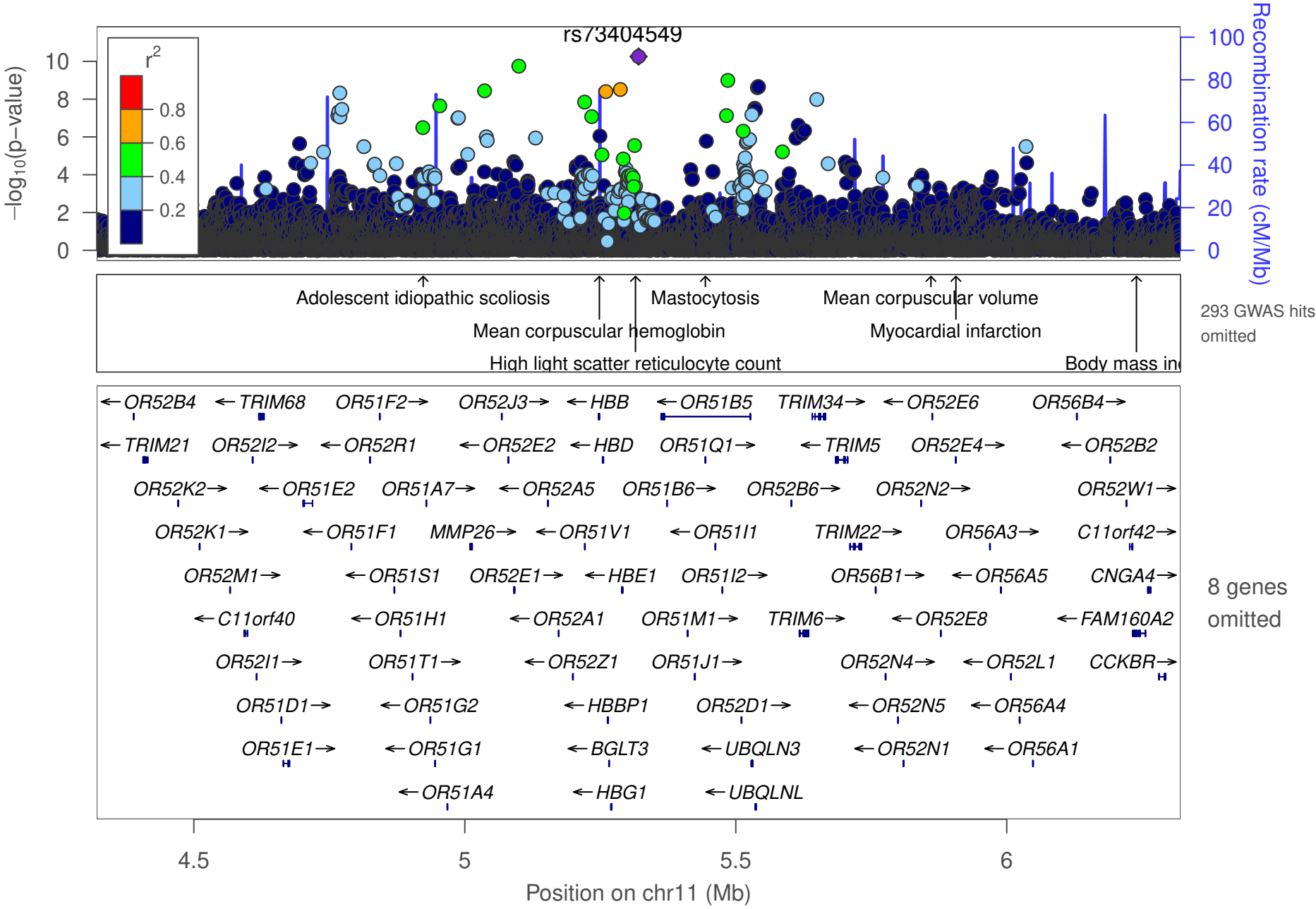


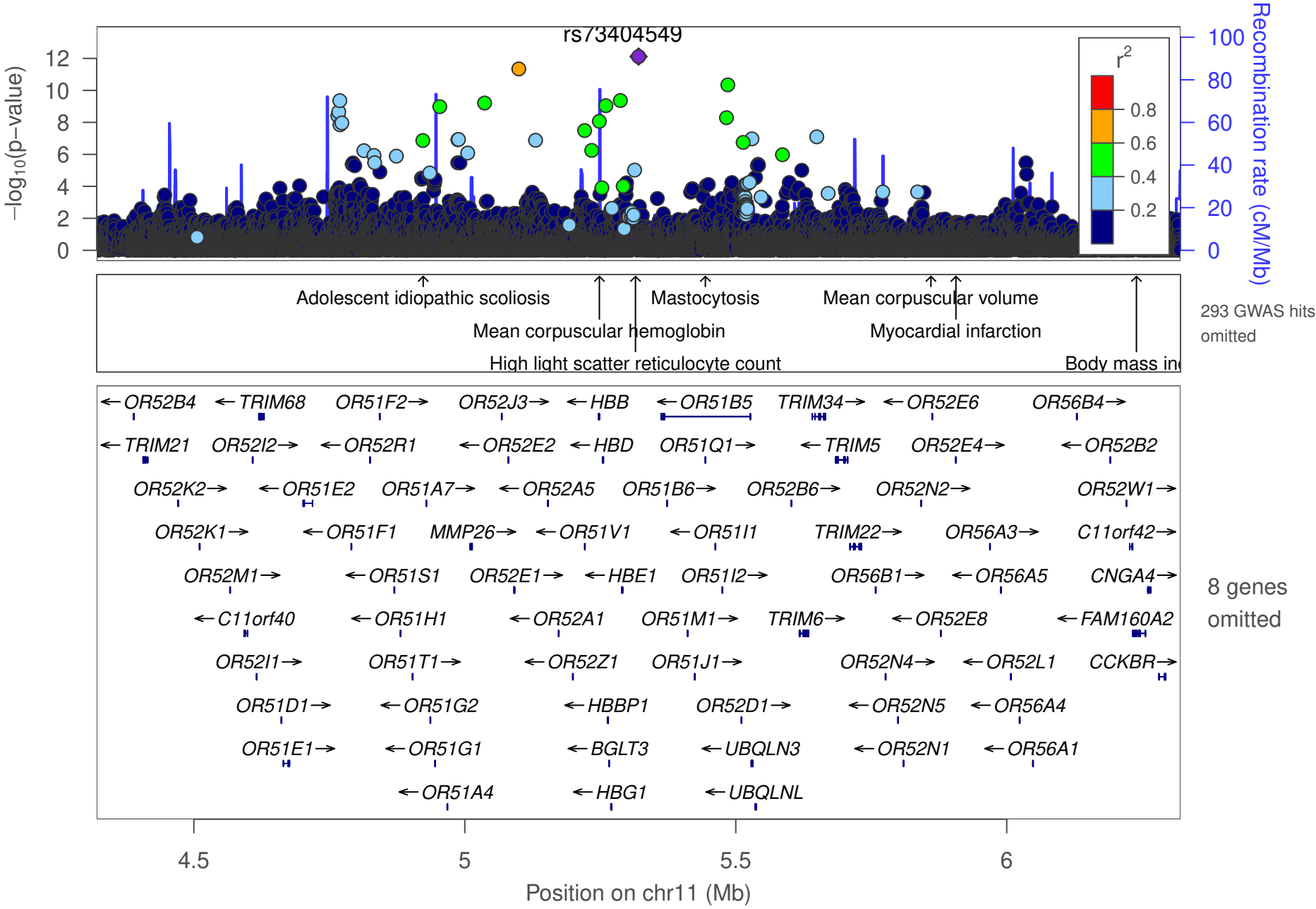

BMP6
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Chromosome

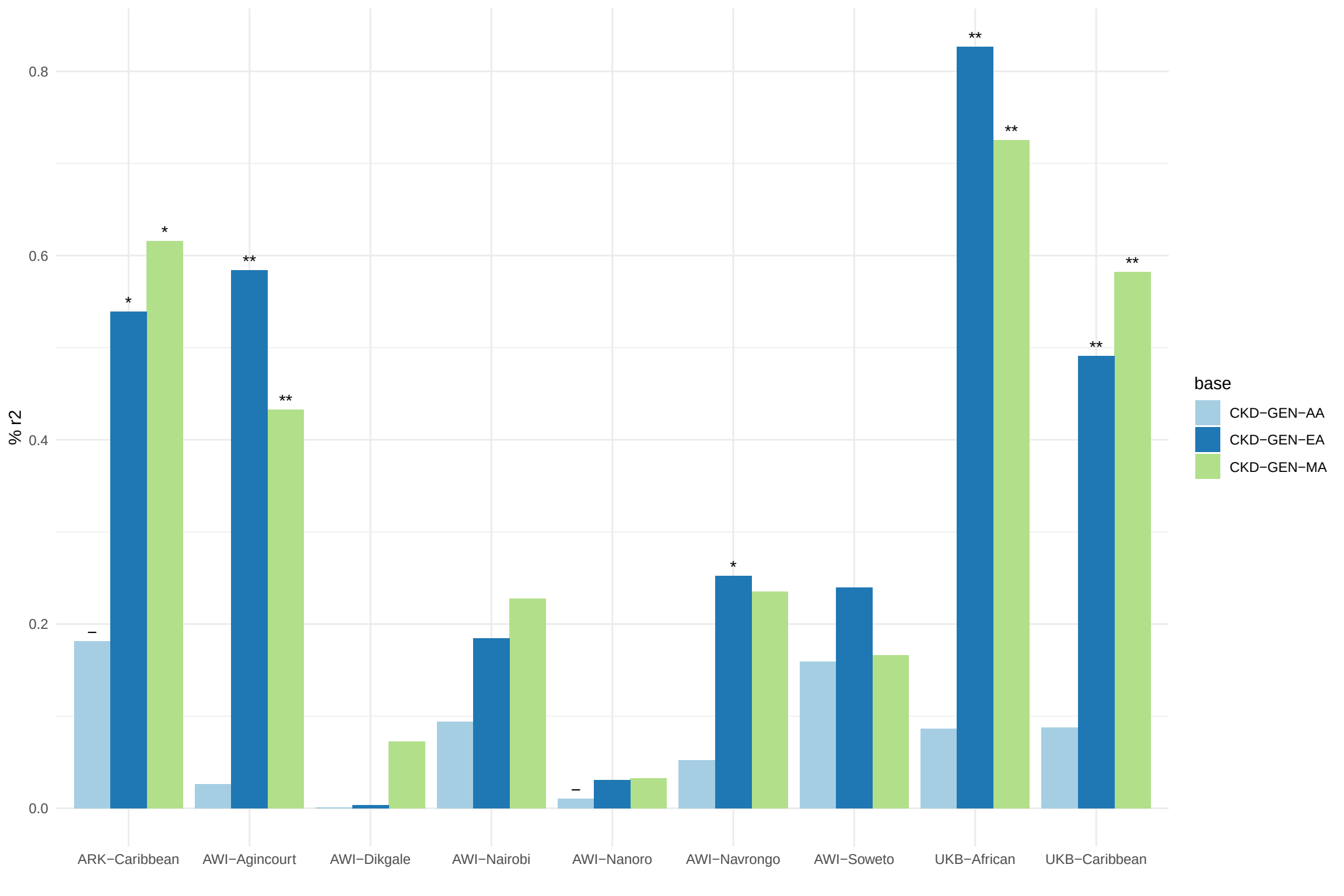

HBE1, OR51B4, HBB
rs79404549
$-\log_{10}(p)$
10
8
6
4
2
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Chromosome


HBE1, OR51B4, HBD
rs73101519
12
10
8
6
4
2
0
$-\log_{10}(p)$
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
Chromosome


rs9505286
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p−value)
Recombination rate (cM/Mb)
10
8
6
4
2
0
100
80
60
40
20
0
phocyte count
PR interval
PR interval
Mean corpuscular hemoglobin
Mean corpuscular hemoglobin
Lung function (FEV1/FVC)
Photic sneeze refle
Smoking initiation
Smoking initiation
423 GWAS hits omitted
RREB1
RIOK1
DSP
BMP6
BLOC1S5
SLC35B3
HULC
SSR1
SNRNP48
BLOC1S5−TXNDC5
LOC100506207
CAGE1
TXNDC5
PIP5K1P1
EEF1E1−BLOC1S5
EEF1E1
SCARNA27
7
7.5
8
8.5
Position on chr6 (Mb)


rs73404549
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
Adolescent idiopathic scoliosis
Mastocytosis
Mean corpuscular volume
Mean corpuscular hemoglobin
Myocardial infarction
High light scatter reticulocyte count
Body mass in
293 GWAS hits omitted
OR52B4 TRIM68 OR51F2 OR52J3 HBB OR51B5 TRIM34 OR52E6 OR56B4
TRIM21 OR52I2 OR52R1 OR52E2 HBD OR51Q1 TRIM5 OR52E4 OR52B2
OR52K2 OR51E2 OR51A7 OR52A5 OR51B6 OR52B6 OR52N2 OR52W1
OR52K1 OR51F1 MMP26 OR51V1 OR51I1 TRIM22 OR56A3 C11orf42
OR52M1 OR51S1 OR52E1 HBE1 OR51I2 OR56B1 OR56A5 CNGA4
C11orf40 OR51H1 OR52A1 OR51M1 TRIM6 OR52E8 FAM160A2
OR52I1 OR51T1 OR52Z1 OR51J1 OR52N4 OR52L1 CCKBR
OR51D1 OR51G2 HBBP1 OR52D1 OR52N5 OR56A4
OR51E1 OR51G1 BGLT3 UBQLN3 OR52N1 OR56A1
OR51A4 HBG1 UBQLNL
8 genes omitted
4.5
5
5.5
6
Position on chr11 (Mb)


rs73404549
−log10(p−value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
Adolescent idiopathic scoliosis
Mastocytosis
Mean corpuscular volume
Mean corpuscular hemoglobin
Myocardial infarction
High light scatter reticulocyte count
Body mass in
293 GWAS hits omitted
OR52B4 TRIM68 OR51F2 OR52J3 HBB OR51B5 TRIM34 OR52E6 OR56B4
TRIM21 OR52I2 OR52R1 OR52E2 HBD OR51Q1 TRIM5 OR52E4 OR52B2
OR52K2 OR51E2 OR51A7 OR52A5 OR51B6 OR52B6 OR52N2 OR52W1
OR52K1 OR51F1 MMP26 OR51V1 OR51I1 TRIM22 OR56A3 C11orf42
OR52M1 OR51S1 OR52E1 HBE1 OR51I2 OR56B1 OR56A5 CNGA4
C11orf40 OR51H1 OR52A1 OR51M1 TRIM6 OR52E8 FAM160A2
OR52I1 OR51T1 OR52Z1 OR51J1 OR52N4 OR52L1 CCKBR
OR51D1 OR51G2 HBBP1 OR52D1 OR52N5 OR56A4
OR51E1 OR51G1 BGLT3 UBQLN3 OR52N1 OR56A1
OR51A4 HBG1 UBQLNL
8 genes omitted
Position on chr11 (Mb)


% r2
0.0
0.2
0.4
0.6
0.8
ARK−Caribbean
AWI−Agincourt
AWI−Dikgale
AWI−Nairobi
AWI−Nanoro
AWI−Navrongo
AWI−Soweto
UKB−African
UKB−Caribbean
base
CKD−GEN−AA
CKD−GEN−EA
CKD−GEN−MA