Whole-genome sequencing reveals contribution of rare and common variation to structural kidney and urinary tract malformations

Melanie MY Chan[1,2], Omid Sadeghi-Alavijeh[1], Seth du Preez[1], Catalin D Voinescu[1], Loes FM van der Zanden[3], Sander Groen in 't Woud[3,4], Michiel F Schreuder[5], Wout Feitz[6], Enrico Mingardo[7], Alina C Hilger[8,9], Heiko Reutter[8,10], Lisanne M Vendrig[11], Rik Westland[11], Glenda M Beaman[12,13], William G Newman[12,13], Adrian S Woolf[14], Horia C Stanescu[1], Adam P Levine[1,15], Detlef Böckenhauer[1,16] and Daniel P Gale[1]

[1]Centre for Kidney and Bladder Health, University College London, UK.
[2]Medical Research Council Laboratory of Medical Sciences, Imperial College London, UK.
[3]IQ Health science department, Radboud university medical center, Nijmegen, The Netherlands.
[4]Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands.
[5]Department of Paediatric Nephrology, Radboud university medical center, Amalia Children's Hospital, Nijmegen, The Netherlands.
[6]Division of Paediatric Urology, Department of Urology, Radboud university medical center, Amalia Children's Hospital, Nijmegen, The Netherlands.
[7]Nutritional Biochemistry, Campus Kulmbach, University of Bayreuth, Germany.
[8]Research Center on Rare Kidney Diseases (RECORD), University Hospital Erlangen, Germany.
[9]Department of Pediatrics and Adolescent Medicine, Friedrich - Alexander University Nürnberg Erlangen, Erlangen, Germany.
[10]Division of Neonatology and Pediatric Intensive Care Medicine, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen, Germany.
[11]Department of Paediatric Nephrology, Emma Children's Hospital, Amsterdam University Medical Center, Amsterdam, The Netherlands.
[12]Manchester Centre for Genomic Medicine, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.
[13]Division of Evolution Infection and Genomics, School of Biological Sciences, Faculty of Biology Medicine and Health, University of Manchester, Manchester, UK.
[14]Division of Cell Matrix Biology and Regenerative Medicine, School of Biological Sciences, Faculty of Biology Medicine and Health, University of Manchester, Manchester, United Kingdom.
[15]Research Department of Pathology, University College London, UK.
[16]Department of Paediatric Nephrology, UZ Leuven, and Department of Cellular and Molecular Physiology, KUL, Belgium.

Keywords: CAKUT, kidney, bladder exstrophy, PUV, GWAS, genetics, WGS

Corresponding author: Dr Melanie Chan, MRC Laboratory of Medical Sciences, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London, W12 0HS, UK. Email: m.chan@lms.mrc.ac.uk

Running head: WGS in CAKUT

Abstract

Introduction

Congenital anomalies of the kidneys and urinary tract (CAKUT) are the commonest cause of kidney failure in children and young adults. Over 50 monogenic causes have been identified, however less than 20% of patients have a genetic diagnosis identified using targeted or whole exome sequencing. We sought to characterise the genomic architecture of CAKUT using whole genome sequencing (WGS).

Methods

Using WGS from 1,052 unrelated individuals with CAKUT recruited to the UK's 100,000 Genomes Project, we determined diagnostic yield and looked for gene-based enrichment of rare variants exome-wide. We performed sequencing based genome-wide association studies (seqGWAS) and used these results to estimate the heritability explained by common variants.

Results

The overall diagnostic yield was 4.9% with family history (P=0.02; OR 2.2; 95% CI 1.1-4.4), consanguinity (P=0.01; OR 3.0; 95%CI 1.2-6.9) and extra-renal features (P=1.1x10^{-4}; OR 3.1; 95% CI 1.7-5.7) independently predicting a monogenic diagnosis. Diagnostic yield was highest in cystic kidney dysplasia (11.1%) and kidney agenesis/hypodysplasia (7.4%). Exome-wide rare variant and genome-wide common variant (minor allele frequency [MAF] $\geq$ 0.5%) association testing in a subset of 813 patients and 25,205 ancestry-matched controls identified significant association at 6q16.3 (rs117473527; P=4.83x10^{-8}; OR 3.13; 95% CI 2.08-4.72; MAF 0.01) which requires replication. Common variants were estimated to explain up to 23% of the phenotypic variance observed in CAKUT in those with European ancestry suggesting that larger studies are needed to recover some of this 'missing heritability'. A genomic risk score for posterior urethral valves was also validated in an independent European cohort of 77 cases and 2,746 controls (P<0.001).

Conclusions

Only a minority of patients in this large, unselected cohort received a monogenic diagnosis, with common variants estimated to account for a substantial proportion of phenotypic heritability. This suggests that non-Mendelian genomic factors may be important for the pathogenesis of CAKUT.

Lay Summary

This study shows that single-gene causes of isolated and non-familial CAKUT are rare, and that genomic testing should be targeted towards those with kidney cysts and/or small kidneys that have not formed properly in the womb. Individuals with a close relative with CAKUT and those with involvement of other organ systems were more likely to receive a genetic diagnosis. These data support a possible polygenic basis for CAKUT, where many common DNA changes cumulatively affect risk, particularly in posterior urethral valves, the most common cause of kidney failure in boys.

Larger collaborative genomic studies are needed to increase our ability to identify these DNA changes and the mechanisms and pathways important for kidney and urinary tract development.

Introduction

Congenital Anomalies of the Kidneys and Urinary Tract (CAKUT) is a widely used diagnostic category that encompasses a range of disorders leading to structural kidney and urinary tract malformations. CAKUT are the commonest cause of chronic kidney disease (CKD) in children, accounting for ~50% of young people requiring dialysis or a kidney transplant to survive.[1,2] This heterogenous group of malformations are caused by defects in embryonic development and include kidney defects (e.g., kidney agenesis, kidney hypodysplasia or cystic dysplasia), collecting system defects (e.g., pelvi-ureteral junction obstruction, primary megaureter or vesicoureteral reflux [VUR]), abnormalities of the lower urinary tract causing bladder dysfunction and/or obstruction (e.g., posterior urethral valves [PUV]) and the bladder-exstrophy-epispadias-complex.

CAKUT can display familial clustering[3–5] with CAKUT-associated syndromes and single gene mouse models implicating a monogenic basis for these disorders. Over 50 monogenic causes of CAKUT have been reported[6] with an enrichment of rare copy number variation (CNV) also suggesting gene dosage is important.[7–9] Many of these genes encode transcription factors, suggesting that disruption of shared transcriptional networks is likely to be a fundamental mechanism underlying these conditions.

However, several observations suggest that CAKUT are more often not the result of a rare single gene defect. First, fewer than 20% of affected individuals have a monogenic cause identified,[10–18] with diagnostic yield strongly influenced by the cohort studied and testing approach employed. Second, CAKUT demonstrate significant genotype-phenotype heterogeneity, incomplete penetrance and variable expressivity meaning the same variant in a gene can be present in individuals with very different malformations. Third, common variation has previously been associated with susceptibility to PUV,[19] VUR[20] and bladder exstrophy.[21] Finally, maternal diabetes[22] and obesity[23–25] have been linked with an increased risk of CAKUT highlighting the potential contribution of genotype-environment interactions to disease risk.[26] Taken together, these observations suggest a more complex genomic basis underlying the pathogenesis of CAKUT and imply disruption of cell-type specific, dynamic and temporo-spatially sensitive gene-regulatory networks.

Whole-genome sequencing (WGS) enables the comprehensive assessment of variation across the allele frequency spectrum, providing uniform genome-wide coverage of coding and non-coding regions as well as improved detection of disease-causing and structural variants.[27,28] Here, we use WGS data to determine its diagnostic utility in 1,052 individuals with CAKUT recruited to the 100,000 Genomes Project.[29] In addition, our unbiased genome-wide approach interrogating both rare and common variant associations across the entire cohort, and stratified by phenotype, reveals evidence of a polygenic contribution to these phenotypically diverse structural malformations.

Methods

The 100,000 Genomes Project

The 100,000 Genomes Project (100KGP) dataset (version 19) consists of WGS data, clinical phenotypes encoded using Human Phenotype Ontology (HPO) codes, and retrospective and prospectively ascertained hospital records for 90,173 individuals with cancer or rare disease, as well

as their unaffected relatives.[29] Full details on the cohort, WGS, alignment, variant calling, and quality control filtering are provided in the Supplementary Methods. The study workflow is shown in Supplementary Figure S1.

Cohort selection

Patients were recruited to the 100KGP as part of the 'Congenital anomalies of the kidneys and urinary tract (CAKUT)' cohort with the following inclusion criteria: CAKUT with syndromic manifestations in other organ systems; isolated CAKUT with a first-degree relative with CAKUT or unexplained CKD; multiple distinct kidney/urinary tract anomalies; CAKUT with unexplained end-stage kidney disease before the age of 25 years. Pre-screening for *HNF1B* and *SALL1* was recommended with a personal or family history of diabetes, or imperforate anus, ear or thumb abnormalities, respectively.

For the association analyses, patients were stratified by HPO terms (Supplementary Table S1) into the following six groups: kidney anomalies (n=237), cystic renal dysplasia (n=112), obstructive uropathy (excluding posterior urethral valves; n=177), VUR (n=174), PUV (n=132) and bladder exstrophy (n=97). Unaffected relatives of non-kidney disease participants from the 100KGP, excluding those with SNOMED-CT or ICD-10 codes consistent with kidney disease/failure (Supplementary Table S2) were used as controls. A cohort of unrelated ancestry-matched cases and controls was generated to minimize the effects of population stratification in this mixed-ancestry cohort (Supplementary Methods and Supplementary Figure S2).

Identification of disease-causing variants

Probands were assessed using the Genomics England clinical interpretation pipeline to determine a genomic diagnosis with results reviewed by a multi-disciplinary team to determine pathogenicity according to American College of Medical Genetics and Genomics (ACMG)[33] criteria (see Supplementary Methods). In those without a molecular diagnosis, the top five ranked Exomiser[32] variants and CNV calls across 286 expertly-curated kidney disease genes and loci (CAKUT panel v1.178 and unexplained kidney failure panel v12.7) were manually reviewed.

Rare variant gene-based analysis

Rare coding single-nucleotide variants (SNVs) and indels with MAF < 0.01% in gnomAD[34] were aggregated by gene and association testing carried out using SKAT-O[36] implemented by SAIGE-GENE+[37] (version 1.3.6). A loss-of-function (LoF) only mask and LoF+missense mask were used. 'High confidence' loss-of-function variants (stop gain, splice site and frameshift) were determined by LOFTEE[34] and missense variants were filtered using REVEL[35] score ≥ 0.75 Variant and sample quality control is detailed in the Supplementary Methods. Sex and the top ten principal components were used as covariates. A Bonferroni adjusted *P* value of 0.05/number of genes tested was used to determine exome-wide significance.

seqGWAS

A WGS genome-wide association study (seqGWAS) of SNVs/indels with MAF ≥ 0.5% was performed with SAIGE[38] (version 1.0.7). seqGWAS was first applied to 813 CAKUT cases and 25,205 ancestry-matched controls and then stratified by phenotype group and ancestry (Supplementary Methods). Sex and the top ten principal components were used as covariates. The conventional genome-wide significance threshold of $5x10^{-8}$ was employed.

Meta-Analysis

GWAS summary statistics from the 100KGP obstructive uropathy cohort (177 cases and 24,451 controls) were meta-analysed with the congenital obstructive uropathy cohort in FinnGen[39] (Release 10; 550 cases and 410,449 controls) using the inverse-variance weighted approach implemented by METAL.[40] 11,647,818 variants with minor allele count (MAC) > 5 were included. Lambda was 1.01 indicating good control of population stratification. The same approach was used to meta-analyse 4,696,412 variants with MAF > 1% in the 100KGP bladder exstrophy cohort (97 cases and 22,037 controls) with 628 European cases and 7,352 controls from Mingardo et al.[21] with lambda of 0.98.

Heritability

Narrow-sense heritability (h^2) was estimated using two different approaches (GREML-LDMS[41] and SumHer[42]) in a subset of the 100KGP cohort with genetically defined European ancestry (623 cases and 20,060 controls). Observed heritability was transformed to a liability threshold model[43] adjusting for case:control ratio and an estimated population prevalence of 0.2%.[44] LD-Score Regression (v1.0.1)[45] was used to estimate enrichment of functional annotations using the European CAKUT seqGWAS summary statistics and 1000 Genomes (Phase 3) reference panel for LD.

PUV Genomic Risk Score

PRS-CS (version 1.1.0)[46] was used to generate a genomic risk score from European male PUV GWAS summary statistics (89 cases and 8,303 controls) using the UK BioBank European LD reference panel. This was applied with PGS Catalog Calculator (version 2.2.0)[47] to an independent cohort of 77 cases from the Dutch AGORA biobank and 2,746 controls from Genomics England using ten principal components as covariates and normalising to Human Genome Diversity Panel and 1000 Genomes Project scores.

Results

A total of 1,052 probands with CAKUT were recruited to the 100KGP: 30.2% were recruited as singletons, 29.8% as duos and 31.0% as trios. Table 1 summarizes the clinical and demographic data for the probands. The cohort had a median age of 20 years. 13.6% had an affected first-degree relative, 6.1% had reported consanguinity and 28.5% had extra-renal manifestations. The most frequently reported HPO terms in the cohort were hydronephrosis (27.2%), vesicoureteral reflux (VUR) (23.7%) and abnormality of the bladder (20.1%) (Supplementary Figure S3).

Table 1. Demographics and clinical characteristics of the CAKUT cohort (n=1,052). Some patients had more than one phenotype. VUR, vesico-ureteral reflux; PUV, posterior urethral valves.

Median age in years (range)		20 (7-86)
Males (%)		662 (62.9)
Self-reported ethnicity		
	White (%)	653 (62.1)
	South Asian (%)	98 (9.3)
	Mixed/Other (%)	59 (5.6)
	Black (%)	28 (2.7)
	Unknown (%)	213 (20.2)
Phenotype		
	Kidney anomaly (%)	393 (37.6)
	Obstructive uropathy (%)	341 (32.4)
	VUR (%)	249 (23.7)
	PUV (%)	173 (16.4)
	Cystic dysplasia (%)	162 (15.4)
	Bladder exstrophy (%)	108 (10.3)
	Duplex kidney (%)	57 (5.4)
	Ectopic kidney (%)	39 (3.7)
Extrarenal manifestations (%)		300 (28.5)
	Neurodevelopmental delay (%)	72 (6.8)
	Skeletal anomalies (%)	40 (3.8)
	Hearing impairment (%)	31 (2.9)
	Endocrine (%)	22 (2.1)
	Eye anomalies (%)	22 (2.1)
	Respiratory (%)	22 (2.1)
	Cardiac anomalies (%)	22 (2.1)
Family history of CAKUT (%)		143 (13.6)
Reported consanguinity (%)		64 (6.1)
Kidney failure (%)		285 (27.1)
Median age of kidney failure in years (range)		12 (0-70)

Monogenic structural kidney and urinary tract malformations are rare

We first assessed the diagnostic utility of WGS in this cohort, largely representative of general nephrology practice. Overall, 4.9% (52/1052) had a genomic diagnosis (Supplementary Table S3), with 71.2% (37/52) of the diagnoses fully explaining the CAKUT phenotype. Excluding individuals with isolated PUV and bladder exstrophy (where a monogenic cause has not been identified or found in a very small number of familial cases), the diagnostic yield was 6.7% (52/771).

40.4% (21/52) of those with a genomic diagnosis had a pathogenic or likely pathogenic variant in a known CAKUT gene, most frequently in *HNF1B* (seven individuals including four with 17q12 deletions; 0.7% of the total cohort). Additional disease-causing variants were reported in *PAX2* (3), *PBX1* (2), *TBX6* (two individuals including one with 16p11.2 microdeletion), *WT1* (2) and *GREB1L* (1) (Figure 1a; Supplementary Table S3). 26.9% (14/52) of the identified pathogenic or likely pathogenic variants affected kidney disease genes which are not usually associated with a CAKUT phenotype: *PKD1* (5), *PKHD1* (2), *CEP290* (2), *NPHP1* (2), *NPHP3* (1), *COL4A5* (1), *CUBN* (1), *SLC3A1* (1) and *CLCN5* (1). These likely represent phenocopies in the case of cystic disease, or possible misclassification during recruitment. An additional 23% (12/52) had a monogenic cause for a non CAKUT-associated syndrome identified which did not fully explain the observed kidney and/or urinary tract anomalies. Excluding these 12 patients with a non-CAKUT syndrome, the five individuals with non-CAKUT diagnoses who may have been mis-classified (*NF1, CLCN5, COL4A5, CUBN and SLC3A1*), and 12 more with cystic/ciliopathy disease (*PKD1, PKHD1, NPHP1, NPHP3, CEP290*) the diagnostic yield fell to 3.1% (23/743). These data demonstrate that the prevalence of CAKUT-associated monogenic disease, at least in this cohort, is low and confirm that CAKUT is both genetically and phenotypically heterogeneous.

Predictors of a genomic diagnosis

Whilst the overall diagnostic yield was low, those with cystic dysplasia and other kidney anomalies (i.e. kidney agenesis, hypoplasia, or dysplasia) were more likely to receive a genomic diagnosis (11.1% and 7.4%, respectively; Figure 1b). Pathogenic variants affecting *HNF1B* (including 17q12 deletions) accounted for 27.8% (5/18) of genomic diagnoses in the cystic dysplasia cohort, however most patients (55.6%) were found to have phenocopies; autosomal dominant polycystic kidney disease (ADPKD), autosomal recessive polycystic kidney disease (ARPKD), or nephronophthisis. Importantly, all patients with cystic dysplasia who had a confirmed genomic diagnosis also reported extra-renal features such as neurodevelopmental or skeletal abnormalities, abnormal liver function tests or hyperuricaemia (Supplementary Table S3). Likewise, in the kidney agenesis/hypodysplasia cohort 76.5% (13/17) of genomic diagnoses were associated with a syndrome or extra-renal manifestations.

Consistent with previous reports from individuals with suspected monogenic kidney disease,[18,46–48] multivariable logistic regression analysis confirmed that extra-renal features (P=1.1x10^{-4}; OR 3.1; 95% CI 1.7-5.7), consanguinity (P=0.01; OR 3.0; 95% CI 1.2-6.9) and family history (P=0.02; OR 2.2; 95% CI 1.1-4.4) were independent predictors of a positive genomic diagnosis (Supplementary Table S4). A diagnosis was identified in 9.3%, 12.5% and 9.1% of those with extra-renal features, consanguinity and a family history, respectively, compared to 5%, 4.5% and 4.3% without. These findings provide evidence to help clinicians identify who might benefit most from genomic testing, particularly relevant in resource-limited settings.

Candidate gene discovery

Given the low prevalence of known monogenic disease in this cohort, an unbiased whole-exome approach was employed to identify candidate genes which may act as non-Mendelian risk factors for disease in a subset of 813 cases and 25,205 ancestry-matched controls. Single variant association

testing can be underpowered when variants are rare and a collapsing strategy which aggregates variants by gene can be adopted to boost power. We therefore combined a) rare (gnomAD MAF < 0.01%) loss-of-function (LoF; protein-truncating or splice-site) and b) rare LoF and predicted deleterious missense (REVEL ≥ 0.75) variants across 18,353 protein-coding genes. A minor allele frequency of 0.01% was chosen to capture incompletely penetrant variants and avoid loss of power associated with using an ultra-rare threshold. We did not identify any genes that were significantly enriched for rare variation after correction for multiple testing (Figure 2a; Supplementary Figure S4a; Supplementary Table S5). The median number of variants tested per gene was 5 (interquartile range [IQR] 7) and 10 (IQR 13), in each analysis respectively. The absence of exome-wide significant gene-based enrichment is likely the result of reduced power due to the genetic and phenotypic heterogeneity of the cohort and the need for multiple testing adjustment.

The mesodermal transcription factor *TBX15* showed the greatest enrichment of rare likely deleterious variation (case minor allele count [MAC] 6, control MAC 18; $P=1.9\times10^{-5}$) and is closely related to known CAKUT gene *TBX18.*[51] *HNF1B* demonstrated non-significant enrichment of rare, LoF and missense variation (case MAC 6, control MAC 25; $P=2.7\times10^{-4}$) but not LoF-only variation (case MAC 1, control MAC 2; $P=0.07$) suggesting that LoF variants are not the primary drivers of the observed signal. Excluding 33 individuals with a known monogenic diagnosis and depleting for PUV and bladder exstrophy cases did not reveal any additional signals in the remaining 539 unsolved cases. The enrichment previously seen in *HNF1B* was attenuated (case MAC 3, control MAC 27; $P=0.01$).

To increase power through more stringent case selection, the rare variant analysis was repeated stratifying by phenotype. No statistically significant associations for a) LoF only or b) LoF and likely deleterious missense variant sets were detected (Supplementary Tables S6-S11). In the cystic dysplasia cohort (n=112), greatest enrichment was seen in *PKD1* (case MAC 4, control MAC 41; $P=1.5\times10^{-5}$), while *PKD2* and *HNF1B* were non-significant ($P=0.85$ and $P=0.72$, respectively), possibly due to limited power and the small size of the cohort. In the congenital obstructive uropathy cohort (n=177), no known monogenic causes of functional or anatomical lower urinary tract obstruction were enriched for rare coding variation: *MYOCD*, $P=0.86$; *TBX18*, $P=0.75$; *BNC2*, $P=0.94$; *FLNA*, $P=0.57$; *HPSE2*, $P=0.83$; *LRIG2*, $P=0.16$; *CHRM3*, $P=0.94$; *CHRNA3*, $P=0.64$. Supplementary Tables S6-11 detail per-gene results for each subgroup: kidney anomalies (n=237), obstructive uropathy (n=177), VUR (n=174), PUV (n=132), cystic dysplasia (n=112) and bladder exstrophy (n=97).

To ensure test statistics were not inflated by the inclusion of individuals from different ancestries, we repeated the rare variant analysis using a subset of 611 cases and 19,576 controls of European ancestry (Supplementary Table S12). Both the lead gene in the mixed-ancestry analysis, *TBX15* (case MAC 3, control MAC 14; $P=4.2\times10^{-3}$) and *HNF1B* (case MAC 5, control MAC 18; $P=3.7\times10^{-4}$) had an attenuated signal, likely due to reduced power. Effect estimates (Spearman's rho=0.9) and *P* values (Spearman's rho=0.7) were consistent across both cohorts indicating that population structure is adequately controlled in the mixed-ancestry cohort (Supplementary Figure S4b and S4c).

seqGWAS

In the absence of significant rare variant enrichment, we next explored whether common and low-frequency variation might contribute to the genomic architecture of CAKUT, performing a sequencing-based GWAS (seqGWAS) in 813 cases and 25,205 ancestry-matched controls testing for association at 10,147,641 SNVs/indels with MAF $\geq$ 0.5% (Figure 2b). Lambda (λ) was 1 indicating minimal evidence of confounding by population stratification in this mixed-ancestry cohort (Supplementary Figure S5a). Two variants reached genome-wide significance. The first, rs117473527 (chr6:102155812:G>C) at 6q16.3 (P=4.83x10^{-8}; OR 3.13; 95% CI 2.08-4.72; MAF 0.01) (Supplementary Figure S5b) was an intergenic variant downstream of *GRIK2* (glutamate ionotropic receptor kainite type subunit 2) which has been associated with neurodevelopmental disorders and urinary tract cancers.[52,53] *Grik2* is expressed in the murine ureteric bud at E10.5-11.5 (Supplementary Figure S5c) when it invades the metanephric mesenchyme and *GRIK2* displays robust expression in bulk-RNAseq data from 8-10 week fetal kidneys[52] indicating it may be important during nephrogenesis. Of note, its family member *GRIK3* is upregulated in dysplastic tubules of *HNF1B* mutant kidneys.[52] Replication of this signal in an independent cohort is needed to robustly validate this association.

The second genome-wide significant variant (a 1 base pair deletion at 1q21.3; rs35251516; P=4.93x10^{-8}; MAF 0.03) was isolated, found in a low-complexity region and likely represents a sequencing artefact. Testing for association in a European-only cohort of 623 cases and 20,060 controls across 11,303,429 variants did not identify any additional signals (Supplementary Figure S6a and S6b; λ 0.998). Significant correlation between P-values and effect sizes in the mixed-ancestry and European-only cohorts confirms that population stratification is well-controlled in both cohorts (Supplementary Figure S6c and S6d).

The absence of additional significant associations in the mixed-ancestry cohort is likely the result of limited power related to the phenotypic heterogeneity of the cohort and sample size. With 813 cases (assuming a prevalence of 1 in 500) this analysis is 80% powered to detect association of common variants (MAF > 5%) with OR > 1.8 or low-frequency variants (MAF > 1%) with OR > 3 (Supplementary Figure S7). To increase power, we aggregated variants across 19,118 autosomal protein-coding genes and 50 hallmark gene-sets curated by MSigDB but did not identify any enrichment in specific genes or biological pathways after correction for multiple-testing (Supplementary Table S13).

seqGWAS of the kidney anomalies, cystic renal dysplasia, obstructive uropathy and VUR cohorts did not identify any statistically significant associations (Supplementary Material) but were underpowered to detect loci other than those with large effect sizes (e.g., OR > 4 for variants with MAF 5%) at these limited sample sizes. To increase power, a meta-analysis of the obstructive uropathy cohort was carried out with GWAS summary statistics from 550 congenital obstructive uropathy cases and 410,449 controls from FinnGen[39] for a total of 727 cases and 434,900 controls across 11,564,803 variants, but no genome-wide significant variants were identified (Supplementary Figure S8). As reported previously,[19] statistically significant association was detected at two loci in 132 individuals with PUV when compared to 23,727 ancestry-matched controls across 19,651,224 SNVs/indels: 12q24.21 near *TBX5* (rs10774740; P=7.81x10^{-12}; OR 0.40; 95% CI 0.31-0.52; MAF 0.37) and 6p21.1 in *PTK7* (rs144171242; P=2.02x10^{-8}; OR 7.20; 95% CI 4.08-12.70; MAF 0.007). Both

associations were replicated in an independent European cohort of 395 PUV patients (333 German and Polish patients from the CaRE for LUTO study and 62 from the UK) and 4,151 unaffected male 100KGP controls.[19]

Bladder Exstrophy seqGWAS

Classic bladder exstrophy (CBE) is the most frequent form of the bladder-exstrophy-epispadias-complex (BEEC) spectrum resulting from a defect in abdominal midline development. It is usually sporadic, not generally associated with other malformations and its pathogenesis is poorly understood. Using the seqGWAS approach we tested 10,103,261 SNV/indels with MAF $\geq$ 0.5% for association in 97 individuals with CBE and 25,921 ancestry-matched controls (Supplementary Figure S9a and S9d). Suggestive evidence of association was identified at 5q35.3 (rs147504710; P=1.11x10^{-7}; Supplementary Figure S9b) and 20p11.22 (P=2.13x10^{-7}; Supplementary Figure S9c; Supplementary Table S14). Of note, these variants had MAF ~1% and a large effect size meaning they may have been missed using conventional genotyping and imputation. rs6106456 at 20p11.22 did not replicate in an independent cohort of 84 unrelated European patients with bladder exstrophy and 10,804 controls from the 100KGP (Cochran-Armitage trend test P=0.52), however the estimated power (0.67) was insufficient to definitively confirm or refute the association. Meta-analysis with a previously published European GWAS of 628 patients with classic bladder exstrophy and 7,352 controls[21] across 4,663,819 variants with MAF $\geq$ 1% failed to identify any additional novel loci (Supplementary Figure S10).

The previously reported *ISL1* association at 5q11.1 was replicated in our study with the lead variant rs9291768 (chr5:51421959:C>T) achieving a P value of 1.31x10^{-3} (OR 1.63; 95% CI 1.21-2.19; MAF 0.35). To determine whether there were any additional rare variants that might be driving this signal we repeated association testing at this locus using variants with MAC > 3 (Supplementary Figure S9e). A rare, intergenic indel rs550737686 (chr5:51494837:CCT>C) demonstrated strongest association (P=2.35x10^{-5}; OR 6.11; 95% CI 3.11-12.03; MAF 0.008). Bayesian fine mapping using functional annotations including conservation, transcription factor binding and *cis*-regulatory elements found rs115201978 (chr5:51035061:G>A) to be likely causal with a high posterior probability > 0.99. rs115201978 is a rare, intergenic variant (MAF 0.006) with a high CADD score of 16.9 reaching a P-value of 1.28x10^{-4} (OR 6.26; 95% CI 2.92-13.44). It is not in linkage disequilibrium (LD) with the lead indel rs550737686 (r^2=0.1) which was excluded from the fine-mapping analysis as it is absent from the 1000 Genomes data.

Heritability and polygenic risk

The proportion of trait heritability attributable to common variation can be evaluated using either WGS data or GWAS summary statistics. In a subset of 623 European CAKUT cases and 20,060 ancestry-matched controls, we used WGS data to estimate the proportion of phenotypic variance attributed to additive common and low-frequency variation (MAF $\geq$ 0.1%) was 23% (SE 11%; Supplementary Table S15). Variants with MAF between 1% and 5% and with low LD scores, meaning they are not usually inherited together with other nearby variants, accounted for > 75% of this estimated heritability (Figure 3a). An alternative model using GWAS summary statistics from the same cohort (MAF $\geq$ 1%) generated a comparable heritability estimate of 24% (standard deviation 15%).

To determine whether specific functional annotations (such as enhancers or promoters) were enriched for common variant heritability in a European-only CAKUT cohort, we used LD Score Regression.[45] This estimated that 1.2 million variants with MAF > 5% accounted for 5% (SE 12%) of heritability on the liability scale, which in conjunction with the estimates above suggests that variants with MAF between 1-5% may account for a substantial proportion of 'missing heritability'. Partitioning this heritability showed nominal ($P < 0.05$) enrichment of 'super enhancers' – large clusters of transcriptional enhancers that drive expression of genes that define cell identity – but this was not significant at a false discovery rate of 10% (Figure 3b).

Finally, to test whether there was a polygenic contribution to PUV we generated a PUV-GRS using 675,086 autosomal variants from the European-only male PUV seqGWAS[19] (MAF > 0.5%; 89 cases and 8,303 controls) and applied this to WGS data from 77 individuals with PUV from the AGORA biobank and 2,746 European male controls from the 100KGP (Figure 3c; Supplementary Table S16). The PUV-GRS was significantly elevated in this independent European cohort compared to controls ($P < 0.001$).

Discussion

In this large cohort of patients with CAKUT, clinical-grade WGS identified a monogenic cause in fewer than 5% of individuals. The highest diagnostic yield was seen in those with kidney agenesis/hypodysplasia, cystic dysplasia, a family history, consanguinity and/or extra-renal features. This figure is at the lower end of estimates from previous studies using targeted gene panels or exome sequencing which report diagnostic yields up to 27%.[11,12,14,16,17,47,48,53–57] This is likely related to the broad definition of phenotypes recruited to 100KGP, including individuals with mild disease, and despite the higher diagnostic yield expected with WGS over exome sequencing.[58] WGS provides a comprehensive assessment of all variant types (SNV, indel, SVs) across the coding and non-coding genome and is less susceptible to sequencing bias.[27,28] While it is also possible individuals with a clearly monogenic cause of CAKUT were not recruited to the 100KGP, genetic testing for CAKUT was not widespread in the UK prior to this study and in a cohort similar to this one, a panel of 208 genes identified a diagnosis in 1.3% in 453 individuals.[10] Overall, these data suggest that only a minority of CAKUT patients in general nephrology practice have a monogenic cause for their condition identified and that genomic testing should be targeted towards those with kidney anomalies (particularly cystic dysplasia), a family history, consanguinity and/or extra-renal manifestations.

Unbiased interrogation of rare variation across the exome did not identify any significant gene-based enrichment. This lack of observed enrichment may be the result of several factors which limit power for discovery; a) the cohort is phenotypically heterogenous, b) this is a genetically heterogenous disorder with over 50 monogenic causes described, c) the sample size is small for a rare variant association study, and d) use of a stringent Bonferroni adjustment for multiple testing.

We used a sequencing-based GWAS approach to interrogate common and low-frequency variants. While statistically significant association was detected at 6q16.3 downstream of *GRIK2*, further replication in an independent cohort is necessary to confirm the association. Importantly however, we show that common genomic variation (with an allele frequency between 1% and 5%) contributes

to a significant proportion of phenotypic variance providing support for a polygenic basis to CAKUT. Furthermore, these common and low-frequency variants may be enriched in *cis*-regulatory elements (such as enhancers) and could potentially impact the transcriptional networks important for kidney and urinary tract development. It should be noted however that the standard errors of these heritability estimates are large and will need to be refined in larger cohorts. Previous work has shown susceptibility to lower urinary tract phenotypes, including PUV,[19] bladder exstrophy,[21,59,60] and VUR,[20,61,62] is influenced by non-coding common variation. This is reinforced by the generation and validation of a common variant PUV-GRS. These data suggest there are likely to be low-frequency variants with effect sizes too small to be detected in this cohort which may contribute to CAKUT. Larger, better-powered cohorts will be essential to identify these variants and the genes/pathways they impact to generate insights into the molecular mechanisms underlying these malformations.

The main strength of this analysis lies in the use of WGS data which enables ancestry-independent variant detection and testing of variants across the allele frequency spectrum. Using cases and controls sequenced on the same platform also minimizes the risk of confounding due to sequencing artefacts or differences in variant calling. We also made our analysis as inclusive as possible, including individuals of all ancestries, using a robust approach to minimize confounding by population stratification.

We acknowledge several limitations to this work. First, the molecular diagnoses are largely reliant on the Genomics England pipeline.[63] Second, the analysis is dependent on accurate phenotyping with HPO terms which was carried out by the recruiting team and patient's clinician. This is highlighted by the inclusion of a few patients with other kidney phenotypes/phenocopies in the CAKUT cohort. Third, we are unable to quantify the number of patients who were not recruited to the 100KGP because a genetic diagnosis had already been made. While this means those with syndromic disease may have been excluded, genetic testing for CAKUT was not widely implemented in the UK prior to the 100KGP. Lastly, is the issue of power. In a cohort of this size, we had 80% power to identify common variants (MAF > 1%) with an OR > 3, but weaker effects or genes accounting for a small proportion of cases are not expected to be reliably detected.

In summary, we provide support for the hypothesis that non-Mendelian genomic factors play an important role in the pathogenesis of CAKUT as evidenced by the lack of an identifiable monogenic cause in most cases and the heritability explained by commoner variants. Larger studies are now needed to try and identify these genomic susceptibility loci and investigate the mechanisms by which they might be impacting the cell-type specific gene regulatory networks important for urinary tract development.

Disclosure Statement

MFS discloses consultant fees from Bayer, Santhera, and Travere Therapeutics.
All the other authors declare no competing interests.

Data Sharing Statement

Supplementary material is available online at www.kidney-international.org. Summary statistics can be found at https://doi.org/10.5281/zenodo.18040141. Genomic and phenotype data from participants recruited to the 100,000 Genomes Project can be accessed by application to Genomics England Ltd at https://www.genomicsengland.co.uk/join-us/. Details of the WGS aggregated dataset used for the analysis can be found at https://re-docs.genomicsengland.co.uk/aggv2/. Code for the case–control ancestry-matching algorithm can be found at https://github.com/APLevine/PCA_Matching. Details of the rare-variant workflow can be found at https://re-docs.genomicsengland.co.uk/avt/. Details of the common-variant GWAS workflow can be found at https://re-docs.genomicsengland.co.uk/gwas/. CAKUT GWAS summary statistics will be made publicly available via the GWAS Catalog on publication (https://www.ebi.ac.uk/gwas/).

Acknowledgements

The authors gratefully acknowledge the participation of the patients, and their families recruited to the 100KGP. Data from the National Genomic Research Library (NGRL) used in this research are available within the secure Genomics England Research Environment. Access to NGRL data is restricted to adhere to consent requirements and protect participant privacy. Access to NGRL data is provided to approved researchers who are members of the Genomics England Research Network, subject to institutional access agreements and research project approval under participant-led governance. For more information on data access, visit: https://www.genomicsengland.co.uk/research.

Funding

MMYC was supported by a Kidney Research UK Clinical Research Fellowship (TF_004_20161125) and the Medical Research Council-National Institute of Health Care Research Rare Disease Research Platform (MR/Y008170/1) with ASW and WGN (MR/Y008340/1). OSA was funded by an MRC Clinical Research Training Fellowship (MR/S021329/1). SGW was supported by a Young Investigator Grant from the ESPN and LvdZ, SGW and LV were supported by a consortium grant from the Dutch Kidney Foundation (20OC002). WGN is supported by the NIHR Manchester BRC (NIHR 203308). APL was supported by an NIHR Academic Clinical Lectureship. DPG is supported by the St Peter's Trust for Kidney, Bladder and Prostate Research.

Author contributions

DPG and MMYC conceived the study and designed the experiments. MMYC, OSA and SP analysed the data and interpreted the results. APL, DB and HCS provided expertise in genomic analysis and contributed to study design. CDV provided computational support. Clinical phenotyping and genotyping/sequencing was carried out by the remaining co-authors (AGORA, ArtDECO Consortium, CaRE for LUTO, bladder exstrophy cohort). The manuscript was drafted by MMYC and critically revised by all authors. All authors approved the final manuscript and agree to be accountable for all aspects of the work.

Supplementary Material

Supplementary Methods

Supplementary Figure S1. Study workflow.

Supplementary Figure S2. Ancestry matching of cases and controls.

Supplementary Figure S3. Human Phenotype Ontology (HPO) terms in the CAKUT cohort.

Supplementary Figure S4. Quantile-quantile plot of CAKUT rare variant gene-based analyses.

Supplementary Figure S5. Quantile-quantile and regional association plot for CAKUT seqGWAS.

Supplementary Figure S6. European-only CAKUT seqGWAS.

Supplementary Figure S7. Statistical power for CAKUT GWAS.

Supplementary Figure S8. Meta-analysis of 727 cases and 434,900 controls with congenital obstructive uropathy.

Supplementary Figure S9. seqGWAS of 97 bladder exstrophy cases and 22,037 ancestry-matched controls.

Supplementary Figure S10. Meta-analysis of 725 cases and 29,389 controls with bladder exstrophy.

Supplementary Table S1. List of human phenotype ontology (HPO) codes used to stratify CAKUT cases.

Supplementary Table S2. List of SNOMED-CT and ICD-10 codes used to filter out controls with known kidney disease.

Supplementary Table S3. Genetic and phenotypic data for CAKUT cases with a genetic diagnosis.

Supplementary Table S4. Predictors of a genetic diagnosis in CAKUT.

Supplementary Table S5. Summary statistics from the rare variant gene-based analysis of 813 cases and 25,205 ancestry-matched controls.

Supplementary Table S6. Summary statistics from the rare variant gene-based analysis of 237 kidney anomaly cases and 22,733 ancestry-matched controls.

Supplementary Table S7. Summary statistics from the rare variant gene-based analysis of 177 congenital obstructive uropathy cases and 24,451 ancestry-matched controls.

Supplementary Table S8. Summary statistics from the rare variant gene-based analysis of 112 cystic dysplasia cases and 24,084 ancestry-matched controls.

Supplementary Table S9. Summary statistics from the rare variant gene-based analysis of 174 VUR cases and 22,562 ancestry-matched controls.

Supplementary Table S10. Summary statistics from the rare variant gene-based analysis of 132 PUV cases and 23,727 ancestry-matched controls.

Supplementary Table S11. Summary statistics from the rare variant gene-based analysis of 97 bladder exstrophy cases and 22,037 ancestry-matched controls.

Supplementary Table S12. Summary statistics from the rare variant gene-based analysis of 611 European CAKUT cases and 19,576 ancestry-matched controls.

Supplementary Table S13. Pathway analysis for 813 CAKUT cases versus 25,205 ancestry-matched controls.

Supplementary Table S14. Lead variants from bladder exstrophy seqGWAS.

Supplementary Table S15. Heritability estimation in 623 European CAKUT cases and 20,060 controls using GREML-LDMS.

Supplementary Table S16. Variants and weights used for PUV Genomic Risk Score.

Supplementary References

References

1. Wühl E, van Stralen KJ, Wanner C, et al. Renal replacement therapy for rare diseases affecting the kidney: an analysis of the ERA-EDTA Registry. *Nephrol Dial Transplant*. 2014;29 Suppl 4(suppl 4):iv1-8.

2. Wühl E, van Stralen KJ, Verrina E, et al. Timing and outcome of renal replacement therapy in patients with congenital malformations of the kidney and urinary tract. *Clin J Am Soc Nephrol*. 2013;8(1):67-74.

3. Viswanathan A, Dawman L, Tiewsoh K, Saxena AK, Dutta S, Suri D. Screening of renal anomalies in first-degree relatives of children diagnosed with non-syndromic congenital anomalies of kidney and urinary tract. *Clin Exp Nephrol*. 2021;25(2):184-190.

4. Bulum B, Birsin Özçakar Z, Üstüner E, et al. High frequency of kidney and urinary tract anomalies in asymptomatic first-degree relatives of patients with CAKUT. *Pediatric Nephrology*. 2013;28(11):2143-2147. doi:10.1007/s00467-013-2530-8

5. Manoharan A, Krishnamurthy S, Sivamurukan P, Ananthakrishnan R, Jindal B. Screening for renal and urinary tract anomalies in asymptomatic first degree relatives of children with Congenital anomalies of the kidney and urinary tract (CAKUT). *Indian J Pediatr*. 2020;87(9):686-691.

6. van der Ven AT, Vivante A, Hildebrandt F. Novel Insights into the Pathogenesis of Monogenic Congenital Anomalies of the Kidney and Urinary Tract. *J Am Soc Nephrol*. 2018;29(1):36-50.

7. Verbitsky M, Westland R, Perez A, et al. The copy number variation landscape of congenital anomalies of the kidney and urinary tract. *Nat Genet*. 2019;51(1):117-127.

8. Caruana G, Wong MN, Walker A, et al. Copy-number variation associated with congenital anomalies of the kidney and urinary tract. *Pediatr Nephrol*. 2015;30(3):487-495.

9. Sanna-Cherchi S, Kiryluk K, Burgess KE, et al. Copy-number disorders are a common cause of congenital kidney malformations. *Am J Hum Genet*. 2012;91(6):987-997.

10. Nicolaou N, Pulit SL, Nijman IJ, et al. Prioritization and burden analysis of rare variants in 208 candidate genes suggest they do not play a major role in CAKUT. *Kidney Int*. 2016;89(2):476-486.

11. van der Ven AT, Connaughton DM, Ityel H, et al. Whole-Exome Sequencing Identifies Causative Mutations in Families with Congenital Anomalies of the Kidney and Urinary Tract. *J Am Soc Nephrol*. 2018;29(9):2348-2361.

12. Hwang DY, Dworschak GC, Kohl S, et al. Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. *Kidney Int*. 2014;85(6):1429-1433.

13. Kohl S, Hwang DY, Dworschak GC, et al. Mild recessive mutations in six Fraser syndrome-related

genes cause isolated congenital anomalies of the kidney and urinary tract. *J Am Soc Nephrol*. 2014;25(9):1917-1922.

14. Ahn YH, Lee C, Kim NKD, et al. Targeted exome sequencing provided comprehensive genetic diagnosis of congenital anomalies of the kidney and urinary tract. *J Clin Med*. 2020;9(3):751.

15. Jung J, Lee JH, Seo GH, et al. Genetic diagnosis of kidney disease by whole exome sequencing and its clinical application. *Clin Genet*. 2023;104(3):298-312.

16. Riedhammer KM, Ćomić J, Tasic V, et al. Exome sequencing in individuals with congenital anomalies of the kidney and urinary tract (CAKUT): a single-center experience. *Eur J Hum Genet*. 2023;31(6):674-680.

17. Rao J, Liu X, Mao J, et al. Genetic spectrum of renal disease for 1001 Chinese children based on a multicenter registration system. *Clin Genet*. 2019;96(5):402-410.

18. Groopman EE, Marasa M, Cameron-Christie S, et al. Diagnostic Utility of Exome Sequencing for Kidney Disease. *N Engl J Med*. 2019;380(2):142-151.

19. Chan MMY, Sadeghi-Alavijeh O, Lopes FM, et al. Diverse ancestry whole-genome sequencing association study identifies TBX5 and PTK7 as susceptibility genes for posterior urethral valves. *Elife*. 2022;11. doi:10.7554/eLife.74777

20. Verbitsky M, Krithivasan P, Batourina E, et al. Copy number variant analysis and genome-wide association study identify loci with large effect for vesicoureteral reflux. *J Am Soc Nephrol*. 2021;32(4):805-820.

21. Mingardo E, Beaman G, Grote P, et al. A genome-wide association study with tissue transcriptomics identifies genetic drivers for classic bladder exstrophy. *Commun Biol*. 2022;5(1):1203.

22. Parimi M, Nitsch D. A Systematic Review and Meta-Analysis of Diabetes During Pregnancy and Congenital Genitourinary Abnormalities. *Kidney Int Rep*. 2020;5(5):678-693.

23. Jadresić L, Au H, Woodhouse C, Nitsch D. Pre-pregnancy obesity and risk of congenital abnormalities of the kidney and urinary tract (CAKUT)-systematic review, meta-analysis and ecological study. *Pediatr Nephrol*. 2021;36(1):119-132.

24. Macumber I, Schwartz S, Leca N. Maternal obesity is associated with congenital anomalies of the kidney and urinary tract in offspring. *Pediatr Nephrol*. 2017;32(4):635-642.

25. Groen In ’t Woud S, Renkema KY, Schreuder MF, et al. Maternal risk factors involved in specific congenital anomalies of the kidney and urinary tract: A case-control study. *Birth Defects Res A Clin Mol Teratol*. 2016;106(7):596-603.

26. Groen In ’t Woud S, van Gelder MMHJ, van Rooij IALM, et al. Genetic and environmental factors driving congenital solitary functioning kidney. *Nephrol Dial Transplant*. 2024;39(3):463-472.

27. Turro E, Astle WJ, Megy K, et al. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*. 2020;583(7814):96-102.

28. Lelieveld SH, Spielmann M, Mundlos S, Veltman JA, Gilissen C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum Mutat*. 2015;36(8):815-822.

29. The National Genomic Research Library v5.1, Genomics England. Published online Spring 5, 2020. doi:10.6084/m9.figshare.4530893.v7

30. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424.

31. Smedley D, Jacobsen JOB, Jäger M, et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat Protoc*. 2015;10(12):2004-2015.

32. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-443.

33. Lee S, Emond MJ, Bamshad MJ, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*. 2012;91(2):224-237.

34. Zhou W, Bi W, Zhao Z, et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat Genet*. Published online September 22, 2022. doi:10.1038/s41588-022-01178-w

35. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am J Hum Genet*. 2016;99(4):877-885.

36. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50(9):1335-1341.

37. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613(7944):508-518.

38. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26(17):2190-2191.

39. Yang J, Bakshi A, Zhu Z, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nat Genet*. 2015;47(10):1114-1120.

40. Speed D, Balding DJ. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat Genet*. 2019;51(2):277-284.

41. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence

among relatives. *Ann Hum Genet*. 1965;29(1):51-76.

42. Public Health England. *National Congenital Anomaly and Rare Disease Registration Service. Congenital Anomaly Statistics 2019*. Dandy Booksellers; 2022.

43. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235.

44. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10(1):1776.

45. Lambert SA, Wingfield B, Gibson JT, et al. Enhancing the Polygenic Score Catalog with tools for score calculation and ancestry normalization. *Nat Genet*. 2024;56(10):1989-1994.

46. Connaughton DM, Kennedy C, Shril S, et al. Monogenic causes of chronic kidney disease in adults. *Kidney Int*. 2019;95(4):914-928.

47. Mann N, Braun DA, Amann K, et al. Whole-exome sequencing enables a precision medicine approach for kidney transplant recipients. *J Am Soc Nephrol*. 2019;30(2):201-215.

48. Jayasinghe K, Stark Z, Kerr PG, et al. Clinical impact of genomic testing in patients with suspected monogenic kidney disease. *Genet Med*. 2021;23(1):183-191.

49. Farin HF, Bussen M, Schmidt MK, Singh MK, Schuster-Gossler K, Kispert A. Transcriptional repression by the T-box proteins Tbx18 and Tbx15 depends on Groucho corepressors. *J Biol Chem*. 2007;282(35):25748-25759.

50. Stolz JR, Foote KM, Veenstra-Knol HE, et al. Clustered mutations in the GRIK2 kainate receptor subunit gene underlie diverse neurodevelopmental disorders. *Am J Hum Genet*. 2021;108(9):1692-1709.

51. Inoue R, Hirohashi Y, Kitamura H, et al. GRIK2 has a role in the maintenance of urothelial carcinoma stem-like cells, and its expression is associated with poorer prognosis. *Oncotarget*. 2017;8(17):28826-28839.

52. Bantounas I, Rooney KM, Lopes FM, et al. Human pluripotent stem cell-derived kidney organoids reveal tubular epithelial pathobiology of heterozygous HNF1B-associated dysplastic kidney malformations. *Stem Cell Reports*. 2024;19(6):859-876.

53. Heidet L, Decramer S, Pawtowski A, et al. Spectrum of HNF1B mutations in a large cohort of patients who harbor renal diseases. *Clin J Am Soc Nephrol*. 2010;5(6):1079-1090.

54. Weber S, Moriniere V, Knüppel T, et al. Prevalence of mutations in renal developmental genes in children with renal hypodysplasia: results of the ESCAPE study. *J Am Soc Nephrol*. 2006;17(10):2864-2870.

55. Thomas R, Sanna-Cherchi S, Warady BA, Furth SL, Kaskel FJ, Gharavi AG. HNF1B and PAX2 mutations are a common cause of renal hypodysplasia in the CKiD cohort. *Pediatr Nephrol*.

2011;26(6):897-903.

56. Vivante A, Hwang DY, Kohl S, et al. Exome sequencing discerns syndromes in patients from consanguineous families with congenital anomalies of the kidneys and urinary tract. *J Am Soc Nephrol*. 2017;28(1):69-75.

57. Ahram DF, Lim TY, Ke J, et al. Rare single nucleotide and copy number variants and the etiology of congenital obstructive uropathy: Implications for genetic diagnosis. *J Am Soc Nephrol*. 2023;34(6):1105-1119.

58. Pandey R, Brennan NF, Trachana K, et al. A meta-analysis of diagnostic yield and clinical utility of genome and exome sequencing in pediatric rare and undiagnosed genetic diseases. *Genet Med*. 2025;27(6):101398.

59. Draaken M, Knapp M, Pennimpede T, et al. Genome-wide association study and meta-analysis identify ISL1 as genome-wide significant susceptibility gene for bladder exstrophy. *PLoS Genet*. 2015;11(3):e1005024.

60. Zhang R, Knapp M, Suzuki K, et al. ISL1 is a major susceptibility gene for classic bladder exstrophy and a regulator of urinary tract development. *Sci Rep*. 2017;7:42170.

61. Darlow JM, Darlay R, Dobson MG, et al. Genome-wide linkage and association study implicates the 10q26 region as a major genetic contributor to primary nonsyndromic vesicoureteric reflux. *Sci Rep*. 2017;7(1):14595.

62. Cordell HJ, Darlay R, Charoen P, et al. Whole-genome linkage and association scan in primary, nonsyndromic vesicoureteric reflux. *J Am Soc Nephrol*. 2010;21(1):113-123.

63. Stark Z, Foulger RE, Williams E, et al. Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *Am J Hum Genet*. 2021;108(9):1551-1557.

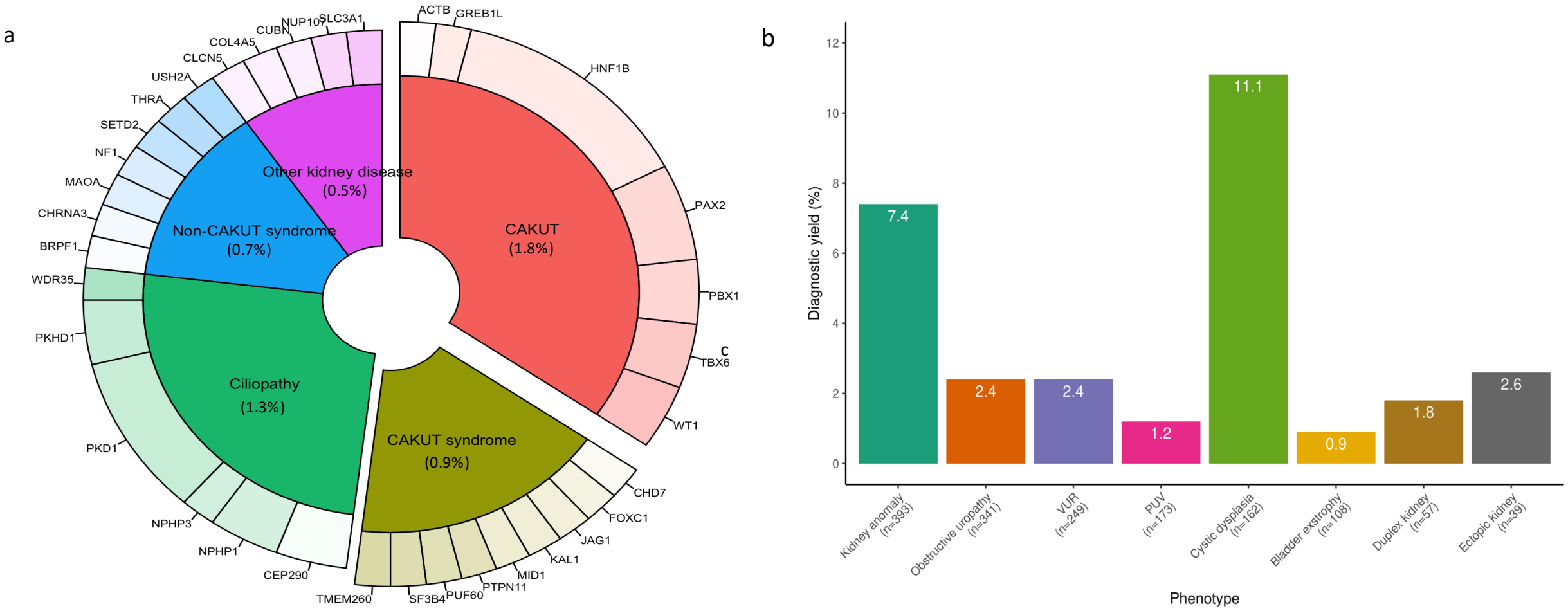
64. Hnisz D, Abraham BJ, Lee TI, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155(4):934-947.

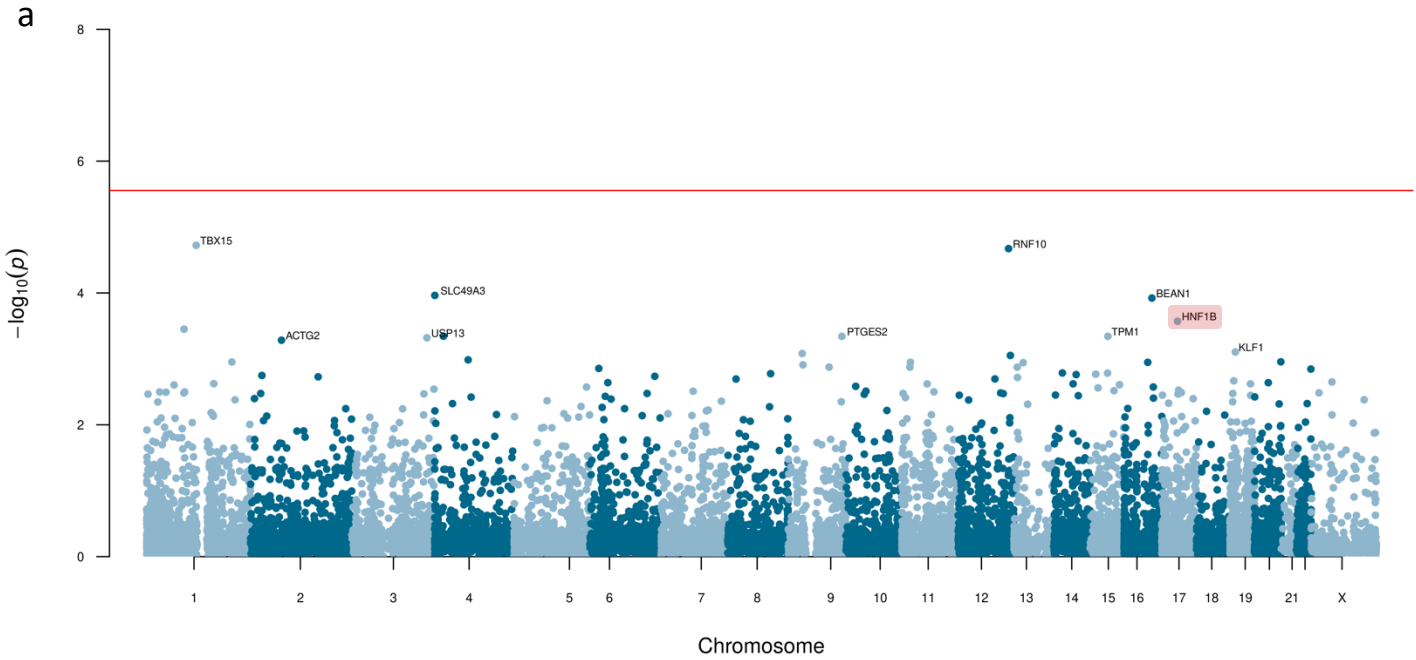
Legends

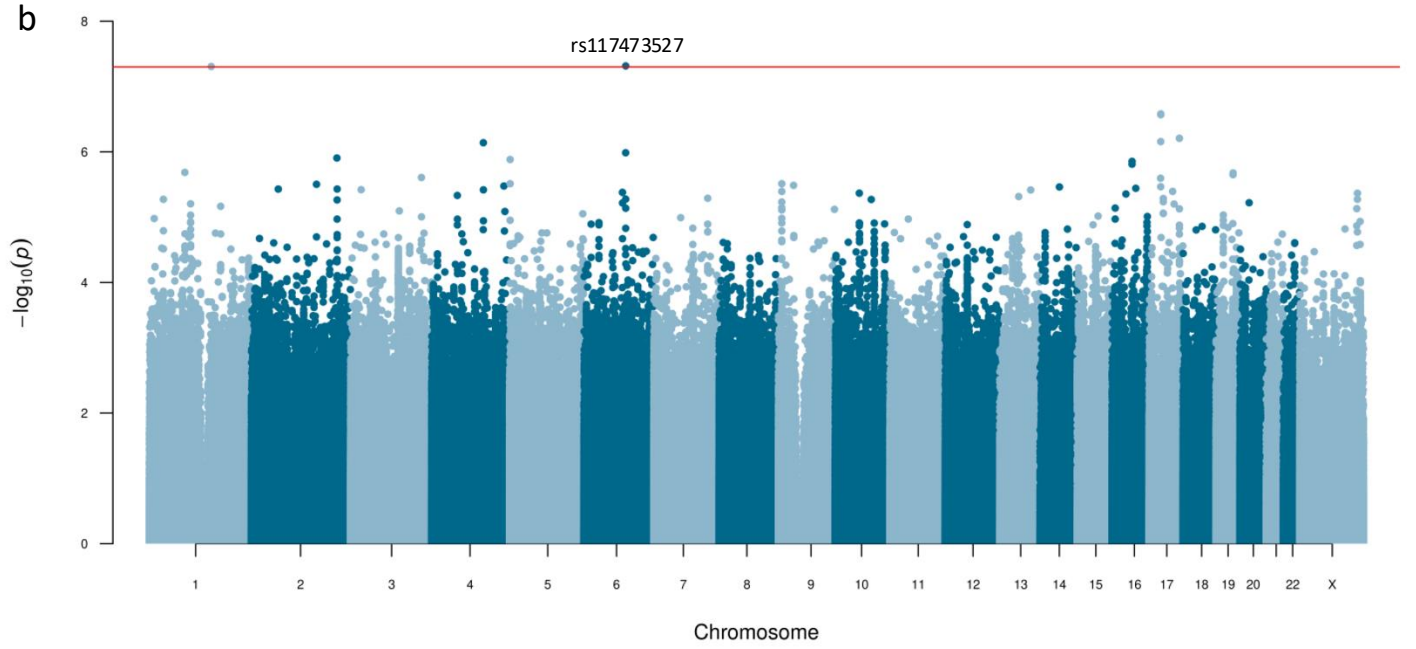
Figure 1. Genomic testing in 1,052 patients with CAKUT. a) Disease-causing genes identified in 4.9% of the cohort. Percentages relate to the entire cohort of 1,052 patients. b) Diagnostic yield by phenotype. CAKUT, congenital anomalies of the kidneys and urinary tract; VUR, vesico-ureteral reflux; PUV, posterior urethral valves.

Figure 2. Rare and common variant association testing in 813 CAKUT cases and 25,205 ancestry-matched controls. a) Manhattan plot of gene-based rare variant association testing using SAIGE-GENE+. Qualifying variants were loss-of-function and likely deleterious missense variants with minor allele frequency < 0.01%. Each dot represents a gene. The red line indicates the exome-wide significance threshold of $P=2.58\times10^{-6}$. Genes with $P < 10^{-3}$ are labelled and known CAKUT-gene associations are highlighted in red. b) Manhattan plot of single variant association testing using SAIGE across 10.1 million variants with MAF ≥ 0.5%. Each dot represents a variant. The red line indicates the genome-wide significance threshold of $P=5\times10^{-8}$.

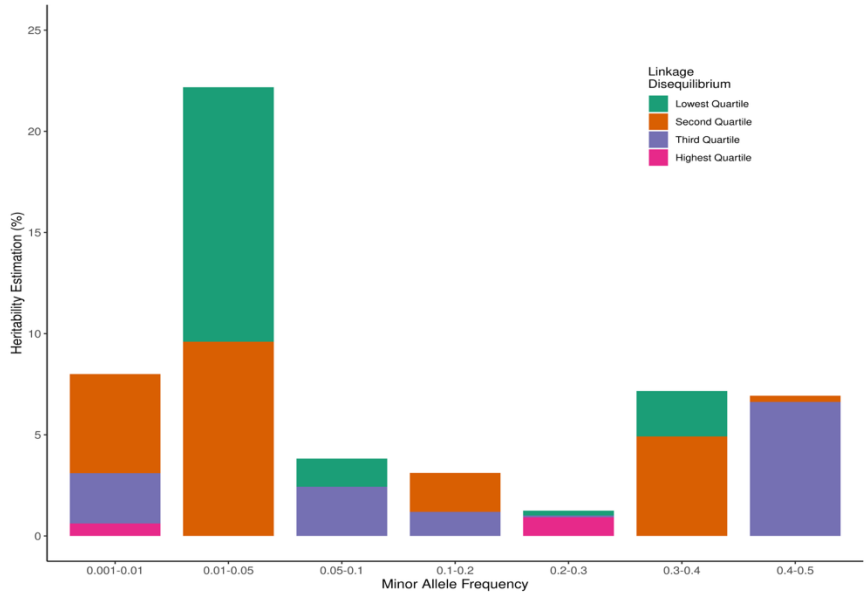
Figure 3. Common variants explain some of the phenotypic variance seen in CAKUT. a) Partitioning of heritability by minor allele frequency (MAF) and linkage disequilibrium in 623 CAKUT cases and 20,060 controls of European ancestry. Narrow-sense heritability (h^2) is represented using the liability threshold model based on a population disease prevalence of 1 in 500. Variants with MAF between 1-5% contribute most of the heritability. b) Enrichment of common variant heritability across functional annotations estimated using the European CAKUT seqGWAS summary statistics. The dashed grey line represents nominal *P*-value of 0.05 but no annotations were significantly enriched using a false discovery rate of 10%. Super-enhancers are large clusters of transcriptional enhancers that drive expression of genes that define cell identity from Hsinz et al.[64] c) Density plot with the distribution of the European PUV-GRS in an independent cohort of 77 cases (AGORA_PUV) and 2,746 controls. *P<0.001. PUV, posterior urethral valves; PRS, polygenic risk score.


a
Other kidney disease
(0.5%)
Non-CAKUT syndrome
(0.7%)
Ciliopathy
(1.3%)
CAKUT
(1.8%)
CAKUT syndrome
(0.9%)
SLC3A1
NUP107
CUBN
COL4A5
CLCN5
USH2A
THRA
SETD2
NF1
MAOA
CHRNA3
BRPF1
WDR35
PKHD1
PKD1
NPHP3
NPHP1
CEP290
TMEM260
SF3B4
PUF60
PTPN11
MID1
KAL1
JAG1
FOXC1
CHD7
WT1
TBX6
PBX1
PAX2
HNF1B
GREB1L
ACTB
b
Diagnostic yield (%)
0
2
4
6
8
10
12
7.4
2.4
2.4
1.2
11.1
0.9
1.8
2.6
Kidney anomaly (n=393)
Obstructive uropathy (n=341)
VUR (n=249)
PUV (n=173)
Cystic dysplasia (n=162)
Bladder exstrophy (n=108)
Duplex kidney (n=57)
Ectopic kidney (n=39)
Phenotype

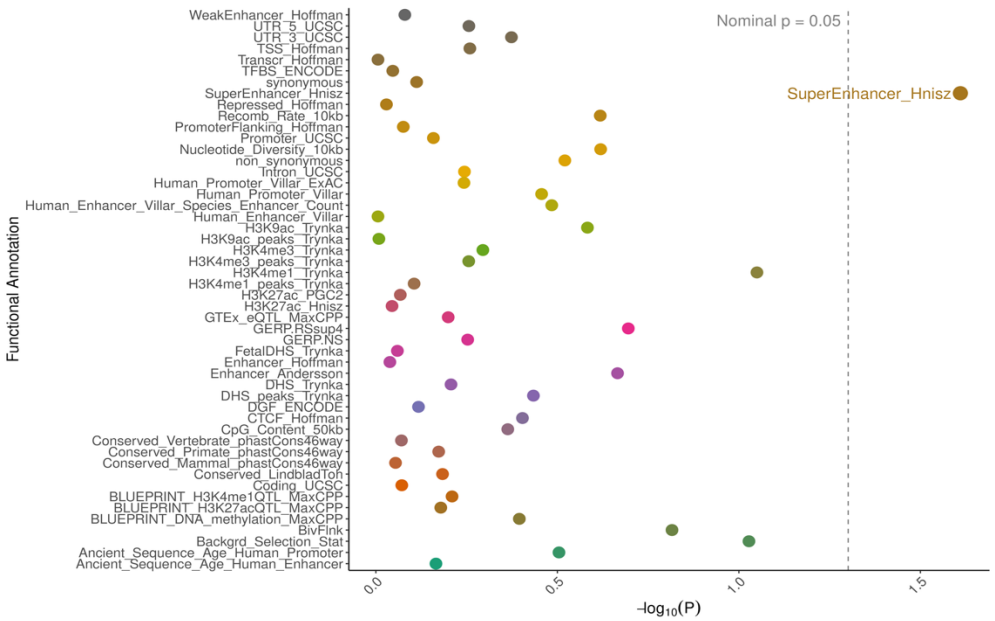

a
−log10(p)
8
6
4
2
0
TBX15
ACTG2
SLC49A3
USP13
PTGES2
RNF10
BEAN1
HNF1B
TPM1
KLF1
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
21
X
Chromosome


b
−log10(p)
8
6
4
2
0
rs117473527
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
22
X
Chromosome

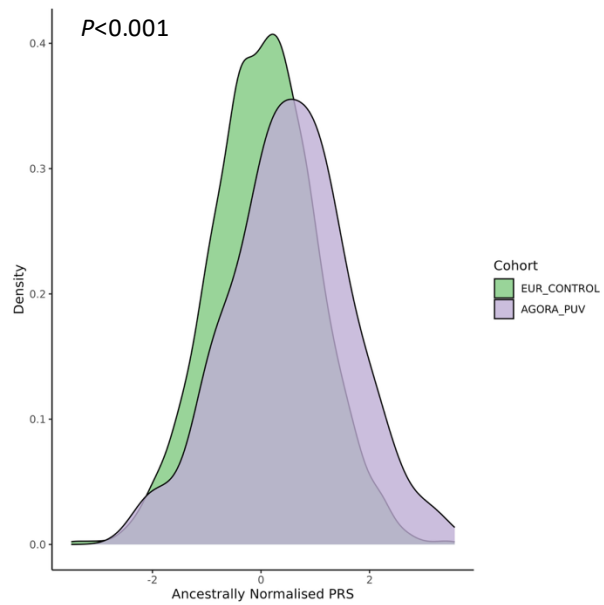
a


Linkage Disequilibrium
Lowest Quartile
Second Quartile
Third Quartile
Highest Quartile
Heritability Estimation (%)
0
5
10
15
20
25
0.001-0.01
0.01-0.05
0.05-0.1
0.1-0.2
0.2-0.3
0.3-0.4
0.4-0.5
Minor Allele Frequency

b


Nominal p = 0.05
SuperEnhancer_Hnisz
Functional Annotation
WeakEnhancer_Hoffman
UTR_5_UCSC
UTR_3_UCSC
TSS_Hoffman
Transcr_Hoffman
TFBS_ENCODE
synonymous
SuperEnhancer_Hnisz
Repressed_Hoffman
Recomb_Rate_10kb
PromoterFlanking_Hoffman
Promoter_UCSC
Nucleotide_Diversity_10kb
non_synonymous
Intron_UCSC
Human_Promoter_Villar_ExAC
Human_Promoter_Villar
Human_Enhancer_Villar_Species_Enhancer_Count
Human_Enhancer_Villar
H3K9ac_Trynka
H3K9ac_peaks_Trynka
H3K4me3_Trynka
H3K4me3_peaks_Trynka
H3K4me1_Trynka
H3K4me1_peaks_Trynka
H3K27ac_PGC2
H3K27ac_Hnisz
GTEx_eQTL_MaxCPP
GERP.RSsup4
GERP.NS
FetalDHS_Trynka
Enhancer_Hoffman
Enhancer_Andersson
DHS_Trynka
DHS_peaks_Trynka
DGF_ENCODE
CTCF_Hoffman
CpG_Content_50kb
Conserved_Vertebrate_phastCons46way
Conserved_Primate_phastCons46way
Conserved_Mammal_phastCons46way
Conserved_Lindblad-Toh
Coding_UCSC
BLUEPRINT_H3K4me1QTL_MaxCPP
BLUEPRINT_H3K27acQTL_MaxCPP
BLUEPRINT_DNA_methylation_MaxCPP
BivFlnk
Backgrd_Selection_Stat
Ancient_Sequence_Age_Human_Promoter
Ancient_Sequence_Age_Human_Enhancer
0.0
0.5
1.0
1.5
−log10(P)

c


P<0.001
Density
0.0
0.1
0.2
0.3
0.4
-2
0
2
Ancestrally Normalised PRS
Cohort
EUR_CONTROL
AGORA_PUV