medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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], 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], Horia C Stanescu[1], Adam P Levine[1,12], Detlef Böckenhauer[1,13] 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]Research Department of Pathology, University College London, UK.
[13]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, Hammersmith Campus, Imperial College London, W12 0NN, UK. Email: m.chan@lms.mrc.ac.uk

Running head: WGS in CAKUT

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Congenital anomalies of the kidneys and urinary tract (CAKUT) are the commonest cause of kidney failure in children and young adults with over 50 monogenic causes identified, largely in cohorts enriched for familial, syndromic, or consanguineous disease. We sought to better characterise the genomic architecture of these conditions using whole genome sequencing data from 992 unrelated individuals recruited to the UK's 100,000 Genomes Project. The overall diagnostic yield was 4.3% with family history (P=7.4x10^{-3}; OR 2.7; 95% CI 1.3-5.4) and extra-renal features (P=2.0x10^{-4}; OR 3.4; 95% CI 1.8-6.6) independently predicting a monogenic diagnosis. Diagnostic yield was highest in cystic kidney dysplasia (10.7%) and kidney agenesis/hypodysplasia (5.9%). Exome-wide rare variant and genome-wide common variant (minor allele frequency ≥ 0.1%) testing was performed in a subset of 813 patients and 25,205 ancestry-matched controls with significant association detected at rs117473527 (P=3.93x10^{-8}; OR 3.17; 95% CI 2.10-4.78; MAF 0.02). Heritability analysis estimated common variants explain 23% (standard error 11%) of phenotypic variance in those with European ancestry. Comparison of phenotype-specific genomic risk scores (GRS) demonstrated shared polygenic aetiology between upper urinary tract phenotypes but distinct patterns for both posterior urethral valves (PUV) and bladder exstrophy. A PUV-GRS consisting of 36,106 variants was validated in an independent European cohort of 77 cases and 2,746 controls (P=1x10^{-4}) accounting for 37% of phenotypic variance. Together, these data demonstrate the importance of non-Mendelian genomic factors in the pathogenesis of CAKUT, evidenced by the fact that only a minority of patients in this large, unselected cohort received a monogenic diagnosis and that a substantial proportion of heritability can be attributed to common variation.

Lay Summary

This study looks at the DNA of patients with congenital anomalies of the kidneys and urinary tract (CAKUT), the most common cause of kidney failure in children. It finds that single gene (monogenic) causes are uncommon, and that genetic testing is more likely to find a cause in people with kidneys that have not formed properly (especially those with cysts), those with affected family members or with involvement of other organ systems. The study also supports the idea that many different common DNA changes (polygenic) together contribute to these conditions, particularly in posterior urethral valves, which is the leading cause of kidney failure in boys. Larger studies are needed to better understand the DNA changes and biological processes involved in kidney and urinary tract development and malformations.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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, less 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 via a broad range of perturbations.

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 992 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 significant polygenic contribution to these phenotypically diverse structural malformations.

Methods

The 100,000 Genomes Project

The 100,000 Genomes Project (100KGP) dataset (version 10) consists of WGS data, clinical phenotypes encoded using Human Phenotype Ontology (HPO) codes, and retrospective and

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

prospectively ascertained hospital records for 89,139 individuals with cancer or rare disease, as well as their unaffected relatives.[29] Ethical approval for the 100KGP was granted by the Research Ethics Committee for East of England – Cambridge South (REC Ref 14/EE/1112). Full details on 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. Testing for other CAKUT genes was not routinely undertaken. Those with a known genetic or chromosomal abnormality were excluded. Individuals with kidney cysts (not dysplasia) were recruited to the 100KGP 'cystic kidney disease' cohort.[30] Five probands had documented negative genetic testing prior to recruitment with either targeted sequencing and/or microarray-based comparative genomic hybridisation (array CGH).

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 subset of unrelated and ancestry-matched cases and controls was then 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. Rare (minor allele frequency [MAF] < 1% for autosomal recessive and MAF < 0.1% for autosomal dominant inheritance), high or moderate impact variants intersecting with an expert-curated panel of 57 CAKUT-associated genes (https://panelapp.genomicsengland.co.uk/panels/234/)[31] and CNV losses impacting 17q12 (encompassing *HNF1B*), 16p11.2 (encompassing *TBX6*), and *NPHP1* were reviewed by a multi-disciplinary team to determine pathogenicity (Supplementary Methods). In patients yet to undergo multi-disciplinary review, variants were prioritised using Exomiser[32] and CNV calls for 17q12, 16p11.2 and *NPHP1* manually reviewed to identify those meeting American College of Medical Genetics and Genomics (ACMG)[33] criteria for pathogenicity.

Rare variant gene-based analysis

Coding single-nucleotide variants (SNVs) and indels with MAF < 0.1% in gnomAD[34] (v3) annotated with missense, in-frame insertion, in-frame deletion, start loss, stop gain, frameshift, splice donor or

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

splice acceptor were extracted. Variants were filtered using a CADD[35] Phred score (version 1.5) ≥ 20, corresponding to the top 1% of predicted deleterious variants in the genome. 'High confidence' loss-of-function variants (stop gain, splice site and frameshift) were determined by LOFTEE.[34] Variant and sample quality control is detailed in the Supplementary Methods. Rare variants were aggregated by gene and association testing carried out using SKAT-O[36] implemented by SAIGE-GENE[37] (version 0.42.1). 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 whole-genome sequencing genome-wide association study (seqGWAS) of SNVs/indels with MAF ≥ 0.1% was performed with SAIGE[38] (version 0.42.1). seqGWAS was first applied to 813 CAKUT cases and 25,205 ancestry-matched controls across 19,193,915 variants and then stratified by phenotype group (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 in a subset of the 100KGP cohort with genetically defined European ancestry (623 cases and 20,060 controls). First, using GREML-LDMS[41] across 22,052,734 WGS variants with MAF ≥ 0.1% and second using SumHer[42] across 1,160,273 variants with MAF ≥ 1%. The observed heritability was transformed to an underlying continuous liability threshold model[43] adjusting for the ratio of cases to controls and an estimated population prevalence of disease of 0.2%.[44]

Genomic Risk Scores

Genomic risk scores (GRS) for each phenotype were generated from 100KGP GWAS summary statistics (MAF ≥ 0.1%) using PRSice-2[45] (version 2.3.3) under an additive model and applied to the remaining phenotypes to determine the phenotypic variance explained by each GRS.

Replication Cohort

WGS data from 127 individuals with obstructive uropathy and 77 individuals with PUV from the Dutch AGORA biobank[46] were used for the obstructive uropathy rare variant replication analysis and validation of the PUV-GRS, respectively. 2,746 male Europeans from the 100KGP with non-urinary tract cancers were used as controls.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Results

A total of 1,091 individuals with CAKUT were recruited to the 100KGP, of whom 992 were probands: 32.5% were recruited as singletons, 29.8% as duos and 30.8% as trios. This is the largest cohort of patients with CAKUT to have undergone WGS to date. Table 1 summarizes the clinical and demographic data for the probands. The cohort had a median age of 17 years, 13.3% had an affected first-degree relative, and 28.5% had extra-renal manifestations. The most frequently reported HPO terms in the cohort were hydronephrosis (27.0%), vesicoureteral reflux (VUR) (22.9%) and abnormality of the bladder (19.0%) (Figure 1a).

Table 1. Demographics and clinical characteristics of the CAKUT cohort (n=992). Some patients had more than one phenotype. VUR, vesico-ureteral reflux; PUV, posterior urethral valves.

Median age in years (range)		17 (3-82)
Males (%)		627 (63.2)
Self-reported ethnicity		
	White (%)	634 (63.9)
	South Asian (%)	92 (9.3)
	Mixed/Other (%)	52 (5.2)
	Black (%)	28 (2.8)
	Chinese (%)	3 (0.3)
	Unknown (%)	183 (18.4)
Phenotype		
	Kidney anomaly (%)	355 (35.8)
	Obstructive uropathy (%)	228 (23.0)
	VUR (%)	227 (22.9)
	PUV (%)	169 (17.0)
	Cystic dysplasia (%)	149 (15.0)
	Bladder exstrophy (%)	108 (10.9)
	Duplex kidney (%)	57 (5.7)
	Ectopic kidney (%)	36 (3.6)
Extrarenal manifestations (%)		283 (28.5)
	Neurodevelopmental delay (%)	61 (6.1)
	Skeletal anomalies (%)	40 (4.0)
	Hearing impairment (%)	27 (2.7)
	Endocrine (%)	22 (2.2)
	Eye anomalies (%)	22 (2.2)
	Respiratory (%)	22 (2.2)
	Cardiac anomalies (%)	21 (2.1)
Family history of CAKUT (%)		132 (13.3)
Reported consanguinity (%)		58 (5.8)

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Kidney failure (%)	211 (21.2)
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. Of 992 probands recruited, 975 (98.3%) had undergone multi-disciplinary review of candidate variants from 57 CAKUT-associated genes, classified according to ACMG[33] guidelines. Overall, 4.3% (43/992) had a genomic diagnosis (Supplementary Table S3), with 74.4% (32/43) 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 6.0% (43/715).

One third (32.6%; 14/43) of those with a genomic diagnosis had a pathogenic or likely pathogenic variant in a known non-syndromic CAKUT gene, most frequently in *HNF1B* (six individuals including four with 17q12 deletions). Additional disease-causing variants were reported in *PAX2* (2), *PBX1* (2), *TBX6* (two individuals including one with 16p11.2 microdeletion), *GREB1L* (1) and *WT1* (1) (Figure 1b; Supplementary Table S3). Another third (32.6%; 14/43) 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), *NPHP3* (1), *COL4A5* (1), *CUBN* (1), *SLC3A1* (1), and *CLCN5* (1). These likely represent phenocopies in the case of cystic disease, or possible mis-phenotyping. An additional 25.6% (11/43) 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 11 patients with a confirmed syndrome, the five individuals with non-CAKUT diagnoses who may have been mis-phenotyped (*NF1, CLCN5, CUBN and SLC3A1*), and eleven more with cystic/ciliopathy disease (*PKD1, PKHD1, NPHP3, CEP290* and *COL4A5*) the diagnostic yield falls to just 2.3% (16/689). These data demonstrate that the prevalence of 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 (10.7% and 5.9%, respectively) (Figure 1c). Pathogenic variants affecting *HNF1B* (including 17q12 deletions) accounted for 25% (4/16) of genomic diagnoses in the cystic dysplasia cohort, however most patients (56.3%) were found to have a phenocopy; 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 80% (8/10) of genomic diagnoses were associated with a syndrome or extra-renal manifestations (Supplementary Table S3). These data suggest that monogenic causes of isolated CAKUT, in the absence of extra-renal clinical manifestations, are rare.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

The diagnostic yield in individuals < 18 years at the time of analysis was significantly higher than in those ≥ 18 years (6.6% vs 1.8%; Fisher's exact P=2.0x10^{-4}; odds ratio [OR] 4.0; 95% confidence interval [CI] 1.8-10.0), potentially reflecting more severe phenotypes or syndromes presenting at an earlier age. Consistent with previous reports from individuals with suspected monogenic kidney disease,[18,47–49] multivariable logistic regression analysis confirmed that extra-renal features (P=2.0x10^{-4}; OR 3.4; 95% CI 1.8-6.6) and family history (P=7.4x10^{-3}; OR 2.7; 95% CI 1.3-5.4) were independent predictors of a positive genomic diagnosis (Supplementary Table S4). These findings can help clinicians prioritize patients for genomic testing, particularly 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 potential candidate genes. 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 < 0.1%), predicted deleterious (protein-truncating, splice-site or missense with CADD score ≥ 20) SNVs and indels across 19,373 protein-coding genes and b) only loss-of-function (LoF) variants across 17,200 genes. We did not identify any genes that were significantly enriched for rare variation when 813 cases were compared to 25,205 ancestry-matched controls, after correction for multiple testing (Figure 2a and Supplementary Figure S3a; Supplementary Table S5). The median number of variants tested per gene was 46 (interquartile range [IQR] 50) and 6 (IQR 7), 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. With this sample size we had 80% power to detect a single gene enriched for loss of function variation accounting for a minimum of 5% of cases.

Four genes which did not reach stringent Bonferroni-adjusted exome-wide significance did however demonstrate an increased burden of likely deleterious variation with a false discovery rate (FDR) q-value < 0.05: *AUTS2* (activator of transcription and developmental regulator; 133 variants; P=3.74x10^{-6}), *ARHGAP5* (Rho GTPase activating protein 5; 160 variants; P=6.04x10^{-6}), known CAKUT gene *HNF1B* (63 variants; P=6.06x10^{-6}), and *ZNF879* (zinc finger protein 879; 43 variants; P=6.96x10^{-6}). Conditioning on the lead variant in each gene attenuated but did not abolish the observed signals (*AUTS2*, P=1.37x10^{-4}; HNF1B, P=2.87x10^{-4}; *ARHGAP5*, P=1.27x10^{-3}; *ZNF879*, P=3.20x10^{-5}), implying the associations are not being driven by single variants. *HNF1B* was not significantly enriched for LoF variation (P=0.06) suggesting that these types of variants are not the primary drivers of the observed signal. Excluding 32 individuals with a known monogenic diagnosis did not reveal any additional signals in the remaining 781 unsolved cases. Whilst these data confirm that *HNF1B* is one of the most common monogenic causes of CAKUT, the other three genes require replication in an independent cohort before association can be definitively established.

Figure 2. Manhattan plot of exome-wide gene-based rare, likely deleterious variant association testing for a) 813 CAKUT cases and 25,205 ancestry-matched controls and b) 177 cases with congenital obstructive uropathy and 24,451 ancestry-matched controls. Each dot represents a gene. The red line indicates the exome-wide significance threshold of P=2.58×10^{-6}. Genes with P < 10^{-4} are labelled and known disease-gene associations are highlighted in red.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

To try and increase power through more stringent case selection, the rare variant analysis was repeated stratifying by phenotype. In the congenital obstructive uropathy group consisting of 177 individuals with pelvi-ureteral junction obstruction (PUJO), vesico-ureteral junction obstruction (VUJO), congenital megaureter, hydronephrosis and hydroureter (but excluding those with PUV or bladder exstrophy) one gene achieved exome-wide significance when compared to 24,451 ancestry-matched controls across 19,406 genes (Figure 2b and Supplementary Figure S3b; Supplementary Table S6): *NCF2* (neutrophil cytosolic factor 2; 53 variants; P=2.55x10^{-6}). 5/177 (2.8%) of the cohort were heterozygous for a rare, likely deleterious *NCF2* variant compared with 84/24,451 (0.3%) controls. Two out of the five cases were heterozygous for the same missense variant (rs777251055) which was absent from controls and has an allele frequency of 6.57x10^{-6} in gnomAD[34] (version 4.0) with a CADD[35] of 22.4. Both individuals had kidney anomalies (renal hypoplasia and cystic renal dysplasia) in addition to hydronephrosis and hydroureter. Conditioning on rs777251055 attenuated but did not abolish the observed signal in *NCF2* (P=4.0x10^{-3}) indicating it is not being driven solely by this variant. Replication of this signal was sought using WGS data from a Dutch cohort with congenital obstructive uropathy (n=127) from the AGORA biobank but no likely deleterious variants in *NCF2* were identified in this European cohort. Interestingly, *NCF2* is associated with reduced risk of genital malignancy in the UK BioBank (P=3.51x10^{-7}; OR 0.12; 95% CI 0.04-0.39) indicating a potential role for this gene in other genitourinary phenotypes.[50] *MYOCD* (myocardin), a monogenic cause of autosomal dominant congenital megabladder was also enriched for rare variation but did not achieve exome-wide significance (90 variants; P=1.11x10^{-5}; Figure 2b). No other known monogenic causes of functional or anatomical lower urinary tract obstruction were enriched for rare coding variation in this cohort: *BNC2, P*=0.38*; FLNA, P*=1*; HPSE2, P*=0.59*; LRIG2, P*=0.47*; CHRM3, P*=0.71*; TSHZ3, P*=0.26.

Rare variant association testing of the remaining phenotypes did not identify statistically significant associations for a) likely deleterious missense and LoF or b) LoF only variant sets. Although *HNF1B* had suggestive evidence of enrichment in the overall CAKUT analysis (P=6.06x10^{-6}), it was not significant in individuals with cystic renal dysplasia (P=0.71), possibly due to limited power and the small size of the cohort (n=112). 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).

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. We therefore performed a sequencing-based GWAS (seqGWAS) in 813 cases and 25,205 ancestry-matched controls testing for association at 19,193,915 SNVs and indels with MAF $\geq$ 0.1% (Supplementary Figure S4a; Supplementary Table S12). The genomic inflation factor (λ) was 1.02 indicating minimal evidence of confounding by population stratification in this mixed-ancestry cohort (Supplementary Figure S4b). Two variants reached genome-wide significance. The first, rs117473527 (chr6:102155812:G>C) at 6q16.3 (P=3.93x10^{-8}; OR 3.17; 95% CI 2.10-4.78; MAF 0.02) (Supplementary Figure S4c) 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.[51,52] *Grick2* is expressed in the murine ureteric bud at E10.5-11.5 (Supplementary Figure S4d)

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

when it invades the metanephric mesenchyme, indicating it may be important during nephrogenesis. Replication of this signal in an independent cohort is needed to validate this association. The second genome-wide significant variant (an indel rs35251516 at 1q21.3; P=4.94x10^{-8}) was isolated, found in a low-complexity (difficult to sequence) region and therefore likely to be an artefact.

The absence of additional significant associations is likely the result of limited power related to the phenotypic heterogeneity of the cohort and sample size. With 813 cases (assuming a disease prevalence of 1 in 500) this analysis is 80% powered to detect association of common variants (MAF > 5%) with OR > 2 or low-frequency variants (MAF > 1%) with OR > 4.5 (Supplementary Figure S5). To increase power, we aggregated variants with MAF $\geq$ 0.1% 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 Tables S13-16) 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 S6; Supplementary Table S18). 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.

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 18,797,149 SNV/indels with MAF $\geq$ 0.1% for association in 97 individuals with CBE and 22,037 ancestry-matched controls (Supplementary Figure S7a; Supplementary Table S19). One isolated variant (rs1571885276 at 1q32.1) achieved genome-wide significance but again was in a low-complexity region suggesting a probable sequencing artefact. Suggestive evidence of association ($P < 10^{-6}$) was identified at 5q35.3, 20p11.22 and 1q32.1 (Supplementary Figure S7b and Supplementary Table S20). Of note, all three lead variants identified had MAF < 1% and a large effect size meaning they would not have been detected using a conventional genotyping and imputation approach. Replication of rs6106456 at 20p11.22 was sought in an independent cohort of 84 unrelated European patients with bladder exstrophy and 10,804 controls from the 100KGP. Although this variant did not replicate (Cochran-Armitage trend test P=0.52), the estimated power (0.67) was insufficient to definitively confirm or refute the

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 ≥ 1% failed to identify any additional novel loci (Supplementary Figure S8 and Supplementary Table S21).

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 1.48×10^{-3} (OR 1.62; 95% CI 1.21-2.17; 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 S7). A rare, intergenic indel rs550737686 (chr5:51494837:CCT>C) demonstrated strongest association ($P=2.35 \times 10^{-5}$; 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 1.28×10^{-4} (OR 6.26; 95% CI 2.92-13.44). It is not in linkage disequilibrium (LD) with the lead indel rs550737686 ($r^2=0.1$). rs550737686 was excluded from the fine-mapping analysis as it is absent from the 1000 Genomes data used to calculate LD. Other than proximity to *ISL1* and localization within the same topologically associating domain (TAD), there was no gene expression or chromatin interaction data to definitively link these potentially causal variants to a specific gene.

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 ≥ 0.1%) was 23% (SE 11%; Supplementary Table S22). 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 ≥ 1%) generated a comparable heritability estimate of 24% (standard deviation 15%).

An individual's genomic susceptibility to a disease can be estimated through a genomic risk score (GRS). A GRS combines information across multiple loci using a weighted sum of the number of risk alleles an individual carries. We hypothesized that each phenotype would have its own distinct GRS and sought to explore whether there was any overlap between groups. Phenotype-specific GRS were generated for each cohort using the mixed ancestry seqGWAS summary statistics (MAF ≥ 0.1%) and evaluated in the remaining phenotypes, removing individuals with overlapping HPO terms (Supplementary Table S23). The phenotypic variance explained by each cohort-specific GRS is shown in Figure 3b. These data show that whilst there is overlap in polygenic risk between kidney anomalies, cystic dysplasia, VUR and obstructive uropathy accounting for 5-10% of phenotypic variance; PUV and bladder exstrophy are genetically distinct from both each other and the upper tract anomalies, explaining <2% of variance seen in the other groups.

Finally, we generated a PUV-GRS using 36,106 variants with $P < 0.05$ from our previously published mixed-ancestry PUV seqGWAS[19] (MAF < 0.1%; 132 cases and 23,727 controls) and applied this to

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

WGS data from 77 individuals with PUV from the AGORA biobank and 2,746 European male controls from the 100KGP (Figure 3c). The PUV-GRS was significantly elevated in this independent European cohort compared to controls (empirical $P=1x10^{-4}$) and predicted to account for 37.4% of phenotypic variance in cases with an AUC of 0.95 (95% CI 0.93-0.98).

Discussion

This study represents the largest cohort of patients with CAKUT to have undergone clinical-grade WGS to date. A monogenic cause was identified in less than 5% of individuals, with a higher diagnostic yield observed in those with kidney agenesis/hypodysplasia, cystic dysplasia, a family history, and/or extra-renal features. This figure is lower than many previous studies using targeted gene panels or exome sequencing in cohorts enriched for consanguinity, severe disease, family history or extra-renal manifestations which report diagnostic yields up to 27%.[11,12,14,16,17,48,49,53–57] This is likely related to the broad definition of phenotypes recruited to 100KGP, including individuals with mild disease, perhaps offering a better representation of general nephrology practice. Whilst it may be possible that individuals with a clear monogenic cause were not recruited, genetic testing for CAKUT was not widespread in the UK prior to the 100KGP. In a cohort similar to this study, a panel of 208 genes had a diagnostic yield of 1.3% in 453 individuals with CAKUT.[10] Overall, these data demonstrate 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 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 modest for a rare variant association study, and d) use of a stringent Bonferroni adjustment for multiple testing. Interestingly, non-significant enrichment of known CAKUT genes *HNF1B* and *MYOCD* was seen in the whole and obstructive uropathy cohorts respectively, providing validation of the testing approach.

This is the first time a whole genome sequencing-based GWAS approach has been used in these phenotypes enabling both common and low-frequency variants to be interrogated. Although statistically significant association was detected at 6q16.3 downstream of *GRIK2*, further replication in an independent cohort is necessary. 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. This is reinforced by validation of a PUV-GRS demonstrating that common variation (in this context MAF $\geq$ 0.1%) explained 37% of phenotypic variance. These data suggest there are likely to be a significant number of low-frequency variants with effect sizes too small to be detected in a cohort of this size which may contribute to an individual's susceptibility to CAKUT. Estimates show a cohort of 5000 cases and 5000 controls has 80% power to detect association of variants with MAF > 1% and OR > 2. Larger, better-powered cohorts are therefore critical to identify these variants and the genes/pathways they impact to provide insight into the molecular mechanisms underlying these malformations.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Previous work has shown susceptibility to lower urinary tract phenotypes, including PUV,[19] bladder exstrophy,[21,58,59] and VUR,[20,60,61] is influenced by non-coding common variation. We expand these findings to show that whilst there is a shared common genomic background explaining up to 10% of heritability of upper tract phenotypes, the polygenic backgrounds of PUV and bladder exstrophy are distinct from both each other and the other phenotypes. This is consistent with what is observed clinically; kidney anomalies, VUR and congenital obstructive phenotypes often coexist whereas PUV and bladder exstrophy usually occur in isolation (although can lead to secondary VUR and subsequent kidney damage). The commonality in polygenic contributors to different upper tract anomalies mirrors the fact that different anomalies can also be associated with the same monogenic defect and points towards underlying shared transcriptional networks. Disruption of these shared gene-regulatory networks through environmental or epigenomic mechanisms may result in a unique phenotype depending on the timing and cell-type specificity of the insult during urinary tract development.

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. This is the largest cohort of patients with CAKUT to have undergone WGS and, by using cases and controls sequenced on the same platform, 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 reliant on the Genomics England pipeline.[31] This has been updated over time with CNV assessment and variant prioritization using Exomiser[32] which may have improved diagnostic yield in more recently analysed samples. In addition, the pipeline did not include novel small structural variants. Second, the analysis is dependent on accurate phenotyping with HPO terms which was carried out by the recruiting team and patient's clinician. 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, although we had 80% power to detect exome-wide significant association of a single gene explaining ≥ 10% of cases and common variants (MAF > 5%) with an OR > 4, weaker effects or genes accounting for a smaller proportion of unexplained 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 a) the lack of an identifiable monogenic cause in most cases, b) the heritability explained by commoner variants and c) the overlap in polygenic contributors to upper tract malformations and their distinction from those implicated in lower tract malformations. Larger studies are now needed to try and identify these individual 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.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

All the other authors declare no competing interests.

Data Statement

Supplementary Tables can be found at https://doi.org/10.5281/zenodo.13834253. 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/about-gecip/joining-research-community/. 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/. GWAS summary statistics will be made publicly available via the GWAS catalog on publication (https://www.ebi.ac.uk/gwas/).

Acknowledgements

This research was made possible through access to data in the National Genomic Research Library, which is managed by Genomics England Limited (a wholly owned company of the Department of Health and Social Care). The National Genomic Research Library holds data provided by patients and collected by the NHS as part of their care and data collected as part of their participation in research. The National Genomic Research Library is funded by the National Institute for Health Research and NHS England. The Wellcome Trust, Cancer Research UK and the Medical Research Council have also funded research infrastructure. The authors gratefully acknowledge the participation of the patients, and their families recruited to the 100KGP.

MMYC was supported by a Kidney Research UK Clinical Research Fellowship (TF_004_20161125) and the Medical Research Council (MR/Y008170/1). OSA is 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). APL is supported by an NIHR Academic Clinical Lectureship. DPG is supported by the St Peter's Trust for Kidney, Bladder and Prostate Research.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Supplementary Material

Supplementary Methods

Supplementary Figure S1. Study workflow.

Supplementary Figure S2. Ancestry matching of 817 cases to 25,718 controls.

Supplementary Figure S3. Quantile-quantile plot of rare variant gene-based analyses.

Supplementary Figure S4. seqGWAS of 813 CAKUT cases and 25,205 ancestry-matched controls.

Supplementary Figure S5. Statistical power for CAKUT GWAS.

Supplementary Figure S6. Meta-analysis of 727 cases and 434,900 controls with congenital obstructive uropathy.

Supplementary Figure S7. seqGWAS of 97 bladder exstrophy cases and 22,037 ancestry-matched controls.

Supplementary Figure S8. 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 43 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 177 congenital obstructive uropathy cases and 24,451 ancestry-matched controls.

Supplementary Table S7. Summary statistics from the rare variant gene-based analysis of 237 kidney anomaly cases and 22,733 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 of the sequencing based GWAS of 813 CAKUT cases versus 25,205 ancestry-matched controls.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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

Supplementary Table S14. Summary statistics of the sequencing based GWAS of 237 kidney anomaly cases and 22,733 ancestry-matched controls.

Supplementary Table S15. Summary statistics of the sequencing based GWAS of 112 cystic dysplasia cases and 24,084 ancestry-matched controls.

Supplementary Table S16. Summary statistics of the sequencing based GWAS of 177 congenital obstructive uropathy cases and 24,451 ancestry-matched controls.

Supplementary Table S17. Summary statistics of the sequencing based GWAS of 174 VUR cases and 22,562 ancestry-matched controls.

Supplementary Table S18. Meta-analysis of 727 cases and 434,900 controls with congenital obstructive uropathy.

Supplementary Table S19. Summary statistics of the sequencing based GWAS of 97 bladder exstrophy cases and 22,037 ancestry-matched controls.

Supplementary Table S20. Lead variants from bladder exstrophy seqGWAS.

Supplementary Table S21. Meta-analysis of 725 cases and 29,389 controls with bladder exstrophy.

Supplementary Table S22. Heritability estimation in 623 European CAKUT cases and 20,060 controls using GREML-LDMS.

Supplementary Table S23. Phenotype-specific genomic risk score analysis.

Supplementary References

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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. Sadeghi-Alavijeh O, Chan MM, Doctor GT, et al. Quantifying variant contributions in cystic kidney disease using national-scale whole genome sequencing. *J Clin Invest*. Published online August 27, 2024. doi:10.1172/JCI181467

31. 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.

32. 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.

33. 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.

34. 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.

35. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894.

36. 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.

37. Zhou W, Zhao Z, Nielsen JB, et al. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nat Genet*. 2020;52(6):634-639.

38. 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.

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

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

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

41. 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.

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

43. Falconer DS. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Ann Hum Genet*. 1965;29(1):51-76.

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

45. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 2019;8(7). doi:10.1093/gigascience/giz082

46. van Rooij IALM, van der Zanden LFM, Bongers EMHF, et al. AGORA, a data- and biobank for birth defects and childhood cancer. *Birth Defects Res A Clin Mol Teratol*. 2016;106(8):675-684.

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

48. 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.

49. 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.

50. Wang Q, Dhindsa RS, Carss K, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature*. 2021;597(7877):527-532.

51. 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.

52. 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.

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.

medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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. 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.

59. 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.

60. 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.

61. 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.

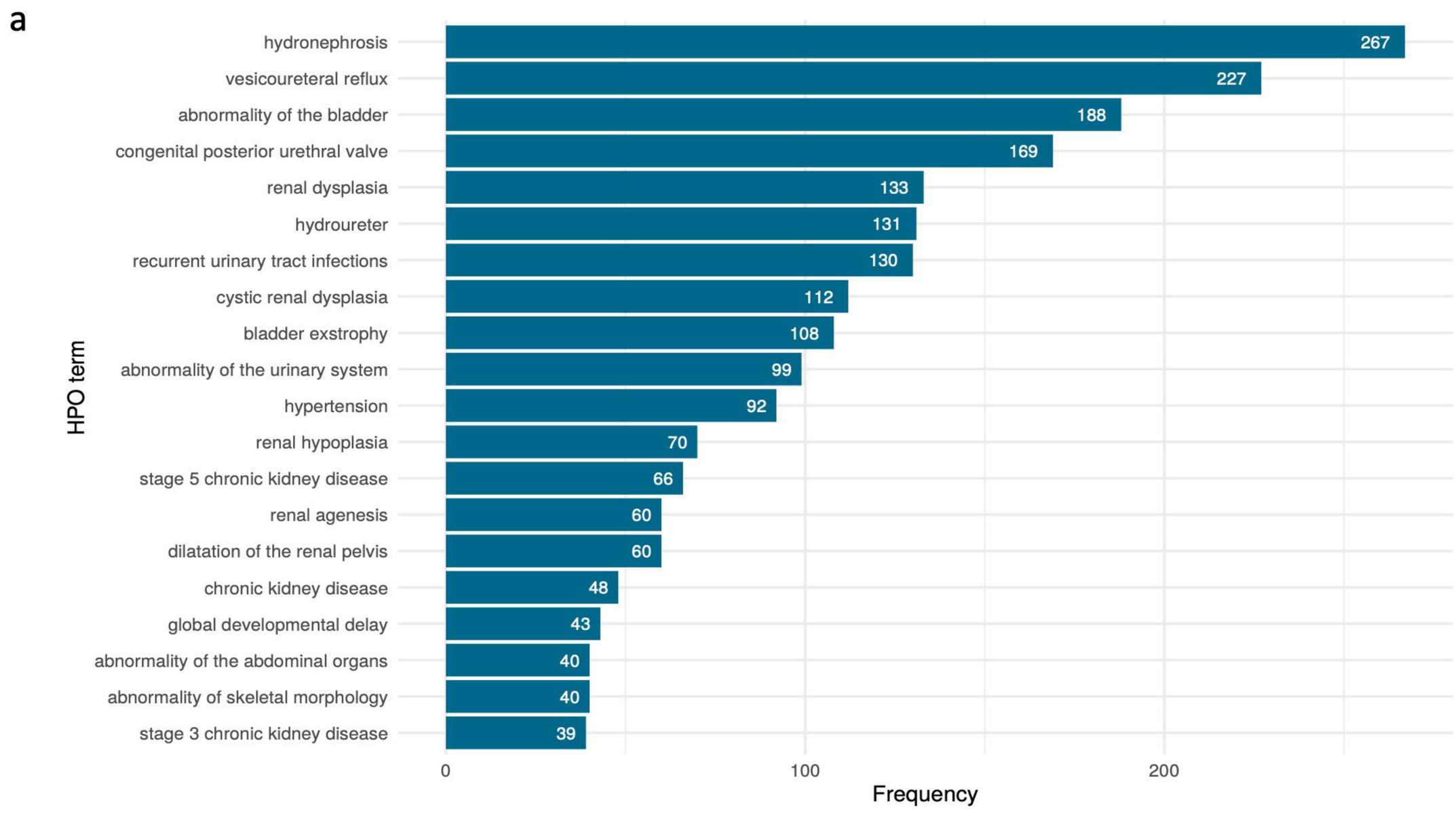
medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

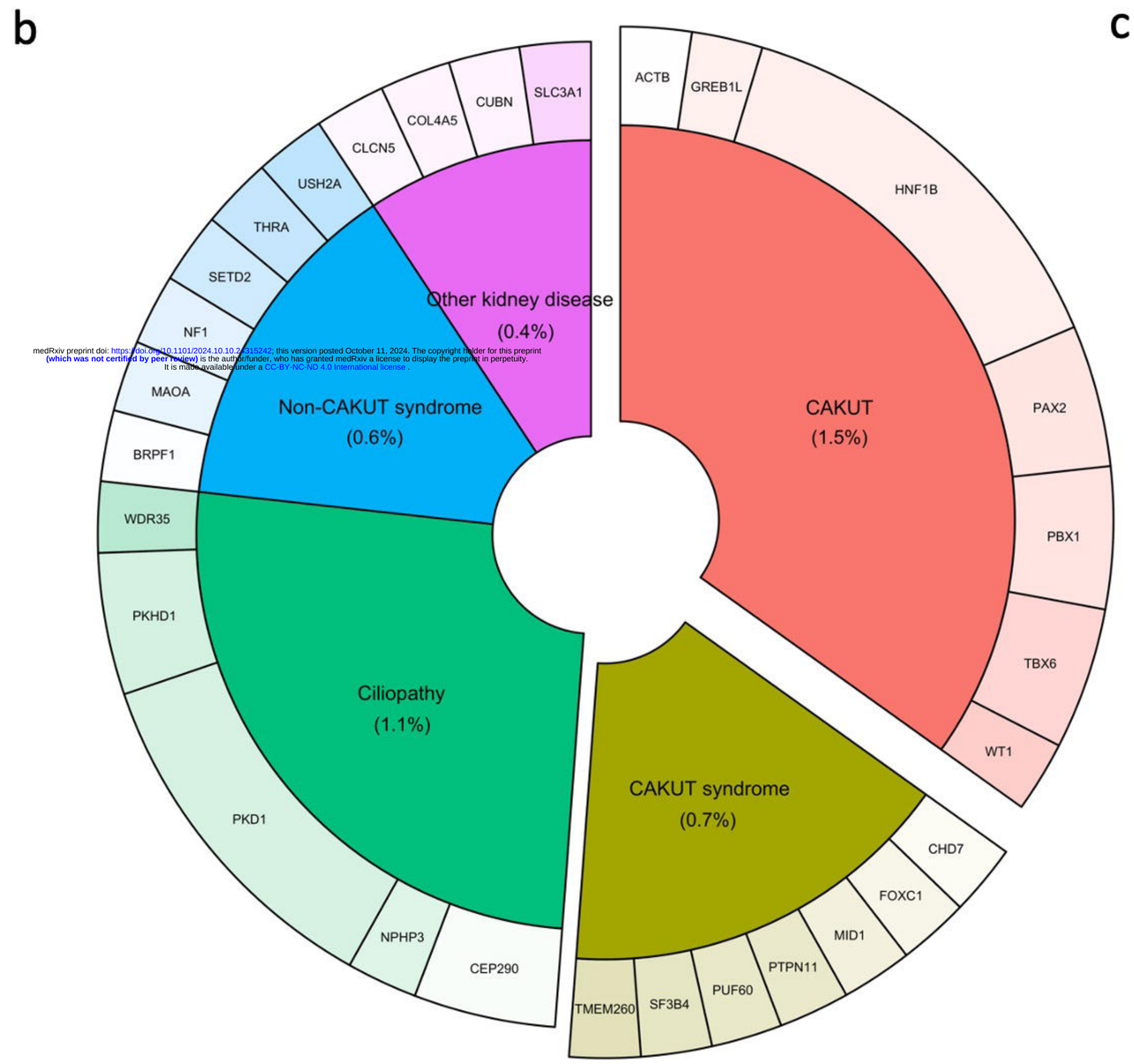
Figure Legends

Figure 1. Genomic testing in 992 patients with CAKUT. a) Most frequently reported Human Phenotype Ontology (HPO) terms in the cohort. Patients could have ≥ 1 associated HPO term. b) Disease-causing genes identified in 4.3% of the cohort. c) Diagnostic yield by phenotype. CAKUT, congenital anomalies of the kidneys and urinary tract; VUR, vesico-ureteral reflux; PUV, posterior urethral valves.

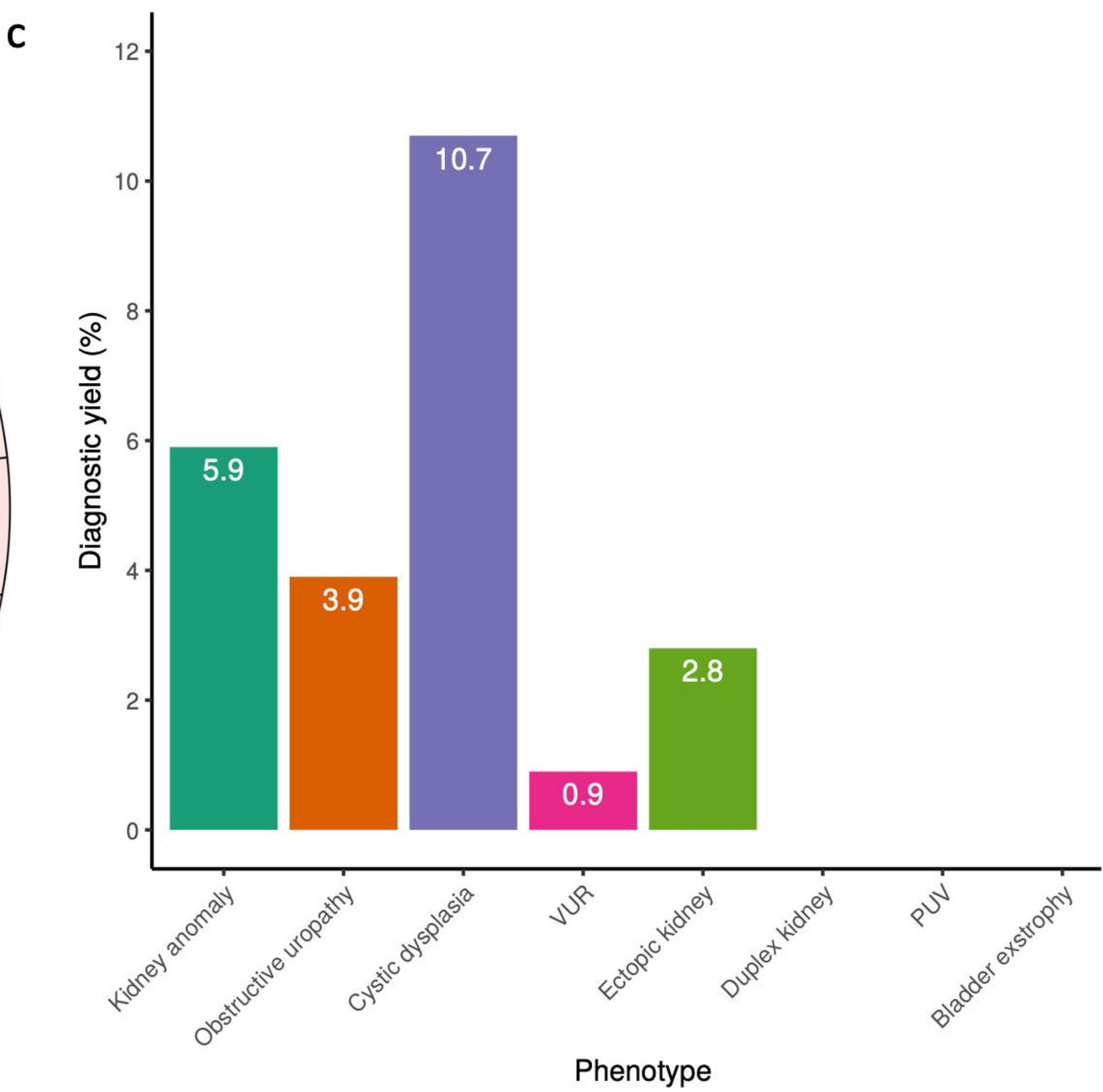
Figure 2. Manhattan plot of exome-wide gene-based rare, likely deleterious variant association testing for a) 813 CAKUT cases and 25,205 ancestry-matched controls and b) 177 cases with congenital obstructive uropathy and 24,451 ancestry-matched controls. Each dot represents a gene. The red line indicates the exome-wide significance threshold of $P=2.58\times10^{-6}$. Genes with $P < 10^{-4}$ are labelled and known disease-gene associations are highlighted in red.

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) Heatmap showing the phenotypic variance explained (R^2, adjusted for disease prevalence on a scale of 0-1) by cohort-specific generated GRS in the other cohorts. Highest overlap is seen for VUR and obstructive uropathy. The bladder exstrophy GRS and PUV-GRS are distinct from each other and the other phenotypes. c) Density plot with the distribution of the PUV-GRS consisting of 36,106 variants in an independent European cohort of 77 cases (AGORA_PUV) and 2,746 controls. $*P=1\text{x}10^{-4}$. PUV, posterior urethral valves; VUR, vesico-ureteral reflux; PRS, polygenic risk score.

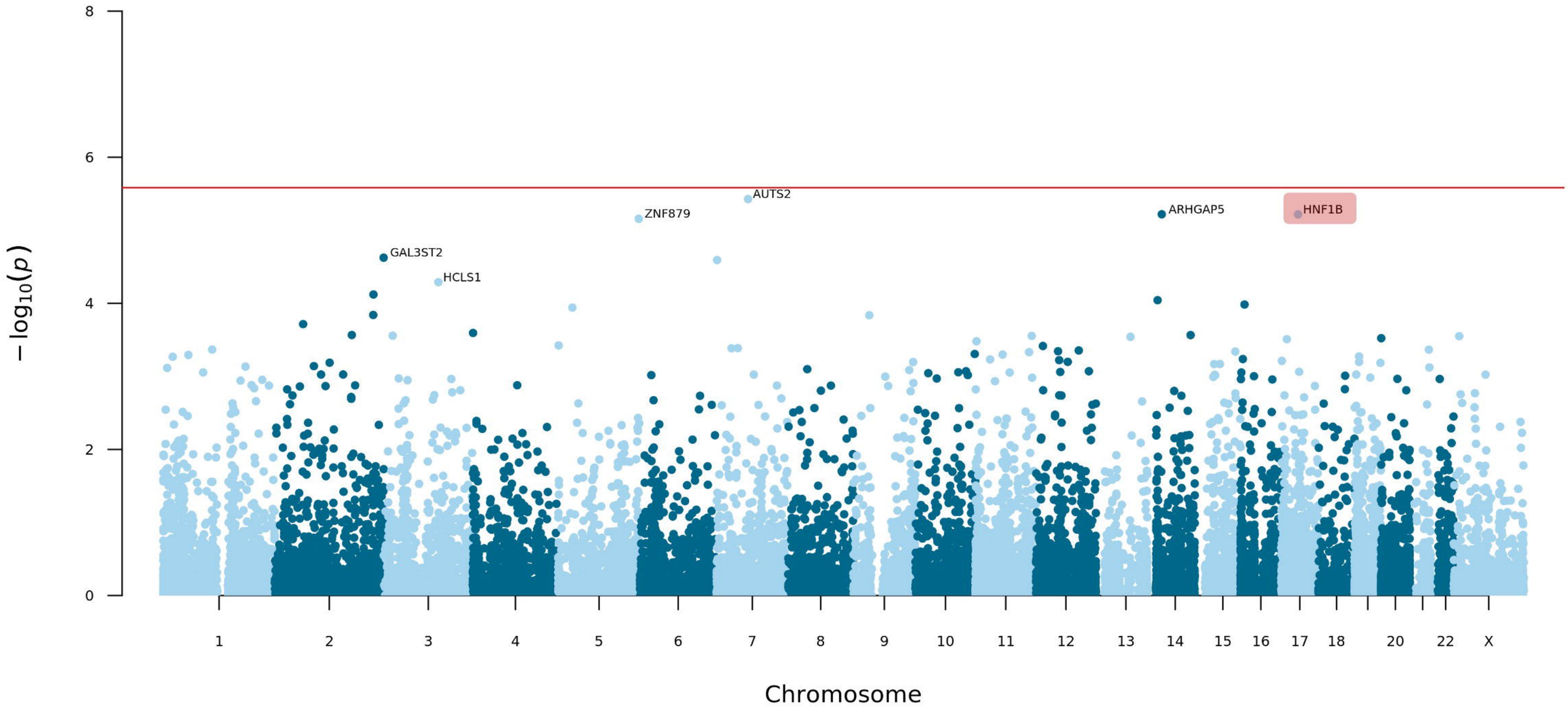




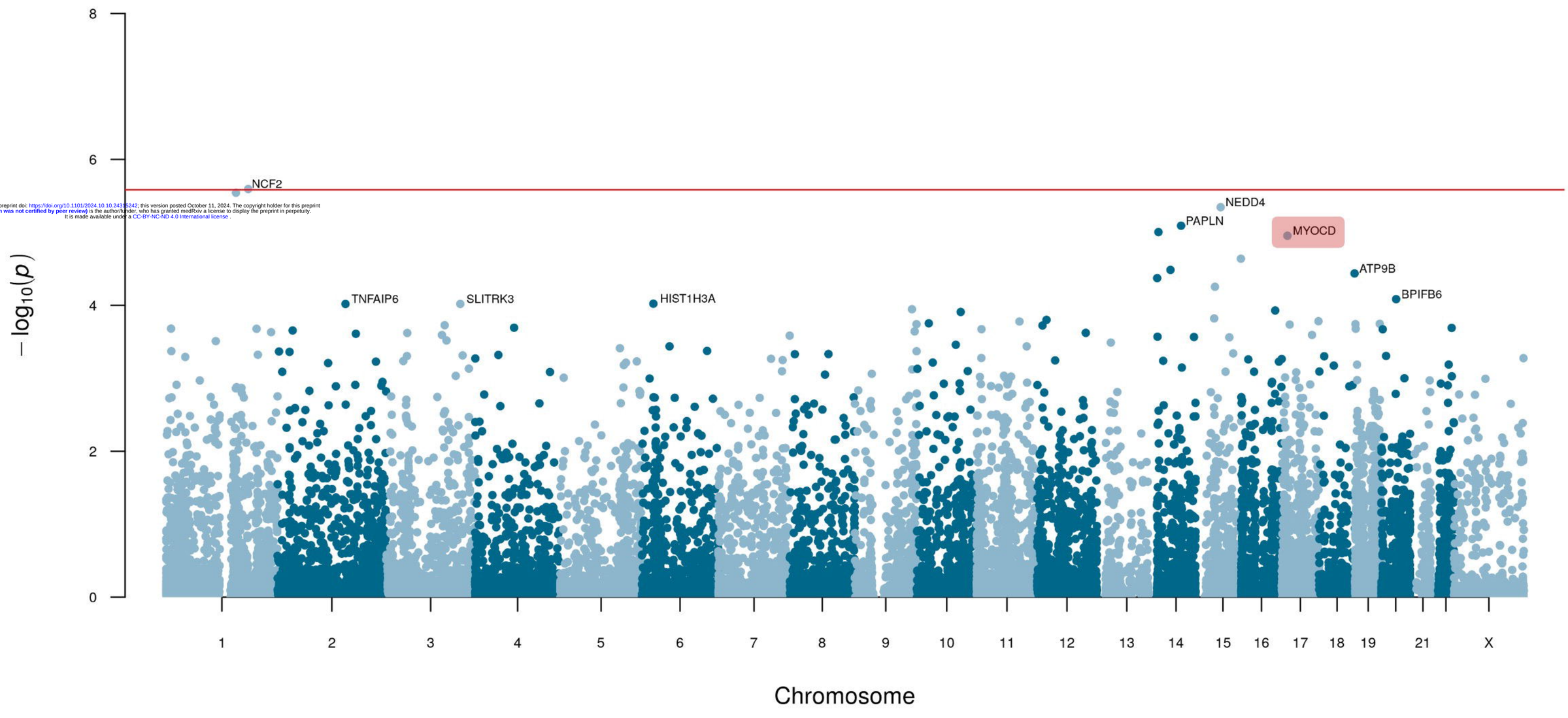
medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



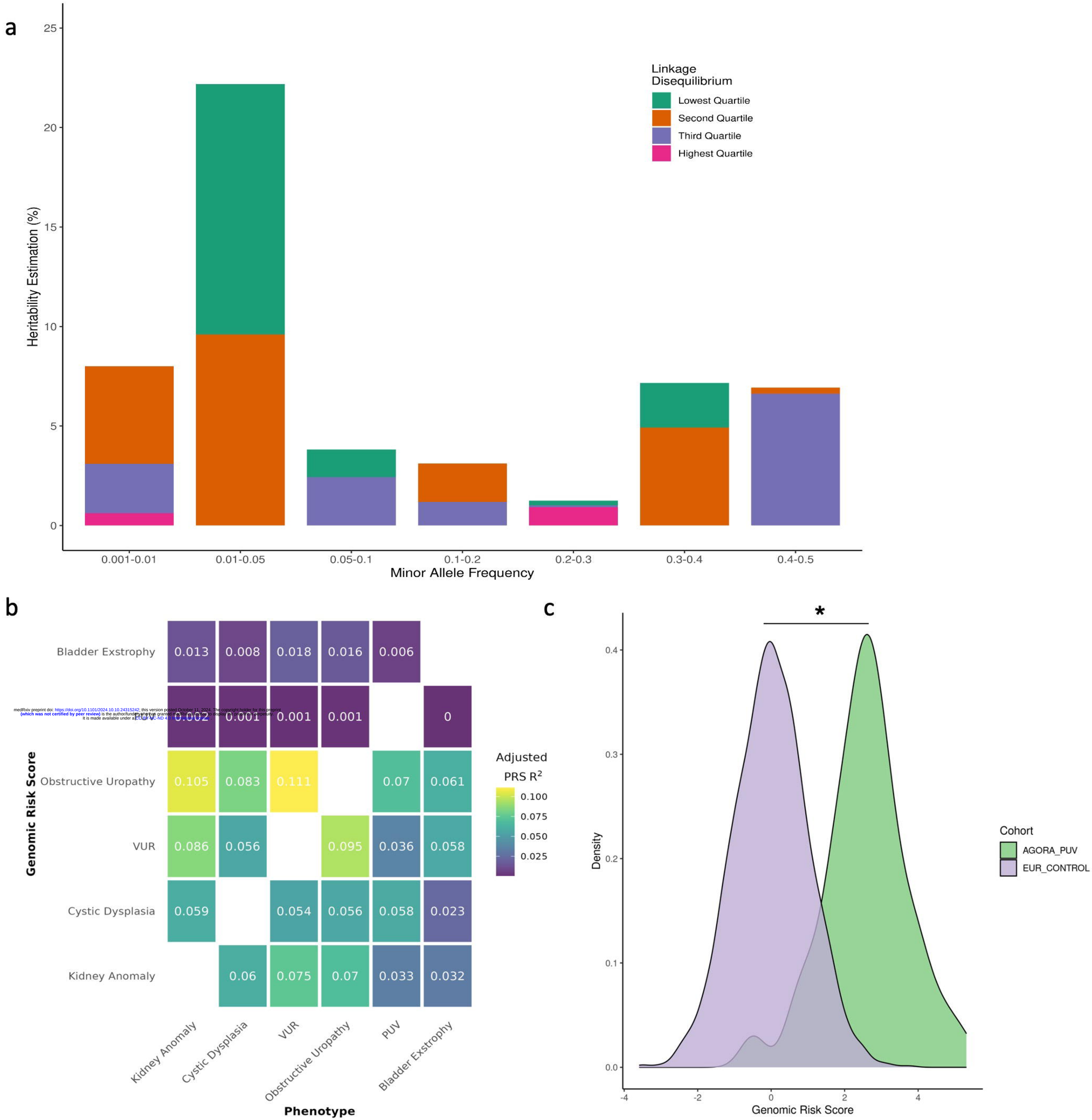
a



b



medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2024.10.10.24315242; this version posted October 11, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.