The Genetic Architecture of Structural Renal and Urinary Tract Malformations

Melanie M.Y. Chan

Division of Medicine

University College London

A thesis submitted for the degree of

Doctor of Philosophy

May 2022

Declaration

'I, Melanie Mai Yee Chan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.'

Abstract

Structural renal and urinary tract malformations are the most common cause of kidney failure in children. These congenital anomalies of the kidneys and urinary tract (CAKUT) are a phenotypically diverse group of malformations that result from defects in embryonic kidney, ureter, and bladder development. A genetic basis for CAKUT has been proposed, with over 50 monogenic causes reported, however, a molecular diagnosis is detected in less than 20% of patients.

In this thesis I used bioinformatics and statistical genetics methodology to investigate the genetic architecture of structural renal and urinary tract malformations using whole-genome sequencing (WGS) data from the 100,000 Genomes Project. Population-based rare and common variant association testing was performed in over 800 cases and 20,000 controls of diverse ancestry seeking enrichment of single-nucleotide/indel and structural variation on a genome-wide, per-gene, and *cis*-regulatory element basis.

Using a sequencing-based genome-wide association study (GWAS) I identified the first robust genetic associations of posterior urethral valves (PUV), the most common cause of kidney failure in boys. Bayesian fine-mapping and functional annotation mapped these two loci to the transcription factor *TBX5* and planar cell polarity gene *PTK7,* with both signals replicated in an independent cohort. Significant enrichment of rare structural variation affecting *cis*-regulatory elements was also detected providing novel insights into the pathogenesis of this poorly understood disorder.

I also demonstrated that the contribution of known monogenic disease to CAKUT has been overestimated and that common and low-frequency variation plays an important role in phenotypic variability. These findings support an omnigenic rather

than monogenic model of inheritance for CAKUT and are consistent with the extensive genotypic-phenotypic heterogeneity, variable expressivity, and incomplete penetrance observed in this condition. Finally, this work demonstrates the value of sequencing-based GWAS methodology in rare disease, beyond conventional monogenic gene discovery, and provides strong support for an inclusive diverse-ancestry approach.

Impact Statement

The findings from this study will have broad impact across multiple disciplines, both inside and outside of nephrology. First, from a patient perspective, this work offers insights into the potential cause and inheritance of PUV, improving understanding of the condition and offering hope for future translational research. Second, these results will be of great interest to other researchers within the field of nephrogenetics, including adult and paediatric nephrologists and clinical geneticists, and will stimulate collaborations to further investigate the genomic complexity of CAKUT. Third, the association of *TBX5* and *PTK7* with PUV expands the known role of these genes in embryonic development and clearly implicates the planar cell polarity pathway in urethral development. These findings will hopefully catalyze future mechanistic and therapeutic studies by developmental biologists.

Finally, the successful use of a rigorously controlled diverse ancestry WGS association study to dissect the genetic contribution of a rare disease is innovative and has wider implications for the field of genomics, particularly as the dividing line between common and rare disease is becoming less well defined. In this thesis I have clearly demonstrated the scientific benefits of an inclusive approach which provided novel insights that would have been missed in a purely European cohort. Such an approach has not previously been applied to rare disease and the methodology used in this study will form an exemplar for subsequent analyses of WGS datasets across any number of rare diseases. Furthermore, this study normalizes the representation of individuals from diverse ancestral backgrounds in genetic association studies which will hopefully contribute to raising awareness of the current inequalities in genomics and promote the recruitment of more diverse cohorts going forward.

The impact of this work will be disseminated primarily through publication in peer-reviewed journals which address a broad audience, with an accompanying online lay summary to increase patient and public engagement. Promotion of the results on social media platforms such as Twitter will aim to increase visibility to other academics and ongoing engagement with the funder Kidney Research UK will ensure that the results of this study are fed back to both patients and charity donors. This work has already been selected for oral presentation at two international conferences (Wellcome Genomics of Rare Disease 2021 and the American Society of Nephrology Kidney Week 2021), two national conferences (Association of Physicians of Great Britain and Ireland Annual Meeting 2021 and UK Kidney Week 2022) and was presented as part of the Genomics England Research Seminar Series in September 2021. This highlights the broad appeal of this work spanning the different fields of genetics, medicine, and nephrology.

Acknowledgements

Finally, to my husband Greg and daughters Mae and Zoe, thank you for putting up with me just 'running one more script'! Without your love, support, and encouragement I would not have been able to complete this thesis.

Publications

During this PhD, I have authored or contributed to the following publications:

Diverse ancestry whole-genome sequencing association study identifies *TBX5* and *PTK7* as susceptibility genes for posterior urethral valves. **Melanie MY Chan**, Omid Sadeghi-Alavijeh, Filipa M Lopes, Alina C Hilger, Horia C Stanescu, Catalin D Voinescu, Glenda M Beaman, William G Newman, Marcin Zaniew, Stefanie Weber, Yee Mang Ho, John O Connolly, Dan Wood, Alexander Stuckey, Athanasios Kousathanas, Genomics England Research Consortium, Robert Kleta, Adrian S Woolf, Detlef Bockenhauer, Adam P Levine and Daniel P Gale. *eLife*. 2022 Sep 20;11:e74777. doi.org/10.7554/eLife.74777. PMID: 36124557.

Genome-Wide Association Study and Meta-Analysis of Intrahepatic Cholestasis of Pregnancy Identifies Multiple Susceptibility Loci in Liver-Specific Genes and Regulatory Elements. Peter H. Dixon, Adam P. Levine, Inês Cebola, **Melanie M. Y. Chan**, Aliya S. Amin, Anshul Aich, Monika Mozere, Hannah Maude, Alice L. Mitchell, Jun Zhang, NIHR BioResource, Genomics England Research Consortium, Jenny Chambers, Argyro Syngelaki, Jennifer Donnelly, Sharon Cooley, Michael Geary, Kypros Nicolaides, Malin Thorsell, William M. Hague, Maria Cecilia Estiu, Hanns-Ulrich Marschall, Daniel P. Gale, Catherine Williamson. *Nature Communications*. 2022 Aug;13(1):4840. doi 10.1038/s41467-022-29931-z. PMID: 35977952.

Genotype-phenotype correlations for *COL4A3-COL4A5* variants resulting in Gly substitutions in Alport syndrome. Joel T Gibson, Mary Huang, Marina Shenelli Croos Dabrera, Krushnam Shukla, Hansjorg Rothe, Pascale Hilbert, Constantinos Deltas, Helen Storey, Beata S Lipska-Ziętkiewicz, **Melanie MY Chan**, Omid Sadeghi-Alavijeh, Daniel P Gale, Genomics England Research Consortium, Agne Cerkauskaite, Judy Savige. *Sci Rep*. 2022 Feb;12(1):2722. doi: 10.1038/s41598-022-06525-9. PMID: 35177655.

Gitelman-Like Syndrome Caused by Pathogenic Variants in mtDNA. Viering D, Schlingmann KP, Hureaux M, Nijenhuis T, Mallett A, **Chan MMY**, van Beek A, van Eerde AM, Coulibaly JM, Vallet M, Decramer S, Pelletier S, Klaus G, Kömhoff M, Beetz R, Patel C, Shenoy M, Steenbergen EJ, Anderson G, Bongers EMHF, Bergmann C, Panneman D, Rodenburg RJ, Kleta R, Houillier P, Konrad M, Vargas-Poussou R, Knoers NVAM, Bockenhauer D, de Baaij JHF; Genomics England

Research Consortium. *J Am Soc Nephrol*. 2022 Feb;33(2):305-325. doi: 10.1681/ASN.2021050596. PMID: 34607911.

Prevalence Estimates of Predicted Pathogenic *COL4A3 - COL4A5* Variants in a Population Sequencing Database and Their Implications for Alport Syndrome. Joel Gibson, Rachel Fieldhouse, **Melanie Chan**, Omid Sadeghi-Alavijeh, Leslie Burnett, Valerio Izzi, Anton Persikov, Daniel Gale, Helen Storey, and Judy Savige. *J Am Soc Nephrol*. 2021 Sep;32(9):2273-2290. doi: 10.1681/ASN.2020071065. PMID: 34400539.

Genetics of renovascular hypertension in children. Viering DHHM, **Chan MMY**, Hoogenboom L, Iancu D, de Baaij JHF, Tullus K, Kleta R, Bockenhauer D. *J Hypertens*. 2020 Oct;38(10):1964-1970. doi: 10.1097/HJH.0000000000002491. PMID: 32890272.

Large-Scale Whole-Genome Sequencing Reveals the Genetic Architecture of Primary Membranoproliferative GN and C3 Glomerulopathy. Levine AP, **Chan MMY**, Sadeghi-Alavijeh O, Wong EKS, Cook HT, Ashford S, Carss K, Christian MT, Hall M, Harris CL, McAlinden P, Marchbank KJ, Marks SD, Maxwell H, Megy K, Penkett CJ, Mozere M, Stirrups KE, Tuna S, Wessels J, Whitehorn D; MPGN/DDD/C3 Glomerulopathy Rare Disease Group; NIHR BioResource, Johnson SA, Gale DP. *J Am Soc Nephrol*. 2020 Feb;31(2):365-373. doi: 10.1681/ASN.2019040433. PMID: 31919107.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

100KGP	100,000 Genomes Project
AD	Autosomal Dominant
ADTKD	Autosomal Dominant Tubulo-interstitial Kidney Disease
AF	Allele Frequency
ALT	Alternate allele
AR	Autosomal Recessive
ATAC-seq	Assay for Transposable-Accessible Chromatin using sequencing
BOR	Branchio-Oto-Renal syndrome
bp	Base-pair
CADD	Combined Annotation Dependent Depletion
CAKUT	Congenital Anomalies of the Kidneys and Urinary Tract
cCRE	Candidate *cis*-Regulatory Element
CGH	Comparative Genomic Hybridization
ChIP-seq	Chromatin ImmunoPrecipitation combined with sequencing
CHR	Chromosome
CI	Confidence Interval
CKD	Chronic Kidney Disease
CNV	Copy Number Variant
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRT	Cyclic Reversible Termination
CTCF	CCCTC-binding factor
DEL	Deletion
dELS	Distal Enhancer-Like Signature
DNA	DeoxyriboNucleic Acid

dNTP	DeoxyNucleotide TriPhosphate
DUP	Duplication
eGFR	Estimated Glomerular Filtration Rate
eQTL	Expression Quantitative Trait Loci
ES	Exome Sequencing
ESKD	End-Stage Kidney Disease
EWAS	Epigenome-Wide Association Study
FDR	False Discovery Rate
FISH	Fluorescence in Situ Hybridization
GLMM	Generalized Logistic Mixed Model
gnomAD	Genome Aggregation Database
GQ	Genotype Quality
GRM	Genomic Relationship Matrix
GT	Genotype
gVCF	Genomic Variant Call Format
GWAS	Genome-Wide Association Study
HDR	Hypoparathyroidism, Deafness, and Renal disease
HEMI	Hemizygous
HET	Heterozygous
Hi-C	High-throughput Chromosome conformation capture
HLA	Human Leucocyte Antigen
HOM	Homozygous
HPO	Human Phenotype Ontology
HR	Hazard Ratio
HWE	Hardy-Weinberg Equilibrium
IBD	Identical by Descent
Indel	Insertion/deletion
INV	Inversion

iPSCs	induced Pluripotent Stem Cells
IQR	Interquartile Range
kb	Kilobase
LCR	Low-Copy Repeat
LMM	Linear Mixed Model
LD	Linkage Disequilibrium
LoF	Loss-of-Function
LUTO	Lower Urinary Tract Obstruction
MAC	Minor Allele Count
MAF	Minor Allele Frequency
Mb	Megabase (1 million bases)
MIM	Mendelian Inheritance in Man
MODY	Mature Onset Diabetes of the Young
NCBI	National Center for Biotechnology Information
NGS	Next-Generation Sequencing
NHS	National Health Service
OR	Odds Ratio
PCA	Principal Components Analysis
PCR	Polymerase Chain Reaction
pELS	Proximal Enhancer-Like Signature
PLS	Promoter-Like Signature
POS	(Chromosomal) Position
PP	Posterior Probability
PUJO	Pelvi-Ureteral Junction Obstruction
PUV	Posterior Urethral Valves
QC	Quality Control
Q-Q	Quantile-Quantile
RCAD	Renal Cysts and Diabetes syndrome

REF	Reference allele
RNA	RiboNucleic Acid
RNAi	RiboNucleic Acid interference
RNA-seq	RiboNucleic Acid sequencing
RR	Relative Risk
SE	Standard Error
SKAT-O	Sequence Kernel Association Test – Optimal
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SPA	Saddlepoint Approximation
SV	Structural Variant
TAD	Topologically Associated Domain
TSS	Transcription Start Site
VACTERL	Vertebral defects, Anal atresia, Cardiac defects, Tracheo-Esophageal fistula, Renal and Limb anomalies
VCF	Variant Call Format
VEP	Variant Effect Predictor
VUJO	Vesico-Ureteral Junction Obstruction
VUR	Vesico-Ureteral Reflux
VUS	Variant of Uncertain Significance
WES	Whole-Exome Sequencing
WGS	Whole-Genome Sequencing
XLD	X-linked Dominant
XLR	X-linked Recessive

Chapter 1: Introduction

1.1 The Genomic Era

The central goal of human genetics is to understand the relationship between genotype and phenotype. Delineating the genetic causes and molecular pathways that influence human disease susceptibility and progression is crucial for the development of targeted therapeutics and the successful implementation of genomic medicine. Twenty years after the ground-breaking publication of the draft human genome sequence (Consortium and International Human Genome Sequencing Consortium 2001, Venter *et al.* 2001), we have seen a revolution in genomic technologies, generating vast datasets of human genetic variation pioneered by the International HapMap Project (The International HapMap 3 Consortium 2010) and the 1000 Genomes Project (1000 Genomes Project Consortium *et al.* 2015). These large international consortia were instrumental in describing common and rare patterns of genetic variation, but have quickly been superseded by biobanks containing large-scale sequencing data from tens to hundreds of thousands of both healthy and diseased individuals (H3Africa Consortium *et al.* 2014, Gudbjartsson *et al.* 2015, Nagai *et al.* 2017, Bycroft *et al.* 2018, GenomeAsia100K Consortium 2019, Taliun *et al.* 2019, Turro *et al.* 2020, 100,000 Genomes Project Pilot Investigators *et al.* 2021).

Initial excitement that publication of the draft human genome would lead to a rapid transformation in clinical care has been tempered somewhat as the true complexity of the human genome became evident. Indeed, it was only very recently that the entire 3.1 billion base-pair (bp) sequence of the human genome was completed (Nurk *et al.* 2022), providing telomere-to-telomere granularity for the first time.

However, the vast array of genomic data that has been generated over the last two decades, combined with innovations in statistical genetics and computational analysis, has advanced our understanding of the underlying biology of many diseases and led to the development of new therapeutic interventions, such as ribonucleic acid interference (RNAi) molecules for primary hyperoxaluria (Garrelfs *et al.* 2021) and PCSK9 inhibitors for primary hypercholesterolemia (Abifadel *et al.* 2003). Substantial benefits are also beginning to be seen in the clinic for patients with cancer and rare disease in the form of prompt and accurate diagnosis, risk stratification by genotype, and personalized treatments and screening programmes. The recent introduction of whole genome sequencing (WGS) into routine clinical care within the NHS means this is truly an exciting time for genomic medicine.

1.1.1 Next-generation sequencing

Walter Gilbert and Frederick Sanger pioneered the sequencing of DNA in 1977, with Sanger's 'chain-termination' method becoming the dominant approach for the next 30 years (Maxam and Gilbert 1977, Sanger *et al.* 1977). With this came the ability to read our genomes for the first time, paving the way for The Human Genome Project (International Human Genome Sequencing Consortium 2004). This collaborative effort used a 'shotgun' approach breaking the genome into ~150 kilobase (kb) segments, then cloning them into bacterial artificial chromosomes for replication before sequencing, taking a total of 13 years at a cost of $2.7 billion. A revolution in DNA sequencing technologies followed, with the development of high-throughput or 'next-generation' sequencing (NGS) in the mid-2000s. Rapid evolution of this massively parallel approach has increased the accuracy, read length, speed and affordability of these technologies meaning an entire genome can now be sequenced in under an hour, for less than £500.

NGS can be broadly divided into short-read (35-700bp) and long-read (20-200Kb) approaches (Goodwin *et al.* 2016). The most widely available and utilized high

throughput platform is provided by Illumina who use the short-read sequencing by synthesis cyclic reversible termination (CRT) method with bridge amplification which has > 99.9% read accuracy. Figure 1.1 describes the principles underlying this technology.
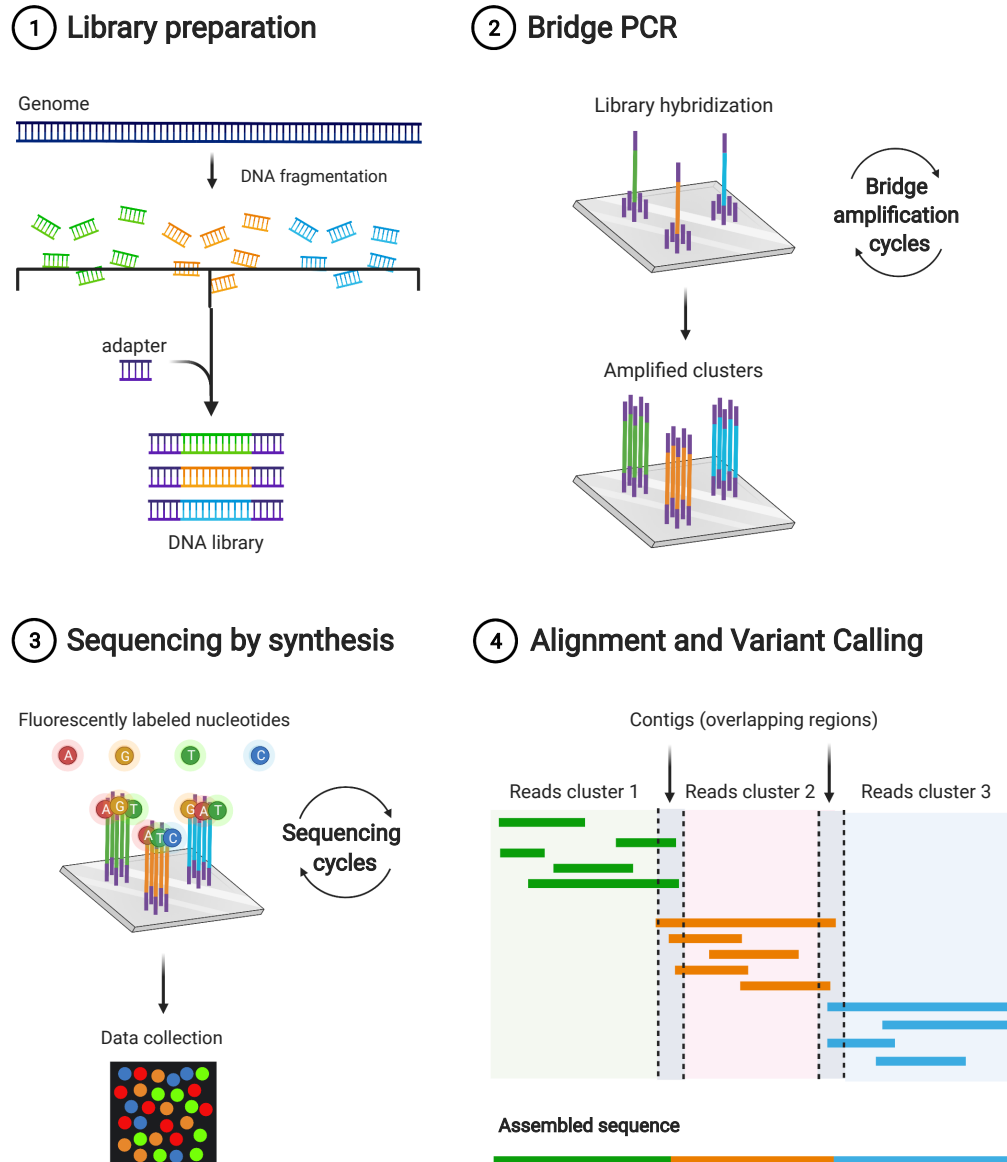
Figure 1.1. Illumina next-generation sequencing-by-synthesis.

1) DNA is randomly fragmented into 200-300bp, and adapters are ligated to the 3' and 5' ends. 2) The sequencing library is loaded into a flow cell where the fragments hybridize onto surface-bound oligonucleotides complementary to the adapters. The free ends interact with nearby primers, forming a bridge. PCR amplification is used to create clonal clusters of the DNA templates before the reverse strands are cleaved and washed away. 3) As each of the four fluorescently labelled reversible terminator-bound deoxynucleotides (dNTPs) are incorporated into DNA template strands, the surface is imaged allowing identification of which dNTP has been incorporated into each cluster. The fluorophore and terminator are then removed and the cycle repeats. 4) The DNA template is sequenced from both directions (paired-end sequencing) to generate a read. The reads are mapped and aligned to the reference genome and assessed for quality. Any differences between the reads and reference genome are called variants and recorded in a standardized variant call file (VCF) or genome variant call file (gVCF).

1.1.3 Long-read sequencing

The last decade has seen the development of 'third-generation', or long-read technologies such as nanopore (Oxford Nanopore Technologies) or single-molecule real-time sequencing (Pacific Biosciences), enabling single DNA molecules to be sequenced without amplification and generating longer reads than conventional short-read NGS. The ultralong read lengths (on average 10-30kb) and improved sequencing of GC-rich or low complexity regions provides more uniform coverage across the genome. This enables better resolution of large, complex structural variants (Ritz *et al.* 2014) and enhances the ability to phase genomes, identifying which chromosome (paternal or maternal) an allele is derived from (Kuleshov *et al.* 2014, Snyder *et al.* 2015). At present these tools are more error-prone than short-read sequencing (with error rates often greater than 10%) but can be used to complement short-read WGS, providing information on medium-to-large SVs and phase that are poorly captured by better base-pair resolution short-read approaches.

1.2   Gene Discovery

One of the main aims of gene discovery is to illuminate the molecular mechanisms of disease, improving our understanding of pathogenesis and prioritizing pathways for possible therapeutic intervention. Gene discovery approaches have proved most successful in Mendelian disorders, in which one or two variants in an individual have a very large effect on the likelihood of a phenotype being detectable. The molecular basis for over 6,000 single-gene disorders have so far been identified (OMIM, February 2022). Traditional gene discovery efforts have typically been family-based, focused on large pedigrees with multiply affected family members using genome-wide linkage analysis. This unbiased statistical genetic mapping approach looks at the transmission of markers through a family to provide an estimate of the distance of each marker from the trait-causing variant(s). Whilst a useful approach for Mendelian disease with large phenotypic effects, it cannot be applied to *de novo* disorders and has reduced power when penetrance is incomplete or there is locus heterogeneity (more than one gene results in the same phenotype). Population-based candidate gene studies (using positional cloning methods and more recently targeted NGS approaches) which select genes based on their biological plausibility in a hypothesis-driven manner have also been used for gene discovery in rare disease but are limited by high false-positive rates and are often difficult to replicate. Both these methods have now largely been superseded by whole-exome sequencing approaches in rare disease, while genome-wide association studies (GWAS) are conventionally used for gene discovery in common, complex traits.

1.2.1 GWAS

The primary aim of a GWAS is to detect statistically significant association between genetic variants and a disease/trait. This unbiased scanning of known (and usually common) variants across the genome enables the identification of previously

unsuspected genes and pathways that might influence disease susceptibility, where the magnitude of effect is less important than the insights they provide into the underlying biology. Because power to detect association in a case-control study is a function of both the strength of effect of an allele (i.e., its odds ratio) and the frequency of that allele in the study population, GWAS has traditionally been used to investigate complex traits and disorders such as diabetes or schizophrenia. The rationale for this is the common disease-common variant hypothesis (Reich and Lander 2001) which posits that common variants (generally those with minor allele frequency [MAF] > 5%) with small-to-modest effects underlie common, complex traits.

The very first landmark GWAS published in 2005 compared just 96 individuals with age-related macular degeneration with 50 controls, identifying significant association with common variation in *Complement Factor H* (*CFH*) (Klein 2005). In the 15 years since, more than 5,700 GWAS have been performed for over 3,300 traits (Uffelmann *et al.* 2021) and its use has expanded beyond simple gene discovery to the generation of clinical risk prediction scores (Khera *et al.* 2018), estimation of heritability (Yang *et al.* 2010) and prioritization of pathways for drug development (Wang *et al.* 2009).

The majority of GWAS use genome-wide single-nucleotide polymorphism (SNP) microarrays covering hundreds of thousands of variants (usually with MAF > 1%) to genotype a cohort of interest for comparison with a suitable control population. Imputation using an ancestry-specific reference panel of haplotypes reconstructed from sequencing data (e.g., the Haplotype Reference Consortium) can then be used to fill in the gaps, capitalizing on patterns of linkage disequilibrium (LD; where alleles are inherited together in a non-random manner) to infer missing variants. These panels, however, are less accurate at imputing variants not in LD with the genotyped variants such as those that are rare (seen in < 1% of the general population), or those seen in non-European populations. These genotyped and imputed variants

are then tested for association with a trait but in most cases are simply markers or indirect proxies of the causal variant in the region.

Because of these limitations, high-coverage WGS data has started to be used for GWAS investigating common diseases or traits (Höglund *et al.* 2019, Zhao *et al.* 2020, Hu *et al.* 2021), demonstrating improved power and sensitivity over conventional techniques, most likely due to the superior variant calling capabilities of WGS. The majority of identified novel associations in these studies were variants that were either rare or ancestry-specific, highlighting the benefits of WGS over traditional genotyping and imputation (Hu *et al.* 2021). For now, however, the widespread adoption of WGS for GWAS remains limited due to cost and sample size constraints.

In response to the exponential increase in published GWAS, stringent guidelines have been proposed and are widely implemented (The Wellcome Trust Case Control Consortium 2007) including strict variant and sample-level quality control filtering to ensure data reliability; adjustment for population structure to prevent spurious genetic associations due to ancestry-specific allele frequency differences; correction for multiple testing and a requirement for replication of any identified loci in an independent cohort. Meta-analysis, where multiple cohorts are combined to produce ever-increasing sample sizes, has also proved a useful tool to boost power and detect association of variants with progressively smaller effect sizes. Recent meta-analyses of estimated glomerular filtration rate (eGFR) and blood pressure for example have involved over one million participants (Evangelou *et al.* 2018, Wuttke *et al.* 2019).

GWAS has now identified tens of thousands of disease/trait associations. These variants are enriched in regulatory regions (Maurano *et al.* 2012) and demonstrate extensive pleiotropy (influence multiple phenotypic traits) (Watanabe *et al.* 2019). Initial successes, such as highlighting the role of autophagy in Crohn's disease

(Hampe *et al.* 2007, Rioux *et al.* 2007), led to much excitement, however, the early promise of GWAS has yet to translate into clinically significant advances in patient care for two main reasons. First, while thousands of variants with small effects have been identified, overall, these make a relatively small contribution to disease risk. This so-called 'missing heritability' (Manolio *et al.* 2009) has been attributed to a) common variants with low effect sizes not captured in current GWAS, b) the contribution of rare or structural variants with larger effects, c) gene-gene interactions (epistasis), and d) parent of origin effects (e.g., genomic imprinting). Second, the vast majority (~90%) of risk alleles identified by GWAS are found in non-coding regions of the genome and functional annotation with cell and context-specific multi-omics data generated by projects such as ENCODE (ENCODE Project Consortium *et al.* 2020), the RoadMap Epigenomics Consortium (Roadmap Epigenomics Consortium *et al.* 2015) and GTEx (GTEx Consortium 2020) is necessary to prioritize the causal variant or gene for functional follow-up. Such an approach has recently been used to successfully fine-map a locus associated with kidney function (eGFR), integrating human kidney methylation and expression quantitative trait loci (eQTL) with single nuclei ATAC-seq data to prioritize *DPEP1* and *CHMP1A* as kidney disease risk genes in the proximal tubule (Guan *et al.* 2021). However, in most cases the complexity, dynamic nature and cell-type specificity of the regulatory genome means meaningful interpretation of these non-coding variants remains a challenge.

Finally, > 95% of GWAS have so far been conducted in individuals of European ancestry (https://gwasdiversitymonitor.com) and it is widely recognized that increasing ancestral diversity in genetic studies is scientifically and ethically necessary (Peterson *et al.* 2019, Fatumo *et al.* 2022). GWAS findings have been shown to replicate across populations in a variety of common diseases (Waters *et al.* 2010, Carlson *et al.* 2013, Coram *et al.* 2013, Marigorta and Navarro 2013, Li and Keating 2014, Liu *et al.* 2015, Kuchenbaecker *et al.* 2019, Lam *et al.* 2019),

suggesting sharing of common causal variants between ancestries despite differences in allele frequency and effect size (Ntzani *et al.* 2012). Furthermore, the benefit of combining population groups has been clearly demonstrated in trans-ancestry meta-analyses (Mahajan *et al.* 2014, Morris *et al.* 2019, Graff *et al.* 2021), where differences in LD structure are specifically utilized to improve the resolution of fine-mapping at significant loci.

1.2.2 Whole-exome sequencing

Ng et al. first reported the targeted capture and parallel sequencing of 12 human exomes in 2009 (Ng *et al.* 2009), demonstrating how this new technology could be used as a cost-efficient and robust gene discovery approach in Mendelian disorders (Ng *et al.* 2010). Whole exome sequencing (WES) focuses on the ~1% of the human genome that encodes proteins, based on the assumption that gene-disrupting variation is more likely to be disease-causing, and has been employed successfully in both a research setting to identify new gene-phenotype relationships as well as for diagnostic purposes in clinical practice (Yang *et al.* 2013, Lee, Deignan, *et al.* 2014, Retterer *et al.* 2016).

Evolutionary theory states that natural selection should eliminate highly deleterious variants before they reach a high frequency in the population, resulting in an enrichment of disease-causing alleles with large effects at low allele frequencies. Population-based exome sequencing cohorts have been used to look for association between rare, likely deleterious variants and a particular disease. However, these studies often lack statistical power for two reasons; a) the power to detect single variant association decreases as the minor allele frequency falls and b) sample sizes have been limited by the cost of sequencing. To overcome this, region-based testing is used to 'collapse' information across a genomic region (e.g., a gene) before testing for association with a phenotype. Initial use of rare variant region-based testing approaches with sequencing data were not very informative, limited by low absolute

power and relatively small sample sizes (The UK10K Consortium *et al.* 2015, Fuchsberger *et al.* 2016). However recently, large-scale exome sequencing projects involving hundreds of thousands of individuals, such as the UK Biobank (Bycroft *et al.* 2018), have shown that when sufficiently powered, a gene-based rare variant collapsing approach can be used successfully to identify novel gene-disease relationships (Akbari *et al.* 2021, Deaton *et al.* 2021, Wang *et al.* 2021).

1.2.3 Whole-genome sequencing

The falling cost of WGS has led to its increasing use as both a research and clinical tool (Lupski *et al.* 2010), spearheaded by national sequencing projects such as the UK's NIHR BioResource (Turro *et al.* 2020) and the 100,000 Genomes Project (100,000 Genomes Project Pilot Investigators *et al.* 2021). The advantages and disadvantages of WGS compared with WES are listed in Table 1.1.

Pioneered by consortia such as the 1000 Genomes Project (1000 Genomes Project Consortium *et al.* 2015) and Iceland's deCODE (Gudbjartsson *et al.* 2015), large-scale sequencing cohorts have now become an integral part of human genetics research. Whole-genome sequencing population datasets such as those generated by TOPMed (Taliun *et al.* 2019) and aggregated by gnomAD (Karczewski *et al.* 2020) provide a reference catalogue of human genetic variation across many different population groups and the use of metrics such as allele frequency and gene constraint (the degree to which a gene can tolerate variation) has proven invaluable for clinical variant interpretation. These large datasets have also provided novel insights into human genetic variation demonstrating an abundance of rare or private variants (only seen in one individual) among the 3-4 million single-nucleotide variants (SNVs) and ~0.5 million indels in the average genome, as well as a higher burden of loss-of-function variation in the human population than previously thought (MacArthur *et al.* 2012, Lek *et al.* 2016).

Table 1.1. Advantages and disadvantages of whole-genome sequencing.

Advantages	Disadvantages
Genome-wide detection of coding and non-coding variation allowing interrogation of regulatory and deep intronic splice variants.	The cost, although falling, is still greater than WES.
Less GC-content bias and more uniform coverage of coding regions improves the accuracy of variant calling (Lelieveld *et al.* 2015). For example, 2.6-10% of disease-causing SNVs detected by WGS are missed by WES (Turro *et al.* 2020).	Less accurate detection of mosaicism (where an individual has at least two populations of cells with distinct genotypes derived from a single fertilized egg) which requires several 100-fold sequencing depths.
Improved resolution and detection of structural variants including balanced rearrangements (e.g., inversions, balanced translocations).	The wealth of data generated (approximately 200GB and 4 million variants per genome) presents a challenge for data storage, data security, and downstream analysis, requiring extensive cloud storage and innovative bioinformatic tools.

1.3   Genetic Architecture

This thesis explores the genetic architecture of structural renal and urinary tract malformations, but what do we mean by genetic architecture? Broadly speaking, this term refers to the underlying genetic variation responsible for the phenotypic variability of a disease or trait. More specifically, it comprises the number and type of variants that influence a phenotype, their allele frequencies as well as the distribution of their effect sizes (Timpson *et al.* 2018). In addition to this, penetrance (the proportion of individuals with a particular genotype that display the associated phenotype), allelic dominance (when one allele overrides the effect of another allele in the same gene) and interactions between genes (epistasis) and the environment all contribute to phenotypic complexity. Delineating the genetic architecture of a disease is key to understanding its aetiology.

Genetic architecture has traditionally been described as monogenic, oligogenic or polygenic meaning that one, few or many genetic variants or genes influence phenotypic variability (Figure 1.2) (Badano and Katsanis 2002). These classifications have largely been dictated by the types of genetic variation studied and the cost and limitations of available technologies. For example, WES is used to identify rare coding variants in Mendelian disease with simple recessive or dominant inheritance whereas common variants are tested for association with complex traits using genotyping arrays. This focus on the most accessible regions of the genome has resulted in an observational bias, with a distinct lack of large-scale population-based genetic studies focused on structural or rare, non-coding variation.

WGS enables the interrogation of all classes of genomic variation, genome-wide, across the entire allele frequency spectrum in a completely unbiased manner. This has challenged the classic definitions of genetic causality in recent years, blurring the distinction between polygenic/complex traits and monogenic/Mendelian disease (Katsanis 2016), with enrichment of Mendelian disease genes near GWAS loci of

phenotypically related complex traits providing support for a shared genetic basis (Freund *et al.* 2018, Sun *et al.* 2022).
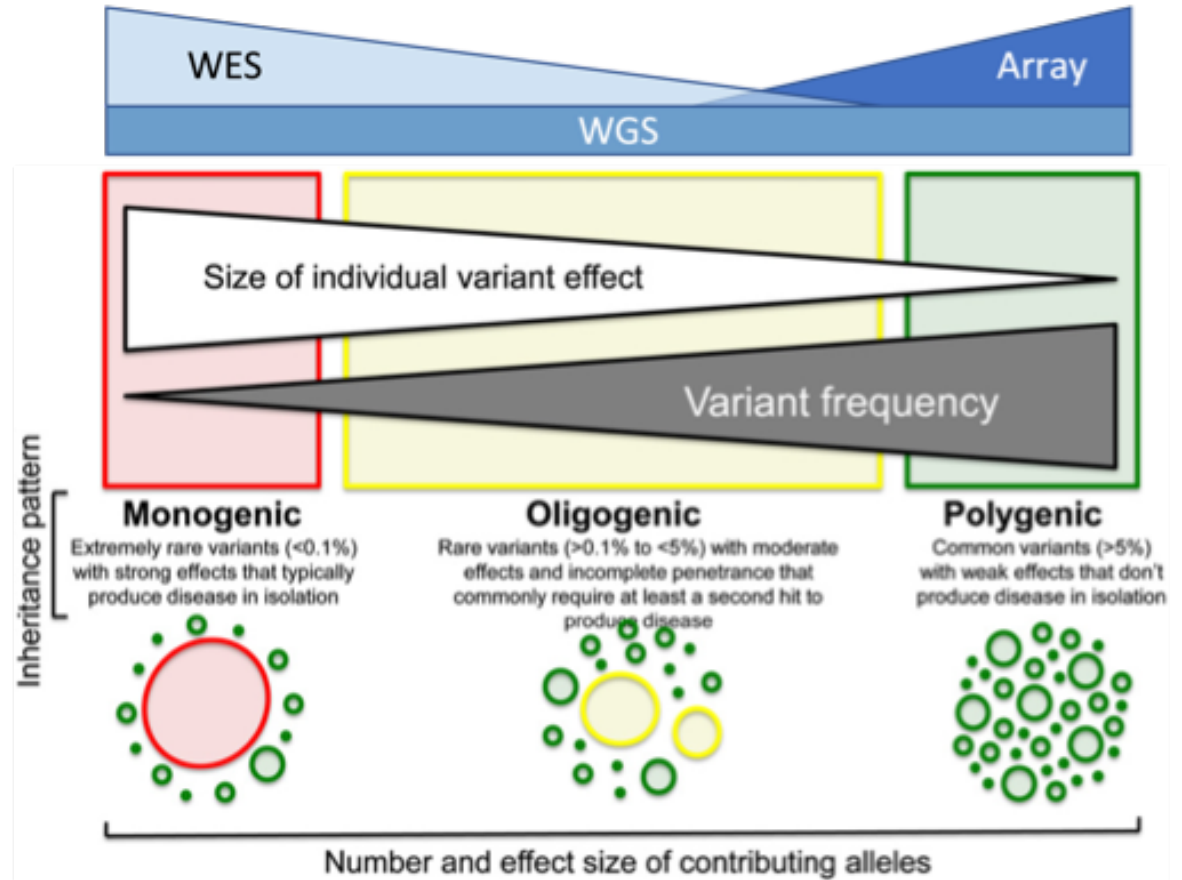
Figure 1.2. Traditional classification of genetic architecture.

Genetic architecture is conventionally classified into monogenic (rare variants with large effect size; red circle), oligogenic (rare and low-frequency variants with moderate effect size; yellow circles) and polygenic (common variants with small effect sizes; green circles). The identification of these variants is dependent on the sequencing technology used. Whole-exome sequencing (WES) detects coding variants and is commonly used to identify rare, likely pathogenic variants in Mendelian disease. Genotyping arrays identify hundreds of thousands of common variants across the genome and are generally used to investigate complex traits. Whole-genome sequencing (WGS) allows genome-wide interrogation of both coding and non-coding variation across the allele frequency spectrum enabling an unbiased assessment of genetic architecture. Adapted from Giudicessi and Ackerman 2013.

Traditional stratification of traits into those associated with common alleles with small effects (odds ratio [OR] 1.1-1.5) or rare alleles with large effects (OR > 2) is now considered too 'reductionist' and there is increasing evidence for a continuous spectrum of rare and common alleles combining to influence a particular phenotype. For example, rare, coding SNVs have been associated with several complex traits (Wang *et al.* 2021, Sun *et al.* 2022) including white blood cell count and plasma low-density lipoprotein cholesterol levels (Cohen *et al.* 2005, Auer *et al.* 2014, Tsai *et al.* 2015) while rare CNVs contribute to disease risk in several complex psychological disorders including schizophrenia and autism (Sebat *et al.* 2007, Marshall *et al.* 2017). Common variants have also been shown to contribute to the risk and clinical presentation of rare Mendelian disease such as severe neurodevelopmental disorders (Niemi *et al.* 2018) and cystic fibrosis (Corvol *et al.* 2015). Furthermore, common alleles with large effect sizes have been associated with a risk of developing kidney disease that approaches Mendelian levels. Homozygosity for *APOL1* risk alleles is associated with an increased risk for HIV-associated nephropathy (OR 29) and focal segmental glomerulosclerosis (OR 17) as well earlier age of onset and faster progression to end-stage kidney disease (ESKD) (Genovese *et al.* 2010, Kopp *et al.* 2011). Similarly, co-inheritance of common human leucocyte antigen (HLA) haplotypes and *PLA2R1* risk alleles confers an 89-fold and 29-fold higher odds of developing membranous nephropathy in East Asian and European populations, respectively, (Stanescu *et al.* 2011, Xie *et al.* 2020) demonstrating the power of epistasis and further highlighting the complexity of genetic inheritance.

The phenotypic variance of several diseases can also be attributed to the combined effect of multiple different allele types and frequencies. For example, in Charcot-Marie-Tooth disease, an increased burden of rare variants in multiple neuropathy-associated genes contributes to phenotypic variability (Gonzaga-Jauregui *et al.* 2015) whereas in autism spectrum disorder common and rare *de novo* variation acts additively to modify disease risk (Weiner *et al.* 2017). Common, non-coding variants

also modify the penetrance of pathogenic coding variants both in *cis* and *trans,* emphasizing the importance of interactions between the coding and non-coding genome (Wu *et al.* 2015, Castel *et al.* 2018).

These observations have led to a shift in emphasis towards the concept of 'genome-wide mutational burden' where deleterious and protective variation of all types and allele frequencies combine to manifest as a unique phenotype (Lupski *et al.* 2011). An extension of this is the recently proposed 'omnigenic' model which hypothesizes that a complex trait is the result of perturbations in a regulatory network of 'core' and 'peripheral' genes where all genes expressed in a cell can contribute to the phenotype, even in the smallest way (Figure 1.3) (Boyle *et al.* 2017). Fortunately, whole genome sequencing now enables comprehensive assessment of the whole spectrum of genetic variation genome-wide, allowing us to fully describe the genetic architecture or 'mutational burden' of a disease for the first time.
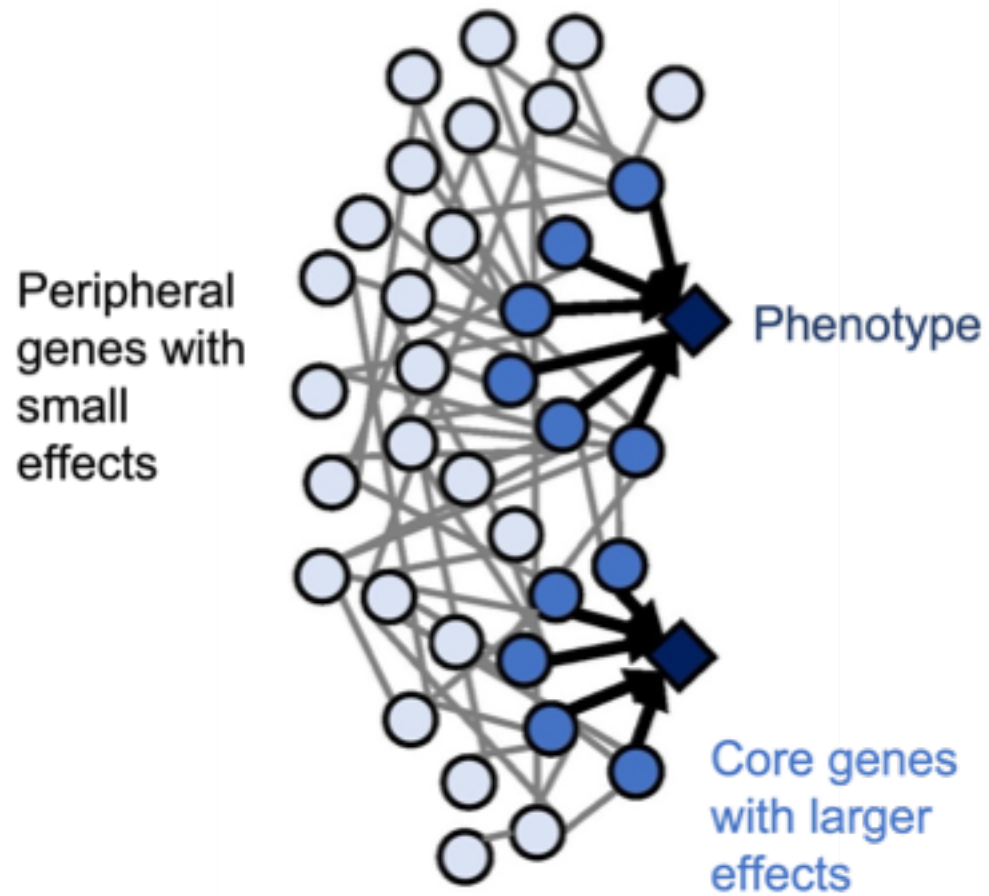
Figure 1.3. The omnigenic model of complex disease.

This model describes a combination of direct effects from core genes (mid-blue) and indirect effects from peripheral genes (light blue) acting in *trans* to manifest as a unique phenotype (dark blue). All genes are considered highly pleiotropic and can affect all traits. Adapted from Fagny and Austerlitz 2021.

1.4   Structural Renal and Urinary Tract Malformations

Structural renal and urinary tract malformations (also referred to as Congenital Anomalies of the Kidneys and Urinary Tract or CAKUT) are the most common cause of chronic kidney disease (CKD) in young children, accounting for 40-50% of those requiring dialysis or a kidney transplant (Harambat *et al.* 2012, Johansen *et al.* 2021, Kramer *et al.* 2021). Paediatric ESKD is associated with significant morbidity and children have a 30-fold higher risk of death than their healthy peers due to predominantly cardiovascular and infection-related causes (Chesnaye *et al.* 2018). This places a significant burden on the affected individual and their family making CAKUT one of the major unmet clinical needs in paediatric nephrology.

CAKUT consists of a heterogenous group of malformations that are characterized by defects in embryonic development and include renal parenchymal malformations (renal agenesis, hypodysplasia or cystic dysplasia), abnormalities in embryonic renal migration or fusion (ectopic, pelvic, or horseshoe kidney), collecting system defects (pelvi-ureteral junction obstruction, duplex kidney, primary megaureter or vesicoureteral reflux) and abnormalities of the lower urinary tract causing bladder dysfunction and/or obstruction (posterior urethral valves, prune belly syndrome and bladder exstrophy). Figure 1.4 illustrates some of the different malformations that are collectively referred to as CAKUT.

Renal hypodysplasia

Vesico-ureteral reflux (VUR)

Renal agenesis

Duplex collecting system

Multicystic dysplastic kidney
(MCDK)

Pelvi-ureteral junction obstruction
(PUJO)

Posterior urethral valves
(PUV)

Primary megaureter
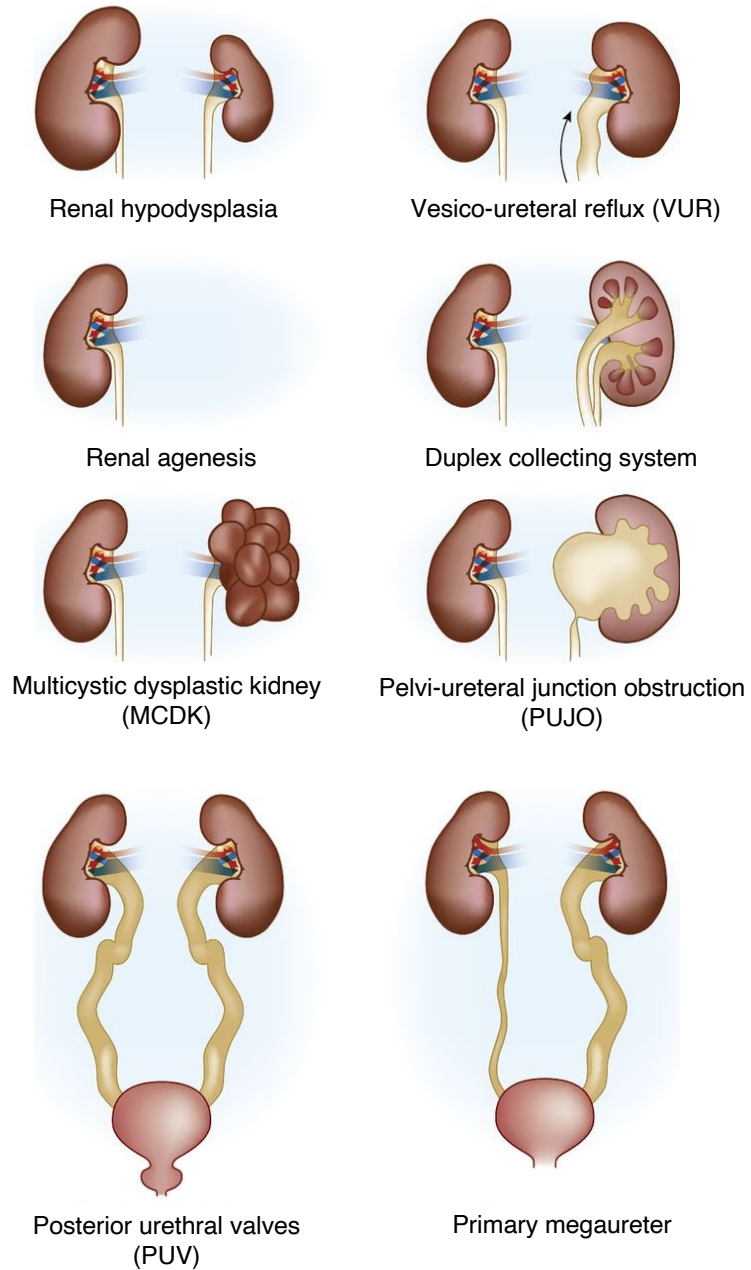
Figure 1.4. Overview of structural renal and urinary tract malformations.

CAKUT comprises a diverse collection of phenotypes affecting the renal parenchyma (renal hypodysplasia, renal agenesis and multicystic dysplastic kidney), collecting system abnormalities (VUR, duplex collecting system, PUJO) and lower urinary tract malformations (PUV, megaureter). Adapted from Westland *et al.* 2020.

1.4.1 Epidemiology

Structural renal and urinary tract malformations occur in approximately 4.2-30 per 10,000 live births (Tain *et al.* 2016, Morris *et al.* 2018) and are the most frequently detected congenital malformation, collectively accounting for 20-30% of all anomalies diagnosed antenatally (Queisser-Luft *et al.* 2002, Nicolaou *et al.* 2015). CAKUT usually occur in isolation, although the co-occurrence of multiple different malformations in the same individual is also observed (Pope *et al.* 1999). In a third of cases, CAKUT is associated with extra-renal features, including other congenital anomalies, diabetes or hearing impairment, and can manifest as part of a syndrome affecting multiple organ systems (Stoll *et al.* 2014). This phenotypic heterogeneity can make the clinical classification of CAKUT challenging. Table 1.2 details the defining clinical features and incidence of the main CAKUT phenotypes.

Table 1.2. The diverse spectrum of CAKUT.

Phenotype	Description	Incidence
Renal parenchymal malformations		
Renal hypodysplasia	Hypoplasia: renal size ≤2 standard deviations of the mean size for age with reduced number of normal nephrons. Usually associated with renal dysplasia: abnormal differentiation of mesenchymal and epithelial elements. Typically produces small, irregular kidneys which may be cystic.	0.2-0.4%
Unilateral renal agenesis	Failure of ureteric bud to develop and induce differentiation of metanephric mesenchyme resulting in an absent kidney.	0.03%
Multicystic kidney dysplasia	Enlarged kidney with multiple non-communicating cysts separated by dysplastic tissue and an absent ureter. Non-functioning and usually involutes by the age of five.	0.02%
Abnormal renal embryonic migration and fusion		
Horseshoe kidney	Renal fusion abnormality with functioning renal parenchyma on both sides of the vertebral column, usually joined at the lower pole. Associated with malrotation and abnormal vascular supply.	0.25%
Ectopic kidney	Kidney fails to ascend during development and can remain in the pelvis, cross the midline, fuse (e.g., crossed fused ectopia) or fail to rotate medially.	0.1%
Anomalies of the collecting system and ureter		
Duplex collecting system	Duplication of renal pelvis and (part of) the ureter. May be associated with VUR, ectopic ureter and ureterocoele.	2%
Primary vesico-ureteral reflux (VUR)	Retrograde flow of urine from the bladder into the upper urinary tract due to an incompetent ureterovesical junction. Usually spontaneously resolves in childhood.	1%
Pelvi-ureteral junction obstruction (PUJO)	Partial or intermittent obstruction between the renal pelvis and ureter impairing drainage of urine. Most common cause of antenatally detected hydronephrosis.	0.2%
Congenital megaureter	Dilated ureter with impaired flow of urine resulting from functional or anatomical abnormality involving the ureterovesical junction.	0.04%
Lower urinary tract malformations		
Posterior urethral valves (PUV)	Persistent membrane in posterior urethra results in congenital bladder outflow obstruction. Only seen in males.	0.03%
Prune belly syndrome	Absence of abdominal wall musculature, bilateral cryptorchidism, hydronephrosis, hydroureter and megacystis. Predominantly in males.	0.004%
Bladder exstrophy	Defect in closure of lower abdominal wall, bladder, and urethra. Part of the bladder exstrophy-epispadias spectrum.	0.002%

1.4.2 Clinical presentation

Many structural malformations of the kidneys and urinary tract are diagnosed during antenatal ultrasound screening. Renal parenchymal malformations may present as absent, echogenic, cystic, or abnormally large or small kidneys. Lower urinary tract obstruction, most frequently associated with PUV in boys, manifests as bilateral hydroureteronephrosis and an enlarged proximal urethra and bladder (resulting in the classic 'keyhole' sign on ultrasound images). At the severe end of the spectrum, lack of functional kidney tissue, as is seen in renal agenesis, results in oligo- or anhydramnios (insufficient or absent amniotic fluid which is generated in part by fetal urine) and the characteristic facies, limb anomalies and pulmonary hypoplasia of the Potter sequence which is associated with significant perinatal mortality and morbidity (Potter 1946).

In those not diagnosed antenatally, CAKUT may manifest as failure to thrive in infancy, recurrent urinary tract infections, voiding difficulties or as part of a CAKUT-associated syndrome. Adults with previously undiagnosed malformations may also present with abnormal urinalysis, kidney stones, hypertension, CKD or with an incidental finding on imaging. Furthermore, it is likely that CAKUT underlies a significant proportion of young adults with 'unexplained' CKD who may present with 'small, scarred kidneys', previously labelled as 'reflux nephropathy'. Management often involves urological intervention to relieve obstruction, prophylactic antibiotics for recurrent urinary tract infections and regular monitoring of blood pressure and kidney function for the development of CKD and its associated complications.

1.4.2.1 HNF1B-associated disease

Pathogenic variation in *HNF1B* is the most frequently identified molecular diagnosis among individuals with CAKUT (Clissold *et al.* 2015). HNF1B is a member of the homeodomain-containing superfamily of transcription factors which mediates the development of the kidneys, liver, pancreas, and urinary tract. Haploinsufficiency

resulting from heterozygous whole-gene deletions (typically a 1.4Mb 17q12 deletion that also includes 14 additional genes) or pathogenic single nucleotide/indel variants result in renal cysts and diabetes syndrome (RCAD; MIM 137920) although the high frequency of *de novo* deletions (~50%) means a family history may not be reported. Mutation hotspots are found in the DNA-binding domain affecting exons 2 and 4, as well as the intron 2 splice site (Chen *et al.* 2010), however there does not appear to be a clear genotype-phenotype correlation. *HNF1B*-related disease is associated with a diverse range of structural renal and urinary tract malformations; renal hypodysplasia, cystic kidneys, single and horseshoe kidneys, malformations of the collecting system as well as autosomal dominant tubulointerstitial kidney disease (ADTKD-*HNF1B*). Extrarenal manifestations include mature-onset diabetes of the young (MODY) type 5, pancreatic hypoplasia, genital malformations, hyperuricemia and early gout, abnormal liver function tests and hypomagnesemia. This highlights the benefits of a molecular diagnosis to guide screening of other organ systems for associated complications.

1.4.3 Clinical outcomes

Clinical outcome is dependent on the type of anomaly involved. Worse renal survival is observed in patients with PUV, solitary kidneys and bilateral renal hypodysplasia (Sanna-Cherchi et al., 2009). Over a third of boys with PUV develop ESKD before the age of 30 years (Heikkilä *et al.* 2011). Coexistent VUR is associated with a higher risk of progression to ESKD (Sanna-Cherchi *et al.* 2009). Although CAKUT has an enormous impact on child health, those affected are more likely to reach ESKD as an adult than a child at a median age of 31 years (Wühl *et al.* 2013), significantly younger than observed in other kidney diseases (median age at ESKD of 61 years). Furthermore, a childhood history of CAKUT, even with normal renal function and blood pressure in adolescence, has been shown to increase the risk of developing ESKD in later life (Calderon-Margalit et al., 2018). This may in part be attributed to

a reduced number of nephrons which has been linked to the development of adult-onset hypertension and CKD in the context of low birth weight and prematurity (Luyckx *et al.* 2013, Crump *et al.* 2019).

For individuals who develop kidney failure, transplantation is the treatment of choice. Short-term outcomes after kidney transplantation are favourable in both children and adults with CAKUT when compared to individuals with primary glomerular or non-CAKUT diseases at 5-10 years, although recipients were more likely to be younger and receive a pre-emptive transplant (Cornwell *et al.* 2021). Over the longer term, an increased risk of graft loss has been reported in individuals with PUV suggesting that graft survival may be limited by bladder dysfunction (McKay *et al.* 2019).

Interestingly, a large historical cohort study identified an increased risk of kidney, ureter and bladder cancer in individuals with CAKUT, although the absolute risk was small (Calderon-Margalit *et al.* 2021). A link between congenital anomalies and increased risk of malignancy has also been suggested by various case reports. For example, horseshoe kidneys have been linked with Wilms tumour and other cancers (Reed and Robinson 1984, Krishnan *et al.* 1997, Huang *et al.* 2004), VACTERL has been associated with Barrett's oesophagus (Su *et al.* 2012), and classic bladder exstrophy with bladder cancer (Smeulders and Woodhouse 2001). In addition, biallelic loss of *HNF1B* is associated with the development of chromophobe renal cell carcinomas (Rebouissou *et al.* 2005, Sun *et al.* 2017). These reports are consistent with the observation that disruption of highly conserved developmental pathways critical for the regulation of cell proliferation, apoptosis and differentiation, may also contribute to carcinogenesis in later life (Dempke *et al.* 2017).

1.5 Embryonic Development of the Kidneys and Urinary Tract

Structural renal and urinary tract malformations were originally thought to result from urinary outflow obstruction causing physical stress on the kidneys (Peters *et al.* 1992) or from ectopic ureteric budding (Mackie and Stephens 1975). However in 2002, Ichikawa *et al.* proposed their paradigm-shifting 'mulitgenic' theory, suggesting that CAKUT was instead the result of an accumulation of minor mutations in multiple specific genes involved in renal morphogenesis (Ichikawa *et al.* 2002). This hypothesis was prescient of the 'mutational burden' or 'omnigenic' models of genetic architecture discussed above and contrasts with the prevailing view that CAKUT is a predominantly monogenic disease.

The development of the renal tract is a dynamic, highly coordinated sequence of morphological events that may be disrupted at any stage and result in pathology that can evolve over time. This results in a diverse array of CAKUT phenotypes depending on the spatiotemporal context of the insult. Although these structural malformations can vary widely, they often share common developmental pathways and regulatory networks which will be reviewed below.

1.5.1 Development of the kidney

The kidney is derived from intermediate mesoderm and its development consists of three sequential stages that proceed in a rostral to caudal (head-to-tail) manner alongside an elongating epithelial tube called the mesonephric (or Wolffian) duct:

- **Pronephros:** develops during the 3rd week of gestation forming a transient, rudimentary, and non-functioning system that degrades by week four. It is an evolutionary artefact, analogous to the kidneys of primitive fish such as lampreys.

- **Mesonephros:** develops during the 4th week of gestation, generating basic functioning nephrons which start to excrete urine, although most of these subsequently degenerate. This structure is like the kidneys of fish and amphibians.

- **Metanephros:** By the 5th week of gestation, the ureteric bud branches from the caudal part of the mesonephric duct into the metanephric mesenchyme to become the metanephros, the precursor to the adult kidney.

Reciprocal induction of the ureteric bud and metanephric mesenchyme is mediated by the GDNF/RET signaling pathway, disruptions of which can lead to failure of, or ectopic, ureteric bud induction resulting in renal agenesis or a duplex or ectopic kidney, respectively (Schedl 2007). Glial cell line-derived neurotrophic factor (GDNF) is a growth factor ligand secreted by the metanephric mesenchyme which binds to the receptor tyrosine kinase RET and its co-receptor glial cell-derived neurotrophic factor family receptor α1 (GFRα1) at the tip of the ureteric bud to stimulate cell proliferation and ureteric bud branching. This key signaling pathway is regulated by several transcription factors including PAX2, GATA3, EYA1, SALL1 and FOXC1, which have all been linked to renal malformations.

Continued reciprocal interaction between the ureteric bud and the metanephric mesenchyme trigger *branching morphogenesis* where iterative tip division forms an epithelial tubule network (Figure 1.5). Disruption of these processes are thought to result in renal hypodysplasia. At the same time, elongation of the ureteric bud outside the metanephric mesenchyme generates the collecting system and ureter.

The metanephric mesenchyme condenses at each ureteric tip to generate the cap mesenchyme containing nephron progenitor cells maintained by SIX2 which undergo *mesenchymal to epithelial transition.* This polarized epithelium lines the

renal vesicle, which twists and elongates to form a comma and then S-shaped body, before forming a primitive nephron (Figure 1.5).
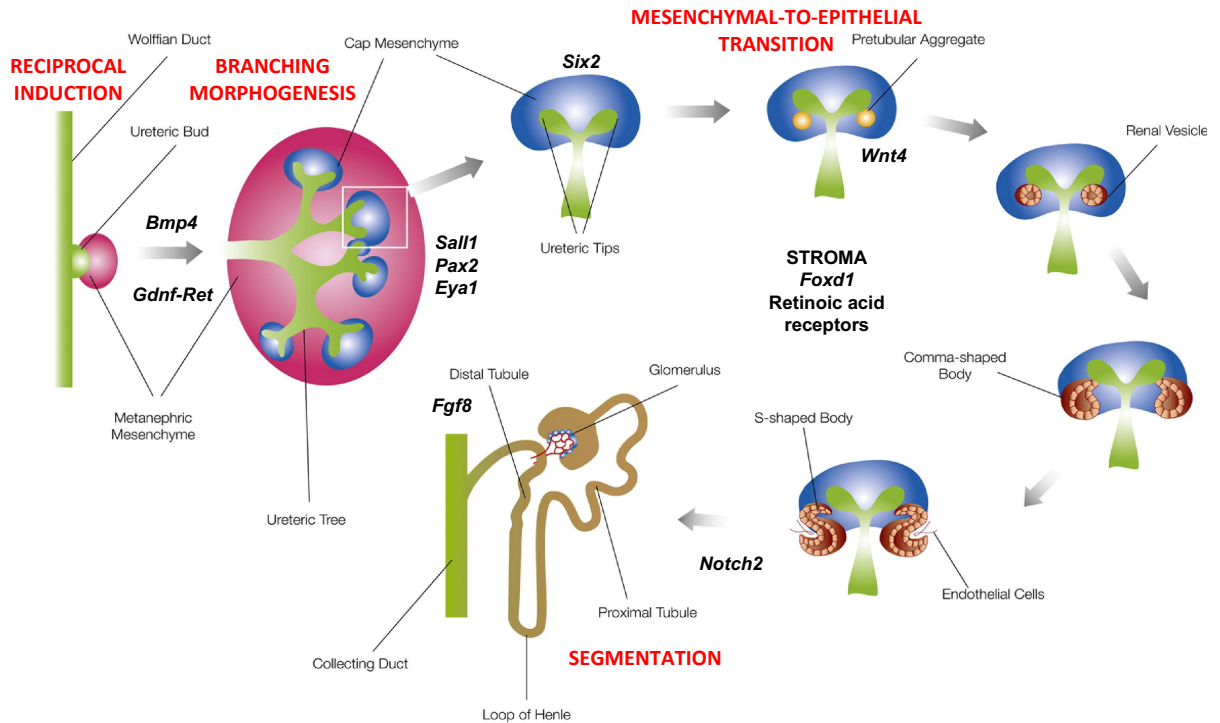
Figure 1.5. The different stages of kidney development.

Reciprocal induction occurs between the ureteric bud and metanephric mesenchyme mediated by the GDNF/RET signaling pathway. The ureteric bud then undergoes branching morphogenesis to form a network of epithelial tubes. Nephron progenitor cells maintained by *SIX2* at the tip of the buds undergo mesenchymal to epithelial transition to form a primitive nephron. Genes with a clearly defined role in murine nephrogenesis are shown. Adapted from Walker and Bertram 2011.

At approximately 8 weeks the metanephros migrates out of the pelvis to its final upper lumbar position. Interruptions to this process can result in pelvic or ectopic kidneys. The metanephros starts to function and produce urine from 9-10 weeks, with nephrogenesis continuing until 36 weeks of gestation. Indeed, 60% of nephrons are formed in the last trimester, which has important clinical implications for preterm

and low-birth-weight infants. Reduced nephron number in these individuals has been associated with an increased long-term risk for hypertension and CKD (Luyckx *et al.* 2013, Crump *et al.* 2019).

1.5.2 Development of the lower urinary tract

At the same time, the lower urinary tract is forming from the endodermal cloaca. The urorectal septum divides the cloaca into ventral and dorsal parts that develop into the urogenital sinus and rectum, respectively (Figure 1.6). The urogenital sinus gives rise to the early bladder, the urethra and vestibule of the vagina in females, and the posterior urethra in males. The urinary bladder is initially drained by the allantois; however, this is obliterated during development and forms a fibrous cord (the urachus) which can be seen as the median umbilical ligament in adults.
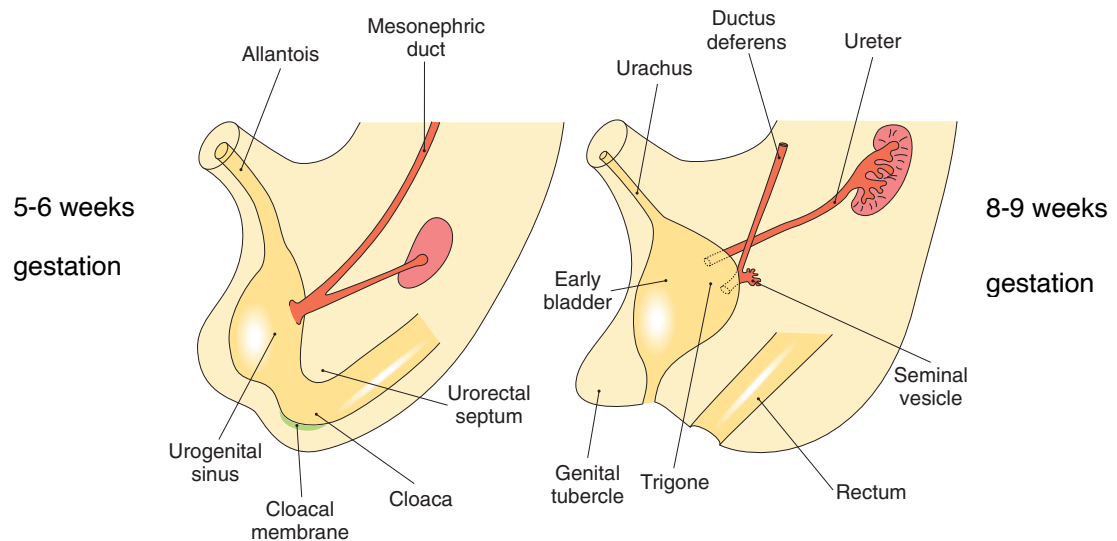
Figure 1.6. Development of the lower urinary tract.

The lower urinary tract is formed from the endodermal cloaca which is divided by the urorectal septum into the ventral urogenital sinus and dorsal rectum. The urogenital sinus forms the bladder, urethra and vaginal vestibule in females, and posterior urethra in males. The mesonephric duct inserts laterally into the cloaca at approximately 4 weeks of gestation. Adapted from Johnson *et al.* 2018.

At approximately 4 weeks gestation the caudal mesonephric duct integrates into the urogenital sinus to form the base of the bladder (trigone) with the resulting ureteric orifices moving cranially as the kidneys ascend. Abnormal insertion of the mesonephric duct into the developing bladder (e.g., in a perpendicular rather than angled manner) can result in an incompetent uretero-vesical junction and VUR. The mesonephric duct is then either absorbed into the prostatic urethra to become the ejaculatory ducts in males or regresses in females. Defective integration of the mesonephric duct into the posterior urethra or persistence of the urogenital membrane have been proposed as possible mechanisms underlying PUV, the most common cause of ESKD in boys, but the exact developmental origin is currently unknown (Krishnan *et al.* 2006).

1.5.3   Molecular pathways of renal morphogenesis

Murine models of CAKUT have provided insights into the developmental pathways required for normal renal tract development (Schedl 2007, van der Ven, Vivante, *et al.* 2018). As mentioned in Section 1.5.1, the GDNF-RET signaling pathway is essential for communication between the ureteric bud and metanephric mesenchyme. Several other key developmental pathways implicated in the pathogenesis of CAKUT are listed below:

- Bone morphogenic protein (BMP) signaling regulates ureteric budding and maintains the nephron progenitor population (Nishinakamura and Sakaguchi 2014).
- Canonical Wnt/β-catenin signaling mediates several critical processes during kidney development including ureteric bud induction (*Wnt11*) and mesenchymal to epithelial transition of the cap mesenchyme to form the renal vesicle (*Wnt4, Wnt9b*) (Schedl 2007).

- The planar cell polarity (PCP) pathway coordinates the orientation of cells across a tissue plane and displays widespread involvement in kidney development (Torban and Sokol 2021).

- Retinoic acid signaling from surrounding stromal cells is important for the insertion of the mesonephric duct into the cloaca (Chia *et al.* 2011) and branching morphogenesis, mediated via its effects on *Ret* expression in the ureteric bud (Batourina *et al.* 2001).

- Hedgehog signaling activity is observed predominantly in the developing collecting ducts and medullary regions. Deletion of *Shh* or its receptor *Ptch1* results in renal agenesis and hypoplasia, respectively (D'Cruz *et al.* 2020).

- Fibroblast growth factor (FGF) signaling is necessary for both branching morphogenesis and distal tubule development (*Fgf8*) (Bates 2011).

- Notch signaling is important for patterning of the proximal tubule (Schedl 2007).

1.6   The Genetic Basis of CAKUT

Monogenic mouse models and over 150 single-gene CAKUT-associated syndromes suggest a monogenic component to CAKUT (van der Ven, Vivante, *et al.* 2018). Case reports detailing familial clustering of renal agenesis and dysplasia (Hack *et al.* 1974, Carter *et al.* 1979, McPherson *et al.* 1987, Murugasu *et al.* 1991, Arfeen *et al.* 1993, Battin *et al.* 1993, McPherson 2007), multicystic dysplastic kidney (Filion *et al.* 1985, Moazin *et al.* 1997, Srivastava *et al.* 1999, Belk *et al.* 2002, Sekine *et al.* 2005, Watanabe *et al.* 2005), VUR (Sirota *et al.* 1986, Van den Abbeele *et al.* 1987, Noe 1992, Connolly *et al.* 1997), PUV (Kroovand *et al.* 1977, Doraiswamy *et al.* 1983, Grajewski and Glassberg 1983, Schreuder *et al.* 2008, Frese *et al.* 2019), classic bladder exstrophy (Shapiro *et al.* 1984, Reutter *et al.* 2003), and Prune Belly syndrome (Lockhart *et al.* 1979) lend additional support to the hypothesis that there

is a significant genetic basis underlying this disorder. Furthermore, ultrasound screening of asymptomatic first-degree relatives can identify structural renal tract abnormalities in 4-23% depending on the phenotype in question and degree of consanguinity (Roodhooft *et al.* 1984, Schwaderer *et al.* 2007, Bulum *et al.* 2013, Manoharan *et al.* 2020, Viswanathan *et al.* 2021), although discordant malformations between the index patient and family member are often observed, highlighting the significant intra-familial variability associated with this disorder.

Candidate gene studies, targeted NGS, and WES have so far identified over 50 monogenic causes of CAKUT (Table 1.3). Many of these genes encode transcription factors or key components of developmental signaling pathways, implying that disruption of the tightly controlled transcriptional networks that govern embryogenesis is critical to the pathogenesis of this condition. However, one of the major challenges facing clinicians and researchers is the substantial genotypic and phenotypic heterogeneity observed in CAKUT, where the same phenotype can result from changes in several different genes and a single gene can be associated with several different phenotypes. Furthermore, variable expressivity (where individuals with the same pathogenic variant display different clinical features) and incomplete penetrance (when not everyone with a pathogenic variant expresses the characteristic phenotype) renders the clinical interpretation of variants challenging and means only a minority of affected individuals receive a molecular diagnosis.

It has also been suggested that syndromic and isolated CAKUT are part of the same phenotypic spectrum. This is based on the observation that biallelic protein-truncating variants in *FRAS1*, *FREM2* and *GRIP1* result in the CAKUT-associated Fraser syndrome (MIM 219000) but biallelic missense variants in the same genes cause isolated CAKUT (Kohl *et al.* 2014). Similarly, likely deleterious missense variants in 12 candidate dominant syndromic CAKUT genes were identified in families with isolated CAKUT (van der Ven, Connaughton, *et al.* 2018), suggesting

that hypomorphic alleles (resulting in partial loss of gene function) in syndromic genes may generate a milder phenotype.

Table 1.3. Monogenic causes of CAKUT.

MIM, Mendelian Inheritance in Man; AD, autosomal dominant; XLR, X-linked recessive; AR, autosomal recessive; XLD, X-linked dominant.

Gene	Phenotype	MIM #	Inheritance	Reference
Transcription factors				
CHD7	CHARGE syndrome: choanal atresia, eye coloboma, cardiac, ear and renal anomalies, deafness, short stature, developmental delay	214800	AD	Vissers *et al.* 2004
EYA1	Branchio-Oto-Renal (BOR) syndrome	113650	AD	Abdelhak *et al.* 1997
FOXP1	Syndromic CAKUT: neurodevelopmental and urinary tract phenotypes.		AD	Bekheirnia *et al.* 2017
GATA3	Hypoparathyroidism, sensorineural Deafness and Renal dysplasia (HDR syndrome)	146255	AD	Van Esch *et al.* 2000
GLI3	Pallister-Hall syndrome: renal agenesis/hypodysplasia	146510	AD	Kang *et al.* 1997
HNF1B	Renal cysts and diabetes syndrome: gout, hypomagnesemia, genital tract anomalies	137920	AD	Lindner *et al.* 1999
MYOCD	Megacystis and Prune Belly Syndrome	618719	AD	Houweling *et al.* 2019
PAX2	Renal coloboma syndrome: optic nerve coloboma and renal hypodysplasia Focal segmental glomerulosclerosis (FSGS)	120330 616002	AD AD	Sanyanusin *et al.* 1995 Barua *et al.* 2014
PBX1	Bilateral and syndromic CAKUT	617641	AD	Heidet *et al.* 2017, Le Tanno *et al.* 2017
SALL1	Townes-Brocks syndrome: imperforate anus, dysplastic ears, thumb malformations, renal agenesis/hypodysplasia	107480	AD	Kohlhase *et al.* 1998

Gene	Phenotype	OMIM	Inheritance	Reference
SIX2	Renal hypodysplasia		AD	Weber *et al.* 2008
SIX5	BOR syndrome	610896	AD	Hoskins *et al.* 2007
TBX18	Renal hypodysplasia and PUJO	143400	AD	Vivante *et al.* 2015
ZIC3	VACTERL: Vertebral anomalies, Anal atresia, Cardiac malformations, TracheoEsophageal fistula, Renal and Limb anomalies	314390	XLR	Wessels *et al.* 2010
ZMYM2	Syndromic CAKUT		AD	Connaughton *et al.* 2020
Extracellular matrix proteins				
ANOS1	Kallman syndrome: hypogonadotropic hypogonadism +/- anosmia, renal agenesis	308700	XLR	Hardelin *et al.* 1992
FRAS1	Fraser syndrome: cryptopthalmos, syndactyly, urogenital defects	219000	AR	McGregor *et al.* 2003
FREM1	Bifid nose +/- renal agenesis and anorectal malformations Manitoba oculotrichoanal syndrome	608980 AR 248450 AR		Alazami *et al.* 2009 Slavotinek *et al.* 2011
FREM2	Fraser syndrome	617666	AR	Jadeja *et al.* 2005
GRIP1	Fraser syndrome	617667	AR	Vogel *et al.* 2012
ITGA8	Bilateral renal agenesis	191830	AR	Humbert *et al.* 2014
NPNT	Bilateral renal agenesis		AR	Dai *et al.* 2021
TNXB	VUR	615963	AD	Gbadegesin *et al.* 2013

Cell signaling pathways				
BMP4	Renal hypodysplasia		AD	Weber *et al.* 2008
CRKL	Di-George syndrome: renal agenesis/hypodysplasia		AD	Lopez-Rivera *et al.* 2017
FGF20	Bilateral renal agenesis	615721	AR	Barak *et al.* 2012
GDNF	Hirschsprung disease and CAKUT	613711	AD	Prato *et al.* 2009
GFRA1	Bilateral renal agenesis		AR	Arora *et al.* 2021
GPC3	Simpson-Golabi-Behmel syndrome: overgrowth syndrome, congenital anomalies	312870	XLR	Pilia *et al.* 1996
GREB1L	Bilateral renal agenesis, renal hypodysplasia	617805	AD	Brophy *et al.* 2017, De Tomasi *et al.* 2017, Sanna-Cherchi *et al.* 2017
GREM1	Renal agenesis		AR	Kohl *et al.* 2014
JAG1	Alagille syndrome: renal dysplasia	118450	AD	Kamath *et al.* 2012
LRP4	Cenani-Lenz syndrome: syndactyly, limb malformations, renal agenesis/hypodysplasia	212780	AR	Li *et al.* 2010
NOTCH2	Alagille syndrome: renal dysplasia	118450	AD	McDaniell *et al.* 2006
NRIP1	Renal hypodysplasia and VUR	618270	AD	Vivante, Mann, *et al.* 2017; Zheng et al., 2022
RET	Renal agenesis		AD	Skinner *et al.* 2008

	Neural guidance/patterning			
LRIG2	Urofacial syndrome: dysfunctional urinary voiding, facial grimace on smiling	615112	AR	Stuart *et al.* 2013
HPSE2	Urofacial syndrome	236730	AR	Daly *et al.* 2010, Pang *et al.* 2010
ROBO1	Syndromic CAKUT		AR	Münch *et al.* 2022
ROBO2	VUR	610878	AD	Lu *et al.* 2007
SLIT2	Cystic dysplastic kidneys, unilateral renal agenesis		AD	Hwang *et al.* 2015
SRGAP1	MCDK, renal hypodysplasia, duplicated collecting system		AD	Hwang *et al.* 2015
Primary cilia proteins				
NPHP3	Renal-hepatic-pancreatic dysplasia syndrome	208540	AR	Bergmann *et al.* 2008
Renin-angiotensin system				
ACE	Renal tubular dysgenesis	267430	AR	Gribouval *et al.* 2005
AGT	Renal tubular dysgenesis	267430	AR	Gribouval *et al.* 2005
AGTR1	Renal tubular dysgenesis	267430	AR	Gribouval *et al.* 2005
REN	Renal tubular dysgenesis ADTKD	267430 613092	AR AD	Gribouval *et al.* 2005 Zivná *et al.* 2009
Other				

ACTG2	Megacystis microcolon intestinal hypoperistalsis syndrome	619431	AD	Tuzovic *et al.* 2015
BNC2	Congenital lower urinary tract obstruction	618612	AD	Kolvenbach *et al.* 2019
CHRM3	Prune-belly-like syndrome, impaired pupillary constriction	100100	AR	Weber *et al.* 2011 Beaman *et al.* 2019
CHRNA3	Autonomic bladder dysfunction with impaired pupillary reflex	191800	AR	Mann, Kause, *et al.* 2019
FAM58A	STAR syndrome: syndactyly telecanthus, anogenital and renal anomalies	300707	XLD	Unger *et al.* 2008
LIFR	Renal agenesis/hypodysplasia		AD	Kosfeld *et al.* 2017
TBC1D1	Renal hypodysplasia, megaureter		AD	Kosfeld *et al.* 2016
TRAP1	VUR, renal agenesis, VACTERL		AR	Saisawat *et al.* 2014

It is important to note that some of the monogenic CAKUT associations reported in the literature are supported by relatively weak evidence. For example, variants in *SIX1* are associated with autosomal dominant deafness (MIM 605192) and branchiootic syndrome (MIM 608389) but *SIX1* has also been designated a cause of CAKUT following a single case report of renal hypodysplasia and VUR (Ruf *et al.* 2004) in combination with evidence from mouse models. Subsequent studies have consistently failed to convincingly identify pathogenic *SIX1* variants in individuals with CAKUT (Krug *et al.* 2011, Hwang *et al.* 2014, Negrisolo *et al.* 2014) calling into question the validity of this association. Another key example is *DSTYK,* which was first identified in a single family with a variable spectrum of anomalies ranging from renal hypodysplasia, PUJO and VUR, in addition to epilepsy (Sanna-Cherchi *et al.* 2013). *DSTYK* variants were then looked for in an additional cohort of 311 CAKUT patients, but the prevalence of these variants was not assessed in a control population. The subsequent release of gnomAD (Karczewski *et al.* 2020) has shown that these variants have a higher allele frequency than would be expected for a rare, monogenic disorder. Similarly, *CHD1L* was proposed as a candidate CAKUT gene after heterozygous missense variants were detected in affected individuals (Brockschmidt *et al.* 2012, Hwang *et al.* 2014), however these variants have subsequently been shown to be relatively common within the general population. These examples clearly demonstrate the importance of an adequately controlled genetic study to ensure that any identified gene-disease associations are robust.

Given over 50 monogenic causes for CAKUT have been reported, what proportion of cases do they explain? Table 1.4 summarizes the diagnostic yield from different studies using a variety of targeted, panel-based and WES approaches. Considerable variability is seen in the estimates which range from 1.3-27%, likely resulting from differences in the number of genes screened, as well as in patient selection, where cohorts with a high proportion of consanguinity, severe disease, family history or extrarenal features are more likely to yield a molecular diagnosis.

Table 1.4. Diagnostic yield of targeted and whole exome sequencing in CAKUT.

N, sample size; ES, exome sequencing; AD, autosomal dominant; AR, autosomal recessive; qPCR, quantitative polymerase chain reaction; MLPA, multiplex-ligation dependent probe amplification.

Study	Country	N	Study population	Method	Diagnostic yield
Hwang et al., 2014	International Multicentre	749	650 families – 15% family history. 38% VUR and 16% renal hypodysplasia.	ES 17 AD gene panel	6.3%
Kohl et al., 2014	International Multicentre	672	590 families with isolated CAKUT. Pre-screened for 17 AD CAKUT genes and *HNF1B* deletions.	ES 12 AR genes	2.5%
Groopman et al., 2019	United States Multicentre	531	Congenital or cystic renal disease. Predominantly unselected adults from the AURORA study (50–80-year-old haemodialysis patients) and a US genetic kidney disease biobank.	ES 625 gene panel	23.9%
Nicolaou et al., 2016	Netherlands Multicentre	453	Unrelated, mostly sporadic CAKUT. 10% had family history, 14% extrarenal manifestations. Predominantly collecting system/lower tract defects (duplex, PUJO, VUR and PUV).	ES 208 gene panel	1.3%
Heidet et al., 2010	France Multicentre	377	Kidney anomalies suggestive of *HNF1B*-related disease. Predominantly paediatric.	Sanger sequencing and qPCR of *HNF1B*	19.9%
van der Ven et al., 2018	United States Multicentre	319	232 families pre-screened for *EYA1, PAX2, HNF1B, GATA3, SIX1, SIX5*. Paediatric cohort with high rate of	ES 404 gene panel	14%

			consanguinity (40%), extrarenal manifestations (25%), syndromic disease (7%), multiplex families (17%).		
Heidet et al., 2017	France Multicentre	204	Bilateral CAKUT or unilateral with extrarenal features or family history (mostly kidney anomalies). 45% were severe fetal cases. 34% pre-screened for *HNF1B, PAX2, EYA1, RET.*	ES 330 gene panel	18%
Sanna-Cherchi et al., 2017	United States Multicentre	202	Renal hypodysplasia.	ES	3.5%
Rao et al., 2019	China Multicentre	159	Paediatric cohort with unspecified CAKUT.	ES	17%
Weber et al., 2006	Europe Multicentre	99	Unrelated probands with renal hypodysplasia and CKD from the ESCAPE study (ESCAPE Trial Group *et al.* 2009). Excluded those with associated bladder anomalies/PUV and syndromic disease. 12% family history.	Sanger sequencing of *HNF1B, PAX2, SALL1, EYA1, SIX1*	17% (15% *HNF1B* or *PAX2*)
Ahn et al., 2020	Korea Single centre	94	Paediatric cohort with predominantly kidney anomalies. 66% extrarenal features. 43% ESKD.	ES 60 gene panel	13.8%
Thomas et al., 2011	North America Multicentre	73	Paediatric cohort with renal hypodysplasia and CKD from CkiD study (Atkinson *et al.* 2021). 30% family history.	Sanger sequencing and MLPA of *HNF1B* and *PAX2*	10%

Bekheirnia et al., 2016	United States Single centre	62	Paediatric cohort with isolated and syndromic (31%) CAKUT. 16% with family history. 42% had kidney anomalies.	ES 35 gene panel	5%
Mann et al., 2019	United States Single centre	55	Paediatric kidney transplant recipients.	ES 396 gene panel	18%
Connaughton et al., 2019	Ireland Multicentre	53	45 families with CAKUT – adults with CKD and family history or extrarenal features.	ES	22%
Tanudisastro et al., 2021	Australia New Zealand Multicentre	40	Adult and paediatric CAKUT patients referred for genetic testing.	ES 230 gene panel	13%
Vivante et al., 2017	United States Multicentre	33	Unrelated probands from consanguineous families with syndromic CAKUT.	Homozygosity mapping and ES	27%
Riedhammer et al., 2020	Germany Single centre	30	CAKUT patients recruited with presentation < 18 years, syndromic disease, family history (30%) and/or consanguinity (7%).	ES	27%
Jayasinghe et al., 2021	Australia Multicentre	14	Adult and paediatric CAKUT with extra-renal features.	ES 336 gene panel	21%

1.6.1 Candidate gene studies

Candidate gene studies incorporate prior information from published literature, similar disease associations and animal models to prioritize genes for investigation. This 'hypothesis-driven' approach is attractive when resources are limited and has been used to identify several novel CAKUT genes. One of the first CAKUT-causing genes identified was the transcription factor *PAX2* (paired box gene 2) in a family with optic nerve colobomas (where the optic nerve has not developed properly), renal hypoplasia, mild proteinuria and VUR, what is now called Renal-coloboma syndrome (Sanyanusin *et al.* 1995). Pathogenic *PAX2* variants have since been identified in individuals with isolated renal hypodysplasia, renal cysts, multicystic dysplastic kidney, and VUR (Bower *et al.* 2012, Rossanti *et al.* 2020) as well as focal segmental glomerulosclerosis (Barua *et al.* 2014).

The candidate gene approach has also been employed successfully in several different CAKUT-associated syndromes, particularly where haploinsufficiency (where a single copy of the gene is insufficient for normal function) is the underlying mechanism. A heterozygous loss-of-function variant in *HNF1B* was identified in a family with MODY and non-diabetic kidney disease (Horikawa et al., 1997), with subsequent detection of loss-of-function variants in additional families (Nishigori *et al.* 1998, Lindner *et al.* 1999, Bingham *et al.* 2001, Kolatsi-Joannou *et al.* 2001) leading to the description of 'Renal cysts and diabetes syndrome' (MIM 37920). Similarly, heterozygous loss-of-function variants in the transcription factors *EYA1* (BOR syndrome; MIM 113650; (Abdelhak *et al.* 1997), *SALL1* (Townes-Brocks syndrome; MIM 107480; (Kohlhase *et al.* 1998) and *GATA3* (HDR syndrome; MIM 146255; (Van Esch *et al.* 2000) were all discovered using a similar approach.

The candidate gene approach is often criticised for its high false positive rate meaning many of the identified genes fail to replicate. An example of this is *SOX17,* originally identified as a candidate CAKUT gene after detection of a *de novo* duplication affecting this locus in a girl with syndromic CAKUT (Gimelli *et al.* 2010,

Brockschmidt *et al.* 2012). Screening individuals with both familial and sporadic VUR identified a recurrent p.Y259N variant that was determined to be likely causal by *in vitro* functional studies. Subsequent genetic cohort studies have been unable to replicate this finding, and the allele frequency of the p.Y259N variant in the general population is in excess of that likely to be disease-causing (Combes *et al.* 2012, Heidet *et al.* 2017). This again highlights the importance of robust comparison with a control population to help determine signal from noise.

1.6.2   Linkage analysis

Genome-wide linkage analysis has proved challenging as a method for gene discovery in CAKUT. A combination of incomplete penetrance, genetic heterogeneity, variable expressivity, and a lack of large, affected pedigrees mean that linkage has limited power to map susceptibility loci, especially those with small effect sizes. Linkage studies in CAKUT have primarily focused on familial VUR (Feather *et al.* 2000, Kelly *et al.* 2007, Conte *et al.* 2008, Briggs *et al.* 2010, Cordell *et al.* 2010). However, the majority have either failed to identify genome-wide significant results or failed to replicate previously detected linkage peaks (Sanna-Cherchi *et al.* 2005, van Eerde *et al.* 2007, Darlow *et al.* 2014). The largest genome-wide linkage study in primary VUR to date combined European cohorts from previous studies (Kelly *et al.* 2007, Cordell *et al.* 2010, Darlow *et al.* 2014), carrying out parametric (assuming dominant inheritance) linkage analysis of 1,062 affected individuals from 460 families (Darlow *et al.* 2017). A significant linkage peak was identified at the 10q26 locus (heterogeneity logarithm of odds [HLOD]=4.90) in an estimated 30% of families, however targeted genomic sequencing of this 9Mb region did not identify any of the 69 genes within it to be likely causal.

Two further linkage studies carried out in seven families with renal hypodysplasia (Sanna-Cherchi *et al.* 2007) and a Somalian family with CAKUT (VUR, PUJO and renal agenesis) (Ashraf *et al.* 2010) reported significant linkage peaks on chromosome 1p32-33 and 8q24, respectively, however neither of these loci, or

genes within these regions, have been replicated since. A genome-wide linkage scan in two pedigrees with classic bladder exstrophy also failed to generate any significant peaks (Ludwig, Rüschendorf, *et al.* 2009). Overall, linkage analysis is not adequately powered for gene discovery in the context of a genetically and phenotypically heterogenous disorder such as CAKUT and has largely been replaced by NGS approaches.

1.6.3 GWAS

GWAS has been used to investigate the contribution of common variation (MAF > 1%) to both primary VUR (Darlow *et al.* 2014, 2017, Verbitsky *et al.* 2021) and classic bladder exstrophy (Reutter *et al.* 2014, Draaken *et al.* 2015) with some success. A recent meta-analysis of 1,395 unrelated European individuals with primary VUR and 5,366 controls looked for association at 6.1 million imputed variants under a range of inheritance models (Verbitsky *et al.* 2021). Three genome-wide significant loci (2p15, 6p12.1, 6q14.1) with relatively large effects (1.41-3.65) were identified, as well as five loci with suggestive association. Of note, all the loci were identified using either a recessive or dominant inheritance model or through sex-specific analysis, rather than the conventional additive model which assumes a uniform linear increase in risk. While some of these loci contain genes which have previously been associated with renal tract development or phenotypes, fine-mapping and targeted sequencing is now needed to robustly establish whether any of these are truly causal.

GWAS has also been used to identify susceptibility loci for classic bladder exstrophy. Draaken *et al.* performed a GWAS in 110 unrelated European patients with classic bladder exstrophy and 1,177 controls, identifying significant association at 5q11.1, which was replicated in a meta-analysis with an additional 98 cases (Draaken *et al.* 2015), and in a further meta-analysis with 268 cases and 92 case-parent trios (Zhang *et al.* 2017). This locus contains the transcription factor *ISL1* which is expressed in the urinary tract and bladder of mouse embryos (Zhang *et al.* 2017) and is therefore

considered a plausible candidate gene for classic bladder exstrophy, although again further functional characterization of this locus is required to robustly attribute causality.

1.6.4 Exome sequencing

Exome sequencing (ES) using either a targeted (based on a panel of known or candidate genes) or whole-exome approach has greatly accelerated gene discovery in recent years and been used successfully in both family- and population-based CAKUT studies. For example, WES in parent-offspring trios identified *de novo* (present in the affected patient but not the unaffected parents) variation in two novel candidate genes: *TBC1D1* (Kosfeld *et al.* 2016) and *FOXP1* (Bekheirnia *et al.* 2017). Large, multiplex (multiple affected individuals) families exhibiting autosomal dominant inheritance have been investigated using a combination of genome-wide linkage and WES to identify candidate CAKUT genes including *DSTYK* (Sanna-Cherchi *et al.* 2013), *TNXB* (Gbadegesin *et al.* 2013) and *TBX18* (Vivante *et al.* 2015). Finally, homozygosity mapping (hypothesis-free genome-wide identification of regions of homozygosity using SNP-based arrays) combined with WES can be used for gene discovery in consanguineous families where inheritance is presumed to follow an autosomal recessive pattern (Saisawat *et al.* 2014, Vivante, Hwang, *et al.* 2017, van der Ven, Kobbe, *et al.* 2018, Saygili *et al.* 2020).

In addition to these family-based approaches, cohorts of predominantly unrelated probands have been interrogated with either targeted or whole ES to reveal novel candidates including *GREB1L* (Sanna-Cherchi *et al.* 2017), *PBX1* (Heidet *et al.* 2017), and *ZMYM2* (Connaughton *et al.* 2020). It is important to note, however, that some of these genes are yet to be robustly replicated which highlights the critical importance of statistical rigour in studies that generate large amounts of data (as discussed in Section 1.6).

Targeted and whole-exome burden analysis using ES data (where rare variants are collapsed into gene-based units and compared between cases and controls) has

also been performed in three studies, none of which identified any single gene with an excess of rare likely deleterious variation (Nicolaou *et al.* 2016, Heidet *et al.* 2017, Sanna-Cherchi *et al.* 2017). This suggests that the contribution of single-gene causes to CAKUT is less than previously suspected, although given the largest cohort reported so far consisted of 434 cases and 498 controls, these studies are likely to be underpowered to detect any genes other than those accounting for a substantial proportion of cases (Nicolaou *et al.* 2016).

1.6.5   Copy number variation

Copy number variation (CNV) describes a type of unbalanced structural variation ($\geq$ 50bp) where regions of DNA are either gained (duplications, triplications) or lost (deletions). These regions may encompass a single gene or multiple genes, resulting in a change in gene dosage, and have been associated with cancer (Yi and Ju 2018) and neurodevelopmental phenotypes including congenital anomalies (Mefford *et al.* 2007, Sebat *et al.* 2007, Stefansson *et al.* 2008, Greenway *et al.* 2009, Cooper *et al.* 2011, Marshall *et al.* 2017). Traditionally, large chromosomal imbalances (500kb-5Mb) were detected using standard karyotyping or fluorescent *in situ* hybridization (FISH), however high-resolution microarrays such as array comparative genomic hybridization (array-CGH) and SNP microarrays are now commonly used in clinical practice and can identify genome-wide copy number variation down to a resolution of ~25kb.

Rare, large (>100kb) CNVs have been reported in 6-16% of patients with CAKUT, with the highest frequency observed in those with kidney anomalies including multicystic kidney dysplasia (Nakayama *et al.* 2010, Caruana *et al.* 2015, Xi *et al.* 2016) and renal hypodysplasia (Nakayama *et al.* 2010, Sanna-Cherchi *et al.* 2012, Westland *et al.* 2015, Siomou *et al.* 2017). While CNVs are often associated with neurodevelopmental phenotypes such as intellectual disability and autism, they can also be observed in isolated CAKUT (Sanna-Cherchi *et al.* 2012, Caruana *et al.* 2015, Siomou *et al.* 2017). Deletions at the 17q12 and 22q11 loci are frequently

detected, resulting in Renal cysts and diabetes syndrome (MIM 137920) and DiGeorge syndrome (MIM 188400), respectively, although copy number variation at 1q21.1, 4p16.1-p16.3 (Wolf-Hirschhorn syndrome; MIM 194190), 6p11.2, and 16p13.11 has also been recurrently identified (Sanna-Cherchi *et al.* 2012, Verbitsky *et al.* 2019).

A microarray-based case-control study performed a comprehensive assessment of copy number variation in patients with CAKUT comparing the burden of rare (<0.1%), large (>100kb), exonic CNVs in 2,824 cases with 21,498 controls (Verbitsky *et al.* 2019). Significant enrichment of CNVs was observed in both syndromic and isolated CAKUT, primarily driven by deletions in individuals with kidney anomalies, but with distinct patterns observed between the different phenotypes, suggesting that renal tract development may be sensitive to perturbations in gene dosage.

While rare copy number variation seems to play a role in the pathogenesis of CAKUT, studies have so far been limited to the assessment of large, coding CNVs detected by microarray-based approaches. The contribution of smaller CNVs (< 100kb), those affecting the non-coding genome, or other types of structural variation (e.g., deletions, duplications, inversions) has yet to be investigated.

1.6.6 Epigenetics

Epigenetics describes the reversible modifications made to a cell's DNA that influence gene expression without altering the DNA sequence itself. Examples include DNA methylation, which typically acts to repress or silence transcription, and histone modifications (e.g., methylation, acetylation, phosphorylation) which regulate chromatin structure and binding of effector molecules. At present little is known about whether these modifications contribute to the pathogenesis of CAKUT, although reports of differentially methylated regions in monozygotic twins discordant for renal agenesis (Jin *et al.* 2014) and hypomethylation of *PAX2* in the ureters of patients with VUR (Zheng *et al.* 2015) suggest epigenomics may play a role. Large-scale epigenome-wide association studies (EWAS) are now feasible and have been used

to identify regions of differential methylation associated with kidney function, CKD and renal fibrosis (Chu *et al.* 2017). However, as the epigenome changes during life, epigenomic differences between cases and controls, even where statistically very different, do not imply causation in the same way that genetic differences do.

1.7 Non-genetic causes of CAKUT

1.7.1 Lower urinary tract obstruction

Urinary tract obstruction *in utero* has been shown to damage the developing kidneys and bladder, with severity dependent on the timing and length of the obstruction (Farrugia 2016). This has been demonstrated in animal models where unilateral ureteral obstruction (UUO) or urethral obstruction results in hydronephrosis, abnormal cortical ureteric duct branching, renal cystic or dysplastic changes, and changes in renal growth and nephron number (Peters *et al.* 1992, Matsell *et al.* 2002). The elevated hydrostatic pressure resulting from lower urinary tract obstruction leads to tubular dilatation and apoptosis, interstitial inflammation, glomerulotubular injury, and progressive interstitial fibrosis (Chevalier *et al.* 2010) with the severity of fibrosis correlating with kidney dysplasia. While there is some debate whether these malformations are solely related to fetal urinary obstruction as opposed to primary developmental defects which affect the entire urinary tract, *in utero* bladder decompression with vesico-amniotic shunting or fetal cystoscopy can be used in select cases to try and reduce the pressure effects on the developing kidneys. These interventions have not however been shown to impact postnatal renal outcomes (Morris *et al.* 2013, Martínez *et al.* 2015) indicating that the renal parenchymal damage was already irreversible at the time of diagnosis.

1.7.2 Environmental factors

Several epidemiological studies have examined the role of the intrauterine environment in the pathogenesis of CAKUT, with specific focus on maternal health.

Maternal diabetes has been linked with an increased risk of any major congenital malformation (Zhao *et al.* 2015) and there is a reported 50% increased risk of CAKUT in the children of mothers with diabetes, with an estimated 4.1% of CAKUT in the UK associated with gestational diabetes (Parimi and Nitsch 2020). The relative risk (RR) is higher in mothers with pre-existing diabetes (RR 1.97) than gestational diabetes (RR 1.39) highlighting the potential importance of normoglycemia during the critical period of organogenesis. Maternal hyperglycaemia in animal models has also been shown to reduce nephron number (Amri *et al.* 1999) and perturb ureteric branching morphogenesis in offspring (Hokke *et al.* 2013), providing further support for the role of abnormal glucose levels in the development of structural renal tract malformations.

Vitamin A, and its active metabolite retinoic acid, play a key role in embryonic development regulating cell proliferation, patterning, and differentiation of tissues. Abnormal maternal vitamin A levels (both deficiency and excess) and retinoic acid receptor defects have been linked to congenital malformations including CAKUT (Wilson *et al.* 1953, Mendelsohn *et al.* 1994, Mark *et al.* 2009). In mouse models vitamin A/retinoic acid has been shown to regulate ureteric bud insertion into the cloaca and branching morphogenesis via its effects on key renal development mediator *Ret* (Batourina *et al.* 2001, Chia *et al.* 2011). Interestingly, variation in *NRIP1* (nuclear receptor interacting protein 1) and *GREB1L* (growth regulation by estrogen in breast cancer 1-like), both cofactors that interact with retinoic acid receptors, has been associated with isolated CAKUT (Vivante, Mann, *et al.* 2017, Zheng *et al.* 2022) and renal agenesis (Brophy *et al.* 2017, De Tomasi *et al.* 2017, Sanna-Cherchi *et al.* 2017), respectively. This provides further support for a role of vitamin A/retinoic acid in the pathogenesis of CAKUT and raises the interesting prospect that maternal dietary modifications may impact susceptibility to CAKUT in offspring.

Pre-pregnancy obesity has also been associated with an increased risk of CAKUT (Slickers *et al.* 2008, Hsu *et al.* 2014, Macumber *et al.* 2017, Jadresić *et al.* 2021)

with a recent meta-analysis reporting an OR of 1.14 (1.02-1.27). Other proposed maternal risk factors for CAKUT include smoking (Källen 1997, Slickers *et al.* 2008), excess alcohol intake (Slickers *et al.* 2008), folic acid use and *in vitro* fertilization (Groen In 't Woud *et al.* 2016). However, the epidemiological literature in this field is somewhat inconsistent and subject to confounding. Further evidence is therefore needed before causality can be inferred.

1.8   Summary

In summary, there is clear evidence for a strong genetic component underlying the pathogenesis of structural renal and urinary tract malformations. While previous investigations have primarily focused on monogenic causes, only a minority of individuals receive a molecular diagnosis. This suggests that while there are likely to be more, yet unidentified, disease-causing genes, it is highly likely that the genetic architecture of this heterogenous disorder is distinctly more complex than previously thought. Furthermore, compound interactions between genetic and *in utero* factors are likely to contribute to the significant phenotypic variability seen in CAKUT, perhaps via common final pathways. The observation that transcription factors and key developmental pathway mediators are recurrently affected, for example, suggests shared mechanisms that disrupt the tightly regulated transcriptional networks governing embryogenesis.

1.9   This Study

In this study I use WGS data from the 100,000 Genomes Project to characterize the genetic architecture of structural renal and urinary tract malformations. Population-based rare and common variant association testing was performed in over 800 cases and 20,000 controls of diverse ancestry seeking enrichment of single-nucleotide/indel and structural variation on a genome-wide, per-gene, and *cis*-regulatory element basis. This is the first time WGS has been used to

comprehensively investigate both coding and non-coding genetic variation in CAKUT, in one of the largest cohorts of patients studied so far, providing a unique opportunity to delineate the genetic basis of this heterogenous disorder.

Chapter 2: Methodology

In this chapter I detail the methods used for cohort selection, data generation and data processing and provide an overview of key statistical genetics concepts including relatedness estimation, control of population structure and adjustment for multiple testing. I also discuss the key principles of SAIGE, the generalized linear mixed model approach used to carry out both single variant and region-based association testing. Custom code used in this thesis can be found at https://github.com/mmyc3/phd_thesis.

2.1 The 100,000 Genomes Project

The 100,000 Genomes Project (100KGP) was launched in the UK in 2012 with the aim of sequencing 100,000 genomes from patients with cancer, rare disease and their unaffected relatives (100,000 Genomes Project Pilot Investigators *et al.* 2021). Participants were recruited via 13 National Health Service (NHS) Genomic Medicine Centres across the UK. Recruitment completed in December 2018 and a total of 118,448 genomes had been sequenced by January 2022. The Genomics England dataset (version 10) consists of WGS data, clinical phenotypes encoded using a standardized vocabulary of phenotypic abnormalities called Human Phenotype Ontology (HPO) codes (Köhler *et al.* 2021), and retrospective and prospectively ascertained NHS hospital records for 89,139 individuals. Ethical approval for the 100KGP was granted by the Research Ethics Committee for East of England – Cambridge South (REC Ref 14/EE/1112). Written informed consent was obtained from all participants or their guardians.

As one of the largest whole-genome sequencing datasets worldwide the 100KGP offers a unique opportunity to combine high-quality, high-coverage genomic data

with rich clinical and phenotypic information from a national health system. Furthermore, the ability to perform population-based association testing where both cases and controls have been sequenced on the same platform minimizes confounding by technical artefacts. Figure 2.1 provides an overview of the study workflow.

2.1.1 Case selection

Cases were recruited by clinicians as part of the 100KGP 'Congenital Anomaly 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 ESKD
- multiple distinct renal/urinary tract anomalies
- CAKUT with unexplained ESKD before the age of 25 years

Those with a clinical or molecular diagnosis of autosomal dominant or autosomal recessive polycystic kidney disease, or who had a known genetic or chromosomal abnormality were excluded. Testing of *HNF1B* and *SALL1* was recommended prior to recruitment if there was a personal or family history of diabetes mellitus or imperforate anus, ear, or thumb abnormalities, respectively. A total of five probands had documented genetic testing prior to recruitment with either targeted sequencing and/or array-CGH with negative results.
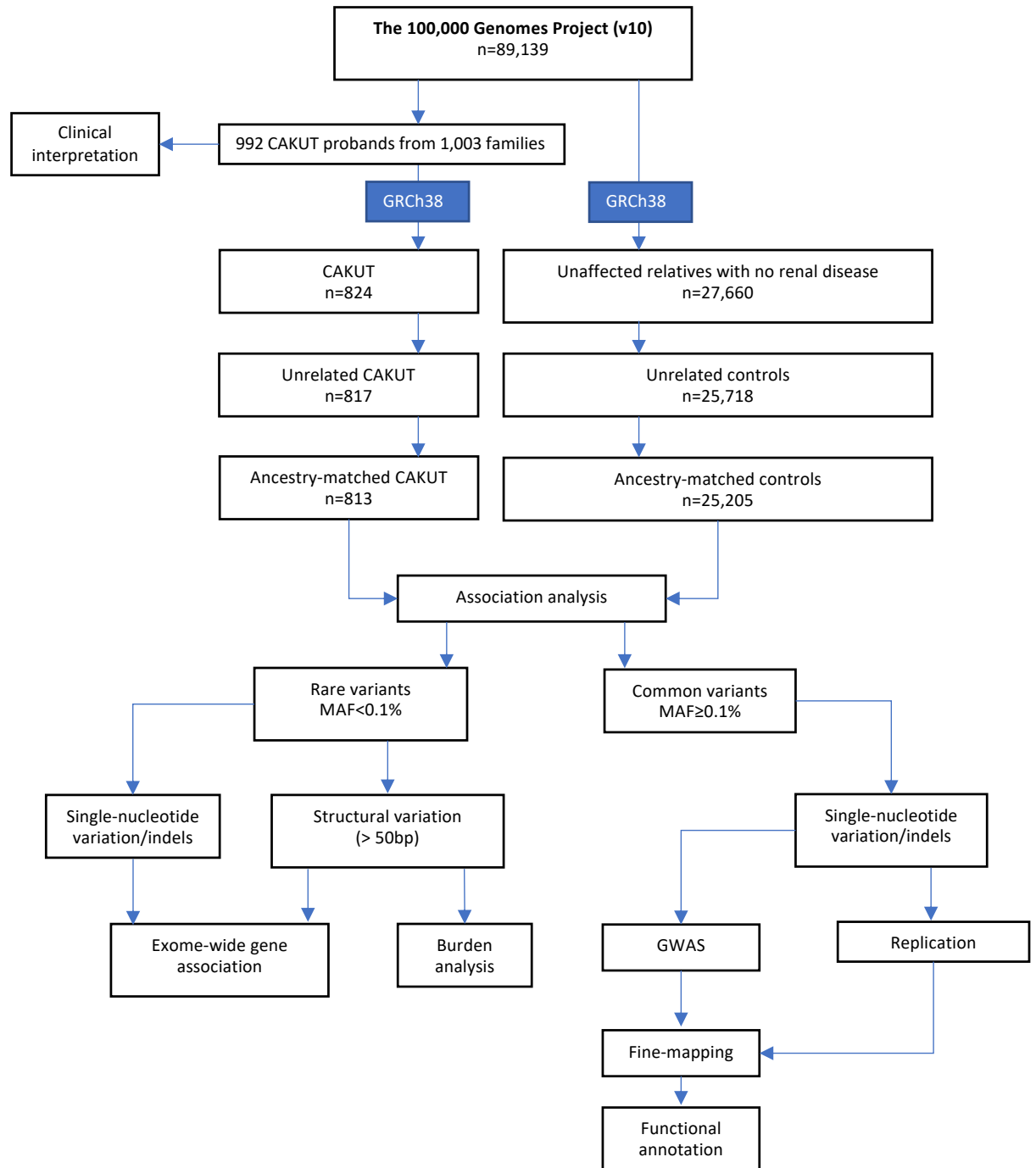
Figure 2.1. Study workflow.

CAKUT, congenital anomalies of the kidneys and urinary tract; MAF, minor allele frequency; GWAS, genome-wide association study. GRCh38 denotes the reference genome build.

2.1.2 Control selection

The control cohort consisted of the unaffected relatives of non-renal rare disease participants from the 100KGP, excluding those with HPO terms and/or hospital episode statistics (HES) data consistent with kidney disease or failure. Given this cohort had not undergone specific ultrasound screening to exclude the presence of renal tract malformations, there remains a possibility that some controls may have asymptomatic CAKUT. However, with over 20,000 controls analyzed it is unlikely that the inclusion of a small number of asymptomatic individuals would significantly affect the outcome.

2.2 Data Generation and Processing

2.2.1 DNA extraction and preparation

DNA collection, extraction and library preparation were performed by Genomics England. In summary, 99% of DNA samples were extracted from blood and prepared using EDTA, with the remaining 1% sourced from saliva, tissue, and fibroblasts. Samples underwent quality control assessment based on concentration, volume, purity, and degradation. Libraries were prepared using the Illumina TruSeq DNA PCR-Free High Throughput Sample Preparation kit to minimize PCR-induced sequencing bias. Where limited DNA was available (<1% samples) the Illumina TruSeq Nano High Throughput Sample Preparation kit was used.

2.2.2 Whole-genome sequencing and alignment

Whole-genome sequencing, alignment, and variant calling were performed by Genomics England. Samples underwent whole-genome sequencing in a single lane of an Illumina HiSeq X instrument generating 150bp paired-end reads which were uniformly processed on the Illumina North Star Version 4 Whole Genome Sequencing Workflow (version 2.6.53.23). With paired-end sequencing both ends of

the 150bp DNA fragment are sequenced, allowing reads to be aligned to the reference genome more precisely, particularly across difficult-to-sequence, repetitive regions of the genome. The reads were mapped to the latest Homo Sapiens NCBI GRCh38 reference assembly and decoys (partially assembled DNA sequences missing from the reference genome) using the Illumina Isaac Aligner (version 03.16.02.19). Alignments had to cover ≥ 95% of the genome at ≥15X with mapping quality > 10 for samples to be retained.

It should be noted that participants recruited to the pilot stages of the 100KGP were originally sequenced and aligned to the NCBI GRCh37 reference genome, however the vast majority of CAKUT probands (83%) have been aligned (or re-aligned) to GRCh38. To exclude batch effects in downstream association analyses only genomes aligned to GRCh38 were used. The diagnostic yield, however, has been calculated using all recruited probands, aligned to either GRCh37 or GRCh38.

Coverage (often denoted with an X) refers to the average number of times a single base is read during sequencing; the more frequently a base is sequenced the more reliable that base call is. Recommended coverage for WGS is >30X but this can be affected by regions with high or low GC-content which can be difficult to sequence. The 100KGP samples achieved a mean of 97.4% coverage at 15X with a median genome-wide coverage of 39X.

High levels of heterozygosity can indicate cross-contamination of samples with DNA from other individuals, leading to false positive results. Samples with >2% cross-contamination as determined by the VerifyBamID algorithm were therefore removed. Male and female subsets were analyzed separately for sex chromosome quality control.

2.2.3 Variant calling

The key challenge in variant calling is distinguishing true genetic variation from technical or sequencing artefacts. Variant calling algorithms incorporate quality

scores and allele counts to predict the likelihood of a genotype at each locus, comparing this to the reference genome and writing the output to a standardized gVCF for downstream analysis. In this study, small variants (SNVs and short insertions/deletions) were called using Illumina's Starling software (version 2.4.7) by Genomics England.

2.2.4 gVCF aggregation and variant-level quality control

For ease of downstream analysis, gVCFs were aggregated by Genomics England using gvcfgenotyper (Illumina, version 2019.02.26) with variants normalized and multi-allelic variants decomposed using vt (Tan *et al.* 2015) (version 0.57721).

Variants were retained if they passed the following filters:

- missingness ≤ 5%
- median depth ≥ 10
- median genotype quality (GQ) ≥ 15
- percentage of heterozygous calls not showing significant allele imbalance for reads supporting the reference and alternate alleles (ABratio) ≥ 25%
- percentage of complete sites (completeGTRatio) ≥ 50%
- *P* value for deviations from Hardy-Weinberg equilibrium (HWE) in unrelated samples of inferred European ancestry ≥$1×10^{-5}$.

HWE is a fundamental principle of population genetics which states that allele and genotype frequencies remain constant from generation to generation if mating is random and if mutation, selection, and migration do not occur. Significant departure from this equilibrium usually indicates genotyping or sequencing errors but may also occur with population stratification or even true associations. For this reason, it is recommended to assess HWE in controls separately to avoid removing potentially true disease associations.

2.2.5   gVCF annotation

Variant annotation was performed using the Ensembl Variant Effect Predictor (McLaren *et al.* 2016) (VEP, version 98.2). Annotations included overall and ancestry-specific allele frequencies from the large publicly available sequencing databases gnomAD (Karczewski *et al.* 2020) (version 3) and TOPMed (Taliun *et al.* 2019) (Freeze 5) as well as Combined Annotation Dependent Depletion (CADD) scores (version 1.5).

CADD is a tool that incorporates >60 different annotations (including evolutionary constraint, epigenetic modifications and functional predictions) via a machine learning model to generate a deleteriousness metric for all ~9 billion potential coding and non-coding SNVs (and some short indels) throughout the human genome (Rentzsch *et al.* 2019). It is widely used to prioritize variants in genetic analyses where a CADD PHRED-scaled score > 20 predicts a variant to be in the top 1% of deleterious variants in the human genome. Numerous other ensemble methods have been released since the development of CADD including DANN (Quang *et al.* 2015) which is trained on the same dataset as CADD, and REVEL for missense variants (Ioannidis *et al.* 2016), however CADD scores offer the most comprehensive genome-wide metric across a wide range of functional categories and genetic architectures and remain the tool of choice for variant prioritization at present.

2.2.6   Bioinformatics tools

Variant filtering was carried out using the command line tools bcftools (version 1.11) (Danecek *et al.* 2021) and BEDTools (version 2.27.1) (Quinlan and Hall 2010). Data manipulation, analysis and plot generation were executed in R (version 4.0.3) using the `dplyr` (https://dplyr.tidyverse.org) and `ggplot2` (Wickham 2011) packages.

The R packages `qqman` (Turner 2014) and `GWASTools` (Gogarten *et al.* 2012) were used to create Manhattan and Quantile-Quantile (Q-Q) plots, respectively, and LocusZoom (version 1.4) (Pruim *et al.* 2010) to visualize regions of interest.

2.3 Relatedness Estimation

Inclusion of related individuals in population-based association studies can result in spurious associations and biased estimates of the standard error of effect sizes if not properly accounted for. It is therefore common practice to perform case-control association studies using 'unrelated' individuals, usually defined as more distant than second-degree relatives.

Given closely related individuals are more likely to share identical alleles, genetic relatedness can be estimated by determining the proportion of loci where a pair of individuals share 0, 1, or 2 alleles inherited from a common ancestor and that are identical-by-descent (IBD). These IBD estimates can be used to calculate a pair-wise kinship coefficient (ϕ), defined as the probability that a randomly selected allele from two individuals is IBD. A kinship coefficient of 0.5, 0.25 and 0.125 is equivalent to monozygotic twin/duplicate, first-degree (parent-offspring or full-sibling) and second-degree relative, respectively.

To determine genetic relatedness within and between cases and controls I used a set of 127,747 high quality autosomal biallelic SNVs with MAF > 1% that had been generated using PLINK (version 1.9) (Purcell *et al.* 2007) by Genomics England. SNVs were included if they met all the following quality control criteria:

- missingness < 1%
- median GQ ≥ 30
- median depth ≥ 30
- AB Ratio ≥ 0.9
- completeness ≥ 0.9

Ambiguous SNVs (A/C or G/T) where it is unclear which allele is on the forward or reverse strand were excluded. To prevent confounding of genetic relatedness estimation, LD pruning was performed using a squared correlation coefficient (r^2) threshold of 0.1 and window of 500kb to remove correlated variants as well as those

in regions of long-range high LD (https://genome.sph.umich.edu/wiki/Regions_of_high_linkage_disequilibrium_(LD). SNVs out of HWE in any of the African (AFR), East Asian (EAS), European (EUR) or South Asian (SAS) 1000 Genomes populations were also removed (pHWE < 1×10^{-5}).

Using this final variant set, I employed the PLINK 2.0 (Chang *et al.* 2015) implementation of the KING-Robust algorithm (Manichaikul *et al.* 2010) which provides robust relationship inference in the presence of population substructure, unlike other algorithms which assume a homogenous population and reliable estimates of allele frequency. Using this approach, I first generated pairwise kinship matrices for case and control cohorts separately and ascertained a subset of unrelated samples using a kinship coefficient threshold of 0.0884 (second degree relationships). I then recalculated the kinship matrix for the combined case-control cohort, removing any controls that had evidence of close genetic relatedness to the cases using a custom Python script (Mr. Catalin Voinescu, UCL).

2.4   Population Stratification

Population structure presents a challenge for genetic association studies. Differences in allele frequencies between populations arise due to non-random mating within geographically isolated groups, genetic drift (random fluctuations in allele frequency) or low rates of migration and gene flow (the transfer of genetic material from one population to another). This can result in inflation of test statistics and false positive associations if cases and controls are sampled from different underlying populations. To address this, various statistical approaches to minimize confounding by population structure can be used either alone or in combination.

2.4.1 Genomic control

Genomic control uses the distribution of test statistics to estimate an overall genome-wide inflation factor (λ) by which the test statistics are then rescaled (Devlin and Roeder 1999). The same correction is applied to all variants irrespective of differences in population allele frequency which can be insufficient for some variants and lead to a loss of power in others.

2.4.2 Principal components analysis (PCA)

PCA uses LD-pruned genome-wide variant data to compute the eigenvectors and eigenvalues of a correlation matrix between individuals (Patterson *et al.* 2006, Price *et al.* 2006). These principal components (PCs) infer continuous axes of genetic variation which can be used to model ancestry differences. Usually, the top ten PCs are included as fixed (non-random) effects in the regression model. While PCA is widely used, it is less reliable at estimating population substructure and loses power when sample sizes are small (Yu *et al.* 2008, Stoltzfus 2011).

2.4.3 Case-control matching

This can be performed using principal components (Luca *et al.* 2008), genetic similarity (Guan *et al.* 2009) or stratification (Epstein *et al.* 2012) scores to optimize genomic similarity between cases and controls.

2.4.4 Linear mixed models (LMM)

LMMs are now standard practice in association studies and can be applied to both continuous and binary traits (Kang *et al.* 2010, Chen, Wang, *et al.* 2016). This variance component approach estimates the genomic similarity between pairs of individuals using an empirical genomic relationship matrix (GRM) generated using genome-wide data. The GRM is then incorporated as a random effect when fitting the null model. Although computationally intensive, LMMs have the added

advantage that they can account for both known and cryptic relatedness as well as inter- and intra-population structure.

2.4.5 Control of population structure

Given the mixed ancestry composition of the cohort, two of the above approaches were chosen to minimize the effects of population structure in the association analyses. First, case-control ancestry matching was performed using a custom R script (https://github.com/APLevine/PCA_Matching). Cases were matched to controls within a distance threshold calculated using the top ten principal components (generated using PLINK 2.0 (Chang *et al.* 2015) from 127,747 high quality autosomal biallelic SNVs with MAF > 1%) weighted by the percentage of genetic variation explained by each component (Figure 2.2). Only controls within a specified distance of a case were included, with each case having to match a minimum of two controls to be included in the final cohort.
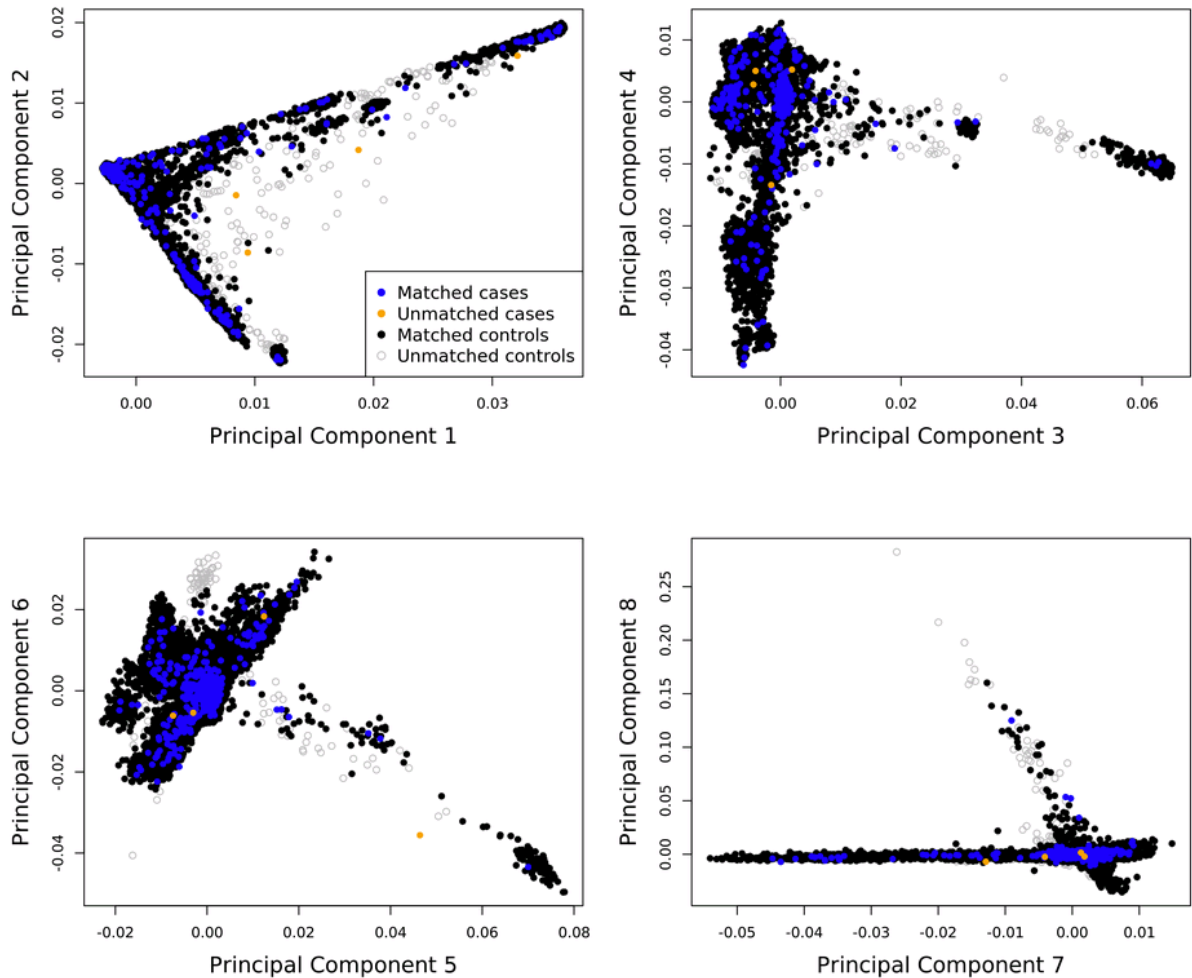
Figure 2.2. Principal component analysis.

Ancestry matching of 817 CAKUT cases to 25,718 controls. Each case had to match a minimum of two controls to be included in the analysis. The orange dots represent the 4 cases that were excluded. 513 controls were removed (grey circles).

Following this, a logistic mixed model approach as implemented by SAIGE (Zhou *et al.* 2018) and SAIGE-GENE (Zhou *et al.* 2020) was used to further control for both population structure and cryptic relatedness in this ancestry-matched cohort.

2.4.6 European cohort selection

A European-only cohort was used for the CAKUT heritability analysis and PUV sequencing-based GWAS. Individuals with ≥ 0.8 probability of European ancestry as determined by a random forest model (Ntrees=400) trained on the first eight principal components generated from 1000 Genomes Project (Phase 3) data (implemented by Genomics England) were extracted. PLINK 2.0 (Chang *et al.* 2015) was used to calculate ten principal components across the European cohort with cases matched to controls using the ancestry-matching algorithm described in Section 2.4 (https://github.com/APLevine/PCA_Matching).

2.5 SAIGE

With the arrival of biobanks containing hundreds of thousands of individuals, large-scale genetic association testing has grown exponentially. This presents substantial challenges to computational and memory requirements as well as difficulties in controlling the inflated type 1 error rates seen with binary traits and when case-control ratios are very unbalanced (case: control < 1:100). Recently, SAIGE (Scalable and Accurate Implementation of GEneralized mixed model), and its extension SAIGE-GENE, have been developed to address these issues and have subsequently been widely adopted by the statistical genetics community for genome-wide single variant and exome-wide region-based association testing in large cohorts (Zhou *et al.* 2018, 2020). The key principles underlying SAIGE are described below.

2.5.1 Generalized logistic mixed model

SAIGE infers and accounts for sample relatedness and population structure using a logistic mixed model which can be written as

$$logit(\mu_i) = X_i\alpha + G_i\beta + b_i$$

where μ_i is the probability of individual i being affected conditional on their covariates, genotype, and random effects. X_i is a vector of covariates (e.g., sex and top ten principal components), α is a vector of fixed covariate effects including the intercept, G_i represents a matrix of allele counts (0,1,2) for each qualifying variant and β is the fixed genotype effect. b_i is a vector of random effects that incorporates relatedness (and consequently population structure) between individuals estimated using an *N* x *N* GRM. SAIGE also employs optimization strategies to reduce the computational cost of fitting the null logistic mixed model meaning this approach can be efficiently applied to large sample sizes.

2.5.2 Saddlepoint approximation

When case-control ratios are unbalanced single variant test statistics do not follow a normal distribution resulting in an inflated type 1 error rate. SAIGE incorporates a robust adjustment called the saddlepoint approximation (SPA) to control for this (Dey *et al.* 2017). The SPA is used to estimate the null distribution, calibrating the distribution of test statistics to generate more accurate *P* values. When variants are rare (minor allele count [MAC] $\leq$ 10) however, this approach is less accurate and efficient resampling is employed, performing permutation testing in only those who carry the minor allele to estimate the sampling distribution and generate empirical *P* values.

2.5.3 Workflow

 SAIGE and SAIGE-GENE (https://saigegit.github.io//SAIGE-doc/) consist of two main steps:

1.   Fitting the null generalized logistic mixed model (GLMM) which includes sex and the top ten principal components as covariates (fixed effects) but no individual genetic variants (see Section 2.5.1). The GRM is constructed using variants with MAF > 1% and the resulting variance components used as random effects.

2.   Testing for association between each variant (using a score test for hypothesis testing under a likelihood-based framework) or qualifying variant set (using SKAT-O) and the phenotype. The saddlepoint approximation is used to account for case-control imbalance.

There are several limitations with the SAIGE approach which are relevant to this study. It has been shown to be slightly conservative when case-control ratios are very unbalanced and estimation of effect sizes (β) is biased for rare variants (when Firth logistic regression should be used instead).

2.6   Stratification by Phenotype

Each association analysis was first carried out using the entire CAKUT cohort before being stratified by phenotype. HPO codes were used to group probands into separate cohorts (Table 2.1). Individuals with more than one anomaly were included in multiple groups.

Table 2.1. HPO codes used to stratify by phenotype.

Cohort	HPO codes
Kidney anomalies	abnormal renal morphology (HP:0012210), renal dysplasia (HP:0000110), renal hypoplasia/aplasia (HP:0008678), renal hypoplasia (HP:0000089), renal agenesis (HP:0000104), unilateral renal atrophy (HP:0008717)
Obstructive uropathy Excluding PUV and bladder exstrophy	abnormal renal collecting system morphology (HP:0004742), abnormality of the renal pelvis (HP:0010944), dilatation of the renal pelvis (HP:0010946), hydronephrosis (HP:0000126), hydroureter (HP:0000072), congenital megaureter (HP:0008676), ureteropelvic junction obstruction (HP:0000074), ureteral obstruction (HP:0006000), ureterovesical junction obstruction (HP:0030735)

VUR	vesicoureteral reflux (HP:0000076)
PUV	congenital posterior urethral valve (HP:0010957)
Cystic dysplasia	multicystic kidney dysplasia (HP:0000003), polycystic kidney dysplasia (HP:0000113), cystic renal dysplasia (HP:0000800), renal cortical cysts (HP:0000803), multiple renal cysts (HP:0005562), renal cyst (HP:0000107)
Bladder exstrophy	bladder exstrophy (HP:0002836)

All individuals included in the association analyses were unrelated and ancestry-matched to controls. Table 2.2 details the final numbers of cases and controls included in each cohort.

Table 2.2. Number of unrelated, ancestry-matched cases and controls per phenotype.

Phenotype	Cases	Controls
CAKUT	813	25,205
Kidney anomalies	237	22,733
Obstructive uropathy	177	24,451
VUR	174	22,562
PUV	132	23,727
Cystic dysplasia	112	24,084
Bladder exstrophy	97	22,037

2.7 Power

Determining power for region-based association testing is challenging due to the large number of parameters that need to be considered, including the allele frequencies and effect sizes of individual variants. To this end, PAGEANT (Derkach

et al. 2018) was developed to generate power calculations for gene-based collapsing tests using the underlying distribution of gene size and MAF of variants from the ExAC dataset (Lek *et al.* 2016). This tool was used to determine the minimum proportion of cases that could be explained by a single gene detected with 80% power in the rare variant analyses (see Chapter 3), under the assumption that 80% of qualifying variants were causal with an exome-wide significance threshold of P=2.5x10^{-6}.

Statistical power for the discovery and replication single-variant association analysis was calculated using the R package `genpwr` (Moore *et al.* 2019) under an additive model using the conventional genome-wide significance threshold of P<5x10^{-8}. Figure 2.3 illustrates the power of the mixed-ancestry GWAS at varying allele frequencies and ORs for the entire CAKUT cohort. With this sample size at an allele frequency of 1%, single variant association testing is sufficiently powered (> 80%) to detect alleles with an OR > 3.
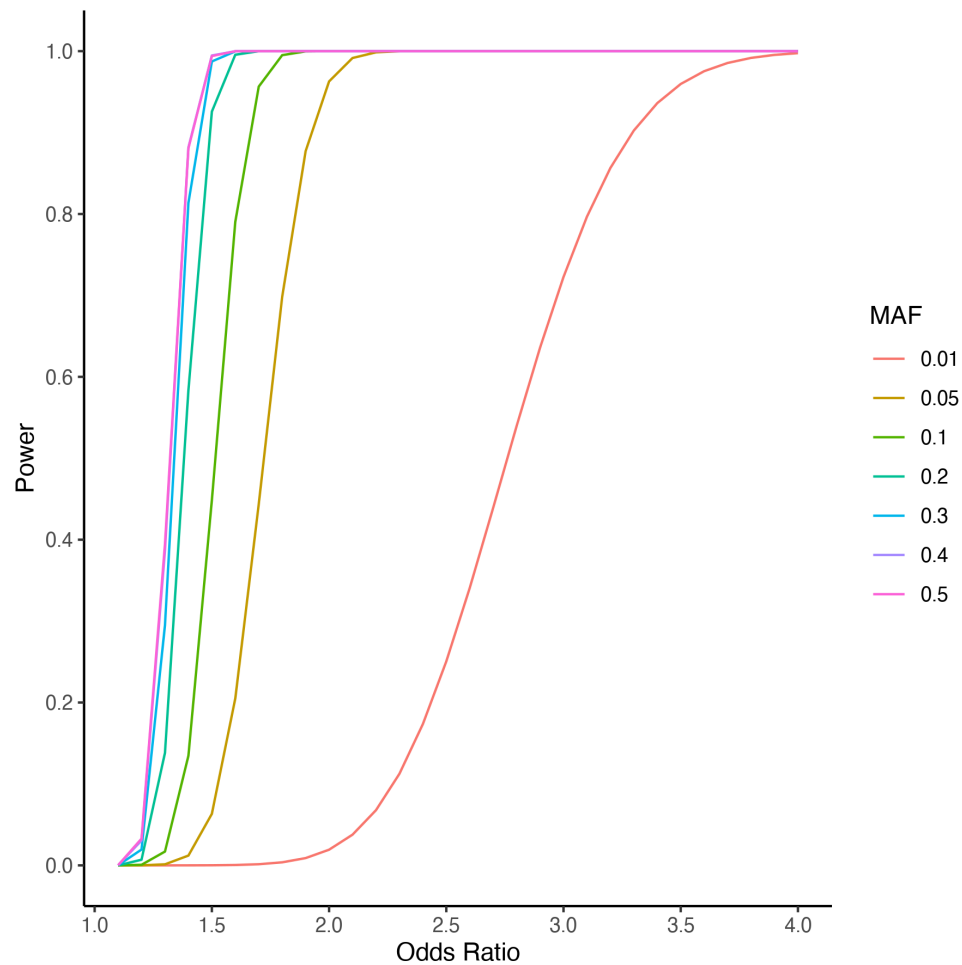
Figure 2.3. Statistical power for CAKUT GWAS.

Power to detect single variant association under an additive model for 813 CAKUT cases and 25,205 controls at a genome-wide significance threshold of 5×10^{-8}. MAF, minor allele frequency.

2.8   Statistical Significance

Determining a threshold of statistical significance that minimizes false positives while maintaining power is critical to making reliable inferences from scientific studies. Genetic association studies present a particular challenge in this regard due to the multiple markers and hypotheses tested; the likelihood of observing a statistically

significant association by chance increases with the number of tests carried out, and with it the likelihood of rejecting the null hypothesis when it is true (Type 1 error). Several different statistical methods can be used to control for multiple testing in association studies each with their own advantages and disadvantages.

2.8.1 Bonferroni correction

The widely used Bonferroni correction sets the significance threshold at α/n, where α is the desired significance level i.e., 0.05 and n is the number of independent tests carried out. This controls the probability of obtaining *at least one* type 1 error but can be overly conservative due to the assumption that every variant is independent of the rest. Use of the Bonferroni correction is appropriate when even a single false positive would be a problem but may result in missed findings due to a high false negative (type 2 error) rate.

2.8.2 False discovery rate

The false discovery rate (FDR) can be controlled by determining an acceptable proportion of false positives among all significant results e.g., 5% or 1% (Benjamini and Hochberg 1995). FDR-controlling procedures have greater power than the Bonferroni correction but at the cost of an increased type 1 error rate. They can however be useful for hypothesis-generating experiments.

2.8.3 Permutation

Permutation testing generates a distribution of the test statistic under the null hypothesis (rather than assuming an underlying distribution) to calculate an empirical P value. This involves resampling the observed data, randomly shuffling case and control labels and calculating a test statistic for each permutation thousands of times. While computationally intensive it can be advantageous when the underlying sampling distribution is unknown.

2.8.4 Bayesian approaches

Bayesian methods are considered more 'intuitive' than classical frequentist approaches which rely purely on null hypothesis significance testing. Bayesian approaches incorporate prior probabilities (e.g., effect size is related to MAF) as well as power to generate a conditional probability based on the observed data and can quantify the degree to which the observed data supports or conflicts with a hypothesis. Bayes Factors (defined as the ratio of the probability of the data under the null and alternative hypotheses) have been proposed as an alternative to *P* values but have not been widely adopted due to computational constraints and concerns regarding the subjectivity of specifying prior probabilities (Wakefield 2009).

2.8.5 Significance thresholds

In this study I used the Bonferroni correction to adjust *P* values for both the rare SNV/indel and structural variant association analyses on the basis that observed signals in these types of study are often difficult to replicate and therefore a stringent statistical significance threshold is required to minimize noise. As mentioned above, one limitation of this approach is that true signals may be missed. Significant associations identified in the structural variant burden testing were confirmed using a permutation approach due to uncertainty regarding the underlying sampling distribution.

For the single variant association analysis, I adopted the conventional and widely implemented GWAS genome-wide significance threshold of 5×10^{-8}. This threshold was originally proposed by the International HapMap Consortium (International HapMap Consortium 2005) based on estimates of the number of common independent variants ($r^2 < 0.8$) with MAF $\geq 5\%$ in a European ancestry population (~1 million). The same threshold is generally applied across all sample sizes and allele frequencies, with no regard to power.

With the recent proliferation of sequencing-based association studies, use of this conventional 5×10^{-8} threshold has come under increased scrutiny. Inclusion of rare and low-frequency variants increases the number of independent tests carried out and consequently the false positive rate. Furthermore, given the greater genetic diversity seen, particularly in individuals of African ancestry, there is an argument to be made for a more conservative significance threshold in non-European populations. In view of this, more stringent P value thresholds for association testing of single variants with MAF $\geq$ 0.1% of 1×10^{-8} and 5×10^{-9} have been proposed, based on simulations using WGS data in European individuals (Fadista *et al.* 2016) and the 1000 Genomes Project (Lin 2019), respectively, but are yet to be widely implemented.

2.9 Summary

In this chapter I have provided an overview of the shared methodology underlying the analyses detailed in this thesis and introduced some fundamental concepts which are key to understanding the statistical genetics approaches employed. More specific methodological considerations and detailed descriptions of the approaches used will be described in the subsequent chapters.

Chapter 3: CAKUT as a monogenic disorder

3.1   Introduction

The importance of rare variants in Mendelian disease is well established with many phenotypes explained by highly penetrant, rare alleles with large effects. Natural selection prevents these damaging alleles from becoming common in the general population, often through a reduction in reproductive fitness. For this reason, rare variants are more likely to be clinically significant and are predicted to provide more direct insights into disease biology. As discussed in Section 1.6, the vast majority of published literature concerning the genetics of CAKUT relates to the identification of rare variants and single-gene defects in individuals and families with renal tract malformations, with these findings cited as support for a predominantly monogenic basis for the disorder (van der Ven, Vivante, *et al.* 2018). In this chapter I explore the contribution of rare (defined here as MAF < 0.1%), coding variants to the genetic architecture of structural renal and urinary tract malformations and discuss whether CAKUT can truly be considered a monogenic disease.

3.2   Aims

1. To determine the prevalence of known monogenic disease in a large cohort of patients with structural renal and urinary tract malformations and identify clinical factors that predict the likelihood of receiving a molecular diagnosis.

2. To discover novel candidate genes using an unbiased exome-wide rare variant association testing approach.

3.3 Methods

3.3.1 Identification of pathogenic variants

All probands underwent assessment via the Genomics England clinical interpretation pipeline to determine a genetic diagnosis (100,000 Genomes Project Pilot Investigators *et al.* 2021). This workflow extracts rare (MAF < 1% for autosomal recessive and MAF < 0.1% for autosomal dominant inheritance), protein-truncating and missense variants that intersect with an expert-curated panel of 57 CAKUT-associated genes (https://panelapp.genomicsengland.co.uk/panels/234/). CNV losses which had ≥ 80% reciprocal overlap with the 17q12 region (encompassing *HNF1B*), 16p11.2 (encompassing *TBX6*), and *NPHP1* were also identified and visualized using Integrative Genomics Viewer (IGV; version 2.9.4). Compound heterozygosity was confirmed where parental DNA was available.

Multi-disciplinary review of candidate variants, considering segregation with disease within a family and mode of inheritance, was undertaken by the local Genomic Medicine Centre with application of the Association for Clinical Genomic Science (ACGS) Best Practice Guidelines for Variant Classification in Rare Disease (https://www.acgs.uk.com/media/11631/uk-practice-guidelines-for-variant-classification-v4-01-2020.pdf) to determine pathogenicity. These well-defined criteria are based on the American College of Medical Genetics and Genomics (ACMG) guidelines (Richards *et al.* 2015) which objectively integrate variant information including population frequency, computational predictions of deleteriousness, functional domain localization, putative mechanism of disease and previous associations with phenotypes in reputable databases to assign one of the following classifications: pathogenic, likely pathogenic, variant of uncertain significance (VUS), likely benign or benign.

Multivariable logistic regression analysis was carried out using the 'glm' function in R to determine significant predictors of a molecular diagnosis using age, sex, family history, consanguinity, and the presence of extra-renal features as factor variables.

3.3.2 Aggregate rare coding variant analysis

3.3.2.1   Overview of rare variant association tests

The analysis of rare variants is more challenging than common variants for two reasons. First, large sample sizes are needed to call rare variants with high confidence and second, single variant association testing is underpowered when variants are rare (Lee, Abecasis, *et al.* 2014). In this cohort, for example, the power to detect association of a variant with an OR of 2 drops from 96% at a MAF of 0.05 to 1.9% at a MAF of 0.01. To address this, several approaches have been proposed to increase power in rare variant association studies including extreme-phenotype sampling or the use of population isolates where rare causal alleles are likely to be enriched, although the generalizability of such results to the underlying population is questionable.

Over the last few years statistical methods often referred to as 'collapsing tests' have become a useful way to boost power where the cumulative effect of multiple variants within a gene, region or pathway can be tested for association with a disease or trait. This approach is useful in the context of allelic heterogeneity where a single variant explains only a small fraction of disease risk and multiple different variants in the same gene can contribute to a phenotype. These collapsing, or rare variant association tests, can be broadly split into burden and variance-component tests; the choice of which is dependent on the underlying genetic architecture of the trait being studied.

3.3.2.1.1   Burden tests

Burden tests collapse rare variants within a region into a single genetic score which is then tested for association with the trait of interest. An example of this using an

additive model would be counting the number of minor alleles across all variants in a region. In such a case the burden test statistic can be represented by:

$$Q_{burden} = \left(\sum_{j=1}^{m} w_j S_j \right)^2$$

where m = the number of variants in the region, w_j = the weight for variant j (e.g., using MAF), and S_j = the score statistic for variant j generated from the sum of allele counts (0,1, or 2) for each individual at variant j, accounting for phenotype. S_j is positive when variant j increases disease risk, and negative when associated with decreased disease risk. A P value is then obtained by comparing the burden test statistic to a x^2 distribution with 1 degree of freedom (Lee, Abecasis, *et al.* 2014).

Various interpretations of the burden test exist including the Cohort Allelic Sums Test (CAST) (Morgenthaler and Thilly 2007), Combined Multivariate and Collapsing (CMC) (Li and Leal 2008) and weighted sum test (WST) (Madsen and Browning 2009), most of which assume a dominant inheritance model. Adaptive burden tests which incorporate variable allele frequency thresholds and functional weights have also been proposed but can be computationally intensive to implement due to the need for permutation testing to estimate P values (Price *et al.* 2010). As a rule, burden tests assume that a large proportion of variants are causal and influence the phenotype in the same direction (i.e., all are deleterious or protective) with the same magnitude of effect (Lee, Abecasis, *et al.* 2014). Violation of these assumptions can result in a loss of power.

3.3.2.1.2   Variance-component tests

Variance-component or dispersion methods test for association by evaluating the distribution of genetic effects for a group of variants. A commonly used approach is the sequence kernel association test (SKAT) which aggregates single variant score test statistics to compute an overall P value, adjusting for covariates to account for

population stratification (Wu *et al.* 2011). The SKAT test statistic consists of the weighted sum of squares of single variant score statistics S_j:

$$Q_{SKAT} = \sum_{j=1}^{m} w_j^2 S_j^2$$

As SKAT collapses S_j^2 rather than S_j (as seen in burden tests), it allows for variants with both positive and negative effects, making it robust to the inclusion of non-causal variants and to the direction and magnitude of effect. However, if a large proportion of variants are causal, variance-component tests lose power and a burden test is preferred.

3.3.2.1.3   Combined tests

As discussed above, the power of burden and variance-component tests can be significantly impacted by the underlying genetic architecture of the disease of interest; however prior knowledge of this is often lacking in practice and can vary from gene to gene. This has led to the development of a data-adaptive optimal test which determines the best linear combination of burden and SKAT tests based on the underlying data to maximise power (Lee, Emond, *et al.* 2012). This SKAT-O combined test statistic is represented by:

$$Q_\rho = (1 - \rho)Q_{SKAT} + \rho Q_{burden}$$

where ρ is a parameter which represents the pair-wise correlation between variant effect coefficients (β). For example, if all variants are acting in the same direction, $\rho = 1$ and the test statistic approximates to Q_{burden}. If the variant coefficients are uncorrelated indicating the direction and magnitude of effects are different, then $\rho = 0$ and the test statistic approximates to Q_{SKAT}. SKAT-O adaptively estimates the optimal value of ρ based on the underlying data and has been shown to perform well across a wide range of disease models.

3.3.2.2   Selection of qualifying variants

A key consideration when using a region-based approach is the choice of 'qualifying' variants. In general, inclusion of variants that are more likely to be disease-causing, i.e., rare, loss-of-function or those predicted *in silico* to be damaging, will increase power to detect association. Based on these assumptions, I extracted coding SNVs and indels to generate two different sets of qualifying variants:

1. Likely deleterious:

- MAF < 0.1% or absent from gnomAD (version 3.1.1)
- Annotated as missense, in-frame insertion, in-frame deletion, start loss, stop gain, frameshift, splice donor or splice acceptor
- CADD (version 1.5) score ≥ 20 corresponding to the top 1% of all predicted deleterious variants in the genome. Indels with missing CADD scores were also kept as most frameshift variants do not have assigned CADD scores.

2. Loss-of-function:

- MAF < 0.1% or absent from gnomAD (version 3.1.1)
- 'High confidence' loss-of-function variants (stop gain, splice site, or frameshift) as determined by LOFTEE (Karczewski *et al.* 2020).

Variants meeting the following quality control (QC) filters were retained: MAC ≤ 20, median site-wide depth in non-missing samples > 20, median GQ ≥ 30. Variants with significantly different missingness between cases and controls ($P < 10^{-5}$) or > 5% missingness overall were excluded. Sample-level QC metrics for each site were set to minimum depth per sample of 10, minimum GQ per sample of 20, ABratio *P* value > 0.001 (allelic imbalance can indicate sequencing errors).

3.3.2.3   SAIGE-GENE

Given that the underlying genetic architecture of CAKUT is largely unknown I chose to use the combined rare variant association test SKAT-O (Lee, Emond, *et al.* 2012)

to maximise power. SKAT-O can be implemented via the R package SAIGE-GENE (version 0.42.1) (Zhou *et al.* 2020), a recently proposed method for region-based association analysis. This approach uses a GLMM to account for population structure and cryptic relatedness and provides a robust adjustment for the high false positive rate seen in case-control imbalanced cohorts with binary traits (see Section 2.5 for more detail).

Chromosome X was analysed separately for males and females. A Bonferroni adjusted *P* value (0.05/number of genes) was used to determine the exome-wide significance threshold.

3.4   Results

3.4.1   Diagnostic yield of WGS in CAKUT

3.4.1.1   Cohort description

A total of 1,091 individuals from 1,003 families with CAKUT were recruited to the 100KGP, of whom 992 were probands: 32.5% were recruited as singletons, 30.8% as trios, and 29.8% as duos. This is one of the largest reported cohorts of CAKUT patients studied to date, and the largest to have undergone WGS. Table 3.1 summarizes the clinical and demographic data for the 992 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.

Table 3.1. Demographics and clinical characteristics of the CAKUT cohort (n=992).

		Number of individuals (%)
Median age (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)
CAKUT phenotypes		
	Kidney anomaly (%)	355 (35.8)
	Obstructive uropathy (%)	228 (23.0)
	VUR (%)	227 (22.9)
	Posterior urethral valves (%)	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)
	Hearing impairment (%)	27 (2.7)
	Cardiac anomalies (%)	21 (2.1)
Family history (%)		132 (13.3)
Reported consanguinity (%)		58 (5.8)
End-stage kidney disease (%)		211 (21.2)
Median age ESKD (range)		12 (0-70)

The most frequently reported HPO terms in the cohort were hydronephrosis (27.0%), VUR (22.9%) and abnormality of the bladder (19.0%) (Figure 3.1).
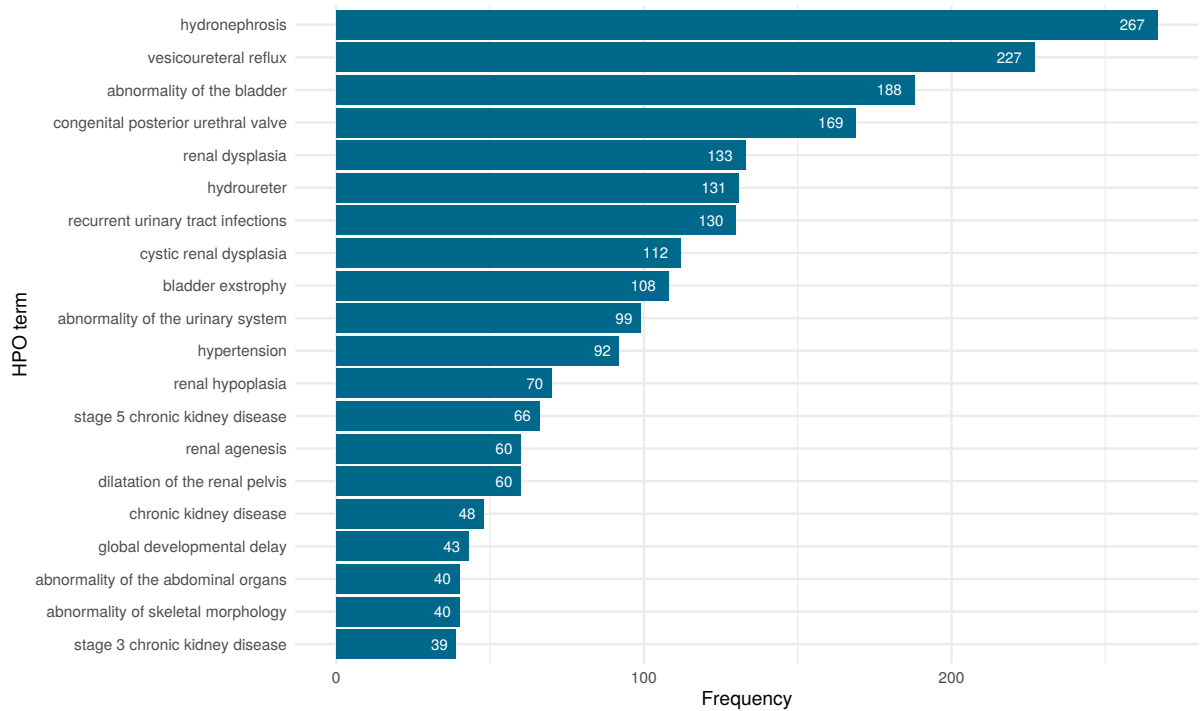
Figure 3.1. Most frequently reported HPO terms in the CAKUT cohort (n=992).

HPO, human phenotype ontology.

3.4.1.2   *Prevalence of monogenic disease*

I first aimed to assess the diagnostic utility of WGS and determine the prevalence of known monogenic disease in this cohort. Of the 992 probands recruited, 975 (98.3%) had undergone multi-disciplinary review of candidate variants from 57 CAKUT-associated genes, classified according to ACGS guidelines. In those who had not yet undergone multi-disciplinary review, I manually curated variants prioritized using Exomiser (https://github.com/exomiser/Exomiser) and reviewed CNV calls for 17q12, 16p11.2 and *NPHP1* to identify any additional potentially disease-causing variants. Overall, 4.3% (43/992) had a molecular diagnosis (Table 3.2), with 74.4% (32/43) of the diagnoses fully explaining the renal/urinary phenotype. Excluding

individuals with PUV and bladder exstrophy, in whom a monogenic cause has not yet been identified, the diagnostic yield was 6.0% (43/715).

In those with isolated CAKUT, pathogenic or likely pathogenic variants in *HNF1B* (2), *PAX2* (2), *PBX1* (2), *TBX6* (1), *GREB1L* (1) and *WT1* (1) were identified. Pathogenic CNVs were detected in five individuals with cystic dysplasia: 17q12 deletions affecting *HNF1B* (4) and a 16p11.2 microdeletion encompassing *TBX6* (1). 32.6% (14/43) had pathogenic or likely pathogenic variants affecting kidney disease genes not classically associated with CAKUT: *PKD1* (5), *PKHD1* (2), *CEP290* (2), *NPHP3* (1), *COL4A5* (1), *CUBN* (1), *SLC3A1* (1), and *CLCN5* (1), which in some cases may represent clinical misclassification or recruitment to the wrong cohort. An additional 25.6% (11/43) had a monogenic cause for a non CAKUT-associated syndrome identified (e.g., Noonan syndrome, Usher syndrome) which did not fully explain the observed renal and/or urinary tract anomalies. Excluding these eleven patients, five individuals with non-CAKUT diagnoses who may have been included in error (*NF1, CLCN5, CUBN, SLC3A1, COL4A5*), and ten more with clear cystic/ciliopathy disease (*PKD1, PKHD1, NPHP3, CEP290*) the diagnostic yield falls to just 2.5% (17/689).

Table 3.2. Clinical and molecular features of 43 CAKUT probands with a genetic diagnosis in the 100KGP.

FH, family history; CONS, consanguinity; ESKD, end-stage kidney disease; CHR, chromosome; POS, position (GRCh38); REF, reference allele; ALT, alternate allele; GT, genotype; ACGS, Association for Clinical Genomic Science; HGVS, human genome variation society; AF, allele frequency; MIM, Mendelian Inheritance in Man; HPO, human phenotype ontology; F, female; M, male; N, no; Y, yes; het, heterozygous; comp het, compound heterozygous; hom, homozygous; hemi, hemizygous; CNV, copy number variant.
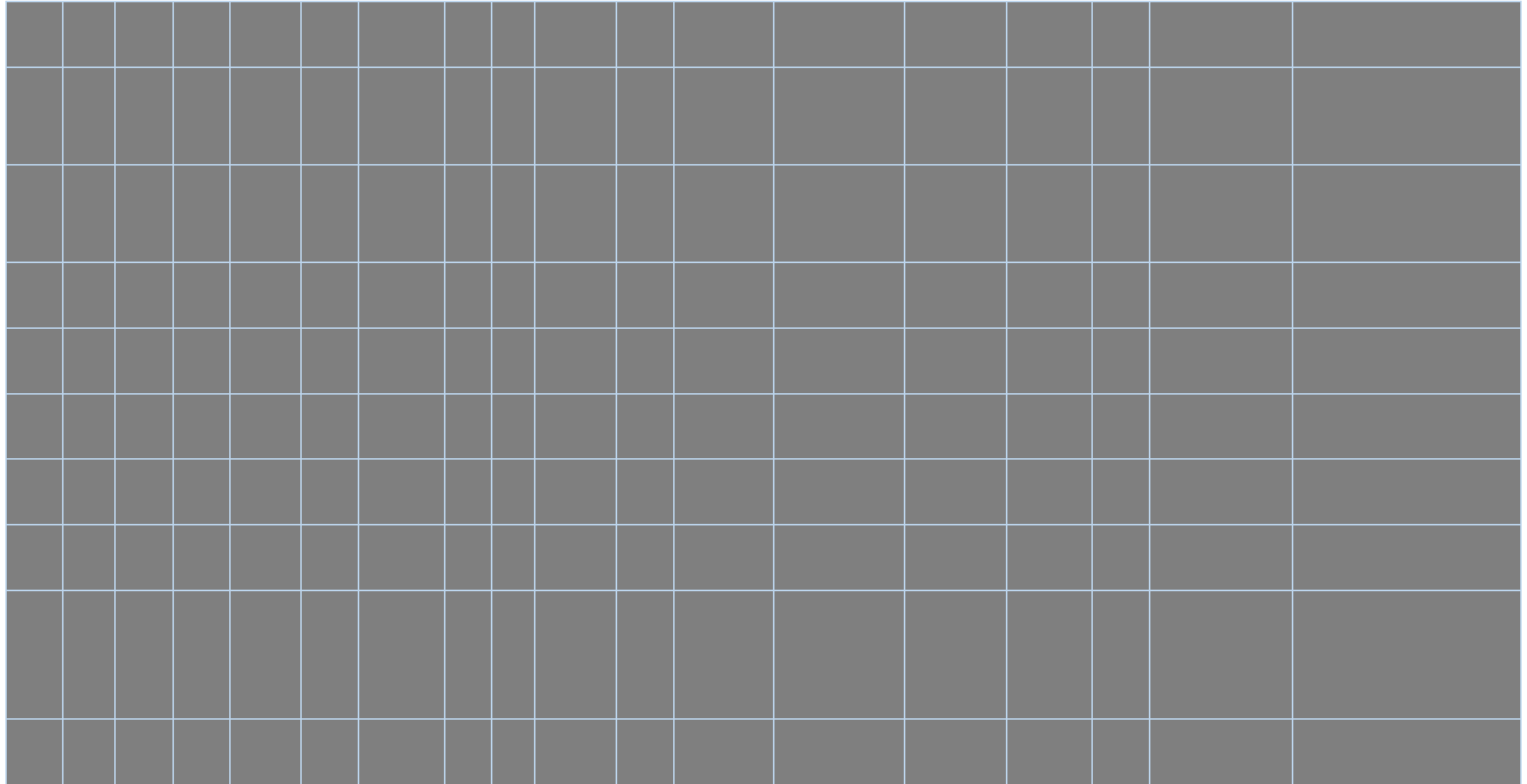
3.4.1.3   Predictors of a positive diagnosis

Although the overall diagnostic yield was low, those with cystic dysplasia and kidney anomalies were more likely to receive a genetic diagnosis (10.7% and 5.9%, respectively), with *HNF1B* and *PKD1* variants accounting for over half of the molecular diagnoses in those with cystic disease (Figure 3.2).
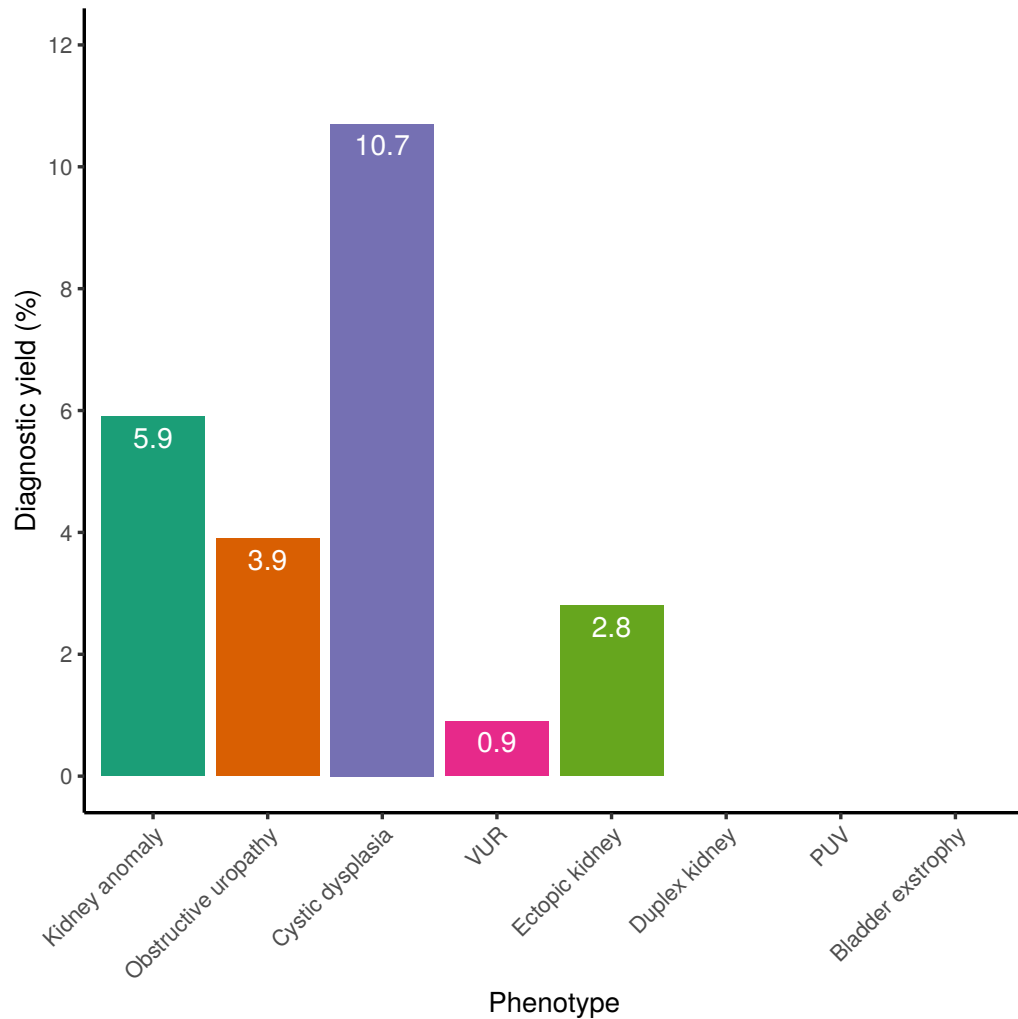
Figure 3.2. Diagnostic yield by CAKUT phenotype.

The white numbers on each bar indicate diagnostic yield (%). VUR, vesico-ureteral reflux; PUV, posterior urethral valves.

The diagnostic yield in individuals < 18 years was significantly higher than in those ≥ 18 years (6.6% vs 1.8%; Fisher's exact P=2.0x10^{-4}; OR 4.0; confidence interval [CI] 1.8-10.0). Multivariable logistic regression also identified extra-renal features (P=1.9x10^{-4}; OR 3.4; CI 1.8-6.6) and family history (P=7.5x10^{-3}; OR 2.7; CI 1.3-5.4) as independent predictors of a positive genetic diagnosis.

3.4.2 Rare variant association testing

3.4.2.1 CAKUT

Given the low prevalence of known monogenic disease in this cohort I next performed rare variant association testing using SKAT-O in a subset of 813 unrelated probands with CAKUT and 25,205 ancestry-matched controls with the aim of identifying novel candidate genes using an unbiased whole-exome approach.

3.4.2.1.1 Likely deleterious variants

Aggregation of rare, predicted deleterious SNVs and indels across 19,168 protein-coding genes did not identify any significant enrichment exome-wide, after correction for multiple testing (Figure 3.3 and Figure 3.4). The median number of variants tested per gene was 46 (interquartile range [IQR] 50) with the greatest number of qualifying variants seen in *TTN* (n=4,687), not unexpectedly, as *TTN* contains both the largest number of exons (363) in any single gene but also the longest single exon (~17kb). While the presence of so many qualifying variants may raise concerns about the introduction of unwanted noise, inclusion of a control cohort that reflects the empirical background variation rate, as well as the adaptive nature of SKAT-O mitigates the potential loss of power caused by including these likely neutral variants.

Although no single gene passed the stringent Bonferroni correction for multiple testing, four genes had a significantly increased burden of likely deleterious variation when using an FDR q-value < 0.05 (as mentioned in Section 2.8.2, FDR thresholds can be more useful for hypothesis generation): *AUTS2* (activator of transcription and

developmental regulator; 133 variants; $P=3.74\times10^{-6}$), *ARHGAP5* (Rho GTPase activating protein 5; 160 variants; $P=6.04\times10^{-6}$), known CAKUT gene *HNF1B* (63 variants; $P=6.06\times10^{-6}$), and *ZNF879* (zinc finger protein 879; 43 variants; $P=6.96\times10^{-6}$). Conditioning on the lead variant in each gene attenuated, but did not abolish, the observed signals (*AUTS2*, $P=1.37\times10^{-4}$; *HNF1B*, $P=2.87\times10^{-4}$; *ARHGAP5*, $P=1.27\times10^{-3}$; *ZNF879*, $P=3.20\times10^{-5}$), implying the detected associations are not being driven by a single variant.
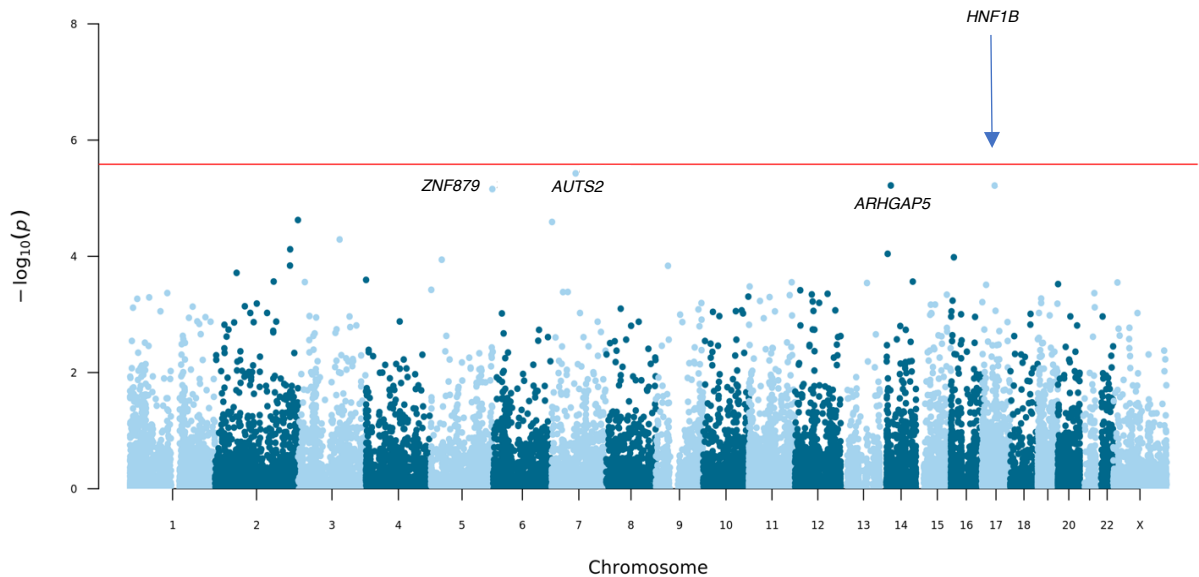
Figure 3.3. Manhattan plot of exome-wide gene-based rare, likely deleterious variant association testing for 813 CAKUT probands and 25,205 ancestry-matched controls.

Each dot represents a gene. The red line indicates the exome-wide significance threshold of $P=2.58\times10^{-6}$. *HNF1B* is highlighted with a blue arrow.
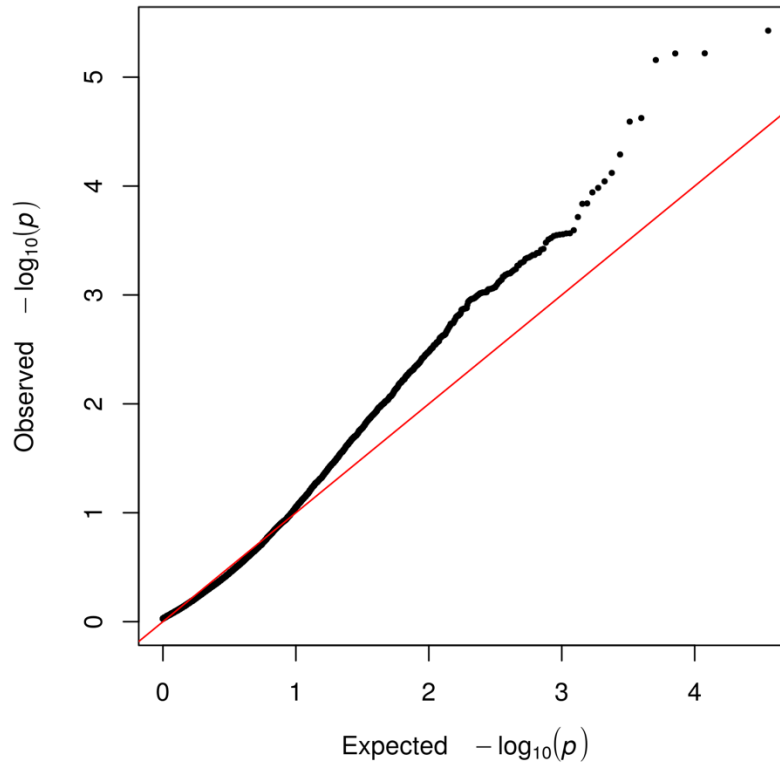
Figure 3.4. Q-Q plot of exome-wide gene-based association testing for 813 CAKUT probands and 25,205 ancestry-matched controls.

SAIGE-GENE was performed for 19,168 genes with loss-of-function and likely deleterious missense variants with MAF < 0.1%. Each dot represents a gene. The red line signifies the observed versus the expected –$\log_{10}(P)$ for each gene tested.

Removal of 32 individuals (with genomes aligned to GRCh38) who had a known monogenic cause did not reveal any exome-wide significant signals in the remaining 781 unsolved CAKUT cases. The top four genes remained the same with attenuated association seen for *AUTS2* (P=4.51x10^{-6}), *HNF1B* (P=1.22x10^{-5}), and *ARHGAP5* (P=1.85x10^{-5}) but stronger association seen for *ZNF879* (P=4.64x10^{-6}).

HNF1B was one of the top three genes to demonstrate evidence of enrichment, consistent with the observation both here and by others that *HNF1B* is one of the most frequently affected genes in CAKUT patients. Interestingly, the lead variant (rs147816724) in *HNF1B* was seen in 4/813 cases (0.5%) compared to 9/25,205 (0.04%) controls (SKAT-O *P*=1.3x10^{-3}). rs147816724 is a missense variant (HGVSp: p.Val61Gly) affecting exon 1 of *HNF1B*, with a CADD score of 22.9 and MAF of 4.34x10^{-4} in gnomAD (version 3.1.2). All four cases with this variant had a diverse range of lower urinary tract phenotypes including PUV, VUR, and bladder abnormalities. rs147816724 has been previously reported in patients with prune belly syndrome (Granberg *et al.* 2012), unilateral multicystic dysplastic kidney as part of the VACTERL association (Hoskins *et al.* 2008), and in an individual with a single kidney (Edghill *et al.* 2006) but has been classified as likely benign/uncertain significance in ClinVar due to its relatively high allele frequency in population databases. Despite being significantly enriched in cases, the wide range of phenotypes associated with this variant and its relatively high population allele frequency suggest it may act as a modifier in conjunction with other additional deleterious alleles, or perhaps be the subject of gene-environment interactions.

The three additional candidate genes with suggestive evidence of enrichment are all biologically plausible, however validation through replication or meta-analysis is necessary before definitive association can be confirmed:

- *AUTS2* is known to regulate the expression of neurodevelopmental genes (Gao *et al.* 2014) and deletions in the gene have been linked to autism spectrum disorder, intellectual disability, and developmental delay (Beunders *et al.* 2013, 2016). Associated congenital malformations including cardiac septal defects, umbilical herniae, and cryptorchidism have also been reported. *AUTS2* is also expressed in the fetal kidney during mid-trimester development, particularly in metanephric cells (Cao *et al.* 2020).

117

- *ARHGAP5* encodes a Rho GTPase-activating protein which mediates cytoskeleton organization by promoting GTP hydrolysis. It regulates cell migration and adhesion and has been linked to several different types of cancer (Gen *et al.* 2009, Yang *et al.* 2021). It is expressed in both ureteric bud and metanephric cells in the developing human fetus (Cao *et al.* 2020).

- *ZNF879* belongs to a large family of transcriptional repressors which recruit co-factors to engage histone modifiers and induce heterochromatin formation (condensed transcriptionally inactive chromatin). It is not highly expressed in mid-trimester fetal kidney (Cao *et al.* 2020).

3.4.2.1.2   Loss-of-function variants

Given the lack of statistically significant gene-based enrichment using both protein-truncating and predicted deleterious missense variants, I repeated the analysis using a subset of loss-of-function (LoF) variants to try and increase power to detect association. However, again no significant enrichment was seen across 17,200 genes after correction for multiple testing using the Bonferroni adjustment (Figure 3.5). A median of 6 (IQR 7) LoF variants per gene were tested. *HNF1B* was not significantly enriched for LoF variation (*P*=0.06) implying that protein-truncating variants are not driving the observed signal seen in the combined analysis.
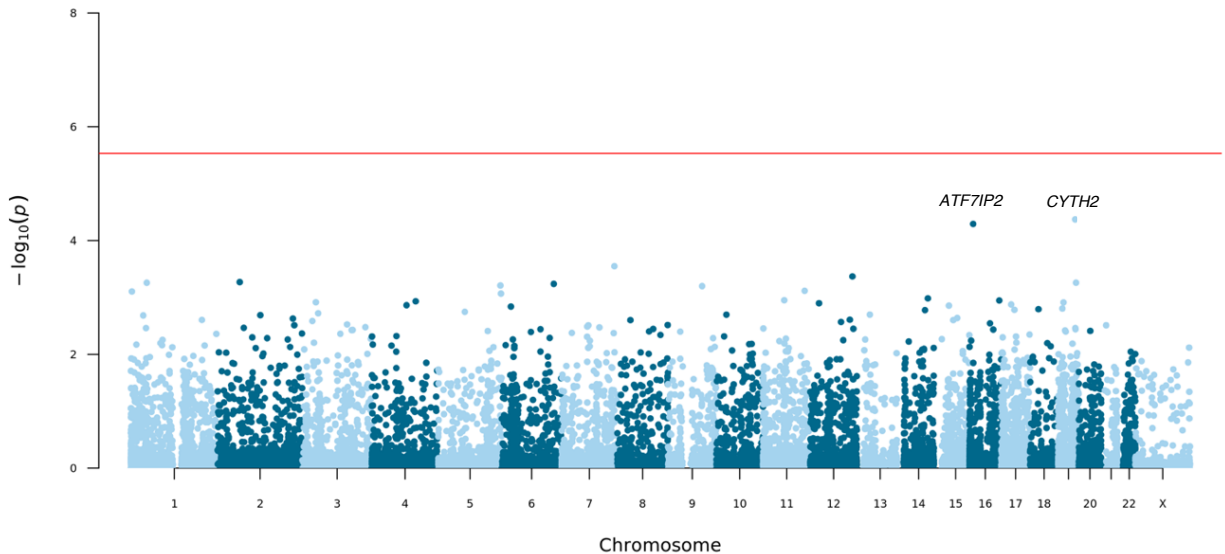
Figure 3.5. Manhattan plot of exome-wide gene-based rare, loss-of-function variant association testing for 813 CAKUT probands and 25,205 ancestry-matched controls.

Each dot represents a gene. The red line indicates the Bonferroni-corrected exome-wide significance threshold of P=2.91×10^{-6}.

The minimum P value observed was 4.25x10^{-5} at *CYTH2* (cytohesin 2) which was non-significant in the combined protein-truncating and missense analysis (P=0.25). Three individuals (0.4%) with PUJO, PUV and bladder exstrophy were heterozygous for loss-of-function *CYTH2* variants compared to 11 controls (0.04%). CYTH2 (also known as ARNO) is a guanine nucleotide exchange factor for the small GTP-binding protein ARF6 and has been shown to activate the epidermal growth factor receptor (EGFR) pathway. It is depleted of loss-of-function variation in gnomAD indicating it is relatively constrained (LOEUF 0.47). Interestingly, *Cyth2* is highly expressed in the embryonic mouse ureteric bud and metanephric mesenchyme at E11.5 and the bladder at E13 (Figure 3.6) although only 5% of cells express *CYTH2* in the mid-trimester human fetal kidney (Cao *et al.* 2020). Inhibition of cytohesins *in vitro* has

been shown to have an antiproliferative effect in both lung and colorectal cancer cell lines (Bill *et al.* 2012, Pan *et al.* 2014) supporting a possible role for *CYTH2* in the regulation of cell proliferation and development.
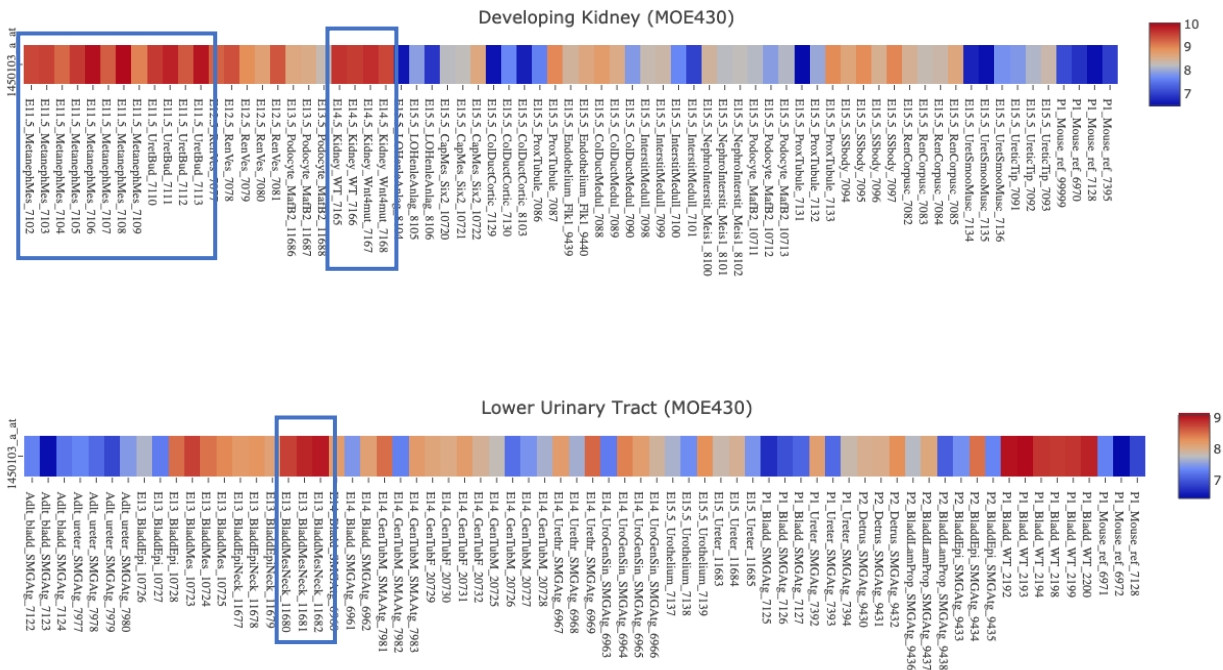
Figure 3.6. Heatmap demonstrating *Cyth2* expression in the embryonic mouse urinary tract.

High expression (dark red) is seen in the metanephric mesenchyme and ureteric bud in the developing kidney as well as in the bladder neck (blue rectangles). Microarray data from GUDMAP (Harding *et al.* 2011).

Enrichment was also observed in *ATF7IP2* (activating transcription factor 7 interacting protein 2)*,* which encodes a protein involved in regulating transcription and chromatin conformation (Ichimura *et al.* 2005). Again, this was not significant after correction for multiple testing (P=5.1x10^{-5}). Four cases (0.5%) were heterozygous for a LoF variant compared to 0.02% of controls. Oddly, three of the cases and two of the controls were heterozygous for the same stop gain variant

(rs148070233) which has a higher-than-expected MAF of 1.32×10^{-4} in gnomAD (version 3.1.2) considering it's predicted LoF effect and CADD score of 33. Further investigation revealed that rs148070233 is inherited in phase with an adjacent synonymous variant (rs141776960) and this multi-nucleotide variant rescues the transcript from nonsense-mediated decay (https://gnomad.broadinstitute.org/variant/16-10551356-CA-AT?dataset=gnomad_r2_1). This explains why such a deleterious variant has a relatively high population allele frequency and highlights the importance of positional context in variant interpretation.

3.4.2.2   Obstructive uropathy

To increase power to detect association I next repeated the rare variant analysis stratifying by phenotype to create a more homogenous cohort. In the obstructive uropathy group consisting of 177 individuals with PUJO, 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 controls (Figure 3.7 and Figure 3.8): *NCF2* (neutrophil cytosolic factor 2; 53 variants; *P*=2.55×10^{-6}).
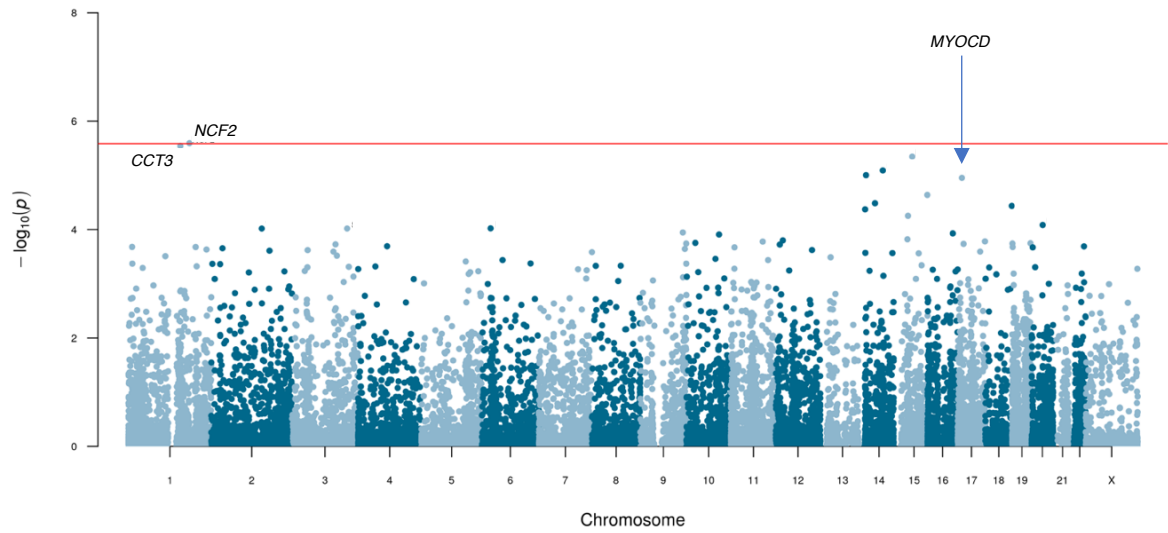
Figure 3.7. Manhattan plot of exome-wide gene-based rare, likely deleterious variant association testing for 177 individuals with obstructive uropathy and 24,451 ancestry-matched controls.

Each dot represents a gene. The red line indicates the Bonferroni-corrected exome-wide significance threshold of P=2.58×10^{-6}. Known CAKUT gene *MYOCD* is highlighted by the blue arrow.
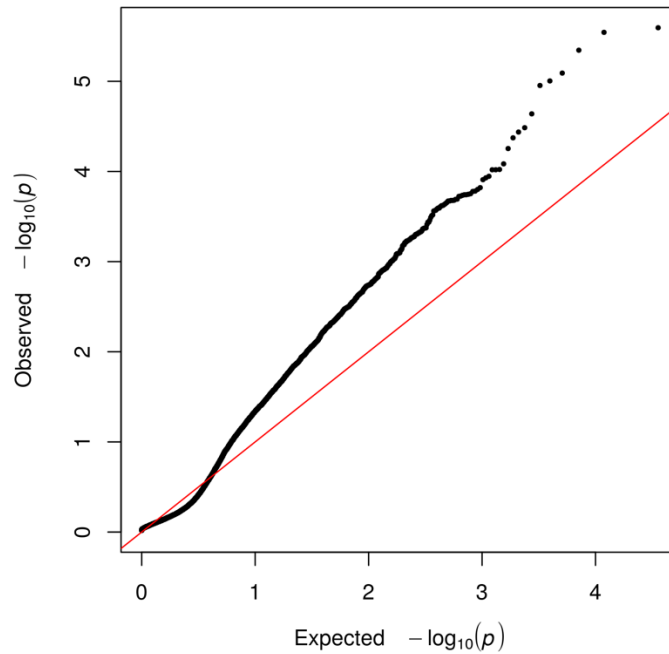
Figure 3.8. Q-Q plot of exome-wide gene-based association testing for 177 obstructive uropathy probands and 24,451 ancestry-matched controls.

SAIGE-GENE was performed for 19,406 genes with loss-of-function and likely deleterious missense variants with MAF < 0.1%. Each dot represents a gene. The red line signifies the observed versus the expected $-\log_{10}(P)$ for each gene tested.

5/177 (2.8%) of the obstructive uropathy 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 *NCF2*:c.1126C>T;p.Arg376Trp (rs777251055) which was absent from controls and has an allele frequency of 6.57×10^{-6} in gnomAD (version 3.1.2) with a CADD score of 22.4. Both individuals had kidney anomalies (renal hypoplasia and cystic renal dysplasia) in addition to hydronephrosis and hydroureter. Conditioning on the lead

variant (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.

NCF2 is a component of the NADPH oxidase complex found in neutrophils, biallelic variants of which cause chronic granulomatous disease (MIM 233710). *NCF2* has also been implicated in the development of certain cancers and studies have shown it regulates cell growth, malignant transformation and differentiation (Zhang *et al.* 2018, Qin *et al.* 2020). Furthermore, *Ncf2* is expressed in the murine bladder at E13 and ureter at E15 as well as in podocytes and endothelium at E13.5-15.5 suggesting a possible role in urinary tract development (Figure 3.9).
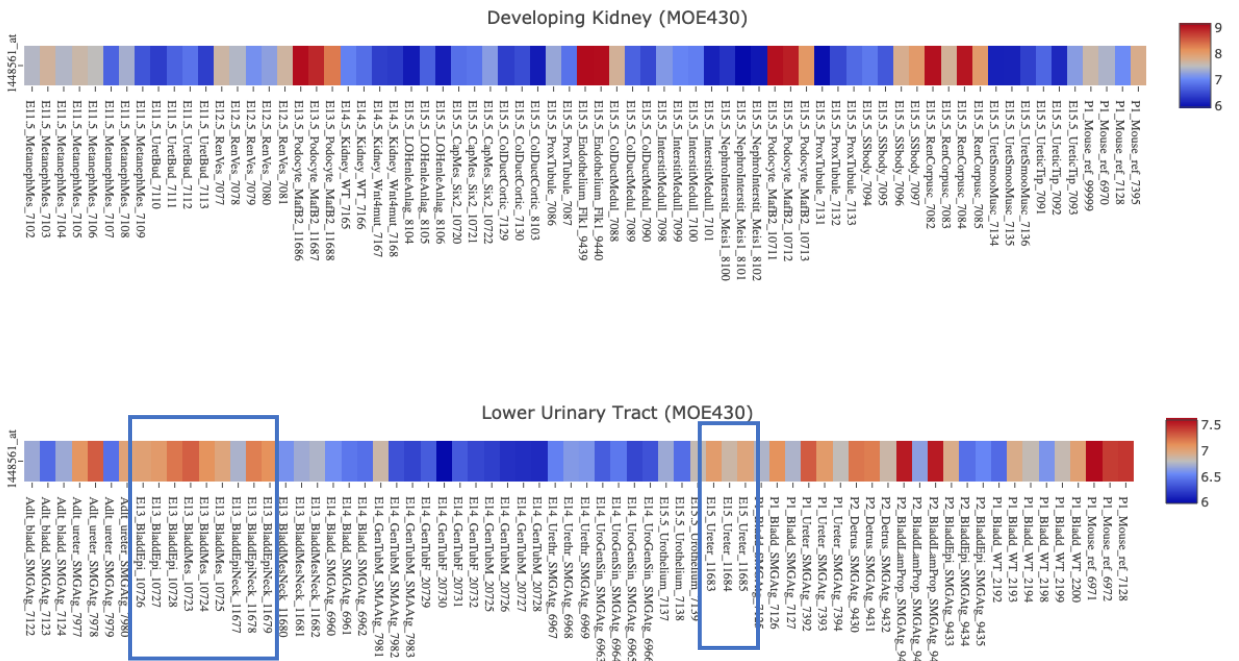
Figure 3.9. Heatmap demonstrating expression of *Ncf2* in the murine developing kidney and lower urinary tract.

Increased expression (red/orange) is seen in the bladder at E13 and ureter at E15 (blue rectangles). Microarray data from GUDMAP (Harding *et al.* 2011).

CCT3 (chaperonin containing T-complex polypeptide 1, subunit 3; 65 variants tested) also had suggestive evidence of association with a *P* value of 2.86x10^-6. 6/177 (3.4%) of cases had at least one missense variant compared with 103/24,451 (0.4%) controls. *CCT3* is a molecular chaperone protein and important subunit of the chaperonin-containing TCP-1 (TRiC) complex which folds numerous proteins, including actin and tubulin. It has been associated with several different cancers (Qian *et al.* 2016, Xu *et al.* 2020) and is widely expressed in both the developing mouse kidney and lower urinary tract (Figure 3.10).
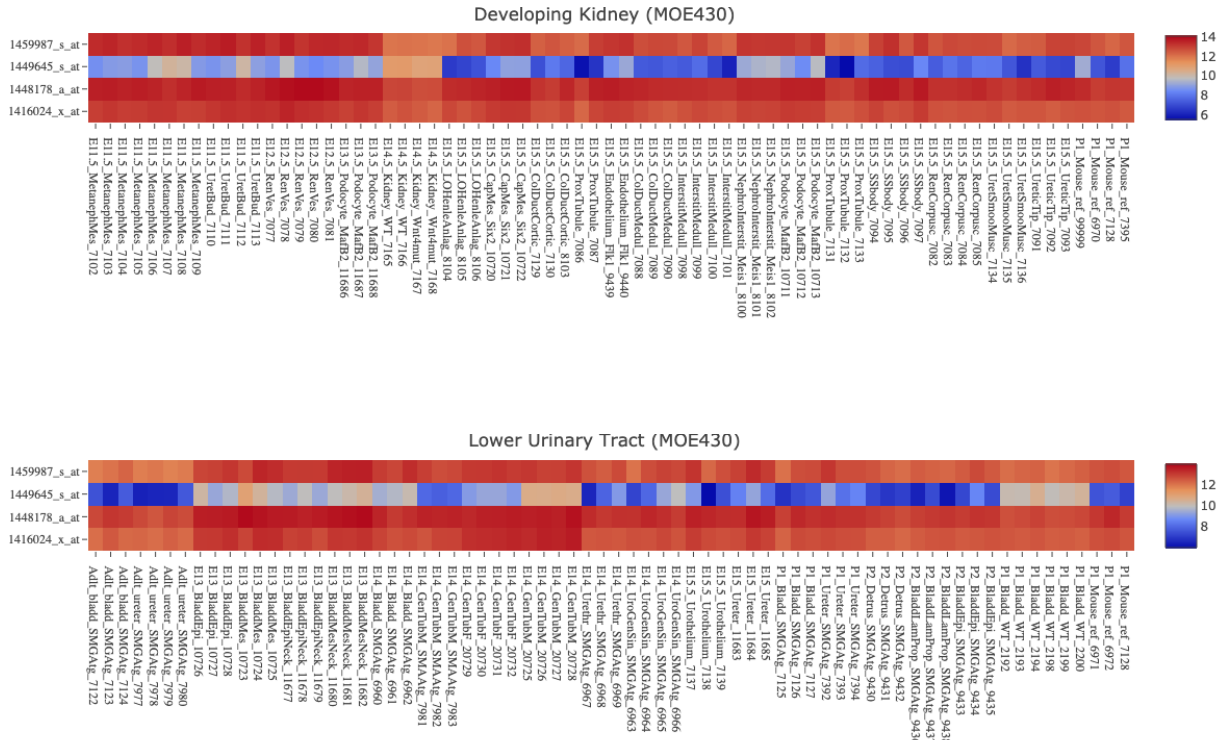
Figure 3.10. Heatmap demonstrating expression of *Cct3* in the murine developing kidney and lower urinary tract.

High expression denoted in red. Microarray data from GUDMAP (Harding *et al.* 2011).

MYOCD (myocardin)*,* a monogenic cause of autosomal dominant congenital megabladder (MIM 618719), was also enriched for rare coding variation in the obstructive uropathy cohort but did not achieve exome-wide significance (90 variants; *P*=1.11×10^{-5}). None of the other known monogenic causes of functional or anatomical lower urinary tract obstruction were found to be enriched for rare coding variation in this cohort: *BNC2, P=0.38; HPSE2, P=0.59; LRIG2, P=0.47; CHRM3, P=0.71; TSHZ3, P=0.26.*

3.4.2.3 Other CAKUT phenotypes

Rare variant association testing of the remaining subgroups did not result in any statistically significant associations for either a) missense and LoF or b) LoF only variant sets (Figure 3.11). Interestingly, although *HNF1B* had suggestive evidence of enrichment in the CAKUT analysis (*P*=6.06×10^{-6}), it was not significant in individuals with cystic renal dysplasia (*P*=0.71).

To assess the power of each subgroup analysis to detect association I calculated the proportion of affected individuals with causal variants in a single gene that could be identified with the sample size in each group (Table 3.3). This demonstrates that significantly larger sample sizes would be needed to identify causal genes that only account for a small proportion of the phenotype.
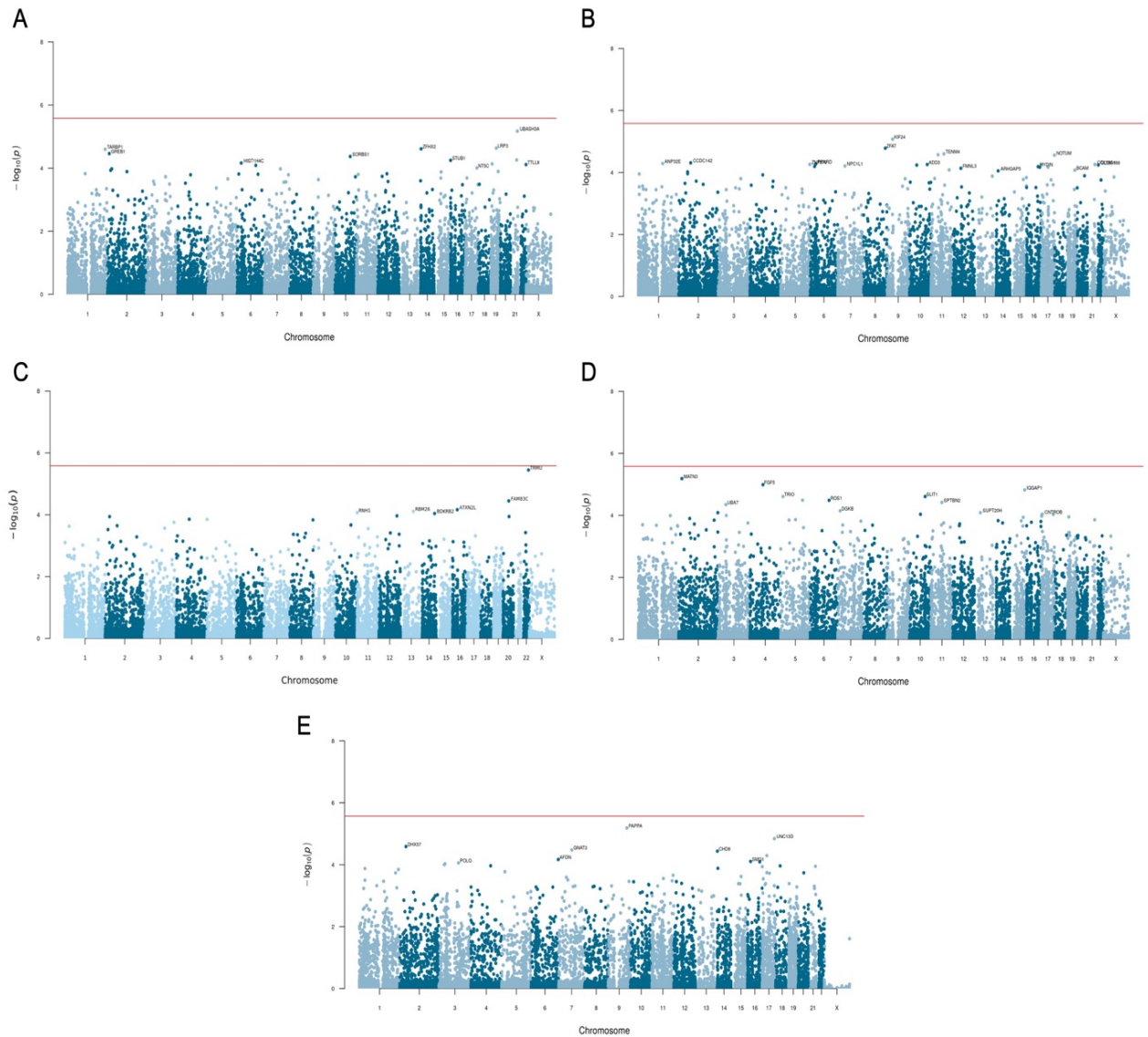
Figure 3.11. Manhattan plots of exome-wide gene-based rare, likely deleterious variant association testing for each CAKUT phenotype.

A, kidney anomalies (n=237); **B**, VUR (n=174); **C**, PUV (n=132); **D**, cystic dysplasia (n=112); **E,** bladder exstrophy (n=97). Each dot represents a gene. The red line indicates the exome-wide significance threshold of P=2.5×10^{-6}. No significant associations were detected.

Table 3.3. The minimum phenotypic variance explained by a single gene detectable in each cohort.

Power calculations were performed using PAGEANT based on a burden test model assuming 80% of qualifying variants are causal, a power of 80%, and an exome-wide significance level of 2.5×10^{-6} (Derkach *et al.* 2018).

Phenotype	Cases	Controls	Proportion of cases explained by LoF variation in a single gene
All CAKUT	813	25,205	5%
Kidney anomaly	237	22,733	16%
Obstructive uropathy	177	24,451	22%
VUR	174	22,562	22%
PUV	132	23,727	29%
Cystic dysplasia	112	24,084	34%
Bladder exstrophy	97	22,037	39%

3.5   Summary

- The contribution of known monogenic disease to CAKUT is much less than previously estimated with < 5% receiving a molecular diagnosis.

- A higher diagnostic yield was observed in those with cystic renal dysplasia or kidney anomalies (renal agenesis, hypoplasia, or dysplasia).

- The presence of affected first-degree relatives or extra-renal manifestations significantly increased the likelihood of a molecular diagnosis.

- Pathogenic *HNF1B* copy number and single nucleotide variation was the most frequently identified genetic diagnosis, although this still accounted for less than 1% of cases overall.

- There was no exome-wide significant gene-based enrichment of rare variation in the CAKUT cohort, however a non-significant increased burden of likely deleterious variation was observed in *HNF1B*.

- Stratification of the cohort by phenotype revealed statistically significant enrichment of rare variation in *NCF2* in individuals with obstructive uropathy including congenital megaureter, PUJO and VUJO.

3.6 Discussion

3.6.1 Prevalence of known monogenic disease

WGS offers a single test that captures nearly all genomic variation in an unbiased manner enabling simultaneous analysis of SNVs and structural variants as well as systematic re-analysis of variants and genes (Costain *et al.* 2018). Several studies have reported a superior diagnostic yield in rare disease using WGS compared with chromosomal microarray, targeted gene panels and WES due to the detection of deep intronic variants and small CNVs, as well as variants affecting non-coding RNAs, mitochondrial DNA and exonic regions with poor coverage on WES (Gilissen *et al.* 2014, Taylor *et al.* 2015, Stavropoulos *et al.* 2016, Lionel *et al.* 2018, Turro *et al.* 2020, 100,000 Genomes Project Pilot Investigators *et al.* 2021).

In this study, however, a monogenic cause of CAKUT was identified in less than 5% of individuals using WGS. This is significantly lower than previous ES studies that report diagnostic yields of up to 27% (Table 1.4) and is likely related to the relatively unselected nature of this cohort which includes a broad definition of CAKUT phenotypes as well as some individuals with relatively mild disease. In addition, it is possible that individuals with a clear monogenic cause for their disease were not entered into the 100KGP.

The largest previously examined CAKUT cohort consisted of 749 individuals from 650 families, the majority of whom had isolated VUR, and a similar proportion of familial disease to this study (Hwang *et al.* 2014, Kohl *et al.* 2014). Targeted ES of just 17 known AD and 12 candidate AR genes identified probable disease-causing

mutations in 6.3% and 2.5% of unrelated families, respectively, although it should be noted that in some of these cases the variant interpretation performed would not pass the stringent ACMG criteria for pathogenicity that were subsequently implemented in 2015 (Richards *et al.* 2015). Other studies report higher diagnostic yields that vary depending on the characteristics of the cohort examined: 14% in a pre-screened paediatric cohort with a high rate of consanguinity (van der Ven, Connaughton, *et al.* 2018); 18% in severe CAKUT, including a high proportion of affected fetuses (Heidet *et al.* 2017); and 24% when congenital and cystic disease are combined (Groopman *et al.* 2019).

In a cohort similar to this study, Nicolaou *et al.* used an ES panel of 208 known and candidate CAKUT genes to investigate 453 unrelated, mostly European individuals with sporadic CAKUT (predominantly collecting system and lower urinary tract defects, including PUV) reporting a diagnostic rate of 1.3% (Nicolaou *et al.* 2016). Given the lower rates of familial disease (10% vs 13%) and extra-renal manifestations (14% vs 29%) as well as the phenotypic make up (fewer individuals with kidney anomalies) of the cohort compared with this study, this lower diagnostic yield is not unexpected.

Individuals with cystic renal dysplasia and kidney anomalies were found to have the highest diagnostic yield of 10.7% and 5.9%, respectively, suggesting that these phenotypes should be prioritized for genetic testing. These estimates are consistent with the lower ranges seen in previous studies which report rates between 3.5-19.9% (Weber *et al.* 2006, Heidet *et al.* 2010, Thomas *et al.* 2011, Sanna-Cherchi *et al.* 2017, Ahn *et al.* 2020). At the other end of the spectrum, the absence of any molecular diagnoses in individuals with lower urinary tract phenotypes such as PUV and bladder exstrophy clearly suggest that these are phenotypically and genetically distinct disorders that should perhaps be considered separate entities from the other CAKUT phenotypes.

Diagnostic yield was significantly higher in paediatric patients. Although age at genetic testing should not theoretically impact the diagnostic yield for a congenital phenotype, it is probable that those who present to nephrology services as a child are more likely to have severe or syndromic disease and therefore a higher diagnostic yield. Both family history and extra-renal manifestations were also significant predictors of a positive genetic diagnosis, consistent with previous data from individuals with suspected monogenic kidney disease (Connaughton *et al.* 2019, Groopman *et al.* 2019, Mann, Braun, *et al.* 2019, Jayasinghe *et al.* 2021). Consanguinity was not significant however, despite being reported previously for both paediatric kidney transplant recipients and CAKUT patients (van der Ven, Connaughton, *et al.* 2018, Mann, Braun, *et al.* 2019), most likely because of the low rates of reported consanguinity and less severe phenotypes seen in this cohort.

This diagnostic yield analysis has several strengths. The 992 probands from 1,003 families examined here make up the largest CAKUT cohort investigated to date and are the first to be assessed using WGS. Candidate variants were comprehensively reviewed by a multi-disciplinary team including the patient's clinician, clinical scientists and clinical geneticists, and assigned pathogenicity using the rigorous ACMG criteria (Richards *et al.* 2015). Furthermore, this group is representative of routine clinical practice, in contrast to research cohorts enriched for consanguinity, severe disease, family history or extra-renal manifestations, meaning the results reported here are generalizable to general nephrology practice and can help inform genomic testing.

There are however several limitations. First, the molecular diagnoses are reliant on the automated bioinformatics pipeline developed by Genomics England which prioritizes variants based on an updateable, crowd-sourced and disease-focused gene panel (PanelApp) (Stark *et al.* 2021). This pipeline has changed over time with the introduction of CNV assessment and the variant prioritization algorithm Exomiser (Smedley *et al.* 2015) in 2019 which may have improved the diagnostic yield in more

recently analysed samples. Furthermore, this diagnostic framework has been shown to have limited sensitivity in the context of another genetically heterogenous rare disease, craniosynostosis, where reliance on a panel-based approach was found to be too restrictive (Hyder *et al.* 2021). Second, the analysis is hugely dependent on accurate data input and phenotyping with regards to HPO terms, clinical features and documentation of family history which will have been carried out by the recruiting team in consultation with the patient's clinician. Finally, the recruitment criteria state that pre-screening for *HNF1B* and *SALL1* should be carried out if clinically indicated but there was no method for capturing how many individuals underwent this screening and received a positive diagnosis, precluding recruitment to the project.

3.6.2 WGS as a diagnostic tool in CAKUT

The success of the 100,000 Genomes Project in the UK has led to the deployment of WGS in routine clinical practice for a wide range of rare diseases and cancer as part of the NHS Genomic Medicine Service. The 100,000 Genomes rare disease pilot study analysed WGS data from 2,183 families with undiagnosed rare disease and reported an overall diagnostic yield of 25%, with a quarter of diagnoses having immediate implications for clinical decision making (100,000 Genomes Project Pilot Investigators *et al.* 2021). This initial analysis included 43 CAKUT patients in whom 7% received a molecular diagnosis, a slightly higher number than observed in this study.

Current eligibility for WGS in renal disorders is limited to cystic kidney disease (including ciliopathies) and unexplained paediatric onset end-stage kidney disease. The phenotypic and genotypic variability associated with CAKUT mean it can be challenging to identify likely causative genes with a high pre-test probability which lends support to an agnostic WGS approach. However, current recommendations for clinical testing are restricted to microarray-based analysis for individuals with clinically significant non-syndromic CAKUT and a family history. This targeted testing

approach is supported by some of the findings in this study which demonstrate that the overall prevalence of monogenic disease in this group is low, that the 17q12 recurrent deletion is the most frequently identified diagnosis, and that individuals with cystic renal dysplasia have the highest diagnostic yield.

However, the prevalence of *HNF1B*-related disease in this cohort and others (Heidet *et al.* 2010, Thomas *et al.* 2011, Hwang *et al.* 2014), the fact that only half of cases are attributable to whole *HNF1B* deletions (and therefore detectable by microarray) and ~50% of variants are *de novo* (Clissold *et al.* 2015) suggest that testing for both CNVs and SNV/indels in *HNF1B,* even in the absence of a family history, may have clinical benefit*.* A positive genetic diagnosis has implications for cascade screening of family members, can inform reproductive counselling, as well as guide surveillance for extra-renal manifestations such as early-onset diabetes. Furthermore, identifying a specific molecular cause for a patient's condition is important for accessing support networks, recruitment to clinical trials, as well as potentially enabling kidney donation from an unaffected relative. On this basis, targeted assessment of *HNF1B* should be considered in any individual with kidney anomalies (renal hypodysplasia, renal agenesis, cystic renal dysplasia), even without a family history, especially if associated abnormalities, such as diabetes, hypomagnesaemia and hyperuricaemia are present (Adalat *et al.* 2009, 2019).

3.6.3 Monogenic causes of CAKUT are rare

Unbiased interrogation of rare variation across the exome did not identify any significant gene-based enrichment in this cohort of CAKUT patients. The absence of observed enrichment may be the accumulation of several factors which limit power for discovery; a) the cohort is phenotypically heterogenous, b) CAKUT 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. Although four genes (*AUTS2*, *ARHGAP5*,

HNF1B and *ZNF879*) demonstrated suggestive enrichment of likely deleterious variation, future meta-analyses will be needed to robustly distinguish the signals observed at these three novel candidate genes from noise. Similarly, identification of a single exome-wide significant gene (*NCF2*) in individuals with obstructive uropathy requires replication in an independent cohort before association can be clearly established. Interestingly, enrichment of the known AD CAKUT genes *HNF1B and MYOCD* was seen in the CAKUT and obstructive uropathy cohorts respectively, acting as 'positive controls' and providing validation of the testing approach.

Previous studies carried out in smaller cohorts of CAKUT patients have attempted rare variant burden testing with limited success, likely due to limited power. Nicolaou *et al.* performed burden testing in 434 unrelated Dutch individuals with predominantly lower urinary tract phenotypes (duplex collecting system, PUJO, PUV and VUR) and 498 unrelated Dutch controls from the Genome of the Netherlands project (GoNL). Using several burden tests (including SKAT) and incorporating rare (MAF < 1%), high and moderate impact coding variants from 208 known and candidate CAKUT genes they found no statistically significant enrichment of rare variation (Nicolaou *et al.* 2016). However, the different sequencing approaches used in cases (SOLiD NGS platform) and controls (Illumina HiSeq 2000), along with a significant difference in average coverage (~130X vs ~14X) may have introduced confounding. Heidet *et al.* employed SKAT-O to compare 168 cases with severe kidney anomalies (pre-screened for *HNF1B*, *PAX2*, *EYA1*, *ANOS1*, *GATA3*, *CHD7*, *PBX1* and *KIF14*) with 426 unrelated controls, all sequenced on the same platform (Heidet *et al.* 2017). Despite interrogating 330 known and candidate CAKUT genes, no gene was significantly enriched for rare, likely deleterious variation.

Lastly, Sanna-Cherchi *et al.* performed exome-wide burden testing (using a Fisher's exact test) in 195 European individuals with unresolved renal hypodysplasia and 6,905 mixed-ethnicity controls, controlling for differences in sequence coverage and

ancestry (Sanna-Cherchi *et al.* 2017). Of note, almost a third of cases (31.2%) had a family history of renal disease, more than double that reported in this study. An enrichment of rare deleterious variants in *GREB1L* was identified under a dominant inheritance model although this did not reach exome-wide significance. Functional studies subsequently confirmed a role for *GREB1L* in renal morphogenesis demonstrating strong mRNA expression in human fetal kidney and bilateral renal agenesis and genital tract anomalies in CRISPR/Cas9 generated *Greb1l* knock-out mice (De Tomasi *et al.* 2017). The identification of an (almost significant) signal in this context, despite the small sample size, was most likely due to the selective nature of the cohort with a high proportion of familial disease and extra-renal features.

Several monogenic causes of both anatomical and functional lower urinary tract obstruction (LUTO) have been identified (see Table 1.3) with monoallelic LoF variants in *MYOCD* associated with congenital megabladder in humans and mice (Houweling *et al.* 2019). Interestingly, in this analysis, a non-significant enrichment of rare, likely deleterious variation (P=1.11x10^{-5}) but not LoF variation (P=1) was seen in *MYOCD*. This suggests that while LoF variation may rarely cause congenital megabladder, missense variants in the same gene may perhaps result in less severe obstructive uropathy phenotypes.

3.6.4 Functional human datasets

The downstream interpretation of candidate genes and variants to try and discern the mechanisms underlying these associations is challenging, particularly in the context of developmental malformations. In recent years, several large multi-omics datasets have been generated and provide a reasonable starting point to investigate new gene-phenotype associations. For example, the Genotype-Tissue Expression (GTEx) project (GTEx Consortium 2020) links variants with healthy tissue-specific gene expression data, although only in tissues from adult donors. Work is now also

underway to generate a Developmental GTEx Project dataset of healthy neonatal, paediatric, and adolescent tissues.

Recently, a human single-cell atlas of fetal gene expression (Cao *et al.* 2020) and chromatin accessibility (Domcke *et al.* 2020) has been published (DESCARTES, Developmental Single Cell Atlas of gene Regulation and Expression; descartes.brotmanbaty.org). This dataset represents a huge leap forward in our understanding of how genes are regulated and expressed in early human development and enables the interrogation of genes and variants that have been associated with developmental conditions. One of the main limitations of this dataset is that all the data generated is from 'mid-gestation' fetal tissue i.e., after the main period of organogenesis in the first trimester. Furthermore, while it includes fetal kidney tissue, the rest of the urinary tract is absent. This limits the conclusions that can be drawn from these datasets and future studies of embryonic urinary tract tissue at the relevant developmental time point will be needed to definitively investigate the regulation and expression of the candidate genes identified.

3.6.5 Strengths and limitations

The strengths of this rare variant association analysis lie in the large sample size and fact that both cases and controls have been sequenced on the same platform. This minimizes the risk of confounding due to sequencing artefacts and means both cohorts are subject to the same quality control metrics with regards to variant calling and coverage. This is also the first time WGS data has been used to perform an exome-wide association test in individuals with CAKUT, providing more uniform coverage of coding regions with its PCR-free approach than WES (Turro *et al.* 2020). In addition, the use of a generalized logistic mixed model and case-control ancestry matching strategy minimized potential confounding by population structure.

The main limitation of this rare variant study is one of power. With this cohort size, I was powered (> 80%) to detect exome-wide significant enrichment of LoF variation in a single gene explaining at least 5% of CAKUT cases. Despite utilizing different strategies to boost power including stratification of the cohort into specific phenotypes, selecting for LoF qualifying variants and using SKAT-O to optimize testing when the genetic architecture is unknown, I only detected a single gene that reached exome-wide significance (*NCF2* in the obstructive uropathy cohort). In addition, SKAT-O tests for association under an additive model (where the risk of disease is presumed to be proportional to the number of risk alleles) which is usually more powerful than a dominant model but may lead to a loss of power in cases of recessive inheritance or in the presence of epistasis. Although future collaborations and rare variant meta-analysis with other cohorts will be useful to try and identify novel genes which affect a smaller proportion of cases, overall, these findings argue against there being a strong monogenic component to CAKUT.

It is also important to note that the association testing performed here focuses purely on rare coding SNVs/indels and does not incorporate additional sources of genomic variation such as structural or non-coding variation. As is evident from *HNF1B*-associated renal disease, both CNVs and SNVs can result in haploinsufficiency and therefore examining the combined burden of both types of variation exome-wide may reveal novel gene associations.

It is also now understood that the non-coding genome is enriched for regulatory elements that affect gene expression. Given ~90% of common variant disease associations occur in the non-coding space it is not unreasonable to hypothesize that rare regulatory variants may also affect phenotypic variance. However, extending rare variant association tests to the non-coding genome has so far proved challenging for two main reasons (Bocher and Génin 2020). First, it is more difficult to predict the functional effects of non-coding compared with protein-coding variants making it challenging to determine which variants should be included in the

'qualifying' set. Second, while genes form clear testing regions, it is not evident how to effectively collapse non-coding variants together. A recent study investigating neurodegenerative disease performed gene-based burden testing using both rare coding and associated non-coding variants (based on cell-specific experimentally predicted regulatory regions) identifying enrichment in *TET2* which would have been missed if only coding variants had been included (Cochran *et al.* 2020). Additional approaches using putative enhancers (Shaffer *et al.* 2019) or even whole topologically associated domains (TADs) (Lumley *et al.* 2018) as functional units have been proposed, however without a clearly defined analysis strategy and a lack of bioinformatic tools for accurate functional prediction, rare non-coding variant association studies remain challenging.

3.7   Conclusion

In this chapter I have presented evidence to suggest that CAKUT, at least in sporadic cases, is not usually a monogenic disease. First, a single-gene cause was identified in only a minority of patients confirming that the prevalence of known monogenic disease is low. Importantly, this is the largest reported cohort of individuals with CAKUT investigated so far and as a group are representative of clinical practice, in contrast to previous studies which have enriched for familial, consanguineous, or severe disease. Second, a lack of statistically significant gene-based enrichment of rare variation suggests that the contribution of previously implicated genes to CAKUT risk has been overestimated. Publication bias of highly penetrant monogenic CAKUT associations has led to an overrepresentation of single-gene causes in the literature however the observed locus and allelic heterogeneity, incomplete penetrance and variable expressivity suggest that the genetic architecture of CAKUT is far more complex. How other types of genomic variation contribute to this complexity is what I aim to investigate in the remainder of this thesis.

Chapter 4: Structural variation

4.1 Introduction

Structural variation (SV) describes several diverse types of genomic variation ≥ 50bp in size that may be balanced or unbalanced in nature (Figure 4.1). Unbalanced SVs include CNVs (which may be deletions, duplications or multi-allelic), insertions and complex SVs usually larger than 1Mb. Inversions and reciprocal translocations result in no net loss or gain of genomic material and are therefore considered balanced variants. Importantly, this type of variation cannot be detected using conventional microarray-based approaches. SVs can result from different mutational mechanisms including DNA recombination-, replication- and repair-associated processes. Recurrent SVs are generated during meiosis by nonallelic homologous recombination (NAHR), essentially a misalignment between two sequences of DNA called low copy repeats (LCRs) which exhibit high (> 95%) sequence similarity. Recombination 'hotspots' in these LCRs result in SVs that recur in multiple individuals (Carvalho and Lupski 2016).
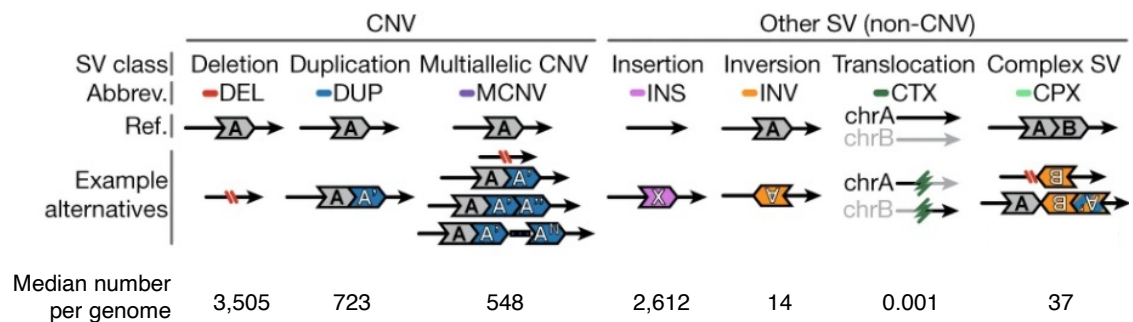
Figure 4.1. Types of structural variation.

Median number of SVs per genome based on short-read WGS detection. CNV, copy number variant; SV, structural variant. Adapted from Collins *et al.* 2020.

Although SVs account for ~0.1% of all variants, their size means they contribute greater diversity between two human genomes at the nucleotide level than any other form of variation (Sudmant *et al.* 2015). A typical human genome is estimated to have 4,500-7,500 SVs (detectable using short-read WGS), most of which are small (median size 331bp) and rare (MAF < 1%), and the majority of which are deletions (~35%), mobile-element insertions (~27%) or tandem duplications (~11%) (Abel *et al.* 2020, Collins *et al.* 2020). The superior sensitivity of long-read WGS has however recently identified over 25,000 SVs per genome, with the increase primarily made up of small (< 2kb), repetitive insertions and large (> 50kb) inversions (Audano *et al.* 2019, Chaisson *et al.* 2019).

In general, SVs tend to have more severe consequences than SNVs, mediating their effects via changes in gene dosage, disruption of gene function, or rearrangement of regulatory elements to alter genomic context (Lappalainen *et al.* 2019). Widespread selection is seen against almost all classes of SV that overlap genes, with modest selection also observed against those affecting non-coding *cis*-regulatory elements (Abel *et al.* 2020, Collins *et al.* 2020). Interestingly, copy-gain duplications do not show evidence of negative selection, consistent with the known evolutionary role of gene duplication events which are the primary force by which new gene functions arise (Dennis and Eichler 2016).

Given this selective pressure, it is perhaps not surprising that rare CNVs have been implicated in the pathogenesis of a wide range of diseases. Charcot Marie Tooth disease type 1A was the first Mendelian disease in which a gene dosage effect was described, where an inherited 1.5Mb tandem duplication encompassing the *PMP22* gene was identified (Lupski *et al.* 1991). Subsequently, rare and *de novo* CNVs have been associated with common complex traits such as autism (Sebat *et al.* 2007) and schizophrenia (Stefansson *et al.* 2008, Marshall *et al.* 2017) as well as developmental delay (Cooper *et al.* 2011) and congenital malformations (Mefford *et al.* 2007, Greenway *et al.* 2009). Somatic SVs are also strongly associated with the development and progression of many cancers, where complex SVs result from processes such as

chromoplexy and the catastrophic chromothripsis or 'chromosome shattering' (Li *et al.* 2020).

SVs are commonly found in repeat-rich segmentally duplicated regions which can be problematic for detection and accurate resolution of breakpoints. Traditionally, low-cost, high-throughput microarray-based approaches (e.g., array CGH or SNP platforms) have been used in clinical diagnostics to identify CNVs down to a resolution of ~25kb but these methods lack the ability to detect a) small SVs, b) balanced SVs such as inversions, and c) SVs not found in the reference genome e.g. novel insertions (Ho *et al.* 2020). The advent of NGS along with the development of computational approaches for SV discovery looking for discordance between sample reads and the reference genome (including read depth, read-pair, split-read and *de-novo* assembly algorithms), has enabled the determination of breakpoints down to single base pair resolution (Ho *et al.* 2020). While still primarily a research tool, the superior sensitivity and resolution of WGS will likely replace conventional arrays for clinical diagnostics in the future.

Previous microarray-based studies have demonstrated an increased burden of large (> 100kb), rare, exonic CNVs in individuals with CAKUT (Sanna-Cherchi *et al.* 2012, Verbitsky *et al.* 2019, 2021). Indeed, in our cohort, the recurrent 1.4Mb 17q12 deletion was the most frequently identified pathogenic variant. However, at present nothing is known about the contribution of smaller SVs to this disorder or whether other types of structural variation (e.g., inversions) may be important. Furthermore, given that tightly regulated transcriptional networks are critical for normal renal tract development it can be hypothesized that SVs might preferentially affect non-coding regulatory regions, thereby altering the genomic context or expression of key developmental genes. In this chapter I therefore take advantage of the superior SV detection capabilities of WGS to investigate the contribution of both rare gene-disrupting and non-coding regulatory structural variation to the pathogenesis of CAKUT.

4.2   Aims

1. To ascertain whether rare gene-disrupting structural variation is enriched in individuals with CAKUT and identify recurrently affected genes or loci using an unbiased collapsing exome-wide approach.

2. To determine whether non-coding *cis*-regulatory elements demonstrate an increased burden of rare structural variation.

4.3   Methods

4.3.1   Variant calling

The SV calling pipeline used by Genomics England incorporates CANVAS (version 1.3.1) to determine copy number (> 10kb), and MANTA (version 0.28.0) to identify SVs $\geq$ 50bp. Both tools are widely-used for SV calling in the bioinformatics community and MANTA has been shown to perform consistently well with regards to precision and recall compared to other SV callers (Cameron *et al.* 2019).

CANVAS scans the genome for regions which have an unexpected number of short read alignments assigning those with fewer or more than expected as losses or gains, respectively (Roller *et al.* 2016).

MANTA combines discordant read-pair and split-read evidence to identify SV regions (Chen, Schulz-Trieglaff, *et al.* 2016). While read-pair approaches are powerful, mapping reads to repetitive regions remains challenging and accurate breakpoint resolution relies on very tight fragment size distributions. Split-read algorithms (where a read has two or more partial alignments to the reference genome i.e. it is split by a breakpoint) can pinpoint exact breakpoints for a wide range of SV types but is only reliable in unique regions of the genome (Alkan *et al.* 2011). While MANTA can detect deletions and tandem duplications < 10kb, inversions, and interchromosomal translocations it cannot reliably identify dispersed duplications, small inversions (< 200bp), fully assembled large

142

insertions (> 2x150bp) or breakends where repeat lengths approach the read size (150 bp). Very few insertions were identified in this cohort using MANTA and in view of this they were excluded from downstream analysis. In addition, variants classified as translocations, single breakends or complex SVs which are more difficult to accurately resolve were filtered out.

4.3.2 Quality control

Variants were retained if they fulfilled the following quality filters:

- CNV length > 10kb and Q-score ≥ Q10 indicating ≥ 90% confidence there is a variant present (CANVAS; (Roller *et al.* 2016)).
- QUAL ≥ 20 indicating ≥ 99% confidence that there is a variant at the site, GQ ≥ 15 indicating ≥ 95% confidence that the genotype assigned to a sample is correct, and MaxMQ0Frac < 0.4 which indicates the proportion of uniquely mapped reads around either breakend (MANTA; (Chen, Schulz-Trieglaff, *et al.* 2016)).

Variants without paired read support, inconsistent ploidy, or depth > 3x the mean chromosome depth near one or both breakends were excluded.

4.3.3 Extraction of exon and *cis*-regulatory element intersecting SVs

For each sample, BEDTools (version 2.27.1) (Quinlan and Hall 2010) was used to extract SVs that intersected by a minimum of 1bp with:

a) at least one exon (GENCODE; version 29) (Frankish *et al.* 2019) or

b) an ENCODE (ENCODE Project Consortium *et al.* 2020) candidate *cis-regulatory* element (cCRE)

ENCODE (ENCODE Project Consortium *et al.* 2020) cCREs are 150-350bp consensus sites of chromatin accessibility (DNase hypersensitivity sites) supported by histone modification (H3K4me3 and/or H3K27ac), or CCCTC-binding factor (CTCF) ChIP-seq data (Figure 4.2). cCREs are defined based on

their epigenetic signatures in at least one biosample and proximity to the nearest annotated GENCODE transcription start site (TSS):
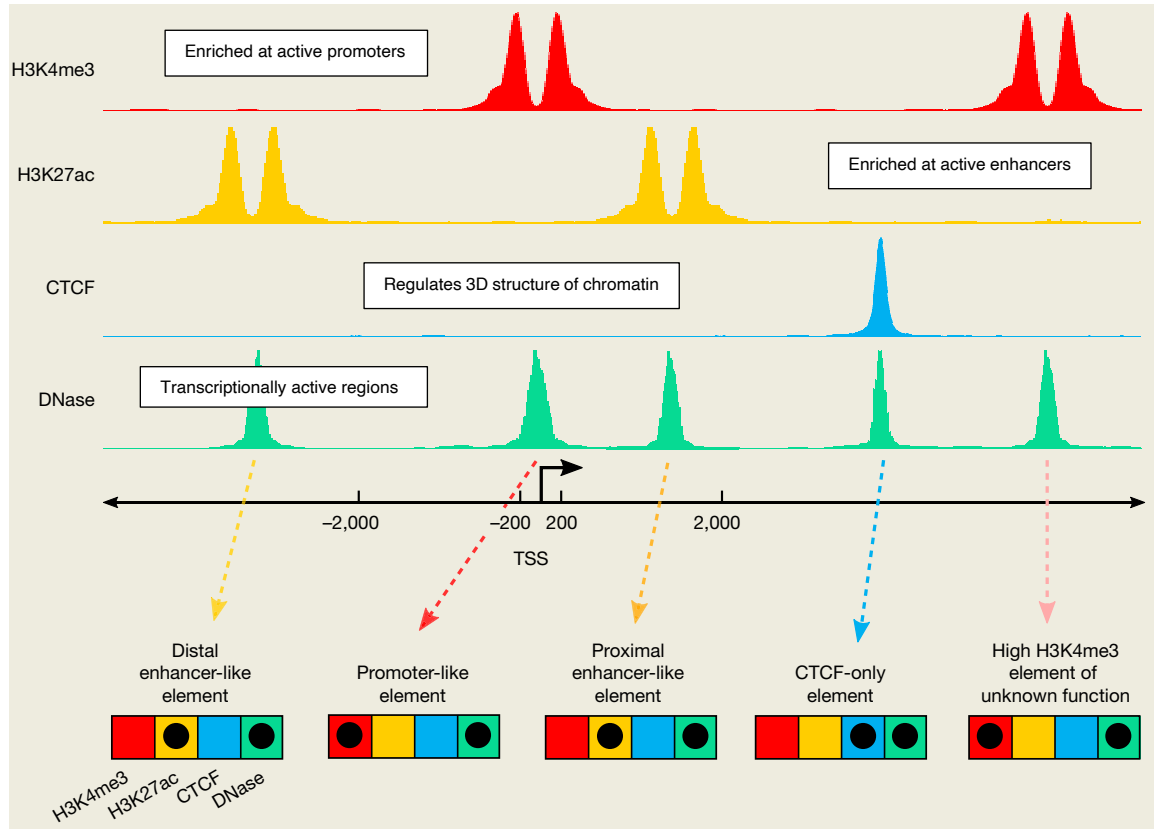
Figure 4.2. Classification of cCREs.

cCREs are assigned to one of five elements (rectangles with coloured blocks) based on their epigenetic signature in at least one biosample and proximity to the nearest transcription start site (TSS). Adapted from ENCODE Project Consortium *et al.* 2020.

A total of 926,535 cCREs encoded by 7.9% of the human genome were downloaded from UCSC Table Browser (Sugnet *et al.* 2002) using the encodeCcreCombined track (updated 20/05/2020). This includes:

- ~668,000 distal enhancer-like signature (dELS) elements > 2kb from TSS

- proximal enhancer-like signature (pELS) elements within

144

- ~57,000 CTCF-only elements which are candidates for insulator and chromatin looping functions

- ~35,000 promoter-like signature (PLS) elements within 200bp (centre to centre) of a TSS

- ~26,000 DNase-H3K4me3 elements with promoter-like signals but are > 200bp from a TSS

4.3.4 Filtering by allele frequency

Variants were first separated by type into CNV, deletion (DEL), duplication (DUP), and inversion (INV) sets before being filtered using BEDTools (version 2.27.1) (Quinlan and Hall 2010) to remove common SVs of the same type. SVs were removed if they had a minimum 70% reciprocal overlap with:

a) the dbVar (Lappalainen *et al.* 2013) NCBI curated dataset of SVs (nstd186) containing variant calls from studies with at least 100 samples and AF > 1% in at least one population, including gnomAD (Karczewski *et al.* 2020), 1000 Genomes (Phase 3) (1000 Genomes Project Consortium *et al.* 2015) and DECIPHER (Firth *et al.* 2009). This dataset does not include inversions.

and/or

 b) a dataset of common (AF > 0.1%) SVs generated from 12,234 cancer patients recruited to the 100KGP. SVs were merged using SURVIVOR (version 1.0.7) (Jeffares *et al.* 2017), allowing a maximum distance of 300bp between pairwise breakpoints and allele frequencies calculated using BCFtools (version 1.11) (Danecek *et al.* 2021). This dataset has the added benefit of being sequenced on the same platform as our case-control cohort.

After removal of overlapping common variants, a custom perl script (Dr Helen Griffin, Newcastle University) was used to calculate allele frequencies for each type of SV across the combined case-control cohort using bins of 10kb across the entire genome. SVs with an AF < 0.1% were retained for further analysis.

4.3.5 Burden analysis

Exome-wide gene-based and genome-wide cCRE-based burden testing was carried out using custom R scripts stratified by SV type: CNV, DEL, DUP and INV. SVs were aggregated across 19,005 autosomal protein-coding genes and five cCRE types (see Figure 4.2). The burden of rare (MAF < 0.1%), autosomal SVs in cases and controls was then enumerated under a dominant inheritance model by comparing the number of individuals with ≥1 SV using a two-sided Fisher's exact test. The Wilcoxon-Mann-Whitney test was used to compare median SV size. Unadjusted P values are reported. The Bonferroni correction for the number of genes (P=0.05/19,005=2.6x10^{-6}) and cCRE/SV combinations (P=0.05/20=2.5x10^{-3}) tested was applied, although with the knowledge that this is likely to be too stringent given the tests are not truly independent (one SV can affect multiple genes or cCREs). Given the underlying null distribution is unknown, empirical P values were also calculated using permutation testing, swapping case-control labels 10,000 times.

4.4 Results

4.4.1 Burden of gene-disrupting structural variation

I first focused on rare, autosomal SVs that intersected with at least one exon and had an AF < 0.1% on the basis that these variants were potentially gene-disrupting and therefore more likely to be deleterious. Analysis of 813 CAKUT cases and 25,205 ancestry-matched controls did not reveal any statistically significant enrichment of CNVs, small deletions and duplications, or inversions (Figure 4.3 and Table 4.).
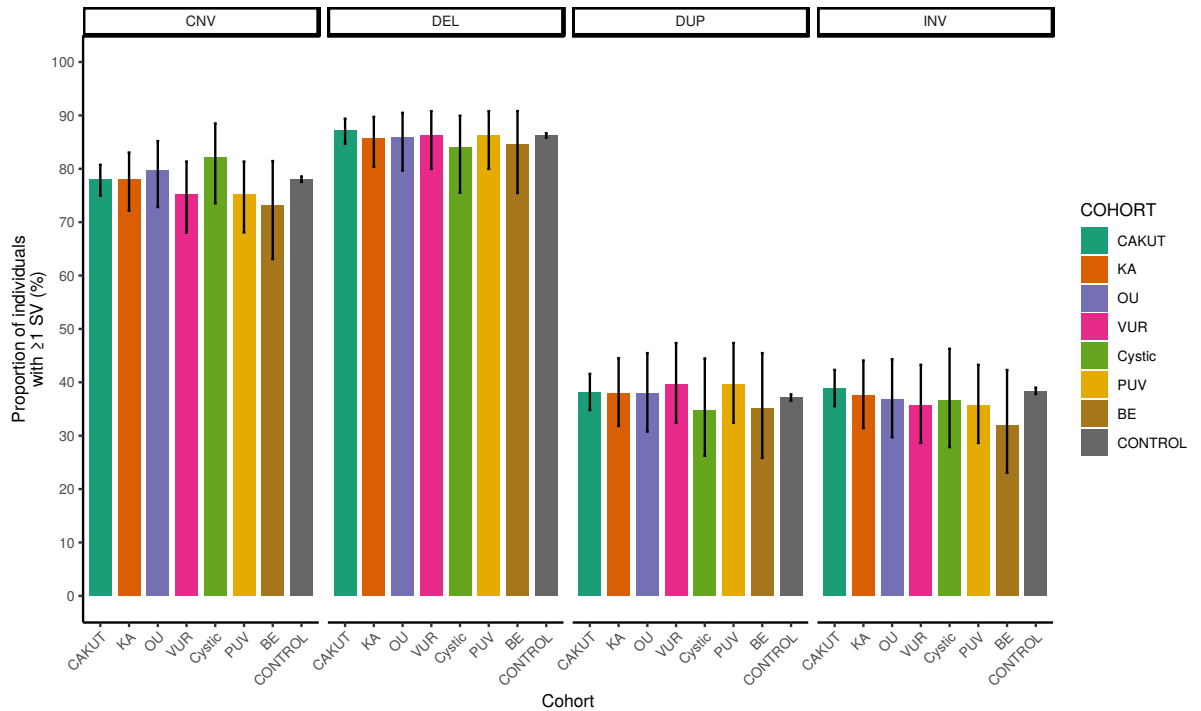
Figure 4.3. Burden of rare, autosomal, exonic structural variants in CAKUT probands stratified by phenotype and 25,205 controls.

Vertical black bars indicate 95% confidence intervals. CNV, copy number variant; DEL, deletion; DUP, duplication; INV, inversion; KA, kidney anomaly; OU, obstructive uropathy; VUR, vesico-ureteral reflux; PUV, posterior urethral valves; BE, bladder exstrophy.

Small deletions (< 10kb) were the most frequently identified SV, seen in over 85% of individuals. The median size of SVs was also similar in cases and controls (Table 4.1).

Table 4.1. Burden of rare, autosomal, exonic structural variants in CAKUT probands, stratified by phenotype.

Although each phenotype was ancestry-matched to a separate subset of controls, data for the largest control cohort analysed is presented here for reference. KA, kidney anomaly; OU, obstructive uropathy; VUR, vesico-ureteral reflux; PUV, posterior urethral valves; BE, bladder exstrophy; CNV, copy number variant; DEL, deletion; DUP, duplication; INV, inversion; OR, odds ratio; CI, 95% confidence interval; IQR, interquartile range.

147

		CAKUT (n=813)	KA (n=237)	OU (n=177)	VUR (n=174)	Cystic (n=112)	PUV (n=132)	BE (n=97)	Controls (n=25,205)
CNV	n (%)	634 (78.0)	185 (78.1)	141 (79.7)	131 (75.3)	92 (82.1)	109 (82.6)	71 (73.2)	19,673 (78.1)
	OR (CI)	1.00 (0.84-1.19)	1.07 (0.78-1.48)	1.17 (0.80-1.74)	0.92 (0.65-1.34)	1.39 (0.85-2.39)	1.52 (0.96-2.50)	0.84 (0.53-1.38)	
	Fisher's exact *P*	0.97	0.76	0.47	0.65	0.22	0.07	0.47	
	Median size (kb) (IQR)	87 (137)	89 (130)	83 (145)	92 (127)	92 (145)	104 (183)	93 (137)	84 (137)
	P (Wilcoxon)	0.51	0.31	0.95	0.16	0.41	0.75	0.38	
DEL	n (%)	709 (87.2)	203 (85.7)	152 (85.9)	150 (86.2)	94 (83.9)	117 (88.6)	82 (84.5)	21,738 (86.2)
	OR (CI)	1.09 (0.88-1.35)	1.06 (0.74-1.58)	1.05 (0.69-1.68)	1.12 (0.72-1.80)	0.94 (0.56-1.65)	1.46 (0.85-2.69)	1.00 (0.57-1.87)	
	Fisher's exact *P*	0.47	0.86	0.92	0.67	0.79	0.19	1	
	Median size (kb) (IQR)	1.5 (4.5)	1.5 (4.7)	1.9 (5.1)	1.2 (4.2)	1.6 (4.7)	1.4 (4.2)	2.6 (4.9)	1.6 (4.5)
	P (Wilcoxon)	0.7	0.99	0.16	0.22	0.76	0.18	0.41	
DUP	n (%)	310 (38.1)	90 (38.0)	67 (37.9)	69 (39.7)	39 (34.8)	59 (44.7)	34 (35.1)	9,361 (37.1)
	OR (CI)	1.04 (0.90-1.21)	1.11 (0.84-1.45)	1.07 (0.77-1.46)	1.20 (0.87-1.65)	0.96 (0.63-1.44)	1.45 (1.01-2.08)	1.02 (0.65-1.57)	
	Fisher's exact *P*	0.58	0.45	0.7	0.233	0.92	0.04	0.92	
	Median size (kb) (IQR)	3.0 (5.6)	3.6 (5.3)	3.6 (5.8)	1.9 (5.1)	3.7 (6.0)	3.7 (5.7)	3.5 (5.1)	2.9 (5.4)
	P (Wilcoxon)	0.41	0.42	0.47	0.18	0.41	0.16	0.99	
INV	n (%)	316 (38.9)	89 (37.6)	65 (36.7)	62 (35.6)	41 (36.6)	66 (50.0)	31 (32.0)	9,679 (38.4)
	OR (CI)	1.02 (0.88-1.18)	1.04 (0.79-1.36)	0.96 (0.70-1.32)	0.96 (0.69-1.32)	0.99 (0.66-1.48)	1.72 (1.20-2.45)	0.84 (0.53-1.31)	
	Fisher's exact *P*	0.8	0.79	0.88	0.87	1	2.1×10^{-3}	0.46	
	Median size (kb) (IQR)	194 (1254)	211 (910)	180 (791)	232 (1346)	194 (1218)	253 (1931)	176 (1346)	175 (1292)
	P (Wilcoxon)	0.15	0.78	0.96	0.70	0.72	0.44	0.67	

Table 4.1. Burden of rare, autosomal, exonic structural variants in CAKUT probands, stratified by phenotype.

Larger CNVs are more likely to be disease causing, however analyzing CNVs by size did not reveal any significant differences between the CAKUT cohort and control group; although slightly more cases had very large ($\geq$ 1Mb) CNVs than controls (3.9% vs 3.3%; P=0.32; OR 1.20; 95% CI 0.81-1.72) (Figure 4.4). The ratio of duplications to deletions was also the same in cases and controls (0.89).

Stratification by phenotype revealed a greater number of individuals with PUV had $\geq$ 1 exon-intersecting SV compared with controls (Table 4.1). In this cohort of 132 cases and 23,727 controls, an increased burden of rare inversions was seen (median size 253kb; P=2.1x10^{-3}; OR 1.72; 95% CI 1.20-2.45), although this was not statistically significant after applying the Bonferroni correction for multiple testing (P=0.05/28=1.8x10^{-3}). None of the other phenotypes had evidence for SV enrichment when compared to ancestry-matched controls.
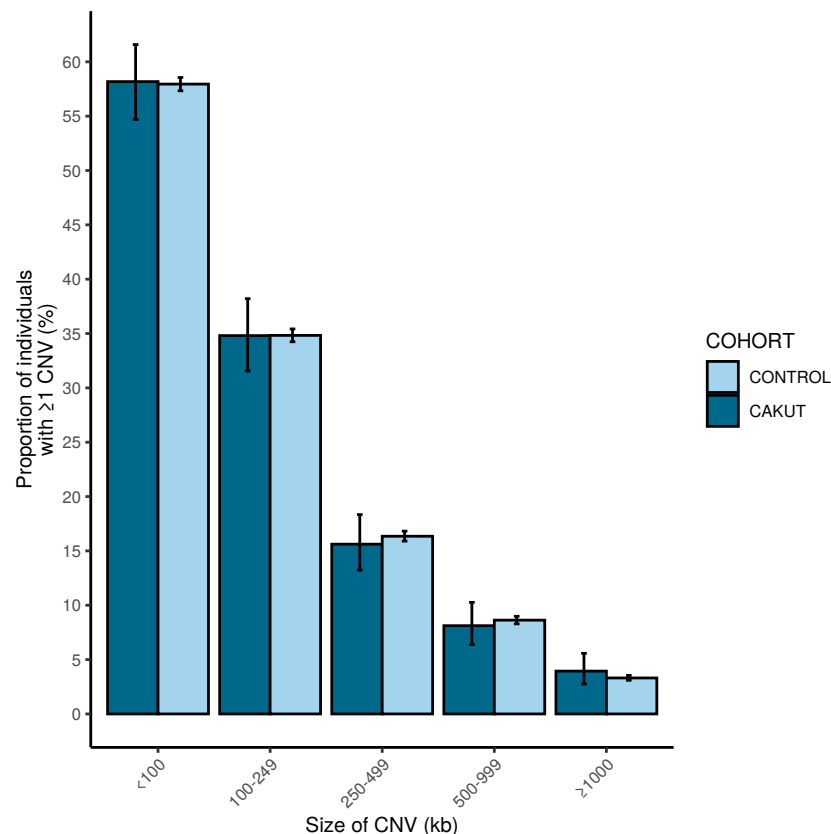
Figure 4.4. Burden of rare, exonic CNVs in 813 CAKUT probands and 25,205 controls stratified by size.

Vertical black bars indicate 95% confidence intervals. CNV, copy number variant.

149

Large, rare, exonic CNV losses have previously been found to be enriched in individuals with kidney anomalies (Sanna-Cherchi *et al.* 2012, Verbitsky *et al.* 2019). As a direct comparison I therefore repeated the analysis in 260 individuals with renal agenesis, hypodysplasia and cystic dysplasia and 25,075 controls using a similar size threshold of $\geq$ 100kb. Although there was evidence of an increased burden of exon intersecting CNV losses this was not statistically significant (26.9% vs 22.0%; *P*=0.06; OR 1.31; 95% CI 0.98-1.73). Interestingly, Verbitsky *et al.* (2019) identified a comparable number of CNV losses in their kidney anomaly cohort consisting of 1,088 individuals (22.5%), however substantially fewer were detected in their 21,498 controls (14.7%).

4.4.2   Exome-wide gene-based burden testing

To ascertain whether any genes were recurrently affected by structural variation I next aggregated the different types of SV across 19,005 protein coding genes. Although there was no significant gene-level enrichment of autosomal rare SVs after correction for multiple testing (*P*=0.05/19,005=2.63×10^{-6}), CNVs affecting *HNF1B* (and the surrounding 17q12 region) demonstrated the strongest association, reaching a minimum *P*-value of 3.1×10^{-4} (Figure 4.5). A greater proportion of individuals with CAKUT had copy number losses affecting the 17q12 region than controls (0.5% vs 0.02%; *P*=5.99×10^{-5}), but no significant difference was seen in copy number gains (0.12% vs 0.04%; *P*=0.32).
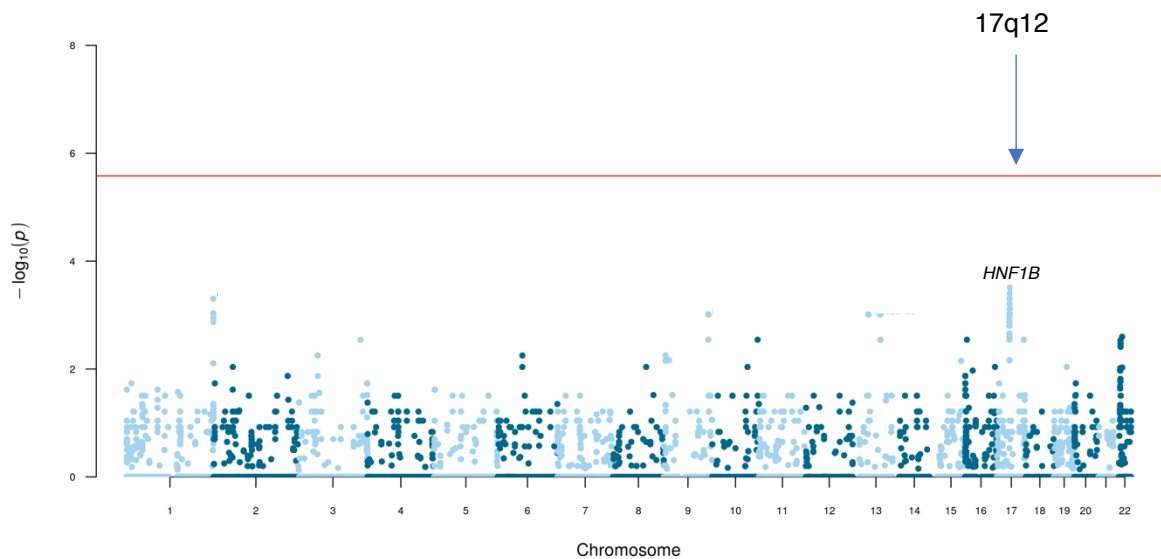
Figure 4.5. Exome-wide burden analysis of rare CNVs in 813 CAKUT probands and 25,205 ancestry-matched controls.

Each dot represents a gene. The red line denotes the Bonferroni corrected exome-wide significance threshold of $P=2.63 \times 10^{-6}$. The 17q12 locus containing *HNF1B* is highlighted with a blue arow.

Stratifying the analysis by phenotype revealed that this signal was driven primarily by the cystic dysplasia cohort where both the 17q12 and 22q11.2 loci were non-significantly enriched for copy number variation (minimum $P=3.1 \times 10^{-4}$ [*TBC1D3G*] and $P=3.23 \times 10^{-4}$ [*SERPIND1 and SNAP29*], respectively) (Figure 4.6).

Interestingly, there was also enrichment of rare CNVs affecting a cluster of olfactory receptor genes on chromosome 1q44 (*OR2T33*; $P=4.98 \times 10^{-4}$) (Figure 4.5). OR genes are commonly found in segmentally duplicated regions and are known to be frequently copy-number variable (Young *et al.* 2008), however the identification of this (non-significant) signal has presumably occurred by chance and is not related to our phenotype. This highlights the perils of statistical noise

and emphasizes the need for stringent multiple testing correction as well as independent replication before robust inferences can be made.
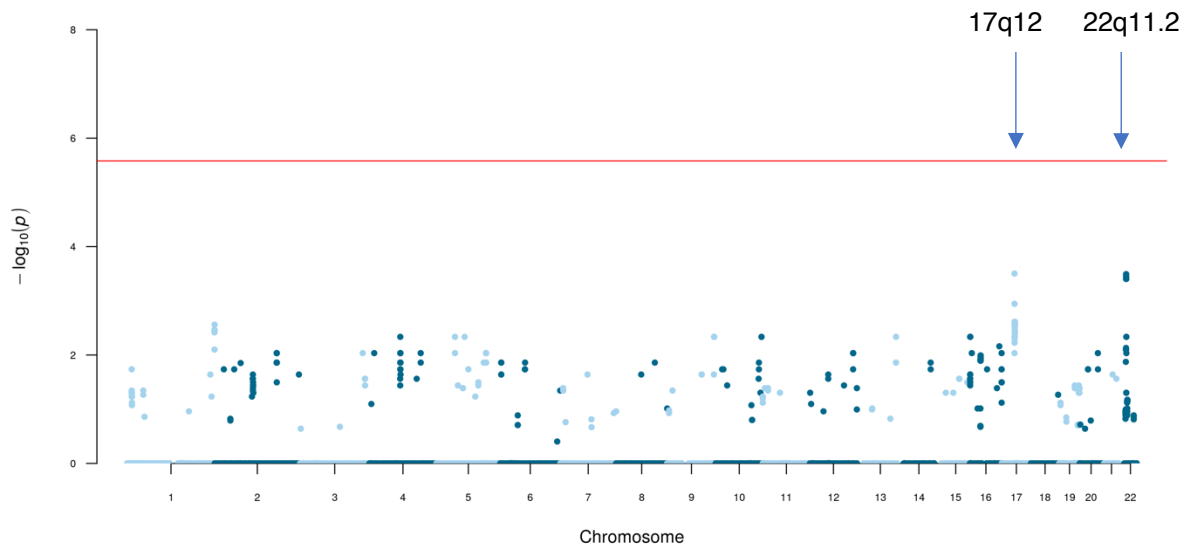
Figure 4.6. Exome-wide burden analysis of rare CNVs in 112 probands with cystic dysplasia and 24,084 ancestry-matched controls.

Each dot represents a gene. The red line denotes the Bonferroni corrected exome-wide significance threshold of P=2.63x10^{-6}. The 17q12 and 22q11.2 loci are highlighted with blue arrows.

4.4.3 Candidate genetic drivers in recurrent deletion syndromes

Examination of genes affected in recurrent deletion syndromes associated with CAKUT can help to prioritize candidate genetic drivers of renal tract anomalies. *HNF1B* has been clearly established as the genetic driver at the 17q12 locus (Mefford *et al.* 2007) and *CRKL* (CRK-like) and *TBX6* (T-box transcription factor 6) have recently been proposed at the 22q11.2 and 16p11.2 loci, respectively (Lopez-Rivera *et al.* 2017, Verbitsky *et al.* 2019, Yang *et al.* 2020). I therefore asked whether an agnostic exome-wide burden testing approach provided any additional support for candidate genes at these loci.

4.4.3.1 22q11.2 Locus

The 22q11.2 locus comprises eight highly homologous LCRs (LCR22A-H) which predispose to recurrent genomic rearrangements (Babcock *et al.* 2007). Deletions affecting this region result in DiGeorge/velocardiofacial syndrome (MIM 188400/192430) which is the most common microdeletion syndrome in humans with an incidence of 1 in 4000 births (Devriendt *et al.* 1998). Heterozygous *de novo* 2.5-3Mb recurrent deletions between LCR22 A and D at 22q11.2 are the most common cause of DiGeorge syndrome (MIM 188400) and result in heart defects, parathyroid and thymic hypoplasia, cleft palate, abnormal facies, and developmental delay. Approximately 40% of patients have associated renal anomalies, predominantly renal agenesis or hypodysplasia (Kobrynski and Sullivan 2007), with a recurrent 370kb region between LCR22 C and D associated with kidney defects. Inactivation of *Crkl* within this locus results in kidney and urinary tract malformations in zebrafish and mouse models and an increased burden of rare *CRKL* variants has been reported in individuals with renal hypodysplasia suggesting a causal role for this gene in the pathogenesis of CAKUT (Lopez-Rivera *et al.* 2017).

Examination of our cohort of 813 CAKUT cases and 25,205 ancestry-matched controls did not identify anyone with the classic 2.5Mb 22q11.2 deletion associated with DiGeorge syndrome. 2/813 (0.2%) cases had a 413kb and 743kb CNV overlapping with the deletion between LCR22 C and D (Figure 4.7), replicating the association of this region with CAKUT. Similar sized CNVs were seen in just 2/25,205 (0.01%) controls (Fisher's exact P=5.6x10^{-3}). Both cases had cystic renal dysplasia and one had a mother with unilateral renal agenesis in whom the 413kb CNV segregated.
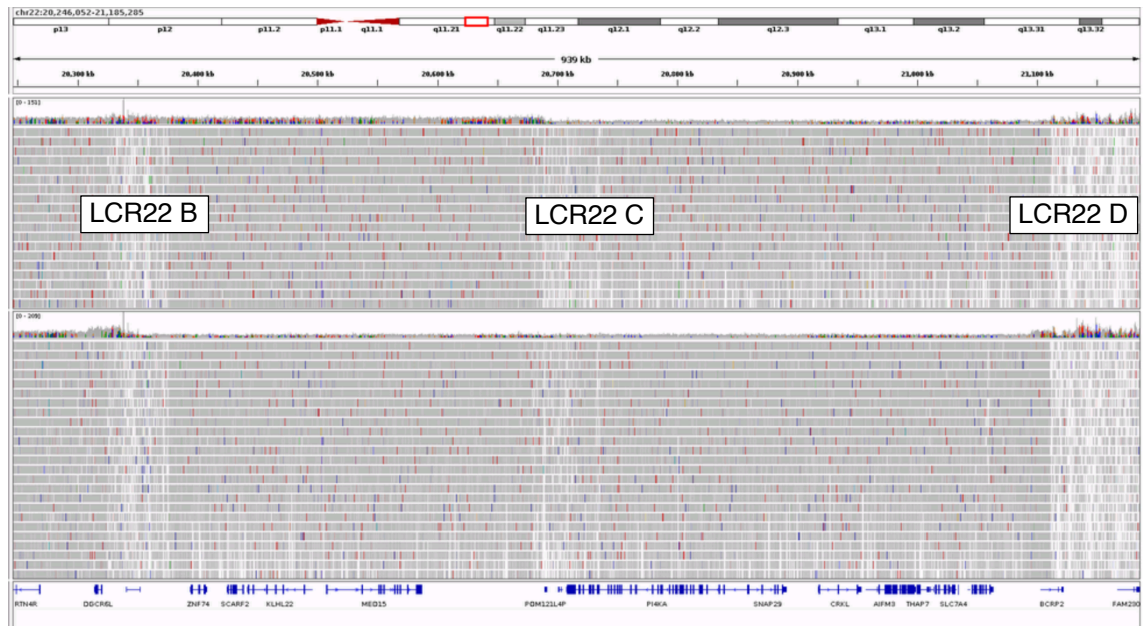
Figure 4.7. Coverage and alignment of the 22q11.2 region in two individuals with cystic renal dysplasia.

Overlapping microdeletions between LCR22 C and D are shown. LCR22, chromosome 22 low-copy repeat. Image generated using Integrated Genome Viewer (IGV).

Further assessment of the rare variant burden in the nine genes found in this recurrent deletion between LCR22 C and D are shown in Table 4.2. CNVs affecting *SERPIND1* and *SNAP29* were most strongly associated with CAKUT although neither of these genes were enriched for SNVs/indels. In contrast to Lopez-Riviera *et al.* (2017), no significant burden of likely deleterious SNV/indel variation affecting *CRKL* was seen in this study but *AIFM3* (apoptosis inducing factor, mitochondria associated 3) reached a minimum *P*-value of 0.01.

AIFM3 induces apoptosis via caspase activation (Xie *et al.* 2005) and is expressed in the developing human kidney (Verbitsky *et al.* 2019, Lozic *et al.* 2021). Furthermore, knockdown of *aifm3* in zebrafish embryos in combination with *snap29* results in renal defects (Verbitsky *et al.* 2019) providing additional support for *AIFM3* as a candidate genetic driver of renal anomalies at the 22q11.2 locus

Table 4.2. Candidate genetic drivers of renal anomalies at the 22q11.2 locus.

Burden of rare variation in 813 CAKUT probands and 25,205 ancestry-matched controls. CNV burden calculated using a two-sided Fisher's exact test. Likely deleterious SNV/indel burden calculated using SKAT-O. Unadjusted *P* values are shown.

Gene	CNV burden	SNV/indel burden
SERPIND1	2.9×10^{-3}	1
SNAP29	2.9×10^{-3}	0.41
CRKL	3.4×10^{-3}	1
PI4KA	3.4×10^{-3}	0.65
AIFM3	3.9×10^{-3}	0.01
LZTR1	3.9×10^{-3}	0.14
P2RX6	3.9×10^{-3}	0.55
SLC7A4	3.9×10^{-3}	0.58
THAP7	3.9×10^{-3}	0.68

4.4.3.2 16p11.2 Locus

I next examined the highly pleiotropic 16p11.2 locus to determine whether there was any statistical evidence to support a particular candidate genetic driver of renal anomalies in this region. 16p11.2 recurrent microdeletion syndrome (MIM 611913) describes a ~600kb deletion encompassing > 25 genes which manifests as developmental delay, intellectual disability, autism spectrum disorder, heart defects and/or vertebral anomalies. A wide range of renal tract anomalies are also observed in up to 40% (Sampson *et al.* 2010, Verbitsky *et al.* 2019).

2/813 (0.2%) CAKUT cases were heterozygous for the recurrent 16p11.2 microdeletion compared with 8/25,205 (0.03%) of controls (*P*=0.04). One of the cases had cystic renal dysplasia and VUR along with other features of the 16p11.2 microdeletion syndrome, while the other had bladder exstrophy, which has not been previously associated with this syndrome. The frequency of this deletion in controls is consistent with an estimated population prevalence of ~3

in 10,000 (Stefansson *et al.* 2014) and suggests that a proportion of these control individuals may have undetected renal tract anomalies or other features of the syndrome such as autism or subtle neurocognitive defects.

Previous deletion mapping of this region in 9 individuals with varied CAKUT phenotypes identified an overlapping ~175kb region containing 19 genes with allelic series in mice implicating *TBX6* in this region as a dosage-dependent driver of renal tract anomalies (Verbitsky *et al.* 2019, Yang *et al.* 2020). Closer examination of this locus in our cohort did not identify any statistically significant enrichment of rare CNVs affecting *TBX6* (P=0.10) although the burden of likely deleterious SNV/indels reached nominal significance (P=0.05). None of the other candidate genes at this locus were enriched for rare variation (Figure 4.8).
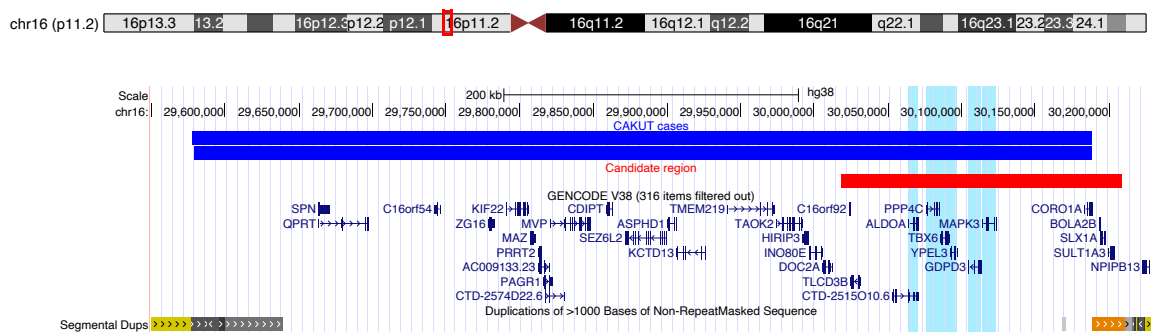
Figure 4.8. 16p11.2 microdeletion seen in two individuals with CAKUT.

CNVs identified in cases are in blue. The previously identified ~175kb candidate CAKUT region is marked in red (Verbitsky *et al.* 2019). Candidate genes are highlighted in light blue. Figure generated using UCSC Genome Browser.

4.4.3   Burden of *cis*-regulatory element structural variation

Given the tightly controlled transcriptional networks that govern embryogenesis I next hypothesized that regulatory regions may be preferentially affected by rare structural variation. To investigate this, I extracted rare, autosomal SVs that

intersected with 926,535 genome-wide cCREs curated from experimental epigenomic data by ENCODE (ENCODE Project Consortium *et al.* 2020). These cCREs include elements with distal enhancer-like signatures (dELS), proximal enhancer-like signatures (pELS) within 2kb of a transcription start site (TSS), promoter-like signatures (PLS) within 200bp of a TSS, CTCF-only elements and DNase-H3K4me3 elements (promoter-like signatures > 200bp from TSS).

Comparing SV types between CAKUT cases and ancestry-matched controls did not reveal any statistically significant enrichment of SVs affecting candidate *cis*-regulatory elements (Figure 4.9 and Table 4.3). Stratification by phenotype identified an enrichment of small duplications (78.8% vs 67.5%; $P=5.0\text{x}10^{-3}$; OR 1.79; 95% CI 1.17-2.83) and inversions (61.4% vs 47.1%; $P=1.2\text{x}10^{-3}$; OR 1.79; 95% CI 1.24-2.59) in 132 PUV cases compared with 23,727 ancestry-matched controls, however after applying a multiple testing correction this was only statistically significant for inversions. The median size of inversions was larger in the PUV cohort compared with other phenotypes (129kb); however, this was not significantly different from matched controls (94kb; $P=0.12$).
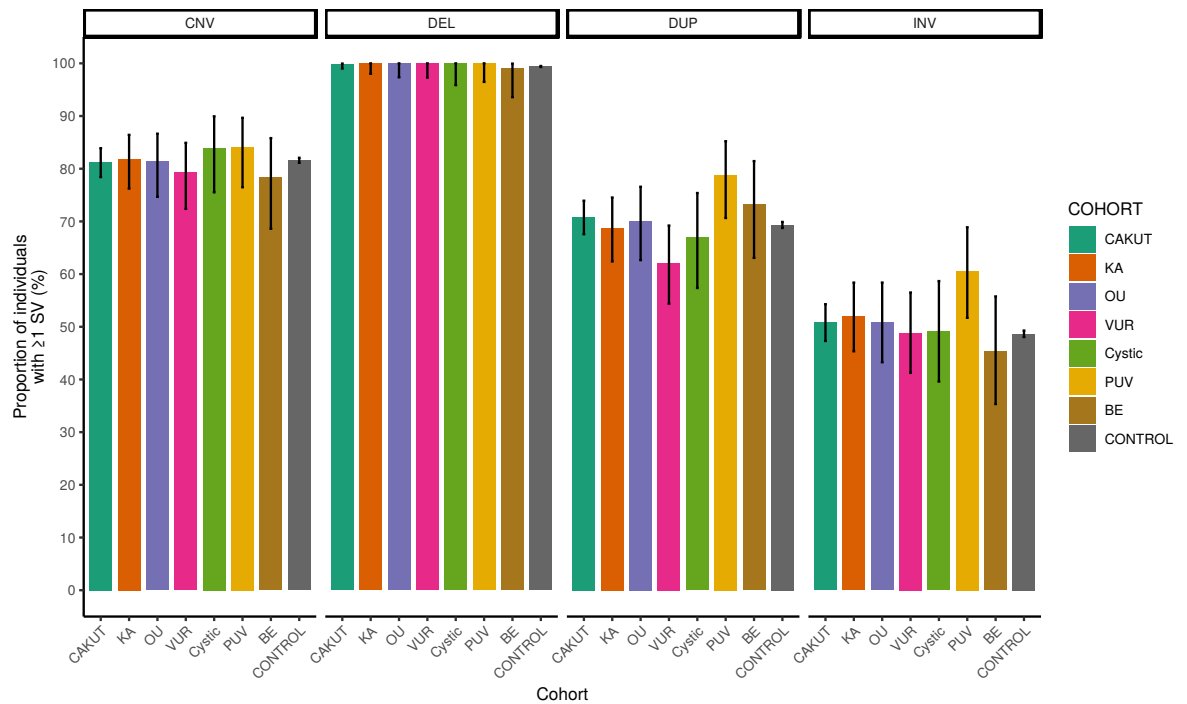
Figure 4.9. Burden of SVs affecting *cis*-regulatory elements in CAKUT probands stratified by phenotype and 25,205 controls.

Vertical black bars indicate 95% confidence intervals. CNV, copy number variant; DEL, deletion; DUP, duplication; INV, inversion; KA, kidney anomaly; OU, obstructive uropathy; VUR, vesico-ureteral reflux; PUV, posterior urethral valves; BE, bladder exstrophy.

Table 4.3. Burden of rare, autosomal candidate *cis*-regulatory element SVs in CAKUT probands stratified by phenotype.

Although each phenotype was ancestry-matched to a separate subset of controls, data for the largest control cohort analyzed is presented here for reference. KA, kidney anomaly; OU, obstructive uropathy; VUR, vesico-ureteral reflux; PUV, posterior urethral valves; BE, bladder exstrophy; CNV, copy number variant; DEL, deletion; DUP, duplication; INV, inversion; OR, odds ratio; CI, 95% confidence interval; IQR, interquartile range.

		CAKUT (n=813)	KA (n=237)	OU (n=177)	VUR (n=174)	Cystic (n=112)	PUV (n=132)	BE (n=97)	Controls (n=25,205)
CNV	n (%)	661 (81.3)	194 (81.9)	144 (81.4)	138 (79.3)	94 (83.9)	111 (84.1)	76 (78.4)	20,567 (81.6)
	OR (CI)	0.98 (0.82-1.18)	1.10 (0.79-1.58)	1.05 (0.72-1.59)	0.94 (0.64-1.39)	1.29 (0.77-2.27)	1.39 (0.87-2.34)	0.90 (0.55-1.54)	
	Fisher's exact P	0.82	0.62	0.85	0.70	0.40	0.20	0.70	
	Median size (kb) (IQR)	76 (121)	79 (113)	73 (132)	79 (119)	79 (117)	80 (165)	84 (117)	75 (122)
	P (Wilcoxon)	0.89	0.62	0.75	0.07	0.65	0.75	0.39	
DEL	n (%)	811 (99.3)	237 (100)	177 (100)	174 (100)	112 (100)	132 (100)	96 (99.0)	25,056 (99.4)
	OR (CI)	2.41 (0.65-20.13)	Inf	Inf	Inf	Inf	Inf	1.62 (0.28-64.9)	
	Fisher's exact P	0.34	0.04	0.12	0.12	0.28	0.04	1	
	Median size (kb) (IQR)	1.7 (4.3)	1.7 (4.3)	1.8 (4.5)	1.6 (4.2)	1.7 (4.4)	1.4 (4.1)	1.8 (4.3)	1.7 (4.3)
	P (Wilcoxon)	0.23	0.84	0.67	0.25	0.92	4.1×10^{-4}	0.68	
DUP	n (%)	576 (50.8)	163 (68.8)	124 (70.1)	108 (62.1)	75 (67.0)	104 (78.8)	71 (73.2)	17,475 (69.3)
	OR (CI)	1.08 (0.92-1.26)	1.05 (0.79-1.41)	1.09 (0.78-1.53)	0.78 (0.57-1.08)	0.96 (0.64-1.47)	1.79 (1.17-2.83)	1.34 (0.85-2.20)	
	Fisher's exact P	0.37	0.78	0.69	0.12	0.84	5.0×10^{-3}	0.23	
	Median size (kb) (IQR)	1.9 (5.0)	1.6 (5.2)	2.7 (5.4)	2.1 (5.1)	1.6 (5.1)	2.0 (5.3)	2.5 (5.2)	2.0 (5.0)
	P (Wilcoxon)	0.35	0.47	0.08	0.70	0.44	0.49	0.70	
INV	n (%)	413 (50.8)	123 (51.9)	90 (50.8)	85 (48.9)	55 (49.1)	81 (61.4)	44 (45.4)	12,266 (48.7)
	OR (CI)	1.09 (0.94-1.26)	1.23 (0.94-1.60)	1.13 (0.83-1.54)	1.09 (0.80-1.48)	1.09 (0.74-1.60)	1.79 (1.24-2.59)	0.98 (0.64-1.48)	
	Fisher's exact P	0.24	0.12	0.45	0.59	0.7	1.2×10^{-3}	0.92	
	Median size (kb) (IQR)	62 (385)	44 (361)	42 (254)	29 (361)	68 (738)	129 (459)	39 (370)	44 (450)
	P (Wilcoxon)	0.26	0.83	0.28	0.49	0.53	0.12	0.75	

Table 4.3. Burden of rare, autosomal candidate *cis*-regulatory element SVs in CAKUT probands stratified by phenotype.

4.4.4 Enrichment of SVs affecting *cis*-regulatory elements in PUV

To further characterize this significant enrichment of SVs observed in individuals with PUV, I next repeated the burden analysis in this cohort stratifying by cCRE subtype (dELS, pELS, PLS, CTCF-only and DNase-H3K4me3 elements). Consistent enrichment was detected across all cCRE types for rare inversions, with the strongest association seen affecting CTCF-only elements (49.2% vs 31.7%; P=3.1x10^{-5}; OR 2.09; 95% CI 1.46-2.99) (Figure 4.10 and Table 4.4). Duplications affecting pELS elements were also significantly enriched in cases compared with controls (29.5% vs 16.8%; P=2.7x10^{-4}; OR 2.08; 95% CI 1.39-3.05). A Bonferroni corrected P-value threshold of 2.5x10^{-3} was used to account for multiple testing (0.05/20).

Given the underlying null distribution is unknown I next calculated empirical P-values using 10,000 permutations for each SV/cCRE combination to confirm whether the observed signal was true. Overall, comparable levels of significance were seen, except in the case of small duplications affecting dELS elements and small deletions affecting pELS elements, which both demonstrated stronger evidence of association using the permutation approach.

On the basis that individuals with African ancestry have greater genetic diversity and more rare SVs per genome than those with European ancestry (median of 468 vs 147) (Collins *et al.* 2020), I also repeated the analysis in a subset of genetically defined European samples to ascertain whether the observed signal was being driven by population structure. A larger proportion of the 88 PUV cases had rare inversions affecting *cis*-regulatory elements than 17,993 controls, but this only reached nominal significance for CTCF-only elements (36.4% vs 26.0%; empirical P=2.8x10^{-2}; OR 1.62; 95% CI 1.02-2.56). This attenuated significance is likely related to reduced power associated with a smaller sample size and suggests that the observed association is not being driven by underlying population structure,
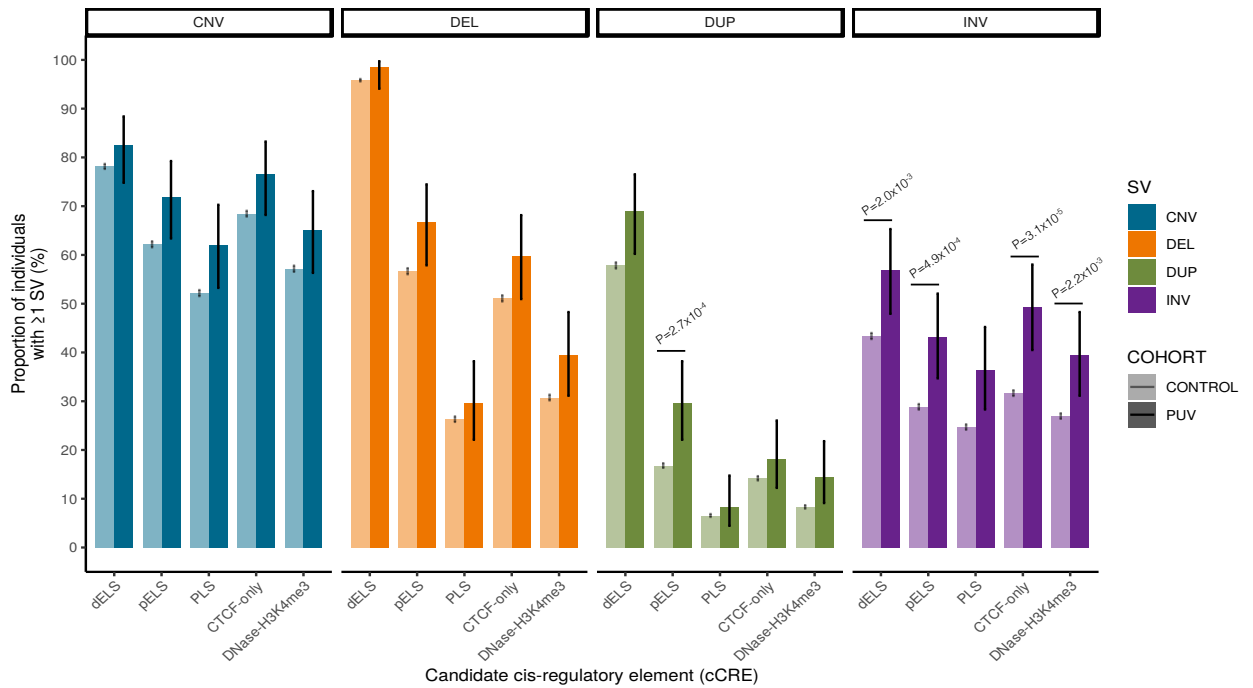
Figure 4.10. The proportion of individuals with ≥1 rare, autosomal SVs intersecting with an ENCODE cCRE in 132 PUV cases and 23,727 ancestry-matched controls.

Vertical black bars indicate 95% confidence intervals. Unadjusted P values shown are significant after correction for multiple testing ($P < 2.5 \times 10^{-3}$). CNV, copy number variant; DEL, deletion; DUP, duplication; INV, inversion; PUV, posterior urethral valves; dELS, distal enhancer-like signature; pELS, proximal enhancer-like signature; PLS, promoter-like signature; cCRE, candidate *cis*-regulatory element.

Table 4.4. Burden of rare, autosomal SVs intersecting with each *cis*-regulatory element type in 132 PUV cases and 23,727 ancestry-matched controls.

Empirical *P*-values were calculated using 10,000 permutations, except for CTCF-only inversions where 100,000 permutations were used. CNV, copy number variant; DEL, deletion; DUP, duplication; INV, inversion; dELS, distal enhancer-like signature; pELS, proximal enhancer-like signature; PLS, promoter-like signature; OR, odds ratio; CI, 95% confidence interval.

		CNV	DEL	DUP	INV
dELS	N of cases (%)	109 (82.6)	130 (98.5)	91 (68.9)	75 (56.8)
	N of controls (%)	18,546 (78.2)	22,743 (95.9)	13,727 (57.9)	10,291 (43.4)
	Fisher's exact *P*	0.25	0.18	0.01	2.0×10^{-3}
	Permutation *P*	0.07	0.04	4.2×10^{-3}	1.9×10^{-3}
	OR (CI)	1.32 (0.84-2.18)	2.81 (0.76-23.5)	1.62 (1.11-2.40)	1.72 (1.20-2.47)
pELS	N of cases (%)	95 (72.0)	88 (66.7)	39 (29.5)	57 (43.2)
	N of controls (%)	14,755 (62.2)	13,446 (56.7)	3,988 (16.8)	6,848 (28.9)
	Fisher's exact *P*	0.02	0.02	2.7×10^{-4}	4.9×10^{-4}
	Permutation *P*	0.02	5.4×10^{-3}	2.0×10^{-4}	5.0×10^{-4}
	OR (CI)	1.56 (1.06-2.35)	1.53 (1.05-2.25)	2.08 (1.39-3.05)	1.87 (1.30-2.68)
PLS	N of cases (%)	82 (62.1)	39 (29.5)	11 (8.3)	48 (36.4)
	N of controls (%)	12,382 (52.2)	6,243 (26.3)	1,557 (6.6)	5,862 (24.7)
	Fisher's exact *P*	0.02	0.43	0.38	3.2×10^{-3}
	Permutation *P*	0.02	0.32	0.27	2.4×10^{-3}
	OR (CI)	1.50 (1.04-2.18)	1.17 (0.79-1.73)	1.29 (0.63-2.41)	1.74 (1.19-2.52)
CTCF-only	N of cases (%)	101 (76.5)	79 (59.8)	24 (18.2)	65 (49.2)
	N of controls (%)	16,247 (68.5)	12,124 (51.1)	3,368 (14.2)	7,512 (31.7)
	Fisher's exact *P*	0.05	0.05	0.21	3.1×10^{-5}
	Permutation *P*	0.03	0.03	0.45	$<1.0 \times 10^{-5}$
	OR (CI)	1.50 (0.99-2.32)	1.43 (0.99-2.06)	1.34 (0.82-2.11)	2.09 (1.46-2.99)
DNase-H3K4me3	N of cases (%)	86 (65.2)	52 (39.3)	19 (14.4)	52 (39.4)
	N of controls (%)	13,567 (57.2)	7,298 (30.8)	1,976 (8.3)	6,397 (27.0)
	Fisher's exact *P*	0.08	0.04	0.02	2.2×10^{-3}
	Permutation *P*	0.05	0.06	0.01	1.9×10^{-3}
	OR (CI)	1.40 (0.97-2.05)	1.46 (1.01-2.10)	1.85 (1.07-3.03)	1.76 (1.22-2.53)

4.5   Summary

- No significant enrichment of rare, gene-disrupting structural variation was identified in individuals with CAKUT as a whole, or when stratified by phenotype.
- The 17q12 and 22q11.2 regions were most frequently affected by copy number variation in CAKUT but seen only in those with cystic renal dysplasia.
- *AIFM3* at the 22q11.2 locus is enriched for both rare CNVs and SNV/indels and is a possible candidate genetic driver of renal anomalies.
- Significant enrichment of rare structural variation affecting *cis*-regulatory elements was observed in individuals with PUV; specifically small duplications affecting candidate enhancer-like elements and inversions affecting CTCF-only elements.

4.6   Discussion

4.6.1   Gene-disrupting SVs are not significantly enriched in CAKUT

This study represents the first detailed examination of genome-wide structural variation in CAKUT carried out using WGS data, with the significant advantage that cases and controls were sequenced on the same platform, therefore minimizing potential bias from technical artefacts. The findings are consistent with previous observations that the 17q12 and 22q11.2 loci are most commonly affected by copy number variation in CAKUT patients (Sanna-Cherchi *et al.* 2012, Lopez-Rivera *et al.* 2017, Verbitsky *et al.* 2019), but no significant enrichment of gene-disrupting structural variation was seen in this cohort as a whole, or when stratified by phenotype.

This is the first time the burden of small deletions, duplications and inversions has been investigated in CAKUT, however there have been several studies analyzing rare CNVs, all published by the same group. The absence of observed CNV burden seen here contrasts with a large study which used SNP-microarrays to identify an enrichment of large ($\geq$ 100kb), rare (AF < 0.1%) autosomal CNVs in 2,824 unrelated individuals with CAKUT compared with 21,498 controls (Verbitsky *et al.* 2019). This signal was primarily driven by known pathogenic CNV losses in the kidney anomaly cohort (n=1,088), consistent with previous data from the same group which identified an increased burden of rare, large (> 500kb) gene-disrupting CNVs in a cohort of 522 individuals with renal hypodysplasia (Sanna-Cherchi *et al.* 2012). Evidence of enrichment was also seen to a lesser extent in those with obstructive uropathy (n=512) and VUR (n=659) but no significant burden of exonic CNVs was seen in those with PUV (n=141), duplex systems (n=233), ectopic or horseshoe kidneys (n=127) or lower urinary tract malformations (n=62). The same authors subsequently confirmed the enrichment of rare, exonic CNVs in VUR using a larger cohort of 1,737 cases and 24,765 controls (Verbitsky *et al.* 2021).

More than double the number of rare, exonic CNVs were seen in this study than observed by Verbitsky *et al.* (78% vs 37%) which were smaller on average (median size 87kb vs 245kb), reflecting the increased sensitivity and resolution of WGS for SV detection compared with microarrays, as well as the difference in size threshold for inclusion ($\geq$ 10kb compared with $\geq$ 100kb). As a direct comparison I increased the size threshold to $\geq$ 100kb but still did not identify any significant difference between cases and controls (49.8% vs 50.7%; *P*=0.64; median size 205kb vs 207kb). A similar proportion of very large (> 1Mb) CNVs were however reported in both this and the study by Verbitsky *et al.* (4% and 5%) demonstrating comparable sensitivity of microarrays and WGS for larger variants.

Two possibilities may explain the difference seen in overall CNV burden between this study and Verbitsky *et al.* (2019). First, although both CAKUT cohorts were similar in phenotypic breakdown and report a similar proportion of familial disease

(14.6%) and extra-renal manifestations (20.2%), Verbitsky *et al.* (2019) had a significantly larger cohort, over a third of whom had kidney anomalies, and therefore were better powered to detect association. Second, the difference may have resulted from the differing methodological approaches. For example, the multiple genotyping arrays used by Verbitsky *et al.* (2019) may have introduced confounding through batch effects and the variant-calling tool used (PennCNV) has been shown to have a false positive rate of up to 24% (Eckel-Passow *et al.* 2011). Furthermore, Verbitsky *et al.* (2019) filtered CNVs using allele frequencies generated from only control samples (rather than the entire cohort) which may introduce additional bias and potentially result in false positive associations. Future collaborative studies involving large cohorts all sequenced on the same platform are therefore necessary to definitively confirm or refute the association.

4.6.2 Regulatory SVs are significantly enriched in PUV

A significant enrichment of rare structural variation affecting cCREs was identified in individuals with PUV. Strong association was observed for inversions affecting CTCF-binding elements and for small duplications (< 10kb) affecting proximal enhancer-like elements. CTCF is a highly conserved, ubiquitously expressed, zinc finger protein which plays a key role in chromatin insulation and gene regulation (Chen *et al.* 2012). While typically the pathogenicity of SVs can be predicted by their effects on gene dosage, SVs can also cause disease by altering the copy number or position of regulatory elements or reshuffling higher-order chromatin structures such as topologically associated domains or TADs (Spielmann *et al.* 2018). This has been clearly demonstrated in the context of congenital limb malformations where CNVs (Flöttmann *et al.* 2018), duplications (Dathe *et al.* 2009), deletions (Lupiáñez *et al.* 2015), and inversions (Lettice *et al.* 2011, Lupiáñez *et al.* 2015) have all been shown to affect non-coding limb enhancers or TAD boundaries. Regulatory SVs have also been implicated in autism (Brandler *et al.* 2018) and neurodevelopmental delay (Redin *et al.* 2017).

Current understanding of the clinical relevance of inversions is limited as the balanced nature and location of breakpoints within complex repeat regions have made detection challenging (Puig *et al.* 2015). Although the vast majority of inversions have no obvious phenotypic consequences, they can increase the risk of further genomic rearrangements (Osborne *et al.* 2001) and recurrent inversions have been associated with haemophilia A (Lakich *et al.* 1993), Hunter syndrome (Bondeson *et al.* 1995), neurodegenerative (Webb *et al.* 2008) and autoimmune disease (Salm *et al.* 2012, Namjou *et al.* 2014). The enrichment of rare inversions affecting cCREs therefore raises the interesting possibility that non-specific perturbation of long-range regulatory networks or TADs could result in developmental anomalies of the lower urinary tract.

4.6.3 Disruption of TADs as a mechanism of disease

The human genome is organized into a hierarchical 3D structure which mediates long-range regulation of gene expression. Underlying this structure are TADs, megabase-sized local chromatin interaction domains with boundaries enriched by CTCF-bound insulator elements (Dixon *et al.* 2012, Nora *et al.* 2017). TADs are formed by chromatin loop extrusion through the ring-shaped cohesin complex until it encounters convergent CTCF-bound sites and are largely conserved across cell types, species and development (Dixon *et al.* 2012, 2015). These structured domains mediate long-range enhancer-promoter interactions via chromatin looping to bring functional elements into proximity and facilitate the assembly of transcriptional machinery (

Figure 4.11).
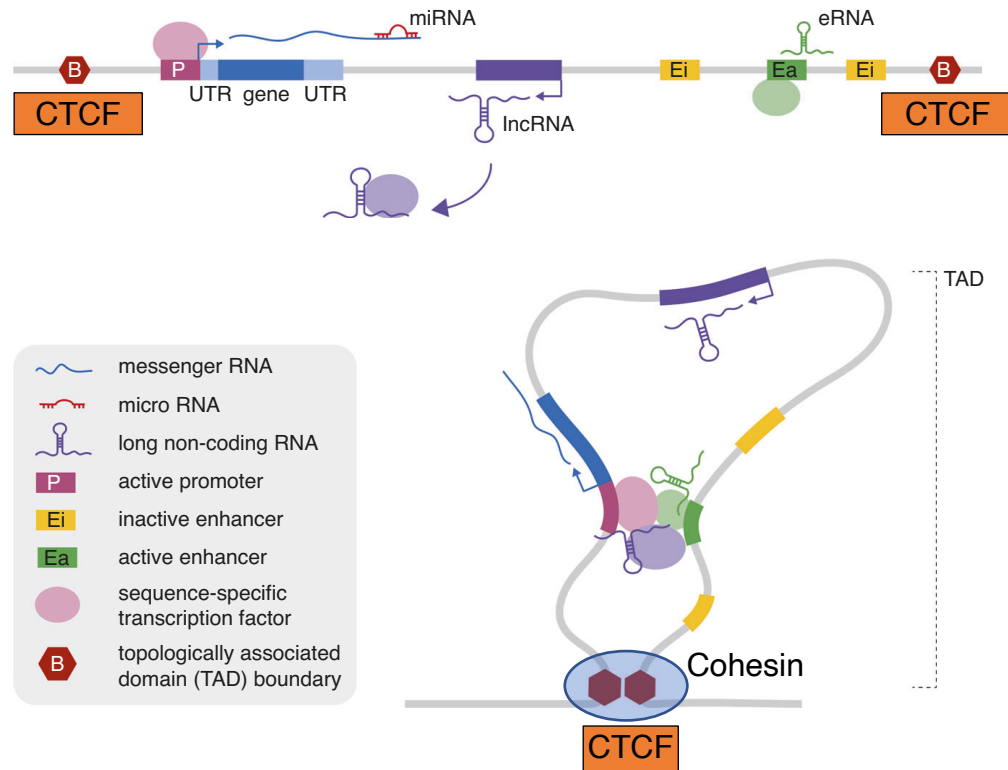
Figure 4.11. The gene regulatory landscape.

The locus is delimited by a cluster of CTCF-binding sites to form a topologically associated domain (TAD). These CTCF-associated TAD boundaries anchor the chromatin loop and enable functional elements to come into proximity allowing assembly of the transcriptional machinery. Adapted from D'haene and Vergult 2021.

TAD disruption is a common phenomenon in cancer development (Hnisz *et al.* 2016, Northcott *et al.* 2017) and is increasingly being recognized as a mechanism of disease in developmental disorders and congenital malformation syndromes. Structural variation affecting CTCF-associated TAD boundaries is subject to negative selection suggesting a deleterious effect (Fudenberg and Pollard 2019, Han *et al.* 2020) and recent studies have demonstrated an association with neurodevelopmental phenotypes (Redin *et al.* 2017, Lowther *et al.* 2022), limb malformations (Lupiáñez *et al.* 2015) and lung agenesis (Melo *et al.* 2021) in

167

humans. Disruption of these boundaries by both balanced and unbalanced structural variation can alter the chromatin structure, result in ectopic enhancer-promoter interactions and subsequently the misexpression of genes (Lupiáñez *et al.* 2015, Symmons *et al.* 2016).
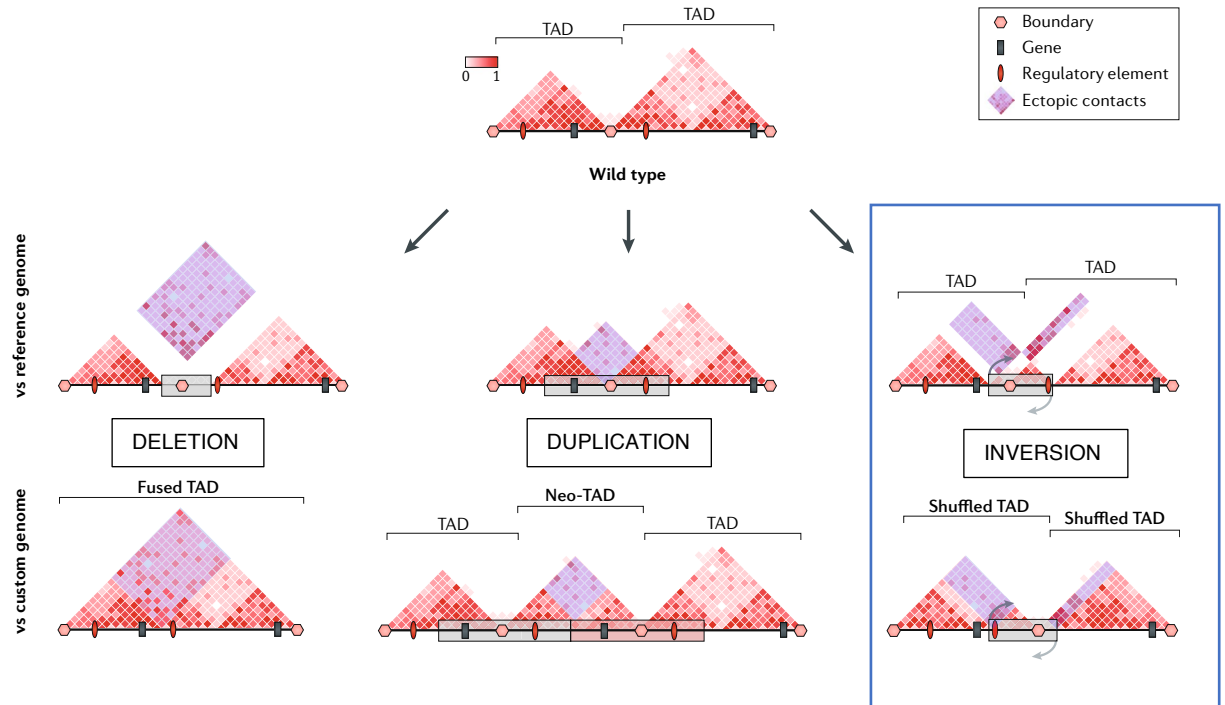
Figure 4.12. Structural variation affecting TAD boundaries.

Changes in chromatin conformation can result in ectopic enhancer-promoter interactions. Inversions that cross TAD boundaries can lead to 'hijacking' of enhancers from a nearby TAD and ectopic gene expression. Adapted from Spielmann *et al.* 2018.

Deletions of CTCF-associated TAD boundaries may lead to fused TADs whereas duplications can result in the formation of neo-TADs that are isolated from the rest of the genome (Figure 4.12) (Franke *et al.* 2016). Inversions that cross TAD boundaries can cause two domains to fuse together (TAD shuffling) and lead to enhancer adoption (or 'hijacking') with ectopic gene expression (Lettice *et al.* 2011). For example, an inversion affecting the enhancer cluster and nearby TAD boundary at the *Epha4* locus resulted in congenital limb malformations as a consequence of

enhancer hijacking (essentially a gain of function mutation) and ectopic expression of *Wnt6* which is usually in a separate TAD (Lupiáñez *et al.* 2015).

Several key functional experiments have demonstrated the importance of CTCF-binding sites in determining 3D chromatin structure and the regulation of developmental gene expression. Inversion of CTCF-binding sites *in vitro* can disrupt chromatin looping and enhancer-promoter interactions highlighting how 3D architecture can be encoded by the location and relative orientation of CTCF-binding sites (Guo *et al.* 2015). Deletion of a CTCF-binding site within the *HoxA* gene cluster resulted in expansion of active chromatin into an adjacent repressive domain and altered gene expression during differentiation of motor neurons highlighting the importance of CTCF in the maintenance of discrete functional domains during development (Narendra *et al.* 2015). Furthermore, *in vivo* studies have demonstrated that CTCF has a key role in the establishment of chromatin structure during human embryogenesis (Chen *et al.* 2019) and is essential for the regulation of gene expression during embryo patterning and organogenesis (Franke *et al.* 2021). These data provide strong support for the essential role of CTCF and chromatin architecture during embryogenesis and support the hypothesis that inversions affecting CTCF-only elements may result in perturbations of 3D chromatin structure leading to the misexpression of genes and subsequently developmental defects of the lower urinary tract.

4.6.4   Strengths and limitations

This is the first comprehensive assessment of both coding and non-coding structural variation in CAKUT carried out using WGS data, providing new insights beyond those generated from array-based CNV studies. WGS offers improved SV resolution and detection over conventional microarrays and using case-control data from more than 20,000 individuals sequenced on the same platform minimized confounding by technical artefacts.

There are however several limitations to this analysis. First, although I used ancestry-matched cases and controls, I was unable to include principal components to adjust for population structure using a Fisher's exact test. Second, the cCRE definitions used for the non-coding burden analysis were a consensus set and not tissue specific. Given the spatio-temporal sensitivity of development they therefore might not be active in the developing embryo in the right place or at the right time. CTCF ChIP-seq data was also not available for mesendoderm cells, although CTCF-binding is largely conserved between cell types and across development. Third, the burden analysis performed was based on a dominant inheritance model and an additive or recessive model may have identified additional signals. Fourth, the SVs have not been validated through long-read sequencing or PCR-based techniques. Fifth, the approach used to calculate the MAF of SV calls used 10kb bins to group variants of the same type. This is therefore an estimation because of the variability in genomic coordinates seen between calls, and the difficulty in determining whether overlapping calls are the same or different variants. This 'spatial uncertainty' may have resulted in some rare SVs being incorrectly grouped together and excluded from the analysis. Sixth, the detected associations have yet to be replicated in an independent cohort.

Finally, short-read WGS has well established limitations in the detection of SVs, with significant read-mapping ambiguity affecting larger variants in complex, repetitive and GC-rich regions. A recent comparison of Illumina's short read and PacBio's long read platforms demonstrated 47% of deletions and 78% of insertions were missed by short-read WGS, the majority of which were intermediate-size variants (50bp-2kb) driven by tandem repeat variation and mobile element insertions (Chaisson *et al.* 2019). False positive calls are also a concern and are dependent on the SV calling algorithm used. While CANVAS (Roller *et al.* 2016) and MANTA (Chen, Schulz-Trieglaff, *et al.* 2016) were used by Genomics England to detect different types of structural variation, it is generally recommended to merge overlapping calls from pairs of algorithms (Kosugi *et al.* 2019) and incorporating additional variant callers

such as LUMPY (Layer *et al.* 2014) or DELLY (Rausch *et al.* 2012) would minimize the risk of false positives in future analyses.

4.7 Conclusion

Recent technological and bioinformatic advances have led to an exponential increase in our ability to detect and characterize structural variation, although how we analyze and interpret these data is still in its infancy. Furthermore, recent insights into the 3D structure of chromatin and how this can be impacted by structural variation has expanded our understanding of mechanisms of disease and is something that appears to be especially important for developmental phenotypes.

In this chapter I identified an enrichment of rare structural variation affecting *cis*-regulatory elements in individuals with PUV suggesting that disruption of the tightly regulated regulatory networks that govern urinary tract development is important in the pathogenesis of this disorder. Furthermore, strong association with rare inversions affecting CTCF-binding elements imply that alterations in chromatin structure may be a key mechanism, perhaps due to the sensitivity of mesonephric duct integration into the posterior urethra to even minor abnormalities of gene expression. Future work will be necessary to validate and replicate these findings and to determine the precise mechanisms involved.

Chapter 5: CAKUT as a complex disorder

5.1 Introduction

The previous two chapters have focused exclusively on rare variation and have shown that in this cohort at least, renal tract malformations cannot usually be explained by single-gene defects detectable by WGS. Furthermore, the considerable genetic and phenotypic heterogeneity associated with CAKUT supports the hypothesis that this is in fact a complex phenotypic spectrum that may be influenced by multiple genetic variants with small effects, consistent with a polygenic rather than monogenic model of inheritance. In this chapter, I therefore explore how both common and low-frequency variants contribute to the genetic landscape of structural renal tract malformations and estimate the proportion of phenotypic variance they explain.

Traditional GWAS have been constrained by cost and technology, relying on array-based platforms and largely European-centric imputation panels which can limit the analysis of rare variants and those found in non-European populations. Furthermore, over 90% of GWAS performed have been carried out in individuals of European ancestry resulting in a drive by the scientific community to increase diversity in genetic association studies. In response to this, sequencing-based GWAS using WGS data have recently been used successfully in mixed ancestry cohorts for common complex traits (e.g., COPD and red blood cell traits) from the NHLBI TOPMed programme (Keramati *et al.* 2019, Zhao *et al.* 2020, Hu *et al.* 2021) and I therefore aimed to establish whether a similar approach could be used effectively in the context of rare disease.

5.2   Aims

1. To assess the contribution of common and low-frequency SNVs and indels (MAF $\geq 0.1\%$) to the genetic architecture of CAKUT as a whole and when stratified by phenotype (kidney anomalies, obstructive uropathy, VUR, PUV, cystic dysplasia and bladder exstrophy) using a sequencing-based GWAS.

2. To ascertain whether any genes or functional pathways are enriched for common and low-frequency variation which might provide insights into the biological mechanisms underlying the pathogenesis of CAKUT.

3. To determine what proportion of phenotypic variance in CAKUT can be explained by common and low-frequency variation.

4. To replicate and fine-map novel and previously identified single variant associations in PUV, VUR and bladder exstrophy.

5. To establish whether a mixed-ancestry sequencing-based GWAS approach can be used for disease locus discovery in a rare disease.

5.3   Methods

5.3.1   Sequencing-based GWAS

A sequencing-based GWAS was carried out using the R package SAIGE (Zhou *et al.* 2018). Sex and the top ten principal components were used as covariates and a score test (Chen, Wang, *et al.* 2016) for association performed. The top five principal components were included when case numbers were less than 120. This was to prevent overfitting (where the model describes random error in the data rather than relationships between variables) and loss of power. A sample size threshold of 120 was chosen based on data from simulation studies that indicate there should be no fewer than ten outcomes per independent variable to minimize the risk of overfitting

173

(Stoltzfus 2011) i.e. inclusion of ten PCs and sex as covariates means a minimum sample size of 110.

The optimum number of PCs for inclusion was also tested in the European only PUV cohort (88 cases and 17,993 controls) by comparing lambda (λ) when using 0, 2, 5, or 10 PCs as covariates. Five PCs achieved a genomic inflation factor closest to 1 (0.91) and this number of covariates was therefore included in the analysis.

Number of principal components	λ
0	0.83
2	0.64
5	0.91
10	0.87

SNVs and indels with MAF $\geq$ 0.1% that passed the following quality control filters were included:

- MAC $\geq$ 20
- missingness < 1%
- HWE $P > 10^{-6}$
- differential missingness $P > 10^{-5}$

One limitation of SAIGE is that the beta values estimated from score tests can be biased at low MACs and therefore ORs for variants with MAF < 1% were calculated separately using case-control allele counts in R.

5.3.2 Conditional analysis and epistasis

At each of the genome-wide significant loci, SAIGE was used to perform:

- conditional analysis on the lead variant to identify potential secondary independent associations

- high resolution single variant analysis using all variants with MAC $\geq$ 3 to ascertain whether the observed signal was being driven by rare variation

Epistasis between the lead variants was assessed using logistic regression in PLINK (version 1.9) (Purcell *et al.* 2007).

5.3.3 Gene and gene-set analysis

Aggregation of variants increases power to detect multiple weaker associations and can test for association with specific biological or functional pathways. MAGMA (version 1.09a) (de Leeuw *et al.* 2015) was used to test the joint association of all variants with MAF $\geq$ 0.1% within a particular gene or gene-set using the WGS data and same quality control filters as the GWAS (Section 5.3.1). Variants were assigned to 17,636 protein coding genes (NCBI build 38, downloaded 29/04/2015 by the developers of MAGMA) with exome-wide significance defined as $P=2.61\times10^{-6}$ (0.05/19,139 genes).

Competitive gene-set enrichment analysis was then performed for 50 hallmark gene sets from MsigDB (version 7.0) (Liberzon *et al.* 2015) using the results of the single gene analysis. These hallmark gene sets summarize information to generate a manually curated set of genes with coordinated expression that represent well-defined biological processes. Competitive analysis tests whether the joint association of genes in a gene-set is stronger than a randomly selected set of similarly sized genes. Bonferroni correction was applied for the total number of tested gene sets ($P=0.05/50=0.001$).

5.3.4 Replication

A replication study was performed for bladder exstrophy and PUV. The replication cohort consisted of 84 individuals with bladder exstrophy (recruited from Manchester, UK) and 398 individuals with PUV; 336 recruited from Poland and Germany as part of the CaRE for LUTO (Cause and Risk Evaluation for Lower

Urinary Tract Obstruction) Study, and 62 from Manchester, UK. None of the individuals in the replication cohorts had been recruited to the 100KGP. All were of self-reported European ancestry.

The replication control cohort consisted of 10,804 genetically determined unrelated European individuals recruited to the cancer arm of the 100KGP, excluding those with urinary tract (kidney, bladder, or prostate) or childhood malignancy.

KASP (Kompetitive Allele-Specific PCR) genotyping of the lead variants at loci with suggestive association ($P < 5 \times 10^{-7}$) was performed by LGC Biosearch Technologies. For bladder exstrophy: rs6106456 at 20p11.22. For PUV: rs10774740 at 12q24.21, rs144171242 at 6p21.1, rs1471950716 at 10q11.21, and rs199975325 at 14q21.1. The location of rs1471950716 at 10q11.21 in a low complexity region caused the genotyping assay to fail and another variant with evidence of association (rs137855548; $P=1.46 \times 10^{-6}$) was used instead.

Allele counts at each variant were compared between cases and controls using a two-sided Cochran-Armitage trend test. A Bonferroni-corrected $P < 0.0125$ (0.05/4) was used to adjust for the number of loci tested in the PUV analysis. Power to detect or refute association at each locus was calculated using the R package `genpwr` as > 0.99.

5.3.5 Bayesian fine-mapping

I applied PAINTOR (v3.1) (Kichaev *et al.* 2014), a statistical fine-mapping method which uses an empirical Bayes prior to integrate functional annotation data, LD patterns and strength of association to estimate the posterior probability (PP) of a variant being causal. Variants within a 100kb window centring on the lead variant at each genome-wide significant locus with $P < 0.05$ were extracted. Z-scores were calculated as effect size (β) divided by standard error. One causal variant was assumed per locus. LD matrices of pairwise correlation coefficients were derived using 1000 Genomes European data (Phase 3) (1000 Genomes Project Consortium

et al. 2015) as a reference, excluding variants with ambiguous alleles (A/T or G/C). Each locus was intersected with the following functional annotations downloaded using the UCSC Table Browser (Sugnet *et al.* 2002): GENCODE (Frankish *et al.* 2019) (v29) transcripts (wgEncodeGencodeBasicV29, updated 2019-02-15), PhastCons (Weinstock *et al.* 2005) (phastConsElements100way, updated 2015-05-08), ENCODE (ENCODE Project Consortium *et al.* 2020) cCREs (encodeCcreCombined, updated 2020-05-20), transcription factor binding clusters (encRegTfbsClustered, updated 2019-05-16), DNase I hypersensitivity clusters (wgEncodeRegDnaseClustered, updated 2019-01-08) and H1 Human embryonic stem cell Hi-C data (h1hescInsitu from (Krietenstein *et al.* 2020)). For bladder exstrophy 74 variants at 5q11.1 were included. A total of 351 variants at 12q24.21 and 166 variants at 6p21.1 were analyzed for PUV.

5.3.6 Heritability

To determine the relative contribution of genetics versus environment, an estimation of heritability can be made, defined as the proportion of phenotypic variance explained by genetic factors (Figure 5.1). Broad-sense heritability (H^2) refers to the phenotypic variance explained by all genetic factors (additive, dominant and epistatic) but is very hard to estimate without making strong assumptions (Visscher *et al.* 2008). For this reason, narrow-sense heritability (h^2) is commonly used instead, estimating purely the contribution of additive genetic variance . In CAKUT, environmental factors (specifically *in utero* and maternal factors) are also likely to contribute significantly to phenotypic variance (Nicolaou *et al.* 2015, Groen In 't Woud *et al.* 2016).
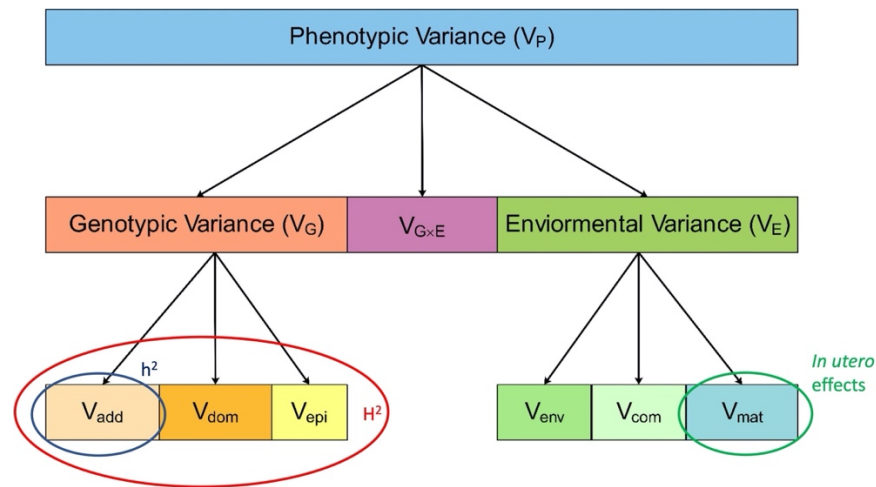
Figure 5.1. Phenotypic variance and heritability.

Genotypic variance (V_G) can be partitioned into additive (V_{add}), dominant (V_{dom}) and epistatic (V_{epi}) effects. Environmental variance (V_E) can be partitioned into common effects (V_{com}) such as living with the same family, maternal effects (V_{mat}) and residual stochastic effects (V_{env}). V_{GxE}, variance due to gene-environment interactions. h^2, narrow-sense heritability; H^2, broad-sense heritability. Adapted from Zhu and Zhou 2020.

To estimate the phenotypic variance explained by common and low-frequency variants in CAKUT, heritability (h^2) analysis was performed with GCTA (version 1.93.1beta) (Yang *et al.* 2011). Variants with MAF $\geq$ 0.1% were included and underwent the same quality control filtering as described in Section 5.3.1. Using the GREML-LDMS approach (Yang *et al.* 2015), variants were stratified into seven different bins based on MAF (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) and for each bin of variants, SNP-based LD scores were calculated over a 200kb region (with 100kb overlap between two adjacent segments). For a given bin of variants defined by MAF, variants were further stratified into quartiles using LD scores. For each of the 28 bins subset by MAF and LD, GCTA was used to produce a GRM from the raw genotype files. Given this method has not been validated in a mixed ancestry cohort, I performed the heritability estimates in a subset of individuals with genetically defined European ancestry (623 cases and

20,060 controls). The REML (restricted maximum likelihood) function was then used to conduct a GREML-LDMS analysis using the 28 GRMs, including the top four principal components as covariates. As the selected cohort was genetically European, the top four principal components were sufficient to account for underlying population substructure without overfitting the model.

When estimating the heritability of a binary trait, the proportion of cases included is usually much higher than the prevalence in the general population making h^2 estimates liable to ascertainment bias (Lee *et al.* 2011). This 'observed' heritability should therefore be transformed to an underlying continuous 'liability threshold' model adjusting for the ratio of cases to controls and the population prevalence of disease which provides a less biased h^2 that can be used for comparison with other traits. In this analysis a CAKUT prevalence of 0.2% was used to transform the observed heritability to a liability threshold model.

5.4   Results

5.4.1 CAKUT

Given the low prevalence of monogenic disease and lack of rare variant enrichment seen in this cohort, I hypothesized that common and low-frequency variation might contribute to the genetic architecture of this complex and heterogenous group of anomalies.

5.4.1.1   *Mixed-ancestry GWAS*

I performed a sequencing-based GWAS in 813 unrelated CAKUT probands and 25,205 ancestry-matched controls testing for association at 19,193,915 SNVs and indels with MAF $\geq 0.1\%$ (Figure 5.2). The genomic inflation factor (λ) was 1.02 indicating minimal evidence of confounding by population structure in this mixed-ancestry cohort (Figure 5.3).

Two single variants reached genome-wide significance. One indel (rs35251516 at 1q21.3; P=4.94x10^{-8}) was in a low-complexity region and deemed to be a sequencing artefact after closer examination of the regional association plot (Figure 5.4).
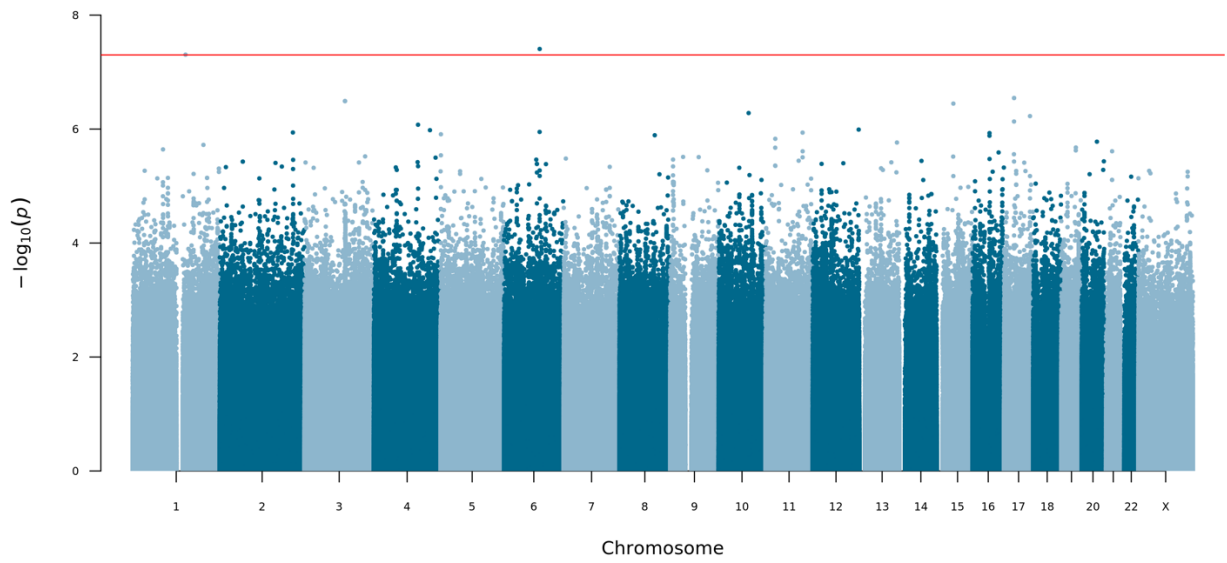
Figure 5.2. Manhattan plot of CAKUT mixed-ancestry GWAS.

A sequencing-based GWAS was carried out in 813 unrelated CAKUT cases and 25,205 controls for 19,193,915 variants with MAF $\geq$ 0.1%. Chromosomal position (GRCh38) is denoted along the x axis and strength of association using a $-\log_{10}(P)$ scale on the y axis. Each dot represents a variant. The red line indicates the conventional threshold for genome-wide significance ($P < $ 5x10^{-8}).
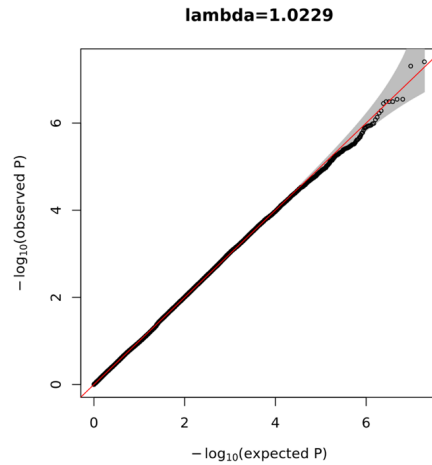
Figure 5.3. Q-Q plot for CAKUT mixed-ancestry GWAS.

Q-Q plot displaying the observed versus the expected $-\log_{10}(P)$ for each variant tested. The grey shaded area represents the 95% confidence interval of the null distribution.
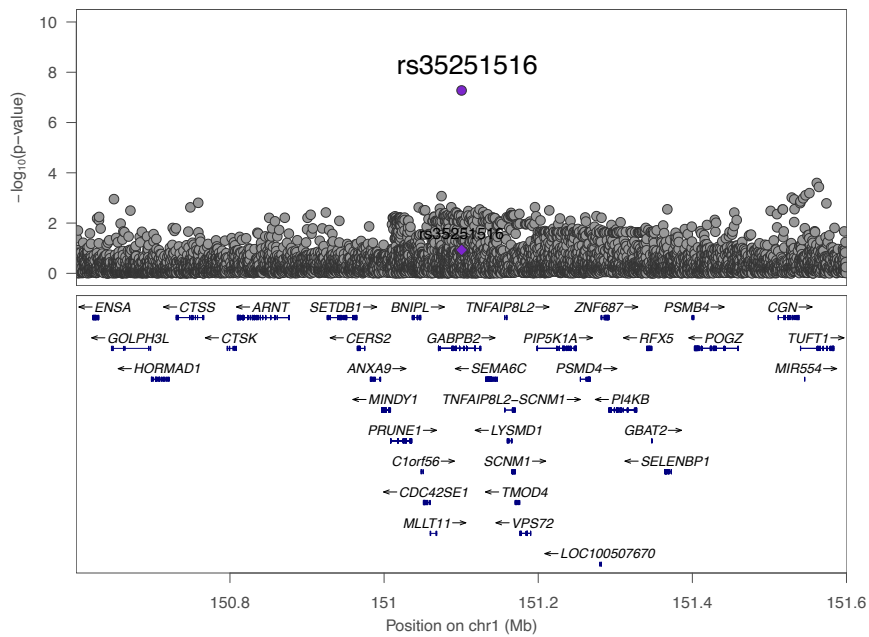
Figure 5.4. Regional association plot of 1q21.3 from CAKUT GWAS.

Each variant is represented by a dot. A single purple variant (rs35251516) is identified in a low complexity region suggesting a sequencing error. The remaining variants are coloured grey due to a lack of data on linkage disequilibrium (LD) from reference populations. Gene names in the region are listed against their chromosomal position (GRCh38).

181

The second variant, rs117473527 (chr6:102155812:G>C) at 6q16.3, reached a *P*-value of 3.93x10^{-8} (OR 3.17; 95% CI 2.10-4.78; MAF 0.02) (Figure 5.5). This intergenic variant is downstream of the gene *GRIK2* (glutamate ionotropic receptor kainite type subunit 2) which has previously been associated with neurodevelopmental disorders (Guzmán *et al.* 2017, Stolz *et al.* 2021) and urinary tract cancers (Inoue *et al.* 2017). It is also expressed in the murine ureteric bud at E10.5-11.5 (

Figure 5.6) when it invades the metanephric mesenchyme (Harding *et al.* 2011), making it biologically plausible that common variation affecting this gene may impact nephrogenesis. Replication in an independent cohort will be necessary to distinguish this association from statistical noise.
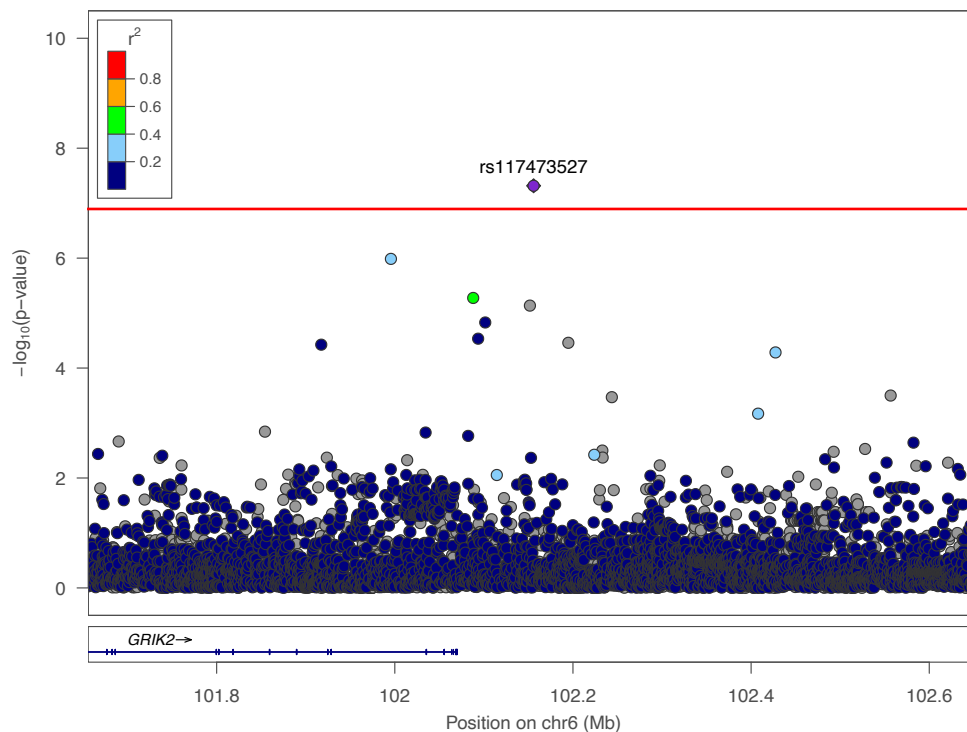
Figure 5.5. Regional association plot of 6q16.3 from CAKUT GWAS.

Each dot is a variant. Variants are coloured according to their linkage disequilibrium (r^2) with the lead variant. The red line indicates the genome-wide significance threshold of 5x10^{-8}. The gene *GRIK2* is shown against its chromosomal position (GRCh38).
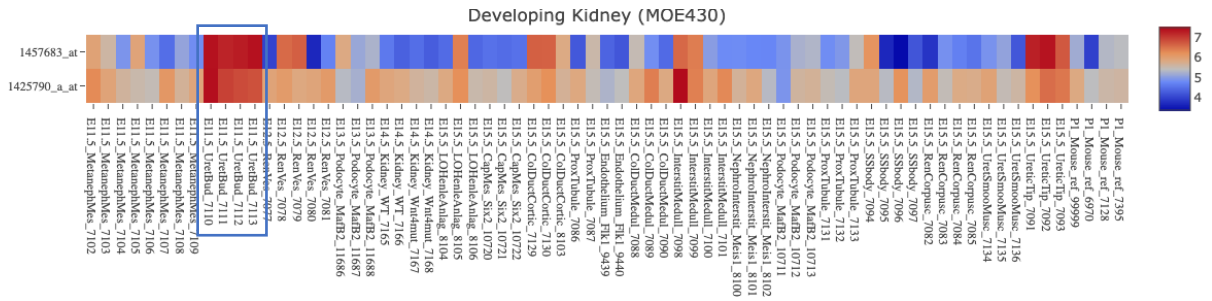
Figure 5.6. Heatmap of *Grik2* expression in the murine developing kidney.

High expression (dark red) is seen in the ureteric bud at E11.5 (blue rectangle). Data from GUDMAP (Harding *et al.* 2011).

Stratification by phenotype did not reveal any genome-wide significant associations for kidney anomalies (n=237), obstructive uropathy (n=177), VUR (n=174) or cystic renal dysplasia (n=112). The absence of significant association is likely the result of limited power related to both phenotypic heterogeneity and small sample size. With 813 CAKUT cases (assuming a disease prevalence of 2 in 1000) this experiment was powered to detect association of common variants (MAF > 5%) with OR > 2 or low-frequency variants (MAF > 1%) with OR > 4.5 (See Figure 2.3).

5.4.1.2   Gene and gene-set analysis

I have previously discussed how collapsing rare variants by gene can increase power to detect association and a similar approach can be used with common variants, collapsing them by gene and by biological pathway to look for enrichment across the genome. I therefore aggregated variants with MAF $\geq 0.1\%$ across 17,636 autosomal genes and 50 gene-sets, comparing the burden between cases and controls. Despite this combined analysis, no single gene reached exome-wide significance (*P*=0.05/17,636 genes=2.84×10^{-6}) with *VTN* (vitronectin) achieving the lowest *P*-value of 1.52×10^{-5} (Figure 5.7).

183

Aggregating variants across 50 hallmark gene-sets curated by MSigDB (Liberzon *et al.* 2015) similarly did not identify any enrichment in a specific biological pathway or process after correction for multiple testing (P=0.05/50=1x10^{-3}). Genes associated with DNA repair had the strongest evidence of association (141 genes; P=1.8x10^{-3}).
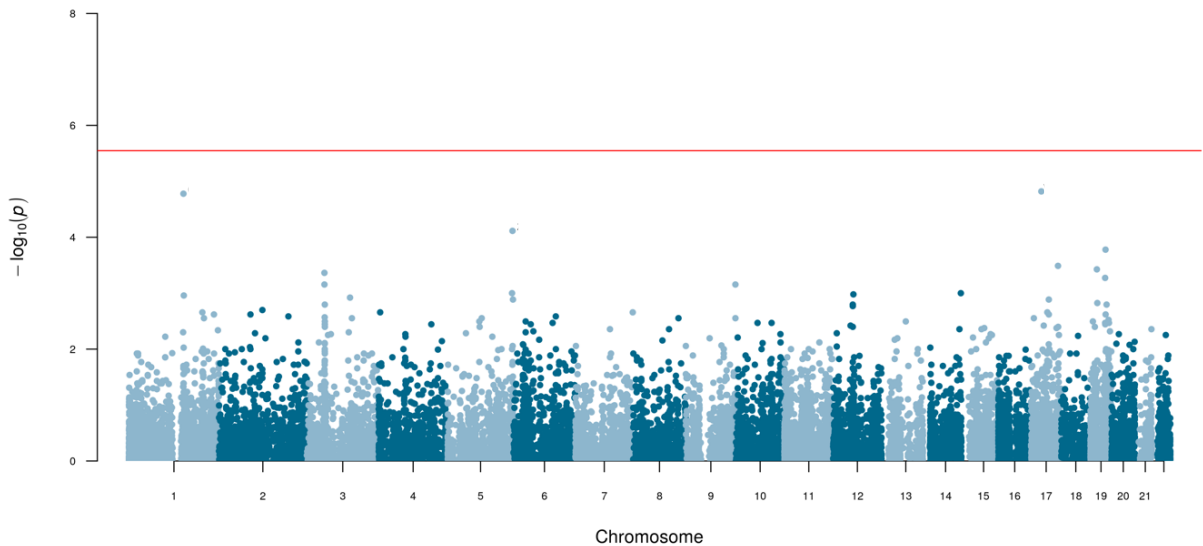
Figure 5.7. Manhattan plot of CAKUT gene-based common variant analysis.

SNVs and indels with MAF $\geq$ 0.1% were aggregated across 17,636 autosomal genes. Each dot represents a gene. The red line denotes the exome-wide significance threshold.

5.4.1.3   Heritability

Traditionally, twin studies have been used to estimate the heritability of human traits helping to differentiate genetic from environmental influences. More recently, the use of mixed-models has enabled genetic relatedness to be estimated from population-based sequencing data (Wainschtein *et al.* 2022). Using this approach, I estimated that the proportion of phenotypic variance in CAKUT explained by additive common and low-frequency variation was 0.23 (SE 0.11). While the large SE observed (because of the small sample size) means this estimate may not be accurate, low-

frequency variants with MAF between 1% and 5% accounted for > 75% of this estimated heritability (Figure 5.8). This suggests that there are likely to be a significant number of contributory low-frequency variants with effect sizes too small to be detected in this cohort.
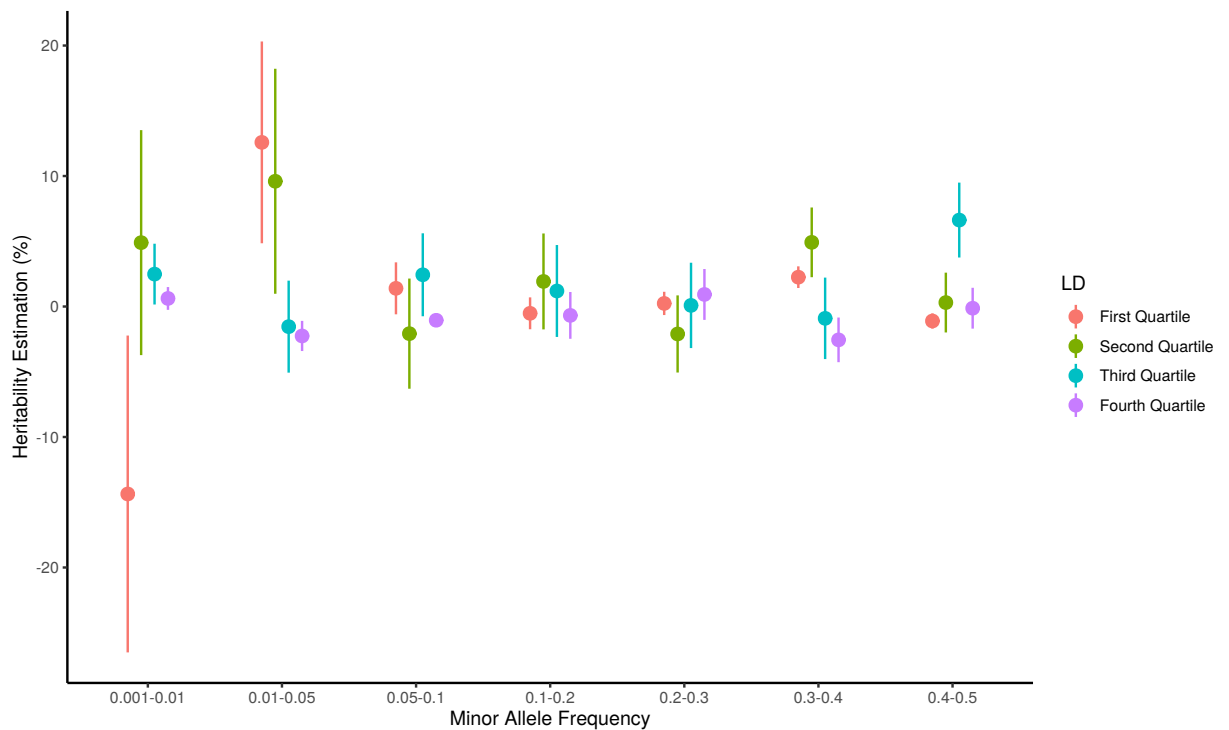
Figure 5.8. Partitioning of heritability by MAF in 623 CAKUT cases and 20,060 controls.

Narrow-sense heritability (h^2) is represented using the liability threshold model based on a population disease prevalence of 1 in 500. Bins indicate MAF of variants tested stratified by LD (First quartile, lowest LD; Fourth quartile, highest LD). Error bars indicate standard error (SE). h^2 follows a normal distribution and therefore unbiased estimates may be negative, as seen here in the 0.001-0.01 bin, particularly if the sample size is small (and the variance large).

5.4.2 Bladder exstrophy

Classic bladder exstrophy (CBE) is part of the bladder exstrophy-epispadias complex (MIM 600057), a spectrum of congenital genitourinary anomalies resulting

from a defect in abdominal midline development. CBE specifically refers to a protrusion of the urinary bladder through a defect on the infraumbilical abdomen, with diastasis of the pubic symphysis and divergent rectus abdominis muscles. It is nearly twice as common in males as females (Cervellione *et al.* 2015). Although a small number of multiply affected families have been reported (Reutter *et al.* 2003, 2007, Ludwig, Reutter, *et al.* 2009), CBE is usually sporadic, not generally associated with other malformations and its pathogenesis is poorly understood.

GWAS have previously associated CBE with common variation at 5q11.1 implicating the transcription factor *ISL1*, however definitive statistical evidence linking this region to *ISL1* is lacking (Draaken *et al.* 2015, Zhang *et al.* 2017). I therefore aimed to use the high resolution and sensitivity afforded by WGS to a) identify novel associations with CBE, b) replicate the association at 5q11.1, and c) perform fine-mapping of 5q11.1 to determine the likely causal variant(s) or genes.

5.4.2.1 Mixed-ancestry GWAS

I carried out a sequencing-based GWAS in 97 unrelated individuals with bladder exstrophy and 22,037 ancestry-matched controls testing 18,797,149 SNVs and indels with MAF $\geq$ 0.1% for association (Figure 5.9). Although only the top five principal components were included as covariates to prevent loss of power given the small sample size, the genomic inflation factor (λ) was slightly deflated at 0.92 (Figure 5.10).
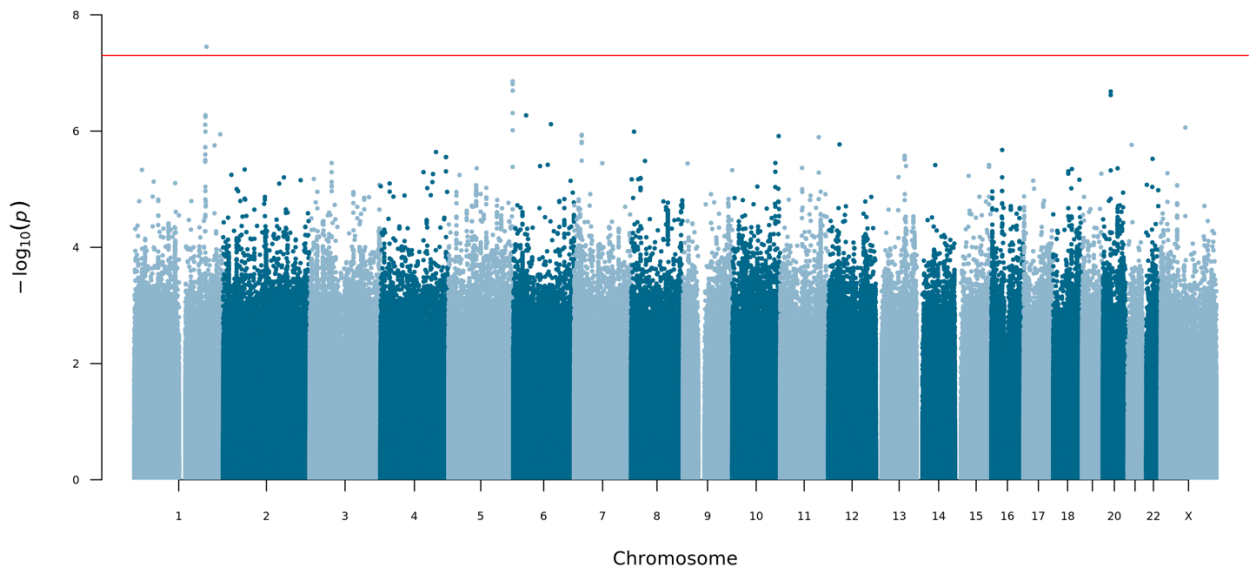
Figure 5.9. Manhattan plot of bladder exstrophy mixed ancestry GWAS.

A sequencing based GWAS was carried out in 97 unrelated bladder exstrophy cases and 22,037 controls for 18,797,149 variants with MAF $\geq$ 0.1%. Chromosomal position (GRCh38) is denoted along the x axis and strength of association using a $-\log_{10}(P)$ scale on the y axis. Each dot represents a variant. The red line indicates the conventional adjusted threshold for genome-wide significance ($P <$ 5×10^{-8}).
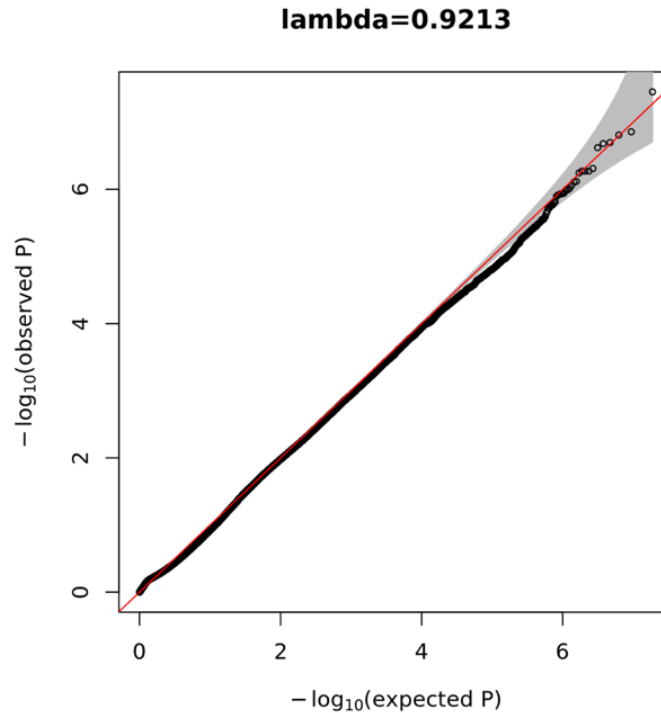
Figure 5.10. Q-Q plot of bladder exstrophy mixed-ancestry GWAS.

The observed versus the expected $-\log_{10}(P)$ for each variant tested is shown. The grey shaded area represents the 95% confidence interval of the null distribution.

One variant reached genome-wide significance: rs1571885276 at 1q32.1 (P=3.53x10^{-8}) however this variant falls in a low complexity region and review of the regional association plot revealed it was likely to be a sequencing artefact (Figure 5.11). Suggestive evidence of association (P < 1 x10^{-6}) was identified at 5q35.3, 20p11.22 and 1q32.1 (

Table 5.1). Of note, all three lead variants identified had MAF < 1% and large effect sizes meaning they would not have been detected using conventional genotyping and imputation.

188
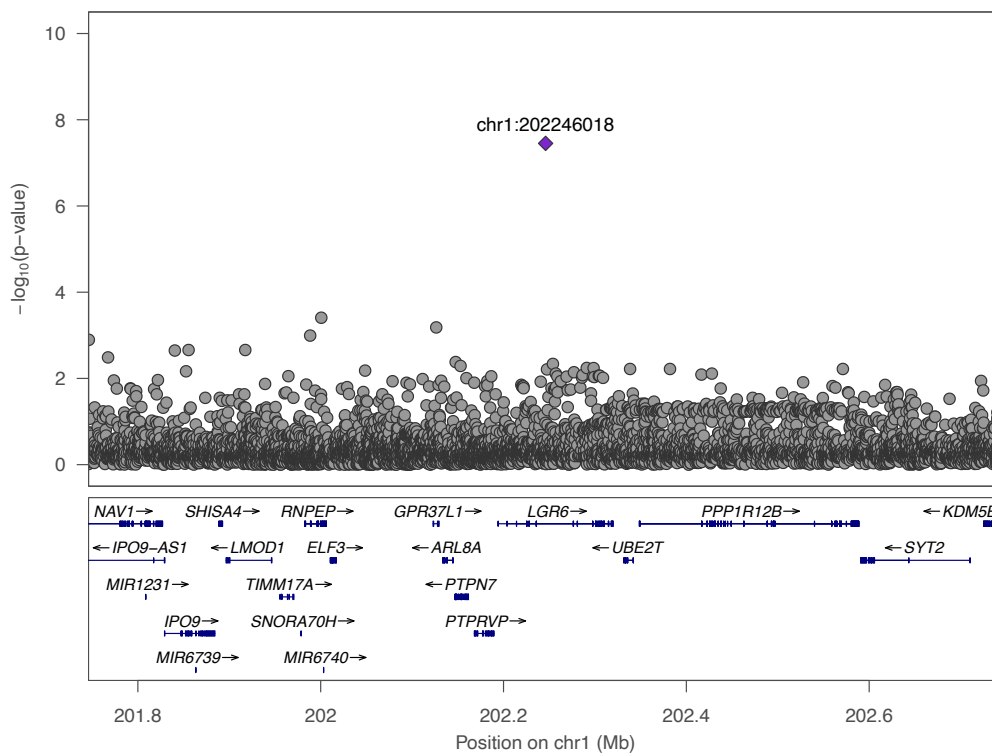
Figure 5.11. Regional association plot of 1q32.1 from bladder exstrophy GWAS.

Each dot represents a variant. rs1571885276 (purple diamond) is in a low complexity region and likely to be a sequencing artefact. The remaining variants are coloured grey due to a lack of data on linkage disequilibrium (LD) from reference populations. Gene names in the region are listed against their chromosomal position (GRCh38).

Table 5.1. Association statistics for bladder exstrophy mixed-ancestry GWAS.

The lead variant with the lowest *P* value at each locus is shown with genome-wide significance defined as P<5×10^{-8}. Genomic positions are with reference to GRCh38. CHR, chromosome; POS, position; OR, odds ratio; CI, confidence interval; EAF, effect allele frequency.

Lead variant	CHR:POS	Effect Allele	Closest gene	*P*	OR (95% CI)	Case EAF	Control EAF
rs147504710	chr5:179964061	C	*RNF130*	1.39×10^{-7}	6.72 (3.80-11.89)	0.067	0.011
rs6106456	chr20:22124633	G	*PAX1*	2.08×10^{-7}	6.72 (3.80-11.89)	0.067	0.010
rs573010426	chr1:199297046	A	*PTPRC*	5.35×10^{-7}	20.96 (8.98-48.92)	0.03	1.5×10^{-3}

5.4.2.2   Replication

Genotyping data was available for rs6106456 at 20p11.22 (Figure 5.12) from an independent cohort of 84 unrelated European patients with bladder exstrophy from Manchester, UK. A replication analysis was therefore performed using this cohort and 10,804 controls from the cancer arm of the 100KGP, excluding those with urinary tract or childhood malignancy. Although this variant did not replicate (Cochran-Armitage trend test *P*=0.52), the estimated power was insufficient to definitively confirm or refute association (power 0.67). Genotyping of rs147504710 and rs573010426 for replication analysis has yet to be performed.
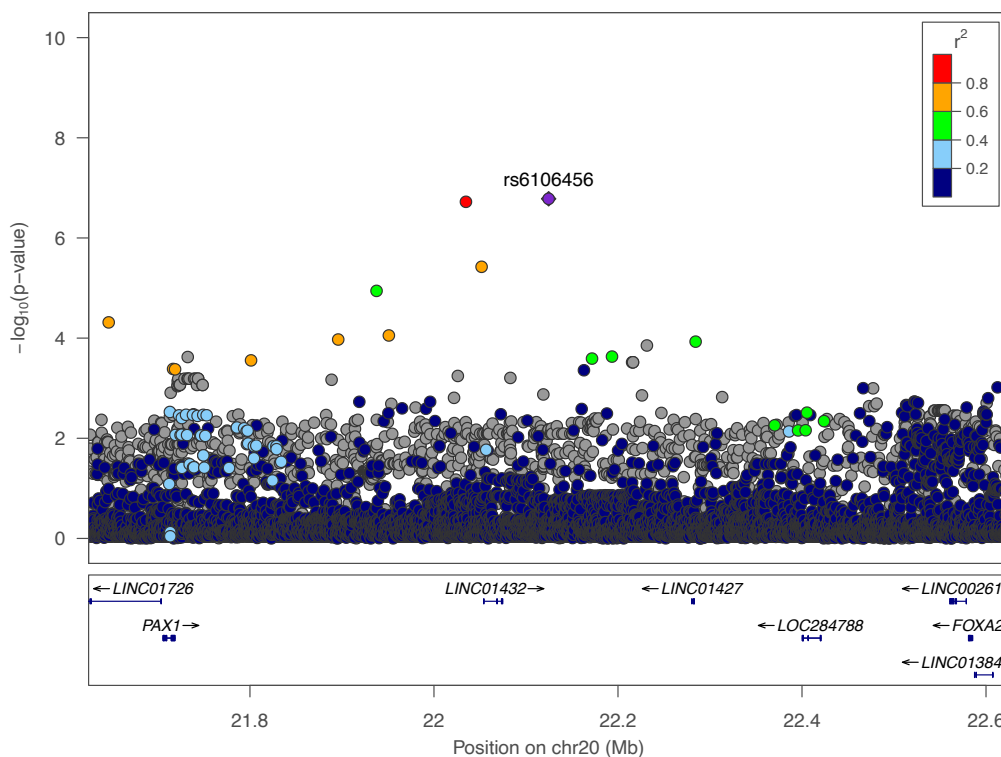
Figure 5.12. Regional association plot of 20p11.22 from bladder exstrophy mixed ancestry GWAS.

Each dot represents a variant. Variants are coloured according to their linkage disequilibrium (r^2) with the lead variant (purple diamond). Gene names in the region are listed against their chromosomal position (GRCh38).

5.4.2.3   Fine-mapping of 5q11.1 locus

The previously reported association at 5q11.1 (Draaken *et al.* 2015) 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 I repeated the analysis at this locus using all variants with MAC $\geq$ 3 (Figure 5.13). A rare intergenic indel rs550737686 (chr5:51494837:CCT>C) demonstrated the strongest evidence of association (*P*=2.35×10^{-5}; OR 6.11; 95% CI 3.11-12.03; MAF 0.008).

191
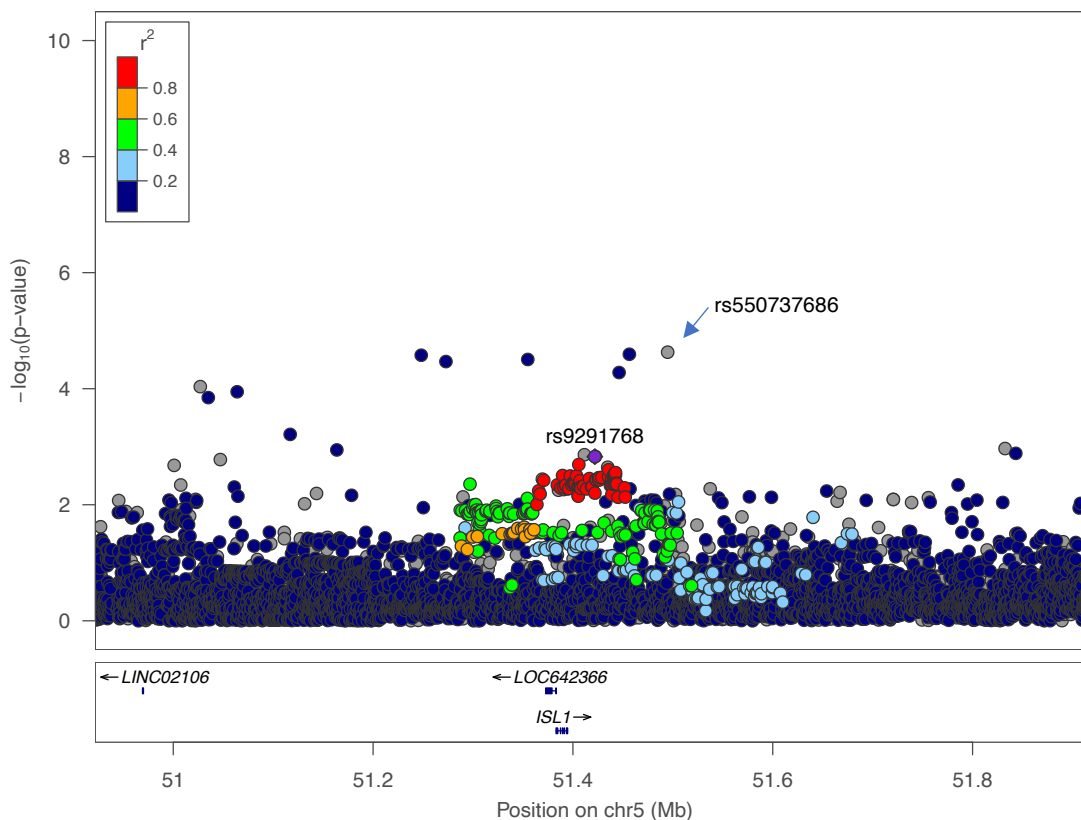
Figure 5.13. Regional association plot of 5q11.1 from bladder exstrophy mixed-ancestry GWAS.

Each dot represents a variant. Variants are coloured according to their linkage disequilibrium (r^2) with the lead variant (purple diamond). Gene names in the region are listed against their chromosomal position (GRCh38). rs9291768 was the lead variant identified in (Draaken *et al.* 2015).

Bayesian fine mapping was performed using PAINTOR (Kichaev *et al.* 2014) integrating the strength of association, LD patterns and functional annotations to derive the posterior probability of a variant being causal. The lead indel was excluded from the analysis as it was not present in the 1000 Genomes data used to calculate LD. Using functional annotations such as conservation (PhastCons elements), transcription factor binding clusters and cCREs (ENCODE Project Consortium *et al.* 2020), rs115201978 (chr5:51035061:G>A) was found to be likely causal with a high

posterior probability > 0.99. This rare variant (MAF 0.006) has an unusually high CADD score of 16.9 for an intergenic variant, reaching a *P*-value of 1.28×10^{-4} in the GWAS (OR 6.26; 95% CI 2.92-13.44) and is not in LD with the lead indel (r^2=0.1). Other than proximity to *ISL1* and localization within the same TAD, there was no gene expression or chromatin interaction data to definitively link these likely causal variants to a specific gene.

5.4.3 Posterior urethral valves

Posterior urethral valves (PUV) are the commonest cause of ESKD in children, affecting 1 in 4,000 male births (Thakkar *et al.* 2014, Brownlee *et al.* 2019) and resulting in congenital bladder outflow obstruction. It is a uniquely male disorder, with over a third of those affected developing ESKD (i.e. requirement for dialysis or kidney transplantation) before the age of 30 years (Sanna-Cherchi *et al.* 2009, Heikkilä *et al.* 2011) and is often associated with renal dysplasia, VUR and bladder dysfunction which are poor prognostic factors for renal survival (Sanna-Cherchi *et al.* 2009). Management involves endoscopic valve ablation to relieve the obstruction; however, the majority of affected children have long-term sequelae related to ongoing bladder dysfunction (DeFoor *et al.* 2008).

The pathogenesis of PUV is poorly understood (Krishnan *et al.* 2006). Although usually sporadic, familial clustering and twin studies suggest an underlying genetic component, although Mendelian inheritance appears rare (Weber *et al.* 2005, Schreuder *et al.* 2008, Chiaramonte *et al.* 2016, Frese *et al.* 2019). Pathogenic heterozygous variants in *BNC2* have been reported in two families (in both males and females) with anatomical congenital bladder outflow obstruction (urethral stenosis, PUV, and pathological voiding) and 2/697 further individuals with lower urinary tract obstruction (Kolvenbach *et al.* 2019). In addition, exon skip-inducing variants in the X-linked gene *FLNA* have been identified in two unrelated males with cardiac anomalies, one of whom had PUV and the other a urethral stricture (Wade *et al.* 2021). However, a definitive monogenic aetiology for isolated PUV has not yet

been identified and the vast majority of affected individuals remain genetically unsolved.

Case reports of chromosomal abnormalities resulting in PUV as part of a wider syndrome (Houcinat *et al.* 2015, Tong *et al.* 2017, Demirkan 2021) and microarray-based studies linking rare CNVs with isolated PUV (Caruana *et al.* 2015, Boghossian *et al.* 2016, Faure *et al.* 2016, Verbitsky *et al.* 2019, Schierbaum *et al.* 2021) suggest structural variation, and in particular duplications (Verbitsky *et al.* 2019, Schierbaum *et al.* 2021), may play a role (although many of these lacked a suitable control population), but no recurrent CNVs have demonstrated a consistent association with PUV. The observation that PUV does not usually follow a classical Mendelian inheritance pattern indicates that the underlying genetic architecture is likely to be complex.

I previously described an increased burden of rare structural variation affecting *cis*-regulatory elements in individuals with PUV suggesting that disruption to gene regulatory networks may be important in this disorder. To determine whether common and low-frequency variation might also impact these networks I performed a sequencing-based GWAS in 132 unrelated male cases and 23,727 ancestry-matched controls.

Table 5.2 details the clinical and demographic details of the cohort.

Table 5.2. Clinical characteristics and genetic ancestry of PUV cohort and ancestry-matched controls.

PUV, posterior urethral valves; PCA, principal components analysis; EUR, European; SAS, South Asian; AFR, African; AMR, Latino/Admixed American; VUR, vesico-ureteral reflux; UTI, urinary tract infection; ESKD, end-stage kidney disease.

		PUV (n=132)	Controls (n=23,727)
Median age (range)		13 (2-66)	
Males (%)		132 (100)	10,425 (43.9)
PCA determined ancestry			
	EUR (%)	89 (67.4)	19,418 (81.8)
	SAS (%)	18 (13.6)	2847 (12.0)
	AFR (%)	11 (8.3)	449 (1.9)
	AMR (%)	0 (0)	7 (0.03)
	Admixed (%)	14 (10.6)	1006 (4.2)
Additional renal/urinary phenotypes			
	Hydronephrosis (%)	56 (42.4)	
	Bladder abnormality (%)	32 (24.2)	
	Hydroureter (%)	30 (22.7)	
	VUR (%)	27 (20.5)	
	Renal dysplasia (%)	16 (12.1)	
	Hypertension (%)	11 (8.3)	
	Renal agenesis (%)	8 (6.1)	
	Recurrent UTIs (%)	5 (3.8)	
	Renal hypoplasia (%)	4 (3.0)	
	Renal duplication (%)	2 (1.5)	
Extrarenal manifestations (%)		35 (26.5)	
	Cardiac anomaly (%)	4 (3.0)	
	Neurodevelopmental disorder (%)	7 (5.3)	
Family history (%)		5 (3.8)	
End-stage kidney disease (%)		23 (17.4)	
Median age ESKD (range)		14 (0-39)	

5.4.3.1   Mixed-ancestry GWAS

19,651,224 SNVs and indels with MAF $\geq$ 0.1% were tested for association using SAIGE in this mixed-ancestry cohort. Statistically significant (P<5x10^{-8}) association was detected at two loci (Figure 5.14 and Table 5.3). At 12q24.21, the lead intergenic variant (rs10774740) was common (MAF 0.37) and reached P=7.81x10^{-12} (OR 0.40; 95% CI 0.31-0.52). A rare (MAF 0.007) variant (rs144171242) at 6p21.1, located in an intron of *PTK7,* was also significant at P=2.02x10^{-8} (OR 7.20; 95% CI 4.08-12.70). The meanDP and meanGQ at these sites were 33.43 and 133.28, respectively, for rs10774740 and 29.34 and 75.59 for rs144171242. The genomic inflation factor (λ) of 1.04 confirmed population stratification was well controlled in this diverse ancestry cohort, although this may represent an underestimate given the low power of the cohort (Figure 5.15). Conditional analysis did not identify secondary independent signals at either locus and epistasis was not detected between the two lead variants (P=0.10).

Table 5.3. Association statistics for significant genome-wide loci in 132 PUV cases and 23,727 ancestry-matched controls.

The lead variant with the lowest P value at each locus is shown with genome-wide significance defined as P<5x10^{-8}. Genomic positions are with reference to GRCh38. CHR, chromosome; POS, position; OR, odds ratio; CI, confidence interval; EAF, effect allele frequency.

Lead variant	CHR:POS	Effect Allele	Closest gene	P	OR (95% CI)	Case EAF	Control EAF
rs10774740	chr12:114228397	T	*TBX5*	7.81x10^{-12}	0.40 (0.31-0.52)	0.19	0.37
rs144171242	chr6:43120356	G	*PTK7*	2.02x10^{-8}	7.20 (4.08-12.70)	0.05	0.007
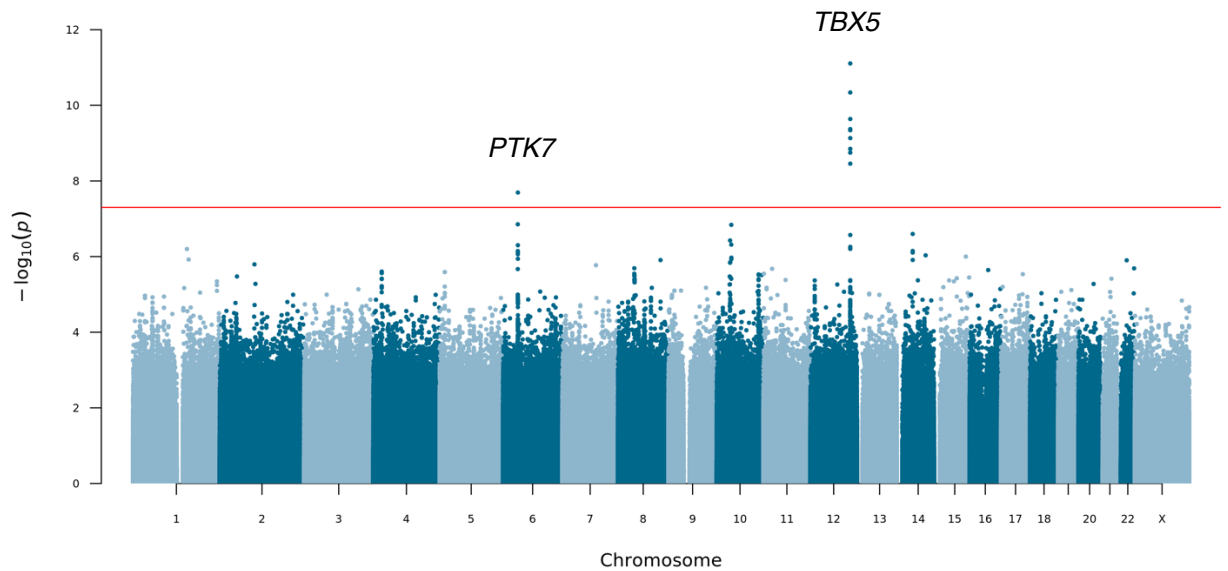
Figure 5.14. Manhattan plot for PUV mixed-ancestry GWAS.

A GWAS was carried out in 132 PUV cases and 23,727 controls for 19,651,224 variants with MAF $\geq$ 0.1%. Chromosomal position (GRCh38) is denoted along the x axis and strength of association using a $-\log_{10}(P)$ scale on the y axis. Each dot represents a variant. The red line indicates the conventional adjusted threshold for genome-wide significance ($P < 5 \times 10^{-8}$). The genes in closest proximity to the lead variant at significant loci are shown.
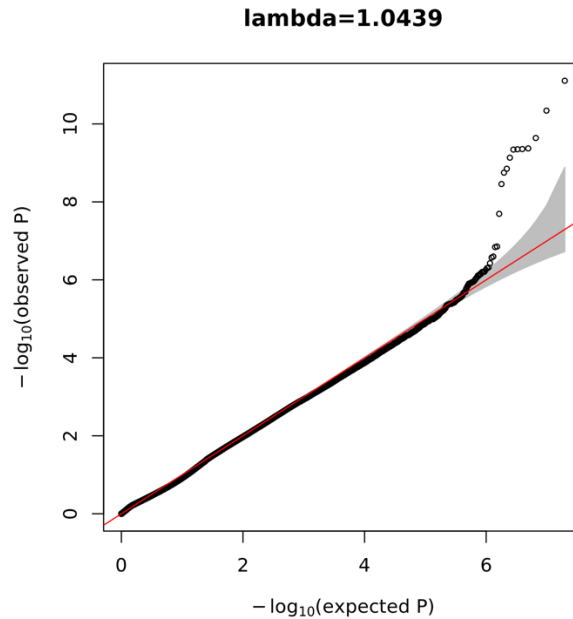
Figure 5.15. Q-Q plot for PUV mixed-ancestry GWAS.

Q-Q plot displaying the observed versus the expected $-\log_{10}(P)$ for each variant tested. The grey shaded area represents the 95% confidence interval of the null distribution.

Due to the uniquely male nature of PUV, I also conducted a sex-specific analysis using only male controls (n=10,425) to determine whether any additional signals could be detected by removing females with potentially undetected genitourinary phenotypes. Both lead variants showed stronger evidence of association: rs10774740 at 12q24.21 (P=7.08x10^{-12}; OR 0.40; 95% CI 0.31-0.52) and rs144171242 at 6p21.1 (P=1.79x10^{-8}; OR 7.40; 95% CI 4.14-13.22). No significant associations on chromosome X were identified.

Gene and gene-set analyses were carried out to assess the joint effect of common and low-frequency variants and identify potential functional pathways associated with PUV, however, no genes or pathways reached statistical significance after correction for multiple testing.

5.4.3.2   Replication

I next carried out a replication study in an independent European cohort consisting of 398 individuals with PUV: 336 from Poland and Germany, recruited through the CaRE for LUTO (Cause and Risk Evaluation for Lower Urinary Tract Obstruction) Study, and 62 from Manchester, UK. 10,804 European individuals recruited to the cancer arm of the 100KGP were used as controls. The UK PUV patients and the 100KGP cancer control cohort had not been included in the discovery analyses. The lead variants at the top four loci with $P<5x10^{-7}$ were tested for replication. Association at both genome-wide significant lead variants was replicated although with smaller effect sizes (Table 5.4). Two further loci with suggestive evidence of association (10q11.2 and 14q21.1) did not replicate.

Table 5.4. Association statistics for the replication study.

The lead variants at the top four loci with $P < 5x10^{-7}$ were genotyped in an independent European cohort of 398 PUV cases and 10,804 controls. P values in the replication cohort were calculated using a one-sided Cochran Armitage Trend test. The genotyping assay failed for variant rs1471950716 (NA) and therefore the next lead variant rs137855548 was used instead. OR, odds ratio; CI, 95% confidence interval.

Locus	Lead variant	Effect Allele	Discovery P value	Replication P value	Discovery OR (95% CI)	Replication OR (95% CI)
12q24.21	rs10774740	T	$7.81x10^{-12}$	$1.9x10^{-3}$	0.40 (0.31-0.52)	0.78 (0.67-0.91)
6p.21.1	rs144171242	G	$2.02x10^{-8}$	$4.5x10^{-3}$	7.20 (4.08-12.70)	2.17 (1.25-3.76)
10q11.21	rs1471950716	A	$1.45x10^{-7}$	NA	3.88 (2.42-6.22)	NA
	rs137855548	G	$1.46x10^{-6}$	0.5471	3.94 (2.36-6.56)	0.84 (0.48-1.47)
14q21.1	rs199975325	G	$2.52x10^{-7}$	0.9636	5.68 (3.22-9.99)	1.02 (0.52-1.98)

5.4.3.3   European-only GWAS

To ascertain whether the observed associations were being driven by a specific ancestry group, I next repeated the GWAS using a subgroup of genetically defined European individuals (88 cases and 17,993 controls) and 16,938,500 variants with MAF $\geq$ 0.1% (Figure 5.16). Although only the top five principal components were included as covariates to prevent loss of power with the small sample size, the genomic inflation factor was deflated at 0.90 (Figure 5.17). This was attributed to the inclusion of rare variants (MAF < 1%) in combination with a small sample size which renders the expected null distribution inaccurate when minor allele counts are low or zero (Figure 5.18).
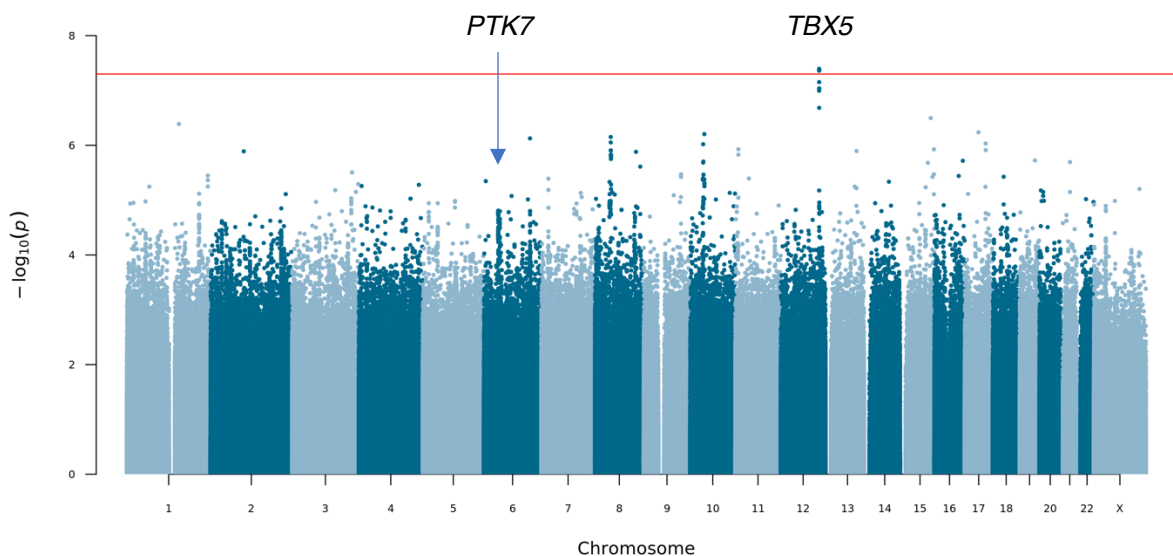
Figure 5.16. Manhattan plot for PUV European-only GWAS.

A GWAS was carried out in 88 cases and 17,993 controls for 16,938,500 variants with MAF $\geq$ 0.1%. The red line indicates the conventional threshold for genome-wide significance ($P < 5\times10^{-8}$). The two genome-wide significant loci from the mixed ancestry GWAS are labelled.
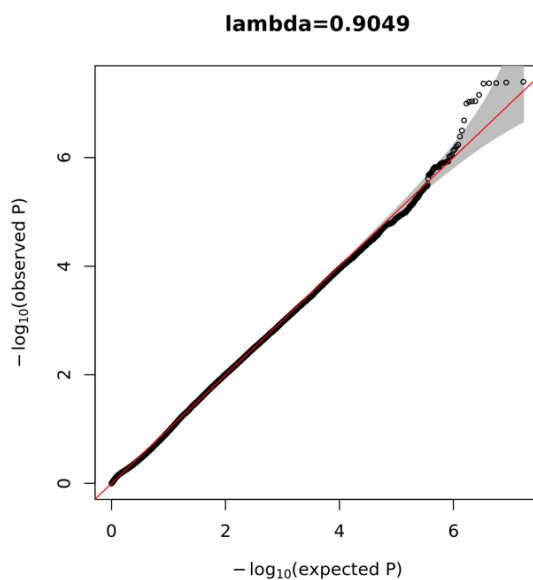
Figure 5.17. Q-Q plot for the PUV European-only GWAS.

The observed versus the expected $-\log_{10}(P)$ for each variant tested. The grey shaded area represents the 95% confidence interval of the null distribution.
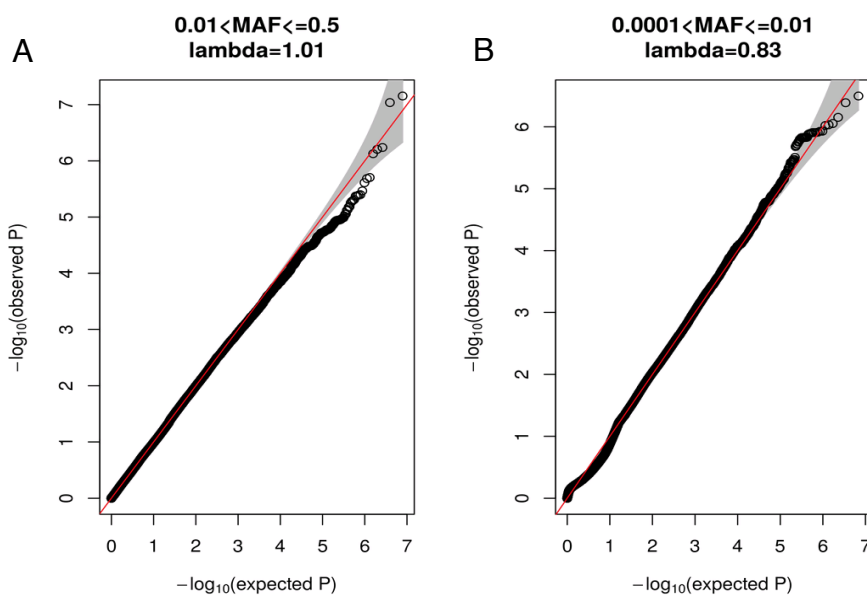
Figure 5.18. Q-Q plots for PUV European GWAS stratified by MAF.

A, MAF between 1% and 50%. **B**, MAF between 0.01% and 1%. The observed versus the expected $-\log_{10}(P)$ for each variant tested is shown. The grey shaded area represents the 95% confidence interval of the null distribution. MAF, minor allele frequency.

The 12q24.21 locus remained genome-wide significant, however the lead variant (rs2555009) in the region showed weaker association (P=4.02x10^{-8}; OR 0.43; 95% CI 0.12-0.73) than rs10774740, the lead variant in the mixed ancestry analysis (Table 5.5). Interestingly the two variants were not in strong linkage disequilibrium in the European population (EUR LD; r^2=0.54). The lead variant at 6p21.1 from the mixed ancestry analysis did not reach genome-wide significance in the European-only study (rs144171242; P=3.60x10^{-5}; OR 5.90; 95% CI 2.88-12.11) suggesting that this signal is being driven partly by non-Europeans or that inclusion of non-Europeans boosts power through increased sample size.

Table 5.5. Comparison of PUV mixed ancestry and European GWAS association statistics.

The lead variants at the top four loci with $P < 5$x10^{-7} are shown. OR, odds ratio; CI, 95% confidence interval.

Lead variant	Effect Allele	Mixed ancestry P value	European P value	Mixed ancestry OR (95% CI)	European OR (95% CI)
rs10774740	T	7.81x10^{-12}	7.03x10^{-8}	0.40 (0.31-0.52)	0.42 (0.10-0.73)
rs144171242	G	2.02x10^{-8}	3.60x10^{-5}	7.20 (4.08-12.70)	5.90 (2.88-12.11)
rs1471950716	A	1.45x10^{-7}	6.24x10^{-5}	3.88 (2.42-6.22)	4.44 (2.68-7.36)
rs199975325	G	2.52x10^{-7}	1.57x10^{-5}	5.68 (3.22-9.99)	4.43 (2.16-9.06)

P values and effect sizes were strongly correlated between the mixed ancestry and European-only GWAS (Figure 5.19) demonstrating that inclusion of individuals from diverse backgrounds to increase sample size can be an effective way to boost power and discover new disease loci, even in a small cohort.
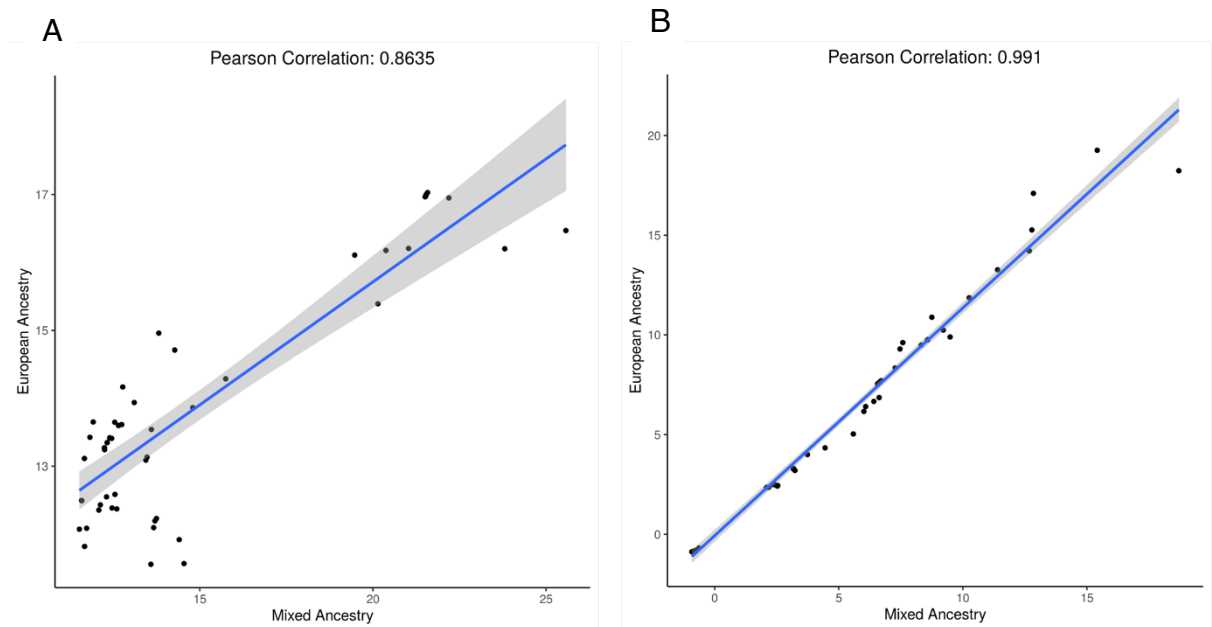
Figure 5.19. Correlation between mixed ancestry and European PUV GWAS.

Comparison of **A**, -log$_{10}$(*P)* and **B**, BETA from the mixed ancestry and European-only ancestry GWAS. All variants with *P* < 10^{-5} in both cohorts are shown. The shaded grey area represents the 95% confidence interval.

5.4.3.4   Ancestry-specific comparison

As the numbers of African, South Asian, and admixed ancestry individuals were too small to reliably carry out subgroup association analyses and subsequent meta-analysis, I instead compared ancestry-specific allele frequencies, effect sizes and directions for each lead variant at 12q24.21 and 6p21.1. rs10774740 (T) had a much

higher allele frequency in individuals of African ancestry (MAF 0.74) compared with European (MAF 0.37) and South Asian (MAF 0.35) populations; however, the effect size and direction of the associations were similar between the groups (Figure 5.20A). rs144171242 (G) was present at a lower allele frequency in South Asian (MAF 0.002) compared with European (MAF 0.008) individuals and was not seen in the African ancestry group. The effect size of this rare variant was higher in the South Asian than European population (Figure 5.20B), which may explain why it only reached genome-wide significance after inclusion of South Asian individuals.
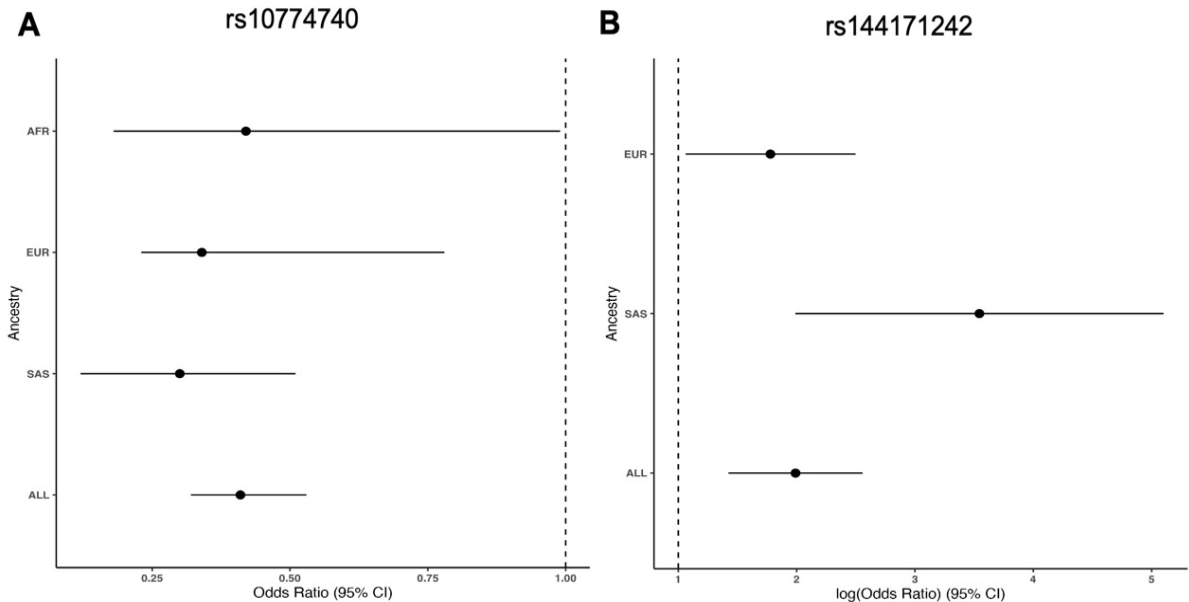
Figure 5.20. Comparison of ancestry-specific odds ratios from PUV GWAS.

GWAS per-ancestry odds ratios for **A**, rs10774740 (T) at 12q24.21 and **B**, rs144171242 (G) at 6p21.1. Error bars represent 95% confidence intervals. The lead variant in B was not present in individuals with African ancestry. AFR, African ancestry (11 cases; 449 controls); EUR, European ancestry (89 cases; 19,418 controls); SAS, South Asian ancestry (18 cases; 2,847 controls); ALL, mixed-ancestry cohort (132 cases; 23,727 controls).

Finally, comparison with population allele frequencies from gnomAD (Karczewski *et al.* 2020) demonstrated that although there is large variation in the allele frequency of rs10774740 between ancestries in non-Europeans this is away from, not towards, the case allele frequency and confirms that the detected associations are not being driven by differences in allele frequency between populations (Figure 5.21).
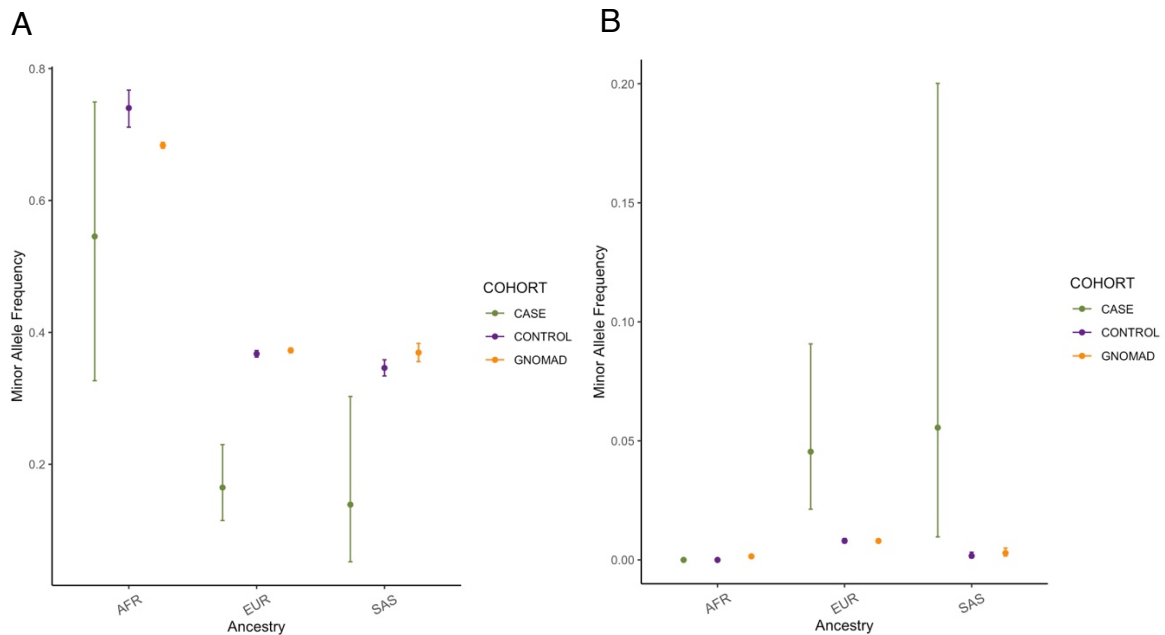
Figure 5.21. Comparison of ancestry-specific allele frequencies from PUV GWAS.

Ancestry-specific minor allele frequencies for **A**, rs10774740 (T) at 12q24.21 and **B**, rs144171242 (G) at 6p21.1. Error bars represent 95% confidence intervals. The lead variant in B was not identified in individuals with African ancestry in this study. AFR, African ancestry (11 cases; 449 controls); EUR, European ancestry (89 cases; 19,418 controls); SAS, South Asian ancestry (18 cases; 2,847 controls). Population allele frequencies from gnomAD (version 3.1.2).

5.4.3.5 Fine-mapping of 12q24.21 and 6p21.1

WGS enables further interrogation of loci of interest at high resolution. I therefore repeated the mixed ancestry analysis at both genome-wide significant loci using all variants with MAC ≥ 3, to determine whether additional ultra-rare variants might be

205

underlying the observed association signals. Both rs10774740 at 12q24.21 and rs144171242 at 6p21.1 remained most strongly associated, suggesting they are likely to be causal. Comparison of the different LD patterns seen across African, European, and South Asian population groups at these loci using 1000 Genomes data (Phase 3) (1000 Genomes Project Consortium *et al.* 2015) demonstrated how a combined ancestry approach can leverage differences in LD to improve the fine mapping of causal variants (Figure 5.22) because the set of alleles in strong LD with the lead marker in all populations included is smaller than the set of alleles in any one of the populations.

I next applied the Bayesian fine-mapping tool PAINTOR (Kichaev *et al.* 2014) to determine the posterior probability of variants at these loci being causal. Using this alternative statistical approach, both lead variants were identified as having a very high probability of being causal under the assumption of one causal variant at each locus: rs10774740 (posterior probability [PP] with no annotations 0.77, PP with annotations > 0.99) and rs144171242 (PP with no annotations 0.83, PP with annotations > 0.99). Conservation and ChIP-seq transcription factor binding clusters had the largest impact on posterior probabilities at 12q24.21 and 6p21.1, respectively. Using European, South Asian, or African (for 12q24.21 only) 1000 Genomes data to calculate LD patterns did not alter the posterior probabilities of the lead variants.

Validation of the lead variants using statistical fine mapping illustrates how the increased sensitivity and improved resolution of WGS compared with genotyping arrays may permit the direct identification of underlying causal variants, particularly in the context of examining rarer variants and non-European populations for which imputation performance may be limiting (Höglund *et al.* 2019, Peterson *et al.* 2019).
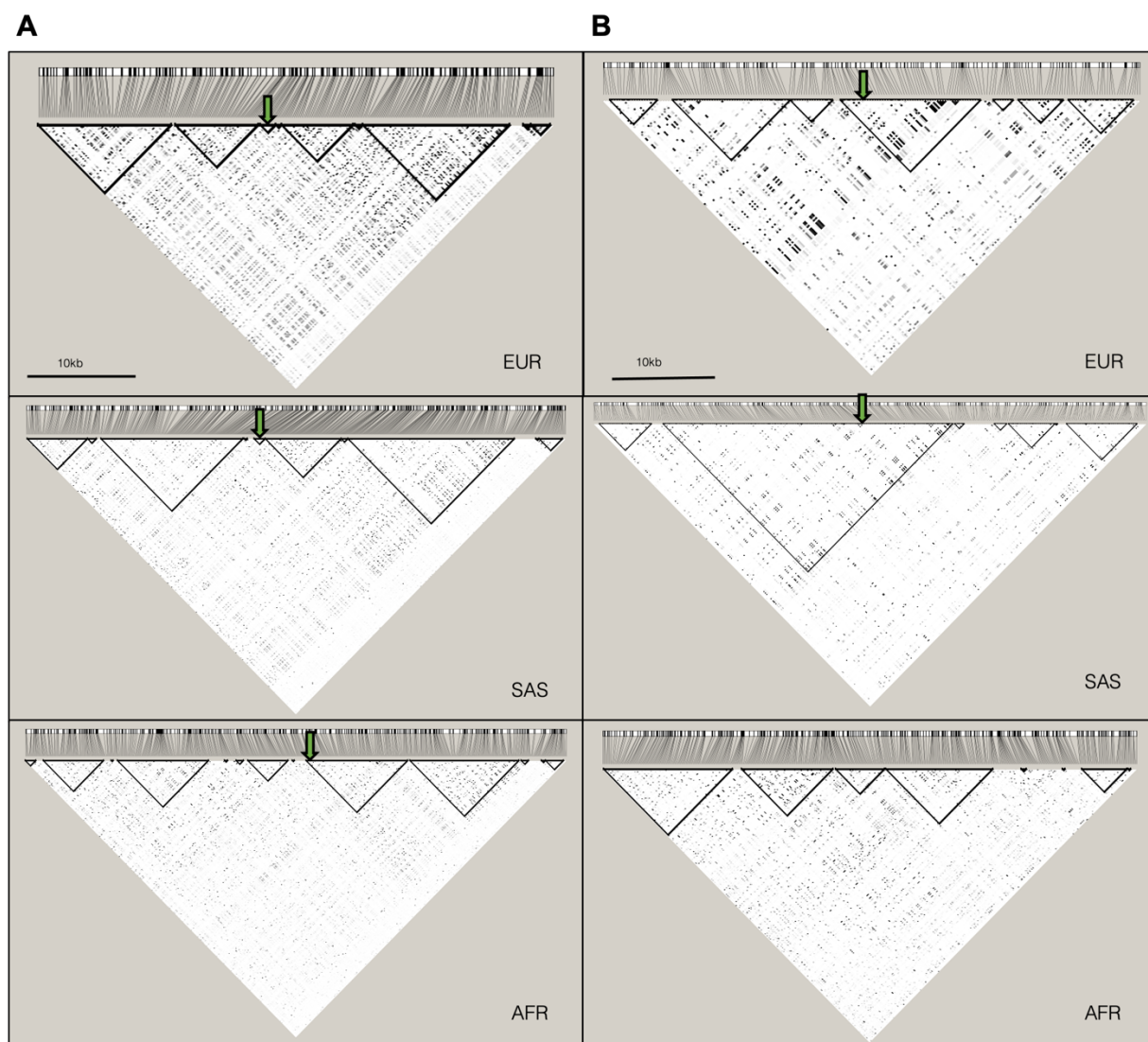
Figure 5.22. Linkage disequilibrium (LD) for reference populations in the 1000 Genomes Project.

LD plots for 503 European (EUR), 489 South Asian (SAS) and 661 African (AFR) ancestry individuals from the 1000 Genomes Project (Phase 3). Haploview (v4.2) was used to compute pairwise LD statistics (r^2) between variants for each population. The darker the shading, the higher the LD between variants. Black outlined triangles indicate haploblocks. **A,** LD plot for chr12:114,641,202-114,691,202 (GRCh37) with the position of the lead variant rs10774740 represented by a green arrow; **B**, LD plot for chr6:43,063,094-43,113,094 (GRCh37) with the position of the lead variant rs144171242 represented by a green arrow. rs144171242 was not seen in the AFR population group.

207

5.5   Summary

- Suggestive association was detected at 6q16.3 for CAKUT and 1q32.1 and 20p11.22 for bladder exstrophy, however meta-analyses with additional cohorts are needed before robust conclusions can be drawn.

- A significant proportion of the estimated heritability of CAKUT is attributable to low-frequency variants ($1\% \leq MAF < 5\%$).

- The previously reported common variant association at 5q11.1 was replicated in our bladder exstrophy cohort. Fine mapping of this locus prioritized additional rare, non-coding variants for functional follow-up.

- Two novel loci, 12q24.21 and 6p21.1, were significantly associated with PUV, both of which replicated in an independent cohort. The lead variants at each locus were predicted to be likely causal using a Bayesian fine-mapping approach.

5.6   Discussion

In this chapter I have used a mixed-ancestry sequencing-based GWAS approach to identify the first robust genetic associations for PUV, replicate and fine-map a locus previously associated with bladder exstrophy and estimate the contribution of common and low-frequency variation to CAKUT. The lack of genomic inflation observed and replication of results in an independent cohort (for PUV) illustrate how a well-controlled mixed-ancestry WGS association study can increase power for disease locus discovery even in a small cohort. Furthermore, the association of rare variants ($MAF < 1\%$) with large effects in both the PUV and bladder exstrophy cohorts would have been missed using conventional array-based GWAS, highlighting the advantage of using WGS for single-variant association studies.

5.6.1 Low-frequency variation contributes to the phenotypic variance of CAKUT

This is the first time common and low-frequency variant associations have been investigated using a sequencing-based approach in CAKUT. Although significant association was detected at 6q16.3 downstream of *GRIK2*, further replication in an independent cohort is necessary before any firm conclusions can be made. Importantly however, I demonstrate that low-frequency variation ($1\% \leq MAF < 5\%$) contributes to a significant proportion of the phenotypic variance associated with renal tract malformations providing evidence for the first time of a polygenic basis for this complex disorder.

While twin studies are seen as the 'gold standard' of heritability estimation, particularly when trying to disentangle genetic and *in utero* influences, very few twin studies examining the heritability of CAKUT have been carried out, presumably due to the relatively low number of reported familial cases. Those that have been performed have focused on phenotypes such as VUR, lower urinary tract obstruction or bladder exstrophy where high concordance rates in monozygotic twins support a genetic basis for these conditions (Kaefer *et al.* 2000, Reutter *et al.* 2007, Frese *et al.* 2019). More recently, heritability estimation using WGS or SNP-array data from unrelated cohorts has been used to provide novel insights into how much impact genetic variation has on a particular complex trait which in turn can help to prioritize further genetic studies (Wainschtein *et al.* 2022). This approach has been used to estimate that common variation accounts for ~15% of phenotypic variance in sporadic VUR, although the reliability of this estimate is not known (Verbitsky *et al.* 2021).

My analysis estimates that ~23% of the phenotypic variance of CAKUT can be explained by common and low-frequency additive genetic variance, with much of this heritability attributed to variants with MAF between 1% and 5%. The large standard error observed however means this figure should be interpreted with caution. Large sample sizes are needed to generate accurate estimates of heritability and these

estimates can be less reliable when disease prevalence is low. The absence of significant association in the CAKUT GWAS indicates that these low-frequency variants are likely to have effect sizes below the threshold that can be detected in a cohort of this size (OR < 4.5) and suggests that larger and better powered studies will be needed to uncover these 'missing' variants in the future.

5.6.2 VUR

Common variant associations have previously been assessed using an array-based approach in a large cohort of patients with VUR. A meta-analysis of 1,395 unrelated VUR patients and 5,366 controls (all European ancestry) demonstrated no genome-wide significant association using a conventional additive model (Verbitsky *et al.* 2021), consistent with the lack of association seen in the mixed-ancestry VUR analysis performed here. Generally, an additive model has reasonable power to detect both additive and dominant effects but may be underpowered to detect some recessive effects. In view of this and based on the observation that VUR is more prevalent in females, Verbitsky *et al.* (2021) performed additional sex-specific analysis using both dominant and recessive models identifying three significant (2p15, 6p12.1, 6q14.1) and five suggestive loci. None of the lead variants at these eight loci replicated in our VUR cohort of 174 patients, although there was sufficient power to refute association at only five of these eight loci using an additive model and I did not investigate sex-specific effects.

The concept of sexual dimorphism of gene expression driven by gene by sex (GxS) interactions is interesting. While the autosomal genome has been assumed to be largely similar in males and females, there are now increasing reports of sex-specific associations with complex traits (Ober *et al.* 2008, Graham *et al.* 2019, Bernabeu *et al.* 2020), renal cell carcinoma (Laskar *et al.* 2019) and autosomal dominant kidney disease (Gale *et al.* 2010, Athanasiou *et al.* 2011). Furthermore, there is some evidence for small but ubiquitous differential gene expression between the sexes, mostly mediated through hormone-related transcription factor binding and epigenetic

changes (Oliva *et al.* 2020). It would therefore be an interesting experiment to perform sex-specific analyses in CAKUT, and specifically bladder exstrophy which has a higher prevalence in boys, although with the caveat that stratifying by sex would severely limit power.

5.6.3   Fine-mapping of 5q11.1 in bladder exstrophy

Our data replicated the previously identified association with 5q11.1 in bladder exstrophy patients (Draaken *et al.* 2015) but failed to detect any additional genome-wide significant loci, although this was not unexpected given the small sample size (97 cases). The LIM homeodomain transcription factor *ISL1* (Islet1) has been proposed as a candidate gene in this region with supportive evidence from mouse models which show strong *Isl1* expression in the genital tubercle at the appropriate developmental time point (Draaken *et al.* 2015, Zhang *et al.* 2017). Furthermore, mice with conditional deletion of *Isl1* develop kidney agenesis or blind ureters (Kaku *et al.* 2013). Fine mapping of this locus using high-resolution WGS in combination with a Bayesian approach identified additional rare variants with stronger evidence of association than the previously reported common variants, however functional annotation with publicly available data could not definitively link these likely causal variants with *ISL1,* other than being in the same TAD. Functional studies are therefore needed to establish whether these variants affect the expression of *ISL1* or indeed disrupt long-range regulatory interactions with a more distant gene.

5.6.4   Two novel loci associated with PUV

Using a mixed-ancestry WGS approach I identified two novel genome-wide significant loci associated with PUV and determined the lead variants at each locus as likely causal. The common variant rs10774740 (T) at 12q24.21 (MAF 0.37) demonstrated a significant protective effect, highlighting the fact that common variants can contribute to an individual's risk of a rare disease, as has recently been reported in the context of neurodevelopmental disorders (Niemi *et al.* 2018). The

effect size and direction of this lead variant were consistent between African, European, and South Asian ancestries, despite differences in allele frequency between the population groups indicating a common, shared causal variant.

The rare variant rs144171242 (G) at 6p21.1 was associated with an increased risk for PUV. Given its rarity (MAF 0.007) this association would not have been detected using a conventional array-based GWAS approach, emphasizing the utility of sequencing-based analysis. The inclusion of South Asian individuals, in whom the effect size of rs144171242 is larger, also increased the power to detect association which was not genome-wide significant in the European-only analysis. Interestingly, this variant was not seen in individuals with African ancestry and is absent from all African populations included in the 1000 Genomes database (1000 Genomes Project Consortium *et al.* 2015) suggesting that it may have arisen after migration from Africa (~50-60,000 years ago).

Association of both lead variants was replicated in an independent European cohort, demonstrating that the observed signal was not being driven by differences in allele frequency within the mixed-ancestry cohort. Of note, the effect sizes for both lead variants were markedly attenuated, likely related to a combination of differences in ancestry (i.e., a mixed ancestry discovery cohort *vs* a European replication cohort) as well as the 'Winner's Curse' phenomenon. Winner's curse is where the effect sizes of alleles in the discovery cohort tend to be overestimated due to regression to the mean; those associations close to the discovery threshold are more likely to have biased overestimates of the variant's true association in the sampled population. Replication in an independent cohort is therefore necessary to not only validate identified associations but also calibrate the effect size estimates.

Only one previous array based GWAS involving PUV patients has previously been published. In contrast to this study, the authors looked specifically for associations with a wide range of kidney injury indicators (reduced eGFR, kidney failure, hypertension, proteinuria, nephrectomy) in patients with obstructive uropathy (PUV

and PUJO) (van der Zanden *et al.* 2021). The authors performed a meta-analysis of Cox regression allelic effects to identify significant association of a common (MAF 0.1) intronic variant in *CDH12* with kidney injury in patients with PUV (rs6874819; P=4.1x10^{-9}; hazard ratio [HR] 2.3; 95% CI 1.7-3.0). As might be expected given the different study objectives and design, neither rs6874819 (P=0.55) nor *CDH12* (P=0.77) were found to be significantly associated with PUV in my analysis.

Of note, PUV is a male-limited phenotype, although urethral abnormalities have been reported in the female relatives of affected males (Kolvenbach *et al.* 2019). This sexual dimorphism is most likely the result of anatomical differences in the development and length of the urethra between males and females, and although females do not develop PUV they may manifest other lower urinary tract phenotypes (see Section 6.4.3). While, X-linked inheritance has been reported in two patients with syndromic PUV/LUTO (Wade *et al.* 2021), I detected no significant common or rare variant associations on chromosome X in this cohort.

5.6.5 Strengths and limitations

The main strength of this analysis lies in the use of WGS data which enables ancestry-independent variant detection and the association testing of variants across the allele frequency spectrum, including multiallelic variants and indels which are not routinely included on genotyping panels. Furthermore, I used a rigorous statistical approach to ensure population structure was adequately controlled in this mixed-ancestry cohort enabling an inclusive and better powered analysis. The lack of genomic inflation and subsequent replication in the PUV analysis indicate that the detected associations are robust.

The main limitations of this study are the relatively small sample size, limiting power to detect variants with small effects and the lack of independent replication performed in the CAKUT and bladder exstrophy cohorts. Future meta-analyses will be necessary to overcome the issue of low power. In addition, most tools used for

post-GWAS analyses (e.g., PAINTOR) are not designed for use in mixed ancestry cohorts and for fine-mapping I was restricted to using 1000 Genomes (1000 Genomes Project Consortium *et al.* 2015) European reference data to determine LD patterns. The drive to include individuals from different backgrounds in genetic association studies means this is an active area of development for the bioinformatics community and is likely to improve in the near future.

5.7　Conclusion

In this chapter I have described how, for the first time, a well-controlled mixed-ancestry sequencing-based GWAS approach can be used successfully for locus discovery in rare disease, increasing power to detect novel associations and enhancing the fine-mapping of likely causal variants. Using this method, I identified the first robust genetic associations with PUV at 12q24.21 and 6p21.1, implicating both common and rare variation in the pathogenesis of a rare disease. Finally, I showed that a significant proportion of phenotypic variance in CAKUT is attributed to low-frequency variation providing evidence for a polygenic basis to this complex disorder and suggesting that better powered studies are likely to uncover novel genetic associations in the future.

Chapter 6: Functional annotation of PUV loci

6.1   Introduction

The pathogenesis of PUV is poorly understood but involves disruption of the dynamic and spatiotemporally specific developmental processes of lower urinary tract development. During embryogenesis, the bladder, prostate, and urethra develop from the endoderm-derived urogenital sinus, while the distal mesonephric (or Wolffian) duct forms the base of the bladder (trigone) before integrating into the prostatic urethra to become the epididymis, vas deferens and seminal vesicles in males. In the absence of testosterone, the mesonephric duct regresses in females. Abnormal integration of the mesonephric duct into the posterior urethra or persistence of the urogenital membrane have both been proposed as possible mechanisms underlying PUV (Krishnan *et al.* 2006), but the exact biological processes involved remain unknown. PUV is a uniquely male disorder, most likely as a result of differences in urethral length and development between males and females, although sex hormones and the epistatic effects of sex chromosomes may also contribute to sex-specific transcriptional and epigenomic profiles during embryogenesis (Deegan and Engel 2019).

In Section 5.4.3 I described the discovery and replication of the first two genomic loci associated with susceptibility to PUV. Fine mapping identified both non-coding lead variants as likely causal, however, identification of these variants is just the first step. Functional annotation of these loci is essential to determine the target gene (or genes) of the lead variants in the relevant cell types and to guide future mechanistic studies into the pathogenesis of PUV.

6.2   Aims

1. To use publicly available functional genomic data to prioritize likely causal genes and provide insights into the possible underlying disease mechanisms.

2. To interrogate large-scale datasets looking for additional phenotypic associations of the lead variants (PheWAS).

3. To determine whether the prioritized genes are expressed in the developing human embryo.

6.3   Methods

6.3.1   Functional annotation

To explore the functional relevance of the prioritized variants I used FUMA (version 1.3.6a) (Watanabe *et al.* 2017) to annotate the genome-wide significant loci. This web-based tool integrates functional gene consequences from ANNOVAR (Wang *et al.* 2010), CADD (Rentzsch *et al.* 2019) scores to predict deleteriousness, RegulomeDB score to indicate potential regulatory function (Boyle *et al.* 2012) and 15-core chromatin state (predicted by ChromHMM for 127 tissue/cell types) (Ernst and Kellis 2012) representing accessibility of genomic regions. Positional mapping (where a variant is physically located within a 10kb window of a gene), GTEx (v8) eQTL data (GTEx Consortium 2020) (using *cis*-eQTLs to map variants to genes up to 1Mb apart) and Hi-C data (to detect long-range 3D chromatin interactions) were used to prioritize genes likely to be affected by the variants of interest. Single-variant GWAS summary statistics were used as input with genomic positions converted to GRCh37 using the UCSC liftOver tool (Sugnet *et al.* 2002).

For tissue-specific annotation I used publicly available data from H1-BMP4 derived mesendoderm cultured cells, given the developing renal tract is derived from both

intermediate mesoderm (kidneys and ureters) and endoderm (bladder and urethra). The prioritized variants were intersected with the following epigenomic datasets generated by the ENCODE Project (ENCODE Project Consortium *et al.* 2020) and Roadmap Epigenomics (Roadmap Epigenomics Consortium *et al.* 2015) Consortia using the UCSC Genome Browser (Sugnet *et al.* 2002):

- Candidate *cis*-regulatory elements (cCREs) - ENCFF918FRW_ENCFF748XLQ_ENCFF313DOD (GRCh38)
- H3K27ac ChIP-seq as a marker of active enhancers - ENCFF918FRW_ENCFF748XLQ_ENCFF313DOD_ENCFF313DOD (GRCh38)
- H3K4me3 ChIP-seq as a marker of active promoters - ENCFF918FRW_ENCFF748XLQ_ENCFF313DOD_ENCFF748XLQ (GRCh38)
- DNase-seq as a marker of chromatin accessibility and potential transcription factor binding - ENCFF918FRW_ENCFF748XLQ_ENCFF313DOD_ENCFF918FRW (GRCh38)
- E004 H1 BMP4 Derived Mesendoderm Cultured Cells ImputedHMM (GRCh37). This dataset uses a model based on imputed data for 12 epigenetic marks across 127 reference epigenomes to assign one of 25 chromatin states to a genomic region e.g., active TSS, promoter, enhancer, transcribed, quiescent/low (Roadmap Epigenomics Consortium *et al.* 2015).

In addition Hi-C data from H1 mesendoderm cells (Dixon *et al.* 2015) were used to map chromatin contacts and visualize TADs for both loci using the 3D Interaction Viewer and Database (http://3div.kr). Hi-C is a high-throughput technique used to analyze genome-wide long-range chromatin interactions by crosslinking DNA-protein complexes with formaldehyde before fragmentation and massively parallel sequencing of the extracted DNA (Lieberman-Aiden *et al.* 2009). This process generates unbiased spatial proximity maps which can be used to potentially link a

217

region of interest to a promoter of a distant gene. Of note, due to high sequencing costs most Hi-C datasets have relatively low resolution (~25kb-40kb) which while useful to define large-scale chromatin architecture (e.g., TADs) are not yet able to identify specific enhancer-promoter interactions.

6.3.2 Transcription factor binding

The JASPAR 2020 (Fornes *et al.* 2020) CORE collection track (UCSC Genome Browser (Sugnet *et al.* 2002), updated 2019-10-13) was utilized to identify significant ($P < 10^{-4}$) predicted transcription factor binding sites (TFBS) that might intersect with the lead variants. The JASPAR database consists of manually curated, non-redundant, experimentally defined transcription factor binding profiles for 746 vertebrates, of which 637 are associated with human transcription factors with known DNA-binding profiles. Sequence logos based on position weight matrices of the DNA binding motifs were downloaded from JASPAR 2020 (Fornes *et al.* 2020).

6.3.3 PheWAS

A phenome-wide association study (PheWAS) is an unbiased method of testing for association between a single variant and many different phenotypes, providing potential insights into pleiotropic effects and related traits. Open Targets Genetics (Ghoussaini *et al.* 2021) was used to interrogate the publicly available NHGRI-EBI GWAS Catalog (MacArthur *et al.* 2017), the UK Biobank (Bycroft *et al.* 2018), and FinnGen (https://www.finngen.fi/) to determine known phenotypic associations of the lead variants at each locus.

6.3.4 Immunohistochemistry

Immunohistochemistry was performed in two seven-week-old human embryos by Dr Filipa Lopes and Professor Adrian Woolf from the University of Manchester. Human embryonic tissues, collected after maternal consent and ethical approval (REC18/NE/0290), were sourced from the Medical Research Council and Wellcome

Trust Human Developmental Biology Resource (https://www.hdbr.org/). Tissue sections were immunostained, as described previously (Kolvenbach *et al.* 2019). Sections were immunostained with the following primary antibodies: TBX5 (https://www.abcam.com/tbx5-antibody-ab223760.html) raised in rabbit; PTK7 (https://www.thermofisher.com/antibody/product/PTK7-Antibody-Polyclonal/PA5-82070) raised in rabbit; and uroplakin 1B (https://www.abcam.com/uroplakin-ibupib-antibody-upk1b3081-ab263454.html) raised in mouse. Primary antibodies were detected with appropriate second antibodies and signals generated with a peroxidase-based system.

6.4   Results

To explore the functional relevance of these loci I interrogated publicly available functional genomic datasets via UCSC Genome Browser (Sugnet *et al.* 2002) and used Functional Mapping and Annotation (FUMA) (Watanabe *et al.* 2017) to prioritize candidate genes. Given the urinary tract is derived from both embryonic mesoderm and endoderm, where possible I used experimental data obtained from male H1 BMP4-derived mesendoderm cultured cells.

6.4.1   12q24.21

The common, non-coding, intergenic lead variant (rs10774740) at the 12q24.21 locus is predicted to be deleterious with a relatively high CADD score of 15.54, which is unusual for an intergenic variant. It intersects with a conserved element (chr12:114228397-114228414; logarithm of odds score 33) that is suggestive of a putative transcription factor binding site (TFBS) (Figure 6.1). However, review of experimentally defined TF binding profiles (Fornes *et al.* 2020) did not identify any known interactions with DNA-binding motifs at this position. Interrogation of epigenomic data from ENCODE (ENCODE Project Consortium *et al.* 2020) revealed rs10774740 is located ~35bp away from a candidate *cis*-regulatory element (cCRE,

EH38E1646218), which although has low-DNase activity in mesendoderm cells, displays a distal enhancer-like signature in cardiac myocytes.
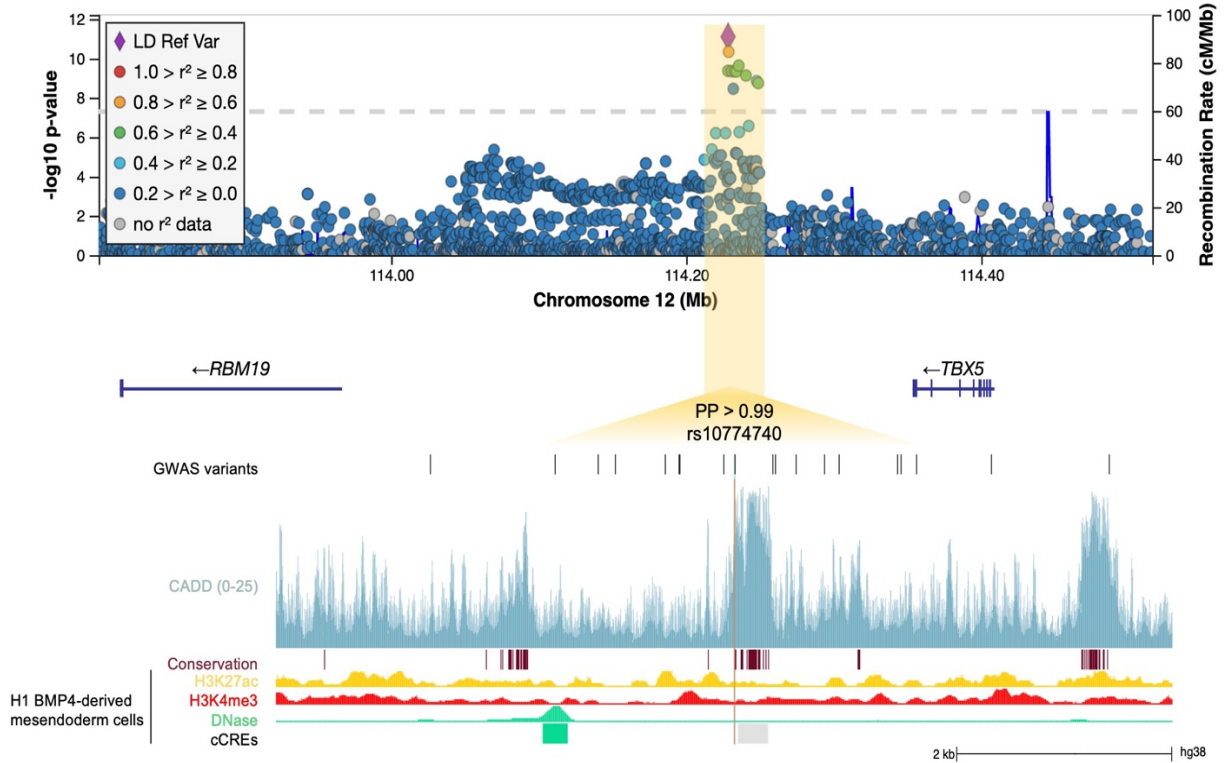
Figure 6.1. Regional association plot of 12q24.21.

Chromosomal position (GRCh38) is denoted along the x axis and strength of association using a –log$_{10}$(P) scale on the y axis. The lead variant (rs10774740) is represented by a purple diamond. Variants are coloured based on their LD with the lead variant using 1000 Genomes data from all population groups. Functional annotation of the lead prioritized variant rs10774740 shows intersection with CADD score (v1.6), PhastCons conserved elements from 100 vertebrates, and ENCODE H3K27ac ChIP-seq, H3K4me3 ChIP-seq and DNase-seq from mesendoderm cells. ENCODE cCREs active in mesendoderm are represented by shaded boxes; low-DNase (grey), DNase-only (green). GWAS variants with $P < 0.05$ are shown. Note that rs10774740 has a relatively high CADD score for a non-coding variant and intersects with a highly conserved region. PP, posterior probability derived using PAINTOR; cCRE, candidate *cis*-regulatory element.

There are no known *cis*-eQTL associations with rs10774740, but using experimental Hi-C data generated from H1 BMP4-derived mesendoderm cells (Dixon *et al.* 2015, Schmitt *et al.* 2016) I was able to determine that this locus is within the same TAD as the transcription factor *TBX5* (Figure 6.2). Chromatin interaction data mapped this intergenic locus directly to the promoter of *TBX5* 180kb away (FDR *q*=2.80x10^{-13}, Figure 6.3).
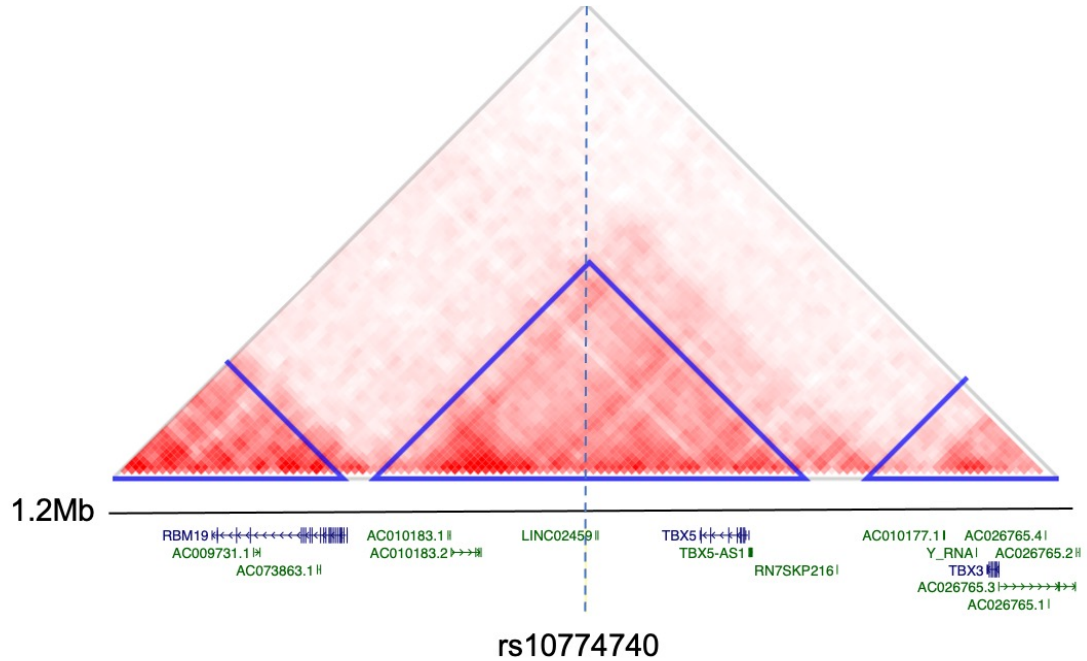
Figure 6.2. Heatmap of Hi-C interactions from H1 BMP4-derived mesendoderm cells.

rs10774740 is located within the same topologically associating domain (TAD) as *TBX5*. TADs are represented by blue triangles. Protein-coding genes are denoted in blue, non-coding genes in green.
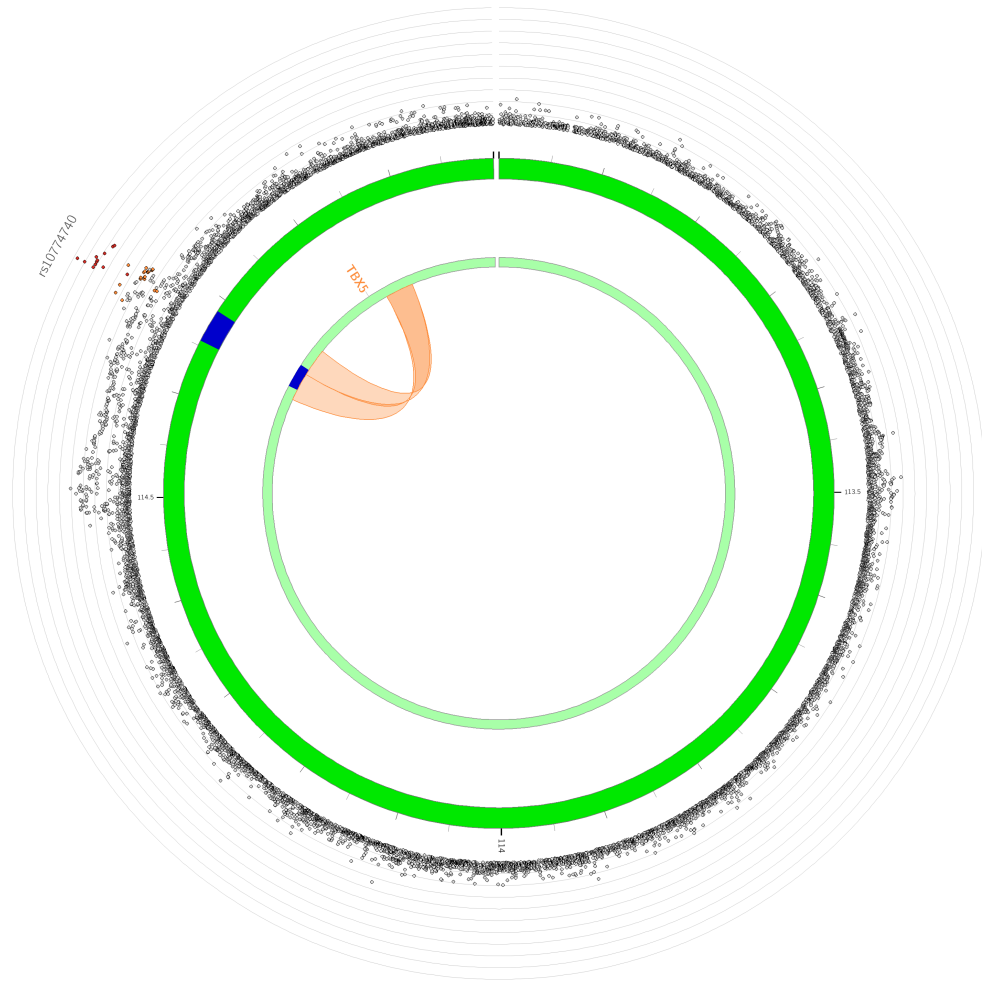
Figure 6.3. Circos plot illustrating significant chromatin interactions between 12q24.21 and the promoter of *TBX5*.

The outer layer represents a Manhattan plot with variants plotted against strength of association. Only variants with $P < 0.05$ are displayed. Genomic risk loci are highlighted in blue in the second layer. Significant chromatin loops detected in H1 BMP4-derived mesendoderm cultured cells are represented in orange.

6.4.2 6p21.1

At the 6p21.1 locus, the non-coding lead variant (rs144171242) is in an intron of the inactive tyrosine kinase *PTK7* (protein tyrosine kinase 7). This rare variant has a low CADD score (0.93) and lacks any relevant chromatin interaction or known eQTL associations, which is not unexpected given its rarity precludes detection by expression-array experiments. Interrogation of epigenomic annotations from ENCODE (ENCODE Project Consortium *et al.* 2020) revealed rs144171242 intersects a cCRE (EH38E2468259) with low DNase activity in mesendoderm cells, but with a distal enhancer-like signature in neurons (Figure 6.4). NIH Roadmap Epigenomics Consortium (Roadmap Epigenomics Consortium *et al.* 2015) data suggests rs144171242 may have regulatory activity in mesendoderm cells, classifying this region as transcribed/weak enhancer (12TxEnhW) using the imputed ChromHMM 25-chromatin state model (Figure 6.4).
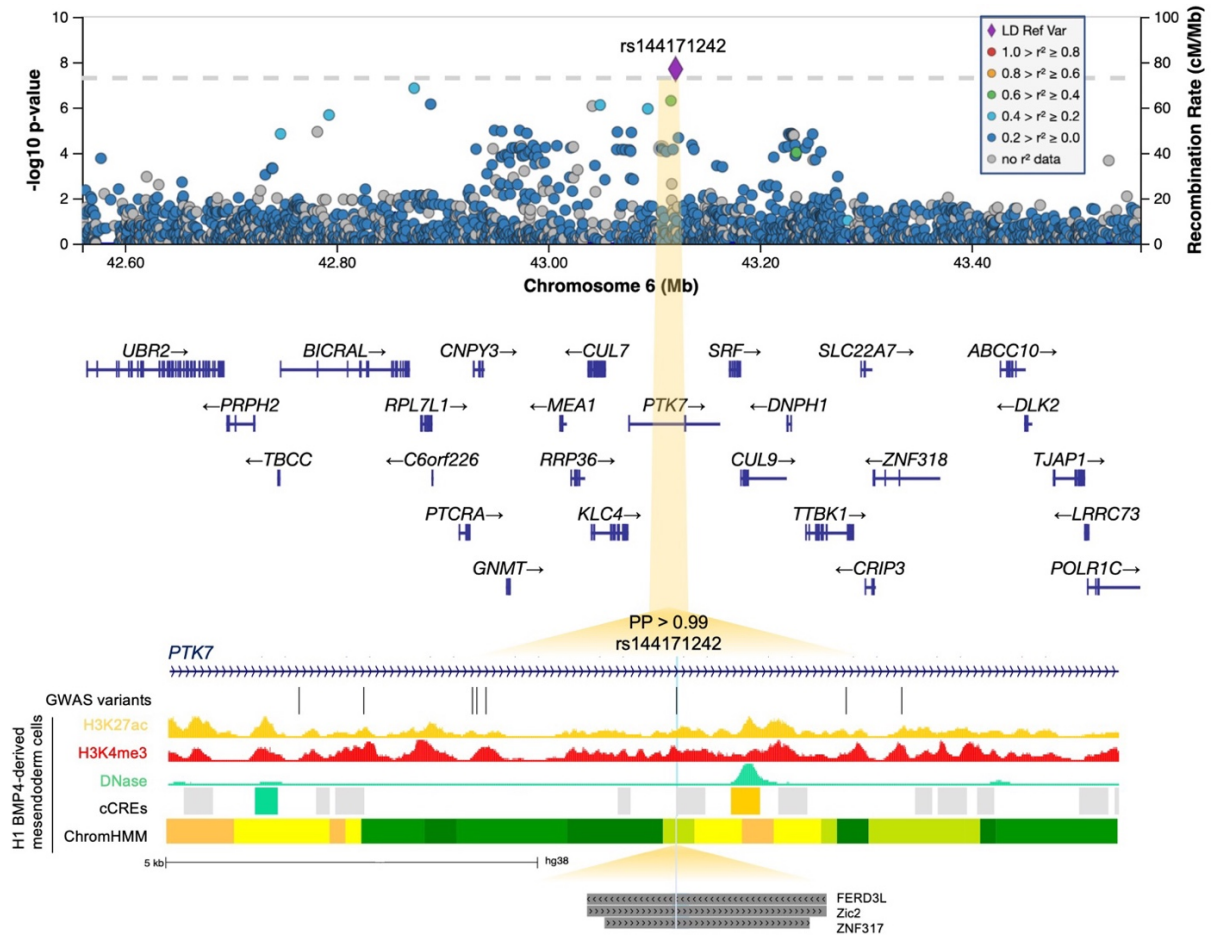
Figure 6.4. Regional association plot of 6p21.1.

Chromosomal position (GRCh38) is along the x axis and strength of association using a $-\log_{10}(P)$ scale on the y axis. The lead variant (rs144171242) is represented by a purple diamond. Variants are coloured based on their LD with the lead variant using 1000 Genomes data from all population groups. Functional annotation of the lead prioritized variant rs144171242 shows intersection with ENCODE H3K27ac ChIP-seq, H3K4me3 ChIP-seq and DNase-seq from mesendoderm cells. ENCODE cCREs active in mesendoderm are represented by shaded boxes; low-DNase (grey), DNase-only (green) and distal enhancer-like (orange). ChromHMM illustrates predicted chromatin states using Roadmap Epigenomics imputed 25-state model for mesendoderm cells; active enhancer (orange), weak enhancer (yellow), strong transcription (green), transcribed and weak enhancer (lime green). Predicted TFBS from the JASPAR 2020 CORE collection are indicated by dark grey shaded boxes. GWAS variants with $P < 0.05$ are shown. Note that rs144171242 intersects with both a predicted regulatory region and TFBS. PP, posterior probability derived using PAINTOR; cCREs, candidate *cis*-regulatory elements.

224

Interrogation of the JASPAR 2020 (Fornes *et al.* 2020) database of experimentally defined TF binding profiles revealed rs144171242 intersects with the DNA-binding motifs of FERD3L, ZNF317 and Zic2 (Figure 6.4), suggesting rs144171242 may potentially affect *PTK7* expression via disruption of TF binding (Figure 6.5). FERD3L (Fer3 Like BHLH Transcription Factor) is a basic helix-loop-helix transcriptional repressor of neurogenesis (Verzi *et al.* 2002).
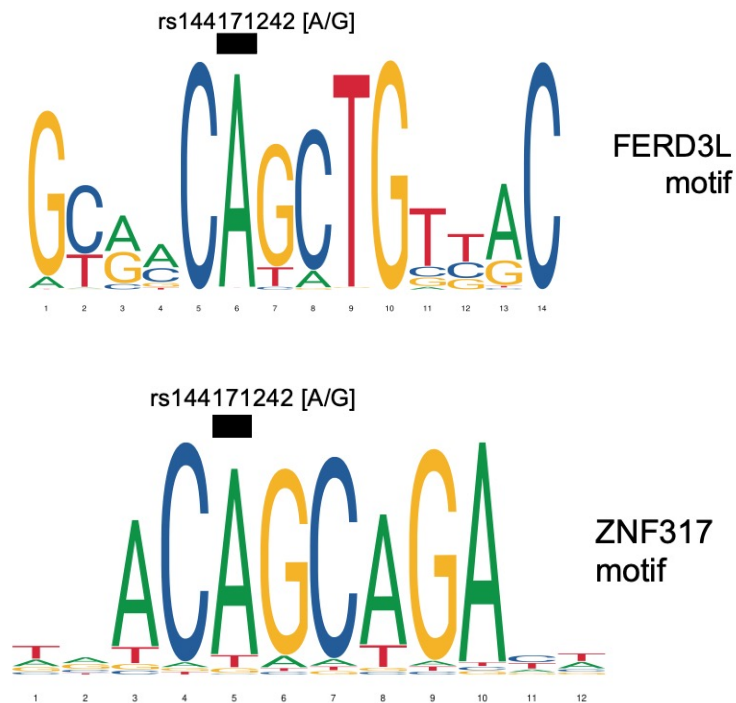
Figure 6.5. Sequence logos representing the DNA-binding motifs of transcription factors FERD3L and ZNF317.

The black boxes indicate where the rs144171242 effect allele [G] may disrupt binding.

6.4.3 PheWAS

Interrogation of the NHGR/EBI GWAS Catalog (MacArthur *et al.* 2017) revealed the risk allele rs10774740 (G) at the *TBX5* locus is associated with prostate cancer aggressiveness (Berndt *et al.* 2015) (P=3x10^{-10}; OR 1.14; 95% CI 1.09-1.18).

PheWAS data from the UK Biobank demonstrated the protective allele rs10774740 (T) also has a protective effect in female genitourinary phenotypes: urinary incontinence (P=8.3x10^{-12}; OR 0.90; 95% CI 0.87-0.92), female stress incontinence (P=7.9x10^{-10}; OR 0.89; 95% CI 0.85-0.92), genital prolapse (P=1.1x10^{-9}; OR 0.92; CI 0.89-0.94) and symptoms involving the female genital tract (P=1.7x10^{-8}; OR 0.90; 95% CI 0.87-0.94). No significant phenotypic associations affecting other organ systems were seen. Figure 6.6 shows the traits associated with rs10774740 (T) in the UK BioBank (Bycroft *et al.* 2018), FinnGen (https://www.finngen.fi/) and the GWAS Catalog (MacArthur *et al.* 2017) providing independent validation for a role of *TBX5* in urogenital development. No known GWAS or PheWAS associations were identified for rs144171242 or any variant in strong LD (EUR r^2 > 0.8) with it at the *PTK7* locus, likely because it is too rare to have been included in previous array-based GWASs.
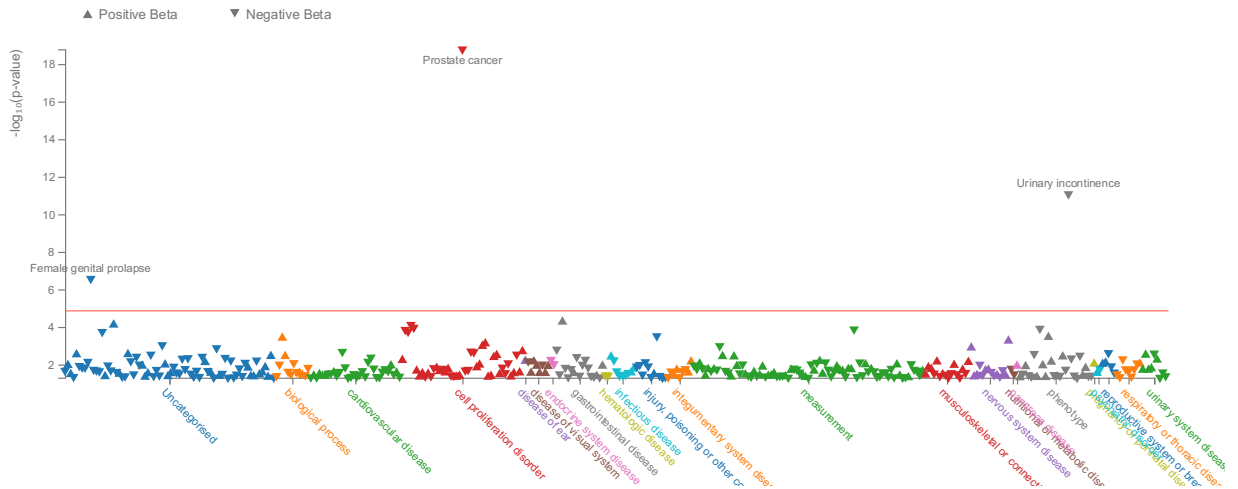
Figure 6.6. Manhattan plot of trait associations for rs10774740 (T).

Data from the UK BioBank, FinnGen and the GWAS Catalog generated by Open Targets Genetics (Ghoussaini *et al.* 2021).

6.4.4   Immunohistochemistry

To determine whether TBX5 and PTK7 are present during urinary tract development, immunohistochemistry was undertaken in two seven-week gestation normal human embryos by Dr Filipa Lopes at Manchester University (Figure 6.7 and Figure 6.8). At this stage of development, the urogenital sinus is a tube composed of epithelia that will differentiate into urothelial cells of the proximal urethra and the urinary bladder. Uroplakin 1B, a water-proofing protein, was detected in urogenital sinus epithelia (Figure 6.7B). PTK7 was detected in epithelia lining the urogenital sinus, and intensely in stromal-like cells surrounding the mesonephric ducts (Figure 6.7C and Figure 6.8A). TBX5 was detected in a nuclear pattern in a subset of epithelial cells lining the urogenital sinus (Figure 6.7D and Figure 6.8B). Omission of primary antibodies resulted in absent signals, as expected (Figure 6.7E).

Figure 6.7. Immunohistochemistry in the developing urinary tract of a 7-week gestation unaffected human embryo.

A, Overview of transverse section of a normal human embryo seven weeks after fertilization. The section has been stained with haematoxylin (blue nuclei). Boxes around the urogenital sinus and the mesonephric duct mark similar areas depicted under high power in **B-E**. In **B-D**, sections were reacted with primary antibodies, as indicated; in **E**, the primary antibody was omitted. **B-E** were counterstained with haematoxylin. In **B-E**, the left-hand frame shows the region around the mesonephric duct, while the right-hand frame shows one lateral horn of the urogenital sinus. **B**, Uroplakin 1b immunostaining revealed positive signal (brown) in the apical aspect of epithelia lining the urogenital sinus (arrows, right frame), the precursor of the urinary bladder and proximal urethra. Uroplakin 1b was also detected in the flat monolayer of mesothelial cells (left frame) that line the body cavity above the mesonephric duct. **C**, There were strong PTK7 signals (brown cytoplasmic staining) in stromal-like cells around the mesonephric duct (left frame), whereas the epithelia of the duct itself were negative. PTK7 was also detected in a reticular pattern in epithelia lining the urogenital sinus (right frame) and in stromal cells near the sinus. **D,** A subset of epithelial cells lining the urogenital sinus (right frame) immunostained for TBX5 (brown nuclei; some are arrowed). The mesothelial cells near the mesonephric duct (left frame) were also positive for TBX5. **E**, This negative control section had the primary antibody omitted; no specific (brown) signal was noted. Bar is 400 µm in **A**, and bars are 100 µm in **B-E**. ugs, urogenital sinus; md, mesonephric duct; hg, hindgut; u, ureter.
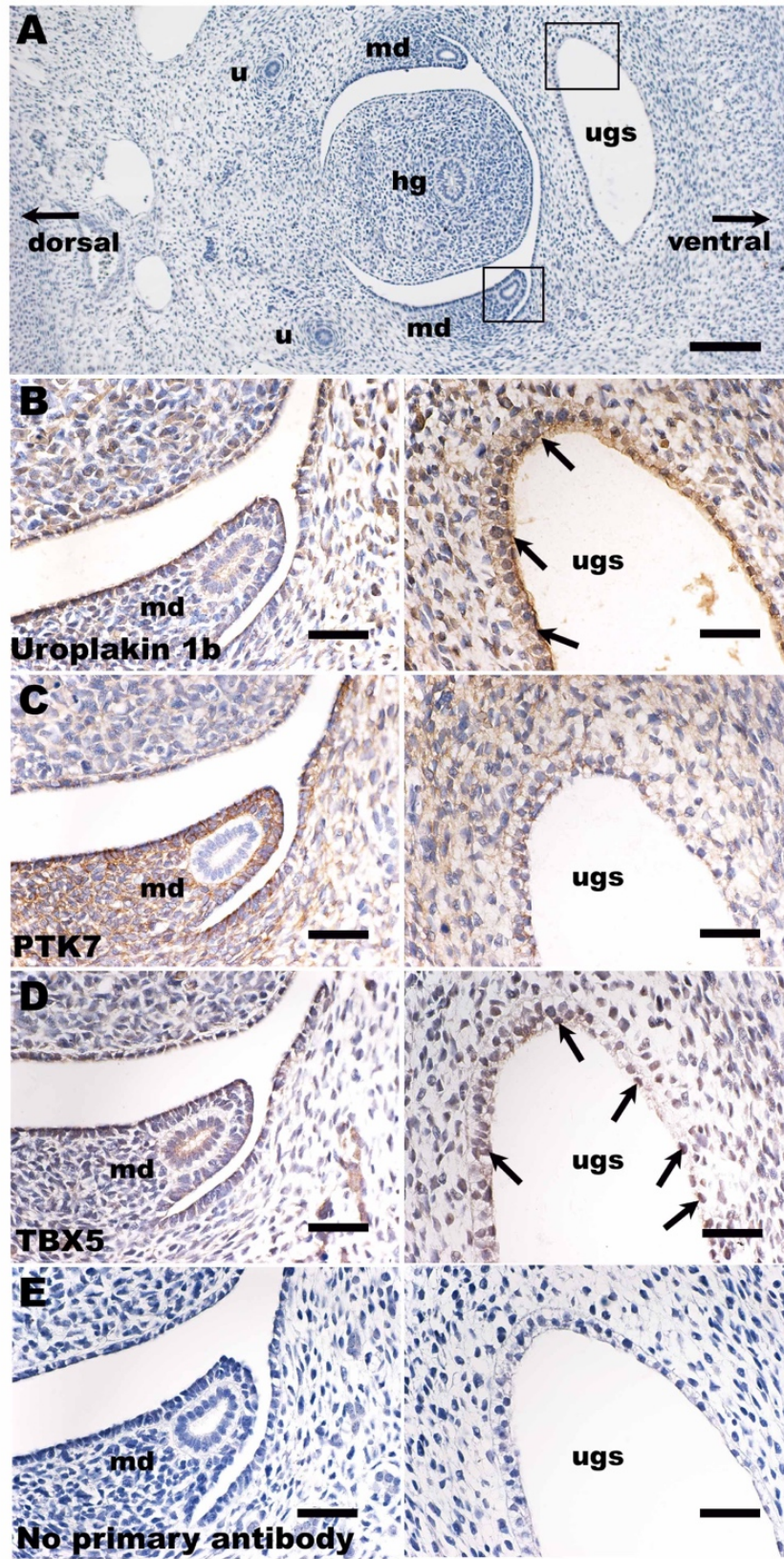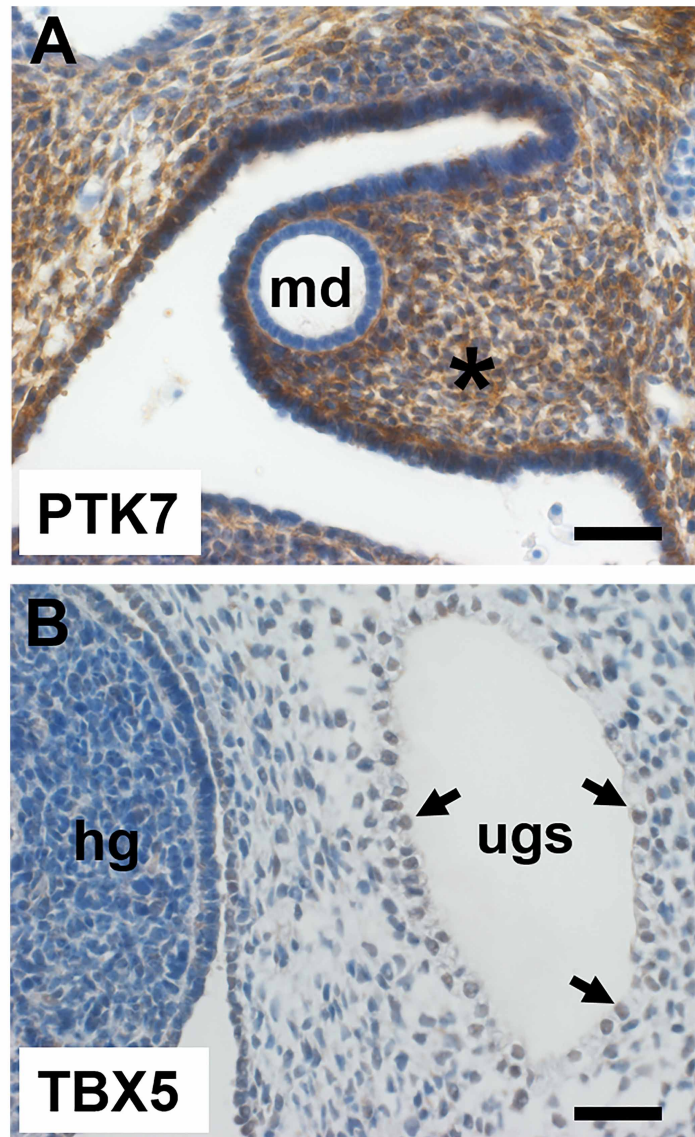
Figure 6.8. Immunohistochemistry of a second seven-week human embryo counterstained with haematoxylin.

A, View of the mesonephric duct (the epithelial tube with *md* in its lumen). Note the prominent signal (brown) for PTK7 in the stromal cells surrounding the duct. **B**, View of the urogenital sinus (*ugs*) with a subset of nuclei (three shown by arrows) in its monolayer epithelium that stain (light brown) for the transcription factor TBX5. The hindgut (*hg*) is nearby. Bars are 100μm.

6.5   Summary

- The common likely causal variant at 12q24.21 (rs10774740) was mapped to the transcription factor *TBX5* using chromatin interaction data from H1-BMP4 mesendoderm cells.

- PheWAS analysis demonstrated that rs10774740 (T) is associated with a reduced risk of aggressive prostate cancer, urinary incontinence and female genital prolapse.

- The rare likely causal variant at 6p21.1 (rs144171242) is in an intron of the planar cell polarity gene *PTK7* and predicted to intersect with transcription factor binding sites in a potentially active regulatory region in H1-BMP4 mesendoderm cells.

- Both *TBX5* and *PTK7* were expressed in the normal developing human urinary tract at the appropriate developmental stage.

6.6   Discussion

As discussed in Section 1.2.1, the vast majority of GWAS discoveries are found in the non-coding genome, suggesting these loci mediate their effects on gene expression through regulatory mechanisms. Mapping these variants to their target genes and cell-types and assigning functional significance, however, remains a major challenge. Conventionally lead variants identified in GWAS are presumed to affect expression of the nearest gene, however studies have shown that this proximity-based mapping approach is reliable in only 56-77% of cases (Mountjoy *et al.* 2021, Nasser *et al.* 2021), with causal variants also found to influence gene expression over long distances (Claussnitzer *et al.* 2015).

The interpretation of non-coding variants has significantly progressed over recent years through the generation of large-scale functional epigenomic, transcriptomic and proteomic datasets for hundreds of different tissues and cell types (Roadmap

Epigenomics Consortium *et al.* 2015, Regev *et al.* 2017, Sun *et al.* 2018, Wang *et al.* 2018, ENCODE Project Consortium *et al.* 2020, GTEx Consortium 2020). Furthermore, functional experiments (such as massively parallel reporter assays and clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9 genome editing) can be used to validate gene-variant relationships in a high-throughput manner (Shalem *et al.* 2015, van Arensbergen *et al.* 2019). However, despite this wealth of publicly available functional genomic data, the mapping of GWAS loci to causal variants to target genes and cell-types and ultimately to mechanisms of disease remains a challenge.

6.6.1 *TBX5* is a transcriptional regulator of embryogenesis

Using experimentally determined chromatin interaction data from mesendoderm cells I mapped the intergenic 12q24.21 locus to the promoter of the T-box transcription factor *TBX5* which was expressed in the nuclei of epithelial cells lining the urogenital sinus in human embryos. *TBX5* is highly constrained (LOEUF 0.14) and has well-defined roles in heart and limb development. Heterozygous pathogenic coding variants in *TBX5* result in autosomal dominant Holt-Oram syndrome (MIM 142900), characterized by congenital cardiac septal defects and upper limb anomalies, typically a triphalangeal thumb and secundum atrial septal defect (ASD) (Basson *et al.* 1997, Li *et al.* 1997). Homozygous variation in an enhancer ~90kb downstream of *TBX5* has also been linked to isolated congenital heart disease (Smemo *et al.* 2012) demonstrating how rare non-coding regulatory variation affecting *TBX5* expression can also cause cardiac malformations.

Animal models have been used extensively to investigate the role of *Tbx5* in heart and lung development. *Tbx5*$^{-/-}$ null mutant mice die between E9.5 and E10.5 with severe cardiac anomalies and failure of lung development while heterozygous mice exhibit the same haploinsufficient features seen in Holt-Oram syndrome (Bruneau *et al.* 2001). Interestingly, studies using *Xenopus* and mouse embryos have demonstrated that *Tbx5* initiates an evolutionarily conserved bi-directional retinoic

acid-Hedgehog-Wnt signaling loop which mediates crosstalk between mesoderm and endoderm and is essential for lung morphogenesis and cardiac septation (Steimle *et al.* 2018, Rankin *et al.* 2021). This reciprocal mesoderm-endoderm signaling loop is mediated via specific transcriptional enhancers, again highlighting the importance of non-coding variation to gene regulatory networks in development. This raises the intriguing possibility that *TBX5* may coordinate other mesoderm-endoderm interactions during embryogenesis, for example regulating the insertion of the distal mesonephric duct into the endoderm-derived cloaca.

The association of the risk allele with prostate cancer aggressiveness in men and genital prolapse and urinary incontinence in women raises the possibility that *TBX5,* which shows moderate expression in the adult bladder, is also associated with lower urinary tract phenotypes in adults. In a similar way that common variation in *TBX5* is associated with adult cardiac phenotypes including atrial fibrillation and cardiac dysrhythmias (https://r5.finngen.fi/gene/TBX5), common variation in the regulatory networks mediating *TBX5* expression may cause subtle structural aberrations of the urinary tract which manifest in later life as genital prolapse or urinary incontinence in women. Furthermore, the association of rs10774740 with prostate cancer is interesting given that PUV is a developmental defect of the prostatic urethra and raises the possibility that reactivation of repressed genes from embryonic development may result in cancer (Ma *et al.* 2010). In support of this theory, variation in genes associated with other developmental anomalies have also been linked to malignancy in the same organ. For example, *WT1* mutations are associated with both CAKUT and Wilms tumour or nephroblastoma (Call *et al.* 1990, Gessler *et al.* 1990), biallelic loss of *HNF1B* has been associated with an aggressive phenotype in chromophobe renal cell carcinoma, and *FOXF1* is linked to both oesophageal development and Barrett's oesophagus, a pre-malignant transformation of oesophageal epithelium (Su *et al.* 2012). These examples highlight the shared molecular pathways driving both development and cancer.

6.6.2 *PTK7* mediates elongation of the mesonephric duct

The likely causal variant at 6p21.1 is in an intron of *PTK7*. *PTK7* is highly constrained (LOEUF 0.28) and encodes a transmembrane receptor required for vertebrate embryonic patterning and morphogenesis. It mediates cell proliferation, migration, adhesion, actin cytoskeleton reorganization, and is a key regulator of planar cell polarity (PCP) via the non-canonical Wnt pathway (Lu *et al.* 2004, Berger *et al.* 2017). Altered expression of *PTK7* was initially observed in cancer (Mossie *et al.* 1995, Dunn and Tolwinski 2016), but rare missense variants in *PTK7* have since been linked to neural tube defects (Wang *et al.* 2015, Lei *et al.* 2019) and scoliosis (Hayes *et al.* 2014) in both humans and animal models, confirming a role for *PTK7* in embryonic development.

The PCP pathway is critical for determining the orientation of cells in the plane of an epithelium, regulating a process called convergent extension whereby cells intercalate by converging in one axis and elongating in the perpendicular axis. This process is mediated by a set of 'core' PCP genes (*Fzd*, *Dvl*, *Celsr*, *Vangl*, *Pk* and *Inversin*) of which multiple homologues exist resulting in functional redundancy (Torban and Sokol 2021). Defective PCP signaling has been shown to affect many aspects of kidney development and mutations in *Vangl2* and *Celsr1* result in neural tube and kidney abnormalities in mice (Yates *et al.* 2010, Brzóska *et al.* 2016). In humans, mutations in several other PCP genes have been associated with kidney malformations usually in the context of additional syndromic features: *FAT4* and renal agenesis (van der Ven *et al.* 2017), *ROR2* and Robinow syndrome (Brunetti-Pierri *et al.* 2008), and *WNT5a* with a duplex collecting system (Pietilä *et al.* 2016). These data highlight the importance of PCP signaling to kidney development but as yet no rare pathogenic variants in *PTK7* have been associated with CAKUT.

Ptk7-knockout mice form hypoplastic kidneys, in addition to neural tube defects (Lu *et al.* 2004), indicating that ureteric bud outgrowth does occur in these mutants. Interestingly, and of specific relevance to PUV, *Ptk7* has also been implicated in the

elongation of the mesonephric duct. Mesoderm-specific conditional deletion of *Ptk7* in mice affected convergent extension and tubular morphogenesis of the mesonephric duct at E18.5, resulting in a short, less coiled duct that affected sperm motility (Xu *et al.* 2016). Furthermore, the smooth muscle/mesenchymal cells surrounding the mesonephric duct were seen to be disorganized and randomly shaped suggesting a defect in radial intercalation (Xu *et al.* 2016). Subsequent studies have demonstrated that *Ptk7* regulates both the integrity of the extra-cellular matrix (ECM) as well as intracellular cytoskeleton mediators such as RAC1 (a Rho GTPase) in epithelial cells and myosin II in smooth muscle/mesenchymal cells (Lee, Andreeva, *et al.* 2012, Andreeva *et al.* 2014, Xu *et al.* 2018). In view of this data, detection of PTK7 in the stromal cells surrounding the mesonephric duct in our study is consistent with the hypothesis that subtle alterations in expression of PTK7 may perturb the delicate process of convergent extension of the mesonephric duct and perhaps lead to abnormal integration of the duct into the posterior urethra.

6.6.3 Strengths and limitations

The main strength of this analysis is the integration of GWAS, epigenomic and chromatin interaction data to ascertain the functional relevance of loci and identify biologically plausible causal genes. A role for the lead variant at the *TBX5* locus in urogenital phenotypes was also independently replicated by PheWAS data. However, although cell-specific annotations were used where possible, these data were generated *in vitro* from cultured mesendoderm cells and are at best an approximation of the complex and dynamic spatiotemporal processes that regulate development. Furthermore, although I have assessed the relevance of the associated loci using bioinformatic approaches and shown that publicly available and our own experimental data support the association, future functional work is needed to determine the precise biological mechanisms involved.

6.7   Conclusion

In this chapter I have used functional genomic data to prioritize the likely causal genes underlying two novel robust genetic associations detected in the PUV GWAS. Integration with publicly available epigenomic and transcription factor binding data, along with gene expression studies, have identified the transcription factor *TBX5* and planar cell polarity gene *PTK7* as candidate susceptibility genes for PUV. Both genes have well established roles in embryonic development, but these data implicate the PCP pathway in the pathogenesis of PUV for the first time. It is hoped that these findings will catalyze future investigations into the biological mechanisms of this important but poorly understood disorder.

Chapter 7: Discussion

The primary objective of this work was to better characterize the genetic architecture of structural renal and urinary tract malformations with the aim of providing insights into the pathogenesis of this genetically and phenotypically heterogenous spectrum of disorders. In this thesis I have described how a WGS population-based approach in individuals of diverse ancestry can be successfully employed to examine the contribution of different types of genomic variation to rare disease. This is the first study of its kind to interrogate both non-coding and coding variation across the entire allele frequency spectrum in an unbiased manner to determine novel genetic associations with a broad range of CAKUT phenotypes. Given our DNA is fixed at conception, the identification of robust genetic associations with a phenotype allows us to draw causal inferences about its pathogenesis (in some cases variants may correlate with environmental factors which have a direct causal effect, meaning causality cannot be proven beyond doubt) and potentially provide mechanistic insights into the underlying disease processes.

The results of this work have been discussed in detail in previous chapters and the key findings are summarized below:

- The contribution of known monogenic disease to CAKUT was small with no significant gene-based enrichment of rare SNV/indel or structural variation exome-wide indicating substantial genetic heterogeneity.
- No clear genome-wide significant common and low-frequency variant associations were identified in individuals with CAKUT, however low-frequency variants ($1\% \leq MAF < 5\%$) accounted for a significant proportion of the estimated heritability supporting a polygenic/complex basis for this disorder.

- Two novel loci, 12q24.21 and 6p21.1, were significantly associated with PUV and the association was replicated in an independent cohort. Bayesian fine mapping and functional annotation identified the likely causal variants and implicated the transcription factor *TBX5* and planar cell polarity gene *PTK7* in the pathogenesis of PUV.

- Significant enrichment of rare structural variation affecting *cis*-regulatory elements was observed in individuals with PUV; specifically small duplications affecting candidate enhancer-like elements and inversions affecting CTCF-only elements. This supports the disruption of long-range regulatory networks and chromatin conformation as a potential mechanism underlying PUV.

In this final chapter I draw together these findings and discuss how they might influence current thinking on the genetic architecture of structural renal and urinary tract disorders. I comment on the methodological implications of this work and the use of WGS as a research tool with particular emphasis on an inclusive approach to genetic association studies that aims to benefit individuals from diverse ancestral backgrounds. Finally, I discuss the wider implications of this research and future work to be performed.

7.1   The genetic architecture of CAKUT

As described in Section 1.3, the genetic architecture of a disease has traditionally been classified as monogenic, oligogenic or polygenic, with the omnigenic model also recently proposed by Boyle *et al.* (Boyle *et al.* 2017). Here, I consider how the findings from this study support or contradict each model and discuss which I think best describes the genetic architecture of CAKUT based on current data.

7.1.1   The monogenic model

The prevailing assumption for many years is that there is a monogenic basis underlying CAKUT (van der Ven, Vivante, *et al.* 2018), driven in part by the publication of genetic studies that have primarily focused on families or small selective cohorts using targeted sequencing approaches to look for rare, highly penetrant disease-causing alleles in a single gene. However, many of these studies have lacked an adequate control population or the genes identified have subsequently failed to replicate, casting doubt on the true causality of these mostly private variants. The focus on familial and severe disease has led to an over-representation of monogenic causes of CAKUT in the literature, compounded by publication bias and resource limitations where large-scale association studies are costly and challenging to perform. While a 'monogenic' cause can be identified in a small proportion of individuals with CAKUT, the incomplete penetrance and variable expressivity associated with even well-described CAKUT syndromes is difficult to reconcile with a simple single-gene model, suggesting that other risk factors or modifiers are likely to be important. Furthermore, despite the advances in NGS made over the last decade, we have not seen a simultaneous exponential increase in CAKUT gene discovery indicating that the 'low-hanging fruit' may have already been picked. The absence of significant gene-based SNV or SV enrichment seen in this study confirms that single-gene drivers of CAKUT are rare, at least in this cohort.

An abundance of monogenic causes of murine CAKUT (over 180 mouse models) is often cited as evidence for a single-gene basis for human CAKUT (van der Ven, Vivante, *et al.* 2018). However, most of these candidate genes have not been shown to impact human phenotypes in the same way. For example, *Bmp7* null mice exhibit renal dysplasia and eye abnormalities (Dudley *et al.* 1995) and *Isl1* deletions result in renal agenesis and hydroureter (Kaku *et al.* 2013), however neither *BMP7* or *ISL1* have been identified as causal genes in human kidney anomalies (although *ISL1* has been implicated in bladder exstrophy). This suggests that species-specific differences in embryonic development, functional redundancy, or differences in

238

genetic background may limit the translation of murine candidate genes to humans and provides further support for the hypothesis that other genetic or environmental factors may influence the final phenotype.

The prevalence of known 'monogenic' disease in this cohort was significantly lower than most previously reported studies, many of which were often established for the purpose of new monogenic gene discovery and so were deliberately enriched for familial, severe, or consanguineous cases. While this attenuated estimate may relate to the inclusion of the full spectrum of CAKUT phenotypes (including PUV and bladder exstrophy) or exclusion of pretested individuals, it may also result from regression to the population mean, a phenomenon often observed in common variant association studies. Specifically, the initial prevalence estimates of monogenic CAKUT have been inflated due to the selection of extreme phenotypes (i.e., severe, familial, or consanguineous disease) and small sample sizes, with subsequent studies such as the one described here, providing more moderate estimates, which regress to the population mean over time. This cohort has similar proportions of familial, consanguineous, and severe disease that might be expected in general nephrology practice in the UK providing a more clinically useful and generalizable (at least among Western countries) estimate of disease driven predominantly by a single gene.

Overall, while a small proportion of CAKUT may be explained by rare, highly penetrant monogenic variation, this is probably the exception rather than the rule and I propose that alternative models of inheritance better explain the genetic and phenotypic heterogeneity seen in this condition.

7.1.2 The oligogenic model

An oligogenic model of disease is considered an intermediate between monogenic and polygenic inheritance where a small number of rare or modest frequency genetic variants of moderate to large effect interact to cause disease. The rare multi-system

ciliopathy Bardet-Biedl syndrome has been proposed as an example of triallelic inheritance where three mutant alleles at two loci were observed in a small number of pedigrees (Katsanis *et al.* 2001), but subsequent studies have not established the clinical importance of these findings. While an oligogenic inheritance model may explain the variable expressivity observed in CAKUT there is only one unvalidated case report in the literature describing a possible oligogenic model of disease in a patient with bilateral cystic renal dysplasia (Schild *et al.* 2013). Although I cannot exclude epistatic gene-gene interactions based on the analyses performed, given the rarity of single-gene drivers it seems unlikely that oligogenic inheritance is an important mechanism in CAKUT.

7.1.3 The polygenic model

A polygenic or complex disease model for structural renal and urinary tract malformations has been proposed but definitive evidence supporting this hypothesis has so far been lacking (Nicolaou *et al.* 2015, 2016). In this study, I demonstrate that a significant proportion of the phenotypic variance of CAKUT (~23%) can be attributed to common and low-frequency variants, accounting for more of the estimated heritability than known monogenic disease (< 5% in this cohort). These data strongly implicate a polygenic basis to CAKUT and suggest that larger association studies will be better powered to detect this 'missing heritability'. The identification of two novel loci associated with PUV, replication of the 5q11.1 locus previously linked with bladder exstrophy (Draaken *et al.* 2015), as well as results from a recent meta-analysis of VUR (Verbitsky *et al.* 2021) provide additional support for a more complex model of disease, at least for lower urinary tract phenotypes.

The polygenic liability threshold model has been suggested to explain susceptibility to a wide range of complex diseases including coronary artery disease, type 2 diabetes, inflammatory bowel disease (Khera *et al.* 2018), schizophrenia (Tansey *et al.* 2016), depression (Wray *et al.* 2018), and autism spectrum disorder (Klei *et al.* 2021). Under this model, the underlying normal distribution of genetic risk is

determined by multiple variants of small effect and there is a liability threshold above which individuals develop disease. How such a model could be applied to the temporally sensitive and cell-specific disruption of embryonic development associated with CAKUT is unclear, although one possibility is that an individual might display a background polygenic liability on which external environmental or *in utero* factors act, ultimately resulting in malformations of the renal tract.

An alternative model where common and low-frequency variants modify the penetrance or expressivity of rare, pathogenic alleles could potentially explain the intra-familial variability and broad range of phenotypes seen. Such polygenic contributions to monogenic risk are being increasingly recognized across a wide range of conditions including inherited cancer syndromes, familial hypercholesterolemia and neurodevelopmental disorders (Paquette *et al.* 2017, Niemi *et al.* 2018, Oetjens *et al.* 2019, Fahed *et al.* 2020, Mars *et al.* 2020). However, given so few individuals with CAKUT are found to have rare, pathogenic alleles this polygenic modification of penetrance and expressivity would likely only apply to a minority of cases. Nonetheless, it would be interesting to test this theory in families with known 'monogenic' disease but discordant phenotypes.

7.1.4   The omnigenic model

The omnigenic model hypothesizes that gene regulatory networks are sufficiently interconnected that *all* genes expressed in disease-relevant cells can affect the function of core disease-related genes and that most of the heritability of a disease is explained by the effects on genes *outside* these core pathways (Boyle *et al.* 2017). Several observations have been made which support this model in complex disease. First, significant GWAS associations explain only a modest fraction of phenotypic variance with the implication that additional variants with effect sizes below the genome-wide significance threshold account for this 'missing heritability' (Manolio *et al.* 2009, Yang *et al.* 2010). Second, GWAS signals are relatively uniformly distributed across the genome (Loh *et al.* 2015, Boyle *et al.* 2017) with the heritability

contributed by each chromosome closely proportional to its physical length (Visscher *et al.* 2006, Shi *et al.* 2016) suggesting that causal variants are evenly spread genome-wide rather than aggregated into specific biological pathways. Lastly, the majority of GWAS signals fall in the non-coding space and are enriched in regions of active chromatin (Maurano *et al.* 2012) which implicates gene regulatory pathways as mediators of disease risk.

How can this omnigenic model be applied to structural renal tract malformations? This study has demonstrated that single-gene drivers of CAKUT are rare and that in most cases this is not a simple monogenic disease. This finding supports a model where well-described 'core' genes (such as *HNF1B*, *PAX2* etc.) or pathways (e.g., planar cell polarity) have a clear biological role and direct effects on disease risk but overall, these core genes only contribute a small fraction of total heritability. The bulk of the heritability is therefore proposed to be mediated by an infinitesimal number of additional genes/variants with non-zero indirect effects that comprise tissue or cell-specific 'peripheral' regulatory networks, modifying the impact of core genes and pathways (Figure 7.1). The finding that common and low-frequency variation makes a significant contribution to the phenotypic variance of CAKUT supports the theory that additional variants with small effects can cumulatively modify disease risk.

Renal and urinary tract development is a highly regulated, dynamic process that must be switched on in the right place at the right time and is dependent on interconnecting hierarchical regulatory networks that include both local enhancer-promoter interactions and large-scale 3D chromatin contacts. Widespread redundancy in these regulatory networks acts to prevent deleterious phenotypic consequences upon the loss of individual regulatory elements adding a further layer of complexity (Osterwalder *et al.* 2018). The discovery in this study that both non-coding SNVs and SVs affecting regulatory elements are associated with PUV highlights how even subtle perturbations in these tightly controlled regulatory networks may impact disease susceptibility.

Remarkably, a multigenic model for CAKUT was originally proposed almost 20 years ago by Ichikawa *et al.* (Ichikawa *et al.* 2002). Noting the incomplete penetrance and variable expressivity associated with the condition, the authors speculated that CAKUT was not a typically Mendelian disease and suggested that it results from the "accumulation of minor mutations in multiple genes, each of which has multiple ontogenic functions". This observation is strangely prescient of the omnigenic model described above and seems a more biologically plausible description of the genetic architecture of structural renal tract malformations than the traditional monogenic view.

Larger-scale WGS association studies and heritability estimates incorporating all genomic variation across the entire allele frequency spectrum will further tease out the proportion of phenotypic variance that can be attributed to genetic variation as opposed to environmental effects in CAKUT. Such studies will undoubtedly also lead to the development of genetic risk scores with the hope that these might help stratify risk of ESKD in affected individuals. However, if an omnigenic model does apply to CAKUT, the fundamental question remains: Will identification of large numbers of variants with very small effects generate significant and clinically translatable insights into the underlying biology of the disease?
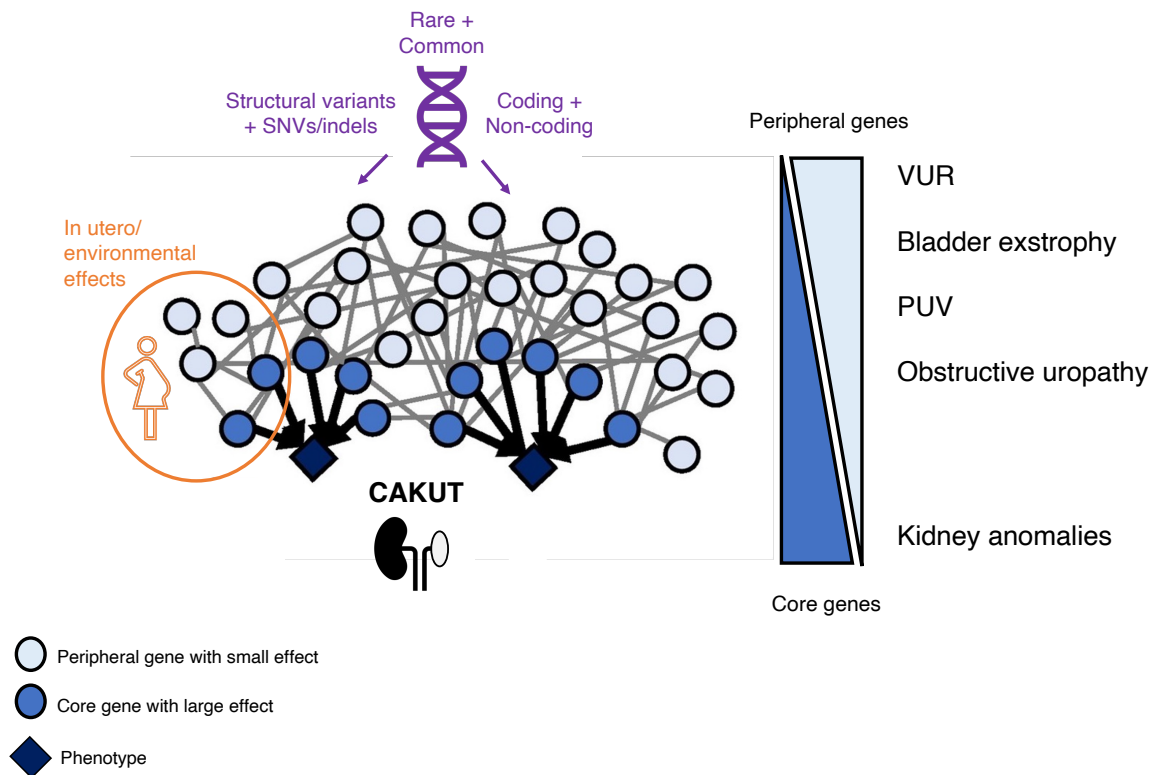
Figure 7.1. An omnigenic model of CAKUT.

This figure details how an omnigenic model might be applied to different CAKUT phenotypes, illustrating how the cumulative 'mutational burden' comprising all types of genomic variation across the full allele frequency spectrum may affect both core genes with large effects and peripheral genes with small effects resulting in different phenotypes. In this model, some CAKUT phenotypes (i.e., kidney anomalies) are further from the omnigenic extreme with larger contributions from a few core genes, whereas other phenotypes (e.g., VUR and bladder exstrophy) are closer to the omnigenic extreme with a larger fraction of heritability attributable to variation in non-core genes. Functional redundancy between regulatory networks and environmental (*in utero*) effects adds further complexity.

7.2   WGS as a diagnostic and research tool

One of the main strengths of this work is the use of clinical-grade WGS which captures nearly all variation across the genome in an unbiased manner enabling simultaneous analysis of SNVs/indels and structural variants across the entire allele

frequency spectrum. Several studies report a superior diagnostic yield in rare disease using WGS compared with chromosomal microarray, targeted gene panels and WES due to the detection of deep intronic variants and small CNVs, as well as variants affecting non-coding RNAs, mitochondrial DNA and exonic regions with poor coverage on WES (Gilissen *et al.* 2014, Taylor *et al.* 2015, Stavropoulos *et al.* 2016, Lionel *et al.* 2018, Turro *et al.* 2020, 100,000 Genomes Project Pilot Investigators *et al.* 2021). PCR-free WGS can also detect > 99.5% of known pathogenic SNVs and indels demonstrating superior sensitivity over WES which shows greater variation in coverage within and between genomic sites and fails to reliably call small exonic deletions (Turro *et al.* 2020).

The recently published 100,000 Genomes rare disease pilot study analyzed WGS data from 2,183 families with undiagnosed rare disease (24% had neurodevelopmental disorders and 88% were of European ancestry), reporting an overall diagnostic yield of 25%, with a quarter of diagnoses having immediate implications for clinical decision making (100,000 Genomes Project Pilot Investigators *et al.* 2021). Of note, 14% of these genetic diagnoses were made after additional research analysis was performed and included *de novo* variants in highly constrained coding regions, mitochondrial, intronic splicing, structural and non-coding variants validated using *in vitro* assays. Re-analysis with updated gene panels increased the diagnostic yield further, highlighting how the periodic and systematic reassessment of WGS can be beneficial, especially in the context of a rapidly evolving field like clinical genetics (Costain *et al.* 2018, 100,000 Genomes Project Pilot Investigators *et al.* 2021).

The advantages of a molecular diagnosis are clear. First, providing a diagnosis for a patient and their family can end the 'diagnostic odyssey' and uncertainty that many experience. For example, in the 100,000 Genomes Project rare disease pilot study the median duration of such a patient pathway was 75 months with a median number of hospital visits of 68 (100,000 Genomes Project Pilot Investigators *et al.* 2021). Furthermore, obtaining a clear diagnosis permits access to patient support groups

and services and can result in eligibility for clinical trials, as well as providing prognostic information. Second, a genetic diagnosis may impact clinical management prompting a change in medication or leading to additional surveillance of the proband or their family. Third, knowledge of the inheritance of a genetic disorder can inform reproductive choices and enable access to pre-implantation genetic diagnosis (PIGD) where embryos created through *in vitro* fertilization (IVF) are tested for a specific disorder and only unaffected embryos selected for implantation in the womb. Fourth, identification of the pathogenic variant allows for pre-symptomatic testing in family members (e.g., in the case of inherited cancer syndromes) and specific to nephrology, enables monogenic disease to be excluded, permitting kidney donation from an unaffected relative. While many of these advantages are not specific to WGS, its superior sensitivity and comprehensive nature mean it can reduce the time to diagnosis, alleviating the anxiety and investigative burden for patients and their families as well as having the potential to reduce overall health care costs.

WGS has now been introduced into routine clinical practice in the NHS and can enhance diagnostic capabilities and precision medicine provision. However, there are ethical, social, and legal considerations that accompany the use of WGS. For example, the maintenance of patient confidentiality and data security is paramount; ensuring the systematic reassessment of VUSs is essential; obtaining valid consent in the context of complex, uncertain, and often changing information is challenging; and clinician responsibilities to family members who may also be affected by the results should be considered. Furthermore, 1-2% of those undergoing sequencing have clinically significant and highly penetrant variants found in one of 73 'medically actionable' genes unrelated to the indication for testing, most of which are associated with cancer or cardiovascular phenotypes (Amendola *et al.* 2015, Olfson *et al.* 2015, Groopman *et al.* 2019, Miller *et al.* 2021). Determining which of these secondary findings should be reported back to the individual is subject to important debate in genomic medicine and requires careful and considered consent. Initiatives to

improve genomic literacy will also be key to aid people's understanding of these complex issues and ensure the successful implementation of WGS for diagnostic testing.

The use of WGS as a research tool is now widespread as a result of large-scale short-read sequencing projects such as the 1000 Genomes Project (Sudmant *et al.* 2015), deCODE (Jónsson *et al.* 2017), TOPMed (Taliun *et al.* 2019), the 100,000 Genomes Project (100,000 Genomes Project Pilot Investigators *et al.* 2021) and the UK Biobank (Bycroft *et al.* 2018) which recently released WGS data for 200,000 individuals. In this analysis I demonstrate clear benefits of using a WGS approach, identifying enrichment of SVs (and inversions) affecting *cis*-regulatory elements in individuals with PUV, which would not have been detected using conventional microarrays. Furthermore, use of WGS data enabled the discovery and fine-mapping of two likely causal variants associated with PUV, where the rarity of one of the variants would have evaded detection by a classic genotyping and imputation approach.

Sequencing-based GWASs (primarily using the ethnically diverse TOPMed dataset) are starting to be employed successfully in an increasing number of common and complex traits and have identified novel associations with rare and low frequency alleles and variants found in non-European populations (Natarajan *et al.* 2018, Raffield *et al.* 2020, Zhao *et al.* 2020, Hu *et al.* 2021, Mikhaylova *et al.* 2021). The next five years will undoubtedly see an explosion of studies utilizing these vast WGS datasets for large-scale population-based association analyses and will likely provide novel insights into the non-coding and regulatory genome. The increasing availability of long-read WGS will also improve our ability to interrogate repetitive regions of the genome to better understand how large and complex SVs contribute to human disease. All in all, this is an incredibly exciting time to be a researcher in genomics.

7.3   Increasing diversity in genetic studies

The majority of genetic association studies are performed in individuals of European ancestry, despite this group making up one of the smallest populations globally. Europeans account for over 95% of GWAS participants (https://gwasdiversitymonitor.com) and most have been recruited from just three countries: the US, UK, and Iceland (Mills and Rahal 2019). Such Euro-centric bias has resulted from practical limitations such as the availability of funding, genotyping technologies, and analytic methods. It is now widely recognized, however, that increasing ancestral diversity in genetic studies has both scientific and ethical advantages; not only expanding our knowledge of the full extent of human genomic variation but also importantly, ensuring that the benefits of genomic medicine are applicable to all (Peterson *et al.* 2019).

Discussions around ancestry, race, ethnicity, and their role in the genetics of human disease have become increasingly prevalent in recent years (Peterson *et al.* 2019, Sirugo *et al.* 2019, Birney *et al.* 2021, Lewis *et al.* 2022). While the concept of race or ethnicity as a meaningful biological construct has been soundly rejected by the human genetics' community, their use as 'proxies' for an individual's social and cultural environment (e.g., access to resources, geographical surroundings, levels of chronic stress) has been suggested and the terms are often still, incorrectly, used interchangeably with ancestry. Genetic ancestry is the preferred term to describe the genetic differences between groups of individuals but even this perpetuates the notion that genetic ancestry is itself a discrete entity. In fact, the considerable genetic variation that is observed between individuals incorporates both recent and ancient population migration, divergence, and admixture, representing a continuum which cannot be categorized into discrete groups (Birney *et al.* 2021, Lewis *et al.* 2022). This multidimensional, continuous view of genetic ancestry is illustrated nicely by PCA using data from this study (Figure 2.2).

Traditionally genetic association studies have assigned individuals to arbitrary population groups based on clustering approaches to enable robust statistical comparisons to be made, the reality however, is that all individuals are 'admixed' to some extent. Exclusion of individuals who do not fit into discrete categories limits both power for discovery and the transferability of results, as has been shown in the case of polygenic risk scores (Duncan *et al.* 2019). The benefit of combining individuals with different genetic ancestry has been clearly demonstrated in a variety of genetic association approaches. First, trans-ancestry meta-analysis specifically utilizes differences in LD structure between population groups to improve the resolution of fine-mapping GWAS loci, reducing the number of candidate causal variants in credible sets for complex traits such as type 2 diabetes, eGFR and lipid levels (Mahajan *et al.* 2014, Morris *et al.* 2019, Chen *et al.* 2021, Graff *et al.* 2021, Graham *et al.* 2021). Second, admixture mapping capitalizes on recent mixing between population groups to correlate local ancestry at genetic loci with a phenotype and has been used successfully to identify the *APOL1* genomic region associated with FSGS and hypertension-related ESKD in individuals of African ancestry (Kao *et al.* 2008, Genovese *et al.* 2010, Shriner 2017). Third, the transferability of polygenic risk scores to predict genetic risk of disease is markedly improved when the discovery study is more diverse due to improvements in fine-mapping and prioritization of multi-ancestry causal variants (Cavazos and Witte 2021, Graham *et al.* 2021). Finally, diverse ancestry rare variant analyses are a useful way to boost power for gene discovery through increased sample size (Cirulli *et al.* 2020, Wang *et al.* 2021), with the 'collapsing' approach used to aggregate rare variants mitigating concerns regarding differing allele frequencies across different genetic ancestries.

GWAS findings have been shown to replicate across populations in a variety of common diseases and complex traits such as diabetes (Waters *et al.* 2010, Marigorta and Navarro 2013), hypertension (Kaur *et al.* 2021), inflammatory bowel disease (Marigorta and Navarro 2013, Liu *et al.* 2015), schizophrenia (Lam *et al.*

2019), blood lipids (Coram *et al.* 2013, Kuchenbaecker *et al.* 2019, Graham *et al.* 2021), and height and body mass index (Guo *et al.* 2021). The Global Biobank Meta-analysis Initiative (GBMI) consisting of 2.1 million people from 19 biobanks across four continents also recently reported 96% concordance in effect size across ancestral groups at > 500 genome-wide significant loci associated with 14 different diseases (Zhou *et al.* 2021) providing further evidence for the sharing of causal variants between ancestries. Where lack of replication in non-European populations has been observed, this has been attributed to differences in LD (where markers in LD with the causal variant in Europeans may not be in LD in non-European populations), differences in genetic architecture because of genetic drift or local selection, as well as epistasis due to differences in genetic background or environmental interactions.

The development of novel statistical tools which model genetic relatedness and mitigate confounding by population structure now enable robust pan-ancestry analysis in genetic association studies. Such an approach has been used to successfully identify novel associations with common phenotypes such as chronic obstructive pulmonary disease (COPD) (Zhao *et al.* 2020), red cell traits (Hu *et al.* 2021), white cell traits (Mikhaylova *et al.* 2021) and lipid levels (Graham *et al.* 2021), however such studies are still the exception rather than the rule. In this study I have demonstrated for the first time how a WGS-based diverse ancestry approach can be used to investigate *rare* disease, beyond conventional monogenic gene discovery. This method capitalizes on the unbiased variant detection permitted by WGS, which is superior to imputation in the identification of variants that are rare and found in non-European populations (Höglund *et al.* 2019). Furthermore, the inclusion of all individuals, regardless of ancestral background, increases both the power for discovery (through increased sample size) and enhances the fine mapping of likely causal variants by leveraging differences in LD patterns (through increased diversity). Further refinement of the statistical methodology underlying pan-ancestry association analysis and robust application to large mixed-ancestry cohorts (with

specific focus on barriers to recruitment) is now necessary to ensure that future studies are as inclusive as possible, for both the scientific advantages they offer and to prevent exacerbation of existing health inequalities.

7.4   Impact and implications

The findings from this work will have an impact across multiple different disciplines. First, from a patient's perspective, clinicians are now able to offer an explanation as to the potential cause of PUV for the first time, offering individuals hope of improved understanding of their disease. Second, these results will be of great interest to other researchers within the nephrogenetics field, adult, and paediatric nephrologists as well as clinical geneticists, and stimulate collaborations to further unravel the genomic complexity of CAKUT. Third, the association of *TBX5* and *PTK7* with PUV provides developmental biologists with new insights into the role of these genes in embryonic development and clearly implicates the planar cell polarity pathway in urethral development. These findings will hopefully catalyze future mechanistic and therapeutic studies. Finally, the successful use of a rigorously controlled pan-ancestry WGS association study to dissect the genetic contribution of a rare and complex disease has wider implications for the genomics community. Here, we have clearly demonstrated the scientific benefits of an inclusive approach which provided novel insights that would have been missed in a purely European cohort. Such an approach has not previously been applied to rare disease and the methodology used in this study will form an exemplar for subsequent analyses of WGS datasets across any number of rare diseases. Furthermore, this study contributes to the normalization of representation of individuals from diverse ancestral backgrounds in genetic association studies which will hopefully raise awareness of the current inequalities in genomics and promote the recruitment of more diverse cohorts going forward.

Looking to the future, the field of regenerative and reconstructive medicine is advancing rapidly. Improving our understanding of the key mechanisms that underlie mammalian kidney organogenesis and urinary tract development will be critical for the generation of mature and functioning organoids with the hope that they may provide options for renal regeneration and bladder reconstruction in the future. This is particularly relevant not only for individuals with ESKD but also for those with impaired bladder function due to myelomeningocele, PUV, bladder exstrophy, cancer or spinal cord injury who often require complex reconstructive surgery. In addition, patients with urethral stricture or who have undergone treatment for prostate cancer often require reconstruction of the posterior urethra which can prove challenging. The use of stem-cell based tissue-engineering to generate organ-specific grafts and the possibility of functional organoids in the future is therefore an exciting prospect (Adamowicz *et al.* 2019).

7.5   Future directions

The main findings of this thesis, that the genes *TBX5* and *PTK7* are associated with PUV, are only the beginning of the story. Further experiments will be necessary to dissect the molecular mechanisms linking the risk alleles to disease pathogenesis and to determine how these genes function in health and disease. Such an undertaking is challenging given the cell-specific dynamic developmental processes involved and functional redundancy of regulatory networks. Some of the key questions generated by this work and how they might be approached are discussed below.

- What effect do the likely causal variants have on gene expression? In which cell types do they act and at which developmental stage?
- What are the regulatory mechanisms that mediate any potential changes in gene expression? Do the variants alter transcription factor binding, promoter-enhancer interactions, or 3D genome organization?

- What impact do the variants have on the phenotype of a model organism such as a zebrafish? Do mice with conditional knockouts of *TBX5* and *PTK7* display genitourinary anomalies?

To answer these questions a combination of functional *in silico*, *in vitro* and *in vivo* approaches could be used. First, to improve the granularity of fine-mapping, integration of the GWAS data with datasets generated from human embryonic urinary tract samples at different developmental stages could help identify associations with cell-specific gene expression (using scRNA-seq) or accessible chromatin (using snATAC-seq). At present, however, such datasets are not publicly available, although work on the 'Human Developmental Cell Atlas' will be critical for understanding congenital disorders in the future (https://www.humancellatlas.org/dca/).

Confirmation that the likely causal variants affect gene expression *in vitro* can be sought using a reporter assay, although with the caveat that this approach does not provide an accurate representation of the complex regulatory networks involved. To assess impact on transcription factor binding, protein-DNA interactions can be interrogated using an electrophoretic mobility shift assay (EMSA) or using genomic footprinting which uses paired ATAC-seq and ChIP-seq to determine transcription factor binding.

For an unbiased assessment of gene expression, a CRISPR-based single-cell functional genomics approach could be applied to induced pluripotent stem cells (iPSCs). Here, single-guide RNAs are tiled across the candidate disease loci containing the variants of interest and scRNA-seq performed to look for potential variant-linked enhancer/silencer target genes showing differential expression compared with isogenic cells.

Recently, patient-derived iPSCs have been used to generate mini 3D kidney and ureteric bud organoids, which in the context of CAKUT, may help to establish the significance of a genetic defect *in vitro* and provide insights into the aberrant

development of the kidney (Schmidt-Ott 2017, Forbes *et al.* 2018, Mae *et al.* 2020, Morais *et al.* 2022). A key benefit of this approach is that the organoids retain their cell-specific genomic background enabling better modelling of the regulatory pathways and networks involved in nephrogenesis.

Considering *in vivo* methods*,* a reverse genetics approach using morpholino-based knockdown or CRISPR-Cas9 genome editing of *TBX5* or *PTK7* in a model vertebrate organism such as zebrafish embryos (*Danio rerio*) would be useful to ascertain the presence of any phenocopies of lower urinary tract obstruction. In addition, conditional knock-out or CRISPR/Cas9 genome editing in embryonic mouse models could be used to look for genitourinary anomalies, although such models are primarily designed to investigate phenotypes associated with loss-of-function variants rather than regulatory variants which may cause more subtle aberrations in gene expression.

This analysis also detected an enrichment of rare SVs affecting *cis*-regulatory elements in individuals with PUV. Given this data is preliminary, validation of these structural variants with long-read sequencing would be desirable in addition to independent replication in a cohort of patients who have similarly undergone WGS. Of particular interest was the association of inversions affecting CTCF-binding regions with susceptibility to PUV. This raises the question:

- Do these inversions impact chromatin conformation and how might this impact the regulation of gene expression?

To investigate this Hi-C, a high throughput method based on chromatin conformation capture, could be used to study the effects of inversions affecting the CTCF-elements on higher order chromatin structure in a genome wide manner. By comparing Hi-C data generated from wildtype and risk allele cell lines (containing the relevant inversions), it would be possible to identify compartment switching (where a region of the genome moves from a B/repressive compartment to an

254

A/active compartment) or changes in TAD boundaries. Integration with scRNA-seq data would quantify any changes in gene expression associated with these variants.

Many questions remain regarding the pathogenesis of CAKUT. For example:

- How can we explain the variable expressivity and incomplete penetrance seen in individuals with a known monogenic cause of CAKUT?
- How much of an impact do maternal and *in utero* factors have on the CAKUT phenotype?
- What role does epigenomics play in the pathogenesis?

Lastly, as with any genetic association study, increasing sample size will increase power to detect association with variants at a lower MAF or with a smaller effect size, as well as genes that contribute to only a small fraction of CAKUT. Collaboration with other researchers in the field will be key to generating further insights into the pathogenesis of this complex disease and plans for meta-analysis of the PUV and bladder exstrophy cohorts with collaborators in Europe are already underway.

7.6   Lessons Learnt

During this project, both the data itself and tools used to analyze it were constantly evolving. For example, in my first year, the WGS data was raw, unfiltered, and difficult to analyze on a large-scale, especially for someone with, at that point, quite limited bioinformatic skills. This improved over time with subsequent data releases from Genomics England and support from the online and Genomics England bioinformatics community. Furthermore, prior to the publication of SAIGE (Zhou *et al.* 2018) and SAIGE-GENE (Zhou *et al.* 2020), the association analyses I performed (using RVTESTS (Zhan *et al.* 2016)) suffered from significant type 1 error rates. With hindsight I spent far too much time looking into candidate genes which hadn't been replicated, many of which were considered 'biologically plausible' which highlights the perils of inadequately controlled studies and statistical noise.

If I were to begin this project again now I would:

1. Use the latest set-based rare-variant association test SAIGE-GENE+ (Zhou *et al.* 2022) which has improved type 1 error control and incorporates additional functional annotations.

2. Perform a joint SNV-SV gene-based analysis to examine the combined burden of these different variants across the exome to identify novel candidate genes.

3. Use the latest SV data generated by Genomics England which has been reprocessed using the DRAGEN pipeline. This should improve the accuracy of SV calling. In addition, SVs would be validated using long-read sequencing technologies which are now becoming more widely available.

4. Use male-only controls when analysing male-limited phenotypes (e.g., PUV) to prevent possible attenuation of association signals.

5. Use tissue-specific *cis*-regulatory elements only for the non-coding analyses.

6. Version control all code.

7. Only investigate replicated and statistically robust signals.

7.7 Conclusion

In this thesis I set out to better characterize the genetic architecture of structural renal and urinary tract anomalies using large-scale WGS data, the first time this approach has been used to dissect this complex and heterogenous spectrum of malformations. I have demonstrated that the monogenic contribution to this disorder is considerably less than previously estimated and that there is likely to be a significant proportion of heritability attributable to common and low-frequency variants. Furthermore, I have identified the first robust genetic associations for PUV, the most common cause of ESKD in boys, demonstrating that common non-coding variation can contribute to rare disease susceptibility. In addition to these findings, I have shown that WGS can be used successfully for the investigation of rare disease

beyond conventional monogenic gene discovery, and that inclusion of individuals from diverse ancestral backgrounds can improve power for disease locus discovery and enhance the fine-mapping of causal variants.

Taking these results together, the genetic architecture of CAKUT does not appear to follow a traditional Mendelian pattern and previous attempts to categorize it as either monogenic or polygenic have been too reductionist. The genotypic and phenotypic heterogeneity of these structural renal and urinary tract malformations, along with the observed variable expressivity and incomplete penetrance, support an omnigenic model of disease where the boundaries between common and rare, coding, and non-coding variation are distinctly blurred. The impact of maternal and environmental factors on this omnigenic model also remains to be explored. The concept of genome-wide 'mutational burden' is an interesting one and determining precisely how different types of variation across the allele frequency spectrum interact in a spatiotemporal specific manner to regulate the intricate and dynamic processes of embryogenesis will be key to understanding this complex yet important disorder.

Bibliography

100,000 Genomes Project Pilot Investigators, Smedley, D., Smith, K.R., Martin, A., Thomas, E.A., McDonagh, E.M., Cipriani, V., Ellingford, J.M., Arno, G., Tucci, A., Vandrovcova, J., Chan, G., Williams, H.J., Ratnaike, T., Wei, W., Stirrups, K., Ibanez, K., Moutsianas, L., Wielscher, M., Need, A., Barnes, M.R., Vestito, L., Buchanan, J., Wordsworth, S., Ashford, S., Rehmström, K., Li, E., Fuller, G., Twiss, P., Spasic-Boskovic, O., Halsall, S., Floto, R.A., Poole, K., Wagner, A., Mehta, S.G., Gurnell, M., Burrows, N., James, R., Penkett, C., Dewhurst, E., Gräf, S., Mapeta, R., Kasanicki, M., Haworth, A., Savage, H., Babcock, M., Reese, M.G., Bale, M., Baple, E., Boustred, C., Brittain, H., de Burca, A., Bleda, M., Devereau, A., Halai, D., Haraldsdottir, E., Hyder, Z., Kasperaviciute, D., Patch, C., Polychronopoulos, D., Matchan, A., Sultana, R., Ryten, M., Tavares, A.L.T., Tregidgo, C., Turnbull, C., Welland, M., Wood, S., Snow, C., Williams, E., Leigh, S., Foulger, R.E., Daugherty, L.C., Niblock, O., Leong, I.U.S., Wright, C.F., Davies, J., Crichton, C., Welch, J., Woods, K., Abulhoul, L., Aurora, P., Bockenhauer, D., Broomfield, A., Cleary, M.A., Lam, T., Dattani, M., Footitt, E., Ganesan, V., Grunewald, S., Compeyrot-Lacassagne, S., Muntoni, F., Pilkington, C., Quinlivan, R., Thapar, N., Wallis, C., Wedderburn, L.R., Worth, A., Bueser, T., Compton, C., Deshpande, C., Fassihi, H., Haque, E., Izatt, L., Josifova, D., Mohammed, S., Robert, L., Rose, S., Ruddy, D., Sarkany, R., Say, G., Shaw, A.C., Wolejko, A., Habib, B., Burns, G., Hunter, S., Grocock, R.J., Humphray, S.J., Robinson, P.N., Haendel, M., Simpson, M.A., Banka, S., Clayton-Smith, J., Douzgou, S., Hall, G., Thomas, H.B., O'Keefe, R.T., Michaelides, M., Moore, A.T., Malka, S., Pontikos, N., Browning, A.C., Straub, V., Gorman, G.S., Horvath, R., Quinton, R., Schaefer, A.M., Yu-Wai-Man, P., Turnbull, D.M., McFarland, R., Taylor, R.W., O'Connor, E., Yip, J., Newland, K., Morris, H.R., Polke, J., Wood, N.W., Campbell, C., Camps, C., Gibson, K., Koelling, N., Lester, T., Németh, A.H., Palles, C., Patel, S., Roy, N.B.A., Sen, A., Taylor, J., Cacheiro, P., Jacobsen, J.O., Seaby, E.G., Davison, V., Chitty, L., Douglas, A., Naresh, K., McMullan, D., Ellard, S., Temple, I.K., Mumford, A.D., Wilson, G., Beales, P., Bitner-Glindzicz, M., Black, G., Bradley, J.R., Brennan, P., Burn, J., Chinnery, P.F., Elliott, P., Flinter, F., Houlden, H., Irving, M., Newman, W., Rahman, S., Sayer, J.A., Taylor, J.C., Webster, A.R., Wilkie, A.O.M., Ouwehand, W.H., Raymond, F.L., Chisholm, J., Hill, S., Bentley, D., Scott, R.H., Fowler, T., Rendon, A., and Caulfield, M., 2021. 100,000 genomes pilot on rare-disease diagnosis in health care - preliminary report. *The New England journal of medicine*, 385 (20), 1868–1880.

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R., 2015. A global reference for human genetic variation. *Nature*, 526 (7571), 68–74.

Abdelhak, S., Kalatzis, V., Heilig, R., Compain, S., Samson, D., Vincent, C., Weil, D., Cruaud, C., Sahly, I., Leibovici, M., Bitner-Glindzicz, M., Francis, M., Lacombe, D., Vigneron, J., Charachon, R., Boven, K., Bedbeder, P., Van Regemorter, N., Weissenbach, J., and Petit, C., 1997. A human homologue of the Drosophila eyes absent gene underlies branchio-oto-renal (BOR) syndrome and identifies a novel gene family. *Nature genetics*, 15 (2), 157–164.

Abel, H.J., Larson, D.E., Regier, A.A., Chiang, C., Das, I., Kanchi, K.L., Layer, R.M., Neale, B.M., Salerno, W.J., Reeves, C., Buyske, S., NHGRI Centers for Common Disease Genomics, Matise, T.C., Muzny, D.M., Zody, M.C., Lander, E.S., Dutcher, S.K., Stitziel, N.O., and Hall, I.M., 2020. Mapping and characterization of structural variation in 17,795 human genomes. *Nature*, 583 (7814), 83–89.

Abifadel, M., Varret, M., Rabès, J.-P., Allard, D., Ouguerram, K., Devillers, M., Cruaud, C., Benjannet, S., Wickham, L., Erlich, D., Derré, A., Villéger, L., Farnier, M., Beucler, I., Bruckert, E., Chambaz, J., Chanu, B., Lecerf, J.-M., Luc, G., Moulin, P., Weissenbach, J., Prat, A., Krempf, M., Junien, C., Seidah, N.G., and Boileau, C., 2003. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nature genetics*, 34 (2), 154–156.

Adalat, S., Hayes, W.N., Bryant, W.A., Booth, J., Woolf, A.S., Kleta, R., Subtil, S., Clissold, R., Colclough, K., Ellard, S., and Bockenhauer, D., 2019. HNF1B mutations are associated with a Gitelman-like tubulopathy that develops during childhood. *Kidney international reports*, 4 (9), 1304–1311.

Adalat, S., Woolf, A.S., Johnstone, K.A., Wirsing, A., Harries, L.W., Long, D.A., Hennekam, R.C., Ledermann, S.E., Rees, L., van't Hoff, W., Marks, S.D., Trompeter, R.S., Tullus, K., Winyard, P.J., Cansick, J., Mushtaq, I., Dhillon, H.K., Bingham, C., Edghill, E.L., Shroff, R., Stanescu, H., Ryffel, G.U., Ellard, S., and Bockenhauer, D., 2009. HNF1B mutations associate with hypomagnesemia and renal magnesium wasting. *Journal of the American Society of Nephrology: JASN*, 20 (5), 1123–1131.

Adamowicz, J., Kuffel, B., Van Breda, S.V., Pokrwczynska, M., and Drewa, T., 2019. Reconstructive urology and tissue engineering: Converging developmental paths. *Journal of tissue engineering and regenerative medicine*, 13 (3), 522–533.

Ahn, Y.H., Lee, C., Kim, N.K.D., Park, E., Kang, H.G., Ha, I.-S., Park, W.-Y., and Cheong, H.I., 2020. Targeted exome sequencing provided comprehensive genetic diagnosis of congenital anomalies of the kidney and urinary tract. *Journal of clinical medicine*, 9 (3), 751.

Akbari, P., Gilani, A., Sosina, O., Kosmicki, J.A., Khrimian, L., Fang, Y.-Y., Persaud, T., Garcia, V., Sun, D., Li, A., Mbatchou, J., Locke, A.E., Benner, C., Verweij,

N., Lin, N., Hossain, S., Agostinucci, K., Pascale, J.V., Dirice, E., Dunn, M., Regeneron Genetics Center, DiscovEHR Collaboration, Kraus, W.E., Shah, S.H., Chen, Y.-D.I., Rotter, J.I., Rader, D.J., Melander, O., Still, C.D., Mirshahi, T., Carey, D.J., Berumen-Campos, J., Kuri-Morales, P., Alegre-Díaz, J., Torres, J.M., Emberson, J.R., Collins, R., Balasubramanian, S., Hawes, A., Jones, M., Zambrowicz, B., Murphy, A.J., Paulding, C., Coppola, G., Overton, J.D., Reid, J.G., Shuldiner, A.R., Cantor, M., Kang, H.M., Abecasis, G.R., Karalis, K., Economides, A.N., Marchini, J., Yancopoulos, G.D., Sleeman, M.W., Altarejos, J., Della Gatta, G., Tapia-Conyer, R., Schwartzman, M.L., Baras, A., Ferreira, M.A.R., and Lotta, L.A., 2021. Sequencing of 640,000 exomes identifies GPR75 variants associated with protection from obesity. *Science (New York, N.Y.)*, 373 (6550), eabf8683.

Alazami, A.M., Shaheen, R., Alzahrani, F., Snape, K., Saggar, A., Brinkmann, B., Bavi, P., Al-Gazali, L.I., and Alkuraya, F.S., 2009. FREM1 mutations cause bifid nose, renal agenesis, and anorectal malformations syndrome. *The American Journal of Human Genetics*, 85 (3), 414–418.

Alkan, C., Coe, B.P., and Eichler, E.E., 2011. Genome structural variation discovery and genotyping. *Nature reviews. Genetics*, 12 (5), 363–376.

Amendola, L.M., Dorschner, M.O., Robertson, P.D., Salama, J.S., Hart, R., Shirts, B.H., Murray, M.L., Tokita, M.J., Gallego, C.J., Kim, D.S., Bennett, J.T., Crosslin, D.R., Ranchalis, J., Jones, K.L., Rosenthal, E.A., Jarvik, E.R., Itsara, A., Turner, E.H., Herman, D.S., Schleit, J., Burt, A., Jamal, S.M., Abrudan, J.L., Johnson, A.D., Conlin, L.K., Dulik, M.C., Santani, A., Metterville, D.R., Kelly, M., Foreman, A.K.M., Lee, K., Taylor, K.D., Guo, X., Crooks, K., Kiedrowski, L.A., Raffel, L.J., Gordon, O., Machini, K., Desnick, R.J., Biesecker, L.G., Lubitz, S.A., Mulchandani, S., Cooper, G.M., Joffe, S., Richards, C.S., Yang, Y., Rotter, J.I., Rich, S.S., O'Donnell, C.J., Berg, J.S., Spinner, N.B., Evans, J.P., Fullerton, S.M., Leppig, K.A., Bennett, R.L., Bird, T., Sybert, V.P., Grady, W.M., Tabor, H.K., Kim, J.H., Bamshad, M.J., Wilfond, B., Motulsky, A.G., Scott, C.R., Pritchard, C.C., Walsh, T.D., Burke, W., Raskind, W.H., Byers, P., Hisama, F.M., Rehm, H., Nickerson, D.A., and Jarvik, G.P., 2015. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome research*, 25 (3), 305–315.

Amri, K., Freund, N., Vilar, J., Merlet-Bénichou, C., and Lelièvre-Pégorier, M., 1999. Adverse effects of hyperglycemia on kidney development in rats: in vivo and in vitro studies. *Diabetes*, 48 (11), 2240–2245.

Andreeva, A., Lee, J., Lohia, M., Wu, X., Macara, I.G., and Lu, X., 2014. PTK7-Src signaling at epithelial cell contacts mediates spatial organization of actomyosin and planar cell polarity. *Developmental cell*, 29 (1), 20–33.

van Arensbergen, J., Pagie, L., FitzPatrick, V.D., de Haas, M., Baltissen, M.P., Comoglio, F., van der Weide, R.H., Teunissen, H., Võsa, U., Franke, L., de Wit, E., Vermeulen, M., Bussemaker, H.J., and van Steensel, B., 2019. High-throughput identification of human SNPs affecting regulatory element activity. *Nature genetics*, 51 (7), 1160–1169.

Arfeen, S., Rosborough, D., Luger, A.M., and Nolph, K.D., 1993. Familial unilateral renal agenesis and focal and segmental glomerulosclerosis. *American journal of kidney diseases: the official journal of the National Kidney Foundation*, 21 (6), 663–668.

Arora, V., Khan, S., El-Hattab, A.W., Dua Puri, R., Rocha, M.E., Merdzanic, R., Paknia, O., Beetz, C., Rolfs, A., Bertoli-Avella, A.M., Bauer, P., and Verma, I.C., 2021. Biallelic pathogenic GFRA1 variants cause autosomal recessive bilateral renal agenesis. *Journal of the American Society of Nephrology: JASN*, 32 (1), 223–228.

Ashraf, S., Hoskins, B.E., Chaib, H., Hoefele, J., Pasch, A., Saisawat, P., Trefz, F., Hacker, H.W., Nuernberg, G., Nuernberg, P., Otto, E.A., and Hildebrandt, F., 2010. Mapping of a new locus for congenital anomalies of the kidney and urinary tract on chromosome 8q24. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association*, 25 (5), 1496–1501.

Athanasiou, Y., Voskarides, K., Gale, D.P., Damianou, L., Patsias, C., Zavros, M., Maxwell, P.H., Cook, H.T., Demosthenous, P., Hadjisavvas, A., Kyriacou, K., Zouvani, I., Pierides, A., and Deltas, C., 2011. Familial C3 glomerulopathy associated with CFHR5 mutations: clinical characteristics of 91 patients in 16 pedigrees. *Clinical journal of the American Society of Nephrology: CJASN*, 6 (6), 1436–1446.

Atkinson, M.A., Ng, D.K., Warady, B.A., Furth, S.L., and Flynn, J.T., 2021. The CKiD study: overview and summary of findings related to kidney disease progression. *Pediatric nephrology (Berlin, Germany)*, 36 (3), 527–538.

Audano, P.A., Sulovari, A., Graves-Lindsay, T.A., Cantsilieris, S., Sorensen, M., Welch, A.E., Dougherty, M.L., Nelson, B.J., Shah, A., Dutcher, S.K., Warren, W.C., Magrini, V., McGrath, S.D., Li, Y.I., Wilson, R.K., and Eichler, E.E., 2019. Characterizing the major structural variant alleles of the human genome. *Cell*, 176 (3), 663-675.e19.

Auer, P.L., Teumer, A., Schick, U., O'Shaughnessy, A., Lo, K.S., Chami, N., Carlson, C., de Denus, S., Dubé, M.-P., Haessler, J., Jackson, R.D., Kooperberg, C., Perreault, L.-P.L., Nauck, M., Peters, U., Rioux, J.D., Schmidt, F., Turcot, V., Völker, U., Völzke, H., Greinacher, A., Hsu, L., Tardif, J.-C., Diaz, G.A., Reiner, A.P., and Lettre, G., 2014. Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. *Nature genetics*, 46 (6), 629–634.

Babcock, M., Yatsenko, S., Hopkins, J., Brenton, M., Cao, Q., de Jong, P., Stankiewicz, P., Lupski, J.R., Sikela, J.M., and Morrow, B.E., 2007. Hominoid lineage specific amplification of low-copy repeats on 22q11.2 (LCR22s) associated with velo-cardio-facial/digeorge syndrome. *Human molecular genetics*, 16 (21), 2560–2571.

Badano, J.L. and Katsanis, N., 2002. Beyond Mendel: an evolving view of human genetic disease transmission. *Nature reviews. Genetics*, 3 (10), 779–789.

Barak, H., Huh, S.-H., Chen, S., Jeanpierre, C., Martinovic, J., Parisot, M., Bole-Feysot, C., Nitschké, P., Salomon, R., Antignac, C., Ornitz, D.M., and Kopan, R., 2012. FGF9 and FGF20 maintain the stemness of nephron progenitors in mice and man. *Developmental cell*, 22 (6), 1191–1207.

Barua, M., Stellacci, E., Stella, L., Weins, A., Genovese, G., Muto, V., Caputo, V., Toka, H.R., Charoonratana, V.T., Tartaglia, M., and Pollak, M.R., 2014. Mutations in PAX2 associate with adult-onset FSGS. *Journal of the American Society of Nephrology: JASN*, 25 (9), 1942–1953.

Basson, C.T., Bachinsky, D.R., Lin, R.C., Levi, T., Elkins, J.A., Soults, J., Grayzel, D., Kroumpouzou, E., Traill, T.A., Leblanc-Straceski, J., Renault, B., Kucherlapati, R., Seidman, J.G., and Seidman, C.E., 1997. Mutations in human cause limb and cardiac malformation in Holt-Oram syndrome. *Nature genetics*, 15 (1), 30–35.

Bates, C.M., 2011. Role of fibroblast growth factor receptor signaling in kidney development. *Pediatric nephrology (Berlin, Germany)*, 26 (9), 1373–1379.

Batourina, E., Gim, S., Bello, N., Shy, M., Clagett-Dame, M., Srinivas, S., Costantini, F., and Mendelsohn, C., 2001. Vitamin A controls epithelial/mesenchymal interactions through Ret expression. *Nature genetics*, 27 (1), 74–78.

Battin, J., Lacombe, D., and Leng, J.J., 1993. Familial occurrence of hereditary renal adysplasia with Müllerian anomalies. *Clinical genetics*, 43 (1), 23–24.

Beaman, G.M., Galatà, G., Teik, K.W., Urquhart, J.E., Aishah, A., O'Sullivan, J., Bhaskar, S.S., Wood, K.A., Thomas, H.B., O'Keefe, R.T., Woolf, A.S., Stuart, H.M., and Newman, W.G., 2019. A homozygous missense variant in CHRM3 associated with familial urinary bladder disease. *Clinical genetics*, 96 (6), 515–520.

Bekheirnia, M.R., Bekheirnia, N., Bainbridge, M.N., Gu, S., Coban Akdemir, Z.H., Gambin, T., Janzen, N.K., Jhangiani, S.N., Muzny, D.M., Michael, M., Brewer, E.D., Elenberg, E., Kale, A.S., Riley, A.A., Swartz, S.J., Scott, D.A., Yang, Y., Srivaths, P.R., Wenderfer, S.E., Bodurtha, J., Applegate, C.D., Velinov, M., Myers, A., Borovik, L., Craigen, W.J., Hanchard, N.A., Rosenfeld, J.A., Lewis, R.A., Gonzales, E.T., Gibbs, R.A., Belmont, J.W., Roth, D.R., Eng, C., Braun, M.C., Lupski, J.R., and Lamb, D.J., 2017. Whole-exome sequencing in the molecular diagnosis of individuals with congenital anomalies of the kidney and urinary tract and identification of a new causative gene. *Genetics in medicine: official journal of the American College of Medical Genetics*, 19 (4), 412–420.

Belk, R.A., Thomas, D.F.M., Mueller, R.F., Godbole, P., Markham, A.F., and Weston, M.J., 2002. A family study and the natural history of prenatally detected unilateral multicystic dysplastic kidney. *The journal of urology*, 167 (2 Pt 1), 666–669.

Benjamini, Y. and Hochberg, Y., 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57 (1), 289–300.

Berger, H., Wodarz, A., and Borchers, A., 2017. PTK7 Faces the Wnt in Development and Disease. *Frontiers in cell and developmental biology*, 5, 31.

Bergmann, C., Fliegauf, M., Brüchle, N.O., Frank, V., Olbrich, H., Kirschner, J., Schermer, B., Schmedding, I., Kispert, A., Kränzlin, B., Nürnberg, G., Becker, C., Grimm, T., Girschick, G., Lynch, S.A., Kelehan, P., Senderek, J., Neuhaus, T.J., Stallmach, T., Zentgraf, H., Nürnberg, P., Gretz, N., Lo, C., Lienkamp, S., Schäfer, T., Walz, G., Benzing, T., Zerres, K., and Omran, H., 2008. Loss of nephrocystin-3 function can cause embryonic lethality, Meckel-Gruber-like syndrome, situs inversus, and renal-hepatic-pancreatic dysplasia. *The American Journal of Human Genetics*, 82 (4), 959–970.

Bernabeu, E., Canela-Xandri, O., Rawlik, K., Talenti, A., Prendergast, J., and Tenesa, A., 2020. Sexual differences in genetic architecture in UK Biobank. *bioRxiv*.

Berndt, S.I., Wang, Z., Yeager, M., Alavanja, M.C., Albanes, D., Amundadottir, L., Andriole, G., Beane Freeman, L., Campa, D., Cancel-Tassin, G., Canzian, F., Cornu, J.-N., Cussenot, O., Diver, W.R., Gapstur, S.M., Grönberg, H., Haiman, C.A., Henderson, B., Hutchinson, A., Hunter, D.J., Key, T.J., Kolb, S., Koutros, S., Kraft, P., Le Marchand, L., Lindström, S., Machiela, M.J., Ostrander, E.A., Riboli, E., Schumacher, F., Siddiq, A., Stanford, J.L., Stevens, V.L., Travis, R.C., Tsilidis, K.K., Virtamo, J., Weinstein, S., Wilkund, F., Xu, J., Lilly Zheng, S., Yu, K., Wheeler, W., Zhang, H., African Ancestry Prostate Cancer GWAS Consortium, Sampson, J., Black, A., Jacobs, K., Hoover, R.N., Tucker, M., and Chanock, S.J., 2015. Two susceptibility loci identified for prostate cancer aggressiveness. *Nature communications*, 6, 6889.

Beunders, G., van de Kamp, J., Vasudevan, P., Morton, J., Smets, K., Kleefstra, T., de Munnik, S.A., Schuurs-Hoeijmakers, J., Ceulemans, B., Zollino, M., Hoffjan, S., Wieczorek, S., So, J., Mercer, L., Walker, T., Velsher, L., Parker, M.J., Magee, A.C., Elffers, B., Kooy, R.F., Yntema, H.G., Meijers-Heijboer, E.J., Sistermans, E.A., and the DDD study, 2016. A detailed clinical analysis of 13 patients with AUTS2 syndrome further delineates the phenotypic spectrum and underscores the behavioural phenotype. *Journal of medical genetics*, 53 (8), 523–532.

Beunders, G., Voorhoeve, E., Golzio, C., Pardo, L.M., Rosenfeld, J.A., Talkowski, M.E., Simonic, I., Lionel, A.C., Vergult, S., Pyatt, R.E., van de Kamp, J., Nieuwint, A., Weiss, M.M., Rizzu, P., Verwer, L.E.N.I., van Spaendonk, R.M.L., Shen, Y., Wu, B.-L., Yu, T., Yu, Y., Chiang, C., Gusella, J.F., Lindgren, A.M., Morton, C.C., van Binsbergen, E., Bulk, S., van Rossem, E., Vanakker, O., Armstrong, R., Park, S.-M., Greenhalgh, L., Maye, U., Neill, N.J., Abbott, K.M., Sell, S., Ladda, R., Farber, D.M., Bader, P.I., Cushing, T., Drautz, J.M., Konczal, L., Nash, P., de Los Reyes, E., Carter, M.T., Hopkins, E., Marshall, C.R., Osborne, L.R., Gripp, K.W., Thrush, D.L., Hashimoto, S., Gastier-Foster, J.M., Astbury, C., Ylstra, B., Meijers-Heijboer, H., Posthuma, D., Menten, B., Mortier, G., Scherer, S.W., Eichler, E.E., Girirajan, S.,

Katsanis, N., Groffen, A.J., and Sistermans, E.A., 2013. Exonic deletions in AUTS2 cause a syndromic form of intellectual disability and suggest a critical role for the C terminus. *The American Journal of Human Genetics*, 92 (2), 210–220.

Bill, A., Schmitz, A., König, K., Heukamp, L.C., Hannam, J.S., and Famulok, M., 2012. Anti-proliferative effect of cytohesin inhibition in gefitinib-resistant lung cancer cells. *PloS one*, 7 (7), e41179.

Bingham, C., Bulman, M.P., Ellard, S., Allen, L.I., Lipkin, G.W., Hoff, W.G., Woolf, A.S., Rizzoni, G., Novelli, G., Nicholls, A.J., and Hattersley, A.T., 2001. Mutations in the hepatocyte nuclear factor-1beta gene are associated with familial hypoplastic glomerulocystic kidney disease. *The American Journal of Human Genetics*, 68 (1), 219–224.

Birney, E., Inouye, M., Raff, J., Rutherford, A., and Scally, A., 2021. The language of race, ethnicity, and ancestry in human genetic research. *arXiv [q-bio.PE]*.

Bocher, O. and Génin, E., 2020. Rare variant association testing in the non-coding genome. *Human genetics*, 139 (11), 1345–1362.

Boghossian, N.S., Sicko, R.J., Kay, D.M., Rigler, S.L., Caggana, M., Tsai, M.Y., Yeung, E.H., Pankratz, N., Cole, B.R., Druschel, C.M., Romitti, P.A., Browne, M.L., Fan, R., Liu, A., Brody, L.C., and Mills, J.L., 2016. Rare copy number variants implicated in posterior urethral valves. *American journal of medical genetics. Part A*, 170 (3), 622–633.

Bondeson, M.-L., Dahl, N., Malmgren, H., Kleijer, W.J., Tönnesen, T., Carlberg, B.-M., and Pettersson, U., 1995. Inversion of the IDS gene resulting from recombination with IDS-related sequences in a common cause of the Hunter syndrome. *Human molecular genetics*, 4 (4), 615–621.

Bower, M., Salomon, R., Allanson, J., Antignac, C., Benedicenti, F., Benetti, E., Binenbaum, G., Jensen, U.B., Cochat, P., DeCramer, S., Dixon, J., Drouin, R., Falk, M.J., Feret, H., Gise, R., Hunter, A., Johnson, K., Kumar, R., Lavocat, M.P., Martin, L., Morinière, V., Mowat, D., Murer, L., Nguyen, H.T., Peretz-Amit, G., Pierce, E., Place, E., Rodig, N., Salerno, A., Sastry, S., Sato, T., Sayer, J.A., Schaafsma, G.C.P., Shoemaker, L., Stockton, D.W., Tan, W.-H., Tenconi, R., Vanhille, P., Vats, A., Wang, X., Warman, B., Weleber, R.G., White, S.M., Wilson-Brackett, C., Zand, D.J., Eccles, M., Schimmenti, L.A., and Heidet, L., 2012. Update of PAX2 mutations in renal coloboma syndrome and establishment of a locus-specific database. *Human mutation*, 33 (3), 457–466.

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., Cherry, J.M., and Snyder, M., 2012. Annotation of functional variation in personal genomes using RegulomeDB. *Genome research*, 22 (9), 1790–1797.

Boyle, E.A., Li, Y.I., and Pritchard, J.K., 2017. An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169 (7), 1177–1186.

Brandler, W.M., Antaki, D., Gujral, M., Kleiber, M.L., Whitney, J., Maile, M.S., Hong, O., Chapman, T.R., Tan, S., Tandon, P., Pang, T., Tang, S.C., Vaux, K.K.,

Yang, Y., Harrington, E., Juul, S., Turner, D.J., Thiruvahindrapuram, B., Kaur, G., Wang, Z., Kingsmore, S.F., Gleeson, J.G., Bisson, D., Kakaradov, B., Telenti, A., Venter, J.C., Corominas, R., Toma, C., Cormand, B., Rueda, I., Guijarro, S., Messer, K.S., Nievergelt, C.M., Arranz, M.J., Courchesne, E., Pierce, K., Muotri, A.R., Iakoucheva, L.M., Hervas, A., Scherer, S.W., Corsello, C., and Sebat, J., 2018. Paternally inherited cis-regulatory structural variants are associated with autism. *Science (New York, N.Y.)*, 360 (6386), 327–331.

Briggs, C.E., Guo, C.-Y., Schoettler, C., Rosoklija, I., Silva, A., Bauer, S.B., Retik, A.B., Kunkel, L., and Nguyen, H.T., 2010. A genome scan in affected sib-pairs with familial vesicoureteral reflux identifies a locus on chromosome 5. *European journal of human genetics: EJHG*, 18 (2), 245–250.

Brockschmidt, A., Chung, B., Weber, S., Fischer, D.-C., Kolatsi-Joannou, M., Christ, L., Heimbach, A., Shtiza, D., Klaus, G., Simonetti, G.D., Konrad, M., Winyard, P., Haffner, D., Schaefer, F., and Weber, R.G., 2012. CHD1L: a new candidate gene for congenital anomalies of the kidneys and urinary tract (CAKUT). *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association*, 27 (6), 2355–2364.

Brophy, P.D., Rasmussen, M., Parida, M., Bonde, G., Darbro, B.W., Hong, X., Clarke, J.C., Peterson, K.A., Denegre, J., Schneider, M., Sussman, C.R., Sunde, L., Lildballe, D.L., Hertz, J.M., Cornell, R.A., Murray, S.A., and Manak, J.R., 2017. A gene implicated in activation of retinoic acid receptor targets is a novel renal agenesis gene in humans. *Genetics*, 207 (1), 215–228.

Brownlee, E., Wragg, R., Robb, A., Chandran, H., Knight, M., McCarthy, L., and BAPS-CASS, 2019. Current epidemiology and antenatal presentation of posterior urethral valves: Outcome of BAPS CASS National Audit. *Journal of pediatric surgery*, 54 (2), 318–321.

Bruneau, B.G., Nemer, G., Schmitt, J.P., Charron, F., Robitaille, L., Caron, S., Conner, D.A., Gessler, M., Nemer, M., Seidman, C.E., and Seidman, J.G., 2001. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell*, 106 (6), 709–721.

Brunetti-Pierri, N., Del Gaudio, D., Peters, H., Justino, H., Ott, C.-E., Mundlos, S., and Bacino, C.A., 2008. Robinow syndrome: phenotypic variability in a family with a novel intragenic ROR2 mutation. *American journal of medical genetics. Part A*, 146A (21), 2804–2809.

Brzóska, H.Ł., d'Esposito, A.M., Kolatsi-Joannou, M., Patel, V., Igarashi, P., Lei, Y., Finnell, R.H., Lythgoe, M.F., Woolf, A.S., Papakrivopoulou, E., and Long, D.A., 2016. Planar cell polarity genes Celsr1 and Vangl2 are necessary for kidney growth, differentiation, and rostrocaudal patterning. *Kidney international*, 90 (6), 1274–1284.

Bulum, B., Birsin Özçakar, Z., Üstüner, E., Düşünceli, E., Kavaz, A., Duman, D., Walz, K., Fitoz, S., Tekin, M., and Yalçınkaya, F., 2013. High frequency of

kidney and urinary tract anomalies in asymptomatic first-degree relatives of patients with CAKUT. *Pediatric Nephrology*.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., and Marchini, J., 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature*, 562 (7726), 203–209.

Calderon-Margalit, R., Efron, G., Pleniceanu, O., Tzur, D., Stern-Zimmer, M., Afek, A., Erlich, T., Derazne, E., Kark, J.D., Keinan-Boker, L., Twig, G., and Vivante, A., 2021. Congenital anomalies of the kidney and urinary tract and adulthood risk of urinary tract cancer. *Kidney international reports*, 6 (4), 946–952.

Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., and Lewis, W.H., 1990. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell*, 60 (3), 509–520.

Cameron, D.L., Di Stefano, L., and Papenfuss, A.T., 2019. Comprehensive evaluation and characterisation of short read general-purpose structural variant calling software. *Nature communications*, 10 (1), 3240.

Cao, J., O'Day, D.R., Pliner, H.A., Kingsley, P.D., Deng, M., Daza, R.M., Zager, M.A., Aldinger, K.A., Blecher-Gonen, R., Zhang, F., Spielmann, M., Palis, J., Doherty, D., Steemers, F.J., Glass, I.A., Trapnell, C., and Shendure, J., 2020. A human cell atlas of fetal gene expression. *Science (New York, N.Y.)*, 370 (6518), eaba7721.

Carlson, C.S., Matise, T.C., North, K.E., Haiman, C.A., Fesinmeyer, M.D., Buyske, S., Schumacher, F.R., Peters, U., Franceschini, N., Ritchie, M.D., Duggan, D.J., Spencer, K.L., Dumitrescu, L., Eaton, C.B., Thomas, F., Young, A., Carty, C., Heiss, G., Le Marchand, L., Crawford, D.C., Hindorff, L.A., Kooperberg, C.L., and PAGE Consortium, 2013. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS biology*, 11 (9), e1001661.

Carter, C.O., Evans, K., and Pescia, G., 1979. A family study of renal agenesis. *Journal of medical genetics*, 16 (3), 176–188.

Caruana, G., Wong, M.N., Walker, A., Heloury, Y., Webb, N., Johnstone, L., James, P.A., Burgess, T., and Bertram, J.F., 2015. Copy-number variation associated with congenital anomalies of the kidney and urinary tract. *Pediatric nephrology* , 30 (3), 487–495.

Carvalho, C.M.B. and Lupski, J.R., 2016. Mechanisms underlying structural variant formation in genomic disorders. *Nature reviews. Genetics*, 17 (4), 224–238.

Castel, S.E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A., Guigo, R., Iossifov, I., Vasileva, A., and Lappalainen, T., 2018. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nature genetics*, 50 (9), 1327–1334.

Cavazos, T.B. and Witte, J.S., 2021. Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *HGG advances*, 2 (1), 100017.

Cervellione, R.M., Mantovani, A., Gearhart, J., Bogaert, G., Gobet, R., Caione, P., and Dickson, A.P., 2015. Prospective study on the incidence of bladder/cloacal exstrophy and epispadias in Europe. *Journal of pediatric urology*, 11 (6), 337.e1–6.

Chaisson, M.J.P., Sanders, A.D., Zhao, X., Malhotra, A., Porubsky, D., Rausch, T., Gardner, E.J., Rodriguez, O.L., Guo, L., Collins, R.L., Fan, X., Wen, J., Handsaker, R.E., Fairley, S., Kronenberg, Z.N., Kong, X., Hormozdiari, F., Lee, D., Wenger, A.M., Hastie, A.R., Antaki, D., Anantharaman, T., Audano, P.A., Brand, H., Cantsilieris, S., Cao, H., Cerveira, E., Chen, C., Chen, X., Chin, C.-S., Chong, Z., Chuang, N.T., Lambert, C.C., Church, D.M., Clarke, L., Farrell, A., Flores, J., Galeev, T., Gorkin, D.U., Gujral, M., Guryev, V., Heaton, W.H., Korlach, J., Kumar, S., Kwon, J.Y., Lam, E.T., Lee, J.E., Lee, J., Lee, W.-P., Lee, S.P., Li, S., Marks, P., Viaud-Martinez, K., Meiers, S., Munson, K.M., Navarro, F.C.P., Nelson, B.J., Nodzak, C., Noor, A., Kyriazopoulou-Panagiotopoulou, S., Pang, A.W.C., Qiu, Y., Rosanio, G., Ryan, M., Stütz, A., Spierings, D.C.J., Ward, A., Welch, A.E., Xiao, M., Xu, W., Zhang, C., Zhu, Q., Zheng-Bradley, X., Lowy, E., Yakneen, S., McCarroll, S., Jun, G., Ding, L., Koh, C.L., Ren, B., Flicek, P., Chen, K., Gerstein, M.B., Kwok, P.-Y., Lansdorp, P.M., Marth, G.T., Sebat, J., Shi, X., Bashir, A., Ye, K., Devine, S.E., Talkowski, M.E., Mills, R.E., Marschall, T., Korbel, J.O., Eichler, E.E., and Lee, C., 2019. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nature communications*, 10 (1), 1784.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*.

Chen, H., Tian, Y., Shu, W., Bo, X., and Wang, S., 2012. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PloS one*, 7 (7), e41374.

Chen, H., Wang, C., Conomos, M.P., Stilp, A.M., Li, Z., Sofer, T., Szpiro, A.A., Chen, W., Brehm, J.M., Celedón, J.C., Redline, S., Papanicolaou, G.J., Thornton, T.A., Laurie, C.C., Rice, K., and Lin, X., 2016. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *American journal of human genetics*, 98 (4), 653–666.

Chen, J., Spracklen, C.N., Marenne, G., Varshney, A., Corbin, L.J., Luan, J., Willems, S.M., Wu, Y., Zhang, X., Horikoshi, M., Boutin, T.S., Mägi, R., Waage, J., Li-Gao, R., Chan, K.H.K., Yao, J., Anasanti, M.D., Chu, A.Y., Claringbould, A., Heikkinen, J., Hong, J., Hottenga, J.-J., Huo, S., Kaakinen, M.A., Louie, T., März, W., Moreno-Macias, H., Ndungu, A., Nelson, S.C., Nolte, I.M., North, K.E., Raulerson, C.K., Ray, D., Rohde, R., Rybin, D., Schurmann, C., Sim, X., Southam, L., Stewart, I.D., Wang, C.A., Wang, Y.,

Wu, P., Zhang, W., Ahluwalia, T.S., Appel, E.V.R., Bielak, L.F., Brody, J.A., Burtt, N.P., Cabrera, C.P., Cade, B.E., Chai, J.F., Chai, X., Chang, L.-C., Chen, C.-H., Chen, B.H., Chitrala, K.N., Chiu, Y.-F., de Haan, H.G., Delgado, G.E., Demirkan, A., Duan, Q., Engmann, J., Fatumo, S.A., Gayán, J., Giulianini, F., Gong, J.H., Gustafsson, S., Hai, Y., Hartwig, F.P., He, J., Heianza, Y., Huang, T., Huerta-Chagoya, A., Hwang, M.Y., Jensen, R.A., Kawaguchi, T., Kentistou, K.A., Kim, Y.J., Kleber, M.E., Kooner, I.K., Lai, S., Lange, L.A., Langefeld, C.D., Lauzon, M., Li, M., Ligthart, S., Liu, J., Loh, M., Long, J., Lyssenko, V., Mangino, M., Marzi, C., Montasser, M.E., Nag, A., Nakatochi, M., Noce, D., Noordam, R., Pistis, G., Preuss, M., Raffield, L., Rasmussen-Torvik, L.J., Rich, S.S., Robertson, N.R., Rueedi, R., Ryan, K., Sanna, S., Saxena, R., Schraut, K.E., Sennblad, B., Setoh, K., Smith, A.V., Sparsø, T., Strawbridge, R.J., Takeuchi, F., Tan, J., Trompet, S., van den Akker, E., van der Most, P.J., Verweij, N., Vogel, M., Wang, H., Wang, C., Wang, N., Warren, H.R., Wen, W., Wilsgaard, T., Wong, A., Wood, A.R., Xie, T., Zafarmand, M.H., Zhao, J.-H., Zhao, W., Amin, N., Arzumanyan, Z., Astrup, A., Bakker, S.J.L., Baldassarre, D., Beekman, M., Bergman, R.N., Bertoni, A., Blüher, M., Bonnycastle, L.L., Bornstein, S.R., Bowden, D.W., Cai, Q., Campbell, A., Campbell, H., Chang, Y.C., de Geus, E.J.C., Dehghan, A., Du, S., Eiriksdottir, G., Farmaki, A.E., Frånberg, M., Fuchsberger, C., Gao, Y., Gjesing, A.P., Goel, A., Han, S., Hartman, C.A., Herder, C., Hicks, A.A., Hsieh, C.-H., Hsueh, W.A., Ichihara, S., Igase, M., Ikram, M.A., Johnson, W.C., Jørgensen, M.E., Joshi, P.K., Kalyani, R.R., Kandeel, F.R., Katsuya, T., Khor, C.C., Kiess, W., Kolcic, I., Kuulasmaa, T., Kuusisto, J., Läll, K., Lam, K., Lawlor, D.A., Lee, N.R., Lemaitre, R.N., Li, H., Lifelines Cohort Study, Lin, S.-Y., Lindström, J., Linneberg, A., Liu, J., Lorenzo, C., Matsubara, T., Matsuda, F., Mingrone, G., Mooijaart, S., Moon, S., Nabika, T., Nadkarni, G.N., Nadler, J.L., Nelis, M., Neville, M.J., Norris, J.M., Ohyagi, Y., Peters, A., Peyser, P.A., Polasek, O., Qi, Q., Raven, D., Reilly, D.F., Reiner, A., Rivideneira, F., Roll, K., Rudan, I., Sabanayagam, C., Sandow, K., Sattar, N., Schürmann, A., Shi, J., Stringham, H.M., Taylor, K.D., Teslovich, T.M., Thuesen, B., Timmers, P.R.H.J., Tremoli, E., Tsai, M.Y., Uitterlinden, A., van Dam, R.M., van Heemst, D., van Hylckama Vlieg, A., van Vliet-Ostaptchouk, J.V., Vangipurapu, J., Vestergaard, H., Wang, T., Willems van Dijk, K., Zemunik, T., Abecasis, G.R., Adair, L.S., Aguilar-Salinas, C.A., Alarcón-Riquelme, M.E., An, P., Aviles-Santa, L., Becker, D.M., Beilin, L.J., Bergmann, S., Bisgaard, H., Black, C., Boehnke, M., Boerwinkle, E., Böhm, B.O., Bønnelykke, K., Boomsma, D.I., Bottinger, E.P., Buchanan, T.A., Canouil, M., Caulfield, M.J., Chambers, J.C., Chasman, D.I., Chen, Y.-D.I., Cheng, C.-Y., Collins, F.S., Correa, A., Cucca, F., de Silva, H.J., Dedoussis, G., Elmståhl, S., Evans, M.K., Ferrannini, E., Ferrucci, L., Florez, J.C., Franks, P.W., Frayling, T.M., Froguel, P., Gigante, B., Goodarzi, M.O., Gordon-Larsen, P., Grallert, H., Grarup, N., Grimsgaard, S., Groop, L., Gudnason, V., Guo, X., Hamsten, A., Hansen, T., Hayward, C., Heckbert, S.R., Horta, B.L.,

Huang, W., Ingelsson, E., James, P.S., Jarvelin, M.-R., Jonas, J.B., Jukema, J.W., Kaleebu, P., Kaplan, R., Kardia, S.L.R., Kato, N., Keinanen-Kiukaanniemi, S.M., Kim, B.-J., Kivimaki, M., Koistinen, H.A., Kooner, J.S., Körner, A., Kovacs, P., Kuh, D., Kumari, M., Kutalik, Z., Laakso, M., Lakka, T.A., Launer, L.J., Leander, K., Li, H., Lin, X., Lind, L., Lindgren, C., Liu, S., Loos, R.J.F., Magnusson, P.K.E., Mahajan, A., Metspalu, A., Mook-Kanamori, D.O., Mori, T.A., Munroe, P.B., Njølstad, I., O'Connell, J.R., Oldehinkel, A.J., Ong, K.K., Padmanabhan, S., Palmer, C.N.A., Palmer, N.D., Pedersen, O., Pennell, C.E., Porteous, D.J., Pramstaller, P.P., Province, M.A., Psaty, B.M., Qi, L., Raffel, L.J., Rauramaa, R., Redline, S., Ridker, P.M., Rosendaal, F.R., Saaristo, T.E., Sandhu, M., Saramies, J., Schneiderman, N., Schwarz, P., Scott, L.J., Selvin, E., Sever, P., Shu, X.-O., Slagboom, P.E., Small, K.S., Smith, B.H., Snieder, H., Sofer, T., Sørensen, T.I.A., Spector, T.D., Stanton, A., Steves, C.J., Stumvoll, M., Sun, L., Tabara, Y., Tai, E.S., Timpson, N.J., Tönjes, A., Tuomilehto, J., Tusie, T., Uusitupa, M., van der Harst, P., van Duijn, C., Vitart, V., Vollenweider, P., Vrijkotte, T.G.M., Wagenknecht, L.E., Walker, M., Wang, Y.X., Wareham, N.J., Watanabe, R.M., Watkins, H., Wei, W.B., Wickremasinghe, A.R., Willemsen, G., Wilson, J.F., Wong, T.-Y., Wu, J.-Y., Xiang, A.H., Yanek, L.R., Yengo, L., Yokota, M., Zeggini, E., Zheng, W., Zonderman, A.B., Rotter, J.I., Gloyn, A.L., McCarthy, M.I., Dupuis, J., Meigs, J.B., Scott, R.A., Prokopenko, I., Leong, A., Liu, C.-T., Parker, S.C.J., Mohlke, K.L., Langenberg, C., Wheeler, E., Morris, A.P., Barroso, I., and Meta-Analysis of Glucose and Insulin-related Traits Consortium (MAGIC), 2021. The trans-ancestral genomic architecture of glycemic traits. *Nature genetics*, 53 (6), 840–860.

Chen, X., Ke, Y., Wu, K., Zhao, H., Sun, Y., Gao, L., Liu, Z., Zhang, J., Tao, W., Hou, Z., Liu, H., Liu, J., and Chen, Z.-J., 2019. Key role for CTCF in establishing chromatin structure in human embryos. *Nature*, 576 (7786), 306–310.

Chen, X., Schulz-Trieglaff, O., Shaw, R., Barnes, B., Schlesinger, F., Källberg, M., Cox, A.J., Kruglyak, S., and Saunders, C.T., 2016. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* , 32 (8), 1220–1222.

Chen, Y.-Z., Gao, Q., Zhao, X.-Z., Chen, Y.-Z., Bennett, C.L., Xiong, X.-S., Mei, C.-L., Shi, Y.-Q., and Chen, X.-M., 2010. Systematic review of TCF2 anomalies in renal cysts and diabetes syndrome/maturity onset diabetes of the young type 5. *Chinese medical journal*, 123 (22), 3326–3333.

Chesnaye, N.C., van Stralen, K.J., Bonthuis, M., Harambat, J., Groothoff, J.W., and Jager, K.J., 2018. Survival in children requiring chronic renal replacement therapy. *Pediatric nephrology (Berlin, Germany)*, 33 (4), 585–594.

Chevalier, R.L., Thornhill, B.A., Forbes, M.S., and Kiley, S.C., 2010. Mechanisms of renal injury and progression of renal disease in congenital obstructive nephropathy. *Pediatric nephrology (Berlin, Germany)*, 25 (4), 687–697.

Chia, I., Grote, D., Marcotte, M., Batourina, E., Mendelsohn, C., and Bouchard, M., 2011. Nephric duct insertion is a crucial step in urinary tract maturation that

is regulated by a Gata3-Raldh2-Ret molecular network in mice. *Development (Cambridge, England)*, 138 (10), 2089–2097.

Chiaramonte, C., Bommarito, D., Zambaiti, E., Antona, V., and Li Voti, G., 2016. Genetic Basis of Posterior Urethral Valves Inheritance. *Urology*, 95, 175–179.

Chu, A.Y., Tin, A., Schlosser, P., Ko, Y.-A., Qiu, C., Yao, C., Joehanes, R., Grams, M.E., Liang, L., Gluck, C.A., Liu, C., Coresh, J., Hwang, S.-J., Levy, D., Boerwinkle, E., Pankow, J.S., Yang, Q., Fornage, M., Fox, C.S., Susztak, K., and Köttgen, A., 2017. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nature communications*, 8 (1), 1286.

Cirulli, E.T., White, S., Read, R.W., Elhanan, G., Metcalf, W.J., Tanudjaja, F., Fath, D.M., Sandoval, E., Isaksson, M., Schlauch, K.A., Grzymski, J.J., Lu, J.T., and Washington, N.L., 2020. Genome-wide rare variant analysis for thousands of phenotypes in over 70,000 exomes from two cohorts. *Nature communications*, 11 (1), 542.

Claussnitzer, M., Dankel, S.N., Kim, K.-H., Quon, G., Meuleman, W., Haugen, C., Glunk, V., Sousa, I.S., Beaudry, J.L., Puviindran, V., Abdennur, N.A., Liu, J., Svensson, P.-A., Hsu, Y.-H., Drucker, D.J., Mellgren, G., Hui, C.-C., Hauner, H., and Kellis, M., 2015. FTO obesity variant circuitry and adipocyte browning in humans. *The New England journal of medicine*, 373 (10), 895–907.

Clissold, R.L., Hamilton, A.J., Hattersley, A.T., Ellard, S., and Bingham, C., 2015. HNF1B-associated renal and extra-renal disease—an expanding clinical spectrum. *Nature reviews. Nephrology*, 11 (2), 102–112.

Cochran, J.N., Geier, E.G., Bonham, L.W., Newberry, J.S., Amaral, M.D., Thompson, M.L., Lasseigne, B.N., Karydas, A.M., Roberson, E.D., Cooper, G.M., Rabinovici, G.D., Miller, B.L., Myers, R.M., Yokoyama, J.S., and Alzheimer's Disease Neuroimaging Initiative, 2020. Non-coding and loss-of-function coding variants in TET2 are associated with multiple neurodegenerative diseases. *The American Journal of Human Genetics*, 106 (5), 632–645.

Cohen, J., Pertsemlidis, A., Kotowski, I.K., Graham, R., Garcia, C.K., and Hobbs, H.H., 2005. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nature genetics*, 37 (2), 161–165.

Collins, R.L., Brand, H., Karczewski, K.J., Zhao, X., Alföldi, J., Francioli, L.C., Khera, A.V., Lowther, C., Gauthier, L.D., Wang, H., Watts, N.A., Solomonson, M., O'Donnell-Luria, A., Baumann, A., Munshi, R., Walker, M., Whelan, C.W., Huang, Y., Brookings, T., Sharpe, T., Stone, M.R., Valkanas, E., Fu, J., Tiao, G., Laricchia, K.M., Ruano-Rubio, V., Stevens, C., Gupta, N., Cusick, C., Margolin, L., Genome Aggregation Database Production Team, Genome Aggregation Database Consortium, Taylor, K.D., Lin, H.J., Rich, S.S., Post, W.S., Chen, Y.-D.I., Rotter, J.I., Nusbaum, C., Philippakis, A., Lander, E., Gabriel, S., Neale, B.M., Kathiresan, S., Daly, M.J., Banks, E., MacArthur,

D.G., and Talkowski, M.E., 2020. A structural variation reference for medical and population genetics. *Nature*, 581 (7809), 444–451.

Combes, P., Planche, V., Eymard-Pierre, E., Sarret, C., Rodriguez, D., Boespflug-Tanguy, O., and Vaurs-Barriere, C., 2012. Relevance of SOX17 variants for hypomyelinating leukodystrophies and congenital anomalies of the kidney and urinary tract (CAKUT). *Annals of human genetics*, 76 (3), 261–267.

Connaughton, D.M., Dai, R., Owen, D.J., Marquez, J., Mann, N., Graham-Paquin, A.L., Nakayama, M., Coyaud, E., Laurent, E.M.N., St-Germain, J.R., Blok, L.S., Vino, A., Klämbt, V., Deutsch, K., Wu, C.-H.W., Kolvenbach, C.M., Kause, F., Ottlewski, I., Schneider, R., Kitzler, T.M., Majmundar, A.J., Buerger, F., Onuchic-Whitford, A.C., Youying, M., Kolb, A., Salmanullah, D., Chen, E., van der Ven, A.T., Rao, J., Ityel, H., Seltzsam, S., Rieke, J.M., Chen, J., Vivante, A., Hwang, D.-Y., Kohl, S., Dworschak, G.C., Hermle, T., Alders, M., Bartolomaeus, T., Bauer, S.B., Baum, M.A., Brilstra, E.H., Challman, T.D., Zyskind, J., Costin, C.E., Dipple, K.M., Duijkers, F.A., Ferguson, M., Fitzpatrick, D.R., Fick, R., Glass, I.A., Hulick, P.J., Kline, A.D., Krey, I., Kumar, S., Lu, W., Marco, E.J., Wentzensen, I.M., Mefford, H.C., Platzer, K., Povolotskaya, I.S., Savatt, J.M., Shcherbakova, N.V., Senguttuvan, P., Squire, A.E., Stein, D.R., Thiffault, I., Voinova, V.Y., Somers, M.J.G., Ferguson, M.A., Traum, A.Z., Daouk, G.H., Daga, A., Rodig, N.M., Terhal, P.A., van Binsbergen, E., Eid, L.A., Tasic, V., Rasouly, H.M., Lim, T.Y., Ahram, D.F., Gharavi, A.G., Reutter, H.M., Rehm, H.L., MacArthur, D.G., Lek, M., Laricchia, K.M., Lifton, R.P., Xu, H., Mane, S.M., Sanna-Cherchi, S., Sharrocks, A.D., Raught, B., Fisher, S.E., Bouchard, M., Khokha, M.K., Shril, S., and Hildebrandt, F., 2020. Mutations of the transcriptional corepressor ZMYM2 cause syndromic urinary tract malformations. *The American Journal of Human Genetics*, 107 (4), 727–742.

Connaughton, D.M., Kennedy, C., Shril, S., Mann, N., Murray, S.L., Williams, P.A., Conlon, E., Nakayama, M., van der Ven, A.T., Ityel, H., Kause, F., Kolvenbach, C.M., Dai, R., Vivante, A., Braun, D.A., Schneider, R., Kitzler, T.M., Moloney, B., Moran, C.P., Smyth, J.S., Kennedy, A., Benson, K., Stapleton, C., Denton, M., Magee, C., O'Seaghdha, C.M., Plant, W.D., Griffin, M.D., Awan, A., Sweeney, C., Mane, S.M., Lifton, R.P., Griffin, B., Leavey, S., Casserly, L., de Freitas, D.G., Holian, J., Dorman, A., Doyle, B., Lavin, P.J., Little, M.A., Conlon, P.J., and Hildebrandt, F., 2019. Monogenic causes of chronic kidney disease in adults. *Kidney international*, 95 (4), 914–928.

Connolly, L.P., Treves, S.T., Connolly, S.A., Zurakowski, D., Share, J.C., Bar-Sever, Z., Mitchell, K.D., and Bauer, S.B., 1997. Vesicoureteral reflux in children: incidence and severity in siblings. *The journal of urology*, 157 (6), 2287–2290.

Consortium, I.H.G.S. and International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. *Nature*.

Conte, M.L., Bertoli-Avella, A.M., de Graaf, B.M., Punzo, F., Lama, G., La Manna, A., Grassia, C., Rambaldi, P.F., Oostra, B.A., and Perrotta, S., 2008. A genome search for primary vesicoureteral reflux shows further evidence for

genetic heterogeneity. *Pediatric nephrology (Berlin, Germany)*, 23 (4), 587–595.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., Abdel-Hamid, H., Bader, P., McCracken, E., Niyazov, D., Leppig, K., Thiese, H., Hummel, M., Alexander, N., Gorski, J., Kussmann, J., Shashi, V., Johnson, K., Rehder, C., Ballif, B.C., Shaffer, L.G., and Eichler, E.E., 2011. A copy number variation morbidity map of developmental delay. *Nature genetics*, 43 (9), 838–846.

Coram, M.A., Duan, Q., Hoffmann, T.J., Thornton, T., Knowles, J.W., Johnson, N.A., Ochs-Balcom, H.M., Donlon, T.A., Martin, L.W., Eaton, C.B., Robinson, J.G., Risch, N.J., Zhu, X., Kooperberg, C., Li, Y., Reiner, A.P., and Tang, H., 2013. Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. *American journal of human genetics*, 92 (6), 904–916.

Cordell, H.J., Darlay, R., Charoen, P., Stewart, A., Gullett, A.M., Lambert, H.J., Malcolm, S., Feather, S.A., Goodship, T.H.J., Woolf, A.S., Kenda, R.B., Goodship, J.A., and UK VUR Study Group, 2010. Whole-genome linkage and association scan in primary, nonsyndromic vesicoureteric reflux. *Journal of the American Society of Nephrology: JASN*, 21 (1), 113–123.

Cornwell, L.B., Ingulli, E.G., Mason, M.D., Ewing, E., and Riddell, J.V., 2021. Renal Transplants Due to Congenital Anomalies of the Kidney and Urinary Tract (CAKUT) Have Better Graft Survival Than Non-CAKUT Controls: Analysis of Over 10,000 Patients. *Urology*.

Corvol, H., Blackman, S.M., Boëlle, P.-Y., Gallins, P.J., Pace, R.G., Stonebraker, J.R., Accurso, F.J., Clement, A., Collaco, J.M., Dang, H., Dang, A.T., Franca, A., Gong, J., Guillot, L., Keenan, K., Li, W., Lin, F., Patrone, M.V., Raraigh, K.S., Sun, L., Zhou, Y.-H., O'Neal, W.K., Sontag, M.K., Levy, H., Durie, P.R., Rommens, J.M., Drumm, M.L., Wright, F.A., Strug, L.J., Cutting, G.R., and Knowles, M.R., 2015. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. *Nature communications*, 6 (1), 8382.

Costain, G., Jobling, R., Walker, S., Reuter, M.S., Snell, M., Bowdin, S., Cohn, R.D., Dupuis, L., Hewson, S., Mercimek-Andrews, S., Shuman, C., Sondheimer, N., Weksberg, R., Yoon, G., Meyn, M.S., Stavropoulos, D.J., Scherer, S.W., Mendoza-Londono, R., and Marshall, C.R., 2018. Periodic reanalysis of whole-genome sequencing data enhances the diagnostic advantage over standard clinical genetic testing. *European journal of human genetics: EJHG*, 26 (5), 740–744.

Crump, C., Sundquist, J., Winkleby, M.A., and Sundquist, K., 2019. Preterm birth and risk of chronic kidney disease from childhood into mid-adulthood: national cohort study. *BMJ (Clinical research ed.)*, 365, l1346.

Dai, L., Li, J., Xie, L., Wang, W., Lu, Y., Xie, M., Huang, J., Shen, K., Yang, H., Pei, C., Zhao, Y., and Zhang, W., 2021. A biallelic frameshift mutation in

nephronectin causes bilateral renal agenesis in humans. *Journal of the American Society of Nephrology: JASN*, 32 (8), 1871–1879.

Daly, S.B., Urquhart, J.E., Hilton, E., McKenzie, E.A., Kammerer, R.A., Lewis, M., Kerr, B., Stuart, H., Donnai, D., Long, D.A., Burgu, B., Aydogdu, O., Derbent, M., Garcia-Minaur, S., Reardon, W., Gener, B., Shalev, S., Smith, R., Woolf, A.S., Black, G.C., and Newman, W.G., 2010. Mutations in HPSE2 cause urofacial syndrome. *The American Journal of Human Genetics*, 87 (2), 309.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., and Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, 10 (2).

Darlow, J.M., Darlay, R., Dobson, M.G., Stewart, A., Charoen, P., Southgate, J., Baker, S.C., Xu, Y., Hunziker, M., Lambert, H.J., Green, A.J., Santibanez-Koref, M., Sayer, J.A., Goodship, T.H.J., Puri, P., Woolf, A.S., Kenda, R.B., Barton, D.E., and Cordell, H.J., 2017. Genome-wide linkage and association study implicates the 10q26 region as a major genetic contributor to primary nonsyndromic vesicoureteric reflux. *Scientific reports*, 7 (1), 14595.

Darlow, J.M., Dobson, M.G., Darlay, R., Molony, C.M., Hunziker, M., Green, A.J., Cordell, H.J., Puri, P., and Barton, D.E., 2014. A new genome scan for primary nonsyndromic vesicoureteric reflux emphasizes high genetic heterogeneity and shows linkage and association with various genes already implicated in urinary tract development. *Molecular genetics & genomic medicine*, 2 (1), 7–29.

Dathe, K., Kjaer, K.W., Brehm, A., Meinecke, P., Nürnberg, P., Neto, J.C., Brunoni, D., Tommerup, N., Ott, C.E., Klopocki, E., Seemann, P., and Mundlos, S., 2009. Duplications involving a conserved regulatory element downstream of BMP2 are associated with brachydactyly type A2. *American journal of human genetics*, 84 (4), 483–492.

D'Cruz, R., Stronks, K., Rowan, C.J., and Rosenblum, N.D., 2020. Lineage-specific roles of hedgehog-GLI signaling during mammalian kidney development. *Pediatric nephrology (Berlin, Germany)*, 35 (5), 725–731.

De Tomasi, L., David, P., Humbert, C., Silbermann, F., Arrondel, C., Tores, F., Fouquet, S., Desgrange, A., Niel, O., Bole-Feysot, C., Nitschké, P., Roume, J., Cordier, M.-P., Pietrement, C., Isidor, B., Khau Van Kien, P., Gonzales, M., Saint-Frison, M.-H., Martinovic, J., Novo, R., Piard, J., Cabrol, C., Verma, I.C., Puri, R., Journel, H., Aziza, J., Gavard, L., Said-Menthon, M.-H., Heidet, L., Saunier, S., and Jeanpierre, C., 2017. Mutations in GREB1L cause bilateral kidney agenesis in humans and mice. *The American Journal of Human Genetics*, 101 (5), 803–814.

Deaton, A.M., Parker, M.M., Ward, L.D., Flynn-Carroll, A.O., BonDurant, L., Hinkle, G., Akbari, P., Lotta, L.A., Baras, A., Nioi, P., Regeneron Genetics Center, and DiscovEHR Collaboration, 2021. Gene-level analysis of rare variants in 363,977 whole exome sequences identifies an association of GIGYF1 loss of function with type 2 diabetes. *bioRxiv*.

Deegan, D.F. and Engel, N., 2019. Sexual dimorphism in the age of genomics: How, when, where. *Frontiers in cell and developmental biology*, 7, 186.

DeFoor, W., Clark, C., Jackson, E., Reddy, P., Minevich, E., and Sheldon, C., 2008. Risk factors for end stage renal disease in children with posterior urethral valves. *The Journal of urology*, 180 (4 Suppl), 1705–8; discussion 1708.

Demirkan, H., 2021. An Unusual Urological Manifestation of Williams-Beuren Syndrome: Posterior Urethral Valve. *Urologia internationalis*, 105 (1–2), 159–162.

Dempke, W.C.M., Fenchel, K., Uciechowski, P., and Chevassut, T., 2017. Targeting developmental pathways: The Achilles heel of cancer? *Oncology*, 93 (4), 213–223.

Dennis, M.Y. and Eichler, E.E., 2016. Human adaptation and evolution by segmental duplication. *Current opinion in genetics & development*, 41, 44–52.

Derkach, A., Zhang, H., and Chatterjee, N., 2018. Power Analysis for Genetic Association Test (PAGEANT) provides insights to challenges for rare variant association studies. *Bioinformatics* , 34 (9), 1506–1513.

Devlin, B. and Roeder, K., 1999. Genomic control for association studies. *Biometrics*, 55 (4), 997–1004.

Devriendt, K., Fryns, J.P., Mortier, G., van Thienen, M.N., and Keymolen, K., 1998. The annual incidence of DiGeorge/velocardiofacial syndrome. *Journal of medical genetics*, 35 (9), 789–790.

Dey, R., Schmidt, E.M., Abecasis, G.R., and Lee, S., 2017. A Fast and Accurate Algorithm to Test for Binary Phenotypes and Its Application to PheWAS. *American journal of human genetics*, 101 (1), 37–49.

D'haene, E. and Vergult, S., 2021. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genetics in medicine: official journal of the American College of Medical Genetics*, 23 (1), 34–46.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V.V., Ecker, J.R., Thomson, J.A., and Ren, B., 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518 (7539), 331–336.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B., 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485 (7398), 376–380.

Domcke, S., Hill, A.J., Daza, R.M., Cao, J., O'Day, D.R., Pliner, H.A., Aldinger, K.A., Pokholok, D., Zhang, F., Milbank, J.H., Zager, M.A., Glass, I.A., Steemers, F.J., Doherty, D., Trapnell, C., Cusanovich, D.A., and Shendure, J., 2020. A human cell atlas of fetal chromatin accessibility. *Science (New York, N.Y.)*, 370 (6518), eaba7612.

Doraiswamy, N.V., Al Badr, M.S., and Freeman, N.V., 1983. Posterior urethral valves in siblings. *British journal of urology*, 55 (4), 448–449.

Draaken, M., Knapp, M., Pennimpede, T., Schmidt, J.M., Ebert, A.-K., Rösch, W., Stein, R., Utsch, B., Hirsch, K., Boemers, T.M., Mangold, E., Heilmann, S.,

Ludwig, K.U., Jenetzky, E., Zwink, N., Moebus, S., Herrmann, B.G., Mattheisen, M., Nöthen, M.M., Ludwig, M., and Reutter, H., 2015. Genome-wide association study and meta-analysis identify ISL1 as genome-wide significant susceptibility gene for bladder exstrophy. *PLoS genetics*, 11 (3), e1005024.

Dudley, A.T., Lyons, K.M., and Robertson, E.J., 1995. A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. *Genes & development*, 9 (22), 2795–2807.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., and Domingue, B., 2019. Analysis of polygenic risk score usage and performance in diverse human populations. *Nature communications*, 10 (1), 3328.

Dunn, N.R. and Tolwinski, N.S., 2016. Ptk7 and Mcc, Unfancied Components in Non-Canonical Wnt Signaling and Cancer. *Cancers*, 8 (7).

Eckel-Passow, J.E., Atkinson, E.J., Maharjan, S., Kardia, S.L.R., and de Andrade, M., 2011. Software comparison for evaluating genomic copy number variation for Affymetrix 6.0 SNP array platform. *BMC bioinformatics*, 12 (1), 220.

Edghill, E.L., Bingham, C., Ellard, S., and Hattersley, A.T., 2006. Mutations in hepatocyte nuclear factor-1beta and their related phenotypes. *Journal of medical genetics*, 43 (1), 84–90.

van Eerde, A.M., Koeleman, B.P.C., van de Kamp, J.M., de Jong, T.P.V.M., Wijmenga, C., and Giltay, J.C., 2007. Linkage study of 14 candidate genes and loci in four large Dutch families with vesico-ureteral reflux. *Pediatric nephrology (Berlin, Germany)*, 22 (8), 1129–1133.

ENCODE Project Consortium, Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E.L., Freese, P., Gorkin, D.U., Shen, Y., He, Y., Mackiewicz, M., Pauli-Behn, F., Williams, B.A., Mortazavi, A., Keller, C.A., Zhang, X.-O., Elhajjajy, S.I., Huey, J., Dickel, D.E., Snetkova, V., Wei, X., Wang, X., Rivera-Mulia, J.C., Rozowsky, J., Zhang, J., Chhetri, S.B., Zhang, J., Victorsen, A., White, K.P., Visel, A., Yeo, G.W., Burge, C.B., Lécuyer, E., Gilbert, D.M., Dekker, J., Rinn, J., Mendenhall, E.M., Ecker, J.R., Kellis, M., Klein, R.J., Noble, W.S., Kundaje, A., Guigó, R., Farnham, P.J., Cherry, J.M., Myers, R.M., Ren, B., Graveley, B.R., Gerstein, M.B., Pennacchio, L.A., Snyder, M.P., Bernstein, B.E., Wold, B., Hardison, R.C., Gingeras, T.R., Stamatoyannopoulos, J.A., and Weng, Z., 2020. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, 583 (7818), 699–710.

Epstein, M.P., Duncan, R., Broadaway, K.A., He, M., Allen, A.S., and Satten, G.A., 2012. Stratification-score matching improves correction for confounding by population stratification in case-control association studies. *Genetic epidemiology*, 36 (3), 195–205.

Ernst, J. and Kellis, M., 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nature methods*, 9 (3), 215–216.

ESCAPE Trial Group, Wühl, E., Trivelli, A., Picca, S., Litwin, M., Peco-Antic, A., Zurowska, A., Testa, S., Jankauskiene, A., Emre, S., Caldas-Afonso, A., Anarat, A., Niaudet, P., Mir, S., Bakkaloglu, A., Enke, B., Montini, G., Wingen, A.-M., Sallay, P., Jeck, N., Berg, U., Caliskan, S., Wygoda, S., Hohbach-Hohenfellner, K., Dusek, J., Urasinski, T., Arbeiter, K., Neuhaus, T., Gellermann, J., Drozdz, D., Fischbach, M., Möller, K., Wigger, M., Peruzzi, L., Mehls, O., and Schaefer, F., 2009. Strict blood-pressure control and progression of renal failure in children. *The New England journal of medicine*, 361 (17), 1639–1650.

Evangelou, E., Warren, H.R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C.P., Karaman, I., Ng, F.L., Evangelou, M., Witkowska, K., Tzanis, E., Hellwege, J.N., Giri, A., Velez Edwards, D.R., Sun, Y.V., Cho, K., Gaziano, J.M., Wilson, P.W.F., Tsao, P.S., Kovesdy, C.P., Esko, T., Mägi, R., Milani, L., Almgren, P., Boutin, T., Debette, S., Ding, J., Giulianini, F., Holliday, E.G., Jackson, A.U., Li-Gao, R., Lin, W.-Y., Luan, J., Mangino, M., Oldmeadow, C., Prins, B.P., Qian, Y., Sargurupremraj, M., Shah, N., Surendran, P., Thériault, S., Verweij, N., Willems, S.M., Zhao, J.-H., Amouyel, P., Connell, J., de Mutsert, R., Doney, A.S.F., Farrall, M., Menni, C., Morris, A.D., Noordam, R., Paré, G., Poulter, N.R., Shields, D.C., Stanton, A., Thom, S., Abecasis, G., Amin, N., Arking, D.E., Ayers, K.L., Barbieri, C.M., Batini, C., Bis, J.C., Blake, T., Bochud, M., Boehnke, M., Boerwinkle, E., Boomsma, D.I., Bottinger, E.P., Braund, P.S., Brumat, M., Campbell, A., Campbell, H., Chakravarti, A., Chambers, J.C., Chauhan, G., Ciullo, M., Cocca, M., Collins, F., Cordell, H.J., Davies, G., de Borst, M.H., de Geus, E.J., Deary, I.J., Deelen, J., Del Greco M, F., Demirkale, C.Y., Dörr, M., Ehret, G.B., Elosua, R., Enroth, S., Erzurumluoglu, A.M., Ferreira, T., Frånberg, M., Franco, O.H., Gandin, I., Gasparini, P., Giedraitis, V., Gieger, C., Girotto, G., Goel, A., Gow, A.J., Gudnason, V., Guo, X., Gyllensten, U., Hamsten, A., Harris, T.B., Harris, S.E., Hartman, C.A., Havulinna, A.S., Hicks, A.A., Hofer, E., Hofman, A., Hottenga, J.-J., Huffman, J.E., Hwang, S.-J., Ingelsson, E., James, A., Jansen, R., Jarvelin, M.-R., Joehanes, R., Johansson, Å., Johnson, A.D., Joshi, P.K., Jousilahti, P., Jukema, J.W., Jula, A., Kähönen, M., Kathiresan, S., Keavney, B.D., Khaw, K.-T., Knekt, P., Knight, J., Kolcic, I., Kooner, J.S., Koskinen, S., Kristiansson, K., Kutalik, Z., Laan, M., Larson, M., Launer, L.J., Lehne, B., Lehtimäki, T., Liewald, D.C.M., Lin, L., Lind, L., Lindgren, C.M., Liu, Y., Loos, R.J.F., Lopez, L.M., Lu, Y., Lyytikäinen, L.-P., Mahajan, A., Mamasoula, C., Marrugat, J., Marten, J., Milaneschi, Y., Morgan, A., Morris, A.P., Morrison, A.C., Munson, P.J., Nalls, M.A., Nandakumar, P., Nelson, C.P., Niiranen, T., Nolte, I.M., Nutile, T., Oldehinkel, A.J., Oostra, B.A., O'Reilly, P.F., Org, E., Padmanabhan, S., Palmas, W., Palotie, A., Pattie, A., Penninx, B.W.J.H., Perola, M., Peters, A., Polasek, O., Pramstaller, P.P., Nguyen, Q.T., Raitakari, O.T., Ren, M., Rettig, R., Rice, K., Ridker, P.M., Ried, J.S., Riese, H., Ripatti, S., Robino, A., Rose, L.M., Rotter, J.I., Rudan, I., Ruggiero, D., Saba, Y., Sala, C.F., Salomaa, V., Samani, N.J., Sarin, A.-

P., Schmidt, R., Schmidt, H., Shrine, N., Siscovick, D., Smith, A.V., Snieder, H., Sõber, S., Sorice, R., Starr, J.M., Stott, D.J., Strachan, D.P., Strawbridge, R.J., Sundström, J., Swertz, M.A., Taylor, K.D., Teumer, A., Tobin, M.D., Tomaszewski, M., Toniolo, D., Traglia, M., Trompet, S., Tuomilehto, J., Tzourio, C., Uitterlinden, A.G., Vaez, A., van der Most, P.J., van Duijn, C.M., Vergnaud, A.-C., Verwoert, G.C., Vitart, V., Völker, U., Vollenweider, P., Vuckovic, D., Watkins, H., Wild, S.H., Willemsen, G., Wilson, J.F., Wright, A.F., Yao, J., Zemunik, T., Zhang, W., Attia, J.R., Butterworth, A.S., Chasman, D.I., Conen, D., Cucca, F., Danesh, J., Hayward, C., Howson, J.M.M., Laakso, M., Lakatta, E.G., Langenberg, C., Melander, O., Mook-Kanamori, D.O., Palmer, C.N.A., Risch, L., Scott, R.A., Scott, R.J., Sever, P., Spector, T.D., van der Harst, P., Wareham, N.J., Zeggini, E., Levy, D., Munroe, P.B., Newton-Cheh, C., Brown, M.J., Metspalu, A., Hung, A.M., O'Donnell, C.J., Edwards, T.L., Psaty, B.M., Tzoulaki, I., Barnes, M.R., Wain, L.V., Elliott, P., Caulfield, M.J., and Million Veteran Program, 2018. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nature genetics*, 50 (10), 1412–1425.

Fadista, J., Manning, A.K., Florez, J.C., and Groop, L., 2016. The (in)famous GWAS P-value threshold revisited and updated for low-frequency variants. *European journal of human genetics: EJHG*, 24 (8), 1202–1205.

Fagny, M. and Austerlitz, F., 2021. Polygenic adaptation: Integrating population genetics and gene regulatory networks. *Trends in genetics: TIG*, 37 (7), 631–638.

Fahed, A.C., Wang, M., Homburger, J.R., Patel, A.P., Bick, A.G., Neben, C.L., Lai, C., Brockman, D., Philippakis, A., Ellinor, P.T., Cassa, C.A., Lebo, M., Ng, K., Lander, E.S., Zhou, A.Y., Kathiresan, S., and Khera, A.V., 2020. Polygenic background modifies penetrance of monogenic variants for tier 1 genomic conditions. *Nature communications*, 11 (1), 3635.

Farrugia, M.-K., 2016. Fetal bladder outlet obstruction: Embryopathology, in utero intervention and outcome. *Journal of pediatric urology*, 12 (5), 296–303.

Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A.R., and Kuchenbaecker, K., 2022. A roadmap to increase diversity in genomic studies. *Nature medicine*.

Faure, A., Bouty, A., Caruana, G., Williams, L., Burgess, T., Wong, M.N., James, P.A., O'Brien, M., Walker, A., Bertram, J.F., and Heloury, Y., 2016. DNA copy number variants: A potentially useful predictor of early onset renal failure in boys with posterior urethral valves. *Journal of pediatric urology*, 12 (4), 227.e1–7.

Feather, S.A., Malcolm, S., Woolf, A.S., Wright, V., Blaydon, D., Reid, C.J., Flinter, F.A., Proesmans, W., Devriendt, K., Carter, J., Warwicker, P., Goodship, T.H., and Goodship, J.A., 2000. Primary, nonsyndromic vesicoureteric reflux and its nephropathy is genetically heterogeneous, with a locus on chromosome 1. *The American Journal of Human Genetics*, 66 (4), 1420–1425.

Filion, R., Grignon, A., and Boisvert, J., 1985. Antenatal diagnosis of ipsilateral multicystic kidney in identical twins. *Journal of ultrasound in medicine: official journal of the American Institute of Ultrasound in Medicine*, 4 (4), 211–212.

Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P., 2009. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *American journal of human genetics*, 84 (4), 524–533.

Flöttmann, R., Kragesteen, B.K., Geuer, S., Socha, M., Allou, L., Sowińska-Seidler, A., Bosquillon de Jarcy, L., Wagner, J., Jamsheer, A., Oehl-Jaschkowitz, B., Wittler, L., de Silva, D., Kurth, I., Maya, I., Santos-Simarro, F., Hülsemann, W., Klopocki, E., Mountford, R., Fryer, A., Borck, G., Horn, D., Lapunzina, P., Wilson, M., Mascrez, B., Duboule, D., Mundlos, S., and Spielmann, M., 2018. Noncoding copy-number variations are associated with congenital limb malformation. *Genetics in medicine: official journal of the American College of Medical Genetics*, 20 (6), 599–607.

Forbes, T.A., Howden, S.E., Lawlor, K., Phipson, B., Maksimovic, J., Hale, L., Wilson, S., Quinlan, C., Ho, G., Holman, K., Bennetts, B., Crawford, J., Trnka, P., Oshlack, A., Patel, C., Mallett, A., Simons, C., and Little, M.H., 2018. Patient-iPSC-derived kidney organoids show functional validation of a ciliopathic renal phenotype and reveal underlying pathogenetic mechanisms. *The American Journal of Human Genetics*, 102 (5), 816–831.

Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., Santana-Garcia, W., Tan, G., Chèneby, J., Ballester, B., Parcy, F., Sandelin, A., Lenhard, B., Wasserman, W.W., and Mathelier, A., 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic acids research*, 48 (D1), D87–D92.

Franke, M., De la Calle-Mustienes, E., Neto, A., Almuedo-Castillo, M., Irastorza-Azcarate, I., Acemel, R.D., Tena, J.J., Santos-Pereira, J.M., and Gómez-Skarmeta, J.L., 2021. CTCF knockout in zebrafish induces alterations in regulatory landscapes and developmental gene expression. *Nature communications*, 12 (1), 5415.

Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.-L., Spielmann, M., Timmermann, B., Wittler, L., Kurth, I., Cambiaso, P., Zuffardi, O., Houge, G., Lambie, L., Brancati, F., Pombo, A., Vingron, M., Spitz, F., and Mundlos, S., 2016. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, 538 (7624), 265–269.

Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I.T., García Girón, C., Gonzalez, J.M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O.G., Lagarde, J., Martin, F.J., Martínez, L., Mohanan, S., Muir, P., Navarro, F.C.P., Parker, A., Pei, B., Pozo,

F., Ruffier, M., Schmitt, B.M., Stapleton, E., Suner, M.-M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J.S., Gerstein, M., Guigó, R., Hubbard, T.J.P., Kellis, M., Paten, B., Reymond, A., Tress, M.L., and Flicek, P., 2019. GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research*, 47 (D1), D766–D773.

Frese, S., Weigert, A., Hoppe, B., Feldkötter, M., Ludwig, M., Weber, S., Kiliś-Pstrusińska, K., Zaniew, M., Reutter, H., and Hilger, A.C., 2019. A classic twin study of lower urinary tract obstruction: Report of 3 cases and literature review. *Lower urinary tract symptoms*, 11 (2), O85–O88.

Freund, M.K., Burch, K., Shi, H., Mancuso, N., Kichaev, G., Garske, K.M., Pan, D.Z., Pajukanta, P., Pasaniuc, G., and Arboleda, V.A., 2018. Phenotype-specific enrichment of Mendelian disorder genes near GWAS regions across 62 complex traits. *bioRxiv*.

Fuchsberger, C., Flannick, J., Teslovich, T.M., Mahajan, A., Agarwala, V., Gaulton, K.J., Ma, C., Fontanillas, P., Moutsianas, L., McCarthy, D.J., Rivas, M.A., Perry, J.R.B., Sim, X., Blackwell, T.W., Robertson, N.R., Rayner, N.W., Cingolani, P., Locke, A.E., Tajes, J.F., Highland, H.M., Dupuis, J., Chines, P.S., Lindgren, C.M., Hartl, C., Jackson, A.U., Chen, H., Huyghe, J.R., van de Bunt, M., Pearson, R.D., Kumar, A., Müller-Nurasyid, M., Grarup, N., Stringham, H.M., Gamazon, E.R., Lee, J., Chen, Y., Scott, R.A., Below, J.E., Chen, P., Huang, J., Go, M.J., Stitzel, M.L., Pasko, D., Parker, S.C.J., Varga, T.V., Green, T., Beer, N.L., Day-Williams, A.G., Ferreira, T., Fingerlin, T., Horikoshi, M., Hu, C., Huh, I., Ikram, M.K., Kim, B.-J., Kim, Y., Kim, Y.J., Kwon, M.-S., Lee, J., Lee, S., Lin, K.-H., Maxwell, T.J., Nagai, Y., Wang, X., Welch, R.P., Yoon, J., Zhang, W., Barzilai, N., Voight, B.F., Han, B.-G., Jenkinson, C.P., Kuulasmaa, T., Kuusisto, J., Manning, A., Ng, M.C.Y., Palmer, N.D., Balkau, B., Stančáková, A., Abboud, H.E., Boeing, H., Giedraitis, V., Prabhakaran, D., Gottesman, O., Scott, J., Carey, J., Kwan, P., Grant, G., Smith, J.D., Neale, B.M., Purcell, S., Butterworth, A.S., Howson, J.M.M., Lee, H.M., Lu, Y., Kwak, S.-H., Zhao, W., Danesh, J., Lam, V.K.L., Park, K.S., Saleheen, D., So, W.Y., Tam, C.H.T., Afzal, U., Aguilar, D., Arya, R., Aung, T., Chan, E., Navarro, C., Cheng, C.-Y., Palli, D., Correa, A., Curran, J.E., Rybin, D., Farook, V.S., Fowler, S.P., Freedman, B.I., Griswold, M., Hale, D.E., Hicks, P.J., Khor, C.-C., Kumar, S., Lehne, B., Thuillier, D., Lim, W.Y., Liu, J., van der Schouw, Y.T., Loh, M., Musani, S.K., Puppala, S., Scott, W.R., Yengo, L., Tan, S.-T., Taylor, H.A., Thameem, F., Wilson, G., Wong, T.Y., Njølstad, P.R., Levy, J.C., Mangino, M., Bonnycastle, L.L., Schwarzmayr, T., Fadista, J., Surdulescu, G.L., Herder, C., Groves, C.J., Wieland, T., Bork-Jensen, J., Brandslund, I., Christensen, C., Koistinen, H.A., Doney, A.S.F., Kinnunen, L., Esko, T., Farmer, A.J., Hakaste, L., Hodgkiss, D., Kravic, J., Lyssenko, V., Hollensted, M., Jørgensen, M.E., Jørgensen, T., Ladenvall, C., Justesen, J.M., Käräjämäki, A., Kriebel, J., Rathmann, W., Lannfelt, L., Lauritzen, T., Narisu, N., Linneberg, A., Melander, O., Milani, L.,

Neville, M., Orho-Melander, M., Qi, L., Qi, Q., Roden, M., Rolandsson, O., Swift, A., Rosengren, A.H., Stirrups, K., Wood, A.R., Mihailov, E., Blancher, C., Carneiro, M.O., Maguire, J., Poplin, R., Shakir, K., Fennell, T., DePristo, M., Hrabé de Angelis, M., Deloukas, P., Gjesing, A.P., Jun, G., Nilsson, P., Murphy, J., Onofrio, R., Thorand, B., Hansen, T., Meisinger, C., Hu, F.B., Isomaa, B., Karpe, F., Liang, L., Peters, A., Huth, C., O'Rahilly, S.P., Palmer, C.N.A., Pedersen, O., Rauramaa, R., Tuomilehto, J., Salomaa, V., Watanabe, R.M., Syvänen, A.-C., Bergman, R.N., Bharadwaj, D., Bottinger, E.P., Cho, Y.S., Chandak, G.R., Chan, J.C.N., Chia, K.S., Daly, M.J., Ebrahim, S.B., Langenberg, C., Elliott, P., Jablonski, K.A., Lehman, D.M., Jia, W., Ma, R.C.W., Pollin, T.I., Sandhu, M., Tandon, N., Froguel, P., Barroso, I., Teo, Y.Y., Zeggini, E., Loos, R.J.F., Small, K.S., Ried, J.S., DeFronzo, R.A., Grallert, H., Glaser, B., Metspalu, A., Wareham, N.J., Walker, M., Banks, E., Gieger, C., Ingelsson, E., Im, H.K., Illig, T., Franks, P.W., Buck, G., Trakalo, J., Buck, D., Prokopenko, I., Mägi, R., Lind, L., Farjoun, Y., Owen, K.R., Gloyn, A.L., Strauch, K., Tuomi, T., Kooner, J.S., Lee, J.-Y., Park, T., Donnelly, P., Morris, A.D., Hattersley, A.T., Bowden, D.W., Collins, F.S., Atzmon, G., Chambers, J.C., Spector, T.D., Laakso, M., Strom, T.M., Bell, G.I., Blangero, J., Duggirala, R., Tai, E.S., McVean, G., Hanis, C.L., Wilson, J.G., Seielstad, M., Frayling, T.M., Meigs, J.B., Cox, N.J., Sladek, R., Lander, E.S., Gabriel, S., Burtt, N.P., Mohlke, K.L., Meitinger, T., Groop, L., Abecasis, G., Florez, J.C., Scott, L.J., Morris, A.P., Kang, H.M., Boehnke, M., Altshuler, D., and McCarthy, M.I., 2016. The genetic architecture of type 2 diabetes. *Nature*, 536 (7614), 41–47.

Fudenberg, G. and Pollard, K.S., 2019. Chromatin features constrain structural variation across evolutionary timescales. *Proceedings of the National Academy of Sciences of the United States of America*, 116 (6), 2175–2180.

Gale, D.P., de Jorge, E.G., Cook, H.T., Martinez-Barricarte, R., Hadjisavvas, A., McLean, A.G., Pusey, C.D., Pierides, A., Kyriacou, K., Athanasiou, Y., Voskarides, K., Deltas, C., Palmer, A., Frémeaux-Bacchi, V., de Cordoba, S.R., Maxwell, P.H., and Pickering, M.C., 2010. Identification of a mutation in complement factor H-related protein 5 in patients of Cypriot origin with glomerulonephritis. *The Lancet*, 376 (9743), 794–801.

Gao, Z., Lee, P., Stafford, J.M., von Schimmelmann, M., Schaefer, A., and Reinberg, D., 2014. An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature*, 516 (7531), 349–354.

Garrelfs, S.F., Frishberg, Y., Hulton, S.A., Koren, M.J., O'Riordan, W.D., Cochat, P., Deschênes, G., Shasha-Lavsky, H., Saland, J.M., Van't Hoff, W.G., Fuster, D.G., Magen, D., Moochhala, S.H., Schalk, G., Simkova, E., Groothoff, J.W., Sas, D.J., Meliambro, K.A., Lu, J., Sweetser, M.T., Garg, P.P., Vaishnaw, A.K., Gansner, J.M., McGregor, T.L., Lieske, J.C., and ILLUMINATE-A Collaborators, 2021. Lumasiran, an RNAi therapeutic for primary hyperoxaluria type 1. *The New England journal of medicine*, 384 (13), 1216–1226.

Gbadegesin, R.A., Brophy, P.D., Adeyemo, A., Hall, G., Gupta, I.R., Hains, D., Bartkowiak, B., Rabinovich, C.E., Chandrasekharappa, S., Homstad, A., Westreich, K., Wu, G., Liu, Y., Holanda, D., Clarke, J., Lavin, P., Selim, A., Miller, S., Wiener, J.S., Ross, S.S., Foreman, J., Rotimi, C., and Winn, M.P., 2013. TNXB mutations can cause vesicoureteral reflux. *Journal of the American Society of Nephrology: JASN*, 24 (8), 1313–1322.

Gen, Y., Yasui, K., Zen, K., Nakajima, T., Tsuji, K., Endo, M., Mitsuyoshi, H., Minami, M., Itoh, Y., Tanaka, S., Taniwaki, M., Arii, S., Okanoue, T., and Yoshikawa, T., 2009. A novel amplification target, ARHGAP5, promotes cell spreading and migration by negatively regulating RhoA in Huh-7 hepatocellular carcinoma cells. *Cancer letters*, 275 (1), 27–34.

GenomeAsia100K Consortium, 2019. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*, 576 (7785), 106–111.

Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., Uzureau, P., Freedman, B.I., Bowden, D.W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., Bernhardy, A.J., Hicks, P.J., Nelson, G.W., Vanhollebeke, B., Winkler, C.A., Kopp, J.B., Pays, E., and Pollak, M.R., 2010. Association of trypanolytic ApoL1 variants with kidney disease in African Americans. *Science (New York, N.Y.)*, 329 (5993), 841–845.

Gessler, M., Poustka, A., Cavenee, W., Neve, R.L., Orkin, S.H., and Bruns, G.A., 1990. Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. *Nature*, 343 (6260), 774–778.

Ghoussaini, M., Mountjoy, E., Carmona, M., Peat, G., Schmidt, E.M., Hercules, A., Fumis, L., Miranda, A., Carvalho-Silva, D., Buniello, A., Burdett, T., Hayhurst, J., Baker, J., Ferrer, J., Gonzalez-Uriarte, A., Jupp, S., Karim, M.A., Koscielny, G., Machlitt-Northen, S., Malangone, C., Pendlington, Z.M., Roncaglia, P., Suveges, D., Wright, D., Vrousgou, O., Papa, E., Parkinson, H., MacArthur, J.A.L., Todd, J.A., Barrett, J.C., Schwartzentruber, J., Hulcoop, D.G., Ochoa, D., McDonagh, E.M., and Dunham, I., 2021. Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics. *Nucleic acids research*, 49 (D1), D1311–D1320.

Gilissen, C., Hehir-Kwa, J.Y., Thung, D.T., van de Vorst, M., van Bon, B.W.M., Willemsen, M.H., Kwint, M., Janssen, I.M., Hoischen, A., Schenck, A., Leach, R., Klein, R., Tearle, R., Bo, T., Pfundt, R., Yntema, H.G., de Vries, B.B.A., Kleefstra, T., Brunner, H.G., Vissers, L.E.L.M., and Veltman, J.A., 2014. Genome sequencing identifies major causes of severe intellectual disability. *Nature*, 511 (7509), 344–347.

Gimelli, S., Caridi, G., Beri, S., McCracken, K., Bocciardi, R., Zordan, P., Dagnino, M., Fiorio, P., Murer, L., Benetti, E., Zuffardi, O., Giorda, R., Wells, J.M., Gimelli, G., and Ghiggeri, G.M., 2010. Mutations in SOX17 are associated with congenital anomalies of the kidney and the urinary tract. *Human mutation*, 31 (12), 1352–1359.

Giudicessi, J.R. and Ackerman, M.J., 2013. Determinants of incomplete penetrance and variable expressivity in heritable cardiac arrhythmia syndromes. *Translational research: the journal of laboratory and clinical medicine*, 161 (1), 1–14.

Gogarten, S.M., Bhangale, T., Conomos, M.P., Laurie, C.A., McHugh, C.P., Painter, I., Zheng, X., Crosslin, D.R., Levine, D., Lumley, T., Nelson, S.C., Rice, K., Shen, J., Swarnkar, R., Weir, B.S., and Laurie, C.C., 2012. GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics* , 28 (24), 3329–3331.

Gonzaga-Jauregui, C., Harel, T., Gambin, T., Kousi, M., Griffin, L.B., Francescatto, L., Ozes, B., Karaca, E., Jhangiani, S.N., Bainbridge, M.N., Lawson, K.S., Pehlivan, D., Okamoto, Y., Withers, M., Mancias, P., Slavotinek, A., Reitnauer, P.J., Goksungur, M.T., Shy, M., Crawford, T.O., Koenig, M., Willer, J., Flores, B.N., Pediaditrakis, I., Us, O., Wiszniewski, W., Parman, Y., Antonellis, A., Muzny, D.M., Baylor-Hopkins Center for Mendelian Genomics, Katsanis, N., Battaloglu, E., Boerwinkle, E., Gibbs, R.A., and Lupski, J.R., 2015. Exome sequence analysis suggests that genetic burden contributes to phenotypic variability and complex neuropathy. *Cell reports*, 12 (7), 1169–1183.

Goodwin, S., McPherson, J.D., and McCombie, W.R., 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature reviews. Genetics*, 17 (6), 333–351.

Graff, M., Justice, A.E., Young, K.L., Marouli, E., Zhang, X., Fine, R.S., Lim, E., Buchanan, V., Rand, K., Feitosa, M.F., Wojczynski, M.K., Yanek, L.R., Shao, Y., Rohde, R., Adeyemo, A.A., Aldrich, M.C., Allison, M.A., Ambrosone, C.B., Ambs, S., Amos, C., Arnett, D.K., Atwood, L., Bandera, E.V., Bartz, T., Becker, D.M., Berndt, S.I., Bernstein, L., Bielak, L.F., Blot, W.J., Bottinger, E.P., Bowden, D.W., Bradfield, J.P., Brody, J.A., Broeckel, U., Burke, G., Cade, B.E., Cai, Q., Caporaso, N., Carlson, C., Carpten, J., Casey, G., Chanock, S.J., Chen, G., Chen, M., Chen, Y.-D.I., Chen, W.-M., Chesi, A., Chiang, C.W.K., Chu, L., Coetzee, G.A., Conti, D.V., Cooper, R.S., Cushman, M., Demerath, E., Deming, S.L., Dimitrov, L., Ding, J., Diver, W.R., Duan, Q., Evans, M.K., Falusi, A.G., Faul, J.D., Fornage, M., Fox, C., Freedman, B.I., Garcia, M., Gillanders, E.M., Goodman, P., Gottesman, O., Grant, S.F.A., Guo, X., Hakonarson, H., Haritunians, T., Harris, T.B., Harris, C.C., Henderson, B.E., Hennis, A., Hernandez, D.G., Hirschhorn, J.N., McNeill, L.H., Howard, T.D., Howard, B., Hsing, A.W., Hsu, Y.-H.H., Hu, J.J., Huff, C.D., Huo, D., Ingles, S.A., Irvin, M.R., John, E.M., Johnson, K.C., Jordan, J.M., Kabagambe, E.K., Kang, S.J., Kardia, S.L., Keating, B.J., Kittles, R.A., Klein, E.A., Kolb, S., Kolonel, L.N., Kooperberg, C., Kuller, L., Kutlar, A., Lange, L., Langefeld, C.D., Le Marchand, L., Leonard, H., Lettre, G., Levin, A.M., Li, Y., Li, J., Liu, Y., Liu, Y., Liu, S., Lohman, K., Lotay, V., Lu, Y., Maixner, W., Manson, J.E., McKnight, B., Meng, Y., Monda, K.L., Monroe, K., Moore, J.H., Mosley, T.H., Mudgal, P., Murphy, A.B., Nadukuru, R., Nalls,

M.A., Nathanson, K.L., Nayak, U., N'Diaye, A., Nemesure, B., Neslund-Dudas, C., Neuhouser, M.L., Nyante, S., Ochs-Balcom, H., Ogundiran, T.O., Ogunniyi, A., Ojengbede, O., Okut, H., Olopade, O.I., Olshan, A., Padhukasahasram, B., Palmer, J., Palmer, C.D., Palmer, N.D., Papanicolaou, G., Patel, S.R., Pettaway, C.A., Peyser, P.A., Press, M.F., Rao, D.C., Rasmussen-Torvik, L.J., Redline, S., Reiner, A.P., Rhie, S.K., Rodriguez-Gil, J.L., Rotimi, C.N., Rotter, J.I., Ruiz-Narvaez, E.A., Rybicki, B.A., Salako, B., Sale, M.M., Sanderson, M., Schadt, E., Schreiner, P.J., Schurmann, C., Schwartz, A.G., Shriner, D.A., Signorello, L.B., Singleton, A.B., Siscovick, D.S., Smith, J.A., Smith, S., Speliotes, E., Spitz, M., Stanford, J.L., Stevens, V.L., Stram, A., Strom, S.S., Sucheston, L., Sun, Y.V., Tajuddin, S.M., Taylor, H., Taylor, K., Tayo, B.O., Thun, M.J., Tucker, M.A., Vaidya, D., Van Den Berg, D.J., Vedantam, S., Vitolins, M., Wang, Z., Ware, E.B., Wassertheil-Smoller, S., Weir, D.R., Wiencke, J.K., Williams, S.M., Williams, L.K., Wilson, J.G., Witte, J.S., Wrensch, M., Wu, X., Yao, J., Zakai, N., Zanetti, K., Zemel, B.S., Zhao, W., Zhao, J.H., Zheng, W., Zhi, D., Zhou, J., Zhu, X., Ziegler, R.G., Zmuda, J., Zonderman, A.B., Psaty, B.M., Borecki, I.B., Cupples, L.A., Liu, C.-T., Haiman, C.A., Loos, R., Ng, M.C.Y., and North, K.E., 2021. Discovery and fine-mapping of height loci via high-density imputation of GWASs in individuals of African ancestry. *American journal of human genetics*.

Graham, S.E., Clarke, S.L., Wu, K.-H.H., Kanoni, S., Zajac, G.J.M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., Locke, A.E., Marouli, E., Hwang, M.Y., Han, S., Narita, A., Choudhury, A., Bentley, A.R., Ekoru, K., Verma, A., Trivedi, B., Martin, H.C., Hunt, K.A., Hui, Q., Klarin, D., Zhu, X., Thorleifsson, G., Helgadottir, A., Gudbjartsson, D.F., Holm, H., Olafsson, I., Akiyama, M., Sakaue, S., Terao, C., Kanai, M., Zhou, W., Brumpton, B.M., Rasheed, H., Ruotsalainen, S.E., Havulinna, A.S., Veturi, Y., Feng, Q., Rosenthal, E.A., Lingren, T., Pacheco, J.A., Pendergrass, S.A., Haessler, J., Giulianini, F., Bradford, Y., Miller, J.E., Campbell, A., Lin, K., Millwood, I.Y., Hindy, G., Rasheed, A., Faul, J.D., Zhao, W., Weir, D.R., Turman, C., Huang, H., Graff, M., Mahajan, A., Brown, M.R., Zhang, W., Yu, K., Schmidt, E.M., Pandit, A., Gustafsson, S., Yin, X., Luan, J., Zhao, J.-H., Matsuda, F., Jang, H.-M., Yoon, K., Medina-Gomez, C., Pitsillides, A., Hottenga, J.J., Willemsen, G., Wood, A.R., Ji, Y., Gao, Z., Haworth, S., Mitchell, R.E., Chai, J.F., Aadahl, M., Yao, J., Manichaikul, A., Warren, H.R., Ramirez, J., Bork-Jensen, J., Kårhus, L.L., Goel, A., Sabater-Lleal, M., Noordam, R., Sidore, C., Fiorillo, E., McDaid, A.F., Marques-Vidal, P., Wielscher, M., Trompet, S., Sattar, N., Møllehave, L.T., Thuesen, B.H., Munz, M., Zeng, L., Huang, J., Yang, B., Poveda, A., Kurbasic, A., Lamina, C., Forer, L., Scholz, M., Galesloot, T.E., Bradfield, J.P., Daw, E.W., Zmuda, J.M., Mitchell, J.S., Fuchsberger, C., Christensen, H., Brody, J.A., Feitosa, M.F., Wojczynski, M.K., Preuss, M., Mangino, M., Christofidou, P., Verweij, N., Benjamins, J.W., Engmann, J., Kember, R.L., Slieker, R.C., Lo, K.S., Zilhao, N.R., Le, P., Kleber, M.E.,

Delgado, G.E., Huo, S., Ikeda, D.D., Iha, H., Yang, J., Liu, J., Leonard, H.L., Marten, J., Schmidt, B., Arendt, M., Smyth, L.J., Cañadas-Garre, M., Wang, C., Nakatochi, M., Wong, A., Hutri-Kähönen, N., Sim, X., Xia, R., Huerta-Chagoya, A., Fernandez-Lopez, J.C., Lyssenko, V., Ahmed, M., Jackson, A.U., Irvin, M.R., Oldmeadow, C., Kim, H.-N., Ryu, S., Timmers, P.R.H.J., Arbeeva, L., Dorajoo, R., Lange, L.A., Chai, X., Prasad, G., Lorés-Motta, L., Pauper, M., Long, J., Li, X., Theusch, E., Takeuchi, F., Spracklen, C.N., Loukola, A., Bollepalli, S., Warner, S.C., Wang, Y.X., Wei, W.B., Nutile, T., Ruggiero, D., Sung, Y.J., Hung, Y.-J., Chen, S., Liu, F., Yang, J., Kentistou, K.A., Gorski, M., Brumat, M., Meidtner, K., Bielak, L.F., Smith, J.A., Hebbar, P., Farmaki, A.-E., Hofer, E., Lin, M., Xue, C., Zhang, J., Concas, M.P., Vaccargiu, S., van der Most, P.J., Pitkänen, N., Cade, B.E., Lee, J., van der Laan, S.W., Chitrala, K.N., Weiss, S., Zimmermann, M.E., Lee, J.Y., Choi, H.S., Nethander, M., Freitag-Wolf, S., Southam, L., Rayner, N.W., Wang, C.A., Lin, S.-Y., Wang, J.-S., Couture, C., Lyytikäinen, L.-P., Nikus, K., Cuellar-Partida, G., Vestergaard, H., Hildalgo, B., Giannakopoulou, O., Cai, Q., Obura, M.O., van Setten, J., Li, X., Schwander, K., Terzikhan, N., Shin, J.H., Jackson, R.D., Reiner, A.P., Martin, L.W., Chen, Z., Li, L., Highland, H.M., Young, K.L., Kawaguchi, T., Thiery, J., Bis, J.C., Nadkarni, G.N., Launer, L.J., Li, H., Nalls, M.A., Raitakari, O.T., Ichihara, S., Wild, S.H., Nelson, C.P., Campbell, H., Jäger, S., Nabika, T., Al-Mulla, F., Niinikoski, H., Braund, P.S., Kolcic, I., Kovacs, P., Giardoglou, T., Katsuya, T., Bhatti, K.F., de Kleijn, D., de Borst, G.J., Kim, E.K., Adams, H.H.H., Ikram, M.A., Zhu, X., Asselbergs, F.W., Kraaijeveld, A.O., Beulens, J.W.J., Shu, X.-O., Rallidis, L.S., Pedersen, O., Hansen, T., Mitchell, P., Hewitt, A.W., Kähönen, M., Pérusse, L., Bouchard, C., Tönjes, A., Chen, Y.-D.I., Pennell, C.E., Mori, T.A., Lieb, W., Franke, A., Ohlsson, C., Mellström, D., Cho, Y.S., Lee, H., Yuan, J.-M., Koh, W.-P., Rhee, S.Y., Woo, J.-T., Heid, I.M., Stark, K.J., Völzke, H., Homuth, G., Evans, M.K., Zonderman, A.B., Polasek, O., Pasterkamp, G., Hoefer, I.E., Redline, S., Pahkala, K., Oldehinkel, A.J., Snieder, H., Biino, G., Schmidt, R., Schmidt, H., Chen, Y.E., Bandinelli, S., Dedoussis, G., Thanaraj, T.A., Kardia, S.L.R., Kato, N., Schulze, M.B., Girotto, G., Jung, B., Böger, C.A., Joshi, P.K., Bennett, D.A., De Jager, P.L., Lu, X., Mamakou, V., Brown, M., Caulfield, M.J., Munroe, P.B., Guo, X., Ciullo, M., Jonas, J.B., Samani, N.J., Kaprio, J., Pajukanta, P., Adair, L.S., Bechayda, S.A., de Silva, H.J., Wickremasinghe, A.R., Krauss, R.M., Wu, J.-Y., Zheng, W., den Hollander, A.I., Bharadwaj, D., Correa, A., Wilson, J.G., Lind, L., Heng, C.-K., Nelson, A.E., Golightly, Y.M., Wilson, J.F., Penninx, B., Kim, H.-L., Attia, J., Scott, R.J., Rao, D.C., Arnett, D.K., Walker, M., Koistinen, H.A., Chandak, G.R., Yajnik, C.S., Mercader, J.M., Tusié-Luna, T., Aguilar-Salinas, C.A., Villalpando, C.G., Orozco, L., Fornage, M., Tai, E.S., van Dam, R.M., Lehtimäki, T., Chaturvedi, N., Yokota, M., Liu, J., Reilly, D.F., McKnight, A.J., Kee, F., Jöckel, K.-H., McCarthy, M.I., Palmer, C.N.A., Vitart, V., Hayward, C., Simonsick, E., van Duijn, C.M., Lu, F., Qu, J., Hishigaki, H., Lin, X., März,

W., Parra, E.J., Cruz, M., Gudnason, V., Tardif, J.-C., Lettre, G., 't Hart, L.M., Elders, P.J.M., Damrauer, S.M., Kumari, M., Kivimaki, M., van der Harst, P., Spector, T.D., Loos, R.J.F., Province, M.A., Psaty, B.M., Brandslund, I., Pramstaller, P.P., Christensen, K., Ripatti, S., Widén, E., Hakonarson, H., Grant, S.F.A., Kiemeney, L.A.L.M., de Graaf, J., Loeffler, M., Kronenberg, F., Gu, D., Erdmann, J., Schunkert, H., Franks, P.W., Linneberg, A., Jukema, J.W., Khera, A.V., Männikkö, M., Jarvelin, M.-R., Kutalik, Z., Cucca, F., Mook-Kanamori, D.O., van Dijk, K.W., Watkins, H., Strachan, D.P., Grarup, N., Sever, P., Poulter, N., Rotter, J.I., Dantoft, T.M., Karpe, F., Neville, M.J., Timpson, N.J., Cheng, C.-Y., Wong, T.-Y., Khor, C.C., Sabanayagam, C., Peters, A., Gieger, C., Hattersley, A.T., Pedersen, N.L., Magnusson, P.K.E., Boomsma, D.I., de Geus, E.J.C., Cupples, L.A., van Meurs, J.B.J., Ghanbari, M., Gordon-Larsen, P., Huang, W., Kim, Y.J., Tabara, Y., Wareham, N.J., Langenberg, C., Zeggini, E., Kuusisto, J., Laakso, M., Ingelsson, E., Abecasis, G., Chambers, J.C., Kooner, J.S., de Vries, P.S., Morrison, A.C., North, K.E., Daviglus, M., Kraft, P., Martin, N.G., Whitfield, J.B., Abbas, S., Saleheen, D., Walters, R.G., Holmes, M.V., Black, C., Smith, B.H., Justice, A.E., Baras, A., Buring, J.E., Ridker, P.M., Chasman, D.I., Kooperberg, C., Wei, W.-Q., Jarvik, G.P., Namjou, B., Hayes, M.G., Ritchie, M.D., Jousilahti, P., Salomaa, V., Hveem, K., Åsvold, B.O., Kubo, M., Kamatani, Y., Okada, Y., Murakami, Y., Thorsteinsdottir, U., Stefansson, K., Ho, Y.-L., Lynch, J.A., Rader, D.J., Tsao, P.S., Chang, K.-M., Cho, K., O'Donnell, C.J., Gaziano, J.M., Wilson, P., Rotimi, C.N., Hazelhurst, S., Ramsay, M., Trembath, R.C., van Heel, D.A., Tamiya, G., Yamamoto, M., Kim, B.-J., Mohlke, K.L., Frayling, T.M., Hirschhorn, J.N., Kathiresan, S., VA Million Veteran Program, Global Lipids Genetics Consortium*, Boehnke, M., Natarajan, P., Peloso, G.M., Brown, C.D., Morris, A.P., Assimes, T.L., Deloukas, P., Sun, Y.V., and Willer, C.J., 2021. The power of genetic diversity in genome-wide association studies of lipids. *Nature*.

Graham, S.E., Nielsen, J.B., Zawistowski, M., Zhou, W., Fritsche, L.G., Gabrielsen, M.E., Skogholt, A.H., Surakka, I., Hornsby, W.E., Fermin, D., Larach, D.B., Kheterpal, S., Brummett, C.M., Lee, S., Kang, H.M., Abecasis, G.R., Romundstad, S., Hallan, S., Sampson, M.G., Hveem, K., and Willer, C.J., 2019. Sex-specific and pleiotropic effects underlying kidney function identified from GWAS meta-analysis. *Nature communications*, 10 (1), 1847.

Grajewski, R.S. and Glassberg, K.I., 1983. The variable effect of posterior urethral valves as illustrated in identical twins. *The journal of urology*, 130 (6), 1188–1190.

Granberg, C.F., Harrison, S.M., Dajusta, D., Zhang, S., Hajarnis, S., Igarashi, P., and Baker, L.A., 2012. Genetic basis of prune belly syndrome: screening for HNF1β gene. *The Journal of urology*, 187 (1), 272–278.

Greenway, S.C., Pereira, A.C., Lin, J.C., DePalma, S.R., Israel, S.J., Mesquita, S.M., Ergul, E., Conta, J.H., Korn, J.M., McCarroll, S.A., Gorham, J.M., Gabriel, S., Altshuler, D.M., Quintanilla-Dieck, M. de L., Artunduaga, M.A., Eavey, R.D.,

Plenge, R.M., Shadick, N.A., Weinblatt, M.E., De Jager, P.L., Hafler, D.A., Breitbart, R.E., Seidman, J.G., and Seidman, C.E., 2009. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nature genetics*, 41 (8), 931–935.

Gribouval, O., Gonzales, M., Neuhaus, T., Aziza, J., Bieth, E., Laurent, N., Bouton, J.M., Feuillet, F., Makni, S., Ben Amar, H., Laube, G., Delezoide, A.-L., Bouvier, R., Dijoud, F., Ollagnon-Roman, E., Roume, J., Joubert, M., Antignac, C., and Gubler, M.C., 2005. Mutations in genes in the renin-angiotensin system are associated with autosomal recessive renal tubular dysgenesis. *Nature genetics*, 37 (9), 964–968.

Groen In 't Woud, S., Renkema, K.Y., Schreuder, M.F., Wijers, C.H.W., van der Zanden, L.F.M., Knoers, N.V.A.M., Feitz, W.F.J., Bongers, E.M.H.F., Roeleveld, N., and van Rooij, I.A.L.M., 2016. Maternal risk factors involved in specific congenital anomalies of the kidney and urinary tract: A case-control study. *Birth defects research. Part A, Clinical and molecular teratology*, 106 (7), 596–603.

Groopman, E.E., Marasa, M., Cameron-Christie, S., Petrovski, S., Aggarwal, V.S., Milo-Rasouly, H., Li, Y., Zhang, J., Nestor, J., Krithivasan, P., Lam, W.Y., Mitrotti, A., Piva, S., Kil, B.H., Chatterjee, D., Reingold, R., Bradbury, D., DiVecchia, M., Snyder, H., Mu, X., Mehl, K., Balderes, O., Fasel, D.A., Weng, C., Radhakrishnan, J., Canetta, P., Appel, G.B., Bomback, A.S., Ahn, W., Uy, N.S., Alam, S., Cohen, D.J., Crew, R.J., Dube, G.K., Rao, M.K., Kamalakaran, S., Copeland, B., Ren, Z., Bridgers, J., Malone, C.D., Mebane, C.M., Dagaonkar, N., Fellström, B.C., Haefliger, C., Mohan, S., Sanna-Cherchi, S., Kiryluk, K., Fleckner, J., March, R., Platt, A., Goldstein, D.B., and Gharavi, A.G., 2019. Diagnostic Utility of Exome Sequencing for Kidney Disease. *The New England journal of medicine*, 380 (2), 142–151.

GTEx Consortium, 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, 369 (6509), 1318–1330.

Guan, W., Liang, L., Boehnke, M., and Abecasis, G.R., 2009. Genotype-based matching to correct for population stratification in large-scale case-control genetic association studies. *Genetic epidemiology*, 33 (6), 508–517.

Guan, Y., Liang, X., Ma, Z., Hu, H., Liu, H., Miao, Z., Linkermann, A., Hellwege, J.N., Voight, B.F., and Susztak, K., 2021. A single genetic locus controls both expression of DPEP1/CHMP1A and kidney disease development via ferroptosis. *Nature communications*, 12 (1), 5078.

Gudbjartsson, D.F., Helgason, H., Gudjonsson, S.A., Zink, F., Oddson, A., Gylfason, A., Besenbacher, S., Magnusson, G., Halldorsson, B.V., Hjartarson, E., Sigurdsson, G.T., Stacey, S.N., Frigge, M.L., Holm, H., Saemundsdottir, J., Helgadottir, H.T., Johannsdottir, H., Sigfusson, G., Thorgeirsson, G., Sverrisson, J.T., Gretarsdottir, S., Walters, G.B., Rafnar, T., Thjodleifsson, B., Bjornsson, E.S., Olafsson, S., Thorarinsdottir, H., Steingrimsdottir, T., Gudmundsdottir, T.S., Theodors, A., Jonasson, J.G., Sigurdsson, A., Bjornsdottir, G., Jonsson, J.J., Thorarensen, O., Ludvigsson, P.,

Gudbjartsson, H., Eyjolfsson, G.I., Sigurdardottir, O., Olafsson, I., Arnar, D.O., Magnusson, O.T., Kong, A., Masson, G., Thorsteinsdottir, U., Helgason, A., Sulem, P., and Stefansson, K., 2015. Large-scale whole-genome sequencing of the Icelandic population. *Nature genetics*, 47 (5), 435–444.

Guo, J., Bakshi, A., Wang, Y., Jiang, L., Yengo, L., Goddard, M.E., Visscher, P.M., and Yang, J., 2021. Quantifying genetic heterogeneity between continental populations for human height and body mass index. *Scientific reports*, 11 (1), 5240.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., Lu, Y., Wu, Y., Jia, Z., Li, W., Zhang, M.Q., Ren, B., Krainer, A.R., Maniatis, T., and Wu, Q., 2015. CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell*, 162 (4), 900–910.

Guzmán, Y.F., Ramsey, K., Stolz, J.R., Craig, D.W., Huentelman, M.J., Narayanan, V., and Swanson, G.T., 2017. A gain-of-function mutation in the GRIK2 gene causes neurodevelopmental deficits. *Neurology. Genetics*, 3 (1), e129.

H3Africa Consortium, Rotimi, C., Abayomi, A., Abimiku, A., Adabayeri, V.M., Adebamowo, C., Adebiyi, E., Ademola, A.D., Adeyemo, A., Adu, D., Affolabi, D., Agongo, G., Ajayi, S., Akarolo-Anthony, S., Akinyemi, R., Akpalu, A., Alberts, M., Alonso Betancourt, O., Alzohairy, A.M., Ameni, G., Amodu, O., Anabwani, G., Andersen, K., Arogundade, F., Arulogun, O., Asogun, D., Bakare, R., Balde, N., Baniecki, M.L., Beiswanger, C., Benkahla, A., Bethke, L., Boehnke, M., Boima, V., Brandful, J., Brooks, A.I., Brosius, F.C., Brown, C., Bucheton, B., Burke, D.T., Burnett, B.G., Carrington-Lawrence, S., Carstens, N., Chisi, J., Christoffels, A., Cooper, R., Cordell, H., Crowther, N., Croxton, T., de Vries, J., Derr, L., Donkor, P., Doumbia, S., Duncanson, A., Ekem, I., El Sayed, A., Engel, M.E., Enyaru, J.C.K., Everett, D., Fadlelmola, F.M., Fakunle, E., Fischbeck, K.H., Fischer, A., Folarin, O., Gamieldien, J., Garry, R.F., Gaseitsiwe, S., Gbadegesin, R., Ghansah, A., Giovanni, M., Goesbeck, P., Gomez-Olive, F.X., Grant, D.S., Grewal, R., Guyer, M., Hanchard, N.A., Happi, C.T., Hazelhurst, S., Hennig, B.J., Hertz-, C., Fowler, Hide, W., Hilderbrandt, F., Hugo-Hamman, C., Ibrahim, M.E., James, R., Jaufeerally-Fakim, Y., Jenkins, C., Jentsch, U., Jiang, P.-P., Joloba, M., Jongeneel, V., Joubert, F., Kader, M., Kahn, K., Kaleebu, P., Kapiga, S.H., Kassim, S.K., Kasvosve, I., Kayondo, J., Keavney, B., Kekitiinwa, A., Khan, S.H., Kimmel, P., King, M.-C., Kleta, R., Koffi, M., Kopp, J., Kretzler, M., Kumuthini, J., Kyobe, S., Kyobutungi, C., Lackland, D.T., Lacourciere, K.A., Landouré, G., Lawlor, R., Lehner, T., Lesosky, M., Levitt, N., Littler, K., Lombard, Z., Loring, J.F., Lyantagaye, S., Macleod, A., Madden, E.B., Mahomva, C.R., Makani, J., Mamven, M., Marape, M., Mardon, G., Marshall, P., Martin, D.P., Masiga, D., Mason, R., Mate-Kole, M., Matovu, E., Mayige, M., Mayosi, B.M., Mbanya, J.C., McCurdy, S.A., McCarthy, M.I., McIlleron, H., Mc'Ligeyo, S.O., Merle, C., Mocumbi, A.O., Mondo, C., Moran, J.V., Motala, A., Moxey-Mims, M., Mpoloka, W.S., Msefula, C.L., Mthiyane, T., Mulder, N., Mulugeta, G.H., Mumba, D., Musuku, J., Nagdee, M., Nash, O.,

Ndiaye, D., Nguyen, A.Q., Nicol, M., Nkomazana, O., Norris, S., Nsangi, B., Nyarko, A., Nyirenda, M., Obe, E., Obiakor, R., Oduro, A., Ofori-Acquah, S.F., Ogah, O., Ogendo, S., Ohene-Frempong, K., Ojo, A., Olanrewaju, T., Oli, J., Osafo, C., Ouwe Missi Oukem-Boyer, O., Ovbiagele, B., Owen, A., Owolabi, M.O., Owolabi, L., Owusu-Dabo, E., Pare, G., Parekh, R., Patterton, H.G., Penno, M.B., Peterson, J., Pieper, R., Plange-Rhule, J., Pollak, M., Puzak, J., Ramesar, R.S., Ramsay, M., Rasooly, R., Reddy, S., Sabeti, P.C., Sagoe, K., Salako, T., Samassékou, O., Sandhu, M.S., Sankoh, O., Sarfo, F.S., Sarr, M., Shaboodien, G., Sidibe, I., Simo, G., Simuunza, M., Smeeth, L., Sobngwi, E., Soodyall, H., Sorgho, H., Sow Bah, O., Srinivasan, S., Stein, D.J., Susser, E.S., Swanepoel, C., Tangwa, G., Tareila, A., Tastan Bishop, O., Tayo, B., Tiffin, N., Tinto, H., Tobin, E., Tollman, S.M., Traoré, M., Treadwell, M.J., Troyer, J., Tsimako-Johnstone, M., Tukei, V., Ulasi, I., Ulenga, N., van Rooyen, B., Wachinou, A.P., Waddy, S.P., Wade, A., Wayengera, M., Whitworth, J., Wideroff, L., Winkler, C.A., Winnicki, S., Wonkam, A., Yewondwos, M., Sen, T., Yozwiak, N., and Zar, H., 2014. Research capacity. Enabling the genomic revolution in Africa. *Science (New York, N.Y.)*, 344 (6190), 1346–1348.

Hack, M., Jaffe, J., Blankstein, J., Goodman, R.M., and Brish, M., 1974. Familial aggregation in bilateral renal agenesis. *Clinical genetics*, 5 (2), 173–177.

Hampe, J., Franke, A., Rosenstiel, P., Till, A., Teuber, M., Huse, K., Albrecht, M., Mayr, G., De La Vega, F.M., Briggs, J., Günther, S., Prescott, N.J., Onnie, C.M., Häsler, R., Sipos, B., Fölsch, U.R., Lengauer, T., Platzer, M., Mathew, C.G., Krawczak, M., and Schreiber, S., 2007. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nature genetics*, 39 (2), 207–211.

Han, L., Zhao, X., Benton, M.L., Perumal, T., Collins, R.L., Hoffman, G.E., Johnson, J.S., Sloofman, L., Wang, H.Z., Stone, M.R., CommonMind Consortium, Brennand, K.J., Brand, H., Sieberts, S.K., Marenco, S., Peters, M.A., Lipska, B.K., Roussos, P., Capra, J.A., Talkowski, M., and Ruderfer, D.M., 2020. Functional annotation of rare structural variation in the human brain. *Nature communications*, 11 (1), 2990.

Harambat, J., van Stralen, K.J., Kim, J.J., and Tizard, E.J., 2012. Epidemiology of chronic kidney disease in children. *Pediatric nephrology*, 27 (3), 363–373.

Hardelin, J.P., Levilliers, J., del Castillo, I., Cohen-Salmon, M., Legouis, R., Blanchard, S., Compain, S., Bouloux, P., Kirk, J., and Moraine, C., 1992. X chromosome-linked Kallmann syndrome: stop mutations validate the candidate gene. *Proceedings of the National Academy of Sciences of the United States of America*, 89 (17), 8190–8194.

Harding, S.D., Armit, C., Armstrong, J., Brennan, J., Cheng, Y., Haggarty, B., Houghton, D., Lloyd-MacGilp, S., Pi, X., Roochun, Y., Sharghi, M., Tindal, C., McMahon, A.P., Gottesman, B., Little, M.H., Georgas, K., Aronow, B.J., Potter, S.S., Brunskill, E.W., Southard-Smith, E.M., Mendelsohn, C., Baldock, R.A., Davies, J.A., and Davidson, D., 2011. The GUDMAP database--an

online resource for genitourinary research. *Development* , 138 (13), 2845–2853.

Hayes, M., Gao, X., Yu, L.X., Paria, N., Henkelman, R.M., Wise, C.A., and Ciruna, B., 2014. ptk7 mutant zebrafish models of congenital and idiopathic scoliosis implicate dysregulated Wnt signalling in disease. *Nature communications*, 5, 4777.

Heidet, L., Decramer, S., Pawtowski, A., Morinière, V., Bandin, F., Knebelmann, B., Lebre, A.-S., Faguer, S., Guigonis, V., Antignac, C., and Salomon, R., 2010. Spectrum of HNF1B mutations in a large cohort of patients who harbor renal diseases. *Clinical journal of the American Society of Nephrology: CJASN*, 5 (6), 1079–1090.

Heidet, L., Morinière, V., Henry, C., De Tomasi, L., Reilly, M.L., Humbert, C., Alibeu, O., Fourrage, C., Bole-Feysot, C., Nitschké, P., Tores, F., Bras, M., Jeanpierre, M., Pietrement, C., Gaillard, D., Gonzales, M., Novo, R., Schaefer, E., Roume, J., Martinovic, J., Malan, V., Salomon, R., Saunier, S., Antignac, C., and Jeanpierre, C., 2017. Targeted Exome Sequencing Identifies PBX1 as Involved in Monogenic Congenital Anomalies of the Kidney and Urinary Tract. *Journal of the American Society of Nephrology: JASN*, 28 (10), 2901–2914.

Heikkilä, J., Holmberg, C., Kyllönen, L., Rintala, R., and Taskinen, S., 2011. Long-term risk of end stage renal disease in patients with posterior urethral valves. *The Journal of urology*, 186 (6), 2392–2396.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., Reddy, J., Borges-Rivera, D., Lee, T.I., Jaenisch, R., Porteus, M.H., Dekker, J., and Young, R.A., 2016. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science (New York, N.Y.)*, 351 (6280), 1454–1458.

Ho, S.S., Urban, A.E., and Mills, R.E., 2020. Structural variation in the sequencing era. *Nature reviews. Genetics*, 21 (3), 171–189.

Höglund, J., Rafati, N., Rask-Andersen, M., Enroth, S., Karlsson, T., Ek, W.E., and Johansson, Å., 2019. Improved power and precision with whole genome sequencing data in genome-wide association studies of inflammatory biomarkers. *Scientific reports*, 9 (1), 16844.

Hokke, S.N., Armitage, J.A., Puelles, V.G., Short, K.M., Jones, L., Smyth, I.M., Bertram, J.F., and Cullen-McEwen, L.A., 2013. Altered ureteric branching morphogenesis and nephron endowment in offspring of diabetic and insulin-treated pregnancy. *PloS one*, 8 (3), e58243.

Hoskins, B.E., Cramer, C.H., 2nd, Tasic, V., Kehinde, E.O., Ashraf, S., Bogdanovic, R., Hoefele, J., Pohl, M., and Hildebrandt, F., 2008. Missense mutations in EYA1 and TCF2 are a rare cause of urinary tract malformations. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association*, 23 (2), 777–779.

Hoskins, B.E., Cramer, C.H., Silvius, D., Zou, D., Raymond, R.M., Orten, D.J., Kimberling, W.J., Smith, R.J.H., Weil, D., Petit, C., Otto, E.A., Xu, P.-X., and

Hildebrandt, F., 2007. Transcription factor SIX5 is mutated in patients with branchio-oto-renal syndrome. *The American Journal of Human Genetics*, 80 (4), 800–804.

Houcinat, N., Llanas, B., Moutton, S., Toutain, J., Cailley, D., Arveiler, B., Combe, C., Lacombe, D., and Rooryck, C., 2015. Homozygous 16p13.11 duplication associated with mild intellectual disability and urinary tract malformations in two siblings born from consanguineous parents. *American journal of medical genetics. Part A*, 167A (11), 2714–2719.

Houweling, A.C., Beaman, G.M., Postma, A.V., Gainous, T.B., Lichtenbelt, K.D., Brancati, F., Lopes, F.M., van der Made, I., Polstra, A.M., Robinson, M.L., Wright, K.D., Ellingford, J.M., Jackson, A.R., Overwater, E., Genesio, R., Romano, S., Camerota, L., D'Angelo, E., Meijers-Heijboer, E.J., Christoffels, V.M., McHugh, K.M., Black, B.L., Newman, W.G., Woolf, A.S., and Creemers, E.E., 2019. Loss-of-function variants in myocardin cause congenital megabladder in humans and mice. *The Journal of clinical investigation*, 129 (12), 5374–5380.

Hsu, C.W., Yamamoto, K.T., Henry, R.K., De Roos, A.J., and Flynn, J.T., 2014. Prenatal risk factors for childhood CKD. *Journal of the American Society of Nephrology: JASN*, 25 (9), 2105–2111.

Hu, Y., Stilp, A.M., McHugh, C.P., Rao, S., Jain, D., Zheng, X., Lane, J., Méric de Bellefon, S., Raffield, L.M., Chen, M.-H., Yanek, L.R., Wheeler, M., Yao, Y., Ren, C., Broome, J., Moon, J.-Y., de Vries, P.S., Hobbs, B.D., Sun, Q., Surendran, P., Brody, J.A., Blackwell, T.W., Choquet, H., Ryan, K., Duggirala, R., Heard-Costa, N., Wang, Z., Chami, N., Preuss, M.H., Min, N., Ekunwe, L., Lange, L.A., Cushman, M., Faraday, N., Curran, J.E., Almasy, L., Kundu, K., Smith, A.V., Gabriel, S., Rotter, J.I., Fornage, M., Lloyd-Jones, D.M., Vasan, R.S., Smith, N.L., North, K.E., Boerwinkle, E., Becker, L.C., Lewis, J.P., Abecasis, G.R., Hou, L., O'Connell, J.R., Morrison, A.C., Beaty, T.H., Kaplan, R., Correa, A., Blangero, J., Jorgenson, E., Psaty, B.M., Kooperberg, C., Walton, R.T., Kleinstiver, B.P., Tang, H., Loos, R.J.F., Soranzo, N., Butterworth, A.S., Nickerson, D., Rich, S.S., Mitchell, B.D., Johnson, A.D., Auer, P.L., Li, Y., Mathias, R.A., Lettre, G., Pankratz, N., Laurie, C.C., Laurie, C.A., Bauer, D.E., Conomos, M.P., Reiner, A.P., and NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, 2021. Whole-genome sequencing association analysis of quantitative red blood cell phenotypes: The NHLBI TOPMed program. *The American Journal of Human Genetics*, 108 (6), 1165.

Huang, E.Y., Mascarenhas, L., and Mahour, G.H., 2004. Wilms' tumor and horseshoe kidneys: a case report and review of the literature. *Journal of pediatric surgery*, 39 (2), 207–212.

Humbert, C., Silbermann, F., Morar, B., Parisot, M., Zarhrate, M., Masson, C., Tores, F., Blanchet, P., Perez, M.-J., Petrov, Y., Khau Van Kien, P., Roume, J., Leroy, B., Gribouval, O., Kalaydjieva, L., Heidet, L., Salomon, R., Antignac, C., Benmerah, A., Saunier, S., and Jeanpierre, C., 2014. Integrin alpha 8

recessive mutations are responsible for bilateral renal agenesis in humans. *The American Journal of Human Genetics*, 94 (5), 799.

Hwang, D.-Y., Dworschak, G.C., Kohl, S., Saisawat, P., Vivante, A., Hilger, A.C., Reutter, H.M., Soliman, N.A., Bogdanovic, R., Kehinde, E.O., Tasic, V., and Hildebrandt, F., 2014. Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. *Kidney international*, 85 (6), 1429–1433.

Hwang, D.-Y., Kohl, S., Fan, X., Vivante, A., Chan, S., Dworschak, G.C., Schulz, J., van Eerde, A.M., Hilger, A.C., Gee, H.Y., Pennimpede, T., Herrmann, B.G., van de Hoek, G., Renkema, K.Y., Schell, C., Huber, T.B., Reutter, H.M., Soliman, N.A., Stajic, N., Bogdanovic, R., Kehinde, E.O., Lifton, R.P., Tasic, V., Lu, W., and Hildebrandt, F., 2015. Mutations of the SLIT2-ROBO2 pathway genes SLIT2 and SRGAP1 confer risk for congenital anomalies of the kidney and urinary tract. *Human genetics*, 134 (8), 905–916.

Hyder, Z., Calpena, E., Pei, Y., Tooze, R.S., Brittain, H., Twigg, S.R.F., Cilliers, D., Morton, J.E.V., McCann, E., Weber, A., Wilson, L.C., Douglas, A.G.L., McGowan, R., Need, A., Bond, A., Tavares, A.L.T., Thomas, E.R.A., Genomics England Research Consortium, Hill, S.L., Deans, Z.C., Boardman-Pretty, F., Caulfield, M., Scott, R.H., and Wilkie, A.O.M., 2021. Evaluating the performance of a clinical genome sequencing program for diagnosis of rare genetic disease, seen through the lens of craniosynostosis. *Genetics in medicine: official journal of the American College of Medical Genetics*.

Ichikawa, I., Kuwayama, F., Pope, J.C., 4th, Stephens, F.D., and Miyazaki, Y., 2002. Paradigm shift from classic anatomic theories to contemporary cell biological views of CAKUT. *Kidney international*, 61 (3), 889–898.

Ichimura, T., Watanabe, S., Sakamoto, Y., Aoto, T., Fujita, N., and Nakao, M., 2005. Transcriptional repression and heterochromatin formation by MBD1 and MCAF/AM family proteins. *The journal of biological chemistry*, 280 (14), 13928–13935.

Inoue, R., Hirohashi, Y., Kitamura, H., Nishida, S., Murai, A., Takaya, A., Yamamoto, E., Matsuki, M., Tanaka, T., Kubo, T., Nakatsugawa, M., Kanaseki, T., Tsukahara, T., Sato, N., Masumori, N., and Torigoe, T., 2017. GRIK2 has a role in the maintenance of urothelial carcinoma stem-like cells, and its expression is associated with poorer prognosis. *Oncotarget*, 8 (17), 28826–28839.

International HapMap Consortium, 2005. A haplotype map of the human genome. *Nature*, 437 (7063), 1299–1320.

International Human Genome Sequencing Consortium, 2004. Finishing the euchromatic sequence of the human genome. *Nature*, 431 (7011), 931–945.

Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L.A., Teerlink, C.C., Stanford, J.L., Isaacs, W.B., Xu, J., Cooney, K.A., Lange, E.M., Schleutker, J., Carpten, J.D., Powell, I.J., Cussenot, O., Cancel-Tassin, G., Giles, G.G., MacInnis, R.J., Maier, C., Hsieh, C.-L., Wiklund, F., Catalona,

W.J., Foulkes, W.D., Mandal, D., Eeles, R.A., Kote-Jarai, Z., Bustamante, C.D., Schaid, D.J., Hastie, T., Ostrander, E.A., Bailey-Wilson, J.E., Radivojac, P., Thibodeau, S.N., Whittemore, A.S., and Sieh, W., 2016. REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *The American Journal of Human Genetics*, 99 (4), 877–885.

Jadeja, S., Smyth, I., Pitera, J.E., Taylor, M.S., van Haelst, M., Bentley, E., McGregor, L., Hopkins, J., Chalepakis, G., Philip, N., Perez Aytes, A., Watt, F.M., Darling, S.M., Jackson, I., Woolf, A.S., and Scambler, P.J., 2005. Identification of a new gene mutated in Fraser syndrome and mouse myelencephalic blebs. *Nature genetics*, 37 (5), 520–525.

Jadresić, L., Au, H., Woodhouse, C., and Nitsch, D., 2021. Pre-pregnancy obesity and risk of congenital abnormalities of the kidney and urinary tract (CAKUT)-systematic review, meta-analysis and ecological study. *Pediatric nephrology (Berlin, Germany)*, 36 (1), 119–132.

Jayasinghe, K., Stark, Z., Kerr, P.G., Gaff, C., Martyn, M., Whitlam, J., Creighton, B., Donaldson, E., Hunter, M., Jarmolowicz, A., Johnstone, L., Krzesinski, E., Lunke, S., Lynch, E., Nicholls, K., Patel, C., Prawer, Y., Ryan, J., See, E.J., Talbot, A., Trainer, A., Tytherleigh, R., Valente, G., Wallis, M., Wardrop, L., West, K.H., White, S.M., Wilkins, E., Mallett, A.J., and Quinlan, C., 2021. Clinical impact of genomic testing in patients with suspected monogenic kidney disease. *Genetics in medicine: official journal of the American College of Medical Genetics*, 23 (1), 183–191.

Jeffares, D.C., Jolly, C., Hoti, M., Speed, D., Shaw, L., Rallis, C., Balloux, F., Dessimoz, C., Bähler, J., and Sedlazeck, F.J., 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nature Communications*.

Jin, M., Zhu, S., Hu, P., Liu, D., Li, Q., Li, Z., Zhang, X., Xie, Y., and Chen, X., 2014. Genomic and epigenomic analyses of monozygotic twins discordant for congenital renal agenesis. *American journal of kidney diseases: the official journal of the National Kidney Foundation*, 64 (1), 119–122.

Johansen, K.L., Chertow, G.M., Foley, R.N., Gilbertson, D.T., Herzog, C.A., Ishani, A., Israni, A.K., Ku, E., Kurella Tamura, M., Li, S., Li, S., Liu, J., Obrador, G.T., O'Hare, A.M., Peng, Y., Powe, N.R., Roetker, N.S., St Peter, W.L., Abbott, K.C., Chan, K.E., Schulman, I.H., Snyder, J., Solid, C., Weinhandl, E.D., Winkelmayer, W.C., and Wetmore, J.B., 2021. US Renal Data System 2020 Annual Data Report: Epidemiology of Kidney Disease in the United States. *American journal of kidney diseases: the official journal of the National Kidney Foundation*, 77 (4 Suppl 1), A7–A8.

Johnson, R.J., Feehally, J., Floege, J., and Tonelli, M., 2018. *Comprehensive Clinical Nephrology*. 6th ed. Philadelphia, PA: Elsevier - Health Sciences Division.

Jónsson, H., Sulem, P., Kehr, B., Kristmundsdottir, S., Zink, F., Hjartarson, E., Hardarson, M.T., Hjorleifsson, K.E., Eggertsson, H.P., Gudjonsson, S.A., Ward, L.D., Arnadottir, G.A., Helgason, E.A., Helgason, H., Gylfason, A.,

Jonasdottir, A., Jonasdottir, A., Rafnar, T., Besenbacher, S., Frigge, M.L., Stacey, S.N., Magnusson, O.T., Thorsteinsdottir, U., Masson, G., Kong, A., Halldorsson, B.V., Helgason, A., Gudbjartsson, D.F., and Stefansson, K., 2017. Whole genome characterization of sequence diversity of 15,220 Icelanders. *Scientific data*, 4, 170115.

Kaefer, M., Curran, M., Treves, S.T., Bauer, S., Hendren, W.H., Peters, C.A., Atala, A., Diamond, D., and Retik, A., 2000. Sibling vesicoureteral reflux in multiple gestation births. *Pediatrics*, 105 (4 Pt 1), 800–804.

Kaku, Y., Ohmori, T., Kudo, K., Fujimura, S., Suzuki, K., Evans, S.M., Kawakami, Y., and Nishinakamura, R., 2013. Islet1 deletion causes kidney agenesis and hydroureter resembling CAKUT. *Journal of the American Society of Nephrology: JASN*, 24 (8), 1242–1249.

Källen, K., 1997. Maternal smoking and urinary organ malformations. *International journal of epidemiology*, 26 (3), 571–574.

Kamath, B.M., Podkameni, G., Hutchinson, A.L., Leonard, L.D., Gerfen, J., Krantz, I.D., Piccoli, D.A., Spinner, N.B., Loomes, K.M., and Meyers, K., 2012. Renal anomalies in Alagille syndrome: a disease-defining feature. *American journal of medical genetics. Part A*, 158A (1), 85–89.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.-Y., Freimer, N.B., Sabatti, C., and Eskin, E., 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature genetics*, 42 (4), 348–354.

Kang, S., Graham, J.M., Jr, Olney, A.H., and Biesecker, L.G., 1997. GLI3 frameshift mutations cause autosomal dominant Pallister-Hall syndrome. *Nature genetics*, 15 (3), 266–268.

Kao, W.H.L., Klag, M.J., Meoni, L.A., Reich, D., Berthier-Schaad, Y., Li, M., Coresh, J., Patterson, N., Tandon, A., Powe, N.R., Fink, N.E., Sadler, J.H., Weir, M.R., Abboud, H.E., Adler, S.G., Divers, J., Iyengar, S.K., Freedman, B.I., Kimmel, P.L., Knowler, W.C., Kohn, O.F., Kramp, K., Leehey, D.J., Nicholas, S.B., Pahl, M.V., Schelling, J.R., Sedor, J.R., Thornley-Brown, D., Winkler, C.A., Smith, M.W., Parekh, R.S., and Family Investigation of Nephropathy and Diabetes Research Group, 2008. MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nature genetics*, 40 (10), 1185–1192.

Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., Gauthier, L.D., Brand, H., Solomonson, M., Watts, N.A., Rhodes, D., Singer-Berk, M., England, E.M., Seaby, E.G., Kosmicki, J.A., Walters, R.K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J.X., Samocha, K.E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A.H., Minikel, E.V., Weisburd, B., Lek, M., Ware, J.S., Vittal, C., Armean, I.M., Bergelson, L., Cibulskis, K., Connolly, K.M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio,

V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M.E., Genome Aggregation Database Consortium, Neale, B.M., Daly, M.J., and MacArthur, D.G., 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581 (7809), 434–443.

Katsanis, N., 2016. The continuum of causality in human genetic disorders. *Genome biology*, 17 (1), 233.

Katsanis, N., Ansley, S.J., Badano, J.L., Eichers, E.R., Lewis, R.A., Hoskins, B.E., Scambler, P.J., Davidson, W.S., Beales, P.L., and Lupski, J.R., 2001. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science (New York, N.Y.)*, 293 (5538), 2256–2259.

Kaur, H., Crawford, D.C., Liang, J., Benchek, P., COGENT BP Consortium, Zhu, X., Kallianpur, A.R., and Bush, W.S., 2021. Replication of European hypertension associations in a case-control study of 9,534 African Americans. *PloS one*, 16 (11), e0259962.

Kelly, H., Molony, C.M., Darlow, J.M., Pirker, M.E., Yoneda, A., Green, A.J., Puri, P., and Barton, D.E., 2007. A genome-wide scan for genes involved in primary vesicoureteric reflux. *Journal of medical genetics*, 44 (11), 710–717.

Keramati, A.R., Chen, M.-H., Rodriguez, B.A.T., Yanek, L.R., Gaynor, B.J., Ryan, K., Brody, J.A., Kammers, K., Kanchan, K., Iyer, K., Kowalski, M.H., Pitsillides, A.N., Cupples, L.A., Shuldiner, A.R., O'Connell, J.R., Mitchell, B.D., Faraday, N., Taub, M.A., Becker, L.C., Lewis, J.P., Mathias, R.A., Johnson, A.D., NHLBI Trans-Omics for Precision (TOPMed) Consortium, and NHLBI TOPMed Hematology and Hemostasis Working Group, 2019. Genome sequencing unveils a new regulatory landscape of platelet reactivity. *bioRxiv*.

Khera, A.V., Chaffin, M., Aragam, K.G., Haas, M.E., Roselli, C., Choi, S.H., Natarajan, P., Lander, E.S., Lubitz, S.A., Ellinor, P.T., and Kathiresan, S., 2018. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nature genetics*, 50 (9), 1219–1224.

Kichaev, G., Yang, W.-Y., Lindstrom, S., Hormozdiari, F., Eskin, E., Price, A.L., Kraft, P., and Pasaniuc, B., 2014. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS genetics*, 10 (10), e1004722.

Klei, L., McClain, L.L., Mahjani, B., Panayidou, K., De Rubeis, S., Grahnat, A.-C.S., Karlsson, G., Lu, Y., Melhem, N., Xu, X., Reichenberg, A., Sandin, S., Hultman, C.M., Buxbaum, J.D., Roeder, K., and Devlin, B., 2021. How rare and common risk variation jointly affect liability for autism spectrum disorder. *Molecular autism*, 12 (1), 66.

Klein, R.J., 2005. Complement factor H polymorphism in age-related macular degeneration. *Science (New York, N.Y.)*, 308 (5720), 385–389.

Kobrynski, L.J. and Sullivan, K.E., 2007. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet*, 370 (9596), 1443–1452.

Kohl, S., Hwang, D.-Y., Dworschak, G.C., Hilger, A.C., Saisawat, P., Vivante, A., Stajic, N., Bogdanovic, R., Reutter, H.M., Kehinde, E.O., Tasic, V., and Hildebrandt, F., 2014. Mild recessive mutations in six Fraser syndrome-related genes cause isolated congenital anomalies of the kidney and urinary tract. *Journal of the American Society of Nephrology: JASN*, 25 (9), 1917–1922.

Köhler, S., Gargano, M., Matentzoglu, N., Carmody, L.C., Lewis-Smith, D., Vasilevsky, N.A., Danis, D., Balagura, G., Baynam, G., Brower, A.M., Callahan, T.J., Chute, C.G., Est, J.L., Galer, P.D., Ganesan, S., Griese, M., Haimel, M., Pazmandi, J., Hanauer, M., Harris, N.L., Hartnett, M.J., Hastreiter, M., Hauck, F., He, Y., Jeske, T., Kearney, H., Kindle, G., Klein, C., Knoflach, K., Krause, R., Lagorce, D., McMurry, J.A., Miller, J.A., Munoz-Torres, M.C., Peters, R.L., Rapp, C.K., Rath, A.M., Rind, S.A., Rosenberg, A.Z., Segal, M.M., Seidel, M.G., Smedley, D., Talmy, T., Thomas, Y., Wiafe, S.A., Xian, J., Yüksel, Z., Helbig, I., Mungall, C.J., Haendel, M.A., and Robinson, P.N., 2021. The Human Phenotype Ontology in 2021. *Nucleic acids research*, 49 (D1), D1207–D1217.

Kohlhase, J., Wischermann, A., Reichenbach, H., Froster, U., and Engel, W., 1998. Mutations in the SALL1 putative transcription factor gene cause Townes-Brocks syndrome. *Nature genetics*, 18 (1), 81–83.

Kolatsi-Joannou, M., Bingham, C., Ellard, S., Bulman, M.P., Allen, L.I.S., Hattersley, A.T., and Woolf, A.S., 2001. Hepatocyte nuclear factor-1beta: a new kindred with renal cysts and diabetes and gene expression in normal human development. *Journal of the American Society of Nephrology: JASN*, 12 (10), 2175–2180.

Kolvenbach, C.M., Dworschak, G.C., Frese, S., Japp, A.S., Schuster, P., Wenzlitschke, N., Yilmaz, Ö., Lopes, F.M., Pryalukhin, A., Schierbaum, L., van der Zanden, L.F.M., Kause, F., Schneider, R., Taranta-Janusz, K., Szczepańska, M., Pawlaczyk, K., Newman, W.G., Beaman, G.M., Stuart, H.M., Cervellione, R.M., Feitz, W.F.J., van Rooij, I.A.L.M., Schreuder, M.F., Steffens, M., Weber, S., Merz, W.M., Feldkötter, M., Hoppe, B., Thiele, H., Altmüller, J., Berg, C., Kristiansen, G., Ludwig, M., Reutter, H., Woolf, A.S., Hildebrandt, F., Grote, P., Zaniew, M., Odermatt, B., and Hilger, A.C., 2019. Rare Variants in BNC2 Are Implicated in Autosomal-Dominant Congenital Lower Urinary-Tract Obstruction. *American journal of human genetics*, 104 (5), 994–1006.

Kopp, J.B., Nelson, G.W., Sampath, K., Johnson, R.C., Genovese, G., An, P., Friedman, D., Briggs, W., Dart, R., Korbet, S., Mokrzycki, M.H., Kimmel, P.L., Limou, S., Ahuja, T.S., Berns, J.S., Fryc, J., Simon, E.E., Smith, M.C., Trachtman, H., Michel, D.M., Schelling, J.R., Vlahov, D., Pollak, M., and Winkler, C.A., 2011. APOL1 genetic variants in focal segmental

glomerulosclerosis and HIV-associated nephropathy. *Journal of the American Society of Nephrology: JASN*, 22 (11), 2129–2137.

Kosfeld, A., Brand, F., Weiss, A.-C., Kreuzer, M., Goerk, M., Martens, H., Schubert, S., Schäfer, A.-K., Riehmer, V., Hennies, I., Bräsen, J.H., Pape, L., Amann, K., Krogvold, L., Bjerre, A., Daniel, C., Kispert, A., Haffner, D., and Weber, R.G., 2017. Mutations in the leukemia inhibitory factor receptor (LIFR) gene and Lifr deficiency cause urinary tract malformations. *Human molecular genetics*, 26 (9), 1716–1731.

Kosfeld, A., Kreuzer, M., Daniel, C., Brand, F., Schäfer, A.-K., Chadt, A., Weiss, A.-C., Riehmer, V., Jeanpierre, C., Klintschar, M., Bräsen, J.H., Amann, K., Pape, L., Kispert, A., Al-Hasani, H., Haffner, D., and Weber, R.G., 2016. Whole-exome sequencing identifies mutations of TBC1D1 encoding a Rab-GTPase-activating protein in patients with congenital anomalies of the kidneys and urinary tract (CAKUT). *Human genetics*, 135 (1), 69–87.

Kosugi, S., Momozawa, Y., Liu, X., Terao, C., Kubo, M., and Kamatani, Y., 2019. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. *Genome biology*, 20 (1), 117.

Kramer, A., Boenink, R., Stel, V.S., Santiuste de Pablos, C., Tomović, F., Golan, E., Kerschbaum, J., Seyahi, N., Ioanou, K., Beltrán, P., Zurriaga, O., Magaz, Á., Slon Roblero, M.F., Gjorgjievski, N., Garneata, L., Arribas, F., Galvão, A.A., Bell, S., Ots-Rosenberg, M., Muñoz-Terol, J.M., Winzeler, R., Hommel, K., Åsberg, A., Spustova, V., Palencia García, M.Á., Vazelov, E., Finne, P., Ten Dam, M.A.G.J., Lopot, F., Trujillo-Alemán, S., Lassalle, M., Kolesnyk, M.O., Santhakumaran, S., Idrizi, A., Andrusev, A., Comas Farnés, J., Komissarov, K., Resić, H., Palsson, R., Kuzema, V., Garcia Bazaga, M.A., Ziginskiene, E., Stendahl, M., Bonthuis, M., Massy, Z.A., and Jager, K.J., 2021. The ERA-EDTA Registry Annual Report 2018: a summary. *Clinical kidney journal*, 14 (1), 107–123.

Krietenstein, N., Abraham, S., Venev, S.V., Abdennur, N., Gibcus, J., Hsieh, T.-H.S., Parsi, K.M., Yang, L., Maehr, R., Mirny, L.A., Dekker, J., and Rando, O.J., 2020. Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular cell*, 78 (3), 554-565.e7.

Krishnan, A., de Souza, A., Konijeti, R., and Baskin, L.S., 2006. The anatomy and embryology of posterior urethral valves. *The Journal of urology*, 175 (4), 1214–1220.

Krishnan, B., Truong, L.D., Saleh, G., Sirbasku, D.M., and Slawin, K.M., 1997. Horseshoe kidney is associated with an increased relative risk of primary renal carcinoid tumor. *The journal of urology*, 157 (6), 2059–2066.

Kroovand, R.L., Weinberg, N., and Emami, A., 1977. Posterior urethral valves in identical twins. *Pediatrics*, 60 (5), 748.

Krug, P., Morinière, V., Marlin, S., Koubi, V., Gabriel, H.D., Colin, E., Bonneau, D., Salomon, R., Antignac, C., and Heidet, L., 2011. Mutation screening of the EYA1, SIX1, and SIX5 genes in a large cohort of patients harboring branchio-

oto-renal syndrome calls into question the pathogenic role of SIX5 mutations. *Human mutation*, 32 (2), 183–190.

Kuchenbaecker, K., Understanding Society Scientific Group, Telkar, N., Reiker, T., Walters, R.G., Lin, K., Eriksson, A., Gurdasani, D., Gilly, A., Southam, L., Tsafantakis, E., Karaleftheri, M., Seeley, J., Kamali, A., Asiki, G., Millwood, I.Y., Holmes, M., Du, H., Guo, Y., Kumari, M., Dedoussis, G., Li, L., Chen, Z., Sandhu, M.S., and Zeggini, E., 2019. The transferability of lipid loci across African, Asian and European cohorts. *Nature Communications*.

Kuleshov, V., Xie, D., Chen, R., Pushkarev, D., Ma, Z., Blauwkamp, T., Kertesz, M., and Snyder, M., 2014. Whole-genome haplotyping using long reads and statistical methods. *Nature biotechnology*, 32 (3), 261–266.

Lakich, D., Kazazian, H.H., Jr, Antonarakis, S.E., and Gitschier, J., 1993. Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. *Nature genetics*, 5 (3), 236–241.

Lam, M., Chen, C.-Y., Li, Z., Martin, A.R., Bryois, J., Ma, X., Gaspar, H., Ikeda, M., Benyamin, B., Brown, B.C., Liu, R., Zhou, W., Guan, L., Kamatani, Y., Kim, S.-W., Kubo, M., Kusumawardhani, A.A.A.A., Liu, C.-M., Ma, H., Periyasamy, S., Takahashi, A., Xu, Z., Yu, H., Zhu, F., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Indonesia Schizophrenia Consortium, Genetic REsearch on schizophreniA neTwork-China and the Netherlands (GREAT-CN), Chen, W.J., Faraone, S., Glatt, S.J., He, L., Hyman, S.E., Hwu, H.-G., McCarroll, S.A., Neale, B.M., Sklar, P., Wildenauer, D.B., Yu, X., Zhang, D., Mowry, B.J., Lee, J., Holmans, P., Xu, S., Sullivan, P.F., Ripke, S., O'Donovan, M.C., Daly, M.J., Qin, S., Sham, P., Iwata, N., Hong, K.S., Schwab, S.G., Yue, W., Tsuang, M., Liu, J., Ma, X., Kahn, R.S., Shi, Y., and Huang, H., 2019. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature genetics*, 51 (12), 1670–1678.

Lappalainen, I., Lopez, J., Skipper, L., Hefferon, T., Spalding, J.D., Garner, J., Chen, C., Maguire, M., Corbett, M., Zhou, G., Paschall, J., Ananiev, V., Flicek, P., and Church, D.M., 2013. DbVar and DGVa: public archives for genomic structural variation. *Nucleic acids research*, 41 (Database issue), D936-41.

Lappalainen, T., Scott, A.J., Brandt, M., and Hall, I.M., 2019. Genomic analysis in the age of human genome sequencing. *Cell*, 177 (1), 70–84.

Laskar, R.S., Muller, D.C., Li, P., Machiela, M.J., Ye, Y., Gaborieau, V., Foll, M., Hofmann, J.N., Colli, L., Sampson, J.N., Wang, Z., Bacq-Daian, D., Boland, A., Abedi-Ardekani, B., Durand, G., Le Calvez-Kelm, F., Robinot, N., Blanche, H., Prokhortchouk, E., Skryabin, K.G., Burdett, L., Yeager, M., Radojevic-Skodric, S., Savic, S., Foretova, L., Holcatova, I., Janout, V., Mates, D., Rascu, S., Mukeria, A., Zaridze, D., Bencko, V., Cybulski, C., Fabianova, E., Jinga, V., Lissowska, J., Lubinski, J., Navratilova, M., Rudnai, P., Świątkowska, B., Benhamou, S., Cancel-Tassin, G., Cussenot, O., Trichopoulou, A., Riboli, E., Overvad, K., Panico, S., Ljungberg, B., Sitaram, R.T., Giles, G.G., Milne, R.L., Severi, G., Bruinsma, F., Fletcher, T., Koppova, K., Larsson, S.C., Wolk, A., Banks, R.E., Selby, P.J., Easton, D.F., Pharoah,

P., Andreotti, G., Beane Freeman, L.E., Koutros, S., Albanes, D., Männistö, S., Weinstein, S., Clark, P.E., Edwards, T.L., Lipworth, L., Carol, H., Freedman, M.L., Pomerantz, M.M., Cho, E., Kraft, P., Preston, M.A., Wilson, K.M., Michael Gaziano, J., Sesso, H.D., Black, A., Freedman, N.D., Huang, W.-Y., Anema, J.G., Kahnoski, R.J., Lane, B.R., Noyes, S.L., Petillo, D., Teh, B.T., Peters, U., White, E., Anderson, G.L., Johnson, L., Luo, J., Chow, W.-H., Moore, L.E., Choueiri, T.K., Wood, C., Johansson, M., McKay, J.D., Brown, K.M., Rothman, N., Lathrop, M.G., Deleuze, J.-F., Wu, X., Brennan, P., Chanock, S.J., Purdue, M.P., and Scelo, G., 2019. Sex specific associations in genome wide association analysis of renal cell carcinoma. *European journal of human genetics: EJHG*, 27 (10), 1589–1598.

Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M., 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome biology*, 15 (6), R84.

Le Tanno, P., Breton, J., Bidart, M., Satre, V., Harbuz, R., Ray, P.F., Bosson, C., Dieterich, K., Jaillard, S., Odent, S., Poke, G., Beddow, R., Digilio, M.C., Novelli, A., Bernardini, L., Pisanti, M.A., Mackenroth, L., Hackmann, K., Vogel, I., Christensen, R., Fokstuen, S., Béna, F., Amblard, F., Devillard, F., Vieville, G., Apostolou, A., Jouk, P.-S., Guebre-Egziabher, F., Sartelet, H., and Coutton, C., 2017. PBX1haploinsufficiency leads to syndromic congenital anomalies of the kidney and urinary tract (CAKUT) in humans. *Journal of medical genetics*, 54 (7), 502–510.

Lee, H., Deignan, J.L., Dorrani, N., Strom, S.P., Kantarci, S., Quintero-Rivera, F., Das, K., Toy, T., Harry, B., Yourshaw, M., Fox, M., Fogel, B.L., Martinez-Agosto, J.A., Wong, D.A., Chang, V.Y., Shieh, P.B., Palmer, C.G.S., Dipple, K.M., Grody, W.W., Vilain, E., and Nelson, S.F., 2014. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA: the journal of the American Medical Association*, 312 (18), 1880.

Lee, J., Andreeva, A., Sipe, C.W., Liu, L., Cheng, A., and Lu, X., 2012. PTK7 regulates myosin II activity to orient planar polarity in the mammalian auditory epithelium. *Current biology: CB*, 22 (11), 956–966.

Lee, S., Abecasis, G.R., Boehnke, M., and Lin, X., 2014. Rare-variant association analysis: study designs and statistical tests. *The American Journal of Human Genetics*, 95 (1), 5–23.

Lee, S., Emond, M.J., Bamshad, M.J., Barnes, K.C., Rieder, M.J., Nickerson, D.A., NHLBI GO Exome Sequencing Project—ESP Lung Project Team, Christiani, D.C., Wurfel, M.M., and Lin, X., 2012. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *American journal of human genetics*, 91 (2), 224–237.

Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M., 2011. Estimating missing heritability for disease from genome-wide association studies. *The American Journal of Human Genetics*, 88 (3), 294–305.

de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D., 2015. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology*, 11 (4), e1004219.

Lei, Y., Kim, S.-E., Chen, Z., Cao, X., Zhu, H., Yang, W., Shaw, G.M., Zheng, Y., Zhang, T., Wang, H.-Y., and Finnell, R.H., 2019. Variants identified in PTK7 associated with neural tube defects. *Molecular genetics & genomic medicine*, 7 (4), e00584.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., Tukiainen, T., Birnbaum, D.P., Kosmicki, J.A., Duncan, L.E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D.N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M.I., Moonshine, A.L., Natarajan, P., Orozco, L., Peloso, G.M., Poplin, R., Rivas, M.A., Ruano-Rubio, V., Rose, S.A., Ruderfer, D.M., Shakir, K., Stenson, P.D., Stevens, C., Thomas, B.P., Tiao, G., Tusie-Luna, M.T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D.M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J.C., Gabriel, S.B., Getz, G., Glatt, S.J., Hultman, C.M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M.I., McGovern, D., McPherson, R., Neale, B.M., Palotie, A., Purcell, S.M., Saleheen, D., Scharf, J.M., Sklar, P., Sullivan, P.F., Tuomilehto, J., Tsuang, M.T., Watkins, H.C., Wilson, J.G., Daly, M.J., MacArthur, D.G., and Exome Aggregation Consortium, 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536 (7616), 285–291.

Lelieveld, S.H., Spielmann, M., Mundlos, S., Veltman, J.A., and Gilissen, C., 2015. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Human mutation*, 36 (8), 815–822.

Lettice, L.A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P.S., Gautier, P., Morrison, H., Fantes, J., Hill, R.E., and FitzPatrick, D.R., 2011. Enhancer-adoption as a mechanism of human developmental disease. *Human mutation*, 32 (12), 1492–1499.

Lewis, A.C.F., Molina, S.J., Appelbaum, P.S., Dauda, B., Di Rienzo, A., Fuentes, A., Fullerton, S.M., Garrison, N.A., Ghosh, N., Hammonds, E.M., Jones, D.S., Kenny, E.E., Kraft, P., Lee, S.S.-J., Mauro, M., Novembre, J., Panofsky, A., Sohail, M., Neale, B.M., and Allen, D.S., 2022. Getting genetic ancestry right for science and society. *Science (New York, N.Y.)*, 376 (6590), 250–252.

Li, B. and Leal, S.M., 2008. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *The American Journal of Human Genetics*, 83 (3), 311–321.

Li, Q.Y., Newbury-Ecob, R.A., Terrett, J.A., Wilson, D.I., Curtis, A.R., Yi, C.H., Gebuhr, T., Bullen, P.J., Robson, S.C., Strachan, T., Bonnet, D., Lyonnet, S., Young, I.D., Raeburn, J.A., Buckler, A.J., Law, D.J., and Brook, J.D., 1997. Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene family. *Nature genetics*, 15 (1), 21–29.

Li, Y., Pawlik, B., Elcioglu, N., Aglan, M., Kayserili, H., Yigit, G., Percin, F., Goodman, F., Nürnberg, G., Cenani, A., Urquhart, J., Chung, B.-D., Ismail, S., Amr, K., Aslanger, A.D., Becker, C., Netzer, C., Scambler, P., Eyaid, W., Hamamy, H., Clayton-Smith, J., Hennekam, R., Nürnberg, P., Herz, J., Temtamy, S.A., and Wollnik, B., 2010. LRP4 mutations alter Wnt/beta-catenin signaling and cause limb and kidney malformations in Cenani-Lenz syndrome. *The American Journal of Human Genetics*, 86 (5), 696–706.

Li, Y., Roberts, N.D., Wala, J.A., Shapira, O., Schumacher, S.E., Kumar, K., Khurana, E., Waszak, S., Korbel, J.O., Haber, J.E., Imielinski, M., PCAWG Structural Variation Working Group, Weischenfeldt, J., Beroukhim, R., Campbell, P.J., and PCAWG Consortium, 2020. Patterns of somatic structural variation in human cancer genomes. *Nature*, 578 (7793), 112–121.

Li, Y.R. and Keating, B.J., 2014. Trans-ethnic genome-wide association studies: advantages and challenges of mapping in diverse populations. *Genome medicine*, 6 (10), 91.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P., 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell systems*, 1 (6), 417–425.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., and Dekker, J., 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326 (5950), 289–293.

Lin, D.-Y., 2019. A simple and accurate method to determine genomewide significance for association tests in sequencing studies. *Genetic epidemiology*, 43 (4), 365–372.

Lindner, T.H., Njolstad, P.R., Horikawa, Y., Bostad, L., Bell, G.I., and Sovik, O., 1999. A novel syndrome of diabetes mellitus, renal dysfunction and genital malformation associated with a partial deletion of the pseudo-POU domain of hepatocyte nuclear factor-1beta. *Human molecular genetics*, 8 (11), 2001–2008.

Lionel, A.C., Costain, G., Monfared, N., Walker, S., Reuter, M.S., Hosseini, S.M., Thiruvahindrapuram, B., Merico, D., Jobling, R., Nalpathamkalam, T., Pellecchia, G., Sung, W.W.L., Wang, Z., Bikangaga, P., Boelman, C., Carter, M.T., Cordeiro, D., Cytrynbaum, C., Dell, S.D., Dhir, P., Dowling, J.J., Heon, E., Hewson, S., Hiraki, L., Inbar-Feigenberg, M., Klatt, R., Kronick, J., Laxer, R.M., Licht, C., MacDonald, H., Mercimek-Andrews, S., Mendoza-Londono, R., Piscione, T., Schneider, R., Schulze, A., Silverman, E., Siriwardena, K., Snead, O.C., Sondheimer, N., Sutherland, J., Vincent, A., Wasserman, J.D., Weksberg, R., Shuman, C., Carew, C., Szego, M.J., Hayeems, R.Z., Basran, R., Stavropoulos, D.J., Ray, P.N., Bowdin, S., Meyn, M.S., Cohn, R.D., Scherer, S.W., and Marshall, C.R., 2018. Improved diagnostic yield compared with targeted gene sequencing panels suggests a role for whole-genome

sequencing as a first-tier genetic test. *Genetics in medicine: official journal of the American College of Medical Genetics*, 20 (4), 435–443.

Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., Abedian, S., Cheon, J.H., Cho, J., Dayani, N.E., Franke, L., Fuyuno, Y., Hart, A., Juyal, R.C., Juyal, G., Kim, W.H., Morris, A.P., Poustchi, H., Newman, W.G., Midha, V., Orchard, T.R., Vahedi, H., Sood, A., Sung, J.Y., Malekzadeh, R., Westra, H.-J., Yamazaki, K., Yang, S.-K., International Multiple Sclerosis Genetics Consortium, International IBD Genetics Consortium, Barrett, J.C., Alizadeh, B.Z., Parkes, M., Bk, T., Daly, M.J., Kubo, M., Anderson, C.A., and Weersma, R.K., 2015. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*, 47 (9), 979–986.

Lockhart, J.L., Reeve, H.R., Bredael, J.J., and Krueger, R.P., 1979. Siblings with prune belly syndrome and associated pulmonic stenosis, mental retardation, and deafness. *Urology*, 14 (2), 140–142.

Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., Schizophrenia Working Group of Psychiatric Genomics Consortium, de Candia, T.R., Lee, S.H., Wray, N.R., Kendler, K.S., O'Donovan, M.C., Neale, B.M., Patterson, N., and Price, A.L., 2015. Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nature genetics*, 47 (12), 1385–1392.

Lopez-Rivera, E., Liu, Y.P., Verbitsky, M., Anderson, B.R., Capone, V.P., Otto, E.A., Yan, Z., Mitrotti, A., Martino, J., Steers, N.J., Fasel, D.A., Vukojevic, K., Deng, R., Racedo, S.E., Liu, Q., Werth, M., Westland, R., Vivante, A., Makar, G.S., Bodria, M., Sampson, M.G., Gillies, C.E., Vega-Warner, V., Maiorana, M., Petrey, D.S., Honig, B., Lozanovski, V.J., Salomon, R., Heidet, L., Carpentier, W., Gaillard, D., Carrea, A., Gesualdo, L., Cusi, D., Izzi, C., Scolari, F., van Wijk, J.A.E., Arapovic, A., Saraga-Babic, M., Saraga, M., Kunac, N., Samii, A., McDonald-McGinn, D.M., Crowley, T.B., Zackai, E.H., Drozdz, D., Miklaszewska, M., Tkaczyk, M., Sikora, P., Szczepanska, M., Mizerska-Wasiak, M., Krzemien, G., Szmigielska, A., Zaniew, M., Darlow, J.M., Puri, P., Barton, D., Casolari, E., Furth, S.L., Warady, B.A., Gucev, Z., Hakonarson, H., Flogelova, H., Tasic, V., Latos-Bielenska, A., Materna-Kiryluk, A., Allegri, L., Wong, C.S., Drummond, I.A., D'Agati, V., Imamoto, A., Barasch, J.M., Hildebrandt, F., Kiryluk, K., Lifton, R.P., Morrow, B.E., Jeanpierre, C., Papaioannou, V.E., Ghiggeri, G.M., Gharavi, A.G., Katsanis, N., and Sanna-Cherchi, S., 2017. Genetic Drivers of Kidney Defects in the DiGeorge Syndrome. *The New England journal of medicine*, 376 (8), 742–754.

Lowther, C., Mehrjouy, M.M., Collins, R.L., Bak, M.C., Dudchenko, O., Brand, H., Dong, Z., Rasmussen, M.B., Gu, H., Weisz, D., Nazaryan-Petersen, L., Fjorder, A.S., Mang, Y., Lind-Thomsen, A., Mendez, J.M.M., Calle, X., Chopra, A., Hansen, C., Bugge, M., Broekema, R.V., Varilo, T., Luukkonen, T., Engelen, J., Vianna-Morgante, A.M., Fonseca, A.C.S., Mazzeu, J.F.,

Dornelles-Wawruk, H., Abe, K.T., Vermeesch, J.R., Van Den Bogaert, K., Sismani, C., Aristidou, C., Evangelidou, P., Schinzel, A.A., Sanlaville, D., Schluth-Bolard, C., Kalscheuer, V.M., Wenzel, M., Kim, H.-G., Õunap, K., Roht, L., Midyan, S., Bonaglia, M.C., Lindstrand, A., Eisfeldt, J., Ottosson, J., Nilsson, D., Pettersson, M., Bastos, E.F., Rajcan-Separovic, E., Silan, F., Sheth, F.J., Novelli, A., Frengen, E., Fannemel, M., Strømme, P., Vokač, N.K., Daumer-Haas, C., Moretti-Ferreira, D., de Souza, D.H., Ramos-Arroyo, M.A., Igoa, M.M., Angelova, L., Kroisel, P.M., Rey, G. del, Vieira, T.A.P., Lewis, S., Hao, W., Drabova, J., Havlovicova, M., Hancarova, M., Sedláček, Z., Vogel, I., Hjortshøj, T.D., Møller, R.S., Tümer, Z., Fagerberg, C., Ousager, L.B., Schönewolf-Greulich, B., Lauridsen, M., Piard, J., Pebrel-Richard, C., Jaillard, S., Ehmke, N., Stefanou, E.G., Marta, C., György, K., Dalal, A., Dutta, U.R., Shukla, R., Lonardo, F., Zuffardi, O., Houge, G., Misceo, D., Baig, S.M., Midro, A., Wawrusiewicz-Kurylonek, N., Carreira, I.M., Melo, J.B., Martinez, L.R., Guitart, M., Lovmar, L., Gullander, J., Hansson, K.B.M., de Almeida Esteves, C., Akkari, Y., Batanian, J.R., Li, X., Lespinasse, J., Silahtaroglu, A., Harding, C.H., Krogh, L.N., Taylor, J., Lehnert, K., Hill, R., Snell, R.G., Samson, C.A., Jacobsen, J.C., Levy, B., Clark, O.A., Toylu, A., Nur, B., Mihci, E., O'Keefe, K., Mohajeri-Stickels, K., Wilch, E.S., Kammin, T., Piña-Aguilar, R.E., Nalbandian, K., Temel, S.G., Sag, S.O., Turkgenc, B., Kamath, A., Ruiz-Herrera, A., Banka, S., Schilit, S.L.P., Currall, B.B., Yachelevich, N., Galloway, S., Chung, W.K., Raskin, S., Maya, I., Orenstein, N., Gilad, N.K., Flamenbaum, K.R., Hay, B.N., Morton, C.C., Liao, E., Choy, K.W., Gusella, J.F., Jacky, P., Aiden, E.L., Bache, I., Talkowski, M.E., Tommerup, N., International Breakpoint Mapping Consortium (IBMC), Danish Cytogenetic Central Registry Study Group, and Developmental Genome Anatomy Project (DGAP), 2022. Balanced chromosomal rearrangements offer insights into coding and noncoding genomic features associated with developmental disorders. *bioRxiv*.

Lozic, M., Minarik, L., Racetin, A., Filipovic, N., Saraga Babic, M., and Vukojevic, K., 2021. CRKL, AIFM3, AIF, BCL2, and UBASH3A during human kidney development. *International journal of molecular sciences*, 22 (17), 9183.

Lu, W., van Eerde, A.M., Fan, X., Quintero-Rivera, F., Kulkarni, S., Ferguson, H., Kim, H.-G., Fan, Y., Xi, Q., Li, Q.-G., Sanlaville, D., Andrews, W., Sundaresan, V., Bi, W., Yan, J., Giltay, J.C., Wijmenga, C., de Jong, T.P.V.M., Feather, S.A., Woolf, A.S., Rao, Y., Lupski, J.R., Eccles, M.R., Quade, B.J., Gusella, J.F., Morton, C.C., and Maas, R.L., 2007. Disruption of ROBO2 is associated with urinary tract anomalies and confers risk of vesicoureteral reflux. *The American Journal of Human Genetics*, 80 (4), 616–632.

Lu, X., Borchers, A.G.M., Jolicoeur, C., Rayburn, H., Baker, J.C., and Tessier-Lavigne, M., 2004. PTK7/CCK-4 is a novel regulator of planar cell polarity in vertebrates. *Nature*, 430 (6995), 93–98.

Luca, D., Ringquist, S., Klei, L., Lee, A.B., Gieger, C., Wichmann, H.-E., Schreiber, S., Krawczak, M., Lu, Y., Styche, A., Devlin, B., Roeder, K., and Trucco, M.,

2008. On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. *The American Journal of Human Genetics*, 82 (2), 453–463.

Ludwig, M., Reutter, H., Rüschendorf, F., Draaken, M., Betz, R., Hübner, N., Saar, K., Schäfer, N., Stein, R., Wolfenbüttel, K.P., and Nöthen, M.M., 2009. Genome-wide linkage scan in a Moroccan family with autosomal-recessive exstrophy of the bladder identifies a novel susceptibility locus on chromosome 3p25.3. *Journal of pediatric urology*, 5, S24–S25.

Ludwig, M., Rüschendorf, F., Saar, K., Hübner, N., Siekmann, L., Boyadjiev, S.A., and Reutter, H., 2009. Genome-wide linkage scan for bladder exstrophy-epispadias complex. *Birth defects research. Part A, Clinical and molecular teratology*, 85 (2), 174–178.

Lumley, T., Brody, J., Peloso, G., Morrison, A., and Rice, K., 2018. FastSKAT: Sequence kernel association tests for very large sets of markers. *Genetic epidemiology*, 42 (6), 516–527.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., Santos-Simarro, F., Gilbert-Dussardier, B., Wittler, L., Borschiwer, M., Haas, S.A., Osterwalder, M., Franke, M., Timmermann, B., Hecht, J., Spielmann, M., Visel, A., and Mundlos, S., 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161 (5), 1012–1025.

Lupski, J.R., Belmont, J.W., Boerwinkle, E., and Gibbs, R.A., 2011. Clan genomics and the complex architecture of human disease. *Cell*, 147 (1), 32–43.

Lupski, J.R., de Oca-Luna, R.M., Slaugenhaupt, S., Pentao, L., Guzzetta, V., Trask, B.J., Saucedo-Cardenas, O., Barker, D.F., Killian, J.M., Garcia, C.A., Chakravarti, A., and Patel, P.I., 1991. DNA duplication associated with Charcot-Marie-Tooth disease type 1A. *Cell*, 66 (2), 219–232.

Lupski, J.R., Reid, J.G., Gonzaga-Jauregui, C., Rio Deiros, D., Chen, D.C.Y., Nazareth, L., Bainbridge, M., Dinh, H., Jing, C., Wheeler, D.A., McGuire, A.L., Zhang, F., Stankiewicz, P., Halperin, J.J., Yang, C., Gehman, C., Guo, D., Irikat, R.K., Tom, W., Fantin, N.J., Muzny, D.M., and Gibbs, R.A., 2010. Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *The New England journal of medicine*, 362 (13), 1181–1191.

Luyckx, V.A., Bertram, J.F., Brenner, B.M., Fall, C., Hoy, W.E., Ozanne, S.E., and Vikse, B.E., 2013. Effect of fetal and child health on kidney development and long-term risk of hypertension and kidney disease. *Lancet*, 382 (9888), 273–283.

Ma, Y., Zhang, P., Wang, F., Yang, J., Yang, Z., and Qin, H., 2010. The relationship between early embryo development and tumourigenesis. *Journal of cellular and molecular medicine*, 14 (12), 2697–2701.

MacArthur, D.G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., Jostins, L., Habegger, L., Pickrell, J.K., Montgomery, S.B., Albers, C.A., Zhang, Z.D., Conrad, D.F., Lunter, G., Zheng, H., Ayub, Q., DePristo, M.A., Banks, E., Hu, M., Handsaker, R.E., Rosenfeld, J.A., Fromer, M., Jin, M., Mu,

X.J., Khurana, E., Ye, K., Kay, M., Saunders, G.I., Suner, M.-M., Hunt, T., Barnes, I.H.A., Amid, C., Carvalho-Silva, D.R., Bignell, A.H., Snow, C., Yngvadottir, B., Bumpstead, S., Cooper, D.N., Xue, Y., Romero, I.G., 1000 Genomes Project Consortium, Wang, J., Li, Y., Gibbs, R.A., McCarroll, S.A., Dermitzakis, E.T., Pritchard, J.K., Barrett, J.C., Harrow, J., Hurles, M.E., Gerstein, M.B., and Tyler-Smith, C., 2012. A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, N.Y.)*, 335 (6070), 823–828.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z.M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., and Parkinson, H., 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research*, 45 (D1), D896–D901.

Mackie, G.G. and Stephens, F.D., 1975. Duplex kidneys: a correlation of renal dysplasia with position of the ureteral orifice. *The journal of urology*, 114 (2), 274–280.

Macumber, I., Schwartz, S., and Leca, N., 2017. Maternal obesity is associated with congenital anomalies of the kidney and urinary tract in offspring. *Pediatric nephrology (Berlin, Germany)*, 32 (4), 635–642.

Madsen, B.E. and Browning, S.R., 2009. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS genetics*, 5 (2), e1000384.

Mae, S.-I., Ryosaka, M., Sakamoto, S., Matsuse, K., Nozaki, A., Igami, M., Kabai, R., Watanabe, A., and Osafune, K., 2020. Expansion of human iPSC-derived ureteric bud organoids with repeated branching potential. *Cell reports*, 32 (4), 107963.

Mahajan, A., Go, M.J., Zhang, W., Below, J.E., Gaulton, K.J., Ferreira, T., Horikoshi, M., Johnson, A.D., Ng, M.C.Y., Prokopenko, I., Saleheen, D., Wang, X., Zeggini, E., Abecasis, G.R., Adair, L.S., Almgren, P., Atalay, M., Aung, T., Baldassarre, D., Balkau, B., Bao, Y., Barnett, A.H., Barroso, I., Basit, A., Been, L.F., Beilby, J., Bell, G.I., Benediktsson, R., Bergman, R.N., Boehm, B.O., Boerwinkle, E., Bonnycastle, L.L., Burtt, N., Cai, Q., Campbell, H., Carey, J., Cauchi, S., Caulfield, M., Chan, J.C.N., Chang, L.-C., Chang, T.-J., Chang, Y.-C., Charpentier, G., Chen, C.-H., Chen, H., Chen, Y.-T., Chia, K.-S., Chidambaram, M., Chines, P.S., Cho, N.H., Cho, Y.M., Chuang, L.-M., Collins, F.S., Cornelis, M.C., Couper, D.J., Crenshaw, A.T., van Dam, R.M., Danesh, J., Das, D., de Faire, U., Dedoussis, G., Deloukas, P., Dimas, A.S., Dina, C., Doney, A.S.F., Donnelly, P.J., Dorkhan, M., van Duijn, C., Dupuis, J., Edkins, S., Elliott, P., Emilsson, V., Erbel, R., Eriksson, J.G., Escobedo, J., Esko, T., Eury, E., Florez, J.C., Fontanillas, P., Forouhi, N.G., Forsen, T., Fox, C., Fraser, R.M., Frayling, T.M., Froguel, P., Frossard, P., Gao, Y., Gertow, K., Gieger, C., Gigante, B., Grallert, H., Grant, G.B., Groop, L.C., Groves, C.J., Grundberg, E., Guiducci, C., Hamsten, A., Han, B.-G., Hara, K., Hassanali, N., Hattersley, A.T., Hayward, C., Hedman, A.K., Herder, C., Hofman, A., Holmen, O.L., Hovingh, K., Hreidarsson, A.B., Hu, C., Hu, F.B.,

Hui, J., Humphries, S.E., Hunt, S.E., Hunter, D.J., Hveem, K., Hydrie, Z.I., Ikegami, H., Illig, T., Ingelsson, E., Islam, M., Isomaa, B., Jackson, A.U., Jafar, T., James, A., Jia, W., Jöckel, K.-H., Jonsson, A., Jowett, J.B.M., Kadowaki, T., Kang, H.M., Kanoni, S., Kao, W.H.L., Kathiresan, S., Kato, N., Katulanda, P., Keinanen-Kiukaanniemi, S.M., Kelly, A.M., Khan, H., Khaw, K.-T., Khor, C.-C., Kim, H.-L., Kim, S., Kim, Y.J., Kinnunen, L., Klopp, N., Kong, A., Korpi-Hyövälti, E., Kowlessur, S., Kraft, P., Kravic, J., Kristensen, M.M., Krithika, S., Kumar, A., Kumate, J., Kuusisto, J., Kwak, S.H., Laakso, M., Lagou, V., Lakka, T.A., Langenberg, C., Langford, C., Lawrence, R., Leander, K., Lee, J.-M., Lee, N.R., Li, M., Li, X., Li, Y., Liang, J., Liju, S., Lim, W.-Y., Lind, L., Lindgren, C.M., Lindholm, E., Liu, C.-T., Liu, J.J., Lobbens, S., Long, J., Loos, R.J.F., Lu, W., Luan, J., Lyssenko, V., Ma, R.C.W., Maeda, S., Mägi, R., Männistö, S., Matthews, D.R., Meigs, J.B., Melander, O., Metspalu, A., Meyer, J., Mirza, G., Mihailov, E., Moebus, S., Mohan, V., Mohlke, K.L., Morris, A.D., Mühleisen, T.W., Müller-Nurasyid, M., Musk, B., Nakamura, J., Nakashima, E., Navarro, P., Ng, P.-K., Nica, A.C., Nilsson, P.M., Njølstad, I., Nöthen, M.M., Ohnaka, K., Ong, T.H., Owen, K.R., Palmer, C.N.A., Pankow, J.S., Park, K.S., Parkin, M., Pechlivanis, S., Pedersen, N.L., Peltonen, L., Perry, J.R.B., Peters, A., Pinidiyapathirage, J.M., Platou, C.G.P., Potter, S., Price, J.F., Qi, L., Radha, V., Rallidis, L., Rasheed, A., Rathmann, W., Rauramaa, R., Raychaudhuri, S., William Rayner, N., Rees, S.D., Rehnberg, E., Ripatti, S., Robertson, N., Roden, M., Rossin, E.J., Rudan, I., Rybin, D., Saaristo, T.E., Salomaa, V., Saltevo, J., Samuel, M., Sanghera, D.K., Saramies, J., Scott, J., Scott, L.J., Scott, R.A., Segrè, A.V., Sehmi, J., Sennblad, B., Shah, N., Shah, S., Samad Shera, A., Shu, X.O., Shuldiner, A.R., Sigurðsson, G., Sijbrands, E., Silveira, A., Sim, X., Sivapalaratnam, S., Small, K.S., So, W.Y., Stančáková, A., Stefansson, K., Steinbach, G., Steinthorsdottir, V., Stirrups, K., Strawbridge, R.J., Stringham, H.M., Sun, Q., Suo, C., Syvänen, A.-C., Takayanagi, R., Takeuchi, F., Tay, W.T., Teslovich, T.M., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tikkanen, E., Trakalo, J., Tremoli, E., Trip, M.D., Tsai, F.J., Tuomi, T., Tuomilehto, J., Uitterlinden, A.G., Valladares-Salgado, A., Vedantam, S., Veglia, F., Voight, B.F., Wang, C., Wareham, N.J., Wennauer, R., Wickremasinghe, A.R., Wilsgaard, T., Wilson, J.F., Wiltshire, S., Winckler, W., Wong, T.Y., Wood, A.R., Wu, J.-Y., Wu, Y., Yamamoto, K., Yamauchi, T., Yang, M., Yengo, L., Yokota, M., Young, R., Zabaneh, D., Zhang, F., Zhang, R., Zheng, W., Zimmet, P.Z., Altshuler, D., Bowden, D.W., Cho, Y.S., Cox, N.J., Cruz, M., Hanis, C.L., Kooner, J., Lee, J.-Y., Seielstad, M., Teo, Y.Y., Boehnke, M., Parra, E.J., Chambers, J.C., Shyong Tai, E., McCarthy, M.I., and Morris, A.P., 2014. Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature genetics*, 46 (3), 234–244.

Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.-M., 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics* , 26 (22), 2867–2873.

Mann, N., Braun, D.A., Amann, K., Tan, W., Shril, S., Connaughton, D.M., Nakayama, M., Schneider, R., Kitzler, T.M., van der Ven, A.T., Chen, J., Ityel, H., Vivante, A., Majmundar, A.J., Daga, A., Warejko, J.K., Lovric, S., Ashraf, S., Jobst-Schwan, T., Widmeier, E., Hugo, H., Mane, S.M., Spaneas, L., Somers, M.J.G., Ferguson, M.A., Traum, A.Z., Stein, D.R., Baum, M.A., Daouk, G.H., Lifton, R.P., Manzi, S., Vakili, K., Kim, H.B., Rodig, N.M., and Hildebrandt, F., 2019. Whole-exome sequencing enables a precision medicine approach for kidney transplant recipients. *Journal of the American Society of Nephrology: JASN*, 30 (2), 201–215.

Mann, N., Kause, F., Henze, E.K., Gharpure, A., Shril, S., Connaughton, D.M., Nakayama, M., Klämbt, V., Majmundar, A.J., Wu, C.-H.W., Kolvenbach, C.M., Dai, R., Chen, J., van der Ven, A.T., Ityel, H., Tooley, M.J., Kari, J.A., Bownass, L., El Desoky, S., De Franco, E., Shalaby, M., Tasic, V., Bauer, S.B., Lee, R.S., Beckel, J.M., Yu, W., Mane, S.M., Lifton, R.P., Reutter, H., Ellard, S., Hibbs, R.E., Kawate, T., and Hildebrandt, F., 2019. CAKUT and autonomic dysfunction caused by acetylcholine receptor mutations. *The American Journal of Human Genetics*, 105 (6), 1286–1293.

Manoharan, A., Krishnamurthy, S., Sivamurukan, P., Ananthakrishnan, R., and Jindal, B., 2020. 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 journal of pediatrics*, 87 (9), 686–691.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., and Visscher, P.M., 2009. Finding the missing heritability of complex diseases. *Nature*, 461 (7265), 747–753.

Marigorta, U.M. and Navarro, A., 2013. High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS genetics*, 9 (6), e1003566.

Mark, M., Ghyselinck, N.B., and Chambon, P., 2009. Function of retinoic acid receptors during embryonic development. *Nuclear receptor signaling*, 7 (1), e002.

Mars, N., Widén, E., Kerminen, S., Meretoja, T., Pirinen, M., Della Briotta Parolo, P., Palta, P., FinnGen, Palotie, A., Kaprio, J., Joensuu, H., Daly, M., and Ripatti, S., 2020. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nature communications*, 11 (1), 6383.

Marshall, C.R., Howrigan, D.P., Merico, D., Thiruvahindrapuram, B., Wu, W., Greer, D.S., Antaki, D., Shetty, A., Holmans, P.A., Pinto, D., Gujral, M., Brandler, W.M., Malhotra, D., Wang, Z., Fajarado, K.V.F., Maile, M.S., Ripke, S., Agartz, I., Albus, M., Alexander, M., Amin, F., Atkins, J., Bacanu, S.A., Belliveau, R.A., Jr, Bergen, S.E., Bertalan, M., Bevilacqua, E., Bigdeli, T.B., Black, D.W., Bruggeman, R., Buccola, N.G., Buckner, R.L., Bulik-Sullivan, B.,

Byerley, W., Cahn, W., Cai, G., Cairns, M.J., Campion, D., Cantor, R.M., Carr, V.J., Carrera, N., Catts, S.V., Chambert, K.D., Cheng, W., Cloninger, C.R., Cohen, D., Cormican, P., Craddock, N., Crespo-Facorro, B., Crowley, J.J., Curtis, D., Davidson, M., Davis, K.L., Degenhardt, F., Del Favero, J., DeLisi, L.E., Dikeos, D., Dinan, T., Djurovic, S., Donohoe, G., Drapeau, E., Duan, J., Dudbridge, F., Eichhammer, P., Eriksson, J., Escott-Price, V., Essioux, L., Fanous, A.H., Farh, K.-H., Farrell, M.S., Frank, J., Franke, L., Freedman, R., Freimer, N.B., Friedman, J.I., Forstner, A.J., Fromer, M., Genovese, G., Georgieva, L., Gershon, E.S., Giegling, I., Giusti-Rodríguez, P., Godard, S., Goldstein, J.I., Gratten, J., de Haan, L., Hamshere, M.L., Hansen, M., Hansen, T., Haroutunian, V., Hartmann, A.M., Henskens, F.A., Herms, S., Hirschhorn, J.N., Hoffmann, P., Hofman, A., Huang, H., Ikeda, M., Joa, I., Kähler, A.K., Kahn, R.S., Kalaydjieva, L., Karjalainen, J., Kavanagh, D., Keller, M.C., Kelly, B.J., Kennedy, J.L., Kim, Y., Knowles, J.A., Konte, B., Laurent, C., Lee, P., Lee, S.H., Legge, S.E., Lerer, B., Levy, D.L., Liang, K.-Y., Lieberman, J., Lönnqvist, J., Loughland, C.M., Magnusson, P.K.E., Maher, B.S., Maier, W., Mallet, J., Mattheisen, M., Mattingsdal, M., McCarley, R.W., McDonald, C., McIntosh, A.M., Meier, S., Meijer, C.J., Melle, I., Mesholam-Gately, R.I., Metspalu, A., Michie, P.T., Milani, L., Milanova, V., Mokrab, Y., Morris, D.W., Müller-Myhsok, B., Murphy, K.C., Murray, R.M., Myin-Germeys, I., Nenadic, I., Nertney, D.A., Nestadt, G., Nicodemus, K.K., Nisenbaum, L., Nordin, A., O'Callaghan, E., O'Dushlaine, C., Oh, S.-Y., Olincy, A., Olsen, L., O'Neill, F.A., Van Os, J., Pantelis, C., Papadimitriou, G.N., Parkhomenko, E., Pato, M.T., Paunio, T., Psychosis Endophenotypes International Consortium, Perkins, D.O., Pers, T.H., Pietiläinen, O., Pimm, J., Pocklington, A.J., Powell, J., Price, A., Pulver, A.E., Purcell, S.M., Quested, D., Rasmussen, H.B., Reichenberg, A., Reimers, M.A., Richards, A.L., Roffman, J.L., Roussos, P., Ruderfer, D.M., Salomaa, V., Sanders, A.R., Savitz, A., Schall, U., Schulze, T.G., Schwab, S.G., Scolnick, E.M., Scott, R.J., Seidman, L.J., Shi, J., Silverman, J.M., Smoller, J.W., Söderman, E., Spencer, C.C.A., Stahl, E.A., Strengman, E., Strohmaier, J., Stroup, T.S., Suvisaari, J., Svrakic, D.M., Szatkiewicz, J.P., Thirumalai, S., Tooney, P.A., Veijola, J., Visscher, P.M., Waddington, J., Walsh, D., Webb, B.T., Weiser, M., Wildenauer, D.B., Williams, N.M., Williams, S., Witt, S.H., Wolen, A.R., Wormley, B.K., Wray, N.R., Wu, J.Q., Zai, C.C., Adolfsson, R., Andreassen, O.A., Blackwood, D.H.R., Bramon, E., Buxbaum, J.D., Cichon, S., Collier, D.A., Corvin, A., Daly, M.J., Darvasi, A., Domenici, E., Esko, T., Gejman, P.V., Gill, M., Gurling, H., Hultman, C.M., Iwata, N., Jablensky, A.V., Jönsson, E.G., Kendler, K.S., Kirov, G., Knight, J., Levinson, D.F., Li, Q.S., McCarroll, S.A., McQuillin, A., Moran, J.L., Mowry, B.J., Nöthen, M.M., Ophoff, R.A., Owen, M.J., Palotie, A., Pato, C.N., Petryshen, T.L., Posthuma, D., Rietschel, M., Riley, B.P., Rujescu, D., Sklar, P., St Clair, D., Walters, J.T.R., Werge, T., Sullivan, P.F., O'Donovan, M.C., Scherer, S.W., Neale, B.M., Sebat, J., and CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium,

2017. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature genetics*, 49 (1), 27–35.

Martínez, J.M., Masoller, N., Devlieger, R., Passchyn, E., Gómez, O., Rodo, J., Deprest, J.A., and Gratacós, E., 2015. Laser ablation of posterior urethral valves by fetal cystoscopy. *Fetal diagnosis and therapy*, 37 (4), 267–273.

Matsell, D.G., Mok, A., and Tarantal, A.F., 2002. Altered primate glomerular development due to in utero urinary tract obstruction. *Kidney international*, 61 (4), 1263–1269.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R., and Stamatoyannopoulos, J.A., 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science (New York, N.Y.)*, 337 (6099), 1190–1195.

Maxam, A.M. and Gilbert, W., 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74 (2), 560–564.

McDaniell, R., Warthen, D.M., Sanchez-Lara, P.A., Pai, A., Krantz, I.D., Piccoli, D.A., and Spinner, N.B., 2006. NOTCH2 mutations cause Alagille syndrome, a heterogeneous disorder of the notch signaling pathway. *The American Journal of Human Genetics*, 79 (1), 169–173.

McGregor, L., Makela, V., Darling, S.M., Vrontou, S., Chalepakis, G., Roberts, C., Smart, N., Rutland, P., Prescott, N., Hopkins, J., Bentley, E., Shaw, A., Roberts, E., Mueller, R., Jadeja, S., Philip, N., Nelson, J., Francannet, C., Perez-Aytes, A., Megarbane, A., Kerr, B., Wainwright, B., Woolf, A.S., Winter, R.M., and Scambler, P.J., 2003. Fraser syndrome and mouse blebbed phenotype caused by mutations in FRAS1/Fras1 encoding a putative extracellular matrix protein. *Nature genetics*, 34 (2), 203–208.

McKay, A.M., Kim, S., and Kennedy, S.E., 2019. Long-term outcome of kidney transplantation in patients with congenital anomalies of the kidney and urinary tract. *Pediatric nephrology* , 34 (11), 2409–2415.

McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F., 2016. The Ensembl Variant Effect Predictor. *Genome biology*, 17 (1), 122.

McPherson, E., 2007. Renal anomalies in families of individuals with congenital solitary kidney. *Genetics in medicine: official journal of the American College of Medical Genetics*, 9 (5), 298–302.

McPherson, E., Carey, J., Kramer, A., Hall, J.G., Pauli, R.M., Schimke, R.N., and Tasin, M.H., 1987. Dominantly inherited renal adysplasia. *American journal of medical genetics*, 26 (4), 863–872.

Mefford, H.C., Clauin, S., Sharp, A.J., Moller, R.S., Ullmann, R., Kapur, R., Pinkel, D., Cooper, G.M., Ventura, M., Ropers, H.H., Tommerup, N., Eichler, E.E.,

and Bellanne-Chantelot, C., 2007. Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *The American Journal of Human Genetics*, 81 (5), 1057–1069.

Melo, U.S., Piard, J., Fischer-Zirnsak, B., Klever, M.-K., Schöpflin, R., Mensah, M.A., Holtgrewe, M., Arbez-Gindre, F., Martin, A., Guigue, V., Gaillard, D., Landais, E., Roze, V., Kremer, V., Ramanah, R., Cabrol, C., Harms, F.L., Kornak, U., Spielmann, M., Mundlos, S., and Van Maldergem, L., 2021. Complete lung agenesis caused by complex genomic rearrangements with neo-TAD formation at the SHH locus. *Human genetics*, 140 (10), 1459–1469.

Mendelsohn, C., Lohnes, D., Décimo, D., Lufkin, T., LeMeur, M., Chambon, P., and Mark, M., 1994. Function of the retinoic acid receptors (RARs) during development (II). Multiple abnormalities at various stages of organogenesis in RAR double mutants. *Development (Cambridge, England)*, 120 (10), 2749–2771.

Mikhaylova, A.V., McHugh, C.P., Polfus, L.M., Raffield, L.M., Boorgula, M.P., Blackwell, T.W., Brody, J.A., Broome, J., Chami, N., Chen, M.-H., Conomos, M.P., Cox, C., Curran, J.E., Daya, M., Ekunwe, L., Glahn, D.C., Heard-Costa, N., Highland, H.M., Hobbs, B.D., Ilboudo, Y., Jain, D., Lange, L.A., Miller-Fleming, T.W., Min, N., Moon, J.-Y., Preuss, M.H., Rosen, J., Ryan, K., Smith, A.V., Sun, Q., Surendran, P., de Vries, P.S., Walter, K., Wang, Z., Wheeler, M., Yanek, L.R., Zhong, X., Abecasis, G.R., Almasy, L., Barnes, K.C., Beaty, T.H., Becker, L.C., Blangero, J., Boerwinkle, E., Butterworth, A.S., Chavan, S., Cho, M.H., Choquet, H., Correa, A., Cox, N., DeMeo, D.L., Faraday, N., Fornage, M., Gerszten, R.E., Hou, L., Johnson, A.D., Jorgenson, E., Kaplan, R., Kooperberg, C., Kundu, K., Laurie, C.A., Lettre, G., Lewis, J.P., Li, B., Li, Y., Lloyd-Jones, D.M., Loos, R.J.F., Manichaikul, A., Meyers, D.A., Mitchell, B.D., Morrison, A.C., Ngo, D., Nickerson, D.A., Nongmaithem, S., North, K.E., O'Connell, J.R., Ortega, V.E., Pankratz, N., Perry, J.A., Psaty, B.M., Rich, S.S., Soranzo, N., Rotter, J.I., Silverman, E.K., Smith, N.L., Tang, H., Tracy, R.P., Thornton, T.A., Vasan, R.S., Zein, J., Mathias, R.A., NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Reiner, A.P., and Auer, P.L., 2021. Whole-genome sequencing in diverse subjects identifies genetic correlates of leukocyte traits: The NHLBI TOPMed program. *The American Journal of Human Genetics*, 108 (10), 1836–1851.

Miller, D.T., Lee, K., Chung, W.K., Gordon, A.S., Herman, G.E., Klein, T.E., Stewart, D.R., Amendola, L.M., Adelman, K., Bale, S.J., Gollob, M.H., Harrison, S.M., Hershberger, R.E., McKelvey, K., Richards, C.S., Vlangos, C.N., Watson, M.S., Martin, C.L., and ACMG Secondary Findings Working Group, 2021. ACMG SF v3.0 list for reporting of secondary findings in clinical exome and genome sequencing: a policy statement of the American College of Medical Genetics and Genomics (ACMG). *Genetics in medicine: official journal of the American College of Medical Genetics*, 23 (8), 1381–1390.

Mills, M.C. and Rahal, C., 2019. A scientometric review of genome-wide association studies. *Communications biology*, 2 (1), 9.

Moazin, M.S., Ahmed, S., and Fouda-Neel, K., 1997. Multicystic kidney in siblings. *Journal of pediatric surgery*, 32 (1), 119–120.

Moore, C.M., Jacobson, S.A., and Fingerlin, T.E., 2019. Power and Sample Size Calculations for Genetic Association Studies in the Presence of Genetic Model Misspecification. *Human heredity*, 84 (6), 256–271.

Morais, M.R.P.T., Tian, P., Lawless, C., Murtuza-Baker, S., Hopkinson, L., Woods, S., Mironov, A., Long, D.A., Gale, D.P., Zorn, T.M.T., Kimber, S.J., Zent, R., and Lennon, R., 2022. Kidney organoids recapitulate human basement membrane assembly in health and disease. *eLife*, 11.

Morgenthaler, S. and Thilly, W.G., 2007. A strategy to discover genes that carry multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test (CAST). *Mutation research*, 615 (1–2), 28–56.

Morris, A.P., Le, T.H., Wu, H., Akbarov, A., van der Most, P.J., Hemani, G., Smith, G.D., Mahajan, A., Gaulton, K.J., Nadkarni, G.N., Valladares-Salgado, A., Wacher-Rodarte, N., Mychaleckyj, J.C., Dueker, N.D., Guo, X., Hai, Y., Haessler, J., Kamatani, Y., Stilp, A.M., Zhu, G., Cook, J.P., Ärnlöv, J., Blanton, S.H., de Borst, M.H., Bottinger, E.P., Buchanan, T.A., Cechova, S., Charchar, F.J., Chu, P.-L., Damman, J., Eales, J., Gharavi, A.G., Giedraitis, V., Heath, A.C., Ipp, E., Kiryluk, K., Kramer, H.J., Kubo, M., Larsson, A., Lindgren, C.M., Lu, Y., Madden, P.A.F., Montgomery, G.W., Papanicolaou, G.J., Raffel, L.J., Sacco, R.L., Sanchez, E., Stark, H., Sundstrom, J., Taylor, K.D., Xiang, A.H., Zivkovic, A., Lind, L., Ingelsson, E., Martin, N.G., Whitfield, J.B., Cai, J., Laurie, C.C., Okada, Y., Matsuda, K., Kooperberg, C., Chen, Y.-D.I., Rundek, T., Rich, S.S., Loos, R.J.F., Parra, E.J., Cruz, M., Rotter, J.I., Snieder, H., Tomaszewski, M., Humphreys, B.D., and Franceschini, N., 2019. Trans-ethnic kidney function association study reveals putative causal genes and effects on kidney-specific disease aetiologies. *Nature communications*, 10 (1), 29.

Morris, J.K., Springett, A.L., Greenlees, R., Loane, M., Addor, M.-C., Arriola, L., Barisic, I., Bergman, J.E.H., Csaky-Szunyogh, M., Dias, C., Draper, E.S., Garne, E., Gatt, M., Khoshnood, B., Klungsoyr, K., Lynch, C., McDonnell, R., Nelen, V., Neville, A.J., O'Mahony, M., Pierini, A., Queisser-Luft, A., Randrianaivo, H., Rankin, J., Rissmann, A., Kurinczuk, J., Tucker, D., Verellen-Dumoulin, C., Wellesley, D., and Dolk, H., 2018. Trends in congenital anomalies in Europe from 1980 to 2012. *PloS one*, 13 (4), e0194986.

Morris, R.K., Malin, G.L., Quinlan-Jones, E., Middleton, L.J., Hemming, K., Burke, D., Daniels, J.P., Khan, K.S., Deeks, J., Kilby, M.D., and Percutaneous vesicoamniotic shunting in Lower Urinary Tract Obstruction (PLUTO) Collaborative Group, 2013. Percutaneous vesicoamniotic shunting versus conservative management for fetal lower urinary tract obstruction (PLUTO): a randomised trial. *Lancet*, 382 (9903), 1496–1506.

Mossie, K., Jallal, B., Alves, F., Sures, I., Plowman, G.D., and Ullrich, A., 1995. Colon carcinoma kinase-4 defines a new subclass of the receptor tyrosine kinase family. *Oncogene*, 11 (10), 2179–2184.

Mountjoy, E., Schmidt, E.M., Carmona, M., Schwartzentruber, J., Peat, G., Miranda, A., Fumis, L., Hayhurst, J., Buniello, A., Karim, M.A., Wright, D., Hercules, A., Papa, E., Fauman, E.B., Barrett, J.C., Todd, J.A., Ochoa, D., Dunham, I., and Ghoussaini, M., 2021. An open approach to systematically prioritize causal variants and genes at all published human GWAS trait-associated loci. *Nature genetics*, 53 (11), 1527–1533.

Münch, J., Engesser, M., Schönauer, R., Hamm, J.A., Hartig, C., Hantmann, E., Akay, G., Pehlivan, D., Mitani, T., Akdemir, Z.C., Tüysüz, B., Shirakawa, T., Dateki, S., Claus, L.R., van Eerde, A.M., Genomics England Research Consortium, Smol, T., Devisme, L., Franquet, H., Attié-Bitach, T., Wagner, T., Bergmann, C., Höhn, A.K., Shril, S., Pollack, A., Wegner, T., Scott, A., Paolucci, S., Buchan, J., Gabriel, G.C., Posey, J.E., Lupski, J.R., Petit, F., McCarthy, A.A., Pazour, G.J., Lo, C.W., Popp, B., and Halbritter, J., 2022. Biallelic pathogenic variants in roundabout guidance receptor 1 associate with syndromic congenital anomalies of the kidney and urinary tract. *Kidney international*.

Murugasu, B., Cole, B.R., Hawkins, E.P., Blanton, S.H., Conley, S.B., and Portman, R.J., 1991. Familial renal adysplasia. *American journal of kidney diseases: the official journal of the National Kidney Foundation*, 18 (4), 490–494.

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., Ninomiya, T., Tamakoshi, A., Yamagata, Z., Mushiroda, T., Murakami, Y., Yuji, K., Furukawa, Y., Zembutsu, H., Tanaka, T., Ohnishi, Y., Nakamura, Y., BioBank Japan Cooperative Hospital Group, and Kubo, M., 2017. Overview of the BioBank Japan Project: Study design and profile. *Journal of epidemiology*, 27 (3S), S2–S8.

Nakayama, M., Nozu, K., Goto, Y., Kamei, K., Ito, S., Sato, H., Emi, M., Nakanishi, K., Tsuchiya, S., and Iijima, K., 2010. HNF1B alterations associated with congenital anomalies of the kidney and urinary tract. *Pediatric nephrology (Berlin, Germany)*, 25 (6), 1073–1079.

Namjou, B., Ni, Y., Harley, I.T.W., Chepelev, I., Cobb, B., Kottyan, L.C., Gaffney, P.M., Guthridge, J.M., Kaufman, K., and Harley, J.B., 2014. The effect of inversion at 8p23 on BLK association with lupus in Caucasian population. *PloS one*, 9 (12), e115614.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D., 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science (New York, N.Y.)*, 347 (6225), 1017–1021.

Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., Mualim, K., Natri, H.M., Weeks, E.M., Munson, G., Kane, M., Kang, H.Y., Cui, A., Ray, J.P., Eisenhaure, T.M., Collins, R.L., Dey, K., Pfister, H., Price, A.L.,

Epstein, C.B., Kundaje, A., Xavier, R.J., Daly, M.J., Huang, H., Finucane, H.K., Hacohen, N., Lander, E.S., and Engreitz, J.M., 2021. Genome-wide enhancer maps link risk variants to disease genes. *Nature*, 593 (7858), 238–243.

Natarajan, P., Peloso, G.M., Zekavat, S.M., Montasser, M., Ganna, A., Chaffin, M., Khera, A.V., Zhou, W., Bloom, J.M., Engreitz, J.M., Ernst, J., O'Connell, J.R., Ruotsalainen, S.E., Alver, M., Manichaikul, A., Johnson, W.C., Perry, J.A., Poterba, T., Seed, C., Surakka, I.L., Esko, T., Ripatti, S., Salomaa, V., Correa, A., Vasan, R.S., Kellis, M., Neale, B.M., Lander, E.S., Abecasis, G., Mitchell, B., Rich, S.S., Wilson, J.G., Cupples, L.A., Rotter, J.I., Willer, C.J., Kathiresan, S., and NHLBI TOPMed Lipids Working Group, 2018. Deep-coverage whole genome sequences and blood lipids among 16,324 individuals. *Nature communications*, 9 (1), 3391.

Negrisolo, S., Centi, S., Benetti, E., Ghirardo, G., Della Vella, M., Murer, L., and Artifoni, L., 2014. SIX1 gene: absence of mutations in children with isolated congenital anomalies of kidney and urinary tract. *Journal of nephrology*, 27 (6), 667–671.

Ng, S.B., Buckingham, K.J., Lee, C., Bigham, A.W., Tabor, H.K., Dent, K.M., Huff, C.D., Shannon, P.T., Jabs, E.W., Nickerson, D.A., Shendure, J., and Bamshad, M.J., 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nature genetics*, 42 (1), 30–35.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E., Bamshad, M., Nickerson, D.A., and Shendure, J., 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, 461 (7261), 272–276.

Nicolaou, N., Pulit, S.L., Nijman, I.J., Monroe, G.R., Feitz, W.F.J., Schreuder, M.F., van Eerde, A.M., de Jong, T.P.V.M., Giltay, J.C., van der Zwaag, B., Havenith, M.R., Zwakenberg, S., van der Zanden, L.F.M., Poelmans, G., Cornelissen, E.A.M., Lilien, M.R., Franke, B., Roeleveld, N., van Rooij, I.A.L.M., Cuppen, E., Bongers, E.M.H.F., Giles, R.H., Knoers, N.V.A.M., and Renkema, K.Y., 2016. Prioritization and burden analysis of rare variants in 208 candidate genes suggest they do not play a major role in CAKUT. *Kidney international*, 89 (2), 476–486.

Nicolaou, N., Renkema, K.Y., Bongers, E.M.H.F., Giles, R.H., and Knoers, N.V.A.M., 2015. Genetic, environmental, and epigenetic factors involved in CAKUT. *Nature reviews. Nephrology*, 11 (12), 720–731.

Niemi, M.E.K., Martin, H.C., Rice, D.L., Gallone, G., Gordon, S., Kelemen, M., McAloney, K., McRae, J., Radford, E.J., Yu, S., Gecz, J., Martin, N.G., Wright, C.F., Fitzpatrick, D.R., Firth, H.V., Hurles, M.E., and Barrett, J.C., 2018. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *Nature*, 562 (7726), 268–271.

Nishigori, H., Yamada, S., Kohama, T., Tomura, H., Sho, K., Horikawa, Y., Bell, G.I., Takeuchi, T., and Takeda, J., 1998. Frameshift mutation, A263fsinsGG, in the

hepatocyte nuclear factor-1beta gene associated with diabetes and renal dysfunction. *Diabetes*, 47 (8), 1354–1355.

Nishinakamura, R. and Sakaguchi, M., 2014. BMP signaling and its modifiers in kidney development. *Pediatric nephrology (Berlin, Germany)*, 29 (4), 681–686.

Noe, H.N., 1992. The long-term results of prospective sibling reflux screening. *The journal of urology*, 148 (5 Part 2), 1739–1742.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G., 2017. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell*, 169 (5), 930-944.e22.

Northcott, P.A., Buchhalter, I., Morrissy, A.S., Hovestadt, V., Weischenfeldt, J., Ehrenberger, T., Gröbner, S., Segura-Wang, M., Zichner, T., Rudneva, V.A., Warnatz, H.-J., Sidiropoulos, N., Phillips, A.H., Schumacher, S., Kleinheinz, K., Waszak, S.M., Erkek, S., Jones, D.T.W., Worst, B.C., Kool, M., Zapatka, M., Jäger, N., Chavez, L., Hutter, B., Bieg, M., Paramasivam, N., Heinold, M., Gu, Z., Ishaque, N., Jäger-Schmidt, C., Imbusch, C.D., Jugold, A., Hübschmann, D., Risch, T., Amstislavskiy, V., Gonzalez, F.G.R., Weber, U.D., Wolf, S., Robinson, G.W., Zhou, X., Wu, G., Finkelstein, D., Liu, Y., Cavalli, F.M.G., Luu, B., Ramaswamy, V., Wu, X., Koster, J., Ryzhova, M., Cho, Y.-J., Pomeroy, S.L., Herold-Mende, C., Schuhmann, M., Ebinger, M., Liau, L.M., Mora, J., McLendon, R.E., Jabado, N., Kumabe, T., Chuah, E., Ma, Y., Moore, R.A., Mungall, A.J., Mungall, K.L., Thiessen, N., Tse, K., Wong, T., Jones, S.J.M., Witt, O., Milde, T., Von Deimling, A., Capper, D., Korshunov, A., Yaspo, M.-L., Kriwacki, R., Gajjar, A., Zhang, J., Beroukhim, R., Fraenkel, E., Korbel, J.O., Brors, B., Schlesner, M., Eils, R., Marra, M.A., Pfister, S.M., Taylor, M.D., and Lichter, P., 2017. The whole-genome landscape of medulloblastoma subtypes. *Nature*, 547 (7663), 311–317.

Ntzani, E.E., Liberopoulos, G., Manolio, T.A., and Ioannidis, J.P.A., 2012. Consistency of genome-wide associations across major ancestral groups. *Human genetics*, 131 (7), 1057–1071.

Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., Aganezov, S., Hoyt, S.J., Diekhans, M., Logsdon, G.A., Alonge, M., Antonarakis, S.E., Borchers, M., Bouffard, G.G., Brooks, S.Y., Caldas, G.V., Chen, N.-C., Cheng, H., Chin, C.-S., Chow, W., de Lima, L.G., Dishuck, P.C., Durbin, R., Dvorkina, T., Fiddes, I.T., Formenti, G., Fulton, R.S., Fungtammasan, A., Garrison, E., Grady, P.G.S., Graves-Lindsay, T.A., Hall, I.M., Hansen, N.F., Hartley, G.A., Haukness, M., Howe, K., Hunkapiller, M.W., Jain, C., Jain, M., Jarvis, E.D., Kerpedjiev, P., Kirsche, M., Kolmogorov, M., Korlach, J., Kremitzki, M., Li, H., Maduro, V.V., Marschall, T., McCartney, A.M., McDaniel, J., Miller, D.E., Mullikin, J.C., Myers, E.W., Olson, N.D., Paten, B., Peluso, P., Pevzner, P.A., Porubsky, D., Potapova, T., Rogaev, E.I., Rosenfeld, J.A., Salzberg, S.L., Schneider, V.A., Sedlazeck, F.J., Shafin, K., Shew, C.J., Shumate, A., Sims,

Y., Smit, A.F.A., Soto, D.C., Sović, I., Storer, J.M., Streets, A., Sullivan, B.A., Thibaud-Nissen, F., Torrance, J., Wagner, J., Walenz, B.P., Wenger, A., Wood, J.M.D., Xiao, C., Yan, S.M., Young, A.C., Zarate, S., Surti, U., McCoy, R.C., Dennis, M.Y., Alexandrov, I.A., Gerton, J.L., O'Neill, R.J., Timp, W., Zook, J.M., Schatz, M.C., Eichler, E.E., Miga, K.H., and Phillippy, A.M., 2022. The complete sequence of a human genome. *Science (New York, N.Y.)*, 376 (6588), 44–53.

Ober, C., Loisel, D.A., and Gilad, Y., 2008. Sex-specific genetic architecture of human disease. *Nature reviews. Genetics*, 9 (12), 911–922.

Oetjens, M.T., Kelly, M.A., Sturm, A.C., Martin, C.L., and Ledbetter, D.H., 2019. Quantifying the polygenic contribution to variable expressivity in eleven rare genetic disorders. *Nature communications*, 10 (1), 4897.

Olfson, E., Cottrell, C.E., Davidson, N.O., Gurnett, C.A., Heusel, J.W., Stitziel, N.O., Chen, L.-S., Hartz, S., Nagarajan, R., Saccone, N.L., and Bierut, L.J., 2015. Identification of medically actionable secondary findings in the 1000 Genomes. *PloS one*, 10 (9), e0135193.

Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A.D.H., Cotter, D.J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., Castel, S.E., Mohammadi, P., Aguet, F., Zou, Y., Khramtsova, E.A., Skol, A.D., Garrido-Martín, D., Reverter, F., Brown, A., Evans, P., Gamazon, E.R., Payne, A., Bonazzola, R., Barbeira, A.N., Hamel, A.R., Martinez-Perez, A., Soria, J.M., Pierce, B.L., Stephens, M., Eskin, E., Dermitzakis, E.T., Segrè, A.V., Im, H.K., Engelhardt, B.E., Ardlie, K.G., Montgomery, S.B., Battle, A.J., Lappalainen, T., Guigó, R., Stranger, B.E., and GTEx Consortium§, 2020. The impact of sex on gene expression across human tissues. *Science (New York, N.Y.)*, 369 (6509), eaba3066.

Osborne, L.R., Li, M., Pober, B., Chitayat, D., Bodurtha, J., Mandel, A., Costa, T., Grebe, T., Cox, S., Tsui, L.C., and Scherer, S.W., 2001. A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. *Nature genetics*, 29 (3), 321–325.

Osterwalder, M., Barozzi, I., Tissières, V., Fukuda-Yuzawa, Y., Mannion, B.J., Afzal, S.Y., Lee, E.A., Zhu, Y., Plajzer-Frick, I., Pickle, C.S., Kato, M., Garvin, T.H., Pham, Q.T., Harrington, A.N., Akiyama, J.A., Afzal, V., Lopez-Rios, J., Dickel, D.E., Visel, A., and Pennacchio, L.A., 2018. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature*, 554 (7691), 239–243.

Pan, T., Sun, J., Hu, J., Hu, Y., Zhou, J., Chen, Z., Xu, D., Xu, W., Zheng, S., and Zhang, S., 2014. Cytohesins/ARNO: the function in colorectal cancer cells. *PloS one*, 9 (3), e90997.

Pang, J., Zhang, S., Yang, P., Hawkins-Lee, B., Zhong, J., Zhang, Y., Ochoa, B., Agundez, J.A.G., Voelckel, M.-A., Fisher, R.B., Gu, W., Xiong, W.-C., Mei, L., She, J.-X., and Wang, C.-Y., 2010. Loss-of-function mutations in HPSE2 cause the autosomal recessive urofacial syndrome. *The American Journal of Human Genetics*, 87 (1), 161.

Paquette, M., Chong, M., Thériault, S., Dufour, R., Paré, G., and Baass, A., 2017. Polygenic risk score predicts prevalence of cardiovascular disease in patients with familial hypercholesterolemia. *Journal of clinical lipidology*, 11 (3), 725-732.e5.

Parimi, M. and Nitsch, D., 2020. A Systematic Review and Meta-Analysis of Diabetes During Pregnancy and Congenital Genitourinary Abnormalities. *Kidney international reports*, 5 (5), 678–693.

Patterson, N., Price, A.L., and Reich, D., 2006. Population structure and eigenanalysis. *PLoS genetics*, 2 (12), e190.

Peters, C.A., Carr, M.C., Lais, A., Retik, A.B., and Mandell, J., 1992. The response of the fetal kidney to obstruction. *The journal of urology*, 148 (2 Part 2), 503–509.

Peterson, R.E., Kuchenbaecker, K., Walters, R.K., Chen, C.-Y., Popejoy, A.B., Periyasamy, S., Lam, M., Iyegbe, C., Strawbridge, R.J., Brick, L., Carey, C.E., Martin, A.R., Meyers, J.L., Su, J., Chen, J., Edwards, A.C., Kalungi, A., Koen, N., Majara, L., Schwarz, E., Smoller, J.W., Stahl, E.A., Sullivan, P.F., Vassos, E., Mowry, B., Prieto, M.L., Cuellar-Barboza, A., Bigdeli, T.B., Edenberg, H.J., Huang, H., and Duncan, L.E., 2019. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell*, 179 (3), 589–603.

Pietilä, I., Prunskaite-Hyyryläinen, R., Kaisto, S., Tika, E., van Eerde, A.M., Salo, A.M., Garma, L., Miinalainen, I., Feitz, W.F., Bongers, E.M.H.F., Juffer, A., Knoers, N.V.A.M., Renkema, K.Y., Myllyharju, J., and Vainio, S.J., 2016. Wnt5a deficiency leads to anomalies in ureteric tree development, tubular epithelial cell organization and basement membrane integrity pointing to a role in kidney collecting duct patterning. *PloS one*, 11 (1), e0147171.

Pilia, G., Hughes-Benzie, R.M., MacKenzie, A., Baybayan, P., Chen, E.Y., Huber, R., Neri, G., Cao, A., Forabosco, A., and Schlessinger, D., 1996. Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nature genetics*, 12 (3), 241–247.

Pope, J.C., 4th, Brock, J.W., 3rd, Adams, M.C., Stephens, F.D., and Ichikawa, I., 1999. How they begin and how they end: classic and new theories for the development and deterioration of congenital anomalies of the kidney and urinary tract, CAKUT. *Journal of the American Society of Nephrology: JASN*, 10 (9), 2018–2028.

Potter, E.L., 1946. Facial characteristics of infants with bilateral renal agenesis. *American journal of obstetrics and gynecology*, 51 (6), 885–888.

Prato, A.P., Musso, M., Ceccherini, I., Mattioli, G., Giunta, C., Ghiggeri, G.M., and Jasonni, V., 2009. Hirschsprung disease and congenital anomalies of the kidney and urinary tract (CAKUT): a novel syndromic association. *Medicine*, 88 (2), 83–90.

Price, A.L., Kryukov, G.V., de Bakker, P.I.W., Purcell, S.M., Staples, J., Wei, L.-J., and Sunyaev, S.R., 2010. Pooled association tests for rare variants in exon-

resequencing studies. *The American Journal of Human Genetics*, 86 (6), 832–838.

Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D., 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38 (8), 904–909.

Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R., and Willer, C.J., 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*.

Puig, M., Casillas, S., Villatoro, S., and Cáceres, M., 2015. Human inversions and their functional consequences. *Briefings in functional genomics*, 14 (5), 369–379.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., and Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics*, 81 (3), 559–575.

Qian, E.-N., Han, S.-Y., Ding, S.-Z., and Lv, X., 2016. Expression and diagnostic value of CCT3 and IQGAP3 in hepatocellular carcinoma. *Cancer cell international*, 16 (1), 55.

Qin, K., Zheng, Z., He, Y., Gao, Y., Shi, H., Mo, S., Zhang, J., and Rong, J., 2020. High expression of neutrophil cytosolic factor 2 (NCF2) is associated with aggressive features and poor prognosis of esophageal squamous cell carcinoma. *International journal of clinical and experimental pathology*, 13 (12), 3033–3043.

Quang, D., Chen, Y., and Xie, X., 2015. DANN: a deep learning approach for annotating the pathogenicity of genetic variants. *Bioinformatics (Oxford, England)*, 31 (5), 761–763.

Queisser-Luft, A., Stolz, G., Wiesel, A., Schlaefer, K., and Spranger, J., 2002. Malformations in newborn: results based on 30,940 infants and fetuses from the Mainz congenital birth defect monitoring system (1990-1998). *Archives of gynecology and obstetrics*, 266 (3), 163–167.

Quinlan, A.R. and Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* , 26 (6), 841–842.

Raffield, L.M., Iyengar, A.K., Wang, B., Gaynor, S.M., Spracklen, C.N., Zhong, X., Kowalski, M.H., Salimi, S., Polfus, L.M., Benjamin, E.J., Bis, J.C., Bowler, R., Cade, B.E., Choi, W.J., Comellas, A.P., Correa, A., Cruz, P., Doddapaneni, H., Durda, P., Gogarten, S.M., Jain, D., Kim, R.W., Kral, B.G., Lange, L.A., Larson, M.G., Laurie, C., Lee, J., Lee, S., Lewis, J.P., Metcalf, G.A., Mitchell, B.D., Momin, Z., Muzny, D.M., Pankratz, N., Park, C.J., Rich, S.S., Rotter, J.I., Ryan, K., Seo, D., Tracy, R.P., Viaud-Martinez, K.A., Yanek, L.R., Zhao, L.P., Lin, X., Li, B., Li, Y., Dupuis, J., Reiner, A.P., Mohlke, K.L., Auer, P.L., TOPMed Inflammation Working Group, and NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, 2020. Allelic heterogeneity at the CRP locus identified by whole-genome sequencing in multi-ancestry cohorts. *The American Journal of Human Genetics*, 106 (1), 112–120.

Rankin, S.A., Steimle, J.D., Yang, X.H., Rydeen, A.B., Agarwal, K., Chaturvedi, P., Ikegami, K., Herriges, M.J., Moskowitz, I.P., and Zorn, A.M., 2021. Tbx5 drives Aldh1a2 expression to regulate a RA-Hedgehog-Wnt gene regulatory network coordinating cardiopulmonary development. *eLife*, 10.

Rausch, T., Zichner, T., Schlattl, A., Stütz, A.M., Benes, V., and Korbel, J.O., 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics (Oxford, England)*, 28 (18), i333–i339.

Rebouissou, S., Vasiliu, V., Thomas, C., Bellanné-Chantelot, C., Bui, H., Chrétien, Y., Timsit, J., Rosty, C., Laurent-Puig, P., Chauveau, D., and Zucman-Rossi, J., 2005. Germline hepatocyte nuclear factor 1alpha and 1beta mutations in renal cell carcinomas. *Human molecular genetics*, 14 (5), 603–614.

Redin, C., Brand, H., Collins, R.L., Kammin, T., Mitchell, E., Hodge, J.C., Hanscom, C., Pillalamarri, V., Seabra, C.M., Abbott, M.-A., Abdul-Rahman, O.A., Aberg, E., Adley, R., Alcaraz-Estrada, S.L., Alkuraya, F.S., An, Y., Anderson, M.-A., Antolik, C., Anyane-Yeboa, K., Atkin, J.F., Bartell, T., Bernstein, J.A., Beyer, E., Blumenthal, I., Bongers, E.M.H.F., Brilstra, E.H., Brown, C.W., Brüggenwirth, H.T., Callewaert, B., Chiang, C., Corning, K., Cox, H., Cuppen, E., Currall, B.B., Cushing, T., David, D., Deardorff, M.A., Dheedene, A., D'Hooghe, M., de Vries, B.B.A., Earl, D.L., Ferguson, H.L., Fisher, H., FitzPatrick, D.R., Gerrol, P., Giachino, D., Glessner, J.T., Gliem, T., Grady, M., Graham, B.H., Griffis, C., Gripp, K.W., Gropman, A.L., Hanson-Kahn, A., Harris, D.J., Hayden, M.A., Hill, R., Hochstenbach, R., Hoffman, J.D., Hopkin, R.J., Hubshman, M.W., Innes, A.M., Irons, M., Irving, M., Jacobsen, J.C., Janssens, S., Jewett, T., Johnson, J.P., Jongmans, M.C., Kahler, S.G., Koolen, D.A., Korzelius, J., Kroisel, P.M., Lacassie, Y., Lawless, W., Lemyre, E., Leppig, K., Levin, A.V., Li, H., Li, H., Liao, E.C., Lim, C., Lose, E.J., Lucente, D., Macera, M.J., Manavalan, P., Mandrile, G., Marcelis, C.L., Margolin, L., Mason, T., Masser-Frye, D., McClellan, M.W., Mendoza, C.J.Z., Menten, B., Middelkamp, S., Mikami, L.R., Moe, E., Mohammed, S., Mononen, T., Mortenson, M.E., Moya, G., Nieuwint, A.W., Ordulu, Z., Parkash, S., Pauker, S.P., Pereira, S., Perrin, D., Phelan, K., Aguilar, R.E.P., Poddighe, P.J., Pregno, G., Raskin, S., Reis, L., Rhead, W., Rita, D., Renkens, I., Roelens, F., Ruliera, J., Rump, P., Schilit, S.L.P., Shaheen, R., Sparkes, R., Spiegel, E., Stevens, B., Stone, M.R., Tagoe, J., Thakuria, J.V., van Bon, B.W., van de Kamp, J., van Der Burgt, I., van Essen, T., van Ravenswaaij-Arts, C.M., van Roosmalen, M.J., Vergult, S., Volker-Touw, C.M.L., Warburton, D.P., Waterman, M.J., Wiley, S., Wilson, A., Yerena-de Vega, M. de la C.A., Zori, R.T., Levy, B., Brunner, H.G., de Leeuw, N., Kloosterman, W.P., Thorland, E.C., Morton, C.C., Gusella, J.F., and Talkowski, M.E., 2017. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nature genetics*, 49 (1), 36–45.

Reed, H.M. and Robinson, N.D., 1984. Horseshoe kidney with simultaneous occurrence of calculi, transitional cell and squamous cell carcinoma. *Urology*, 23 (1), 62–64.

Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I., Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Göttgens, B., Hacohen, N., Haniffa, M., Hemberg, M., Kim, S., Klenerman, P., Kriegstein, A., Lein, E., Linnarsson, S., Lundberg, E., Lundeberg, J., Majumder, P., Marioni, J.C., Merad, M., Mhlanga, M., Nawijn, M., Netea, M., Nolan, G., Pe'er, D., Phillipakis, A., Ponting, C.P., Quake, S., Reik, W., Rozenblatt-Rosen, O., Sanes, J., Satija, R., Schumacher, T.N., Shalek, A., Shapiro, E., Sharma, P., Shin, J.W., Stegle, O., Stratton, M., Stubbington, M.J.T., Theis, F.J., Uhlen, M., van Oudenaarden, A., Wagner, A., Watt, F., Weissman, J., Wold, B., Xavier, R., Yosef, N., and Human Cell Atlas Meeting Participants, 2017. The Human Cell Atlas. *eLife*, 6.

Reich, D.E. and Lander, E.S., 2001. On the allelic spectrum of human disease. *Trends in genetics: TIG*, 17 (9), 502–510.

Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M., 2019. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic acids research*, 47 (D1), D886–D894.

Retterer, K., Juusola, J., Cho, M.T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K.G., McKnight, D., Bai, R., Suchy, S., Friedman, B., Tahiliani, J., Pineda-Alvarez, D., Richard, G., Brandt, T., Haverfield, E., Chung, W.K., and Bale, S., 2016. Clinical application of whole-exome sequencing across clinical indications. *Genetics in medicine: official journal of the American College of Medical Genetics*, 18 (7), 696–704.

Reutter, H., Draaken, M., Pennimpede, T., Wittler, L., Brockschmidt, F.F., Ebert, A.-K., Bartels, E., Rösch, W., Boemers, T.M., Hirsch, K., Schmiedeke, E., Meesters, C., Becker, T., Stein, R., Utsch, B., Mangold, E., Nordenskjöld, A., Barker, G., Kockum, C.C., Zwink, N., Holmdahl, G., Läckgren, G., Jenetzky, E., Feitz, W.F.J., Marcelis, C., Wijers, C.H.W., Van Rooij, I.A.L.M., Gearhart, J.P., Herrmann, B.G., Ludwig, M., Boyadjiev, S.A., Nöthen, M.M., and Mattheisen, M., 2014. Genome-wide association study and mouse expression data identify a highly conserved 32 kb intergenic region between WNT3 and WNT9b as possible susceptibility locus for isolated classic exstrophy of the bladder. *Human molecular genetics*, 23 (20), 5536–5544.

Reutter, H., Qi, L., Gearhart, J.P., Boemers, T., Ebert, A.-K., Rösch, W., Ludwig, M., and Boyadjiev, S.A., 2007. Concordance analyses of twins with bladder exstrophy-epispadias complex suggest genetic etiology. *American journal of medical genetics. Part A*, 143A (22), 2751–2756.

Reutter, H., Shapiro, E., and Gruen, J.R., 2003. Seven new cases of familial isolated bladder exstrophy and epispadias complex (BEEC) and review of the literature. *American journal of medical genetics. Part A*, 120A (2), 215–221.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H.L., and ACMG Laboratory Quality Assurance Committee, 2015. 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. *Genetics in medicine: official journal of the American College of Medical Genetics*, 17 (5), 405–424.

Rioux, J.D., Xavier, R.J., Taylor, K.D., Silverberg, M.S., Goyette, P., Huett, A., Green, T., Kuballa, P., Barmada, M.M., Datta, L.W., Shugart, Y.Y., Griffiths, A.M., Targan, S.R., Ippoliti, A.F., Bernard, E.-J., Mei, L., Nicolae, D.L., Regueiro, M., Schumm, L.P., Steinhart, A.H., Rotter, J.I., Duerr, R.H., Cho, J.H., Daly, M.J., and Brant, S.R., 2007. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature genetics*, 39 (5), 596–604.

Ritz, A., Bashir, A., Sindi, S., Hsu, D., Hajirasouliha, I., and Raphael, B.J., 2014. Characterization of structural variants with single molecule and hybrid sequencing approaches. *Bioinformatics (Oxford, England)*, 30 (24), 3458–3466.

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., and Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. *Nature*, 518 (7539), 317–330.

Roller, E., Ivakhno, S., Lee, S., Royce, T., and Tanner, S., 2016. Canvas: versatile and scalable detection of copy number variants. *Bioinformatics* , 32 (15), 2375–2377.

Roodhooft, A.M., Birnholz, J.C., and Holmes, L.B., 1984. Familial Nature of Congenital Absence and Severe Dysgenesis of Both Kidneys. *New England Journal of Medicine*.

Rossanti, R., Morisada, N., Nozu, K., Kamei, K., Horinouchi, T., Yamamura, T., Minamikawa, S., Fujimura, J., Nagano, C., Sakakibara, N., Ninchoji, T., Kaito,

H., Ito, S., Tanaka, R., and Iijima, K., 2020. Clinical and genetic variability of PAX2-related disorder in the Japanese population. *Journal of human genetics*, 65 (6), 541–549.

Ruf, R.G., Xu, P.-X., Silvius, D., Otto, E.A., Beekmann, F., Muerb, U.T., Kumar, S., Neuhaus, T.J., Kemper, M.J., Raymond, R.M., Jr, Brophy, P.D., Berkman, J., Gattas, M., Hyland, V., Ruf, E.-M., Schwartz, C., Chang, E.H., Smith, R.J.H., Stratakis, C.A., Weil, D., Petit, C., and Hildebrandt, F., 2004. SIX1 mutations cause branchio-oto-renal syndrome by disruption of EYA1-SIX1-DNA complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (21), 8090–8095.

Saisawat, P., Kohl, S., Hilger, A.C., Hwang, D.-Y., Yung Gee, H., Dworschak, G.C., Tasic, V., Pennimpede, T., Natarajan, S., Sperry, E., Matassa, D.S., Stajić, N., Bogdanovic, R., de Blaauw, I., Marcelis, C.L.M., Wijers, C.H.W., Bartels, E., Schmiedeke, E., Schmidt, D., Märzheuser, S., Grasshoff-Derr, S., Holland-Cunz, S., Ludwig, M., Nöthen, M.M., Draaken, M., Brosens, E., Heij, H., Tibboel, D., Herrmann, B.G., Solomon, B.D., de Klein, A., van Rooij, I.A.L.M., Esposito, F., Reutter, H.M., and Hildebrandt, F., 2014. Whole-exome resequencing reveals recessive mutations in TRAP1 in individuals with CAKUT and VACTERL association. *Kidney international*, 85 (6), 1310–1317.

Salm, M.P.A., Horswell, S.D., Hutchison, C.E., Speedy, H.E., Yang, X., Liang, L., Schadt, E.E., Cookson, W.O., Wierzbicki, A.S., Naoumova, R.P., and Shoulders, C.C., 2012. The origin, global distribution, and functional impact of the human 8p23 inversion polymorphism. *Genome research*, 22 (6), 1144–1153.

Sampson, M.G., Coughlin, C.R., 2nd, Kaplan, P., Conlin, L.K., Meyers, K.E.C., Zackai, E.H., Spinner, N.B., and Copelovitch, L., 2010. Evidence for a recurrent microdeletion at chromosome 16p11.2 associated with congenital anomalies of the kidney and urinary tract (CAKUT) and Hirschsprung disease. *American journal of medical genetics. Part A*, 152A (10), 2618–2622.

Sanger, F., Nicklen, S., and Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74 (12), 5463–5467.

Sanna-Cherchi, S., Caridi, G., Weng, P.L., Dagnino, M., Seri, M., Konka, A., Somenzi, D., Carrea, A., Izzi, C., Casu, D., Allegri, L., Schmidt-Ott, K.M., Barasch, J., Scolari, F., Ravazzolo, R., Ghiggeri, G.M., and Gharavi, A.G., 2007. Localization of a Gene for Nonsyndromic Renal Hypodysplasia to Chromosome 1p32-33. *The American Journal of Human Genetics*.

Sanna-Cherchi, S., Khan, K., Westland, R., Krithivasan, P., Fievet, L., Rasouly, H.M., Ionita-Laza, I., Capone, V.P., Fasel, D.A., Kiryluk, K., Kamalakaran, S., Bodria, M., Otto, E.A., Sampson, M.G., Gillies, C.E., Vega-Warner, V., Vukojevic, K., Pediaditakis, I., Makar, G.S., Mitrotti, A., Verbitsky, M., Martino, J., Liu, Q., Na, Y.-J., Goj, V., Ardissino, G., Gigante, M., Gesualdo, L., Janezcko, M., Zaniew, M., Mendelsohn, C.L., Shril, S., Hildebrandt, F., van Wijk, J.A.E., Arapovic, A., Saraga, M., Allegri, L., Izzi, C., Scolari, F., Tasic,

V., Ghiggeri, G.M., Latos-Bielenska, A., Materna-Kiryluk, A., Mane, S., Goldstein, D.B., Lifton, R.P., Katsanis, N., Davis, E.E., and Gharavi, A.G., 2017. Exome-wide association study identifies GREB1L mutations in congenital kidney malformations. *The American Journal of Human Genetics*, 101 (6), 1034.

Sanna-Cherchi, S., Kiryluk, K., Burgess, K.E., Bodria, M., Sampson, M.G., Hadley, D., Nees, S.N., Verbitsky, M., Perry, B.J., Sterken, R., Lozanovski, V.J., Materna-Kiryluk, A., Barlassina, C., Kini, A., Corbani, V., Carrea, A., Somenzi, D., Murtas, C., Ristoska-Bojkovska, N., Izzi, C., Bianco, B., Zaniew, M., Flogelova, H., Weng, P.L., Kacak, N., Giberti, S., Gigante, M., Arapovic, A., Drnasin, K., Caridi, G., Curioni, S., Allegri, F., Ammenti, A., Ferretti, S., Goj, V., Bernardo, L., Jobanputra, V., Chung, W.K., Lifton, R.P., Sanders, S., State, M., Clark, L.N., Saraga, M., Padmanabhan, S., Dominiczak, A.F., Foroud, T., Gesualdo, L., Gucev, Z., Allegri, L., Latos-Bielenska, A., Cusi, D., Scolari, F., Tasic, V., Hakonarson, H., Ghiggeri, G.M., and Gharavi, A.G., 2012. Copy-number disorders are a common cause of congenital kidney malformations. *American journal of human genetics*, 91 (6), 987–997.

Sanna-Cherchi, S., Ravani, P., Corbani, V., Parodi, S., Haupt, R., Piaggio, G., Innocenti, M.L.D., Somenzi, D., Trivelli, A., Caridi, G., Izzi, C., Scolari, F., Mattioli, G., Allegri, L., and Ghiggeri, G.M., 2009. Renal outcome in patients with congenital anomalies of the kidney and urinary tract. *Kidney international*, 76 (5), 528–533.

Sanna-Cherchi, S., Reese, A., Hensle, T., Caridi, G., Izzi, C., Kim, Y.Y., Konka, A., Murer, L., Scolari, F., Ravazzolo, R., Ghiggeri, G.M., and Gharavi, A.G., 2005. Familial vesicoureteral reflux: testing replication of linkage in seven new multigenerational kindreds. *Journal of the American Society of Nephrology: JASN*, 16 (6), 1781–1787.

Sanna-Cherchi, S., Sampogna, R.V., Papeta, N., Burgess, K.E., Nees, S.N., Perry, B.J., Choi, M., Bodria, M., Liu, Y., Weng, P.L., Lozanovski, V.J., Verbitsky, M., Lugani, F., Sterken, R., Paragas, N., Caridi, G., Carrea, A., Dagnino, M., Materna-Kiryluk, A., Santamaria, G., Murtas, C., Ristoska-Bojkovska, N., Izzi, C., Kacak, N., Bianco, B., Giberti, S., Gigante, M., Piaggio, G., Gesualdo, L., Vukic, D.K., Vukojevic, K., Saraga-Babic, M., Saraga, M., Gucev, Z., Allegri, L., Latos-Bielenska, A., Casu, D., State, M., Scolari, F., Ravazzolo, R., Kiryluk, K., Al-Awqati, Q., D'Agati, V.D., Drummond, I.A., Tasic, V., Lifton, R.P., Ghiggeri, G.M., and Gharavi, A.G., 2013. Mutations in DSTYK and dominant urinary tract malformations. *The New England journal of medicine*, 369 (7), 621–629.

Sanyanusin, P., Schimmenti, L.A., McNoe, L.A., Ward, T.A., Pierpont, M.E., Sullivan, M.J., Dobyns, W.B., and Eccles, M.R., 1995. Mutation of the PAX2 gene in a family with optic nerve colobomas, renal anomalies and vesicoureteral reflux. *Nature genetics*, 9 (4), 358–364.

Saygili, S., Atayar, E., Canpolat, N., Elicevik, M., Kurugoglu, S., Sever, L., Caliskan, S., and Ozaltin, F., 2020. A homozygous HOXA11 variation as a potential

novel cause of autosomal recessive congenital anomalies of the kidney and urinary tract. *Clinical genetics*, 98 (4), 390–395.

Schedl, A., 2007. Renal abnormalities and their developmental origin. *Nature reviews. Genetics*, 8 (10), 791–802.

Schierbaum, L.M., Schneider, S., Herms, S., Sivalingam, S., Fabian, J., Reutter, H., Weber, S., Merz, W.M., Tkaczyk, M., Miklaszewska, M., Sikora, P., Szmigielska, A., Krzemien, G., Zachwieja, K., Szczepanska, M., Taranta-Janusz, K., Kroll, P., Polok, M., Zaniew, M., and Hilger, A.C., 2021. Genome-wide survey for microdeletions or -duplications in 155 patients with lower urinary tract obstructions (LUTO). *Genes*, 12 (9), 1449.

Schild, R., Knüppel, T., Konrad, M., Bergmann, C., Trautmann, A., Kemper, M.J., Wu, K., Yaklichkin, S., Wang, J., Pestell, R., Müller-Wiefel, D.E., Schaefer, F., and Weber, S., 2013. Double homozygous missense mutations in DACH1 and BMP4 in a patient with bilateral cystic renal dysplasia. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association*, 28 (1), 227–232.

Schmidt-Ott, K.M., 2017. How to grow a kidney: patient-specific kidney organoids come of age. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association*, 32 (1), 17–23.

Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L., and Ren, B., 2016. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell reports*, 17 (8), 2042–2059.

Schreuder, M.F., van der Horst, H.J.R., Bökenkamp, A., Beckers, G.M.A., and van Wijk, J.A.E., 2008. Posterior urethral valves in three siblings: a case report and review of the literature. *Birth defects research. Part A, Clinical and molecular teratology*, 82 (4), 232–235.

Schwaderer, A.L., Bates, C.M., McHugh, K.M., and McBride, K.L., 2007. Renal anomalies in family members of infants with bilateral renal agenesis/adysplasia. *Pediatric nephrology (Berlin, Germany)*, 22 (1), 52–56.

Sebat, J., Lakshmi, B., Malhotra, D., Troge, J., Lese-Martin, C., Walsh, T., Yamrom, B., Yoon, S., Krasnitz, A., Kendall, J., Leotta, A., Pai, D., Zhang, R., Lee, Y.-H., Hicks, J., Spence, S.J., Lee, A.T., Puura, K., Lehtimäki, T., Ledbetter, D., Gregersen, P.K., Bregman, J., Sutcliffe, J.S., Jobanputra, V., Chung, W., Warburton, D., King, M.-C., Skuse, D., Geschwind, D.H., Gilliam, T.C., Ye, K., and Wigler, M., 2007. Strong association of de novo copy number mutations with autism. *Science*, 316 (5823), 445–449.

Sekine, T., Namai, Y., Yanagisawa, A., Shirahama, H., Tashiro, Y., Terahara, M., Nagata, M., Harita, Y., Fukuoka, U., Inatomi, J., and Igarashi, T., 2005. A familial case of multicystic dysplastic kidney. *Pediatric nephrology (Berlin, Germany)*, 20 (9), 1245–1248.

Shaffer, J.R., LeClair, J., Carlson, J.C., Feingold, E., Buxó, C.J., Christensen, K., Deleyiannis, F.W.B., Field, L.L., Hecht, J.T., Moreno, L., Orioli, I.M., Padilla,

C., Vieira, A.R., Wehby, G.L., Murray, J.C., Weinberg, S.M., Marazita, M.L., and Leslie, E.J., 2019. Association of low-frequency genetic variants in regulatory regions with nonsyndromic orofacial clefts. *American journal of medical genetics. Part A*, 179 (3), 467–474.

Shalem, O., Sanjana, N.E., and Zhang, F., 2015. High-throughput functional genomics using CRISPR–Cas9. *Nature reviews. Genetics*, 16 (5), 299–311.

Shapiro, E., Lepor, H., and Jeffs, R.D., 1984. The inheritance of the exstrophy-epispadias complex. *The journal of urology*, 132 (2), 308–310.

Shi, H., Kichaev, G., and Pasaniuc, B., 2016. Contrasting the genetic architecture of 30 complex traits from summary association data. *The American Journal of Human Genetics*, 99 (1), 139–153.

Shriner, D., 2017. Overview of admixture mapping. *et al [Current protocols in human genetics]*, 94 (1), 1.23.1-1.23.8.

Siomou, E., Mitsioni, A.G., Giapros, V., Bouba, I., Noutsopoulos, D., and Georgiou, I., 2017. Copy-number variation analysis in familial nonsyndromic congenital anomalies of the kidney and urinary tract: Evidence for the causative role of a transposable element-associated genomic rearrangement. *Molecular medicine reports*, 15 (6), 3631–3636.

Sirota, L., Hertz, M., Laufer, J., Jonas, P., and Boichis, H., 1986. Familial vesicoureteral reflux: a study of 16 families. *Urologic radiology*, 8 (1), 22–24.

Sirugo, G., Williams, S.M., and Tishkoff, S.A., 2019. The missing diversity in human genetic studies. *Cell*, 177 (4), 1080.

Skinner, M.A., Safford, S.D., Reeves, J.G., Jackson, M.E., and Freemerman, A.J., 2008. Renal aplasia in humans is associated with RET mutations. *The American Journal of Human Genetics*, 82 (2), 344–351.

Slavotinek, A.M., Baranzini, S.E., Schanze, D., Labelle-Dumais, C., Short, K.M., Chao, R., Yahyavi, M., Bijlsma, E.K., Chu, C., Musone, S., Wheatley, A., Kwok, P.-Y., Marles, S., Fryns, J.-P., Maga, A.M., Hassan, M.G., Gould, D.B., Madireddy, L., Li, C., Cox, T.C., Smyth, I., Chudley, A.E., and Zenker, M., 2011. Manitoba-oculo-tricho-anal (MOTA) syndrome is caused by mutations in FREM1. *Journal of medical genetics*, 48 (6), 375–382.

Slickers, J.E., Olshan, A.F., Siega-Riz, A.M., Honein, M.A., Aylsworth, A.S., and National Birth Defects Prevention Study, 2008. Maternal body mass index and lifestyle exposures and the risk of bilateral renal agenesis or hypoplasia: the National Birth Defects Prevention Study. *American journal of epidemiology*, 168 (11), 1259–1267.

Smedley, D., Jacobsen, J.O.B., Jäger, M., Köhler, S., Holtgrewe, M., Schubach, M., Siragusa, E., Zemojtel, T., Buske, O.J., Washington, N.L., Bone, W.P., Haendel, M.A., and Robinson, P.N., 2015. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nature protocols*, 10 (12), 2004–2015.

Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C., and Nobrega, M.A., 2012. Regulatory variation in a TBX5 enhancer leads to

isolated congenital heart disease. *Human molecular genetics*, 21 (14), 3255–3263.

Smeulders, N. and Woodhouse, C.R., 2001. Neoplasia in adult exstrophy patients. *BJU international*, 87 (7), 623–628.

Snyder, M.W., Adey, A., Kitzman, J.O., and Shendure, J., 2015. Haplotype-resolved genome sequencing: experimental methods and applications. *Nature reviews. Genetics*, 16 (6), 344–358.

Spielmann, M., Lupiáñez, D.G., and Mundlos, S., 2018. Structural variation in the 3D genome. *Nature reviews. Genetics*, 19 (7), 453–467.

Srivastava, T., Garola, R.E., and Hellerstein, S., 1999. Autosomal dominant inheritance of multicystic dysplastic kidney. *Pediatric nephrology (Berlin, Germany)*, 13 (6), 481–483.

Stanescu, H.C., Arcos-Burgos, M., Medlar, A., Bockenhauer, D., Kottgen, A., Dragomirescu, L., Voinescu, C., Patel, N., Pearce, K., Hubank, M., Stephens, H.A.F., Laundy, V., Padmanabhan, S., Zawadzka, A., Hofstra, J.M., Coenen, M.J.H., den Heijer, M., Kiemeney, L.A.L.M., Bacq-Daian, D., Stengel, B., Powis, S.H., Brenchley, P., Feehally, J., Rees, A.J., Debiec, H., Wetzels, J.F.M., Ronco, P., Mathieson, P.W., and Kleta, R., 2011. Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy. *The New England journal of medicine*, 364 (7), 616–626.

Stark, Z., Foulger, R.E., Williams, E., Thompson, B.A., Patel, C., Lunke, S., Snow, C., Leong, I.U.S., Puzriakova, A., Daugherty, L.C., Leigh, S., Boustred, C., Niblock, O., Rueda-Martin, A., Gerasimenko, O., Savage, K., Bellamy, W., Lin, V.S.K., Valls, R., Gordon, L., Brittain, H.K., Thomas, E.R.A., Taylor Tavares, A.L., McEntagart, M., White, S.M., Tan, T.Y., Yeung, A., Downie, L., Macciocca, I., Savva, E., Lee, C., Roesley, A., De Fazio, P., Deller, J., Deans, Z.C., Hill, S.L., Caulfield, M.J., North, K.N., Scott, R.H., Rendon, A., Hofmann, O., and McDonagh, E.M., 2021. Scaling national and international improvement in virtual gene panel curation via a collaborative approach to discordance resolution. *The American Journal of Human Genetics*, 108 (9), 1551–1557.

Stavropoulos, D.J., Merico, D., Jobling, R., Bowdin, S., Monfared, N., Thiruvahindrapuram, B., Nalpathamkalam, T., Pellecchia, G., Yuen, R.K.C., Szego, M.J., Hayeems, R.Z., Shaul, R.Z., Brudno, M., Girdea, M., Frey, B., Alipanahi, B., Ahmed, S., Babul-Hirji, R., Porras, R.B., Carter, M.T., Chad, L., Chaudhry, A., Chitayat, D., Doust, S.J., Cytrynbaum, C., Dupuis, L., Ejaz, R., Fishman, L., Guerin, A., Hashemi, B., Helal, M., Hewson, S., Inbar-Feigenberg, M., Kannu, P., Karp, N., Kim, R., Kronick, J., Liston, E., MacDonald, H., Mercimek-Mahmutoglu, S., Mendoza-Londono, R., Nasr, E., Nimmo, G., Parkinson, N., Quercia, N., Raiman, J., Roifman, M., Schulze, A., Shugar, A., Shuman, C., Sinajon, P., Siriwardena, K., Weksberg, R., Yoon, G., Carew, C., Erickson, R., Leach, R.A., Klein, R., Ray, P.N., Meyn, M.S., Scherer, S.W., Cohn, R.D., and Marshall, C.R., 2016. Whole genome

sequencing expands diagnostic utility and improves clinical management in pediatric medicine. *npj genomic medicine*, 1 (1).

Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., Magnusdottir, B., Morgen, K., Arnarsdottir, S., Bjornsdottir, G., Walters, G.B., Jonsdottir, G.A., Doyle, O.M., Tost, H., Grimm, O., Kristjansdottir, S., Snorrason, H., Davidsdottir, S.R., Gudmundsson, L.J., Jonsson, G.F., Stefansdottir, B., Helgadottir, I., Haraldsson, M., Jonsdottir, B., Thygesen, J.H., Schwarz, A.J., Didriksen, M., Stensbøl, T.B., Brammer, M., Kapur, S., Halldorsson, J.G., Hreidarsson, S., Saemundsen, E., Sigurdsson, E., and Stefansson, K., 2014. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*, 505 (7483), 361–366.

Stefansson, H., Rujescu, D., Cichon, S., Pietiläinen, O.P.H., Ingason, A., Steinberg, S., Fossdal, R., Sigurdsson, E., Sigmundsson, T., Buizer-Voskamp, J.E., Hansen, T., Jakobsen, K.D., Muglia, P., Francks, C., Matthews, P.M., Gylfason, A., Halldorsson, B.V., Gudbjartsson, D., Thorgeirsson, T.E., Sigurdsson, A., Jonasdottir, A., Jonasdottir, A., Bjornsson, A., Mattiasdottir, S., Blondal, T., Haraldsson, M., Magnusdottir, B.B., Giegling, I., Möller, H.-J., Hartmann, A., Shianna, K.V., Ge, D., Need, A.C., Crombie, C., Fraser, G., Walker, N., Lonnqvist, J., Suvisaari, J., Tuulio-Henriksson, A., Paunio, T., Toulopoulou, T., Bramon, E., Di Forti, M., Murray, R., Ruggeri, M., Vassos, E., Tosato, S., Walshe, M., Li, T., Vasilescu, C., Mühleisen, T.W., Wang, A.G., Ullum, H., Djurovic, S., Melle, I., Olesen, J., Kiemeney, L.A., Franke, B., GROUP, Sabatti, C., Freimer, N.B., Gulcher, J.R., Thorsteinsdottir, U., Kong, A., Andreassen, O.A., Ophoff, R.A., Georgi, A., Rietschel, M., Werge, T., Petursson, H., Goldstein, D.B., Nöthen, M.M., Peltonen, L., Collier, D.A., St Clair, D., and Stefansson, K., 2008. Large recurrent microdeletions associated with schizophrenia. *Nature*, 455 (7210), 232–236.

Steimle, J.D., Rankin, S.A., Slagle, C.E., Bekeny, J., Rydeen, A.B., Chan, S.S.-K., Kweon, J., Yang, X.H., Ikegami, K., Nadadur, R.D., Rowton, M., Hoffmann, A.D., Lazarevic, S., Thomas, W., Boyle, E.A., Horb, M.E., Luna-Zurita, L., Ho, R.K., Kyba, M., Jensen, B., Zorn, A.M., Conlon, F.L., and Moskowitz, I.P., 2018. Evolutionarily conserved Tbx5–Wnt2/2b pathway orchestrates cardiopulmonary development. *Proceedings of the National Academy of Sciences of the United States of America*, 115 (45), E10615–E10624.

Stoll, C., Dott, B., Alembik, Y., and Roth, M.-P., 2014. Associated nonurinary congenital anomalies among infants with congenital anomalies of kidney and urinary tract (CAKUT). *European journal of medical genetics*, 57 (7), 322–328.

Stoltzfus, J.C., 2011. Logistic regression: a brief primer. *Academic emergency medicine: official journal of the Society for Academic Emergency Medicine*, 18 (10), 1099–1104.

Stolz, J.R., Foote, K.M., Veenstra-Knol, H.E., Pfundt, R., Ten Broeke, S.W., de Leeuw, N., Roht, L., Pajusalu, S., Part, R., Rebane, I., Õunap, K., Stark, Z., Kirk, E.P., Lawson, J.A., Lunke, S., Christodoulou, J., Louie, R.J., Rogers,

R.C., Davis, J.M., Innes, A.M., Wei, X.-C., Keren, B., Mignot, C., Lebel, R.R., Sperber, S.M., Sakonju, A., Dosa, N., Barge-Schaapveld, D.Q.C.M., Peeters-Scholte, C.M.P.C.D., Ruivenkamp, C.A.L., van Bon, B.W., Kennedy, J., Low, K.J., Ellard, S., Pang, L., Junewick, J.J., Mark, P.R., Carvill, G.L., and Swanson, G.T., 2021. Clustered mutations in the GRIK2 kainate receptor subunit gene underlie diverse neurodevelopmental disorders. *The American Journal of Human Genetics*, 108 (9), 1692–1709.

Stuart, H.M., Roberts, N.A., Burgu, B., Daly, S.B., Urquhart, J.E., Bhaskar, S., Dickerson, J.E., Mermerkaya, M., Silay, M.S., Lewis, M.A., Olondriz, M.B.O., Gener, B., Beetz, C., Varga, R.E., Gülpınar, O., Süer, E., Soygür, T., Ozçakar, Z.B., Yalçınkaya, F., Kavaz, A., Bulum, B., Gücük, A., Yue, W.W., Erdogan, F., Berry, A., Hanley, N.A., McKenzie, E.A., Hilton, E.N., Woolf, A.S., and Newman, W.G., 2013. LRIG2 mutations cause urofacial syndrome. *American journal of human genetics*, 92 (2), 259–264.

Su, Z., Gay, L.J., Strange, A., Palles, C., Band, G., Whiteman, D.C., Lescai, F., Langford, C., Nanji, M., Edkins, S., van der Winkel, A., Levine, D., Sasieni, P., Bellenguez, C., Howarth, K., Freeman, C., Trudgill, N., Tucker, A.T., Pirinen, M., Peppelenbosch, M.P., van der Laan, L.J.W., Kuipers, E.J., Drenth, J.P.H., Peters, W.H., Reynolds, J.V., Kelleher, D.P., McManus, R., Grabsch, H., Prenen, H., Bisschops, R., Krishnadath, K., Siersema, P.D., van Baal, J.W.P.M., Middleton, M., Petty, R., Gillies, R., Burch, N., Bhandari, P., Paterson, S., Edwards, C., Penman, I., Vaidya, K., Ang, Y., Murray, I., Patel, P., Ye, W., Mullins, P., Wu, A.H., Bird, N.C., Dallal, H., Shaheen, N.J., Murray, L.J., Koss, K., Bernstein, L., Romero, Y., Hardie, L.J., Zhang, R., Winter, H., Corley, D.A., Panter, S., Risch, H.A., Reid, B.J., Sargeant, I., Gammon, M.D., Smart, H., Dhar, A., McMurtry, H., Ali, H., Liu, G., Casson, A.G., Chow, W.-H., Rutter, M., Tawil, A., Morris, D., Nwokolo, C., Isaacs, P., Rodgers, C., Ragunath, K., MacDonald, C., Haigh, C., Monk, D., Davies, G., Wajed, S., Johnston, D., Gibbons, M., Cullen, S., Church, N., Langley, R., Griffin, M., Alderson, D., Deloukas, P., Hunt, S.E., Gray, E., Dronov, S., Potter, S.C., Tashakkori-Ghanbaria, A., Anderson, M., Brooks, C., Blackwell, J.M., Bramon, E., Brown, M.A., Casas, J.P., Corvin, A., Duncanson, A., Markus, H.S., Mathew, C.G., Palmer, C.N.A., Plomin, R., Rautanen, A., Sawcer, S.J., Trembath, R.C., Viswanathan, A.C., Wood, N., Trynka, G., Wijmenga, C., Cazier, J.-B., Atherfold, P., Nicholson, A.M., Gellatly, N.L., Glancy, D., Cooper, S.C., Cunningham, D., Lind, T., Hapeshi, J., Ferry, D., Rathbone, B., Brown, J., Love, S., Attwood, S., MacGregor, S., Watson, P., Sanders, S., Ek, W., Harrison, R.F., Moayyedi, P., de Caestecker, J., Barr, H., Stupka, E., Vaughan, T.L., Peltonen, L., Spencer, C.C.A., Tomlinson, I., Donnelly, P., Jankowski, J.A.Z., Esophageal Adenocarcinoma Genetics Consortium, and Wellcome Trust Case Control Consortium 2, 2012. Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. *Nature genetics*, 44 (10), 1131–1136.

Sudmant, P.H., Rausch, T., Gardner, E.J., Handsaker, R.E., Abyzov, A., Huddleston, J., Zhang, Y., Ye, K., Jun, G., Fritz, M.H.-Y., Konkel, M.K., Malhotra, A., Stütz, A.M., Shi, X., Casale, F.P., Chen, J., Hormozdiari, F., Dayama, G., Chen, K., Malig, M., Chaisson, M.J.P., Walter, K., Meiers, S., Kashin, S., Garrison, E., Auton, A., Lam, H.Y.K., Mu, X.J., Alkan, C., Antaki, D., Bae, T., Cerveira, E., Chines, P., Chong, Z., Clarke, L., Dal, E., Ding, L., Emery, S., Fan, X., Gujral, M., Kahveci, F., Kidd, J.M., Kong, Y., Lameijer, E.-W., McCarthy, S., Flicek, P., Gibbs, R.A., Marth, G., Mason, C.E., Menelaou, A., Muzny, D.M., Nelson, B.J., Noor, A., Parrish, N.F., Pendleton, M., Quitadamo, A., Raeder, B., Schadt, E.E., Romanovitch, M., Schlattl, A., Sebra, R., Shabalin, A.A., Untergasser, A., Walker, J.A., Wang, M., Yu, F., Zhang, C., Zhang, J., Zheng-Bradley, X., Zhou, W., Zichner, T., Sebat, J., Batzer, M.A., McCarroll, S.A., 1000 Genomes Project Consortium, Mills, R.E., Gerstein, M.B., Bashir, A., Stegle, O., Devine, S.E., Lee, C., Eichler, E.E., and Korbel, J.O., 2015. An integrated map of structural variation in 2,504 human genomes. *Nature*, 526 (7571), 75–81.

Sugnet, C.W., Furey, T.S., Roskin, K.M., and Pringle, T.H., 2002. The human genome browser at UCSC. *Research/ a journal of science and its applications*.

Sun, B.B., Kurki, M.I., Foley, C.N., Mechakra, A., Chen, C.-Y., Marshall, E., Wilk, J.B., Biogen Biobank Team, Chahine, M., Chevalier, P., Christé, G., FinnGen, Palotie, A., Daly, M.J., and Runz, H., 2022. Genetic associations of protein-coding variants in human disease. *Nature*.

Sun, B.B., Maranville, J.C., Peters, J.E., Stacey, D., Staley, J.R., Blackshaw, J., Burgess, S., Jiang, T., Paige, E., Surendran, P., Oliver-Williams, C., Kamat, M.A., Prins, B.P., Wilcox, S.K., Zimmerman, E.S., Chi, A., Bansal, N., Spain, S.L., Wood, A.M., Morrell, N.W., Bradley, J.R., Janjic, N., Roberts, D.J., Ouwehand, W.H., Todd, J.A., Soranzo, N., Suhre, K., Paul, D.S., Fox, C.S., Plenge, R.M., Danesh, J., Runz, H., and Butterworth, A.S., 2018. Genomic atlas of the human plasma proteome. *Nature*, 558 (7708), 73–79.

Sun, M., Tong, P., Kong, W., Dong, B., Huang, Y., Park, I.Y., Zhou, L., Liu, X.-D., Ding, Z., Zhang, X., Bai, S., German, P., Powell, R., Wang, Q., Tong, X., Tannir, N.M., Matin, S.F., Rathmell, W.K., Fuller, G.N., McCutcheon, I.E., Walker, C.L., Wang, J., and Jonasch, E., 2017. HNF1B loss exacerbates the development of chromophobe renal cell carcinomas. *Cancer research*, 77 (19), 5313–5326.

Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., and Spitz, F., 2016. The shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Developmental cell*, 39 (5), 529–543.

Tain, Y.-L., Luh, H., Lin, C.-Y., and Hsu, C.-N., 2016. Incidence and risks of congenital anomalies of kidney and urinary tract in newborns. *Medicine*, 95 (5), e2659.

Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Gagliano Taliun, S.A., Corvelo, A., Gogarten, S.M., Kang, H.M., Pitsillides, A.N., LeFaive, J., Lee, S.-B., Tian, X., Browning, B.L., Das, S., Emde, A.-K., Clarke, W.E., Loesch, D.P., Shetty, A.C., Blackwell, T.W., Wong, Q., Aguet, F., Albert, C., Alonso, A., Ardlie, K.G., Aslibekyan, S., Auer, P.L., Barnard, J., Graham Barr, R., Becker, L.C., Beer, R.L., Benjamin, E.J., Bielak, L.F., Blangero, J., Boehnke, M., Bowden, D.W., Brody, J.A., Burchard, E.G., Cade, B.E., Casella, J.F., Chalazan, B., Chen, Y.-D.I., Cho, M.H., Choi, S.H., Chung, M.K., Clish, C.B., Correa, A., Curran, J.E., Custer, B., Darbar, D., Daya, M., de Andrade, M., DeMeo, D.L., Dutcher, S.K., Ellinor, P.T., Emery, L.S., Fatkin, D., Forer, L., Fornage, M., Franceschini, N., Fuchsberger, C., Fullerton, S.M., Germer, S., Gladwin, M.T., Gottlieb, D.J., Guo, X., Hall, M.E., He, J., Heard-Costa, N.L., Heckbert, S.R., Irvin, M.R., Johnsen, J.M., Johnson, A.D., Kardia, S.L.R., Kelly, T., Kelly, S., Kenny, E.E., Kiel, D.P., Klemmer, R., Konkle, B.A., Kooperberg, C., Köttgen, A., Lange, L.A., Lasky-Su, J., Levy, D., Lin, X., Lin, K.-H., Liu, C., Loos, R.J.F., Garman, L., Gerszten, R., Lubitz, S.A., Lunetta, K.L., Mak, A.C.Y., Manichaikul, A., Manning, A.K., Mathias, R.A., McManus, D.D., McGarvey, S.T., Meigs, J.B., Meyers, D.A., Mikulla, J.L., Minear, M.A., Mitchell, B., Mohanty, S., Montasser, M.E., Montgomery, C., Morrison, A.C., Murabito, J.M., Natale, A., Natarajan, P., Nelson, S.C., North, K.E., O'Connell, J.R., Palmer, N.D., Pankratz, N., Peloso, G.M., Peyser, P.A., Post, W.S., Psaty, B.M., Rao, D.C., Redline, S., Reiner, A.P., Roden, D., Rotter, J.I., Ruczinski, I., Sarnowski, C., Schoenherr, S., Seo, J.-S., Seshadri, S., Sheehan, V.A., Benjamin Shoemaker, M., Smith, A.V., Smith, N.L., Smith, J.A., Sotoodehnia, N., Stilp, A.M., Tang, W., Taylor, K.D., Telen, M., Thornton, T.A., Tracy, R.P., Van Den Berg, D.J., Vasan, R.S., Viaud-Martinez, K.A., Vrieze, S., Weeks, D.E., Weir, B.S., Weiss, S.T., Weng, L.-C., Willer, C.J., Zhang, Y., Zhao, X., Arnett, D.K., Ashley-Koch, A.E., Barnes, K.C., Boerwinkle, E., Gabriel, S., Gibbs, R., Rice, K.M., Rich, S.S., Silverman, E., Qasba, P., Gan, W., Trans-Omics for Precision Medicine (TOPMed) Program, TOPMed Population Genetics Working Group, Papanicolaou, G.J., Nickerson, D.A., Browning, S.R., Zody, M.C., Zöllner, S., Wilson, J.G., Adrienne Cupples, L., Laurie, C.C., Jaquish, C.E., Hernandez, R.D., O'Connor, T.D., and Abecasis, G.R., 2019. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Cold Spring Harbor Laboratory*.

Tan, A., Abecasis, G.R., and Kang, H.M., 2015. Unified representation of genetic variants. *Bioinformatics* , 31 (13), 2202–2204.

Tansey, K.E., Rees, E., Linden, D.E., Ripke, S., Chambert, K.D., Moran, J.L., McCarroll, S.A., Holmans, P., Kirov, G., Walters, J., Owen, M.J., and O'Donovan, M.C., 2016. Common alleles contribute to schizophrenia in CNV carriers. *Molecular psychiatry*, 21 (8), 1085–1089.

Taylor, J.C., Martin, H.C., Lise, S., Broxholme, J., Cazier, J.-B., Rimmer, A., Kanapin, A., Lunter, G., Fiddy, S., Allan, C., Aricescu, A.R., Attar, M., Babbs, C., Becq, J., Beeson, D., Bento, C., Bignell, P., Blair, E., Buckle, V.J., Bull, K.,

Cais, O., Cario, H., Chapel, H., Copley, R.R., Cornall, R., Craft, J., Dahan, K., Davenport, E.E., Dendrou, C., Devuyst, O., Fenwick, A.L., Flint, J., Fugger, L., Gilbert, R.D., Goriely, A., Green, A., Greger, I.H., Grocock, R., Gruszczyk, A.V., Hastings, R., Hatton, E., Higgs, D., Hill, A., Holmes, C., Howard, M., Hughes, L., Humburg, P., Johnson, D., Karpe, F., Kingsbury, Z., Kini, U., Knight, J.C., Krohn, J., Lamble, S., Langman, C., Lonie, L., Luck, J., McCarthy, D., McGowan, S.J., McMullin, M.F., Miller, K.A., Murray, L., Németh, A.H., Nesbit, M.A., Nutt, D., Ormondroyd, E., Oturai, A.B., Pagnamenta, A., Patel, S.Y., Percy, M., Petousi, N., Piazza, P., Piret, S.E., Polanco-Echeverry, G., Popitsch, N., Powrie, F., Pugh, C., Quek, L., Robbins, P.A., Robson, K., Russo, A., Sahgal, N., van Schouwenburg, P.A., Schuh, A., Silverman, E., Simmons, A., Sørensen, P.S., Sweeney, E., Taylor, J., Thakker, R.V., Tomlinson, I., Trebes, A., Twigg, S.R.F., Uhlig, H.H., Vyas, P., Vyse, T., Wall, S.A., Watkins, H., Whyte, M.P., Witty, L., Wright, B., Yau, C., Buck, D., Humphray, S., Ratcliffe, P.J., Bell, J.I., Wilkie, A.O.M., Bentley, D., Donnelly, P., and McVean, G., 2015. Factors influencing success of clinical genome sequencing across a broad spectrum of disorders. *Nature genetics*, 47 (7), 717–726.

Thakkar, D., Deshpande, A.V., and Kennedy, S.E., 2014. Epidemiology and demography of recently diagnosed cases of posterior urethral valves. *Pediatric research*, 76 (6), 560–563.

The International HapMap 3 Consortium, 2010. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467 (7311), 52–58.

The UK10K Consortium, Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R.B., Xu, C., Futema, M., Lawson, D., Iotchkova, V., Schiffels, S., Hendricks, A.E., Danecek, P., Li, R., Floyd, J., Wain, L.V., Barroso, I., Humphries, S.E., Hurles, M.E., Zeggini, E., Barrett, J.C., Plagnol, V., Brent Richards, J., Greenwood, C.M.T., Timpson, N.J., Durbin, R., Soranzo, N., Bala, S., Clapham, P., Coates, G., Cox, T., Daly, A., Danecek, P., Du, Y., Durbin, R., Edkins, S., Ellis, P., Flicek, P., Guo, X., Guo, X., Huang, L., Jackson, D.K., Joyce, C., Keane, T., Kolb-Kokocinski, A., Langford, C., Li, Y., Liang, J., Lin, H., Liu, R., Maslen, J., McCarthy, S., (co-chair) Muddyman, D., Quail, M.A., Stalker, J., (co-chair), Sun, J., Tian, J., Wang, G., Wang, J., Wang, Y., Wong, K., Zhang, P., Barroso, I., Birney, E., Boustred, C., Chen, L., Clement, G., Cocca, M., Danecek, P., Davey Smith, G., Day, I.N.M., Day-Williams, A., Down, T., Dunham, I., Durbin, R., Evans, D.M., Gaunt, T.R., Geihs, M., Greenwood, C.M.T., Hart, D., Hendricks, A.E., Howie, B., Huang, J., Hubbard, T., Hysi, P., Iotchkova, V., Jamshidi, Y., Karczewski, K.J., Kemp, J.P., Lachance, G., Lawson, D., Lek, M., Lopes, M., MacArthur, D.G., Marchini, J., Mangino, M., Mathieson, I., McCarthy, S., Memari, Y., Metrustry, S., Min, J.L., Moayyeri, A., Muddyman, D., Northstone, K., Panoutsopoulou, K., Paternoster, L., Perry, J.R.B., Quaye, L., Brent Richards, J., (co-chair), Ring, S., Ritchie, G.R.S., Schiffels, S., Shihab, H.A., Shin, S.-Y., Small, K.S., Soler Artigas, M., Soranzo, N., (co-chair), Southam, L., Spector, T.D., St

Pourcain, B., Surdulescu, G., Tachmazidou, I., Timpson, N.J., (co-chair), Tobin, M.D., Valdes, A.M., Visscher, P.M., Wain, L.V., Walter, K., Ward, K., Wilson, S.G., Wong, K., Yang, J., Zeggini, E., Zhang, F., Zheng, H.-F., Anney, R., Ayub, M., Barrett, J.C., Blackwood, D., Bolton, P.F., Breen, G., Collier, D.A., Craddock, N., Crooks, L., Curran, S., Curtis, D., Durbin, R., Gallagher, L., Geschwind, D., Gurling, H., Holmans, P., Lee, I., Lönnqvist, J., McCarthy, S., McGuffin, P., McIntosh, A.M., McKechanie, A.G., McQuillin, A., Morris, J., Muddyman, D., O'Donovan, M.C., Owen, M.J., (co-chair), Palotie, A., (co-chair), Parr, J.R., Paunio, T., Pietilainen, O., Rehnström, K., Sharp, S.I., Skuse, D., St Clair, D., Suvisaari, J., Walters, J.T.R., Williams, H.J., Barroso, I., (co-chair), Bochukova, E., Bounds, R., Dominiczak, A., Durbin, R., Farooqi, I.S., (co-chair), Hendricks, A.E., Keogh, J., Marenne, G., McCarthy, S., Morris, A., Muddyman, D., O'Rahilly, S., Porteous, D.J., Smith, B.H., Tachmazidou, I., Wheeler, E., Zeggini, E., Al Turki, S., Anderson, C.A., Antony, D., Barroso, I., Beales, P., Bentham, J., Bhattacharya, S., Calissano, M., Carss, K., Chatterjee, K., Cirak, S., Cosgrove, C., Durbin, R., Fitzpatrick, D.R., (co-chair), Floyd, J., Reghan Foley, A., Franklin, C.S., Futema, M., Grozeva, D., Humphries, S.E., Hurles, M.E., (co-chair), McCarthy, S., Mitchison, H.M., Muddyman, D., Muntoni, F., O'Rahilly, S., Onoufriadis, A., Parker, V., Payne, F., Plagnol, V., Lucy Raymond, F., Roberts, N., Savage, D.B., Scambler, P., Schmidts, M., Schoenmakers, N., Semple, R.K., Serra, E., Spasic-Boskovic, O., Stevens, E., van Kogelenberg, M., Vijayarangakannan, P., Walter, K., Williamson, K.A., Wilson, C., Whyte, T., Ciampi, A., Greenwood, C.M.T., (co-chair), Hendricks, A.E., Li, R., Metrustry, S., Oualkacha, K., Tachmazidou, I., Xu, C., Zeggini, E., (co-chair), Bobrow, M., Bolton, P.F., Durbin, R., Fitzpatrick, D.R., Griffin, H., Hurles, M.E., (co-chair), Kaye, J., (co-chair), Kennedy, K., Kent, A., Muddyman, D., Muntoni, F., Lucy Raymond, F., Semple, R.K., Smee, C., Spector, T.D., Timpson, N.J., Charlton, R., Ekong, R., Futema, M., Humphries, S.E., Khawaja, F., Lopes, L.R., Migone, N., Payne, S.J., Plagnol, V., (chair), Pollitt, R.C., Povey, S., Ridout, C.K., Robinson, R.L., Scott, R.H., Shaw, A., Syrris, P., Taylor, R., Vandersteen, A.M., Barrett, J.C., Barroso, I., Davey Smith, G., Durbin, R., (chair), Farooqi, I.S., Fitzpatrick, D.R., Hurles, M.E., Kaye, J., Kennedy, K., Langford, C., McCarthy, S., Muddyman, D., Owen, M.J., Palotie, A., Brent Richards, J., Soranzo, N., Spector, T.D., Stalker, J., Timpson, N.J., Zeggini, E., Amuzu, A., Pablo Casas, J., Chambers, J.C., Cocca, M., Dedoussis, G., Gambaro, G., Gasparini, P., Gaunt, T.R., Huang, J., Iotchkova, V., Isaacs, A., Johnson, J., Kleber, M.E., Kooner, J.S., Langenberg, C., Luan, J., Malerba, G., März, W., Matchan, A., Min, J.L., Morris, R., Nordestgaard, B.G., Benn, M., Ring, S., Scott, R.A., Soranzo, N., Southam, L., Timpson, N.J., Toniolo, D., Traglia, M., Tybjaerg-Hansen, A., van Duijn, C.M., van Leeuwen, E.M., Varbo, A., Whincup, P., Zaza, G., Zeggini, E., Zhang, W., Writing group, Production group, Cohorts group, Neurodevelopmental disorders group, Obesity group, Rare disease group, Statistics group, Ethics group, Incidental

330

findings group, Management committee, Lipid meta-analysis group, and The UCLEB Consortium, 2015. The UK10K project identifies rare variants in health and disease. *Nature*, 526 (7571), 82–90.

The Wellcome Trust Case Control Consortium, 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447 (7145), 661–678.

Thomas, R., Sanna-Cherchi, S., Warady, B.A., Furth, S.L., Kaskel, F.J., and Gharavi, A.G., 2011. HNF1B and PAX2 mutations are a common cause of renal hypodysplasia in the CKiD cohort. *Pediatric nephrology (Berlin, Germany)*, 26 (6), 897–903.

Timpson, N.J., Greenwood, C.M.T., Soranzo, N., Lawson, D.J., and Richards, J.B., 2018. Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature reviews. Genetics*, 19 (2), 110–124.

Tong, C.C., Duffy, K.A., Chu, D.I., Weiss, D.A., Srinivasan, A.K., Canning, D.A., and Kalish, J.M., 2017. Urological Findings in Beckwith-Wiedemann Syndrome With Chromosomal Duplications of 11p15.5: Evaluation and Management. *Urology*, 100, 224–227.

Torban, E. and Sokol, S.Y., 2021. Planar cell polarity pathway in kidney development, function and disease. *Nature reviews. Nephrology*, 17 (6), 369–385.

Tsai, C.-W., North, K.E., Tin, A., Haack, K., Franceschini, N., Saroja Voruganti, V., Laston, S., Zhang, Y., Best, L.G., MacCluer, J.W., Beaty, T.H., Navas-Acien, A., Kao, W.H.L., and Howard, B.V., 2015. Both rare and common variants in PCSK9 influence plasma low-density lipoprotein cholesterol level in American Indians. *The journal of clinical endocrinology and metabolism*, 100 (2), E345-9.

Turner, S.D., 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Cold Spring Harbor Laboratory*.

Turro, E., Astle, W.J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H.L., Sanchis-Juan, A., Frontini, M., Thys, C., Stephens, J., Mapeta, R., Burren, O.S., Downes, K., Haimel, M., Tuna, S., Deevi, S.V.V., Aitman, T.J., Bennett, D.L., Calleja, P., Carss, K., Caulfield, M.J., Chinnery, P.F., Dixon, P.H., Gale, D.P., James, R., Koziell, A., Laffan, M.A., Levine, A.P., Maher, E.R., Markus, H.S., Morales, J., Morrell, N.W., Mumford, A.D., Ormondroyd, E., Rankin, S., Rendon, A., Richardson, S., Roberts, I., Roy, N.B.A., Saleem, M.A., Smith, K.G.C., Stark, H., Tan, R.Y.Y., Themistocleous, A.C., Thrasher, A.J., Watkins, H., Webster, A.R., Wilkins, M.R., Williamson, C., Whitworth, J., Humphray, S., Bentley, D.R., NIHR BioResource for the 100,000 Genomes Project, Kingston, N., Walker, N., Bradley, J.R., Ashford, S., Penkett, C.J., Freson, K., Stirrups, K.E., Raymond, F.L., and Ouwehand, W.H., 2020. Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, 583 (7814), 96–102.

Tuzovic, L., Tang, S., Miller, R.S., Rohena, L., Shahmirzadi, L., Gonzalez, K., Li, X., LeDuc, C.A., Guo, J., Wilson, A., Mills, A., Glassberg, K., Rotterdam, H.,

Sepulveda, A.R., Zeng, W., Chung, W.K., and Anyane-Yeboa, K., 2015. New insights into the genetics of fetal megacystis: ACTG2 mutations, encoding γ-2 smooth muscle actin in megacystis microcolon intestinal hypoperistalsis syndrome (berdon syndrome). *Fetal diagnosis and therapy*, 38 (4), 296–306.

Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., and Posthuma, D., 2021. Genome-wide association studies. *Nature Reviews Methods Primers*, 1 (1).

Unger, S., Böhm, D., Kaiser, F.J., Kaulfuss, S., Borozdin, W., Buiting, K., Burfeind, P., Böhm, J., Barrionuevo, F., Craig, A., Borowski, K., Keppler-Noreuil, K., Schmitt-Mechelke, T., Steiner, B., Bartholdi, D., Lemke, J., Mortier, G., Sandford, R., Zabel, B., Superti-Furga, A., and Kohlhase, J., 2008. Mutations in the cyclin family member FAM58A cause an X-linked dominant disorder characterized by syndactyly, telecanthus and anogenital and renal malformations. *Nature genetics*, 40 (3), 287–289.

Van den Abbeele, A.D., Treves, S.T., Lebowitz, R.L., Bauer, S., Davis, R.T., Retik, A., and Colodny, A., 1987. Vesicoureteral reflux in asymptomatic siblings of patients with known reflux: Radionuclide cystography. *The journal of urology*, 138 (2), 464–464.

Van Esch, H., Groenen, P., Nesbit, M.A., Schuffenhauer, S., Lichtner, P., Vanderlinden, G., Harding, B., Beetz, R., Bilous, R.W., Holdaway, I., Shaw, N.J., Fryns, J.P., Van de Ven, W., Thakker, R.V., and Devriendt, K., 2000. GATA3 haplo-insufficiency causes human HDR syndrome. *Nature*, 406 (6794), 419–422.

van der Ven, A.T., Connaughton, D.M., Ityel, H., Mann, N., Nakayama, M., Chen, J., Vivante, A., Hwang, D.-Y., Schulz, J., Braun, D.A., Schmidt, J.M., Schapiro, D., Schneider, R., Warejko, J.K., Daga, A., Majmundar, A.J., Tan, W., Jobst-Schwan, T., Hermle, T., Widmeier, E., Ashraf, S., Amar, A., Hoogstraaten, C.A., Hugo, H., Kitzler, T.M., Kause, F., Kolvenbach, C.M., Dai, R., Spaneas, L., Amann, K., Stein, D.R., Baum, M.A., Somers, M.J.G., Rodig, N.M., Ferguson, M.A., Traum, A.Z., Daouk, G.H., Bogdanović, R., Stajić, N., Soliman, N.A., Kari, J.A., El Desoky, S., Fathy, H.M., Milosevic, D., Al-Saffar, M., Awad, H.S., Eid, L.A., Selvin, A., Senguttuvan, P., Sanna-Cherchi, S., Rehm, H.L., MacArthur, D.G., Lek, M., Laricchia, K.M., Wilson, M.W., Mane, S.M., Lifton, R.P., Lee, R.S., Bauer, S.B., Lu, W., Reutter, H.M., Tasic, V., Shril, S., and Hildebrandt, F., 2018. Whole-Exome Sequencing Identifies Causative Mutations in Families with Congenital Anomalies of the Kidney and Urinary Tract. *Journal of the American Society of Nephrology: JASN*, 29 (9), 2348–2361.

van der Ven, A.T., Kobbe, B., Kohl, S., Shril, S., Pogoda, H.-M., Imhof, T., Ityel, H., Vivante, A., Chen, J., Hwang, D.-Y., Connaughton, D.M., Mann, N., Widmeier, E., Taglienti, M., Schmidt, J.M., Nakayama, M., Senguttuvan, P., Kumar, S., Tasic, V., Kehinde, E.O., Mane, S.M., Lifton, R.P., Soliman, N., Lu, W., Bauer, S.B., Hammerschmidt, M., Wagener, R., and Hildebrandt, F., 2018. A

homozygous missense variant in VWA2, encoding an interactor of the Fraser-complex, in a patient with vesicoureteral reflux. *PloS one*, 13 (1), e0191224.

van der Ven, A.T., Shril, S., Ityel, H., Vivante, A., Chen, J., Hwang, D.-Y., Laricchia, K.M., Lek, M., Tasic, V., and Hildebrandt, F., 2017. Whole-Exome Sequencing Reveals FAT4 Mutations in a Clinically Unrecognizable Patient with Syndromic CAKUT: A Case Report. *Molecular syndromology*, 8 (5), 272–277.

van der Ven, A.T., Vivante, A., and Hildebrandt, F., 2018. Novel Insights into the Pathogenesis of Monogenic Congenital Anomalies of the Kidney and Urinary Tract. *Journal of the American Society of Nephrology: JASN*, 29 (1), 36–50.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Gabor Miklos, G.L., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J.,

Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.H., Coyne, M., Dahlke, C., Mays, A., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X., 2001. The sequence of the human genome. *Science*, 291 (5507), 1304–1351.

Verbitsky, M., Krithivasan, P., Batourina, E., Khan, A., Graham, S.E., Marasà, M., Kim, H., Lim, T.Y., Weng, P.L., Sánchez-Rodríguez, E., Mitrotti, A., Ahram, D.F., Zanoni, F., Fasel, D.A., Westland, R., Sampson, M.G., Zhang, J.Y., Bodria, M., Kil, B.H., Shril, S., Gesualdo, L., Torri, F., Scolari, F., Izzi, C., van Wijk, J.A.E., Saraga, M., Santoro, D., Conti, G., Barton, D.E., Dobson, M.G., Puri, P., Furth, S.L., Warady, B.A., Pisani, I., Fiaccadori, E., Allegri, L., Degl'Innocenti, M.L., Piaggio, G., Alam, S., Gigante, M., Zaza, G., Esposito, P., Lin, F., Simões-E-Silva, A.C., Brodkiewicz, A., Drozdz, D., Zachwieja, K., Miklaszewska, M., Szczepanska, M., Adamczyk, P., Tkaczyk, M., Tomczyk, D., Sikora, P., Mizerska-Wasiak, M., Krzemien, G., Szmigielska, A., Zaniew, M., Lozanovski, V.J., Gucev, Z., Ionita-Laza, I., Stanaway, I.B., Crosslin, D.R., Wong, C.S., Hildebrandt, F., Barasch, J., Kenny, E.E., Loos, R.J.F., Levy, B., Ghiggeri, G.M., Hakonarson, H., Latos-Bieleńska, A., Materna-Kiryluk, A., Darlow, J.M., Tasic, V., Willer, C., Kiryluk, K., Sanna-Cherchi, S., Mendelsohn, C.L., and Gharavi, A.G., 2021. Copy number variant analysis and genome-wide association study identify loci with large effect for vesicoureteral reflux. *Journal of the American Society of Nephrology: JASN*, 32 (4), 805–820.

Verbitsky, M., Westland, R., Perez, A., Kiryluk, K., Liu, Q., Krithivasan, P., Mitrotti, A., Fasel, D.A., Batourina, E., Sampson, M.G., Bodria, M., Werth, M., Kao, C., Martino, J., Capone, V.P., Vivante, A., Shril, S., Kil, B.H., Marasà, M., Zhang, J.Y., Na, Y.-J., Lim, T.Y., Ahram, D., Weng, P.L., Heinzen, E.L., Carrea, A., Piaggio, G., Gesualdo, L., Manca, V., Masnata, G., Gigante, M., Cusi, D., Izzi, C., Scolari, F., van Wijk, J.A.E., Saraga, M., Santoro, D., Conti, G., Zamboli, P., White, H., Drozdz, D., Zachwieja, K., Miklaszewska, M., Tkaczyk, M., Tomczyk, D., Krakowska, A., Sikora, P., Jarmoliński, T., Borszewska-Kornacka, M.K., Pawluch, R., Szczepanska, M., Adamczyk, P., Mizerska-Wasiak, M., Krzemien, G., Szmigielska, A., Zaniew, M., Dobson, M.G., Darlow, J.M., Puri, P., Barton, D.E., Furth, S.L., Warady, B.A., Gucev, Z., Lozanovski, V.J., Tasic, V., Pisani, I., Allegri, L., Rodas, L.M., Campistol, J.M., Jeanpierre, C., Alam, S., Casale, P., Wong, C.S., Lin, F., Miranda, D.M., Oliveira, E.A., Simões-E-Silva, A.C., Barasch, J.M., Levy, B., Wu, N., Hildebrandt, F., Ghiggeri, G.M., Latos-Bielenska, A., Materna-Kiryluk, A.,

Zhang, F., Hakonarson, H., Papaioannou, V.E., Mendelsohn, C.L., Gharavi, A.G., and Sanna-Cherchi, S., 2019. The copy number variation landscape of congenital anomalies of the kidney and urinary tract. *Nature genetics*, 51 (1), 117–127.

Verzi, M.P., Anderson, J.P., Dodou, E., Kelly, K.K., Greene, S.B., North, B.J., Cripps, R.M., and Black, B.L., 2002. N-twist, an evolutionarily conserved bHLH protein expressed in the developing CNS, functions as a transcriptional inhibitor. *Developmental biology*, 249 (1), 174–190.

Visscher, P.M., Hill, W.G., and Wray, N.R., 2008. Heritability in the genomics era--concepts and misconceptions. *Nature reviews. Genetics*, 9 (4), 255–266.

Visscher, P.M., Medland, S.E., Ferreira, M.A.R., Morley, K.I., Zhu, G., Cornes, B.K., Montgomery, G.W., and Martin, N.G., 2006. Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS genetics*, 2 (3), e41.

Vissers, L.E.L.M., van Ravenswaaij, C.M.A., Admiraal, R., Hurst, J.A., de Vries, B.B.A., Janssen, I.M., van der Vliet, W.A., Huys, E.H.L.P.G., de Jong, P.J., Hamel, B.C.J., Schoenmakers, E.F.P.M., Brunner, H.G., Veltman, J.A., and van Kessel, A.G., 2004. Mutations in a new member of the chromodomain gene family cause CHARGE syndrome. *Nature genetics*, 36 (9), 955–957.

Viswanathan, A., Dawman, L., Tiewsoh, K., Saxena, A.K., Dutta, S., and Suri, D., 2021. Screening of renal anomalies in first-degree relatives of children diagnosed with non-syndromic congenital anomalies of kidney and urinary tract. *Clinical and experimental nephrology*, 25 (2), 184–190.

Vivante, A., Hwang, D.-Y., Kohl, S., Chen, J., Shril, S., Schulz, J., van der Ven, A., Daouk, G., Soliman, N.A., Kumar, A.S., Senguttuvan, P., Kehinde, E.O., Tasic, V., and Hildebrandt, F., 2017. Exome sequencing discerns syndromes in patients from consanguineous families with congenital anomalies of the kidneys and urinary tract. *Journal of the American Society of Nephrology: JASN*, 28 (1), 69–75.

Vivante, A., Kleppa, M.-J., Schulz, J., Kohl, S., Sharma, A., Chen, J., Shril, S., Hwang, D.-Y., Weiss, A.-C., Kaminski, M.M., Shukrun, R., Kemper, M.J., Lehnhardt, A., Beetz, R., Sanna-Cherchi, S., Verbitsky, M., Gharavi, A.G., Stuart, H.M., Feather, S.A., Goodship, J.A., Goodship, T.H.J., Woolf, A.S., Westra, S.J., Doody, D.P., Bauer, S.B., Lee, R.S., Adam, R.M., Lu, W., Reutter, H.M., Kehinde, E.O., Mancini, E.J., Lifton, R.P., Tasic, V., Lienkamp, S.S., Jüppner, H., Kispert, A., and Hildebrandt, F., 2015. Mutations in TBX18 cause dominant urinary tract malformations via transcriptional dysregulation of ureter development. *The American Journal of Human Genetics*, 97 (2), 291–301.

Vivante, A., Mann, N., Yonath, H., Weiss, A.-C., Getwan, M., Kaminski, M.M., Bohnenpoll, T., Teyssier, C., Chen, J., Shril, S., van der Ven, A.T., Ityel, H., Schmidt, J.M., Widmeier, E., Bauer, S.B., Sanna-Cherchi, S., Gharavi, A.G., Lu, W., Magen, D., Shukrun, R., Lifton, R.P., Tasic, V., Stanescu, H.C., Cavaillès, V., Kleta, R., Anikster, Y., Dekel, B., Kispert, A., Lienkamp, S.S.,

and Hildebrandt, F., 2017. A dominant mutation in nuclear receptor interacting protein 1 causes urinary tract malformations via dysregulation of retinoic acid signaling. *Journal of the American Society of Nephrology: JASN*, 28 (8), 2364–2376.

Vogel, M.J., van Zon, P., Brueton, L., Gijzen, M., van Tuil, M.C., Cox, P., Schanze, D., Kariminejad, A., Ghaderi-Sohi, S., Blair, E., Zenker, M., Scambler, P.J., Ploos van Amstel, H.K., and van Haelst, M.M., 2012. Mutations in GRIP1 cause Fraser syndrome. *Journal of medical genetics*, 49 (5), 303–306.

Wade, E.M., Jenkins, Z.A., Morgan, T., Gimenez, G., Gibson, H., Peng, H., Sanchez Russo, R., Skraban, C.M., Bedoukian, E., and Robertson, S.P., 2021. Exon skip-inducing variants in FLNA in an attenuated form of frontometaphyseal dysplasia. *American journal of medical genetics. Part A*, 185 (12), 3675–3682.

Wainschtein, P., Jain, D., Zheng, Z., TOPMed Anthropometry Working Group, NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Cupples, L.A., Shadyab, A.H., McKnight, B., Shoemaker, B.M., Mitchell, B.D., Psaty, B.M., Kooperberg, C., Liu, C.-T., Albert, C.M., Roden, D., Chasman, D.I., Darbar, D., Lloyd-Jones, D.M., Arnett, D.K., Regan, E.A., Boerwinkle, E., Rotter, J.I., O'Connell, J.R., Yanek, L.R., de Andrade, M., Allison, M.A., McDonald, M.-L.N., Chung, M.K., Fornage, M., Chami, N., Smith, N.L., Ellinor, P.T., Vasan, R.S., Mathias, R.A., Loos, R.J.F., Rich, S.S., Lubitz, S.A., Heckbert, S.R., Redline, S., Guo, X., Chen, Y.-D.I., Laurie, C.A., Hernandez, R.D., McGarvey, S.T., Goddard, M.E., Laurie, C.C., North, K.E., Lange, L.A., Weir, B.S., Yengo, L., Yang, J., and Visscher, P.M., 2022. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nature genetics*.

Wakefield, J., 2009. Bayes factors for genome-wide association studies: comparison with P-values. *Genetic epidemiology*, 33 (1), 79–86.

Walker, K.A. and Bertram, J.F., 2011. Kidney development: core curriculum 2011. *American journal of kidney diseases: the official journal of the National Kidney Foundation*, 57 (6), 948–958.

Wang, K., Li, M., and Hakonarson, H., 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research*, 38 (16), e164.

Wang, K., Zhang, H., Kugathasan, S., Annese, V., Bradfield, J.P., Russell, R.K., Sleiman, P.M.A., Imielinski, M., Glessner, J., Hou, C., Wilson, D.C., Walters, T., Kim, C., Frackelton, E.C., Lionetti, P., Barabino, A., Van Limbergen, J., Guthery, S., Denson, L., Piccoli, D., Li, M., Dubinsky, M., Silverberg, M., Griffiths, A., Grant, S.F.A., Satsangi, J., Baldassano, R., and Hakonarson, H., 2009. Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *The American Journal of Human Genetics*, 84 (3), 399–405.

Wang, M., De Marco, P., Merello, E., Drapeau, P., Capra, V., and Kibar, Z., 2015. Role of the planar cell polarity gene Protein tyrosine kinase 7 in neural tube

defects in humans. *Birth defects research. Part A, Clinical and molecular teratology*, 103 (12), 1021–1027.

Wang, Q., Dhindsa, R.S., Carss, K., Harper, A.R., Nag, A., Tachmazidou, I., Vitsios, D., Deevi, S.V.V., Mackay, A., Muthas, D., Hühn, M., Monkley, S., Olsson, H., AstraZeneca Genomics Initiative, Wasilewski, S., Smith, K.R., March, R., Platt, A., Haefliger, C., and Petrovski, S., 2021. Rare variant contribution to human disease in 281,104 UK Biobank exomes. *Nature*, 597 (7877), 527–532.

Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., Hardison, R., Wang, T., and Yue, F., 2018. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome biology*, 19 (1), 151.

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D., 2019. A global overview of pleiotropy and genetic architecture in complex traits. *Nature genetics*, 51 (9), 1339–1348.

Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D., 2017. Functional mapping and annotation of genetic associations with FUMA. *Nature communications*, 8 (1), 1826.

Watanabe, T., Yamazaki, A., Kurabayashi, T., and Hanaoka, J.-I., 2005. Familial multicystic dysplastic kidney. *Pediatric nephrology (Berlin, Germany)*, 20 (8), 1200.

Waters, K.M., Stram, D.O., Hassanein, M.T., Le Marchand, L., Wilkens, L.R., Maskarinec, G., Monroe, K.R., Kolonel, L.N., Altshuler, D., Henderson, B.E., and Haiman, C.A., 2010. Consistent association of type 2 diabetes risk variants found in europeans in diverse racial and ethnic groups. *PLoS genetics*, 6 (8).

Webb, A., Miller, B., Bonasera, S., Boxer, A., Karydas, A., and Wilhelmsen, K.C., 2008. Role of the tau gene region chromosome inversion in progressive supranuclear palsy, corticobasal degeneration, and related disorders. *Archives of neurology*, 65 (11), 1473–1478.

Weber, S., Mir, S., Schlingmann, K.P., Nürnberg, G., Becker, C., Kara, P.E., Ozkayin, N., Konrad, M., Nürnberg, P., and Schaefer, F., 2005. Gene locus ambiguity in posterior urethral valves/prune-belly syndrome. *Pediatric nephrology* , 20 (8), 1036–1042.

Weber, S., Moriniere, V., Knüppel, T., Charbit, M., Dusek, J., Ghiggeri, G.M., Jankauskiené, A., Mir, S., Montini, G., Peco-Antic, A., Wühl, E., Zurowska, A.M., Mehls, O., Antignac, C., Schaefer, F., and Salomon, R., 2006. Prevalence of mutations in renal developmental genes in children with renal hypodysplasia: results of the ESCAPE study. *Journal of the American Society of Nephrology: JASN*, 17 (10), 2864–2870.

Weber, S., Taylor, J.C., Winyard, P., Baker, K.F., Sullivan-Brown, J., Schild, R., Knüppel, T., Zurowska, A.M., Caldas-Alfonso, A., Litwin, M., Emre, S.,

Ghiggeri, G.M., Bakkaloglu, A., Mehls, O., Antignac, C., Network, E., Schaefer, F., and Burdine, R.D., 2008. SIX2 and BMP4 mutations associate with anomalous kidney development. *Journal of the American Society of Nephrology: JASN*, 19 (5), 891–903.

Weber, S., Thiele, H., Mir, S., Toliat, M.R., Sozeri, B., Reutter, H., Draaken, M., Ludwig, M., Altmüller, J., Frommolt, P., Stuart, H.M., Ranjzad, P., Hanley, N.A., Jennings, R., Newman, W.G., Wilcox, D.T., Thiel, U., Schlingmann, K.P., Beetz, R., Hoyer, P.F., Konrad, M., Schaefer, F., Nürnberg, P., and Woolf, A.S., 2011. Muscarinic Acetylcholine Receptor M3 Mutation Causes Urinary Bladder Disease and a Prune-Belly-like Syndrome. *American journal of human genetics*, 89 (5), 668–674.

Weiner, D.J., iPSYCH-Broad Autism Group, Wigdor, E.M., Ripke, S., Walters, R.K., Kosmicki, J.A., Grove, J., Samocha, K.E., Goldstein, J.I., Okbay, A., Bybjerg-Grauholm, J., Werge, T., Hougaard, D.M., Taylor, J., Skuse, D., Devlin, B., Anney, R., Sanders, S.J., Bishop, S., Mortensen, P.B., Børglum, A.D., Smith, G.D., Daly, M.J., Robinson, E.B., and Psychiatric Genomics Consortium Autism Group, 2017. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nature genetics*, 49 (7), 978–985.

Weinstock, G.M., Wilson, R.K., Gibbs, R.A., and Kent, W.J., 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada*.

Wessels, M.W., Kuchinka, B., Heydanus, R., Smit, B.J., Dooijes, D., de Krijger, R.R., Lequin, M.H., de Jong, E.M., Husen, M., Willems, P.J., and Casey, B., 2010. Polyalanine expansion in the ZIC3 gene leading to X-linked heterotaxy with VACTERL association: a new polyalanine disorder? *Journal of medical genetics*, 47 (5), 351–355.

Westland, R., Renkema, K.Y., and Knoers, N.V.A.M., 2020. Clinical integration of genome diagnostics for congenital anomalies of the kidney and urinary tract. *Clinical journal of the American Society of Nephrology: CJASN*, 16 (1), 128–137.

Westland, R., Verbitsky, M., Vukojevic, K., Perry, B.J., Fasel, D.A., Zwijnenburg, P.J.G., Bökenkamp, A., Gille, J.J.P., Saraga-Babic, M., Ghiggeri, G.M., D'Agati, V.D., Schreuder, M.F., Gharavi, A.G., van Wijk, J.A.E., and Sanna-Cherchi, S., 2015. Copy number variation analysis identifies novel CAKUT candidate genes in children with a solitary functioning kidney. *Kidney international*, 88 (6), 1402–1410.

Wickham, H., 2011. Ggplot2. *Wiley interdisciplinary reviews. Computational statistics*, 3 (2), 180–185.

Wilson, J.G., Roth, C.B., and Warkany, J., 1953. An analysis of the syndrome of malformations induced by maternal vitamin A deficiency. Effects of restoration of vitamin A at various times during gestation. *The American journal of anatomy*, 92 (2), 189–217.

Wray, N.R., eQTLGen, Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A.F.T., Bigdeli, T.B., Binder, E.B., Blackwood, D.R.H., Bryois, J., Buttenschøn, H.N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J.H., Clarke, T.-K., Coleman, J.I.R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G.E., Crowley, C.A., Dashti, H.S., Davies, G., Deary, I.J., Degenhardt, F., Derks, E.M., Direk, N., Dolan, C.V., Dunn, E.C., Eley, T.C., Eriksson, N., Escott-Price, V., Kiadeh, F.H.F., Finucane, H.K., Forstner, A.J., Frank, J., Gaspar, H.A., Gill, M., Giusti-Rodríguez, P., Goes, F.S., Gordon, S.D., Grove, J., Hall, L.S., Hannon, E., Hansen, C.S., Hansen, T.F., Herms, S., Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.-J., Hougaard, D.M., Hu, M., Hyde, C.L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Lane, J.M., Li, Y., Li, Y., Lind, P.A., Liu, X., Lu, L., MacIntyre, D.J., MacKinnon, D.F., Maier, R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S.E., Mehta, D., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mill, J., Mondimore, F.M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Purcell, S.M., Quiroz, J.A., Qvist, P., Rice, J.P., Riley, B.P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.B.C., Smit, J.H., Smith, D.J., Stefansson, H., Steinberg, S., Stockmeier, C.A., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van der Auwera, S., van Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., de Geus, E.C.J., DePaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Hinds, D.A., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.F.A., Magnusson, P.K., Martin, N.G., McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M., Werge, T., Winslow, A.R., Lewis, C.M., Levinson, D.F., Breen, G., Børglum, A.D., Sullivan, P.F., 23andMe, and the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 2018. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics*, 50 (5), 668–681.

Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X., 2011. Rare-variant association testing for sequencing data with the sequence kernel association test. *The American Journal of Human Genetics*, 89 (1), 82–93.

Wu, N., Ming, X., Xiao, J., Wu, Z., Chen, X., Shinawi, M., Shen, Y., Yu, G., Liu, J., Xie, H., Gucev, Z.S., Liu, S., Yang, N., Al-Kateb, H., Chen, J., Zhang, J., Hauser, N., Zhang, T., Tasic, V., Liu, P., Su, X., Pan, X., Liu, C., Wang, L., Shen, J., Shen, J., Chen, Y., Zhang, T., Zhang, J., Choy, K.W., Wang, J., Wang, Q., Li, S., Zhou, W., Guo, J., Wang, Y., Zhang, C., Zhao, H., An, Y., Zhao, Y., Wang, J., Liu, Z., Zuo, Y., Tian, Y., Weng, X., Sutton, V.R., Wang, H., Ming, Y., Kulkarni, S., Zhong, T.P., Giampietro, P.F., Dunwoodie, S.L., Cheung, S.W., Zhang, X., Jin, L., Lupski, J.R., Qiu, G., and Zhang, F., 2015. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *The New England journal of medicine*, 372 (4), 341–350.

Wühl, E., van Stralen, K.J., Verrina, E., Bjerre, A., Wanner, C., Heaf, J.G., Zurriaga, O., Hoitsma, A., Niaudet, P., Palsson, R., Ravani, P., Jager, K.J., and Schaefer, F., 2013. Timing and outcome of renal replacement therapy in patients with congenital malformations of the kidney and urinary tract. *Clinical journal of the American Society of Nephrology: CJASN*, 8 (1), 67–74.

Wuttke, M., Li, Y., Li, M., Sieber, K.B., Feitosa, M.F., Gorski, M., Tin, A., Wang, L., Chu, A.Y., Hoppmann, A., Kirsten, H., Giri, A., Chai, J.-F., Sveinbjornsson, G., Tayo, B.O., Nutile, T., Fuchsberger, C., Marten, J., Cocca, M., Ghasemi, S., Xu, Y., Horn, K., Noce, D., van der Most, P.J., Sedaghat, S., Yu, Z., Akiyama, M., Afaq, S., Ahluwalia, T.S., Almgren, P., Amin, N., Ärnlöv, J., Bakker, S.J.L., Bansal, N., Baptista, D., Bergmann, S., Biggs, M.L., Biino, G., Boehnke, M., Boerwinkle, E., Boissel, M., Bottinger, E.P., Boutin, T.S., Brenner, H., Brumat, M., Burkhardt, R., Butterworth, A.S., Campana, E., Campbell, A., Campbell, H., Canouil, M., Carroll, R.J., Catamo, E., Chambers, J.C., Chee, M.-L., Chee, M.-L., Chen, X., Cheng, C.-Y., Cheng, Y., Christensen, K., Cifkova, R., Ciullo, M., Concas, M.P., Cook, J.P., Coresh, J., Corre, T., Sala, C.F., Cusi, D., Danesh, J., Daw, E.W., de Borst, M.H., De Grandi, A., de Mutsert, R., de Vries, A.P.J., Degenhardt, F., Delgado, G., Demirkan, A., Di Angelantonio, E., Dittrich, K., Divers, J., Dorajoo, R., Eckardt, K.-U., Ehret, G., Elliott, P., Endlich, K., Evans, M.K., Felix, J.F., Foo, V.H.X., Franco, O.H., Franke, A., Freedman, B.I., Freitag-Wolf, S., Friedlander, Y., Froguel, P., Gansevoort, R.T., Gao, H., Gasparini, P., Gaziano, J.M., Giedraitis, V., Gieger, C., Girotto, G., Giulianini, F., Gögele, M., Gordon, S.D., Gudbjartsson, D.F., Gudnason, V., Haller, T., Hamet, P., Harris, T.B., Hartman, C.A., Hayward, C., Hellwege, J.N., Heng, C.-K., Hicks, A.A., Hofer, E., Huang, W., Hutri-Kähönen, N., Hwang, S.-J., Ikram, M.A., Indridason, O.S., Ingelsson, E., Ising, M., Jaddoe, V.W.V., Jakobsdottir, J., Jonas, J.B., Joshi, P.K., Josyula, N.S., Jung, B., Kähönen, M., Kamatani, Y., Kammerer, C.M., Kanai, M., Kastarinen, M., Kerr, S.M., Khor, C.-C., Kiess, W., Kleber, M.E., Koenig, W., Kooner, J.S., Körner, A., Kovacs, P., Kraja, A.T., Krajcoviechova, A., Kramer, H., Krämer, B.K., Kronenberg, F., Kubo, M.,

Kühnel, B., Kuokkanen, M., Kuusisto, J., La Bianca, M., Laakso, M., Lange, L.A., Langefeld, C.D., Lee, J.J.-M., Lehne, B., Lehtimäki, T., Lieb, W., Lifelines Cohort Study, Lim, S.-C., Lind, L., Lindgren, C.M., Liu, J., Liu, J., Loeffler, M., Loos, R.J.F., Lucae, S., Lukas, M.A., Lyytikäinen, L.-P., Mägi, R., Magnusson, P.K.E., Mahajan, A., Martin, N.G., Martins, J., März, W., Mascalzoni, D., Matsuda, K., Meisinger, C., Meitinger, T., Melander, O., Metspalu, A., Mikaelsdottir, E.K., Milaneschi, Y., Miliku, K., Mishra, P.P., V. A. Million Veteran Program, Mohlke, K.L., Mononen, N., Montgomery, G.W., Mook-Kanamori, D.O., Mychaleckyj, J.C., Nadkarni, G.N., Nalls, M.A., Nauck, M., Nikus, K., Ning, B., Nolte, I.M., Noordam, R., O'Connell, J., O'Donoghue, M.L., Olafsson, I., Oldehinkel, A.J., Orho-Melander, M., Ouwehand, W.H., Padmanabhan, S., Palmer, N.D., Palsson, R., Penninx, B.W.J.H., Perls, T., Perola, M., Pirastu, M., Pirastu, N., Pistis, G., Podgornaia, A.I., Polasek, O., Ponte, B., Porteous, D.J., Poulain, T., Pramstaller, P.P., Preuss, M.H., Prins, B.P., Province, M.A., Rabelink, T.J., Raffield, L.M., Raitakari, O.T., Reilly, D.F., Rettig, R., Rheinberger, M., Rice, K.M., Ridker, P.M., Rivadeneira, F., Rizzi, F., Roberts, D.J., Robino, A., Rossing, P., Rudan, I., Rueedi, R., Ruggiero, D., Ryan, K.A., Saba, Y., Sabanayagam, C., Salomaa, V., Salvi, E., Saum, K.-U., Schmidt, H., Schmidt, R., Schöttker, B., Schulz, C.-A., Schupf, N., Shaffer, C.M., Shi, Y., Smith, A.V., Smith, B.H., Soranzo, N., Spracklen, C.N., Strauch, K., Stringham, H.M., Stumvoll, M., Svensson, P.O., Szymczak, S., Tai, E.-S., Tajuddin, S.M., Tan, N.Y.Q., Taylor, K.D., Teren, A., Tham, Y.-C., Thiery, J., Thio, C.H.L., Thomsen, H., Thorleifsson, G., Toniolo, D., Tönjes, A., Tremblay, J., Tzoulaki, I., Uitterlinden, A.G., Vaccargiu, S., van Dam, R.M., van der Harst, P., van Duijn, C.M., Velez Edward, D.R., Verweij, N., Vogelezang, S., Völker, U., Vollenweider, P., Waeber, G., Waldenberger, M., Wallentin, L., Wang, Y.X., Wang, C., Waterworth, D.M., Bin Wei, W., White, H., Whitfield, J.B., Wild, S.H., Wilson, J.F., Wojczynski, M.K., Wong, C., Wong, T.-Y., Xu, L., Yang, Q., Yasuda, M., Yerges-Armstrong, L.M., Zhang, W., Zonderman, A.B., Rotter, J.I., Bochud, M., Psaty, B.M., Vitart, V., Wilson, J.G., Dehghan, A., Parsa, A., Chasman, D.I., Ho, K., Morris, A.P., Devuyst, O., Akilesh, S., Pendergrass, S.A., Sim, X., Böger, C.A., Okada, Y., Edwards, T.L., Snieder, H., Stefansson, K., Hung, A.M., Heid, I.M., Scholz, M., Teumer, A., Köttgen, A., and Pattaro, C., 2019. A catalog of genetic loci associated with kidney function from analyses of a million individuals. *Nature genetics*, 51 (6), 957–972.

Xi, Q., Zhu, X., Wang, Y., Ru, T., Dai, C., Wang, Z., Li, J., and Hu, Y., 2016. Copy number variations in multicystic dysplastic kidney: update for prenatal diagnosis and genetic counseling. *Prenatal diagnosis*, 36 (5), 463–468.

Xie, J., Liu, L., Mladkova, N., Li, Y., Ren, H., Wang, W., Cui, Z., Lin, L., Hu, X., Yu, X., Xu, J., Liu, G., Caliskan, Y., Sidore, C., Balderes, O., Rosen, R.J., Bodria, M., Zanoni, F., Zhang, J.Y., Krithivasan, P., Mehl, K., Marasa, M., Khan, A., Ozay, F., Canetta, P.A., Bomback, A.S., Appel, G.B., Sanna-Cherchi, S., Sampson, M.G., Mariani, L.H., Perkowska-Ptasinska, A., Durlik, M., Mucha,

K., Moszczuk, B., Foroncewicz, B., Pączek, L., Habura, I., Ars, E., Ballarin, J., Mani, L.-Y., Vogt, B., Ozturk, S., Yildiz, A., Seyahi, N., Arikan, H., Koc, M., Basturk, T., Karahan, G., Akgul, S.U., Sever, M.S., Zhang, D., Santoro, D., Bonomini, M., Londrino, F., Gesualdo, L., Reiterova, J., Tesar, V., Izzi, C., Savoldi, S., Spotti, D., Marcantoni, C., Messa, P., Galliani, M., Roccatello, D., Granata, S., Zaza, G., Lugani, F., Ghiggeri, G., Pisani, I., Allegri, L., Sprangers, B., Park, J.-H., Cho, B., Kim, Y.S., Kim, D.K., Suzuki, H., Amoroso, A., Cattran, D.C., Fervenza, F.C., Pani, A., Hamilton, P., Harris, S., Gupta, S., Cheshire, C., Dufek, S., Issler, N., Pepper, R.J., Connolly, J., Powis, S., Bockenhauer, D., Stanescu, H.C., Ashman, N., Loos, R.J.F., Kenny, E.E., Wuttke, M., Eckardt, K.-U., Köttgen, A., Hofstra, J.M., Coenen, M.J.H., Kiemeney, L.A., Akilesh, S., Kretzler, M., Beck, L.H., Stengel, B., Debiec, H., Ronco, P., Wetzels, J.F.M., Zoledziewska, M., Cucca, F., Ionita-Laza, I., Lee, H., Hoxha, E., Stahl, R.A.K., Brenchley, P., Scolari, F., Zhao, M.-H., Gharavi, A.G., Kleta, R., Chen, N., and Kiryluk, K., 2020. The genetic architecture of membranous nephropathy and its potential to improve non-invasive diagnosis. *Nature communications*, 11 (1), 1600.

Xie, Q., Lin, T., Zhang, Y., Zheng, J., and Bonanno, J.A., 2005. Molecular cloning and characterization of a human AIF-like gene with ability to induce apoptosis. *The journal of biological chemistry*, 280 (20), 19673–19681.

Xu, B., Santos, S.A.A., and Hinton, B.T., 2018. Protein tyrosine kinase 7 regulates extracellular matrix integrity and mesenchymal intracellular RAC1 and myosin II activities during Wolffian duct morphogenesis. *Developmental biology*, 438 (1), 33–43.

Xu, B., Washington, A.M., Domeniconi, R.F., Ferreira Souza, A.C., Lu, X., Sutherland, A., and Hinton, B.T., 2016. Protein tyrosine kinase 7 is essential for tubular morphogenesis of the Wolffian duct. *Developmental biology*, 412 (2), 219–233.

Xu, G., Bu, S., Wang, X., Zhang, H., and Ge, H., 2020. Suppression of CCT3 inhibits the proliferation and migration in breast cancer cells. *Cancer cell international*, 20 (1), 218.

Yang, C., Wu, S., Mou, Z., Zhou, Q., Zhang, Z., Chen, Y., Ou, Y., Chen, X., Dai, X., Xu, C., Liu, N., and Jiang, H., 2021. Transcriptomic analysis identified ARHGAP family as a novel biomarker associated with tumor-promoting immune infiltration and nanomechanical characteristics in bladder cancer. *Frontiers in cell and developmental biology*, 9, 657219.

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A.A.E., Lee, S.H., Robinson, M.R., Perry, J.R.B., Nolte, I.M., van Vliet-Ostaptchouk, J.V., Snieder, H., LifeLines Cohort Study, Esko, T., Milani, L., Mägi, R., Metspalu, A., Hamsten, A., Magnusson, P.K.E., Pedersen, N.L., Ingelsson, E., Soranzo, N., Keller, M.C., Wray, N.R., Goddard, M.E., and Visscher, P.M., 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature genetics*, 47 (10), 1114–1120.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., and Visscher, P.M., 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics*, 42 (7), 565–569.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M., 2011. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88 (1), 76–82.

Yang, N., Wu, N., Dong, S., Zhang, L., Zhao, Y., Chen, W., Du, R., Song, C., Ren, X., Liu, J., Pehlivan, D., Liu, Z., Rao, J., Wang, C., Zhao, S., Breman, A.M., Xue, H., Sun, H., Shen, J., Zhang, S., Posey, J.E., Xu, H., Jin, L., Zhang, J., Liu, P., Sanna-Cherchi, S., Qiu, G., Wu, Z., Lupski, J.R., and Zhang, F., 2020. Human and mouse studies establish TBX6 in Mendelian CAKUT and as a potential driver of kidney defects associated with the 16p11.2 microdeletion syndrome. *Kidney international*, 98 (4), 1020–1030.

Yang, Y., Muzny, D.M., Reid, J.G., Bainbridge, M.N., Willis, A., Ward, P.A., Braxton, A., Beuten, J., Xia, F., Niu, Z., Hardison, M., Person, R., Bekheirnia, M.R., Leduc, M.S., Kirby, A., Pham, P., Scull, J., Wang, M., Ding, Y., Plon, S.E., Lupski, J.R., Beaudet, A.L., Gibbs, R.A., and Eng, C.M., 2013. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *The New England journal of medicine*, 369 (16), 1502–1511.

Yates, L.L., Papakrivopoulou, J., Long, D.A., Goggolidou, P., Connolly, J.O., Woolf, A.S., and Dean, C.H., 2010. The planar cell polarity gene Vangl2 is required for mammalian kidney-branching morphogenesis and glomerular maturation. *Human molecular genetics*, 19 (23), 4663–4676.

Yi, K. and Ju, Y.S., 2018. Patterns and mechanisms of structural variations in human cancer. *Experimental & molecular medicine*, 50 (8), 1–11.

Young, J.M., Endicott, R.M., Parghi, S.S., Walker, M., Kidd, J.M., and Trask, B.J., 2008. Extensive copy-number variation of the human olfactory receptor gene family. *The American Journal of Human Genetics*, 83 (2), 228–242.

Yu, K., Wang, Z., Li, Q., Wacholder, S., Hunter, D.J., Hoover, R.N., Chanock, S., and Thomas, G., 2008. Population substructure and control selection in genome-wide association studies. *PloS one*, 3 (7), e2551.

van der Zanden, L.F.M., van Rooij, I.A.L.M., Quaedackers, J.S.L.T., Nijman, R.J.M., Steffens, M., de Wall, L.L.L., Bongers, E.M.H.F., Schaefer, F., Kirchner, M., Behnisch, R., Bayazit, A.K., Caliskan, S., Obrycki, L., Montini, G., Duzova, A., Wuttke, M., Jennings, R., Hanley, N.A., Milmoe, N.J., Winyard, P.J.D., Renkema, K.Y., Schreuder, M.F., Roeleveld, N., and Feitz, W.F.J., 2021. CDH12 as a candidate gene for kidney injury in posterior urethral valve cases: A genome-wide association study among patients with obstructive uropathies. *European urology open science*, 28, 26–35.

Zhan, X., Hu, Y., Li, B., Abecasis, G.R., and Liu, D.J., 2016. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics (Oxford, England)*, 32 (9), 1423–1426.

Zhang, J.-X., Chen, Z.-H., Chen, D.-L., Tian, X.-P., Wang, C.-Y., Zhou, Z.-W., Gao, Y., Xu, Y., Chen, C., Zheng, Z.-S., Weng, H.-W., Ye, S., Kuang, M., Xie, D., and Peng, S., 2018. LINC01410-miR-532-NCF2-NF-kB feedback loop promotes gastric cancer angiogenesis and metastasis. *Oncogene*, 37 (20), 2660–2675.

Zhang, R., Knapp, M., Suzuki, K., Kajioka, D., Schmidt, J.M., Winkler, J., Yilmaz, Ö., Pleschka, M., Cao, J., Kockum, C.C., Barker, G., Holmdahl, G., Beaman, G., Keene, D., Woolf, A.S., Cervellione, R.M., Cheng, W., Wilkins, S., Gearhart, J.P., Sirchia, F., Di Grazia, M., Ebert, A.-K., Rösch, W., Ellinger, J., Jenetzky, E., Zwink, N., Feitz, W.F., Marcelis, C., Schumacher, J., Martinón-Torres, F., Hibberd, M.L., Khor, C.C., Heilmann-Heimbach, S., Barth, S., Boyadjiev, S.A., Brusco, A., Ludwig, M., Newman, W., Nordenskjöld, A., Yamada, G., Odermatt, B., and Reutter, H., 2017. ISL1 is a major susceptibility gene for classic bladder exstrophy and a regulator of urinary tract development. *Scientific reports*, 7, 42170.

Zhao, E., Zhang, Y., Zeng, X., and Liu, B., 2015. Association between maternal diabetes mellitus and the risk of congenital malformations: A meta-analysis of cohort studies. *Drug discoveries & therapeutics*, 9 (4), 274–281.

Zhao, X., Qiao, D., Yang, C., Kasela, S., Kim, W., Ma, Y., Shrine, N., Batini, C., Sofer, T., Taliun, S.A.G., Sakornsakolpat, P., Balte, P.P., Prokopenko, D., Yu, B., Lange, L.A., Dupuis, J., Cade, B.E., Lee, J., Gharib, S.A., Daya, M., Laurie, C.A., Ruczinski, I., Cupples, L.A., Loehr, L.R., Bartz, T.M., Morrison, A.C., Psaty, B.M., Vasan, R.S., Wilson, J.G., Taylor, K.D., Durda, P., Johnson, W.C., Cornell, E., Guo, X., Liu, Y., Tracy, R.P., Ardlie, K.G., Aguet, F., VanDenBerg, D.J., Papanicolaou, G.J., Rotter, J.I., Barnes, K.C., Jain, D., Nickerson, D.A., Muzny, D.M., Metcalf, G.A., Doddapaneni, H., Dugan-Perez, S., Gupta, N., Gabriel, S., Rich, S.S., O'Connor, G.T., Redline, S., Reed, R.M., Laurie, C.C., Daviglus, M.L., Preudhomme, L.K., Burkart, K.M., Kaplan, R.C., Wain, L.V., Tobin, M.D., London, S.J., Lappalainen, T., Oelsner, E.C., Abecasis, G.R., Silverman, E.K., Barr, R.G., NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, TOPMed Lung Working Group, Cho, M.H., and Manichaikul, A., 2020. Whole genome sequence analysis of pulmonary function and COPD in 19,996 multi-ethnic participants. *Nature communications*, 11 (1), 5182.

Zheng, B., Wang, C., Seltzsam, S., Schneider, S., Schierbaum, L., Wu, W., Dai, R., Connaughton, D.M., Nakayama, M., Mann, N., Bauer, S.B., Awad, H.S., Eid, L.A., Tasic, V., Shril, S., and Hildebrandt, F., 2022. A truncating NRIP1 variant in an Arabic family with congenital anomalies of the kidneys and urinary tract. *American journal of medical genetics. Part A*, 188 (1), 310–313.

Zheng, Y., Xu, J., Guo, W., Xu, H., Chen, J., Shen, Q., Zhang, X., and Zhai, Y., 2015. The significance of Pax2 expression in the ureter epithelium of children with vesicoureteric reflux. *Human pathology*, 46 (7), 963–970.

Zhou, W., Bi, W., Zhao, Z., Dey, K.K., Jagadeesh, K.A., Karczewski, K.J., Daly, M.J., Neale, B.M., and Lee, S., 2022. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nature genetics*.

Zhou, W., Kanai, M., Wu, K.-H.H., Humaira, R., Tsuo, K., Hirbo, J.B., Wang, Y., Bhattacharya, A., Zhao, H., Namba, S., Surakka, I., Wolford, B.N., Faro, V.L., Lopera-Maya, E.A., Läll, K., Favé, M.-J., Chapman, S.B., Karjalainen, J., Kurki, M., Mutaamba, M., Brumpton, B.M., Chavan, S., Chen, T.-T., Daya, M., Ding, Y., Feng, Y.-C.A., Gignoux, C.R., Graham, S.E., Hornsby, W.E., Ingold, N., Johnson, R., Laisk, T., Lin, K., Lv, J., Millwood, I.Y., Palta, P., Pandit, A., Preuss, M., Thorsteinsdottir, U., Uzunovic, J., Zawistowski, M., Zhong, X., Campbell, A., Crooks, K., De Bock, G. h., Douville, N.J., Finer, S., Fritsche, L.G., Griffiths, C.J., Guo, Y., Hunt, K.A., Konuma, T., Marioni, R.E., Nomdo, J., Patil, S., Rafaels, N., Richmond, A., Shortt, J.A., Straub, P., Tao, R., Vanderwerff, B., Barnes, K.C., Boezen, M., Chen, Z., Chen, C.-Y., Cho, J., Smith, G.D., Finucane, H.K., Franke, L., Gamazon, E., Ganna, A., Gaunt, T.R., Ge, T., Huang, H., Huffman, J., Lajonchere, C., Law, M.H., Li, L., Lindgren, C.M., Loos, R.J.F., MacGregor, S., Matsuda, K., Olsen, C.M., Porteous, D.J., Shavit, J.A., Snieder, H., Trembath, R.C., Vonk, J.M., Whiteman, D., Wicks, S.J., Wijmenga, C., Wright, J., Zheng, J., Zhou, X., Awadalla, P., Boehnke, M., Cox, N.J., Geschwind, D.H., Hayward, C., Hveem, K., Kenny, E.E., Lin, Y.-F., Mägi, R., Martin, H.C., Medland, S.E., Okada, Y., Palotie, A.V., Pasaniuc, B., Sanna, S., Smoller, J.W., Stefansson, K., van Heel, D.A., Walters, R.G., Zoellner, S., Martin, A.R., Willer, C.J., Daly, M.J., Neale, B.M., Japan, B., BioMe, BioVU, Canadian Partnership for Tomorrow, China Kadoorie Biobank Collaborative Group, Colorado Center for Personalized Medicine, deCODE Genetics, Biobank, E., FinnGen, Scotland, G., Genes & Health, LifeLines, Biobank, M.G.B., Michigan Genomics Initiative, QIMR Berghofer Biobank, Biobank, T., The HUNT Study, UCLA ATLAS Community Health Initiative, and UK Biobank, 2021. Global Biobank Meta-analysis Initiative: powering genetic discovery across human diseases. *bioRxiv*.

Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., Bastarache, L.A., Wei, W.-Q., Denny, J.C., Lin, M., Hveem, K., Kang, H.M., Abecasis, G.R., Willer, C.J., and Lee, S., 2018. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature genetics*, 50 (9), 1335–1341.

Zhou, W., Zhao, Z., Nielsen, J.B., Fritsche, L.G., LeFaive, J., Gagliano Taliun, S.A., Bi, W., Gabrielsen, M.E., Daly, M.J., Neale, B.M., Hveem, K., Abecasis, G.R., Willer, C.J., and Lee, S., 2020. Scalable generalized linear mixed model for region-based association tests in large biobanks and cohorts. *Nature genetics*, 52 (6), 634–639.

Zhu, H. and Zhou, X., 2020. Statistical methods for SNP heritability estimation and partition: A review. *Computational and structural biotechnology journal*, 18, 1557–1568.

Zivná, M., Hůlková, H., Matignon, M., Hodanová, K., Vylet'al, P., Kalbácová, M., Baresová, V., Sikora, J., Blazková, H., Zivný, J., Ivánek, R., Stránecký, V., Sovová, J., Claes, K., Lerut, E., Fryns, J.-P., Hart, P.S., Hart, T.C., Adams, J.N., Pawtowski, A., Clemessy, M., Gasc, J.-M., Gübler, M.-C., Antignac, C., Elleder, M., Kapp, K., Grimbert, P., Bleyer, A.J., and Kmoch, S., 2009. Dominant renin gene mutations associated with early-onset hyperuricemia, anemia, and chronic kidney failure. *The American Journal of Human Genetics*, 85 (2), 204–213.