Title

The copy number variation landscape of congenital anomalies of the kidney and urinary tract.

Journal

Nature Genetics, 51(1)

Authors

Verbitsky, Miguel
Westland, Rik
Perez, Alejandra
et al.

Publication Date

2019

DOI

10.1038/s41588-018-0281-y

Peer reviewed

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2019 July 31.

Published in final edited form as:

Nat Genet. 2019 January ; 51(1): 117–127. doi:10.1038/s41588-018-0281-y.

The copy number variation landscape of congenital anomalies of the kidney and urinary tract

A full list of authors and affiliations appears at the end of the article.

Abstract

Congenital anomalies of the kidney and urinary tract (CAKUT) are a major cause of pediatric kidney failure. We performed a genome-wide analysis of copy number variants (CNVs) in 2,824 cases and 21,498 controls. Affected individuals carried a significant burden of rare exonic (i.e. affecting coding regions) CNVs and were enriched for known genomic disorders (GD). Kidney anomaly (KA) cases were most enriched for exonic CNVs, encompassing GD-CNVs and novel deletions; obstructive uropathy (OU) had a lower CNV burden and an intermediate prevalence of GD-CNVs; vesicoureteral reflux (VUR) had the fewest GD-CNVs but was enriched for novel exonic CNVs, particularly duplications. Six loci (1q21, 4p16.1-p16.3, 16p11.2, 16p13.11, 17q12, and 22q11.2) accounted for 65% of patients with GD-CNVs. Deletions at 17q12, 4p16.1-p16.3, and 22q11.2 were specific for KA; the 16p11.2 locus showed extensive pleiotropy. Using a multidisciplinary approach, we identified *TBX6* as a driver for the CAKUT subphenotypes in the 16p11.2 microdeletion syndrome.

INTRODUCTION

Congenital anomalies of the kidney and urinary tract (CAKUT) have devastating impact on childhood renal survival[1–4]. A better understanding of the pathogenesis of CAKUT is imperative to improve the prognosis of affected children[5]. CAKUT encompasses a broad

$To whom correspondence should be addressed, Simone Sanna-Cherchi, M.D., Division of Nephrology, Columbia University College of Physicians and Surgeons, 1150 Street Nicholas Avenue, Russ Berrie Pavilion #412D, New York, New York 10032, USA. ss2517@cumc.columbia.edu., Ali G. Gharavi, M.D., Division of Nephrology, Columbia University College of Physicians and Surgeons, 1150 Street Nicholas Avenue, Russ Berrie Pavilion #412C, New York, New York 10032, USA. ag2239@cumc.columbia.edu., Cathy L. Mendelsohn, Ph.D., Department of Urology, Columbia University College of Physicians and Surgeons, 1130 Street Nicholas Avenue, IRCC #311, New York, New York 10032, USA. clm20@cumc.columbia.edu., Virginia E. Papaioannou, Ph.D., Department of Genetics and Development, Columbia University Medical Center, 701 W 168th Street, HHSC #1602, New York, New York, 10032, USA. vep1@cumc.columbia.edu.

*Both authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

S.S-C. directed the project. V.E.P., C.L.M., A.G.G. and S.S-C. designed the project. M.V., R.W., A.P., Q.L., P.K., D.A.F., E.B., M.W., J.M., V.P.C., Y.N., T.L., D.A., H.W. performed the experiments and/or data generation. M.G.S., M.G.D., J.M.D., P.P., D.E.B., S.L.F., B.A.W., C.J., D.M.M., E.A.O., A.C.S-e-S, F.H. and H.H. contributed array genotype data for CNV analyses. M.V., R.W., P.K., A.P., E.B., A.M., V.E.P., C.L.M. and S.S-C analyzed the data. K.K., J.M.B., and B.L. provided critical intellectual content to the design of the study. All other authors (A.M., M.B., C.K., A.V., S.S., B.K, M.M., J.Y.Z., P.L.W., E.L.H., A.C., G.P., L.G., V.M., G.M., M.G., D.C., C.I., F.S., J.A.E.vW., M.S., D.S., G.C., P.Z., D.D., K.Z., M.M., M.T., D.T., A.K., P.S., T.J., M.B.-K, R.P., M.Sz., P.A., M.M-W., G.K., A.S., M.Z., Z.G., V.J.L., V.T., I.P., L.A., L.M.R., J.M.C., S.A., P.C., F.L., W.N., G.M.G., A.L-B., A.M-K. and F.Z.) recruited cases and submitted clinical information for the study. M.V., R.W., V.E.P., C.L.M., A.G.G., and S.S-C. wrote the draft of the manuscript. All authors critically revised the manuscript.

Competing Financial Interests Statement

The authors have no competing financial interests related to this manuscript.

spectrum of phenotypes, which can result from early disruptions in transcription factors and signaling molecules such as *PAX2*, *EYA1*, *RET*, *BMP4*[5,6] and others, or are directly related to spatiotemporal interactions of the outgrowing ureteric bud and metanephric mesenchyme[7–9]. Early disruption of these interactions leads to renal agenesis, hypoplasia or -dysplasia, whereas later perturbations in the outgrowth of the ureteric bud result in obstructive uropathy (OU), vesicoureteric reflux (VUR) or ectopic or horseshoe kidney (EK-HK)[6,10–13]. Maldevelopment of the lower urinary tract can result in epispadias or hypospadias (LUTM) or posterior urethral valves (PUV)[14]. Genetic manipulation in mice indicates that disruption of the same cellular pathways can lead to multiple different genitourinary phenotypes[15–19]. Similarly, mutations in genes associated with Mendelian forms of CAKUT can lead to different subphenotypes in individuals from the same families[18,20,21], suggesting that a single genetic lesion can have pleiotropic manifestations across the spectrum of CAKUT. Conversely, differences in prevalence and severity of structural malformations point towards distinct molecular basis and genetic architecture[22,23]. To date, there are over 50 single-gene disorders underlying isolated and non-isolated (i.e. syndromic) CAKUT[5,24,25]. Furthermore, a significant number of CAKUT patients carry copy number variants (CNV) that have been previously associated with a syndrome diagnosis or are large and extremely rare in the general population[26–28]. Nevertheless, a molecular diagnosis can be established in less than 20% of affected individuals[15–17,29–32], emphasizing that large studies across the entire phenotypic spectrum of CAKUT are indispensable to identify those genes and allelic variants that are either specific to subcategories of disease or those that have pleiotropic effects across the entire genitourinary tract, and to discover novel cellular pathways that are implicated in kidney and urinary development.

Here, we show both the presence of a distinct genetic architecture as well as pleiotropic mutations for the different subphenotypes of CAKUT. Our CNV analysis of nearly 3,000 cases across the phenotypic spectrum of CAKUT sheds light on the genomic architecture of disease, and implicates *TBX6* as a main driver for the various CAKUT phenotypes in the 16p11.2 microdeletion syndrome.

RESULTS

Burden of rare CNVs is high in CAKUT

We conducted a study in 2,824 CAKUT cases and 21,498 population controls (Supplementary Tables 1 and 2 and Supplementary Figure 1) to compare the prevalence of rare CNVs that intersect genes. The case cohort represented common CAKUT subcategories: kidney anomalies (KA; including renal agenesis, hypoplasia, dysplasia, and multicystic dysplasia), vesicoureteral reflux (VUR), obstructive uropathy (OU; including congenital hydronephrosis, ureteropelvic junction obstruction, ureterovesical junction obstruction, and congenital megaureter), duplicated collecting system (DCS; including duplications of the ureter or kidney, partial and complete), posterior urethral valves (PUV), ectopic kidney or horseshoe kidney (EK-HK), and other lower urinary tract malformations (LUTM; including anomalies of the bladder and anterior urethra). Our analysis focused on

large (≥100 kb) CNVs that are present in less than 1:1,000 population controls, across different ancestries as estimated by principal component analysis (Supplementary Figure 2).

This analysis revealed marked enrichment for large, rare CNVs in CAKUT compared to controls ($P = 1.04 \times 10^{-24}$; Figure 1A), consistent across virtually all metrics examined, including the the number of individuals with large, rare CNVs, the median size and total CNV span per genome, and the fraction of GD-CNVs, indicating an important role for gene-disrupting CNVs across the whole CAKUT spectrum (Table 1 and Supplementary Table 3). This signal was driven by cases with and without extrarenal manifestations, since the burden difference was still highly significant when analyzing cases without extrarenal defects separately ($P = 3.21 \times 10^{-8}$; Figure 1A–B and Supplementary Figures 3 and 4). Even when considering only simplex isolated CAKUT cases (i.e. cases without extrarenal manifestation and with no additional CAKUT phenotypes other than the primary one), there was still an excess burden of large, rare CNVs compared to controls ($P = 1.15 \times 10^{-8}$; Supplementary Figure 5). Comparison of burden metrics for cases and controls indicated a population attributable risk of 4.1% for large, rare CNVs ≥500 kb in CAKUT (OR 1.64, 95% CI 1.44–1.87; $P = 3.23 \times 10^{-13}$). This excess burden is predominantly attributable to exonic deletions, most prominently in the KA cases (Supplementary Tables 4 and 5). Interestingly, we observed an enrichment for rare duplications compared to deletions in VUR and PUV cases when compared to controls ($P = 3.33 \times 10^{-2}$ and $P = 8.67 \times 10^{-4}$, respectively; Supplementary Figure 6). Secondary analysis also showed enrichment of the number of genes per individual genome that were affected by rare CNVs for nearly all CAKUT subcategories (Supplementary Table 6 and Supplementary Figure 7).

Genomic disorders inform about the genetic architecture of CAKUT

We cross-annotated all rare CNVs with a curated list of known genomic disorders (Supplementary Table 7)[33,34] and identified 45 distinct known genomic disorders (GD) in 112 (4.0%) independent cases (Supplementary Table 8, Figure 1C). Five cases carried more than one known GD-CNV, resulting in a total number of 117 known GD-CNVs in our cohort (Supplementary Table 9); in comparison, known GD-CNVs were found in only 134/21,498 (0.6%) population controls (OR 6.58, 95% CI 5.05–8.55; $P = 7.53 \times 10^{-41}$; Supplementary Table 10).

Further annotation identified 54 large, rare, exonic CNVs in an additional 47 CAKUT cases (1.7%) that fulfilled ACMG classification[35,36] as likely pathogenic imbalances (Online Methods and Supplementary Table 11). Among these, CNVs that can be classified as pathogenic were an atypical deletion at the 16p11.2 locus (see below), a 300 kb deletion involving *PAX2*, a 100 kb duplication containing *TBX18,* a 571 kb deletion spanning *PBX1,* and a 6.8 Mb duplication including *BMP4* (Supplementary Figure 8A–D). We also identified overlapping duplications at the 15p11.2 locus in 5 cases with ureteric defects (VUR, OU) and PUV, as well as two KA cases with deletions proximal to the 16p11.2 microdeletion syndrome. In line with the burden test results, the distribution of deletions and duplications at known and novel GD-CNV loci (Figure 1C,D) was significantly different among CAKUT categories (6 × 2 Fisher's Exact Test $P = 8.93 \times 10^{-5}$). In particular, subjects with KA and

OU were enriched for deletion syndromes, contrary to PUV and DCS that showed an excess of duplications at the same genomic loci.

When we examined associations with disease severity, cases who carried a known GD-CNV were more likely to have multiple sites of the urinary tract affected (OR 1.60, 95% CI 1.05–2.42; $P = 0.02$) and more frequently harbored extrarenal malformations (OR 4.79, 95% CI 3.21–7.17; $P = 6.00 \times 10^{-15}$) compared to cases without a known GD-CNV. However, consistent with the burden tests conducted above, analysis of simplex isolated cases still showed increased burden compared to controls (OR 3.12, 95% CI 2.06–4.61; $P = 1.86 \times 10^{-7}$), strongly implicating the significance of GD-CNVs in milder forms of CAKUT. To test more complex genetic models for modes of disease determination, we examined cases and controls for second site CNVs. Among the 159 cases with GD-CNVs (112 known, 47 new/likely pathogenic: altogether called “diagnostic CNVs” (DCNVs)) eleven (6.9%) carried more than one DCNV (Supplementary Table 12). In cases, the presence of zero, one, or more than one DCNV increased the likelihood of extrarenal malformations (Chi-square test for 3 × 2 table: $P = 6.94 \times 10^{-7}$; Supplementary Figure 9).

KA, OU, and VUR show distinct genomic characteristics

Comparison of CNV landscape between the largest CAKUT subcategories revealed both commonalities and differences between KA, OU, and VUR (Table 2). We found significant CNV enrichment for all three phenotypes, which was most uniformly shown by a larger proportion of cases carrying exonic CNVs ≥100 kb compared to controls. KA cases had the highest CNV burden ($P = 9.01 \times 10^{-25}$; Figure 1B), as evidenced by median CNV size and total span as well as proportion of individuals with large imbalances that could be classified as pathogenic GD-CNVs (80 cases with GD-CNVs, OR 12.65, 95% CI 9.40–16.94; $P = 8.53 \times 10^{-50}$; Table 2 and Figure 1C). After removal of individuals with a known GD-CNV, the strength of the excess CNV burden in KA was still detectable but markedly attenuated ($P = 1.27 \times 10^{-3}$), consistent with the major role for GD-CNVs in the pathogenesis of KA (Supplementary Figure 10). In contrast, VUR cases were also affected by a high CNV burden ($P = 7.56 \times 10^{-7}$), but these CNVs predominantly involved duplications and were less likely to be classified as GD-CNVs (7 cases with GD-CNVs, OR 1.71; 95% CI 0.67–3.64; $P = 0.20$; Figure 1C). The large rare CNVs in VUR were mostly classified as likely pathogenic, potentially indicating that novel syndromes account for the molecular basis of this disorder (Table 2, Figure 1D and Supplementary Figures 3 and 10). Finally, OU cases fell in an intermediate category, with CNVs that were not statistically larger compared to controls, yet the excess CNV burden ($P = 0.01$; Supplementary Figure 3) was reflected by a greater proportion of individuals with large imbalances that were also more likely to be classified as pathogenic GD-CNVs (11 cases with GD-CNVs, OR 3.50, 95% CI 1.70–6.52; $P = 5.96 \times 10^{-4}$; Table 3 and Supplementary Figure 3). These data suggest a distinct genomic architecture among CAKUT subcategories, with enrichment of GD-CNVs in KA, and enrichment for novel, large or intermediate size imbalances in VUR and OU, respectively.

Six GD loci account for 65% of CAKUT cases with known GD-CNVs

We conducted a literature search, including a survey of databases of genomic variants such as DECIPHER[37,38] and ISCA[35], and found either a known link to CAKUT[5,23,26,39–46] or case series or reports in which CAKUT was part of the clinical phenotype[47–53] in 43 out of 45 GD-CNVs. This finding underlines that these GD-CNVs are causally related to CAKUT.

While the majority of GD-CNVs affected unique or few cases (Supplementary Table 8), six loci explained 73 of the 112 (65%) cases who carried a known GD-CNV (Figure 2). These common GD loci included chromosome 1q21.1 (7 cases = 6.2% of cases with a GD-CNV; five deletions, two duplications), chromosome 4p16.1-p16.3 (Wolf-Hirschhorn syndrome; 5 cases = 4.4% of cases with a GD-CNV; all deletions), chromosome 16p11.2 (9 cases = 8.0% of cases with a GD-CNV; eight deletions, one duplication), chromosome 16p13.11 (9 cases = 8.0% of cases with a GD-CNV; four deletions, five duplications), chromosome 17q12 (26 cases = 23.0% of cases with a GD-CNV; 23 deletions, three duplications), and chromosome 22q11.2 (17 cases = 15.0% of cases with a GD-CNV; 14 deletions, three duplications) loci. Genotype-phenotype correlations identified that microdeletions, but not duplications, were enriched in upper urinary tract defects, especially KA (1q21.1, 4p16.1-p16.3, 17q12, 22q11.2), whereas the 16p11.2 microdeletion syndrome was identified in all CAKUT subcategories, suggesting a high pleiotropic effect. These observations provide further support that some genetic lesions result in specific CAKUT subphenotypes while other genetic lesions, such as the 16p11.2 microdeletion syndrome, have high pleiotropic effects across the genitourinary tract. These loci provide a list of regions that are likely to encompass critical regulators of kidney and urinary tract development in humans, offering a unique opportunity for gene discovery.

Exome sequencing indicates haploinsufficiency underlying CNV deletions

To test the mechanism by which pathogenic CNVs confer risk to CAKUT, we conducted whole exome sequencing (WES) in 23 patients harboring pathogenic microdeletions at 14 independent loci (Supplementary Table 13). Based on recessive loss-of-function (LOF) inheritance, WES would uncover a hemizygous LOF mutation on the non-deleted allele (unmasking effect). WES was performed as previously described[18,23,54]. We retrieved all identified candidate hemizygous LOF variants located within the case-specific deletion loci. Overall we identified only one LOF variant in *EFCAB12* (p.Q437*; observed once in heterozygosity in the ExAC database[55,56]) in a patient affected by unilateral KA and multiple extrarenal manifestations, including neurodevelopmental delay, epilepsy, corpus callosum agenesis, left radial bone agenesis and a patent ductus arteriosus, who harbored a 14.9 Mb deletion at chromosome 3q13.22–13.1 (Supplementary Figure 11)[57]. The *EFCAB12* variant was inherited from a heterozygous unaffected mother, while the CNV occurred *de novo* (Supplementary Figure 11)[57]. Query of 15,469 control individual exomes at the Institute of Genomic Medicine at Columbia University did not reveal any homozygous or compound heterozygous truncating variants, suggesting that biallelic truncating mutations are likely not tolerated in humans and propose recessive mutations in *EFCAB12* as contributors to the developmental syndrome of this individual. In addition, we performed clinical annotation of genes that are known to be implicated in Mendelian forms of CAKUT, by querying WES data for an in-house gene list as described[58]. We did not find any

pathogenic variants in known CAKUT genes for any of the 23 deletion carrier patients. Hence, our WES studies suggest haploinsufficiency as main pathogenetic mechanism for the CAKUT-associated deletion CNVs.

CAKUT is a common phenotype in the 16p11.2 microdeletion syndrome

We identified nine CAKUT cases and five controls with overlapping 16p11.2 microdeletions (0.32% vs 0.02%; OR 13.7, 95% CI 4.1–52.2; $P = 4.39 \times 10^{-6}$; Figure 2, Table 3 and Supplementary Tables 9 and 10), implicating CAKUT as an important feature of this syndrome. To better estimate the prevalence of genitourinary malformations in individuals who harbor this GD-CNV, we first analyzed the clinical reports from 186 cases with 16p11.2 microdeletion syndrome in the DECIPHER database (Supplementary Table 14). The most prevalent associated conditions were abnormalities of the nervous system (92.5%), abnormalities of head or neck (26.9%), growth defects (23.7%), and abnormalities of the limbs (14.5%). Abnormalities of the genitourinary system (including KA, hydronephrosis, and VUR) were reported in only 10 cases (5.4%), but these estimates likely reflect the standard clinical indication for ordering a DNA microarray for diagnosis of a syndromic disease, i.e. neurodevelopmental delay and dysmorphic features. In fact, abnormalities of the skeletal system, which are a hallmark for the 16p11.2 microdeletion syndrome[59,60], were reported in only 19 cases (10.2%), since, similarly to individuals with CAKUT, these patients are rarely referred for genetic testing. To obtain a more unbiased estimate of prevalence of genitourinary defects in patients with 16p11.2 microdeletion, we queried the data warehouse of the Center for Applied Genomics (CAG) at the Children Hospital of Philadelphia (CHOP). We analyzed medical records for 42 children with the 16p11.2 microdeletion syndrome (Supplementary Table 15). Neurodevelopmental defects and brain anomalies, and spine defects were the most prevalent conditions, recorded in 22 (52.4%) and 6 (14.2%) cases, respectively. Imaging studies and clinical data assessing the kidney and urinary tract were available for 15 out of 42 and genitourinary anomalies were present in 6 out of these 15 cases (40.0%), highlighting that CAKUT, similar to scoliosis, is a common feature in this syndrome characterized by incomplete penetrance. Consistent with our observation of a high pleiotropic effect for the 16p11.2 microdeletion, the phenotype was also variable in this cohort: two patients with OU, two with VUR, one with PUV, and one with LUTM. These data implicate CAKUT as a significant feature of the chromosome 16p11.2 microdeletion syndrome and provide an opportunity to identify novel key regulators of kidney and urinary tract development.

TBX6 is a driver for CAKUT in the 16p11.2 microdeletion syndrome

We recently demonstrated that the combinatorial use of genomic analyses with functional modeling in vertebrates can lead to the identification of key drivers of CAKUT phenotypes from microdeletion syndromes[23]. Analysis of breakpoints in 9 CAKUT cases with typical or atypical 16p11.2 microdeletions identified a ~175 Kb minimal region of overlap (MRO) predicted to harbor the genetic driver(s) for CAKUT (Supplementary Figure 12). Among the 19 genes included in the MRO, *TBX6* appeared as a strong candidate: it is included in the list of essential genes from the International Mouse Phenotyping Consortium[61]; it is depleted from LOF mutations in subjects from the ExAC database (aggregate prevalence of LOF <0.001); it has a role in paraxial and intermediate mesoderm development[62,63]; and its

inactivation in humans and rodents leads to spine defects including congenital scoliosis and spondylocostal dysostosis[59,62,64–66]. The connection of *TBX6* mutations with vertebral anomalies is of particular interest because of the historically well-known clinical association between congenital scoliosis and CAKUT[67–69].

To provide a functional link between inactivation of *TBX6* and CAKUT, we generated an allelic series by cross breeding mice with two different alleles: a *Tbx6* null allele (*Tbx6*^{tm2PA})[70] and a *Tbx6* hypomorphic allele, the spontaneous mutant *Tbx6*^{rv} (hereafter referred to as *Tbx6*$^{-}$ and *Tbx6*^{rv}, respectively) previously studied for vertebral development and spine defects[71]. Since the homozygous null mutation is embryonically lethal at E9.5, precluding analysis of the developing urinary tract, we first studied compound heterozygous embryos (*Tbx6*$^{rv/-}$), which retain sufficient residual expression of *Tbx6* for survival past E9.5. We studied 19 *Tbx6*$^{rv/-}$ embryos at E17.5-E18.5, and observed full penetrance of CAKUT with variable expressivity of phenotypes characterized by unilateral or bilateral renal agenesis, unilateral or bilateral renal hypoplasia and dysplasia, and obstructive uropathy characterized by hydronephrosis and hydroureter (Table 4, Figure 3A–L). Microscopic analysis of renal tissue from *Tbx6*$^{rv/-}$ embryos at stages from E13.5 to E18.5 showed variable severity of kidney anomalies (Figure 3D–L). Severe phenotypes included unilateral or bilateral renal agenesis, rudimentary kidneys and undeveloped renal parenchyma embedded in the paraspinal musculature (Figure 3F,I,L). Milder phenotypes included unilateral and bilateral renal hypoplasia with hydroureter, tubule dilation and hydronephrosis (Figure 3E,H,K). These data strongly implicate *TBX6* as a major driver for CAKUT phenotypes. Comparative analysis was performed on wildtype and *Tbx6*$^{rv/-}$ embryos at E11.5 after staining with E-cadherin (Cdh1), an epithelial marker, and Pax2, which labels mesenchymal progenitors that produce nephrons, the ureteric bud which gives rise to the renal collecting duct system, and the common nephric duct (CND)[72–74]. This analysis revealed that the events leading to CAKUT (renal parenchyma abnormalities as well as obstruction) are present at very early stages during development (Figure 3M,N). In wildtype E11.5 embryos, the ureteric bud has invaded the metanephric blastema and has undergone a round of branching (Figure 3N). In the *Tbx6*$^{rv/-}$ mutant (Figure 3N), the ureteric bud has not fully invaded the metanephric blastema and had not branched, a defect that is predicted to lead to KA. Additional Cdh1-positive cells in Figure 3N might represent persistent mesonephros or ectopic ureteric buds. In this latter possibility, ectopic buds would explain the duplication of kidney and ureter phenotype as observed in Figure 4.

Since more subtle CAKUT phenotypes like OU or DCS may require a minimum glomerular filtration rate or might be masked in more severe models, we examined a milder model that might be closer to the genetic architecture of humans harboring the 16p11.2 microdeletion syndrome. We therefore analyzed embryos homozygous for the *Tbx6* hypomorphic allele (*Tbx6*$^{rv/rv}$). We generated 25 homozygous *Tbx6*$^{rv/rv}$ animals at E15.5 (n=10), E18.5 (n=10), and P0 (n=5) and observed incomplete penetrance and variable expressivity of multiple phenotypes that are typical of human CAKUT, including mild unilateral or bilateral hypoplasia with asymmetric kidneys, bifid ureter or DCS, and OU with profound hydronephrosis (Table 4 and Figure 4). This pleiotropy of CAKUT phenotypes, ranging from renal parenchyma defects (KA) to hydronephrosis (OU), duplication of ureters (DCS) is highly reminiscent of the pleiotropic manifestations of the 16p11.2 microdeletion in

humans, suggesting that *TBX6* gene dosage is a major determinant for the pathogenesis of CAKUT and the observed variable expressivity of phenotypes in this syndrome.

DISCUSSION

CAKUT has profound impact on child health and alone accounts for about 50% of kidney failure requiring dialysis and transplantation in children[1,3,75–78]. Moreover, CAKUT is often accompanied by extrarenal comorbidities, such as neurodevelopmental and cardiovascular diseases, contributing further to the disease burden in affected children[5,26,27,79]. Understanding the genetic architecture of CAKUT has important implications for the development of therapeutic tools that aim to slow down progression of kidney disease and to mitigate the associated neurocognitive and cardiac disease. In that respect, besides the value in achieving molecular diagnosis and hence solving the etiology of disease. In this respect, it, is relevant to address two features of CAKUT: incomplete penetrance and variable expressivity of disease. While prediction of a targeted treatment that prevents or reverses the disease is still limited by current knowledge, understanding the intrinsic inter-individual mechanisms of compensation or amplification of disease are of critical for devising treatments that might slow down or halt progression of kidney disease and associated extrarenal malformations.

The study of CNVs has provided enormous insight into the genetic architecture of many developmental traits[80–8889]. We conducted a large study on rare CNVs in nearly 3,000 CAKUT cases and over 21,000 controls. By studying the burden of rare CNVs, we identified that KA, OU, and VUR are particularly enriched for large and rare structural variants that affect coding regions of the genome. When examining known GD-CNVs, we found particular enrichment in CAKUT cases with KA, OU, DCS, and PUV when compared to controls. Overall, we identified 45 distinct GDs at 37 independent genomic loci in 4.1% of the CAKUT cases, indicating substantial genetic heterogeneity. We identified novel GD-CNVs in an additional ~2% of CAKUT cases, providing multiple novel susceptibility loci to kidney disease. With respect to mechanism, WES in cases harboring deletions was consistent with haploinsufficiency as the most frequent pathogenic mechanism in individuals affected by GD-associated deletions. Moreover, increased burden of second site CNVs was associated with KA and increased prevalence of extrarenal malformations. These data indicate a role for background genomic burden in variable penetrance and expressivity of disease. CNVs at six of the known 37 loci accounted for ~65% of the cases with a known GD, thus identifying major susceptibility CNVs for CAKUT. At these six loci, deletions were associated with KA, whereas duplications were enriched in cases with ureteric and lower tract defects, such as DCS and PUV. This preliminary observation gives rise to the hypothesis that CAKUT subcategories such as ureteric and lower urinary tract defects (DCS, PUV) may represent, at a molecular level, mirror traits of conditions affecting the upper urinary tract (KA, OU). This “mirroring phenomenon” has been recognized in autism spectrum disorder caused by structural variation at the 16p11.2 in which patients with deletions show macrocephaly and patients with duplications are microcephalic[90,91]. Larger human cohorts and experimental data will be required to investigate this hypothesis. A closer look at common GD loci assigns specificity to CAKUT subcategories, such as deletions, but not duplications, at 4p (Wolf-Hirschhorn syndrome)[92], 17q12 (renal cysts and diabetes

syndrome)[40], and 22q11.2 (DiGeorge syndrome)[93], that were nearly exclusively identified in cases with upper urinary tract malformations (mainly KA). On the other end, microdeletions at 16p11.2 were observed across the whole phenotypic spectrum of CAKUT, underlining the highly pleiotropic effect of this genomic region on human kidney and urinary tract development. The finding that pleiotropic urogenital defects are an important feature of this syndrome was replicated in two independent series from DECIPHER and CHOP.

We previously reported on the potential of large-scale genetic studies coupled with functional modeling in vertebrates in the identification of genetic drivers for kidney phenotypes of microdeletion syndromes[23]. Here we focused on the 16p11.2 microdeletion syndrome, because its pleiotropic effect and incomplete penetrance provide an ideal scenario to identify genetic factors that might be amenable for devising therapeutic intervention. Deletion mapping and prioritization analyses pointed to *TBX6* as the main genetic driver for CAKUT in cases with this syndrome. *Tbx6* mouse allelic series showed that compound heterozygous embryos for a null (*Tbx6*^{tm2Pa}) and a hypomorphic (*Tbx6*^{rv}) allele displayed fully penetrant CAKUT with variable expressivity of phenotypes, from bilateral renal agenesis to hypodysplasia and obstructive uropathy. Analysis of E11.5 embryonic tissue stained for renal progenitors and epithelial markers suggested that the initiating events for the CAKUT phenotypes are likely to occur very early in development in these mice. These data are in line with recent fate mapping studies that show that Tbx6 is expressed in renal progenitor cells prior to commitment to ureteric bud or metanephric mesenchyme lineages[94]. When we analyzed embryos and mice with milder Tbx6 inactivation (*Tbx6*$^{rv/rv}$) as a closer model to the human 16p11.2 microdeletion, we observed incomplete penetrance of multiple kidney and urinary tract malformations. Interestingly, similar to our 16p11.2 microdeletion cases, mutant mice showed phenotypes across the phenotypic spectrum of CAKUT, including KA, OU and DCS. Overall, our mouse data at least partly intimate that the variable expressivity and incomplete penetrance of CAKUT in cases with the 16p11.2 microdeletion is attributable to a fine regulation of *TBX6* gene expression during organogenesis.

In summary, with this study on children and young adults with kidney and urinary tract malformations we provide significant insight into the genomic landscape of human CAKUT. We identify several susceptibility genetic loci and genes, and we highlight *TBX6* as a main genetic driver for CAKUT in cases with the chromosome 16p11.2 microdeletion syndrome, with significant implications for understanding, and potentially modifying, penetrance of disease.

URLs

Annovar: http://annovar.openbioinformatics.org/en/latest/

ATAV: https://redmine.igm.cumc.columbia.edu/projects/atav/wiki

DECIPHER: https://decipher.sanger.ac.uk/

Exome aggregation consortium: https://exac.broadinstitute.org

ISCA: https://www.iscaconsortium.org/

Python Software Foundation: https://www.python.org/

- lifelines: https://github.com/CamDavidsonPilon/lifelines/
- pandas: https://pandas.pydata.org/

R Foundation for Statistical Computing: https://www.R-project.org/

- dplyr package: https://CRAN.R-project.org/package=dplyr
- survival package: https://CRAN.R-project.org/package=survival

Seattle seq: http://snp.gs.washington.edu/SeattleSeqAnnotation138/

UCSC genome browser: https://genome.ucsc.edu

METHODS

Study cohorts and Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

ONLINE METHODS

Study subjects

CAKUT cases consisted 2,824 affected, unrelated individuals with different subphenotypes across the entire CAKUT spectrum. Cases were recruited in the United States, Europe and Brazil (for an overview of recruitment sites, see Supplementary Figure 1). Baseline characteristics of the study cohort are shown in Supplementary Table 1. One in every four cases presented with more than one CAKUT phenotype (i.e. complex CAKUT), whereas a significant proportion of patients had extrarenal manifestations such as neurocognitive defects, congenital heart disease or dysmorphic features. A positive family history for renal disease was identified in 15% of the cases. Cases included CAKUT patients from the CKiD[27] and KIMONO[28] studies; genotyping results of 823 (29%) subjects had been partially reported in previous publications[23,26–28]. The control population consisted of 21,498 individuals recruited as part of genome-wide genotyping studies of complex traits that are not associated to nephropathy or developmental defects (Supplementary Table 2).

CNV discovery and annotation

Genomic DNA was obtained from peripheral blood samples or, in the case of CKiD participants, lymphoblastoid cell lines derived from peripheral blood samples. Genome-wide genotyping was performed in all cases and controls using HumanHap550 or higher density Illumina (San Diego, CA, USA) or Affymetrix SNP6.0 (Santa Clara, CA, USA) microarrays (Supplementary Table 2). Raw data processing and subsequent analyses were performed in the same fashion for both cases and controls to avoid bias.

Raw data was first processed with Affymetrix Power Tools and the PennCNV-Affy protocol or with Illumina GenomeStudio v2011, to obtain probe level logR-ratio and b allele frequency values. Raw intensity data were processed in GenomeStudio v2011 (Illumina).

PennCNV software[95] was used to determine CNV calls. PennCNV and PLINK software[96] were used for quality control. Principal component analysis (PCA) was conducted using smartPCA[97] based on SNP genotypes derived from the same genotyping arrays. CNV calling and all analyses were performed on hg18 coordinates, CNVs were then mapped to hg19 using UCSC liftover software (see URLs). Only high quality CNVs with confidence scores ≥30 were included in the analyses based on independent experimental validation from our prior CNV study on kidney anomalies (KA)[26].

CNVs in cases were compared to those in controls and to known CNV coordinates, and annotated with RefGene (see URLs) and curated sets of genes using custom Perl code[98]. Known CNVs were based on the DECIPHER[37,38] and the International Standards for Cytogenomic Arrays (ISCA) databases[35] and the literature (Supplementary Table 7). Curated sets of genes included those known to be associated with kidney disease and/or development from the OMIM and MGI databases and the literature. Two CNVs were considered to be identical when they had the same copy number value and had a reciprocal overlap equal to or greater than 70%. All reported CNVs were visually inspected using Illumina Genome Viewer 1.9.0 or Affymetrix ChAS to exclude potential artifacts.

CNV classification

CNVs were classified as known genomic disorders (GD-CNV) or likely pathogenic CNVs ("novel GD-CNV"). A CNV was defined as a known GD-CNV when it overlapped at least 70% of a known syndromic CNV. The criteria to define a likely pathogenic CNV were adapted from prior recommendations for interpretation of microarray data[26–28,35]. A CNV was classified as likely pathogenic if it (1) intersected at least one exon, (2) was at least 100 kb in size, (3) had a frequency in controls of at most 0.02%, (4) did not overlap (<70%) with a benign or likely benign CNV in the ISCA database and fulfilled at least one of the following additional criteria: (a) ≥70% overlap with a reported pathogenic or likely pathogenic CNV in the ISCA database, (b) intersected a causative autosomal dominant gene for CAKUT in humans or mice and/or (c) is the reciprocal of a known GD-CNV (coordinates with ≥70% overlap). Known GD-CNV and likely pathogenic taken together were termed as 'diagnostic CNVs'. CNVs that did not meet the criteria for diagnostic CNVs were defined as variants of unknown significance, if they were at least 100 kb in size and had a frequency of at most 0.1% in controls, including homo-/hemizygous deletions of loci that underlie a recessive disorder for human disease with a renal phenotype.

CNV burden analysis

We restricted burden analyses to autosomal CNV of size ≥100 kb in cases and controls, with a frequency ≤1% in the whole control dataset and in any one control cohort comprising it. CNVs were further filtered at a frequency ≤0.1% in control population subgroups based on PCA, to avoid including CNVs that might be relatively common within ancestry groups represented in controls. Prior to determining the CNV burden, seven cases (5 KA, 1 vesicoureteral reflux and 1 lower urinary tract malformations) and 8 controls were removed because they were unmatched outliers in PCA, yielding a dataset of 2,817 cases and 21,490 controls for burden analyses.

Whole exome sequencing

Whole exome sequencing was performed on 23 CAKUT patients harboring 14 distinct known or novel GD-CNVs at the New York Genome Center (NYGC). Briefly, for each capture experiment, 1 μg of genomic DNA was fragmented, linkers were ligated to the ends and a library was prepared. Next, genomic DNA was annealed to Agilent V4 capture probes, and bound genomic DNA was eluted and subjected to Next-Generation sequencing performed on an Illumina HiSeq 2500 machine. Sequence reads were converted to FASTQ format and mapped to the reference genome. We used uniform procedures for variant calling to prevent technical bias. Samples were processed using a consistent alignment and variant calling pipeline, consisting of primary alignment using bwa-0.5.10, duplicate removal using Picard tools, index realignment and variant calling using GATK 3.6, and variant annotation using snpEff-3.3, AnnoVar (see URLs), and SeattleSeq (see URLs), with Ensembl-GRCh37.73 annotations.

After variant calling, resultant calls and their underlying quality statistics were then stored in a database of variants (AnnoDB) that is used by Analysis Tool for Annotated Variants (ATAV) analyses (see URLs). Next, we queried ATAV for the individual deleted regions of all 23 CAKUT cases and annotated all hemizygous variants (minimal coverage >8X, minimal genotype quality score >30) with a minor allele frequency of <1% (autosomal recessive model) in the Exome Aggregation Consortium database (ExAC) and in 3,653 ethnically-matched controls available in-house controls, and with a Combined Annotation Dependent Depletion (CADD) score ≥20 (i.e. strongly predicted to be deleterious). We retained one homozygous variant in one individual CAKUT case: a homozygous truncating mutation in *EFCAB12* found in a case with a 14.9 Mb deletion at chromosome 3q13.22–13. The variant was confirmed by Sanger sequencing according to standard protocols. Finally, we queried the ExAC database as well as the data warehouse of the Columbia Institute of Genomic Medicine (IGM), which contains whole exome data for 15,469 control individuals that are recruited for other reasons than chronic kidney disease. We did not find any homozygous or compound heterozygous truncating variants in *EFCAB12*. We performed clinical annotation of genes that are known to be implicated in Mendelian forms of CAKUT, by querying WES data for an in-house gene list as previously described[58]. We did not find any pathogenic variants in known CAKUT genes for all deletion carrier patients

Generation and analysis of *Tbx6* mutant mice and embryos

The null allele null expression reporter allele, $Tbx6^{tm2Pa}$, which deletes exon 2, part of exon 3 and has an *H2B-EYFP* fusion gene inserted in frame into exon 1[70], was maintained on a mixed genetic background of 129 and ICR (Taconic). B6L-$Tbx6^{rv}$/J mice (JAX) were mated with C57BL/6Tac mice (Taconic) and maintained in a small closed colony. Embryos were collected from timed matings of mice heterozygous for either allele, with noon on the day of the plug considered embryonic day (E) 0.5. Embryos were genotyped by PCR using the following primers pairs for $Tbx6^{tm2Pa}$ (1) 5'-GTACCATCCACGAGAGTTGTAC-3', (2) 5'-GGGAAGAATGAGGATCCAGG-3' to obtain a 220bp wild type allele fragment; and (3) 5'-ATTGCACGCAGGTTCTCCGG-3'; (4) 5'-GTCACGACGAGATCCTCGCC-3' to obtain a 550bp mutant allele fragment. For $Tbx6^{rv}$ (1) 5'-CTCGCAGCTTCACTAGTCC-3', (2) 5'-

GTGTCTGGCGTATCAGCTCA-3' to obtain a 322bp wild type allele fragment and 504bp mutant allele fragment.

We analyzed different allele combinations for the *Tbx6* null (*Tbx6*^{tm2Pa}) and hypomorphic (*Tbx6*^{V}) alleles. Hematoxylin and eosin (H&E) staining: embryos were fixed overnight in 4% PFA and dehydrated through a graded ethanol series (25–50-75–100%) followed by Xylene washes. The embryos were incubated in 2 changes of paraffin wax at 55°C under vacuum. The embryos were then embedded in paraffin wax and sectioned at 10μm. Sections were dewaxed in Xylene, rehydrated through ethanol, stained with Harris haematoxylin and Eosin-Y, and finally dehydrated before mounting in Cytoseal. Immunostaining: embryos were fixed in 4% PFA. Tissues were dehydrated into 100% methanol post-fixation. For sectioning, tissues were embedded in paraffin or frozen in OCT and sectioned at various thicknesses (5–20μm). Paraffin sections were deparaffinised with Histoclear and rehydrated via an ethanol-series. Cryosections were prepared for immunolabelling by removing OCT by washing in PBS. Antibodies: Cdh1: Goat R&D AF741 at 1:400 dilution. Pax2: Rabbit Zymed 716000 1:70.

Statistical analysis

All burden metrics calculations, statistical tests and plots were performed with R v3.1 software (R Core Team 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, see URLs) and its dplyr (Hadley Wickham and Romain Francois (2016). dplyr: A Grammar of Data Manipulation, see URLs), survival (Therneau T (2015). A Package for Survival Analysis in S. version 2.38, and ggplot2[99] packages, and the Python v2.7 language (see URLs) and its package pandas[100] and lifelines (see URLs). Unadjusted P-values were reported for all tests. Summary statistics (mean, standard deviation, median, interquartile range) were calculated for CNV metrics such as size and total span, and non-parametric Wilcoxon test was used for comparison between cases and controls. Proportions were compared using 2-sided Fisher's Exact test. Kaplan-Meier survival curves were calculated to represent the largest CNV per genome and non-parametric log-rank test was used to compare the survival distributions between cases and controls. Logistic regression was used to test if the number of genes intersected by large and rare CNVs detected (quantitative predictor) was associated with the case-control status (outcome). The reciprocal cumulative distribution of the number of intersected genes in cases versus controls were plotted.

Data Availability Statement

Raw data that support the findings of these study will in part be available from the corresponding authors upon reasonable request and in part available from dbGaP (https://www.ncbi.nlm.nih.gov/gap; accession pending). Some restrictions may apply according to participants' consent and privacy protection. All images generated from mouse experiments reported in this study will also be available from the corresponding authors upon reasonable request.

Code Availability Statement

Custom Perl code used for CNV comparison and annotation as well as R and Python code used in burden analyses are available upon request from the corresponding authors.

Ethical Statement

All aspects of the study involving human research participants adhered to the principles of the Declaration of Helsinki and the study protocol was approved by the Institutional Review Boards (IRB) of Columbia University Medical Center and each participating recruitment site. Signed written informed consent by the participant and/or their parents or guardians was obtained according to the protocol of local IRBs.

All animal experiments followed protocols approved by the Institutional Animal Care and Use Committee at Columbia University.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Miguel Verbitsky[1,*], Rik Westland[1,2,*], Alejandra Perez[1], Krzysztof Kiryluk[1], Qingxue Liu[1], Priya Krithivasan[1], Adele Mitrotti[1], David A. Fasel[1], Ekaterina Batourina[3], Matthew G. Sampson[4], Monica Bodria[5], Max Werth[1], Charlly Kao[6], Jeremiah Martino[1], Valentina P. Capone[1], Asaf Vivante[7,8], Shirlee Shril[7], Byum Hee Kil[1], Maddalena Marasa[1], Jun Y. Zhang[1], Young-Ji Na[1], Tze Y. Lim[1], Dina Ahram[1], Patricia L. Weng[9], Erin L. Heinzen[10], Alba Carrea[5], Giorgio Piaggio[5], Loreto Gesualdo[11], Valeria Manca[12], Giuseppe Masnata[12], Maddalena Gigante[11], Daniele Cusi[13], Claudia Izzi[14], Francesco Scolari[15], Joanna A.E. van Wijk[2], Marijan Saraga[16,17], Domenico Santoro[18], Giovanni Conti[19], Pasquale Zamboli[20], Hope White[1], Dorota Drozdz[21], Katarzyna Zachwieja[21], Monika Miklaszewska[22], Marcin Tkaczyk[23], Daria Tomczyk[23], Anna Krakowska[23], Przemyslaw Sikora[24], Tomasz Jarmoliński[25], Maria K. Borszewska-Kornacka[26], Robert Pawluch[26], Maria Szczepanska[26], Piotr Adamczyk[26], Malgorzata Mizerska-Wasiak[27], Grazyna Krzemien[27], Agnieszka Szmigielska[27], Marcin Zaniew[28], Mark G. Dobson[29,30], John M. Darlow[29,30], Prem Puri[30,31], David E. Barton[29,32], Susan L. Furth[33], Bradley A. Warady[34], Zoran Gucev[35], Vladimir J. Lozanovski[35,36], Velibor Tasic[35], Isabella Pisani[37], Landino Allegri[37], Lida M. Rodas[38], Josep M. Campistol[38], Cécile Jeanpierre[39], Shumyle Alam[40], Pasquale Casale[40,41], Craig S. Wong[42], Fangming Lin[43], Débora M. Miranda[44], Eduardo A. Oliveira[44], Ana Cristina Simões-e-Silva[44], Jonathan M. Barasch[1], Brynn Levy[45], Nan Wu[46,47], Friedhelm Hildebrandt[7], Gian Marco Ghiggeri[5], Anna Latos-Bielenska[48], Anna Materna-Kiryluk[48], Feng Zhang[49], Hakon Hakonarson[6], Virginia E. Papaioannou[50,$], Cathy L. Mendelsohn[3,$], Ali G. Gharavi[1,$], Simone Sanna-Cherchi[1,$]

Affiliations

[1]Division of Nephrology, Department of Medicine, Columbia University, New York, New York, USA. [2]Department of Pediatric Nephrology, VU University Medical Center, Amsterdam, The Netherlands. [3]Department of Urology, Columbia University College of Physicians and Surgeons, New York, New York, USA. [4]University of Michigan School of Medicine, Department of Pediatrics-Nephrology, Ann Arbor, Michigan, USA. [5]Division of Nephrology, Dialysis, Transplantation, and Laboratory on Pathophysiology of Uremia, Istituto G. Gaslini, Genoa, Italy. [6]Center for Applied Genomics, The Children's Hospital of Philadelphia and Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania, USA. [7]Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [8]Pediatric Department B and Pediatric Nephrology Unit, Edmond and Lily Safra Children's Hospital, Chaim Sheba Medical Center, Tel Hashomer and the Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [9]Department of Pediatric Nephrology, UCLA Medical Center and UCLA Medical Center-Santa Monica, Los Angeles, California, USA. [10]Institute for Genomic Medicine, Columbia University Medical Center, New York, New York, USA. [11]Section of Nephrology, Department of Emergency and Organ Transplantation, University of Bari, Bari, Italy. [12]Department of Pediatric Urology, Azienda Ospedaliera Brotzu, Cagliari, Italy. [13]National Research Council of Italy, Inst. Biomedical Technologies Milano Bio4dreams Scientific Unit, Milano, Italy. [14]Dipartimento Ostetrico-Ginecologico e Seconda Divisione di Nefrologia ASST, Spedali Civili e Presidio di Montichiari, Brescia, Italy. [15]Cattedra di Nefrologia, Università di Brescia, Seconda Divisione di Nefrologia, Azienda Ospedaliera Spedali Civili di Brescia Presidio di Montichiari, Brescia, Italy. [16]Department of Pediatrics, University Hospital of Split, Split, Croatia. [17]School of Medicine, University of Split, 21000, Split, Croatia. [18]Dipartimento di Medicina Clinica e Sperimentale, Università degli Studi di Messina, Messina, Italy. [19]Department of Pediatric Nephrology, Azienda Ospedaliera Universitaria "G. Martino", Messina, Italy. [20]Division of Nephrology, University of Campania "Luigi Vanvitelli", Naples, Italy. [21]Department of Pediatric Nephrology and Hypertension, Dialysis Unit, Jagiellonian University Medical College, Krakow, Poland [22]Department of Pediatric Nephrology, Jagiellonian University Medical College, Krakow, Poland. [23]Department of Pediatrics, Immunology and Nephrology, Polish Mother's Memorial Hospital Research Institute, Lodz, Poland. [24]Department of Pediatric Nephrology Medical University of Lublin, Lublin, Poland. [25]Children's Department, Międzyrzecki Hospital, Międzyrzecz, Poland [26]Department of Pediatrics, School of Medicine with the Division of Dentistry in Zabrze, Medical University of Silesia in Katowice, Poland. [27]Department of Pediatrics and Nephrology, Medical University of Warsaw, Poland. [28]Department of Pediatrics, University of Zielona Góra, Zielona Góra, Poland. [29]Department of Clinical Genetics, Our Lady's Children's Hospital Crumlin, Dublin, 12, Ireland [30]National Children's Research Centre, Our Lady's Children's Hospital Crumlin, Dublin, 12, Ireland. [31]National Children's Hospital Tallaght, Dublin, 24, Ireland. [32]University College Dublin UCD School of Medicine, Our Lady's Children's Hospital

Crumlin, Dublin, 12, Ireland. [33]Departments of Pediatrics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Division of Nephrology, Children's Hospital of Philadelphia (CHOP), Philadelphia, Pennsylvania, USA. [34]Department of Pediatrics, University of Missouri-Kansas City School of Medicine, Division of Nephrology, Children's Mercy Kansas City, Kansas City, Missouri, USA [35]University Children's Hospital, Medical Faculty of Skopje, Skopje, Macedonia. [36]University Clinic for General, Visceral and Transplantation Surgery, University of Heidelberg, Im Neuenheimer Feld 110, D-69120 Heidelberg, Germany [37]Department of Medicine and Surgery, University of Parma, Italy. [38]ICNU-Nephrology and Urology Department, Barcelona, Spain. [39]Laboratory of Hereditary Kidney Diseases, Inserm UMR 1163, Imagine Institute, Paris Descartes-Sorbonne Paris Cité University, Paris, France. [40]Department of Pediatric Urology, Columbia University College of Physicians and Surgeons, New York, New York, USA. [41]Mount Sinai Medical Center, Kravis Children's Hospital, New York, New York, USA. [42]Division of Pediatric Nephrology, University of New Mexico Children's Hospital, Albuquerque, New Mexico, USA. [43]Division of Pediatric Nephrology, Department of Pediatrics, Columbia University, New York, New York, USA. [44]Department of Pediatrics, Unit of Pediatric Nephrology, Interdisciplinary Laboratory of Medical Investigation, Faculty of Medicine, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil. [45]Department of Pathology and Cell Biology, Columbia University, New York, New York, USA. [46]Department of Orthopedic Surgery, Beijing Key Laboratory for Genetic Research of Skeletal Deformity, Medical Research Center of Orthopedics, all at Peking Union Medical College Hospital, Peking Union Medical College & Chinese Academy of Medical Sciences, Beijing, China. [47]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA. [48]Department of Medical Genetics, Poznan University of Medical Sciences, and NZOZ Center for Medical Genetics GENESIS, Poznan, Poland. [49]Obstetrics and Gynecology Hospital, Fudan University, Shanghai, China. [50]Department of Genetics and Development, Columbia University Medical Center, New York, New York, USA.

ACKNOWLEDGEMENTS

We thank all patients and family members for participating to this study. We thank Dr. James R. Lupski for critical review of this manuscript. This work was supported by grants (1R01DK103184, 1R21DK098531, and UL1 TR000040, to Dr. Sanna-Cherchi; 2R01DK080099, to Dr. Gharavi; 3U54DK104309, to Drs. Gharavi and Barasch; R37HD033082, to Dr. Papaioannou; and 1R01DK105124 to Dr. Kiryluk) from the National Institutes of Health (NIH); a grant-in-aid (13GRNT14680075, to Dr. Sanna-Cherchi) from the American Heart Association; a grant (RF-2010–2307403, to Drs. Sanna-Cherchi and Ghiggeri) from the Joint Italian Ministry of Health and NIH Young Investigators Finalized Research; a grant (to Dr. Ghiggeri) from the Fondazione Malattie Renali nel Bambino; grants to Dr Barton and Mr Puri from the National Children's Research Centre and the Irish Health Research Board (HRA-POR-2014–693); a grant (AAE07007KSA, to Dr. Jeanpierre) from the GIS-Institut des Maladies Rares; by the Polish Ministry of Health (to Drs. Materna-Kiryluk and Latos-Bielenska); by the Polish Kidney Genetics Network (POLYGENES), the Polish Registry of Congenital Malformations (PRCM), and the NZOZ Center for Medical Genetics (GENESIS); by grants (to the Chronic Kidney Disease in Children Study) from the National Institute of Diabetes and Digestive and Kidney Diseases and the Eunice Kennedy Shriver National Institute of Child Health and Human Development; by grants (U01DK66143, U01DK66174, U01DK082194, U01DK66116, and RO1DK082394) from the National Heart, Lung, and Blood Institute; and by the Paul Marks Scholar Award (to Dr. Sanna-Cherchi); a CNPq Grant #460334/2014–0 and a FAPEMIG Grant #PPM-005555–15 (to Drs Miranda, Oliveira and Simões-e-Silva); and a Kolff Postdoc Fellowship Abroad grant (15OKK95, to Dr. Westland) from the Dutch Kidney Foundation. Dr. Sanna-Cherchi is the Florence Irving Assistant Professor of Medicine at Columbia

University. We thank David B. Goldstein for providing infrastructure for whole-exome sequencing at the Institute for Genomic Medicine (IGM) at Columbia University, and for his critical review of the manuscript. Acknowledgments to the investigators who contributed of whole-exome sequencing data for 15,469 controls from the IGM warehouse can be found in the Supplementary Appendix.

REFERENCES

1. Wuhl E et al. Timing and outcome of renal replacement therapy in patients with congenital malformations of the kidney and urinary tract. Clin J Am Soc Nephrol 8, 67–74 (2013). [PubMed: 23085722]
2. Chesnaye NC et al. Mortality risk disparities in children receiving chronic renal replacement therapy for the treatment of end-stage renal disease across Europe: an ESPN-ERA/EDTA registry analysis. Lancet 389, 2128–2137 (2017). [PubMed: 28336050]
3. Sanna-Cherchi S et al. Renal outcome in patients with congenital anomalies of the kidney and urinary tract. Kidney Int 76, 528–33 (2009). [PubMed: 19536081]
4. Goodyer PR Renal dysplasia/hypoplasia. in Pediatric nephrology, 5th ed (eds. Avner ED, Harmon WE & Niaudet P) 83–91 (Lippincott Williams & Wilkins, Philadelphia 2004).
5. Sanna-Cherchi S, Westland R, Ghiggeri GM & Gharavi AG Genetic basis of human congenital anomalies of the kidney and urinary tract. J Clin Invest 128, 4–15 (2018). [PubMed: 29293093]
6. Schedl A Renal abnormalities and their developmental origin. Nat Rev Genet 8, 791–802 (2007). [PubMed: 17878895]
7. Costantini F & Kopan R Patterning a complex organ: branching morphogenesis and nephron segmentation in kidney development. Dev Cell 18, 698–712 (2010). [PubMed: 20493806]
8. Short KM & Smyth IM The contribution of branching morphogenesis to kidney development and disease. Nat Rev Nephrol 12, 754–767 (2016). [PubMed: 27818506]
9. dos Santos Junior AC, de Miranda DM & Simoes e Silva AC Congenital anomalies of the kidney and urinary tract: an embryogenetic review. Birth Defects Res C Embryo Today 102, 374–81 (2014). [PubMed: 25420794]
10. Chen F Genetic and developmental basis for urinary tract obstruction. Pediatr Nephrol 24, 1621–32 (2009). [PubMed: 19085015]
11. Vainio S & Lin Y Coordinating early kidney development: lessons from gene targeting. Nat Rev Genet 3, 533–43 (2002). [PubMed: 12094231]
12. Sanna-Cherchi S et al. Genetic approaches to human renal agenesis/hypoplasia and dysplasia. Pediatr Nephrol 22, 1675–84 (2007). [PubMed: 17437132]
13. Nicolaou N, Renkema KY, Bongers EM, Giles RH & Knoers NV Genetic, environmental, and epigenetic factors involved in CAKUT. Nat Rev Nephrol 11, 720–31 (2015). [PubMed: 26281895]
14. Georgas KM et al. An illustrated anatomical ontology of the developing mouse lower urogenital tract. Development 142, 1893–908 (2015). [PubMed: 25968320]
15. Thomas R et al. HNF1B and PAX2 mutations are a common cause of renal hypodysplasia in the CKiD cohort. Pediatr Nephrol 26, 897–903 (2011). [PubMed: 21380624]
16. Weber S et al. Prevalence of mutations in renal developmental genes in children with renal hypodysplasia: results of the ESCAPE study. J Am Soc Nephrol 17, 2864–70 (2006). [PubMed: 16971658]
17. Nicolaou N et al. Prioritization and burden analysis of rare variants in 208 candidate genes suggest they do not play a major role in CAKUT. Kidney Int 89, 476–86 (2016). [PubMed: 26489027]
18. Sanna-Cherchi S et al. Mutations in DSTYK and dominant urinary tract malformations. N Engl J Med 369, 621–9 (2013). [PubMed: 23862974]
19. Vivante A et al. Mutations in TBX18 Cause Dominant Urinary Tract Malformations via Transcriptional Dysregulation of Ureter Development. Am J Hum Genet 97, 291–301 (2015). [PubMed: 26235987]
20. Schimmenti LA et al. Further delineation of renal-coloboma syndrome in patients with extreme variability of phenotype and identical PAX2 mutations. Am J Hum Genet 60, 869–78 (1997). [PubMed: 9106533]

21. Bekheirnia MR et al. 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. Genet Med 19, 412–420 (2017). [PubMed: 27657687]
22. Ichikawa I, Kuwayama F, Pope J.C.t., Stephens FD & Miyazaki Y Paradigm shift from classic anatomic theories to contemporary cell biological views of CAKUT. Kidney Int 61, 889–98 (2002). [PubMed: 11849443]
23. Lopez-Rivera E et al. Genetic Drivers of Kidney Defects in the DiGeorge Syndrome. N Engl J Med 376, 742–754 (2017). [PubMed: 28121514]
24. Hwang DY et al. Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. Kidney Int 85, 1429–33 (2014). [PubMed: 24429398]
25. Vivante A, Kohl S, Hwang DY, Dworschak GC & Hildebrandt F Single-gene causes of congenital anomalies of the kidney and urinary tract (CAKUT) in humans. Pediatr Nephrol 29, 695–704 (2014). [PubMed: 24398540]
26. Sanna-Cherchi S et al. Copy-number disorders are a common cause of congenital kidney malformations. Am J Hum Genet 91, 987–97 (2012). [PubMed: 23159250]
27. Verbitsky M et al. Genomic imbalances in pediatric patients with chronic kidney disease. J Clin Invest 125, 2171–8 (2015). [PubMed: 25893603]
28. Westland R et al. Copy number variation analysis identifies novel CAKUT candidate genes in children with a solitary functioning kidney. Kidney Int 88, 1402–1410 (2015). [PubMed: 26352300]
29. Ulinski T et al. Renal phenotypes related to hepatocyte nuclear factor-1beta (TCF2) mutations in a pediatric cohort. J Am Soc Nephrol 17, 497–503 (2006). [PubMed: 16371430]
30. Madariaga L et al. Severe prenatal renal anomalies associated with mutations in HNF1B or PAX2 genes. Clin J Am Soc Nephrol 8, 1179–87 (2013). [PubMed: 23539225]
31. Hoskins BE et al. Missense mutations in EYA1 and TCF2 are a rare cause of urinary tract malformations. Nephrol Dial Transplant 23, 777–9 (2008).
32. Heidet L et al. Spectrum of HNF1B mutations in a large cohort of patients who harbor renal diseases. Clin J Am Soc Nephrol 5, 1079–90 (2010). [PubMed: 20378641]
33. Carvalho CM & Lupski JR Mechanisms underlying structural variant formation in genomic disorders. Nat Rev Genet 17, 224–38 (2016). [PubMed: 26924765]
34. Lupski JR Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. Trends Genet 14, 417–22 (1998). [PubMed: 9820031]
35. Miller DT et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. Am J Hum Genet 86, 749–64 (2010). [PubMed: 20466091]
36. South ST et al. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. Genet Med 15, 901–9 (2013). [PubMed: 24071793]
37. Swaminathan GJ et al. DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. Hum Mol Genet 21, R37–44 (2012). [PubMed: 22962312]
38. Firth HV et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 84, 524–33 (2009). [PubMed: 19344873]
39. Grisaru S, Ramage IJ & Rosenblum ND Vesicoureteric reflux associated with renal dysplasia in the Wolf-Hirschhorn syndrome. Pediatr Nephrol 14, 146–8 (2000). [PubMed: 10684366]
40. Mefford HC et al. Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. Am J Hum Genet 81, 1057–69 (2007). [PubMed: 17924346]
41. Portnoi MF Microduplication 22q11.2: a new chromosomal syndrome. Eur J Med Genet 52, 88–93 (2009). [PubMed: 19254783]
42. Kobrynski LJ & Sullivan KE Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. Lancet 370, 1443–52 (2007). [PubMed: 17950858]

43. Weber S et al. Mapping candidate regions and genes for congenital anomalies of the kidneys and urinary tract (CAKUT) by array-based comparative genomic hybridization. Nephrol Dial Transplant 26, 136–43 (2011). [PubMed: 20605837]
44. Hildebrandt F et al. A novel gene encoding an SH3 domain protein is mutated in nephronophthisis type 1. Nat Genet 17, 149–53 (1997). [PubMed: 9326933]
45. Willatt L et al. 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. Am J Hum Genet 77, 154–60 (2005). [PubMed: 15918153]
46. Mattina T, Perrotta CS & Grossfeld P Jacobsen syndrome. Orphanet J Rare Dis 4, 9 (2009). [PubMed: 19267933]
47. Sampson MG et al. Evidence for a recurrent microdeletion at chromosome 16p11.2 associated with congenital anomalies of the kidney and urinary tract (CAKUT) and Hirschsprung disease. Am J Med Genet A 152A, 2618–22 (2010). [PubMed: 20799338]
48. Fontes MI et al. Genotype-phenotype correlation of 16p13.3 terminal duplication and 22q13.33 deletion: Natural history of a patient and review of the literature. Am J Med Genet A 170, 766–72 (2016). [PubMed: 26638882]
49. Goh ES et al. Definition of a critical genetic interval related to kidney abnormalities in the Potocki-Lupski syndrome. Am J Med Genet A 158A, 1579–88 (2012). [PubMed: 22639462]
50. Yamamoto T et al. A large interstitial deletion of 17p13.1p11.2 involving the Smith-Magenis chromosome region in a girl with multiple congenital anomalies. Am J Med Genet A 140, 88–91 (2006). [PubMed: 16333830]
51. Kim YM et al. Phelan-McDermid syndrome presenting with developmental delays and facial dysmorphisms. Korean J Pediatr 59, S25–S28 (2016). [PubMed: 28018439]
52. Ozgun MT et al. Prenatal diagnosis of a fetus with partial trisomy 7p. Fetal Diagn Ther 22, 229–32 (2007). [PubMed: 17228165]
53. Trachoo O, Assanatham M, Jinawath N & Nongnuch A Chromosome 20p inverted duplication deletion identified in a Thai female adult with mental retardation, obesity, chronic kidney disease and characteristic facial features. Eur J Med Genet 56, 319–24 (2013). [PubMed: 23542666]
54. Westland R et al. Phenotypic expansion of DGKE-associated diseases. J Am Soc Nephrol 25, 1408–14 (2014). [PubMed: 24511134]
55. Karczewski KJ et al. The ExAC browser: displaying reference data information from over 60 000 exomes. Nucleic Acids Res 45, D840–D845 (2017). [PubMed: 27899611]
56. Lek M et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536, 285–91 (2016). [PubMed: 27535533]
57. Materna-Kiryluk A et al. The emerging role of genomics in the diagnosis and workup of congenital urinary tract defects: a novel deletion syndrome on chromosome 3q13.31–22.1. Pediatr Nephrol 29, 257–67 (2014). [PubMed: 24292865]
58. Lata S et al. Whole-Exome Sequencing in Adults With Chronic Kidney Disease: A Pilot Study. Ann Intern Med 168, 100–109 (2018). [PubMed: 29204651]
59. Wu N et al. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. N Engl J Med 372, 341–50 (2015). [PubMed: 25564734]
60. Al-Kateb H et al. Scoliosis and vertebral anomalies: additional abnormal phenotypes associated with chromosome 16p11.2 rearrangement. Am J Med Genet A 164A, 1118–26 (2014). [PubMed: 24458548]
61. Dickinson ME et al. High-throughput discovery of novel developmental phenotypes. Nature 537, 508–514 (2016). [PubMed: 27626380]
62. Chapman DL & Papaioannou VE Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. Nature 391, 695–7 (1998). [PubMed: 9490412]
63. Chapman DL, Agulnik I, Hancock S, Silver LM & Papaioannou VE Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. Dev Biol 180, 534–42 (1996). [PubMed: 8954725]
64. Abe K et al. Novel ENU-Induced Mutation in Tbx6 Causes Dominant Spondylocostal Dysostosis-Like Vertebral Malformations in the Rat. PLoS One 10, e0130231 (2015).

65. Lefebvre M et al. Autosomal recessive variations of TBX6, from congenital scoliosis to spondylocostal dysostosis. Clin Genet 91, 908–912 (2017). [PubMed: 27861764]
66. Sparrow DB et al. Autosomal dominant spondylocostal dysostosis is caused by mutation in TBX6. Hum Mol Genet 22, 1625–31 (2013). [PubMed: 23335591]
67. MacEwen GD, Winter RB, Hardy JH & Sherk HH Evaluation of kidney anomalies in congenital scoliosis. 1972. Clin Orthop Relat Res, 4–7 (2005).
68. Cowell HR, MacEwen GD & Hubben C Incidence of abnormalities of the kidney and ureter in congenital scoliosis. Birth Defects Orig Artic Ser 10, 142–5 (1974). [PubMed: 4470887]
69. MacEwen GD, Winter RB & Hardy JH Evaluation of kidney anomalies in congenital scoliosis. J Bone Joint Surg Am 54, 1451–4 (1972). [PubMed: 4653629]
70. Hadjantonakis AK, Pisano E & Papaioannou VE Tbx6 regulates left/right patterning in mouse embryos through effects on nodal cilia and perinodal signaling. PLoS One 3, e2511 (2008).
71. Watabe-Rudolph M, Schlautmann N, Papaioannou VE & Gossler A The mouse rib-vertebrae mutation is a hypomorphic Tbx6 allele. Mech Dev 119, 251–6 (2002). [PubMed: 12464437]
72. Batourina E et al. Distal ureter morphogenesis depends on epithelial cell remodeling mediated by vitamin A and Ret. Nat Genet 32, 109–15 (2002). [PubMed: 12195422]
73. Batourina E et al. Vitamin A controls epithelial/mesenchymal interactions through Ret expression. Nat Genet 27, 74–8 (2001). [PubMed: 11138002]
74. Batourina E et al. Apoptosis induced by vitamin A signaling is crucial for connecting the ureters to the bladder. Nat Genet 37, 1082–9 (2005). [PubMed: 16186816]
75. Harambat J, van Stralen KJ, Kim JJ & Tizard EJ Epidemiology of chronic kidney disease in children. Pediatr Nephrol 27, 363–73 (2012). [PubMed: 21713524]
76. Westland R, Kurvers RA, van Wijk JA & Schreuder MF Risk factors for renal injury in children with a solitary functioning kidney. Pediatrics 131, e478–85 (2013). [PubMed: 23319536]
77. Westland R, Schreuder MF, Bokenkamp A, Spreeuwenberg MD & van Wijk JA Renal injury in children with a solitary functioning kidney--the KIMONO study. Nephrol Dial Transplant 26, 1533–41 (2011). [PubMed: 21427076]
78. Westland R, Schreuder MF, van Goudoever JB, Sanna-Cherchi S & van Wijk JA Clinical implications of the solitary functioning kidney. Clin J Am Soc Nephrol 9, 978–86 (2014). [PubMed: 24370773]
79. Verbitsky M et al. Genomic Disorders and Neurocognitive Impairment in Pediatric CKD. J Am Soc Nephrol 28, 2303–2309 (2017). [PubMed: 28348065]
80. Cooper GM et al. A copy number variation morbidity map of developmental delay. Nat Genet 43, 838–46. [PubMed: 21841781]
81. Greenway SC et al. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. Nat Genet 41, 931–5 (2009). [PubMed: 19597493]
82. Osoegawa K et al. Identification of novel candidate genes associated with cleft lip and palate using array comparative genomic hybridisation. J Med Genet 45, 81–6 (2008). [PubMed: 17873121]
83. Serra-Juhe C et al. Contribution of rare copy number variants to isolated human malformations. PLoS One 7, e45530 (2012).
84. Brunetti-Pierri N et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. Nat Genet 40, 1466–71 (2008). [PubMed: 19029900]
85. Sanders SJ et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70, 863–85 (2011). [PubMed: 21658581]
86. Sebat J et al. Strong association of de novo copy number mutations with autism. Science 316, 445–9 (2007). [PubMed: 17363630]
87. Stefansson H et al. Large recurrent microdeletions associated with schizophrenia. Nature 455, 232–6 (2008). [PubMed: 18668039]
88. Yu L et al. De novo copy number variants are associated with congenital diaphragmatic hernia. J Med Genet 49, 650–9 (2012). [PubMed: 23054247]

89. Mannik K et al. Copy number variations and cognitive phenotypes in unselected populations. JAMA 313, 2044–54 (2015). [PubMed: 26010633]

90. Golzio C & Katsanis N Genetic architecture of reciprocal CNVs. Curr Opin Genet Dev 23, 240–8 (2013). [PubMed: 23747035]

91. Golzio C et al. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. Nature 485, 363–7 (2012). [PubMed: 22596160]

92. Gandelman KY, Gibson L, Meyn MS & Yang-Feng TL Molecular definition of the smallest region of deletion overlap in the Wolf-Hirschhorn syndrome. Am J Hum Genet 51, 571–8 (1992). [PubMed: 1379774]

93. Driscoll DA, Budarf ML & Emanuel BS A genetic etiology for DiGeorge syndrome: consistent deletions and microdeletions of 22q11. Am J Hum Genet 50, 924–33 (1992). [PubMed: 1349199]

94. Concepcion D et al. Cell lineage of timed cohorts of Tbx6-expressing cells in wild-type and Tbx6 mutant embryos. Biol Open 6, 1065–1073 (2017). [PubMed: 28606934]

METHODS ONLY REFERENCES

95. Wang K et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 17, 1665–74 (2007). [PubMed: 17921354]

96. Purcell S et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81, 559–75 (2007). [PubMed: 17701901]

97. Price AL et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904–9 (2006). [PubMed: 16862161]

98. Fasel D, Verbitsky M & Sanna-Cherchi S CNVkit – Software Tools for Analyzing Genomic Structural Variants. J Am Soc Nephrol 26(2015).

99. Wickham H ggplot2: Elegant Graphics for Data Analysis, 213 (Springer-Verlag, 2009).

100. McKinney W. van der Walt S. Millman J. Data Structures for Statistical Computing in Python; Proceedings of the 9th Python in Science Conference; Austin, Texas. .: 2010. 51–56.

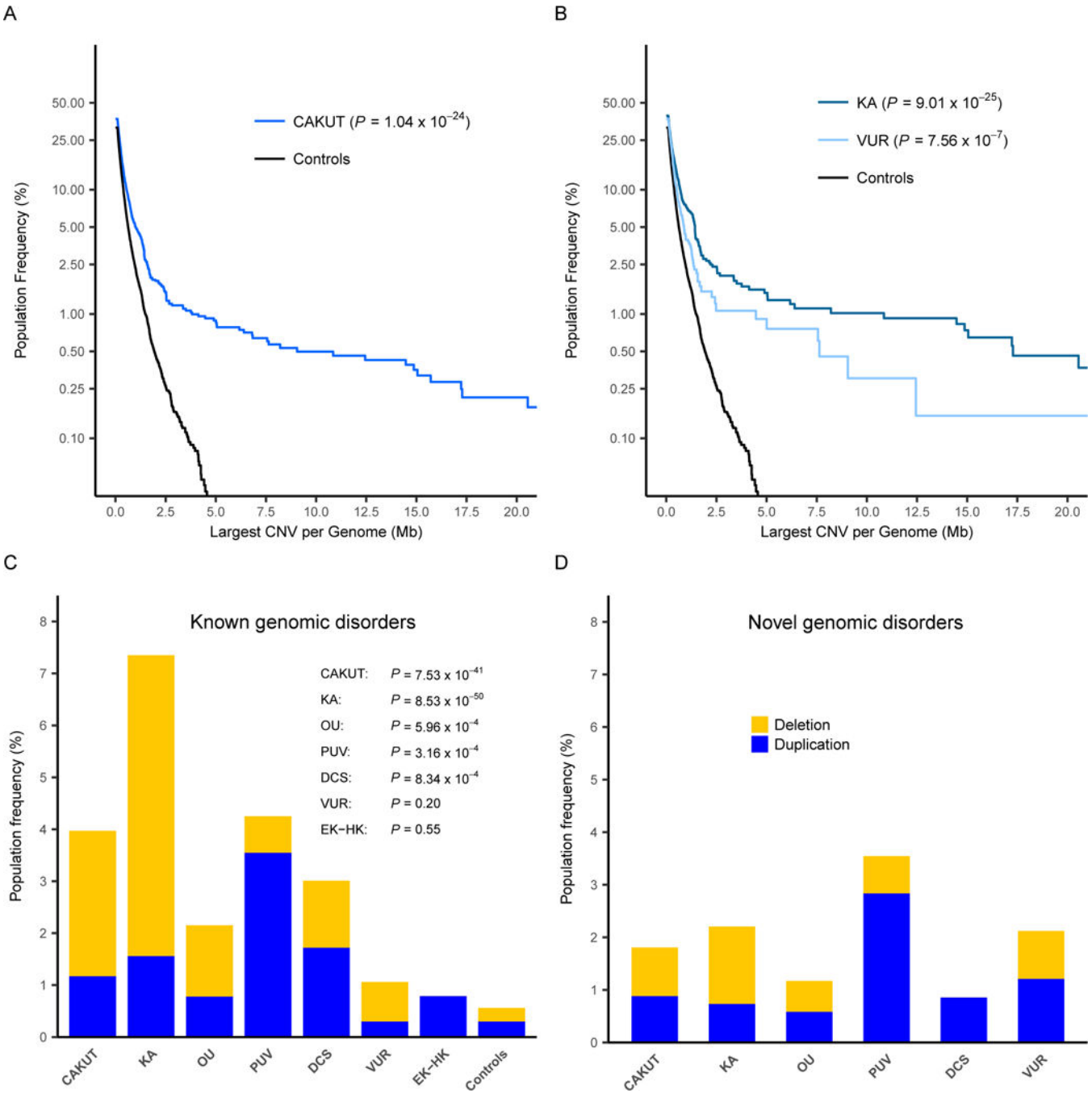

FIGURE 1.

Burden of rare copy number variations in CAKUT cases compared to controls. **A, B.** Burden of large, rare, exonic CNVs in all CAKUT cases and controls (**A**) and in KA and VUR cases and controls **(B)**. **C, D**. Prevalence of known genomic disorders (**C**) and novel likely pathogenic copy number variants (**D**) in CAKUT cases and controls. Deletions are marked in red, duplications are marked in green. KA, OU, PUV and DCS were significantly enriched for genomic disorders. The genomic architecture of KA cases was predominantly devised by deletions, while the genetic basis of PUV and DCS cases mostly constituted of

duplications. CAKUT = congenital anomalies of the kidney and urinary tract, DCS = duplex collecting system, EK-HK = ectopic kidney/horseshoe kidney, ERM = extrarenal malformations, KA = kidney anomaly, Mb = megabases, OU = obstructive uropathy, PUV = posterior urethral valves, VUR = vesicoureteral reflux.

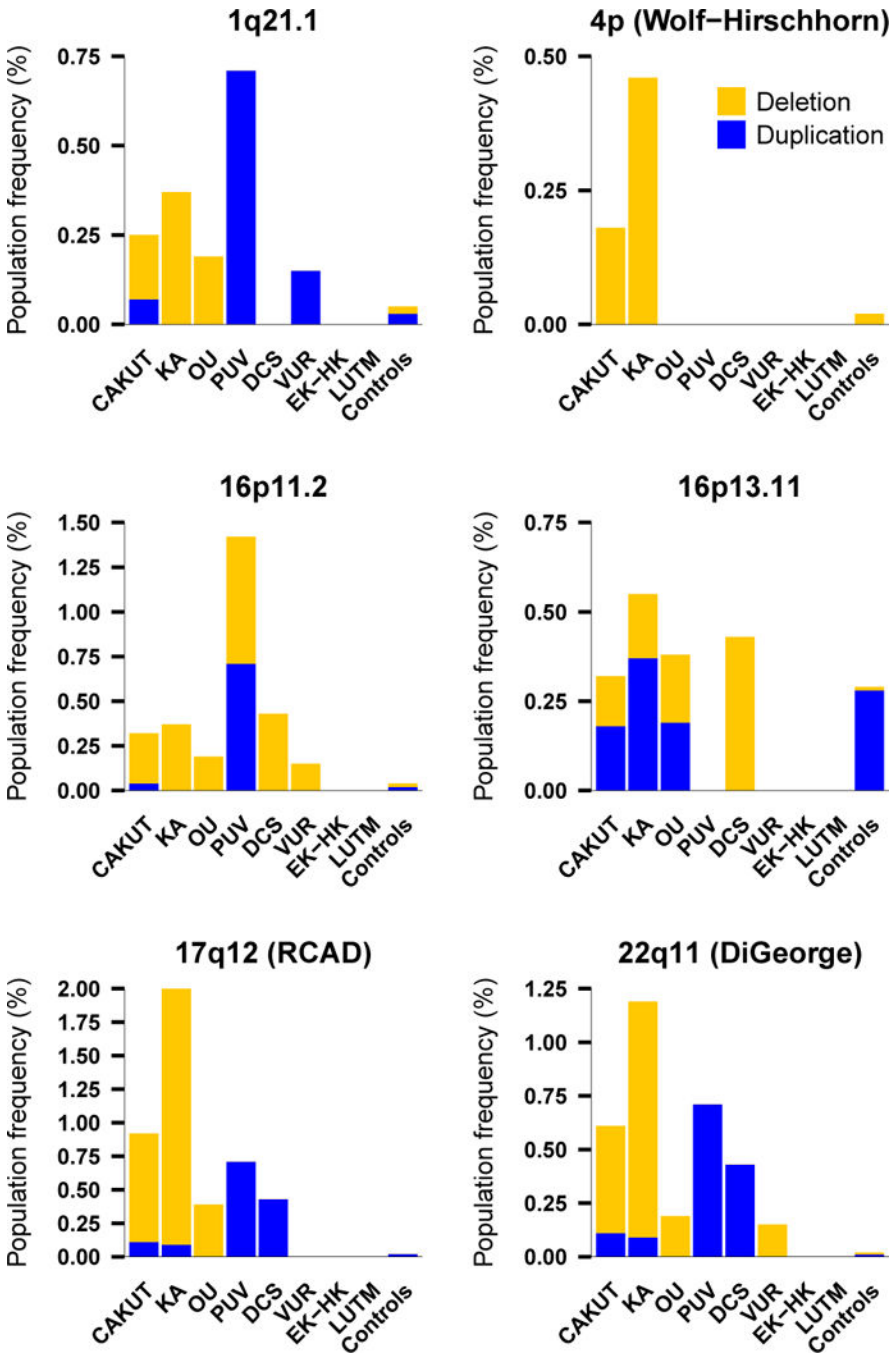

FIGURE 2.

Common genomic disorders loci in CAKUT cases and their prevalence in controls. Deletions are marked in red, duplications are marked in green. Among these common genomic loci, the chromosome 16p11.2 locus showed high pleiotropy, whereas the Wolf-Hirschhorn, 17q12 and 22q11.2 loci were mostly identified in KA cases. CAKUT = congenital anomalies of the kidney and urinary tract, Ctrls = controls, DCS = duplex collecting system, EK-HK = ectopic kidney/horseshoe kidney, KA = kidney anomaly,

LUTM = other lower urinary tract malformation, OU = obstructive uropathy, PUV = posterior urethral valves, RCAD = renal cysts and diabetes, VUR = vesicoureteral reflux.

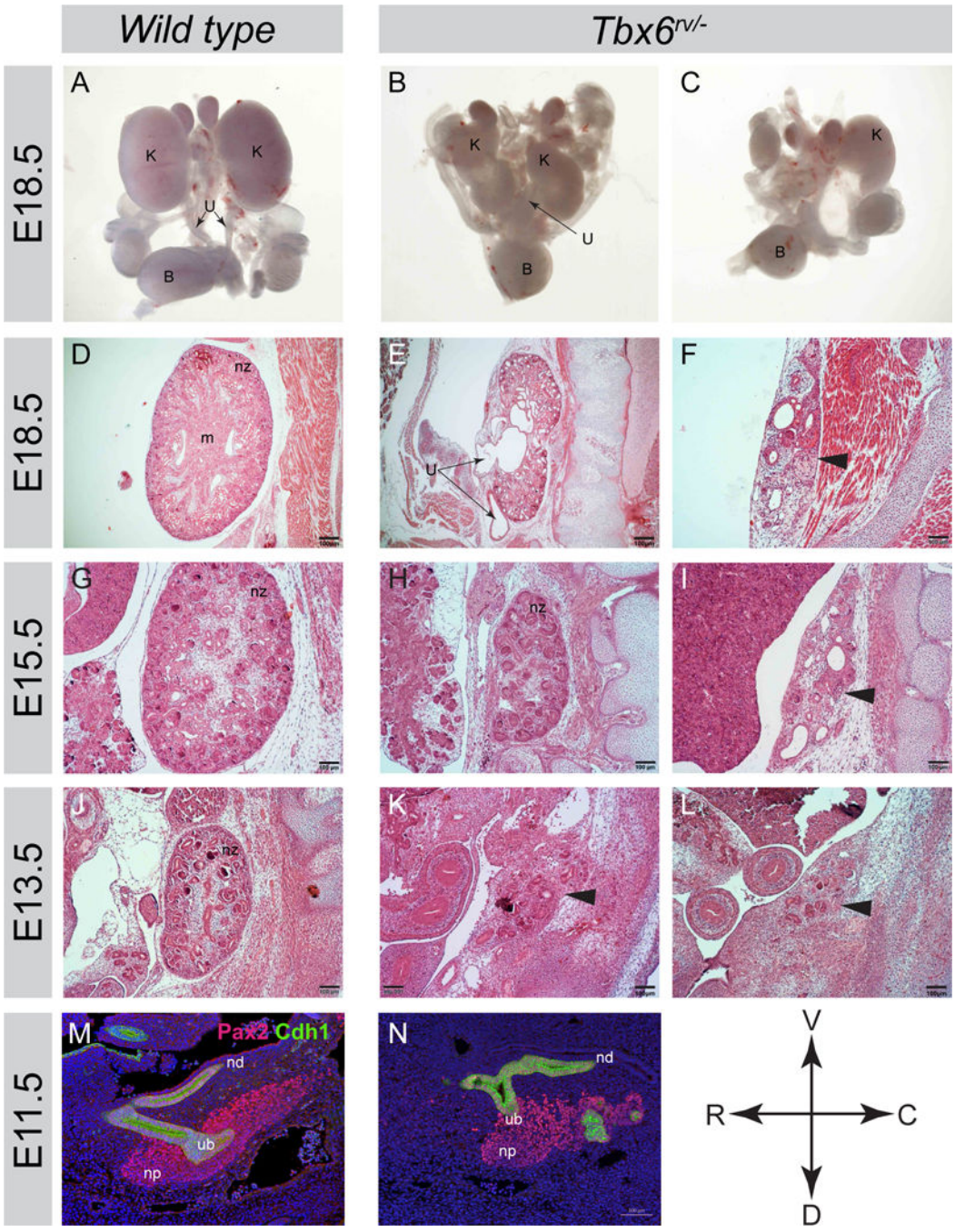

FIGURE 3.

Analysis of urinary tract phenotypes in *Tbx6*$^{rv/-}$ mutants. **A-C**. Whole mounts of urogenital tracts isolated from E18.5 wild type (**A**) and *Tbx6*$^{rv/-}$ mutants showing severe bilateral renal hypoplasia (**B**) and unilateral renal agenesis with contralateral renal hypoplasia (**C**); K = kidney, U = ureter, B = bladder. **D-F**. H&E stained sagittal sections from E18.5 wild type (**D**) and *Tbx6*$^{rv/-}$ mutants (**E,F**). In the wild type it is appreciable the normal developing nephrogenic zone (nz) and kidney medulla (m). The arrows in **E** point to the dilated renal pelvis (upper arrow) and ureter (lower arrow), which are indicative of hydronephrosis and

hydroureter, respectively. In E the kidney parenchyma also appears severely hypoplastic. The arrowhead in F points to the rudimentary kidney, which is embedded in paraspinal musculature. Few dilated tubule and microcysts are present. **G-I**. H&E stained kidney from and E15.5 wild type embryo (**G**) and E15.5 *Tbx6*$^{rv/-}$ mutant embryos (**H, I**). The mutants show moderate to severe hypoplasia with reduction of nephrogenic zone (nz) (**H**); the arrowhead indicated severely underdeveloped kidney tissue with tubule dilation and microcysts (**I**). **J-L**. H&E histological analysis of kidneys from E13.5 wild type embryos (**J**) and *Tbx6*$^{rv/-}$ mutants (K,L). The arrowheads point to the rudimentary kidneys that are embedded in the body wall. **M,N**. Immunostaining of E11.5 wild type and *Tbx6*$^{rv/-}$ mutant embryos stained with Pax2 (red) and Cdh1 (green) showing the ureteric bud (ub) nephron progenitors (np), and nephric duct (nd). Note in the wild type embryo, the ureteric bud has invaded the metanephric mesenchyme and branched, while in the mutant, the ureteric bud has not fully invaded the metanephric mesenchyme.

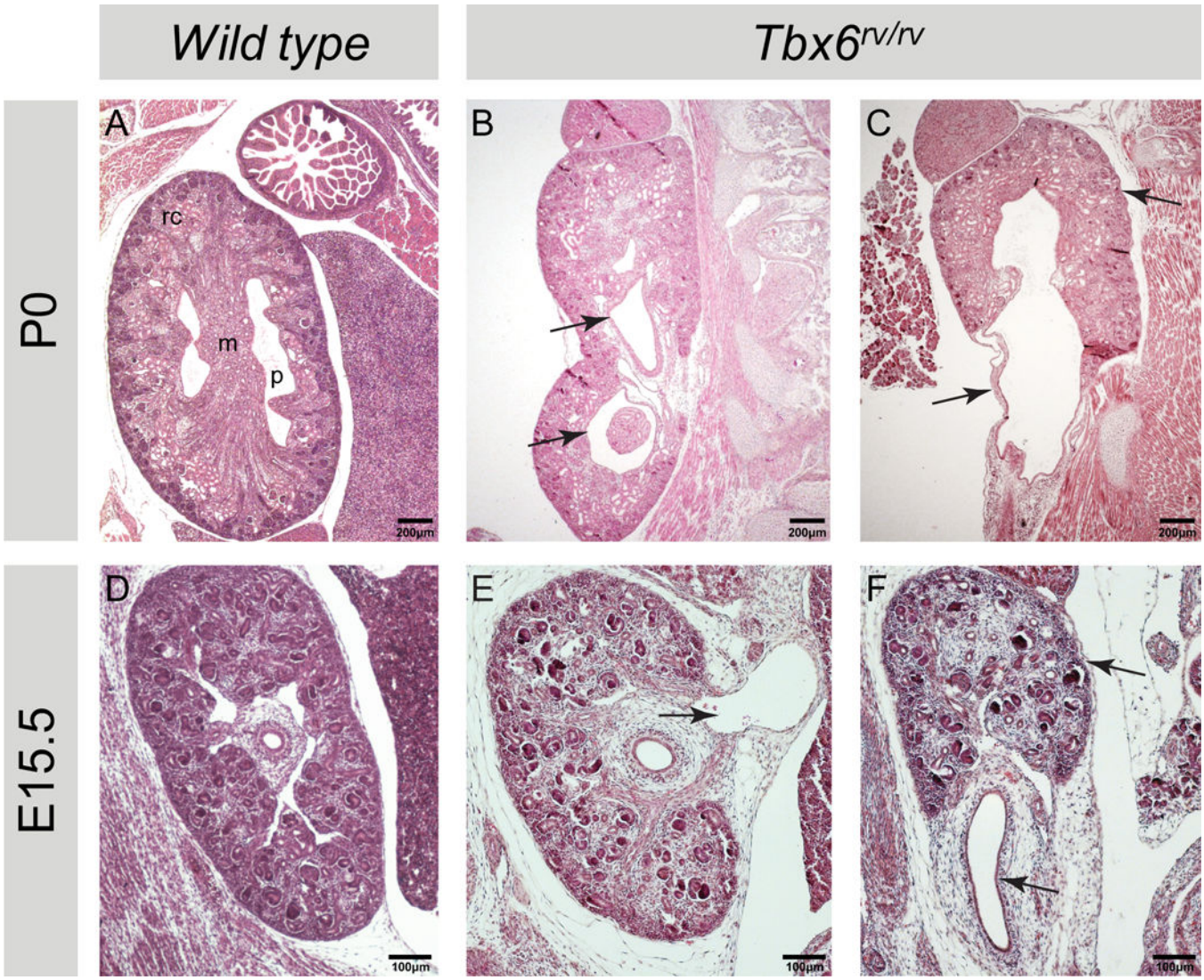

FIGURE 4.

Analysis of urinary tract phenotypes in *Tbx6*$^{rv/rv}$ mutants. **A-C**. H&E stained sagittal sections from a wild type P0 pup (**A**) and P0 *Tbx6*$^{rv/rv}$ pups (**B,C**). In the wild type can be appreciated the renal cortex (rc) resulting from the development of the nephrogenic zone; the medulla (m) and the renal pelvis (p). The arrows in B point to the duplicated kidneys. The arrows in C point to hypoplastic kidney and the dilated renal pelvis and proximal ureter. **D-F**. H&E stained sections from E15.5 wild type embryo (**D**) and *Tbx6*$^{rv/rv}$ mutant embryos (**E,F**). The arrow in E points to the dilated renal pelvis and proximal ureter. The arrows in F points to the hypoplastic kidney and dilated ureter.

TABLE 1.

Distribution of largest, rare, exonic CNV per individual across different size thresholds

	All CAKUT	KA	OU	PUV	DCS	VUR	EK-HK	LUTM	Controls
Total n	**2,817**	**1,083**	**512**	**141**	**233**	**659**	**127**	**62**	**21,490**
≥100 kb: n (%)	1,044 (37.1)	426 (39.4)	199 (38.9)	46 (32.6)	66 (28.3)	245 (37.2)	46 (36.2)	16 (25.8)	6,767 (31.5)
OR (CI)	1.28 (1.18–1.39)	1.41 (1.24–1.60)	1.38 (1.15–1.66)	1.05 (0.72–1.52)	0.86 (0.64–1.15)	1.29 (1.09–1.52)	1.24 (0.84–1.80)	0.76 (0.40–1.36)	-
P Fisher's exact	4.01×10^{-9}	1.02×10^{-17}	5.25×10^{-4}	0.79	0.32	2.20×10^{-5}	0.25	0.41	-
≥250 kb: n (%)	615 (21.8)	260 (24.0)	107 (20.9)	26 (18.4)	38 (16.3)	147 (22.3)	25 (19.7)	12 (19.4)	3,510 (16.3)
OR (CI)	1.43 (1.30–1.58)	1.62 (1.40–1.87)	1.35 (1.15–1.66)	1.16 (0.72–1.79)	1.00 (0.68–1.42)	1.47 (1.21–1.78)	1.26 (0.78–1.96)	1.23 (0.60–2.34)	-
P Fisher's exact	1.35×10^{-12}	3.13×10^{-10}	7.82×10^{-3}	0.49	1.00	9.61×10^{-5}	0.33	0.49	-
≥500 kb: n (%)	316 (11.2)	154 (14.2)	45 (8.8)	13 (9.2)	16 (6.9)	71 (10.8)	12 (9.4)	5 (8.1)	1,535 (7.1)
OR (CI)	1.64 (1.44–1.87)	2.20 (1.79–2.58)	1.25 (1.08–1.68)	1.32 (0.68–2.35)	0.96 (0.54–1.60)	1.57 (1.20–2.02)	1.36 (0.68–2.47)	1.14 (0.36–2.83)	-
P Fisher's exact	3.23×10^{-13}	5.67×10^{-15}	0.16	0.32	1.00	7.62×10^{-5}	0.29	0.80	-
≥1,000 kb: n (%)	141 (5.0)	79 (7.3)	16 (3.1)	7 (5.0)	8 (3.4)	26 (3.9)	2 (1.6)	3 (4.8)	502 (2.3)
OR (CI)	2.20 (1.81–2.67)	3.30 (2.54–4.22)	1.35 (0.76–2.23)	2.18 (0.86–4.66)	1.49 (0.63–3.00)	1.72 (1.10–2.57)	0.67 (0.08–2.48)	2.13 (0.42–6.56)	-
P Fisher's exact	4.41×10^{-14}	4.82×10^{-17}	0.24	4.94×10^{-2}	0.27	1.29×10^{-2}	1.00	0.18	-

Proportion of individuals with their largest, rare CNV at least as large as the indicated size threshold. CAKUT = congenital anomalies of the kidney and urinary tract, CI = 95% confidence interval, CNV = copy number variation, DCS = duplication of the collecting system and/or ureter, EK-HK = ectopic kidney-horseshoe kidney, KA = kidney anomaly, kb = kilobases, LUTM = other lower urinary tract malformation, OR = odds ratio, OU = obstructive uropathy, PUV = posterior urethral valves, VUR = vesicoureteral reflux.

TABLE 2.

Comparison of CNV landscape across major CAKUT subcategories

	CAKUT (n=2,824)	KA (n=1,088)	OU (n=512)	VUR (n=660)	Controls (n=21,498)
Median CNV size (kb) (IQR)‡	245 (315)	258 (394)	223 (244)	248 (273)	223 (248)
***P* value**	3.4×10^{-6}	3.6×10^{-6}	0.87	1.9×10^{-2}	-
Median total CNV span (kb) per genome (IQR)‡	350 (598)	414 (815)	306 (451)	353 (503)	300 (437)
***P* value**	1.4×10^{-7}	1.9×10^{-8}	0.61	1.7×10^{-2}	-
% individuals with at least 1 large, rare CNVs‡	37.1	39.3	38.9	37.2	31.5
***P* value**	4.0×10^{-9}	1.0×10^{-7}	5.3×10^{-4}	2.2×10^{-3}	-
% individuals with at least 2 large, rare CNVs‡	10.5	12.6	10.4	9.7	8.1
***P* value**	1.6×10^{-5}	1.0×10^{-6}	7.1×10^{-2}	0.12	-
% individuals with known GDs	4.1	7.4	2.1	1.1	0.6
***P* value**	7.5×10^{-41}	8.5×10^{-50}	6.0×10^{-4}	0.20	-
Ratio of No. of large, rare deletions: duplications‡	0.90	1.02	1.05	0.74	0.94
***P* value**	0.34	0.35	0.37	3.3×10^{-2}	-

CNVs refer to exonic CNVs ≥100 kb with frequency of < 1:1,000 in controls.

‡Metrics derived from burden analysis.

CAKUT = congenital anomalies of the kidney and urinary tract, CNV = copy number variation, GD = genomic disorder, IQR = interquartile range, KA = kidney anomaly, OU = obstructive uropathy, VUR = vesicoureteral reflux.

TABLE 3.

Clinical characteristics of CAKUT cases affected by the heterozygous 16p11.2 microdeletion.

Case	Sex	Ethnicity	Chr	Start (Mb)	End (Mb)	Size (Mb)	Number of genes	Age of Diagnosis (years)	Phenotype	Additional CAKUT	Extrarenal malformations	Family history	eGFR (ml/min/1.73m^2)
CAKUT1	M	White	16	28.4	30.1	1.7	69	9	KA	N	N	NA	NA
CAKUT2	M	Black	16	29.5	30.1	0.6	31	NA	OU	N	N	NA	NA
CAKUT3	F	White	16	29.5	30.1	0.6	31	NA	KA	N	N	NA	NA
CAKUT4	M	White	16	29.5	30.1	0.6	31	0	PUV	Y (VUR, KA bilateral)	N	Y	158
CAKUT5	F	White	16	29.5	30.1	0.6	30	NA	VUR	N	Y (scoliosis)	NA	NA
CAKUT6	M	White	16	29.5	30.1	0.6	40	NA	KA	N	Y (craniofacial dysmorphia, radial agenesis, bilateral thumb aplasia)	N	44
CAKUT7	M	Admixed	16	29.6	30.1	0.5	28	NA	DCS	Y	N	NA	NA
CAKUT8	M	White	16	29.6	30.1	0.5	34	NA	KA	N	Y (cryptorchidism)	NA	NA
CAKUT9	M	White	16	29.9	32.6	2.7	94	N	KA	N	Y (craniofacial dysmorphias, abnormality of the feet, cardiac defect, corpus callosum hypoplasia)	NA	5

CAKUT = congenital anomalies of the kidney and urinary tract, Chr = chromosome, DCS = duplicated collecting system and/or ureter, eGFR = estimated glomerular filtration rate, F = female, KA = kidney anomaly, M = male. Mb = megabases, N = no, NA = not available, OU = obstructive uropathy, PUV = posterior urethral valves, VUR = vesicoureteral reflux; Y = yes.

TABLE 4.

Tbx6 dosage-dependent kidney and urinary tract phenotypes

Phenotype	*Tbx6*$^{rv/-}$	*Tbx6*$^{rv/rv}$			*Tbx6*$^{+/-}$	*Tbx6*$^{rv/+}$	*Tbx6*$^{+/+}$
	E17.5–E18.5*	E15.5	E18.5	P0-P1	E18.5*	E18.5*	E18.5
Bilateral renal agenesis	4/19	0/8	0/5	0/3	0/4	0/6	0/12
Unilateral renal agenesis	6/19	0/8	0/5	0/3	0/4	0/6	0/12
Bilateral renal hypoplasia/dysplasia	8/19	3/8	0/5	0/3	0/4	0/6	0/12
Unilateral renal hypoplasia/dysplasia	3/19	3/8	4/5	1/3	0/4	0/6	0/12
Hydronephrosis/ hydroureter	5/19	1/8	1/5	3/3	0/4	0/6	0/12
Duplex kidney/ureter	1/19	0/8	4/5	0/3	0/4	0/6	0/12
Total embryos with defects (%)	19/19 (100)	6/8 (75)	4/5 (80)	3/3 (100)	0/4 (0)	0/6 (0)	0/12 (0)

*Based on gross morphology only. We analyzed four *Tbx6*$^{+/-}$ and six Tbx6$^{+/rv}$ E18.5 embryos by gross morphology thus hampering ability to assess milder phenotypes and incomplete penetrance in these embryos.