Corrected: Author Correction

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

Miguel Verbitsky [1,51], Rik Westland[1,2,51], 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 Marasà [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]* and Simone Sanna-Cherchi [1]*

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 (that is, 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; and 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.

CAKUT has a 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 spectrum of phenotypes, which can result from early disruptions in transcription factors and signaling molecules, such as *PAX2, EYA1, RET, BMP4* (refs. [5,6]) and others, or are directly related to spatiotemporal interactions of the outgrowing ureteric bud and metanephric mesenchyme[7-9]. Early disturbance of these interactions leads to renal agenesis, hypoplasia or hypodysplasia, whereas later perturbations in the outgrowth of the ureteric bud result in obstructive uropathy (OU), vesicoureteral 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 has indicated 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], thus suggesting that a single genetic lesion can have pleiotropic manifestations across the

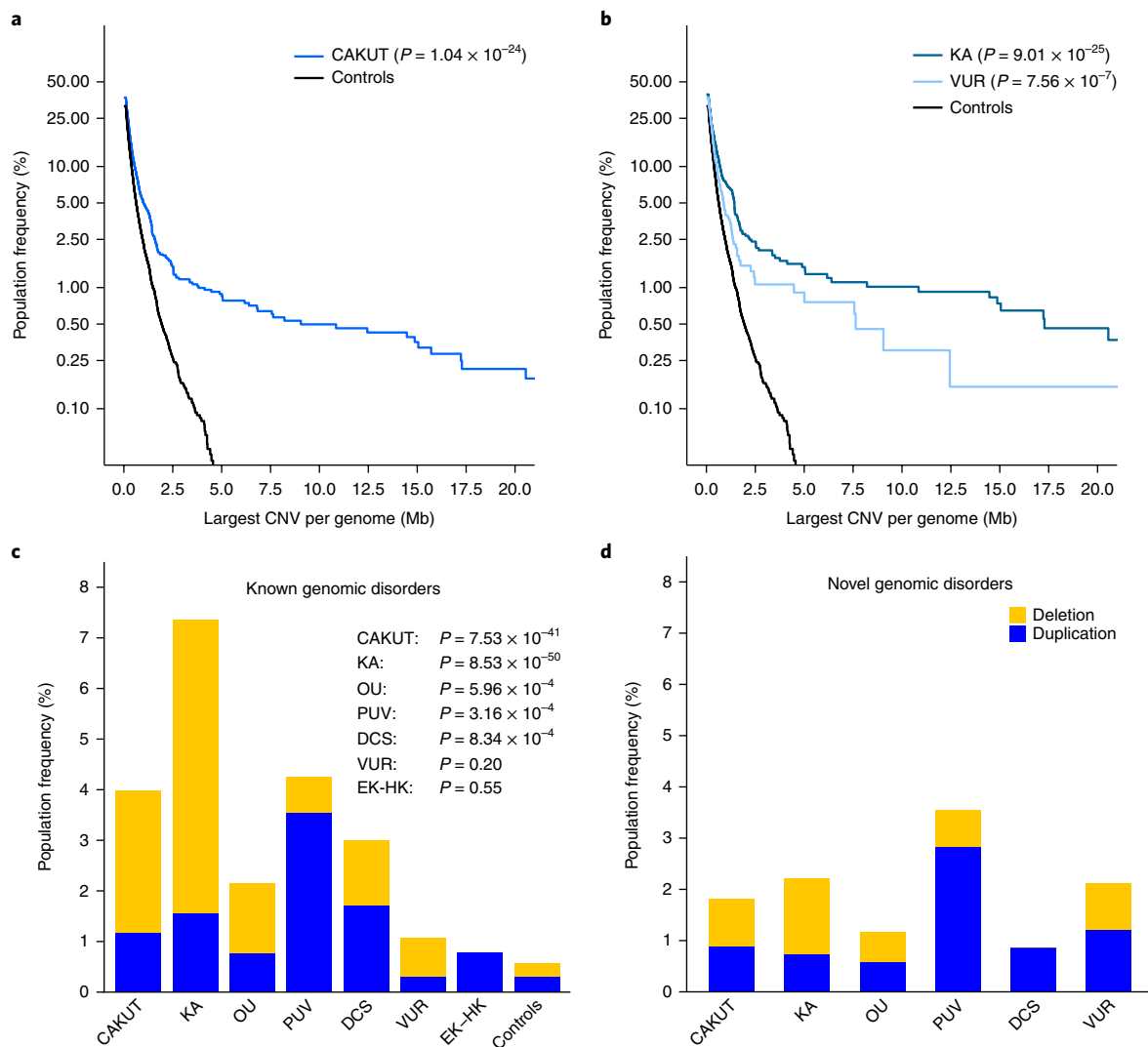A full list of affiliations appears at the end of the paper.

Fig. 1 | Burden of rare copy number variants in CAKUT cases compared with 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 CNVs (**d**) in CAKUT cases and controls. Deletions are marked in yellow; duplications are marked in blue. KA, OU, PUV and DCS were significantly enriched for genomic disorders. Deletions predominated in KA, whereas duplications were more frequent in PUV and DCS. P values are shown for comparison between cases and controls.

spectrum of CAKUT. Conversely, differences in the prevalence and severity of structural malformations point toward a distinct molecular basis and genetic architecture[22,23]. To date, there are more than 50 single-gene disorders known to underlie isolated and nonisolated (that is, syndromic) CAKUT[5,24,25]. Furthermore, a substantial number of CAKUT patients carry copy number variants (CNV) that were 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], thus emphasizing that large studies across the entire phenotypic spectrum of CAKUT are indispensable to identify genes and allelic variants that either are specific to subcategories of disease or 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 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 Fig. 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 fewer than 1:1,000 population controls across

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
	n = 2,817	n = 1,083	n = 512	n = 141	n = 233	n = 659	n = 127	n = 62	n = 21,490
CNV ≥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)	–
Fisher's exact P	4.01×10^{-9}	1.02×10^{-17}	5.25×10^{-4}	0.79	0.32	2.20×10^{-5}	0.25	0.41	–
CNV ≥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)	–
Fisher's exact P	1.35×10^{-12}	3.13×10^{-10}	7.82×10^{-3}	0.49	1.00	9.61×10^{-5}	0.33	0.49	–
CNV ≥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)	–
Fisher's exact P	3.23×10^{-13}	5.67×10^{-15}	0.16	0.32	1.00	7.62×10^{-5}	0.29	0.80	–
CNV ≥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)	–
Fisher's exact P	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, comparing cases to controls.

different ancestries, as estimated by principal component analysis (Supplementary Fig. 2).

This analysis revealed a marked enrichment for large, rare CNVs in CAKUT compared with controls ($P = 1.04 \times 10^{-24}$; Fig. 1a), which was consistent across virtually all metrics examined, including the number of individuals with large, rare CNVs, the median size and total CNV span per genome, and the fraction of GD-CNVs, thus indicating an important role for gene-disrupting CNVs across the entire CAKUT spectrum (Table 1 and Supplementary Table 3). This signal was driven by cases with and without extrarenal manifestations, because the burden difference was still highly significant in analysis of cases without extrarenal defects separately ($P = 3.21 \times 10^{-8}$; Fig. 1a,b and Supplementary Figs. 3 and 4). Even when only simplex isolated CAKUT cases (that is, cases without extrarenal manifestation and with no additional CAKUT phenotypes other than the primary one) were considered, there was still an excess burden of large, rare CNVs compared with controls ($P = 1.15 \times 10^{-8}$; Supplementary Fig. 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 (odds ratio (OR) 1.64, 95% confidence interval (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 with deletions in VUR and PUV cases compared with controls ($P = 3.33 \times 10^{-2}$ and $P = 8.67 \times 10^{-4}$, respectively; Supplementary Fig. 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 Fig. 7).

Genomic disorders inform about the genetic architecture of CAKUT. We cross-annotated all rare CNVs with a curated list of known GDs[33,34] (Supplementary Table 7) and identified 45 distinct known GDs in 112 (4.0%) independent cases (Fig. 1c and Supplementary Table 8). 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 criteria[35,36] as likely pathogenic imbalances (Methods and Supplementary Table 11). Among these, the CNVs classified as pathogenic were an atypical deletion at the 16p11.2 locus (described 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 Fig. 8a–d). We also identified overlapping duplications at the 15p11.2 locus in five 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 (Fig. 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, which showed an excess of duplications at the same genomic loci.

When we examined associations with disease severity, cases with 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; $P = 6.00 \times 10^{-15}$) than did cases without a known GD-CNV. However, in agreement with the burden tests conducted above, analysis of simplex isolated cases still showed a greater burden than that of controls (OR 3.12, 95% CI 2.06–4.61; $P = 1.86 \times 10^{-7}$), thus strongly implicating the importance 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)), 11 (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 Fig. 9).

KA, OU and VUR show distinct genomic characteristics. A comparison of CNV landscapes 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

Table 2 | Comparison of CNV landscapes 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)[a]	245 (315)	258 (394)	223 (244)	248 (273)	223 (248)
P (Wilcoxon)	3.4×10^{-6}	3.6×10^{-6}	0.87	1.9×10^{-2}	–
Median total CNV span (kb) per genome (IQR)[a]	350 (598)	414 (815)	306 (451)	353 (503)	300 (437)
P (Wilcoxon)	1.4×10^{-7}	1.9×10^{-8}	0.61	1.7×10^{-2}	–
% individuals with at least one large, rare CNVs[a]	37.1	39.3	38.9	37.2	31.5
P (Fisher's exact)	4.0×10^{-9}	1.0×10^{-7}	5.3×10^{-4}	2.2×10^{-3}	–
% individuals with at least two large, rare CNVs[a]	10.5	12.6	10.4	9.7	8.1
P (Fisher's exact)	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 (Fisher's exact)	7.5×10^{-41}	8.5×10^{-50}	6.0×10^{-4}	0.20	–
Ratio of number of large, rare deletions: duplications[a]	0.90	1.02	1.05	0.74	0.94
P (Fisher's exact)	0.34	0.35	0.37	3.3×10^{-2}	–

CNVs here refer to autosomal, exonic CNVs ≥100 kb with a frequency of <1:1,000 in controls. [a]Metrics derived from burden analysis. CAKUT cases were compared with controls. P values were derived from the indicated tests.IQR, interquartile range.

Table 3 | Clinical characteristics of CAKUT cases affected by the heterozygous 16p11.2 microdeletion

Case	Sex	Ancestry	Chromosome	Start (Mb)	End (Mb)	Size (Mb)	Number of genes	Age of diagnosis (years)	Phenotype	Additional CAKUT	Extrarenal malformations	Family history	eGFR (ml/min/ 1.73m²)
CAKUT1	M	White	16	28.4	30.1	1.7	69	9	KA	N	N	N	<5
CAKUT2	M	Black	16	29.5	30.1	0.6	31	0.5	OU	U	U	U	39
CAKUT3	F	White	16	29.5	30.1	0.6	31	0	KA	U	U	U	49
CAKUT4	M	White	16	29.5	30.1	0.6	31	0	PUV	Y (VUR, KA bilateral)	N	U	158
CAKUT5	F	White	16	29.5	30.1	0.6	30	U	VUR	N	Y (scoliosis)	U	>90
CAKUT6	M	White	16	29.5	30.1	0.6	40	2	KA	N	Y (craniofacial dysmosphism radial agenesis, bilateral thumb aplasia)	N	44
CAKUT7	M	Admixed	16	29.6	30.1	0.5	28	0.5	DCS	Y (VUR)	N	N	110
CAKUT8	M	White	16	29.6	30.1	0.5	34	9	KA	N	Y (cryptorchidism, febrile seizures)	N	113
CAKUT9	M	White	16	29.9	32.6	2.7	94	Prenatal	KA	N	Y (craniofacial dysmorphism, abnormality of the feet, cardiac defect, corpus callosum hypoplasia)	N	<5

eGFR, estimated glomerular filtration rate at presentation; F, female; M, male; N, no; Y, yes; U, unknown.

proportion of cases carrying exonic CNVs ≥100 kb than controls. KA cases had the highest CNV burden ($P = 9.01 \times 10^{-25}$; Fig. 1b), as evidenced by median CNV size and total span, as well as the 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 Fig. 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 ($P = 1.27 \times 10^{-3}$), in agreement with the major role for GD-CNVs in the pathogenesis of KA (Supplementary Fig. 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 (seven cases with GD-CNVs, OR 1.71; 95% CI 0.67–3.64; $P = 0.20$; Fig. 1c). The large rare CNVs in VUR were mostly classified as likely pathogenic, thus potentially indicating that novel syndromes account for the molecular basis of this disorder (Table 2, Fig. 1d and Supplementary Figs. 3 and 10). Finally, OU cases fell in an intermediate category, with CNVs that were not statistically larger than those of controls, yet the excess CNV burden ($P = 0.01$; Supplementary Fig. 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 Fig. 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-sized 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
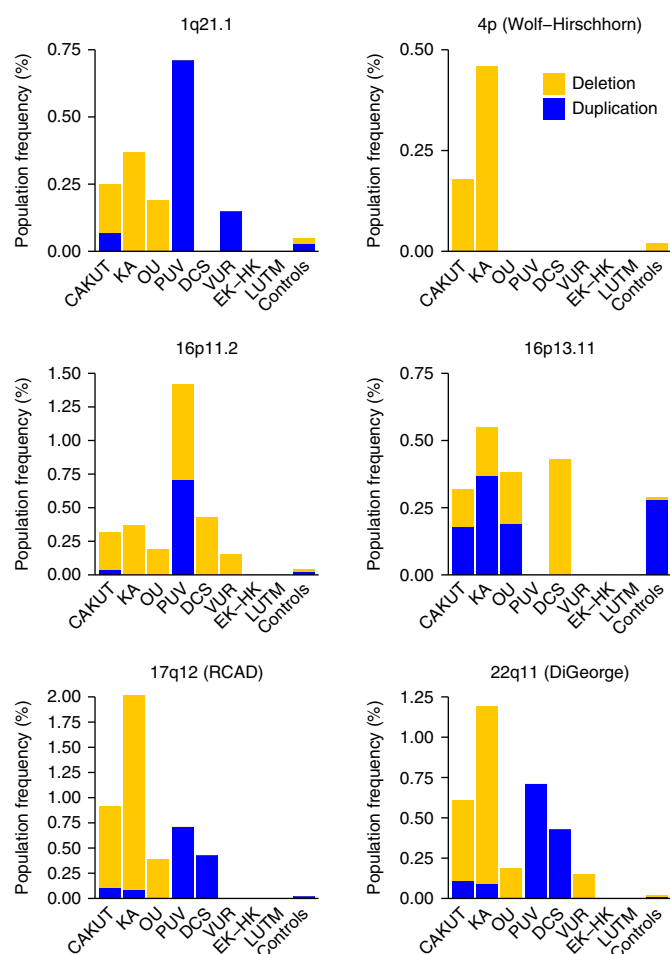
Fig. 2 | Common genomic disorders loci in CAKUT cases and their prevalence in controls. Deletions are marked in yellow; duplications are marked in blue. Among these common genomic loci, the chromosome 16p11.2 deletion showed high pleiotropy, whereas the Wolf–Hirschhorn, 17q12 and 22q11.2 deletions were mostly identified in KA cases. RCAD, renal cysts and diabetes.

databases of genomic variants such as DECIPHER[37,38] and ISCA[35] and found 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.

Although most GD-CNVs affected unique or few cases (Supplementary Table 8), six loci explained 73 of the 112 (65.2%) subjects who carried a known GD-CNV (Fig. 2). These common GD loci included chromosome 1q21.1 (seven (6.3%) cases with a GD-CNV; five deletions, two duplications), chromosome 4p16.1-p16.3 (Wolf–Hirschhorn syndrome; five (4.5%) cases with a GD-CNV; all deletions), chromosome 16p11.2 (nine (8.0%) cases with a GD-CNV; eight deletions, one duplication), chromosome 16p13.11 (nine (8.0%) cases with a GD-CNV; four deletions, five duplications), chromosome 17q12 (26 (23.2%) cases with a GD-CNV; 23 deletions, three duplications) and chromosome 22q11.2 (17 (15.2%) cases with a GD-CNV; 14 deletions, three duplications) loci. Genotype–phenotype correlations indicated 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, thus suggesting a high pleiotropic effect. These observations provide further support that some genetic

lesions result in specific CAKUT subphenotypes, whereas 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 through which pathogenic CNVs confer risk for CAKUT, we conducted whole-exome sequencing (WES) in 23 patients with pathogenic microdeletions at 14 independent loci (Supplementary Table 13). On the basis of recessive loss-of-function (LOF) inheritance, WES would uncover a hemizygous LOF mutation on the nondeleted 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 Exome Aggregation Consortium database (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 had a 14.9-Mb deletion at chromosome 3q13.22–13.1 (ref. [57]) (Supplementary Fig. 11). The *EFCAB12* variant was inherited from a heterozygous unaffected mother, whereas the CNV occurred de novo[57] (Supplementary Fig.11). A 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, thus suggesting that biallelic truncating mutations are probably not tolerated in humans. This finding further points toward recessive mutations in *EFCAB12* as contributors to the developmental syndrome of this individual. Additionally, we performed clinical annotation of genes 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 any of the 23 deletion carrier patients. Hence, our WES studies suggest haploinsufficiency as the 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% versus 0.02%; OR 13.7, 95% CI 4.1–52.2; $P = 4.39 \times 10^{-6}$; Fig. 2, Table 3 and Supplementary Tables 9 and 10), thus implicating CAKUT as an important feature of this syndrome. To better estimate the prevalence of genitourinary malformations in individuals with 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 the 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 ten cases (5.4%), but these estimates probably reflect the standard clinical indication for ordering a DNA microarray for diagnosis of a syndromic disease, that is, neurodevelopmental delay and dysmorphic features. Abnormalities of the skeletal system, which are a hallmark of the 16p11.2 microdeletion syndrome[59,60], were reported in only 19 cases (10.2%), because, 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's Hospital of Philadelphia (CHOP). We analyzed the medical records of 42 children with the 16p11.2 microdeletion syndrome (Supplementary

Table 4 | *Tbx6* dosage-dependent kidney and urinary tract phenotypes

Phenotype	$Tbx6^{rv/-}$	$Tbx6^{rv/rv}$	$Tbx6^{rv/rv}$	$Tbx6^{rv/rv}$	$Tbx6^{+/-}$	$Tbx6^{rv/+}$	$Tbx6^{+/+}$
	E17.5–E18.5[a]	E15.5	E18.5	P0–P1	E18.5[a]	E18.5[a]	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)

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

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 cases, and genitourinary anomalies were present in 9 of 15 cases (40.0%), thus highlighting that CAKUT, similarly to scoliosis, is a common feature in this syndrome and is characterized by incomplete penetrance. In agreement 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 major 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 of CAKUT in the 16p11.2 microdeletion syndrome. We have 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 nine CAKUT cases with typical or atypical 16p11.2 microdeletions identified a ~175-kb minimal region of overlap predicted to harbor the genetic driver(s) for CAKUT (Supplementary Fig. 12). Among the 19 genes included in the minimal region of overlap, *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 the 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), which has been studied for vertebral development and spine defects[71]. Because the homozygous null mutation is lethal at embryonic day (E) 9.5, thus 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 and Fig. 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 (Fig. 3d–l). Severe phenotypes included unilateral or bilateral renal agenesis, rudimentary kidneys and undeveloped renal parenchyma embedded in the paraspinal musculature (Fig. 3f,i,l). Milder phenotypes included unilateral and bilateral renal hypoplasia with hydroureter, tubule dilation and hydronephrosis (Fig. 3e,h,k). These data strongly implicate *TBX6* as a major driver of CAKUT phenotypes. Comparative analysis was performed on wild-type and *Tbx6^{rv/−}* embryos at E11.5 after staining with E-cadherin (Cdh1), an epithelial marker, which labels the common nephric duct and the ureteric bud, which gives rise to the collecting duct system, and Pax2, which prevalently labels mesenchymal progenitors that produce nephrons[72–74]. This analysis revealed that the events leading to CAKUT (renal parenchyma abnormalities and obstruction) are present at early stages during development (Fig. 3m,n). In wild-type E11.5 embryos, the ureteric bud had invaded the metanephric blastema and had undergone a round of branching (Fig. 3n). In the *Tbx6^{rv/−}* mutant (Fig. 3n), the ureteric bud had not fully invaded the metanephric blastema and had not branched, a defect predicted to lead to KA. Additional Cdh1-positive cells in Fig. 3n might represent persistent mesonephros or ectopic ureteric buds; in that possibility, ectopic buds would explain the duplication of kidney and ureter phenotype, as observed in Fig. 4.

Because more subtle CAKUT phenotypes such as 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 with 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 postnatal day (P) 0 ($n = 5$) 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 Fig. 4). This pleiotropy of CAKUT phenotypes, ranging from renal parenchyma defects (KA) to hydronephrosis (OU) or duplication of ureters (DCS), is highly reminiscent of the pleiotropic manifestations of the 16p11.2 microdeletion in humans, thus suggesting that *TBX6* gene dosage is a major determinant of the pathogenesis of CAKUT and the observed variable expressivity of phenotypes in this syndrome.

Discussion

CAKUT has a profound effect on child health and alone accounts for about 50% of kidney failure requiring dialysis and transplantation
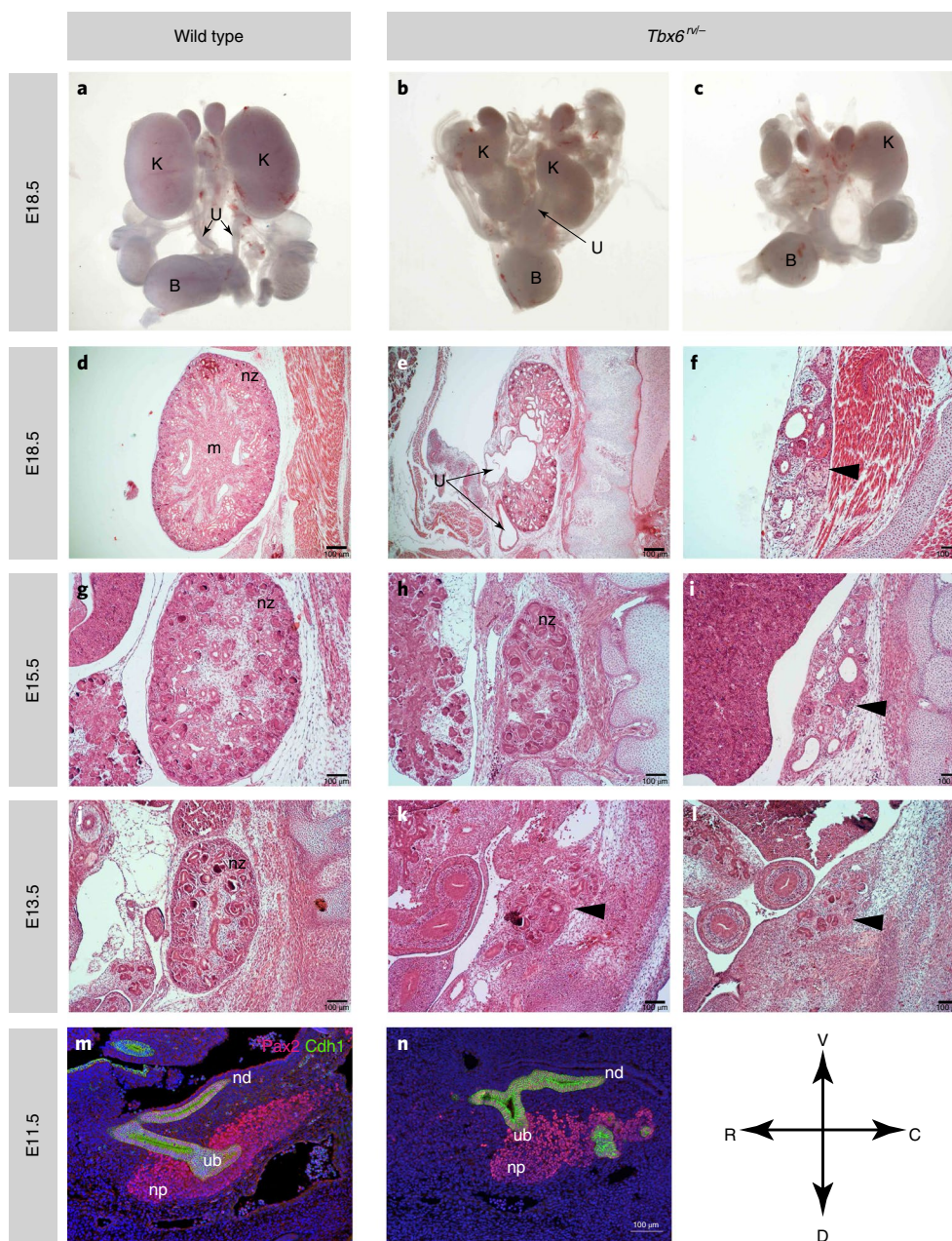
Fig. 3 | Analysis of urinary tract phenotypes in *Tbx6^rv/-* mutants. a–c, Whole mounts of urogenital tracts. isolated from E18.5 wild type (**a**), and *Tbx6^rv/-*
mutants showing severe bilateral renal hypoplasia (**b**) and unilateral renal agenesis with contralateral renal hypoplasia (**c**); K, kidney; U, ureter; B, bladder.
d–f, Hematoxylin and eosin (H&E)-stained sagittal sections from E18.5 wild type (**d**) and *Tbx6^rv/-* mutants (**e,f**). The normal developing nephrogenic zone
(nz) and kidney medulla (m) are clearly discernible in the wild type but not the mutant mice. 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 an 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). In the wild type embryo, the ureteric bud has invaded the
metanephric mesenchyme and branched, whereas in the mutant, the ureteric bud has not fully invaded the metanephric mesenchyme.

in children[1,3,75–78]. Moreover, CAKUT is often accompanied by extrarenal comorbidities, such as neurodevelopmental and cardiovascular diseases, thus 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 the progression of

kidney disease and to mitigate the associated neurocognitive and cardiac disease. In this respect, it is relevant to address two features of CAKUT: incomplete penetrance and variable expressivity of disease. Whereas prediction of a targeted treatment that prevents or reverses the disease is still limited by current knowledge, understanding the intrinsic interindividual mechanisms of
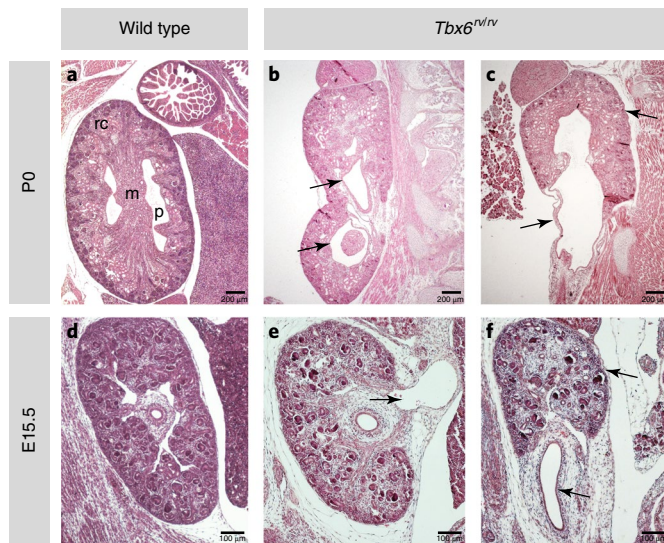
Fig. 4 | Analysis of urinary tract phenotypes in *Tbx6^rv/rv^* mutants. a–c, H&E stained sagittal sections from a wild type P0 pup (**a**) and P0 *Tbx6^rv/rv^* pups (**b,c**). The renal cortex (rc), which originates from the nephrogenic zone, is clearly distinguishable in the wild-type but not the mutant mice. The pelvis (p) and the medulla (m) are indicated. The arrows in **b** point to the duplicated kidneys. The arrows in **c** point to the hypoplastic kidney, dilated renal pelvis and proximal ureter. **d–f**, H&E-stained sections from an 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.

compensation or amplification of disease are critical for devising treatments that might slow or halt the progression of kidney disease and associated extrarenal malformations.

The study of CNVs has provided enormous insight into the genetic architecture of many developmental traits[80–89]. 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 compared with controls. Overall, we identified 45 distinct GDs at 37 independent genomic loci in 4% of the CAKUT cases, thus 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 with 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], which 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 entire phenotypic spectrum of CAKUT, thus 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 aid in 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 results from recent fate-mapping studies showing that *Tbx6* is expressed in renal progenitor cells before 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, similarly to the observations in our 16p11.2 microdeletion cases, we observed that the mutant mice showed phenotypes across the phenotypic spectrum of CAKUT, including KA, OU and DCS. Overall, our mouse data at least partly suggest that the variable expressivity and incomplete penetrance of CAKUT in cases with the 16p11.2 microdeletion are 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 substantial 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 subjects with the chromosome 16p11.2 microdeletion syndrome, with notable implications for understanding, and potentially modifying, disease penetrance.

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, http://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/.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41588-018-0281-y.

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).
2. Chesnaye, N. C. 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).
3. Sanna-Cherchi, S. et al. Renal outcome in patients with congenital anomalies of the kidney and urinary tract. *Kidney Int.* **76**, 528–533 (2009).
4. Goodyer, P. R. Renal dysplasia/hypoplasia. in *Pediatric Nephrology* 5th edn. (eds. Avner, E. D., Harmon, W. E. & Niaudet, P.) 83–91 (Lippincott Williams & Wilkins, Philadelphia, 2004).
5. Sanna-Cherchi, S., Westland, R., Ghiggeri, G. M. & Gharavi, A. G. Genetic basis of human congenital anomalies of the kidney and urinary tract. *J. Clin. Invest.* **128**, 4–15 (2018).
6. Schedl, A. Renal abnormalities and their developmental origin. *Nat. Rev. Genet.* **8**, 791–802 (2007).
7. Costantini, F. & Kopan, R. Patterning a complex organ: branching morphogenesis and nephron segmentation in kidney development. *Dev. Cell.* **18**, 698–712 (2010).
8. Short, K. M. & Smyth, I. M. The contribution of branching morphogenesis to kidney development and disease. *Nat. Rev. Nephrol.* **12**, 754–767 (2016).
9. dos Santos Junior, A. C., de Miranda, D. M. & Simoes e Silva, A. C. Congenital anomalies of the kidney and urinary tract: an embryogenetic review. *Birth Defects Res. C. Embryo Today* **102**, 374–381 (2014).
10. Chen, F. Genetic and developmental basis for urinary tract obstruction. *Pediatr. Nephrol.* **24**, 1621–1632 (2009).
11. Vainio, S. & Lin, Y. Coordinating early kidney development: lessons from gene targeting. *Nat. Rev. Genet.* **3**, 533–543 (2002).
12. Sanna-Cherchi, S. et al. Genetic approaches to human renal agenesis/hypoplasia and dysplasia. *Pediatr. Nephrol.* **22**, 1675–1684 (2007).
13. Nicolaou, N., Renkema, K. Y., Bongers, E. M., Giles, R. H. & Knoers, N. V. Genetic, environmental, and epigenetic factors involved in CAKUT. *Nat. Rev. Nephrol.* **11**, 720–731 (2015).
14. Georgas, K. M. et al. An illustrated anatomical ontology of the developing mouse lower urogenital tract. *Development* **142**, 1893–1908 (2015).
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).
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**, 2864–2870 (2006).
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**, 476–486 (2016).
18. Sanna-Cherchi, S. et al. Mutations in DSTYK and dominant urinary tract malformations. *N. Engl. J. Med.* **369**, 621–629 (2013).
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).
20. Schimmenti, L. A. et al. Further delineation of renal-coloboma syndrome in patients with extreme variability of phenotype and identical PAX2 mutations. *Am. J. Hum. Genet.* **60**, 869–878 (1997).
21. Bekheirnia, M. R. 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).
22. Ichikawa, I., Kuwayama, F., Pope, J. Ct, Stephens, F. D. & Miyazaki, Y. Paradigm shift from classic anatomic theories to contemporary cell biological views of CAKUT. *Kidney Int.* **61**, 889–898 (2002).
23. Lopez-Rivera, E. et al. Genetic drivers of kidney defects in the DiGeorge syndrome. *N. Engl. J. Med.* **376**, 742–754 (2017).
24. Hwang, D. Y. et al. Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. *Kidney Int.* **85**, 1429–1433 (2014).
25. Vivante, A., Kohl, S., Hwang, D. Y., Dworschak, G. C. & Hildebrandt, F. Single-gene causes of congenital anomalies of the kidney and urinary tract (CAKUT) in humans. *Pediatr. Nephrol.* **29**, 695–704 (2014).
26. Sanna-Cherchi, S. et al. Copy-number disorders are a common cause of congenital kidney malformations. *Am. J. Hum. Genet.* **91**, 987–997 (2012).
27. Verbitsky, M. et al. Genomic imbalances in pediatric patients with chronic kidney disease. *J. Clin. Invest.* **125**, 2171–2178 (2015).
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).
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).
30. Madariaga, L. et al. Severe prenatal renal anomalies associated with mutations in HNF1B or PAX2 genes. *Clin. J. Am. Soc. Nephrol.* **8**, 1179–1187 (2013).
31. Hoskins, B. E. et al. Missense mutations in EYA1 and TCF2 are a rare cause of urinary tract malformations. *Nephrol. Dial. Transplant.* **23**, 777–779 (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**, 1079–1090 (2010).
33. Carvalho, C. M. & Lupski, J. R. Mechanisms underlying structural variant formation in genomic disorders. *Nat. Rev. Genet.* **17**, 224–238 (2016).
34. Lupski, J. R. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet.* **14**, 417–422 (1998).
35. Miller, D. T. 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**, 749–764 (2010).
36. South, S. T. et al. ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013. *Genet. Med.* **15**, 901–909 (2013).
37. Swaminathan, G. J. et al. DECIPHER: web-based, community resource for clinical interpretation of rare variants in developmental disorders. *Hum. Mol. Genet.* **21**, R37–R44 (2012).
38. Firth, H. V. et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am. J. Hum. Genet.* **84**, 524–533 (2009).
39. Grisaru, S., Ramage, I. J. & Rosenblum, N. D. Vesicoureteric reflux associated with renal dysplasia in the Wolf-Hirschhorn syndrome. *Pediatr. Nephrol.* **14**, 146–148 (2000).
40. Mefford, H. C. et al. Recurrent reciprocal genomic rearrangements of 17q12 are associated with renal disease, diabetes, and epilepsy. *Am. J. Hum. Genet.* **81**, 1057–1069 (2007).
41. Portnoi, M. F. Microduplication 22q11.2: a new chromosomal syndrome. *Eur. J. Med. Genet.* **52**, 88–93 (2009).
42. Kobrynski, L. J. & Sullivan, K. E. Velocardiofacial syndrome, DiGeorge syndrome: the chromosome 22q11.2 deletion syndromes. *Lancet* **370**, 1443–1452 (2007).
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**, 136–143 (2011).
44. Hildebrandt, F. et al. A novel gene encoding an SH3 domain protein is mutated in nephronophthisis type 1. *Nat. Genet.* **17**, 149–153 (1997).
45. Willatt, L. et al. 3q29 microdeletion syndrome: clinical and molecular characterization of a new syndrome. *Am. J. Hum. Genet.* **77**, 154–160 (2005).
46. Mattina, T., Perrotta, C. S. & Grossfeld, P. Jacobsen syndrome. *Orphanet. J. Rare. Dis.* **4**, 9 (2009).
47. Sampson, M. G. 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**, 2618–2622 (2010).
48. Fontes, M. I. 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**, 766–772 (2016).
49. Goh, E. S. et al. Definition of a critical genetic interval related to kidney abnormalities in the Potocki-Lupski syndrome. *Am. J. Med. Genet. A* **158A**, 1579–1588 (2012).
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).
51. Kim, Y. M. et al. Phelan-McDermid syndrome presenting with developmental delays and facial dysmorphisms. *Korean J. Pediatr.* **59**, S25–S28 (2016).
52. Ozgun, M. T. et al. Prenatal diagnosis of a fetus with partial trisomy 7p. *Fetal. Diagn. Ther.* **22**, 229–232 (2007).

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**, 319–324 (2013).

54. Westland, R. et al. Phenotypic expansion of DGKE-associated diseases. *J. Am. Soc. Nephrol.* **25**, 1408–1414 (2014).

55. Karczewski, K. J. et al. The ExAC browser: displaying reference data information from over 60,000 exomes. *Nucleic Acids Res.* **45**, D840–D845 (2017).

56. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).

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**, 257–267 (2014).

58. Lata, S. et al. Whole-exome sequencing in adults with chronic kidney disease: A pilot study. *Ann. Intern. Med.* **168**, 100–109 (2018).

59. Wu, N. et al. TBX6 null variants and a common hypomorphic allele in congenital scoliosis. *N. Engl. J. Med.* **372**, 341–350 (2015).

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**, 1118–1126 (2014).

61. Dickinson, M. E. et al. High-throughput discovery of novel developmental phenotypes. *Nature* **537**, 508–514 (2016).

62. Chapman, D. L. & Papaioannou, V. E. Three neural tubes in mouse embryos with mutations in the T-box gene Tbx6. *Nature* **391**, 695–697 (1998).

63. Chapman, D. L., Agulnik, I., Hancock, S., Silver, L. M. & Papaioannou, V. E. Tbx6, a mouse T-Box gene implicated in paraxial mesoderm formation at gastrulation. *Dev. Biol.* **180**, 534–542 (1996).

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).

66. Sparrow, D. B. et al. Autosomal dominant spondylocostal dysostosis is caused by mutation in TBX6. *Hum. Mol. Genet.* **22**, 1625–1631 (2013).

67. MacEwen, G. D., Winter, R. B., Hardy, J. H. & Sherk, H. H. Evaluation of kidney anomalies in congenital scoliosis. 1972. *Clin. Orthop. Relat. Res.* **434**, 4–7 (2005).

68. Cowell, H. R., MacEwen, G. D. & Hubben, C. Incidence of abnormalities of the kidney and ureter in congenital scoliosis. *Birth Defects. Orig. Artic. Ser.* **10**, 142–145 (1974).

69. MacEwen, G. D., Winter, R. B. & Hardy, J. H. Evaluation of kidney anomalies in congenital scoliosis. *J. Bone Joint Surg. Am.* **54**, 1451–1454 (1972).

70. Hadjantonakis, A. K., Pisano, E. & Papaioannou, V. E. 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, V. E. & Gossler, A. The mouse rib-vertebrae mutation is a hypomorphic Tbx6 allele. *Mech. Dev.* **119**, 251–256 (2002).

72. Batourina, E. et al. Distal ureter morphogenesis depends on epithelial cell remodeling mediated by vitamin A and Ret. *Nat. Genet.* **32**, 109–115 (2002).

73. Batourina, E. et al. Vitamin A controls epithelial/mesenchymal interactions through Ret expression. *Nat. Genet.* **27**, 74–78 (2001).

74. Batourina, E. et al. Apoptosis induced by vitamin A signaling is crucial for connecting the ureters to the bladder. *Nat. Genet.* **37**, 1082–1089 (2005).

75. Harambat, J., van Stralen, K. J., Kim, J. J. & Tizard, E. J. Epidemiology of chronic kidney disease in children. *Pediatr. Nephrol.* **27**, 363–373 (2012).

76. Westland, R., Kurvers, R. A., van Wijk, J. A. & Schreuder, M. F. Risk factors for renal injury in children with a solitary functioning kidney. *Pediatrics* **131**, e478–e485 (2013).

77. Westland, R., Schreuder, M. F., Bokenkamp, A., Spreeuwenberg, M. D. & van Wijk, J. A. Renal injury in children with a solitary functioning kidney: the KIMONO study. *Nephrol. Dial. Transplant.* **26**, 1533–1541 (2011).

78. Westland, R., Schreuder, M. F., van Goudoever, J. B., Sanna-Cherchi, S. & van Wijk, J. A. Clinical implications of the solitary functioning kidney. *Clin. J. Am. Soc. Nephrol.* **9**, 978–986 (2014).

79. Verbitsky, M. et al. Genomic disorders and neurocognitive impairment in pediatric CKD. *J. Am. Soc. Nephrol.* **28**, 2303–2309 (2017).

80. Cooper, G. M. et al. A copy number variation morbidity map of developmental delay. *Nat. Genet.* **43**, 838–846 (2011).

81. Greenway, S. C. et al. De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat. Genet.* **41**, 931–935 (2009).

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**, 81–86 (2008).

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**, 1466–1471 (2008).

85. Sanders, S. J. et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).

86. Sebat, J. et al. Strong association of de novo copy number mutations with autism. *Science* **316**, 445–449 (2007).

87. Stefansson, H. et al. Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232–236 (2008).

88. Yu, L. et al. De novo copy number variants are associated with congenital diaphragmatic hernia. *J. Med. Genet.* **49**, 650–659 (2012).

89. Mannik, K. et al. Copy number variations and cognitive phenotypes in unselected populations. *JAMA* **313**, 2044–2054 (2015).

90. Golzio, C. & Katsanis, N. Genetic architecture of reciprocal CNVs. *Curr. Opin. Genet. Dev.* **23**, 240–248 (2013).

91. Golzio, C. et al. KCTD13 is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363–367 (2012).

92. Gandelman, K. Y., Gibson, L., Meyn, M. S. & Yang-Feng, T. L. Molecular definition of the smallest region of deletion overlap in the Wolf-Hirschhorn syndrome. *Am. J. Hum. Genet.* **51**, 571–578 (1992).

93. Driscoll, D. A., Budarf, M. L. & Emanuel, B. S. A genetic etiology for DiGeorge syndrome: consistent deletions and microdeletions of 22q11. *Am. J. Hum. Genet.* **50**, 924–933 (1992).

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).

Acknowledgements

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.-J.N., T.Y.L., D.A. and 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 for the design of the study. All other authors (A.M., M.B., C.K., A.V., S.S., B.H.K., M.M., J.Y.Z., P.L.W., E.L.H., A.C., G.P., L.G., V.M., G.M., M.G.,

Additional information

[1]Division of Nephrology, Department of Medicine, Columbia University, New York, NY, USA. [2]Department of Pediatric Nephrology, Amsterdam UMC, Amsterdam, the Netherlands. [3]Department of Urology, Columbia University College of Physicians and Surgeons, New York, NY, USA. [4]University of Michigan School of Medicine, Department of Pediatrics-Nephrology, Ann Arbor, MI, 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, PA, USA. [7]Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, 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, CA, USA. [10]Institute for Genomic Medicine, Columbia University Medical Center, New York, NY, 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, 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 Vanvitell", 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, Katowice, Poland. [27]Department of Pediatrics and Nephrology, Medical University of Warsaw, 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, Ireland. [30]National Children's Research Centre, Our Lady's Children's Hospital Crumlin, Dublin, Ireland. [31]National Children's Hospital Tallaght, Dublin, Ireland. [32]University College Dublin UCD School of Medicine, Our Lady's Children's Hospital Crumlin, Dublin, 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, PA, USA. [34]Department of Pediatrics, University of Missouri-Kansas City School of Medicine, Division of Nephrology, Children's Mercy Kansas City, Kansas City, MO, USA. [35]University Children's Hospital, Medical Faculty of Skopje, Skopje, Macedonia. [36]University Clinic for General, Visceral and Transplantation Surgery, University of Heidelberg, Heidelberg, Germany. [37]Department of Medicine and Surgery, University of Parma, Parma, Italy. [38]Renal Division, Hospital Clinic, IDIBAPS, University of Barcelona, 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, NY, USA. [41]Mount Sinai Medical Center, Kravis Children's Hospital, New York, NY, USA. [42]Division of Pediatric Nephrology, University of New Mexico Children's Hospital, Albuquerque, NM, USA. [43]Division of Pediatric Nephrology, Department of Pediatrics, Columbia University, New York, NY, 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, NY, 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, TX, 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, NY, USA. [51]These authors contributed equally: Miguel Verbitsky, Rik Westland.
*e-mail: vep1@cumc.columbia.edu; clm20@cumc.columbia.edu; ag2239@cumc.columbia.edu; ss2517@cumc.columbia.edu

Methods

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 from the participant and/or their parents or guardians was obtained according to the protocols of the local IRBs.

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

Study subjects. CAKUT cases consisted of 2,824 affected, unrelated individuals with different subphenotypes across the entire CAKUT spectrum. Individuals were recruited in the United States, Europe and Brazil (overview of recruitment sites in Supplementary Fig. 1). The baseline characteristics of the study cohort are shown in Supplementary Table 1. One in every four cases presented with more than one CAKUT phenotype (that is, complex CAKUT), whereas a substantial 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 with 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 with 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 were 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. 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 by 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 in 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 KA[26].

CNVs in subjects were compared with those in controls and known CNV coordinates and were annotated with RefGene (see URLs) and curated sets of genes with 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 ≥70%. All reported CNVs were visually inspected in 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[26–28,35,36]. A CNV was classified as likely pathogenic if it (i) intersected at least one exon, (ii) was at least 100 kb in size, (iii) had a frequency in controls of at most 0.02%, (iv) 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: (i) had ≥70% overlap with a reported pathogenic or likely pathogenic CNV in the ISCA database, (ii) intersected a causative autosomal-dominant gene for CAKUT in humans or mice, and/or (iii) was the reciprocal of a known GD-CNV (coordinates with ≥70% overlap). Known GD-CNVs and likely pathogenic CNVs together were termed '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- and hemizygous deletions of loci underlying a recessive disorder for human disease with a renal phenotype.

CNV burden analysis. We restricted burden analyses to autosomal CNVs ≥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 on the basis of PCA, to avoid including CNVs that might be relatively common within ancestry groups represented in controls. Before the CNV burden was determined, seven cases (five KA, five vesicoureteral reflux and one lower urinary tract malformations) and eight controls were removed because they were unmatched outliers in PCA, thus yielding a dataset of 2,817 cases and 21,490 controls for burden analyses.

Whole-exome sequencing. Whole-exome sequencing was performed on 23 patients with CAKUT with 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 with a consistent alignment and variant-calling pipeline consisting of primary alignment with bwa-0.5.10, duplicate removal with Picard tools, index realignment and variant calling with GATK 3.6 and variant annotation with snpEff-3.3, AnnoVar (see URLs) and SeattleSeq (see URLs), with Ensembl-GRCh37.73 annotations.

After variant calling, the 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 >8×, minimal genotype quality score >30) with a minor-allele frequency of <1% (autosomal-recessive model) in the ExAC database and in 3,653 ancestrally matched available in-house controls and with a combined annotation-dependent depletion (CADD) score ≥20 (that is, 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 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 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 has deletion of exon 2 and part of exon 3 and has an *H2B-EYFP* fusion gene inserted in frame into exon 1 (ref. [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, and noon on the day of the plug was considered E0.5. Embryos were genotyped via PCR with the following primers pairs for *Tbx6[tm2Pa]* (i) 5′-GTACCATCCACGAGAGTTGTAC-3′;
(ii) 5′-GGGAAGAATGAGGATCCAGG-3′, to obtain a 220-bp wild-type allele fragment; (iii) 5′-ATTGCACGCAGGTTCTCCGG-3′; and (iv) 5′-GTCACGACGAGATCCTCGCC-3′, to obtain a 550-bp mutant allele fragment. For *Tbx6[rv]*: (i) 5′-CTCGCAGCTTCACTAGTCC-3′ and (ii) 5′-GTGTCTGGCGTATCAGCTCA-3′, to obtain a 322-bp wild-type allele fragment and 504-bp mutant allele fragment.

We analyzed different allele combinations for the *Tbx6*-null (*Tbx6[tm2Pa]*) and hypomorphic (*Tbx6[rv]*) alleles. For H&E staining: embryos were fixed overnight in 4% paraformaldehyde and dehydrated through a graded ethanol series (25–50–75–100%), then washed with xylene. The embryos were incubated in two changes of paraffin wax at 55 °C under vacuum. The embryos were then embedded in paraffin wax and sectioned at 10 µm. Sections were dewaxed in xylene, rehydrated through ethanol, stained with Harris H&E-Y and dehydrated before mounting in Cytoseal. For immunostaining, embryos were fixed in 4% araformaldehyde. Tissues were dehydrated into 100% methanol postfixation. For sectioning, tissues were embedded in paraffin or frozen in OCT and sectioned at various thicknesses (5–20 µm). Paraffin sections were deparaffinized with Histoclear and rehydrated via an ethanol series. Cryosections were prepared for immunolabeling by removal of OCT by washing in PBS. Antibodies were as follows: Cdh1, goat, R&D AF741, 1:400 dilution; Pax2, rabbit, Zymed 716000, 1:70.

Statistical analysis. All burden metrics calculations, statistical tests and plots were performed or generated in R v3.1 software (see URLs) and its dplyr (see URLs), survival (see URLs) and ggplot2 (ref. [99]) packages, and the Python v2.7 language (see URLs) and its packages pandas[100] and lifelines (see URLs). Unadjusted *P* values were reported for all tests. Summary statistics (mean, standard deviation, median and interquartile range) were calculated for CNV metrics such as size and total span, and the nonparametric Wilcoxon test was used for comparison between cases and controls. Proportions were compared with a two-sided Fisher's exact test. Kaplan–Meier survival curves were calculated to represent the largest CNV per genome, and a nonparametric log-rank test was used to compare the survival distributions between cases and controls. Logistic regression was used to test whether the number of genes intersected by large and rare CNVs detected (quantitative predictor) was associated with the case–control status (outcome). The reciprocal cumulative distributions of the number of intersected genes in cases versus controls were plotted.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

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

Data availability

Raw data that support the findings of this study will in part be available from the corresponding authors upon reasonable request and are 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.

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**, 1665–1674 (2007).
96. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
97. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
98. Fasel, D., Verbitsky, M. & Sanna-Cherchi, S. CNVkit: software tools for analyzing genomic structural variants. *J. Am. Soc. Nephrol.* **26** (Abstract edition), 443A (2015).
99. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer, New York, 2009).
100. McKinney, W. Data Structures for Statistical Computing in Python. in *Proc. 9th Python in Sci. Conference* (eds. van der Walt, S. & Millman, J.) 51–56 (Austin, Texas, USA, 2010).

nature research

Corresponding author(s):   Dr. Simone Sanna-Cherchi

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Olympus cellSense Standard 1.2
Data analysis	Commercial software used: Illumina GenomeStudio and GenomeViewer, Affymtetrix Power Tools and Chromosome Analysis Suite, Sequencher. Open source software used: ATAV, PennCNV, Plink, Python2.7, smartPCA, R.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was deemed to be appropriate based on the fact we were able to detect statistically significant differences and effects in similar studies, previously published, involving smaller cohorts. The present study reports the largest sample sizes to date for these phenotypes.
Data exclusions	Cases with clinically recognizable chromosomal abnormalities (e.g. Down syndrome) were excluded. Pre-established QC criteria were applied to the data (e.g. call rate, PennCNV confidence scores).
Replication	Findings on 16p11.2 locus were replicated in the CHOP dataset and the DECIPHER data.
Randomization	n/a
Blinding	Investigators were blind to genotypes in all mouse experiments.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☐ Unique biological materials
☐	☐ Antibodies
☐	☐ Eukaryotic cell lines
☐	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☐	☐ ChIP-seq
☐	☐ Flow cytometry
☐	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All materials (e.g. antibodies) are available from standard commercial sources.

Antibodies

Antibodies used	Cdh1: Goat R&D AF741; Pax2: Rabbit Zymed 716000
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	*State the source of each cell line used.*
Authentication	*Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.*

Mycoplasma contamination	*Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.*
Commonly misidentified lines (See ICLAC register)	*Name any commonly misidentified cell lines used in the study and provide a rationale for their use.*

Palaeontology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*
Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mice carrying the null expression reporter alleleTbx6tm2Pa and B6L-Tbx6rv/J mice can be purchased from The Jackson Laboratory;
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*

Human research participants

Policy information about studies involving human research participants

Population characteristics	Described in supplementary material.
Recruitment	Described in supplementary material.

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*

Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

Flow Cytometry

Plots

Confirm that:

☐ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*
Volume censoring	*Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings	*Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*
Effect(s) tested	*Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	*Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*
Correction	*Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity	*Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*
Graph analysis	*Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*
Multivariate modeling and predictive analysis	*Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*