

## ARTICLE OPEN



# Clinical exome sequencing efficacy and phenotypic expansions involving non-isolated congenital anomalies of kidney and urinary tract (CAKUT+)

E. Andres Rivera-Munoz<sup>1,2</sup> , Xiaonan E. Zhao<sup>1,3</sup> , Jill A. Rosenfeld<sup>1,3</sup> , Pamela N. Luna<sup>1</sup>, Chad A. Shaw<sup>1</sup>, Jennifer E. Posey<sup>1,4</sup> and Daryl A. Scott<sup>1,3,4</sup> ✉

© The Author(s) 2025

Congenital Anomalies of Kidney and Urinary Tract (CAKUT) can occur in isolation or in conjunction with one or more non-CAKUT associated congenital anomalies or neurodevelopmental disorders (CAKUT+). A molecular cause is not identified in most individuals with CAKUT+. This is due, in part, to uncertainty regarding the efficacy of genetic testing and an incomplete understanding of the genes that cause CAKUT+. Here, we use data from 515 individuals with CAKUT+ ( $n = 500$ ) or isolated CAKUT ( $n = 15$ ) to determine the efficacy of clinical exome sequencing (cES) and to identify new phenotype expansions that involve CAKUT. We determined that cES established a molecular diagnosis in 27.4% (141/515) of individuals in this cohort. No statistically significant difference in efficacy was seen with regards to age, sex, CAKUT phenotype, or associated organ system abnormality. Only 3.5% (5/144) to 14.6% (21/144) of the individual diagnoses made in our cohort could have been identified using one of four clinically available CAKUT gene panels. We then used a machine-learning approach to confirm that *PHIP* is a CAKUT gene and to implicate *ADNP* and *SETD5* genes associated with an increased risk of CAKUT. These findings lead us to conclude that cES should be considered in individuals with CAKUT+ for whom a molecular diagnosis has not been identified, that cES has the potential to identify many diagnoses in individuals with CAKUT+ that would be missed using a CAKUT gene panel, and that individuals with *ADNP*-, *PHIP*-, and *SETD5*-related disorders may present with CAKUT phenotypes.

*European Journal of Human Genetics*; <https://doi.org/10.1038/s41431-025-01929-3>

## INTRODUCTION

Congenital Anomalies of Kidney and Urinary Tract (CAKUT) affect 4–100 per 10,000 newborns and account for 20–30% of pediatric malformations [1–3]. The spectrum of CAKUT phenotypes is highly heterogeneous [4–7]. While CAKUT may manifest in isolation, it frequently is accompanied by one or more non-CAKUT associated congenital anomalies or neurodevelopmental disorders (CAKUT+) [8–10].

CAKUT remains a leading cause of kidney failure in children and young adults [6, 8]. Undiagnosed CAKUT in children and young adults also plays a role in susceptibility to renal disease later in life [8, 11–13]. Identifying a molecular etiology for CAKUT+ in an individual has the potential to change medical management and is a prerequisite for accurate genetic counseling and prognostication [8, 14–16].

Over 100 genes have been implicated in monogenic forms of CAKUT and CAKUT+, and clinical ES (cES) has become a commonly ordered genetic test for individuals with a wide range of medical conditions [5, 9, 13, 15, 17, 18]. However, cES is not universally ordered for individuals with CAKUT+ who do not have a molecular diagnosis. Limited information regarding the diagnostic efficacy of cES in CAKUT+, and an incomplete understanding of the genes that

contribute to CAKUT+ phenotypes, has likely limited the use of cES among these patients.

In the present study, we sought to address these deficiencies by evaluating the diagnostic efficacy of cES in individuals with CAKUT and CAKUT+, determining if a similar efficacy could have been achieved using commercially available CAKUT gene panels, and identifying novel phenotypic expansion involving CAKUT using clinical and molecular data from a large clinical cohort.

## SUBJECTS AND METHODS

### Cohort Identification and diagnostic review

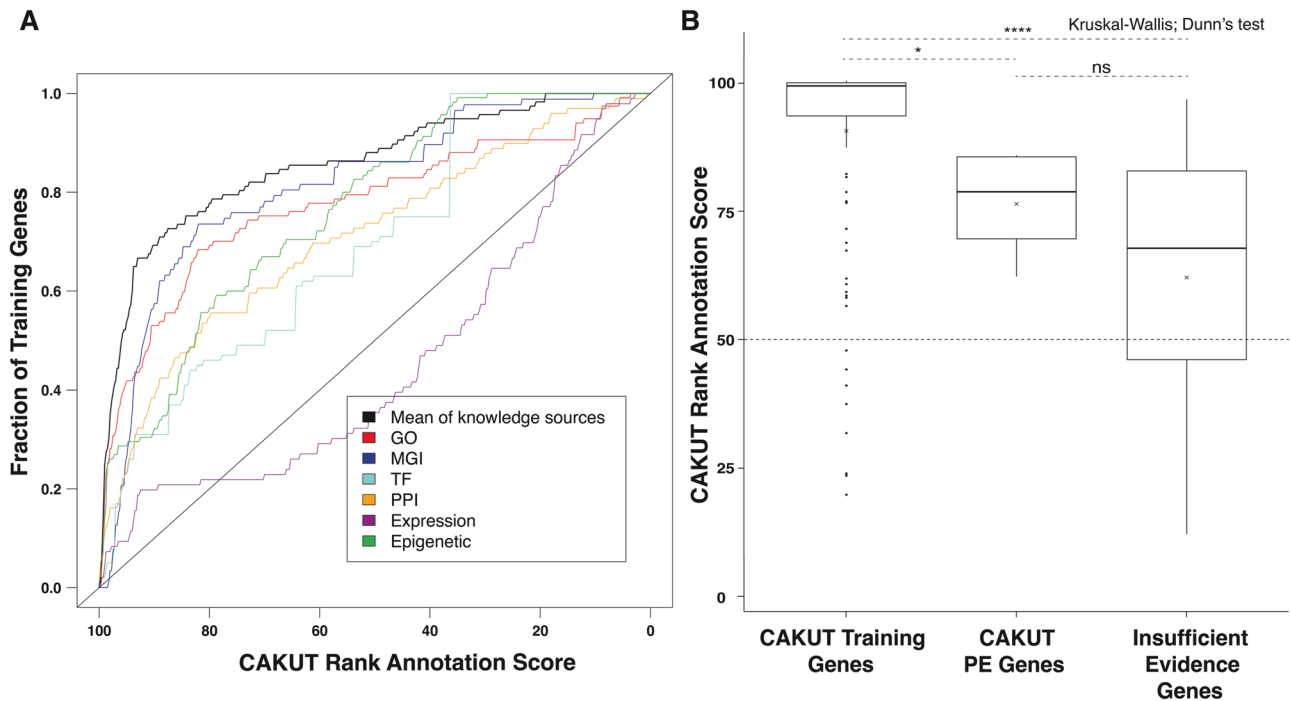
We identified 515 individuals (CAKUT\_P1–P515) with isolated CAKUT ( $n = 15$ ) or CAKUT+ ( $n = 500$ ) referred to Baylor Genetics for cES (Table S1). Clinical phenotypes were annotated with Human Phenotype Ontology (HPO) terms [19–22] and manually reviewed and grouped into seven categories: renal agenesis, renal duplication, lower urinary tract obstruction or anomalies, obstructive ureteral phenotypes, renal ectopy/malrotation/fusion, vesicoureteral reflux, multicystic dysplastic kidney, or multiple categories [2, 13].

Each case was reviewed to establish a level of confidence in the molecular diagnosis (definitive, probable, or provisional) based on the assessed pathogenicity of the molecular findings in the genetics report,

<sup>1</sup>Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA. <sup>2</sup>Genetics and Genomics Graduate Program, Baylor College of Medicine, Houston, TX, USA. <sup>3</sup>Baylor Genetics, Houston, TX, USA. <sup>4</sup>These authors contributed equally: Jennifer E. Posey, Daryl A. Scott. ✉email: [dscott@bcm.edu](mailto:dscott@bcm.edu)

Received: 8 February 2025 Revised: 26 June 2025 Accepted: 19 August 2025

Published online: 06 September 2025



**Fig. 1** Ranking all RefSeq genes based on their similarity to a set of known CAKUT genes. A set of 117 genes clearly associated with CAKUT (Table S2) which were used to train a previously published machine learning algorithm that integrates annotation data from various genome-scale knowledge sources [24]. **A** A leave-one-out validation was performed, and the resulting receiver operating characteristic (ROC) style curves were generated for each knowledgebase (colored lines) and an omnibus curve (black line) was generated using fit data from all knowledge sources. The area under each curve demonstrates performance above chance (black diagonal line). **B** Box plots comparing the CAKUT-specific rank annotation scores of genes in the training set, 4 high priority/phenotype expansion (PE) CAKUT candidate genes, and 55 other CAKUT candidate genes with insufficient evidence to suggest a phenotype expansion. Comparisons between groups were performed using a Wilcoxon rank-sum test to determine whether the distribution of scores significantly differed between these groups. GO Gene Ontology, MGI Mouse Genome Informatics, TF Transcription factor, PPI Protein-protein interaction. NS:  $p > 0.05$ . \* $p \leq 0.05$ . \*\* $p \leq 0.01$ . \*\*\* $p \leq 0.001$ . \*\*\*\* $p \leq 0.0001$ .

expected mode of inheritance, and available phenotypic information as previously described (Table S1, Supplemental Materials) [23].

### Generating CAKUT-specific rank annotation scores

We employed a previously published machine learning algorithm to generate CAKUT-specific rank annotation scores [24]. This tool integrates annotations from various genome-scale knowledge sources to construct a pattern in genomic feature space based on a set of phenotype-associated training genes, then ranks all RefSeq genes with respect to their similarity to that pattern using quantitative similarity metrics [25–31]. Based on each gene's annotation similarity, the algorithm generates a gene-, phenotype-specific rank annotation score between 0 and 100%.

A curated list of 117 known CAKUT genes from the literature and commercial diagnostic CAKUT panels (Table S2) was used to train the algorithm. The resulting omnibus curve, produced using fit data from all knowledge sources, validates that the algorithm distinguishes between the CAKUT-associated genes in the training set and all other RefSeq genes better than random chance (Fig. 1A).

We generated CAKUT-specific rank annotation scores for all RefSeq genes (Table S3). The rank annotation scores of the training set genes ranged from 20% to 100% with a median score of 99% (Fig. 1B), compared to a median score of 50% across all RefSeq genes, demonstrating, again, that the algorithm can distinguish between the 117 CAKUT-associated genes and all other RefSeq genes.

For further details, please see the Supplemental Subjects and Materials.

## RESULTS

### Efficacy of cES in individuals with CAKUT phenotypes

We identified 515 individuals with CAKUT who were referred to a clinical diagnostic laboratory for cES (Table S1). Most (500/515; 97.1%) had one or more non-CAKUT associated congenital anomalies or neurodevelopmental disorders and were designated

as CAKUT+ (Table 1). These 515 individuals were divided into those with variants in genes putatively relevant to their phenotypes which were reported back to physicians (positive,  $n = 203$ ), and those without putatively relevant variants identified (negative,  $n = 312$ ) (Fig. S1). The distributions of age at genetic testing and sex between positive and negative cases were not significantly different ( $p$  value = 0.9004, T-test, and  $p$  value = 0.4075, Chi-square test, respectively).

In our cohort, 27.4% (141/515) of individuals had a variant(s) associated with a definitive or probable molecular diagnosis (Fig. S1). Of these 141 individuals, 3 had dual definitive and/or probable molecular diagnoses for a total of 144 diagnoses. An additional 12.0% (62/515) of individuals in our cohort carried a variant(s) associated with a provisional diagnosis.

We further evaluated the diagnostic rate of cES across age-at-testing groups, CAKUT phenotypes, and additional affected organ systems (CAKUT+) (Fig. 2; Table 1). Most individuals in our cohort had cES during childhood (2–11 years old). We did not observe a statistically significant difference between receiving a definitive or probable diagnosis between age groups (Table 1;  $p$  value = 0.0680, Chi-square).

In our cohort, 50.5% of individuals were male, 41.4% were female, and 8.2% were unknown/prenatal cases. We did not observe a significant difference in the likelihood of receiving definitive or probable diagnoses compared to receiving provisional or unsolved diagnoses by reported sex (Table 1;  $p$  value = 0.0572, Chi-Square).

Next, we compared the odds of receiving definitive or probable diagnoses between individual CAKUT phenotypes compared to all other CAKUT cases and did not observe a statistically significant difference by CAKUT phenotype (Fig. 2A,  $p$  value = 0.3714;

**Table 1.** Overview of 515 CAKUT cases referred to a diagnostic laboratory for cES.

	Definitive	Probable	Provisional	Unsolved	Significance ( $P < 0.05$ )
Total (%)	94 (18.3)	47 (9.1)	62 (12.0)	312 (60.6)	
Mean age (SD)	5.1 (5.7)	7.7 (10.6)	4.0 (4.9)	5.3 (7.3)	NS
Age Group (%)					NS
Prenatal	5 (11.4)	1 (2.3)	8 (18.2)	30 (68.2)	
Neonatal (<1 mo)	8 (16.3)	1 (2.0)	7 (14.3)	33 (67.3)	
Infant (1 mo–1 yr)	16 (16.0)	8 (8.0)	10 (10.0)	66 (66.0)	
Toddler (1–2 yr)	9 (17.6)	7 (13.7)	8 (15.7)	27 (52.9)	
Childhood (2–11 yr)	40 (22.1)	21 (11.6)	23 (12.7)	97 (53.6)	
Adolescent (11–21 yr)	13 (18.6)	6 (8.6)	6 (8.6)	45 (64.3)	
Adulthood (21+ yr)	2 (10.5)	3 (15.8)	0 (0.0)	14 (73.7)	
Unknown	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	
Sex (%)					NS
Female	39 (18.3)	20 (9.4)	30 (14.1)	124 (58.2)	
Male	51 (19.6)	26 (10)	24 (9.2)	159 (61.2)	
Unknown (Fetal)	4 (9.5)	1 (2.4)	8 (19.0)	29 (69.0)	

SD = Standard deviation, NS = Not significant, mo = months, yr = years.

Kruskal-Wallis Rank Sum). Finally, there was no statistical difference in the odds of receiving a definitive or probable diagnosis between individuals with isolated CAKUT vs. CAKUT+ ( $p$  value = 0.0801; Fisher's exact), or between individuals with CAKUT+ based on other organ systems affected ( $p$  value = 1; Kruskal-Wallis Rank Sum; Fig. 2B).

#### Coverage of commercially available CAKUT gene panels

In our CAKUT cohort, cES revealed 144 definitive or probable diagnoses including dual molecular diagnoses in three individuals. Based on gene coverage, between 3.5% (5/144) and 14.6% (21/144) of these diagnoses could have been made using one of four commercially available CAKUT gene panels. From these 144 definitive or probable diagnoses, 120 unique genes were involved in our CAKUT cohort. Between 4.2% (5/120) and 8.3% (10/120) of these genes were covered in each of the four commercially available CAKUT gene panels.

#### Review of reported disease genes association with CAKUT

We generated and reviewed CAKUT-specific rank annotation scores for the 267 genes listed in the reports of CAKUT individuals (Table S3). A subset of our cohort had variants in genes known to cause CAKUT based on data from commercial panels, OMIM/HPO, or the literature. Among these 110 known CAKUT genes (Fig. 3), *ACTB*, *BRIP1*, *CDK13*, *CDKL5*, *CREBBP*, *CTNNB1*, *EP300*, *FOXP1*, *KAT6B*, *KMT2D*, *PTPN11*, *SON*, and *TFAP2A* were associated with a definitive or probable diagnosis in 2 or more individuals in our cohort. Despite being clearly associated with CAKUT, *BRIP1*, *CDKL5*, *CREBBP*, *CTNNB1*, *EP300*, and *FOXP1* do not have CAKUT phenotypes listed in the clinical phenotypes associated with their various disorders in OMIM (December 2024) [32].

The remaining variants reported back to physicians were in 157 genes for which an association with CAKUT had not yet been clearly established. These genes were designated as CAKUT candidate genes.

#### CAKUT phenotypic expansions

We evaluated each of the CAKUT candidate genes to determine whether there was sufficient evidence to support a phenotype expansion involving CAKUT based on: 1) whether the variant(s) in the gene led to a definitive/probable diagnosis in one or more patients in our CAKUT cohort, 2) the gene's CAKUT-specific

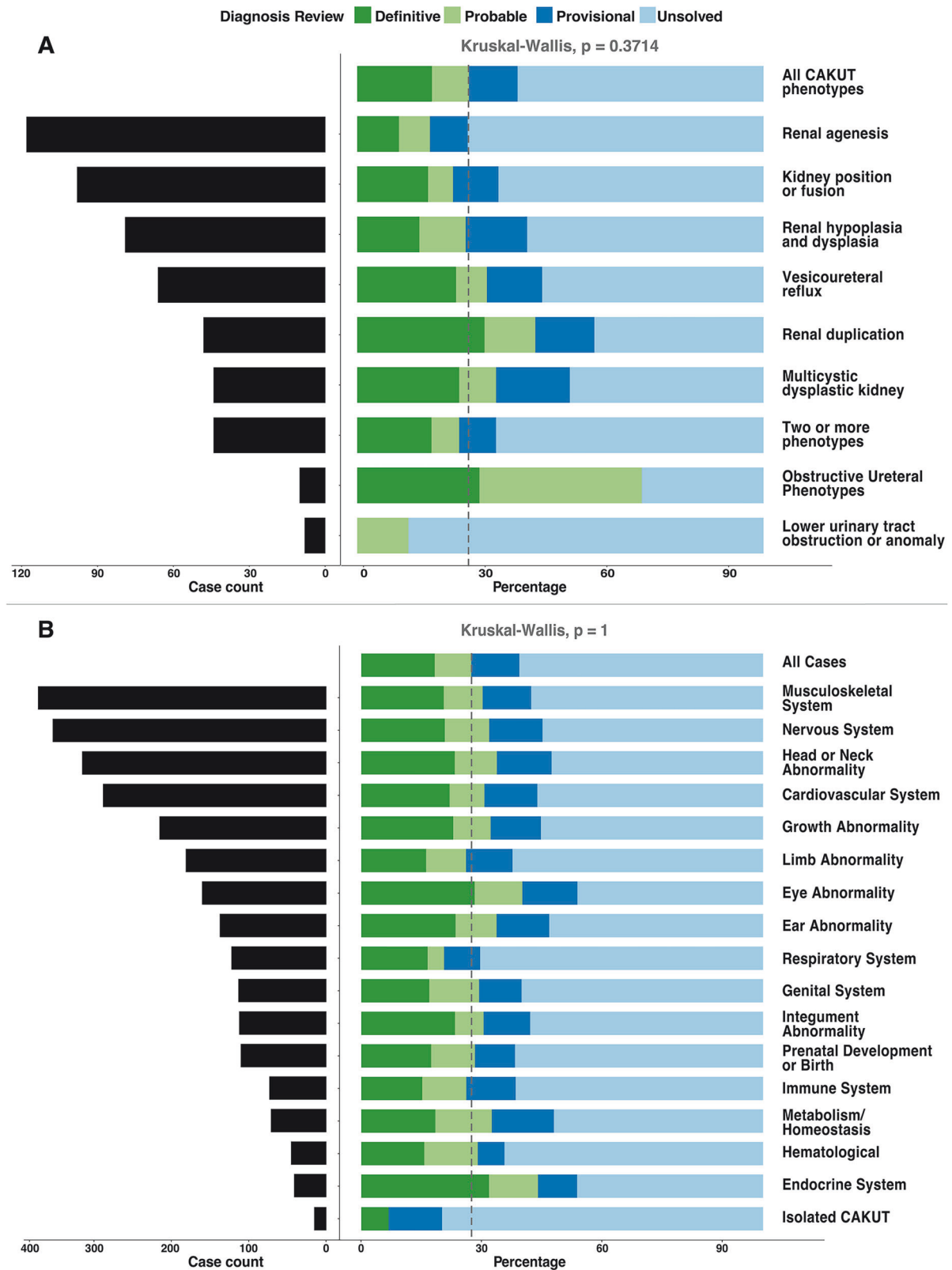
rank annotation score, 3) published reports in which CAKUT phenotypes were documented in individuals carrying putative pathogenic variants in the gene, 4) the presence of CAKUT phenotypes in transgenic mouse models involving the gene's homolog, and/or 5) expression of the gene's mouse homolog in the developing and/or adult mouse urinary tract.

This evaluation yielded four high-priority CAKUT candidate genes—*ADNP*, *CBL*, *PHIP*, and *SETD5*. We then performed gene-specific queries of the Baylor Genetics clinical database for each high priority candidate gene to identify additional individuals with likely damaging/pathogenic variants and renal/urinary phenotypes. This review identified 5 individuals (CAKUT\_P516-P520; Table S1) not included in our primary cohort whose molecular and clinical data provide additional support for phenotypic expansions involving CAKUT. Other individuals—CAKUT\_P349 (*ADNP*), CAKUT\_P337 and CAKUT\_P303 (*PHIP*), and CAKUT\_P306 (*SETD5*) with variants in these genes were in the original cohort. However, these variants were not reported back to the ordering physician and were not taken into consideration when determining cES efficacy.

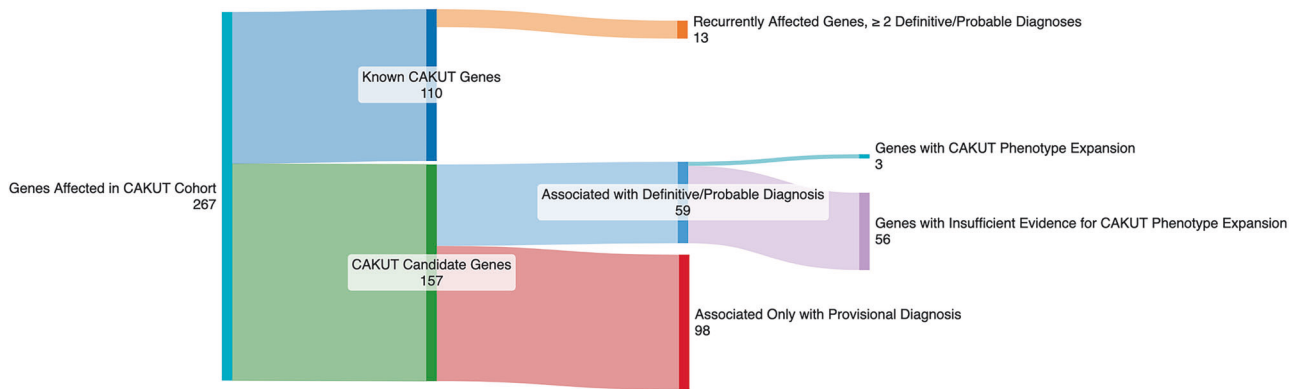
The homologs of *ADNP*, *CBL*, *PHIP*, and *SETD5* are expressed in the mouse urinary tract. In addition, these genes had 1) highly positive ( $\geq 85\%$ ) CAKUT-specific rank annotation scores and  $\geq 3$  definitive/probable cases reported in our cohort or in the literature, or 2) a positive ( $> 50\%$ ) CAKUT-specific rank annotation score and  $\geq 6$  definitive/probable cases reported in our cohort or in the literature. The CAKUT-specific rank annotation scores for these genes ranged from 62.2% to 85.6% with a mean score of 76.2% and median score of 78.6% (Fig. 1B).

We then determined the incidence of CAKUT in individuals in the Baylor Genetics database whose cases were solved by variants in each of our high-priority CAKUT candidate genes. This was then compared to a high estimate of CAKUT incidence in the general population; 1:100. As shown in Table 2, the incidence of CAKUT in individuals with *ADNP*-, *SETD5*- and *PHIP*-related disorders in the Baylor Genetics clinical database were significantly higher than the rate of CAKUT in the general population with  $P$  values of 0.0444, 0.0079, and 0.0002, respectively. We conclude that *ADNP*, *SETD5*, and *PHIP* are CAKUT associated genes (Table 3; Fig. 3, 'Genes with CAKUT Phenotypic Expansion').

In contrast, the incidence of CAKUT in individuals with *CBL*-related disorder was not statistically different from that of the



**Fig. 2 Diagnostic rate of cES for CAKUT. A** Diagnostic rate of 515 individuals with CAKUT by diagnostic category—definitive, probable, provisional, and unsolved—and by distinct CAKUT subtype. **B** Diagnostic rate of 500 CAKUT+ (syndromic CAKUT) and 15 isolated CAKUT individuals by specific organ system involvement.



**Fig. 3 Sanke plot demonstrating the breakdown of 267 genes carrying reported variants from our CAKUT cohort.** Of the 267 genes carrying variants reported back to physicians, 110 were genes known to cause CAKUT and 13 of these genes were altered in  $\geq 2$  individuals in our cohort. Of the remaining 157 CAKUT candidate genes, 98 were only associated with provisional diagnoses in our cohort. Of the 59 associated with a definitive or probable diagnosis, 3 had sufficient evidence to support a phenotypic expansion involving CAKUT, and 56 did not.

general population ( $P = 0.0758$ ). Hence, we consider it only a high priority CAKUT candidate gene (Fig. 3: 'Insufficient Evidence').

#### Other CAKUT candidate genes

We identified 55 other CAKUT candidate genes associated with a definitive or probable diagnoses in our cohort. The CAKUT-specific rank annotation scores for these genes ranged from 12.5% to 96.4% with a mean score of 61.7% and a median score of 67.7% (Fig. 1B). Within this group, *MED13L* and *WNT7A* were of particular interest. We performed gene-specific queries of the Baylor Genetics clinical database for each of these genes and identified two individuals in our original cohort with variants in *MED13L*—CAKUT\_P252 and CAKUT\_P385—which were not reported to their ordering physicians. However, even with these findings, there was insufficient evidence to support a phenotypic expansion involving CAKUT for *MED13L* or *WNT7A*.

The rank annotation scores of training genes were significantly higher compared to those of *ADNP*, *CBL*, *PHIP*, and *SETD5*, and the other 55 CAKUT genes with insufficient evidence for phenotypic expansion ( $P$  values = 0.039 and 0.0001, respectively, Dunn's test). However, there was no significant difference in the median scores between genes in the strong and insufficient evidence category ( $P$  value = 0.597).

## DISCUSSION

### Diagnostic efficacy of cES in individuals with CAKUT

cES has emerged as a powerful tool with the potential to provide a precise molecular diagnosis that informs clinical management and improves genetic counseling for affected individuals and their families [33–36]. The evaluation of cES diagnostic efficacy is key in supporting the decision-making process for physicians and setting expectations for patients and other stakeholders [37, 38].

In our CAKUT cohort, cES provided a definitive or probable diagnosis in 27.4% (141/515) of cases, consistent with previous retrospective studies [5, 15, 39]. When provisional diagnoses are included, the diagnostic yield of cES in our cohort increases to 39.4% (203/515). Most provisional cases stem from variants classified as variants of uncertain significance (VUS).

Most individuals in our cohort were referred for cES in childhood. We did not observe a difference in the diagnostic yield between different age groups. Similarly, the diagnostic yield did not differ significantly by specific CAKUT phenotype (Fig. 2A). However, we acknowledge that this may be due, in part, to the paucity of cES referrals for individuals in some categories, particularly obstructive ureteral phenotypes, and lower urinary tract obstruction or anomalies.

Similarly, we did not observe a significant difference between individuals with CAKUT and CAKUT+, likely due to the relatively low number of isolated CAKUT cases (2.9%, 15/515). We also note the possibility that the cES efficacy in isolated CAKUT cases may have been inflated due to selection bias, a failure to report phenotypes that could have resulted in reclassification as CAKUT+, or a positive family history that was not fully documented in their indications for testing.

Among those with CAKUT+, the most common organ system abnormalities were musculoskeletal (372/515, 72.2%). There was no significant difference in diagnostic yield by affected organ system (Fig. 2B).

### Comparisons of cES and CAKUT gene panel efficacy

The ability of cES to interrogate all disease-relevant genes, regardless of their known association with a specific phenotype, leads to an increased ability to make a molecular diagnosis in individuals with CAKUT when compared to a gene panel [40]. Specifically, we found that of the 144 definitive or probable diagnoses in our CAKUT cohort, only 3.5–14.6% (5–21/144) could have been made using one of four commercially available CAKUT gene panels. In an alternative comparison, we found that only 4.2–8.3% (5–10/120) of the genes implicated in a definitive or probable diagnosis in our cohort were covered in one of the four commercially available CAKUT gene panels.

Based on these results, we recommend that cES be considered in individuals with CAKUT+ for whom a molecular diagnosis has not been identified even if gene panel testing was negative. We would expect that the diagnostic yield of clinical genome sequencing (cGS) would exceed that of cES and could serve as a first-tier test replacing the use of both chromosome microarray (CMA) and cES [41, 42].

We note that the European Rare Kidney Disease Reference Network (ERKNet) Working Group on Kidney Malformations suggests that genetic testing for CAKUT should begin with a next generation sequencing gene panel or exome sequencing with applied filters that limit the sequence variant results to those found in a prioritized list of seven genes. If negative, the filters can be adjusted to report sequence variants in genes that have been associated with kidney phenotypes in humans or animal models [43]. If no causative gene is identified in these tests, they indicate that trio cES or cGS should be considered for patients for which there is a high suspicion for inherited kidney disease. Our results suggest that the recommendation to consider cES or cGS should be extended to all individuals with CAKUT+ regardless of the clinical suspicion for an inherited kidney disease. This suggestion seems especially appropriate given the high number of individuals



**Table 2.** Individuals with a molecular diagnosis, with and without CAKUT, in the Baylor Genetics clinical database.

Gene	Phenotypes (MIM#), Inheritance	Individuals with a definitive or probable diagnosis and CAKUT	Individuals with a definitive or probable diagnosis without CAKUT	Percentage with CAKUT	95% Confidence Interval	P value <sup>a</sup>
<i>ADNP</i>	Helsmoortel-van der Aa syndrome (615873), AD	2	12	14.3%	1.8 to 42.8%	0.0394 <sup>b</sup>
<i>CBL</i>	Noonan syndrome-like disorder with or without juvenile myelomonocytic leukemia (613563), AD	1	3	25.0%	0.6 to 80.6%	0.0758 (NS)
<i>PHIP</i>	Chung-Jansen syndrome (617991), AD	3	2	60.0%	14.7 to 94.7%	0.0002 <sup>b</sup>
<i>SETD5</i>	Intellectual disability, autosomal dominant 23, (615761), AD	3	11	21.4%	4.7 to 50.8%	0.0056 <sup>b</sup>

NS Not Significant.

<sup>a</sup>Each gene is compared to a general population estimate of CAKUT in 1/100 patients.<sup>b</sup>Significant = *P* value < 0.05.**Table 3.** Genes with sufficient evidence to support a phenotypic expansion involving CAKUT.

Gene	CAKUT-Specific Rank Annotation	Phenotypes (MIM#), Inheritance	CAKUT Cases (This Study)	CAKUT Cases (Literature)	Mouse Urinary Tract Expression (MGI)	Transgenic Mice with CAKUT Phenotypes? (MGI/DMDD/IMPC)
<i>ADNP</i>	85.6%	Helsmoortel-van der Aa syndrome (615873), AD	Total 4: (1D) Renal duplication (1Pv) Polycystic kidney disease (1Pv) Vesicoureteral reflux (1D) Bilateral hydronephrosis	Total 7: (1D) Renal agenesis <sup>48</sup> (6l) Renal anomalies <sup>51</sup>	Yes	Not Reported
<i>PHIP</i>	62.2%	Chung-Jansen syndrome (617991), AD	Total 5: (1D) Vesicoureteral reflux (1D) Horseshoe kidney, ureteropelvic junction obstruction (1Pr) Ureteral hypoplasia (1Pv) Renal malrotation (1Pv) Hydronephrosis	Total 12: (2D-1Pr) Vesicoureteral reflux <sup>57,58</sup> (1Pr) Renal agenesis <sup>59</sup> (1Pr) Ureterocele <sup>55</sup> (1D-2l) Renal hypoplasia <sup>57,60</sup> (1D) Ureterovesical junction obstruction <sup>57</sup> (1D-1l) Horseshoe kidney <sup>58</sup> (1Pr) Atrophic kidney <sup>58</sup>	Yes	Yes: Enlarged bladder
<i>SETD5</i>	85.2%	Intellectual disability, autosomal dominant 23, (615761), AD	Total 4: (1D) Renal agenesis (1Pv) Multicystic dysplastic kidney, renal dysplasia, vesicoureteral reflux (1D) Hydronephrosis (1D) Kidney abnormalities	Total 1: (1D) Posterior urethral valves <sup>63</sup>	Yes	Yes: Small kidneys

AD Autosomal dominant, D Definitive, DMDD Deciphering the Mechanisms of Developmental Disorders, I Insufficient information to determine diagnostic certainty, IMPC International Mouse Phenotyping Consortium, MGI Mouse Genome Informatics, MIM Online Mendelian Inheritance in Man, Pr Probable, Pv Provisional.

with CAKUT+ who were diagnosed based on the identification of a de novo variant (Table S1).

We feel that the potential benefits of identifying a molecular diagnosis through the use of cES or cGS in individuals with CAKUT + outweigh concerns that might be caused by the identification of incidental findings. Specifically, the benefits of identifying a molecular diagnosis can include the development of individualized medical care plans, access to emerging therapies, accurate genetic counseling, improved prognostication, psychological relief, increased levels of social and emotional support through engagement with those who share the molecular diagnosis, and the avoidance of medical waste in the form of unnecessary testing, imaging, and/or medical procedures [44].

### Recognition of CAKUT phenotypes associated with known disease genes

Understanding the full phenotypic range of known disease genes is crucial for delivering appropriate care to individuals for whom a molecular diagnosis has been made [45]. Online Mendelian Inheritance in Man (OMIM) has served as a valuable resource allowing physicians to quickly determine the phenotypes associated with disease genes and their respective genetic disorders [46]. It is unreasonable to expect that OMIM's clinical synopses will contain descriptions of all possible phenotypes associated with a particular gene/disorder.

Among the genes affected in more than 2 individuals in our cohort, we found 6 genes with clear evidence of an association with CAKUT but whose clinical synopses in OMIM do not contain CAKUT phenotypes. This underscores the importance of performing a detailed investigation to avoid inaccurately concluding that CAKUT phenotypes must be due to genetic factors other than the variants already identified in known disease genes.

### Phenotypic expansions involving CAKUT

A subset of individuals in our cohort had putatively damaging variants in genes that are not currently known to cause CAKUT. Among these CAKUT candidate genes, we identified three for which there is currently sufficient evidence to support a phenotypic expansion involving CAKUT: *ADNP*, *PHIP*, and *SETD5*.

#### ADNP

Helsmoortel-van der Aa syndrome (MIM #615873), an autosomal dominant disorder characterized by impaired intellectual development, autism spectrum disorder, motor delay, hypotonia, visual difficulties, congenital heart disease, gastrointestinal phenotypes, and facial dysmorphisms, is caused by heterozygous variants in activity-dependent neuroprotective homeobox (*ADNP*; MIM #611386) [47]. Three individuals in our original cohort had *ADNP* variants whose phenotypes provide support for a phenotypic expansion involving CAKUT. CAKUT\_P132 had a definitive diagnosis of Helsmoortel-van der Aa syndrome and duplicated kidney. CAKUT\_P517 and CAKUT\_P349 had provisional diagnoses of Helsmoortel-van der Aa syndrome, and polycystic kidney disease and vesicoureteral reflux, respectively. We identified a fourth individual in our clinical database, CAKUT\_P516, who also had a definitive diagnosis of Helsmoortel-van der Aa syndrome who had bilateral hydronephrosis.

Asegaonkar et al. described a fetus with a de novo c.2619\_2620delCA, p.(D873Efs\*7) pathogenic variant in *ADNP* who had a solitary, enlarged kidney [48]. This fetus also carried a c.1868C>T, p.(A623V) VUS (PM2, BP4) in *NR1P1*, a gene implicated in CAKUT (MIM# 618270) [49]. Although this VUS was inherited from the father, who did not have evidence for CAKUT by ultrasound, this case potentially represents an example of a blended phenotype, where both variants may be contributing to the development of CAKUT [50].

Van Dijck et al. summarized the phenotypes of a large cohort of individuals with *ADNP*-associated neurodevelopmental disorder

and found that 6 individuals out of 48 had "renal anomalies" which they described as "narrow ureters, bilateral vesicoureteral reflux that was surgically repaired." The *ADNP* variants present in these individuals were not specified [51].

*ADNP* has a high CAKUT-specific rank annotation score of 85.6%, and *Adnp* is expressed in the mouse urinary system from TS19 (E11) through TS28 (P4-Adult). The incidence of CAKUT among individuals with Helsmoortel-van der Aa syndrome in the Baylor Genetics clinical database is statistically higher than that of the general population (Table 3). This suggests that individuals with Helsmoortel-van der Aa syndrome can present with a variety of CAKUT phenotypes [26].

#### PHIP

Pleckstrin homology domain-interacting protein (*PHIP*; MIM# 612870) encodes multiple proteins through alternative splicing, stimulates cell proliferation, has anti-apoptotic activity, and plays a role in the regulation of cell morphology and cytoskeletal organization [52, 53]. Loss-of-function variants in *PHIP* are the cause of Chung-Jansen syndrome (CHUJANS; MIM# 617991) which is characterized by intellectual disability, behavioral problems, obesity, and dysmorphic features [54, 55].

We report five individuals with CAKUT who carry variants in *PHIP*. CAKUT\_P74 and CAKUT\_P337 have definitive diagnoses of CHUJANS, and CAKUT\_P205 has a probable diagnosis of CHUJANS. Their CAKUT phenotypes are vesicoureteral reflux, horseshoe kidney and left ureteropelvic junction obstruction, and left ureteral hypoplasia, respectively. Provisional diagnoses of CHUJANS were made in CAKUT\_P303 and CAKUT\_P518 who had a malrotated left kidney and hydronephrosis, respectively.

The incidence of CAKUT in individuals with CHUJANS in the Baylor Genetics database is statistically higher than that of the general population (Table 3). Data from the International Mouse Phenotyping Consortium (IMPC) suggest that 50% (2/4) of early adult male mice carrying a single null-allele of *Phip* had enlarged urinary bladders [56]. We also note that *Phip* is expressed in the mouse urinary system from TS19 (E11) through TS28 (P4-Adult) [26], and that *PHIP* has a positive CAKUT-specific rank annotation score of 62.2%. This led us to conclude that *PHIP* was a CAKUT gene.

During our investigation, Fallois et al. independently concluded that *PHIP* is a CAKUT gene based on novel and published cases of CAKUT associated with deleterious variants in *PHIP* [55, 57–60]. They also provide evidence of *PHIP*'s expression in the developing human and mouse kidney. Hence, our findings provide confirmatory evidence that *PHIP* is a CAKUT gene.

#### SETD5

Set domain-containing protein 5 (*SETD5*; MIM# 615743) is a methyltransferase that targets histone H3lys36 for trimethylation (H3K36me3) and thereby affects transcription in neural progenitor cells and their derivatives [61]. Intellectual developmental disorder, autosomal dominant 23 (IDDD23; MIM# 615761) is caused by heterozygous loss-of-function variants in *SETD5* [62].

In our original cohort, CAKUT\_P111 has a definitive diagnosis of IDDD23 and unilateral renal agenesis. CAKUT\_P306 has a provisional diagnosis of IDDD23 and a left multicystic dysplastic kidney, right-sided renal dysplasia, and a history of vesicoureteral reflux and hydronephrosis. In the Baylor Genetics clinical database we also identified CAKUT\_P519 and CAKUT\_P520 who have definitive diagnoses of IDDD23 and hydronephrosis and kidney abnormalities, respectively. Szczałuba et al. previously described a family segregating a p.(S973\*) [NM\_001080517.2] pathogenic variant in *SETD5* in which an affected male child had posterior urethral valves [63].

*Setd5* is expressed in the developing mouse urinary system from TS19 (E11) through TS28 (P4-Adult) [26], and the Deciphering the Mechanisms of Developmental Disorders (DMDD) consortium has reported small kidneys in mice that are heterozygous for a null

variant in *Setd5* [64]. These data, and SETD5's high CAKUT-specific rank annotation score of 85.6%, provide strong evidence to suggest that individuals with IDDD23 can present with CAKUT phenotypes.

### CAKUT candidate genes of interest

Although there was insufficient evidence to conclude that deleterious variants in *CBL*, *MED13L* and *WNT7A*, were associated with an increased risk of CAKUT, these genes remain CAKUT candidate genes of interest.

### CBL

Casitas B-lineage lymphoma protooncogene (*CBL*; MIM# 165360) is an E3 ubiquitin-protein ligase that plays a crucial role in cell signal transduction by targeting receptor tyrosine kinases (RTKs) for ubiquitination and degradation [65]. *CBL* variants have been associated with Noonan-like disorder with or without juvenile myelomonocytic leukemia (NSLL, MIM# 613563) also known as CBL syndrome [66]. This disorder is characterized by developmental delay, hematological abnormalities and dysmorphic facial features.

CAKUT\_P126 has renal duplication and a definitive diagnosis of NSLL. There are five other published individuals with NSLL and CAKUT phenotypes: one with unilateral renal agenesis and a c.1144A>G, p.(K382E) pathogenic variant [66], one with hydronephrosis and a c.1096-1G>T (previously reported as c.1098-1G>T) pathogenic splice variant [67], one with bilateral hydronephrosis and hydroureters and a c.1096-4\_1096-1del pathogenic variant [68], and identical twins with p.(Q358fs) pathogenic variants; one with hypoplastic kidneys and chronic severe vesicoureteral reflux, and the other with renal dysplasia/hypoplasia and severe ureteral reflux [67]. As with most germline *CBL* variants that cause NSLL, these variants are predicted to cause changes in the linker helix region and RING-finger domains of *CBL* [68].

*CBL*'s CAKUT-specific rank annotation score is 71.9%. Additionally, *Cbl* is expressed in the mouse urinary system from TS19 (E11) through TS28 (P4-Adult). Although we did not demonstrate an increased risk of CAKUT among individuals with NSLL in the Baylor Genetics clinical database, these data suggest that *CBL* should be considered a high-priority CAKUT candidate gene.

### MED13L and WNT7A

There is currently insufficient evidence to suggest that the other 55 CAKUT candidate genes can cause CAKUT phenotypes. However, their CAKUT-specific rank annotation scores of these genes are higher than would be expected by random chance (Fig. 1B). This suggests that this group of genes may be enriched for true CAKUT genes. Within this group, *MED13L* and *WNT7A* currently have the greatest amount of evidence in support of a phenotypic expansion involving CAKUT.

Impaired intellectual development and distinctive facial features with or without cardiac defects (MIM# 616789), also referred to as Asadollahi-Rauch syndrome (ARS), is caused by heterozygous variants in *MED13L*. CAKUT\_P147 has a definitive diagnosis of ARS and vesicoureteral reflux, and CAKUT\_P252 and CAKUT\_P385 have provisional diagnoses of ARS bilateral renal hypoplasia with end stage renal disease, and crossed fused renal ectopia, respectively. Three individuals with CAKUT phenotypes and de novo *MED13L* variants have been previously described [69–71]. *MED13L* has a positive CAKUT rank annotation score of 73.6%, and *Med13l* is expressed in the mouse urinary system from TS19 (E11) through TS28 (P4-Adult).

Autosomal recessive *WNT7A* variants associated with partial loss of function cause the abnormal limb development characteristic of Fuhrmann syndrome (FS; MIM# 228930), and autosomal recessive variants leading to a complete loss of function cause the more severe Al-Awadi/Raas-Rothschild/Schinzal phocomelia syndrome (AA/RRS; MIM# 276820) [72].

CAKUT\_P193 has a provisional diagnosis of FS, renal agenesis, and other phenotypes consistent with this disorder including an absent thumb, an absent radius, and hip dysplasia. Two individuals with phenotypes suggestive of AA/RRS and CAKUT have been previously described [73, 74]. *WNT7A* has a highly positive CAKUT rank annotation score of 92.5%, and *Wnt7a* is expressed in the mouse urinary system from TS21 (E12.5–14) to TS23 (E15) and possibly at TS28 (P4-adult).

### Clinical practice recommendations

Our findings demonstrate the diagnostic efficacy of cES in patients with CAKUT and broaden the phenotypic features associated with genes already known to cause human genetic disorders. We conclude that cES or cGS should be considered in individuals with CAKUT+ for whom a molecular diagnosis has not been identified, and that cES has the potential to identify many diagnoses in individuals with CAKUT+ that would be missed by using a commercial CAKUT panel. Our data also suggest that additional testing aimed at identifying an independent cause of CAKUT may not be warranted in individuals with an *ADNP*-, *PHIP*-, or *SETD5*-related disorder.

### DATA AVAILABILITY

The data generated during this study can be found within the published article and its supplementary files. All variants reported here have been submitted to the ClinVar database (<https://www.ncbi.nlm.nih.gov/clinvar/>).

### REFERENCES

- Capone VP, Morello W, Taroni F, Montini G. Genetics of congenital anomalies of the kidney and urinary tract: the current state of play. *Int J Mol Sci*. 2017;18:796.
- Hays T, Thompson MV, Bateman DA, Sahni R, Tolia VN, Clark RH, et al. The prevalence and clinical significance of congenital anomalies of the kidney and urinary tract in preterm infants. *JAMA Netw Open*. 2022;5:e2231626.
- Loane M, Dolk H, Kelly A, Teljeur C, Greenlees R, Densem J, et al. Paper 4: EUROCAT statistical monitoring: identification and investigation of ten year trends of congenital anomalies in Europe. *Birth Defects Res A Clin Mol Teratol*. 2011;91:S31–43.
- Verbitsky M, Westland R, Perez A, Kiryluk K, Liu Q, Krithivasan P, et al. The copy number variation landscape of congenital anomalies of the kidney and urinary tract. *Nat Genet*. 2019;51:117–27.
- van der Ven AT, Connaughton DM, Ityel H, Mann N, Nakayama M, Chen J, et al. Whole-exome sequencing identifies causative mutations in families with congenital anomalies of the kidney and urinary tract. *J Am Soc Nephrol*. 2018;29:2348–61.
- Sanna-Cherchi S, Ravani P, Corbani V, Parodi S, Haupt R, Piaggio G, et al. Renal outcome in patients with congenital anomalies of the kidney and urinary tract. *Kidney Int*. 2009;76:528–33.
- Nicolaou N, Renkema KY, Bongers EM, Giles RH, Knoers NV. Genetic, environmental, and epigenetic factors involved in CAKUT. *Nat Rev Nephrol*. 2015;11:720–31.
- Vivante A, Hildebrandt F. Exploring the genetic basis of early-onset chronic kidney disease. *Nat Rev Nephrol*. 2016;12:133–46.
- Nigam A, Knoers N, Renkema KY. Impact of next generation sequencing on our understanding of CAKUT. *Semin Cell Dev Biol*. 2019;91:104–10.
- Sanna-Cherchi S, Westland R, Ghiggeri GM, Gharavi AG. Genetic basis of human congenital anomalies of the kidney and urinary tract. *J Clin Invest*. 2018;128:4–15.
- Neild GH. What do we know about chronic renal failure in young adults? I. Primary renal disease. *Pediatr Nephrol*. 2009;24:1913–9.
- Chevalier RL. CAKUT: a pediatric and evolutionary perspective on the leading cause of CKD in childhood. *Pediatr Rep*. 2023;15:143–53.
- Murugapopathy V, Gupta IR. A primer on congenital anomalies of the kidneys and urinary tracts (CAKUT). *Clin J Am Soc Nephrol*. 2020;15:723–31.
- Malinowski J, Miller DT, Demmer L, Gannon J, Pereira EM, Schroeder MC, et al. Systematic evidence-based review: outcomes from exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability. *Genet Med*. 2020;22:986–1004.
- Bekheirnia MR, Bekheirnia N, Bainbridge MN, Gu S, Coban Akdemir ZH, Gambin T, 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*. 2017;19:412–20.



16. Groopman EE, Rasouly HM, Gharavi AG. Genomic medicine for kidney disease. *Nat Rev Nephrol*. 2018;14:83–104.
17. Kohl S, Habbig S, Weber LT, Liebau MC. Molecular causes of congenital anomalies of the kidney and urinary tract (CAKUT). *Mol Cell Pediatr*. 2021;8:2.
18. Liu P, Meng L, Normand EA, Xia F, Song X, Ghazi A, et al. Reanalysis of clinical exome sequencing data. *N Engl J Med*. 2019;380:2478–80.
19. Parikh JR, Genetti CA, Aykanat A, Brownstein CA, Schmitz-Abe K, Danowski M, et al. A data-driven architecture using natural language processing to improve phenotyping efficiency and accelerate genetic diagnoses of rare disorders. *HGG Adv*. 2021;2:100035.
20. Robinson PN, Haendel MA. Ontologies, knowledge representation, and machine learning for translational research: recent contributions. *Yearb Med Inf*. 2020;29:159–62.
21. Liu C, Peres Kury FS, Li Z, Ta C, Wang K, Weng C. Doc2Hpo: a web application for efficient and accurate HPO concept curation. *Nucleic Acids Res*. 2019;47:W566–W70.
22. Greene D, Richardson S, Turro E. ontologyX: a suite of R packages for working with ontological data. *Bioinformatics*. 2017;33:1104–6.
23. Scott TM, Campbell IM, Hernandez-Garcia A, Lalani SR, Liu P, Shaw CA, et al. Clinical exome sequencing data reveal high diagnostic yields for congenital diaphragmatic hernia plus (CDH+) and new phenotypic expansions involving CDH. *J Med Genet*. 2021;59:270–8.
24. Campbell IM, Rao M, Arredondo SD, Lalani SR, Xia Z, Kang SH, et al. Fusion of large-scale genomic knowledge and frequency data computationally prioritizes variants in epilepsy. *PLoS Genet*. 2013;9:e1003797.
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*. 2000;25:25–9.
26. Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT, Mouse Genome Database G. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res*. 2011;39:D842–8.
27. Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, Grimmond SM, et al. PINA v2.0: mining interactome modules. *Nucleic Acids Res*. 2012;40:D862–5.
28. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res*. 2010;38:D355–60.
29. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. 2005;120:15–20.
30. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci USA*. 2004;101:6062–7.
31. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol*. 2010;28:1045–8.
32. Online Mendelian Inheritance in Man, OMIM®[online database], McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). <https://omim.org> [Accessed Dec 2024].
33. Schofield D, Alam K, Douglas L, Shrestha R, MacArthur DG, Davis M, et al. Cost-effectiveness of massively parallel sequencing for diagnosis of paediatric muscle diseases. *NPJ Genom Med*. 2017;2:4.
34. Posey JE, O'Donnell-Luria AH, Chong JX, Harel T, Jhangiani SN, Coban Akdemir ZH, et al. Insights into genetics, human biology and disease gleaned from family based genomic studies. *Genet Med*. 2019;21:798–812.
35. Tarailo-Graovac M, Shyr C, Ross CJ, Horvath GA, Salvarinova R, Ye XC, et al. Exome sequencing and the management of neurometabolic disorders. *N Engl J Med*. 2016;374:2246–55.
36. Palmer EE, Schofield D, Shrestha R, Kandula T, Macintosh R, Lawson JA, et al. Integrating exome sequencing into a diagnostic pathway for epileptic encephalopathy: Evidence of clinical utility and cost effectiveness. *Mol Genet Genom Med*. 2018;6:186–99.
37. Griesemer I, Staley BS, Lightfoot AF, Bain L, Byrd D, Conway C, et al. Engaging community stakeholders in research on best practices for clinical genomic sequencing. *Per Med*. 2020;17:435–44.
38. Mackley MP, Fletcher B, Parker M, Watkins H, Ormondroyd E. Stakeholder views on secondary findings in whole-genome and whole-exome sequencing: a systematic review of quantitative and qualitative studies. *Genet Med*. 2017;19:283–93.
39. Knoers N, Antignac C, Bergmann C, Dahan K, Giglio S, Heidet L, et al. Genetic testing in the diagnosis of chronic kidney disease: recommendations for clinical practice. *Nephrol Dial Transpl*. 2022;37:239–54.
40. Ahram DF, Aggarwal VS, Sanna-Cherchi S. Phenocopies, phenotypic expansion, and coincidental diagnoses: time to abandon targeted gene panels? *Am J Kidney Dis*. 2020;76:451–3.
41. Posey JE. Genome sequencing and implications for rare disorders. *Orphanet J Rare Dis*. 2019;14:153.
42. Brlek P, Bulic L, Bracic M, Projic P, Skaro V, Shah N, et al. Implementing whole genome sequencing (WGS) in clinical practice: advantages, challenges, and future perspectives. *Cells*. 2024;13:504.
43. Kohl S, Avni EE, Boor P, Capone V, Clapp WL, De Palma D, et al. Definition, diagnosis and clinical management of non-obstructive kidney dysplasia: a consensus statement by the ERKNet Working Group on Kidney Malformations. *Nephrol Dial Transpl*. 2022;37:2351–62.
44. Fraiman YS, Wojcik MH. The influence of social determinants of health on the genetic diagnostic odyssey: who remains undiagnosed, why, and to what effect? *Pediatr Res*. 2021;89:295–300.
45. Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*. 2013;369:1502–11.
46. McKusick VA. Mendelian Inheritance in Man and its online version, OMIM. *Am J Hum Genet*. 2007;80:588–604.
47. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, Vandeweyer G, Rooms L, van den Ende J, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat Genet*. 2014;46:380–4.
48. Asegaonkar P, Kotecha U, Dongre M, Mistri M, Sharda S. Expanding the spectrum of ADNP-related disorder-Antenatally diagnosed congenital diaphragmatic hernia and a novel de novo mutation in ADNP gene. *Am J Med Genet A*. 2023;191:275–9.
49. Vivante A, Mann N, Yonath H, Weiss AC, Getwan M, Kaminski MM, et al. A dominant mutation in nuclear receptor interacting protein 1 causes urinary tract malformations via dysregulation of retinoic acid signaling. *J Am Soc Nephrol*. 2017;28:2364–76.
50. Posey JE, Harel T, Liu P, Rosenfeld JA, James RA, Coban Akdemir ZH, et al. Resolution of disease phenotypes resulting from multilocus genomic variation. *N Engl J Med*. 2017;376:21–31.
51. Van Dijk A, Vulto-van Silfhout AT, Cappuyns E, van der Werf IM, Mancini GM, Tzschach A, et al. Clinical presentation of a complex neurodevelopmental disorder caused by mutations in ADNP. *Biol Psychiatry*. 2019;85:287–97.
52. Farhang-Fallah J, Randhawa VK, Nimnual A, Klip A, Bar-Sagi D, Rozakis-Adcock M. The pleckstrin homology (PH) domain-interacting protein couples the insulin receptor substrate 1 PH domain to insulin signaling pathways leading to mitogenesis and GLUT4 translocation. *Mol Cell Biol*. 2002;22:7325–36.
53. Bai SW, Herrera-Abreu MT, Rohn JL, Racine V, Tajadura V, Suryavanshi N, et al. Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration. *BMC Biol*. 2011;9:54.
54. Jansen S, Hoischen A, Coe BP, Carvill GL, Van Esch H, Bosch DGM, et al. A genotype-first approach identifies an intellectual disability-overweight syndrome caused by PHIP haploinsufficiency. *Eur J Hum Genet*. 2018;26:54–63.
55. Craddock KE, Okur V, Wilson A, Gerkes EH, Ramsey K, Heeley JM, et al. Clinical and genetic characterization of individuals with predicted deleterious PHIP variants. *Cold Spring Harb Mol Case Stud*. 2019;5:a004200.
56. Groza T, Gomez FL, Mashhadi HH, Munoz-Fuentes V, Gunes O, Wilson R, et al. The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res*. 2023;51:D1038–D45.
57. de Fallois J, Sieckmann T, Schönauer R, Petzold F, Münch J, Pauly M, et al. Pathogenic PHIP variants are variably associated with CAKUT. *Kidney Int Rep*. 2024;9:2484–97.
58. Sudnawa KK, Calamia S, Geltzeiler A, Chung WK. Clinical phenotypes of individuals with Chung-Jansen syndrome across age groups. *Am J Med Genet A*. 2024;194:e63471.
59. Tang Y, Chen Y, Wang J, Zhang Q, Wang Y, Xu Y, et al. Clinical characteristics and genetic expansion of 46,XY disorders of sex development children in a Chinese prospective study. *Endocr Connect*. 2023;12:e230029.
60. Kaur H, Panigrahi I. Chung-Jansen Syndrome with obesity. *Obes Res Clin Pr*. 2021;15:303–5.
61. Sessa A, Fagnocchi L, Mastrototaro G, Massimino L, Zaghi M, Indrigo M, et al. SETD5 regulates chromatin methylation state and preserves global transcriptional fidelity during brain development and neuronal wiring. *Neuron*. 2019;104:271–89.e13.
62. Kuechler A, Zink AM, Wieland T, Ludecke HJ, Cremer K, Salviati L, et al. Loss-of-function variants of SETD5 cause intellectual disability and the core phenotype of microdeletion 3p25.3 syndrome. *Eur J Hum Genet*. 2014;23:753–60.
63. Szczaluba K, Brzezinska M, Kot J, Rydzanicz M, Walczak A, Stawinski P, et al. SETD5 loss-of-function mutation as a likely cause of a familial syndromic intellectual disability with variable phenotypic expression. *Am J Med Genet A*. 2016;170:2322–7.
64. Mohun T, Adams DJ, Baldock R, Bhattacharya S, Copp AJ, Hemberger M, et al. Deciphering the Mechanisms of Developmental Disorders (DMDD): a new programme for phenotyping embryonic lethal mice. *Dis Model Mech*. 2013;6:562–6.

65. Mohapatra B, Ahmad G, Nadeau S, Zutshi N, An W, Scheffe S, et al. Protein tyrosine kinase regulation by ubiquitination: critical roles of Cbl-family ubiquitin ligases. *Biochim Biophys Acta*. 2013;1833:122–39.
66. Martinelli S, De Luca A, Stellacci E, Rossi C, Checquolo S, Lepri F, et al. Heterozygous germline mutations in the CBL tumor-suppressor gene cause a Noonan syndrome-like phenotype. *Am J Hum Genet*. 2010;87:250–7.
67. Hecht A, Meyer JA, Behnert A, Wong E, Chehab F, Olshen A, et al. Molecular and phenotypic diversity of CBL-mutated juvenile myelomonocytic leukemia. *Haematologica*. 2022;107:178–86.
68. Martinelli S, Stellacci E, Pannone L, D'Agostino D, Consoli F, Lissewski C, et al. Molecular diversity and associated phenotypic spectrum of germline CBL mutations. *Hum Mutat*. 2015;36:787–96.
69. Caro-Llopis A, Rosello M, Orellana C, Oltra S, Monfort S, Mayo S, et al. De novo mutations in genes of mediator complex causing syndromic intellectual disability: mediatoropathy or transcriptomopathy? *Pediatr Res*. 2016;80:809–15.
70. Smol T, Petit F, Piton A, Keren B, Sanlaville D, Afenjar A, et al. MED13L-related intellectual disability: involvement of missense variants and delineation of the phenotype. *Neurogenetics*. 2018;19:93–103.
71. Yi Z, Zhang Y, Song Z, Pan H, Yang C, Li F, et al. Report of a de novo c.2605C > T (p.Pro869Ser) change in the MED13L gene and review of the literature for MED13L-related intellectual disability. *Ital J Pediatr*. 2020;46:95.
72. Woods CG, Stricker S, Seemann P, Stern R, Cox J, Sherridan E, et al. Mutations in WNT7A cause a range of limb malformations, including Fuhrmann syndrome and Al-Awadi/Raas-Rothschild/Schinzle phocomelia syndrome. *Am J Hum Genet*. 2006;79:402–8.
73. Garavelli L, Wischmeijer A, Rosato S, Gelmini C, Reverberi S, Sassi S, et al. Al-Awadi-Raas-Rothschild (limb/pelvis/uterus-hypoplasia/aplasia) syndrome and WNT7A mutations: genetic homogeneity and nosological delineation. *Am J Med Genet A*. 2011;155A:332–6.
74. Eyaid W, Al-Qattan MM, Al Abdulkareem I, Fetaini N, Al Balwi M. A novel homozygous missense mutation (c.610G>A, p.Gly204Ser) in the WNT7A gene causes tetra-amelia in two Saudi families. *Am J Med Genet A*. 2011;155A:599–604.

## AUTHOR CONTRIBUTIONS

EARM – Conceptualization, Formal analysis, Data curation, Writing original draft, Methodology, Validation. XEZ – Formal analysis, Data curation, Review & editing. JAR – Resources, Review & editing. PNL – Software, Resources. CS – Software, Resources. JEP – Conceptualization, Review & editing, Supervision, Methodology, Funding acquisition. DAS – Conceptualization, Review & editing, Supervision, Methodology, Funding acquisition.

## FUNDING

This work was supported by the National Institutes of Health grants T32 NIH-NIGMS GM136554 (Clinical Translational Research-CAQ) and U01 NHGRI HG011758

(Genomics Research to Elucidate the Genetics of Rare disease; GREGoR), and National Institutes of Health/Eunice Kennedy Shriver National Institute of Child Health and Human Development grant R01 HD098458.

## COMPETING INTERESTS

The Department of Molecular & Human Genetics at Baylor College of Medicine receives revenue from clinical genetic testing completed at Baylor Genetics.

## ETHICAL APPROVAL

This work was approved by the Institutional Review Board of Baylor College of Medicine (protocol H-47546) and was conducted in accordance with international standards on human subjects research. Individual consent was not obtained since only anonymized/coded molecular and clinical data were used in this study.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41431-025-01929-3>.

**Correspondence** and requests for materials should be addressed to Daryl A. Scott.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025