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

Mir Reza Bekheirnia, MD, FACMG1,2,3,4, Nasim Bekheirnia, MBS, MS1,2,4, Matthew N. Bainbridge, PhD5, Shen Gu, PhD1, Zeynep Hande Coban Akdemir, PhD1, Tomek Gambin, PhD1, Nicolette K. Janzen, MD3,4, Shalini N. Jhangiani, MS5, Donna M. Muzny, MS5, Mini Michael, MD4,6, Eileen D. Brewer, MD4,6, Ewa Elenberg, MD4,6, Arundhati S. Kale, MD4,6, Alyssa A. Riley, MD4,6, Sarah J. Swartz, MD4,6, Daryl A. Scott, MD, PhD1,4, Yaping Yang, PhD1, Poyyapakkam R. Srivaths, MD4,6, Scott E. Wenderfer, MD, PhD4,6, Joann Bodurtha, MD, MPH7, Carolyn D. Applegate, MS7, Milen Velinov, MD, PhD7, Angela Myers, MD7, Lior Borovik, MS7, William J. Craigie, MD, PhD1,4, Neil A. Hanchard, MD, PhD1,4, Jill A. Rosenfeld, MS1, Richard Alan Lewis, MD1,4,10, Edmond T. Gonzales, MD3,4, Richard A. Gibbs, PhD1,5, John W. Belmont, MD, PhD1,4, David R. Roth, MD3,4, Christine Eng, MD1, Michael C. Braun, MD4,6, James R. Lupski, MD, PhD1,4,5,11 and Dolores J. Lamb, PhD2,3,12

Purpose: To investigate the utility of whole-exome sequencing (WES) to define a molecular diagnosis for patients clinically diagnosed with congenital anomalies of kidney and urinary tract (CAKUT).

Methods: WES was performed in 62 families with CAKUT. WES data were analyzed for single-nucleotide variants (SNVs) in 35 known CAKUT genes, putatively deleterious sequence changes in new candidate genes, and potentially disease-associated copy-number variants (CNVs).

Results: In approximately 5% of families, pathogenic SNVs were identified in PAX2, HNF1B, and EYA1. Observed phenotypes in these families expand the current understanding about the role of these genes in CAKUT. Four pathogenic CNVs were also identified using two CNV detection tools. In addition, we found one deleterious de novo SNV in FOXP1 among the 62 families with CAKUT. The clinical database of the Baylor Miraca Genetics laboratory was queried and seven additional unrelated individuals with novel de novo SNVs in FOXP1 were identified. Six of these eight individuals with FOXP1 SNVs have syndromic urinary tract defects, implicating this gene in urinary tract development.

Conclusion: We conclude that WES can be used to identify molecular etiology (SNVs, CNVs) in a subset of individuals with CAKUT. WES can also help identify novel CAKUT genes.

Key Words: CAKUT; WES; FOXP1; PAX2; HNF1B

INTRODUCTION

Diagnosis of congenital anomalies of the kidney and urinary tract (CAKUT) is based on the recognition of a broad spectrum of renal and urinary tract malformations that, in aggregate, constitute the most common cause of end-stage renal disease in children.1,2 CAKUT may result in chronic kidney disease that leads to severe impairment of physical and psychosocial development.1 Socioeconomically, CAKUT poses a substantial economic burden to families and health-care systems. CAKUT is a clinically heterogeneous phenotype that encompasses renal agenesis, renal hypoplasia/dysplasia, multicystic kidney dysplasia, cross-fused ectopia, duplex renal collecting system, ureteropelvic junction obstruction, mega-ureter, posterior urethral valves, and vesicoureteral reflux (VUR).

Multiple lines of evidence suggest that genetic factors contribute to CAKUT. This evidence includes familial segregation of CAKUT cases and the identification of causative genes.3 Discovery of an underlying genetic etiology facilitates...
molecular diagnosis and can aid physicians and family members by clarifying associated risks and allowing improved genetic counseling.

In the past, genetic diagnosis was limited to the analysis of individual candidate genes, but whole-exome sequencing (WES) provides an opportunity to arrive at an accurate molecular diagnosis with a single test.\(^5\)\(^6\) WES is able to identify single-nucleotide variants (SNVs); however, recently, it has also been used to uncover even small CNVs encompassing a single gene or even one exon.\(^7\) Extraction of CNV information from WES data is challenging, partly due to the potential artifacts introduced during the exon targeting and amplification steps of WES.\(^8\) Moreover, WES can enhance gene discovery for novel potential contributory genes.

Initial reports of clinical WES at a Clinical Laboratory Improvement Amendments–certified laboratory indicated a molecular diagnostic rate of 25% for patients referred for genetic evaluation.\(^5\)\(^6\) However, neurological phenotypes constituted 80% of the patient population in that study. The clinical utility of using WES in common, sporadic birth defects is undergoing active investigation. Current evaluations of CAKUT patients involve diagnostic imaging, but involvement of other organs may go undiagnosed during such evaluations. Here, we investigated the utility of WES to define a molecular diagnosis (SNVs and CNVs) in patients clinically diagnosed with CAKUT.

**MATERIALS AND METHODS**

Patients and their families were recruited from the pediatric urology and renal diseases clinics at the Texas Children’s Hospital in Houston, Texas. Inclusion criteria included individuals with nonsyndromic forms of CAKUT (as defined above) and individuals with syndromic forms of CAKUT for which a genetic etiology had not been identified. Exclusion criteria included individuals with syndromic forms of CAKUT in which an underlying genetic etiology was known and individuals with nonsyndromic and nonfamilial forms of VUR. Therefore, individuals with syndromic features without a known diagnosis were included in the study. The study protocol was approved by the Institutional Review Board for the Protection of Human Subjects at Baylor College of Medicine. Standard procedures were used to recruit subjects for this study. Demographics of the families and phenotypic details of subjects with CAKUT are summarized in Table 1 and Supplementary Table S1 online. Blood samples or saliva-based specimens were collected by standard procedures and according to the families’ wishes. DNA extraction was performed with a QIAmp kit (Qiagen) per the manufacturer’s instructions. DNA was quantified with nanodrop and 1 µg of DNA was used for WES. In familial multiplex cases, WES was performed on the affected available family members most distantly related and observed in their respective pedigrees (see Figure 1). Among apparently isolated singleton cases, we performed WES for the proband and the two apparently unaffected parents (case–parent trios in 20 families) when both parents were available. In all other cases, WES was performed only for the proband.

### Table 1 Demographics and different phenotypes of the 62 families with CAKUT who underwent whole-exome sequencing

| Age of probands (mean, SD) | 6.5 years, 5.8 years |
|----------------------------|---------------------|
| Sex of probands (M/F)      | 1.29                |
| Other organ affected (syndromic) | 19 (31%)*           |
| Mode of analysis           |                     |
| Familial (more than one affected individual in the family) | 10 (16%)            |
| Sporadic: case–parent trios | 20 (32%)            |
| Sporadic: single cases     | 32 (52%)            |
| Diagnosis of probands      |                     |
| Renal dysplasia            | 14 (23%)            |
| Renal agenesis/hypoplasia  | 12 (19%)            |
| Posterior urethral valve   | 10 (16%)            |
| Vesicoureteral reflux (VUR)| 9 (15%)             |
| Duplicated collecting system (DCS) | 7 (11%)           |
| Ureteropelvic junction obstruction (UPIO) | 5 (8%) |
| Fusion anomaly: horseshoe kidney and crossed fused ectopia | 4 (6%) |
| Anterior urethral valve    | 1 (2%)              |
| Total number of families   | 62 (100%)           |

*Two more syndromic families were identified after whole-exome sequencing results became available; accordingly, total syndromic = 21 (34%).

**Whole-exome sequencing analysis**

WES analysis started with conversion of raw sequencing data (bcl files) to the fastq format by Casava. Then, the short reads were mapped to a human genome reference sequence (GRCh37) by the Burrows-Wheeler Alignment. Subsequently, the recalibration was performed by GATK, and variant calling was performed by the Atlas2 suite.\(^10\) The Mercury pipeline is available in the Cloud via DNANexus (http://blog.dnanexus.com/2013-10-22-run-mercury-variant-calling-pipeline/).

**SNV prioritization and filtering workflow**

After detection of all biallelic (homozygous or compound heterozygous) and de novo variants from WES data, we established a SNV prioritization workflow. This included sequential analysis of biallelic predicted loss of function variants (stopgain, frameshift indels, and splicing), biallelic missense variants, de novo truncating variants, and de novo missense variants. Finally, we further examined the shared rare variants among affected family members and parents to detect potential mosaic variants in parents. This SNV prioritization workflow was followed by subsequent filtering of variants based on their frequencies (minor allele frequency ≤0.1%) in internal and external databases, including the Baylor-Hopkins Center for Mendelian Genomics (BHCMG), the Exome Aggregation Consortium, the Exome Sequencing Project (ESP), the 1000 Genomes Project, and the Atherosclerosis Risk in Communities Study (ARIC) databases. To retrieve potentially deleterious and conserved missense changes, we utilized various bioinformatics tools, including the Phylp conservation score and Mutation Taster, SIFT (the
Copy-number variant inference
To identify copy-number variants (CNVs), our WES data were analyzed using CoNIFER software and CoNVex algorithms. In CoNIFex, as a first step, read depth information from WES data was extracted. Then, the general additive model correction method was performed to remove the systemic bias from the read depth information. The Smith-Waterman algorithm was used to infer the CNV state and score the detected CNV regions. Each potential CNV region was assigned a confidence score. We further filtered CoNVex-detected CNV calls by selecting those that have an associated confidence score ≥5 and ≥5 probes. Afterward, these detected CNV calls by CoNVex were overlapped with CNV calls detected by CoNIFER by using Granges function in R Bioconductor GenomicRanges Package. Overlapping CNVs in previous studies were subjected to validation by array comparative genomic hybridization (aCGH). Nonoverlapping CNVs were also investigated. Among them, rare CNVs—not present or present with low frequency per the Database of Genomic Variants (http://dgv.tcag.ca)—involving genes potentially contributing to kidney abnormalities were selected for aCGH validation. The flowchart of CNV discovery from WES data is provided in Supplementary Figure S1a online.

SNV and CNV interpretation criteria
SNV interpretation was based on the most recent guidelines published by the American College of Medical Genetics and Genomics (ACMG). Accordingly, only variants that met strict criteria were called pathogenic. CNV interpretation was based on size, gene content, overlap with known disease-associated regions, and phenotype overlap according to the ACMG guidelines for postnatal CNV calling.

Known CAKUT-associated genes
In this study, Codified software (https://www.sciencedxchange.com/labs/codified-genomics) was used to search WES data for pathogenic SNVs, variants of uncertain clinical significance (VUS), and benign SNVs in the following 19 dominantly inherited genes reported to be associated with CAKUT: BMP4, BMP7, CDC51, CHD11, DSYKY, EYA1, GATA3, HNF1B, KAL1, PAX2, RET, ROBO2, SALL1, SIX1, SIX2, SIX5, SOX17, TNXB, and UPK3A. The following recessive CAKUT genes were interrogated for two SNVs: AGT, ACE, REN, AGTR1, FRAS1, FREM2, GRIP1, HPS2, LR4P, and ROR2. In addition, we searched for pathogenic SNVs in the following six dominantly inherited genes: GLI3, JAG1, NOTCH2, TFAP2A, TBX18, and WNT4.

Codified Genomics software was used to annotate, filter, and prioritize variants. Variants were filtered as previously described. Annotations were generated by Annovar against the UCSC, RefSeq, and Ensembl gene models. Variants and genes were further annotated using dbNSFP Illumina body map, Uniprot, HPO, and OMIM databases, among others. Variants were prioritized based on patient phenotype similarity to known disease genes and mutation type, and nonsynonymous variants were prioritized based on predicted deleteriousness.

RESULTS
We performed WES for 112 individuals from 62 families with CAKUT (Table 1 and Supplementary Table S1 online). Proband-specific variants were mostly children and young adults ranging in

Figure 1 Pedigrees and genotypes of the families with pathogenic single-nucleotide variants (SNVs) in genes known to cause CAKUT and a novel CAKUT gene (family 38, FOXP1). *Individuals for whom whole-exome sequencing was performed. NT, not tested. Family 1: Solid black fill shows renal dysplasia and solid gray fill shows proteinuria. CG/CG is normal and CG/C− denotes heterozygous deletion of G (c.70delG) in PAX2. Family 2: Proband has cystic renal dysplasia. C/C is normal and C/CC denotes heterozygous duplication of C (c.1132dupC) in HNF1B. Family 3: Proband has vesicoureteral reflux and multicystic dysplastic kidney. G/G is normal and G/T denotes heterozygous de novo FOXP1 p.P225T SNV.

“sorting intolerant from tolerant” algorithm, and PolyPhen-2 prediction scores. Next, these potential rare causative variants were analyzed in terms of (1) gene function and the associated phenotype in OMIM and PubMed; (2) gene-associated animal models; (3) tissue expression of the encoded protein; (4) association with the already known gene/genes linked to the patient's phenotype in terms of (i) gene networks, (ii) gene families, (iii) coexpression, (iv) physical protein–protein interaction, (v) predicted protein–protein interaction, and (vi) molecular pathways; and (5) location of the variant with respect to functional protein domains. The resulting most promising candidate variants were further confirmed and segregated by Sanger sequencing. Finally, the confirmed variants in candidate genes were interrogated in the BHCMG and Baylor Miraca Genetics Laboratories (BMGL) databases and/or through GeneMatcher for the identification of additional affected cases with similar phenotypes.

Copy-number variant inference
To identify copy-number variants (CNVs), our WES data were analyzed using CoNIFER software and CoNVex algorithms. In CoNIFex, as a first step, read depth information from WES data was extracted. Then, the general additive model correction method was performed to remove the systemic bias from the read depth information. Next, the Smith-Waterman algorithm was used to infer the CNV state and score the detected CNV regions. Each potential CNV region was assigned a confidence score. We further filtered CoNVex-detected CNV calls by selecting those that have an associated confidence score ≥5 and ≥5 probes. Afterward, these detected CNV calls by CoNVex were overlapped with CNV calls detected by CoNIFER by using Granges function in R Bioconductor GenomicRanges Package. Overlapping CNVs in previous studies were subjected to validation by array comparative genomic hybridization (aCGH). Nonoverlapping CNVs were also investigated. Among them, rare CNVs—not present or present with low frequency per the Database of Genomic Variants (http://dgv.tcag.ca)—involving genes potentially contributing to kidney abnormalities were selected for aCGH validation. The flowchart of CNV discovery from WES data is provided in Supplementary Figure S1a online.

SNV and CNV interpretation criteria
SNV interpretation was based on the most recent guidelines published by the American College of Medical Genetics and Genomics (ACMG). Accordingly, only variants that met strict criteria were called pathogenic. CNV interpretation was based on size, gene content, overlap with known disease-associated regions, and phenotype overlap according to the ACMG guidelines for postnatal CNV calling.

Known CAKUT-associated genes
In this study, Codified software (https://www.sciencedxchange.com/labs/codified-genomics) was used to search WES data for pathogenic SNVs, variants of uncertain clinical significance (VUS), and benign SNVs in the following 19 dominantly inherited genes reported to be associated with CAKUT: BMP4, BMP7, CDC51, CHD11, DSYKY, EYA1, GATA3, HNF1B, KAL1, PAX2, RET, ROBO2, SALL1, SIX1, SIX2, SIX5, SOX17, TNXB, and UPK3A. The following recessive CAKUT genes were interrogated for two SNVs: AGT, ACE, REN, AGTR1, FRAS1, FREM2, GRIP1, HPS2, LR4P, and ROR2. In addition, we searched for pathogenic SNVs in the following six dominantly inherited genes: GLI3, JAG1, NOTCH2, TFAP2A, TBX18, and WNT4.

Codified Genomics software was used to annotate, filter, and prioritize variants. Variants were filtered as previously described. Annotations were generated by Annovar against the UCSC, RefSeq, and Ensembl gene models. Variants and genes were further annotated using dbNSFP Illumina body map, Uniprot, HPO, and OMIM databases, among others. Variants were prioritized based on patient phenotype similarity to known disease genes and mutation type, and nonsynonymous variants were prioritized based on predicted deleteriousness.

RESULTS
We performed WES for 112 individuals from 62 families with CAKUT (Table 1 and Supplementary Table S1 online). Proband-specific variants were mostly children and young adults ranging in
Table 2  Pathogenic single-nucleotide variants in 35 known genes identified in 62 families with CAKUT by whole-exome sequencing

| Family ID | Gene | Sex | Ethnicity | Renal phenotype | Genome build | Chr: position | Nucleotide change | Amino acid change | Other organ defects | Seen in CMG | ESP EA/AA | CADD score | Frequency in ExAc | Number of LOF in ExAc | Family segregation | Comment |
|-----------|------|-----|-----------|-----------------|--------------|---------------|-----------------|------------------|-------------------|-------------|------------|-------------|------------------|-------------------|---------------------|---------|
| 1         | PAX2 | M,F | Spanish   | RHD             | Hg19         | 10:102509528  | c.70delG        | p.G24fs          | Optic nerve coloboma (previously undiagnosed) | no          | NR; N/R   | 35          | 0                 | 5                  | Yes; multiple individuals, AD | Known pathogenic variant, impacts medical management |
| 2         | HNF1B| M   | Mixed Caucasian and African Caribbean | CRD             | Hg19         | 17:36070584   | c.1132dupC      | p.Q378fs         | Gout, elevated LFTs (recent diagnosis) and increased echogenicity of pancreas (not noted prior to WES) | no          | NR; N/R   | 35          | 0                 | 0                  | De novo by WES and Sanger | Novel pathogenic variant, impacts medical management |
| 3         | EYA1 | F   | AA        | VUR, MCDK       | Hg19         | 8:72183988    | c.867+5G>A     | n/a              | No                | no          | NR; N/R   | 16.61       | 0                 | 4                  | De novo by WES and Sanger | Known recurrent pathogenic variant in BOR; MCDK in this patient with VUR likely due to this variant, impacts medical management |

AA, African American; AD, autosomal dominant; BOR, branchio-oto-renal syndrome; CADD, combined annotation dependent depletion; Chr, chromosome; CMG, Center for Mendelian Genomics; CRD, cystic renal dysplasia; ESP, Exome Sequencing Project; ExAc, Exome Aggregation Consortium; LFT, liver-function tests; LOF, loss-of-function; MCDK, multicystic dysplastic kidney; NR, none reported; RHD, renal hypodysplasia; VUR, vesicoureteral reflux; WES, whole-exome sequencing.
age from 2 months to 24 years. In 31% of the probands, more
than one organ (other than the kidney and urinary tract) was
involved, which suggests that these patients potentially harbor a
syndromic form of CAKUT. In approximately 16% of the fami-
lies, WES was performed in a familial mode because more than
one individual was affected with CAKUT. The most common
phenotypic indications were “renal dysplasia” and “agenesis/
hypoplasia.”

WES results were interrogated for SNVs in known genes that
cause CAKUT as described in Methods. Pathogenic SNVs were
identified in one individual from this cohort (family 38) and 7 additional (cases 2–8 from clinical whole-exome sequencing database) individuals with pathogenic novel de novo SNVs

| Case number | Genome build | Chr: position | SNV | Protein change | CADD score | Frequency in ExAc | Number of LOF in ExAc | Segregation | Phenotype |
|-------------|--------------|---------------|-----|----------------|------------|-------------------|----------------------|-------------|-----------|
| Case 1 (family 38) | Hg19 | 3: 71090675 | c.C673A | p.P225T | 24.7 | 0 | 2 | De novo | Hydrocephaly, GDD, and unilateral renal agenesis |
| Case 2 | Hg19 | 3: 71027085 | c.1240_1241del | p.L414fs | 35 | 0 | 2 | De novo | Hydrocephaly, brain atrophy, seizure, DD, VSD, neurogenic bladder, hydronephrosis, tethered cord, and behavior problems |
| Case 3 | Hg19 | 3: 71090502 | c.844-845del | p.V283fs | 35 | 0 | 2 | De novo | Bilateral ventricular enlargement of brain, ID, DD, hypotonia, horseshoe kidney, UDT, hypospadias, and facial dysmorphic features |
| Case 4 | Hg19 | 3: 71021784 | c.1574G>A | p.R525Q | 34 | 0 | 2 | De novo | GDD, mega cisterna magna, seizure, hypotonia, concealed penis, retractable testis, and skeletal abnormalities |
| Case 5 | Hg19 | 3: 71021815 | c.1543C>G | p.H515D | 27.7 | 0 | 2 | De novo | GDD, hip contractures, dysmorphic features, short stature, mild scoliosis, abnormally positioned thumbs, dysmorphic features, and UDT |
| Case 6 | Hg19 | 3: 71019958 | c.1653-2A>T | N/A | 24.8 | 0 | 2 | De novo | DD, ataxia, mild distal arthrogryposis, ASD, dysmorphic facial features, strabismus, vertical nystagmus and snoring, right duplicated collecting system, and trabeculation of the posterior urinary bladder wall |
| Case 7 | Hg19 | 3: 71027034 | c.1291_1292del | p.T431fs | 35 | 0 | 2 | De novo | Macrocephaly, GDD, ID, dysmorphic features, hypextensibility, joint contractures (3rd finger camptodactyly), droopy eyelids, and two café-au-lait spots |
| Case 8 | Hg19 | 3: 71021735 | c.1606_1622dup | p.F541fs | 35 | 0 | 2 | De novo | Structural brain abnormalities, GDD, aggressive behavior and hyperactivity, hyperopic astigmatism, hypotonia, and ankle tightness |

Six out of eight cases have syndromic forms of kidney and genitourinary tract defects. This strongly suggests an important role for FOXP1 in kidney and genitourinary tract development.
identified in three known genes (EYA1, HNF1B, and PAX2) in three families (approximately 5%) (Table 2). Two of these variants were frameshift variants and one was a splice site variant, with each suggesting a loss-of-function mechanism. The frameshift variant in HNF1B is a novel pathogenic allele. Pedigrees of these families are illustrated as Figure 1. Among the pathogenic SNVs identified in these three families, two were de novo from trio analyses and one was inherited from an affected parent. All selected SNVs identified in the probands and their parents (when available) were confirmed by Sanger sequencing.

Among the three families with pathogenic SNVs, clinical assessment had not identified anomalies of any other organs prior to WES. Importantly, WES elicited further clinical assessment and the delineation of the additional organ system involvement in families 1 and 2 retrospectively. In family 3, defects in other organs have not been observed clinically.

The initial diagnosis of family 1 with p.G24fs SNV in PAX2 was a familial form of renal dysplasia and membranous nephropathy. After the familial variant was identified and in recognition of the current understanding of the phenotype of patients with PAX2 variants (renal-coloboma syndrome; MIM 120330), the first-degree relatives were referred to an ophthalmologist with expertise in the diagnosis and management of genetic disorders. This clinical evaluation revealed optic nerve colobomata and other congenital optic nerve abnormalities in those first-degree relatives who were proven to be variant carriers.

After the initial clinical diagnosis of cystic renal dysplasia (CRD) in family 2, and in light of the WES results (p.Q378fs in HNF1B), the patient’s clinical presentation was further reviewed. The patient had a recent diagnosis of gout and elevated liver-function test (LFT) results. The patient also had increased echogenicity of the pancreas (one of the known signs of HNF1B variants) noted previously by abdominal ultrasound; its significance became apparent after the genetic analyses. A known recurrent de novo intronic variant (c.867+5G>A) was identified in EYA1 in the proband of family 3, who has VUR and multicystic dysplastic kidney. Both parents were negative for this variant based on both trio WES and Sanger sequencing.

In this study, we also defined VUSs in 19 dominantly inherited known CAKUT genes (Supplementary Table S2 online). In this cohort, SNVs were not identified in SIX1, SOX17, GATA3, or UPK3A. Benign SNVs and VUSs were identified in BMP7, CDC5L, CHD1L, SALL1, SIX5, SIX2, ROBO2, BMP4, KAL1, TNXB, RET, PAX2, EYA1, and DSTYK (Supplementary Table S2 online). Further allele frequencies and prediction data for all SNVs identified in known CAKUT genes in this cohort are summarized in Supplementary Table S3 online. Probability of loss-of-function score (pLI) is also provided in this table. The closer the pLI score is to 1 (unity), the more LoF (loss-of-function)-intolerant the gene appears to be (http://exac.broadinstitute.org). We attempted to confirm all VUSs with Sanger sequencing. Details of confirmation are provided in Supplementary Table S2 online.

**Novel CAKUT gene identification: Forkhead Box P1 (FOXP1)**

Trio analyses consisting of WES for the proband and both biological parents to evaluate for new mutations were performed for 20 families. We confirmed relationships (paternity and maternity) in the trios by review of the de novo SNVs in each family. There was no proband with more than the expected number of de novo SNVs (>2) in the coding exonic region of the genome, which was well within the expected rate of \(1.20 \times 10^{-8}\) per nucleotide per generation. In this study, we also defined VUSs in 19 dominantly inherited known CAKUT genes (Supplementary Table S2 online). In this cohort, SNVs were not identified in SIX1, SOX17, GATA3, or UPK3A. Benign SNVs and VUSs were identified in BMP7, CDC5L, CHD1L, SALL1, SIX5, SIX2, ROBO2, BMP4, KAL1, TNXB, RET, PAX2, EYA1, and DSTYK (Supplementary Table S2 online). Further allele frequencies and prediction data for all SNVs identified in known CAKUT genes in this cohort are summarized in Supplementary Table S3 online. Probability of loss-of-function score (pLI) is also provided in this table. The closer the pLI score is to 1 (unity), the more LoF (loss-of-function)-intolerant the gene appears to be (http://exac.broadinstitute.org). We attempted to confirm all VUSs with Sanger sequencing. Details of confirmation are provided in Supplementary Table S2 online.

**Table 4** Copy-number variants (CNVs) identified (from whole-exome sequencing data of 62 families) that are relevant to the patient’s phenotype

| Family ID | Chromosomal region | CNV type | Start (Mb) | End (Mb) | Size (Mb) | Number of genes | Syndrome | Phenotype | Parental studies | Comments |
|-----------|------------------|----------|-----------|---------|-----------|----------------|----------|-----------|-----------------|----------|
| Family 34 | 22q11            | Trp      | 16.63     | 18.64   | 2.01      | 33             | Cat-eye syndrome | VUR, further details in text (multiple anomalies) | De novo  | Pathogenic     |
| Family 39 | 16p13.11         | Dup      | 15.12     | 16.29   | 1.17      | 19             | 16p13.11 dup | MCDK, facial dysmorphic features | Unknown | Pathogenic     |
| Family 10 | 16p11.2          | Dup      | 29.68     | 30.20   | 0.52      | 35             | 16p11.2 dup | Solitary kidney, psychiatric disorder, hypothyroidism | Unknown | Pathogenic     |
| Family 31 | 16p11.2          | Del      | 28.83     | 29.04   | 0.21      | 13             | 16p11.2 del | VUR, seizure, DD, LD, and optic edema | Unknown | Pathogenic     |
| Family 33 | 3q29             | Dup      | 197.51    | 197.59  | 0.08      | 2              | –         | VUR, cataract, and growth delay | Unknown | VUS            |
| Family 25 | 2p24.3           | Del      | 15.30     | 15.38   | 0.08      | 1              | –         | PUV, heterotaxy | Inherited | VUS            |
| Family 21 | 4q35.1           | Dup      | 185.99    | 189.11  | 3.12      | 30             | –         | PUV | Inherited | VUS            |

| Family ID | Chromosomal region | CNV type | Start (Mb) | End (Mb) | Size (Mb) | Number of genes | Syndrome | Phenotype | Parental studies | Comments |
|-----------|------------------|----------|-----------|---------|-----------|----------------|----------|-----------|-----------------|----------|
| Family 34 | 22q11            | Trp      | 16.63     | 18.64   | 2.01      | 33             | Cat-eye syndrome | VUR, further details in text (multiple anomalies) | De novo  | Pathogenic     |
| Family 39 | 16p13.11         | Dup      | 15.12     | 16.29   | 1.17      | 19             | 16p13.11 dup | MCDK, facial dysmorphic features | Unknown | Pathogenic     |
| Family 10 | 16p11.2          | Dup      | 29.68     | 30.20   | 0.52      | 35             | 16p11.2 dup | Solitary kidney, psychiatric disorder, hypothyroidism | Unknown | Pathogenic     |
| Family 31 | 16p11.2          | Del      | 28.83     | 29.04   | 0.21      | 13             | 16p11.2 del | VUR, seizure, DD, LD, and optic edema | Unknown | Pathogenic     |
| Family 33 | 3q29             | Dup      | 197.51    | 197.59  | 0.08      | 2              | –         | VUR, cataract, and growth delay | Unknown | VUS            |
| Family 25 | 2p24.3           | Del      | 15.30     | 15.38   | 0.08      | 1              | –         | PUV, heterotaxy | Inherited | VUS            |
| Family 21 | 4q35.1           | Dup      | 185.99    | 189.11  | 3.12      | 30             | –         | PUV | Inherited | VUS            |
Laboratory at BMGL was queried for other de novo SNVs in FOXP1. We identified seven more de novo SNVs in this gene among approximately 5,000 patients (Table 3). Relationships (paternity and maternity) were confirmed by inheritance of rare SNVs from each parent in cases 3–8. In case 2, paternity was confirmed by inheritance of rare SNVs from the father. Maternity, however, could not be genetically confirmed per consent and was verified by pregnancy history.

All eight individuals had neurodevelopmental phenotypes consistent with loss-of-function variants in FOXP1 (MIM 613670). However, four out of eight individuals also had upper urinary tract defects, and five had defects in the lower genital-urinary (GU) tract, including undescended testis, hypospadias, and neurogenic bladder. In addition, these patients have brain and heart involvement, which is consistent with the role of FOXP1 in development of these two organs.22,23 CNS malformations including hydrocephaly and cardiac defects were among the phenotypes of the patients in this study. The genotypes and phenotypes of these individuals (6 out of 8 with upper or lower urinary tract defects) are summarized in Table 3. The pLI score of FOXP1 is 1, which suggests this gene is intolerant to loss-of-function variants.

**CNV discovery from WES data**

CNVs were inferred from WES data as described in the Materials and Methods section. Pathogenic CNVs and CNVs of uncertain clinical significance are summarized in Table 4. A de novo 22q11.1q11.21 triplication was identified in family 34; the proband had syndromic VUR. This triplication is proximal to the DiGeorge region, consistent with the gain of genetic material seen with type I supernumerary inv dup(22)(q11), which is associated with cat-eye syndrome25 (MIM 115470) (family 34). This patient’s phenotype has overlap with Goldenhar or oculo-auriculo-vertebral spectrum (OAV; MIM 164210) and VATER Association (MIM 192350). The patient was an 11-year-old Latin American boy with short stature, imperforate anus, thumb anomaly, severe gastroesophageal reflux disease (GERD), VUR, neurogenic bladder, right renal hypoplasia with evidence of scarring/renal damage, bilateral ear tags, ocular Duane anomaly, and left microphthalmia. The patient had normal cognition, although with some difficulty in mathematics.

Three other pathogenic CNVs were found in regions associated with known syndromes, namely 16p11.2 deletion, 16p11.2 duplication, and 16p13.11 duplication. CNVs in all individuals in Table 4 were validated by aCGH. Parental studies were also performed by aCGH. In family 34, in which samples were available from both parents, the CNV was found to be de novo. The flowchart of copy-number data inference, CoNIFER, and aCGH data for de novo 22q11 triplication are shown in Supplementary Figure S1 online.

**DISCUSSION**

In the past 4 years, WES has become a powerful clinical test for defining both recognized and previously undefined genes and potential variant susceptibilities to establish molecular diagnoses for birth defects. Clinical WES identifies pathogenic SNVs in approximately 25% of pediatric cases (mostly syndromic) that represent diagnostic dilemmas refractory to clinical diagnosis despite previous extensive medical evaluations.5,6 Nevertheless, the utility of WES for molecular diagnosis in isolated birth defects including CAKUT remains uncertain. This study shows that WES can be used in the diagnostic setting to define the molecular defects that underlie CAKUT and to reveal additional insights into the clinical presentation of the disorder. In addition, WES can be used for the identification of new candidate genes.

In family 1 (Figure 1 and Table 2), the diagnosis of renal coloboma syndrome (MIM 120330) was possible only after WES results became available. Prior to WES, the proband was using immunosuppressants for proteinuria but after molecular diagnosis management was changing to a tapering dose of immunosuppressive therapy. This allows avoidance of unnecessary immunosuppression because the etiology of kidney disease is not immunologic in this family. The molecular diagnosis of this family obtained by WES thus affected clinical decision making for the patient and the prognosis and management for family members. These data also expand the phenotype related to PAX2 pathogenic variants because the proband has membranous nephropathy and other family members have proteinuria and renal dysplasia. To date, focal segmental glomerulosclerosis has been reported with PAX2 variants26; membranous nephropathy is a novel finding. In family 2, a novel frameshift pathogenic SNV (p.Q378fs) was identified in HNF1B. This molecular diagnosis concluded by WES further substantiated the clinical phenotype as described in the results, thus minimizing the necessity for additional diagnostic evaluation. The variant in this family is novel, which adds to our current knowledge of diseases related to the HNF1B gene.

The phenotype of family 3 with the de novo EYA1 variant suggests that underlying genetic predisposition can lead to or at least exacerbate renal pathology in patients with VUR. Variants in EYA1 can cause branchio-oto-renal syndrome (BOR; MIM 113650), an autosomal dominant disorder characterized by sensorineural, conductive, or mixed hearing loss, structural defects of the outer, middle, and inner ear, branchial fistulas or cysts, and renal abnormalities ranging from mild hypoplasia to agenesis. The c.867+5G>A SNV does not affect the invariant splice site; nevertheless, RNA analysis of samples from patients with BOR showed that this SNV affects EYA1 splicing, producing an aberrant mRNA transcript that lacks exon 8 and results in premature termination in exon 9.27 The proband in this family will be evaluated for hearing impairment because the SNV in this individual causes BOR. This family provides another example that WES can improve the clinical diagnosis of syndromic forms of CAKUT beyond clinical evaluation alone.

Families 1, 2, and 3 exemplify the effects of WES on the clinical management of the patients and families because identification of the SNVs in PAX2, HNF1B, and EYA1, respectively, warranted further investigations in other organ systems. We identified only a fraction of families (3/62 = 4.8%) with...
pathogenic SNVs, similar to a recent large study that evaluated 749 individuals from 650 families with CAKUT for variants in 17 known CAKUT genes (6.3%).

CAKUT is a clinically heterogeneous clinical spectrum; therefore, many more genes and causal variants are likely to be identified. Next-generation sequencing and specifically WES have improved discovery of novel causative genes.7,29–34 We have identified a novel gene (FOXPI) that likely contributes to the CAKUT phenotype. We found eight novel different de novo pathogenic SNVs from both clinical and research WES in FOXPI in unrelated individuals. As summarized in Table 3, the phenotypes observed in these individuals suggest a clinical pattern that may be potentially recognizable. Structural brain anomalies (including hydrocephaly), intellectual disability, developmental delay, cardiac defects, hypotonia, behavior problems, and renal/GU defects are some of the more common features of this syndrome. Six out of eight individuals in this study (Table 3) have a known renal/GU phenotype in addition to other organ involvement. Although de novo disruptions in FOXPI were recently discovered to cause intellectual disability (OMIM 613670),23,24 here we defined a new syndrome that is characterized by hydrocephalus/brain malformation, cognitive impairment, cardiac defects, and CAKUT attributable to a single gene with pleiotropic effect. We recommend that patients with pathogenic or likely pathogenic variants in FOXPI should undergo renal ultrasound. Upper tract defects may remain undiagnosed if ultrasound is not performed.

Although FOXPI has been shown to have important roles in the developmental process of key organs including lung, heart, and brain,22,24,38 there are no data regarding the role of this master transcription factor in kidney and urinary tract development. In this study, we showed the role of FOXPI in CAKUT and lower urinary tract defects. All the FOXPI SNVs identified in this study were de novo and novel variants. These variants included frameshift, as documented for other birth defects such as the megacystis microcolon intestinal hypoperistalsis syndrome due to de novo SNVs in ACTG2.36 Approximately 6.5% (4/62) of our cohort have pathogenic CNVs that were extracted from WES data. In some families, organ involvement beyond CAKUT was sought retrospectively, after the review of WES results. We identified previously unrecognized genes and genetic variants (both SNVs and CNVs) in this cohort and expanded the phenotype of several known genes. Pathogenic SNVs in FOXPI in individuals with GU/renal phenotype strongly suggest an important role for this gene in urinary tract development.

**Supplementary Material**

Supplementary material is linked to the online version of the paper at http://www.nature.com/gim

**Acknowledgments**

This study was supported in part by K12 DK083014 Multidisciplinary K12 Urologic Research Career Development Program and R01 DK078121 from the National Institute of Diabetes and Digestive and Kidney Diseases awarded to D.J.L. This work was funded in part by grant R01 NS058529 (JRL) from the National Institute of Neurological Disorders and Stroke and U54 HG006542 from the National Human Genome Research Institute/National Heart Lung and Blood Institute to the Baylor Hopkins Center for Mendelian Genomics. Supported in part also by a National Institutes of Human Genome Research (NHGRI) (U54HG003273) awarded to R.A.G. R.A.L. is a Senior Scientific Investigator of Research to Prevent Blindness whose unrestricted funding supported part of this study. DNA extraction was performed in laboratory of translational genomics (LTG) at Baylor College of Medicine by Patricia Hernandez and Gladys Zapata. The analysis of the data was performed by software generously available by Codified Genomics Company. We also express our sincere gratitude to our patients and their families for their willing participation in this study.
DISCLOSURE
The Department of Molecular and Human Genetics at Baylor College of Medicine derives revenue from clinical exome sequencing offered by the Baylor Miraca Genetlics Laboratories and Whole Genome Laboratory. Authors who are faculty members in the Department of Molecular and Human Genetics at Baylor College of Medicine are identified as such in the affiliation section. The authors declare no conflict of interest with the following exceptions: M.N.B. is the founder of Codified Genomics LLC, a genomic interpretation company; R.A.G. is CSO of the Baylor Miraca Genetics Laboratories; D.J.L. is on the Scientific Advisory Board of Cellmatrix Inc.; J.R.L. has stock ownership in 23andMe and Lasergen, is a paid consultant for Regeneron, and is a coinventor of US and European patents related to molecular diagnostics for inherited neuropathies, eye diseases, and bacterial genomic fingerprinting.

REFERENCES
1. Sanna-Cherchi S, Ravani P, Corbani V, et al. Renal outcome in patients with congenital anomalies of the kidney and urinary tract. Kidney Int 2009;76:528–533.
2. Mansoor O, Chandar J, Rodriguez MM, et al. Long-term risk of chronic kidney disease in unilateral multicystic dysplastic kidney. Pediatr Nephrol 2011;26:597–603.
3. Groothoff JW. Long-term outcomes of children with end-stage renal disease. Pediatr Nephrol 2005;20:849–853.
4. Vivante A, Kohl S, Hwang DY, Dworschak GC, Hildebrandt F. Single-gene causes of congenital anomalies of the kidney and urinary tract (CAKUT) in humans. Pediatr Nephrol 2014;29:695–704.
5. Yang Y, Muzny DM, Reid JG, et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med 2013;369:1502–1511. http://www.ncbi.nlm.nih.gov/pubmed2408041.
6. Yang Y, Muzny DM, Xia F, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA 2014;312:1870–1879.
7. Karaca E, Harel T, Pehlivan D, et al. Genes that affect brain structure and function identified by rare variant analyses of mendelian neurologic disease. Neuron 2015;88:499–513.
8. Jiang Y, Oldridge DA, Diskin SJ, Zhang NR. CODEX: A normalization and copy number variation detection method for whole exome sequencing. Nucleic Acids Res 2015;43:e39.
9. McKenna A, Hanna M, Banks E, et al. Genome Analysis Toolkit: a unified algorithm for variation discovery and genotyipng of sequence variants. Genome Res 2010;20:1297–1303.
10. Chollis D, Yu J, Evani US, et al. An integrative variant analysis suite for whole exome next-generation sequencing data. BMC Bioinformatics 2012;13:8.
11. Krumm N, Sudmant PH, Ko A, et al.; NHLBI Exome Sequencing Project. Copy number variation detection and genotyping from exome sequence data. Genome Res 2012;22:1525–1532.
12. Vijayaragakannan P, Fitzgerald T. Detection of copy number variation from exomes in the DDD and UK10K projects. UK10KOrg. 2012;10:7319. http://www.uk10k.org/assets/ashx_vijayaragakannan_et_al_2012.pdf.
13. Sanna-Cherchi S, Kinyuk K, Burgess KE, et al. Copy-number disorders are a common cause of congenital kidney malformations. Am J Hum Genet 2012;91:987–997.
14. Verbitisky M, Sanna-Cherchi S, Fasel DA, et al. Genomic imbalances in pediatric patients with chronic kidney disease. J Clin Invest 2015;125:2171–2178.
15. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–423.
16. Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med 2011;13:680–685.
17. Nicolaou N, Renkema KY, Bongers EM, Giles RH, Knoers NV. Genetic, environmental, and epigentic factors involved in CAKUT. Nat Rev Nephrol 2015;11:720–731.
18. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.
19. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 2010;26:2069–2070.
20. Liu X, Jian X, Boerwinkle E. dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions. Hum Mutat 2011;32:894–899.
21. Bezinner F, Frigge ML, Masson G, et al. Rate of de novo mutations and the importance of father’s age to disease risk. Nature 2012;488:471–475.
22. Chang SW, Mislankan M, Misra C, et al. Genetic abnormalities in FOXP1 are associated with congenital heart defects. Hum Mutat 2013;34:1226–1230.
23. Hamdan FF, Daud H, Rochefort D, et al. De novo mutations in FOXP1 in cases with intellectual disability, autism, and language impairment. Am J Hum Genet 2016;87:671–678.
24. Solis E, Graham SA, Vino A, et al. Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. Hum Mol Genet 2016;25:546–557.
25. McTaggart KE, Budarf ML, Driscoll CA, Emanuels BS, Ferreia P, McDermid HE. Cat eye syndrome chromosome breakpoint clustering: identification of two intervals also associated with 22q11 deletion syndrome breakpoints. Cytogenet Cell Genet 1998;81:222–228.
26. Barua M, Stellacci E, Stella L, et al. Mutations in PXK2 associate with adult-onset FSGS. J Am Soc Nephrol 2014;25:1942–1953.
27. Stockley TL, Mendoza-Londono R, Propst EJ, Sodhi S, Lupsis L, Papsin BC. A recurrent EYA1 mutation causing alternative RNA splicing in brachio-oto-renal syndrome: implications for molecular diagnostics and disease mechanism. Am J Med Genet A 2009;149A:322–327.
28. Hwang DY, Dworschak GC, Kohl S, et al. Mutations in 12 known dominant disease-causing genes clarify many congenital anomalies of the kidney and urinary tract. Kidney Int 2014;85:1429–1433.
29. Sanna-Cherchi S, Sampogna RV, Papena N, et al. Mutations in DSTYK and dominant urinary tract malformations. N Engl J Med 2013;369:621–629.
30. Saisavat T, Kohl S, Hilger AC, et al. Whole-exome sequencing reveals recessive mutations in TRAP1 in individuals with CAKUT and VACTERL association. Kidney Int 2014;85:1310–1317.
31. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annu Rev Med 2012;63:35–61.
32. Vivante A, Kleppa MJ, Schulz J, et al. Mutations in TBX18 cause dominant urinary tract malformations via transcriptional dysregulation of ureter development. Am J Hum Genet 2015;97:291–301.
33. Gonzaga-Jauregui C, Harel T, Gambin T, et al.; Baylor-Hopkins Center for Mendelian Genomics. Heterozygous de novo and inherited mutations in the development.
34. Yuan B, Pehlivan D, Karaca E, et al. Global transcriptional disturbances underlie Cornelia de Lange syndrome and related phenotypes. J Clin Invest 2015;125:636–651.
35. Li W, Lu MM, Zhang Y, Tucker PW, Zhou D, Morrissey EY. Foxp2 and Foxp1 cooperate to regulate lung and esophagus development. Development 2007;134:1991–2000.
36. Wangler MF, Gonzaga-Jauregui C, Gambin T, et al.; Baylor-Hopkins Center for Mendelian Genomics. Heterozygous de novo and inherited mutations in the smooth muscle actin (ACTG2) gene underlie megacystis-microcolon-intestinal hypoperistalsis syndrome. PLoS Genet 2014;10:e1004258.
37. Retterer K, Scuffins J, Schmidt D, et al. Assessing copy number from sequencing and exome array CGH based on CNV spectrum in a large clinical cohort. Genet Med 2015;17:623–629.
38. Rosas PR, Sijstems JM, Theunsen MM, et al. Phenotypic variability of the cat eye syndrome. Case report and review of the literature. Genet Couns 2001;12:273–282.
39. Knijnenburg J, van Bever Y, Hulsman LO, et al. A 600 kb triplication in the cat eye syndrome critical region causes anorectal, renal and preauricular anomalies in a three-generation family. Eur J Hum Genet 2012;20:986–989.