KI REPORTS
KIReports.org
CLINICAL RESEARCH

The Importance of Copy Number Variant Analysis in Patients with Monogenic Kidney Disease

Laura R. Claus[1], Robert F. Ernst[1], Martin G. Elferink[1], Hanneke W.M. van Deutekom[1], Bert van der Zwaag[1] and Albertien M. van Eerde[1]

[1]Department of Genetics, University Medical Centre Utrecht, Utrecht, the Netherlands

Introduction: Genetic testing can reveal monogenic causes of kidney diseases, offering diagnostic, therapeutic, and prognostic benefits. Although single nucleotide variants (SNVs) and copy number variants (CNVs) can result in kidney disease, CNV analysis is not always included in genetic testing.

Methods: We investigated the diagnostic value of CNV analysis in 2432 patients with kidney disease genetically tested at the University Medical Centre Utrecht between 2014 and May 2022. We combined previous diagnostic testing results, encompassing SNVs and CNVs, with newly acquired results based on retrospective CNV analysis. The reported yield considers both the American College of Medical Genetics and Genomics (ACMG) classification and whether the genotype actually results in disease.

Results: We report a diagnostic yield of at least 23% for our complete diagnostic cohort. The total diagnostic yield based solely on CNVs was 2.4%. The overall contribution of CNV analysis, defined as the proportion of positive genetic tests requiring CNV analysis, was 10.5% and varied among different disease subcategories, with the highest impact seen in congenital anomalies of the kidney and urinary tract (CAKUT) and chronic kidney disease at a young age. We highlight the efficiency of exome-based CNV calling, which reduces the need for additional diagnostic tests. Furthermore, a complex structural variant, likely a *COL4A4* founder variant, was identified. Additional findings unrelated to kidney diseases were reported in a small percentage of cases.

Conclusion: In summary, this study demonstrates the substantial diagnostic value of CNV analysis, providing insights into its contribution to the diagnostic yield and advocating for its routine inclusion in genetic testing of patients with kidney disease.

Kidney Int Rep (2024) **9,** 2695–2704; https://doi.org/10.1016/j.ekir.2024.06.026
KEYWORDS: CNV analysis; *COL4A4*; diagnostic yield; ExomeDepth; genetic testing; nephrogenetics

Genetic testing in patients with kidney disease can reveal a monogenic cause for the patient's disease. A genetic diagnosis can have a diagnostic, therapeutic, and prognostic impact for patients and can also be of importance for counselling family members.[1] Both SNVs, including small insertions and deletions, and CNVs can result in monogenic kidney disease (MKD). CNVs are known to provide a significant part of the genomic burden in several MKD, such as in ciliopathies and CAKUT.[1] In a recent review, we found that CNVs are increasingly recognized as potential contributor to MKD, however even in recent publications ∼25% of studies did not include CNV analysis.[1] With read depth-based tools for CNV detection (e.g., ExomeDepth) emerging, massively parallel sequencing can detect both SNVs and CNVs.[2] This makes CNV detection more efficient because it does not require a separate test (e.g., SNP-array, and MLPA) for CNV detection.

In this study we aim to determine the diagnostic yield of CNV analysis in a cohort of patients with kidney disease. Our cohort comprises all 2432 probands that received a kidney disease gene panel at the genome diagnostic laboratory of the University Medical Centre Utrecht. It is an unselected cohort representing a cross-section of all clinical subcategories of nephropathy. By investigating the diagnostic yield of CNV analysis in this cohort, we aim to determine the contribution of CNVs to the diagnostic yield in patients with kidney disease. In doing so, we not only present the diagnostic outcomes of CNV analysis but also offer a comprehensive overview of our expert

Correspondence: Albertien M. van Eerde or Laura R. Claus, Department of Genetics, University Medical Centre Utrecht, KC.04.084.2, PO Box 85090, Utrecht 3508 AB, the Netherlands. E-mail: A.vanEerde@umcutrecht.nl or L.R.Claus-3@umcutrecht.nl
Received 24 May 2024; accepted 17 June 2024; published online 22 June 2024

center's diagnostic results over the past 8 years, covering both SNVs and CNVs.

METHODS

Patient Selection

All patients who received a gene panel focused on MKD from 2014 until May 12, 2022 at the genome diagnostic laboratory of the University Medical Centre Utrecht were included in this analysis. This includes genetic testing requested by physicians from inside the University Medical Centre Utrecht academic hospital, as well as from other hospitals. The laboratory started offering kidney disease gene panels in 2014. The panel content over the years can be found in Supplementary Table S1 (largest panel contained 495 genes in 2022). In 2018, exome-based gene panel sequencing was implemented as standard practice, enabling read depth-based CNV calling. Exome-based CNV calling using ExomeDepth[2] was validated for diagnostic use in our center in April 2021 (Supplementary Methods) and has been part of genetic testing reports since 2020 (still undergoing validation at that time). The exome-based gene panels allowed us to add retrospective CNV analysis for all patient samples sent since 2018.

CNV Calling and Assessment

CNVs were called for patients who underwent exome-based gene panel sequencing. CNV detection was performed using a modified version of ExomeDepth[2,3] and in-house developed scripts.[4] ExomeDepth uses an algorithm to determine CNVs based on the read depth information of a patient compared with a matched reference set (Supplementary Methods).

Only CNVs overlapping with at least 1 gene included in the requested diagnostic gene panel were analyzed (based on the most recent gene panel composition of 2022; Supplementary Table 1), because CNV calling outside of the diagnostic region was not part of the diagnostic consent. The assessment of CNVs was conducted using an in-house pipeline based on the ACMG criteria.[5,6] This in-house pipeline includes a quality control of the CNV data (Supplementary Methods). All resulting potential disease-causing variants were discussed with a clinical laboratory geneticist specialized in nephrogenetics to ensure that the reported CNVs are true positives in accordance with the diagnostic analysis criteria.

Identified pathogenic CNVs not yet reported in diagnostics were reported back to the referring physician, to assess whether a diagnosis could still be relevant for the patient.

Data Extraction for all Patients

Genetic testing results from gene panels focused on MKD for all patients who met patient selection criteria were extracted. The extracted data combined previous diagnostic testing results, with the newly acquired results based on retrospective CNV analysis. For the included patients that underwent other genetic tests (e.g., other than multigene panel for MKD) at our department or laboratory, genetic test results positive for a MKD causing genotype were also extracted and included in the analysis. For each variant, the annotation of whether this was an SNV (including small insertions and deletions detected with regular massively parallel sequencing) or CNV, zygosity, and ACMG classification was recorded. If for a specific variant ACMG classification differed between 2 patients, the most recent classification was used.

(Likely) pathogenic variants, as per ACMG criteria, underwent further interpretation regarding their potential for causing disease. The patient's genotype was categorized as "could be disease-causing," if zygosity matched the known inheritance pattern (based on Online Mendelian Inheritance in Man,[7] Clinical Genomic Database[8] [both downloaded in November 2022], and/or PubMed), and variants were confirmed to be in trans for compound heterozygosity. If segregation was not available to confirm that the variants were on separate alleles, but all other beforementioned criteria applied, it was concluded to be plausible that the genotype could be disease causing. This category also included homozygous/compound heterozygous *PKD1* variants that included a hypomorphic allele, a case with 2 *SLC12A3* variants of which 1 was 30% mosaic, and a case with a hemizygous *AVPR2* variant that has previously been reported in control database Exome Aggregation Consortium,[9] but never in a hemizygous state. A separate category was created for variants known as risk factors or for which incomplete penetrance or variable expression was known, as well as a separate category for additional findings unrelated to kidney disease.

Furthermore, the detection method was described for both SNVs and CNVs. For exome-based CNV calling using ExomeDepth, this included whether it was done at the time of diagnostic testing or retrospectively.[10] For each patient, the number of CNVs detected per panel, the size of the diagnostically relevant CNV, and the overlapping gene that was determined to be relevant were also included in the overview. The requested MKD gene panel, age (categorized as adult, pediatric, or prenatal) at time of testing, type of hospital, and specialty of the requesting physician were also extracted, as well as whether the individual was considered the proband of the family (detailed in Supplementary Methods).

Analysis of Diagnostic Yield

The number of probands was used to determine the overall diagnostic yield for the entire cohort, based on (L)P variants, and highlighting (i) the proportion of the yield that could be disease causing, (ii) the proportion for which disease causality was plausible (but segregation/causality not confirmed) , (iii) the proportion in which part of the phenotype was explained by the genotype, (iv) the proportion that carried a risk factor or a variant for which incomplete penetrance or variable expression is known, and (v) additional findings unrelated to kidney disease, ranked in that order. If a patient had >1 relevant result, the highest ranked category was used for further analysis. The overall yield of genetic testing in this diagnostic cohort was plotted, as well as per age at time of testing, year, and requested gene panel. For these last 2, it should be noted that some patients had multiple gene panels requested, which might have occurred in different years, yielding the same diagnostic results. The yield that is shown represents the diagnostic yield for that specific panel or year. For example, a patient receiving an MKD panel for Alport syndrome and CKDY could yield the same *COL4A3* pathogenic variant. The overall yield for the diagnostic cohort is calculated by counting each proband only once. The diagnostic yield per variant type was also calculated. Furthermore, the contribution of CNV analysis (defined as the proportion of positive genetic tests requiring CNV analysis), was determined and displayed. The different CNV testing methods were visualized using pie charts. For each gene panel, the most prevalent causative genes were determined. This was also determined for the CNVs separately. Excel (Version 2302 Build 16.0) PivotTables were used for analysis and visualization.

RESULTS

Diagnostic testing using MKD gene panels was performed in 2622 individuals from 2014 till May 2022 (Figure 1a). This included 2432 probands and 190 family. In the probands 2918 multigene panels were requested, with multiple panels performed in 393 individuals. At least 1 (L)P variant was detected in 903 individuals. Also, in 301 individuals variants of uncertain significance (VUS) were detected for which the zygosity matched the known inheritance for that gene. In 47 of these individuals there was also a relevant (L)P variant reported. The probands and variants were further assessed based on zygosity (including whether the variants were confirmed to be in trans), disease penetrance and disease causality (Figure 1a).

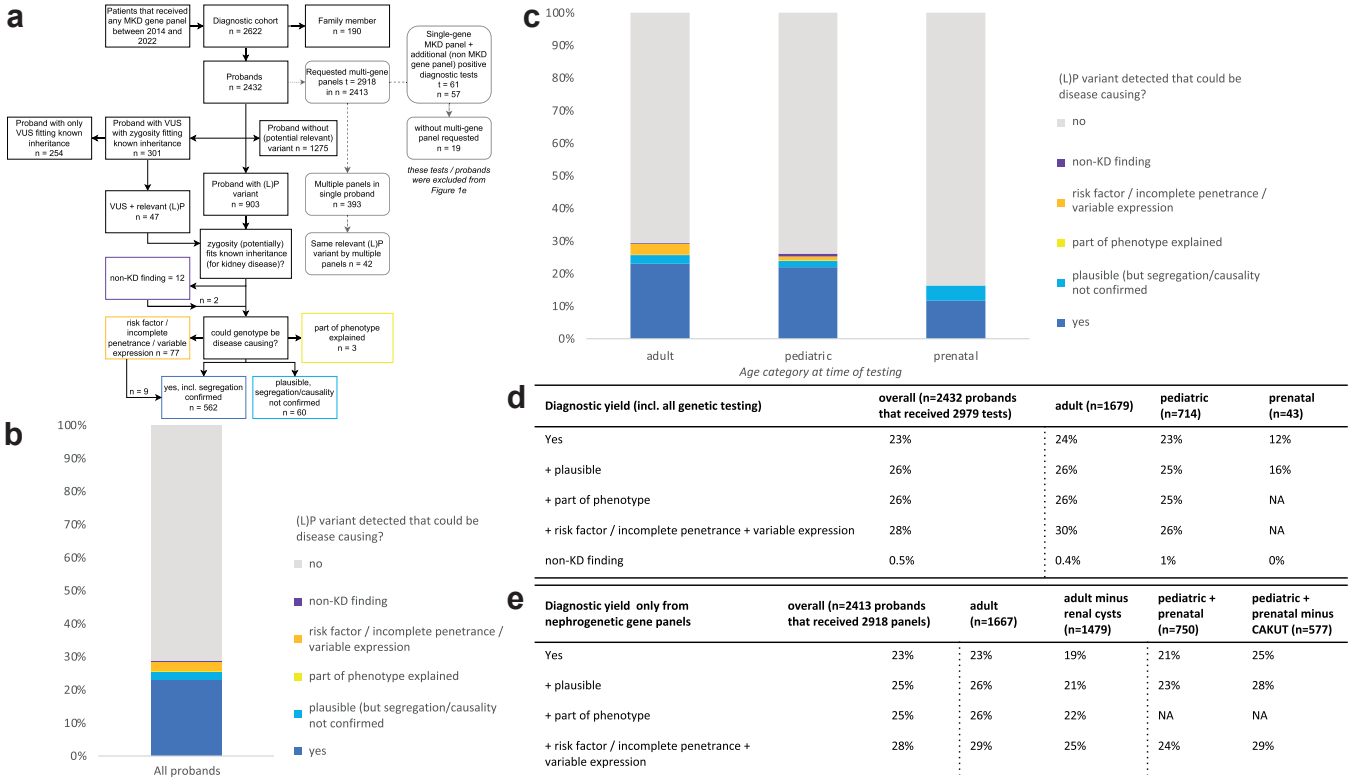
Diagnostic Yield in Diagnostic Cohort

Considering only (L)P variants that could be disease causing (including confirmation of compound heterozygosity) the diagnostic yield of our diagnostic cohort was 23% (Figure 1b and d). When we add (L)P variants that were plausible to be disease causing, the yield increased to 26% and remained 26% with adding patients for which part of the phenotype is explained. Adding also (L)P variants that are known as risk factors or known to have incomplete penetrance or variable expression the yield increased to 28% (Figure 1b and d).

The diagnostic yield was comparable in the adult and pediatric group, which is determined at the time of testing (Figure 1c). The prenatal group for which an MKD gene panel was requested had a diagnostic yield of 16% when (L)P variants plausible to be disease causing were also included. The CAKUT gene panel was requested most often in pediatric and prenatal cases (Supplementary Figure S1), being in 26% of requested tests. The yield of this panel was only 5%, which is comparable to previous reports.[1] When we remove the CAKUT gene panel and recalculate, the diagnostic yield increased by ∼5% (Figure 2e). For adults, the panel "renal cysts (adults)/autosomal dominant tubulointerstitial kidney disease" had a relatively high impact with a diagnostic yield of 52% (Supplementary Figure S2), which comprises 12% of requested tests. Removing this panel decreased the yield by ∼5% for adults (Figure 2e).

CNV Contribution to Diagnostic Yield

The diagnostic yield based solely on CNVs was 2.4% for our complete diagnostic cohort, considering all diagnostic result categories (Figure 2a). For 0.6% of the diagnostic yield the reported result included both a SNV and a CNV. Altogether, CNV analysis was necessary in 10.5% of reported diagnostic results (Figure 2a). How this contribution of CNV analysis, defined as the proportion of positive genetic tests requiring CNV analysis, was distributed per diagnostic result category is shown in Figure 2c, including what variant types were detected. Figure 2d shows the variant type and contribution of CNV analysis per MKD panel. In gene panels CKDY, CAKUT, and renal ciliopathies the highest contribution of CNV analysis to the diagnostic yield was found, followed by electrolyte abnormalities and nephrocalcinosis/nephrolithiasis (Figure 2d). The different CNV detection methods are shown in the pie chart in Figure 2b. Nearly half of these variants were detected with a separate test (i.e., with MLPA and SNP-array). However, >75% were detected with (retrospective) ExomeDepth CNV analysis, with 25% of those also being identified through a separate test.



d

Diagnostic yield (incl. all genetic testing)	overall (n=2432 probands that received 2979 tests)	adult (n=1679)	pediatric (n=714)	prenatal (n=43)
Yes	23%	24%	23%	12%
+ plausible	26%	26%	25%	16%
+ part of phenotype	26%	26%	25%	NA
+ risk factor / incomplete penetrance + variable expression	28%	30%	26%	NA
non-KD finding	0.5%	0.4%	1%	0%

e

Diagnostic yield only from nephrogenetic gene panels	overall (n=2413 probands that received 2918 panels)	adult (n=1667)	adult minus renal cysts (n=1479)	pediatric + prenatal (n=750)	pediatric + prenatal minus CAKUT (n=577)
Yes	23%	23%	19%	21%	25%
+ plausible	25%	26%	21%	23%	28%
+ part of phenotype	25%	26%	22%	NA	NA
+ risk factor / incomplete penetrance + variable expression	28%	29%	25%	24%	29%

Figure 1. Overall yield of genetic testing in diagnostic cohort. (a) Flowchart of the diagnostic cohort: patients receiving any MKD panel were included. The chart displays the number of probands (indicated with n), broken down by the number of probands with VUS and (likely) pathogenic ((L)P) variants. Both SNV and CNV were considered. The right side of this chart shows the number of requested tests (indicated with t) and instances of multiple panel requests and concordant relevant variants. For probands with a (L)P variant it was determined whether that genotype could be disease causing resulting in the categories "yes, incl. segregation confirmed," "plausible (but segregation/causality not confirmed)," "part of phenotype explained" and "risk factor/incomplete penetrance/variable expression." Nine probands had both a risk factor variant and a clear disease-causing variant. Findings unrelated to the patient's kidney disease (non-KD findings) were reported in 12 patients. In 2 patients this was in addition to another relevant (L)P variant. (b) Graph displaying overall diagnostic yield in all probands regardless of whether the diagnosis was reached through MKD panel testing. The highest-ranking category is displayed when (L)P variants were present in multiple categories. The colors of the categories correspond to the categories in the flowchart from a. "Yes" is short for "Yes, genotype could be disease causing, incl. segregation confirmed when applicable." (c) Graph displaying overall diagnostic yield in probands distributed by age at time of testing. Note that 26% of pediatric/prenatal panels are CAKUT panels, yielding 5% diagnostic variants, whereas 12% of adults received the "renal cysts (adults)/ ADTKD" gene panel, yielding 52% diagnostic variants (Supplementary Figure S1 and S2). Removal of these panels is shown in e. The highest-ranking category is displayed when (L)P variants were present in multiple categories. The colors of the categories correspond to the categories in the flowchart from a. "Yes" is short for "Yes, genotype could be disease causing, incl. segregation confirmed when applicable." (d) Diagnostic yield by (L)P variant category as displayed in b and c. Note that four patients transitioned from the pediatric/prenatal category to the adult category, with one obtaining a diagnosis as an adult. (e) Diagnostic yield based only on monogenic kidney disease panels (i.e., excluding additional tests as indicated on the right side of the flow chart). The diagnostic yield after excluding the most impactful panels for adults and pediatric/prenatal is also depicted (also see Supplementary Figure S1 and S2, showing panel-specific diagnostic yields). ADTKD, autosomal dominant tubulointerstitial kidney diesease; CAKUT, congenital anomalies of the kidney and urinary tract; CNV, copy number variants; (L)P variant, likely pathogenic; MKD, monogenic kidney disease; SNV, single nucleotide variants; VUS, variants of uncertain significance.

ExomeDepth CNV analysis

ExomeDepth CNV analysis was performed in 1841 probands (Supplementary Figure S3A). In 2% of probands the quality control metrics for CNV data did not meet the predefined criteria. However, in 17% of these samples this was caused by an additional finding unrelated to the patient's kidney disease (i.e., large chromosomic rearrangements often result in a failing quality control). CNVs were only called when these overlapped with a gene from the requested gene panel. There were 0 to 8 CNV calls detected per requested diagnostic test (Supplementary Figure S3B).

Recurrent CNVs revealing likely COL4A4 founder variant

In 3 patients we found a complex COL4A4 rearrangement. The CNV that was called in all 3 patients was a 17.7 kb duplication on chromosome 2, overlapping with *COL4A4*. Based on the orientation of the reads, we suspected this duplication to be part of a more complex

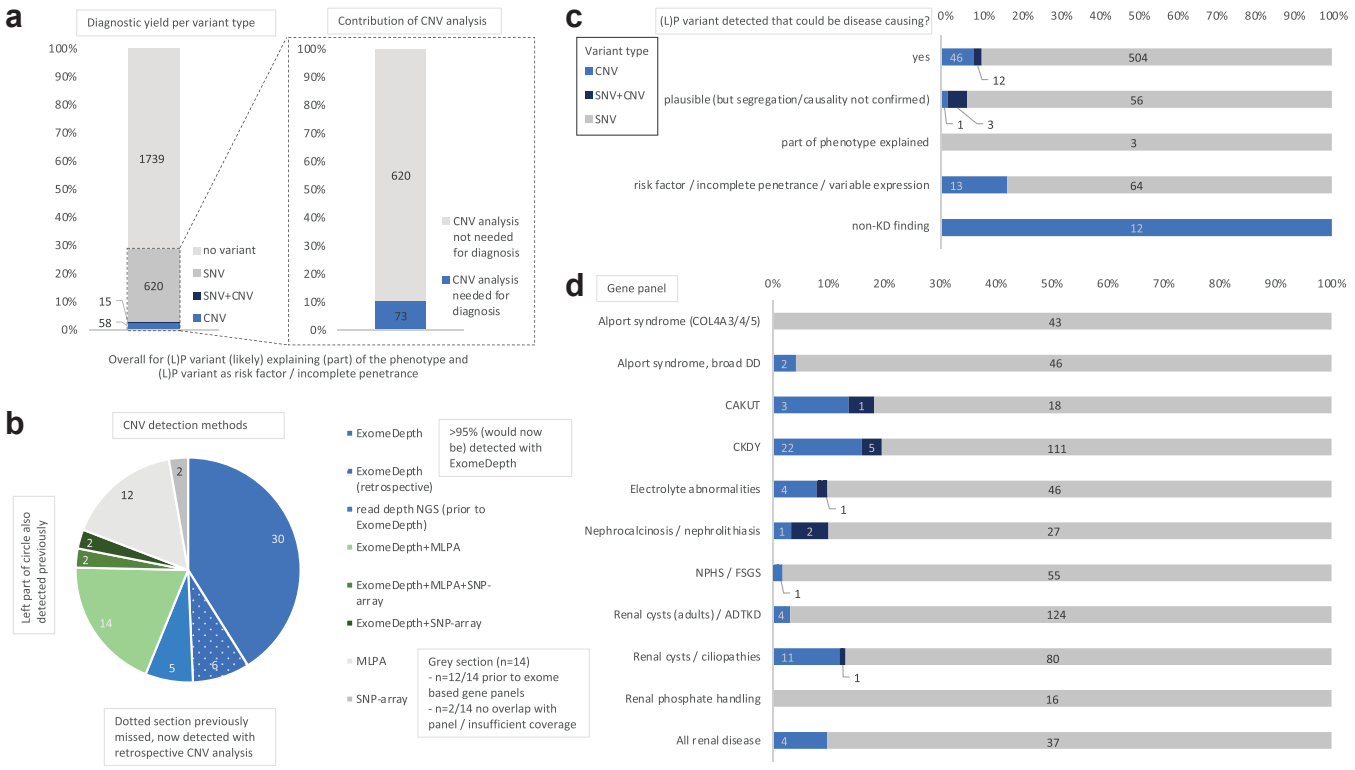


Figure 2. Analysis of CNV in diagnostic cohort.e The categories from Figure 1: "yes, incl. segregation confirmed," "plausible (but segregation/causality not confirmed)," "part of phenotype explained," and "risk factor/incomplete penetrance/variable expression" were aggregated for panel a, b, and d. Only the highest-ranking category was included when (L)P variants were present in multiple categories in a single proband. (a) On the left the diagnostic yield per variant type for the total cohort. On the right the proportion of the diagnostic yield for which CNV analysis was necessary. (b) This pie chart provides an overview of the methods employed for CNV detection. (c) This section illustrates the types of variants (SNV vs. CNV) associated with reported (L)P variants per categorization of variants as in Figure 1. (d) This graph presents the distribution of variant types across gene panels for which the laboratory reported more than ten variants. ADTKD, autosomal dominant tubulointerstitial kidney disease; CAKUT, congenital anomalies of the kidney and urinary tract; CKDY, chronic kidney disease at a young age; DD, differential diagnosis; FSGS, focal segmental glomerulosclerosis; KD, kidney disease; MLPA, multiplex ligation-dependent probe amplification; NPHS, nephrotic syndrome; NGS, next-generation sequencing; SNP-array, single nucleotide polymorphism array.

rearrangement. Using low-pass whole genome sequencing we were able to elucidate the exact structure of this rearrangement, which consist of a duplication of *COL4A4* and *DNER,* which is then inverted including the region in-between (Supplementary Figure S4). This results in disruption of 1 of the 2 *COL4A4* alleles leading to loss of function of this gene copy. It is unlikely that this exact rearrangement happened multiple times by chance. This complex *COL4A4* rearrangement is therefore likely a founder variant. Other recurrent CNVs in our diagnostic cohort are the commonly known recurrent full gene deletions of *NPHP1* and *HNF1B* and the recurrent deletion overlapping *CHFR1* and *CFHR3* (Supplementary Table S2).[10-13]

Additional Findings Unrelated to Kidney Disease

Findings unrelated to the patient's kidney disease were reported in 12 patients, which is 0.5% of our diagnostic cohort (Figure 1d). Two of these patients had both an additional finding and a diagnostic (L)P variant in an MKD gene (Figure 1a). All additional findings unrelated to kidney disease were identified through CNV analysis (Figure 2c). The reported additional findings were Klinefelter syndrome (XXY) in 5 patients, trisomy 21 in 4 patients, 2 patients with trisomy X, and 1 patient with an unbalanced XY translocation. For most, these diagnoses were already known and previously confirmed with genetic testing (Supplementary Table S2).

Core Genes and Possible Phenocopies

Relevant (L)P variants were detected in 104 different genes. However, only 6 genes (i.e., *PKD1*, *COL4A3*, *COL4A4*, *PKD2*, *SLC12A3*, and *COL4A5*) were responsible for >50% of the cases with such a diagnostic variant (Figure 3a). Considering only reported CNVs, we found that these were most often reported in *NPHP1*, *HNF1B*, *SLC3A1*, and *CFHR1* + *CHFR3* (Figure 3b). A detailed overview of the number of genes with a reported variant per requested gene panel

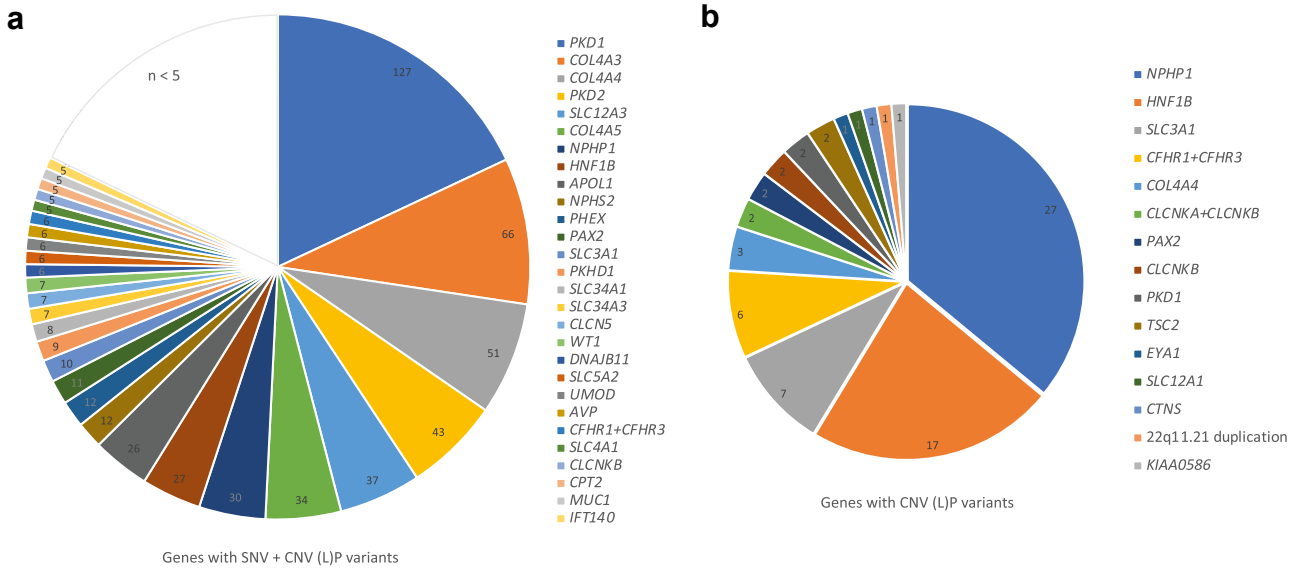


Figure 3. Genes in which (likely) pathogenic variants were detected. The categories from Figure 1: "yes, incl. segregation confirmed", "plausible (but segregation/causality not confirmed)," "part of phenotype explained," and "risk factor/incomplete penetrance/variable expression." were aggregated The number of probands with a (L)P variant in a gene is shown. (a) Genes with (L)P variants detected. This includes both SNV and CNV. Supplementary Table S3 provides the detailed list of genes in which (L)P variants were identified for each requested test. (b) Genes with (L)P CNV detected. Supplementary Table S4 provides the detailed list of genes in which (L)P CNV were identified for each requested test. CNV, copy number variants; (L)P variant, likely pathogenic; SNV, single nucleotide variants.

can be found in Supplementary Table S3 and Supplementary Table S4, with the latter containing the breakdown of CNVs. Some phenocopies can be suspected from this. For example, *COL4A3*, *COL4A4*, and *CLCN5* variants were reported in 7, 8, and 2 probands, respectively, who received a nephrotic syndrome/focal segmental glomerulosclerosis gene panel. Moreover, a *COL4A4* variant was detected in patients for whom a renal cysts/ciliopathy gene panel was requested. Furthermore, *NPHP1* and *NPHP4* disease causing variants were identified in probands receiving a CAKUT gene panel.

DISCUSSION

In our extensive cohort, comprising diagnostic testing results from over 2400 probands who received MKD gene panel testing within the past 8 years, we identified the pivotal role of CNV analysis in establishing a diagnosis in 10.5% of probands with a positive genetic test. This is comparable to other genetic diseases.[14–16] Our analysis revealed the paramount contribution of CNV analysis to the diagnostic yield in patients who underwent gene panel testing for CKDY, CAKUT, or renal ciliopathies, a finding consistent with previously reported data.[1] When undiagnosed patients with these phenotypes have not had CNV analysis performed, it is recommended to consider (retrospective) CNV analysis.

The use of recently developed exome-based CNV detection tools, such as ExomeDepth, obviates the need for a separate diagnostic CNV test when exome-based gene panels are performed. We expect that ExomeDepth CNV analysis would have been able to detect >95% of reported CNVs in our study, because these regions are now sufficiently covered in our exome-based gene panels with a minimum of 3 probes. Other common CNV detection methods are MLPA, chromosomal microarray (e.g., SNP-array), and whole-genome sequencing-based CNV detection. Advantages and disadvantages of the different methods are discussed in Table 1. Exome-based CNV detection is currently most advantageous, because exome sequencing is already routinely used in diagnostic laboratories and requires no additional tests. It makes genetic testing more cost efficient and represents another step toward a single genetic test that can detect most genetic diagnoses.[17] However, it is crucial to remain aware of genetic variants that cannot be detected with exome-based gene panels, such as most cytosine insertions in the variable number tandem repeat region in *MUC1* or single exon CNVs with insufficient coverage.[18]

Because ExomeDepth CNV calling was introduced in our center in 2020, we performed retrospective analysis using ExomeDepth for all patients who had an exome-based gene panel from 2018 onwards. This yielded a novel diagnosis in 6 patients, including a clear disease-causing variant in 2 patients (e.g., *TSC2* and a *PKD1* + *TSC2* deletion), and a risk factor based on a *SLC3A1* CNV in 4 patients. The number of missed diagnoses before the introduction of ExomeDepth CNV calling was thus

Table 1. Advantages and disadvantages of common CNV detection methods

CNV detection method	Resolution[a]	Advantages	Disadvantages
MLPA	Single exon	- High resolution, accurate detection of single exon deletions/duplications - No incidental findings	- Single test per gene - Limited by availability of MLPA kits
Chromosomal micro-arrays	Single/ multi gene/ multi exon	- Cost-effective when CNV as cause is highly suspicious	- Possible incidental findings - Separate diagnostic test
Exome-based CNV detection	Multi exon	- No additional diagnostic test needed for combined SNV/CNV analysis - Exome sequencing is currently routinely performed in diagnostics - Sporadic detection of exact breakpoints possible	- Possible incidental findings - Bioinformatics expertise required for implementation
Genome-based CNV detection	Multi exon (single exon)	- No additional test needed for combined SNV/CNV analysis - Noncoding regions covered - Detection of exact breakpoints often possible	- Possible incidental findings - Sequencing costs (but declining) - Bioinformatics expertise required for implementation

CNV, copy number variant; SNV, single nucleotide variant.
[a]Minimal resolution reported. Resolution is dependent on assay design, probe density and stability, additionally the resolution may also be influenced by factors such as sample quality and complexity of the genomic region under scrutiny.

limited in our center. Hence, we conclude that before the introduction of ExomeDepth, Dutch clinicians ordered a separate CNV test, or specific CNV testing had been part of the requested panel (i.e., MLPAs for commonly deleted genes had been part of MKD gene panels in our center [Supplementary Table S1]).

The additional testing also explains why there was not a clear increase in overall diagnostic yield after ExomeDepth was introduced in our center (Supplementary Figure S5A). One would expect an increase in diagnostic yield over time merely based on the discovery of new genes. However, we did not observe this, which might be because of genetic testing being offered to more patients and not only to patients with a high suspicion of a genetic disease. Also, genetic testing in the context of living related kidney donation may have an impact. In fact, there was an increase in requested tests over the past years (Supplementary Figure S5B and S5C), likely because of (inter)national guidelines and a large observational study.[19-22]

One potential drawback of CNV analysis could be the identification of additional findings unrelated to the patient's kidney disease. In our cohort most of these findings had previously been detected with another genetic test. Because we only had limited access to the patient's phenotype data, we could not determine which of these findings were incidental findings, and it could be possible that we underestimated the number of additional findings. We could for instance have detected a (L)P variant in a kidney disease gene that did not actually explain this patient's phenotype and should have been marked as a finding unrelated to the patient's kidney disease.

This study offers a complete overview of our expert center's diagnostic results over the past 8 years. It is an unselected cohort representing a cross-section of all clinical subcategories of nephropathy. Because genetic testing is reimbursed in the Netherlands, we do not expect that the (perceived) cost of testing influenced the selection of patients. We provide an unbiased estimate of the contribution of CNVs to the diagnostic yield in patients with MKD. Retrospective CNV analysis for patients from 2018 to 2021 only resulted in an additional diagnosis for a few patients. Hence, we do not expect the probands from before 2018 without retrospective ExomeDepth CNV analysis, which comprise 24% of the cohort, having a big impact on the reported diagnostic yield in our cohort. However, this outcome may vary in other centers. Whether there is a duty to recontact and reanalyze has extensively been explored in literature.[23-25] In our center we instruct patients with negative testing results to recontact us after 3 to 5 years, or sooner when a couple wishes to conceive, additional symptoms emerge, or family members are diagnosed with kidney disease. It should be noted that we only analyzed CNVs overlapping with the requested gene panel as covered by the patient's consent. Therefore, CNVs outside of the gene panel were not considered, unless an additional test (e.g., SNP-array) was requested. When requesting genetic testing, we recommend verifying whether exome-based CNV testing is performed and to what extent, because practices vary among genome diagnostic laboratories.

Another strength of this study is that we based our diagnostic yield not only on the ACMG classification of variants, but also interpreted whether the genotype could be disease causing. Because we did not have access to detailed phenotype data, we were unable to do this on an individual patient level, which could affect the ACMG classification. However, we did provide on individual variant level whether a genotype could be disease causing or was only known as a risk factor for MKD (e.g., *APOL1*-variants or *SLC3A1*-variants) and clearly show these different categories in Figure 1. Furthermore, we did not have access to segregation data for interpreting individual variants of uncertain significance, resulting in a potential underestimation of the yield. Note that, in a diagnostic

setting it is important to always check whether a (L)P variant explains a patient's phenotype and to further analyze individual variants of uncertain significance (in family members) or refer a patient to a genetic professional. The diagnostic yield reported per phenotype group should be interpreted within the context of a clinical diagnostic setting, because there were no stringent scientific inclusion criteria. To more accurately determine the yield for a specific phenotype group, all patients with this specific well-characterized phenotype should be genetically tested. A limitation of our study could be that we included patients based on having received an MKD panel. We therefore did not include diagnoses that resulted from whole exome analysis as a first-tier test, which is for instance often done in prenatal cases.

Interestingly, our CNV analysis identified a likely founder variant in *COL4A4* resulting from a complex rearrangement. This variant illustrates the importance of identifying the boundaries of a CNV and using other techniques such as low-pass whole-genome sequencing to confirm. More recurrent variants are present in this dataset (Supplementary Table S2). These might be founder variants, but could also be recurrent because of other reasons, such as mutational hotspots based on CpG sites or CNVs resulting from nonallelic homologous recombination. Even if assuming that these patients would have hidden familial relationships, this would negatively affect the diagnostic yield by a maximum of 4% (excluding the known recurrent variants in *NPHP1*, *HNF1B*, *CFHR1* + *CFHR3*, and *APOL1*).

We also offer an analysis of gene panel requests categorized by requesting medical specialty over time (Supplementary Figure S5B and S5C). These data clearly demonstrate the increasing trend in genetic testing requests in recent years.

Finally, we show in which genes (L)P variants were most often reported providing a list of core genes per phenotype group (Supplementary Figure S2 and S3). The fact that a limited number of genes was responsible for >50% of diagnostic results, has previously been reported.[1,26] This supports a tiered approach (using exome-based virtual gene panels) in which a panel of genes known to be associated with the patient's phenotype is analyzed first, before potentially moving on to a complete set of known kidney disease genes, possibly followed by whole-exome analysis in selected cases. Potential reasons to deviate from this approach are prenatal cases, an urgent need for diagnosis and/or indistinct phenotypes. A tiered approach, as recommended by the ERA-EDTA Working Group for Inherited Kidney Diseases and the Molecular Diagnostics Taskforce of the European Rare Kidney Disease Reference Network,[20] limits the number of reported individual variants of uncertain significance and incidental findings and is 1 of the reasons we provide an extensive set of phenotype-driven gene panels in our center (Supplementary Table S1). However, one should be aware of phenocopies and consider additional testing after an initial negative result. We highlighted some potential phenocopies which have been associated with the respective phenotypes before but were constrained by limited phenotype information.[27–32] Similar phenocopies involving different genes are also reported in literature.[33–35] Detailed analyses of subgroups of our patients who consented to reporting on detailed phenotyping will prove valuable in elucidating these types of observations further. Patients can be approached to participate in the GeNepher data and biobank, which includes broad consent, allowing for this type of research.[36] For future research, we have also implemented a protocol using a "no objection" arrangement that allows for variant detection outside of the diagnostic region using an aggregated anonymous approach.[36]

In summary, our study demonstrates the substantial diagnostic value of CNV analysis in kidney diseases, providing insights into its contribution to the diagnostic yield and advocating for its inclusion in genetic testing of kidney disease patients. We also offer a comprehensive overview of our expert center's diagnostic results over the past eight year covering both SNVs and CNVs.

DISCLOSURES

All the authors declared no competing interests.

DATA AVAILABILITY STATEMENT

Variants identified in this diagnostic cohort are available in Supplementary Table S2. The genome diagnostic laboratory makes variants publicly available by uploading them to the Leiden Open Variation Database (LOVD) through the Dutch Society for Clinical Genetic Laboratory Diagnostics (VKGL).[37] Onsite access to the CNV data for verification purposes can be arranged.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the authors used ChatGPT to improve language (analogous to hiring a native speaking editor). After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the full staff and personnel of the UMC Utrecht genome diagnostic laboratory who contributed to generating the individual results. The authors of this publication are members of the European Reference Network for Rare Kidney Diseases (ERKNet). This work was supported by the Dutch Kidney Foundation (18OKG19 to AMvE). UMC Utrecht participates in the IMAGEN project (grant LSHM20009), which is co-funded by the PPP Allowance made available by Health Holland (Top Sector Life Sciences & Health).

AUTHOR CONTRIBUTIONS

LRC and AMvE designed the study. RFE and MGE did the bioinformatic analysis. HvD was involved in the design of the study and the bioinformatic analysis. LRC performed retrospective CNV analysis. BvdZ was the laboratory specialist involved that performed initial classification of the diagnostic dataset and checked relevant CNVs resulting from the retrospective analysis. LRC combined results from the diagnostic cohort and conducted further descriptive analysis. All authors critically assessed the paper. LRC drafted and revised the paper. All authors approved the final version of the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary Methods (PDF)

Supplementary material (xlsx)

Figure S1. Diagnostic yield of probands per panel for pediatric / prenatal age at time of testing.

Figure S2. Diagnostic yield of probands per panel for adult age at time of testing.

Figure S3. ExomeDepth CNV analysis: quality control and number of CNVs detected per requested test.

Figure S4. Complex COL4A4 rearrangement.

Figure S5. Diagnostic testing practices over time.

Table S1. Panel content over the years for monogenic kidney disease (MKD).

Table S2. Complete overview of detected variants per patient per requested test.

Table S3. Overview of the number of genes with a reported variant per requested gene panel.

Table S4. Overview of the number of genes with a reported copy number variant per requested gene panel.

RECORD checklist

REFERENCES

1. Claus LR, Snoek R, Knoers NVAM, van Eerde AM. Review of genetic testing in kidney disease patients: diagnostic yield of single nucleotide variants and copy number variations evaluated across and within kidney phenotype groups. *Am J Med Genet C Semin Med Genet*. 2022;190:358–376. https://doi.org/10.1002/ajmg.c.31995
2. Plagnol V, Curtis J, Epstein M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. *Bioinformatics*. 2012;28:2747–2754. https://doi.org/10.1093/bioinformatics/bts526
3. UMCUGenetics. GitHub - UMCUGenetics/ExomeDepth. ExomeDepth R package for the detection of copy number variants in exomes and gene panels using high throughput DNA sequencing data. GitHub. Accessed August 8, 2022. https://github.com/UMCUGenetics/ExomeDepth
4. UMCUGenetics. Dx_resources/ExomeDepth at master-UMCUGenetics/Dx_resources. GitHub. Accessed August 8, 2022. https://github.com/UMCUGenetics/Dx_resources/tree/master/ExomeDepth
5. 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;17:405–424. https://doi.org/10.1038/gim.2015.30
6. Riggs ER, Andersen EF, Cherry AM, et al. Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). *Genet Med*. 2020;22:245–257. https://doi.org/10.1038/s41436-019-0686-8
7. Amberger J, Bocchini CA, Scott AF, Hamosh A. McKusick's Online Mendelian Inheritance in Man (OMIM(R)). *Nucleic Acids Res*. 2009;37:D793–D796. https://doi.org/10.1093/nar/gkn665
8. Solomon BD, Nguyen AD, Bear KA, Wolfsberg TG. Clinical genomic database. *Proc Natl Acad Sci U S A*. 2013;110:9851–9855. https://doi.org/10.1073/pnas.1302575110
9. Karczewski KJ, Weisburd B, Thomas B, et al. The ExAC browser: displaying reference data information from over 60 000 exomes. *Nucleic Acids Res*. 2017;45:D840–D845. https://doi.org/10.1093/nar/gkw971
10. Saunier S, Calado J, Benessy F, et al. Characterization of the NPHP1 locus: mutational mechanism involved in deletions in familial juvenile nephronophthisis. *Am J Hum Genet*. 2000;66:778–789. https://doi.org/10.1086/302819
11. Yuan B, Liu P, Gupta A, et al. Comparative genomic analyses of the human NPHP1 locus reveal complex genomic architecture and its regional evolution in Primates. *PLoS Genet*. 2015;11:e1005686. https://doi.org/10.1371/journal.pgen.1005686
12. Mitchel MW, Moreno-De-Luca D, Myers SM, et al. 17q12 Recurrent Deletion Syndrome. In: Adam MP, Feldman J, Mirzaa GM, et al., eds. *GeneReviews*. Published online; 2016:1993–2023.
13. Cantsilieris S, Nelson BJ, Huddleston J, et al. Recurrent structural variation, clustered sites of selection, and disease risk for the complement factor H (*CFH*) gene family. *Proc Natl Acad Sci U S A*. 2018;115:E4433–E4442. https://doi.org/10.1073/pnas.1717600115
14. Zampaglione E, Kinde B, Place EM, et al. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. *Genet Med*. 2020;22:1079–1087. https://doi.org/10.1038/s41436-020-0759-8
15. Pennings M, Meijer RPP, Gerrits M, et al. Copy number variants from 4800 exomes contribute to ~7% of genetic

diagnoses in movement disorders, muscle disorders and neuropathies. *Eur J Hum Genet*. 2023;31:654–662. https://doi.org/10.1038/s41431-023-01312-0

16. Vulto-van Silfhout AT, Hehir-Kwa JY, van Bon BWM, et al. Clinical significance of de novo and inherited copy-number variation. *Hum Mutat*. 2013;34:1679–1687. https://doi.org/10.1002/humu.22442
17. Louw N, Carstens N, Lombard Z. Incorporating CNV analysis improves the yield of exome sequencing for rare monogenic disorders-an important consideration for resource-constrained settings. *Front Genet*. 2023;14:1277784. https://doi.org/10.3389/FGENE.2023.1277784/BIBTEX
18. Kirby A, Gnirke A, Jaffe DB, et al. Mutations causing medullary cystic kidney disease type 1 lie in a large VNTR in MUC1 missed by massively parallel sequencing. *Nat Genet*. 2013;45:299–303. https://doi.org/10.1038/ng.2543
19. KDIGO Conference Participants. Genetics in chronic kidney disease: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. *Kidney Int*. 2022;101:1126–1141. https://doi.org/10.1016/j.kint.2022.03.019
20. Knoers N, Antignac C, Bergmann C, et al. Genetic testing in the diagnosis of chronic kidney disease: recommendations for clinical practice. *Nephrol Dial Transplant*. 2022;37:239–254. https://doi.org/10.1093/ndt/gfab218
21. Snoek R, Bongers EHF, Brooks A, et al. Handreiking Erfelijke Nieraandoeningen en Urineafwijkingen en Nefrologische Verwijsindicatie Klinische Genetica. Nederlandse Federatie voor Nefrologie (NFN, Dutch Federation for Nephrology). 2018. Accessed March 14, 2024. https://www.nefro.nl/richtlijnen/handreiking-genetische-diagnostiek-bij-nierziekten-2018
22. de Haan A, Eijgelsheim M, Vogt L, et al. Diagnostic yield of massively parallel sequencing in patients with chronic kidney disease of unknown etiology: rationale and design of a national prospective cohort study. *BMJ Open*. 2022;12:e057829. https://doi.org/10.1136/bmjopen-2021-057829
23. Otten E, Plantinga M, Birnie E, et al. Is there a duty to recontact in light of new genetic technologies? A systematic review of the literature. *Genet Med*. 2015;17:668–678. https://doi.org/10.1038/GIM.2014.173
24. Appelbaum PS, Parens E, Berger SM, Chung WK, Burke W. Is there a duty to reinterpret genetic data? The ethical dimensions. *Genet Med*. 2020;22:633–639. https://doi.org/10.1038/S41436-019-0679-7
25. van der Geest MA, Maeckelberghe ELM, van Gijn ME, et al. Systematic reanalysis of genomic data by diagnostic laboratories: a scoping review of ethical, economic, legal and (psycho) social implications. *Eur J Hum Genet*. 2024;32:489–497. https://doi.org/10.1038/S41431-023-01529-Z. Published online 2024.
26. Groopman EE, Marasa M, Cameron-Christie S, et al. Diagnostic utility of exome sequencing for kidney disease. *N Engl J Med*. 2019;380:142–151. https://doi.org/10.1056/nejmoa1806891
27. Malone AF, Phelan PJ, Hall G, et al. Rare hereditary COL4A3/COL4A4 variants may be mistaken for familial focal segmental glomerulosclerosis. *Kidney Int*. 2014;86:1253–1259. https://doi.org/10.1038/ki.2014.305
28. Savige J, Harraka P. Pathogenic variants in the genes affected in Alport syndrome (COL4A3-COL4A5) and their association with other kidney conditions: a review. *Am J Kidney Dis*. 2021;78:857–864. https://doi.org/10.1053/j.ajkd.2021.04.017
29. Valina MR, Larsen CP, Kanosky S, Suchy SF, Nield LS, Onder AM. A novel CLCN5 mutation in a boy with asymptomatic proteinuria and focal global glomerulosclerosis. *Clin Nephrol*. 2013;80:377–384. https://doi.org/10.5414/CN107429
30. Chen Q, Li Y, Wu X. Dent disease manifesting as nephrotic syndrome, 01125 *Intractable Rare Dis Res. 2023*. 2022;12:67–70. https://doi.org/10.5582/irdr.2022.01125
31. Kubo K, Aizawa T, Watanabe S, et al. Does Dent disease remain an underrecognized cause for young boys with focal glomerulosclerosis? *Pediatr Int*. 2016;58:747–749. https://doi.org/10.1111/ped.12944
32. van der Ven AT, Connaughton DM, Ityel H, et al. Whole-exome sequencing identifies causative mutations in families with congenital anomalies of the kidney and urinary tract. *J Am Soc Nephrol*. 2018;29:2348–2361. https://doi.org/10.1681/ASN.2017121265
33. Pantel D, Mertens ND, Schneider R, et al. Copy number variation analysis in 138 families with steroid-resistant nephrotic syndrome identifies causal homozygous deletions in PLCE1 and NPHS2 in two families. *Pediatr Nephrol*. 2024;39:455–461. https://doi.org/10.1007/s00467-023-06134-2
34. Halawi AA, Burgmaier K, Buescher AK, et al. Clinical characteristics and courses of patients with autosomal recessive polycystic kidney disease-mimicking phenocopies. *Kidney Int Rep*. 2023;8:1449–1454. https://doi.org/10.1016/j.ekir.2023.04.006
35. Gee HY, Otto EA, Hurd TW, et al. Whole-exome resequencing distinguishes cystic kidney diseases from phenocopies in renal ciliopathies. *Kidney Int*. 2014;85:880–887. https://doi.org/10.1038/ki.2013.450
36. Claus LR, Lekkerkerker I, van der Zwaag B, et al. *GeNepher data- and biobank for patients with (suspected) genetic kidney disease: rationale, design and status update*. 2. Rare; 2024:100030. https://doi.org/10.1016/J.RARE.2024.100030
37. Fokkema IFAC, Taschner PEM, Schaafsma GCP, Celli J, Laros JFJ, den Dunnen JT. LOVD v.2.0: the next generation in gene variant databases. *Hum Mutat*. 2011;32:557–563. https://doi.org/10.1002/humu.21438