Title: Diagnostic Utility of Exome Sequencing for Polycystic Kidney Disease

Alexander R. Chang*[‡1,2], Bryn Moore*[3], Jonathan Z. Luo[3], Gino Sartori[4], Brian Fang[1], Steven Jacobs[2], Yoosif Abdalla[2], Mohammed Taher[1], Regeneron Genetics Center[5], William Triffo[4], Gurmukteshwar Singh[1,2], Tooraj Mirshahi[3‡]

[1]Kidney Health Research Institute, [2]Department of Nephrology, [3]Department of Molecular and Functional Genomics, [4]Department of Radiology, Geisinger, Danville, PA

[5]Regeneron Genetics Center, Sawmill Road, Tarrytown, NY

*Co-first author

[‡]Correspondence:

Tooraj Mirshahi, Ph.D.
E-mail: tmirshahi@geisinger.edu

Department of Molecular and Functional Genomics
Geisinger Clinic
100 North Academy Avenue
Danville, PA 17822-2621
Phone: (570) 271-5967

Alexander R. Chang M.D.
e-mail: achange@geisinger.edu

Department of Nephrology
Kidney Health Research Institute
Geisinger Clinic
100 North Academy Avenue
Danville, PA 17822-2621
Phone: (614)271-6393

Running Title: Whole exome sequencing for PKD

Abstract

Background: Large clinical cohorts with electronic health records (EHR) and genetic data allow estimating population prevalence of rare genetic disorders like polycystic kidney disease (PKD) and deciphering their genetics, allowing more precise diagnosis and management.

Methods: We performed genotype- and phenotype-based analyses to gain insight into the prevalence and genetic basis of PKD in an unselected health system-based cohort of 173,954 patients with EHR and exome sequencing data. We determined the diagnostic rate of PKD amongst patients with PKD1/2 variants, reviewed EHR data including imaging and family history when available to phenotype PKD cases of various severity, determined disease prevalence, and identified the genetic basis for most of these cases.

Results: In genotype-first analyses, individuals with a protein-truncating (PTV) or copy number variant deletion (CNV) in PKD1 and PKD2 had ADPKD rates of 94.3% (66/70) and 97.7% (43/44) respectively. For individuals with missense or in-frame deletion variants previously classified as likely pathogenic (LP) or hypomorphic, ADPKD diagnosis was infrequent (34/425 for PKD1 and 0 /12 for PKD2). Phenotype-first analyses identified 235 ADPKD cases (1.35/1,000), which were further subclassified by imaging findings (196 moderate/severe typical ADPKD, 16 mild typical ADPKD, 23 atypical ADPKD). The genetic diagnostic rate for moderate/severe typical ADPKD was 69% (135/196) and significantly lower for mild typical and atypical ADPKD (25.0% and 21.7% respectively). Specifically, for typical ADPKD, 125/196 (63.7%) had a diagnostic (CNV/PTV/LP) variant in PKD1 or PKD2, two had PKD1 missense variants previously classified as VUS, and 24 had novel PKD1 or PKD2 missense variants, 8 of which segregated with ADPKD in pedigree analysis. Genetic diagnostic yield was much higher in those with a family history of ADPKD than those without a family history.

Conclusions: Data from a large carefully phenotyped clinical cohort provides the population-based prevalence of PKD and establishes the utility of exome sequencing **in** ADPKD diagnosis which should lead to earlier diagnosis and improved care for PKD.

Introduction

Autosomal dominant polycystic kidney disease (ADPKD) is the most common genetic cause of kidney failure. Prevalence estimates range from 3.96 per 10,000 in a study examining clinically-determined cases in the European Union(1) to 1 per 1000 in large cohorts with whole genome sequencing(2). The most common genetic variants that cause ADPKD are in *PKD1* (77%) and *PKD2* (15%), which encode polycystin-1 and polycystin-2, respectively(3). Milder cases of polycystic kidney disease (PKD) without family history are often observed in clinical practice and are not well-represented in ADPKD research cohorts. As studies of other Mendelian diseases have demonstrated, the population penetrance and phenotypic severity of ADPKD may be lower than those described in disease-specific cohorts, which tend to be overrepresented by patients with more severe presentations (4-6). An analysis of the Toronto Genetic Epidemiology of Polycystic Kidney Disease study has revealed that roughly one half of participants without family history of ADPKD do not have a detectable *PKD1* or *PKD2* variant(7). It is possible some of these unexplained cases may be due to variants in other genes (e.g., *ALG9, GANAB, DNAJB11, HNF1B) (8-10).* However, the penetrance and expressivity of variants in these atypical PKD genes remains largely unexplored.

Exome sequencing and next generation sequencing (NGS) allow for exploration of all the genetic variants in a population to better understand penetrance and expressivity for rare diseases. For a select group of genes where prior knowledge of genetic variants can lead to preventive clinical action, a genotype-based approach can be used to inform participants of their genetic findings(11). The American College of Medical Genetics (ACMG) has developed guidelines for identifying actionable genetic variants and reporting them back to providers and participants(12, 13). Despite recognition of high penetrance of pathogenic variants in *PKD1 and PKD2*, especially truncating *PKD1* variants(14, 15), a primarily genotype-based approach for early ADPKD has not yet been implemented.

Early recognition of ADPKD may be important as an approved treatment for ADPKD exists(16), and there are several ongoing clinical trials testing new therapeutic options(17-19). Currently, diagnosis of ADPKD for most at-risk patients with family history of ADPKD is done by abdominal imaging(20), and genetic testing is generally only performed when ADPKD is suspected in persons without family history or with equivocal kidney imaging findings, in the setting of related kidney donation, or for family planning purposes(21).

NGS panels and exome sequencing services are increasingly available and affordable for kidney disease and other medical conditions. As a result, many incidental findings are expected to occur in individuals unaware of renal cystic disease. There is an urgent need to better understand the significance of protein-truncating variants (PTVs) in atypical cystic disease genes and clarify the clinical significance of missense variants in *PKD1* and *PKD2* to improve counseling and guide management decisions. However, genetic testing for *PKD1* can be complicated due to its large size, high guanine, and cytosine (GC) content, and the presence of 6 pseudogenes encompassing exons 1-33(22).

In this study, we examine the utility of WES for deciphering monogenic causes of ADPKD in a large health-system based cohort including many patients without family history. We use a genotype-based approach to evaluate the prevalence of ADPKD and kidney and liver cyst diagnoses in PTV carriers of *PKD1*, *PKD2, ALG8, ALG9, DNAJB11, GANAB, HNF1B, IFT40, LRP5, PKHD1, PRKCSH, SEC61B,* and *SEC63.* In parallel, we identified clinically-confirmed ADPKD cases in our largely unselected, health system-based research cohort gaining insight into true population prevalence and the genetic causes of ADPKD in these patients.

Results

Genotype-first analyses for pLOFs in *PKD1* and *PKD2*

From exome sequencing data among 173,954 MyCode cohort participants, we identified 3,511 variants in *PKD1*. Allele frequencies for *PKD1* variants ranges from 0.19 to 2.88×10^{-6} (1,415 variants in single patients). Among the variants, 1,919 are also in gnomAD exome sequencing and an additional 344 are in gnomAD genome sequencing. ClinVar interpretation is available for 460 of the variants we identified in our cohort. Using American College of Medical Genetics (ACMG) criteria as assessed by VarSome (www.varsome.com), *PKD1* variants were classified as 58 benign, 667 likely benign (LB), 2,265 Variant of Uncertain Significance (VUS), 384 Likely Pathogenic (LP), and 137 as Pathogenic. For *PKD2,* we found 433 variants (AF range $0.36\text{-}2.88 \times 10^{-6}$) with 177 variants present in single patients. Among the *PKD2* variants, 216 are in gnomAD exome sequencing and 47 are in gnomAD genome sequencing cohort. Variants were classified using ACMG criteria per Varsome as 7 benign, 39 LB, 339 VUS, 19 LP, and 29 pathogenic. ClinVar had pathogenic or LP with two-star review status classification for only 16 *PKD1* and 7 *PKD2* found variants in our cohort.

We identified 135 individuals with protein truncating variants (PTVs) in *PKD1* or *PKD2*, which include early terminations, frameshift, and splice site variants, as well as large deletions (CNVs) that encompassed *PKD1* or *PKD2* (**Figure 1A**). We used two diagnostic approaches, first relying solely on "polycystic kidney" ICD codes in the EHR by using ICD9 codes 753.12 (Polycystic kidney, unspecified type), and 753.13 (Polycystic kidney, autosomal dominant) as well as ICD-10 codes Q61.2 (Polycystic kidney, adult type) and Q61.3 (Polycystic kidney, unspecified). In a second more in depth analysis, expert chart review by two nephrologists and reviews of available imaging by two radiologists, blinded to genotype, were performed to diagnose ADPKD. Comparing the ICD code-only analysis to chart review, we excluded 21 participants with ICD codes who lacked sufficient clinical or imaging information to make a determination about ADPKD of whom 11 were under the age of 40. Among these 21 patients, 16 had *PKD1* PTVs, 4 *PKD2* PTVs, and 1 had a *PKD2* CNV.

ADPKD diagnosis among *PKD1* PTV and CNV

When using ICD code-alone diagnoses, 3/3 (100%) patients with large genomic deletions in *PKD1* had ADPKD and 63/83 (75.9%) with *PKD1* PTVs had ADPKD. Among 67 patients with chart and imaging review confirmed diagnosis and a *PKD1* PTV variant, 63 had ADPKD (94%). Therefore, 4 individuals with *PKD1* PTVs had imaging sufficient to rule out ADPKD and did not have a family history of ADPKD. Among the 66 with confirmed ADPKD and radiologist imaging review, 100% had typical ADPKD on imaging, with moderate-severe replacement of kidney tissue with cysts in a bilateral and diffuse distribution. Overall, from 70 patients with a *PKD1* PTV or CNV, 66 (94.3%) had radiologically confirmed ADPKD indicating approximately 6% false positive rate for ADPKD for *PKD1* truncating, or large genomic deletion variants detected from exome sequencing.

ADPKD diagnosis among *PKD2* PTV and CNV

For *PKD2*, 7 out of 10 patients with large genomics deletions, and 30/39 patients with PTVs had ADPKD ICD codes **(Figure 1A)**. After expert chart review, we excluded 5 individuals, 4 with *PKD2* PTVs and 1 with *PKD2* CNV, all of whom were under the age of 40 and lacked sufficient clinical or imaging information to make a determination about ADPKD. Furthermore, 5 *PKD2* PTV and 1 *PKD2* CNV deletion carriers that did not have ADPKD ICD codes were diagnosed as ADPKD upon chart and imaging review**.** Among patients with large genomic deletions in *PKD2,* 8 of 9 (88.9%) had ADPKD and among those with *PKD2* PTVs 35 of 35 (100%) had ADPKD. Overall, from 44 patients with a *PKD2* PTV or CNV, 43 (97.7%) had radiologically confirmed ADPKD indicating an almost perfect diagnostic rate for ADPKD among *PKD2* truncating and large genomic deletion variants carriers identified from exome sequencing. *PKD2* variant carriers showed a wider spectrum of ADPKD severity with 74% showing typical ADPKD, 17% with mild PKD, and 9% with atypical PKD with unilateral atrophy (see Methods for detailed description of each PKD type).

ADPKD diagnosis among carriers of known *PKD1* and *PKD2* missense variants

The Mayo clinic maintains PKDB (http://pkdb.pkdcure.org), a large repository of variants in *PKD1* and *PKD2* that is curated to include information such as variant type, clinical significance, per ACMG classification, and information on the number of pedigrees and publications associated with each variant. We examined the presence of rare missense variants in *PKD1* or *PKD2* in our cohort that had also been reported in the PKDB (**Figure 1B**). For those with in-frame deletion in *PKD1*, 6/6 (100%) had a diagnosis of ADPKD per ICD code and chart and imaging review. Among those with *PKD1* missense variants designated as LP in PKDB, 22/98 (22.4%) had an ADPKD diagnosis. We found no in-frame deletions in *PKD2* in our cohort, and the one participant with a *PKD2* LP missense variant listed in PKDB, did not have an ADPKD diagnosis. After expert chart review, we removed 21 patients who had insufficient data to diagnose ADPKD. Among the remaining patients 24/77 (31.2%) with a *PKD1* LP variant, 6/6 with in-frame deletions had ADPKD, and the one patient with a *PKD2* LP variant had ADPKD.

Genotype-first analyses for *ALG8, ALG9, DNAJB11, GANAB, HNF1B, IFT40, LRP5, PKHD, PRKCSH, SEC61B, SEC63*

While ADPKD has traditionally been interpreted based on genetic variation in *PKD1* and *PKD2*, several more genes have recently been associated with development of kidney and liver cysts, and possibly ADPKD (*10, 23*). We assessed exome sequencing data for PTVs and CNVs among the following genes: *ALG8, ALG9, DNAJB11, GANAB, HNF1B, IFT40, LRP5, PKHD, PRKCSH, SEC61B, SEC63* and examined the association of PTVs in these genes with kidney/liver cysts. Based on our and others' prior studies examining *ALG9* and *DNAJB11(8, 23)*, we anticipated that ADPKD and PLD diagnoses would be rare and therefore used a broader composite outcome of any kidney/liver cyst ICD code, which included ADPKD, cystic kidney disease, acquired renal cyst, and liver cyst diagnoses. There were 3,793 patients with rare PTVs in these 11 genes (243 *ALG8*, 30 *ALG9*, 11

DNAJB11, 22 *GANAB*, 7 *HNF1B*, 300 *IFT40*, 2187 *LRP5*, 754 *PKHD1*, 102 *PRKCSH*, 4 *SEC61B*, 44 *SEC63*). One *ALG8* PTV carrier, 1 *GANAB* PTV carrier and 4 *PKHD1* PTV carriers had ADPKD ICD diagnosis in the absence of a diagnostic *PKD1* or *PKD2* variant. Additionally, there were 61 CNV carriers in these genes (2 *ALG8*, 1 *ALG9*, 26 *HNF1B*, 21 *IFT40*, 2 *LRP5*, 3 *PKHD1*, 1 *SEC61B*, 6 *SEC63*). **Table 1** shows data comparing prevalence of PKD diagnosis and any kidney/liver cyst diagnosis among PTV and CNV deletion carriers for these genes as well as *PKD1* and *PKD2*. There were too few *SEC61B* PTV carriers for analysis. Most of the *LRP5* PTV's are from a splice site variant in the last exon which may not cause loss of function. For this analysis we removed first and second-degree related individuals to avoid false positive. Having a PTV or CNV deletion in *PKD1* or *PKD2* was strongly associated with the PKD diagnosis as well as kidney and liver cystic phenotypes as expected and consistent with the detail from **Figure 1**. For PTV or CNV deletion in the other genes, *GANAB*, *HNF1B* and *PKHD1* were significantly associated with PKD. PTV or CNV deletion in *ALG9, ALG8, GANAB, HNF1B, and SEC63* were significantly associated with for the composite outcome of any kidney or liver cyst ICD diagnosis. Only two *17q12* deletion carriers (that encompasses *HNF1B*) and two *PKHD1* CNV deletion carriers had ADPKD diagnoses in their EHR.

Overall, these data show that a genotype-first approach using high quality exome sequencing and well curated EHR to determine susceptibility to ADPKD is feasible with high positive predictive value. Our data from a large clinical cohort also show that PTV variants in other genes that have been associated with development of ADPKD, other than *PKD1* and *PKD2*, are much less penetrant and associated with milder cystic phenotypes.

Prevalence and phenotypic landscape of PKD

Of the 173,954 individuals with WES data, there were a total of 303 participants with at least 1 ADPKD ICD code in their EHR. Closer inspection of the EHR showed that 42 of the 303 either did not have enough data or were negative for ADPKD, so it is not clear why they received ADPKD ICD

diagnoses **(Figure 2A)**. We chart reviewed the remaining 261 patients and found 235 with ADPKD of some degree, with an additional 26 who had phenotypes more closely aligned with other kidney phenotypes (6 with suspected ARPKD, 2 with tuberous sclerosis, 2 with CAKUT), 4 who had only ADPKD family history but no personal history of ADPKD, and 10 who had kidney cysts, but not enough for an ADPKD diagnosis per expert review (**Figure 2B)**. We broke down the 235 with ADPKD and sufficient data, into 3 groups: 196 Typical ADPKD, 23 atypical ADPKD, 16 mild-ADPKD (**Figure 2A**, see methods for description). **Supplemental Table 1** provides specific information for the 16 patients in the other kidney diseases category, including, demographics, EHR data, genetic information and prevalence and pathogenicity assertion from previous cohorts or public data where available.

***PKD1*/2 PTVs and CNVs account for most of ADPKD cases**

Of the 196 patients with typical ADPKD, 99 had PTVs or CNVs in *PKD1* or *PKD2*. Specifically, 69 had a *PKD1* PTV [large deletion, stop gain, frame shift, splices site variant], 2 had chromosomal deletions encompassing *PKD1*, 23 had a *PKD2* PTV, and 5 had large chromosomal deletions encompassing *PKD2* (**Figure 3A**). Using genetically determined relatedness data we built pedigrees enabling genotype-phenotype segregation analyses for cases who had relatives in the cohort. We performed pedigree analysis for most of the cases with typical ADPKD and a PTV/CNV for *PKD1* or *PKD2*. **Figure 3B** is an example of pedigree analysis for carriers of the protein truncating variant *PKD1_Glu1061Ter*, where the variant perfectly segregates with ADPKD diagnosis. Kidney CT scan images are shown for the 4 variant carriers (**Figure 3C**).

Identifying known missense *PKD1*/*PKD2* variants in typical ADPKD

Utilizing PKDB, we identified carriers of *PKD1* and *PKD2* inframe deletion or missense variants classified as likely pathogenic (LP). We found 26 participants with a previously reported LP missense

or in-frame deletion in *PKD1* (**Figure 4A**), and none for *PKD2*. Pedigree analysis of a family with the likely pathogenic *PKD1_Arg3719Gln* variant shows segregation with ADPKD diagnosis (**Figure 4B**). Polycystic kidneys are observed in both young and older carriers of *PKD1_Arg3719Gln* and polycystic liver disease can be seen in the older carrier (**Figure 4C**).

Using family and cohort data to classify VUS and novel missense variants

We identified 25 ADPKD patients who had *PKD1* or *PKD2* variants that were reported as VUS or likely benign in PKDB or were not reported in that database (**Figure 5A**). The list of these patients and their variants in shown in **Supplemental Table 2**. For 9 of these patients, we found sufficient clinical and genetic data, from EHR and other clinical genetic testing, that corroborated presence of ADPKD. The phenotype and additional genetic data strongly suggest these variants are likely pathogenic. For instance, we identified a carrier of *PKD1_Ala2511Pro* who had a diagnosis of ADPKD who had several family members in our electronic health records that were not part of the MyCode exome sequencing project (**Figure 5B, C**). Among the family members, 5 had a diagnosis of ADPKD, and 2 had prior clinical genetic testing identifying the same variant as the proband in the exome sequencing cohort.

We also found that several ADPKD patients had novel *PKD1* and *PKD2* variants, not reported in PKDB. We noticed that several of these patients carried the same variant (e.g., *PKD1_Gly2034Val*, *PKD1_Asp3304Asn*, *PKD1_Ser3591Phe*, *PKD1_Cys51Tyr* and *PKD2_Arg320Leu*), so we investigated all carriers of those variants for ADPKD. We identified 4 carriers of *PKD1_Gly2034Val*; 3 of whom had ADPKD (2 with ICD diagnosis, 1 on additional imaging review), and 1 had insufficient clinical information as she was less than 25 years of age. Pedigrees and abdominal imaging are shown for two of the *PKD1_Gly2034Val* patients (**Figure 5D, E**). We identified 11 carriers of *PKD1_Asp3304Asn* in 2 families; 7 had imaging data, of which 6 had ADPKD (4 had PKD ICD code in EHR, 2 on additional imaging review). Two pedigrees and associated

imaging data are shown (**Figure 5 F, G**). All 3 carriers of the *PKD1_Ser3591Phe* in two families had ADPKD, and none of the unaffected family members of these carriers had ADPKD (**Figure 5 H, I**).

Further detail for the rest of VUS, novel or LB from PKDB are in **Supplemental Table 2**. Of 2 unrelated individuals with the *PKD1_Cys51Tyr* variant, 1 had ADPKD and the other individual had congenital solitary kidney with multiple cysts. The two carriers of *PKD2_Arg320Leu* are unrelated and had no known relatives, so pedigrees analysis was not possible. Both *PKD1_Cys51Tyr* and *PKD2_Arg320Leu* are classified as VUS with minor pathogenic evidence per ACMG criteria in VarSome (https://doi.org/10.1093/bioinformatics/bty897). Our data suggest that these VUS need to be classified as likely pathogenic.

Atypical PKD genes in those diagnosed with ADPKD

Finally, 46 individuals with ADPKD did not have a rare variant in *PKD1* or *PKD2* (**Figure 6A**). We therefore searched for rare variants in the "atypical" ADPKD genes *ALG8, ALG9, DNAJB11, GANAB, HNF1B, IFT40, LRP5, PKHD1, PRKCSH, SEC63,* and *SEC61B.* We identified 9 patients with rare variants in any of these genes, of whom 4 had a missense VUS in *ALG9*, *HNF1B*, or *PKHD1*, one had a previously unreported *GANAB* variant and 4 had missense variants in *IFT40* (**Figure 6B**). *GANAB_Asp647Val* segregates with ADPKD phenotype in one family and is present in an affected individual in another family with two unaffected non-carriers (**Figure 6C, D**). In addition to this *GANAB Asp647Val* carrier with typical ADPKD, there were 7 other carriers of *GANAB Asp647Val,* 4 had mild or atypical ADPKD, 1 had multiple hypodensities in kidneys and liver, 1 had bilateral kidney cysts before age 20, and 1 had no imaging available for review. CT images for 4 carriers of *IFT40* missense variants are shown in **Figure 6E**. Whether variants in other atypical ADPKD genes identified in this group are the genetic cause of ADPKD remains undetermined. Overall, among 235 patients with confirmed typical, atypical, and mild ADPKD, 69 (29.3%) had no identified candidate variant in the 12 cystic genes examined to explain their ADPKD diagnoses.

Mild and Atypical ADPKD

In addition to the 196 participants with typical ADPKD, we found 16 participants with mild ADPKD, and 23 with atypical ADPKD (**Figure 7A)**, Definitions for mild and atypical ADPKD are based on specific clinical and imaging criteria (details provided in the methods). Breakdown of genetic variants among these patients is shown in **Figure 7B**. Four subjects with mild or atypical ADPKD had PTVs in *PKD2*, 1 had a LP variant in *PKD1*, 1 had a VUS variant in *PKD1*, 1 had the *PKD1_ Gly2034Val* variant described in **Figure 5**, and 4 had unreported *PKD1* variants. Three more carriers of the *GANAB_Asp647Val* were identified with mild or atypical ADPKD in two unrelated families (see **Figure 6C, D**). Out of 39 participants with mild or atypical ADPKD 18 (46%) did not have a candidate variant identified.

ARPKD, 17q12 syndrome, tuberous sclerosis, and CAKUT participants with ICD codes for PKD

Among the 6 patients with clinical phenotype consistent with ARPKD, there were 2 related patients who had a large deletion in *PKHD1* along with *PKHD1*_Gln3407Ter, as expected with the recessive inheritance of ARPKD. One patient was heterozygous for a *PKHD1* stop gain (Arg496Ter), 1 was heterozygous for a *PKHD1* missense variant classified as pathogenic in ClinVar, and 2 had no protein-changing variants in the genes examined. Two patients had 17q12 syndrome, a large chromosomal deletion encompassing *HNF1B*, and exhibited the clinical phenotype renal cysts and diabetes syndrome (**Supplemental Table 1**). The 2 participants with a tuberous sclerosis clinical phenotype had *TSC2* variants detected. Among the 3 participants with CAKUT, 1 had a large deletion in *PKD2*, 1 had 2 rare missense variants in *FRAS1* and a rare missense variant in *SALL1*, and 1 had a pathogenic variant in *PKHD1*.

Discussion

Utilizing both a phenotype-based and genotype-based approach in a large, unselected patient population, we show that exome sequencing has high diagnostic accuracy for ADPKD. Our findings from the genotype-first approach show a high predictive value for exome sequencing for ADPKD in carriers of pLOFs (PTV/CNV deletions) in *PKD2* and *PKD1*. We see almost perfect diagnostic accuracy among *PKD2* PTV/CNV deletion as well as *PKD1* CNV deletion carriers, and only 6% false positive among *PKD1* PTV individuals. Since CLIA certified Sanger sequencing is almost always used in clinic prior to taking clinical action, the ~6% false positive *PKD1* rate can be easily excluded by sequencing. Therefore >95% of those with PKD1 or PKD2 PTV or CNV identified in exome sequencing will receive a correct prediction from exome sequencing with significant clinical value.

The size of our cohort allowed us to examine atypical PKD genes and whether we can extract clinically valuable information from identifying PTVs among these genes. We find that carriers of atypical PKD genes have rare diagnoses of ADPKD but increased prevalence of a composite of any kidney/liver cyst diagnoses, reflecting the lower severity of disease caused by PTVs in these genes.

Interestingly, very few individuals who had missense variants in *PKD1* or *PKD2* previously described as likely pathogenic had ADPKD diagnoses, and therefore the potential for false positives exists among carriers of that class of variants. This discordance may be reflective of a lower penetrance among these types of variants. Alternatively, interpretation from older studies identifying these missense variants as the causal variant for ADPKD may be incorrect since those were likely reported from isolated cases with limited genetic data. Data from increasingly larger clinic cohorts, such as ours, will continue to address this discordance and help identify and verify causal *PKD1* and *PKD2* missense variants.

Our data clearly show that exome sequencing is a great starting point for population-based identification of the vast majority of ADPKD patients, many of whom may be undiagnosed or unaware of their susceptibility to developing ADPKD. A prior study found limited utility in exome sequencing *PKD1,* showing only 50% sensitivity due to issues with low read depth and sequencing quality in

exons 1-33, which are also present in *PKD1P1-PKD1P6* pseudogenes (22). In our cohort, we had excellent read depth and sequencing quality throughout most of the *PKD1* gene. Out of 58 *PKD1* PTVs that we identified, only 2 variants were misassigned to a pseudogene (see **supplemental Figure1 A and B**), for 1 variant our splice prediction didn't agree with spliceAI (https://spliceailookup.broadinstitute.org/) and 2 variants had either poor sequencing quality or occurred in a DNA region with a string of Gs, which was difficult to interpret. Therefore 53/58 variants were correctly assigned to *PKD1*. Therefore, good chromosome capture and high-quality deep sequencing, augmented with manual sequence curation as needed for some variants, maximizes the utility of exome sequencing in identifying causal *PKD1* variants.

We also show that the phenotypic spectrum of *PKD1* and *PKD2* may be wider than previously appreciated as the prevalence of ADPKD was very low for those with likely hypomorphic and likely pathogenic missense variants in *PKD1* and *PKD2*. This was also noted in a study examining the Exome Aggregation Consortium (ExAC), where there was 6.9-fold higher than expected rate of previously reported disease-causing variants in *PKD1* and *PKD2* (25). Thus, additional work is needed to assess pathogenicity of many missense variants in *PKD1* and *PKD2*.

In our cohort of 196 typical ADPKD participants, 118 (60.2%) of participants had self-reported family history of ADPKD and 78 (39.8%) had no self-reported family history. In the positive family history group, 110/118 (93.2%) had a variant of some kind in *PKD1* or *PKD2*, with 96 being CNV/PTV or reported LP variants and 7 that co-segregated within families with the ADPKD phenotype. One patient had *GANAB_As647Val* that segregated with typical ADPKD. Only 40/78 (51.2%) with no family history had a *PKD1/PKD2* variant. Among these 40 patients, 29 had a CNV/PTV or reported LP variants and 1 had a variant that co-segregated in a family with ADPKD. The lower diagnostic yield for PKD with no family history was similar to what was seen in an analysis of Toronto Genetic Epidemiology Study of Polycystic Kidney Disease (26).

In addition, we demonstrate that several, but not all, putative atypical ADPKD/PLD genes are associated with increased risk of cystic kidney/liver disease. The overall prevalence of diagnosed ADPKD was very low, and zero for many atypical ADPKD/PLD genes. In analyses using expanded kidney or liver cystic ICD codes, we found that *ALG8, ALG9, DNAJB11, GANAB, and SEC63* carriers were at increased risk, compared to non-carriers. However, neither *LRP5* nor *PRKCSH* were associated with increased risk of kidney/liver cysts. The implications of mild cystic kidney disease are unclear at this time and our study was not designed to thoroughly describe prevalence of cystic kidney and liver disease that was not deemed severe enough to justify an ICD diagnosis by a clinician. For example, while no patients with an *ALG9* PTV had a PKD ICD code, we showed that 88% of patients above the age of 50 had at least 4 cysts on imaging (8). Future studies are needed to better understand the significance of mild cystic disease and whether cystic disease caused by atypical genes are responsive to therapies to slow ADPKD progression such as vasopressin receptor 2 antagonism (16).

Most individuals with ADPKD have not undergone genetic testing in real-world clinical practice. However, new clinical genetic testing panels are becoming more common place in nephrology practice, many of them using exome capture technology to report exonic variants in tens or hundreds of genes. Our data demonstrate that most of these approaches, as well as sequencing of the entire exome in a clinic setting like ours, will have good positive predictive value to identify causal variants in *PKD1 and PKD2*, if followed by additional bioinformatic steps and Sanger sequencing in some cases. In addition to an increase in exome sequencing services for medical genetic testing and research purposes, there has been an increase in direct-to-consumer genetic testing with some including exome sequencing in their packages and with some offering return of results for certain actionable conditions (27). Careful curation of any of these types of exome sequencing data should facilitate early identification of at-risk individual for ADPKD.

Acknowledgments: This work was supported by NIDDK106515 to A.C and GM111913 to T.M. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability: The data supporting the findings of this study are available within the article and its Supplementary Data files. Additional information for reproducing the results described in the article is available upon reasonable request and subject to a data use agreement.

Author Contributions: Conceptualization: A.C, B.S.M., G.S., T.M.; Data curation: A.C., B.S.M, J.Z.L., B.F., S.J., Y.A., M.T., W.J.T., G.S. R.G.C.; Formal Analysis: A.C., B.S.M., J.Z.L; Funding acquisition: A.C., T.M.; Writing – original draft: A.C., B.S.M.; Writing – review & editing: A.C., B.S.M., J.Z.L., T.M.

Competing Interest: Authors declare no conflict of interest.

Ethics Declaration: This research was approved by the Geisinger Clinic Institutional Review Board and included 173,954 participants in the MyCode Health Initiative who have exome sequencing data obtained as part of the Geisinger-Regeneron DiscovEHR collaboration. All participants provided written informed consent, and all experiments were performed in accordance with relevant guidelines and regulations. The authors did not have access to any identifying information for the participants. The human phenotype and genotype data in this study were all deidentified by a “data broker” who was not involved in the study before any analysis was performed. De-identified clinical data was obtained from electronic health records (EHR).

References

1. Willey CJ, Blais JD, Hall AK, Krasa HB, Makin AJ, Czerwiec FS. Prevalence of autosomal dominant polycystic kidney disease in the European Union. Nephrol Dial Transplant. 2017;32(8):1356-63.
2. Lanktree MB, Haghighi A, Guiard E, Iliuta IA, Song X, Harris PC, et al. Prevalence Estimates of Polycystic Kidney and Liver Disease by Population Sequencing. J Am Soc Nephrol. 2018;29(10):2593-600.
3. Bergmann C, Guay-Woodford LM, Harris PC, Horie S, Peters DJM, Torres VE. Polycystic kidney disease. Nat Rev Dis Primers. 2018;4(1):50.
4. Manickam K, Buchanan AH, Schwartz MLB, Hallquist MLG, Williams JL, Rahm AK, et al. Exome Sequencing-Based Screening for BRCA1/2 Expected Pathogenic Variants Among Adult Biobank Participants. JAMA Netw Open. 2018;1(5):e182140.
5. Carruth ED, Beer D, Alsaid A, Schwartz MLB, McMinn M, Kelly MA, et al. Clinical Findings and Diagnostic Yield of Arrhythmogenic Cardiomyopathy Through Genomic Screening of Pathogenic or Likely Pathogenic Desmosome Gene Variants. Circ Genom Precis Med. 2021;14(2):e003302.
6. Adler A, Novelli V, Amin AS, Abiusi E, Care M, Nannenberg EA, et al. An International, Multicentered, Evidence-Based Reappraisal of Genes Reported to Cause Congenital Long QT Syndrome. Circulation. 2020;141(6):418-28.
7. Iliuta IA, Kalatharan V, Wang K, Cornec-Le Gall E, Conklin J, Pourafkari M, et al. Polycystic Kidney Disease without an Apparent Family History. J Am Soc Nephrol. 2017;28(9):2768-76.
8. Besse W, Chang AR, Luo JZ, Triffo WJ, Moore BS, Gulati A, et al. ALG9 Mutation Carriers Develop Kidney and Liver Cysts. J Am Soc Nephrol. 2019;30(11):2091-102.
9. Besse W, Dong K, Choi J, Punia S, Fedeles SV, Choi M, et al. Isolated polycystic liver disease genes define effectors of polycystin-1 function. J Clin Invest. 2017;127(9):3558.
10. Cornec-Le Gall E, Torres VE, Harris PC. Genetic Complexity of Autosomal Dominant Polycystic Kidney and Liver Diseases. J Am Soc Nephrol. 2018;29(1):13-23.
11. Schwartz MLB, McCormick CZ, Lazzeri AL, Lindbuchler DM, Hallquist MLG, Manickam K, et al. A Model for Genome-First Care: Returning Secondary Genomic Findings to Participants and Their Healthcare Providers in a Large Research Cohort. Am J Hum Genet. 2018;103(3):328-37.
12. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med. 2013;15(7):565-74.
13. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med. 2017;19(2):249-55.
14. Hwang YH, Conklin J, Chan W, Roslin NM, Liu J, He N, et al. Refining Genotype-Phenotype Correlation in Autosomal Dominant Polycystic Kidney Disease. J Am Soc Nephrol. 2016;27(6):1861-8.
15. Cornec-Le Gall E, Audrézet MP, Rousseau A, Hourmant M, Renaudineau E, Charasse C, et al. The PROPKD Score: A New Algorithm to Predict Renal Survival in Autosomal Dominant Polycystic Kidney Disease. J Am Soc Nephrol. 2016;27(3):942-51.
16. Torres VE, Chapman AB, Devuyst O, Gansevoort RT, Grantham JJ, Higashihara E, et al. Tolvaptan in patients with autosomal dominant polycystic kidney disease. N Engl J Med. 2012;367(25):2407-18.
17. Capuano I, Buonanno P, Riccio E, Amicone M, Pisani A. Therapeutic advances in ADPKD: the future awaits. J Nephrol. 2021.
18. Pirson Y. Extrarenal manifestations of autosomal dominant polycystic kidney disease. Adv Chronic Kidney Dis. 2010;17(2):173-80.
19. Flahault A, Joly D. Screening for Intracranial Aneurysms in Patients with Autosomal Dominant Polycystic Kidney Disease. Clin J Am Soc Nephrol. 2019;14(8):1242-4.

20. Pei Y, Obaji J, Dupuis A, Paterson AD, Magistroni R, Dicks E, et al. Unified criteria for ultrasonographic diagnosis of ADPKD. Journal of the American Society of Nephrology : JASN. 2009;20(1):205-12.
21. Lanktree MB, Iliuta IA, Haghighi A, Song X, Pei Y. Evolving role of genetic testing for the clinical management of autosomal dominant polycystic kidney disease. Nephrol Dial Transplant. 2019;34(9):1453-60.
22. Ali H, Al-Mulla F, Hussain N, Naim M, Asbeutah AM, AlSahow A, et al. PKD1 Duplicated regions limit clinical Utility of Whole Exome Sequencing for Genetic Diagnosis of Autosomal Dominant Polycystic Kidney Disease. Sci Rep. 2019;9(1):4141.
23. Huynh VT, Audrézet MP, Sayer JA, Ong AC, Lefevre S, Le Brun V, et al. Clinical spectrum, prognosis and estimated prevalence of DNAJB11-kidney disease. Kidney Int. 2020;98(2):476-87.
24. Mallawaarachchi AC, Hort Y, Cowley MJ, McCabe MJ, Minoche A, Dinger ME, et al. Whole-genome sequencing overcomes pseudogene homology to diagnose autosomal dominant polycystic kidney disease. Eur J Hum Genet. 2016;24(11):1584-90.
25. Mallawaarachchi AC, Furlong TJ, Shine J, Harris PC, Cowley MJ. Population data improves variant interpretation in autosomal dominant polycystic kidney disease. Genet Med. 2019;21(6):1425-34.
26. Lanktree MB, Guiard E, Li W, Akbari P, Haghighi A, Iliuta IA, et al. Intrafamilial Variability of ADPKD. Kidney Int Rep. 2019;4(7):995-1003.
27. Hunter JE, Irving SA, Biesecker LG, Buchanan A, Jensen B, Lee K, et al. A standardized, evidence-based protocol to assess clinical actionability of genetic disorders associated with genomic variation. Genet Med. 2016;18(12):1258-68.
28. Carey DJ, Fetterolf SN, Davis FD, Faucett WA, Kirchner HL, Mirshahi U, et al. The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. Genet Med. 2016;18(9):906-13.
29. Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science. 2016;354(6319).
30. Packer JS, Maxwell EK, O'Dushlaine C, Lopez AE, Dewey FE, Chernomorsky R, et al. CLAMMS: a scalable algorithm for calling common and rare copy number variants from exome sequencing data. Bioinformatics. 2016;32(1):133-5.
31. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016;17(1):122.
32. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, 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(5):405-24.
33. Gout AM, Martin NC, Brown AF, Ravine D. PKDB: Polycystic Kidney Disease Mutation Database--a gene variant database for autosomal dominant polycystic kidney disease. Hum Mutat. 2007;28(7):654-9.
34. Irazabal MV, Rangel LJ, Bergstralh EJ, Osborn SL, Harmon AJ, Sundsbak JL, et al. Imaging classification of autosomal dominant polycystic kidney disease: a simple model for selecting patients for clinical trials. J Am Soc Nephrol. 2015;26(1):160-72.
35. Reynolds DM, Falk CT, Li A, King BF, Kamath PS, Huston J, 3rd, et al. Identification of a locus for autosomal dominant polycystic liver disease, on chromosome 19p13.2-13.1. American journal of human genetics. 2000;67(6):1598-604.
36. Qian Q, Li A, King BF, Kamath PS, Lager DJ, Huston J, 3rd, et al. Clinical profile of autosomal dominant polycystic liver disease. Hepatology. 2003;37(1):164-71.
37. Staples J, Qiao D, Cho MH, Silverman EK, Nickerson DA, Below JE. PRIMUS: rapid reconstruction of pedigrees from genome-wide estimates of identity by descent. Am J Hum Genet. 2014;95(5):553-64.

Methods

Study Population

The Geisinger Institutional Review Board approved this study. Informed consent was waived as participants were previously consented in the MyCode™ Community Health Initiative as part of the Geisinger-Regeneron DiscovEHR collaboration(28). For this study we included 173,954 participants from the MyCode cohort who had exome sequencing and health record data available.

Whole exome sequencing

Exome sequencing was performed in collaboration with Regeneron Genetics Center as previously described (28). A modified version of the xGEN probe from Integrated DNA Technologies (IDT) was used for target sequence capture (29). Sequencing was performed by paired end 75bp reads on either an Illumina HiSeq2500 or NovaSeq. Coverage depth was sufficient to provide more than 90% coverage of the targeted bases in for 99% of samples. Alignments and variant calling were based on GRCh38 human genome reference sequence. Copy number variants for *PKD1*, *PKD2*, and atypical ADPKD genes were estimated using lattice-aligned mixture models (CLAMMS) software(30).

Nonsense or PTVs (protein truncating variants) are defined in this study as variants that cause a start-loss, frameshift, or early termination/stop-gain of the encoded protein. Variants in *PKD1, PKD2, ALG8, ALG9, DNAJB11, GANAB, HNF1B, IFT40, LRP5, PKHD1,* and *SEC61* were annotated using the Ensembl Variant Effect Predictor with Ensembl 90 definitions (31). We included PTV variants that had GQ score >20, AD_ALT>3, AD_ALT/AD_REF ≥0.5, and MAF<0.01 in the MyCode population. Missense variants in *PKD1* and *PKD2* were included using the same criteria, but only if previously reported in the PKDB as Likely Pathogenic, or Likely Hypomorphic. We cross-referenced all variants with ClinVar interpretation and stars rating system (32). *PKD1* and *PKD2* were additionally cross-referenced with the PKDB (33) as well as Varsome (https://doi.org/10.1093/bioinformatics/bty897), a

database that considers case-control, computational, functional, and family segregation data to provide ACMG-guided classification of variants' pathogenicity(32). For patients with clinical diagnosis of ADPKD, we also cataloged any accompanying rare (MAF <0.001) missense variants in *PKD1 and PKD2* and rare pLOF or missense variants in 11 other genes that have been described in patients with cystic kidney disease (*ALG8, ALG9, DNAJB11, GANAB, HNF1B, IFT40, LRP5, PKHD1, PRKCSH, SEC61B, SEC63*).

Phenotyping using EHR

We used data from the electronic health record (EHR) to determine whether participants had been diagnosed with PKD using 9th and 10th International Classification Diseases (ICD) diagnosis codes (Q61.2, Q61.3, 753.13, 753.12). Additional diagnoses examined included other cystic kidney diseases (Q61.5, Q61.8, Q61.9, 753.10), acquired kidney cyst (N28.1, 593.2), congenital kidney cyst (Q61.00, Q61.01, Q61.02, 753.11, 753.19), or liver cystic disease (Q44.6, 573.8); a composite outcome of "any kidney/liver cyst diagnosis" included all the ICD codes.

Chart Reviews

To enhance genotype-phenotype analyses, additional chart review was performed on 1) patients with ICD diagnosis codes for PKD; 2) patients who had *PKD1* or *PKD2* variants that were PTVs, or missense variants previously described as LP or likely hypomorphic in PKDB; 3) family members of patients with chart-confirmed PKD who also had available exome sequencing data. Chart review including review of imaging was done by at least 1 nephrologist with focus on kidney and liver imaging, nephrolithiasis, cerebral aneurysms, history of dialysis, transplant, family history of ADPKD or cerebral aneurysms, and clinical genetic testing. Additionally, blinded review of imaging was done by at least 1 radiologist with additional review by a senior radiologist with expertise in abdominal imaging for questionable cases, and cases were discussed until consensus was achieved.

Participants who had a PKD ICD code but whose clinical diagnosis/phenotype was more consistent with autosomal recessive polycystic kidney disease (ARPKD), tuberous sclerosis complex (TSC), or congenital abnormalities of kidney and urinary tract (CAKUT) were examined separately from the ADPKD cohort.

Defining typical, atypical, and mild ADPKD

Imaging phenotypes were classified per the Mayo clinic imaging classification. Most cases were assigned to Class 1: typical ADPKD (bilateral and diffuse distribution, with mild, moderate, or severe replacement of kidney tissue by cysts, where all cysts contribute similarly to total kidney volume) or Class 2 atypical ADPKD (34). As there was variability in the type of imaging obtained in this clinical cohort, total kidney volume could not be determined for all patients; Class 1 ADPKD was classified as moderate/severe or mild by 2 radiologists. Individuals who only had imaging after nephrectomy or only had imaging after ESKD was reached were excluded from image phenotyping. If patients lacked imaging that was available for radiologist review, we considered them to have typical ADPKD if they had clinical history consistent with typical ADPKD (e.g. ESKD due to ADPKD, imaging report and family history consistent with Ravine-Pei criteria) (20, 35, 36).

Statistical Analyses

Genotype-based approach. First, we examined the prevalence of participants with rare PTV variants (MAF<0.01), or large deletions detected by CNV analysis in *PKD1*, *PKD2*, and any of the 11 other cystic disease-associated genes (*ALG8, ALG9, DNAJB11, GANAB, HNF1B, IFT40, LRP6, PKHD1, PRKCSH, SEC61B, SEC63B)*. We also examined the proportion of participants with rare missense mutations in *PKD1* or *PKD2* listed in the PKDB listed as likely pathogenic, or likely hypomorphic. We then evaluated whether carriers of PTVs or large deletions in each of the 11 other cystic disease-associated genes were at increased risk of ICD-diagnosed PKD, liver cystic disease, or any

kidney/liver cystic disease. Logistic regression was used to assess the association of PTVs with ADPKD determined by ICD code. The severely imbalanced nature of the rare genotypic data may result in bias in the maximum likelihood estimations. Additionally, the strong predictive power of variables such as the truncation of *PKD1* or *PKD2* for diagnosis of ADPKD could result in the problem of separation. For this reason, Firth logistic regression was used which reduces the bias by penalizing the likelihood function by Jeffrey's invariant prior. The regression was adjusted for age, sex, year of first outpatient encounter at Geisinger, and genetically-determined ancestry. First and second degree relatives were removed in this analysis to avoid false positives due to presence of related individuals. For each gene, the logistic regression coefficient represents the expected change in log-odds of the outcome occurring if a variant is present in the gene. A larger coefficient may be interpreted as a larger increase in the relative risk of the outcome.

Phenotype-based approach. After chart review was performed to confirm PKD on individuals with at least one code for PKD in the EHR, we examined exome sequencing data for PTVs and CNV deletions in *PKD1/PKD2*, reported missense variants in the PKDB, rare and novel *PKD1/PKD2* missense/inframe deletion variants, and rare variants in the 11 other putative cystic genes. Previously reported variants were then grouped based on the PKDB classifications. After identifying first- and second-degree relatives of each participant with ADPKD using PRIMUS(37) and EHR-documented family history, we reconstructed pedigrees of patients with ADPKD. We then reviewed charts of family members of ADPKD patients to evaluate evidence for co-segregation of the variants in question with disease.

Figure 1. Genotype first approach- carriers of *PKD1* or *PKD2* **A)** CNV/PTVs or **B)** Inframe deletions/missense variants classified as likely pathogenic in PKDB. 173,954 participants with whole exome sequencing were evaluated for copy number variants (CNVs) or protein truncating variants (PTVs) in *PKD1* and *PKD2*. Electronic health records (EHR) of carriers of these variants were examined for ADPKD ICD codes (Q61.2, Q61.3, 751.12, 751.13), and chart reviewed to confirm presence of ADPKD.

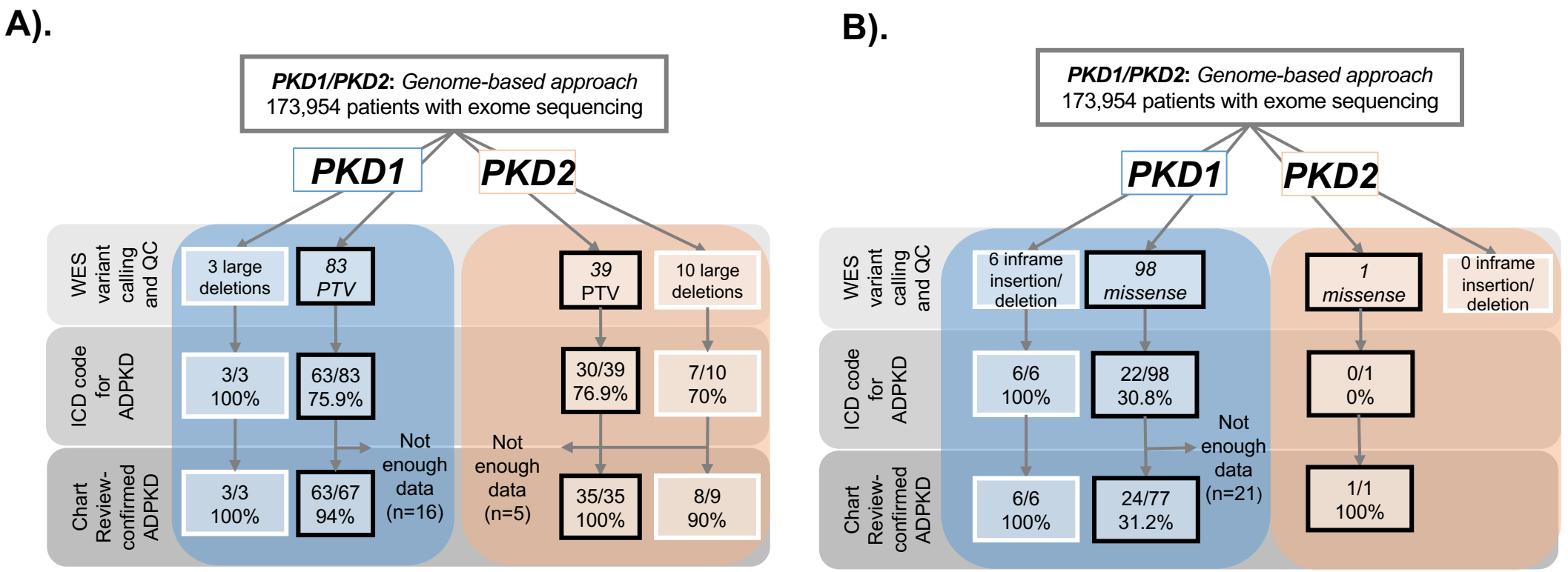


Table 1. Risk of ADPKD or kidney/liver cyst ICD codes among those with PTV or CNVs in atypical APDPKD genes *ALG8, ALG9, GANAB, HNF1B, IFT40,* and *SEC63* are increased.

Firth's logistic regression was used to assess the association of PTV and CNVs with ADPKD and kidney/liver cyst determined by ICD code. The regression was adjusted for age, sex, year of first outpatient encounter at clinic, and genetically-determined ancestry. For these analysis, first- or second-degree relatives in the cohort were removed. For each gene, the logistic regression coefficient represents the expected change in log-odds of the outcome occurring if a PTV is present in the gene. A larger coefficient may be interpreted as a larger increase in the relative risk of the outcome (see statistical methods for regression details).

		PKD Diagnosis code				Any Kidney or Liver cyst			
		N	Co-eff (95%CI)	P-Value	Chi-Sq	N	Co-eff (95%CI)	P-Value	Chi-Sq
	No PTV in Cystic Genes (N=42,485)	34	Ref	-	-	2033	Ref	-	-
Gene	ALG8 (n=159)	0	1.21 (-3.66, 3.25)	4.83E-01	0.49	17	0.95 (0.41, 1.44)	**1.06E-03**	10.73
	ALG9 (n=25)	0	3.19 (-1.67, 5.23)	1.37E-01	2.21	4	1.13 (-0.05, 2.09)	**5.91E-02**	3.56
	DNAJB11 (n=7)	0	4.5 (-0.39, 6.69)	6.30E-02	3.46	1	1.49 (-0.82, 3.15)	1.72E-01	1.86
	GANAB (n=14)	1	4.83 (2.59, 6.32)	**7.93E-04**	11.26	3	1.71 (0.32, 2.84)	**1.97E-02**	5.44
	HNF1B (n=16)	1	4.79 (2.52, 6.3)	**8.88E-04**	11.05	3	1.95 (0.54, 3.11)	**9.75E-03**	6.68
	IFT140 (n=205)	5	3.49 (2.47, 4.33)	**2.23E-07**	26.82	28	1.18 (0.75, 1.57)	**6.30E-07**	24.82
	LRP5 (n=1414)	0	-0.82 (-5.65, 1.13)	5.08E-01	0.44	70	0.14 (-0.11, 0.38)	2.72E-01	1.21
	PKD1 (n=54)	42	8.39 (7.64, 9.23)	**0.00E+00**	**Inf**	42	4.73 (4.11, 5.43)	**0.00E+00**	**Inf**
	PKD2 (n=24)	17	8.42 (7.41, 9.55)	**0.00E+00**	**Inf**	21	5.23 (4.16, 6.58)	**0.00E+00**	**Inf**
	PKHD1 (n=522)	2	1.8 (0.21, 2.89)	**3.07E-02**	4.67	27	0.13 (-0.28, 0.5)	5.18E-01	0.42
	PRKCSH (n=74)	0	2.23 (-2.62, 4.2)	2.49E-01	1.33	3	-0.05 (-1.35, 0.89)	9.23E-01	0.01
	SEC63 (n=81)	0	2.1 (-2.74, 4.07)	2.69E-01	1.22	8	0.96 (0.16, 1.63)	**2.07E-02**	5.35

Figure 2. Phenotype first approach. **A)** A total of **173,954** participants with whole exome sequencing were evaluated for ADPKD ICD codes (Q61.2, Q61.3, 751.12, 751.13), followed by chart review of 303 participants. 196 participants had typical ADPKD, 16 had mild ADPD, 23 had atypical-ADPKD, 26 had other kidney diseases and 42 did not have enough data to diagnose any kidney disease. **B)** Of the 26 people who had other kidney diseases, 6 had ARPKD, 2 had TSC, 2 had CAKUT, 2 had 17q12 syndrome, and 4 had a family history of ADPKD.

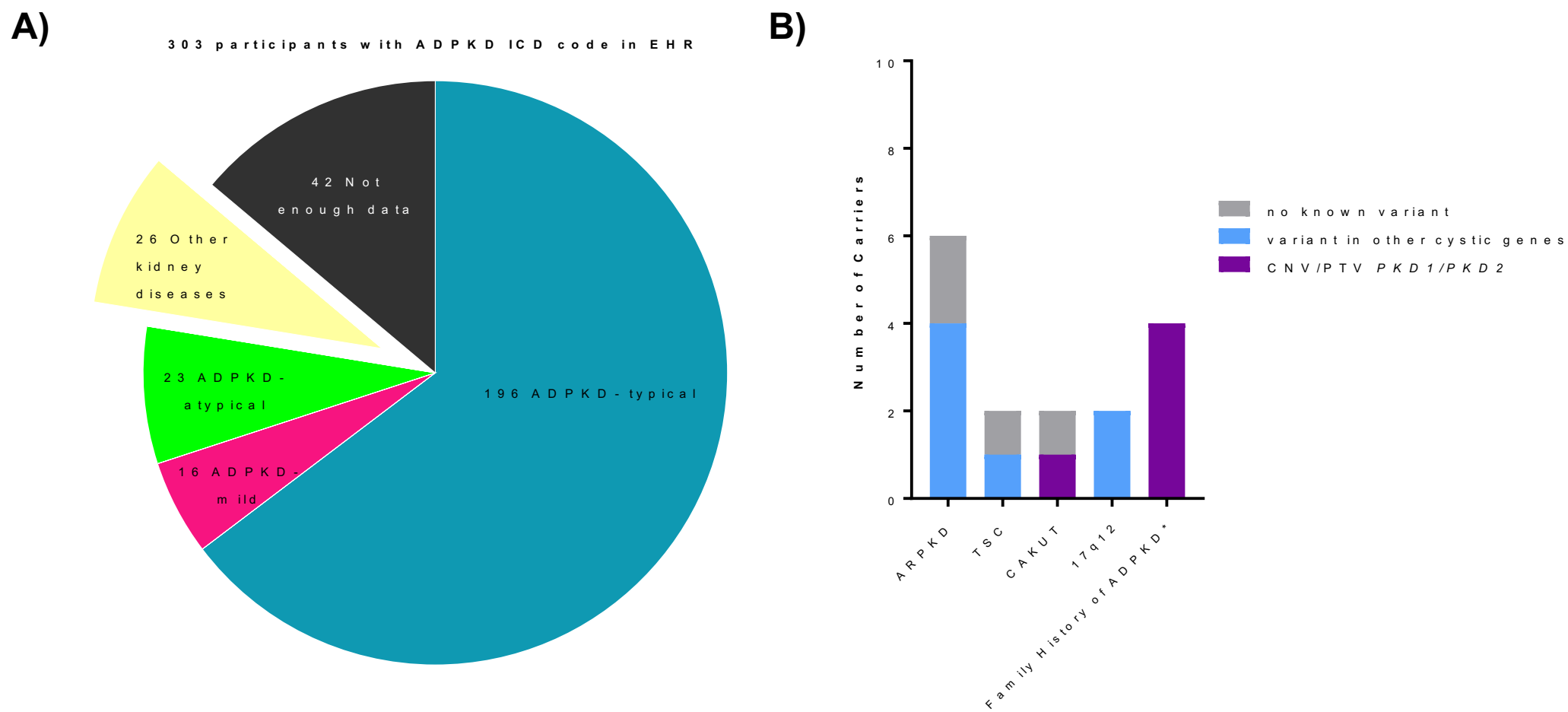


Figure 3. Copy Number Variants (CNVs) or Protein Truncating Variants (PTVs) in *PKD1/PKD2*. **A)** 99 out of 196 individuals with ADPKD were carriers of PTVs or CNVs of *PKD1* or *PKD2*. Number of each type of CNV/PTV in PKD1/PKD2 are shown in the bar graph. **C)** Example pedigree of family with PTV *PKD1_Glu1061Ter* carriers. **D)** Example kidney CT scan imaging for participant ADPKD130, ADPKD144, ADPKD155833 and ADPKD5.

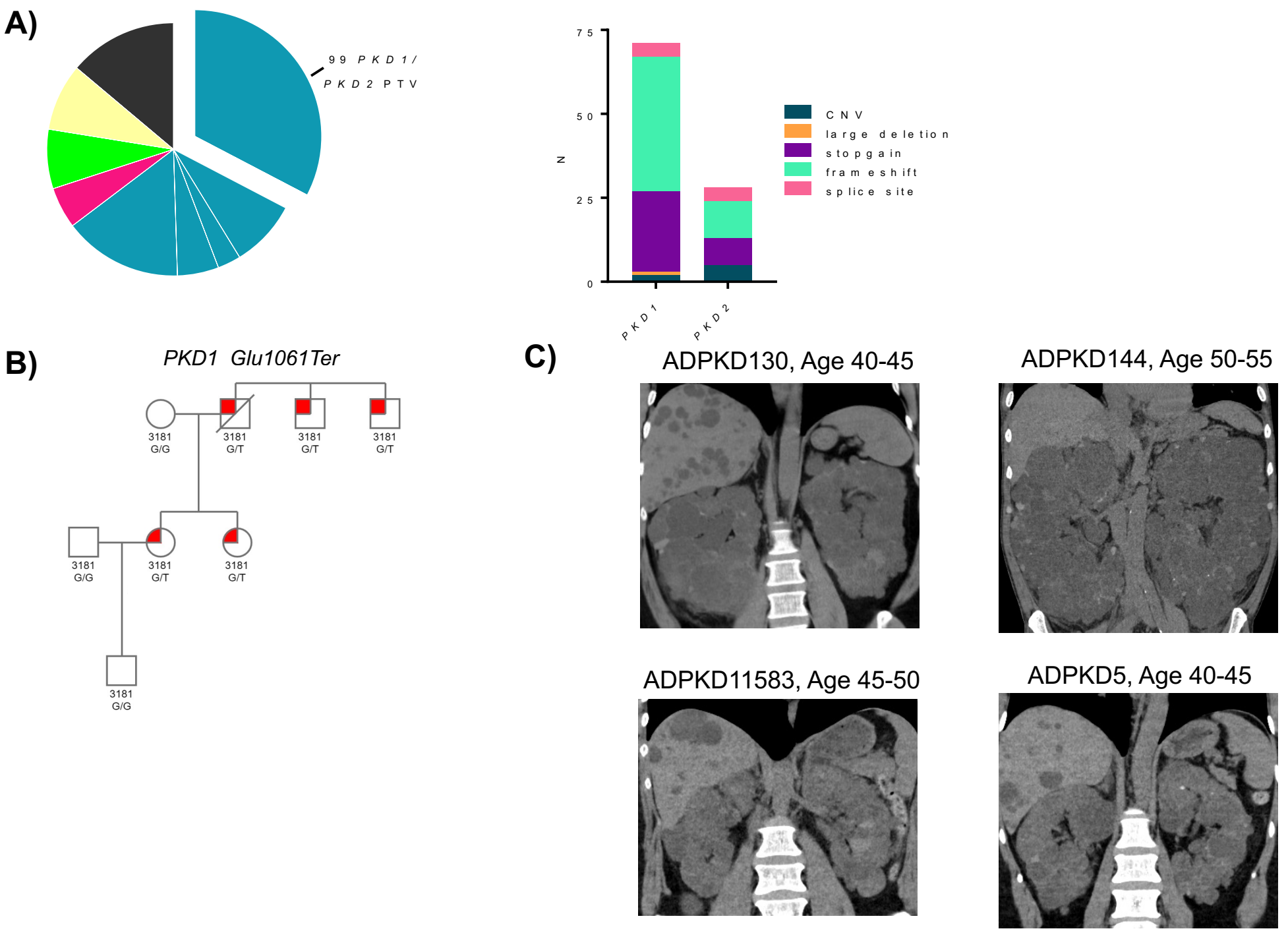
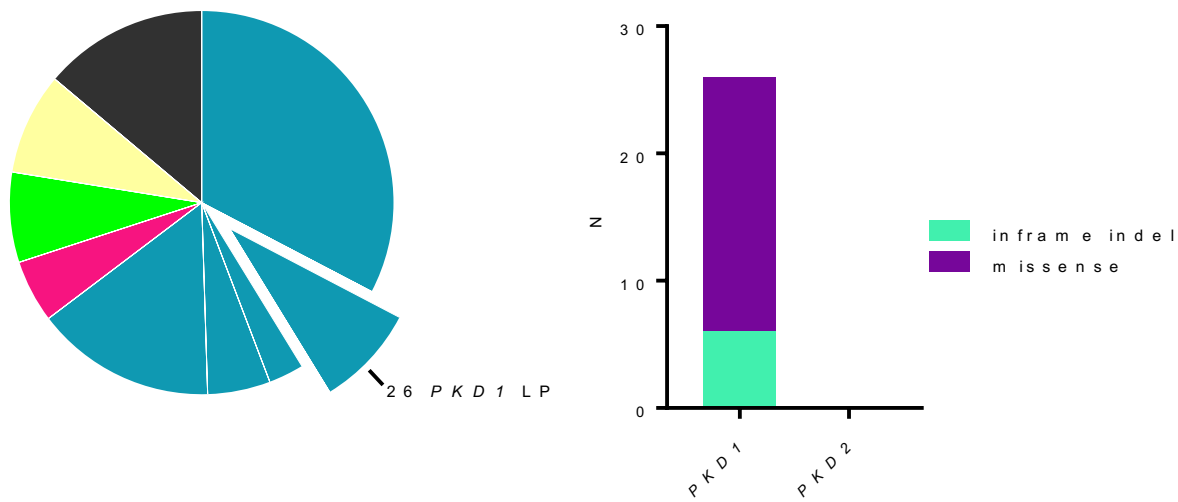
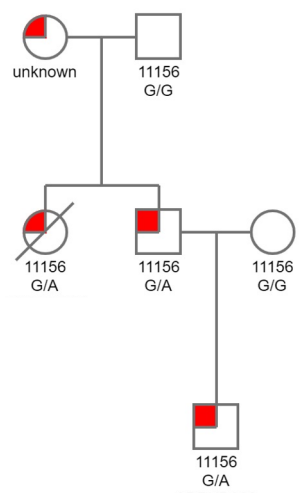


Figure 4. *PKD1/PKD2* reported as Likely Pathogenic (LP) in PKDB. **A)** 26 out of 196 individuals with ADPKD were found to be carriers of LP variants of *PKD1* or *PKD2 with n*umber of inframe insertion/deletions and missense variants classified as likely pathogenic by PKDB. **B)** Example pedigree of family with LP variant *PKD1_Arg3719Gln* carriers. **C)** Example kidney CT imaging for participant ADPKD339 and ADPKD100.

A)

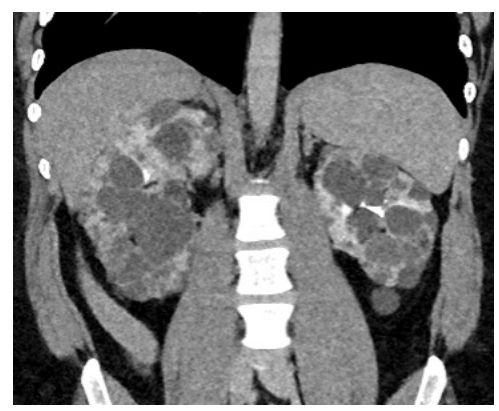


B) *PKD1_Arg3719Gln*



C)

ADPKD339, Age 25-30



ADPKD100, Age 50-55

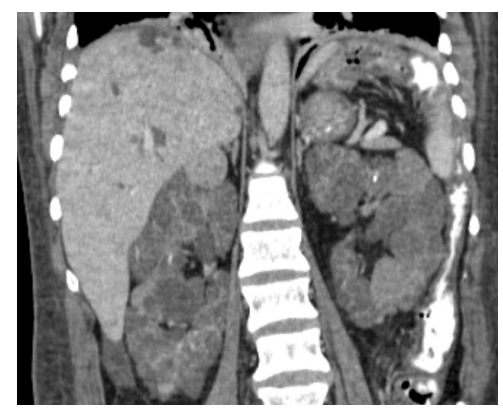


Figure 5. Carriers of variants in *PKD1/PKD2* that are novel or indeterminate or likely benign in the PKDB. **A)** 26 out of 196 individuals with ADPKD were found to be carriers of variants in *PKD1* or *PKD2* that are classified in PKDB as indeterminate or likely benign or not present in PKDB and therefore "novel". **B)** Example pedigree of carriers of variant *PKD1 Ala2511Pro*, which had been listed in the PKDB as "likely benign". The proband has a two first degree relatives with ADPKD who underwent cascade testing and were found to have the Ala2511Pro as well, suggesting this variant should be reclassified. **C**) Kidney CT imaging of ADPKD331. **D, E)** Pedigrees and imaging of carriers of novel *PKD1 Gly2034Val* missense variant. Three out of four carriers have ADPKD, while the one without current evidence of ADPKD is <25 years old (*). Non-carrier relatives in either family do not have an ADPKD phenotype**. F, G)** Pedigree and available CT imaging for carriers of *PKD1 Asp3304Asn.* In total, 6/10 have imaging evidence of ADPKD, 3 are shown. Of the 4 carriers with no evidence, 3 do not have abdominal imaging and one is under age 40. **H, I)** Pedigree and imaging for carriers of *PKD1Ser3591Phe, of whom* 3 of 3 have ADPKD. While non-carrier family members do not have ADPKD.

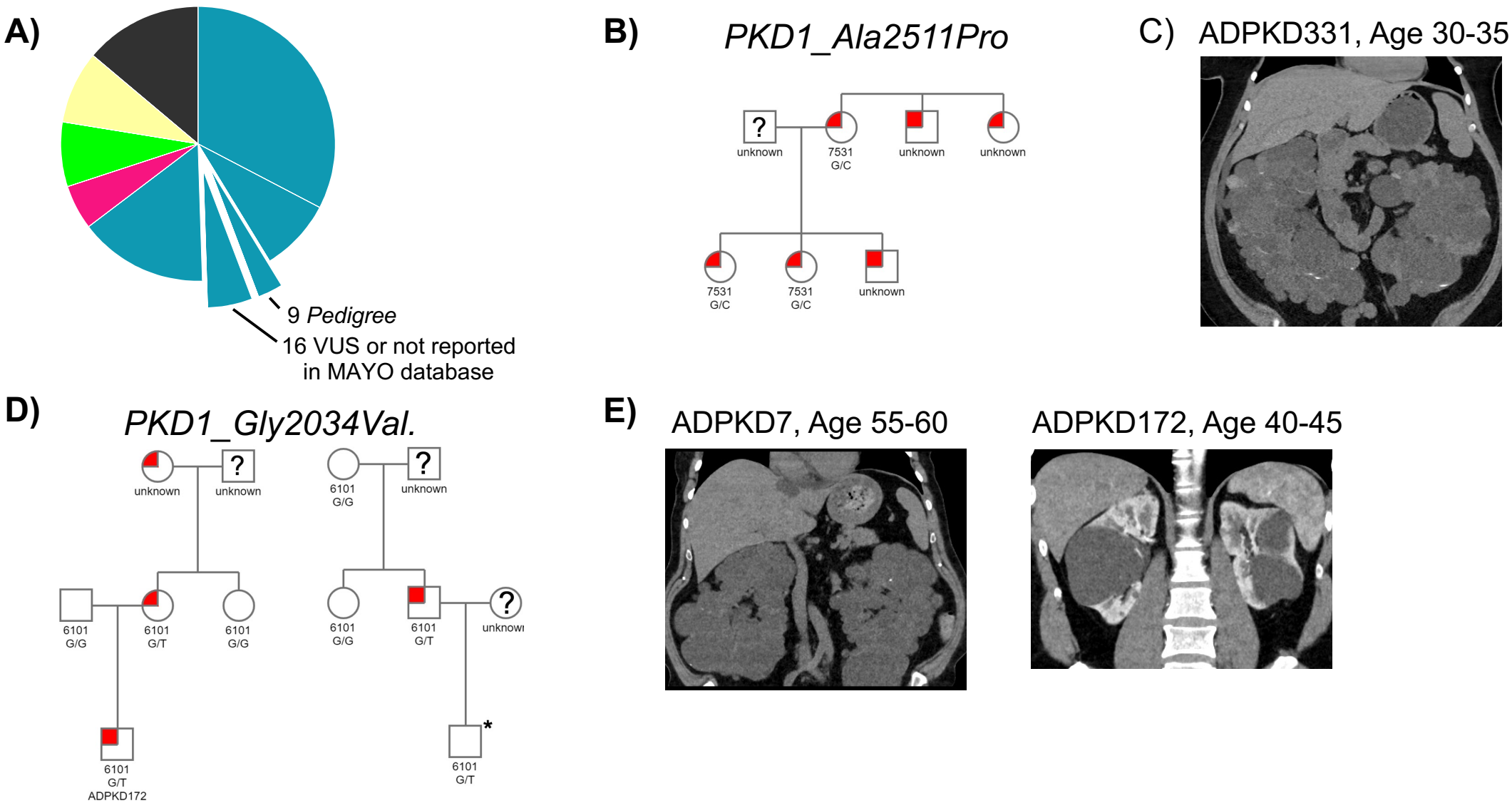
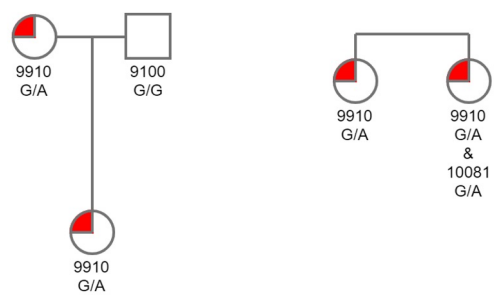


Figure 5, Continued:

F)

PKD1_Asp3304Asn.

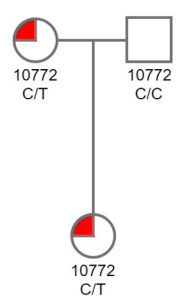


G)

ADPKD3, Age 60-65 ADPKD207, Age 60-65 ADPKD336, Age 70-75

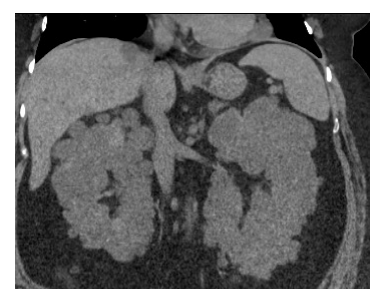
H)

PKD1_Ser3591Phe



I)

ADPKD118, Age 40-45



ADPKD198 , Age 40-45

Figure 6. Carriers of variants in genes other than *PKD1/PKD2*. **A)** From the initial 247 participants, 82 did not have any rare variants detected in *PKD1* or *PKD2*. WES was evaluated for any variant in *ALG8, ALG9, DNAJB11, GANAB, HNF1b, IFT40 LRP5, PKHD1, PRKCSH, SEC63, SEC61b.* **B)** Four patients have a previously unreported *GANAB_Asp647Val* variant, 2 participants have a PTV and 6 more have a missense variant in one of these genes. Whether these variants are the genetic cause of ADPKD is undetermined. Additionally, 69 people (27.8%) have no identified candidate variant to explain their ADPKD. **C)** Pedigrees of carriers of *GANAB_Asp647Val*. Within two families with at least one carrier of *GANAB_Asp647Val*, only those carriers have ADPKD. In total 5/7 carriers of *GANAB_Asp647Val* have evidence of ADPKD (though may not have an ICD code in their EHR), 1 without imaging and 1 multiple hypodensities in kidneys and liver, that don't quite qualify as ADPKD. **D)** CT Imaging from participants with *GANAB_Asp647Val.* **E)** CT Imaging from participants with *IFT140* variants.

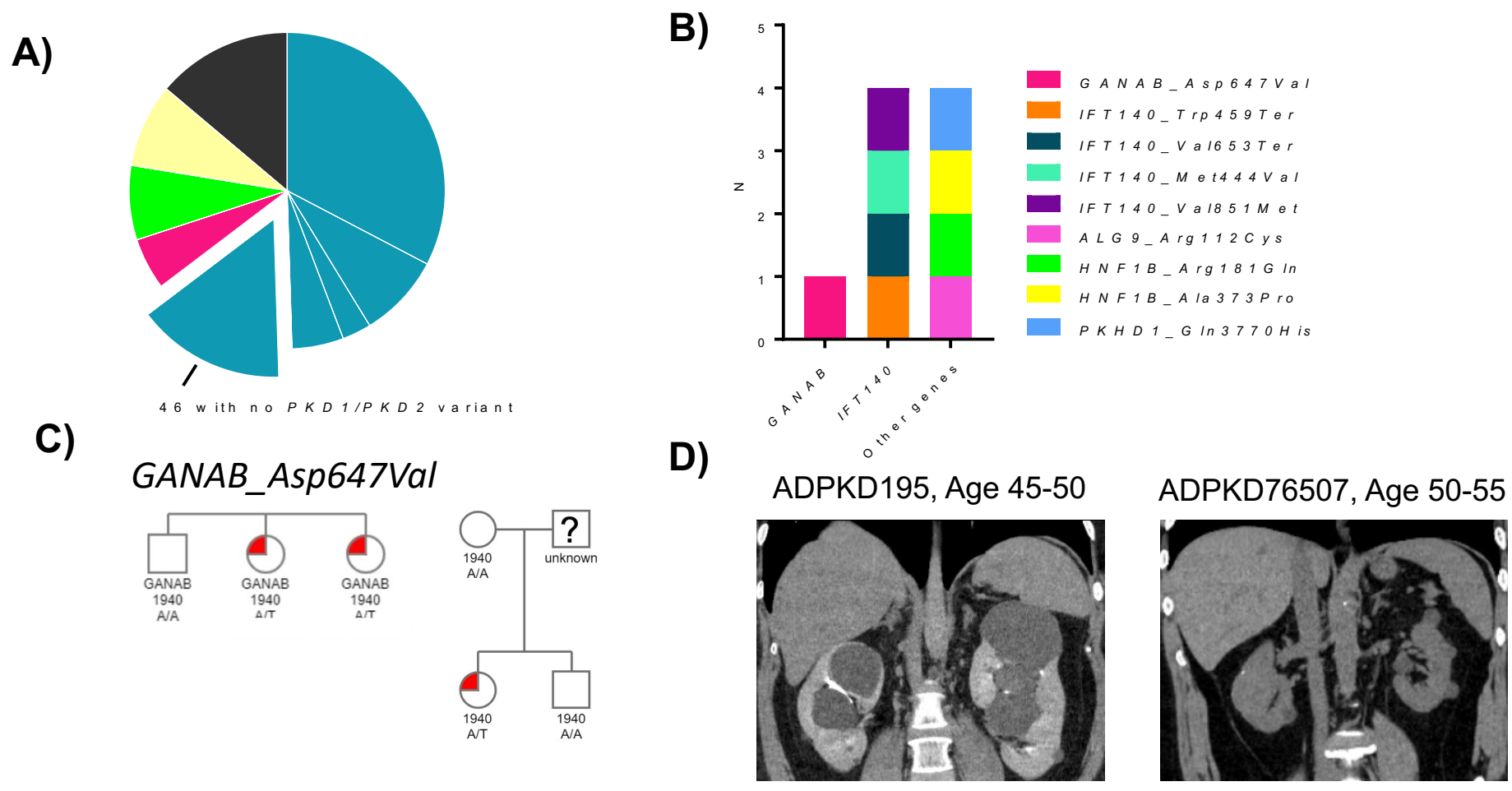


Figure 6. continued

E)

ADPKD136869
Age 50-55
IFT140 Trp653Ter
1959 G/A

ADPKD283
Age 80-85
IFT140 Trp459Ter
1377 G/A

ADPKD83
Age 55-60
IFT140 Arg1404Gln
4211G/A

ADPKD15
Age 55-60
IFT140 Met444Val
1330 A/G

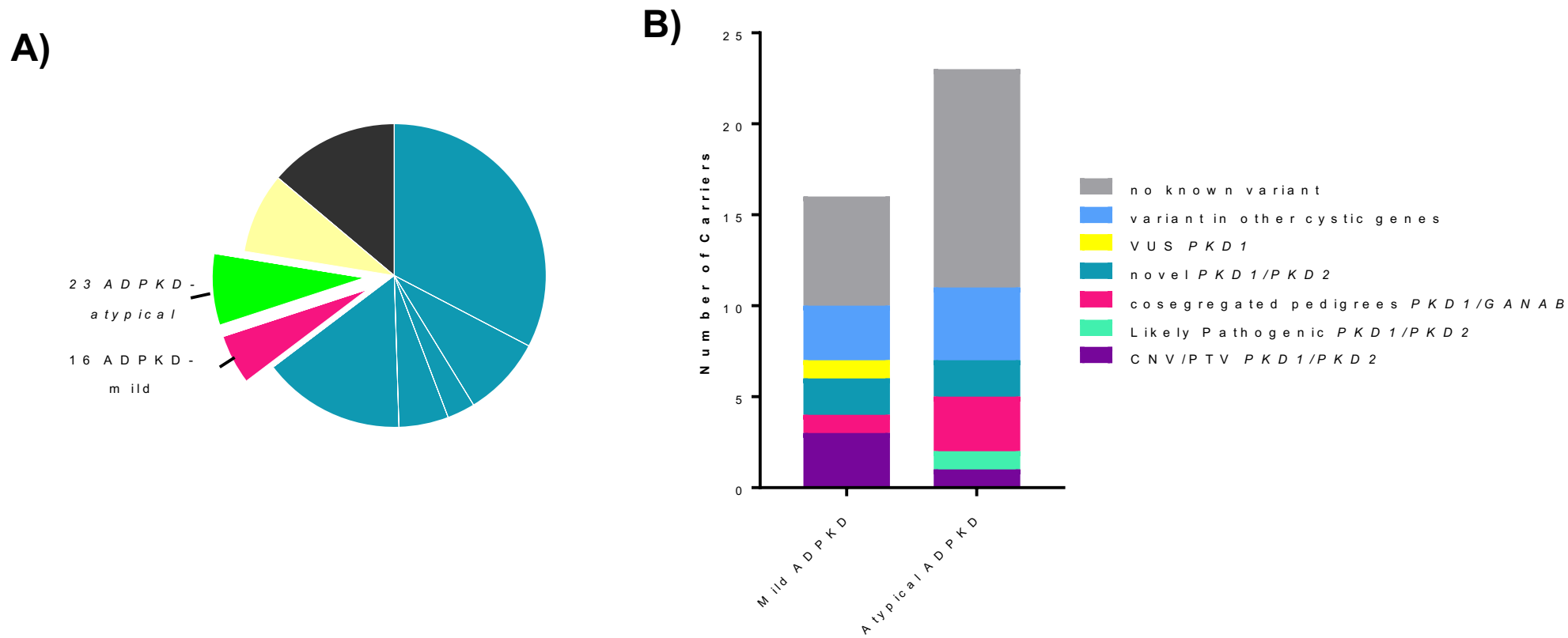



Figure 7. Individuals with atypical or mild forms of ADPKD are much less likely to have a potential variant identified as a contributor to their ADPKD.