Prevalence Estimates of Polycystic Kidney and Liver Disease by Population Sequencing

Matthew B. Lanktree,[1] Amirreza Haghighi,[1] Elsa Guiard,[1] Ioan-Andrei Iliuta,[1] Xuewen Song,[1] Peter C. Harris,[2] Andrew D. Paterson,[3,4,5] and York Pei[1]

[1]Division of Nephrology, University Health Network, University of Toronto, Toronto, Ontario, Canada; [2]Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota; [3]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; and Divisions of [4]Epidemiology and [5]Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

ABSTRACT

Background Estimating the prevalence of autosomal dominant polycystic kidney disease (ADPKD) is challenging because of age-dependent penetrance and incomplete clinical ascertainment. Early studies estimated the lifetime risk of ADPKD to be about one per 1000 in the general population, whereas recent epidemiologic studies report a point prevalence of three to five cases per 10,000 in the general population.

Methods To measure the frequency of high-confidence mutations presumed to be causative in ADPKD and autosomal dominant polycystic liver disease (ADPLD) and estimate lifetime ADPKD prevalence, we used two large, population sequencing databases, gnomAD (15,496 whole-genome sequences; 123,136 exome sequences) and BRAVO (62,784 whole-genome sequences). We used stringent criteria for defining rare variants in genes involved in ADPKD (*PKD1*, *PKD2*), ADPLD (*PRKCSH*, *SEC63*, *GANAB*, *ALG8*, *SEC61B*, *LRP5*), and potential cystic disease modifiers; evaluated variants for quality and annotation; compared variants with data from an ADPKD mutation database; and used bioinformatic tools to predict pathogenicity.

Results Identification of high-confidence pathogenic mutations in whole-genome sequencing provided a lower boundary for lifetime ADPKD prevalence of 9.3 cases per 10,000 sequenced. Estimates from whole-genome and exome data were similar. Truncating mutations in ADPLD genes and genes of potential relevance as cyst modifiers were found in 20.2 cases and 103.9 cases per 10,000 sequenced, respectively.

Conclusions Population whole-genome sequencing suggests a higher than expected prevalence of ADPKD-associated mutations. Loss-of-function mutations in ADPLD genes are also more common than expected, suggesting the possibility of unrecognized cases and incomplete penetrance. Substantial rare variation exists in genes with potential for phenotype modification in ADPKD.

J Am Soc Nephrol 29: 2593–2600, 2018. doi: https://doi.org/10.1681/ASN.2018050493

Recent advances in our understanding of autosomal dominant polycystic kidney disease (ADPKD) have highlighted the important contribution of both gene locus and allelic heterogeneity to variable age-dependent penetrance and disease severity.[1,2] Specifically, patients with *PKD1* mutations have more severe disease with larger kidneys and earlier onset of ESRD compared with patients with *PKD2* mutations or no mutation detected.[1–5] The *PKD1* mutation class also predicts kidney disease severity, with the most severe disease associated with truncating mutations (non-sense, frameshift, canonical splice site mutations, and large gene insertions/deletions) and in-frame insertions/deletions (indels), and generally milder disease with missense mutations.[1,2] However, substantial within-family

Received May 8, 2018. Accepted July 27, 2018.

A.D.P. and Y.P. are cosenior authors.

Published online ahead of print. Publication date available at www.jasn.org.

Correspondence: Dr. York Pei, Division of Nephrology, University Health Network, University of Toronto, 8N838, 585 University Avenue, Toronto, ON M5G 2N2, Canada. Email: york.pei@uhn.ca

disease variability is well documented in ADPKD, suggesting the presence of genetic and environmental modifiers.[6]

Liver cysts are common in ADPKD, ranging from a few cysts to innumerable cysts resulting in massive polycystic liver.[7] By contrast, autosomal dominant polycystic liver disease (ADPLD) is a distinct clinical entity diagnosed by the presence of 20 or more liver cysts, with few or no kidney cysts.[7,8] A common genetic pathway for ADPKD and ADPLD has been discovered. Specifically, mutations in one of five genes (*PRKCSH*, *SEC63*, *GANAB*, *ALG8*, and *SEC61B*) result in reduced polycystin-1 biogenesis, post-translational modification, and trafficking to the cell membrane, leading to ADPLD.[9] In the setting of ADPKD, the presence of a germline mutation in any one of these ADPLD genes may potentially modify the cystic disease severity by further lowering the functional polycystin-1 dosage.[4]

Estimating the prevalence of ADPKD has been challenging because of variable age-dependent penetrance and incomplete clinical ascertainment in the general population. Two early epidemiologic studies of ADPKD have estimated a lifetime risk of one per 1000[10] and an annualized incidence of 2.75 per 100,000 person-years.[11] By contrast, more recent epidemiologic studies reported point prevalence of ADPKD at 4.76 per 10,000 in the province of Modena in Italy,[12] 3.96 per 10,000 by the European Dialysis and Transplant Association Registry,[13] and 5.73 per 10,000 by 60 years of age in southwest Germany.[14] Severe isolated ADPLD, presenting with numerous liver cysts and abdominal distention, is reported to be a very rare condition, with a prevalence of less than one per 100,000.[15,16] However, as ADPLD is rarely symptomatic except in extreme cases, it is likely underrecognized,[16] and an autopsy study reported multiple liver cysts in ten out of 10,104 retrospectively-collected consecutive autopsies and four out of 95 prospectively-collected consecutive autopsies.[17]

Next-generation sequencing has revolutionized our ability to identify rare genetic variants in populations.[18] The development of large population-based sequencing databases provides an important resource to advance clinical molecular diagnostics of ADPKD.[18,19] Although phenotype information is unavailable from participants, large population sequencing databases allow for the cataloguing and estimation of mutation carrier rates in cystic disease genes. There is a concern about the diagnostic accuracy of whole-exome sequencing (WES) in ADPKD because of the presence of six pseudogenes with a high degree of sequence identity with *PKD1*; inadvertent capture of sequence reads from pseudogene regions could result in false positive and negative variant calls.[3] However, two recent studies using WES reported low rates of false positive mutation calls.[20,21] Whole-genome sequencing (WGS) avoids sequence capture and utilizes long paired-end reads to improve variant calling to further avoid errant calls and is now available in large research cohorts.[22] Here, we report our analysis of rare variants in genes for ADPKD and ADPLD as well as other genes with potential relevance as cystic disease modifiers found in large WGS and WES databases.

Significance Statement

Many patients with autosomal dominant polycystic kidney disease (ADPKD) and autosomal dominant polycystic liver disease (ADPLD) remain asymptomatic as the disease progresses, and estimating prevalence of these conditions is challenging because of incomplete clinical ascertainment. The authors' analysis of population-based sequencing data of genes involved in ADPKD and ADPLD found that protein-truncating and clinically confirmed mutations provide a lifetime risk of ADPKD of at least nine cases per 10,000. Protein-truncating mutations in genes causative of ADPLD are as common as 20 cases per 10,000. Individually rare variants in genes of potential relevance as ADPKD modifiers are cumulatively common, but most have uncertain clinical significance. Further research using exome or targeted gene resequencing in patients with ADPKD will help evaluate these rare variants.

METHODS

Data Acquisition

Data for this study was obtained from two large public sequence databases that were collected to catalog the spectrum of genetic variation across the general population: gnomAD (release 2.0.2, available at gnomad.broadinstitute.org; n=15,496 diploid genome sequences and n=123,136 diploid exome sequences)[18] and BRAVO (powered by TOPMed Freeze5 on genome reference consortium human build 38 patch release 12, available at bravo.sph.umich.edu/freeze5/hg38/#; n=62,784 diploid whole-genome sequences). These databases were developed to improve clinical interpretation of rare variants in Mendelian disease (*i.e.*, common alleles cannot be causative), identify genes with poor tolerance to mutations, and identify human knockouts with homozygous loss-of-function mutations.[18,19] Participants were ascertained from a wide array of family-based, case-control, cohort, and randomized, controlled trial populations. Both gnomAD and BRAVO exclude related individuals and patients with pediatric-onset disease present in the index case or a first-degree family member, but do include participants with adult-onset disease, including cardiovascular disease, diabetes, and psychiatric conditions (Supplemental Appendix 1). Both databases include participants from multiple ethnicities including European, Finnish, African, South Asian, and Latino. GnomAD provides ethnic specific allele counts, but BRAVO is unable to provide the ethnicity of rare variant carriers. This study included analysis of data published in online repositories and ethics approvals were obtained by contributing studies. Both sites were accessed November 12, 2017 for data downloads.

Variant Assessment

Variants were filtered to exclude those observed with minor allele frequency (MAF) >0.01% (Supplemental Figure 1). Published observations report that >97% of disease-causing alleles in 77 well characterized genes were at population frequencies <0.01%.[23] As homozygous large effect variants in ADPKD have not been reported and likely result in embryonic lethality,[24] variants with a homozygous observation were

excluded. Intronic and synonymous variants were also excluded. Variants were filtered by their effect on the canonical transcript according to GTEx (gtexportal.org, version 7).[25] Rare variants were compared with data curated in the Mayo Polycystic Kidney Disease database (PKDB, version 3.1; pkdb.mayo.edu). In-frame indels were evaluated with VEST4.[26] Pathogenic variants were defined as protein-truncating mutations, including non-sense mutations, frameshift mutations, and mutations to canonical splice donor or acceptor sites, in-frame indels deemed pathogenic in VEST4 ($P<0.005$), or those listed as "definitely" or "highly likely" pathogenic mutations in PKDB. Likely pathogenic variants included "likely pathogenic" mutations in PKDB and in-frame indels deemed likely pathogenic in VEST4 ($0.05<P<0.005$), whereas bioinformatic missense variants were predicted deleterious by more than 12 out of 16 bioinformatic prediction algorithms (SIFT, PolyPhen2-HumDiv, PolyPhen2-HumVar, MutationTaster, MutationAssessor, LRT, FATHMM, FATHMM-MKL, PROVEAN, VEST3, MetaSVM, MetaLR, M-CAP, CADD, DANN, and GenoCanyon) in the web-based ANNOVAR (wannovar.wglab.org) representing the most pathogenic fifth percentile of all variants.[27]

Genes for ADPLD (*SEC63*, *PRKCSH*, *GANAB*, *ALG8*, *SEC61B*, *LRP5*) were similarly examined in gnomAD and BRAVO. We also examined genes with potential to cause other cystic disease or modify disease severity (*PKHD1*, *DZIP1L*, *HNF1B*, *TSC1*, *TSC2*, *VHL*, *COL4A1*, *COL4A3*, *COL4A4*, *COL4A5*). For recessive conditions, a MAF threshold of 0.1% was used and the disease prevalence was estimated by squaring the sum of the pathogenic alleles (as per Hardy–Weinberg equilibrium). To provide a lower boundary for the disease prevalence, we report the prevalence of protein-truncating mutations.

Statistical Analyses

Cumulative frequencies were obtained by summing the number of observed alleles in each database separately. There is a possibility of sample overlap between the two databases (Supplemental Appendix 1), so both populations were analyzed separately and a weight-based meta-analysis was performed to obtain a conservative estimate of the prevalence of disease causing alleles. We calculated 95% confidence intervals (95% CIs) by the normal approximation (Wald) interval

$\hat{p} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$, with z=1.96.[28] Carrier frequency was estimated with Hardy–Weinberg equilibrium ($2pq$).

RESULTS

Genetic Lifetime Prevalence Estimates for ADPKD from WGS

WGS of 62,784 BRAVO participants identified 6369 *PKD1* variants and 808 *PKD2* variants with MAF<0.01%, including 2226 and 295 nonsynonymous variants, respectively, whereas WGS of 15,496 gnomAD participants identified 3343 *PKD1* variants and 433 *PKD2* variants with MAF<0.01%, including 1109 and 116 nonsynonymous variants, respectively (Supplemental Figure 2, Supplemental Appendix 2). Variants were distributed uniformly across exons of *PKD1* and *PKD2* without evidence of a mutational "hotspot," and between the duplicated and unduplicated regions of *PKD1* without apparent enrichment in any coding region (Supplemental Figures 3–5). Excluded noncanonical variants included: 11 mutations in BRAVO and four mutations in gnomAD that were protein-truncating in noncanonical transcript annotations with altered reading frames but produced synonymous changes in the canonical transcript; and 19 mutations in BRAVO and four mutations in gnomAD in noncanonical splice donor or acceptor sites (Supplemental Appendices 3 and 4). All observed mutations were examined for presence in the Mayo PKDB. The overall prevalence of high-confidence pathogenic *PKD1* mutations was 6.8 per 10,000 (95% CI, 5.0 to 8.6) and pathogenic *PKD2* mutations was 2.6 per 10,000 (95% CI, 1.4 to 3.7), yielding a cumulative prevalence of 9.3 per 10,000 (95% CI, 7.2 to 11.5) (Figure 1). The frequency of likely pathogenic *PKD1* and *PKD2* mutations was 14.1 per 10,000 and 3.3 per 10,000, respectively. The prevalence of bioinformatic-predicted *PKD1* and *PKD2* mutations was 32.3 per 10,000 and 12.7 per 10,000, respectively.

Of the observed high-confidence pathogenic alleles, 13 were *PKD1* truncating mutations, nine were in-frame indels, 27 were PKDB highly likely *PKD1* missense mutations, and 18 were truncating or PKDB highly likely *PKD2* missense mutations (Supplemental Appendices 3 and 4). Of 42 variants, 31 (74%) were singletons, seven (17%) were observed twice, and two were observed three times. The remaining two highly likely pathogenic variants were p.Asp2557Gly (rs757619679, seven observations; MAF_{WGS}=0.0044%; MAF_{WES}=0.0028%) and p.Gly2858Ser (rs755522953, nine observations; MAF_{WGS}=0.0057%; MAF_{WES}=0.0053%). None of the observed protein-truncating mutations in *PKD1* or *PKD2* were located in the last exon. We also examined if the pathogenic alleles were present above our MAF cutoff of 0.01%. One highly likely *PKD1* PKDB missense variant (p.Cys259Tyr), from a single published case, was found in seven alleles (MAF=0.22%) in gnomAD and 20 alleles (MAF=0.28%) in BRAVO. No truncating mutations or large effect indels were observed above the MAF threshold of 0.01%.

Examining the "likely hypomorphic" alleles in PKDB, three out of nine were not in BRAVO or gnomAD (Supplemental Appendix 5). The most commonly observed was p.Arg2765-Cys, in 734 heterozygotes and four homozygotes in BRAVO (MAF=0.58%) and 136 heterozygotes and two homozygotes in gnomAD (MAF=0.44%). Another hypomorphic *PKD1* allele (p.R3277C), which has been experimentally validated,[4] was observed in 33 participants in BRAVO (MAF=0.026%) and 14 participants in gnomAD (MAF=0.045%), with no observed homozygotes.

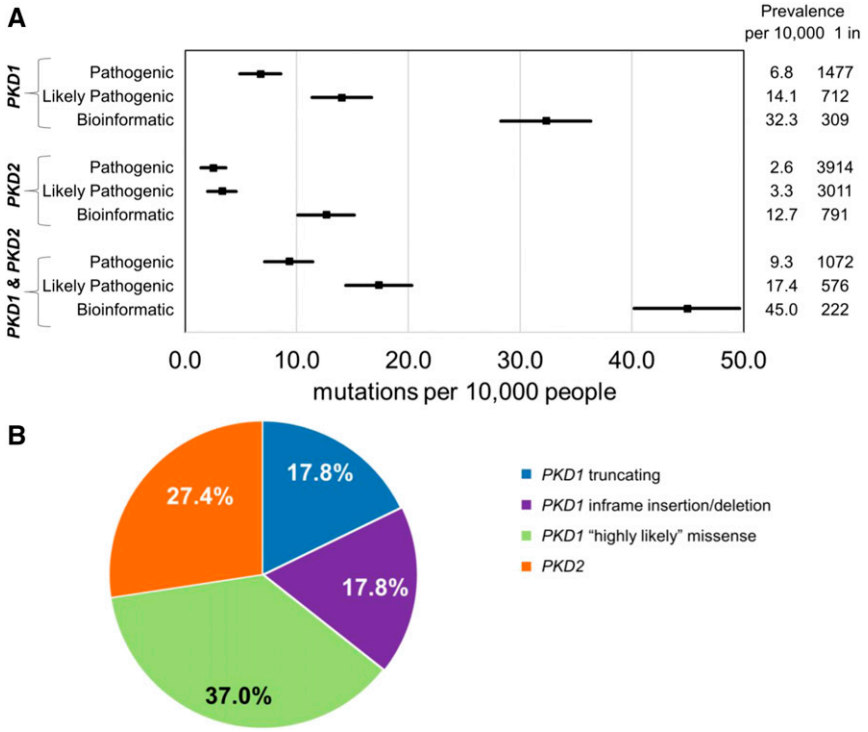

Figure 1. Observed prevalence of ADPKD mutations in population sequencing. (A) Cumulative frequency of high-confidence pathogenic (non-sense, frameshift, and canonical splice site mutations, high-risk in-frame insertions or deletions, or those listed as "definitely" or "highly likely" pathogenic mutations in Mayo PKDB), likely pathogenic ("likely pathogenic" mutations in Mayo PKDB and moderate risk in-frame insertions or deletions), or bioinformatic predicted pathogenic (predicted pathogenic by more than 12 out of 16 bioinformatic algorithms in ANNOVAR) from WGS data. Error bars represent 95% CIs. (B) Proportion of high-confidence pathogenic mutations by mutation class in WGS databases.

Genetic Prevalence Estimates for ADPKD from WES

GnomAD includes more samples and better representation of ethnicities in its WES than WGS data, but we were concerned about the accuracy of *PKD1* variant calling in WES. Qualitatively, the coverage of *PKD1* appears better with WGS than WES, with the entire gene obtaining read depths >20× for all participants (Supplemental Figures 6 and 7). There was no significant difference between the observed mutation frequency in either the duplicated or unduplicated region of *PKD1* using WGS or WES (Figure 2). As expected, the given coverage is good for *PKD2* in WES, and we observed no difference in *PKD2* mutation frequency between WES and WGS. Because variants observed in PKDB could be over-represented for alleles present in Europeans, we compared the frequency of protein-truncating mutations and pathogenic in-frame indels between populations. With >7500 people sequenced, there appeared to be a slightly lower prevalence of *PKD1* mutations in the Finnish population, but no significant variation in mutation prevalence across ethnicities overall (Figure 3). No protein-truncating mutation or indel was observed in more than one ethnic group.

Genetic Prevalence Estimates of ADPLD and Other Potential Cystic Disease Modifier Genes

Mutations in six genes have been identified to cause ADPLD, with five of them affecting polycystin-1 maturation and trafficking.[4] On the basis of truncating mutations in the combined WGS and WES dataset, we found that three well characterized ADPLD genes (*PRKCSH*, *SEC63*, *GANAB*) cumulatively yield a prevalence of 9.2 per 10,000 (one in 1082). Including all six ADPLD genes, truncating mutations were observed in 20.2 per 10,000, corresponding to a prevalence of one in 496 (Figure 4). The cumulative allele frequency of all individually rare (MAF<0.01%) missense variants in the six ADPLD genes was 2.17% (95% CI, 2.12% to 2.21%). Next, we examined the prevalence of truncating mutations in genes of potential

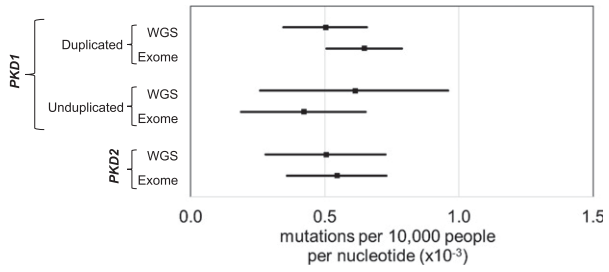

Figure 2. The density of mutations is similar in *PKD1* duplicated and unduplicated regions and *PKD2*. Mutation frequency estimates, corrected for transcript length, from WGS (*n*=78,280) and WES (*n*=123,136) data. Error bars represent 95% CIs.

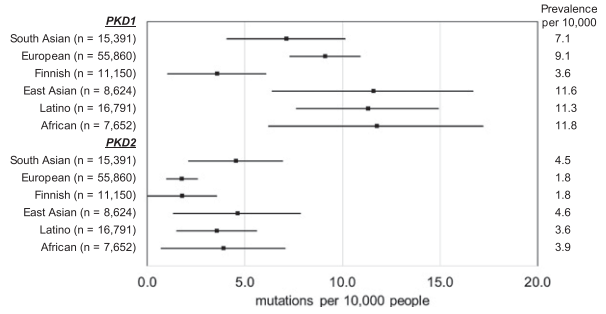

Figure 3. No difference in ADPKD mutation prevalence was observed between ethnicities. Prevalence of truncating mutations and severe in-frame insertion/deletions in exome sequencing of *PKD1* and *PKD2* across ethnic groups. Error bars represent 95% CIs.

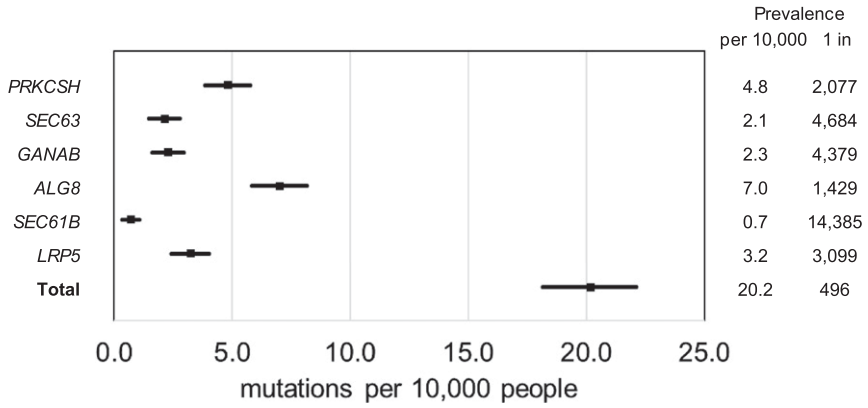

Figure 4. ADPLD protein-truncating mutations are more prevalent than expected. Error bars represent 95% CIs.

relevance as renal cystic disease modifiers (Table 1). The cumulative allele frequency of truncating mutations in *PKHD1* and *DZIP1L* was 0.22% and 0.05%, with a Hardy–Weinberg predicted carrier frequency of 0.45% (one in 224) and 0.10% (one in 959), respectively. Truncating mutations in *TSC1* and *TSC2* were cumulatively observed in 4.3 per 10,000, corresponding to a prevalence of tuberous sclerosis complex at one in 2315. Rare truncating mutations were observed in genes associated with Von Hippel–Lindau disease (one in 6603), and hereditary angiopathy with nephropathy, aneurysms, and muscle cramps syndrome (one in 11,188). Mutations in *COL4A3*, *COL4A4*, and *COL4A5* cause familial hematuric nephropathies including thin basement membrane disease and Alport syndrome in autosomal dominant, recessive, and sex-linked inheritance patterns, and could modify the severity of ADPKD. Truncating *COL4A3* and *COL4A4* mutations were observed in 11.5 per 10,000 and 7.6 per 10,000, respectively. These data provide a lower bound prevalence estimate of thin basement membrane disease at 0.2% in the general population. By contrast, only ten *COL4A5* truncating alleles were observed, including one homozygous female and one hemizygous male with previously unreported splice donor mutations (c.231+2T>C). Summing the truncating mutations across all genes of potential relevance as cystic disease modifiers yielded a carrier frequency of 103.9 per 10,000 (95% CI, 99.5 to 108.3) or approximately 1% of the general population. The cumulative allele frequency of rare missense variants in all potential modifier genes was 13.4% (95% CI, 13.3% to 14.5%) (Supplemental Appendix 6).

DISCUSSION

This study measured the frequency of high-confidence mutations presumed to be causative of polycystic kidney and liver disease in large, public WGS and WES databases. Study participants were unrelated, free from childhood-onset conditions, and represent the spectrum of genomic variation in the general population. Variants that predispose to adult-onset conditions would be expected to be present at the same frequency as the general population; however, specific cohorts/studies may have intentionally or unintentionally excluded patients with severe preexisting medical conditions, including ADPKD or ADPLD. Specifically, we suspect patients with severe ADPKD with advanced stages of kidney failure were

Table 1. Prevalence of truncating mutations in potential cystic disease modifier genes

			gnomAD-WGS		BRAVO WGS		gnomAD-Exome		Overall	
			n=15,847		*n*=62,784		*n*=123,117		*n*=201,388	
									PT Carriers	
Gene	Disease	MOI	PT	MAF	PT	MAF	PT	MAF	Per 10,000	One in
PKHD1	ARPKD	AR	36	0.12%	226	0.18%	640	0.26%	44.7	223
DZIP1L	ARPKD	AR	17	0.06%	87	0.07%	106	0.04%	10.4	959
HNF1B	RCAD	AD	0	0.00%	4	0.00%	3	0.00%	0.3	28,770
TSC1	TSC	AD	1	0.00%	7	0.01%	10	0.00%	0.9	11,188
TSC2	TSC	AD	3	0.01%	14	0.01%	52	0.02%	3.4	2919
VHL	VHL	AD	8	0.03%	28	0.02%	25	0.01%	3.0	3301
COL4A1	HANAC	AD	2	0.01%	12	0.01%	22	0.01%	1.8	5594
COL4A3	TBMD	AR/AD	16	0.05%	80	0.06%	135	0.06%	11.5	872
COL4A4	TBMD	AR/AD	4	0.01%	56	0.05%	93	0.04%	7.6	1316
COL4A5	AS	XL	1	0.00%	2	0.00%	7	0.00%	0.5	20,139

PT, protein-truncating mutations; MOI, mode of inheritance; *PKHD1*, polycystic kidney and hepatic disease 1 or fibrocystin; AR, autosomal recessive; *DZIP1L*, DAZ interacting zinc finger protein 1 like; *HNF1B*, hepatic nuclear factor 1β; RCAD, renal cysts and diabetes syndrome; AD, autosomal dominant; *TSC1*, tuberous sclerosis complex 1; TSC, tuberous sclerosis complex; *TSC2*, tuberous sclerosis complex 2; VHL, Von Hippel–Lindau disease; *COL4A1–A5*, collagen 4 chains; HANAC, hereditary angiopathy with nephropathy, aneurysms, and muscle cramps syndrome; TBMD, thin basement membrane disease; AS, Alport syndrome; XL, X-linked recessive.

less likely to participate in the sequencing studies. Among patients enriched with clinical risk factors for progression, such as those enrolled in the Consortium for Radiologic Imaging Studies of Polycystic Kidney Disease (CRISP), Genkyst, and Halt Progression of Polycystic Kidney Disease (HALT-PKD) studies, 51%–57% had a *PKD1* truncation mutation,[1] whereas in a less selected cohort such as the Toronto Genetic Epidemiology Study of PKD, 38.3% of patients had a truncating *PKD1* mutation.[2] By contrast, only 19.4% of the pathogenic mutations identified here were truncating. Additionally, up to 5% of mutations found in ADPKD are due to large deletions in *PKD1* or *PKD2*, which would have been missed in this analysis.[29] With these caveats, we found a lower bound genetic lifetime prevalence estimate of ADPKD at 9.3 per 10,000 (one out of 1072). This is somewhat higher than expected, given point prevalence estimates of approximately one in 2000 by recent epidemiologic studies.[12–14] Adjusting for the missing cases because of selection bias against severe truncating mutations and large gene rearrangements missed by sequencing would lead to an even higher genetic prevalence estimate.

Moreover, we did not include the cumulatively common but individually rare missense variants because of uncertainty in assessing their pathogenicity, further underestimating the true genetic prevalence of all of the cystic diseases reported here. The American College of Medical Genetics has standards for declaring pathogenicity of rare variants, including assessment of frequency in population data, identification of familial cosegregation, presence in clinically ascertained mutation databases, and functional and bioinformatic data, but most rare variants remain of uncertain clinical significance.[18] False attribution of a phenotype to an observed rare variant in a sequenced case can occur, leading to the overabundance of reportedly pathogenic variants in clinically ascertained sequencing databases.[30–32] Consistent with this notion, the high cumulative prevalence of likely pathogenic mutations for ADPKD (17.3 per 10,000, or one out of 576) suggests that many are not pathogenic. We did not find any high-confidence mutations at a prevalence incompatible with an ADPKD diagnosis (*i.e.*, MAF>0.01%) and most were singleton observations. We found bioinformatic prediction tools further overestimated the effect of rare variants, resulting in a further inflation of prevalence estimate. Raising the threshold for a prediction algorithm to call a particular variant pathogenic may increase the specificity, but would also reduce sensitivity.

False positive or false negative genotype calls may occur in the pseudogene region of *PKD1* (*i.e.*, exons 1–33) by DNA sequence capture in exome sequencing. However, these concerns have been allayed by two recent studies. Trujillano *et al.*[20] reported a 97.2% positive diagnostic rate in 36 previously genetically resolved families *via* Sanger sequencing by exome sequencing, with no false positive spurious calls. Similarly, Eisenberger *et al.*[21] reported a 99% positive diagnostic rate, on the basis of 681 out of 683 Sanger-verified sequencing variants (including pathogenic mutations) from 55 patient samples. WGS can avoid DNA sequence capture bias and false pseudogene mapping. Mallawaarachchi *et al.*[22] used WGS and found a positive diagnostic rate of 86% in 24 previously genetically uncharacterized patients, and importantly, there were no false positive calls. We found no significant differences in pathogenic mutation frequency between WGS and WES, enabling further examination of the larger WES dataset. Founder populations with higher prevalence of ADPKD are possible. However, looking at large outbred ethnic population subgroups in gnomAD, including South Asian, Latino, African American, Finnish, and European, we did not find evidence to support large ethnic differences in disease prevalence.

A responsible mutation is identified in only approximately 50% of patients with clinically ascertained ADPLD.[9] Yet, loss-of-function variants in genes mutated in ADPLD were much more frequent (one in 496) than would be expected by the reported prevalence of clinically recognized ADPLD cases, likely because of incomplete clinical ascertainment and incomplete penetrance. Disease severity can be quite variable in ADPLD, and in a substantial portion of cases, liver cysts in ADPLD are an incidental finding of little clinical relevance.[16] Any role mutations in these genes play to predisposition to simple cysts remains unknown. As five out of six ADPLD genes are known regulators of polycystin-1 maturation, they are ideal candidates to serve as modifiers of ADPKD. The predicted prevalence of ARPKD was inconsistent with observed annualized incidence of one in 26,485,[33] likely due to the embryonic or early infancy lethality of the condition.[34] The prevalence of *HNF1B* truncating mutations was low, at one per 28,770, but many cases are because of large deletions that would have been missed in this study.[35] Our estimated prevalence for tuberous sclerosis complex (one in 2315) is somewhat higher than published estimates of one per 5000–10,000, but the 3.8:1 ratio of *TSC2:TSC1* is consistent with previous observations.[36,37]

There are limitations in our study. Because individual-level genotype information is unavailable for public sequencing databases, we were unable to perform haplotype analysis to determine if variants with multiple observations were identical-by-descent because of a founder effect or recurrent mutation. Similarly, we were unable to determine if a single person is a carrier of mutations in more than one gene or were compound heterozygotes for two variants in a single gene. Furthermore, phase information is not available, which would be required to determine if mutations are found in *cis* or *trans.* Finally, phenotype information is also unavailable to determine if a sequenced study participant had previously unrecognized kidney or liver cysts.

In conclusion, by examining large population sequencing datasets we document a minimal genetic prevalence of ADPKD at 9.3 per 10,000. Cumulatively, we found up to 1% of the population carry a truncating mutation and up to 23% carry a rare missense variant in a gene of potential relevance as a cystic disease modifier. A similar prevalence of truncating

and missense variants from these genes would be expected in patients who also have a *PKD1 or PKD2* mutation, potentially contributing to phenotype modification. Further research using exome or targeted gene resequencing in patients with ADPKD will allow for evaluation of these rare variants. Rare cases of severe ADPKD due to bilineal inheritance of *PKD1* and *PKD2* mutations have been reported,[2,4,38] but evidence of a modifying effect from mutations of other cystic disease genes has not, except for *HNF1B* and *COL4A1*.[39,40] Exome and targeted gene resequencing will allow the identification rare variants in multiple cystic disease modifiers in individual patients with ADPKD. High-throughput *in vitro* functional screens using CRISPR-Cas9 editing will enable functional testing of the identified rare variants in cell or organoid-based assays.[41] These efforts are expected to advance precision medicine in ADPKD by providing more accurate prognostication to guide clinical management in an era of mechanism-based therapeutics.

ACKNOWLEDGMENTS

M.B.L., A.D.P., and Y.P. designed the study. M.B.L. collected and analyzed the data, produced the figures, and drafted the paper. All authors contributed to variant evaluation and edited and approved the final version of the manuscript.

M.B.L. received a Ben Lipps postdoctoral fellowship from the American Society of Nephrology and is a fellow of the Kidney Research Scientist Core Education and National Training program, a national kidney research training partnership of the Kidney Foundation of Canada, the Canadian Society of Nephrology, and the Canadian Institutes of Health Research. This work was supported in part by the Canadian Institutes of Health Research Strategy for Patient Oriented Research program grant in CKD (CAN-Solve-CKD, to Y.P.) and from the National Institutes of Health for the Molecular Genetics and Proteomics Core of the Mayo Translational PKD Center (DK090728, to P.C.H.).

DISCLOSURES

M.B.L. and Y.P. received compensation for participating in advisory and consultancy boards with Otsuka Pharmaceuticals. There are no additional competing interests or disclosures.

REFERENCES

1. Heyer CM, Sundsbak JL, Abebe KZ, Chapman AB, Torres VE, Grantham JJ, et al.: HALT PKD and CRISP Investigators: Predicted mutation strength of nontruncating PKD1 mutations aids genotype-phenotype correlations in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol* 27: 2872–2884, 2016
2. Hwang Y-H, 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* 27: 1861–1868, 2016
3. Song X, Haghighi A, Iliuta I-A, Pei Y: Molecular diagnosis of autosomal dominant polycystic kidney disease. *Expert Rev Mol Diagn* 17: 885–895, 2017
4. Cornec-Le Gall E, Torres VE, Harris PC: Genetic complexity of autosomal dominant polycystic kidney and liver diseases. *J Am Soc Nephrol* 29: 13–23, 2018
5. Harris PC, Bae KT, Rossetti S, Torres VE, Grantham JJ, Chapman AB, et al.: Cyst number but not the rate of cystic growth is associated with the mutated gene in autosomal dominant polycystic kidney disease. *J Am Soc Nephrol* 17: 3013–3019, 2006
6. Paterson AD, Magistroni R, He N, Wang K, Johnson A, Fain PR, et al.: Progressive loss of renal function is an age-dependent heritable trait in type 1 autosomal dominant polycystic kidney disease. *J Am Soc Nephrol* 16: 755–762, 2005
7. Van Keimpema L, De Koning DB, Van Hoek B, Van Den Berg AP, Van Oijen MGH, De Man RA, et al.: Patients with isolated polycystic liver disease referred to liver centres: Clinical characterization of 137 cases. *Liver Int* 31: 92–98, 2011
8. Lantinga MA, Gevers TJG, Drenth JPH: Evaluation of hepatic cystic lesions. *World J Gastroenterol* 19: 3543–3554, 2013
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* 127: 1772–1785, 2017
10. Dalgaard OZ: Bilateral polycystic disease of the kidneys; a follow-up of two hundred and eighty-four patients and their families. *Acta Med Scand Suppl* 328: 1–255, 1957
11. Iglesias CG, Torres VE, Offord KP, Holley KE, Beard CM, Kurland LT: Epidemiology of adult polycystic kidney disease, Olmsted County, Minnesota: 1935-1980. *Am J Kidney Dis* 2: 630–639, 1983
12. Solazzo A, Testa F, Giovanella S, Busutti M, Furci L, Carrera P, et al.: The prevalence of autosomal dominant polycystic kidney disease (ADPKD): A meta-analysis of European literature and prevalence evaluation in the Italian province of Modena suggest that ADPKD is a rare and underdiagnosed condition. *PLoS One* 13: e0190430, 2018
13. 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* 32: 1356–1363, 2017
14. Neumann HPH, Jilg C, Bacher J, Nabulsi Z, Malinoc A, Hummel B, et al.: Else-Kroener-Fresenius-ADPKD-Registry: Epidemiology of autosomal-dominant polycystic kidney disease: An in-depth clinical study for south-western Germany. *Nephrol Dial Transplant* 28: 1472–1487, 2013
15. D'Agnolo HMA, Kievit W, van Munster KN, van der Laan JJH, Nevens F, Drenth JPH: Center is an important indicator for choice of invasive therapy in polycystic liver disease. *Transpl Int* 30: 76–82, 2017
16. Qian Q, Li A, King BF, Kamath PS, Lager DJ, Huston J 3rd, et al.: Clinical profile of autosomal dominant polycystic liver disease. *Hepatology* 37: 164–171, 2003
17. Karhunen PJ, Tenhu M: Adult polycystic liver and kidney diseases are separate entities. *Clin Genet* 30: 29–37, 1986
18. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al.: ACMG Laboratory Quality Assurance Committee: 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* 17: 405–424, 2015
19. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al.: Exome Aggregation Consortium: Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536: 285–291, 2016
20. Trujillano D, Bullich G, Ossowski S, Ballarín J, Torra R, Estivill X, et al.: Diagnosis of autosomal dominant polycystic kidney disease using efficient PKD1 and PKD2 targeted next-generation sequencing. *Mol Genet Genomic Med* 2: 412–421, 2014
21. Eisenberger T, Decker C, Hiersche M, Hamann RC, Decker E, Neuber S, et al.: An efficient and comprehensive strategy for genetic diagnostics of polycystic kidney disease. *PLoS One* 10: e0116680, 2015
22. 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* 24: 1584–1590, 2016

23. Kobayashi Y, Yang S, Nykamp K, Garcia J, Lincoln SE, Topper SE: Pathogenic variant burden in the ExAC database: An empirical approach to evaluating population data for clinical variant interpretation. *Genome Med* 9: 13, 2017
24. Paterson AD, Wang KR, Lupea D, St George-Hyslop P, Pei Y: Recurrent fetal loss associated with bilineal inheritance of type 1 autosomal dominant polycystic kidney disease. *Am J Kidney Dis* 40: 16–20, 2002
25. Battle A, Brown CD, Engelhardt BE, Montgomery SB; GTEx Consortium: Genetic effects on gene expression across human tissues. *Nature* 550: 204–213, 2017
26. Douville C, Masica DL, Stenson PD, Cooper DN, Gygax DM, Kim R, et al.: Assessing the pathogenicity of insertion and deletion variants with the Variant Effect Scoring Tool (VEST-Indel). *Hum Mutat* 37: 28–35, 2016
27. Wang K, Li M, Hakonarson H: ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 38: e164, 2010
28. Harper R, Reeves B: Reporting of precision of estimates for diagnostic accuracy: A review. *BMJ* 318: 1322–1323, 1999
29. Ariyurek Y, Lantinga-van Leeuwen I, Spruit L, Ravine D, Breuning MH, Peters DJM: Large deletions in the polycystic kidney disease 1 (PKD1) gene. *Hum Mutat* 23: 99, 2004
30. de Andrade KC, Mirabello L, Stewart DR, Karlins E, Koster R, Wang M, et al.: Higher-than-expected population prevalence of potentially pathogenic germline TP53 variants in individuals unselected for cancer history. *Hum Mutat* 38: 1723–1730, 2017
31. Kim J, Field A, Schultz KAP, Hill DA, Stewart DR: The prevalence of DICER1 pathogenic variation in population databases. *Int J Cancer* 141: 2030–2036, 2017
32. Walsh R, Thomson KL, Ware JS, Funke BH, Woodley J, McGuire KJ, et al.: Exome Aggregation Consortium: Reassessment of Mendelian gene pathogenicity using 7,855 cardiomyopathy cases and 60,706 reference samples. *Genet Med* 19: 192–203, 2017
33. Alzarka B, Morizono H, Bollman JW, Kim D, Guay-Woodford LM: Design and implementation of the hepatorenal fibrocystic disease core center clinical database: A centralized resource for characterizing autosomal recessive polycystic kidney disease and other hepatorenal fibrocystic diseases. *Front Pediatr* 5: 80, 2017
34. Furu L, Onuchic LF, Gharavi A, Hou X, Esquivel EL, Nagasawa Y, et al.: Milder presentation of recessive polycystic kidney disease requires presence of amino acid substitution mutations. *J Am Soc Nephrol* 14: 2004–2014, 2003
35. Verhave JC, Bech AP, Wetzels JFM, Nijenhuis T: Hepatocyte nuclear factor 1β-associated kidney disease: More than renal cysts and diabetes. *J Am Soc Nephrol* 27: 345–353, 2016
36. Northrup H, Krueger DA; International Tuberous Sclerosis Complex Consensus Group: Tuberous sclerosis complex diagnostic criteria update: Recommendations of the 2012 Iinternational Tuberous Sclerosis Complex Consensus Conference. *Pediatr Neurol* 49: 243–254, 2013
37. O'Callaghan FJ: Tuberous sclerosis. *BMJ* 318: 1019–1020, 1999
38. Pei Y, Paterson AD, Wang KR, He N, Hefferton D, Watnick T, et al.: Bilineal disease and trans-heterozygotes in autosomal dominant polycystic kidney disease. *Am J Hum Genet* 68: 355–363, 2001
39. Bergmann C, von Bothmer J, Ortiz Brüchle N, Venghaus A, Frank V, Fehrenbach H, et al.: Mutations in multiple PKD genes may explain early and severe polycystic kidney disease. *J Am Soc Nephrol* 22: 2047–2056, 2011
40. Cornec-Le Gall E, Chebib FT, Madsen CD, Senum SR, Heyer CM, Lanpher BC, et al.: HALT Progression of Polycystic Kidney Disease Group Investigators: The value of genetic testing in polycystic kidney diseases illustrated by a family with PKD2 and COL4A1 mutations. *Am J Kidney Dis* 72: 302–308, 2018
41. Cai Y, Fedeles SV, Dong K, Anyatonwu G, Onoe T, Mitobe M, et al.: Altered trafficking and stability of polycystins underlie polycystic kidney disease. *J Clin Invest* 124: 5129–5144, 2014

This article contains supplemental material online at http://jasn.asnjournals.org/lookup/suppl/doi:10.1681/ASN.2018050493/-/DCSupplemental.