HHS Public Access

Author manuscript

Clin Genet. Author manuscript; available in PMC 2024 March 18.

Published in final edited form as:

Clin Genet. 2022 December ; 102(6): 483–493. doi:10.1111/cge.14214.

A Common Intronic Single Nucleotide Variant Modifies *PKD1* Expression Level

Zhengmao Zhang[1], **Jon Blumenfeld**[2,4], **Andrew Ramnauth**[1], **Irina Barash**[2,4], **Pengbo Zhou**[1], **Daniel Levine**[3,4], **Thomas Parker**[3,4], **Hanna Rennert**[*,1]

[1]Departments of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY

[2]Department of Medicine, Weill Cornell Medicine, New York, NY

[3]Department of Biochemistry, Weill Cornell Medicine, New York, NY

[4]The Rogosin Institute, New York, NY

Abstract

Autosomal dominant polycystic kidney disease (ADPKD), caused by mutations in *PKD1* and *PKD2* (*PKD1/2*), has unexplained phenotypic variability likely affected by environmental and other genetic factors. Approximately 10% of individuals with ADPKD phenotype have no causal mutation detected, possibly due to unrecognized risk variants of *PKD1/2*. This study was designed to identify risk variants of PKD genes through population genetic analyses.

We used Wright's F-statistics (Fst) to evaluate common single nucleotide variants (SNVs) potentially favored by positive natural selection in *PKD1* from 1000 Genomes Project (1KG) and genotyped 388 subjects from the Rogosin Institute ADPKD Data Repository. The variants with >90th percentile Fst scores underwent further investigation by *in silico* analysis and molecular genetics analyses. We identified a deep intronic SNV, rs3874648G>A, located in a conserved binding site of the splicing regulator Tra2-β in *PKD1* intron 30. Reverse-transcription PCR (RT-PCR) of peripheral blood leukocytes (PBL) from an ADPKD patient homozygous for rs3874648-A identified an atypical *PKD1* splice form. Functional analyses demonstrated that rs3874648-A allele increased Tra2-β binding affinity and activated a cryptic acceptor splice-site, causing a frameshift that introduced a premature stop codon in mRNA, thereby decreasing *PKD1* full-length transcript level. *PKD1* transcript levels were lower in PBL from rs3874648-G/A carriers than in rs3874648-G/G homozygotes in a small cohort of normal individuals and patients with *PKD2* inactivating mutations.

Our findings indicate that rs3874648G>A is a *PKD1* expression modifier attenuating *PKD1* expression through Tra2-β, while the derived G allele advantageously maintains *PKD1* expression and is predominant in all subpopulations.

[*]Correspondence to: Hanna Rennert, Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, 525 East 68 St., F544, New York, NY 10065, Phone: 212-746-6412, Fax: 212-746-8302, har2006@med.cornell.edu.

Declaration of interests

None.

Keywords

ADPKD; single nucleotide variant; hypomorphic variant; intron; next-generation sequencing

INTRODUCTION

Autosomal dominant polycystic kidney disease (ADPKD [MIM: 173900 and 613095]) is the most common inherited kidney disease with a prevalence of approximately 0.1%–0.25% of the population, causing 5%–10% of end-stage kidney disease (ESKD) worldwide [1,2]. Pathogenic mutations of *PKD1* (polycystin 1, PC1, [MIM: 601313]) and *PKD2* (polycystin 2, PC2, [MIM: 173910]) (*PKD1/2*), account for about 78% and 13% of pedigrees, respectively. [3,4] In the remaining ~10% of cases, pathogenic germline mutations of *PKD1* and *PKD2* were not identified. [3–7] There is considerable heterogeneity in the rate of disease progression, even among affected family members, despite their shared pathogenic *PKD* gene mutations. This phenotypic heterogeneity has been attributed to various factors, including unrecognized risk variants or PKD gene expression modifiers. [6–9] However, these factors remain incompletely defined.

PKD1/2 are highly polymorphic genes, with numerous sequence variants. [5] To date, 7,577 variants of *PKD1* (ENSG00000008710.20) and 1,064 variants of *PKD2* (ENSG00000118762.8) have been marked in human genomes (gnomAD v3.1.2, Accessed April 4, 2022) [10]. Mutations occur throughout *PKD1/2,* and no significant hotspot region has been identified. [5,11,12] The biological significance of many of these *PKD1/2* variants is undefined. [5,11–14]

The predominant genetic mechanism of cyst formation (cystogenesis) is a “two-hit” process involving germline and somatic inactivating mutations of *PKD1/2*. [5,12,15] The likelihood of cystogenesis increases when the functional levels of *PKD1/2* products, (i.e., polycystin 1, and polycystin 2, respectively) are below a critical threshold. [9] Reduced gene dosage was reported in orthologous ADPKD models and patients with incompletely penetrant hypomorphic alleles that reduced the level of *PKD1/2* gene expression and caused mild-to-severe ADPKD, with a phenotype that can be indistinguishable from a “two-hit” mechanism. [13,16,17]

Genes that are crucial to human fitness are thought to be under positive natural selection. [18] Mutations occur randomly and those most commonly observed are considered neutral variants, [19] remaining in genomes for generations, with the allele frequency varying by chance. [19] However, if an allele confers a phenotype with better fitness or adaptation to environmental change, then the frequency of this “advantageous allele” will increase in human populations. [18,20] The strength of natural selection can be evaluated by comparing the variation of allele frequency within and among populations. [21,22] Wright’s F-statistics (Fst) is a commonly used method to infer the action of natural selection upon human genomes [22] and to screen and identify phenotype-related variants in distinct loci of human genomes. [23,24]

PKD1/2 genes are essential to human fitness [25]. We proposed that variants in the *PKD1/2* loci have undergone positive selection during human migration and the advantageous variants have high Fst values. Given that the pathogenic variants of *PKD1* account for approximately 80% of ADPKD pedigrees, in the current study, we preferentially evaluated the Fst values of common single nucleotide variants (SNVs) in *PKD1* across subpopulations (African, Eurasian, and American) from the 1000 Genomes Project (1KG) [26]. Of the top 10% variants with the highest Fst values, we identified a novel deep intronic variant functioning as a *PKD1* expression modifier.

MATERIALS AND METHODS

Study Subjects

Study subjects were participants in the Rogosin Institute ADPKD Data Repository, a single-center, longitudinal study of genotype and phenotype characteristics of individuals with ADPKD (http://clinicaltrials.gov identifier: NCT00792155). All subjects provided written informed consent. The normal individuals were healthy blood donor volunteers as previously described. [27] The studies were approved by the Institutional Review Board Committees at Weill Cornell Medicine (WCM) and Rockefeller University, respectively (New York, NY).

PKD1/2 Gene Testing for Pathogenic Mutations

Genomic DNA from 388 ADPKD subjects was analyzed for *PKD1* and *PKD2* mutations at Athena Diagnostics, Inc. (Worcester, MA), or Weill Cornell Medicine (WCM) New York, NY, using long range-PCR (LR-PCR) Sanger sequencing and/or next-generation sequencing (NGS). [28] Mutation-negative patients were further tested by multiplex ligation-dependent probe amplification (MLPA) for copy number variation by Prevention Genetics (Marshfield, WI).

Population Genetic and *In-silico* Analyses

Population genetic analyses were performed on 2,403 variants of *PKD1* (ENSG00000008710) from 2,504 individuals of 26 subpopulations from Africa, East Asia, Europe, South Asia, and the Americas of the 1KG (phase-3, GRCh37/hg19) [26]. The common variants (minor allele frequency >0.01) were subjected to Fst analysis using PLINK. [29] Variants with high Fst-values were further evaluated by Combined Annotation-Dependent Depletion (CADD) tool with 99% confidence interval (CI). [21,22,25,30]. The effects of intronic variants on binding affinity of RNA-splicing factors/regulators were predicted by using RBPmap with a high stringency setting. [31] The binding protein profiles were cross-checked using Human Splicing Finder V3.1 (HSF) [32] and the predicted splice donor and acceptor sites were further evaluated by MaxEntScan. [33]

Minigene Splicing Assay

PKD1 exons 30 to 34 (2.6Kb) were amplified from peripheral blood leukocytes (PBL) DNA of ADPKD patients and normal controls (Supplemental Table 1). [28] PCR products were subcloned into the pcDNA3-Flag mammalian expression vector. *PKD1* rs3874648G>A changes in the rescue assay were introduced by directed mutagenesis. *PKD1*-minigene constructs were transiently transfected into Human Embryonic Kidney (HEK293T) cells

using Lipofectamine 2000™ Reagent as described in the Supplemental Methods. RNA splicing was detected by RT-PCR using minigene-specific primers (Supplemental Table 1). Testing was performed twice in triplicate. All constructs and splicing products were verified by Sanger sequencing. Recombinant DNA manipulations were done following WCM Biological Safety Program DNA Recombination regulations.

RT-PCR assay

Total RNA was extracted from PBL pellets or transfected HEK293T cells using TRIzol Reagent and subjected to first-strand cDNA synthesis using Applied Biosystems™ High-Capacity cDNA Reverse-Transcription Kit following the manufacturer's instructions (ThermoFisher). Amplification of *PKD1* SNV-containing and minigene splicing products was performed in the PrimeSTAR GXL system (Takara Bio), containing 10% Dimethyl sulfoxide and 500mM Betaine on a Biometra TRIO Thermocycler. Amplicons were analyzed by agarose-gel electrophoresis with ethidium-bromide staining.

Quantitative RT-PCR

Complementary DNA was synthesized from 1μg PBL RNA using random primers and the High Capacity cDNA Archive Kit (ThermoFisher). [34] *PKD1*expression level relative to the housekeeping gene Ribonuclease P Protein Subunit p30 (*RPP30*) was quantified using *PKD1*-specific primers (Supplemental Table 1) and the ABI 7500 Real-Time thermocycler. Data analysis was performed by the comparative threshold cycle (*Ct*) method using the SDS v2.2.2 software (ThermoFisher) as previously described. [34] Experiments were performed in duplicate, and each sample was analyzed in triplicate.

TaqMan allelic discrimination assay

PKD1:c.10051-239G>A (rs3874648:G>A) SNP genotyping was performed using a ThermoFisher Custom TaqMan® Assay as described in the Supplemental Material.

RNA pull-down assay

Pull-down bait RNAs were transcribed *in vitro* in RNA Synthesis system (E2070) from DNA fragments containing variants of interest. Amplified-bait RNAs were purified using RNeasy kits (QIAGEN), biotinylated by T4 RNA ligase and coupled with Streptavidin-Magnetic Beads (Pierce™). Bait-RNA-magnetic beads were incubated with HEK293T cell nuclear extracts in the presence of protease inhibitors, washed with incubating buffer, and suspended in 1X Laemmli Sample Buffer, then heated at 100°C for 10min, and stored at −80°C (Supplemental Figure 1). The Bait-RNA-magnetic-beads-bound proteins were detected by Western blot with rabbit anti-Tra2-β (GeneTex, GTX114752) or rabbit anti-SNRPA (Small Nuclear Ribonucleoprotein Polypeptide A) (GeneTex, GTX101664) primary antibody, and fluorescence-conjugated secondary antibodies.

Statistical analysis

An unpaired t-test was used to compare the difference in the relative mRNA expression level analyzed between patients and normal individuals. A two-sided p-value <0.05 was considered significant. [34] One-way analysis of variance and contingency analysis were

used to compare the differences in variables among the -GA variants (JMP Pro Statistical Software16.1.0, Cary, NC).

RESULTS

Identification of putative functional *PKD1* variants

The 211 common variants from 2,403 *PKD1* SNVs of 2,504 individuals were subjected to Fst scoring to quantify the selection pressure upon the *PKD1* locus (Figure 1A). [21,22,29] Of the 22 variants ranking in the upper 10% Fst scores, ten have been classified as "Benign" and "Likely Neutral" in ClinVar [35] and Mayo ADPKD Mutation databases [36], respectively, while the remaining variants were unclassified (Table 1). We then performed MSC-associated CADD evaluation on these unclassified variants to prioritize the variants for further analysis (Supplemental Table 2). [25,30] Of these, four were excluded since they were located either in 3'-UTR or in the low-complexity genomic regions. [10] (Figure 1B and Supplemental Table 2). The remaining variants, located in introns, were evaluated using RBPmap for RNA splicing factor/regulator profile analysis. [31] The results indicated that variants rs7206195C>T, rs7185040A>C and rs11861948A>G did not alter while rs116189075A>G, rs114796022T>C, rs12926737T>C, and rs58999880G>T only slightly altered RNA-binding protein affinity to these sites (Supplemental Table 3). By contrast, variant rs3874648G>A (*PKD1* c.10051-239G>A) dramatically enhanced SRSF10 (human/mouse) and Tra2 (drosophila) binding to intron 30 (Supplemental Table 4). Consistently, HSF analysis confirmed that substitution of G to A either strengthens the binding affinity of Transformer 2 Beta (Tra2-β) by 13%~42%, alters the 9G8 binding site to become a Tra2-β-preferred binding site, or creates a new binding site for Tra2-β (Table 2). HSF also showed that binding affinities of other splicing enhancers or silencers to this flanking region are unaffected (Supplemental Tables 5 and 6).

RNA expression studies and qRT-PCR analysis

To identify the RNA expression profile associated with these variants, we performed RT-PCR analysis in PBL RNAs. RT-PCR analysis with primer set (PKD1-Ex30-F1/PKD1-Ex31-R1) demonstrated a higher molecular weight extra band (~270bp) in PBLs of the ADPKD patient homozygous for the rs3874648-A allele that was not detectable in the normal control carrying the rs3874648-G alleles (Figure 2A). Sequence analysis of this distinct band demonstrated an insertion of 41bp of intronic sequence resulting in premature termination of translation at codon K3350 (*PKD1* p.V3351X) (Figure 2B and 2C). No abnormal splicing events were detected in the remaining intronic variants (rs116189075A>G, rs114796022T>C, rs11861948A>G, and rs11862600A>G) of the same RNA sample (Supplemental Figure 2).

In normal individuals, the qRT-PCR analysis demonstrated that *PKD1* mRNA levels were ~10% ($p<0.05$) lower in PBLs heterozygous for rs3874648G>A compared to PBLs bearing rs3874648-G/G alleles (Figure 2D). This observation was consistent with the Genotype-Tissue Expression survey (see Discussion). [37] To determine whether variant rs3874648G>A affected *PKD1* expression in ADPKD patients, we evaluated *PKD1* mRNA levels by qRT-PCR in *PKD2*-mutation-positive patients. *PKD1* expression levels in *PKD2* mutation-

positive patients homozygous for rs3874648-G were significantly lower (~70%) than in normal subjects carrying the G/G genotype (Figure 2E). Furthermore, in these ADPKD patients, *PKD1* mRNA levels were ~30% (p<0.05) lower in PBLs bearing the rs3874648-G/A variant than in rs3874648-G homozygotes. No PBL was available from rs3874648-A/A normal controls or ADPKD patients with a *PKD2* mutation for this analysis.

Functional characterization of *PKD1* rs3874648G>A

Computational analysis of *PKD1* intron 30 demonstrated the presence of a putative cryptic acceptor splice-site (gagcagGT) distal from the authentic acceptor site of exon 31 with a slightly lower consensus value compared to the canonical acceptor site (ctgcagGT) (87.07 vs 93.71 by HSF and 7.51 vs 11.85 by MaxENTScan) (Table 3 and Supplemental Table 7, respectively). [32] To determine the functional effect of rs387464G>A on *PKD1* RNA cryptic splicing, we conducted a minigene splicing assay, spanning exon 30 to intron 33 (Figure 3A). RT-PCR analysis of the minigene transcripts in HEK293T cells revealed that the pcDNA3.1-rs3874648-A construct generated a distinct larger size band of ~750bp compared to the product size (~700bp) obtained with the pcDNA3.1-rs3874648-G allele (Figure 3B). A confirmation assay substituting the G>A in pcDNA3.1-rs3874648-G (pcDNA3.1-rs3874648-G>A) and A>G in pcDNA3.1-rs3874648-A (pcDNA3.1-rs3874648-A>G) affirmed that rs3874648G>A was the only SNV involved in regulating intron 30 cryptic splicing event (Figure 3B, Supplemental Figure 3).

The rs3874648G>A variant is located in a conserved adenine-rich binding site of Tra2-β, a critical RNA splicing regulator protein [38,39], therefore, we performed RNA-protein-pulldown in HEK293T cell nuclear extracts to evaluate the interaction between Tra2-β and rs3874648:G>A variant using RNA baits (Figure 3C and Supplemental Figure 1). Compared to the motif containing the rs3874648-G allele, Tra2-β binding to rs3874648-A containing-motif was 4-fold stronger (Figure 3D and 3E, Figure 4). By contrast, the affinity of SNRPA to its binding motifs in the proximal region of the RNA baits was comparable for both variants, indicating a specific effect of Tra2-β (Figure 3D).

Assessing variant prevalence in ADPKD cohort

Our results indicate that the rs3874648-A allele increased the binding affinity of Tra2-β to *PKD1* intron 30 leading to lower levels of *PKD1* mRNA by partial intron retention. We assessed the prevalence of rs3874648G>A in our ADPKD cohort by allelic discriminating assay. Overall, the frequency of the rs3874648-A allele was 0.1095 in the ADPKD cohort (Supplemental Table 8). In total, we identified eight homozygotes for the rs3874648-A allele. Of these, three carried pathogenic mutations in *PKD1* and two patients had pathogenic mutations in *PKD2*. In the other three subjects, no pathogenic mutations in *PKD1 or PKD2* were identified by LR-PCR-NGS and MLPA. Although the rs3874648-A allele distributes normally in the ADPKD cohort according to Hardy-Weinberg Equilibrium, the prevalence of rs3874648-A homozygotes in the patients without a detectable *PKD1/2* mutation was slightly but significantly higher than predicted (p=0.009) (Supplemental Table 9). In a multivariate analysis, we analyzed the variables that have a strong impact on disease severity (ht-TKV, GFR, or hypertension). We found no association between rs3874648-G/A and patient characteristics known to be associated with ADPKD severity, e.g, total kidney

volume, Mayo Clinic Classification, age of onset of ESKD, or hypertension (Supplemental Figure 4).

DISCUSSION

Many SNVs in *PKD1/2* genes have been previously reported, but their pathogenicity, penetrance, and molecular mechanisms are incompletely defined. [5] In this study, we identified a common *PKD1* variant, rs3874648G>A, which can alter the expression of full-length *PKD1* by modifying its binding affinity to splicing factor Tra2-β and activating a cryptic splicing site leading to *PKD1* premature termination. To our knowledge, our report is the first to show a significant effect of a common SNV on *PKD1* expression in ADPKD patients by using a combination of population genomics and molecular genetics analyses. There was no statistically significant association between rs3874648-G/A variant type and clinical characteristics or Mayo Classification. However, the number of rs3874648-A homozygotes in our cohort was very small to allow the drawing of meaningful conclusions.

rs3874648G>A is located in a conserved Tra2-β binding site in *PKD1* intron 30. Using a minigene functional assay, we showed that rs3874648-A disrupted normal splicing by engaging an adjacent cryptic acceptor splice site predicted to truncate the *PKD1* protein, thereby decreasing levels of full-length *PKD1* transcript. This adverse effect of the -A allele was “rescued” by substituting it with the -G allele, thereby restoring normal *PKD1* mRNA transcript levels. Accordingly, *PKD1* transcript levels were ~10% lower in PBL from rs3874648-G/A carriers compared to rs3874648-G/G carriers in a small cohort of normal individuals. Moreover, in our ADPKD cohort with a *PKD2* mutation, *PKD1* expression was ~30% lower in rs3874648-G/A carriers than in rs3874648-G/G carriers. Notably, *PKD1* expression levels were ~70% lower in ADPKD patients having *PKD2* mutations who also had the rs3874648-G/G genotype compared to non-ADPKD controls with this variant. This may reflect interdependent regulation of *PKD1*/2, previously reported in human and mouse cell lines, and attributed to *PKD2* depletion-induced anomalies in *PKD1* (post-)transcription regulations. [40,41]

The comparative genomics analysis revealed that the allele rs3874648-A was shared in primates (Supplemental Figure 5) and Neanderthals and Denisovans (Supplemental Figure 6). Compared with the rs3874648-A allele, the derived rs3874648-G-allele in *PKD1* has a lower binding affinity to Tra2-β, enabling the latter to maintain higher *PKD1* transcript levels. These results are consistent with the Genotype-Tissue Expression (GTEx) dataset of a large cohort of normal participants that reported an association of rs3874648-G with higher *PKD1* transcript levels in numerous tissue types. [37] Compared to rs3874648-A carriers, the mean *PKD1* mRNA level in those with the rs3874648-G/G genotype was 33% higher in kidneys and 18% higher in non-kidney tissues (accessed on Oct. 01, 2021) (Supplemental Figure 7). [37]

The association of higher *PKD1* expression levels with rs3874648-G was significant in tissues with relatively high expression levels of Tra2-β. [37] Tra2-β is also known as splicing factor arginine/serine (RS)-rich 10 (SFRS10), a transformer 2 (Tra2) homolog in *Drosophila*, and one of the strongest splicing regulators found in RS-rich protein family

members [38], regulating alternative splice-site selection in a dosage-dependent manner. [38,39] A comprehensive transcriptome-wide Tra2-β binding site analysis in human cells indicated that Tra2-β preferentially binds to adenine-rich motifs. [42] Increased expression levels of Tra2-β were reported during oxidative stress [43], which can be magnified by hyperglycemia in diabetes, angiotensin II in hypertension, and hypoxia in ischemia injury. [44–46] In these situations, elevated Tra2-β protein levels might further decrease *PKD1* expression in carriers of the rs3874648-A allele. Accordingly, reduced functioning renal mass due to acute or chronic kidney disease, or a hypomorphic *PKD1/2* allele might further reduce PC1 production through the previously proposed "third-hit" mechanism [47] and accelerate the progression of ADPKD in those with the rs3874648-A-allele. Altogether, these data support the hypothesis that lower *PKD1* transcript levels associated with the rs3874648-A-allele could increase the risk for ADPKD by maintaining the dosage of PC1 at or below a critical threshold in susceptible individuals. [9,13,14,17].

The common variant rs3874648 is present across all populations with an overall prevalence of 0.189 in 1KG and 0.194 in gnomAD (Supplemental Table 10 and 11). [10] The prevalence of the *PKD1* rs3874648-A variant in our cohort was 0.1095, in agreement with gnomAD for Caucasians in Europe (0.1094), but lower than reported in African/African Americans (0.5024). This likely reflects the higher prevalence of Caucasian subjects in our cohort (74%). Because of the disproportionately high prevalence of the rs3874648-A variant, a higher rate of ADPKD in the African population is anticipated. Although the prevalence of ADPKD in the African population is likely to be underdiagnosed, [48] in California, the crude prevalence of ADPKD for Blacks was 73.0, non-Hispanic Whites 63.2, Hispanics 39.9, and Asian/Pacific Islanders 48.9 per 100,000 ($P<0.001$) [49]. However, our study was not designed to assess the association of this variant with the prevalence of ADPKD in various populations.

Genetic variants that affect RNA splicing have been reported in PKD genes. [50] Previously, we characterized a novel missense mutation in *PKD2* (p.L441CfsX4) that abolished a conserved acceptor splice site of intron 5, resulting in premature translation termination, to a lesser magnitude, in PBLs from other ADPKD patients as well as normal individuals. [27] Even under physiological conditions, two long polypyrimidine regions in human *PKD1* introns 21 and 22 were associated with abnormal splicing across these introns and early transcript termination, potentially reducing *PKD1* expression levels below the "cystogenic threshold". [51] Xie et al. have recently identified eight rare *PKD1* intronic variants extracted from the Mayo Clinic ADPKD and ClinVar databases shown to alter RNA splicing by minigene assays.[52]

Strengths of this study include the use of well-established LR-PCR-NGS techniques with high sensitivity and call rates, and the recruitment of ADPKD patients with standard selection criteria in a single health system. Results from population and molecular genetics strongly support the functional significance of the rs3874648G>A variant in regulating *PKD1* expression.

A limitation of this study was the small cohort (n=8) homozygous for the rs3874648-A-allele available for clinical traits association and functional studies (Supplemental Table 12).

PBLs were available from only two ADPKD patients with the rs3874648-G/A variant, and from one (MC9-002572) who was homozygous for rs3874648-A. No mutation of *PKD1/2* was identified in the latter patient, who demonstrated a ~40% lower *PKD1* transcript level in PBL, likely reflecting either an undetected mutation in *PKD1/2* or another gene causing an ADPKD phenotype (Supplemental Figure 8). Nonetheless, we were able to identify a higher-than-expected prevalence of rs3874648-A-allele homozygotes in the *PKD1/2* mutation-negative cohort and demonstrate the functional significance of the rs3874648-A-allele and the pathogenic mechanism of its action. Additional studies are required to clarify the impact of this variant on ADPKD phenotype including in patients with no identified exonic *PKD1/2* mutations.

In summary, through population genomics and functional analyses, we identified a common deep intronic variant rs3874648G>A in *PKD1* that acts as a gene expression modifier by increasing the binding affinity of splicing regulator Tra2-β, activating a cryptic splicing site, and reducing levels of full-length *PKD1* mRNA by disrupting normal splicing. Molecular genetics analysis supports the derived rs3874648-G allele as the advantageous allele predominating in all subpopulations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We are grateful to all patients and their families for their invaluable participation. We thank Dr. Yanlin Wang in the Department of Medicine, University of Connecticut Health Center for sharing the pcDNA3.1-Flag mammalian expression vector. Research reported in this publication was supported by the National Center for Advancing Translational Science of the National Institute of Health under award number UL1TR002384.

REFERENCES

1. Paul BM, Vanden Heuvel GB. Kidney: polycystic kidney disease. Wiley Interdiscip Rev Dev Biol. Nov-Dec 2014;3(6):465–87. doi:10.1002/wdev.152 [PubMed: 25186187]
2. Chapman AB, Devuyst O, Eckardt KU, et al. Autosomal-dominant polycystic kidney disease (ADPKD): executive summary from a Kidney Disease: Improving Global Outcomes (KDIGO) Controversies Conference. Kidney Int. Jul 2015;88(1):17–27. doi:10.1038/ki.2015.59 [PubMed: 25786098]
3. Harris PC, Torres VE. Genetic mechanisms and signaling pathways in autosomal dominant polycystic kidney disease. J Clin Invest. Jun 2014;124(6):2315–24. doi:10.1172/JCI72272 [PubMed: 24892705]
4. Cornec-Le Gall E, Audrezet MP, Le Meur Y, Chen JM, Ferec C. Genetics and pathogenesis of autosomal dominant polycystic kidney disease: 20 years on. Hum Mutat. Dec 2014;35(12):1393–406. doi:10.1002/humu.22708 [PubMed: 25263802]
5. Tan AY, Zhang T, Michaeel A, et al. Somatic Mutations in Renal Cyst Epithelium in Autosomal Dominant Polycystic Kidney Disease. J Am Soc Nephrol. Aug 2018;29(8):2139–2156. doi:10.1681/ASN.2017080878 [PubMed: 30042192]
6. Lanktree MB, Haghighi A, di Bari I, Song X, Pei Y. Insights into Autosomal Dominant Polycystic Kidney Disease from Genetic Studies. Clin J Am Soc Nephrol. Jul 20 2020;doi:10.2215/CJN.02320220

7. Senum SR, Li YSM, Benson KA, et al. Monoallelic IFT140 pathogenic variants are an important cause of the autosomal dominant polycystic kidney-spectrum phenotype. Am J Hum Genet. Jan 6 2022;109(1):136–156. doi:10.1016/j.ajhg.2021.11.016 [PubMed: 34890546]
8. Rossetti S, Harris PC. Genotype-phenotype correlations in autosomal dominant and autosomal recessive polycystic kidney disease. J Am Soc Nephrol. May 2007;18(5):1374–80. doi:10.1681/ASN.2007010125 [PubMed: 17429049]
9. Bergmann C, Guay-Woodford LM, Harris PC, Horie S, Peters DJM, Torres VE. Polycystic kidney disease. Nat Rev Dis Primers. Dec 6 2018;4(1):50. doi:10.1038/s41572-018-0047-y [PubMed: 30523303]
10. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. May 2020;581(7809):434–443. doi:10.1038/s41586-020-2308-7 [PubMed: 32461654]
11. Rossetti S, Strmecki L, Gamble V, et al. Mutation analysis of the entire PKD1 gene: genetic and diagnostic implications. Am J Hum Genet. Jan 2001;68(1):46–63. doi:10.1086/316939 [PubMed: 11115377]
12. Zhang Z, Bai H, Blumenfeld J, et al. Detection of PKD1 and PKD2 Somatic Variants in Autosomal Dominant Polycystic Kidney Cyst Epithelial Cells by Whole-Genome Sequencing. J Am Soc Nephrol. Oct 29 2021;doi:10.1681/ASN.2021050690
13. Rossetti S, Kubly VJ, Consugar MB, et al. Incompletely penetrant PKD1 alleles suggest a role for gene dosage in cyst initiation in polycystic kidney disease. Kidney Int. Apr 2009;75(8):848–55. doi:10.1038/ki.2008.686 [PubMed: 19165178]
14. Vujic M, Heyer CM, Ars E, et al. Incompletely penetrant PKD1 alleles mimic the renal manifestations of ARPKD. J Am Soc Nephrol. Jul 2010;21(7):1097–102. doi:10.1681/ASN.2009101070 [PubMed: 20558538]
15. Watnick TJ, Torres VE, Gandolph MA, et al. Somatic mutation in individual liver cysts supports a two-hit model of cystogenesis in autosomal dominant polycystic kidney disease. Mol Cell. Aug 1998;2(2):247–51. doi:10.1016/s1097-2765(00)80135-5 [PubMed: 9734362]
16. Lantinga-van Leeuwen IS, Dauwerse JG, Baelde HJ, et al. Lowering of Pkd1 expression is sufficient to cause polycystic kidney disease. Hum Mol Genet. Dec 15 2004;13(24):3069–77. doi:10.1093/hmg/ddh336 [PubMed: 15496422]
17. Hopp K, Ward CJ, Hommerding CJ, et al. Functional polycystin-1 dosage governs autosomal dominant polycystic kidney disease severity. J Clin Invest. Nov 2012;122(11):4257–73. doi:10.1172/JCI64313 [PubMed: 23064367]
18. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. Natural selection has driven population differentiation in modern humans. Nat Genet. Mar 2008;40(3):340–5. doi:10.1038/ng.78 [PubMed: 18246066]
19. Kern AD, Hahn MW. The Neutral Theory in Light of Natural Selection. Mol Biol Evol. Jun 1 2018;35(6):1366–1371. doi:10.1093/molbev/msy092 [PubMed: 29722831]
20. Sabeti PC, Schaffner SF, Fry B, et al. Positive natural selection in the human lineage. Science. Jun 16 2006;312(5780):1614–20. doi:10.1126/science.1124309 [PubMed: 16778047]
21. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev Genet. 2013;47:97–120. doi:10.1146/annurev-genet-111212-133526 [PubMed: 24274750]
22. Holsinger KE, Weir BS. Genetics in geographically structured populations: defining, estimating and interpreting F(ST). Nat Rev Genet. Sep 2009;10(9):639–50. doi:10.1038/nrg2611 [PubMed: 19687804]
23. Hamblin MT, Di Rienzo A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am J Hum Genet. May 2000;66(5):1669–79. doi:10.1086/302879 [PubMed: 10762551]
24. Bersaglieri T, Sabeti PC, Patterson N, et al. Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet. Jun 2004;74(6):1111–20. doi:10.1086/421051 [PubMed: 15114531]
25. Itan Y, Shang L, Boisson B, et al. The mutation significance cutoff: gene-level thresholds for variant predictions. Nat Methods. Feb 2016;13(2):109–10. doi:10.1038/nmeth.3739 [PubMed: 26820543]

26. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. Nature. Oct 1 2015;526(7571):75–81. doi:10.1038/nature15394 [PubMed: 26432246]
27. Tan YC, Blumenfeld J, Michaeel A, et al. Aberrant PKD2 splicing due to a presumed novel missense mutation in autosomal-dominant polycystic kidney disease. Clin Genet. Sep 2011;80(3):287–92. doi:10.1111/j.1399-0004.2010.01555.x [PubMed: 20950398]
28. Tan AY, Michaeel A, Liu G, et al. Molecular diagnosis of autosomal dominant polycystic kidney disease using next-generation sequencing. J Mol Diagn. Mar 2014;16(2):216–28. doi:10.1016/j.jmoldx.2013.10.005 [PubMed: 24374109]
29. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. Sep 2007;81(3):559–75. doi:10.1086/519795 [PubMed: 17701901]
30. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic acids research. Jan 8 2019;47(D1):D886–D894. doi:10.1093/nar/gky1016 [PubMed: 30371827]
31. Paz I, Kosti I, Ares M Jr., Cline M, Mandel-Gutfreund Y. RBPmap: a web server for mapping binding sites of RNA-binding proteins. Nucleic acids research. Jul 2014;42(Web Server issue):W361–7. doi:10.1093/nar/gku406 [PubMed: 24829458]
32. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic acids research. May 2009;37(9):e67. doi:10.1093/nar/gkp215 [PubMed: 19339519]
33. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2–3):377–94. doi:10.1089/1066527041410418 [PubMed: 15285897]
34. Rennert H, Sadowl C, Edwards J, et al. An alternative spliced RNASEL variant in peripheral blood leukocytes. J Interferon Cytokine Res. Nov 2006;26(11):820–6. doi:10.1089/jir.2006.26.820 [PubMed: 17115900]
35. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research. Jan 2014;42(Database issue):D980–5. doi:10.1093/nar/gkt1113 [PubMed: 24234437]
36. 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. Jul 2007;28(7):654–9. doi:10.1002/humu.20474 [PubMed: 17370309]
37. Consortium GT. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. Sep 11 2020;369(6509):1318–1330. doi:10.1126/science.aaz1776 [PubMed: 32913098]
38. Sciabica KS, Hertel KJ. The splicing regulators Tra and Tra2 are unusually potent activators of pre-mRNA splicing. Nucleic acids research. 2006;34(22):6612–20. doi:10.1093/nar/gkl984 [PubMed: 17135210]
39. Chen YC, Chang JG, Jong YJ, Liu TY, Yuo CY. High expression level of Tra2-beta1 is responsible for increased SMN2 exon 7 inclusion in the testis of SMA mice. PLoS One. 2015;10(3):e0120721. doi:10.1371/journal.pone.0120721 [PubMed: 25781985]
40. Cruz NM, Song X, Czerniecki SM, et al. Organoid cystogenesis reveals a critical role of microenvironment in human polycystic kidney disease. Nat Mater. Nov 2017;16(11):1112–1119. doi:10.1038/nmat4994 [PubMed: 28967916]
41. Friedrich S, Muller H, Riesterer C, et al. Identification of pathological transcription in autosomal dominant polycystic kidney disease epithelia. Sci Rep. Jul 23 2021;11(1):15139. doi:10.1038/s41598-021-94442-8 [PubMed: 34301992]
42. Fontrodona N, Aube F, Claude JB, et al. Interplay between coding and exonic splicing regulatory sequences. Genome Res. May 2019;29(5):711–722. doi:10.1101/gr.241315.118 [PubMed: 30962178]
43. Kajita K, Kuwano Y, Kitamura N, et al. Ets1 and heat shock factor 1 regulate transcription of the Transformer 2beta gene in human colon cancer cells. J Gastroenterol. Nov 2013;48(11):1222–33. doi:10.1007/s00535-012-0745-2 [PubMed: 23361474]

44. Yan LJ. Pathogenesis of chronic hyperglycemia: from reductive stress to oxidative stress. J Diabetes Res. 2014;2014:137919. doi:10.1155/2014/137919 [PubMed: 25019091]
45. Wright E Jr., Scism-Bacon JL, Glass LC. Oxidative stress in type 2 diabetes: the role of fasting and postprandial glycaemia. Int J Clin Pract. Mar 2006;60(3):308–14. doi:10.1111/j.1368-5031.2006.00825.x [PubMed: 16494646]
46. Forrester SJ, Kikuchi DS, Hernandes MS, Xu Q, Griendling KK. Reactive Oxygen Species in Metabolic and Inflammatory Signaling. Circ Res. Mar 16 2018;122(6):877–902. doi:10.1161/CIRCRESAHA.117.311401 [PubMed: 29700084]
47. Takakura A, Contrino L, Zhou X, et al. Renal injury is a third hit promoting rapid development of adult polycystic kidney disease. Hum Mol Genet. Jul 15 2009;18(14):2523–31. doi:10.1093/hmg/ddp147 [PubMed: 19342421]
48. Arogundade FA, Akinbodewa AA, Sanusi AA, Okunola O, Hassan MO, Akinsola A. Clinical presentation and outcome of autosomal dominant polycystic kidney disease in Nigeria. Afr Health Sci. Sep 2018;18(3):671–680. doi:10.4314/ahs.v18i3.25 [PubMed: 30603000]
49. Aung TT, Bhandari SK, Chen Q, et al. Autosomal Dominant Polycystic Kidney Disease Prevalence among a Racially Diverse United States Population, 2002 through 2018. Kidney360. Dec 30 2021;2(12):2010–2015. doi:10.34067/KID.0004522021 [PubMed: 35419536]
50. Wang K, Zhao X, Chan S, et al. Evidence for pathogenicity of atypical splice mutations in autosomal dominant polycystic kidney disease. Clin J Am Soc Nephrol. Feb 2009;4(2):442–9. doi:10.2215/CJN.00980208 [PubMed: 19158373]
51. Lea WA, Parnell SC, Wallace DP, et al. Human-Specific Abnormal Alternative Splicing of Wild-Type PKD1 Induces Premature Termination of Polycystin-1. J Am Soc Nephrol. Oct 2018;29(10):2482–2492. doi:10.1681/ASN.2018040442 [PubMed: 30185468]
52. Xie S, Leng X, Tao D, et al. Identification of novel single-nucleotide variants altering RNA splicing of PKD1 and PKD2. J Hum Genet. Jan 2022;67(1):27–34. doi:10.1038/s10038-021-00959-1 [PubMed: 34257392]

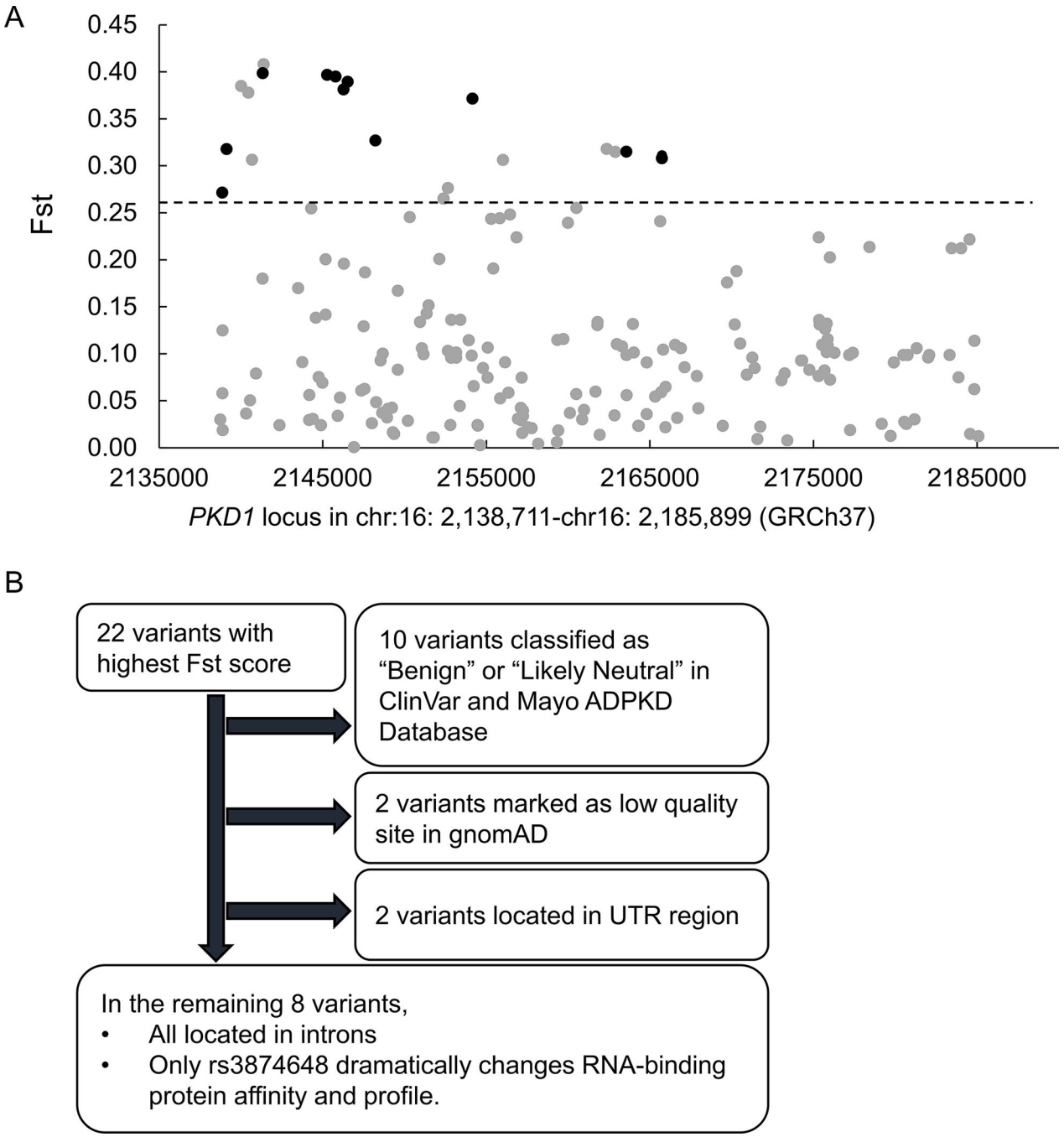

Figure 1. Evaluation of common variants in *PKD1*.

(A) 211 common SNPs within *PKD1* locus (47kb in GRCh37 from chr:16: 2,138,711 to chr16: 2,185,899) were subjected to Wright's F-statistics (Fst) analyses. Of the top 10% variants (n=22) with the highest Fst scores (range: 0.26 to 0.41), twelve variants, shown in dark circles, were unclassified in the ClinVar and Mayo ADPKD Mutation database. These variants were more likely to be subjected to positive natural selection. (B) Analysis workflow of the variants with the highest Fst scores.

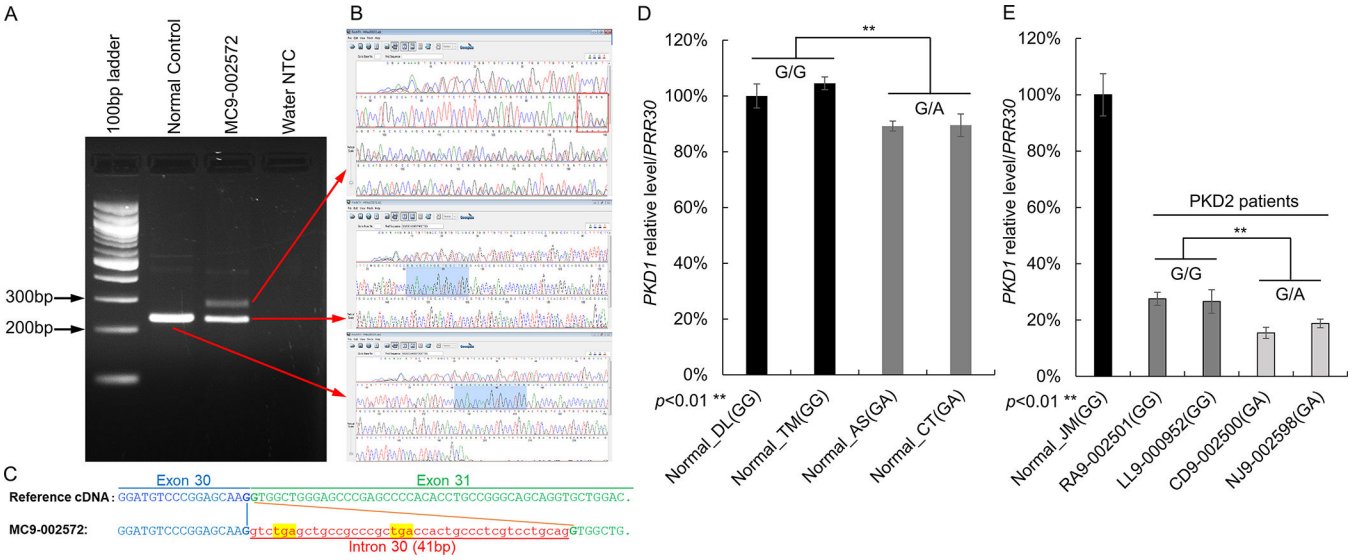

Figure 2. RNA expression and qRT-PCR analysis of intron 30 in *PKD1*.

(**A**) RT-PCR analysis of RNA PBL with primer set (PKD1-Ex30-F1/PKD1-Ex31-R1) demonstrated a higher weight extra band (~270bp) in the ADPKD patient homozygous for rs3874648-A that was not detectable in PBL from the normal control rs3874648-GG. (**B, C**) Sequence analysis of the abnormal product (upper panel) demonstrated an insertion of 41bp fragment of the 3'end of intron 30 in the mature mRNA of *PKD1,* resulting in a frameshift at codon K3350 of the translated product (*PKD1* p.V3351X). (**D**) *PKD1* mRNA expression in PBL from normal individuals and ADPKD patients was measured by quantitative real-time RT-PCR (qRT-PCR). *PKD1* expression levels, relative to the *RPP30* housekeeping gene, were determined by comparing Ct values as detailed in the Methods section. In normal individuals, *PKD1* transcript levels decreased by ~10% ($p<0.05$) in PBL cells (AS and CT) heterozygous for the rs3874648-G/A variant compared with the levels observed in cells (DL and TM) homozygous for rs3874648-G. (**E**) *PKD1* mRNA levels in PBL from ADPKD patients with truncating *PKD2* mutation were ~30% ($p<0.05$) lower in PBL cells bearing the rs3874648-G/A (CD-9 and NJ-9) variant compared to cells with the rs3874648-G/G (RA-9 and LL-9) variant. PBL, peripheral blood lymphocytes.

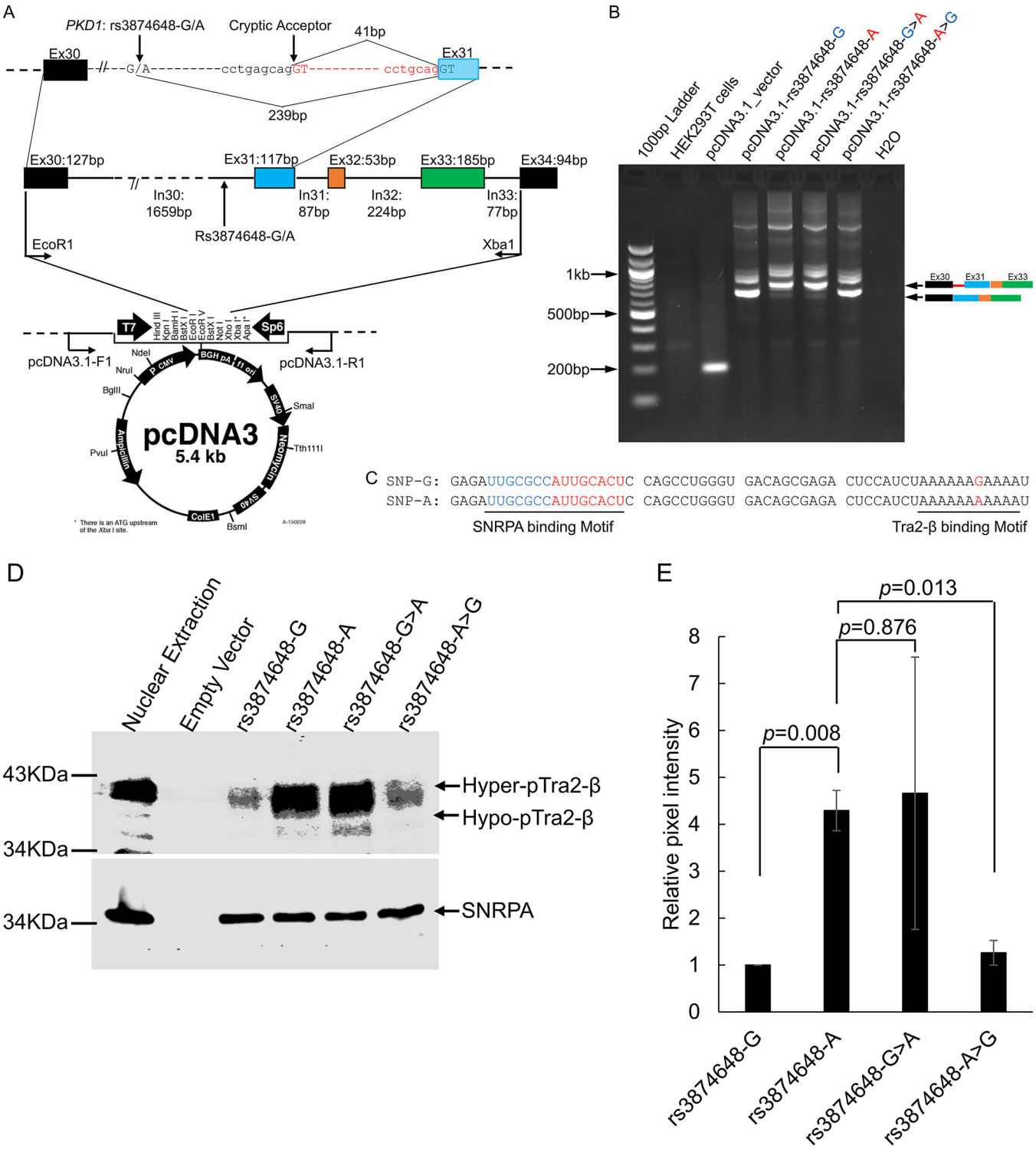

Figure 3. Functional analysis of PKD1 rs387468-G/A variants using minigene assay and RNA-Pulldown assay.

(**A**) Schematic representation of the *PKD1* minigene constructs used in this study. *PKD1* exon 30-intron 33 PCR products from a normal individual and ADPKD patient (MC2572) were inserted between EcoRI and XbaI site of pcDNA3.1 vector. (**B**) Detection of *PKD1*

exon30-intron33 minigenes by RT-PCR. The splicing products of the minigene carrying the rs3874648-A allele were larger (~750bp) compared to the product size obtained for the minigene carrying the rs3874648-G allele (~700bp). The point mutation rescue assays showed that pcDNA3.1-rs3874648A>G mimicked the splicing events occurring in the cells transfected with pcDNA3.1-rs3874648-G and *vice versa*. All the *PKD1* minigene constructs and transcripts were confirmed by Sanger sequencing. (**C**). The interaction between variant-containing RNA and Tra2-β was verified and measured by RNA-pulldown assay in nuclear extracts of HEK293T cells, a transformed non-cancer human kidney cell line with a high Tra2-β protein level. The schematic diagram shows two-tandem SNRPA binding motifs located in the proximal upstream region within a 30-nucleotide interval of rs3874648G>A. (**D**). The RNA-binding proteins were detected by Western blotting using anti-Tra2-β and anti-SNRPA antibodies. The alleles rs3874648-G and rs3874648-A exhibited distinct binding affinities to Tra2-β, whereas the binding between SNRPA and its motif was not affected. Phosphorylation of Tra2-β is required for its activity and the multiple bands of Tra2-β are indicative of the hypo- and hyper-phosphorylation forms. (**E**). The binding affinity was measured by comparing fluorescent intensity between blots. The rs3874648-A allele had a ~4-times higher binding affinity compared with the rs3874648-G allele.

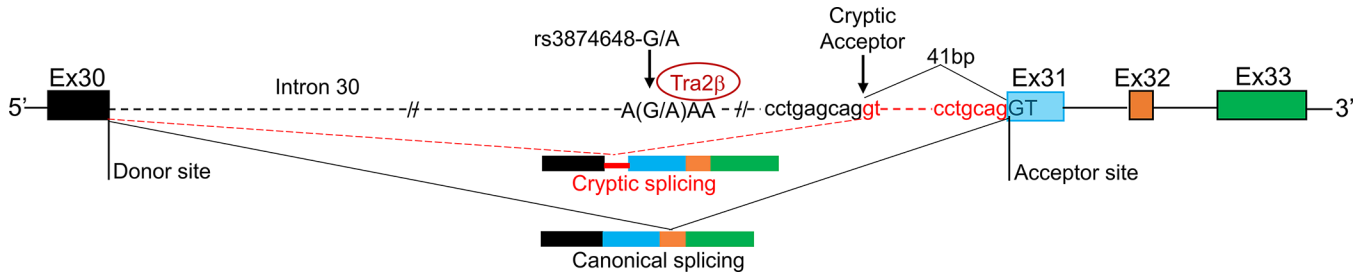

Figure 4. rs3874648G>A modifies *PKD1* expression levels via binding to the splicing regulator Tra2-β.

The intronic variant rs3874648G>A in intron 30 of *PKD1* is located within a conserved regulatory element specifically recognized by the serine and arginine-rich protein (SR protein) Splicing Regulator Tra2-β, 239bp upstream of the canonical splicing acceptor site. Binding of Tra2-β to rs3874648-G allele enables normal splicing using the canonical splice sites, while the rs3874648-A allele increases Tra2-β binding affinity and activates a cryptic acceptor splicing site 41 bp downstream from exon 31 canonical splice acceptor, resulting in partial intron retention, premature termination codon (PTC) introduction.

Table 1.

Top 10% of *PKD1* variants with the highest Fst score

SNP	POSITION (GRCh37)	Annotations	Exon/ Intron	Alternative Allele	Fst Score	AFR	AMR	EAS	EUR	SAS	ClinVar Annotation	Mayo Annotation
rs11866494	2141396	PKD1 c.11712+28C>G	Intron42	G	0.408	72.70%	20.20%	0.00%	18.60%	6.30%	Not Reported in ClinVar	Likely Neutral
rs11862600	2141343	PKD1c.11712+81A>G	Intron42	G	0.398	71.40%	19.60%	0.00%	18.60%	6.00%	Not Reported in ClinVar	NA
rs7206195	2145280	PKD1 c.10499+1362C>T	Intron34	T	0.397	72.10%	21.20%	0.00%	19.90%	6.10%	Not Reported in ClinVar	NA
rs7185040	2145787	PKD1 c.10500–1069A>C	Intron34	C	0.395	72.10%	21.20%	0.00%	19.50%	6.60%	Not Reported in ClinVar	NA
rs58999880	2146528	PKD1 c.10499+621G>T	Intron34	T	0.389	71.90%	21.20%	0.00%	19.90%	7.00%	Not Reported in ClinVar	NA
rs7203729	2140010	PKD1 c.12630A>G, (p.Pro4210=)	Exon46	G	0.385	72.50%	21.00%	0.00%	20.30%	8.60%	Benign	Likely Neutral
rs12926737	2146282	PKD1 c.10499+867T>C	Intron34	C	0.381	71.10%	21.00%	0.00%	19.90%	7.00%	Not Reported in ClinVar	NA
rs3087632	2140454	PKD1 c.12276T>C, (p.Ala4092=)	Exon45	C	0.378	71.90%	21.00%	0.00%	19.90%	9.00%	Benign	Likely Neutral
rs61374883	2154158	PKD1 c.8162–263_8162–262insAG	Intron22	AG	0.371	89.60%	33.60%	11.70%	26.70%	24.70%	Not Reported in ClinVar	NA
rs3874648	2148224	PKD1 c.10051-239G>A	Intron30	A	0.327	57.10%	10.40%	0.40%	9.40%	0.60%	Not Reported in ClinVar	NA
rs2549677	2162361	PKD1 c.3275A>G, (p.Met1092Thr)	Exon14	G	0.318	53.80%	14.30%	0.00%	9.40%	1.40%	Benign	Likely Neutral
rs3087631	2139127	PKD1 c.*601A>T,3’ UTR	3’UTR	T	0.318	65.20%	20.50%	0.00%	20.00%	9.20%	Not Reported in ClinVar	NA
rs2369068	2162887	PKD1 c.3063A>G, (p.Gly1021=)	Exon13	G	0.315	53.60%	14.30%	0.00%	9.40%	1.40%	Benign	Likely Neutral
rs11861948	2163562	PKD1 c.2854–269A>G	Intron11	G	0.315	53.60%	14.30%	0.00%	9.40%	1.40%	Not Reported in ClinVar	NA
rs114796022	2165740	PKD1 c.1850–114T>C	Intron9	C	0.311	54.40%	15.10%	0.00%	10.50%	1.80%	Not Reported in ClinVar	NA
rs116189075	2165737	PKD1 c.1850–111A>G	Intron9	G	0.308	54.10%	15.10%	0.00%	10.50%	1.80%	Not Reported in ClinVar	NA
rs10960	2140680	PKD1 c.12133T>C, (p.Ile4045Val)	Exon44	C	0.306	63.50%	20.50%	0.00%	19.70%	8.50%	Benign	Likely Neutral

SNP	POSITION (GRCh37)	Annotations	Exon/ Intron	Alternative Allele	Fst Score	AFR	AMR	EAS	EUR	SAS	ClinVar Annotation	Mayo Annotation
rs28575767	2156021	PKD1 c.7708A>G, (p.Leu2570=)	Exon20	G	0.306	52.10%	13.80%	0.00%	9.00%	1.40%	Benign	Likely Neutral
rs9928278	2152651	PKD1 c.8949–17T>C	Intron24	C	0.276	56.90%	17.60%	0.00%	16.40%	6.70%	Benign	Likely Neutral
rs13332377	2138869	PKD1 c.*859C>T,3’ UTR	3’UTR	T	0.271	50.20%	14.60%	0.00%	11.80%	2.00%	NA	NA
rs77028972	2152387	PKD1 c.9196A>G, (p.Phe3066Leu)	Exon25	G	0.265	55.50%	17.60%	0.00%	16.40%	6.60%	Benign	Likely Neutral
rs9935834	2152388	PKD1 c.9195C>G, (p.Val3065=)	Exon25	G	0.265	55.50%	17.60%	0.00%	16.40%	6.60%	Benign	NA

Of the 22 variants ranking in the upper 10% Fst scores, three were nonsynonymous variants, five were synonymous variants, two were in the 3’-UTR, and the remainder were in *PKD1* intronic regions. Of these, eight were in genic regions, and fourteen were in the nongenic regions. Of the 22 variants, ten were classified as “Benign” or “Likely Neutral” in the public database, while the remaining twelve variants were unclassified and their significance unknown. AFR: African, AMR: Admixed American, EAS: East Asian, EUR: European, SAS: South Asian. NA: not available

Table 2.

Computational analysis of Tra2-β binding splicing regulators binding motifs

Sequence Position (GRCh38)	Linked ESE protein	Reference Motif (value 0–100)	Linked ESE protein	Mutant Motif (value 0–100)	Variation
chr16:2098227	Tra2-β	aaaa**g** (83.21)	Tra2-β	aaaa**a** (94.14)	13.14%
chr16:2098226	9G8	aa**g**aa (66.18)	Tra2-β	aaaaa (94.14)	42.25%
chr16:2098224			Tra2-β	aaaaa (94.14)	New site
chr16:2098223	9G8	**g**aaaat (66.38)	Tra2-β	**a**aaaa (94.14)	41.82%

Tra2-β binding site computational analysis. Analysis of splicing regulators' binding motifs in intron 30 of *PKD1* using Human Splicing Finder demonstrated that rs3874648G>A substitution is predicted to act by either increasing the affinity or creating a new binding site for Tra2-β in the intron 30 of *PKD1*. Tra2-β: Transformer2-β; 9G8: Serine/Arginine-Rich Splicing Factor 7.

Clin Genet. Author manuscript; available in PMC 2024 March 18.

Table 3.

Predicted strength of authentic and cryptic *PKD1* intron 30 splice sites

Sequence Position (GRCh38)	Splice site type	Motif	Potential splice site	Consensus value (0–100)	HSF result interpretation
chr16:2098025	Acceptor	tgtcctgagcaggt	tgtcctgagcagGT	87.07	Alternative 3’ splicing site
chr16:2097984	Acceptor	ctcgtcctgcagGT	ctcgtcctgcagGT	93.71	Canonical 3’ splicing site

Splice site analysis using Human Splicing Finder identified a cryptic 3’ acceptor splicing site 41bp from the canonical 3’ splicing site. This cryptic site yielded a high consensus value similar to that of the canonical 3’ splice site (87.07 vs 93.71).