Gene 516 (2013) 93–100

Contents lists available at SciVerse ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Methods Paper

Genetic diagnosis of autosomal dominant polycystic kidney disease by targeted capture and next-generation sequencing: Utility and limitations

Xiao-Ping Qi [a,*,1], Zhen-Fang Du [b,1], Ju-Ming Ma [a,1], Xiao-Ling Chen [b], Qing Zhang [c], Jun Fei [a], Xiao-Ming Wei [c], Dong Chen [d], Hai-Ping Ke [e], Xuan-Zhu Liu [c], Feng Li [a], Zhen-Guang Chen [a], Zheng Su [c], Hang-Yang Jin [a], Wen-Ting Liu [b], Yan Zhao [b], Hu-Ling Jiang [b], Zhang-Zhang Lan [c], Peng-Fei Li [c], Ming-Yan Fang [c], Wei Dong [c,**], Xian-Ning Zhang [b,***]

[a] Department of Urologic Surgery, the 117th PLA Hospital, Hangzhou, Zhejiang Province, China
[b] Department of Cell Biology and Medical Genetics, National Education Base for Basic Medical Sciences, Institute of Cell Biology, Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, China
[c] BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, Guangdong Province, China
[d] Department of Evolution, Ecology and Behavior, University of Liverpool, Liverpool, UK
[e] Department of Biology, Ningbo College of Health Sciences, Ningbo, Zhejiang Province, China

ARTICLE INFO

Article history:
Accepted 3 December 2012
Available online 21 December 2012

Keywords:
Autosomal dominant polycystic kidney disease
Genetic diagnosis
PKD1 gene
PKD2 gene
Targeted DNA-HiSeq
Whole-exome sequencing

ABSTRACT

Mutation-based molecular diagnostics of autosomal dominant polycystic kidney disease (ADPKD) is complicated by genetic and allelic heterogeneity, large multi-exon genes, and duplication sequences of *PKD1*. Recently, targeted resequencing by pooling long-range polymerase chain reaction (LR-PCR) amplicons has been used in the identification of mutations in ADPKD. Despite its high sensitivity, specificity and accuracy, LR-PCR is still complicated. We performed whole-exome sequencing on two unrelated typical Chinese ADPKD probands and evaluated the effectiveness of this approach compared with Sanger sequencing. Meanwhile, we performed targeted gene and next-generation sequencing (targeted DNA-HiSeq) on 8 individuals (1 patient from one family, 5 patients and 2 normal individuals from another family). Both whole-exome sequencing and targeted DNA-HiSeq confirmed c.11364delC (p.H3788QfsX37) within the unduplicated region of *PKD1* in one proband; in the other family, targeted DNA-HiSeq identified a small insertion, c.401_402insG (p.V134VfsX79), in *PKD2*. These methods do not overcome the screening complexity of homology. However, the true positives of variants confirmed by targeted gene and next-generation sequencing were 69.4%, 50% and 100% without a false positive in the whole coding region and the duplicated and unduplicated regions, which indicated that the screening accuracy of *PKD1* and *PKD2* can be largely improved by using a greater sequencing depth and elaborate design of the capture probe.

1. Introduction

Autosomal dominant polycystic kidney disease (ADPKD; OMIM ID: 173900), the most common inherited kidney disease with an incidence of 1 in 400 to 1000, is characterized by the development and progressive enlargement of cysts in the kidney (Harris and Rossetti, 2010; Koptides and Deltas, 2000). Extra-renal manifestations include cysts in the liver, pancreas, seminal vesicles, and arachnoid membranes (Harris and Torres, 2009). Approximately 50% of affected individuals progress to end-stage renal disease (ESRD) by the age of 60 years and these patients comprise 4–5% of the dialysis population in the United States (Torres and Harris, 2009). Mutations in *PKD1* and *PKD2* were thought to account for 85% and 15% of cases, respectively, while more recent population-based series suggest a higher prevalence of *PKD2* ranging from 26% to 36% (Koptides et al., 1998; Rossetti et al., 2007). Large deletions and duplications account for ~4% of the pathogenic mutations reported for *PKD1*

Abbreviations: ADPKD, autosomal dominant polycystic kidney disease; CT, computed tomography; ESRD, end-stage renal disease; LR-PCR, long-range polymerase chain reaction; MRI, magnetic resonance imaging; ORF, open reading frame; targeted DNA-HiSeq, targeted gene and next-generation sequencing; WES, whole-exome sequencing.

* Correspondence to: X.-P. Qi, Department of Urologic Surgery, the 117th PLA Hospital, 40 Jichang Road, Hangzhou, Zhejiang Province, 310004, China. Tel.: +86 571 87340912; fax: +86 571 87348500.
** Correspondence to: W. Dong, BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen, Guangdong Province, 518083, China. Tel./fax: +86 755 25273620.
*** Correspondence to: X.-N. Zhang, Department of Cell Biology and Medical Genetics, Zhejiang University School of Medicine, 866 Yuhangtang Road, Hangzhou, Zhejiang Province, 310058, China. Tel.: +86 571 88208367.
E-mail addresses: qxplmd@163.com (X.-P. Qi), dongw@genomics.cn (W. Dong), zhangxianning@zju.edu.cn (X.-N. Zhang).
[1] These authors contributed equally to this work.

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2012.12.060

(Ariyurek et al., 2004; Consugar et al., 2008). Estimates of the variability in ESRD development indicate that 18–50% may be due to heritable modifying factors, and Liu et al. (2011) suggest that genetic variation of *DKK3* may modify the severity of ADPKD resulting from *PKD1* mutations.

The diagnosis of ADPKD is typically determined by renal imaging with the ultrasound diagnostic criterion of age-related cyst number (Connor et al., 2008). Computed tomography (CT) and magnetic resonance imaging (MRI) can also quantify renal cystic disease, and a recent trial showed that MRI is a reliable means of monitoring disease progression by renal enlargement (Chapman et al., 2003). These diagnostics are highly reliable in older individuals (>30 years) but less certain in young adults and in patients with a negative family history (Chapman et al., 2003; Connor et al., 2008). This is a significant issue when renal donation is being considered before the age of 30, particularly for *PKD2* patients in whom the sensitivity of these criteria is only 67% (Nicolau et al., 1999). Linkage analysis requires accurate diagnosis and the availability and willingness of sufficient numbers of affected family members to be tested; this is feasible in <50% of families (Parfrey et al., 1990). The identification of pathogenic genes for ADPKD has allowed molecular diagnosis (Hughes et al., 1995; Mochizuki et al., 1996). Molecular testing plays a vital role in the evaluation of related potential kidney donors with doubtful imaging data, in individuals with a negative family history, and in cases of early-onset ADPKD (Harris and Rossetti, 2010). Furthermore, the characterization of mutations in clinical trial cohorts provides genetic stratification for the evaluation of such trials (Hogan et al., 2010; Rossetti et al., 2007).

PKD1 is a large gene consisting of 46 exons with an open reading frame (ORF) of ~13 kb. Its entire 5′ region up to exon 33 is duplicated six times on chromosome 16p, and the presence of these highly homologous pseudogenes has made genetic analysis of *PKD1* challenging (Bogdanova et al., 2001). By contrast, *PKD2* is a single-copy gene consisting of 15 exons with an ORF of ~3 kb (Rossetti et al., 2007; Wang et al., 2012). High levels of allelic heterogeneity exist for both genes. There are no mutation hot spots for both *PKD1* and *PKD2*, meaning mutations are usually private, highly variable, and spread throughout the entire gene (Gout et al., 2007). The availability of protocols for long-range and locus-specific amplification of *PKD1* has enabled complete mutation screening of this complex gene (Rossetti et al., 2002). The largest obstacle in completing this genetic analysis for *PKD1* has been distinguishing it from a family of homologs that map elsewhere on chromosome 16. Approximately 70% of this gene, beginning with its 5′ terminus, is duplicated in at least three other loci located more proximally on the chromosome. The sequence identity is >95% in the regions of similarity (The International Polycystic Kidney Disease Consortium, 1995). Mutation analysis for diagnostics has also been challenged because of the multi-exon structures of the genes, high GC content, marked allelic heterogeneity, and common missense variants (Rossetti et al., 2007; Tan et al., 2009). Thus, few complete screens of both genes have been described, and molecular testing by direct DNA sequencing is now possible with likely mutations identified in <90% of patients (Kumar et al., 1991; Rossetti et al., 2007; Tan et al., 2009). Although the demand for molecular diagnostics is limited, the development of therapies is likely to transform the situation as a firm diagnosis in young adults is needed before significant renal changes occur (Torres and Harris, 2006).

Whole-exome sequencing (WES) is a direct, reproducible and robust method for the confirmation of novel pathogenic genes and the genetic diagnosis of both Mendelian and complex diseases (Bonnefond et al., 2010; Choi et al., 2009). Here, we used WES, coupling the Agilent whole-exome capture system to the Illumina HiSeq 2000 DNA sequencing platform, to attempt the genetic diagnosis of ADPKD. Meanwhile, we used targeted gene and next-generation sequencing (targeted DNA-HiSeq), a custom high-density, solid-phase array to capture 222 distinct genes involved in more than 100 different genetic diseases, including ADPKD (Wei et al., 2011). And this targeted DNA-HiSeq technology has been demonstrated to successfully detect variants of disease-causing genes with high fidelity, high throughput, high speed and low cost (Wei et al., 2011). After enrichment, high-throughput sequencing using the Illumina HiSeq2000 Analyzer was performed to identify the variants of disease-causing genes in different patient samples in one sequencing lane using the sample bar-coding method. Next, we used the Sanger DNA sequencing technique to evaluate the accuracy of WES and targeted DNA-HiSeq. The true positives in the duplicated region of *PKD1* by WES was 28.6%, which warned us to use this method cautiously in the screening of exonic duplicated regions, while the results from targeted DNA-HiSeq indicated that by increasing the sequencing depth and improving the design of the capture probe, it is possible to analyze the gene variant profiles of monogenic diseases with high sensitivity and throughput.

2. Materials and methods

2.1. Subjects

The diagnosis of ADPKD was based on (i) a clinical history consistent with ADPKD and (ii) renal ultrasound and/or CT/MRI findings consistent with ADPKD. Two unrelated families with typical ADPKD and 100 unrelated healthy matched controls were included (Fig. 1).

This study was conducted in conformity with the Declaration of Helsinki and was approved by the Ethics Committees of the 117th PLA Hospital and Zhejiang University. Written informed consent was given by all subjects.

2.2. DNA sequencing, data filtering and analysis pipeline

Peripheral blood genomic DNA samples from each of 2 ADPKD family members (FA II-1 and FB II-2) were prepared for WES. The whole-exome capture array design, library construction, next-generation sequencing, data filtering and analysis pipeline followed previously described protocols (Qi et al., 2011; Shi et al., 2011).

For targeted DNA-HiSeq, a custom capture array (NimbleGen, Roche) was designed to capture all exons (3093), splice sites and the immediately adjacent intron sequences of 222 genes known to be associated with multiple types of hereditary diseases according to GeneReviews (NCBI), including inherited metabolic disorders, tuberous sclerosis, and spastic paraplegia. The methods used for DNA target capture, enrichment and elution followed previously described protocols (Wei et al., 2011). We used this method on samples from 8 individuals (FA II-1, and FB II-2, II-4, II-9, II-12, III-1, III-5 and VI-1) from two unrelated ADPKD families (Fig. 1).

2.3. Confirmation of variants in PKD1 and PKD2

Sanger sequencing of the entire *PKD1* and *PKD2* coding sequence was used as the gold standard to evaluate the accuracy of variants in *PKD1* and *PKD2* disclosed by WES and targeted DNA-HiSeq. The duplicated region of *PKD1* was amplified as five PCR fragments that were either anchored in single-copy DNA or mismatched with its pseudogenes (Phakdeekitcharoen et al., 2001). LR-PCR fragments were amplified as previously outlined, and exons were amplified from these fragments after a dilution of at least 1:1000 to avoid genomic DNA carryover (Knierim et al., 2011; Phakdeekitcharoen et al., 2001).

2.4. Pathogenic mutation confirmation

In an effort to assess the functional significance of the gene variations identified in this study, we used public databases as well as web-based software. The pathogenic potential was evaluated computationally by examining interspecies sequence variations, using PolyPhen (http://genetics.bwh.harvard.edu/pph) (Ramensky et al., 2002), Sorting Intolerant from Tolerant (SIFT; http://sift.bii.a-star.edu.sg/) (Ng and