Development and Validation of a Mass Spectrometry–Based Assay for the Molecular Diagnosis of Mucin-1 Kidney Disease

Citation: Blumenstiel, Brendan et al. "Development and Validation of a Mass Spectrometry–Based Assay for the Molecular Diagnosis of Mucin-1 Kidney Disease." The Journal of Molecular Diagnostics 18.4 (2016): 566–571. © 2016 American Society for Investigative Pathology and the Association for Molecular Pathology

As Published: http://dx.doi.org/10.1016/j.jmoldx.2016.03.003

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/106520

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

The Journal of Molecular Diagnostics, Vol. 18, No. 4, July 2016

ELSEVIER

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Development and Validation of a Mass Spectrometry–Based Assay for the Molecular Diagnosis of Mucin-1 Kidney Disease

Brendan Blumenstiel,* Matthew DeFelice,* Ozge Birsoy,*† Anthony J. Bleyer,‡ Stanislav Kmoch,§ Todd A. Carter,* Andreas Gnirke,* Kendrah Kidd,‡ Heidi L. Rehm,*† Lucienne Ronco,* Eric S. Lander,* Stacey Gabriel,* and Niall J. Lennon*

From the Broad Institute of MIT and Harvard, Cambridge, Massachusetts; the Partners Healthcare Laboratory for Molecular Medicine,† Cambridge, Massachusetts; the Nephrology Section,‡ Wake Forest Baptist Medical Center, Winston-Salem, North Carolina; and the Institute for Inherited Metabolic Disorders,§ Charles University in Prague, Prague, Czech Republic*

Accepted for publication March 9, 2016.

Address correspondence to Brendan Blumenstiel, B.A., Broad Institute of MIT and Harvard, 320 Charles St., Cambridge, MA 02141. E-mail: brendan@broadinstitute.org.

Mucin-1 kidney disease, previously described as medullary cystic kidney disease type 1 (MCKD1, OMIM 174000), is an autosomal dominant tubulointerstitial kidney disease recently shown to be caused by a single-base insertion within the variable number tandem repeat region of the *MUC1* gene. Because of variable age of disease onset and often subtle signs and symptoms, clinical diagnosis of mucin-1 kidney disease and differentiation from other forms of hereditary kidney disease have been difficult. The causal insertion resides in a variable number tandem repeat region with high GC content, which has made detection by standard next-generation sequencing impossible to date. The inherently difficult nature of this mutation required an alternative method for routine detection and clinical diagnosis of the disease. We therefore developed and validated a mass spectrometry–based probe extension assay with a series of internal controls to detect the insertion event using 24 previously characterized positive samples from patients with mucin-1 kidney disease and 24 control samples known to be wild type for the variant. Validation results indicate an accurate and reliable test for clinically establishing the molecular diagnosis of mucin-1 kidney disease with 100% sensitivity and specificity across 275 tests called. *(J Mol Diagn 2016, 18: 566–571; http://dx.doi.org/10.1016/j.jmoldx.2016.03.003)*

Previously described as medullary cystic kidney disease type 1 (MCKD1), mucin-1 kidney disease (MKD) is a genetic subclassification of a larger family of autosomal dominant tubulointerstitial kidney diseases.[1] Although mucin-1 is expressed in many tissues in the body, the only clinical consequence of MKD is slowly progressive deterioration of kidney function, leading to the need for dialysis or kidney transplantation.[1,2] Furthermore, the pathological changes that occur in the kidney are nonspecific, meaning that pathological studies of kidney biopsy specimens from these patients will also not reveal a specific diagnosis. Because there are no other signs of this condition, a genetic assay is the only way to diagnose MKD and differentiate it from many other causes of kidney disease. MKD is considered a rare disease, but its low prevalence may be in part due to current difficulties in diagnosis.

Once MKD is diagnosed within a family, it still remains difficult to predict disease risk in individual family members. The loss of kidney function is variable, with some patients starting dialysis in the third decade of life, whereas other family members may have relatively preserved kidney function into their 70s. The only clinical method available to detect disease risk in a family member is to test the serum creatinine level. However, this test is not sensitive for early disease stages because kidney function can be reduced before an increase in serum creatinine level.[3–5] It also is not specific because many individuals in the population have mild elevations in their serum creatinine from other conditions. Thus, clinical diagnosis of this condition often is not

Supported by the Slim Initiative for Genomic Medicine, a project funded by the Carlos Slim Foundation of Mexico.

Disclosures: B.B., M.D., and A.G. are named as inventors on a patent application covering the method described in this article.

http://dx.doi.org/10.1016/j.jmoldx.2016.03.003

possible, leading to uncertainty regarding disease status in many family members. In addition, as kidney function worsens, many patients with MKD will require kidney transplantation. Siblings are often the best candidates for kidney donation, but in the past many siblings have not been allowed to donate because of the inability to rule out disease risk. In some cases siblings have donated and later been found to have MKD. Thus, the development of a genetic assay will allow family members to learn whether they are at risk for MKD and also will allow for the transplantation of many affected individuals through the ability to rule out MKD in potential family donors.[6–8]

MKD is caused by a single-base insertion within a large variable number tandem repeat (VNTR) unit (60 bp) of the *MUC1* gene, which encodes the transmembrane protein Mucin 1. The single-base cytosine insertion event resides in repeat region with high GC content (>80%) and expands a 7-bp cytosine homopolymer (*7C*) to an 8-bp homopolymer (*8C*). The addition of the cytosine residue causes a frameshift, resulting in a premature stop codon and a truncated protein lacking the downstream self-cleavage module expressed in the wild-type version.[9]

Three factors make detection of the variant by next-generation sequencing and traditional genotyping methods extremely difficult. First, the insertion event is believed to occur in a different yet specific repeat unit, depending on family. Coupled with the fact that the repeat number can vary from 20 to 125 for an individual, directly targeting the variant 60-bp unit with anchored primers outside the VNTR is unlikely.[10] Because of the highly skewed allelic ratio of mutant to wild-type copies of the repeat, detecting a 0.8% to 5% alternate signal within a high (95% to 99.2%) wild-type background makes standard genotyping assays unreliable for a clinical setting. Second, the high GC nature of the repeat domain leads to underrepresentation in both targeted exome and whole-genome sequencing. Third, because the insertion expands a *7C* homopolymer to *8Cs* (*+C*), even if the region is sufficiently covered with next-generation sequencing reads, the low allelic ratio variant would likely be interpreted as a sequencing error.[11,12]

In early efforts to validate the *+C* single-base insertion as the causal variant responsible for MKD, a genotyping method that would enrich for and detect the rare mutation within a strong wild-type background was needed. Fortuitously, it was discovered that the *+C* insertion destroyed a restriction site cleaved by endonuclease MwoI. Digestion of the wild-type 60-bp repeat domain enabled preenrichment of the intact *+C* domain for downstream interrogation. To detect the *+C* insertion, genomic DNA (gDNA) was digested with MwoI, and the remaining intact repeat units were amplified using tailed PCR primers flanking the variant site. A second MwoI digestion further digested any remaining amplified wild-type fragments. Interrogation of the *+C* base was performed by probe extension and detection by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry using the Sequenom MassArray system (Sequenom, San Diego, CA).

During initial assay development, six MKD-affected families, consisting of 62 symptomatic and 79 unaffected individuals, were genotyped along with 500 control individuals from the Japanese, Chinese, Yoruban, and Tuscan HapMap 3 populations. Results revealed that the *+C* insertion segregated perfectly with the disease risk haplotype within the affected families and was completely absent in all 500 controls.

Our aim was to further develop and validate this robust assay to be used as a clinical molecular diagnostic test for MKD. To establish the analytical validity of the optimized assay, we retrospectively tested 24 samples with known clinical family history of MKD, which were found to be positive for the *+C* insertion within a research setting, and 24 negative Hap Map control samples.

Materials and Methods

Patients, Samples, and DNA Extraction

We obtained purified genomic DNA extracted from blood for 24 individuals enrolled in an ongoing MKD research project in collaboration with Anthony J. Bleyer, M.D., and Wake Forest University Baptist Medical Center. All 24 individuals consented for research and had a family history of kidney disease. Previous testing under the research protocol revealed that all individuals harbored the MKD-associated *+C* insertion. Samples were deidentified and assigned a unique alphanumeric code for sample tracking. Genomic DNA derived from 24 HapMap cell lines were used as negative controls for the validation (Coriell Cell Repository, Camden, NJ).

Multiplex Primer Design and Enrichment

Fifty nanograms of genomic DNA was digested to completion using 10 units of restriction endonuclease MwoI (New England Biolabs, Beverly, MA), which selectively cleaves the wild-type 7C (GCCCCCCCAGC) sequence repeats while leaving intact mutant *8C* fragments containing the *+C* insertion (GCCCCCCCCAGC). After a 2× SPRI cleanup using Ampure XP beads (Beckman Coulter, Brea, CA), tailed primers nested within the 60-bp repeat unit were then used to PCR amplify the remaining intact VNTR fragments, thus enriching for duplication-containing fragments over wild-type background. As a control for gDNA template addition and enzymatic steps, a nonpolymorphic region of the housekeeping gene *GAPDH* was multiplexed in the same PCR reaction. PCR was performed in a 25-μL reaction volume using 0.8 units of Hot Start Plus Taq Polymerase (Qiagen, Hilden, Germany), supplied reaction buffer at 1×, 1 mmol/L $MgCl_2$, 0.76 mmol/L dNTP, 0.2 mmol/L of each primer, MKD forward (5′-CTGG-GAATCGCACCAGCGTGTGGCCCCGGGCTCCACC-3′),

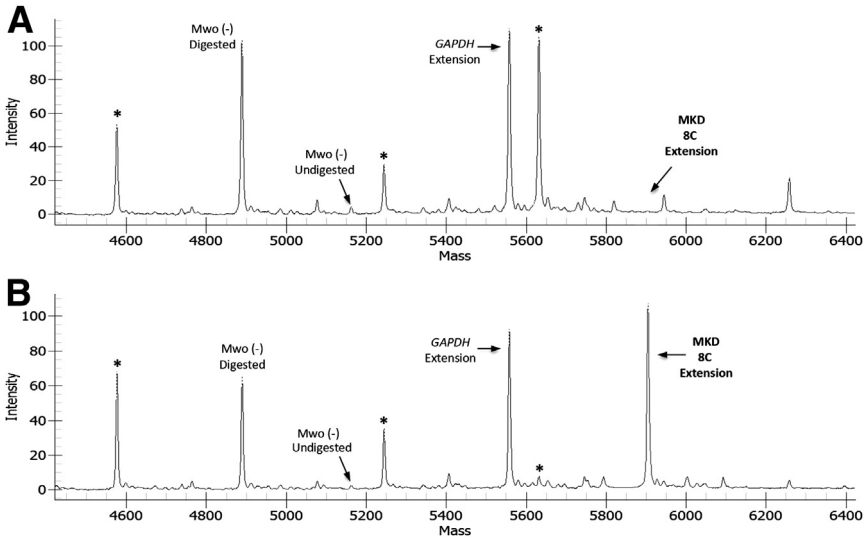

Figure 1 Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry spectra readout of a wild-type negative sample (**A**) and $+C$ positive sample (**B**). Unextended probe products are indicated by **asterisks**. Successful restriction digestion is indicated by a strong peak at 4889.21 Da and low-intensity peak at 5162.4 Da (MwoI control). Reaction performance is indicated by the *GAPDH* control extension product at 5557.63 Da. $+C$ insertion event indicated by a peak at 5904.83 Da (**shaded area**).

MKD reverse (5′-CGTGGATGAGGAGCCGCAGTGTCC-GGGGCCGAGGTGACA-3′), *GAPDH* forward (5′-CG-CGGTAAAACGACGGCCAGAAGGCTGTGGGCAAG-GTC-3′), *GAPDH* reverse (5′-GGCCCGACAGGAAA-CACGTATGACCCTGCTTCACCACCTTCTTG-3′). PCR was performed using the following cycling conditions: initial denaturation of 5 minutes at 95°C followed by 45 cycles of (30 seconds at 94°C, 30 seconds at 62°C, and 1 minute at 72°C) and a final extension of 10 minutes per 72°C. Then samples were stored at 4°C until further processing. PCR reactions were then cleaned again using 2× SPRI. Once amplicons were eluted, a second digestion with MwoI was used to further digest any amplified wild-type repeat units, further enriching for mutant $+C$ product.

MwoI Digestion Control

To control for successful MwoI digestion, a nonhuman 132-bp amplicon was spiked in to the second digestion reaction. This amplicon contained a single MwoI restriction site, which when cut produces two fragments 101 and 31 bp in length. Successful digestion was assayed via probe extension multiplexed with the MKD and *GAPDH* assays in the following steps.

Probe Extension and Analyte Detection

Oligonucleotide probes were specifically designed to interrogate the $+C$ insertion and the *GAPDH* and MwoI digestion controls. Probe extension was performed in a 10-μL reaction volume using 1.28 units of Thermosequenase (Sequenom); 0.7× PCR buffer (Qiagen); 0.22× iPlex Buffer Plus (Sequenom); 0.6 mmol/L $MgCl_2$; 0.19 μL SAP Buffer (Sequenom); 0.22 mmol/L each deoxyadenosine triphosphate, 2',3'-dideoxycytidine-5'-triphosphate, and 2',3'-dideoxyguanosine-5'-triphosphate; 0.6 μmol/L each oligo probe, MKD probe (5′-CGGGCTCCACCGCCCCC-CC-3′), *GAPDH* probe (5′-CCTGAGCTGAACGGGAA-3′), and MwoI probe (5′-AATCGTGATCTAGCA-3′). Probe extension was performed using the following cycling conditions: 2 minutes at 94°C followed by 75 cycles of (5 seconds at 94°C, 5 seconds at 52°C, and 5 seconds at 72°C) and a final extension of 7 minutes at 72°C. Then samples were stored at 4°C until further processing. Reactions were then desalted using 6 mg of Clean Resin cationic exchange resin (Sequenom).

Analyte pools for each sample were then co-crystalized with organic matrix (3-hydroxypicolinic acid) on a 384-spot silica chip using a pin tool and detected by MALDI-TOF mass spectrometry using the Sequenom Mass Array platform. Spectral data were generated, and analyte intensity and signal to noise values (SNR) were recorded.

Analysis and Genotype Calling

Genotype calling is performed by analyzing the resulting extension products for the three multiplexed probes (Figure 1). Thresholds were set based on empirical observation of signal during research and development runs.

GAPDH Control

To control for template addition, PCR amplification, SPRI clean-up, and probe extension reactions, a 151-bp region of the *GAPDH* gene (chr12:6,646,867-6,647,017) is PCR amplified and interrogated with a 17-bp extension probe. An SNR >15 for an extension product at 5244.43 Da represents a passing reaction.

Mwo-Negative Control

To control for successful MwoI digestion, a nonhuman, synthetic, 132-bp amplicon, which contains a single MwoI restriction site, is spiked into the second digestion reaction. Digestion is interrogated with a 15-bp extension probe positioned a single base upstream of the cut site. A successful restriction digestion is indicated by a single

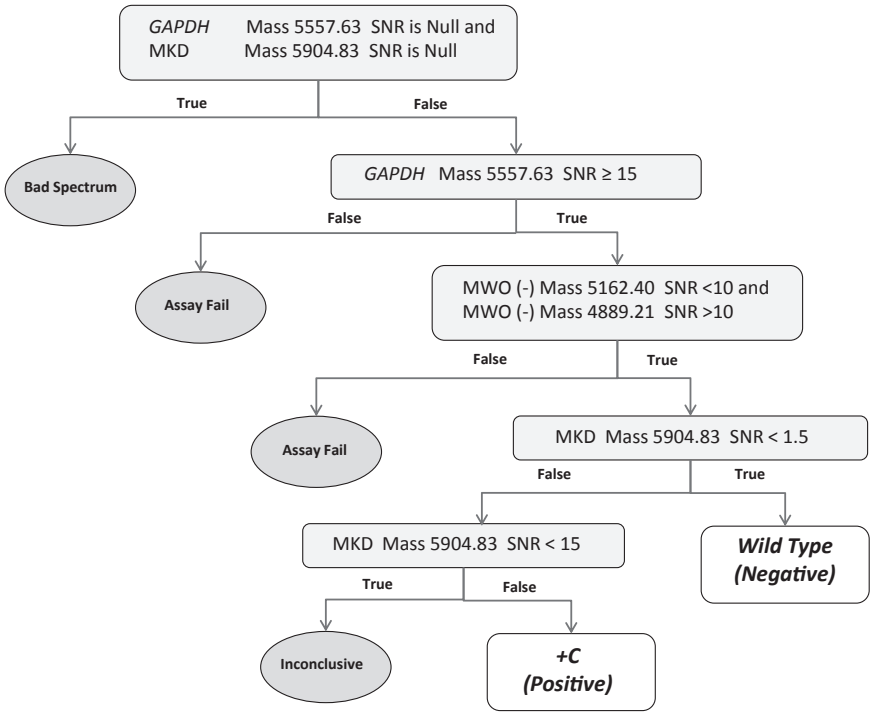

Figure 2 Flowchart of mucin-1 kidney disease (MKD) calling. Signal to noise ratios (SNRs) for four potential peaks are used to determine the test result. Conclusive calls of wild-type (negative) or *+C* (positive) are only reported when both the *GAPDH* and MwoI (−) controls pass the set thresholds. Samples with low but detectable signal for the MKD extension product (SNR, 1.5 to 15) are called inconclusive.

deoxyadenosine triphosphate termination product at 4889.21 Da with a SNR >10 and an SNR <10 for a two-base addition 2',3'-dideoxycytidine-5'-triphosphate termination product at 5162.40 Da.

MKD Assay

For the detection of the *+C* insertion, a 20-bp probe is positioned just upstream of the insertion site. If the *+C* insertion event is present, the probe will be terminated with a 2',3'-dideoxyguanosine-5'-triphosphate, resulting in an extension analyte at 5904.83 Da.

There are four possible call outcomes for this assay, and they are determined for each sample by assessing the SNR for the four possible extension analytes (Figure 2). SNR thresholds for each call type have been established by values from previously performed replicate experiments and were reassessed and confirmed in this validation.

Assay Fail

GAPDH extension analyte SNR at 5557.62 Da is <15, MwoI control extension analyte SNR at 4889.21 Da is <10, and analyte SNR at 5162.40 Da is >10.

+C (Positive)

MKD extension analyte SNR at 5904.83 Da is ≥15 with passing MwoI and *GAPDH* controls.

Wild Type (Negative)

MKD extension analyte SNR at 5904.83 Da is <1.5 with passing MwoI and *GAPDH* controls.

Inconclusive

MKD extension analyte SNR at 5904.83 Da is <15 and ≥1.5 with passing MwoI and *GAPDH* controls.

Results

The 48 validation samples were analyzed by two independent technologists (B.B., M.D.), both performing the testing in duplicate for each sample in each run. One technologist repeated the experiment a second time for a total of six replicate runs of the sample set. Using the calling thresholds described in *MKD Assay*, all six replicate runs were called (Figure 3). Call rate, precision, accuracy, and analytical sensitivity and specificity were determined.

Of 288 total samples detected, three failed either the *GAPDH* or Mwo (negative) control and were called assay fail for a rate of 1%. Of the 285 passing samples, 10 samples were called inconclusive for a rate of 3.5%. Of the 275 samples that generated conclusive calls of either *+C* (positive) or wild type (negative), no false-negative or false-positive calls were made against the known reference, indicating analytical sensitivity and specificity of the test to be 100%. All conclusive calls made were concordant with their known reference for an accuracy of 100%. Although conclusive call rates varied from 83% to 100% across the six replicate runs, performed by two technologists (B.B., M.D.) across three different days, no discrepant conclusive calls were made, and assay precision was maintained with 100% repeatability and reproducibility (Table 1).

In addition to the standard 50-ng DNA input, we also tested assay performance across a range of DNA input amounts. The 48 validation samples were tested in a single run with both 20-ng and 100-ng total DNA input. No discordant calls were made against sample reference for 100% accuracy. Analytical sensitivity and specificity were

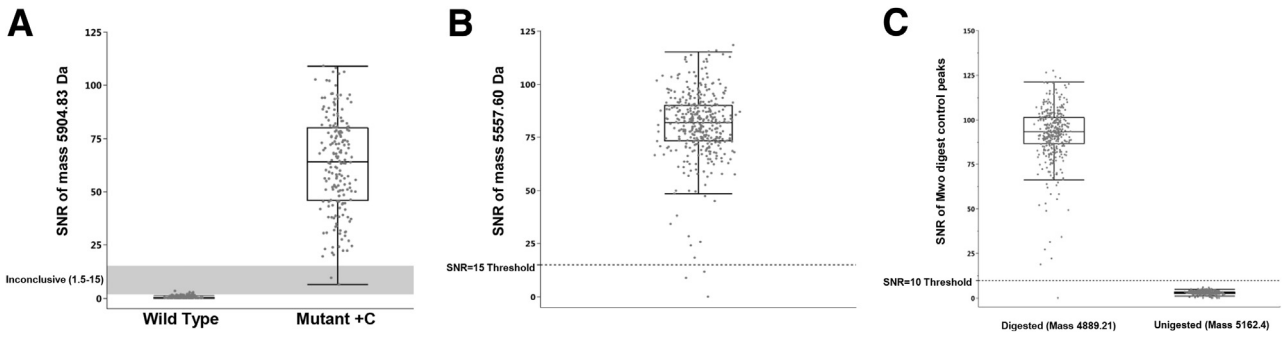

Figure 3 **A:** The distribution of peak signal to noise ratios (SNRs) for the mucin-1 kidney disease *+C* analyte at 5904.83 Da was observed for all 384 validation reactions. Results binned by known sample reference. Eleven of 384 reactions generated a SNR between 1.5 and 15 (**shaded area**) and were called inconclusive. **B:** Distribution of *GAPDH* control extension analyte SNR at 5557.63 Da for 384 assays performed during validation. Three of 384 wells failed because the signal was below the set SNR threshold of 15 (**dotted line**). **C:** Successful MwoI restriction digestion is determined by assessing the two potential extension products of a probe interrogating the restriction site of a synthetic amplicon spike in. One of 384 assays failed the MwoI control with a signal below the SNR threshold of 10 (**dotted line**) at mass 4889.21 Da. Both peak values are plotted for 384 assay wells run in the validation. Data presented as distributions with outlier box plots.

100%. Conclusive call rates for 20-ng and 100-ng inputs were 98% and 100%, respectively, demonstrating that the test performance is consistent across a fivefold range of input material.

Discussion

The inherent nature of this causal mutation for MKD makes it extremely difficult to detect using standard diagnostic detection methods. The high GC content of approximately 85%, the variable length of the VNTR, and a mutation involving a single-base expansion of a cytosine homopolymer make standard sequencing technologies unreliable for rapid and accurate diagnosis in the clinic. The restriction enzyme MwoI cuts wild-type *7C* copies of the 60-bp repeat unit while leaving intact the mutated *8C* units, allowing us to develop a detection method around the enzymatic enrichment of mutant signal over wild-type background. By relying on separation of analytes by molecular mass using MALDI-TOF mass spectrometry, we are able to clearly distinguish signal intensity derived from both mutant and wild-type products with very high specificity. Although the presence of the *+C* insertions is quite easy to resolve, with a typically strong analyte peak at the expected size in the readout spectra, calling a wild-type or negative patient relies on interpreting the lack of a signal. This could lead to a higher likelihood of false-negative calls, resulting from technical or process errors, such as no gDNA template addition, thermocycler failure, or poor elution during reaction cleanup. Incorporation of a multiplexed assay targeting the *GAPDH* housekeeping gene allows us to control for all steps in the process and is built into the result calling logic. By automating the assay using robotic liquid handlers, we can simultaneously process up to 89 patient samples along with seven control wells in standard 96-well format. From DNA receipt to result generation, we can typically generate batch results in approximately five days, allowing us to quickly report the molecular diagnosis of a patient back to the ordering physician.

The test that was validated in this study only looks for a *+C* insertion; however, other insertions or deletions within the VNTR could potentially result in the same truncated form of the Mucin 1 protein and thus manifest as MKD. Interestingly, no other mutations in *MUC1* have thus far been reported as causal for MKD, suggesting that if alternative mutations exist, they are rare. In addition, the position of the mutation within the VNTR (early or late unit of the entire repeat region) could possibly contribute to the variable age of onset seen in affected patients. Unfortunately,

Table 1 Reproducibility of Assay Across Eight Replicates of the 48 Sample Validation Set

Validation set	Technologist	Inconclusive (%)	Assay fail (%)	True-positive rate (%)	True negative rate (%)	Analytical sensitivity (%)	Analytical specificity (%)	Accuracy (%)
Plate 1, replicate 1	A	2	0	100	100	100	100	100
Plate 1, replicate 2	A	0	0	100	100	100	100	100
Plate 2, replicate 1	A	0	2	100	100	100	100	100
Plate 2, replicate 2	A	0	2	100	100	100	100	100
Plate 3, replicate 1	B	15	2	100	100	100	100	100
Plate 3, replicate 2	B	4	0	100	100	100	100	100
Mean		3.5	1	100	100	100	100	100
Plate 4, 20 ng	A	2	0	100	100	100	100	100
Plate 4, 100 ng	A	0	0	100	100	100	100	100

Although one replicate run shows a higher inconclusive rate of approximately 15%, analytical sensitivity, specificity, and accuracy all remain at 100%.

although the test presented here detects the presence of the causal variant in one single copy of the VNTR sequence, it does not identify the specific repeat unit from which it originates. Ultimately, long read sequencing that can accurately read through the entire 20- to 125-copy VNTR region will be required to positively identify the position of the mutation in most patients.

The development of a successful genetic assay, for a condition in which a clinical diagnosis often cannot be made, will lead to immediate improvements in the care of patients with MKD. Although there is no specific therapy available for MKD, there are a number of tangible benefits. First, patients can obtain a genetic diagnosis without the risk of a kidney biopsy that would often fail to provide a diagnosis. Second, siblings of affected individuals can be tested and appropriately cleared for kidney transplantation. Third, the progression of kidney disease in this disorder, and factors affecting it, can be better studied. Fourth, the condition can be studied in individuals who are minimally affected. Fifth, patients can be identified for interventional trials.

References

1. Eckardt KU, Alper SL, Antignac C, Bleyer AJ, Chauveau D, Dahan K, Deltas C, Hosking A, Kmoch S, Rampoldi L, Weisener M, Wolf M, Devuyst O: Autosomal dominant tubulointerstitial kidney disease: diagnosis, classification, and management: a KDIGO consensus report. Kidney Int 2015, 88:676−683
2. Stravrou C, Deltas C, Christophides TC, Pierides A: Outcome of kidney transplantation in autosomal dominant medullary cystic kidney disease type 1. Nephrol Dial Transplant 2003, 18:2165−2169
3. Stravrou C, Koptides M, Tombazos C, Psara E, Patsias C, Zouvani I, Kyriacou K, Hildebrandt F, Christophides T, Pierides A, Deltas C: Autosomal-dominant medullary cyctic kidney disease type 1: clinical and molecular findings in six large Cypriot families. Kidney Int 2002, 62:1385−1394
4. Bostom AG, Kronenberg F, Ritz E: Predictive performance of renal function equations for patients with chronic kidney disease and normal serum creatinine levels. J Am Soc Nephrol 2002, 13:2140−2144
5. Shemesh O, Golbetz H, Kriss JP, Myers BD: Limitations of creatinine as a filtration marker in glomerulopathic patients. Kidney Int 1985, 28: 830−838
6. Bleyer AJ, Kmoch S, Antignac C, Robins V, Kidd K, Kelsoe JR, Hladik G, Klemmer P, Knohl SJ, Scheinman SJ, Vo N, Santi A, Harris A, Canaday O, Weller N, Hulick PJ, Vogel K, Rahbari-Oskoui FF, Tuazon J, Deltas C, Somers D, Megarbane A, Kimmel PL, Sperati CJ, Orr-Urtreger A, Ben-Shachar S, Waugh DA, McGinn S, Bleyer AJ, Hodaňová K, Vylet'al P, Živná M, Hart TC, Hart SP: Variable clinical presentation of an MUC1 mutation causing medullary cystic kidney disease type 1. Clin J Am Soc Nephrol 2014, 9:527−535
7. Bleyer AJ, Hart SP, Kmoch S: Hereditary interstitial kidney disease. Semin Nephrol 2010, 30:366−373
8. Kiser RL, Wolf M, Martin JL, Zalewskib I, Attanasio M, Hildebrandt F, Klemmer P: Medullary cystic kidney disease type 1 in a large Native-American kindred. Am J Kidney Dis 2004, 44:611−617
9. Kirby A, Gnirke A, Jaffe DB, Barešová V, Pochet N, Blumenstiel B, et al: Mutations causing medullary cystic kidney disease type 1 (MCKD1) lie in a large VNTR in MUC1 missed by massively parallel sequencing. Nat Genet 2013, 45:299−303
10. Hanisch F, Muller S: MUC1: the polymorphic appearance of a human mucin. Glycobiology 2000, 10:439−449
11. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB: Characterizing and measuring bias in sequence data. Genome Biol 2013, 14:R51
12. Laehnemann D, Borkhardt A, McHardy AC: Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. Brief Bioinform 2016, 17:154−179