Supplementary information

Supplementary Materials and Methods

Genomic DNA extraction and Sanger sequencing

Genomic DNA (gDNA) was extracted from whole blood samples using the QIAamp DNA Blood Maxi Kit (Qiagen, Valencia, CA), from skin and semen samples using the QIAamp DNA Mini Kit (Qiagen) and from saliva using the ORAgene^{TM} DNA self-collection kit (DNA Genotek Inc, Ottawa, Canada) following each manufacturer's protocol.

Mutation analysis of the coding regions and exon-intron boundaries of *KRT1* and *KRT10* for gDNA from peripheral blood of the proband was performed by direct sequencing. First, standard PCR amplification procedures were employed, with high-fidelity polymerase, Platinum Taq DNA Polymerase (Invitrogen, Carlsbad, CA, USA) and 37.5 ng genomic DNA as the template in 50 μL volume. The PCR fragments were amplified with specific primers. Then, PCR products were purified by QIAquick PCR Purification Kit (Qiagen) and sequenced directly with the same primers as used for

each PCR amplification to identify mutations by Sanger sequencing with an Applied Biosystems 3730 DNA Analyzer.

Quantification of allele frequencies of the mutant c.446C>T (p.Arg156Cys) in *KRT10* in gDNA samples from the proband's father, including in semen, by next-generation sequencing

To quantify the allele frequencies of the mutant c.446C>T (p.Arg156Cys) in *KRT10* in gDNA samples from the proband's father, including in semen, we performed amplicon sequencing of gDNA samples from the father's peripheral blood, sperm and lesional skin by using the MiSeq Sequencing System and the MiSeq Reagent Kit v3 (Illumina, San Diego, CA).

To confirm the accuracy of this system, we also measured the mutant allele frequencies of 6 standard samples prepared by a mixture of gDNA from the proband and a healthy control (Table S2).

To prepare the sample for MiSeq sequencing, standard PCR amplification procedures were employed, with high-fidelity polymerase, KOD Plus Ver. 2

polymerase (Toyobo, Tokyo, Japan) and 37.5 ng gDNA as the template in 50-μl volume. The PCR were performed in triplicate using each primer set with sample- and PCR-specific tags (Table S1). The thermal conditions were the following: 94 ºC for 2 min, followed by 33 cycles at 98 ºC for 10 s, 65 ºC for 20 s. PCR products were purified by QIAquick Gel Purification Kit (Qiagen). The concentration of each purified PCR product was measured by Qubit™ 2.0 Fluorometer (Invitrogen, Carlsbad, CA). To obtain more than 100,000 reads per sample, we mixed all the samples. (The content of each sample was 1% of the total sample.) Mixed PCR products were injected through the inlet on the case of the MiSeq Reagent Kit v3. MiSeq Sequencing was performed according to the manufacturer's protocol. Obtained raw data was analyzed using the CLC genomics workbench (Qiagen). Sequence data were transferred to CLC genomics workbench and each sequence of the amplicon was mapped on the reference sequence. The sequences mapped to the desired position were collected and the unmapped sequences were discarded. Mapped sequence were exported as a CSV file and the mutation number was counted by using Microsoft Excel (Microsoft, Redmond, WA).

Supplementary results

Confirmation of the accuracy of mutant allele frequency measurements by next-generation sequencing

6 points of the mutant allele frequencies of standard samples prepared by mixing gDNA from the proband and a healthy control were measured. The results are summarized in Table S2. Part of the deep sequencing results for one PCR product from gDNA in the father's semen is shown in Figure S1.

Supplementary Tables

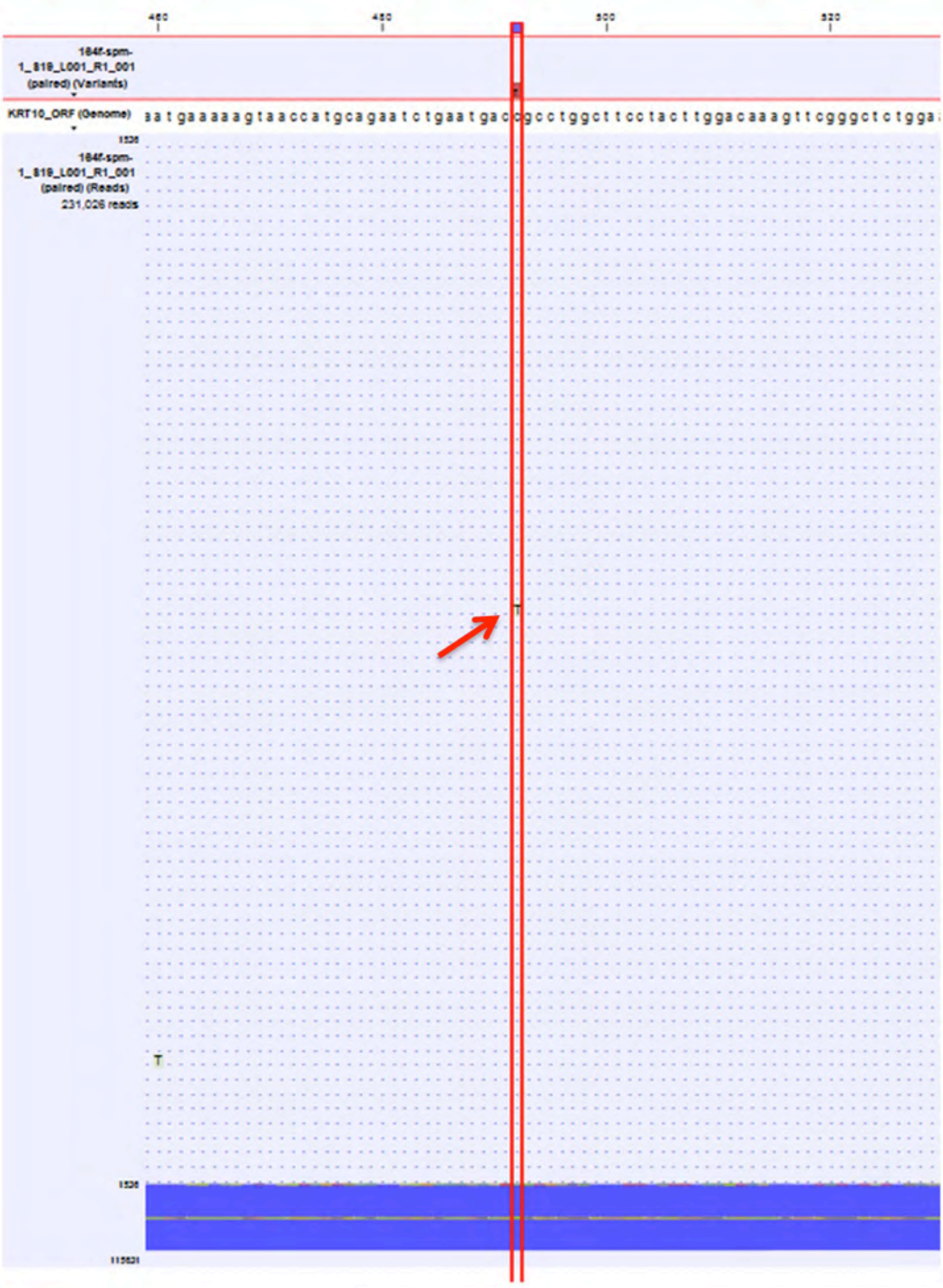
Table S1. Primer sequences with sample- and PCR-specific tags for preparation of PCR products for next-generation sequencing

Primer	Sequence
KRT10_c466CT_i501.f	AATGATACGGCGACCACCGAGATCTACAC-TATAGCCT-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-TGGTGGCCTTCTCTCTGGAAATGA
KRT10_c466CT_i502.f	AATGATACGGCGACCACCGAGATCTACAC-ATAGAGGC-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-TGGTGGCCTTCTCTCTGGAAATGA
KRT10_c466CT_i503.f	AATGATACGGCGACCACCGAGATCTACAC-CCTATCCT-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-TGGTGGCCTTCTCTCTGGAAATGA
KRT10_c466CT_i701.r	CAAGCAGAAGACGGCATACGAGAT-CGAGTAAT-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i702.r	CAAGCAGAAGACGGCATACGAGAT-TCTCCGGA-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i703.r	CAAGCAGAAGACGGCATACGAGAT-AATGAGCG-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i704.r	CAAGCAGAAGACGGCATACGAGAT-GGAATCTC-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i705.r	CAAGCAGAAGACGGCATACGAGAT-TTCTGAAT-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i706.r	CAAGCAGAAGACGGCATACGAGAT-ACGAATTC-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i707.r	CAAGCAGAAGACGGCATACGAGAT-AGCTTCAG-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i708.r	CAAGCAGAAGACGGCATACGAGAT-GCGCATTA-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i709.r	CAAGCAGAAGACGGCATACGAGAT-CATAGCCG-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA
KRT10_c466CT_i710.r	CAAGCAGAAGACGGCATACGAGAT-TTCGCGGA-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT-CCTTGATTTTGCCTTCCAGCTCATA

Tags are in red.

Table S2. Summary of the theoretical values and measured values of the allele frequencies in standard gDNA samples and gDNA samples from the proband and her father.

sample	average total reads number	average mutant reads number	mutant allele frequency theoretical value (%)	mutant allele frequency, measured value (%)
standard 12.8%	227181 ± 7743	25302 ± 915	12.8	11.1 ± 0.29
standard 5.16%	67009 ± 1047	2912 ± 14	5.16	4.35 ± 0.09
standard 2.35%	83495 ± 6122	1618 ± 148	2.35	1.93 ± 0.05
standard 1.12%	96667 ± 2298	820 ± 143	1.12	0.84 ± 0.14
standard 0.55%	116863 ± 6644	565 ± 43	0.55	0.49 ± 0.05
standard 0.27%	102738 ± 4372	259 ± 43	0.27	0.25 ± 0.03
gDNA from father's blood	99853 ± 5391	5258 ± 206	-	5.28 ± 0.11
gDNA from father's semen	122653 ± 5256	4839 ± 419	-	3.93 ± 0.17
gDNA from father's lesional skin	108441 ± 1508	12993 ± 381	-	12.0 ± 0.47
gDNA from patient's saliva	100825 ± 4650	50004 ± 2754	-	49.6 ± 0.47



Chromosome	Region	Type	Reference	Allele	Count	Coverage	Frequency
KRT10_ORF	492	SNV	C	T	4166	114114	3.65
KRT10_ORF	492	SNV	C	C	109838	114114	96.25

Figure S1. Part of the readout panel for one PCR product from gDNA in the father's semen. 114,114 sequences from deep sequencing of one triplicate amplicon from gDNA in the father's semen were mapped on the reference sequence including the mutation site c.446C in exon 1 of *KRT10*. 4,166 mutant sequences with the mutation c.446C>T (p.Arg156Cys) were detected among total 114,114 sequences. 74 representative reads out of 114,114 reads of amplicon sequencing are shown in the present panel. Only variant/mutant sequences are shown as letters, and sequences identical to the reference sequences are indicated by dots. Only one mutant sequence T (arrow) is seen in c.446 in this panel.