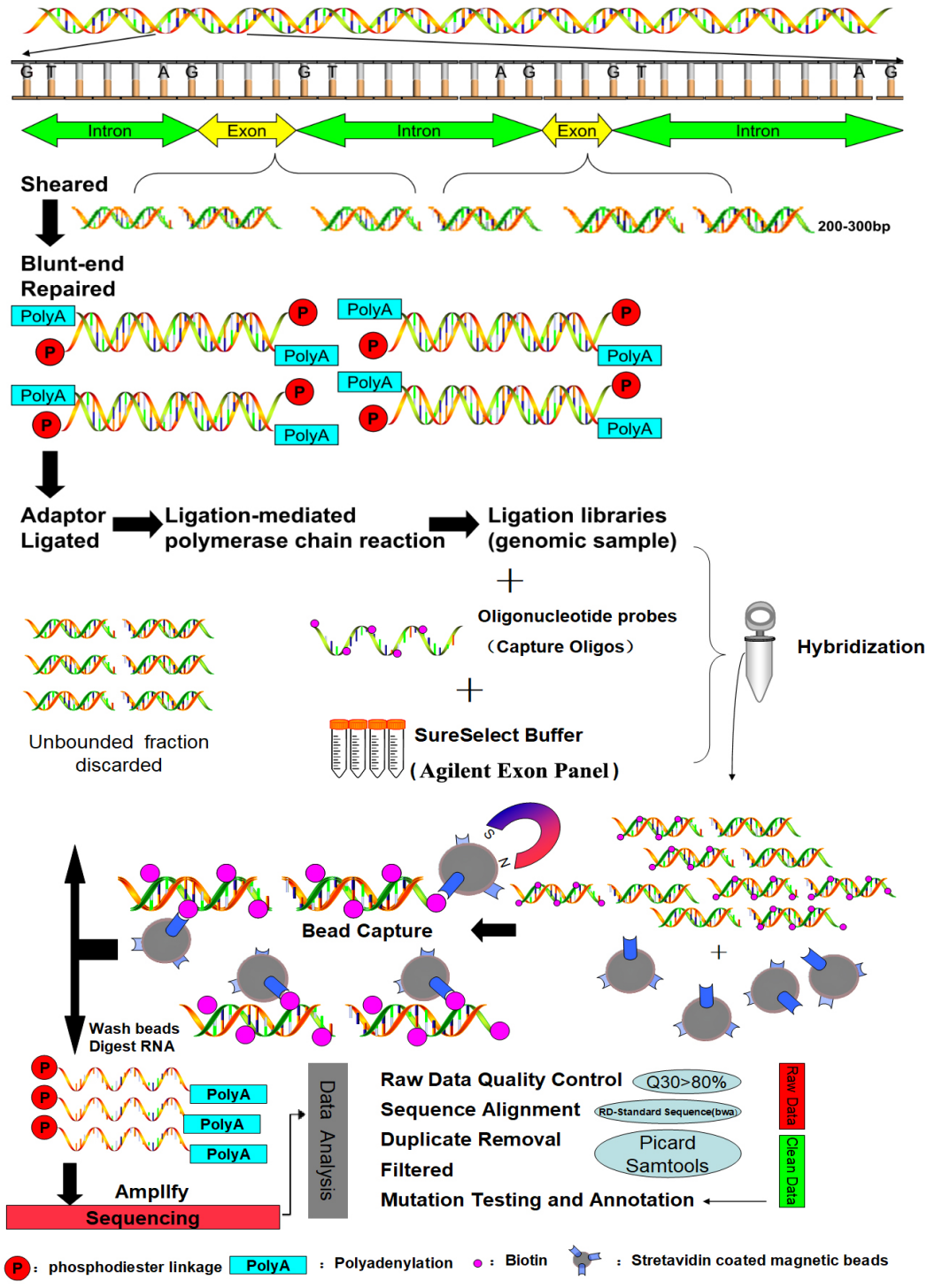
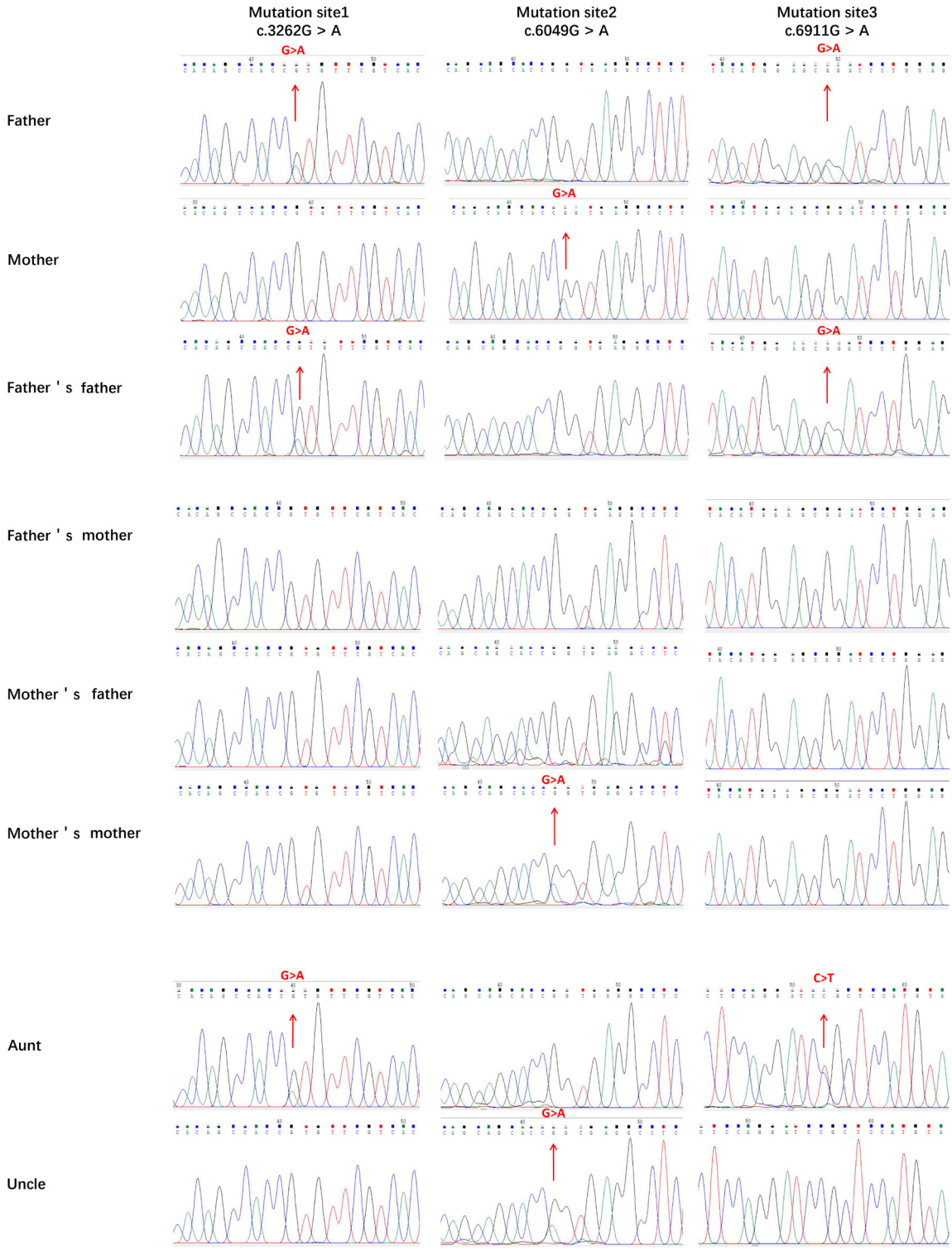


The standard workflow of cWES protocol

SUPPLEMENTARY FIG. 1. Flowchart of clinical whole exome sequencing. Genomic DNAs were isolated from peripheral blood samples of the proband, and her family members with a Blood DNA Kit (Tian Gen Biotech, Beijing, China) following the manufacturer's instructions. The integrity and quantity of the extracted leukocyte DNAs were checked via a NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) according to the manufacturer's protocol, and a 1% agarose gel electrophoresis respectively. Target enrichment was performed via the Agilent Human Customerized Exon Panel Target Enrichment System (Agilent Technologies, Inc., Santa Clara, CA, U.S.A.) followed by ligation-mediated amplification. The panel covers 3877 genes associated with 4269 inherited diseases, the lists of which are provided in the supplementary materials. The constructed DNA library was immediately sequenced on BioelectronSeq4000 in Capitalbio Medlab (Beijing, China), and the raw data were analyzed with a standard pipeline, including data cleaning, filtering, base calling and etc. Then, an array of annotation tools was subsequently employed to seek for the candidate genes with risk variations.

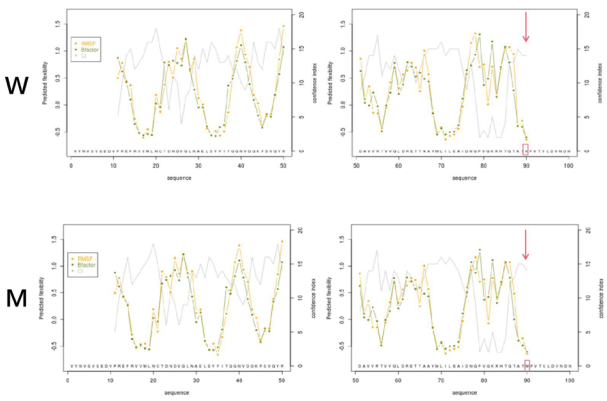
cWES, clinical whole exome sequencing.



SUPPLEMENTARY FIG. 2. Sanger sequencing on the familial members. The amplified products were recovered from agaroses with a PCR purification kit (TaKaRa, Japan), and sent to a commercial laboratory for Sanger sequencing (Sangon Biotech, Shanghai, China). The sequencing data were analyzed via Chromas software 2.23 and aligned against the wild-type *CDH23* sequences retrieved from the NCBI database. The sequencing chromatograms of the proband's familial members were all presented. The corresponding variations in *CDH23* were indicated by the red arrows, and the transmitted pattern of the compound heterozygous variations had been inferred from the sequencing results and the clinical phenotypes (also see Figure 1).

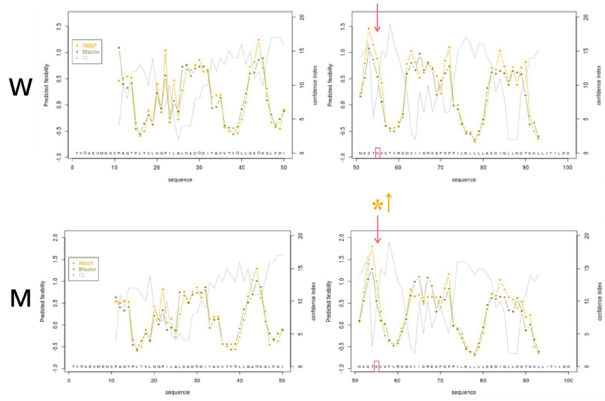
CDH23, cadherin-related 23.

a



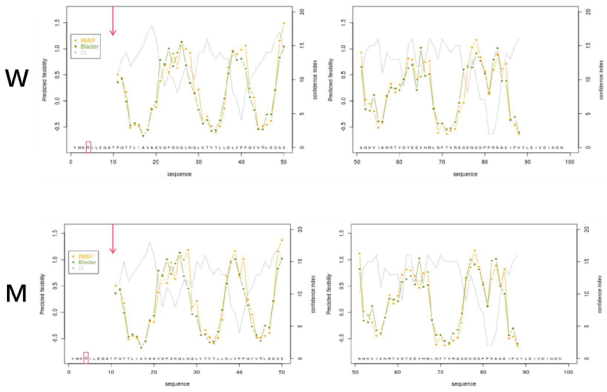
The flexibility of EC10

b



The flexibility of EC19*

c



The flexibility of EC21

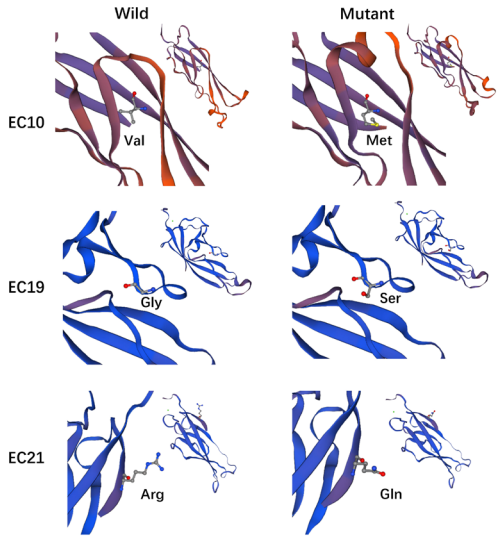
d

Gene	Domain	Sample	RSMD	CHARMm Energy
CDH23	EC10	wild	0.076	-4830.44145
		c.3262G>A p.Val1088Met		-4761.15461
	EC19*	wild	0.254	-1933.87297
		c.6049G>A p.Glu2017Ser		-1922.72739
	EC21	wild	0.059	-2409.00986
		c.6049G>A p.Arg2304Gln		-2274.36995

SUPPLEMENTARY FIG. 3. Effects of amino acidic residue substitutions on the flexibility of EC domains. PredyFlexy was used to deduce the impact of mutations on molecular flexibility of the *CDH23* EC domains with two parameters, including the B-factor and RMSF (root mean square fluctuation). The mutant residues are indicated by red arrows. The upper case letters W and M refer to the wild-type and mutant EC domains respectively, and the gray line represents the confidence index. Molecular flexibility of the *CDH23* p.Val1088Met, Glu2017Ser and p.Arg2304Gln mutants compared with the wild-type ones shown in (a), (b) and (c), respectively. (d) Calculations of the RMSD and Gibbs free energy (kJ/mol) for these structures. The protocol running molecular dynamic stimulation in Discovery Studio 2.5 was executed for conformational optimization by calculating the Gibbs free energy. Then, SWISS PDB Viewer was used to compute the RMSD.

CDH23, cadherin-related 23; EC, extracellular; RMSD, root-mean-square deviation.

a



The 3D structural models

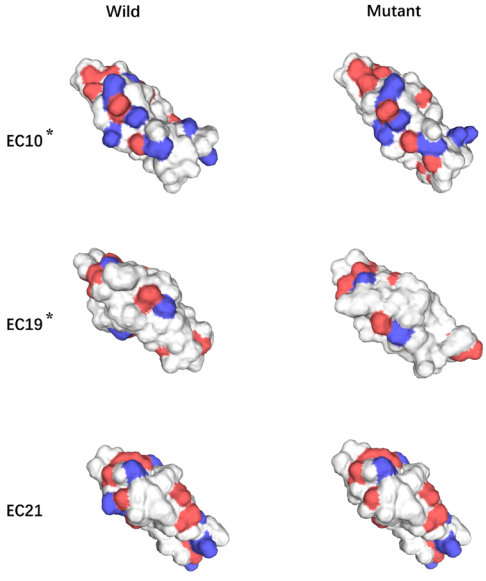
b

Calculation of PI and Mw

Gene	Domain	Sample	PI (isoelectric point)	Mw (molecular weight)
CDH23	EC10	wild	4.47	10913.2
		c.3262G>A p.Val1088Met	4.47	10945.26
	EC19	wild	3.68	10939.38
		c.6049G>A p.Glu2017Ser	3.68	10969.4
	EC21*	wild	4.2	10629.93
		c.6049G>A p.Arg2304Gln	4.08	10601.88
	EC10-EC21	wild	4.42	185286.58
		c.3262G>A、p.Val1088Met c.6049G>A、p.Arg2304Gln	4.41	185320.61

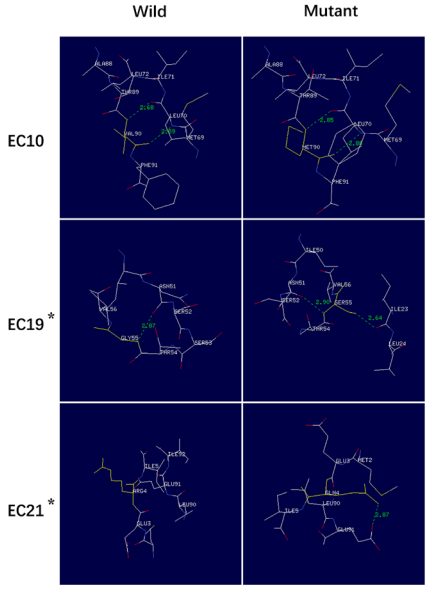
Isoelectric point and molecular weight

c



The surface charge distribution

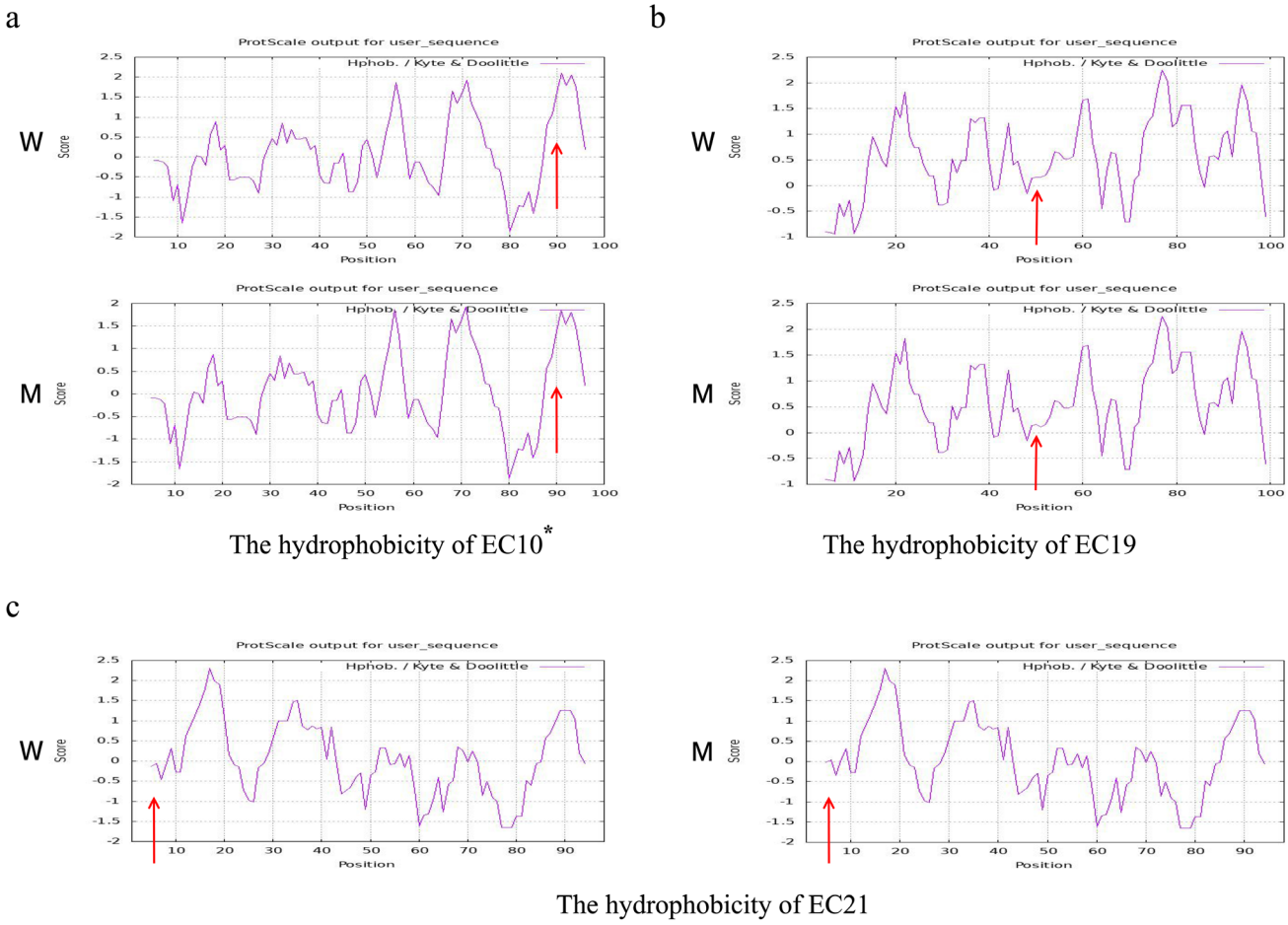
d



Analyses of hydrogen bonds

SUPPLEMENTARY FIG. 4. Structural bioinformatic analyses. (a) The 3D structures of the *CDH23* extracellular domain 10, 19 and 21 built via homology modeling. (b) Calculations of the isoelectric points and molecular weights of these polypeptides. (c) Changes in surface charge distribution (blue for positive, red for negative, white for neutral). (d) The influence of the substitutions on the hydrogen bonds. The green dashed lines represent hydrogen bonds, and the green Arabic numbers label the hydrogen bond distances (Å). In wild-type EC10, H-bonds formed between Val 90 and Leu 70 as well as Ile 71 having bond lengths of 2.68 Å and 2.59 Å, respectively. In mutant EC10, H-bonds formed between Met 90 and Leu 70 as well as Ile 71 having bond lengths of 2.85 Å and 2.88 Å, respectively. In wild-type EC19, H-bonds formed between Gly 55 and Asn 51, with a bond length of 2.87 Å. In mutant EC19, H-bonds formed between Ser 55 and Asn 51, as did Ile 23, with bond lengths of 2.9 Å and 2.64 Å, respectively. In native EC21, no H-bond formed between Arg 4 and other amino acidic residues, but in mutant EC21, a nascent H-bond formed between Gln 4 and Glu 91, with a bond length of 2.87 Å.

CDH23, cadherin-related 23; 3D, three-dimensional, EC, extracellular.



SUPPLEMENTARY FIG. 5. The hydrophobicity analyses were performed with ProtScale. (a) Hydrophobicity of the *CDH23* p.Val1088Met mutant domain compared with the wild-type domain. (b) Hydrophobicity of the *CDH23* p.Glu2017Ser mutant domain compared with the wild-type domain. (c) Hydrophobicity of the *CDH23* p.Arg2304Gln mutant domain compared with the wild-type domain. The mutant sites, including those in EC10, EC19 and EC20, are all indicated by red arrows. The uppercase letters of W and M denote the wild-type and mutant *CDH23* EC domains, respectively.

EC, extracellular; CDH23, cadherin-related 23