(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2014 (03.01.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/005076 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2013/048688

(22) International Filing Date:
28 June 2013 (28.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/666,423 29 June 2012 (29.06.2012) US
61/672,916 18 July 2012 (18.07.2012) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF MICHIGAN** [US/US]; 1600 Huron Parkway, 2nd Floor, Ann Arbor, Michigan 48109 (US).

(72) Inventor; and
(71) Applicant : **HILDEBRANDT, Friedhelm** [DE/US]; 580 Riverview Drive, Ann Arbor, Michigan 48104 (US).

(74) Agent: **ARENSON, Tanya A.**; Casimir Jones, S.C., 2275 Deming Way, Suite 310, Middleton, Wisconsin 53562 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

WO 2014/005076 A2

(54) **Title**: METHODS AND BIOMARKERS FOR DETECTION OF KIDNEY DISORDERS

(57) **Abstract**: Provided herein are compositions, kits, methods and biomarkers for detection and characterization and diagnosis of kidney disorders (e.g., one or more of nephronophthisis-related ciliopathies (NPHP-RC), congenital abnormalities of the kidney and urinary tract (CAKUT)), and karyomegalic interstitial nephritis (KIN)) in biological samples (e.g., tissue samples, blood samples, plasma samples, cell samples, serum samples).

METHODS AND BIOMARKERS FOR DETECTION OF KIDNEY DISORDERS

This application claims priority to provisional applications 61/666,423, filed June 29, 2012 and 61/672,916, filed July 18, 2012, each of which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

Provided herein are compositions, kits, methods and biomarkers for detection and characterization and diagnosis of kidney disorders (e.g., one or more of nephronophthisis-related ciliopathies (NPHP-RC), congenital abnormalities of the kidney and urinary tract (CAKUT)), and karyomegalic interstitial nephritis (KIN)) in biological samples (e.g., tissue samples, blood samples, plasma samples, cell samples, serum samples).

BACKGROUND OF THE INVENTION

Nephronophthisis (NPHP) is a recessive cystic kidney disease that represents the most frequent genetic cause of end-stage kidney disease in the first three decades of life. NPHP-related ciliopathies (NPHP-RC) are single-gene recessive disorders that affect kidney, retina, brain and liver by prenatalonset dysplasia or by organ degeneration and fibrosis in early adulthood. Identification of recessive mutations in more than 10 different genes (NPHP1-NPHP10) (Attanasio et al., Nat Genet 39, 1018-1024 2007; Delous et al., Nat Genet 39, 875-881 2007; Hildebrandt et al., Nat Genet 17, 149-153 1997; Mollet et al., Nat Genet 32, 300-305 2002; Olbrich et al., Nat Genet 34, 455-459 2003; Otto et al., Am J Hum Genet 71, 1167-1171 2002; Otto, Nat Genet 37, 282-288 2005; Otto et al., Nat Genet 42, 840-850 2010a; Otto et al., Nat Genet 34, 413-420 2003; Otto et al., J. Am Soc Nephrol 19, 587-592 2008; Sayer et al., Nat Genet 38, 674-681 2006; Valente et al., Nat Genet 38, 623-625 2006) revealed that their gene products share localization at the primary cilia-centrosomes complex and mitotic spindle poles in a cell cycle dependent manner, characterizing them as retinal-renal "ciliopathies" (Ansley et al., Nature 425, 628-633 2003; Hildebrandt et al., N Engl J Med 364, 1533-1543 2011). Multiple signaling pathways downstream of cilia have been implicated in the disease mechanisms of NPHP-RC, including Wnt signaling (Germino, Nat Genet 37, 455-457 2005; Simons et al., Nat Genet 37, 537-543 2005) and Shh signaling (Huangfu and Anderson, Proc Natl Acad Sci U S A 102, 11325-11330 2005; Huangfu et al., Nature 426, 83-87 2003). However, despite convergence of ciliopathy pathogenesis at cilia and centrosomes it remains largely unknown what signaling pathways downstream of cilia

and centrosome function operate in the disease mechanisms that generate the NPHP-RC phenotypes.

SUMMARY OF THE INVENTION

Provided herein are compositions, kits, methods and biomarkers for detection and characterization and diagnosis of kidney disorders (e.g., one or more of nephronophthisis-related ciliopathies (NPHP-RC), congenital abnormalities of the kidney and urinary tract (CAKUT)), and karyomegalic interstitial nephritis (KIN)) in biological samples (e.g., tissue samples, blood samples, plasma samples, cell samples, serum samples).

Embodiments of the present invention provide a kit for detecting gene variants associated with nephronophthisis-related ciliopathies (NPHP-RC) in a subject, comprising (e.g., consisting essentially of): a) a first NPHP-RC informative detection reagent for identification of one or more variants in a first gene selected from, for example, meiotic recombination 11 homolog (MRE11), zinc finger protein 423 (ZNF423), or centrosomal protein 164kDa (CEP164); and b) a second NPHP-RC informative detection reagent for identification of one or more variants in a second gene selected from, for example, MRE11, ZNF423, CEP164, wherein the second gene is different than the first gene. In some embodiments, the variations result in loss of function mutations in the one or more genes. In some embodiments, the MRE11 mutation is c.1897C>T (p.R633X), the ZNF423 mutation is selected from, for example, c.2738C>T, c1518delC, or c.3829C>T (p.P913L, p.P506fzX43, or p.H1277Y)), and the CEP164 mutations are selected from, for example, c.32A>C, c.277C>T, c.1573C>T, c.1726C>T, or c.4383A>G (p.Q11P, p.R93W, p.Q525X, p.R576X, and p.X1460W). In some embodiments, the determining comprises detecting variant nucleic acids or polypeptides (e.g., using a nucleic acid detection method selected from, for example, sequencing, amplification or hybridization). In some embodiments, the first gene is MRE11 and the second gene is ZNF423 or CEP164. In some embodiments, the first gene is CEP164 and the second gene is ZNF423. In some embodiments, kits, uses and methods provide a third reagent that identifies variants in a third distinct gene. For example, in some embodiments, reagents identify MRE11, ZNF423, and CEP164. In some embodiments, reagents identify one or more mutations (e.g., 1, 2, 3, 4, 5 or more) in the above described genes. In some embodiments, the first and second reagents are nucleotide probes that specifically bind to the variants. In some embodiments, the first and second reagents are antibodies that specifically bind to polypeptides encoded by the variants. In some embodiments, the first reagent is a pair of primers for amplifying the first gene and the second reagent is a pair of primers for

amplifying the second gene. In some embodiments, the first and second reagents are sequence primers for sequencing the first and second genes.

Further embodiments provide methods and uses (e.g., utilizing the kits described herein) for detecting gene variants associated with nephronophthisis-related ciliopathies (NPHP-RC) in a subject; and b) diagnosing the subject with NPHP-RC when the variants are present in the sample. In some embodiments, the sample is, for example, a tissue sample, a cell sample, or a blood sample. In some embodiments, the determining comprises a computer implemented method. In some embodiments, the computer implemented method comprises analyzing variant information and displaying the information to a user. In some embodiments, the method further comprises the step of treating the subject for NPHP-RC (e.g., under condition such that at least one symptom of NPHP-RC is diminished or eliminated).

Additional embodiments of the present invention provide a kit for detecting gene variants associated with congenital abnormalities of the kidney and urinary tract (CAKUT) in a subject, comprising (e.g., consisting essentially of): a) a first CAKUT informative detection reagent for identification of one or more variants in a first gene selected from, for example, fraser syndrome 1 (FRAS1), FRAS1 related extracellular matrix protein 2 (FREM2), Ret proto-oncogene, (RET), or bone morphogenetic protein 4 (BMP4); and b) a second CAKUT informative detection reagent for identification of one or more variants in a second gene selected from, for example, FRAS1, FREM1, RET, or BMP4, wherein the second gene is different than the first gene. In some embodiments, the FRAS1 mutation is c.776T>G, c.299G>T, c.981G>A, c.7861C>T, or c.11268C>A (p.L259R, p.D998Y, p.R3273H, p.R2621X, or p.H3757Q), the FREM2 mutation is c.7013C>T (p.T2338I), the RET mutation is c.2110G>T (p.V704F), and the BMP4 mutation is c.362A>G (p.H121R). In some embodiments, the first gene is FRAS1 and the second gene is FREM1, RET or BMP4. In some embodiments, the first gene is FREM1 and the second gene is RET or BMP4. In some embodiments, the first gene is RET and the second gene is BMP4. In some embodiments, kits further comprise a third reagent that identifies variants in a third distinct gene and optionally a fourth reagent that identifies variants in a fourth distinct gene. For example, in some embodiments, reagents identify FRAS1, FREM2, and RET; FRAS1, FREM2, and BMP4; FREM2, RET, and BMP4; or FRAS1, FREM2, RET and BMP4. In some embodiments, reagents identify one or more mutations (e.g., 1, 2, 3, 4, 5 or more) in the above described genes.

In some embodiments, the present invention provides methods and uses for detecting gene variants associated with CAKUT, comprising a) contacting a sample from a subject with a variant detection assay (e.g., the above described kits), under conditions that the presence of a variant associated with CAKUT is determined; and b) diagnosing the subject with CAKUT when the variants are present in the sample.

In certain embodiments, the present invention provides kits, methods, and uses for identifying KIN in a subject. In some embodiments, the kits, methods and uses utilize a first KIN informative detection reagent for identification of a first mutation in a KAN gene; and optionally b) a second KIN informative detection reagent for identification of a second mutation in a KAN1 gene, wherein the second mutation is different than the first mutation (e.g., to identify, treat, or characterize KIN in a subject). In some embodiments, the first and second mutations are, for example, c.1234+2T>A, c.2036_7delGA, c.2245C>T, c.1375+1G>A, c.2616delA, c.1606C>T, c.2878G>A, c.2120G>A, c.2786A>C, c.2611T>C, c.2774_5delTT, or c.2810G>A (e.g., c.1234+2T>A and c.2036_7delGA; c.1234+2T>A and c.2245C>T; c.1375+1G>A and c.2616delA; c.1606C>T and c.2786A>C; c.1606C>T and c.2878G>A; c.2611T>C and c.2878G>A; or c.2774_5delTT and c.2810G>A). In some embodiments, the first and second mutations encode a polypeptide selected from, for example, p.Arg679Thrfs*5, p.Arg749*, p.Asp873Thrfs*17, p.Arg536*, p.Cys871Arg, p.Trp707*, p.Leu925Profs*25, p.Gln929Pro, p.Gly937Asp, p.Asp960Asn, or and a splice site mutation (e.g., a splice site mutation and p.Arg679Thrfs*5; a splice site mutation and p.Arg749*; a splice site mutation and p.Asp873Thrfs*17; p.Arg536* and p.Gln929Pro; p.Arg536* and p.Asp960Asn; p.Cys871Arg and p.Asp960Asn; or p.Leu925Profs*25 and p.Gly937Asp).

Additional embodiments will be apparent to persons skilled in the relevant art based on the teachings contained herein.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows identification of recessive mutations in MRE11, ZNF423 and CEP164 in NPHP-RC using homozygosity mapping and WER. Data regarding homozygosity mapping and mutations are shown for family A3471 with MRE11 mutation (A-B), family F874 with ZNF423 mutation (C-D), and family KKESH001 with CEP164 mutation (E-F). (A, C, E) Non-parametric lod scores (NPL) are plotted across the human genome in 3 families (A3471, F874 and KKESH001) with NPHP-RC. (B, D, F) Homozygous mutations

detected in families with NPHP-RC. Family number (underlined), mutation (arrowheads) and predicted translational changes (in parenthesis) are indicated.

Figure 2 shows two ZNF423 mutations have dominant negative characteristics (A-D), ZNF423 mutation abrogates interaction with PARP1 (E), and ZNF423 directly interacts with the NPHPRC protein CEP290/NPHP6 (F-G). (A) Amino acid residues altered by NPHP-RC mutations in ZNF423 are drawn in relation to functional annotation of its 30 Zn-fingers. (B-D) S-phase index assay (fraction of transfected cells incorporating BrdU) for P19 cells transfected with either wildtype or mutant ZNF423. (B) Representative field of cells transfected with wildtype ZNF423 shows high frequency of BrdU+ FLAG+ double-positive cells. (C) ZNF423–H1277Y transfected cells exhibits fewer FLAG–positive cells and a lower proportion that are double positive. (D) S-phase index measured in duplicate transfections for each of three DNA preparations per construct. (E) ZNF423 interacts with PARP1. (F-G) ZNF423 directly interacts with CEP290/NPHP6. (F) A human fetal brain yeast two-hybrid library screened with human CEP290/NPHP6 (JAS2; aa 1917-2479) fused to the DNA-binding domain of GAL4 (pDEST32) identified ZNF423 as a direct interaction partner of CEP290/NPHP6. The interaction was confirmed using direct yeast two-hybrid assay where 1 and 2 represent colony growth of CEP290 bait with ZNF423 prey. (G) HEK293T were cotransfected with human V5-tagged partial human V5-CEP290 clone and GFPtagged full-length human ZNF423 clone.

Figure 3 shows (A-D) expression of mutant CEP164 in renal epithelial cells abrogates localization to centrosomes. (A) Immunofluorescence using α-SDCCAG8/NPHP10-CG antibody, labels both centrioles, whereas α-CEP164-ENR antibody demonstrates CEP164 staining at the mother centriole only. (B) Inducible overexpression of N-terminally GFP-tagged human full-length CEP164 isoform 1 (NGFPCEP164-WT) in IMCD3 cells demonstrates, in addition to a cytoplasmic expression pattern, localization at one of the two centrioles (inset, arrow heads). (C) In contrast, the centrosomal signal is abrogated upon overexpression of an N-terminally GFP-tagged truncated CEP164 construct representing the mutation p.Q525X. (D) The number of centrosomes positive for CEP164 is reduced upon overexpression of C-terminally GFPtagged human full-length CEP164 isoform 1 (CGFP-hCEP164-WT), which mimics the mutation p.X1460fs57 that causes a read-through of the stop-codon X1460, adding 57 aberrant amino acid residues to the Cterminus of CEP164. (E-H) Knockdown of Cep164 disrupts ciliary frequency. (E) Depletion of Cep164 by siRNA (F) causes a ciliary defect in 3D spheroid growth assays. (G) Nuclei and cilia were scored within a single spheroid to generate ciliary frequencies. siCep164 transfected cells manifest lower

cilia frequencies (33%) compared to control transfected IMCD3 cells (49%). (H) Ciliary frequency is not rescued by mutant CEP164. Ciliary frequencies are reduced in siCep164 transfected IMCD3 cells (39%) compared to control siCtrl transfected IMCD3 cells (54%).

Figure 4 shows (A-P) colocalization upon immunofluorescence of the NPHP-RC proteins SDCCAG8/NPHP10, ZNF423 and CEP164 to nuclear foci that are positive for the DDR signaling proteins SC35, TIP60 and Chk1 in hTERT-RPE cells. (A-G) Colocalization of NPHP-RC proteins with SC35 in nuclear foci. SDCCAG8/NPHP10 (A-C) and ZNF423 (D) fully colocalize to nuclear foci with SC35, and (E) CEP164 partially colocalizes with SC35. SDCCAG8/ NPHP10 also colocalizes with the newly identified NPHP-RC proteins ZNF423 (F) and CEP164 (G). (H-J) Colocalization of NPHP-RC proteins with the DDR protein TIP60 I nuclear foci. (H) TIP60 fully colocalizes with SC35. (I) TIP60 partially colocalizes with CEP164. (J) Chk1 fully colocalizes with SC35/ SRSF2. DNA is stained in blue with DAPI. (K-P) Colocalization of DDR and NPHP proteins upon induction of DDR by UV radiation in HeLa cells. (K) Following irradiation of HeLa cells with UV light at 20 J/m2 a strong immunofluorescence signal of an anti-γH2AX antibody indicates activation of DDR. (L-M) Upon irradiation with UV light, CEP164-positive nuclear foci condense and colocalize with newly appearing TIP60 foci of similar size. (N-O) In untreated cells (N) a pattern of broad CEP164 speckles, which are CHK1-negative and locate to DAPI-negative domains, changes to a pattern of multiple smaller foci (O) that are doublepositive for both, CEP164-N11 and CHK1. (P) p317-CHK1 fully colocalizes with TIP60 to nuclear foci and to the centrosome.

Figure 5 shows knockdown of Cep164 causes anaphase lag and retarded cell growth. (A-B) Knockdown of CEP164 causes anaphase lag. (B). Bars represent SEM, p values (student T-test) are indicated above the bar graph. (C-D) Transient knockdown of Cep164 inhibits proliferation, which is rescued by wild type but not mutant CEP164. (C) IMCD3 cells depleted of murine Cep164 grew more slowly (siRNA) than non-depleted cells (control) or the non-depleted cells induced to express human wild type CEP164 (Dox). Expression of WT Cep164 in siRNA-depleted cells rescued the slow growth phenotype of Cep164 depletion (siRNA+dox). (D) As in C, except mutant Cep164 cDNA (CEP164-Q525X) was expressed under doxycyclin control. (E-H) The effect of roscovitine on UV-induced DDR. (E-F) Immunofluorescence analysis showed that roscovitine triggered uniform nuclear distribution of γH2AX (activated H2AX phosphorylated at Ser139) in non UVirradiated cells showing partial DDR activation (E). UV radiation caused enhanced γH2Ax staining with a prominent

nuclear foci pattern, characteristic of strong DDR activation (F). (G-H) The effect of roscovitine on UV-triggered subcellular localization of CEP164 and Chk1.

Figure 6 shows that knockdown of cep164 in zebrafish embryos results in ciliopathy phenotypes, and knockdown of Cep164 or Zfp423(Znf423) causes sensitivity to DNA damage. (A-E) Whereas p53 MO injection (n=67) did not produce any phenotype (A), coinjection of cep164 MO at 28 hpf caused the mild ciliopathy phenotype of ventral body axis curvature in 48% of embryos (60/125) (B). 50% of embryos (62/125) showed severe cell death throughout the body as judged by grey-appearing cells in the head region (C). Embryos with severe cell death also showed increased expression of phosphorylated γH2AX (D) compared to p53 MO control (E). Most embryos with massive cell death did not survive beyond 48 hpf. (F-I) At 48 hpf, surviving cep164 morphants displayed the ciliopathy phenotype of laterality defects. Whereas p53 MO did not cause any abnormal heart looping (F,G), cep164 MO caused inverted heart looping (H) or ambiguous heart looping (I). (A, atrium; L, left; V, ventricle). (J-M) At 72 hpf, embryos developed further ciliopathy phenotypes. When compared to p53 MO controls (J), pronephric tubules (arrow heads) exhibited cystic dilation (K, asterisks) in 25% (7/28) of embryos. In addition, when compared to p53 MO controls (L), 0% (0/67) of embryos showed kidney cysts, hydrocephalus (asterisk), or retinal dysplasia (brackets) (M). (N) At 0.2 mM, cep164 MO knockdown effectively altered mRNA processing as revealed by RT-PCR. (O) Quantification of γ-H2AX levels in cep164 MO morphants. (P-Q) Cep164-deficient IMCD3 cells exhibit radiation sensitivity. (R-T) Zfp423(Znf423)-deficient P19 cells exhibit radiation sensitivity.

Figure 7 shows links of newly identified nephronophthisis-related ciliopathy (NPHP-RC) proteins to DNA damage response (DDR) signaling and cell cycle control.Columns depict two pathways of DDR signaling, the ATM (ataxia-telangiectasia mutated) pathway (A), and the ATR (ATM-and-Rad-related) pathway (C). Rows depict stages of DDR signaling including DNA damage sensing and repair (A, C), as well as outcomes regarding checkpoint activation, cell cycle arrest, and apoptosis (D).

Figure 8 shows alternative transcripts, knockdown targets, and human mutations of CEP164 with interacting domains, interaction partners, and antigens of the Cep164 protein. (A) The CEP164 gene extends over 85.4 kb, contains 33 exons (vertical hatches) and an alternative exon 5a used in isoform 2. (B) Exon structure of human CEP164 cDNA. Positions of start codon (ATG) and of stop codon (TGA) are indicated. (C) Domain structure of the Cep164 protein. (D) Minimal segment to which CEP164 the interaction partner ATRIP has

been mapped. (E) Extent of antigens used for α-CEP164 antibody production. (F) CEP164 mutations detected in 4 families with NPHP-RC.

Figure 9 shows subcellular localization of CEP164. (A) Characterization of anti-CEP164 antibodies by immunoblotting. (B) Characterization of α-CEP164 antibodies by overexpression of N-terminally GFP-labeled human isoform 1 wild type construct EGFP-CEP164-WT and immunofluorescence in cell lines. (C) CEP164 in GFP-Centrin-2 mouse photoreceptors. (D-G) Expression of mutant CEP164 in hTERT-RPE cells abrogates localization to centrosomes. (D) Doxycyclin (Dox)-inducible overexpression of N-terminally GFP-tagged human full-length CEP164 wild type isoform 1 (NGFP-hCEP164-WT) in hTERT-RPE (human retinal pigment epithelium) cells demonstrates, in addition to a cytoplasmic expression pattern, localization at centrosomes. (E) In contrast, the signal at centrosomes is abrogated upon overexpression of an N-terminally GFP-tagged truncated CEP164 construct (NGFP-CEP164-Q525X) representing the mutation p.Q525X occurring in NPHP-RC family F59. (F) Expression of GFP alone as a negative control yielded no centrosomal expression in IMCD3 or hTERTRPE cells. (G) Immunoblot of IMCD3 and hTERT-RPE cells expressing inducible GFP-CEP164 constructs confirmed doxycyclin-inducible NGFP-CEP164 fusion protein expression (~191 kDa) both with α-CEP164-NR and α-GFP antibody blotting (arrow heads).

Figure 10 shows subcellular localization of CEP164 and SDCCAG8 by immunofluorescence. (A) CEP164 colocalizes with the mother centriole (labeled with γ-tubulin), the mitotic spindle poles (actetylated tubulin), and the midbody throughout the cell cycle in hTERT-RPE cells. (B-D) Upon immunofluorescence (IF) in hTERT-RPE cells antibody α-SDCCAG8-CG recognizes nuclear foci that are absent upon transient SDCCAG8/NPHP10 knockdown. (B) hTERT cells transiently transfected with GFP-labeled negative control shRNA construct exhibit nuclear foci upon IF with α-SDCCAG8-CG (arrow heads in right panel), whereas in cells transfected with GFP-labeled SDCCAG8 shRNA knockdown constructs pGIPZ-259547 (C) or pGIPZ-90858 (D) nuclear foci are absent (asterisks in right panels of C and D), demonstrating specificity of the nuclear foci signal detected by α-SDCCAG8-CG.

Figure 11 shows (A-B) Immuofluorescence imaging of the NPHP-RC protein SDCCAG8/NPHP10 with other proteins of nuclear foci in hTERT-RPE cells, and (C-J) identification of CCDC92 and TTKB2 as direct interaction partners of CEP164. (A-B) SDCCAG8 labeled with antibody α-SDCCAG8-CG exhibits a similar number and size of nuclear foci in comparison to signals from antibodies against the nucler foci markers

promyelocytic leukemia protein (PML) (A) and centromere protein C (CENP-C) (B). (C-J) Identification of CCDC92 and TTKB2 as direct interaction partners of CEP164. (C) Four baits, BD-CEP164fl, BD-CEP1641-550, BD-CEP164551-1100, and BD-CEP1641101-1640 were used to screen two retinal cDNA libraries, a human oligo-dT primed library and a bovine randomly primed library, employing a GAL-4 based yeast two-hybrid system. (D) Lysates from COS-1 cells transfected with 3xFlag-CCDC92fl or 3xFlag-TTBK2fl were used in a pulldown assay of GST fusion proteins GST-CEP164fl (191kDa), GST-CEP1-550 (87kDa), GST-CEP164551-1100 (92kDa), and GST-CEP1641101-1640 (68kDa). (E-F) For co-immunoprecipitation 3xHA-CEP164fl in combination with 3xFlag-CCDC92fl were overexpressed in COS-1 cells (10% protein input shown). (G-H) For co-immunoprecipitation 3xHA-CEP164fl in combination with 3xFlag-TTBK2fl were overexpressed in COS-1 cells (10% protein input shown). (I-J) (I) The α-CCDC92 antibody signal fully colocalizes with α -CEP164-M26 at the mother centriole upon immunofluorescence imaging in hTERT-RPE cells. (J) TTBK2 weakly colocalizes with α -CEP164-M26 at one of the centrioles, but yields a strong signal at the mid body in dividing hTERT-RPE cells.

Figure 12 shows molecular interaction of NPHP-RC gene products with DDR proteins. (A-B) Interaction of NPHP3 with CEP164. (C-D) Interaction of NPHP2 with DDB1. (E-H) Cep164 and Dvl3 are in a precipitable complex.

Figure 13 shows sequence chromatograms of 11 confirmed variants identified in 11 individuals with unilateral renal agenesis.

Figure 14 shows renal histology in individuals with KIN. (a–c) Renal histology of individuals with FAN1 mutation shows the characteristic triad of nephronophthisis, with cystic dilation of renal tubules (asterisks), interstitial infiltrations (dotted circle in a), and widespread fibrosis (b). Karyomegaly is observed (arrowheads in b and c) in tubules that have lost epithelial cells at their circumference (black arrowheads in b and c). The tubular basement membrane is thickened (double arrows in c) as well as attenuated (black arrow in c). a and b are from individual A4393-21; c is from individual A4466-21.

Figure 15 shows phenotypes of FAN1-mutant cells. (a) Protein blot analysis with antibody raised against the N-terminal 90 aa of FAN1. (b) Examples of metaphase spreads of the indicated cell lines after exposure to 50 nM MMC. (c,d) Sensitivity to ICL of the indicated FAN1-mutant (mut) cell lines in comparison to FANCA-mutant and wild-type cell lines. Primary fibroblasts (c) or LCLs (d) were treated in triplicate with increasing concentrations of MMC (left) or DEB (right). (e) Cell cycle analysis of the indicated fibroblast cell lines after treatment with 100 nM MMC or 0.1 μg/ml of DEB.

Figure 16 shows complementation of FAN1-mutant cells with FAN1 cDNA and epistasis analysis with genes implicated in ICL resistance. (a) Complementation of MMC sensitivity in fibroblasts of individual A1170-22 with KIN. (b) Immunoblot showing expression of FAN1 mutants in A1170-22 fibroblasts used in the MMC sensitivity assay in a. (c) MMC sensitivity in A1170-22 fibroblasts transfected with the indicated small interfering RNAs (siRNAs) (d) Immunoblots of expression of the indicated proteins after siRNA-mediated knockdown in A1170-22 fibroblasts from the experiment in c.

Figure 17 shows phenotypes caused by loss of fan1 function in zebrafish. (a–c) Knockdown of fan1 in zebrafish causes developmental abnormalities. (d,e) Knockdown of fan1 in zebrafish induces widespread apoptosis. (f–h) Knockdown of fan1 on the background of p53 morphants reveals pronephric cysts.

Figure 18 shows differential expression of FANCD2 and FAN1 in human tissues and greater DDR in CKD. (a,b) Expression of FANCD2 and FAN1 in 25 different human tissues. (c–h) DDR is pronounced in CKD.

DEFINITIONS

To facilitate an understanding of the present invention, a number of terms and phrases are defined below:

As used herein, the term "sensitivity" is defined as a statistical measure of performance of an assay (e.g., method, test), calculated by dividing the number of true positives by the sum of the true positives and the false negatives.

As used herein, the term "specificity" is defined as a statistical measure of performance of an assay (e.g., method, test), calculated by dividing the number of true negatives by the sum of true negatives and false positives.

As used herein, the term "informative" or "informativeness" refers to a quality of a marker or panel of markers, and specifically to the likelihood of finding a marker (or panel of markers) in a positive sample.

As used herein, the terms "NPHP-RC informative reagent" "KIN informative reagent" or "CAKUT informative reagent" refers to reagents that are informative for identification of variants or mutations in one or more of the markers described herein. In some embodiments, reagents are primers, probes or antibodies for detection of mutant or variant alleles of the markers described herein.

As used herein, the term "adverse outcome" refers to an undesirable outcome in a patient diagnosed with NPHP-RC, KIN, or CAKUT. In some embodiments, the patient is

undergoing or has undergone treatment for NPHP-RC, KIN, or CAKUT. Examples of adverse outcome include but are not limited to, recurrence of NPHP-RC, KIN or CAKUT, progression of disease, disability, or death.

As used herein, the term "amplicon" refers to a nucleic acid generated using one or more primers (e.g., two primers). The amplicon is typically single-stranded DNA (e.g., the result of asymmetric amplification), however, it may be RNA or dsDNA.

The term "amplifying" or "amplification" in the context of nucleic acids refers to the production of multiple copies of a polynucleotide, or a portion of the polynucleotide, typically starting from a small amount of the polynucleotide (*e.g.*, a single polynucleotide molecule), where the amplification products or amplicons are generally detectable.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced (*e.g.*, in the presence of nucleotides and an inducing agent such as a biocatalyst (*e.g.*, a DNA polymerase or the like) and at a suitable temperature and pH). The primer is typically single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is generally first treated to separate its strands before being used to prepare extension products. In some embodiments, the primer is an oligodeoxyribonucleotide. The primer is sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method. In certain embodiments, the primer is a capture primer.

A "sequence" of a biopolymer refers to the order and identity of monomer units (e.g., nucleotides, etc.) in the biopolymer. The sequence (e.g., base sequence) of a nucleic acid is typically read in the 5' to 3' direction.

As used herein, the term "subject" refers to any animal (*e.g.*, a mammal), including, but not limited to, humans, non-human primates, rodents, and the like, which is to be the recipient of a particular treatment. Typically, the terms "subject" and "patient" are used interchangeably herein in reference to a human subject.

As used herein, the term "non-human animals" refers to all non-human animals including, but are not limited to, vertebrates such as rodents, non-human primates, ovines, bovines, ruminants, lagomorphs, porcines, caprines, equines, canines, felines, aves, etc.

The term "locus" as used herein refers to a nucleic acid sequence on a chromosome or on a linkage map and includes the coding sequence as well as 5' and 3' sequences involved in regulation of the gene.

DETAILED DESCRIPTION OF THE INVENTION

Provided herein are compositions, kits, methods and biomarkers for detection and characterization and diagnosis of kidney disorders (e.g., one or more of nephronophthisis-related ciliopathies (NPHP-RC), congenital abnormalities of the kidney and urinary tract (CAKUT)), and karyomegalic interstitial nephritis (KIN)) in biological samples (e.g., tissue samples, blood samples, plasma samples, cell samples, serum samples).

Recessive null mutations in certain NPHP genes cause severe, congenital-onset phenotypes (Meckel syndrome) of dysplasia and malformation in kidney (polycystic dysplastic kidneys), eye (coloboma/microphthalmia), cerebellum (vermis hypoplasia in Joubert syndrome), and liver (cysts, ductal plate malformation), whereas hypomorphic mutations in the same gene cause late-onset, degenerative phenotypes such as renal tubular degeneration with fibrosis (nephronophthisis), retinal degeneration (Senior-Loken syndrome; SLSN), and liver fibrosis (Chaki et al., Kidney Int 2011; Hildebrandt et al., N Engl J Med 364, 1533-1543 2011). The more than 10 known NPHP genes explain less than 50% of all cases with NPHP-RC, and that many of the single-gene causes of NPHP-RC are still unknown (Otto et al., J. Med Genet 2010b). Some of the recently identified genetic causes of NPHP-RC are exceedingly rare (Attanasio et al., Nat Genet 39, 1018-1024 2007).

Embodiments of the present invention utilized homozygosity mapping with whole exome resequencing (WER) (Otto et al., Nat Genet 42, 840-850 2010a) to identify multiple different causes of NPHP-RC within a short time frame. Experiments described herein demonstrate that NPHP-RC proteins play a role in DDR signaling by demonstrating, i) mutation of MRE11, ZNF423 and CEP164 in NPHP-RC individuals; ii) abrogation by an NPHP-RC mutation of the ZNF423 interaction with the DDR protein PARP1; iii) colocalization of the centrosomal proteins ZNF423, CEP164, and SDCCAG8/NPHP10 to nuclear foci with the DDR protein TIP60; iv) cellular sensitivity to DNA damaging agents upon knockdown of ZNF423 or CEP164; and v) occurrence of dysregulated DDR and an NPHP-RC phenotype in zebrafish upon knockdown of cep164.

Further experiments conducted during the course of development of embodiments of the present invention identified seven different mutations of four different genes in 7 out of 40 patients with CAKUT, using massively parallel exon resequencing. Two heterozygous

mutations were novel mutations in known human CAKUT-causing genes: BMP4 and RET. Five heterozygous mutations were in two candidate genes that have never been reported in nonsyndromic CAKUT: FRAS1 and FREM2. McGregor et al. (Nat Genet 2003; 34: 203–208) and Vrontou et al. (Nat Genet 2003; 34: 209–214) identified recessive mutations in FRAS1 using linkage analysis in families that were affected with Fraser syndrome, a rare multi-organ disorder characterized by cryptophthalmos, cutaneous syndactyly, and renal agenesis.

Fras1 encodes Fras1, a protein containing repeats of the chondroitin sulfate proteoglycan core domain, the function of which is to maintain epithelial cell integrity. Fras1 protein was detected in several developing tissues such as limb, lung, gut, and kidney. In metanephros, Fras1 was detected in the extracellular matrix underlying the basal surface of outgrowing ureter and the basement membrane of collecting tubule. All FRAS1 mutations that have been reported in Fraser syndrome individuals were either homozygous or compound heterozygous, indicating a recessive genetic mechanism. Most mutations were truncating (McGregor et al., supra; Slavotinek et al., J Med Genet 2002; 39: 623–633; van Haelst et al., Am J Med Genet A 2008; 146A: 2252–2257; van Haelst et al., Am J Med Genet A 2007; 143A: 3194–3203).

Frem2, a FRAS1-related extracellular matrix protein 2 gene, was identified as segregating in the mouse myUcl strain that showed phenotypes similar to that of Fras1/mutant mice. Frem2 was expressed in mesonephric and metanephric epithelium, especially in the ureteric bud (Jadeja et al., Nat Genet 2005; 37: 520–52). Frem2 shares several structural domains with Fras1, including the core region of 12 chondroitin sulfate proteoglycans. Jadeja et al. (supra) detected a homozygous mutation in FREM2 in two families with Fraser syndrome not linked to FRAS1. To date, there have been only two different homozygous mutations of FREM2 reported, representing three unrelated families with Fraser syndrome (Jadeja et al., supra; Shafeghati et al., Am J Med Genet A 2008; 146A: 529–531).

Nephronophthisis (NPHP)-related ciliopathies are a heterogeneous group of recessive diseases that cause CKD through chronic fibrosis and cyst development in the kidney (Hildebrandt et al., N. Engl. J. Med. 364, 1533–1543 (2011)). To identify additional causative genes for NPHP, experiments described herein of homozygosity mapping (Hildebrandt, F. et al. PLoS Genet. 5, e1000353 (2009)) and exome sequencing (Otto, E.A. et al.Nat. Genet. 42, 840–850 (2010)) in two siblings (from family A1170) of Maori descent (Table 3) showed that renal histology in these individuals was indistinguishable from that associated with NPHP,

except for the presence of enlarged nuclei, as seen in karyomegalic interstitial nephritis (KIN) (Fig. 14). Variants in FANCD2/FANCI-associated nuclease 1 (KAN1) were identified as associated with KIN. KIN is a kidney disease with renal tubular degeneration of unknown origin that was first described in 1974 (Burry, J. Pathol. 113, 147–150 (1974)). KIN causes CKD through renal histological changes that are characteristic of NPHP2, including tubular basement membrane degeneration, atrophic tubules, tubular microcysts, interstitial infiltrations and pronounced fibrosis (Fig. 14a) (Zollinger, H.U. et al. Helv. Paediatr. Acta 35, 509–530 (1980)). The main feature that distinguishes KIN from NPHP is the presence of karyomegaly (Fig. 14b,c), which can also be present in the lung, liver and brain (Spoendlin, M. et al. Karyomegalic interstitial nephritis: further support for a distinct entity and evidence for a genetic defect. Am. J. Kidney Dis. 25, 242–252 (1995)). To date, only 12 families with KIN have been described (Palmer et al., Diagn. Cytopathol. 35, 179–182 (2007); Baba et al., Pathol. Res. Pract. 202, 555–559 (2006)), which is compatible with an autosomal recessive mode of inheritance. Variants of KAN1 include, but are not limited to, p.Trp707*, p.Gln929Pro, p.Gly937Asp and p.Asp960Asn.

Diagnostic and Screening Applications

Embodiments of the present invention provide diagnostic, prognostic, and screening compositions, kits, and methods. In some embodiments, compositions, kits, and methods characterize and diagnose NPHP-RC and/or CAKUT. Exemplary, non-limiting reagents and methods vof identifying mutations associated with NPHP-RC and/or CAKUT are described below.

A. Mutations

Embodiments of the present invention provide compositions and methods for detecting mutations in one or more genes (e.g., to identify or diagnose NPHP-RC and/or CAKUT). The present invention is not limited to particular mutations. In some embodiments, mutations are loss of function mutations (e.g., truncation, nonsense, missense, or frameshift mutations).

Exemplary mutations include, but are not limited to, mutations in MRE11, ZNF423, and CEP164 associated with HPHP-RC, FRAS1, FREM2, RET and BMP4 associated with CAKUT, and KAN1 associated with KIN. For example, in some embodiments, the MRE11 mutation is c.1897C>T (p.R633X), the ZNF423 mutation is selected from, for example, c.2738C>T, c1518delC, or c.3829C>T (p.P913L, p.P506fzX43, or p.H1277Y)), and the

CEP164 mutations are selected from, for example, c.32A>C, c.277C>T, c.1573C>T, c.1726C>T, or c.4383A>G (p.Q11P, p.R93W, p.Q525X, p.R576X, and p.X1460W), the FRAS1 mutation is c.776T>G, c.299G>T, c.981G>A, c.7861C>T, or c.11268C>A (p.L259R, p.D998Y, p.R3273H, p.R2621X, or p.H3757Q), the FREM2 mutation is c.7013C>T (p.T2338I), the RET mutation is c.2110G>T (p.V704F), and the BMP4 mutation is c.362A>G (p.H121R). In some embodiments, the KAN1 variants include, but are not limited to, p.Trp707*, p.Gln929Pro, p.Gly937Asp and p.Asp960Asn.

While the present invention exemplifies several markers specific for detecting one or more of NPHP-RC, KIN, and CAKUT, any marker that is correlated with the presence or absence or prognosis of one or more of NPHP-RC, KIN, and CAKUT may be used. A marker, as used herein, includes, for example, nucleic acid(s) whose production or mutation or lack of production is characteristic of one or more of NPHP-RC, KIN, and CAKUT and mutations that cause the same effect (e.g., deletions, truncations, etc).

In some embodiments, one or more (e.g., 2, 3, 4, or all) of the above-described mutations are identified in order to diagnose or characterize one or more of NPHP-RC, KIN, and CAKUT. In some embodiments, mutations are identified in combination with one or more additional markers of one or more of NPHP-RC, KIN, and CAKUT. In some embodiments, multiple markers are detected in a panel or multiplex format.

For example, in some embodiments, one or more of the following combinations is utilized to identify NPHP-RC: a) one or more mutations in MRE11 and one or more mutations in ZNF423, one or more mutations in MRE11 in combination with one or more mutations in CEP164, one or more mutations in ZNF423 in combination with one or more mutations in CEP164, and one or more mutations in MRE11 in combination with one or more mutations in ZNF423 and one or more mutations in CEP164.

In some embodiments, one or more of the following combinations is utilized to identify CAKUT: a) one or more mutations in FRAS1 and one or more mutations in FREM2, one or more mutations in FRAS1 in combination with one or more mutations in RET, one or more mutations in FRAS1 in combination with one or more mutations in BMP4, one or more mutations in FREM2 and one or more mutations in RET, one or more mutations in FREM2 in combination with one or more mutations in BMP4, one or more mutations in RET in combination with one or more mutations in BMP4 , one or more mutations in FRAS1 in combination with one or more mutations in FREM2 and one or more mutations in RET, one or more mutations in FRAS1 in combination with one or more mutations in FREM2 and one or more mutations in BMP4, one or more mutations in BMP4 in combination with one or

more mutations in FREM2 and one or more mutations in RET, and one or more mutations in FRAS1 in combination with one or more mutations in FREM2, one or more mutations in RET, and one or more mutations in BMP4. One or more (e.g., 1, 2, 3, 4, 5, or all) mutations are assayed in each gene, alone or in combination.

In some embodiments, one or more (e.g., 1, 2, 3, 4, or all) KAN1 variants including, but are not limited to, c.1234+2T>A, c.2036_7delGA, c.2245C>T, c.1375+1G>A, c.2616delA, c.1606C>T, c.2878G>A, c.2120G>A, c.2786A>C, c.2611T>C, c.2774_5delTT, or c.2810G>A (e.g., c.1234+2T>A and c.2036_7delGA; c.1234+2T>A and c.2245C>T; c.1375+1G>A and c.2616delA; c.1606C>T and c.2786A>C; c.1606C>T and c.2878G>A; c.2611T>C and c.2878G>A; or c.2774_5delTT and c.2810G>A). In some embodiments, the varaints encode a polypeptide selected from, for example, p.Arg679Thrfs*5, p.Arg749*, p.Asp873Thrfs*17, p.Arg536*, p.Cys871Arg, p.Trp707*, p.Leu925Profs*25, p.Gln929Pro, p.Gly937Asp, p.Asp960Asn, or and a splice site mutation (e.g., a splice site mutation and p.Arg679Thrfs*5; a splice site mutation and p.Arg749*; a splice site mutation and p.Asp873Thrfs*17; p.Arg536* and p.Gln929Pro; p.Arg536* and p.Asp960Asn; p.Cys871Arg and p.Asp960Asn; or p.Leu925Profs*25 and p.Gly937Asp). In some embodiments, one or more of the KAN1 variants in Table 3 is used to identify KIN.

Particular combinations of markers may be used that show optimal function with different ethnic groups or sex, different geographic distributions, different stages of disease, different degrees of specificity or different degrees of sensitivity. Particular combinations may also be developed which are particularly sensitive to the effect of therapeutic regimens on disease progression. Subjects may be monitored after a therapy and/or course of action to determine the effectiveness of that specific therapy and/or course of action.

B. Detection of Variant Alleles

In some embodiments, the present invention provides methods of detecting the presence of wild type or variant (*e.g.*, mutant or polymorphic) nucleic acids or polypeptides. The detection of variant alleles finds use in the diagnosis of disease (*e.g.*, one or more of NPHP-RC, KIN, and CAKUT), research, and selection of appropriate treatment and/or monitoring regimens.

Accordingly, the present invention provides compositions, kits, and methods for determining whether a patient has a mutation profile associated with one or more of NPHP-RC, KIN, and CAKUT.

A number of methods maybe be used for analysis of variant (*e.g.*, mutant or polymorphic) nucleic acid sequences. Assays for detecting variants (*e.g.*, polymorphisms or mutations) fall into several categories, including, but not limited to direct sequencing assays, fragment polymorphism assays, hybridization assays, and computer based data analysis. In some embodiments, assays are performed in combination or in hybrid (*e.g.*, different reagents or technologies from several assays are combined to yield one assay). The following assays are useful in the present invention.

Any patient sample containing nucleic acids or polypeptides may be tested according to the methods of the present invention. By way of non-limiting examples, the sample may be tissue, blood, urine, semen, or a fraction thereof (e.g., plasma, serum, whole blood, cells, etc.).

The patient sample may undergo preliminary processing designed to isolate or enrich the sample for the variant nucleic acids or polypeptides or cells that contain the variant nucleic acids. Centrifugation; immunocapture; cell lysis; sequence capture; and, nucleic acid target capture maybe used for such purposes.

i. DNA and RNA Detection

The variants of the present invention may be detected as genomic DNA or mRNA with techniques including but not limited to: nucleic acid sequencing; nucleic acid hybridization; and, nucleic acid amplification.

1. Sequencing

Illustrative non-limiting examples of nucleic acid sequencing techniques include, but are not limited to, chain terminator (Sanger) sequencing and dye terminator sequencing. Those of ordinary skill in the art will recognize that because RNA is less stable in the cell and more prone to nuclease attack experimentally RNA is usually reverse transcribed to DNA before sequencing.

Chain terminator sequencing uses sequence-specific termination of a DNA synthesis reaction using modified nucleotide substrates. Extension is initiated at a specific site on the template DNA by using a short radioactive, fluorescent or other labeled, oligonucleotide primer complementary to the template at that region. The oligonucleotide primer is extended using a DNA polymerase, standard four deoxynucleotide bases, and a low concentration of one chain terminating nucleotide, most commonly a di-deoxynucleotide. This reaction is repeated in four separate tubes with each of the bases taking turns as the di-deoxynucleotide.

Limited incorporation of the chain terminating nucleotide by the DNA polymerase results in a series of related DNA fragments that are terminated only at positions where that particular di-deoxynucleotide is used. For each reaction tube, the fragments are size-separated by electrophoresis in a slab polyacrylamide gel or a capillary tube filled with a viscous polymer. The sequence is determined by reading which lane produces a visualized mark from the labeled primer as you scan from the top of the gel to the bottom.

Dye terminator sequencing alternatively labels the terminators. Complete sequencing can be performed in a single reaction by labeling each of the di-deoxynucleotide chain-terminators with a separate fluorescent dye, which fluoresces at a different wavelength.

Some embodiments of the present invention utilize next generation or high-throughput sequencing. A variety of nucleic acid sequencing methods are contemplated for use in the methods of the present disclosure including, for example, chain terminator (Sanger) sequencing, dye terminator sequencing, and high-throughput sequencing methods. See, e.g., Sanger et al., Proc. Natl. Acad. Sci. USA 74:5463-5467 (1997); Maxam et al., Proc. Natl. Acad. Sci. USA 74:560-564 (1977); Drmanac, et al., Nat. Biotechnol. 16:54-58 (1998); Kato, Int. J. Clin. Exp. Med. 2:193-202 (2009); Ronaghi et al., Anal. Biochem. 242:84-89 (1996); Margulies et al., Nature 437:376-380 (2005); Ruparel et al., Proc. Natl. Acad. Sci. USA 102:5932-5937 (2005), and Harris et al., Science 320:106-109 (2008); Levene et al., Science 299:682-686 (2003); Korlach et al., Proc. Natl. Acad. Sci. USA 105:1176-1181 (2008); Branton et al., Nat. Biotechnol. 26(10):1146-53 (2008); Eid et al., Science 323:133-138 (2009); each of which is herein incorporated by reference in its entirety.

In some embodiments, sequencing technology including, but not limited to, pyrosequencing, sequencing-by-ligation, single molecule sequencing, sequence-by-synthesis (SBS), massive parallel clonal, massive parallel single molecule SBS, massive parallel single molecule real-time, massive parallel single molecule real-time nanopore technology, etc. Morozova and Marra provide a review of some such technologies in *Genomics*, **92**: 255 (2008), herein incorporated by reference in its entirety. Those of ordinary skill in the art will recognize that because RNA is less stable in the cell and more prone to nuclease attack experimentally RNA is usually reverse transcribed to DNA before sequencing.

A number of DNA sequencing techniques may be utilized, including fluorescence-based sequencing methodologies (See, e.g., Birren et al., Genome Analysis: Analyzing DNA, 1, Cold Spring Harbor, N.Y.; herein incorporated by reference in its entirety). In some embodiments, the sequence techniques are parallel sequencing of partitioned amplicons (PCT Publication No: WO2006084132 to Kevin McKernan et al., herein incorporated by reference

in its entirety). In some embodiments, the technology finds use in DNA sequencing by parallel oligonucleotide extension (See, e.g., U.S. Pat. No. 5,750,341 to Macevicz et al., and U.S. Pat. No. 6,306,597 to Macevicz et al., both of which are herein incorporated by reference in their entireties). Additional examples of sequencing techniques in which the technology finds use include the Church polony technology (Mitra et al., 2003, Analytical Biochemistry 320, 55-65; Shendure et al., 2005 Science 309, 1728-1732; U.S. Pat. No. 6,432,360, U.S. Pat. No. 6,485,944, U.S. Pat. No. 6,511,803; herein incorporated by reference in their entireties), the 454 picotiter pyrosequencing technology (Margulies et al., 2005 Nature 437, 376-380; US 20050130173; herein incorporated by reference in their entireties), the Solexa single base addition technology (Bennett et al., 2005, Pharmacogenomics, 6, 373-382; U.S. Pat. No. 6,787,308; U.S. Pat. No. 6,833,246; herein incorporated by reference in their entireties), the Lynx massively parallel signature sequencing technology (Brenner et al. (2000). Nat. Biotechnol. 18:630-634; U.S. Pat. No. 5,695,934; U.S. Pat. No. 5,714,330; herein incorporated by reference in their entireties), and the Adessi PCR colony technology (Adessi et al. (2000). Nucleic Acid Res. 28, E87; WO 00018957; herein incorporated by reference in its entirety).

Next-generation sequencing (NGS) methods share the common feature of massively parallel, high-throughput strategies, with the goal of lower costs in comparison to older sequencing methods (see, e.g., Voelkerding *et al.*, *Clinical Chem.*, *55:* 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbiol.*, *7:* 287-296; each herein incorporated by reference in their entirety). NGS methods can be broadly divided into those that typically use template amplification and those that do not. Amplification-requiring methods include pyrosequencing commercialized by Roche as the 454 technology platforms (e.g., GS 20 and GS FLX), the Solexa platform commercialized by Illumina, and the Supported Oligonucleotide Ligation and Detection (SOLiD) platform commercialized by Applied Biosystems. Non-amplification approaches, also known as single-molecule sequencing, are exemplified by the HeliScope platform commercialized by Helicos BioSciences, and emerging platforms commercialized by VisiGen, Oxford Nanopore Technologies Ltd., Life Technologies/Ion Torrent, and Pacific Biosciences, respectively.

In pyrosequencing (Voelkerding *et al.*, *Clinical Chem.*, *55:* 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbiol.*, *7:* 287-296; U.S. Pat. No. 6,210,891; U.S. Pat. No. 6,258,568; each herein incorporated by reference in its entirety), template DNA is fragmented, end-repaired, ligated to adaptors, and clonally amplified in-situ by capturing single template molecules with beads bearing oligonucleotides complementary to the adaptors. Each bead

bearing a single template type is compartmentalized into a water-in-oil microvesicle, and the template is clonally amplified using a technique referred to as emulsion PCR. The emulsion is disrupted after amplification and beads are deposited into individual wells of a picotitre plate functioning as a flow cell during the sequencing reactions. Ordered, iterative introduction of each of the four dNTP reagents occurs in the flow cell in the presence of sequencing enzymes and luminescent reporter such as luciferase. In the event that an appropriate dNTP is added to the 3′ end of the sequencing primer, the resulting production of ATP causes a burst of luminescence within the well, which is recorded using a CCD camera. It is possible to achieve read lengths greater than or equal to 400 bases, and 10^6 sequence reads can be achieved, resulting in up to 500 million base pairs (Mb) of sequence.

In the Solexa/Illumina platform (Voelkerding *et al.*, *Clinical Chem.*, *55:* 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbiol.*, *7:* 287-296; U.S. Pat. No. 6,833,246; U.S. Pat. No. 7,115,400; U.S. Pat. No. 6,969,488; each herein incorporated by reference in its entirety), sequencing data are produced in the form of shorter-length reads. In this method, single-stranded fragmented DNA is end-repaired to generate 5′-phosphorylated blunt ends, followed by Klenow-mediated addition of a single A base to the 3′ end of the fragments. A-addition facilitates addition of T-overhang adaptor oligonucleotides, which are subsequently used to capture the template-adaptor molecules on the surface of a flow cell that is studded with oligonucleotide anchors. The anchor is used as a PCR primer, but because of the length of the template and its proximity to other nearby anchor oligonucleotides, extension by PCR results in the "arching over" of the molecule to hybridize with an adjacent anchor oligonucleotide to form a bridge structure on the surface of the flow cell. These loops of DNA are denatured and cleaved. Forward strands are then sequenced with reversible dye terminators. The sequence of incorporated nucleotides is determined by detection of post-incorporation fluorescence, with each fluor and block removed prior to the next cycle of dNTP addition. Sequence read length ranges from 36 nucleotides to over 50 nucleotides, with overall output exceeding 1 billion nucleotide pairs per analytical run.

Sequencing nucleic acid molecules using SOLiD technology (Voelkerding *et al.*, *Clinical Chem.*, *55:* 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbiol.*, *7:* 287-296; U.S. Pat. No. 5,912,148; U.S. Pat. No. 6,130,073; each herein incorporated by reference in their entirety) also involves fragmentation of the template, ligation to oligonucleotide adaptors, attachment to beads, and clonal amplification by emulsion PCR. Following this, beads bearing template are immobilized on a derivatized surface of a glass flow-cell, and a primer complementary to the adaptor oligonucleotide is annealed. However, rather than utilizing this

primer for 3′ extension, it is instead used to provide a 5′ phosphate group for ligation to interrogation probes containing two probe-specific bases followed by 6 degenerate bases and one of four fluorescent labels. In the SOLiD system, interrogation probes have 16 possible combinations of the two bases at the 3′ end of each probe, and one of four fluors at the 5′ end. Fluor color, and thus identity of each probe, corresponds to specified color-space coding schemes. Multiple rounds (usually 7) of probe annealing, ligation, and fluor detection are followed by denaturation, and then a second round of sequencing using a primer that is offset by one base relative to the initial primer. In this manner, the template sequence can be computationally re-constructed, and template bases are interrogated twice, resulting in increased accuracy. Sequence read length averages 35 nucleotides, and overall output exceeds 4 billion bases per sequencing run.

In certain embodiments, the technology finds use in nanopore sequencing (see, e.g., Astier et al., J. Am. Chem. Soc. 2006 Feb 8; 128(5):1705–10, herein incorporated by reference). The theory behind nanopore sequencing has to do with what occurs when a nanopore is immersed in a conducting fluid and a potential (voltage) is applied across it. Under these conditions a slight electric current due to conduction of ions through the nanopore can be observed, and the amount of current is exceedingly sensitive to the size of the nanopore. As each base of a nucleic acid passes through the nanopore, this causes a change in the magnitude of the current through the nanopore that is distinct for each of the four bases, thereby allowing the sequence of the DNA molecule to be determined.

In certain embodiments, the technology finds use in HeliScope by Helicos BioSciences (Voelkerding *et al.*, *Clinical Chem.*, *55:* 641-658, 2009; MacLean *et al.*, *Nature Rev. Microbiol.*, *7:* 287-296; U.S. Pat. No. 7,169,560; U.S. Pat. No. 7,282,337; U.S. Pat. No. 7,482,120; U.S. Pat. No. 7,501,245; U.S. Pat. No. 6,818,395; U.S. Pat. No. 6,911,345; U.S. Pat. No. 7,501,245; each herein incorporated by reference in their entirety). Template DNA is fragmented and polyadenylated at the 3′ end, with the final adenosine bearing a fluorescent label. Denatured polyadenylated template fragments are ligated to poly(dT) oligonucleotides on the surface of a flow cell. Initial physical locations of captured template molecules are recorded by a CCD camera, and then label is cleaved and washed away. Sequencing is achieved by addition of polymerase and serial addition of fluorescently-labeled dNTP reagents. Incorporation events result in fluor signal corresponding to the dNTP, and signal is captured by a CCD camera before each round of dNTP addition. Sequence read length ranges from 25–50 nucleotides, with overall output exceeding 1 billion nucleotide pairs per analytical run.

The Ion Torrent technology is a method of DNA sequencing based on the detection of hydrogen ions that are released during the polymerization of DNA (see, e.g., *Science* **327**(5970): 1190 (2010); U.S. Pat. Appl. Pub. Nos. 20090026082, 20090127589, 20100301398, 20100197507, 20100188073, and 20100137143, incorporated by reference in their entireties for all purposes). A microwell contains a template DNA strand to be sequenced. Beneath the layer of microwells is a hypersensitive ISFET ion sensor. All layers are contained within a CMOS semiconductor chip, similar to that used in the electronics industry. When a dNTP is incorporated into the growing complementary strand a hydrogen ion is released, which triggers a hypersensitive ion sensor. If homopolymer repeats are present in the template sequence, multiple dNTP molecules will be incorporated in a single cycle. This leads to a corresponding number of released hydrogens and a proportionally higher electronic signal. This technology differs from other sequencing technologies in that no modified nucleotides or optics are used. The per-base accuracy of the Ion Torrent sequencer is ~99.6% for 50 base reads, with ~100 Mb generated per run. The read-length is 100 base pairs. The accuracy for homopolymer repeats of 5 repeats in length is ~98%. The benefits of ion semiconductor sequencing are rapid sequencing speed and low upfront and operating costs.

The technology finds use in another nucleic acid sequencing approach developed by Stratos Genomics, Inc. and involves the use of Xpandomers. This sequencing process typically includes providing a daughter strand produced by a template-directed synthesis. The daughter strand generally includes a plurality of subunits coupled in a sequence corresponding to a contiguous nucleotide sequence of all or a portion of a target nucleic acid in which the individual subunits comprise a tether, at least one probe or nucleobase residue, and at least one selectively cleavable bond. The selectively cleavable bond(s) is/are cleaved to yield an Xpandomer of a length longer than the plurality of the subunits of the daughter strand. The Xpandomer typically includes the tethers and reporter elements for parsing genetic information in a sequence corresponding to the contiguous nucleotide sequence of all or a portion of the target nucleic acid. Reporter elements of the Xpandomer are then detected. Additional details relating to Xpandomer-based approaches are described in, for example, U.S. Pat. Pub No. 20090035777, entitled "High Throughput Nucleic Acid Sequencing by Expansion," filed June 19, 2008, which is incorporated herein in its entirety.

Other single molecule sequencing methods include real-time sequencing by synthesis using a VisiGen platform (Voelkerding *et al.*, *Clinical Chem.*, *55:* 641–58, 2009; U.S. Pat. No. 7,329,492; U.S. Pat. App. Ser. No. 11/671956; U.S. Pat. App. Ser. No. 11/781166; each

herein incorporated by reference in their entirety) in which immobilized, primed DNA template is subjected to strand extension using a fluorescently-modified polymerase and florescent acceptor molecules, resulting in detectible fluorescence resonance energy transfer (FRET) upon nucleotide addition.

In some embodiments, capillary electrophoresis (CE) is utilized to analyze amplification fragments. During capillary electrophoresis, nucleic acids (e.g., the products of a PCR reaction) are injected electrokinetically into capillaries filled with polymer. High voltage is applied so that the fluorescent DNA fragments are separated by size and are detected by a laser/camera system. In some embodiments, CE systems from Life Technogies (Grand Island, NY) are utilized for fragment sizing (See e.g., US 6706162, US8043493, each of which is herein incorporated by reference in its entirety).

2. Hybridization

Illustrative non-limiting examples of nucleic acid hybridization techniques include, but are not limited to, in situ hybridization (ISH), microarray, and Southern or Northern blot. In situ hybridization (ISH) is a type of hybridization that uses a labeled complementary DNA or RNA strand as a probe to localize a specific DNA or RNA sequence in a portion or section of tissue (in situ), or, if the tissue is small enough, the entire tissue (whole mount ISH). DNA ISH can be used to determine the structure of chromosomes. RNA ISH is used to measure and localize mRNAs and other transcripts within tissue sections or whole mounts. Sample cells and tissues are usually treated to fix the target transcripts in place and to increase access of the probe. The probe hybridizes to the target sequence at elevated temperature, and then the excess probe is washed away. The probe that was labeled with either radio-, fluorescent- or antigen-labeled bases is localized and quantitated in the tissue using either autoradiography, fluorescence microscopy or immunohistochemistry, respectively. ISH can also use two or more probes, labeled with radioactivity or the other non-radioactive labels, to simultaneously detect two or more transcripts.

3. Microarrays

In some embodiments, microarrays are utilized for detection of variant nucleic acid sequences. Examples of microarrays include, but not limited to: DNA microarrays (e.g., cDNA microarrays and oligonucleotide microarrays); protein microarrays; tissue microarrays; transfection or cell microarrays; chemical compound microarrays; and, antibody microarrays. A DNA microarray, commonly known as gene chip, DNA chip, or biochip, is a

collection of microscopic DNA spots attached to a solid surface (e.g., glass, plastic or silicon chip) forming an array for the purpose of expression profiling or monitoring expression levels for thousands of genes simultaneously. The affixed DNA segments are known as probes, thousands of which can be used in a single DNA microarray. Microarrays can be used to identify disease genes by comparing gene expression in disease and normal cells. Microarrays can be fabricated using a variety of technologies, including but not limiting: printing with fine-pointed pins onto glass slides; photolithography using pre-made masks; photolithography using dynamic micromirror devices; ink-jet printing; or, electrochemistry on microelectrode arrays.

Arrays can also be used to detect copy number variations at al specific locus. These genomic micorarrys detect microscopic deletions or other variants that lead to disease causing alleles.

Southern and Northern blotting is used to detect specific DNA or RNA sequences, respectively. DNA or RNA extracted from a sample is fragmented, electrophoretically separated on a matrix gel, and transferred to a membrane filter. The filter bound DNA or RNA is subject to hybridization with a labeled probe complementary to the sequence of interest. Hybridized probe bound to the filter is detected. A variant of the procedure is the reverse Northern blot, in which the substrate nucleic acid that is affixed to the membrane is a collection of isolated DNA fragments and the probe is RNA extracted from a tissue and labeled.

4. Amplification

Variant nucleic acid may be amplified prior to or simultaneous with detection. Illustrative non-limiting examples of nucleic acid amplification techniques include, but are not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA). Those of ordinary skill in the art will recognize that certain amplification techniques (e.g., PCR) require that RNA be reversed transcribed to DNA prior to amplification (e.g., RT-PCR), whereas other amplification techniques directly amplify RNA (e.g., TMA and NASBA).

The polymerase chain reaction (U.S. Pat. Nos. 4,683,195, 4,683,202, 4,800,159 and 4,965,188, each of which is herein incorporated by reference in its entirety), commonly referred to as PCR, uses multiple cycles of denaturation, annealing of primer pairs to opposite

strands, and primer extension to exponentially increase copy numbers of a target nucleic acid sequence. In a variation called RT-PCR, reverse transcriptase (RT) is used to make a complementary DNA (cDNA) from mRNA, and the cDNA is then amplified by PCR to produce multiple copies of DNA. For other various permutations of PCR see, e.g., U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159; Mullis et al., Meth. Enzymol. 155: 335 (1987); and, Murakawa et al., DNA 7: 287 (1988), each of which is herein incorporated by reference in its entirety.

Transcription mediated amplification (U.S. Pat. Nos. 5,480,784 and 5,399,491, each of which is herein incorporated by reference in its entirety), commonly referred to as TMA, synthesizes multiple copies of a target nucleic acid sequence autocatalytically under conditions of substantially constant temperature, ionic strength, and pH in which multiple RNA copies of the target sequence autocatalytically generate additional copies. See, e.g., U.S. Pat. Nos. 5,399,491 and 5,824,518, each of which is herein incorporated by reference in its entirety. In a variation described in U.S. Publ. No. 20060046265 (herein incorporated by reference in its entirety), TMA optionally incorporates the use of blocking moieties, terminating moieties, and other modifying moieties to improve TMA process sensitivity and accuracy.

The ligase chain reaction (Weiss, R., Science 254: 1292 (1991), herein incorporated by reference in its entirety), commonly referred to as LCR, uses two sets of complementary DNA oligonucleotides that hybridize to adjacent regions of the target nucleic acid. The DNA oligonucleotides are covalently linked by a DNA ligase in repeated cycles of thermal denaturation, hybridization and ligation to produce a detectable double-stranded ligated oligonucleotide product.

Strand displacement amplification (Walker, G. et al., Proc. Natl. Acad. Sci. USA 89: 392-396 (1992); U.S. Pat. Nos. 5,270,184 and 5,455,166, each of which is herein incorporated by reference in its entirety), commonly referred to as SDA, uses cycles of annealing pairs of primer sequences to opposite strands of a target sequence, primer extension in the presence of a dNTPαS to produce a duplex hemiphosphorothioated primer extension product, endonuclease-mediated nicking of a hemimodified restriction endonuclease recognition site, and polymerase-mediated primer extension from the 3' end of the nick to displace an existing strand and produce a strand for the next round of primer annealing, nicking and strand displacement, resulting in geometric amplification of product. Thermophilic SDA (tSDA) uses thermophilic endonucleases and polymerases at higher temperatures in essentially the same method (EP Pat. No. 0 684 315).

Other amplification methods include, for example: nucleic acid sequence based amplification (U.S. Pat. No. 5,130,238, herein incorporated by reference in its entirety), commonly referred to as NASBA; one that uses an RNA replicase to amplify the probe molecule itself (Lizardi et al., BioTechnol. 6: 1197 (1988), herein incorporated by reference in its entirety), commonly referred to as Qβ replicase; a transcription based amplification method (Kwoh et al., Proc. Natl. Acad. Sci. USA 86:1173 (1989)); and, self-sustained sequence replication (Guatelli et al., Proc. Natl. Acad. Sci. USA 87: 1874 (1990), each of which is herein incorporated by reference in its entirety). For further discussion see Persing, David H., "In Vitro Nucleic Acid Amplification Techniques" in Diagnostic Medical Microbiology: Principles and Applications (Persing et al., Eds.), pp. 51-87 (American Society for Microbiology, Washington, DC (1993)).

5. Detection Methods

Non-amplified or amplified nucleic acids can be detected by a variety of techniques. For example, nucleic acid can be detected by hybridization with a detectably labeled probe and measurement of the resulting hybrids. Illustrative non-limiting examples of detection methods are described below.

One illustrative detection method, the Hybridization Protection Assay (HPA) involves hybridizing a chemiluminescent oligonucleotide probe (e.g., an acridinium ester-labeled (AE) probe) to the target sequence, selectively hydrolyzing the chemiluminescent label present on unhybridized probe, and measuring the chemiluminescence produced from the remaining probe in a luminometer. See, e.g., U.S. Pat. No. 5,283,174 and Norman C. Nelson et al., Nonisotopic Probing, Blotting, and Sequencing, ch. 17 (Larry J. Kricka ed., 2d ed. 1995, each of which is herein incorporated by reference in its entirety).

Another illustrative detection method provides for quantitative evaluation of the amplification process in real-time. Evaluation of an amplification process in "real-time" involves determining the amount of amplicon in the reaction mixture either continuously or periodically during the amplification reaction, and using the determined values to calculate the amount of target sequence initially present in the sample. The amount of initial target sequence present in a sample based on real-time amplification may be determined using methods including those disclosed in U.S. Pat. Nos. 6,303,305 and 6,541,205, each of which is herein incorporated by reference in its entirety. Another method for determining the quantity of target sequence initially present in a sample, but which is not based on a real-time

amplification, is disclosed in U.S. Pat. No. 5,710,029, herein incorporated by reference in its entirety.

Amplification products may be detected in real-time through the use of various self-hybridizing probes, most of which have a stem-loop structure. Such self-hybridizing probes are labeled so that they emit differently detectable signals, depending on whether the probes are in a self-hybridized state or an altered state through hybridization to a target sequence. By way of non-limiting example, "molecular torches" are a type of self-hybridizing probe that includes distinct regions of self-complementarity (referred to as "the target binding domain" and "the target closing domain") which are connected by a joining region (e.g., non-nucleotide linker) and which hybridize to each other under predetermined hybridization assay conditions. In a preferred embodiment, molecular torches contain single-stranded base regions in the target binding domain that are from 1 to about 20 bases in length and are accessible for hybridization to a target sequence present in an amplification reaction under strand displacement conditions. Under strand displacement conditions, hybridization of the two complementary regions, which may be fully or partially complementary, of the molecular torch is favored, except in the presence of the target sequence, which will bind to the single-stranded region present in the target binding domain and displace all or a portion of the target closing domain. The target binding domain and the target closing domain of a molecular torch include a detectable label or a pair of interacting labels (e.g., luminescent/quencher) positioned so that a different signal is produced when the molecular torch is self-hybridized than when the molecular torch is hybridized to the target sequence, thereby permitting detection of probe:target duplexes in a test sample in the presence of unhybridized molecular torches. Molecular torches and a variety of types of interacting label pairs are disclosed in U.S. Pat. No. 6,534,274, herein incorporated by reference in its entirety.

Another example of a detection probe having self-complementarity is a "molecular beacon." Molecular beacons include nucleic acid molecules having a target complementary sequence, an affinity pair (or nucleic acid arms) holding the probe in a closed conformation in the absence of a target sequence present in an amplification reaction, and a label pair that interacts when the probe is in a closed conformation. Hybridization of the target sequence and the target complementary sequence separates the members of the affinity pair, thereby shifting the probe to an open conformation. The shift to the open conformation is detectable due to reduced interaction of the label pair, which may be, for example, a fluorophore and a quencher (e.g., DABCYL and EDANS). Molecular beacons are disclosed in U.S. Pat. Nos. 5,925,517 and 6,150,097, herein incorporated by reference in its entirety.

Other self-hybridizing probes include probe binding pairs having interacting labels, such as those disclosed in U.S. Pat. No. 5,928,862 (herein incorporated by reference in its entirety) might be adapted for use in the present invention. Probe systems used to detect single nucleotide polymorphisms (SNPs) might also be utilized in the present invention. Additional detection systems include "molecular switches," as disclosed in U.S. Publ. No. 20050042638, herein incorporated by reference in its entirety. Other probes, such as those comprising intercalating dyes and/or fluorochromes, are also useful for detection of amplification products in the present invention. See, e.g., U.S. Pat. No. 5,814,447 (herein incorporated by reference in its entirety).

ii. Detection of Variant Proteins

In other embodiments, variant polypeptides are detected. Any suitable method may be used to detect truncated or mutant polypeptides including, but not limited to, those described below.

1. Antibody Binding

In some embodiments, antibodies (See below for antibody production) are used to determine if an individual contains an allele encoding a variant polypeptides. In preferred embodiments, antibodies are utilized that discriminate between variant (*i.e.*, truncated proteins); and wild-type proteins. In some embodiments, the antibodies are directed to the C-terminus of proteins. Proteins that are recognized by the N-terminal, but not the C-terminal antibody are truncated. In some embodiments, quantitative immunoassays are used to determine the ratios of C-terminal to N-terminal antibody binding. In other embodiments, identification of variants is accomplished through the use of antibodies that differentially bind to wild type or variant forms of the polypeptides described herein.

Antibody binding is detected by techniques such as radioimmunoassay, ELISA (enzyme-linked immunosorbant assay), "sandwich" immunoassays, immunoradiometric assays, gel diffusion precipitation reactions, immunodiffusion assays, *in situ* immunoassays (*e.g.*, using colloidal gold, enzyme or radioisotope labels, for example), Western blots, precipitation reactions, agglutination assays (*e.g.,* gel agglutination assays, hemagglutination assays, etc.), complement fixation assays, immunofluorescence assays, protein A assays, and immunoelectrophoresis assays, etc.

In one embodiment, antibody binding is detected by detecting a label on the primary antibody. In another embodiment, the primary antibody is detected by detecting binding of a

secondary antibody or reagent to the primary antibody. In a further embodiment, the secondary antibody is labeled.

In some embodiments, an automated detection assay is utilized. Methods for the automation of immunoassays include those described in U.S. Patents 5,885,530, 4,981,785, 6,159,750, and 5,358,691, each of which is herein incorporated by reference. In some embodiments, the analysis and presentation of results is also automated. For example, in some embodiments, software that generates a prognosis based on the result of the immunoassay is utilized. In other embodiments, the immunoassay described in U.S. Patents 5,599,677 and 5,672,480; each of which is herein incorporated by reference.

C. Kits for Detecting Mutant or Variant Alleles

The present invention also provides kits for determining whether an individual contains a wild-type or variant (*e.g.*, mutant or polymorphic) allele of one or more of the genes described herein as being associated with one or more of NPHP-RC, KIN, and CAKUT. In some embodiments, the kits are useful for determining whether the subject has one or more of NPHP-RC, KIN, and CAKUT or to provide a prognosis to an individual diagnosed with one or more of NPHP-RC, KIN, and CAKUT. The diagnostic kits are produced in a variety of ways. In some embodiments, the kits contain at least two reagents (e.g., one or more of NPHP-RC, KIN, and CAKUT informative reagent) useful, necessary, or sufficient for specifically detecting two or more distinct mutant or variant allele or protein. In some embodiments, the kits contain reagents for detecting a truncation in the polypeptide encoded by the variant nucleic acid. In preferred embodiments, the reagent is a nucleic acid that hybridizes to nucleic acids containing the mutation and that does not bind to nucleic acids that do not contain the mutation. In other embodiments, the reagent is a pair of primers for amplifying the region of DNA containing the mutation. In still other embodiments, the reagents are antibodies that preferentially bind either the wild-type or truncated or variant proteins. In some embodiments, reagents are one or more sequencing primers.

In some embodiments, the kits include ancillary reagents such as buffering agents, nucleic acid stabilizing reagents, protein stabilizing reagents, and signal producing systems (*e.g.*, florescence generating systems as Fret systems), and software (e.g., data analysis software). The test kit may be packages in any suitable manner, typically with the elements in a single container or various containers as necessary along with a sheet of instructions for carrying out the test. In some embodiments, the kits also preferably include a positive control sample.

In some embodiments, markers (e.g., those described herein) are detected alone or in combination with other markers in a panel or multiplex format. For example, in some embodiments, a plurality of markers are simultaneously detected in an array or multiplex format (e.g., using the detection methods described herein).

5

D. Bioinformatics

For example, in some embodiments, a computer-based analysis program is used to translate the raw data generated by the detection assay (*e.g.*, the presence, absence, or amount of a given allele or polypeptide) into data of predictive value for a clinician. The clinician 10 can access the predictive data using any suitable means. Thus, in some preferred embodiments, the present invention provides the further benefit that the clinician, who may not be trained in genetics or molecular biology, need not understand the raw data. The data is presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information in order to optimize the care of the subject.

15 The present invention contemplates any method capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information providers, medical personal, and subjects. For example, in some embodiments of the present invention, a sample (*e.g.*, a biopsy or a blood or serum sample) is obtained from a subject and submitted to a profiling service (*e.g.*, clinical lab at a medical facility, genomic profiling 20 business, etc.), located in any part of the world (*e.g.*, in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center, or subjects may collect the sample themselves (*e.g.*, a urine sample) and directly send it to a profiling center. 25 Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (*e.g.*, an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center using an electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced (*i.e.*, presence of 30 wild type or mutant allele), specific for the screening, diagnostic or prognostic information desired for the subject.

The profile data is then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw data, the prepared format may represent a diagnosis or risk assessment (*e.g.*, diagnosis or prognosis of NPHP-RC and/or CAKUT) for

the subject, along with recommendations for particular treatment options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can be printed for the clinician (*e.g.*, at the point of care) or displayed to the clinician on a computer monitor.

In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data is then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data is stored in a central facility with uniform security protocols), speed, and uniformity of data analysis. The central processing facility can then control the fate of the data following treatment of the subject. For example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers.

In some embodiments, the subject is able to directly access the data using the electronic communication system. The subject may chose further intervention or counseling based on the results. In some embodiments, the data is used for research use. For example, the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition or stage of disease.

In some embodiments, the methods disclosed herein are useful in monitoring the treatment of one or more of NPHP-RC, KIN, and CAKUT. For example, in some embodiments, the methods may be performed immediately before, during and/or after a treatment to monitor treatment success. In some embodiments, the methods are performed at intervals on disease free patients to ensure treatment success.

The present invention also provides a variety of computer-related embodiments. Specifically, in some embodiments the invention provides computer programming for analyzing and comparing a pattern of one or more of NPHP-RC, KIN, and CAKUT -specific marker detection results in a sample obtained from a subject to, for example, a library of such marker patterns known to be indicative of the presence or absence of one or more of NPHP-RC, KIN, and CAKUT, or a particular stage or prognosis of one or more of NPHP-RC, KIN, and CAKUT.

In some embodiments, the present invention provides computer programming for analyzing and comparing a first and a second pattern of one or more of NPHP-RC, KIN, and CAKUT -specific marker detection results from a sample taken at least two different time points.

In yet another embodiment, the invention provides computer programming for analyzing and comparing a pattern of one or more of NPHP-RC, KIN, and CAKUT -specific marker detection results from a sample to a library of one or more of NPHP-RC, KIN, and CAKUT -specific marker patterns known to be indicative of the presence or absence of one or more of NPHP-RC, KIN, and CAKUT, wherein the comparing provides, for example, a differential diagnosis between an aggressive and a less aggressive one or more of NPHP-RC, KIN, and CAKUT.

The methods and systems described herein can be implemented in numerous ways. In one embodiment, the methods involve use of a communications infrastructure, for example the internet. Several embodiments of the invention are discussed below. It is also to be understood that the present invention may be implemented in various forms of hardware, software, firmware, processors, distributed servers (e.g., as used in cloud computing) or a combination thereof. The methods and systems described herein can be implemented as a combination of hardware and software. The software can be implemented as an application program tangibly embodied on a program storage device, or different portions of the software implemented in the user's computing environment (e.g., as an applet) and on the reviewer's computing environment, where the reviewer may be located at a remote site (e.g., at a service provider's facility).

For example, during or after data input by the user, portions of the data processing can be performed in the user-side computing environment. For example, the user-side computing environment can be programmed to provide for defined test codes to denote platform, carrier/diagnostic test, or both; processing of data using defined flags, and/or generation of flag configurations, where the responses are transmitted as processed or partially processed responses to the reviewer's computing environment in the form of test code and flag configurations for subsequent execution of one or more algorithms to provide a results and/or generate a report in the reviewer's computing environment.

The application program for executing the algorithms described herein may be uploaded to, and executed by, a machine comprising any suitable architecture. In general, the machine involves a computer platform having hardware such as one or more central processing units (CPU), a random access memory (RAM), and input/output (I/O) interface(s). The computer platform also includes an operating system and microinstruction code. The various processes and functions described herein may either be part of the microinstruction code or part of the application program (or a combination thereof) which is executed via the

operating system. In addition, various other peripheral devices may be connected to the computer platform such as an additional data storage device and a printing device.

As a computer system, the system generally includes a processor unit. The processor unit operates to receive information, which generally includes test data (e.g., specific gene products assayed), and test result data (e.g., the pattern of gastrointestinal neoplasm-specific marker detection results from a sample). This information received can be stored at least temporarily in a database, and data analyzed in comparison to a library of marker patterns known to be indicative of the presence or absence of one or more of NPHP-RC, KIN, and CAKUT.

Part or all of the input and output data can also be sent electronically; certain output data (e.g., reports) can be sent electronically or telephonically (e.g., by facsimile, e.g., using devices such as fax back). Exemplary output receiving devices can include a display element, a printer, a facsimile device and the like. Electronic forms of transmission and/or display can include email, interactive television, and the like. In some embodiments, all or a portion of the input data and/or all or a portion of the output data (e.g., usually at least the library of the pattern of gastrointestinal neoplasm-specific marker detection results known to be indicative of the presence or absence of one or more of NPHP-RC, KIN, and CAKUT) are maintained on a server for access, e.g., confidential access. The results may be accessed or sent to professionals as desired.

A system for use in the methods described herein generally includes at least one computer processor (e.g., where the method is carried out in its entirety at a single site) or at least two networked computer processors (e.g., where detected marker data for a sample obtained from a subject is to be input by a user (e.g., a technician or someone performing the assays)) and transmitted to a remote site to a second computer processor for analysis (e.g., where the pattern of one or more of NPHP-RC, KIN, and CAKUT -specific marker) detection results is compared to a library of patterns known to be indicative of the presence or absence of one or more of NPHP-RC, KIN, and CAKUT), where the first and second computer processors are connected by a network, e.g., via an intranet or internet). The system can also include a user component(s) for input; and a reviewer component(s) for review of data, and generation of reports, including detection of a one or more of NPHP-RC, KIN, and CAKUT. Additional components of the system can include a server component(s); and a database(s) for storing data (e.g., as in a database of report elements, e.g., a library of marker patterns known to be indicative of the presence or absence of one or more of NPHP-RC, KIN, and CAKUT and/or known to be indicative of a grade and/or a stage of one or more of NPHP-

RC, KIN, and CAKUT, or a relational database (RDB) which can include data input by the user and data output. The computer processors can be processors that are typically found in personal desktop computers (e.g., IBM, Dell, Macintosh), portable computers, mainframes, minicomputers, tablet computer, smart phone, or other computing devices.

The input components can be complete, stand-alone personal computers offering a full range of power and features to run applications. The user component usually operates under any desired operating system and includes a communication element (e.g., a modem or other hardware for connecting to a network using a cellular phone network, Wi-Fi, Bluetooth, Ethernet, etc.), one or more input devices (e.g., a keyboard, mouse, keypad, or other device used to transfer information or commands), a storage element (e.g., a hard drive or other computer-readable, computer-writable storage medium), and a display element (e.g., a monitor, television, LCD, LED, or other display device that conveys information to the user). The user enters input commands into the computer processor through an input device. Generally, the user interface is a graphical user interface (GUI) written for web browser applications.

The server component(s) can be a personal computer, a minicomputer, or a mainframe, or distributed across multiple servers (e.g., as in cloud computing applications) and offers data management, information sharing between clients, network administration and security. The application and any databases used can be on the same or different servers. Other computing arrangements for the user and server(s), including processing on a single machine such as a mainframe, a collection of machines, or other suitable configuration are contemplated. In general, the user and server machines work together to accomplish the processing of the present invention.

Where used, the database(s) is usually connected to the database server component and can be any device which will hold data. For example, the database can be any magnetic or optical storing device for a computer (e.g., CDROM, internal hard drive, tape drive). The database can be located remote to the server component (with access via a network, modem, etc.) or locally to the server component.

Where used in the system and methods, the database can be a relational database that is organized and accessed according to relationships between data items. The relational database is generally composed of a plurality of tables (entities). The rows of a table represent records (collections of information about separate items) and the columns represent fields (particular attributes of a record). In its simplest conception, the relational database is a collection of data entries that "relate" to each other through at least one common field.

Additional workstations equipped with computers and printers may be used at point of service to enter data and, in some embodiments, generate appropriate reports, if desired. The computer(s) can have a shortcut (e.g., on the desktop) to launch the application to facilitate initiation of data entry, transmission, analysis, report receipt, etc. as desired.

In certain embodiments, the present invention provides methods for obtaining a subject's risk profile for developing one or more of NPHP-RC, KIN, and CAKUT or having aggressive one or more of NPHP-RC, KIN, and CAKUT. In some embodiments, such methods involve obtaining a blood or blood product sample from a subject (e.g., a human at risk for developing one or more of NPHP-RC, KIN, and CAKUT; a human undergoing a routine physical examination, or a human diagnosed with one or more of NPHP-RC, KIN, and CAKUT), detecting the presence or absence of the variants described herein associated with one or more of NPHP-RC, KIN, and CAKUT in the sample, and generating a risk profile for developing one or more of NPHP-RC, KIN, and CAKUT. For example, in some embodiments, a generated profile will change depending upon specific markers and detected as present or absent or at defined threshold levels. The present invention is not limited to a particular manner of generating the risk profile. In some embodiments, a processor (e.g., computer) is used to generate such a risk profile. In some embodiments, the processor uses an algorithm (e.g., software) specific for interpreting the presence and absence of specific exfoliated epithelial markers as determined with the methods of the present invention. In some embodiments, the presence and absence of specific variants as determined with the methods of the present invention are imputed into such an algorithm, and the risk profile is reported based upon a comparison of such input with established norms (e.g., established norm for various risk levels for developing one or more of NPHP-RC, KIN, and CAKUT, established norm for subjects diagnosed with various variations of one or more of NPHP-RC, KIN, and CAKUT). In some embodiments, the risk profile indicates a subject's risk for developing one or more of NPHP-RC, KIN, and CAKUT. In some embodiments, the risk profile indicates a subject to be, for example, a very low, a low, a moderate, a high, and a very high chance of developing one or more of NPHP-RC, KIN, and CAKUT or having a poor prognosis (e.g., likelihood of long term survival) from one or more of NPHP-RC, KIN, and CAKUT. In some embodiments, a health care provider will use such a risk profile in determining a course of treatment or intervention.

EXPERIMENTAL

The following examples are provided in order to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

Example 1

EXPERIMENTAL PROCEDURES

Research subjects. Blood samples and pedigrees were obtained following informed consent from individuals with NPHP-RC and/or their parents. The diagnosis of NPHP-RC was based on published clinical criteria (Chaki et al., Kidney Int 2011). Rat studies were performed according to Dutch Animal Welfare laws and were locally reviewed by the University of Utrecht Animal Ethical Committee (DEC). Human subjects provided informed consent to the use of their tissue for research purposes.

Linkage analysis. For genome-wide homozygosity mapping the GeneChip® Human Mapping 250k StyI Array from Affymetrix was used. Non-parametric LOD scores were calculated using a modified version of the program GENEHUNTER 2.1 (Kruglyak et al., Am J Hum Genet 58, 1347-1363 1996; Strauch et al., Am J Hum Genet 66, 1945-1957 2000) through stepwise use of a sliding window with sets of 110 SNPs and the program ALLEGRO (Gudbjartsson et al., Nat Genet 25, 12-13 2000) in order to identify regions of homozygosity as described (Hildebrandt et al., PloS Genetics 5, 31000353 2009c; Sayer et al., Nat Genet 38, 674-681 2006) using a disease allele frequency of 0.0001 and Caucasian marker allele frequencies.

Bioinformatics. Genetic location is according to the February 2009 Human Genome Browser data. Statistical analysis. Student‟s two-tailed non-paired t-tests and normal distribution two-tailed z-tests were carried out using pooled standard error and s.d. values to determine the statistical significance of different cohorts.

Whole-Exome Sequencing. Whole exome capture. Exome enrichment was conducted largely following the manufacturer‟s protocol for the NimbleGen SeqCap EZ Exome v2 (Roche NimbleGen, version 2). Briefly, three micrograms of genomic DNA was fragmented by sonication using the Covaris S2 system to achieve a uniform distribution of fragments with a mean size of 300 bp. The sonicated DNA was purified using AMPure XP Solid Phase

Reversible Immobilization paramagnetic (SPRI) beads (Agencourt) followed by polishing of the DNA ends by removing the 3‟ overhangs and filling in the 5‟ overhangs resulting from sonication using T4 DNA polymerase and Klenow fragment (New England Biolabs). Following end polishing, a single "A"-base was added to the 3‟ end of the DNA fragments using Klenow fragment (3‟ to 5‟ exo minus). This prepares the DNA fragments for ligation to specialized adaptors that have a "T"-base overhang at their 3‟ends. The end-repaired DNA with a single "A"-base overhang was ligated to paired-end adaptors (Illumina) in a standard ligation reaction using T4 DNA ligase and 2 μM – 4 μM final adaptor concentration, depending on the DNA yield following purification after the addition of the "A"-base (a 10-fold molar excess of adaptors is used in each reaction). Following ligation, the samples were purified using SPRI beads, amplified by six cycles of PCR to maintain complexity and avoid bias due to amplification and quality controlled by library size assessment on the Agilent Bioanalyzer and quantitation using PicoGreen reagent (Invitrogen).

One microgram of amplified, purified DNA (DNA library) was prepared for hybridization by adding COT1 DNA and blocking oligonucleotides to the DNA library, desiccating the DNA completely and resuspending the material in NimbleGen hybridization buffer. The resuspended material was denatured at 95°C prior to addition of the exome capture library bait material. The DNA library and biotin-labeled capture library were hybridized by incubation at 47°C for 68 h. Following hybridization, streptavidin coated magnetic beads were used to purify the DNA:DNA hybrids formed between the capture library and sequencing library during hybridization. The purified sequencing library was amplified directly from the purification beads using 8 cycles of PCR using Pfx DNA polymerase (Invitrogen). The libraries were purified following amplification and the library size was assessed using the Agilent Bioanalyzer. A single peak between 350-400 bp indicates a properly constructed and amplified library ready for sequencing. Final quantitation of the library was performed using the Kapa Biosciences Real-time PCR assay and appropriate amounts loaded onto the Illumina flowcell for sequencing by paired-end 50 nt sequencing on the Illumina HiSeq2000 sequencer.

Sequencing. Sequencing was performed largely as described in (Bentley et al., Nature 456, 53-59 2008). Following dilution to 10 nM final concentration based on the real-time PCR and bioanalyzer results, the final library stock is then used in paired-end (PE) cluster generation at a final concentration of 6-8 pM to achieve a cluster density of 600,000/mm2 (on the Illumina HiSeq2000 instrument, v2.5 reagents). Following cluster generation, 100nt

paired-end sequencing was performed using the standard Illumina protocols. Sequencing each sample in a single lane at paired-end 100 nt conditions on the Illumina HiSeq (v2.5 reagents) generated an average of 19.1Gb of pass-filter sequence data with 94.96% aligning to the genome. This amount of sequence corresponded to an average coverage of 181x over the 44 Mb of capture region. 80.41% of sequencing reads fell with 200 bp of the target region. 98.39% of the capture region was sequenced to a depth of at least 1x, 95.6% to at least 8x, 92.17% to at least 20x, and 89.42% to at least 30x.

Data Analysis. Raw sequencing data for each individual were mapped to the human reference genome (build hg19) using the Burrows-Wheeler Aligner (BWA v0.5.81536)14. The BWA aligned sequencing reads were processed by Picard to label the PCR duplicates. The Genome Analysis Toolkit (GATK, version 5091) was then used to remove duplicates, perform local realignment and map quality score recalibration to produce a "cleaned" BAM file for each individual. SNP calls were made by the Unified Genotyper module in GATK using the "cleaned" BAM files in batch fashion (90 samples per batch). The resulting Variant Call Format (VCF, version 4.0) files were annotated using the GenomicAnnotator module in GATK to identify and label the called variants that are within the targeted coding regions and overlap with known and likely benign SNPs reported in dbSNP v132 with allele frequencies >5%. The annotated VCF files were then filtered using the GATK variant filter module with a hard filter setting and a custom script for initial filtering. Variant calls that failed to pass the following filters were eliminated from the call set: i) MQ0 >= 4 && ((MQ0 / (1.0 * DP)) > 0.1); ii) QUAL < 30.0 || QD < 5.0 || HRun > 5 || SB > 0.00; iii) Cluster size 10; iv) not contain dbSNP id; v) inside the targeted regions. Combined VCF files were then split into individual files and remaining variants were tested for recessive segregation in the respective families. For each of the three families with mutation of MRE11, ZNF423 or CEP164 there was only one gene in each family in any of the homozygous regions for which the homozygous variant passed Polyphen 1 and 2 scores, segregated within the family, and was absent from >270 controls, thus validating the pathogenic relevance of the mutation.

CEP164 mutation analysis. Exon-PCR and Sanger sequencing of all 31 coding exons for one affected individual in each of 856 different NPHP-RC families was performed. The number of families examined (by increasing number of organs involved) was: isolated nephronophthisis (5), isolated retinal degeneration (100), Senior-Loken syndrome (168), Joubert syndrome (240), Bardet-Biedl syndrome (195), Meckel syndrome (52), and other

severe unsolved NPHP-RC cases (96). 480 individuals from additional NPHP-RC families were also examined using massively-parallel sequencing of exon-PCR from pooled DNA samples (Otto et al., Nat Genet 34, 413-420 2010).

Yeast-two-hybrid screening with ZNF423. A subclone of human CEP290 was prepared using high fidelity Taq polymerase, spanning 1770 bp, (562 amino acids, 1917-2479) and including wild-type stop codon and cloned into pENTR-TOPO as previously described in (Sayer et al., 2006, supra) (clone named "JAS2"). The CEP290 subclone insert was switched to destination vector pDEST32 (binding domain containing yeast-2-hydrid vector, "bait") (Invitrogen). CEP290 (JAS2) was used as bait, fused to the GAL4 DNA binding domain in the pDEST32 vector, and a human fetal brain expression library was screened and cloned into pEXPAD22 GAL4 activation domain fusion vector (Invitrogen). Approximately 1 x 10^6 clones were screened after cotransforming plasmids into competent MaV203 yeast cells (lithium acetate method) and plating onto His, Leu and Trp deficient medium containing mM 3-aminotriazole. Colonies were replica plated on restrictive media and surviving colonies were used for cDNA extraction. Five ml cultures were grown at 30°C overnight. cDNA was extracted using RPM yeast plasmid isolation kitTM (Bio 101 systems). cDNA was transformed into E. coli, purified and directly sequenced using vector specific primers. Sequence analysis allowed prediction of amino acid sequences (ORFinder), which were then identified by BLAT analysis. Direct yeast-2-hybrid interaction experiments allowed colony growth to be compared to 2 negative controls (respective plasmids without insert) and 5 positive control yeast strains for different interaction strength.

Coimmunoprecipitation with ZNF423. Full length human ZNF423 was subcloned into pcDNA3.1-NT-GFP using long range PCR (Expand, Roche) and IMAGE clone 100072717 as template. Clones were sequenced to confirm orientation and fidelity. A partial length human CEP290 cDNA spanning 1770 bp as described above (JAS2) was subcloned into a DEST-V5 vector. HEK293 cells were transiently co-transfected with full length GFP-ZNF423 and partial length CEP290-V5 using Lipofectamine 2000. After 24 h cells were washed in phosphate-buffered saline (PBS), pelleted, and lysed in NP-40 buffer (150 mM sodium chloride, 0.5% NP-40, 50 mM Tris pH 7.4, phenylmethanesulphonylfluoride, and protease inhibitors). The lysate was centrifuged for 20 min at 10,000 g at 4°C and the supernatant was precleared with protein G sepharose beads (GE Healthcare, Giles, Buckinghamshire, UK). After removal of protein G the supernatant was then incubated

overnight with protein G and 1 μg of either anti-V5 (Invitrogen) or mouse IgG (Sigma) at 4°C. The beads were washed extensively in lysis buffer and bound proteins resolved by 10% SDS polyacrylamide gel electrophoresis as described above. GFP-tagged proteins were detected with anti-GFP-conjugated to HRP (Santa Cruz) 1:5,000 and V5-tagged proteins were detected with anti-V5 conjugated to HRP (Invitrogen) 1:5,000.

Spheroid assays.

mIMCD3 cell culture. mIMCD3 is a mouse inner medullary collecting duct cell line. The cells were cultured in Dulbecco"s Modified Eagle"s Medium (DMEM):F12 (1:1) (GlutaMAX, Gibco), supplemented with 10% Fetal Calf Serum (FCS) and penicillin and streptomycin (1% P/S). Cells were incubated at 37°C in 5% carbon dioxide (CO2) to approximately 90% confluence.

Transfections. Before exposing the cells to different experimental conditions, cells were seeded in 6-, 24- or 48-well plates, depending on the experimental setup. After at least 6 h, the cells were transfected with Lipofectamine RNAimax (Invitrogen, 13778-075), according to the supplier's protocol. Opti-MEM (Invitrogen, 31985-062) was used to dilute the ON-TARGETplus siRNA SMARTpools (Thermo Scientific Dharmacon) for Non-targeting pool (D-001810-10) or Cep164 (L-057068-01).

Antibodies. The following primary antibodies were used: mouse anti-acetylated-tubulin (Sigma, T7451), 1:20000; human anti-Nuclear ANA-Centromere Autoantibody (CREST) (Cortex Biochem, CS1058) 1:1000; mouse γH2AX anti-phospho-histone H2A.X (ser139) (Millipore, DAM 1493341. Alexa-488, Alexa-555, Alexa-564 and Alexa-633 and 647 conjugated secondary antibodies were obtained from Invitrogen.

mIMCD3 spheroid growth assay. Cells were trypsinized 24 hour post-transfection and resuspended cells were then mixed 1:1 with growth factor-depleted matrigel (BD Bioscience). After the matrigel polymerized for 20 minutes at 37°C, warm medium was dropped over the matrix until just covered. The IMCD3 cells formed spheroids with cleared lumens 3 days later. Medium was removed by pipetting and the gels were washed three times for 10 minutes with warm PBS supplemented with calcium and magnesium. The gels were then fixed in fresh 4% PFA for 30 minutes RT. After washing three times in PBS after fixation, the cells were permeabilized for 15 minutes in gelatin dissolved in warm PBS (350

mg/50 mL) and 0.5% triton-X100. Primary antibody (mouse anti-acetylated tubulin, Sigma, at 1:20,000) was diluted in the permeabilization gelatin buffer and incubated at 4°C overnight. After washing the spheroids 3 times for 30 minutes in permeabilization buffer, goat anti-mouse Alexa-555 secondary antibody (Invitrogen) was diluted 1:500 in permeabilization buffer and incubated overnight at 4°C. The next day, spheroids were washed 3 times in permeabilization buffer and mounted with Dapi (1:2,000) in Fluoroumount-G (Cell Lab, Beckman Coulter). Images were taken with a Zeiss LSM510 confocal microscope and 50 spheroids per condition were scored. GraphPad Prism 5.0 was used to perform two-tailed student t-tests.

RT-qPCR for Cep164. IMCD3 cells were transfected with non-targeting siControl or siCep164 oligonucleotides (Dharmacon, ON-TARGETplus SMARTpool Cep164) using Lipofectamine RNAimax (Invitrogen) in 6 well plates seeded with cells at 50% confluency. Cells were lysed 24, 48 and 72 h after transfection and total RNA was isolated (RNeasy Mini Kit, Qiagen, 74106) and measured (NanoDrop spectrophotometer ND-1000, Thermo Fischer Scientific Inc.). cDNA was synthesized from 500 ng RNA template using the iScript cDNA Synthesis Kit (Bio-Rad, 170-8891) according to the supplier's protocol. Dilutions were made for RT-QPCR analysis to determine the expression of Cep164, normalized against reference gene RPL27. The primers (Sigma) used are mCep164 forward 5"-AGAGTGACAACCAGAGTGTCC, mCep164 reverse 5"-GGAGACTCCTCGTACTCAAAGTT, mRPL27 forward 5"-CGCCCTCCTTTCCTTTCTGC and mRPL27 reverse 5"-GGTGCCATCGTCAATGTTCTTC. The iQ SYBR Green Supermix (Bio-Rad, 170-8880) was used to multiply and measure the cDNA with a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). All samples were run in triplicate in 20 µl reactions. The following PCR program was used: 95°C for 3 min, followed by 40 cycles of 10 s at 95°C, 30s at 51.5/61°C and 30 s at 72°C, then 10 s at 95°C followed by a melt of the product from 65°C-95°C. The ΔΔCT method was used for statistical analysis to determine gene expression levels.

Immunofluorescence and confocal microscopy. For immunostaining mIMCD3 cells, cells were grown on coverslips and fixed for 30 minutes in 4%PFA followed by a 15 minutes permeabilization step in 0.5%Triton-X100/1%BSA/PBS. Primary antibody incubations (human anti-CREST at 1:1,000, mouse anti-acetylated tubulin at 1:20,000) were performed overnight in 1% BSA/PBS. Secondary antibody incubations were performed for 1 hour at RT. DAPI incubations were performed for 10' at RT. Coverslips were mounted in

Fluormount G (Cell Lab, Beckman Coulter). Confocal imaging was performed using Zeiss LSM510 Confocal laser microscope and images were processed with the LSM software. Approximately 250 events per condition were scored. GraphPad Prism 5.0 was used to perform two-tailed student t-tests.

Cell cycle studies.

Generation of IMCD3 and hTERT-RPE cell lines that are doxycycline-inducible for N-GFPCEP164 constructs. Both IMCD3 and hTERT-RPE cells were stably transfected with N-terminally GFP tagged human CEP164 constructs in a retroviral vector for doxycycline (Dox)-inducible expression (pRetroX-Tight-Pur). Following selection on puromycin for two weeks, cells were induced with 10 ng/ml of doxycyclin for 20 h. Based on the GFP expression levels upon immunofluorescence, clonal cells were generated for IMCD3 cells.

Cell cycle studies. IMCD3 cells Dox-inducible for human N-GFP-Cep164-WT (wild type) or N-GFPCep164-Q525X (mutant) were transfected with either non-targeting control siRNA (50 nM, Dharmacon D-001206-14-20) or mouse Cep164 siRNA (50 nM, Smartpool Dharmacon L-057068-01-0020) using Polyplus™ transfection reagents. Cells were then re-plated and treated for double thymidine (2 mM) blocks beginning at 24 h post transfection. At the same time, cells were also induced with doxycycline (10 ng/ml) for N-GFP-Cep164-WT or N-GFP-Cep164-Q525X. Cells were released from second thymidine block for 6 h and fixed with 2% PFA and stained with PI/RNase staining solution (BD Biosciences). FACS analysis for cell cycle histograms was performed and data were then analyzed using Modfit™ software. Mean and SD of % DNA amount for different phases (triplicate samples) were calculated and plotted as bar diagram.

Cell Proliferation studies. IMCD3 cells Dox-inducibly expressing human N-GFP-CEP164-WT or NGFP-CEP164-Q525X were transfected with either non-targeting control siRNA (50 nM, Dharmacon D-001206-14-20) or mouse Cep164 siRNA (50 nM, Smartpool Dharmacon L-057068-01-0020) using Polyplus™ transfection reagents. Cells were then re-plated in triplicate of 10,000 cells for each group and treated for thymidine (2 mM) blocks beginning at 24 h post transfection. At the same time, cells were also induced with doxycycline (10 ng/ml) for N-GFP-CEP164-WT or N-GFP-CEP164-Q525X expression. Cells were counted at 48, 72, and 96 h post transfection.

Roscovitine studies.

Cell culture and cell irradiation. Sh-Neg-IMCD3 cells were maintained in DMEM:F12 supplemented with 10% FCS, 1% Penicillin/Streptomycin and 5 μg/ml Puromycin (all from Gibco/Invitrogen, Carlsbad, CA) and maintained in a 5% CO_2 humidified atmosphere at 37°C. Roscovitine (Chem Partners, China) was incubated at 80 μM for 24 h. Ultraviolet (UV) irradiation was performed using a UV StratalinkerTM (Stratagene/Agilent, Santa Clara, CA) at 30 J/m2 dose; cells were left to recover for 1h at 37°C before further analyses.

Immunoblotting, immunofluorescence and subcellular fractionation. Cells were lysed as previously described (Bukanov et al., Nature 444, 949-952 2006). Subcellular fractionation was performed using Nuclear Extract Kit according to the manufacturer's instructions (Active Motif, Carlsbad, CA). Protein concentration was determined by BCA protein assay (Pierce/Thermofisher, Rockford, IL). Proteins were resolved on SDS PAGE 5-12% gradient and transferred using iBlot system (Invitrogen). The following antibodies were used: CEP164 (Novus Biologicals, Littleton, CO), Phospho-Chk1 (Ser317) and phospho-H2AX (Ser139) (Cell Signaling, Danvers, MA), Chk1 (Upstate/Millipore, Billerica, MA), H2AX and 13.3.3 (AbCam, Cambridge, MA) and Sam-68 (Santa Cruz Biotech. Santa Cruz, CA). Primary antibodies were detected with anti-rabbit or anti-mouse horseradish peroxidase-conjugated secondary antibodies (Promega, Madison, WI) and revealed by ECL (Amersham, Little Chalfont, Buckinghamshire, UK) as previously described (Bukanov et al., 2006, surpa). For immunofluorescence, cells were fixed for 20 min in 4% paraformaldehyde (Electron Microscopy Science, Hatfield, PA) in PBS (pH 7.2) and were permeabilized with 0.1 % Triton X100 in PBS. Immunofluorescence was then performed as previously described (Smith et al., J Am Soc Nephrol 17, 2821-2831 2006).

Gene expression analysis. Gene expression analysis was performed as described before (Smith et al, JASN, 2006, 17-pp2821-2831) using TAQ-Man gene expression assays (Applied Biosystems/Invitrogen, Carslbad, CA).

Interaction studies of CEP164 with TTBK2 and CCDC92.

Yeast two-hybrid. A GAL4-based yeast two-hybrid system was used to screen for binary CEP164 interactors by using methods described by Letteboer and Roepman (Letteboer and Roepman, 2008). Bait constructs expressing full length CEP164 (CEP164fl) and several

fragments thereof (CEP1641-550, CEP164551-1100 and CEP1641101-1460) were used to screen two retinal cDNA libraries: a bovine library of randomly primed retina cDNA and a human library of oligo-dT primed retina cDNA (Letteboer and Roepman, Methods Mol Biol 484, 145-159 2008).

GST pulldown. Full length 3xFlag-CCDC92 and 3xFlag-TTBK2 were expressed in COS-1 cells. CEP164fl, CEP1641-550, CEP164551-1100 and CEP1641101-1640 were cloned in pDEST15 and the resulting expression constructs were transformed in BL21DE3 to express glutathione S-transerase (GST) fusion proteins. The GST pulldown was performed as described by Coene et al. (Coene et al., Hum Mol Genet 20, 3592-3605 2011). The results were assessed by immunoblotting, and Flag-tagged proteins were detected using anti-Flag primary antibodies (Sigma-Aldrich) and goat-anti-mouse IRDye800 secondary antibodies (Rockland, Immunochemicals, Gilbertsville, PA, USA). A Li-Cor Odyssey 2.1 scanner was used for imaging.

Co-immunoprecipitation. 3xHA-CEP164fl was expressed in combination with either 3xFlag- CCDC92fl or 3xFlag-TTBK2fl in COS-1 cells (and vice versa with swapped tags). 3xFlag-dNp63 or 3xHA-dNp63 was used as a control for specificity. The co-immunoprecipitation was performed as described by Coene et al. Transfected COS-1 cells were lysed and incubated with α-M2-agarose from mouse (Sigma-Aldrich) or α-HA affinity matrix from rat (Roche) for 2 h at 4°C. After washing, sample buffer was added to the beads and the sample was heated. Beads were then precipitated by centrifugation and the supernatant was analyzed by immunoblotting to assess if TTKB2 and CCDC92 indeed coimmunoprecipitated with CEP164.

Interaction studies of CEP164 with DVL3. Immunoprecipitation coupled to mass-spectrometry. Mouse embryonal fibroblasts were grown in DMEM supplemented with 10% fetal bovine serum at 37˚C to confluency and cell pellets are frozen at -80˚C. Cells were then lysed in lysis buffer (0.5% NP40; 150mM NaCl; 50 mM Tris pH 7.4) and subjected to immunoprecipitation with antibodies against Dvl3 (sc-8027); Dvl2 (sc-8026) or control IgG (sc-2025; Santa Cruz Biotechnology). Proteomics grade trypsin (Sigma) was added to the beads directly (10 ng/µl) and left at 37˚C for 12 h. Tryptic digested peptides were desalted using ZipTip C18 column (Millipore) based on the manufacturer's protocol. LC-MS/MS was performed on a NanoAcquity UPLC (Waters) on-line coupled to an ESI Q-TOF Premier

(Waters) mass spectrometer. Trypsin digested peptides eluted from ZipTip column were diluted in autoclaved M.Q. water (Millipore) and loaded onto a 180 μm x 20 mm nanoAcquity UPLC Symmetry trap column (Waters) packed with 5 μm BEH C-18 beads. After 1 minute of trapping, peptides were eluted through a 75 μm x 150 mm nanoAcquity (Waters) analytical column packed with 1.7 μm BEH C-18 beads at a flow rate of 400 nL/min using a gradient of 3 – 40% acetonitril with 0.1% formic acid for 35 minutes at a temperature of 35°C. Effluent was directly fed into the ESI source of the mass spectrometer. Raw data was acquired in data independent MSE Identity (Waters) mode. Precursor ion spectra were acquired with collision energy 5 V and fragment ion spectra with a collision energy 20-35 V ramp in alternating 1 sec scans. Raw data was then subjected to a database search using species specific Uniprot and NCBI mouse protein database by the PLGS2.3 software (Waters). Acetyl N-terminal, Deamidation N and Q, Carbamidomethyl C and Oxidation M were set as variable modifications. Peptide accuracy and MS/MS fragment mass accuracy was less than 20 ppm.

Western blotting and immunoprecipitation. Immunoblotting and sample preparations were performed as previously described (Bryja et al., J Cell Sci 120, 586-595 2007b). HEK293 cells were transfected with the indicated constructs and lysed 48 h later. Samples were analyzed using SDS-PAGE and Western blotting, or subjected to immunoprecipitation. The antibodies used include the following: Dvl3 (sc-8027), Dvl2 (sc-8026), mouse IgG (sc-2025), from Santa Cruz Biotech, CEP164 (4533.00.02) from Sdix, HA.11 (MMS-101P) from Covance, GFP (3H9), RFP (3F5) from Chromotek and FLAG M2 (F1804), GST (G1160) from Sigma. Immunoprecipitation was performed as previously described (Bryja et al., Proc Natl Acad Sci USA 104, 6690-6695 2007a). The antibodies used for immunoprecipitation were RFP Trap from Chromotek, HA.11 (MMS-101P) from Covance, anti FLAG (F1804; Sigma), anti Dvl3 (sc-8027), anti Dvl2 (sc-8026) (all Santa Cruz Biotechnology).

GST pulldown. Production of recombinant GST-tagged fragments of Cep164 (Sivasubramaniam et al., Genes Dev 22, 587-600 2008) was induced by adding 0.2 mM IPTG and grown for 4 h at 37°C. The bacteria were spun down at 4000 g/4°C/10 min and the pellet was resuspended in 10 ml of GST lysis buffer (50 mM Tris-Cl pH 7.5, 150 mM NaCl, 5mM MgCl2, protease inhibitors (Roche)) and stored at -80°C. Then the solution was thawed, sonicated for 3x30 seconds and spun down 15000 g/4°C/15 min. The supernatant was then used for incubation with GST beads for 2 h at rotator (100 μl of beads per 100 ml of original

bacterial culture). After incubation beads were washed 3 times with 1 ml of GST lysis buffer and frozen at -80°C as 25% slurry in GST lysis buffer +20 % glycerol. HEK293T cells were transfected according to the scheme, grown for 24 or 48 h and lysed in 0.5 % NP 40 lysis buffer (0.5 % NP40, 50 mM Tris pH 7.4, 1 mM EDTA, 150 mM NaCl) with added protease inhibitors (Roche) and phosphatase inhibitors (Calbiochem). The lysate was spun down at 16100 g/4°C/10 minutes and the supernatant was used for overnight incubation with 15 µl of solid GST beads containing GST recombinant proteins on the rotator. After the incubation samples were washed with 800 µl 0.5 % NP 40 lysis buffer, the GST beads were collected at 0.1 g/4°C/1 min, the supernatant was aspirated out and this washing was repeated 6 times. Proteins were eluted with 45 µl of 2x Laemmli buffer.

Immunohistochemistry. HEK-293 cells were seeded at approx. 2x105 cells/well on collagen coated coverslips in 24-well plates. Cells were fixed in fresh 4% paraformaldehyde, permeabilized with 0.05% Triton-X100, blocked with PBTA (3% BSA, 0,25% Triton, 0,01% NaN3) for 1 hour and incubated overnight with primary antibodies (CEP164 (4533.00.02) from Sdix and Dvl2 (sc-8026 from Santa Cruz Biotechnology). Next day, coverslips were washed in PBS, and incubated with secondary antibodies: Alexa 488, Alexa 568 (Invitrogen), washed with PBS, stained with DAPI (1:5000) and mounted on coverslips. Cells were visualized using a Leica TCS SP-5 confocal microscope.

zebrafish ruvbl1 morphants. ruvbl1hi1055b/+ heterozygous fish were acquired from the Zebrafish International Resource Center (ZIRC, Oregon). For histological analyses, embryos were fixed at 72 hpf with 4% PFA/PBS and embedded in JB-4 resin (PolySciences) following the manufacturer‟s protocol. 6-µm sections were obtained using a Leica R2265 microtome and stained with methylene blue following published procedures (Zhou et al., Am J Physiol Renal Physiol 299, F55-62 2010). For γH2aX staining, embryos were fixed at 27 hpf with 4% PFA/PBS +1% DMSO for overnight, permeablized with acetone at -20°C for 7 minutes, and stained with antibody against phosphorylated zebrafish H2AX (1:1,000, from Amatruda lab at UT Southwestern). Alex488-anti rabbit IgG was used at 1:1,000. The IF procedure followed a standard protocol. After washing off secondary antibody, embryos were mounted in Vector-Shield Mounting Media and imaged with Leica SP5X confocal system. Z-stacks were processed to reduce noise and overlayed to show the representative images of IF staining for γH2AX. Morpholinos were: cep164 MO: 5‟-TATATGCTCTTCTCCATCACCTCAT; p53 MO: 5‟-GCGCCATTGCTTTGCAAGAATTG.

Coimmunoprecipitation with Ruvbl1. HEK293T cells were transiently transfected using the calcium phosphate method and the indicated plasmids and co-immunopresipitation was performed as described previously (Habbig et al., J Cell Biol 193, 633-642 2011). Enhanced chemiluminescence was detected with a digital imaging system (Fusion, Peqlab). Human Ruvbl1 and human NPHP5 were cloned from a human cDNA library. Human NPHP1 and NPHP4, mouse NPHP2 and human EPS15L11-225 have been described previously (Habbig et al., 2011, supra; Liebau et al., J Biol Chem 286, 14237-14245 2011; Otto et al., Nat Genet 34, 413-420 2003).

Multicolor Competition Assay (MCA). MCA was performed as described previously (Smogorzewska et al., Cell 129, 289-301 2007). Briefly, 1.7 x 10^5 U2OSGFP were reverse transfected with 64 pmol siRNA duplex using Lipofectamine RNAiMAX (Invitrogen) as per manufactures instructions. 48 h post transfection siRNA treated U2OS-GFP were mixed with anti-Luciferase siRNA treated U2OS-RFP cells at a 1:1 ratio. 72 h post transfection cells were treated with DNA damaging agents as depicted. Cells were harvested for FACS analysis 7 days post treatment. siRNA duplex sequences were as follows:
Luc: CGTACGCGGAATACTTCGA (siRNA targeting firefly luciferase mRNA), CEP164#1 GGACCATCCATGTGACGAA,
CEP164#2: GGCTGGAACGTGTCAAGAA
CEP164#3: GAGTTGGAGTCTCAACAGA,
ZNF423#1: GGAGAACCACAAGAACATT
ZNF423#2: CGAGTGCAGTGTCAAGTTT,
ZNF423#3: GCATCAACCACGAGTGTAA, all from Ambion.
BRCA2: (Stealth siRNA, Invitrogen) used as a combination of three siRNAs:
GGAACCAAATGATACTGATCCATTA, GGAGGACTCCTTATGTCCAAATTTA, GAGCGCAAATATATCTGAAACTTC;
ATM (Stealth siRNA, Invitrogen) used as a combination of three siRNAs):
GCGCAGTGTAGCTACTTCTTCTATT, GGGCCTTTGTTCTTCGAGACGTTAT, GCAACATTTGCCTATATCAGCAATT. anti-CEP164 antibody for immunoblot was from Novus (cat #45330002).

Quantitative RT-PCR. Intron spanning primer pairs were designed for quantitative RT-PCR of ZNF423 and beta-actin as control. RNA was extracted from the siRNA treated

U2OS cells using QIAGEN RNeasy Plus Mini Kit. First-strand cDNAs were then synthesized using the Invitrogen SuperScript III Reverse Transcriptase kit. Quantitative RT-PCR was performed using Platinum SybrGreen Super Mix (Invitrogen) according to the manufacturer's instructions. Sequences of primer pairs were: Beta-actin forward: GCTACGAGCTGCCTGACG, Beta-actin reverse: GGCTGGAAGAGTGCCTCA, ZNF423 forward: GTCTCTGGCAGACCTGACG, ZNF423 reverse: AGAGTTGTGGGTCGTCATCA.

ZNF423 DNA damage response. Lentivirus constructs were obtained in a pLKO vector (Sigma) modified to express Emerald GFP in place of the Puromycin marker. Transduction efficiencies were ~70% by GFP fluorescence. For DDR assays, transduced cells were plated at 1x105 per well in 12-well plate containing poly-L-lysine coated glass cover slides one day before irradiation with a calibrated source (Moores UCSD Cancer Center). Two h after irradiation, cells were fixed with 4% PFA. Mouse anti-γH2AX (Upstate, clone JBW30) and rabbit anti-ZNF423 antibodies were added at room temperature for 2 h. Donkey anti-mouse (Alexa fluor 555) and anti-rabbit (Alexa fluor 488) antibodies (Jackson Immunoresearch) were incubated at room temperature for 1 h. Images were captured on an Olympus FV1000 confocal microscope. Signal intensities of γH2AX were quantified from image files in ImageJ software. Distribution of foci in Zfp423-knockdown cells did not meet the conditions of a ttest and were analyzed by the Mann-Whitney U (Wilcoxon rank sum) test implemented in R 2.8.1.

FACS analysis of γH2AX. Both control and Cep164 stable knockdown IMCD3 cells were irradiated with indicated doses (Fig. 6Q). Cells were then fixed 60 min post-irradiation with 2% paraformaldehyde, permeabilized with 0.1% SDS and stained with rabbit γH2AX (Cell Signaling). Alexa-fluor-488 conjugated secondary antirabbit antibody was used to analyze the mean fluorescent intensity (MFI) of samples in triplicate in the FACS facility (Cancer Center, University of Michigan). Results shown are representation of mean and SD of triplicate samples.

Statistical analysis. Student's two-tailed non-paired t-tests and normal distribution two-tailed z-tests were carried out using pooled standard error and s.d. values to determine the statistical significance of different cohorts.

RESULTS

Whole exome resequencing accelerates discovery of NPHP-RC genes. Identification of monogenic causes of ciliopathies is limited by their rarity (Attanasio et al., Nat Genet 39, 1018-1024 2007), necessitating methods to identify ciliopathy-causing genes in single families, which include whole exome resequencing (WER). However, WER typically yields hundreds of variants from normal reference sequence as "candidate mutations" (Ng et al., Nature 461, 272-276 2009), whereas only a single-gene mutation will represent the disease cause. To overcome this limitation, WER was combined with homozygosity mapping (Hildebrandt et al., PloS Genetics 5, 31000353 2009c) in sib pairs affected with NPHP-RC and performed functional analysis of the identified genes (Otto et al., Nat Genet 42, 840-850 2010a).

Homozygosity mapping yielded positional candidate regions of homozygosity by descent (Hildebrandt et al., 2009c) in families A3471 (2 regions), F874 (9 regions), and KKESH001-7 (14 regions) (Figure 1), who had one or more features of NPHP-RC, including NPHP, retinal degeneration, liver fibrosis, or cerebellar degeneration/hypoplasia (Table 1). WER was then performed in one affected individual of each of the three NPHP-RC families (Ng et al., 2009, supra; Otto et al., 2010a, supra). Each of three NPHP-RC genes consecutively identified by this approach, MRE11, ZNF423 and CEP164, indicated a functional connection to the DDR pathway (Figure 1, Table 1).

A mutation of MRE11 causes progressive cerebellar degeneration. In family F3471 two siblings had cerebellar vermis hypoplasia (CVH), a central feature of NPHPRC (Table 1). Mapping regions of homozygosity by descent yielded 2 candidate loci (Figure 1A). WER detected a homozygous truncation mutation (p.R633X) of MRE11 (Figure 1B; Table 1) previously described for CVH in a Pakistani family (Stewart et al., 1999). Family F3471 is also from Pakistan, indicating a founder effect for this allele. MRE11 is an essential component of the ATMChk2 pathway of DDR (Figure 7), where it recruits ATM (ataxia telangiectasia-mutated) to sites of DNA double-strand breaks (Figure 7A). Rediscovery of this MRE11 mutation in family F3471 thus generated an unexpected link between NPHP-RC phenotype and the ATM pathway of DDR signaling (Figure 7A).

Patients with the NPHP-RC Joubert syndrome have defects in ZNF423. Another link of NPHP-RC to the ATM pathway of DDR signaling emerged from homozygosity mapping

and WER in two siblings (F874) with infantile onset NPHP, CVH, and situs inversus (Table 1). SNP mapping yielded nine candidate regions of homozygosity by descent (Figure 1C). Both affected individuals had a homozygous missense mutation (p.P913L; conserved in vertebrates) of ZNF423 (Figure 1D). In addition, when examining 96 additional Joubert syndrome subjects, two heterozygous-only mutations of ZNF423 were detected: p.P506fsX43 in family A106 and p.H1277Y in individual A111-21 (Table 1). Mutations of the mouse ortholog Zfp423 cause reduced proliferation and abnormal development of midline neural progenitors resulting in a loss of the cerebellar vermis (Alcaraz et al., Proc Natl Acad Sci U S A 103, 19424-19429 2006; Cheng et al., Dev Biol 307, 43-52 2007) similar to that seen in Joubert syndrome patients with CVH.

ZNF423 encodes a protein with 30 zinc fingers (Figure 2A). A mouse mutation disrupting only the last zinc finger, similar to p.H1277Y, is sufficient to cause a severe phenotype (Cheng et al., 2007, supra). Mouse models also display phenotypic variability that is subject to modifier genes, environment, and stochastic effects (Alcaraz et al., Hum Mol Genet 2011; Alcaraz et al., 2006, supra), consistent with the variable presentations of NPHP-RC patients. The homozygous mutation p.P913L, located between zinc fingers 21 and 22 (Figure 2A), may exert recessive loss-of-function, analogous to the Zfp423 mouse models. Of the heterozygous-only mutations, one (p.N506fsX43) truncates the protein before the 11th zinc finger, whereas the other mutation (p.H1277Y) abrogates a zinc coordinating histidine that is part of the "knuckle" in the last zinc finger, which is required for interactions with EBF family transcription factors (Tsai and Reed, Mol Cell Biol 18, 6447-6456 1998), and is important for ZNF423 function (Figure 2A).

It was next examined whether the heterozygous-only mutations (Table 1) lead to loss of function via a dominant mechanism, using a proliferation assay in P19 cells (Figure 2B-D). Mutations were engineered into a FLAG-tagged ZNF423 cDNA and assayed by a S-phase index, defined as the proportion of transfected cells that incorporate BrdU in 1 h, 48 h after transfection. Simple loss of function alleles should not interfere with endogenous Zfp423 activity in this assay. Indeed, overexpression of either wild-type or the homozygous p.P913L allele had no effect (Figure 2D). However, transfection with either the p.P506fsX43 frame-shifting allele, which removes the zinc fingers required for SMAD and EBF interactions, and the H1277Y substitution allele, which destroys the terminal zinc finger required for EBF interaction, reduced the mitotic index to little more than half that of cells transfected with GFP control vector or other alleles of ZNF423 (Figure 2B-D). A dominant mechanism is plausible for the two heterozygous mutations, as each is predicted to interfere selectively with

a subset of interaction domains (Figure 2A). Neither subject had siblings, and DNA from parents was not available to determine whether the mutations occurred de novo.

Five additional putative mutations were identified in highly conserved (including histidine knuckle) residues of ZNF423 among Joubert syndrome families (Table 2). While these mutations have not been confirmed functionally, the high incidence of predicted deleterious mutations found in patients but absent from 270 healthy control individuals, dbSNP, and 1,000 Genomes Project data further support identification of ZNF423 as a causal gene in NPHP-RC and JBS.

ZNF423/OAZ was recently shown to interact with the DNA ds-damage sensor PARP1 (poly-ADP ribosyl polymerase 1) (Ku et al., 2003), which recruits MRE11 and ATM to sites of DNA damage (Figure 7A). This indirectly linked ZNF423 to the ATM pathway of DNA damage signaling (Figure 7A). It was therefore tested whether ZNF423 mutations affect interaction between ZNF423 and PARP1. Coimmunoprecipitation verified the association of ZNF423 and PARP1 in reciprocal assays (Figure 2E). The truncating mutation P506fsX43, which was detected in a JBTS patient (Table 1), abrogates this interaction (Figure 2E), while H1277Y inhibits multimerization of ZNF423 (Figure 2E). In addition, depletion of ZNF423 mRNA caused sensitivity to DNA damaging agents (see below). Furthermore, ZNF423 was identified as a direct interaction partner of CEP290/NPHP6, which is mutated in NPHP-RC (Sayer et al., Nat Genet 38, 674-681 2006; Valente et al., Nat Genet 38, 623-625 2006). In a yeast two-hybrid screen of human fetal brain library with a CEP290 (JAS2; amino acids 1917-2479) 'bait' 3 in-frame “prey” sequences corresponding to ZNF423 (amino acids 178-406) were found. This interaction was confirmed in a direct yeast two-hybrid assay (Figure 2F) and by co-immunoprecipitation of CEP290/NPHP6 with ZNF423 expressed in HEK293T cells (Figure 2G). CEP290/NPHP6 is known to interact with the NPHP-RC protein NPHP5 (Schafer et al., 2008) and localizes to the ciliary transition zone (Sang et al., Cell 145, 513-528 2011). A nuclear function of CEP290/NPHP6 is likely: it contains a nuclear localization sequence (NLS), binds the transcription factor ATF4, and localizes to the nucleus by cell fractionation (Sayer et al., 2006, supra). Direct binding between ZNF423 and CEP290/NPHP6, whose mutation also causes a CVH phenotype (Sayer et al., 2006; Valente et al., 2006), is consonant with other NPHP-RC protein interactions in dynamic complexes (Sang et al., 2011, supra).

Mutations of CEP164 cause NPHP-RC. Leber congenital amaurosis (LCA) is an early-onset form of isolated retinal degeneration (RD) that can be allelic with NPHP-RC. For

example, null mutations of CEP290/NPHP6 cause severe multiorgan NPHP-RC variants of JBTS and MKS syndromes (Helou et al., J Med Genet 44, 657-663 2007), whereas hypomorphic mutations cause LCA only (Chang et al., Hum Mol Genet 15, 1847-1857 2006; den Hollander et al., Am J Hum Genet 79, 556-561 2006). By homozygosity mapping in a Saudi family (KKESH001) of first-cousin parents with a child who had LCA with nystagmus, hyperopic discs, vascular attenuation, diffuse retinal pigment epithelium atrophy, and non-recordable ERG (Table 1), 14 candidate regions (Figure 1E) were identified. By whole exome resequencing a homozygous point mutation in CEP164 (centrosomal protein 164 kDa) was identified that abolished the termination codon, adding 57 amino acid residues to the open reading frame (p.X1460WfsX57) (Figure 1F, Table 1). The mutation was absent from 96 Saudi healthy controls and from 224 North American LCA patients who lack mutations in other known LCA genes.

CEP164 was also considered as a candidate gene for NPHP-RC, because it is part of the human centrosomal proteome (Andersen et al., Nature 426, 570-574 2003). Exon-PCR and Sanger sequencing of all 31 coding exons for one affected individual in each of 856 different NPHP-RC families was performed. Both mutated CEP164 alleles were detected in each of 3 additional families with NPHP-RC (Table 1; Figure 8). Specifically, i) a homozygous missense mutation (p.Q11P; conserved to Chlamydomonas) was detected in two siblings of family F319 with Senior-Loken syndrome (NPHP with retinal degeneration) (Table 1), ii) compound heterozygosity for a missense mutation (p.R93W; conserved to Chlamydomonas) and a truncating mutation (p.Q525X) in three siblings of family F59 with Senior-Loken syndrome, one of whom also had seizures and intellectual disability and, iii) a homozygous nonsense mutation (p.R576X) in individual NPH-505 with Senior-Loken syndrome, partial vermis aplasia and bronchiectasis (Table 1) were detected. In individuals with LCA heterozygous mutations, for which a second recessive allele was not detected, were identified. These were p.R93W, also present in F59, and p.Q1410X, which was detected in two families with phenotypic features of Joubert syndrome and oral-facial-digital syndrome (OFG6), and heterozygously in two out of 96 healthy control individuals. All other mutations were absent from >96 ethnically matched healthy control individuals and from healthy controls of the 1000 genomes project (Table 1). Recessive mutations of CEP164 were identified as a new cause of NPHP-RC. Because of the significant overlap of phenotypic features with other forms of NPHP-RC, the alias "NPHP14" was identified for the CEP164 protein.

Although the number of families with CEP164 mutation is small, the findings indicate a genotype phenotype correlation (Table 1). First, the patient with the mildest phenotype had isolated LCA only and a homozygous mutation (p.X1460WfsX57) that removes the termination codon, adding 57 new amino acids to the C-terminus. This may reduce protein function, because overexpression of full length CEP164 with a C-terminal GFP tag strongly reduced centrosomal labeling compared to an Nterminal GFP tag (Figure 3A-D). Second, the homozygous missense allele (p.Q11P) caused a multiorgan phenotype of NPHP with retinal degeneration (Senior-Loken syndrome; SLSN). Third, compound heterozygosity for one missense and one truncating allele (p.R93W; p.Q525X) caused SLSN with additional central nervous system involvement. Fourth, a homozygous truncating mutation (p.R576X) caused a severe phenotype that combined Joubert syndrome (NPHP, retinal degeneration and vermis hypoplasia) with bronchiectasis, a classic feature of primary ciliary dyskinesia (PCD), a disease of motile cilia. These data further support the gradient of genotype-phenotype correlations characteristic of NPHP-RC, in which null mutations cause the severe dysplastic phenotypes of Meckel syndrome and Joubert syndrome, whereas hypomorphic alleles cause the milder degenerative phenotypes of NPHP and SLSN (Hildebrandt et al., N Engl J Med 364, 1533-1543 2011).

CEP164 is transcribed into 3 common isoforms (Figure 8A-C). In addition to the centrosome, CEP164 is part of the photoreceptor sensory cilium proteome (Liu et al., Mol Cell Proteomics 6, 1299-1317 2007). Homozygous truncating mutations found in exons common to isoforms 1 and 3, but not in isoform 2 support the idea that isoforms 1 and/or 3 are relevant for the NPHP-RC phenotype (Figure 8B). The deduced CEP164 amino acid sequence has a WW domain (aa 57-89), which interacts with the DDR protein ATRIP (Figure 8D). CEP164 also contains a segment of 6 coiled-coil domains (Figure 7C), which are frequently found in NPHP-RC genes (Hildebrandt et al., 2009a). CEP164 is conserved across species including the green alga Chlamydomonas reinhardtii, indicating a conserved function with strong sequence constraints. To study expression and subcellular localization of the CEP164 protein antibodies against human CEP164 were utilized by immunoblotting and immunofluorescence (Figure 9).

Mutation of CEP164 abrogates mother centriole localization. By confocal microscopy of GFP-labeled CEP164 protein with other labels, it was shown that CEP164 colocalizes in hTERT-RPE cells with the mother centriole, with the mitotic spindle poles, and with the abscission structure in a cell cycle-dependent way (Figure 10), a feature characteristic of

proteins involved in single-gene ciliopathies (Otto et al., Nat Genet 42, 840-850 2010a) (Graser et al., J Cell Biol 179, 321-330 2007). Subcellular localization of CEP164 in mouse retina (Figure 9C) as well as in MDCK2 (Madin Darby canine kidney) cell lines, IMCD3 (mouse kidney inner medullary collecting duct) cell lines (Figure 3), and hTERT-RPE (human retinal pigment epithelium) cell lines (Figure 9D-F, Figure 11A) that stably express GFP-tagged human full length CEP164 wild type isoform were generated from a doxycyclin-inducible stable expression construct (NGFP-hCEP164-WT). It was found using the α-CEP164-NR antibody in MDCK cells as well as upon induction with doxycyclin (10 ng/ml) of N-terminally GFP tagged NGFP-hCEP164-WT in IMCD3 cells, CEP164 localizes to mother centrioles of cells (Figure 3A-B). In contrast, the signal at centrosomes was abrogated upon overexpression of an N-terminally GFP-tagged truncated CEP164 construct representing the mutation p.Q525X (Figure 3C). Furthermore, the signal occurred at a reduced number of centrosomes upon overexpression of C-terminally GFP-tagged human full-length construct (Figure 3D), which mimics the mutation p.X1460WfsX57 occurring in NPHP-RC family KKESH001 that causes a read-through of the stop-codon X1460, adding 57 aberrant amino acid residues to the Cterminus of CEP164 (Table 1). Corresponding data were obtained upon CEP164 expression in hTERT_RPE cells (Figure 9D-E). A lack of centrosomal localization for the truncating mutation p.Q525X and for an equivalent of the p.Q1460WextX57 mutation was demonstrated.

Endogenous Cep164 levels regulate cilia in 3D epithelial spheroid structures. Loss of function of several genes that cause nephronophthisis in NPHP-RC cause disruption of 3D architecture of renal epithelial cell culture (Otto et al., Nat Genet 42, 840-850 2010a; Sang et al., Cell 145, 513-528 2011). To evaluate CEP164 by this criterion, murine kidney IMCD3 cells were transfected with siRNA oligonucleotides against murine Cep164, or random sequences (Ctrl) in 3D spheroid growth assays. After 3 days, siCtrl transfected cells formed spheroid structures with a clear lumen, apical cilia, defined tight junctions and clear basolateral structure. Cells transfected with siCep164 developed spheroids with overall normal architecture and size, but with markedly reduced frequency of cilia (Figure 3E-H). While 49% of the spheroid cells transfected with siCtrl generated detectable cilia, only 33% of the siCep164 transfected cells grown in 3D cultures were ciliated ($p<0.0001$). It was concluded that Cep164 affects ciliogenesis or maintenance, but that the overall architecture of renal 3D growths is not as grossly affected as previously seen for knockdown of other NPHP-RC genes (Sang et al., 2011, supra).

To exclude the possibility that the effect observed was due to off-target effects, two IMCD3 clones stably transfected with inducible full-length human CEP164 (IMCD3-NGFP-CEP164- WT clones 2 and 8) were utilized. Both lines form spheroids in 3D cultures. siRNA of the endogenous murine Cep164 was performed and no irregularities except for reduced cilia frequency were observed. However, after doxycycline induction of siCep164-resistant human NGFP-CEP164, ciliary frequencies were restored, with 57% of the siCep164 transfected cells ciliated ($p=<0.0001$) (Figure 3G). Rescue was not observed for IMCD3 cells expressing a patient mutation (NGFP-CEP164-Q525X, Figure 3H). Reducing cellular levels of Cep164 affects renal ciliation frequencies without gross disturbances to tubular architecture, a phenomenon that can be rescued with WT but not mutant CEP164.

NPHP-RC proteins colocalize with the DDR protein TIP60 to nuclear foci. A non-centrosomal localization for CEP164 was recently described by demonstrating its translocation to nuclear foci in response to DNA damage (Pan and Lee, Cell Cycle 8, 655-664 2009; Sivasubramaniam et al., Genes Dev 22, 587-600 2008). In this context CEP164 is thought to play a role in DNA damage-response (DDR) signaling where it interacts with the DDR protein ATRIP (Figure S2D), is activated by the DDR proteins ATM and ATR, and is necessary for checkpoint-1 (Chk1) activation. Abrogation of CEP164 function leads to loss of G2/M cell cycle checkpoint and aberrant nuclear divisions (Sivasubramaniam et al., 2008, supra). Because the NPHP-RC gene products MRE11 and CEP164 are essential components of DDR signaling, and because ZNF423 interacts with the bona fide DDR protein PARP1 (Ku et al., Biochem Biophys Res Commun 311, 702-707 2003), the identification of mutations in these genes support a role of DDR signaling in the pathogenesis of NPHP-RC (Figure 7). Nuclear foci localization for the newly identified NPHP-RC causing gene products ZNF423 and CEP164 and for previously identified NPHP-RC gene products was examined.

Localization of SDCCAG8 (alias NPHP10), in which NPHP-RC mutations were previously identified (Otto et al., 2010a, supra), shows nuclear foci in hTERT-RPE cells in addition to its centrosomal localization (Figure 4B-D). Transient shRNA knockdown confirmed specificity of the signal (Figure 10B-D). SDCCAG8/NPHP10 did not colocalize with markers for PLM bodies (Janderova-Rossmeislova et al., J Struct Biol 159, 56-70 2007) or CENP-C (marking chromosomal centromeres) (Figure 11A). In contrast SDCCAG8/NPHP10 fully colocalized with SC35 in hTERT-RPE cells (Figure 4A-C). SC35, also known as serine/arginine-rich splicing factor 2 (SRSF2), is a splicing factor that plays a

role in DDR by controlling cell fate decisions in response to DNA damaging agents (Edmond et al., EMBO J. 2010; Reinhardt et al., Cell Cycle 10, 23-27 2011). SC35 marks hubs of enhanced gene expression (Szczerbal and Bridger, Chromosome Res 18, 887-895 2010), is phosphorylated by topoisomerase I (Elias et al., Exp Cell Res 291, 176-188 2003), and is required for genomic stability during mammalian organogenesis (Xiao et al., Mol Cell Biol 27, 5393-5402 2007). Moreover, ZNF423 also fully colocalizes (Figure 4D), and CEP164 partially colocalizes (Figure 4E) with SC35 in nuclear foci. Consequently, ZNF423 and CEP164 also colocalize with SDCCAG8/NPHP10 in SC35-positive nuclear foci (Figure 4F,G).

SC35 functions within a TIP60 complex, in which TIP60 acetylates SC35 on lysine 52 (Figure 7B), modifying the role of SC35 in the promotion of apoptosis and inhibition of G2/M arrest (Edmond et al., EMBO J. 2010), which is regulated by the checkpoint proteins Chk1 and Chk2 (Figure S1D). The TIP60 protein, together with the heterotrimeric MRN complex (of which MRE11 is a component) constitutes the major activator of ATM within the ATM pathway of DDR signaling (Ciccia and Elledge, Mol Cell 40, 179-204 2010) (Figure 7A). In hTERT-RPE cells the ATM activator TIP60 colocalizes to nuclear foci with SC5/SRSF2 (Figure 4H) and partially with the newly identified NPHP-RC protein CEP164 (Figure 4I). A new group of NPHP-RC proteins that colocalize to nuclear foci with the DDR proteins TIP60 and SC35 is thus identified. These gene products include the newly identified NPHP-RC proteins in ZNF423 and CEP164 as well as SDCCAG8/NPHP10. The protein OFD1, which is mutated in the ciliopathy oral-facial-digital syndrome, is part of the TIP60 complex, which is a major ATM activator within DDR signaling; OFD1 has been identified as a direct interaction partner of SDCCAG8/NPHP10 (Figure 7B) (Otto et al., 2010a, supra).

CEP164CEP164 colocalizes with the DDR proteins TIP60 and CHK1 in nuclear foci upon DNA damage. Because one of the central mechanisms controlled by DDR signaling is cell cycle regulation through phosphorylation of checkpoint-1 (Chk1) and checkpoint-2 (Chk2) proteins (Figure 7D), it was tested whether checkpoint proteins are recruited to SC25/SRSF2-positive nuclear foci. SC35 and p317-Chk1 colocalize to nuclear foci in hTERT-RPE cells (Figure 4J). As colocalization conveys only a static image of DDR components, it was then tested whether localization of CEP164 to nuclear foci was inducible by DNA damage as suggested (Sivasubramaniam et al., 2008, supra). Following irradiation with 20-50 J/m2 of UV light, CEP164-positive nuclear foci condensed to larger size and colocalized with newly appearing TIP60 foci of similar size (Figure 4K-M). Similarly, a pattern of broad CEP164 speckles, which were CHK1-negative and locate to DAPI-negative

domains in untreated cells (Figure 4N), changed to a pattern of multiple smaller foci that were positive for both CEP164 and CHK1 (Figure 4O). Colocalization of TIP60 and p317-CHK1 in these foci (Figure 4P) demonstrated that the UV-inducible foci observed in HeLa cells are likely equivalent to the foci positive for TIP60 and SC35 that were observed as positive for NPHP-RC proteins in hTERT-RPE cells (Figure 4A-J). It was thus demonstrated that CEP164 translocates in response to DNA damage to nuclear foci that contain the DDR proteins TIP60 and CHK1.

Reducing endogenous levels of Cep164 causes anaphase lagging chromosomes. Lagging chromosomes on anaphase spindles ("anaphase lag") are a hallmark of many mutations that affect mitotic checkpoint integrity. Acute reduction of Cep164 by siCep164 knockdown in IMCD3 cells exacerbated the incidence of anaphase lag from 1% in siCtrl controls to 21% in siCep164-treated cells, in over 250 anaphases scored from unsynchronized cells in a total of five independent experiments (Figure 5A-B, p=0.04). CREST antiserum and DAPI confirmed the presence of incomplete mitotic congression and unattached kinetochores during late anaphase. This phenomenon was specific, since doxycycline-inducible expression of WT-CEP164 during Cep164 siRNA knockdown reduced the incidence of anaphase lag to just 4% (Figure 5B). Untransfected (non-siRNA treatment) IMCD3 cells had no detectable anaphase lag (0%). These data indicate a requirement for Cep164 at the G2/M checkpoint.

Human wild type CEP164 but not its NPHP-RC truncation rescues IMCD3 cell proliferation. In clonally selected IMCD3 cells expressing wild type human CEP164 cDNA construct N-GFPCEP164-WT under doxycycline (Dox) control, depletion of endogenous mouse Cep164 retarded proliferation in comparison to either undepleted control cells or undepleted cells that were Doxinduced to overexpress N-GFP-CEP164-WT alone (Figure 5C). Cep164-depleted growth was rescued by Dox-induced expression of human N-GFP-CEP164-WT (Figure 5C). Cells expressing truncated cDNA construct N-GFP-CEP164-Q525X, modeling the NPHP-RC mutation in family F59, exhibited retarded growth, even when the endogenous Cep164 was present (Figure 5D), consistent with a dominant negative effect. Further depletion of the endogenous Cep164 in N-GFP-CEP164-Q525X expressing cells showed an additive effect on growth retardation, confirming the dominant negative effect of N-GFP-CEP164-Q525X in this experimental system (Figure 5D).

CDK inhibition translocates Cep164 to the nucleus. CEP164 is required for DNA damage-induced phosphorylation of Chk1, and down regulation of CEP164 significantly reduces DDR (Maude and Enders, Cancer Res 65, 780-786 2005; Sivasubramaniam et al., 2008, supra). On the other hand, the DDR pathway can also be activated by the small molecule CDK inhibitor roscovitine, which in addition reduces Chk1 expression (Maude and Enders, 2005, supra). Roscovitine also reduces the development of kidney cysts in the Nphp9 mouse model, Jck (Bukanov et al., Nature 444, 949-952 2006). The influence of roscovitine (targeting CDK2, 5, 7 and 9) on DDR activation in IMCD3 cells was assayed. Immunofluorescence shows increased uniform distribution of γH2AX (activated H2AX phosphorylated at Ser139) in the nucleus of IMCD3 cells upon roscovitine treatment in irradiated cells, indicating partial DDR activation (Figure 5E). Second, in cells treated with roscovitine, UV irradiation caused enhanced γH2AX staining with a prominent nuclear foci pattern, characteristic of strong DDR activation (Figure 5F). Immunoblotting showed that roscovitine decreased the amount of CEP164 present in both control and UV-irradiated cells (Figure 5G-H). This was most likely due to translocation of CEP164 into the nucleus upon roscovitine treatment, as shown by subcellular fractionation (Figure 5H). UV radiation increased phosphorylation of Chk1 at Ser317 (p-Chk1) (Figure 5G), and roscovitine decreased Chk1 protein expression and abrogated UV-induced p-Chk1 in both cytoplasm and nucleus (Figure 5G-H). Together, these data indicate that CDK inhibition by roscovitine causes nuclear translocation of CEP164 and inhibits Chk1 activation. γH2AX activation by roscovitine may restore cell cycle control by Chk2 activation instead (Maude and Enders, Cancer Res 65, 780-786 2005).

CEP164 directly interacts with CCDC92 and TTBK2. NPHP-RC proteins are known to interact with other NPHP-RC proteins in dynamic complexes that have been termed the "NPHP-JBTS-MKS interaction network" (Sang et al., Cell 145, 513-528 2011). In order to identify novel direct interaction partners of CEP164 yeast two-hybrid screening of both a random-primed bovine retina and an oligo-dT primed human retina cDNA library was performed. Full length and partial cDNA clones of CEP164 were used as baits: CEP164fl (encoding the full length protein), CEP1641-550 (encoding amino acids 1-550), CEP164551-1100 (encoding amino acids 551-1100) and CEP1641101-1460 (encoding amino acids 1101-1460). CCDC92 (coiled coil domain containing protein 92) and TTBK2 (tau tubulin kinase 2) were identified as direct interactors of CEP164 (Figure 11CJ). CCDC92 was the protein with the most positive clones in the screen with the human oligo-dT retinal cDNA library, while

TTBK2 appeared to be the main CEP164 interactor in the bovine library. Two overlapping CCDC92 clones were found with baits CEP164fl, CEP1641-550 and CEP1641101-1460 in the human library, however, no CCDC92 clones were identified with the bovine library (Figure 11C). TTBK2 was identified in both screens, with 12 hits (4 different clones) in the bovine, and a single clone in the human retina cDNA library (Figure S5C). Interactions between CEP164 and each of these new partners were validated by GST pulldown (Figure 11D) and co-immunoprecipitation (Figure 11E-H). Immunofluorescence showed that CCDC92 fully colocalizes with CEP164 at the mother centriole (Figure 11I). TTBK2 weakly colocalizes with CEP164 at one of the centrioles, but yields a strong signal at the mid body in dividing hTERT-RPE cells (Figure 11J). Tau tubulin kinase 2 (TTBK2) is a member of the casein kinase family encoded by the gene mutated in cerebellar ataxia type 11 (Houlden et al., 2007). TTBK2 can phosphorylate tubulin, and its kinase activity is required for the phosphorylation of tau by GSK-3β. Tau is a microtubule associated protein that is also found in the nucleus, where it is a key player in the early stress response/DNA damage protection of neurons (Sultan et al., Biol Chem 286, 4566-4575 2011).

CEP164 interacts with NPHP3 and DVL3. To determine whether CEP164 interacts with known NPHP-RC proteins, HEK 293T cells were cotransfected with N-terminally V5-tagged human full-length CEP164 and the seven different FLAGtagged human full length proteins NPHP1-NPHP5, NPHP8, NPHP9 or the control protein CD2AP. NPHP proteins were precipitated, using anti-Flag M2 beads. Interaction of CEP164 with NPHP3 and weakly with NPHP4 was detected (Figure 11A-B), demonstrating that CEP164 is in a complex with other known NPHP-RC proteins (Figure 7A-B). The DDR protein DDB1 interacted with NPHP2 (Figure 12C-D).

The dishevelled protein (Dvl) is a central component of the Wnt pathway and it has been shown that NPHP2/inversin interacts with Dvl targeting it for proteasomal degradation, thereby liberating the β-catenin destruction complex and triggering a switch from canonical to non-canonical Wnt signaling (Germino, Nat Genet 37, 455-457 2005; Simons et al., 2005). In loss of function of the NPHP-RC protein NPHP2/INVS, which causes NPHP type 2 in humans (Otto et al., Nat Genet 34, 413-420 2003), this switch is lacking (Simons et al., Nat Genet 37, 537-543 2005). By immunoprecipitation coupled to mass spectrometry using endogenous Dvl3 as bait interaction between Dvl3 and CEP164 (Figure S6E-H) was identified. Immunocytochemistry revealed that endogenous Dvl3 and CEP164 share centrosomal localization in most cells analyzed (Figure 12A). To further study CEP164-Dvl3

interaction a GST-pull down assay in HEK293T cells, as well as domain mapping, using a set of Dvl3 deletion constructs (Angers et al., Nat Cell Biol 8, 348-357 2006), was performed. It was demonstrated that GST-CEP164 (aa2-195), which was shown to interact with the DDR protein ATRIP (Sivasubramaniam et al., 2008, supra), is sufficient to pull down endogenous Dvl3 from the cellular lysate (Figure 12B). Domain mapping for Dvl3 indicates that CEP164 interacts with the proline-rich region of Dvl3, because only mutants containing this sequence efficiently co-immunoprecipitate with CEP164-GFP (Figure 12C). Only wild type CEP164-mCherryRFP but not the NPHP-RC causing mutant CEP164-Q525X detected in family F59 (Table 1) can be efficiently immunoprecipitated with Dvl3 (Figure 12D), further supporting its pathogenic role.

cep164 loss of function causes NPHP-RC and DDR activation in zebrafish. To test in a vertebrate animal model whether loss of cep164 function results in both, an NPHP-RC phenotype as well as DDR activation, cep164 knockdown was performed in zebrafish embryos using a morpholino-oligonucleotide (MO) that targets the exon 7 splice donor site of cep164 (Fig. 6). A p53 MO was injected to reduce off-target MO effects (Robu et al., PLoS Genet 3, e78 2007). At 28 hours post fertilization (hpf) the ciliopathy phenotypes of ventral body axis curvature and cell death were observed (Fig. 6AC). Embryos showed increased expression of phosphorylated γH2AX (Fig. 6D-E). At 48 hpf, cep164 morphants displayed the typical ciliopathy phenotype of abnormal heart looping as a laterality defect (Fig. 6F-I). Furthermore, at 72 hpf, embryos developed further NPHP-RC phenotypes, including pronephric tubule cysts (Fig. 6J-K), as well as hydrocephalus and retinal dysplasia (Fig. 6L-O). To test in a vertebrate animal model whether loss of cep164 function results in both, an NPHP-RC phenotype as well as DDR activation, cep164 knockdown was performed in zebrafish embryos using a morpholino-oligonucleotide (MO) that targets the exon 7 splice donor site of cep164 (Fig. 6). A p53 MO was injected to reduce off-target MO effects (Robu et al., 2007,supra). At 28 hours post fertilization (hpf) the ciliopathy phenotypes of ventral body axis curvature and cell death were observed (Fig. 6AC). Embryos showed increased expression of phosphorylated γH2AX (Fig. 6D-E). At 48 hpf, cep164 morphants displayed the typical ciliopathy phenotype of abnormal heart looping as a laterality defect (Fig. 6F-I). Furthermore, at 72 hpf, embryos developed further NPHP-RC phenotypes, including pronephric tubule cysts (Fig. 6J-K), as well as hydrocephalus and retinal dysplasia (Fig. 6L-O).

Depletion of CEP164 or ZNF423(Zfp423) causes sensitivity to DNA damaging agents. To assess whether depletion of CEP164 causes sensitivity to DNA damage, Cep164 expression was stably suppressed in the mouse renal cell line IMCD3 (Fig. 6P-Q). Cep164 knockdown resulted in a dose-dependent increase of γH2AX intensity levels in a FACS analysis, signifying increased radiation sensitivity to IR and perturbed DDR. Cellular sensitivity to IR was also seen in cells depleted of CEP164 using a multicolor competition assay (MCA) (Smogorzewska et al., 2007) (Figure S7A-B).

To test whether ZNF423(Zfp423) affects DDR, P19 cells, which express high levels of endogenous Zfp423 (Fig. 6R-T) were examined. Replicate cultures infected with lentivirus expressing either scrambled control or Zfp423-targeted shRNA were exposed to 0-10 Gy of X-irradiation and imaged for Zfp423 and nuclear γH2AX foci (Fig. 6R). Quantification showed significantly increased γH2AX intensities in Zfp423-depleted cells at lower (0.5 and 1.0 Gy) exposures (Fig. 12), but the effect was non-significant when corrected for number of exposures. To determine whether sensitivity to lower dose is reproducible, 32 additional cultures were exposed at 1.0 Gy (Fig. 6T). Normalized γH2AX fluorescence in Zfp423 knockdown had both higher mean (9.6 vs. 4.7) and median (6.6 vs. 5.2) values than control (Fig. 6T). These data replicate the radiation sensitivity with high significance ($p=0.018$, Mann-Whitney U test, 2 tails), indicating that P19 cells require Zfp423 for quantitatively normal DDR.

Table 1. Mutations of *MRE11A*, *ZNF423* and *CEP164* in families with NPHP-RC.

Family -Individual	Ethnic origin	Nucleotide alteration[a,b] (Hg19 position)	Deduced protein change	Exon/ Intron (state)	Continuous amino acid sequence conservation	Parental consan-guinity	Kidney (age at ESKF)	Eye (age at RD)	Other (at age)
MRE11									
F3471 -21 -22	Pakistani	c.1897C>T (Chr11: 94,170,372)	p.R633X	18 (hom)	N/A	Yes	No renal failure	normal	-21: CVA (MRI), ataxia, dysarthria, myoclonus -22: CVA (MRI), ataxia
ZNF423									
F874 -21 -22	Turkey	c.2738C>T (Chr16: 48,570,325)	p.P913L	5 (hom)	(*Danio rerio*)	Yes	NPHP	ND	-21 and -22: CVH Infantile NPHP *Situs inversus*
A188 -21 -22	Iceland	c.1516delC (Chr16: 48,571,545)	p.P506fsX43	5 (het)	(*Xenopus trop*)	No	PKD	LCA	CVH (Joubert)
A111 -21	?	c.3826C>T (Chr16: 48,525,232)	p.H1277Y	9 (het)	(*Danio rerio*)	?	PKD	RD	CVH, NPHP, perinatal breathing abnormality, tongue tumor
CEP164									
F313 -21 -22	Turkey	c.32A>C (Chr11: 117,289,334)	p.Q11P	3 (hom)	*Ch. R.*[c]	Yes	NPHP, no Bx -21: (5yr) -22: (8yr)	-21: RD (EM 1yr, not yet blind) -22: no RP at 8 yrs	-21: obesity? no LF -22: obesity? LF?
F59 -21 -22 -23	USA (Europe)	c.2770C>T, (Chr11: 117,222,588) c.1573C>T (Chr11: 117,262,580)	p.R838X, p.Q525X	5 (het) 13 (het)	*Ch. R.*[c] N/A	No	NPHP, no Bx -21: (8 yr) -22. (8 yr) -23: normal	-21: RD (5 yr) -22: LCA (legally blind at 5 mo) -23: (2 yr)	-22: NY (birth), mild AI -23: seizures[d], substantial DD, mild ID
NPH505	ND	c.1728C>T (Chr11: 117,267,920)	p.R576X	15 (hom)	N/A	Yes	NPHP, Bx (8 yr)	RD and flat ERG (not blind)	CVH, PD, bilateral PD, bronchiectasis (1 mo), abnormal LFT, obesity
KKESH001-7	Saudi	c.4383A>G (Chr11: 117,262,884)	p.X1460W fsX57	33 (hom)	N/A	Yes	normal	(RD) LCA, flat ERG (blind <2 yr)	N/A

[a]All mutations were absent from >270 healthy control individuals and from the ESP Exome Variant Server data base, except the *CEP164* variant p.R578X (allele frequency in European Americans 1/7,019).
[b]cDNA mutation numbering is based on human reference sequences NM_014956.4 for *MRE11*, NM_015069.3 for *ZNF423*, and NP_055771 for *CEP164*, where +1 corresponds to the A of ATG start translation codon.
[c]*Ch.R. Chlamydomonas Reinhardtii*.
[d]Seizures were intractable, generalized and/or partial complex.
AI, aortic insufficiency; Bx, Kidney biopsy; CVH, cerebeller vermis hypoplasia; DD, developmental delay; ERG, electroretinogram; ESKF, end-stage kidney failure; FD, facial dysmorphism; het, heterozygous; hom, homozygous; ID, intellectual disability; LCA, Leber congenital amaurosis; LF, liver fibrosis; LFT, liver function tests; MRI, magnetic resonance imaging; N/A, not applicable; ND, no data; NPHP, nephronophthisis; NPHP-RC, nephronophthisis-related ciliopathies, NY, nystagmus; PD, polydactyly; RD, retinal degeneration. SS, short stature; yr, year/years.

Example 2

MATERIALS AND METHODS

Human subjects. Blood or DNA samples were obtained from 40 individuals diagnosed with CAKUT from 38 different families. The diagnosis was made by pediatric nephrologists on the basis of imaging studies. After informed consent was obtained, detailed clinical data and pedigree information was referred to us by the specialists through a standardized clinical questionnaire. The cohort of 40 individuals included 29 (72.5%) individuals diagnosed with unilateral renal agenesis, 3 (7.5%) individuals with renal hypodysplasia, 7 (17.5%) individuals with vesicoureteral reflux, and 1 (2.5%) individual with bilateral duplex systems.

Whole-genome amplification and DNA pooling. To normalize various DNA samples, whole-genome amplification was performed on DNA of 40 different individuals and 96 healthy control samples using Phi29-based DNA polymerase strand displacement

amplification (GenomiPhi DNA amplification kit, GE Healthcare, Piscataway, NJ). Subsequently, whole-genome amplified DNA was purified using 96-well spin columns (Rapid 96TM PCR purification system, Marligen Biosciences, Reutlingen, Germany). DNA of 20 individuals were pooled at 2 mg each and diluted to 60 ng/ml. Pooling was repeated to represent 40 individuals. In addition, an equimolar DNA pool was generated by pooling 96 DNA samples derived from healthy individuals of Caucasian origin (Human Random Control DNA Panel-1 (HRC-1); European Collection of Cell Cultures, Salisbury, UK).

PCR amplification and massively parallel exon resequencing. All 313 exons (342 amplicons) of 30 candidate genes, ATGR2, BMP4, CDC5L, EMX2, EYA1, FOXC1, FRAS1, FREM2, GATA1, GDF11, GDNF, GFRA1, GREM1, HOX11, HOXA11, KAL1, PAX2, RET, ROBO2, SALL1, SIX1, SIX2, SIX4, SIX5, SLIT2, SPRY1, TCF2 (HNF1b), UPK3A, and USF2, were individually amplified using exon-flanking primers on each of the DNA pools. PCR products from the same pool were combined and enzymatically modified. The modified PCR fragment mixture was then used to construct an Illumina sequencing library using 'Genomic DNA Sample Prep Kit' as previously described (Otto et al., J Med Genet 2010). Each library was run on a single lane of a Solexa/Illumina Genome Analyzer GAII platform, generating about 15–20 million single-end sequence reads of 39 bases each.

Sequence analysis and mutation carrier identification. Sequence alignment was performed with CLC Genomics Workbench software (CLC-bio, Aarhus, Denmark) using imported and annotated human reference genome assembly NCBI36/hg18 as a reference. Sequence reads were mapped to exonic coding regions plus adjacent 100 bp intronic sequence of all 313 exons. Variant calls were obtained using the following filter parameters: Coverage ≥400-fold, variant frequency ≥1%, and a minimum variant count of five reads. The variant analysis included coordinates of obligatory splice sites, and all variants predicted to change the amino-acid sequence (missense, nonsense, and coding indels). Variants present in dbSNP130, the '1000 Genomes Project' (270 control individuals), or in the healthy control pool of 96 individuals (HRC-1) were excluded from further analysis. To prioritize for downstream analysis, missense variants were scored according to the information of evolutionary conservation and the likelihood of a potential protein-damaging effect using PolyPhen software predictions (Ramensky et al., Nucleic Acids Res 2002; 30: 3894–3900). All variants with a predicted 'probably damaging' or 'possibly damaging' effect and a score above 1.4 were further analyzed. The selected variants were amplified by PCR for each

individual in the corresponding DNA pool with subsequent Sanger sequencing to identify the mutation carrier.

RESULTS

Two times 20 DNA samples of individuals with CAKUT (29 with unilateral renal agenesis and 11 with other forms of CAKUT) were pooled and exon PCR was performed in 30 candidate genes. Massively parallel exon resequencing of 342 different exon PCR products generated was carried out on an Illumina/Solexa GAII platform (one pool per lane of a flow cell) delivering about 5.3 million short reads of 39 bases (5.22 million in pool no. 1 and 5.37 million in pool no. 2). Following gapped alignment to the human normal reference sequence, 75% of all reads mapped back to one of the 342 target sequences that contained the exons plus 100-bp adjacent intronic sequence. The sequence concatenation of all 342 amplicons amounts to a total length of 175.7, 68.2 kb of which were exonic coding regions. The median coverage depth for the coding regions was 1228-fold (mean 1402-fold). On average, about 95% of nucleotides in targeted coding regions had at least 400-fold coverage depth. This translates into a depth of at least 20-fold per patient, which is sufficient to cause a heterozygous change.

Mutation carrier identification by Sanger sequencing. Massively parallel exon resequencing of all PCR products of 40 individuals revealed initially a total of 114 variants from normal reference sequence within the coding regions and splice sites of the 30 candidate genes analyzed, 47 of which were known single-nucleotide polymorphisms (SNPs). Eight additional variants were present in a cohort of 96 Caucasian healthy control individuals and are thought to be either as yet unannotated SNPs or false calls due to software base calling or alignment artifacts. Of the remaining 59 variants, 11 were predicted to truncate the protein products and 16 others had a PolyPhen score higher than 1.4 (predicted to be 'possibly damaging'). These 27 variants were assumed to affect the function of encoded proteins, and thus were followed by direct Sanger sequencing to identify the mutation carrier(s) out of the respective two pools of 20 individuals each. This carrier analysis led to the confirmation of 10 of the 27 variants ('true positives') in 11 individuals from 11 different families, whereas 17 of the 27 variants could not be confirmed ('false positives'). Segregation study was conducted in all confirmed cases where DNA of relatives was available and 3 out of 10 variants were not segregating.

The remaining seven variants (Table 2, Figure 13) were heterozygous missense mutations in four genes. Four mutations were in FRAS1 (p.L259R, p.D998Y, p.R3273H, and p.H3757Q) and one in FREM2 (p.T2338I). Mutations in FRAS1 and FREM2 have never been reported in nonsyndromic CAKUT in humans. All five mutations were absent from 96 healthy control individuals of Caucasian origin and 270 control individuals from the '1000 Genomes project'. Two mutations in FRAS1 (p.D998Y and p.H3757Q) and one mutation in FREM2 (p.T2338I), which were detected in individuals of Arabic and Indian origin, respectively, were additionally screened and were absent in 141 ethnically matched healthy controls. As FRAS1 and FREM2 were known to inherit recessively, all other coding exons in respective mutation carriers were amplified and sequenced to exclude the presence of the second mutation. No further mutations were found except in A2381 II-2 (FRAS1; R3273H), where an additional nonsense mutation was detected (R2621X). The R2621X change was 2349-fold covered with frequency of 4.5% but had been systematically masked out. The two mutations (R3273H and R2621X) were in trans as R3273H as heterozygously present only in the mother, whereas R2621X was heterozygously present only in the father. The elder female sibling of A2381 II-2 had died in utero of bilateral renal agenesis and absent bladder; however, the DNA was not available for study.

Two other variants were novel mutations in known CAKUT genes: RET (p.V704F) and BMP4 (p.H121R; Table 2, Figure 13). Family A1143 and A1184 (both Macedonian) shared the same novel change in BMP4 (p.H121R). Both mutations were absent from 96 healthy control individuals and 270 control individuals from the '1000 Genomes project'.

Table 2

Table 1 | Genotypes and phenotypes of eight individuals with unilateral renal agenesis (eight families) with eight different heterozygous missense mutations in one of the 30 candidate genes

Family	Phenotype	Pattern of inheritance	Origin	Gene[a]	Nucleotide change (zygosity state)[b]	Amino-acid change	PolyPhen PICS score[c]	Count/coverage (frequency, %)	Segregation
A2435 II-1	URA	Sporadic	Macedonia	*FRAS1*	c.776T>G (h)	p.L259R	1.692	16/648 (2.5)	NA
A1808 II-1	URA	Sporadic	Kuwait	*FRAS1*	c.2992G>T (h)	p.D998Y	2.336	32/1,388 (2.3)	NA
A2381 II-2	URA	Familial[d]	Austria	*FRAS1*	c.9815G>A (h)	p.R3273H	1.685	32/1,085 (2.9)	Maternal
				FRAS1	c.7861C>T (h)	p.R2621X	NA	105/2349 (4.5)	Paternal
A1425 II-1	URA	Sporadic	Lebanon	*FRAS1*	c.11268C>A (h)	p.H3757Q	2.011	16/1,320 (1.2)	NA
A1623 II-1	URA	Sporadic	India	*FREM2*	c.7013C>T (h)	p.T2338I	2.050	29/1,067 (2.7)	NA
A1077 II-1	URA	Sporadic	Kuwait	*RET*	c.2110G>T (h)	p.V704F	1.537	8/494 (1.6)	NA
A1143 II-1	URA	Sporadic	Macedonia	*BMP4*	c.362A>G (h)	p.H121R	2.001	26/826 (3.1)	NA
A1184 II-1	URA	Sporadic	Macedonia	*BMP4*	c.362A>G (h)	p.H121R	2.001	800/826 (3.1)	Paternal

Abbreviations: h, heterozygous; HRC-1, Human Random Control DNA Panel-1; NA, not available; PICS, predicted impact of coding single-nucleotide polymorphisms; URA, unilateral renal agenesis.

[a]Accession numbers: *FRAS1*, NM_025074; *FREM2*, NM_207361; *RET*, NM_020975; *BMP4*, NM_001202.

[b]Mutation numbering is based on the cDNA position in reference sequences. All changes were absent from 96 healthy control (HRC-1) DNA pool, 270 individuals of the '1000 Genomes project' (http://www.1000genomes.org/page.php), and from additional ethnically matched healthy controls.

[c]PICS score above 1.4 is considered 'probably damaging' (http://genetics.bwh.harvard.edu/pph/).

[d]A2381 II-1 (older female sibling, died in utero) had bilateral renal agenesis with absent urinary bladder. DNA is not available for study.

Example 3

Methods

Research subjects. Blood samples and pedigrees were obtained after obtaining informed consent, from individuals with KIN. Approval for research on human subjects was obtained from the University of Michigan Institutional Review Board and the other institutions involved. Diagnosis with NPHP-related ciliopathy was based on published clinical criteria. The clinical features of many individuals with KIN have been reported (Table 3) (Palmer et al., Diagn. Cytopathol. 35, 179–182 (2007); Spoendlin, et al. Am. J. Kidney Dis. 25, 242–252 (1995); Baba et al., Pathol. Res. Pract. 202, 555–559 (2006); Godin, et al., Verine et al., Ann. Pathol. 30, 240–242 (2010); Moch et al., Pathologe 15, 44–48 (1994)).

Homozygosity mapping. For genome-wide homozygosity mapping (Hildebrandt, F. et al. PLoS Genet. 5, e1000353 (2009)), the Human Mapping 250k StyI array and the Genome-wide Human SNP 6.0 Array from Affymetrix were used. Genomic DNA samples were hybridized and scanned using the manufacturer's standard protocol at the University of Michigan Core Facility. Non-parametric logarithm of odds (LOD) scores were calculated using a modified version of the GENEHUNTER 2.1 program (Kruglyak, et al., Am. J. Hum. Genet. 58, 1347–1363 (1996); Strauch, K. et al. Am. J. Hum. Genet. 66, 1945–1957 (2000)) through stepwise use of a sliding window with sets of 110 SNPs using ALLEGRO35. Genetic regions of homozygosity by descent (homozygosity peaks) were plotted across the genome as candidate regions for recessive disease-causing genes (Fig. 14a) as described (Hildebrandt, F. et al. PLoS Genet. 5, e1000353 (2009); Otto, E.A. et al. Nat. Genet. 42, 840–850 (2010)). Disease allele frequency was set at 0.0001, and European ancestry marker allele frequencies were used.

Whole-exome sequencing. Exome enrichment was conducted following the manufacturer's protocol for NimbleGen SeqCap EZ Exome v2' beads (Roche NimbleGen). The kit interrogates a total of approximately 30,000 genes (~330,000 coding sequence (CCDS) exons). Massively parallel sequencing was performed as described (Bentley, D.R. et al. Nature 456, 53–59 (2008)).

Mutation calling. Following exome sequencing, mutation calling was performed using CLC Genomics Workbench software. The minimum length fraction with which a read had to match the reference sequence was set to 90%. For SNP detection, the minimum quality score of the central base and the minimum average quality score of surrounding bases were kept at

default values (20 and 15, respectively). Quality assessment was performed within a window of 11 bases. Only reads that uniquely aligned to the reference genome were used for variant SNP or deletion/insertion polymorphism (DIP) calling. In individuals with evidence of homozygosity by descent, the threshold for the number of reads (minor allele frequency) was set to >55%.

Filtering of variants from normal reference sequence (VRS). For DIPs and SNPs, we used the following a priori criteria to restrict the high number of VRS (average of 53,272 for DIPs and 315,372 for SNPs). (i) We retained exonic variants (missense, nonsense and indels) and obligatory splice-site variants only. (ii) We included only VRSs that were not listed in the SNP129 database of innocuous polymorphisms. (iii) We evaluated exonic changes only within genomic regions in which homozygosity mapping showed linkage for both affected siblings (retaining on average 38 for DIPs and 169 for SNPs). (iv) Variants were analyzed using the BLAT program at the UCSC human genome Bioinformatics Browser for the presence of paralogous genes, pseudogenes, misalignments at ends of sequence reads and for whether the variant was a known variant in dbSNP132 with an allele frequency of >1% in populations of European ancestry. In families in whom mapping showed homozygosity by descent, we retained only homozygous variants and examined all of them in the sequence alignments within CLC Genomics Workbench software for the presence of mismatches indicating potential false alignments or poor sequence quality. (v) Sanger sequencing was performed to confirm the remaining variants in original DNA samples and to test for intrafamily segregation in a recessive mode. (vi) Finally, remaining variants were ranked by whether mutations truncated the conceptual reading frame (nonsense, frameshift and obligatory splice variants), by analysis of evolutionary conservation of missense variants, by using web-based programs predicting the impact of disease candidate variants on the encoded protein and by whether variants were known disease-causing mutations.

Segregation analysis by Sanger sequencing. Sanger dideoxy- terminator sequencing was applied for confirmation and segregation of potential disease-causing variants in the respective affected subjects, their affected siblings and their parents. In affected subjects in whom only one heterozygous mutation was detected by exome capture and massively parallel sequencing, all exons and flanking intronic sequences of the respective gene(s) were analyzed by Sanger sequencing. PCR was performed using a touchdown protocol described previously (Otto, E.A. et al. Hum. Mutat. 29, 418–426 (2008)). Sequencing was performed using the BigDye Terminator v3.1 Cycle Sequencing Kit on an ABI 3730 XL sequencer (Applied

Biosystems). Sequence traces were analyzed using Sequencher (version 4.8) software (Gene Codes Corporation).

Web-based variant analysis. Predictions of the possible impact of an amino-acid substitution on chemical change, evolutionary conservation and protein function were obtained using the following web-based programs: PolyPhen, PolyPhen-2, SIFT and MutationTaster. GERP calculation was performed at the SeattleSeq website.

Cell culture. Human fibroblasts were grown in 3% oxygen in DMEM supplemented with 15% FBS, 100 U/ml of penicillin, 0.1 mg/ml streptomycin and nonessential amino acids (all from Invitrogen). BJ cells are normal foreskin fibroblasts and were obtained from ATCC. Fibroblasts were immortalized using pWZLhTERT and/or pMSCVNeo HPV16E6E7 plasmids. LCLs were immortalized using Epstein-Barr virus (EBV) and were grown in RPMI supplemented as above, except with 20% FBS.

DNA damage sensitivity assay. Cells were plated in a 6-well plate in triplicate at a density of 5×10^4 cells per well for primary fibroblasts and LCLs or 2.5×10^4 cells per well for transformed fibroblasts. Immediately after plating for LCLs or 24 h later for fibroblasts, MMC or DEB was added at a final concentration of 0–100 nM for MMC or 0.75–1 μg/ml for DEB. After 6–8 d of culture, cell numbers were determined using a Z2 Coulter Counter (Beckman Coulter). Cell number after MMC or DEB treatment was normalized to cell number in the untreated sample to give the percentage of survival.

RNA interference (RNAi). For siRNA experiments, A1170-22 E6E7/hTERT cells were transfected with a pool of three siRNAs using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's instruction, with the final concentration of total siRNA at 25 nM.

Mutagenesis. Mutagenesis of FAN1 was performed on a pDONR223 FAN1 cDNA construct (Smogorzewska, A. et al. Mol. Cell 39, 36–47 (2010)) using a multisite mutagenesis kit (Agilent). Other FAN1 mutants were previously described (Smogorzewska et al., supra).

Antibodies. Antibody to FAN1 (RC394) was raised in a rabbit using GST-FAN1aa1–90 as an antigen and was affinity purified against HIS-FAN1aa1–90. Commercial antibodies were purchased to HA (Covance, MMS-101R), XPF/ERCC4 (Bethyl Laboratory, A301-315A), MUS81 (Sigma, M1445) and FANCD2 (Novus, NB100-182).

Breakage analysis. Cells were exposed to 0.01 or 0.1 μg/ml DEB for 72 h or with 50 nM MMC for 24 h, arrested with colcemid (0.167 μg/ml) for 2 h, harvested, incubated for 10 min at 37 °C in 0.075 M KCl and fixed in freshly prepared methanol:glacial acidic acid (3:1).

Cells were stored at 4 °C and, when needed, were dropped onto wet slides and air dried at 40 °C for 1 h before staining with Karyomax Giemsa (Invitrogen) Gurr Buffer for 3 min. After rinsing with fresh Gurr Buffer and then with distilled water, slides were fully dried at 40 °C for 1 h and were scanned using the Metasystems Metafer application.

Cell cycle analysis. Cells were left untreated, were treated with 100 nM MMC and grown for 48 h, or were treated with 0.1 μg/ml of DEB and grown for 72 h. Collected cells were resuspended in 300 μl of PBS. While vortexing, 700 μl of ice-cold 100% ethanol was added dropwise, and suspensions were stored at −20 °C at least overnight. Thirty minutes before fluorescence-activated cell sorting (FACS), cells were spun down, resuspended in propidium iodine mix (1 ml of PBS, 10 μl of RNase (from a stock solution of 20 mg/ml) and 10 ml of propidium iodine (from a stock solution of 1 mg/ml)) and analyzed using a FACSCalibur instrument (Becton Dickinson). Cell cycle analysis was performed using FlowJo software (Tree Star).

Morpholino-mediated knockdown of fan1 in zebrafish. Morpholino oligonucleotides were obtained from Gene Tools. Morpholinos (fan1D7 at 0.1 mM, standard control morpholino at 0.2 mM and p53 morpholino at 0.2 mM) were injected into zebrafish embryos at the 1–4 cell stages. Embryos were then fixed at 27 h.p.f. with 4% paraformaldehyde in PBS with 1% DMSO overnight, were permeabilized with acetone at −20 °C for 7 min and were stained with antibody against phosphorylated zebrafish γH2AX (1:1,000 dilution; a gift from Amatruda (UT Southwestern), or antibody against cleaved Caspase-3 (1:200 dilution; BD Biosciences). Alex568-conjugated secondary antibody to rabbit IgG was used at a 1:2,000 and a 1:1,000 dilution, respectively, for each primary antibody. The immunofluorescence procedure followed a standard protocol. After washing off secondary antibody, embryos were mounted in Vector-Shield Mounting Media and were imaged with a Leica SP5X confocal system. Z stacks were processed to reduce noise and were overlayed to obtain representative images of immunofluorescence for γH2AX and cleaved caspase. DAPI was used to stain nuclei.

Quantitative RT-PCR. cDNA from 48 human tissues was purchased from OriGene (Tissue SCANTM Normal Tissue qPCR Arrays, HMRT502). Quantitative RT-PCR was performed using the TaqMan Gene Expression Assay kit (Applied Biosystems) according to the manufacturer's instructions. Briefly, 1 μl of cDNA was mixed with 10 μl of 2× TaqMan Universal Master Mix and 1 μl of 20× TaqMan Gene Expression Assay, bringing the total volume to 20 μl with RNase-free water. Target amplification was performed in 96-well plates using the StepOnePlus Real-Time PCR System (Applied Biosystems). TaqMan probes for

FAN1 (Hs00429686_m1), FANCD2 (Hs00276992_m1) and GAPDH (Hs02758991_g1) were purchased from Applied Biosystems. PCR thermal cycling conditions included an initial 10-min hold at 95 °C to activate the AmpliTaq Gold DNA polymerase followed by 40 cycles of denaturation (15 s at 95°) and annealing and primer extension (15 s at 60 °C). More than three RT-PCR analyses were executed for each sample, and the obtained threshold cycle values were averaged. Relative RNA expression levels were calculated via a comparative threshold cycle (CT) method using GAPDH as control, where ΔCT = CT(GAPDH) − CT(FAN1 or FANCD2). Fold change in gene expression, normalized to GAPDH and relative to the control sample (FAN1 expression in the kidney), was calculated as 2−ΔΔCT□

Immunocytochemistry in the FHH rat and in humans with kidney disease. Kidney tissue coupes from FHH rats30,38 (n = 10) embedded in paraffin were deparafinized, treated with PO block for 15 min and incubated at 100 °C in citrate-HCl buffer for 20 min. The coupes were stained with mouse antibody to γH2AX (1:200 dilution) overnight at 4 °C. Samples were incubated with horseradish peroxidase (HRP)-conjugated secondary polyclonal rabbit antibody to mouse (1:100 diltuion; Dako, P0260) for 30 min at room temperature. Finally, samples were incubated with BrightVision Poly HRP-Anti Rabbit IgG (Immunologic, DPVR55HRP) for 1 h at room temperature. The Nova RED substrate kit for Peroxidase (Vector, SK-4800) was used, and samples were counterstained with hematoxylin. Analysis was performed using Aperio ImageScope software. Ten random tubular fields in the cortex (approximately 270 cells per field) were analyzed for positively stained nuclei using an in-house algorithm macro. GraphPad Prism 5.0 was used to calculate the R2 and P values.

Statistical analysis. Student's two-tailed nonpaired t tests and normal distribution two-tailed z tests were carried out using pooled standard error and s.d. values to determine the statistical significance of differences between cohorts.

Bioinformatics. Genetic locations are annotated according to March 2006 Human Genome Browser data.

Results

By homozygosity mapping in family A1170, we defined seven candidate regions of homozygosity by descent. Exome sequencing identified a homozygous nonsense mutation (coding for p.Trp707*) in FAN1 (which encodes the Fanconi anemia–associated nuclease 1 protein) in both affected siblings (Table 3). No additional homozygous truncating mutations were detected in any other genes within the mapped candidate regions in family A1170. DNA samples were obtained from five published families with KIN and five unpublished cases

(Table 3). Clinical phenotypes have been published for families A4385 (Godin, M. et al. Am. J. Kidney Dis. 27, 166 (1996)), A4393 (Verine, et al., Ann. Pathol. 30, 240–242 (2010)), A1170 (Auerbach, A.D. & Wolman, Nature 261, 494–496 (1976)), A4433 (Spoendlin, M. et al. Am. J. Kidney Dis. 25, 242–252 (1995)) and A4333 (Baba, et al. Pathol. Res. Pract. 202, 555–559 (2006)). With Sanger sequencing of all FAN1 exons, we found 12 different mutations of FAN1 in 9 of the 10 families with KIN (Table 3), detecting both mutated alleles in 9 families (Table 3). Eight of the 12 mutations truncated the conceptual reading frame (Table 3). Three missense mutations (encoding p.Gln929Pro, p.Gly937Asp and p.Asp960Asn) altered amino-acid residues that have been conserved throughout evolution and are located in the FAN1 nuclease domain (VRR-NUC) (Smogorzewska, et al., supra). All mutations were absent from 96 healthy controls. Recessive mutations of FAN1 were identified as a major cause of KIN, an NPHP-like fibrotic kidney disease.

FAN1 is considered to be an effector of the Fanconi anemia pathway, a DDR signaling pathway involved in the repair of ICL damage (Knipscheer, P. et al. Science 326, 1698–1701 (2009)). Individuals with Fanconi anemia are characterized by developmental abnormalities, bone marrow failure and predisposition to cancer (Auerbach, A.D. Mutat. Res. 668, 4–10 (2009)). However, no FAN1 mutations have been detected in individuals with Fanconi anemia of unassigned complementation groups. The FAN1 protein is recruited to sites of ICL damage by interacting with a monoubiquitinated FANCI-FANCD2 complex through its UBZ domain (Kratz, K. et al. Cell 142, 77–88 (2010); Liu et al., Science 329, 693–696 (2010); MacKay, C. et al. Cell 142, 65–76 (2010); Smogorzewska, A. et al. Mol. Cell 39, 36–47 (2010)). In vitro, FAN1 has nuclease activity.

FAN1 expression was examined in fibroblasts and lymphoblastoid cell lines (LCLs) from individuals with KIN (Fig. 15a). No FAN1 protein was detected in the three individuals (A1170-22, A4385-22 and A4466-21) who had two truncating mutations in FAN1 (Fig. 15a). Conversely, the protein was detected in the cell line from individual A4486-23 with a missense mutation (encoding p.Asp960Asn) in the nuclease domain of FAN1 (Fig. 15a).

As depletion of FAN1 sensitizes human cell lines to ICL-inducing agents (Mihatsch et al., supra; Kratz et al., supra; Liu et al., supra; MacKay et al., supra; Smogorzewska et al., supra), FAN1-mutant cells from individuals with KIN were examined for genome instability upon exposure to mitomycin C (MMC) (Fig. 15b). Chromatid breaks and radial chromosomes were observed on metaphase spreads (Fig. 15b), which is consistent with a role for FAN1 in genome maintenance and DDR. The levels of genome instability observed in KIN cell lines were not as high as in Fanconi anemia cell lines that lack FANCA gene

function (RA3087 and RA3157), but were above background levels seen in wild-type cells. The results of the classic test for Fanconi anemia, diepoxybutane (DEB) breakage (Auerbach, A.D. & Wolman, S.R. Nature 261, 494–496 (1976)), were negative in all FAN1-mutant cell lines tested but positive in the control FANCA-mutant cell lines (RA3087 and RA3157).

Despite the differences in chromosomal instability following MMC and DEB exposure, survival of FAN1-mutant cell lines from affected individuals was severely compromised after exposure to either ICL-inducing agent (Fig. 15c,d). ICL sensitivity was observed in multiple cell lines from affected individuals (Fig. 15c,d). In contrast, cell cycle arrest in late-S/G2 phase, which is characteristic of Fanconi anemia cells, was seen in FAN1-mutant cells only after MMC and not after DEB exposure (Fig. 15e). These observed differences between FANCA-and FAN1-mutant cells can be explained by differential engagement of FAN1 versus other Fanconi anemia pathway–directed nucleases in the repair of different ICL lesions or by different processing of the same kind of lesion by these distinct nucleases. Unlike the Fanconi anemia pathway defect, the FAN1 deficiency clearly resulted in high ICL sensitivity in the survival assays but did not lead to the profound genomic instability seen in Fanconi anemia cells. These differences in the cellular phenotype of FAN1- and FANCA-deficient cell lines may explain the lack of phenotypic similarity between FAN1-deficient individuals, who present with KIN, and individuals with Fanconi anemia, who have bone marrow failure and cancer predisposition.

A different phenotype for KIN is also consistent with the finding that FAN1 is not necessary for activation of the Fanconi anemia pathway, as judged by the presence of normal FANCD2 ubiquitination in FAN1-deficient cells.

To complement the FAN1 defect, fibroblasts of individual A1170-22 were transduced with wild-type FAN1 cDNA or with FAN1 cDNA carrying KIN-associated mutations or mutations known to inhibit nuclease activity (encoding p.Glu975Ala/Lys977Ala) or interaction with FANCD2 (p.Cys44Ala/Cys47Ala)6 (Fig. 16a,b). No FAN1 protein was detected in immunoblotting of fibroblasts expressing the two truncating variants (p.Trp707* and p.Arg679Thrfs*5), indicating that they are unstable, whereas p.Leu925Profs*25 yielded a shortened protein product (Fig. 16b). Whereas wild-type FAN1 did complement the MMC sensitivity defect, none of the cDNAs carrying mutations from individuals with KIN rescued the defect, with the exception of the encoded p.Cys871Arg variant, which partially complemented cell survival upon MMC exposure, indicating that it represents a hypomorphic allele of FAN1 (Fig. 16a). The FAN1 variant p.Cys44Ala/Cys47Ala that abolishes FANCD2-dependent localization of FAN1 to sites of DNA damage (Smogorzewska, A. et al, supra)

was fully capable of rescuing the MMC resistance defect, indicating that FAN1 activity is independent of the Fanconi anemia pathway in cell lines from individuals with KIN. The p.Glu975Ala/Lys977Ala mutant lacking nuclease activity could not complement the FAN1 defect.

It was previously shown in DT40 chicken cells that deletion of FAN1 has an effect on ICL sensitivity additive to the effect seen upon deletion of the Fanconi anemia–associated genes FANCC and FANCJ (Yoshikiyo, K. et al. Proc. Natl. Acad. Sci. USA 107, 21553–21557 (2010)). To test whether inhibition of the Fanconi anemia pathway in the FAN1-mutant cells led to increased ICL sensitivity, transcripts of FANCD2, SLX4 or the SLX4-associated nucleases XPF and MUS81 were depleted (Fig. 16c,d). Depletion gave rise to profound MMC sensitivity that was greater than in cells with FAN1 deficiency alone, indicating that FAN1 can work independently of the Fanconi anemia pathway to repair ICL damage.

The fibrotic and cystic kidney phenotypes observed in NPHP-related ciliopathies are still mostly unknown (Hildebrandt et al., N. Engl. J. Med. 364, 1533–1543 (2011); Simons, M. et al. Nat. Genet. 37, 537–543 (2005); Huangfu, D. et al. Nature 426, 83–87 (2003)). Because most DDR pathways as well as NPHP-related ciliopathy phenotypes are conserved in zebrafish (Otto, E.A. et al. Nat. Genet. 42, 840–850 (2010); Otto, E.A. et al. Nat. Genet. 34, 413–420 (2003); Zhou et al., Am. J. Physiol. Renal Physiol. 299, F55–F62 (2010); Schäfer, T. et al. Hum. Mol. Genet. 17, 3655–3662 (2008); Sayer, J.A. et al. Nat. Genet. 38, 674–681 (2006)), it was evaluated whether loss of FAN1 function through morpholino oligonucleotide knockdown of fan1 in zebrafish embryos would cause both disturbance of DDR signaling and NPHP-like phenotypes. Injection into 1–4 cell stage embryos of a morpholino (fan1D7) that targets the splice-acceptor site of exon 7 caused the characteristic NPHP-like phenotype of shortened body axis (Fig. 17a,b) but also the DDR phenotypes of microcephaly, microphthalmia and massive cell death throughout the embryo (Fig. 17c). Cell death was from apoptosis, as shown by greater amounts of activated Caspase-3 (Fig. 14d). DDR signaling was activated, with increased signal for γH2AX shown in immunofluorescence analysis (Fig. 17e). These finding are similar to the ones described for knockdown of the Fanconi anemia–associated gene fancd2 in zebrafish (Liu, T.X. et al. Dev. Cell 5, 903–914 (2003); Zeng et al., Zebrafish 6, 405–415 (2009)). In addition, the fan1D7 morphants showed phenotypes characteristic of NPHP-related ciliopathies (Otto, E.A. et al. Nat. Genet. 42, 840–850 (2010), including ventral body axis curvature and renal cysts when apoptosis was suppressed through knockdown of p53 function (Fig. 17f–h). Specifically,

zebrafish embryos injected with fan1D7 and p53 morpholinos at 72 h post-fertilization (h.p.f.) showed pronephric kidney cysts (19 ± 3%) (Fig. 17f,h) and body curvature (45 ± 4%), whereas p53 morpholino alone did not cause pronephric cysts (Fig. 17g,h) and caused body axis curvature in a significantly smaller fraction of embryos (11 ± 2%) (Fig. 17h). Thus, loss of fan1 function results in ciliopathy-related phenotypes, which are further revealed when p53 function is inhibited. Masking of the renal cystic phenotype in the presence of p53 function is most likely due to the fact that fan1 morphants had embryonic deformity (Fig. 17b,c) and highly elevated levels of apoptosis (Fig. 18d), which prevented the observation of kidney cysts that developed in later embryonic stages. The data show that loss of fan1 function in zebrafish embryos leads to a dual set of phenotypes: one representing activation of DDR and apoptosis (microcephaly and microphthalmia) and the other mimicking NPHP-like ciliopathy phenotypes that are revealed when p53-dependent apoptosis is suppressed. It was found that whereas knockdown of fancd2 in zebrafish led to increased γH2AX staining, it did not cause kidney cysts or other ciliopathy phenotypes seen following fan1 knockdown.

The clinical phenotypes caused by FAN1 and FANCD2 mutations differ substantially. The former causes KIN and histological karyomegaly in the liver and brain, whereas the latter causes Fanconi anemia. To evaluate whether these phenotypic differences can be partially attributed to differential tissue expression, we performed quantitative PCR of 48 different human tissue sources and observed significant differences in expression levels in 25 sources. Notably, FAN1 expression exceeded that of FANCD2 in parenchymatous organs, including the kidney, liver, neuronal tissue and female reproductive organs (Fig. 18a), whereas FANCD2 expression levels far exceeded those of FAN1 in six different lymphatic or bone marrow–derived sources, as well as in skin and testis (Fig. 18b). FAN1 mutations cause KIN and karyomegaly in parenchymatous organs in the absence of Fanconi anemia–like blood dyscrasias, whereas FANCD2 mutation causes Fanconi anemia with pancytopenia, skin involvement and male infertility. FAN1 expression by protein blotting was particularly high in the kidney, indicating that the kidney may depend on FAN1 for its normal function.

It was further explored whether a defect in DDR signaling represents a broader pathogenic mechanism that applies also to other forms of CKD. Renal tissue sections from a standard congenic rat model were evaluated for progressive chronic renal failure, the fawn-hooded hypertensive (FHH) rat, which has well-defined physiological parameters. Ten animals were selected on the basis of disease progression at 9–10 months of age, as measured by proteinuria (Koeners et al., Am. J. Physiol. Regul. Integr. Comp. Physiol. 294, R1847–R1855 (2008)).

To determine the load of DDR in the kidneys of these animals, the criterion of increased nuclear γH2AX staining in immunohistochemistry was applied (Fig. 18c–e). After staining, whole slides (two to four kidney sections per animal) were digitally scanned, ten randomized cortical fields were analyzed per section for positively stained nuclei using a modified algorithm, and results were calibrated for diagnostic pathology. In total, 2,000–3,000 cells were scored per animal by a researcher blinded to the disease status of the animals. A positive correlation ($R2 = 0.64$, $P < 0.0054$) was found between disease progression and DDR in this assay (Fig. 18c). Having established a quantitative correlation in a well-defined congenic model for kidney disease, genetically heterogeneous affected humans were examined. A small pilot study using renal transplant biopsies from individuals that were clinically well characterized by a nephropathologist and nephrologist supported the correlation found in the rat model. γH2AX immunostaining was examined in a specimen from a transplant kidney 4 months after transplantation (histology report without injury and no proteinuria) (Fig. 18f), a specimen from a transplant kidney 16 years after transplantation (pathology report of chronic damage) (Fig. 18g) and a transplant kidney 10 years after transplantation (pathology report of chronic tissue damage) (Fig. 18h). These data support in human CKD the relationship seen in the FHH rat model between the extent of kidney damage and DNA lesions (γH2AX). They indicate that DDR contributes to renal damage in pediatric NPHP-related ciliopathies and in CKD of adults.

Recessive mutations of FAN1 were identified as the cause of KIN, which leads to CKD with fibrotic degeneration of the kidney. Thus, DDR, which has a known role in cellular senescence (Mallette, F.A. & Ferbeyre, G. Cell Cycle 6, 1831–1836 (2007)), contributes to the premature aging phenotype of renal fibrosis in NPHP-like diseases. A KIN-like phenotype has also been described in humans and animal models that were exposed to ochratoxin A (Godin, M. et al. Karyomegalic interstitial nephritis. Am. J. Kidney Dis. 27, 166 (1996)), busulfan or pyrrolizidine alkaloids (Burry, A.F. J. Pathol. 113, 147–150 (1974)), all of which cause ICL. Thus, FAN1 mutation represents the genetic equivalent of environmental genotoxic causes of KIN by the shared pathogenic mechanism of defective ICL repair. In this context, it is notable that, among the 12 individuals with KIN-associated and FAN1 mutations, CKD presented by a median age at 45 years and that 7 of the 12 individuals with KIN carried 2 truncating mutations or missense mutations affecting the nuclease domain of FAN1, which represent null alleles (Fig. 16a and Table 3). A high percentage of adult individuals treated within chronic dialysis programs suffer from fibrotic nephropathy of unknown cause. Furthermore, recessive mutations in FAN1 confer susceptibility to MMC

and DEB, FAN1 mutations may also sensitize to other ICL-causing environmental genotoxins, such as ochratoxin A, which are abundant (Verine et al., Ann. Pathol. 30, 240–242 (2010)).

Table 3

Family	Individual	Ancestry	Nucleotide alteration[a,b]	Deduced protein change	Exon or intron (state)	Continuous amino-acid sequence conservation	Parental consanguinity
A4385	21 (no DNA) 32	French	**c.1234+2T>A** **c.2936_7delGA**	Splice site p.Arg979Thrfs*5	2 (het) 7 (het)	N/A N/A	No
A4466	21	French	**c.1234+2T>A** c.2245C>T	Splice site p.Arg749*	2 (het) 9 (het)	N/A N/A	No
A4393	21	French	c.1375+1G>A **c.2616delA**	Splice site p.Asp873Thrfs*17	3 (het) 12 (het)	N/A N/A	No
A4605	31	German	c.1606C>T c.2786A>C	p.Arg536* p.Gln929Pro	5 (het) 12 (het)	N/A *D. rerio*[c]	No
A4486	22 23	German	c.1606C>T **c.2878G>A**	p.Arg536* p.Asp960Asn	5 (het) 13 (het)	N/A *S. pombe*[c]	No
A1170	22 26	New Zealand Maori	c.2120G>A	p.Trp707*	8 (hom)	N/A	distant?
A4433	21	Spanish	**c.2616delA**	p.Asp873Thrfs*17	12 (hom)	N/A	No
A4407	24	Swiss	c.2611T>C **c.2878G>A**	p.Cys871Arg p.Asp960Asn	12 (het) 13 (het)	(Never R)[d] *S. pombe*[c]	No
A4333	21	N/D	c.2774_5delTT c.2810G>A	p.Leu925Profs*25 p.Gly937Asp	12 (het) 13 (het)	N/A *C. intestinalis*[c]	No

N/A, not applicable; N/D, no data; het, heterozygous; hom, homozygous.
[a]Four different European founder mutations are marked in bold. [b]*FAN1* cDNA mutations are annotated according to human cDNA reference NM_014967.4, where +1 corresponds to the A of the ATG start translation codon. Exon 1 is non-coding. [c]Amino-acid residue is continually conserved throughout evolution, including in *Danio rerio*, *Saccharomyces pombe* or *Ciona intestinalis*, as indicated. [d]Amino-acid residue is not conserved in evolution but is never substituted by an 'R'.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the medical sciences are intended to be within the scope of the following claims.

CLAIMS

1. A kit for detecting gene variants associated with nephronophthisis-related ciliopathies (NPHP-RC) in a subject, consisting essentially of:

a) a first NPHP-RC informative detection reagent for identification of one or more variants in a first gene selected from the group consisting of meiotic recombination 11 homolog (MRE11), zinc finger protein 423 (ZNF423), and centrosomal protein 164kDa (CEP164); and

b) a second NPHP-RC informative detection reagent for identification of one or more variants in a second gene selected from the group consisting of MRE11, ZNF423, CEP164, wherein said second gene is different than said first gene.

2. The kit of claim 1, wherein said variants result in a loss of function of said one or more genes.

3. The kit of claim 2, wherein said MRE11 mutation is c.1897C>T.

4. The kit of claim 3, wherein said mutation encodes a p.R633X protein mutation.

5. The kit of claim 2, wherein said ZNF423 mutation is selected from the group consisting of c.2738C>T, c1518delC, and c.3829C>T.

6. The kit of claim 5, wherein said mutation encodes a protein mutation selected from the group consisting of p.P913L, p.P506fzX43, and p.H1277Y.

7. The kit of claim 2, wherein said CEP164 mutations are selected from the group consisting of c.32A>C, c.277C>T, c.1573C>T, c.1726C>T, and c.4383A>G.

8. The kit of claim 7, wherein said mutation encodes a protein mutation selected from the group consisting of p.Q11P, p.R93W, p.Q525X, p.R576X, and p.X1460W.

9. The kit of claim 1, wherein said first gene is MRE11 and said second gene is ZNF423.

10. The kit of claim 1, wherein said first gene is MRE11 and said second gene is CEP164.

11. The kit of claim 1, wherein said first gene is CEP164 and said second gene is ZNF423.

12. The kit of claim 1, further comprising a third reagent that identifies variants in a third distinct gene.

13. The kit of any one of claims 1 to 12, wherein said first and second reagents are nucleotide probes that specifically bind to said variants.

14. The kit of any one of claims 1 to 12, wherein said first and second reagents are antibodies that specifically bind to polypeptides encoded by said variants.

15. The kit of any one of claims 1 to 12, wherein said first reagent is a pair of primers for amplifying said first gene and second reagent is a pair of primer for amplifying said second gene.

16. The kit of any one of claims 1 to 12, wherein said first and second reagents are sequence primers for sequencing said first and second genes.

17. A method for detecting gene variants associated with NPHP-RC in a subject, comprising:

a) contacting a sample from a subject with a variant detection assay comprising the reagents of any one of claims 1 to 16; and

b) diagnosing said subject with NPHP-RC when said variants are present in said sample.

18. A method for detecting gene variants associated with NPHP-RC in a subject, comprising:

a) contacting a sample from a subject with a variant detection assay, wherein said variant detection assay comprises i) a first NPHP-RC informative detection reagent for identification of one or more variants in a first gene selected from the group consisting of MRE11, ZNF423, and CEP164; and ii) a second NPHP-RC informative detection reagent for identification of one or more variants in a second gene selected from the group consisting of

MRE11, ZNF423, and CEP164, wherein said second gene is different than said first gene, under conditions that the presence of a variant associated with NPHP-RC is determined; and

b) diagnosing said subject with NPHP-RC when said variants are present in said sample.

19. The method of claim 18, wherein said variations result in loss of function mutations in said one or more genes.

20. The method of claim 19, wherein said MRE11 mutation is c.1897C>T.

21. The method of claim 20, wherein said mutation encodes a p.R633X protein mutation.

22. The method of claim 19, wherein said ZNF423 mutation is selected from the group consisting of c.2738C>T, c1518delC, and c.3829C>T.

23. The method of claim 22, wherein said mutation encodes a protein mutation selected from the group consisting of p.P913L, p.P506fzX43, and p.H1277Y.

24. The method of claim 19, wherein said CEP164 mutations are selected from the group consisting of c.32A>C, c.277C>T, c.1573C>T, c.1726C>T, and c.4383A>G.

25. The method of claim 24, wherein said mutation encodes a protein mutation selected from the group consisting of p.Q11P, p.R93W, p.Q525X, p.R576X, and p.X1460W.

26. The method of claim 18, wherein said determining comprises detecting variant nucleic acids or polypeptides.

27. The method of claim 26, wherein said detecting variant nucleic acids comprises one or more nucleic acid detection method selected from the group consisting of sequencing, amplification and hybridization.

28. The method of claim 18, wherein said first gene is MRE11 and said second gene is ZNF423.

29. The method of claim 18, wherein said first gene is MRE11 and said second gene is CEP164.

30. The method of claim 18, wherein said first gene is CEP164 and said second gene is ZNF423.

31. The method of claim 18, further comprising a third reagent that identifies variants in a third distinct gene.

32. The method of claim 18, wherein said first and second reagents are nucleotide probes that specifically bind to said variants.

33. The method of claim 18, wherein said first and second reagents are antibodies that specifically bind to polypeptides encoded by said variants.

34. The method of claim 18, wherein said first reagent is a pair of primers for amplifying said first gene and second reagent is a pair of primer for amplifying said second gene.

35. The method of claim 18, wherein said first and second reagents are sequence primers for sequencing said first and second genes.

36. The method of Claim 18, wherein said biological sample is selected from the group consisting of a tissue sample, a cell sample, and a blood sample.

37. The method of claim 18, wherein said determining comprises a computer implemented method.

38. The method of claim 37, wherein said computer implemented method comprises analyzing variant information and displaying said information to a user.

39. The method of claim 18, further comprising the step of treating said subject for NPHP-RC.

40. The method of claim 18, further comprising the step of treating said subject for NPHP-RC under condition such that at least one symptom of said NPHP-RC is diminished or eliminated.

41. Use of first and second HPHP-RC informative reagents that detect two variant MRE11, ZNF423, or CEP164 nucleic acid or polypeptide for identifying NPHP-RC in a subject.

42. The use of claim 41, wherein said MRE11, ZNF423, or CEP164 variant encodes a loss of function mutation.

43. The use of claim 41, wherein said MRE11 mutation is c.1897C>T, said ZNF423 mutation is selected from the group consisting of c.2738C>T, c1518delC, and c.3829C>T, said CEP164 mutations are selected from the group consisting of c.32A>C, c.277C>T, c.1573C>T, c.1726C>T, and c.4383A>G, and said CEP164 mutations are selected from the group consisting of c.32A>C, c.277C>T, c.1573C>T, c.1726C>T, and c.4383A>G.

44. A kit for detecting gene variants associated with congenital abnormalities of the kidney and urinary tract (CAKUT) in a subject, consisting essentially of:

a) a first CAKUT informative detection reagent for identification of one or more variants in a first gene selected from the group consisting of fraser syndrome 1 (FRAS1), FRAS1 related extracellular matrix protein 2 (FREM2), Ret proto-oncogene, (RET), bone morphogenetic protein 4 (BMP4); and

b) a second CAKUT informative detection reagent for identification of one or more variants in a second gene selected from the group consisting of FRAS1, FREM1, RET, and BMP4, wherein said second gene is different than said first gene.

45. The kit of claim 44, wherein said variants result in a loss of function of said one or more genes.

46. The kit of claim 45, wherein said FRAS1 mutations are selected from the group consisting of c.776T>G, c.299G>T, c.981G>A, c.7861C>T, and c.11268C>A.

47. The kit of claim 46, wherein said mutation encodes a protein mutation selected from the group consisting of p.L259R, p.D998Y, p.R3273H, p.R2621X, and p.H3757Q.

48. The kit of claim 45, wherein said FREM2 mutation is c.7013C>T.

49. The kit of claim 48, wherein said mutation encodes a p.T2338I protein mutation.

50. The kit of claim 45, wherein said RET mutation is c.2110G>T.

51. The kit of claim 50, wherein said mutation encodes a p.V704F protein mutation.

52. The kit of claim 45, wherein said BMP4 mutation is c.362A>G.

53. The kit of claim 52, wherein said mutation encodes a p.H121R protein mutation.

54. The kit of claim 45, wherein said first gene is FRAS1 and said second gene is FREM1.

55. The kit of claim 45, wherein said first gene is FRAS1 and said second gene is RET.

56. The kit of claim 45, wherein said first gene is FRAS1 and said second gene is BMP4.

57. The kit of claim 45, wherein said first gene is FREM1 and said second gene is RET.

58. The kit of claim 45, wherein said first gene is FREM1 and said second gene is BMP4.

54. The kit of claim 45, wherein said first gene is RET and said second gene is BMP4.

55. The kit of claim 45, further comprising a third reagent that identifies variants in a third distinct gene and optionally a fourth reagent that identifies variants in a fourth distinct gene.

56. The kit of any one of claims 45 to 55, wherein said first and second reagents are nucleotide probes that specifically bind to said variants.

57. The kit of any one of claims 45 to 55, wherein said first and second reagents are antibodies that specifically bind to polypeptides encoded by said variants.

58. The kit of any one of claims 45 to 55, wherein said first reagent is a pair of primers for amplifying said first gene and second reagent is a pair of primer for amplifying said second gene.

59. The kit of any one of claims 45 to 55, wherein said first and second reagents are sequence primers for sequencing said first and second genes.

60. A method for detecting gene variants CAKUT in a subject, comprising:

a) contacting a sample from a subject with a variant detection assay of any one of claims 45 to 59; and

b) diagnosing said subject with CAKUT when said variants are present in said sample.

61. A method for detecting gene variants CAKUT in a subject, comprising:

a) contacting a sample from a subject with a variant detection assay, wherein said variant detection assay comprises i) a first CAKUT informative detection reagent for identification of one or more variants in a first gene selected from the group consisting of FRAS1, FREM2, RET, BMP4; and ii) a second CAKUT informative detection reagent for identification of one or more variants in a second gene selected from the group consisting of FRAS1, FREM1, RET, and BMP4, wherein said second gene is different than said first gene under conditions that the presence of a variant associated with CAKUT is determined; and

b) diagnosing said subject with CAKUT when said variants are present in said sample.

62. The method of claim 62, wherein said variations result in loss of function mutations in said one or more genes.

63. The method of claim 62, wherein said FRAS1 mutations are selected from the group consisting of c.776T>G, c.299G>T, c.981G>A, c.7861C>T, and c.11268C>A.

64. The method of claim 63, wherein said mutation encodes a protein mutation selected from the group consisting of p.L259R, p.D998Y, p.R3273H, p.R2621X, and p.H3757Q.

65. The method of claim 62, wherein said FREM2 mutation is c.7013C>T.

66. The method of claim 65, wherein said mutation encodes a p.T2338I protein mutation.

67. The method of claim 62, wherein said RET mutation is c.2110G>T.

68. The method of claim 67, wherein said mutation encodes a p.V704F protein mutation.

69. The method of claim 62, wherein said BMP4 mutation is c.362A>G.

70. The method of claim 69, wherein said mutation encodes a p.H121R protein mutation.

71. The method of claim 61, wherein said determining comprises detecting variant nucleic acids or polypeptides.

72. The method of claim 61, wherein said detecting variant nucleic acids comprises one or more nucleic acid detection method selected from the group consisting of sequencing, amplification and hybridization.

73. The method of claim 61, wherein said first gene is FRAS1 and said second gene is FREM1.

74. The method of claim 61, wherein said first gene is FRAS1 and said second gene is RET.

75. The method of claim 61, wherein said first gene is FRAS1 and said second gene is BMP4.

76. The method of claim 61, wherein said first gene is FREM1 and said second gene is RET.

77. The method of claim 61, wherein said first gene is FREM1 and said second gene is BMP4.

78. The method of claim 61, wherein said first gene is RET and said second gene is BMP4.

79. The method of claim 61, further comprising a third reagent that identifies variants in a third distinct gene.

80. The method of claim 61, wherein said first and second reagents are nucleotide probes that specifically bind to said variants.

81. The method of claim 61, wherein said first and second reagents are antibodies that specifically bind to polypeptides encoded by said variants.

82. The method of claim 61, wherein said first reagent is a pair of primers for amplifying said first gene and second reagent is a pair of primer for amplifying said second gene.

83. The method of claim 61, wherein said first and second reagents are sequence primers for sequencing said first and second genes.

84. The method of Claim 61, wherein said biological sample is selected from the group consisting of a tissue sample, a cell sample, and a blood sample.

85. The method of claim 61, wherein said determining comprises a computer implemented method.

86. The method of claim 85, wherein said computer implemented method comprises analyzing variant information and displaying said information to a user.

87. The method of claim 61, further comprising the step of treating said subject for CAKUT.

88. The method of claim 61, further comprising the step of treating said subject for CAKUT under condition such that at least one symptom of said CAKUT is diminished or eliminated.

89. Use of first and second CAKUT informative reagents that detect two variant FRAS1, FREM2, RET, or BMP4 nucleic acid or polypeptide for identifying CAKUT in a subject.

90. The use of claim 36, wherein said FRAS1, FREM2, RET, or BMP4 variant encodes a loss of function mutation.

91. The use of claim 37, wherein FRAS1 mutations are selected from the group consisting of c.776T>G, c.299G>T, c.981G>A, c.7861C>T, and c.11268C>A, said FREM2 mutation is c.7013C>T, said RET mutation is c.2110G>T, and said BMP4 mutation is c.362A>G.

92. A kit for detecting gene variants associated with karyomegalic interstitial nephritis (KIN) in a subject, consisting essentially of:

a) a first KIN informative detection reagent for identification of a first mutation in a KAN gene; and

b) a second KIN informative detection reagent for identification of a second mutation in a KAN1 gene, wherein said second mutation is different than said first mutation.

93. The kit of claim 92, wherein said first and second mutations are selected from the group consisting of c.1234+2T>A, c.2036_7delGA, c.2245C>T, c.1375+1G>A, c.2616delA, c.1606C>T, c.2878G>A, c.2120G>A, c.2786A>C, c.2611T>C, c.2774_5delTT, and c.2810G>A.

94. The kit of claim 93, wherein said first mutation is c.1234+2T>A and said second mutation is c.2036_7delGA.

95. The kit of claim 93, wherein said first mutation is c.1234+2T>A and said second mutation is c.2245C>T.

96. The kit of claim 93, wherein said first mutation is c.1375+1G>A and said second mutation is c.2616delA.

97. The kit of claim 93, wherein said first mutation is c.1606C>T and said second mutation is c.2786A>C.

98. The kit of claim 93, wherein said first mutation is c.1606C>T and said second mutation is c.2878G>A.

99. The kit of claim 93, wherein said first mutation is c.2611T>C and said second mutation is c.2878G>A.

100. The kit of claim 93, wherein said first mutation is c.2774_5delTT and said second mutation is c.2810G>A.

101. The kit of any one of claims 92 to 100, wherein said first and second mutations encode a polypeptide mutation selected from the group consisting of p.Arg679Thrfs*5, p.Arg749*, p.Asp873Thrfs*17, p.Arg536*, p.Cys871Arg, p.Trp707*, p.Leu925Profs*25, p.Gln929Pro, p.Gly937Asp, p.Asp960Asn, and a splice site mutation.

102. The kit of claim 101, wherein said first mutation is a splice site mutation and said second mutation is p.Arg679Thrfs*5.

103. The kit of claim 101, wherein said first mutation is a splice site mutation and said second mutation is p.Arg749*.

104. The kit of claim 101, wherein said first mutation is a splice site mutation and said second mutation is p.Asp873Thrfs*17.

105. The kit of claim 101, wherein said first mutation is p.Arg536* and said second mutation is p.Gln929Pro.

106. The kit of claim 101, wherein said first mutation is p.Arg536* and said second mutation is p.Asp960Asn.

107. The kit of claim 101, wherein said first mutation is p.Cys871Arg and said second mutation is p.Asp960Asn.

108. The kit of claim 101, wherein said first mutation is p.Leu925Profs*25 and said second mutation is p.Gly937Asp.

109. The kit of any one of claims 92 to 108, wherein said variants result in a loss of function of said KAN1 gene.

110. The kit of any one of claims 92 to 100, wherein said first and second reagents are nucleotide probes that specifically bind to said variants.

111. The kit of any one of claims 100 to 108, wherein said first and second reagents are antibodies that specifically bind to said polypeptides.

112. The kit of any one of claims 92 to 100, wherein said first reagent is a pair of primers for amplifying said first gene and second reagent is a pair of primer for amplifying said second gene.

113. The kit of any one of claims 92 to 100, wherein said first and second reagents are sequence primers for sequencing said first and second genes.

114. A method for detecting gene variants associated with KIN in a subject, comprising:

a) contacting a sample from a subject with a variant detection assay, wherein said variant detection assay comprises i) a first KIN informative detection reagent for identification of a first variant in a KAN gene; and optionally ii) a second KIN informative detection reagent for identification of a second variant in a KAN1 gene, wherein said second variant is different than said first variant under conditions that the presence of a variant associated with KIN is determined; and b) diagnosing said subject with KIN when said variant is present in said sample.

115. The method of claim 114, wherein said variations result in loss of function mutations in said one or more genes.

116. The method of claim 114, wherein said first and second mutations are selected from the group consisting of c.1234+2T>A, c.2036_7delGA, c.2245C>T, c.1375+1G>A,

c.2616delA, c.1606C>T, c.2878G>A, c.2120G>A, c.2786A>C, c.2611T>C, c.2774_5delTT, and c.2810G>A.

117. The method of claim 114, wherein said first mutation is c.1234+2T>A and said second mutation is c.2036_7delGA.

118. The method of claim 114, wherein said first mutation is c.1234+2T>A and said second mutation is c.2245C>T.

119. The method of claim 114, wherein said first mutation is c.1375+1G>A and said second mutation is c.2616delA.

120. The method of claim 114, wherein said first mutation is c.1606C>T and said second mutation is c.2786A>C.

121. The method of claim 114, wherein said first mutation is c.1606C>T and said second mutation is c.2878G>A.

122. The method of claim 114, wherein said first mutation is c.2611T>C and said second mutation is c.2878G>A.

123. The method of claim 114, wherein said first mutation is c.2774_5delTT and said second mutation is c.2810G>A.

124. The method of any one of claims 116 to 123, wherein said first and second mutations encode a polypeptide mutation selected from the group consisting of p.Arg679Thrfs*5, p.Arg749*, p.Asp873Thrfs*17, p.Arg536*, p.Cys871Arg, p.Trp707*, p.Leu925Profs*25, p.Gln929Pro, p.Gly937Asp, p.Asp960Asn, and a splice site mutation.

125. The method of claim 124, wherein said first mutation is a splice site mutation and said second mutation is p.Arg679Thrfs*5.

126. The method of claim 124, wherein said first mutation is a splice site mutation and said second mutation is p.Arg749*.

127. The method of claim 124, wherein said first mutation is a splice site mutation and said second mutation is p.Asp873Thrfs*17.

128. The method of claim 124, wherein said first mutation is p.Arg536* and said second mutation is p.Gln929Pro.

129. The method of claim 124, wherein said first mutation is p.Arg536* and said second mutation is p.Asp960Asn.

130. The method of claim 124, wherein said first mutation is p.Cys871Arg and said second mutation is p.Asp960Asn.

131. The method of claim 124, wherein said first mutation is p.Leu925Profs*25 and said second mutation is p.Gly937Asp.

132. The method of claim 112, wherein said detecting variant nucleic acids comprises one or more nucleic acid detection method selected from the group consisting of sequencing, amplification and hybridization.

133. The method of any one of claims 113 to 123, wherein said first and second reagents are nucleotide probes that specifically bind to said variants.

134. The method of any one of claims 114 to 131, wherein said first and second reagents are antibodies that specifically bind to polypeptides encoded by said variants.

135. The method of any one of claims 113 to 123, wherein said first reagent is a pair of primers for amplifying said first gene and second reagent is a pair of primer for amplifying said second gene.

136. The method of any one of claims 113 to 123, wherein said first and second reagents are sequence primers for sequencing said first and second genes.

137. The method of Claim 112, wherein said biological sample is selected from the group consisting of a tissue sample, a cell sample, and a blood sample.

138. The method of claim 112, wherein said determining comprises a computer implemented method.

139. The method of claim 138, wherein said computer implemented method comprises analyzing variant information and displaying said information to a user.

140. The method of claim 112, further comprising the step of treating said subject for KIN.

141. The method of claim 140, further comprising the step of treating said subject for KIN under condition such that at least one symptom of said KIN is diminished or eliminated.

142. Use of at least a first KIN informative reagents that detects at least one variant KAN1 nucleic acids or polypeptides for identifying KIN in a subject.

143. The use of claim 142, wherein said KAN1 variant nucleic acids are selected from the group consisting of c.1234+2T>A, c.2036_7delGA, c.2245C>T, c.1375+1G>A, c.2616delA, c.1606C>T, c.2878G>A, c.2120G>A, c.2786A>C, c.2611T>C, c.2774_5delTT, and c.2810G>A.

144. The use of claim142, wherein said variant KAN1 polypeptides are selected from the group consisting of p.Arg679Thrfs*5, p.Arg749*, p.Asp873Thrfs*17, p.Arg536*, p.Cys871Arg, p.Trp707*, p.Leu925Profs*25, p.Gln929Pro, p.Gly937Asp, p.Asp960Asn, and a splice site mutation.

Figure 1

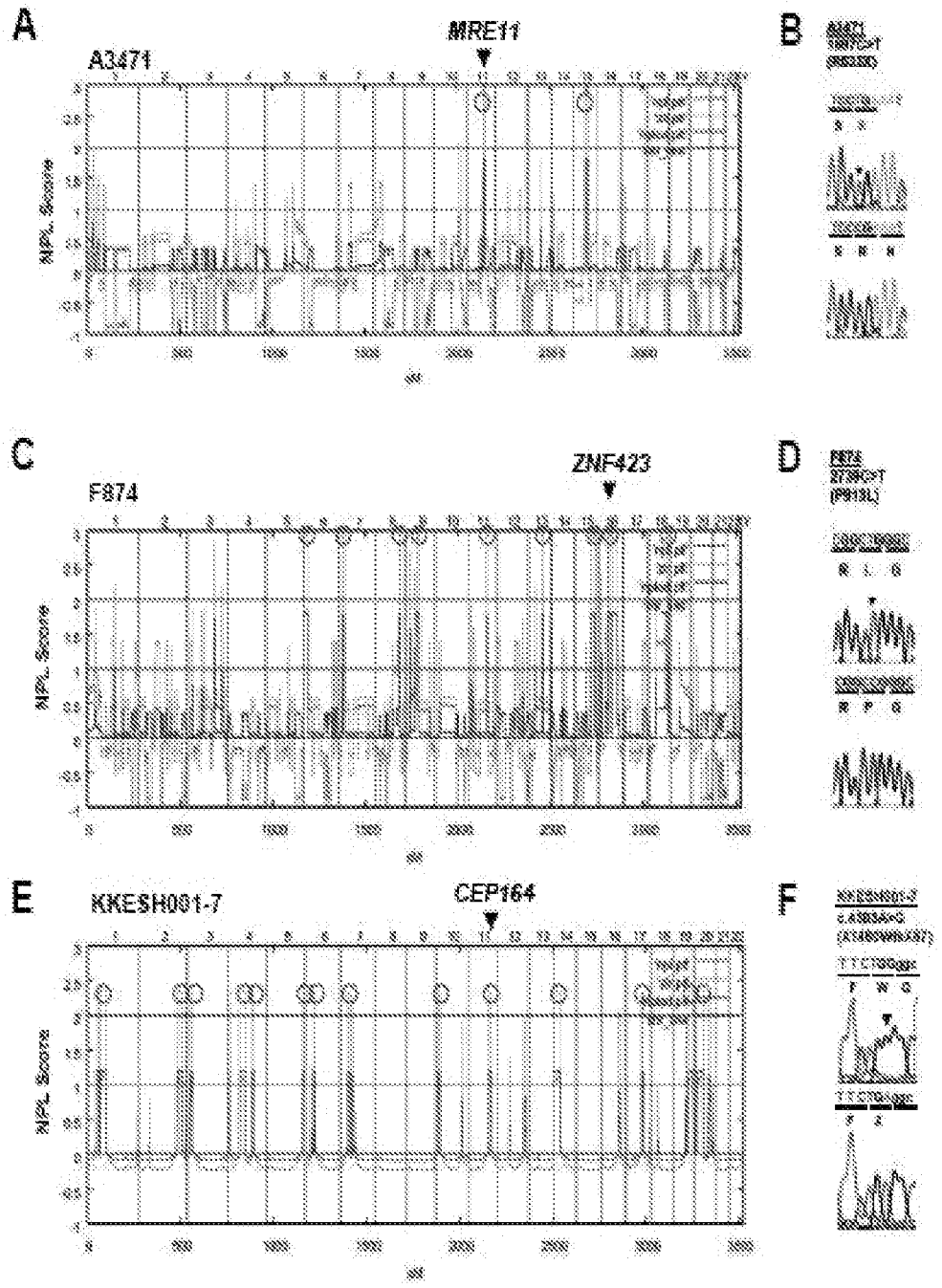

A
MRE11
A3471
NPL Score
B
C
ZNF423
F874
NPL Score
D
E
CEP164
KKESH001-7
NPL Score
F

Figure 2

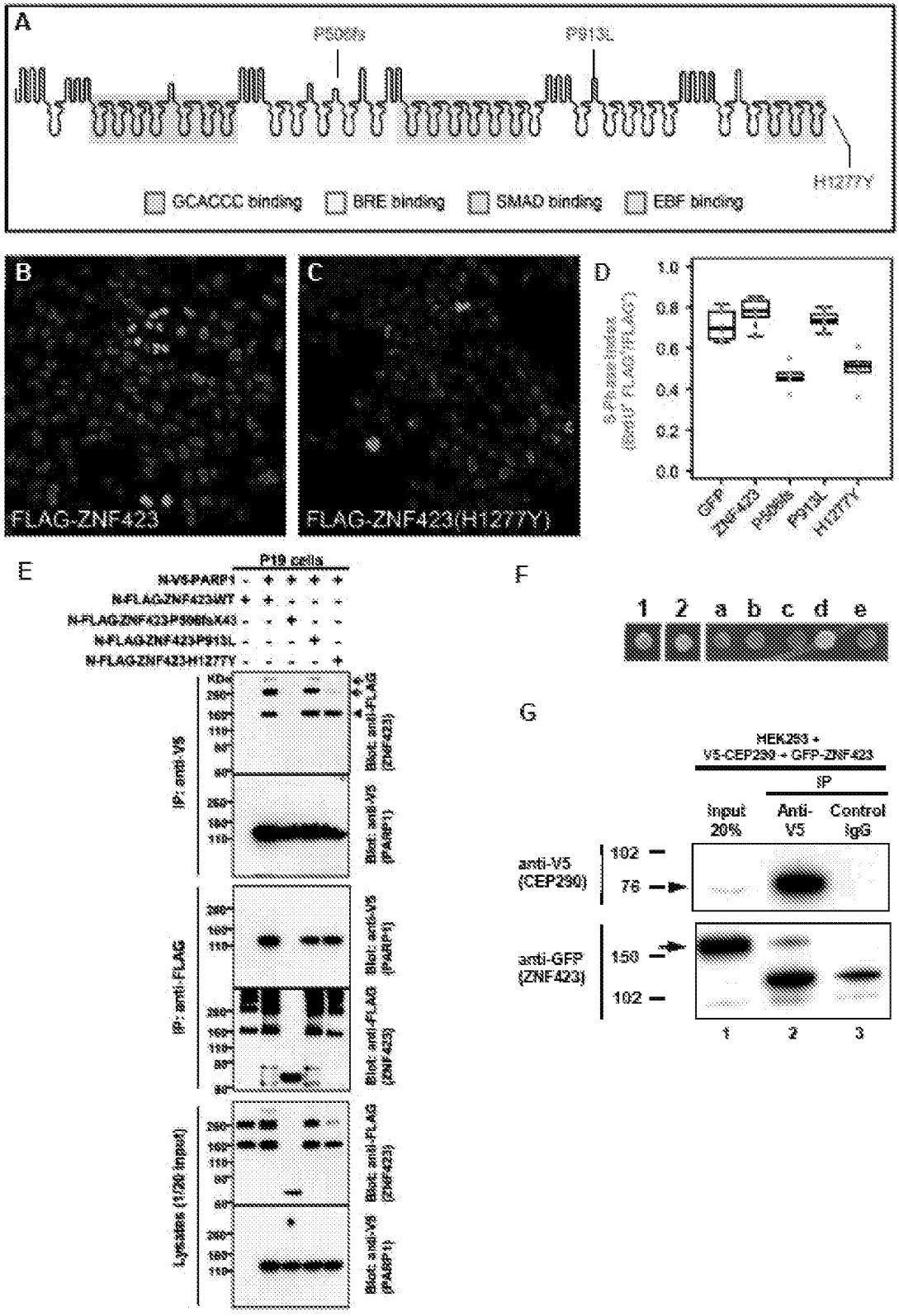

A
P508S
P913L
H1277Y
GCACCC binding
BRE binding
SMAD binding
EBF binding
B
FLAG-ZNF423
C
FLAG-ZNF423(H1277Y)
D
GFP
ZNF423
P508S
P913L
H1277Y
E
P19 cells
N-V5-PARP1
N-FLAG-ZNF423WT
N-FLAG-ZNF423P508delX43
N-FLAG-ZNF423P913L
N-FLAG-ZNF423H1277Y
IP: anti-V5
IP: anti-FLAG
Lysates (1/20 input)
F
1 2 a b c d e
G
HEK293 +
V5-CEP290 + GFP-ZNF423
IP
Input 20%
Anti-V5
Control IgG
anti-V5 (CEP290)
anti-GFP (ZNF423)
1 2 3

Figure 3

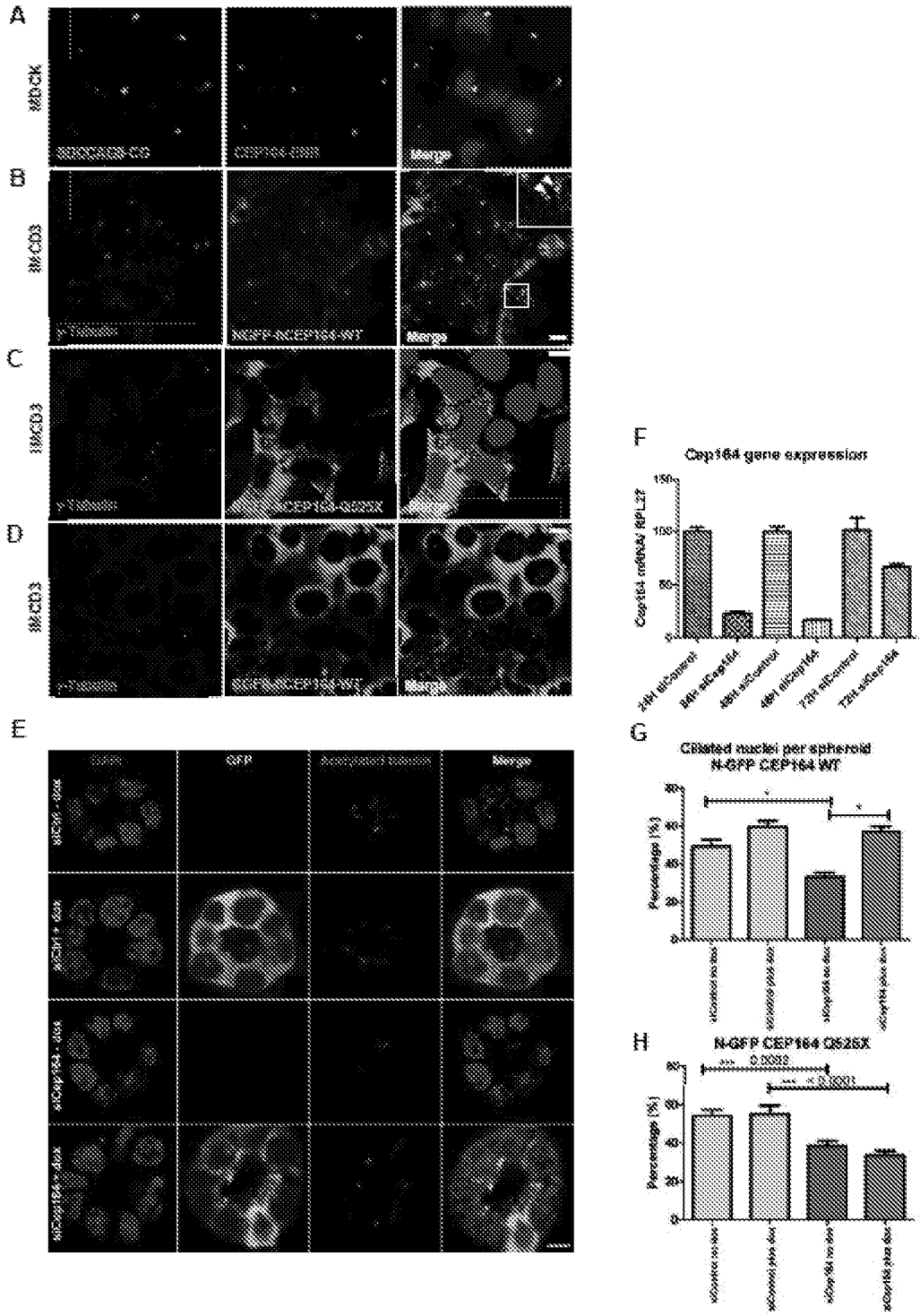

A
B
C
D
E
F
Cep164 gene expression
G
Ciliated nuclei per spheroid
N-GFP CEP164 WT
H
N-GFP CEP164 Q525X

Figure 4

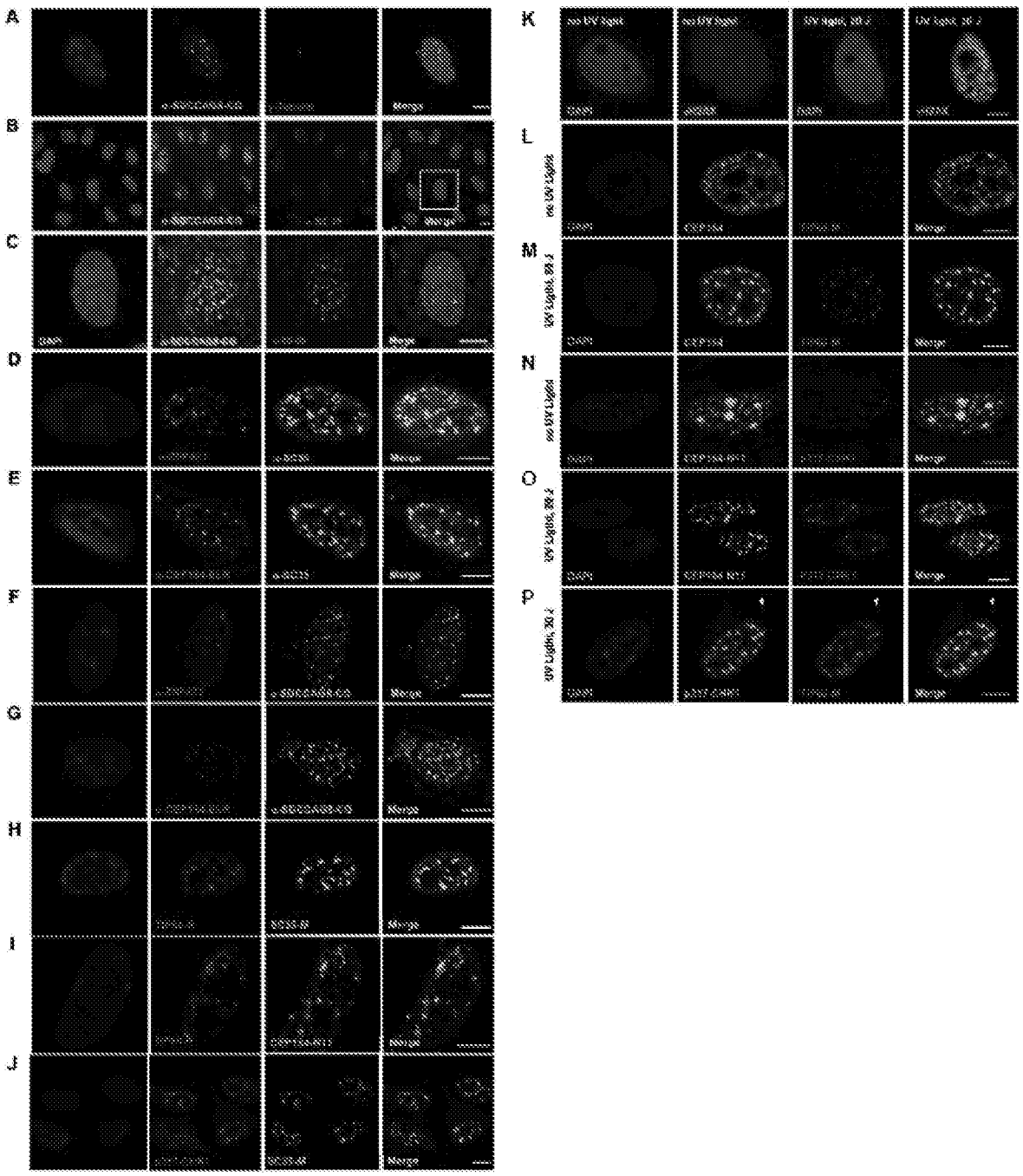


Figure 5

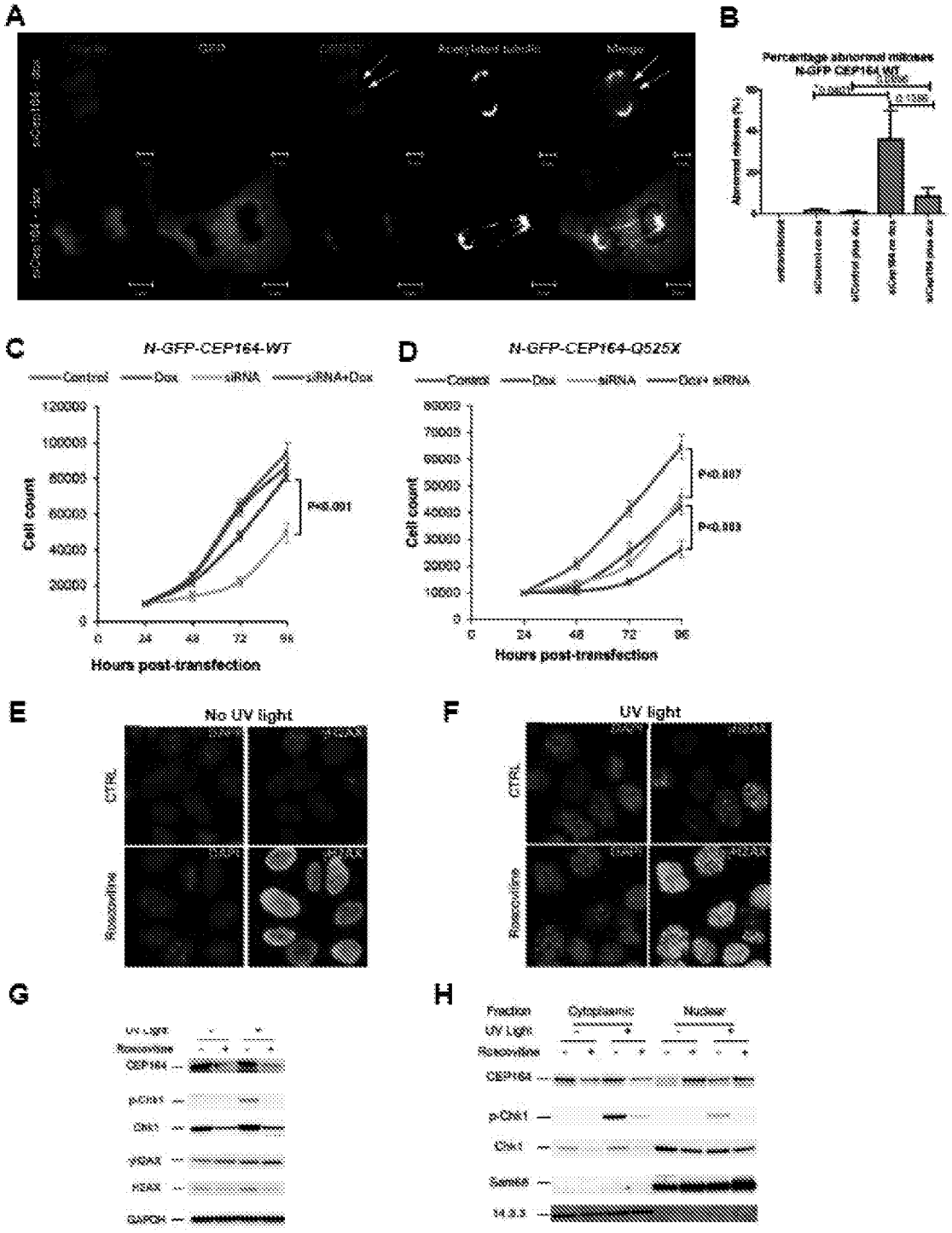

A
B
C
N-GFP-CEP164-WT
Control
Dox
siRNA
siRNA+Dox
Cell count
Hours post-transfection
D
N-GFP-CEP164-Q525X
E
No UV light
CTRL
Roscovitine
F
UV light
G
H
Fraction
Cytoplasmic
Nuclear
CEP164
p-Chk1
Chk1
Sumo3
14-3-3

Figure 6

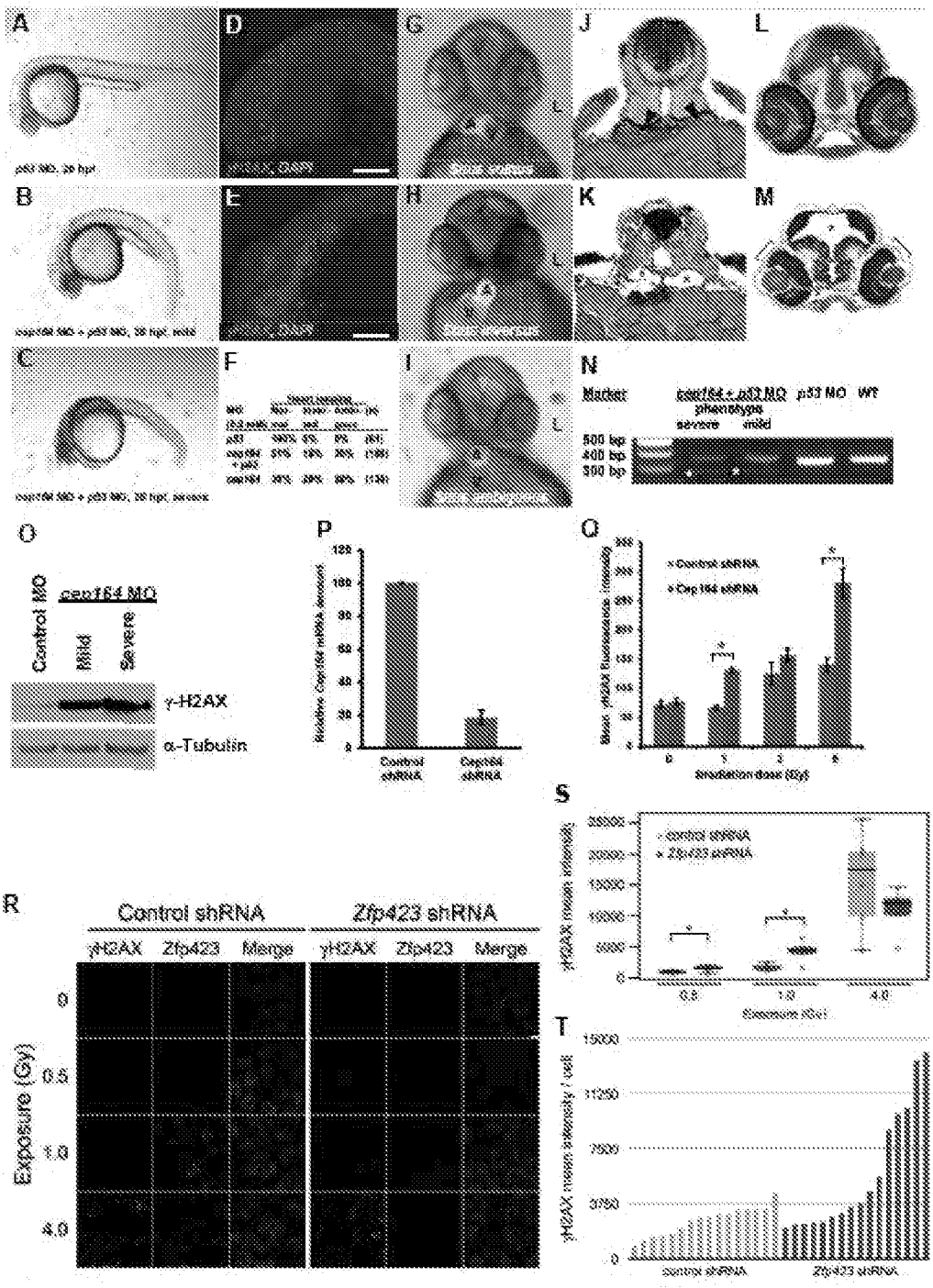

A
D
G
J
L
B
E
H
K
M
C
F
I
N
Marker
cep164 + p53 MO
p53 MO
WT
phenotype
severe
mild
500 bp
400 bp
300 bp
O
Control MO
cep164 MO
Mild
Severe
γ-H2AX
α-Tubulin
P
Relative Cep164 mRNA amount
Control shRNA
Cep164 shRNA
Q
Control shRNA
Cep164 shRNA
Irradiation dose (Gy)
S
Control shRNA
Zfp423 shRNA
γH2AX mean intensity
Exposure (Gy)
R
Control shRNA
Zfp423 shRNA
γH2AX
Zfp423
Merge
Exposure (Gy)
0
0.5
1.0
4.0
T
γH2AX mean intensity / cell
Control shRNA
Zfp423 shRNA

Figure 7

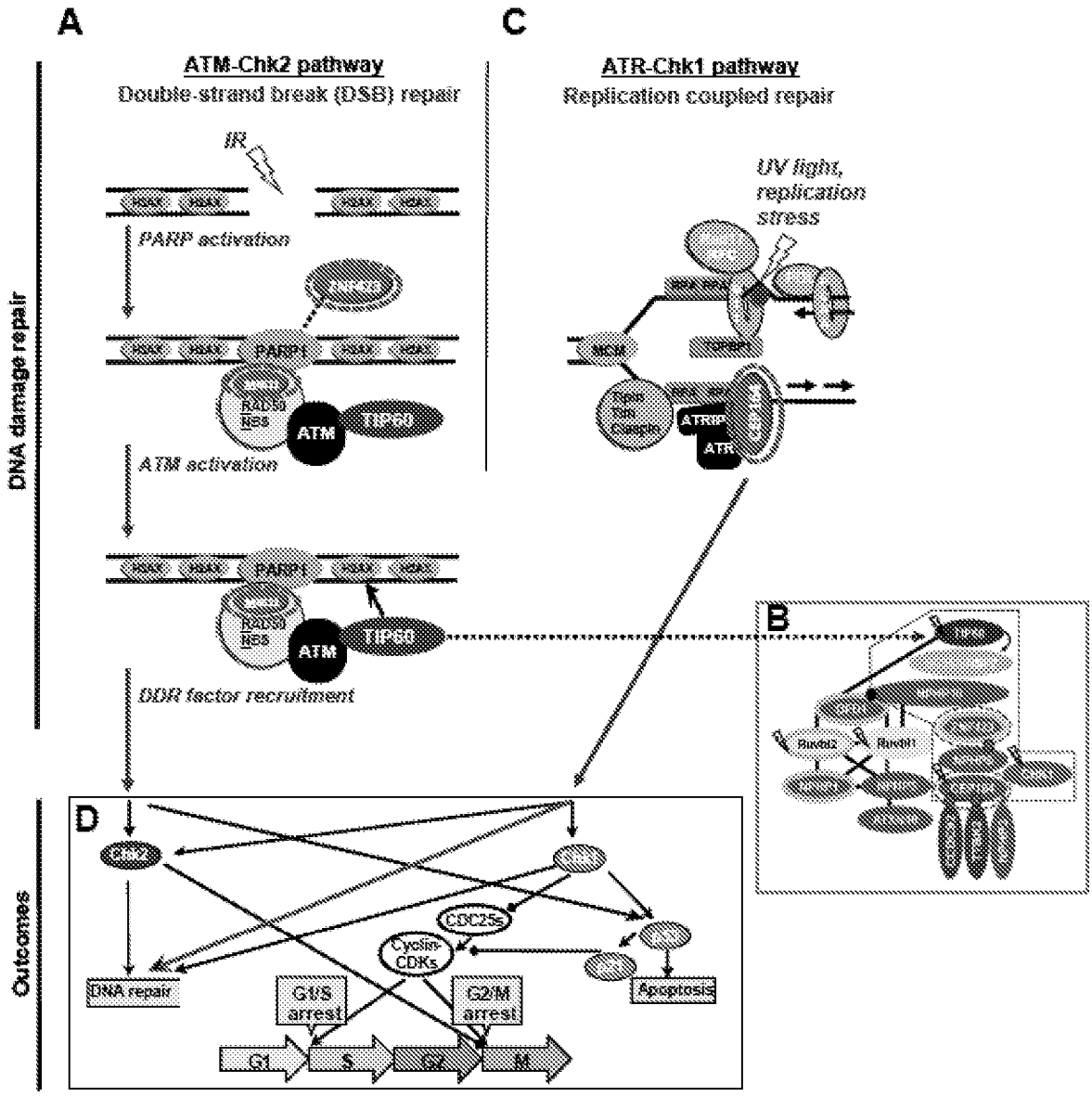

A
ATM-Chk2 pathway
Double-strand break (DSB) repair
IR
PARP activation
PARP1
ATM
TIP60
ATM activation
DDR factor recruitment
C
ATR-Chk1 pathway
Replication coupled repair
UV light, replication stress
MCM
ATRIP
ATR
B
D
CHK2
CHK1
CDC25s
Cyclin-CDKs
DNA repair
G1/S arrest
G2/M arrest
Apoptosis
G1
S
G2
M
DNA damage repair
Outcomes

Figure 8

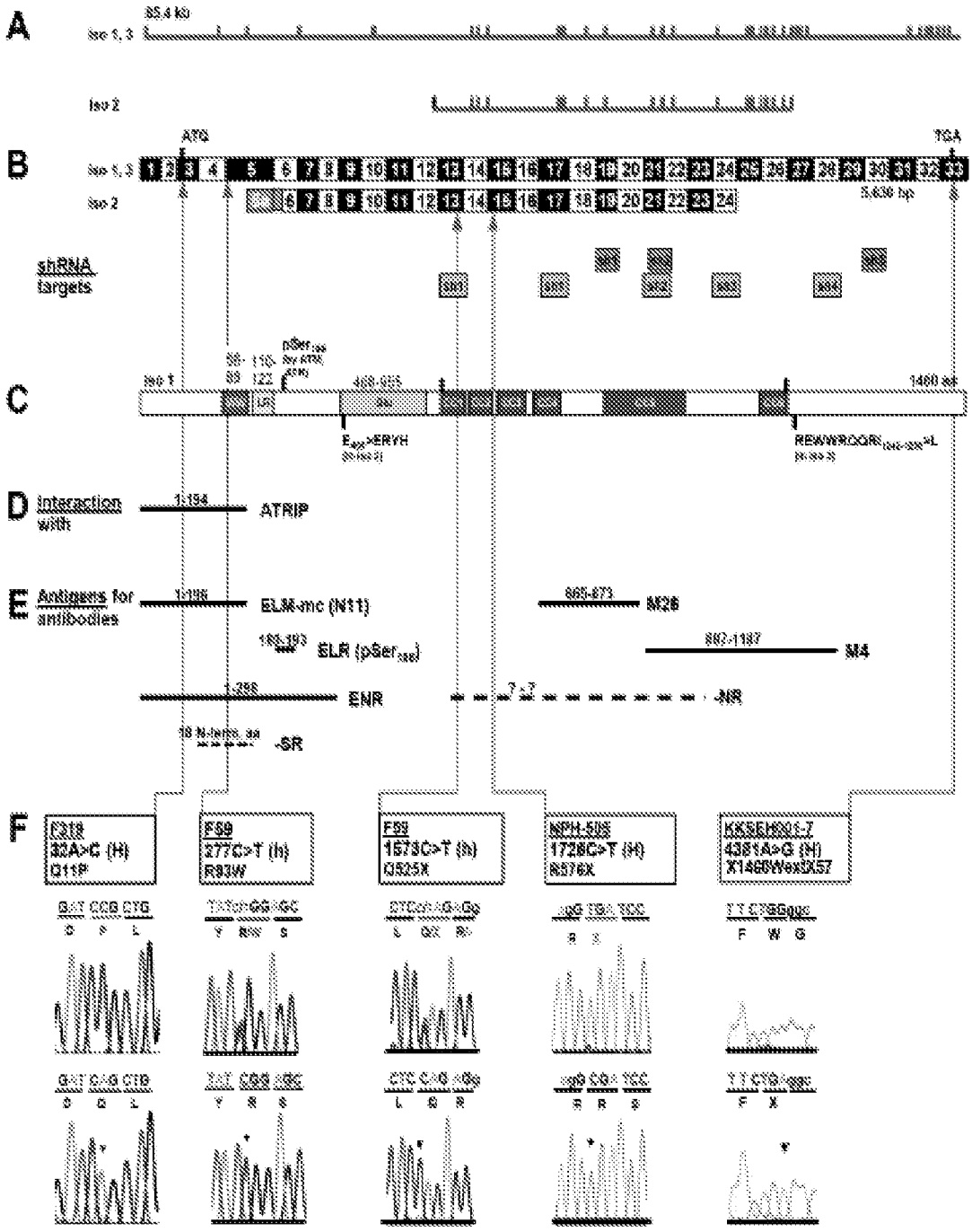

A
B
C
D
E
F
ATG
TGA
shRNA targets
Interaction with
ATRIP
Antigens for antibodies
ELM-mc (N11)
M26
M4
ELR (pSer)
ENR
-NR
-SR

Figure 9

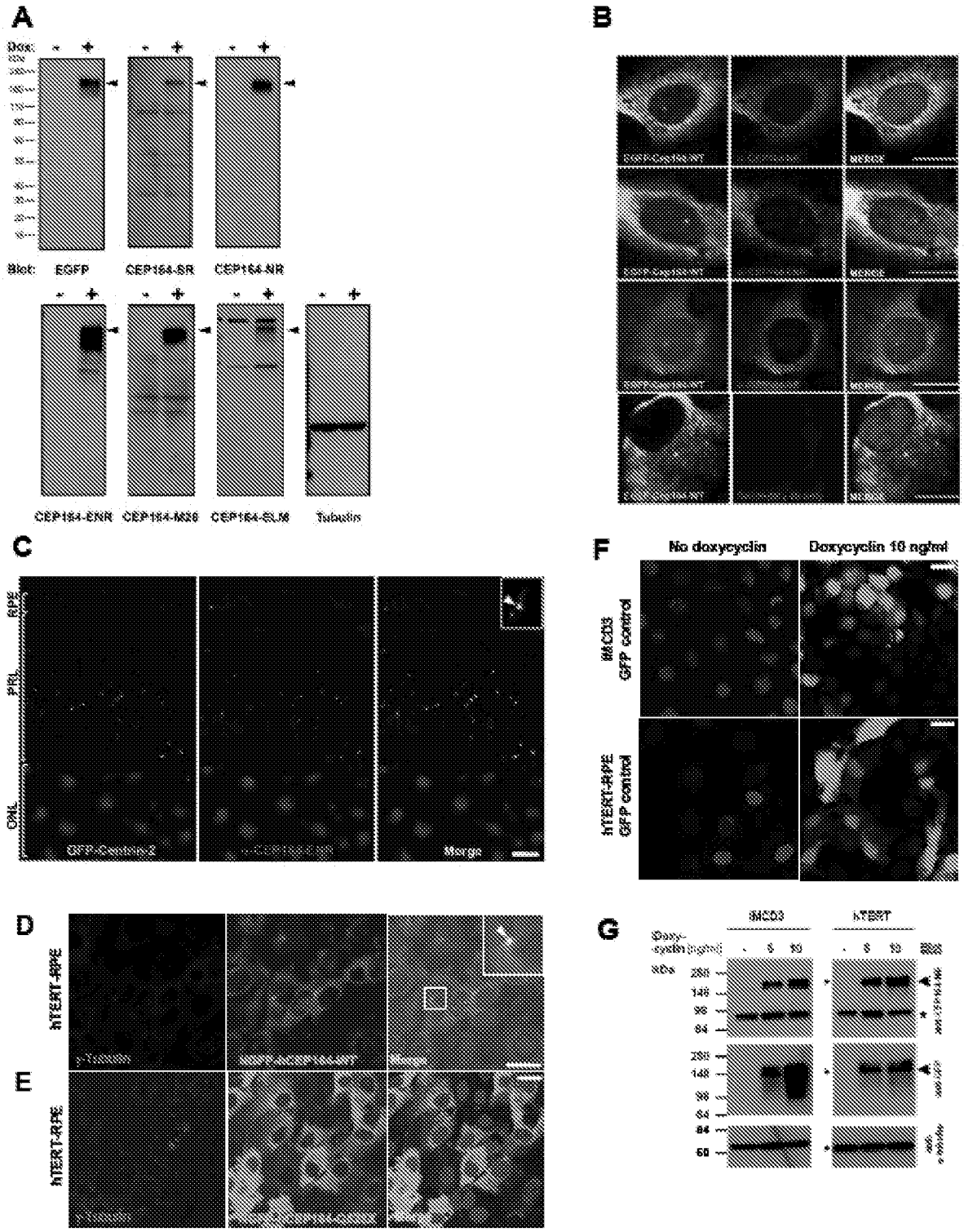

A
Dox:
Blot: EGFP
CEP164-SR
CEP164-NR
CEP164-ENR
CEP164-M26
CEP164-SL/8
Tubulin
B
C
RPE
FRL
ONL
GFP-Centrin-2
Merge
D
hTERT-RPE
E
hTERT-RPE
F
No doxycyclin
Doxycyclin 16 ng/ml
hTERT-RPE GFP control
G
hTERT
Doxy-cyclin (ng/ml)
kDa

Figure 10

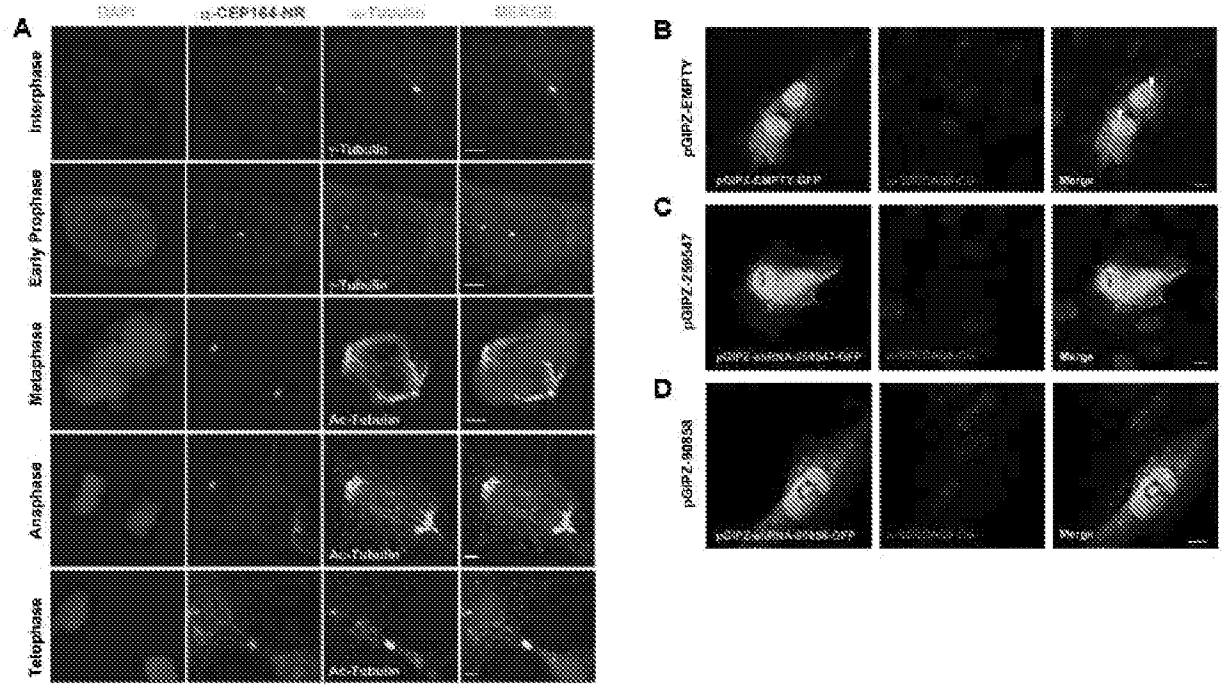
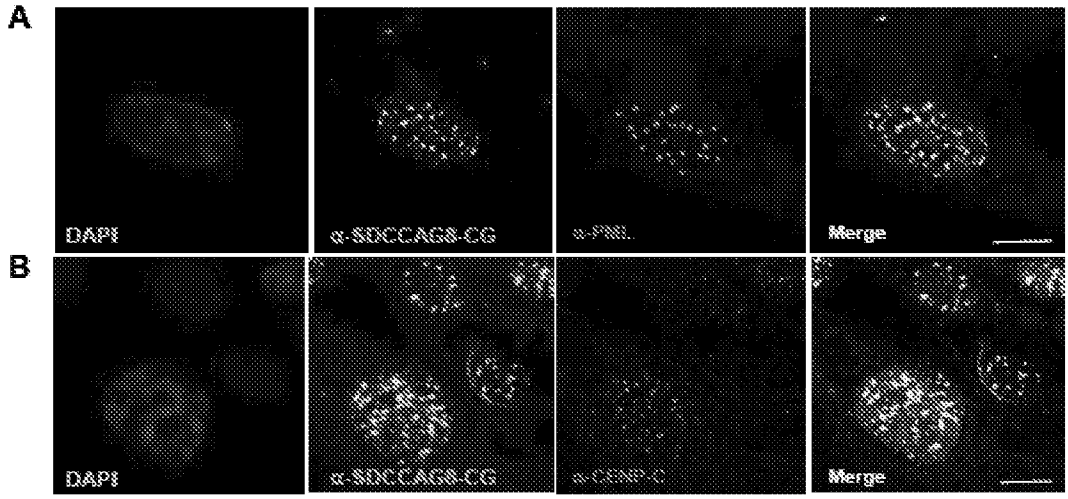

A
Interphase
Early Prophase
Metaphase
Anaphase
Telophase
B
C
D

Figure 11

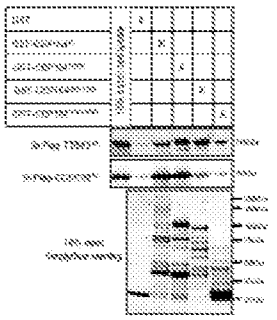


C

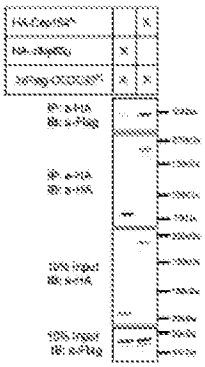
Bait	Prey	# clones	cDNA library
CEP164[1]	hCCDC92[1-903]	2	Human oligo-dT primed
CEP164[1-930]	hCCDC92[1-903]	3	Human oligo-dT primed
CEP164[1-930]	hCCDC92[1-274]	1	Human oligo-dT primed
CEP164[1-930]	hCCDC92[275-574]	1	Human oligo-dT primed
CEP164[1-930]	bTTBK2[1-1243]	1	Human oligo-dT primed
CEP164[1]	bTTBK2[1-1243]	1	Bovine random primed
CEP164[1]	bTTBK2[1-1243]	1	Bovine random primed
CEP164[1-930]	bTTBK2[1-1243]	9	Bovine random primed
CEP164[1-930]	bTTBK2[1-1243]	1	Bovine random primed

h = Human sequence; b = Bovine sequence; * = Reading frame could not be determined; sequence may be longer at the 5' end. The numbers in superscript indicate which region was identified after sequencing.

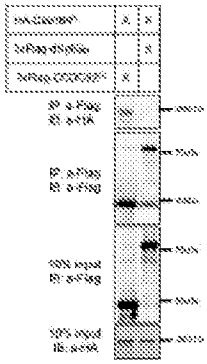
D



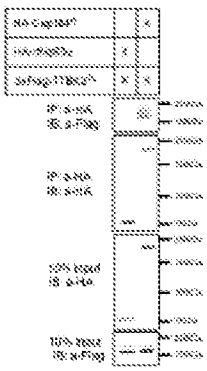
E



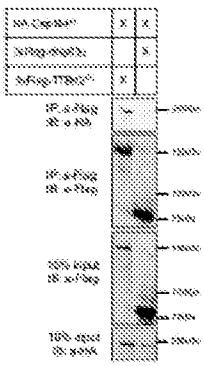
F



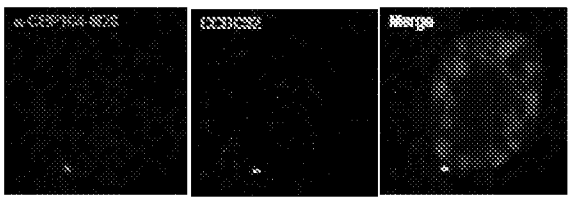
G



H



I



J

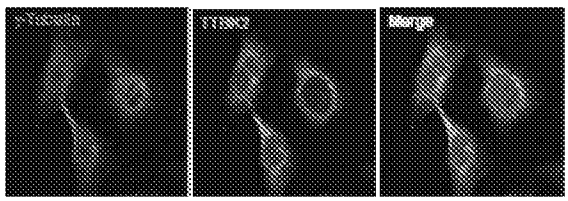


Figure 12

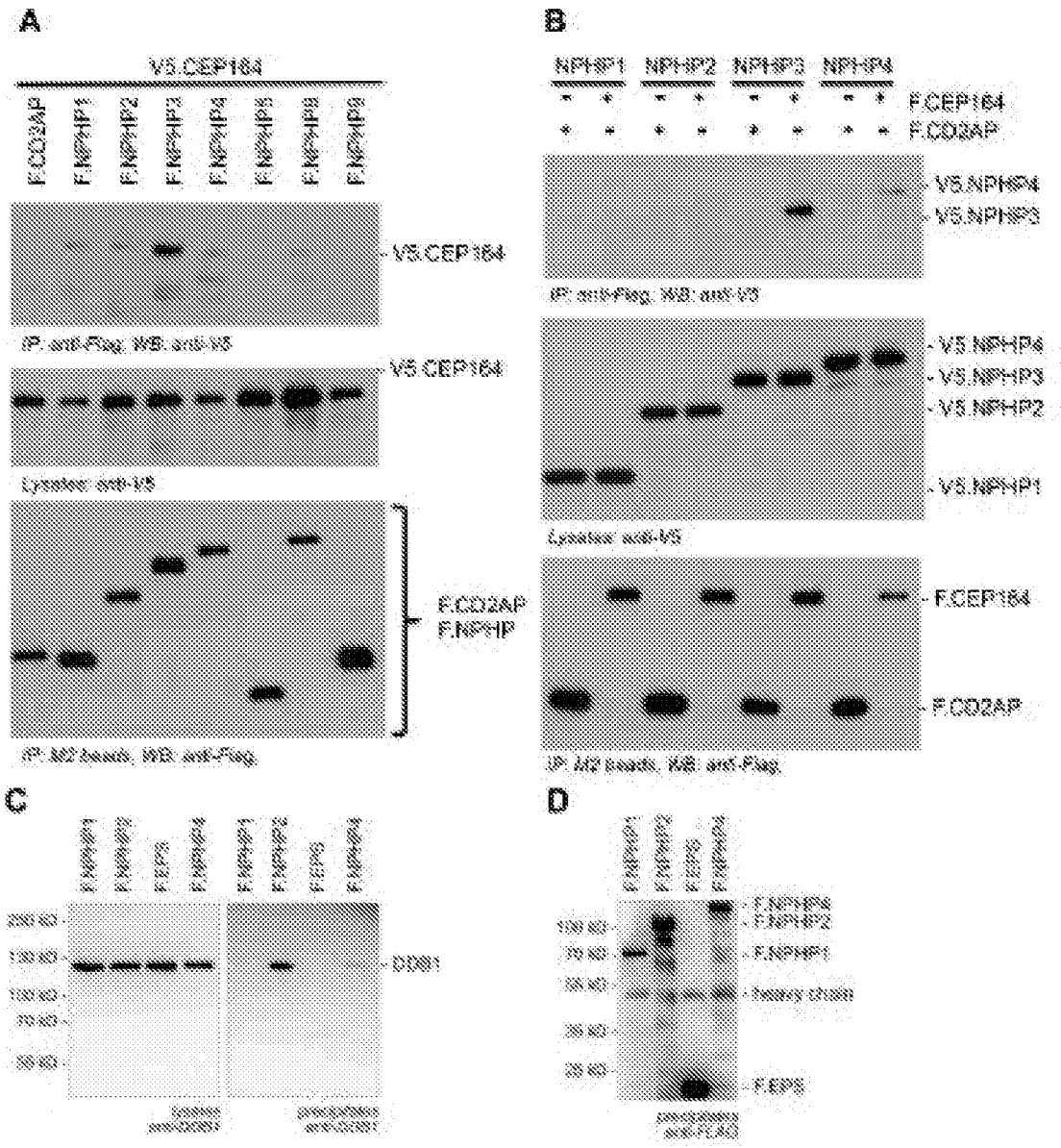

A
V5.CEP164
F.CD2AP
F.NPHP1
F.NPHP2
F.NPHP3
F.NPHP4
F.NPHP5
F.NPHP6
F.NPHP9
V5.CEP164
IP: anti-Flag, WB: anti-V5
V5.CEP164
Lysates: anti-V5
F.CD2AP
F.NPHP
IP: M2 beads, WB: anti-Flag
B
NPHP1
NPHP2
NPHP3
NPHP4
F.CEP164
F.CD2AP
V5.NPHP4
V5.NPHP3
IP: anti-Flag, WB: anti-V5
V5.NPHP4
V5.NPHP3
V5.NPHP2
V5.NPHP1
Lysates: anti-V5
F.CEP164
F.CD2AP
IP: M2 beads, WB: anti-Flag
C
250 kD
130 kD
100 kD
70 kD
55 kD
CD81
Lysates
anti-CD81
anti-CD81
D
100 kD
70 kD
55 kD
35 kD
25 kD
F.NPHP4
F.NPHP2
F.NPHP1
heavy chain
F.EPS
anti-FLAG

Figure 12 Continued

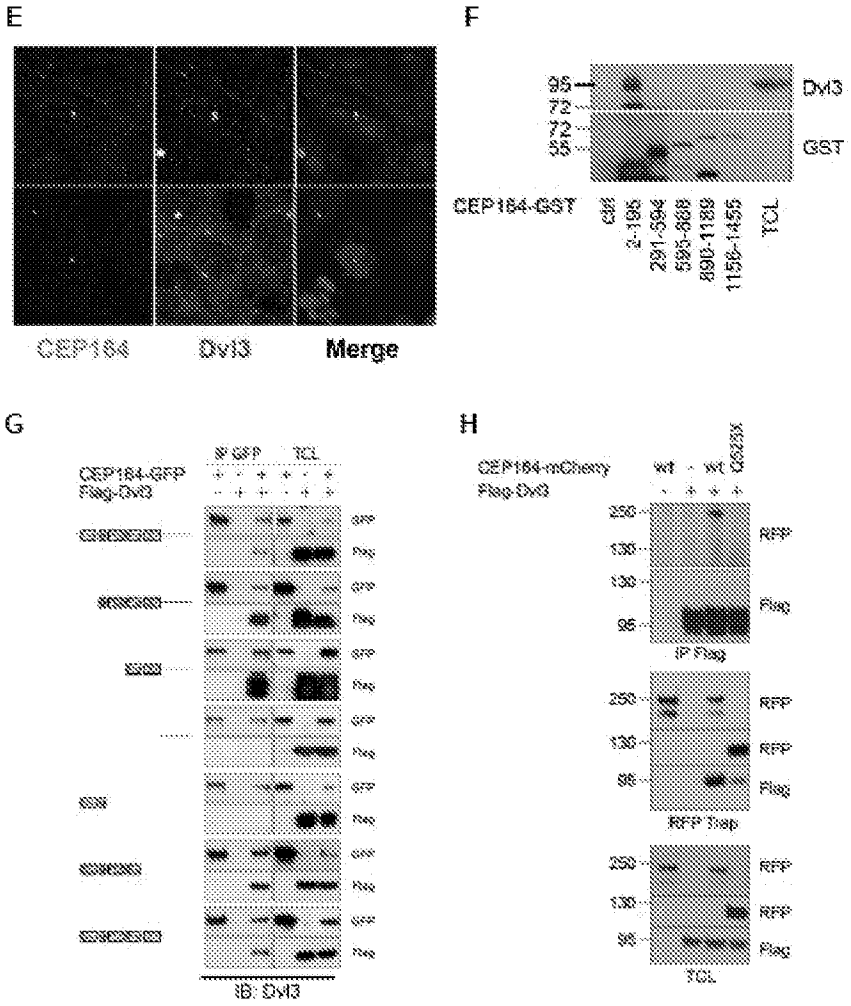


Figure 13

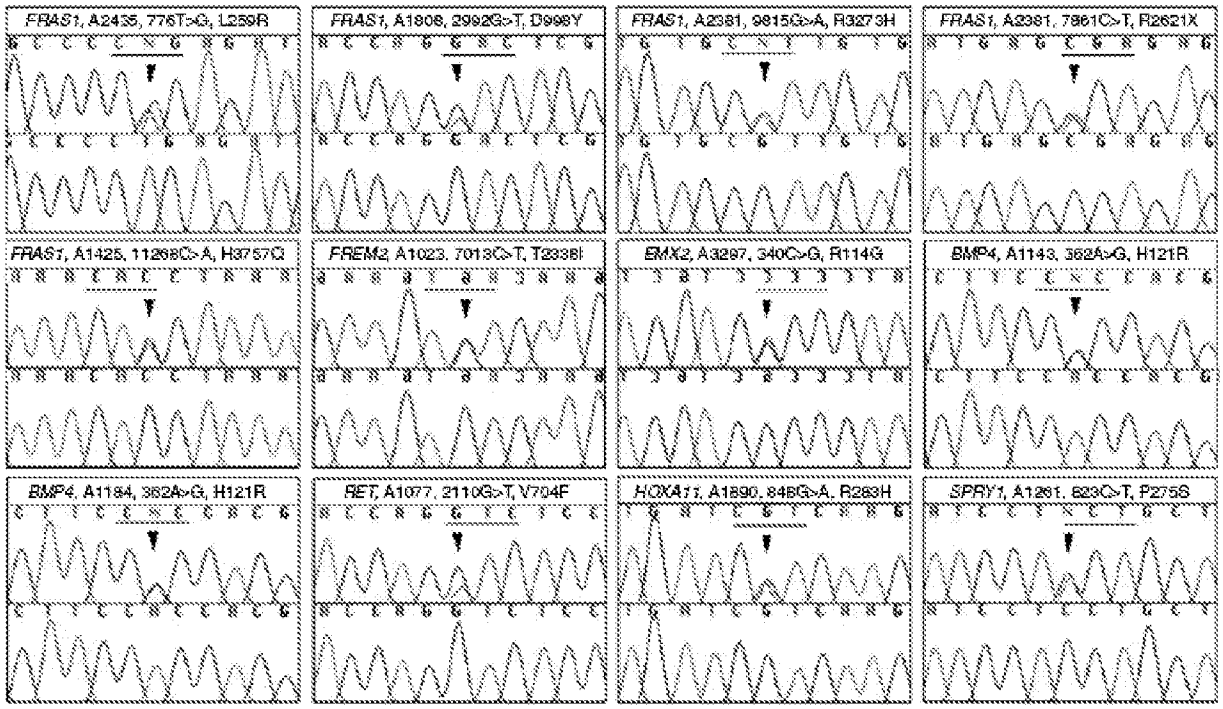

FRAS1, A2435, 776T>G, L259R
FRAS1, A1808, 2992G>T, D998Y
FRAS1, A2381, 9815G>A, R3273H
FRAS1, A2381, 7861C>T, R2621X
FRAS1, A1425, 11268C>A, H3757Q
FREM2, A1023, 7013C>T, T2338I
EMX2, A3297, 340C>G, R114G
BMP4, A1143, 362A>G, H121R
BMP4, A1184, 362A>G, H121R
RET, A1077, 2110G>T, V704F
HOXA11, A1890, 848G>A, R283H
SPRY1, A1261, 823C>T, P275S

Figure 14

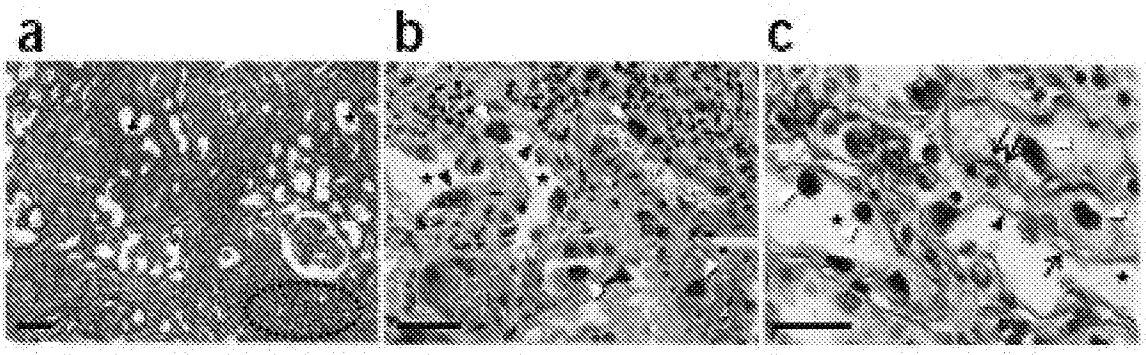


Figure 15

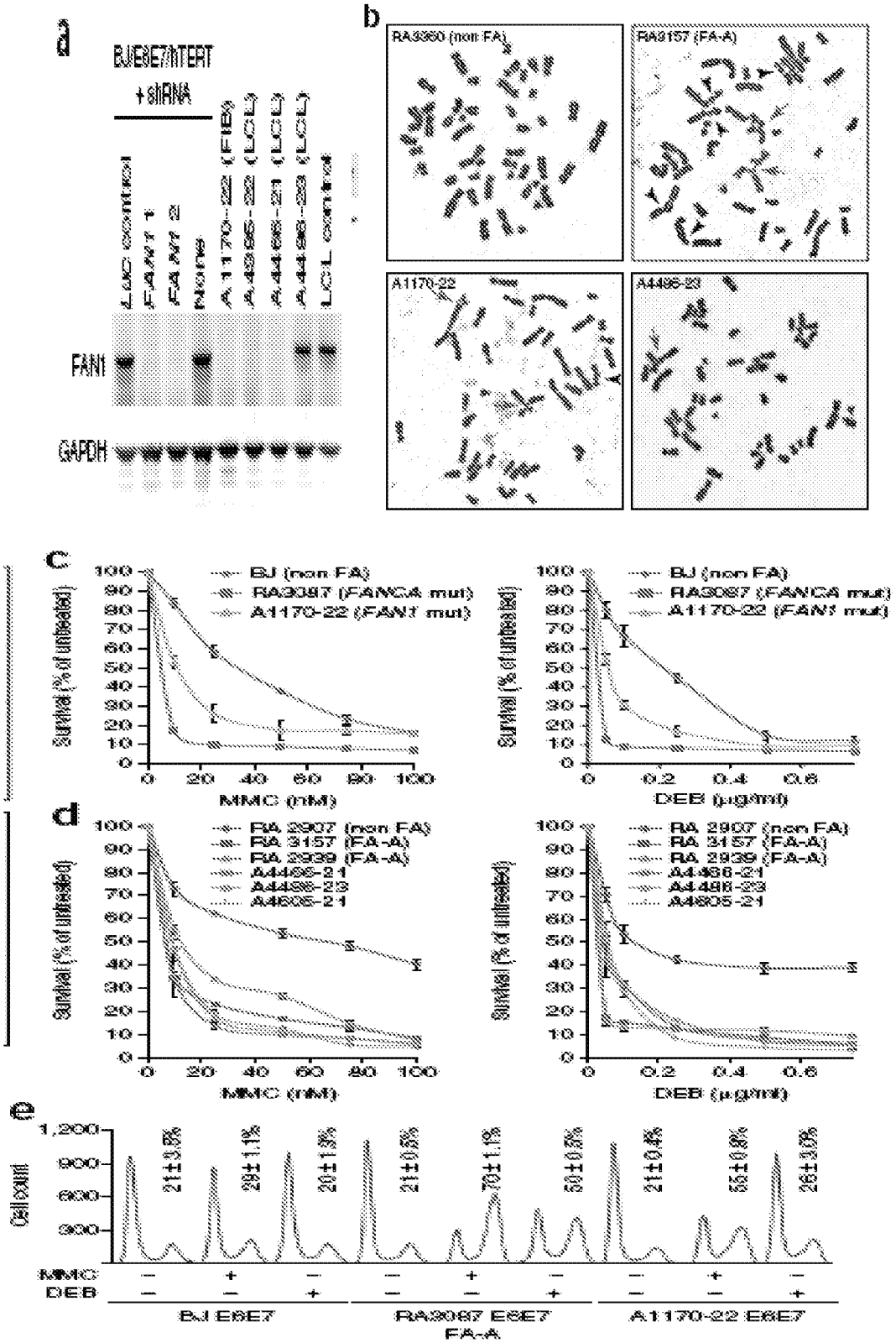

a
BJ/E6E7/hTERT
+siRNA
LMC control
FANT 1
FANT 2
None
A1170-22 (FIB)
A4386-22 (LCL)
A4486-21 (LCL)
A4486-23 (LCL)
LCL control
FANI
GAPDH
b
RA3360 (non FA)
RA3157 (FA-A)
A1170-22
A4486-23
c
BJ (non FA)
RA3087 (FANCA mut)
A1170-22 (FANI mut)
Survival (% of untreated)
MMC (nM)
DEB (μg/ml)
d
RA 2907 (non FA)
RA 3157 (FA-A)
RA 2939 (FA-A)
A4486-21
A4486-23
A4605-21
e
Cell count
21±9.8%
29±1.1%
20±1.3%
21±0.5%
70±1.1%
50±0.5%
21±0.4%
55±0.8%
26±3.0%
MMC
DEB
BJ E6E7
RA3087 E6E7
A1170-22 E6E7
FA-A

Figure 16

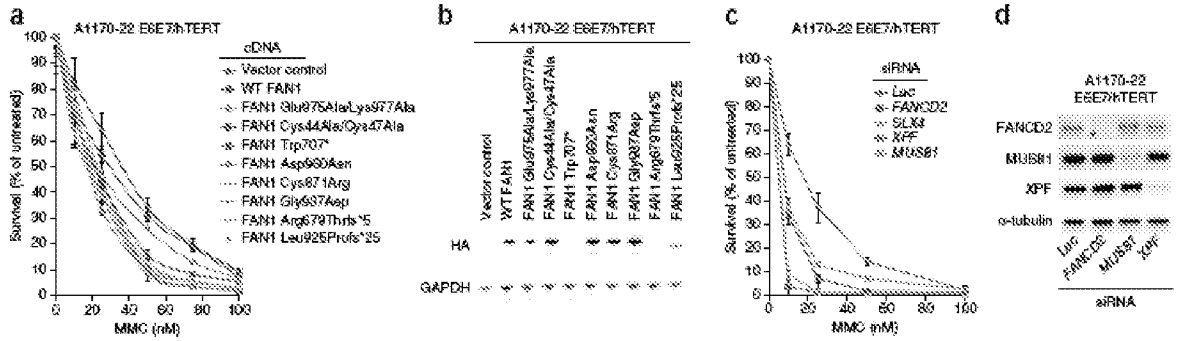


Figure 17

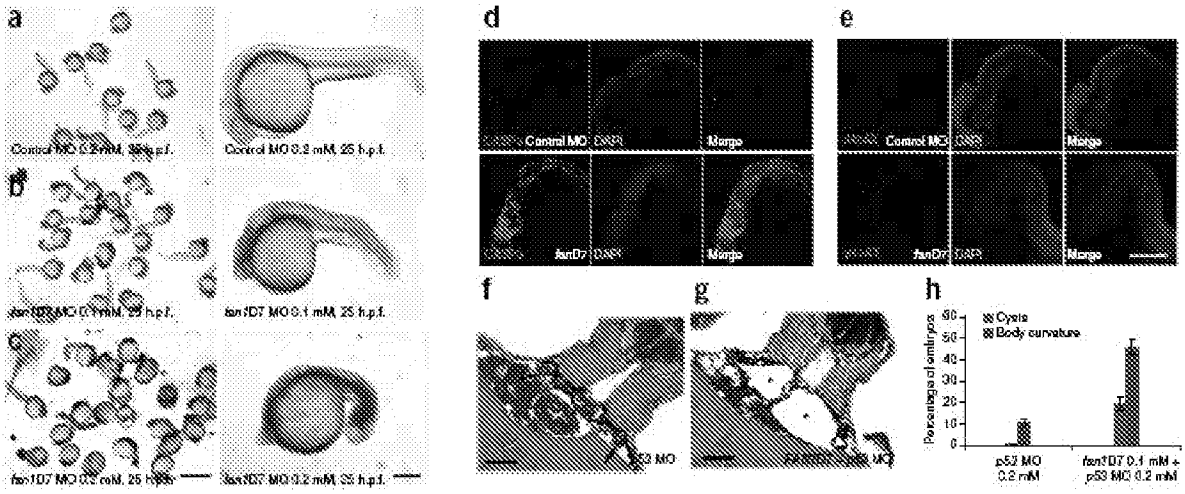


Figure 18