SUPPLEMENTARY MATERIAL

Online Methods

Table S1. Filtering process for variants from normal reference sequence (VRS) following whole exome resequencing in 10 sib pairs with a nephronophthisis-related ciliopathy (NPHP-RC).

Figure S1. Homozgosity mapping and position of primary causative gene mutations in 10 sib pairs with nephronophthisis-related ciliopathies (NPHP-RC).

Figure S2. Bioinformatics pipeline flowchart implemented for whole exome resequncing (WER).

Supplementary material is linked to the online version of the paper at http://www.nature.com/ki.

SUPPLEMENTARY MATERIAL

ONLINE METHODS

Homozygosity mapping. Non-parametric LOD scores were calculated for both affected siblings together as described in Hildebrandt et al.,[1] using ALLEGRO and assuming first-degree cousin consanguinity of the parents, regardless of actual consanguinity status. Nonparametric ZLR scores (minor allele frequency >0.2) were plotted over genetic distance across the genome, where chromosomal positions are concatenated from p to q-arm (left to right) (see **Figure S1**). "Homozygosity peaks" exceeding the empirical cut-off value of 2.0 represent possible segments of homozygosity by descent, one of which (black arrowhead) harbors the homozygous disease-causing gene mutation in each patient (see **Figure S1**).

Whole exome capture. Briefly, three μg of genomic DNA were fragmented by sonication using the Covaris™ S2 system to achieve a uniform distribution of fragments with a mean size of 300 bp. The fragmented DNA was purified using Agencourt's AMPure XP Solid Phase Reversible Immobilization paramagnetic (SPRI) beads followed by polishing of the DNA ends by removing the 3' overhangs and filling in the 5' overhangs that resulted 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 Illumina paired-end adaptors in a standard ligation reaction using T4 DNA ligase and 2-4 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 was 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 to the DNA library COT1 DNA and blocking oligonucleotides, desiccating the DNA completely and resuspending the material in NimbleGen hybridization buffer. The resuspended material was denatured at 95ºC prior to addition of the exome capture library bait material. The DNA library and biotin-labeled capture library were then hybridized by incubation at 47ºC for 68 hours. 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 100 nt sequencing on the Illumina HiSeq2000.

Massively parallel sequencing was performed largely as described in Bentley et al.[2] Briefly, following dilution of exome capture libraries to 10 nM final concentration based on the real-time PCR and bioanalyzer results, the final library stock was used in paired-end (PE) cluster generation at a final concentration of 6-8 pM to achieve a cluster density of 600,000/mm^2 on the Illumina HiSeq2000

sequencer (v2.5 reagents). Following cluster generation, 100 nt paired-end sequencing was performed using the standard Illumina protocols.

Mutation calling. Following whole exome sequencing, mutation calling was performed using the CLC Genomic Workbench™ software. Minimum length fraction of a read to match the reference sequence was set to 90%. For SNP detection, the minimum quality score of the central base as well as and the minimum average quality score of surrounding bases were kept at default (20 and 15, respectively). Quality assessment was performed within a window of 11 bases. Only reads which uniquely aligned to the reference genome were used for variant SNP or DIP (deletion/insertion polymorphism) calling. In patients with evidence of homozygosity by descent, the threshold for the number of reads (minor allele frequency) was set to >55%. In individuals lacking significant homozygous regions, the presence of a compound heterozygous mutation was considered more likely (see A2841-21 and F838 in **Figure S1**), and therefore minor allele frequency was set to >20%. The threshold coverage for minimum valid reads ('minimum variant count'), which displays the variant at a given position was set to 2 reads.

Filtering of variants from normal reference sequence (VRS).[3] For deletions/insertions (DIPs) and single nucleotide polymorphisms (SNPs) we used the following *a priori* criteria to restrict the high number of VRS (average of 55,701 for DIPs and 317,353 for SNPs) as follows (see **Table S1** and **Figure S2**):

i) We retained *exonic* variants (missense, nonsense, indels) and obligatory splice site variants only (retained average of 387 DIPs and 7,116 SNPs).
ii) We included only VRSs that are not listed in the data base 'SNP132' of innocuous polymorphisms (retained average of 213 DIPs and 1,050 SNPs).
iii) We evaluated exonic changes only within genomic regions, in which homozygosity mapping showed linkage for both affected siblings (retained average 51 for DIPs and 210 for SNPs). (These numbers were higher in the two families (A2841 and F838) without homozygosity by descent (see **Table S1** and **Figure S2**).
iv) Variants were analyzed using the program BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat?command=start) at the UCSC human genome Bioinformatics Browser (http://genome.ucsc.edu/) for the presence of paralogous genes, pseudogenes, misalignments at ends of sequence reads, and whether the variant is a known dbSNP132 with an allele frequency >1% in Caucasian populations. In families, in whom mapping demonstrated homozygosity by descent, we retained only homozygous variants and scrutinized all of them in the sequence alignments within CLC Genomic Workbench™ software for the presence of mismatches, indicating potential false alignments or poor sequence quality (retained variants ranged from 0 – 5 for DIPs and 0 – 13 for SNPs [in a heterozygous family this number was 62]).
v) Sanger sequencing was performed to confirm the remaining variants in original DNA samples and to test for intrafamilial segregation in a recessive mode.
vi) Finally, remaining variants were ranked by the criteria whether mutations were truncating the conceptual reading frame (non-sense, frameshift and obligatory splice variants) or by evolutionary conservation analysis of missense variants, and by using web-based programs predicting the impact of disease candidate variants on the encoded protein or whether they were known disease-causing mutations (see **Table S1** and **Figure S2**).

This approach led to identification of the recessive disease-causing gene in 7 of 10 sibs (see **Table S1**).

Segregation analysis by Sanger sequencing. We applied Sanger dideoxy-terminator sequencing for confirmation and segregation of potential disease-causing variants in the respective patients, their affected siblings and their parents. In patients, 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. Polymerase chain reaction (PCR) was performed using a touchdown protocol described previously.[4] Sequencing was performed using BigDye® Terminator v3.1 Cycle Sequencing Kit on an ABI 3730 XL sequencer (Applied Biosystem). Sequence traces were analyzed using Sequencher (version 4.8) software (Gene Codes Corporation).

Web-based variant analysis. Predictions on the possible impact of an amino acid substitution on the chemical change, evolutionary conservation, and protein function were obtained by using the following web-based programs: PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/), SIFT, (Sorting Intolerant from Tolerant, http://sift.jcvi.org/), and Mutation Taster (http://www.mutationtaster.org/). GERP calculation was performed using http://gvs.gs.washington.edu/SeattleSeqAnnotation/.

GenBank accession numbers. The following GenBank accession numbers were used for annotation of recessive mutations: *NPHP2*, NM_014425.2; *NPHP4*, NM_015102.3; *BBS1,* NM_024649.4; *BBS9*, NM_198428.2; *ALMS1,* NM_015120.4; *SLC4A1,* NM_000342.3; *AGXT,* NM_000030.2; *IPPN5E*, NM_019892.4.

REFERENCES

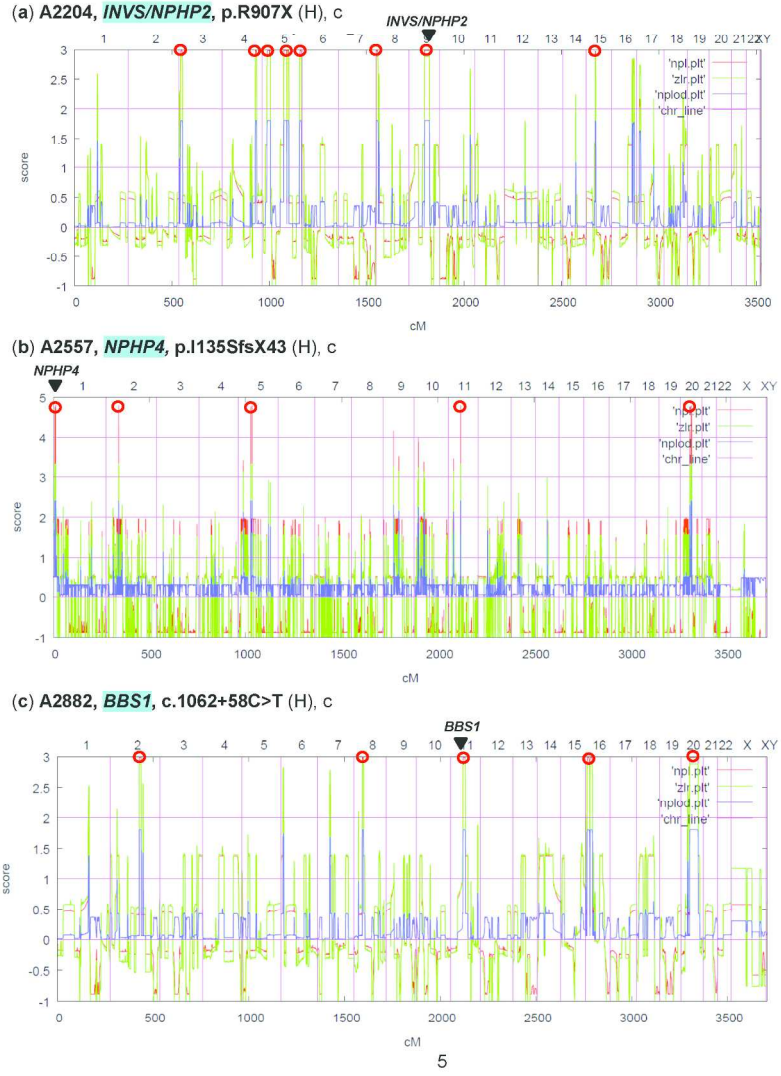
1. Hildebrandt F, Heeringa SF, Rüschendorf F*, et al.* A Systematic Approach to Mapping Recessive Disease Genes in Individuals from Outbred Populations. *PloS Genetics* 2009; **5:** 31000353.
2. Bentley DR, Balasubramanian S, Swerdlow HP*, et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 2008; **456:** 53-59.
3. Otto EA, Hurd TW, Airik R*, et al.* Candidate exome capture identifies mutation of SDCCAG8 as the cause of a retinal-renal ciliopathy. *Nat Genet* 2010.
4. Otto EA, Helou J, Allen SJ*, et al.* Mutation analysis in nephronophthisis using a combined approach of homozygosity mapping, CEL I endonuclease cleavage, and direct sequencing. *Hum Mutat* 2008; **29:** 418-426.

Table S1. Filtering process for variants from normal reference sequence (VRS) following whole exome resequencing in 10 sib pairs with a nephronophthisis-related ciliopathy (NPHP-RC).

[a]AFFECTED SIBS	Consanguinity	[b]# of homozygosity peaks	Cumulative homozygosity [Mb]	[c]Hypothesis from mapping: homozygous (H), heterozygous (h)	Total sequence reads (Mil.)	Matched Reads	Total DIPs	Exonic DIPs	% exonic / total DIPs	DIPs not present in dbSNP129	DIPs in linked region	DIPs after inspection and not present in dbSNP132 (MAF > 1%)	[d]Sanger confirmation / Segregation	Total SNPs	Exonic SNPs	% exonic / total SNPs	SNPs not present in dbSNP129	SNPs in linked region	SNPs after inspection and not present in dbSNP132 (MAF > 1%)	[d]Sanger confirmation / Segregation	Causative gene	Mutation effect on gene product	Comments
A2204	Y	8	190	H	91	96.2%	26,375	298	1.13%	119	28	0	-	263,033	5,589	2.12%	434	76	3	1	*INVS*	R907X (H)	[e]Known mutation
A2557-21	Y	5	10	H	144	98.2%	41,117	318	0.77%	142	21	1	1	312,020	5,500	1.76%	505	91	7	0	*NPHP4*	I135SfsX43 (H)	Novel mutation
A2882-21	Y	5	86	H	123	96.7%	69,898	359	0.51%	129	31	0	-	336,133	5,897	1.75%	615	52	0	-	*BBS1*	c.106Z+58C>T (H) cryptic splice site	[e]Known intronic mutation found by homozygosity mapping
A2888-22	Y	1	10	H	108	96.7%	83,896	427	0.51%	262	43	3	0	334,566	6,261	1.87%	658	61	13	0	*BBS9*	c.153A>G (H) splice donor	Novel intronic mutation found by homozygosity mapping
A2841-21	N	0	-	[f]h	163	97.2%	71,040	682	0.96%	451	116	5	1	278,884	13,127	4.71%	3,620	925	5	1	*ALMS1*	S1967X (h) L2797fsX3 (h)	Two novel mutations
F855-21	Y	3	41	H	156	96.7%	115,184	407	0.35%	357	142	1	1	448,939	5,877	1.31%	726	117	6	1	*SLC4A1*	delF524 (H)	Novel mutation
A3254-21	Y	3	17	H	60	93.0%	37,552	258	0.69%	107	[b]1	1	0	466,986	5,690	1.22%	511	[b]13	6	2	*AGXT*	M195R (H)	[e]Known mutation
F93-25	Y	1	1	H	112	97.7%	48,036	291	0.61%	134	43	3	0	274,961	5,537	2.01%	478	143	4	0	-	-	No mutations found, but mapping to *PKHD1* locus only
F838-21	N	0	-	[b]h	98	98.0%	22,033	522	2.37%	268	66	2	1	173,672	12,176	7.01%	2,522	604	62	0	*INPP5E*	Q309X (h)	Only 1 heterozygous mutation detected
A2059-21	N	1	2	H	123	98.4%	31,879	303	0.95%	160	16	0	-	284,340	5,502	1.94%	432	26	11	0	-	-	No segregating mutation found
Average		3	36		118	96.9%	54,701	387	0.89%	213	51	2	0 or 1	317,353	7,116	2.57%	1,050	210	12	1			
Median		2	13.5		117.5	97.0%	44,577	339	0.01%	151	43	1	0	298,180	5,784	1.90%	563	91	6	0.5			
Range		0-8	1-190		60-163	93%-98.4	18,977-115,184	258-682	0.35%-2.37%	107-451	1-142	0-5	0-1	173,672-466,986	5,500-13,127	1.22-7.01%	432-3,620	13-925	0-62	0-2			

[a]see Table 1
[b]see Figure S1 in Supplementary Material
[c]evaluation for heterozygous variant was done in regions of linkage for 2 sibs (1/4 of the genome) and at least one truncating mutation (DIP) was required.
[d]red numbers denote number of filtered-down variants that contained the disease causing gene.
[e]published in Otto et al. Nat Genet 34:413-20, 2003
[f]published in Abu Safieh et al. Med Genet 47:236-41, 2010
[g]published in Frishberg et al. Am J Nephrol 25:269, 2005
[h]numbers are low because mapping in cousins restricts to smaller candidate regions.
"-" , not applicable; db, database; DIP, deletion/insertion polymorphism; H, homozygous; h, heterozygous; MAF, minor allele frequency; N, no; SNP, single nucleotide polymorphism; Y, yes.

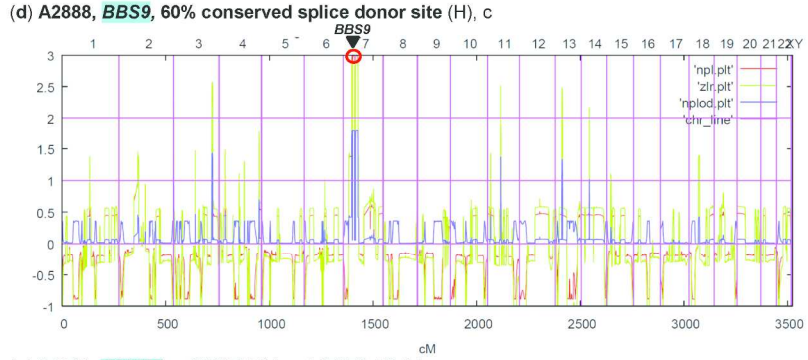
279x215mm (300 x 300 DPI)



5

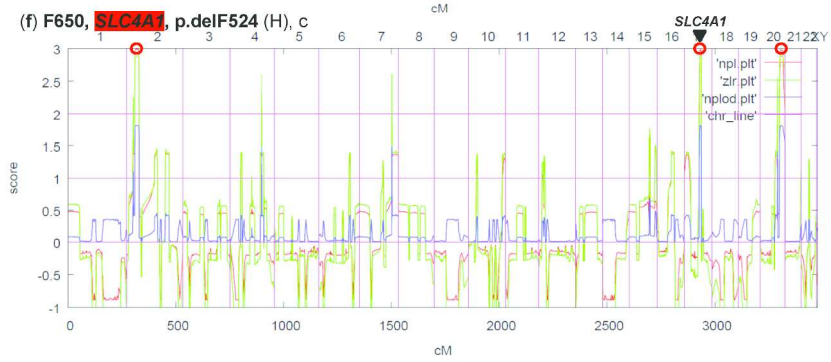
215x279mm (300 x 300 DPI)

(d) A2888, *BBS9*, 60% conserved splice donor site (H), c



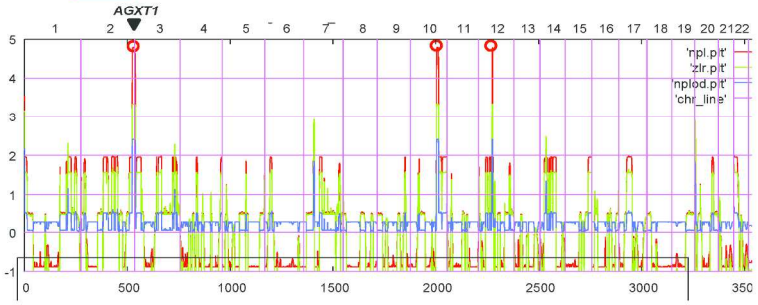
(e) A2841, *ALMS1*, p.S1967X (h); p.L2797fsX3 (h), nc

(f) F650, *SLC4A1*, p.delF524 (H), c

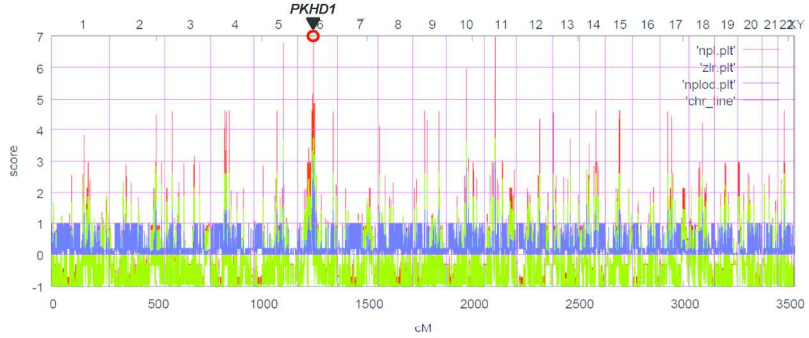


215x279mm (300 x 300 DPI)

(g) A3254, *AGXT1*, p.M195R (H), c



(h) F93, homozygosity mapped to a single gene locus, *PKHD1*, c



(i) F838, *INPP5E*, p.Q309X (h), nc

7

215x279mm (300 x 300 DPI)

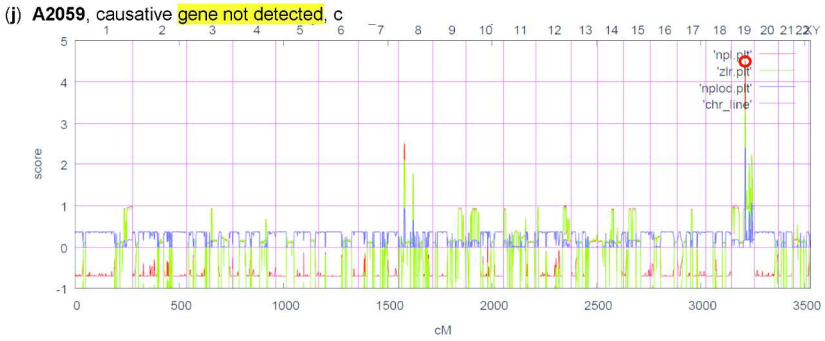


Figure S1. Homozgosity mapping and position of causative gene mutations in 10 sib pairs with nephronophthisis-related ciliopathies (NPHP-RC).

Homozygosity profiles (see **Methods**) are shown for 10 sib pairs with an NPHP-RC. Non-parametric LOD scores were calculated for both affected siblings together as described in **Methods** and in Hildebrandt et al.,[1] using ALLEGRO and assuming first-degree cousin consanguinity of the parents. Nonparametric LOD scores were plotted over genetic distance across the genome, where chromosomal positions are concatenated from p to q-arm (left to right). "Homozygosity peaks"[1] (red circles) represent possible segments of homozygosity by descent, one of which (arrow head) harbors the disease-causing gene in each patient (see **Table 1**). Plots are listed in the same order of families as in **Table 1** and **Supplementary Table S1**. Each upper left corner depicts family name, *gene symbol*, effect of mutation on translation product, (H) for homozygous, (h) for heterozygous, "c" for consanguineous, and "nc" for non-consanguineous, respectively. Genes are highlighted in blue for known NPHP-RC genes, red for known NPHP-phenocopying genes, and yellow for unsolved cases.

215x279mm (300 x 300 DPI)

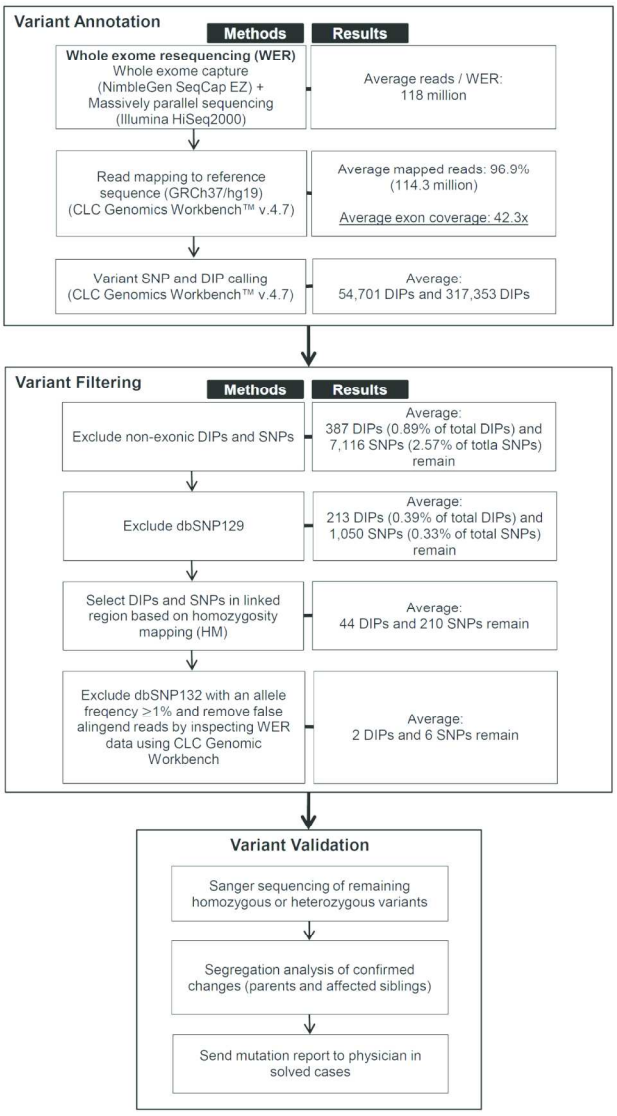


Figure S2. **Bioinformatics pipeline flowchart implemented for whole exome resequncing (WER).** SNP, single nucleoride polymorphism; DIP, deletion insertion polymorphism.

215x279mm (300 x 300 DPI)