SUPPLEMENTARY INFORMATION

Legends

Figure 1S. Use of TGE&NGS in clinical diagnostics requires a multistep, closely integrated pipeline. A) Sample Registration: Upon receipt, samples are logged into the MORL Clinical Database and DNA is extracted and tested for quality. B) TGE&NGS: Samples that pass DNA quality control are carried forward for library preparation and targeted enrichment of select genomic regions identified for sequence analysis in the design of the GRP. Massively parallel sequencing (MPS, often used interchangeably with NGS) is done using either Illumina HiSeq or MiSeq instruments. C) NGS Data Process: If more than 1.9 million total reads (per sample) are generated, the sample is moved forward for data processing. Data are archived as fastq files and are analyzed on dedicated computing resources using locally implemented open-source Galaxy software on a high-performance computing cluster. The workflow for variant calling integrates multiple publicly available tools. After variant filtering (minor allele frequency less than 1% is one filtering metric), variant annotation is performed with a customized Annotation and Reporting Tool developed by our bioinformatics team. D) CNV Analysis: Concurrent to the TGE&NGS step, copy number variants are scored across the *CFHR3-CFHR1* genomic region using a multiplex ligation-dependent probe amplification assay. The identification of CNVs is also incorporated into the NGS Data Process, but by including MLPA, CNVs are identified and confirmed with an orthogonal technology. E) Variant Validation: Variants that are reported as pathogenic, likely pathogenic, or of uncertain significance are Sanger confirmed. In addition, exons 20-22 of *CFH* are Sanger sequenced in all cases. F) Multidisciplinary Team Meeting and Final Report Generation: Unique to the GRP pipeline is the Renal Group Meeting, a multidisciplinary meeting where genetic findings are discussed in light of clinical data to generate a meaningful

report for healthcare providers. All e-mail enquiries are reviewed and discussed in this forum to generate a consensus response.

Figure 2S. The GRP was used to screen 193 patients, most of whom were diagnosed with a TMA or C3G (see Table 1). Total number of NGS reads exceeded our minimal threshold of 1.9 million. When *CFHR3-CFHR1* are homozygously deleted, the 30X coverage data relative to the total target region (percentage of bases) drops below 94% reflecting the absence of these two genes.

Figure 3S. All rare and novel variants are shown, plotting each variant as a function of its MAF (as reported in the Exome Aggregation Consortium) relative to variant impact. As a general rule, variants that are KNOWN to be pathogenic are rare and impact conserved amino acid positions. Note that nearly all 'pathogenic' variants (red dots) fall in the bottom of the graph. Variants labeled 'likely pathogenic' (blue dots, see Figure 2) also fall in this area and have a composite pathogenicity score of 5 or greater (for missense variants), change a canonical splice site (+2 or -2 bp), or cause loss-of-function, in addition to having an ultra-low MAF (less than 0.1%). VUSs are more common (higher MAF), not predicted to be pathogenic (low composite pathogenicity score), or both. Variants with a MAF>1% are either 'likely benign' or 'benign' (see Figure 4).

Figure 4S. **A)** Arg148Gln in the C3β chain (yellow) has a pathogenicity score of 6, which is supported by a change from a native salt bridge with Asp133 on FB (purple) to a weaker hydrogen bond. **B)** Trp1034Arg (purple), with a pathogenicity score of 6, is changed from a neutral to a charged amino acid. This variant is located at an intersection of the C3α and C3β chains, and is predicted to cause destabilization in folding and intramolecular interactions. **C)** Leu1318Arg (purple), with a pathogenicity

score of 5, is changed from a neutral to a charged amino acid. Its proximity to FB and the C3α and C3β chains is likely to disrupt both inter- and intramolecular interactions.

Table 1S. Novel variants identified in this study

Diag. group	Gene	Position	cDNA change	AA change	Allele count in RVD*	Allele number in RVD
TMA	*CFH*	chr1:196658654:T>C	NM_000186:c.1069T>C	p.Cys357Arg	1	1186
TMA	*CFH*	chr1:196695654:->A	NM_000186:c.1928_1929insA		1	1186
TMA	*CFH*	chr1:196695745:T>A	NM_000186:c.2019T>A	p.Cys673Stop	1	1186
TMA	*CFH*	chr1:196697628:T>C	NM_000186:c.2389T>C	p.Trp797Arg	1	1186
TMA	*CFH*	chr1:196706761:G>A	NM_000186:c.2753G>A	p.Gly918Glu	1	1186
TMA	*CFHR5*	chr1:196971761:G>A	NM_030787:c.1297G>A	p.Asp433Asn	1	578
TMA	*CD46*	chr1:207940405:G>A	NM_002389:c.721G>A	p.Gly241Arg	2	1144
TMA	*CD46*	chr1:207940420:T>A	NM_002389:c.736T>A	p.Phe246Ile	2	1144
TMA	*CD46*	chr1:207940452:C>A	NM_002389:c.768C>A	p.Cys256Stop	2	1144
TMA	*DGKE*	chr17:54921378:A>G	NM_003647:c.465-2A>G		2 (1 homozygous)	350
TMA	*C3*	chr19:6690672:C>T	NM_000064:c.3457G>A	p.Glu1153Lys	1	864
TMA	*C3*	chr19:6694496:A>T	NM_000064:c.3100T>A	p.Trp1034Arg	1	864
C3G	*C3*	chr19:6711150:C>A	NM_000064:c.1327G>T	p.Ala443Ser	1	742
C3G	*C3*	chr19:6718166:C>T	NM_000064:c.443G>A	p.Arg148Gln	1	742
TMA	*CFI*	chr4:110682703:C>T	NM_000204:c.628G>A	p.Ala210Thr	1	1156
TMA	*CFI*	chr4:110687885:C>A	NM_000204:c.153G>T	p.Trp51Cys	1	1156

* RVD (Renal Variant Database) is an in-house database that records clinical and research genetic testing results.

Table 2S. Refined molecular modeling using the Force Field X (FFX) software package

Disease	Gene	Variant	ExAC MAF	PS#	Molecular Interactions
TMA	*C3*	p.Ser1619Arg	0.001096	3	Located on a non-binding surface of C3α chain opposite FB
TMA	*C3*	p.Thr1383Asn	0.000091	1	Located on a non-binding surface*
C3G	*C3*	p.Leu1318Arg	0.000082	5	Located at interface of C3α and β chains near FB
TMA	*C3*	p.Glu1153Lys	0	5	Located in interior; destabilizes folding
C3G	*C3*	p.Leu1100Pro	0.000008	6	Destabilizes an alpha helix fold
TMA	*C3*	p.Trp1034Arg	0	6	Destabilizes folding of a hydrophobic domain
C3G	*C3*	p.Ala443Ser	0	5	Located in the interior*
C3G & TMA	*C3*	p.Arg161Trp	0	3	Destroys Arg161 C3β chain intramolecular salt bridge with Glu37
C3G	*C3*	p.Arg148Gln	0	6	Intermolecular salt bridge with Asp133 becomes a hydrogen bond
C3G	*CFB*	p.Ile242Leu	0.000988	1	Buried in a hydrophobic domain*
C3G	*CFB*	p.Asp279Glu	0.000305	2	Located at the surface near the C3β chain
TMA	*CFB*	p.Glu326Asp	0.000766	0	Located on the surface; forms a salt bridge with Lys323*

Pathogenicity Score: based on GERP++, PhyloP, MutationTaster, PolyPhen2, SIFT, and LRT;

*Conservative mutations expected to cause minimal thermodynamic changes with respect to either protein folding stability and/or intermolecular interactions

Figure 1S

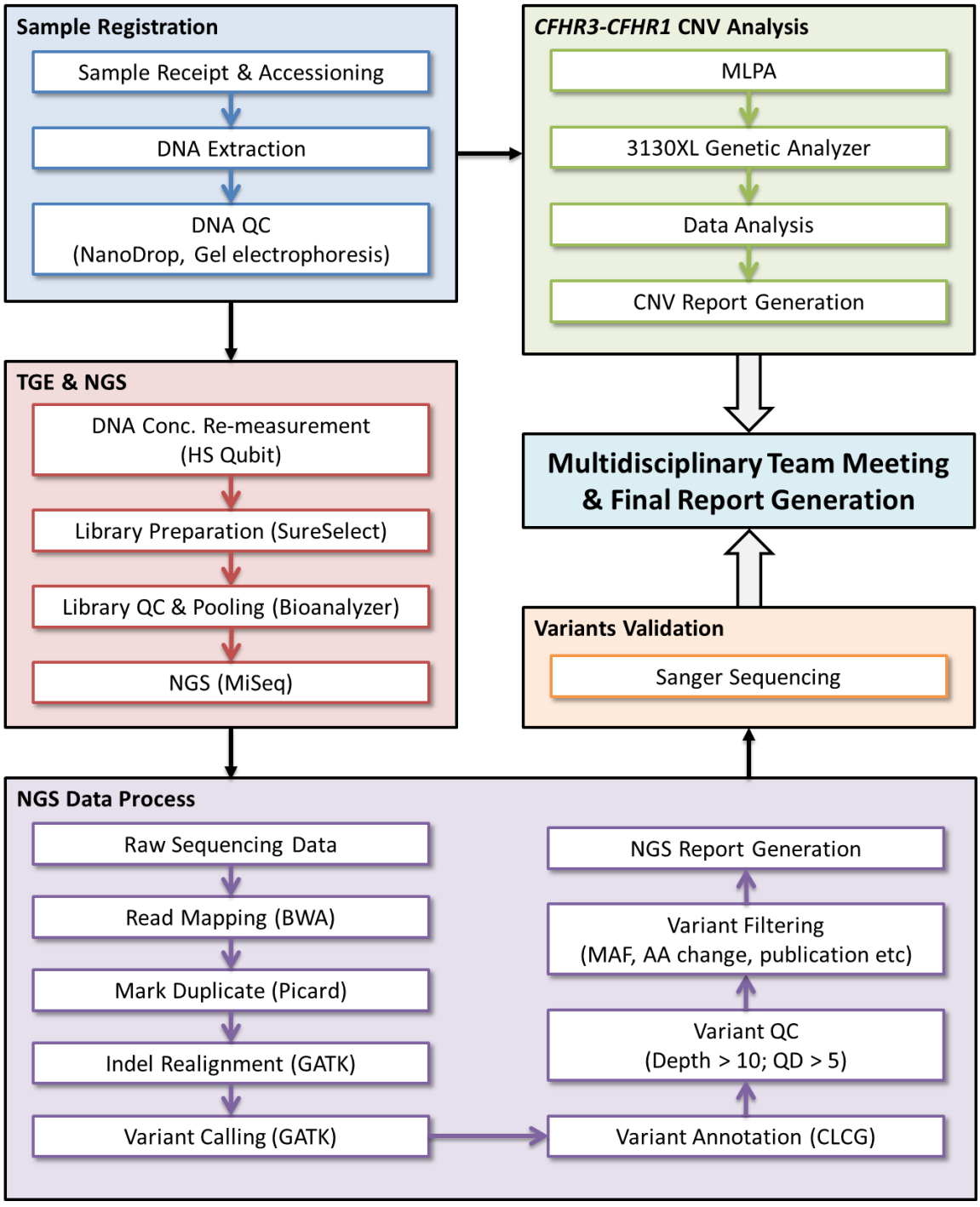
Sample Registration
Sample Receipt & Accessioning
DNA Extraction
DNA QC
(NanoDrop, Gel electrophoresis)
CFHR3-CFHR1 CNV Analysis
MLPA
3130XL Genetic Analyzer
Data Analysis
CNV Report Generation
TGE & NGS
DNA Conc. Re-measurement
(HS Qubit)
Library Preparation (SureSelect)
Library QC & Pooling (Bioanalyzer)
NGS (MiSeq)
Multidisciplinary Team Meeting
& Final Report Generation
Variants Validation
Sanger Sequencing
NGS Data Process
Raw Sequencing Data
Read Mapping (BWA)
Mark Duplicate (Picard)
Indel Realignment (GATK)
Variant Calling (GATK)
NGS Report Generation
Variant Filtering
(MAF, AA change, publication etc)
Variant QC
(Depth > 10; QD > 5)
Variant Annotation (CLCG)

Figure 2S

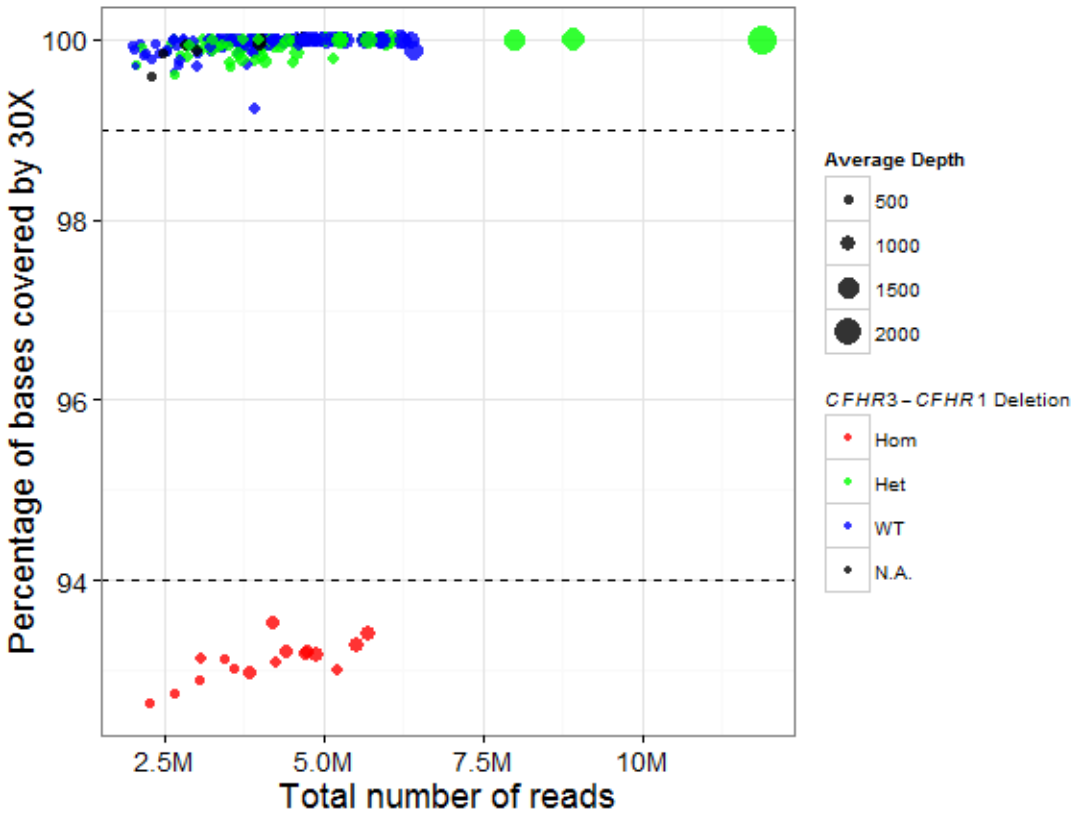
Percentage of bases covered by 30X
100
98
96
94
2.5M
5.0M
7.5M
10M
Total number of reads
Average Depth
500
1000
1500
2000
CFHR3 – CFHR1 Deletion
Hom
Het
WT
N.A.

Figure 3S

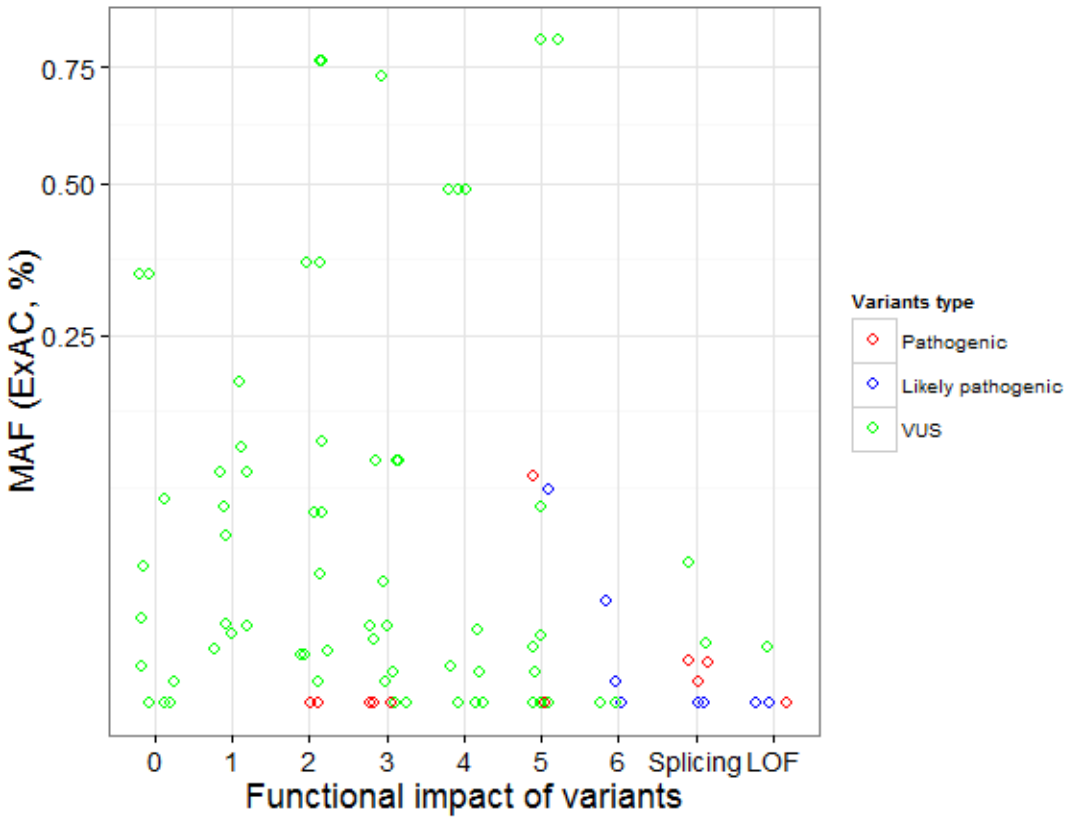
0.75
0.50
0.25
MAF (ExAC, %)
0
1
2
3
4
5
6
Splicing
LOF
Functional impact of variants
Variants type
Pathogenic
Likely pathogenic
VUS

Figure 4S