The Journal of Molecular Diagnostics, Vol. 19, No. 6, November 2017

ELSEVIER

the Journal of Molecular Diagnostics
jmd.amjpathol.org

Clinical Validation of Copy Number Variant Detection from Targeted Next-Generation Sequencing Panels

Jennifer Kerkhof,* Laila C. Schenkel,† Jack Reilly,* Sheri McRobbie,* Erfan Aref-Eshghi,† Alan Stuart,* C. Anthony Rupar,†‡ Paul Adams,§ Robert A. Hegele,¶‖ Hanxin Lin,*† David Rodenhiser,**††‡‡§§ Joan Knoll,*† Peter J. Ainsworth,*† and Bekim Sadikovic*†

From the Molecular Genetics Laboratory and the Biochemical Genetics Laboratory,‡ Molecular Diagnostics Division, London Health Sciences Centre, London, Ontario; the Departments of Pathology and Laboratory Medicine,† Gastroenterology,§ Medicine,¶ Biochemistry,** Paediatrics,†† and Oncology‡‡ and the Robarts Research Institute,‖ Western University, London, Ontario; and the London Regional Cancer Center Program,§§ the Children's Health Research Institute, London, Ontario, Canada*

Accepted for publication July 31, 2017.

Address correspondence to Bekim Sadikovic, Ph.D., DABMGG, FACMG, Department of Pathology and Laboratory Medicine, Victoria Hospital, London Health Sciences Centre, 800 Commissioner's Rd. E., B10-104, London, ON, Canada N6A 5W9. E-mail: bekim.sadikovic@lhsc.on.ca.

Next-generation sequencing (NGS) technology has rapidly replaced Sanger sequencing in the assessment of sequence variations in clinical genetics laboratories. One major limitation of current NGS approaches is the ability to detect copy number variations (CNVs) approximately >50 bp. Because these represent a major mutational burden in many genetic disorders, parallel CNV assessment using alternate supplemental methods, along with the NGS analysis, is normally required, resulting in increased labor, costs, and turnaround times. The objective of this study was to clinically validate a novel CNV detection algorithm using targeted clinical NGS gene panel data. We have applied this approach in a retrospective cohort of 391 samples and a prospective cohort of 2375 samples and found a 100% sensitivity (95% CI, 89%—100%) for 37 unique events and a high degree of specificity to detect CNVs across nine distinct targeted NGS gene panels. This NGS CNV pipeline enables stand-alone first-tier assessment for CNV and sequence variants in a clinical laboratory setting, dispensing with the need for parallel CNV analysis using classic techniques, such as microarray, long-range PCR, or multiplex ligation—dependent probe amplification. This NGS CNV pipeline can also be applied to the assessment of complex genomic regions, including pseudogenic DNA sequences, such as the *PMS2CL* gene, and to mitochondrial genome heteroplasmy detection. *(J Mol Diagn 2017, 19: 905—920; http://dx.doi.org/10.1016/j.jmoldx.2017.07.004)*

Potentially deleterious changes in DNA sequences involve single-nucleotide polymorphisms (SNPs) and/or structural variants of rearrangements that affect >50 bp, including small insertions and deletions, copy number variations (CNVs), and large structural variants. The ability to detect CNVs with a high degree of sensitivity and specificity is fundamental to a comprehensive gene analysis by a modern clinical laboratory. Inherited and somatically acquired CNVs account for a substantial proportion of genetic variation in the human genome and have been associated with a significant number of human disorders.[1–5] By definition, CNV refers to an intermediate scale structural variant, with copy number changes ranging from 1 Kb to 5 Mb of DNA[6]; however, clinically significant structural variants can range from nucleotide-level insertions/deletions to entire chromosomes and are therefore included in our definition of a CNV. Large deletions and duplications that involve dosage-sensitive developmental genes are also known to be related to the presentation of well-characterized microdeletion and microduplication syndromes, such as Charcot-Marie-Tooth and DiGeorge syndromes.[6]

Several approaches have been developed for CNV assessment, including fluorescent *in situ* hybridization,[7] multiplex ligation—dependent probe amplification (MLPA),[8] comparative

Supported by the London Health Sciences Molecular Genetics Laboratory research and development fund.

Disclosures: None declared.

Copyright © 2017 American Society for Investigative Pathology and the Association for Molecular Pathology. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmoldx.2017.07.004

Table 1 Next-Generation Sequencing Target Panels and Respective Genes

Assorted gene panel			
ACADM	*MECP2*	*RET*	*TP53*
GJB2	*MEN1*	*SCN4A*	*TTR*
GJB6	*NOTCH3*	*SPTLC1*	
BRCA*			
BRCA1	*BRCA2*		
Cancer			
APC	*CDH1*	*MSH6*	*RAD51D*
ATM	*CDK4*	*MUTYH*	*SMAD4*
BARD1	*CDKN2A*	*NBN*	*STK11*
BMPR1A	*CHEK2*	*PALB2*	*TP53*
BRCA1	*EPCAM*	*PMS2*	
BRCA2	*MLH1*	*PTEN*	
BRIP1	*MSH2*	*RAD51C*	
Charcot-Marie-Tooth syndrome			
EGR2	*HSPB1*	*MPZ*	*SH3TC2*
FIG4	*HSPB8*	*NEFL*	*TRPV4*
GARS	*LITAF*	*PMP22*	
GDAP1	*LMNA*	*PRX*	
GJB1	*MFN2*	*RAB7A*	
Dyslipidemia			
ABCA1	*CAV1*	*LDLR*	*PAX4*
ABCC8	*CEL*	*LDLRAP1*	*PCSK4*
ABCG5	*CETP*	*LEP*	*PDX1*
ABCG8	*CIDEC*	*LEPR*	*PIK3R1*
ADIPOQ	*DYRK1B*	*LIPA*	*PLIN1*
AGPAT2	*FBN1*	*LIPC*	*POLD1*
AKT2	*FTO*	*LIPE*	*POMC*
ANGPTL3	*GCK*	*LIPG*	*PPARG*
APOA1	*GPD1*	*LMF1*	*PSMB8*
APOA5	*GPIHBP1*	*LMNA*	*PTRF*
APOB	*HNF1A*	*LMNB2*	*RXRG*
APOC2	*HNF1B*	*LPIN1*	*SAR1B*
APOC3	*HNF4A*	*LPL*	*SCARB1*
APOE	*INS*	*MC3R*	*STAP1*
ATF6	*KCNJ11*	*MC4R*	*TBC1D4*
BLK	*KCNJ6*	*MTTP*	*USF1*
BSCL2	*KLF11*	*NEUROD1*	*WRN*
C5AR2	*LCAT*	*OSBPL10*	*ZMPSTE24*
Epilepsy			
ALDH7A1	*GATM*	*NECAP1*	*SCN8A*
AMT	*GLDC*	*NEU1*	*SCN9A*
ARX	*GOSR2*	*NHLRC1*	*SLC2A1*
ASAH1	*GRIN2A*	*NRXN1*	*SLC6A8*
ATP1A2	*GRIN2B*	*PCDH19*	*SLC9A6*
ATP1A3	*HCN1*	*PHGDH*	*SPTAN1*
CDKL5	*KCNC1*	*PLCB1*	*STXBP1*
CERS1	*KCNJ10*	*PNKP*	*SUOX*
CHD2	*KCNJ11*	*PNPO*	*SYNGAP1*
CHRNA7	*KCNQ2*	*POLG*	*TBC1D24*
CNTNAP2	*KCNQ3*	*PRICKLE2*	*TCF4*
CSTB	*KCNT1*	*PRRT2*	*TSC1*
DNM1	*KCTD7*	*PSAT1*	*TSC2*
DOCK7	*LMNB2*	*PSPH*	*UBE3A*
EPM2A	*MBD5*	*SCARB2*	*ZEB2*
FOLR1	*MECP2*	*SCN1A*	
FOXG1	*MEF2C*	*SCN1B*	
GAMT	*MOCS1*	*SCN2A*	
Hyperferritinemia			
ALAS2	*FTH1*	*HFE2*	*STEAP3*
B2M	*FTL*	*SEC23B*	*TF*
CDAN1	*HAMP*	*SLC25A38*	*TFR2*
CP	*HFE*	*SLC40A1*	
Lysosomal storage/urea cycle disorder			
AGA	*CTSK*	*GUSB*	*NEU1*
ARG1	*DNAJC5*	*HEXA*	*NPC1*
ARSA	*FUCA1*	*HEXB*	*NPC2*
ARSB	*GAA*	*HGSNAT*	*OTC*
ASAH1	*GALC*	*HYAL1*	*PPT1*
ASL	*GALNS*	*IDS*	*PSAP*
ASS1	*GBA*	*IDUA*	*SGSH*
CA5A	*GLA*	*LAMP2*	*SLC17A5*
CLN3	*GLB1*	*LIPA*	*SLC24A2*
CLN5	*GLUD1*	*MAN2B1*	*SLC25A13*
CLN6	*GLUL*	*MANBA*	*SLC25A15*
CLN8	*GM2A*	*MCOLN1*	*SLC7A7*
CPS1	*GNPTAB*	*MFSD8*	*SMPD1*
CTNS	*GNPTG*	*NAGA*	*SUMF1*
CTSA	*GNS*	*NAGLU*	*TPP1*
CTSD	*GRN*	*NAGS*	
Mitochondrial DNA			
MT-ATP6	*MT-ND4L*	*MT-TI*	*MT-TS2*
MT-ATP8	*MT-ND5*	*MT-TK*	*MT-TT*
MT-CO1	*MT-ND6*	*MT-TL1*	*MT-TV*
MT-CO2	*MT-TA*	*MT-TL2*	*MT-TW*
MT-CO3	*MT-TC*	*MT-TM*	*MT-TY*
MT-CYB	*MT-TD*	*MT-TN*	*MT-RNR1*
MT-ND1	*MT-TE*	*MT-TP*	*MT-RNR2*
MT-ND2	*MT-TF*	*MT-TQ*	
MT-ND3	*MT-TG*	*MT-TR*	
MT-ND4	*MT-TH*	*MT-TS1*	

*BRCA panel description given by Schenkel et al.[12]

genomic hybridization microarrays, and SNP arrays.[9] Although large chromosomal and segmental rearrangements are detectable by fluorescent *in situ* hybridization, most CNVs identified in the human genome are below the resolution of current fluorescent *in situ* hybridization technology. To date, MLPA and microarray-based technologies have been the most reliable and effective methods for discovering copy number alterations. Although both comparative genomic hybridization and SNP array techniques provide a genome-wide CNV screening capability, and SNP arrays also allow allelotyping, generally small deletions/duplications need confirmation by an alternate method or are not detectable.[10] MLPA is a semiquantitative PCR-based technique that can detect deletions and duplications for up to 50 genetic loci in one assay.[11] Because of its low cost, high sensitivity and specificity, and medium throughput, the MLPA technique has become the gold standard diagnostic tool. However, drawbacks of MLPA include inability to provide information regarding the exact location of a duplicated sequence or its orientation, lack of sensitivity for regions not directly encompassed by the

Table 2 Summary of Panel Sequencing Read Depth Threshold and Variability

Panel	Panel size, bp	No. of panel regions	Minimum RD QC threshold*	Avg. RD QC threshold	Panel regions meeting QC threshold, %†	Mean ± SEM avg. RD‡	Mean ± SEM minimum RD§	Mean minimum RD range	Mean ± SEM on-target percentage
Assorted	30,279	133	100	500	100	1280 ± 484	1052 ± 397	125–1630	24.1 ± 8.2
BRCA	17,769	48	200	1000	100	7619 ± 2749	6645 ± 2441	2998–8055	88.3 ± 1.7
Cancer	90,140	385	100	300	100	726 ± 291	578 ± 235	114–1048	40.4 ± 4.4
Charcot-Marie-Tooth syndrome	34,304	142	100	500	100	1044 ± 437	844 ± 346	241–1054	20.7 ± 6.4
Dyslipidemia	170,595	808	100	500	99.4 (5)	1138 ± 314	914 ± 262	4.2–1651	56.7 ± 10.5
Epilepsy	219,783	1018	100	300	99.3 (7)	857 ± 120	686 ± 106	4–1319	55.4 ± 2
Hyperferritinemia	31,368	160	100	300	100	936 ± 384	811 ± 338	203–1080	18.6 ± 7.7
Lysosomal storage/urea cycle disorder	129,620	723	100	500	99.6 (3)	1540 ± 407	1272 ± 349	24–2298	46.4 ± 2.3
Mitochondrial DNA	15,416	37	500	1000	100	17,287 ± 11,731	15,153 ± 10,309	5177–18,905	82.6 ± 7.3

*Copy number variation analysis not performed if coverage fails to meet minimum threshold.
†Number of failed regions indicated in brackets.
‡Average nucleotide read depth across the entire panel for all samples tested.
§Average minimum nucleotide read depth across the entire panel for all samples tested.
QC, quality control; RD, sequencing read depth.

probe sets used, and false-positive results attributable to polymorphism-induced allele dropouts.[8] Taken together, these assays have their unique advantages and disadvantages, related to limited coverage, low resolution, and high cost.

In contrast, recent advancements in next-generation sequencing (NGS) technologies have provided more cost-effective, rapid, and high-throughput methods to allow identification of sequence variants and CNVs.[12,13] Clinical use of NGS enables simultaneous assessment of targeted gene panels and entire exomes or genomes using a limited quantity of biological samples.[14] Some of the characteristics of using NGS for the detection of CNVs include a high resolution, the ability to detect novel small CNVs and balanced genomic rearrangements, and accurate estimation of copy number.[15–17]

CNV assessment from NGS multigene panel data most commonly uses the sequencing read depth (RD) assessment approach, which is based on the assumption that the RD signal is proportional to the number of copies of chromosomal segments present in that specimen.[18] Several bioinformatics tools for CNV assessment have been developed, including XHMM,[19] CoNIFER,[20] ExomeDepth,[21] and CONTRA.[22] However, a review evaluating these RD-based algorithms has demonstrated several limitations concerning sensitivity and specificity.[18] Another bioinformatic algorithm, CoNVaDING, which focuses on specific target region and uses selected control samples for RD comparison, has demonstrated high sensitivity and specificity to detect CNVs using targeted NGS.[23]

In this report, we describe the validation and clinical use of a noncomputationally intensive CNV detection algorithm using targeted clinical NGS gene panel data. We have demonstrated high positive predictive values and sensitivity for the detection of CNVs, >50 bp in nine distinct clinical gene panels, with a stand-alone first-tier pipeline for the routine analysis of CNVs and sequence variants in a clinical laboratory. This approach dispenses with the need for parallel CNV analyses. In addition to the assessment of the unique gene sequences and CNV, this approach also enables the assessment of complex genomic regions, including pseudogenic DNA sequences, such as the *PMS2CL* gene in the hereditary cancer panel, and in the quantification of mitochondrial genome heteroplasmy.

Materials and Methods

Sample Collection and DNA Isolation

The retrospective validation cohort was composed of 391 patient specimens previously analyzed using MLPA analysis, Southern blot methods, and/or Sanger sequencing at the Molecular Genetics Laboratory and/or Biochemical Genetics Laboratory at the London Health Science Centre and included 24 research specimens for dyslipidemia analyzed at the Robarts Research Institute, Western University. The prospective cohort includes 2375 patient specimens screened using NGS analysis at the Molecular Genetics Laboratory as part of routine clinical testing, where the samples identified with pathogenic, likely pathogenic, and/or variants of unknown clinical significance were confirmed by Sanger sequencing, MLPA, long-range PCR (LR-PCR), quantitative fluorescence PCR, real-time quantitative PCR, or a combination. All patients gave consent (implied or written) and were counseled for clinical testing as part of their clinical genetics assessment. The 24 dyslipidemia research specimens were obtained under Western University Health Science Research Ethics Review Board

Table 3 CNV Identified by NGS in the Retrospective Cohort ($n = 391$)

Panel	Gene	Transcript	No. of patients	Chromosome	Exon(s) involved	Coverage plot size, bp	Genomic size, bp	Copy number	CNV identified by NGS	CNV identified before NGS
Assorted	*GJB6*	NM_001110219.2	1	13	6	443	>443	1	c.-21-?_443del	del1309kb
Assorted	*MECP2*	NM_004992.3	1	X	4	548	548	1	c.523_1186delins85	Equivalent
Assorted	*MECP2*	NM_004992.3	1	X	4	527	535	1	c.721_1255delins CCAAGCCT	Equivalent
Assorted	*MEN1*	NM_130799.2	1	11	7	27	58	1	c.824+35_836del*	Equivalent
Assorted	*MEN1*	NM_130799.2	1	11	4–11	1708	>3806	1	c.446-?_1883+?del*	Equivalent
Assorted	*MEN1*	NM_130799.2	1	11	4–7, 9–11	1531	>3629	1	c.446-?_912+?del(;) 1050-?_1833+?del*	Equivalent
BRCA	*BRCA1*	NM_007294.3	1	17	2–24	6472	>78,459	1	c.(?_-21)_(*21_?)del	Ex1-24del
BRCA	*BRCA1*	NM_007294.3	1	17	2–24	6472	>78,459	3	c.(?_-21)_(*21_?)dup	Ex1-24del
BRCA	*BRCA1*	NM_007294.3	3	17	13	212	>212	3	c.4186-?_4357+?dup	Equivalent
BRCA/cancer	*BRCA1*	NM_007294.3	1	17	18–20	323	>6940	3	c.5075-?_5277+?dup	Equivalent
BRCA	*BRCA1*	NM_007294.3	1	17	24	165	>165	1	c.5468-?_5592+?del	Equivalent
BRCA/cancer	*BRCA2*	NM_000059.3	1	13	8–10, 12–13	1644	>6364	1	c.632-?_1909+?del(;) 6842-?_7007+?del	Equivalent
BRCA	*BRCA2*	NM_000059.3	1	13	19–20	381	>738	1	c.8332-?_8632+?del	Equivalent
CMT	*GJB1*	NM_001097642.2	1	X	3	892	~150 M	3	c.(?_-21)_(*21_?)dup	XXX
CMT	*MPZ*	NM_000530.6	1	1	1–6	987	>4070	4	c.ins(?_-21)_(*21_?)[2]	Equivalent
CMT	*PMP22*	NM_000304.3	2	17	2–5	643	>29,851	1	c.(?_-21)_(*21_?)del	Equivalent
CMT	*PMP22*	NM_000304.3	2	17	2–5	643	>29,851	3	c.(?_-21)_(*21_?)dup	Equivalent
Cancer	*MLH1*	NM_000249.3	1	3	12	411	>411	1	c.1039-?_1409+?del	Equivalent
Cancer	*MSH2*	NM_000251.2	1	2	12–14	819	>3535	1	c.1760_2458del	Equivalent
Cancer	*PMS2*	NM_000535.5	2	7	1–5	738	>6607	1	c.-21-?_537+?del	Equivalent
Cancer	*PMS2*	NM_000535.5	2	7	3–7	840	>6772	1	c.164-?_803+?del	Ex5-7del
Cancer	*PMS2*	NM_000535.5	1	7	14	210	>210	1	c.2276-?_2445+?del	Equivalent
Cancer	*PMS2CL*	NR_002217.1	1	7	4–6	535	>535	1	n.1115-?_1549+?del	Equivalent
Dyslipidemia	*ABCA1*	NM_005502.3	1	9	4	182	>182	1	c.161-?_302+?del	—
Dyslipidemia	*AGPAT2*	NM_006412.3	1	9	3–4	283	1037	0	c.366_588+534del	?1 kb deletion
Dyslipidemia	*LDLR*	NM_000527.4	1	19	1	107	>107	1	c.-21-?_67+?del	Equivalent
Dyslipidemia	*LDLR*	NM_000527.4	1	19	3–6	910	>4891	1	c.191-?_940+?del	Equivalent
Dyslipidemia	*LDLR*	NM_000527.4	1	19	16–18	392	>3349	1	c.2312-?_2583+?del	Equivalent
Dyslipidemia	*LPIN1*	NM_001261428.1	1	2	23	155	1763	1	c.2550-865_2665-29del	—
Dyslipidemia	*MTTP*	NM_000253.2	1	4	11–16	1064	11,601	0	c.1237-204_2080del	Equivalent
Dyslipidemia	*MTTP*	NM_000253.2	1	4	11–16	1064	11,601	1	c.1237-204_2080del	Equivalent
LSD/UCD	*CTNS*	NM_001031681.2	1	17	3–10	1147	>17,984	0	c.-21-?_847del	57-kb deletion
LSD/UCD	*GALC*	NM_000153.3	1	14	9–10	219	7474	1	c.752+3257_910del	Equivalent
mtDNA	—	NC_012920.1	1	m	NA	7850	7850	80%	m.6249_14098del	Equivalent −80%
mtDNA	—	NC_012920.1	1	m	NA	5906	5906	60%	m.8475_14380del	Equivalent −60%
mtDNA	—	NC_012920.1	1	m	NA	~6000	~6000	22%	~6-kb deletion	Equivalent −30%
mtDNA	—	NC_012920.1	1	m	NA	~6000	~6000	18%	~6-kb deletion	Equivalent −20%

*Normalized CNV plot showing *MEN1* deletions is presented in Supplemental Figure S1.

CMT, Charcot-Marie-Tooth syndrome; CNV, copy number variation; LSD/UCD, lysosomal storage/urea cycle disorder; mtDNA, mitochondrial DNA; NGS, next-generation sequencing.

protocol 07920E. Samples were composed of extracted DNA specimens or were peripheral blood samples from which genomic DNA was isolated by standard protocols using the MagNA Pure system (Roche Diagnostics, Laval, QC, Canada) at London Health Sciences Centre. DNA was quantified by the measurement of absorbance with a DTX 880 Multimode Detector (Beckman Coulter, Brea, CA).

Retrospective Cohort CNV Assessment

The retrospective cohort had been previously assessed by MLPA analysis, Southern blot, and/or Sanger sequencing. Briefly for MLPA assessment, 100 ng of genomic DNA was amplified according to the manufacturer's recommendations using a SALSA MLPA kit (P002-BRCA1-D1, P087-BRCA1-C1,

Table 4 Summary of CNVs Identified by NGS and Confirmation Technique in Retrospective and Prospective Cohorts

Panel	Total No. of tests	Detected by MLPA*	Detected by NGS	No. of unique variants by NGS	False-negative results	Sensitivity (95% CI), %	No. (%) of false-positive results	Analytic PPV† (95% CI), %
Retrospective								
Assorted	48	6	6	6	0	100 (52—100)	—	—
BRCA	120	9	9	7	0	100 (62—100)	—	—
Cancer	60	10	10	8	0	100 (65—100)	—	—
CMT	46	6	6	4	0	100 (51—100)	—	—
Dyslipidemia	24	6	8‡	8	0	100 (51—100)	—	—
Epilepsy	0	—	—	—	—	—	—	—
Hyperferritinemia	0	—	—	—	—	—	—	—
LSD/UCD	22	2	2	2	0	100 (19—100)	—	—
mtDNA	71	4	4	4	0	100 (39—100)	—	—
Total	391	43	45	39	0	100 (89—100)	—	—
Prospective								
Assorted	421	4	7	6	—	—	3 (0.7)	57 (20—88)
BRCA	502	1	4	3	—	—	3 (0.6)	25 (1—78)
Cancer	576	6	17	11	—	—	11 (1.9)	35 (15—61)
CMT	619	88	92	10	—	—	4 (0.6)	95 (88—98)
Dyslipidemia	0	—	—		—	—	—	—
Epilepsy	36	0	0	0	—	—	—	—
Hyperferritinemia	116	0	3	2	—	—	3 (2.6)	0 (0—69)
LSD/UCD	12	0	0	0	—	—	—	—
mtDNA	93	7	7	6	—	—	0	100 (56—100)
Total	2375	106	130	38	—	—	24 (1)	81 (73—87)
X chromosome dosage								
Total	1272	1272	1272	—	0	100 (99—100)	0	100 (99—100)

*Detected by MLPA or other confirmation test.

†Calculated based on analytical results alone and not adjusted for clinical prevalence.

‡Two additional cases were detected by NGS and later confirmed positive by long-range PCR.

CMT, Charcot-Marie-Tooth syndrome; CNV, copy number variation; LSD/UCD, lysosomal storage/urea cycle disorder; MLPA, multiplex ligation—dependent probe amplification; mtDNA, mitochondrial DNA; NGS, next-generation sequencing; PPV, positive predictive value.

P090-BRCA2-A4, P077-BRCA2-A2, P008-PMS2-C1, P033-CMT1-B, P015-MECP2-F1, P017-MEN1-C1, P003-MLH1/MSH2-C1; MRC Holland, Amsterdam, the Netherlands). PCR products were separated by capillary electrophoresis on an ABI 3730 (Life Technologies, Thermo Fisher Scientific, Waltham, MA), and copy number alterations were analyzed with Coffalyser.Net software version 131211.1524 (MRC Holland).

Southern Blot analysis was performed for mitochondrial deletion determination. Genomic DNA (0.25 to 11.25 μg) was digested with PvuII and BamHI and transferred to a nylon membrane overnight, with fixing performed for 30 minutes at 120°C. Overnight hybridization was performed at 65°C with 300 ng of digoxigenin-labeled probe prepared according to the High Prime DNA labeling kit (Life Technologies), digested with ApaI. Detection was performed according to the manufacturer's recommendations with the digoxigeninluminescent detection kit (Life Technologies), with a final exposure time between 15 and 60 minutes. Deletion size was determined with the included molecular weight marker, and heteroplasmy levels were assigned based on densitometry with Quantity One software version 4.6.1 (Bio-Rad Laboratories Inc., Hercules, CA).

Sanger sequencing was performed with the BigDye Terminator version 1.1 cycle sequencing kit (Life Technologies). Sequencing products were separated by capillary electrophoresis on an ABI 3730 (Life Technologies) and were analyzed with Mutation Surveyor software version 4.0.7 (SoftGenetics, LLC, State College, PA).

NGS Library Design

Custom sequence capture probes were designed using the SeqCap EZ Choice Library system (Roche NimbleGen, Inc., Madison, WI). The design included enrichment for all coding exons and 20 bp of the 5′ and 3′ flanking intronic regions. This design can include up to 2.1 million different probes that massively overlap each other across the target region, thereby introducing significant redundancy and the ability to capture complex, CG-rich, and polymorphic genomic regions. The SeqCap EZ Choice Library are proprietary designs that involve a Nimblegen-designed, bioinformatically targeted, probe coverage of the region of interest, which normally involves high-density tiling of the targeted region. If needed, each specific design could be

Table 5 CNV Identified by NGS in the Prospective Cohort ($n = 2375$)

Panel	Gene	Transcript	No. of patients	Chromosome	Exon(s) involved	Coverage plot size, bp	Genomic size, bp	Copy number	CNV identified by NGS	CNV identified by non-NGS
Assorted	*GJB6*	NM_001110219.2	2	13	6	443	>443	1	c.-21-?_443del	del1309kb
Assorted	*MECP2*	NM_004992.3	1	X	1	102	NA	1	c.-21-?_62+?del	False positive
Assorted	*MECP2*	NM_004992.3	1	X	3–4	1515	>2230	1	c.27-?_1461+?del	Equivalent
Assorted	*MECP2*	NM_004992.3	1	X	3–4	1222	6392	1	c.27-4474_1188del	Equivalent
Assorted	*MEN1*	NM_130799.2	1	11	10	205	NA	1	c.1186-?_1350+?del	False positive
Assorted	*SPTLC1*	NM_006415.3	1	9	1	35	NA	3	c.43_57+?dup	False positive
BRCA/cancer	*BRCA1*	NM_007294.3	2	17	2	120	>120	3	c.-21-?_80+?dup	Ex1-2dup
BRCA	*BRCA1*	NM_007294.3	1	17	5	118	NA	1	c.135-?_212+?del	False positive
BRCA	*BRCA2*	NM_000059.3	2	13	21	162	NA	3	c.8633-?_8754+?dup	False positive
CMT	*FIG4*	NM_014845.5	1	6	2	139	NA	1	c.67-?_165+?del	False positive
CMT	*FIG4*	NM_014845.5	1	6	17	99	NA	3	c.1890-?_1948+?dup	False positive
CMT	*GJB1*	NM_001097642.2	1	X	3	892	~150 M	3	c.(?_-21)_(*21_?)dup	XXX
CMT	*GJB1*	NM_001097642.2	1	X	3	892	~150 M	3	c.(?_-21)_(*21_?)dup	XXY
CMT	*HSPB1*	NM_001540.3	1	7	2	45	NA	1	c.404_428+?del	False positive
CMT	*HSPB1*	NM_001540.3	1	7	3	116	NA	3	c.523_618+?dup	False positive
CMT	*PMP22*	NM_000304.3	29	17	2–5	643	>29,851	1	c.(?_-21)_(*21_?)del	Equivalent
CMT	*PMP22*	NM_000304.3	55	17	2–5	643	>29,851	3	c.(?_-21)_(*21_?)dup	Equivalent
CMT	*PMP22*	NM_000304.3	1	17	4–5	385	>8735	1	c.179-?_*21+?del	Equivalent
Cancer	*APC*	NM_001127510.2	1	5	15	2499	4646	1	c.1958+241_4457del	Equivalent
Cancer	*BRCA1*	NM_007294.3	1	17	3–16	5426	>44,892	1	c.81-?_4986+?del	Equivalent
Cancer	*BRCA2*	NM_000059.3	1	13	22	74	NA	3	c.8920_8953+?dup	False positive
Cancer	*CHEK2*	NM_007194.3	1	22	14–15	251	>1359	1	c.1462-?_1632+?del	Equivalent
Cancer	*MSH6*	NM_000179.2	1	2	4	106	NA	3	c.769_874dup	False positive
Cancer	*PMS2*	NM_000535.5	1	7	5	126	NA	1	c.432_537+?del	False positive
Cancer	*PMS2*	NM_000535.5	1	7	5–7	570	>5351	1	c.354-?_803+?del	Equivalent
Cancer	*PMS2*	NM_000535.5	1	7	14	210	NA	1	c.2276-?_2445+?del	False positive
Cancer	*PMS2*	NM_000535.5	7	7	15	184	NA	1	c.2446-?_2589+?del	False positive
Cancer	*PMS2CL*	NR_002217.1	1	7	3–6	743	>9748	1	n.947-?_1549+?del	Equivalent
Hyperferritinemia	*ALAS2*	NM_000032.4	1	X	10	217	NA	3	c.1169-?_1365dup	False positive
Hyperferritinemia	*CP*	NM_000096.3	2	3	14	75	NA	3	c.2500_2554+?dup	False positive
mtDNA	—	NC_012920.1	1	m	NA	~4700	~4700	10%	~4.7 kb deletion	Equivalent
mtDNA	—	NC_012920.1	2	m	NA	~5000	~5000	15%	~5 kb deletion	Equivalent
mtDNA	—	NC_012920.1	1	m	NA	3895	3895	55%	m.548_4442del	Equivalent
mtDNA	—	NC_012920.1	1	m	NA	7850	7850	75%	m.6249_14098del	Equivalent
mtDNA	—	NC_012920.1	1	m	NA	7125	7125	98%	m.8483_13459del	Equivalent
mtDNA	—	NC_012920.1	1	m	NA	5938	5938	20%	m.10130_16067del	Equivalent

CMT, Charcot-Marie-Tooth syndrome; CNV, copy number variation; mtDNA, mitochondrial DNA; NA, not applicable; NGS, next-generation sequencing.

adjusted for probe concentrations on empirical assessment on patient samples to ensure a uniform depth of coverage.

The following panels were designed with the method described above and include (in brackets) the total number of nucleotides in the panel and the number of regions of interest, respectively; assorted gene panel (30279, 133), BRCA panel (17769, 48), hereditary cancer (cancer) (90140, 385), Charcot-Marie-Tooth syndrome (34304, 142), dyslipidemia (170595, 808), epilepsy (219783, 1018), hyperferritinemia (31368, 160), lysosomal storage/urea cycle disorder (129620, 723), and mitochondrial DNA (mtDNA) (15416, 37). The full list of genes analyzed for each panel can be seen in Table 1.

Library Preparation and Target Capture Sequencing

Libraries were prepared with 100 ng of genomic DNA randomly fragmented at 180 to 220 bp using a Covaris E220 Series Focused ultrasonicator (Covaris, Inc., Woburn, MA). Each sample library was ligated with a specific barcode index according to the manufacturer's protocol (Roche NimbleGen, Inc.) and then assessed for quantification and size distribution with the Qubit fluorometer (Life Technologies) and 2200 TapeStation (Agilent Technologies, Santa Clara, CA), respectively. DNA libraries were then pooled as a 12-plex (dyslipidemia, epilepsy, and lysosomal storage/urea cycle disorder panels) or 24-plex panel (assorted,

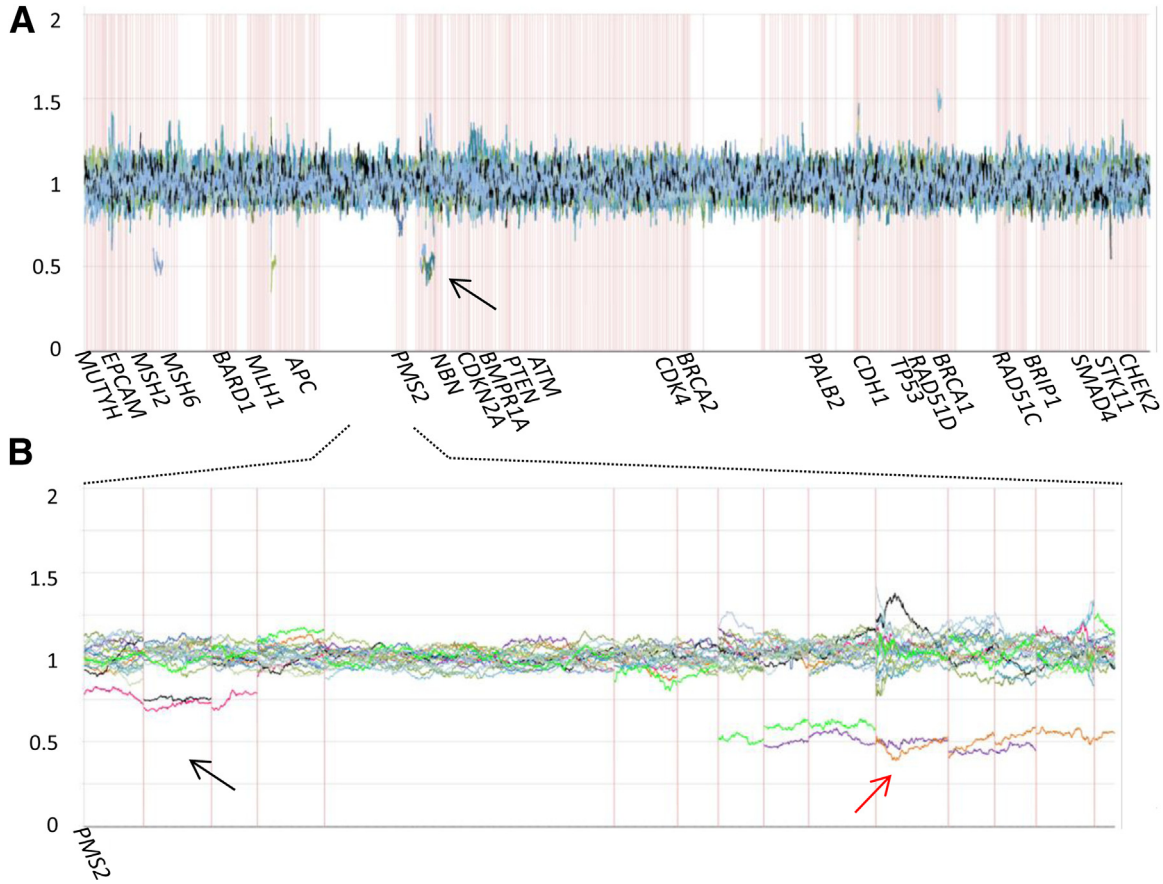


Figure 1 Hereditary cancer panel: *PMS2* deletion detection. **A:** Normalized copy number variant (CNV) plot demonstrating deletion detection at the *PMS2* gene (**arrow**). Additional CNVs are also shown in *MSH2* (c.1760_2458del), *MLH1* (c.1038-?_1409+?del), and *BRCA1* (c.5074-?_5277+?dup). **B:** Zoomed-in view of the *PMS2* gene shows three *PMS2* CNVs identified in the 5′ region of the gene with a ratio of 0.5 (**red arrow**), whereas another two (**black arrow**) are in the region of high homology with *PMS2CL* and are identified by a deletion ratio of 0.75. *PMS2/PMS2CL* deletions sizes range from 210 to 840 bp. *x* axis indicates gene-exon locations. **Red lines** indicate exon boundaries. *y* axis represents quantile normalized copy number data (for unique autosomal genes, 0.5 indicates 1 copy; 1, 2 copies; and 1.5, 3 copies; for homologous autosomal genes with their pseudogene, 0.75 indicates 3 copies; 1, 4 copies; and 1.25, 5 copies). Constitutional deletions are defined by a mean ratio of $\leq$0.65, and duplications are defined by a ratio of $\geq$1.35. Homologous region *PMS2/PMS2CL* deletions and duplications are assessed by a ratio of $<$0.8 and $>$1.2, respectively.

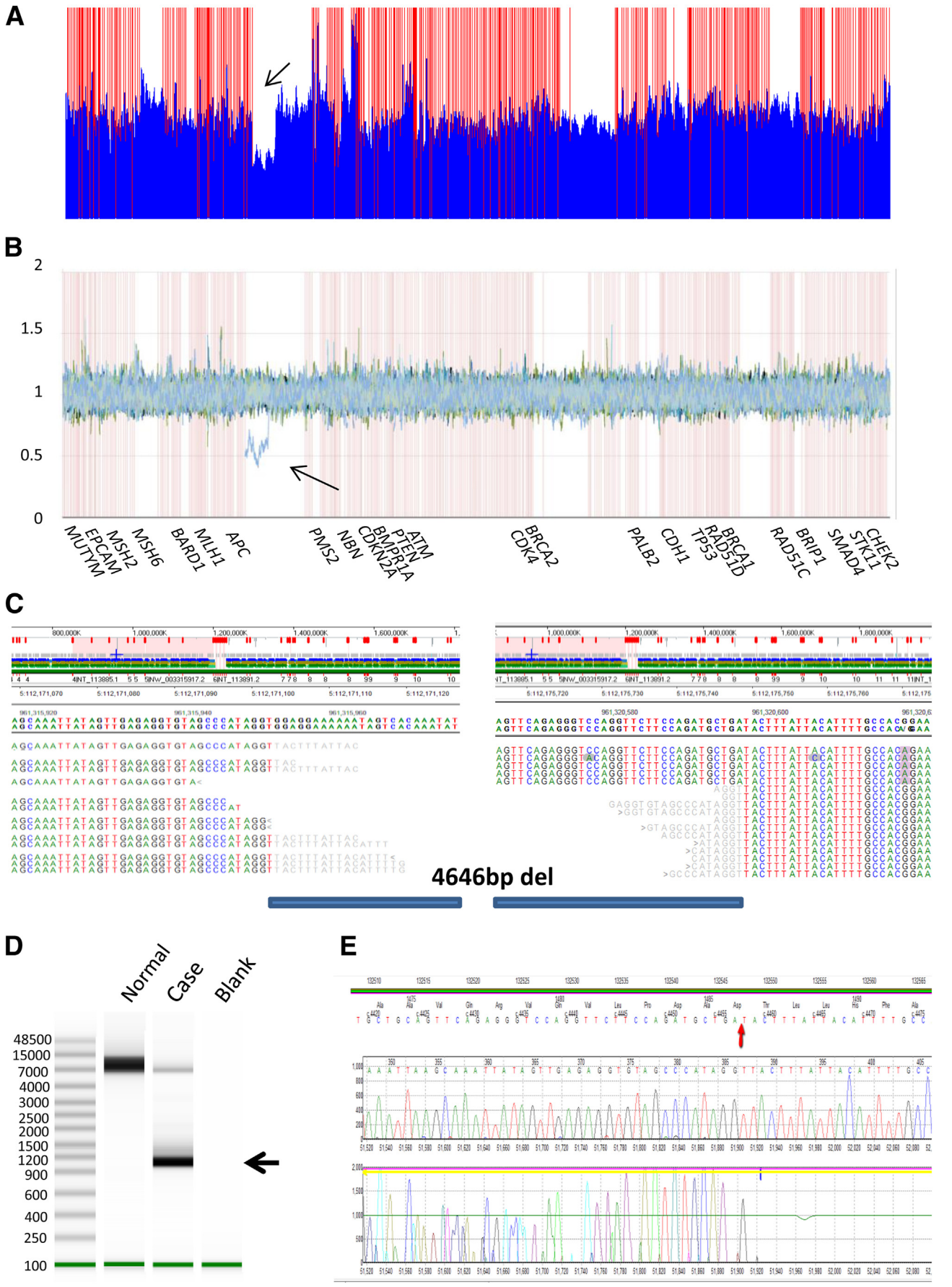
BRCA, hereditary cancer, Charcot-Marie-Tooth syndrome, hyperferritinemia, and mtDNA panels) and captured using the SeqCap EZ Choice Library system (Roche NimbleGen, Inc.). As with the sample libraries, captured libraries were assessed for quantification and size distribution to determine molarity and were diluted to a concentration of 4 nmol/L to process for sequencing, according to the manufacturer's instructions (Illumina, San Diego, CA). The final captured library concentration for sequencing was 10 pmol/L with a 1% PhiX spike-in. Libraries were sequenced using the MiSeq version 2 reagent kit to generate 2 $\times$ 150-bp paired-end reads using the MiSeq fastq generation mode (Illumina).

NGS Alignment Parameters

Sequence alignment and coverage distribution were performed with NextGene software version 2.4.1 (SoftGenetics, LLC) using standard alignment settings (allowable mismatch bases, 1; allowable ambiguous alignments, 50; seed bases, 30; move step bases, 5; allowable alignments, 100; matching base percentage, $>$85%). BAM and VCF files were imported into Geneticist Assistant software version 1.1.5 (SoftGenetics, LLC) for quality control assessment (minimum base coverage; mean region coverage).

NGS Analysis of CNVs

Reports for base coverage distribution were generated using NextGene software version 2.4.1 (SoftGenetics, LLC). Single-nucleotide coverage for each patient was normalized (see below). Briefly, the sum of all sample sequencing reads divided by the number of patients equals the total mean coverage per patient (mean of sums). The normalization factor was then calculated by dividing the sum of reads for each patient by the mean of sums. Finally, each read per nucleotide per patient was divided by the normalization factor and by the average read of each nucleotide in each of


A
B
2
1.5
1
0.5
0
APC
MLH1
BARD1
MSH6
MSH2
EPCAM
MUTYM
ATM
PTEN
BMPR1A
CDKN2A
NBN
PMS2
BRCA2
CDK4
PALB2
CDH1
BRCA1
RAD51D
TP53
BRIP1
RAD51C
CHEK2
STK11
SMAD4
C
4646bp del
D
Normal
Case
Blank
48500
15000
7000
4000
3000
2500
2000
1500
1200
900
600
400
250
100
E

the samples, resulting in the normalized reads per nucleotide per patient. The normalized data were then presented in a graph, allowing visualization of CNVs >50 bp (ie, deletions and duplications) at exon and subexon levels (Excel version 14.0.6129.5000; Microsoft Corporation, Redmond, WA). Constitutional deletions were defined by a mean ratio of $\leq$0.65 (1/2 alleles), and duplications were defined by a ratio of $\geq$1.35 (3/2 alleles). These cutoff values were determined with internal laboratory reference analysis in our retrospective cohort of specimens with no known copy number alterations, achieving between 95% and 99% specificity, depending on the specific panel (low false-positive rate). Mitochondrial deletions for heteroplasmy detection were defined by a mean ratio of $\leq$0.9 over at least 1 kbp, with cutoffs determined with internal laboratory reference analysis as described above. CNVs detected by our NGS CNV pipeline were confirmed by a second method (confirmation CNV testing of prospective cohort):

$$\frac{\sum \text{Reads}}{\#\text{Patients}} = \text{Mean of Sums}$$

$$\frac{\sum \text{Reads per individual patient}}{\text{Mean of Sums}} = \text{NF}$$

$$\frac{\text{Reads per nucleotide per patient}}{\text{NF} * \text{Average Reads per nucleotide}} = \text{NR}$$

CNV Analysis of Homologous Regions

When a panel gene contains a pseudogene (or another homologous region) that lacks any difference within a 30-bp stretch (the seeding size used during alignment), a four-allele normalization method was adapted. The coverage for the four alleles was totaled at each nucleotide position and underwent the normalization algorithm described above. Deletions were defined by a mean ratio of $\leq$0.8 (3/4 alleles), whereas duplications were defined by a ratio of $\geq$1.2 (5/4 alleles). These cutoff values were determined with internal laboratory reference analysis in our retrospective cohort of specimens with no known copy number alterations, achieving between 95% and 99% specificity. This method identifies the presence of a CNV, which then requires localization testing with a combination of MLPA SNP-specific probes and/or LR-PCR with Sanger sequencing (as described above). Regions in our targeted gene panels that were assessed with these parameters include exons 9 and 11 to 15 of *PMS2* and exons 2, 3, 6, 8, 9, 11, and 12 of *GBA*.

Assessment of Normalized Data

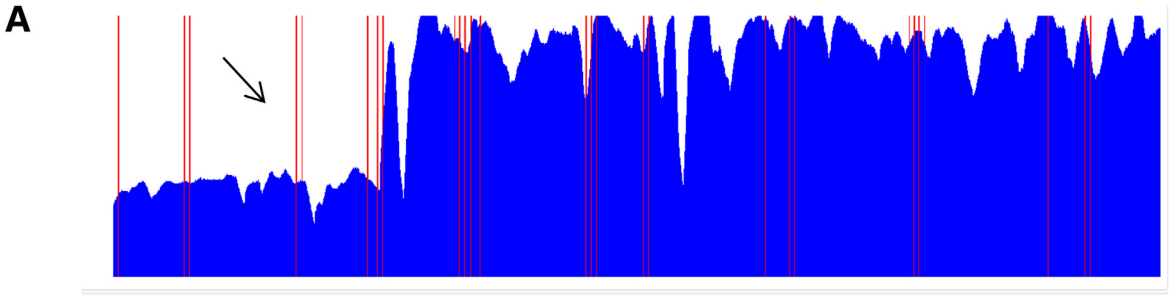
Because of some variability among batches, intrarun normalization is always performed on a minimum of 12 samples and only assessed on regions meeting a minimum RD threshold as defined in Table 2. First analysis includes all samples for normalization regardless of depth of coverage to identify homozygous deletions. The first review assesses sample quality based on intrarun normalized values falling within the defined threshold ratios (0.65 to 1.35 for constitutional, 0.8 to 1.2 for homologous regions, and <0.9 for mitochondria). Any sample with greater than two regions outside the defined boundaries is assessed as poor quality and flagged for repeat processing (or repeat sampling after two failures). Samples that met criteria for repeat analysis or fall below minimum coverage threshold without a homozygous deletion were removed from the normalized data, and visual assessment of the remaining samples was performed by the first reviewer to indicate potential positive regions for follow-up. A second reviewer then assessed the poor and good quality normalized charts, and any discrepancies underwent a third review. At completion of confirmation testing using an alternate technique, an additional reviewer verifies all data sets before final report sign-out.

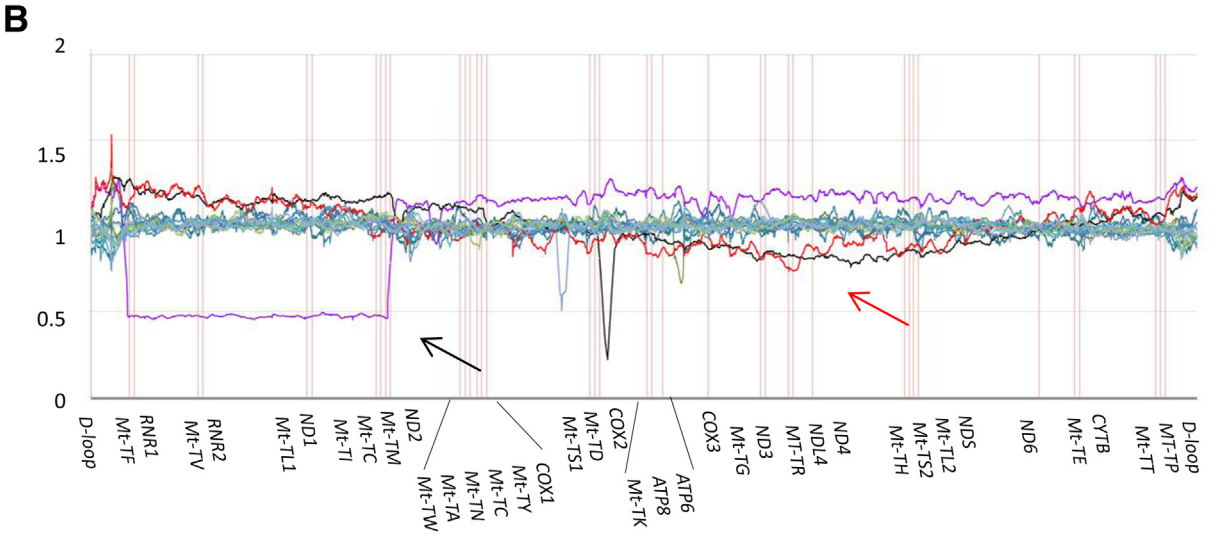
Confirmation CNV Testing of the Prospective Cohort

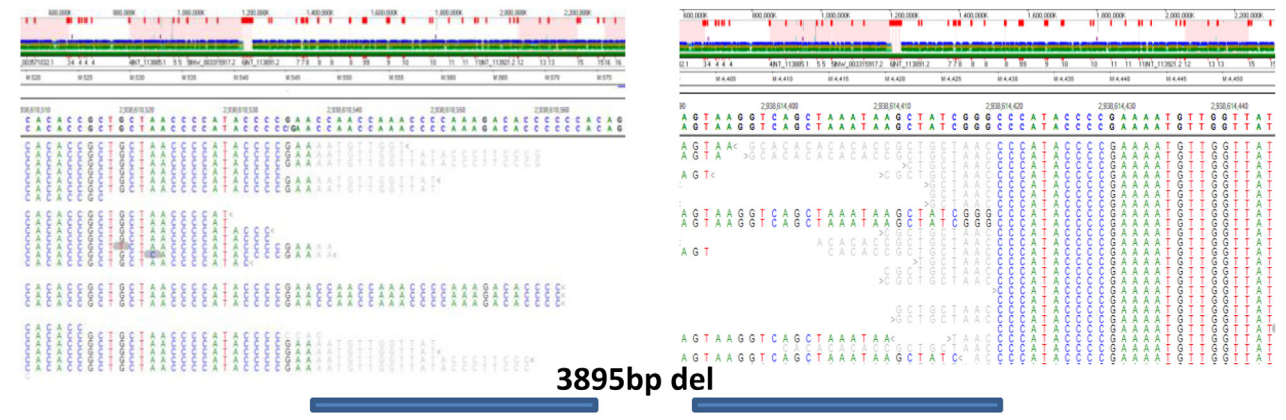
Clinically validated laboratory techniques were performed to confirm and characterize all CNVs identified in the prospective cohort. When available, MLPA was performed as described above. In the absence of MLPA, primers were designed for LR-PCR, and the region of interest was amplified with the SequalPrep Long PCR kit (Life Technologies) and separated by electrophoresis on the 2200 TapeStation (Agilent Technologies). Breakpoints were determined by Sanger sequencing, as described above (in section *Retrospective Cohort CNV Assessment*).

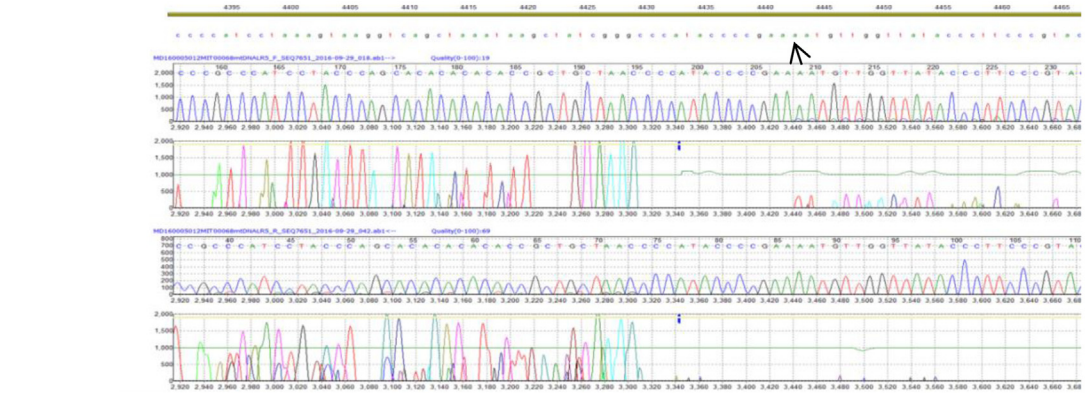
In some cases, real-time quantitative PCR was performed by designing primers for target and housekeeping genes, amplified with the LightCycler 480 High Resolution Melting Master kit, cycled with the LightCycler 480 PCR system, and analyzed with LightCycler 480 software version 1.5.1.62 (Roche Diagnostics).

Figure 2 Hereditary cancer panel: *APC* exon deletion and breakpoint determination. **A:** Sequence read depth for the hereditary cancer panel in a patient with an *APC* deletion, Lower coverage (**arrow**) suggests a large heterozygous deletion. **Red lines** indicate exon boundaries. *x* axis indicates nucleotide positions on corresponding genes and exons. *y* axis indicates depth of sequence coverage. **B:** Normalized copy number variant plot demonstrating deletion detection at *APC* gene (**arrow**). *x* axis indicates gene-exon locations. Red lines indicate exon boundaries. *y* axis represents quantile normalized copy number data (for autosomal genes, 0.5 indicates 1 copy; 1, 2 copies; and 1.5, 3 copies). Constitutional deletions are defined by a mean ratio of $\leq$0.65, and duplications were defined by a ratio of $\geq$1.35. **C:** Breakpoint detection of the *APC* gene deletion by pileup analysis using the next-generation sequencing software at the 5′ breakpoint (**left**) and 3′ breakpoint (**right**). **D:** Agarose gel electrophoresis of the long-range PCR. **Arrow** denotes the allele with deletion. **E:** Sanger sequencing electropherogram analysis of the reverse direction. The breakpoint is indicated by the **red arrow**.


A


B
2
1.5
1
0.5
0
D-loop
Mt-TF
RNR1
Mt-TV
RNR2
Mt-TL1
ND1
Mt-TI
Mt-TC
Mt-TM
ND2
Mt-TW
Mt-TA
Mt-TN
Mt-TC
Mt-TY
COX1
Mt-TS1
Mt-TD
COX2
Mt-TK
ATP8
ATP6
COX3
Mt-TG
ND3
MT-TR
NDL4
ND4
Mt-TH
Mt-TS2
Mt-TL2
ND5
ND6
Mt-TE
CYTB
MT-TP
Mt-TT
D-loop


C
3895bp del


D

For the two cases that indicated a duplication of *GJB1* on the X chromosome, the entire chromosome dosage was assessed to rule out or in the presence of an extra copy of the entire chromosome. Genomic DNA was amplified according to the manufacturer's recommendations using the Aneufast quantitative fluorescence PCR S1/S2 and MXY kits (Aneufast, Basel, Switzerland). PCR products were separated by capillary electrophoresis on an ABI 3730 (Life Technologies) and copy number alterations analyzed with GeneMapper software version 5 (Life Technologies).

Breakpoint Determination with NGS Data

For constitutional CNVs or mitochondrial deletions with heteroplasmy >25%, breakpoints were determined when the CNV plot indicated at least one breakpoint within the targeted region. The deleted allele aligns to the reference genome when the matching base percentage is >85%, and the pileup of these sequencing reads can then be inspected in NextGene software version 2.4.1 (SoftGenetics, LLC). The unmatched region is indicated on the pileup in gray and is used to search in Alamut Visual software version 2.7.2 (Interactive Biosoftware, Rouen, France) or MITOMAP (*www.mitomap.org*) to determine the location of the second breakpoint. When the second location is determined or when both breakpoints are within the targeted region, the sequencing pileup for that region is also reviewed.

NGS Heteroplasmy Analysis

The retrospective cohort contained four mitochondrial samples with heteroplasmy levels of 20% to 80% as previously determined by quantitative Southern blot analysis and were accurately detected by the NGS normalized value for the deleted copies. Mitochondrial heteroplasmy was calculated for deletions >1 Kbp as an average of one minus the mean ratio determined using the formula above.

Results

CNV Assessment in the Retrospective Cohort

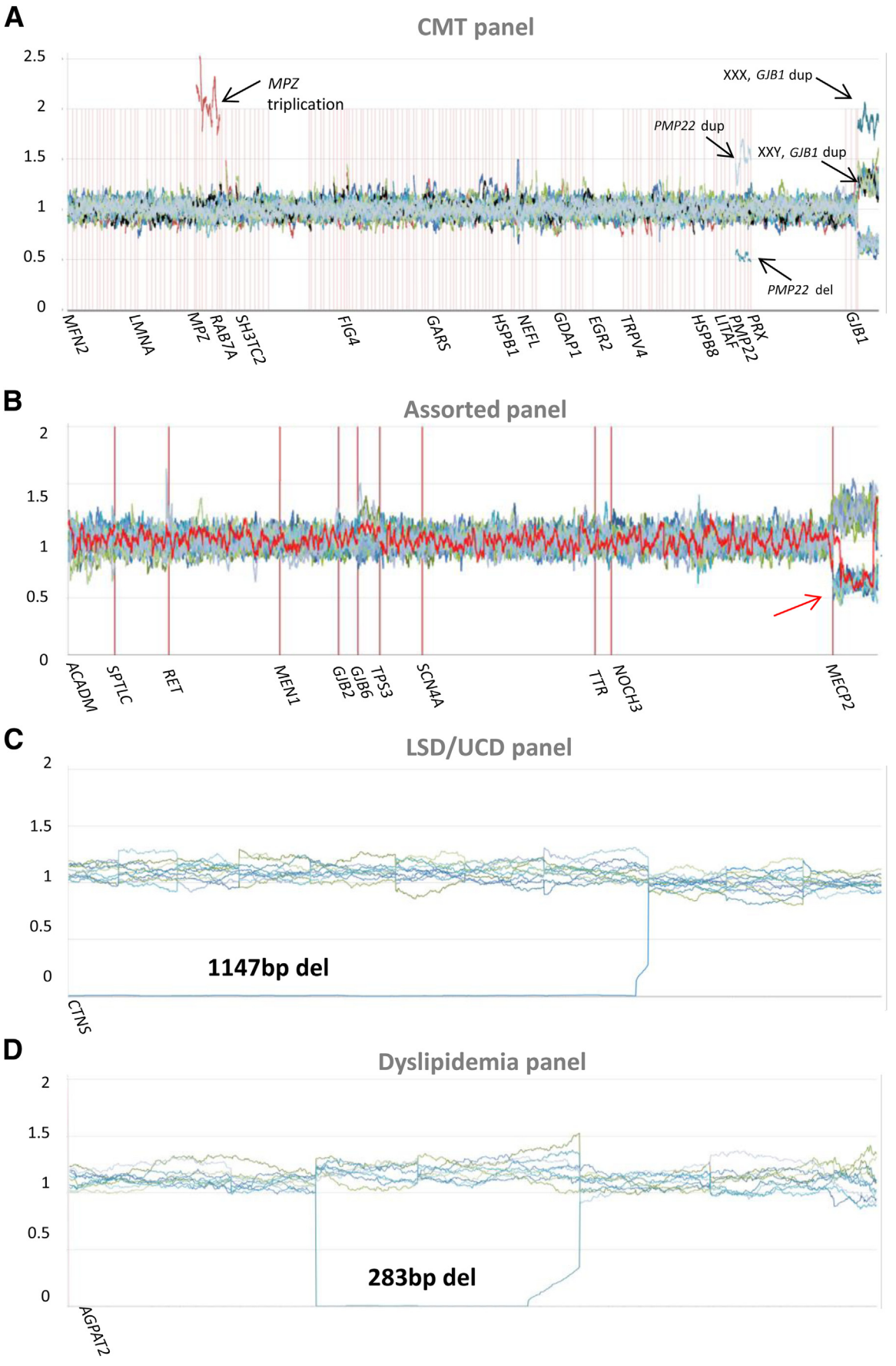
The retrospective cohort included 391 patient samples previously analyzed by Sanger sequencing and MLPA or Southern blot. Within this cohort, 43 samples (37 unique variants) were identified with CNVs by MLPA. NGS of retrospective samples were performed using one of seven gene panels that ranged in size from 16 (mtDNA) to 170 Kb (dyslipidemia panel) (Tables 2 and 3; Supplemental Figure S1). NGS CNV analysis accurately identified all 37 CNVs, demonstrating 100% sensitivity (95% CI, 89%–100%) (calculated according to the efficient-score method, corrected for continuity) for all panels analyzed, as well as identifying an additional two large deletions that had not previously been detected because of a lack of an available MLPA protocol (two dyslipidemia cases) (Tables 3 and 4). These two novel deletions identified were confirmed by LR-PCR. Of the 391 samples, 140 had also been analyzed for X chromosome dosage because of the presence of the X-linked gene on the corresponding gene panel, and NGS CNV analysis accurately correlated these samples to the expected sex. Together, NGS CNV analysis correctly assigned copy number changes and X-chromosome dosage, with no false-negative results, yielding 100% sensitivity.

CNV Assessment in the Prospective Cohort

The prospective cohort included 2375 patient samples assessed with one of the eight clinical NGS panels, ranging in size from 15 (mtDNA) to 220 Kb (epilepsy) (Tables 2 and 5), as part of routine clinical testing. Five of these panels, involving 1132 patient samples, included a gene on the X chromosome, which allowed for an internal control of chromosome copy number assessment. Of the 1132 samples, 1130 accurately matched the expected sex of the patient, with the discordant samples indicating XX and XXX genotypes in a male and a female patient, respectively. Follow-up testing with quantitative fluorescence PCR confirmed the NGS findings of an XXY male and XXX female, indicating 100% accuracy in detecting X chromosome dosage.

A total of 130 patients had a potential CNV identified by the NGS CNV tool (Table 5). Potential CNVs underwent follow-up analysis with MLPA, Sanger, real-time quantitative PCR, LR-PCR, or some combination of each. Follow-up analysis confirmed 106 of the 130 potential CNVs, resulting in 24 false-positive cases and an overall false-positive rate of 1.0% (24/2375). The false-positive rate of each panel can be seen in Table 4 and ranged from 0% in the mtDNA panel to 2.6% in the hyperferritinemia panel.

Figure 3 Mitochondrial DNA (mtDNA) panel: detection of uncommon 5′ deletions and low heteroplamy deletions. **A:** Sequence read depth for mtDNA screen in a patient with a large 5′ mitochondrial deletion (**arrow**). Red lines indicate exon boundaries. *x* axis indicates nucleotide positions on corresponding genes and exons. *y* axis indicates depth of sequence coverage. **B:** Normalized copy number variant plot demonstrating large 5′ deletion (m.548_4442del) of 55% heteroplasmy outside the common mitochondrial deletion syndromes region (**black arrow**). The other two samples show deletions of approximately 15% heteroplasmy in the common deletion region associated with Kearns-Sayre syndrome (**red arrow**). *x* axis indicates gene-exon locations. **Red lines** indicate exon boundaries. *y* axis represents quantile normalized copy number data (for mitochondrial genome, 0.25 indicates 75% heteroplasmy; 0.5, 50% heteroplasmy; 0.75, 25% heteroplasmy; and 1, homoplasmy). Mitochondrial deletions for heteroplasmy detection were defined by a mean ratio of ≤0.9 over at least 1 Kbp. **C:** Breakpoint detection of the large 5′ mitochondrial deletion by pileup analysis using the next-generation sequencing software at the 5′ breakpoint (**left**) and 3′ breakpoint (**right**). **D:** Sanger sequencing electropherogram analysis showing a preferential amplification of the deleted allele. Arrow indicates breakpoint.


A
CMT panel
2.5
2
1.5
1
0.5
0
MPZ
triplication
XXX, GJB1 dup
PMP22 dup
XXY, GJB1 dup
PMP22 del
MFN2
LMNA
MPZ
RAB7A
SH3TC2
FIG4
GARS
HSPB1
NEFL
GDAP1
EGR2
TRPV4
HSPB8
LITAF
PMP22
PRX
GJB1
B
Assorted panel
2
1.5
1
0.5
0
ACADM
SPTLC
RET
MEN1
GJB2
GJB6
TPS3
SCN4A
TTR
NOCH3
MECP2
C
LSD/UCD panel
2
1.5
1
0.5
0
1147bp del
CTNS
D
Dyslipidemia panel
2
1.5
1
0.5
0
283bp del
AGPAT2

Low-quality DNA samples having too much noise in the CNV copy plot or called with multiple suspicious exon duplications or deletions underwent a repeat NGS analysis. From our entire combined (retrospective and prospective) cohort, 0.9% (26/2766) of the cases were evaluated as needing repeat processing. The results from a repeat analysis revealed a complete resolution of the noise or a repetitive pattern of noise in which case a repeat sampling of the patient was requested (2% to 5% of the cases with noise, depending on the panel).

Assessment of Homologous Regions

Eight of the CNVs identified by NGS in the retrospective and prospective cohorts were confirmed as true *PMS2* or *PMS2CL* gene deletions (Tables 3 and 5), demonstrating the ability of our CNV algorithm to incorporate pseudogene analysis to allow for accurate CNV identification. Figure 1A shows five samples (four retrospective and one prospective) identified with *PMS2* deletions; three *PMS2* CNVs were identified in the 5′ region of the gene with a ratio of 0.5, whereas the other two were in the region of high homology with *PMS2CL* and were identified by a deletion ratio of 0.75 attributable to the normalization algorithm being applied to four alleles (Figure 1B). An additional 10 false-positive *PMS2* CNVs were detected by NGS, resulting in a false-positive rate for the initial NGS assessment for our hereditary cancer panel of 1.7% (10/576). Nine of these false-positive samples were the result of overcautiously assessing the 3′ region of *PMS2* for CNV ratios of 0.8 to 1.2, which is used to avoid false-negative results. However, all CNVs detected by NGS undergo confirmation testing by MLPA to provide a final and accurate result.

In addition to the homologous *PMS2* gene, a patient with the CNV in the *CHEK2* gene was also detected. Exons 11 to 15 of *CHEK2* have 96.8% homology with the pseudogene *CHEK2P2*. Our NGS-CNV pipeline accurately detected a deletion of exons 14 and 15 (data not shown).

Breakpoint Determination

In samples in which the deletion proved to map within the exons, the significant sequence depth in these targeted NGS panels allowed for an accurate breakpoint determination. For example, a sample with an *APC*:c.1958+241_4457del variant was found to have a loss of one of the two copies based on the low panel read coverage (Figure 2A) and CNV analysis (Figure 2B), a finding that was confirmed by MLPA analysis. A closer look at the NGS sequence reads at the approximate breakpoint area allowed us to precisely identify the true breakpoint (Figure 2C) and facilitated a precise sequence primer design to allow confirmation by LR-PCR (Figure 2D) and Sanger resequencing (Figure 2E).

Heteroplasmy Analysis

By using RD to determine CNVs, we were able to expand the application to precisely determine the level of heteroplasmy in the mitochondrial genome. Retrospective cohort analysis confirmed four mitochondrial deletions ranging from 18% to 80% heteroplasmy (Table 3), whereas the prospective analysis identified seven samples with CNVs of 10% to 98% heteroplasmy (Table 5). Figure 3 shows a characterization of three of the prospective samples, including a large 5′ deletion (m.548_4442del) outside the common mitochondrial DNA deletion syndromes region. This deletion is easily detected by reviewing the panel sequence depth coverage (Figure 3A) and by NGS CNV analysis (Figure 3B). Figure 3B also shows two samples with approximately 15% heteroplasmy for the common mitochondrial deletion that is frequently associated with Kearns-Sayre syndrome. By assessing the sequence reads near the approximate CNV boundaries in a similar manner applied to the nuclear gene panels, definitive breakpoints (Figure 3C) used for Sanger sequencing or LR-PCR confirmation (Figure 3D) can be identified.

Rare Dosage Variants

NGS CNV analysis in the prospective cohort identified a number of constitutional CNVs outside the more commonly found heterozygous deletions and duplications genotype with ideal normalized ratios of 0.5 and 1.5, respectively. Gene triplication was observed at the autosomal *MPZ* gene, as indicated by a 2.0 copy number ratio (Figure 4A). Copy number analysis of the X-linked *GJB1* gene shows a potential duplication in both a female (1.75 ratio instead of 1.25) and male patient (1.25 ratio instead 0.75). X chromosome dosage analysis confirms the *GJB1* duplications represent an XXX female and an XXY male, respectively. Assessment of another X-linked gene, *MECP2*, shows identification of a partial gene deletion (c.27-4474_1188del) when reviewed in combination with the expected sex of the patient (Figure 4B). In addition, Figure 4, C and D shows patients with homozygous deletions in the *CTNS* gene on the lysosomal storage/urea cycle disorder panel (c.(?_-21)

Figure 4 **A:** Charcot-Marie-Tooth syndrome (CMT) panel: normalized copy number variant plot demonstrating a *MPZ* gene triplication and *PMP22* gene duplication and deletion (**arrows**). Copy number analysis of the *GJB1* gene in an XXX female (1.75 ratio instead of 1.25) and in an XXY male (1.25 ratio instead 0.75; sample in black) (**arrows**). No Y-chromosome genes are present on this panel. **B:** Assorted panel: identification of the *MECP2*:c.27-4474_1188del partial gene deletion in a female patient (**red arrow**, patient in red). **C:** Lysosomal storage/urea cycle disorder (LSD/UCD) panel: identification of a homozygous *CTNS* deletion that is a subsection of the larger 57-Kb deletion (sample in blue). **D:** Dyslipidemia panel: identification of homozygous deletion, *AGPAT2*:c.366_588+534del, in a patient (sample in blue). *x* axis indicates gene-exon locations. **Red lines** indicate exon boundaries. *y* axis represents quantile normalized copy number data (for autosomal genes, 0 indicates no copies; 0.5, 1 copy; 1, 2 copies; and 1.5, 3 copies). Constitutional deletions were defined by a mean ratio of $\leq$0.65, and duplications were defined by a ratio of $\geq$1.35.

_847del) and in the *AGPAT2* gene on the dyslipidemia panel (c.366_588+534del).

Discussion

NGS is increasingly being used in the molecular diagnosis of constitutional disorders, a significant proportion of which are driven by structural variations and CNVs across the genome. An optimized and validated assay design is crucial for accurate detection of such molecular causes to facilitate accurate diagnosis and subsequent decision making in a clinical setting. In this article, clinical validation of NGS data is defined as the ability of a targeted NGS gene panel assay to sensitively reproduce data generated by the gold standard of Sanger and MLPA analysis on clinical specimens (retrospective cohort) and the ability of clinically validated Sanger and MLPA analysis to reproduce NGS-derived data using our custom targeted gene panels (prospective cohort). In the present study, we report the development and validation of a CNV assessment tool for targeted NGS gene panel data as a replacement of the traditional methods for clinical detection of CNVs and demonstrate the efficacy of this tool using nine distinct clinical NGS panels. Using the retrospective cohort of patients with known CNVs, we have demonstrated 100% accuracy and sensitivity (95% CI, 89%−100%) of the NGS pipeline for CNV detection. Furthermore, two additional CNVs were identified that were not detected by the classical methods. Assessment of the prospective cohort also demonstrated high positive predictive values and a low false-positive rate. However, all molecular assays for CNV detection, including the standard techniques of MLPA and copy number arrays, suffer from the possibility of false-negative results. Our method demonstrates at least equal sensitivity to these established techniques. Furthermore, it is not limited by the requirement of the targeted probe placement across the CNV region for sensitive detection, which may decrease the specificity of the probe-based assays, such as MLPA or microarrays. Our method analyzes NGS reads of 150 bp in size and requires 85% matching for accurate alignment. The reliability of this alignment is strengthened by the RD in which a minimum of 200 reads are required. With these analysis parameters, we report evidence that some small CNVs that are too large to be detected by the sequencing algorithm's indel detector can result in the loss of the mapped reads and the consequent copy number alteration on the CNV plot. CNV alterations can also present as a result of read misalignment attributable to technical issues, such as read size, nearby variants, insertions, inversions, or other complex sequence changes. These most commonly present as deletions by NGS CNV analysis and therefore always require confirmatory assessment by another validated laboratory technique, such as Sanger sequencing or MLPA analysis. As an example, Supplemental Figure S2 shows a CNV plot demonstrating evidence of a small deletion at a <50% level, as well as an apparent alignment artifact on the NGS pileup in that same region, which was subsequently demonstrated to be a 78-bp in-frame deletion in the *PRX* gene. We have recently demonstrated, albeit in a smaller sample size, a 100% sensitivity on the BRCA targeted panel compared with the parallel MLPA analysis.[12] However, no molecular technique, including our approach, can claim to have 100% sensitivity to all possible structural rearrangements, including rare balanced rearrangements that may not be detectable by standard molecular techniques, such as MLPA, microarrays, or NGS CNV algorithms.

To date, at least a dozen CNV prediction algorithms for NGS-generated RD have been made available.[18−23] Most of these algorithms use complex computational methods for CNV prediction and annotation for both whole genome and whole exome sequencing reads.[17] The outcomes from these algorithms are not always consistent because they assume different parameters and statistical reasoning in the prediction. Therefore, their potential implementation in a clinical setting is limited, and most clinical CNV detection panels still rely on traditional methods. Applying a novel NGS CNV algorithm in a clinical setting requires rigorous validation that includes a large number of previously identified clinical specimens that are positive for relatively rare copy number alterations. Another limitation relates to the use of these algorithms in assessment of genomic, comparatively low-sequence-depth approaches, with inherently higher intersample variability, resulting in high-level false-positive and false-negative rates, not amenable for routine clinical service. Therefore, targeted panel high-sequence-depth approaches, similar to this random fragmentation sequence capture−based deep sequencing approach, are well suited for implementation of NGS CNV analysis in a clinical setting. This study and other studies[13,24,25] have demonstrated the clinical use of an NGS CNV method capable of replacing the conventional methods. One advantage of our algorithm is its simplicity, which allows it to be implemented by using a basic statistical package, such as Microsoft Excel. The key requirement enabling this sensitive CNV detection is deep and uniform sequence coverage (Table 2, Supplemental Figure S3), together with the ability to multiplex a large number of patients (12 to 24) per single NGS run, which can then be used as references in an intrarun fashion for normalization. We have recently described clinical validation of this targeted sequencing pipeline, demonstrating sequencing uniformity, depth, and quality using the genomic fragmentation and target panel enrichment approach amenable to this CNV analysis.[12]

Another advantage of using NGS to detect CNV is the ability to avoid allele dropout, commonly associated with other PCR-based methods in which interference with PCR primer binding over DNA polymorphisms may produce a false-positive result attributable to loss of adequate amplification of the affected allele. Alternatively, a deletion may not be detected by an MLPA assay if it does not directly

overlap one of the MLPA probes, which can cause a false-negative result. In addition, a sequence (ie, SNP) variation at or near the MLPA probes ligation site may generate a false-positive result. The power of our targeted NGS capture method relies on the highly overlapping/tiled probe capture design (up to 2.1 million probes available for each panel) in combination with highly redundant and random genomic fragmentation. This unique design is able to avoid such biases and results in highly uniform sequence coverage in an intrasample and intersample fashion, which enables the clinical use of this CNV algorithm. However, use of this algorithm in the more variable NGS data sets (low coverage, high intersample variability) may not be reliable. In addition to being able to detect CNVs, high and uniform sequence coverage along with the high level of redundancy and staggering of the individual NGS reads allows precise mapping of chromosomal breakpoints, when a CNV is localized within the target captured sequence. This in turn facilitates primer design for confirmation/follow-up testing and significantly reduces the workload and turnaround times associated with confirmatory testing and reporting of these variants.

Mitochondrial deletion syndromes are a broad spectrum of clinical symptoms that occur because of deletions in the mitochondrial genome.[26] The traditional method to screen for common mitochondrial deletions is Southern blot and/or LR-PCR; however, the RD (minimum of 1000×) and uniformity of the sequence coverage within an NGS run enable a precise estimation of the degree of mitochondrial CNV heteroplasmy to approximately 5% to 10%, meeting the minimum required heteroplasmic rate for clinical interpretation.[27] In this study, we demonstrated the validity of our algorithm to sensitively detect 18% heteroplasmy in the retrospective cohort as confirmed by previous Southern blot analysis and translated this to the prospective cohort with detection levels of 10% heteroplasmy. Our NGS-based CNV detection method is able to identify rare mitochondrial CNVs in addition to the more common 5 Kb (95% CI, 8.5−13.5 Kb) mitochondrial deletion. Indeed, our NGS method has identified a patient with a large mitochondrial deletion that is situated outside the region of the common deletion that is usually tested for by Southern blot (Figure 3).

Pseudogenes are a challenge in CNV and sequence assessment from NGS data and by classic approaches. For example, our hereditary cancer panel includes the *PMS2* gene that has a pseudogene *PMS2CL* with significant sequence homology, including nearly identical sequence in exons 9 and 11 to 15.[28] In our retrospective cohort, we have confirmed the ability to accurately detect four different CNVs within the *PMS2* gene. The 5′ region of the gene has enough variability enabling the sequence alignment algorithm to accurately map the reads to the gene or pseudogene. However, a lack of sequence variation in the 3′ region, combined with conversion events between *PMS2* and *PMS2CL,* presents a challenge in accurate assignment of the NGS reads. One way to address this issue is to add the total sequence depth for each corresponding nucleotide in exons 9 and 11 to 15 of *PMS2* and *PMS2CL* and apply the normalization algorithm as described above. Applying the cutoffs of 0.8 and 1.2 in this region has allowed us to detect the imbalance that results from the presence of three or five compared with the normal total of four alleles, although follow-up MLPA and Sanger sequencing are still needed to determine whether the CNV or sequence variants are located in the gene or pseudogene.

The work described above builds on our previously reported application of this pipeline to the hereditary breast cancer panel[12] and demonstrates the ability of this pipeline to be applied more broadly for CNV detection on a number of distinct gene panels, although some of the gene panels with no previous clinical data, such as epilepsy and hyperferritinemia, require further empirical assessment. Implementation of NGS technologies into clinical molecular diagnostics requires extensive validation and comparisons with conventional methods to assess the consistency, reliability, sensitivity, specificity, turnaround time improvement (including service contracts and institutional overhead), and per-sample costs. For example, using the simplest BRCA two-gene targeted panel with 24 patients per MiSeq run compared with the classic Sanger sequencing and MLPA testing, approximately 70% reduction in overall total test costs for reagents, labor, and significant reduction in turnaround time can be achieved. Cost-effectiveness can be enhanced further through increased automation and scalability (100 patients per run instead of 24) and use of comprehensive gene panels (eg, hereditary cancer panel), enabling significant improvements in efficiency of delivery of patient care and genetic services with a net positive financial effect on the health care system. In the present study, we have retrospectively and prospectively assessed data from nearly 3000 patient samples, describing the design and analytical validation of a high-throughput NGS pipeline that enables simultaneous screening of CNVs in nine distinct clinical NGS panels. Our data reveal high sensitivity and specificity, which, combined with the economic advantages, provide an approach that outperforms conventional methods that involve parallel CNV detection methods.

Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.jmoldx.2017.07.004*.

References

1. Wang Y, Yao X, Li SN, Suo AL, Tian T, Ruan ZP, Guo H, Yao Y: Detection of prostate cancer related copy number variations with SNP genotyping array. Eur Rev Med Pharmacol Sci 2013, 17:2916−2922
2. Shlien A, Malkin D: Copy number variations and cancer. Genome Med 2009, 1:62

3. Pfarr N, Penzel R, Klauschen F, Heim D, Brandt R, Kazdal D, Jesinghaus M, Herpel E, Schirmacher P, Warth A, Weichert W, Endris V, Stenzinger A: Copy number changes of clinically actionable genes in melanoma, non-small cell lung cancer and colorectal cancer-A survey across 822 routine diagnostic cases. Genes Chromosomes Cancer 2016, 55:821−833
4. Butchbach ME: Copy number variations in the survival motor neuron genes: implications for spinal muscular atrophy and other neurodegenerative diseases. Front Mol Biosci 2016, 3:7
5. Grayton HM, Fernandes C, Rujescu D, Collier DA: Copy number variations in neurodevelopmental disorders. Prog Neurobiol 2012, 99:81−91
6. Freeman JL, Perry GH, Feuk L, Redon R, McCarroll SA, Altshuler DM, Aburatani H, Jones KW, Tyler-Smith C, Hurles ME, Carter NP, Scherer SW, Lee C: Copy number variation: new insights in genome diversity. Genome Res 2006, 16:949−961
7. Buysse K, Delle Chiaie B, Van Coster R, Loeys B, De Paepe A, Mortier G, Speleman F, Menten B: Challenges for CNV interpretation in clinical molecular karyotyping: lessons learned from a 1001 sample experience. Eur J Med Genet 2009, 52:398−403
8. Shen Y, Wu BL: Designing a simple multiplex ligation-dependent probe amplification (MLPA) assay for rapid detection of copy number variants in the genome. J Genet Genomics 2009, 36:257−265
9. Carter NP: Methods and strategies for analyzing copy number variation using DNA microarrays. Nat Genet 2007, 39(7 Suppl):S16−S21
10. Szuhai K, Vermeer M: Microarray techniques to analyze copy-number alterations in genomic DNA: array comparative genomic hybridization and single-nucleotide polymorphism array. J Invest Dermatol 2015, 135:e37
11. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G: Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. Nucleic Acids Res 2002, 30:e57
12. Schenkel LC, Kerkhof J, Stuart A, Reilly J, Eng B, Woodside C, Levstik A, Howlett CJ, Rupar AC, Knoll JH, Ainsworth P, Waye JS, Sadikovic B: Clinical next-generation sequencing pipeline outperforms a combined approach using Sanger sequencing and multiplex ligation-dependent probe amplification in targeted gene panel analysis. J Mol Diagn 2016, 18:657−667
13. Nord AS, Lee M, King MC, Walsh T: Accurate and exact CNV identification from targeted high-throughput sequence data. BMC Genomics 2011, 12:184
14. Cottrell CE, Al-Kateb H, Bredemeyer AJ, Duncavage EJ, Spencer DH, Abel HJ, Lockwood CM, Hagemann IS, O'Guin SM, Burcea LC, Sawyer CS, Oschwald DM, Stratman JL, Sher DA, Johnson MR, Brown JT, Cliften PF, George B, McIntosh LD, Shrivastava S, Nguyen TT, Payton JE, Watson MA, Crosby SD, Head RD, Mitra RD, Nagarajan R, Kulkarni S, Seibert K, Virgin HW 4th, Milbrandt J, Pfeifer JD: Validation of a next-generation sequencing assay for clinical molecular oncology. J Mol Diagn 2014, 16:89−105
15. Alkan C, Coe BP, Eichler EE: Genome structural variation discovery and genotyping. Nat Rev Genet 2011, 12:363−376
16. Meyerson M, Gabriel S, Getz G: Advances in understanding cancer genomes through second-generation sequencing. Nat Rev Genet 2010, 11:685−696
17. Zhao M, Wang Q, Jia P, Zhao Z: Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 2013, 14(Suppl 11):S1
18. Tan R, Wang Y, Kleinstein SE, Liu Y, Zhu X, Guo H, Jiang Q, Allen AS, Zhu M: An evaluation of copy number variation detection tools from whole-exome sequencing data. Hum Mutat 2014, 35:899−907
19. Fromer M, Moran JL, Chambert K, Banks E, Bergen SE, Ruderfer DM, Handsaker RE, McCarroll SA, O'Donovan MC, Owen MJ, Kirov G, Sullivan PF, Hultman CM, Sklar P, Purcell SM: Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. Am J Hum Genet 2012, 91:597−607
20. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, Quinlan AR, Nickerson DA, Eichler EE: Copy number variation detection and genotyping from exome sequence data. Genome Res 2012, 22:1525−1532
21. Plagnol V, Curtis J, Epstein M, Mok KY, Stebbings E, Grigoriadou S, Wood NW, Hambleton S, Burns SO, Thrasher AJ, Kumararatne D, Doffinger R, Nejentsev S: A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics 2012, 28:2747−2754
22. Li J, Lupat R, Amarasinghe KC, Thompson ER, Doyle MA, Ryland GL, Tothill RW, Halgamuge SK, Campbell IG, Gorringe KL: CONTRA: copy number analysis for targeted resequencing. Bioinformatics 2012, 28:1307−1313
23. Johansson LF, van Dijk F, de Boer EN, van Dijk-Bos KK, Jongbloed JD, van der Hout AH, Westers H, Sinke RJ, Swertz MA, Sijmons RH, Sikkema-Raddatz B: CoNVaDING: single exon variation detection in targeted NGS data. Hum Mutat 2016, 37:457−464
24. Judkins T, Leclair B, Bowles K, Gutin N, Trost J, McCulloch J, Bhatnagar S, Murray A, Craft J, Wardell B, Bastian M, Mitchell J, Chen J, Tran T, Williams D, Potter J, Jammulapati S, Perry M, Morris B, Roa B, Timms K: Development and analytical validation of a 25-gene next generation sequencing panel that includes the BRCA1 and BRCA2 genes to assess hereditary cancer risk. BMC Cancer 2015, 15:215
25. Lincoln SE, Kobayashi Y, Anderson MJ, Yang S, Desmond AJ, Mills MA, Nilsen GB, Jacobs KB, Monzon FA, Kurian AW, Ford JM, Ellisen LW: A systematic comparison of traditional and multigene panel testing for hereditary breast and ovarian cancer genes in more than 1000 patients. J Mol Diagn 2015, 17:533−544
26. Sadikovic B, Wang J, El-Hattab A, Landsverk M, Douglas G, Brundage EK, Craigen WJ, Schmitt ES, Wong LJ: Sequence homology at the breakpoint and clinical phenotype of mitochondrial DNA deletion syndromes. PLoS One 2010, 5:e15687
27. Tuppen HA, Blakely EL, Turnbull DM, Taylor RW: Mitochondrial DNA mutations and human disease. Biochim Biophys Acta 2010, 1797:113−128
28. Li J, Dai H, Feng Y, Tang J, Chen S, Tian X, Gorman E, Schmitt ES, Hansen TA, Wang J, Plon SE, Zhang VW, Wong LJ: A comprehensive strategy for accurate mutation detection of the highly homologous PMS2. J Mol Diagn 2015, 17:545−553