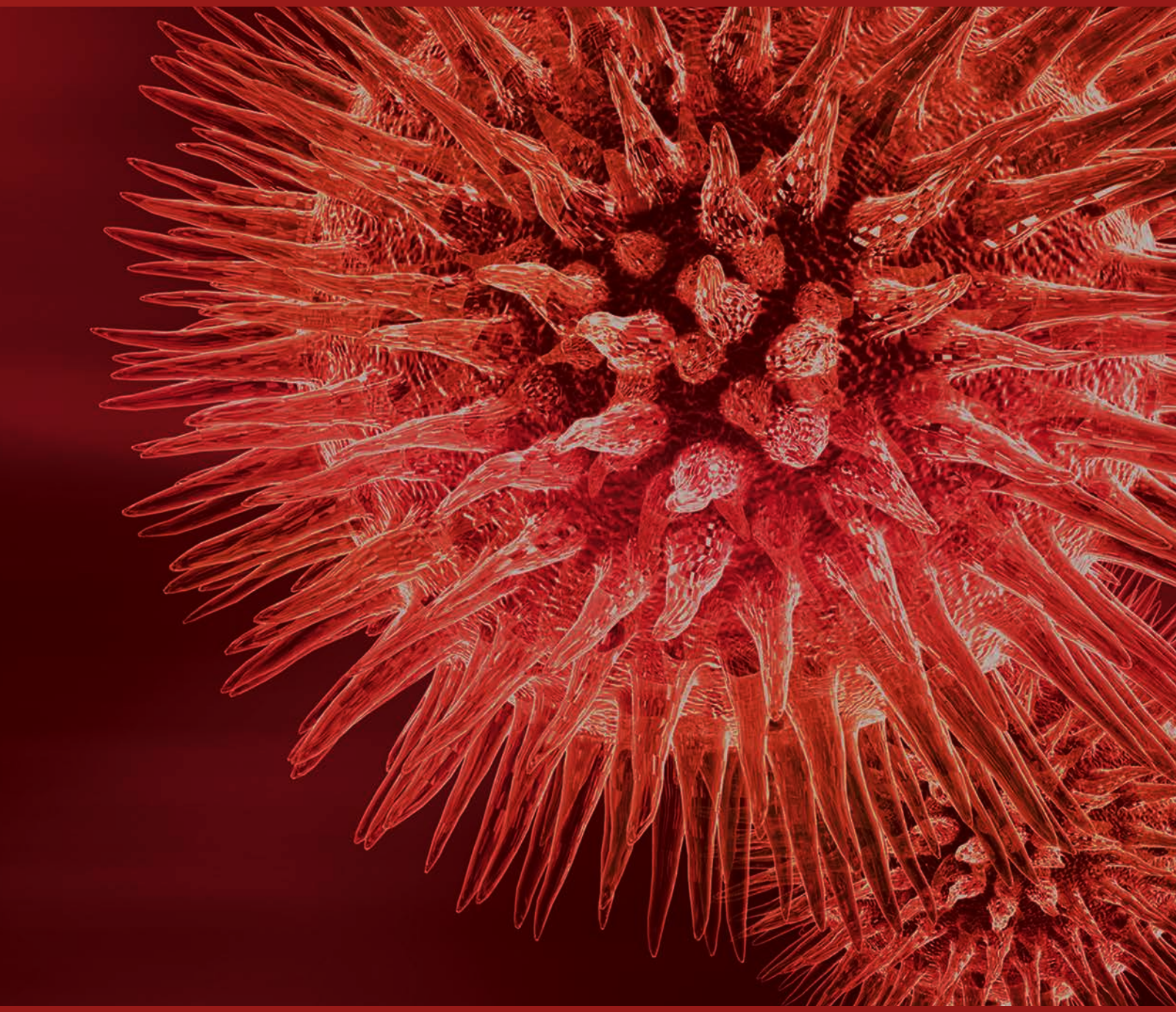
BioMed Research International

Laboratory Genetic Testing in Clinical Practice 2016

Guest Editors: Ozgur Cogulu, Jacqueline Schoumans, Gokce Toruner, Urszula Demkow, Emin Karaca, and Asude Alpman Durmaz



Laboratory Genetic Testing in Clinical Practice 2016

BioMed Research International

Laboratory Genetic Testing in Clinical Practice 2016

Guest Editors: Ozgur Cogulu, Jacqueline Schoumans, Gokce Toruner, Urszula Demkow, Emin Karaca, and Asude Alpman Durmaz

Contents

Hindawi Publishing Corporation
BioMed Research International
Volume 2017, Article ID 5798714, 2 pages
http://dx.doi.org/10.1155/2017/5798714

Editorial

Laboratory Genetic Testing in Clinical Practice 2016

Ozgur Cogulu,[1] Jacqueline Schoumans,[2] Gokce Toruner,[3] Urszula Demkow,[4] Emin Karaca,[1] and Asude Alpman Durmaz[1]

[1] *Department of Medical Genetics, Faculty of Medicine, Ege University, 35100 Izmir, Turkey*
[2] *Cancer Cytogenetic Unit, Department of Medical Genetics, Lausanne University Hospital, 1011 Lausanne, Switzerland*
[3] *Institute of Genomic Medicine, UMDNJ-NJ Medical School, Newark, NJ 07103, USA*
[4] *Department of Laboratory Diagnostics and Clinical Immunology, Medical University of Warsaw, 00-576 Warsaw, Poland*

Correspondence should be addressed to Ozgur Cogulu; ozgur.cogulu@ege.edu.tr

Received 21 November 2016; Accepted 22 November 2016; Published 4 January 2017

The genetic testing in various clinical conditions emerges to have an important role in both diagnosis and treatment. Recently, a revolution in genomic technologies from the first-generation Sanger method to next-generation high throughput sequencing and microarrays has occurred. All of these technologies have been widely applied for genome, exome, transcriptome sequencing, and epigenomics. The conclusions from basic research resulted in the development of new protocols with great potential for clinical application. Selected examples of their clinical use include breakthroughs in prenatal screening, identification of rare genetic variants associated with monogenic Mendelian disorders, and efficient detection of either inherited or somatic mutations in cancer genes. This special issue is addressing laboratory genetic testing in practice, moving beyond classical concept of patient approach. Personalized medicine, which may provide accurate and effective treatment option in most of the human diseases in the future, will offer the promise of altering conventional medicine. Diagnosis, clinical findings, and treatment options vary in every individual. Thus, it is important to clarify wide spectrum of clinical and laboratory findings in examined patients. In this special issue, researchers present various cases and studies emphasizing the importance of clinical, laboratory, and genetic findings which will be beneficial in clinical practice. Among them there is ischemic stroke (IS) which is one of the important causes of morbidity and mortality. The authors showed that Vitamin K epoxide reductase complex subunit 1- (VKORC1-) 1639A allele can be a possible genetic risk factor for IS in Ukrainian population. On the other hand precision medicine mentioned above is one of the current models, which tries to explain the genetic indicators to improve the quality of medical care in recent days. Another paper points to molecular diagnostics for precision medicine in colorectal cancer (CRC). The authors discuss the future perspectives of CRC heterogeneity associated with anti-EGFR resistance and immune checkpoint blockage therapy. Authors of another paper find NGS data analysis problematic when the differentiation of the indel errors and false positive mutations is needed. The authors propose new software AGSA helping to detect false positive mutations in homopolymeric sequences at lower costs and in shorted time while increasing reliability, notably for homopolymer tracts. Tissue sampling and microdissection are very important steps in the molecular genetic analysis of cancer samples. This was described on the basis of microdissection strategy in pulmonary tumor. The authors demonstrate the importance of microdissection in morphologically different tumor components for pyrosequencing in KRAS and BRAF mutations. Despite rapid technological advances in medicine, essential hypertension (EH) etiology remains largely unknown. We publish the first research assessing the atrial natriuretic peptide gene polymorphisms in both EH and type 2 diabetes mellitus patients among Malaysian population. Copy number variations (CNVs) have attracted increasing attention as cancer susceptibility regulators. Further paper describes CNV-67048 of WW domain-containing oxidoreductase which is shown to be a risk factor of epithelial ovarian cancer in Chinese population. CNV based microarray is an important

technology which provides to investigate the CNV of whole genome. The detection rate and pathogenic yield of chromosomal microarray analysis (CMA) in neurodevelopmental disorders are described. The authors established the detection rate and pathogenic yield of CMA as depending on the primary indications for testing, the age of the individuals tested, and the specialty of the ordering doctor. Targeted sequencing is the method of choice for examining genes in specific pathways or in genetic heterogeneity which is responsible for many systemic disorders. Further article describes targeted sequencing in inherited retinal dystrophies (IRD). The study demonstrates that NGS represents a comprehensive cost-effective approach for IRDs molecular diagnosis. The identification of the genetic alterations underlying the phenotype enabled the clinicians to achieve a more accurate diagnosis. The results emphasize the importance of molecular diagnosis coupled with clinic information to unravel the extensive phenotypic heterogeneity of these diseases.

Ozgur Cogulu
Jacqueline Schoumans
Gokce Toruner
Urszula Demkow
Emin Karaca
Asude Alpman Durmaz

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 6341870, 14 pages
http://dx.doi.org/10.1155/2016/6341870

Research Article

Unravelling the Complexity of Inherited Retinal Dystrophies Molecular Testing: Added Value of Targeted Next-Generation Sequencing

Isabella Bernardis,[1,2] Laura Chiesi,[3] Elena Tenedini,[1,2] Lucia Artuso,[1,2] Antonio Percesepe,[4] Valentina Artusi,[1] Maria Luisa Simone,[1,5] Rossella Manfredini,[5,6] Monica Camparini,[7] Chiara Rinaldi,[7] Antonio Ciardella,[8] Claudio Graziano,[9] Nicole Balducci,[8] Antonia Tranchina,[9] Gian Maria Cavallini,[3] Antonello Pietrangelo,[1,2] Valeria Marigo,[5] and Enrico Tagliafico[1,2]

[1] *Center for Genome Research, University of Modena and Reggio Emilia, Modena, Italy*
[2] *Department of Medical and Surgical Sciences, University of Modena and Reggio Emilia, Modena, Italy*
[3] *Institute of Ophthalmology, University of Modena and Reggio Emilia, Modena, Italy*
[4] *Medical Genetics Unit, Azienda Ospedaliero-Universitaria di Parma, Parma, Italy*
[5] *Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy*
[6] *Centre for Regenerative Medicine, University of Modena and Reggio Emilia, Modena, Italy*
[7] *Ophthalmology, S.Bi.Bi.T. Department, University of Parma, Parma, Italy*
[8] *Ophthalmology Unit, Policlinico S. Orsola-Malpighi, Bologna, Italy*
[9] *Medical Genetics Unit, Policlinico S. Orsola-Malpighi, Bologna, Italy*

Correspondence should be addressed to Enrico Tagliafico; enrico.tagliafico@unimore.it

Received 15 June 2016; Revised 30 September 2016; Accepted 20 October 2016

Academic Editor: Ozgur Cogulu

To assess the clinical utility of targeted Next-Generation Sequencing (NGS) for the diagnosis of Inherited Retinal Dystrophies (IRDs), a total of 109 subjects were enrolled in the study, including 88 IRD affected probands and 21 healthy relatives. Clinical diagnoses included Retinitis Pigmentosa (RP), Leber Congenital Amaurosis (LCA), Stargardt Disease (STGD), Best Macular Dystrophy (BMD), Usher Syndrome (USH), and other IRDs with undefined clinical diagnosis. Participants underwent a complete ophthalmologic examination followed by genetic counseling. A custom AmpliSeq™ panel of 72 IRD-related genes was designed for the analysis and tested using Ion semiconductor Next-Generation Sequencing (NGS). Potential disease-causing mutations were identified in 59.1% of probands, comprising mutations in 16 genes. The highest diagnostic yields were achieved for BMD, LCA, USH, and STGD patients, whereas RP confirmed its high genetic heterogeneity. Causative mutations were identified in 17.6% of probands with undefined diagnosis. Revision of the initial diagnosis was performed for 9.6% of genetically diagnosed patients. This study demonstrates that NGS represents a comprehensive cost-effective approach for IRDs molecular diagnosis. The identification of the genetic alterations underlying the phenotype enabled the clinicians to achieve a more accurate diagnosis. The results emphasize the importance of molecular diagnosis coupled with clinic information to unravel the extensive phenotypic heterogeneity of these diseases.

1. Introduction

Inherited Retinal Dystrophies (IRDs) are a heterogeneous group of eye disorders characterized by rod and/or cone photoreceptor cells degeneration, which include Retinitis Pigmentosa (RP), Leber Congenital Amaurosis (LCA), Stargardt Disease (STGD), Best Macular Dystrophy (BMD), and syndromic forms such as Usher Syndrome (USH).

Table 1: Patients cohort.

Clinical diagnosis	Number of cases	Healthy relatives	Familiar Cases (number of families)	Presumed inheritance in family				Sex		Age at genetic counseling	
				Sporadic	AD	AR	XL	M	F	Range	Median
BMD	4		2 (1)		4			1	3	12–65	58
LCA	5	5		1		4		2	3	5–85	9
STGD	14		6 (3)			14		5	9	8–59	28
RP	45	12	9 (4)	14	6	20	5	25	20	2–73	47.5
USH	3					3		2	1	33–53	51
nd IRD	17	4	6 (2)	6	6	5		13	4	2–62	35
Total	88	21	23 (10)	21	16	46	5	48	40	2–85	37

BMD: Best Macular Dystrophy; LCA: Leber Congenital Amaurosis; STGD: Stargardt disease; RP: Retinitis Pigmentosa; USH: Usher syndrome; nd IRD: inherited retinal degeneration not otherwise specified without precisely defined diagnosis; AD: autosomal dominant; AR: autosomal recessive; XL: X-linked; M: male; F: female.

The overall prevalence of these disorders is ~1 in 4,000 individuals for RP, ~1 in 90,000 individuals for LCA and USH, ~1 in 5,000–10,000 individuals for STGD, and 1/5000–1/67000 for BMD (http://www.orpha.net). Classification of IRDs considers the principal site of retinal dysfunction (rod, cone, retinal pigment epithelium, or inner retina), the mode of inheritance, the underlying gene defect, typical age of onset, rate of progression, and association with systemic syndromes. The genetic bases of IRDs are highly heterogeneous, with almost 150 genes currently known [RetNet, https://sph.uth.edu/retnet/] and a wide clinical and genetic overlap among the different disorders, with high phenotypic variability and genes associated with more than one phenotype. The inheritance of these diseases is also complex, with autosomal dominant (AD), autosomal recessive (AR), X-linked (XL), and even digenic patterns [1]. The extensive clinical and genetic heterogeneity in IRD, along with the variable age of onset, the incomplete penetrance, and unclear inheritance, hamper clinical diagnosis.

Recently, Next-Generation Sequencing (NGS) has been used for the genetic diagnosis of retinal diseases [2–6] and has been reported as a cost-effective approach [7, 8] with a wide range of reported mutation detection rates related to differences in number of genes analyzed, NGS platform, and cohort size but above all composition of the study case phenotypes. We therefore present a multidisciplinary approach coupled with a comprehensive NGS amplicon-based strategy to explore IRD genetic complexity and evaluate genotype-phenotype correlations.

2. Patients and Methods

This study was approved by the ethics committee (Comitato Etico di Modena, Modena, Italy). The procedures followed were in accordance with the Helsinki Declaration of 1975, as revised in 2000, and samples were obtained after patients had provided written informed consent.

A total of 109 samples were collected, including 88 IRDs affected probands with unknown molecular diagnosis and 21 healthy family members (Table 1). Subjects were recruited at the Medical Genetics Unit of the University Hospital of Modena (70 samples), at the Medical Genetics Unit of Parma University Hospital (15 samples) and Medical Genetics Unit of Policlinico Sant'Orsola Malpighi, Bologna (24 samples). All subjects underwent a complete ophthalmologic examination (visual acuity, anterior segment and fundus examination, spectral domain-optical coherence tomography, electroretinogram, and/or electrooculogram) followed by genetic counseling. When indicated fundus autofluorescence imaging and visual field were also performed. Clinical information for the patients with identified pathogenic mutations is shown in Supplementary Table 1 (in Supplementary Material available online at http://dx.doi.org/10.1155/2016/6341870). Clinical diagnoses of participating subjects included RP, USH (hearing impairment + RP), LCA, STGD, BMD, and IRDs not otherwise specified or with imprecisely defined clinical diagnosis. Four control patients with known molecular diagnosis were used to validate our method.

2.1. AmpliSeq Panel Design and Ion Torrent™ PGM™ Library Preparation and Sequencing. The Ion AmpliSeq technology (Life Technologies Ltd., Paisley, UK) was used to design a panel of 72 genes (Supplementary Table 2) associated with the following IRD forms: RP, LCA, STGD, BMD, and USH [RetNet, https://sph.uth.edu/retnet/]. The Ion AmpliSeq Designer tool (https://www.ampliseq.com/browse.action) generated an optimized primers design encompassing the coding DNA sequence of the selected genes, for a total of 1.649 amplicons divided into two pools to optimize coverage and multiplex PCR conditions. Libraries were prepared using the Ion AmpliSeq Library Kit 2.0 starting from 15 ng of gDNA/pool according to manufacturer's recommendations. Template preparation was performed using an Ion OneTouch™ 2 System following the latest version of the manufacturer's manuals. The template positive Ion Sphere Particles (ISPs+) were sequenced on an Ion Torrent Personal Genome Machine® (PGM) System (Life Technologies Ltd., Paisley, UK) using the Ion 318™ Chip kit v2 following the Ion PGM Sequencing 200 Kit v2 manual.

2.2. Sanger Sequencing. Sanger sequencing was performed to validate *CNGB1* c.875-5_891dup mutation (identified with

an anomalous distribution of NGS reads attributable to amplification problems due to the insertion itself located at the end of the target region) and to sequence *RPGR* ORF15 partially uncovered by the NGS panel. Primers for PCR and sequencing are shown in Supplementary Table 3. The following conditions were used: a 50 μL PCR reaction containing 100 ng of DNA, 100 pmol of forward and reverse primers, 5 μL of buffer, and 0.5 μL of Taq Expand High Fidelity™ DNA Polymerase (Roche). PCR amplification (see Supplementary Table 3) was performed using a Gene Amp PCR System 9700 (Applied Biosystems, California, USA). The resultant amplicons were purified using High Pure PCR Product Purification Kit (Roche). Additional primers for *RPGR* sequencing were used. The sequencing reactions were performed with BigDye Terminator v1.0 (Life Technologies) and run on ABI PRISM® 3130XL Genetic Analyzer (Life Technologies). Due to sequence composition and technical difficulties, part of *RPGR* ORF15 (~250 bp, chrX: 38145343–38145593) could not be accurately sequenced with Sanger sequencing.

2.3. Data Analysis. Samples were processed using the Ion Torrent Suite™ (TS) Software for raw data processing and sequence alignment to the human genome reference sequence hg19. The TS Variant Caller was used for the detection of germline variants that were subsequently analyzed using the following optimized filtering and annotation pipeline. Annovar [9] and Variant Effect Predictor (VEP) [10] were used to functionally annotate the detected variants, retrieving RefSeq gene annotation, dbSNP rs identifiers, ClinVar accession, and allele frequency observed in the population (1000-Genome Project, NHLBI GO Exome Sequencing Project ESP6500SI-V2, Exome and Aggregation Consortium ExAC 0.3). Variants with low coverage or low frequency (<30 reads or <30%, resp.) were filtered out. The synonymous variants and variants having an allele frequency greater than 1% reported in the population were discarded as well. In addition, an internal database, built with all variants present in our cohort of processed samples, allowed recognizing and classifying as polymorphisms variants not listed in public databases. Variants were further annotated with conservation scores and functional predictions listed in dbNSFP [11–13], a database which compiles scores from various prediction algorithms, among which are SIFT, Polyphen2, LRT, MutationTaster, MutationAssessor, and FATHMM. Retina International (http://www.retina-international.org/), RPGR database (http://rpgr.hgu.mrc.ac.uk/index.php?select_db=RPGR), CEP290base (http://cep290base.cmgg.be/), and BEST1 LOVD database (http://www-huge.uni-regensburg.de/BEST1_database) were used to explore additional annotations and literature information, if present. Splice-altering predictions were obtained using the online tools Human Splicing Finder (HSF 3.0) [14] and NNSPLICE 0.9 [15] and the databases dbscSNV [16] and SPIDEX [17], which provide predicted effects for all of the potential variants within splicing consensus regions or across the entire genome, respectively. For the prioritization of pathogenetic mutations, the evaluation of inheritance mode was taken into account, along with segregation information coming from the sequencing of healthy family members, if available.

NGS procedure and data analysis were tested on the four control samples with known molecular diagnosis as proof of concept. In all cases the previously identified variants were correctly detected and prioritized as pathogenic variants.

3. Results

A cohort of 109 samples (Table 1), including 88 IRDs affected probands without molecular diagnosis and 21 unaffected family members, was analyzed by the newly developed system based on NGS and data analysis. A total of 19 sequencing runs were performed (6 samples/Ion Chip 318), obtaining on average a mean coverage of 450 mapped reads, with 92% mean uniformity and 97.6% (SD ± 1.4) of target regions covered at least 30x (96.2% > 50x). For each sample, 242 raw variants were detected on average. Annotation and filtering procedure resulted in the identification of possibly causative mutations in 59.1% of patients ($n = 52/88$) (Table 2, Figure 1). The majority of the obtained molecular diagnoses were consistent with the subject's clinical presentation and family history.

We found pathogenic mutations in 16 genes, with the most recurrent being *ABCA4* for STGD and *USH2A* for RP/USH patients. The majority of the mutated genes were inherited with an AR pattern (78.9%), followed in order by AD (11.5%) and XL (9.6%) inheritance. The majority of cases displaying recessive inheritance were compound heterozygous of two different pathogenic variants, in line with the low frequency of consanguineous marriages in Italy

Identified candidate pathogenic mutations are shown in Table 3. Overall, 63 different mutations were identified: 62.5% of variants were already reported in previous studies, while 37.5% were novel. Among the list of novel variants, 56% were missense predicted to have deleterious protein functional effect by the prediction algorithms described in the Patients and Methods (predicted to be damaging by at least three of the applied algorithms), and 44% were frameshift, nonsense, or splice-site mutations that might severely affect protein function. Notably, 12% of identified variants were located within splicing consensus regions, and additional 12% were exonic variants predicted to alter splicing through enhancer/silencer motif modification or the creation of new potential donor/acceptor sites.

Table 2 summarizes the mutation detection rates obtained for the different clinical subtypes of our study cohort. The highest diagnostic yields were achieved for BMD, LCA, USH, and STGD patients with well-defined clinical diagnosis, where the number of known genes associated with each disease is relatively limited.

For BMD cases, all diagnosed patients were heterozygous for mutations on *BEST1*. Three patients (mother and son) were found to harbour a novel *BEST1* missense mutation c.80G>C (p.Ser27Thr) located in the immediate N-terminus, in one of the four mutational hotspots regions in the highly conserved N-terminal half of the protein [18] and predicted to be deleterious by all interrogated algorithms.

Figure 1: The chart summarizes the diagnostic yields obtained for the clinical subtypes of this study. The different levels of circles (from inner to outside) specify clinical diagnoses, inheritance mode, mutated genes, and clinical reassessment.

Table 2: Diagnostic yields for the clinical subtypes of this study.

Clinical diagnosis	Cases (n)	Genetic diagnosis (n)	Unsolved cases (n)	Clinical reassessment (final diagnosis)	Diagnostic yield (%)
BMD	4	4	—		100
LCA	5	4	1		80
STGD	14	11	3		78.5
RP	45	27	18	2 (USH)	60.0
USH	3	3	—		100
nd IRD	17	3	14	3 (ACHM, LCA, STGD)	17.6
Total	88	52	36	5	59.1

BMD: Best Macular Dystrophy; LCA: Leber Congenital Amaurosis; STGD: Stargardt Disease; RP: Retinitis Pigmentosa; USH: Usher Syndrome; nd IRD: inherited retinal degeneration not otherwise specified without precisely defined diagnosis; ACHM: Achromatopsia.

Table 3

Patient ID	Family	Clinical diagnosis	Clinical reassessment	Genotype	Inheritance	Gene	Mutation type	Region	cds change
IRD027		STGD		Comp Het	ar	*ABCA4*	Splice_region	INTRON_40	c.5714+5G>A
						ABCA4	Frameshift	EXON_11	c.1375delA
IRD036		STGD		Comp Het	ar	*ABCA4*	Stop_gained	EXON_14	c.2099G>A
	Familiar case					*ABCA4*	Splice_region syn	EXON_6	c.768G>T
IRD037		STGD		Comp Het	ar	*ABCA4*	Stop_gained	EXON_14	c.2099G>A
						ABCA4	Splice_region syn	EXON_6	c.768G>T
IRD042		STGD		Comp Het	ar	*ABCA4*	Missense	EXON_42	c.5882G>A
	Familiar case					*ABCA4*	Missense	EXON_6	c.634C>T
IRD043		STGD		Comp Het	ar	*ABCA4*	Missense	EXON_42	c.5882G>A
						ABCA4	Missense	EXON_12	c.1622T>C
IRD050		STGD		Comp Het	ar	*ABCA4*	Missense	EXON_16	c.2461T>A
						ABCA4	Missense	EXON_15	c.2300T>A
IRD054		STGD		Comp Het	ar	*ABCA4*	Stop_gained	EXON_47	c.6445C>T
						ABCA4	Missense	EXON_42	c.5882G>A
IRD055		STGD		Comp Het	ar	*ABCA4*	Missense	EXON_19	c.2842C>T
						ABCA4	Missense	EXON_15	c.2300T>A
IRD061		STGD		Comp Het	ar	*ABCA4*	Missense	EXON_42	c.5882G>A
						ABCA4	Missense	EXON_28	c.4139C>T
IRD062		STGD		Comp Het	ar	*ABCA4*	Missense	EXON_42	c.5882G>A
						ABCA4	Missense	EXON_16	c.2549A>G
IRD073		nd IRD	STGD	Hom	ar	*ABCA4*	Missense	EXON_19	c.2894A>G
IRD077		STGD		Comp Het	ar	*ABCA4*	Missense	EXON_37	c.5285C>A
						ABCA4	Missense	EXON_15	c.2300T>A
IRD047		BMD		Het	ad	*BEST1*	Missense	EXON_2	c.73C>T
IRD057	Familiar case	BMD		Het	ad	*BEST1*	Missense	EXON_2	c.80G>C
IRD058		BMD		Het	ad	*BEST1*	Missense	EXON_2	c.80G>C
IRD064		BMD		Het	ad	*BEST1*	Missense	EXON_2	c.80G>C
IRD010		LCA		Comp Het	ar	*CEP290*	Missense	EXON_33	c.4237G>C
						CEP290	Frameshift	EXON_23	c.2390delA
IRD066		RP		Comp Het	ar	*CEP290*	Stop_gained	EXON_48	c.6640A>T
						CEP290	Frameshift	EXON_14	c.1219_1220delAT

Table 3: Continued.

IRD072		nd IRD	LCA	Comp Het	ar	*CEP290*	Missense	EXON_14	c.1298A>G
						CEP290	Frameshift	EXON_3	c.164_167delCTCA
IRD039		RP		Hom	ar	*CNGB1*	Frameshift	EXON_13	c.875-5_891dup
IRD052		RP		Comp Het	ar	*CNGB1*	Missense	EXON_29	c.2957A>T
						CNGB1	Frameshift	EXON_13	c.875-5_891dup
IRD068		RP		Comp Het	ar	*CNGB1*	Splicing, syn	EXON_26	c.2526C>T
						CNGB1	Missense	EXON_21	c.2153G>C
IRD085		RP		Hom	ar	*CNGB1*	Missense	EXON_23	c.2284C>T
IRD032		nd IRD	ACHM	Comp Het	ar	*CNGB3*	Splice_donor	INTRON_13	c.1578+1G>A
						CNGB3	Frameshift	EXON_10	c.1148delC
IRD029		RP		Hom	ar	*CRB1*	Missense	EXON_5	c.2200G>A
IRD030	Familiar case	RP		Hom	ar	*CRB1*	Missense	EXON_5	c.2200G>A
IRD031		RP		Hom	ar	*CRB1*	Missense	EXON_5	c.2200G>A
IRD035		LCA		Het	ad	*CRX*	Frameshift	EXON_4	c.514delC
IRD008		RP		Hom	ar	*PDE6B*	Splice_region	EXON_18	c.2193+1delG
IRD013		RP		Comp Het	ar	*PDE6B*	Missense	EXON_4	c.794G>A
					ar	*PDE6B*	Intron	INTRON_8	c.1108-10G>A
IRD026		RP		Het	ad	*RHO*	Missense	EXON_3	c.568G>T
IRD016		RP		Comp Het	ar	*ROM1*	Missense	EXON_1	c.178C>A
						ROM1	Missense	EXON_1	c.323C>T
IRD033		RP		Hem	xl	*RP2*	Frameshift	EXON_2	c.382_383delTT
IRD076		RP		Hom	ar	*RPE65*	Missense	EXON_2	c.65T>C
IRD001		RP		Comp Het	ar	*RPE65*	Missense	EXON_2	c.65T>C
						RPE65	Frameshift	EXON_9	c.893delA
IRD074		LCA		Hom	ar	*RPE65*	Missense	EXON_5	c.430T>G
IRD002		LCA		Comp Het	ar	*RPGRIP1*	Frameshift	EXON_15	c.2225_2226delGA
						RPGRIP1	Frameshift	EXON_17	c.2795_2796insT
IRD012		RP		Hem	xl	*RPGR*	Missense	EXON_8	c.785C>G
IRD067		RP		Hem	xl	*RPGR*	Missense	EXON_8	c.814G>T
IRD075		RP		Hem	xl	*RPGR*	Missense, Splice_region	EXON_2	c.154G>A
IRD017		RP		Hem	De novo	*RPGR*	Frameshift	EXON_2	c.89delT
IRD059	Familiar case	RP		Comp Het	ar	*TULP1*	Missense	EXON_15	c.1590C>G
						TULP1	Missense	EXON_13	c.1255C>T
IRD060		RP		Comp Het	ar	*TULP1*	Missense	EXON_15	c.1590C>G
						TULP1	Missense	EXON_13	c.1255C>T
IRD041		RP		Comp Het	ar	*TULP1*	Splice_region	INTRON_14	c.1496-6C>A
						TULP1	Missense	EXON_14	c.1445G>A

Table 3: Continued.

IRD007		USH		Comp Het	ar	*USH2A*	Missense	EXON_63	c.12420T>G
						USH2A	splice_region, syn	EXON_28	c.5775A>T
IRD009		USH		Comp Het	ar	*USH2A*	Missense	EXON_63	c.13546G>T
						USH2A	splice_region, Missense	EXON_10	c.1645T>C
IRD021		RP		Comp Het	ar	*USH2A*	Missense	EXON_69	c.14995A>G
						USH2A	Missense	EXON_8	c.1481A>G
IRD023	Familiar case	RP	USH	Comp Het	ar	*USH2A*	Missense	EXON_13	c.2296T>C
						USH2A	Frameshift	EXON_3	c.545_548delAAGA
IRD024		RP	USH	Comp Het	ar	*USH2A*	Missense	EXON_13	c.2296T>C
						USH2A	Frameshift	EXON_3	c.545_548delAAGA
IRD038		RP		Comp Het	ar	*USH2A*	Missense	EXON_13	c.2296T>C
						USH2A	Missense	EXON_13	c.2276G>T
IRD084		USH		Hom	ar	*USH2A*	Frameshift	EXON_69	c.14977_14978delTT
IRD034		RP		Hom	ar	*USH2A*	Missense	EXON_63	c.12574C>T

Table 3: Continued.

Patient ID	Protein change	Frequency (%)	Coverage (# reads)	Segregation and unaffected siblings	Functional predictions (dbNSFP)	Splicing predictions			Reference
						Human Splicing Finder	dbscSNV	SPIDEX	
IRD027		44.9	514			Broken WT Donor Site	0.999|0.988	−3.21	PMID: 15494742
IRD036	p.Thr459GlnfsX2	47.7	1179						PMID: 21911583
	p.Trp700X	48.2	303		.|..|N|A|.|.|.|.|.|D				PMID: 11702214
IRD037	p.Val256Val	47.2	53			Broken WT Donor Site	1.000|0.952	−2.43	PMID: 12037008
	p.Trp700X	44.5	110		.|..|N|A|.|.|.|.|.|D	New Acceptor Site		−5.41	PMID: 11702214
IRD042	p.Val256Val	48.3	29			Broken WT Donor Site	1.000|0.952	−2.43	PMID: 12037008
	p.Gly1961Glu	47.1	1325		D|DD|D|D|N|D|D|D|D|D				PMID: 9295268
	p.Arg212Cys	49.1	432		D|DD|D|A|M|D|D|D|D|D				PMID: 11726554
IRD043	p.Gly1961Glu	46.9	796		D|DD|D|D|N|D|D|D|D|D				PMID: 9295268
	p.Leu541Pro	51.9	727		D|DD|D|A|M|D|D|D|D|D				PMID: 11527935
IRD050	p.Trp821Arg	43.8	309		D|DD|D|D|H|T|D|D|D|D				PMID: 11527935
	p.Val767Asp	46.3	452		D|BB|D|D|M|D|T|D|D|D				PMID: 15494742
IRD054	p.Arg2149X	49.1	422		.|..|D|A|.|.|.|.|.|D	New ESS site		−58.3	PMID: 12202497
	p.Gly1961Glu	49.4	1448		D|DD|D|D|N|D|D|D|D|D				PMID: 9295268
IRD055	p.Arg948Cys	52.0	175		T|BB|N|D|L|D|T|T|N|N				This study
	p.Val767Asp	51.5	437		D|BB|D|D|M|D|T|D|D|D				PMID: 15494742
IRD061	p.Gly1961Glu	50.0	729		D|DD|D|D|N|D|D|D|D|D				PMID: 9295268
	p.Pro1380Leu	55.8	437		D|DP|N|A|M|D|D|D|D|D	New ESS site		−5.44	PMID: 11726554
IRD062	p.Gly1961Glu	100	787		D|DD|D|D|N|D|D|D|D|D				PMID: 9295268
	p.Tyr850Cys	49.4	176		D|DD|D|D|M|T|D|D|D|D				PMID: 23096905
IRD073	p.Asn965Ser	100	225		D|DD|D|D|L|D|D|D|D|D				PMID: 9054934
IRD077	p.Ala1762Asp	50.8	259		D|DD|D|A|M|D|D|D|D|D				PMID: 15192030
	p.Val767Asp	51.4	752		D|BB|D|D|M|D|T|D|D|D				PMID: 15494742
IRD047	p.Arg25Trp	56.0	348		D|DD|U|D|M|D|D|D|D|D	New Donor Site, New ESS site			PMID: 10798642
IRD057	p.Ser27Thr	46.8	344		D|DD|U|D|H|D|D|D|D|D				This study
IRD058	p.Ser27Thr	45.5	317		D|DD|U|D|H|D|D|D|D|D				This study
IRD064	p.Ser27Thr	47.1	453		D|DD|U|D|H|D|D|D|D|D				This study
IRD010	p.Asp1413His	49.2	413		D|BB|D|D|N|T|T|T|N|D				ClinVar: RCV000082249.5
IRD066	p.Lys797SerfsX2	30.1	163						This study
	p.Lys2214X	47.5	705		.|..|D|A|.|.|.|.|.|D	ESE Site Broken		−86.6	This study
	p.Met407GlufsX14	51.1	225						PMID: 17724218

Table 3: Continued.

IRD072	p.Asp433Gly	53.4	116		T|DP|D|D|L|T|T|T|D|D	New ESS site, New donor site			This study
	p.Thr55SerfsX3	43.2	243						PMID: 20690115
IRD039	p.Gly298CysfsX13	100 *							This study
IRD052	p.Asn986Ile	51.7	471		D|DD|D|D|M|D|D|D|D|D				PMID: 21147909
	p.Gly298CysfsX13	26, 7*	258						This study
IRD068	Thr842Thr	52.1	431			ESE Site Broken			This study
	p.Gly718Ala	47.1	153		D|PP|D|D|M|T|T|T|D|D				This study
IRD085	p.Arg762Cys	100	57		D|DD|D|D|H|D|D|D|D|D				This study
IRD032		47.8	907			Broken WT Donor Site		−8.56	PMID: 15657609
	p.Thr383IlefsX13	46.5	588						PMID: 15657609
IRD029	p.Gly734Arg	100	397		D|DD|.|D|M|T|D|D|D|D				This study
IRD030	p.Gly734Arg	100	397		D|DD|.|D|M|T|D|D|D|D				This study
IRD031	p.Gly734Arg	100	397		D|DD|.|D|M|T|D|D|D|D				This study
IRD035	p.Pro172LeufsX15	50.5	521						This study
IRD008		100	395	Brother: Het		Broken WT Donor Site			This study
IRD013	p.Arg265Gln	51.7	319	n.a.	T|DD|D|D|L|T|T|T|N|D				ClinVar: RCV000178068.1
		54.7	75	Mother: Het			0.001|0.096		PMID: 8698075
IRD026	p.Asp190Tyr	44.6	168		D|DD|D|D|M|T|T|T|D|D				PMID: 8401533
IRD016	p.Pro60Thr	56.1	278		T|BB|N|N|L|T|T|T|N|N				PMID: 8595413
	p.Thr108Met	52.8	108		T|PB|N|D|L|T|T|T|N|D				PMID: 8595413
IRD033	p.Leu129ValfsX9	100	392						This study
IRD076	p.Leu22Pro	100	495		T|BB|D|D|M|D|D|D|N|D				PMID: 9801879
IRD001	p.Leu22Pro	46.3	257	Brother: wt	T|BB|D|D|M|D|D|D|N|D				PMID: 9801879
	p.Lys298SerfsX27	98	150	Brother: wt					PMID: 11462243
IRD074	p.Tyr144Asp	100	430	Father: Het	D|DD|D|D|M|D|D|D|D|D				PMID: 11462243
IRD002	p.Glu743ArgfsX24	48.8	570	Father: Het					This study
	p.Glu933X	48.8	400	Mother: Het					This study
IRD012	p.Ala262Gly	100	280		T|BB|N|N|L|D|T|T|N|N				This study
IRD067	p.Gly272Cys	100	155		D|DD|D|D|H|D|D|D|D|D				This study
IRD075	p.Gly52Arg	100	348		D|DP|U|D|M|T|T|T|D|D	Broken WT Donor Site			PMID: 15364249
IRD017	p.Phe30SerfsX38	100	113	Brother: wt Female twin: wt					This study
IRD059	p.Ile530Met	50.6	682		D|DD|D|D|H|D|D|D|D|N				This study
	p.Arg419Trp	49.5	645		D|DD|D|D|H|D|D|D|D|D				PMID: 25342620
IRD060	p.Ile530Met	51.0	655		D|DD|D|D|H|D|D|D|D|N				This study
	p.Arg419Trp	45.3	575		D|DD|D|D|H|D|D|D|D|D				PMID: 25342620
IRD041		54.1	727	Father: Het			0.005|0.419		PMID: 9660588
	p.Arg482Gln	48.5	485	Mother: Het	D|DD|D|D|H|D|D|D|D|D	New Acceptor Site		−1.28	PMID: 22665969

Table 3: Continued.

IRD007	p.Cys4140Trp	50.5	214	D|DD|D|D|M|T|T|T|D|D				This study
		49.5	398		Broken WT Donor Site	0.998|0.986	−4.24	This study
IRD009	p.Gly4516Trp	53.8	239	D|DD|U|D|H|T|D|D|D|D				This study
	p.Cys549Arg	49.2	566	D|DD|U|D|H|D|D|D|D|D		0.417|0.520		This study
IRD021	p.Thr4999Ala	51.0	400	D|DD|U|D|M|T|T|T|D|D				This study
	p.Tyr494Cys	49.0	400	D|DD|N|D|L|T|T|T|D|D				This study
IRD023	p.Cys766Arg	39.0	82	D|DD|D|D|H|D|D|D|D|D				PMID: 23591405
	p.Lys182ArgfsX9	61.4	202					This study
IRD024	p.Cys766Arg	43.5	124	D|DD|D|D|H|D|D|D|D|D				PMID: 23591405
	p.Lys182ArgfsX9	48.0	225					This study
IRD038	p.Cys766Arg	47.2	89	D|DD|D|D|H|D|D|D|D|D				PMID: 23591405
	p.Cys759Phe	51.1	90	D|DD|D|A|H|D|D|D|D|D				PMID: 10775529
IRD084	p.Phe4993ProfsX7	100	483					PMID: 24944099
IRD034	p.Arg4192Cys	100	515	D|DP|N|D|M|D|D|D|D|D				PMID: 24498627

ACHM: Achromatopsia; ad: autosomal dominant; ar: autosomal recessive; BMD: best macular disease; Comp Het: compound heterozygous; ESE: exonic splicing enhancer; ESS: exonic splicing silencer; Hem: Hemizygous; Het: heterozygous; Hom: homozygous; LCA: Leber Congenital Amaurosis; nd IRD: inherited retinal degeneration not otherwise specified without precisely defined diagnosis; RP: Retinitis Pigmentosa; STGD: Stargardt Disease; USH: Usher Syndrome; wt: wild-type; xl: X-linked. For nonsynonymous variants, predictions from dbNSFP are reported, comprising scores from the following alghoritms: SIFT | Polyphen2HDIV Polypehn2HVAR | LRT | MutationTaster | MutationAssessor | FATHMM | MetaSVM | MetaLR | PROVEAN | fathmm-MKL. For splicing variants, predictions from Human Splicing Finder, dbscSNV (ada_score|rf_score) and SPIDEX are reported. For SPIDEX, max dPSI is shown if lower than −1 (maximum mutation − induced change in the percentage of transcripts with the exon spliced in). Familiar case: the patients were from the same family. * Sanger sequencing was performed to confirm mutation frequency.

For STGD patients, genetic diagnosis was achieved in 11 out of 14 (78.5% of the cases). All diagnosed patients in our cohort carried mutations on *ABCA4*. In 75% of the unsolved cases at least one *ABCA4* pathogenic allele was identified, suggesting the presence of disease-causing mutations lying outside the coding sequence covered by our panel, as reported in a previous study [19].

In LCA patients, causative mutations were identified in *CEP290, RPE65, RPGRIP1,* and *CRX* genes, and only one case remained unsolved (20% of the total LCA cases), whereas all Usher 2 syndrome cases were found to carry mutations in USH2A gene.

For RP patients, genetic diagnosis was achieved in 27 out of 45 (60% of the cases), involving mutations in 11 different genes: confirming that these phenotypes are genetically heterogeneous (Figure 1). Dominant mutations were identified in *RHO* gene, whereas *USH2A, CNGB1,* and *TULP1* were the most recurrently mutated genes in ARRP. X-linked inheritance was established for 5 RP male patients (4 probands had mutations in *RPGR*, whereas one had a mutation in RP2). The identification of *USH2A* as the defective gene in patients with initial clinical diagnosis of RP was followed by audiometric testing to establish if there were any hearing deficiencies. A hearing impairment was found in 2 cases out of 5 leading to clinical reassessment and final diagnosis of USH (Table 2).

For patients with IRD without a defined clinical diagnosis or with unclear disease manifestations, we identified causative mutations in 7 out of 17 probands (23.5% of the total IRD cases). In two cases the molecular results allowed a refined clinical diagnosis: a compound heterozygosity of two mutations in *CEP290* led to a genetic diagnosis of LCA in a patient with initial diagnosis of North Carolina or Stargardt macular dystrophy, whereas a homozygous pathogenic variant in *ABCA4* was found in a patient with tapetoretinal degeneration.

In 36 patients (12 familiar and 24 sporadic) the molecular analysis did not achieve any definitive result, even after the analysis of the healthy family members, which was performed in 8 cases. Half of the cases with a negative test result (18 out of 36) were affected by RP. The additional analysis of the *RPGR* ORF15 (a mutational hotspot which was nonsufficiently covered in our panel) for the male patients with a sporadic or suspected X-linked pattern of inheritance (10 patients) by Sanger sequencing yielded no additional mutations.

4. Discussion

The results of the present study confirm that high-throughput Next-Generation Sequencing represents a comprehensive cost-effective approach for the molecular diagnosis of Inherited Retinal Dystrophies (IRDs), achieving a molecular diagnosis for 59.1% of the studied cases. More specifically, among the different clinical phenotypes, the highest detection rates were achieved for BMD, LCA, USH, and STGD patients, in whom the genetic test clearly confirmed the clinical diagnoses (Table 2). The results of the RP and of the not defined IRD cohorts, instead, demonstrated the high genetic heterogeneity of this diseases and the essential contribution of our NGS analysis to achieving an accurate diagnosis, with the involvement of 12 different genes in 28 sporadic cases. Revision of the initial diagnosis, performed for 9.6% of the genetically diagnosed patients, further emphasizes the importance of a comprehensive genotype/phenotype analysis to unravel the extensive heterogeneity of these diseases. Notably, a remarkable fraction of identified variants are splice-altering mutations (25% of the total mutation burden, 16 out of 64), located within splicing consensus regions, or exonic variants predicted to cause enhancer/silencer motif modification or the creation of new potential donor/acceptor, which are amenable to the antisense-mediated splicing-correction approaches, as recently reported for several genetic diseases, including *CEP290*-caused LCA [20, 21].

The prevalence of IRD and most importantly the frequency of gene mutations causing those diseases are not well characterized in Italy and only few data have been reported [22–24]. *RPE65, CRB1,* and *GUCY2D* were identified as the most prevalent mutated genes in Italian LCA patients [22] and *RHO* was reported to be the gene most commonly responsible for ADRP [23] and *EYS* the most recurrent for nonsyndromic ARRP and sporadic cases [24]. Our study contributes only partially to the knowledge of the gene mutation frequencies, since each IRD type is represented by small cohorts of cases (i.e., the LCA and dominant RP phenotypes were accounted for by 5 and 6 cases, resp.), and some probands of other ethnicities have been included too. Indeed, regarding LCA, we identified mutations in CEP290, *RPE65*, *CRX*, and *RPGRIP1* genes.

For ADRP, *RHO* was identified to be responsible for the phenotype in one case, whereas, in ARRP and sporadic RP, *USH2A, CNGB1,* and *TULP1* were the most recurrently mutated genes. *RPE65* mutations were found in two ARRP cases: in one more case, still unsolved, a single *RPE65* heterozygous pathogenic variant was found. *ROM1* compound heterozygosity was established in one RP proband, suggesting a mechanism of recessive inheritance for this gene associated with dominant and digenic forms. X-linked inheritance was established for 5 RP affected probands, with *RPGR* and *RP2* identified as the disease-causing gene in 4 cases and 1 case, respectively. All BMD diagnosed patients were heterozygous for mutations on *BEST1* gene, the major gene responsible for Best's juvenile form [25], whereas the 78.5% of patients with clinically diagnosed STGD carried pathogenic variants on *ABCA4* [26].

Similarly to a recent study [6], the clinical sensitivity of our NGS analysis was not uniform, with the highest diagnostic yields obtained in conditions where the disease-causing genes have been nearly all identified.

Direct comparison of our findings with other recently published NGS studies [2–6, 27] is not straightforward, due to differences in the number of genes analyzed but especially due to composition and relative representation of the different phenotypes in the patients cohorts. However, the finding of *USH2A* and *ABCA4* as the most mutated genes for RP/USH and STGD patients is consistent with previous reports [27–29]. In our RP cohort, *USH2A* is followed by *CNGB1* and *RPGR*. These two genes, already reported among the most frequently mutated genes in IRD patients [29], were not highly frequently altered in the Saudi population [6] or

in a large cohort of Western European and South Asian individuals [27]. Also, we did not find any alteration in *EYS*, one of the top three genes contributing to IRD in other populations [28, 29].

The different gene alterations identified in our LCA cohort (*CEP290, RPE65, RPGRIP1,* and *CRX* genes) were consistent with the different disease manifestations of the analyzed patients, in accordance with the specific clinical features described for each of the LCA-associated genes [30, 31]. Less direct is the correlation between the genes involved and the phenotypic features in RP, due to the known contribution of environmental factors to late-childhood- and adult-onset-diseases.

Allelic heterogeneity, with different mutations in the same gene causing different phenotypes, is evident also in *USH2A*-related retinal disease. Genotype-phenotype correlations observed in our cohort were in accordance with the allelic hierarchy proposed in a recent study [32], supporting the model that USH represents the null phenotype consequent upon severe *USH2A* defects, whereas milder mutations in at least one allele result in a pure retinal phenotype associated with normal auditory function.

IRD genetic heterogeneity, reflected in the identification of mutations in many genes with a considerable number of previously undescribed alterations, supported the conclusion that molecular diagnosis of these disorders should rely on massive parallel multigene sequencing. Nevertheless, for 36 probands, including 12 familiar cases and 24 unrelated probands, our NGS procedure did not result in the identification of a clear genetic cause of the disease. Some subjects may have mutations that cannot be detected by our amplicon-based approach, such as deep intronic mutations, copy-number variations, or large deletions. In the perspective of the design of a more complete new version of the panel, additional deep intronic regions reported in the literature as carrying disease-causing mutations [19, 33, 34] or a higher exon padding (5 bp in our design, up to 100 bp available in the current pipeline version of the Ion AmpliSeq Designer tool) could be implemented. Moreover, technical limitations, including the difficult amplification of *RPGR* ORF15, a mutational hotspot for X-linked RP, may have accounted for some of the missed diagnosis (our panel is presently covering only 30% of this critical exon), but the addition of the specific analysis by Sanger sequencing of the ORF15 of the *RPGR* gene in 10 males patients, with sporadic/X-linked RP and previously testing negative for pathogenic mutations using our NGS panel, did not reveal any mutation in the analyzed region. Finally, as an improvement to further support the pathogenicity of novel mutations identified in probands, the analysis of both affected and unaffected family member should be performed, when possible.

In some of the patients who tested negative we however identified single potentially pathogenic heterozygous mutations in recessive genes or novel heterozygous missense variants in dominant genes with unknown significance, lacking the appropriate level of evidence to classify them as disease-causing or not in concordance with patients' clinical presentations or family data. The contribution of these variants in combination with deep intronic mutations or large deletions is suspected but could not be demonstrated with the present technique.

Database incompleteness further complicates variant interpretation. Two probands with BMD phenotype and *BEST1* mutation were found to harbour also heterozygous mutation in *RHO* (c.578C>T, p.Thr193Met), which was predicted to be damaging and listed as associated with ADRP in a public database [http://www.retina-international.org/sci-news/databases/mutation-database] but in our cohort was carried also by healthy subject, reinforcing the need of a critical interpretation of the molecular findings in view of the phenotypic features of the patients with IRD until a more thorough knowledge of the frequency of the variants and a critical amount of data present in the public disease databases are reached.

In conclusion, by presenting profoundly different mutation rates varying according to the clinical diagnosis and by reporting 9.61% of cases of reassessment of the initial diagnosis on the basis of the results of the test, our study reinforces the need of a multidisciplinary work-up before and after the genetic testing, due to the implications of the results in terms of risk assessment for family members and inclusion in gene-based clinical trials.

Abbreviations/Acronyms

AD: Autosomal dominant
AR: Autosomal recessive
BMD: Best Macular Dystrophy
IRDs: Inherited Retinal Dystrophies
LCA: Leber Congenital Amaurosis
NGS: Next-Generation Sequencing
RP: Retinitis Pigmentosa
STGD: Stargardt Disease
USH: Usher Syndrome
XL: X-linked.

Competing Interests

No conflicting relationship exists for any author.

Authors' Contributions

Isabella Bernardis and Laura Chiesi contributed equally.

Acknowledgments

This study was supported by Regione Emilia Romagna "RARER," Area1 (E35E09000880002). The authors thank Professor Sandro Banfi for kindly providing the control samples used to validate their procedure. Programma di Ricerca Regione-Università 2010–2012 "Next-Generation Sequencing and Gene Therapy to Diagnose and Cure Rare Diseases in Regione Emilia Romagna (RARER)," Area 1, Strategic Programmes (E35E09000880002), is acknowledged. The funding organization participated in the design of the study.

References

[1] T. P. Dryja, L. B. Hahn, K. Kajiwara, and E. L. Berson, "Dominant and digenic mutations in the peripherin/RDS and ROM1 genes in retinitis pigmentosa," *Investigative Ophthalmology and Visual Science*, vol. 38, no. 10, pp. 1972–1982, 1997.

[2] K. Neveling, R. W. J. Collin, C. Gilissen et al., "Next-generation genetic testing for retinitis pigmentosa," *Human Mutation*, vol. 33, no. 6, pp. 963–972, 2012.

[3] T. Eisenberger, C. Neuhaus, A. O. Khan et al., "Increasing the yield in targeted next-generation sequencing by implicating CNV analysis, non-coding exons and the overall variant load: the example of retinal dystrophies," *PLoS ONE*, vol. 8, no. 11, Article ID e78496, 2013.

[4] J. Wang, V. W. Zhang, Y. Feng et al., "Dependable and efficient clinical utility of target capture-based deep sequencing in molecular diagnosis of retinitis pigmentosa," *Investigative Ophthalmology & Visual Science*, vol. 55, no. 10, pp. 6213–6223, 2014.

[5] M. B. Consugar, D. Navarro-Gomez, E. M. Place et al., "Panel-based genetic diagnostic testing for inherited eye diseases is highly accurate and reproducible, and more sensitive for variant detection, than exome sequencing," *Genetics in Medicine*, vol. 17, no. 4, pp. 253–261, 2015.

[6] N. Patel, M. A. Aldahmesh, H. Alkuraya et al., "Expanding the clinical, allelic, and locus heterogeneity of retinal dystrophies," *Genetics in Medicine*, vol. 18, no. 6, pp. 554–556, 2015.

[7] J. P.-W. Chiang, T. Lamey, T. McLaren, J. A. Thompson, H. Montgomery, and J. De Roach, "Progress and prospects of next-generation sequencing testing for inherited retinal dystrophy," *Expert Review of Molecular Diagnostics*, vol. 15, no. 10, pp. 1269–1275, 2015.

[8] J. P.-W. Chiang and K. Trzupek, "The current status of molecular diagnosis of inherited retinal dystrophies," *Current Opinion in Ophthalmology*, vol. 26, no. 5, pp. 346–351, 2015.

[9] K. Wang, M. Li, and H. Hakonarson, "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data," *Nucleic Acids Research*, vol. 38, no. 16, article e164, 2010.

[10] W. McLaren, B. Pritchard, D. Rios, Y. Chen, P. Flicek, and F. Cunningham, "Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor," *Bioinformatics*, vol. 26, no. 16, pp. 2069–2070, 2010.

[11] X. Liu, X. Jian, and E. Boerwinkle, "dbNSFP: a lightweight database of human nonsynonymous SNPs and their functional predictions," *Human Mutation*, vol. 32, no. 8, pp. 894–899, 2011.

[12] X. Liu, C. Wu, C. Li, and E. Boerwinkle, "dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs," *Human Mutation*, vol. 37, no. 3, pp. 235–241, 2016.

[13] C. Dong, P. Wei, X. Jian et al., "Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies," *Human Molecular Genetics*, vol. 24, no. 8, pp. 2125–2137, 2015.

[14] F.-O. Desmet, D. Hamroun, M. Lalande, G. Collod-Bëroud, M. Claustres, and C. Béroud, "Human Splicing Finder: an online bioinformatics tool to predict splicing signals," *Nucleic Acids Research*, vol. 37, no. 9, article e67, 2009.

[15] M. G. Reese, F. H. Eeckman, D. Kulp, and D. Haussler, "Improved splice site detection in Genie," *Journal of Computational Biology*, vol. 4, no. 3, pp. 311–323, 2009.

[16] X. Jian, E. Boerwinkle, and X. Liu, "*In silico* prediction of splice-altering single nucleotide variants in the human genome," *Nucleic Acids Research*, vol. 42, no. 22, pp. 13534–13544, 2014.

[17] H. Y. Xiong, B. Alipanahi, L. J. Lee et al., "The human splicing code reveals new insights into the genetic determinants of disease," *Science*, vol. 347, no. 6218, Article ID 1254806, 2015.

[18] V. M. Milenkovic, E. Röhrl, B. H. F. Weber, and O. Strauss, "Disease-associated missense mutations in bestrophin-1 affect cellular trafficking and anion conductance," *Journal of Cell Science*, vol. 124, no. 17, pp. 2988–2996, 2011.

[19] T. A. Braun, R. F. Mullins, A. H. Wagner et al., "Non-exomic and synonymous variants in ABCA4 are an important cause of Stargardt disease," *Human Molecular Genetics*, vol. 22, no. 25, Article ID ddt367, pp. 5136–5145, 2013.

[20] R. W. Collin, A. I. den Hollander, S. D. van der Velde-Visser, J. Bennicelli, J. Bennett, and F. P. Cremers, "Antisense oligonucleotide (AON)-based therapy for leber congenital amaurosis caused by a frequent mutation in CEP290," *Molecular Therapy Nucleic Acids*, vol. 1, article e14, 2012.

[21] N. Bacchi, S. Casarosa, and M. A. Denti, "Splicing-correcting therapeutic approaches for retinal dystrophies: where endogenous gene regulation and specificity matter," *Investigative Ophthalmology and Visual Science*, vol. 55, no. 5, pp. 3285–3294, 2014.

[22] F. Simonelli, C. Ziviello, F. Testa et al., "Clinical and molecular genetics of Leber's congenital amaurosis: a multicenter study of Italian patients," *Investigative Ophthalmology & Visual Science*, vol. 48, no. 9, pp. 4284–4290, 2007.

[23] C. Ziviello, F. Simonelli, F. Testa et al., "Molecular genetics of autosomal dominant retinitis pigmentosa (ADRP): a comprehensive study of 43 Italian families," *Journal of Medical Genetics*, vol. 42, no. 7, p. e47, 2005.

[24] C. O. Pierrottet, M. Zuntini, M. Digiuni et al., "Syndromic and non-syndromic forms of retinitis pigmentosa: a comprehensive Italian clinical and molecular study reveals new mutations," *Genetics and Molecular Research*, vol. 13, no. 4, pp. 8815–8833, 2014.

[25] F. Krämer, K. White, D. Pauleikhoff et al., "Mutations in the VMD2 gene are associated with juvenile-onset vitelliform macular dystrophy (Best disease) and adult vitelliform macular dystrophy but not age-related macular degeneration," *European Journal of Human Genetics*, vol. 8, no. 4, pp. 286–292, 2000.

[26] V. C. Sheffield and E. M. Stone, "Genomics and the eye," *The New England Journal of Medicine*, vol. 364, no. 20, pp. 1932–1942, 2011.

[27] J. M. Ellingford, S. Barton, S. Bhaskar et al., "Molecular findings from 537 individuals with inherited retinal disease," *Journal of Medical Genetics*, vol. 53, no. 11, pp. 761–767, 2016.

[28] X.-F. Huang, F. Huang, K.-C. Wu et al., "Genotype-phenotype correlation and mutation spectrum in a large cohort of patients with inherited retinal dystrophy revealed by next-generation sequencing," *Genetics in Medicine*, vol. 17, no. 4, pp. 271–278, 2015.

[29] Z. Ge, K. Bowles, K. Goetz et al., "NGS-based Molecular diagnosis of 105 eyeGENE(®) probands with Retinitis Pigmentosa," *Scientific Reports*, vol. 5, p. 18287, 2015.

[30] S. Hanein, I. Perrault, S. Gerber et al., "Leber congenital amaurosis: comprehensive survey of the genetic heterogeneity, refinement of the clinical definition, and genotype-phenotype correlations as a strategy for molecular diagnosis," *Human Mutation*, vol. 23, no. 4, pp. 306–317, 2004.

[31] F. Coppieters, I. Casteels, F. Meire et al., "Genetic screening of LCA in Belgium: predominance of CEP290 and identification of potential modifier alleles in AHI1 of CEP290-related phenotypes," *Human Mutation*, vol. 31, no. 10, pp. E1709–E1766, 2010.

[32] E. Lenassi, A. Vincent, Z. Li et al., "A detailed clinical and molecular survey of subjects with nonsyndromic USH2A retinopathy reveals an allelic hierarchy of disease-causing variants," *European Journal of Human Genetics*, vol. 23, no. 10, pp. 1318–1327, 2015.

[33] A. Liquori, C. Vaché, D. Baux et al., "Whole USH2A gene sequencing identifies several new deep intronic mutations," *Human Mutation*, vol. 37, no. 2, pp. 184–193, 2016.

[34] A. I. den Hollander, R. K. Koenekoop, S. Yzer et al., "Mutations in the CEP290 (NPHP6) gene are a frequent cause of Leber congenital amaurosis," *The American Journal of Human Genetics*, vol. 79, no. 3, pp. 556–561, 2006.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 3284534, 7 pages
http://dx.doi.org/10.1155/2016/3284534

Research Article

Clinical Performance of an Ultrahigh Resolution Chromosomal Microarray Optimized for Neurodevelopmental Disorders

Karen S. Ho,[1,2] Hope Twede,[1] Rena Vanzo,[1] Erin Harward,[1] Charles H. Hensel,[1] Megan M. Martin,[1] Stephanie Page,[1] Andreas Peiffer,[1,2] Patricia Mowery-Rushton,[1] Moises Serrano,[1] and E. Robert Wassman[1]

[1]*Lineagen, Inc., Salt Lake City, UT, USA*
[2]*Department of Pediatrics, University of Utah, Salt Lake City, UT, USA*

Correspondence should be addressed to E. Robert Wassman; bwassman@lineagen.com

Received 17 June 2016; Revised 27 September 2016; Accepted 20 October 2016

Academic Editor: Urzsula Demkow

Copy number variants (CNVs) as detected by chromosomal microarray analysis (CMA) significantly contribute to the etiology of neurodevelopmental disorders, such as developmental delay (DD), intellectual disability (ID), and autism spectrum disorder (ASD). This study summarizes the results of 3.5 years of CMA testing by a CLIA-certified clinical testing laboratory 5487 patients with neurodevelopmental conditions were clinically evaluated for rare copy number variants using a 2.8-million probe custom CMA optimized for the detection of CNVs associated with neurodevelopmental disorders. We report an overall detection rate of 29.4% in our neurodevelopmental cohort, which rises to nearly 33% when cases with DD/ID and/or MCA only are considered. The detection rate for the ASD cohort is also significant, at 25%. Additionally, we find that detection rate and pathogenic yield of CMA vary significantly depending on the primary indications for testing, the age of the individuals tested, and the specialty of the ordering doctor. We also report a significant difference between the detection rate on the ultrahigh resolution optimized array in comparison to the array from which it originated. This increase in detection can significantly contribute to the efficient and effective medical management of neurodevelopmental conditions in the clinic.

1. Introduction

Neurodevelopmental disabilities, including developmental delay (DD), intellectual disability (ID), and autism spectrum disorder (ASD), affect up to 15% of children [1]. However, in the majority of cases, a child's clinical presentation does not allow for a definitive etiological diagnosis. Copy number variants (CNVs) contribute significantly to the etiology of neurodevelopmental disorders, as well as syndromes of multiple congenital anomalies (MCA). The clinical utility of chromosomal microarray analysis (CMA) for the detection of CNVs associated with these disorders has been recognized by multiple professional societies and has been deemed the first-tier clinical diagnostic test for the evaluation of these disorders [2–6].

Microarrays of various designs and reflective of variable genomic content have been applied to the clinical care of individuals with these conditions; as such, there are varying degrees of diagnostic yield with an increase over time as arrays have evolved [7–17]. The ACMG issued a guideline in 2011 on the optimal design of CMAs and recommended inclusion of additional probe content in areas of known relevance [18]. Most studies reporting on the clinical performance of CMA have been on populations enriched by virtue of the nature of the reporting institution and relative indications for testing.

This study summarizes the results of routine clinical CMA testing in a CLIA-certified laboratory using an array specifically designed to increase detection of CNVs in genomic regions of demonstrated relevance to DD/ID/ASD over a period of 3.5 years.

2. Materials and Methods

2.1. Patient Ascertainment. Data were obtained from a consecutive series of routine clinical samples referred for CMA

to a CLIA-licensed laboratory for etiological diagnosis of DD/ID/ASD and MCAs between July 2012 and December 2015. Patients selectively ascertained and tested as a part of research studies were excluded from these analyses to preclude bias in the observed rates of diagnosis. A second smaller series of 1194 CMAs performed on the same cohort (i.e., identical referral base and underlying patient demographics) with the Affymetrix CytoScan® HD array run during development and local regulatory approval periods is compared here as well to control for the likely ascertainment bias present in previously published reports. Testing indications used here to group patients are defined by the codes routinely provided by referring physicians when ordering tests and are derived from the International Classification of Diseases, Clinical Modification, Revisions 9 or 10, (ICD-9 and ICD-10) from the Centers for Medicare & Medicaid Services (https://www.cms.gov/).

2.2. Microarray Design. The custom microarray [FirstStepDX PLUS® (FSDX PLUS®), Lineagen, Inc.] was utilized in this study in all cases except where specified, and its analytical and clinical validation has been described in detail elsewhere [19]. It is an expanded whole genome chromosomal microarray (CMA) built upon the ultrahigh resolution Affymetrix CytoScan HD platform plus 88,435 custom probes targeting genomic regions strongly associated with ID/DD/ASD [15–24] added under good manufacturing practices (GMP) by Affymetrix using their previously described microarray design process [16]. This resulted in a grand total of 2,784,985 probes. Both copy number (CNV) and single nucleotide polymorphic (SNP) probes are included in the array, which is consistent with the ACMG guideline for CMA design, as is the "enrichment of probes targeting dosage-sensitive genes known to result in phenotypes consistent with common indications for a genomic screen" [18]. Such critical regions that did not contain ≥1 probe/1000 bp on the baseline array were supplemented with additional probe content to provide improved detection of smaller deletions and duplications. Additional probe enrichment targeted genomic regions identified by our prior studies and identified elsewhere in the medical literature. These regions included published copy number variants and individual genes associated with DD/ID/ASD [20–29]. The increase in analytical sensitivity resulting from this additional 3.3% probe content has been calculated to be 2.6% [19].

2.3. CMA Performance and Interpretation. CMA was routinely performed on DNA extracted by standard methodologies from buccal swab samples (ORAcollect®) in a CLIA-certified laboratory. CMA reagents and equipment were as specified by Affymetrix. The established standard cytogenetic criteria for interpretation were routinely applied [30] with minimum of 25-consecutive impacted probes as the baseline determinant for deletions and 50 probes for duplications. Rare CNVs (<1% overall population frequency) were determined to be "pathogenic" if there was sufficient published clinical evidence (at least two independent publications) to indicate that haploinsufficiency or triplosensitivity of the region or gene(s) involved is causative of clinical features. If, however, such clinical evidence was insufficient, but at least some preliminary evidence existed for a causative role for the region or gene(s) therein, and they were not previously categorized as normal population variants in the Database of Genomic Variants (DGV) [31], they were classified as variants of unknown significance (VOUS). Areas of absence of heterozygosity (AOH) were also classified as VOUS if they were of sufficient size and location to increase the risk for conditions with autosomal recessive inheritance or conditions with parent-of-origin/imprinting effects. Cases with no CNVs or only CNVs determined by these criteria to most likely represent normal population variants, for example, contained in databases such as DGV documenting presumptively benign CNVs, were reported as normal.

Table 1: Overall diagnostic yield of 5487 chromosomal microarrays in a routine clinical population.

	CMAs	Pathogenic (% yield)	VOUS (% yield)	Normal (% yield)
Total	5487	506 (9.2)	1109 (20.2)	3872 (70.6)
Female	1558	217 (13.5)	325 (20.2)	1065 (66.3)
Male	3929	342 (8.6)	797 (20.1)	2825 (71.3)

3. Results

3.1. Overall Findings and Diagnostic Yield. A total of 5487 FSDX PLUS CMAs were performed in this time period. There were 1558 females and 3929 males (M : F: 2.5 : 1) tested with a mean age of 7.2 years (median 5.5 years) (Table 1). While largely targeting a pediatric population, a subset of 225 patients was comprised of adults over 18 years old (parental and sibling studies excluded). Based on ICD-9 and ICD-10 codes at the time of referral, 3134 cases represented patients with intellectual (ID) or developmental (DD) disability of varying degrees, 3016 cases represented patients with ASD with or without other features, 743 cases represented patients with multiple congenital anomalies, and 1507 cases represented patients with speech/language delay. Referring physicians were pediatricians (15.0%), medical geneticists (11.2%), pediatric neurologists (40.2%), developmental pediatricians (31.6%), psychiatrists (1.7%), and other medical practitioners (0.4%).

The most common pathogenic findings detected in this unselected population of individuals with neurodevelopmental disorders are shown in Figure 1.

Overall, there were 506 (9.2%) pathogenic abnormalities and 1109 (20.2%) VOUS observed or a 29.4% overall CNV diagnostic yield for potentially abnormal findings (Table 1). However, the yield of pathogenic findings varies significantly on a multivariate basis including but not limited to referring physician specialty, age of patient at testing, patient gender, and referring indication or combination of indications. In addition, a single individual with a reported CNV may have more than one pathogenic CNV, a pathogenic CNV as well as a VOUS, or multiple VOUS findings in the same patient. Patients with any reportable finding had on average 1.2 CNVs

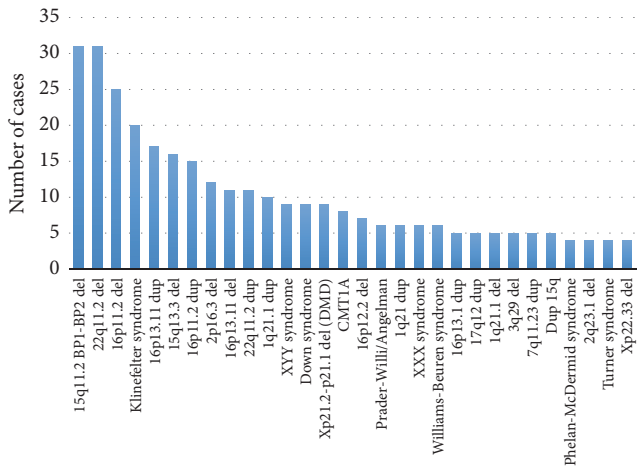


Figure 1: Most common pathogenic findings on 5487 chromosomal microarrays (FSDX PLUS).

Table 2: Multiple CNV are observed in individual patients (mean 1.2 per patient).

	Total number of individual CNVs detected	CNVs per 100 tests (N = 5487)
ABN	734	13.4
VOUS	1272	23.2

per report. Of these, there were 13.4 CNVs classified as pathogenic and 23.2 CNVs classified as a VOUS per 100 CMAs (Table 2).

Rates vary significantly by the specialty of the ordering physician (Table 3), but, regardless of specialty expertise, clinically significant rates of detection were observed in all specialties as well as in the primary care setting. At the extremes were psychiatrists (5.5% diagnostic yield) and medical geneticists (15.5% diagnostic yield), and these groups also differed significantly in the rate of VOUSs (30% and 20%, resp.).

Reported duplications are significantly larger than deletions on average (Table 4). For both duplications and deletions, the average size of pathogenic CNVs was significantly larger than CNVs classified as a VOUS ($p < 0.0001$, two-tailed unpaired t-test).

3.2. Detection Rates by Indication and Age. In patients where the indication for testing was either DD/ID or MCA, the rate of pathogenic CNVs was highest in the first year of life at 16.8% and 21.3%, respectively (Tables 5 and 6). Values were lower but consistent throughout the remainder of childhood but peaked again in the small subset of adult patients tested at levels similar to the first year of life (16.8% and 20.0%, resp.).

Due to the age of clinical recognition, indications including ASD and speech/language deficits were not stratified as to the first year of life separately, but rather with a 0–3.4-year range as the lowest cohort considered. Patients with indications of speech/language deficits demonstrated a gradual rise in the rate of pathogenic findings from the 0–3.4-year-old group (6.7%) to peak in later childhood (12.8%), then dropping slightly in adolescence (10.8%) and reaching a maximum in the adults tested (19.1%) (Tables 7 and 8). VOUS rates were the highest in the youngest cohort (22.2%) and relatively constant in the other age groups but distinctly the lowest in the adults (14.9%).

Individuals with ASD as an indication for testing had a lower pathogenic yield but comparable VOUS rates to other categories (Table 7). The pathogenic rate rose gradually from 3.8% in the youngest cohort (0–3.4 years) to a peak at 8.7% in adolescence. The overall reported CNV rate for individuals with ASD ranged within 22%–29%, again with the peak in adolescents tested. Those with ASD not only had lower, albeit substantial, pathogenic CNV rates than those with other indications but also clearly lowered the rate for all other indications when it was an additional indication; for example, DD/ID/MCA rate when ASD ICD-9/ICD-10 code was excluded was 13.4% (Table 9). The diagnostic yield excluding ASD is significantly higher ($p < 0.0001$) than for the ASD cohort (13.4% compared to 5.9%, resp.).

VOUS rates tended to be relatively constant across groups and with age (18–22%) with the exception of a significantly lower rate in the first year of life for those with DD/ID indication (15.8%), which could be due to the small sample size (Table 5), and adults with MCAs or speech/language deficits (14.3% and 14.9%, resp.) (Tables 6 and 8). Those with MCAs also showed higher peak rates of 24.6% and 25% in the early childhood (1–3.4 years) and late childhood (5.5–10.1 years) cohorts and a dip, again potentially due to small sample size in this group, to 15.8% between these ages (Table 6).

3.3. Comparison to Detection on Baseline Array. Detection rates in the same overall cohort (i.e., same referral base, underlying patient demographics timeframe, laboratories, and interpretation process and criteria) on the CytoScan HD array (N = 1194), which was the baseline for FSDX PLUS, were lower than those in this series diagnosed on the custom FSDX PLUS array (9.0% pathogenic CNV and 14.2% VOUS compared to 9.2% and 20.2%, resp.) (Table 10).

4. Discussion

CMA is the guideline-recognized first-tier test in the evaluation of MCA, DD/ID, and ASD, [2–6] and yields significant rates of abnormal or potentially abnormal (VOUS) results [7–17] with clinical utility for the management of individuals with these disorders [28, 29, 32–39]. Since the introduction of this technology, the total genomic content in terms of probes on CMAs has progressively increased, leading to higher diagnostic yields and resolution of abnormalities [10, 14–17] with corresponding increases in clinical value of these tests [32–40]. In addition to guidelines on the clinical indications for CMA, ACMG has issued guidance on the appropriate content and design of such arrays and specifically opined that "it is desirable to have enrichment of probes targeting dosage-sensitive genes known to result in phenotypes consistent with common indications for a genomic screen (e.g., intellectual disability, developmental delays, autism, and congenital anomalies)" [18]. We report here on over three years' experience

Table 3: Diagnostic yield and mean patient age vary significantly by the specialty of the ordering physician.

Specialty	% total CMAs ordered	Average age (years) [6.4 overall]	Pathogenic % yield	VOUS % yield	Normal %
Pediatric neurology	40.2%	6.5	8.2%	20.0%	71.8%
Developmental and behavioral pediatrics	31.6%	6.1	7.1%	20.6%	72.3%
Pediatrics	15.0%	6.8	11.2%	17.6%	71.2%
Genetics	11.2%	6.0	15.5%	20.1%	64.4%
Psychiatry	1.7%	10.7	5.5%	29.7%	64.8%
Other specialties	0.4%	8.3	13.6%	18.2%	68.2%

Table 4: Clinically reported duplications are significantly larger than deletions on average.

		Deletions	Duplications
Pathogenic CNVs	Average size (kb)	3,284 (N = 474)	8,105 (N = 258)
	Median size (kb)	1,418	1,680
VOUS CNVs	Average size (kb)	308 (N = 584)	528 (N = 751)
	Median size (kb)	129	357

Table 5: Diagnostic yield by age in ID/DD (986 females and 2148 males, total n = 3134).

Age in years	Total tests	Pathogenic (% yield)	VOUS (% yield)	Normal (%)
0-1	95	16 (16.8%)	15 (15.8%)	64 (67.4%)
1–3.4	950	87 (9.2%)	188 (19.8%)	675 (71.1%)
3.5–5.4	572	54 (9.4%)	103 (18.0%)	415 (72.6%)
5.5–10.0	775	92 (11.9%)	152 (19.6%)	531 (68.5%)
10.1–18	623	65 (10.4%)	117 (18.8%)	441 (70.8%)
18+	119	20 (16.8%)	26 (21.8%)	73 (61.3%)
Total	3134	334 (10.7%)	601 (19.2%)	2199 (70.2%)

Table 6: Diagnostic yield by age in MCA (289 females and 454 males, total n = 743).

Age buckets	Total tests	Pathogenic (% yield)	VOUS (% yield)	Normal (%)
0-1 years	122	26 (21.3%)	23 (18.9%)	73 (59.8%)
1–3.4 years	179	29 (16.2%)	44 (24.6%)	106 (59.2%)
3.5–5.4 years	95	14 (14.7%)	15 (15.8%)	66 (69.5%)
5.5–10.4 years	164	30 (18.3%)	41 (25.0%)	93 (56.7%)
10.5–18	148	28 (18.9%)	29 (19.6%)	91 (61.5%)
18+	35	7 (20.0%)	5 (14.3%)	23 (65.7%)
Total	743	134 (18.0%)	157 (21.1%)	452 (60.8%)

with a unselected clinical referral base on a CMA specifically designed to extend the scope of detection for individuals with ASD and other neurodevelopmental disorders through the addition of probes targeting genomic regions more recently identified as of pathogenic relevance to these disorders.

Table 7: Diagnostic yield by age in ASD (622 females and 2394 males, total n = 3016).

Age in years	Number of tests	Pathogenic (% yield)	VOUS (% yield)	Normal (%)
0–3.4	735	28 (3.8%)	134 (18.2%)	573 (78.0%)
3.5–5.4	688	33 (4.8%)	121 (17.6%)	534 (77.6%)
5.5–10	789	50 (6.3%)	158 (20.0%)	581 (73.6%)
10.1–18	679	59 (8.7%)	138 (20.3%)	482 (71.0%)
18+	125	8 (6.4%)	25 (20%)	92 (73.6%)
Total	3016	178 (5.9%)	576 (19%)	2262 (75%)

Table 8: Diagnostic yield by age in speech/language deficits (427 females and 1080 males, total n = 1507).

Age buckets	Total tests	Pathogenic (% yield)	VOUS (% yield)	Normal (%)
0–3.4 years	449	30 (6.7%)	100 (22.2%)	319 (71.0%)
3.5–5.4 years	331	27 (8.2%)	63 (19.0%)	241 (72.8%)
5.5–10.4 years	420	52 (12.4%)	89 (21.2%)	279 (66.4%)
10.5–18	260	28 (10.8%)	50 (19.2%)	182 (70.0%)
18+	47	9 (19.1%)	7 (14.9%)	31 (66.0%)
Total	1507	146 (9.7%)	309 (20.5%)	1052 (69.8%)

Our data demonstrate that diagnostic yield is a complex multivariate function dependent upon several clinical variables including the patient's clinical diagnosis/presentation, age at testing, and referring physician specialty training. An unselected consecutive referral base, with a substantial non-specialty physician referral component, lack of bias toward selected subgroups (e.g., exclusion of research enriched population of WHS/4p-cohort in the present series) [41], and the active offering of testing to the most recent clinical indication for CMA, ASD, which has an expectably lower rate of such findings [13–15], would be expected to result in a lower overall diagnostic yield in the present series. However, the overall detection rate for clinically established pathogenic CNVs of 9.2% is equivalent or higher than other reported series/platforms [7–17] despite the inherent bias toward lower rates based on the unselected referral base and focus on ASD. An internal comparison to cases run on the standard array (CytoScan HD) which was the baseline for development of

Table 9: Diagnostic yield by age in neurodevelopmental disorders and/or MCA, excluding ASD (females = 909; males = 1486; total $n = 2395$).

Age in years	Total (excluding ASD)	Pathogenic (% yield)	VOUS (% yield)	Normal (%)
0–1	204	38 (18.6%)	37 (18.1%)	129 (63.2%)
1–3.4	699	84 (12.0%)	146 (20.9%)	469 (67.1%)
3.5–5.4	344	43 (12.5%)	63 (18.3%)	238 (69.2%)
5.5–10	589	82 (13.9%)	121 (20.5%)	386 (65.5%)
10.1–18	461	55 (11.9%)	83 (18.0%)	323 (70.1%)
18+	98	19 (19.4%)	15 (15.3%)	64 (65.3%)
Total	2395	321 (13.4%)	465 (19.4%)	1609 (67.2%)

Table 10: Comparison of FSDX ($N = 5487$) to CytoScan HD ($N = 1194$) arrays performed on same ascertainment base and interpretation paradigm.

Array	Pathogenic yield	VOUS yield	Normal
FSDX PLUS ($N = 5487$)	9.2%	20.2%	70.6%
CytoScan HD ($N = 1172$)	9.0%	14.2%	76.7%

the FSDX PLUS array showed a slight, but not significant, increase in detection rate for pathogenic variants from 9.0% to 9.2% over the same referral base and underlying patient demographics, using the same interpretation paradigm. The same comparative analysis showed a highly significant differential in detection of VOUS from 14.2% to 20.2% (Chi-squared p value < 0.0001). The analytical sensitivity of the FSDX PLUS array was recently calculated to be at least 2.6% greater than the baseline array, which is generally consistent with the observed increase in the overall rate of reportable CNVs (pathogenic plus VOUS) [19].

When individuals with ASD are excluded so as to more closely match populations reported for other CMA platforms/series, the diagnostic yield is further differentiated with diagnostic yields of 13.4% pathogenic and 19.4% VOUS and a total detection rate for potentially causative variants, of nearly 33%. It is likely that, even after this correction, other enrichment biases remain in comparing other series to this one.

While significantly lower than the overall population or the ASD-excluded subpopulation ($p < 0.0001$), the diagnostic yield in ASD cases of 5.9% pathogenic and 19.0% VOUS exceeds those previously reported [13–15] and supports the value of incremental targeted content for areas of clinical relevance in this important setting for CMA.

The variations in diagnostic yield evident in subgroup analyses may in turn contain clues for future research and causation. For example, the rise in rate of detected abnormalities in the ASD population with age suggests that earlier use of CMA and perhaps other genetic testing may be important. It is estimated that at least 20% of ASD individuals have an underlying genetic syndrome, but a survey of a large autism center showed that less than 10% of their population had received any form of genetic evaluation [42, 43].

Not surprisingly, patients who are tested in their first year of life for most "indication" groups have the highest diagnostic yield. This is likely due to the probability that increased severity of features would prompt physician investigation earlier in life. It is, however, remarkable that adults (>18 years old) tested also have such a high pathogenic CNV rate observed. This could be due to the relatively small size of this cohort. Alternatively, it may be more reflective of severity in that particular age group. For example, clinicians/families might believe that testing is not as valuable for adults but perform it anyway when the individual is considered to be relatively severely impaired.

In addition to clinically well-defined pathogenic CNVs, a variety of CNVs of less obvious correlation with causation are routinely found on all CMAs. Efforts to better identify and biologically define the relevance of VOUS in these disorders have critical importance to understanding disease mechanisms and, ultimately, give insight to appropriate medical management in the future. An increased rate of CNVs classified as VOUS is therefore of potential clinical importance. Furthermore, VOUS results have been clearly demonstrated to be of great importance to parents of patients with DD/ID/ASD [44–47].

While earlier literature did not typically consider VOUS in the diagnostic yield, this was due to inconsistent criteria for reporting, lack of established databases of normal population variants, and limited sharing of data [12, 13]. Today with these tools better established, it is common and reasonable to consider VOUS in an overall diagnostic yield [9, 32] as many of these variants will evolve into clearly pathogenic finding based on emerging clinical experience and represent an exciting and abundant opportunity to better understand the full range of genomic abnormalities contributing to the neurodevelopmental phenotypes.

Numerous studies have now demonstrated the clinical actionability and utility of CMA testing [32–40]. The increased yield of an optimized array as described here will extend the range and scope of this utility, and it is readily demonstrated through relevant case studies and series to date [35–40]. Of critical importance is the ongoing evaluation of novel methods to assess the potential role of VOUS findings in the underlying pathology of individual patients to realize the maximum benefit of the increased detection rate achieved through array and interpretation optimization.

Competing Interests

All authors are employees of Lineagen, Inc., which is a clinical reference laboratory performing genetic testing for individuals with neurodevelopmental disorders.

Acknowledgments

The authors wish to thank the remarkable patients, families, and providers who participated in this work, Suresh Venkatasubramanian for developing custom code to aid in data analysis, Sean Dixon and Kenny Lentz for database management and technical support, and the entire Lineagen team for their advocacy and dedication to those with neurodevelopmental disorders.

References

[1] C. A. Boyle, S. Boulet, L. A. Schieve et al., "Trends in the prevalence of developmental disabilities in US children, 1997–2008," *Pediatrics*, vol. 127, no. 6, pp. 1034–1042, 2011.

[2] M. Manning and L. Hudgins, "Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities," *Genetics in Medicine*, vol. 12, no. 11, pp. 742–745, 2010.

[3] G. B. Schaefer and N. J. Mendelsohn, "Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions," *Genetics in Medicine*, vol. 15, no. 5, pp. 399–407, 2013.

[4] F. Volkmar, M. Siegel, M. Woodbury-Smith, B. King, J. McCracken, and M. State, "Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder," *Journal of the American Academy of Child and Adolescent Psychiatry*, vol. 53, no. 2, pp. 237–257, 2014.

[5] J. B. Moeschler, M. Shevell, R. A. Saul et al., "Comprehensive evaluation of the child with intellectual disability or global developmental delays," *Pediatrics*, vol. 134, no. 3, pp. e903–e918, 2014.

[6] D. J. Michelson, M. I. Shevell, E. H. Sherr, J. B. Moeschler, A. L. Gropman, and S. Ashwal, "Evidence Report: genetic and metabolic testing on children with global developmental delay: report of the quality standards Subcommittee of the American Academy of Neurology and the Practice Committee of the Child Neurology Society," *Neurology*, vol. 77, no. 17, pp. 1629–1635, 2011.

[7] K. M. Heil and C. P. Schaaf, "The genetics of autism spectrum disorders—a guide for clinicians," *Current Psychiatry Reports*, vol. 15, no. 1, article 334, 2013.

[8] L. Bernardini, V. Alesi, S. Loddo et al., "High-resolution SNP arrays in mental retardation diagnostics: how much do we gain," *European Journal of Human Genetics*, vol. 18, no. 2, pp. 178–185, 2010.

[9] S. G. McGrew, B. R. Peters, J. A. Crittendon, and J. Veenstra-VanderWeele, "Diagnostic yield of chromosomal microarray analysis in an autism primary care practice: which guidelines to implement?" *Journal of Autism and Developmental Disorders*, vol. 42, no. 8, pp. 1582–1591, 2012.

[10] K. B. Howell, A. J. Kornberg, A. S. Harvey et al., "High resolution chromosomal microarray in undiagnosed neurological disorders," *Journal of Paediatrics and Child Health*, vol. 49, no. 9, pp. 716–724, 2013.

[11] A. Battaglia, V. Doccini, L. Bernardini et al., "Confirmation of chromosomal microarray as a first-tier clinical diagnostic test for individuals with developmental delay, intellectual disability, autism spectrum disorders and dysmorphic features," *European Journal of Paediatric Neurology*, vol. 17, no. 6, pp. 589–599, 2013.

[12] D. T. Miller, M. P. Adam, S. Aradhya et al., "Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies," *American Journal of Human Genetics*, vol. 86, no. 5, pp. 749–764, 2010.

[13] Y. Shen, K. A. Dies, I. A. Holm et al., "Clinical genetic testing for patients with autism spectrum disorders," *Pediatrics*, vol. 125, no. 4, pp. e727–e735, 2010.

[14] L. Edelmann and K. Hirschhorn, "Clinical utility of array CGH for the detection of chromosomal imbalances associated with mental retardation and multiple congenital anomalies," *Annals of the New York Academy of Sciences*, vol. 1151, pp. 157–166, 2009.

[15] A. L. Beaudet, "The utility of chromosomal microarray analysis in developmental and behavioral pediatrics," *Child Development*, vol. 84, no. 1, pp. 121–132, 2013.

[16] H. Mason-Suares, W. Kim, L. Grimmett et al., "Density matters: comparison of array platforms for detection of copy-number variation and copy-neutral abnormalities," *Genetics in Medicine*, vol. 15, no. 9, pp. 706–712, 2013.

[17] R. Pfundt, K. Kwiatkowski, A. Roter et al., "Clinical performance of the CytoScan Dx Assay in diagnosing developmental delay/intellectual disability," *Genetics in Medicine*, vol. 18, no. 2, pp. 168–173, 2016.

[18] H. M. Kearney, S. T. South, D. J. Wolff, A. Lamb, A. Hamosh, and K. W. Rao, "American College of Medical Genetics recommendations for the design and performance expectations for clinical genomic copy number microarrays intended for use in the postnatal setting for detection of constitutional abnormalities," *Genetics in Medicine*, vol. 13, no. 7, pp. 676–679, 2011.

[19] C. H. Hensel, R. Vanzo, M. Martin et al., "Analytical and clinical validity study of FirstStep^{Dx} PLUS: a chromosomal microarray optimized for patients with neurodevelopmental conditions," *PLoS Currents*, 2016.

[20] N. Matsunami, D. Hadley, C. H. Hensel et al., "Identification of rare recurrent copy number variants in high-risk autism families and their prevalence in a large ASD population," *PLoS ONE*, vol. 8, no. 1, Article ID e52239, 2013.

[21] J. Sebat, B. Lakshmi, D. Malhotra et al., "Strong association of de novo copy number mutations with autism," *Science*, vol. 316, no. 5823, pp. 445–449, 2007.

[22] C. R. Marshall, A. Noor, J. B. Vincent et al., "Structural variation of chromosomes in autism spectrum disorder," *American Journal of Human Genetics*, vol. 82, no. 2, pp. 477–488, 2008.

[23] S. L. Christian, C. W. Brune, J. Sudi et al., "Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder," *Biological Psychiatry*, vol. 63, no. 12, pp. 1111–1117, 2008.

[24] J. T. Glessner, K. Wang, G. Cai et al., "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes," *Nature*, vol. 459, no. 7246, pp. 569–573, 2009.

[25] M. Bucan, B. S. Abrahams, K. Wang et al., "Genome-wide analyses of exonic copy number variants in a family-based study point to novel autism susceptibility genes," *PLoS Genetics*, vol. 5, no. 6, Article ID e1000536, 2009.

[26] D. Pinto, A. T. Pagnamenta, L. Klei et al., "Functional impact of global rare copy number variation in autism spectrum disorders," *Nature*, vol. 466, no. 7304, pp. 368–372, 2010.

[27] P. Szatmari, A. D. Paterson, L. Zwaigenbaum et al., "Mapping autism risk loci using genetic linkage and chromosomal rearrangements," *Nature Genetics*, vol. 39, no. 3, pp. 319–328, 2007.

[28] L. A. Weiss, Y. Shen, J. M. Korn et al., "Association between microdeletion and microduplication at 16p11.2 and autism," *New England Journal of Medicine*, vol. 358, no. 7, pp. 667–675, 2008.

[29] M.-L. Jacquemont, D. Sanlaville, R. Redon et al., "Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders," *Journal of Medical Genetics*, vol. 43, no. 11, pp. 843–849, 2006.

[30] S. T. South, C. Lee, A. N. Lamb, A. W. Higgins, and H. M. Kearney, "ACMG Standards and Guidelines for constitutional cytogenomic microarray analysis, including postnatal and prenatal applications: revision 2013," *Genetics in Medicine*, vol. 15, no. 11, pp. 901–909, 2013.

[31] J. R. MacDonald, R. Ziman, R. K. C. Yuen, L. Feuk, and S. W. Scherer, "The database of genomic variants: a curated collection of structural variation in the human genome," *Nucleic Acids Research*, vol. 42, no. 1, pp. D986–D992, 2014.

[32] J. L. Roberts, K. Hovanes, M. Dasouki, A. M. Manzardo, and M. G. Butler, "Chromosomal microarray analysis of consecutive individuals with autism spectrum disorders or learning disability presenting for genetic services," *Gene*, vol. 535, no. 1, pp. 70–78, 2014.

[33] J. Saam, J. Gudgeon, E. Aston, and A. R. Brothman, "How physicians use array comparative genomic hybridization results to guide patient management in children with developmental delay," *Genetics in Medicine*, vol. 10, no. 3, pp. 181–186, 2008.

[34] M. E. Coulter, D. T. Miller, D. J. Harris et al., "Chromosomal microarray testing influences medical management," *Genetics in Medicine*, vol. 13, no. 9, pp. 770–776, 2011.

[35] J. W. Ellison, J. B. Ravnan, J. A. Rosenfeld et al., "Clinical utility of chromosomal microarray analysis," *Pediatrics*, vol. 130, no. 5, pp. e1085–e1095, 2012.

[36] E. R. Riggs, K. E. Wain, D. Riethmaier et al., "Chromosomal microarray impacts clinical management," *Clinical Genetics*, vol. 85, no. 2, pp. 147–153, 2014.

[37] L. B. Henderson, C. D. Applegate, E. Wohler, M. B. Sheridan, J. Hoover-Fong, and D. A. S. Batista, "The impact of chromosomal microarray on clinical management: a retrospective analysis," *Genetics in Medicine*, vol. 16, no. 9, pp. 657–664, 2014.

[38] V. Q. Tao, K. Y. K. Chan, V. W. Y. Chu et al., "The clinical impact of chromosomal microarray on paediatric care in Hong Kong," *PLoS ONE*, vol. 9, no. 10, Article ID e109629, 2014.

[39] M. R. Sdano, R. J. Vanzo, M. M. Martin et al., "Clinical utility of chromosomal microarray analysis of dna from buccal cells: detection of mosaicism in three patients," *Journal of Genetic Counseling*, vol. 23, no. 6, pp. 922–927, 2014.

[40] M. M. Martin, R. J. Vanzo, M. R. Sdano, A. L. Baxter, and S. T. South, "Mosaic deletion of 20pter due to rescue by somatic recombination," *American Journal of Medical Genetics Part A*, vol. 170, no. 1, pp. 243–248, 2016.

[41] K. S. Ho, S. T. South, A. Lortz et al., "Chromosomal microarray testing identifies a 4p terminal region associated with seizures in Wolf–Hirschhorn syndrome," *Journal of Medical Genetics*, vol. 53, no. 4, pp. 256–263, 2016.

[42] F. Gurrieri, "Working up autism: the practical role of medical genetics," *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, vol. 160, no. 2, pp. 104–110, 2012.

[43] T. L. Wenger, C. Kao, D. M. McDonald-McGinn et al., "The role of mGluR copy number variation in genetic and environmental forms of syndromic autism spectrum disorder," *Scientific Reports*, vol. 6, article 19372, 2016.

[44] M. Reiff, B. A. Bernhardt, S. Mulchandani et al., "'What does it mean?': uncertainties in understanding results of chromosomal microarray testing," *Genetics in Medicine*, vol. 14, no. 2, pp. 250–258, 2012.

[45] M. Reiff, E. Giarelli, B. A. Bernhardt et al., "Parents' perceptions of the usefulness of chromosomal microarray analysis for children with autism spectrum disorders," *Journal of Autism and Developmental Disorders*, vol. 45, no. 10, pp. 3262–3275, 2015.

[46] S. Jez, M. Martin, S. South, R. Vanzo, and E. Rothwell, "Variants of unknown significance on chromosomal microarray analysis: parental perspectives," *Journal of Community Genetics*, vol. 6, no. 4, pp. 343–349, 2015.

[47] E. J. Wilkins, A. D. Archibald, M. A. Sahhar, and S. M. White, "'It wasn't a disaster or anything': Parents' experiences of their child's uncertain chromosomal microarray result," *American Journal of Medical Genetics Part A*, vol. 170, no. 11, pp. 2895–2904, 2016.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 1298198, 10 pages
http://dx.doi.org/10.1155/2016/1298198

Research Article

G-1639A but Not C1173T *VKORC1* Gene Polymorphism Is Related to Ischemic Stroke and Its Various Risk Factors in Ukrainian Population

Yevhen I. Dubovyk,[1] Viktoriia Yu. Harbuzova,[2] and Alexander V. Ataman[1]

[1] *Department of Physiology, Pathophysiology and Medical Biology, Sumy State University, Sumy 40007, Ukraine*
[2] *Scientific Laboratory of Molecular Genetic Research, Sumy State University, Sumy 40007, Ukraine*

Correspondence should be addressed to Yevhen I. Dubovyk; janitor@ukr.net

Received 28 May 2016; Revised 25 July 2016; Accepted 17 August 2016

Academic Editor: Emin Karaca

Vitamin K epoxide reductase complex subunit 1 (VKORC1) is integral 163-amino acid long transmembrane protein which mediates recycling of vitamin K 2,3-epoxide to vitamin K hydroquinone and it is necessary for activation of vitamin K-dependent proteins (VKDPs). Herein, the association between G-1639A (rs9923231) and C1173T (rs9934438) single-nucleotide polymorphisms (SNPs) of the *VKORC1* gene and ischemic stroke (IS) was tested in Ukrainian population. Genotyping was performed in 170 IS patients and 124 control subjects (total 294 DNA samples) using PCR-RFLP (polymerase chain reaction with following restriction fragment length polymorphism analysis) method. Our data showed that G-1639A but not C1173T polymorphism was related to IS, regardless of adjustment for age, sex, body mass index, smoking status, and arterial hypertension. The risk for IS in -1639A allele carriers (OR = 2.138, $P = 0.015$) was higher than in individuals with G/G genotype. Haplotype analysis demonstrated that -1639G/1173T and -1639A/1173C were related to increased risk for IS (OR = 3.813, $P = 0.010$, and OR = 2.189, $P = 0.011$, resp.), while -1639G/1173C was a protective factor for IS (OR = 0.548, $P < 0.001$). Obtained results suggested that -1639A allele can be a possible genetic risk factor for IS in Ukrainian population.

1. Introduction

A large number of proteins require posttranslational modification for further activation. The one way of such modification is a change of multiple glutamic acid residues to γ-carboxyglutamate in the peptide sequences of proteins (γ-carboxylation). The biochemical system which is responsible for carrying out this modification is called the vitamin K cycle [1]. The functioning of this cycle results in the oxidation of vitamin K hydroquinone to vitamin K 2,3-epoxide and it is impossible without recycling of vitamin K 2,3-epoxide to vitamin K hydroquinone. The enzyme which mediates recycling of vitamin K 2,3-epoxide to vitamin K hydroquinone is called vitamin K epoxide reductase complex subunit 1 (VKORC1).

VKORC1 is integral 163-amino acid long transmembrane protein (18 kDa) which is widely expressed in many organs and tissues of the human and animal organisms (liver, salivary gland, prostate, lung, kidney, brain, bone, skeletal muscle, heart, etc.) [2]. It is necessary for activation of vitamin K-dependent proteins (VKDPs), which undergo posttranslation modification in vitamin K cycle. It is known that VKDPs include a number of clotting factors involved in the coagulation cascade (factors II, VII, IX, and X), anticoagulants (proteins C, S, and Z), proteins involved in bone and soft-tissue mineralization (matrix Gla-protein (MGP), Gla-rich protein (GRP), and osteocalcin) [3, 4], protein involved in differentiation of vascular smooth muscle cells (VSMCs), and platelet activation (growth arrest-specific 6 (GAS6)) [5]. Consequently, it can be assumed that dysfunction of VKORC1 might cause activity reduction of VKDPs and thus might lead to thrombosis, calcification, and inflammation of the vascular wall, and so forth. Such changes are essential steps for development of atherosclerotic lesions of the brain arteries, which often lead to ischemic stroke.

Ischemic stroke (IS) is a multifactorial disease at which development is determined by environmental and genetic factors. Since the discovery of the *VKORC1* gene in 2004 [6, 7], numerous studies about the relation of various single-nucleotide polymorphisms (SNPs) of the *VKORC1* gene to development of cardiovascular and cerebrovascular diseases were conducted [8–12]. There are some studies where association of the *VKORC1* G-1639A, C1173T, and T2255C polymorphisms with IS in a Chinese population has been investigated [13–15], but the data obtained in other ethnic groups remain controversial [16–20]. Role studying of the *VKORC1* G-1639A and C1173T polymorphisms in development of IS in Ukrainian and other Slavic populations has not been conducted. Thus, we have performed a case-control study on representatives of Ukrainian population with the aim of investigating the possible association of the *VKORC1* G-1639A and C1173T SNPs with IS in individuals who had different risk factors of atherosclerosis.

2. Materials and Methods

2.1. Subjects. In present study we selected 170 unrelated Ukrainian patients (42.4% women and 57.6% men) from 40 to 85 years of age (mean age (±SD) 64.7±9.5) who had IS and had been under medical surveillance and outpatient treatment in the 5th Sumy Clinical Hospital since 2009 to 2011. A final diagnosis of IS was established on the basis of clinical, computed tomography, and magnetic resonance imaging investigations. Each case of IS was assessed according to the TOAST (Trial of Org 10172 in Acute Stroke Treatment) criteria [21] on the basis of anamnestic data and peculiarities of the clinical disease circuit, as well as the data of ultrasonic Doppler sonography of the main head artery and electrocardiograms. The patients with IS of cardioembolic origin and undetermined etiology were excluded from the studied group. The clinical characteristics of patients with IS included arterial blood pressure (BP), body mass index (BMI), composition of blood plasma lipoproteins, and indices of blood coagulation. The total cholesterol, HDL-cholesterol, LDL-cholesterol, and triglyceride levels determination was done in 157 with IS. Thus, the analysis of VKORC1 polymorphisms influence on lipid metabolism in stroke patients was performed only between these 157 cases.

The control group included 124 individuals with the absence of ischemic stroke and other cerebrovascular pathologies, which was verified using amnestic data, ECG test, blood pressure measurement, and carrying out generally accepted neurologic researches. It is well known that warfarin and similar oral anticoagulants are inhibitors of VKORC1. To avoid distortion of results about association between *VKORC1* gene polymorphism and IS only individuals who have never taken anticoagulant therapy were included to the case and control groups.

The patients of both groups were divided into the pairs of subgroups defined by sex, BMI (BMI $< 25\,\mathrm{kg/m^2}$ and $\geq 25\,\mathrm{kg/m^2}$), and BP (nonhypertensive or hypertensive: systolic BP > 140 mmHg, diastolic BP > 90 mmHg).

The study was complied with the Declaration of Helsinki and approved by the Ethic Committee of Medical Institute of Sumy State University. Written informed consent from all the subjects was obtained before enrollment.

2.2. Genotyping of SNPs. Genomic DNA was extracted from white cells using GeneJET Whole Blood Genomic DNA Purification Mini Kit (Thermo Fisher Scientific, USA) according to the manufacturer's protocol. *VKORC1* promoter G-1639A (rs9923231) and first intron C1173T (rs9934438) polymorphisms genotyping was performed using PCR-RFLP. We used primers synthesized by Metabion (Germany). The reaction mixture of 25 mL volume contained 50–100 ng of DNA, 1.5 mM magnesium sulfate, 200 mM of each dNTP, 5 μL 5x PCR-buffer, 20 pM of each primer, and 0.5 U of Taq DNA polymerase (Thermo Fisher Scientific, USA). PCR was carried out in a thermocycler GeneAmp PCR System 2700 (Thermo Fisher Scientific, USA).

The sequence of nucleotides in specific primers for *VKORC1* promoter G-1639A SNP was as follows: forward – 5′-GCCAGCAGGAGAGGGAAATA-3′ and reverse – 5′-AGTTTGGACTACAGGTGCCT-3′. Thermocycling conditions consisted of 94°C for 5 min, followed by 33 cycles of 94°C for 50 s, 61°C for 45 s, and 72°C for 50 s with a final extension step of 72°C for 5 min. For restriction analysis 6 μL of the amplification products was incubated at 37°C for 20 h with 5 U *Msp*I (*Hpa*II) (Thermo Fisher Scientific, USA). In the case of guanine at position -1639 of the *VKORC1* promoter amplified fragment, which consisted of 290, bps was cut by *Msp1* into two fragments of 168 and 122 bps. Guanine to adenine substitution resulted in the loss of *Msp1* restriction site and fragment of the promoter (290 bps) could not be cleaved.

The polymorphism of the *VKORC1* first intron (C1173T) was analyzed using the following primers: forward – 5′-AAGATGAAAAGCAGGGCCTAC-3′, reverse – 5′-CCG-AGAAAGGTGATTTCCAA-3′. Thermocycling conditions consisted of 94°C for 5 min, followed by 33 cycles of 94°C for 50 s, 60°C for 50 s, and 72°C for 55 s with a final extension step of 72°C for 5 min. 6 μL of the amplification products (195 bps) was incubated at 37°C for 18 h with 3 U *Sty*I (*Eco130*I) (Thermo Fisher Scientific, USA). The presence of cytosine at position 1173 of the gene prevented the restriction and in the case of substitution for thymine *Sty*I cleaved the amplified fragment into two fragments 125 and 70 bps in length.

The restriction fragments were separated by horizontal electrophoresis (electrical field strength 10 V/cm) in 1.5% agarose gel containing 10 mg/mL ethidium bromide. Visualization of DNA fragments after electrophoresis was performed using ultraviolet transillumination.

2.3. Statistical Analysis. Most statistical analyses were performed using Statistical Package for Social Science software (SPSS, version 17.0, Chicago, IL, USA). Continuous data are expressed as mean ± SD; categorical data are presented as number and percentage value. All continuous variables were normally distributed (Kolmogorov-Smirnov and Shapiro-Wilk tests); thus the comparison between the groups was performed using ANOVA or two-tailed Student's t-test. Bonferroni's correction was used for multiple comparisons. Each SNP was tested for deviation from

Table 1: Clinical characteristics of the study groups.

Parameter	IS group (N = 170)	Control group (N = 124)	P
Age, years	64.7 ± 9.5	76.7 ± 10.2	<0.001
Sex, male/female	72/98	45/79	0.294
Current smokers, N (%)	50 (29.4)	31 (25.0)	0.403
BMI, kg/m^2	28.2 ± 4.3	27.6 ± 5.0	0.279
Systolic BP, mmHg	167 ± 29.2	152.6 ± 23.4	<0.001
Diastolic BP, mmHg	95.4 ± 15.6	86.3 ± 12.4	<0.001
Fasting glucose, mmol/L	5.92 ± 1.5	5.29 ± 0.7	<0.001

Categorical variables were compared by χ^2 test and continuous variables by t-test.

Hardy-Weinberg equilibrium using the Online Encyclopedia for Genetic Epidemiology Studies (http://www.oege.org/software/hardy-weinberg.html). The χ^2 test was used to compare genotype and haplotype distributions of *VKORC1* SNPs between case and control groups. To estimate the risk we calculated the odds ratio (OR) and 95% confidence interval (CI) for the four models of inheritance: dominant (major homozygous genotype as a reference), recessive (genotypes with major allele as a reference), overdominant (major and minor homozygous genotypes as a reference), and additive (heterozygous genotype and minor homozygous genotype with major homozygous genotype as a reference). The Akaike information criterion (AIC) was used for selecting the most probable inheritance model. Such risk factors for IS like age, sex, BMI, smoking status, and arterial hypertension were incorporated as covariates by multivariable logistic regression analysis. Linkage disequilibrium (LD) and haplotype frequencies were analyzed by Arlequin (version 3.1, University of Berne, Bern, Switzerland). All statistical tests were two-sided; $P < 0.05$ was considered significant.

3. Results

The clinical characteristics of 170 cases and 124 controls are shown in Table 1. No significant differences between the groups with respect to gender, smoking status, and BMI were noted (P = 0.294, 0.403, and 0.279, resp.), but the average age of the control group (76.7 ± 10.2) was considerably higher than in the case group ($P < 0.001$). The average level of fasting glucose and the average meaning of systolic and diastolic BP were higher in IS group than in controls ($P < 0.001$).

The genotype distributions of the two SNPs (G-1639A and C1173T) in controls (minor allele frequency (MAF) = 0.371 and MAF = 0.327, resp.) and cases (MAF = 0.476 and MAF = 0.412, resp.) were consistent with the Hardy-Weinberg equilibrium ($P > 0.05$).

The results of *VKORC1* G-1639A genotyping are shown in Table 2. The difference in the distribution of three genotypes (G/G, G/A, and A/A) between the cases and controls was significant (P = 0.027). When analyzed in women and men independently, the significant difference in genotype distribution was not revealed (P = 0.228 and P = 0.119, resp.). Division of IS patients into subgroups according to the presence or absence of known atherosclerosis risk factors allowed carrying out a comparative analysis of their genotype frequencies. Statistically significant differences for groups with overweight (BMI ≥ 25 kg/m^2) and arterial hypertension were established (P = 0.025 and P = 0.003, resp.).

Analysis of G-1639A genotypic association with IS under the four common models of inheritance is presented in Table 3. Significant association was established in total group before and after adjusting for age, gender, smoking status, BMI, and arterial hypertension under dominant (P_{obs} = 0.009, P_{adj} = 0.015) and additive model (P_{obs} = 0.032, P_{adj} = 0.041 for G/A genotype, and P_{obs} = 0.017, P_{adj} = 0.028 for A/A genotype). Relative risk analysis estimated an increased risk for IS in minor allele carriers (OR_{adj} = 2.138, 95% CI = 1.157–3.953) and separately in patients with G/A (OR_{adj} = 1.979, 95% CI = 1.029–3.805) and A/A (OR_{adj} = 2.621, 95% CI = 1.110–6.191) genotypes compared for individuals with G/G genotype. Genotypic association of G-1639A was also revealed in women after adjustment for covariates of age, BMI, smoking status, and arterial hypertension under dominant (P_{adj} = 0.038, OR_{adj} = 2.848, 95% CI = 1.058–7.665) and additive (P_{adj} = 0.049, OR_{adj} = 2.888, 95% CI = 1.006–8.293) model. In men significant difference was present only in the crude additive model (P_{obs} = 0.049, OR_{obs} = 2.240, 95% CI = 1.002–5.007) but was lost after adjustment (P_{adj} = 0.077). The association between G-1639A SNP and IS was also found to be significant in individuals with BMI ≥ 25 kg/m^2 under dominant model with or without the adjustment for gender, age, smoking, and arterial hypertension (P_{obs} = 0.008, P_{adj} = 0.016, OR_{adj} = 2.391, 95% CI = 1.180–4.843). Under observed additive model genotypic association was revealed for both G/A (P_{obs} = 0.029, OR_{obs} = 1.997, 95% CI = 1.072–3.723) and A/A (P_{obs} = 0.020, OR_{obs} = 2.478, 95% CI = 1.155–5.317) genotypes, but after adjusting for the risk factors the genotypic association remained for A/A genotype (P_{adj} = 0.021, OR_{adj} = 3.304, 95% CI = 1.199–9.106), and was lost for G/A genotype (P_{adj} = 0.054). The frequencies of G-1639A genotypes were different between the cases and controls with arterial hypertension either before or after adjustment for the covariates of gender, age, BMI, and smoking status under dominant (P_{obs} = 0.002, P_{adj} = 0.029, and OR_{adj} = 2.374, 95% CI = 1.091–5.166) and recessive (P_{obs} = 0.015, P_{adj} = 0.049, and OR_{adj} = 2.862, 95% CI = 1.003–8.169) model. Significant association under additive model after adjusting was revealed only for A/A genotype (P_{adj} = 0.029, OR_{adj} = 4.029, 95% CI = 1.153–14.077). No genotypic associations between *VKORC1* G-1639A polymorphism and IS in subjects with BMI < 25 kg/m^2 and normal BP were revealed ($P > 0.05$). Dominant

Table 2: Genotypes distribution of *VKORC1* G-1639A polymorphism in patients with IS and control subjects with different risk factors.

Group	N	Genotype			P
		G/G (%) (95% CI)	G/A (%) (95% CI)	A/A (%) (95% CI)	
Total					
IS	170	49 (28.8) (22.0–35.6)	79 (46.5) (40.0–54.0)	42 (24.7) (18.2–31.2)	0.027
Control	124	54 (43.6) (34.8–52.3)	49 (39.5) (30.9–48.1)	21 (16.9) (10.3–23.5)	
Gender					
Women					
IS	72	18 (25.0) (15.0–35.0)	39 (54.2) (42.7–65.7)	15 (20.8) (11.5–30.2)	0.228
Control	45	18 (40.0) (25.7–54.3)	20 (44.4) (29.9–59.0)	7 (15.6) (5.0–26.2)	
Men					
IS	98	31 (31.6) (22.4–40.8)	40 (40.8) (31.1–50.6)	27 (27.6) (18.7–36.4)	0.119
Control	79	36 (45.6) (34.6–56.6)	29 (36.7) (26.1–47.3)	14 (17.7) (9.3–26.1)	
BMI					
BMI < 25 kg/m^2					
IS	41	10 (24.4) (11.3–37.5)	22 (53.7) (38.4–68.9)	9 (22.0) (9.3–34.6)	0.629
Control	38	13 (34.2) (19.1–49.3)	18 (47.4) (31.5–63.2)	7 (18.4) (6.1–30.8)	
BMI ≥ 25 kg/m^2					
IS	129	39 (30.2) (22.3–38.2)	57 (44.2) (35.6–52.8)	33 (25.6) (18.1–33.1)	0.025
Control	85	41 (48.2) (37.6–58.9)	30 (35.3) (25.1–45.5)	14 (16.5) (8.6–24.4)	
Arterial blood pressure					
Nonhypertensive					
IS	42	14 (33.3) (19.1–47.6)	19 (45.2) (30.2–60.3)	9 (21.5) (9.0–33.8)	0.852
Control	48	17 (35.4) (21.9–49.0)	19 (39.6) (25.8–53.4)	12 (25.0) (12.8–37.3)	
Hypertensive					
IS	128	35 (27.3) (19.6–35.1)	60 (46.9) (38.2–55.5)	33 (25.8) (18.2–33.4)	0.003
Control	73	36 (49.3) (37.9–60.8)	29 (39.7) (28.5–51.0)	8 (11.0) (3.8–18.1)	

N: number of subjects in the subgroups.
P: the likelihood of differences between IS patients and control group by the χ^2-criterion.
95% CI: 95% confidence interval.

model had the lowest value of AIC in most subgroups (Table 3).

Table 4 indicates the results of *VKORC1* C1173T polymorphism case-control genotyping. The distribution of three genotypes (C/C, C/T, and T/T) between the cases and controls was similar ($P = 0.178$). Statistically significant differences in the C1173T genotypes distribution were also absent when subjects of comparison groups were divided by gender, BMI, and blood pressure.

Analysis of C1173T genotypic association with IS under the four common models of inheritance is summarized in Table 5. The link with IS was not found either in total group or subgroups by gender, BMI, and BP under different inheritance models. The association was absent both without and with adjustment ($P_{obs} > 0.05$, $P_{adj} > 0.05$). The lowest AIC value in most subgroups was observed for dominant model (Table 5).

Clinical characteristics of the subgroups stratified by *VKORC1* G-1639A genotypes in IS subjects are shown in Table 6. The statistically significant difference was revealed only for thrombin time (16.48 ± 3.2 s, 17.25 ± 4.1 s, and 15.26 ± 2.5 s, $P = 0.013$). Bonferroni's correction allowed establishing the significant difference between individuals with A/A and G/A genotypes ($P = 0.010$). In Table 7 clinical characteristics of IS patients according to *VKORC1* C1173T genotypes are presented. As shown, statistically significant differences for any comparison were not found.

The next step of the present study was the calculation of linkage disequilibrium (LD) between G-1639A/C1173T SNPs pair. Significant high LD ($D' = 0.809$, $r^2 = 0.518$) was revealed. Therefore, estimation of haplotype frequencies was performed. Analysis of the G-1639A/C1173T haplotype distribution in case and control groups is presented in Table 8. Significant difference in frequencies of -1639G/1173T and -1639A/1173C haplotypes between IS subjects and controls was established; herewith individuals with these haplotypes had an increased risk for IS ($P = 0.010$, OR = 3.813, 95% CI = 1.268–11.298, and $P = 0.011$, OR = 2.189, 95% CI = 1.185–4.045, resp.). In contrast, -1639G/1173C haplotype frequency was significantly higher in the control group than in IS patients and it decreased the risk for ischemic stroke ($P < 0.001$, OR = 0.548, 95% CI = 0.393–0.765). Frequency of -1639A/1173T haplotype in both groups was similar ($P = 0.218$).

4. Discussion

The data obtained in present work demonstrated that *VKORC1* G-1639A (A risk allele frequency 47.6%) but not C1173T (T risk allele frequency 41.2%) polymorphism was

Table 3: Analysis of G-1639A genotypic association with IS under four common models of inheritance.

Model	P_{obs}	OR_{obs} (95% CI)	P_{adj}	OR_{adj} (95% CI)	AIC
		Total			
Dominant	0.009	1.905 (1.172–3.097)	0.015	2.138 (1.157–3.953)	19.27
Recessive	0.111	1.609 (0.897–2.888)	0.142	1.780 (0.824–3.847)	23.46
Overdominant	0.235	1.329 (0.831–2.125)	0.226	1.434 (0.800–2.571)	24.66
Additive[a]	0.032	1.777 (1.050–3.006)	0.041	1.979 (1.029–3.805)	20.83
	0.017	2.204 (1.149–4.227)	0.028	2.621 (1.110–6.191)	
		Gender			
Women					
Dominant	0.090	2.000 (0.899–4.452)	0.038	2.848 (1.058–7.665)	15.94
Recessive	0.479	1.429 (0.533–3.832)	0.520	1.498 (0.437–5.135)	18.31
Overdominant	0.307	1.477 (0.699–3.124)	0.133	2.046 (0.804–5.207)	17.78
Additive	0.122	1.950 (0.836–4.549)	0.049	2.888 (1.006–8.293)	17.91
	0.178	2.143 (0.706–6.501)	0.149	2.747 (0.697–10.826)	
Men					
Dominant	0.058	1.809 (0.979–3.344)	0.082	2.102 (0.911–4.582)	18.08
Recessive	0.126	1.766 (0.853–3.656)	0.196	1.958 (0.708–5.415)	19.27
Overdominant	0.578	1.189 (0.646–2.187)	0.521	1.292 (0.591–2.823)	21.37
Additive	0.173	1.602 (0.813–3.154)	0.166	1.872 (0.770–4.551)	19.40
	0.049	2.240 (1.002–5.007)	0.077	2.830 (0.894–8.959)	
		BMI			
BMI < 25 kg/m^2					
Dominant	0.339	1.612 (0.606–4.288)	0.519	1.521 (0.426–5.434)	14.98
Recessive	0.697	1.246 (0.413–3.758)	0.898	0.907 (0.203–4.062)	15.75
Overdominant	0.577	1.287 (0.531–3.116)	0.500	1.481 (0.474–4.625)	15.59
Additive	0.380	1.589 (0.565–4.465)	0.479	1.615 (0.428–6.086)	16.98
	0.434	1.671 (0.462–6.051)	0.801	1.252 (0.218–7.183)	
BMI ≥ 25 kg/m^2					
Dominant	0.008	2.150 (1.219–3.793)	0.016	2.391 (1.180–4.843)	18.11
Recessive	0.118	1.743 (0.869–3.498)	0.090	2.230 (0.882–5.638)	22.61
Overdominant	0.196	1.451 (0.825–2.552)	0.282	1.457 (0.734–2.892)	23.47
Additive	0.029	1.997 (1.072–3.723)	0.054	2.107 (0.989–4.489)	19.81
	0.020	2.478 (1.155–5.317)	0.021	3.304 (1.199–9.106)	
		Arterial blood pressure			
Nonhypertensive					
Dominant	0.836	1.097 (0.458–2.625)	0.228	1.966 (0.655–5.904)	15.77
Recessive	0.690	0.818 (0.306–2.191)	0.612	0.717 (0.199–2.588)	15.66
Overdominant	0.588	1.261 (0.545–2.918)	0.114	2.382 (0.811–6.997)	15.52
Additive	0.689	1.214 (0.469–3.143)	0.161	2.268 (0.722–7.123)	17.49
	0.870	0.911 (0.298–2.782)	0.725	0.778 (0.192–3.154)	
Hypertensive					
Dominant	0.002	2.585 (1.417–4.717)	0.029	2.374 (1.091–5.166)	19.72
Recessive	0.015	2.822 (1.255–6.501)	0.049	2.862 (1.003–8.169)	22.64
Overdominant	0.327	1.339 (0.747–2.399)	0.574	1.230 (0.598–2.528)	28.46
Additive	0.021	2.128 (1.119–4.046)	0.095	2.042 (0.844–4.720)	19.26
	0.002	4.243 (1.722–10.453)	0.029	4.029 (1.153–14.077)	

CI: confidence interval; AIC: Akaike information criterion; P_{obs}: observed P value; OR_{obs}: observed odds ratio; P_{adj}: P value adjusted for covariates of age, gender, smoking status, body mass index, and hypertension in total group; for age, smoking status, body mass index, and hypertension in gender group; for age, gender, smoking status, and hypertension in BMI group; for age, gender, smoking status, and body mass index in hypertension group; OR_{ads}: odds ratio adjusted for covariates.

[a]First row in additive model describes comparison of heterozygous genotype with major homozygous genotype and second row comparison of minor homozygous genotype with major homozygous genotype.

Table 4: Genotypes distribution of *VKORC1* C1173T polymorphism in patients with IS and control subjects with different risk factors.

Group	*N*	Genotype			*P*
		C/C (%) (95% CI)	C/T (%) (95% CI)	T/T (%) (95% CI)	
Total					
IS	170	63 (37.1) (29.8–44.3)	74 (43.5) (36.1–51.0)	33 (19.4) (13.5–25.4)	0.178
Control	124	59 (47.6) (38.8–56.4)	47 (37.9) (29.4–46.4)	18 (14.5) (8.3–20.7)	
Gender					
Women					
IS	72	24 (33.3) (22.4–44.2)	32 (44.4) (33.0–56.0)	16 (22.2) (12.6–31.8)	0.154
Control	45	22 (48.9) (34.3–63.5)	18 (40.0) (25.7–54.3)	5 (11.1) (1.9–20.3)	
Men					
IS	98	39 (39.8) (30.1–49.5)	42 (42.9) (33.1–52.7)	17 (17.3) (9.9–24.9)	0.626
Control	79	37 (46.8) (35.8–57.8)	29 (36.7) (26.1–47.3)	13 (16.5) (8.3–24.6)	
BMI					
BMI < 25 kg/m^2					
IS	41	16 (39.0) (24.1–54.0)	16 (39.0) (24.1–54.0)	9 (22.0) (9.3–34.6)	0.568
Control	38	19 (50.0) (34.1–65.9)	11 (28.9) (14.5–43.4)	8 (21.1) (8.1–34.0)	
BMI ≥ 25 kg/m^2					
IS	129	47 (36.4) (28.1–44.8)	58 (45.0) (36.4–53.6)	24 (18.6) (11.9–25.3)	0.212
Control	85	40 (47.1) (36.5–57.7)	35 (41.2) (30.7–51.6)	10 (11.8) (4.9–18.6)	
Arterial blood pressure					
Nonhypertensive					
IS	42	16 (38.1) (23.4–52.8)	19 (45.2) (30.2–60.3)	7 (16.7) (5.4–27.9)	0.725
Control	48	22 (45.8) (31.7–59.9)	18 (37.5) (23.8–51.2)	8 (16.7) (6.1–27.2)	
Hypertensive					
IS	128	47 (36.7) (28.4–45.1)	55 (43.0) (34.4–51.5)	26 (20.3) (13.3–27.3)	0.140
Control	73	37 (50.7) (39.2–62.2)	26 (35.6) (24.6–46.6)	10 (13.7) (5.8–21.6)	

See Table 2.

associated with IS in Ukrainian population. It has been shown that the risk for IS in patients with A/A and G/A genotypes (G-1639A polymorphism) was higher than for individuals with G/G genotype. It should also be noted that the risk for IS was increased when combining A/A or G/A genotype with hypertension or overweight. Moreover, haplotype analysis revealed that individuals with -1639G/1173T and -1639A/1173C haplotypes had an enhanced risk for IS; conversely the -1639G/1173C haplotype was related to reduced risk for IS.

VKORC1 C1173T polymorphic variant (rs9934438) is located in the first intron and leads to substitution of cytosine to thymine at 1173 position of the gene. Our results demonstrated high LD between G-1639A/C1173T SNPs pair, which was consistent with the data obtained in other studies [13, 14, 22]. The G-1639A (rs9923231) single-nucleotide polymorphism is located in the second nucleotide of the E-Box (CA/GGGTG) of *VKORC1* promoter region and leads to guanine/adenine conversion at position -1639 of the gene. Such -1639G>A substitution creates the E-box binding site (from C<u>G</u>GGTG to C<u>A</u>GGTG), which attracts repressive E-box binding proteins [23]. Therefore, this polymorphism leads to changes in the *VKORC1* promoter activity and causes reduction of *VKORC1* mRNA production and enzyme expression. According to this, few studies showed that *VKORC1* mRNA expression was higher in tissues of subjects with G-1639G/C1173C genotypes compared to subjects with G-1639A/C1173T and A-1639A/T1173T genotypes [23, 24].

It allows suggesting that inhibition of vitamin K recycling in -1639A and 1173T allele carriers may cause insufficient γ-carboxylation of protein C, protein S, protein Z, MGP, GRP, and osteocalcin. This would result in increased risk of clot formation, calcification of arteries, and atherosclerotic plaques contributing to the atherosclerosis and its complications development. According to this, Teichert et al. [8] showed that T-allele of the *VKORC1* C1173T polymorphism was associated with a significantly higher risk of aortic calcification in the Caucasian. Tavridou et al. [12] revealed association between *VKORC1* G-1639A SNP and maximum carotid intima-media thickness in type 2 diabetes mellitus, which was explained by the higher prevalence of calcification in individuals with -1639A allele.

Previously, we also investigated an association of *VKORC1* T2255C polymorphism with IS in the same Ukrainian population [25]. It was shown that carriers of C/C genotype had a significant higher risk for IS than individuals with T/T genotype. Our findings were consistent with results demonstrated by Du et al. [15], who reported that *VKORC1* G-1639A and T2255C were associated with susceptibility to cardiovascular and cerebrovascular diseases (CCVD) in Chinese population. Herewith, individuals with A and C allele had an increased risk for CCVD.

Table 5: Analysis of C1173T genotypic association with IS under four common models of inheritance.

Model	P_{obs}	OR_{obs} (95% CI)	P_{adj}	OR_{adj} (95% CI)	AIC
		Total			
Dominant	0.071	1.542 (0.963–2.467)	0.054	1.914 (0.989–3.449)	18.96
Recessive	0.275	1.418 (0.757–2.657)	0.701	1.165 (0.535–2.538)	21.00
Overdominant	0.333	1.263 (0.787–2.026)	0.055	1.811 (0.989–3.318)	21.28
Additive	0.135	1.475 (0.886–2.455)	0.078	2.058 (0.912–3.926)	20.76
	0.117	1.717 (0.874–3.373)	0.260	1.620 (0.700–3.749)	
		Gender			
Women					
Dominant	0.096	1.913 (0.892–4.102)	0.109	2.115 (0.846–5.286)	16.83
Recessive	0.135	2.286 (0.774–6.751)	0.243	2.162 (0.593–7.879)	17.17
Overdominant	0.637	1.200 (0.563–2.556)	0.454	1.427 (0.563–3.617)	19.41
Additive	0.242	1.630 (0.720–3.690)	0.217	1.869 (0.692–5.046)	17.79
	0.069	2.933 (0.921–9.347)	0.131	2.886 (0.730–11.413)	
Men					
Dominant	0.347	1.333 (0.732–2.426)	0.165	1.744 (0.796–3.823)	17.38
Recessive	0.875	1.066 (0.483–2.352)	0.794	0.872 (0.311–2.444)	18.24
Overdominant	0.407	1.293 (0.704–2.375)	0.101	1.998 (0.874–4.569)	17.58
Additive	0.340	1.374 (0.715–2.640)	0.099	2.091 (0.871–5.018)	19.33
	0.619	1.241 (0.530–2.905)	0.756	1.192 (0.394–3.602)	
		BMI			
BMI < 25 kg/m^2					
Dominant	0.328	1.562 (0.639–3.818)	0.496	1.476 (0.481–4.527)	15.23
Recessive	0.923	1.055 (0.360–3.089)	0.536	0.637 (0.153–2.655)	16.18
Overdominant	0.347	1.571 (0.613–4.025)	0.208	2.213 (0.642–7.626)	15.30
Additive	0.292	1.727 (0.626–4.769)	0.260	2.112 (0.575–7.763)	17.06
	0.625	1.336 (0.418–4.268)	0.820	0.839 (0.184–3.830)	
BMI ≥ 25 kg/m^2					
Dominant	0.122	1.551 (0.889–2.706)	0.053	2.142 (0.974–4.310)	18.40
Recessive	0.184	1.714 (0.774–3.796)	0.346	1.598 (0.603–4.236)	18.94
Overdominant	0.585	1.167 (0.671–2.031)	0.141	1.683 (0.842–3.367)	20.50
Additive	0.257	1.410 (0.778–2.556)	0.054	2.085 (0.989–4.397)	19.65
	0.099	2.043 (0.873–4.777)	0.118	2.322 (0.806–6.868)	
		Arterial blood pressure			
Nonhypertensive					
Dominant	0.459	1.375 (0.592–3.194)	0.295	1.730 (0.620–4.828)	15.44
Recessive	1.000	1.000 (0.329–3.038)	0.628	0.701 (0.166–2.955)	15.99
Overdominant	0.457	1.377 (0.593–3.199)	0.157	2.167 (0.743–6.318)	15.43
Additive	0.423	1.451 (0.584–3.610)	0.181	2.157 (0.700–6.648)	17.34
	0.763	1.203 (0.362–4.001)	0.978	0.978 (0.209–4.584)	
Hypertensive					
Dominant	0.055	1.771 (0.989–3.173)	0.071	1.960 (0.944–4.066)	17.64
Recessive	0.242	1.606 (0.726–3.553)	0.362	1.571 (0.594–4.152)	19.93
Overdominant	0.307	1.362 (0.753–2.465)	0.246	1.558 (0.736–3.299)	20.31
Additive	0.115	1.665 (0.883–3.142)	0.116	1.901 (0.853–4.236)	19.42
	0.097	2.047 (0.877–4.775)	0.164	2.097 (0.740–5.946)	

See Table 3.

Table 6: Clinical characteristics of IS patients with different *VKORC1* G-1639A genotypes.

Parameter	G/G	G/A	A/A	Total	*P*
N	49	79	42	170	—
BMI	27.9 ± 3.9	28.6 ± 4.9	27.8 ± 3.6	28.2 ± 4.3	0.533
Systolic BP, mmHg	164.1 ± 32.5	169.7 ± 29.3	165.2 ± 24.8	166.9 ± 29.2	0.522
Diastolic BP, mmHg	94.9 ± 15.3	96.4 ± 15.4	93.9 ± 16.4	95.4 ± 15.6	0.692
Total cholesterol[a], mmol/L	5.09 ± 1.5	5.02 ± 1.5	5.06 ± 1.6	5.05 ± 1.5	0.961
HDL-cholesterol[a], mmol/L	1.07 ± 0.3	1.01 ± 0.3	0.99 ± 0.3	1.02 ± 0.3	0.452
LDL-cholesterol[a], mmol/L	3.29 ± 1.4	3.24 ± 1.4	3.24 ± 1.5	3.26 ± 1.4	0.979
Triglyceride[a], mmol/L	1.62 ± 0.7	1.69 ± 0.8	1.80 ± 0.9	1.70 ± 0.8	0.588
Prothrombin time, s	9.68 ± 1.9	9.51 ± 2.0	9.17 ± 2.2	9.48 ± 2.0	0.483
Thrombin time, s	16.48 ± 3.2	17.25 ± 4.1	15.26 ± 2.5	16.54 ± 3.6	0.013[b]
Fibrinogen, g/L	3.79 ± 1.3	3.89 ± 1.3	4.17 ± 1.1	3.93 ± 1.2	0.301
Fasting glucose, mmol/L	5.82 ± 1.4	6.11 ± 1.6	5.70 ± 1.5	5.92 ± 1.5	0.312

N: number of subjects; HDL: high density lipoprotein; LDL: low density lipoprotein.
[a] $N = 47$ for G/G genotype, $N = 70$ for G/A genotype, and $N = 40$ for A/A genotype.
[b] Significant difference between A/A and G/A genotypes ($P = 0.010$) by Bonferroni's correction.

Table 7: Clinical characteristics of IS patients with different *VKORC1* C1173T genotypes.

Parameter	C/C	C/T	T/T	Total	*P*
N	63	74	33	170	—
BMI, kg/m^2	28.2 ± 5.0	28.3 ± 3.8	28.0 ± 4.0	28.2 ± 4.3	0.961
Systolic BP, mmHg	166.6 ± 28.9	167.6 ± 32.3	166.4 ± 22.6	167.0 ± 29.2	0.973
Diastolic BP, mmHg	94.1 ± 13.9	96.6 ± 18.3	95.0 ± 11.7	95.4 ± 15.6	0.657
Total cholesterol[a], mmol/L	5.09 ± 1.6	4.98 ± 1.4	5.13 ± 1.6	5.05 ± 1.5	0.862
HDL-cholesterol[a], mmol/L	1.05 ± 0.3	1.01 ± 0.3	1.00 ± 0.3	1.02 ± 0.3	0.670
LDL-cholesterol[a], mmol/L	3.25 ± 1.5	3.22 ± 1.4	3.35 ± 1.5	3.26 ± 1.4	0.910
Triglyceride[a], mmol/L	1.74 ± 0.8	1.66 ± 0.8	1.72 ± 0.9	1.70 ± 0.8	0.820
Prothrombin time, s	9.56 ± 2.0	9.61 ± 2.1	8.99 ± 1.9	9.48 ± 2.0	0.313
Thrombin time, s	16.59 ± 3.5	16.73 ± 3.6	16.00 ± 3.6	16.54 ± 3.6	0.616
Fibrinogen, g/L	3.85 ± 1.2	3.85 ± 1.3	4.26 ± 1.1	3.93 ± 1.2	0.227
Fasting glucose, mmol/L	6.02 ± 1.6	5.85 ± 1.4	5.90 ± 1.7	5.92 ± 1.5	0.795

N: number of subjects; HDL: high density lipoprotein; LDL: low density lipoprotein.
[a] $N = 59$ for C/C genotype, $N = 69$ for C/T genotype, and $N = 29$ for T/T genotype.

On the other hand, Zhang et al. [14] investigated the contribution of *VKORC1* G-1639A and C1173T SNPs to ischemic cerebrovascular disease (ICVD) in Chinese Han population and reported that subjects carrying the -1639G (G risk allele frequency 11.4%) or 1173C (C risk allele frequency 7.4%) allele might be at increased risk of ICVD. Furthermore, the 1639G-1173C haplotype was a risk factor for ICVD, and 1639A-1173T was a protective factor. The researchers suggested that -1639G allele, which can increase *VKORC1* mRNA production, was associated with low sensitivity to vitamin K antagonists and higher risk of thrombosis. Similar result was obtained by Wang et al. [13], who identified natural haplotype block in *VKORC1* gene, which included five common noncoding SNPs (G-1639A, C1173T, C1542G, T2255C, and G3730A) with strong LD. Authors discovered that *VKORC1* 2255C allele, which can reflect G-C-G-C-A haplotype, increased almost twice the risk of stroke, coronary heart disease, and aortic dissection in Chinese population. In support of this, Shyu et al. [18] showed that -1639A allele (VKORC1 G-1639A polymorphism) had protective effect on the development of large-artery atherosclerotic stroke and was associated with reduced stroke risk in Taiwan population. It was explained that minor allele carriers may have lower concentrations of blood coagulation factors, leading to protection against vascular thrombosis and, consequently, to a reduced susceptibility for stroke.

At the same time, the most case-control studies in Europe and North America population did not reveal association between *VKORC1* SNPs and cerebrovascular diseases development. Thus, Ragia et al. [17] did not find significant difference in the *VKORC1* G-1639A (A risk allele frequency 42.6%) genotypes distribution among Greek Caucasian IS patients and matched controls. Moreover, minor -1639A allele was not associated with occurrence and clinical aspects of ischemic stroke. Authors proposed that lack of such association could be explained by combined effects of -1639A allele on vitamin K-dependent hemostatic and nonhemostatic proteins, affecting both clot formation and vascular calcification. Hindorff et al. [19] did not reveal significant association of five common *VKORC1* SNPs and haplotypes with myocardial

Table 8: Analysis of G-1639A/C1173T haplotype distribution in IS and control groups.

Haplotype	IS group		Control group		P	OR	95% CI
	2N	Frequency	2N	Frequency			
G-T	20	0.059	4	0.016	0.010	3.813	1.268–11.298
A-C	42	0.124	15	0.061	0.011	2.189	1.185–4.045
G-C	158	0.464	152	0.613	<0.001	0.548	0.393–0.765
A-T	120	0.353	77	0.310	0.281	1.211	0.854–1.717

N: number of subjects; OR: odds ratio; CI: confidence interval.

infarction, ischemic stroke, and venous thrombosis on large scale study in North American population. In accordance with the above, two case-control studies carried out in Belgian and Southern German population by Lemmens et al. [20] and Arnold et al. [16], respectively, also did not show any association between *VKORC1* haplotypes and different subtypes of ischemic stroke.

Recently, a meta-analysis conducted by Li et al. [26] demonstrated that G-1639A and T2255C SNPs in *VKORC1* gene might contribute to the risk of cerebrovascular and cardiovascular diseases. Herewith, the above data showed that minor alleles in some populations can be the wild alleles in other populations, which leads to different interpretation of the results. It turns out that *VKORC1* genetic polymorphisms can lead to reduction of the thrombosis and atherosclerosis risk in some cases and increase the risk of clot formation and atherogenesis in other cases, contributing to development of cerebrovascular and cardiovascular diseases (as has been shown in our study). The mechanism of this duality is not fully clear. One explanation for this could be the widespread vitamin K deficiency among individuals of different populations [27]. In such case, the transport systems provide preferential targeting of phylloquinone to the liver to preserve coagulation, while less important Gla-proteins, which are synthesized in the extrahepatic tissues, do not receive vitamin K (according to the triage theory by McCann and Ames [28]). In this way, VKORC1 activity reduction due to genetic polymorphisms exacerbates the shortage of the vitamin K in the extrahepatic tissues and is compensated by the vitamin K in the liver.

Our case-control study has few limitations, which have to be taken into account during interpretation the results. First of all, study groups included quite low number of individuals and may not represent the general Ukrainian population. Small size sample is explained by the difficulty of the subject selection. Only individuals who have never been on oral anticoagulant therapy were included to case and control groups. Such people are quite rare but were necessary for our study (the most part of oral anticoagulants inhibits VKORC1 and may obscure the impact of *VKORC1* gene polymorphism on IS development). Second, it should also be noted that average age of the control group was significantly higher than in the stroke patients. It was a condition of the study design, because it allowed suggesting that control individuals had the presumably reduced risk for IS in the future. Additionally, both groups were similar in smoking and body mass index, but the arterial hypertension prevalence was higher among IS patients. This fact makes it impossible to draw firm conclusion about the impact of this risk factor on IS development in individuals with different *VKORC1* genotypes. Finally, more people should be enrolled in the study even as other *VKORC1* SNPs should be analyzed in order to make a definitive conclusion about *VKORC1* association with ischemic stroke in Ukrainian population. It will be the focus of our further research.

5. Conclusion

In summary, this is the first report investigating the association between *VKORC1* G-1639A and C1173T polymorphisms and IS in Ukrainian population. Obtained results revealed that G-1639A but not C1173T polymorphism was related to IS. The risk for IS in -1639A allele carriers was higher than in major allele homozygotes. Moreover, -1639G/1173T and -1639A/1173C haplotypes were risk factors for IS, and -1639G/1173C haplotype was a protective factor for IS. Subsequent studies with larger number of participants are required to confirm our present results.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Authors' Contributions

Yevhen I. Dubovyk performed the clinical research and wrote the manuscript. Viktoriia Yu. Harbuzova performed genotyping and biostatistics. Alexander V. Ataman designed the research plan and organized the study.

Acknowledgments

The study was a part of scientific project "Association of 'Ectopic Calcification Genes' Allelic Polymorphism with the Development of Common Cardiovascular Diseases and Their Complications," supported by the Ministry of Education and Science of Ukraine (no. 0115U000688).

References

[1] J. Oldenburg, M. Marinova, C. Müller-Reible, and M. Watzka, "The vitamin K Cycle," *Vitamins and Hormones*, vol. 78, pp. 35–62, 2008.

[2] M. Caspers, K. J. Czogalla, K. Liphardt et al., "Two enzymes catalyze vitamin K 2,3-epoxide reductase activity in mouse:

VKORC1 is highly expressed in exocrine tissues while VKORC1L1 is highly expressed in brain," *Thrombosis Research*, vol. 135, no. 5, pp. 977–983, 2015.

[3] M. S. El Asmar, J. J. Naoum, and E. J. Arbid, "Vitamin K dependent proteins and the role of vitamin K2 in the modulation of vascular calcification: a review," *Oman Medical Journal*, vol. 29, no. 3, pp. 172–177, 2014.

[4] C. S. B. Viegas, M. S. Rafael, J. L. Enriquez et al., "Gla-rich protein acts as a calcification inhibitor in the human cardiovascular system," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 35, no. 2, pp. 399–408, 2015.

[5] S. Laurance, C. A. Lemarié, and M. D. Blostein, "Growth arrest-specific gene 6 (gas6) and vascular hemostasis," *Advances in Nutrition*, vol. 3, no. 2, pp. 196–203, 2012.

[6] T. Li, C.-Y. Chang, D.-Y. Jin, P.-J. Lin, A. Khvorova, and D. W. Stafford, "Identification of the gene for vitamin K epoxide reductase," *Nature*, vol. 427, no. 6974, pp. 541–544, 2004.

[7] S. Rost, A. Fregin, V. Ivaskevicius et al., "Mutations in VKORC1 cause warfarin resistance and multiple coagulation factor deficiency type 2," *Nature*, vol. 427, no. 6974, pp. 537–541, 2004.

[8] M. Teichert, L. E. Visser, R. H. N. van Schaik et al., "Vitamin K Epoxide Reductase Complex Subunit 1 (VKORC1) polymorphism and aortic calcification," *Arteriosclerosis, Thrombosis, and Vascular Biology*, vol. 28, no. 4, pp. 771–776, 2008.

[9] M. N. Watzka, A. Nebel, E. Mokhtari et al., "Functional promoter polymorphism in the VKORC1 gene is no major genetic determinant for coronary heart disease in Northern Germans," *Thrombosis and Haemostasis*, vol. 97, no. 6, pp. 998–1002, 2007.

[10] D. M. Smadja, M.-A. Loriot, L. A. Hindorff, L. Mellottee, P. Gaussem, and J. Emmerich, "No clear link between VKORC1 genetic polymorphism and the risk of venous thrombosis or peripheral arterial disease," *Thrombosis and Haemostasis*, vol. 99, no. 5, pp. 970–972, 2008.

[11] D. Fodor, C. Bondor, A. Albu, R. Popp, I. V. Pop, and L. Poanta, "Relationship between VKORC1 single nucleotide polymorphism 1173C>T, bone mineral density & carotid intima-media thickness," *Indian Journal of Medical Research*, vol. 137, no. 4, pp. 734–741, 2013.

[12] A. Tavridou, I. Petridis, M. Vasileiadis et al., "Association of *VKORC1* –1639G>A polymorphism with carotid intima-media thickness in type 2 diabetes mellitus," *Diabetes Research and Clinical Practice*, vol. 94, no. 2, pp. 236–241, 2011.

[13] Y. Wang, W. Zhang, Y. Zhang et al., "VKORC1 haplotypes are associated with arterial vascular diseases (stroke, coronary heart disease, and aortic dissection)," *Circulation*, vol. 113, no. 12, pp. 1615–1621, 2006.

[14] H. Zhang, L. Yang, Q. Feng, Y. Fan, H. Zheng, and Y. He, "Association between VKORC1 gene polymorphisms and ischemic cerebrovascular disease in Chinese Han population," *Journal of Molecular Neuroscience*, vol. 53, no. 2, pp. 166–170, 2014.

[15] J. Du, Z. Zhang, Y. Ge, J. Zhen, J. Leng, and J. Wang, "VKORC1 and CD-14 genetic polymorphisms associate with susceptibility to cardiovascular and cerebrovascular diseases," *International Journal of Clinical and Experimental Medicine*, vol. 8, no. 11, pp. 20444–20453, 2015.

[16] M.-L. Arnold, C. Lichy, I. Werner, A. Radbruch, S. Wagner, and C. Grond-Ginsbach, "Single nucleotide polymorphisms in the VKORC1 gene and the risk of stroke in the Southern German population," *Thrombosis and Haemostasis*, vol. 100, no. 4, pp. 614–617, 2008.

[17] G. Ragia, S. Marousi, J. Ellul, V. G. Manolopoulos, and A. Tavridou, "Association of functional VKORC1 promoter polymorphism with occurrence and clinical aspects of ischemic stroke in a Greek population," *Disease Markers*, vol. 35, no. 6, pp. 641–646, 2013.

[18] H.-Y. Shyu, C.-S. Fong, Y.-P. Fu et al., "Genotype polymorphisms of GGCX, NQO1, and VKORC1 genes associated with risk susceptibility in patients with large-artery atherosclerotic stroke," *Clinica Chimica Acta*, vol. 411, no. 11-12, pp. 840–845, 2010.

[19] L. A. Hindorff, S. R. Heckbert, N. Smith, K. D. Marciante, and B. M. Psaty, "Common VKORC1 variants are not associated with arterial or venous thrombosis," *Journal of Thrombosis and Haemostasis*, vol. 5, no. 10, pp. 2025–2027, 2007.

[20] R. Lemmens, S. Abboud, L. Vanhees, A. Goris, and V. Thijs, "Lack of association between variants in the VKORC1 gene and cerebrovascular or coronary heart disease," *Journal of Thrombosis and Haemostasis*, vol. 6, no. 12, pp. 2220–2223, 2008.

[21] H. P. Adams and B. H. Bendixen, "Classification of subtype of acute ischemic stroke definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment," *Stroke*, vol. 24, no. 1, pp. 35–41, 1993.

[22] G. D'Andrea, R. L. D'Ambrosio, P. Di Perna et al., "A polymorphism in the VKORC1 gene is associated with an interindividual variability in the dose-anticoagulant effect of warfarin," *Blood*, vol. 105, no. 2, pp. 645–649, 2005.

[23] D. Wang, H. Chen, K. M. Momary et al., "Regulatory polymorphism in vitamin K epoxide reductase complex subunit 1 (VKORC1) affects gene expression and warfarin dose requirement," *Blood*, vol. 112, no. 4, pp. 1013–1021, 2008.

[24] H.-Y. Yuan, J.-J. Chen, M. T. M. Lee et al., "A novel functional VKORC1 promoter polymorphism is associated with inter-individual and inter-ethnic differences in warfarin sensitivity," *Human Molecular Genetics*, vol. 14, no. 13, pp. 1745–1751, 2005.

[25] V. Y. Garbuzova, D. A. Stroy, V. E. Dosenko et al., "Association of allelic polymorphisms of genes matrix Gla-protein system with ischemic atherothrombotic stroke," *Fiziolohichnyĭ Zhurnal*, vol. 61, no. 1, pp. 19–27, 2015.

[26] Y. Li, J. Zhu, and J. Q. Ding, "VKORC1 rs2359612 and rs9923231 polymorphisms correlate with high risks of cardiovascular and cerebrovascular diseases," *Genetics and Molecular Research*, vol. 14, no. 4, pp. 14731–14744, 2015.

[27] S. L. Booth and A. A. Rajabi, "Determinants of vitamin K status in humans," *Vitamins and Hormones*, vol. 78, pp. 1–22, 2008.

[28] J. C. McCann and B. N. Ames, "Vitamin K, an example of triage theory: is micronutrient inadequacy linked to diseases of aging?" *The American Journal of Clinical Nutrition*, vol. 90, no. 4, pp. 889–907, 2009.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 9850690, 12 pages
http://dx.doi.org/10.1155/2016/9850690

Review Article

Molecular Diagnostics for Precision Medicine in Colorectal Cancer: Current Status and Future Perspective

Guoli Chen,[1] Zhaohai Yang,[1] James R. Eshleman,[2] George J. Netto,[2] and Ming-Tseh Lin[2]

[1]*Department of Pathology, Penn State College of Medicine, Milton S. Hershey Medical Center, Hershey, PA, USA*
[2]*Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA*

Correspondence should be addressed to Ming-Tseh Lin; mlin36@jhmi.edu

Received 31 May 2016; Accepted 10 August 2016

Academic Editor: Gokce Toruner

Precision medicine, a concept that has recently emerged and has been widely discussed, emphasizes tailoring medical care to individuals largely based on information acquired from molecular diagnostic testing. As a vital aspect of precision cancer medicine, targeted therapy has been proven to be efficacious and less toxic for cancer treatment. Colorectal cancer (CRC) is one of the most common cancers and among the leading causes for cancer related deaths in the United States and worldwide. By far, CRC has been one of the most successful examples in the field of precision cancer medicine, applying molecular tests to guide targeted therapy. In this review, we summarize the current guidelines for anti-EGFR therapy, revisit the roles of pathologists in an era of precision cancer medicine, demonstrate the transition from traditional "one test-one drug" assays to multiplex assays, especially by using next-generation sequencing platforms in the clinical diagnostic laboratories, and discuss the future perspectives of tumor heterogeneity associated with anti-EGFR resistance and immune checkpoint blockage therapy in CRC.

1. Introduction

Colorectal cancer (CRC), predominantly referring to colorectal adenocarcinoma, is one of the most common malignant neoplasms and a leading cause for cancer related deaths worldwide [1]. In 2014, there are nearly 140,000 newly diagnosed patients in the United States where it also ranks in the second place as a cause of cancer related mortality in men and women combined [2]. Therefore, studies aimed at understanding pathogenic mechanisms and optimizing clinical management of CRC have been intensively and devotedly conducted.

In the past two decades, major progress in understanding the genetic alterations of diseases has been achieved and accordingly successful examples of utilizing such information in clinical management are accumulating. These advances have paved the way for the emergence of a new concept, precision medicine, essentially offering individualized medical care to patients based on their unique molecular/genetic profiling and other personalized information. This is in contrast to cohort-based therapy specifically treating patients based on successful therapy of a cohort of similar patients treated previously. In the field of oncology, therapies targeting specific genetic alterations have been proven to be a successful example of practicing precision medicine by significantly improving clinical outcomes compared to conventional chemotherapy and/or radiotherapy. By far, a rapidly growing list of drugs targeting different genetic alterations have been approved by the Food and Drug Administration (FDA) in the United States for treatment of advanced-stage solid tumors [3]. Most of the drugs work through inhibiting kinase activity. For example, *BRAF* inhibitors (vemurafenib and dabrafenib) [4, 5] and *MEK* inhibitor (trametinib) [6] were approved for patients with melanoma bearing *BRAF* p.V600E mutation, anti-EGFR monoclonal antibodies (cetuximab and panitumumab) for CRC without *RAS* mutations [7, 8], EGFR tyrosine kinase inhibitors (gefitinib and erlotinib) targeting certain *EGFR* mutations for non-small-cell lung cancers (NSCLC) [9, 10], and ALK tyrosine kinase inhibitor (crizotinib) for NSCLC carrying the *ALK* gene translocations [11]. Molecular testing of targeted mutations has become essential to select patients for these therapies [12, 13].

To explore more useful "targets" for clinical management of cancers, numerous potential biomarkers have been proposed and investigated with tremendous effort. However, only a limited number of them have so far been proven to be clinically meaningful and subsequently become or potentially become a part of standard patient care. In this review, we focus on the molecular diagnostics currently used in established standard care of CRC, especially those related to targeted therapy or expected to be so shortly.

2. Current Guidelines for Targeted Therapy in CRC

In 2009, the American Society of Clinical Oncology (ASCO) issued a recommendation on molecular analysis for *KRAS* gene mutations in patients with metastatic CRC to predict response to anti-EGFR therapy [12]. Following the initial focus on common *KRAS* mutations at codons 12 and 13, recent data have revealed that mutations at codons 59, 61, 117, and 146 and *NRAS* gene mutations are also associated with anti-EGFR resistance [7, 8, 12, 14, 15]. Based on reviews of currently available evidences, ASCO recently updated their provisional clinical opinions: both *KRAS* and *NRAS* exons 2 (codons 12 and 13), 3 (codons 59 and 61), and 4 (codons 117 and 146) (so-called extended *RAS* testing) should be screened for mutations in all patients with metastatic CRC who are candidates for anti-EGFR therapy [16]. Similarly, a provisional guideline from the Association of Clinical Pathologists Molecular Pathology and Diagnostics Group in the United Kingdom also recommends that at least *KRAS* codons 12, 13, 59, 61, 117, and 146 and *NRAS* codons 12, 13, 59, and 61 should be included for molecular analysis in CRC patients [17]. European Society of Medical Oncology and Japanese Society of Medical Oncology recently also revised/updated their clinical guidelines to recommend testing of extended *KRAS/NRAS* mutations [18, 19]. In addition to *RAS*, there is increasing evidence suggesting that the *BRAF* p.V600E mutation makes response to anti-EGFR therapy highly unlikely [7, 20–22]. The Colon/Rectal Cancer Panel from National Cancer Comprehensive Network (NCCN) recently revised its guideline (Version 2.2016) for anti-EFGR therapy by recommending genotyping of tumor tissues in all patients with metastatic CRC for the extended *RAS* mutations as well as *BRAF* mutations (http://www.nccn.org/professionals/physician_gls/f_guidelines.asp). Also, mutations in exon 20 of the *PIK3CA* gene may be associated with anti-EGFR resistance in *KRAS* wild-type cancer [7]. However, *PIK3CA* mutations are often accompanied by a *KRAS* mutation in CRC [7, 23]. The benefit of testing *PIK3CA* mutations to guide anti-EGFR therapy requires further studies with a large cohort of CRC patients carrying wild-type *KRAS* and *NRAS* genes.

In addition to the above guidelines/recommendations largely limited on "what to test" in order to guide oncologists for targeted therapy, more details on "how to test" thus with more practical meaning to molecular diagnostic laboratories and pathologists are also emerging. Using lung cancer as an example, in 2013, the College of American Pathologists (CAP), International Association for the Study of Lung Cancer (IASLC), and Association for Molecular Pathology (AMP) jointly released a guideline regarding molecular diagnostics in lung cancers [24–26]. In addition to clarifying testing for *EGFR* mutations and *ALK* translocations to guide targeted therapy with EGFR or ALK inhibitors, respectively, in all patients with advanced-stage adenocarcinoma, they also offer recommendations and/or expert consensus opinions to address questions such as "how *EGFR* and *ALK* testing should be performed" and "how the molecular testing should be implemented and operationalized." Soon upon availability, this guideline was endorsed by ASCO [27]. Similarly, a colorectal cancer expert panel from American Society for Clinical Pathology (ASCP), CAP, ASCO, and AMP has drafted a summary of recommendations for guideline on the evaluation of molecular markers for CRC (http://www.amp.org/committees/clinical_practice/CRCOpenComment.cfm). These draft guidelines were opened for comments in 2015. Briefly, they intend to offer "recommendations" and "no recommendations" for questions including not only "which molecular tests should be performed for CRC," but also "what the appropriate sample for the tests is," "how testing should be performed," and so forth. The final version is expected shortly.

3. Roles of Pathologists

Currently molecular diagnostics of CRC is largely conducted on tissue specimens embedded in paraffin. The responsibility of surgical pathologists and cytopathologists is not restricted to only making a histological diagnosis. Pathologists, indeed, play crucial roles in preanalytic specimen preparation for molecular diagnostics, including standardizing operating protocols for tissue sampling and processing, requesting mandatory tests, selecting appropriate tissue blocks, and designating adequate areas with sufficient amount and proportion of tumor cells for nuclei acid extraction.

Pathologists also need to estimate tumor cellularity within the designated areas to ensure that the tumor cell percentage is more than the analytic sensitivity (or limits of detection) of the requested molecular assay [28–30]. They also play a critical role in protecting the tissue blocks from unnecessary testing so that critical tests (even future ones) can be performed.

3.1. Standard Operating Protocols for Tissue Sampling and Processing. Adequate tissue sampling and processing are critical, not only for histopathological interpretation but also for molecular diagnostics. In our retrospective quality assessment of a next-generation sequencing (NGS) assay, referred specimens experienced a significantly higher failure rate than in-house specimens, presumably due in part to a standardized tissue processing protocol applied to in-house specimens [31]. In this era of precision cancer medicine, pathologists are responsible for revisiting of the standard operating protocols for tissue sampling at the grossing stage and tissue processing in the histopathology laboratories to ensure adequate quality and quantity of nucleic acids for molecular diagnostics. Ten percent neutral-buffered formalin is recommended to fix surgical pathology specimens or cell pellet specimens from fine needle aspiration (FNA) or effusion for 6–48 hours

depending on the size of the specimens [25, 26]. Fixative containing heavy metals such as B5 solution or acids such as Bouin's fixative should be avoided if molecular tests are expected to be requested. Bone specimens represent a challenge for molecular diagnosis since bone is common and sometimes the only metastatic site of solid tumors and decalcification step for bone tissues typically using strong acids can damage nucleic acids [32–34]. Therefore, a standard operating protocol for tissue sampling and processing should be established, specifically for bone specimens [35], in the histopathology laboratories to preserve tissues for potential molecular diagnosis.

3.2. Selection of Appropriate Blocks and Designation of Adequate Areas for DNA Extraction. In the clinical diagnostic setting, specimens containing low tumor cellularity are not uncommon [36, 37] and may lead to false negative results, particularly for assays with a lower analytic sensitivity, such as Sanger sequencing [38–40]. Assays with a higher analytic sensitivity are preferred for specimens with low tumor cellularity. Nevertheless, selection of appropriate specimens and designation of adequate areas for DNA extraction by the pathologists may be the most cost-effective option for accurate molecular diagnosis [28]. If possible, specimens with a low tumor cell percentage should be avoided, such as areas with a prominent desmoplastic reaction or inflammatory cell infiltration, resected or biopsied specimens of lymph nodes with subcapsular and/or infiltrative metastasis without nodular formation, and FNA specimens with neoplastic cells intermingled with prominent nonneoplastic tissues. Also, regions rich in potential PCR inhibitors such as mucin and necrotic debris should be avoided as well. Prior medical treatment can interfere with tumor tissue adequacy for molecular testing. For example, neoadjuvant therapy for rectal cancers and some high-risk colon cancers can downstage the tumoral lesion; however the accompanied significant depletion of tumor cellularity may lead to false negative results of *RAS* mutations [28, 41, 42]. For patients undergoing neoadjuvant therapy, pathologists may need to seek alternative specimens, such as pretreatment biopsies if available for molecular tests. In addition, tissue blocks from patients with metastatic CRC should be saved for molecular diagnostics, especially those taken by core biopsy or FNA. Assay feasibility of specimens with limited tissues can be improved by limiting block trimming when preparing slides and by avoiding extensive immunohistochemical workups.

3.3. Estimation of Tumor Cellularity and Postanalytic Quality Assessment. An accurate estimation of tumor cellularity by pathologists provides a preanalytic measure to ensure specimen adequacy for the analytic sensitivity of the assay. A note should be provided in the report to indicate a potential false result when the tumor cellularity is insufficient. Tumor cellularity also provides an independent parameter for postanalytic quality assessment of the assay [43]. In our retrospective quality assessment of a pyrosequencing assay designed to detect the *BRAF* p.V600E mutation for melanomas, much lower than the expected mutant allele frequencies (equivalent to half of the estimated tumor cellularity assuming a heterozygous mutation) were observed in two specimens. Quality assurance investigation revealed that our original pyrosequencing assay, similar to cobas 4800 *BRAF* V600 mutation test, revealed a false, weak p.V600E signal in specimens with a p.V600K mutation. Since the assay was revised, we confirmed that p.V600K and other non-p.V600E mutations are common in melanomas. Correlation between observed and expected mutant allele frequencies may predict tumor heterogeneity and mutant allele-specific imbalance in CRC [36]. A lower than expected mutant allele frequency indicates tumor heterogeneity while a higher than expected mutant allele frequency indicates mutant allele-specific imbalance. We have prospectively and retrospectively confirmed several CRCs with *RAS* mutation present in a subset of tumor cells by analysis of subareas in cases with lower than expected mutant allele frequency. Resistant clones with acquired *RAS* mutations may arise from a small subpopulation present within the original tumor before anti-EGFR therapy or as a consequence of continued mutagenesis over the course of targeted treatment [44]. Presence of *KRAS*-mutant subpopulations may correlate with inferior progression-free survival in CRC patients treated with anti-EGFR therapy [45]. If confirmed, a note should be added in the report to indicate that a *RAS* mutation is present in a subpopulation of the tumor. A note of mutant allele-specific imbalance of the *RAS* gene may also be needed if the information is clinically relevant in the future.

However, assessing tumor cell percentage may not be always precise or accurate [46, 47]. Therefore, a sample with tumor cellularity estimated as borderline adequacy for an assay (e.g., 10–20% of tumor cellularity for a pyrosequencing assay which carries an analytic sensitivity of 10% of tumor cellularity or 5% of mutant allele frequency) should be reevaluated by a second pathologist, preferentially one with both molecular and histological experience. Ideally, pathologists who interpret molecular test results also assess tumor cellularity since they are more aware of the analytic sensitivity of the requested assay. In addition, they can also compare the mutant allele frequency and the estimated tumor cellularity for postanalytic quality assessment of the assay and results as mentioned above. Multiple reasons may account for a perceived discrepancy in correlating the estimated tumor cellularity with the observed mutant allele frequency, one of which is how pathologists determine tumor cellularity. Pathologists are accustomed to using simple linear measurements for microscopic evaluation of sizes. Therefore, they may estimate tumor cellularity by the surface area of tumor cells rather than the ratio of tumor nuclei. This often results in overestimation of tumor cellularity, because cancer cells are commonly larger and visually more impressive than a tiny lymphocyte or stromal cell yet the total DNA amount in both is almost the same.

4. Molecular Assays for Targeted Therapy

Since specimens containing low tumor cellularity are common in the clinical diagnostic setting [36, 37], macro- or microdissection of unstained slides or coring of the tissue

blocks from the area(s) designated by pathologists has become a routine preanalytic approach to enhance tumor cellularity in many clinical laboratories [29, 30]. However, specimens with a small cluster of tumor cells surrounded by nonneoplastic tissue, such as CRC specimens with prior neoadjuvant therapy, lymph node specimens without formation of distinct metastatic tumor nodules, and cytopathologic specimens with scattered tumor cells are still problematic [28, 48]. In the quality assessment of a NGS assay implemented for clinical mutation detection in CRC, a significant portion of mutations were detected with a low mutant allele frequency [36]. With an analytic sensitivity of 10–20%, Sanger sequencing as the prior gold standard could have missed 8% (with less than 10% mutant allele frequency) or even 23% (with less than 20% mutant allele frequency) of the mutations in this series. Therefore, even with all of the preanalytic efforts, it is still not uncommon to have inadequate or marginally adequate materials. This necessitates assays with higher analytic sensitivity.

4.1. Traditional "One Gene-One Drug" Assays. Traditional molecular diagnostics assays usually focus on one mutation or one gene for one drug. Examples include a few assays approved by the FDA of the United States for companion testing of targeted therapy: for example, cobas 4800 BRAF V600 mutation test (for *BRAF* p.V600E mutation), cobas 4800 KRAS mutation test, and therascreen KRAS test (for 7 most common *KRAS* mutations in codons 12 and 13) [49–52]. The cobas 4800 BRAF V600 mutation test, designed specifically for the p.V600E mutation, may also detect non-p.V600E codon 600 mutations at a lower analytic sensitivity. These assays may not be suitable for comprehensive mutational profiling in core biopsy or fine needle aspiration specimens containing limited tissue. A variety of assays have also been applied in molecular diagnostics laboratories to detect *KRAS* or *BRAF* mutations with a limit of detection ranging from 10–20% mutant allele for Sanger sequencing, approximately 5% for pyrosequencing and high-resolution melting curve analysis, to 1–5% for real-time PCR-based assay [53–57]. Results from different assays are usually concordant except for specimens with poor tumor cellularity for which more sensitive assays are needed to prevent false negative results. In addition to molecular assays, immunohistochemistry (IHC) stain is also a highly sensitive and cost-effective assay for detection of *BRAF* mutation [58, 59]. IHC stain is especially valuable for specimens with scattered tumor cells intermixed with abundant nontumor cells. However, monoclonal antibodies against the BRAF p.V600E protein adopted in current IHC assays do not detect other mutations with adequate analytic sensitivity and specificity.

4.2. Multiplex Assays. With continuous expansion of predictive markers for targeted therapeutics, molecular testing has been in a transition from the traditional "one test-one drug" model to a multiplex genotyping platform to simultaneously test a panel of genes for a specific cancer [60, 61]. Primer extension-based multiplex assays, such as the multiplex SNaPshot assay or the Sequenom MassARRAY system, are capable of testing multiple targets in a single reaction while retaining an analytic sensitivity of 5% or below [62, 63]. More excitingly, massively parallel sequencing or NGS technology has revolutionized genome research and also will soon become the most cost-effective multiplex sequencing platform in the clinical diagnostic setting as more and more biomarkers join standard patient care [60]. The US Centers for Disease Control and Prevention (CDC) has convened the Next-Generation Sequencing: Standardization of Clinical Testing (Nex-StoCT) workgroup and published guidelines to address the 4 components of quality management of NGS assays in clinical laboratories: test validation, quality control, proficient test, and reference materials [64]. The 6 analytic performance characteristics for clinical validation of NGS assays were also defined. Although the workgroup focused on heritable genetic diseases, the same principles can also be applied to precision cancer medicine. Subsequently, the workgroup (Nex-StoCT II) also published recommendations to design, optimize, and implement informatics pipelines for clinical NGS assays [65]. In 2015, CAP issued laboratory standards for NGS clinical tests including the new checklist requirements [66].

4.3. Clinical Validation and Implementation of NGS Platforms. Mutational profiling of cancer specimens based on NGS assays has been validated and implemented for prospective standard patient care or clinical trial in the clinical molecular diagnostic laboratories. The spectrum of mutations examined (reportable range) ranges from a panel of genes for a specific tumor or a group of tumors [67–70] to a larger panel of targetable/actionable genes or oncogenes/tumor suppression genes [37, 71] and whole exome sequencing [72, 73]. Currently, treatment guided by the comprehensive analysis of whole exome sequencing is still limited in a small fraction of patients. The ability to generate large amounts of data of unproven significance, therefore, should not take precedence over the timely generation of clinical useful data. When an extensive NGS panel is offered for clinical diagnosis, it is recommended that an assay with a shorter turnaround time is reported first, followed by a more comprehensive assay which may take longer time to complete [24–26].

We have validated and implemented a NGS platform using the AmpliSeq Cancer Hotspot Panel and Personal Genome Machine in a *Clinical Laboratory Improvement Amendments*- (CLIA-) certified laboratory [67]. In a retrospective quality assessment study, we surveyed the performance characteristics of the NGS assay conducted in 310 CRC specimens [36]. NGS demonstrated a high analytic sensitivity (2% or lower mutant allele frequency), broad reportable ranges of mutation spectrum relevant to anti-EGFR therapy in *KRAS*, *NRAS*, *BRAF*, and *PIK3CA* genes, capacity for quantitative measurement of mutant allele frequencies, and simultaneous detection of concomitant mutations [36]. The test feasibility was approximately 98% (2% of combined rejection rate and failure rate) [31] with a turnaround time of 3–6 working days for 90% or more of specimens as the assay was conducted twice a week (unpublished data). Seventeen percent of *KRAS* mutations were outside codons 12 and 13 and 48% of *PIK3CA* mutations were outside the 3 most common mutated codons 542, 545, and 1047. The

incidence of tumors with predicted resistance to anti-EGFR therapy increased as reportable ranges became wider, from 40% if only mutations in *KRAS* exon 2 were tested to 47% if exons 2–4 were included, 48% if *KRAS* and *NRAS* exons 2–4 were included, 58% if also including *BRAF* codon 600 mutations, and 59% if adding *PIK3CA* exon 20 mutations. Interestingly, right-sided CRCs were found with a higher risk of predicted anti-EGFR resistance. The advantage with broader reportable ranges may also be helpful in elucidating the clinical significance of uncommon mutations [36, 74]. For example, we identified *BRAF* mutations with reduced or silent kinase activity, such as mutations affecting codon 594. CRC patients with kinase-impaired *BRAF* mutations may respond to anti-EGFR therapy [7]. Kinase-impaired *BRAF* mutants showed a significantly higher incidence of concomitant activating *KRAS* or *NRAS* mutations [36, 75]. In the presence of oncogenic RAS proteins, kinase-impaired BRAF forms a complex with CRAF, which leads to hyperactivation of the CRAF/MEK/ERK cascade [76, 77]. Therefore, CRC patients with coexisting kinase-impaired *BRAF* mutation and activating *RAS* mutation may benefit from MEK inhibitors [75].

5. Future Perspectives

With the inspiration of achieved success in precision cancer medicine, tremendous efforts have been devoted to discovering more biomarkers for potential usage in clinical diagnosis, prediction, and prognostication of CRC [78]. In addition to mutations within the mitogen-activated protein kinase (MAPK or RAS/RAF/MEK/ERK) pathway and the phosphatidylinositol 3-kinase (PI3K/AKT/mTOR) pathway, the potential molecular markers also include epigenetic alterations and microRNA expression. Also, exploration of "old" markers for new utilization is feasible and economical given that a single marker can be used for multiple purposes or in different situations. For example, *BRAF* mutation status has several clinical implications in CRC. Detecting the p.V600E mutation has been a part of the algorithm to distinguish sporadic CRC with microsatellite instability (MSI) from hereditary nonpolyposis colorectal cancers (HNPCC or Lynch syndrome) since this *BRAF* mutation is present only in sporadic cases [79, 80]. *BRAF* mutations along with MSI status can also offer information for risk stratification of CRC. Patients with a p.V600E mutation may carry an inferior prognosis in microsatellite-stable (MSS) CRC but not CRC with MSI [81, 82]; however, a study suggests that in advanced (metastatic) disease patients with a p.V600E mutation carry an inferior prognosis also in MSI CRC [83]. MSI status also has multiple implications. Besides the well-known role in screening for HNPCC, MSI is associated with a lower stage of CRC at diagnosis and a favorable stage-specific prognosis [84], although conflicting results in stage IV patients exist [85]. In terms of predictive significance, MSI patients may not benefit from 5-FU based adjuvant therapy [86–88], though benefit has been observed in stage III patients with suspected HNPCC [89]. Further studies are needed to clarify the predictive value of MSI status [90]. More importantly, a recent clinical trial demonstrated the utility of MSI status as a predictive marker for responsiveness to PD-1 blockade immunotherapy in advanced CRC patients [91]. As expected, there are many ongoing clinical trials regarding targeted therapy, immunotherapy, and combinatorial therapy for CRC patients. Undoubtedly there will be prosperous progress of precision medicine in CRC in the near future. For the rest of this review, we will focus on recent advance in the field of tumor heterogeneity associated with anti-EGFR resistance and immunotherapy.

5.1. Acquired Resistance to Anti-EGFR Therapy. Resistance to targeted therapy can be classified into intrinsic (primary) or acquired (secondary) resistance. Intrinsic resistance is usually defined as immediate failure of treatment, whereas acquired resistance is defined as the disease progression following a period of clinical response. By far, best studied is the resistance to EGFR tyrosine kinase inhibitors in lung cancers [92]. In additional to testing for *EGFR* mutations for first-line targeted therapy, examination of the most prevalent acquired resistance mutation, *EGFR* p.T790M, has become a common clinical practice to select patients for third-generation tyrosine kinase inhibitors [92]. The whole picture of acquired resistance mechanisms in CRC was not fully understood until the last few years, presumably because of lack of second-line targeted therapy to overcome acquired resistance and concern of risk for tissue biopsy. Mutational profiling of cell-free circulating tumor DNA (ctDNA) in the blood (so-called liquid biopsy) provided an alternative noninvasive approach to uncover the genetic landscape of acquired resistance mechanisms in CRC [44, 93–96].

Several mechanisms underlying acquired anti-EGFR resistance in CRC have been reported [97–99]. While mutations responsible for acquired resistance to small molecule kinase inhibitors often occur within the kinase domain, there is significant overlap between the intrinsic and acquired genomic alterations leading to anti-EGFR resistance in CRC. These include mutations in the *KRAS*, *NRAS*, *BRAF*, and *MAP2K1* within the MAPK pathway [97–99], mutations involving codons 464, 465, 467, 491, 492, and 494 within the extracellular domain [93, 100, 101] and codons 714 and 794 within the kinase domain of EGFR [102], and amplification of the *KRAS*, *MET*, and *EBRR2* genes [93–95]. Acquired resistance mutations commonly involve codons 12, 13, and 61 of the *KRAS* gene and codon 61 of the *NRAS* gene in contrast to codons 12 and 13 of the *KRAS* gene in the primary resistance setting. The reported acquired *EGFR* mutations were rare or not seen in CRCs prior to anti-EGFR therapy [103]. Recently, whole exome sequencing accompanied with in vitro study of anti-EGFR sensitivity has recognized, albeit infrequently, mutations in the *ERBB2*, *EGFR*, *FGFR1*, *PDGFRA*, and *MAP2K1* as potential mechanisms underlying primary resistance as well as the tyrosine receptor adaptor gene, *IRS2*, in tumors with increased sensitivity to anti-EGFR therapy [102].

5.2. Tumor Heterogeneity Associated with Anti-EGFR Resistance. A phenomenon relevant to drug resistance is tumor heterogeneity [104, 105]. Comprehensive analysis by NGS has revealed remarkable genetic variation between and within

tumors. Genomic heterogeneity seen in tumors prior to or after targeted therapy poses a major challenge to precision cancer medicine. Tumor heterogeneity associated with primary or acquired anti-EGFR resistance has been well documented in lung cancers. Presence of the *EGFR* p.T790M mutation in a minor subpopulation of tyrosine kinase inhibitor-naïve tumors predicts an inferior response to the first-line EGFR tyrosine kinase inhibitors [106, 107]. Some of the best evidence of tumor heterogeneity associated with acquired resistance in lung cancers is the reciprocal relationship between the *EGFR* p.T790M mutation and transformation of small cell carcinoma. In patients with multiple resistant metastases showing separate adenocarcinoma and small cell carcinoma components, p.T790M was often detected only in the adenocarcinoma component [108, 109], while small cell carcinoma component carried *RB1* mutations, but not p.T790M [108–110].

As mentioned previously, CRC with "acquired" *RAS* mutations may arise from a small subpopulation present within the original tumor before anti-EGFR therapy or as a consequence of continued mutagenesis over the course of targeted treatment [44]. Tumor heterogeneity associated with intrinsic *KRAS* mutations carries an inferior response to anti-EGFR therapy [45]. The observation of multiple acquired resistance mutations in plasma ctDNA from the same patients also suggested tumor heterogeneity associated with acquired anti-EGFR resistance in CRC [44, 93, 94, 96]. Recently, Russo et al. confirmed that tumor heterogeneity associated with acquired anti-EGFR resistance may affect lesion-specific response to the second-line targeted therapy and that plasma ctDNA is a better source of specimens for comprehensive capture and dynamic monitoring of resistance mutations than tissue specimens which are subjected to risk and spatial selection bias of a core biopsy [99].

Strategies have been proposed to overcome acquired resistance associated with tumor heterogeneity. One potential strategy is to combine anti-EGFR therapy with MEK or ERK inhibitors [94, 97]. Although tumor heterogeneity may lead to metastatic lesions demonstrating different resistance mechanisms, these mutations are biologically convergent on the MAPK pathway to sustain activation of MEK and ERK. This was supported by a study of ctDNA showing one or more acquired resistance mutations in the genes involved in the MAPK pathway in 23 of 24 patients [96]. Another potential strategy is to identify intrinsic resistance mutations associated with tumor heterogeneity by using ultrasensitive assays to examine plasma ctDNA. Preclinical studies have shown that combining anti-EGFR therapy with other kinase inhibitors may be effective in tumors harboring mutations or amplification in *ERBB2*, *MET*, *EGFR*, *FGFR1*, and *PDGFRA* genes [93, 102].

5.3. Anti-PD-1 Immunotherapy in CRC Patients. Immune checkpoint blockage inhibition has provided an alternative option for treatment of metastatic solid tumors [111–115]. In particular, antibody-mediated blockade of programmed death 1 (PD-1) or programmed death ligand 1 (PD-L1) can induce durable disease-free survivals, albeit only in a portion of patients with advanced melanomas, lung cancers, and bladder cancers [114–116]. Several clinical trials targeting immune checkpoint have also been conducted in gastrointestinal cancers, including CRC [117]. In 2015, Le and colleagues reported that MSI is a marker to predict the benefit of pembrolizumab, an anti-PD-1 immune checkpoint inhibitor, in CRC and other solid tumors [91]. These results are consistent with the pathogenesis of MSI in hereditary or sporadic CRC with defect in the mismatch repair machinery. Defective mismatch repair causes hypermutation of the genome, including MSI, and generates tumor neoantigens [118, 119]. CRC with MSI also demonstrated highly upregulated expression of multiple immune checkpoint molecules, including PD-1 and PD-L1 [119]. These may explain the profound lymphocyte infiltration as well as better outcomes and response to anti-PD-1 monoclonal antibody in CRCs with MSI [91, 120, 121].

5.4. Predictive Biomarkers for Immune Checkpoint Blockage Therapy. Currently, researchers have been aggressively exploring potential markers to select candidates who will benefit from immune checkpoint blockage therapy. These include expressional levels and gene amplification of PD-L1 [116, 122–124], MSI status [91], and genomic hypermutation [91, 111, 125, 126]. Although strong PD-L1 expression has been reported in a subset of CRCs [127, 128] and appears to predict repose to anti-PD-L1 therapy [116, 122–124], the definition of PD-L1 positive CRC needs further standardization and validation in a clinical diagnostic setting. In addition, tumors in response to immune checkpoint blockage may not have strong PD-L1 expression [117]. MSI test is a routine assay in most molecular diagnostics laboratories for screening HNPCC. The revised NCCN guideline (version 2.2016) has suggested screening for HNPCC (Lynch syndrome) in CRC patients aged 70 year or younger and those older than 70 who meet the Bethesda guidelines (http://www.nccn.org/professionals/physician_gls/f_guidelines.asp). Further studies are warranted to elucidate if the MSI assay or immunohistochemical stain of the mismatch repair machinery components (MLH1, MSH2, MSH6, and PMS2) may become a standard of care for those solid tumors with a higher incidence of MSI, such as CRC and endometrial cancer [129].

Mutation load may prove to be the most important predictor for immune checkpoint blockage therapy [91, 111, 125, 126]. MSI occurs in a subset of hypermutated tumors with defective mismatch repair machinery. There are many mechanisms causing a mutator phenotype with hypermutations, such as exposure to external mutagens (e.g., cigarette smoke and UV radiation), endogenous mutagens (e.g., reactive oxygen species), and mutations in the *POLE* or *POLD1* genes encoding the DNA polymerases [130]. Tumors with hypermutation caused by these alternative mechanisms are anticipated to have enhanced sensitivity to checkpoint blockade. Germline or somatic mutations in the regions encoding exonuclease domain of POLE and POLD1 impair polymerase proofreading and lead to an exceedingly high rate of base substitution mutations [131–133]. Results from the Cancer Genome Atlas showed hypermutation in 16% of CRC, including three-quarters with MSI and one-quarter with *POLE*

Table 1: Molecular diagnostics markers of CRC currently used in established standard care or potentially being used in the near future.

Clinical utility	Markers*
Diagnostic markers	*BRAF* [79, 80]; MSI [129]
Predictive markers	
(i) Primary resistance to anti-EGFR mAb	*KRAS* and *NRAS* [7, 8, 12, 14–19]; *BRAF* [7, 20–22]; *PIK3CA* [7]; other potential markers (*ERBB2*, *EGFR*, *FGFR1*, *PDGFRA*, and *MAP2K1*) [102]
(ii) Secondary resistance to anti-EGFR mAb	Mutations in the MAPK pathway (*KRAS*, *NRAS*, *BRAF*, and *MAP2K1*) [97–99]; *EGFR* [93, 100–102]; amplification of *KRAS*; *MET; ERBB2* [93–95]
(iii) Immune checkpoint blockage therapy	PD-L1 expression [116, 122–124]; MSI [91]; genomic hypermutation [91, 111, 125, 129]
Prognostic markers	*BRAF* [81–83]; MSI [84, 85]; *POLE* [135]

*The numbers in the parentheses indicate the references cited in this article.

mutations [134]. A recently study [135] investigating association between *POLE* mutations and prognosis in more than 4500 stage II/III CRC patients shows the pathogenic somatic *POLE* mutations were detected in approximately 1.0% of CRCs. *POLE* mutations were mutually exclusive with MSI. Compared with MSS *POLE* wild-type CRCs, *POLE*-mutant CRCs showed increased CD8+ lymphocyte infiltration and expression of cytotoxic T-cell markers and effector cytokines, with a level similar to that observed in immunogenic MSI tumors. Both *POLE* mutations and MSI status were associated with significantly reduced risk of recurrence compared with MSS CRCs in multivariable analysis. Higher *POLE* mutation rates (7–12%) have also been reported in endometrial cancers [136, 137], similarly characterized by a robust intratumoral T-cell response [138], and carry an excellent outcome [136, 137]. *POLE* mutations in CRC, though uncommon, may be associated with a favorable response to anti-PD-1 or anti-PD-L1 immunotherapy. However, the excellent prognosis demonstrated in this group of patients also underscores the importance of POLE mutations in precision medicine.

In summary, MSI is a routine clinical assay not only being suitable for screening for HNPCC but also being potentially predictive for immune checkpoint blockage therapy. However, testing for MSI status picks up only those CRCs with defective mismatch repair machinery, but not other hypermutated tumors. Whole exome sequencing is certainly a robust approach to define hypermutation. Currently, it may not be practical to perform whole exome sequencing as a daily clinical routine to select patients for immune checkpoint blockage therapy. Alternatively, NGS assays may be designed to include a panel of genes to identify tumors with a high mutation load for potential treatment with immunotherapies as well as driver mutations for targeted therapy [139]. Further clinical trials are also needed to guide through the complexity in selecting targeted therapeutic agents in combination or in sequence with immune checkpoint blockage inhibitors [140].

6. Conclusion

The established and/or potential key molecular markers in molecular diagnosis of CRC are briefly summarized in Table 1. Current guidelines from several organizations recommend testing extended *RAS* genes to select CRC patients for anti-EGFR targeted therapy. The *BRAF* p.V600E mutation is also highly likely to predict for anti-EGFR resistance. More studies are needed to elucidate the role of *PIK3CA* mutations. The convergence of recent advances in molecular technology and rapid expansion of targeted therapeutics is transforming the approach in clinical molecular diagnostic laboratories from the traditional "one test-one drug" paradigm to the multiplexed genotyping platform, especially NGS. In the clinical diagnostic setting, NGS assays demonstrate a high sensitivity, a broad reportable range, and a precise measurement of mutant allele frequency. Eventually, NGS will become the most cost-effective assay as more and more genetic alterations have demonstrated clinical utility. In this era of precision cancer medicine, pathologists play crucial roles in molecular diagnostics. They are responsible for preserving tumor tissues with adequate quantity and quality of nucleic acids for molecular tests. Tumor cellularity estimated by pathologists provides an important parameter for postanalytic quality assessment of the assays and results. Recently studies have identified more potential intrinsic resistance mutations and uncovered the landscape of acquired resistance mechanisms. Tumor heterogeneity associated with anti-EGFR resistance poses challenges to targeted therapeutics in CRC. Strategies have been proposed to overcome anti-EGFR resistance resulting from tumor heterogeneity. Immune checkpoint blockage therapy has emerged as a critical alternative option for metastatic cancers with durable progression-free survival, although long term follow-up has not been achieved. In addition to the MSI test, further efforts to develop assays amenable in clinical molecular diagnostic laboratories are warranted to select CRC patients for targeted therapy and/or immunotherapy.

Competing Interests

The authors declare that they have no competing interests.

References

[1] J. Ferlay, I. Soerjomataram, R. Dikshit et al., "Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012," *International Journal of Cancer*, vol. 136, no. 5, pp. E359–E386, 2015.

[2] R. Siegel, C. Desantis, and A. Jemal, "Colorectal cancer statistics, 2014," *CA: A Cancer Journal for Clinicians*, vol. 64, no. 2, pp. 104–117, 2014.

[3] H. Gharwan and H. Groninger, "Kinase inhibitors and monoclonal antibodies in oncology: clinical implications," *Nature Reviews Clinical Oncology*, vol. 13, pp. 209–227, 2016.

[4] J. A. Sosman, K. B. Kim, L. Schuchter et al., "Survival in BRAF V600-mutant advanced melanoma treated with vemurafenib," *The New England Journal of Medicine*, vol. 366, no. 8, pp. 707–714, 2012.

[5] A. Hauschild, J.-J. Grob, L. V. Demidov et al., "Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial," *The Lancet*, vol. 380, no. 9839, pp. 358–365, 2012.

[6] K. T. Flaherty, C. Robert, P. Hersey et al., "Improved survival with MEK inhibition in BRAF-mutated melanoma," *The New England Journal of Medicine*, vol. 367, no. 2, pp. 107–114, 2012.

[7] W. De Roock, B. Claes, D. Bernasconi et al., "Effects of KRAS, BRAF, NRAS, and PIK3CA mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis," *The Lancet Oncology*, vol. 11, no. 8, pp. 753–762, 2010.

[8] J.-Y. Douillard, K. S. Oliner, S. Siena et al., "Panitumumab-FOLFOX4 treatment and RAS mutations in colorectal cancer," *The New England Journal of Medicine*, vol. 369, no. 11, pp. 1023–1034, 2013.

[9] T. J. Lynch, D. W. Bell, R. Sordella et al., "Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib," *The New England Journal of Medicine*, vol. 350, no. 21, pp. 2129–2139, 2004.

[10] J. G. Paez, P. A. Jänne, J. C. Lee et al., "EGFR mutations in lung, cancer: correlation with clinical response to gefitinib therapy," *Science*, vol. 304, no. 5676, pp. 1497–1500, 2004.

[11] E. L. Kwak, Y.-J. Bang, D. R. Camidge et al., "Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer," *The New England Journal of Medicine*, vol. 363, no. 18, pp. 1693–1703, 2010.

[12] C. J. Allegra, J. M. Jessup, M. R. Somerfield et al., "American society of clinical oncology provisional clinical opinion: Testing for KRAS gene mutations in patients with metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy," *Journal of Clinical Oncology*, vol. 27, no. 12, pp. 2091–2096, 2009.

[13] V. L. Keedy, S. Temin, M. R. Somerfield et al., "American Society of clinical oncology provisional clinical opinion: epidermal growth factor receptor (EGFR) mutation testing for patients with advanced non-small-cell lung cancer considering first-line EGFR tyrosine kinase inhibitor therapy," *Journal of Clinical Oncology*, vol. 29, no. 15, pp. 2121–2127, 2011.

[14] M. J. Sorich, M. D. Wiese, A. Rowland, G. Kichenadasse, R. A. McKinnon, and C. S. Karapetis, "Extended RAS mutations and anti-EGFR monoclonal antibody survival benefit in metastatic colorectal cancer: a meta-analysis of randomized, controlled trials," *Annals of Oncology*, vol. 26, no. 1, pp. 13–21, 2015.

[15] L. S. Schwartzberg, F. Rivera, M. Karthaus et al., "PEAK: a randomized, multicenter phase II study of panitumumab plus modified fluorouracil, leucovorin, and oxaliplatin (mFOLFOX6) or bevacizumab plus mFOLFOX6 in patients with previously untreated, unresectable, wild-type KRAS exon 2 metastatic colorectal cancer," *Journal of Clinical Oncology*, vol. 32, no. 21, pp. 2240–2247, 2014.

[16] C. J. Allegra, R. Bryan Rumble, and R. L. Schilsky, "Extended RAS gene mutation testing in metastatic colorectal carcinoma to predict response to anti-epidermal growth factor receptor monoclonal antibody therapy: American society of clinical oncology provisional clinical opinion update 2015 summary," *Journal of Oncology Practice*, vol. 12, no. 2, pp. 180–181, 2016.

[17] N. A. C. S. Wong, D. Gonzalez, M. Salto-Tellez et al., "RAS testing of colorectal carcinoma—a guidance document from the Association of Clinical Pathologists Molecular Pathology and Diagnostics Group," *Journal of Clinical Pathology*, vol. 67, no. 9, pp. 751–757, 2014.

[18] H. Taniguchi, K. Yamazaki, T. Yoshino et al., "Japanese Society of Medical Oncology Clinical Guidelines: RAS (KRAS/NRAS) mutation testing in colorectal cancer patients," *Cancer Science*, vol. 106, no. 3, pp. 324–327, 2015.

[19] E. Van Cutsem, A. Cervantes, R. Adam et al., "ESMO consensus guidelines for the management of patients with metastatic colorectal cancer," *Annals of Oncology*, vol. 27, no. 8, pp. 1386–1422, 2016.

[20] F. Di Nicolantonio, M. Martini, F. Molinari et al., "Wild-type BRAF is required for response to panitumumab or cetuximab in metastatic colorectal cancer," *Journal of Clinical Oncology*, vol. 26, no. 35, pp. 5705–5712, 2008.

[21] C. Bokemeyer, E. V. Cutsem, P. Rougier et al., "Addition of cetuximab to chemotherapy as first-line treatment for KRAS wild-type metastatic colorectal cancer: pooled analysis of the CRYSTAL and OPUS randomised clinical trials," *European Journal of Cancer*, vol. 48, no. 10, pp. 1466–1475, 2012.

[22] F. Pietrantonio, F. Petrelli, A. Coinu et al., "Predictive role of BRAF mutations in patients with advanced colorectal cancer receiving cetuximab and panitumumab: A meta-analysis," *European Journal of Cancer*, vol. 51, no. 5, article no. 9390, pp. 587–594, 2015.

[23] G. Cathomas, "PIK3CA in colorectal cancer," *Frontiers in Oncology*, vol. 4, article 35, 2014.

[24] N. I. Lindeman, P. T. Cagle, M. B. Beasley et al., "Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology," *Journal of Thoracic Oncology*, vol. 8, no. 7, pp. 823–859, 2013.

[25] N. I. Lindeman, P. T. Cagle, M. B. Beasley et al., "Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the study of lung cancer, and Association for Molecular Pathology," *Archives of Pathology and Laboratory Medicine*, vol. 137, no. 6, pp. 828–860, 2013.

[26] N. I. Lindeman, P. T. Cagle, M. B. Beasley et al., "Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology," *The Journal of Molecular Diagnostics*, vol. 15, no. 4, pp. 415–453, 2013.

[27] N. B. Leighl, N. Rekhtman, W. A. Biermann et al., "Molecular testing for selection of patients with lung cancer for epidermal growth factor receptor and anaplastic lymphoma kinase tyrosine kinase inhibitors: American Society of Clinical Oncology endorsement of the College of American Pathologists/International Association for the Study of Lung Cancer/Association for Molecular Pathology Guideline," *Journal of Clinical Oncology*, vol. 32, no. 32, pp. 3673–3679, 2014.

[28] J. Dudley, L.-H. Tseng, L. Rooper et al., "Challenges posed to pathologists in the detection of KRAS mutations in colorectal

cancers," *Archives of Pathology and Laboratory Medicine*, vol. 139, no. 2, pp. 211–218, 2015.

[29] J. L. Hunt and S. D. Finkelstein, "Microdissection techniques for molecular testing in surgical pathology," *Archives of Pathology & Laboratory Medicine*, vol. 128, no. 12, pp. 1372–1378, 2004.

[30] J. H. J. M. Van Krieken, A. Jung, T. Kirchner et al., "KRAS mutation testing for predicting response to anti-EGFR therapy for colorectal carcinoma: proposal for an European quality assurance program," *Virchows Archiv*, vol. 453, no. 5, pp. 417–431, 2008.

[31] G. Zheng, H. Tsai, L. Tseng et al., "Test feasibility of next-generation sequencing assays in clinical mutation detection of small biopsy and fine needle aspiration specimens," *American Journal of Clinical Pathology*, vol. 145, no. 5, pp. 696–702, 2016.

[32] V. M. Singh, R. C. Salunga, V. J. Huang et al., "Analysis of the effect of various decalcification agents on the quantity and quality of nucleic acid (DNA and RNA) recovered from bone biopsies," *Annals of Diagnostic Pathology*, vol. 17, no. 4, pp. 322–326, 2013.

[33] T. Reineke, B. Jenni, M.-T. Abdou et al., "Ultrasonic decalcification offers new perspectives for rapid FISH, DNA, and RT-PCR analysis in bone marrow trephines," *The American Journal of Surgical Pathology*, vol. 30, no. 7, pp. 892–896, 2006.

[34] C. L. Wickham, P. Sarsfield, M. V. Joyner, D. B. Jones, S. Ellard, and B. Wilkins, "Formic acid decalcification of bone marrow trephines degrades DNA: alternative use of EDTA allows the amplification and sequencing of relatively long PCR products," *Molecular Pathology*, vol. 53, no. 6, p. 336, 2000.

[35] G. Zheng, M. Lin, P. M. Lokhandwala et al., "Clinical mutational profiling of bone metastases of lung and colon carcinoma and malignant melanoma using next-generation sequencing," *Cancer Cytopathology*, 2016.

[36] L. Haley, L.-H. Tseng, G. Zheng et al., "Performance characteristics of next-generation sequencing in clinical mutation detection of colorectal cancers," *Modern Pathology*, vol. 28, no. 10, pp. 1390–1399, 2015.

[37] G. M. Frampton, A. Fichtenholtz, G. A. Otto et al., "Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing," *Nature Biotechnology*, vol. 31, no. 11, pp. 1023–1031, 2013.

[38] B. Angulo, E. García-García, R. Martínez et al., "A commercial real-time PCR kit provides greater sensitivity than direct sequencing to detect KRAS mutations: a morphology-based approach in colorectal carcinoma," *The Journal of Molecular Diagnostics*, vol. 12, no. 3, pp. 292–299, 2010.

[39] J. Tol, J. R. Dijkstra, M. E. Vink-Börger et al., "High sensitivity of both sequencing and real-time PCR analysis of KRAS mutations in colorectal cancer tissue," *Journal of Cellular and Molecular Medicine*, vol. 14, no. 8, pp. 2122–2131, 2010.

[40] W. Weichert, C. Schewe, A. Lehmann et al., "KRAS genotyping of paraffin-embedded colorectal cancer tissue in routine diagnostics: comparison of methods and impact of histology," *The Journal of Molecular Diagnostics*, vol. 12, no. 1, pp. 35–42, 2010.

[41] S. L. Ondrejka, D. F. Schaeffer, M. A. Jakubowski, D. A. Owen, and M. P. Bronner, "Does neoadjuvant therapy alter KRAS and/or MSI results in rectal adenocarcinoma testing?" *The American Journal of Surgical Pathology*, vol. 35, no. 9, pp. 1327–1330, 2011.

[42] F. Boissière-Michot, E. Lopez-Crapez, H. Frugier et al., "KRAS genotyping in rectal adenocarcinoma specimens with low tumor cellularity after neoadjuvant treatment," *Modern Pathology*, vol. 25, no. 5, pp. 731–739, 2012.

[43] J. C. Dudley, G. T. Gurda, L.-H. Tseng et al., "Tumor cellularity as a quality assurance measure for accurate clinical detection of braf mutations in melanoma," *Molecular Diagnosis and Therapy*, vol. 18, no. 4, pp. 409–418, 2014.

[44] L. A. Diaz Jr., R. T. Williams, J. Wu et al., "The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers," *Nature*, vol. 486, no. 7404, pp. 537–540, 2012.

[45] P. Laurent-Puig, D. Pekin, C. Normand et al., "Clinical relevance of KRAS-mutated subclones detected with picodroplet digital PCR in advanced colorectal cancer treated with anti-EGFR therapy," *Clinical Cancer Research*, vol. 21, no. 5, pp. 1087–1097, 2015.

[46] H. Viray, K. Li, T. A. Long et al., "A prospective, multi-Institutional diagnostic trial to determine pathologist accuracy in estimation of percentage of malignant cells," *Archives of Pathology & Laboratory Medicine*, vol. 137, no. 11, pp. 1545–1549, 2013.

[47] A. J. J. Smits, J. A. Kummer, P. C. De Bruin et al., "The estimation of tumor cell percentage for molecular testing by pathologists is not accurate," *Modern Pathology*, vol. 27, no. 2, pp. 168–174, 2014.

[48] G. Chen, J. Dudley, L.-H. Tseng et al., "Lymph node metastases of melanoma: challenges for BRAF mutation detection," *Human Pathology*, vol. 46, no. 1, pp. 113–119, 2015.

[49] S. Anderson, K. J. Bloom, D. U. Vallera et al., "Multisite analytic performance studies of a real-time polymerase chain reaction assay for the detection of BRAF V600E mutations in formalin-fixed, paraffin- embedded tissue specimens of malignant melanoma," *Archives of Pathology and Laboratory Medicine*, vol. 136, no. 11, pp. 1385–1391, 2012.

[50] H. Halait, K. Demartin, S. Shah et al., "Analytical performance of a real-time PCR-based assay for V600 mutations in the BRAF gene, used as the companion diagnostic test for the novel BRAF inhibitor vemurafenib in metastatic melanoma," *Diagnostic Molecular Pathology*, vol. 21, no. 1, pp. 1–8, 2012.

[51] C. T. Harbison, C. E. Horak, J.-M. Ledeine et al., "Validation of companion diagnostic for detection of mutations in codons 12 and 13 of the KRAS gene in patients with metastatic colorectal cancer: analysis of the NCIC CTG CO.17 trial," *Archives of Pathology & Laboratory Medicine*, vol. 137, no. 6, pp. 820–827, 2013.

[52] A. Harlé, B. Busser, M. Rouyer et al., "Comparison of COBAS 4800 KRAS, TaqMan PCR and High Resolution Melting PCR assays for the detection of KRAS somatic mutations in formalin-fixed paraffin embedded colorectal carcinomas," *Virchows Archiv*, vol. 462, no. 3, pp. 329–335, 2013.

[53] A. C. Tsiatis, A. Norris-Kirby, R. G. Rich et al., "Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications," *The Journal of Molecular Diagnostics*, vol. 12, no. 4, pp. 425–432, 2010.

[54] H. Do, M. Krypuy, P. L. Mitchell, S. B. Fox, and A. Dobrovic, "High resolution melting analysis for rapid and sensitive EGFR and KRAS mutation detection in formalin fixed paraffin embedded biopsies," *BMC Cancer*, vol. 8, article 142, 2008.

[55] T. Kobunai, T. Watanabe, Y. Yamamoto, and K. Eshima, "The frequency of KRAS mutation detection in human colon carcinoma is influenced by the sensitivity of assay methodology: a comparison between direct sequencing and real-time PCR," *Biochemical and Biophysical Research Communications*, vol. 395, no. 1, pp. 158–162, 2010.

[56] K. Oliner, T. Juan, S. Suggs et al., "A comparability study of 5 commercial KRAS tests," *Diagnostic Pathology*, vol. 5, article 23, 2010.

[57] P. Szankasi, N. S. Reading, C. P. Vaughn, J. T. Prchal, D. W. Bahler, and T. W. Kelley, "A quantitative allele-specific PCR test for the BRAF V600E mutation using a single heterozygous control plasmid for quantitation: a model for qPCR testing without standard curves," *The Journal of Molecular Diagnostics*, vol. 15, no. 2, pp. 248–254, 2013.

[58] G. V. Long, J. S. Wilmott, D. Capper et al., "Immunohistochemistry is highly sensitive and specific for the detection of V600E BRAF mutation in melanoma," *American Journal of Surgical Pathology*, vol. 37, no. 1, pp. 61–65, 2013.

[59] C. A. Routhier, M. C. Mochel, K. Lynch, D. Dias-Santagata, D. N. Louis, and M. P. Hoang, "Comparison of 2 monoclonal antibodies for immunohistochemical detection of BRAF V600E mutation in malignant melanoma, pulmonary carcinoma, gastrointestinal carcinoma, thyroid carcinoma, and gliomas," *Human Pathology*, vol. 44, no. 11, pp. 2563–2570, 2013.

[60] R. Dienstmann, J. Rodon, J. Barretina, and J. Tabernero, "Genomic medicine frontier in human solid tumors: prospects and challenges," *Journal of Clinical Oncology*, vol. 31, no. 15, pp. 1874–1884, 2013.

[61] K. K. Deeb, J. P. Sram, H. Gao, and M. G. Fakih, "Multigene assays in metastatic colorectal cancer," *Journal of the National Comprehensive Cancer Network*, vol. 11, supplement 4, pp. S9–S17, 2013.

[62] W. O. Greaves, S. Verma, K. P. Patel et al., "Frequency and spectrum of BRAF mutations in a retrospective, single-institution study of 1112 cases of melanoma," *Journal of Molecular Diagnostics*, vol. 15, no. 2, pp. 220–226, 2013.

[63] S. Magnin, E. Viel, A. Baraquin et al., "A multiplex SNaPshot assay as a rapid method for detecting KRAS and BRAF mutations in advanced colorectal cancers," *The Journal of Molecular Diagnostics*, vol. 13, no. 5, pp. 485–492, 2011.

[64] A. S. Gargis, L. Kalman, M. W. Berry et al., "Assuring the quality of next-generation sequencing in clinical laboratory practice," *Nature Biotechnology*, vol. 30, no. 11, pp. 1033–1036, 2012.

[65] A. S. Gargis, L. Kalman, D. P. Bick et al., "Good laboratory practice for clinical next-generation sequencing informatics pipelines," *Nature Biotechnology*, vol. 33, no. 7, pp. 689–693, 2015.

[66] N. Aziz, Q. Zhao, L. Bry et al., "College of American pathologists' laboratory standards for next-generation sequencing clinical tests," *Archives of Pathology & Laboratory Medicine*, vol. 139, no. 4, pp. 481–493, 2015.

[67] M.-T. Lin, S. L. Mosier, M. Thiess et al., "Clinical validation of KRAS, BRAF, and EGFR mutation detection using next-generation sequencing," *American Journal of Clinical Pathology*, vol. 141, no. 6, pp. 856–866, 2014.

[68] A. G. Hadd, J. Houghton, A. Choudhary et al., "Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens," *Journal of Molecular Diagnostics*, vol. 15, no. 2, pp. 234–247, 2013.

[69] R. R. Singh, K. P. Patel, M. J. Routbort et al., "Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes," *The Journal of Molecular Diagnostics*, vol. 15, no. 5, pp. 607–622, 2013.

[70] C. E. Cottrell, H. Al-Kateb, A. J. Bredemeyer et al., "Validation of a next-generation sequencing assay for clinical molecular oncology," *The Journal of Molecular Diagnostics*, vol. 16, no. 1, pp. 89–105, 2014.

[71] C. C. Pritchard, S. J. Salipante, K. Koehler et al., "Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens," *The Journal of Molecular Diagnostics*, vol. 16, no. 1, pp. 56–67, 2014.

[72] H. Beltran, K. Eng, J. M. Mosquera et al., "Whole-exome sequencing of metastatic cancer and biomarkers of treatment response," *JAMA Oncology*, vol. 1, no. 4, pp. 466–474, 2015.

[73] D. W. Parsons, A. Roy, Y. Yang et al., "Diagnostic yield of clinical tumor and germline whole-exome sequencing for children with solid tumors," *JAMA Oncology*, vol. 2, no. 5, pp. 616–624, 2016.

[74] J. Carter, L.-H. Tseng, G. Zheng et al., "Non-p.V600E BRAF mutations are common using a more sensitive and broad detection tool," *American Journal of Clinical Pathology*, vol. 144, no. 4, pp. 620–628, 2015.

[75] G. Zheng, L.-H. Tseng, G. Chen et al., "Clinical detection and categorization of uncommon and concomitant mutations involving BRAF," *BMC Cancer*, vol. 15, article 779, 2015.

[76] P. T. C. Wan, M. J. Garnett, S. M. Roe et al., "Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF," *Cell*, vol. 116, no. 6, pp. 855–867, 2004.

[77] S. J. Heidorn, C. Milagre, S. Whittaker et al., "Kinase-dead BRAF and oncogenic RAS cooperate to drive tumor progression through CRAF," *Cell*, vol. 140, no. 2, pp. 209–221, 2010.

[78] A. N. Bartley and S. R. Hamilton, "Select biomarkers for tumors of the gastrointestinal tract: present and future," *Archives of Pathology & Laboratory Medicine*, vol. 139, no. 4, pp. 457–468, 2015.

[79] G. Deng, I. Bell, S. Crawley et al., "BRAF mutation is frequently present in sporadic colorectal cancer with methylated hMLH1, but not in hereditary nonpolyposis colorectal cancer," *Clinical Cancer Research*, vol. 10, no. 1 I, pp. 191–195, 2004.

[80] W. K. Funkhouser Jr., I. M. Lubin, F. A. Monzon et al., "Relevance, Pathogenesis, and testing algorithm for mismatch repair-defective colorectal carcinomas: a report of the association for molecular pathology," *Journal of Molecular Diagnostics*, vol. 14, no. 2, pp. 91–103, 2012.

[81] P. Lochhead, A. Kuchiba, Y. Imamura et al., "Microsatellite instability and braf mutation testing in colorectal cancer prognostication," *Journal of the National Cancer Institute*, vol. 105, no. 15, pp. 1151–1156, 2013.

[82] A. I. Phipps, P. J. Limburg, J. A. Baron et al., "Association between molecular subtypes of colorectal cancer and patient survival," *Gastroenterology*, vol. 148, no. 1, pp. 77.e2–87.e2, 2015.

[83] S. Venderbosch, I. D. Nagtegaal, T. S. Maughan et al., "Mismatch repair status and BRAF mutation status in metastatic colorectal cancer patients: a pooled analysis of the CAIRO, CAIRO2, COIN, and FOCUS studies," *Clinical Cancer Research*, vol. 20, no. 20, pp. 5322–5330, 2014.

[84] S. Popat, R. Hubner, and R. S. Houlston, "Systematic review of microsatellite instability and colorectal cancer prognosis," *Journal of Clinical Oncology*, vol. 23, no. 3, pp. 609–618, 2005.

[85] J. Goldstein, B. Tran, J. Ensor et al., "Multicenter retrospective analysis of metastatic colorectal cancer (CRC) with high-level microsatellite instability (MSI-H)," *Annals of Oncology*, vol. 25, no. 5, pp. 1032–1038, 2014.

[86] C. M. Ribic, D. J. Sargent, M. J. Moore et al., "Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer," *The New England Journal of Medicine*, vol. 349, no. 3, pp. 247–257, 2003.

[87] J. M. Carethers, E. J. Smith, C. A. Behling et al., "Use of 5-fluorouracil and survival in patients with microsatellite-unstable colorectal cancer," *Gastroenterology*, vol. 126, no. 2, pp. 394–401, 2004.

[88] D. J. Sargent, S. Marsoni, G. Monges et al., "Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer," *Journal of Clinical Oncology*, vol. 28, no. 20, pp. 3219–3226, 2010.

[89] F. A. Sinicrope, N. R. Foster, S. N. Thibodeau et al., "DNA mismatch repair status and colon cancer recurrence and survival in clinical trials of 5-fluorouracil-based adjuvant therapy," *Journal of the National Cancer Institute*, vol. 103, no. 11, pp. 863–875, 2011.

[90] E. M. Webber, T. L. Kauffman, E. O'Connor, and K. A. Goddard, "Systematic review of the predictive effect of MSI status in colorectal cancer patients undergoing 5FU-based chemotherapy," *BMC Cancer*, vol. 15, article 156, 2015.

[91] D. T. Le, J. N. Uram, H. Wang et al., "PD-1 blockade in tumors with mismatch-repair deficiency," *The New England Journal of Medicine*, vol. 372, no. 26, pp. 2509–2520, 2015.

[92] C.-S. Tan, D. Gilligan, and S. Pacey, "Treatment approaches for EGFR-inhibitor-resistant patients with non-small-cell lung cancer," *The Lancet Oncology*, vol. 16, no. 9, pp. e447–e459, 2015.

[93] G. Siravegna, B. Mussolin, M. Buscarino et al., "Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients," *Nature Medicine*, vol. 21, pp. 795–801, 2015.

[94] S. Misale, S. Arena, S. Lamba et al., "Blockade of EGFR and MEK intercepts heterogeneous mechanisms of acquired resistance to anti-EGFR therapies in colorectal cancer," *Science Translational Medicine*, vol. 6, no. 224, Article ID 224ra26, 2014.

[95] C. Bettegowda, M. Sausen, R. J. Leary et al., "Detection of circulating tumor DNA in early- and late-stage human malignancies," *Science Translational Medicine*, vol. 6, no. 224, 2014.

[96] S. Misale, R. Yaeger, S. Hobor et al., "Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer," *Nature*, vol. 486, no. 7404, pp. 532–536, 2012.

[97] S. Misale, F. Di Nicolantonio, A. Sartore-Bianchi, S. Siena, and A. Bardelli, "Resistance to anti-EGFR therapy in colorectal cancer: From heterogeneity to convergent evolution," *Cancer Discovery*, vol. 4, no. 11, pp. 1269–1280, 2014.

[98] G. Bronte, N. Silvestris, M. Castiglia et al., "New findings on primary and acquired resistance to anti-EGFR therapy in metastatic colorectal cancer: do all roads lead to RAS?" *Oncotarget*, vol. 6, no. 28, pp. 24780–24796, 2015.

[99] M. Russo, G. Siravegna, L. S. Blaszkowsky et al., "Tumor heterogeneity and lesion-specific response to targeted therapy in colorectal cancer," *Cancer Discovery*, vol. 6, no. 2, pp. 147–153, 2016.

[100] C. Montagut, A. Dalmases, B. Bellosillo et al., "Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer (vol 18, pg 221, 2012)," *Nature Medicine*, vol. 18, no. 9, p. 1445, 2012.

[101] S. Arena, G. Siravegna, B. Mussolin et al., "MM-151 overcomes acquired resistance to cetuximab and panitumumab in colorectal cancers harboring EGFR extracellular domain mutations," *Science Translational Medicine*, vol. 8, no. 324, Article ID 324ra14, 2016.

[102] A. Bertotti, E. Papp, S. Jones et al., "The genomic landscape of response to EGFR blockade in colorectal cancer," *Nature*, vol. 526, no. 7572, pp. 263–267, 2015.

[103] C. Esposito, A. M. Rachiglio, M. L. La Porta et al., "The S492R EGFR ectodomain mutation is never detected in KRAS wild-type colorectal carcinoma before exposure to EGFR monoclonal antibodies," *Cancer Biology and Therapy*, vol. 14, no. 12, pp. 1143–1146, 2013.

[104] M. Gerlinger, A. J. Rowan, S. Horswell et al., "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing," *The New England Journal of Medicine*, vol. 366, no. 10, pp. 883–892, 2012.

[105] R. A. Burrell, N. McGranahan, J. Bartek, and C. Swanton, "The causes and consequences of genetic heterogeneity in cancer evolution," *Nature*, vol. 501, no. 7467, pp. 338–345, 2013.

[106] S. Maheswaran, L. V. Sequist, S. Nagrath et al., "Detection of mutations in EGFR in circulating lung-cancer cells," *The New England Journal of Medicine*, vol. 359, no. 4, pp. 366–377, 2008.

[107] K.-Y. Su, H.-Y. Chen, K.-C. Li et al., "Pretreatment Epidermal Growth Factor Receptor (EGFR) T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small-cell lung cancer," *Journal of Clinical Oncology*, vol. 30, no. 4, pp. 433–440, 2012.

[108] M. Furugen, K. Uechi, J. Hirai et al., "An autopsy case of two distinct, acquired drug resistance mechanisms in epidermal growth factor receptor-mutant lung adenocarcinoma: small cell carcinoma transformation and epidermal growth factor receptor T790M mutation," *Internal Medicine*, vol. 54, no. 19, pp. 2491–2496, 2015.

[109] K. Suda, I. Murakami, K. Sakai et al., "Small cell lung cancer transformation and T790M mutation: complimentary roles in acquired resistance to kinase inhibitors in lung cancer," *Scientific Reports*, vol. 5, Article ID 14447, 2015.

[110] M. J. Niederst, L. V. Sequist, J. T. Poirier et al., "RB loss in resistant EGFR mutant lung adenocarcinomas that transform to small-cell lung cancer," *Nature Communications*, vol. 6, article 6377, 2015.

[111] A. Snyder, V. Makarov, T. Merghoub et al., "Genetic basis for clinical response to CTLA-4 blockade in melanoma," *The New England Journal of Medicine*, vol. 371, no. 23, pp. 2189–2199, 2014.

[112] F. S. Hodi, S. J. O'Day, D. F. McDermott et al., "Improved survival with ipilimumab in patients with metastatic melanoma," *The New England Journal of Medicine*, vol. 363, no. 8, pp. 711–723, 2010.

[113] C. Robert, L. Thomas, I. Bondarenko et al., "Ipilimumab plus dacarbazine for previously untreated metastatic melanoma," *The New England Journal of Medicine*, vol. 364, no. 26, pp. 2517–2526, 2011.

[114] J. R. Brahmer, S. S. Tykodi, L. Q. M. Chow et al., "Safety and activity of anti-PD-L1 antibody in patients with advanced cancer," *The New England Journal of Medicine*, vol. 366, no. 26, pp. 2455–2465, 2012.

[115] S. L. Topalian, F. S. Hodi, J. R. Brahmer et al., "Safety, activity, and immune correlates of anti-PD-1 antibody in cancer," *The New England Journal of Medicine*, vol. 366, no. 26, pp. 2443–2454, 2012.

[116] T. Powles, J. P. Eder, G. D. Fine et al., "MPDL3280A (anti-PD-L1) treatment leads to clinical activity in metastatic bladder cancer," *Nature*, vol. 515, no. 7528, pp. 558–562, 2014.

[117] E. de Guillebon, P. Roussille, E. Frouin, and D. Tougeron, "Anti program death-1/anti program death-ligand 1 in digestive cancers," *World Journal of Gastrointestinal Oncology*, vol. 7, no. 8, pp. 95–101, 2015.

[118] J. R. Eshleman, E. Z. Lang, G. K. Bowerfind et al., "Increased mutation rate at the hprt locus accompanies microsatellite instability in colon cancer," *Oncogene*, vol. 10, no. 1, pp. 33–37, 1995.

[119] N. J. Llosa, M. Cruise, A. Tam et al., "The vigorous immune microenvironment of microsatellite instable colon cancer is

balanced by multiple counter-inhibitory checkpoints," *Cancer Discovery*, vol. 5, no. 1, pp. 43–51, 2015.

[120] F. Pagès, A. Berger, M. Camus et al., "Effector memory T cells, early metastasis, and survival in colorectal cancer," *The New England Journal of Medicine*, vol. 353, no. 25, pp. 2654–2666, 2005.

[121] S. D. Brown, R. L. Warren, E. A. Gibb et al., "Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival," *Genome Research*, vol. 24, no. 5, pp. 743–750, 2014.

[122] S. M. Ansell, A. M. Lesokhin, I. Borrello et al., "PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma," *The New England Journal of Medicine*, vol. 372, no. 4, pp. 311–319, 2015.

[123] R. S. Herbst, J.-C. Soria, M. Kowanetz et al., "Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients," *Nature*, vol. 515, no. 7528, pp. 563–567, 2014.

[124] F. Passiglia, G. Bronte, V. Bazan et al., "PD-L1 expression as predictive biomarker in patients with NSCLC: a pooled analysis," *Oncotarget*, 2016.

[125] N. A. Rizvi, M. D. Hellmann, A. Snyder et al., "Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer," *Science*, vol. 348, no. 6230, pp. 124–128, 2015.

[126] E. M. Van Allen, D. Miao, B. Schilling et al., "Genomic correlates of response to CTLA-4 blockade in metastatic melanoma," *Science*, vol. 350, no. 6257, pp. 207–211, 2015.

[127] H. Dong, S. E. Strome, D. R. Salomao et al., "Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion," *Nature Medicine*, vol. 8, no. 8, pp. 793–800, 2002.

[128] R. A. Droeser, C. Hirt, C. T. Viehl et al., "Clinical impact of programmed cell death ligand 1 expression in colorectal cancer," *European Journal of Cancer*, vol. 49, no. 9, pp. 2233–2242, 2013.

[129] J. C. Dudley, M. Lin, D. T. Le, and J. R. Eshleman, "Microsatellite instability as a biomarker for PD-1 Blockade," *Clinical Cancer Research*, vol. 22, no. 4, pp. 813–820, 2016.

[130] S. A. Roberts and D. A. Gordenin, "Hypermutation in human cancer genomes: footprints and mechanisms," *Nature Reviews Cancer*, vol. 14, no. 12, pp. 786–800, 2014.

[131] C. Palles, J. B. Cazier, K. M. Howarth et al., "Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas," *Nature Genetics*, vol. 45, no. 2, pp. 136–144, 2013.

[132] L. Valle, E. Hernández-Illán, F. Bellido et al., "New insights into POLE and POLD1 germline mutations in familial colorectal cancer and polyposis," *Human Molecular Genetics*, vol. 23, no. 13, pp. 3506–3512, 2014.

[133] S. Briggs and I. Tomlinson, "Germline and somatic polymerase ϵ and δ mutations define a new class of hypermutated colorectal and endometrial cancers," *The Journal of Pathology*, vol. 230, no. 2, pp. 148–153, 2013.

[134] The Cancer Genome Atlas Network, "Comprehensive molecular characterization of human colon and rectal cancer," *Nature*, vol. 487, no. 7407, pp. 330–337, 2012.

[135] E. Domingo, L. Freeman-Mills, E. Rayner et al., "Somatic POLE proofreading domain mutation, immune response, and prognosis in colorectal cancer: a retrospective, pooled biomarker study," *The Lancet Gastroenterology & Hepatology*, 2016.

[136] E. Stelloo, T. Bosse, R. A. Nout et al., "Refining prognosis and identifying targetable pathways for high-risk endometrial cancer; A TransPORTEC initiative," *Modern Pathology*, vol. 28, no. 6, pp. 836–844, 2015.

[137] M. K. McConechy, A. Talhouk, S. Leung et al., "Endometrial carcinomas with POLE exonuclease domain mutations have a favorable prognosis," *Clinical Cancer Research*, vol. 22, no. 12, pp. 2865–2873, 2016.

[138] I. C. Van Gool, F. A. Eggink, L. Freeman-Mills et al., "POLE proofreading mutations elicit an antitumor immune response in endometrial cancer," *Clinical Cancer Research*, vol. 21, no. 14, pp. 3347–3355, 2015.

[139] E. I. Lin, L.-H. Tseng, C. D. Gocke et al., "Mutational profiling of colorectal cancers with microsatellite instability," *Oncotarget*, vol. 6, no. 39, pp. 42334–42344, 2015.

[140] M. B. Atkins and J. Larkin, "Immunotherapy combined or sequenced with targeted therapy in the treatment of solid tumors: current perspectives," *Journal of the National Cancer Institute*, vol. 108, no. 6, Article ID djv414, 2016.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 5623089, 11 pages
http://dx.doi.org/10.1155/2016/5623089

Research Article

Improved Efficiency and Reliability of NGS Amplicon Sequencing Data Analysis for Genetic Diagnostic Procedures Using AGSA Software

Axel Poulet,[1] Maud Privat,[1,2] Flora Ponelle,[1,2] Sandrine Viala,[1,2] Stephanie Decousus,[1] Axel Perin,[1] Laurence Lafarge,[1,2] Marie Ollier,[2] Nagi S. El Saghir,[3] Nancy Uhrhammer,[1,2] Yves-Jean Bignon,[1,2] and Yannick Bidet[2]

[1] *Département d'Oncogénétique, Centre Jean Perrin, 63011 Clermont-Ferrand, France*
[2] *EA 4677 ERTICa, Université Clermont Auvergne, 63000 Clermont-Ferrand, France*
[3] *Naef K. Basile Cancer Institute (NKBCI), Breast Center of Excellence, American University of Beirut Medical Center, Beirut 1107 2020, Lebanon*

Correspondence should be addressed to Yves-Jean Bignon; yves-jean.bignon@cjp.fr

Received 25 April 2016; Accepted 28 June 2016

Academic Editor: Urzsula Demkow

Screening for *BRCA* mutations in women with familial risk of breast or ovarian cancer is an ideal situation for high-throughput sequencing, providing large amounts of low cost data. However, 454, Roche, and Ion Torrent, Thermo Fisher, technologies produce homopolymer-associated indel errors, complicating their use in routine diagnostics. We developed software, named AGSA, which helps to detect false positive mutations in homopolymeric sequences. Seventy-two familial breast cancer cases were analysed in parallel by amplicon 454 pyrosequencing and Sanger dideoxy sequencing for genetic variations of the *BRCA* genes. All 565 variants detected by dideoxy sequencing were also detected by pyrosequencing. Furthermore, pyrosequencing detected 42 variants that were missed with Sanger technique. Six amplicons contained homopolymer tracts in the coding sequence that were systematically misread by the software supplied by Roche. Read data plotted as histograms by AGSA software aided the analysis considerably and allowed validation of the majority of homopolymers. As an optimisation, additional 250 patients were analysed using microfluidic amplification of regions of interest (Access Array Fluidigm) of the BRCA genes, followed by 454 sequencing and AGSA analysis. AGSA complements a complete line of high-throughput diagnostic sequence analysis, reducing time and costs while increasing reliability, notably for homopolymer tracts.

1. Background

Mutations in the *BRCA1* and *BRCA2* genes are associated with high risk of breast and ovarian cancer [1, 2]. Mutation screening in families with multiple breast and/or ovarian cancer cases (hereditary breast and ovarian cancer (HBOC)) or exceptionally young cases has revealed a high frequency of germline mutations in these genes [3, 4]. To detect these mutations, genomic DNA sequencing remains the gold standard, as pathogenic variants can occur throughout the gene and the entire gene coding sequence must therefore be screened. Historically performed by PCR amplification and Sanger dideoxy sequencing, *BRCA* resequencing is time-consuming and expensive. The major technological advances of new generation sequencing provide high-throughput strategies to reduce the time and expense of direct sequence analysis [5–7]. Furthermore, various methods of multiplex or microfluidic PCR reduce the time necessary to prepare samples before sequencing.

Among massively parallel sequencers, the GS-FLX and GS-Junior (Roche) and the PGM and Proton (Life Technologies) have the advantage of generating read lengths of up to 500 bases. Amplicons resembling or identical to those currently used for dideoxy sequencing can thus be analysed.

On the other hand, these sequencers have the inconvenience of generating numerous reading errors in homopolymeric sequences, which complicates the analysis. Many software approaches have been developed to correct NGS calling errors [8]. Some produce good results, but false indel detection in homopolymers remains challenging. Moreover, these software solutions often require significant computational power and bioinformatics skills that are difficult to maintain in a routine medical diagnostic laboratory.

This study addresses the suitability of pyrosequencing technology associated with in-house developed software for molecular diagnosis, in terms of performance, robustness, and reliability. In preparation for high-throughput analysis of HBOC families, our results on cases blindly analysed with both techniques validate sequencing on the GS-FLX for mutation detection in the *BRCA1* and *BRCA2* genes.

2. Methods

2.1. Family Selection. 72 HBOC family samples used for technical comparisons were identified in the Oncogenetics consultation of the Centre Jean Perrin. 250 HBOC samples used for microfluidic amplification were identified in the Oncogenetics consultation of the Naef K. Basile Cancer Institute (NKBCI) at the American University of Beirut Medical Center (AUBMC, Beirut, Lebanon). DNA was isolated from peripheral blood by standard techniques, using NucleoSpin® Blood XL kit (Macherey-Nagel) at the Centre Jean Perrin and QIAmp DNA isolation kit (Qiagen) at the AUBMC.

2.2. Ethical Approval and Consent. All patients gave informed consent for analysis of breast cancer predisposition genes. Lebanese HBOC patients were recruited in a study of *BRCA1* and *BRCA2* mutations in Lebanon approved by IRB at AUBMC, granted by Ethnic Research Initiative (ERI) and sponsored by GlaxoSmithKline (GSK).

2.3. Amplicon Design. The amplicons traditionally used for Sanger dideoxy sequencing by our laboratory for the analysis of the *BRCA* genes ranged from 195 to 592 base pairs and included the entire coding sequence plus all intron-exon junctions. The *BRCA1* coding sequence (7,224 bp) was covered in 31 amplicons (10,823 nucleotides of sequence) and *BRCA2* (11,386 bp) in 44 amplicons (16,513 nucleotides of sequence). Large exons were covered by overlapping amplicons. For dideoxy sequencing, all forward primers carried an M13 forward extension and all reverse primers an M13 reverse extension.

For GS-FLX pyrosequencing, amplicons exceeding 500 bp were redesigned, resulting in 40 amplicons for *BRCA1* (10,006 nucleotides of sequence) and 44 for *BRCA2* (15,396 nucleotides of sequence), from 142 to 501 bp in length. The 454 "A" and "B" sequencing primer extensions were included at the 5′ ends of all forward and reverse PCR primers, respectively. Primers were also designed to homogenize amplification conditions.

For more uniform coverage, some poorly represented sequences were amplified in duplicate in the Access Array.

2.4. Sanger Dideoxy Sequencing. Amplicons were sequenced by standard techniques, using BigDye v3 (Applied Biosystems) reagents, and were resolved on a 3130XL capillary sequencer (Applied Biosystems). Sequences were aligned with reference sequences NM_007294.2 (*BRCA1*) and NM_000059.3 (*BRCA2*) using Seqman software (DNAStar). All variants were noted; deleterious mutations were confirmed by independent analysis of second samples.

2.5. Pyrosequencing. For standard amplification, 72 amplicon libraries were prepared from genomic DNA by amplification in 96-well plates, one plate per patient. After verification on agarose gels and AmpureXT purification according to Beckman Coulter's recommendations, amplicons were quantified with PicoGreen (Invitrogen) on an Infinite 200 plate reader (Tecan), according to Invitrogen's guidelines. Isomolar amplicon pools were prepared for each sample, and these libraries were diluted to 2×10^5 molecules/μL.

For microfluidic amplification, we used the Access Array System (Fluidigm) according to the manufacturer's instructions. This technology allows amplification of 48 amplicons for 48 patients at once. During the same PCR, Roche MID oligonucleotide adapters were added to allow multiplexing and identification of patients.

Emulsion-PCR, bead recovery and enrichment, and pyrosequencing were performed as recommended (emPCR LibA and sequencing kit, Roche).

2.6. Data Processing. Pyrosequencing data processing was performed using the software provided by Roche: raw images were automatically converted into sequences by gs Run Processor v2.3 and the sequences were aligned to the *BRCA* references by Amplicon Variant Analyzer (AVA v2.3). AVA uses a package called "newbler" developed specifically for 454 sequencing that does both alignment and variant calling [9]. It is then able to display variants graphically, with a histogram indicating both the number of reads at each position and the percentage of variant reads. Variations are also accessible with a color-coded multiple alignment which highlights regions and bases differing from the reference sequence. Read frequencies of variants were reported in a summary table for each sample.

2.7. Data Analysis. We developed a Java pipeline called AGSA which takes as input a calculated AVA project and the Genbank file of the gene(s) of interest. AGSA was developed to detect and annotate variants and generate easy-to-read results in a color-coded Excel result file.

Results were validated according to several adjustable parameters, such as read depth, the percentage of reads presenting a variant, and the presence of variants on both strands.

The minimum read depth was set to 40 for each nucleotide to validate an amplicon, and the minimum percentage of mutated reads to 20% by default (however, a variant can be detected in a nonvalidated amplicon). Furthermore, it was possible to specify the region of interest for each exon (default value −20 to +6 to include splicing sites). For a relevant comparison, the same settings were

used in the SeqNext Software. HGVS nomenclature was obtained for each variant thanks to a request to the Alamut servers (Interactive Biosoftware, Rouen, France). After this annotation, a homopolymer analysis step was added to help users decide if a detected indel variant was an artefact due to the technology or a real variant.

2.8. Homopolymer Analysis. The homopolymer analysis consists of the construction of a histogram showing two representations of the same dataset, for each putative variant located in a homopolymer. The distribution of the measured light intensity at the position considered (flowgram values) was represented in blue with a step of 1 to mimic AVA's interpretation of the sequence and in red with a step of 0.1 (Figure 2). Datasets were extracted from the sff files located in the AVA project.

2.9. Presentation of "AGSA" Software. AGSA was developed in order to facilitate 454 sequencing analysis of one or more gene(s) of interest. AGSA is a graphic interface software with few parameters to adjust: boundaries of the region of interest for each exon, threshold value for the minimum depth to validate an amplicon, and threshold frequency to accept a variant (screenshot in Supplementary Figure 1A, in Supplementary Material available online at http://dx.doi.org/10.1155/2016/5623089). To keep it simple, AGSA uses Amplicon Variant Analyser (AVA) output files as raw data and the Genbank file(s) of the studied gene(s) (Figure 1). During processing, several output files are created to control the sequencing quality by the coverage of each amplicon and each sample. Read depth is determined for each base of the region of interest, but only the minimum is used to consider an amplicon correctly sequenced or not. Detection of variants is calculated independently of read depth, to make sure that variants are not missed, even in poorly sequenced amplicons that should be repeated. To be validated and kept in the final report, a variant must be present on both strands (unless there were reads for only one strand) and in at least 4 reads (to eliminate random sequencing errors). The variant must also exceed the minimum percentage of reads defined in the parameters. These settings permit mutation calling in insufficiently covered amplicons, without generating many false positives. Then the program tests the nucleotides just before and after the position of the variant, to check if it is in a homopolymer context. When the variant is detected within a homopolymer, AGSA searches the sequence flowgram file for the values of all reads of the sample at the variant position. A histogram is created and saved as a .jpeg file. For a heterozygous insertion or deletion, the distribution of read values is split into two populations, showing that some reads have n identical bases and others $n + 1$ (insertion) or $n - 1$ (deletion) identical bases (Figure 2(a)). On the contrary, in case of artefacts, a single population is observed between n and $n + 1$ (or $n - 1$), resulting from poor quantification of the strong homopolymer signal (Figure 2(b)). In case of homozygous variation, a single population is observed too but it is centred on $n + 1$ (or $n - 1$), showing that all reads have the same number of bases in the homopolymer and that this number is different from the wild type. The graphs aid the interpretation of the data as one or two different alleles; the mono- or bimodal distribution of flowgram values may also be evaluated statistically (though we have not yet developed this function). AGSA extracts and graphs the relevant signal information from the raw data for all detected variants in a homopolymer context, for each sample. Variant annotation is then performed using Alamut Batch software v1.1.6 (Interactive Biosoftware, Rouen, France). Annotation includes the nomenclature recommended by the Human Genome Variation Society (HGVS), as well as predictions about the mutation impact determined by different algorithms. Finally, AGSA automatically annotates variants identified as neutral in a list supplied by the users (here the BIC and UMD-*BRCA* databases were used) and also reports the amplicons without variants. Any variants that have not been automatically classified must be interpreted manually by the user. To assist this step we developed a graphic interface with all information needed to annotate the variant (position, homopolymer context, graphic of homopolymer, variant frequency, coverage depth, and so forth, Supplementary Figure 1B). When all variants are sorted, the user can generate a report to summarize the annotation of the gene(s) of interest for all samples.

This software was created in interaction with end-users in order to simplify the software utilization and graphic interface. It is thus an easy-to-use tool, with very few parameters to manage. AGSA software was initially developed to analyse *BRCA1* and *BRCA2* for diagnosis purposes, but any combination of genes can be analysed.

3. Results and Discussion

3.1. Comparison of BRCA Analysis Using Sanger Sequencing and Roche Pyrosequencing with AGSA Analysis. One run of 8 samples and four runs of 16 samples for a total of 72 different patients were performed using GS-FLX chemistry. These 72 samples were analysed in parallel by Sanger dideoxy sequencing. All 72 samples were coded for blind analysis. Dideoxy sequencing and pyrosequencing were performed by different technicians. The Titanium technology was developed to produce 400 bp long reads. Although it was able to produce longer reads, a random sampling of 10% of the samples showed that the quality dropped quickly after 400 nucleotides (Supplementary Figure 2). 77% of nucleotides of the 400 bp long reads presented a quality score above Q20, which is the threshold used for Sanger sequencing, and 53% were above Q30. The longest fragments were covered on both ends by high quality base calls, with a large overlap in the middle.

Our own observations, as well as discussions with other groups and expert committees [7, 10], suggested limits for the analysis of pyrosequencing data. Amplicons were validated when no nucleotide in the region of interest was read less than 40 times, and heterozygous variants were validated when present in at least 20% of reads and represented on both strands. This latter threshold is lower than the smallest value found for the known mutations of the composite sample described below (see Section 3.3).

Because dideoxy sequencing remains the gold standard for diagnostic purposes, variants were classified according to

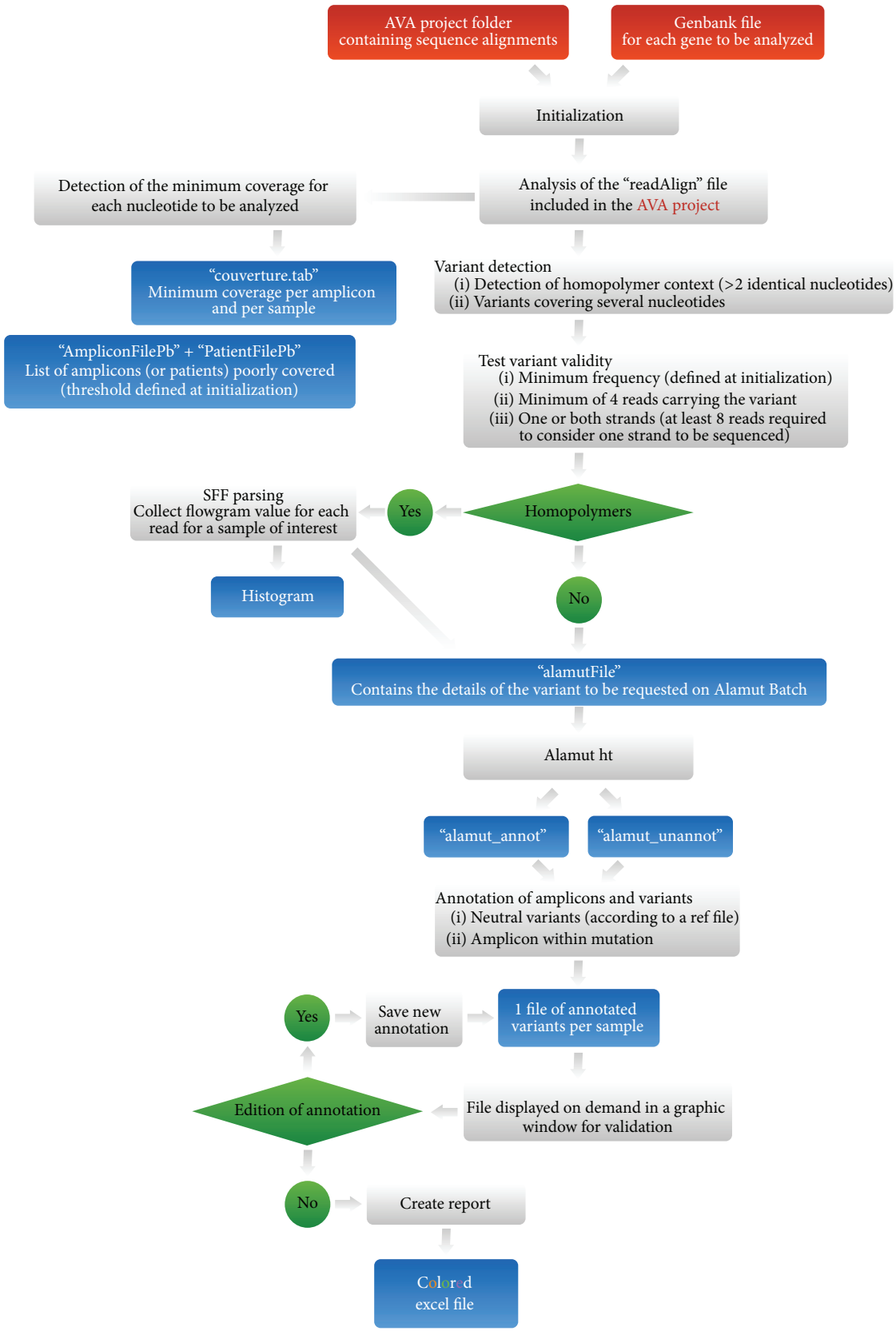


Figure 1: Organization of the AGSA software. The diagram represents the operational flow of the AGSA software. The red boxes represent input files required to operate the software. The blue boxes represent output files generated by AGSA. The green boxes are steps where the software performs a test.

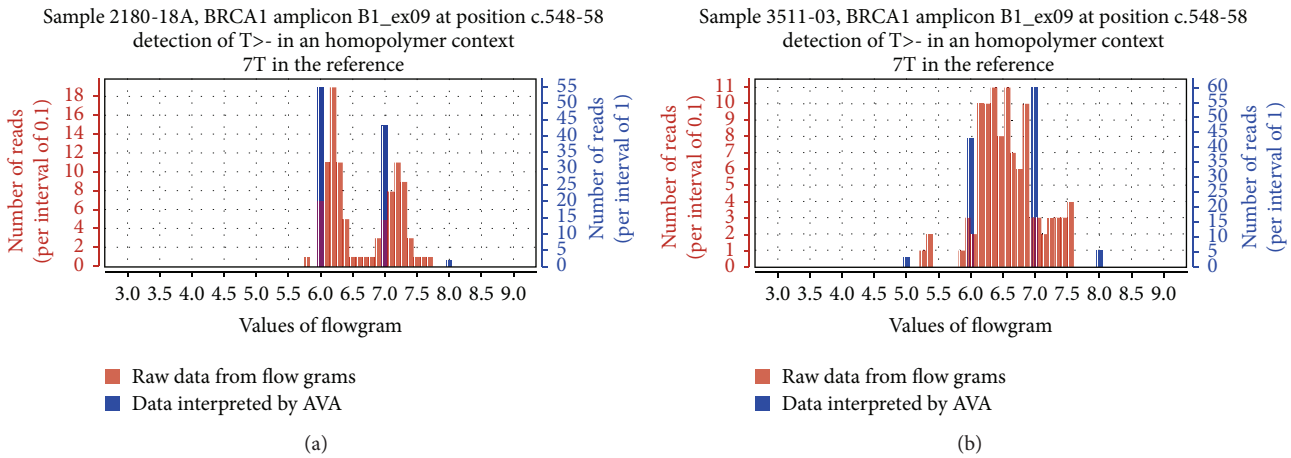


Figure 2: Graphic representations of homopolymer flowgrams. Individual flowgrams of indel variants were generated by AGSA. The X-axis represents the signal intensity computed during pyrosequencing. Red bars represent the number of reads for each intensity interval of 0.1; blue bars represent the percentage of variation as shown by AVA (intensity interval of 1). Distribution of standard flowgrams discriminates real indel (a) from artefacts (b).

their Sanger status: only variants detected by the method of reference were considered real. 564 variants were reported in the 72 patients tested by dideoxy sequencing (Figure 3). These variants consisted of substitutions and a wide range of insertions/deletions (1 to 29 nucleotides). They included frequent polymorphisms, deleterious mutations, and variants of unknown significance. Pyrosequencing followed by analysis with AGSA software reported 599 variants. As a comparison, analysis only with AVA generated a list of 3800 putative variants (data not shown). Despite the overall higher number of variants called after pyrosequencing, 61 variants detected by the Sanger technique were missing in the pyrosequencing reports. 56 of these were in amplicons not validated due to insufficient read depth and automatically targeted to be resequenced by standard techniques. The remaining 5 variants were Sanger false positives since a second Sanger analysis of these samples revealed a wild-type sequence. These false positives were due to information transfer errors or to poor quality Sanger sequences; none were potentially deleterious mutations.

Conversely, pyrosequencing revealed 95 variants reported by AGSA but not by Sanger analysis. When the Sanger traces were reviewed, 32 variants were actually present and can thus be considered false negatives of Sanger analysis. Most were information transfer errors; some were due to sequences more difficult to interpret. Excepting one unclassified variant in *BRCA2*, all were frequent polymorphisms. In total, pyrosequencing allowed correcting 37 errors in Sanger reports. We insist on the fact that these false positives or negatives were not generated by the Sanger technique itself. They were due to inevitable small rate of human errors when manually analysing large volumes of data in routine diagnostic procedures. The other 63 variants were falsely reported after analysis of the pyrosequencing data. Because Sanger sequencing is the technique of reference, these were considered artefactual variants due to poor quality reads in some amplicons (mainly in *BRCA2*), although, unlikely, some of them could however be real variants not detected by Sanger sequencing.

Nevertheless, most of the false positive variants represent technical limits of the 454 NGS sequencing in homopolymers. These were specifically treated by AGSA software.

3.2. AGSA Software Eliminates Most False Positives without Generating False Negatives. One bias of 454 pyrosequencing is the misreading of homopolymeric sequences that generates many false positive variants. To respond to this bias, AGSA software aims to determine if a reported variant in a homopolymer is a reading mistake or a true variant and presents this information to the user as explained above. For example, for a heterozygous deletion in a series of 7 A's, the distribution of raw data falls into two populations, showing that some reads have 7 identical bases and others 6 identical bases (Figure 2(a)). When this same sequence is misread, only one population is observed between 6 and 7 nucleotides resulting from poor quantitation of the strong homopolymer signal (Figure 2(b)). At this time, the classification of the graphs as representative of an artefact (monomodal distribution) or a real variant (bimodal distribution) is determined by the user.

To evaluate the accuracy of the result delivered by AGSA, we blindly assessed 299 putative variants in six different homopolymeric sequences (of which one is a common polymorphism with frequent heterozygotes and both classes of homozygotes), detected in our 72 patients. The histograms were interpreted independently by two persons. Although subjective, the interpretation of the histograms generated by AGSA was robust, as more than 90% of the cases were assessed concordantly. Most of the discordances related to one reader being not sure of the interpretation (Supplementary Figure 3). Discordances were not critical, however, as all doubtful sequences are flagged for further confirmation. 256 samples were defined as homozygous wild type and 43 as heterozygous (Figure 4). Dideoxy sequencing confirmed all

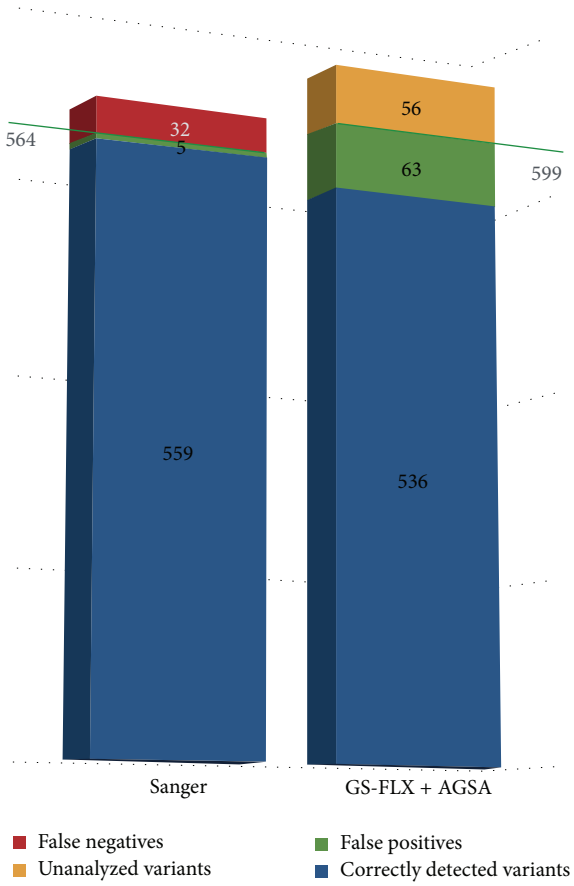


Figure 3: AGSA detected efficiently all the variants reported in Sanger analysis. Efficiencies of Sanger sequencing versus Roche pyrosequencing analysed with AGSA. The blue bars represent confirmed variants. The green bars represent false positive variants (technical artefacts). For pyrosequencing, false positives are defined as variants not confirmed by Sanger sequencing. For Sanger sequencing, false positives were not found by pyrosequencing and were not confirmed by a second Sanger run of the same sample. The red bars represent false negative variants. No false negative was found by pyrosequencing. False negatives for Sanger analysis were detected by pyrosequencing and they were actually found on a second Sanger run of the same sample. The yellow bar represents variants that were not called by AGSA because of poor coverage (inducing a number of variant reads < 4).

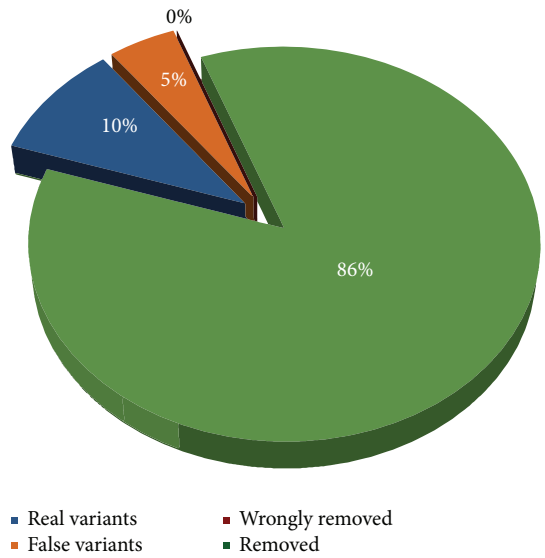


Figure 4: Performance of AGSA software for evaluation of homopolymers. 299 indel variants were found by AGSA in homopolymer sequences. After analysis of individual flowgrams, 246 (86%) were classified as false positive variants and 43 (14%) as true variants. Sanger sequencing confirmed that the 246 AGSA-classified false positives were actually wild-type sequences. Among the 43 potentially real variants, 29 (10%) were confirmed with Sanger analysis and 14 (5%) were actually wild-type Sanger sequences.

wild-type classified homopolymers as homozygous. Among the 43 heterozygous calls at the polymorphic homopolymer, 29 were heterozygous and 14 were homozygous.

With help of the homopolymer histograms, 86% of the indel variants in homopolymer contexts were eliminated, while no false negatives were created (Figure 4). Of the remaining variants, 10% were real variants and 5% were false positives.

3.3. Comparison of AGSA to SeqNext Software. One of the samples analysed in the test phase was a composite sample including 20 deleterious mutations and 8 polymorphisms. This sample was generated by pooling appropriate PCR products from different patients, with known *BRCA* variants ranging from single nucleotide substitutions to a 29-base-pair deletion, including transitions, transversions, single nucleotide insertions and deletions, and a deletion of one nucleotide in a homopolymer of eight. All expected variants were detected by AGSA (Figure 5 and Table 1), as well as 10 false positive variants that were not present in the Sanger analyses. Six of these were in amplicons with less than 40 reads, which could explain loss of specificity. However, seeking variants in amplicons with low coverage allowed detection of two real variants in *BRCA1* (c.19_47del and c.212+1G>A).

To compare the performance of AGSA to commercial software for NGS data analysis, we tested the SeqNext module of Sequence Pilot from JSI Medical Systems (version 4.1.2). To stay close to AGSA parameters and after discussions with the developer, we set parameters to read depth ≥ 40, 20% variant reads, and regions of interest from −20 to +6 per exon. With SeqNext, one false negative and 28 false positive variants were detected (Figure 5 and Table 1). Most of these false positives were in poorly sequenced amplicons. AGSA appears to deal better with these regions. The missed variant was an insertion of 39 nucleotides. According to the developer, it was missed because the sequence downstream of the insertion is too short to allow realignment. In this specific case, mutated reads are ignored. One nomenclature mistake was also detected with SeqNext (c.3839_3844delinsAGGCG instead of c.3839_3843delinsAGGC).

To validate the results obtained on this composite sample, SeqNext analysis was performed on a sampling of 39 *BRCA1* and 33 *BRCA2* analyses from the 72 test patients (Supplementary Figure 4). SeqNext detected all the variants found by

Table 1: Comparison of the variants detected by Sanger, by GS-Flx with SeqNext analysis, and by GS-Flx with AGSA analysis.

Gene	HGVS nomenclature	Sanger	AVA + AgsA	SeqNext
BRCA1	c.19-47del29	h	h (56%)**	—**
	c.81-12delC	—	—	h (41%)†
	c.124delA	h	h (48%)	h (48%)
	c.212+1G>A	h	h (72%)**	—**
	c.342-343delTC	h	h (36%)	h (37%)
	c.671-11dup	—	h (45%)†	h (43%)†
	c.798-799del	h	h (48%)	h (48%)
	c.1116G>A	h	h (43%)	h (43%)
	c.1390dupA	h	h (49%)	h (48%)
	c.1823-1826del	h	h (46%)	h (46%)
	c.1953_1956delGAAA	h	h (35%)	h (34%)
	c.2077G>A	h	h (60%)	h (62%)
	c.2082C>T	H	**H** (100%)	**H** (100%)
	c.2269delG	h	h (66%)	h (66%)
	c.2612C>T	h	h (42%)	h (42%)
	c.3113A>G	h	h (51%)	h (52%)
	c.3548A>G	h	h (49%)	h (49%)
	3839-3843del5ins4	h	h (52%)	h (52%)
	c.4127del	h	h (56%)	h (44%)
	c.4214-4215delIns5	—	—	h (23%)†
	c.4221delins9	—	—	h (26%)†
	C.4227-4237delins16	—	—	h (24%)†
	c.4243-4244delGA	—	—	h (26%)†
	c.4281_4282ins39	h	h (44%)	—††
	c.4308T>C	h	h (55%)	h (49%)
	c.4575-4585del11	h	h (46%)	h (43%)
	c.4810C>T	h	h (58%)	h (53%)
	c.5266dupC	h	h (59%)	h (59%)
	c.5333-20_5333-19insT	—	—	h (25%)†
BRCA2	c.37_44del8	h	h (25%)**	h (26%)
	c.1114A>C	h	h (47%)	h (51%)
	c.1246A>G	h	h (47%)	h (49%)
	c.1553_1554insT	—	h (31%)**†	—**
	c.1748_1749insA	—	h (47%)**†	h (26%)†
	c.1759-1761delinsC	—	—	h (25%)†
	c.1774delT	—	—	h (33%)†
	c.1804-1806delins3	—	—	h (21%)†
	c.1803dupA	—	—	h (43%)†
	c.1815dupA	—	h (68%)†	h (31%)†
	c.1823dupA	—	—	h (33%)†
	c.1833dupA	—	—	h (21%)†
	c.2589T>A	—	—	h (34%)†
	c.2803G>A	h	h (39%)	h (40%)
	c.3479G>A	—	h (31%)**†	—
	c.3807T>C	h	h (44%)	h (42%)
	c.4332-4333delTA	—	—	h (66%)†
	c.4350dupT	—	—	h (44%)†
	c.4781delins3	—	—	h (22%)†
	c.5073dupA	h	h (42%)	h (41%)
	c.5385dupA	—	—	h (22%)†
	c.5459_5460insA	—	h (32%)**†	—
	c.7977-10dup	—	—	h (70%)†
	c.8125dupA	—	—	h (23%)†
	c.8147-8148insA	—	—	h (29%)†
	c.8574dup	—	h (38%)†	h (30%)†
	c.8797del	—	H†	—
	c.8800del	—	h (80%)**†	—
	c.8823dupA	—	—	h (28%)†
	c.8946dup	—	h (27%)**†	—

Table 1: Continued.

Gene	HGVS nomenclature	Sanger	AVA + AgsA	SeqNext
	c.10083del	—	—	h (21%)[†]
	c.10115dupC	—	—	h (50%)[†]
	c.10122delC	—	—	h (32%)[†]
False negative (out of 64)[††]		0 (reference)[††]	0[††]	1[†]
False positive (out of 64)[†]		0 (reference)[†]	10[†]	28[††]

% of reads carrying the variant is given in parentheses.
[**]Depth of the variant was <40 reads.
[†]Cells highlight false positives; [††]cells highlight false negatives.

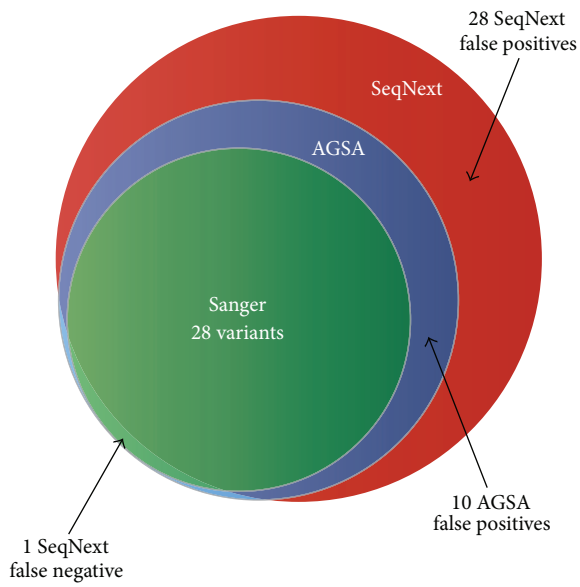


Figure 5: A composite sample including 28 variants validated in Sanger was analysed both with AGSA software and with SeqNext, using the same threshold of 20%. AGSA detected all 28 variants and 10 false positive variants whereas SeqNext missed 1 real variant and reported 28 false positives.

Sanger sequencing. Nevertheless, SeqNext gave about twice as many false positives as AGSA (177 for SeqNext versus 85 for AGSA).

AGSA software thus seems to be more efficient for Roche pyrosequencing analysis as it generates fewer false negative and false positive variants.

3.4. Protocol Optimization. To challenge the robustness of our method of *BRCA* pyrosequencing, we performed a larger series of 250 patients. To optimize the cost and duration of the analysis, we used microfluidic amplification on the Access Array Fluidigm system, generating 48 amplicons for 48 patients simultaneously. The analysis of *BRCA1* and *BRCA2* sequences of 250 Lebanese patients was performed using AGSA; variants were confirmed by dideoxy sequencing.

We first examined the coverage of each amplicon across the six runs performed. Around 10% of all amplicons had to be repeated by dideoxy sequencing. This included 7.5% that did not reach the 40-read threshold and 3.2% of homopolymers that were not eliminated by AGSA (Figures 6(a) and 6(b)). The percentage of homopolymers flagged for confirmation decreased from the first to last runs (3.2% versus 9% (22 + 5 for 299 variants), Supplementary Figure 3), suggesting that confidence in reading the histograms generated by AGSA increased with experience. With Sanger analysis, the first run of sequencing for each patient resulted in a low percentage of failed amplification or sequencing. In our laboratory this percentage was evaluated on our most recent 500 patients analysed using the Sanger technique, with a mean of 15% in *BRCA1* and 13% in *BRCA2*. Thus Sanger sequencing and Roche pyrosequencing generate similar rates of technical failure, with no correlation between the two techniques for samples either partially or fully analysed by both techniques.

The cost and duration of analyses were both improved. Reagent costs were reduced by 67% and technician time was reduced by 71% (Figure 6(c)). Finally, the analysis of 96 patients was estimated at about 34 working days for GS-FLX sequencing versus 60 for Sanger sequencing (Figure 6(d)). This time saving is essential since the delay for reporting results can be very long for *BRCA* sequencing. With the development of specific cancer therapies targeted to *BRCA* mutation carriers, there is a crucial interest in lowering these turnaround times.

4. Conclusions

Although NGS methods are starting to be used routinely in many molecular genetic laboratories [7, 11–13], Sanger dideoxy sequencing remains the gold standard technique. This study aims to validate our NGS method for constitutional genetics diagnosis. This method combines 454 sequencing and analysis with AGSA in-house developed software. Comparing NGS to Sanger sequencing for 72 samples, we showed that all variants found in standard Sanger method were also found by NGS when the conditions of analysis were set to a minimum of 40 reads and 20% of reads carrying a variant. This method is thus at least as sensitive as Sanger sequencing. Moreover we showed that this method allows automation of the sequence reading and thus decreased human potential error rate. This is particularly interesting in a goal of increasing the quality of routine diagnostic procedures.

Some false positive variants in homopolymers were found in 454 pyrosequencing compared to Sanger analyses. This is a recurrent problem for Ion Torrent (Life Technology) and 454 (Roche) sequencing. Several authors developed and tested different analysis workflows in order to correct false indel

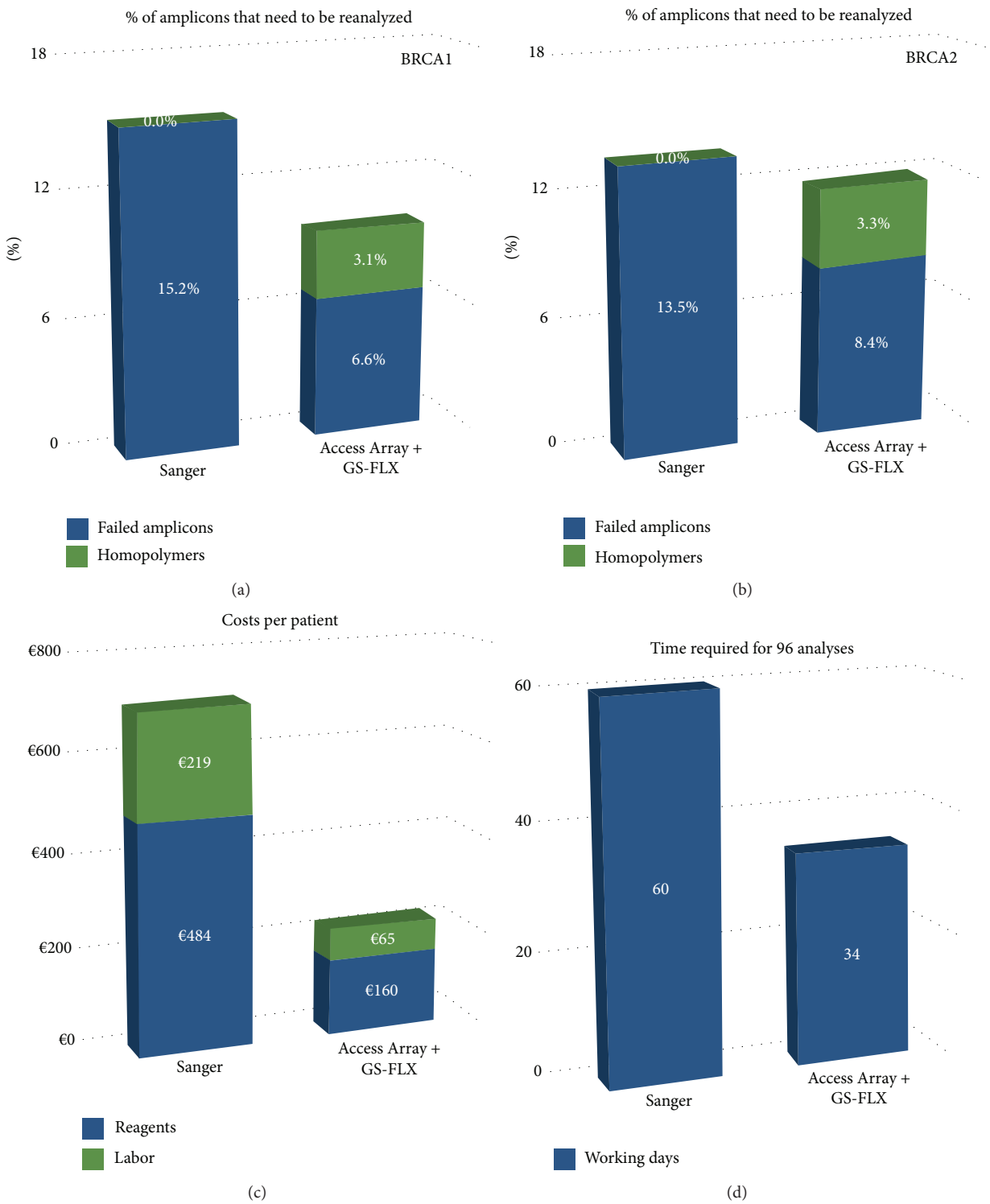


Figure 6: 250 patients were studied for *BRCA* mutations by Access Array Fluidigm combined with 454 pyrosequencing and AGSA analysis. This sequencing methodology was compared to Sanger analysis in terms of percentage of amplicons to be reanalysed for *BRCA1* (a) and *BRCA2* (b), cost per patient (c), and time required to analyse 96 patients (d).

detections [8, 14–17]. Coral is so far the state-of-the-art error corrector [17]. It is very efficient for substitutions, but it does not account for homopolymer context when interpreting indels, leading to calling errors in these sequences. HECTOR appears as a more optimized approach to deal with indels in homopolymers [16]. Although particularly powerful, this software is designed to correct genome-wide sequences, which implies improvements in runtimes but also a relative tolerance for false negatives, making it not adapted for gene-specific diagnosis purposes. In contrast, AmpliconNoise was designed to correct indel errors in PCR-based pyrosequences [18]. It is, however, computing-power demanding and it is hardly usable by nonbioinformaticians.

Our home-made software AGSA allows easily visualizing the data of each problematic homopolymer and quickly eliminating most of these false positive variants and generates easy-to-use results tables of validated amplicons and annotated variants.

Globally our proposed method of 454 sequencing is thus well adapted to constitutional sequencing diagnostics, since it is very sensitive, faster, and less expensive than Sanger sequencing.

Data Access

Because supporting data are genomic data of patients, they must remain confidential and they cannot be made available to the scientific community. Other data supporting the results of this paper are included within the paper.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Nancy Uhrhammer, Yves-Jean Bignon, and Yannick Bidet conceived the study. Both Axel Poulet and Flora Ponelle coded the software. Yves-Jean Bignon and Nagi S. El Saghir recruited the patients. Sandrine Viala, Stephanie Decousus, and Laurence Lafarge produced the data and the diagnostic analysis. Maud Privat and Yannick Bidet analysed the data and wrote the paper. Nancy Uhrhammer revised the paper. Axel Poulet and Maud Privat contributed equally to the paper.

Acknowledgments

This work was supported by funding from the Institut National du Cancer, as well as by the sponsors of the GINA (Génotypage INtensif en Auvergne) technical platform: the Université d'Auvergne, Clermont Communauté, the Conseil Régional d'Auvergne, the Ligue Contre le Cancer, the Groupement d'Entreprises Françaises Lutte contre le Cancer, Fonds Européens de Développement en Région, and Ethnic Research Initiative- (ERI-) GlaxoSmith Kline (GSK) for the American University of Beirut Medical Center of Lebanon.

References

[1] A. Antoniou, P. D. Pharoah, S. Narod et al., "Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies," *The American Journal of Human Genetics*, vol. 72, no. 5, pp. 1117–1130, 2003.

[2] K. Offit, "BRCA mutation frequency and penetrance: new data, old debate," *Journal of the National Cancer Institute*, vol. 98, no. 23, pp. 1675–1677, 2006.

[3] N. Loman, O. Johannsson, U. Kristoffersson, H. Olsson, and Å. Borg, "Family history of breast and ovarian cancers and BRCA1 and BRCA2 mutations in a population-based series of early-onset breast cancer," *Journal of the National Cancer Institute*, vol. 93, no. 16, pp. 1215–1223, 2001.

[4] J. N. Weitzel, V. I. Lagos, C. A. Cullinane et al., "Limited family structure and BRCA gene mutation status in single cases of breast cancer," *Journal of the American Medical Association*, vol. 297, no. 23, pp. 2587–2595, 2007.

[5] O. Harismendy, P. C. Ng, R. L. Strausberg et al., "Evaluation of next generation sequencing platforms for population targeted sequencing studies," *Genome Biology*, vol. 10, no. 3, article R32, 2009.

[6] T. Tengs, H. Zhang, A. Holst-Jensen et al., "Characterization of unknown genetic modifications using high throughput sequencing and computational subtraction," *BMC Biotechnology*, vol. 9, article 1472, 2009.

[7] G. Michils, S. Hollants, L. Dehaspe et al., "Molecular analysis of the breast cancer genes *BRCA1* and *BRCA2* using amplicon-based massive parallel pyrosequencing," *Journal of Molecular Diagnostics*, vol. 14, no. 6, pp. 623–630, 2012.

[8] X. Yang, S. P. Chockalingam, and S. Aluru, "A survey of error-correction methods for next-generation sequencing," *Briefings in Bioinformatics*, vol. 14, no. 1, Article ID bbs015, pp. 56–66, 2013.

[9] M. Margulies, M. Egholm, W. E. Altman et al., "Genome sequencing in microfabricated high-density picolitre reactors," *Nature*, vol. 437, pp. 376–380, 2005.

[10] G. Bentley, R. Higuchi, B. Hoglund et al., "High-resolution, high-throughput HLA genotyping by next-generation sequencing," *Tissue Antigens*, vol. 74, no. 5, pp. 393–403, 2009.

[11] K. De Leeneer, J. Hellemans, J. De Schrijver et al., "Massive parallel amplicon sequencing of the breast cancer genes BRCA1 and BRCA2: opportunities, challenges, and limitations," *Human Mutation*, vol. 32, no. 3, pp. 335–344, 2011.

[12] M. Chan, S. M. Ji, Z. X. Yeo et al., "Development of a next-generation sequencing method for BRCA mutation screening: a comparison between a high-throughput and a benchtop platform," *Journal of Molecular Diagnostics*, vol. 14, no. 6, pp. 602–612, 2012.

[13] L. Feliubadaló, A. Lopez-Doriga, E. Castellsagué et al., "Next-generation sequencing meets genetic diagnostics: development of a comprehensive workflow for the analysis of BRCA1 and BRCA2 genes," *European Journal of Human Genetics*, vol. 21, no. 8, pp. 864–870, 2013.

[14] J. L. Costa, S. Sousa, A. Justino et al., "Nonoptical massive parallel DNA sequencing of *BRCA1* and *BRCA2* genes in a diagnostic setting," *Human Mutation*, vol. 34, no. 4, pp. 629–635, 2013.

[15] Z. X. Yeo, J. C. L. Wong, S. G. Rozen, and A. S. G. Lee, "Evaluation and optimisation of indel detection workflows for

ion torrent sequencing of the BRCA1 and BRCA2 genes," *BMC Genomics*, vol. 15, no. 1, article 516, 2014.

[16] A. Wirawan, R. S. Harris, Y. Liu, B. Schmidt, and J. Schröder, "HECTOR: a parallel multistage homopolymer spectrum based error corrector for 454 sequencing data," *BMC Bioinformatics*, vol. 15, article 131, 2014.

[17] L. Salmela and J. Schröder, "Correcting errors in short reads by multiple alignments," *Bioinformatics*, vol. 27, no. 11, Article ID btr170, pp. 1455–1461, 2011.

[18] C. Quince, A. Lanzen, R. J. Davenport, and P. J. Turnbaugh, "Removing noise from pyrosequenced amplicons," *BMC Bioinformatics*, vol. 12, article 38, 2011.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 8759267, 5 pages
http://dx.doi.org/10.1155/2016/8759267

Research Article

Necessity of Microdissecting Different Tumor Components in Pulmonary Tumor Pyrosequencing

Dahui Qin, Zhong Zheng, Shanxiang Shen, Prudence Smith, and Farah K. Khalil

Department of Pathology, Moffitt Cancer Center, 12902 USF Magnolia Drive, Tampa, FL 33612, USA

Correspondence should be addressed to Dahui Qin; dahui.qin@moffitt.org

Received 15 February 2016; Revised 20 June 2016; Accepted 25 July 2016

Academic Editor: Urzsula Demkow

Microdissection is a useful method in tissue sampling prior to molecular testing. Tumor heterogeneity imposes new challenges for tissue sampling. Different microdissecting methods have been employed in face of such challenge. We improved our microdissection method by separately microdissecting the morphologically different tumor components. This improvement helped the pyrosequencing data analysis of two specimens. One specimen consisted of both adenocarcinoma and neuroendocrine components. When both tumor components were sequenced together for KRAS (Kirsten rat sarcoma viral oncogene homolog) gene mutations, the resulting pyrogram indicated that it was not a wild type, suggesting that it contained KRAS mutation. However, the pyrogram did not match any KRAS mutations and a conclusion could not be reached. After microdissecting and testing the adenocarcinoma and neuroendocrine components separately, it was found that the adenocarcinoma was positive for KRAS G12C mutation and the neuroendocrine component was positive for KRAS G12D mutation. The second specimen consisted of two morphologically different tumor nodules. When microdissected and sequenced separately, one nodule was positive for BRAF (v-raf murine sarcoma viral oncogene homolog B1) V600E and the other nodule was wild type at the BRAF codon 600. These examples demonstrate that it is necessary to microdissect morphologically different tumor components for pyrosequencing.

1. Introduction

Microdissection has been used to obtain tumor tissue for molecular testing with the primary goal of separating tumor and normal tissue to increase the amount of tumor DNA and, therefore, increasing test sensitivity. Tumor heterogeneity imposes new challenges for tissue sampling. We need to microdissect tumor not only from normal tissue but also from different intratumor elements. Different methods have been used for microdissection [1, 2].

Pyrosequencing is a sequencing technology that is more sensitive than Sanger sequencing [3, 4]. Pyrosequencing results are usually depicted as a series of peaks called a pyrogram, which reflects the DNA sequence. Pyrosequencing has been used in detecting different mutations, including KRAS (Kirsten rat sarcoma viral oncogene homolog) [5]. In our lab, it is used to screen for mutations in KRAS, EGFR (epidermal growth factor receptor), and BRAF (v-raf murine sarcoma viral oncogene homolog B1) since these gene mutations are associated with gene targeted therapies [6, 7]. For example, EGFR exon 19 deletions are associated with the tumor susceptibility to EGFR targeted therapy while KRAS mutations are associated with the tumor resistance to EGFR targeted therapy. BRAF V600E mutation is associated with the tumor susceptibility BRAF targeted therapy [6, 7]. In this study, we report unusual molecular test results from two different specimens (specimens 1 and 2), which demonstrate the necessity of microdissecting morphologically different components prior to pyrosequencing.

2. Material and Methods

Two separate specimens were examined for this study with distinctly different morphologies:

(1) Adenocarcinoma with neuroendocrine differentiation (specimen 1).

(2) Adenocarcinoma with two morphologically different nodules (specimen 2).

2.1. Microdissection of the Adenocarcinoma with Neuroendocrine Differentiation. Specimen 1 consists of a 3-cm pulmonary nodule, containing adenocarcinoma and neuroendocrine components. The formalin fixed paraffin embedded (FFPE) tumor tissue sections were microdissected manually using two methods: one combining both adenocarcinoma and neuroendocrine components and the second separating the adenocarcinoma and neuroendocrine components. The three microdissected samples from specimen 1 were sequenced separately.

2.2. Microdissection of the Adenocarcinoma with Morphologically Different Nodules. Specimen 2 consists of adenocarcinoma with two morphologically different pulmonary nodules. One tumor nodule is invasive adenocarcinoma, solid predominant type [8], and the other is minimally invasive adenocarcinoma, nonmucinous tumor [8]. Each nodule was microdissected separately and sequenced separately.

2.3. DNA Extraction. DNA was extracted from the microdissected tissue, using the QIAamp DNA FFPE Tissue Kit (Qiagen, Cat# 56404, Valencia, CA 91355, USA) and QiaCube instrument (Qiagen, Valencia, CA 91355, USA). The microdissected tissue was transferred into a 2 mL tube containing 180 μL of ATL buffer. Next, twenty microliters of proteinase K was then added to the tube and mixed by vortexing. The tissue mixture was incubated at 56°C for 2 hours followed by 90°C incubation for 1 hour. Following incubation the sample was loaded onto the QiaCube to complete the rest of DNA extraction process using the gDNA Extraction Program for FFPE tissue. Finally, the DNA extract was eluted into 100 μL ATE buffer to be used for pyrosequencing.

2.4. Pyrosequencing. Pyrosequencing for KRAS, EGFR, and BRAF mutation was performed according to the manufacturer's instructions [9–11] with some modifications. The targeted DNA sequences that included KRAS codons 12, 13, and 61, EGFR exons 18, 19, 20, and 21, and BRAF codon 600 were amplified using Qiagen KRAS, EGFR, and BRAF mutation test kits. The PCR products were used as templates for sequencing. Sequencing primers from the Qiagen KRAS, EGFR, and BRAF kits are designed to hybridize to the sequence near the targeted mutations, usually within a few nucleotides.

The targeted sequence for KRAS codons 12 and 13 is GGTGGCGTAGG and the dispensing order is ACTGTACGTGATCGTAGCA (Figure 1(a), x-axis). The targeted sequence for KRAS codon 61 is ACAGCAGGTCAAGAG. Since the sequencing primer for KRAS codon 61 is a reverse primer, the pyrosequencing reading is the reverse complement of the target sequence, CTCTTGACCTGCTGT. Therefore the dispensing order is GCTCGATACGACCT (data not shown). A typical pyrogram peak pattern for wild type KRAS codons 12 and 13 is shown in Figure 1(a).

The targeted sequence for BRAF codon 600 is CTAGCTACAGTG. The sequence primer for BRAF codon 600 is also a reverse primer; therefore the pyrosequencing reading for BRAF is the reverse complement of the targeted sequence, CACTGTAGCTAG. The dispensing order is TCGTATCTGTAG (Figure 1(e), x-axis). A typical pyrogram peak pattern for wild type BRAF codon 600 is shown in Figure 1(e).

EGFR exons 18, 19, 20, and 21 were also sequenced (data not shown). The targeted sequences were GGCTCCGGTGC (exon 18), TATCAAGGAATTAAGAGAAGCAACATCTCCGAAAG (exon 19), CAGCGTGG and ATCACGCAG (exon 20, codons 768 and 790), and CTGGCCAAACTGCTGGGT (exon 21) respectively. The dispensing orders for EGFR exons are CATGTCACTCGTG (for exon 18), CTATCACTGTCAGCTCGATCGTCATCGTCACGC (for exon 19), GCAGTACGTGTCGTGTACGTGACCACACTG and GATCATCTG (for exon 20, codons 768 and 790 resp.), and ACGTGTCACATGTC (for exon 21).

3. Result

3.1. Morphology and KRAS Mutations of the Adenocarcinoma with Neuroendocrine Differentiation (Specimen 1). The tumor morphology of specimen 1 is shown in Figures 2(a) and 2(b). The tumor consists of both glandular component (Figure 2(a)) and solid nested neuroendocrine component (Figure 2(b)). Immunohistochemical stains, performed with appropriate controls, revealed positivity for neuroendocrine marker, chromogranin, in the solid nests, and no reactivity for the marker in the glandular component (data not shown).

The pyrosequencing result of the entire tumor from specimen 1, including both adenocarcinoma and neuroendocrine components, is shown in Figure 1(b). The pyrogram of this result is different from that of wild type (shown in Figure 1(a)), suggesting that this pattern may reflect a mutated KRAS. The pattern, however, does not match the peak pattern of any KRAS mutations. Sequencing was then performed on each of the microdissected adenocarcinoma and neuroendocrine components. The pyrograms now show that the adenocarcinoma is positive for KRAS G12C mutation (Figure 1(c)) and the neuroendocrine component is positive for KRAS G12D mutation (Figure 1(d)). Both tumor components were negative for BRAF and EGFR mutations (data not shown).

3.2. Morphology and BRAF Mutation of the Adenocarcinoma with Morphologically Different Nodules. Specimen 2 consists of two tumor nodules, invasive adenocarcinoma, solid predominant type [8] (Figure 2(c)), and minimally invasive adenocarcinoma, nonmucinous tumor [8] (Figure 2(d)). Immunohistochemical stains performed with appropriate controls on both tumor nodules reveal immunoreactivity for TTF-1 (Thyroid Transcription Factor-1, data not shown), indicating pulmonary primary. Pyrosequencing result of the first nodule is positive for BRAF V600E mutation (Figure 1(f)), and the second nodule is negative for BRAF codon 600 mutation (Figure 1(e)). Both tumor nodules are negative for KRAS and EGFR mutations (data not shown).

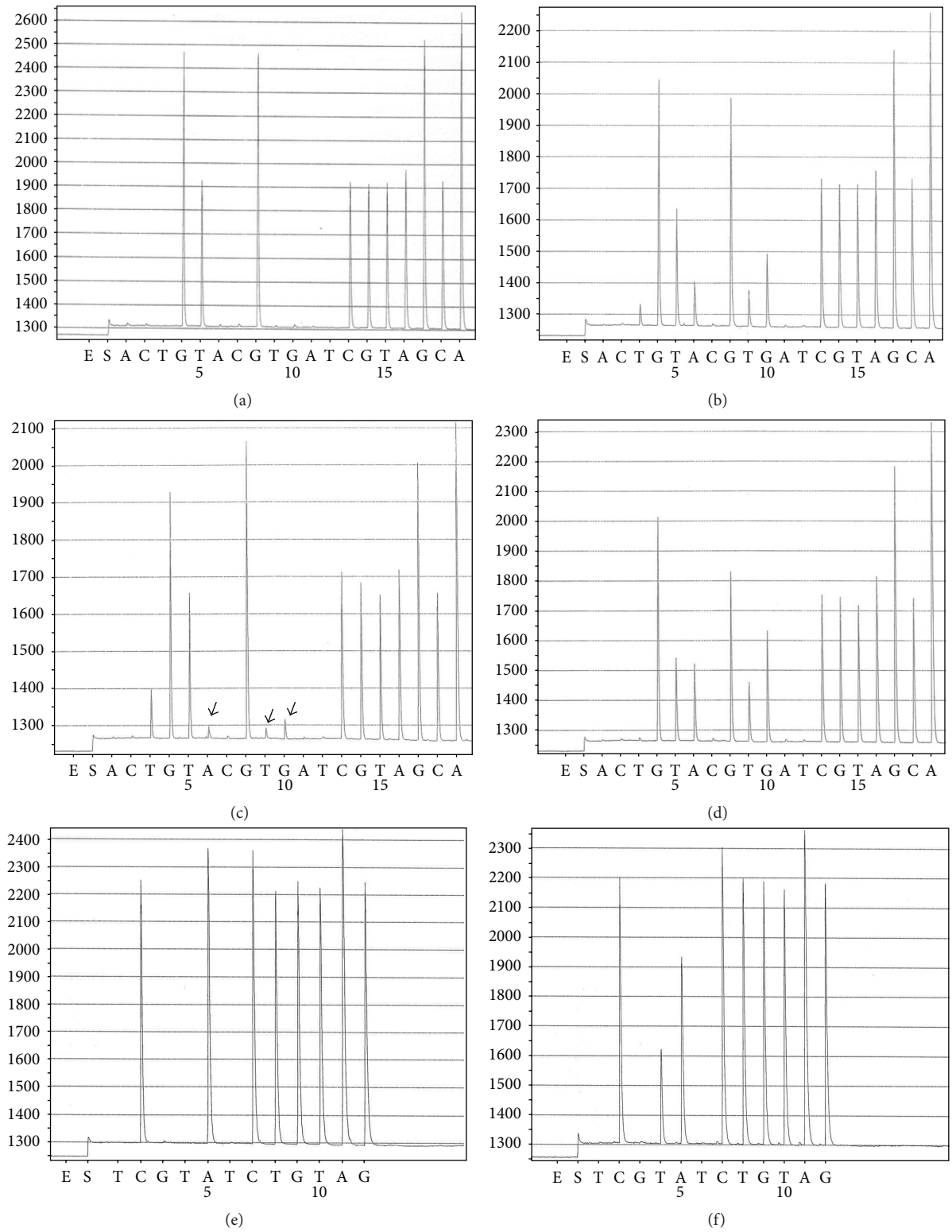


Figure 1: Six pyrograms, panels (a, b, c, d, e, and f), are present. The x-axis represents pyrosequencing dispensing order. The y-axis represents peak height. Panel (a) shows a pyrogram of wild type KRAS codons 12 and 13. Panel (b) shows a pyrogram from the whole tumor of an adenocarcinoma with neuroendocrine differentiation. The pyrogram peak pattern is different from a wild type pattern, but the pattern cannot match any KRAS mutation patterns. Panel (c) shows a pyrogram from the adenocarcinoma component that is shown in Figure 2(a), indicating KRAS G12C mutation. Three small peaks (indicated by black arrows) reflect the contamination of adenocarcinoma by neuroendocrine component due to imperfect microdissection. Panel (d) shows a pyrogram from the neuroendocrine component that is shown in Figure 2(b), indicating KRAS G12D mutation. Panel (e) shows a pyrogram from a tumor nodule with features of minimally invasive adenocarcinoma that is shown in Figure 2(d), indicating a wild type BRAF. Panel (f) shows a pyrogram from an invasive adenocarcinoma, solid predominant, poorly differentiated carcinoma nodule, which is shown in Figure 2(c), indicating BRAF V600E mutation.

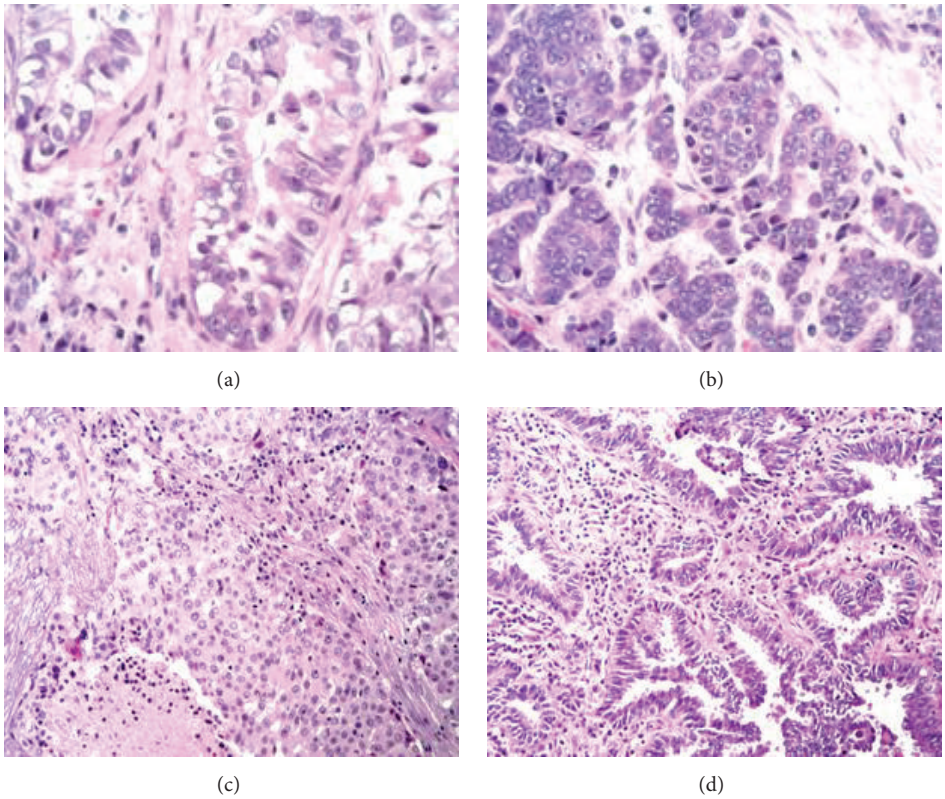


Figure 2: H&E stain of two specimens is shown. Panel (a) shows the glandular tumor component from specimen 1. Panel (b) shows the neuroendocrine tumor component from specimen 1. Panel (c) shows an invasive adenocarcinoma, solid predominant, poorly differentiated tumor, from specimen 2. Panel (d) shows the minimally invasive adenocarcinoma, nonmucinous type, from specimen 2.

4. Discussion

Tumor heterogeneity has been previously recognized and vigorously studied. The heterogeneity involves different levels of tumor clonal evolution, including cellular morphology, gene mutations, and biological responses to therapies [12, 13]. Tumor heterogeneity imposes a challenge to tissue sampling in molecular testing. In our practice, it has been noted that a tumor specimen may have morphologically different components, like specimen 1 in this report or morphologically different tumor nodules, as in specimen 2.

Specimen 1 consists of adenocarcinoma and neuroendocrine components. Morphologically, the two components are intermingled in some areas of the tumor but are distinct in other areas of the tumor (Figures 2(a) and 2(b)). When the entire tumor was microdissected altogether for mutation analysis, the data was difficult to interpret (Figure 1(b)). The pyrogram pattern was different from wild type pattern (Figure 1(a)), but it did not match any KRAS mutation pattern. In fact, the pyrogram pattern could not be interpreted. When the adenocarcinoma and neuroendocrine components were microdissected separately and tested separately for KRAS mutations, the adenocarcinoma component was found to harbor a KRAS G12C mutation (Figure 1(c)), and the neuroendocrine component was found to harbor a KRAS G12D mutation (Figure 1(d)). Retrospectively, the pyrogram from the whole tumor (Figure 1(b)) actually reflects the overlap of KRAS G12C and G12D pyrograms. This explains why the pyrogram from the whole tumor could not be interpreted as any specific KRAS mutation. Only when the two tumor components were tested separately did it become clear that each component harbored a different KRAS mutation.

In Figure 1(c), three small peaks are present (indicated by arrows). These small peaks are due to imperfect microdissection and reflect contamination of the adenocarcinoma by the neuroendocrine component. The results of these molecular tests indicate that it is necessary to microdissect morphologically different components separately prior to pyrosequencing to ensure accurate interpretation.

Specimen 2 consists of two morphologically different tumor nodules, an invasive adenocarcinoma, solid predominant type [8] (Figure 2(c)), which is positive for BRAF V600E mutation (Figure 1(f)) and a minimally invasive adenocarcinoma, nonmucinous tumor (Figure 2(d)), which is negative for BRAF codon 600 mutation (Figure 1(e)). If only the minimally invasive adenocarcinoma had been tested, we would have mistakenly assumed that the patient's tumor had no BRAF mutation. Likewise, if only the invasive adenocarcinoma, solid predominant nodule, had been tested, we would have believed that both tumor nodules contained the identified BRAF mutation. In either case, the molecular test report would not accurately reflect whole picture of the

BRAF gene mutation status, again emphasizing the necessity to microdissect each tumor nodule/component separately for mutation testing.

In current pathology practice, both small specimen size and tumor heterogeneity complicate the sampling process, indicating an increasing need for different microdissecting methods. A microdissection of different tumor components, as discussed in this report, addresses a portion of the difficulties in tissue sampling. This approach works when the heterogeneous components are morphologically different. Different methods have been and are still being developed to address different aspects of the tissue sampling issue. For example, microdissecting heterogeneous tumor components based on immunohistochemical phenotypes has been recently reported [2].

Different methods of microdissecting have improved the quality of clinical molecular testing and have a direct impact on patient care as indicated by the results from specimen 1. This specimen was an invasive adenocarcinoma with neuroendocrine differentiation. At the beginning, the entire tumor was microdissected altogether and the result was not interpretable. The puzzle was not resolved until the two tumor components were microdissected and tested separately. Then it became clear that two components bear different KRAS mutations.

Specimen 2 is an example that microdissecting different tumor components may have impact on patient management. This case presents two tumor nodules of different morphology. One was an invasive adenocarcinoma, solid predominant type, and the second nodule was a minimally invasive adenocarcinoma. The former was positive for BRAF V600E mutation and the latter was wild type for BRAF. Treatment for tumors with or without BRAF mutation could be different. Microdissecting and testing each tumor component separately can provide more precise mutation information for the patient's personalized management.

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors thank Mike Gruidl, Ph.D., for reviewing and editing the paper.

References

[1] S. R. Chowdhuri, L. Xi, T. H.-T. Pham et al., "EGFR and KRAS mutation analysis in cytologic samples of lung adenocarcinoma enabled by laser capture microdissection," *Modern Pathology*, vol. 25, no. 4, pp. 548–555, 2012.

[2] A. Z. Rosenberg, M. D. Armani, P. A. Fetsch et al., "High-throughput microdissection for next-generation sequencing," *PLoS ONE*, vol. 11, no. 3, Article ID e0151775, 2016.

[3] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlén, and P. Nyrén, "Real-time DNA sequencing using detection of pyrophosphate release," *Analytical Biochemistry*, vol. 242, no. 1, pp. 84–89, 1996.

[4] M. Ronaghi, M. Uhlén, and P. Nyrén, "A sequencing method based on real-time pyrophosphate," *Science*, vol. 281, no. 5375, pp. 363–365, 1998.

[5] S. Dufort, M.-J. Richard, and F. de Fraipont, "Pyrosequencing method to detect KRAS mutation in formalin-fixed and paraffin-embedded tumor tissues," *Analytical Biochemistry*, vol. 391, no. 2, pp. 166–168, 2009.

[6] R. A. Lilenbaum and L. A. Horn, "Management of EGFR mutation-positive non-small cell lung cancer," *Journal of the National Comprehensive Cancer Network*, vol. 14, no. 5, supplement, pp. 672–674, 2016.

[7] D. S. Ettinger, D. E. Wood, W. Akerley et al., "NCCN guidelines insights: non-small cell lung cancer, version 4. 2016," *Journal of the National Comprehensive Cancer Network*, vol. 14, no. 3, pp. 255–264, 2016.

[8] W. D. Travis, E. Brambilla, M. Noguchi et al., "International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma," *Journal of Thoracic Oncology*, vol. 6, no. 2, pp. 244–285, 2011.

[9] Qiagen, *EGFR Pyro Handbook*, Qiagen, 2010.

[10] Qiagen, PyroMark® KRAS v2.0 Handbook, March 2010.

[11] Qiagen, RAF Pyro® Handbook, September 2010.

[12] G. H. Heppner, "Tumor heterogeneity," *Cancer Research*, vol. 44, no. 6, pp. 2259–2265, 1984.

[13] L. I. Shlush and D. Hershkovitz, "Clonal evolution models of tumor heterogeneity," *American Society of Clinical Oncology Educational Book*, pp. e662–e665, 2015.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 6712529, 8 pages
http://dx.doi.org/10.1155/2016/6712529

Research Article

Genetic Analysis of the Atrial Natriuretic Peptide Gene Polymorphisms among Essential Hypertensive Patients in Malaysia

Nooshin Ghodsian, Patimah Ismail, Salma Ahmadloo, Narges Eskandarian, and Ali Etemad

Genetic Research Group, Department of Biomedical Science, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, Malaysia

Correspondence should be addressed to Patimah Ismail; patimahismail@gmail.com

Received 7 March 2016; Revised 26 April 2016; Accepted 4 May 2016

Academic Editor: Asude A. Durmaz

Background. Atrial natriuretic peptide (*ANP*) considerably influences blood pressure regulation through water and sodium homoeostasis. Several of the studies have utilized anonymous genetic polymorphic markers and made inconsequent claims about the *ANP* relevant disorders. Thus, we screened Insertion/Deletion (ID) and G191A polymorphisms of *ANP* to discover sequence variations with potential functional significance and to specify the linkage disequilibrium pattern between polymorphisms. The relationships of detected polymorphisms with EH with or without Type 2 Diabetes Mellitus (T2DM) status were tested subsequently. *Method*. *ANP* gene polymorphisms (*I/D* and *A191G*) were specified utilizing mutagenically separated Polymerase Chain Reaction (PCR) in 320 subjects including 163 EH case subjects and 157 controls. *Result*. This case-control study discovered a significant association between *I/D* polymorphisms of *ANP* gene in EH patient without T2DM. However, the study determined no association between G191A polymorphisms of *ANP* in EH with or without T2DM. In addition, sociodemographic factors in the case and healthy subjects exhibited strong differences ($P < 0.05$). *Conclusion*. As a risk factor, *ANP* gene polymorphisms may affect hypertension. Despite the small sample size in this study, it is the first research assessing the *ANP* gene polymorphisms in both EH and T2DM patients among Malaysian population.

1. Introduction

Hypertension, known as high blood pressure or a silent killer, is classified into 2 types: primary (essential) and secondary (pulmonary) form of hypertension. Essential hypertension (EHT) accounts for about 95% of people with hypertension and is used to describe hypertension when no specific cause is found according to the World Health Organization [1]. Hypertension is a significant public health issue for its damaging consequences globally [2]. According to World Health Organization, hypertension is widespread in plenty of developing economies. Globally, 1.56 billion people (29%) of adult population are predicted to suffer from hypertension by the year of 2025. Nearly seven million deaths annually might be affected by hypertension as stated by Singh et al. [3]. The heightened risk of stroke, kidney disease, heart attack, and heart failure will result as the blood pressure increases. In another study by Chobanian et al. [4], several risk factors (e.g., diabetes, high cholesterol levels) also raise the CVD risk from hypertension.

Previous prospective and case control studies have shown that hypertension progression is an independent predictor of type 2 diabetes [5]. Several possible factors are likely to be causes of the association between Type 2 Diabetes Mellitus (T2DM) and hypertension. The relationships between hypertension and diabetes were also obtained by other research groups [6]. Considering the fact that 88 individuals out of 163 hypertension cases suffered from diabetes, a strong relation was established between hypertension and diabetes in the study.

Genetics is claimed to contribute to hypertension. Genetic evidence influencing blood pressure comes from various sources. Lifton et al. [7] mention that high and low blood pressure, as one of rare Mendelian forms, are considerably

influenced by single genes. The genetic influence or heritability estimation on blood pressure (BP) variation displays remarkable range (30 to 50%) [8]. Genetic variations can considerably affect EHT genesis which significantly exhibits risk factor for progressive renal damage, stroke, ischemic heart disease, and peripheral vascular disease [9]. Most recent studies have conducted an investigation of genetic causes of essential hypertension associated with analysis of candidate genes. The investigated candidate genes include the following.

Atrial natriuretic peptide (*ANP*) is located on chromosome 1p36.2. Its main product (atrial natriuretic peptide protein) acts as cardiac hormone which is synthesized substantially within the heart and stored in the atrial myocyte as prohormones for rapid release in response to stimuli [10]. *ANP* is a 28-amino acid peptide with a 17-amino acid ring in the middle of the molecule. The ring is formed by a disulfide bond between two cysteine residues at positions 7 and 23 [11]. The heart is considerably affected by salt and water balance regulation as reported by Lee and Burnett [12]. Atrial natriuretic peptide (*ANP*) receptor, as a cardiac hormone, is involved in the physiological maintenance of blood volume and arterial blood pressure [13].

Insertion/deletion (*I*/*D*) polymorphism, which is 8-bp biallelic, is located in the second intron of the human atrial natriuretic peptide gene. The deletion variant might participate in the functional impairment of natriuretic peptide system defining an increased genetic susceptibility to hypertension [14].

The G191A polymorphism, which appears similarly in the previous studies as G664A but is used as G191A polymorphism and is mapped in a hydrophobic leader segment, is removed from the mature *ANP* (exon 1). In recent studies on G191A polymorphism, strong association with the cardiovascular disorders (including hypertension) was reported. The results with null findings were obtained for G664A including hypertension and other cardiovascular disorders [15]. However, the other studies indicated a significant association between G191A and hypertension in black Africans with the positive results for *Hpa* II RFLP [16]. Previous studies reported the association of *A191G* polymorphisms in European Americans and in black Africans, which was detected across the three ethnic groups but not for Japanese population. Given these findings, it will be of interest to investigate whether some of the ANP polymorphisms reported to date have arisen after human population differentiation [15].

Based on literature review, most of the studies were done on ANP gene polymorphisms associated with EHT (but their interactions within T2DM were not investigated). So this study was the first research conducted to determine genetic polymorphism of the I/D and G191A among EHT subjects with or without T2DM in Malaysian subjects.

2. Material and Method

2.1. Study Subject. In the current study, 163 Malaysian case subjects and 157 controls were analyzed. The controls were recruited based on the following criteria: (1) no history of EHT and Type 2 Diabetes Mellitus; (2) Systolic Blood Pressure (SBP) ≤ 140 mm Hg and DBP ≤ 90 mm Hg as measured with a digital sphygmomanometer; and (3) no recent symptoms of heart and renal disorders. The case subjects were recruited based on the following criteria: EHT history; SBP > 140 mm Hg and/or Diastolic Blood Pressure (DBP) > 90 mm Hg measured by a digital sphygmomanometer; and biological or clinical signs of pulmonary hypertension. Controls were selected from 168 consecutive volunteers without EHT histories. Among these, 11 subjects were eliminated for missing DNA extraction. The case subjects were selected in the Seremban Hospital. Between December 2011 and June 2012, we specified 168 patients of whom 5 subjects were excluded because they did not fit the blood pressure criterion. Ethical approval was acquired from Universiti Putra Malaysia and Seremban Hospital. All participants were asked to fill in informed consent questionnaires. The samples were used with reference number UPM.FPSK.PADS/T7-MJETIKAPer/F01-JSB-Mac.

2.2. Measurement. Body mass index (BMI) was calculated by measuring of case and control subjects' height and weight. Blood pressure was evaluated by measuring SBP and DBP with a safe, reproducible, accurate, and noninvasive method to screen Malaysian populations.

2.3. Biochemical Analysis. The mean of three consecutive measurements was computed. Plasma was extracted for determination of DNA extraction and standard biochemical measurements at the end of this procedure. Peripheral venous blood samples were collected after an overnight fasting in control subjects who were participating. In this section, serum electrolytes were utilized to examine the lipid profiles which included triglycerides (TG), total cholesterol (TCH), low-density lipoprotein (LDL), and high-density lipoprotein (HDL). Fasting Blood Sugar (FBS) is also measured with standard laboratory techniques. It was noticeable that we had referred to the hospital to assess biochemical information for cases' documents.

2.4. Genotype Investigations. In the study, the buccal and blood cells were collected from study group (hypertensive patients) and controls, respectively. The blood was kept in ethylenediaminetetraacetic acid (EDTA) tube and stored at 4°C for a maximum of three days before utilizing. The DNA was applied for amplification after extracting from buccal and blood cell samples using Qiagen Kit (Germany); then, it was stored at −20°C for later usage. DNA was qualified right after all primers were optimized by PCR method. By utilizing the nanodrop in two optical density (OD) wave lengths (260 nm and 280 nm), the extracted DNA concentration was examined.

Genomic DNA was amplified by multiplex-PCRs. Utilizing the mutagenically separated PCR technique, I/D (in intron 2) and the G191A polymorphisms (in exon 1) were genotyped. Each reaction was composed of 6x master mix (including DNA polymerase, $MgCl_2$, dNTPs, and reaction buffers), 0.6 μL relative primers (0.3 μL forward and 0.3 μL reverse), and 1 μL of genomic DNA and ultimately distilled water was added to a final volume of 25 μL. For each

Table 1: Oligonucleotides for amplification and screening of polymorphisms.

ANP gene polymorphism	Forward and reverse primer	PCR cycling conditions	PCR product (bp)	Reference
A191G	FP 5-AGAGGGGAACCAGAGAGGAACCAG-3 FP 5-CCATCAGGTCTGCGTTGGATAC-3	5 min 95°C 60 sec 55°C 90 sec 72°C 60 sec 94°C ×45 10 sec 72°C 4°C	189 and 197	Kato et al., [26]
I/D	FP 5-GCAGTCCAGCCTAGGTGATA-3 RP 5-TCCGGAGTAGCTAGGACTTACA-3	5 min 95°C 60 sec 55°C 90 sec 72°C 60 sec 94°C ×45 10 sec 72°C 4°C	230 and 238	Kato et al., [26]

PCR cycling conditions are represented as temperature and time of initial denaturation, denaturation, annealing, extension, and final extension × number of cycles.

polymorphism, several PCR and DNA amplifications were carried out with forward and reverse primers at the specific temperatures in several conditions (Table 1). Amplified PCR products were analyzed by gel electrophoresis methods with agarose.

2.5. Sequencing. In order to confirm the genotyping results, random samples were used and repeated with the same PCR conditions. To receive a final confirmation of the nucleotide sequences, purified PCR products were sent to Research Biolabs Malaysia. The sequencing results were aligned with the gene sequence from the NCBI-GenBank by the MEGA4 software.

2.6. Statistical Analysis. The statistical analysis in the present study was conducted by statistical package for the social science (SPSS version 22). Utilizing two-tailed Student's t-test and one-way ANOVA test, all variables among the groups and the group means were being contrasted ($P < 0.05$ was viewed to be significant statistically). The distribution of genotypes with Hardy-Weinberg expectations was calculated using a chi-squared test. Allelic frequencies were analyzed by gene-counting method. In order to detect the effects of high risk alleles, odds ratios (OR) with 95% confidence intervals (CI) were checked as well.

3. Result

3.1. Sociodemographic Factors. In this study, three different races' population were selected for our searching program including Malay, Chinese, and Indian subjects. It was also divided into three groups including EHT ($n = 75$), EHT + T2DM ($n = 88$), and control group ($n = 157$). The associations of clinical characteristics as major risk factors with ANP gene polymorphisms were investigated among EHT subjects with or without T2DM in Malaysia.

The mean and the standard deviation (SD) of the clinical characteristics were clearly indicated in Table 2. The mean age of EHT and EHT + T2DM subjects was nearly equal (59.45 ± 10.34, 59.05 ± 11.10) but higher compared to the controls (52.51±9.41). Considering age, Systolic Blood Pressure (SBP), Diastolic Blood Pressure (DBP), Body Mass Index (BMI), Fasting Blood Sugar (FBS), low-density lipoprotein (LDL), and triglyceride (TG), the subjects' clinical characteristics illustrated significant differences for mentioned parameters between hypertensive and normotensive subjects ($P < 0.05$). Noticeably, EHT subjects showed the major amount of SBP and DBP (152.01 ± 23.10, 94.12 ± 10.00). Besides, EHT + T2DM subjects showed the major amount in BMI and FBS which were 28.04 ± 4.25 and 7.89 ± 1.35, respectively. The amount of TCH and HDL did not show any significant differences between EHT and control subjects; however, EHT + T2DM subjects maintained minimum record of TCH (4.53 ± 1.04) in comparison with controls. Besides, LDL (3.21 ± 1.10) in controls' mean displayed the highest parameter in contrast to other groups. According to Table 2, the significant difference was obtained in the level of age (P value = 0.000), SBP (P value = 0.000), DBP (P value = 0.000), BMI (P value = 0.000), FBS (P value = 0.000), and TG (P value = 0.000) between EHT subjects with or without T2DM and control.

3.2. I/D. Analysis of *ANP*'s *I/D* mutation with 230 and 238 bp PCR product with mutation (I to D) was amplified by PCR. The *I/D* mutation of *ANP* gene was detected using 5% agarose gel electrophoresis. Figure 1 displays three different bands referring to three different genotypes of *I/D* polymorphism (II, ID, and DD). To analyze the *I/D* variation of *ANP* gene, we contrasted the EHT subjects with or without T2DM to the controls.

Table 3 shows the distribution of *I/D* polymorphic genotypes and allele frequencies of *ANP* gene between EHT, EHT with diabetes, and control subjects. There was a significant value of genotype ($P = 0.027$) and allele frequencies ($P = 0.015$) between EHT subjects and controls ($P < 0.05$). There was no significant difference between EHT + T2DM subjects

Table 2: Clinical characteristics of EHT, EHT + T2DM, and control subjects.

Parameter	EHT (N = 75)	EHT + T2DM (N = 88)	Control (N = 157)
AGE	59.45 ± 10.34	59.05 ± 11.10	52.51 ± 9.41
SBP	152.01 ± 23.10	150.51 ± 20.00	122.21 ± 11.24
DBP	94.12 ± 10.00	93.00 ± 9.34	76.20 ± 9.00
BMI	26.21 ± 6.02	28.04 ± 4.25	25.00 ± 4.00
FBS	5.46 ± 1.11	7.89 ± 1.35	5.04 ± 0.50
TCH	5.02 ± 1.10	4.53 ± 1.04*	5.10 ± 1.31
LDL	3.01 ± 1.10	3.00 ± 1.00	3.21 ± 1.10
HDL	1.34 ± 0.34*	1.21 ± 0.30*	1.30 ± 0.52
TG	1.50 ± 0.52	2.00 ± 0.53	1.23 ± 0.55

EHT refers to essential hypertensive patients, EHT + T2DM refers to essential hypertensive patients with Type 2 Diabetes Mellitus, and control refers to healthy subjects. P value is obtained by the comparisons of means between EHT and control subjects as well as between EHT + T2DM and control subjects, in the same row significant. Value mean ± standard deviation. *Nonsignificant $P > 0.05$.

Table 3: Genotypes and allele frequencies distribution of gene polymorphisms between two patient groups and control subjects.

Genotype and allele frequency	EHT n (%)	EHT+T2DM n (%)	Control n (%)
Insertion/deletion			
II	57 (76)	71 (80.7)	133 (84.7)
ID	12 (16)	14 (15.9)	21 (13.4)
DD	6 (8)	3 (3.4)	3 (1.9)
I	126 (84)	156 (88.6)	287 (91.4)
D	12 (16)	20 (11.4)	27 (8.6)
P value	0.027*/0.015*	0.219/0.200	—
Post hoc test			
II vs ID	0.749	—	—
II vs DD	0.056	—	—
ID vs DD	0.195	—	—
Odds ratio (95% CI)	0.494 (0.274–0.890)	0.734 (0.399–0.890)	—
A191G			
GG	54 (72)	69 (78.4)	124 (79)
GA	17 (22.7)	17 (19.3)	31 (19.7)
AA	4 (5.3)	2 (2.3)	2 (1.3)
G	125 (83.3)	155 (89.1)	279 (88.9)
A	25 (16.7)	19 (10.9)	35 (11.1)
P value	0.761/0.067	0.451/0.534	—
Odds ratio (95% CI)	0.627 (0.360–1.093)	1.023 (0.566–1.850)	—

EHT refers to essential hypertensive patients, EHT + T2DM refers to essential hypertensive patients with Type 2 Diabetes Mellitus, and control refers to healthy subjects. *P value < 0.05.

and controls (P value > 0.05). After post hoc test, neither genotype nor allele frequencies showed association between EHT subjects and controls. No significant value was achieved between EHT + T2DM subjects and controls.

3.3. A191G. Analysis of *ANP*'s *A191G* polymorphism with 189 and 197 bp PCR product with mutation (G to A) was amplified by PCR. In Figure 2, analysis of *A191G* polymorphism amplification product displayed three different variations (GG, GA, and AA) in *A191G* polymorphism.

The distribution of *A191G* polymorphic genotypes of *ANP* gene between the EHT, EHT + T2DM, and control subjects is indicated in Table 3. No significant difference was obtained in genotypes and allele frequencies between EHT patients and control group with P value of 0.761 and 0.067, respectively. Moreover, there were no significant differences of genotype and allele frequencies between EHT + T2DM subjects and control group which were 0.451 and 0.534, respectively.

3.4. Genotype Analysis Based on Race. As this study is a heterogeneous research, race was examined to specify for any significant role among genotypes. The genotypes of each groups based on race were analyzed in Table 4. Analyzing the ID genotypes base on race demonstrated significant

Table 4: Genotype analysis based on race.

Race	EHT $N = 75$			EHT + T2DM $N = 88$			Control $N = 157$		
	II	ID	DD	II	ID	DD	II	ID	DD
Malay	20 (69.0)	7 (24.1)	2 (6.9)	27 (79.4)	6 (17.6)	1 (2.9)	52 (78.8)	13 (19.7)	1 (1.5)
Chinese	27 (93.1)	0 (0.0)	2 (6.9)	20 (90.9)	1 (4.5)	1 (4.5)	54 (93.1)	3 (5.2)	1 (1.7)
Indian	10 (58.8)	5 (29.4)	2 (11.8)	24 (75.0)	7 (21.9)	1 (3.1)	27 (81.8)	5 (15.2)	1 (3.0)
P value		0.033			0.570			0.189	
	GG	GA	AA	GG	GA	AA	GG	GA	AA
Malay	19 (65.5)	9 (31.0)	1 (3.4)	25 (73.5)	8 (23.5)	1 (2.9)	46 (69.7)	20 (30.3)	0 (0.0)
Chinese	26 (89.7)	2 (6.9)	1 (3.4)	20 (90.9)	1 (4.5)	1 (4.5)	52 (89.7)	5 (8.6)	1 (1.7)
Indian	9 (52.9)	6 (35.3)	2 (11.8)	24 (75.0)	8 (25.0)	0 (0.0)	26 (78.8)	6 (18.2)	1 (3.0)
P value		0.046			0.264			0.021	

EHT refers to essential hypertensive patients, EHT + T2DM refers to essential hypertensive patients with Type 2 Diabetes Mellitus, and control refers to healthy subjects. P value < 0.05.

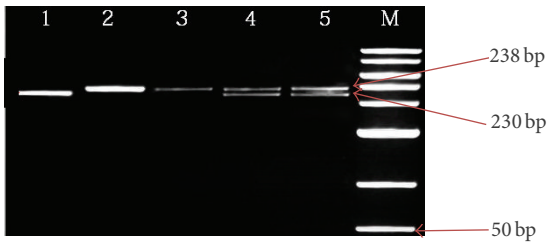


Figure 1: I/D polymorphism detection of ANP gene in 5% agarose gel electrophoresis. The picture represents homozygote (DD) in lane 1; lanes 2 and 3 show homozygote (II) and lane 4 represents heterozygote (I/D). Lane 5 genotype has been used as a control and to determine I/D. M represents a 50 bp DNA Ladder Plus (Bioline).

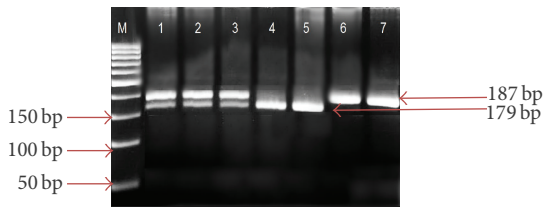


Figure 2: PCR product amplification of *ANP* gene in 5% agarose gel electrophoresis. The figure demonstrates the PCR product amplification of *A191G* polymorphism of *ANP* gene. Lanes 1, 2, and 3 show GA (189 and 197 bp); lanes 4 and 5 show GG; and lanes 6 and 7 show AA. M represents a 50 bp DNA Ladder Plus (Bioline).

difference between genotypes in EHT subjects ($P < 0.033$); besides, no significant differences were achieved among genotypes in EHT + T2DM subjects and/or in controls. Significant differences *were* observed between race and genotype of A191G in two groups which were EHT subjects (0.046) and control (0.021). However, no significant difference was observed between race and genotype in EHT + T2DM subjects in this table.

3.5. Allele Frequency Analysis Based on Gender. Regarding the importance of gender in the study, gender and allele frequency for *I*/*D* polymorphism were analyzed among EHT subjects with or without T2DM. Table 5 indicates the allele frequencies in each of the three groups. No significant differences were observed between the gender and each polymorphism.

4. Discussion

The aim of the present case-control study was to evaluate the effect of two polymorphism variants (*I*/*D* and *G191A*) on elements of the atrial natriuretic peptide and blood pressure. Blood pressure was evaluated by measuring SBP and DBP with a safe, reproducible, accurate, and noninvasive method to screen Malaysian populations. The study affirms that blood pressure and age are major determinants of hypertension. Negative association was achieved between EHT and HDL or TCH in both case and control subjects in the study; however, findings were also inconsistent in some other populations in the other studies. These discrepancies may be explained by several factors: the population under study may be affected by the sex and age distributions. The other possibility is that these inconsistencies might be the physiological mechanisms' results which act distinctly depending on sex and age.

High blood pressure is an important risk factor for CVD that is 26.4% around the world and it is increasing up to 29.2% in 2025. According to NHANES, the percentage is about 29.3% in United States (2003-2004). In Malaysia, the prevalence of EHT was about 32.2% in 2006 among the elderly. However, the prevalence of hypertension in the present study is 50.90%, which is higher in comparison to the study conducted by NHMS [17].

In many instances, genetic, environmental, and demographic complicated interactions lead to hypertension. Older age group is more likely to develop hypertension. Besides, it tends to increase rapidly with aging. Camm [18] also states that blood pressure increases with age, except where salt intake is low, physical activity is high, and obesity is not present. In this study, we have found a significant difference in hypertensive (where the hypertensive age was higher with $P < 0.05$) in comparison to normotensive subjects.

Table 5: Allele frequency analysis based on gender.

Alleles of gene polymorphisms		EHT $N = 75$		EHT + T2DM $N = 88$		Control $N = 157$	
		Male $N = 47$	Female $N = 28$	Male $N = 43$	Female $N = 45$	Male $N = 74$	Female $N = 83$
Insertion/deletion	I	77	49	76	80	132	155
	D	17	7	10	10	16	11
P value		0.456		0.551		0.228	
Odds ratio		0.647 (0.250–1.674)		0.950 (0.374–2.410)		0.585 (0.263–1.306)	
A191G	G	77	48	75	80	128	151
	A	17	8	11	10	20	15
P value		0.653		0.818		0.215	
Odds ratio		0.755 (0.303–1.884)		852 (0.342–2.123)		0.636 (0.313–1.293)	

EHT refers to essential hypertensive patients, EHT + T2DM refers to essential hypertensive patients with Type 2 Diabetes Mellitus, and control refers to healthy subjects. *P* value > 0.05.

Significant differences in age, SBP, DBP, BMI, FBS, LDL, and TG were found between hypertensive and normotensive subjects ($P < 0.05$), although HDL and TCH levels showed no significant differences in hypertensive and controls ($P > 0.05$). Significant differences of clinical characteristics (BMI, HDL, LDL, TG, TCH, SBP, and DBP) were identified between healthy individuals and EHT subjects as reported previously. Considerable variations were observed in triglycerides and total cholesterol between the groups similar to Asian Indian population study [19]. However, the risk factors (e.g., DBP, LDL, and HDL) have not shown any significant difference between case and control results in previous study of Asian Indian population. According to the finding, the increase of triglycerides and HDL reduction increase the chance of initial development and occurrence of the disease [20]. Hence, we were required identifying the etiological factors associated with hypertension disorders among the Malaysian population.

Gene searches, which affect the primary hypertension development in population, resulted from genetic analysis effective techniques (particularly genome-wide linkage analysis). Hsueh et al. [21] state that fundamental links of blood pressure to many chromosomal regions (as regions linked to familial combined hyperlipidemia) have been statistically identified by technique applications. Thus, there are several genetic loci that influence blood pressure in general population according to these findings. Nevertheless, Harrap et al. [22] studies indicate that recognizable single genes which affect hypertension are found to be unusual and match up with a multifactorial cause of hypertension. The results of Dominiczak et al. [23] studies showed relation between genes and blood pressure changes in 50% of cases.

The aim of the present case-control study was to evaluate the effect of two polymorphism variants (I/D and G191A) on elements of the atrial natriuretic peptide in hypertension to specify potential association affecting high blood pressure determination. The study affirms that blood pressure and age are major determinants of hypertension.

4.1. ANP I/D Polymorphism. The natriuretic peptides (specifically *ANP*) are increasingly recognized to play a fundamental role in BP regulation [24]. The role in BP regulation reflects the pluripotent cardiorenal actions of *ANP*, which include diuresis, enhancement of renal blood flow and glomerular filtration rate, systemic vasodilatation, suppression of aldosterone, and inhibition of the sympathetic nervous system. In addition to recent human studies, these actions of *ANP* demonstrate an association between higher plasma of *ANP* and a lower risk of EH [25].

Notably, strong association was observed between the *ANP I/D* polymorphism and hypertension in current study. The existence of the *ANP* D allele indicated relationship with increased SBP and age in both genders in the hypertensive subjects. The fact that the *ANP* polymorphism is related to SBP (merely in case) may be related to a hormone of atrial natriuretic peptide influences on the existence of hypertension.

In a previous study [26], the *ANP* system genetic variants are involved in the EHT etiology. *ANP*, which is in response to increased blood volume, acts to reduce water, sodium, and adipose loads on the circulatory system, thereby reducing blood pressure. It is mainly synthesized by atrial myocytes and released by relaxing blood vessels, where it acts by reducing peripheral resistance, increasing the glomerular filtration rate inhibiting renin release. It causes significant increases in urinary sodium excretion and urine output; this mechanism is involved in the regulation of BP [25].

4.2. ANP G191A Polymorphism. In humans, different single nucleotide polymorphisms (SNPs) have been identified in *ANP* gene. They seem to be in relation to high blood pressure and low plasma *ANP* levels and leave ventricular hypertrophy in subjects. Some other SNPs, characterized for EHT by the deletion variant, might participate in the functional impairment of natriuretic peptide system defining an increased genetic susceptibility to EHT [14, 27].

Our findings are in agreement with those of earlier studies demonstrating distribution of the *ANP G191A* genotypes

in case and control subjects. Homozygote (GG genotype) displayed higher frequency in case and control subjects of both sexes in comparison to other genotypes. Providing no justification with this astonishing observation, it may be artifactual to some degree because patients were chosen based on blood pressure levels.

5. Conclusion

Taken together, earlier single studies were expanded on EHT by proposing that *ANP* gene *I/D* and *G191A* polymorphisms may influence the occurrence of hypertension, particularly in population-based studies. Our observations also keep the inquiry open considering the heterogeneous influence of *ANP* gene *I/D* and *G191A* in variant ethnic populations. More functional and genetic studies are guaranteed to clarify the relation between *ANP* gene *I/D* and *G191A* polymorphisms and EHT as well as between the *ANP* gene mechanisms and hypertension. The pathophysiological relevance of *ANP* in relation to hypertension-associated phenotypes requires additional studies in various ethnic groups.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Nooshin Ghodsian conceived the study and participated in the experimental design, data acquisition and analysis, and interpretation of results and drafted the paper. Nooshin Ghodsian, Patimah Ismail, Salma Ahmadloo, Narges Eskandarian, and Ali Etemad interpreted the results and critically reviewed the study for important intellectual content. All authors approved the final version of the paper.

Acknowledgments

The authors fully acknowledge the cooperation of the volunteers and all participants who generate their generous support. They are also grateful to Mr. Mohammad Akhlaghi who edited this paper grammatically and in that regard improved the paper significantly.

References

[1] World Health Organization, *Research for Universal Health Coverage: World Health Report*, World Health Organization, Geneva, Switzerland, 2013.

[2] M. Ezzati, S. Oza, G. Danaei, and C. J. L. Murray, "Trends and cardiovascular mortality effects of state-level blood pressure and uncontrolled hypertension in the United States," *Circulation*, vol. 117, no. 7, pp. 905–914, 2008.

[3] T. P. Singh, M. Rohit, A. Grover, S. Malhotra, and R. Vijayvergiya, "A randomized, placebo-controlled, double-blind, crossover study to evaluate the efficacy of oral sildenafil therapy in severe pulmonary artery hypertension," *American Heart Journal*, vol. 151, no. 4, pp. 851.e1–851.e5, 2006.

[4] A. V. Chobanian, G. L. Bakris, H. R. Black et al., "Seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure," *Hypertension*, vol. 42, no. 6, pp. 1206–1252, 2003.

[5] D. Conen, P. M. Ridker, S. Mora, J. E. Buring, and R. J. Glynn, "Blood pressure and risk of developing type 2 diabetes mellitus: the Women's Health Study," *European Heart Journal*, vol. 28, no. 23, pp. 2937–2943, 2007.

[6] J. R. Banegas, E. López-García, A. Graciani et al., "Relationship between obesity, hypertension and diabetes, and health-related quality of life among the elderly," *European Journal of Cardiovascular Prevention & Rehabilitation*, vol. 14, no. 3, pp. 456–462, 2007.

[7] R. P. Lifton, A. G. Gharavi, and D. S. Geller, "Molecular mechanisms of human hypertension," *Cell*, vol. 104, no. 4, pp. 545–556, 2001.

[8] D. S. Timberlake, D. T. O'Connor, and R. J. Parmer, "Molecular genetics of essential hypertension: recent results and emerging strategies," *Current Opinion in Nephrology and Hypertension*, vol. 10, no. 1, pp. 71–79, 2001.

[9] F. Cambien and L. Tiret, "Genetics of cardiovascular diseases: from single mutations to the whole genome," *Circulation*, vol. 116, no. 15, pp. 1714–1724, 2007.

[10] D. L. Vesely, "Natriuretic peptides and acute renal failure," *American Journal of Physiology—Renal Physiology*, vol. 285, no. 2, pp. F167–F177, 2003.

[11] W. Yan, F. Wu, J. Morser, and Q. Wu, "Corin, a transmembrane cardiac serine protease, acts as a pro-atrial natriuretic peptide-converting enzyme," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 15, pp. 8525–8529, 2000.

[12] C. Y. W. Lee and J. C. Burnett Jr., "Natriuretic peptides and therapeutic applications," *Heart Failure Reviews* , vol. 12, no. 2, pp. 131–142, 2007.

[13] J. R. Dietz, "Mechanisms of atrial natriuretic peptide secretion from the atrium," *Cardiovascular Research*, vol. 68, no. 1, pp. 8–17, 2005.

[14] K. Lucarelli, M. Iacoviello, P. Dessì-Fulgheri et al., "Peptidi natriuretici e ipertensione arteriosa essenziale," *Italian Heart Journal, Supplement*, vol. 3, no. 11, pp. 1085–1091, 2002.

[15] A. I. Lynch, S. A. Claas, and D. K. Arnett, "A review of the role of atrial natriuretic peptide gene polymorphisms in hypertension and its sequelae," *Current Hypertension Reports*, vol. 11, no. 1, pp. 35–42, 2009.

[16] D. R. Rutledge, Y. Sun, and E. A. Ross, "Polymorphisms within the atrial natriuretic peptide gene in essential hypertension," *Journal of Hypertension*, vol. 13, no. 9, pp. 953–955, 1995.

[17] L. Rampal, S. Rampal, M. Z. Azhar, and A. R. Rahman, "Prevalence, awareness, treatment and control of hypertension in Malaysia: a national study of 16,440 subjects," *Public Health*, vol. 122, no. 1, pp. 11–18, 2008.

[18] J. Camm, "Medical management of atrial fibrillation: state of the art," *Journal of Cardiovascular Electrophysiology*, vol. 17, supplement 2, pp. S2–S6, 2006.

[19] G. Bansal, To study the prevalence of asymptomatic coronary artery disease in type 2 diabetes mellitus in KLES hospital and MRC, Belgaum-one year cross sectional study, 2008.

[20] W. C. S. Swee, T. G. Poh, T. S. Yean et al., "Relationship between body composition and bone mineral density in healthy postmenopausal Chinese women in Malaysia," *Jurnal Sains Kesihatan Malaysia*, vol. 5, no. 2, pp. 29–38, 2007.

[21] W.-C. Hsueh, B. D. Mitchell, J. L. Schneider et al., "QTL influencing blood pressure maps to the region of PPH1 on chromosome 2q31-34 in Old Order Amish," *Circulation*, vol. 101, no. 24, pp. 2810–2816, 2000.

[22] S. B. Harrap, A. D. Cumming, D. L. Davies et al., "Glomerular hyperfiltration, high renin, and low-extracellular volume in high blood pressure," *Hypertension*, vol. 35, no. 4, pp. 952–957, 2000.

[23] A. F. Dominiczak, D. C. Negrin, J. S. Clark, M. J. Brosnan, M. W. McBride, and M. Y. Alexander, "Genes and hypertension: from gene mapping in experimental models to vascular gene transfer strategies," *Hypertension*, vol. 35, no. 1, pp. 164–172, 2000.

[24] P. M. McKie, T. Ichiki, and J. C. Burnett Jr., "M-atrial natriuretic peptide: a novel antihypertensive protein therapy," *Current Hypertension Reports*, vol. 14, no. 1, pp. 62–69, 2012.

[25] G. Hu, X. Xu, X. Liang et al., "Associations of plasma atrial natriuretic peptide and electrolyte levels with essential hypertension," *Experimental and Therapeutic Medicine*, vol. 5, no. 5, pp. 1439–1443, 2013.

[26] N. Kato, T. Sugiyama, H. Morita et al., "Genetic analysis of the atrial natriuretic peptide gene in essential hypertension," *Clinical Science*, vol. 98, no. 3, pp. 251–258, 2000.

[27] S. Rubattu, G. Bigatti, A. Evangelista et al., "Association of atrial natriuretic peptide and type a natriuretic peptide receptor gene polymorphisms with left ventricular mass in human essential hypertension," *Journal of the American College of Cardiology*, vol. 48, no. 3, pp. 499–505, 2006.

Hindawi Publishing Corporation
BioMed Research International
Volume 2016, Article ID 6594039, 7 pages
http://dx.doi.org/10.1155/2016/6594039

Research Article

WWOX CNV-67048 Functions as a Risk Factor for Epithelial Ovarian Cancer in Chinese Women by Negatively Interacting with Oral Contraceptive Use

Yongxiu Chen, Xiaochang Tan, Yongli Ding, Bi Mai, Xiaowen Huang, Guiying Hu, and Xiping Luo

Gynecology Department, Guangdong Women and Children Hospital, No. 521, Xingnan Road, Panyu District, Guangzhou 511442, China

Correspondence should be addressed to Xiping Luo; luoxiping07@aliyun.com

Received 29 January 2016; Accepted 22 March 2016

Academic Editor: Ozgur Cogulu

Copy number variations (CNVs) have attracted increasing evidences to represent their roles as cancer susceptibility regulators. However, little is known about the role of CNV in epithelia ovarian cancer (EOC). Recently, the CNV-67048 of WW domain-containing oxidoreductase (*WWOX*) was reported to alter cancer risks. Considering that *WWOX* also plays a role in EOC, we hypothesized that the CNV-67048 was associated with EOC risk. In a case-control study of 549 EOC patients and 571 age (±5 years) matched cancer-free controls, we found that the low copy number of CNV-67048 (1-copy and 0-copy) conferred a significantly increased risk of EOC (OR = 1.346, 95% CI = 1.037–1.747) and it determined the risk by means of copy number-dependent dosage effect ($P = 0.009$). Data from TCGA also confirmed the abovementioned association as the frequency of low copies in EOC group was 3.68 times more than that in healthy group ($P = 0.023$). The CNV also negatively interacted with oral contraceptive use on EOC risk ($P = 0.042$). Functional analyses further showed a lower mRNA level of WWOX in tissues with the 0-copy or 1-copy than that in those with the 2-copy ($P = 0.045$). Our data suggested the CNV-67048 to be a risk factor of EOC in Chinese women.

1. Introduction

Ovarian cancer (OC) is one of the most common tumors of genital system and causes of cancer-related death among females. In America, it was estimated that there were 21,290 new OC patients and 14,180 deaths caused by it in 2015 [1]. In China, OC ranked the tenth cancer incidence with a rate of $6.89/10^5$ in 2011 [2]. Despite the fact that treatments for OC have greatly improved these years, high mortality caused by it was still observed as above data presented. The poor prognosis depends on a late diagnosis as more than 50% of patients were diagnosed with advanced OC. Prevention is the first-rank policy to decrease OC detriment such as etiological intervention, early detection, and early diagnosis, of which the foremost step is to discriminate high risk group of OC with respect to these risk factors of OC that included both environmental and genetic factors.

To date, multiple association studies have established several genetic factors, most of which are single nucleotide polymorphisms (SNPs), to be susceptible loci for OC [3–7]. For example, three genome-wide association studies (GWASs) had revealed seven SNPs at loci 9p22, 8q24, 2q31, 19p13, 3q25, and 17q21 to be risk alleles of OC. However, as the relatively small increments in risk are exerted, these SNPs can only represent a small proportion of OC heritability, which would cause low accuracy on predicting individuals' onset risk of OC in a population [8]. Thus, to reveal missing heritability of OC is impending. Missing heritability can be explained by other variants in genomics of human, such as copy number variants (CNVs), del/ins, and DNA transversion. Among them, CNV is the second major category of genetic variants after SNP, which makes up 10%~13% of human variants [9]. More and more evidences had indicated that CNVs are associated with various cancer

risks [10–14]. However, little is known about CNV on OC risk.

Recently, two studies had reported that one germline CNV-67048 that is located in a tumor suppressor gene *WW domain-containing oxidoreductase* (*WWOX*) contributed increased risks of lung cancer and gliomas in the Chinese [15, 16]. Also, CNVs across *WWOX* were reported to have potential contributions to breast cancer initiation and progression [17]. Functioning as a proapoptotic molecular and RNA-binding protein, WWOX protein plays a suppressor role in OC development [18–20]. As reported, WWOX can inhibit OC stem cells proliferation by downregulating expression of cell cycle proteins cyclin E-CDK2 and cyclin D1-CDK4 and promote cell apoptosis by upregulating expression of Wnt-5α, JNK, and caspase-3 [20]. WWOX also involves epithelial-mesenchymal transition of human OC stem cells [18]. Moreover, loss of WWOX expression was observed in OC, especially epithelial ovarian cancer (EOC) [21, 22]. Because previous studies showed that CNV-67048 influences WWOX expression in tumor tissues, considering the function and abnormal expression of WWOX in EOC, we hypothesized that the CNV-67048 contributes to EOC development. To test this hypothesis, we performed a case-control study including 549 EOC patients and 571 age (±5 years) matched cancer-free controls among Chinese women. We also performed functional assays to reveal the function of the CNV in EOC.

2. Materials and Methods

2.1. Study Subject. We conducted a case-control study in Guangzhou city in China. During September 2011 and July 2015, 549 EOC patients and 571 age (±5 years) matched cancer-free controls were recruited from the Guangdong Provincial Maternity and Child Care Center. EOC diagnosis is performed according to the International Federation of Gynecology and Obstetrics (FIGO). Individuals with tumor history were excluded. After a written informed consent was obtained, each participant was asked to denote 3 mL peripheral blood sample and complete a questionnaire to collect their data on sociodemographic, smoking status, alcohol consumption, menstrual and reproductive histories, and contraceptive use. This study was approved by the institutional review boards of Guangdong Provincial Maternity and Child Care Center.

2.2. Genotyping. Genomic DNA was extracted from 3 mL peripheral blood sample. Each DNA was then diluted into a concentration of 50 ng/μL. The copy number of CNV-67048 was tested with the TaqMan® Copy Number Assay according to the protocol of Applied Biosystems by Life Technologies. Briefly, a 10 μL reaction system including 1 μL DNA, 5 μL TaqMan Master Mix, 1 μL special probes and primers for CNV-67048 (cat# Hs03922779, Applied Biosystems), 0.5 μL control RNase P probe (Applied Biosystems), and 2.5 μL deionized water for each subject was prepared and run on the ABI 7900 system. Then the copy number was directly calculated by the CopyCaller® Software 2.1. 5% of the samples were randomly selected to repeat genotyping and the results were 98% in agreement.

2.3. Bioinformatics Analysis. To validate the association between the CNV and EOC risk as well as the possible effect of the CNV on WWOX expression, we downloaded available EOC germline CNVs data and WWOX expression value of 20 Asian EOC individuals from TCGA (https://tcga-data.nci.nih.gov/tcga/). The data of Asian common CNVs in healthy controls were also downloaded from one previously published study [23].

2.4. WWOX mRNA Level Estimation. A total of 31 EOC and 22 normal ovarian tissues were collected from the Guangdong Provincial Maternity and Child Care Center. Total RNA was extracted using the RNAiso reagent (Takara, Japan) and then reverse-transcribed into cDNA with the PrimeScript RT Master Mix (Takara, Japan). The SYBR-Green real-time PCR was used to assess the mRNA level of WWOX in the abovementioned tissues with the primers as suggested by previously published study, 5′-TGG GTT TAC TAC GCC AAT C-3′ (forward) and 5′-GTC CGT TCT CAT CAG TTT CT-3′ (reverse) [16]. The β-actin was used as an internal reference. Method of $2^{\Delta Ct}$ was used to demonstrate the mRNA level of WWOX. All analyses were performed in a blinded fashion with the laboratory persons unaware of genotyping data and each assay was done in triplicate.

2.5. Statistical Analysis. The chi-square test was used to assess differences in the distributions of CNV-67048 copy number between EOC cases and controls. The unconditional logistic regression model with or without adjustment for surrounding factors including age, age at menarche, number of births, menstrual history, oral contraceptive use, family history of cancer, smoking status, and alcohol intake was used to infer odds ratio (OR) and 95% confidence interval (95% CI) for each association between the CNV-67048 and EOC risk. The multiplicative interaction model was used to assess the possible interaction between the CNV-67048 and selected variables on cancer risk. The Kruskal-Wallis test was used to evaluate the effect of CNV-67048 on WWOX expression in tissues. All tests were two-sided by using the IBM SPSS software (version 22.0). $P < 0.05$ was considered to be statistically significant.

3. Results

3.1. Characteristics of the Study Subjects. As shown in Table 1, age matched well between EOC cases and controls with no significant difference ($P = 0.382$). Also, there was no significant difference in frequency distribution of menstrual history, family history of cancer, smoking status, and alcohol intake between the two groups ($P > 0.05$ for all). However, significantly higher frequency of menarche age less than 15 years, births number no less than 4, and reported null oral contraceptive use were observed in EOC cases than in controls (P values are 0.039, <0.001, and 0.039 in turn). Moreover, EOC cases were more likely to be heavy smokers with no less than 20 pack-years smoked than controls ($P = 0.029$).

Table 1: Frequency distributions of selected variables among EOC cases and controls.

Variables	Case (n = 549) n (%)	Control (n = 571) n (%)	P^a
Age (years)			
≤55	277 (50.46)	303 (53.06)	0.382
>55	272 (49.54)	268 (46.94)	
Age at menarche (years)			
<15	211 (53.55)	257 (46.73)	
≥15	183 (46.45)	293 (53.27)	0.039*
Unclear	155 (—)	21 (—)	
Number of births			
0	16 (2.91)	11 (1.93)	
1–3	409 (74.50)	490 (85.81)	<0.001
≥4	124 (22.59)	70 (12.26)	
Menstrual history			
Premenopause	152 (27.69)	131 (22.94)	0.068
Menopause	397 (72.31)	440 (77.06)	
Oral contraceptive use			
Never	414 (82.31)	399 (75.86)	
Seldom	56 (11.13)	78 (14.83)	0.039*
Often	33 (6.56)	49 (9.32)	
Unclear	46 (—)	45 (—)	
Family history of cancer			
Yes	40 (7.29)	38 (6.65)	0.678
No	509 (92.71)	533 (93.35)	
Smoking status			
Ever	68 (12.39)	54 (9.46)	0.116
Never	481 (87.61)	517 (90.54)	
Pack-years smoked			
≥20	36 (6.56)	18 (3.15)	
<20	32 (5.83)	36 (6.30)	0.029
0	481 (87.61)	517 (90.54)	
Alcohol intake			
Ever	34 (6.19)	49 (8.58)	0.127
Never	515 (93.81)	522 (91.42)	
Staging			
I + II	43 (7.84)		
III	420 (76.50)		
IV	86 (15.66)		

[a] P values for a χ^2 test.
* Statistical analysis excluded subjects with unclear or unknown data.

3.2. Contribution of WWOX CNV-67048 to EOC Development. As shown in Table 2, three types of copy number of CNV-67048, which are 2-copy, 1-copy, and 0-copy, were detected. The alter frequency of loss allele in the current study (10.3%) was equivalent to that in Asian individuals as reported (10.0%) [23]. The CNV was related to EOC susceptibility as its frequency distributions of copy number were significantly different between EOC cases and controls (P = 0.005). Results from the unconditional logistic regression model without adjustment for surrounding factors presented significant increases in EOC risk in both 1-copy carriers (OR = 1.325, 95% CI = 1.024–1.714) and 0-copy carriers (OR = 2.425, 95% CI = 1.261–4.665) compared to 2-copy carriers. A tendency for an increased EOC risk was further observed accompanied by decreased copy number (P = 0.002). Moreover, after adjustment for surrounding factors, the 0-copy (OR = 2.198, 95% CI = 1.111–4.348) and a combination of 1-copy and 0-copy (OR = 1.346, 95% CI = 1.037–1.747) still conferred significantly increased risk of EOC. The tendency is also significant (P = 0.009). However, we did not find any statistically significant associations between CNV-67048 and EOC stages in case only study (Table 2). In addition, in order to validate our result, we compared the Asian germline CNVs data from TCGA database of EOC and Asian common CNVs data of health population as published [23]; the frequency of low copy number in EOC group was 3.68 times more than that in healthy group with statistical significance (P = 0.023).

3.3. Associations between CNV-67048 and EOC Risk Stratified by Selected Variables. As shown in Table 3, the contributions of CNV-67048 on EOC risk were only significant in subgroups of 1 to 3 births' number, null oral contraceptive use, no family history of cancer, never smokers, heavy smokers, and never drinkers. However, the nonsignificant effect in the corresponding subgroups may be due to the limited sample size. Interestingly, the interaction analysis showed that the CNV-67048 significantly interacted with oral contraceptive use on EOC risk. As the CNV exerted a risk effect and oral contraceptive use exhibited a protective effect on EOC, their interactions are negative (P = 0.042). We further used the additive interaction model to show detailed interaction effect between them in Figure 1. As shown, compared to subjects who carried 2-copy CNV-67048 and never used oral contraceptive, those who carried 1-copy or 0-copy and never used oral contraceptive harbored the highest EOC risk (OR = 1.786, 95% CI = 1.317–2.423).

3.4. Effect of the CNV-67048 on WWOX Expression. As shown in Figure 2, in the total 53 cases of ovarian tissues, the mRNA levels of WWOX were significantly lower in tissues with the 0-copy (median: 0.0342) or 1-copy (median: 0.0347) of CNV-67048 than in those with the 2-copy (median: 0.0617; P = 0.045). We also queried the TCGA database and downloaded the 20 cases of Asians gene expression data; the gene expression of *WWOX* was also higher in tissues with 2-copy than in those with 0-copy or 1-copy. The difference between them is 0.082 (2-copy, 0.725 ± 0.100, versus 0-copy and 1-copy, 0.643 ± 0.053).

4. Discussion

In the current study, we found that the CNV-67048 of *WWOX* was significantly related to the risk of EOC in Chinese women. The CNV also negatively interacted with oral contraceptive use, because it could significantly tarnish the protective role of oral contraceptive use on EOC

Table 2: Associations between *WWOX* CNV-67048 copy numbers and EOC risk and stages.

CNV-67048 genotypes	*n* (%)	*n* (%)	*P*	Crude OR (95% CI)	Adjusted OR (95% CI)[a]
Case-control study	*Patients*	*Controls*			
Total number	549	571			
2-copy	339 (61.75)	397 (69.53)		1.000 (ref.)	1.000 (ref.)
1-copy	181 (32.97)	160 (28.02)	0.005	1.325 (1.024–1.714)	1.268 (0.967–1.663)
0-copy	29 (5.28)	14 (2.45)		2.425 (1.261–4.665)	2.198 (1.111–4.348)
1 + 0-copy	210 (38.25)	174 (30.47)		1.413 (1.103–1.811)	1.346 (1.037–1.747)
Trend test *P* value				0.002	0.009
Case only study	*Stages III + IV*	*Stages I + II*			
Total number	506	43	0.308		
2-copy	309 (61.07)	30 (69.77)		1.000 (ref.)	1.000 (ref.)
1-copy	168 (33.20)	13 (30.23)		1.255 (0.637–2.470)	1.330 (0.655–2.701)
0-copy	29 (5.73)	0 (0.00)		—	—
1 + 0-copy	197 (38.93)	13 (30.23)		1.471 (0.749–2.888)	1.554 (0.771–3.129)
Trend test *P* value				0.129	0.112

[a] Adjusted in a logistic regression model that included age, age at menarche, number of births, menstrual history, oral contraceptive use, family history of cancer, smoking status, and alcohol intake.

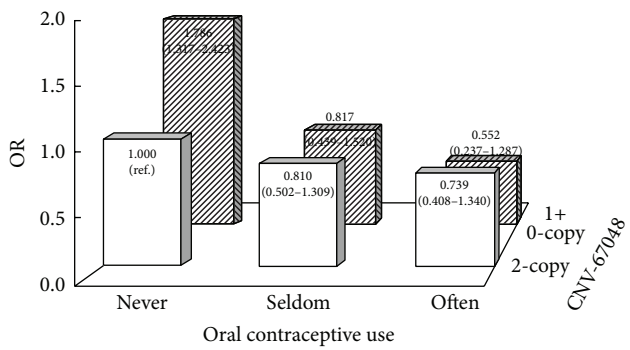


Figure 1: Negative interaction between the CNV-67048 and oral contraceptive use on EOC risk. The subjects who carried 2-copy CNV-67048 and never used oral contraceptive were defined as reference. The loss copy of CNV-67048 significantly interacted with null oral contraceptive use on EOC risk.

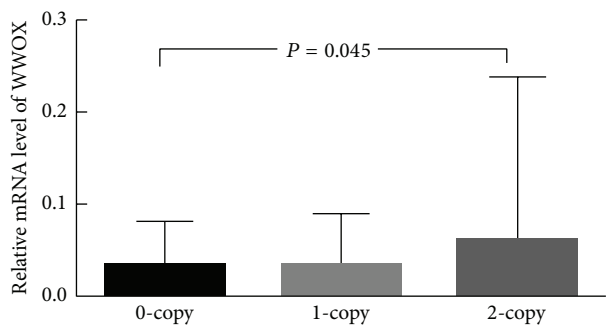


Figure 2: Effect of the CNV-67048 on WWOX mRNA expression in ovary tissues. The Kruskal-Wallis test was used to evaluate the effect of CNV-67048 on WWOX expression in tissues.

development. However, this study did not show any significant association between the CNV-67048 and EOC stages.

The WWOX protein is a kind of broad-spectrum tumor suppressor involving many kinds of human cancers [24]. Functional suppression of WWOX prevents apoptotic cell death induced by a variety of stress stimuli, such as tumor necrosis factor, UV radiation, and chemotherapeutic drug treatment [25]. Through protein-protein interaction, WWOX could directly bond onto a lot of well-known cancer-related molecules such as the p53, p73, Jun, and ErbB4 to enhance apoptosis [26]. WWOX also participates in the cellular metabolism and affects tumor metabolism and thus inhibits tumorigenesis [27]. Alteration of WWOX has been observed in many tumors, including breast [28], ovarian [22], prostate [29], lung [16], hepatocellular [30], gastric [31], and other cancers [32–35], and loss or reduction of its expression is reported to be correlated with worse clinical prognosis such as breast and ovarian cancer [28, 36]. Similarly in EOC, lower WWOX expression was found in tumors compared to normal ovaries [22]. Previous studies had showed that the 0-copy or 1-copy could cause a lower WWOX expression in human tissues than the 2-copy; thus individuals carrying low copy might be more predisposed to develop EOC in response to carcinogenic stimulation than individuals carrying 2-copy because of their innate differences in WWOX expression.

We also found a negative interaction between the CNV-67048 and oral contraceptive use on EOC risk. It is well-established that oral contraceptive plays a protective role in EOC risk [37]. Oral contraceptive can inhibit ovulation frequency and thus decrease the risk of EOC, because chronic stimulus caused by several ovulations can result in abnormal cell proliferation and repair and further tumorigenesis. As mentioned above, WWOX participates in the cell process in response to such stimulus and plays a role in cell repair.

Table 3: Associations between CNV-67048 and EOC risk stratified by selected variables.

	EOC Patients (n = 549)		Controls (n = 571)		Adjusted OR (95% CI)[a]	P_{inter}[b]
	2-copy n (%)	1 + 0-copy n (%)	2-copy n (%)	1 + 0-copy n (%)	1 + 0-copy versus 2-copy	
Age (years)						
≤55	177 (63.90)	100 (36.10)	206 (67.99)	97 (32.01)	1.175 (0.817–1.692)	0.396
>55	162 (59.56)	110 (40.44)	191 (71.27)	77 (28.73)	1.474 (0.990–2.194)	
Age at menarche (years)						
<15	133 (63.03)	78 (36.97)	177 (68.87)	80 (31.13)	1.315 (0.878–1.971)	0.564
≥15	124 (67.76)	59 (32.24)	207 (70.65)	86 (29.35)	1.193 (0.775–1.839)	
Number of births						
0	10 (62.50)	6 (43.75)	7 (63.64)	4 (36.36)	0.414 (0.043–4.002)	
1–3	238 (58.19)	171 (41.81)	342 (69.80)	148 (30.20)	1.556 (1.164–2.081)	0.075
≥4	91 (73.39)	33 (26.61)	48 (68.57)	22 (31.43)	0.873 (0.442–1.722)	
Menstrual history						
Premenopause	89 (58.55)	63 (41.45)	92 (70.23)	39 (29.77)	1.451 (0.788–2.671)	0.563
Menopause	250 (62.97)	147 (37.03)	305 (69.32)	135 (30.68)	1.262 (0.931–1.710)	
Oral contraceptive use						
Never	239 (57.73)	175 (42.27)	283 (70.93)	116 (29.07)	1.825 (1.348–2.471)	
Seldom	35 (62.50)	21 (37.50)	51 (65.38)	27 (34.62)	1.019 (0.454–2.291)	0.042
Often	23 (69.70)	10 (30.30)	33 (67.35)	16 (32.65)	0.464 (0.110–1.953)	
Family history of cancer						
Yes	23 (57.50)	17 (42.50)	25 (65.79)	13 (34.21)	2.440 (0.826–7.208)	0.654
No	316 (62.08)	193 (37.92)	372 (69.79)	161 (30.21)	1.333 (1.016–1.748)	
Smoking status						
Ever	43 (63.24)	25 (36.76)	37 (68.52)	17 (31.48)	1.353 (0.513–3.573)	0.488
Never	296 (61.54)	185 (38.46)	360 (69.63)	157 (30.37)	1.528 (1.165–2.003)	
Pack-years smoked						
≥20	23 (63.89)	13 (36.11)	14 (77.78)	4 (22.22)	10.067 (1.110–91.293)	
<20	20 (62.50)	12 (37.50)	23 (63.89)	13 (36.11)	0.676 (0.175–2.616)	0.695
0	296 (61.54)	185 (38.46)	360 (69.63)	157 (30.37)	1.528 (1.165–2.003)	
Alcohol intake						
Ever	24 (70.59)	10 (29.41)	36 (73.47)	13 (26.53)	1.548 (0.383–6.248)	0.455
Never	315 (61.17)	200 (38.83)	361 (69.16)	161 (30.84)	1.401 (1.070–1.835)	

[a] Adjusted in a logistic regression model that included age, age at menarche, number of births, menstrual history, oral contraceptive use, family history of cancer, smoking status, and alcohol intake.
[b] P value from a multiple interaction analysis.

Thus, it is possible that the low expression of WWOX driven by low copy number of CNV-67048 enhances the cellular malignant change and suppresses the protective effect of oral contraceptive use.

WWOX CNV-67048 has been reported to be associated with lung cancer and gliomas risk in previous studies [15, 16]. Here, consistently it was found to be related to EOC risk. However, although loss of WWOX expression correlates with advanced EOC stages as reported, we did not find that the CNV was correlated with the stages. By now, only four studies have tested the effect of CNV on EOC development in large populations. Gonzalez Bosquet et al. used the integrating high-throughput data from TCGA including CNVs to construct a molecular signature to predict chemoresponse in EOC [38]. Kamieniak et al. presented a CNV hallmark of BRCA1 and BRCA2 EOC [39]. Huang et al. identified special copy number landscapes for EOC histotypes [40]. Fridley et al. tested genome-wide CNVs but found no association between inherited CNVs and ovarian cancer survival [41]. It is to be observed that these above studies did not study the effect of CNV on EOC risk. To the best of our knowledge, our study is the first study to show that a CNV is related to the risk of EOC.

Although this case-control study presented a significant association between the *WWOX* CNV-67048 and EOC risk and got TCGA support to be reliable and functional, there

were some unavoidable limitations in the current study. As a hospital-based case-control study, there must be a selection bias such as Berkson bias and an information bias including recall bias. Meanwhile, our study sample size is relatively small. Additionally, for those interviews completed, data on some variables were not obtained or were unclear. This may cause bias on the estimation of the abovementioned association. Thus, further study with large sample size in other ethnics is warranted.

In conclusion, our data revealed the CNV-67048 and its interaction with oral contraceptive use to be in association with EOC risk in Chinese women. This CNV might be a genetic risk factor of EOC in Chinese women.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

Yongxiu Chen and Xiaochang Tan performed the experiments, collected the data, and wrote the paper; Bi Mai, Xiaowen Huang, and Guiying Hu performed the investigation and collected the samples; Xiping Luo conceived the study and supervised all aspects of the study. Yongxiu Chen and Xiaochang Tan contributed equally to this work.

Acknowledgments

This study was supported by the Guangdong Medical Scientific Research Grants no. A2010068 (Xiping Luo) and Scientific Research Grants of Guangdong Provincial Population Family Planning Commission no. 2010241 (Xiping Luo).

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2015," *CA: A Cancer Journal for Clinicians*, vol. 65, no. 1, pp. 5–29, 2015.

[2] W. Chen, R. Zheng, H. Zeng, and S. Zhang, "The updated incidences and mortalities of major cancers in China, 2011," *Chinese Journal of Cancer*, vol. 34, p. 53, 2015.

[3] L. E. Kelemen, K. Lawrenson, J. Tyrer et al., "Genome-wide significant risk associations for mucinous ovarian carcinoma," *Nature Genetics*, vol. 47, no. 8, pp. 888–897, 2015.

[4] K. B. Kuchenbaecker, S. J. Ramus, J. Tyrer et al., "Identification of six new susceptibility loci for invasive epithelial ovarian cancer," *Nature Genetics*, vol. 47, no. 2, pp. 164–171, 2015.

[5] K. Chen, H. Ma, L. Li et al., "Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women," *Nature Communications*, vol. 5, article 4682, 2014.

[6] F. J. Couch, X. Wang, L. McGuffog et al., "Genome-wide association study in *BRCA1* mutation carriers identifies novel loci associated with breast and ovarian cancer risk," *PLoS Genetics*, vol. 9, no. 3, Article ID e1003212, 2013.

[7] P. D. Pharoah, Y.-Y. Tsai, S. J. Ramus et al., "GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer," *Nature Genetics*, vol. 45, no. 4, pp. 362–370, 2013.

[8] A. Gusev, G. Bhatia, N. Zaitlen et al., "Quantifying missing heritability at known GWAS loci," *PLoS Genetics*, vol. 9, no. 12, Article ID e1003993, 2013.

[9] P. Stankiewicz and J. R. Lupski, "Structural variation in the human genome and its role in disease," *Annual Review of Medicine*, vol. 61, pp. 437–455, 2010.

[10] Y. Song, L. Li, Y. Ou et al., "Identification of genomic alterations in oesophageal squamous cell cancer," *Nature*, vol. 508, no. 7498, pp. 91–95, 2014.

[11] L. Yang, X. Lu, F. Qiu et al., "Duplicated copy of CHRNA7 increases risk and worsens prognosis of COPD and lung cancer," *European Journal of Human Genetics*, vol. 23, no. 8, pp. 1019–1024, 2015.

[12] L. Yang, B. Liu, F. Qiu et al., "The effect of functional MAPKAPK2 copy number variation CNV-30450 on elevating nasopharyngeal carcinoma risk is modulated by EBV infection," *Carcinogenesis*, vol. 35, no. 1, pp. 46–52, 2014.

[13] B. Liu, L. Yang, B. Huang et al., "A functional copy-number variation in MAPKAPK2 predicts risk and prognosis of lung cancer," *American Journal of Human Genetics*, vol. 91, no. 2, pp. 384–390, 2012.

[14] X. Li, X. Chen, G. Hu et al., "Combined analysis with copy number variation identifies risk loci in lung cancer," *BioMed Research International*, vol. 2014, Article ID 469103, 9 pages, 2014.

[15] K. Yu, J. Fan, X. Ding et al., "Association study of a functional copy number variation in the WWOX gene with risk of gliomas among Chinese people," *International Journal of Cancer*, vol. 135, no. 7, pp. 1687–1691, 2014.

[16] L. Yang, B. Liu, B. Huang et al., "A functional copy number variation in the WWOX gene is associated with lung cancer risk in Chinese," *Human Molecular Genetics*, vol. 22, no. 9, pp. 1886–1894, 2013.

[17] A. L. Masson, B. A. Talseth-Palmer, T.-J. Evans, D. M. Grice, G. N. Hannan, and R. J. Scott, "Expanding the genetic basis of copy number variation in familial breast cancer," *Hereditary Cancer in Clinical Practice*, vol. 12, no. 1, article 15, 2014.

[18] H. Yan and Y. Sun, "Evaluation of the mechanism of epithelial-mesenchymal transition in human ovarian cancer stem cells transfected with a WW domain-containing oxidoreductase gene," *Oncology Letters*, vol. 8, no. 1, pp. 426–430, 2014.

[19] C. Gourley, A. J. W. Paige, K. J. Taylor et al., "WWOX gene expression abolishes ovarian cancer tumorigenicity in vivo and decreases attachment to fibronectin via integrin $\alpha 3$," *Cancer Research*, vol. 69, no. 11, pp. 4835–4842, 2009.

[20] H. Yan, J. Tong, X. Lin, Q. Han, and H. Huang, "Effect of the WWOX gene on the regulation of the cell cycle and apoptosis in human ovarian cancer stem cells," *Molecular Medicine Reports*, vol. 12, no. 2, pp. 1783–1788, 2015.

[21] H. Yan and J. Sun, "Methylation status of WWOX gene promoter CpG islands in epithelial ovarian cancer and its clinical significance," *Biomedical Reports*, vol. 1, no. 3, pp. 375–378, 2013.

[22] C. Lan, W. Chenggang, B. Yulan, D. Xiaohui, Z. Junhui, and W. Xiao, "Aberrant expression of WWOX protein in epithelial ovarian cancer: a clinicopathologic and immunohistochemical study," *International Journal of Gynecological Pathology*, vol. 31, no. 2, pp. 125–132, 2012.

[23] H. Park, J.-I. Kim, Y. S. Ju et al., "Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing," *Nature Genetics*, vol. 42, no. 5, pp. 400–405, 2010.

[24] U. Lewandowska, M. Zelazowski, K. Seta, M. Byczewska, E. Pluciennik, and A. K. Bednarek, "WWOX, the tumour suppressor gene affected in multiple cancers," *Journal of Physiology and Pharmacology*, vol. 60, supplement 1, pp. 47–56, 2009.

[25] J.-Y. Lo, Y.-T. Chou, F.-J. Lai, and L.-J. Hsu, "Regulation of cell signaling and apoptosis by tumor suppressor WWOX," *Experimental Biology and Medicine*, vol. 240, no. 3, pp. 383–391, 2015.

[26] Z. Salah, R. Aqeilan, and K. Huebner, "WWOX gene and gene product: tumor suppression through specific protein interactions," *Future Oncology*, vol. 6, no. 2, pp. 249–259, 2010.

[27] M. Abu-Remaileh and R. I. Aqeilan, "The tumor suppressor WW domain-containing oxidoreductase modulates cell metabolism," *Experimental Biology and Medicine*, vol. 240, no. 3, pp. 345–350, 2015.

[28] M. I. Nunez, J. Ludes-Meyers, M. C. Abba et al., "Frequent loss of WWOX expression in breast cancer: correlation with estrogen receptor status," *Breast Cancer Research and Treatment*, vol. 89, no. 2, pp. 99–105, 2005.

[29] H. R. Qin, D. Iliopoulos, S. Semba et al., "A role for the WWOX gene in prostate cancer," *Cancer Research*, vol. 66, no. 13, pp. 6477–6481, 2006.

[30] C. Huang, Y. Tian, R. Peng et al., "Association of downregulation of WWOX with poor prognosis in patients with intrahepatic cholangiocarcinoma after curative resection," *Journal of Gastroenterology and Hepatology (Australia)*, vol. 30, no. 2, pp. 421–433, 2015.

[31] N. Maeda, S. Semba, S. Nakayama, K. Yanagihara, and H. Yokozaki, "Loss of WW domain-containing oxidoreductase expression in the progression and development of gastric carcinoma: clinical and histopathologic correlations," *Virchows Archiv*, vol. 457, no. 4, pp. 423–432, 2010.

[32] J. Yang, D. Cogdell, D. Yang et al., "Deletion of the WWOX gene and frequent loss of its protein expression in human osteosarcoma," *Cancer Letters*, vol. 291, no. 1, pp. 31–38, 2010.

[33] M. G. Diniz, E. R. Borges, F. J. Pimenta et al., "Evidence of molecular alterations in the tumour suppressor gene WWOX in benign and malignant bone related lesions of the jaws," *Oncology Reports*, vol. 25, no. 2, pp. 499–502, 2011.

[34] M.-F. Chiang, S.-T. Chen, C.-P. Lo, C.-I. Sze, N.-S. Chang, and Y.-J. Chen, "Expression of WW domain-containing oxidoreductase WOX1 in human nervous system tumors," *Analytical Cellular Pathology*, vol. 36, no. 5-6, pp. 133–147, 2013.

[35] M. Nowakowska, E. Płuciennik, W. I. Wujcicka et al., "The correlation analysis of WWOX expression and cancer related genes in neuroblastoma- a real time RT-PCR study," *Acta Biochimica Polonica*, vol. 61, no. 1, pp. 91–97, 2014.

[36] M. I. Nunez, D. G. Rosen, J. H. Ludes-Meyers et al., "WWOX protein expression varies among ovarian carcinoma histotypes and correlates with less favorable outcome," *BMC Cancer*, vol. 5, article 64, 2005.

[37] J. B. Greer, F. Modugno, G. O. Allen, and R. B. Ness, "Short-term oral contraceptive use and the risk of epithelial ovarian cancer," *American Journal of Epidemiology*, vol. 162, no. 1, pp. 66–72, 2005.

[38] J. Gonzalez Bosquet, D. C. Marchion, H. Chon, J. M. Lancaster, and S. Chanock, "Analysis of chemotherapeutic response in ovarian cancers using publicly available high-throughput data," *Cancer Research*, vol. 74, no. 14, pp. 3902–3912, 2014.

[39] M. M. Kamieniak, I. Muñoz-Repeto, D. Rico et al., "DNA copy number profiling reveals extensive genomic loss in hereditary BRCA1 and BRCA2 ovarian carcinomas," *British Journal of Cancer*, vol. 108, no. 8, pp. 1732–1742, 2013.

[40] R. Y. Huang, G. B. Chen, N. Matsumura et al., "Histotype-specific copy-number alterations in ovarian cancer," *BMC Medical Genomics*, vol. 5, article 47, 2012.

[41] B. L. Fridley, P. Chalise, Y.-Y. Tsai et al., "Germline copy number variation and ovarian cancer survival," *Frontiers in Genetics*, vol. 3, article 142, 2012.