

*Original Research Article*

## Development of a Semiconductor Sequencing-based Panel for Screening Individuals with Lynch Syndrome

Ryia-Illani Mohd Yunos<sup>1</sup>, Nurul-Syakima Ab Mutalib<sup>1</sup>, Janice Khor Sheau Sean<sup>2</sup>, Sazuita Saidin<sup>1</sup>, Mohd Ridhwan Abd Razak<sup>1</sup>, Isa Md. Rose<sup>3</sup>, Ismail Sagap<sup>4</sup>, Luqman Mazlan<sup>4</sup>, Rahman Jamal<sup>1\*</sup>

**Article History**

**Received:** 21 June 2022;

**Received in Revised Form:**  
23 July 2022;

**Accepted:** 28 July 2022;

**Available Online:** 31 July  
2022

<sup>1</sup>UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia; ryia.yunos@ppukm.ukm.edu.my (RIMY); syakima@ppukm.ukm.edu.my (NS-AM); sazuita@ukm.edu.my (SS); ridhwanrazak@ppukm.ukm.edu.my (MRAR)

<sup>2</sup>Life Sciences Solutions, Thermo Fisher Scientific Inc.; Janice.Khor@thermofisher.com (JK)

<sup>3</sup>Department of Pathology, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia; isa@ppukm.ukm.edu.my (IR)

<sup>4</sup>Department of Surgery, Faculty of Medicine, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia; ismailisagap@ppukm.ukm.edu.my (IS); luqman@ppukm.ukm.edu.my (LM)

\*Corresponding author: Rahman Jamal; UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia; rahmanj@ppukm.ukm.edu.my (RJ)

**Abstract:** Lynch syndrome is a genetic disorder associated with mutations in mismatch repair (MMR) genes that are linked to the development of colorectal cancer. Individuals with this condition have a lifetime risk of developing cancer at around 20% to 65%. Due to the autosomal dominant inheritance pattern, close biological relatives are also at high risk. Early detection of CRC may lead to better health outcomes and considerable savings in treatment costs. Therefore, our objective is to develop a rapid screening method for LS. We designed an Ion Ampliseq™ Custom Panel, which includes four MMR genes associated with LS (MLH1, MLH2, MSH6, and PSM2), a downstream gene (EPCAM), and a gene that indicates sporadic CRC (BRAF) for sequencing on the Ion Torrent PGM™ sequencer. Sequencing was performed on 16 DNA samples derived from various stages of CRC. The sensitivity and specificity of the identified mutations were determined by sequencing the serially diluted DNA from two human cancer cell lines, HCT 116 and LN-18. An average of 92 % of reads were mapped to the target region with 98 % uniformity. There was no amplicons dropout across all samples, and 58 variants were chosen for validation in 19 samples using MassARRAY and Sanger sequencing. We achieved 87% specificity, 97% accuracy, and 100% sensitivity for detecting variants at an allele frequency of more than 13% using this LS gene panel. With the development of this method, LS CRC can be detected at an early stage using this rapid and sensitive approach.

**Keywords:** Lynch Syndrome, Colorectal Cancer, MMR Genes, Next Generation Sequencing, Gene Panel

## 1. Introduction

Colorectal cancer (CRC) is the third most common cancer in men (10.9% of the total cases) and the second in women (9.5% of the total cases)<sup>[1]</sup>. CRC is a major global health burden and causes significant morbidity and mortality. Almost 55% of CRC cases occur in developed regions<sup>[1]</sup>. In Malaysia, CRC is the commonest cancer among males and the second most common cancer among females<sup>[2]</sup>. This cancer is widely preventable through the various interventions that the community can adopt, such as adopting a healthy lifestyle and regular medical screening<sup>[3,4]</sup>.

In general, there are three classifications of CRC based on the increasing hereditary risk of cancer<sup>[5,6]</sup>. The most common type is sporadic CRC, which accounts for about 70% of all cases. The other two are familial and hereditary CRCs<sup>[5,6]</sup>. Familial CRC (20%) refers to patients with at least one biological family diagnosed with CRC, but there is no specific germline mutation or obvious pattern of inheritance<sup>[6]</sup>. Hereditary CRC (10%), also known as Lynch Syndrome (LS), is caused by the inheritance of germline mutations in highly penetrant cancer susceptibility genes. LS is an autosomal dominant disease, in which if one of the parents has the disease, there is a 50% chance that the offspring will inherit it. This syndrome is marked by an increased risk of developing CRC and other cancers, such as the endometrium, ovary, stomach, hepatobiliary tract, urinary tract, brain, and skin [7, 8]. Besides, the risk of developing a second primary CRC is high (approximately 16% within ten years), and the risk of cancer in a first or second-degree family member is around 45% for men and 35% for women by the age of 70 [9]. While the last group is the least common, this group is crucial in understanding the molecular mechanisms of carcinogenesis. It has important implications for the screening and follow-up of patients and their families [6].

Mutations in DNA mismatch repair (MMR) genes; namely, MutL homologue 1 (MLH1), MutS homologues 2 and 6 (MSH2 and MSH6), and postmeiotic segregation increased 2 (PMS2), are the main causes of Lynch Syndrome<sup>[10]</sup>. Mutations in the MMR genes have prevented the repair of base mismatches, minor insertions, and deletions that might result in cancers, which is the reason for the loss of MMR function in a cell.<sup>[10]</sup> Global data show that MLH1 accounts for 39% of entries in the International Society for Gastrointestinal Hereditary Tumours (InSiGHT) database ([www.insight-group.org/mutations/](http://www.insight-group.org/mutations/)), MSH2 for 34%, MSH6 for 20%, and PMS2 for 8%<sup>[11]</sup>. Identification of MMR gene mutation confirms the clinical diagnosis of LS and therefore identifies the individuals who should undergo routine surveillance for associated cancers. Another gene, EPCAM, which was not categorised in the MMR group but is located adjacent to one of the MMR genes (MSH2), was also identified as the cause of this disease. Mutation of the 3' end of EPCAM can affect MSH2 gene expression, whereby EPCAM 3' deletion can lead to inactivation of the MSH2 gene<sup>[12,13]</sup>. Therefore, EPCAM should also be included in the screening process for LS. Besides, a marker such as the BRAF V600E mutation may help to distinguish sporadic from hereditary or familial tumours<sup>[14,15]</sup>. The BRAF gene, which belongs to the RAF-RAS gene family, encodes a cytoplasmic serine/threonine kinase, a key component of the mitogen-activated protein kinase signalling pathway. In 15% of sporadic CRCs, somatic mutations in

the BRAF gene, primarily at codon 600, are found. Because BRAF gene mutations are mostly linked to sporadic CRC, testing of this gene will essentially rule out the diagnosis of LS [15].

Experts in pharmacoeconomics agree that CRC is linked to increased economic burden [16, 17]. The annual incidence is forecasted to increase by ~80 % from the current 1.2 million estimated cases over the next two decades. A majority of the increase is expected to be contributed by less developed regions [16]. The long-term cost of managing CRC was estimated to go up to \$50,175 per patient in 2008 [16]. Hence, the efforts to control CRC are becoming critical [18, 19]. The current standard for diagnosis of LS includes mutation detection using Polymerase Chain Reaction (PCR) followed by Sanger sequencing. Using this technique, mutations in the six genes mentioned earlier must be screened one at a time. One diagnostic laboratory charges around US\$1 000 per gene. Testing six genes will easily cost the patient more than US\$6,000 for such tests, excluding any pre-screening test (including MSI analysis, IHC, and BRAF test for V600E), before the genetic screening (<http://fightlynch.org/physicians-info/genetic-testing/>). Hence, this approach is expensive, low throughput, and time-consuming, primarily because at least five or six genes may have to be analysed, and their mutational spectra are extensive [19]. Developing a benchtop Next Generation Sequencing (NGS) instrument that allows the analysis of multiple genes simultaneously is a superior method to Sanger sequencing. The higher throughput capabilities of this technology remarkably reduced the cost of mutation screening and shortened the turnaround time. Several diagnostics service centres abroad, such as Invitae and the University of Chicago (Genetic Services Laboratory) in the USA, have already been offering this screening service but at approximately US\$1500 to US\$3000 per patient. However, if we develop this test locally in Malaysia, the cost will be around US\$628. To the best of our knowledge, our country has no diagnostic service provider yet that offers the mutation screening of these six genes using NGS technology. Therefore, we aim to develop a relatively affordable, fast, specific, and sensitive screening for LS using the NGS approach.

## 2. Materials and Methods

### 2.1 Tumour Specimens and Cell Lines

Sixteen specimens of CRC with known mutations were selected from the UMBI-PPUKM Biobank. These samples have been analysed for variant detection via whole exome sequencing [20] and denaturing High-Performance Liquid Chromatography (dHPLC) in our previous studies [21]. The average age of patients was approximately 66 years old (range 44 – 75 years). With regards to cancer stage, 13% (n=2) of the patients were of Dukes' A, 38% (n=6) were Dukes' B, 43% (n=7) were Dukes' C, and the remaining 6% (n=1) were Dukes' D. HCT 116 is a cell line with a heterozygous mutation of MLH1: p.S252X, c.755C>A while the LN-18 cell line is negative for this mutation. This mutation is a verified COSMIC variant. The DNA extraction was performed using the QIAamp DNA mini kit (Qiagen, Germany) according to the manufacturer's protocol. The quality and quantity of the extracted DNA were assessed using the Qubit Fluorometer (Invitrogen, USA), NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, US), and agarose gel electrophoresis.

## 2.2 Ion Ampliseq Custom Panel Design

Ion Ampliseq Custom Panels (Life Technologies, US) covering the entire coding region of MLH1, MSH2, MSH6, PMS2, BRAF, and 3' untranslated region (UTR) of EPCAM was designed using the Ion Ampliseq Designer v4 (<https://ampliseq.com>). The designed gene panel contains a pool of 115 primers for multiplex amplification of genomic regions of interest.

## 2.3 Ion Ampliseq Library Preparation

Libraries were constructed using Ion Ampliseq™ Library Kit 2.0 and Ion Ampliseq™ Custom Panel (both from Life Technologies, US) using ten ng of genomic DNA according to the protocol suggested by the manufacturer. Genomic DNA was amplified by PCR for each primer pool (two primer pools for this panel), followed by partial digestion of primer sequences with FuPa reagent. The amplicons were then ligated to the adapters with the addition of a barcode from the Ion Xpress™ Barcode Adapters 1-16 before being subjected to amplification and purification. Amplified libraries were assessed on the Agilent® Bioanalyzer® instrument using the High Sensitivity DNA kit (both from Agilent Technologies, US). Libraries were diluted to 60 pM, and eight libraries were combined in one pool before proceeding with template preparation.

## 2.4 Template Preparation and Sequencing

Enriched template-positive Ion Sphere™ Particles were prepared for sequencing on Ion Chef™ System (Life Technologies, US). Combined libraries (eight libraries) were loaded onto an Ion 318™ BC chip and subsequently sequenced on Ion Torrent PGM (IT-PGM) sequencer (Life Technologies, US). Ion PGM™ HiQ™ Sequencing kit (Life Technologies, US) was used for sequencing up to ~400 bp library inserts to improve the indel sequencing accuracy of targeted resequencing panels.

## 2.5 Sensitivity Analysis

The IT-PGM platform's mutation detection sensitivity was tested by sequencing serially diluted DNA from a human colorectal adenocarcinoma cell line, HCT 116. The DNA from HCT 116 was diluted in the DNA obtained from a malignant glioma cell line, LN-18 (CRL-2610; ATCC, US) in the ratio of 1:3, 1:9, and 1:19, resulting in 25%, 10%, and 5% dilution.

## 2.6 Validation of Mutations by the Agena MassARRAY System and Sanger Sequencing

Fifty-eight variants identified across 19 samples (including three from cell lines) were validated via MassARRAY (Agena Bioscience, CA). The four wells containing 58 assays for variant screening were developed in our laboratory using the MassARRAY® Assay Design Suite to achieve universal thermal cycling conditions for all assays. DNA was amplified and subjected to single-base primer extension using the iPLEX Gold kit (Agena Bioscience, CA) and analysed using the MALDI-TOF (mass-assisted laser desorption/ionisations-time of flight) mass spectrometry (Agena Bioscience, CA). Our four-well multiplex primer extension assay was designed to assess the mutational status of 58 variants identified in MLH1, MSH2, MSH6, PMS2, EPCAM, and BRAF by IT-PGM. The target regions were amplified using 20 ng of DNA per well, followed by dephosphorylation of the unincorporated nucleotides by treatment with alkaline phosphatase. The amplified product was used as the template for a

locus-specific single base extension with mass-modified dideoxynucleotides. Primers for PCR amplification and single base extension were designed using the MassARRAY® Assay Design Suite and were synthesised by Integrated DNA Technologies, IDT (Coralville, IA). MALDI-TOF analysed the mass of products of a single base extension for single nucleotide polymorphism detection using Typer Analyzer software (Agena Bioscience, CA).

Variants that were not successfully validated by MassARRAY (Agena Bioscience, CA) were subjected to Sanger sequencing for further validation. PCR was performed using genomic DNA as a template and primer pairs flanking the variant sites. The PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Germany) according to the manufacturer's instructions. The cycle sequencing was performed using the Big Dye Terminator V3.1 (Life Technologies, US). The cycle sequencing products were then purified using ethanol precipitation, and sequencing was carried out using the ABI 3130xl Genetic Analyser capillary electrophoresis (Life Technologies, US). The results were analysed using the Basic Local Alignment System Tool (BLAST) <sup>[22]</sup> and Sequence Scanner (Applied Biosystem, US).

## 2.7 Data Analysis

We analysed the sequencing data using three pipelines to perform the alignment and variant calling. First, the sequencing data were processed using Ion Torrent Suite™ Software v4.4.2.1 running on the Torrent Server. The pipeline included signal processing, base calling, adapter trimming, PCR duplicate removal, alignment to the human genome 19 reference (hg19), quality control of mapping quality, coverage analysis, and variant calling. Coverage analysis and variant calling were performed using Torrent Variant Caller plugin software v4.4.2.1 in the Torrent Server. The variant caller parameter setting was germline PGM low stringency. We obtained the FASTQ files from Torrent Suite Software and subjected them to quality assessment and pre-processing using FastQC software <sup>[23]</sup>. FastQC: a quality control tool for high throughput sequence data. Available online at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. It is important to ensure only high-quality bases were used for the variant analysis. We performed alignment and variant calling from here using another two additional pipelines. Firstly, the FASTQ files were mapped to the human reference genome using Bowtie2 v2.2.6 <sup>[24]</sup>, and variant calling was performed using SAMtools/BCFtools (mpileup) v0.1.19 <sup>[25]</sup>. Secondly, the alignment was performed using Burrows-Wheeler Aligner (BWA) v0.7.12 <sup>[26]</sup>, and indel realignment, base recalibration and variant calling were performed using GATK-Picard pipeline 3.4-46 <sup>[27]</sup>. The variants were further annotated using ANNOVAR <sup>[28]</sup> against alternative allele frequency data in 1000 Genomes Project (1000g2014oct\_all) (The 1000 Genomes Project Consortium, 2015) <sup>[29]</sup>, dbSNP version 138 (snp138) <sup>[30]</sup>, CLINVAR (clinvar\_20150629) <sup>[31]</sup> and COSMIC version 70 (cosmic70) <sup>[32]</sup>.

## 2.8 Determination of Specificity, Sensitivity, and Accuracy of Ion Torrent PGM Gene Panel Assay

We applied a method by Parikh et al. (2008) <sup>[33]</sup> to determine sensitivity (probability of being tested positive when disease present), specificity (probability of being tested negative

when disease absent), and accuracy (measured by sensitivity and specificity). True positive, true negative, false positive, and false negative variants were analysed across 19 samples by Sanger sequencing and MassARRAY. Below is the equation used to determine the specificity, sensitivity, and accuracy:

$$\text{Specificity} = \frac{\text{Number of true negatives}}{\text{Number of true negatives} + \text{Number of false positives}}$$

$$\text{Sensitivity} = \frac{\text{Number of true positives}}{\text{Number of true positives} + \text{Number of false negatives}}$$

### 3. Results

#### 3.1 The NGS Panel Performance

The six genes (MLH1, MSH2, MSH6, PMS2, EPCAM, and BRAF) were sequenced in the 19 samples (including three samples from cell lines), generating a mean of 664 260 reads per sample. On average, 92% of these reads mapped to the targeted regions, with uniformity across all targets being 98%. The percentage of reads mapped to the target was a bit low due to the paralogous gene to PMS, which is PMS2CL. Nevertheless, other genes were well-covered. No amplicons dropout was observed across all samples. However, to determine whether coverage was substantially lower for any particular region, we calculated the proportion of amplicons covered less than 40X. At about 500X sequencing coverage, 2.6% of the amplicon was covered less than 40X. At about 4000X sequencing coverage, the percentage of amplicons covered less than 40X decreased to 1.7%. No amplicons were covered at less than 40X when sequencing coverage reached 9000X. The amplicons covered less than 40X covered MLH1 and MSH6 genes. Hence, variants identified in these amplicons will require careful interpretation. On the other hand, an evenly distributed mean depth of coverage 4000-5000X for all six genes across 19 samples was achieved.

#### 3.2 Comparison of Three Different Data Analysis Pipelines

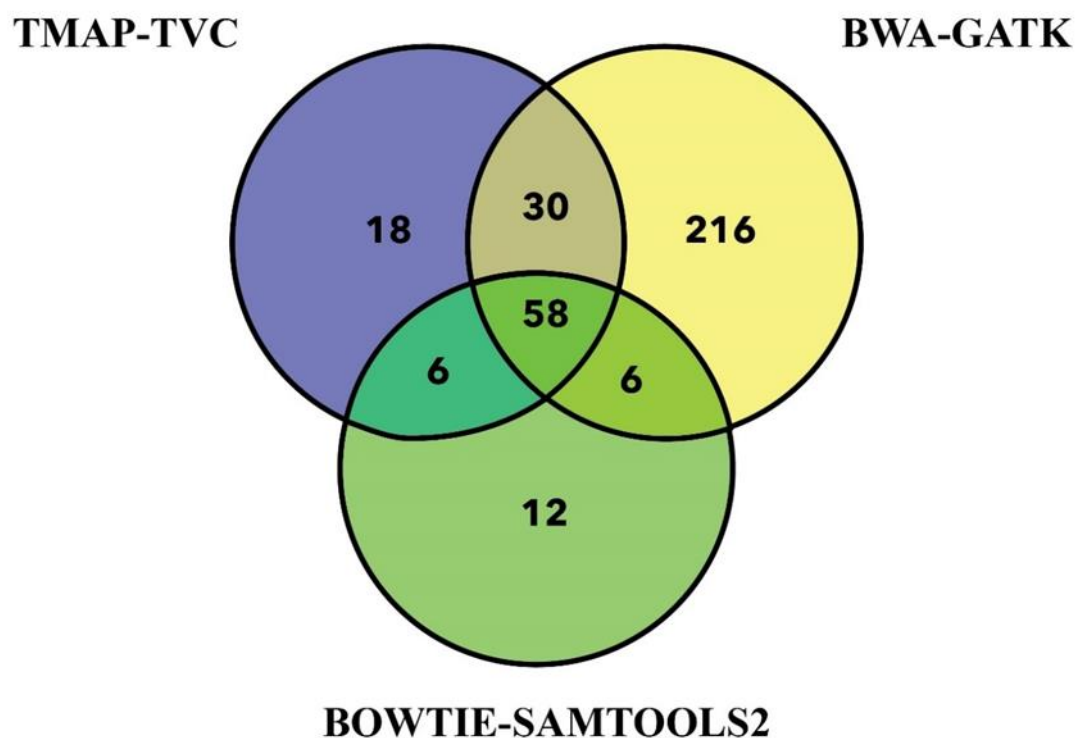
The data were analysed using three different pipelines for comparison. From the analysis, we found that an average of 99.9% of the reads were mapped to the genome. TMAP gave the highest percentage of mapped reads to the target region, 92%. In terms of the number of called variants, GATK gave the highest number of variants compared to the other two pipelines (Table 1).

**Table 1.** Comparison of mapped reads and variants called between three different pipelines.

<b>Mapping and Alignment</b>	<b>BWA</b>	<b>Bowtie2</b>	<b>TMAP</b>
Percentage of mapped reads to the reference genome	99.9	99.9	99.7
Percentage of mapped reads to target region	86.9	86.8	92
<b>Variant Calling</b>	<b>BWA-GATK</b>	<b>Bowtie2-Samtools</b>	<b>TMAP-TVC</b>
SNPs	105	68	93
INDELS	209	14	26

### 3.3 Sequencing on Ion Torrent PGM and Variants Detection

Upon the completion of data analysis using three different pipelines, 58 variants were found to be overlapped and chosen for validation in 19 samples (resulting in a total of 381 variants) (Figure 1). This analysis classified six variants as pathogenic according to ClinVar<sup>[31]</sup> (Assessed on 13th May 2016), as shown in Table 2. The Ion Torrent PGM identified two missense mutations in MLH1 and MSH2 genes in patients C360T and C76T, respectively. Both are pathogenic variants as classified by ClinVar. A single base-pair substitution in the MSH2 gene (C to T) with a variant frequency of 52% was identified in patient C76T and validated by the MassARRAY (Figure 2A). BRAF V600E mutation was also clearly identified by the Ion Torrent PGM in patient C41T. This mutation is a single base pair substitution in the BRAF gene resulting in a substitution of valine with a glutamic acid residue, and the MassARRAY confirmed this. The reliability of the mutation call was evident by the high total sequencing coverage of 11,223X (Figure 2B). Patient C36T harbour at least one pathogenic missense mutation in MSH6 with a variant frequency of 50%. We compared one of the pathogenic variants identified in patient C337T with the in-house exome data. Interestingly, the mutation we detected using the Lynch panel was somatic in our previous exome study<sup>[20]</sup>. The variant was covered at about 30% in the exome and targeted panel. This was also confirmed by MassARRAY (Figure 3).

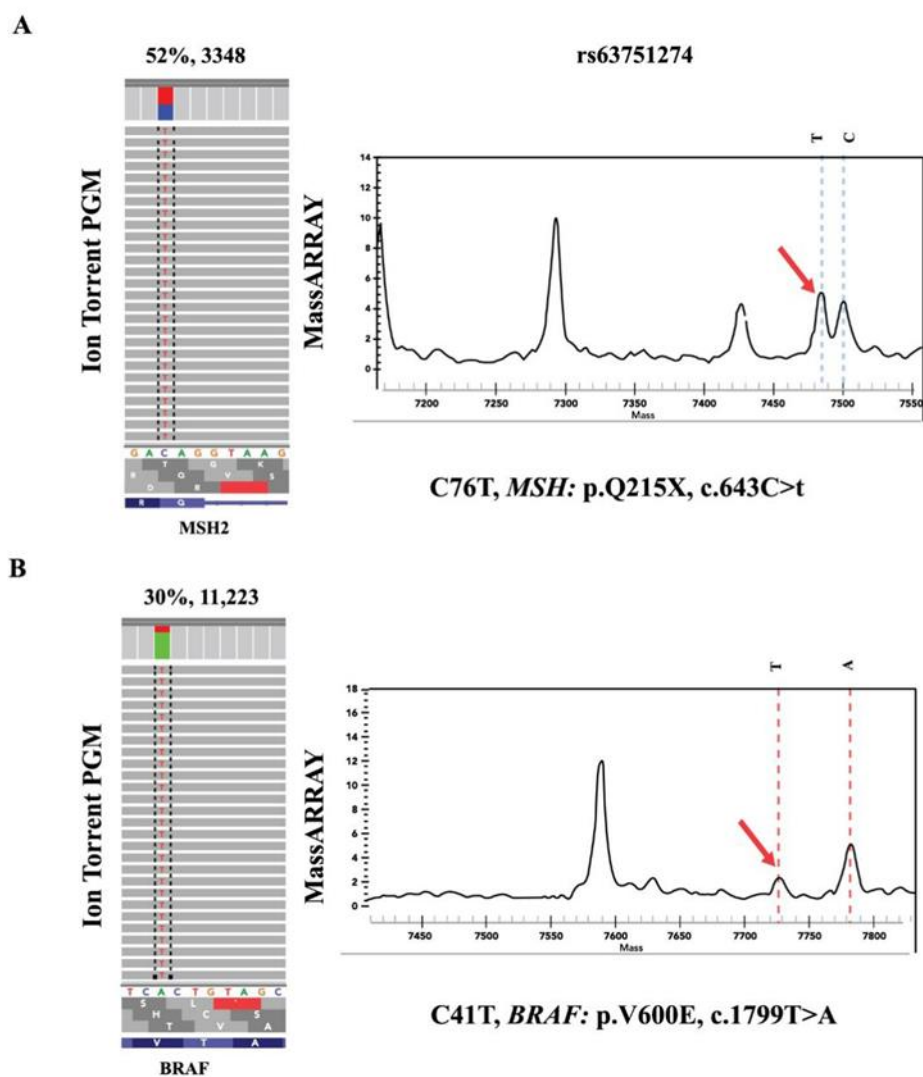


**Figure 1.** Overlapped variants were identified from three different data analysis pipelines.

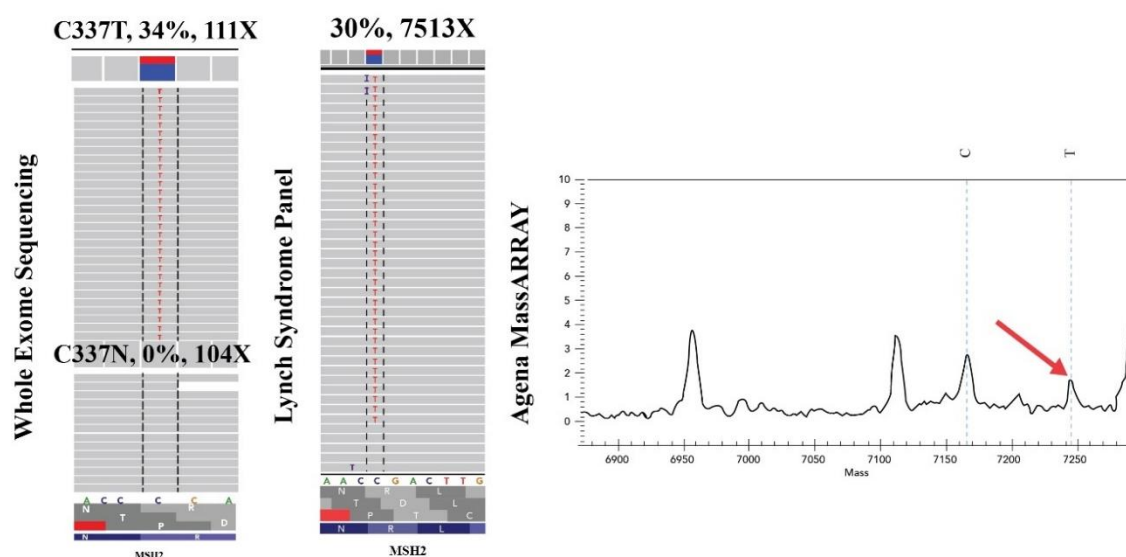
**Table 2.** List of identified pathogenic variants.

Sample ID	Gene	Protein Change	DNA Change	Variants Frequency	Coverage	Variants Coverage	dbSNP ID	ClinVar
C41T	BRAF	p.V600E	c.1799T>A	30%	11223	3348	rs113488022	Pathogenic
C360T	MLH1	p.C77Y	c.230G>A	68%	3522	2381	rs63750437	Pathogenic
HCT116_LN-18-5%				13%	4793	632		
HCT116_LN-18-25%	MLH1	p.S252X	c.755C>A	30%	5645	1670	rs63750198	Pathogenic
C76T	MSH2	p.Q215X	c.643C>T	52%	3348	1748	rs63751274	Pathogenic
C337T	MSH2	p.R389X	c.1165C>T	30%	7513	2235	rs587779075	Pathogenic
C36T	MSH6	p.Y214Y	c.642C>T	50%	4000	1992	rs1800937	Pathogenic





**Figure 2(A).** A pathogenic variant was identified in the *MSH2* gene of patient C76T. **Figure 2(B).** A pathogenic variant was identified in the *BRAF* gene of patient C41T.



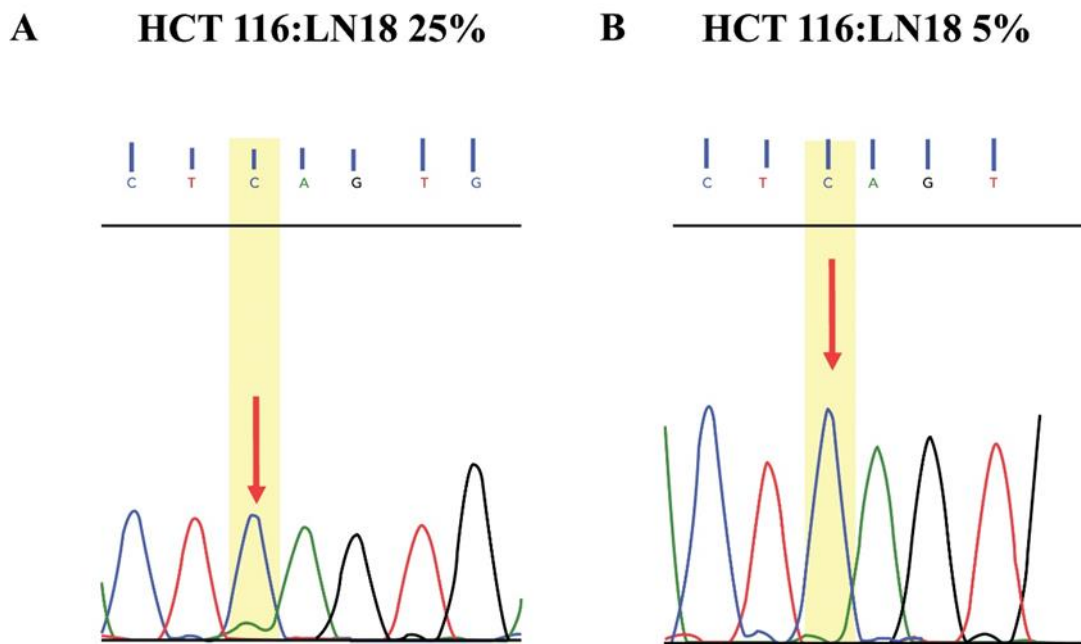
**Figure 3.** A pathogenic variant in the MSH2 gene of patient C337T, comparing whole exome sequencing (matched normal and tumour sample) and Lynch Panel.

### 3.4 Sensitivity analysis of Ion Torrent PGM Gene Panel Assay

Serially diluted DNA from two human cancer cell lines: HCT 116 and LN-18, were sequenced and used to determine the sensitivity of the Ion Torrent PGM technology for variant identification. DNA from HCT 116 was diluted into DNA from LN-18, resulting in 25%, 10%, and 5% dilution, respectively. From the sequencing, we managed to identify the mutation with a variant frequency of about 25 to 30% and 10-13% for 25% dilution and 10% dilution, respectively, using three different pipelines (Table 3). We have validated the variant detected in our sensitivity sample using Sanger sequencing. However, due to the limit of detection by Sanger sequencing, we could not identify as low as 5% of allele frequency (Figure 4).

**Table 3.** Variant frequency of MLH1, c.755C>A in serially diluted DNA HCT 116 and LN-18.

Samples	Details	Bowtie2-Samtools	BWA-GATK	TMAP-TVC
HCT116_LN-18 25%	Coverage	5,603	2757	1,670
	Frequency (Ref)	70.79%	74.95%	70%
	Frequency (Alt)	29.21%	25.05%	30%
HCT116_LN-18 10%	Coverage	Not detected	2,403	632
	Frequency (Ref)		89.01%	87%
	Frequency (Alt)		10.99%	13%



**Figure 4.** Validation of MLH1, c.755C>A using Sanger sequencing in (A) serially diluted DNA from HCT 116 to LN-18 at 25% and (B) serially diluted DNA from HCT 116 to LN-18 at 5%.

### 3.5 Determination of Specificity, Sensitivity, and Accuracy of Ion Torrent PGM Gene Panel Assay

The develop panel's specificity, sensitivity, and accuracy were determined by comparing the number of variants identified by the Ion Torrent PGM with variants detected by MassARRAY or Sanger sequencing. Our LS panel achieved 87% specificity, 100% sensitivity, and 97% accuracy.

## 4. Discussion

Lynch syndrome (LS) is inherited in an autosomal dominant manner, which means that a parent with LS has a 50% (1 in 2) chance of passing the condition on to their children. It also will not skip a generation, meaning the grandchildren will not be affected if the children do not inherit LS [34]. It is important to note that people who inherit LS have a significantly increased risk of developing cancer, not the disease itself. Not all people who inherit mutations in these mentioned genes will develop cancer. However, screening of LS may help in the early detection of CRC and save the treatment cost. It also helps in managing the psychological impact and emotions of the individuals. Non-carriers may avoid unnecessary surveillance programs and experience relief from worries. The carriers can minimise their risk by applying a healthy lifestyle and getting annually screened via colonoscopy. Therefore, developing a rapid and sensitive method to screen LS is strategic, especially in countries aiming to reduce the incidence of CRC.

According to the American College of Medical Genetics and Genomics (ACMG) guideline [35], the standard procedure for evaluating tumour tissue for MSI is immunohistochemistry of four MMR proteins. However, MSI status alone is insufficient to diagnose LS as 10-15% of sporadic CRC exhibit MSI. Methylation test on MLH1 and somatic BRAF pathogenic variants may help identify those tumours more likely to be

sporadic than hereditary <sup>[36]</sup>. The molecular genetic testing of the MMR genes can be performed to identify a germline pathogenic variant when findings from these two tests are consistent with LS.

We ventured into the development of rapid tests using NGS based panel to screen individuals with LS. As BRAF gene mutation is predominantly associated with sporadic CRC, we've included the BRAF gene in our panel to rule out the diagnosis of LS. It requires a small amount of DNA and provides simultaneous detection of variants in six genes with high accuracy and sensitivity. Upon analysis using three different pipelines, we shortlisted 58 highly confident variants to be validated via Sanger sequencing and MassARRAY. Six of these 58 variants were pathogenic based on ClinVar in five patients. We observed that the TMAP-TVC pipeline outperformed the other two pipelines, Bowtie2-Samtools and BWA-GATK, regarding alignment to the genome and mapping to the target.

It is known that ion semiconductor sequencing platforms, such as Ion Torrent PGM, suffer from inaccuracy in detecting variants in the homopolymer regions <sup>[37]</sup>. These homopolymer errors often lead to inaccurate local alignment results, and false-negative detection and careful interpretation of variants located in this region are critical. However, using open-source software such as GATK and SAMtools, we showed that the false negatives could be minimised using appropriate bioinformatics analysis <sup>[37]</sup>. In this study, we aim to determine the developed panel's sensitivity, specificity, and accuracy rather than the pipeline used. To overcome the issue of false-negative detection in homopolymer regions, we performed sequencing on the Ion PGM™ system using Ion PGM™ Sequencing 400 Kit (Thermo Fisher Scientific) and the Ion PGM™ Hi-Q™ Sequencing Kit (Thermo Fisher Scientific). HiQ chemistry was claimed to reduce mapping errors to 49 errors for 10kbp mapped compared to 89 errors using the former kit. Additionally, reduction of insertion and deletion (indel) errors, including homopolymer errors, by 80% across 400 base pair read lengths on the Ion PGM™ system, resulting in higher data quality <sup>[38]</sup>. This was also supported by a study in the forensic genetic laboratories by Churchill et al. in 2016. This suggests that reliable and accurate data can be generated, and the identification of variants in the homopolymeric region can be improved using Ion Torrent Hi-Q™ Sequencing Chemistry <sup>[39]</sup>.

The lifetime risk of CRC in LS has been estimated in various ways, and it appears to depend on gender and the mutated MMR gene <sup>[40]</sup>. Most studies estimate the lifetime risks of CRC for MLH1 and MSH2 gene mutation carriers to be between 30 and 74% of CRC. Patients with MSH6 mutations, on the other hand, have a decreased lifetime risk of colorectal cancer, ranging from 10% to 22%, compared to 15% to 20% in those with PMS2 mutations <sup>[41]</sup>. In LS patients, the average age of CRC diagnosis is 44 to 61 years, compared to 69 years in sporadic cases of CRC <sup>[42]</sup>. CRC is becoming more common in young people, with one out of every ten new cases affecting those under 50 years old <sup>[43]</sup>.

We observed a pathogenic heterozygous variant in the MSH2 gene of patient C76T, substitution from C to T. Since we have documented the presence of a mutation in this 44-year-old man who was diagnosed with Dukes' D CRC, testing for at-risk individuals in the family is possible. Genetic counselling is recommended for this patient and for other family members at risk for carrying this mutation. Extensive investigation of MMR gene mutation status in the family may be required to rule out LS if they meet the Amsterdam criteria or Bethesda guidelines.

On top of mutations identified in MMR genes, we observed a BRAF V600E mutation in 81 years old women with colorectal cancer Dukes' B (moderately differentiated adenocarcinoma). It is known that mutations in the four MMR genes (MLH1, MSH2, MSH6, and PMS2) account for only half of the LS cases identified by pedigree criteria <sup>[44]</sup>. Most sporadic colorectal tumours show no mismatch repair defects <sup>[45]</sup>. On the other hand, 12–15% of CRC with defective MMR and MSI-H phenotype is primarily due to hypermethylation of the MLH1 gene promoter, not to the germline. In more detailed analyses, BRAF mutations were not detected in those cases with a germline mutation in either MLH1 or MSH2 mutations <sup>[45]</sup>. Domingo and colleagues 2004 reported that 40% of sporadic MSI-H tumours harbour BRAF V600E, but none was found in 111 tested LS tumours <sup>[36]</sup>. Thus, we conclude that this 81-year-old woman probably has a sporadic form of CRC. We also compared the data obtained from the NGS panel with the whole exome sequencing (WES) data of the same patients we analysed. Whole exome sequencing of the tumour and matched normal tissue of patient C337T was previously performed. We identified one pathogenic variant, MSH2: p.R389X, c.1165C>T, in the tumour tissue via WES and targeted sequencing using our developed panel. However, this variant was not observed in the matched normal tissue of the C337T patient. Thus, we suggest that this mutation is most probably a somatic mutation. Our finding is also supported by a study from Haraldsdottir et al. in 2014, suggesting that deficiency can arise from somatic mutations. In the study, they observed some patients with MMR deficiency during screening for LS, and 70% of these patients acquired somatic mutations rather than germline mutations in MMR genes <sup>[46]</sup>. Thus, it is recommended that patients with CRC and MMR deficiency not explained by germline mutations might undergo analysis for somatic mutations in MMR genes to guide future surveillance guidelines. However, we would like to highlight some of the limitations and shortcomings of the present study. One main concern is the limited number of Dukes' A CRC patients, which restricts the generality of this applied method for an early stage of CRC with LS. Samples from tumour tissues and cell lines were insufficient to address its accuracy. Results of detected variants from adjacent normal tissues should be employed to confirm its efficacy for early detection of LS.

## 5. Conclusions

In summary, multiplex PCR followed by NGS is helpful for screening individuals with LS by detecting germline mutations in the MMR genes. Including BRAF gene mutation screening in our panel will assist in differentiating sporadic CRC from LS. We achieved 87% specificity, 97% accuracy, and 100% sensitivity for detecting variants at more than 13% allele frequency. Overall, this Ampliseq-based panel was specific and sensitive enough for mutation analysis of MMR genes and can be incorporated into daily clinical practice. However, further investigation is warranted on detecting Indels in MMR genes and PMS2 (due to highly homology to PMS2L).

**Author Contributions:** RIMY, NSAM, and JKSS conceived and designed the experiments; NSAM and JKSS designed the gene panel; RIMY performed the next generation sequencing and validation using MassArray; SS and MRAR performed Sanger sequencing; RIMY and KJS analysed the data; IS and LM provide the tissue samples, IMR is the pathologist who assessed the tumour percentage of the samples, RIMY prepared the manuscript, figures, and tables, NSAM and RJ provide critical insights of the scientific contents. All authors read and approved the final manuscript.

**Funding:** This work was funded by Universiti Kebangsaan Malaysia Dana Pembangunan Penyelidikan Grant (DPP-2015-119).

**Acknowledgements:** We thank Mr Ang Mia Yang and Ms Pei Sin Chong from Codon Genomic Sdn Bhd for partly being involved in the bioinformatics analysis.

**Conflicts of Interest:** The authors declare no conflict of interest

## References

1. Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; 68(6): 394-424.
2. Hashimah B, Nirmal K, Nabihah A, et al. Malaysia National Cancer Registry 2012-2016. World Cancer Research Fund/American Institute of Cancer Research (WCRF/AICR), Food, nutrition, physical activity, and the prevention of cancer: a global perspective 2007.
3. Hampel H. Genetic counselling and cascade genetic testing in Lynch syndrome. *Fam Cancer* 2016; 17(3): 423-427, 2016.
4. Ivanovich JL, Read TE, Ciske DJ, et al. A practical approach to familial and hereditary colorectal cancer. *Am J Med* 1999; 107(1):68-77.
5. Kheirelseid EAH, Miller N, and Kerin MJ. Molecular biology of colorectal cancer: Review of the literature. *Am J Mol Biol* 2013; 3: 72-80, 2013.
6. Lynch H and Chapelle A de la. Hereditary Colorectal Cancer. *N Engl J Med* 2003; 17: 31-36.
7. Wang Y, Wang Y, Li J, et al. Lynch syndrome-related endometrial cancer: Clinical significance beyond the endometrium. *J Hematol Oncol* 2013; 6 (1): 22.
8. Vasen HFA and De Vos WH. Lynch syndrome — how should colorectal cancer be managed? *Nat Publ Gr* 2011; 8(4): 184-186.
9. Hampel H, Frankel WL, Martin E, et al. Screening for the Lynch syndrome (hereditary non-polyposis colorectal cancer). *N Engl J Med* 2005; 352(18): 1851-1860.
10. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. *J Natl Cancer Inst* 2004; 96 (4): 261-268.
11. Plazzer JP, Sijmons RH, Woods MO, et al. The InSiGHT database: Utilising 100 years of insights into Lynch Syndrome. *Fam Cancer* 2013; 12 (2): 175-180.
12. Kuiper RP, Vissers LELM, Venkatachalam R, et al. Recurrence and variability of germline EPCAM deletions in Lynch syndrome. *Hum Mutat* 2011; 84-108.
13. Tutlewska K, Lubinski J, and Kurzawski G. Germline deletions in the EPCAM gene as a cause of Lynch syndrome - literature review. *Hered Cancer Clin Pract.* 2013; 11 (1): 9.
14. Koinuma K, Shitoh K, Miyakura Y, et al. Mutations of BRAF are associated with extensive hMLH1 promoter methylation in sporadic colorectal carcinomas. *Int J Cancer* 2004; 108 (2): 237-242.

15. Thiel A, Heinonen M, Kantonen J, et al. BRAF mutation in sporadic colorectal cancer and Lynch syndrome. *Virchows Arch* 2013; 463(5): 613-621.
16. Kriza C, Emmert M, Wahlster P, et al. Cost of Illness in Colorectal Cancer: An International Review. *Pharmacoeconomics* 2013; 31(7): 577-588.
17. Luengo-Fernandez R, Leal J, Gray A, et al. Economic burden of cancer across the European Union: A population-based cost analysis. *Lancet Oncol* 2013; 14(12): 1165-1174.
18. Karsa LV, Lignini TA, Patnick J, et al. The dimensions of the CRC problem. *Best Pract Res Clin Gastroenterol* 2010; 24(4): 381-396.
19. Vasen HFA, Moslein G, Alonso A, et al. Guidelines for the clinical management of Lynch syndrome (hereditary non-polyposis cancer). *J Med Genet* 2007; 44(6): 353-362.
20. Mohd Yunus RI, Ab Mutalib NS, Khor SS, et al. Whole exome sequencing identifies genomic alterations in proximal and distal colorectal cancer. *Prog Microbes Mol Biol* 2019; 2(1), 1-14.
21. Abdul Murad NA, Othman Z, Khalid M, et al. Missense Mutations in MLH1, MSH2, KRAS, and APC Genes in Colorectal Cancer Patients in Malaysia. *Dig Dis Sci* 2012; 57 (11): 2863-2872.
22. Altschul SF, Gish W, Miller W, et al. Basic local alignment search tool. *J Mol Biol* 1990; 215(3): 403-410.
23. S. Andrews. FastQC: A quality control tool for high throughput sequence data. Babraham Bioinforma. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/>.
24. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinforma*. 2010; no. SUPP.32.
25. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; 25 (16): 2078-2079.
26. Li H and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010; 26 (5): 589-595.
27. Van der Auwera GA, Carneiro MO, Hartl C, et al. GATK Best Practices. *Curr. Protoc. Bioinformatics* 2002; 11(1110): 11.10.1-11.10.33.
28. Wang K, Li M, and Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010; 38 (16): 164.
29. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature* 2015; 526 (7571): 68-74.
30. Sherry ST. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001; 29 (1): 308-311.

31. Landrum MJ, Lee JM, Benson M, et al. ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* 2016; 44: D862-D868.
32. Forbes SA, Beare D, Gunasekaran P, Leung K, et al. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015; 43: D805-D811.
33. Parikh R, Mathai A, Parikh S, Chandra SG, et al. Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol* 2008; 56(1): 45-50.
34. Palmquist AEL, Koehly LM, Peterson SK, et al. The Cancer Bond. Exploring the Formation of Cancer Risk Perception in Families with Lynch Syndrome. *J Genet Couns* 2010; 19 (5): 473-486, 2010.
35. Hegde M, Ferber M, Mao R, et al. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). *Genet Med* 2014; 16(1): 101-116.
36. Domingo E, Laiho P, Ollikainen M, et al. BRAF screening as a low-cost effective strategy for simplifying HNPCC genetic testing. *J Med Genet* 2004; 41(9): 664-668.
37. Loman NJ, Misra RV, Dallman TJ, et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 2012; 30(5): 434-439.
38. Pereira FL, Soares SC, Dorella FA, et al. Evaluating the efficacy of the new Ion PGM Hi-Q Sequencing Kit applied to bacterial genomes. *Genomics* 2016; 107(5): 189-198.
39. Churchill JD, King JL, Chakraborty R, et al. Effects of the Ion PGMTM Hi-Q™ sequencing chemistry on sequence data quality. *Int J Legal Med* 2016; 130(5): 1169-1180.
40. Barrow E, Alduaij W, Robinson L, et al. Colorectal cancer in HNPCC: Cumulative lifetime incidence, survival and tumour distribution. A report of 121 families with proven mutations. *Clin Genet* 2008; 74(4): 233-242.
41. Choi YH, Cotterchio M, McKeown-Eyssen G, et al. Penetrance of colorectal cancer among MLH1/MSH2 carriers participating in the colorectal cancer familial registry in Ontario. *Hered Cancer Clin Pract* 2009; 7(1): 14.
42. Belot A, Grosclaude P, Bossard N, et al. Cancer incidence and mortality in France over the period 1980-2005. *Rev Epidemiol Sante Publique* 2008; 56(3): 159-175.
43. Meyer JE, Narang T, Schnoll-Sussman FH, et al. Increasing incidence of rectal cancer in patients aged younger than 40 years: An analysis of the surveillance, epidemiology, and end results database. *Cancer*. 2010; 116(18): 4354-4359.



44. Vasen HF, Watson P, Mecklin JP, et al. New clinical criteria for hereditary non-polyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* 1999; 116(6): 1453-1456.
45. Poulogiannis G, Frayling IM, and Arends MJ. DNA mismatch repair deficiency in sporadic colorectal cancer and Lynch syndrome. *Histopathology* 2010; 56(2): 167-169.
46. Haraldsdottir S, Hampel H, Tomsic J, et al. Colon and endometrial cancers with mismatch repair deficiency can arise from somatic rather than germline mutations. *Gastroenterology* 2014; 147 (6): 1308–1316.



Author(s) shall retain the copyright of their work and grant the Journal/Publisher right for the first publication with the work simultaneously licensed under:

Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). This license allows for the copying, distribution and transmission of the work, provided the correct attribution of the original creator is stated. Adaptation and remixing are also permitted.