

Faculty of Medicine and Health Sciences
School of Medicine

DEVELOPMENT OF MOLECULAR DIAGNOSTIC ASSAYS FOR GASTRIC AND COLORECTAL CANCERS MANAGEMENT

By

Henry Okuchukwu Ebili
(MBBS, MSc, FMCPath)

Thesis submitted to the University of Nottingham

for the degree of Doctor of Philosophy

August 2021

Declaration

This thesis entitled “Development of molecular diagnostic assays of clinical utility for gastric and colorectal cancers management” is the result of my own investigations and observations, unless otherwise stated.

Acknowledgements

I wish to acknowledge the support and supervision rendered by my Supervisor, Prof Emad Rakha during this PhD study. I would also like to thank Prof Mohammed Ilyas for his supervision of the earlier part of this study. My love also goes to my friends and colleagues in the Division of Cancer and Stem Cells, specifically the members of the Molecular Pathology Research Group and the Breast Cancer Research Team, University of Nottingham, for their support and kind words during the stressful periods of the study. I wish to specifically mention Dr Wakkas Mohammed Fadhil who offered his friendship and invaluable assistance to me during the period of my study. I also want to thank Drs Fei Sang and Victoria Wright, of the DeepSeq Nottingham, who graciously gave their assistance in acquiring the genomic data used in chapter six of this thesis. My special thanks go to Prof A.O.J. Agboola of the Department of Morbid Anatomy and Histopathology, Olabisi Onabanjo University, Nigeria for his active support and guidance during the entire period of the study. Finally, I would like to thank my sponsor, the Olabisi Onabanjo University for funding my PhD study.

Dedication

This work is dedicated to my eternally patient wife, Nike, and my children George, Vanessa, and Elizabeth, for the unconditional love, support and sacrifices they have given over these years to enable me to complete the PhD study.

Abstract

The primary need for molecular diagnostic tests to achieve the utmost of performance indices, the rapid development of new targeted therapy, the emergence of new diagnostic histopathology entities requiring differentiation from similar disease entities, and the increasing cancer patient population have necessitated the continuous development of molecular diagnostic assays to meet the challenges and needs of cancer management.

In this thesis I present four (4) molecular diagnostics assays which I developed with the aim of improving certain aspects of their performance indices: Squirrel PCR, mHRMA, MSI-HRMA and the MSI-WES assays. The general hypothesis for this study is that the performance indices of a molecular diagnostics test can be improved by altering certain components of the test or by simply using a different test or detection platform.

The Squirrel PCR assay was developed with the objective of directly sequencing ultrashort DNA segments by Sanger method. The assay resolves the challenges associated with direct sequencing of sub-100bp DNA segments: poor quality sequencing and inaccurate mutation calling. It uses Squirrel primers to elongate - by PCR - the amplicons of ultrashort DNA molecules, thereby enabling good-quality and accurate sequencing by Sanger method. The preliminary data obtained for the Squirrel PCR assay showed that the assay enabled the achievement of 100% genomic and target sequence qualities by Sanger method compared to the design sequences, and 100% accuracy of mutation detection by Sanger method in standard samples. The Squirrel PCR assay allows the circumvention of the use of next-generation sequencing (NGS), pyrosequencing, and sequencing-by-cloning methods, the current laborious, time-consuming, and expensive methods of sequencing short DNA segments. The Squirrel PCR assay therefore would improve the turnaround time and cost-effectiveness of sequencing short DNA molecules and finds application

in small sequencing projects in which the available DNA templates are naturally degenerate or degraded.

The consequence of the rapid development of targeted therapy for cancer management is that more genetic targets need to be tested in each patient's tumour for evidence of predictive variants. For most molecular diagnostics laboratories where limited testing is the routine it is impractical to acquire expensive high throughput mutation detection platforms, and small improvements in performance (in this case, the speed of testing) of existing assays may be all that is required. The mHRMA assay is a high-resolution melting analysis-based assay which combines the exquisite sensitivity of high-resolution melting analysis (HRMA) and the developed capacity to multiplex targets in the HRMA platform, to achieve rapid detection of mutations in colorectal cancer (CRC). The mHRMA assay utilises differential tagging of amplicons with GC-tails – by using GC-tagged primers - to achieve adequate separation of the melting temperatures of amplicons and enable robust duplexing and triplexing of targets. The capacity for multiplexing in the HRMA detection platform makes mHRMA a faster, more cost-effective, and less laborious mutation detection assay than the conventional, single target HRMA. In this development, preliminary data showed that the mHRMA assay can achieve a perfect concordance and 100% sensitivity, specificity, and accuracy compared with the single target HRMA tests.

Microsatellite instability (MSI) testing recently achieved additional relevance in cancer management because of its newly discovered predictive value in the use of the highly effective immune checkpoint inhibitors. The implication of its new relevance means that more cancer patients would undergo MSI testing to determine their suitability for immunotherapy. Thus, the need for a faster, less laborious, and probably more cost-effective MSI testing method. The MSI-

HRMA assay tests MSI in CRC using a combination of established (BAT-25, BAT-26, MONO-27) and novel (TYMS-26, BCAT-25, MYB, ANGEL2, TP53 (BAT34CA), FBXO46 and TCF4) microsatellite markers in the HRMA detection platform. The selected panels of the MSI-HRMA markers achieved sensitivities, specificities, and accuracies of 91.3-100%, 100% and 100%, respectively, compared to the gold standards. However, because of its detection platform MSI-HRMA is easier, faster, and more cost-effective to set up and run than either capillary electrophoresis (CE) or immunohistochemistry (IHC). Also, as shown by the mHRMA assay, the MSI-HRMA *can be* multiplexed to further reduce the labour, cost and turnaround time associated with MSI testing.

One advantage of using NGS for mutation detection is that huge numbers of variants, including clinically relevant single nucleotide variants (SNVs) as well as microsatellite-based indels, can be tested in single NGS experiments, thereby obviating the necessity for separate experiments for SNV detection and for MSI testing. The MSI-WES was designed as a simple strategy for testing MSI in gastric cancer (GCa) using level 2 whole exome sequencing (WES) data as recorded in variant call format (vcf) files. MSI-WES simply counts the insertions and deletions (indels) occurring within microsatellite contexts and uses the count-based indices as MSI scores. Based on a well-defined threshold of MSI scores MSI-WES robustly classified each GCa case into MSI or MSS. MSI-WES achieved a near-perfect concordance with the PCR-based MSI test, and sensitivity, specificity, and accuracy of 100%, 99.03% and 99.22%, respectively. The attraction for MSI-WES is its simplicity of MSI-testing strategy devoid of complex statistical permutations. Furthermore, MSI testing is performed on the same processed vcf files from which other clinically relevant mutation types are extracted.

Each of the four assays were validated to variable extents as discussed in the thesis.

PhD study-related publications, conference presentations and awards

Publications

1. **Ebili HO**, Agboola AOJ, Rakha E. MSI-WES: a simple approach for microsatellite instability testing using whole exome sequencing. Future Oncol. 2021 doi: 10.2217/fon-2021-0132 (in press).
2. Ham-Karim HA, **Ebili HO**, Manger K, Fadhil W, Ahmad NS, Richman SD, Ilyas M. Targeted Next-Generation Sequencing Validates the Use of Diagnostic Biopsies as a Suitable Alternative to Resection Material for Mutation Screening in Colorectal Cancer. Mol Diagn Ther. 2019 Jun;23(3):383-393. doi: 10.1007/s40291-019-00388-z. PMID: 30741375.
3. Ham-Karim HA, **Ebili HO**, Bradshaw K, Richman SD, Fadhil W, Domingo E, Tomlinson I, Ilyas M. Targeted next generation sequencing reveals a common genetic pathway for colorectal cancers with chromosomal instability and those with microsatellite and chromosome stability. Pathol Res Pract. 2019 Jul;215(7):152445. doi: 10.1016/j.prp.2019.152445. Epub 2019 May 20. PMID: 31153694.
4. Susanti S, Fadhil W, Murtaza S, Hassall JC, **Ebili HO**, Oniscu A, Ilyas M. Positive association of *PIK3CA* mutation with *KRAS* mutation but not *BRAF* mutation in colorectal cancer suggests co-selection is gene-specific but not pathway-specific. J Clin Pathol. 2019 Mar;72(3):263-264. doi: 10.1136/jclinpath-2018-205483. Epub 2018 Oct 25. PMID: 30361395.
5. Susanti, S., Fadhil, W., **Ebili, H.O.**, Asiri, A., Nestarenkaite, A., Hadjimichael, E., Ham-Karim, H.A., Field, J., Stafford, K., Matharoo-Ball, B., Hassall, J.C., Sharif, A., Oniscu,

A., Ilyas, M. (2018). N_LyST: A simple and rapid test for Lynch Syndrome. Journal of Clinical Pathology; 71(8). DOI: 10.1136/jclinpath-2018-205013

6. **Ebili, H.O.**, Hassall, J.C., Fadhil, W., Ham-Karim, H. A., Asiri, A., Agboola, A.O.J., Ilyas M. (2017). QMC-PCRx: a novel method for rapid mutation detection. Journal of Clinical Pathology; 70 (8):702-711. DOI: 10.1136/jclinpath-2016-204264. Epub 2017 Feb 2
7. Raposo, T.P., **Ebili, H.O.**, Ilyas, M. (2017). A refined method to study gene dosage changes in-vitro using CRISPR/Cas9. Journal of Clinical Pathology; 0:1-8. DOI:10.1136/jclinpath-2017-204558
8. **Ebili, H.O.**, Hassall, J.C., Fadhil, W., Ham-Karim, H., Asiri, A., Raposo, T.P., Agboola, A.O.J., Ilyas, M. (2017) "Squirrel" primer-based PCR assay for direct and targeted Sanger-sequencing of short genomic segments. Journal of Biomolecular Techniques; 28(3). DOI: 10.7171/JBT
9. Ham-Karim H., **Ebili H.O.**, Fadhil W., Asiri A., Hassall J., Ilyas M. (2017). COLD-HRM: A Combination of Methods to Infer the Nature of Somatic Mutations. Advances in Cytology and Pathology; 2(2): 00017. DOI: 10.15406/acp.2017.02.00017

Conference Proceedings

1. **Ebili, H**., Fadhil, W., Dalleywater, W.J., Ilyas, M. Validating a tissue microarray linked colorectal cancer cohort [poster abstract]. In: *Leeds Pathology 2019*. 12th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 2–4 July 2019. Poster abstract nr 153. https://path.org.uk/wp-content/uploads/2019/09/714463-PSoc-Path19-A4-ABST-rgb-3-POST-v12w.pdf

2. Dalleywater, W.J., Fadhil, W., **Ebili, H.,** Ilyas, M. Stage is Not Predictive of Time to Recurrence in Colorectal Cancer [poster abstract]. In: *Leeds Pathology 2019*. 12th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 2–4 July 2019. Poster abstract nr 96. https://path.org.uk/wp-content/uploads/2019/09/714463-PSoc-Path19-A4-ABST-rgb-3-POST-v12w.pdf

3. Susanti, S., Fadhil, W., **Ebili, H.O**., Asiri, A., Nestarenkaite, A., et al. N_LyST: A simple and rapid screening test for Lynch Syndrome [oral abstract]. In: *Maastricht Pathology, 2018*. 11th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 19th – 22nd June 2018. Poster Abstract nr 69. www.path.org.uk/wp-content/uploads/2018/06/Poster-Abstracts_Final.pdf

4. Hassall, J.C., **Ebili, H.O**., Ilyas, M. An Exquisitely Sensitive Method for Mutation Detection [oral abstract]. In: *Maastricht Pathology, 2018*. 11th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 19th – 22nd June 2018. Oral Abstract nr 8. https://www.path.org.uk/wp-content/uploads/2018/08/Oral-Plenary-Oral-Abstracts_Final.pdf

5. Murtaza, S., Fadhil, W.M., Hassall, J., **Ebili, H.O.**, Ilyas, M. Testing for PIK3CA and BRAF mutations in colorectal cancer with microsatellite instability [oral abstract]. In: In: *Maastricht Pathology, 2018*. 11th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain &

Ireland, 19th – 22nd June 2018. Poster Abstract nr 71. www.path.org.uk/wp-content/uploads/2018/06/Poster-Abstracts_Final.pdf

6. Fadhil, W., Susanti, S., **Ebili, H.O**., Field, J., Stafford, K., Hadjimichael, H., Ilyas, M. Introducing a new panel of markers for microsatellite instability (MSI) detection in colorectal cancer by high resolution melting analysis [oral abstract]. In: *Belfast Pathology 2017*. 10th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 20-23 June, 2017. Oral Abstract nr 08. www.path.org.uk/wp-content/uploads/2017/06/BP2017-PLENARY-ORAL-ABSTRACTS-05.06.17.pdf

7. Asiri, A.A., Hassall, J.H., Ham-Karim, H.H.K., **Ebili H.**, Fadhil, W.F., Ng, O.N., Fardouly, P.F., Acheson, A.A., Ilyas, M.I. Detection of mutant circulating free DNA as a biomarker of colorectal cancer [poster abstract]. In: *Belfast Pathology 2017*. 10th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 20-23 June, 2017. Abstract nr 30. www.path.org.uk/wp-content/uploads/2017/06/BP2017-POSTER-ABSTRACTS-16.06.17.pdf

8. **Ebili, H.O**., Hassall, J., Asiri, A., Ham-Karim, H.A., Ilyas, M. A novel method for screening for multiple targets in tumours using high resolution melting [poster abstract]. In: *Nottingham Pathology 2016*. 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 28 June – 1 July 2016. Abstract nr 126. (Also published in *The Journal of Pathology*. 2016;240:S1-S48). http://onlinelibrary.wiley.com/doi/10.1002/path.2016.240.issue-S1/issuetoc

9. Ham-Karim, H.A., **Ebili, H.O.**, Asiri, A., Hassall, J., Ilyas, M. Aberrant high resolution melting pattern: germline variant or somatic mutation? [Poster abstract]. In: *Nottingham Pathology 2016*. 9th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain &

Ireland, 28 June – 1 July 2016. Abstract nr 128. (Also published in *The Journal of Pathology. 2016*; 240: S1-S48). http://onlinelibrary.wiley.com/doi/10.1002/path.2016.240.issue-S1/issuetoc

10. Asiri, A., Ham-Karim, H., Ng, O, Hassall, J., Fadhil, W., Acheson, A., Ilyas, M, **Ebili, H**. Circulating free DNA in colorectal cancer as a marker of surgical clearance [Poster abstract]. In: *BASO – NCRI Joint Meeting Abstracts. European Journal of Surgical Oncology*, 2016; 42(11): S242-243. Abstract nr 323. https://doi.org/10.1016/j.ejso.2016.07.100

Awards

1. 2nd poster prize, Maastricht Pathology 2018, 11th Joint Meeting of the British Division of the International Academy of Pathology and the Pathological Society of Great Britain & Ireland, 19-22 June 2018, Maastricht, Netherland.
2. Nigerian Tertiary Education Trust Fund Scholarship for a PhD in Pathology at the University of Nottingham: 2015 – 2020.

Patent

1. Ilyas M, Hassall JC, Ebili HO (2019). Detection of nucleic acid sequences. WO/2019/145734; PTC/GB2019/050221.

Table of Contents

Declaration……..i

Acknowledgement…….....ii

Dedication…….iii

Abstract…….iv

PhD study-related publications, conference proceedings, awards, and patent……vii

Table of Contents……xii

List of Tables…….xvii

List of Figures……... xix

List of Abbreviations……xxi

1. Chapter 1: General Introduction……...1

1.1. Colorectal Cancer……2

1.1.1. Epidemiology of CRC……...2

1.1.2. Clinical features of CRC……4

1.1.3. Molecular Pathogenesis of CRC……5

1.2. Gastric Cancer……14

1.2.1. Epidemiology of GC……14

1.2.2. Clinical Features of GC……16

1.2.3. Molecular Pathogenesis of GC……17

1.3. The Role of Diagnostic Molecular Pathology in Oncology……23

1.3.1. Molecular Pathology in Cancer Diagnosis……26

1.3.2. Molecular Pathology for Prediction of Therapeutic Response…….29

1.3.3. Molecular Pathology for Prediction of Cancer Recurrence……..35

1.3.4. Molecular Alterations of Clinical Importance……..37

1.4. Development of Molecular Diagnostic Assays……..40

1.5. Study Hypothesis and Aims……44

2. Chapter 2: Materials and Methods……46
2.1. Materials……47
2.1.1. CRC Cell Lines……..47
2.1.2. CRC Clinical Cohorts……47
2.1.3. Genomic Data for GC Cohort……48
2.2. Methods……49
2.2.1. DNA Extraction……..49
2.2.2. Primer Design……51
2.2.3. Polymerase Chain Reaction……52
2.2.4. QMC-PCR……..53
2.2.5. High Resolution Melting Analysis……53
2.2.6. Primer Optimisation……54
2.2.7. Primer Validation……59
2.2.8. Immunohistochemistry……63
2.2.8.1. Tissue Microarray Preparation……..63
2.2.8.2. IHC Protocol……..64
2.2.8.3. Whole Slide Imaging……....65
2.2.9. Capillary Electrophoresis……66
2.2.10. Sanger Sequencing……...67
2.2.11. Genomic Data Analysis……67
2.2.11.1. Annotated Somatic Mutation Data……67
2.2.11.2. Gene Expression Analysis……70
2.2.11.3. DNA Methylation Analysis……...70
2.2.12. Statistical Analysis……..72
2.3. Ethical approval……...73

3. Chapter 3: Squirrel PCR Assay for Robust Sequencing of Ultra-short DNA Segments……...……74
3.1. Background……..75
3.2. Materials and Methods……....78
3.2.1. Principles/Concepts……..78
3.2.2. Squirrel Primer Design……79
3.2.3. Sequence Analysis……84
3.3. Results……85
3.3.1. Squirrel PCR Assay versus Binladen et al Strategy……86
3.3.2. Target Sequence Quality……...89
3.3.3. Genomic Sequence Quality……..89
3.3.4. Squirrel Primer Tail Length versus Genomic and Target Sequence Qualities….90
3.3.5. Analytical Performance of the Squirrel PCR Assay…….93
3.4. Discussion……100
4. Chapter 4: mHRMA – Multiplex HRMA for Rapid Mutation Detection……104
4.1. Background……...105
4.2. Materials and Methods…….....111
4.2.1. Principles of mHRMA…….....111
4.2.2. Materials……..111
4.2.3. mHRMA and Tm Peaks Separation……111
4.2.4. Separation of Target Tm Peaks Using GC-tagged Primers…….....112
4.2.5. Validation of mHRMA Assay……...……..113
4.2.6. Statistical Analysis……..115
4.3. Results……...116
4.3.1. Tm Peak Intersection Produces Spurious Melting Patterns……116
4.3.2. GC-tagged Primers Enhance Separation of Multiple Tm Peaks for Robust mHRMA ……..116
4.4. Discussion……124

5. Chapter 5: MSI-HRMA for microsatellite instability testing in CRC……128
5.1. Background……129
5.2. Materials and Methods……132
5.2.1. Materials……132
5.2.2. Microsatellite Markers Identification and Selection……135
5.2.3. Primer Design, Primer Optimisation and Validation and PCR Strategy…….137
5.2.4. Cohort Testing…….....137
5.2.5. Combination of Microsatellite Markers in Panels……138
5.2.6. Validation of MSI-HRMA Assay……139
5.2.7. Statistical Analysis……...139
5.3. Results……140
5.3.1. QMC-PCR versus single-stage PCR……140
5.3.2. Training Cohort……140
5.3.3. Validation Cohort……145
5.4. Discussion……...148
6. Chapter 6: MSI-WES: a simple approach for microsatellite instability testing using whole exome sequencing……152
6.1. Background……153
6.2. Materials and Methods……156
6.2.1. Study Cohort……156
6.2.2. Genomic Data Analyses……158
6.2.3. Quality Control of Somatic Mutation Data……...159
6.2.4. Study Approach and Principles……159
6.2.5. Validation of NGS-based MSI indices……..161
6.2.6. Statistical Analysis……162
6.3. Results……163
6.3.1. Confirmation of microsatellite Status of Small Indels……..163
6.3.2. NGS-based MSI Indices Determination……163

6.3.3. Threshold for Dichotomising MSI Index_ngs…….……163
6.3.4. Validation of MSI Indices……166
6.3.4.1 MMR gene expression……..166
6.3.4.2 The MTM index……167
6.3.4.3 Performance tests: MSI status_pcr versus MSI status_ngs……...167
6.4. Discussion……...170
7. Chapter 7: General Discussion……175
7.1. Introduction……176
7.2. The Squirrel PCR Assay……176
7.3. HRMA-based Assays……..178
7.3.1. mHRMA Assay……180
7.3.2. MSI-HRMA Assay……182
7.4. MSI-WES……184
7.5. The Extent of Validations Achieved for the Assays……186
References……..190
Appendix……235
Appendix 1: Comparisons of cost estimates of deploying new assays versus standard assays……..235
Appendix 2: Comparison of turnaround time estimates for deploying new assays versus standard assays……..241

List of Tables

Table 1.1: Molecular subtypes of gastric cancer by Lei et al., 2013……21

Table 3.1: Squirrel and USeq primers characteristics……...82

Table 3.2: Amplicon, genomic and target sequence sizes……83

Table 3.3: Squirrel PCR assay versus Binladen et al sequencing strategy……87

Table 3.4: Sequencing quality of the target sequences……91

Table 3.5: Sequencing quality of the genomic sequences……92

Table 3.6: KRAS exon 2 mutation statuses in cell lines……...95

Table 3.7: HRMA patterns of CRC samples at mutation hotspots……...95

Table 3.8: Confirmation of HRMA results by Squirrel PCR-Sanger sequencing……96

Table 4.1: List of primers and their product Tm's……..114

Table 4.2: mHRMA versus QMC PCR……..123

Table 5.1: Characteristics of genomic markers selected for MSI testing……136

Table 5.2: MSI testing of the training cohort by selected microsatellite markers……..142

Table 5.3: Analytical performance of the eleven microsatellite markers……...143

Table 5.4: Concordance tests for microsatellite markers panels……144

Table 5.5: AUROC analysis of the microsatellite marker panels……...144

Table 5.6: Analytical performance of the microsatellite marker panels……144

Table 5.7: MSI testing of the validation cohort by MSI-HRMA validation panel……145

Table 6.1: DNA mismatch repair gene expression and MSI indices……..166

Table 6.2: Relationship between MTM index and the MSI indices……167

Table 6.3: Relationship among the MLH1 methylation probes……..168

Table 6.4: Correlations of MLH1 methylation……169

Appendix Supplementary Table 1: Comparison of cost estimates of designed assays versus standard tests ……..238

Appendix Supplementary Table 2: Comparison of turnaround times for designed assays versus standard tests……243

Appendix Table 5.1: Features of MSI-HRMA primers……..246

Appendix Table 5.2: Concordance between QMC-PCR and single stage PCR for MSI-HRMA……...248

Appendix Table 5.3: MSI testing by MSI-HRMA Panel 1……249

Appendix Table 5.4: MSI testing by MSI-HRMA Panel 2……250

Appendix Table 5.5: MSI testing by MSI-HRMA Panel 3……251

Appendix Table 5.6: MSI testing by MSI-HRMA Panel 4……252

Appendix Table 5.7: MSI testing by MSI-HRMA Panel 5……253

Appendix Table 5.8: MSI testing by MSI-HRMA Panel 6……254

Appendix Table 6.1: Genomic coordinates of the mutational targets of MSI……255

Appendix Table 6.2: ENSEMBL gene identifier for MMR genes……255

Appendix Table 6.3: MSI and MTM indices……..256

Appendix Table 6.4: DNA mismatch repair gene expression in FPKM……267

List of Figures

Figure 1.1: Molecular pathogenesis and morphological evolution of sporadic CRC……13

Figure 1.2: Summary molecular classifications of gastric cancer……..20

Figure 1.3: Algorithm for molecular classification of gastric cancer by Setia et al., 2016……22

Figure 2.1: Optimsation of the Squirrel primers……...56

Figure 2.2: Gradient PCR-HRMA of the mHRMA assay primers……57

Figure 2.3: Gradient PCR-HRMA of some of the the MSI-HRMA assay primers……58

Figure 2.4: Squirrel primer validation……60

Figure 2.5: mHRMA assay primers validation……..61

Figure 2.6: MSI-HRMA primers validation……...80

Figure 3.1. Squirrel PCR assay strategy……80

Figure 3.2: Electropherogram of PTEN exon 3 sequence……..86

Figure 3.3: Squirrel PCR vs Binladen et al strategy……...88

Figure 3.4: Sanger sequencing by the Squirrel PCR strategy –
KRAS2Seq vs KRAS2LSeq……93

Figure 3.5: Alignment of sequencing data with the GenBank Homo sapiens
GRCh38.p7 Primary Assembly……..94

Figure 3.6: KRAS codons 12 and 13 mutation detection
by Squirrel PCR assay……..……..97

Figure 3.7: PIK3CA, BRAF and PTEN mutation detection
by Squirrel PCR-based Sanger method……...98

Figure 3.8: KRAS codons 61 mutation detection
by Squirrel PCR-enabled Sanger method……..….99

Figure 4.1: Singleplex or single target HRMA……110

Figure 4.2: HRM analysis of duplex targets TP53 E8b and TP53 E4d……...117

Figure 4.3: HRM analysis of duplex targets SMAD4 E9a and TP53 E4d……..118

Figure 4.4: Demonstrations of GC-tagged primers' capacity
to achieve adequate separation of Tm's……120

Figure 4.5: mHRMA of PTEN E3/PIK3CA E20/KRAS E3 triplex……121

Figure 4.6: mHRMA of PTEN E7/ NRAS E2 duplex…….122

Figure 5.1: Capillary electropherogram showing MSI status by fragment analysis of mononucleotide markers……..133

Figure 5.2: Immunohistochemistry of expression of four mismatch repair (MMR) proteins……..134

Figure 5.3: HRMA of the microsatellite markers in the training cohort of CRC……143

Figure 6.1: Re-classification of the MSI status of gastric cancer using the MSI-versus-MSS scheme…….158

Figure 6.2: A list of potential microsatellite-based indels in variant call formats……...160

Figure 6.3: Scatter plots showing the relationship between MSI index_ngs and the PCR-based MSI status…….165

List of Abbreviations

AAPC Attenuated adenomatous polyposis coli

AKT AKR mouse Thymoma

APC Adenomatous polyposis coli

ARID1A AT-Rich Interaction Domain 1A

ASR Age-standardised rates

ATM Ataxia Telangectasia Mutated

ATP4A H+/K+ ATPase Alpha Subunit

ATP4B ATPase subunit beta

ATR Ataxia Telangectasia and Rad3 Related

BARD1 BRCA1-associated RING domain protein-1

BAT25 Big Adenine Tract 25

BAT26 Big Adenine Tract 26

BAT40 Big Adenine Tract 40

BAX BCL2 Associated X

BCAT25 Beta Catenin 25

BRAF v-Raf murine sarcoma viral oncogene homolog B

BRCA1 Breast Cancer 1

BRCA2 Breast Cancer 2

BRIP1 BRCA1 Interaction protein 1

CASP5 Caspase 5

CCND1 Cyclin D1

CCND2 Cyclin D2

CCNE1 Cyclin E1

CDH1 Cadherin 1

CDH2 Cadherin 2

CDK1A Cyclin-dependent kinase inhibitor 1A

CDK4 Cyclin-dependent kinase 4

CDKN2A Cyclin-dependent kinase inhibitor 2A

CDKN2B Cyclin-dependent kinase inhibitor 2B

CE Capillary electrophoresis

CGH Comparative Genome Hybridisation

CHFR Checkpoint with Forkhead and Ring Finger Domains

CHGA Chromogranin A

CIMP CpG Island Methylation Pathway

CIN Chromosomal Instability

CMMRD Constitutional Mismatch Repair Deficiency

COX-2 Cyclooxygenase-2

CRAC1/HMPS Cartilage acidic protein 1/ Hereditary mixed polyposis syndrome

CRC Colorectal Cancer

CTNNB1 Beta Catennin

DAB 3,3'-Diaminobenzidine

DNA Deoxyribonucleic Acid

EBV Epstein-Barr Virus

ECM Extracellular Matrix

ECOG2197 Eastern Cooperative Oncology Group

EGF Epidermal growth factor

EGFR Epidermal growth factor receptor

EMT Epithelial Mesenchymal Transition

EPCAM Epithelial Cell Adhesion Molecule

ERG ETS-related gene

EWSR1 Ewing Sarcoma Breakpoint Region 1

FAP Familial Adenomatous Polyposis

FBXO46 F-Box protein 46

FCCTX Familial CRC type X

FGFR2 Fibroblast Growth Factor Receptor 2

FISH Fluorescent In Situ Hybridisation

FKHR Forkhead in rhabdomyosarcoma

FLI1 Friend leukemia integration 1 transcription factor

FOLFOXIRI Folinic acid, 5-fluorouracil, oxaliplatin and irinotecan

GC Gastric Cancer

GDC Genome Data Commons

GIF Gastric Intrinsic Factor

GIST Gastrointestinal Stromal Tumour

GLOBOCAN Global Cancer Incidence, Mortality and Prevalence

HDGC Hereditary Diffuse Gastric Cancer

HNPCC Hereditary Non-Polyposis Colon Cancer

HRAS Harvey rat sarcoma viral oncogene homolog

HRMA High Resolution Melting Analysis

IARC International Agency for Research on Cancer

ICGC International Cancer Genome Consortium

IGFIIR Insulin-like growth Factor 2 Receptor

IHC Immunohistochemistry

JAK Janus kinase 2

KRAS Kirsten rat sarcoma viral oncogene homolog

LOH Loss of heterozygosity

LS Lynch syndrome

MAP MUTYH-Associated Polyposis

MAPK mitogen-activated protein kinase

MBD4 Methyl-CpG-binding domain protein 4

MDM2 Mouse double minute 2 homolog

MET Mesenchymal Epithelial Transition

MGMT O-6-methylguanine DNA methyltransferase

MLH1 MutL homolog 1

MLH3 MutL homolog 3

MLPA Multiplex Ligation-dependent Probe Amplification

MMR Mismatch Repair

MPNST Malignant Peripheral Nerve Sheath Tumour

MSH2 MutS homolog 2

MSH3 MutS Homolog 3

MSH6 MutS Homolog 6

MSI Microsatellite Instability

MS-PCR Methylation-Sensitive PCR

MSS Microsatellite Stable

mTOR Mechanistic Target of Rapamycin

MUC5AC Mucin 5AC

MUC6 Mucin 6

MUTYH MutY homolog

MYB V-myb avian myeloblastosis viral oncogene

MYC Myelocytomatosis oncogene

NAP NTHL1-Associated Polyposis

NCCN National Comprehensive Cancer Network

NFKB Nuclear Factor Kappa B

NGS Next generation Sequencing

NICE National Institute for Health and Care Excellencce

NRAS Neuroblastoma RAS viral oncogene homolog

NSCLC Non-Small Cell Lung Cancer

PAX3 Paired Box 3

PAX7 Paired Box 7

PAX8 Paired Box 8

PDE4D Phosphodiesterase D

PDGFRA Platelet-Derived Growth Factor Receptor Alpha

PIK3CA Phosphatidylinositol 3-Kinase, Catalytic

PKC Protein Kinase C

PLCγ Phospholipase C, gamma 1

PMS2 PostMeiotic Segregation increased 2

PNET Primitive Neuroectodermal Tumour

POLD1 Polymerase Delta 1

POLE Polymerase Epsilon

PPARD Peroxisome proliferator-activated Receptor Delta

PPARγ Peroxisome proliferator-activated Receptor Gamma

PTC Papillary Thyroid carcinoma oncogene

PTEN Phosphatase and Tensin homolog

PTPRD Protein Tyrosine Phosphatase Receptor D

QMC-PCR Quick Multiplex Consensus PCR

qPCR Quantitative PCR

RAD50 RAD50 homolog

RARβ Retinoic acid receptor beta

RAS RAS viral oncogene homolog

RB1 Retinoblastoma 1

RET Rearranged during Transfection

RIZ Retinoblastoma Interacting Zinc finger

RNASeq RNA sequencing

RT-PCR Reverse transcription PCR

RUNX3 Runt-related transcription factor 3

SHH Sonic Hedgehog

SMAD4 SMAD family member 4

SMAD7 SMAD family member 7

SRC V-src sarcoma viral oncogene

SSX1 Synovial Sarcoma, X breakpoint 1

SSX2 Synovial Sarcoma, X breakpoint 2

STAT Signal Transducer and Activator of Transcription

STK11/LKB1 Serine/Threonine Kinase 11/Liver Kinase 1

SWOG SouthWest Oncology Group

SYT Synaptotagmin 1

TBX3 T-box 3

TCF4 Transcription factor 4

TCGA The Cancer Genome Atlas

TFF2 Trefoil Factor 2

TGFB Transforming Growth Factor Beta

TGFBRII Transforming Growth Factor Beta Receptor 2

TGFα Transforming Growth Factor Alpha

TILs Tumour-Infiltrating Lymphocytes

TIMP-3 Tissue Inhibitor of Metalloproteinase 3

TMA Tissue Microarray Array

CHAPTER 1

GENERAL INTRODUCTION

1.1 Colorectal Cancer

1.1.1 Epidemiology of colorectal cancer

According to the 2020 GLOBOCAN statistics, colorectal cancer (CRC) is the 3rd most common malignancy in the world after female breast and lung cancers, accounting for over 1.9 million new cases (9.8% of all malignancies) worldwide (Sung et al., 2021, Ferlay et al., 2021). It is also the 2nd most common cause of cancer deaths with 915,880 new cases (9.2% of all malignancies) in 2020 (Ferlay et al., 2021; Sung et al., 2021). In the male gender, it is the 3rd commonest cancer in 2020 after lung and prostate cancers, and the 3rd most common cause of cancer deaths after lung and liver cancers (Sung et al., 2021; Ferlay et al., 2021). Among the female gender CRC ranks 2nd in the most frequent malignancies after breast cancer, as well as the 3rd most common cause of cancer deaths after breast and lung cancers (Sung et al., 2021; Ferlay et al., 2021). CRC shows a very wide variation in geographical incidence. Its incidence is highest in Europe, North America, Australia/New Zealand, and Eastern Asia where Age-standardised rates (ASR) incidence of 25.9 – 33.6 per 100,000 population were recorded, and lowest in African regions, Central America, Melanesia, and South-Central Asia with ASR incidence of 5.5 – 13.7/100,000 population (Sung et al., 2021; Ferlay et al., 2021). The mortality rates generally followed the incidence rate patterns, with the highest ASR mortality of 8.2-11.8/100,000 population observed in the world regions with the highest incidence rates, while the lowest ASR mortality of 3.2-7.2/100,000 population were observed in the regions with the lowest incidence rates (Sung et al., 2021; Ferlay et al., 2021).

Worldwide, CRC showed a male predominance pattern in 2020 – as before - with a male-to-female incidence ratio of 1.23 (Bray et al., 2018; Ferlay et al., 2019; Ferlay et al., 2021; Sung et al., 2021). This male predominance was, however, reversed in regions such as South America,

Caribbean, and Eastern Africa where the female incidence slightly predominated (Ferlay et al., 2021; Sung et al., 2021). Furthermore, CRC showed a predominantly elderly (60 years of age and above) with an elderly-to-young age ratio of 2.64. This pattern of aged predominance is maintained when the CRC population was examined separately for the male and female gender groups (Ferlay et al., 2021; Sung et al., 2021). Worldwide, the peak age incidence of CRC in 2020 was in the 60-69 age group in both genders. The 60-69 years age-group was also the peak age incidence for the male gender (Ferlay et al., 2021; Sung et al., 2021). However, the 70-79 years age group was the peak age incidence for the female gender, being only slightly higher than the 60-69 years age group (Ferlay et al., 2021; Sung et al., 2021). The peak age-group for CRC incidence showed some geographical variation among the 50-59 years, 60-69 years, and 70-79 years age-groups. While the peak age group for CRC incidence for Western Europe, Southern Europe, Northern Europe, Australia and New Zealand and the Caribbean was found to be in the 70-79 years age-group (Ferlay et al., 2021; Sung et al., 2021), the peak age-group for CRC incidence was found to be in the 50-59 age-group for the Western African region (Ferlay et al., 2021; Sung et al., 2021). The peak age-group for the rest of the world was found in the 60-69 age-group (Ferlay et al., 2021; Sung et al., 2021).

The risk factors for the development of CRC can generally be divided into modifiable and non-modifiable (also known as genetic/hereditary) factors. The modifiable adverse risk factors for the development of CRC include obesity, sedentary lifestyle, excessive consumption of diets high in red and processed meat, calorie, haem iron and fat, and low in fibre; high alcohol intake and tobacco smoking (Turner, 2015; American Institute of Cancer Research, 2018; Sullivan et al., 2015). Others are high blood insulin levels and inflammatory bowel diseases (Ulcerative Colitis and Crohn's disease) (Turner, 2015). Conversely, the protective risk factors against CRC include

high consumption of fish and fish oil, fibre-rich diets, fruits and vegetables; high intake of vitamins C, D and calcium; habitual exercise, and regular use of aspirin (Turner, 2015; American Institute of Cancer Research, 2018; Sullivan et al., 2015).

The genetic/hereditary risk factors of CRC will be discussed under the molecular pathogenesis of CRC.

1.1.2 Clinical features of CRC

Three modes of clinical presentation have been recognised for CRC patients. These include (i) asymptomatic presentation in which asymptomatic patients are discovered to have CRC during routine screening, (ii) symptomatic CRC in which patients present with the classic and other symptoms of CRC, (iii) emergency presentation in which patients present in the emergency department with acute symptomatology including intestinal obstruction, acute gastrointestinal haemorrhage, or peritonitis (Turner, 2015; Baer et al., 2017; Moreno et al., 2016).

About 10% of CRC cases are diagnosed during routine screening for CRC- either by screening colonoscopy or faecal occult blood testing, faecal immunochemical test or stool DNA test (Moreno et al., 2016; National Institute for Health and Care Excellence, 2017a; Carethers, 2020). Routine screening programmes have enabled the early detection of CRC in a substantial number of patients who would otherwise have presented at later stages of the disease (Turner, 2015; Moreno et al., 2016). However, about 70-90% of CRC patients present with classic symptoms of CRC, including abdominal pain, passage of melaena, haematochezia, iron-deficiency anaemia, changes in bowel habit, abdominal or rectal mass and weight loss (Turner, 2015; Moreno et al., 2016; Hamilton et al., 2005; Fisher & Daniels, 2007; Cleary et al., 2007; Elzouki et al., 2014; Thompson et al., 2017). CRC patients also commonly present with metastatic disease. The most

common sites of CRC metastases include regional lymph nodes, liver, lungs and the peritoneum via lymphatics, blood vessels as well as transcoelomically (Turner, 2015; Holch et al., 2017; Manfredi et al., 2006; Riihimäki et al., 2016). The clinical presentation of metastatic disease includes right upper quadrant pain, abdominal distention, early satiety, and cervical lymphadenopathy and peri-umbilical nodules, vesico-enteric and entero-enteric fistulae, pyrexia of unknown origin (Belhassen-García et al., 2013; Guend et al., 2015; Khangura et al., 2015; Maslov & Bragg, 2019; Skierucha et al., 2018; Wadhwani & Diwakar, 2018; Xie et al., 2021). About 5% of CRC patients present with acute symptomatology (as mentioned above), necessitating emergency surgery (Baer et al., 2017; Moreno et al., 2016).

1.1.3 Molecular pathogenesis of CRC

CRC is a multistep process involving the inactivation of tumour suppressor genes and DNA repair genes, and simultaneous activation of oncogenes, and leading to stepwise accumulation of mutations and epimutations (Turner, 2015; Arnold et al., 2005; Gonzalez et al., 2017; Yamagishi et al., 2016). CRC generally presents in three molecular pathogenetic patterns: (i) hereditary CRC syndromes which account for about 5% of all CRC cases, (ii) non-syndromic or familial CRC which is responsible for 20-30% of all CRC cases, and (iii) sporadic CRC which make up about 70% of CRC cases (Armelao & de Pretis, 2014; Arnold et al., 2005; Boardman, 2002; Gonzalez et al., 2017; Provenzale et al., 2016; Talseth-Palmer, 2017; Yamagishi et al., 2016).

The most common hereditary CRC syndromes include Lynch (also known as hereditary non-polyposis colon cancer, HNPCC) and the familial adenomatous polyposis (FAP) syndromes (Arnold et al., 2005; Boardman, 2002; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014). Variants of Lynch syndrome include Muir-Torre syndrome, constitutional

mismatch repair deficiency (CMMRD) syndrome, and Turcot syndrome (Gonzalez et al., 2017; Leoz et al., 2015). Variants of the FAP syndrome include Gardner, Turcot and hereditary flat adenoma (or, attenuated adenomatous polyposis coli, AAPC) syndromes (Boardman, 2002; Cattaneo et al., 2007; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Talseth-Palmer, 2017). Less common hereditary CRC syndromes include *MUTYH*-associated polyposis (MAP), and *NTHL1*-associated polyposis (NAP) (Arnold et al., 2005; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Talseth-Palmer, 2017). Other syndromes which frequently present with CRC include Peutz-Jeghers, Juvenile Polyposis, Cowden, Bannayan-Riley-Ruvalcada, Proteus and Proteus-like, and Cronkite-Canada syndromes (Boardman, 2002; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Provenzale et al., 2016; Talseth-Palmer, 2017).

Lynch syndrome (LS) is the most common hereditary CRC syndrome at a prevalence of 2-5% of all CRC (Arnold et al., 2005; Boardman, 2002; Gonzalez et al., 2017; Provenzale et al., 2016). It is inherited as an autosomal dominant defect and caused by mutations in the DNA mismatch repair (MMR) genes, most commonly *MLH1, MSH2, PMS2, MSH6,* etc., and the *MSH2* neighbour, *EPCAM* (Arnold et al., 2005; Boardman, 2002; Gonzalez et al., 2017; Provenzale et al., 2016; National Institute for Health and Care Excellence, 2017b) . These genes are involved in the repair of DNA mismatches which usually accumulate during DNA replication (Arnold et al., 2005; Boardman, 2002; Gonzalez et al., 2017; Provenzale et al., 2016; National Institute for Health and Care Excellence, 2017b). Mutations of the MMR genes lead to genome-wide accumulations of mutations and confers a thousand-fold mutation rates in the tumours of LS patients (Boardman, 2002). These mutations – short insertions and deletions, or indels - are most notable in genomic regions with repetitive DNA sequences called microsatellites, and lead to a

phenomenon called microsatellite instability (MSI) (Boardman, 2002; Gonzalez et al., 2017). Microsatellites are more commonly found in non-coding than in the coding regions of the genome. MSI causes frameshift mutations in growth-regulatory and other genes, and lead to development of tumours (Boardman, 2002; Gonzalez et al., 2017). LS patients have a lifetime risk of 52-82% for developing CRC (de Paula et al., 2021; Dominguez-Valentin et al., 2020; Provenzale et al., 2016). Classically, LS causes a familial clustering of neoplasms in the colon, rectum, stomach, endometrium, ovary, ureters, brain, small bowel, the hepatobiliary tract, pancreas, and skin (Arnold et al., 2005; Boardman, 2002; de Paula et al., 2021; Dominguez-Valentin et al., 2020; Gonzalez et al., 2017; Provenzale et al., 2016; National Institute for Health and Care Excellence, 2017b). LS patients usually present at a younger age than patients with the sporadic form of CRC (Arnold et al., 2005; Boardman, 2002; Gonzalez et al., 2017; Provenzale et al., 2016). Due to the autosomal dominant pattern and high prevalence of LS and the risk of developing metachronous LS tumours, the National Comprehensive Cancer Network (NCCN) and National Institute for Health and Care Excellence (NICE) have recommended that all CRC patients be screened for LS (National Comprehensive Cancer Network, 2020; Provenzale et al., 2016; National Institute for Health and Care Excellence, 2017b). The benefits of this screening programme include the improved chances of early detection of potential tumours at other sites, as well as the tumour detection and/or follow-up of the relatives of the patients (Provenzale et al., 2016).

Familial adenomatous polyposis is an autosomal dominant inherited disorder characterised by early onset of hundreds to thousands of adenomatous polyps throughout the colon and rectum (Arnold et al., 2005; Boardman, 2002; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014; Talseth-Palmer, 2017). About 75-80% of FAP is caused by

transmitted germline mutations in the adenomatous polyposis coli gene, *APC*. A further 20-30% is caused by *de novo* mutations in the germline of one of the parents of the affected individual (Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014). The age of onset of classical FAP is 16-19 years (Arnold et al., 2005; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Talseth-Palmer, 2017). Almost all classical FAP patients develop CRC by the age of 35-40 years, if untreated by prophylactic colectomy (Arnold et al., 2005; Boardman, 2002; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014; Talseth-Palmer, 2017). Classical FAP syndrome also include other features such as gastric fundic polyp, gastric adenomas, duodenal adenomas, peri-ampullary adenomas, congenital hypertrophy of retinal pigment epithelium and cribriform-morular papillary thyroid carcinoma (Arnold et al., 2005; Boardman, 2002; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014; Talseth-Palmer, 2017). Gardner (colonic polyposis, skeletal osteomas, epidermoid cysts, and desmoid fibromatosis), Turcot (colonic polyposis, medulloblastoma), attenuated FAP (fewer polyps (about 30-35), presenting in older patients (about 36 years of age), and at the proximal colon) are variants of FAP, and are all caused by mutations in *APC* (Arnold et al., 2005; Boardman, 2002; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014; Talseth-Palmer, 2017). *APC*, a 15-exon large, tumour suppressor gene that maps to 5q21-22 (Arnold et al., 2005; Boardman, 2002; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014; Talseth-Palmer, 2017). More than a thousand mutations - including point mutations, short deletions, and large rearrangement - have been described in classical FAP and its variants (Gonzalez et al., 2017; Leoz et al., 2015). Most of *APC* mutations that cause the severe forms of FAP are clustered in the codons 1250-1464 of exon 15 (Arnold et al., 2005; Boardman, 2002; Dinarvand et al., 2019;

Gonzalez et al., 2017; Leoz et al., 2015; Talseth-Palmer, 2017). The mutations responsible for the milder forms of the disease (fewer polyps and older onset of disease) are scattered across exons 1-14 (Arnold et al., 2005; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Talseth-Palmer, 2017). *APC* is a multifunctional gene that functions in metaphase chromosome alignment by stabilising microtubules during metaphase. It also functions in the *WNT* signalling pathway where it sequesters β-catenin in the cell membrane and prevents its overexpression. *APC* mutations cause loss of inhibition of β-catenin, leading to the translocation of β-catenin to the nucleus where it upregulates the transcription of genes which regulates cell cycle progression: *C-MYC, CCND1* and *PPARD* (Arnold et al., 2005; Boardman, 2002; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014; Talseth-Palmer, 2017). Thus, *APC* mutations lead to uncontrolled cell proliferation, which provides the appropriate conditions for accumulation of other mutations, such as in *TP53* and *KRAS*, and progression to malignant transformation (Arnold et al., 2005; Boardman, 2002; Dinarvand et al., 2019; Gonzalez et al., 2017; Leoz et al., 2015; Stoffel & Kastrinos, 2014; Talseth-Palmer, 2017). Thus, CRC develop in FAP via the chromosomal instability (CIN) pathway (Gonzalez et al., 2017; Talseth-Palmer, 2017).

MUTYH-associated polyposis is currently regarded as a clinical variant of FAP (Arnold et al., 2005; Dinarvand et al., 2019; Stoffel & Kastrinos, 2014). It is characterised by the presence of tens to hundreds of colonic polyps, which onset is about 45 years of age. MAP usually presents as CRC at the time of diagnosis (Arnold et al., 2005; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Talseth-Palmer, 2017). Other features of MAP include the presence of duodenal polyps and gastric polyps in one-fifth of affected patients. But there is no increased risk of cancer in other sites or organs (Arnold et al., 2005; Cattaneo et al., 2007; Gonzalez et al.,

2017; Leoz et al., 2015; Talseth-Palmer, 2017). It is caused by bi-allelic mutations in the base excision repair gene, *MUTYH*, which is normally involved in the repair of non-bulky DNA damage caused by endogenous and exogenous oxidative agents (Arnold et al., 2005; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Talseth-Palmer, 2017). Specifically, *MUTYH* is involved in the repair of oxidised guanine residue (oxo-8-G) in DNA, the loss of which *MUTYH* leads to genome-wide G>T transversion mutations (Arnold et al., 2005; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Talseth-Palmer, 2017).

Familial, non-syndromic CRC, or familial CRC type X (FCCTX) is a heterogeneous group of CRCs in which the underlying molecular mechanisms are unknown (Armelao & de Pretis, 2014; Stoffel & Kastrinos, 2014; Xu et al., 2020). It is, by definition, characterised by clustering of CRC without evidence for any of the known inherited syndromes (Armelao & de Pretis, 2014; Stoffel & Kastrinos, 2014; Xu et al., 2020). It has been postulated to be either some unrecognised hereditary syndrome or sporadic forms of CRC that aggregate in families (Xu et al., 2020). Although the molecular mechanism is unknown, FCCTX is postulated to result from co-inheritance of multiple low-penetrance genetic alterations, some of which are common (Armelao & de Pretis, 2014; Stoffel & Kastrinos, 2014; Xu et al., 2020). The common genetic loci for familial CRC include 2q32.3, 8q24, 8q23.3, 10p14, 12p13 (*CCND2*), 12q24.21 (*TBX3*), 15q13.3 (*CRAC1/HMPS*), 16q22.2, 18q21 (*SMAD7*) (Abulí et al., 2010; Broderick et al., 2007; Haiman et al., 2007; Jaeger et al., 2008; Peters et al., 2013; Tenesa et al., 2008; Tomlinson et al., 2007; Tomlinson et al., 2008; Zanke et al., 2007). More recently, low-penetrance variants in *MUTYH, BRCA1, BRCA2, POLE, POLD1, TP53, BARD1, ATR, ATM, APC,* and *BRIP1* have been shown to contribute to the molecular pathogenesis of FCCTX (Armelao & de Pretis, 2014; Stoffel & Kastrinos, 2014; Xu et al., 2020). Compared to Lynch syndrome, FCCTX occur in older patients,

are more commonly found in the rectum and in male patients, exhibit less amount of metachronous carcinoma, are more likely well to moderately differentiated tumours with frequent dirty necrosis, and have a low risk of extra-colorectal tumours (Armelao & de Pretis, 2014; Stoffel & Kastrinos, 2014; Xu et al., 2020).

Sporadic CRC develop through three genetic pathways. About 70-85% of sporadic CRC arise via the chromosomal instability (CIN) pathway, 10-15% through the microsatellite instability (MSI) pathway, while less than 5% arise through the CpG island methylation pathway (CIMP) (Turner, 2015; Arnold et al., 2005; Pino & Chung, 2010; Yamagishi et al., 2016). CRC that arises by the CIN pathway are characterised by numerical and structural chromosomal abnormalities leading to aneuploidies, loss of heterozygosity (LOH) and gross chromosomal rearrangements (amplification, duplication, translocations, etc.) (Turner, 2015; Arnold et al., 2005; Pino & Chung, 2010; Yamagishi et al., 2016). The CIN pathway of sporadic colorectal carcinogenesis is associated with accumulation of mutations in *APC, KRAS, PIK3CA, BRAF, TP53*, and *SMAD4* (Turner, 2015; Arnold et al., 2005; Pino & Chung, 2010; Yamagishi et al., 2016). The MSI pathway results from epigenetic silencing of the MMR genes, especially the *MLH1* gene (in more than 80% of cases). Epigenetic silencing of *MLH1* by promoter methylation leads to loss of *MLH1* expression and MSI (Turner, 2015;Arnold et al., 2005; Weisenberger et al., 2006; Yamagishi et al., 2016). Most MSI CRCs are diploid or near-diploid, but have frequent frameshift mutations, especially in the so-called CRC genetic targets of MSI (Arnold et al., 2005; Jonchere et al., 2018; Yamagishi et al., 2016). These genetic targets include *TGFBRII, IGFIIR, BAX, MSH3, MSH6, MLH3, MBD4, PTEN, WISP3, RAD50, RIZ, CASP5, GARE, E2F4*, etc. (Arnold et al., 2005; Yamagishi et al., 2016). Also, mutations in *APC* and *BRAF* are common, but *KRAS* and *TP53* mutations are infrequent (Turner, 2015; Jass, 2007; Yamagishi et al., 2016).

MSI CRC are characterised by right-sidedness, mucinous and medullary histology, infiltration by tumour-infiltrating lymphocytes (TILs), earlier-stage disease and better prognosis (Jenkins et al., 2007; Yamagishi et al., 2016). CRC arising via CIMP have widespread CpG island methylation (Turner, 2015; Barault et al., 2008; Yamagishi et al., 2016). Thirty to forty percent of them occur in the proximal colon, while 3-12% are found in the distal colon (Barault et al., 2008; Yamagishi et al., 2016). On the basis of *MLH1, CDKN2A, MINTS1, MINTS2 and MINTS31* methylation, CRC has been classified as CIMP-high, CIMP-low and CIMP-negative (Arnold et al., 2005; Barault et al., 2008; Jass, 2007; Yamagishi et al., 2016). A subset of CIMP-positive CRC is MSI-high (MSI-H) due to *MLH1* methylation. However, more than 50% of CIMP-positive tumours are microsatellite stable (MSS) (Barault et al., 2008; Jass, 2007; Sullivan et al., 2015). CIMP-high tumours characteristically are MSI-H and have *BRAF* mutations, while CIMP-low tumours have *KRAS* mutations. CIMP-negative CRC are MSS and have high frequency of *TP53* mutations (Turner, 2015; Jass, 2007; Yamagishi et al., 2016).

As can be surmised from the foregoing, the three pathways are not mutually exclusive (Turner, 2015; Jass, 2007; Pino & Chung, 2010; Yamagishi et al., 2016). For example, a subset (25%) of MSI has CIN (Pino & Chung, 2010; Yamagishi et al., 2016). Whilst CIMP-positive tumours are often MSI-H, and 12% of CIN are MSI, 25-33% of CIMP-positive CRC has CIN (Jass, 2007; Pino & Chung, 2010; Simons et al., 2013).

The genetic pathway of colorectal carcinogenesis is related to the evolution of morphology of normal colon to adenocarcinoma. For example, the traditional adenoma-carcinoma sequence (50-70% of CRC), whereby normal colonic epithelium transforms to tubular adenoma and then to adenocarcinoma, occurs more significantly in the distal colon and is associated with *APC* and *TP53* mutations, and the CIN pathway (Jass, 2007; Pancione et al., 2012; Yamagishi et al.,

2016). The serrated pathway – normal epithelium to sessile serrated lesion (SSL) and sessile serrated lesion with dysplasia (SSLD) to adenocarcinoma – occurs in 10-20% of CRC and in the proximal colon and is associated with MSI, *MLH1* methylation, CIMP and *BRAF* mutations (Kim and Kang, 2020; Crockett and Nagtegaal, 2019). The alternative pathway, whereby normal epithelium undergoes transformation to villous, tubulovillous, traditional serrated or serrated

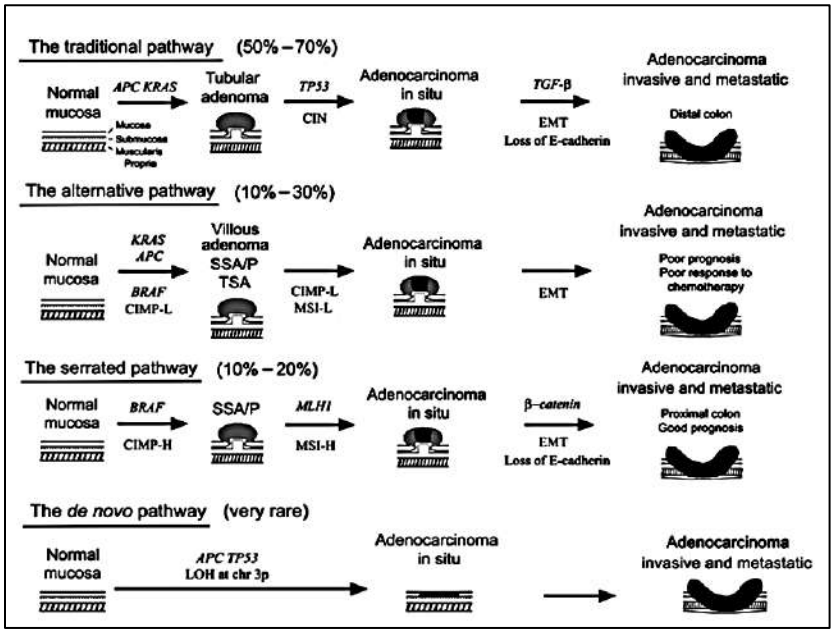


Figure 1.1: Molecular pathogenesis and morphological evolution of sporadic CRC identifies 3 main pathways of colorectal carcinogenesis. A 4th pathway known as the de novo is rare and involve evolution from superficial-type adenoma to superficial-type carcinoma, associated with frequent *TP53* and *APC* mutations, but no *KRAS* mutations. Figure 1.1 was adapted from Yamagishi et al., 2016. The SSA/P (sessile serrated adenomas/polyps) are now correctly known as sessile serrated lesions (Kim and Kang, 2020; Crockett and Nagtegaal, 2019)

tubulovillous adenomas then to adenocarcinoma, occurs in 10-30% of CRC and associated with CIMP, *KRAS*, *BRAF* and *APC* mutations (Jass, 2007; Pancione et al., 2012; Yamagishi et al., 2016; Kim and Kang, 2020; Crockett and Nagtegaal, 2019).

In 2015, the CRC Subtyping Consortium (CRCSC) described the consensus molecular subtypes (CMS) of CRC. Four subgroups of CRC – CMS1 to CMS4 – were recognised (Guinney et al., 2015). CMS1 tumours comprise 14% of the consensus molecular types of CRC and are essentially hypermutated MSI tumours with immune activation and low somatic copy number alterations (SCNA). CMS2 tumours show high SCNA with copy number gains and losses,

comprise 37% of CRCs and have chromosomal instability with marked WNT and MYC activation. CMS3 tumours, comprising 15% of CRCs, are described as metabolic and epithelial since they show molecular features of metabolic dysregulation. They have few SCNA and higher CIMP-low cluster. About 30% of them are also hypermutated. CMS4 tumours (23% of CRCs) are described as mesenchymal since they show prominent TGFB activation, stromal invasion and angiogenesis. Furthermore, whilst CMS1 tumours show significant rates of BRAF mutations, receptor tyrosine kinase activation, the CMS2 cases show consistent enrichment in HNF4A amplification, and not in other genetic loci. CMS3 tumours have significant rates of KRAS mutations and MAPK pathway activation. From the clinicopathological standpoint, CMS1 tumours are more frequently found in females, are right-sided and have higher histological grades. CMS2 tumours, on the other hand, are left-sided lesions, whilst CMS4 tumours are significantly diagnosed at the advanced stages (stages III and IV). Furthermore, whilst the patients with the CMS4 tumours show worse overall survival and relapse-free survival, the CMS2 patients show superior survival indices after relapse. The CMS1 patients are characterised by poor survival after relapse (Guinney et al., 2015).

1.2 Gastric Cancer

1.2.1 Epidemiology of gastric cancer

Gastric cancer (GCa) is the 5th most common cancer worldwide after breast, lung, colorectal and prostate cancers, accounting for 1,089,103 new cases of cancer (5.6% of all cancers) in 2020 (Ferlay et al., 2021; Sung et al., 2021). It is also the 4th most common cause of cancer deaths after lung, colorectal and liver cancers, accounting for 768,793 (7.7%) cancer deaths in 2020. In the male gender, it is the 4th commonest cancer in 2020 after lung, prostate and colorectal

cancers, and the 4th most common cause of cancer deaths after lung, liver, and colorectal cancers (Ferlay et al., 2021; Sung et al., 2021). Among the female gender GCa is 7th in the most frequent malignancies as well as the 5th most common cause of cancer deaths after breast, lung, colorectal and cervical cancers (Ferlay et al., 2021; Sung et al., 2021). Like CRC, GCa shows a very wide variation in geographical incidence. Its incidence is highest in Eastern Asia, Central and Eastern Europe, South America, Polynesia and Western Asia where Age-standardised rates (ASR) incidence range from 8.5 – 22.4 per 100,000 population, and lowest in Western Europe, the Caribbean, Central America, Southern Europe and Melanesia with ASR incidence of 5.9 – 7.9/100,000 population (Ferlay et al., 2021; Sung et al., 2021). The mortality rates generally followed the incidence rate patterns, with the highest ASR mortality of 6.3-14.6/100,000 population observed in the world regions with the highest incidence rates, while the lowest ASR mortality of 4.8-5.6/100,000 population were observed in the regions with the lowest incidence rates (Ferlay et al., 2021; Sung et al., 2021).

As with CRC, GCa showed a male predominance pattern in 2020 with a male-to-female incidence ratio of 1.95. This male predominance was, however, reversed in Eastern Africa region where the female incidence slightly predominated (Ferlay et al., 2021; Sung et al., 2021). Furthermore, GCa showed a predominantly elderly (60 years of age and above) with an elderly-to-young age ratio of 2.59. This pattern of aged predominance is maintained when the CRC population was examined separately for the male and female gender groups (2.63 and 2.52, respectively) (Ferlay et al., 2021; Sung et al., 2021). The peak age incidence of GCa in 2020 was in the 60-69 age group in both genders. The 60-69 years age-group was also the peak age incidence for either the male and female genders, respectively (Ferlay et al., 2021; Sung et al., 2021).

The risk factors for the development of GCa can be categorised into hereditary/genetic and environmental risk factors. The environmental risk factors include Helicobacter pylori infection, antioxidant deficiencies, high salt intake, dried or pickled foods, nitroso compounds, Epstein-Barr infection, tobacco and alcohol consumption, low socioeconomic class, and atrophic gastritis both of which are associated with *H. pylori* infection, low standard of hygiene, smoking and a diet lacking fruits and vegetables but rich in starch and preserved meat (Ajani et al., 2016; Carl-Mcgrath et al., 2007; Fuccio et al., 2010; IARC Monograph, 1994; Malila et al., 2002; Plummer et al., 2007; Rawla & Barsouk, 2019; Taylor, 2007). The genetic risk factors are discussed in the molecular pathogenesis of GCa.

1.2.2 Clinical features of GCa

There is a general lack symptomatology in early gastric carcinoma. In 50% of cases the patients present with non-specific gastrointestinal symptoms like dyspepsia. However, if the tumour is cardiac or pyloric the patient may present early with clinical features of gastric outlet or gastroesophageal junction obstruction (Carl-Mcgrath et al., 2007; Dicken et al., 2005).

Advanced GCa, on the other hand, present with more severe symptomatology such as abdominal swelling and pain, bleeding, vomiting, anorexia, dysphagia, severe weight loss and evidence of invasion of adjacent structures/metastasis to lymph nodes and other organs/structures (Carl-Mcgrath et al., 2007; Dicken et al., 2005).

Uncommon presenting features of GCa include paraneoplastic conditions, including microangiopathic haemolytic anaemia, membranous nephropathy, Trousseau's syndrome, the Leser–Trélat sign, acanthosis nigricans, and in rare cases, dermatomyositis (Dourmishev & Draganov, 2009; Lechner & Obermeier, 2012; Lefaucheur et al., 2006; Thrumurthy et al., 2011).

In most parts of the world, most patients present with advanced GCa, whereas in Japan and Korea the increased use of mass screening of patients has resulted in the identification of a large proportion of early cases, with a consequential improvement in survival rates (Ajani et al., 2016; Carl-Mcgrath et al., 2007; Rawla & Barsouk, 2019).

1.2.3 Molecular pathogenesis of GCa

At the molecular level, GCa is a heterogeneous group of malignant epithelial neoplasm of the gastric mucosa characterised by genetic, epigenetic, and chromosomal alterations. Inherited germline mutations in *CDH1* and the DNA mismatch repair genes, *MLH1, MSH2,* etc., predispose to the hereditary diffuse gastric cancer (HDGC)- and the Lynch syndrome-associated GCa, respectively, while germline mutations in *APC* and *STK11/LKB1* are associated with the familial adenomatous polyp (FAP) and the Peutz-Jeghers syndrome respectively (Capelle et al., 2010; Hansford et al., 2015; Shibata et al., 2013; van Lier et al., 2011). Other inherited risk factors for GC include Li-Fraumeni syndrome (*TP53*), and the familial breast and ovarian cancer (*BRCA1 and BRCA2*) (Grabsch & Tan, 2013).

Point mutations in growth regulatory genes such as *KRAS, BRAF, APC, HER2, TP53, PTEN, PIK3CA, ARID1A, CTNNB1, CDH1,* etc. have been described in GCa. Copy number alterations in *KRAS, FGFR2, CCNE1, CCND1,PIK3CA, EGFR, HER2, MET, RB1, CDKN2A, CDKN2B,* etc., epigenetic silencing of tumour-suppressor genes *MLH1, APC, CDH1, MGMT, RARβ, COX-2, TIMP-3, CHFR, RUNX3, CDKN2A, CDKN2B,* etc., chromosomal deletions and amplifications involving 1p, 1q, 3p, 5q, 7q, 12p, 13q, 17q, 18p, 18q, 19p, 20pq, 21q, etc., and microsatellite instability (MSI) have also been described in GCa (An et al., 2012; Deng et al., 2012; Fang et al., 2012; Hu & He, 2013; Kim et al., 2000; Kim et al., 2013; Kimura et al., 2004; Lee et al.,

2012; Nadauld & Ford, 2013; Peng et al., 2003; Satherley et al., 2015; Sato & Meltzer, 2006; Shi et al., 2012; Takahashi et al., 2014; Tamura, 2006; Tay et al., 2003; K. Wang et al., 2011). While some of these molecular alterations are acquired early in the course of the disease and are regarded as the founding alterations in the oncogenic pathway, others occur late in the course of the disease by the sequential accumulation of more alterations which confer additional malignant phenotypes to the tumour cells (H. S. Kim et al., 2000).

GCa shows a good genotype-phenotype correlation. For example, MSI-H GCa is predominantly of the intestinal-type morphology (An et al., 2012; Fang et al., 2012; J.-Y. Kim et al., 2013), while GCa with *CDH1* mutations is associated with the diffuse-type histomorphology (Nadauld & Ford, 2013). Furthermore, some alterations have prognostic/ therapeutic significance, e.g., *MET, FGFR2, EGFR, HER2* amplification and MSI (An et al., 2012; Deng et al., 2012; Fang et al., 2012; Kim et al., 2013).

In recent times molecular profiling of tumours has enabled the discovery of molecular markers for targeted therapy, which has been shown to improve outcomes in certain cancers (Wei et al., 2005). Also, identification of drug resistance biomarkers has not only enhanced the understanding of mechanisms of drug resistance in cancer but has aided clinical decision-making (Davidson, 2016). Furthermore, molecular profiling of tumours has aided the subtyping of patients into prognostic and therapeutic groups, which in turn has led to improved management of patients in cancer (Kittaneh et al., 2013; Perez-Villamil et al., 2012).

In the last decade attempts have been made to subtype gastric cancer into molecular groups using the genetic/molecular profile of gastric cancer sample sets.

In 2013, Lei et al derived a molecular classification for GCa using expression profiling (Lei et al., 2013). The study identified 3 molecular subtypes of GCa with differential pathway genes

expression: Mesenchymal (focal adhesion and ECM-receptor interaction), Proliferative (cell cycle and DNA replication), and Metabolic (various metabolic processes, e.g., digestion) (Figure 1.2). The genetic pathways most altered (i.e., genes highly expressed or downregulated) in these molecular subtypes are shown in Table 1.1. This subtyping scheme showed some good correlations with clinicopathological features such as tumour grade and Lauren classification. The mesenchymal subtype of gastric cancer more than the proliferative tend to be high grade. The diffuse type of GCa tends to be mesenchymal than either the proliferative or the metabolic, while the intestinal type tend to be proliferative and metabolic than mesenchymal. A good therapeutic significance was observed with this subtyping scheme. Patients with the metabolic subtype of GCa derived benefit from 5FU treatment. The metabolic subtype cell lines were more sensitive to 5FU, while the mesenchymal subtype GCa cell lines were sensitive to compounds that target the PI3K-AKT-mTOR pathway. In the Singaporean and Australian cohorts, the metabolic subtype showed better cancer-specific and disease-free survival following treatment with 5FU (100% 5-year cancer-specific survival among the Singaporean cohorts). However, in the absence of treatment, there was no difference in survival among the 3 subtypes, although the proliferative subtype patients had worse progression-free survival in multivariate analysis. Furthermore, this study did not define a minimal predictor gene set for translation into clinical use (Lei et al., 2013).

In 2014, the Cancer Genome Atlas Group described 4 molecular subtypes of GCa, including the Epstein-Barr Virus (EBV), microsatellite instability (MSI), genome stable (GS) and chromosomal instability (CIN) subtypes using array-based somatic copy number analysis, whole exome sequencing, array-based DNA methylation profiling, messenger RNA sequencing, microRNA sequencing and reverse-phase protein array (Figure 1.2) (Bass et al., 2014). This

classification showed subtype-specific characteristics such as extreme DNA hypermethylation, *PIK3CA* mutations and PDL1/2 expression (EBV), *MLH1* silencing (MSI), diffuse histology and *CDH1* and *RHOA* mutations (GS), intestinal histology, high recurrent focal gene alterations, i.e., amplification and deletions and *TP53* mutations (CIN). It also identified potential targets for therapy both within and across the four subtypes. However, this scheme did not define a minimum biomarker set for clinical translation or show any prognostic significance of the molecular classification scheme.

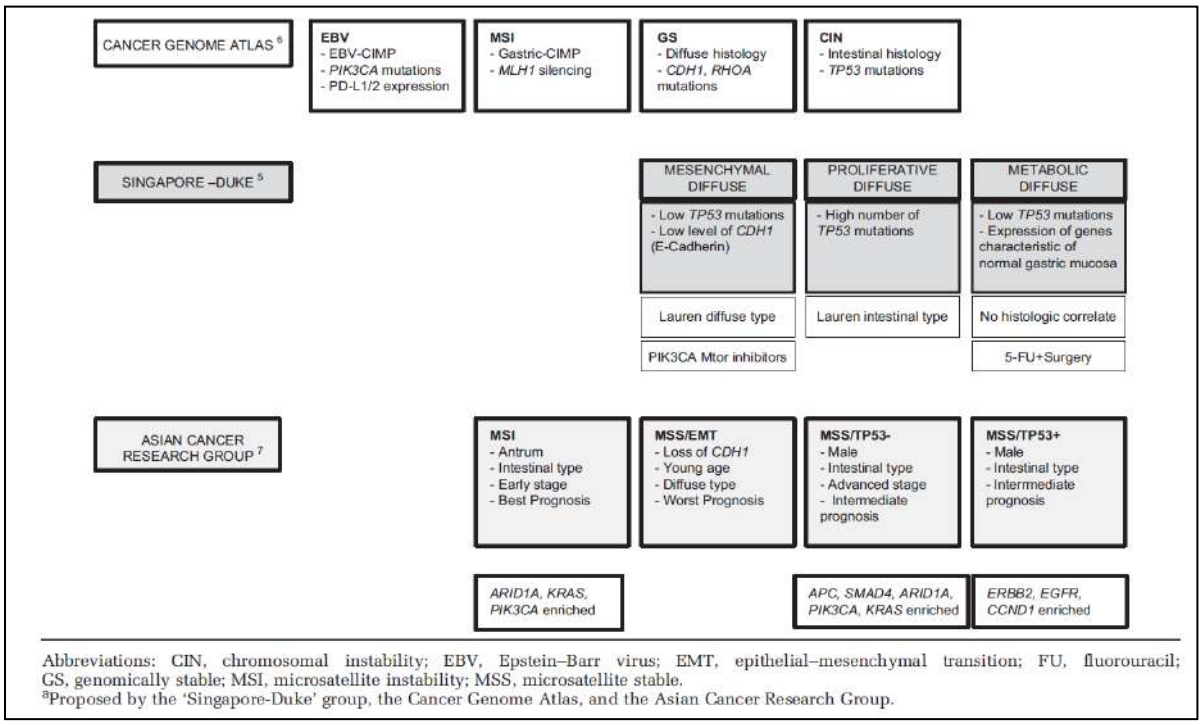


Figure 1.2: Summary molecular classifications of gastric cancer. Adapted from Setia et al., 2016.

In 2015, the Asian Cancer Research Group described 4 four molecular subtypes of GCa using expression profiling, gene chips for copy number variation, targeted gene sequencing and whole gene sequencing: microsatellite instability (MSI), microsatellite stable with epithelial mesenchymal transition (MSS/EMT), microsatellite stable with active P53 (MSS/P53+ve), and microsatellite stable with inactive P53 (MSS/P53-ve) (Figure 1.2) (Cristescu et al., 2015). This

classification scheme has prognostic importance as the MSI subtype showed the best prognosis followed by the MSS/P53+ve then the MSS/P53-ve subtypes. The MSS/EMT subtype has the worst prognosis. Also, the scheme has good clinical correlates as the MSS/EMT is characterised by preponderance of younger patients, advance stage disease (stages III and IV) and diffuse histological type; the MSI subtype occurs predominantly in the antrum, shows intestinal type morphology and stages I and II disease. The MSS/P53+ve subtype is associated with EBV infection. At the molecular level, the MSI subtype shows hypermutation involving *KRAS, PI3-PTEN-mTOR* pathway, *ALK* and *ARID1A*. The MSS/EMT on the other hand shows lower mutation rate. The MSS/P53-ve subtype has high *TP53* mutation rates and high rate of focal gene alteration such as *ERBB2, EGFR, CCNE1, CCND1*, etc., while the MSS/P53+ve subtype shows higher mutation rates of *APC, ARID1A, KRAS, PIK3CA* and *SMAD4*. The study also defined a minimum predictor immunohistochemistry biomarker set for potential clinical translation: MLH1 (MSI), VIM/ZEB1/CDH1 (MSS/EMT), MDM2 and CDK1A for defining P53 activity (MSS/P53+ve or MSS/P53-ve)(Cristescu et al., 2015).

Table 1.1: Molecular subtypes of gastric cancer and their most commonly associated altered gene expression as found by Lei et al., 2013

PROLIFERATIVE	METABOLIC	MESENCHYMAL
CCNE1 Up-regulation	*TFF2* Up-regulation	*CDH2* Up-regulation
E2F "	*MUC5AC* "	*TGFB* "
MYC "	*MUC6* "	*VEGF* "
RAS "	*ATP4A* "	*NFKB* "
ERBB2 "	*ATP4B* "	*MTOR* "
PDE4D Down-regulation	*GIF* "	*SHH* "
PTPRD "	*CHGA* "	*CDH1* Down-regulation
		High *TP53* mutations rates

In 2016, Setia et al described 5 subtypes of GCa using in situ hybridisation and immunohistochemistry: Epstein Barr Virus positive (EBV +ve), Microsatellite instability High

(MSI-H), E-cadherin aberrant, P53 aberrant and Normal P53 subtypes. EBV +ve tumours have prominent lymphoid infiltrate, strong expression of PDL1 and good survival characteristics (Setia et al., 2016). The MSI-H tumours have lower frequency of lymph node metastases, tendency to better survival, loss of MLH1 and PMS2. The tumours in the aberrant E-cadherin expression group show diffuse type of morphology have lower aberrant P53 expression and belong to the older age-group. The aberrant-P53-expression group are characterised by intestinal morphology, higher lymph node stage and increased HER2 expression. Lastly, the normal P53 group are mostly of the intestinal-type morphology and have increased MUC6 expression.

It is important to note that these groups are not mutually exclusive, but each patient's tumour is assigned into a group based on a diagnostic algorithm (Figure 1.3). This classification showed prognostic relevance. It also identified targets for therapy in the EBV +ve (PDL1), and aberrant P53 expression (HER2) subtypes. The MSI is a biomarker of 5-Fluorouracil-based chemotherapy and immunotherapy response (S. T. Kim et al., 2018; S. Y. Kim et al., 2015; Smyth et al., 2017).

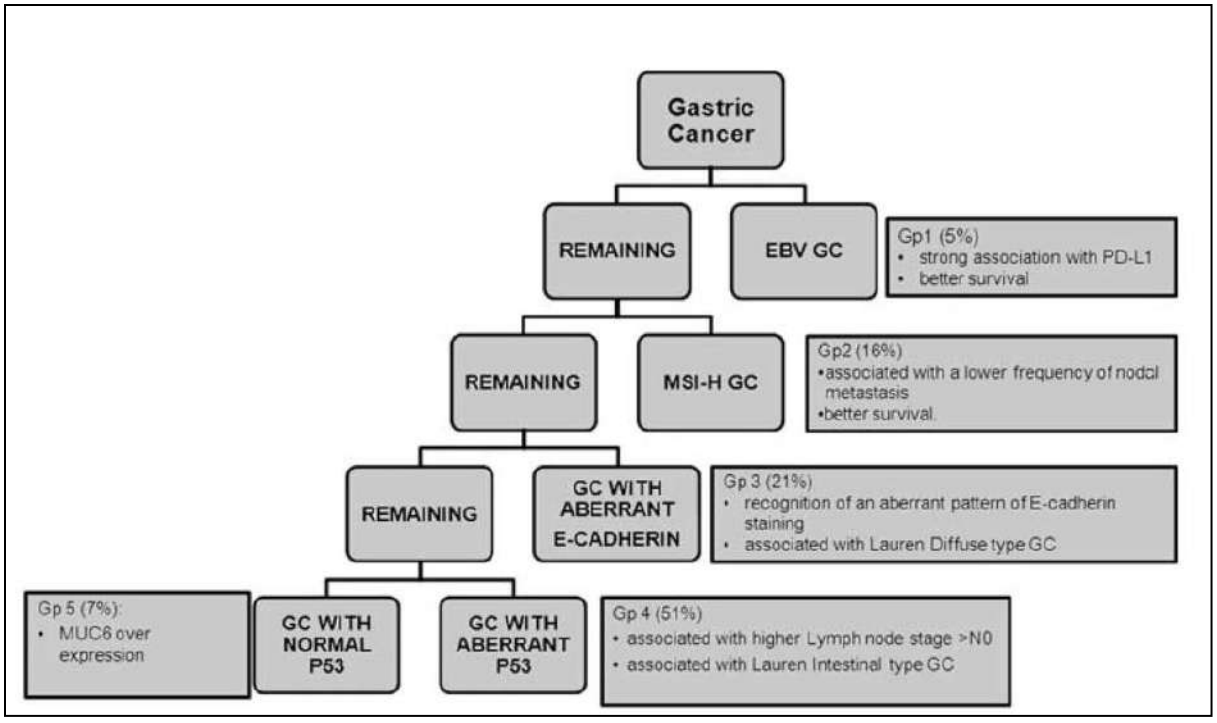


Figure 1.3: Algorithm for molecular classification of gastric cancer by Setia et al., 2016.

1.3 The role of diagnostic molecular pathology in oncology

Nowhere else is the importance of molecular diagnostics/ molecular pathology in health care, including cancer care, more manifest than in the UK healthcare sector. The UK Accreditation Service (UKAS) is the UK body that performs accreditation and certification (by ISO 15189) of services offered by medical laboratories, including molecular diagnostic laboratories, and that ensures that these laboratories conform to and comply with standards and regulations through testing, verification, inspection, and calibration (UK Accreditation Services, 2018; Public Health England, 2017; Saunders et al, 2013). This is so that laboratory tests delivered to the UK public are reliable and fit for the purposes for which they are designed. Over the years, the UK has developed and implemented a rigorous External Quality Assessment scheme (EQA) for molecular testing in oncology and other specialties via both the Cytogenomics External Quality Assessment Service (CEQAS) and UK National External Quality Assessment Service (UK NEQAS) for Molecular Genetics, now both merged as one Genomics Quality Assessment (GenQA) (CEQAS, 2018; United Kingdom National External Quality Assessment Service, 2019; Richman et al, 2017; Richman et al, 2018). These schemes ensure that molecular diagnostics laboratories all around the country participate in regular Quality Assessments to ensure continuous validity of the molecular tests offered to the British public (Richman et al, 2017; Richman et al, 2018). Recently, the UK government, recognising the importance of molecular testing in oncology and other specialties, made a bold move to consolidate genomic testing in the UK. Through the Department of Health and Social Care seven Genomic Laboratory Hubs (GLHs) and a national testing network were established in October 2018 to consolidate and enhance the existing molecular laboratories provisions, with the aim of creating a world class genomic testing resource (Leong, 2022; NHS England, 2021a). It also established the National

Genomics Informatics System to enable direct access to test ordering by patients (Leong, 2022; UK Department of Health & Social Care, 2018); as well as established the National Genomics Directory for Cancer and Rare Diseases which outlines all tests that would be performed by the GLHs (Leong, 2022; NHS England, 2021b). The ultimate purposes of these attempts are to (i) standardise molecular testing services across the UK, (ii) improve collaborations among clinical experts, (iii) maximise the use of equipment, and (iv) ultimately provide better treatment and access for the patients (Leong, 2022; UK Department of Health & Social Care, 2020). It is understood that genomic testing, by a combination of targeted next-generation sequencing, whole exome sequencing (WES) and whole genome sequencing (WGS), would be deployed as standard of care under the new Genomic Medicine policy (UK Department of Health & Social Care, 2020). In 2020, the UK government, as a follow-up on the above, published the "Genome UK- the future of healthcare", a policy which set out a ten-year strategy to create the most advanced healthcare system in the world, delivering better health outcomes at lower costs (UK Department of Health & Social Care, 2020). It backed this policy by committing to increase public spending on Research and Development by £22 billion by 2025 (UK Department of Health & Social Care, 2020). And in 2021 it announced the release of new funding of £37 million for genomics projects and data-driven initiatives under the Genome UK Implementation Plan and the UK Functional Genomics Initiative (Rees, 2021). It was projected that the Genomic UK strategy would receive £17 million funding from the £37 million above to, among other aspects of the Genome UK plan, support a next-generation approach to cancer diagnosis (Rees, 2021). So, the entire medical science world now looks forward to seeing how much implementing a comprehensive molecular testing strategy would improve cancer care and prolong the lives of cancer patients.

The roles of molecular pathology and genetic testing in the management of cancer patients can be summarised as follows: (i) identification of cancer predisposition syndromes, (ii) confirmation of diagnoses and resolution of differential diagnoses, (iii) prognostication of cancer (iv) prediction of response to therapy (Agarwal et al., 2015; Boardman, 2002; Bovée & Hogendoorn, 2010; Cattaneo et al., 2007; Demetri et al., 2010; Demicco, 2013; Dubbink et al., 2014; Gallon et al., 2021; Gonzalez et al., 2017; Hsiao, 2019; Kopetz et al., 2019; Künstlinger et al., 2014; Leoz et al., 2015; Murtuza et al., 2019; Nabors, 2020; National Comprehensive Cancer Network, 2020, 2021g, 2021f, 2021e, 2021b; Pandith et al., 2018; Provenzale et al., 2016; Wieduwilt & Moasser, 2008; Wojdacz & Dobrovic, 2007).

Diagnostic molecular pathology, an emerging discipline in Pathology, deals with the determination of the molecular aberrations present in tumours for diagnostic, prognostic, or predictive purposes (Demicco, 2013; Dubbink et al., 2014; Hsiao, 2019). It encompasses the development of molecular and genetic testing for the diagnosis and classification of diseases, the discovery of biomarkers and design of assays or tests for the prediction of treatment response, drug toxicity, disease progression, and the susceptibility of individuals to disease development (Bovée & Hogendoorn, 2010; Dubbink et al., 2014; Hsiao, 2019).

Diagnostic molecular pathology currently plays a significant role in the clinical management of cancers. Molecular pathology approaches have been employed in all the levels of cancer management, from diagnosis and resolution of differential diagnoses to prediction of treatment response, prognostication of disease and disease recurrence monitoring (Boardman, 2002; Bovée & Hogendoorn, 2010; Cattaneo et al., 2007; Demicco, 2013; Dubbink et al., 2014; Gallon et al., 2021; Gonzalez et al., 2017; Hsiao, 2019; Leoz et al., 2015; Provenzale et al., 2016). Although

molecular pathology has found applications in all types of cancers, the focus of the sections that follow will be on its role in management of solid cancers.

1.3.1 Molecular pathology in cancer diagnosis

Although tumour morphology and immunohistochemistry remain the bases of histopathological diagnosis of cancer, the detection of tumour-specific genetic alterations has had significant roles in the diagnosis and resolution of differential diagnoses of many cancer types in recent years (Boardman, 2002; Bovée & Hogendoorn, 2010; Cattaneo et al., 2007; Demicco, 2013; Dubbink et al., 2014; Gallon et al., 2021; Gonzalez et al., 2017; Hsiao, 2019; Leoz et al., 2015; Provenzale et al., 2016). The tumour-specific genetic alterations tested for cancer diagnosis range from point mutations, small indels, gene amplifications and deletions, to large chromosomal rearrangements such as translocations, and aneuploidies (Bovée & Hogendoorn, 2010; Demicco, 2013; Dubbink et al., 2014; Hsiao, 2019).

Molecular diagnostics have impacted the diagnosis of the sarcomas more than any other group of malignancies. As many as 20% of sarcoma diagnoses depend on molecular diagnostics (Bovée & Hogendoorn, 2010; Demicco, 2013). The need for characterisation of the genetic alterations in sarcoma arise from the frequent difficulty experienced by pathologists in differentiating one sarcoma type from the other based on tumour morphology and immunohistochemistry alone (Bovée & Hogendoorn, 2010; Demicco, 2013). Many sarcomas are characterised by specific recurrent gene fusions resulting from chromosomal translocations. These translocations are detected by fluorescent in-situ hybridisation (FISH); and as these gene fusions give rise to fusion transcripts, they are also readily detected by breakpoint reverse transcriptase PCR (RT-PCR) (Bovée & Hogendoorn, 2010; Demicco, 2013; Dubbink et al., 2014; Hsiao, 2019). For example, testing of chromosome 12q13-25 amplification by FISH is used as an ancillary diagnostic test for

the diagnosis of atypical lipomatous tumour/well differentiated liposarcoma and de-differentiated liposarcoma when immunohistochemistry of MDM2 and CDK4 are equivocal or negative, and to resolve the differential diagnosis between well differentiated liposarcoma from retroperitoneal lipoma (Bovée & Hogendoorn, 2010; Demicco, 2013). The chromosome 22q12 translocations involving the *EWS* gene (also known as *EWSR1*) is a characteristic alteration of the Ewing sarcoma/Primitive Neuroectodermal Tumour (PNET) group. The diagnosis of Ewing sarcoma/PNET presents some diagnostic difficulties because Ewing sarcoma/PNET is morphologically similar to other types of the small round blue cell tumours, including alveolar rhabdomyosarcoma, neuroblastoma, synovial sarcoma and non-Hodgkin lymphoma (Bovée & Hogendoorn, 2010; Demicco, 2013). However, the translocation between the *EWS* and *FLI1* (chromosome 11q24) genes, i.e. t(11;22)(q24;q12), in 85% of Ewing sarcoma/PNET, and *EWS* and *ERG* (chromosome 21q22) genes, i.e. t(21;22)(q22;q12), in a further 5-10% of cases, enables the resolution of differential diagnoses of Ewing sarcoma by FISH and RT-PCR when the result of immunohistochemistry with CD99 is equivocal, such as in the differential diagnosis from monophasic synovial sarcoma (Bovée & Hogendoorn, 2010). Furthermore, the characteristic t(X;18)(p11.2; q11) and t(X;18)(p11.21;q11) translocations, which are present in 80-90% of synovial sarcoma and which give rise to the *SYT-SSX1* and *SYT-SSX2* fusion genes, have been exploited for the resolution of the differential diagnoses from a number of other entities such as malignant peripheral nerve sheath tumour (MPNST), solitary fibrous tumour, spindle cell rhabdomyosarcoma, fibrosarcoma, etc. (Bovée & Hogendoorn, 2010; Demicco, 2013). The t(18:X) translocations, which give rise to an aberrant transcription factor, can be detected using FISH or RT-PCR (Bovée & Hogendoorn, 2010; Demicco, 2013; Dubbink et al., 2014; Hsiao, 2019). Moreover, the diagnoses of alveolar rhabdomyosarcoma could sometimes depend on the

FISH, RT-PCR and genomic breakpoint PCR detection of the *PAX3-FKHR* and *PAX7-FKHR* gene fusions which are present in 85% of that tumour. Accurate differentiation of alveolar rhabdomyosarcoma from embryonal rhabdomyosarcoma, desmoplastic small round cell tumour, Ewing sarcoma, pleomorphic rhabdomyosarcoma and monophasic synovial sarcoma is important for therapeutic reasons (Bovée & Hogendoorn, 2010; Demicco, 2013).

The diagnoses of carcinomas as well as syndromes associated with carcinomas have also been impacted by molecular diagnostics. For example, for the diagnosis of Lynch syndrome the patient's tumour DNA is first tested for the presence of microsatellite instability (MSI) using fluorescent PCR followed by fragment analysis by capillary electrophoresis (CE) and /or IHC of MMR proteins MLH1, MSH2, PMS2 and MSH6 (Boardman, 2002; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Provenzale et al., 2016). An MSI-H case is then selected for further genetic testing such as methylation study of *MLH1* and *MSH2* by methylation-specific PCR, and *BRAF* V600E mutation by allele-specific PCR to exclude a somatic cause of MSI (Gonzalez et al., 2017; Provenzale et al., 2016). The failure to exclude a somatic cause of MSI/MMR-deficiency would warrant testing of germline mutations of MMR genes by Sanger sequencing, MLPA, NGS, etc. (Boardman, 2002; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Provenzale et al., 2016).

In 2009, following confirmation from large prospective studies that genetic markers *BRAF, RAS, RET/PTC* and *PAX8/PPARγ* fusions, mRNA expression marker GEC 167 and immunohistochemistry marker galectin-3 can improve the pre-operative diagnostic accuracy for patients with indeterminate thyroid nodules, the American Thyroid Association's *Revised Management Guidelines for Patients with Thyroid Nodules and Differentiated Thyroid Cancer* recommended the use of molecular markers as an ancillary approach to the diagnosis of papillary

thyroid carcinoma (Cooper et al., 2009). This recommendation was re-endorsed following validation of the 2009 findings by new 2016 studies (Haugen et al., 2016).

1.3.2 Molecular pathology for therapeutic response prediction

By far the most common application of molecular pathology is in the prediction of response to a wide range of cancer therapeutics, including targeted therapy and conventional cytotoxic therapy (Agarwal et al., 2015; Bovée & Hogendoorn, 2010; Demetri et al., 2010; Dubbink et al., 2014; Kopetz et al., 2019; Künstlinger et al., 2014; Y.-P. Liu et al., 2014; Murtuza et al., 2019; Nabors, 2020; National Comprehensive Cancer Network, 2021c, 2021e, 2021b, 2021g, 2021f, 2020, 2021d; Pandith et al., 2018; Patel et al., 2020; Wieduwilt & Moasser, 2008; Wojdacz & Dobrovic, 2007). "Companion molecular diagnostics" is a term that has been coined to describe molecular pathology assays which are developed to test for the presence of genetic or molecular markers of tumour response to specific therapeutic agents. The list of molecular pathology assays that have so far been developed for this purpose is long and still growing. Only a few will be described in the following paragraphs (Agarwal et al., 2015).

KIT exons 9, 11, 13 and 17 and *PDGFRA* exons 12, 14 and 18 harbour mutations (including deletions, insertions and substitutions) which cause constitutive activation of the kinase receptors and downstream signalling, and lead to cell proliferation in gastrointestinal tumours (GISTs) (Demetri et al., 2010; National Comprehensive Cancer Network, 2021e). Whilst about 70-80% of GISTs harbours the *KIT* mutations, 5-10% has the *PDGFRA* mutations. These mutations, which drive tumorigenesis in GISTs and cause over-expression of the proteins, make the tumours targetable by the small molecule inhibitor of kinases imatinib (Gleevec), are detected by PCR-based methods, such as high-resolution melting analysis (HRMA) and direct Sanger sequencing

(Künstlinger et al., 2014). KIT protein expression is detected by immunohistochemistry (Demetri et al., 2010; National Comprehensive Cancer Network, 2021e).

The therapy of high-grade gliomas relies on the chemotherapeutic drug temozolomide which alkylates guanines at the O^6 positions (Nabors, 2020; Pandith et al., 2018; Wojdacz & Dobrovic, 2007). However, sustained response of tumour cells to the alkylating actions of temozolomide is dependent on the expression and activity of *MGMT*, an enzyme which de-alkylates guanines at the O^6 positions. *MGMT* activity renders temozolomide – and other drugs with similar mechanism of action – ineffective (Nabors, 2020; Pandith et al., 2018; Wojdacz & Dobrovic, 2007). The epigenetic silencing of *MGMT* expression by hypermethylation - present in 35-55% of high-grade gliomas - has been demonstrated to improve response to temozolomide (Nabors, 2020; Pandith et al., 2018; Wojdacz & Dobrovic, 2007). *MGMT* methylation status by methylation-sensitive PCR (MS-PCR), methylation-sensitive high-resolution melting (MS-HRM) analysis and MGMT IHC is now routinely determined for patients with high-grade glioma (Nabors, 2020; Pandith et al., 2018; Wojdacz & Dobrovic, 2007).

EGFR encodes a transmembrane receptor belonging to the ERBB family group of receptors which include ERBB1 (HER1), ERBB2 (HER2), ERBB3 (HER3) and ERBB4 (HER4) (Wieduwilt & Moasser, 2008). All family members contain an extracellular ligand-binding domain with two cysteine-rich regions, a single membrane-spanning region and a cytoplasmic tyrosine kinase domain. Binding of ligand (EGF, Amphiregulin, Epiregulin and TGFα) to the extracellular binding domain activate the cytoplasmic tyrosine kinase domain, which in turn activates downstream signalling pathways such as *RAS-RAF-MEK-MAPK, PI3K -AKT-MTOR*, the SRC family of kinases, PLCγ-PKC and *JAK/STAT* pathways (Wieduwilt & Moasser, 2008).

These pathways, in turn, activate several cellular processes such as cell proliferation, invasion, migration, angiogenesis and survival. Constitutive activation of *EGFR* or its downstream effectors, and /or de-activation of pathway inhibitors (such as *PTEN*) by molecular alterations (point mutations, changes in gene copy numbers or gene expression levels) have been established as culprits in tumorigenesis and cancer progression (Wieduwilt & Moasser, 2008). Targeted anti-EGFR therapy, including monoclonal antibodies (e.g., cetuximab and panitumumab) and small molecule inhibitors of kinase (e.g., gefitinib and erlotinib) are routinely used for the treatment of *EGFR1*-mutated lung and other cancer patients (Murtuza et al., 2019; National Comprehensive Cancer Network, 2021f; National Institute for Health and Care Excellence, 2013b). *EGFR* mutations are present in about 10-20% of non-small cell lung cancers (NSCLC). Along with testing for *EGFR* mutations that specify favourable responses to anti-EGFR therapy, molecular testing detects anti-EGFR therapy resistance-inducing mutations such as the T79M mutation in *EGFR*, the extended *RAS* mutations as well as *BRAF* mutations (Murtuza et al., 2019; National Comprehensive Cancer Network, 2021f; National Institute for Health and Care Excellence, 2013b; Wieduwilt & Moasser, 2008). The presence of *EGFR* T79M mutation in NSCLC not only indicates acquired resistance to first- and second-generation EGFR TKI, but it also an indication to switch treatment to osimertinib, a third-generation EGFR TKI which is effective against the T79M mutations in patients with acquired anti-EGFR resistance (Murtuza et al., 2019). Additionally, anti-EGFR therapy is now part of the treatment options for CRC. Anti-EGFR drugs such as cetuximab and panitumumab are currently part of the modified FOLFOXIRI plus panitumumab/cetuximab treatment options for stage IV, unresectable colorectal tumours which are *RAS* and *BRAF* wild type (National Comprehensive Cancer Network, 2021g; 2021b; National Institute for Health and Care Excellence, 2012). Furthermore,

colorectal tumours with *BRAF* V600E mutations are now being targeted with a BRAF-MEK-EGFR inhibitors combination (dabrafenib (BRAF inhibitor) + trametinib (MEK inhibitor) + cetuximab or panitumumab (EGFR antibody), or encorafenib (BRAF inhibitor) + binimetinib (MEK inhibitor) + cetuximab or panitumumab) after the drug combination were shown to give superior survival profile over standard chemotherapy in patients with *BRAF* V600E mutation. *BRAF*-mutated CRC have hitherto had dim survival characteristics (Kopetz et al., 2019; National Comprehensive Cancer Network, 2020; National Institute for Health and Care Excellence, 2021e).

Moreover, *BRAF* V600E mutation is a predictive marker for targeted anti-BRAF therapy in melanoma. About 40-50% of melanoma harbour *BRAF* mutations, 90% which are located in codon 600 (*BRAF* V600). Of the *BRAF* V600 mutations, V600E constitutes 90%, whilst V600K makes up 5%. V600D, V600E2 and V600R make up the rest (National Comprehensive Cancer Network, 2021c; Patel et al., 2020). Targeted *BRAF*-mutated therapy with vemurafenib, cobimetinib, dabrafenib and trametinib, either as single agents or in combination, are now first- or second-line therapy for different stages of melanoma, including metastatic melanoma (National Comprehensive Cancer Network, 2021c; Patel et al., 2020).

EGFR, RAS (*KRAS*, *NRAS* and *HRAS*) and *BRAF* mutations are commonly tested by PCR-based techniques such as mutation-specific (or allele-specific PCR), PCR followed by HRMA, or by direct Sanger sequencing (Y.-P. Liu et al., 2014; National Comprehensive Cancer Network, 2021b; 2021g; 2021c; National Institute for Health and Care Excellence, 2013b; Richman et al, 2017; Richman et al, 2018).

Trastuzumab (Herceptin), an effective monoclonal antibody for breast, gastric and colorectal cancer treatment, targets the IV domain of extracellular fragment of HER2 protein. *HER2*, which

is frequently amplified – and overexpressed - in colorectal, gastric and breast cancers, plays significant roles in proliferation, invasion, metastases, and enhanced cancer cell survival (National Comprehensive Cancer Network, 2021b, 2021g, 2021d, 2021a). Trastuzumab treatment is currently implemented in the management of *HER2*-amplified or HER2-overexpressing colorectal, gastric and breast cancers (National Comprehensive Cancer Network, 2021b; 2021g; 2021d; 2021a; National Institute for Health and Care Excellence, 2010). *HER2* amplification is tested by FISH or by qPCR whilst HER2 protein over-expression is detected by immunohistochemistry (National Comprehensive Cancer Network, 2021b, 2021g, 2021d, 2021a; National Institute for Health and Care Excellence, 2010).

Besides its use to exclude or include Lynch syndrome in gastric and colorectal cancer patients, MSI testing is currently recommended for stage II GCa and CRC cases to determine patients' benefit from fluoropyrimidine-based therapy. It has been demonstrated that the patients with MSI-H, stage II CRC and stages II and III GCa have no benefit from that therapy (Kim et al., 2015; Sargent et al., 2010). Furthermore, the FDA gave an expedited approval of pembrolizumab for the treatment of MSI-H cancers (GCa and CRC inclusive) based on a landmark study which found that patients with MSI-H/MMR-defective tumours showed a significantly better response to immune checkpoint inhibition than those with MSI-negative /MMR-proficient tumours, irrespective of tumour site origin (tissue-agnostic) (Lemery et al., 2017). Currently the NCCN and National Institute for Health and Care Excellence recommend MSI testing for all gastric, metastatic CRC (mCRC) and endometrial cancer (EC) to guide immunotherapy (National Comprehensive Cancer Network, 2021g, 2021b, 2021d; National Institute for Health and Care Excellence, 2021a; 2022). Recently, PD-L1has emerged as a predictive biomarker for prediction

of response to immunotherapy, specifically immune checkpoint inhibitors (ICIs) such as pembrolizumab and is currently recommended by the NICE for prediction of response to pembrolizumab and other immune checkpoint inhibitors in various cancers (National Institute for Health and Care Excellence, 2017c; 2017d; 2017e; 2018a; 2018b; 2020; 2021b; 2021c; 2021d. Davis and Patel, 2020; Akinleye and Rasool, 2019, Twomey and Zhang, 2021). PDL1 expression is tested on tumour tissue sections using IHC (Davis and Patel, 2020; Akinleye and Rasool, 2019, Twomey and Zhang, 2021) and shows strong correlations to MSI in various cancer types (Morihiro et al, 2019; Zhao et al, 2019; Cho et al, 2021; Luchini et al, 2019). In the landmark study that saw the approval of pembrolizumab for CRC therapy, it was found that in addition to the 60% of MSI cases that showed favourable response to immunotherapy 40% of MSS cases also showed favourable response to the drug (Lemery et al., 2017). This throws up the question whether MSI is an adequate biomarker for predicting ICIs response in tumours with MSI subset (e.g., CRC, GCa, and EC). Furthermore, in many non-MSI tumour types that show good response to ICIs PDL1 expression is the recommended biomarker (National Institute for Health and Care Excellence, 2017c; 2017d; 2017e; 2018a; 2018b; 2020; 2021c; Davis and Patel, 2020; Akinleye and Rasool, 2019, Twomey and Zhang, 2021). Thus, overall, PDL1 has emerged as the dominant biomarker for the prediction of response to ICIs. However, PD-L1 expression suffers from the same poor response predictability drawbacks as MSI status (Davis and Patel, 2020). For example, PD-L1 expression was found not to be predictive of ICIs response in 53.3% of the clinical trials investigating ICIs therapy for FDA approval in various cancers (Davis and Patel, 2020). Plus, in recent years the relevance of PD-L1 as a predictive biomarker for ICIs therapy appears to be waning with recent approvals (2021) by the FDA and NCCN that immune checkpoint inhibition can be administered to gastric and GEJ cancer patients regardless of PD-L1

combined positive scores (CPS) (Sahin et al, 2022; US Food and Drug Administration, 2022a, 2022b, National Comprehensive Cancer Network, 2021d; 2022; Bristol Myers Squibb, 2021; Shitara et al, 2020). And MSI is still recommended by both the NCCN and NICE as the predictive biomarker for ICIs in CRC, GCa, EC and oesophageal cancer (National Comprehensive Cancer Network, 2021g; 2021b; 2021d; 2022; National Institute for Health and Care Excellence, 2021a; 2022).

Recently, the NCCN recommended the use of larotrectinib as second-line treatment option for metastatic CRC with *NTRK* gene fusion (National Comprehensive Cancer Network, 2021b, 2021g). This recommendation was preceded by an FDA tissue-agnostic approval of the drug after a clinical study showed that patients with *NTRK* gene fusion-positive tumours had an 80% response rate to Larotrectinib (Drilon et al., 2018). Gene fusions are commonly tested by either FISH or RT-PCR (Bovée & Hogendoorn, 2010; Demicco, 2013; Dubbink et al., 2014; Gonzalez et al., 2017; National Comprehensive Cancer Network, 2021b, 2021g, 2021d).

1.3.3 Molecular pathology for prediction of cancer recurrence

A few molecular pathology assays have been clinically validated for prediction of recurrence in early-stage cancer, notably in breast cancer. These assays offer estimated recurrence risks and prognostic information independent of that obtained from clinical and pathological indices (National Comprehensive Cancer Network, 2015; National Institute for Health and Care Excellence, 2013a) .

The Oncotype Dx genomic assay, an expression profiling-based test, is currently utilised for the prediction of likelihood of breast cancer recurrence and can be used to determine which patient would benefit from systemic chemotherapy to delay recurrence (McVeigh & Kerin, 2017; Paik et

al., 2004, 2006). The assay derives a recurrence score (RS) to stratify patients into low-risk (0-17), intermediate risk (18-30) and high-risk (≥31) groups (Paik et al., 2004, 2006). These scores are based on the expression of 16 cancer-related genes relative to 5 house-keeping genes. These genes include those which serve cellular functions such as proliferation, invasion, cell cycle, apoptosis, intracellular signalling, etc.(McVeigh & Kerin, 2017; Paik et al., 2004). Patients with low-risk RS have a 6.8% risk of distant metastases at 10 years, while intermediate- and high-risk patients have 14.3% and 30.5% risks, respectively, of developing distant metastases (McVeigh & Kerin, 2017; Paik et al., 2004). The Oncotype Dx RS has also been shown to be an independent predictor of breast cancer-related mortality in tamoxifen-treated patients with ER-positive, *node-negative* disease. The 10-year risks of death are 2.8%, 10.7%, and 15.5% in low-, intermediate-, and high-risk patients, respectively (Habel et al., 2006; McVeigh & Kerin, 2017). Furthermore, the Southwest Oncology Group (SWOG)-8814 trial provided evidence to the effect that the Oncotype Dx RS risk categories could stratify those *node-positive* patients at greatest risk of disease relapse and mortality at 10 years, and therefore can discriminate between patients with *node-positive*, ER-positive disease who will derive chemotherapy benefits from those who will not (Albain et al., 2010; Habel et al., 2006; McVeigh & Kerin, 2017). The TransATAC and ECOG2197 studies have provided further evidence that the Oncotype DX RS significantly correlates with recurrence and with time to distant recurrence within both node-positive and node-negative subgroups of invasive breast cancer (Dowsett et al., 2010; Goldstein et al., 2008). The introduction of Oncotype Dx RS into clinical practice has had significant impact on the management of breast cancer. For example, there has been a change in the management plan for 31- 44% of patients in the form of changing from chemotherapy-plus-endocrine therapy regime to endocrine therapy alone. There has been an overall reduction in the use of chemotherapy in

low-risk individuals with the attendant reduction in cost of their management and the avoidance of the side-effects of chemotherapy (Asad et al., 2008; Henry et al., 2009; Lo et al., 2010; Loncaster et al., 2017; McVeigh & Kerin, 2017; Rutter et al., 2016).

Other gene expression signatures which are currently used to guide chemotherapy, predict overall survival, risk of metastasis, recurrence, and benefits or otherwise from chemotherapy include Mammaprint, Prosigna (PM50), Endopredict, Breast Cancer index, and Genomic grade. Whilst Endopredict, Breast Cancer index and Genomic grade are reverse transcriptase PCR-based signature, Prosigna is Nanostring-based, and Mammaprint is DNA microarray-based signature (Brandão et al., 2018; National Institute for Health and Care Excellence, 2013a).

1.3.4 Molecular alterations of clinical importance in GCa and CRC

To summarise, only a limited number of genetic markers have clinical use in GCa and CRC currently. These markers are important for the identification of hereditary cancer syndromes, diagnosis, prognostication, and prediction of response to therapy (Boardman, 2002; Bovée & Hogendoorn, 2010; Cattaneo et al., 2007; Demicco, 2013; Dubbink et al., 2014; Gallon et al., 2021; Gonzalez et al., 2017; Hsiao, 2019; Leoz et al., 2015; National Comprehensive Cancer Network, 2020, 2021b, 2021g, 2014; National Health Service, 2022; Provenzale et al., 2016). These include *APC*, *MUTYH*, MSI, *KRAS, NRAS*, *BRAF*, *HER2, MLH1*methylation, *MLH1, MSH2, PMS2, and MSH6* (Gonzalez et al., 2017; National Comprehensive Cancer Network, 2020, 2021b, 2021g, 2021d; Provenzale et al., 2016; National Health Service, 2022). Whilst MSI, *MLH1* methylation and somatic *BRAF* mutation, and germline mutations and rearrangements in *APC*, *MUTYH, MLH1, MSH2, PMS2, MSH6* and *EPCAM* are tested for the diagnosis of hereditary cancer syndromes (Boardman, 2002; Gonzalez et al., 2017; Leoz et al.,

2015; National Comprehensive Cancer Network, 2014, 2020, 2021d; National Institute for Health and Care Excellence, 2017b; Provenzale et al., 2016), MSI, and somatic *KRAS, NRAS,* and *BRAF* mutations have prognostic relevance (Beghelli et al., 2006; Koncina et al., 2020; Smyth et al., 2017), MSI, somatic *KRAS, NRAS,* and *BRAF* mutations and *HER2* amplifications have important therapeutic response prediction values (Koncina et al., 2020; National Comprehensive Cancer Network, 2020; 2021b; 2021g; 2021d; National Institute for Health and Care Excellence, 2010; 2012; 2021a; 2021e).

MSI is tested by fluorescent PCR followed by fragment analysis using capillary electrophoresis (Umar et al., 2004). Somatic *KRAS, NRAS, BRAF* mutations are tested by PCR-based techniques such as mutation-specific (or allele-specific PCR), PCR followed by HRMA, or by direct Sanger sequencing (Liu et al., 2014; National Comprehensive Cancer Network, 2021b, 2021g, 2021e; National Institute for Health and Care Excellence, 2013b; Richman et al, 2017; Richman et al, 2018), *HER2* amplification or overexpression by FISH, qPCR, qRT-PCR or IHC (National Comprehensive Cancer Network, 2021b, 2021g, 2021d, 2021a), while *MLH1* promoter methylation is tested by methylation-specific PCR or NGS (Gonzalez et al., 2017; Provenzale et al., 2016). Germline mutations in *APC, MUTYH, MLH1, MSH2, PMS2, and MSH6* are tested by direct Sanger sequencing, and more recently by NGS (Boardman, 2002; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Provenzale et al., 2016). Genomic rearrangements of the *MLH1*, *MSH2* and *EPCAM* genes, which are responsible for up to 50% of molecular alterations in Lynch syndrome patients, are tested using multiplex ligation-dependent probe amplification (MLPA), Southern blot hybridisation, multiplex amplifiable probe hybridisation, quantitative PCR analysis, and gene-targeted array-based comparative genomic hybridisation (aCGH) (Boardman, 2002; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Provenzale et

al., 2016). The genetic markers of clinical importance in GCa management include *CDH1* (E-cadherin), MSI, *HER2* amplification or overexpression, *MLH1* methylation, *MLH1, MSH2, PMS2, and MSH6* (Koncina et al., 2020; National Comprehensive Cancer Network, 2021d; National Institute for Health and Care Excellence, 2010). *CDH1* mutation testing enables the identification of the hereditary diffuse gastric cancer syndrome (Koncina et al., 2020; National Comprehensive Cancer Network, 2021d).

The use of Artificial Intelligence (AI) prediction tools is currently gaining traction in health care, especially in the clinical research setting. One area that the use of AI appeared recently is in the prediction of the presence of molecular alterations, including MSI, in patients' tumour by simply analysing the H and E tissue morphology of the tumours (Echle et al, 2021, Bilal et al, 2021, Kather et al, 2019; 2020). The question of the use of AI to predict MSI borders on obviating the need for universal MSI testing and thereby reducing the cost and labour involved in MSI testing, rather than on replacing entirely molecular testing of MSI with AI-based MSI prediction. (Yamashita et al, 2021; Kather et al, 2019; Echle et al, 2022; Echle et al, 2020; Cao et al, 2020; Lee et al, 2021). In other words, is the automation (by AI) of the process of using the known histo-morphological features of MSI to exclude certain patients (i.e., MSS patients) from molecular testing feasible? Molecular testing remains the definitive diagnostic tool for MSI determination. Histological features such as the level of tumour differentiation, presence of signet ring cells, mucinous histology, cribriforming, medullary-type growth pattern, sponge-like mucinous growth, pushing invasive borders, Crohn's-like lymphoid reactions, intra-tumoural lymphocytic infiltrate, intra-epithelial lymphocytes and histological heterogeneity have been found to correlate with the molecular diagnosis of MSI (Alexander et al, 2001; Brazowski et al, 2012; Greenson et al, 2003; Halvarson et al, 2008; Hoproapan and Somran, 2020). At least in the

research settings, pathologists have attempted to predict MSI from the examination of H and E tissue sections of tumours using the aforementioned criteria (Alexander et al, 2001; Brazowski et al, 2012; Greenson et al, 2003; Halvarson et al, 2008; Hoproapan and Somran, 2020). In many studies AI has outperformed pathologists in predicting – or excluding – MSI. Although the level of performances so far achieved with the AI-based MSI prediction tools is desirable (Yamashita et al, 2021; Kather et al, 2019; Echle et al, 2022; Echle et al, 2020; Cao et al, 2020; Lee et al, 2021), further training of the AI models, perhaps with the addition of clinical and gross pathological data, may help improve their performances. However, using AI-based tools for pre-screening patients for molecular testing can be described as 'digging one hole to fill another' in terms of overall costs and labour involved in cancer management. The question is, does the cost and labour of universal testing significantly outweigh the cost and labour involved in deploying AI for MSI pre-screening?

1.4 Development of molecular pathology assays

The rationale for the development of new molecular pathology tests or assays stems from the following factors:

i. The rapid development of new targeted therapy leading to a growing demand for new predictive molecular tests (Agarwal et al., 2015; Dubbink et al., 2014),

ii. The emergence of new diagnostic histopathology entities with unique molecular features that require differentiation from other entities by molecular methods (Demicco, 2013; Trpkov et al., 2021),

iii. The increasing cancer patient population which requires new molecular tests, or modification of existing ones, to improve turnaround time (Bray et al., 2018; Jemal et al., 2011; Sung et al., 2021; Torre et al., 2015),

iv. The primary need to improve the performance indices of existing molecular pathology tests to meet clinical needs: improved accuracy, sensitivity, specificity, cost-effectiveness, ease and speed of deployment and operation of the assay (both of which improve the test turnaround time) and robustness of the assay against tissue types and nucleic acid template types (Association for Molecular Pathology Clinical Practice Committee, 2009, 2014; Association of Molecular Pathology, 1999; Mattocks et al., 2010; Prence, 1999; Saunders et al, 2013; Public Health England, 2017).

The process of introducing a laboratory-developed test into the laboratory entails (i) development of the new laboratory test and (ii) extensive or full and continuous validation of the test (Mattocks et al., 2010).

The aim of test development is to establish the testing procedure and to show in general that the test or assay is valid for the purpose for which it was designed (Association for Molecular Pathology Clinical Practice Committee, 2009, 2014; Association of Molecular Pathology, 1999; Mattocks et al., 2010, Saunders et al, 2013; Public Health England, 2017). Factors that come into considerations in the design and development of a molecular pathology test include the type of molecular/genetic alterations to be tested, the throughput required, the type and quality of analyte(s) to be used, the technology platform(s) available for the test design, the cost-effectiveness of deploying the test and the test performance indices of the new test (which

include accuracy, sensitivity, specificity, precision, positive predictive value, negative predictive value, limit of detection, ease and speed of operation and turnaround time) (Association for Molecular Pathology Clinical Practice Committee, 2009, 2014; Association of Molecular Pathology, 1999; Mattocks et al., 2010, Saunders et al, 2013; Public Health England, 2017).

The type of mutation-detection strategy deployed for testing depends on the types of genetic alterations and whether the mutations are known or unknown (Dubbink et al., 2014). When the mutation is known, point mutations and short indels are generally tested on genomic DNA using allele-specific PCR or HRMA followed by confirmatory Sanger sequencing (Dubbink et al., 2014). However, direct Sanger sequencing or next-generation sequencing can be used when the mutations are unknown. NGS – whether targeted or whole exome or genome – is also used when there is a need to for a higher throughput (Dubbink et al., 2014). Gene copy number variations are usually tested by quantitative PCR, MLPA, FISH, etc. (Boardman, 2002; Cattaneo et al., 2007; Gonzalez et al., 2017; Leoz et al., 2015; Provenzale et al., 2016). And as gene copy number variation give rise to changes in gene expression, gene expression levels can be tested using qRT-PCR, or for higher throughput operations, expression profiling by DNA microarray and NGS (RNASeq) using cDNA (Fey, 2011; Iida & Nishimura, 2002; Saddala et al., 2020; VanGuilder et al., 2008). Genomic rearrangements, such as chromosomal translocations, and chromosomal gains and losses are tested on genomic DNA with FISH, MLPA, array CGH, etc. (Boardman, 2002; Cattaneo et al., 2007; Demicco, 2013; Dubbink et al., 2014; Gonzalez et al., 2017; Leoz et al., 2015; Provenzale et al., 2016; Thunnissen et al., 2012), or on cDNA using the high-resolution RT-PCR for gene fusion caused by chromosomal translocations (Demicco, 2013; Thunnissen et al., 2012). Although Sanger sequencing is the gold standard for single gene mutation detection, it has only a modest limit of detection (Dubbink et al., 2014; Tsiatis et al.,

2010). Allele-specific PCR, pyrosequencing and HRMA are more sensitive, have faster turnaround time and less hands-on time than Sanger sequencing (Dubbink et al., 2014; Tsiatis et al., 2010). However, allele-specific PCR, pyrosequencing and HRMA only identify known mutations, and the PCR-based methods, especially HRMA, often require confirmation by other methods (Dubbink et al., 2014; Tsiatis et al., 2010). FISH is fast to perform, affords the opportunity to examine cellular and tissue morphology, and is amenable to automation. However, the microscopic interpretation of FISH is tedious, and interpretation of intra-chromosomal rearrangements can be particularly difficult (Dubbink et al., 2014). Compared to Sanger sequencing, NGS has higher sensitivity at detecting mutant allele frequency, and only requires low DNA inputs for all the targets to be tested and covers all kinds of mutations (Dubbink et al., 2014). However, the cost of testing and turnaround time are high when only a small number of samples are processed at a time (Dubbink et al., 2014). Furthermore, NGS cannot detect mutations and gene rearrangements in one assay, and it sequences high G+C content and repetitive genomic regions very poorly (Chen et al., 2013; Hagemann, 2015; Kieleczawa, 2006; Tørresen et al., 2019; Treangen & Salzberg, 2012).

The quality of the available analyte also dictates to a large extent the design of the molecular pathology test. FFPE nucleic acid templates are the most available templates for molecular testing in oncology (Dietrich et al., 2013). However, FFPE templates are degraded, and usually the design of the molecular test that use this template type limits the PCR amplicon design to 100-200 bp (Dubbink et al., 2014; Groenen et al., 2011). Also, consideration is given to the fact that clinical nucleic acid templates such as from tumour biopsies contain mixtures of normal and tumour DNA (Dubbink et al., 2014). While this mixture reduces the detection of mutant alleles when using such methods as allele-specific PCR, Sanger sequencing and NGS, it enhances

heteroduplex formation on which the detection capability of HRMA depends (Dubbink et al., 2014; Garritano et al., 2009).

The cost-effectiveness of any designed assay entails not only the low direct costs of running the more efficient new test but also the exclusion of indirect costs that might result from using less-effective tests and the reduction of the overall costs of managing the disease by utilising the new test, even if the direct costs of running the new test may be higher than that of the existing one (Association of Molecular Pathology, 1999).

The performance indices used to establish that the new test is valid for its designed purpose include accuracy (the amount of agreement between the information from the index test and the reference standard), sensitivity (the ability of a test to detect a mutation when that mutation is present), specificity (the ability of a test to give a normal or negative results in specimens without the mutation), precision (getting the same results with repetition of the test or assay), positive predictive value (probability that a positive test result correctly identifies the presence of the mutation), negative predictive value (probability that a negative test result correctly identifies the absence of the mutation being tested), limit of detection (the lowest concentration of the mutant nucleic acid that the assay can detect) (Association for Molecular Pathology Clinical Practice Committee, 2009, 2014; Association of Molecular Pathology, 1999; Mattocks et al., 2010; Prence, 1999, Saunders et al, 2013; Public Health England, 2017).

In this thesis I describe the development of four mutation-detection assays, the Squirrel PCR assay (for Sanger-sequencing of ultrashort DNA segment), mHRMA (for multiplexing of mutation targets on the HRMA platform), MSI-HRMA (a HRMA-based test for MSI testing) and

the MSI-WES (an NGS-based MSI detecting strategy) and show that these assays have high analytical performances.

1.5 The hypothesis and aims of the study

The work in this thesis is based on the hypothesis that "**the development of new molecular diagnostics assays for the detection of clinically relevant molecular alterations in gastric and colorectal cancers, or the enhancement of the performance indices of existing assays - analytical performance, turnaround time, and the cost-effectiveness of the assays - can be accomplished by altering the components and /or the detection methods of the existing assays, or by the utilisation of entirely different mutation testing platforms.**"

The aims of the study are as follows:

i. Squirrel PCR assay – to develop a PCR strategy that enables direct, good-quality, and accurate Sanger-sequencing of ultrashort DNA segments by using long-tailed (or Squirrel) primers that extend the sizes of DNA segments post-PCR.

ii. mHRMA assay – to develop an HRMA-based multiplex assay that enables fast and accurate detection of mutations in CRC by using tagged amplicons to ensure adequate separation of the Tm's of targets along the HRMA temperature axis.

iii. MSI-HRMA – to develop an HRMA-based test for the detection of microsatellite instability (MSI) in CRC.

iv. MSI-WES – to devise a simple and accurate MSI-testing strategy that utilises level 2 whole exome sequencing data.

The principles, further aims and objectives of the individual studies in this thesis are expounded in the respective chapters.

CHAPTER 2

MATERIALS AND METHODS

2.1 Materials

2.1.1 CRC cell lines

Six (6) CRC cell lines with known mutation statuses in *KRAS* exon 2, namely, DLD1, HUTU80, HCT116, LS1034, RKO, and SW620 were used to develop and optimise the Squirrel PCR assay for targeted sequencing of ultrashort DNA segments.

The following six (6) cell lines were included in the development and optimisation of the mHRMA assay: Vaco5, SW984, HCT116, RKO, SW620 and Vaco10MS. These cell lines have been previously validated for mutations (Fadhil et al., 2010).

Furthermore, nine (9) CRC cell lines with known microsatellite statuses were included in the study that demonstrated the utility of the HRMA platform for microsatellite instability (MSI) testing in CRC. These cell lines include those with stable microsatellite loci, MSS (SW480, SW620, HUTU80 and SW837) and those with unstable microsatellite loci, MSI (DLD1, HCT 116, LoVo, LS1034 and RKO).

All the cell lines used in this study were gifted to Molecular Pathology Research Group by Prof Ian Tomlinson, Cancer Research UK (CRUK) (Seth et al, 2009).

2.1.2 Clinical CRC samples

Seventeen (17) formalin-fixed, paraffin-embedded (FFPE) CRC cases from the Pathology archives of the Queen's Medical Centre were included in this study. These cases had previously been screened with the QMC-PCR method for mutations in *PTEN* exons 3, 5 and 7, *KRAS* exons 2, 3 and 4, *NRAS* exon 2, *BRAF* exon 15, *PIK3CA* exons 1, 9 and 20. The templates from these clinical samples had been enriched for the gene targets using the pre-diagnostic multiplex (PDM) stage of the QMC PCR method (Fadhil et al., 2010). The enriched templates from all 17 cases were included in development of the mHRMA assay.

PDM-enriched templates from thirteen (13) FFPE CRC cases – a subset of the mHRMA cohort – with known mutation statuses in *PTEN* exons 3, 5 and 7, *KRAS* exons 2, 3 and 4, *NRAS* exon 2, *BRAF* exon 15, *PIK3CA* exons 1, 9 and 20 were included in the study that validated the Squirrel PCR assay for Sanger sequencing (Fadhil et al., 2010).

A further 166 FFPE samples, including 13 normal and 153 CRC samples, from the Queen's Medical Centre Pathology archives were included in the development of the MSI-HRMA assay. The 166 samples were divided into a training cohort which comprised 44 samples - 31 CRC cases with known microsatellite status based on immunohistochemistry and MSI-PCR (i.e., capillary electrophoresis of standard PCR-amplified microsatellite targets) data, and 13 normal colonic samples – and a validation cohort of 122 CRC cases with known DNA mismatch repair (MMR) status.

All the DNA used in all the studies were obtained from resection samples. DNA was extracted from single blocks per case. Enrichment of tumour DNA by macro-dissection was not performed for any of the samples used in the studies. Tumour blocks included in all the studies had tumour cellularity of 50% and above.

2.1.3 Genomic data for whole exome sequencing-based MSI testing (MSI-WES)

The study cohorts comprised 441 GCa cases with somatic mutations, DNA methylation, gene expression and PCR-derived MSI data. Somatic mutation profile, gene expression and DNA methylation data which were generated by whole exome sequencing (WES), RNASeq and methylation array technologies, respectively, under the auspices of The Cancer Genome Atlas, were retrieved from the National Cancer Institute's Genomic Data Commons (GDC) repository (www.portal.gdc.cancer.gov/repository). The PCR-based MSI status data (MSI status_pcr)

which was generated by PCR and capillary electrophoresis using the MSI-Mono-Dinucleotide Assay marker panel (comprising four mononucleotide markers BAT25, BAT26, BAT40 and TGFRBII; and three CA dinucleotide repeat markers D2S123, D5S346 and D17S250) was obtained from reference study (Y. Liu et al., 2018).

2.2 Methods

2.2.1 DNA Extraction

DNA was extracted from cell lines using the GenElute™ Mammalian Genomic DNA Miniprep Kit (Sigma Aldrich, UK) as per manufacturer's instructions. Briefly, CRC cell pellets in Eppendorf tubes were re-suspended in 200µl of Resuspension Solution. Then 20µl of RNase A Solution was added to the cell suspension which was incubated for 2 minutes at room temperature. Cell lysis was achieved by adding 20µl of Proteinase K and 200µl of Lysis Solution C followed by thorough vortexing for about 15 minutes and incubation at 70 °C for 10 minutes. DNA binding to the spin column was facilitated by column preparation: 500µl of the Column Preparation Solution was added to each pre-assembled GenElute Miniprep Binding Column and spun at 12,000g for a minute, following which the flow-through liquid was discarded.

Homogenisation of the cell lysate was achieved by adding 200µl of 95% ethanol to the lysate in the Eppendorf tube with thorough vortexing, following which the homogenised lysate was loaded into the activated binding columns and centrifuged at 6500g for 1 minute. The flow-through liquid was discarded while the binding column was transferred to another collection tube. A two-step wash procedure was accomplished using 500µl of Wash Solution with

centrifugation both times. DNA elution was achieved with 200µl of Elution Solution and centrifugation.

DNA was extracted from the 17 mHRMA study samples and the 31 CRC and 13 normal samples of the MSI-HRMA study using the QIAmp FFPE Tissue Kit (Qiagen, Germany) with some minor modifications to the protocol. Sections of 20 µm thickness were cut from the tissue blocks and placed in 1.5 mL Eppendorf tubes. The sections were then de-paraffinised using mineral oil with incubation at 80°C for 1 min. Tissue lysis and protein digestion were achieved via an overnight incubation with proteinase K at 56°C in a thermomixer. DNA precipitation was accomplished with 100% ethanol in buffer AL following protein digestion. The precipitated DNA, bound to the MinElute column, underwent a two-step washing process. Elution of the bound DNA was achieved via the addition of 40 mL ATE buffer to the column with a 5 min incubation step at room temperature. All other steps were performed in line with the Qiagen protocol.

DNA was extracted from the validation cohort of the MSI-HRMA study using the Qiagen AllPrep DNA/RNA FFPE kit. Four freshly cut 10µm FFPE CRC and normal tissue sections were placed in an Eppendorf tube and de-paraffinized using 320µl of Qiagen Deparaffinized Solution (cat. no. 19093) following the manufacturer's protocol. The pelleted tissue was re-suspended in Buffer PKD and tissue lysis was accomplished by the incubation with 10µl of Proteinase K solution at 56^0C for 15 minutes, following which the sample was cooled on ice for 3 minutes. The lysate was centrifuged at 20.000g for 15 minutes and the resulting supernatant was collected for RNA purification (RNA not used in this PhD study). The pellets were re-suspended in 180µl of Buffer ATL to which 40µl of Proteinase K solution had been added and incubated overnight to achieve complete lysis. Following lysis DNA was precipitated with Buffer AL (200µl) in 95%

ethanol (200µl). The lysate was then transferred to the QIAamp MinElute spin column, and a two-step wash procedure was applied to the sample using 700 µl each of Buffer AW1 and Buffer AW2. DNA elution was accomplished using 100µl of Buffer ATE.

2.2.2 Primer Design

Primer3 (https://primer3.ut.ee/) was used to design the outer primers used in the PDM stages of the QMC-PCR and mHRMA studies (Fadhil et al., 2010; see chapter 4), while a combination of freely available online databases and software were used to design the inner primers for the MSD stage of the mHRMA, the Squirrel PCR assay and the MSI-HRMA assay.

Information on the mutation hotspots of the targets was obtained from the Catalogue of Somatic Mutations in Cancer (COSMIC) (https://cancer.sanger.ac.uk/cosmic) (Forbes et al., 2015). The sequences around the mutation hotspots were sought and obtained from ENSEMBL (www.ensembl.org) (Aken et al., 2016). Primers were then 'manually' designed using a combination of UCSC *in silico* PCR (https://genome.ucsc.edu/cgi-bin/hgPcr), MFEprimer (http://www.biocompute.bmi.ac.cn/CZlab/MFEprimer-2.0/), Thermofisher Scientific Tm Calculator (https://www.thermofisher.com/sa/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html), and Reverse complement (http://www.bioinformatics.org) (Breslauer et al., 1986; Qu et al., 2012; Speir et al., 2016; Stothard, 2000). Additionally, the sequences of primers previously designed and used in our research group (such as those of *KRAS, BRAF, PIK3CA*, etc.) were utilised to perform *in silico* PCR in order to obtain the gene sequences. New primers are then re-designed from those sequences. Furthermore, for those genes whose mutations are yet to be captured on COSMIC, primer sequences were copied from published literatures and used to re-design new primers as described above. A combination of *in*

silico PCR and UMELT (https://www.dna.utah.edu/umelt/umb.php) (Dwight et al., 2011) were used to predict the melting temperature, Tm, of the designed PCR product.

Sequences and primers for microsatellite-based targets were obtained from the published literatures and from the SelTar database (www.seltarbase.org) (Woerner et al., 2010). New primers were designed from these retrieved sequences using a combination of UCSC *in silico* PCR, MFEprimer, Thermofisher Scientific Tm Calculator, and Reverse complement.
As the specific characteristics of the primers are essential components of the assays described in this thesis the full details of the primer designs for the mHRMA (GC-tagged primers) and Squirrel PCR assays (Squirrel primers) are described in the respective chapters.

2.2.3 Polymerase chain reaction

All PCR performed in this PhD project, including the optimisation steps, were done in the ABI 7500 (Applied BioSystems), Strategene Mx3005 Pro (Agilent) and the Primus96 machines. Both single-stage PCR and the two-stage, nested Quick Multiplex Consensus PCR (QMC-PCR) were applied in the amplification of targets in this project. All single-stage PCRs, as well as the singleplex specific diagnostic (SSD) stages of the QMC-PCR were performed in reaction volumes of 15µl and with primer concentrations of 250nM, unless otherwise stated. The Diamond Hotshot HRM mastermix was used for PCRs unless otherwise stated. Eva Green as 10% of PCR volume was as saturating dye for all real-time PCR and HRM-based reactions. **All the samples included in all the PCR experiments were run in duplicates, except otherwise indicated.**

2.2.4 Quick Multiplex Consensus PCR

The QMC-PCR is a nested PCR comprised of two stages: the pre-diagnostic multiplex (PDM) stage and the SSD stage. The first PDM stage uses outer primers in a multiplex reaction, while the second SSD stage uses nested or inner primers in singleplex reactions.

Enriched templates used for the mHRMA and the Squirrel PCR assays were available in our research group as starting nucleic acid templates. Enrichment had been achieved using the PDM stage of the QMC-PCR method. Eleven targets comprising *PTEN* exons 3, 5 and 7, *KRAS* exons 2, 3 and 4, *NRAS* exon 2, *BRAF* exon 15, *PIK3CA* exons 1, 9 and 20 were amplified with outer primer pairs in 20µl multiplex reactions which consist of 10µl of Diamond Hotshot master mix, 400nM of each outer primer and 20ng of template. The cycling parameters include 1 cycle of 95ºC for 5 minutes followed by 30 cycles of 95ºC for 3 seconds and 55ºC for 10 seconds. The PCR amplicons were diluted 1:100 in nuclease-free water and preserved in cold storage until use. The training cohort for the MSI-HRMA assay was also amplified with the QMC-PCR method and was compared with using single-stage PCR for the assay (chapter 5). The DNA templates of the MSI-HRMA training cohort were enriched using the PDM reactions. Eleven microsatellite targets, including BAT25, BAT26, BCAT25, *ANGEL2*, *EWSR1*, *FBXO46, MYB, TCF4*, *TP53* (BAT34CA), TYMS26, and MONO27, were also amplified using outer primers with the same reaction conditions and cycling parameters as above. The subsequent SSD stage was performed using nested primers and 1:100 dilutions of the PDM products.

2.2.5 HRM Analysis

PCR products were subjected to HRM and analyses on the LightScanner-96 platform. The products were first transferred into wells in the LightScanner HRM plate, and a 20µL mineral oil

overlay was placed in each well. The PCR products were then spun down in a Megafuge centrifuge for 5 minutes at 2500 rpm. Melting was performed using the following parameters: the temperature range of the instrument was set at 65^0C – 95^0C and to 'Auto' exposure with a ramp rate of 0.1^0C/sec. Sample equilibration was performed at 62^0C. The acquired melting data were analysed with the LightScanner Call-IT software version 2.0.0.1.331 using the Expert scanning and Auto scanning modules. The negative filter was utilised to exclude the negative controls and poorly amplified samples from further analyses. To reduce the variation of fluorescence among samples due to difference in starting template or optics, the highest and lowest fluorescence were normalised at 100% and 0% respectively, as per manufacturer's instructions. To overcome the effect of absolute temperature variation from position-to-position across the samples, the normalised data was shifted at the 0.05 level (by software default), while sensitivity was set to normal at zero level. Both the 'Auto group' and 'Common vs Variant' functions were used to group the normalised and shifted melting curves according to their melting patterns. The results were viewed in the 'Shifted melt curves' and 'Difference curves' outputs.

2.2.6 Primer Optimisation

Gradient PCR followed by high resolution melting analysis (or gradient PCR-HRMA) was routinely utilised to optimise the primers designed in all the studies. Optimisation, the process of determining the optimum annealing temperatures for primer pairs, is conventionally done with gel electrophoresis post-gradient PCR. However, the use of HRMA in place of gel electrophoresis for annealing temperature optimisation is already in place (Vossen et al., 2009). The interpretation of the gradient PCR is easy and the HRMA process is faster. It is also

probably more sensitive because HRMA is likely to detect minute quantities of non-specific products which may not be visible to the naked eye when viewed in a gel.

To perform gradient PCR-HRMA, the Tm's of the PCR products were first predicted using the UMELT software (www.dna.utah.edu/umelt/umelt.html) as follows: *in silico* PCRs were first performed on MFEprimer software using the specific primer sequences. The *in silico* amplicon sequences were then cut and pasted in UMELT for *in silico* melt curve derivation, following which the Tm's of the individual products were obtained. For tailed primers, the non-specific long tail forward sequences were tagged to 5' end of the product sequence, while the reverse complements of the long tail reverse sequences were tagged to the 3' end of the PCR product sequence (after conversion on Reverse Complement), as MFEprimer cannot perform *in silico* PCR when the lengths of the non-specific tails are 40nt-long or above. The full sequences (specific PCR product sequences and non-specific tail sequences) were then cut and pasted on UMELT for *in silico* melt curve derivation, following which the Tm's of the tailed primers' products were obtained. Experimental gradient PCRs were performed with cell line DNA using the primer pairs in the presence of Eva green HRMA dye. The products were subjected to HRMA on the LightScanner. The optimum annealing temperatures of the gradient were regarded as those with the highest yields (as shown by the height of the melt curve peaks at or near the predicted Tm's) with absence of non-specific products (i.e., no additional peaks).

Figures 2.1, 2.2 and 2.3 show the gradient PCR-HRMA results of optimisation of some of the primers utilised in the design of the Squirrel PCR, mHRMA and the MSI-HRMA assays. The optimum annealing temperature of the designed primer pair was regarded as that annealing temperature that has a single melting temperature (Tm) peak which is also the highest peak in the

temperature gradient at or near its predicted Tm – the Tm of the designed product was first predicted on UMELT. As the presence of mutation(s) in the target DNA regions could give rise to double-peaked curves even at the actual annealing temperature, an important precaution I took was to perform gradient PCR-HRMA with DNA templates from cell lines which are known to be wild type at the target region.

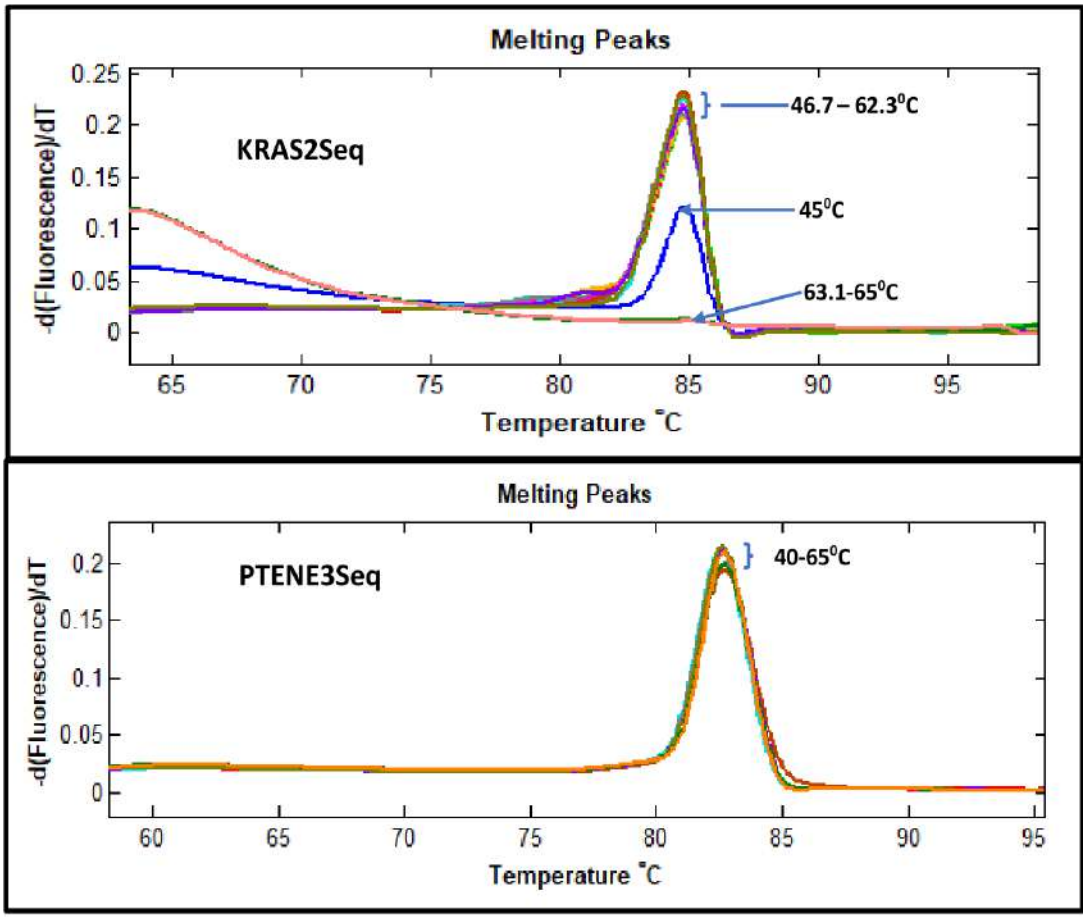


Figure 2.1: Optimisation of the Squirrel primers KRAS2Seq and PTEN3Seq using gradient PCR-HRMA show that primers with long sequences have a wide range of optimum annealing temperatures, Ta's.

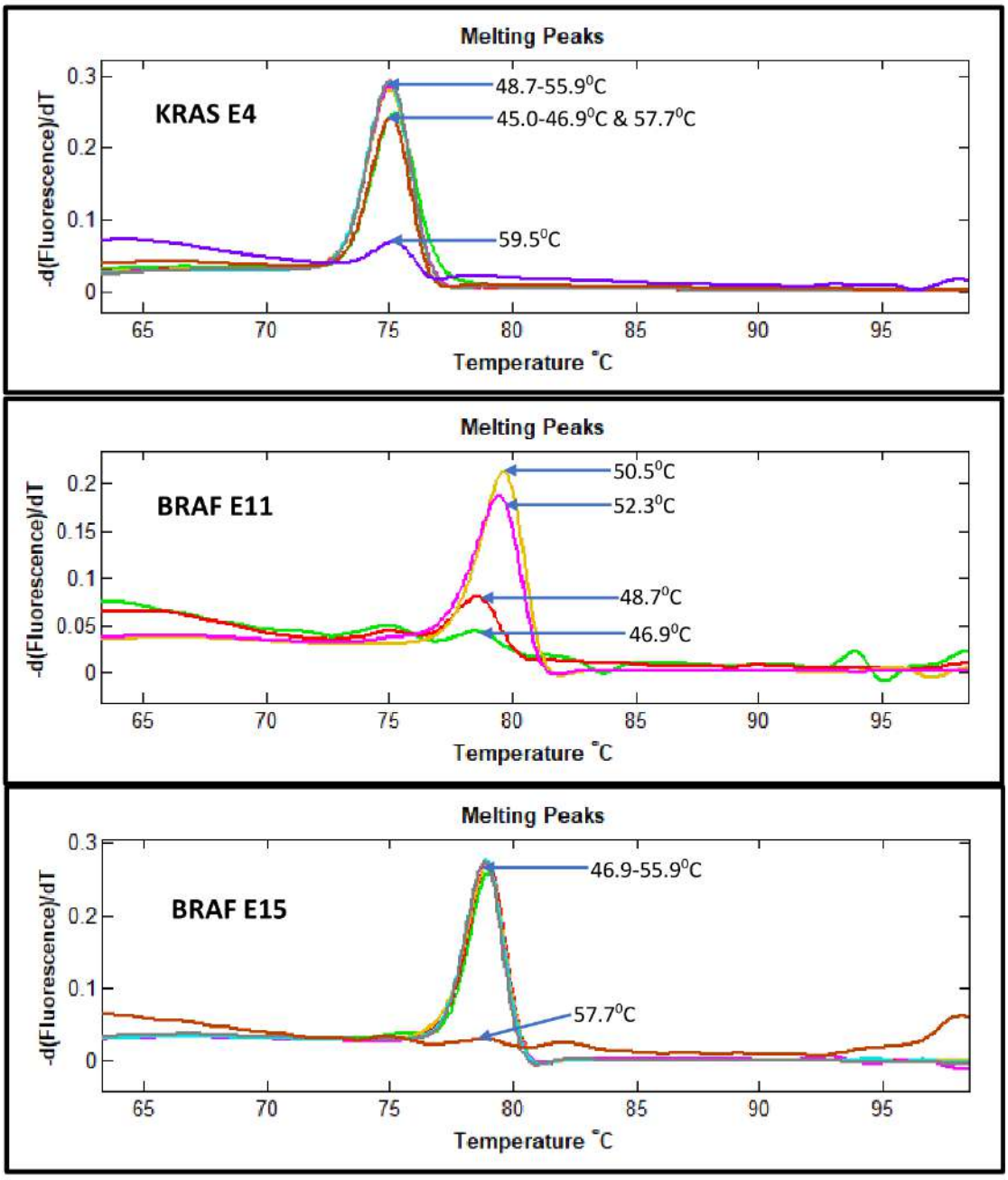


Figure 2.2: Gradient PCR-HRMA of some of the primers used in the development of the mHRMA assay. The optimum Ta's are the temperature curves with the highest peaks.

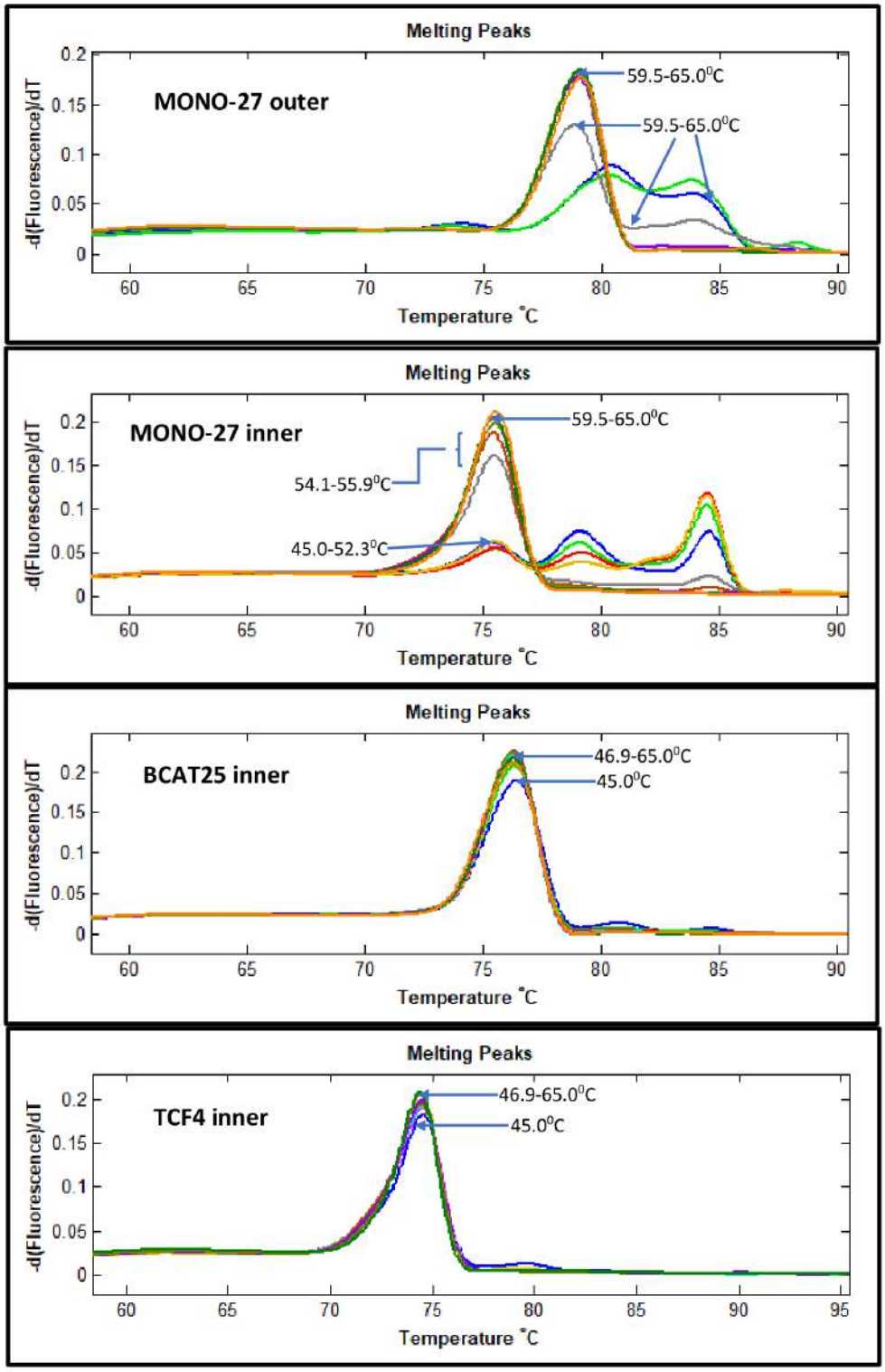


Figure 2.3: Gradient PCR-HRMA of some of the primers used in the development of the MSI-HRMA assay. The smaller peaks seen in some of the temperature curves represent non-specific PCR products when PCR is run at the annealing temperatures represented by curves with the multiple peaks.

2.2.7 Primer Validation

Primer validation is the process of ensuring that the primer pairs can detect the mutations for which they are designed. This process involves the amplification of DNA templates with known mutations – usually cell lines with comprehensive mutation profiles or clinical samples with known mutations – using the optimised primer pairs to detect the specific mutations. In this PhD project known CRC cell line and clinical samples with known mutation profiles were used to validate the designed primers using PCR followed by HRMA detection (Figures 2.4, 2.5, 2.6). Primer validation represents the analytical validation aspect of the assay development. It tests the analytical accuracy, precision, sensitivity (limit of detection) and specificity of the primers. All the primer validation procedures done in this work included positive, negative (or normal), as well as no-template controls. Also, the samples were run in replicates in all the experiments performed, except otherwise indicated.

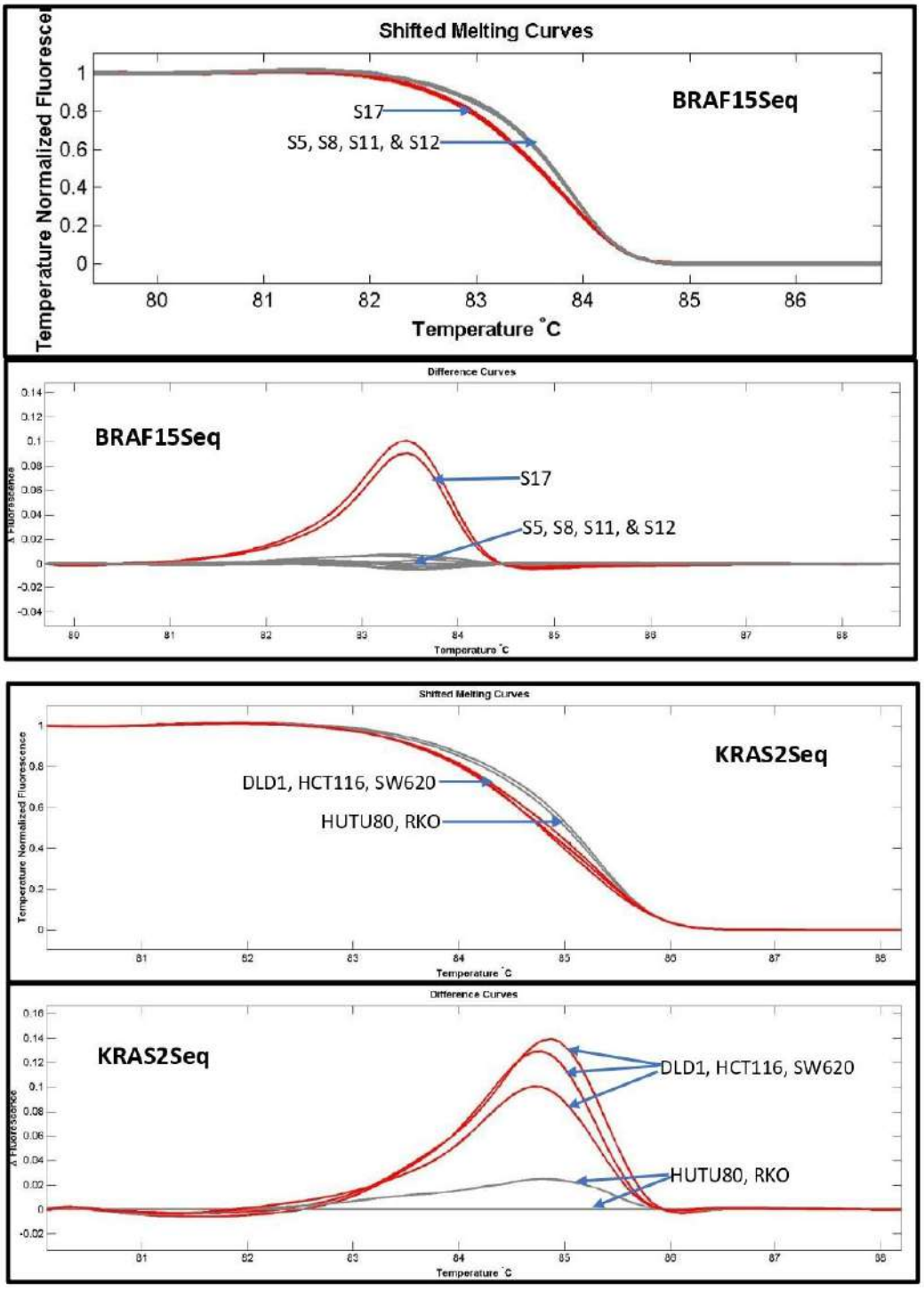


Figure 2.4: Squirrel primer validation was performed to determine if the designed primers can detect the target mutations when such are present in the samples. BRAF15Seq and KRAS2Seq accurately detected mutations (or the absence of it) in standard clinical samples and cell lines.

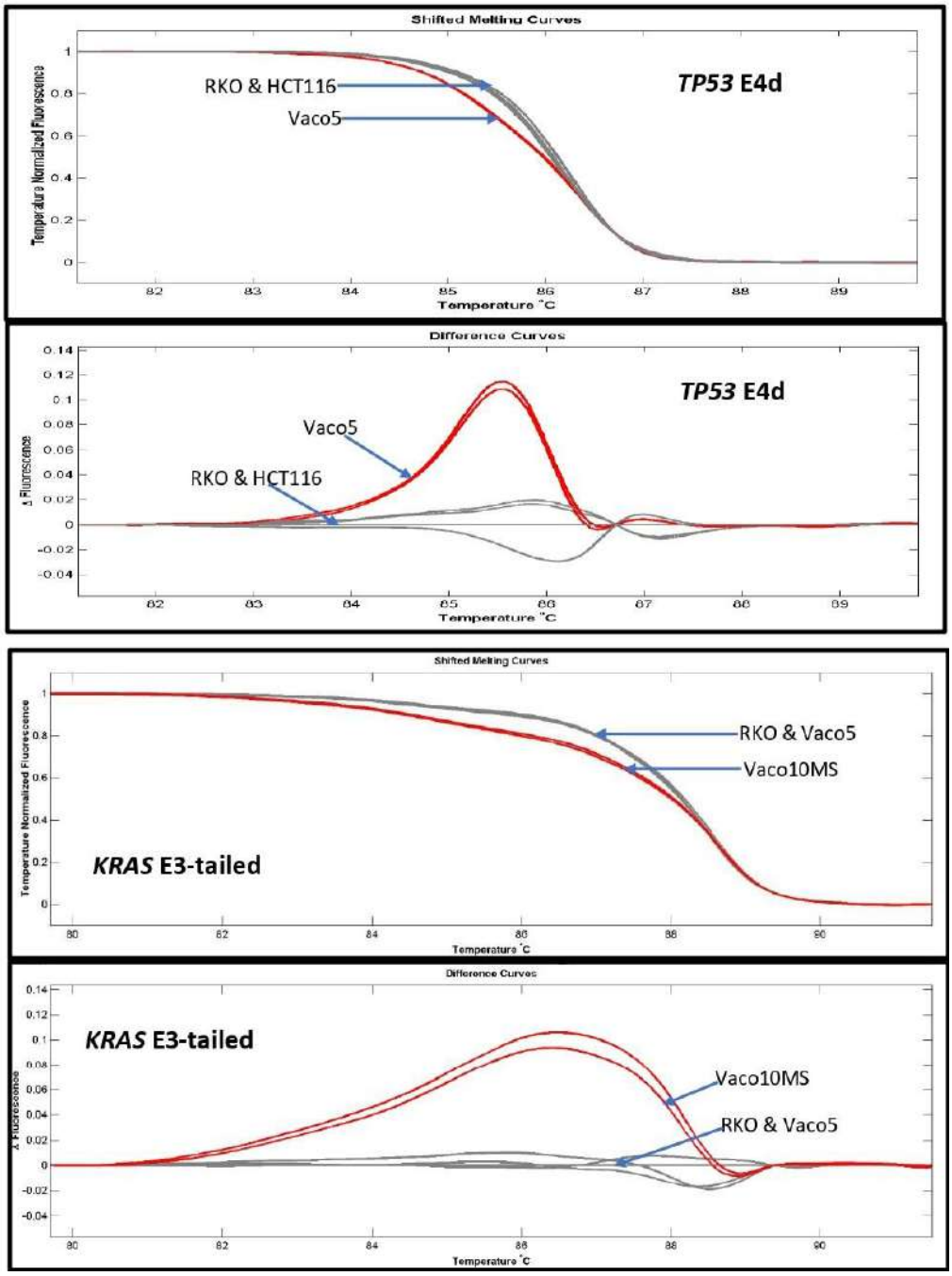


Figure 2.5: Validation of the primers used in the development of the mHRMA assay by PCR-HRMA. Only the results of the validation of *TP53* exon 4d and tailed *KRAS* exon 3 primers are shown here. See Appendix images for more results.

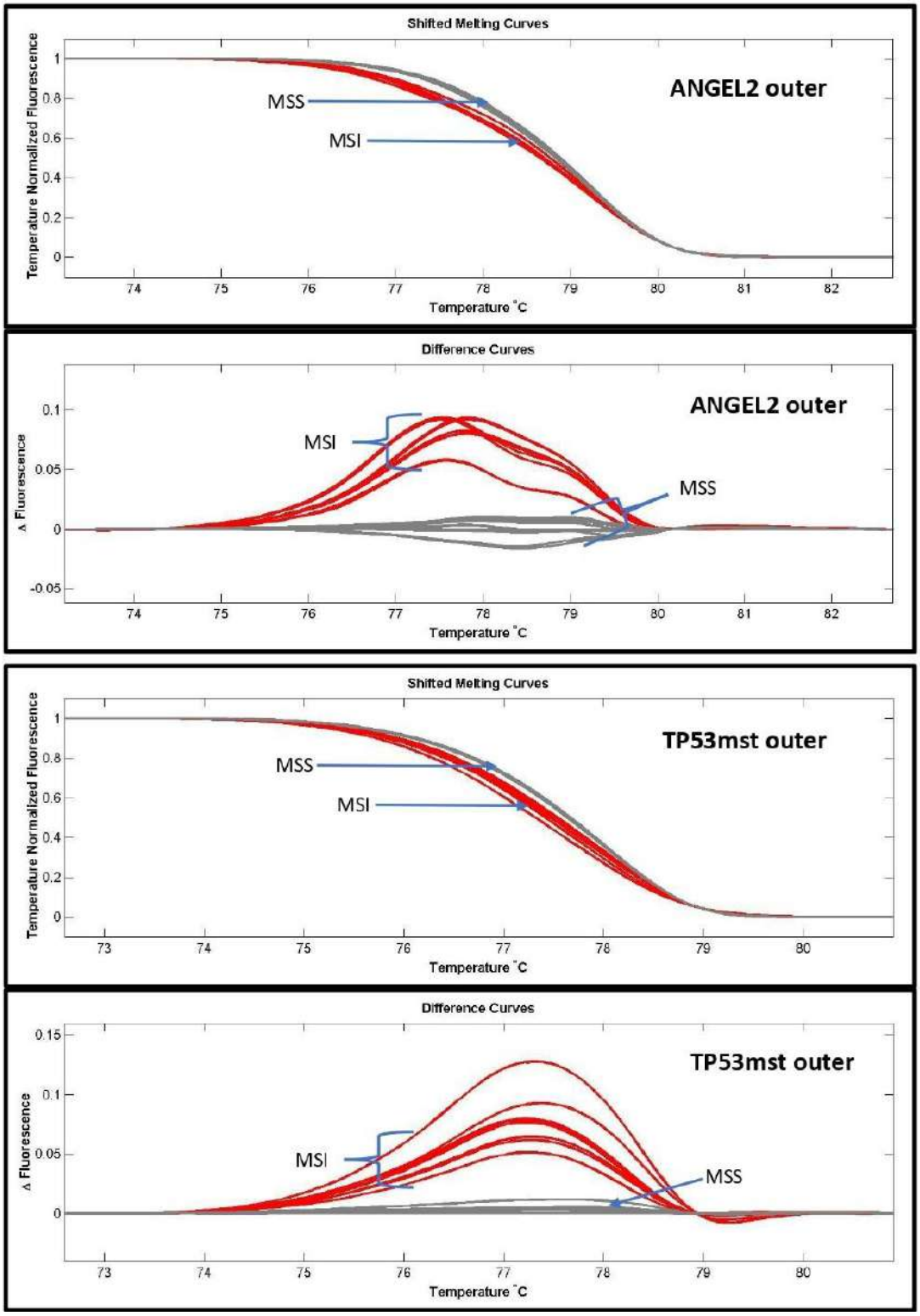


Figure 2.6: MSI-HRMA primers validation. PCR-HRMA validation of the ANGEL2 and TP53mst microsatellite markers outer primer pairs using MSI and MSS cell lines.

2.2.8 Immunohistochemistry

Only the data from immunohistochemistry (IHC) works were analysed for this PhD study. This candidate did not perform the actual work of tissue microarray preparation or staining of the sections.

IHC was performed at the Diagnostic Histopathology Department of the Nottingham University Hospitals NHS Trust, Nottingham, UK, using compact polymer technology on an automated BOND-MAX (Leica Microsystems, Germany). The MMR panel comprises four antibodies - MLH1 (clone ES05; Leica, cat. no.NCL-L-MLH1), MSH2 (clone 25D12, Novocastra, cat. no.NCL-MSH2), MSH6 (clone 44/MSH6, BD Transduction Labs, cat. no.610918), and PMS2 (clone A16-4, BD Pharmingen, cat. no.556415).

Whilst the MMR panel immune-staining of the MSI-HRMA training cohort was done as part of the routine clinical work-up of CRC cases, the staining of the validation cohort for MMR panel was specifically performed for our group research – the Molecular Pathology Research Group headed by Prof Mohammad Ilyas - as part of staining of a larger panel of biomarkers.

2.2.8.1 Tissue Microarray Array (TMA) Preparation

TMA blocks were produced from the FFPE tissue blocks of the MSI-HRMA validation cohort using the fully automated TMA arrayer, TMA GrandMaster® (3D HISTECH®, Budapest, Hungary). Haematoxylin and eosin-stained slides of the validation cohort (n=122 cases) of the MSI-HRMA assay were reviewed for tumour areas. The luminal, middle, and advancing edge of tumours, and non-tumour colon tissues were marked on the slides and on the corresponding FFPE tissue blocks (donor blocks). Array recipient blocks were prepared by applying molten paraffin into a mould. Then, the paraffin was mounted on histopathological tissue cassettes and

incubated on ice to set. The excess wax was trimmed off until the recipient blocks became flat surfaced and ready to receive the cores. The blocks were then placed into the TMA arrayer. Three cores of tumour (plus one core from adjacent normal tissue, if available), with a core size of 0.6 mm were taken from each donor block and mounted and assembled into the recipient block.

2.2.8.2 IHC protocol

The IHC protocol at the Diagnostic Histopathology Department of the Nottingham University Hospitals NHS Trust is fully automated, and all the immuno-staining steps from de-waxing to counterstaining were performed by the BOND-MAX machine. Three micron of CRC tissue sections were cut, placed on microscopic glass slides, air-dried at room temperature for 20 minutes, then incubated at 60°C for 20 min before they were loaded into the automated BOND strainers.

The following optimal dilutions for staining were obtained for the MMR panel: 1:100 (MLH1, Epitope Retrieval Solution 1 (ER1) applied for 30minutes), 1:50 (MSH2, ER2 for 30 minutes), 1:100 (MSH6, ER1 for 20 minutes), 1:100 (PMS2, ER2 for 30 minutes). Sections of the appendix were used as positive controls for the MMR panel.

The protocol used for the MMR panel staining is as follows: de-waxing with Leica Dewax solution (AR9222) was set for 30s at 72°C, followed by antigen retrieval step with the Epitope Retrieval solutions (see above); washing was accomplished in Leica Bond Wash Solution (AR9590), following which the slide sections were immersed in peroxide block solution (Kit DS9800) for 5 minutes at room temperature to block endogenous peroxidase; incubation of the sections with primary antibody was carried out for 15 minutes at room temperature, while

incubation with post-primary block and the polymer detection system (ChromoPlex 1 Dual Detection system) were each applied for 8 minutes and at room temperature; following incubation with polymer detection system, 3,3'-Diaminobenzidine (DAB) was applied to the slide sections for 10 minutes at room temperature followed by the DAB enhancer (AR9432) for 5 minutes at room temperature; the slide sections were counterstained with haematoxylin for 5 minutes at room temperature.

Following counterstaining with haematoxylin, the slide sections were removed from the BOND staining machine, dehydrated in three baths of 100% industrial methylated spirits (Genta Medical), cleared in xylene (Genta Medical) and permanently mounted under glass coverslips with Pertex (Histolab).

All wash steps between each major step were performed using Leica Bond Wash solution (AR9590). Peroxidase block, post-primary block, polymer, DAB and haematoxylin were all supplied in the Leica BOND Refine Detection kit (DS9800).

2.2.8.3 Whole Slide Imaging and scoring of the MMR panel-stained TMA sections

The stained TMA sections were scanned (at X40 magnification) using the NanoZoomer Digital slide scanner (Hamamatsu, Japan). Slide digital sections were viewed and scored with the 3DHISTECH CaseViewer.

The IHC slides for the validation cohort was scored by using the following stringent criteria as previously described (Fadhil and Ilyas, 2012, Klarskov et al., 2010, Shia et al., 2011). Each of the antibodies was scored individually and only the nuclei staining were assessed. Individual cores were scored as “positive” if the nuclear staining in the tumour cells was stronger than the nuclear staining in the adjacent stromal cells, “negative” if there was no nuclear staining in the

tumour cells but there was positive nuclear staining in the adjacent stromal cells, or "unclassifiable" if there was no nuclear staining in either the stromal cells or the tumour cells. A sample or case was deemed deficient for an MMR protein if that protein expression was negative in the tumour cells of ALL THREE CORES. Once the scoring of all the antibodies was completed, the entire MMR protein panel was assessed for each case or sample. The established patterns of expression of the binding partners MLH1-PMS2 and MSH2-MSH6 were used as a second level of scoring criteria: loss of nuclear MLH1 expression is almost always associated with loss of PMS2 but not vice versa, while the loss of MSH2 expression is almost always associated with the loss of MSH6 but not vice versa. Therefore, a tumour sample was only regarded as negative for MLH1 expression if it was also negative for PMS2. Similarly, negative staining for MSH2 was only accepted if there was negative MSH6 staining. However, individual loss of either PMS2 or MSH6 was accepted if the other criteria were fulfilled. A sample or case was deemed as MMR-deficient (dMMR) if it showed negative expression of one or more of the MMR proteins as described above, and MMR-proficient if it showed positive expression of all four MMR proteins.

2.2.9 Capillary Electrophoresis

Standard MSI analysis data for the MSI-HRMA training cohort was obtained from the Molecular Genetics Laboratory at Nottingham University Hospitals NHS trust. These data had been generated from the routine clinical work of the laboratory with fluorescent PCR followed by capillary electrophoresis (CE) by using the Promega MSI system v1.2 (5 microsatellite markers: BAT-25, BAT-26, NR-21, NR-24, MONO-27; and 2 identity markers: Penta C and Penta D). CE was performed on the Applied Biosystems 3130xl Genetic Analyzer (Life Technologies) using

the kit internal lane standard. Data analysis was accomplished with the GeneMapper software. Any sample with instability at ≥2 mononucleotide loci was adjudged as microsatellite unstable (i.e., MSI).

2.2.10 Sanger sequencing

All the Sanger sequencing works done for this PhD project were performed at the University of Nottingham DNA Sequencing Laboratory, School of Life Sciences, Queen's Medical Centre, Nottingham. Sanger sequencing was performed with the dye terminator chemistry (BigDye version 3.1) on the 3130xl ABI PRISM Genetic Analyzer (Applied Biosystems, United Kingdom). The sequencing data was viewed and analysed using the freely available online Finch TV software.

2.2.11 Genomic Data Analysis

The annotated somatic mutation files in vcf formats, the gene expression files, and the DNA methylation files were all downloaded alongside their respective sample sheets. File sorting and genomic data analysis was performed in cygwin.

2.2.11.1 Annotated somatic mutation data

File sorting

For the annotated somatic mutation data, vcf files were sorted for individual cases by first matching the file names to the TCGA case ID's in the sample sheet. Then the files were moved into individual case folders (directories) named with TCGA case ID by using the following codes:

```
mv filename.vcf  TCGA-corresponding-caseid
```

From each individual case folder (directory) the VarScan2 files for the 441 cases were selected from among the four variant callers using the following codes:

```
cd TCGA-corresponding-caseid  && cp $(grep -lir 'varscan2' *.vep.vcf) TCGA- corresponding-caseid.varscan2.vcf  &&  cd ..
```

Only the varscan2 files were used to demonstrate the MSI-WES approach.

File processing and quality control

The varscan2 files were renamed GCa#1 to GCa#441 for the purpose of this study by using the following code:

```
cp TCGA-corresponding-caseid.varscan2.vcf  GCa#.varscan2.vcf
```

To enable faster computation, the varscan2 vcf files were stripped of non-essential information, including the headers. For quality control only the variants marked "SOMATIC" and have a "PASS" filter were included for analysis. The following combination of codes was used to strip the vcf files of non-essential and to assure quality:

```
grep –v "##"  GCa#.vcf.varscan2.vcf  >  GCa#.vcf
grep -w   '#CHROM\|PASS'  GCa#.vcf  | grep '#CHROM\|SOMATIC'  > GCa#.vcf.clean.vcf && cut -f1,2,4,5  GCa#.vcf.clean.vcf  > GCa#.vcf.4column.vcf
```

With the above codes, only the information about the genomic coordinates (chromosome numbers and positions) and the reference and alternative (mutated) bases were left in the processed files for further analyses.

Verification of the microsatellite contexts of the small indels

All 441 processed and quality-controlled vcf files (*.4column.vcf) were moved into one folder (directory) for further analyses. Then, a list of potential vcf-based indels was compiled and stored as .txt file (MSILesions.txt), following which the genomic coordinates of all the small indels were retrieved from the 441 files using the code:

```
grep –wf  MSILesions.txt  *.4column.vcf  > MSI.coordinates.csv
```

The obtained coordinates were input into online software, RefSeq Downloader (https://github.com/Ghahfarokhi/NCBI_Batch_Download/blob/master/1_RefSeq_Downloader_v1.xlsm), to retrieve the flanking sequences of the small indels from the NCBI genome database.

Per case microsatellite-based indels counting

The indels obtained per case from the computation above (in the MSI.coordinates.csv file) were counted using the code below:

```
sed 's/\:/\t/g' MSI.coordinates.csv  > MSI.coordinates.tab.csv  && cut –f1
MSI.coordinates.tab.csv  | sort | uniq –c  > MSI.count.csv
```

The per-case MSI counts were expressed as a fraction of 3088Mb (the size of the genome that was sequences) to obtain the MSI index_ngs.

Mutation targets of MSI

The somatic mutations of the genes *TGFBRII, IGFIIR, TCF4, RIZ, BAX, CASPASE5, FAS, BCL10, APAF1, hMSH6, hMSH3, MED1, RAD50, BLM, ATR,* and *MRE11* (Ratti et al., 2018) were retrieved per case from the 441 processed file using their genomic coordinates which were obtained from ENSEMBL (www.ensembl.org/) (Appendix Table 6.1). The codes below enabled the retrieval of the somatic variants of *TGFBRII* from the vcf files:

```
awk ‘$3==”chr3” && $2>30606600 && $2<30694142) {print $1, $2, $3, $4}’
GCa#.vcf.4column.vcf  > GCa#.vcf.TGFBRII.mutations.vcf
```

```
cat *.TGFBRII.mutations.vcf  > TGFBRII.mutations.vcf
```

Somatic mutations of the *TGFBRII* were then counted using the code below:

```
cut –f1 TGFBRII.mutations.vcf  | sort | uniq –c  > TGFBRII.mutations.count.csv
```

Somatic mutations in other MTM genes were similarly retrieved and counted.

2.2.11.2 Gene Expression data

File sorting

Level 3 RNASeq data was available for 379 cases and were downloaded .txt file format. Each .txt file contained data for 60,483 transcripts. The individual case files were sorted by matching the file names to the TCGA case ID's in the sample sheet. The files were then renamed with the appropriate study identifier:

```
cp filename.FPKM.txt TCGA-corresponding-caseid.FPKM.txt
cp TCGA-corresponding-caseid.FPKM.txt  GCa#.FPKM.txt
```

MMR gene expression

The ENSEMBL gene identifier of *MLH1, MSH2, MSH3, MSH6, PMS1,* and *PMS2* (Appendix Table 6.2) were retrieved from www.ensembl.org/ and used to retrieve the gene expression levels in FPKM. The following code retrieved the gene expression from all the renamed case files:

```
grep -w  'ENSG00000076242'   *.FPKM.txt  >  MLH1_expression.csv
grep -w  'ENSG00000095002'   *.FPKM.txt  >  MSH2_expression.csv
grep -w  'ENSG00000113318'   *.FPKM.txt  >  MSH3_expression.csv
grep -w  'ENSG00000116062'   *.FPKM.txt  >  MSH6_expression.csv
grep -w  'ENSG00000122512'   *.FPKM.txt  >  PMS2_expression.csv
grep -w  'ENSG00000064933'   *.FPKM.txt  >  PMS1_expression.csv
```

2.2.11.3 DNA Methylation Data

File sorting

DNA methylation beta value data was available for all 441 cases and 27 normal samples as .txt files. File sorting was achieved by matching the filenames to the TCGA case ID's on the sample

sheet, as before. The files were then renamed with the matching TCGA case ID's using the code below:

```
cp filename.txt TCGA-corresponding-caseid.methylation.txt
cp TCGA-corresponding-caseid.methylation.txt  GCa#.methylation.txt
```

Then to reduce computation time only the information essential to further analysis were included in the methylation file per case.

```
cut –f1,2,6,10,11 GCa#.methylation.txt  > GCa#.methylation.5column.txt
```
(#leaves only the probe numbers, beta values, gene symbols, CpG Island coordinates and feature type [or type and position of the regulatory elements] in 5 columns).

MLH1 promoter methylation

The per-case DNA methylation beta values record contains multiple entries for per gene, each entry representing different gene promoter – or other regulatory – regions that has been probed for methylation status. The methylation status per region is stated as a beta value and range from 0.0 to 1.0, with the lower values denoting normo- or hypo-methylation while the higher values denote hypermethylation. The specific promoter region is represented with a numbered probe. As not all the promoter or regulatory regions control gene expression a good strategy that has been adopted for methylation study by many researchers is to probe for promoter regions that consistently show hypo- or normo-methylation (using a beta value of 0.3 as a threshold for normo-methylation (Losi et al., 2018; Warden et al., 2013) in normal DNA samples, and then test these regions or probes in tumour DNA. In this study the following codes were used to retrieve the probes (promoter or regulatory regions) with consistent normo-methylation in normal DNA files:

```
grep –w "MLH1"  *.methylation.5column.normal.txt  > MLH1.methylation.normal.txt
```
(# this codes outputs all the *MLH1* entries in all the normal sample files into one file).

sed 's/\:/\t/g' MLH1normprobes.methylation.normal.txt > MLH1.methylation.normal.tab.txt &&
sed 's/\,/\t/g' MLH1.methylation.normal.tab.txt > MLH1.methylation.normal.tab1.txt &&
awk '$2<=0.3 {print $1, $2, $3, $4, $5}' MLH1.methylation.normal. tab1.txt >
MLH1normprobes.methylation.normal. tab1txt && cut –f2
MLH1normprobes.methylation.normal.tab1.txt | sort | uniq –c > MLH1normoprobe.count.txt
(# retrieves all the MLH1 methylation probes with beta values of <=0.3 (normo-methylation) and counts how many times the normo-methylated probe is found in the cohort of normal DNA samples, and thus show which probes have normo-methylation consistency or otherwise).
The beta values of the identified probes were sought in tumour files using the codes below and thus tumour samples were scored as either hypermethylated or hypo-/normo-methylated at these probe regions:

rep –w "cg00893636" *.methylation.5column.tumour.txt >
cg00893636.MLH1.methylation.csv

grep –w "cg02279071" *.methylation.5column.tumour.txt >
cg02279071.MLH1.methylation.csv

grep –w "cg13846866" *.methylation.5column.tumour.txt >
cg02279071.MLH1.methylation.csv
grep –w "cg18320188" *.methylation.5column.tumour.txt >
cg02279071.MLH1.methylation.csv

(# the probes cg00893636, cg02279071, cg13846866 and cg18320188 were identified in this study).

2.2.12 Statistical Analyses

All statistical analyses for the PhD study were performed in SPSS version 21. Associations between two or more categorical variables were evaluated using the Chi Square (or Fisher's) test, while correlations between continuous variables were tested with Univariate or Multivariate Analyses. Correlations between categorical and continuous variables were sought with the Eta statistics while concordances between newly designed and existing assays were calculated using the Cohen's Kappa statistics. Crude concordances (mHRMA assay) and the Quality of Sequence (Squirrel PCR assay) were manually calculated. Analytical performances of designed assays

were calculated using the online MEDCALC statistical software at https://www.medcalc.org/calc/diagnostic_test.php. Unless otherwise stated, a *P* level of <0.05 was regarded as significant. Correction for multiple testing at a false discovery rate of 5% was accomplished with Benjamini-Hochberg test, where applicable using the online software Seed-based *d* Mapping software at https://www.sdmproject.com/utilities/?show=FDR.

2.3 Ethical Approval

This project was carried out in accordance with the guidelines of the Research Ethics Committee, and all experimental works were undertaken with local institutional approval. The clinical cases used in the development of the Squirrel PCR and mHRMA assays were kindly provided Prof Mohammad Ilyas (Professor of Pathology, School of Medicine, University of Nottingham, Division of Pathology; and Queen's Medical Centre, Nottingham University Hospitals NHS Trust). Blocks from CRC resection specimens were chosen to contain at least 50% tumour tissue, and ethical approval was obtained for the use of patient materials (Research Ethics Committee reference 05/Q1605/66) (Fadhil et al, 2012). Access to tissues and ethics approval for the development of the MSI-HRMA assay were granted by Nottingham Health Sciences Biobank, which has approval as an IRB from Northwest—Greater Manchester Central Research Ethics Committee (REC reference: 15/NW/0685) (Susanti et al, 2018). Furthermore, genomic data access and analyses were approved by the International Cancer Genome Consortium (ICGC), dbGap and the University of Nottingham (ICGC reference number DAC0-1081299, NIH approval number # 88362-1).

CHAPTER 3

SQUIRREL PCR ASSAY FOR ROBUST SEQUENCING OF ULTRA-SHORT DNA SEGMENTS

3.1 Background

Sanger method, also known as the chain termination method, was developed in 1977 by Frederick Sanger, an English biochemist, to determine the order of nucleotide bases in a piece of DNA molecule (Sanger et al., 1977; Shendure et al., 2019). Over the years the detection technique for the Sanger method has evolved from gel electrophoresis to its current automated or semi-automated capillary electrophoresis form (Shendure et al., 2019). In this era of massively parallel sequencing or next-generation sequencing (NGS) Sanger sequencing has remained relevant for sequencing single genes or amplicon targets of up to 1000 bases (Crossley et al., 2020; Rajendra, 2012; Rehm et al., 2013; Sikkema-Raddatz et al., 2013). In its present form Sanger method is the most practical approach for small sequencing projects involving a few samples and a few gene targets (Crossley et al., 2020; A. Hamilton et al., 2016), and it is the most robust technique for gene fragment analysis (Lian et al., 2021; Long et al., 2019; H. Zhang et al., 2015) and short tandem repeat analysis (Aranake-Chrisinger et al., 2016; Goh et al., 2009; M. Lee et al., 2019), and sequencing of disease-associated genome regions with extended segments of repeat units or high guanine-cytosine (GC) contents, which genomic regions are poorly sequenced by NGS (Chen et al., 2013; Hagemann, 2015; Kieleczawa, 2006). However, Sanger method has also been applied to high-scale sequencing, such as was used in the Human Genome Project (Dewey et al., 2012; França et al., 2002; Lander et al., 2001; Rajendra, 2012). It is the gold standard for the detection of single nucleotide polymorphisms (SNPs), somatic point mutations and insertion-deletion (indel) mutations in clinical oncology practice (Crossley et al., 2020; Dewey et al., 2012; A. Hamilton et al., 2016); as long as the number of per-tumour or per-patient gene targets for personalised cancer therapy or disease prognosis is still limited to one or a few gene targets (Jin et al., 2019; Lima et al., 2019; Malone et al., 2020). Therefore, Sanger

method remain the most cost-effective and practical approach for molecular diagnostic laboratories in both resource-sufficient and –constrained settings to accomplish their nucleic acid sequencing needs. Furthermore, even where NGS is utilised for nucleic acid sequencing current best practices demand that validation of critical NGS findings is performed with Sanger method (Dewey et al., 2012; A. Hamilton et al., 2016; Mu et al., 2016; Rehm et al., 2013).

To perform standard Sanger sequencing of a genomic segment of interest, the segment is first amplified by PCR, and the generated amplicons are then subjected to the chain termination method using the PCR primers as the sequencing primers (CD Genomics, 2020; Clark & Pazdernik, 2013; Crossley et al., 2020; Hagemann, 2015; Heather & Chain, 2016). Sanger method is used for sequencing genomic segments that are 100-1000bp long (Crossley et al., 2020; Rajendra, 2012). However, the characteristic poor-quality peaks observed in the first 15-40 bases of any DNA molecule sequenced by Sanger method places some limitations on the template size that can be sequenced by the chain termination method (Binladen et al., 2007; Crossley et al., 2020; Rajendra, 2012). Standard Sanger sequencing techniques cannot be used to sequence nucleic acid molecules that are below 100 bases in length because the method is error-prone in the first 40 odd bases - in the forward and reverse directions – and gives rise to poor quality base peaks that are difficult to interpret on a Sanger sequencing electropherogram (Binladen et al., 2007; Crossley et al., 2020; Rajendra, 2012). This causes extensive loss of data and creates a huge challenge for the Sanger validation or confirmation of mutations detected, for example by PCR-based methodologies, using degraded nucleic acid templates such as from formalin-fixed, paraffin-embedded (FFPE) tissue (Dietrich et al., 2013; Gilbert et al., 2007; Ludyga et al., 2012; Srinivasan et al., 2002) and cell-free and circulating tumour DNA (ctDNA) (Alcaide et al., 2020; X. Liu et al., 2020; Mouliere & Rosenfeld, 2015; Underhill et al., 2016).

Conventionally, to sequence such ultra-short DNA molecules laboratories rely on either direct sequencing by NGS and Pyrosequencing method (Binladen et al., 2007; Fakruddin & Chowdhury, 2012; Gharizadeh et al., 2007; L. Liu et al., 2012; Rehm et al., 2013), or indirect sequencing by the plasmid cloning technique (Crossley et al., 2020; Jajesniak & Wong, 2015; Li et al., 2011; Tao et al., 2020). However, the use of both NGS and Pyrosequencing require huge investments in expensive new equipment and sequencing reagents, which is impractical for most molecular diagnostic laboratories considering that their routine operations only target a limited number of genes (Binladen et al., 2007; Jin et al., 2019; Lima et al., 2019; L. Liu et al., 2012; Malone et al., 2020) using Sanger sequencing. Plus, the limited versatility of the pyrosequencing platform, i.e., limited to DNA sequencing, relative to capillary electrophoresis platforms, makes it less attractive from the economic point of view. For the same reason of prohibitive cost of sequencing, the reliance on the NGS would stifle the progress of current and emerging molecular research and diagnostic fields in which the available templates are naturally short (Alcaide et al., 2020; Dietrich et al., 2013; Gilbert et al., 2007; X. Liu et al., 2020; Ludyga et al., 2012; Mouliere & Rosenfeld, 2015; Srinivasan et al., 2002; Underhill et al., 2016). Sequencing by plasmid cloning is accomplished by the insertion of the DNA segment of interest into a plasmid vector with which bacteria - usually E. coli - are then transformed. The transformed bacteria are then grown exponentially in culture, following which the replicated plasmids are extracted and the DNA inserts are sequenced using plasmid-specific primers. Insertion of the DNA segment of interest into the plasmid serves two purposes: (i) the flanking plasmid sequences ‘elongate’ the DNA target and serve as matrices for its sequencing, and (ii) multiplication of the plasmid resulting from the replication of the bacteria creates many copies of the DNA insert and an adequate amount of templates for sequencing (Crossley et al., 2020; Jajesniak & Wong, 2015; Li

et al., 2011; Rajendra, 2012; Tao et al., 2020). However, as a technique plasmid cloning is cumbersome, laborious, and prone to repeated failures (Jajesniak & Wong, 2015; Li et al., 2011; Tao et al., 2020).

In this chapter I describe a novel Sanger sequencing-purposed PCR assay: the Squirrel PCR assay for enabling targeted and direct sequencing of ultra-short DNA segments. In this assay pairs of long-tailed primers – Squirrel primers – were used to PCR-amplify sub-100 bp DNA segments, following which the universal sequencing primers – USeq primers - which bind to the 5' ends of the long tails were used to perform Sanger sequencing of the amplified DNA segments. I showed that this assay can permit the robust sequencing of 51 to 93 bp DNA segments.

3.2 Materials and Methods

3.2.1 Principles/Concepts

As Sanger sequencing shows poor quality electropherograms in the initial 15-40 bases, i.e., at the 5' ends of the forward and reverse sequences, it would be impractical to directly sequence DNA segments of 50-100 bp using standard Sanger sequencing strategy, as this would produce unreliable and probably un-interpretable electropherograms (Binladen et al., 2007; Fakruddin & Chowdhury, 2012; Rajendra, 2012). However, by elongating the 5' ends of the forward and reverse sequences of the short DNA segments using PCR primers with long tails of non-specific sequences, the target DNA region is shifted further away from the 5' ends of both the forward and reverse strands in the resulting PCR amplicons.

As sequencing reads commence at the primer binding site of the sequencing primers, the use of a short (20 bases or less) sequencing primer pair which bind at the 5' regions of the long tails of

the PCR primers would ensure that the read commences within the non-specific 5' end of the PCR amplicons and produce good quality reads over the target regions (which would now be farther away from the 3' binding site of the sequencing primers in the forward and reverse sequences of the amplicons) (Figure 3.1A).

3.2.2 Primer Design

The "Squirrel" primers consist of 18 to 26-nt long specific sequences and 40 to 65-nt long non-specific 5' tail sequences (Figure 3.1A). The specific sequences were designed manually using a combination of ENSEMBL(Aken et al., 2016), MFEprimer (Qu et al., 2012), UCSC in silico PCR (Speir et al., 2016) and Reverse Complement (Stothard, 2000), to amplify short genomic segments which encompass codons 12 and 13 of *KRAS* exon 2, codon 61 of *KRAS* exon 3, codons 592 to 600 of *BRAF* exon 15, codons 1043 and 1047 of *PIK3CA* exon 20, and codons 60 and 61 of *PTEN* exon 3.

Four Squirrel primer tail lengths were designed to test the ability of the assay to permit Sanger sequencing of short genome segments: L40 (40-base long tail), L50 (50 bases), L58 (58 bases) and L65 (65 bases). The L40 forward sequence was copied from published literature (Binladen et al., 2007; Lindblad-Toh et al., 2000; Sanchez et al., 2003). The reverse sequence was derived by the scrambling 40 bases together. The L50, L58 and L65 sequences were designed by the addition of non-specific sequences to the 3'ends of the L40. L40 –L65 sequences were then checked on BLASTn (Johnson et al., 2008) to exclude any specificities, in the absence of which they were 'tagged' to (and synthesised with) the specific primer sequences to form "Squirrel" primers.

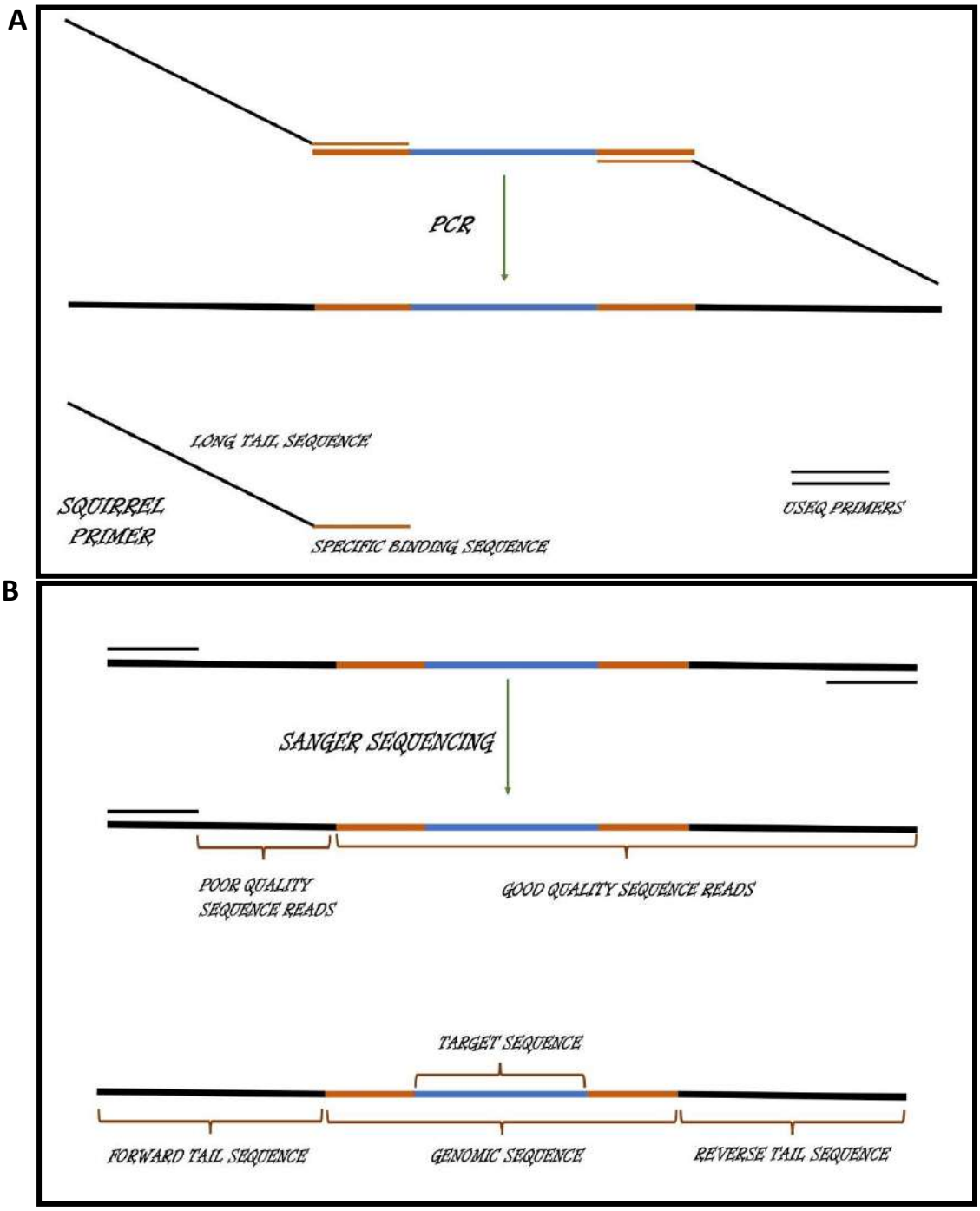


Figure 3.1. Squirrel PCR assay strategy **A**. Diagram showing the elongation of the target DNA segment using a pair of Squirrel primers in a PCR reaction and parts of the Squirrel primer **B**. Diagram showing the sequencing of the PCR product using USeq primers and definition of terminologies used in this chapter.

To test the robustness of the Squirrel PCR against different gene targets with different primer tail lengths, the L40 tail was tagged to the *KRAS* exon 2, *KRAS* exon 3 and *PIK3CA* exon 20 primers to form respectively the KRAS2Seq, KRAS3Seq and PIK20Seq squirrel primers; L50 was tagged to the *BRAF* exon 15 primer (BRAF15Seq), while L58 and L65 were tagged to the *PTEN* exon 3 (PTEN3Seq) and *KRAS* exon 2 (KRAS2LSeq) primers, respectively. Sequencing primer pairs (USeq primers) were also designed to bind to the 5'region of the long tail sequences. The genomic regions, primer sequences, primer tail lengths, sizes of genomic DNA and target sequences (see 'Sequencing Analysis' below for definitions) and sizes of the final PCR products are shown in Tables 3.1 and 3.2.

Table 3.1: Squirrel and USeq primers characteristics

Primer names	*Primer sequences	Tail lengths	Genomic region amplified
KRAS2Seq	F:5'-aactgactaaactaggtgccacgtcgtgaaagtctgacaaTGAATATAAACTTGTGGTAGTTGGR:5'-tagactcctgatcccttcattgccctgcatctgacacgcaGCTGTATCGTCAAGGCACTCT	40nt each	*KRAS* exon 2
BRAF15Seq	F:5'-aactgactaaactaggtgccacgtcgtgaaagtctgacaagggccggcccTTCATGAAGACCTCACAGTAAA R:5'-tagactcctgatcccttcattgccctgcatctgacacgcagggccggcccGACCCACTCCATCGAGAT	50nt each	*BRAF* exon 15
KRAS3Seq	F:5'-aactgactaaactaggtgccacgtcgtgaaagtctgacaaCTTGGATATTCTCGACACAGCA R:5'-tagactcctgatcccttcattgccctgcatctgacacgcaTCCCTCATTGCACTGTACTCCT	40nt each	*KRAS* exon 3
PIK20Seq	F:5-aactgactaaactaggtgccacgtcgtgaaagtctgacaaGCAAGAGGCTTTGGAGTATTTCA R:5'-tagactcctgatcccttcattgccctgcatctgacacgcaCCATTTTTGTTGTCCAGCCA	40nt each	*PIK3CA* exon 20
PTEN3Seq	F:5'-aactgactaaactaggtgccacgtcgtgaaagtctgacaagtcgtgaaagtctgacaaTGTTTGTTTGTTTTGTTTTAAG R:5'- tagactcctgatcccttcattgccctgcatctgacacgcaccctgcatctgacacgcaGCATACAAATAAGAAAACATACTTAC	58nt each	*PTEN* exon 3
KRAS2LSeq	F:5'- aactgactaaactaggtgccacgtcgtgaaagtctgacaagtcgtgaaagtctgacaaccggcccTGAATATAAACTTGTGGTAGTTGG R:5'- tagactcctgatcccttcattgccctgcatctgacacgcaccctgcatctgacacgcactagctaGCTGTATCGTCAAGGCACTCT	65nt each	*KRAS* exon 2
USeq	F: 5'-ctgactaaactaggtgccacgt R: 5'-tagactcctgatcccttcat	-	-

*Tail sequences in lower case and specific primer sequences in upper case

Table 3.2: Amplicon, genomic and target sequence sizes

Primer names	Predicted Tm's[#]	Sizes of PCR products	Sizes of genomic sequences*	Sizes of target sequences*
KRAS2Seq	F:69.6 ^{0}C R:76.3^{0}C	142 bp	62 bp	17 bp
BRAF15Seq	F:76.5^{0}C R:82.0^{0}C	170 bp	70 bp	30 bp
KRAS3Seq	F: 72.3^{0}C R:76.0^{0}C	131 bp	51 bp	7 bp
PIK20Seq	F:72.4^{0}C R:75.6^{0}C	151 bp	71 bp	28 bp
PTEN3Seq	F: 71.1^{0}C R:75.5^{0}C	209 bp	93 bp	45 bp
KRAS2LSeq	F: 71.5°C R: 79.5°C	192 bp	62 bp	17 bp

[#]As predicted by Thermofisher Scientific Tm Calculator (https://www.thermofisher.com/ng/en/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html.)

*Defined in the Material and Method section under 'Sequence Analysis'.

3.2.3 Sequence Analysis

The sequencing data was viewed and analysed using Finch TV software. BLASTn operations for highly matched sequences were directly executed from the Finch TV software to determine the sequence alignments scores (using Homo sapiens GRCh38.p7 Primary Assembly) for genomic and target sequences. That is, the size of the forward and reverse genomic and target sequences that were accurately captured by the Sanger sequencing reads were noted. The electropherograms were also 'manually analysed' to confirm the results of the BLASTn operations, excluding automatic call error, and confirming the presence or absence of mutations and missed/added base peaks.

To objectively assess how robust the Squirrel PCR assay was at enabling sequencing, two (2) indices were derived: the genomic sequence quality and the target sequence quality. For the purpose of this study "genomic sequence" was defined as the sequence in the final PCR product (or amplicon) that is specific to the human genome, i.e. the portion of the amplicon sequence which is flanked by the tails of the Squirrel primers; while the "target sequence" is the genomic sequence minus the specific primer sequences, i.e. the portion of the genomic sequence which is flanked by the specific primer sequences (Figure 3.1B). The genomic/target sequence quality of the sequencing data is the proportion of the sizes of the genomic and target sequences – in the forward and reverse directions - that were accurately identified by the BLASTn operations, relative to sequence design sizes. The identified genomic sequence lengths were read off the BLASTn alignment identities, while the target sequence lengths were manually counted from the aligned query and subject sequences in the BLASTn alignment output. Moreover, the effect of the length of the Squirrel primer tail on the quality of the sequencing data was examined using Pearson's correlation analysis (SPSS version 21). Furthermore, the Squirrel PCR assay was

assessed for its ability to enable the accurate detection of mutations by Sanger sequencing. To do this Cohen's Kappa statistics (in SPSS version 21) was utilised to compare the established mutation statuses of the CRC cell lines and clinical samples (the gold standard) to the results of the Squirrel PCR-enabled Sanger sequencing. The analytical performance of the Squirrel PCR-Sanger sequencing was also calculated using the online MEDCALC statistical software (https://www.medcalc.org/calc/diagnostic_test.php).

3.3 Results

Five short (<100bp) genomic sequences representing mutational hotspots in CRC were amplified in 4 CRC cell lines with known mutations status and in 13 FFPE CRC samples with known HRM patterns using Squirrel primers, and then sequenced with the USeq primers. A total of thirty-four (34) amplicons, including 17 *KRAS* exon 2 genomic sequences (13 KRAS2Seq and 4 KRAS2LSeq), 5 *KRAS* exon 3, and 4 each of *BRAF* exon 15, *PIK3CA* exon 20 and *PTEN* exon 3, were amplified by the Squirrel PCR assay (n=30) and by the Binladen et al method (n=4) and subsequently sequenced by Sanger method.

The results showed that the Squirrel PCR assay enabled direct and targeted sequencing of mutation hotspots in genomic sequences which are less than 100 bp long. Figure 3.2 shows the sequence of a Squirrel primer-based PCR-amplified product comprising the partial tail sequence of the forward primer, the specific genomic sequence (*PTEN* exon 3), and the reverse primer tail sequence.

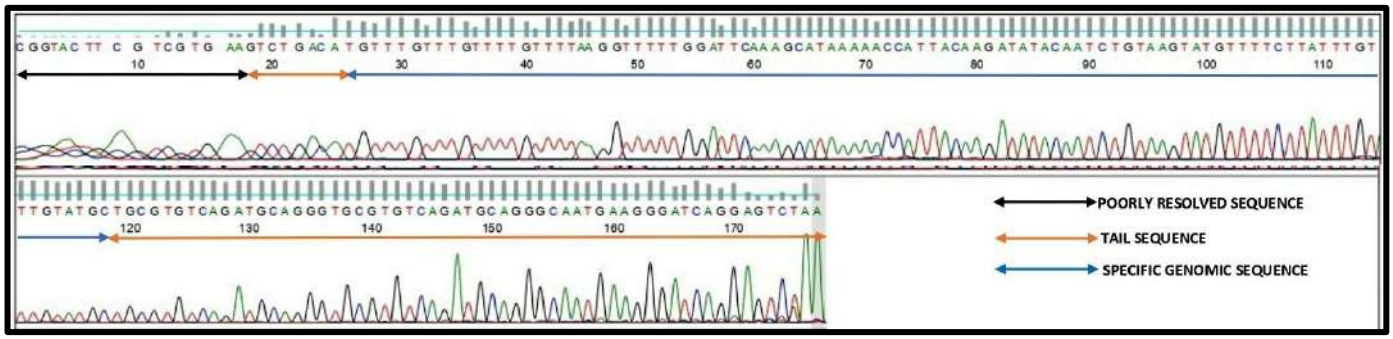


Figure 3.2: Electropherogram of *PTEN* exon 3 sequence showing segments of the poorly resolved sequence, tail sequences and the specific genomic sequence.

3.3.1 Squirrel PCR assay versus the Binladen et al sequencing strategy

Binladen et al used the tailed primers as the sequencing primer for the PCR products which had been amplified using conventional PCR primers. To compare the Binladen et al sequencing strategy with the Squirrel PCR assay, the L40 primer pairs were used to sequence 62bp *KRAS* exon 2 genomic sequence from DLD1, LS1034, RKO and SW620, and genomic and target sequence qualities obtained were compared with the results of the Squirrel PCR strategy (Table 3.3). The results showed that while the genomic and target sequence qualities obtained with the Binladen et al sequencing strategy were 51.6% - 66.1% and 58.8% - 100.0%, respectively, the Squirrel PCR assay yielded 87.1%-100% and 100% respectively for genomic and target sequence qualities (genomic sequence quality forward strand: mean difference = 0.400, 95% confidence interval of difference [95% CI diff] = 0.316 – 0.485, **P<0.001**; genomic sequence quality reverse strand: mean difference = 0.358, 95% CI diff = 0.323 – 0.392, **P<0.001**; average genome sequence quality: mean difference=0.379, 95% CI diff = 0.320 - 0.438, **P<0.001;** target sequence quality forward strand**:** mean difference = 0.339**,** 95% CI diff = 0.186 – 0.491, **P=0.001**; target sequence quality reverse strand: mean difference = 0.089, 95% CI diff = 0.027 –

0.150, **P=0.011**; average target sequence quality: mean difference = 0.214, 95% CI diff = 0.107 - 0.320, **P=0.002**).

The loss of *target sequence quality* was due to the inability of the Sanger method to capture the bases in the 5' region of the target sequence when the Binladen strategy was used. The sequencing products of the Squirrel PCR assay, on the other hand, were completely free of these artefacts. The Squirrel PCR assay also showed superior genomic sequence quality than the Binladen strategy, even though its overall genome sequence quality was less than 100% with the L40 primers (Figure 3.3).

Table 3.3: Squirrel PCR assay versus Binladen et al sequencing strategy - genomic and target sequence sizes & proportions

	Squirrel PCR assay		Binladen et al strategy	
Samples	Genomic sequence size (and quality)	Target sequence size (and quality)	Genomic sequence size (and quality)	Target sequence size (and quality)
DLD1	61/58 (98.4%/93.5%)	17/17 (100%/100%)	32/38 (51.6%/61.3%)	10/15 (58.8%/88.2%)
LS1034	54/61 (87.1%/98.4%)	17/17 (100%/100%)	32/38 (51.6%/61.3%)	10/15 (58.8%/88.2%)
RKO	62/61 (100%/98.4%)	17/17 (100%/100%)	39/41 (62.9%/66.1%)	15/17 (88.2%/100%)
SW620	60/61 (96.8%/98.4%)	17/17 (100%/100%)	32/39 (51.6%/62.9%)	10/15 (58.8%/88.2%)

Values for the forward and reverse sequences are separated by the "/" sign

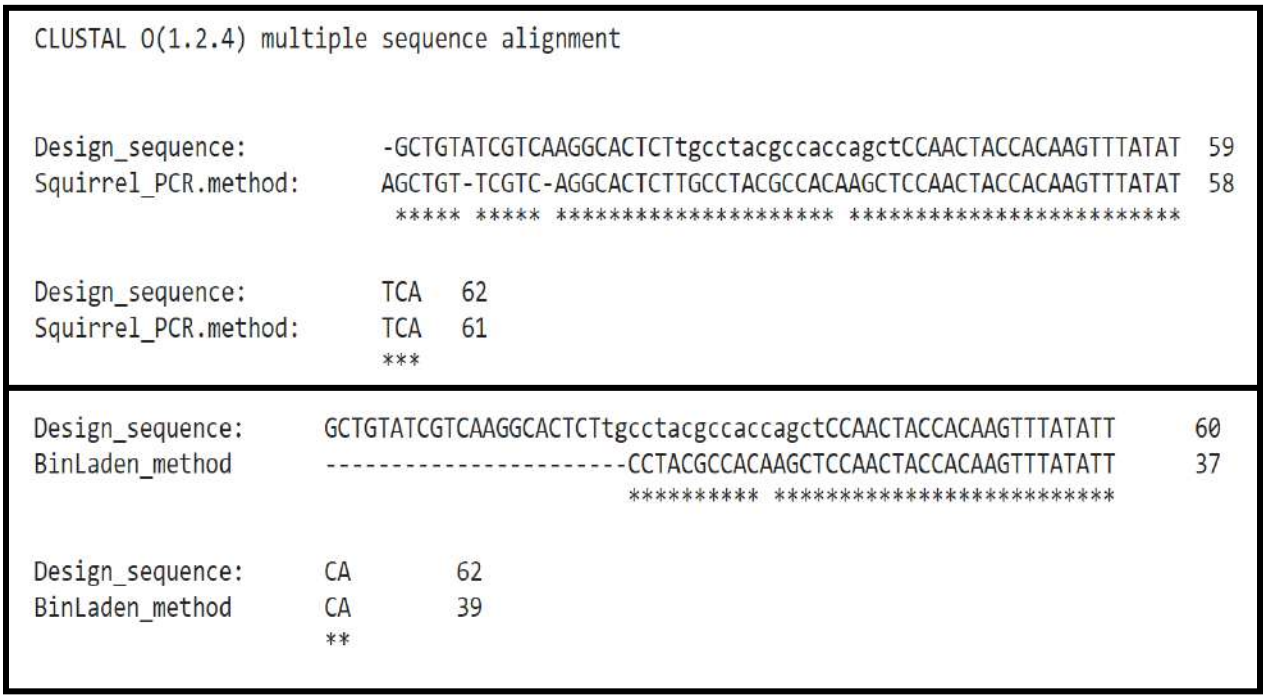


Figure 3.3: Sanger sequencing of the 62bp *KRAS* exon 2 sequence in DLD1 cell line using the Squirrel PCR assay (with the KRAS2Seq primers) and the BinLaden method. The Squirrel PCR assay captured 100% of the target sequence (lower-case in the design sequence) with the accurate identification of the c.35G>T mutation whilst the BinLaden strategy showed partial loss of target sequence. Also, the product of the BinLaden method showed loss of the entire forward primer sequence, whilst the Squirrel PCR assay product only had missed bases in the forward primer sequence. The A nucleotide at the 5' end of the Squirrel PCR product is part of the L40 tail sequence of the KRAS2Seq primers. Sequence alignment accomplished with ClustalW multiple sequence alignment software at https://www.ebi.ac.uk/Tools/msa/clustalo/.

3.3.2 Target Sequence Quality

The target sequence quality obtained by the Squirrel PCR assay was 100% in all the samples tested, irrespective of the template type used or target sequence length. That is, Sanger sequencing accurately captured all the target sequences in all the samples in both the forward and reverse directions (Table 3.4). Furthermore, in all the target sequences tested the mutation hotspots in both the forward and reverse directions were free of sequencing artefacts.

3.3.3 Genomic Sequence Quality

The overall average genomic sequence quality obtained with the Squirrel PCR assay for the gene targets in the entire sample cohort was 98.3% (range=93.6% -100%, forward strand=98.1%, reverse strand=98.5%). The overall average genomic sequence quality for cell lines was 98.6% (range=93.6%-100%, forward strand =98.0%, reverse strand=98.6%) while that for the pre-amplified FFPE samples was 98.3% (range = 96.1-100%, forward strand =98.1%, reverse strand=98.4%) (Table 3.5). In all the samples with less than 100% quality, the genomic sequences were interrupted by missed/ added bases in the 5' regions. The independent-sample T-test showed that the genome sequence quality was independent of the template type, i.e., cell line genomic DNA versus pre-amplified FFPE DNA (genomic sequence quality forward strand: 95% confidence interval of difference [95% CI diff] = -0.053 to 0.017, P=0.291; genomic sequence quality reverse strand: 95% CI diff = -0.040 to 0.033, P=0.825; average genome sequence quality: 95% CI diff = -0.025 to 0.008, P=0.282). However, the genomic sequence quality in the reverse strand, but not in the forward strand or the average quality, showed a positive correlation with the genome and target sequence length independent of Squirrel primer tail length (genomic sequence quality reverse strand [vs genomic sequence length: R=0.430, **P=0.020**; vs target sequence length: R=0.530, **P=0.003**]; genomic sequence quality forward strand [vs genomic

sequence length: R= -0.003, P=0.989; vs target sequence length: R=-0.008, P=0.968]; average genome sequence quality [vs genomic sequence length: R=0.298, P=0.116; vs target sequence length: R=0.360, P=0.055]).

3.3.4 Squirrel Primer Tail Length versus Genomic and Target Sequence Qualities

Next, I sought to determine how increasing the length of the Squirrel primer tails would affect the genomic and target sequence qualities of the Squirrel PCR assay. Whereas the L40 and L50 primers (*KRAS* exon 2, *KRAS* exon 3, *BRAF* exon 15 and *PIK3CA* exon 20) produced amplicons with genomic sequences showing 5' missed/added bases and bad electropherogram peaks in at least one direction and therefore a shortening of the artefact-free size of the genomic sequences, the amplicons of the L58 and L65 primers (PTEN3Seq and KRAS2LSeq) were completely free of these artefacts and had full design sizes (Tables 3.4 and 3.5, Figures 3.4 and 3.5). Pearson correlation analysis of the 30 amplicons sequenced by the Squirrel PCR assay showed a direct correlation between primer tail length and genomic sequence quality independent of genome/target sequence lengths and template type (genomic sequence quality forward strand: R=0.717, P<0.001; genomic sequence quality reverse strand: R=0.706, P=0.003; average genome sequence quality: R=0.751, P<0.001), confirming that the longer the Squirrel primer tail, the better the sequencing data quality.

Table 3.4: Sequencing Quality of the Target Sequences

Samples	Target sequence sizes & proportions				
	KRAS exon 2 (17bp)	*KRAS* exon 3 (7bp)	*BRAF* exon 15 (30bp)	*PIK3CA* exon 20 (28bp)	*PTEN* exon 3 (45bp)
DLD1	100%/100%	-	-	-	-
LS1034	100%/100%	-	-	-	-
RKO	100%/100%	-	-	-	-
SW620	100%/100%	-	-	-	-
S1	100%/100%	-	-	-	-
S3		-	-	100%/100%	-
S5	100%/100%	100%/100%	-	100%/100%	-
S6		-	-	-	100%/100%
S7		-	-	-	100%/100%
S8	100%/100%	100%/100%	-	100%/100%	-
S9		-	100%/100%	100%/100%	100%/100%
S11		100%/100%	100%/100%	-	-
S12		100%/100%	100%/100%	-	-
S13	100%/100%	-	-	-	-
S16		-	-	-	100%/100%
S17		100%/100%	100%/100%	-	-
Rx10	100%/100%	-	-	-	-

Values for the forward and reverse sequences are separated by the "/ " sign

Table 3.5: Sequencing Quality of the Genomic Sequences

Samples		Genomic sequence sizes & proportions				
	KRAS exon 2 (62bp) (KRAS2Seq)	*KRAS* exon 3 (51bp) (KRAS3Seq)	*PIK3CA* exon 20 (71bp) (PIK20Seq)	*BRAF* exon 15 (70bp) (BRAF15Seq)	*PTEN* exon 3 (94bp) (PTEN3Seq)	*KRAS* exon 2 (62 bp) (KRAS2LSeq)
DLD1	61/58 (98.4%/93.5%)	-	-	-	-	62/62 (100%/100%)
LS1034	54/61 (87.1%/98.4%)	-	-	-	-	62/62 (100%/100%)
RKO	62/61 (100%/98.4%)	-	-	-	-	62/62 (100%/100%)
SW620	60/61 (96.8%/98.4%)	-	-	-	-	62/62 (100%/100%)
S1	60/60 (96.8%/96.8%)	-	-	-	-	-
S3	-	-	67/71(94.4%/100%)	-	-	-
S5	61/59 (98.4%/95.2%)	51/49(100%/96.1%)	70/71(98.6%/100%)	-	-	-
S6	-	-		-	93/93(100%/100%)	-
S7	-	-		-	93/93(100%/100%)	-
S8	60.5/59.5 (97.6%/96.0%)	50/49 (98.0%/96.1%)	67/71(94.4%/100%)	-	-	-
S9	-		70/71(98.6%/100%)	69/70(98.6%/100%)	93/93(100%/100%)	-
S11	-	50/50 (98.0%/98.0%)	-	69/70(98.6%/100%)	-	-
S12	-	49/49 (96.1.0%/96.1%)	-	69/70(98.6%/100%)	-	-
S13	61/60 (98.4%/96.8%)	-	-	-	-	-
S16	-	-	-	-	93/93(100%/100%)	-
S17	-	50/49 (98.0%/96.1%)	-	69/70(98.6%/100%)	-	-
Rx10	59.5/60.5 (96.0%/97.6%)	-	-	-	-	-

Values for the forward and reverse sequences are separated by the "/" sign



Figure 3.4: Sanger sequencing by the Squirrel PCR strategy: comparison of the sequences of the KRAS2Seq and KRAS2LSeq primers products of RKO cell line DNA template. Whilst the KRAS2Seq amplicon showed partial loss of forward primer sequences, the KRAS2LSeq product showed a full design sequence. The extra 5' and 3' bases in the Squirrel PCR products represent parts of the primer tail sequences.

3.3.5 Analytical Performance of the Squirrel PCR assay

Cohen's Kappa test showed a perfect concordance between the known mutation status of the samples and the Squirrel primer-enabled Sanger sequencing results. The results of the Squirrel primer-based sequencing were concordant with the known *KRAS* exon 2 mutations status of CRC cell lines and confirmed the HRMA patterns observed with all the mutation targets for all the clinical (i.e., pre-amplified FFPE) samples (Table 3.6-3.8 and Figures 3.6-3.8). The performance indices of the Squirrel PCR-Sanger sequencing assay showed a sensitivity of 100% (95% C.I. = 73.54% to 100.00%) at mutation detection, and a specificity of 100% (95 % C.I. = 80.49% to 100.00%). Positive and negative predictive values for mutation detection by the Squirrel PCR-Sanger sequencing assay were both 100% while the accuracy of the assay was 100% (95% C.I. = 88.06% to 100.00%).

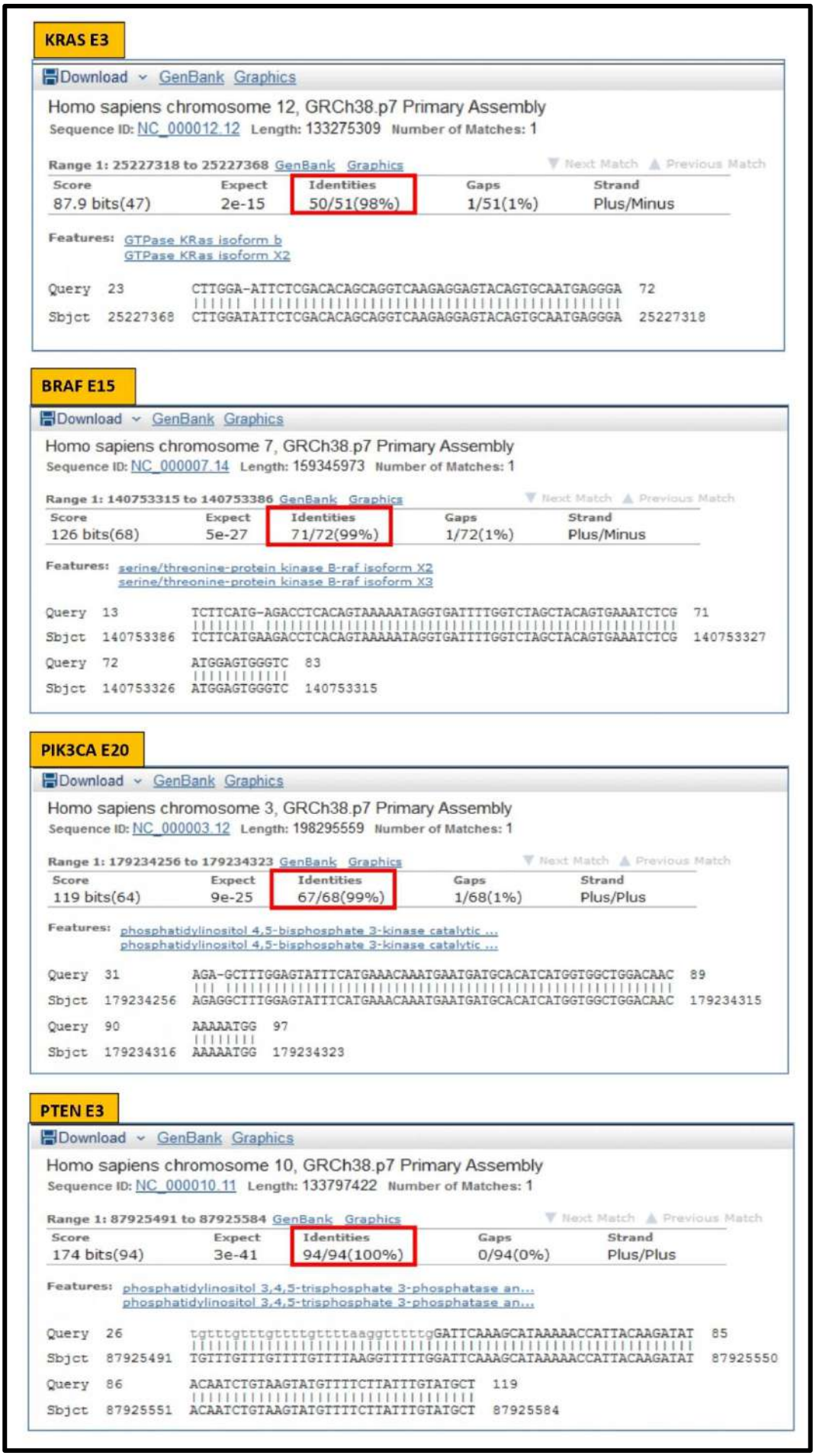


Figure 3.5: Alignment of sequencing data with the GenBank Homo sapiens GRCh38.p7 Primary Assembly. The alignment scores are indicated in the red box. The Gap scores are >0% for the sequences of the L40 and L50 primers (KRAS3Seq, BRAF15Seq and PIK20Seq), with evidence of 5'missed bases. The L58 primer (PTEN3Seq), on the other hand, has a Gap score of 0% with no 5' missed/added bases.

Table 3.6: *KRAS* exon 2 mutation status in cell lines

Cell lines	***KRAS* exon 2 mutation status**
DLD1	c.38G>A
HCT116	c.38G>A
RKO	wt
SW620	c.35G>T
HUTU80	wt

Table 3.7: HRMA patterns of CRC samples at mutation hotspots

Samples	*KRAS* exon 2	*KRAS* exon 3	*BRAF* exon 15	*PIK3CA* exon 20	*PTEN* exon 3
S1	Normal	-	-	-	-
S3	-	-	-	Aberrant	-
S5	Aberrant	Normal	-	Normal	-
S6	-	-	-	-	Normal
S7	-	-	-	-	Normal
S8	Aberrant	Normal	-	Normal	-
S9	-	--	Normal	Aberrant	Aberrant
S11	-	Normal	Normal	-	-
S12	-	Normal	Normal	-	-
S13	Aberrant	-	-	-	-
S16	-	-	-	-	Normal
S17	-	Normal	Aberrant	-	-
Rx10	Aberrant	-	-	-	-

Table 3.8: Confirmation of HRMA results by Squirrel PCR-Sanger sequencing

Samples	KRAS exon 2		KRAS exon 3		BRAF exon 15		PIK3CA exon 20		PTEN exon 3	
	HRMA	Seq	HRMA	Seq	HRMA	Seq	HRMA	Seq	HRMA	Seq
S1	wt	wt	-	-	-	-	-	-	-	-
S3	-	-	-	-	-	-	Aberrant	c.3140A>G	-	-
S5	Aberrant	c.34G>T	wt	wt	-	-	wt	wt	-	-
S6	-	-	-	-	-	-	-	-	wt	wt
S7	-	-	-	-	-	-	-	-	wt	wt
S8	Aberrant	c.38G>A	wt	wt	-	-	wt	wt	-	-
S9	-	-	-	-	wt	wt	Aberrant	c.3140A>G	Aberrant	c.180G>T
S11	-	-	wt	wt	wt	wt	-	-	-	-
S12	-	-	wt	wt	wt	wt	-	-	-	-
S13	Aberrant	c.35G>A	-	-	-		-	-	-	-
S16	-	-	-	-	-		-	-	wt	wt
S17	-	-	wt	wt	Aberrant	c.1799T>A	-	-	-	-
Rx10	Aberrant	c.35G>T	-	-			-	-	-	-

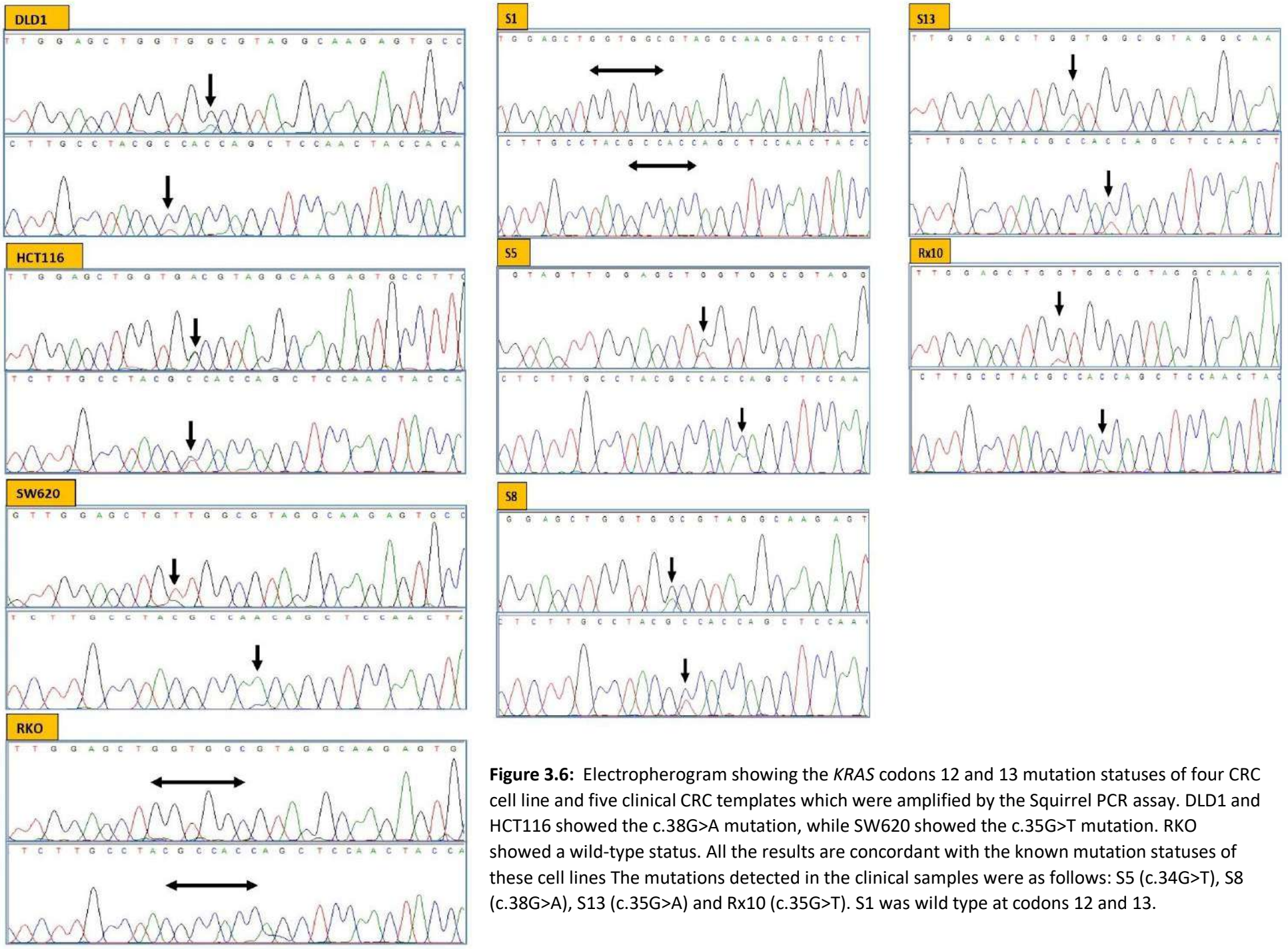


Figure 3.6: Electropherogram showing the *KRAS* codons 12 and 13 mutation statuses of four CRC cell line and five clinical CRC templates which were amplified by the Squirrel PCR assay. DLD1 and HCT116 showed the c.38G>A mutation, while SW620 showed the c.35G>T mutation. RKO showed a wild-type status. All the results are concordant with the known mutation statuses of these cell lines The mutations detected in the clinical samples were as follows: S5 (c.34G>T), S8 (c.38G>A), S13 (c.35G>A) and Rx10 (c.35G>T). S1 was wild type at codons 12 and 13.

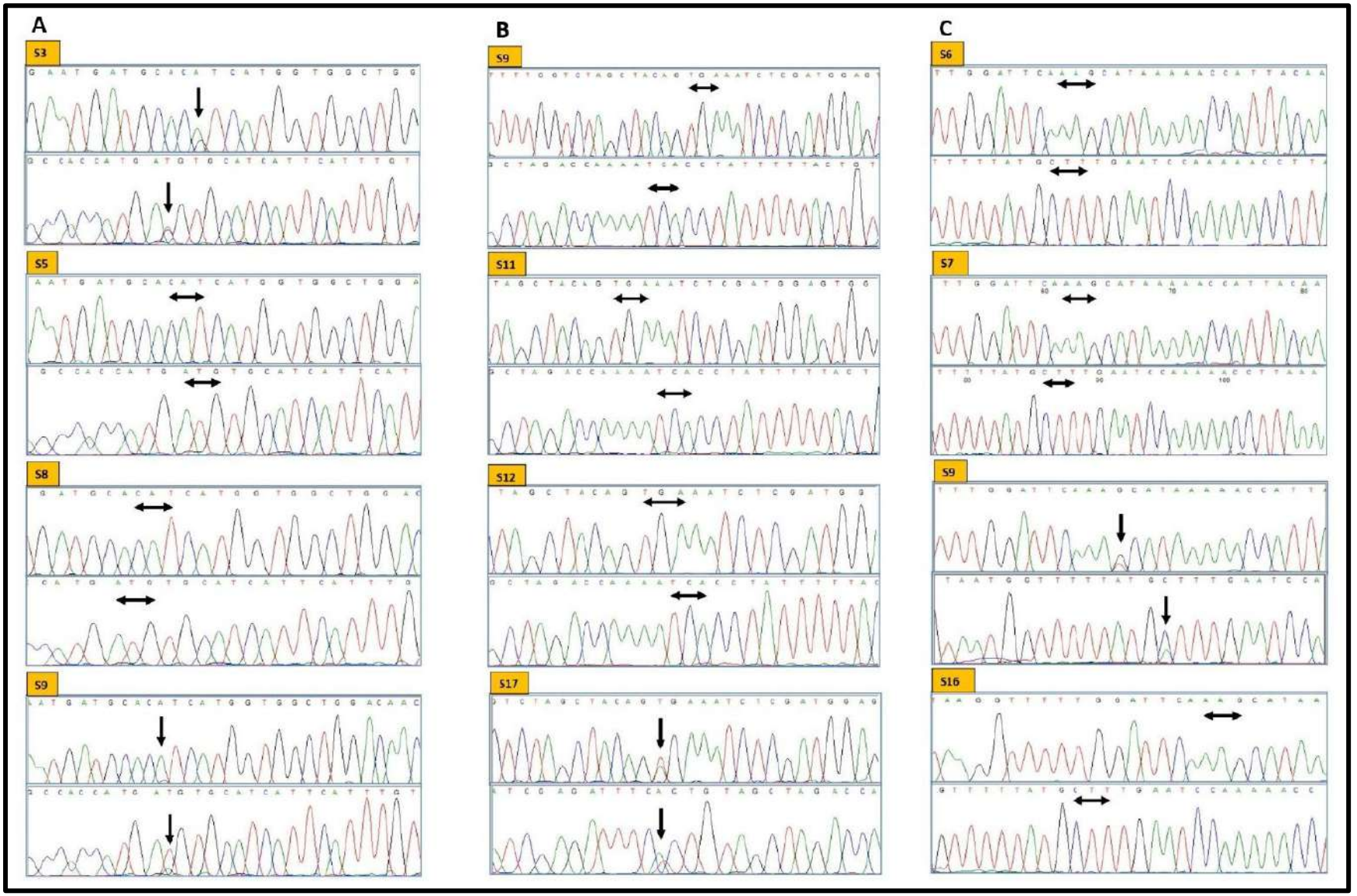


Figure 3.7: Mutation detection by Squirrel PCR-based Sanger method A. Electropherogram showing *PIK3CA* codon 1047 mutation status of five clinical CRC sample templates which were amplified by the Squirrel primer-based assay. Samples S3 and S9 showed the c.3140A>G mutation, while the other samples were wild type at this locus. B. Electropherogram showing *BRAF* codons 600 mutation statuses of four clinical CRC sample templates which were amplified by the novel PCR assay. Sample S17 showed the c.1799T>A (V600E) mutation while the other samples were wild type at this locus C. Electropherogram showing *PTEN* codons 60 and 61mutation status of five clinical CRC sample templates which were amplified by the novel PCR assay. The c.180G>T (codon 60) mutation was found in sample S9; the other samples were wild type at this locus and no mutations were found in codon 61 for any sample.

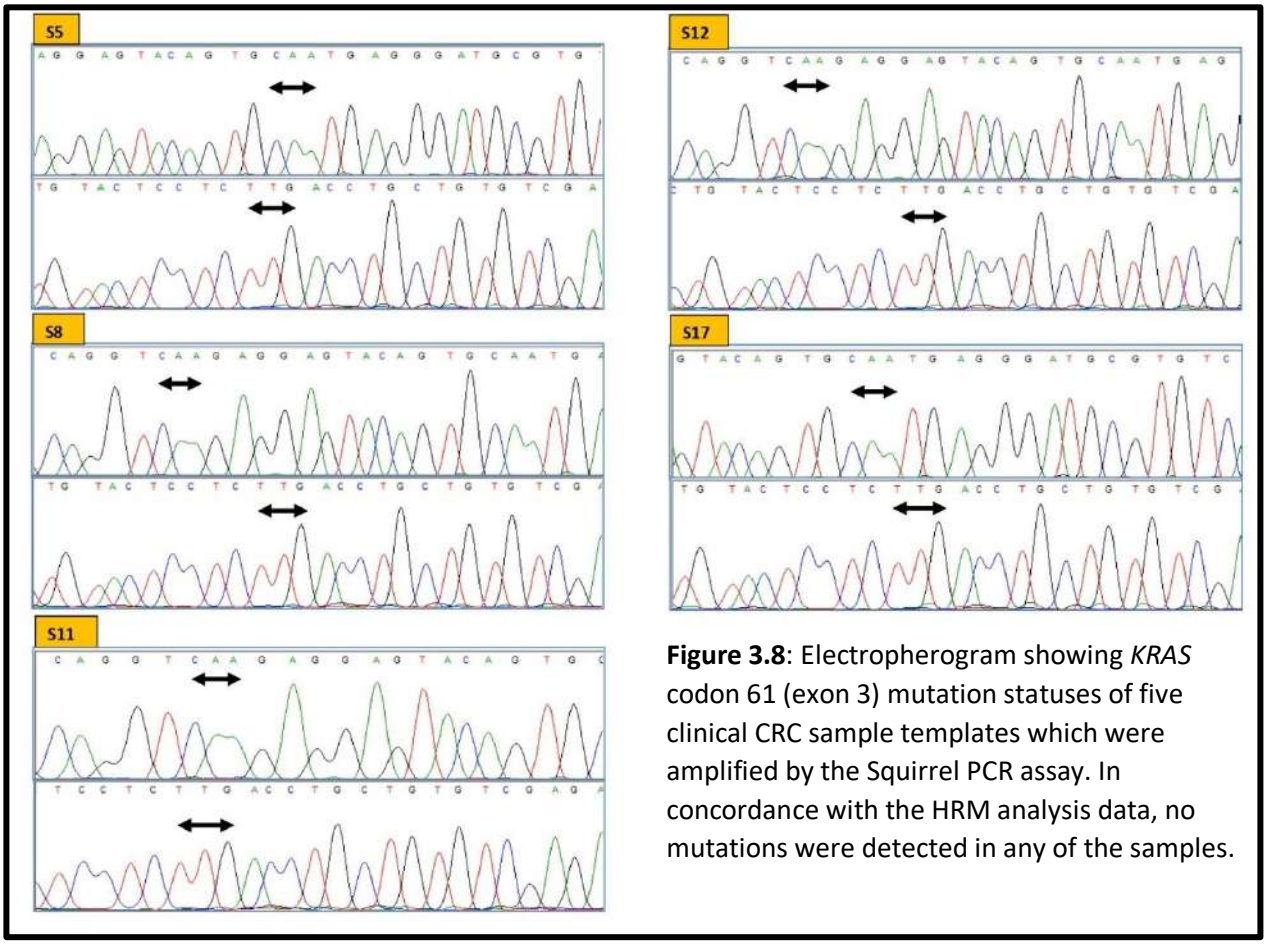


Figure 3.8: Electropherogram showing *KRAS* codon 61 (exon 3) mutation statuses of five clinical CRC sample templates which were amplified by the Squirrel PCR assay. In concordance with the HRM analysis data, no mutations were detected in any of the samples.

3.4 Discussion

The Squirrel PCR assay permitted direct and targeted sequences of ultra-short genomic segments, with complete capture of target sequences and accurate mutation detection both in the forward and reverse directions. The Squirrel PCR-Sanger sequencing assay also achieved high genomic quality irrespective of the template type or genome sequence length. However, the genome sequence quality was significantly improved by primer tail length of 58 bases and above, evidence that the longer the primer tail the better the sequence quality. This is in keeping with the observation that Sanger sequencing is error-prone in the initial 15-40 bases of the electropherogram. When the L40 or L50 primers were used to amplify the genomic region of interest and the 3' end of the 20-base long USeq primers bind in or near the middle of the long tails, the sequencing errors would likely extend into the genomic sequence of the amplicons. However, with the L58 or L65 primers the USeq primers bind farther away from the genomic sequences and reduced the likelihood of sequencing errors extending into the genomic sequences. Therefore, the longer the Squirrel primer tail the better the sequencing quality. To robustly obtain perfect genome sequence quality from ultra-short DNA molecules I recommend Squirrel and USeq primers design that would give a 40- to 45-base allowance between USeq primer 3' binding site and the 5' end of the specific primer sequence.

Tailed primers are routinely used for Sanger sequencing. For example, the M13 complementary sequences are commonly incorporated into plasmid-specific sequences in which they serve as matrices for Sanger sequencing using M13 primers. Using M13-tailed primers in cloning-to-sequencing simplifies large sequencing projects (Li et al., 2011). Furthermore, Williams-Woods et al (Williams-Woods et al., 2011) used M13 tailed primers to achieve improvements in the sequencing of 84-213 bp DNA segments. It achieved 95% alignment score with the 213bp

product. However, data for the 84-89bp segments were not presented in the published article. Moreover, Brisotto et al (Brisotto et al., 2015) used 18-19mer M13 tailed primers to perform normalised bisulfite sequencing PCR and achieved some improvement in the quantitation of C-T ratios in 169bp and 246bp *CHD1* and miR-200/miR-141 CG-rich promoter sequences, respectively. In addition, Binladen et al (Binladen et al., 2007) used 40bases- and 60bases-tailed (L40 and L60) primers as sequencing primers for sub-150bp products amplified by conventional primers and found improvements in the quality of sequencing data obtained. However, I could not replicate the Binladen et al sequencing strategy as the electropherogram I obtained using their strategy showed poor sequencing data in the genomic and target sequences. The Squirrel PCR method differs from the one described by Binladen et al since the tail is incorporated into the amplicons post-PCR, rather than into the sequencing primers. As electropherogram reads begin at the 3' primer binding site of the sequencing primers, the use of a shorter sequencing primer pair (USeq primers) that binds farther upstream of the non-specific tail sequences would produce better quality genome-specific sequence data in comparison to using sequencing primers which incorporate the PCR primer sequences. Our research group successfully sequenced the region C of the *MLH1* promoter using the Squirrel PCR strategy (Susanti et al., 2018). Compared to NGS or sequencing-by-cloning, the Squirrel PCR sequencing assay is much faster, easier to perform and much more cost-effective for sequencing single or few gene targets (Crossley et al., 2020; Li et al., 2011) (see Appendices 1 and 2, and the accompanying Supplementary Tables 1 and 2 for a comparison of the cost estimates and turnaround times for deploying the newly developed assays relative to established assays). Although pyrosequencing may be best suited for sequencing sub-100bp targets in terms of techniques, cost of operation and speed (Fakkrudin et al, 2013), for a routine molecular diagnostic laboratory that already

possesses the more versatile capillary electrophoresis machine – that performs bisulfite sequencing, DNA fragment analysis, STR analysis and DNA sequencing, etc., or for one that seeks to add a capillary electrophoresis machine to its armamentarium due to the machine's versatility – the addition of a pyrosequencing platform just for sequencing short DNA segments may be unconscionable from the economics standpoint.

The Squirrel PCR assay can be applied to the sequencing of PCR products from any DNA template type, including FFPE and cell line DNA. As this study has shown, sequencing quality is not dependent on the template type. A particularly important application for the Squirrel PCR assay would be in the sequencing of PCR amplicons derived from ctDNA template. Application of the Squirrel PCR assay to the sequencing of ctDNA-derived product is supremely important because of the role ctDNA has been projected to play in the near-future management of cancer patients (Cheng et al., 2021; Lissa & Robles, 2019; Osumi et al., 2020). ctDNA mutation profiling, in principle, gives a portrait of the molecular changes occurring real-time as the cancer progresses in the patient. Additionally, the process of obtaining ctDNA from the patient, i.e., liquid biopsy, is comparatively much less invasive than tissue biopsy (Cheng et al., 2021; Lissa & Robles, 2019; Osumi et al., 2020). These advantages make it feasible and necessary to repeatedly sample ctDNA for prediction of response to therapy, prediction of prognosis and monitoring response to therapy (Cheng et al., 2021; Lissa & Robles, 2019; Osumi et al., 2020). Currently, the NCCN recommends ctDNA profiling for disease-monitoring or therapy response-prediction purposes in lung, invasive breast, and gastro-oesophageal cancers when a patient with confirmed diagnosis is medically unfit to undergo invasive tissue sampling (Cheng et al., 2021). With improvements in ctDNA isolation technology, and so long as the limited-gene-target-profiling approach remains the mainstay of per-tumour patient management, it is expected that

the utilisation of ctDNA mutation profiling will rise and become a commonplace test for cancer management (Cheng et al., 2021; Lissa & Robles, 2019; Osumi et al., 2020). The Squirrel PCR sequencing assay is most suited to the ctDNA templates as they are short with mean fragment length of approximately 150bp (Alcaide et al., 2020; X. Liu et al., 2020; Mouliere & Rosenfeld, 2015; Underhill et al., 2016).

Furthermore, archival tissue (i.e., FFPE) DNA templates, which are usually degraded and fragmented, have been invaluable in molecular pathology research and diagnostics (Dietrich et al., 2013; Gilbert et al., 2007; Ludyga et al., 2012; Srinivasan et al., 2002). The Squirrel PCR sequencing assay would therefore find useful applications in research and diagnostics which utilise ctDNA and FFPE tumour DNA as starting materials for molecular profiling of cancer. This applicability would, not only be in resource-poor or budget-restricted settings, but also in the setting where limited gene targets are the goals for sequencing.

However, the limit of detection of this assay was not determined due to time constraints of the PhD work. Furthermore, many clinically relevant variants, from among *KRAS* codons 12/13/59/61/117/146, *NRAS* codon 12/13/59/61, *BRAF* codon 600, etc. were not included in the development of the assay. In addition, overall, the actual number of events tested by the Squirrel PCR assay was too small for a thorough assessment of the assay performance to be made.

In conclusion, the data obtained from the design and development Squirrel PCR assay prove the concept that the Sanger method can be manipulated to sequence sub-100 bp DNA segments, if such segments are elongated using the Squirrel PCR strategy. The data also showed that the Squirrel PCR-Sanger sequencing combination can be robust against DNA template types and can have high performance indices. Further validation is warranted to test the performance of the assay at detecting clinically relevant targets and to confirm its high-performance indices.

CHAPTER 4

mHRMA: MULTIPLEX HRMA FOR RAPID MUTATION DETECTION

4.1 Background

Somatic point mutations and small insertion-deletions (indels) in proto-oncogenes and tumour suppressor genes together constitute one of the commonest causes of molecular alterations in cancers (Greenman et al., 2007; Mészáros et al., 2016; Vogelstein et al., 2013). The detection of these mutations is currently an indispensable component of personalised medicine and the targeted cancer therapy modality (Andre et al., 2014; Senft et al., 2017; X. Wang, 2016). As the use of the targeted therapy approach has generally achieved significant improvement in cancer outcomes (Awad et al., 2019; Camidge, 2014; Dong et al., 2019; Heinemann et al., 2014; Kabbinavar et al., 2005; Kantarjian et al., 2012; Van Cutsem et al., 2011), tumour mutation profiling has become an important aspect of the general work-up of patients for therapy (Brown & Elenitoba-Johnson, 2020; El-Deiry et al., 2019; Malone et al., 2020; Presley et al., 2018). Therapy-purpose somatic mutation detection serves three objectives: 1) identification of patients who are likely to respond to the targeted therapy due to the mutation harboured by their tumour, 2) identification of patients who are potentially non-responders to the targeted therapy due to the mutations present in their tumours, and 3) monitoring of acquired resistance to targeted therapy (Brown & Elenitoba-Johnson, 2020; El-Deiry et al., 2019; Malone et al., 2020; Presley et al., 2018; Van Emburgh et al., 2016). Although genetic mutations in any individual patient's tumour may be multitudinous, the number of per-tumour type mutations of current clinical or targeted-therapy relevance is limited to a handful of mutations in established proto-oncogenes and tumour suppressor genes (Martinelli et al., 2020; National Comprehensive Cancer Network, 2021e). For example, according to the 2021-2022 UK National Genomic Test Directory for Cancer, only ten gene targets are currently clinically actionable for sporadic colorectal, gastric, and oesophageal

cancers combined. The investigation of other targets such as mutations in *MSH2, MLH1, PMS2, MSH6, POLE* and *POLD1* are undertaken for Lynch syndrome screening (National Health Service, 2022). Furthermore, there are about eleven drugs approved by the FDA for targeted therapy in colorectal, gastric, and oesophageal cancers (El-Deiry et al., 2019; National Cancer Institute, 2021) and only seven by the NICE for the same cancers (National Institute for Health and Care Excellence, 2010; 2012; 2021a; 2021e; Cancer Research UK, 2022a; 2022b). The companion diagnostic assays for four of these drugs test for point mutations in *EGFR* (Cetuximab and Panitumumab), *KRAS* (Cetuximab and Panitumumab), *NRAS* (Cetuximab and Panitumumab), and *BRAF* (Encorafenib) and indels (i.e., instability) in established microsatellite regions (Pembrolizumab) (El-Deiry et al., 2019; National Cancer Institute, 2021, National Institute for Health and Care Excellence, 2012; 2021a; 2021e). Moreover, mutations in *PIK3CA* and *PTEN* have been shown to have predictive values for anti-EGFR therapy in CRC (Therkildsen et al., 2014). Furthermore, mutations in *PIK3CA* are relevant to the therapeutic responses in preclinical models of colorectal and gastric cancers (K.-J. Kim et al., 2020; Willis et al., 2020).

As the number of molecular targets currently tested routinely for therapy purpose are small, a cost-effective and practical approach for mutation detection in the clinical setting - especially for standard molecular diagnostics laboratories without large investments in research and development - would be to use a PCR-based mutation detection and Sanger sequencing methods (Awad et al., 2019; Brown & Elenitoba-Johnson, 2020; Mancini et al., 2015; Presley et al., 2018). Many PCR-based assays are currently available as laboratory-developed tests. Most of these assays utilise the single-stranded conformation polymorphism analysis (SSCP), the denaturing high performance liquid chromatography (dHPLC), Amplification Refractory

Mutation Specific (ARMS) PCR and High Resolution Melting Analysis (HRMA) strategies for the detection of mutations in the clinical laboratories (Mancini et al., 2015).

HRMA is a rapid, exquisitely sensitive, and accurate, easy-to-perform, and cost-effective method of mutation detection. It has found application in single nucleotide mutations or polymorphisms (SNPs) mutation screening and genotyping, and small indel detection, microsatellite instability analysis, methylation analysis, and gene copy number analysis (Dobrowolski et al., 2012; Farrar & Wittwer, 2017; Guo et al., 2012; Janavicius et al., 2010; Reed et al., 2007; Wojdacz & Dobrovic, 2007). Although, HRMA by itself cannot detect the specific base mutations, but by inclusion of the appropriate spike-in wild-type controls in the assay, HRMA can be used to define the genotype of a mutation or a SNP, a technique called quantitative heteroduplex analysis (Garritano et al., 2009; Palais et al., 2005). In addition, HRMA is a quantitative method and can be used to assess variant allele frequency, VAF (or, mutant allele frequency MAF) if standard quantitative controls are included in the experiments (Wojdacz et al, 2007; Ham-Karim et al, 2017; Fadhil et al, 2010; Ibrahem et al, 2010; Ishige et al, 2018). The VAF of test samples is determined by matching the heights of their Difference Curves to those of the standard quantitative controls. VAF, in turn, can be used to assess the Heterogeneity score (HS), an index used for determining clonality of a mutation. If the tumour cellularity (or tumour percentage) of any DNA sample is known, the heterogeneity score of a mutation can be determined by simply applying the following equation: HS = (VAF X 2)/ tumour cellularity. Mutations with HS <1 are regarded as clonal, those with HS =1 are total, while mutations with HS > 1 are regarded as occurring in the setting of copy number changes, either by uniparental disomy or by deletion of wild type allele (Li et al, 2018; Normanno et al, 2015; Li et al, 2017). If the tumour cellularity is known, HRMA can be used to determine VAF, HS and clonality. Furthermore, HRMA can be a

confirmatory test for the presence/absence of a mutation when equivocal results are obtained by other methods (Mancini et al., 2015).

HRMA detects DNA sequence variation of PCR amplicons on the bases of (i) differences between the melting properties of the standard (known, usually normal or wild type DNA) and the target (or test) sequences; this sequence variation is usually as small as one-base difference between the known and unknown sequences. The higher the number of base differences the more divergent the melting properties of the standard and target sequences and the easier the sequence variation detection by HRMA (ii) the presence of mutated and wild type sequences in starting DNA template – such as from a clinical tumour sample - cause the formation of heteroduplexes, which melts faster than either wild-type or mutant homoduplex sequences, and enhances the divergence of the melting properties between wild type sample and mutated sample (Farrar & Wittwer, 2017; Reed et al., 2007; Tucker & Huynh, 2014). To detect sequence variation by HRMA, PCR is first performed with a saturating fluorescent dye such as Eva Green, LC Green, and SYTO9 (Farrar & Wittwer, 2017; Reed et al., 2007; Tucker & Huynh, 2014). The amplicons are then subjected to melting in a PCR machine with HRMA capability or a purpose-built HRMA machine – instruments capable of melting DNA at a rate of 0.1-1^{0}C/s and acquiring up to 65 fluorescence data points per ^{0}C changes in temperature (Farrar & Wittwer, 2017; Reed et al., 2007; Tucker & Huynh, 2014). Then the difference in melting patterns between samples is detected in the (i) normalised melting curve or shifted melting curve, which is obtained by plotting the temperature-normalised fluorescence (y-axis) against the temperature (x-axis), (ii) the difference curve, which is obtained by plotting the *difference in change of fluorescence* between the standard and test sequences over the melting temperature range (Reed et al., 2007; Tucker & Huynh, 2014). Usually only one target – represented by a single melting peak or a

melting curve with a single flexure or bend - is screened per tube (Figure 4.1). Thus, HRMA is conventionally performed as a singleplex target assay.

I reasoned that if the number of target amplicons in a single tube could be increased or multiplexed, then the speed of operation of HRMA could be increased whilst the costs of manpower and consumable would be reduced. In this chapter I describe the development of the mHRMA, a multiplexed HRMA-based assay for rapid detection of mutation in cancer, using CRC cell lines and pre-amplified clinical CRC samples.

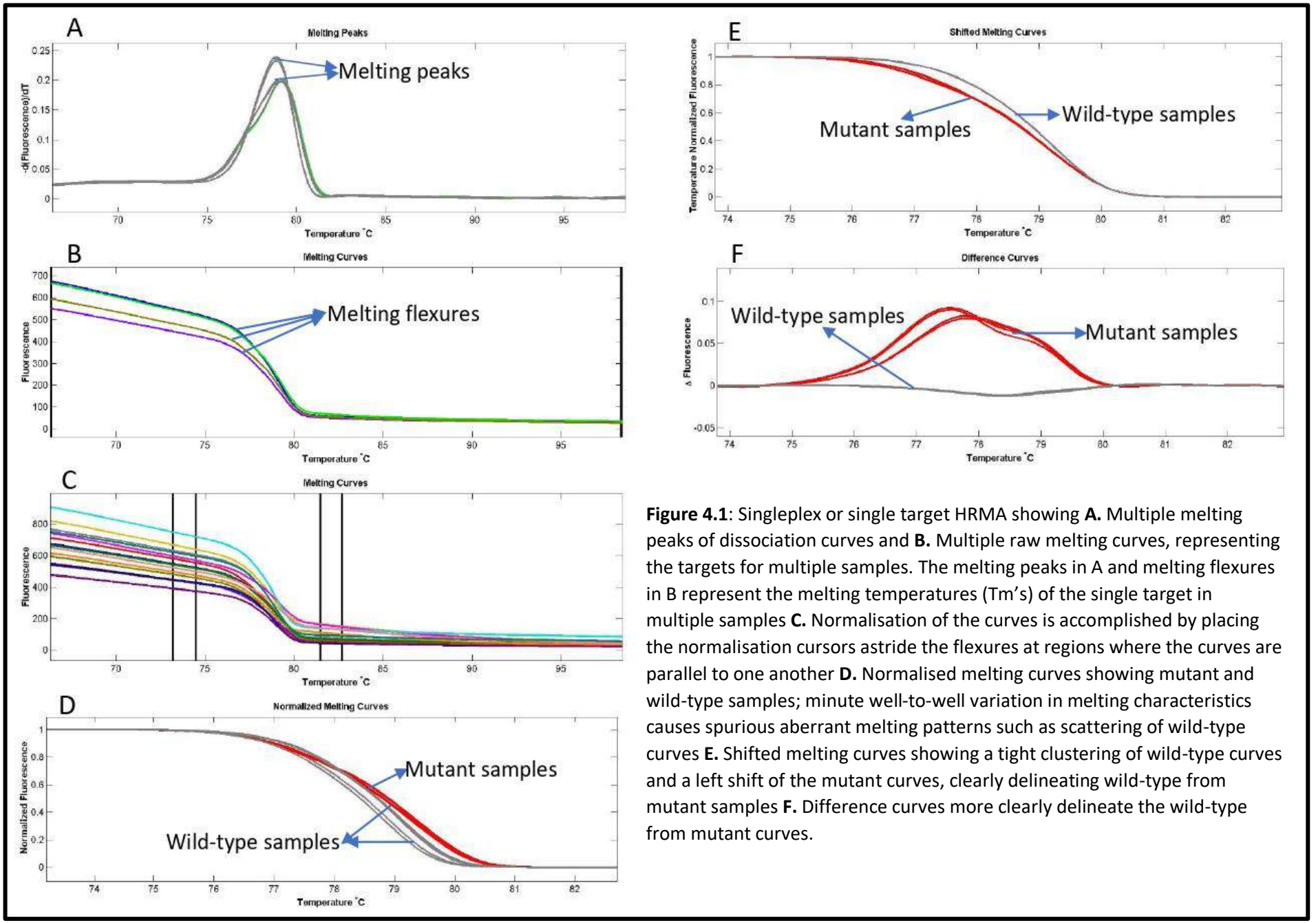


Figure 4.1: Singleplex or single target HRMA showing **A.** Multiple melting peaks of dissociation curves and **B.** Multiple raw melting curves, representing the targets for multiple samples. The melting peaks in A and melting flexures in B represent the melting temperatures (Tm's) of the single target in multiple samples **C.** Normalisation of the curves is accomplished by placing the normalisation cursors astride the flexures at regions where the curves are parallel to one another **D.** Normalised melting curves showing mutant and wild-type samples; minute well-to-well variation in melting characteristics causes spurious aberrant melting patterns such as scattering of wild-type curves **E.** Shifted melting curves showing a tight clustering of wild-type curves and a left shift of the mutant curves, clearly delineating wild-type from mutant samples **F.** Difference curves more clearly delineate the wild-type from mutant curves.

4.2 Material and Methods

4.2.1 Principles of mHRMA

An HRMA assay can be multiplexed if the melting temperature (Tm) peaks (or curve flexures) of the target amplicons are sufficiently spaced apart along the temperature axis to preclude their co-interference during analyses.

To prove/confirm this principle I needed to show that (i) overlapping amplicon Tm peaks produce spurious melting patterns, while sufficiently separated Tm peaks achieve correct melting patterns, (ii) GC-tagged primers, which produce amplicons with raised Tm's can be used to vary the Tm's of the PCR amplicons and to achieve sufficiently spaced Tm's along the temperature axis.

4.2.2 Materials

Fresh DNA, derived from the CRC cell lines Vaco5, SW948, HCT116, SW620 and Vaco10MS, was used for the development and optimisation of the methodology. Seventeen (17) pre-amplified clinical CRC samples with known mutations and melting patterns – as previously determined by the QMC PCR method (Fadhil et al., 2010) - were used to validate the mHRMA assay.

4.2.3 mHRMA and Tm peaks separation

Targets in *TP53* and *SMAD4* were used to ascertain whether sufficient separation of amplicon Tm peaks affect melting patterns. Five CRC cell lines, including Vaco5 (containing *TP53* E282W), SW 984 (containing *TP53* codon 117, G del), HCT116, SW620 and Vaco10MS (wild type at those *TP53* loci) were used to demonstrate the effect of Tm peak separation on accurate melting patterns. All 5 cell lines are wild type at *SMAD4* exon 9a. Three different targets: *TP53*

exon 8b (encompassing the codon 282; predicted Tm of 84.3 ^{0}C, *TP53* exon 4d (encompassing codon 117; predicted Tm of 87.8 ^{0}C) and *SMAD4* exon 9a targets (encompassing codon 330; predicted Tm of 76.9 ^{0}C) were tested in different duplex combinations using the cycling conditions: 1 cycle of 95ºC for 5 minutes followed by 45 cycles of 95ºC for 5 seconds and 60^{0}C for 30 seconds and 72^{0}C for 20 seconds. The cycling was followed by a 5-minute final extension, and a dissociative curve stage comprising 30 seconds each at 95^{0}C, 55^{0}C and 95^{0}C. The PCR primer sequences for this optimisation step are shown in Table 4.1. For the multiplex HRM analysis data normalisation was achieved by placing the cursors astride individual group of peaks or raw melt curves flexures (see Results). Other melting and analyses parameters applied were the same as for the single product analysis (as described above in the Introduction section).

4.2.4 Achieving adequate separation of target Tm peaks using GC-tagged primers

My initial experiments from above showed that (i) the validity of mHRMA results is dependent on sufficiently spaced melting peaks of the individual PCR amplicons (ii) the combination of the natural Tm's of the PCR amplicons does not consistently achieve adequate separation of the Tm peaks. Thus, I sought to test how the modification of the primers by addition of non-specific GC-tags to the 5' ends of the specific primers would enhance the Tm peaks of the PCR amplicons and facilitate adequate separation of the multiplex amplicons along the temperature axis. A combination of Bioinformatics tools, including the UCSC *in silico* PCR (https://genome.ucsc.edu/cgi-bin/hgPcr), MFEprimer (www.biocompute.bmi.ac.cn/CZlab/MFEprimer-2.0/) and UMELT batch 2.0 (https://www.dna.utah.edu/umelt/umb.php) were used to demonstrate *in silico* that adding a GC-tail to some primers could be utilised to alter the Tm and thereby enhance separation of multiplex amplicons in order to ensure reliable mHRMA (Dwight et al., 2011; Qu et al., 2012; Speir et al.,

2016) as follows. *In silico* PCRs were performed on UCSC *In silico* PCR or MFEprimer software using the designed primer sequences; then the product sequences were cut and pasted in the Batch mode of the UMELT, and an *in silico* melting was performed, following which the Tm's of the products were visualised. The separation of the peaks of the individual targets was achieved by adding GC-tags *in silico* to the ends of one or the other sequences (see Results). This was confirmed experimentally with the primer sequences shown in Table 4.1.

4.2.5 Validation of mHRMA

Pre-amplified clinical samples with known melting patterns at *PTEN* exons 3, 5 and 7, *KRAS* exons 2, 3 and 4, *NRAS* exon 2, *BRAF* exon 15, and *PIK3CA* exons 1, 9 and 20 mutation hotspots were used to validate the mHRMA. UMELT software was utilised to predict the targets with sufficiently spaced Tm peaks, and which would therefore be suitable for mHRMA. Five targets were combined as follows for mHRMA: *PTEN* exon 3/*PIK3CA* exon 20/*KRAS* exons 3 were combined in triplex reactions while *PTEN* exon 7/ *NRAS* exon 2 were chosen for duplex reactions. One of the primer pairs, *KRAS* exon 3, was tailed in order to enhance Tm separation. PCR volume was undertaken in a final volume of 15µl. Two µl of the pre-amplified templates were used for the multiplex reactions, while the primer concentration was 250nM for all the target exons except for the *NRAS* exon-2 target for which an optimised primer concentration of 75nM was used. The cycling parameters for the multiplex reactions were as follows: 1 cycle of 95ºC for 5 minutes, followed by 45 cycles of 95ºC for 10 seconds, 60ºC (for the triplex) and 55^0C (for the duplex) for 30 seconds, 72ºC for 20 seconds, and a final extension of 72ºC for 5 minutes.

Table 4.1: List of primers and their product Tm's

Genes/Exons	Primer sequences	Product sizes	Product Tm's
TP53 E4d	F: GGCAGCTACGGTTTCC R: CCCTCAGGGCAACTGA	80bp	84.3^{0}C
TP53 E8b	F: GCCTGTCCTGGGAGAGAC R: CCTTTCTTGCGGAGATTCT	53bp	87.8^{0}C
SMAD4 E9a	F: GTATTGGTGTTCCATTGCTTACT R: TAAATGTCTCTCCTACCTGAACAT	56bp	76.9^{0}C
KRAS E2	F: TGAATATAAACTTGTGGTAGTTG R: GCTGTATCGTCAAGGCACTCT	62bp	84.1^{0}C
KRAS E3	F: CTTGGATATTCTCGACACAGCA R: TCCCTCATTGCACTGTACTCCT TAGGED F: cccgggcgggccggcccCTTGGATATTCTCGACACAGCA R: cccgggcgggccggcccTCCCTCATTGCACTGTACTCCT	51bp 85bp	81.5^{0}C 92.0^{0}C
KRAS E4	F: TGGAATTCCTTTTATTGAAACATC R: TTTCAGTGTTACTTACCTGTCTTGTCT	56bp	75.1^{0}C
BRAF E11	F: TGGGCAGATTACAGTGGGA R: GCCACTTTCCCTTGTAGACTG	68bp	81.5^{0}C
BRAF E15	F: TTCATGAAGACCTCACAGTAAA R: GACCCACTCCATCGAGAT TAGGED F: gggccggcccTTCATGAAGACCTCACAGTAAA R: gggccggcccGACCCACTCCATCGAGAT	70bp 90bp	80.5^{0}C 87.0^{0}C
PTEN E3	F: TTAAGGTTTTTGGATTCAAAG R: CAAATAAGAAAACATACTTACAGATT	71bp	75.0^{0}C
PTEN E7	F: CCAGCTAAAGGTGAAGATA R: CAAAGTACATGAACTTGTCTTC	71bp	82.5^{0}C
PIK3CA E20	F: GCAAGAGGCTTTGGAGTATTTCA R: CCATTTTTGTTGTCCAGCCA TAGGED F: ggccggcccGCAAGAGGCTTTGGAGTATTTCA R: ggccggcccCCATTTTTGTTGTCCAGCCA	71bp 89bp	80.5^{0}C 87.0^{0}C
NRAS E2	F: TACAAACTGGTGGTGGTTGG R: CACTGGGCCTCACCTCTATG	115	87.5^{0}C

4.2.6 Statistical Analyses

The results of the mHRMA and QMC-PCR were compared using crude percentage concordance calculations. Furthermore, the assay agreement between the mHRMA and the QMC-PCR was also tested with Cohen's Kappa statistics in SPSS. Analytical performances of the mHRMA assay were calculated using the online MEDCALC statistical software at https://www.medcalc.org/calc/diagnostic_test.php.

4.3 Results

4.3.1 Tm peaks intersection produces spurious melting patterns on mHRMA

While the Tm peaks of *TP53* exons 4d and 8b amplicons intersected or overlapped with each other close to their peaks, both were each sufficiently separate from the melt curve of the *SMAD4* exon 9a amplicon (Figures 4.2 and 4.3). With overlapping peaks, there was difficulty with attempts at normalisation of individual peaks or melt curve flexures. The Shifted Melt Curves and Difference Curves from multiplex targets with overlapping peaks or flexures showed spurious melting results which suggested that samples with known wild-type genotypes have mutations (Figure 4.2). On the other hand, when the melt curves were sufficiently separate and there was minimal or no overlap between the curves, the Call-IT displayed the correct or valid melting patterns for each of the samples (Figure 4.3). The multiplex PCRs were therefore performed with primers which amplify products with adequately separated Tm's.

4.3.2 GC-tagged primers increase Tm and enhance separation of multiplex peaks for robust mHRMA

Having shown that intersecting Tm peaks resulted in artefactual melting patterns, I proposed that that the addition of non-specific GC tags to the 5' end of natural primer sequences could alter the Tm's of the PCR products, enhance separation of multiplex Tm's and improve mHRM analysis. This proposal was tested *in-silico* and verified experimentally for a *KRAS* exons 2, 3 and 4 triplex and a *BRAF* exons 11 and 15 duplex. Primer pairs for *KRAS* exon3 and *BRAF* exon 15 were tagged with non-specific, GC-rich, 5' tails and used to perform *in silico* PCR using MFEprimer. The *in silico* PCR amplicon sequences of each multiplex groups were pasted on UMELT Batch 2.0 mode and melting was performed *in silico*.

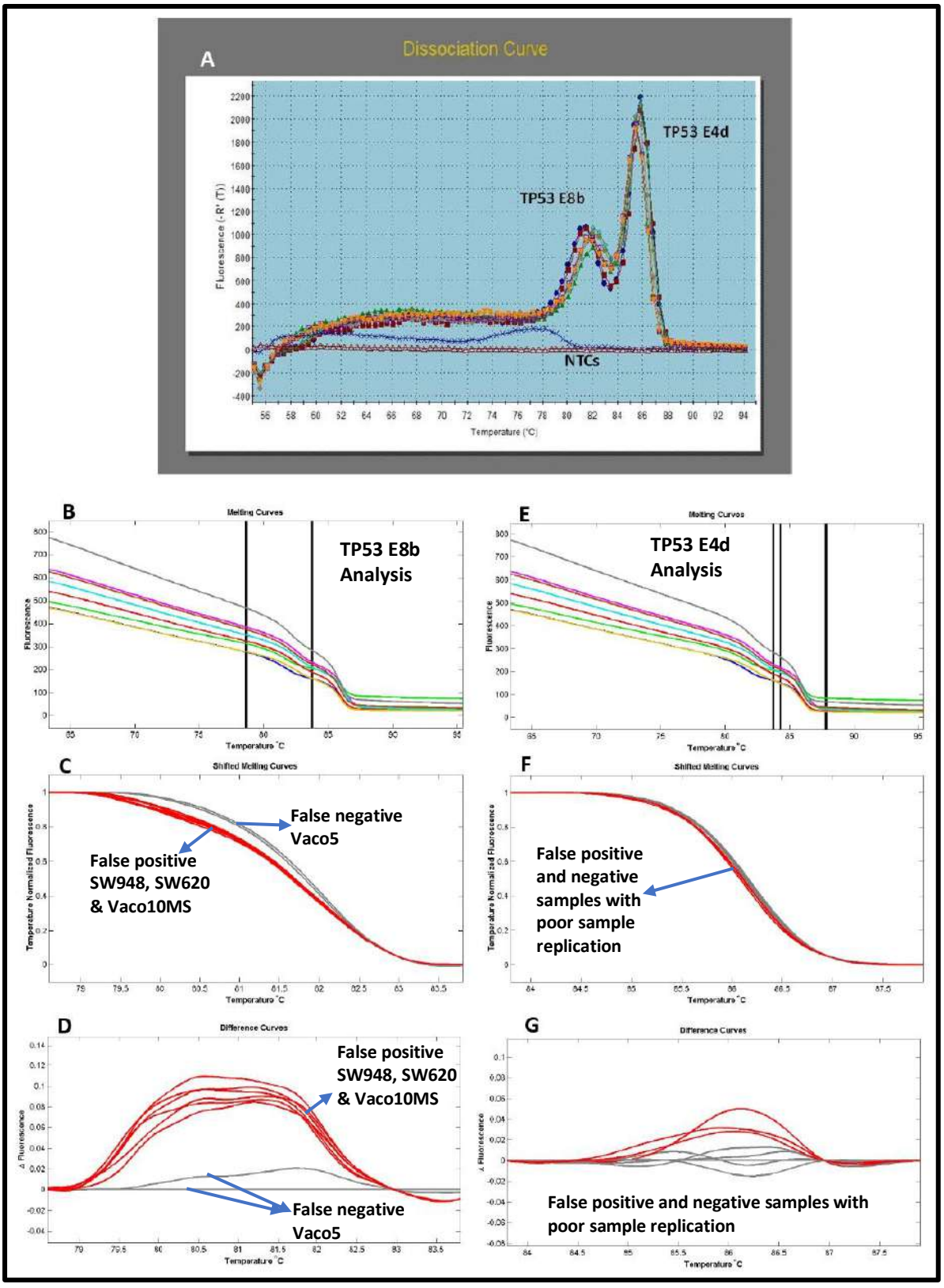


Figure 4.2: Vaco5, SW948, SW620 and Vaco10MS were utilised to demonstrate the effect of overlapping target peaks (TP53 E8b and TP53 E4d) on HRM analysis. Overlapping of target peaks (**A**) leads to difficult normalisation of individual target peaks (**B** and **E**). Poor normalisation, in turn, leads to spurious melting patterns and loss of precision (**C**, **D**, **F** and **G**).

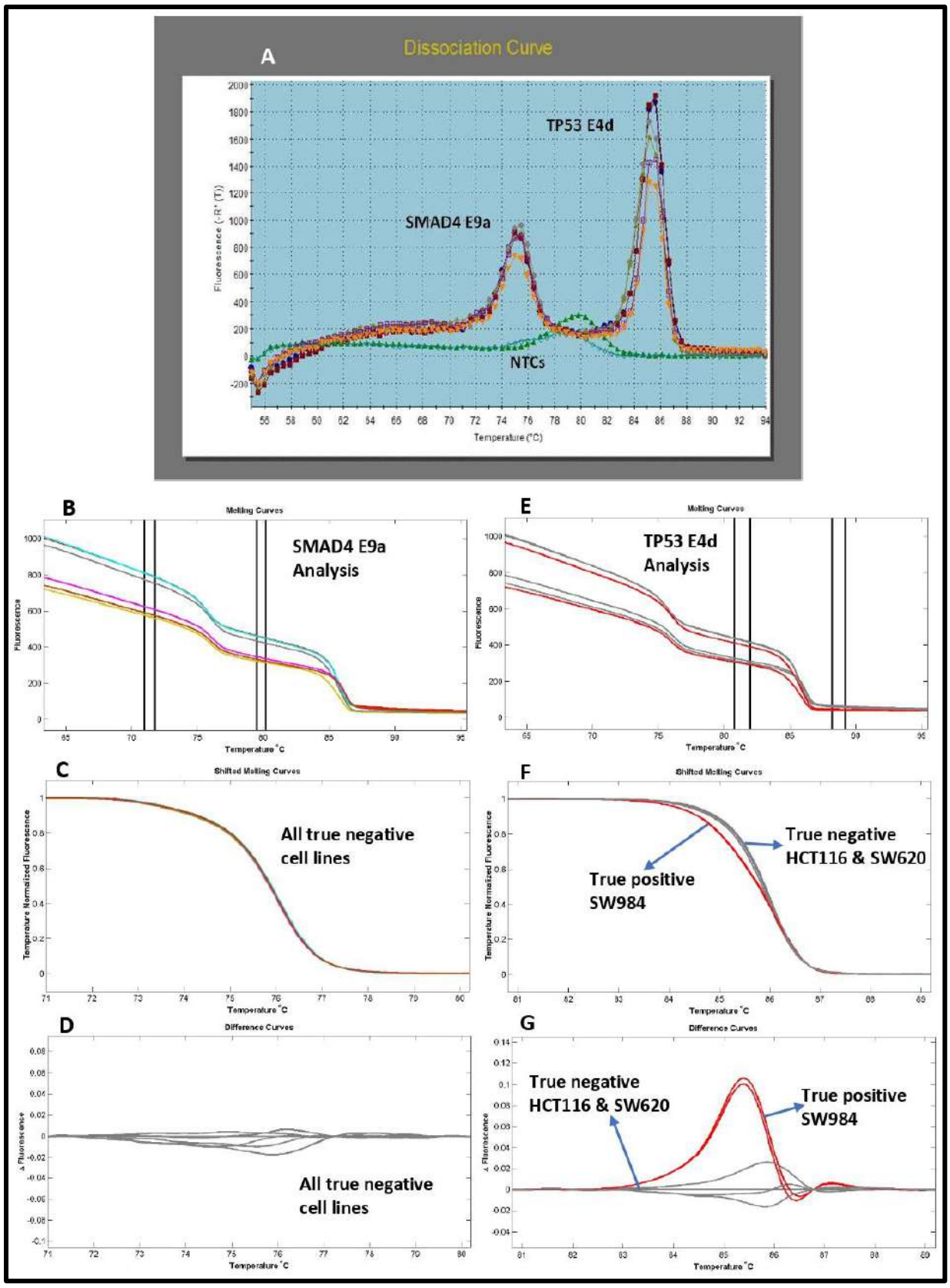


Figure 4.3: HRM analysis of SW948, HCT116 and SW620 duplex targets SMAD4 E9a and TP53 E4d which have discrete peaks. Sufficiently separate peaks (**A**) allow correct normalisation of individual target peaks (**B** and **E**) and produces accurate and reproducible high resolution melting patterns (**C**, **D**, **F** and **G**).

The results showed that the use of GC-rich primers for one or more targets of the multiplex groups enhanced the separation of the melting peaks of multiple amplicons through the increases in sequence Tm (Figure 4.4A). The *in silico* predictions were verified experimentally (Figure 4.4B).

4.3.3 mHRMA versus QMC-PCR

To validate mHRMA assay, I tested five targets each in the seventeen pre-amplified clinical samples, using a triplex (*PTEN* exon 3/*PIK3CA* exon 20/*KRAS* exons 3) and a duplex (*PTEN* exon 7/ *NRAS* exon 2) combination. The component targets of the multiplex groups were selected based on their individual Tm predictions. The *PTEN* exon-7/ *NRAS* exon-2 duplex showed adequate Tm separation using the standard primers. However, the adequate separation of the Tm peak triplex was only achieved when the primers of the *PIK3CA* exon 20 and *KRAS* exons 3 targets were tagged with GC-tails (see Table 4.1 and Figure 4.5).

A total of 85 targets were tested in the seventeen samples. There was a concordance of 100% between the mHRMA and QMC-PCR results. Cohen's test showed a Kappa of 1.00 (95% confidence interval, CI = 1.00 to 1.00, SE kappa=0.000). Analytical performance showed the following: Sensitivity =100.00% (95% CI = 54.07% to 100.00%), Specificity = 100.00% (95% CI=95.44% to 100.00%), Accuracy =100.00% (95% CI= 95.75% to 100.00%), Positive Predictive Value =100.00%, Negative Predictive Value = 100.00%.

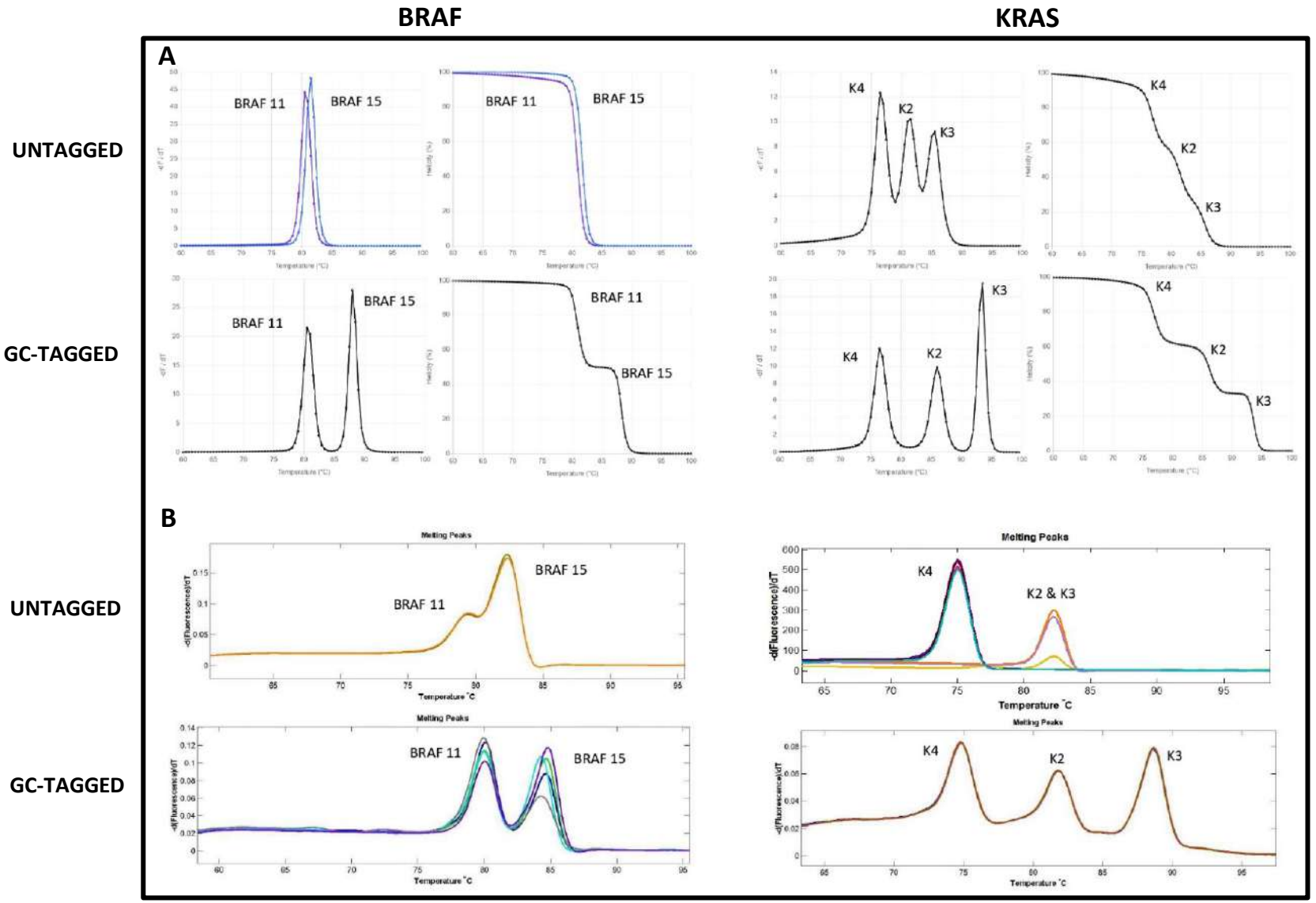


Figure 4.4: A. *In silico* demonstration of the capacity of GC-tagged primers to achieve adequate separation of Tm's for the components of a multiplex reaction. K4, K2 and K3 are *KRAS* exon 4, *KRAS* exon 2 and *KRAS* exon 3, respectively. In this analysis, *BRAF* exon 15 and *KRAS* exon 3 primer pairs were tagged with GC tails. **B.** Experimental demonstration of the capacity of GC-tagged primers to achieve adequate separation of Tm's for the components of a multiplex reaction. K4, K2 and K3 are *KRAS* exon 4, *KRAS* exon 2 and *KRAS* exon 3, respectively. The GC-tagged primer pairs were those of *BRAF* exon 15 and *KRAS* exon 3.

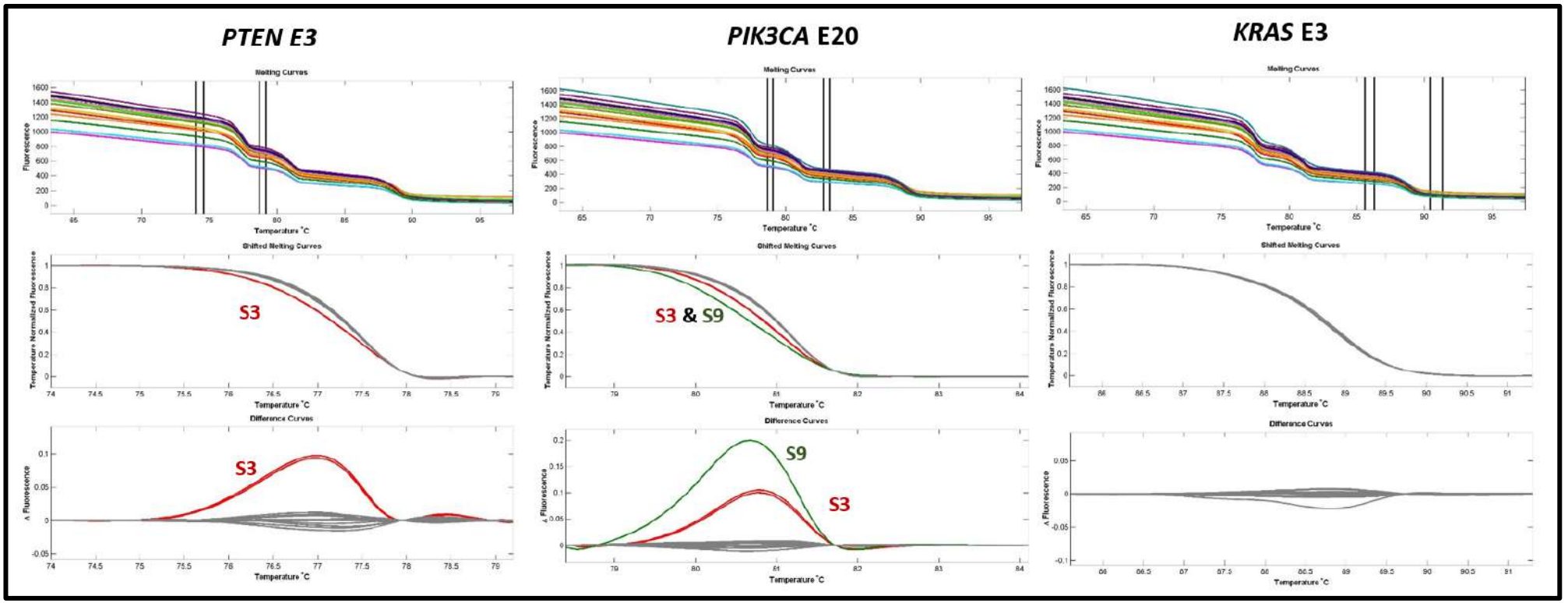


Figure 4.5: mHRMA showing melt curve data normalisation for each target; and detection of *PTEN* exon 3 and *PIK3CA* exon 20 mutations in sample 3, S3 (*PTEN* exon 3 and *PIK3CA* exon 20) and sample 9, S9 (PIK3CA exon 20). *PTEN* E3 =*PTEN* exon 3, *PIK3CA* E20 = *PIK3CA* exon 20, *KRAS* E3=*KRAS* exon 3. Adequate separation of the *PIK3CA* exon 20 and *KRAS* exon 3 targets was achieved by tagging the *KRAS* exon 3 primer pair with a GC-rich tail.

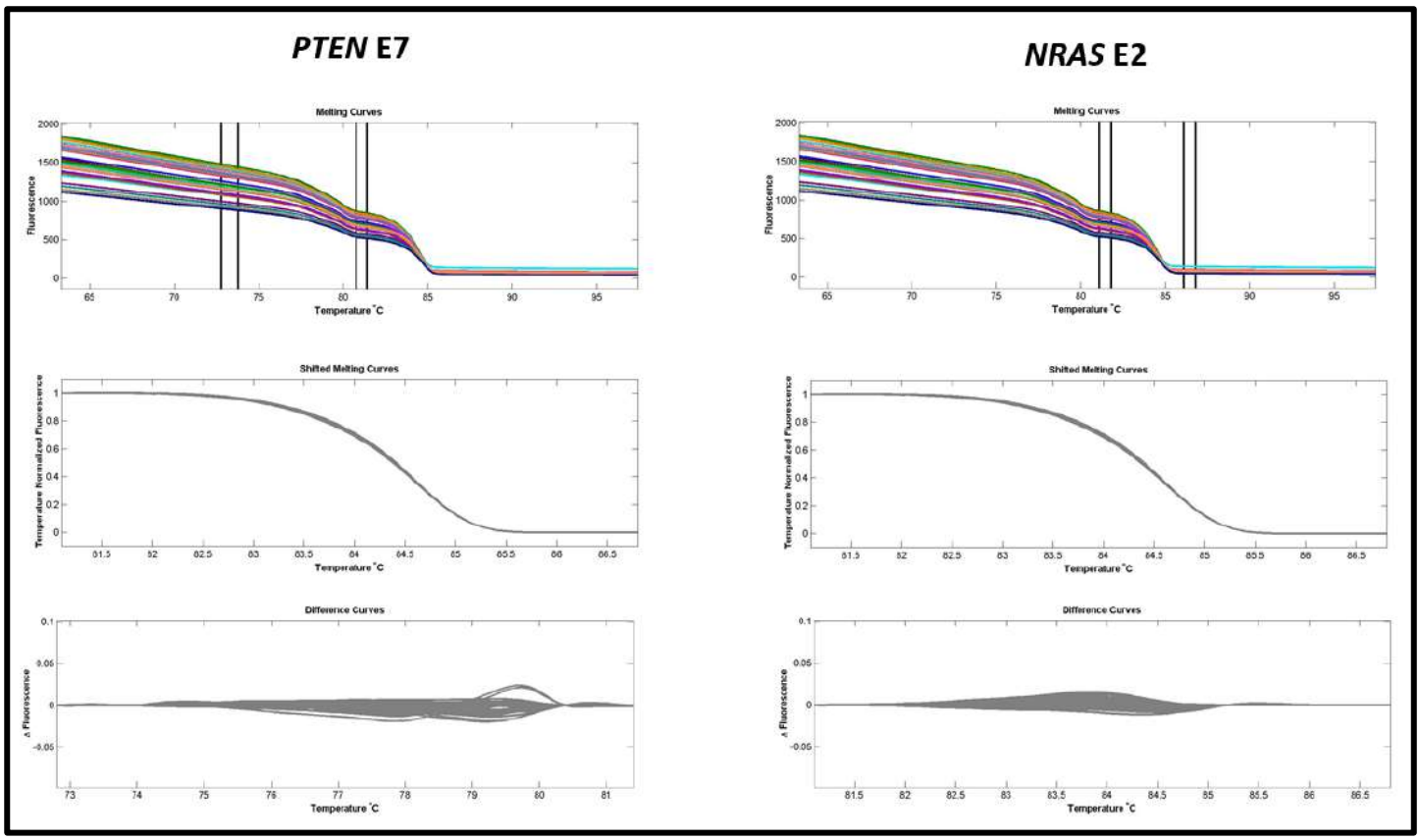


Figure 4.6: mHRMA showing no mutation was detected in the *PTEN* exon 7/ *NRAS* exon 2 duplex.

Table 4.2: mHRMA versus QMC PCR

TARGETS SAMPLES	*PTEN* E3		*PIK3CA* E20		*KRAS* E3		*NRAS* E2		*PTEN* E7	
	QMC PCR	mHRMA	QMC PCR	mHRMA	QMC PCR	mHRMA	QMC PCR	mHRMA	QMC PCR	mHRMA
S1	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S2	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S3	WT	WT	**M**	**M**	WT	WT	WT	WT	WT	WT
S4	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S5	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S6	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S7	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S8	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S9	**M**	**M**	**M**	**M**	WT	WT	WT	WT	WT	WT
S10	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
SS11	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S12	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S13	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S14	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S15	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S16	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT
S17	WT	WT	WT	WT	WT	WT	WT	WT	WT	WT

*WT = wild-type, M=mutant

4.4 Discussion

In the development of the mHRMA assay I showed that overlapping melting peaks results in spurious melting patterns and bogus mutation scanning results, and correct detection of mutation by mHRMA was only feasible when the target amplicons have widely varying Tm peaks that do not overlap or intersect. And I also showed that GC-tagged primers can be used to increase the melting temperatures of PCR amplicons, and this can be exploited to achieve adequate separation of melting peaks of the multiplex targets and enable correct normalisation of melt curves and valid HRM analysis. Finally, I validated the mHRMA assay against the QMC-PCR assay using 85 well-defined gene targets in 17 CRC cases. The mHRMA assay showed a perfect concordance and a high analytical performance relative to the QMC-PCR-HRMA assay. Although pre-amplified templates were used to demonstrate the validity of the mHRMA assay, the use of genomic DNA from CRC cell lines for the optimisation of the assay showed that the mHRMA can be formulated as a single stage PCR assay.

As HRM uses a single dye chemistry and its analysis depends on differences in melting temperature characteristics of amplicons, HRM multiplexing can only be achieved if the Tm of the target amplicons are discrete (Vossen et al., 2009). The maximum number of targets that can achieved per tube with the mHRMA assay was three (3). However, I believe that with further optimisation of melting peak separation strategy, one or two more targets can be added *in tube*. Previous studies have described assays which used HRM to detect mutations or SNPs in multiplexed targets. For example, while Seipp et al used a combination of GC-tailed, AT-tailed, and locked nucleic acid (LNA)-incorporating primers to separate the melt curves of the target amplicons, Pereyna et al and Helsmoortel et al simply used the differences in product sizes of 50-250 bp amplicons to adequately vary the Tm of the target amplicons, utilising conventional,

non-modified primer pairs (Helsmoortel et al., 2016; Pereyra et al., 2012; Seipp et al., 2008). More recently – about 4 years after the publication of this data in the Journal of Clinical Pathology under the title "QMC-PCRx : a novel method for rapid mutation detection" – Xiu et al used GC-tailed primers to achieve target amplicon Tm peaks spacing in a multiplex HRM assay that detected antibiotics-resistant strains of clinically important bacterial organisms (Xiu et al., 2020).

Multiplexing of targets in any PCR assay improves the speed of testing by reducing the on-the-bench preparation time since the number of reactions needed to be set up are reduced (Markoulatos et al., 2002). For example, only 34 reactions needed to be set up in the mHRMA assay to test a total of 85 targets in 17 samples. Comparatively, 85 QMC-PCR assay reactions were used to test the same number of targets in the same number of samples. Furthermore, multiplexing minimises the volume of reagents used for testing, and therefore reduces the cost of testing. While about 2,550 µL of master mix was expended to test the 85 targets in the SSD stage of the QMC-PCR assay, only about 1,020 µL was used in the mHRMA assay to test the same number of targets. (Refer to Appendices 1 and 2, and the accompanying Supplementary Tables 1 and 2 for a comparison of the cost estimates and turnaround times for deploying the newly developed assays relative to established assays).

The design and optimisation of the mHRMA assay was fairly easy and was greatly facilitated by bioinformatics prediction tool support (Dwight et al., 2011; Qu et al., 2012; Speir et al., 2016). The bioinformatics prediction of Tm's and Tm peaks separation closely matched the experimental findings (see Figure 4.4). Optimisation of PCR conditions for the mHRMA posed minimal challenges. For example, a primer concentration of 250nM was generally used throughout the experimentation. Only with the *NRAS* exon-2 target was further primer

concentration optimisation necessary to achieve optimum PCR results. Furthermore, optimisation of annealing temperature (Ta) for the multiplex reactions was not necessary as the individual primer pairs were initially designed to have an optimised Ta of wide ranges (see chapter 2). Thus, for the multiplex reactions I simply used the Ta of the primer pair with the lowest Ta in each of the multiplex groups.

The mHRMA assay was designed to detect point mutations and indels in multiple targets per run, and thus can be further optimised to serve as companion diagnostic tests for targeted therapeutic agents. The assay is scalable in terms of the number of gene targets per tube and in the number of tubes per sample, and thus may be able to accommodate – to a certain limit – the increasing number of targets that are continuously being discovered for therapy. Furthermore, as it uses an inexpensive mutation detection platform – HRM function is in-built in many real-time PCR machines in use in many standard molecular diagnostics laboratories (thereby obviating the need for investments in new equipment) and utilises cheap reagents, i.e., saturating fluorescent dyes – I project that the mHRMA assay would be a more cost-effective assay than the current commercially available multiplexing assays. Furthermore, aspects of mHRMA – from DNA extraction to PCR amplification - which are common to many other assays are amenable to automation as has been achieved with the ARMS-PCR in the Biocartis Idylla platform (Johnston et al, 2018). Moreover, the melting temperature data from HRMA can be used to train an Artificial Intelligence algorithm to achieve automation of the analysis aspect of HRMA (Echle et al, 2021, Bilal et al, 2021, Kather et al, 2019; 2020).

However, time constraints precluded a more comprehensive validation of the mHRMA. For example, the limits of detection of the multiplex primers could not be determined, and more mutated targets, especially the clinically actionable targets such as *KRAS* codons

12/13/59/61/117/146, *NRAS* codon 12/13/59/61, *BRAF* codon 600, etc. could not be included to increase the power of the assay. Only five mutated targets were included in the preliminary validation.

In conclusion, the development of the mHRMA has validated the concept that multiplexing of mutation targets can be achieved on the HRMA platform by the selective utilisation of GC-tagged primers to separate the temperature curves of target amplicons. The development also shows that mHRMA can have excellent performance indices. However, further validation of mHRMA is warranted to conclusively prove its performance, especially with respect to detecting clinically actionable mutations.

CHAPTER 5

MSI-HRMA FOR MICROSATELLITE INSTABILITY TESTING IN COLORECTAL CANCER

5.1 Background

Microsatellite instability (MSI), a well-established mechanism of tumourigenesis in cancer and a genetic maker of clinical importance, is caused by genetic and epigenetic alterations in the mismatch (MMR) repair genes *MLH1, MSH2, PMS2, MSH6,* etc. (Boland & Goel, 2010; Kanopiene et al., 2014; Muzny et al., 2012; Ratti et al., 2018; Yamamoto et al., 2020). MSI causes genome-wide insertions and deletions (indels) of one or more bases within microsatellites – genomic regions characterised by 5 or more repeats of one to six bases (Bonneville et al., 2017; Fujimoto et al., 2020; Schlötterer, 1998). The presence of MSI in any patient's tumour is an indication for further genetic evaluation to exclude or include Lynch syndrome (Moreira et al., 2012; Shia, 2008; L. Zhang, 2008; National Institute for Health and Care Excellence, 2017b). Furthermore, it is fairly established that gastric and colorectal cancers (GCa and CRC) with MSI have better prognoses than those lacking this feature (Beghelli et al., 2006; Gryfe et al., 2000; Polom et al., 2018; Popat et al., 2005; Smyth et al., 2017). Moreover, MSI is a predictive marker of response to 5-fluorouracil-based chemotherapy in CRC of stage II and below (Carethers et al., 2004; Ribic et al., 2003; Sargent et al., 2010). MSI has acquired added clinical relevance since 2017 when the Food and Drug Administration approved the use of the immune checkpoint inhibitors (ICIs) for the management of MSI-H cancers (Lemery et al., 2017; Overman et al., 2017, 2018, National Institute for Health and Care Excellence, 2021a; 2022).

Traditionally, MSI is detected in tumours using fluorescent PCR by which 5-7 microsatellite markers – including BAT-25, BAT-26, D5S346, D2S123, and D17S250 for the Bethesda panel (Boland et al., 1998); BAT-25, BAT-26, NR-21, NR-22,and NR-24 for the revised Bethesda panel (Suraweera et al., 2002; Umar et al., 2004); and BAT-25, BAT-26, NR-21, NR-24, MONO-27, Penta C and Pent D for the MSI Analysis System panel (Bacher et al., 2004; Murphy

et al., 2006) - are amplified and the PCR products detected by most commonly capillary electrophoresis (CE) (Bacher et al., 2004; Murphy et al., 2006). Furthermore, the detection of expression of the mismatch repair proteins MLH1, MSH2, PMS2 and MSH6 by immunohistochemistry (IHC) is used as a surrogate test for MSI (Shia, 2008; L. Zhang, 2008). The presence of instability in more than 40% of markers – instability is adjudged as differences in allele size of the marker in the tumour relative to normal DNA – is regarded as the cut-off for MSI in the patient's tumour (Bacher et al., 2004; Bocker et al., 1997; Boland et al., 1998; Murphy et al., 2006; Suraweera et al., 2002; Umar et al., 2004).

In this chapter I describe the development of MSI-HRMA, a high-resolution melting analysis-based assay for testing MSI in gastrointestinal cancers. Previously, Janavicius et al and Fadhil et al showed that HRMA is a feasible method for MSI testing (Fadhil et al., 2012; Janavicius et al., 2010) I expand on those previous works by showing that several other markers besides the more established microsatellite markers – BAT-25 and BAT-26 – can be utilised for MSI testing. My hypothesis is that any combination of 5 sensitive microsatellite markers tested on the HRMA platform is sufficient to classify a tumour sample into MSI or microsatellite-stable (MSS) tumour. This hypothesis is based on three principles (i) MSI is a genome-wide phenomenon in mismatch repair-deficient tumours (Bonneville et al., 2017; Fujimoto et al., 2020; Schlötterer, 1998), (ii) some microsatellite markers are mutated more frequently than others in specific tumour types (Boland & Goel, 2010; Ratti et al., 2018), (iii) HRMA is an exquisitely sensitive and accurate method of detection of short insertions and deletions (Gerini et al., 2014).

The rationale for developing a new MSI testing assay is to circumvent certain challenges encountered in the use of the gold standard capillary electrophoresis detection method (Bacher et al., 2004). These challenges include the high cost of acquisition of DNA analysers, as well as the

relatively high recurrent cost incurred in the use of fluorescent-labelled primers for STR (including MSI) analysis, the need to improve the sensitivity of detection method, ease of operation and ease of interpretation of MSI data (Lian et al., 2021). HRMA offers all these advantages, as have been mentioned previously in chapter four.

The aim of this development is to show that MSI can be tested using combinations of sensitive microsatellite markers on the HRMA platform. The objectives include (i) to identify sensitive microsatellite markers which can be grouped into different combinations for MSI testing (ii) to test the performance of the panels of five microsatellite markers against the CE- and IHC-based tests in a training cohort of clinical CRC cases, and (iii) to validate a panel of the most sensitive microsatellite markers using a large validation cohort of clinical CRC cases.

5.2 Materials and Methods

5.2.1 Materials

Nine (9) CRC cell lines with known microsatellite statuses were utilised to validate the newly designed MSI testing PCR primers. These include MSS cell lines SW480, SW620, HUTU80 and SW837, and MSI cell lines DLD1, HCT 116, LoVo, LS1034 and RKO.

A training cohort of 44 FFPE samples were used to identify frequently mutated microsatellite loci in CRC, and to test the different combinations of microsatellite markers for sensitivity at detection of MSI. Of these 44 samples 23 were classed as MSI by CE and IHC while 8 were MSS. Thirteen (13) of the 44 samples were normal DNA from FFPE tissues. The validation cohort, comprising 122 FFPE CRC samples – with known MMR status by IHC - was used to validate the MSI-HRMA assay. All the CRC samples had known MSI/MMR statuses determined by IHC (all 153 cases) and CE (33/44 training cohort cases) (Figure 5.1 and 5.2).

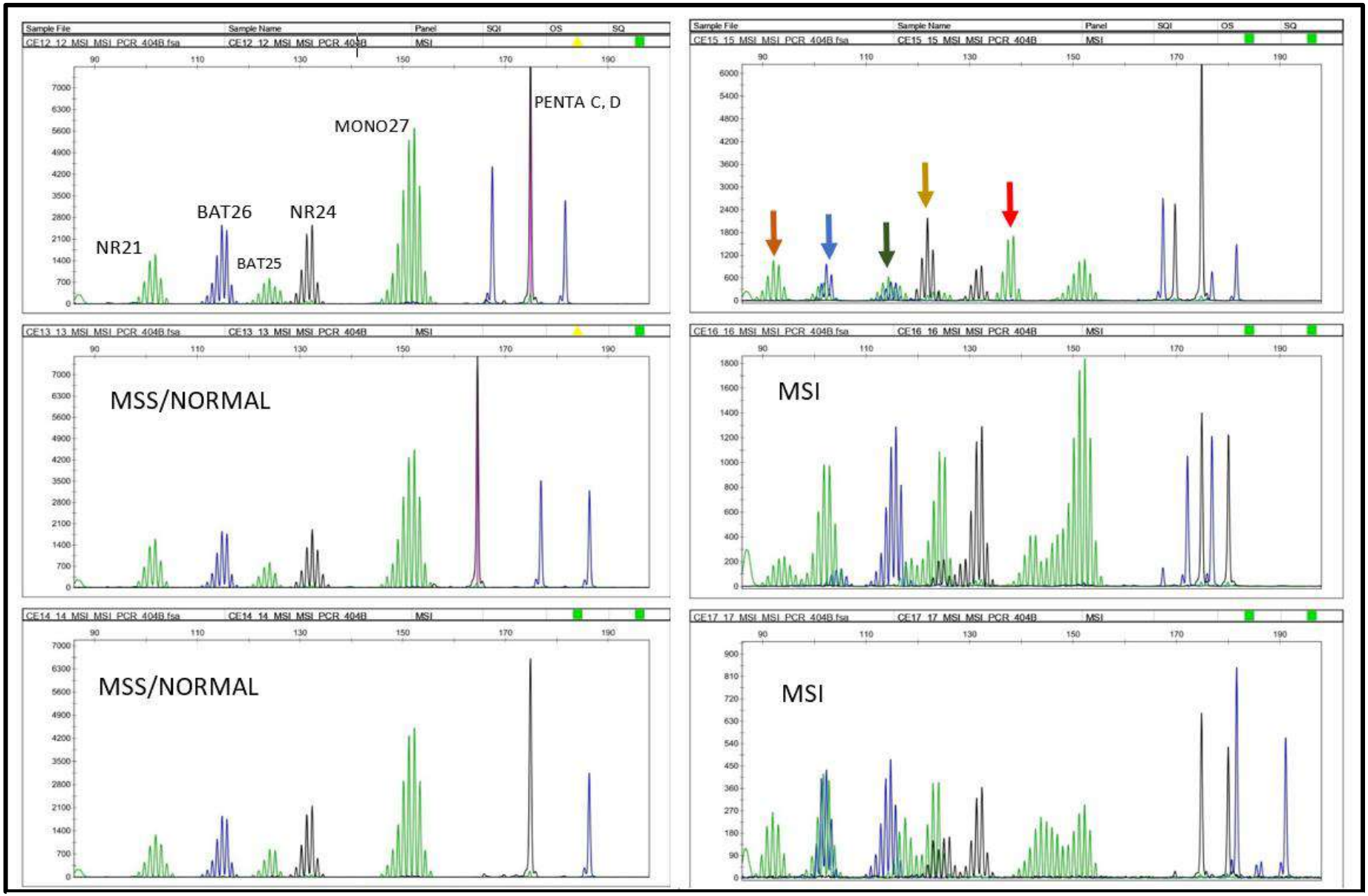


Figure 5.1: Capillary electropherogram showing MSI status by fragment analysis of mononucleotide markers NR21 (100bp), BAT26 (115bp), BAT25 (125bp), NR24 (130bp) and MONO27 (150bp). The Penta C and D markers are gender markers for patient identification A. Typical electropherogram findings in MSS/normal cases are single distinct peaks for each microsatellite marker B. Electropherogram for MSI cases showing multiple peaks for each of the markers with allelic size shift for NR21 (orange arrow), BAT26 (blue arrow), BAT25 (green arrow), NR24 (yellow arrow) and MONO27 (red arrow).

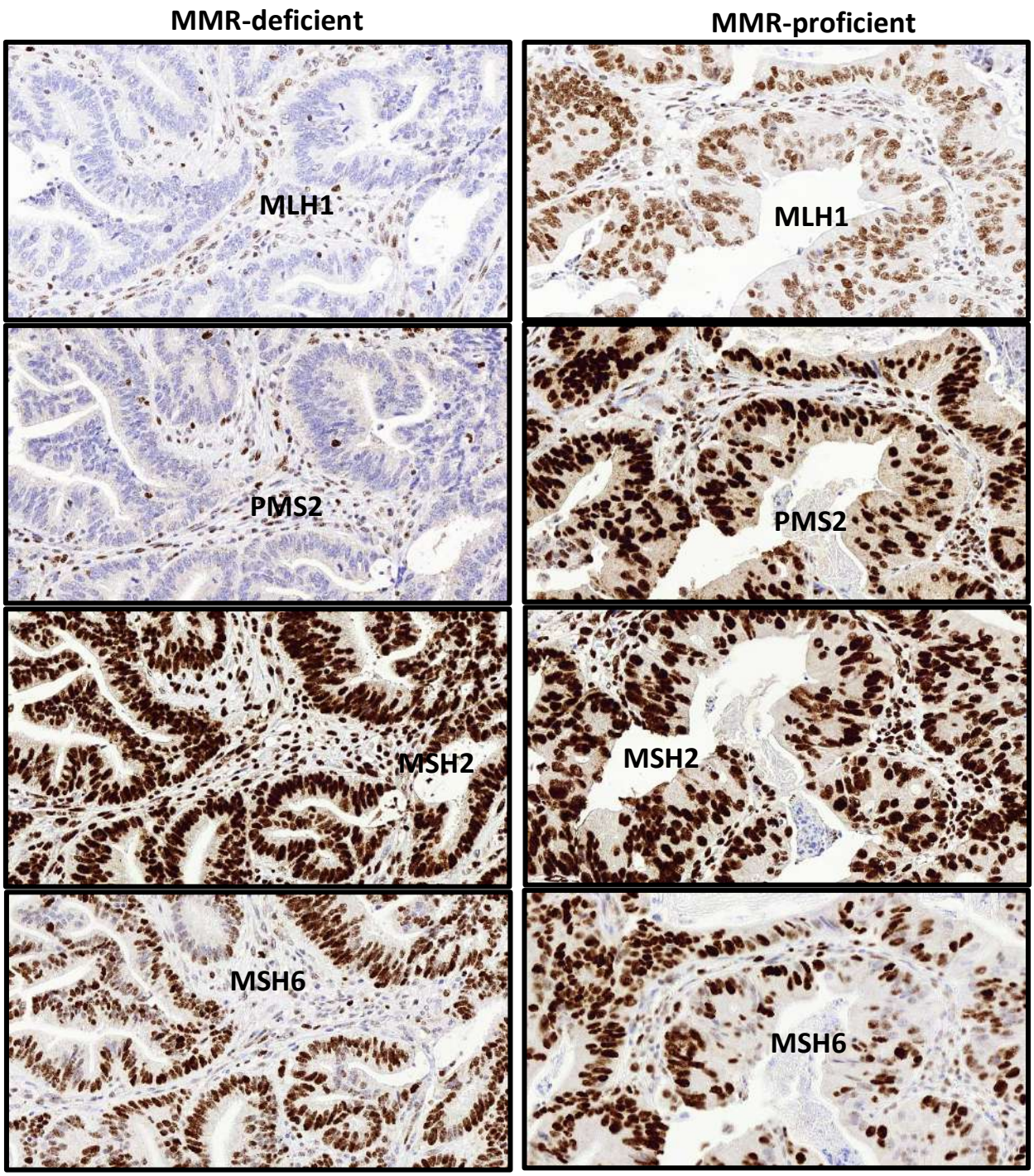


Figure 5.2: Immunohistochemistry of expression of four mismatch repair (MMR) proteins MLH1, PMS2, MSH2 and MSH6 in CRC. The MMR-deficient CRC shows simultaneous loss of expression of MLH1 and PMS2, while the MMR-proficient case show expression of all MMR proteins. Loss of PMS2 was used as a negative control for MLH1 loss, but not vice versa. Both internal (stromal cells in the tumour body) and external (glandular and stromal cells in normal colon tissue) were included in the interpretation of IHC data.

5.2.2 Microsatellite markers identification and selection

A total of eleven (11) mononucleotide microsatellite markers were identified from published literature and from databases. BAT-25, BAT-26 and MONO-27 are already established markers for MSI testing (Bacher et al., 2004; Boland et al., 1998; Murphy et al., 2006; Suraweera et al., 2002; Umar et al., 2004), while BCAT-25 was formerly identified by our research group as a sensitive microsatellite marker (Fadhil et al., 2012). TYMS-26 and EWSR1 were identified from literature search (Ghosh et al., 2012; Kishore et al., 2014; Kondelin et al., 2015), while MYB, ANGEL2, TP53 (BAT34CA), FBXO46 and TCF4 were retrieved from the SelTar database of altered mononucleotide repeats in colorectal, gastric and endometrial cancers (www.seltarbase.org) (Woerner et al., 2010) (Table 5.1).The individual markers were selected from SelTar database if (i) the mononucleotide repeat size is greater than 10, and (ii) if the published rate of mutation in MSI tumours (CRC, GCa and EC) is greater than 80%. For all the markers except TYMS-26, the published primer sequences were retrieved, either from the published articles or from the SelTar database, then in silico PCRs was performed on MFEprimer. The resulting amplicon sequence was then used to redesign *HRMA-specific* primers as stated in chapter 2. For the TYMS-26 marker, the genomic coordinate of the TYMS T26 5'-UTR microsatellite was retrieved from the Ghosh et al., 2012 article, and was used to retrieve the microsatellite and flanking sequences from UCSC Genome using the code https://genome.ucsc.edu/cgi-bin/das/hg19/dna?segment=chr18:656907,656984 (26 bp on either side of the microsatellite). The TYMSS-26 primers were designed from the sequence as before.

Table 5.1: Characteristics of genomic markers selected for MSI testing

MARKERS	GENOME LOCATIONS	GENES/POSITIONS	SOURCES OF MARKERS*	PUBLISHED MUTATION RATES	PCR PRODUCT SIZES	REPEAT SIZES	OPTIMUM PRIMER Ta
BCAT 25	3p22.1	*CTNNB1*/3'UTR	Literature	Unknown	100bp	T25	50-60^{0}C
BAT 25	4q12	*CKIT*/intron 13	Literature	96 -100%	85bp	T25	46-64^{0}C
BAT 26	2p21-p16.3	*MSH2*/intron 5	Literature	92.5 -100%	77bp	A26	45-65^{0}C
ANGEL2	1q32.3	*ANGEL2*/transcribed pseudogene	www.seltarbase.org	97.8%	65bp	T17	
EWSR1	22q12.2	*EWSR1*/3'UTR	Literature	99.5%	70bp	T16	45-60^{0}C
FBXO46	19q13.32	*FBXO46*/3'UTR	www.seltarbase.org	95 -100%	57bp	T13	
TP53(BAT34CA)	17p13.1	*TP53*/3'UTR	www.seltarbase.org	87.5%	69bp	A18	47-65^{0}C
TCF4	18q21.2	*TCF4*/intronic	www.seltarbase.org	83.5%	54bp	T13	47-65^{0}C
MONO27	2p22.1	*MAP4K3*/intron 13	Literature Also on www.seltarbase.org	98.3%	78bp	A27	55-65^{0}C
MYB	6q23.3	*MYB*/intronic	www.seltarbase.org	100%	75bp	T23	47-65^{0}C
TYMS26	18p11.32	*TYMS*/5'UTR	Literature	Unknown	78bp	A26	

5.2.3 Primer design, optimisation and validation and PCR strategy

In order to determine which PCR strategy would give the optimum results for MSI testing of FFPE DNA using the MSI-MRMA assay, I compared the conventional single stage PCR with QMC-PCR protocol. QMC-PCR was designed by our research group to achieve optimum amplification results from degraded DNA templates such as those obtained from FFPE tissues (Fadhil et al., 2010). Therefore, outer and inner nested primer pairs design for all the microsatellite markers was accomplished manually as stated in chapter 2, whilst putting into consideration all the features necessary to accomplish optimum QMC-PCR experiments (Fadhil et al., 2010). All the primers utilised for the development of the MSI-HRMA assay are shown in Appendix Table 5.1.

Primer optimisation was performed with gradient PCR-HRMA as described in chapter 2 (Figure 2.3, chapter 2). Gradient PCR was performed on the Primus96 machine whilst all HRMA experiments were performed with the LightScanner-96 instrument. PCR experiments for primer validation and cohort testing were carried out on the MxPro 3005 real time PCR machine.

Microsatellite marker primer pairs were validated using the MSI and MSS cell lines (Figure 2.6, chapter 2).

5.2.4 Cohort testing

Cohort testing was carried out in order to (i) test the sensitivity and specificity of the individual microsatellite markers, and to enable the grouping of the eleven markers into panels, (ii) compare the quality of amplification between QMC-PCR and the standard PCR as measured by the accuracy of MSI testing by HRMA. Of the 44 samples that comprised the training cohort, 24 CRC cases (23 MSI and 1 MSS) and 2 normal samples were amplified at all eleven

microsatellite marker loci using the QMC-PCR and then subjected to HRMA in the LightScanner. Furthermore, the same 26 samples were amplified at the newly selected microsatellite markers (TYMS-26, EWSR1, MYB, ANGEL2, TP53 (BAT34CA), FBXO46 and TCF4) and MONO-27 using conventional, single stage PCR. The amplicons were subjected to HRMA, and the results compared with data obtained with QMC-PCR amplification.

Having determined that there was no significant superiority of amplification quality of the QMC-PCR over standard PCR for the MSI-HRMA assay, a further 13 CRC cases (all MSS) and 5 normal DNA samples belonging to the training cohort were amplified at the TYMS-26, EWSR1, MYB, ANGEL2, TP53 (BAT34CA), FBXO46 and TCF4 and MONO-27 loci using only the single stage PCR.

Moreover, the validation of the assay was performed with standard single stage PCR on 122 CRC cases.

5.2.5 Combination of microsatellite markers into panels

All the eleven microsatellite markers were grouped into six (6) five-marker MSI panels. MSI panels 1-4 comprised random assortments of the markers whilst panels 5 and 6 comprised more established/most sensitive markers and the least sensitive and/or specific markers, respectively. A sample was regarded as MSI when it showed positivity for at least 4 out of the 5 markers in each of the panels. The data obtained from each of the panels were compared to the CE and IHC data using the appropriate statistical tests.

5.2.6 Validation of MSI-HRMA assay

Validation of the MSI-HRMA assay was accomplished using markers BAT-25, BAT-26, BCAT-25, TP53, EWSR1 and MYB on the 122 cases that comprised the validation cohort. The threshold for MSI-positivity was taken as 5 out of 6 markers. The data obtained from this validation stage were compared to the IHC data using the appropriate statistics.

5.2.7 Statistical Analyses

The MSI-HRMA and CE/IHC data were input into SPSS version 24 and analysed with the appropriate statistical tests. Analytical performances (sensitivity, specificity, negative predictive value, positive predictive value, and accuracy) of the 11 individual microsatellite markers were tested using the indices true negative, true positive, false negative, and false positive all relative to the results of the "gold standards" IHC and CE. The analytical performance indices were calculated using online MEDCALC statistical software at https://www.medcalc.org/calc/diagnostic_test.php and the results were output as value + 95% confidence interval. Then, the 11 individual markers were grouped into 6 5-marker panels using criteria described in section 5.2.5 above, and the agreement between the IHC/CE gold standards and the MSI-HRMA was tested using the Cohen's Kappa statistics and the Area Under the Receiver Operating characteristic (AUROC) curve, both of which are performed in SPSS. Furthermore, the analytical performances (sensitivity, specificity, negative predictive value, positive predictive value, and accuracy) of the individual panel of markers were calculated using online MEDCALC and the results were output as value + 95% confidence interval.

5.3 Results

5.3.1 QMC-PCR versus single stage PCR

The data obtained by single stage PCR-based MSI-HRMA was compared to that obtained with the QMC-PCR-based assay. There was 100% concordance between the QMC PCR-based MSI-HRMA and the standard PCR-based MSI-HRMA assays (Appendix Table 5.2), evidence that the standard PCR gave an optimum quality amplification that enabled robust HRMA. In terms of detection of MSI both PCR strategies were similar, with no superiority of one over the other. But standard PCR was chosen for the MSI-HRMA assay because it is a single-stage PCR method and would therefore be easier to operate than using the QMC-PCR – which is a two-stage PCR technique – for the MSI-HRMA assay.

5.3.2 Training cohort

In total, 26 of the 44 samples (23 MSI, 1 MSS and 2 normal samples) were amplified at BCAT-25, BAT-25 and BAT-26 loci, while all 44 samples (23 MSI, 8 MSS and 13 normal samples) were amplified at TP53, EWSR1, MYB, FBXO46, ANGEL2, MONO-27, TYMS26 and TCF4 microsatellite loci (Table 5.2). All the eleven selected microsatellite markers showed high analytical performances at detecting MSI in the training cohort. Whilst the BCAT-25, BAT-25 and BAT-26 markers individually showed the highest sensitivity, specificity, and accuracy at MSI detection, as did some of the newly selected markers such as TP53, EWSR1 and MYB, the remaining markers, namely FBXO46, ANGEL2, MONO27, TYMS26 and TCF4 all showed a relatively reduced - albeit very high - performances. Figure 5.3 and Table 5.3 show the results of testing with the selected markers in relation to CE and IHC data.

A total of 26 samples (including 23 MSI, 1 MSS and 2 normal samples) had complete data for MSI panels 1-3 and 5, while all 44 samples had complete data for panels 4 and 6 (Appendix

Tables 5.3 – 5.8). As stated in section 5.2.5 above, MSI panels 1-4 comprised random assortments of the markers whilst panels 5 and 6 comprised more established/most sensitive markers and the least sensitive and/or specific markers, respectively. A sample was regarded as MSI when it showed positivity for at least 4 out of the 5 markers in each of the panels. The Kappa statistics, AUROC analyses and analytical performances of MSI panels 1 - 6 relative to the CE/IHC data show excellent performances of the MSI panels (Tables 5.4 – 5.6), confirming the hypothesis that any combination of 5 sensitive microsatellite markers tested on the HRMA platform is sufficient to classify tumour samples into MSI or MSS.

Table 5.2: MSI testing of the training cohort by selected microsatellite markers

SERIAL	ID	IHC	BCAT25	hBAT25	hBAT26	ANGEL2	EWSR1	FBXO46	MONO27	MYB	TCF4	TP53	TYMS26
1	Si #1	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
2	Si #2	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
3	Si #3	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
4	Si #4	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
5	Si#5	**dMSH2**	MSI	MSS	MSS	MSS	MSI	MSI	MSS	MSI	MSS	MSI	MSS
6	Si #6	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
7	Si #7	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
8	Si #8	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
9	Si #9	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	**MSS**	MSI	MSI
10	Si #10	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
11	Si #11	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	**MSS**	MSI	MSI	**MSS**
12	Si #12	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
13	Si #13	**dMLH1**	MSI	MSI	MSI	MSI	**MSS**	MSI	**MSS**	MSI	MSI	MSI	MSI
14	Si #14	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
15	Si #15	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
16	Si #16	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
17	Si #17	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
18	Si #18	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
19	Si #19	**dMLH1**	MSI	MSI	MSI	**MSS**	MSI	MSI	MSI	MSI	MSI	MSI	MSI
20	Si #20	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
21	Si #21	**dMSH2**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
22	Si #22	**dMSH2**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
23	Si #23	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI	MSI
24	Rx8	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
25	Rx24	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
26	Rx266	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
27	Rx 27	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
28	bx24	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
29	bx30	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
30	SS7	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
31	SS8	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
32	SS9	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
33	SS10	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
34	SS11	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
35	SS12	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
36	SS14	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
37	SS15	**p-MMR**	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
38	BX36	**p-MMR**	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
39	BX25	**p-MMR**	MSS	MSS	MSS	MSS	MSS	**MSI**	MSS	MSS	MSS	MSS	MSS
40	BX20	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	**MSI**	MSS	MSS	MSS	MSS
41	BX21	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
42	BX14	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	MSS	MSS	MSS	MSS	MSS
43	BX28	**p-MMR**	-------	-------	-------	MSS	MSS	MSS	**MSI**	MSS	MSS	MSS	MSS
44	BX31	**p-MMR**	-------	-------	-------	MSS	MSS	**MSI**	MSS	MSS	MSS	MSS	MSS

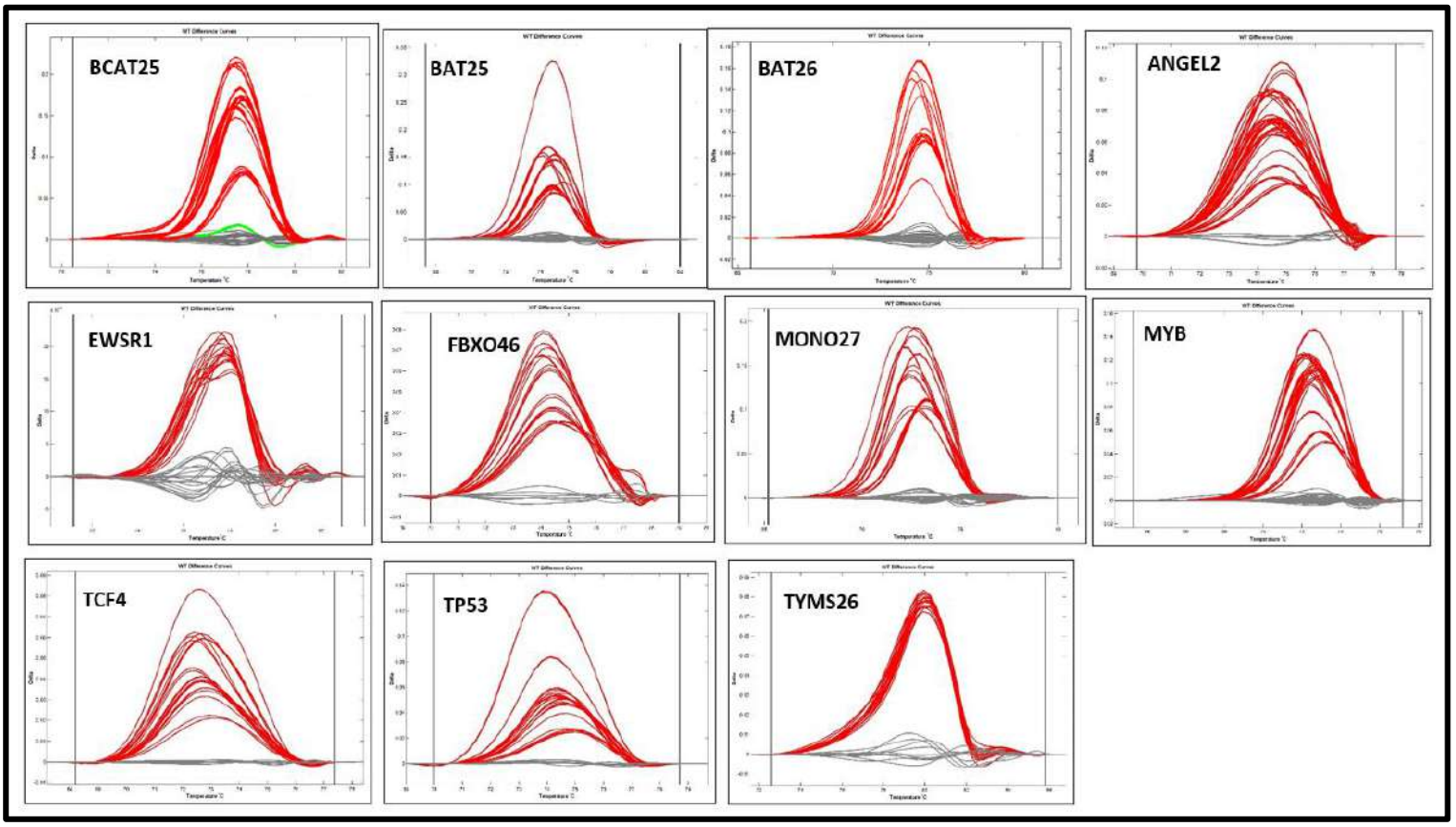


Figure 5.3: HRMA of the microsatellite markers in the training cohort of CRC show high performance indices. Only the Difference Curves are shown here.

Table 5.3: Analytical performance of the eleven microsatellite markers

MICROSATELLITE MARKER	TRUE POSITIVE	TRUE NEGATIVE	FALSE POSITIVE	FALSE NEGATIVE	SENSITIVITY (%)	SPECIFICITY (%)	PPV (%)	NPV (%)	ACCURACY (%)
BCAT25	23	3	0	0	100	100	100	100	100
BAT25	22	3	0	1	95.7	100	100	99.2	99.4
BAT26	22	3	0	1	95.7	100	100	99.2	99.4
ANGEL2	21	21	0	2	91.3	100	100	98.5	98.7
EWSR1	22	21	0	1	95.7	100	100	99.2	99.4
FBXO46	23	19	2	0	100	90.5	65.0	100	91.9
TCF4	21	21	0	2	91.3	100	100	98.5	98.7
TP53	23	21	0	0	100	100	100	100	100
MYB	22	21	0	1	95.7	100	100	99.2	99.4
MONO27	21	19	2	2	91.3	90.5	62.9	98.3	90.6
TYMS26	21	21	0	2	91.3	100	100	98.5	98.7

*PPV = positive predictive value; *NPV = negative predictive value

Table 5.4: Concordance* tests for microsatellite markers panels

MSI PANEL	Sample size	Kappa value	Standard error	*P*-value
MSI Panel 1	25	0.835	0.159	0.002
MSI Panel 2	25	0.835	0.159	0.002
MSI Panel 3	25	1.000	0.000	<0.001
MSI Panel 4	44	0.909	0.062	<0.001
MSI Panel 5	25	0.835	0.159	0.002
MSI Panel 6	44	0.955	0.045	<0.001

*Relative to IHC and fluorescent PCR-CE

Table 5.5: AUROC analysis of the microsatellite marker panels

MSI PANEL	Sample size	Area under curve (95% C.I.)	Standard Error	*P*-value
MSI Panel 1	25	0.978 (0.924 -1.000)	0.028	0.008
MSI Panel 2	25	0.978 (0.924 -1.000)	0.028	0.008
MSI Panel 3	25	1.000 (1.000 – 1.000)	0.000	0.006
MSI Panel 4	44	0.957 (0.878 – 1.000)	0.040	0.011
MSI Panel 5	25	0.978 (0.924 -1.000)	0.028	0.008
MSI Panel 6	44	0.978 (0.924 -1.000)	0.028	0.008

Table 5.6: Analytical performance of the microsatellite marker panels

MSI PANEL	Sample size	Sensitivity (95% C.I.)	Specificity (95% C.I.)	PPV	NPV (95% C.I.)	Accuracy (95% C.I.)
MSI Panel 1	25	95.65% (78.05 -99.89%)	100% (29.24-100%)	100%	99.24% (95.04-99.89%)	99.35% (85.58-100.0%)
MSI Panel 2	25	95.65% (78.05 -99.89%)	100% (29.24-100%)	100%	99.24% (95.04-99.89%)	99.35% (85.58-100.0%)
MSI Panel 3	25	100% (85.18 – 100%)	100% (29.24-100%)	100%	100%	100% (86.77 – 100.0%)
MSI Panel 4	44	91.30% (71.96 – 98.93%)	100% (83.89 -100%)	100%	98.49% (94.55-99.59%)	98.7% (89.58– 100.0%)
MSI Panel 5	25	95.65% (78.05 -99.89%)	100% (29.24-100%)	100%	99.24% (95.04-99.89%)	99.35% (85.58-100.0%)
MSI Panel 6	44	95.65% (78.05 -99.89%)	100% (83.89 -100%)	100%	99.24% (95.04-99.89%)	99.35% (90.73-100.0%)

5.3.3 Validation cohort

The 122 CRC cases in the validation cohort were amplified at the BAT-25, BAT-26, BCAT-25, TP53, EWSR1 and MYB microsatellite loci using single stage PCR and then subjected to HRMA. The results obtained for the individual markers were compared with IHC data. The markers showed the similarly high levels of sensitivity and specificity at MSI detection as was obtained with the training cohort (Table 5.7). Using a cut-off of 5 out 6 markers, the panel of markers showed 100% concordance and a kappa of 1.00 (SE=0.000) with the IHC data, showing that MSI-HRMA is a robust and accurate assay for MSI testing in CRC. The analytical performance of the MSI-HRMA in comparison to the IHC showed the following the following indices: Sensitivity =100.00% (95% CI = 81.47% to 100.00%), Specificity = 100.00% (95% CI= 96.52% to 100.00%), Accuracy =100.00% (95% CI= 97.02% to 100.00%), Positive Predictive Value =100.00%, Negative Predictive Value = 100.00%.

Table 5.7: MSI testing of the validation cohort by MSI-HRMA validation panel

SERIAL NO	SAMPLE ID	IHC STATUS	BAT25	BAT26	BCAT25	EWSR1	MYB	TP53	MSI-HRMA
1	1	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
2	20	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
3	23	p-MMR	MSS	MSS	MSI	MSS	MSS	MSS	**MSS**
4	32	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
5	33	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
6	51	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
7	71	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
8	76	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
9	87	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
10	93	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
11	95	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
12	119	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
13	129	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
14	130	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
15	141	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
16	155	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
17	167	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
18	178	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
19	179	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
20	183	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**

21	186	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
22	187	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
23	198	p-MMR	MSS	MSS	MSI	MSS	MSS	MSS	**MSS**
24	215	p-MMR	MSS	MSS	MSS	MSS	MSI	MSS	**MSS**
25	221	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
26	226	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
27	230	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
28	242	p-MMR	MSS	MSS	MSI	MSS	MSI	MSS	**MSS**
29	250	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
30	253	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
31	255	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
32	257	p-MMR	MSS	MSS	MSI	MSS	MSS	MSS	**MSS**
33	261	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
34	262	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
35	265	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
36	267	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
37	271	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
38	276	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
39	287	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
40	297	p-MMR	MSS	MSS	MSI	MSS	MSS	MSS	**MSS**
41	299	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
42	303	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
43	321	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
44	336	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
45	371	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
46	376	**dMSH6**	MSS	MSI	MSI	MSI	MSI	MSI	**MSI**
47	385	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
48	392	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
49	393	p-MMR	MSS	MSS	MSI	MSS	MSS	MSS	**MSS**
50	394	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
51	404	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
52	411	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
53	417	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
54	419	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
55	420	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
56	442	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
57	443	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
58	446	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
59	466	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
60	477	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
61	478	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
62	483	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
63	485	p-MMR	MSS	MSS	MSS	MSS	MSI	MSS	**MSS**
64	501	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
65	503	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
66	506	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
67	509	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
68	520	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
69	522	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
70	525	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
71	530	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
72	537	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
73	550	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**

74	557	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
75	606	**dMSH2**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
76	608	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
77	614	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
78	615	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
79	640	p-MMR	MSS	MSS	MSS	MSS	MSI	MSS	**MSS**
80	646	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
81	648	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
82	658	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
83	661	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
84	663	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
85	673	p-MMR	MSS	MSS	MSI	MSS	MSS	MSS	**MSS**
86	701	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
87	729	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
88	743	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
89	748	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
90	758	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
91	765	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
92	768	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
93	774	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
94	778	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
95	779	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
96	782	**dMSH2**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
97	784	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
98	794	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
99	808	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
100	814	**dMSH2**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
101	821	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
102	824	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
103	825	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
104	828	p-MMR	MSS	MSS	MSS	MSS	MSI	MSS	**MSS**
105	880	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
106	883	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
107	887	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
108	893	**dMLH1**	MSI	MSI	MSI	MSI	MSI	MSI	**MSI**
109	896	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
110	900	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
111	901	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
112	917	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
113	922	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
114	931	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
115	932	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
116	940	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
117	948	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
118	957	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
119	967	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
120	971	p-MMR	MSS	MSS	MSI	MSS	MSS	MSS	**MSS**
121	975	p-MMR	MSS	MSS	MSS	MSS	MSS	MSS	**MSS**
122	982	p-MMR	MSS	MSS	MSI	MSS	MSS	MSS	**MSS**

5.4 Discussion

This study has not only validated the two previous studies that tested the feasibility of MSI testing on the HRMA platform, but it has also shown that besides the established and commercially validated microsatellite markers, MSI testing can be accomplished using novel, well-selected microsatellite markers, since MSI is a genome-wide phenomenon in tumours with defective MMR (Fadhil et al., 2012; Janavicius et al., 2010).

In this study I have opted to select and validate only mononucleotide markers for the MSI-HRMA assay due to the established fact that these markers are far more easily mutated and therefore much more sensitive than the higher order repeats (Murphy et al., 2006; Suraweera et al., 2002; Umar et al., 2004). The performance indices of the novel microsatellite markers are comparable to those of the established markers, at least on the HRMA platform. These performance levels were retained when the markers were grouped into panels of differing marker combinations. It must be noted that the obtained performance indices of MSI panels 2, 4 and 5 would have been higher if the stringent cut-off of 4 out of 5 markers were relaxed to the conventional 2 out of 5 markers. However, I have opted to use the 4-out-of-5 cut-off because of the potential existence of single nucleotide polymorphisms (SNPs) at or near the microsatellite repeats (Press et al., 2014; Sonay et al., 2015; Willems et al., 2014), which SNPs can give rise to false positive tests with one or more individual markers. The use of a higher threshold for calling MSI positivity reduces the probability of such false positive results.

Case number 5 (si#5) in the training cohort reduced the performance scores of the markers and panels when applying the stringent MSI-HRMA scoring criterion. According to the IHC data the defective MMR protein in si#5 was MSH2, however only 5 markers showed MSI positivity with the MSI-HRMA assay. The same equivocal result was obtained with MSI detection by capillary

electrophoresis. A plausible explanation for this observation is that intra-tumour clonal heterogeneity, a well-recognised cause of intra-tumour diversity (McGranahan & Swanton, 2017), resulted in MSI heterogeneity, which, combined with the possible differences in the limits of detection of the MSI markers, produced an equivocal MSI status for si#5. However, it must be noted that si#5 would be classifiable as MSI by 4 of the 6 MSI panels, including the panel of least sensitive/specific markers (panel 6) (see Appendix Tables 5.3-5.8) if the stringent MSI-HRMA criterion were relaxed to the standard ≥2/5 markers threshold (Fadhil et al., 2012; Janavicius et al., 2010).

The current standard test for MSI is short tandem repeat analysis of the panel of microsatellite markers by fluorescent PCR and CE (Bacher et al., 2004; Boland et al., 1998; Murphy et al., 2006; Suraweera et al., 2002; Umar et al., 2004). Moreover, the use of NGS-based MSI tests have recently been recommended and is being gradually introduced into clinical and research practice (Bonneville et al., 2020; Y. Lee et al., 2021; National Comprehensive Cancer Network, 2021b; Zheng et al., 2020). However, MSI-HRMA can find a place in MSI testing in molecular diagnostic laboratories because of its favourable profile (Nestarenkaite et al., 2020; Susanti et al., 2018). Apart from its analytical performance shown in this study, the MSI-HRMA test is easier to deploy both in terms of cost of initial investment in equipment and recurrent costs of reagents and consumables. For example, MSI-HRMA may not require additional investments in instruments beyond the real-time PCR machine that is available in every molecular diagnostic laboratory, in contrast to the deployment of CE and NGS for MSI testing, which would entail the purchase of the costly DNA Analyzer and NGS machine, respectively (CADTH, 2014; Lian et al., 2021; L. Liu et al., 2012). Furthermore, MSI-HRMA is easier to operate and has shorter turnaround time than either CE or NGS, as it requires neither prolonged and cumbersome post-

PCR procedures or complex bioinformatics analyses (Bonneville et al., 2020; Gerini et al., 2014; Y. Lee et al., 2021; Lian et al., 2021; Zheng et al., 2020). (Refer to Appendices 1 and 2, and the accompanying Supplementary Tables 1 and 2 for a comparison of the cost estimates and turnaround times for deploying the newly developed assays relative to established assays). In addition, as a PCR-based test MSI-HRMA may be technically more suitable for microsatellite analysis than NGS as deep sequencing may induce spurious mutations at microsatellite loci due to the known difficulty of NGS in sequencing repeats regions of the genome (Tørresen et al., 2019; Treangen & Salzberg, 2012).

Cohen's Kappa statistics is used to test inter-observer or inter-rater reliability. The statistics represents the extent to which the diagnostic test results represent correctly the indices measured. Cohen's Kappa factors chance agreements between two diagnostic assessments into its equations. In practice, a novel diagnostic test with a Kappa value less than 0.80 are deemed unacceptable (moderate, weak, minimal and no levels of agreements) because only 63% or less of data generated by such a test is reliable. On the other hand, Kappa values of 0.8-0.9 are classed as having "strong levels of agreements" with the gold standards, while more than 0.9 is regarded as "almost perfect level of agreements". The data generated by a diagnostic test with strong agreement is deemed to have a data reliability of 64-81% while that with almost perfect agreement with the gold standards is 82-100% reliable. (McHugh, 2012). The AUROC statistics is used to determine the diagnostic ability of a novel diagnostic test relative to a gold standard. An AUC value of 0.5 suggests that the new test has no power to discriminate between normal and the disease condition under evaluation - or in the context of evaluating unstable microsatellites in CRC - between presence (MSI) and absence (MSS) of microsatellite

instability; 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding (Mandrekar, 2010).

MSI-HRMA is amenable to the multiplexing of the MSI markers. In chapter 3 of this thesis, I described the mHRMA assay which allows the multiplexing of two to three markers of sufficiently separate melting temperatures (Tm's) *per tube* (Ebili, Hassall, Asiri, et al., 2017). The multiplexing of the MSI markers for the HRMA platform would potentially further reduce the cost and labour of MSI testing, as well as improve the turnaround time of the MSI-HRMA assay.

In the development of this assay, I investigated which of the two PCR methods - the QMC-PCR and conventional single stage PCR - would produce a quality of amplification that would be more suitable for the MSI-HRMA assay and found that the conventional single stage PCR produced PCR amplification which is robust enough for microsatellite analysis. Furthermore, the use of a single stage PCR for the MSI-HRMA has advantages over the QMC-PCR in terms of cost, labour and turnaround around time. However, far more DNA templates were used for the single-stage PCR than would otherwise be used if the QMC-PCR had been deployed for amplification (Fadhil et al., 2010).

In conclusion, this chapter has described the development of the MSI-HRMA and has shown that combinations of sensitive microsatellite markers on the HRMA platform can be utilised for MSI testing.

CHAPTER 6

MSI-WES: A SIMPLE APPROACH FOR MICROSATELLITE INSTABILITY TESTING USING WHOLE EXOME SEQUENCING

6.1 Background

Gastric cancer (GCa) is the 5th most common cancer worldwide and the 4th most common cause of cancer deaths, contributing 7.7% of cancer mortality (Sung et al., 2021). The subtyping of GCa using different subtyping schemes has revealed a recurrent molecular subtype: the microsatellite instability (MSI) subtype (Bass et al., 2014; Cristescu et al., 2015; Setia et al., 2016). This molecular subtype of gastric cancer constitutes 15-26.4% of gastric cancers (Bass et al., 2014; Cristescu et al., 2015; Polom et al., 2018; Setia et al., 2016; Vos et al., 2021) and displays defined clinicopathological features, including age (older patients) (Bass et al., 2014; Cristescu et al., 2015; Polom et al., 2018; Setia et al., 2016; Vos et al., 2021) and gender (female) (Bass et al., 2014; Polom et al., 2018; Vos et al., 2021) biases, the intestinal histological type according to the Lauren classification (Cristescu et al., 2015; Polom et al., 2018; Vos et al., 2021), distal (pyloric) location (Cristescu et al., 2015; Polom et al., 2018; Vos et al., 2021), early-stage disease (Cristescu et al., 2015; Polom et al., 2018), and paucity of lymph node metastasis (Cristescu et al., 2015; Polom et al., 2018). But most significantly, the MSI GCa subtype shows a peculiar pattern of prognosis which is related to chemotherapy and disease stage. MSI-high GCa subtypes exhibit better survival profiles than MSS/MSI-L subtypes only in stages II and III tumours when the patients are treated with surgery without chemotherapy (S. Y. Kim et al., 2015). These favourable profiles are attenuated when chemotherapy is included in the patients' treatment regimen. Furthermore, the MSI-H GCa subtype was recently shown to exhibit a much better response to immunotherapy with pembrolizumab compared to the MSS/MSI-L GCa subtype (S. T. Kim et al., 2018). MSI testing is therefore an important aspect of the management of GCa.

Routine MSI testing is achieved with fluorescent PCR followed by capillary electrophoresis (Bacher et al., 2004; Bocker et al., 1997; Boland et al., 1998; Murphy et al., 2006; Suraweera et al., 2002; Umar et al., 2004). However, with the gradual translation of next generation sequencing (NGS) into clinical practice it has become possible to test hundreds to thousands of clinically relevant molecular markers of diseases in single experiments (Kamps et al., 2017). NGS technologies – including targeted and whole exome sequencing – are gradually gaining relevance in the clinical management of cancers and other disorders (Kamps et al., 2017). Recently, the NCCN and NICE approved the use of targeted and whole exome sequencing for MSI testing, especially in patients with metastatic diseases (National Comprehensive Cancer Network, 2021b, UK Department of Health & Social Care, 2020; NHS England, 2021a, 2021b; Richman et al, 2017; Richman et al, 2018).

Informative NGS data from microsatellite loci provide veritable discriminative power between MSI and MSS tumours (Kautto et al., 2017; Niu et al., 2014; Pabla et al., 2019; Salipante et al., 2014; Trabucco et al., 2019). The use of NGS for MSI testing offers certain advantages over the conventional methods. Perhaps the most important advantage would be that MSI analysis can be easily incorporated into already existing NGS workflow or pipeline for mutation analysis without the need for additional or separate experiments (Kautto et al., 2017; Salipante et al., 2014). This contrasts with the conventional methods in which such additional, wet lab experiments would be required for MSI testing (Kautto et al., 2017). In addition, as NGS tests several clinically relevant variants in single experiments, the assessment of hundreds to thousands of microsatellite loci, rather than the five to seven loci tested by conventional MSI testing assays, would reduce the risk of misclassification of MSI cases and increase sensitivity and specificity of the MSI assay (Kautto et al., 2017; Salipante et al., 2014). Furthermore, the fact that the gene targets of

MSI differ from one cancer type to another makes the use of the five- to seven-marker PCR-based MSI testing methods grossly inadequate for investigating MSI in previously uncharacterised cancer types (Kautto et al., 2017; Salipante et al., 2014). In contrast, an individual NGS-based approach for MSI testing could be sufficiently applied to uncharacterised cancer types since it assesses hundreds to thousands of microsatellite loci simultaneously (de la Chapelle & Hampel, 2010).

A few methods for MSI testing by NGS have been described in the literature, and these have used different approaches of varying complexities and NGS data of different levels of processing (Hause et al., 2016; Kautto et al., 2017; Lu et al., 2013; Ni Huang et al., 2015; Niu et al., 2014; Pabla et al., 2019; Salipante et al., 2014; Trabucco et al., 2019). Whilst some NGS-based MSI-testing approaches have used the raw data from BAM files (Kautto et al., 2017; Niu et al., 2014; Pabla et al., 2019; Salipante et al., 2014; Trabucco et al., 2019) and have computed MSI scores from the fraction or proportion of allelic sizes at microsatellite loci by utilising different statistical methods (Hause et al., 2016; Kautto et al., 2017; Lu et al., 2013; Ni Huang et al., 2015; Niu et al., 2014; Pabla et al., 2019; Salipante et al., 2014; Trabucco et al., 2019), some others have used processed/partially processed (level 2) data in mutation annotation format (maf) files (Ni Huang et al., 2015). In particular, Huang *et al* applied the approach of simply counting the indels within simple repeat sequences in maf files (Ni Huang et al., 2015). Although the use of raw or processed data in combination with statistical methods of different complexities have all shown good performances in MSI testing (Hause et al., 2016; Kautto et al., 2017; Lu et al., 2013; Ni Huang et al., 2015; Niu et al., 2014; Salipante et al., 2014), the use of partially processed NGS data, in my view, would entail a faster operation than the use of level 1 data (Ni Huang et al., 2015).

In this study, I aimed to demonstrate that MSI-WES, an approach which directly counts insertions and deletions within microsatellite loci in partially processed (level 2) WES data in vcf files, is a valid and adequately sensitive approach for MSI testing in cancers. The study hypothesis is that simply counting the mutated microsatellites in adequately quality-controlled somatic mutation data generated by WES can be utilised to test for MSI in tumours.

6.2 Materials and Methods

6.2.1 Study Cohorts

This study analysed the genomic data of 441 gastric cancer cases. All the genomic data was generated by The Cancer Genome Atlas (TCGA) Network's Pan-Cancer Atlas initiative (Y. Liu et al., 2018) and deposited in the National Cancer Institute's Genomic Data Commons (GDC) repository (www.portal.gdc.cancer.gov/repository) from where they were downloaded and analysed. The genomic coordinates of all the somatic mutation data used in this study were based on the GRCh38 (hg38) genome assembly.

Levels 2 and 3 data comprising annotated somatic mutation data in variant call format (vcf) (level 2), DNA methylation beta values and gene expression quantification datasets (level 3) were included in the study. The study cohorts comprised 441 GCa cases with somatic mutations, DNA methylation, gene expression and PCR-derived MSI data (Y. Liu et al., 2018).

Next generation sequencing technology was used to generate the somatic mutation (whole exome sequencing, WES), gene expression (RNASeq) while Methylation Array was utilised to produce the DNA methylation data. The PCR-based MSI status data (MSI status_pcr) was generated experimentally by PCR and capillary electrophoresis using the MSI-Mono-Dinucleotide Assay

marker panel (comprising four mononucleotide markers BAT25, BAT26, BAT40 and TGFRBII; and three CA dinucleotide repeat markers D2S123, D5S346 and D17S250) (Y. Liu et al., 2018).

The DNA methylation and gene expression data are open-access and were freely obtained from the GDC repository. The annotated somatic mutation vcf files are controlled-access and were also obtained from the GDC repository following the approvals of the International Cancer Genome Consortium (ICGC), dbGap and the University of Nottingham (ICGC reference number DAC0-1081299, NIH approval number # 88362-1). The PCR-based MSI status data were obtained from the reference (Y. Liu et al., 2018).

A total of 383 cases of the GCa were previously classified into the microsatellite status groups microsatellite stable (MS, n=256), microsatellite unstable-low (MSI-L, n=52), and microsatellite unstable-high (MSI-H, n=75). A total of 58 cases had unknown microsatellite status (Y. Liu et al., 2018). However, for the purpose of this study the tumours were reclassified as microsatellite stable MSS (MS and MSI-L, n=308), and microsatellite unstable, MSI (MSI-H, n=75) (Figure 6.1). This classification is based on the finding that MS and MSI-L tumours have shown no significant biological, clinical, or molecular differences (Hause et al., 2016; Kim et al., 2013; Ni Huang et al., 2015; Pawlik et al., 2004).

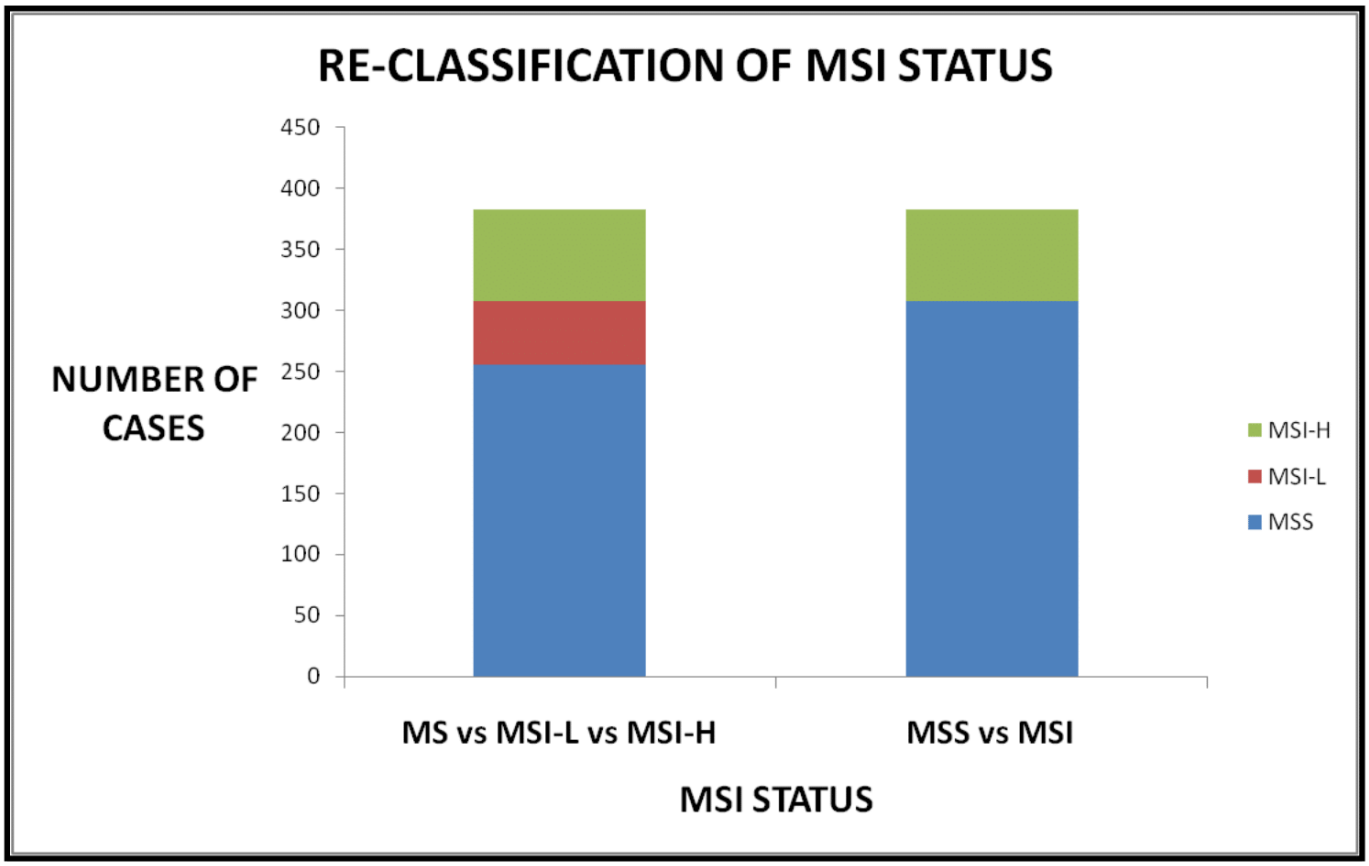


Figure 6.1: Re-classification of the MSI status of gastric cancer using the MSI-versus-MSS scheme

6.2.2 Genomic Data Analyses

Data analysis was accomplished in this study using codes and scripts which were written by using a combination of simple unix commands such as 'grep', 'awk', 'cut', 'paste', 'cat', 'sed', 'sort', 'uniq', and 'wc' in cygwin, a programming and runtime environment designed to run in Microsoft Windows (Racine, 2000) (see chapter 2 for the codes used in genomic data analyses).

6.2.3 Quality Control of mutation data

The vcf files generated by the VarScan2 variant caller were used to demonstrate this MSI-testing approach. For quality control of the VarScan2 vcf files only the variants marked “SOMATIC” and have a “PASS” FILTER were included in the input files.

6.2.4 Study Approach and Principles

The MSI-WES approach entails counting the number of small indels for each case and expressing this count as a fraction (MSI index_ngs) of the total size of genome sequenced (3088 Mb). The MSI-WES is enabled by the characteristic format in which indels are recorded in the vcf files. Deleted or inserted bases are recorded with their 5' flanking bases in the reference (deletions) or alternative (insertions) columns. This format facilitates easy retrieval of microsatellite-based indels from the data. A list of potential microsatellite-based indel formats was compiled (Figure 6.2) and then used to retrieve the mutated microsatellite loci from each of the vcf files. However, as about 90% of small indels caused by MSI are single-base mutations (Meier et al., 2018) whose microsatellite sequence contexts are not captured in the vcf files, an essential step to using the MSI-WES is to verify the sequence context of the one- to four-base indels by retrieving their flanking sequences from the NCBI Genome database. For the purpose of this study a microsatellite was defined as a sequence of one to six bases repeated five or more times, i.e., a mononucleotide repeated 5 or more times.

TG	AC	TA	CG	GA	CT
TGG	ACC	TAA	CGG	GAA	CTT
TGGG	ACCC	TAAA	CGGG	GAAA	CTTT
TGGGG	ACCCC	TAAAA	CGGGG	GAAAA	CTTTT
TGGGGG	ACCCCC	TAAAAA	CGGGGG	GAAAAA	CTTTTT
TGGGGGG	ACCCCCC	TAAAAAA	CGGGGGG	GAAAAAA	CTTTTTT
TGGGGGGG	ACCCCCCC	TAAAAAAA	CGGGGGGG	GAAAAAAA	CTTTTTTT
TGGGGGGGG	ACCCCCCCC	TAAAAAAAA	CGGGGGGGG	GAAAAAAAA	CTTTTTTTT
TGGGGGGGGG	ACCCCCCCCC	TAAAAAAAAA	CGGGGGGGGG	GAAAAAAAAA	CTTTTTTTTT
AG	GC	GT	TC	CA	AT
AGG	GCC	GTT	TCC	CAA	ATT
AGGG	GCCC	GTTT	TCCC	CAAA	ATTT
AGGGG	GCCCC	GTTTT	TCCCC	CAAAA	ATTTT
AGGGGG	GCCCCC	GTTTTT	TCCCCC	CAAAAA	ATTTTT
AGGGGGG	GCCCCCC	GTTTTTT	TCCCCCC	CAAAAAA	ATTTTTT
AGGGGGGG	GCCCCCCC	GTTTTTTT	TCCCCCCC	CAAAAAAA	ATTTTTTT
AGGGGGGGG	GCCCCCCCC	GTTTTTTTT	TCCCCCCCC	CAAAAAAAA	ATTTTTTTT

Figure 6.2: A list of potential microsatellite-based indels in variant call formats.

The underlying principles of the MSI-WES approach is that if the derived NGS-based MSI indices were valid measures of the MSI status, then they should show good correlations with DNA mismatch repair (MMR) gene expression, methylation status, and the molecular targets of MSI (MTM) index as previously established; and most importantly, they should show very good concordance, in the least, with the PCR-based MSI status (MSI status_pcr).

The MSI index_ngs was calculated from the mutated microsatellite counts and then dichotomised into a nominal variable, the MSI status_ngs. Both MSI index_ngs and MSI status_ngs were validated using the MSI status_pcr, MMR gene methylation, MMR gene expression, and the MTM index (see below).

6.2.5 Validation of NGS-based MSI Indices

The established somatic mutations targets of MSI in gastric cancer (*TGFBRII, IGFIIR, TCF4, RIZ, BAX, CASPASE5, FAS, BCL10, APAF1, hMSH6, hMSH3, MED1, RAD50, BLM, ATR, MRE11;* see Appendix Table 6.1) (Ratti et al., 2018) were retrieved from the somatic mutation datasets and used to generate the MTM index for each of the cases in the vcf datasets thus: mutations in any target gene was given a score of 1, and the total number of scores were tallied for each case and expressed as a fraction of the total number of target genes probed. The MSI indices and the MTM indices were compared for any correlation using Pearson's correlation analysis.

The ENSEMBL gene identifiers of the MMR genes (*MLH1, MSH2, MSH3, MSH6, PMS1,* and *PMS2*) were retrieved from www.ensembl.org/ (Appendix Table 6.2) and used to probe the gene expression files of the GCa cohort. The gene expression levels (in fragments per kilobase of transcript per million mapped reads, FPKM) were compared with the MSI indices for any correlations using Pearson's correlation analysis.

In addition, the methylation statuses of the MMR genes were determined by selecting the DNA methylation probes with consistent normo-methylation levels (using a beta value of 0.3 or less (Losi et al., 2018; Warden et al., 2013) in normal DNA. Then, statistical associations between all MSI measures (PCR- and NGS-based) and methylation status of the selected probes were sought by the chi square and Eta tests. The levels of associations and correlations were compared among the MSI measures.

Finally, the level of concordance between the NGS-based and the PCR-based MSI statuses were sought using the Cohen's Kappa test.

6.2.6 Statistical Analyses

Mutated microsatellite counts were output in Excel spreadsheet from cygwin to enable further direct calculations of MSI and MTM indices scores, and statistical analyses with *MLH1* expression and *MLH1* promoter methylation. All statistical analyses were performed in SPSS version 21 and a *P* level of <0.05 was regarded as significant. Correction for multiple testing was accomplished with Benjamini-Hochberg test, where applicable. Associations between two or more categorical variables were evaluated using the Chi Square (or Fisher's) test, while correlations between continuous variables were tested with Univariate or Multivariate Analyses. Correlations between categorical and continuous variables were sought with the Eta statistics while concordance between two qualitative tests were calculated using the Cohen's Kappa statistics. Analytical performance of the MSI-WES was calculated using the online MEDCALC statistical software at https://www.medcalc.org/calc/diagnostic_test.php.

Figures 6.1 and 6.2 were produced in Excel spreadsheet and Word, respectively; while Figure 6.3 was produced using the online scatter plot maker scatterplot.online at https://scatterplot.online/.

6.3 Results

6.3.1 Confirmation of the microsatellite status of small indels

A total of 70,149 small indels were retrieved from the entire GCa cohort. Their genomic coordinates were used to retrieve the flanking sequences of the one- to four-base indels from the UCSC Genome database. The results showed that over 90% of the small indels occur within perfect mononucleotide microsatellites. The remaining 10% include mixed mononucleotide repeats (such as AAAAGGG, GGGAAAA, etc), dinucleotide repeats (GTGTGT, ACACAC), imperfect or interrupted mononucleotide repeats, and trinucleotide repeats. The implication of this finding is that all the small indels in GCa occur within microsatellite repeats and are therefore caused by MSI.

6.3.2 NGS-based MSI index determination

Having confirmed the microsatellite status of the small indels, the list of potentially mutated microsatellite loci was used to probe, retrieve, and count the number of mutated microsatellites in the individual vcf files, from which the MSI index_ngs scores were calculated per case (Appendix Tables 6.3) The MSI index_ngs clearly distinguished the MSI from the MSS cases as determined by PCR (Figure 6.3). While the MSS cases tightly clustered together, the MSI cases were widely dispersed; evidence that the MSI group is highly heterogeneous with respect to the MSI index_ngs scores. The MSI index also showed a good correlation with the PCR-based MSI status (R=0.869, Eta test).

6.3.3 Threshold for dichotomising MSI index_ngs

A one-dimensional k-means cluster analysis was performed with the 383 cases with PCR-based MSI data to determine the threshold for dichotomising the MSI index_ngs into MSI status_ngs. Using the MSI index_ngs as the only variable, the number of clusters was progressively

increased until no change was observed in the number of items (cases) in the largest cluster. Chi square test confirmed that all but one case in this cluster were PCR-based MSS cases. On the other hand, no PCR-based MSS case was found in the other clusters. The mean (± 3 standard deviations, 3SD) of the MSI index_ngs score in the MSS group was 0.0112 ±0.0174 (or a range of -0.0062 to 0.0286). All cases with MSI index_ngs scores within this range were regarded as MSS_ngs while those cases with an MSI index_ngs score above 0.0286 were regarded as MSI_ngs. Based on this threshold 91/441 MSI_ngs and 350/441 MSS_ngs cases were found in the GCa cohort.

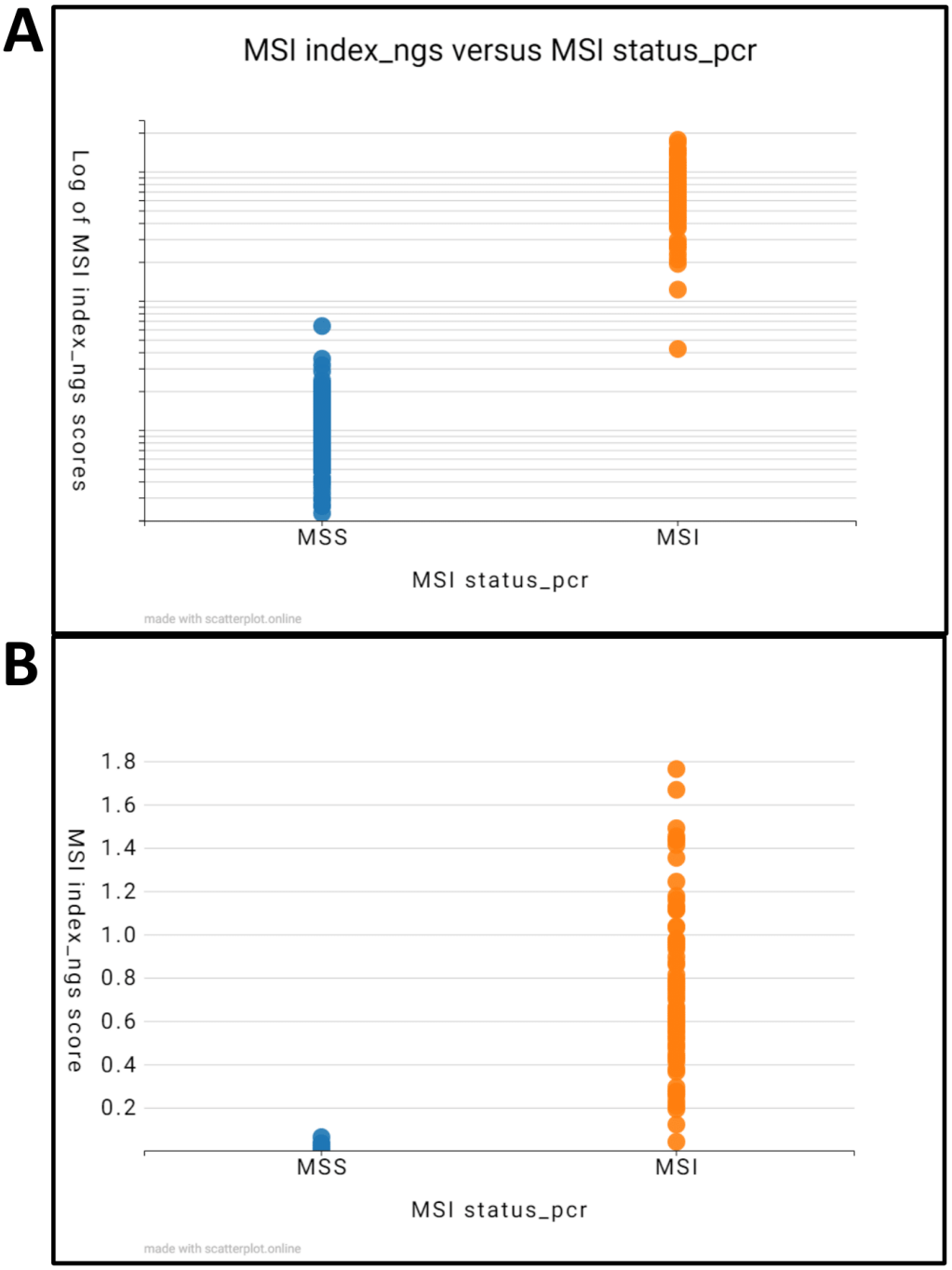


Figure 6.3: Scatter plots showing the relationship between MSI index_ngs and the PCR-based MSI status **A.** Log of MSI index_ngs scores clearly distinguishes MSI (n=75/383) from MSS (n=308/383) cases **B.** MSI index_ngs scores of MSS cases are tightly clustered together, while those of MSI cases are dispersed.

6.3.4 Validation of MSI indices

6.3.4.1 MMR gene expression

Gene expression data was available for only 379/441 cases in the GDC repository. The gene expression levels of *MLH1, MSH2, MSH3, MSH6, PMS1,* and *PMS2* are shown in Appendix Table 6.4. Only *MLH1* expression showed any significant correlation with all MSI measures. Specifically, *MLH1* expression showed an inverse correlation with MSI index_ngs in keeping with the established relationship between *MLH1* expression and loss of mismatch repair. Furthermore, there was a comparable level of *MLH1* expression correlation between the PCR-based MSI status and the NGS-based MSI indices which were derived from simple counts of mutated microsatellites (Table 6.1).

Table 6.1: DNA mismatch repair gene expression and MSI indices

MMR gene expression	MSI index_ngs	MSI status_ngs	MSI status_pcr	MTM index
MLH1	**R= -0.527** **P<0.001** N=379	**R=0.578** N=379	**R=0.591** N=379	**R=-0.519** **P<0.001** N=379
MSH2	R=0.078 P=0.132 N=379	R=0.093 N=379	R=0.087 N=379	R=0.085 P=0.097 N=379
MSH3	R=0.010 P=0.847 N=379	R=0.026 N=379	R=0.021 N=379	R=0.017 P=0.744 N=379
MSH6	R=0.044 P=0.397 N=379	R=0.059 N=379	R=0.049 N=379	R=0.048 P=0.354 N=379
PMS1	R=0.049 P=0.343 N=379	R=0.109 N=379	R=0.058 N=379	R=0.044 P=0.389 N=379
PMS2	R=0.062 P=0.227 N=379	R=0.064 N=379	R=0.078 N=379	R=0.083 P=0.107 N=379

*Significant relations shown in bold fonts

6.3.4.2 The MTM index

The MTM index per case was derived by summing up the number of somatic mutations in 16 of the most common genes that are mutated in an MSI background in GCa, then expressing the sum total as a fraction of 16. The results showed that only 195/441 cases have at least one mutation in the molecular targets of MSI (Appendix Table 6.3). Eta test showed a good correlation between the MTM index and the MSI status_pcr at a level comparable to the correlation between MTM index and MSI status_ngs (Table 6.2).

Table 6.2: Relationship between MTM index and the MSI indices

	MSI index_ngs	MSI status_ngs	MSI status_pcr
MTM index	**R=0.940** **P<0.001** N=441	**R=0.834** N=441	**R=0.831** N=383
MSI index_ngs		**R=0.839** N=441	**R=0.869** N=383
MSI status_ngs			**P<0.001** N=383

6.3.4.3 *MLH1* promoter methylation status

Four *MLH1* DNA methylation probes showed consistent normo-methylation patterns across the 28 normal DNA samples: cg00893636, cg02279071, cg13846866 and cg18320188 in 27/28, 27/28, 27/28 and 24/28 normal samples, respectively. All 441 GCa cases had DNA methylation data for the selected probes. The beta values of these probes were retrieved for the tumour samples, and on the basis of these values the samples were categorised into *MLH1*-normomethylated and –hypermethylated. Pearson's analyses showed good beta value correlation among the *MLH1* DNA probes and validated the relationship between *MLH1* promoter methylation and *MLH1* expression in this GCa cohort (Tables 6.3 and 6.4). Chi square test

showed a strong association between *MLH1* methylation state and MSI statuses (MSI status_pcr and MSI status_ngs) (Table 6.4), evidence that a simple count of mutated microsatellites in level 2 data is a sufficient measure of MSI.

Table 6.3: Relationship among the *MLH1* methylation probes

	cg0227907 methylation status	cg13846866 methylation status	cg18320188 methylation status	cg02279071 beta value	cg13846866 beta value	cg18320188 beta value
cg00893636 methylation status	R=0.801 P<0.001 N=441	R=0.836 P<0.001 N=441	R=0.628 P<0.001 N=441			
cg02279071 methylation status		R=0.728 P<0.001 N=441	R=0.770 P<0.001 N=441			
cg13846866 methylation status			R=0.561 P<0.001 N=441			
cg00893636 beta value				R=0.869 P<0.001 N=441	R=0.923 P<0.001 N=441	R=0.691 P<0.001 N=441
cg02279071 beta value					R=0.807 P<0.001 N=441	R=0.838 P<0.001 N=441
cg13846866 beta value						R=0.638 P<0.001 N=441

6.3.5 Performance tests: MSI status_pcr versus MSI status_ngs

Cohen's Kappa test was used to measure the level of agreement between MSI status_pcr and MSI status_ngs at determining MSI in the GCa cohort. The result showed a near perfect concordance (Kappa =0.976) between the PCR- and NGS-based methods

The performance of the MSI status_ngs relative to the PCR gold standard was calculated using the online MEDCALC statistical software as follows: Sensitivity = 100% (95% confidence

interval, CI = 95.2% to 100%), Specificity = 99.03% (95% CI = 97.18% to 99.80%), Positive Predictive Value, PPV = 96.15% (95% CI = 89.02% to 98.72%), Negative Predictive Value, NPV = 100%, Accuracy =99.22% (95% CI = 97.73% to 99.84%).

Table 6.4: Correlations of *MLH1* methylation

	MLH1 expression	MTM index	MSI index_ngs	MSI status_ngs	MSI status_pcr
cg00893636 methylation status	R=0.510 N=379	R=0.542 N=441	R=0.574 N=441	R=0.654 P<0.001 N=441	R=0.688 P<0.001 N=383
cg02279071 methylation status	R=0.431 N=379	R=0.511 N=441	R=0.501 N=441	R=0.580 P<0.001 N=441	R=0.601 P<0.001 N=383
cg13846866 methylation status	R=0.434 N=379	R=0.536 N=441	R=0.558 N=441	R=0.626 P<0.001 N=441	R=0.678 P<0.001 N=383
cg18320188 methylation status	R=0.247 N=379	R=0.360 N=441	R=0.351 N=441	R=0.403 P<0.001 N=441	R=424 P<0.001 N=383
cg00893636 beta value	R=-0.524 P<0.001 N=379	R=0.563 P<0.001 N=441	R=0.602 P<0.001 N=441	R=0.669 N=441	R=0.702 N=383
cg02279071 beta value	R=-0.440 P<0.001 N=379	R=0.527 P<0.001 N=441	R=0.534 P<0.001 N=441	R=0.590 N=441	R=0.613 N=383
cg13846866 beta value	R=-0.517 P<0.001 N=379	R=0.563 P<0.001 N=441	R=0.607 P<0.001 N=441	R=0.674 N=441	R=0.717 N=383
cg18320188 beta value	R=-0.291 P<0.001 N=379	R=0.393 P<0.001 N=441	R=0.397 P<0.001 N=441	R=0.437 N=441	R=0.415 N=383

6.4 Discussion

The high concordance and significant correlations between the NGS-based MSI indices on the one hand, and the combination of MSI status_pcr, MTM index, *MLH1* expression and *MLH1* methylation on the other, validates the MSI-WES approach to MSI testing. A simple count of microsatellite-based indels per tumour sample is adequate and sensitive for MSI testing using WES.

The use of NGS for MSI detection in tumours has previously been reported by other groups. For example, Niu et al developed the MSISensor software tool which utilises a BAM file input for the detection of MSI from exome sequencing data of tumour in comparison to normal tissue. The MSISensor first retrieves homopolymers and microsatellites of 5bp and above from a reference genome and stores these in a reusable file for subsequent analysis. A k-mer distribution of microsatellite lengths at each of the reference loci for tumour and paired normal is obtained by tallying microsatellite repeat length to a total of 20 or more reads. Then, chi square test with correction for multiple testing is utilised to identify the per locus differences between tumour and paired normal. Any locus with significant difference in distribution between tumour and normal is regarded as 'somatic' or mutated. Next, the total number of somatic sites is counted for each tumour sample and scored as a percentage of the total number of microsatellite sites examined for the pair (Niu et al., 2014).

Kautto et al developed the MANTIS tool for determination of MSI in tumour. It also uses a tumour-normal sample comparison approach with stringent quality control of input sequencing data in BAM file format. The MANTIS tool retrieves microsatellite loci from reference genome and stores them in a BED file format. Microsatellites repeat motifs (repeat lengths) are identified separately for tumour and paired normal at each microsatellite locus, and the read count for each

repeat motif is determined and normalised to the total reads for that locus. Then, the stepwise differences (distance) between the tumour distribution of normalised repeat motif counts and those of paired normal are calculated per locus. This value represents the per-locus instability. Finally, the average value for all the examined loci is then taken to represent the instability value of the tumour sample (Kautto et al., 2017).

Furthermore, Salipante et al described the mSINGS software tool developed to analyse MSI in tumour samples. The tool accepts variant call files purposedly processed from BAM files for MSI detection. mSINGS first identified unstable microsatellite loci from 3 different datasets of known MSI-H samples. The mean number of repeat lengths (alleles) at each of the identified loci was then calculated in known MSI-negative cases (or microsatellite stable, MSS) and peripheral blood samples and used as the baseline reference values. Next, the total number of alleles exceeding 5% read frequency was tallied per locus. Any locus with value exceeding the mean plus 3 standard deviation of the baseline value was regarded as unstable for that tumour sample. Then, the total number of unstable loci was counted per sample and expressed as a fraction of the total number of loci analysed. Any sample with a fraction of 0.2 (or 20%) unstable loci was regarded as microsatellite unstable (Salipante et al., 2014).

Moreover, Pabla et al described an approach which also utilises BAM file input and calculates the Euclidean distance between a test sample and the centrioid of an original training k-mean cluster by using the number of indels and the average length of the indels per microsatellite loci for 29 highly prevalent microsatellite loci (a total 58 features) which were identified from principal component analysis of 522 loci. The Euclidean distance was then used to assign a microsatellite status (MSI or MSS) to the test sample. This approach does not require comparison with matched normal samples (Pabla et al., 2019).

In addition, Trabucco et al used the principal component analysis of the average microsatellite length and variance of lengths for each microsatellite loci to generate an MSI score for the stratification of MSI and MSS cases. Their approach does not require matched normal tissue for comparison, and either targeted, whole exome or whole genome sequencing data can be used for this MSI detection approach (Trabucco et al., 2019).

Lu et al calculated the proportions of microsatellite-based indels relative to total indels (i.e., PI, proportion of indels; and PD, proportion of deletions) using Tandem Repeat Finder and DINDEL, respectively, for calling microsatellites and total indels from aligned RNASeq data. Then the PI:PD ratio was used as the MSI score (Lu et al., 2013).

Ni Huang et al, on the other hand, developed the MSIseq-based NGSclassifier, which directly counted microsatellite-based indels from maf files and expressed the count as a fraction or proportion of the total length of the genome captured for hybridisation and sequenced (Ni Huang et al., 2015). The difference between the MSIseq and MSI-WES is in the input file format used by either method. The input files also differentiate the strategy for counting the microsatellite-based indels. But more importantly, the vcf file contains more variants and therefore has more genomic coverage than the maf file (Zhang et al., 2021). The use of the vcf file input would therefore give a better classification of MSI status than that of the maf file, at least theoretically.

All the NGS-based MSI detection tools and approaches described above showed high performances in comparison to the PCR-based gold standards.

MSI-WES uses partially processed (level 2) data from vcf files. As such MSI analysis is performed on the same vcf files from which other clinically relevant variants are extracted. This

simplifies the incorporation of MSI analysis into existing pipelines of mutational analyses. In addition, this MSI detection approach does not require analysis of matched normal data.

In the demonstration of this NGS-based MSI detection approach two pertinent issues have arisen. The first issue pertains to the cut-off value or threshold for calling MSI. Inasmuch as a gastric cancer cohort has been used to demonstrate the validity and accuracy of this approach, and an MSI threshold has been established for the NGS-based analysis using a combination of k-means clustering analysis and Cohen's Kappa test for concordance – and it is envisaged that the threshold would not change significantly for a larger cohort of GCa cases during the process of standardisation – it remains to be seen if this one threshold would be directly applicable to other cancer types. Different MSI-calling thresholds may need to be established for different cancer types. The second issue – which is related to the first – concerns which variant caller to use for the standardisation process for this NGS-based approach. Whilst of the four TCGA variant callers only the VarScan2 and MuTect2 record indels, the MuTect2 records a higher number of variants than the VarScan2. This has implications for standardisation of MSI thresholds: either only one type of variant caller should be standardised for MSI analysis or different thresholds should be established for the different variant callers.

The principles of the MSI-WES assay are easily adaptable to whole genome sequencing (WGS) data for MSI detection because WGS data, similar to WES data, are recorded in vcf files. As previously explained, for MSI detection using the MSI-WES mutated microsatellite loci are extracted from vcf files. These vcf files also contain several germline and somatic variants such as *BRAF*, *KRAS*, etc. Hence MSI-WES (or MSI-WGS) can be adapted into the same workflow that can also detect somatic and germline mutations in the MMR and other genes and used for Lynch syndrome screening. Furthermore, although gene methylation studies require a separate

type of wet lab experiments (Y. Liu et al., 2018), methylation files generated from those experiments can be included in the same workflow for end-to-end Lynch syndrome screening.

In conclusion, a simple count of the mutated microsatellites in adequately quality-controlled somatic mutation data generated by WES is a valid and accurate approach for MSI testing, at least in gastric cancer. A freely accessible MSI_locator algorithm is currently being developed using the described NGS-based MSI detection approach (https://github.com/marcos-diazg/MSI_locator). [1]

[1] This chapter of the thesis has been published in its entirety in Future Oncology, 2021 Jul 22. doi: 10.2217/fon-2021-0132. Online ahead of print.

CHAPTER 7

GENERAL DISCUSSION

7.1 Introduction

The overall aim of this thesis is to develop simple molecular diagnostic assays with improved performance indices for mutation detection. I have accomplished this aim by designing and developing four molecular pathology tests: the Squirrel PCR assay for enabling the quality and accurate sequencing of ultrashort DNA segments; the mHRMA assay, an HRMA-based multiplex assay for accurate and fast detection of mutations; the MSI-HRMA, an HRMA-based test for MSI; and the MSI-WES, a whole exome sequencing strategy for measuring MSI in GCa using level 2 somatic point mutation data recorded in vcf formats. The description of these four assays (Squirrel PCR, mHRMA, MSI-HRMA and MSI-WES assays) have been published in peer-reviewed journals (Ebili et al., 2021; Ebili, Hassall, Asiri, et al., 2017; Ebili, Hassall, Fadhil, et al., 2017; Susanti et al., 2018).

The development of these assays was accomplished by one or a combination of two or more of the following strategies (i) altering the components of the assays, (ii) altering the detection method, (iii) altering the assay platform. The purpose of altering the component of an existing assay, or of using an entirely new test is to achieve improvements in one or more aspects of the test performance indices.

7.2 The Squirrel PCR assay

The Squirrel PCR assay, which was shown to enable good-quality, accurate and direct sequencing of ultrashort DNA segments by Sanger method, circumvents the need for using the expensive, cumbersome, and failure-prone NGS or sequencing-by-cloning methods for sequencing such short DNA segment. The Squirrel PCR assay is an example of how altering the components of an existing assay can improve accuracy, and ease and speed of operation of a

detection method. Standard Sanger sequencing is accomplished by amplifying the DNA segment of interest by PCR and then using the PCR primers as sequencing primers in the Sanger method. However, this strategy is not suitable for ultrashort DNA segments as it leads to unreliable and inaccurate sequencing reads. Therefore, by elongating the DNA segment of interest using the Squirrel PCR assay, the poor quality and inaccurate reads which would otherwise have resulted from using standard PCR are eliminated. The Squirrel primer pairs add non-specific sequences at the 5' ends of the resultant amplicons, thereby providing matrices for a separate sequencing primer pair (USeq primers) which prime the sequencing from a site much farther away from the genomic DNA of interest and facilitates the direct sequencing of short DNA segments. The alterations of assay components effected in the Squirrel PCR assay is the use of the non-conventional Squirrel primer pairs for PCR and the addition of the USeq primer pairs as sequencing primers. As the Squirrel PCR assay enables direct sequencing of short DNA segments, both the cost-effectiveness and turnround time for sequencing are improved.

The use of Squirrel primers (with 40-65 base-long, non-specific tail sequence) for PCR is possible because the minimum useful primer length (number of consecutive 3' primer bases that must anneal to the template sequence for PCR to proceed) was found to be nine (9) bases (Williams et al., 1990). The useful primer lengths for the Squirrel primers were 18-26 bases. In my experience, the long non-specific tails did not adversely affect the annealing temperatures of the Squirrel primers. On the contrary, they appeared to enhance primer annealing over a wide range of annealing temperatures, probably due to their high Tm, and thereby making the primers much more efficient than their non-tailed counterparts (Arif et al., 2014; Kalendar et al., 2011; Timmons et al., 2013; Vandenbussche et al., 2016). The Squirrel PCR assay was successfully applied in the sequencing of the C-region of the *MLH1* promoter by our research group (Susanti

et al., 2018).

The Squirrel PCR assay uses the Squirrel primer technique to improve the sequencing of ultrashort DNA segments by Sanger method. Therefore, any concerns – valid as they truly are – about the level of sensitivity of the Squirrel PCR-Sanger sequencing combination is a commentary on the sensitivity of Sanger sequencing per se, and not on the Squirrel PCR assay. The sensitivity of the Squirrel PCR assay is limited by the LOD of Sanger method. The LOD of Sanger sequencing is generally considered to be about 10-20% (Hageman, 2015; Ishige et al, 2018) but improvements to 5% or less has been achieved with the introduction of the new Minor Variant Finder by the Applied Biosystems (Schreiber et al, 2016; Kikuchi et al, 2018; Applied Biosystems, 2016; Subbotina et al, 2020). Furthermore, our group had developed a modified CO-amplification at Lower Denaturation temperature PCR (COLD PCR) cycling strategy that enables enrichment of minor variants in clinical tumour samples (Ham-Karim et al, 2017). With further validation, this cycling strategy could have been applied to the Squirrel PCR assay to enrich the minor allele for easy detection by Sanger method as did previous study (Li et al, 2009). Despite this sensitivity limitations of Sanger sequencing, it is listed in the NICE guidelines as one of the methods of mutation detection in cancer genetics, and many molecular diagnostic laboratories in the UK currently use Sanger sequencing for mutation detection (National Institute for Health and Care Excellence, 2013b; Richman et al, 2017; Richman et al, 2018).

7.3 HRMA-based assays

Although HRMA detection platform was developed as a screening or “mutation scanning” test for point mutations, the results of which screening require confirmation by Sanger sequencing

(Fitarelli-Kiehl et al., 2016), HRMA can be deployed as a diagnostic test. For example, the MSI-HRMA assay which uses a panel of microsatellite markers can be a screening test for Lynch syndrome, but a diagnostic test for immunotherapy response. Furthermore, an HRMA-based test can be used as a diagnostic test for point mutations under certain conditions. The *BRAF* V600 mutations are markers of response to BRAF inhibitors. Any mutation at V600 is considered a positive predictive test. Therefore, an altered HRM pattern produced by a well-designed HRMA-based test is virtually diagnostic as it is unnecessary, under this circumstance, to differentiate V600E from V600K or V600R/M/D/G (National Comprehensive Cancer Network, 2021h; Patel et al., 2020). Moreover, in cases where the definitive genotype of the mutation is required from the test, the addition of an appropriate positive control for that genotype can be included in the experiments. For example, the *KRAS* G12C mutation is now currently targeted for therapy in lung cancer management using adagrasib and sotorasib (Hong et al., 2020; Jänne et al., 2020; Skoulidis et al., 2021). Therefore, to diagnose a *KRAS* G12C mutation using an HRMA-based test, a *KRAS* G12C-mutated positive control can be included in HRMA run to define a *KRAS* G12C genotype among the test samples. The exquisite sensitivity of HRMA makes it suitable to detect - by differential melting patterns - multiple genotypes which differ from one another by only single nucleotides (Palais et al., 2005).

The detection limit for HRMA is 1-10%, depending on the study (Wojdacz et al, 2007; Ham-Karim et al, 2017; Fadhil et al, 2010; Ibrahem et al, 2010; Ishige et al, 2018). In our research group HRMA has achieved a sensitivity (LOD) of 1.5-3% (Ibrahem et al, 2010; Ham-Karim et al, 2017; Fadhil et al, 2010). In my experience and in those of others the sensitivity of any HRMA assay depends on the size of the PCR amplicons or products (Ishige et al, 2018). The shorter the amplicons the lower the LOD (or the higher the sensitivity). With only two

exceptions, the amplicons for the mHRMA and the MSI-HRMA were designed to be less than 100bp. Although LOD for the primer pairs could not be determined for the assays due to time constraints, the limits of detection for the mHRMA and MSI-HRMA may be between 1-3%. This estimation is based on the LOD of primer pairs which amplified similar amplicon sizes in other studies from our research group (Ham-Karim et al, 2017; Fadhil et al, 2010; Ibrahem et al, 2010).

7.3.1 mHRMA assay

The successful development of the mHRMA assay demonstrated that the multiplexing of targets on the HRMA platform is not only feasible but it is an accurate, sensitive, specific and fast strategy for mutation detection. In the development of the mHRMA I showed that the speed of mutation testing on the HRMA platform can be further improved significantly by multiplexing of the targets. As HRMA depend on temperature melting profile, the adequate separation of targets along the melting temperature axis ensured robust HRM analysis and accurate calling of mutations. This separation of targets was achieved by manipulating the melting temperatures of the targets using GC-tagged primers. With respect to conventional HRMA, the assay components that was altered to improve test performance include the increase of the number of targets that can be amplified and analysed simultaneously, as well as the use of GC-tagged primers to properly separate the targets.

It is therefore possible to predict the size of the GC tags required to adequately separate one target from the next along the temperature axis by using online Tm prediction tools such as UMELT (Dwight et al., 2011). It is also possible to determine how many targets can fit into the range of the temperature axis with this tool. I have devised a protocol for designing mHRMA that can accommodate up to four (4) targets *in tube*. This protocol entails the following (i)

designing all amplicons or targets to have a size of 30-40bp and roughly equal Tm's *in silico*, (ii) then by use of varying numbers of G- and C-tags, increasing the Tm's of the individual targets *in silico* to achieve adequate separation of the Tm's of the targets. Adequate separation of the Tm's for HRM analysis should be defined as a difference of 1^0C along the temperature axis between the descending slope of a preceding curve and the ascending slope of the succeeding curve.

Understandably, with the multiplex design of 3 or more targets the individual target primers will, more likely than not, have varying annealing temperatures. The use of a theoretical annealing temperature for the multiplex reaction will most likely lead to unequal amplification of the targets in the multiplex, or failure of one or more targets to amplify (Carlson et al., 2013; Markoulatos et al., 2002). However, the use of gradient PCR-HRMA can help to define an empirical annealing temperature for the multiplex. The advantage of using HRMA for gradient PCR, rather than the conventional gel-based detection, is that the experimenter can define the optimal annealing temperature that ensures equal amplification of all the targets and at the same time be able to see if the targets' Tm's are sufficiently separated.

The ability of GC-tagged primers to alter the Tm's of primers has been exploited by our research group to design and develop the highly optimised annealing temperature amplification refractory mutation system (HOT_ARMS) PCR, a modification of the ARMS PCR (or allele-specific PCR). In this assay, the forward and reverse primers, one of which is allele- or mutation-specific (i.e., can anneal to the template when a mutation is present) are tagged with GC bases to raise the primer Tm above 70^0C so that a high annealing temperature of about 65^0C can be applied to the PCR to achieve specificity. When a mutation is present at the hotspot, the primers anneal efficiently even at high temperature due to their high Tm's and amplify the mutated template. However, in the absence of a mutation there is poor annealing of the allele-specific primer to the

hotspot at the primer 3' end, which poor annealing is worsened at higher temperature, leading to failure to amplify the wild-type template. The HOT_ARMS PCR assay was found to be exquisitely sensitive and specific as it overcomes non-specific priming, a common cause of non-specific amplification in conventional ARMS PCR (Imyanitov et al., 2002; J. Liu et al., 2012; Morris, 2008). The HOT_ARMS PCR assay is currently undergoing the process of patenting (https://www.freepatentsonline.com/20210040550.pdf).

7.3.2 MSI-HRMA assay

Analysis of the microsatellite loci of GCa WES data showed that individual MSI-H cases can have more than two alleles representing a combination of deletions, insertions and different insertion and deletion alleles in any one microsatellite locus. This multiplicity of alleles can lead to formation of multiple heteroduplexes in individual microsatellite marker loci and improve the sensitivity of HRMA-based detection of MSI (Garritano et al., 2009; Palais et al., 2005). The MSI-HRMA is an example of an assay in which the detection method has been altered to ease operation of the test, improve the speed of testing, achieve reduction in the cost of running the test, and improve analytical performance. The gold standard for MSI testing is fluorescent PCR followed by fragment analysis by capillary electrophoresis. The switch from CE to HRMA detection reduces the cost of testing, and the laboriousness associated with CE. Another alteration, which is of necessity to using the HRMA detection method for MSI testing is the alteration of primer design from fluorescent-tagged primers to conventional primer design, but with addition of HRMA-suitable fluorescent dye to the PCR mix. This primer design alteration contributes to reduction in running cost of MSI test.

Further alterations that could be made to the MSI test methodology is the multiplexing of the microsatellite markers. In chapter 4 I demonstrated with the mHRMA assay the feasibility of

duplexing and triplexing of targets on the HRMA detection platform. Multiplexing of the panel of MSI markers would lead to a further improvement in testing speed and reduction in cost and labour (Markoulatos et al., 2002).

MSI-based assays are amenable to automation to enable deployment in a point-of-care (POC) setting. All aspects of the mHRMA and MSI-HRMA assays from DNA extraction to result output, can be automated to enable focused and speedy delivery of clinical mutation testing services, similar to the automation of the ARMS (or mutation-specific or allele-specific) PCR on the Biocartis Idylla platform (Johnston et al, 2018). The Biocartis Idylla is automated ARMS-PCR assay designed for POC testing of point mutations in clinically relevant targets. Whilst mHRMA is more cost-effective than "manual" ARMS-PCR – based on the cost-savings resulting from multiplexing of targets – the comparison of mHRMA with the automated Biocartis Idylla is false equivalency. However, both mHRMA and MSI-HRMA can be automated to the same extent that ARMS-PCR has been automated in the Biocartis Idylla, at least hypothetically: FFPE tissue sections are placed directly into the Idylla system cartridge, with no pre-processing. The cartridge is then inserted into the Idylla platform, where high-intensity focused ultrasound (HiFu) technology, buffers, reagents, and heat induces deparaffinisation, tissue disruption, cell lysis and subsequent release of nucleic acid. *Allele-specific primers and fluorescent probes then perform real-time PCR amplification and mutation detection*. All required consumables are provided in the cartridge (Johnston et al, 2018). By replacing allele-specific primers and fluorescent probes with HRMA-specific primers and saturating fluorescent dyes in the cartridge, as well as by installing the HRMA software on the machine, HRMA can be automated for multiplexing of mutation targets and for microsatellite testing in a single run.

Similarly, the miniaturisation of Sanger sequencing platforms can enable automation of the Squirrel primer assay strategy for POC testing (Furutani et al, 2022). For a clinical laboratory that already possesses a Sanger platform that performs focused and limited DNA sequencing for cancer mutation detection, automation of the steps - from DNA extraction to Squirrel PCR - to enable speedy production PCR templates for sequencing on the available Sanger machine can be undertaken. To reiterate, the Squirrel PCR strategy simply enables Sanger method to sequence ultrashort DNA segments as found, for example, in ctDNA. According to the 2017 and 2018 UK EQA report, Sanger sequencing is still a common platform for cancer mutation detection in the UK. Although the dominant method for sequencing is now NGS (National Institute for Health and Care Excellence, 2013b; Richman et al, 2017; Richman et al, 2018).

7.4 MSI-WES

Next-generation sequencing is already playing critical roles in clinical practice and research. In my opinion, its role in clinical research supersedes its current role in actual clinical practice, and this is due to the capacity of NGS to generate huge amount of mutation data most of which, however, has no current clinical use. Currently, the most important factor militating against the widespread and full deployment of NGS in clinical practice is cost, coupled with the fact that the relatively few clinically relevant mutations can be detected by more cost-effective methods. However, the cost of deployment of NGS is gradually falling, and in the near future will fall to a level comfortable enough for easy acquisition (UK Department of Health & Social Care, 2020). This is especially as the miniaturisation of equipment and devices, as well as emergence of more competition into the NGS market are on the rise (Huo et al, 2021; UK Department of Health & Social Care, 2020; Mayday et al, 2019; Suckling et al, 2019; Kulski, 2016, Liu et al, 2012).

Furthermore, the number of clinically actionable molecular biomarkers is expected to rise to match the number of targeted therapies that are being discovered and developed daily (Falzone et al, 2018; Bandino, 2015; Zhong et al, 2021; Hsu and Lu, 2016). Together these factors will cause the rate of uptake of NGS for routine clinical mutation detection to surpass those of current methods. In preparation for these near-future events I have elected to add a chapter on the use of NGS for MSI detection to my thesis.

The MSI-WES method exemplifies the case in which an entirely different platform than the gold standard is utilised for molecular testing. In contrast to the conventional PCR-based capillary electrophoresis method used for MSI testing the MSI-WES uses level 2 whole exome sequencing data generated with a sequencing-by-synthesis NGS technology from genomic DNA templates of tumour samples. The differences between the MSI-WES and other NGS-based MSI-testing methods have been discussed in chapter 6. But to reiterate here, MSI-WES uses annotated somatic mutation data recorded in vcf file formats for counting altered microsatellites. MSI-WES exploits the format in which microsatellite-based indels are recorded in vcf files to *retrieve* the indels. This strategy of *retrieving* the altered microsatellites is one of the differentiators of MSI-WES from the method described by Ni Huang et al. (Ni Huang et al., 2015). Ni Huang et al. also counted altered microsatellites (or simple repeat sequences) by matching annotated indels ('INS' and 'DEL') in maf files to the DNA sequence context of the indels, and then determining the *S.ind*, which is the number of indels in the simple repeat sequences per Mb of the genome sequenced, for each sample. As can be seen from the discussion above, while the NGS-classifier uses maf file input the MSI-WES uses vcf file.

The molecular targets of MSI differ among tumours, hence the MSI panel developed for one tumour type in non-NGS platforms may be inadequate for another tumour type. Furthermore, the

frequencies of mutations of microsatellite markers in MSI differ from tumour to tumour. The MSI-Mono-Dinucleotide Assay panel which was used for MSI testing in the TCGA cohort contains 4 mononucleotide markers BAT25, BAT26, BAT40 and TGFRBII. An analysis of these markers on the SelTar database show that they have sensitivities of 89.9%, 92.3%, 78.7% and 64.5%, respectively, for the detection of MSI in gastric cancer. In comparison the sensitivities of same markers at detecting MSI in CRC and endometrial cancer are 90.1%, 92.5%, 90.1%, 73.8% and 79.8%, 79.9%, 82.5%, 14.5%, respectively (http://www.seltarbase.org/). On the other hand, MMR status analysis by IHC have been reported to present some limitations, especially with gastric cancer. Between 5-11% of gastric cancer with MSI do not show loss of MMR protein because of the phenomenon of non-functional protein with retained antigenicity (Ratti et al, 2018, Murphy et al, 2003). Missense mutations in *MLH1* and *MSH6* cause translation of proteins with reduced enzymatic activities. The use of IHC alone in these cases can lead to false negative dMMR results (Ratti et al, 2018, Murphy et al, 2003). Hence NGS should play a role, not only in the direct diagnosis of MSI but in the research and development of tumour-specific MSI marker panels for use in non-NGS MSI detection platforms.

7.5 The extent of validation achieved for the assays

Limited validation was performed for all the assays developed in this PhD study. Aspects of both analytical and clinical validations were attempted. The indices of analytical validation performed for the assays included analytical specificity (all experiments were run with no-template controls), analytical accuracy (positive and negative controls were included in all the experiments), repeatability or short-term precision (all samples were run in replicates in all the experiments) and intermediate precision, or between-run precision (the same positive, negative

and no-template controls were included in successive experiments in all cases), and robustness of the assay against DNA template types as shown in chapters 2, 3 and 4. However, analytical sensitivity, or limit of detection, could not be performed for any of the assays due to time constraints. Furthermore, it was practically impossible to perform analytical validation for the MSI-WES method (Association for Molecular Pathology Clinical Practice Committee, 2009, 2014; Association of Molecular Pathology, 1999; Mattocks et al., 2010; Prence, 1999).

The indices of clinical validation tested for the developed assays include sensitivity and specificity, accuracy, concordance, NPV and PPV. The newly designed assays were compared with the already existing gold standard tests using clinical samples of CRC and GCa (Association for Molecular Pathology Clinical Practice Committee, 2009, 2014; Association of Molecular Pathology, 1999; Mattocks et al., 2010; Prence, 1999). However, it was not practical within the constraints of the PhD study to perform inter-laboratory precision (reproducibility) for the assays.

Although the Squirrel PCR assay was developed with only 4 CRC cell lines and 13 pre-amplified or PCR-enriched clinical CRC samples with known mutation patterns – a total of 34 amplicons were tested – the assay has been applied to the sequencing of the *MLH1* promoter region C in a further study using genomic DNA from 20 clinical CRC samples (Susanti et al., 2018), bringing the total amplicons tested so far to 54. The preliminary data obtained with the development of the Squirrel PCR assay prove the concept that Sanger sequencing can be used to sequence sub-100bp DNA fragments if such fragments are elongated by Squirrel PCR. The data also show that the assay *can be* robust against different nucleic acid template types and *can have* high performance indices. However, the total number of variants tested with the Squirrel PCR assay thus far is limited and therefore further validation by upgrade of the number of test cohort is

required. This limitation is evident from the performance indices of accuracy, clinical sensitivity, and clinical specificity values calculated from the experimental data. The wide range of 95% CI values obtained for these indices [sensitivity of 100% (95% C.I. = 73.54% to 100.00%); specificity of 100% (95 % C.I. = 80.49% to 100.00%); accuracy of 100% (95% C.I. = 88.06% to 100.00%)] show that the adequate number of known mutations and wild-type cases were not included in the cohort tested. Furthermore, from the point of view of clinical application of this assay, many of the clinically actionable mutations from among *KRAS* codons 12/13/59/61/117/146, *NRAS* codon 12/13/59/61, *BRAF* codon 600, etc. were not tested by the Squirrel PCR assay. The Squirrel PCR assay is therefore still a work in progress.

The mHRMA assay was developed using 5 CRC cell lines and 17 pre-amplified clinical CRC samples to a total of 85 targets in the clinical samples (Ebili et al., 2017a). The preliminary data obtained with the mHRMA assay proved that mutation targets can be multiplexed on the HRMA platform by using GC-tagged primers to achieve adequate separation of the curves on the melting temperature axis. Although the assay showed robustness against the cell line and pre-amplified DNA templates, the mHRMA has not been applied directly to clinical genomic DNA templates. Furthermore, the number of variants that were included in the development was grossly limited, hence the wide range of 95% CI values obtained for the sensitivity index [sensitivity =100.00% (95% CI = 54.07% to 100.00%), specificity = 100.00% (95% CI=95.44% to 100.00%), accuracy =100.00% (95% CI= 95.75% to 100.00%)]. In addition, the clinically actionable mutations such as *KRAS* codons 12/13/59/61/117/146, *NRAS* codon 12/13/59/61, *BRAF* codon 600, etc. were not included in the development of the assay. Therefore, further validation using increased sample size and especially by increasing the proportion of clinically relevant variants in the test cohorts is warranted but could not be performed due to time constraints.

Within the limits of the PhD study the MSI-HRMA assay was well-validated. Analytical validation was performed with 9 known MSI and MSS CRC cell lines. Clinical validation was performed with a total of 153 CRC cases with known MSI/MMR-D status - as determined by IHC and PCR-based MSI method - and 13 normal samples. In the development process we showed that the MSI-HRMA was robust against cell line and genomic DNA template types.

The MSI-WES method showed high performance indices with the PCR-based data obtained for the 441 GCa cases. It was also validated using *MLH1* expression and methylation data. The final stage of developing the MSI-WES method would have been to create a software that tests MSI using the MSI-WES strategy. However, this may not be possible within the limits of the PhD study due to ethical constraints. Furthermore, it would have been ideal to test the MSI-WES method in a CRC cohort, as part of a furtherance of the validation process. But this too was not attempted due to time constraints.

In conclusion this thesis describes the design, development, and limited validation of four mutation-detection assays with improved performances. Certain further validations are required for some of the assays before they can be ready for “prime time”.

REFERENCES

Abulí, A., Bessa, X., González, J. R., Ruiz-Ponte, C., Cáceres, A., Muñoz, J., Gonzalo, V., Balaguer, F., Fernández-Rozadilla, C., González, D., de Castro, L., Clofent, J., Bujanda, L., Cubiella, J., Reñé, J. M. A., Morillas, J. D., Lanas, A., Rigau, J., García, A. M. A., … Andreu, M. (2010). Susceptibility genetic variants associated with colorectal cancer risk correlate with cancer phenotype. Gastroenterology, 139(3), 788–796, 796.e1-6. https://doi.org/10.1053/j.gastro.2010.05.072

Agarwal, A., Ressler, D., & Snyder, G. (2015). The current and future state of companion diagnostics. Pharmacogenomics and Personalised Medicine, 8, 99–110. https://doi.org/10.2147/PGPM.S49493

Ajani, J. A., D'Amico, T. A., Almhanna, K., Bentrem, D. J., Chao, J., Das, P., Denlinger, C. S., Fanta, P., Farjah, F., Fuchs, C. S., Gerdes, H., Gibson, M., Glasgow, R. E., Hayman, J. A., Hochwald, S., Hofstetter, W. L., Ilson, D. H., Jaroszewski, D., Johung, K. L., … Sundar, H. (2016). Gastric Cancer, Version 3.2016, NCCN Clinical Practice Guidelines in Oncology. Journal of the National Comprehensive Cancer Network: JNCCN, 14(10), 1286–1312. https://doi.org/10.6004/jnccn.2016.0137

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., Fernandez Banet, J., Billis, K., García Girón, C., Hourlier, T., Howe, K., Kähäri, A., Kokocinski, F., Martin, F. J., Murphy, D. N., Nag, R., Ruffier, M., Schuster, M., Tang, Y. A., … Searle, S. M. J. (2016). The Ensembl gene annotation system. Database: The Journal of Biological Databases and Curation, 2016, baw093. https://doi.org/10.1093/database/baw093

Akinleye, A., & Rasool, Z. (2019). Immune checkpoint inhibitors of PD-L1 as cancer therapeutics. Journal of Hematology & Oncology, 12(1), 92. https://doi.org/10.1186/s13045-019-0779-5

Albain, K. S., Barlow, W. E., Shak, S., Hortobagyi, G. N., Livingston, R. B., Yeh, I.-T., Ravdin, P., Bugarini, R., Baehner, F. L., Davidson, N. E., Sledge, G. W., Winer, E. P., Hudis, C., Ingle, J. N., Perez, E. A., Pritchard, K. I., Shepherd, L., Gralow, J. R., Yoshizawa, C., … Hayes, D. F. (2010). Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. The Lancet Oncology, 11(1), 55–65. https://doi.org/https://doi.org/10.1016/S1470-2045(09)70314-6

Alcaide, M., Cheung, M., Hillman, J., Rassekh, S. R., Deyell, R. J., Batist, G., Karsan, A., Wyatt, A. W., Johnson, N., Scott, D. W., & Morin, R. D. (2020). Evaluating the quantity, quality and size distribution of cell-free DNA by multiplex droplet digital PCR. Scientific Reports, 10(1), 12564. https://doi.org/10.1038/s41598-020-69432-x

Alexander, J., Watanabe, T., Wu, T. T., Rashid, A., Li, S., & Hamilton, S. R. (2001). Histopathological identification of colon cancer with microsatellite instability. The American Journal of Pathology, 158(2), 527–535. https://doi.org/10.1016/S0002-9440(10)63994-6

American Institute of Cancer Research. (2018). Diet, nutrition, physical activity and colorectal cancer. Continuous Update Project, 42–44. http://www.aicr.org/continuous-update-project/reports/breast-cancer-report-2017.pdf

An, J. Y., Kim, H., Cheong, J.-H., Hyung, W. J., Kim, H., & Noh, S. H. (2012). Microsatellite instability in sporadic gastric cancer: its prognostic role and guidance for 5-FU based chemotherapy after R0 resection. International Journal of Cancer, 131(2), 505–511. https://doi.org/10.1002/ijc.26399

Andre, F., Mardis, E., Salm, M., Soria, J.-C., Siu, L. L., & Swanton, C. (2014). Prioritising targets for precision cancer medicine. Annals of Oncology, 25(12), 2295–2303. https://doi.org/10.1093/annonc/mdu478

Applied Biosystems. (2016). Low-level somatic variant detection in tumor FFPE samples by Sanger sequencing. Application Note: Minor Variant Software. https://tools.thermofisher.com/content/sfs/brochures/variant-detection-tumor-ffpe-samples-sanger-sequencing.pdf

Aranake-Chrisinger, J., Huettner, P. C., Hagemann, A. R., & Pfeifer, J. D. (2016). Use of short tandem repeat analysis in unusual presentations of trophoblastic tumors and their mimics. Human Pathology, 52, 92–100. https://doi.org/https://doi.org/10.1016/j.humpath.2016.01.005

Arif, M., Aguilar-Moreno, G. S., Wayadande, A., Fletcher, J., & Ochoa-Corona, F. M. (2014). Primer modification improves rapid and sensitive in vitro and field-deployable assays for detection of high plains virus variants. Applied and Environmental Microbiology, 80(1), 320–327. https://doi.org/10.1128/AEM.02340-13

Armelao, F., & de Pretis, G. (2014). Familial colorectal cancer: a review. World Journal of Gastroenterology, 20(28), 9292–9298. https://doi.org/10.3748/wjg.v20.i28.9292

Arnold, C. N., Goel, A., Blum, H. E., & Boland, C. R. (2005). Molecular pathogenesis of colorectal cancer: implications for molecular diagnosis. Cancer, 104(10), 2035–2047. https://doi.org/10.1002/cncr.21462

Asad, J., Jacobson, A. F., Estabrook, A., Smith, S. R., Boolbol, S. K., Feldman, S. M., Osborne, M. P., Boachie-Adjei, K., Twardzik, W., & Tartter, P. I. (2008). Does oncotype DX recurrence score affect the management of patients with early-stage breast cancer? American Journal of Surgery, 196(4), 527–529. https://doi.org/10.1016/j.amjsurg.2008.06.021

Association for Molecular Pathology Clinical Practice Committee. (2009). Molecular Diagnostic Assay Validation. In Association for Molecular Pathology (Issue October). www.amp.org

Association for Molecular Pathology Clinical Practice Committee. (2014). Update to the 2009 AMP Molecular Diagnostic Assay Validation White Paper.

Association of Molecular Pathology. (1999). Association of Molecular Pathology Statemment: Recommendations for In-House Development and Operation of Molecular Diagnostic Tests. In American Journal of Clinical Pathology (Vol. 111, Issue 4). https://doi.org/10.1093/ajcp/111.4.449

Awad, K., Dalby, M., Cree, I., Challoner, B., Ghosh, S., & Thurston, D. (2019). The Precision Medicine Approach to Cancer Therapy: Part 1 — Solid Tumours. Pharmaceutical Journal. https://doi.org/10.1211/pj.2019.20207119

Bacher, J. W., Flanagan, L. A., Smalley, R. L., Nassif, N. A., Burgart, L. J., Halberg, R. B., Megid, W. M. A., & Thibodeau, S. N. (2004). Development of a fluorescent multiplex assay for detection of MSI-High tumors. Disease Markers, 20(4–5), 237–250. https://doi.org/10.1155/2004/136734

Baer, C., Menon, R., Bastawrous, S., & Bastawrous, A. (2017). Emergency Presentations of Colorectal Cancer. The Surgical Clinics of North America, 97(3), 529–545. https://doi.org/10.1016/j.suc.2017.01.004

Barault, L., Charon-Barra, C., Jooste, V., de la Vega, M. F., Martin, L., Roignot, P., Rat, P., Bouvier, A.-M., Laurent-Puig, P., Faivre, J., Chapusot, C., & Piard, F. (2008). Hypermethylator Phenotype in Sporadic Colon Cancer: Study on a Population-Based Series of 582 Cases. Cancer Research, 68(20), 8541 LP – 8546. https://doi.org/10.1158/0008-5472.CAN-08-1171

Bass, A. J., Thorsson, V., Shmulevich, I., Reynolds, S. M., Miller, M., Bernard, B., Hinoue, T., Laird, P. W., Curtis, C., Shen, H., Weisenberger, D. J., Schultz, N., Shen, R., Weinhold, N., Kelsen, D. P., Bowlby, R., Chu, A., Kasaian, K., Mungall, A. J., … Center, U. N. C. L. C. C. (2014). Comprehensive molecular characterisation of gastric adenocarcinoma. Nature, 513(7517), 202–209. https://doi.org/10.1038/nature13480

Baudino, T. A. (2015). Targeted Cancer Therapy: The Next Generation of Cancer Treatment. Current Drug Discovery Technologies, 12(1), 3–20. https://doi.org/10.2174/1570163812666150602144310

Beghelli, S., de Manzoni, G., Barbi, S., Tomezzoli, A., Roviello, F., Di Gregorio, C., Vindigni, C., Bortesi, L., Parisi, A., Saragoni, L., Scarpa, A., & Moore, P. S. (2006). Microsatellite instability in gastric cancer is associated with better prognosis in only stage II cancers. Surgery, 139(3), 347–356. https://doi.org/10.1016/j.surg.2005.08.021

Belhassen-García, M., Velasco-Tirado, V., López-Bernus, A., Alonso-Sardón, M., Carpio-Pérez, A., Fuentes-Pardo, L., Pardo-Lledías, J., Alvela-Suárez, L., Romero-Alegría, Á., Iglesias-Gomez, A., & Sánchez, M. C. (2013). Fever of unknown origin as the first manifestation of colonic pathology. Clinical Medicine, 13(2), 141 LP – 145. https://doi.org/10.7861/clinmedicine.13-2-141

Bilal, M., Raza, S. E. A., Azam, A., Graham, S., Ilyas, M., Cree, I. A., Snead, D., Minhas, F., & Rajpoot, N. M. (2021). Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal

cancer from routine histology images: a retrospective study. The Lancet. Digital Health, 3(12), e763–e772. https://doi.org/10.1016/S2589-7500(21)00180-1

Binladen, J., Thomas, M., Gilbert, P., Campos, P. F., & Willerslev, E. (2007). 5′-Tailed sequencing primers improve sequencing quality of PCR products. BioTechniques, 42(2), 174–176. https://doi.org/10.2144/000112316

Boardman, L. A. (2002). Heritable colorectal cancer syndromes: recognition and preventive management. Gastroenterology Clinics of North America, 31(4), 1107–1131. https://doi.org/https://doi.org/10.1016/S0889-8553(02)00049-3

Bocker, T., Diermann, J., Friedl, W., Gebert, J., Holinski-Feder, E., Karner-Hanusch, J., von Knebel-Doeberitz, M., Koelble, K., Moeslein, G., Schackert, H.-K., Wirtz, H.-C., Fishel, R., & Rüschoff, J. (1997). Microsatellite Instability Analysis: A Multicenter Study for Reliability and Quality Control. Cancer Research, 57(21), 4739 LP – 4743. http://cancerres.aacrjournals.org/content/57/21/4739.abstract

Boland, C. R., & Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. Gastroenterology, 138(6), 2073-2087.e3. https://doi.org/10.1053/j.gastro.2009.12.064

Boland, C. R., Thibodeau, S. N., Hamilton, S. R., Sidransky, D., Eshleman, J. R., Burt, R. W., Meltzer, S. J., Rodriguez-Bigas, M. A., Fodde, R., Ranzani, G. N., & Srivastava, S. (1998). A National Cancer Institute Workshop on Microsatellite Instability for Cancer Detection and Familial Predisposition: Development of International Criteria for the Determination of Microsatellite Instability in Colorectal Cancer. Cancer Research, 58(22), 5248 LP – 5257. http://cancerres.aacrjournals.org/content/58/22/5248.abstract

Bonneville, R., Krook, M. A., Chen, H.-Z., Smith, A., Samorodnitsky, E., Wing, M. R., Reeser, J. W., & Roychowdhury, S. (2020). Detection of Microsatellite Instability Biomarkers via Next-Generation Sequencing. Methods in Molecular Biology (Clifton, N.J.), 2055, 119–132. https://doi.org/10.1007/978-1-4939-9773-2_5

Bonneville, R., Krook, M. A., Kautto, E. A., Miya, J., Wing, M. R., Chen, H.-Z., Reeser, J. W., Yu, L., & Roychowdhury, S. (2017). Landscape of Microsatellite Instability Across 39 Cancer Types. JCO Precision Oncology, 1, 1–15. https://doi.org/10.1200/PO.17.00073

Bovée, J. V. M. G., & Hogendoorn, P. C. W. (2010). Molecular pathology of sarcomas: concepts and clinical implications. Virchows Archiv : An International Journal of Pathology, 456(2), 193–199. https://doi.org/10.1007/s00428-009-0828-5

Brandão, M., Pondé, N., & Piccart-Gebhart, M. (2018). MammaprintTM: a comprehensive review. Future Oncology, 15(2), 207–224. https://doi.org/10.2217/fon-2018-0221

Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6), 394–424. https://doi.org/https://doi.org/10.3322/caac.21492

Brazowski, E., Rozen, P., Pel, S., Samuel, Z., Solar, I., & Rosner, G. (2012). Can a gastrointestinal pathologist identify microsatellite instability in colorectal cancer with reproducibility and a high degree of specificity? Familial Cancer, 11(2), 249–257. https://doi.org/10.1007/s10689-012-9508-8

Breslauer, K. J., Frank, R., Blöcker, H., & Marky, L. A. (1986). Predicting DNA duplex stability from the base sequence. Proceedings of the National Academy of Sciences of the United States of America, 83(11), 3746–3750. https://doi.org/10.1073/pnas.83.11.3746

Brisotto, G., di Gennaro, A., Damiano, V., Armellin, M., Perin, T., Maestro, R., & Santarosa, M. (2015). An improved sequencing-based strategy to estimate locus-specific DNA methylation. BMC Cancer, 15(1), 639. https://doi.org/10.1186/s12885-015-1646-6

Bristol Myers Squibb. (2021, September 27). U.S. Food and Drug Administration accepts Bristol Myers Squibb's applications for Opdivo (nivolumab) + Yervoy (ipilimumab) and Opdivo + chemotherapy for unresectable advanced, recurrent or metastatic esophageal squamous cell carcinoma. Bristol Myers Squibb Corporate/Financial News. https://news.bms.com/news/corporate-financial/2021/U.S.-Food-and-Drug-Administration-Accepts-Bristol-Myers-Squibbs-Applications-for-Opdivo-nivolumab--Yervoy-ipilimumab-and-Opdivo--Chemotherapy-for-Unresectable-Advanced-Recurrent-or-Metastatic-Esophageal

Broderick, P., Carvajal-Carmona, L., Pittman, A. M., Webb, E., Howarth, K., Rowan, A., Lubbe, S., Spain, S., Sullivan, K., Fielding, S., Jaeger, E., Vijayakrishnan, J., Kemp, Z., Gorman, M., Chandler, I., Papaemmanuil, E., Penegar, S., Wood, W., Sellick, G., … Houlston, R. S. (2007). A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nature Genetics, 39(11), 1315–1317. https://doi.org/10.1038/ng.2007.18

Brown, N. A., & Elenitoba-Johnson, K. S. J. (2020). Enabling Precision Oncology Through Precision Diagnostics. Annual Review of Pathology: Mechanisms of Disease, 15(1), 97–121. https://doi.org/10.1146/annurev-pathmechdis-012418-012735

CADTH. (2014). Next Generation DNA Sequencing: A Review of the Cost Effectiveness and Guidelines (Issue February). https://pubmed.ncbi.nlm.nih.gov/25654150/

Camidge, D. R. (2014). Targeted therapy vs chemotherapy: which has had more impact on survival in lung cancer? Does targeted therapy make patients live longer? Hard to prove, but impossible to ignore. Clinical Advances in Hematology & Oncology: H&O, 12(11), 763–766.

Cancer Research UK. (2022). Cancer drugs A to Z list. Cancer Research UK. https://www.cancerresearchuk.org/about-cancer/cancer-in-general/treatment/cancer-drugs/drugs

Cancer Research UK. (2022). Targeted and immunotherapy drugs for advanced bowel cancer. Cancer Research UK. https://www.cancerresearchuk.org/about-cancer/bowel-cancer/advanced/treatment/targeted-cancer-drugs-treatment

Cao, R., Yang, F., Ma, S.-C., Liu, L., Zhao, Y., Li, Y., Wu, D.-H., Wang, T., Lu, W.-J., Cai, W.-J., Zhu, H.-B., Guo, X.-J., Lu, Y.-W., Kuang, J.-J., Huan, W.-J., Tang, W.-M., Huang, K., Huang, J., Yao, J., & Dong, Z.-Y. (2020). Development and interpretation of a pathomics-based model for the prediction of microsatellite instability in Colorectal Cancer. Theranostics, 10(24), 11080–11091. https://doi.org/10.7150/thno.49864

Capelle, L. G., Van Grieken, N. C. T., Lingsma, H. F., Steyerberg, E. W., Klokman, W. J., Bruno, M. J., Vasen, H. F. A., & Kuipers, E. J. (2010). Risk and epidemiological time trends of gastric cancer in Lynch syndrome carriers in the Netherlands. Gastroenterology, 138(2), 487–492. https://doi.org/10.1053/j.gastro.2009.10.051

Carethers, J. M. (2020). Fecal DNA Testing for Colorectal Cancer Screening. Annual Review of Medicine, 71, 59–69. https://doi.org/10.1146/annurev-med-103018-123125

Carethers, J. M., Smith, E. J., Behling, C. A., Nguyen, L., Tajima, A., Doctolero, R. T., Cabrera, B. L., Goel, A., Arnold, C. A., Miyai, K., & Boland, C. R. (2004). Use of 5-fluorouracil and survival in patients with microsatellite-unstable colorectal cancer. Gastroenterology, 126(2), 394–401. https://doi.org/10.1053/j.gastro.2003.12.023

Carl-Mcgrath, S., Ebert, M., & Röcken, C. (2007). Gastric adenocarcinoma: epidemiology, pathology and pathogenesis. Cancer Therapy, 5, 879-896

Carlson, C. S., Emerson, R. O., Sherwood, A. M., Desmarais, C., Chung, M.-W., Parsons, J. M., Steen, M. S., LaMadrid-Herrmannsfeldt, M. A., Williamson, D. W., Livingston, R. J., Wu, D., Wood, B. L., Rieder, M. J., & Robins, H. (2013). Using synthetic templates to design an unbiased multiplex PCR assay. Nature Communications, 4(1), 3680. https://doi.org/10.1038/ncomms3680

Cattaneo, F., Molatore, S., Mihalatos, M., Apessos, A., Venesio, T., Bione, S., Grignani, P., Nasioulas, G., & Ranzani, G. N. (2007). Heterogeneous molecular mechanisms underlie attenuated familial adenomatous polyposis. Genetics in Medicine, 9(12), 836–841. https://doi.org/10.1097/GIM.0b013e31815bf940

CD Genomics. (2020). Sanger Sequencing: Introduction, Principle, and Protocol. https://www.cd-genomics.com/blog/sanger-sequencing-introduction-principle-and-protocol/

CEQAS. (2018). Welcome to CEQAS now operating under the name of GenQA. CEQAS. https://www.ceqas.org/

Chen, Y.-C., Liu, T., Yu, C.-H., Chiang, T.-Y., & Hwang, C.-C. (2013). Effects of GC Bias in Next-Generation-Sequencing Data on De Novo Genome Assembly. PLOS ONE, 8(4), e62856. https://doi.org/10.1371/journal.pone.0062856

Cheng, M. L., Pectasides, E., Hanna, G. J., Parsons, H. A., Choudhury, A. D., & Oxnard, G. R. (2021). Circulating tumor DNA in advanced solid tumors: Clinical relevance and future directions. CA: A Cancer Journal for Clinicians, 71(2), 176–190. https://doi.org/https://doi.org/10.3322/caac.21650

Cho, Y. A., Lee, H., Kim, D. G., Kim, H., Ha, S. Y., Choi, Y.-L., Jang, K.-T., & Kim, K.-M. (2021). PD-L1 Expression Is Significantly Associated with Tumor Mutation Burden and Microsatellite Instability Score. Cancers, 13(18). https://doi.org/10.3390/cancers13184659

Clark, D. P., & Pazdernik, N. J. (2013). Chapter 8 - DNA Sequencing (D. P. Clark & N. J. B. T.-M. B. (Second E. Pazdernik (eds.); pp. 227–247). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-378594-7.00008-1

Cleary, J., Peters, T. J., Sharp, D., & Hamilton, W. (2007). Clinical features of colorectal cancer before emergency presentation: a population-based case–control study. Family Practice, 24(1), 3–6. https://doi.org/10.1093/fampra/cml059

Cooper, D. S., Doherty, G. M., Haugen, B. R., Kloos, R. T., Lee, S. L., Mandel, S. J., Mazzaferri, E. L., McIver, B., Pacini, F., Schlumberger, M., Sherman, S. I., Steward, D. L., & Tuttle, R. M. (2009). Revised American Thyroid Association Management Guidelines for Patients with Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid, 19(11), 1167–1214. https://doi.org/10.1089/thy.2009.0110

Cristescu, R., Lee, J., Nebozhyn, M., Kim, K.-M., Ting, J. C., Wong, S. S., Liu, J., Yue, Y. G., Wang, J., Yu, K., Ye, X. S., Do, I.-G., Liu, S., Gong, L., Fu, J., Jin, J. G., Choi, M. G., Sohn, T. S., Lee, J. H., … Aggarwal, A. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nature Medicine, 21(5), 449–456. https://doi.org/10.1038/nm.3850

Crockett, S. D., & Nagtegaal, I. D. (2019). Terminology, Molecular Features, Epidemiology, and Management of Serrated Colorectal Neoplasia. Gastroenterology, 157(4), 949-966.e4. https://doi.org/10.1053/j.gastro.2019.06.041

Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., Clement, T., & Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. Journal of Veterinary Diagnostic Investigation, 32(6), 767–775. https://doi.org/10.1177/1040638720905833

Davidson, B. (2016). Recently identified drug resistance biomarkers in ovarian cancer. Expert Review of Molecular Diagnostics, 16(5), 569–578. https://doi.org/10.1586/14737159.2016.1156532

Davis, A. A., & Patel, V. G. (2019). The role of PD-L1 expression as a predictive biomarker: an analysis of all US Food and Drug Administration (FDA) approvals of immune checkpoint inhibitors. Journal for Immunotherapy of Cancer, 7(1), 278. https://doi.org/10.1186/s40425-019-0768-9

de la Chapelle, A., & Hampel, H. (2010). Clinical relevance of microsatellite instability in colorectal cancer. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 28(20), 3380–3387. https://doi.org/10.1200/JCO.2009.27.0652

de Paula, A. E., Galvão, H. de C. R., Bonatelli, M., Sabato, C., Fernandes, G. C., Berardinelli, G. N., Andrade, C. E. M., Neto, M. C., Romagnolo, L. G. C., Campacci, N., Scapulatempo-Neto, C., Reis, R. M., & Palmero, E. I. (2021). Clinicopathological and molecular characterisation of Brazilian families at risk for Lynch syndrome. Cancer Genetics, 254–255, 82–91. https://doi.org/https://doi.org/10.1016/j.cancergen.2021.02.003

Demetri, G. D., Mehren, M. von, Antonescu, C. R., DeMatteo, R. P., Ganjoo, K. N., Maki, R. G., Pisters, P. W. T., Raut, C. P., Riedel, R. F., Schuetze, S., Sundar, H. M., Trent, J. C., & Wayne, J. D. (2010). NCCN Task Force Report: Update on the Management of Patients with Gastrointestinal Stromal Tumors. Journal of the National Comprehensive Cancer Network J Natl Compr Canc Netw, 8(Suppl_2), S-1-S-41. https://doi.org/10.6004/jnccn.2010.0116

Demicco, E. G. (2013). Sarcoma diagnosis in the age of molecular pathology. Advances in Anatomic Pathology, 20(4), 264–274. https://doi.org/10.1097/PAP.0b013e31829c2c7b

Deng, N., Goh, L. K., Wang, H., Das, K., Tao, J., Tan, I. B., Zhang, S., Lee, M., Wu, J., Lim, K. H., Lei, Z., Goh, G., Lim, Q.-Y., Tan, A. L.-K., Sin Poh, D. Y., Riahi, S., Bell, S., Shi, M. M., Linnartz, R., … Tan, P. (2012). A comprehensive survey of genomic alterations in gastric cancer reveals systematic patterns of molecular exclusivity and co-occurrence among distinct therapeutic targets. Gut, 61(5), 673–684. https://doi.org/10.1136/gutjnl-2011-301839

Dewey, F. E., Pan, S., Wheeler, M. T., Quake, S. R., & Ashley, E. A. (2012). DNA sequencing: clinical applications of new DNA sequencing technologies. Circulation, 125(7), 931–944. https://doi.org/10.1161/CIRCULATIONAHA.110.972828

Dicken, B. J., Bigam, D. L., Cass, C., Mackey, J. R., Joy, A. A., & Hamilton, S. M. (2005). Gastric adenocarcinoma: review and considerations for future directions. Annals of Surgery, 241(1), 27–39. https://doi.org/10.1097/01.sla.0000149300.28588.23

Dietrich, D., Uhl, B., Sailer, V., Holmes, E. E., Jung, M., Meller, S., & Kristiansen, G. (2013). Improved PCR performance using template DNA from formalin-fixed and paraffin-embedded tissues by overcoming PCR inhibition. PloS One, 8(10), e77771–e77771. https://doi.org/10.1371/journal.pone.0077771

Dinarvand, P., Davaro, E. P., Doan, J. V, Ising, M. E., Evans, N. R., Phillips, N. J., Lai, J., & Guzman, M. A. (2019). Familial Adenomatous Polyposis Syndrome: An Update and Review of Extraintestinal Manifestations. Archives of Pathology & Laboratory Medicine, 143(11), 1382–1398. https://doi.org/10.5858/arpa.2018-0570-RA

Dobrowolski, S. F., Pham, H. T., Downes, F. P., Prior, T. W., Naylor, E. W., & Swoboda, K. J. (2012). Newborn screening for spinal muscular atrophy by calibrated short amplicon melt profiling. Clinical Chemistry, 58(6), 1033–1039. https://doi.org/10.1373/clinchem.2012.183038

Dominguez-Valentin, M., Sampson, J. R., Seppälä, T. T., ten Broeke, S. W., Plazzer, J.-P., Nakken, S., Engel, C., Aretz, S., Jenkins, M. A., Sunde, L., Bernstein, I., Capella, G.,

Balaguer, F., Thomas, H., Evans, D. G., Burn, J., Greenblatt, M., Hovig, E., de Vos tot Nederveen Cappel, W. H., … Møller, P. (2020). Cancer risks by gene, age, and gender in 6350 carriers ofpathogenic mismatch repair variants: findings from the Prospective Lynch SyndromeDatabase. Genetics in Medicine, 22(1), 15–25. https://doi.org/10.1038/s41436-019-0596-9

Dong, J., Li, B., Lin, D., Zhou, Q., & Huang, D. (2019). Advances in Targeted Therapy and Immunotherapy for Non-small Cell Lung Cancer Based on Accurate Molecular Typing. Frontiers in Pharmacology, 10, 230. https://doi.org/10.3389/fphar.2019.00230

Dourmishev, L. A., & Draganov, P. V. (2009). Paraneoplastic dermatological manifestation of gastrointestinal malignancies. World Journal of Gastroenterology, 15(35), 4372–4379. https://doi.org/10.3748/wjg.15.4372

Dowsett, M., Cuzick, J., Wale, C., Forbes, J., Mallon, E. A., Salter, J., Quinn, E., Dunbier, A., Baum, M., Buzdar, A., Howell, A., Bugarini, R., Baehner, F. L., & Shak, S. (2010). Prediction of risk of distant recurrence using the 21-gene recurrence score in node-negative and node-positive postmenopausal patients with breast cancer treated with anastrozole or tamoxifen: a TransATAC study. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 28(11), 1829–1834. https://doi.org/10.1200/JCO.2009.24.4798

Drilon, A., Laetsch, T. W., Kummar, S., DuBois, S. G., Lassen, U. N., Demetri, G. D., Nathenson, M., Doebele, R. C., Farago, A. F., Pappo, A. S., Turpin, B., Dowlati, A., Brose, M. S., Mascarenhas, L., Federman, N., Berlin, J., El-Deiry, W. S., Baik, C., Deeken, J., … Hyman, D. M. (2018). Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children. The New England Journal of Medicine, 378(8), 731–739. https://doi.org/10.1056/NEJMoa1714448

Dubbink, H. J., Deans, Z. C., Tops, B. B. J., van Kemenade, F. J., Koljenović, S., van Krieken, H. J. M., Blokx, W. A. M., Dinjens, W. N. M., & Groenen, P. J. T. A. (2014). Next generation diagnostic molecular pathology: critical appraisal of quality assurance in Europe. Molecular Oncology, 8(4), 830–839. https://doi.org/10.1016/j.molonc.2014.03.004

Dwight, Z., Palais, R., & Wittwer, C. T. (2011). uMELT: prediction of high-resolution melting curves and dynamic melting profiles of PCR products in a rich web application. Bioinformatics (Oxford, England), 27(7), 1019–1020. https://doi.org/10.1093/bioinformatics/btr065

Ebili, H. O., Agboola, A. O. J., & Rakha, E. (2021). MSI-WES: a simple approach for microsatellite instability testing using whole exome sequencing. Future Oncology. https://doi.org/10.2217/fon-2021-0132

Ebili, H. O., Hassall, J. C., Fadhil, W., Ham-Karim, H., Asiri, A., Raposo, T. P., Agboola, A. J., & Ilyas, M. (2017a). Squirrel Primer-Based PCR assay for direct and targeted sanger sequencing of short genomic segments. Journal of Biomolecular Techniques, 28(3), 97–110. https://doi.org/10.7171/jbt.17-2803-001

Ebili, H. O., Hassall, J., Asiri, A., Ham-Karim, H., Fadhil, W., Agboola, A. J., & Ilyas, M. (2017b). QMC-PCRx: A novel method for rapid mutation detection. Journal of Clinical Pathology, 70(8), 702–711. https://doi.org/10.1136/jclinpath-2016-204264

Echle, A., Ghaffari Laleh, N., Quirke, P., Grabsch, H. I., Muti, H. S., Saldanha, O. L., Brockmoeller, S. F., van den Brandt, P. A., Hutchins, G. G. A., Richman, S. D., Horisberger, K., Galata, C., Ebert, M. P., Eckardt, M., Boutros, M., Horst, D., Reissfelder, C., Alwers, E., Brinker, T. J., … Kather, J. N. (2022). Artificial intelligence for detection of microsatellite instability in colorectal cancer-a multicentric analysis of a pre-screening tool for clinical application. ESMO Open, 7(2), 100400. https://doi.org/10.1016/j.esmoop.2022.100400

Echle, A., Grabsch, H. I., Quirke, P., van den Brandt, P. A., West, N. P., Hutchins, G. G. A., Heij, L. R., Tan, X., Richman, S. D., Krause, J., Alwers, E., Jenniskens, J., Offermans, K., Gray, R., Brenner, H., Chang-Claude, J., Trautwein, C., Pearson, A. T., Boor, P., … Kather, J. N. (2020). Clinical-Grade Detection of Microsatellite Instability in Colorectal Tumors by Deep Learning. Gastroenterology, 159(4), 1406-1416.e11. https://doi.org/10.1053/j.gastro.2020.06.021

Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., & Kather, J. N. (2021). Deep learning in cancer pathology: a new generation of clinical biomarkers. British Journal of Cancer, 124(4), 686–696. https://doi.org/10.1038/s41416-020-01122-x

El-Deiry, W. S., Goldberg, R. M., Lenz, H.-J., Shields, A. F., Gibney, G. T., Tan, A. R., Brown, J., Eisenberg, B., Heath, E. I., Phuphanich, S., Kim, E., Brenner, A. J., & Marshall, J. L. (2019). The current state of molecular testing in the treatment of patients with solid tumors, 2019. CA: A Cancer Journal for Clinicians, 69(4), 305–343. https://doi.org/10.3322/caac.21560

Elzouki, A.-N., Habel, S., Alsoaeiti, S., Abosedra, A., & Khan, F. (2014). Epidemiology and clinical findings of colorectal carcinoma in two tertiary care hospitals in Benghazi, Libya. Avicenna Journal of Medicine, 4(4), 94–98. https://doi.org/10.4103/2231-0770.140659

Fadhil, W., & Ilyas, M. (2012). Immunostaining for mismatch repair (MMR) protein expression in colorectal cancer is better and easier to interpret when performed on diagnostic biopsies. In Histopathology (Vol. 60, Issue 4, pp. 653–655). https://doi.org/10.1111/j.1365-2559.2011.04021.x

Fadhil, W., Ibrahem, S., Seth, R., & Ilyas, M. (2010). Quick-multiplex-consensus (QMC)-PCR followed by high-resolution melting: a simple and robust method for mutation detection in formalin-fixed paraffin-embedded tissue. Journal of Clinical Pathology, 63(2), 134 LP – 140. https://doi.org/10.1136/jcp.2009.070508

Fadhil, W., Ibrahem, S., Seth, R., AbuAli, G., Ragunath, K., Kaye, P., & Ilyas, M. (2012). The utility of diagnostic biopsy specimens for predictive molecular testing in colorectal cancer. Histopathology, 61(6), 1117–1124. https://doi.org/https://doi.org/10.1111/j.1365-2559.2012.04321.x

Fakruddin, M., & Chowdhury, A. (2012). Pyrosequencing-An Alternative to Traditional Sanger Sequencing. American Journal of Biochemistry and Biotechnology, 8(1 SE-Research Article). https://doi.org/10.3844/ajbbsp.2012.14.20

Fakruddin, Mazumdar, R. M., Chowdhury, A., Hossain, N., Mahajan, S., & Islam, S. (2013). Pyrosequencing-A next generation sequencing technology. World Applied Sciences Journal, 24(12), 1558–1571. https://doi.org/10.5829/idosi.wasj.2013.24.12.2972

Falzone, L., Salomone, S., & Libra, M. (2018). Evolution of Cancer Pharmacological Treatments at the Turn of the Third Millennium. Frontiers in Pharmacology, 9, 1300. https://doi.org/10.3389/fphar.2018.01300

Fang, W.-L., Chang, S.-C., Lan, Y.-T., Huang, K.-H., Chen, J.-H., Lo, S.-S., Hsieh, M.-C., Li, A. F.-Y., Wu, C.-W., & Chiou, S.-H. (2012). Microsatellite instability is associated with a better prognosis for gastric cancer patients after curative surgery. World Journal of Surgery, 36(9), 2131–2138. https://doi.org/10.1007/s00268-012-1652-7

Farrar, J. S., & Wittwer, C. T. (2017). High-Resolution Melting Curve Analysis for Molecular Diagnostics. In Molecular Diagnostics: Third Edition. Elsevier Ltd. https://doi.org/10.1016/B978-0-12-802971-8.00006-7

Ferlay, J, Colombet, M., Soerjomataram, I., Mathers, C., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2019). Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. International Journal of Cancer, 144(8), 1941–1953. https://doi.org/10.1002/ijc.31937

Ferlay, Jacques, Colombet, M., Soerjomataram, I., Parkin, D. M., Piñeros, M., Znaor, A., & Bray, F. (2021). Cancer statistics for the year 2020: An overview. International Journal of Cancer, 149(4), 778–789. https://doi.org/https://doi.org/10.1002/ijc.33588

Fey, M. F. (2011). VI. DNA microarray technology. Principles and application to the analysis of malignant tumours (with special emphasis on lymphoma). Annals of Oncology, 22, iv44–iv46. https://doi.org/10.1093/annonc/mdr172

Fisher, S. E., & Daniels, I. R. (2007). The clinical presentation of colorectal cancer. In G. Brown (Ed.), Colorectal Cancer (pp. 1–14). Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511902468.002

Fitarelli-Kiehl, M., Macedo, G. S., Schlatter, R. P., Koehler-Santos, P., Matte, U. da S., Ashton-Prolla, P., & Giacomazzi, J. (2016). Comparison of multiple genotyping methods for the identification of the cancer predisposing founder mutation p.R337H in TP53. Genetics and Molecular Biology, 39(2), 203–209. https://doi.org/10.1590/1678-4685-GMB-2014-0351

Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., Ding, M., Bamford, S., Cole, C., Ward, S., Kok, C. Y., Jia, M., De, T., Teague, J. W., Stratton, M. R., McDermott, U., & Campbell, P. J. (2015). COSMIC: exploring the world's

knowledge of somatic mutations in human cancer. Nucleic Acids Research, 43(Database issue), D805–D811. https://doi.org/10.1093/nar/gku1075

França, L. T. C., Carrilho, E., & Kist, T. B. L. (2002). A review of DNA sequencing techniques. Quarterly Reviews of Biophysics, 35(2), 169–200. https://doi.org/DOI: 10.1017/S0033583502003797

Fuccio, L., Eusebi, L. H., & Bazzoli, F. (2010). Gastric cancer, Helicobacter pylori infection and other risk factors. World Journal of Gastrointestinal Oncology, 2(9), 342–347. https://doi.org/10.4251/wjgo.v2.i9.342

Fujimoto, A., Fujita, M., Hasegawa, T., Wong, J. H., Maejima, K., Oku-Sasaki, A., Nakano, K., Shiraishi, Y., Miyano, S., Yamamoto, G., Akagi, K., Imoto, S., & Nakagawa, H. (2020). Comprehensive analysis of indels in whole-genome microsatellite regions and microsatellite instability across 21 cancer types. Genome Research, 30(3), 334–346. https://doi.org/10.1101/gr.255026.119

Furutani, S., Furutani, N., Kawai, Y., Nakayama, A., & Nagai, H. (2022). Rapid DNA Sequencing Technology Based on the Sanger Method for Bacterial Identification. In Sensors (Vol. 22, Issue 6). https://doi.org/10.3390/s22062130

Gallon, R., Gawthorpe, P., Phelps, R. L., Hayes, C., Borthwick, G. M., Santibanez-Koref, M., Jackson, M. S., & Burn, J. (2021). How Should We Test for Lynch Syndrome? A Review of Current Guidelines and Future Strategies. In Cancers (Vol. 13, Issue 3). https://doi.org/10.3390/cancers13030406

Garritano, S., Gemignani, F., Voegele, C., Nguyen-Dumont, T., Le Calvez-Kelm, F., De Silva, D., Lesueur, F., Landi, S., & Tavtigian, S. V. (2009). Determining the effectiveness of High Resolution Melting analysis for SNP genotyping and mutation scanning at the TP53 locus. BMC Genetics, 10(1), 5. https://doi.org/10.1186/1471-2156-10-5

Gerini, C., Dal Canto, M., & Porfirio, B. (2014). High resolution melting analysis of deletion/insertion polymorphisms: A new method for the detection and quantification of mixed chimerism in allogeneic stem cell transplantation. Molecular and Cellular Probes, 28(1), 19–24. https://doi.org/https://doi.org/10.1016/j.mcp.2013.09.001

Gharizadeh, B., Ghaderi, M., & Nyeren, P. (2007). Pyrosequencing Technology for Short DNA Sequencing and Whole Genome Sequencing. Seibutsu Butsuri, 47(2), 129–132. https://doi.org/10.2142/biophys.47.129

Ghosh, S., Hossain, M. Z., Borges, M., Goggins, M. G., Ingersoll, R. G., Eshleman, J. R., Klein, A. P., & Kern, S. E. (2012). Analysis of Polymorphisms and Haplotype Structure of the Human Thymidylate Synthase Genetic Region: A Tool for Pharmacogenetic Studies. PLOS ONE, 7(4), e34426. https://doi.org/10.1371/journal.pone.0034426

Gilbert, M. T. P., Haselkorn, T., Bunce, M., Sanchez, J. J., Lucas, S. B., Jewell, L. D., Van Marck, E., & Worobey, M. (2007). The isolation of nucleic acids from fixed, paraffin-

embedded tissues-which methods are useful when? PloS One, 2(6), e537–e537. https://doi.org/10.1371/journal.pone.0000537

Goh, R.-Y., Cho, S.-S., Song, Y.-J., Heo, K., Oh, S.-Y., Kim, S.-H., Kwon, H.-C., Kim, H.-J., & Han, J.-Y. (2009). Clinical Utility of Chimerism Status Assessed by Lineage-Specific Short Tandem Repeat Analysis: Experience from Four Cases of Allogeneic Stem Cell Transplantation. Annals of Laboratory Medicine, 29(4), 277–281. https://doi.org/10.3343/kjlm.2009.29.4.277

Goldstein, L. J., Gray, R., Badve, S., Childs, B. H., Yoshizawa, C., Rowley, S., Shak, S., Baehner, F. L., Ravdin, P. M., Davidson, N. E., Sledge, G. W. J., Perez, E. A., Shulman, L. N., Martino, S., & Sparano, J. A. (2008). Prognostic utility of the 21-gene assay in hormone receptor-positive operable breast cancer compared with classical clinicopathologic features. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 26(25), 4063–4071. https://doi.org/10.1200/JCO.2007.14.4501

Gonzalez, R. S., Washington, K., & Shi, C. (2017). Current applications of molecular pathology in colorectal carcinoma. Applied Cancer Research, 37(1), 13. https://doi.org/10.1186/s41241-017-0020-1

Grabsch, H. I., & Tan, P. (2013). Gastric Cancer Pathology and Underlying Molecular Mechanisms. Digestive Surgery, 30(2), 150–158. https://doi.org/10.1159/000350876

Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., Edkins, S., O'Meara, S., Vastrik, I., Schmidt, E. E., Avis, T., Barthorpe, S., Bhamra, G., Buck, G., Choudhury, B., … Stratton, M. R. (2007). Patterns of somatic mutation in human cancer genomes. Nature, 446(7132), 153–158. https://doi.org/10.1038/nature05610

Greenson, J. K., Bonner, J. D., Ben-Yzhak, O., Cohen, H. I., Miselevich, I., Resnick, M. B., Trougouboff, P., Tomsho, L. D., Kim, E., Low, M., Almog, R., Rennert, G., & Gruber, S. B. (2003). Phenotype of microsatellite unstable colorectal carcinomas: Well-differentiated and focally mucinous tumors and the absence of dirty necrosis correlate with microsatellite instability. The American Journal of Surgical Pathology, 27(5), 563–570. https://doi.org/10.1097/00000478-200305000-00001

Groenen, P. J. T. A., Blokx, W. A. M., Diepenbroek, C., Burgers, L., Visinoni, F., Wesseling, P., & van Krieken, J. H. J. M. (2011). Preparing pathology for personalised medicine: possibilities for improvement of the pre-analytical phase. Histopathology, 59(1), 1–7. https://doi.org/https://doi.org/10.1111/j.1365-2559.2010.03711.x

Gryfe, R., Kim, H., Hsieh, E. T. K., Aronson, M. D., Holowaty, E. J., Bull, S. B., Redston, M., & Gallinger, S. (2000). Tumor Microsatellite Instability and Clinical Outcome in Young Patients with Colorectal Cancer. New England Journal of Medicine, 342(2), 69–77. https://doi.org/10.1056/NEJM200001133420201

Guend, H., Patel, S., & Nash, G. M. (2015). Abdominal metastases from colorectal cancer: intraperitoneal therapy. Journal of Gastrointestinal Oncology, 6(6), 693–698. https://doi.org/10.3978/j.issn.2078-6891.2015.078

Guinney, J., Dienstmann, R., Wang, X., de Reyniès, A., Schlicker, A., Soneson, C., Marisa, L., Roepman, P., Nyamundanda, G., Angelino, P., Bot, B. M., Morris, J. S., Simon, I. M., Gerster, S., Fessler, E., De Sousa E Melo, F., Missiaglia, E., Ramay, H., Barras, D., … Tejpar, S. (2015). The consensus molecular subtypes of colorectal cancer. Nature Medicine, 21(11), 1350–1356. https://doi.org/10.1038/nm.3967

Guo, Q., Xiao, L., & Zhou, Y. (2012). Rapid Diagnosis of Aneuploidy by High-Resolution Melting Analysis of Segmental Duplications. Clinical Chemistry, 58(6), 1019–1025. https://doi.org/10.1373/clinchem.2011.178475

Habel, L. A., Shak, S., Jacobs, M. K., Capra, A., Alexander, C., Pho, M., Baker, J., Walker, M., Watson, D., Hackett, J., Blick, N. T., Greenberg, D., Fehrenbacher, L., Langholz, B., & Quesenberry, C. P. (2006). A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. Breast Cancer Research, 8(3), R25. https://doi.org/10.1186/bcr1412

Hagemann, I. S. (2015). Overview of Technical Aspects and Chemistries of Next-Generation Sequencing. In Clinical Genomics. Elsevier Inc. https://doi.org/10.1016/B978-0-12-404748-8.00001-0

Haiman, C. A., Le Marchand, L., Yamamato, J., Stram, D. O., Sheng, X., Kolonel, L. N., Wu, A. H., Reich, D., & Henderson, B. E. (2007). A common genetic risk factor for colorectal and prostate cancer. Nature Genetics, 39(8), 954–956. https://doi.org/10.1038/ng2098

Halvarsson, B., Anderson, H., Domanska, K., Lindmark, G., & Nilbert, M. (2008). Clinicopathologic factors identify sporadic mismatch repair-defective colon cancers. American Journal of Clinical Pathology, 129(2), 238–244. https://doi.org/10.1309/0PP5GDRTXUDVKAWJ

Hamilton, A., Tétreault, M., Dyment, D. A., Zou, R., Kernohan, K., Geraghty, M. T., Consortium, F. C., Consortium, C. C., Hartley, T., & Boycott, K. M. (2016). Concordance between whole-exome sequencing and clinical Sanger sequencing: implications for patient care. Molecular Genetics & Genomic Medicine, 4(5), 504–512. https://doi.org/https://doi.org/10.1002/mgg3.223

Hamilton, W., Round, A., Sharp, D., & Peters, T. J. (2005). Clinical features of colorectal cancer before diagnosis: a population-based case-control study. British Journal of Cancer, 93(4), 399–405. https://doi.org/10.1038/sj.bjc.6602714

Ham-Karim H., Ebili H.O., Fadhil W., Asiri A., Hassall J., Ilyas M. (2017). COLD-HRM: A Combination of Methods to Infer the Nature of Somatic Mutations. Advances in Cytology and Pathology, 2(2), 00017. DOI: 10.15406/acp.2017.02.00017

Hansford, S., Kaurah, P., Li-Chang, H., Woo, M., Senz, J., Pinheiro, H., Schrader, K. A., Schaeffer, D. F., Shumansky, K., Zogopoulos, G., Santos, T. A., Claro, I., Carvalho, J., Nielsen, C., Padilla, S., Lum, A., Talhouk, A., Baker-Lange, K., Richardson, S., … Huntsman, D. G. (2015). Hereditary Diffuse Gastric Cancer Syndrome: CDH1 Mutations and Beyond. JAMA Oncology, 1(1), 23–32. https://doi.org/10.1001/jamaoncol.2014.168

Haugen, B. R., Alexander, E. K., Bible, K. C., Doherty, G. M., Mandel, S. J., Nikiforov, Y. E., Pacini, F., Randolph, G. W., Sawka, A. M., Schlumberger, M., Schuff, K. G., Sherman, S. I., Sosa, J. A., Steward, D. L., Tuttle, R. M., & Wartofsky, L. (2016). 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid: Official Journal of the American Thyroid Association, 26(1), 1–133. https://doi.org/10.1089/thy.2015.0020

Hause, R. J., Pritchard, C. C., Shendure, J., & Salipante, S. J. (2016). Classification and characterisation of microsatellite instability across 18 cancer types. Nature Medicine, 22(11), 1342–1350. https://doi.org/10.1038/nm.4191

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics, 107(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Heinemann, V., von Weikersthal, L. F., Decker, T., Kiani, A., Vehling-Kaiser, U., Al-Batran, S.-E., Heintges, T., Lerchenmüller, C., Kahl, C., Seipelt, G., Kullmann, F., Stauch, M., Scheithauer, W., Hielscher, J., Scholz, M., Müller, S., Link, H., Niederle, N., Rost, A., … Stintzing, S. (2014). FOLFIRI plus cetuximab versus FOLFIRI plus bevacizumab as first-line treatment for patients with metastatic colorectal cancer (FIRE-3): a randomised, open-label, phase 3 trial. The Lancet Oncology, 15(10), 1065–1075. https://doi.org/10.1016/S1470-2045(14)70330-4

Helsmoortel, C., Kooy, R. F., & Vandeweyer, G. (2016). Multiplexed High Resolution Melting Assay for Versatile Sample Tracking in a Diagnostic and Research Setting. The Journal of Molecular Diagnostics, 18(1), 32–38. https://doi.org/10.1016/j.jmoldx.2015.06.011

Henry, L. R., Stojadinovic, A., Swain, S. M., Prindiville, S., Cordes, R., & Soballe, P. W. (2009). The influence of a Gene expression profile on breast cancer decisions. Journal of Surgical Oncology, 99(6), 319–323. https://doi.org/https://doi.org/10.1002/jso.21244

Holch, J. W., Demmer, M., Lamersdorf, C., Michl, M., Schulz, C., von Einem, J. C., Modest, D. P., & Heinemann, V. (2017). Pattern and Dynamics of Distant Metastases in Metastatic Colorectal Cancer. Visceral Medicine, 33(1), 70–75. https://doi.org/10.1159/000454687

Hong, D. S., Fakih, M. G., Strickler, J. H., Desai, J., Durm, G. A., Shapiro, G. I., Falchook, G. S., Price, T. J., Sacher, A., Denlinger, C. S., Bang, Y.-J., Dy, G. K., Krauss, J. C., Kuboki, Y., Kuo, J. C., Coveler, A. L., Park, K., Kim, T. W., Barlesi, F., … Li, B. T. (2020). KRAS(G12C) Inhibition with Sotorasib in Advanced Solid Tumors. The New England Journal of Medicine, 383(13), 1207–1217. https://doi.org/10.1056/NEJMoa1917239

Horpaopan, S., & Somran, J. (2020). The utility of histopathological features predicting microsatellite high (MSI-H) colorectal cancer in a limited setting. Int J Clin Exp Med. 13(3):1878-1887.

Hsiao, S. J. (2019). Sources of error in molecular diagnostic analyses. In: A. Dasgupta & J. Sepulveda (Eds.), Accurate Results in the Clinical Laboratory (2nd ed., pp. 337–347). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-813776-5.00021-2

Hsu, J. C., & Lu, C. Y. (2016). Longitudinal trends in use and costs of targeted therapies for common cancers in Taiwan: a retrospective observational study. BMJ Open, 6(6), e011322. https://doi.org/10.1136/bmjopen-2016-011322

Hu, X.-T., & He, C. (2013). Recent progress in the study of methylated tumor suppressor genes in gastric cancer. Chinese Journal of Cancer, 32(1), 31–41. https://doi.org/10.5732/cjc.011.10175

Huo, W., Ling, W., Wang, Z., Li, Y., Zhou, M., Ren, M., Li, X., Li, J., Xia, Z., Liu, X., & Huang, X. (2021). Miniaturised DNA Sequencers for Personal Use: Unreachable Dreams or Achievable Goals. In Frontiers in Nanotechnology (Vol. 3). https://www.frontiersin.org/article/10.3389/fnano.2021.628861

IARC Monograph. (1994). Schistosomes, Liver Flukes and Helicobacter Pylori. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans (No. 61). https://www.ncbi.nlm.nih.gov/books/NBK487782/pdf/Bookshelf_NBK487782.pdf

Ibrahem, S., Seth, R., O'Sullivan, B., Fadhil, W., Taniere, P., & Ilyas, M. (2010). Comparative analysis of pyrosequencing and QMC-PCR in conjunction with high resolution melting for KRAS/BRAF mutation detection. International Journal of Experimental Pathology, 91(6), 500–505. https://doi.org/10.1111/j.1365-2613.2010.00733.x

Iida, K., & Nishimura, I. (2002). Gene expression profiling by DNA microarray technology. Critical Reviews in Oral Biology and Medicine: An Official Publication of the American Association of Oral Biologists, 13(1): 35–50. https://doi.org/10.1177/154411130201300105

Imyanitov, E. N., Buslov, K. G., Suspitsin, E. N., Kuligina, E. S., Belogubova, E. V, Grigoriev, M. Y., Togo, A. V, & Hanson, K. P. (2002). Improved Reliability of Allele-Specific PCR. BioTechniques, 33(3), 484–490. https://doi.org/10.2144/02333bm04

Ishige, T., Itoga, S., & Matsushita, K. (2018). Locked Nucleic Acid Technology for Highly Sensitive Detection of Somatic Mutations in Cancer. Advances in Clinical Chemistry, 83, 53–72. https://doi.org/10.1016/bs.acc.2017.10.002

Jaeger, E., Webb, E., Howarth, K., Carvajal-Carmona, L., Rowan, A., Broderick, P., Walther, A., Spain, S., Pittman, A., Kemp, Z., Sullivan, K., Heinimann, K., Lubbe, S., Domingo, E., Barclay, E., Martin, L., Gorman, M., Chandler, I., Vijayakrishnan, J., … Tomlinson, I. (2008). Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3

influence colorectal cancer risk. Nature Genetics, 40(1), 26–28. https://doi.org/10.1038/ng.2007.41

Jajesniak, P., & Wong, T. S. (2015). QuickStep-Cloning: a sequence-independent, ligation-free method for rapid construction of recombinant plasmids. Journal of Biological Engineering, 9(1), 15. https://doi.org/10.1186/s13036-015-0010-3

Janavicius, R., Matiukaite, D., Jakubauskas, A., & Griskevicius, L. (2010). Microsatellite Instability Detection by High-Resolution Melting Analysis. Clinical Chemistry, 56(11), 1750–1757. https://doi.org/10.1373/clinchem.2010.150680

Jänne, P. A., Rybkin, I. I., Spira, A. I., Riely, G. J., Papadopoulos, K. P., Sabari, J. K., Johnson, M. L., Heist, R. S., Bazhenova, L., Barve, M., Pacheco, J. M., Leal, T. A., Velastegui, K., Cornelius, C., Olson, P., Christensen, J. G., Kheoh, T., Chao, R. C., & Ou, S. H. I. (2020). KRYSTAL-1: Activity and Safety of Adagrasib (MRTX849) in Advanced/ Metastatic Non-Small Cell Lung Cancer (NSCLC) Harboring KRAS G12C Mutation. European Journal of Cancer, 138, S1–S2. https://doi.org/10.1016/S0959-8049(20)31076-5

Jass, J. R. (2007). Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology, 50(1), 113–130. https://doi.org/https://doi.org/10.1111/j.1365-2559.2006.02549.x

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. CA: A Cancer Journal for Clinicians, 61(2), 69–90. https://doi.org/https://doi.org/10.3322/caac.20107

Jenkins, M. A., Hayashi, S., O'Shea, A.-M., Burgart, L. J., Smyrk, T. C., Shimizu, D., Waring, P. M., Ruszkiewicz, A. R., Pollett, A. F., Redston, M., Barker, M. A., Baron, J. A., Casey, G. R., Dowty, J. G., Giles, G. G., Limburg, P., Newcomb, P., Young, J. P., Walsh, M. D., … Registry, C. C. F. (2007). Pathology features in Bethesda guidelines predict colorectal cancer microsatellite instability: a population-based study. Gastroenterology, 133(1), 48–56. https://doi.org/10.1053/j.gastro.2007.04.044

Jin, J., Wu, X., Yin, J., Li, M., Shen, J., Li, J., Zhao, Y., Zhao, Q., Wu, J., Wen, Q., Cho, C. H., Yi, T., Xiao, Z., & Qu, L. (2019). Identification of Genetic Mutations in Cancer: Challenge and Opportunity in the New Era of Targeted Therapy. Frontiers in Oncology, 9, 263. https://doi.org/10.3389/fonc.2019.00263

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: a better web interface. Nucleic Acids Research, 36(Web Server issue), W5–W9. https://doi.org/10.1093/nar/gkn201

Johnston, L., Power, M., Sloan, P., Long, A., Silmon, A., Chaffey, B., Lisgo, A. J., Little, L., Vercauteren, E., Steiniche, T., Meyer, T., & Simpson, J. (2018). Clinical performance evaluation of the Idylla NRAS-BRAF mutation test on retrospectively collected formalin-fixed paraffin-embedded colorectal cancer tissue. Journal of Clinical Pathology, 71(4), 336–343. https://doi.org/10.1136/jclinpath-2017-204629

Jonchere, V., Marisa, L., Greene, M., Virouleau, A., Buhard, O., Bertrand, R., Svrcek, M., Cervera, P., Goloudina, A., Guillerm, E., Coulet, F., Landman, S., Ratovomanana, T., Job, S., Ayadi, M., Elarouci, N., Armenoult, L., Merabtene, F., Dumont, S., … Duval, A. (2018). Identification of Positively and Negatively Selected Driver Gene Mutations Associated With Colorectal Cancer With Microsatellite Instability. Cellular and Molecular Gastroenterology and Hepatology, 6(3), 277–300. https://doi.org/https://doi.org/10.1016/j.jcmgh.2018.06.002

Kabbinavar, F. F., Hambleton, J., Mass, R. D., Hurwitz, H. I., Bergsland, E., & Sarkar, S. (2005). Combined Analysis of Efficacy: The Addition of Bevacizumab to Fluorouracil/Leucovorin Improves Survival for Patients With Metastatic Colorectal Cancer. Journal of Clinical Oncology, 23(16), 3706–3712. https://doi.org/10.1200/JCO.2005.00.232

Kalendar, R., Lee, D., & Schulman, A. H. (2011). Java web tools for PCR, in silico PCR, and oligonucleotide assembly and analysis. Genomics, 98(2), 137–144. https://doi.org/https://doi.org/10.1016/j.ygeno.2011.04.009

Kamps, R., Brandão, R. D., Bosch, B. J. van den, Paulussen, A. D. C., Xanthoulea, S., Blok, M. J., & Romano, A. (2017). Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. International Journal of Molecular Sciences, 18(2), 308. https://doi.org/10.3390/ijms18020308

Kanopiene, D., Vidugiriene, J., Povilas Valuckas, K., Smailyte, G., Uleckiene, S., & Bacher, J. (2014). Endometrial cancer and microsatellite instability status. Open Medicine, 10(1). https://doi.org/doi:10.1515/med-2015-0005

Kantarjian, H., O'Brien, S., Jabbour, E., Garcia-Manero, G., Quintas-Cardama, A., Shan, J., Rios, M. B., Ravandi, F., Faderl, S., Kadia, T., Borthakur, G., Huang, X., Champlin, R., Talpaz, M., & Cortes, J. (2012). Improved survival in chronic myeloid leukemia since the introduction of imatinib therapy: a single-institution historical experience. Blood, 119(9), 1981–1987. https://doi.org/10.1182/blood-2011-08-358135

Kather, J. N., Heij, L. R., Grabsch, H. I., Loeffler, C., Echle, A., Muti, H. S., Krause, J., Niehues, J. M., Sommer, K. A. J., Bankhead, P., Kooreman, L. F. S., Schulte, J. J., Cipriani, N. A., Buelow, R. D., Boor, P., Ortiz-Brüchle, N.-N., Hanby, A. M., Speirs, V., Kochanny, S., … Luedde, T. (2020). Pan-cancer image-based detection of clinically actionable genetic alterations. Nature Cancer, 1(8), 789–799. https://doi.org/10.1038/s43018-020-0087-6

Kather, J. N., Pearson, A. T., Halama, N., Jäger, D., Krause, J., Loosen, S. H., Marx, A., Boor, P., Tacke, F., Neumann, U. P., Grabsch, H. I., Yoshikawa, T., Brenner, H., Chang-Claude, J., Hoffmeister, M., Trautwein, C., & Luedde, T. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nature Medicine, 25(7), 1054–1056. https://doi.org/10.1038/s41591-019-0462-y

Kautto, E. A., Bonneville, R., Miya, J., Yu, L., Krook, M. A., Reeser, J. W., & Roychowdhury, S. (2017). Performance evaluation for rapid detection of pan-cancer microsatellite

instability with MANTIS. Oncotarget, 8(5), 7452–7463. https://doi.org/10.18632/oncotarget.13918

Khangura, R. K., Khangura, C. K., Desai, A., Goyert, G., & Sangha, R. (2015). Metastatic Colorectal Cancer Resembling Severe Preeclampsia in Pregnancy. Case Reports in Obstetrics and Gynecology, 2015, 487824. https://doi.org/10.1155/2015/487824

Kieleczawa, J. (2006). Fundamentals of sequencing of difficult templates--an overview. Journal of Biomolecular Techniques: JBT, 17(3), 207–217. https://pubmed.ncbi.nlm.nih.gov/16870712

Kikuchi, A., Sawamura, T., & Nonaka, K. (2018). Detection of Genetic Mutations Using Minor Variant Finder Software. American Journal of Clinical Pathology, 150(suppl_1), S132–S132. https://doi.org/10.1093/ajcp/aqy102.316

Kim, H. S., Woo, D. K., Bae, S. I., Kim, Y. I., & Kim, W. H. (2000). Microsatellite instability in the adenoma-carcinoma sequence of the stomach. Laboratory Investigation; a Journal of Technical Methods and Pathology, 80(1), 57–64. https://doi.org/10.1038/labinvest.3780008

Kim, J. H., & Kang, G. H. (2020). Evolving pathologic concepts of serrated lesions of the colorectum. Journal of Pathology and Translational Medicine, 54(4), 276–289. https://doi.org/10.4132/jptm.2020.04.15

Kim, J.-Y., Shin, N. R., Kim, A., Lee, H.-J., Park, W.-Y., Kim, J.-Y., Lee, C.-H., Huh, G.-Y., & Park, D. Y. (2013). Microsatellite instability status in gastric cancer: a reappraisal of its clinical significance and relationship with mucin phenotypes. Korean Journal of Pathology, 47(1), 28–35. https://doi.org/10.4132/KoreanJPathol.2013.47.1.28

Kim, K.-J., Kim, J.-W., Sung, J. H., Suh, K. J., Lee, J. Y., Kim, S. H., Lee, J.-O., Kim, J. W., Kim, Y. J., Kim, J. H., Bang, S.-M., Lee, J. S., Kim, H. K., & Lee, K.-W. (2020). PI3K-targeting strategy using alpelisib to enhance the antitumor effect of paclitaxel in human gastric cancer. Scientific Reports, 10(1), 12308. https://doi.org/10.1038/s41598-020-68998-w

Kim, S. T., Cristescu, R., Bass, A. J., Kim, K.-M., Odegaard, J. I., Kim, K., Liu, X. Q., Sher, X., Jung, H., Lee, M., Lee, S., Park, S. H., Park, J. O., Park, Y. S., Lim, H. Y., Lee, H., Choi, M., Talasaz, A., Kang, P. S., … Kang, W. K. (2018). Comprehensive molecular characterisation of clinical responses to PD-1 inhibition in metastatic gastric cancer. Nature Medicine, 24(9), 1449–1458. https://doi.org/10.1038/s41591-018-0101-z

Kim, S. Y., Choi, Y. Y., An, J. Y., Shin, H. B., Jo, A., Choi, H., Seo, S. H., Bang, H.-J., Cheong, J.-H., Hyung, W. J., & Noh, S. H. (2015). The benefit of microsatellite instability is attenuated by chemotherapy in stage II and stage III gastric cancer: Results from a large cohort with subgroup analyses. International Journal of Cancer, 137(4), 819–825. https://doi.org/https://doi.org/10.1002/ijc.29449

Kim, T.-M., Laird, P. W., & Park, P. J. (2013). The landscape of microsatellite instability in colorectal and endometrial cancer genomes. Cell, 155(4), 858–868. https://doi.org/10.1016/j.cell.2013.10.015

Kimura, Y., Noguchi, T., Kawahara, K., Kashima, K., Daa, T., & Yokoyama, S. (2004). Genetic alterations in 102 primary gastric cancers by comparative genomic hybridisation: gain of 20q and loss of 18q are associated with tumor progression. Modern Pathology, 17(11), 1328–1337. https://doi.org/10.1038/modpathol.3800180

Kishore, S., Piscuoglio, S., Kovac, M. B., Gylling, A., Wenzel, F., Trapani, F., Altermatt, H. J., Mele, V., Marra, G., Peltomäki, P., Terracciano, L., Zavolan, M., & Heinimann, K. (2014). 3′-UTR Poly(T/U) Tract Deletions and Altered Expression of *EWSR1* Are a Hallmark of Mismatch Repair–Deficient Cancers. Cancer Research, 74(1), 224 – 234. https://doi.org/10.1158/0008-5472.CAN-13-2100

Kittaneh, M., Montero, A. J., & Glück, S. (2013). Molecular profiling for breast cancer: a comprehensive review. Biomarkers in Cancer, 5, 61–70. https://doi.org/10.4137/BIC.S9455

Klarskov, L., Ladelund, S., Holck, S., Roenlund, K., Lindebjerg, J., Elebro, J., Halvarsson, B., von Salomé, J., Bernstein, I., & Nilbert, M. (2010). Interobserver variability in the evaluation of mismatch repair protein immunostaining. Human Pathology, 41(10), 1387–1396. https://doi.org/10.1016/j.humpath.2010.03.003

Koncina, E., Haan, S., Rauh, S., & Letellier, E. (2020). Prognostic and Predictive Molecular Biomarkers for Colorectal Cancer: Updates and Challenges. In Cancers (Vol. 12, Issue 2). https://doi.org/10.3390/cancers12020319

Kondelin, J., Tuupanen, S., Gylfe, A. E., Aavikko, M., Renkonen-Sinisalo, L., Järvinen, H., Böhm, J., Mecklin, J.-P., Andersen, C. L., Vahteristo, P., Pitkänen, E., & Aaltonen, L. A. (2015). 3′-UTR poly(T/U) repeat of EWSR1 is altered in microsatellite unstable colorectal cancer with nearly perfect sensitivity. Familial Cancer, 14(3), 449–453. https://doi.org/10.1007/s10689-015-9804-1

Kopetz, S., Grothey, A., Yaeger, R., Van Cutsem, E., Desai, J., Yoshino, T., Wasan, H., Ciardiello, F., Loupakis, F., Hong, Y. S., Steeghs, N., Guren, T. K., Arkenau, H.-T., Garcia-Alfonso, P., Pfeiffer, P., Orlov, S., Lonardi, S., Elez, E., Kim, T.-W., … Tabernero, J. (2019). Encorafenib, Binimetinib, and Cetuximab in BRAF V600E–Mutated Colorectal Cancer. New England Journal of Medicine, 381(17), 1632–1643. https://doi.org/10.1056/NEJMoa1908075

Kulski, J. K. (2016). Next-Generation Sequencing — An Overview of the History, Tools, and "Omic" Applications. In Next Generation Sequencing - Advances, Applications and Challenges [Internet] (pp. 3–60). IntechOpen. https://doi.org/10.5772/61964

Künstlinger, H., Binot, E., Merkelbach-Bruse, S., Huss, S., Wardelmann, E., Buettner, R., & Schildhaus, H.-U. (2014). High-resolution melting analysis is a sensitive diagnostic tool to detect imatinib-resistant and imatinib-sensitive PDGFRA exon 18 mutations in

gastrointestinal stromal tumors. Human Pathology, 45(3), 573–582. https://doi.org/https://doi.org/10.1016/j.humpath.2013.10.025

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., … Trust, T. W. (2001). Initial sequencing and analysis of the human genome. Nature, 409(6822), 860–921. https://doi.org/10.1038/35057062

Lechner, K., & Obermeier, H. L. (2012). Cancer-related microangiopathic hemolytic anemia: clinical and laboratory features in 168 reported cases. Medicine, 91(4), 195–205. https://doi.org/10.1097/MD.0b013e3182603598

Lee, H., Seo, J., Lee, G., Park, J., Yeo, D., & Hong, A. (2021). Two-Stage Classification Method for MSI Status Prediction Based on Deep Learning Approach. In Applied Sciences (Vol. 11, Issue 1). https://doi.org/10.3390/app11010254

Lee, J., van Hummelen, P., Go, C., Palescandolo, E., Jang, J., Park, H. Y., Kang, S. Y., Park, J. O., Kang, W. K., MacConaill, L., & Kim, K.-M. (2012). High-Throughput Mutation Profiling Identifies Frequent Somatic Mutations in Advanced Gastric Adenocarcinoma. PLOS ONE, 7(6), e38892. https://doi.org/10.1371/journal.pone.0038892

Lee, M., Chun, S.-M., Sung, C. O., Kim, S. Y., Kim, T. W., Jang, S. J., & Kim, J. (2019). Clinical Utility of a Fully Automated Microsatellite Instability Test with Minimal Hands-on Time. J Pathol Transl Med, 53(6), 386–392. https://doi.org/10.4132/jptm.2019.09.25

Lee, Y., Lee, J. A., Park, H. E., Han, H., Kim, Y., Bae, J. M., Kim, J. H., Cho, N.-Y., Kim, H.-P., Kim, T.-Y., & Kang, G. H. (2021). Targeted next-generation sequencing-based detection of microsatellite instability in colorectal carcinomas. PLOS ONE, 16(2), e0246356. https://doi.org/10.1371/journal.pone.0246356

Lefaucheur, C., Stengel, B., Nochy, D., Martel, P., Hill, G. S., Jacquot, C., & Rossert, J. (2006). Membranous nephropathy and cancer: Epidemiologic evidence and determinants of high-risk cancer association. Kidney International, 70(8), 1510–1517. https://doi.org/10.1038/sj.ki.5001790

Lei, Z., Tan, I. B., Das, K., Deng, N., Zouridis, H., Pattison, S., Chua, C., Feng, Z., Guan, Y. K., Ooi, C. H., Ivanova, T., Zhang, S., Lee, M., Wu, J., Ngo, A., Manesh, S., Tan, E., Teh, B. T., So, J. B. Y., … Rozen, S. G. (2013). Identification of molecular subtypes of gastric cancer with different responses to PI3-kinase inhibitors and 5-fluorouracil. Gastroenterology, 145(3), 554–565. https://doi.org/10.1053/j.gastro.2013.05.010

Lemery, S., Keegan, P., & Pazdur, R. (2017). First FDA Approval Agnostic of Cancer Site — When a Biomarker Defines the Indication. New England Journal of Medicine, 377(15), 1409–1412. https://doi.org/10.1056/NEJMp1709968

Leong, K. W. (2020). The State of Centralisation of Pathology and Molecular Testing in England. https://www.art25capital.com/en/news/the-state-of-centralisation-of-pathology-and-molecular-testing-in-england

Leoz, M. L., Carballal, S., Moreira, L., Ocaña, T., & Balaguer, F. (2015). The genetic basis of familial adenomatous polyposis and its implications for clinical practice and risk management. Application of Clinical Genetics, 8, 95–107. https://doi.org/10.2147/TACG.S51484

Li, C., Wen, A., Shen, B., Lu, J., Huang, Y., & Chang, Y. (2011). FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. BMC Biotechnology, 11, 92. https://doi.org/10.1186/1472-6750-11-92

Li, J., Milbury, C. A., Li, C., & Makrigiorgos, G. M. (2009). Two-round coamplification at lower denaturation temperature-PCR (COLD-PCR)-based sanger sequencing identifies a novel spectrum of low-level mutations in lung adenocarcinoma. Human Mutation, 30(11), 1583–1590. https://doi.org/10.1002/humu.21112

Li, W., Qiu, T., Guo, L., & Ying, J. (2017). Major challenges related to tumor biological characteristics in accurate mutation detection of colorectal cancer by next-generation sequencing. Cancer Letters, 410, 92–99. https://doi.org/10.1016/j.canlet.2017.09.014

Li, W., Qiu, T., Ling, Y., Gao, S., & Ying, J. (2018). Subjecting appropriate lung adenocarcinoma samples to next-generation sequencing-based molecular testing: challenges and possible solutions. Molecular Oncology, 12(5), 677–689. https://doi.org/10.1002/1878-0261.12190

Lian, D.-S., Chen, X.-Y., Zeng, H.-S., & Wang, Y.-Y. (2021). Capillary electrophoresis based on nucleic acid analysis for diagnosing inherited diseases. Clinical Chemistry and Laboratory Medicine (CCLM), 59(2), 249–266. https://doi.org/doi:10.1515/cclm-2020-0186

Lima, Z. S., Ghadamzadeh, M., Arashloo, F. T., Amjad, G., Ebadi, M. R., & Younesi, L. (2019). Recent advances of therapeutic targets based on the molecular signature in breast cancer: genetic mutations and implications for current treatment paradigms. Journal of Hematology & Oncology, 12(1), 38. https://doi.org/10.1186/s13045-019-0725-6

Lindblad-Toh, K., Winchester, E., Daly, M. J., Wang, D. G., Hirschhorn, J. N., Laviolette, J.-P., Ardlie, K., Reich, D. E., Robinson, E., Sklar, P., Shah, N., Thomas, D., Fan, J.-B., Gingeras, T., Warrington, J., Patil, N., Hudson, T. J., & Lander, E. S. (2000). Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. Nature Genetics, 24(4), 381–386. https://doi.org/10.1038/74215

Lissa, D., & Robles, A. I. (2019). Comprehensive genomic analysis of circulating tumor DNA for patients with advanced non-small cell lung cancer. Annals of Translational Medicine, 7(5), 80. https://doi.org/10.21037/atm.2018.12.57

Liu, J., Huang, S., Sun, M., Liu, S., Liu, Y., Wang, W., Zhang, X., Wang, H., & Hua, W. (2012). An improved allele-specific PCR primer design method for SNP marker analysis and its application. Plant Methods, 8(1), 34. https://doi.org/10.1186/1746-4811-8-34

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of Next-Generation Sequencing Systems. Journal of Biomedicine and Biotechnology, 2012, 251364. https://doi.org/10.1155/2012/251364

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of Next-Generation Sequencing Systems. Journal of Biomedicine and Biotechnology, 2012, 251364. https://doi.org/10.1155/2012/251364

Liu, X., Lang, J., Li, S., Wang, Y., Peng, L., Wang, W., Han, Y., Qi, C., Song, L., Yang, S., Zhang, K., Zang, G., Pei, H., Lu, Q., Peng, Y., Xi, S., Wang, W., Yuan, D., Bing, P., … Tian, G. (2020). Fragment Enrichment of Circulating Tumor DNA With Low-Frequency Mutations. Frontiers in Genetics, 11, 147. https://doi.org/10.3389/fgene.2020.00147

Liu, Y., Sethi, N. S., Hinoue, T., Schneider, B. G., Cherniack, A. D., Sanchez-Vega, F., Seoane, J. A., Farshidfar, F., Bowlby, R., Islam, M., Kim, J., Chatila, W., Akbani, R., Kanchi, R. S., Rabkin, C. S., Willis, J. E., Wang, K. K., McCall, S. J., Mishra, L., … Laird, P. W. (2018). Comparative Molecular Analysis of Gastrointestinal Adenocarcinomas. Cancer Cell, 33(4), 721-735.e8. https://doi.org/10.1016/j.ccell.2018.03.010

Liu, Y.-P., Wu, H.-Y., Yang, X., Xu, H.-Q., Chen, D., Huang, Q., & Fu, W.-L. (2014). Diagnostic accuracy of high resolution melting analysis for detection of KRAS mutations: a systematic review and meta-analysis. Scientific Reports, 4(1), 7521. https://doi.org/10.1038/srep07521

Lo, S. S., Mumby, P. B., Norton, J., Rychlik, K., Smerage, J., Kash, J., Chew, H. K., Gaynor, E. R., Hayes, D. F., Epstein, A., & Albain, K. S. (2010). Prospective Multicenter Study of the Impact of the 21-Gene Recurrence Score Assay on Medical Oncologist and Patient Adjuvant Breast Cancer Treatment Selection. Journal of Clinical Oncology, 28(10), 1671–1676. https://doi.org/10.1200/JCO.2008.20.2119

Loncaster, J., Armstrong, A., Howell, S., Wilson, G., Welch, R., Chittalia, A., Valentine, W. J., & Bundred, N. J. (2017). Impact of Oncotype DX breast Recurrence Score testing on adjuvant chemotherapy use in early breast cancer: Real world experience in Greater Manchester, UK. European Journal of Surgical Oncology: The Journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology, 43(5), 931–937. https://doi.org/10.1016/j.ejso.2016.12.010

Long, F., Kong, M., Wu, S., Zhang, W., Liao, Q., Peng, Z., Nan, L., Liu, Y., Wang, M., He, C., Wu, Y., Lu, X., & Kang, M. (2019). Development and validation of an advanced fragment analysis-based assay for the detection of 22 pathogens in the cerebrospinal fluid of patients with meningitis and encephalitis. Journal of Clinical Laboratory Analysis, 33(3), e22707. https://doi.org/https://doi.org/10.1002/jcla.22707

Losi, L., Fonda, S., Saponaro, S., Chelbi, S. T., Lancellotti, C., Gozzi, G., Alberti, L., Fabbiani, L., Botticelli, L., & Benhattar, J. (2018). Distinct DNA Methylation Profiles in Ovarian Tumors: Opportunities for Novel Biomarkers. In International Journal of Molecular Sciences, 19(6). https://doi.org/10.3390/ijms19061559

Lu, Y., Soong, T. D., & Elemento, O. (2013). A Novel Approach for Characterising Microsatellite Instability in Cancer Cells. PLOS ONE, 8(5), e63056. https://doi.org/10.1371/journal.pone.0063056

Luchini, C., Bibeau, F., Ligtenberg, M. J. L., Singh, N., Nottegar, A., Bosse, T., Miller, R., Riaz, N., Douillard, J.-Y., Andre, F., & Scarpa, A. (2019). ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. Annals of Oncology: Official Journal of the European Society for Medical Oncology, 30(8), 1232–1243. https://doi.org/10.1093/annonc/mdz116

Ludyga, N., Grünwald, B., Azimzadeh, O., Englert, S., Höfler, H., Tapio, S., & Aubele, M. (2012). Nucleic acids from long-term preserved FFPE tissues are suitable for downstream analyses. Virchows Archiv, 460(2), 131–140. https://doi.org/10.1007/s00428-011-1184-9

Malila, N., Taylor, P. R., Virtanen, M. J., Korhonen, P., Huttunen, J. K., Albanes, D., & Virtamo, J. (2002). Effects of alpha-tocopherol and beta-carotene supplementation on gastric cancer incidence in male smokers (ATBC Study, Finland). Cancer Causes & Control: CCC, 13(7), 617–623. https://doi.org/10.1023/a:1019556227014

Malone, E. R., Oliva, M., Sabatini, P. J. B., Stockley, T. L., & Siu, L. L. (2020). Molecular profiling for precision cancer therapies. Genome Medicine, 12(1), 8. https://doi.org/10.1186/s13073-019-0703-1

Mancini, I., Pinzani, P., Simi, L., Brandslund, I., Vermeersch, P., Di Resta, C., Schwab, M., Marc, J., van Schaik, R., Pazzagli, M., & (WG-PLM), =European Federation of Clinical Chemistry and Laboratory Medicine (EFLM)–European Society of Pharmacogenomics and Theranostics (ESPT) joint Working Group Personalised Laboratory Medicine. (2015). Implementation of a companion diagnostic in the clinical laboratory: the BRAF example in melanoma. Clinica Chimica Acta; International Journal of Clinical Chemistry, 439, 128–136. https://doi.org/10.1016/j.cca.2014.10.020

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. Journal of Thoracic Oncology: Official Publication of the International Association for the Study of Lung Cancer, 5(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Manfredi, S., Bouvier, A. M., Lepage, C., Hatem, C., Dancourt, V., & Faivre, J. (2006). Incidence and patterns of recurrence after resection for cure of colonic cancer in a well-defined population. British Journal of Surgery, 93(9), 1115–1122. https://doi.org/10.1002/bjs.5349

Markoulatos, P., Siafakas, N., & Moncany, M. (2002). Multiplex polymerase chain reaction: a practical approach. Journal of Clinical Laboratory Analysis, 16(1), 47–51. https://doi.org/10.1002/jcla.2058

Martinelli, E., Ciardiello, D., Martini, G., Troiani, T., Cardone, C., Vitiello, P. P., Normanno, N., Rachiglio, A. M., Maiello, E., Latiano, T., De Vita, F., & Ciardiello, F. (2020). Implementing anti-epidermal growth factor receptor (EGFR) therapy in metastatic colorectal cancer: challenges and future perspectives. Annals of Oncology, 31(1), 30–40. https://doi.org/https://doi.org/10.1016/j.annonc.2019.10.007

Maslov, D., & Bragg, J. (2019). Left Cervical Lymphadenopathy Presentation of Metastatic Colorectal Adenocarcinoma. The Ochsner Journal, 19(4), 410–412. https://doi.org/10.31486/toj.18.0121

Mattocks, C. J., Morris, M. A., Matthijs, G., Swinnen, E., Corveleyn, A., Dequeker, E., Müller, C. R., Pratt, V., Wallace, A., & Group, E. V. (2010). A standardised framework for the validation and verification of clinical molecular genetic tests. European Journal of Human Genetics: EJHG, 18(12), 1276–1288. https://doi.org/10.1038/ejhg.2010.101

Mayday, M. Y., Khan, L. M., Chow, E. D., Zinter, M. S., & DeRisi, J. L. (2019). Miniaturisation and optimisation of 384-well compatible RNA sequencing library preparation. PLOS ONE, 14(1), e0206194. https://doi.org/10.1371/journal.pone.0206194

McGranahan, N., & Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell, 168(4), 613–628. https://doi.org/10.1016/j.cell.2017.01.018

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. Biochemia Medica, 22(3), 276–282.

McVeigh, T. P., & Kerin, M. J. (2017). Clinical use of the Oncotype DX genomic test to guide treatment decisions for patients with invasive breast cancer. Breast Cancer (Dove Medical Press), 9, 393–400. https://doi.org/10.2147/BCTT.S109847

Meier, B., Volkova, N. V., Hong, Y., Schofield, P., Campbell, P. J., Gerstung, M., & Gartner, A. (2018). Mutational signatures of DNA mismatch repair deficiency in C. elegans and human cancers. Genome Research, 28(5), 666–675. https://doi.org/10.1101/gr.226845.117

Mészáros, B., Zeke, A., Reményi, A., Simon, I., & Dosztányi, Z. (2016). Systematic analysis of somatic mutations driving cancer: uncovering functional protein regions in disease development. Biology Direct, 11(1), 23. https://doi.org/10.1186/s13062-016-0125-6

Moreira, L., Balaguer, F., Lindor, N., de la Chapelle, A., Hampel, H., Aaltonen, L. A., Hopper, J. L., Le Marchand, L., Gallinger, S., Newcomb, P. A., Haile, R., Thibodeau, S. N., Gunawardena, S., Jenkins, M. A., Buchanan, D. D., Potter, J. D., Baron, J. A., Ahnen, D. J., Moreno, V., … Consortium, E. (2012). Identification of Lynch syndrome among patients with colorectal cancer. JAMA, 308(15), 1555–1565. https://doi.org/10.1001/jama.2012.13088

Moreno, C. C., Mittal, P. K., Sullivan, P. S., Rutherford, R., Staley, C. A., Cardona, K., Hawk, N. N., Dixon, W. T., Kitajima, H. D., Kang, J., Small, W. C., Oshinski, J., & Votaw, J. R. (2016). Colorectal Cancer Initial Diagnosis: Screening Colonoscopy, Diagnostic Colonoscopy, or Emergent Surgery, and Tumor Stage and Size at Initial Presentation. Clinical Colorectal Cancer, 15(1), 67–73. https://doi.org/https://doi.org/10.1016/j.clcc.2015.07.004

Morihiro, T., Kuroda, S., Kanaya, N., Kakiuchi, Y., Kubota, T., Aoyama, K., Tanaka, T., Kikuchi, S., Nagasaka, T., Nishizaki, M., Kagawa, S., Tazawa, H., & Fujiwara, T. (2019). PD-L1 expression combined with microsatellite instability/CD8+ tumor infiltrating lymphocytes as a useful prognostic biomarker in gastric cancer. Scientific Reports, 9(1), 4633. https://doi.org/10.1038/s41598-019-41177-2

Mouliere, F., & Rosenfeld, N. (2015). Circulating tumor-derived DNA is shorter than somatic DNA in plasma. Proceedings of the National Academy of Sciences, 112(11): 3178 – 3179. https://doi.org/10.1073/pnas.1501321112

Mu, W., Lu, H.-M., Chen, J., Li, S., & Elliott, A. M. (2016). Sanger Confirmation Is Required to Achieve Optimal Sensitivity and Specificity in Next-Generation Sequencing Panel Testing. The Journal of Molecular Diagnostics, 18(6), 923–932. https://doi.org/10.1016/j.jmoldx.2016.07.006

Murphy, K. M., Zhang, S., Geiger, T., Hafez, M. J., Bacher, J., Berg, K. D., & Eshleman, J. R. (2006). Comparison of the Microsatellite Instability Analysis System and the Bethesda Panel for the Determination of Microsatellite Instability in Colorectal Cancers. The Journal of Molecular Diagnostics, 8(3): 305–311. https://doi.org/https://doi.org/10.2353/jmoldx.2006.050092

Murtuza, A., Bulbul, A., Shen, J. P., Keshavarzian, P., Woodward, B. D., Lopez-Diaz, F. J., Lippman, S. M., & Husain, H. (2019). Novel Third-Generation EGFR Tyrosine Kinase Inhibitors and Strategies to Overcome Therapeutic Resistance in Lung Cancer. Cancer Research, 79(4), 689–698. https://doi.org/10.1158/0008-5472.CAN-18-1281

Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., Kovar, C. L., Lewis, L. R., Morgan, M. B., Newsham, I. F., Reid, J. G., Santibanez, J., Shinbrot, E., Trevino, L. R., Wu, Y.-Q., Wang, M., Gunaratne, P., Donehower, L. A., Creighton, C. J., … group, T. source sites and disease working. (2012). Comprehensive molecular characterisation of human colon and rectal cancer. Nature, 487(7407), 330–337. https://doi.org/10.1038/nature11252

Nabors, L. B. (2020). Management of Gliomas: Individualised Treatment Options. Journal of the National Comprehensive Cancer Network J Natl Compr Canc Netw, 18(7.5): 985–988. https://doi.org/10.6004/jnccn.2020.5008

Nadauld, L. D., & Ford, J. M. (2013). Molecular profiling of gastric cancer: toward personalised cancer medicine. Journal of clinical oncology: official journal of the American Society of Clinical Oncology, 31(7):838–839. https://doi.org/10.1200/JCO.2012.47.1714

National Cancer Institute. (2021). Targeted Cancer Therapies Fact Sheet. https://www.cancer.gov/about-cancer/treatment/types/targeted-therapies/targeted-therapies-fact-sheet

National Comprehensive Cancer Network. (2014). Clinical practice guidelines in oncology: gastric cancer. In Cancer/Radiothérapie,14 (S1). https://doi.org/10.1016/S1278-3218(10)70011-1

National Comprehensive Cancer Network. (2015). Clinical practice guidelines in oncology: breast cancer screening and diagnosis. In NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®).

National Comprehensive Cancer Network. (2020). Clinical practice guidelines in oncology: colorectal cancer screening (version 2.2020). http://www.nccn.org/professionals/physician_gls/pdf/colorectal_screening.pdf

National Comprehensive Cancer Network. (2021a). Clinical practice guidelines in oncology: breast cancer (version 2.2021). https://www.nccn.org/professionals/physician_gls/pdf/breast.pdf

National Comprehensive Cancer Network. (2021b). Clinical practice guidelines in oncology: colon cancer (version 2.2021). https://www.nccn.org/professionals/physician_gls/pdf/colon.pdf

National Comprehensive Cancer Network. (2021c). Clinical practice guidelines in oncology: cutaneous melanoma.

National Comprehensive Cancer Network. (2021d). Clinical practice guidelines in oncology: gastric cancer (version 2.2021). https://www.nccn.org/professionals/physician_gls/pdf/gastric.pdf

National Comprehensive Cancer Network. (2021e). Clinical practice guidelines in oncology: gastrointestinal stromal tumours (version 1.2021). https://www.nccn.org/professionals/physician_gls/pdf/gist.pdf

National Comprehensive Cancer Network. (2021f). Clinical practice guidelines in oncology: non-small cell lung cancer (version 4.2021).

National Comprehensive Cancer Network. (2021g). Clinical practice guidelines in oncology: rectal cancer (version 2.2021). https://www.nccn.org/professionals/physician_gls/pdf/rectal.pdf

National Comprehensive Cancer Network. (2021h). Clinical practice guidelines in oncology - melanoma: cutaneous (version 2.2021). https://www.nccn.org/professionals/physician_gls/pdf/cutaneous_melanoma.pdf

National Comprehensive Cancer Network. (2022). Clinical practice guidelines in oncology: esophageal and esophagogastric junction cancers. NCCN Clinical Practice Guidelines in Oncology. https://www.nccn.org/professionals/physician_gls/pdf/esophageal.pdf

National Institute for Health and Clinical Excellence. (2010). Trastuzumab for the treatment of HER2-positive metastatic gastric cancer. NICE Technology Appraisal Guidance [TA208]. https://www.nice.org.uk/guidance/ta208.

National Institute for Health and Care Excellence. (2012). Cetuximab, bevacizumab and panitumumab for the treatment of metastatic colorectal cancer after first- line chemotherapy: Cetuximab (monotherapy or combination chemotherapy), bevacizumab (in combination with non-oxaliplatin chemotherapy) and panitumumab (m. NICE Technology Appraisal Guidance [TA242]. https://www.nice.org.uk/guidance/ta242/resources/cetuximab-bevacizumab-and-panitumumab-for-the-treatment-of-metastatic-colorectal-cancer-after-firstline-chemotherapy-cetuximab-monotherapy-or-combination-chemotherapy-bevacizumab-in-combination-with--pdf-82600427700421

National Institute for Health and Care Excellence. (2013a). Gene expression profiling and expanded immunohistochemistry tests for guiding adjuvant chemotherapy decisions in early breast cancer management: MammaPrint, Oncotype DX, IHC4 and Mammostrat NICE Diagnostics guidance (Issue February).

National Institute for Health and Care Excellence. (2013b). EGFR-TK mutation testing in adults with locally advanced or metastatic non-small-cell lung cancer. NICE Diagnostic Guidance [DG9]. https://www.nice.org.uk/guidance/dg9

National Institute for Health and Care Excellence. (2017a). Quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care Diagnostics guidance. NICE Diagnostic Guidance [DG30]. https://www.nice.org.uk/guidance/dg30/resources/quantitative-faecal-immunochemical-tests-to-guide-referral-for-colorectal-cancer-in-primary-care-pdf-1053744003781

National Institute for Health and Care Excellence. (2017b). Molecular testing for Lynch syndrome in people with colorectal cancer. NICE Diagnostic Guidance [DG27]. https://www.nice.org.uk/guidance/dg27/documents/diagnostics-assessment-report

National Institute for Health and Care Excellence. (2017c). Atezolizumab for untreated PD-L1-positive locally advanced or metastatic urothelial cancer when cisplatin is unsuitable. NICE Technology Appraisal Guidance [TA429]. https://www.nice.org.uk/guidance/ta492

National Institute for Health and Care Excellence. (2017d). Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer. NICE Technology Appraisal Guidance [TA447]. https://www.nice.org.uk/guidance/ta447

National Institute for Health and Care Excellence. (2017e). Pembrolizumab for treating PD-L1-positive non-small-cell lung cancer after chemotherapy. NICE Technology Appraisal Guidance [TA428]. https://www.nice.org.uk/guidance/ta428

National Institute for Health and Care Excellence. (2018a). Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer. NICE Technology Appraisal Guidance [TA531]. https://www.nice.org.uk/guidance/ta531

National Institute for Health and Care Excellence. (2018b). Pembrolizumab for untreated PD-L1-positive locally advanced or metastatic urothelial cancer when cisplatin is unsuitable. NICE Technology Appraisal Guidance [TA552]. https://www.nice.org.uk/guidance/ta522

National Institute for Health and Care Excellence. (2020). Atezolizumab with nab- paclitaxel for untreated PD- L1-positive, locally advanced or metastatic, triple-negative breast cancer. NICE Technology Appraisal Guidance [TA639]. https://www.nice.org.uk/guidance/ta639

National Institute for Health and Care Excellence. (2021a). Pembrolizumab for untreated metastatic colorectal cancer with high microsatellite instability or mismatch repair deficiency. NICE Technology Appraisal Guidance [TA709]. https://www.nice.org.uk/guidance/ta709/resources

National Institute for Health and Care Excellence. (2021b). Pembrolizumab for untreated locally advanced or metastatic urothelial cancer when cisplatin is unsuitable. NICE Technology Appraisal Guidance [TA674]. https://www.nice.org.uk/guidance/ta674

National Institute for Health and Clinical Excellence. (2021c). Atezolizumab for untreated PD-L1-positive advanced urothelial cancer when cisplatin is unsuitable. NICE Technology Appraisal Guidance [TA739]. https://www.nice.org.uk/guidance/ta739

National Institute for Health and Care Excellence. (2021d). Pembrolizumab with platinum- and chemotherapy for untreated advanced oesophageal and gastro-oesophageal junction cancer. NICE Technology Appraisal Guidance [TA737]. https://www.nice.org.uk/guidance/ta737/resources/pembrolizumab-with-platinum-and-fluoropyrimidinebased-chemotherapy-for-untreated-advanced-oesophageal-and-gastrooesophageal-junction-cancer-pdf-82611259544005

National Institute for Health and Care Excellence. (2021e). Encorafenib plus cetuximab for previously treated BRAF metastatic colorectal cancer. NICE Technology Appraisal Guidance [TA668]. https://www.nice.org.uk/guidance/ta668/resources/encorafenib-plus-cetuximab-for-previously-treated-braf-v600e-mutationpositive-metastatic-colorectal-cancer-pdf-82609265839813

National Institute for Health and Care Excellence. (2022). Dostarlimab for previously treated advanced or recurrent endometrial cancer with high microsatellite instability or mismatch repair deficiency. NICE Technology Appraisal Guidance [TA779]. https://www.nice.org.uk/guidance/ta779

Nestarenkaite, A., Fadhil, W., Rasmusson, A., Susanti, S., Hadjimichael, E., Laurinaviciene, A., Ilyas, M., & Laurinavicius, A. (2020). Immuno-Interface Score to Predict Outcome in Colorectal Cancer Independent of Microsatellite Instability Status. Cancers, 12(10), 2902. https://doi.org/10.3390/cancers12102902

NHS England. (2021a). Genomic Laboratory Hubs. NHS England Genomics. https://www.england.nhs.uk/genomics/genomic-laboratory-hubs/

NHS England. (2021b). National Genomic Test Directory. NHS England Publications. https://www.england.nhs.uk/publication/national-genomic-test-directories/

Ni Huang, M., McPherson, J. R., Cutcutache, I., Teh, B. T., Tan, P., & Rozen, S. G. (2015). MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations. Scientific Reports, 5(1), 13321. https://doi.org/10.1038/srep13321

Niu, B., Ye, K., Zhang, Q., Lu, C., Xie, M., McLellan, M. D., Wendl, M. C., & Ding, L. (2014). MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics (Oxford, England), 30(7), 1015–1016. https://doi.org/10.1093/bioinformatics/btt755

Normanno, N., Rachiglio, A. M., Lambiase, M., Martinelli, E., Fenizia, F., Esposito, C., Roma, C., Troiani, T., Rizzi, D., Tatangelo, F., Botti, G., Maiello, E., Colucci, G., & Ciardiello, F. (2015). Heterogeneity of KRAS, NRAS, BRAF and PIK3CA mutations in metastatic colorectal cancer and potential effects on therapy in the CAPRI GOIM trial. Annals of Oncology: Official Journal of the European Society for Medical Oncology, 26(8), 1710–1714. https://doi.org/10.1093/annonc/mdv176

Osumi, H., Shinozaki, E., & Yamaguchi, K. (2020). Circulating Tumor DNA as a Novel Biomarker Optimising Chemotherapy for Colorectal Cancer. In Cancers, 12(6). https://doi.org/10.3390/cancers12061566

Overman, M. J., Lonardi, S., Wong, K. Y. M., Lenz, H.-J., Gelsomino, F., Aglietta, M., Morse, M. A., Van Cutsem, E., McDermott, R., Hill, A., Sawyer, M. B., Hendlisz, A., Neyns, B., Svrcek, M., Moss, R. A., Ledeine, J.-M., Cao, Z. A., Kamble, S., Kopetz, S., & André, T. (2018). Durable Clinical Benefit With Nivolumab Plus Ipilimumab in DNA Mismatch Repair–Deficient/Microsatellite Instability–High Metastatic Colorectal Cancer. Journal of Clinical Oncology, 36(8), 773–779. https://doi.org/10.1200/JCO.2017.76.9901

Overman, M. J., McDermott, R., Leach, J. L., Lonardi, S., Lenz, H.-J., Morse, M. A., Desai, J., Hill, A., Axelson, M., Moss, R. A., Goldberg, M. V, Cao, Z. A., Ledeine, J.-M., Maglinte, G. A., Kopetz, S., & André, T. (2017). Nivolumab in patients with metastatic DNA mismatch repair-deficient or microsatellite instability-high colorectal cancer (CheckMate 142): an open-label, multicentre, phase 2 study. The Lancet. Oncology, 18(9), 1182–1191. https://doi.org/10.1016/S1470-2045(17)30422-9

Pabla, S., Andreas, J., Lenzo, F. L., Burgher, B., Hagen, J., Giamo, V., Nesline, M. K., Wang, Y., Gardner, M., Conroy, J. M., Papanicolau-Sengos, A., Morrison, C., & Glenn, S. T. (2019). Development and analytical validation of a next-generation sequencing based microsatellite instability (MSI) assay. Oncotarget; Vol 10, No 50. https://www.oncotarget.com/article/27142/text/

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., Hiller, W., Fisher, E. R., Wickerham, D. L., Bryant, J., & Wolmark,

N. (2004). A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. New England Journal of Medicine, 351(27), 2817–2826. https://doi.org/10.1056/NEJMoa041588

Paik, S., Tang, G., Shak, S., Kim, C., Baker, J., Kim, W., Cronin, M., Baehner, F. L., Watson, D., Bryant, J., Costantino, J. P., Geyer, C. E., Wickerham, D. L., & Wolmark, N. (2006). Gene Expression and Benefit of Chemotherapy in Women With Node-Negative, Estrogen Receptor–Positive Breast Cancer. Journal of Clinical Oncology, 24(23), 3726–3734. https://doi.org/10.1200/JCO.2005.04.7985

Palais, R. A., Liew, M. A., & Wittwer, C. T. (2005). Quantitative heteroduplex analysis for single nucleotide polymorphism genotyping. Analytical Biochemistry, 346(1), 167–175. https://doi.org/https://doi.org/10.1016/j.ab.2005.08.010

Pancione, M., Remo, A., & Colantuoni, V. (2012). Genetic and epigenetic events generate multiple pathways in colorectal cancer progression. Pathology Research International, 2012, 509348. https://doi.org/10.1155/2012/509348

Pandith, A. A., Qasim, I., Zahoor, W., Shah, P., Bhat, A. R., Sanadhya, D., Shah, Z. A., & Naikoo, N. A. (2018). Concordant association validates MGMT methylation and protein expression as favorable prognostic factors in glioma patients on alkylating chemotherapy (Temozolomide). In Scientific reports (Vol. 8, Issue 1, p. 6704). https://doi.org/10.1038/s41598-018-25169-2

Patel, H., Yacoub, N., Mishra, R., White, A., Long, Y., Alanazi, S., & Garrett, J. T. (2020). Current Advances in the Treatment of BRAF-Mutant Melanoma. Cancers, 12(2), 482. https://doi.org/10.3390/cancers12020482

Pawlik, T. M., Raut, C. P., & Rodriguez-Bigas, M. A. (2004). Colorectal carcinogenesis: MSI-H versus MSI-L. Disease Markers, 20(4–5), 199–206. https://doi.org/10.1155/2004/368680

Peng, D.-F., Sugihara, H., Mukaisho, K., Tsubosa, Y., & Hattori, T. (2003). Alterations of chromosomal copy number during progression of diffuse-type gastric carcinomas: metaphase- and array-based comparative genomic hybridisation analyses of multiple samples from individual tumours. The Journal of Pathology, 201(3), 439–450. https://doi.org/https://doi.org/10.1002/path.1459

Pereyra, S., Velazquez, T., Bertoni, B., & Sapiro, R. (2012). Rapid multiplex high resolution melting method to analyse inflammatory related SNPs in preterm birth. BMC Research Notes, 5, 69. https://doi.org/10.1186/1756-0500-5-69

Perez-Villamil, B., Romera-Lopez, A., Hernandez-Prieto, S., Lopez-Campos, G., Calles, A., Lopez-Asenjo, J.-A., Sanz-Ortega, J., Fernandez-Perez, C., Sastre, J., Alfonso, R., Caldes, T., Martin-Sanchez, F., & Diaz-Rubio, E. (2012). Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior. BMC Cancer, 12, 260. https://doi.org/10.1186/1471-2407-12-260

Peters, U., Jiao, S., Schumacher, F. R., Hutter, C. M., Aragaki, A. K., Baron, J. A., Berndt, S. I., Bézieau, S., Brenner, H., Butterbach, K., Caan, B. J., Campbell, P. T., Carlson, C. S., Casey, G., Chan, A. T., Chang-Claude, J., Chanock, S. J., Chen, L. S., Coetzee, G. A., … Consortium, C. C. F. R. and the G. and E. of C. C. (2013). Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. Gastroenterology, 144(4), 799-807.e24. https://doi.org/10.1053/j.gastro.2012.12.020

Pino, M. S., & Chung, D. C. (2010). The chromosomal instability pathway in colon cancer. Gastroenterology, 138(6), 2059–2072. https://doi.org/10.1053/j.gastro.2009.12.065

Plummer, M., Vivas, J., Lopez, G., Bravo, J. C., Peraza, S., Carillo, E., Cano, E., Castro, D., Andrade, O., Sánchez, V., Garcia, R., Buiatti, E., Aebischer, C., Franceschi, S., Oliver, W., & Muñoz, N. (2007). Chemoprevention of precancerous gastric lesions with antioxidant vitamin supplementation: a randomised trial in a high-risk population. Journal of the National Cancer Institute, 99(2), 137–146. https://doi.org/10.1093/jnci/djk017

Polom, K., Marrelli, D., Smyth, E. C., Voglino, C., Roviello, G., Pascale, V., Varas, J., Vindigni, C., & Roviello, F. (2018). The Role of Microsatellite Instability in Positive Margin Gastric Cancer Patients. Surgical Innovation, 25(2), 99–104. https://doi.org/10.1177/1553350617751461

Popat, S., Hubner, R., & Houlston, R. S. (2005). Systematic Review of Microsatellite Instability and Colorectal Cancer Prognosis. Journal of Clinical Oncology, 23(3), 609–618. https://doi.org/10.1200/JCO.2005.01.086

Prence, E. M. (1999). A Practical Guide for the Validation of Genetic Tests. Genetic Testing, 3(2), 201–205. https://doi.org/10.1089/gte.1999.3.201

Presley, C. J., Tang, D., Soulos, P. R., Chiang, A. C., Longtine, J. A., Adelson, K. B., Herbst, R. S., Zhu, W., Nussbaum, N. C., Sorg, R. A., Agarwala, V., Abernethy, A. P., & Gross, C. P. (2018). Association of Broad-Based Genomic Sequencing With Survival Among Patients With Advanced Non-Small Cell Lung Cancer in the Community Oncology Setting. JAMA, 320(5), 469–477. https://doi.org/10.1001/jama.2018.9824

Press, M. O., Carlson, K. D., & Queitsch, C. (2014). The overdue promise of short tandem repeat variation for heritability. Trends in Genetics: TIG, 30(11), 504–512. https://doi.org/10.1016/j.tig.2014.07.008

Provenzale, D., Gupta, S., Ahnen, D. J., Bray, T., Cannon, J. A., Cooper, G., David, D. S., Early, D. S., Erwin, D., Ford, J. M., Giardiello, F. M., Grady, W., Halverson, A. L., Hamilton, S. R., Hampel, H., Ismail, M. K., Klapman, J. B., Larson, D. W., Lazenby, A. J., … Darlow, S. (2016). Genetic/Familial High-Risk Assessment: Colorectal Version 1.2016, NCCN Clinical Practice Guidelines in Oncology. Journal of the National Comprehensive Cancer Network: JNCCN, 14(8), 1010–1030. https://doi.org/10.6004/jnccn.2016.0108

Public Health England. (2017). Evaluations, validations and verifications of diagnostic tests (No. Q1). UK Standards for Microbiology Investigations. https://www.gov.uk/uk-standards-for-microbiology-investigations-smi-quality-andconsistency-in-clinical-laboratories

Qu, W., Zhou, Y., Zhang, Y., Lu, Y., Wang, X., Zhao, D., Yang, Y., & Zhang, C. (2012). MFEprimer-2.0: a fast thermodynamics-based program for checking PCR primer specificity. Nucleic Acids Research, 40(Web Server issue), W205–W208. https://doi.org/10.1093/nar/gks552

Racine, J. (2000). The Cygwin tools: a GNU toolkit for Windows. Journal of Applied Econometrics, 15(3), 331–341. https://doi.org/10.1002/1099-1255(200005/06)15:3<331::aid-jae558>3.0.co;2-g

Rajendra, B. (2012). DNA Representation. DNA Sequencing - Methods and Applications. https://doi.org/10.5772/36956

Ratti, M., Lampis, A., Hahne, J. C., Passalacqua, R., & Valeri, N. (2018). Microsatellite instability in gastric cancer: molecular bases, clinical perspectives, and new treatment approaches. Cellular and Molecular Life Sciences: CMLS, 75(22), 4151–4162. https://doi.org/10.1007/s00018-018-2906-9

Rawla, P., & Barsouk, A. (2019). Epidemiology of gastric cancer: global trends, risk factors and prevention. Przeglad Gastroenterologiczny, 14(1), 26–38. https://doi.org/10.5114/pg.2018.80001

Reed, G. H., Kent, J. O., & Wittwer, C. T. (2007). High-resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics, 8(6), 597–608. https://doi.org/10.2217/14622416.8.6.597

Rees, V. (2021, April 30). UK Government to fund £37 million into genomics projects. Drug Target Review. https://www.drugtargetreview.com/news/90361/uk-government-to-fund-37-million-into-genomics-projects/

Rehm, H. L., Bale, S. J., Bayrak-Toydemir, P., Berg, J. S., Brown, K. K., Deignan, J. L., Friez, M. J., Funke, B. H., Hegde, M. R., Lyon, E., & Commitee, W. G. of the A. C. of M. G. and G. L. Q. A. (2013). ACMG clinical laboratory standards for next-generation sequencing. Genetics in Medicine: Official Journal of the American College of Medical Genetics, 15(9), 733–747. https://doi.org/10.1038/gim.2013.92

Ribic, C. M., Sargent, D. J., Moore, M. J., Thibodeau, S. N., French, A. J., Goldberg, R. M., Hamilton, S. R., Laurent-Puig, P., Gryfe, R., Shepherd, L. E., Tu, D., Redston, M., & Gallinger, S. (2003). Tumor Microsatellite-Instability Status as a Predictor of Benefit from Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. New England Journal of Medicine, 349(3), 247–257. https://doi.org/10.1056/NEJMoa022289

Richman, S. D., Fairley, J., Butler, R., & Deans, Z. C. (2017). RAS screening in colorectal cancer: a comprehensive analysis of the results from the UK NEQAS colorectal cancer external quality assurance schemes (2009–2016). Virchows Archiv, 471(6), 721–729. https://doi.org/10.1007/s00428-017-2162-7

Richman, S. D., Fairley, J., Hall, J. A., Nataraj, N., Bhide, M., Lau, A., Norman, K. L., & Deans, Z. C. (2018). Results of the UK NEQAS for Molecular Genetics reference sample

analysis. Journal of Clinical Pathology, 71(11), 989–994. https://doi.org/10.1136/jclinpath-2018-205277

Riihimäki, M., Hemminki, A., Sundquist, J., & Hemminki, K. (2016). Patterns of metastasis in colon and rectal cancer. Scientific Reports, 6, 29765. https://doi.org/10.1038/srep29765

Rutter, C. E., Yao, X., Mancini, B. R., Aminawung, J. A., Chagpar, A. B., Saglam, O., Hofstatter, E. W., Abu-Khalaf, M., Gross, C. P., & Evans, S. B. (2016). Influence of a 21-Gene Recurrence Score Assay on Chemotherapy Delivery in Breast Cancer. Clinical Breast Cancer, 16(1), 59–62. https://doi.org/10.1016/j.clbc.2015.09.008

Saddala, M. S., Lennikov, A., Bouras, A., & Huang, H. (2020). RNA-Seq reveals differential expression profiles and functional annotation of genes involved in retinal degeneration in Pde6c mutant Danio rerio. BMC Genomics, 21(1), 132. https://doi.org/10.1186/s12864-020-6550-z

Sahinn, I., Kunal, K. C., & Salem, M. E. (2022, February 2). Checkpoint Inhibitors for All or Only for High PD-L1 Expression: Exploring First-Line Therapy for Patients With Metastatic Gastroesophageal Malignancies. ASCO Daily News. https://dailynews.ascopubs.org/do/10.1200/ADN.22.200830/full/?utm_source=TrendMD&utm_medium=cpc&utm_campaign=ASCO_Daily_News_TrendMD_0

Salipante, S. J., Scroggins, S. M., Hampel, H. L., Turner, E. H., & Pritchard, C. C. (2014). Microsatellite Instability Detection by Next Generation Sequencing. Clinical Chemistry, 60(9), 1192–1199. https://doi.org/10.1373/clinchem.2014.223677

Sanchez, J. J., Børsting, C., Hallenberg, C., Buchard, A., Hernandez, A., & Morling, N. (2003). Multiplex PCR and minisequencing of SNPs—a model with 35 Y chromosome SNPs. Forensic Science International, 137(1), 74–84. https://doi.org/https://doi.org/10.1016/S0379-0738(03)00299-8

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America, 74(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Sargent, D. J., Marsoni, S., Monges, G., Thibodeau, S. N., Labianca, R., Hamilton, S. R., French, A. J., Kabat, B., Foster, N. R., Torri, V., Ribic, C., Grothey, A., Moore, M., Zaniboni, A., Seitz, J.-F., Sinicrope, F., & Gallinger, S. (2010). Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 28(20), 3219–3226. https://doi.org/10.1200/JCO.2009.27.1825

Satherley, L. K., Ye, L., Jiang, W. G., & Hargest, R. (2015). Tumour suppressor genes in gastric carcinogenesis: an up-to-date review of genetic and epigenetic changes in the development of gastric adenocarcinoma. Translational Gastrointestinal Cancer; Vol 4, No 4 (July 2015): Translational Gastrointestinal Cancer. https://tgc.amegroups.com/article/view/6997

Sato, F., & Meltzer, S. J. (2006). CpG island hypermethylation in progression of esophageal and gastric cancer. Cancer, 106(3), 483–493. https://doi.org/10.1002/cncr.21657

Saunders, N., Zambon, M., Sharp, I., Siddiqui, R., Bermingham, A., Ellis, J., Vipond, B., Sails, A., Moran-Gilad, J., Marsh, P., Guiver, M., & Division, H. P. A. M. S. (2013). Guidance on the development and validation of diagnostic tests that depend on nucleic acid amplification and detection. Journal of Clinical Virology: The Official Publication of the Pan American Society for Clinical Virology, 56(3), 260–270. https://doi.org/10.1016/j.jcv.2012.11.013

Schlötterer, C. (1998). Genome evolution: Are microsatellites really simple sequences? Current Biology, 8(4), R132–R134. https://doi.org/10.1016/S0960-9822(98)70989-3

Schreiber, E. H., Leong, H., Schneider, S. J., Jeff, M., George, W., Sangha, H., Lim, Y. P., Chan, E., Berosik, S., Storm, M., & Colburn, J. (2016). Abstract 5269: Minor Variant Finder: New software for detecting somatic mutations at low level in Sanger sequencing traces. Cancer Research, 76(14_Supplement), 5269. https://doi.org/10.1158/1538-7445.AM2016-5269

Seipp, M. T., Pattison, D., Durtschi, J. D., Jama, M., Voelkerding, K. V, & Wittwer, C. T. (2008). Quadruplex Genotyping of F5, F2, and MTHFR Variants in a Single Closed Tube by High-Resolution Amplicon Melting. Clinical Chemistry, 54(1), 108–115. https://doi.org/10.1373/clinchem.2007.097121

Senft, D., Leiserson, M. D. M., Ruppin, E., & Ronai, Z. A. (2017). Precision Oncology: The Road Ahead. Trends in Molecular Medicine, 23(10), 874–898. https://doi.org/10.1016/j.molmed.2017.08.003

Seth, R., Keeley, J., Abu-Ali, G., Crook, S., Jackson, D., & Ilyas, M. (2009). The putative tumour modifier gene ATP5A1 is not mutated in human colorectal cancer cell lines but expression levels correlate with TP53 mutations and chromosomal instability. Journal of Clinical Pathology, 62(7), 598 LP – 603. https://doi.org/10.1136/jcp.2009.064436

Setia, N., Agoston, A. T., Han, H. S., Mullen, J. T., Duda, D. G., Clark, J. W., Deshpande, V., Mino-Kenudson, M., Srivastava, A., Lennerz, J. K., Hong, T. S., Kwak, E. L., & Lauwers, G. Y. (2016). A protein and mRNA expression-based classification of gastric cancer. Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc, 29(7), 772–784. https://doi.org/10.1038/modpathol.2016.55

Shendure, J., Balasubramanian, S., Church, G. M., Gilbert, W., Rogers, J., Schloss, J. A., & Waterston, R. H. (2019). Publisher Correction: DNA sequencing at 40: past, present and future. In Nature (Vol. 568, Issue 7752, p. E11). https://doi.org/10.1038/s41586-019-1120-8

Shi, J., Yao, D., Liu, W., Wang, N., Lv, H., Zhang, G., Ji, M., Xu, L., He, N., Shi, B., & Hou, P. (2012). Highly frequent PIK3CA amplification is associated with poor prognosis in gastric cancer. BMC Cancer, 12(1), 50. https://doi.org/10.1186/1471-2407-12-50

Shia, J. (2008). Immunohistochemistry versus Microsatellite Instability Testing For Screening Colorectal Cancer Patients at Risk For Hereditary Nonpolyposis Colorectal Cancer Syndrome: Part I. The Utility of Immunohistochemistry. The Journal of Molecular Diagnostics, 10(4), 293–300. https://doi.org/https://doi.org/10.2353/jmoldx.2008.080031

Shia, J. (2008). Immunohistochemistry versus Microsatellite Instability Testing For Screening Colorectal Cancer Patients at Risk For Hereditary Nonpolyposis Colorectal Cancer Syndrome: Part I. The Utility of Immunohistochemistry. The Journal of Molecular Diagnostics, 10(4), 293–300. https://doi.org/https://doi.org/10.2353/jmoldx.2008.080031

Shibata, C., Ogawa, H., Miura, K., Naitoh, T., Yamauchi, J., & Unno, M. (2013). Clinical characteristics of gastric cancer in patients with familial adenomatous polyposis. The Tohoku Journal of Experimental Medicine, 229(2), 143–146. https://doi.org/10.1620/tjem.229.143

Shitara, K., Van Cutsem, E., Bang, Y.-J., Fuchs, C., Wyrwicz, L., Lee, K.-W., Kudaba, I., Garrido, M., Chung, H. C., Lee, J., Castro, H. R., Mansoor, W., Braghiroli, M. I., Karaseva, N., Caglevic, C., Villanueva, L., Goekkurt, E., Satake, H., Enzinger, P., … Tabernero, J. (2020). Efficacy and Safety of Pembrolizumab or Pembrolizumab Plus Chemotherapy vs Chemotherapy Alone for Patients With First-line, Advanced Gastric Cancer: The KEYNOTE-062 Phase 3 Randomised Clinical Trial. JAMA Oncology, 6(10), 1571–1580. https://doi.org/10.1001/jamaoncol.2020.3370

Sikkema-Raddatz, B., Johansson, L. F., de Boer, E. N., Almomani, R., Boven, L. G., van den Berg, M. P., van Spaendonck-Zwarts, K. Y., van Tintelen, J. P., Sijmons, R. H., Jongbloed, J. D. H., & Sinke, R. J. (2013). Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics. Human Mutation, 34(7), 1035–1042. https://doi.org/10.1002/humu.22332

Simons, C. C. J. M., Hughes, L. A. E., Smits, K. M., Khalid-de Bakker, C. A., de Bruïne, A. P., Carvalho, B., Meijer, G. A., Schouten, L. J., van den Brandt, P. A., Weijenberg, M. P., & van Engeland, M. (2013). A novel classification of colorectal tumors based on microsatellite instability, the CpG island methylator phenotype and chromosomal instability: implications for prognosis. Annals of Oncology: Official Journal of the European Society for Medical Oncology, 24(8), 2048–2056. https://doi.org/10.1093/annonc/mdt076

Skierucha, M., Barud, W., Baraniak, J., & Krupski, W. (2018). Colovesical fistula as the initial manifestation of advanced colon cancer: A case report and review of literature. World Journal of Clinical Cases, 6(12), 538–541. https://doi.org/10.12998/wjcc.v6.i12.538

Skoulidis, F., Li, B. T., Dy, G. K., Price, T. J., Falchook, G. S., Wolf, J., Italiano, A., Schuler, M., Borghaei, H., Barlesi, F., Kato, T., Curioni-Fontecedro, A., Sacher, A., Spira, A., Ramalingam, S. S., Takahashi, T., Besse, B., Anderson, A., Ang, A., … Govindan, R. (2021). Sotorasib for Lung Cancers with KRAS p.G12C Mutation. New England Journal of Medicine. https://doi.org/10.1056/NEJMoa2103695

Smyth, E. C., Wotherspoon, A., Peckitt, C., Gonzalez, D., Hulkki-Wilson, S., Eltahir, Z., Fassan, M., Rugge, M., Valeri, N., Okines, A., Hewish, M., Allum, W., Stenning, S., Nankivell, M., Langley, R., & Cunningham, D. (2017). Mismatch Repair Deficiency, Microsatellite Instability, and Survival: An Exploratory Analysis of the Medical Research Council Adjuvant Gastric Infusional Chemotherapy (MAGIC) Trial. JAMA Oncology, 3(9), 1197–1203. https://doi.org/10.1001/jamaoncol.2016.6762

Sonay, B. T., Carvalho, T., Robinson, M. D., Greminger, M. P., Krützen, M., Comas, D., Highnam, G., Mittelman, D., Sharp, A., Marques-Bonet, T., & Wagner, A. (2015). Tandem repeat variation in human and great ape populations and its impact on gene expression divergence. Genome Research, 25(11), 1591–1599. https://doi.org/10.1101/gr.190868.115

Speir, M. L., Zweig, A. S., Rosenbloom, K. R., Raney, B. J., Paten, B., Nejad, P., Lee, B. T., Learned, K., Karolchik, D., Hinrichs, A. S., Heitner, S., Harte, R. A., Haeussler, M., Guruvadoo, L., Fujita, P. A., Eisenhart, C., Diekhans, M., Clawson, H., Casper, J., … Kent, W. J. (2016). The UCSC Genome Browser database: 2016 update. Nucleic Acids Research, 44(D1), D717–D725. https://doi.org/10.1093/nar/gkv1275

Srinivasan, M., Sedmak, D., & Jewell, S. (2002). Effect of fixatives and tissue processing on the content and integrity of nucleic acids. The American Journal of Pathology, 161(6), 1961–1971. https://doi.org/10.1016/S0002-9440(10)64472-0

Stoffel, E. M., & Kastrinos, F. (2014). Familial colorectal cancer, beyond Lynch syndrome. Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association, 12(7), 1059–1068. https://doi.org/10.1016/j.cgh.2013.08.015

Stothard, P. (2000). The sequence manipulation suite: JavaScript programs for analysing and formatting protein and DNA sequences. BioTechniques, 28(6), 1102,1104. https://doi.org/10.2144/00286ir01

Subbotina, T. N., Maslyukova, I. E., Faleeva, A. A., Nikolaeva, P. A., Khazieva, A. S., Dunaeva, E. A., Mironov, K. O., Polushkina, L. B., Martynkevich, I. S., Vereshchagina, S. V., & Barankin, B. V. (2020). Using the Minor Variant Finder software to identify and quantify the allelic burden level of somatic mutations in oncohematologic diseases. Oncogematologiya, 15(2), 85–91. https://doi.org/10.17650/1818-8346-2020-15-2-85-91

Suckling, L., McFarlane, C., Sawyer, C., Chambers, S. P., Kitney, R. I., McClymont, D. W., & Freemont, P. S. (2019). Miniaturisation of high-throughput plasmid DNA library preparation for next-generation sequencing using multifactorial optimisation. Synthetic and Systems Biotechnology, 4(1), 57–66. https://doi.org/10.1016/j.synbio.2019.01.002

Sullivan, T., Sullivan, R., & Ginsburg, O. M. (2015). Screening for Cancer: Considerations for Low- and Middle-Income Countries. In: H. Gelband, P. Jha, R. Sankaranarayanan, & S. Horton (Eds.), Disease Control Priorities, (3rd Edition, Volume 3: Cancer, November 2015, pages 211-222). World Bank Group. https://doi.org/10.1596/978-1-4648-0349-9_ch12

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/https://doi.org/10.3322/caac.21660

Suraweera, N., Duval, A., Reperant, M., Vaury, C., Furlan, D., Leroy, K., Seruca, R., Iacopetta, B., & Hamelin, R. (2002). Evaluation of tumor microsatellite instability using five quasimonomorphic mononucleotide repeats and pentaplex PCR. Gastroenterology, 123(6), 1804–1811. https://doi.org/https://doi.org/10.1053/gast.2002.37070

Susanti, S., Fadhil, W., Ebili, H. O., Asiri, A., Nestarenkaite, A., Hadjimichael, E., Ham-Karim, H. A., Field, J., Stafford, K., Matharoo-Ball, B., Hassall, J. C., Sharif, A., Oniscu, A., & Ilyas, M. (2018). N_LyST: a simple and rapid screening test for Lynch syndrome. Journal of Clinical Pathology, 71(8), 713–720. https://doi.org/10.1136/jclinpath-2018-205013

Takahashi, N., Yamada, Y., Taniguchi, H., Fukahori, M., Sasaki, Y., Shoji, H., Honma, Y., Iwasa, S., Takashima, A., Kato, K., Hamaguchi, T., & Shimada, Y. (2014). Clinicopathological features and prognostic roles of KRAS, BRAF, PIK3CA and NRAS mutations in advanced gastric cancer. BMC Research Notes, 7(1), 271. https://doi.org/10.1186/1756-0500-7-271

Talseth-Palmer, B. A. (2017). The genetic basis of colonic adenomatous polyposis syndromes. Hereditary Cancer in Clinical Practice, 15(1), 5. https://doi.org/10.1186/s13053-017-0065-x

Tamura, G. (2006). Alterations of tumor suppressor and tumor-related genes in the development and progression of gastric cancer. World Journal of Gastroenterology, 12(2), 192–198. https://doi.org/10.3748/wjg.v12.i2.192

Tao, C.-C., Yang, Y., Li, F., Qiao, L., Wu, Y., Sun, X.-D., Zhang, Y.-Y., & Li, C.-L. (2020). Cloning short DNA into plasmids by one-step PCR. Thoracic Cancer, 11(11), 3409–3415. https://doi.org/10.1111/1759-7714.13660

Tay, S. T., Leong, S. H., Yu, K., Aggarwal, A., Tan, S. Y., Lee, C. H., Wong, K., Visvanathan, J., Lim, D., Wong, W. K., Soo, K. C., Kon, O. L., & Tan, P. (2003). A combined comparative genomic hybridisation and expression microarray analysis of gastric cancer reveals novel molecular subtypes. Cancer Research, 63(12), 3309–3316.

Taylor, P. R. (2007). Prevention of Gastric Cancer: A Miss. JNCI: Journal of the National Cancer Institute, 99(2), 101–103. https://doi.org/10.1093/jnci/djk026

Tenesa, A., Farrington, S. M., Prendergast, J. G. D., Porteous, M. E., Walker, M., Haq, N., Barnetson, R. A., Theodoratou, E., Cetnarskyj, R., Cartwright, N., Semple, C., Clark, A. J., Reid, F. J. L., Smith, L. A., Kavoussanakis, K., Koessler, T., Pharoah, P. D. P., Buch, S., Schafmayer, C., … Dunlop, M. G. (2008). Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nature Genetics, 40(5), 631–637. https://doi.org/10.1038/ng.133

Therkildsen, C., Bergmann, T. K., Henrichsen-Schnack, T., Ladelund, S., & Nilbert, M. (2014). The predictive value of KRAS, NRAS, BRAF, PIK3CA and PTEN for anti-EGFR treatment in metastatic colorectal cancer: A systematic review and meta-analysis. Acta Oncologica, 53(7), 852–864. https://doi.org/10.3109/0284186X.2014.895036

Thompson, M. R., O'Leary, D. P., Flashman, K., Asiimwe, A., Ellis, B. G., & Senapati, A. (2017). Clinical assessment to determine the risk of bowel cancer using Symptoms, Age, Mass and Iron deficiency anaemia (SAMI). The British Journal of Surgery, 104(10), 1393–1404. https://doi.org/10.1002/bjs.10573

Thrumurthy, S. G., Anuruddha, A. H. M. P., De Zoysa, M. I. M., & Samarasekera, D. N. (2011). Unexpected outcome from Trousseau syndrome. BMC Surgery, 11(1), 1. https://doi.org/10.1186/1471-2482-11-1

Thunnissen, E., Bubendorf, L., Dietel, M., Elmberger, G., Kerr, K., Lopez-Rios, F., Moch, H., Olszewski, W., Pauwels, P., Penault-Llorca, F., & Rossi, G. (2012). EML4-ALK testing in non-small cell carcinomas of the lung: a review with recommendations. Virchows Archiv: An International Journal of Pathology, 461(3), 245–257. https://doi.org/10.1007/s00428-012-1281-4

Timmons, C., Dobhal, S., Fletcher, J., & Ma, L. I. M. (2013). Primers with 5′ Flaps Improve the Efficiency and Sensitivity of Multiplex PCR Assays for the Detection of Salmonella and Escherichia coli O157:H7. Journal of Food Protection, 76(4), 668–673. https://doi.org/10.4315/0362-028X.JFP-12-428

Tomlinson, I. P. M., Webb, E., Carvajal-Carmona, L., Broderick, P., Howarth, K., Pittman, A. M., Spain, S., Lubbe, S., Walther, A., Sullivan, K., Jaeger, E., Fielding, S., Rowan, A., Vijayakrishnan, J., Domingo, E., Chandler, I., Kemp, Z., Qureshi, M., Farrington, S. M., … Houlston, R. S. (2008). A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nature Genetics, 40(5), 623–630. https://doi.org/10.1038/ng.111

Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar, S., Chandler, I., Gorman, M., Wood, W., Barclay, E., Lubbe, S., Martin, L., Sellick, G., Jaeger, E., Hubner, R., Wild, R., Rowan, A., Fielding, S., … Houlston, R. (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nature Genetics, 39(8), 984–988. https://doi.org/10.1038/ng2085

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. CA: A Cancer Journal for Clinicians, 65(2), 87–108. https://doi.org/https://doi.org/10.3322/caac.21262

Tørresen, O. K., Star, B., Mier, P., Andrade-Navarro, M. A., Bateman, A., Jarnot, P., Gruca, A., Grynberg, M., Kajava, A. V, Promponas, V. J., Anisimova, M., Jakobsen, K. S., & Linke, D. (2019). Tandem repeats lead to sequence assembly errors and impose multi-level challenges for genome and protein databases. Nucleic Acids Research, 47(21), 10994–11006. https://doi.org/10.1093/nar/gkz841

Trabucco, S. E., Gowen, K., Maund, S. L., Sanford, E., Fabrizio, D. A., Hall, M. J., Yakirevich, E., Gregg, J. P., Stephens, P. J., Frampton, G. M., Hegde, P. S., Miller, V. A., Ross, J. S., Hartmaier, R. J., Huang, S.-M. A., & Sun, J. X. (2019). A Novel Next-Generation Sequencing Approach to Detecting Microsatellite Instability and Pan-Tumor Characterisation of 1000 Microsatellite Instability-High Cases in 67,000 Patient Samples. The Journal of Molecular Diagnostics: JMD, 21(6), 1053–1066. https://doi.org/10.1016/j.jmoldx.2019.06.011

Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics, 13(1), 36–46. https://doi.org/10.1038/nrg3117

Trpkov, K., Hes, O., Williamson, S. R., Adeniran, A. J., Agaimy, A., Alaghehbandan, R., Amin, M. B., Argani, P., Chen, Y.-B., Cheng, L., Epstein, J. I., Cheville, J. C., Comperat, E., da Cunha, I. W., Gordetsky, J. B., Gupta, S., He, H., Hirsch, M. S., Humphrey, P. A., … Gill, A. J. (2021). New developments in existing WHO entities and evolving molecular concepts: The Genitourinary Pathology Society (GUPS) update on renal neoplasia. Modern Pathology, 34, 1392–1424. https://doi.org/10.1038/s41379-021-00779-w

Tsiatis, A. C., Norris-Kirby, A., Rich, R. G., Hafez, M. J., Gocke, C. D., Eshleman, J. R., & Murphy, K. M. (2010). Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. The Journal of Molecular Diagnostics: JMD, 12(4), 425–432. https://doi.org/10.2353/jmoldx.2010.090188

Tucker, E. J., & Huynh, B. L. (2014). Genotyping by high-resolution melting analysis. Methods in Molecular Biology, 1145(3), 59–66. https://doi.org/10.1007/978-1-4939-0446-4_5

Turner, J. (2015). Chapter 17 - The gastrointestinal tract. In J. C. A. & J. A. P. V. Kumar, A.K. Abbas (Ed.), Robbins and Cotran pathologic Basis of Diseases (9th ed., pp. 749–820). Elsevier Saunders.

Twomey, J. D., & Zhang, B. (2021). Cancer Immunotherapy Update: FDA-Approved Checkpoint Inhibitors and Companion Diagnostics. The AAPS Journal, 23(2), 39. https://doi.org/10.1208/s12248-021-00574-0

U.S. Food and Drug Administration. (2021a). FDA approves first immunotherapy for initial treatment of gastric cancer. US FDA News and Events. https://www.fda.gov/news-events/press-announcements/fda-approves-first-immunotherapy-initial-treatment-gastric-cancer

U.S. Food and Drug Administration. (2021b). FDA approves pembrolizumab for esophageal or GEJ carcinoma. US FDA News and Events. https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-pembrolizumab-esophageal-or-gej-carcinoma

UK Accreditation Services. (2018). ISO 15189: Medical Laboratory Accreditation. UKAS Accreditation Standards. https://www.ukas.com/accreditation/standards/medical-laboratory-accreditation/

UK Department of Health & Social Care. (2018). Government Response to the House of Commons Science and Technology Committee's Third Report of Session 2017-19, 'Genomics and Genome Editing in the NHS. UK Government Policy Paper. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/725632/Government_Response_to_the_Genomics_and_Genome_....pdf

UK Department of Health & Social Care. (2020). Genome UK: the future of healthcare. UK Government Policy Paper. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/920378/Genome_UK_-_the_future_of_healthcare.pdf

Umar, A., Boland, C. R., Terdiman, J. P., Syngal, S., Chapelle, A. de la, Rüschoff, J., Fishel, R., Lindor, N. M., Burgart, L. J., Hamelin, R., Hamilton, S. R., Hiatt, R. A., Jass, J., Lindblom, A., Lynch, H. T., Peltomaki, P., Ramsey, S. D., Rodriguez-Bigas, M. A., Vasen, H. F. A., … Srivastava, S. (2004). Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. JNCI: Journal of the National Cancer Institute, 96(4), 261–268. https://doi.org/10.1093/jnci/djh034

Underhill, H. R., Kitzman, J. O., Hellwig, S., Welker, N. C., Daza, R., Baker, D. N., Gligorich, K. M., Rostomily, R. C., Bronner, M. P., & Shendure, J. (2016). Fragment Length of Circulating Tumor DNA. PLoS Genetics, 12(7), e1006162–e1006162. https://doi.org/10.1371/journal.pgen.1006162

United Kingdom National External Quality Assessment Service. (2019). 2019/20 Services Available List. UK NEQAS International Quality Expertise. https://ukneqas.org.uk/wp-content/uploads/2019/05/UK-NEQAS-Services-Available-Leaflet-2019-2020-FINAL.pdf

Van Cutsem, E., Köhne, C.-H., Láng, I., Folprecht, G., Nowacki, M. P., Cascinu, S., Shchepotin, I., Maurel, J., Cunningham, D., Tejpar, S., Schlichting, M., Zubel, A., Celik, I., Rougier, P., & Ciardiello, F. (2011). Cetuximab Plus Irinotecan, Fluorouracil, and Leucovorin As First-Line Treatment for Metastatic Colorectal Cancer: Updated Analysis of Overall Survival According to Tumor KRAS and BRAF Mutation Status. Journal of Clinical Oncology, 29(15), 2011–2019. https://doi.org/10.1200/JCO.2010.33.5091

Van Emburgh, B. O., Arena, S., Siravegna, G., Lazzari, L., Crisafulli, G., Corti, G., Mussolin, B., Baldi, F., Buscarino, M., Bartolini, A., Valtorta, E., Vidal, J., Bellosillo, B., Germano, G., Pietrantonio, F., Ponzetti, A., Albanell, J., Siena, S., Sartore-Bianchi, A., … Bardelli, A. (2016). Acquired RAS or EGFR mutations and duration of response to EGFR blockade in colorectal cancer. In Nature communications (Vol. 7, p. 13665). https://doi.org/10.1038/ncomms13665

van Lier, M. G. F., Westerman, A. M., Wagner, A., Looman, C. W. N., Wilson, J. H. P., de Rooij, F. W. M., Lemmens, V. E. P. P., Kuipers, E. J., Mathus-Vliegen, E. M. H., & van Leerdam, M. E. (2011). High cancer risk and increased mortality in patients with Peutz-Jeghers syndrome. Gut, 60(2), 141–147. https://doi.org/10.1136/gut.2010.223750

Vandenbussche, F., Mathijs, E., Lefebvre, D., De Clercq, K., & Van Borm, S. (2016). A Tale of Tails: Dissecting the Enhancing Effect of Tailed Primers in Real-Time PCR. PLOS ONE, 11(10), e0164463. https://doi.org/10.1371/journal.pone.0164463

VanGuilder, H. D., Vrana, K. E., & Freeman, W. M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. BioTechniques, 44(5), 619–626. https://doi.org/10.2144/000112776

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz Jr, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. Science (New York, N.Y.), 339(6127), 1546–1558. https://doi.org/10.1126/science.1235122

Voisey, J. & Morris, C. P. (2008). SNP Technologies for Drug Discovery: A Current Review. In Current Drug Discovery Technologies, 5(3):230–235. https://doi.org/http://dx.doi.org/10.2174/157016308785739811

Vos, E. L., Maron, S. B., Krell, R. W., Nakauchi, M., Fiasconaro, M., Walch, H. S., Capanu, M., Ku, G. Y., Ilson, D. H., Janjigian, Y. Y., Vanderbilt, C., Tang, L. H., & Strong, V. E. (2021). The interaction between microsatellite instability high (MSI-high) gastric cancer and chemotherapy on survival. Journal of Clinical Oncology, 39(3_suppl), 244. https://doi.org/10.1200/JCO.2021.39.3_suppl.244

Vossen, R. H. A. M., Aten, E., Roos, A., & den Dunnen, J. T. (2009). High-resolution melting analysis (HRMA): more than just sequence variant screening. Human Mutation, 30(6), 860–866. https://doi.org/10.1002/humu.21019

Wadhwani, N., & Diwakar, D. K. (2018). Localised perforation of locally advanced transverse colon cancer with spontaneous colocutaneous fistula formation: a clinical challenge. BMJ Case Reports, 2018, bcr-2018-224668. https://doi.org/10.1136/bcr-2018-224668

Wang, K., Kan, J., Yuen, S. T., Shi, S. T., Chu, K. M., Law, S., Chan, T. L., Kan, Z., Chan, A. S. Y., Tsui, W. Y., Lee, S. P., Ho, S. L., Chan, A. K. W., Cheng, G. H. W., Roberts, P. C., Rejto, P. A., Gibson, N. W., Pocalyko, D. J., Mao, M., … Leung, S. Y. (2011). Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. Nature Genetics, 43(12), 1219–1223. https://doi.org/10.1038/ng.982

Wang, X. (2016). Gene mutation-based and specific therapies in precision medicine. Journal of Cellular and Molecular Medicine, 20(4), 577–580. https://doi.org/10.1111/jcmm.12722

Warden, C. D., Lee, H., Tompkins, J. D., Li, X., Wang, C., Riggs, A. D., Yu, H., Jove, R., & Yuan, Y.-C. (2013). COHCAP: an integrative genomic pipeline for single-nucleotide resolution DNA methylation analysis. Nucleic Acids Research, 41(11), e117. https://doi.org/10.1093/nar/gkt242

Wei, T., Liao, B., Ackermann, B. L., Jolly, R. A., Eckstein, J. A., Kulkarni, N. H., Helvering, L. M., Goldstein, K. M., Shou, J., Estrem, S. T., Ryan, T. P., Colet, J.-M., Thomas, C. E., Stevens, J. L., & Onyia, J. E. (2005). Data-driven analysis approach for biomarker

discovery using molecular-profiling technologies. Biomarkers, 10(2–3), 153–172. https://doi.org/10.1080/13547500500107430

Weisenberger, D. J., Siegmund, K. D., Campan, M., Young, J., Long, T. I., Faasse, M. A., Kang, G. H., Widschwendter, M., Weener, D., Buchanan, D., Koh, H., Simms, L., Barker, M., Leggett, B., Levine, J., Kim, M., French, A. J., Thibodeau, S. N., Jass, J., … Laird, P. W. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nature Genetics, 38(7), 787–793. https://doi.org/10.1038/ng1834

Wieduwilt, M. J., & Moasser, M. M. (2008). The epidermal growth factor receptor family: biology driving targeted therapeutics. Cellular and Molecular Life Sciences: CMLS, 65(10), 1566–1584. https://doi.org/10.1007/s00018-008-7440-8

Willems, T., Gymrek, M., Highnam, G., Consortium, 1000 Genomes Project, Mittelman, D., & Erlich, Y. (2014). The landscape of human STR variation. Genome Research, 24(11), 1894–1904. https://doi.org/10.1101/gr.177774.114

Williams, J. G., Kubelik, A. R., Livak, K. J., Rafalski, J. A., & Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research, 18(22), 6531–6535. https://doi.org/10.1093/nar/18.22.6531

Williams-Woods, J., González-Escalona, N., & Burkhardt, W. (2011). Direct sequencing of hepatitis A virus and norovirus RT-PCR products from environmentally contaminated oyster using M13-tailed primers. Journal of Virological Methods, 178(1), 253–257. https://doi.org/https://doi.org/10.1016/j.jviromet.2011.09.014

Willis, O., Choucair, K., Alloghbi, A., Stanbery, L., Mowat, R., Charles Brunicardi, F., Dworkin, L., & Nemunaitis, J. (2020). PIK3CA gene aberrancy and role in targeted therapy of solid malignancies. Cancer Gene Therapy, 27(9), 634–644. https://doi.org/10.1038/s41417-020-0164-0

Woerner, S. M., Yuan, Y. P., Benner, A., Korff, S., von Knebel Doeberitz, M., & Bork, P. (2010). SelTarbase, a database of human mononucleotide-microsatellite mutations and their potential impact to tumorigenesis and immunology. Nucleic Acids Research, 38(Database issue), D682–D689. https://doi.org/10.1093/nar/gkp839

Wojdacz, T. K., & Dobrovic, A. (2007). Methylation-sensitive high resolution melting (MS-HRM): a new approach for sensitive and high-throughput assessment of methylation. Nucleic Acids Research, 35(6), e41–e41. https://doi.org/10.1093/nar/gkm013

Xie, T.-H., Su, P., Hong, J.-G., & Zhang, H. (2021). Cervical lymph node enlargement as the initial manifestation of rectal cancer. BMC Gastroenterology, 21(1), 57. https://doi.org/10.1186/s12876-021-01628-5

Xiu, L., Li, Y., Wang, F., Zhang, C., Li, Y., Zeng, Y., Yin, Y., & Peng, J. (2020). Multiplex High-Resolution Melting Assay for Simultaneous Identification of Molecular Markers Associated with Extended-Spectrum Cephalosporins and Azithromycin Resistance in

Neisseria gonorrhoeae. The Journal of Molecular Diagnostics, 22(11), 1344–1355. https://doi.org/10.1016/j.jmoldx.2020.08.003

Xu, Y., Li, C., Zhang, Y., Guo, T., Zhu, C., Xu, Y., & Liu, F. (2020). Comparison Between Familial Colorectal Cancer Type X and Lynch Syndrome: Molecular, Clinical, and Pathological Characteristics and Pedigrees. In Frontiers in Oncology, 10, 1603. https://www.frontiersin.org/article/10.3389/fonc.2020.01603

Yamagishi, H., Kuroda, H., Imai, Y., & Hiraishi, H. (2016). Molecular pathogenesis of sporadic colorectal cancers. Chinese Journal of Cancer, 35, 4. https://doi.org/10.1186/s40880-015-0066-y

Yamamoto, H., Watanabe, Y., Maehata, T., Imai, K., & Itoh, F. (2020). Microsatellite instability in cancer: a novel landscape for diagnostic and therapeutic approach. Archives of Toxicology, 94(10), 3349–3357. https://doi.org/10.1007/s00204-020-02833-z

Yamashita, R., Long, J., Longacre, T., Peng, L., Berry, G., Martin, B., Higgins, J., Rubin, D. L., & Shen, J. (2021). Deep learning model for the prediction of microsatellite instability in colorectal cancer: a diagnostic study. The Lancet. Oncology, 22(1), 132–141. https://doi.org/10.1016/S1470-2045(20)30535-0

Zanke, B. W., Greenwood, C. M. T., Rangrej, J., Kustra, R., Tenesa, A., Farrington, S. M., Prendergast, J., Olschwang, S., Chiang, T., Crowdy, E., Ferretti, V., Laflamme, P., Sundararajan, S., Roumy, S., Olivier, J.-F., Robidoux, F., Sladek, R., Montpetit, A., Campbell, P., … Dunlop, M. G. (2007). Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nature Genetics, 39(8), 989–994. https://doi.org/10.1038/ng2089

Zhang, H., Cheng, H., Wang, Q., Zeng, X., Chen, Y., Yan, J., Sun, Y., Zhao, X., Li, W., Gao, C., Gong, W., Li, B., Zhang, R., Nan, L., Wu, Y., Bao, S., Han, J.-D. J., & Zheng, H. (2015). An advanced fragment analysis-based individualised subtype classification of pediatric acute lymphoblastic leukemia. Scientific Reports, 5(1), 12435. https://doi.org/10.1038/srep12435

Zhang, L. (2008). Immunohistochemistry versus microsatellite instability testing for screening colorectal cancer patients at risk for hereditary nonpolyposis colorectal cancer syndrome. Part II. The utility of microsatellite instability testing. The Journal of Molecular Diagnostics: JMD, 10(4), 301–307. https://doi.org/10.2353/jmoldx.2008.080062

Zhang, Z., Hernandez, K., Savage, J., Li, S., Miller, D., Agrawal, S., Ortuno, F., Staudt, L. M., Heath, A., & Grossman, R. L. (2021). Uniform genomic data analysis in the NCI Genomic Data Commons. Nature Communications, 12(1), 1226. https://doi.org/10.1038/s41467-021-21254-9

Zhao, P., Li, L., Jiang, X., & Li, Q. (2019). Mismatch repair deficiency/microsatellite instability-high as a predictor for anti-PD-1/PD-L1 immunotherapy efficacy. Journal of Hematology & Oncology, 12(1), 54. https://doi.org/10.1186/s13045-019-0738-1

Zheng, K., Wan, H., Zhang, J., Shan, G., Chai, N., Li, D., Fang, N., Liu, L., Zhang, J., Du, R., Wu, Q., Li, X., & Zhang, C. (2020). A novel NGS-based microsatellite instability (MSI) status classifier with 9 loci for colorectal cancer patients. Journal of Translational Medicine, 18(1), 215. https://doi.org/10.1186/s12967-020-02373-1

Zhong, L., Li, Y., Xiong, L., Wang, W., Wu, M., Yuan, T., Yang, W., Tian, C., Miao, Z., Wang, T., & Yang, S. (2021). Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. Signal Transduction and Targeted Therapy, 6(1), 201. https://doi.org/10.1038/s41392-021-00572-w

APPENDIX

APPENDIX 1: COMPARISONS OF COST ESTIMATES OF DEPLOYING NEW ASSAYS VERSUS STANDARD ASSAYS

The fairest way to make a cost comparison between the assays designed for this thesis and the other established assays is to compare the costs of the reagents/kits used in the experiments for both the designed assays and the established ones. Doing otherwise, such as comparing the amount a laboratory charges for the tests with the cost of the reagents I used for the designed assays, would be unfair. This is the because the costs a laboratory charges for its tests would most likely include other incidental charges such as cost of labour and rent and profit margins for rendering the service of deploying the test. Additionally, comparing the costs of the equipment used in deploying the assays is fair way of comparing cost-effectiveness of the assays. Below is a list of the designed assays and the standard tests to which they are compared. The costs of the reagents used in both the designed assays and the standard tests are found in the web addresses indicated in parenthesis after the tests. Some of the websites may require registration and login to be able to view the costs of the reagents. The table that follows gives a comparison of the cost estimates of deploying the designed assays versus standard tests using the cost of reagents only.

- **Squirrel PCR assay** (https://www.thermofisher.com/order/catalog/product/A38073)

  versus

  - **Next generation sequencing** (https://www.illumina.com/products/by-type/sequencing-kits/library-prep-kits/ampliseq-cancer-hotspot-panel.html)
  - **Pyrosequencing** (https://www.qiagen.com/gb/products/discovery-and-translational-research/pyrosequencing/consumables/pyromark-q24-mdx-accessories-and-reagents/; https://www.qiagen.com/gb/products/discovery-and-translational-

research/pyrosequencing/instruments/pyromark-q24-and-advanced/?catno=978703; https://www.sigmaaldrich.com/GB/en/substance/streptavidinsepharosehighperformance1234598765?gclid=CjwKCAiAg6yRBhBNEiwAeVyL0NANaLlMr9CNcKIJ991XRBlRxNfAazz4Ajs2c3smUo1D-kAgtipbUhoCjVAQAvD_BwE)

- **Sequencing by cloning** (https://www.thermofisher.com/order/catalog/product/450030)

- **MSI-HRMA** (https://www.clentlifescience.co.uk/products/hotshot-diamond-pcr-mastermix/; https://biotium.com/product/evagreen-dye-20x-in-water/) versus
  - **Fluorescent PCR followed by capillary electrophoresis** (https://www.proquinorte.com/en/products/truemark%E2%84%A2-msi-assay; https://www.fishersci.com/shop/products/promega-microsatellite-instability-analysis-version-1-2/PRMD1641)
  - **MMR IHC** (https://www.abcam.com/mismatch-repair-msh6-pms2-mlh1-msh2-antibody-panel-human-ab252190.html; https://www.sigmaaldrich.com/SA/en/technical-documents/protocol/protein-biology/immunohistochemistry/immunohistochemistry-protocol)
  - **Next generation sequencing**

- **mHRMA** (https://www.clentlifescience.co.uk/products/hotshot-diamond-pcr-mastermix/; https://biotium.com/product/evagreen-dye-20x-in-water/) versus
  - **Standard HRMA** (https://www.clentlifescience.co.uk/products/hotshot-diamond-pcr-mastermix/; https://biotium.com/product/evagreen-dye-20x-in-water/)

- **Standard ARMS-PCR** (same as standard HRMA but uses 3 primers per target)
- **Next generation sequencing** (as before)
- **Pyrosequencing** (as before)

The calculation of the costs of an assay is much more complex than simply considering the cost of the reagents. The final cost of any assay depends on the total cost of services involved in delivering the assay as well as the profit margin placed on the assay by the manufacturer or service provider. Hence, one assay may have cheaper reagents, but the final cost of deploying the assay may be higher than that of the next assay with more expensive reagents. For the same reason of cost of deploying services, the cost of the same IHC assay may vary widely from centre to centre.

Supplementary Table 1: Comparison of cost estimates of designed assays versus standard tests

Task	Assay	Cost of reagents	Cost of assay per reaction	Remark
Cost of sequencing ultrashort DNA segments	**Squirrel PCR assay for Sanger sequencing**	Squirrel primers: £70/10,000 reactions = £0.007/reaction PCR master mix: £220/500 reactions = £0.44/reaction Eva green dye: £214/10,000 reactions= £0.0214/reaction Useq primers: £15/10,000 reactions=£0.0015/reaction Sanger sequencing kit: £1,288/200 reactions =£6.44/reaction	£6.8885/reaction per marker	Total cost of sequencing is £75.7735 for the 11 odd markers recommended by NCCN/NICE
	Next-generation sequencing	AmpliSeq Cancer Hotspot Panel for Illumina (50 genes): £527/24 reactions = £21.96/reaction AmpliSeq Library Prep kit: £2515/24 reaction = £104.79/reaction AmpliSeq Adapters: £136/24 reactions = £5.67/reaction	£132.42/reaction for 50 markers	Total cost of sequencing 11 NCCN/NICE markers (and microsatellites) remains £132.42*
	Pyrosequencing	Biotinylated PCR primers=£0.132/reaction Conventional primers: £15/10,000 reactions = £0.0015/reaction Qiagen PyroMark Q24 PCR kit = £189/200 reactions = £0.945/reaction Streptavidin Sepharose High Performance, GE Healthcare =£467/2,500 reactions = £0.1868/reaction Sequencing primers=£15/10,000 reactions = £0.0015/reaction	£5.46/reaction per marker	Total cost of sequencing all 11 NCCN/NICE markers is £60.06

		PyroMark Q24 Validation Oligo = £560/3 assays = £186.67/assay Qiagen PyroMark Q24 Gold Reagents (enzymes, dNTPs and template prep buffers: £503.24/120 reactions = £4.19/reaction		
	Sequencing by cloning	Thermofisher TOPO™ TA Cloning™ Kit for Sequencing, without competent cells: £280/25 reactions = £3.94/reaction PCR primers=£15/10,000 reactions = £0.0015/reaction Sanger sequencing kit: £1,288/200 reactions = £6.44/reaction	£10.38/reaction per marker	Total cost of sequencing 11 NCCN/NICE markers is £114.18
Cost of MSI Analysis	**MSI-HRMA versus**	Conventional primers: £15/10,000 reactions = £0.0015/reaction PCR master mix: £220/500 reactions = £0.44/reaction Eva green dye: £214/10,000 reactions= £0.0214/reaction	£0.4629/reaction per marker	Total cost of testing the complete MSI_HRMA panel in duplicate reactions per sample is £5.5548
	Fluorescent PCR followed by capillary electrophoresis	Applied Biosystems TrueMark™ MSI Assay: £1,916/100 reactions = £19.16/reaction	£19.16/reaction	£19.16/reaction for all 5 markers
	Mismatch repair immunohistochemistry	Abcam Mismatch Repair (MSH6, PMS2, MLH1, MSH2) Antibody Panel: £395/20 samples = £19.75/sample	£19.75/sample for all 4 IHC markers	£19.75/sample for all 4 IHC markers
	Next-generation sequencing	As before	As before	Total cost of sequencing 11 NCCN/NICE markers (and microsatellites) remains £132.42*

Cost of multiplexing HRMA versus cost of standard tests	**mHRMA (duplex PCRs)**	Conventional primers: £30/10,000 reactions = £0.0030/reaction = £0.0015/target PCR master mix: £220/333 reactions* = £0.660/reaction = £0.330/target Eva green dye: £214/6,666 reactions= £0.032/reaction = £0.016/target	£0.3278/reaction	Each reaction tests two markers giving a cost of £0.1639 per marker and £3.6058 for duplicate reactions for all 11 NCCN/NICE markers
	Standard HRMA	Conventional primers: £15/10,000 reactions = £0.0015/reaction = £0.0015/target PCR master mix: £220/500 reactions = £0.44/reaction = £0.44/target Eva green dye: £214/10,000 reactions= £0.0214/reaction = £0.0214/target	£0.4629/reaction	£0.4629 per marker and £10.1838 for duplicate reactions for all 11 NCCN/NICE markers
	Standard ARMS-PCR	Conventional primers: £22.5/10,000 reactions = £0.00225/reaction = £0.00225/target PCR master mix: £220/500 reactions = £0.44/reaction = £0.44/target Eva green dye: £214/10,000 reactions= £0.0214/reaction = £0.0214/target	£0.4839/ reaction	£0.4839 per marker and £10.6458 for duplicate reactions for all 11 NCCN/NICE markers
	Next-generation sequencing	As before	As before	Total cost of sequencing 11 NCCN/NICE markers (and microsatellites) remains £132.42.

APPENDIX 2: COMPARISON OF TURNAROUND TIME ESTIMATES FOR DEPLOYING NEW ASSAYS VERSUS STANDARD ASSAYS

The terms 'faster' and/or 'easier to perform' were used to compare the described molecular assays to existing ones, in terms of the turnaround time for performing the test at the bench (completion of the experiment and analysis of the experimental data), and the amount of labour that goes into performing the experiment and analysing the data to produce the results, respectively. The most objective way of comparing speed and ease of operation among the assays is to use the amount of time spent from experiments to output of results. These times were curated from literatures describing and comparing these assays.

- Squirrel PCR assay versus
    - Next-generation sequencing (Liu et al, 2012)
    - Pyrosequencing (as above; Fakhruddin et al, 2013)
    - Sequencing by cloning (Jajesniak & Wong, 2015; Li et al., 2011; Tao et al., 2020)
- MSI-HRMA versus
    - Flourescent PCR followed by capillary electrophoresis
    - IHC (https://www.sigmaaldrich.com/GB/en/technical-documents/protocol/protein-biology/immunohistochemistry/immunohistochemistry-protocol; https://www.biolegend.com/en-us/protocols/immunohistochemistry-protocol-for-paraffin-embedded-sections)
    - NGS (as before)
- mHRMA versus
    - Standard HRMA (Fadhil et al, 2010, Ibrahem et al, 2010)

- Standard ARMS-PCR (same as standard HRMA)
- NGS (as before)

Supplementary Table 2: Comparison of turnaround times for designed assays versus standard tests

Task	Assay	Workflow	Turnaround time	Remark
Sequencing of ultrashort DNA segments	**Squirrel PCR assay for Sanger sequencing**	PCR =1.5-1.75 hours Sanger sequencing = 20 Mins -3 Hours	1hr 50 mins to 4.75 hrs for all 11 NCCN/NICE genetic markers	"Multiplexing" of capillary electrophoresis enables up to 16 targets to be sequenced at once
	Next-generation sequencing	Sample preparation = 2-8 hours Sequencing = 2 hours to 27 hours for paired end reads* depending on the platform Bioinformatic analysis = 1 hour	5 hours to more than 36 hours in total for one as well as for all 11 NCCN/NICE markers	
	Pyrosequencing	PCR = 1.5-1.75 hours Sample preparation= 15 mins – 8 hours, depending on the platform Pyrosequencing = 24 mins to 24 hours, depending on the platform Bioinformatic analysis =1 hour	3 hours 11 mins to 34 hour 4 mins in total for all 11 NCCN/NICE markers, depending on the platform.	
	Sequencing by cloning	Cloning experiments = 24 hours Sanger sequencing = 20 Mins -3 Hours		The fact that only one marker can be cloned at a time makes cloning cumbersome for routing clinical diagnosis
MSI testing	**MSI-HRMA versus**	PCR =1.5-1.75 hours HRMA =15-20 minutes	1.75 – 2.05 hours for all 5-6 markers/patient	

	Fluorescent PCR followed by capillary electrophoresis	Multiplex PCR =1.5 -1.75 hours Capillary electrophoresis =20 Mins - 3 Hours	1hr 50 mins to 4.75 hrs for all 5-6 MSI markers/patient	
	Mismatch repair immunohistochemistry	Deparaffinization and Rehydration = 1 hour 10 mins Antigen retrieval =35-40 mins ± peroxidase inactivation = 7-10 mins ±Blocking buffer=1 hour Primary antibody reaction = 1 hour 15 mins Secondary antibody reaction= 35 mins – 1 hour Development = 5-10 mins Counterstaining = 10 mins Slide reading by pathologist =15 mins	4 hours 5 mins to 5 hours 52 mins for all 4 markers stained simultaneously	This does not take into consideration the time spent on several other minor activities in-between the major procedures of IHC
	Next-generation sequencing	Sample preparation = 2-8 hours Sequencing = up to 27 hours for paired end reads*	More than 33 hours in total all microsatellite markers	
Mutation profiling	**mHRMA (duplex PCRs)**	PCR =1.5-1.75 hours HRMA =15-20 minutes	1.75 – 2.05 hours per 2 markers = 50 mins -1 hour per marker OR 9.17 – 11 hours for all 11 NCCN/NICE markers	
	Standard HRMA	PCR =1.5-1.75 hours HRMA =15-20 minutes	1.75 – 2.05 hours per marker OR 19.25 – 22.55 hours for all 11 NCCN/NICE markers	

	Standard ARMS-PCR	PCR =1.5-1.75 hours Gel electrophoresis = 45mins -1hour	2.25 -2.75 hours per marker OR 24.75 – 30.25 hours for all 11 NCCN/NICE markers	
	Next-generation sequencing	Sample preparation = 2-8 hours Sequencing = up to 27 hours for paired end reads*	More than 33 hours in total for one as well as for all 11 NCCN/NICE markers	
	Pyrosequencing	PCR = 1.5-1.75 hours Sample preparation= 15 mins – 8 hours, depending on the platform Pyrosequencing = 24 mins to 24 hours, depending on the platform Bioinformatic analysis =1 hour	3 hours 11 mins to 34 hour 4 mins in total for all 11 NCCN/NICE markers, depending on the platform.	

*Sequencing time for NGS can extend from one to 14 days depending on the platform. Sequencing using newer third generation platforms tend to be faster

CHAPTER 5

MSI-HRMA FOR MICROSATELLITE INSTABILITY TESTING IN COLORECTAL CANCER

Table 5.1: Features of MSI-HRMA primers

Microsatellites	Primer sequences	Product size	Tm
BAT25	OUTER F: GAGTGATTCTCTAAAGAGTTTTGTGTTTTGT R: TATGGCTCTAAAATGCTCTGTTCTC NESTED F: GTGATTCTCTAAAGAGTTTTGTGTTTTGT R: TGGCTCTAAAATGCTCTGTTCTCAA	89bp 85bp	76^0C 75.5^0C
BAT26	OUTER F: TCAGAGCCCTTAACCTTTTTCAGGT R: GAAAACATTTTTTAACCATTCAACATTTTTAA NESTED F: GCCCTTAACCTTTTTCAGGTAA R: CATTTTTTAACCATTCAACATTTTTAA	87bp 77bp	76^0C 74.5^0C
BCAT25	OUTER F: TCTGTAATGGTACTGACTTTGCT R: AACTTAACACTACGAGAGACTTAAAA NESTED F: GACTTTGCTTGCTTTGAAGTAGCTCT R: CACTACGAGAGACTTAAAAAACAGTTACTGCA	102bp 81bp	79.5^0C 77.5^0C
ANGEL2	OUTER F: CAAGTTTGGAATCACTGGGTACAATGCAAC R: TTGGGCGACAGAGTGAGACCCTGTCT NESTED F: GAATCACTGGGTACAATGCAAC R: CAGAGTGAGACCCTGTCTCAAA	81bp 65bp	81^0C 77^0C
TP53	OUTER F: GGCAACAAAGCGAGACCCAGTCT R: CCAAGACTTGTTTTATGCTCAGGGTCA NESTED F: GCGAGACCCAGTCTCAAAG R: TATGCTCAGGGTCAATTTCTT	 91bp 69bp	 77^0C 76.5^0C
MONO27	OUTER F: GGAGACAGAGCAAGACTCTGCCT R: GCCAACTAAAAAAGAAAAAAAAGTAAAACCAGGATTT NESTED F: AGATTTATTTTGGGCTTCACTCTT R: GTGACAGAGCTAGACTCTGTCTC	 85bp 72bp	 77^0C 76.5^0C
MYB	OUTER F: AGACTTACACCTCTGGGCTTTATA R: CAAGAACTACAAAAATGTAAAACACGA NESTED F: CAACTAAAAAAGAAAAAAAAGTAAAACCAG R: CAGAGCAAGACTCTGCCTCAA	87bp 78bp	77^0C 75.5^0C
TYMS26	OUTER		

	GGGCTCAAACAAAAAACCTCGGAAAAGCCCT CTCCAGCCTGGGCGACAGAGCAAGACTGTCCCA NESTED F: AAACCTCGGAAAAGCCCTGGCGGTCT R: GCCTGGGCGACAGAGCAAGACTGTCCCA	96bp 78bp	84.5^0C 83^0C
ESWR1	OUTER F: cccgcgcACTGTAACAATGTTCATGGTTGTGATGT R: cccgcgcgTGACTCTTTATAAACATTTGGAATTTTA NESTED F: TAACAATGTTCATGGTTGTGATGT R: TGACTCTTTATAAACATTTGGAATTTTA	89bp 70bp	79.5^0C 73^0C
FBXO46	OUTER F: ACCAGGGCCGTGGACCCTCT R: GGGCTGAATGGAGAAGAAAGTGAGATGCT NESTED F: TGGACCCTCTGAGAAGGGT R: GGAGAAGAAAGTGAGATGCTGA	76bp 57bp	83^0C 77.5^0C
TCF4	OUTER F: GAATACTTCCCTCGGGT R: GCCCCTTTCCACTTCT NESTED F: CTCGGGTGAAACTGAAGTTAA R: ACTTCTCACCTTGTGTTTGGA	 74bp 54bp	 80.5^0C 75.5^0C

USeq F: 5’-ctgactaaactaggtgccacgt

R: 5’-tagactcctgatcccttcat

40L F: aactgactaaactaggtgccacgtcgtgaaagtctgacaa
40L R: tagactcctgatcccttcattgccctgcatctgacacgca

Table 5.2: Concordance between QMC-PCR and single stage PCR for MSI-HRMA

QMC PCR		SINGLE STAGE PCR		CONCORDANCE
		ANGEL2		
		MSI*	MSS*	
ANGEL2	MSI	21	0	
	MSS	0	5	100%
		MYB		
		MSI	MSS	
MYB	MSI	22	0	
	MSS	0	4	100%
		EWSR1		
		MSI	MSS	
EWSR1	MSI	22	0	
	MSS	0	4	100%
		TP53		
		MSI	MSS	
TP53	MSI	23	0	
	MSS	0	3	100%
		TYMS26		
		MSI	MSS	
TYMS26	MSI	21	0	
	MSS	0	5	100%
		FBXO46		
		MSI	MSS	
FBXO46	MSI	25	0	
	MSS	0	1	100%
		TCF4		
		MSI	MSS	
TCF4	MSI	21	0	
	MSS	0	5	100%
		MONO27		
		MSI	MSS	
MONO27	MSI	21	0	
	MSS	0	5	100%

Table 5.3: MSI testing by MSI-HRMA Panel 1

SERIAL	ID	IHC	BAT25	ANGEL2	TCF4	TP53	TYMS26	MARKER	MSI-HRMA
1	Si #1	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
2	Si #2	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
3	Si #3	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
4	Si #4	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
5	Si#5	dMSH2	**MSS**	**MSS**	**MSS**	MSI	**MSS**	**1/5**	**MSS**
6	Si #6	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
7	Si #7	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
8	Si #8	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
9	Si #9	dMLH1	MSI	MSI	**MSS**	MSI	MSI	4/5	**MSI**
10	Si #10	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
11	Si #11	dMLH1	MSI	MSI	MSI	MSI	**MSS**	4/5	**MSI**
12	Si #12	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
13	Si #13	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
14	Si #14	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
15	Si #15	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
16	Si #16	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
17	Si #17	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
18	Si #18	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
19	Si #19	dMLH1	MSI	**MSS**	MSI	MSI	MSI	4/5	**MSI**
20	Si #20	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
21	Si #21	dMSH2	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
22	Si #22	dMSH2	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
23	Si #23	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
37	SS15	p-MMR	**MSS**	**MSS**	**MSS**	**MSS**	**MSS**	0/5	**MSS**
38	BX36	p-MMR	**MSS**	**MSS**	**MSS**	**MSS**	**MSS**	0/5	**MSS**
39	BX25	p-MMR	**MSS**	**MSS**	**MSS**	**MSS**	**MSS**	0/5	**MSS**

Table 5.4: MSI testing by MSI-HRMA Panel 2

SERIAL	ID	IHC	BAT26	EWSR1	FBXO46	TCF4	TP53	MARKER	MSI-HRMA
1	Si #1	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
2	Si #2	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
3	Si #3	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
4	Si #4	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
5	Si#5	**dMSH2**	MSS	MSI	MSI	MSS	MSI	**3/5**	**MSS**
6	Si #6	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
7	Si #7	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
8	Si #8	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
9	Si #9	**dMLH1**	MSI	MSI	MSI	**MSS**	MSI	4/5	**MSI**
10	Si #10	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
11	Si #11	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
12	Si #12	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
13	Si #13	**dMLH1**	MSI	**MSS**	MSI	MSI	MSI	4/5	**MSI**
14	Si #14	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
15	Si #15	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
16	Si #16	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
17	Si #17	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
18	Si #18	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
19	Si #19	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
20	Si #20	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
21	Si #21	**dMSH2**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
22	Si #22	**dMSH2**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
23	Si #23	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
37	SS15	**p-MMR**	MSS	MSS	MSS	MSS	MSS	0/5	**MSS**
38	BX36	**p-MMR**	MSS	MSS	MSS	MSS	MSS	0/5	**MSS**
39	BX25	**p-MMR**	MSS	MSS	**MSI**	MSS	MSS	1/5	**MSS**

Table 5.5: MSI testing by MSI-HRMA Panel 3

SERIAL	ID	IHC	BCAT25	ANGEL2	EWSR1	FBXO46	MYB	MARKER	MSI-HRMA
1	Si #1	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
2	Si #2	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
3	Si #3	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
4	Si #4	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
5	Si#5	**dMSH2**	MSI	MSS	MSI	MSI	MSI	4/5	**MSI**
6	Si #6	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
7	Si #7	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
8	Si #8	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
9	Si #9	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
10	Si #10	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
11	Si #11	**dMLH1**	MSI	MSI	MSI	MSI	**MSS**	4/5	**MSI**
12	Si #12	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
13	Si #13	**dMLH1**	MSI	MSI	**MSS**	MSI	MSI	4/5	**MSI**
14	Si #14	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
15	Si #15	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
16	Si #16	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
17	Si #17	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
18	Si #18	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
19	Si #19	**dMLH1**	MSI	**MSS**	MSI	MSI	MSI	4/5	**MSI**
20	Si #20	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
21	Si #21	**dMSH2**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
22	Si #22	**dMSH2**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
23	Si #23	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
37	SS15	**p-MMR**	MSS	MSS	MSS	MSS	MSS	0/5	MSS
38	BX36	**p-MMR**	MSS	MSS	MSS	MSS	MSS	0/5	MSS
39	BX25	**p-MMR**	MSS	MSS	MSS	**MSI**	MSS	1/5	MSS

Table 5.6: MSI testing by MSI-HRMA Panel 4

SERIAL	ID	IHC	ANGEL2	MONO27	MYB	TP53	TYMS26	MARKER	MSI-HRMA
1	Si #1	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
2	Si #2	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
3	Si #3	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
4	Si #4	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
5	Si#5	dMSH2	MSS	MSS	MSI	MSI	MSS	2/5	MSS
6	Si #6	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
7	Si #7	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
8	Si #8	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
9	Si #9	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
10	Si #10	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
11	Si #11	dMLH1	MSI	MSI	MSS	MSI	MSS	3/5	MSS
12	Si #12	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
13	Si #13	dMLH1	MSI	MSS	MSI	MSI	MSI	4/5	MSI
14	Si #14	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
15	Si #15	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
16	Si #16	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
17	Si #17	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
18	Si #18	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
19	Si #19	dMLH1	MSS	MSI	MSI	MSI	MSI	5/5	MSI
20	Si #20	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
21	Si #21	dMSH2	MSI	MSI	MSI	MSI	MSI	5/5	MSI
22	Si #22	dMSH2	MSI	MSI	MSI	MSI	MSI	5/5	MSI
23	Si #23	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
24	Rx8	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
25	Rx24	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
26	Rx26	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
27	Rx 27	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
28	bx24	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
29	bx30	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
30	SS7	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
31	SS8	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
32	SS9	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
33	SS10	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
34	SS11	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
35	SS12	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
36	SS14	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
37	SS15	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
38	BX36	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
39	BX25	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
40	BX20	p-MMR	MSS	MSI	MSS	MSS	MSS	1/5	MSS
41	BX21	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
42	BX14	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
43	BX28	p-MMR	MSS	MSI	MSS	MSS	MSS	1/5	MSS
44	BX31	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS

Table 5.7: MSI testing by MSI-HRMA Panel 5

SERIAL	ID	IHC	BCAT25	BAT25	BAT26	MYB	TP53	MARKER	MSI-HRMA
1	Si #1	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
2	Si #2	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
3	Si #3	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
4	Si #4	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
5	Si#5	**dMSH2**	MSI	MSS	MSS	MSI	MSI	**3/5**	MSS
6	Si #6	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
7	Si #7	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
8	Si #8	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
9	Si #9	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
10	Si #10	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
11	Si #11	**dMLH1**	MSI	MSI	MSI	**MSS**	MSI	4/5	**MSI**
12	Si #12	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
13	Si #13	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
14	Si #14	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
15	Si #15	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
16	Si #16	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
17	Si #17	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
18	Si #18	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
19	Si #19	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
20	Si #20	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
21	Si #21	**dMSH2**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
22	Si #22	**dMSH2**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
23	Si #23	**dMLH1**	MSI	MSI	MSI	MSI	MSI	5/5	**MSI**
37	SS15	**p-MMR**	MSS	MSS	MSS	MSS	MSS	0/5	MSS
38	BX36	**p-MMR**	MSS	MSS	MSS	MSS	MSS	0/5	MSS
39	BX25	**p-MMR**	MSS	MSS	MSS	MSS	MSS	0/5	MSS

Table 5.8: MSI testing by MSI-HRMA Panel 6

SERIAL	ID	IHC	ANGEL2	FBXO46	MONO27	TCF4	TYMS26	MARKER	MSI-HRMA
1	Si #1	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
2	Si #2	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
3	Si #3	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
4	Si #4	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
5	Si#5	dMSH2	MSS	MSI	MSS	MSS	MSS	1/5	MSS
6	Si #6	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
7	Si #7	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
8	Si #8	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
9	Si #9	dMLH1	MSI	MSI	MSI	MSS	MSI	4/5	MSI
10	Si #10	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
11	Si #11	dMLH1	MSI	MSI	MSI	MSI	MSS	4/5	MSI
12	Si #12	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
13	Si #13	dMLH1	MSI	MSI	MSS	MSI	MSI	4/5	MSI
14	Si #14	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
15	Si #15	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
16	Si #16	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
17	Si #17	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
18	Si #18	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
19	Si #19	dMLH1	MSS	MSI	MSI	MSI	MSI	4/5	MSI
20	Si #20	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
21	Si #21	dMSH2	MSI	MSI	MSI	MSI	MSI	5/5	MSI
22	Si #22	dMSH2	MSI	MSI	MSI	MSI	MSI	5/5	MSI
23	Si #23	dMLH1	MSI	MSI	MSI	MSI	MSI	5/5	MSI
24	Rx8	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
25	Rx24	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
26	Rx266	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
27	Rx 27	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
28	bx24	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
29	bx30	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
30	SS7	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
31	SS8	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
32	SS9	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
33	SS10	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
34	SS11	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
35	SS12	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
36	SS14	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
37	SS15	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
38	BX36	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
39	BX25	p-MMR	MSS	MSI	MSS	MSS	MSS	1/5	MSS
40	RX24	p-MMR	MSS	MSS	MSI	MSS	MSS	1/5	MSS
41	RX26	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
42	RX27	p-MMR	MSS	MSS	MSS	MSS	MSS	0/5	MSS
43	BX24	p-MMR	MSS	MSS	MSI	MSS	MSS	1/5	MSS
44	BX30	p-MMR	MSS	MSI	MSS	MSS	MSS	1/5	MSS

CHAPTER 6

MSI-WES: A SIMPLE APPROACH FOR MICROSATELLITE INSTABILITY TESTING USING WHOLE EXOME SEQUENCING

Table 6.1: Genomic coordinates of the mutational targets of MSI

GENE	GENOMIC COORDINATES
TGFBR2	Chromosome 3: 30,606,601-30,694,142
IGF2R	Chromosome 6: 159,969,082-160,113,507
TCF4	Chromosome 18: 55,222,185-55,664,787
RIZ (PRDM2)	Chromosome 1: 13,700,198-13,825,079
BAX	Chromosome 19: 48,954,815-48,961,798
CASPASE5(CASP5)	Chromosome 11: 104,994,235-105,023,168
FAS	Chromosome 10: 88,990,531-89,017,059
BCL10	Chromosome 1: 85,265,776-85,276,632
APAF1	Chromosome 12: 98,645,290-98,735,433
hMSH6 (MSH6)	Chromosome 2: 47,695,530-47,810,063
hMSH3 (MSH3)	Chromosome 5: 80,654,652-80,876,815
MED1	Chromosome 17: 39,404,285-39,451,272
RAD50	Chromosome 5: 132,556,019-132,646,349
BLM	Chromosome 15: 90,717,346-90,816,166
ATR	Chromosome 3: 142,449,007-142,578,733
MRE11	Chromosome 11: 94,415,570-94,493,885

Table 6.2: ENSEMBL gene identifier of MMR genes

GENE	ENSEMBL IDENTIFIER
MLH1	ENSG00000076242
MSH2	ENSG00000095002
MSH3	ENSG00000113318
MSH6	ENSG00000116062
PMS1	ENSG00000064933
PMS2	ENSG00000122512

Table 6.3: MSI and MTM indices

Case ID	MSI index_ngs	MSI status_ngs	MSI status_pcr	MTM index
GCa#1	0.014572539	MSS_ngs	MSS_pcr	0
GCa#2	0.01003886	MSS_ngs	MSS_pcr	0.0625
GCa#3	0.754533679	MSI_ngs	MSI_pcr	1
GCa#4	0.019430052	MSS_ngs	Unknown_pcr	0.125
GCa#5	0.366256477	MSI_ngs	MSI_pcr	0.6875
GCa#6	0.04242228	MSS_ngs	MSS_pcr	0.25
GCa#7	0.00615285	MSS_ngs	Unknown_pcr	0
GCa#8	0.015220207	MSS_ngs	MSS_pcr	0
GCa#9	0.01003886	MSS_ngs	Unknown_pcr	0
GCa#10	1.35265544	MSI_ngs	Unknown_pcr	1.3125
GCa#11	0.009715026	MSS_ngs	MSS_pcr	0
GCa#12	0.01003886	MSS_ngs	Unknown_pcr	0
GCa#13	0.011658031	MSS_ngs	MSS_pcr	0.0625
GCa#14	0.958873057	MSI_ngs	MSI_pcr	1.25
GCa#15	0.007124352	MSS_ngs	Unknown_pcr	0
GCa#16	0.010362694	MSS_ngs	MSS_pcr	0
GCa#17	0.569624352	MSI_ngs	MSI_pcr	0.375
GCa#18	0.761334197	MSI_ngs	MSI_pcr	0.875
GCa#19	0.009067358	MSS_ngs	MSS_pcr	0
GCa#20	0.007772021	MSS_ngs	MSS_pcr	0.0625
GCa#21	0.792098446	MSI_ngs	MSI_pcr	0.6875
GCa#22	1.12111399	MSI_ngs	MSI_pcr	0.9375
GCa#23	0.007772021	MSS_ngs	MSS_pcr	0
GCa#24	0.020401554	MSS_ngs	MSS_pcr	0
GCa#25	1.435556995	MSI_ngs	MSI_pcr	1.75
GCa#26	1.129857513	MSI_ngs	MSI_pcr	0.8125
GCa#27	0.643782383	MSI_ngs	MSI_pcr	0.6875
GCa#28	0.012629534	MSS_ngs	MSS_pcr	0
GCa#29	1.032059585	MSI_ngs	MSI_pcr	1.1875
GCa#30	0.014572539	MSS_ngs	MSS_pcr	0.0625
GCa#31	0.86107513	MSI_ngs	MSI_pcr	1.1875
GCa#32	0.016839378	MSS_ngs	MSS_pcr	0.0625
GCa#33	1.452720207	MSI_ngs	MSI_pcr	1.875
GCa#34	0.010362694	MSS_ngs	MSS_pcr	0
GCa#35	0.009067358	MSS_ngs	MSS_pcr	0.0625
GCa#36	0.013601036	MSS_ngs	Unknown_pcr	0
GCa#37	0.008419689	MSS_ngs	MSS_pcr	0
GCa#38	0.007124352	MSS_ngs	MSS_pcr	0
GCa#39	0.017487047	MSS_ngs	MSS_pcr	0.0625
GCa#40	0.009067358	MSS_ngs	MSS_pcr	0

GCa#41	0.006476684	MSS_ngs	MSS_pcr	0
GCa#42	0.009715026	MSS_ngs	MSS_pcr	0
GCa#43	0.584520725	MSI_ngs	MSI_pcr	0.625
GCa#44	0.009391192	MSS_ngs	MSS_pcr	0
GCa#45	0.009067358	MSS_ngs	Unknown_pcr	0.0625
GCa#46	0.008095855	MSS_ngs	Unknown_pcr	0
GCa#47	0.007448187	MSS_ngs	Unknown_pcr	0
GCa#48	0.006800518	MSS_ngs	Unknown_pcr	0
GCa#49	0.006800518	MSS_ngs	MSS_pcr	0
GCa#50	0.229274611	MSI_ngs	MSI_pcr	0.375
GCa#51	0.006476684	MSS_ngs	MSS_pcr	0
GCa#52	0.580634715	MSI_ngs	MSI_pcr	0.75
GCa#53	0.006800518	MSS_ngs	Unknown_pcr	0
GCa#54	0.009391192	MSS_ngs	MSS_pcr	0
GCa#55	0.272992228	MSI_ngs	MSI_pcr	0.4375
GCa#56	0.011010363	MSS_ngs	Unknown_pcr	0.0625
GCa#57	0.44753886	MSI_ngs	MSI_pcr	0.6875
GCa#58	0.009391192	MSS_ngs	MSS_pcr	0.125
GCa#59	0.005181347	MSS_ngs	MSS_pcr	0.0625
GCa#60	0.005829016	MSS_ngs	MSS_pcr	0.0625
GCa#61	0.007124352	MSS_ngs	MSS_pcr	0
GCa#62	0.007772021	MSS_ngs	MSS_pcr	0
GCa#63	0.475712435	MSI_ngs	MSI_pcr	0.6875
GCa#64	0.007448187	MSS_ngs	MSS_pcr	0
GCa#65	0.002914508	MSS_ngs	MSS_pcr	0
GCa#66	0.007124352	MSS_ngs	Unknown_pcr	0.0625
GCa#67	0.016515544	MSS_ngs	MSS_pcr	0.125
GCa#68	0.002590674	MSS_ngs	MSS_pcr	0
GCa#69	0.268458549	MSI_ngs	MSI_pcr	0.3125
GCa#70	0.01392487	MSS_ngs	Unknown_pcr	0.1875
GCa#71	0.007772021	MSS_ngs	MSS_pcr	0
GCa#72	1.490608808	MSI_ngs	MSI_pcr	1
GCa#73	0.008419689	MSS_ngs	MSS_pcr	0
GCa#74	0.524935233	MSI_ngs	MSI_pcr	0.8125
GCa#75	0.122085492	MSS_ngs	Unknown_pcr	0.3125
GCa#76	0.006800518	MSS_ngs	Unknown_pcr	0
GCa#77	0.01392487	MSS_ngs	MSS_pcr	0
GCa#78	0.012305699	MSS_ngs	MSS_pcr	0
GCa#79	0.002590674	MSS_ngs	MSS_pcr	0
GCa#80	0.009391192	MSS_ngs	Unknown_pcr	0
GCa#81	0.008419689	MSS_ngs	Unknown_pcr	0
GCa#82	0.006800518	MSS_ngs	MSS_pcr	0
GCa#83	0.8125	MSI_ngs	Unknown_pcr	0.9375

GCa#84	1.159326425	MSI_ngs	MSI_pcr	0.9375
GCa#85	0.045012953	MSS_ngs	Unknown_pcr	0.1875
GCa#86	0.62888601	MSI_ngs	MSI_pcr	0.875
GCa#87	0.002914508	MSS_ngs	MSS_pcr	0
GCa#88	0.010686528	MSS_ngs	MSS_pcr	0
GCa#89	0.005505181	MSS_ngs	Unknown_pcr	0.0625
GCa#90	0.008743523	MSS_ngs	MSS_pcr	0.0625
GCa#91	0.870466321	MSI_ngs	MSI_pcr	0.75
GCa#92	0.007772021	MSS_ngs	MSS_pcr	0
GCa#93	0.014248705	MSS_ngs	Unknown_pcr	0.0625
GCa#94	0.010362694	MSS_ngs	MSS_pcr	0
GCa#95	0.974740933	MSI_ngs	MSI_pcr	0.8125
GCa#96	0.018782383	MSS_ngs	MSS_pcr	0
GCa#97	0.004209845	MSS_ngs	MSS_pcr	0
GCa#98	0.011334197	MSS_ngs	Unknown_pcr	0.0625
GCa#99	0.495142487	MSI_ngs	MSI_pcr	0.8125
GCa#100	0.008095855	MSS_ngs	MSS_pcr	0.0625
GCa#101	0.019430052	MSS_ngs	MSS_pcr	0
GCa#102	0.006800518	MSS_ngs	MSS_pcr	0
GCa#103	0.014896373	MSS_ngs	MSS_pcr	0.0625
GCa#104	0.008743523	MSS_ngs	MSS_pcr	0
GCa#105	1.353950777	MSI_ngs	MSI_pcr	1.875
GCa#106	0.008095855	MSS_ngs	MSS_pcr	0.125
GCa#107	0.022668394	MSS_ngs	Unknown_pcr	0.3125
GCa#108	0.014572539	MSS_ngs	MSS_pcr	0
GCa#109	1.037564767	MSI_ngs	MSI_pcr	0.8125
GCa#110	0.003562176	MSS_ngs	MSS_pcr	0
GCa#111	0.009067358	MSS_ngs	Unknown_pcr	0
GCa#112	0.006476684	MSS_ngs	MSS_pcr	0
GCa#113	0.015544041	MSS_ngs	MSS_pcr	0
GCa#114	0.012629534	MSS_ngs	MSS_pcr	0.125
GCa#115	0.063795337	MSS_ngs	Unknown_pcr	1.625
GCa#116	0.009715026	MSS_ngs	MSS_pcr	0
GCa#117	0.007448187	MSS_ngs	MSS_pcr	0
GCa#118	0.005505181	MSS_ngs	Unknown_pcr	0.0625
GCa#119	0.010362694	MSS_ngs	MSS_pcr	0
GCa#120	0.014572539	MSS_ngs	Unknown_pcr	0.0625
GCa#121	0.004857513	MSS_ngs	MSS_pcr	0
GCa#122	0.01003886	MSS_ngs	MSS_pcr	0
GCa#123	0.011658031	MSS_ngs	MSS_pcr	0
GCa#124	0.012953368	MSS_ngs	MSS_pcr	0.0625
GCa#125	0.013277202	MSS_ngs	Unknown_pcr	0.0625
GCa#126	0.014248705	MSS_ngs	MSS_pcr	0.0625

GCa#127	0.012629534	MSS_ngs	MSS_pcr	0.0625
GCa#128	0.005505181	MSS_ngs	MSS_pcr	0
GCa#129	0.002266839	MSS_ngs	MSS_pcr	0
GCa#130	0.008419689	MSS_ngs	MSS_pcr	0.0625
GCa#131	0.00615285	MSS_ngs	MSS_pcr	0
GCa#132	0.014248705	MSS_ngs	MSS_pcr	0.0625
GCa#133	0.008419689	MSS_ngs	MSS_pcr	0
GCa#134	0.012953368	MSS_ngs	Unknown_pcr	0
GCa#135	0.011334197	MSS_ngs	MSS_pcr	0
GCa#136	0.006476684	MSS_ngs	MSS_pcr	0
GCa#137	0.00388601	MSS_ngs	MSS_pcr	0
GCa#138	0.01003886	MSS_ngs	MSS_pcr	0
GCa#139	0.008095855	MSS_ngs	MSS_pcr	0
GCa#140	0.010686528	MSS_ngs	MSS_pcr	0
GCa#141	0.018782383	MSS_ngs	MSS_pcr	0
GCa#142	0.01003886	MSS_ngs	MSS_pcr	0
GCa#143	0.009715026	MSS_ngs	MSS_pcr	0
GCa#144	1.159974093	MSI_ngs	Unknown_pcr	1.0625
GCa#145	0.002590674	MSS_ngs	Unknown_pcr	0
GCa#146	0.011010363	MSS_ngs	MSS_pcr	0
GCa#147	0.009067358	MSS_ngs	MSS_pcr	0.0625
GCa#148	0.035621762	MSS_ngs	MSS_pcr	0
GCa#149	0.011981865	MSS_ngs	Unknown_pcr	0
GCa#150	0.011334197	MSS_ngs	Unknown_pcr	0
GCa#151	0.009067358	MSS_ngs	MSS_pcr	0
GCa#152	0.007448187	MSS_ngs	MSS_pcr	0
GCa#153	0.019106218	MSS_ngs	Unknown_pcr	0
GCa#154	0.017487047	MSS_ngs	MSS_pcr	0.0625
GCa#155	0.013601036	MSS_ngs	Unknown_pcr	0
GCa#156	0.010686528	MSS_ngs	MSS_pcr	0
GCa#157	0.047279793	MSS_ngs	Unknown_pcr	0
GCa#158	0.016839378	MSS_ngs	MSS_pcr	0.0625
GCa#159	0.01392487	MSS_ngs	Unknown_pcr	0.0625
GCa#160	0.009715026	MSS_ngs	MSS_pcr	0.0625
GCa#161	0.017810881	MSS_ngs	MSS_pcr	0.125
GCa#162	0.007124352	MSS_ngs	MSS_pcr	0
GCa#163	0.010362694	MSS_ngs	Unknown_pcr	0.125
GCa#164	0.009715026	MSS_ngs	MSS_pcr	0
GCa#165	0.016839378	MSS_ngs	MSS_pcr	0
GCa#166	0.657707254	MSI_ngs	MSI_pcr	0.4375
GCa#167	0.010362694	MSS_ngs	MSS_pcr	0
GCa#168	0.013601036	MSS_ngs	MSS_pcr	0
GCa#169	0.007448187	MSS_ngs	MSS_pcr	0

GCa#170	0.018134715	MSS_ngs	MSS_pcr	0.0625
GCa#171	0.014572539	MSS_ngs	MSS_pcr	0
GCa#172	1.121761658	MSI_ngs	MSI_pcr	1.3125
GCa#173	0.009391192	MSS_ngs	MSS_pcr	0.125
GCa#174	0.379533679	MSI_ngs	MSI_pcr	0.5
GCa#175	0.52007772	MSI_ngs	MSI_pcr	0.5
GCa#176	0.013277202	MSS_ngs	Unknown_pcr	0.0625
GCa#177	0.011658031	MSS_ngs	MSS_pcr	0
GCa#178	0.011981865	MSS_ngs	MSS_pcr	0
GCa#179	0.975712435	MSI_ngs	MSI_pcr	1.125
GCa#180	0.948510363	MSI_ngs	MSI_pcr	1
GCa#181	0.013277202	MSS_ngs	MSS_pcr	0
GCa#182	0.022668394	MSS_ngs	Unknown_pcr	0
GCa#183	0.02234456	MSS_ngs	MSS_pcr	0
GCa#184	0.01392487	MSS_ngs	MSS_pcr	0
GCa#185	0.010686528	MSS_ngs	MSS_pcr	0
GCa#186	1.436204663	MSI_ngs	MSI_pcr	1.1875
GCa#187	0.006800518	MSS_ngs	MSS_pcr	0
GCa#188	0.013601036	MSS_ngs	MSS_pcr	0
GCa#189	0.01619171	MSS_ngs	MSS_pcr	0
GCa#190	0.01003886	MSS_ngs	Unknown_pcr	0
GCa#191	0.997085492	MSI_ngs	Unknown_pcr	0.875
GCa#192	0.010686528	MSS_ngs	MSS_pcr	0
GCa#193	0.600712435	MSI_ngs	MSI_pcr	0.375
GCa#194	0.019430052	MSS_ngs	MSS_pcr	0
GCa#195	0.024287565	MSS_ngs	MSS_pcr	0.0625
GCa#196	0.014896373	MSS_ngs	MSS_pcr	0
GCa#197	0.008095855	MSS_ngs	MSS_pcr	0.0625
GCa#198	0.007124352	MSS_ngs	MSS_pcr	0
GCa#199	0.028497409	MSS_ngs	MSS_pcr	0
GCa#200	0.006476684	MSS_ngs	Unknown_pcr	0
GCa#201	0.008743523	MSS_ngs	MSS_pcr	0
GCa#202	0.018458549	MSS_ngs	MSS_pcr	0
GCa#203	0.010362694	MSS_ngs	MSS_pcr	0
GCa#204	0.009067358	MSS_ngs	MSS_pcr	0
GCa#205	1.76392487	MSI_ngs	MSI_pcr	2.0625
GCa#206	0.009391192	MSS_ngs	Unknown_pcr	0
GCa#207	1.413860104	MSI_ngs	MSI_pcr	1.125
GCa#208	0.016515544	MSS_ngs	MSS_pcr	0
GCa#209	0.039183938	MSS_ngs	Unknown_pcr	0.0625
GCa#210	1.668069948	MSI_ngs	MSI_pcr	1.6875
GCa#211	0.021696891	MSS_ngs	Unknown_pcr	0.0625
GCa#212	1.369170984	MSI_ngs	Unknown_pcr	1.3125

GCa#213	0.020401554	MSS_ngs	Unknown_pcr	0
GCa#214	0.011010363	MSS_ngs	MSS_pcr	0.0625
GCa#215	1.333225389	MSI_ngs	Unknown_pcr	0.75
GCa#216	0.01619171	MSS_ngs	MSS_pcr	0.0625
GCa#217	0.011334197	MSS_ngs	Unknown_pcr	0
GCa#218	0.021373057	MSS_ngs	MSS_pcr	0.0625
GCa#219	0.012953368	MSS_ngs	Unknown_pcr	0.125
GCa#220	0.011658031	MSS_ngs	MSS_pcr	0
GCa#221	0.006476684	MSS_ngs	MSS_pcr	0
GCa#222	0.014248705	MSS_ngs	MSS_pcr	0
GCa#223	0.007772021	MSS_ngs	MSS_pcr	0.0625
GCa#224	0.008095855	MSS_ngs	MSS_pcr	0.0625
GCa#225	0.007448187	MSS_ngs	MSS_pcr	0.0625
GCa#226	0.01619171	MSS_ngs	MSS_pcr	0
GCa#227	0.014248705	MSS_ngs	MSS_pcr	0.125
GCa#228	0.014896373	MSS_ngs	MSS_pcr	0
GCa#229	0.022668394	MSS_ngs	Unknown_pcr	0.0625
GCa#230	0.01619171	MSS_ngs	MSS_pcr	0
GCa#231	0.008095855	MSS_ngs	MSS_pcr	0.0625
GCa#232	0.009715026	MSS_ngs	MSS_pcr	0
GCa#233	0.009391192	MSS_ngs	MSS_pcr	0
GCa#234	0.018458549	MSS_ngs	MSS_pcr	0.0625
GCa#235	0.010686528	MSS_ngs	Unknown_pcr	0
GCa#236	0.012629534	MSS_ngs	MSS_pcr	0
GCa#237	0.01392487	MSS_ngs	Unknown_pcr	0
GCa#238	0.006476684	MSS_ngs	MSS_pcr	0
GCa#239	0.011010363	MSS_ngs	MSS_pcr	0
GCa#240	0.012953368	MSS_ngs	MSS_pcr	0.0625
GCa#241	0.012629534	MSS_ngs	MSS_pcr	0.0625
GCa#242	0.007448187	MSS_ngs	MSS_pcr	0
GCa#243	0.010362694	MSS_ngs	MSS_pcr	0.1875
GCa#244	0.008419689	MSS_ngs	MSS_pcr	0
GCa#245	0.546955959	MSI_ngs	MSI_pcr	0.5
GCa#246	0.009391192	MSS_ngs	Unknown_pcr	0.0625
GCa#247	0.273963731	MSI_ngs	MSI_pcr	0.3125
GCa#248	0.012305699	MSS_ngs	MSS_pcr	0.125
GCa#249	0.009391192	MSS_ngs	Unknown_pcr	0
GCa#250	0.017487047	MSS_ngs	MSS_pcr	0.0625
GCa#251	0.008743523	MSS_ngs	MSS_pcr	0
GCa#252	0.66515544	MSI_ngs	MSI_pcr	0.25
GCa#253	0.012305699	MSS_ngs	Unknown_pcr	0.125
GCa#254	0.012305699	MSS_ngs	MSS_pcr	0
GCa#255	0.005505181	MSS_ngs	MSS_pcr	0

GCa#256	0.009715026	MSS_ngs	MSS_pcr	0
GCa#257	0.009715026	MSS_ngs	MSS_pcr	0
GCa#258	0.015867876	MSS_ngs	MSS_pcr	0.0625
GCa#259	0.006476684	MSS_ngs	Unknown_pcr	0
GCa#260	0.011010363	MSS_ngs	Unknown_pcr	0
GCa#261	0.013277202	MSS_ngs	MSS_pcr	0.0625
GCa#262	0.018134715	MSS_ngs	MSS_pcr	0
GCa#263	1.24384715	MSI_ngs	MSI_pcr	1.5
GCa#264	0.608484456	MSI_ngs	MSI_pcr	0.4375
GCa#265	0.00615285	MSS_ngs	MSS_pcr	0
GCa#266	0.255505181	MSI_ngs	MSI_pcr	0.5625
GCa#267	0.007124352	MSS_ngs	MSS_pcr	0
GCa#268	0.004209845	MSS_ngs	MSS_pcr	0
GCa#269	0.014248705	MSS_ngs	MSS_pcr	0.0625
GCa#270	0.003238342	MSS_ngs	Unknown_pcr	0.0625
GCa#271	0.005181347	MSS_ngs	Unknown_pcr	0
GCa#272	0.013277202	MSS_ngs	MSS_pcr	0.0625
GCa#273	0.004857513	MSS_ngs	MSS_pcr	0
GCa#274	0.005505181	MSS_ngs	MSS_pcr	0.0625
GCa#275	0.01619171	MSS_ngs	MSS_pcr	0
GCa#276	0.011010363	MSS_ngs	MSS_pcr	0.0625
GCa#277	0.722150259	MSI_ngs	Unknown_pcr	0.8125
GCa#278	0.009067358	MSS_ngs	MSS_pcr	0
GCa#279	0.008419689	MSS_ngs	MSS_pcr	0
GCa#280	0.010686528	MSS_ngs	MSS_pcr	0
GCa#281	0.006800518	MSS_ngs	Unknown_pcr	0.125
GCa#282	0.412564767	MSI_ngs	MSI_pcr	0.375
GCa#283	0.006476684	MSS_ngs	MSS_pcr	0.0625
GCa#284	0.013277202	MSS_ngs	Unknown_pcr	0.0625
GCa#285	0.008743523	MSS_ngs	MSS_pcr	0
GCa#286	1.177137306	MSI_ngs	MSI_pcr	1.3125
GCa#287	0.208225389	MSI_ngs	Unknown_pcr	0.5
GCa#288	0.019430052	MSS_ngs	Unknown_pcr	0
GCa#289	0.017487047	MSS_ngs	Unknown_pcr	0
GCa#290	0.015220207	MSS_ngs	Unknown_pcr	0
GCa#291	0.015220207	MSS_ngs	Unknown_pcr	0
GCa#292	0.015220207	MSS_ngs	MSS_pcr	0
GCa#293	1.117551813	MSI_ngs	MSI_pcr	1.3125
GCa#294	0.012305699	MSS_ngs	MSS_pcr	0.0625
GCa#295	0.015867876	MSS_ngs	Unknown_pcr	0
GCa#296	0.009067358	MSS_ngs	MSS_pcr	0
GCa#297	0.00615285	MSS_ngs	Unknown_pcr	0
GCa#298	0.008419689	MSS_ngs	MSS_pcr	0

GCa#299	0.011010363	MSS_ngs	Unknown_pcr	0
GCa#300	0.749028497	MSI_ngs	MSI_pcr	0.8125
GCa#301	0.700777202	MSI_ngs	MSI_pcr	0.4375
GCa#302	0.006800518	MSS_ngs	MSS_pcr	0.0625
GCa#303	0.012305699	MSS_ngs	MSS_pcr	0
GCa#304	0.014248705	MSS_ngs	MSS_pcr	0
GCa#305	0.010362694	MSS_ngs	Unknown_pcr	0.0625
GCa#306	0.006476684	MSS_ngs	Unknown_pcr	0
GCa#307	0.601036269	MSI_ngs	MSI_pcr	0.5625
GCa#308	0.015544041	MSS_ngs	MSS_pcr	0
GCa#309	0.949158031	MSI_ngs	MSI_pcr	1
GCa#310	0.896373057	MSI_ngs	MSI_pcr	0.6875
GCa#311	0.553108808	MSI_ngs	MSI_pcr	0.625
GCa#312	0.60265544	MSI_ngs	MSI_pcr	0.75
GCa#313	0.005181347	MSS_ngs	MSS_pcr	0
GCa#314	0.004857513	MSS_ngs	Unknown_pcr	0
GCa#315	0.010686528	MSS_ngs	MSS_pcr	0
GCa#316	0.712759067	MSI_ngs	MSI_pcr	0.6875
GCa#317	0.193329016	MSI_ngs	MSI_pcr	0.3125
GCa#318	0.011658031	MSS_ngs	MSS_pcr	0
GCa#319	0.010686528	MSS_ngs	MSS_pcr	0
GCa#320	0.784974093	MSI_ngs	MSI_pcr	0.9375
GCa#321	0.007772021	MSS_ngs	Unknown_pcr	0
GCa#322	0.015220207	MSS_ngs	MSS_pcr	0.1875
GCa#323	0.009391192	MSS_ngs	MSS_pcr	0.1875
GCa#324	0.009067358	MSS_ngs	MSS_pcr	0
GCa#325	0.018458549	MSS_ngs	MSS_pcr	0.125
GCa#326	0.018134715	MSS_ngs	MSS_pcr	0.125
GCa#327	0.012629534	MSS_ngs	MSS_pcr	0
GCa#328	0.006800518	MSS_ngs	MSS_pcr	0
GCa#329	0.018782383	MSS_ngs	Unknown_pcr	0
GCa#330	0.008743523	MSS_ngs	MSS_pcr	0.0625
GCa#331	0.015867876	MSS_ngs	MSS_pcr	0
GCa#332	0.01392487	MSS_ngs	Unknown_pcr	0.0625
GCa#333	0.018458549	MSS_ngs	MSS_pcr	0.0625
GCa#334	0.008419689	MSS_ngs	MSS_pcr	0
GCa#335	0.01003886	MSS_ngs	MSS_pcr	0.0625
GCa#336	0.013277202	MSS_ngs	MSS_pcr	0
GCa#337	0.008095855	MSS_ngs	MSS_pcr	0
GCa#338	0.008095855	MSS_ngs	MSS_pcr	0
GCa#339	0.008743523	MSS_ngs	MSS_pcr	0.125
GCa#340	0.004857513	MSS_ngs	MSS_pcr	0
GCa#341	0.018782383	MSS_ngs	Unknown_pcr	0

GCa#342	0.008743523	MSS_ngs	MSS_pcr	0
GCa#343	0.007772021	MSS_ngs	Unknown_pcr	0
GCa#344	0.010362694	MSS_ngs	Unknown_pcr	0
GCa#345	0.031735751	MSS_ngs	MSS_pcr	0
GCa#346	0.007124352	MSS_ngs	MSS_pcr	0
GCa#347	0.78011658	MSI_ngs	MSI_pcr	0.875
GCa#348	0.009067358	MSS_ngs	MSS_pcr	0
GCa#349	0.011334197	MSS_ngs	MSS_pcr	0
GCa#350	0.012629534	MSS_ngs	MSS_pcr	0
GCa#351	0.006476684	MSS_ngs	MSS_pcr	0.0625
GCa#352	0.007124352	MSS_ngs	MSS_pcr	0
GCa#353	0.432966321	MSI_ngs	MSI_pcr	0.5
GCa#354	0.008419689	MSS_ngs	MSS_pcr	0
GCa#355	0.005181347	MSS_ngs	Unknown_pcr	0.0625
GCa#356	0.009715026	MSS_ngs	MSS_pcr	0.0625
GCa#357	0.009067358	MSS_ngs	MSS_pcr	0
GCa#358	0.01003886	MSS_ngs	Unknown_pcr	0
GCa#359	0.009715026	MSS_ngs	MSS_pcr	0
GCa#360	0.014572539	MSS_ngs	MSS_pcr	0.125
GCa#361	0.005505181	MSS_ngs	MSS_pcr	0
GCa#362	0.005505181	MSS_ngs	MSS_pcr	0
GCa#363	0.004209845	MSS_ngs	MSS_pcr	0
GCa#364	0.006800518	MSS_ngs	MSS_pcr	0.0625
GCa#365	0.00388601	MSS_ngs	MSS_pcr	0
GCa#366	0.008095855	MSS_ngs	MSS_pcr	0.0625
GCa#367	0.005829016	MSS_ngs	MSS_pcr	0.0625
GCa#368	0.012305699	MSS_ngs	MSS_pcr	0
GCa#369	0.012953368	MSS_ngs	MSS_pcr	0.0625
GCa#370	0.269106218	MSI_ngs	MSI_pcr	0.375
GCa#371	0.429404145	MSI_ngs	MSI_pcr	0.375
GCa#372	0.008095855	MSS_ngs	MSS_pcr	0.0625
GCa#373	0.025582902	MSS_ngs	Unknown_pcr	0
GCa#374	0.012305699	MSS_ngs	MSS_pcr	0
GCa#375	0.007772021	MSS_ngs	MSS_pcr	0.0625
GCa#376	0.012305699	MSS_ngs	MSS_pcr	0.0625
GCa#377	0.00615285	MSS_ngs	MSS_pcr	0
GCa#378	0.015867876	MSS_ngs	MSS_pcr	0
GCa#379	0.008095855	MSS_ngs	MSS_pcr	0
GCa#380	0.484132124	MSI_ngs	MSI_pcr	0.375
GCa#381	0.015544041	MSS_ngs	MSS_pcr	0.0625
GCa#382	0.951748705	MSI_ngs	MSI_pcr	1.5
GCa#383	0.011658031	MSS_ngs	MSS_pcr	0
GCa#384	0.013601036	MSS_ngs	MSS_pcr	0

GCa#385	0.006476684	MSS_ngs	MSS_pcr	0
GCa#386	0.625647668	MSI_ngs	MSI_pcr	0.625
GCa#387	0.012305699	MSS_ngs	MSS_pcr	0
GCa#388	0.011981865	MSS_ngs	MSS_pcr	0
GCa#389	0.008743523	MSS_ngs	MSS_pcr	0
GCa#390	0.008095855	MSS_ngs	Unknown_pcr	0
GCa#391	0.015544041	MSS_ngs	MSS_pcr	0
GCa#392	0.011981865	MSS_ngs	MSS_pcr	0.0625
GCa#393	0.01003886	MSS_ngs	MSS_pcr	0
GCa#394	0.007448187	MSS_ngs	MSS_pcr	0
GCa#395	0.730893782	MSI_ngs	MSI_pcr	0.625
GCa#396	0.932966321	MSI_ngs	MSI_pcr	0.8125
GCa#397	0.007124352	MSS_ngs	MSS_pcr	0
GCa#398	0.011981865	MSS_ngs	Unknown_pcr	0
GCa#399	1.008419689	MSI_ngs	Unknown_pcr	0.75
GCa#400	0.006800518	MSS_ngs	MSS_pcr	0.0625
GCa#401	0.295012953	MSI_ngs	MSI_pcr	0.375
GCa#402	0.01392487	MSS_ngs	Unknown_pcr	0
GCa#403	0.008419689	MSS_ngs	Unknown_pcr	0
GCa#404	0.012629534	MSS_ngs	MSS_pcr	0
GCa#405	1.112046632	MSI_ngs	MSI_pcr	1.25
GCa#406	0.360751295	MSI_ngs	Unknown_pcr	0.5625
GCa#407	0.758095855	MSI_ngs	MSI_pcr	0.875
GCa#408	0.008419689	MSS_ngs	MSS_pcr	0
GCa#409	0.012305699	MSS_ngs	MSS_pcr	0.125
GCa#410	0.008419689	MSS_ngs	MSS_pcr	0.0625
GCa#411	0.014896373	MSS_ngs	MSS_pcr	0.0625
GCa#412	0.013277202	MSS_ngs	MSS_pcr	0
GCa#413	0.011334197	MSS_ngs	Unknown_pcr	0.0625
GCa#414	0.021696891	MSS_ngs	MSS_pcr	0
GCa#415	0.007448187	MSS_ngs	MSS_pcr	0
GCa#416	0.014248705	MSS_ngs	MSS_pcr	0
GCa#417	0.016839378	MSS_ngs	MSS_pcr	0
GCa#418	0.007772021	MSS_ngs	Unknown_pcr	0
GCa#419	0.814119171	MSI_ngs	MSI_pcr	1
GCa#420	0.010686528	MSS_ngs	MSS_pcr	0
GCa#421	0.007772021	MSS_ngs	MSS_pcr	0
GCa#422	0.008419689	MSS_ngs	MSS_pcr	0
GCa#423	0.01392487	MSS_ngs	MSS_pcr	0
GCa#424	0.005829016	MSS_ngs	MSS_pcr	0
GCa#425	0.005829016	MSS_ngs	MSS_pcr	0
GCa#426	0.014572539	MSS_ngs	MSS_pcr	0
GCa#427	0.008419689	MSS_ngs	MSS_pcr	0

GCa#428	0.012953368	MSS_ngs	MSS_pcr	0
GCa#429	0.009067358	MSS_ngs	MSS_pcr	0
GCa#430	0.011658031	MSS_ngs	MSS_pcr	0
GCa#431	0.011334197	MSS_ngs	Unknown_pcr	0
GCa#432	0.01003886	MSS_ngs	MSS_pcr	0.0625
GCa#433	0.010362694	MSS_ngs	MSS_pcr	0
GCa#434	0.208873057	MSI_ngs	MSI_pcr	0.125
GCa#435	0.008743523	MSS_ngs	MSS_pcr	0
GCa#436	0.01003886	MSS_ngs	MSS_pcr	0
GCa#437	0.014572539	MSS_ngs	MSS_pcr	0
GCa#438	0.012305699	MSS_ngs	MSS_pcr	0.0625
GCa#439	0.011981865	MSS_ngs	MSS_pcr	0
GCa#440	0.009067358	MSS_ngs	MSS_pcr	0
GCa#441	0.008095855	MSS_ngs	MSS_pcr	0.0625

Table 6.4: DNA mismatch repair gene expression in FPKM

Case ID	MSI status_ngs	MLH1	MSH2	MSH3	MSH6	PMS1	PMS2
GCa#1	MSS_ngs	7.277185191	11.81631	2.705865	3.592925	3.91324	1.596363
GCa#2	MSS_ngs	6.332527466	4.963435	6.202378	2.707221	2.750859	1.784809
GCa#3	MSI_ngs						
GCa#4	MSS_ngs	7.488398581	4.278732	2.19947	3.325181	1.071119	2.965172
GCa#5	MSI_ngs	2.744127833	8.226434	2.174746	4.47364	2.496866	2.223008
GCa#6	MSS_ngs	0.839190965	7.79765	2.826238	3.242212	3.557917	2.095132
GCa#7	MSS_ngs	7.417842365	6.475472	3.107386	4.607448	2.592479	2.112661
GCa#8	MSS_ngs	4.323295307	5.020121	3.333125	2.48545	2.201457	2.422085
GCa#9	MSS_ngs						
GCa#10	MSI_ngs						
GCa#11	MSS_ngs	9.402150342	4.037074	6.33	3.908041	2.663968	2.742515
GCa#12	MSS_ngs						
GCa#13	MSS_ngs	8.892794218	6.852008	5.321327	4.064183	3.226021	3.389537
GCa#14	MSI_ngs	4.424050853	6.632377	2.491119	5.280374	4.395438	2.445123
GCa#15	MSS_ngs	18.13106659	11.58969	5.535544	8.888909	5.430243	3.839179
GCa#16	MSS_ngs						
GCa#17	MSI_ngs	5.655233149	6.310843	4.602429	5.716965	3.659537	2.56804
GCa#18	MSI_ngs	2.474634578	9.108967	6.35508	5.568372	2.178896	2.823815
GCa#19	MSS_ngs	9.799771767	6.06812	8.278705	4.404527	1.738893	3.339456
GCa#20	MSS_ngs	9.63601579	6.288654	6.10437	4.845156	3.085176	1.571483
GCa#21	MSI_ngs	1.263186765	8.054028	5.007988	5.849884	3.246505	2.476224
GCa#22	MSI_ngs						
GCa#23	MSS_ngs	12.44691273	3.629993	4.24415	5.209328	2.902569	2.339107
GCa#24	MSS_ngs	10.01666802	10.84827	3.949049	4.248315	4.877836	2.119165
GCa#25	MSI_ngs	3.293115884	12.45526	2.933357	5.218683	4.995636	2.413265
GCa#26	MSI_ngs						
GCa#27	MSI_ngs	3.850807908	9.673757	4.010458	5.09807	4.042792	3.56062
GCa#28	MSS_ngs	14.862551	8.562975	3.784167	3.98441	1.698839	2.428238
GCa#29	MSI_ngs	4.255411183	9.726867	4.59114	6.938524	2.734614	4.300056
GCa#30	MSS_ngs	10.36849723	13.65238	4.966743	6.104152	4.798085	5.487396
GCa#31	MSI_ngs	7.531951272	5.726436	3.913356	7.117686	5.683444	5.063133
GCa#32	MSS_ngs	7.078539877	8.854391	6.147837	4.563167	3.470795	4.281344
GCa#33	MSI_ngs	2.284353995	10.17766	5.789926	4.692728	4.477246	2.915011
GCa#34	MSS_ngs	7.442098804	3.636643	4.850958	4.552141	2.530388	3.848426
GCa#35	MSS_ngs	6.758642942	4.577206	2.675434	2.634792	1.563449	1.855653
GCa#36	MSS_ngs	16.26949423	11.30522	3.179412	5.188459	2.067484	3.214997
GCa#37	MSS_ngs	6.548375677	15.4942	3.569118	6.804729	3.405721	3.277106
GCa#38	MSS_ngs	7.04279399	2.041539	2.31589	8.050104	2.359645	3.54624
GCa#39	MSS_ngs	9.38323325	9.617666	3.917136	2.31645	0.814239	2.061459
GCa#40	MSS_ngs	5.97417572	2.574202	3.638082	8.458459	2.139194	4.87904

GCa#41	MSS_ngs	6.625810306	1.942894	2.943644	2.3015	1.541703	1.861604
GCa#42	MSS_ngs	7.295253302	14.00043	2.42695	2.706886	1.227311	1.620465
GCa#43	MSI_ngs	1.339111486	3.082621	3.339473	7.798898	2.333267	5.058622
GCa#44	MSS_ngs	7.20246044	3.911446	1.741858	2.622955	1.55173	2.024062
GCa#45	MSS_ngs		8.986134	2.23829	3.588786	1.993438	2.640332
GCa#46	MSS_ngs						
GCa#47	MSS_ngs	8.719880277	6.338521	3.502634	4.84493	2.889901	2.041928
GCa#48	MSS_ngs	8.755213248	1.631316	3.706573	8.967607	2.182174	4.066858
GCa#49	MSS_ngs	7.211700574	3.586495	2.410021	1.794531	1.131566	1.972056
GCa#50	MSI_ngs	5.332273685	2.223376	3.031814	1.93093	1.221524	3.131669
GCa#51	MSS_ngs	5.021556687	3.222255	2.009232	1.109589	1.59751	0.686808
GCa#52	MSI_ngs	6.866720792	6.946912	2.780681	2.745879	1.596795	2.219389
GCa#53	MSS_ngs	2.933878772	7.078675	2.373076	2.626059	2.53124	3.261978
GCa#54	MSS_ngs	8.360156286	7.586025	2.60196	5.32093	2.22387	2.625718
GCa#55	MSI_ngs	6.664542396	3.847718	3.001298	3.690768	3.035371	6.672569
GCa#56	MSS_ngs	9.355883653	2.987089	4.32362	3.132252	1.89277	2.358941
GCa#57	MSI_ngs	7.64370125	7.031884	3.640152	3.270866	2.250856	1.900992
GCa#58	MSS_ngs	1.083134439	8.673796	3.504591	6.982335	1.792872	2.238608
GCa#59	MSS_ngs	6.506843732	6.010885	3.15525	5.550897	4.231716	3.917107
GCa#60	MSS_ngs	6.659774959	3.743232	2.98701	3.854031	1.57363	3.703115
GCa#61	MSS_ngs	4.287081568	2.823568	1.763605	2.142652	1.613082	1.527905
GCa#62	MSS_ngs	5.521808458	15.8211	2.470306	3.089468	2.153883	2.046084
GCa#63	MSI_ngs	8.418369034	13.10886	5.907098	11.38108	3.505977	3.93725
GCa#64	MSS_ngs	1.842828352	5.01173	4.859777	6.310622	2.716671	3.897199
GCa#65	MSS_ngs	4.786970661	4.126373	2.587687	3.454621	2.046857	3.165531
GCa#66	MSS_ngs	6.087778303	7.019657	5.356888	3.867863	2.924823	3.243512
GCa#67	MSS_ngs	9.934885484	6.258146	2.719798	4.728607	2.656721	2.905308
GCa#68	MSS_ngs	5.065589692	5.154708	4.099241	4.17132	3.105027	2.69394
GCa#69	MSI_ngs	5.903825571	2.992114	5.308943	3.665806	2.262792	3.226835
GCa#70	MSS_ngs	6.923832014	6.558316	4.566023	1.75963	2.087095	3.04048
GCa#71	MSS_ngs	6.618742063	7.067418	3.816588	5.694705	3.60297	3.728587
GCa#72	MSI_ngs	4.490058528		3.278962	2.918562	2.271185	2.067727
GCa#73	MSS_ngs						
GCa#74	MSI_ngs	6.83350866	7.446252	4.813964	5.432046	2.046471	3.703585
GCa#75	MSS_ngs	3.301584861	8.261796	2.840284	6.365921	3.747639	2.255192
GCa#76	MSS_ngs	10.14770077	1.606192	6.728109	2.354338	3.211636	3.970886
GCa#77	MSS_ngs						
GCa#78	MSS_ngs	9.348212477	14.97767	4.257926	6.574329	3.292949	3.688587
GCa#79	MSS_ngs	3.208930549	6.691558	3.6764	3.684705	3.15407	2.945003
GCa#80	MSS_ngs	6.029950588	5.058795	4.714883	3.28722	2.543074	2.298273
GCa#81	MSS_ngs	7.312255112	16.30217	2.042236	12.65865	3.041205	2.985835
GCa#82	MSS_ngs	6.940599409	7.76534	3.428617	3.990678	3.082782	4.092263
GCa#83	MSI_ngs	6.406222816	5.773854	2.74047	3.836476	2.792103	3.369508

GCa#84	MSI_ngs						
GCa#85	MSS_ngs	1.102447793	11.36521	2.74586	7.286983	5.042539	6.727172
GCa#86	MSI_ngs						
GCa#87	MSS_ngs						
GCa#88	MSS_ngs	4.950934173	3.839056	3.848062	3.86507	2.222931	3.295406
GCa#89	MSS_ngs	7.285036101	4.14782	3.960947	3.961474	2.265149	2.174536
GCa#90	MSS_ngs	8.612430188	5.472909	4.252688	4.347987	2.161035	2.26486
GCa#91	MSI_ngs	6.530681252	6.19702	9.241166	4.286307	2.892123	3.805624
GCa#92	MSS_ngs	1.004727022	12.09778	4.518861	4.87964	1.765323	4.080065
GCa#93	MSS_ngs	3.588990874	6.996679	4.187452	4.477842	3.089187	3.111064
GCa#94	MSS_ngs						
GCa#95	MSI_ngs	8.003195366	3.047432	3.387395	3.013182	2.247309	2.467796
GCa#96	MSS_ngs	2.370284773	11.08749	7.190373	3.129655	6.171121	2.994672
GCa#97	MSS_ngs	4.728982483	7.122807	3.104251	5.235064	3.243998	4.149219
GCa#98	MSS_ngs	7.59404055	5.397811	4.518705	4.131145	2.352788	2.902958
GCa#99	MSI_ngs	8.77334953	10.2085	6.210615	5.690485	3.867391	4.935117
GCa#100	MSS_ngs	1.600513485	5.442101	2.745024	2.355012	3.264903	2.909058
GCa#101	MSS_ngs	7.195790328	3.634688	4.099859	3.840765	2.745592	2.728705
GCa#102	MSS_ngs	4.895495882	8.825672	3.091019	4.321032	4.339327	4.014517
GCa#103	MSS_ngs	6.463684532	8.800625	4.035575	5.19072	3.345361	2.968166
GCa#104	MSS_ngs	7.10843112	8.122876	5.125191	6.44359	3.768728	2.8918
GCa#105	MSI_ngs	6.562929666	7.12708	3.094998	4.249944	2.127411	4.3163
GCa#106	MSS_ngs	4.792914045	6.306013	2.488488	2.405057	3.004198	2.996892
GCa#107	MSS_ngs	6.005885524	7.631394	3.723258	3.959227	2.270431	1.958012
GCa#108	MSS_ngs	12.20284091	11.8142	5.485469	7.460866	4.683347	3.540931
GCa#109	MSI_ngs	9.456994134	5.389494	5.070458	4.234243	2.816876	4.904634
GCa#110	MSS_ngs	1.896637948	8.706997	1.997549	3.603095	2.596884	3.888355
GCa#111	MSS_ngs	7.740480425	4.053768	4.861065	3.645962	2.858936	3.004979
GCa#112	MSS_ngs	7.502312344	8.801788	3.287003	3.030456	4.887218	2.338489
GCa#113	MSS_ngs	7.211114491	5.803571	4.530571	4.684111	3.050625	3.553334
GCa#114	MSS_ngs	11.56548135	19.85614	4.56424	11.69281	4.642371	3.746859
GCa#115	MSS_ngs	6.167265862	9.375216	6.269817	5.24795	4.364599	3.396291
GCa#116	MSS_ngs	4.865739602	10.45422	5.240758	8.156751	3.197019	4.004901
GCa#117	MSS_ngs	6.779069345	4.271489	4.302104	2.701969	2.545874	3.919062
GCa#118	MSS_ngs	5.979316107	7.929864	4.333601	4.823057	3.19732	4.105755
GCa#119	MSS_ngs	9.207228781	7.348392	4.560055	4.660382	2.99492	3.307735
GCa#120	MSS_ngs	4.479226258	5.446	4.459373	4.188988	1.398527	3.578806
GCa#121	MSS_ngs	4.238802304	6.033361	3.633516	2.569053	2.061048	2.309403
GCa#122	MSS_ngs	7.65904498	4.018181	2.248218	3.284002	3.286778	2.270622
GCa#123	MSS_ngs	3.165685096	5.464302	1.949268	2.057391	3.3448	1.883776
GCa#124	MSS_ngs						
GCa#125	MSS_ngs						
GCa#126	MSS_ngs						

GCa#127	MSS_ngs	9.688257776	19.58296	5.411985	11.44073	5.201973	2.40389
GCa#128	MSS_ngs	5.382904273	5.470912	1.902157	3.209323	4.539195	6.531233
GCa#129	MSS_ngs						
GCa#130	MSS_ngs	7.500737552	2.7878	3.67328	3.892104	2.705177	1.77088
GCa#131	MSS_ngs						
GCa#132	MSS_ngs						
GCa#133	MSS_ngs						
GCa#134	MSS_ngs						
GCa#135	MSS_ngs	5.20545726	7.226166	5.222489	4.75262	2.619757	9.125125
GCa#136	MSS_ngs	7.60326461	3.997611	3.753542	2.997686	2.706317	2.19644
GCa#137	MSS_ngs	5.854766939	6.579384	3.934782	3.016188	4.779146	2.420937
GCa#138	MSS_ngs	6.07852918	3.567585	3.507092	3.25657	2.864012	3.037626
GCa#139	MSS_ngs	4.624459837	4.370205	2.341046	2.3511	2.847311	1.953912
GCa#140	MSS_ngs	7.516488816	2.250627	2.864568	2.819102	2.345797	2.165602
GCa#141	MSS_ngs						
GCa#142	MSS_ngs						
GCa#143	MSS_ngs	7.188362589	8.210175	8.312404	6.376733	3.546409	3.096568
GCa#144	MSI_ngs						
GCa#145	MSS_ngs	0.480030215	11.29297	2.827878	1.104255	2.515606	1.543552
GCa#146	MSS_ngs						
GCa#147	MSS_ngs	5.615686223	2.380059	2.605108	2.324355	1.523592	1.143366
GCa#148	MSS_ngs	6.443689522	4.834004	2.653402	3.5548	1.939451	1.653302
GCa#149	MSS_ngs	8.788658184	7.349131	1.986512	3.427065	1.753939	3.100782
GCa#150	MSS_ngs	5.156382082	4.961402	2.058455	4.552008	3.801506	2.177095
GCa#151	MSS_ngs	9.461290488	6.677344				
GCa#152	MSS_ngs	5.516626787	3.611291	4.092329	4.223147	2.086156	2.581543
GCa#153	MSS_ngs	7.093723748	4.432007	1.676668	2.651534	1.021855	2.428492
GCa#154	MSS_ngs	5.625442321	5.371991	3.204885	2.698523	2.965017	3.935444
GCa#155	MSS_ngs	5.873771535	4.849885	2.856864	3.155819	2.279392	2.32244
GCa#156	MSS_ngs	7.372805465	12.26085	3.79441	8.668313	4.570958	3.856746
GCa#157	MSS_ngs	1.20572511	3.860729	3.347352	3.546144	1.584593	1.272902
GCa#158	MSS_ngs	2.914836435	8.638953	4.811498	8.295777	5.768242	4.251269
GCa#159	MSS_ngs	6.715720353	5.079075	3.614894	3.675756	2.212794	4.253932
GCa#160	MSS_ngs	5.22494563	4.278104	3.476469	4.35166	2.534352	3.477444
GCa#161	MSS_ngs	11.15329526	9.406922	10.91261	11.55534	2.255875	7.709193
GCa#162	MSS_ngs	8.941127479	5.805419	5.373939	3.308954	2.201843	3.897834
GCa#163	MSS_ngs	9.613777725	22.40472	8.010998	12.27448	4.525462	9.746359
GCa#164	MSS_ngs	8.228844061	7.15114	4.058753	4.701188	2.028667	5.107909
GCa#165	MSS_ngs	13.48376734	8.003843	3.468003	4.915438	5.422022	6.830908
GCa#166	MSI_ngs						
GCa#167	MSS_ngs	5.092055392	4.684453	1.835657	2.123802	2.304676	3.101477
GCa#168	MSS_ngs	5.662758081	5.979514	3.092227	2.719611	3.66739	2.269675
GCa#169	MSS_ngs	8.351492562	4.438877	3.243393	3.303435	3.411051	2.768719

GCa#170	MSS_ngs	6.668042619	3.868041	3.930449	2.61062	3.492729	3.187741
GCa#171	MSS_ngs	7.677963998	6.051066	2.014142	3.638564	2.793096	2.476965
GCa#172	MSI_ngs	1.630485397	7.753137	2.57323	3.371524	2.741195	2.749123
GCa#173	MSS_ngs	3.547244956	6.552938	2.46987	3.127314	2.342544	3.209822
GCa#174	MSI_ngs						
GCa#175	MSI_ngs						
GCa#176	MSS_ngs						
GCa#177	MSS_ngs	8.377163237	5.772171	4.775206	4.249156	3.454511	2.249329
GCa#178	MSS_ngs	9.22334932	6.274788	2.711162	3.844162	3.615732	3.843233
GCa#179	MSI_ngs	3.88721609	7.540448	3.607143	4.673935	3.444172	4.740406
GCa#180	MSI_ngs	3.083781477	5.959427	1.977242	7.315175	4.619236	2.436226
GCa#181	MSS_ngs	8.816527378	7.184867	5.718253	5.468631	5.846473	4.736823
GCa#182	MSS_ngs	5.969911495	4.185452	1.698023	3.164514	2.206976	2.518059
GCa#183	MSS_ngs	8.734539134	9.872618	4.479919	4.309904	4.003678	2.778214
GCa#184	MSS_ngs	11.47990876	15.73937	2.963195	6.491771	5.665185	2.837705
GCa#185	MSS_ngs	9.437489907	6.771492	4.049403	4.786369	2.404021	2.256984
GCa#186	MSI_ngs	1.43699579	7.781673	5.516802	7.255287	5.554017	2.933948
GCa#187	MSS_ngs	9.841795788	8.739519	2.883678	3.727977	4.507428	2.081368
GCa#188	MSS_ngs	7.567659264	22.91229	3.344543	6.394992	3.467132	2.542731
GCa#189	MSS_ngs						
GCa#190	MSS_ngs						
GCa#191	MSI_ngs	10.37286371	6.034678	4.791667	5.888817	4.817067	2.764486
GCa#192	MSS_ngs	10.12940909	3.884524	3.422245	3.898236	2.240474	2.74074
GCa#193	MSI_ngs	2.532955045	10.51584	7.798605	6.064856	4.271115	2.358336
GCa#194	MSS_ngs	8.673552221	12.09056	2.133456	6.974734	6.571436	2.39262
GCa#195	MSS_ngs	10.50451511	10.85558	5.393055	6.086428	6.629634	5.135942
GCa#196	MSS_ngs						
GCa#197	MSS_ngs						
GCa#198	MSS_ngs	11.45294009	10.07405	3.394698	8.52535	3.886599	2.584834
GCa#199	MSS_ngs	10.07048843	10.50995	4.278367	5.234538	3.168955	4.450469
GCa#200	MSS_ngs	9.991954489	10.45145	3.412321	8.356748	6.522173	3.158496
GCa#201	MSS_ngs	12.10547623	4.723294	3.344178	3.766184	2.087395	2.020225
GCa#202	MSS_ngs	7.554798176	4.871843	3.548473	3.309446	1.986023	2.48852
GCa#203	MSS_ngs	7.018792362	6.617385	1.159909	4.656608	2.193126	3.047997
GCa#204	MSS_ngs	7.818441016	3.335925	2.816354	2.531865	1.520999	1.495861
GCa#205	MSI_ngs	9.33927608	5.694467	3.376609	3.478012	0.913871	1.670512
GCa#206	MSS_ngs	4.792031794	0.897508	1.81144	1.065917	0.865269	1.129222
GCa#207	MSI_ngs	8.384044476	5.102704	3.710593	4.682401	2.329487	2.142892
GCa#208	MSS_ngs	2.966923594	4.688225	2.79387	2.973425	2.026451	1.982219
GCa#209	MSS_ngs	6.25081282	4.854162	2.402438	4.63352	1.304977	2.117386
GCa#210	MSI_ngs	12.87932514	15.44886	2.839323	7.132745	2.048067	7.544995
GCa#211	MSS_ngs	0.416548767	6.869178	2.586048	4.992214	1.857836	3.832507
GCa#212	MSI_ngs						

GCa#213	MSS_ngs	6.464662028	2.122987	1.971727	2.684672	1.52167	2.00993
GCa#214	MSS_ngs	4.085219132	1.676934	2.141597	1.129086	0.968267	1.687104
GCa#215	MSI_ngs	10.11028927	4.733709	1.840393	3.167762	1.395522	2.532106
GCa#216	MSS_ngs	5.488919618	2.010763	2.20256	1.895361	1.067981	1.630156
GCa#217	MSS_ngs	4.827774477	2.047802	2.00022	1.728415	1.186381	0.924211
GCa#218	MSS_ngs	8.166725991	5.91737	2.459101	4.0399	2.556801	2.078622
GCa#219	MSS_ngs	7.318342067	6.07394	3.627573	4.570981	1.292992	1.913975
GCa#220	MSS_ngs						
GCa#221	MSS_ngs						
GCa#222	MSS_ngs	5.315246764	5.001375	3.274473	4.042897	2.569127	1.991977
GCa#223	MSS_ngs	6.950858024	5.468473	3.027733	3.385984	1.435515	3.017972
GCa#224	MSS_ngs	5.022820828	4.095826	2.577683	3.091992	1.348981	2.758692
GCa#225	MSS_ngs	7.263176664	3.052133	2.268901	2.702815	1.240434	1.836841
GCa#226	MSS_ngs	8.239851257	4.102789	3.84961	3.559816	2.095045	2.610909
GCa#227	MSS_ngs	6.933353899	7.33043	2.933229	5.012558	2.902946	2.888245
GCa#228	MSS_ngs	7.969924904	6.548692	2.242061	4.990209	3.271513	3.193974
GCa#229	MSS_ngs	10.66022296	4.207782	2.454745	4.228289	0.918127	2.111579
GCa#230	MSS_ngs	8.344021254	9.901894	2.431611	5.406848	4.823915	6.232858
GCa#231	MSS_ngs	9.164610257	7.561052	3.464161	5.712339	2.781373	4.668776
GCa#232	MSS_ngs						
GCa#233	MSS_ngs	6.94099015	5.71434	3.704209	5.878391	1.73506	2.643463
GCa#234	MSS_ngs						
GCa#235	MSS_ngs	13.51959918	6.684908	3.57613	6.145786	3.582073	4.346421
GCa#236	MSS_ngs	7.526312099	6.938678	2.490037	4.658115	2.329359	3.869008
GCa#237	MSS_ngs	10.64484603	10.17143	6.465208	7.246955	3.364425	4.558419
GCa#238	MSS_ngs	9.539754611	9.691693	5.913516	6.998777	3.184756	2.636626
GCa#239	MSS_ngs	6.422201583	4.407694	4.152579	3.863748	3.265427	3.715687
GCa#240	MSS_ngs	8.106339701	6.695799	9.499891	6.421135	2.3575	4.614172
GCa#241	MSS_ngs	8.214466885	4.175727	4.341691	3.264555	2.956417	4.360787
GCa#242	MSS_ngs	4.854601096	4.152737	3.996376	3.421246	2.012204	3.745716
GCa#243	MSS_ngs	6.698749854	2.962046	3.633755	3.327311	2.510319	3.991106
GCa#244	MSS_ngs						
GCa#245	MSI_ngs	14.06601533	19.99491	4.720436	8.100131	6.911285	4.506049
GCa#246	MSS_ngs						
GCa#247	MSI_ngs	11.12067574	9.674614	4.897427	7.160898	3.320086	3.593414
GCa#248	MSS_ngs						
GCa#249	MSS_ngs	6.678017131	2.948892	4.225558	2.670989	2.165752	1.92301
GCa#250	MSS_ngs						
GCa#251	MSS_ngs	1.492532205	14.09331	2.829422	5.529059	5.858869	2.266322
GCa#252	MSI_ngs	7.819012066	11.92564	1.245863	4.518403	5.184064	2.335787
GCa#253	MSS_ngs	2.866875516	8.164292	5.421523	4.953437	2.432034	2.951207
GCa#254	MSS_ngs	12.37381967	10.57428	4.149645	8.986955	3.756475	2.776812
GCa#255	MSS_ngs	7.4362983	15.04699	2.71304	9.761157	3.215433	3.3716

GCa#256	MSS_ngs	4.602553547	9.802552	2.442072	6.89921	2.188007	2.107438
GCa#257	MSS_ngs	6.669533668	3.150026	2.329357	3.678678	2.243837	2.325152
GCa#258	MSS_ngs	4.406426473	2.145052	2.405758	2.046356	1.678394	1.521351
GCa#259	MSS_ngs	8.296879281	4.317718	3.632488	2.542446	3.26686	2.017597
GCa#260	MSS_ngs						
GCa#261	MSS_ngs						
GCa#262	MSS_ngs	6.68832285	6.283343	4.038098	4.727231	3.397518	2.783907
GCa#263	MSI_ngs						
GCa#264	MSI_ngs	0.71318654	8.203194	4.483059	3.838154	3.375804	4.858056
GCa#265	MSS_ngs	3.335478784	7.869703	3.379414	4.495727	2.452011	3.459295
GCa#266	MSI_ngs	8.327484872	8.856744	1.532341	4.142286	4.865293	3.931084
GCa#267	MSS_ngs	2.709735706	5.201624	3.291531	3.53016	1.782552	1.970604
GCa#268	MSS_ngs	7.830221389	6.128436	3.029323	3.865133	3.159731	2.136814
GCa#269	MSS_ngs	7.60713749	1.892234	2.100568	2.501008	1.463728	1.482449
GCa#270	MSS_ngs	8.172984305	4.966606	2.377264	3.797252	1.786304	1.709012
GCa#271	MSS_ngs	7.806830685	7.758969	4.866218	4.741135	2.533301	3.724688
GCa#272	MSS_ngs	11.3525762	6.597818	4.347972	5.791127	2.584085	4.079331
GCa#273	MSS_ngs	5.122373078	9.854047	5.611075	3.666141	2.885448	4.302169
GCa#274	MSS_ngs	7.354008462	3.019936	3.66883	3.900531	2.345571	2.716863
GCa#275	MSS_ngs	8.453696618	4.922978	4.306407	4.352887	2.977236	3.069955
GCa#276	MSS_ngs	6.126697346	9.391739	3.100926	6.607148	2.886474	2.984443
GCa#277	MSI_ngs	3.493213441	7.721255	3.665835	4.043918	2.095831	5.795
GCa#278	MSS_ngs	7.06291533	4.807099	3.996095	3.868296	2.891576	2.877031
GCa#279	MSS_ngs	12.53955149	3.66335	3.313795	2.656708	1.954458	1.978412
GCa#280	MSS_ngs	10.0704261	9.395958	3.948997	6.117463	4.927609	3.667968
GCa#281	MSS_ngs	5.984329492	5.047428	4.736552	3.062039	1.363793	2.173252
GCa#282	MSI_ngs	4.013120468	8.39149	2.662269	4.731636	2.335014	3.604135
GCa#283	MSS_ngs	10.58382173	11.97846	4.752401	9.997719	1.725381	3.208319
GCa#284	MSS_ngs	5.932034342	9.768825	6.60898	4.952063	2.211795	3.209817
GCa#285	MSS_ngs						
GCa#286	MSI_ngs	0.53945453	8.437159	3.790333	3.371326	2.585292	1.96481
GCa#287	MSI_ngs	1.018460915	5.075345	4.636694	4.319901	2.139511	3.797389
GCa#288	MSS_ngs	6.061189633	3.299065	2.93277	2.382995	2.393314	1.528163
GCa#289	MSS_ngs						
GCa#290	MSS_ngs	5.295419576	6.981275	4.792774	3.406235	2.806987	2.37835
GCa#291	MSS_ngs						
GCa#292	MSS_ngs	6.055903276	9.886798	3.888964	8.762386	2.647301	3.365411
GCa#293	MSI_ngs	1.450936787	10.02057	2.698301	6.457655	3.070237	3.987995
GCa#294	MSS_ngs	9.759993955	7.933843	3.950178	6.644257	3.535734	2.623126
GCa#295	MSS_ngs	12.62162323	10.6416	5.932307	7.344641	4.11817	4.109219
GCa#296	MSS_ngs	9.046747473	6.802679	2.443557	4.155896	4.273621	4.346413
GCa#297	MSS_ngs	11.35294882	8.556862	3.877509	4.407258	4.692929	1.321613
GCa#298	MSS_ngs	6.716768072	3.676252	5.780057	2.629787	3.161307	1.877578

GCa#299	MSS_ngs						
GCa#300	MSI_ngs	2.155074134	12.78163	2.577535	8.244336	3.752075	3.059985
GCa#301	MSI_ngs	0.314807663	6.90576	3.592435	4.579035	2.696948	2.570705
GCa#302	MSS_ngs	4.950247288	5.711163	2.497549	2.317233	3.295844	2.572117
GCa#303	MSS_ngs	5.535858747	8.473497	3.612242	3.823279	3.593627	4.379797
GCa#304	MSS_ngs	7.094902438	12.47357	2.980744	6.868735	3.180493	4.022819
GCa#305	MSS_ngs	8.548225556	3.17739	3.740113	3.164103	2.527889	1.887798
GCa#306	MSS_ngs	13.98041445	8.736378	5.018582	8.156847	4.260231	4.382309
GCa#307	MSI_ngs	7.035180804	3.318741	3.166609	2.342114	1.718737	1.690766
GCa#308	MSS_ngs	8.561868843	8.342999	3.891555	5.772468	2.373694	1.647495
GCa#309	MSI_ngs	2.128871239	5.931455	5.944051	1.444448	3.575773	3.538549
GCa#310	MSI_ngs	0.588737046	8.670398	5.977469	4.220364	3.394721	2.099236
GCa#311	MSI_ngs	0.584285439	15.91712	6.132913	5.943648	3.298162	2.555273
GCa#312	MSI_ngs	1.597685263	9.05941	2.770854	2.945477	2.983006	3.308768
GCa#313	MSS_ngs	9.739259535	2.333135	3.215069	3.578238	2.603478	2.027022
GCa#314	MSS_ngs	15.44718268	18.73454	4.579825	10.74191	5.699759	4.131863
GCa#315	MSS_ngs	8.603189844	6.869385	2.118621	4.94023	4.284687	5.541275
GCa#316	MSI_ngs	0.820263743	9.189879	2.478472	2.565833	2.308731	2.519394
GCa#317	MSI_ngs	2.03845179	6.010388	2.549712	4.61006	3.757389	3.332076
GCa#318	MSS_ngs	9.925166186	7.747059	5.001853	5.08245	6.083399	3.548042
GCa#319	MSS_ngs	4.812682918	6.03858	2.775875	3.972895	2.55571	2.468794
GCa#320	MSI_ngs	0.823775749	9.441972	4.759756	5.239163	3.787178	3.277309
GCa#321	MSS_ngs	7.82274262	1.929487	3.209167	3.372103	1.989511	1.643513
GCa#322	MSS_ngs	6.324862585	8.641945	2.599007	3.552059	2.479022	2.870245
GCa#323	MSS_ngs	4.466077094	3.423435	2.75389	2.52166	1.152663	2.280703
GCa#324	MSS_ngs	7.641534856	5.378196	3.615071	5.086837	3.431646	2.757321
GCa#325	MSS_ngs	4.703665604	3.369535	2.587701	1.722328	1.749039	2.384927
GCa#326	MSS_ngs	4.541749274	3.68081	3.879349	2.077214	2.0014	1.892749
GCa#327	MSS_ngs	7.556393299	7.666211	2.742537	2.496456	1.30988	2.997499
GCa#328	MSS_ngs	3.802934922	2.772399	2.924252	2.132878	1.664358	2.088816
GCa#329	MSS_ngs	6.211172651	4.754438	4.784281	4.720101	4.308973	2.696682
GCa#330	MSS_ngs	6.948835268	2.164721	3.129135	2.518969	2.947515	1.715367
GCa#331	MSS_ngs	8.238611712	7.993471	1.645482	6.89136	3.479293	2.935066
GCa#332	MSS_ngs						
GCa#333	MSS_ngs	5.431097943	5.456963	2.015275	3.727926	4.511103	1.772232
GCa#334	MSS_ngs	4.197197377	10.06076	2.851189	5.392793	6.546072	3.712909
GCa#335	MSS_ngs	5.493068538	5.458426	2.545907	3.18847	2.099775	1.906596
GCa#336	MSS_ngs	26.23134898	14.12192	1.768141	8.461426	3.491116	5.67215
GCa#337	MSS_ngs	6.327185749	4.031378	2.801803	2.341739	2.271869	1.594137
GCa#338	MSS_ngs	6.359112686	4.75336	4.952408	3.638197	2.344047	3.355401
GCa#339	MSS_ngs	9.371831329	9.221832	4.666329	5.933826	3.21401	3.840991
GCa#340	MSS_ngs	9.365646915	8.962571	4.821376	6.728784	3.791993	2.779918
GCa#341	MSS_ngs						

GCa#342	MSS_ngs	8.562027756	9.171436	2.038651	7.450864	3.640928	3.693536
GCa#343	MSS_ngs	6.385262882	3.663761	4.257765	3.681296	2.277768	2.647385
GCa#344	MSS_ngs	7.20866587	8.236751	3.027854	3.26829	5.106187	2.616251
GCa#345	MSS_ngs	7.468683784	6.962772	5.22964	4.370661	1.686088	3.0464
GCa#346	MSS_ngs	7.481093686	5.715669	6.549136	4.263945	3.017468	3.062853
GCa#347	MSI_ngs	3.212383605	6.567927	1.576899	4.623301	3.009559	3.060762
GCa#348	MSS_ngs	8.606333297	13.06603	3.911254	9.751485	2.196301	2.598079
GCa#349	MSS_ngs	9.16936145	21.79754	3.889286	12.33701	4.13261	5.038121
GCa#350	MSS_ngs	3.779455712	4.343367	1.103271	3.14739	2.536354	2.905711
GCa#351	MSS_ngs	5.537793503	6.839963	3.130142	3.82213	3.53594	2.869696
GCa#352	MSS_ngs	6.856700833	8.973803	3.327409	7.111699	5.424146	5.077376
GCa#353	MSI_ngs	2.985502546	3.765263	1.663469	2.07985	2.187462	2.31335
GCa#354	MSS_ngs	8.304183665	5.191618	2.738537	3.690229	1.981729	2.072491
GCa#355	MSS_ngs						
GCa#356	MSS_ngs	8.197749614	5.045013	4.313253	3.263715	2.541692	2.376771
GCa#357	MSS_ngs	6.701321212	3.707617	2.972101	2.565002	2.786174	2.405393
GCa#358	MSS_ngs	9.179093733	9.182482	4.777082	4.833757	4.402502	3.143395
GCa#359	MSS_ngs	6.736950442	3.341772	3.200278	3.311248	2.67778	2.367803
GCa#360	MSS_ngs	9.441352274	5.795004	2.716342	3.889171	4.056262	2.617552
GCa#361	MSS_ngs	8.32623963	4.337504	3.864975	3.363312	3.171919	3.429449
GCa#362	MSS_ngs	5.337489341	4.643476	2.781785	3.541006	2.784866	1.495285
GCa#363	MSS_ngs	8.607382733	4.74982	4.516469	4.343928	3.956861	2.450932
GCa#364	MSS_ngs	8.106107644	8.354408	4.806864	5.527492	3.330403	4.33328
GCa#365	MSS_ngs	8.376809362	2.037583	3.155915	3.072202	2.401704	2.486538
GCa#366	MSS_ngs	9.457060489	3.438428	3.092015	3.662577	2.626365	2.090358
GCa#367	MSS_ngs	5.639565097	2.189789	3.960891	2.615748	2.018814	2.27925
GCa#368	MSS_ngs	6.799310652	5.407214	3.110281	3.336766	3.385446	3.115674
GCa#369	MSS_ngs	8.167222144	5.309109	4.043061	4.090666	2.923963	2.141673
GCa#370	MSI_ngs	3.31468695	7.453775	1.910771	4.62301	4.20329	2.439431
GCa#371	MSI_ngs	0.834828468	4.128405	2.942979	3.245541	3.269022	1.977729
GCa#372	MSS_ngs	5.345230711	5.7744	2.026754	4.070518	2.584671	4.00157
GCa#373	MSS_ngs						
GCa#374	MSS_ngs	7.95291218	5.993073	4.454953	3.286894	2.655625	2.76306
GCa#375	MSS_ngs	6.035856761	5.01058	3.86164	3.271675	5.234573	2.010425
GCa#376	MSS_ngs	4.759596008	8.027044	1.784975	4.034581	2.942084	2.345675
GCa#377	MSS_ngs	8.522409937	5.2651	3.098105	2.338198	2.107058	3.005977
GCa#378	MSS_ngs	7.883762195	6.167871	7.299692	3.41548	1.581614	2.341578
GCa#379	MSS_ngs	10.06407055	9.13249	5.080725	6.409837	3.194744	2.30624
GCa#380	MSI_ngs	1.99672005	16.95829	7.150366	7.303223	4.400164	6.607403
GCa#381	MSS_ngs	3.968348004	6.274814	2.7459	4.389587	2.131104	2.514035
GCa#382	MSI_ngs	0.611857231	15.18428	4.380652	5.951221	2.891458	3.126436
GCa#383	MSS_ngs	7.191604387	6.490363	4.746922	3.485062	3.223999	1.895603
GCa#384	MSS_ngs	7.332782305	7.980259	3.85932	5.572857	3.543346	2.49481

GCa#385	MSS_ngs	7.651174445	4.938685	5.696675	3.455377	2.743245	2.212275
GCa#386	MSI_ngs	1.30855237	14.13086	5.605263	8.732639	4.461617	4.306443
GCa#387	MSS_ngs	9.484210805	6.739838	4.078041	3.930489	3.740224	4.279348
GCa#388	MSS_ngs	9.060366454	7.900005	5.312054	4.176035	3.480209	3.124432
GCa#389	MSS_ngs	9.884412208	13.67022	5.835469	6.463606	4.18963	3.368638
GCa#390	MSS_ngs	11.84560057	9.714616	4.30588	6.528333	3.386394	3.80346
GCa#391	MSS_ngs	6.197551506	9.09677	3.092498	6.117059	4.1106	3.707614
GCa#392	MSS_ngs	5.977878131	3.612058	3.162446	2.29021	1.807271	2.843755
GCa#393	MSS_ngs	5.639427873	7.194274	3.049668	4.636348	3.341359	2.329969
GCa#394	MSS_ngs	8.543405674	5.000596	5.421641	3.300604	2.578149	2.954246
GCa#395	MSI_ngs	2.336470518	10.17333	4.559337	3.882531	3.783145	2.138236
GCa#396	MSI_ngs	1.663846183	6.860162	5.079222	4.129466	2.895176	2.79065
GCa#397	MSS_ngs	9.734828755	11.4304	5.025593	6.268188	3.767186	4.731528
GCa#398	MSS_ngs						
GCa#399	MSI_ngs						
GCa#400	MSS_ngs	8.966388216	4.182482	3.977403	3.694322	2.462115	3.460276
GCa#401	MSI_ngs	2.56425591		5.05764	2.554293	4.020119	3.032563
GCa#402	MSS_ngs						
GCa#403	MSS_ngs						
GCa#404	MSS_ngs	11.58724144	4.870432	4.890514	4.526133	2.054425	3.056179
GCa#405	MSI_ngs	4.23732541	9.008399	1.040988	3.566424	3.020288	3.069946
GCa#406	MSI_ngs	2.765126448	9.534839	8.606946	3.960499	2.985704	3.737702
GCa#407	MSI_ngs	1.421973146	5.038591	3.348539	3.07057	3.163205	2.049605
GCa#408	MSS_ngs	10.12448446	10.0953	6.124741	6.569459	11.14503	3.43099
GCa#409	MSS_ngs	9.520266982	6.062178	6.318503	5.777174	3.717041	3.308731
GCa#410	MSS_ngs	4.867764958	9.4122	5.660702	6.106292	4.061835	2.816698
GCa#411	MSS_ngs	8.603965981	11.10715	3.314537	5.036137	3.981522	3.470648
GCa#412	MSS_ngs	5.49483424	6.541744	5.019018	3.194926	2.471986	1.567289
GCa#413	MSS_ngs						
GCa#414	MSS_ngs	4.214676794	8.376905	8.018585	3.147111	2.38036	2.71291
GCa#415	MSS_ngs	6.777899076	6.93404	5.060873	4.487364	2.288411	2.546179
GCa#416	MSS_ngs	9.534154607	7.594665	4.622398	4.742842	4.73806	2.452807
GCa#417	MSS_ngs	7.320712087	12.19161	3.288127	6.253434	2.474885	2.596965
GCa#418	MSS_ngs	10.46809233	9.393156	4.598633	5.375542	3.817453	3.773696
GCa#419	MSI_ngs	1.463503457	12.16475	1.385853	5.885548	3.24966	2.259947
GCa#420	MSS_ngs	5.113156089	5.757561	4.105523	4.604196	3.557973	2.274756
GCa#421	MSS_ngs	7.726020329	5.025242	6.757569	3.243094	3.254115	2.349011
GCa#422	MSS_ngs	10.50711037	7.045933	5.105752	8.080395	3.742409	2.830394
GCa#423	MSS_ngs	7.449700414	8.968963	3.517348	7.11051	2.264858	3.479967
GCa#424	MSS_ngs	10.15074997	10.93111	3.513725	4.507267	3.744749	4.321942
GCa#425	MSS_ngs	8.553424477	3.39068	3.447867	4.179404	3.017512	1.674927
GCa#426	MSS_ngs	7.511711894	8.874188	5.051431	5.432692	4.23254	4.477274
GCa#427	MSS_ngs	6.604844166	7.109145	5.065962	6.826448	3.94682	2.652099

GCa#428	MSS_ngs	5.821192409	12.0405	6.748051	6.528834	4.745615	5.078796
GCa#429	MSS_ngs	6.542510011	8.048819	5.872476	4.101028	3.679539	4.756537
GCa#430	MSS_ngs	10.94037841	11.33891	2.33362	5.15547	2.296098	5.318513
GCa#431	MSS_ngs	13.54202907	29.24334	12.28293	15.85123	5.729134	4.150839
GCa#432	MSS_ngs	7.887352483	9.692536	3.784677	5.441662	2.407714	5.564126
GCa#433	MSS_ngs						
GCa#434	MSI_ngs	0.542143568	9.240293	3.704527	6.143825	2.108867	3.607497
GCa#435	MSS_ngs	6.713491243	6.855321	3.242577	4.027179	2.601501	3.290737
GCa#436	MSS_ngs	7.29887559	4.273334	4.958491	3.075004	2.099813	1.479435
GCa#437	MSS_ngs						
GCa#438	MSS_ngs	9.740132743	9.140422	3.738979	5.010196	3.626012	4.236923
GCa#439	MSS_ngs	7.066712483	5.810398	4.949001	3.98655	2.432461	2.252763
GCa#440	MSS_ngs	7.148162222	4.755482	3.007517	3.704572	1.870147	2.964315
GCa#441	MSS_ngs	5.827546953	6.929003	2.698311	2.957359	3.3491	2.045137