**UNIVERSITÀ DEGLI STUDI DELL'INSUBRIA**

**DIPARTIMENTO DI MEDICINA E INNOVAZIONE TECNOLOGICA**

PhD in Experimental and Translational Medicine

# Precision oncology in endometrioid endometrial carcinoma integrating genetic and epigenetic features

Tutor: Prof.ssa Daniela Furlan

Co-Tutor: Prof. Antonio Travaglino

PhD Thesis of:

Muhammad Adnan

ID: 755741

Academic Year 2022/2025

*Blank for Dedication*

**Contents**

## ABSTRACT OF THE RESEARCH

Endometrial carcinoma (EC) displays marked molecular heterogeneity driven by complex interactions between genetic and epigenetic mechanisms. Among these, somatic mutations, microsatellite instability (MSI) and DNA methylation play key diagnostic and biological roles, yet the molecular drivers that shape the methylation landscape and deficiency of DNA mismatch repair system (dMMR) or MSI remain partly understood. This thesis integrates two complementary studies aimed at exploring the link between genetic and epigenetic alterations; and using computational approaches to optimise Next-Generation Sequencing (NGS)-based MSI detection to advance molecular diagnostics and clinical application.

The first study interrogated the interplay between driver mutations, global DNA methylation and MSI status with particular focus on *MLH1* promoter methylation in endometrioid EC (EEC). We combined *in-silico* analysis of public datasets with an institutional cohort characterised by NGS, MMR immunohistochemistry (IHC), PCR-based MSI testing, MS-MLPA for *MLH1* methylation and long interspersed nuclear element-1 (LINE-1) pyrosequencing. Our findings demonstrate that *KRAS*-activating mutations are significantly associated with MSI and *MLH1* promoter hypermethylation. Furthermore, co-occurrence of *KRAS* and *ARID1A* mutations correlated with a hypermethylated phenotype, suggesting a cooperative role in epigenetic remodeling. LINE-1 analysis indicated that *KRAS* mutations contribute not only to locus-specific but also to global DNA hypermethylation, highlighting mechanistic links between oncogenic signaling and epigenetic regulation.

The second study focused on optimising and validating an NGS-based MSI detection panel. Starting from 122 microsatellite loci, we identified a refined subset of 51 loci that retained discriminatory performance for MSI status through model-based clustering and threshold optimisation in colorectal carcinoma (CRC) and EC cohorts. NGS-derived MSI scores demonstrated high concordance with conventional gold standard methods, validating a streamlined workflow capable of integrating mutational and MSI assessment in a single assay.

Collectively, these studies emphasize the biological and clinical significance of *KRAS* and *ARID1A* mutations in shaping MSI and methylation landscapes and establishing a robust NGS-based framework for accurate MSI detection across tumor types. The findings support the potential utility of integrated molecular profiling for improved risk stratification, therapeutic decision-making and identification of actionable subset of EC, including candidates for *KRAS*-targeted or immunotherapy-based interventions.

**Keywords:** Endometrial cancer, NGS, *KRAS/ARID1A* mutations, LINE-1 methylation, MSI, computational optimisation, model-based clustering.

## ABBREVIATIONS

| | |
|---|---|
| AF | Allele Frequency |
| AIRTUM | Associazione Italiana Registri Tumori |
| AUC | Area Under the Curve |
| CAH | Complex Atypical Hyperplasia |
| CCEC | Clear Cell Endometrial Carcinoma |
| CE | Capillary Electrophoresis |
| CIMP | CpG Island Methylator Phenotype |
| CN-H | Copy Number-High |
| CN-L | Copy Number-Low |
| CNV | Copy Number Variation |
| CRC | Colorectal Carcinoma |
| DFS | Disease-Free Survival |
| dMMR | Deficiency of Mismatch Repair Protein System |
| DNMT | DNA Methyltransferase |
| EBRT | External Beam Radiotherapy |
| EC | Endometrial Carcinoma |
| EEC | Endometrioid Endometrial Carcinoma |
| EIN | Endometrial Intraepithelial Neoplasia |
| ER | Oestrogen Receptor |
| ESGO | European Society of Gynaecological Oncology |
| ESMO | European Society for Medical Oncology |
| ESP | European Society of Pathology |
| ESTRO | European Society for Radiotherapy and Oncology |
| FFPE | Formalin-Fixed and Paraffin-Embedded |
| FIGO | International Federation of Gynecology and Obstetrics |
| FN | False Negative |
| FP | False Positive |
| FPR | False-Positive Ratio |

| | |
|---|---|
| GMM | Gaussian Mixture Model |
| HS | High Sensitivity |
| ICI | Immune Checkpoint Inhibitor |
| IHC | Immunohistochemistry |
| indel | Insertion-Deletion |
| IQR | Interquartile Range |
| LINE-1 | Long Interspersed Nuclear Element-1 |
| LVSI | Lymphovascular Space Invasion |
| MBC | Molecular Barcode |
| MMR | DNA Mismatch Repair Protein System |
| MMRd | DNA Mismatch Repair-Deficient |
| MNR | Mononucleotide Repeat |
| MSI | Microsatellite Instability |
| MSI-H | Microsatellite Instability-High |
| MSI-L | Microsatellite Instability-Low |
| MS-MLPA | Methylation-Specific Multiplex Ligation-Dependent Probe Amplification |
| MSS | Microsatellite Stable |
| NGS | Next-Generation Sequencing |
| NSMP | No Specific Molecular Profile |
| OR | Odds Ratio |
| ORF | Open Reading Frame |
| OS | Overall Survival |
| p53abn | p53-abnormal |
| PCA | Principal Component Analysis |
| pMMR | Proficiency of Mismatch Repair Protein System |
| *POLE*mut | *POLE*-Ultramutated |
| PR | Progesterone Receptor |
| QC | Quality Control |
| ROC | Receiver-Operating Characteristic |

| | |
|---|---|
| SALSA | Selective Adaptor Ligation, Selective Amplification |
| SD | Standard Deviation |
| SEC | Serous Endometrial Carcinoma |
| SNV | Single Nucleotide Variation |
| SWI/SNF | SWItch/Sucrose Non-Fermentable Complex |
| TCGA | The Cancer Genome Atlas |
| TF | Transcription Factor |
| TIL | Tumour-Infiltrating Lymphocyte |
| TMB | Tumour Mutation Burden |
| UCS | Uterine Carcinosarcoma |
| UEC/DEC | Undifferentiated/Dedifferentiated Endometrial Carcinoma |
| UTR | Untranslated Region |
| VA | Varese |
| VUS | Variant Of Unknown Significance |
| WHO | World Health Organisation |

# GENERAL INTRODUCTION

## Epidemiology of EC

Endometrial carcinoma (EC), a malignancy arising from the inner lining of the uterine corpus, stands as the most frequently diagnosed gynaecological cancer in developed countries with approximately 400,000 new cases diagnosed annually worldwide [1]. In Italy, data from the *Italian Registry of Tumours* (Associazione Italiana Registri Tumori - AIRTUM) for 2024 estimated an annual burden of approximately 8,000 new cases, making it the fourth most common cancer among Italian women [2]. The incidence has been observed to increase over recent decades, largely attributed to aging and prevalence of key risk factors, notably obesity, progesterone deficiency, unopposed exposure to oestrogen, early menarche, late menopause, nulliparity and certain genetic syndromes such as Lynch syndrome [3]. The majority of patients are postmenopausal, with a peak incidence in women aged 60-70 years [3], however up to 15% of cases are presented in premenopausal females with around 4% in younger than 40 years and having better prognosis [4]. Clinically, postmenopausal bleeding is the sentinel symptom that prompts investigation in more than 90% cases of EC, leading to a diagnosis at an early stage [5]. However, a significant minority of patients are presented with aggressive, high-grade histological subtypes such as, serous EC (SEC) or clear cell carcinomas (CCEC), with advanced-stage disease [5].

## Histological characterisation

The histological characterisation of EC is the basis of its diagnosis and initial risk stratification, built upon the dualistic model of classification, which divides EC into two main types, with distinct morphological features and clinical behaviours [6, 7]. Type I carcinomas, accounting for approximately 80% of cases, are typically endometrioid ECs (EEC). They are oestrogen-driven, low-grade tumours that exhibit glandular architectures resembling normal endometrium, often arising from a precancerous lesion known as Endometrial Intraepithelial Neoplasia (EIN) [6-8]. In contrast, Type II carcinomas are non-oestrogen dependent, high-grade neoplasms with aggressive clinical courses. This category includes SECs, characterised by complex papillary and glandular patterns with marked cytological atypia; clear cell and mixed cell carcinomas, which are rare subtypes [6, 7].

The World Health Organisation (WHO) classification further refines histological subtyping, distinguishing between low-grade (grades 1 and 2) and high-grade (grade 3) ECs. Grades 1 and 2 comprise EECs while the grade 3 encompasses EEC, SEC, CCEC, undifferentiated or dedifferentiated EC (UEC/DEC) and uterine carcinosarcoma (UCS) [7, 9].

Furthermore, International Federation of Gynecology and Obstetrics (FIGO) staging classifies ECs into four stages I to IV, often bifurcated into early (I-II) and advanced (III-IV) categories [10]. In 2023, stages were re-classified by FIGO based on histological type, depth of myometrial invasion and lymphovascular space invasion (LVSI), dividing stages into subcategories as A, B and C, shown in table 1 [10-12]. The accurate distinction between these types and their histological mimics is critical and has limitations, including significant inter-observer variability in tumour grading and the recognition of ambiguous cases with mixed or overlapping features [13]. This propelled the integration of molecular classification to provide a more objective and prognostically powerful framework.

**Table 1: Framework of FIGO staging criteria 2023 of EC summarising key features of substages based on tumour confinement, invasion depth, LVSI and metastasis.**

| Stage | Key features and substages |
|---|---|
| **Stage I** | Tumour confined to the uterine corpus and ovary.<br>**IA:** Non-aggressive; limited to endometrium; <50% myometrial invasion; no or focal LVSI.<br>• *IA1*: Confined to endometrium or polyp.<br>• *IA2*: <50% myometrial invasion; no/focal LVSI.<br>• *IA3*: Low-grade EEC limited to uterus and ovary.<br>**IB:** ≥50% myometrial invasion; no/focal LVSI.<br>**IC:** Aggressive histology confined to endometrium or polyp. |
| **Stage II** | Cervical stromal invasion or substantial LVSI, no extrauterine spread.<br>**IIA:** Cervical stromal invasion (non-aggressive).<br>**IIB:** Substantial LVSI (non-aggressive).<br>**IIC:** Aggressive histology with any myometrial invasion. |
| **Stage III** | Local/regional spread beyond uterus.<br>**IIIA:** Invasion of uterine serosa and/or adnexa.<br>• *IIIA1*: Ovarian/fallopian involvement.<br>• *IIIA2*: Uterine serosal/subserosal invasion.<br>**IIIB:** Vaginal, parametrial, or pelvic peritoneal metastasis.<br>• *IIIB1*: Vaginal/parametrial spread.<br>• *IIIB2*: Pelvic peritoneal metastasis.<br>**IIIC:** Pelvic/para-aortic lymph node metastasis.<br>• *IIIC1*: Pelvic nodes metastasis (i: micro, ii: macro).<br>• *IIIC2*: Para-aortic nodes metastasis (i: micro, ii: macro). |
| **Stage IV** | Invasion of bladder/intestinal mucosa or distant metastasis.<br>**IVA:** Involvement of bladder and/or bowel mucosa.<br>**IVB:** Peritoneal metastasis beyond pelvis.<br>**IVC:** Distant metastasis (lung, liver, brain, bone or distant lymph nodes). |

**Molecular classification**

Recently, The Cancer Genome Atlas (TCGA) research network defined four distinct molecular

subgroups of EC through comprehensive genomic profiling with profound prognostic implications[14]. This classification includes: 1) *POLE*-ultramutated (*POLE*mut), characterised by pathogenic mutations in the exonuclease domain of the *POLE* gene, resulting in an extremely high tumour mutation burden (TMB) ($232\times10^{-6}$ mutations per Mb). This group, despite its high-grade histological features, exhibits an excellent prognosis. 2) Microsatellite unstable (MSI-hypermutated), with defects in DNA mismatch repair (MMR) protein system (also called MMR-deficient subtype, MMRd), leading to a high mutation rate ($18\times10^{-6}$ mutations per Mb). 3) Copy-number low (CN-L) or no specific molecular profile (NSMP), mostly encompasses the low-grade EECs with a relatively stable genome and lower TMB ($2.9\times10^{-6}$ mutations per Mb), frequently harbouring *PTEN* and *CTNNB1* mutations. 4) Copy-number high (CN-H) or p53-abnormal (p53abn), is strongly associated with serous-like carcinomas with low TMB ($2.3\times10^{-6}$ mutations per Mb). This molecular class is characterised by frequent *TP53* mutations and a highly unstable genome and is linked to the worst prognosis [14, 15].

The main driver genes identified in EC include *POLE, POLD1, PTEN, PIK3CA, PIK3R1, KRAS, FGFR2, ARID1A, TP53, CTNNB1 and FBXW7* [14, 16]. The molecular classification is increasingly being integrated into clinical practice to guide adjuvant therapy decisions, highlighting the critical role of molecular diagnostics [15].

**Molecular evolution of EC**

While TCGA provided a robust molecular aspect of EC, a critical knowledge gap persists regarding the precise sequence and interplay of molecular mechanisms that drive the disease's natural history. In low-grade EEC, prolonged oestrogen stimulation promotes the progression of EIN precursor lesions into malignant tumours. Early mutational events reported to be involved in this malignant transformation are loss of *PTEN*, *KRAS* and *ARID1A* mutations [17-19]. *PTEN* inactivation mutations, along with mutations in PI3K/AKT pathway genes, are often related with the early events of initiation of precancerous lesion leading to invasive tumour in type I carcinomas [20]. *KRAS* mutations have been found in endometrial hyperplasia and early-stage EEC, suggesting an early-event role [20, 21]. *ARID1A* mutations also emerge in early EECs but are associated frequently with progression rather than initiation, indicating their role in tumour evolution [22, 23]. Interestingly, *ARID1A* loss has also been associated with the CpG island methylator phenotype (CIMP) in EC and other cancers, suggesting interplay between genetic and epigenetic regulation [24].

In parallel, the development of MSI, often by epigenetic silencing of *MLH1* or by inherited deficiency of MMR (dMMR), contributes as an early event to the evolution of MSI/MMRd

subtype [12]. Clinically, MSI/dMMR status carries multiple implications. Firstly, MSI-positive ECs have a favourable prognosis in early-stage disease, likely due to high immunogenicity and tumour-infiltrating lymphocytes (TILs) [16]. Secondly, MSI status has become a key predictive biomarker for immune checkpoint inhibitor (ICI) therapy (e.g., anti-PD1/PDL1 antibodies) across tumour types, including EC [25]. Due to the significance of MSI/dMMR in EC, European Society for Medical Oncology (ESMO) guidelines recommend MSI detection to be a part of clinical report.

In summary, EEC evolves through a multistep process, as shown in figure 1, involving driver genetic alterations (e.g., *PTEN, KRAS, ARID1A* mutations) integrated with epigenetic dysregulation, the latter encompassing aberrant global or gene-specific DNA methylation [26] (see subsequent paragraph).

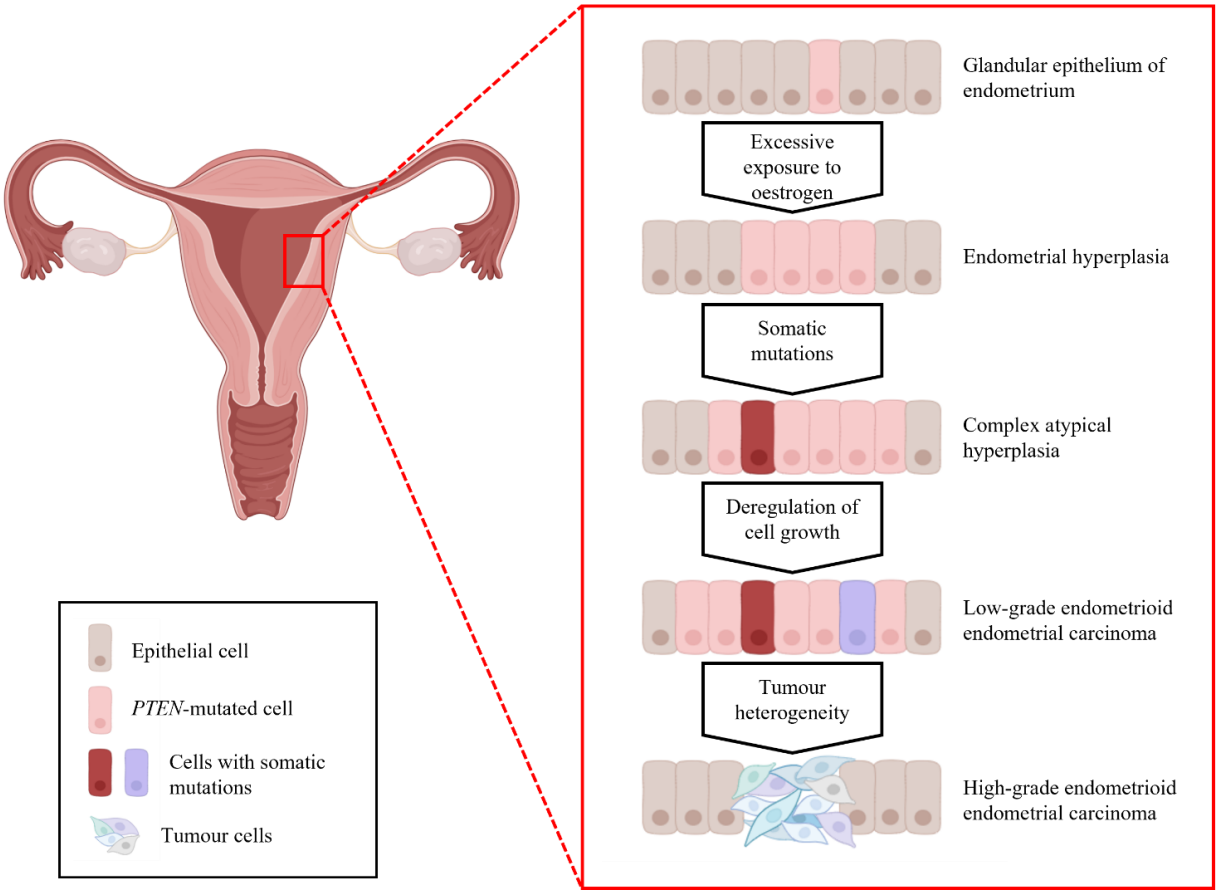



**Figure 1: Overview of initiation of endometrioid endometrial carcinoma (EEC).** The stepwise morphological and molecular evolution of endometrial glandular epithelium toward EEC initiates under prolonged oestrogen stimulation, leading to development of endometrial hyperplasia. Progressive accumulation of somatic mutations, most commonly in *PTEN, PIK3CA, PIK3R1, KRAS* genes, drives the emergence of complex atypical hyperplasia (CAH), characterised by nuclear atypia. Subsequent acquisition of additional mutations and epigenetic changes promotes clonal expansion and the transition to low-grade EEC and eventually to high-grade EEC. *Adapted from Urick and Bell, Nature Reviews Cancer (2019) [26], created with BioRender.com.*

Among other mutations *CTNNB1* and *TP53* are prominent in ECs. *CTNNB1* encodes β-catenin, a key component of the Wnt/β-catenin signalling pathway, which regulates cell proliferation, adhesion and differentiation. Driver mutations in *CTNNB1* exon 3 prevent β-catenin degradation,

leading to nuclear accumulation and activation of oncogenic transcription [27]. This overexpression of Wnt pathway is involved in the development of EIN, a precursor to EEC [28]. *CTNNB1* mutations are typically associated with a low risk profile; however, they have been implicated in recurrence and poor outcomes with reduced overall survival [14, 29].

On the other hand, *TP53* is a key tumour-suppressor gene, the alteration of which is a hallmark of aggressive disease. Although *TP53* mutations are relatively uncommon in overall EEC (found in approximately around 10%, predominantly in Grade 3 tumours), these alterations are more prevalent in SEC, affecting around 90% of cases. In these aggressive tumours, *TP53* mutation acts as a primary driver event [14].

**Epigenetic regulation in evolution of EC**

Beyond genetic alterations, epigenetic dysregulation is increasingly recognised as playing a central role in pathogenesis of EC. DNA methylation changes include hypermethylation of tumour-suppressor genes like *PTEN*, *MLH1* and Oestrogen and Progesterone Receptor (ER and PR respectively); hypomethylation of proto-oncogenes causing their overexpression or contributing to chromosomal instability and activation of repetitive elements [30, 31].

In ECs with MSI, alongside the hypermethylation of *MLH1* in sporadic cases [32, 33], *KRAS* mutations are also frequently observed [34]. While *BRAF* mutations are typically established as the main drivers of CIMP in colorectal carcinomas (CRC), alterations in the *KRAS*-mediated MAPK pathway, as well as in *ARID1A*, have been linked to CIMP subtypes, indicating a possible role of these genes in methylation events in EC [24, 35].

*ARID1A* is one of the key components of the SWItch/Sucrose Non-Fermentable (SWI/SNF) complex, a master regulator of chromatin remodelling that operates through an ATP-dependent mechanism [36]. The clinical relevance of these alterations is underscored by its high frequency across cancers: Cajuso *et al*. reported frequent alterations in the *ARID* gene family in MSI CRC with mutations in *ARID1A* in around 40% cases [37]. Mutations or loss of *ARID1A* function are also frequently observed in ECs. This disruption impairs SWI/SNF complex function, resulting in aberrant gene expression patterns that strongly promote oncogenesis [38].

In parallel to these genetic events, the epigenetic silencing of ER and PR is reported to be intrinsically linked with type I ECs and can lead to progression and aggressive behaviour of tumours [39]. This loss of hormone receptor signaling can facilitate tumour progression and lead to a more aggressive phenotype [39]. Specifically, the loss of expression, often due to promoter hypermethylation, is a clinically significant event: it not only causes acquired resistance to

standard progesterone therapies [40] but has also been functionally linked with metastatic spread [41]. Beyond these gene-specific changes, global DNA methylation status may represent another critical facet of epigenetic dysregulation in EC.

One of the established markers of global DNA methylation is long interspersed nuclear element-1 (LINE-1) retrotransposon. LINE-1 contains the only elements in humans that can still autonomously transpose and accounts for 17% of the human genome [42]. Functional retro-transposition events originate from full-length LINE-1 elements, typically around 6 kb in size. These elements possess two open reading frames (ORFs) bordered by 5′ and 3′ untranslated regions (UTRs) [43]. The 5′ UTR promoter region is characterised by the presence of multiple transcription factor (TF) binding motifs, notably including those recognised by SOX family TFs. These motifs are instrumental in dictating the cell type-specific expression patterns of LINE-1 [44]. Recently, thanks to the rise of massively parallel DNA sequencing, scientists can study the role of somatic retro-transposition in cancer on a scale never achieved. A recent whole-genome sequencing of 2,954 cancer genomes from 38 different tissue types dramatically advanced this understanding showing that new insertions varied significantly across different cancer types and that LINE-1 seems to be more expressed in retrotransposition-high cancers, such as epithelial tumours [45]. Given this high expression and potential for autonomous transposition, the epigenetic control of LINE-1 is critical. Its hypomethylation in cancers has been associated with genomic instability, chromosomal rearrangements, tumour progression and worse prognosis. The full length of LINE-1 sequence is 6 kb, however, most of the LINE-1 sequences are truncated and thus present in inactivated form. The full-length sequences have their promoter sequences epigenetically inactivated by methylation. The transcription and subsequent translation of LINE-1 sequence produce two proteins ORF1p and ORF2p. ORF2p has both endonuclease and reverse transcriptase activity. After their translation, ORF1p, ORF2p and LINE-1-mRNA forms a complex that is translocated to nucleus [46, 47]. LINE-1 activation by hypomethylation is reported in EC, CRC and gastric cancers [48].

The collective evidence regarding gene-specific (e.g., *MLH1*) and global (e.g., LINE−1) epigenetic deregulation, alongside key genetic drivers (*KRAS, ARID1A*), highlights the complex, integrated molecular etiology of EC.

**Therapeutics and prognosis of EC**

Treatment of EC is largely stage-dependent with surgery as the primary management across all stages. According to the European Society of Gynaecological Oncology (ESGO), the European SocieTy for Radiotherapy and Oncology (ESTRO) and the European Society of Pathology (ESP)

guidelines, for early stage (I-II), surgery consisting of a hysterectomy and bilateral salpingo-oophorectomy is performed; adjuvant radiotherapy can be administered based on pathological risk factors [11, 49]. In premenopausal younger G1-IA1 EEC patients, a fertility-conservative approach is often employed if LVSI, myometrial and adnexal invasion is absent [50]. Depending on the extent of spread of stage III tumours, surgery is combined with radiotherapy and chemotherapy, whereas Stage IV metastatic disease is managed with radiotherapy or surgery tailored to the patient's condition [11].

Within molecular classes, *POLE*mut tumours, are primarily treated with surgery and may not require adjuvant therapy at early stages [49]. In advance or recurrent *POLE*mut tumours as well as MSI/MMRd subtype, ICIs such as pembrolizumab and dostarlimab can be used due to presence of high TMB and PD-L1 on immune cells [25, 51]. Patients with p53abn (CN-H) tumours have been reported to respond to combined chemotherapy and radiotherapy [15] while HER2-positive p53abn tumours, respond favourably to trastuzumab [52]. The heterogeneity of the CN-L group makes it challenging to define an optimal therapeutic strategy, however, adjuvant radiotherapy and chemoradiotherapy are commonly used [15].

The overall five-year survival rate for EC is approximately 80%, but this figure has disparities between different histological and molecular subtypes [53]. Although the prognosis for early-stage, low-grade EEC is generally excellent (five-year survival around 90%), the survival rates for advanced-stage, high-grade tumours (including SEC) and recurrent disease remain poor (five-year survival around 40%) [53]. Similarly, among molecular classes, *POLE*mut shows the best prognosis, followed by MSI and CN-L, while CN-H shows the worst prognosis [13].

**Inhibitors targeting *KRAS***

*KRAS* being the most frequently mutated gene in RAS family in human cancers, promotes proliferation, survival and metastasis [54-56]. *KRAS* was considered undruggable because of its high affinity for GTP/GDP, lack of obvious binding pockets and central role in many signalling pathways [55, 57], until the development of allele-specific inhibitors for $KRAS^{G12C}$ mutation [58]. Two such selective inhibitors of $KRAS^{G12C}$ include sotorasib and adagrasib, that bind irreversibly to *KRAS* and keep it in an inactive state [59]. Clinical trials have shown good overall response with $KRAS^{G12C}$ inhibitors in solid tumours; however, a meta-analysis showed better efficacy in non-small cell lung cancer (NSCLC) than CRC [60]. A more recently developed $KRAS^{G12C}$ inhibitor is divarasib with 5-20 times more potency and around 50 times more selectivity than sotorasib and adagrasib [61]. Resistance to *KRAS* monotherapy has led to the studies using combination with other targeted therapies and immunotherapies to overcome resistance-barrier [62]. Since the

development of *KRAS*$^{G12C}$ inhibitors, next generation inhibitors targeting other mutations (*KRAS*$^{G12D}$ and *KRAS*$^{G12V}$) are also under investigation [63]. Though *KRAS* inhibitors are still under investigation, the presence of frequent *KRAS* alterations in ECs suggests potential use in future and could be promising in combination with immunotherapies in *KRAS*-mutated/MSI patients.

**Cutting-edge techniques in cancer research**

The histopathological diagnosis of EC has been fundamentally transformed since TCGA molecular classification and the integration of immunohistochemistry (IHC) assays and MSI testing with Next-Generation Sequencing (NGS) for mutational profiling.

IHC serves as a rapid, cost-effective and widely available technique for identifying alterations in four DNA MMR proteins (MSH2, MSH6, MLH1 and PMS2) and p53 protein [14]. Loss of nuclear expression of MMR proteins (typical of dMMR tumours) is highly correlated with the MSI molecular subgroup [14]. MSI testing is performed by PCR-based fragment analysis of specific mononucleotide and dinucleotide repeat markers. Since it assesses the functional consequence of dMMR, it is highly concordant with MMR-IHC and serves as a crucial confirmatory test, especially in cases with equivocal IHC results [64]. Moreover, MMR-IHC and MSI testing are recommended by ESMO guidelines as a screening tool for Lynch syndrome and a potential for response to immunotherapy [65]. IHC for p53 protein is performed to identify *TP53* mutation and classify a case as CN-H with a highly unstable genome, conferring a poor prognosis [14, 15]. These methods have established diagnostic accuracy but also limitations. IHC may miss rare MMR defects or yield ambiguous results and PCR-based panels analyse a relatively small number of loci, may require matched normal tissue and do not integrate comprehensive genomic context.

In recent advancements, NGS represents the most comprehensive technique, allowing simultaneous high-throughput detection of genomic alterations, including single nucleotide variants (SNVs), insertions/deletions (indels), copy number variations (CNVs), structural variants, TMB, together with MSI status [66]. The ability to integrate multiple biomarkers in a single assay enhances the efficiency and cost-effectiveness of molecular pathology workflows. NGS panels can incorporate the information of molecular classification (*POLE* mutations for *POLE*mut, *TP53* mutations for CN-H, MSI testing for MSI subtype) in a single assay [67]. Beyond classification, NGS identifies targetable genomic alterations in PI3K/AKT/mTOR, DNA repair (*POLE*, MMR) and receptor tyrosine kinase (*ERBB2*, *FGFR2*) pathways guiding personalised therapy [26, 66]. NGS can quantify TMB, that predicts susceptibility of high TMB subtypes to ICIs such as response of MSI to pembrolizumab shown in KEYNOTE-158 study [68, 69].

Specifically, for MSI detection, a number of studies have demonstrated that NGS-based approaches can accurately assess MSI status by analysing a much larger set of microsatellite loci than PCR-based methods, often without the need for matched normal tissue. Marques *et al*. achieved more than 96% sensitivity and 100% specificity in tumour-only samples of CRC, EC and stomach cancer using the NGS [70]. A recent study performed on different tumour types found 98% concordance of NGS-based MSI analysis with standard PCR/IHC methods [71].

Nonetheless, NGS-based MSI detection is progressing, but it also faces important challenges: tumour purity, coverage depth, sample quality especially in formalin-fixed paraffin-embedded (FFPE) samples, panel design (number and type of microsatellite loci), bioinformatic algorithms and interpretation of borderline MSI cases. However, the integration of genomic profiling and MSI testing has evolved NGS as a rapid, accurate and standalone assay thereby supporting precision oncology.

These applications have significant therapeutic implications and prognostic value, as they enable the identification of molecular alterations that can guide targeted treatment strategies and predict disease outcomes.

**Scope of thesis**

EC is a common and increasingly prevalent malignancy in females with clinical heterogeneity and molecular complexity. The incorporation of molecular classification has reshaped our understanding of tumour biology and clinical management. The two studies presented in this thesis contribute to the evolving landscape by 1) dissecting genetic drivers of methylation within the frequent EEC subtype and 2) use of advancing NGS-based diagnostics for MSI testing, a clinically actionable marker.

The first manuscript focuses on the interplay of somatic *KRAS*/*ARID1A* mutations, *MLH1* promoter methylation (a key driver of sporadic MSI) and LINE-1 methylation and their association with MSI status in EEC.

The hypothesis arises from evidence that *KRAS* mutations, common in low-grade EEC, promote oncogenic signalling through the RAS/MAPK pathway. This pathway is also critically involved in CRC, where it is frequently driven by *BRAF* mutations, which are strongly associated with the CpG Island Methylator Phenotype (CIMP) and widespread epigenetic dysregulation. whereas *ARID1A* loss, a biomarker of chromatin remodelling defects, contributes to epigenetic deregulation and genomic instability. Similarly, in EEC, we hypothesize that *KRAS* activation, coupled with *ARID1A* loss (a biomarker of chromatin remodelling defects), contributes to epigenetic

deregulation and genomic instability.

By quantifying LINE-1 methylation as indicator of global DNA methylation, the study aims to uncover how the co-occurrence of *KRAS* and *ARID1A* mutations may outline the epigenetic landscape of EEC and whether these alterations correlate with MSI status, reflecting underlying dMMR or *MLH1* promoter methylation. This integrative analysis seeks to define molecular subsets driven by concurrent genetic and epigenetic aberrations, providing new insight into the biological heterogeneity of EEC and its potential diagnostic and prognostic implications.

The second manuscript focuses on diagnostic and clinical utility of an NGS-based MSI detection panel in CRC and EC, validated against conventional MMR-IHC and PCR-based MSI testing using capillary electrophoresis (MSI-CE). Through biostatistical and model-based computational approaches, this study refines a 122-loci panel to a 51-loci subset with improved diagnostic accuracy and sensitivity thereby offering a practical, high-throughput alternative to standard MSI testing with simultaneous mutational profiling.

Together, these studies aim to enhance the molecular stratification of EEC by integrating genetic drivers, epigenetic alterations and diagnostic innovation to facilitate the translation of integrative genomic analyses into routine clinical practice.

# AIMS OF THE RESEARCH

## Manuscript 1: KRAS/ARID1A mutations and methylation study

1. To determine the frequency of *KRAS* mutations and co-occurrence with commonly mutated genes in EEC.

2. To quantify global LINE-1 methylation levels and MSI status to evaluate their relationship with *KRAS* mutations and *MLH1* methylation.

3. To assess the association of *KRAS* and *ARID1A* mutations with distinct methylation profiles (including *MLH1* methylation) or MSI status.

4. To explore the potential biological and prognostic implications of genetic-epigenetic interactions in the molecular evolution of EEC.

## Manuscript 2: NGS-based MSI detection and validation

1. To validate the performance of an NGS-based MSI detection panel across EC and CRC samples.

2. To compare the NGS-based MSI results with conventional IHC for MMR proteins and PCR-based MSI testing using mononucleotide repeat markers.

3. To refine and optimise the MSI detection panel using statistical and model-based clustering to identify a robust subset of loci with maximal diagnostic accuracy.

4. To evaluate the applicability of NGS-based MSI assessment for routine molecular diagnostics and its integration into molecular classification.

## Integrative Aim

The collective aim of this thesis is to advance precision oncology in EEC by establishing targeted NGS as a comprehensive and clinically meaningful molecular diagnostic platform. By enabling the simultaneous detection of key driver mutations (such as *POLE, TP53, KRAS, ARID1A, CTNNB1*) and the accurate assessment of MSI status, this work seeks to unify multiple current diagnostic methodologies into a single high-throughput NGS approach. Through the integration of genetic alterations with epigenetic biomarkers, including gene-specific (*MLH1* silencing associated with dMMR) and global (LINE-1) methylation changes, the thesis aims to deepen the biological understanding of MSI development and its heterogeneity within EEC. In parallel, the diagnostic performance of a targeted NGS-based MSI detection panel is evaluated in both EEC

and CRC samples to define a robust subset of informative loci with improved discriminatory accuracy. Ultimately, the integrative molecular characterisation aims to demonstrate the feasibility of a single-platform NGS to improve tumour classification, refine prognostic stratification and support translation into precision oncology workflows, offering a scalable alternative to existing multimodal testing strategies used in routine clinical management.

## MANUSCRIPT 1

# Co-mutation of *KRAS* and *ARID1A* correlates with the MSI/LINE-1 hypermethylation profile in endometrial carcinoma

## Abstract

Epigenetic deregulation and microsatellite instability (MSI) are key molecular features of endometrial carcinoma (EC). However, the interaction between oncogenic signaling and the methylation landscape in EC remains poorly defined. Somatic alterations in *KRAS* have been implicated in both oncogenic transformation and epigenetic dysregulation, while *ARID1A*, a chromatin remodeling gene, contributes to transcriptional reprogramming and altered DNA methylation. This study investigated the relationship of pathogenic *KRAS* and *ARID1A* mutations with *MLH1* promoter methylation and global methylation patterns in a consecutive institutional cohort of endometrioid ECs (EECs) compared with a public dataset.

EEC cohort was analysed by integrating next generation sequencing (NGS) for mutation profiling, immunohistochemistry (IHC) for DNA mismatch repair (MMR) proteins and capillary electrophoresis-based MSI testing. *MLH1* promoter methylation was assessed using MS-MLPA, while global methylation levels were quantified through LINE-1 pyrosequencing.

*KRAS* mutations were enriched in *MLH1*-hypermethylated MSI tumours as well as in LINE-1 hypermethylated tumours, suggesting a functional link between oncogenic activation and epigenetic silencing. Furthermore, co-occurrence of *KRAS*-*ARID1A* mutations showed association with MSI and *MLH1* promoter methylation. Notably, *KRAS*-*ARID1A* co-mutant tumours displayed higher LINE-1 methylation levels than wildtype or single-mutant cases, indicating a CpG Island Methylator Phenotype (CIMP)-like phenotype.

The coexistence of *KRAS* activation and *ARID1A* loss defines an EEC subset characterised by *MLH1* hypermethylation, MSI and increased global methylation. These findings suggest the convergence of RAS signaling and chromatin remodeling pathways in shaping epigenetic instability and promoting MMR deficiency (dMMR). This integrated mechanism provides comprehension of methylation-associated EC evolution and may refine diagnostic and therapeutic implications.

**Keywords:** Endometrial carcinoma, *KRAS*, *ARID1A*, *MLH1* methylation, MSI, LINE-1, chromatin remodeling, epigenetic modulation.

## Introduction

Endometrial carcinoma (EC) is the most frequent gynaecological malignancy globally, with the endometrioid subtype (EEC) accounting for the majority of cases and generally conferring a favourable prognosis [1, 2]. Despite this overall positive outlook, a substantial subset of EEC exhibits aggressive behaviour and disease recurrence, reflecting the profound molecular heterogeneity of this tumour type. This clinical reality has driven a paradigm shift from traditional histopathological classification to the integrated molecular classification defined by The Cancer Genome Atlas (TCGA) and refined by Leon-Castillo *et al*., which stratifies ECs into four prognostic subgroups: *POLE*-ultramutated (*POLE*mut), DNA mismatch repair-deficient (MMRd) or microsatellite instability (MSI) hypermutated, p53-abnormal (p53abn) or copy number-high (CN-H) and no specific molecular profile (NSMP) or copy number-low (CN-L) [3, 4].

While the TCGA molecular subtypes provide broad risk stratification, a deeper understanding of gene-specific drivers within MMRd/MSI and NSMP/CN-L subgroups is crucial. Focusing on these subtypes, *KRAS* mutations, occurring in approximately 15-30% of EECs, define a major molecular subset [3, 5]. Functionally, *KRAS* drives proliferative signalling [6] and *KRAS*-mutated ECs are significantly associated with endometrioid histology [3, 7]. Recently, the development of covalent inhibitors for $KRAS^{G12C}$ has shown the potential to selectively suppress oncogenic RAS signalling, creating new opportunities for precision therapy [8].

Crucially, EEC pathogenesis also involves significant epigenetic changes. Approximately 20-40% of ECs are of MMRd subtype, most commonly due to hypermethylation of the *MLH1* promoter in sporadic cases [9]. Concurrently, global DNA methylation, often assessed via long interspersed nuclear element-1 (LINE-1) sequence methylation, serves as a biomarker for overall genomic stability, with its loss traditionally associated with increased instability [10-13]. Therefore, understanding how known genetic drivers interact with these key epigenetic processes is essential for a comprehensive molecular portrait. However, the specific interaction between the *KRAS* oncogene and these crucial epigenetic mechanisms (*MLH1* methylation and global methylation status) remains poorly defined in EEC.

A growing body of evidence concerning the co-mutation of *KRAS* and the chromatin remodeler *ARID1A* in EEC [3, 14-16], suggests a complex interplay between RAS/MAPK signalling and epigenetic dysregulation. The precise molecular link between activation and the establishment of an MSI phenotype, distinct from the *KRAS* mutual exclusivity seen in colorectal carcinoma (CRC) [17, 18], represents a critical gap in the literature.

In this context, our study aims to integrate molecular and epigenetic profiling to characterise the *KRAS*-mutated subgroup of EEC. Specifically, we assessed the frequency and co-occurrence of *KRAS* mutations, MSI and LINE-1 methylation and investigated their clinicopathological associations. By combining *in-silico* analysis of a large public dataset with molecular profiling of a consecutive, single-institution EEC cohort, we seek to elucidate potential molecular drivers of methylation and identify a distinct *KRAS*-defined MMRd subgroup with specific prognostic and therapeutic implications.

## Materials and methods

The study was conducted on a consecutive cohort of 83 EEC cases collected during 2018-2019 from the Unit of Anatomic Pathology, ASST-Sette Laghi/University of Insubria, Varese (VA), Italy. All patients in this cohort provided informed consent for the molecular classification tests, which were performed for clinical relevance and potential prognostic/therapeutic application. To test our hypothesis, we initially performed an *in-silico* analysis using the publicly available data from cBioPortal (https://www.cbioportal.org/) [19, 20]. The uterine cancer datasets were downloaded and filtered to include only EECs [3, 21-32]. To ensure data consistency, we retained cases with a single sample per patient from primary tumours and profiled for mutations, resulting in a subset of 1300 cases. Clinicopathological and survival data was collected for VA EEC cohort to compare with cBioPortal data, in addition, data for type of surgical procedure, immunohistochemistry (IHC) of MMR proteins and treatment method was also available for VA EECs. Clinicopathologic variables were systematically collected and analysed for all patients included in the study cohort (n=83). Demographic characteristics included age at diagnosis, analysed both as a continuous variable and dichotomised at 65 years. Histopathologic features encompassed tumour grade (low-grade [G1–G2] vs. high-grade [G3]), FIGO stage (early stage [I–II] vs. advanced stage [III–IV]) and type of hysterectomy (simple vs. radical). The depth of myometrial invasion (<50% vs. ≥50%) and the presence or absence of lymph node metastases were recorded. Treatment-related data included the administration and type of adjuvant therapy, specifying external beam radiotherapy (EBRT), brachytherapy and/or chemotherapy. Associations between *KRAS* mutational status and clinicopathologic or molecular features were explored using appropriate statistical tests, as detailed below.

### Molecular analysis

Genomic DNA was extracted from three representative 8 µm sections of formalin-fixed and paraffin-embedded (FFPE) tissues using the Maxwell® RSC FFPE Plus DNA Kit (Promega, USA). After extraction, DNA was quantified using Qubit® dsDNA High Sensitivity Assay Kit and Qubit® 2.0 Fluorometer (Invitrogen™, Thermo Fisher Scientific Inc., USA).

### NGS analysis for mutation profiles

To analyse the mutational profile, Next-Generation Sequencing (NGS) was performed for 83 EECs using the Archer® VariantPlex® Solid Tumour kit targeting 67 genes, customised to include three spike-in genes *POLE, POLD1* and *ARID1A,* for single nucleotide variants (SNVs) and insertions and deletions (indels) in solid tumours (supplementary table 1). Followed by PreSeq DNA quality control (QC) assay, 50-100 ng DNA was subjected to fragmentation, end repairing,

ligation with molecular barcode (MBC) adapters and nested PCR to prepare libraries. Libraries were quantified with KAPA qPCR, normalised and sequenced on Illumina® MiSeq Sequencing System (Illumina Inc, USA) according to the manufacturers' instructions. The detailed version of the Archer® VariantPlex® NGS protocol is available at http://www.idtdna.com/. After sequencing, the FASTQ files were analysed using Archer Analysis software version 7.1 (https://analysis.archerdx.com/). The detected gene variants were carefully selected by taking account of sample metrices, read depth, allele frequency (AF) and pathogenicity. Only pathogenic and likely-pathogenic variants were considered as mutations in genes, while variant of unknown significance (VUS), likely-benign and benign were considered as wildtype.

**MSI status analysis**

The MSI status of EECs was assessed with OncoMate™ MSI Dx Analysis System (Promega, USA) by amplifying five microsatellite markers (NR-21, BAT-25, BAT-26, NR-24 and MONO-27) using 10 ng DNA on Veriti thermocycler (Applied Biosystems®, Thermo Fisher Scientific Inc., USA). It was followed by the size-specific separation of markers by capillary electrophoresis using SeqStudio™ Genetic Analyzer (Applied Biosystems®, Thermo Fisher Scientific Inc., USA) and the results were analysed using GeneMapper™ version 6.0 (Thermo Fisher Scientific Inc., USA). The samples were designated as MSI or microsatellite stable (MSS) according to the established criteria of the manufacturer. The IHC status of MMR proteins was already available and integrated with the molecular testing of MSI status for confirmation.

***MLH1* methylation analysis**

In addition to molecular testing of MSI and MMR proteins status with IHC, we analysed the methylation of MMR genes using SALSA® (Selective Adaptor Ligation, Selective Amplification) MS-MLPA® (Methylation-Specific Multiplex Ligation-dependent Probe Amplification) Probemix ME011 (MRC-Holland, The Netherlands) to identify sporadic epigenetic silencing of *MLH1,* thereby distinguishing from possible inherited deficiency of MMR (dMMR) system. 25 tumours were selected for *MLH1* methylation testing which included: 21 dMMR cases assessed negative for MLH1 protein with IHC; and four additional cases, two cases with sub-clonal loss of MLH1 protein but MSS and two dMMR determined as MLH1 positive by IHC but MSI by molecular analysis. The detail of ME011 probes is listed in (supplementary table 2). 80-100 ng DNA of 25 patient samples and 3 reference samples was denatured and hybridised with MS-MLPA probes and subsequently amplified by PCR. Following PCR, the samples were assessed for difference in fragment sizes by capillary electrophoresis using SeqStudio™ Genetic Analyzer (Applied Biosystems®, Thermo Fisher Scientific Inc., USA). The results were analysed using

Coffalyser.Net version 220513 (MRC-Holland, The Netherlands).

**LINE-1 methylation analysis**

To assess the methylation status of LINE-1 sequences, the DNA sample of 83 patients was first subjected to bisulfite conversion using EZ DNA Methylation-Gold™ Kit (Zymo Research®, USA) according to the manufacturer's protocol. Then the converted DNA was amplified by qRT-PCR using LINE-1 specific forward and reverse primers and GoTaq® DNA Polymerase (Promega). Subsequent to amplification, the samples were run for pyrosequencing in Qiagen PyroMark Q96 ID and were analysed with PyroMark Q96 ID software version 2.5. For each sample, percentage of methylation of four CpG promoter sites (CpG2, CpG3, CpG5 and CpG6) in LINE-1 were noted and mean was taken to get the methylation level of LINE-1. The detailed protocol is described in the work of Stefanoli *et al*. [33]. A data-driven cutoff of 59% was estimated using model-based clustering [34] algorithm (R library *mclust*) to consider a case as either hypomethylated (LINE-1 <59%) or hypermethylated (LINE-1 ≥59%). Comparative analysis was performed between LINE-1 methylation level and *KRAS*-mutated cases.

**Statistical analysis**

All results were statistically evaluated using GraphPad Prism version 8.0.1. Associations between categorical variables were tested using the Fisher's exact test, continuous variables with Student's *t*-test, survival analysis was performed with log-rank (Mantel-Cox) tests and plotted using Kaplan-Meier curves.

## Results

### Overview of cohorts

We analysed two independent cohorts of EEC: the publicly available dataset from cBioPortal filtered to retain 1300 cases with single primary tumour per patient analysed for mutations and our VA cohort of 83 consecutive cases collected during 2018-2019. The evaluation of clinical and molecular parameters was performed for both datasets by segregating based on the genomic status of *KRAS* as either mutated or wildtype. Clinicopathological and molecular features for cBioPortal data and VA cohort are reported in table 1 and 2, respectively.

The cBioPortal dataset included 1300 EEC cases, comprising 320 (24.6%) *KRAS*-mutated and 980 (75.4%) *KRAS*-wildtype tumours. The age of diagnosis was 62.4 ± 11.5 (mean ± SD) years with approximately 40% cases reported in elderly (≥65 years) without any significant association with age of onset. The dataset was predominantly composed of low-grade (72.3%) , with *KRAS*-mutated group showing a significantly higher proportion of low-grade tumours compared to *KRAS*-wildtype group (*p=0.029*). The cohort consisted mainly of early-stage (I-II, 81.5%) compared to advanced-stage (III-IV, 18.5%) cases with similar incidence of staging in *KRAS*-segregated groups. Overall, 39.5% cases were MSI however, *KRAS*-mutated tumours were dominant with 53.8% MSI over 46.2% MSS, whereas *KRAS*-wildtype were mostly MSS (65.3%) compared to MSS (34.7%) showing significant *p-value* of 0.003. *MLH1* silencing by methylation was present in 37.3% cases, comparison within *KRAS* groups showed highly significant distribution of *MLH1* silencing with 56.2% cases silenced vs. 43.8% not-silenced in *KRAS*-mutated and 30.8% cases silenced vs. 69.2% not-silenced in *KRAS*-wildtype group (*p=0.0001*).

**Table 2: Clinicopathological data of cases from cBioPortal (n=1300).**

| Feature | All (n=1300) | *KRAS*-mut (n=320) | *KRAS*-wt (n=980) | *P*-value |
|---|---|---|---|---|
| Diagnosis Age, years | | | | |
| Mean ± SD | 62.4 ± 11.5 | 62.2 ± 10.6 | 62.4 ± 11.8 | ns |
| Range | 31-90 | 35-87 | 31-90 | |
| ≥65 | 154 (39.9%) | 32 (35.6%) | 122 (41.2%) | ns |
| <65 | 232 (60.1%) | 58 (64.4%) | 174 (58.8%) | |
| NA | 914 | 230 | 684 | |
| Grade | | | | |
| Low (G1/2) | 914 (72.3%) | 243 (77.1%) | 671 (70.6%) | **0.029** |
| High (G3) | 351 (27.7%) | 72 (22.9%) | 279 (29.4%) | |
| NA | 35 | 5 | 30 | |
| Stage | | | | |
| Early (I-II) | 233 (81.5%) | 61 (82.4%) | 172 (81.1%) | ns |
| Advanced (III-IV) | 53 (18.5%) | 13 (17.6%) | 40 (18.9%) | |
| NA | 1014 | 246 | 768 | |
| MSI status | | | | |
| MSI | 124 (39.5%) | 42 (53.8%) | 82 (34.7%) | **0.003** |
| MSS | 190 (60.5%) | 36 (46.2%) | 154 (65.3%) | |
| NA | 986 | 242 | 744 | |
| *MLH1* Silencing | | | | |
| Yes | 107 (37.3%) | 41 (56.2%) | 66 (30.8%) | **0.0001** |
| No | 180 (62.7%) | 32 (43.8%) | 148 (69.2%) | |
| NA | 1013 | 247 | 766 | |

*Associations between continuous variables were tested with Student's t-test and for categorical variables using the Fisher's exact test. mut, mutated; wt, wildtype; SD, standard deviation; MSI, microsatellite instability; MSS, microsatellite stable.*

The VA cohort composed of 83 consecutive EECs included 17 (20.5%) *KRAS*-mutated and 66 (79.5%) *KRAS*-wildtype cases. The cohort presented diagnosis age of 64.5 ± 9.4 years (mean ± SD) including 54.2% elderly patients. Both *KRAS*-mutated and *KRAS*-wildtype groups exhibited comparable age distribution and equal proportion of younger and older patients. Most tumours in VA cohort were low-grade (G1/2, 84.3%) and early-stage (I-II, 97.6%), characteristic of EEC subtype, compared to high-grade (G3, 15.7%) and advanced-stage (III-IV, 2.4%). *KRAS*-mutated group was composed of 94.1% low-grade and exclusively of early-stage tumours, while wildtype group had 81.8% low-grade and 97% early-stage tumours. Among early-stage cases, stage IA represented the majority with 44 tumours (8 *KRAS*-mutated, 36 wildtype), followed by stage IB with 28 tumours (6 mutated, 22 wildtype) and stage II with 9 tumours (3 mutated, 6 wildtype). Only two cases were classified as stage III (advanced), both *KRAS*-wildtype, emphasizing the dominance of early-stage disease across the cohort. Surgical procedure and myometrial invasion included data of 70 patients with simple vs. radical hysterectomy performed for 97.1% and 2.9% cases respectively while deep (≥50%) vs. shallow infiltration (<50%) seen in 72.9% and 23.1%

cases. Lymph node evaluation of 49 cases showed metastasis in 8.2% cases.

Molecular testing performed for MSI status analysis confirmed 32 (38.6%) MSI and 51 (61.4%) MSS tumours with similar distribution of MSI and MSS in *KRAS* groups. Among the 25 cases evaluated for *MLH1* promoter methylation, 76% were detected silenced including all six analysed cases as *KRAS*-mutated. 19/22 (86.4%) MLH1-deficient dMMR tumours were found *MLH1* methylated, indicating a predominance of sporadic dMMR. Considering all MSI cases (*MLH1*-methylated and unmethylated), the association with *KRAS* mutation status was far from statistical significance, however, a trend emerged when focusing specifically on *MLH1*-methylated MSI tumours, consistent with observations in cBioPortal cohort. LINE-1 methylation analysis revealed a mean methylation level of 61.9% for EECs and with bifurcation at cutoff of 59%, hypermethylated (≥59%) cases were frequent with 75.9% prevalence than 24.1% of hypomethylated (<59%). Although not significant, *KRAS*-mutated group showed more cases with LINE-1 hypermethylation (88.2%) than *KRAS*-wildtype cases (72.7%) ($p=0.2$).

Due to the limited sample size of VA cohort, most clinicopathological associations did not reach statistical significance. However, similar to cBioPortal cohort, we observed trends towards occurrence of *KRAS* mutations in low-grade, early-stage and hypermethylated tumours.

**Association of *KRAS* mutations with *MLH1* methylation**

To investigate the genomic and epigenomic relationship, we compared *KRAS* mutations with *MLH1* methylation in sporadic EECs of cBioPortal cohort. Among 97 *MLH1*-methylated MSI tumours, *KRAS* mutations were identified in 38.1% (37/97), whereas in remaining 179 *MLH1*-unmethylated MSI/MSS tumours, the mutation frequency was 17.9% (32/179). This resulted an odds ratio (OR) of 2.8, indicating that *KRAS* mutations and *MLH1*-methylation significantly co-occurred nearly three times ($p=0.0004$).

Similarly, we assessed the 19 sporadic MSI cases (MLH1-negative and *MLH1*-methylated) in the VA cohort. Within this *MLH1*-methylated group, 31.6% (6/19) carried *KRAS* mutations, whereas in rest of the 64 tumours (13 MSI including 3 MLH1-negative and *MLH1*-unmethylated; 51 MSS), 17.2% (11/64) tumours were *KRAS*-mutated. This corresponded to an OR of 2.2, indicating that *KRAS* mutations were more than twice as frequent in *MLH1*-methylated tumours, although the association did not reach statistical significance ($p=0.2$).

**Table 3: Clinicopathological data of VA EEC cohort (n=83).**

| Feature | All (n=83) | *KRAS*-mut (n=17) | *KRAS*-wt (n=66) | *P*-value |
|---|---|---|---|---|
| Diagnosis Age, years | | | | |
| Mean ± SD | 64.5 ± 9.4 | 62.4 ± 10.5 | 65.1 ± 9.2 | ns |
| Range | 43-85 | 47-85 | 43-83 | |
| ≥65 | 45 (54.2%) | 8 (47.1%) | 37 (56.1%) | ns |
| <65 | 38 (45.8%) | 9 (52.9%) | 29 (43.9%) | |
| Grade | | | | |
| Low (G1/2) | 70 (84.3%) | 16 (94.1%) | 54 (81.8%) | ns |
| High (G3) | 13 (15.7%) | 1 (5.9%) | 12 (18.2%) | |
| Stage | | | | |
| Early (I-II) | 81 (97.6%) | 17 (100%) | 64 (97%) | ns |
| Advanced (III-IV) | 2 (2.4%) | 0 (0%) | 2 (3%) | |
| Hysterectomy | | | | |
| Simple | 68 (97.1%) | 12 (100%) | 56 (96.6%) | ns |
| Radical | 2 (2.9%) | 0 (0%) | 2 (3.4%) | |
| NA | 13 | 5 | 8 | |
| Myometrial infiltration | | | | |
| <50% | 19 (27.1%) | 3 (25%) | 16 (27.6%) | ns |
| ≥50% | 51 (72.9%) | 9 (75%) | 42 (72.4%) | |
| NA | 13 | 5 | 8 | |
| Lymph node metastasis | | | | |
| Presence | 4 (8.2%) | 1 (12.5%) | 3 (7.3%) | ns |
| Absence | 45 (91.8%) | 7 (87.5%) | 38 (92.7%) | |
| NA | 34 | 9 | 25 | |
| MSI status | | | | |
| MSI | 32 (38.6%) | 7 (41.2%) | 25 (37.9%) | ns |
| MSS | 51 (61.4%) | 10 (58.8%) | 41 (62.1%) | |
| MMR (IHC) | | | | |
| dMMR | 31 (37.3%) | 7 (41.2%) | 24 (36.4%) | ns |
| pMMR | 52 (62.7%) | 10 (58.8%) | 42 (63.6%) | |
| *MLH1* Silencing | | | | |
| Yes | 19 (76%) | 6 (100%) | 13 (68.4%) | ns |
| No | 6 (24%) | 0 (0%) | 6 (31.6%) | |
| NA | 58 | 11 | 47 | |
| LINE-1 methylation | | | | |
| Hypomethylation <59% | 20 (24.1%) | 2 (11.8%) | 18 (27.3%) | ns |
| Hypermethylation ≥59% | 63 (75.9%) | 15 (88.2%) | 48 (72.7%) | |
| Therapy | | | | |
| Adjuvant | 33 | 7 | 26 | |
| Radiotherapy EBRT/Brachytherapy | 30 | 7 | 23 | |
| Chemotherapy | 10 | 3 | 7 | |

*Associations between continuous variables were tested with Student's t-test and for categorical variables using the Fisher's exact test. mut, mutated; wt, wildtype; SD, standard deviation; MSI, microsatellite instability; MSS, microsatellite stable; MMR, DNA mismatch repair protein system; dMMR, MMR deficiency; pMMR, MMR proficiency; EBRT, external beam radiotherapy.*

**Association of *KRAS* mutations with LINE-1 methylation**

In addition to *MLH1* methylation, global methylation profiling by LINE-1 pyrosequencing revealed a wide range of methylation percentages (48-70%, mean ± SD = 61.9 ± 5.1%). When stratified by *KRAS* mutation status, *KRAS*-mutated tumours (n=17) exhibited higher LINE-1 methylation levels (mean ± SD = 64.4 ± 3.6%) compared with *KRAS*-wildtype tumours (n=66) (mean ± SD = 61.3 ± 5.3%) (*p=0.0079*). The direction of this association suggests that *KRAS* activation is linked with a global hypermethylated profile (figure 1).

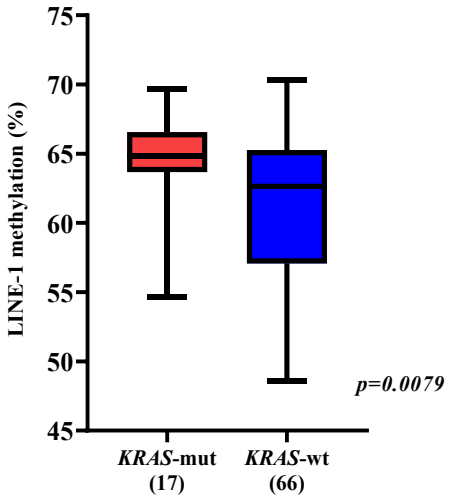



**Figure 2: LINE-1 methylation levels by *KRAS* status in VA EEC cohort.** Comparison of LINE-1 methylation levels was done between *KRAS*-mutated (n=17) and *KRAS*-wildtype (n=66), showing a significant relation (p=0.0079). *P*-value was calculated using two-tailed t-test with Welch's correction. mut, mutated; wt, wildtype.

After estimation of LINE-1 methylation cutoff of 59% using model-based clustering, we segregated the LINE-1 hypermethylated (≥59%) and hypomethylated (<59%) cases according to the *KRAS* status (table 1) and found a higher proportion of *KRAS*-mutated tumours showing hypermethylation. Although hypermethylation of LINE-1 in *KRAS*-mutated tumours did not reach statistical significance (*p=0.2)*, the trend was in parallel to enrichment of *MLH1*-methylation in *KRAS*-mutated group, indicating that *KRAS* may promote a methylator phenotype in EECs.

Despite the limited number, these cases point to a distinct molecular pattern in which oncogenic activation and epigenetic deregulation coexist. This convergence suggests that *KRAS*-driven signalling may promote both focal promoter hypermethylation (*MLH1*) leading to mismatch repair deficiency and global DNA hypermethylation (LINE-1).

**Spectrum of *KRAS* mutations**

In cBioPortal cohort, we found 328 *KRAS* variants: G12D and G12V were the most frequent with 90 (27.4%) and 82 (25%) respectively, followed by G13D/C/V in 60 (18.3%), G12A in 37 (11.3%), G12C in 21 (6.4%), Q61H/L/R in 10 (3%), G12S in 7 (2.1%) and others in 21 (6.4%),

shown in figure 2a.

In VA cohort, 17 *KRAS* variants were observed: G12D, G12V and G12C, each were present in 4 (23.5%), G13D/C in 3 (17.6%), both G12A and Q61H in 1 (5.9%), shown in figure 2b.

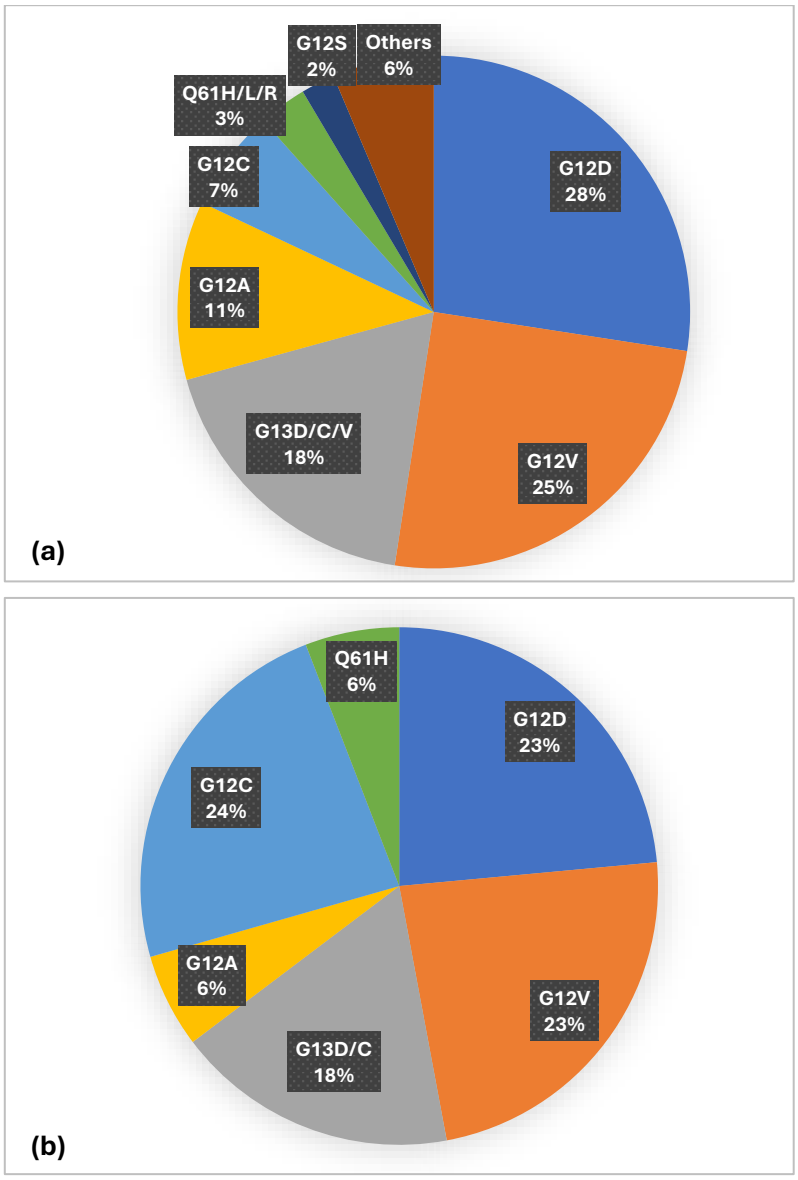



**Figure 3: Pie-chart showing frequency of (a) 328 *KRAS* variants in cBioPortal dataset and (b) 17 *KRAS* variants in VA EEC cohort.**

## Co-occurrence of *KRAS* with other gene mutations

We analysed the mutation landscape of EECs in both cohorts by classifying cases as mutated for a specific gene when at least one pathogenic or likely-pathogenic variants was detected. In the cBioPortal dataset, the most frequently mutated genes were *PTEN* (80.5%), *ARID1A* (59.3%), *PIK3CA* (51.2%), *PIK3R1* (36.2%), *CTNNB1* (26.7%), *KRAS* (24.6%), *TP53* (17.1%), *FBXW7* (13.7%) and *POLE* (11.7%). In VA EEC cohort, *PTEN* (78.3%), *ARID1A* (60.2%) and *PIK3CA* (48.2%) mutations predominated with similar frequencies, while *KRAS* mutations were observed

in 20.5% of cases.

Analysis of co-occurrence and mutual exclusivity within *KRAS*-mutated tumours revealed a consistent pattern across both datasets. The incidence of this pattern for frequently mutated genes is illustrated in figures 3a and 3b in cBioPortal and VA cohort, respectively.

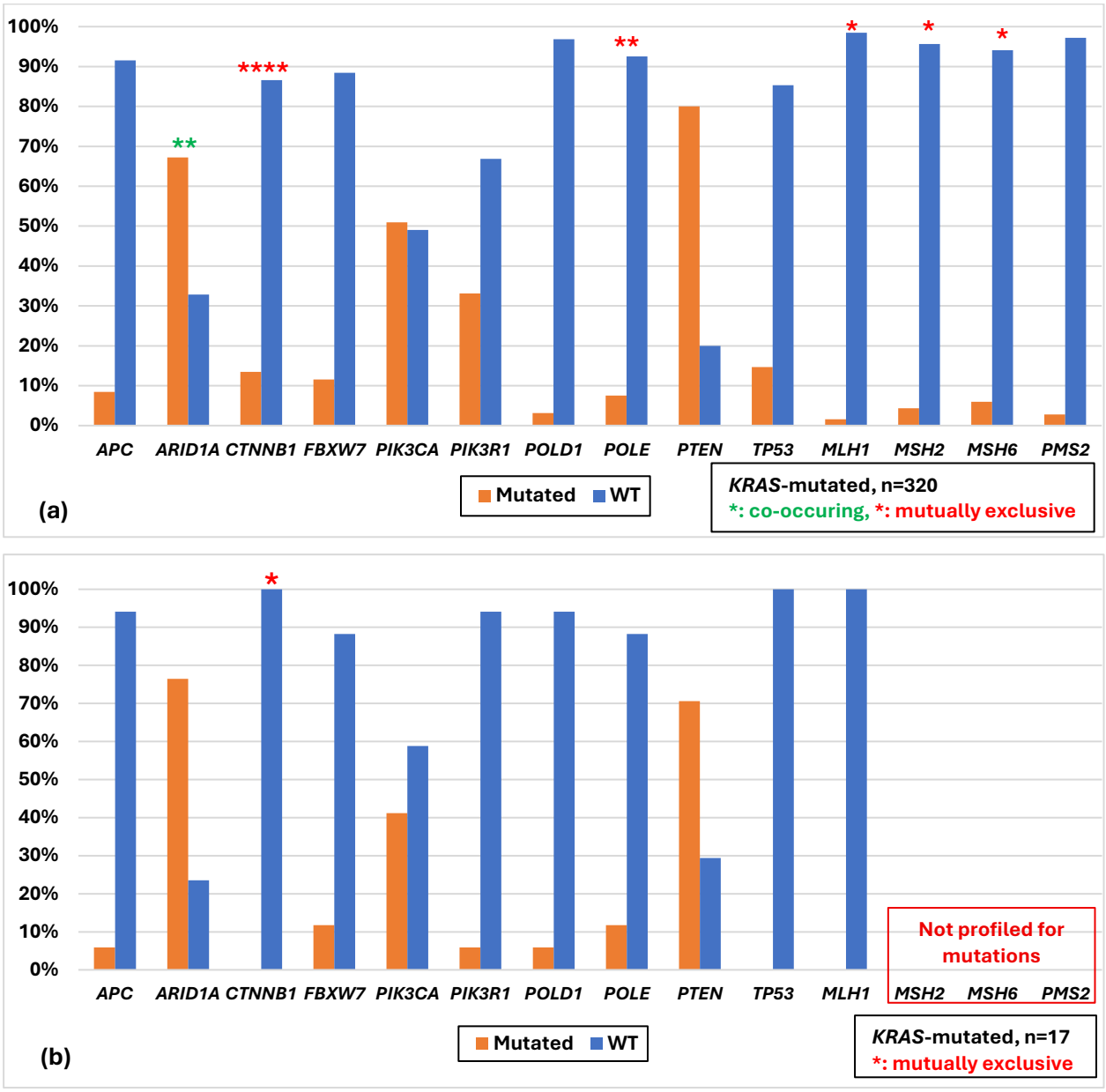



**Figure 4: Comparison of co-occurrence and mutual exclusivity of gene mutations in *KRAS*-mutated cases in (a) cBioPortal (n=320) and (b) in VA EEC cohort (n=17). (a)** Mutational data of 320 cases was available for all genes except for *POLD1*, for which data was available for 318 cases. **(b)** Mutational data was compared between all 17 *KRAS*-mutated cases. 3 out of 4 MMR genes (*MSH2, MSH6* and *PMS2*) were not profiled as they were not included in NGS panel (Archer® VariantPlex® Solid Tumour), however the data was available for cBioPortal cohort. Significant associations are marked with asterisks. *P*-value was calculated using Fisher's exact test.

In *KRAS*-mutated cases of cBioPortal dataset (n=320), significant co-occurrence and mutual exclusivity patterns were observed between with other frequently altered genes. *ARID1A* mutations were strongly co-occurring with *KRAS* mutations (*p=0.001*). In contrast, *CTNNB1*

(*p<0.0001*), *MLH1* (*p=0.0169*), *MSH2* (*p=0.0106*), *MSH6* (*p=0.0404*) and *POLE* (*p=0.0067*) mutations were significantly mutually exclusive with *KRAS*. Notably, *PTEN* mutations showed high co-occurrence with *KRAS* (80%) but without statistical significance, aligning with their frequent joint activation of the PI3K/AKT/mTOR pathway in EECs (figure 3a).

In *KRAS*-mutated cases of VA EEC cohort (n=17), similar but less pronounced trends were observed due to the smaller sample size. *ARID1A* mutations were present in 76% of *KRAS*-mutated tumours, consistent with the frequency seen in the larger cBioPortal dataset but did not reach statistical significance (*p=0.168*). *CTNNB1* mutations were significantly mutually exclusive with *KRAS* (*p=0.0336*), whereas *TP53* displayed a borderline exclusivity (*p=0.0607*) (figure 3b).

**_KRAS/ARID1A_ mutations as epigenetic modulators**

The presence of significant co-occurrence of *KRAS* and *ARID1A* mutations in the cBioPortal cohort intrigued us to explore their combined potential as oncogenic and epigenetic drivers. Therefore, we evaluated the overlap of *KRAS* and *ARID1A* mutations with MSI and *MLH1* status (figure 4a and 4b). Among 52 cases with concurrent mutations, 27 (52%) showed MSI whereas among 122 cases, wildtype for both genes, mostly were MSS (90, 74%). The distribution of cases with dual (*KRAS* and *ARID1A*) and individual (either *KRAS* or *ARID1A*) mutations compared to MSI status is shown in figure 4a with significant *p-value* (*0.0008*). Likewise, among 48 *KRAS-ARID1A* mutated cases analysed for *MLH1* methylation, 27 (56%) were *MLH1*-methylated whereas 86 (74%) were *MLH1*-unmethylated among 116 *KRAS-ARID1A* wildtype cases. Figure 4b shows the frequency of mutations significantly associated with *MLH1* methylation (*p=0.0005*). In figure 4c, the overlapping of mutational status of *KRAS* and *ARID1A* along with MSI and *MLH1* is presented. The visual representation highlights patterns of co-occurrence, exclusivity and overall prevalence of molecular alterations, showing the relation of altered genes with epigenetic features in the cohort.

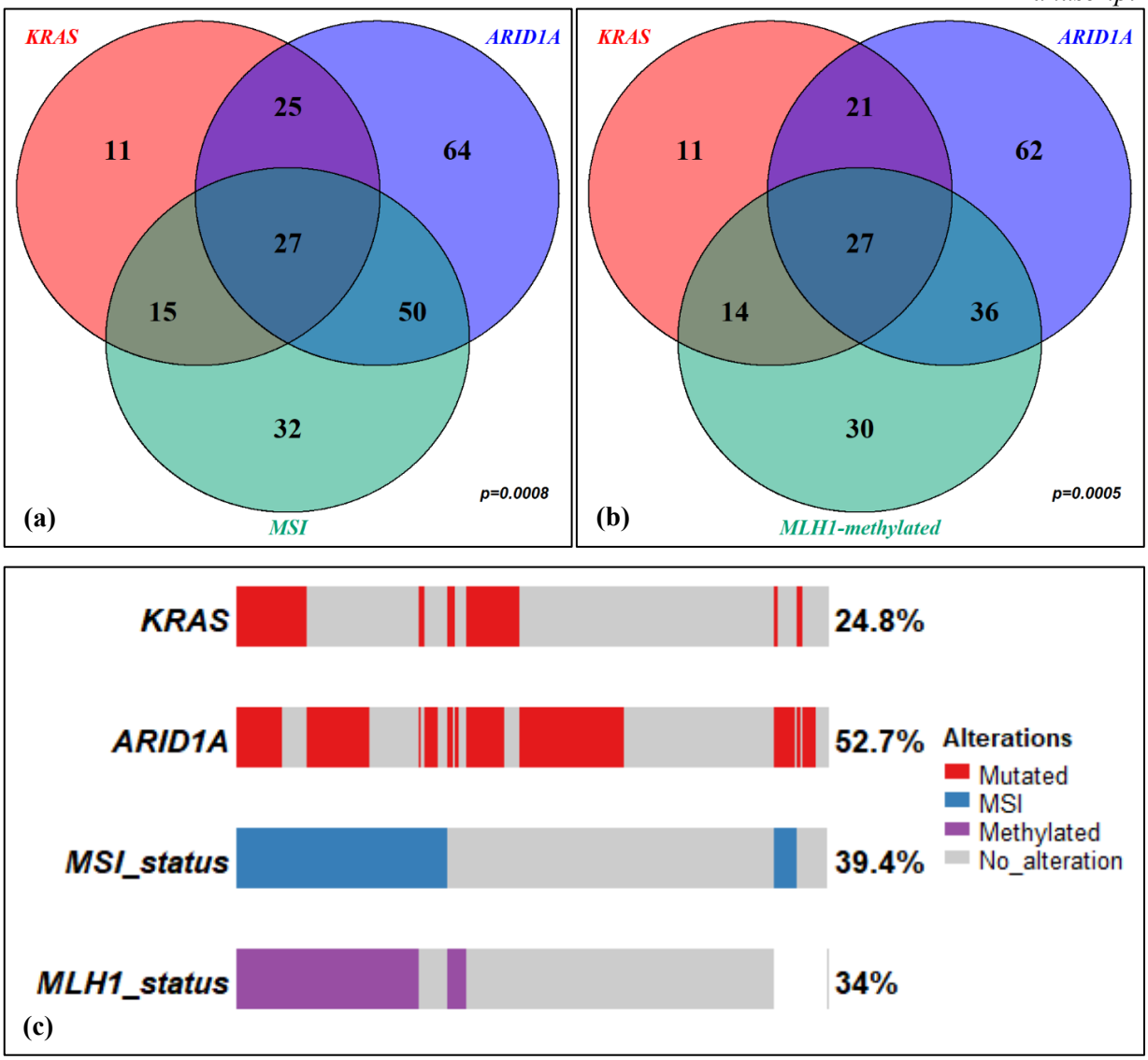



**Figure 5: Venn diagram showing the overlap of *KRAS* and *ARID1A* mutations with (a) MSI status and (b) *MLH1* methylation status in the cBioPortal dataset, (c) oncoprint for mutational, MSI and *MLH1* methylation status.** Analysis performed with Fisher's exact test between *KRAS/ARID1A* mutations and MSI and *MLH1* status showed significant co-occurrence with the majority of double-mutated cases belonging to the (a) MSI subgroup (p=0.0008) and (b) *MLH1*-methylated subgroup, supporting the combined role of *KRAS-ARID1A* mutations as potential oncogenic and epigenetic drivers. **(c)** Each row corresponds to a gene or molecular feature and each column represents a single sample. Mutated samples are indicated in red, MSI in blue, *MLH1*-methylated in purple and grey represents wildtype or unaltered status. Percentages on the left denote the proportion of samples showing alterations in each feature.

We examined the similar interaction in VA EEC cohort for mutations of *KRAS* and *ARID1A* with MSI and *MLH1* status. Although the relation was not significant due to smaller numbers, however we observed a trend similar to cBioPortal cohort with 5 (38%) of mutated cases as MSI and 22 (76%) of wildtype cases as MSS. Similarly, within cases tested for *MLH1* methylation, all 4 (100%) mutated tumours were *MLH1*-methylated. Moreover, we assessed the LINE-1 methylation levels in combination with mutational status of *KRAS* and *ARID1A* (figure 5). LINE-1 methylation was significantly higher in group with dual mutations (median 64.8 with interquartile range (IQR) 63.3-66.3) compared with wildtype having median of 61.8 (IQR: 58.7-64.7) (*p=0.0474*). *KRAS*-

only mutated tumours showed highest median (65.5, IQR: 64.1-68.9) and the most significant value (*p=0.0163*). A significant difference was also observed between *KRAS*-mut and *ARID1A*-mut (63.1, IQR: 55.9-66.1) groups (*p=0.0235*), indicating a dominant hypermethylation effect of *KRAS* mutations.

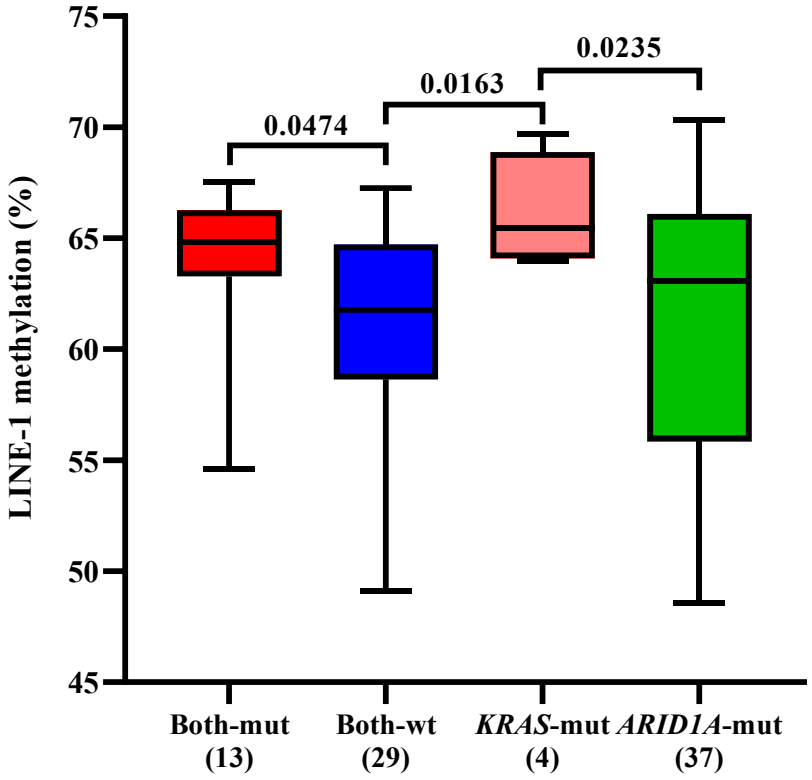



**Figure 6: Boxplot showing association between LINE-1 methylation and *KRAS/ARID1A* mutational status in VA EEC cohort.** Tumours harbouring *KRAS-ARID1A* mutations (Both-mut, n=13), displayed significantly higher LINE-1 methylation compared with wildtype (Both-wt, n=29) (p=0.0474), indicating a trend toward a hypermethylated phenotype associated with dual mutations. The *KRAS*-only mutated subgroup (*KRAS*-mut, n=4) exhibited the highest methylation level with a significant difference from *ARID1A*-only mutated group (*ARID1A*-mut, n=37).

## Survival analysis

Survival data were available for 422 and 379 patients of the cBioPortal cohort for overall survival (OS) and disease-free survival (DFS), respectively, with 48 events in OS and 60 events in DFS cases. Kaplan-Meier survival analysis of OS comparing *KRAS*-mutated and *KRAS*-wildtype tumours did not show any significance, however, *KRAS*-mutated tumours exhibited a trend toward longer DFS exhibiting borderline significance value (*p=0.0594*) (figure 6a).

For VA cohort, survival data were available for 69 patients, with 1 event in OS and 6 events in DFS, reflecting the favourable prognosis of early-stage EECs. Although OS and DFS between *KRAS*-mutated and *KRAS*-wildtype groups were not significant due to majority of cases being censored, a trend for better DFS was seen in *KRAS*-mutated tumours (*p=0.1771*) (figure 6b).

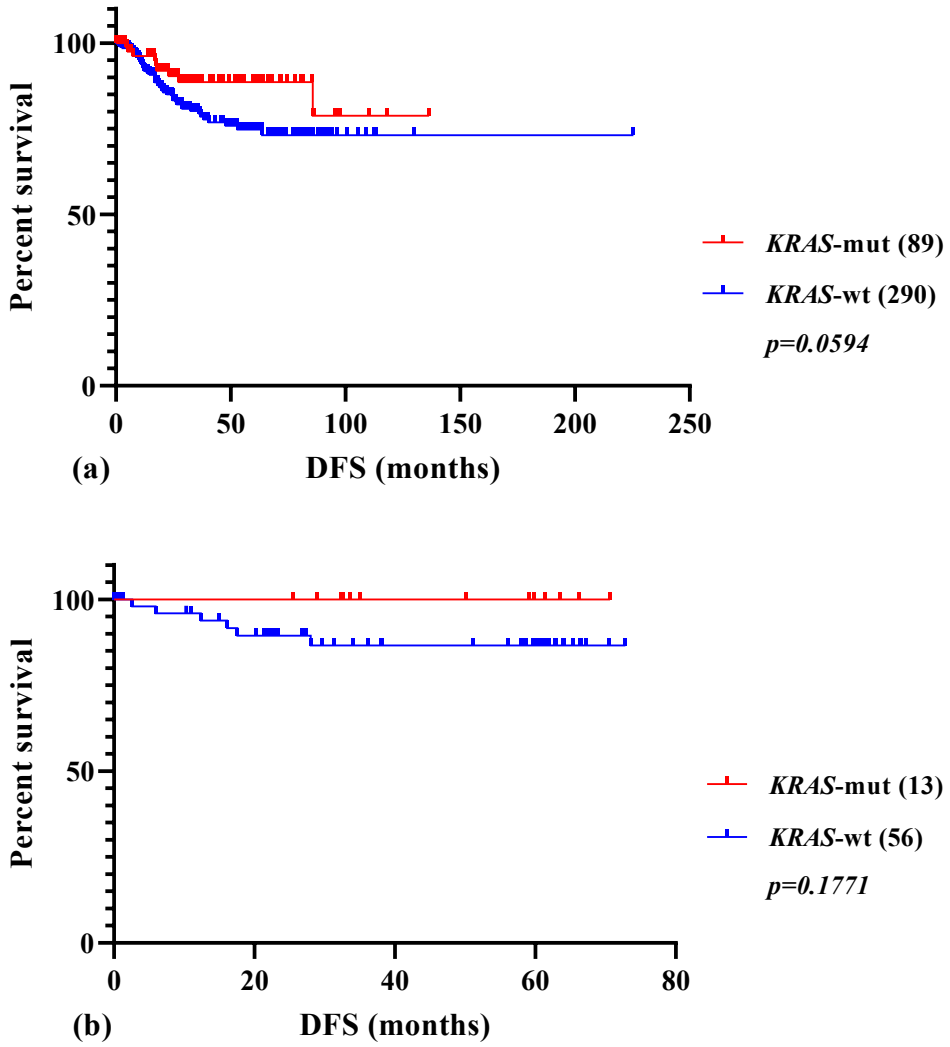



**Figure 7: Kaplan-Meier curves showing disease-free survival (DFS) in (a) cBioPortal dataset (n=379) and (b) VA cohort (n=69) comparing *KRAS*-mutated and *KRAS*-wildtype cases.** Though not significant, *KRAS*-mutated group showed better prognosis in both cohorts with longer DFS. Log-rank (Mantel-Cox) test was performed to calculate the *p*-value.

In the cBioPortal cohort, survival analysis revealed a favourable prognostic trend associated with *ARID1A* mutations. Kaplan-Meier curves demonstrated significantly improved OS in *ARID1A*-mutated tumours compared with wildtype counterparts ($p=0.0188$) (figure 7a) and a borderline trend toward longer DFS ($p=0.058$) (figure 7b). Comparison of MSI status and LINE-1 methylation between *KRAS*-mutated and *KRAS*-wildtype in VA cohort did not show significance ($p>0.05$).

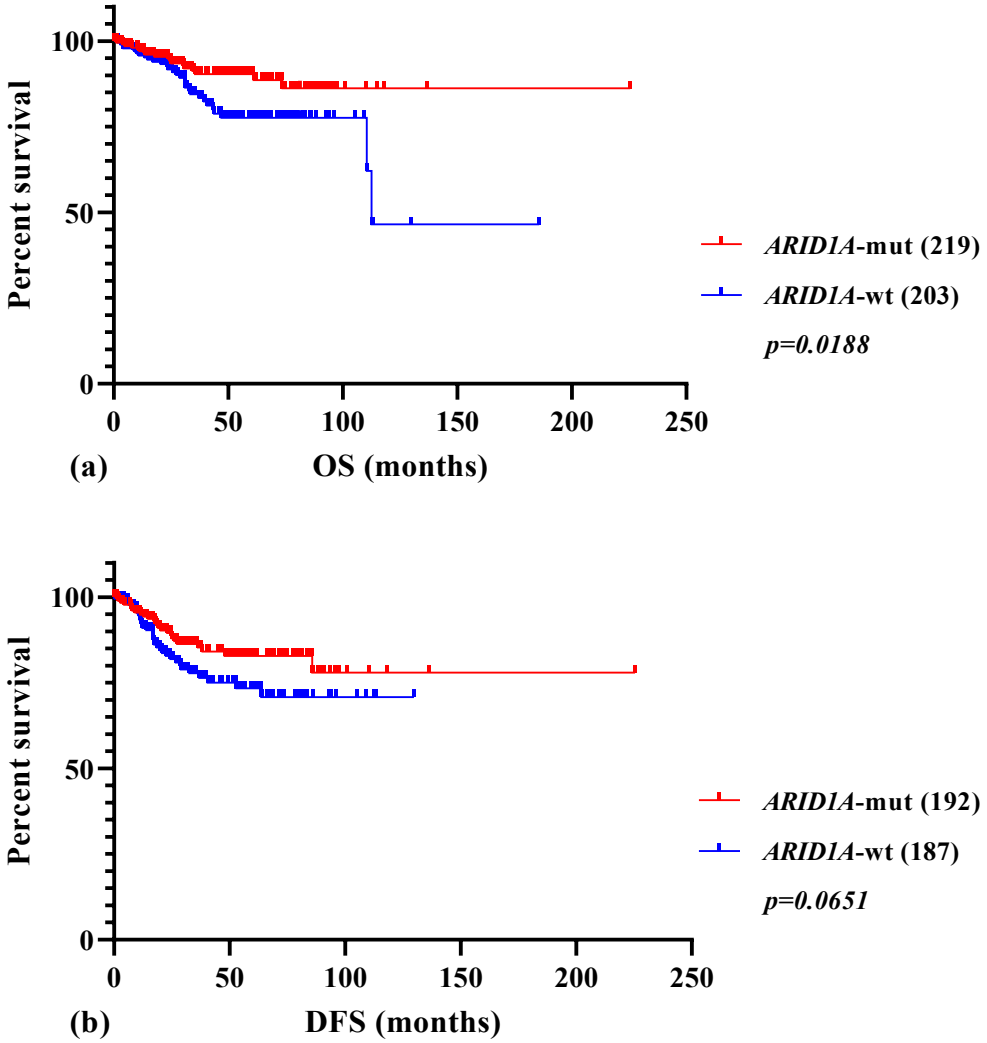



**Figure 8: Kaplan-Meier curves showing (a) overall survival (OS, n=422) and (b) disease-free survival (DFS, n=379) comparing *ARID1A*-mutated and *ARID1A*-wildtype cases in cBioPortal dataset.** Patients with *ARID1A* mutations demonstrated significantly better OS (p=0.0188) and a trend toward longer DFS (p=0.0651). Log-rank (Mantel-Cox) test was performed to calculate the *p*-value.

**Discussion**

Our study provides integrative insights into the molecular and epigenetic landscape of *KRAS*-mutated EECs, revealing distinct biological and clinical patterns. By combining *in-silico* analysis of a large public dataset for discovery with experimental validation using a single-institution cohort, we demonstrate that *KRAS* mutations predominantly occur in low-grade, early-stage EECs, co-associate with *ARID1A* mutations and *MLH1* promoter methylation and are characterised by LINE-1 hypermethylation and favourable prognosis.

Consistent with prior studies, *KRAS* mutations were mainly found in low-grade and early-stage tumours, supporting their role in the initiation of well-differentiated endometrioid carcinogenesis [6, 35]. This pattern distinguishes *KRAS*-mutated EECs from *TP53*-mutated non-endometrioid carcinomas, which typically occur in high-grade and advanced stages [3]. The observed mutual exclusivity between *KRAS* and *TP53* or *CTNNB1* mutations is coherent with the idea that these driver events activate alternative molecular routes [3, 36].

Furthermore, the frequent co-mutation of *ARID1A* with *KRAS* aligns with the phenomenon that SWItch/Sucrose Non-Fermentable (SWI/SNF) complex disruption and RAS/MAPK-pathway activation often cooperate in endometrial tumorigenesis [14, 37]. *ARID1A* encodes a subunit of the SWI/SNF chromatin remodelling complex and its loss has been associated with increased genomic accessibility and enhanced antitumour immune signalling [38, 39]. Interestingly, despite its role as a tumour suppressor, several studies have reported that *ARID1A*-mutated EECs tend to display more favourable outcomes compared with wildtype tumours [40-42]. Evidence shows that *ARID1A* support *KRAS*-driven transcription in CRC [43], yet in other cellular contexts its loss enhances the oncogenic effects of mutant *KRAS* [44, 45], indicating that the interaction between *ARID1A* dysfunction and RAS/MAPK signalling is context-dependent.

Our findings suggest that *KRAS* activation is associated with specific alterations in the epigenetic landscape of EEC. Notably, the *KRAS* mutation is significantly associated with *MLH1* promoter methylation and MSI in the *in-silico* data, a relationship also supported by a strong correlative trend in our VA EEC cohort. The observed OR of 2.8 in *in-silico* cohort and OR of 2.2 in VA cohort reveals that *KRAS*-mutated tumours exhibit two- to three-fold increased likelihood of showing sporadic dMMR due to *MLH1* methylation.

This observation is central: we hypothesize that *KRAS* may contribute to the emergence of a hypermethylator phenotype in EEC, mirroring the pattern of oncogenic signalling driving methylation seen in CRC, where the *BRAF* mutation is strongly associated with the CpG Island

Methylator Phenotype (CIMP) [46, 47]. This potential mechanistic link could imply that *KRAS* signalling, possibly in conjunction with *ARID1A* loss, facilitates the silencing of *MLH1* by hypermethylation [48, 49].

The dominance of $KRAS^{G12D}$ and $KRAS^{G12V}$ mutations observed in both cohorts, which typically result in strong and sustained MAPK pathway activation [50, 51], lends plausibility to a persistent oncogenic signal driving chronic epigenetic changes. This dependency on persistent MAPK signalling has made *KRAS* a prospective therapeutic target and the ongoing development of *KRAS* inhibitors (against $KRAS^{G12C}$, $KRAS^{G12D}$ and $KRAS^{G12V}$) demonstrates the clinical feasibility of suppressing *KRAS* signalling and downstream proliferative pathways [8, 52-54].

This finding stands in contrast to the mutual exclusivity of *KRAS* and MSI typically seen in CRC but parallels with the well-established association between $BRAF^{V600E}$ mutations and the CIMP in CRC sporadic MSI [55], highlighting the tissue-specific interplay between oncogenic signalling and epigenetic regulation.

This *KRAS* associated hypermethylator phenotype extends beyond focal gene silencing to global DNA methylation. We performed LINE-1 methylation analysis, a surrogate of global methylation levels. Our data revealed the novel finding that *KRAS*-mutated EECs exhibit significantly higher LINE-1 methylation levels compared to *KRAS*-wildtype tumours. This contrasts with the classical model of global hypomethylation linked to *KRAS* signalling in cancer progression [56-58]. Instead, our results suggest that *KRAS*-mutated endometrial tumours maintain a relatively epigenetically stable phenotype by hypermethylating repetitive elements, which may be supported by *ERK*-dependent chromatin regulation. This distinctive hypermethylated profile further defines the *KRAS*-mutated EEC as a unique molecular entity.

Clinically, *KRAS*-mutated EECs exhibited a better prognosis compared to wildtype cases, consistent with their lower-grade, early-stage and relatively stable methylation profile [59]. Studies indicate that *KRAS* mutations are associated with poorer survival in CRC [60, 61], however data regarding survival in EC remain inconsistent. Ito *et al.* found that *KRAS* mutations had significant poor survival in older patients [62]; another study reported that though *KRAS* amplification/overexpression was associated with poor outcome, *KRAS* mutations had no significant association with survival [63]; while a more recent work showed that *KRAS*-mutated ECs are genotypically distinct and tend to have longer mean overall survival [16]. These heterogeneous outcomes emphasize the integration of co-occurring genomic and epigenetic features when assessing patient prognosis.

The coexistence of *KRAS* mutation, MSI and LINE-1 hypermethylation may identify a subgroup with intermediate mutational burden and a partially maintained epigenetic control. From a therapeutic perspective, the presence of MSI along with *KRAS* activation might suggest potential responsiveness to immune checkpoint inhibitors (ICIs), while the maintenance of methylation raises the possibility of epigenetic plasticity influencing immune evasion and treatment sensitivity.

A major strength of this study lies in its integrative design, combining a large public cohort with an independent institutional dataset. The use of comprehensive profiling simultaneously assessing mutational status, MSI status, *MLH1* promoter methylation and global LINE-1 methylation within the VA cohort provides a robust framework to investigate interplay between genetics and epigenetics in EEC. Our study also presents a novel exploration of *KRAS-ARID1A* co-mutated tumours in relation to MSI and methylation patterns, providing biologically relevant insights with potential therapeutic implications. The consistency of trends across the two datasets supports the strength and reproducibility of these findings.

Nonetheless, the study was limited by the modest size of institutional cohort and the number of *KRAS*-mutated cases, reducing the statistical power to detect subtle associations. Additionally, survival analyses were constrained by few events due to early-stage disease, restricting prognostic interpretations. LINE-1 methylation, while informative of global patterns, is not genome-wide and may introduce bias. Lastly, the single-centre nature of institutional highlights the need for larger, multi-centre studies to further validate and expand our results.

In summary, our results indicate that *KRAS*-mutated EECs represent a molecularly and epigenetically distinct subset characterised by *ARID1A* co-mutation, co-occurrence of MSI/*MLH1* methylation and LINE-1 hypermethylation. These features collectively associate with favourable clinicopathological features and prognosis. Integrating *KRAS* mutation status and global methylation assessment into current molecular classification may improve risk stratification and reveal mechanistic insights into endometrial tumorigenesis.

**References**

1. WHO: *WHO Classification of Tumours: Female Genital Tumours.* 5 edn: International Agency for Research on Cancer; 2020.

2. Yen TT, Wang TL, Fader AN, Shih IM, Gaillard S: **Molecular Classification and Emerging Targeted Therapy in Endometrial Cancer.** *Int J Gynecol Pathol* 2020, **39:**26-35.

3. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, et al: **Integrated genomic characterization of endometrial carcinoma.** *Nature* 2013, **497:**67-73.

4. León-Castillo A, de Boer SM, Powell ME, Mileshkin LR, Mackay HJ, Leary A, Nijman HW, Singh N, Pollock PM, Bessette P, et al: **Molecular Classification of the PORTEC-3 Trial for High-Risk Endometrial Cancer: Impact on Prognosis and Benefit From Adjuvant Therapy.** *J Clin Oncol* 2020, **38:**3388-3397.

5. Son J, Zhang Y, Lin H, Mirallas O, Alvarez Ballesteros P, Nardo M, Clark N, Hillman RT, Campbell E, Holla V, et al: **Clinical and Genomic Landscape of RAS Mutations in Gynecologic Cancers.** *Clin Cancer Res* 2024, **30:**2986-2995.

6. Sideris M, Emin EI, Abdullah Z, Hanrahan J, Stefatou KM, Sevas V, Emin E, Hollingworth T, Odejinmi F, Papagrigoriadis S, et al: **The Role of KRAS in Endometrial Cancer: A Mini-Review.** *Anticancer Res* 2019, **39:**533-539.

7. Maru Y, Tanaka N, Tatsumi Y, Nakamura Y, Itami M, Hippo Y: **Kras activation in endometrial organoids drives cellular transformation and epithelial-mesenchymal transition.** *Oncogenesis* 2021, **10:**46.

8. Parikh K, Banna G, Liu SV, Friedlaender A, Desai A, Subbiah V, Addeo A: **Drugging KRAS: current perspectives and state-of-art review.** *J Hematol Oncol* 2022, **15:**152.

9. Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, Kolodner R: **Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines.** *Cancer Res* 1997, **57:**808-811.

10. Chalitchagorn K, Shuangshoti S, Hourpai N, Kongruttanachok N, Tangkijvanich P, Thong-ngam D, Voravud N, Sriuranpong V, Mutirangura A: **Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis.** *Oncogene* 2004, **23:**8841-8846.

11. Pavicic W, Joensuu EI, Nieminen T, Peltomäki P: **LINE-1 hypomethylation in familial and sporadic cancer.** *J Mol Med (Berl)* 2012, **90:**827-835.

12. Sunami E, de Maat M, Vu A, Turner RR, Hoon DS: **LINE-1 hypomethylation during primary colon cancer progression.** *PLoS One* 2011, **6:**e18884.

13. Gezer U, Özgür E, Yörüker EE, Polatoglou E, Holdenrieder S, Bronkhorst A: **LINE-1 cfDNA Methylation as an Emerging Biomarker in Solid Cancers.** *Cancers (Basel)* 2024, **16**.

14. Takeda T, Banno K, Okawa R, Yanokura M, Iijima M, Irie-Kunitomi H, Nakamura K, Iida M, Adachi M, Umene K, et al: **ARID1A gene mutation in ovarian and endometrial cancers (Review).** *Oncol Rep* 2016, **35:**607-613.

15. Toumpeki C, Liberis A, Tsirkas I, Tsirka T, Kalagasidou S, Inagamova L, Anthoulaki X, Tsatsaris G, Kontomanolis EN: **The Role of ARID1A in Endometrial Cancer and the Molecular Pathways Associated With Pathogenesis and Cancer Progression.** *In Vivo* 2019, **33:**659-667.

16. Kilowski KA, Dietrich MF, Xiu J, Baca Y, Hinton A, Ahmad S, Herzog TJ, Thaker P, Holloway RW: **KRAS mutations in endometrial cancers: Possible prognostic and treatment implications.** *Gynecol Oncol* 2024, **191:**299-306.

17. Li W, Zhi W, Zou S, Qiu T, Ling Y, Shan L, Shi S, Ying J: **Distinct Clinicopathological Patterns of Mismatch Repair Status in Colorectal Cancer Stratified by KRAS Mutations.** *PLoS One* 2015, **10:**e0128202.

18. Martianov AS, Mitiushkina NV, Ershova AN, Martynenko DE, Bubnov MG, Amankwah P, Yanus GA, Aleksakhina SN, Tiurin VI, Venina AR, et al: **and MSI Status in a Large Consecutive Series of Colorectal Carcinomas.** *Int J Mol Sci* 2023, **24**.

19. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, Heuer ML, Larsson E, et al: **The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data.** *Cancer Discov* 2012, **2:**401-404.

20. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, Sun Y, Jacobsen A, Sinha R, Larsson E, et al: **Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal.** *Sci Signal* 2013, **6:**pl1.

21. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V, et al: **Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer.** *Cell* 2018, **173:**291-304.e296.

22. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, Hess J, Ma S, Chiotti KE, McLellan M, et al: **Scalable Open Science Approach for Mutation Calling of Tumor Exomes Using Multiple Genomic Pipelines.** *Cell Syst* 2018, **6:**271-281.e277.

23. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, Schumacher SE, Wang C, Hu H, Liu J, et al: **Genomic and Functional Approaches to Understanding Cancer Aneuploidy.** *Cancer Cell* 2018, **33:**676-689.e673.

24. Gao Q, Liang WW, Foltz SM, Mutharasu G, Jayasinghe RG, Cao S, Liao WW, Reynolds SM, Wyczalkowski MA, Yao L, et al: **Driver Fusions and Their Implications in the Development and Treatment of Human Cancers.** *Cell Rep* 2018, **23:**227-238.e223.

25. Ding L, Bailey MH, Porta-Pardo E, Thorsson V, Colaprico A, Bertrand D, Gibbs DL, Weerasinghe A, Huang KL, Tokheim C, et al: **Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics.** *Cell* 2018, **173:**305-320.e310.

26. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghafinia S, et al: **Oncogenic Signaling Pathways in The Cancer Genome Atlas.** *Cell* 2018, **173:**321-337.e310.

27. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV, et al: **An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics.** *Cell* 2018, **173:**400-416.e411.

28. Bhandari V, Hoey C, Liu LY, Lalonde E, Ray J, Livingstone J, Lesurf R, Shiah YJ, Vujcic T, Huang X, et al: **Molecular landmarks of tumor hypoxia across cancer types.** *Nat Genet* 2019, **51:**308-318.

29. Poore GD, Kopylova E, Zhu Q, Carpenter C, Fraraccio S, Wandro S, Kosciolek T, Janssen S, Metcalf J, Song SJ, et al: **Microbiome analyses of blood and tissues suggest cancer diagnostic approach.** *Nature* 2020, **579:**567-574.

30. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen HZ, Reeser JW, Yu L, Roychowdhury S: **Landscape of Microsatellite Instability Across 39 Cancer Types.** *JCO Precis Oncol* 2017, **2017**.

31. Lakbir S, Lahoz S, Cuatrecasas M, Camps J, Glas RA, Heringa J, Meijer GA, Abeln S, Fijneman RJA: **Tumour break load is a biologically relevant feature of genomic instability with prognostic value in colorectal cancer.** *Eur J Cancer* 2022, **177:**94-102.

32. Weigelt B, Marra A, Selenica P, Rios-Doria E, Momeni-Boroujeni A, Berger MF, Arora K, Nemirovsky D, Iasonos A, Chakravarty D, et al: **Molecular Characterization of Endometrial Carcinomas in Black and White Patients Reveals Disparate Drivers with Therapeutic Implications.** *Cancer Discov* 2023, **13:**2356-2369.

33. Stefanoli M, La Rosa S, Sahnane N, Romualdi C, Pastorino R, Marando A, Capella C, Sessa F, Furlan D: **Prognostic relevance of aberrant DNA methylation in g1 and g2 pancreatic neuroendocrine tumors.** *Neuroendocrinology* 2014, **100:**26-34.

34. Fraley C, Raftery AE: **Model-Based Clustering, Discriminant Analysis, and Density Estimation.** *Journal of the American Statistical Association* 2002, **97:**611--631.

35. Chou AJ, Bing RS, Ding DC: **Endometrial Atypical Hyperplasia and Risk of Endometrial Cancer.** *Diagnostics (Basel)* 2024, **14**.

36. León-Castillo A, Gilvazquez E, Nout R, Smit VT, McAlpine JN, McConechy M, Kommoss S, Brucker SY, Carlson JW, Epstein E, et al: **Clinicopathological and molecular characterisation of 'multiple-classifier' endometrial carcinomas.** *J Pathol* 2020, **250:**312-322.

37. Mandal J, Yu ZC, Shih IM, Wang TL: **ARID1A loss activates MAPK signaling via DUSP4 downregulation.** *J Biomed Sci* 2023, **30:**94.

38. Wilson BG, Roberts CW: **SWI/SNF nucleosome remodellers and cancer.** *Nat Rev Cancer* 2011, **11:**481-492.

39. Shen J, Ju Z, Zhao W, Wang L, Peng Y, Ge Z, Nagel ZD, Zou J, Wang C, Kapoor P, et al: **ARID1A deficiency promotes mutability and potentiates therapeutic antitumor immunity unleashed by immune checkpoint blockade.** *Nat Med* 2018, **24:**556-562.

40. Mao TL, Ardighieri L, Ayhan A, Kuo KT, Wu CH, Wang TL, Shih IM: **Loss of ARID1A expression correlates with stages of tumor progression in uterine endometrioid carcinoma.** *Am J Surg Pathol* 2013, **37:**1342-1348.

41. Samartzis EP, Gutsche K, Dedes KJ, Fink D, Stucki M, Imesch P: **Loss of ARID1A expression sensitizes cancer cells to PI3K- and AKT-inhibition.** *Oncotarget* 2014, **5:**5295-5303.

42. Kato MK, Yoshida H, Tanase Y, Uno M, Ishikawa M, Kato T: **Loss of ARID1A Expression as a Favorable Prognostic Factor in Early-Stage Grade 3 Endometrioid Endometrial Carcinoma Patients.** *Pathol Oncol Res* 2021, **27:**598550.

43. Sen M, Wang X, Hamdan FH, Rapp J, Eggert J, Kosinsky RL, Wegwitz F, Kutschat AP, Younesi FS, Gaedcke J, et al: **ARID1A facilitates KRAS signaling-regulated enhancer activity in an AP1-dependent manner in colorectal cancer cells.** *Clin Epigenetics* 2019, **11:**92.

44. Livshits G, Alonso-Curbelo D, Morris JP, Koche R, Saborowski M, Wilkinson JE, Lowe SW: **Arid1a restrains Kras-dependent changes in acinar cell identity.** *Elife* 2018, **7**.

45. Guo B, Friedland SC, Alexander W, Myers JA, Wang W, O'Dell MR, Getman M, Whitney-Miller CL, Agostini-Vulaj D, Huber AR, et al: **Arid1a mutation suppresses TGF-β signaling and induces cholangiocarcinoma.** *Cell Rep* 2022, **40:**111253.

46. Nagasaka T, Koi M, Kloor M, Gebert J, Vilkin A, Nishida N, Shin SK, Sasamoto H, Tanaka N, Matsubara N, et al: **Mutations in both KRAS and BRAF may contribute to the methylator phenotype in colon cancer.** *Gastroenterology* 2008, **134:**1950-1960, 1960.e1951.

47. Serra RW, Fang M, Park SM, Hutchinson L, Green MR: **A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype.** *Elife* 2014, **3:**e02313.

48. Levine AJ, Phipps AI, Baron JA, Buchanan DD, Ahnen DJ, Cohen SA, Lindor NM, Newcomb PA, Rosty C, Haile RW, et al: **Clinicopathologic Risk Factor Distributions for MLH1 Promoter Region Methylation in CIMP-Positive Tumors.** *Cancer Epidemiol Biomarkers Prev* 2016, **25:**68-75.

49. Cohen J, R W, BY K, C W: **A KRAS hot spot mutation correlates with MLH1 methylation in endometrial carcinomas with microsatellite instability: A potential triage tool for Lynch syndrome evaluation.** *Gynecologic Oncology* 2014, **133:**79-80.

50. Rachagani S, Senapati S, Chakraborty S, Ponnusamy MP, Kumar S, Smith LM, Jain M, Batra SK: **Activated KrasG$^{12}$D is associated with invasion and metastasis of pancreatic cancer cells through inhibition of E-cadherin.** *Br J Cancer* 2011, **104:**1038-1048.

51. Xin T, Gallini S, Wei H, Gonzalez DG, Matte-Martone C, Machida H, Fujiwara H, Pasolli HA, Suozzi KC, Gonzalez LE, et al: **Oncogenic Kras induces spatiotemporally specific tissue deformation through converting pulsatile into sustained ERK activation.** *Nat Cell Biol* 2024, **26:**859-867.

52. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM: **K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.** *Nature* 2013, **503:**548-551.

53. Sacher AG, Miller WH, Patel MR, Paz-Ares L, Santoro A, Ahn MJ, Dziadziuszko R, Freres P, Luo J, Bowyer S, et al: **Single-Agent Divarasib in Patients With.** *J Clin Oncol* 2025, **43:**3249-3253.

54. D'Alessio-Sands L, Gaynier J, Michel-Milian V, Agbowuro AA, Brackett CM: **Current Strategies and Future Dimensions in the Development of KRAS Inhibitors for Targeted Anticancer Therapy.** *Drug Dev Res* 2025, **86:**e70042.

55. Kim JH, Bae JM, Cho NY, Kang GH: **Distinct features between MLH1-methylated and unmethylated colorectal carcinomas with the CpG island methylator phenotype: implications in the serrated neoplasia pathway.** *Oncotarget* 2016, **7:**14095-14111.

56. Ugai S, Yao Q, Takashima Y, Zhong Y, Matsuda K, Kawamura H, Imamura Y, Okadome K, Mima K, Arima K, et al: **Clinicopathological, molecular, and prognostic features of colorectal carcinomas with KRAS c.34G>T (p.G12C) mutation.** *Cancer Sci* 2024, **115:**3455-3465.

57. Tew BY, Durand JK, Bryant KL, Hayes TK, Peng S, Tran NL, Gooden GC, Buckley DN, Der CJ, Baldwin AS, Salhia B: **Genome-wide DNA methylation analysis of KRAS mutant cell lines.** *Sci Rep* 2020, **10:**10149.

58. Li Q, Geng S, Luo H, Wang W, Mo YQ, Luo Q, Wang L, Song GB, Sheng JP, Xu B: **Signaling pathways involved in colorectal cancer: pathogenesis and targeted therapy.** *Signal Transduct Target Ther* 2024, **9:**266.

59. Ring KL, Yates MS, Schmandt R, Onstad M, Zhang Q, Celestino J, Kwan SY, Lu KH: **Endometrial Cancers With Activating KRas Mutations Have Activated Estrogen Signaling and Paradoxical Response to MEK Inhibition.** *Int J Gynecol Cancer* 2017, **27:**854-862.

60. Phipps AI, Buchanan DD, Makar KW, Win AK, Baron JA, Lindor NM, Potter JD, Newcomb PA: **KRAS-mutation status in relation to colorectal cancer survival: the joint impact of correlated tumour markers.** *Br J Cancer* 2013, **108:**1757-1764.

61. Aljehani MA, Bien J, Lee JSH, Fisher GA, Lin AY: **KRAS Sequence Variation as Prognostic Marker in Patients With Young- vs Late-Onset Colorectal Cancer.** *JAMA Netw Open* 2023, **6:**e2345801.

62. Ito K, Watanabe K, Nasim S, Sasano H, Sato S, Yajima A, Silverberg SG, Garrett CT: **K-ras point mutations in endometrial carcinoma: effect on outcome is dependent on age of patient.** *Gynecol Oncol* 1996, **63:**238-246.

63. Birkeland E, Wik E, Mjøs S, Hoivik EA, Trovik J, Werner HM, Kusonmano K, Petersen K, Raeder MB, Holst F, et al: **KRAS gene amplification and overexpression but not mutation associates with aggressive and metastatic endometrial cancer.** *Br J Cancer* 2012, **107:**1997-2004.

## MANUSCRIPT 2

## A data-driven NGS approach for MSI assessment: improving diagnostic performance in colorectal and endometrial cancers

### Abstract

Microsatellite instability (MSI) is a key biomarker for Lynch syndrome screening and immunotherapy response prediction in colorectal (CRC) and endometrial carcinoma (EC). Conventional MSI testing based on PCR-based fragment analysis of Bethesda panel and immunohistochemistry (IHC) of DNA mismatch repair (MMR) proteins remains limited by resolution and assay scope. Next-Generation Sequencing (NGS) provides a scalable alternative; however, the accuracy of MSI calls depends on marker selection and analytical thresholds. In this perspective, we aimed to optimise and validate an NGS-based panel for robust MSI detection across CRC and EC.

Myriapod® NGS Cancer Probe PLUS panel targeting 50 genes and 122 microsatellite loci was evaluated in 65 CRC and 66 EC cases with known MMR and MSI status by IHC and PCR-based capillary electrophoresis (CE), respectively. MSI scores derived from indel variability across loci were analysed by unsupervised hierarchical clustering, principal component analysis (PCA) and model-based clustering. Performance metrics were compared against gold standard assays by ROC curve analysis to assess diagnostic concordance and classification accuracy. In addition, mutation rates were compared to investigate the biological context in MSI and microsatellite stable (MSS) tumours.

Data-driven threshold refinement and feature reduction identified a 51-loci subset with optimal discriminatory power and improved sensitivity for MSI/dMMR tumours, particularly in EC. The refined panel achieved high concordance with IHC/CE showing 100% and >86% overlap for CRC and EC, respectively, confirming its analytical robustness and cross-tumour applicability. High mutation frequencies of *BRAF* (47%) and *PIK3CA* (47%) in CRC; and *KRAS* (32%) and *PIK3CA* (56%) in EC, showed expected enrichment in MSI tumors. Discordant cases in EC displayed borderline instability and mutation rates, reflecting intermediate or mixed MSI phenotypes.

In conclusion, NGS-based panel offers both precise and scalable MSI detection and can resolve biologically discordant cases through integration of mutational landscape. This work demonstrates and supports the value of model-driven optimisation of NGS for cross-tumour MSI assessment and clinical applicability as well as gives insights on refinement of automated MSI-calling

pipelines.

**Keywords:** MSI, NGS, colorectal cancer, endometrial cancer, model-based clustering, diagnostic validation.

## Introduction

Microsatellite instability (MSI) is characterised by variation in the number of short repetitive sequences and serves as a critical biomarker for Lynch syndrome screening, tumour prognosis and predicting response to immunotherapy. MSI resulting from defects in DNA mismatch repair (MMR) protein system can arise due to germline or somatic aberrations. While MMR protein test is routinely performed by immunohistochemistry, MSI status is traditionally determined using PCR-based fragment analysis of mononucleotide repeats (MNRs) using capillary electrophoresis (CE) or RT-PCR based assays [1]. Despite exhibiting high concordance, traditional methods present limitations in sensitivity and scope, notably in tumours with equivocal immunohistochemistry (IHC) results or low MSI burden and, in any case, required additional tumour slides and the application of a second test. For these reasons, although these methods are still considered a gold standard, Next-Generation Sequencing (NGS) approach is creating new opportunities for MSI evaluation because it may offer the ability to analyse a larger number of microsatellite loci and simultaneously detect clinically relevant mutations specific to different molecular profiles [2].

Despite its benefits, NGS approach for MSI detection has not yet been integrated into standard diagnostic practice because of several critical points that must be carefully managed to ensure accurate and reliable results. Firstly, various algorithms have been developed for NGS-based MSI detection [2-9] but each algorithm uses a different strategy to quantify MSI detection (i.e., comparing repeat lengths to a reference genome or a panel of normal samples). The choice of algorithm can significantly affect the result and a consensus on a standardised approach is still ongoing. Moreover, the presence of normal DNA in tumour lesions could dilute the MSI signal, especially in samples with low tumour content. Bioinformatics algorithms must be robust enough to account for this heterogeneity and specific enough to discriminate the presence of insertion-deletion (indel) errors introduced by sequencing, which can be mistaken for genuine instability [10]. Finally, NGS-based methods can analyse hundreds or thousands of microsatellite loci, but nowadays there is no universally accepted expanded panel beyond the Bethesda one [11]. The selection of informative markers is crucial, as some loci are more prone to instability than others and their performance can vary depending on the tumour type. Nevertheless, the potential of leveraging MSI data from routine mutation testing is highly promising and desirable, driving pathologists to overcome these obstacles.

Our study investigates the impact of including additional MNRs, beyond the clinically approved markers, for the determination of MSI status in colorectal (CRC) and endometrial carcinoma (EC).

The analysis is based on data generated using a hybrid-capture NGS panel (Myriapod® NGS Cancer Probe PLUS, Diatech Pharmacogenetics) designed for the detection of genomic variants across 50 cancer-related genes. This panel evaluates 122 microsatellite loci, including the eight standard markers currently adopted in clinical practice utilised by the data analysis software for the determination of MSI status (BAT-25, BAT-26, CAT-25, MONO-27, NR-21, NR-22, NR-24 and NR-27) [12, 13].

We explored the potential contribution to MSI status detection of the 114 additional MNRs, currently excluded from standard MSI scoring, along with the eight standard markers. Thus, we developed an integrated analytical pipeline utilising unsupervised clustering, principal component analysis (PCA) and model-based clustering [14] for optimising classification cut-offs. Crucially, this optimisation was performed using a cohort of well-characterised clinical samples with confirmed MMR status, ensuring robust and reliable new MSI detection criteria. We evaluated the performance of this new MSI analysis pipeline, comparing our NGS-based MSI calls to gold standard IHC and CE-based MSI testing methods. This data-driven approach, designed to identify informative loci and establish precise thresholds, aims to facilitate a more reliable and widespread implementation of NGS-based MSI assessment in routine molecular pathology.

## Materials and Methods

### Study Cohort

A cohort of 131 formalin-fixed and paraffin-embedded (FFPE) tumour samples (comprising 65 consecutive CRCs and 66 consecutive ECs) was selected from the Molecular Oncology Laboratory (Fondazione Edo ed Elvo Tempia, Biella, Italy) and the Unit of Anatomic Pathology (Varese, Italy) during 2024-2025. All patients in this cohort provided informed consent for the molecular classification tests, which were performed for clinical relevance and potential prognostic/therapeutic application. For these cases, reporting had already been performed using IHC and/or MSI testing using PCR followed by CE of five MNR markers (BAT-25, BAT-26, MONO-27, NR-21 and NR-24).

The cases were subsequently subjected to NGS analysis to assess both the mutation profile and MSI status. A comprehensive dataset was compiled for both tumour groups integrating results from MMR-IHC, MSI testing and NGS analysis (supplementary table 1).

### NGS Analysis

NGS was performed using the Myriapod® NGS Cancer Probe PLUS, which allows the genomic profiling of 50 cancer-related genes including single nucleotide variations (SNVs), indels, copy number variations (CNVs), gene fusions and 122 microsatellite loci, including eight standard MNR markers (BAT-25, BAT-26, CAT-25, MONO-27, NR-21, NR-22, NR-24 and NR-27) (supplementary table 2). Prior to library preparation, DNA quantity and integrity were assessed using the EasyPGX® qPCR system (Diatech Pharmacogenetics) to ensure sample suitability for NGS. Libraries were prepared from 50 ng (optimal range 50-100 ng) of DNA extracted from FFPE tissue using the MagnisDx NGS Prep Station (Agilent Technologies), an automated platform designed for hybrid-capture workflows. Library preparation was performed according to the manufacturer's instructions using the Myriapod® NGS Cancer Probe PLUS kit. Briefly, the process included enzymatic DNA fragmentation, ligation of unique pair of sequencing adapters having indexed barcodes to each fragment of a sample, target amplification via multiplex PCR and purification of libraries using magnetic beads. Prepared libraries were quantified using the Qubit™ dsDNA high sensitivity (HS) Assay Kit and the Qubit™ 4 Fluorometer. Equimolar amounts of each library were pooled, to ensure homogenous coverage of each sample and then sequenced on the Illumina MiSeq or NextSeq 550dx platform. The raw reads generated as FASTQ files were processed and analysed using the Myriapod® NGS Data Analysis Software version 5.0.11 (Diatech Pharmacogenetics). During the processing of raw reads, proprietary variant and MSI-calling pipelines aligned the reads of each sample to human reference genome (hg19) to identify

SNVs/indels and quantified a score for each one of the 122 microsatellite loci. The proprietary MSI caller is based on the score of the 8 standard MNR microsatellites BAT-25, BAT-26, CAT-25, MONO-27, NR-21, NR-22, NR-24 and NR-27. The specific algorithm used to generate the MSI scores and the cutoff thresholds for each locus are considered a proprietary right of the company and are not disclosed. At the end of the analysis, a graphical interface is generated for each sample with the information of sample metrics, variants and MSI status. Specifically, for MSI status, there were three possible results: MSI-high (MSI-H) if two or more standard loci were scored as "unstable"; MSI-low (MSI-L) if one locus was "unstable" and microsatellite stable (MSS) if no locus was unstable.

**Statistical analyses for validating a new NGS pipeline for MSI detection**

Firstly, we compared the performance of NGS-based MSI detection against established reference methods on the cohort of samples tested. Concordance ratio was determined by receiver-operating characteristic (ROC) curve analysis in Stata 17 to calculate the sensitivity, specificity, false-positive ratio (FPR) and area under the curve (AUC). Discordant cases (false positive, FP; false negative, FN) were subjected to in-depth review, including re-examination of IHC slides, electropherograms generated from fragment analysis and evaluation of NGS coverage metrics.

Unsupervised hierarchical clustering was performed by using built-in *heatmap* function of RStudio to identify a subset of informative loci in individual cohorts contributing to the determination of MSI-status. After confirmation, a two-tailed unpaired *t-test* was performed to create boxplots in GraphPad Prism 8 on MSI and MSS cases concordant with reference methods using the MSI-scores of selected loci to identify the statistically significant ($p < 0.05$) loci differentiating MSI from MSS group. To confirm the contribution of identified loci, PCA was performed using Stata 17 to compare the graphical distribution of cases by 122 loci and informative loci.

After selection of informative loci, model-based clustering (Gaussian mixture model, GMM), using *mclust* library in RStudio, was conducted on concordant MSI and MSS cases to estimate the cut-off thresholds for each of the informative locus. Threshold values obtained for CRC and EC were compared to check if the identified loci have robust values for both subsets. The derived thresholds were then applied to the entire cohorts of CRC and EC to re-assess the MSI-status. To designate the MSI-status after revision, we defined the criteria using identified standard and additional informative loci collectively and labelled the cases as MSI and MSS. Subsequently, ROC curve analysis was performed between data obtained by the new determined model and MMR-IHC/molecular testing using Stata17.

## Results

### NGS-based MSI testing in CRCs and ECs: a comparative analysis with standard methods

Our cohort comprised 131 FFPE tumour samples, with 65 CRCs and 66 ECs cases. In CRC samples, 15 were classified as deficient in MMR (dMMR) and 50 as proficient in MMR (pMMR) by gold standard methods (IHC/MSI test). Among ECs, dMMR and pMMR tumours were 29 and 37, respectively (supplementary table 1).

All the cases were subjected to NGS analysis to assess both mutation profiles and MSI status. For MSI status determination, the Myriapod® NGS cancer probe PLUS's proprietary MSI-calling pipeline utilised the 8 conventional microsatellite loci, disregarding the scores from the additional 114 microsatellite loci. Using these criteria, the NGS analysis generated on samples cohort yielded the following results: for the CRC samples, 12 of 15 dMMR cases were correctly classified as MSI, while 3 were classified as MSS. Of the 50 pMMR cases, 48 were correctly identified as MSS, with 2 being classified as MSI. The ROC curve analysis showed a sensitivity of 80% and specificity of 96% for detecting dMMR in CRCs, with an AUC of 0.88 (2 FP and 3 FN). In the EC cohort, the NGS panel identified 15 of 29 dMMR cases as MSI and 14 as MSS. Among the 37 pMMR cases, 36 were confirmed as MSS, with a single case being MSI. For ECs, sensitivity was lower at 52%, though specificity remained high at 97%, with an AUC of 0.75 (1 FP and 14 FN, supplementary table 1).

### Defining common informative loci for MSI detection in CRCs and ECs

With the scope of defining our new NGS MSI caller strategy, we firstly aimed to select a restricted panel of loci, among the 122 microsatellite loci included in the Myriapod® NGS cancer probe PLUS panel analysis, able to demonstrate good concordance with results by gold standard methods on our sample cohort in both CRCs and ECs. Firstly, we performed unsupervised hierarchical clustering using the built-in *heatmap* function of RStudio, which applied Euclidean distance and complete linkage to MSI scores across 122 loci. With heatmap we obtained 6 clusters in CRC and 7 clusters in EC (supplementary figure 1). The loci with differential instability scores clustered together and allowed us to select loci strongly associated with MSI status (cluster 1, 2, 3, 4 and 6 in CRC; cluster 1, 2, 3, 4, 5 and 7 in EC) from loci with minimal discriminatory power (cluster 5 in CRC and cluster 6 in EC). Through this analysis, we identified 59 loci in CRCs and 61 loci in ECs in selected clusters that most effectively distinguished MSI from MSS cases (supplementary figure 1). To further refine the selection of informative loci, a two-tailed *t-test* conducted on concordant cases (excluding the FP and FN cases described earlier) identified 51 common loci significantly associated with MSI status in both tumour types ($p < 0.05$). These included the 8

standard MNR markers (BAT-25, BAT-26, CAT-25, MONO-27, NR-21, NR-22, NR-24 and NR-27), as well as 43 loci from 114 additional loci (table 1). The results of *t-test* performed for complete set of 122 loci are available in supplementary table 3 and respective box plots for CRC and EC are presented in supplementary figure 2 and 3, respectively. To validate the discriminative power of this subset, PCA was performed separately for CRCs and ECs, comparing this subset against the full set of 122 microsatellite loci. PCA confirmed that the 51 loci panel primarily contributed for the majority of variance associated with MSI status and effectively separated MSI and MSS cases (supplementary figure 4).

**Table 4: List of 51 informative microsatellite loci for MSI Detection across CRCs and ECs.** The list includes statistically significant loci ($p<0.05$) identified through two-tailed *t-test* using MSI scores each locus. The *t-test* was performed separately on concordant MSI and MSS samples of CRC (n=60) and EC (n=51).

| Serial | Locus Name | Chromosome | Start Position (hg19) | Reference Repeat Motif and Length | Gene Association | Locus Category | *P*-value (CRC) | *P*-value (EC) |
|---|---|---|---|---|---|---|---|---|
| 1 | MONO-27 | Chr 2 | 39573063 | (T)27 | MAP4K3 | Standard | <0.0001 | <0.0001 |
| 2 | BAT-26 | Chr 2 | 47641560 | (T)26 | MSH2 | Standard | <0.0001 | <0.0001 |
| 3 | NR-24 | Chr 2 | 95849362 | (T)24 | ZNF2 | Standard | <0.0001 | <0.0001 |
| 4 | BAT-25 | Chr 4 | 55598212 | (T)25 | KIT | Standard | 0.0001 | 0.0001 |
| 5 | CAT-25 | Chr 7 | 143003343 | (T)25 | CASP2 | Standard | <0.0001 | <0.0001 |
| 6 | NR-27 | Chr 11 | 102193509 | (T)27 | BIRC3 | Standard | <0.0001 | <0.0001 |
| 7 | NR-22 | Chr 11 | 125490766 | (T)22 | STT3A | Standard | <0.0001 | <0.0001 |
| 8 | NR-21 | Chr 14 | 23652347 | (T)21 | SLC7A8 | Standard | <0.0001 | 0.013 |
| 9 | MSI-002 | Chr 1 | 16572709 | (T)14 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 10 | MSI-006 | Chr 1 | 33402335 | (A)17 | RNF19B | Non-standard | <0.0001 | <0.0001 |
| 11 | MSI-009 | Chr 1 | 145002892 | (A)19 | PDE4DIP | Non-standard | <0.0001 | <0.0001 |
| 12 | MSI-010 | Chr 1 | 145260856 | (T)16 | NOTCH2NLR | Non-standard | <0.0001 | <0.0001 |
| 13 | MSI-011 | Chr 1 | 149900986 | (A)16 | MTMR11 | Non-standard | <0.0001 | <0.0001 |
| 14 | MSI-012 | Chr 1 | 159290623 | (T)14 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 15 | MSI-013 | Chr 1 | 165793509 | (T)15 | TMCO1 | Non-standard | <0.0001 | <0.0001 |
| 16 | MSI-014 | Chr 1 | 170733810 | (T)22 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 17 | MSI-015 | Chr 1 | 180075942 | (T)14 | CEP350 | Non-standard | <0.0001 | <0.0001 |
| 18 | MSI-016 | Chr 1 | 201754411 | (T)17 | NAV1 | Non-standard | <0.0001 | <0.0001 |
| 19 | MSI-017 | Chr 1 | 204084143 | (T)15 | SOX13 | Non-standard | <0.0001 | <0.0001 |
| 20 | MSI-018 | Chr 1 | 220318426 | (A)15 | IARS2 | Non-standard | <0.0001 | <0.0001 |
| 21 | MSI-019 | Chr 1 | 227618722 | (A)15 | LINC01641 | Non-standard | <0.0001 | 0.0014 |
| 22 | MSI-020 | Chr 1 | 235486356 | (T)17 | ARID4B | Non-standard | <0.0001 | <0.0001 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 23 | MSI-021 | Chr 2 | 39564894 | (T)28 | MAP4K3 | Non-standard | <0.0001 | 0.0007 |
| 24 | MSI-023 | Chr 2 | 62063094 | (A)17 | FAM161A | Non-standard | <0.0001 | <0.0001 |
| 25 | MSI-024 | Chr 2 | 97309469 | (T)16 | FER1L5 | Non-standard | <0.0001 | <0.0001 |
| 26 | MSI-038 | Chr 4 | 39501723 | (A)18 | UGDH | Non-standard | <0.0001 | <0.0001 |
| 27 | MSI-043 | Chr 5 | 172421761 | (T)15 | ATP6V0E1 | Non-standard | <0.0001 | <0.0001 |
| 28 | MSI-045 | Chr 6 | 43021977 | (G)12 | MRPL2 | Non-standard | <0.0001 | <0.0001 |
| 29 | MSI-047 | Chr 6 | 142691951 | (T)17 | ADGRG6 | Non-standard | <0.0001 | <0.0001 |
| 30 | MSI-050 | Chr 8 | 17665798 | (T)15 | MTUS1-DT | Non-standard | <0.0001 | <0.0001 |
| 31 | MSI-058 | Chr 10 | 12381073 | (A)16 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 32 | MSI-059 | Chr 10 | 12424938 | (T)15 | CAMK1D | Non-standard | <0.0001 | <0.0001 |
| 33 | MSI-062 | Chr 10 | 79635929 | (A)14 | DLG5 | Non-standard | <0.0001 | <0.0001 |
| 34 | MSI-066 | Chr 11 | 62703438 | (T)15 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 35 | MSI-077 | Chr 12 | 85285921 | (A)17 | SLC6A15 | Non-standard | <0.0001 | <0.0001 |
| 36 | MSI-080 | Chr 13 | 31722621 | (A)17 | HSPH1 | Non-standard | <0.0001 | <0.0001 |
| 37 | MSI-085 | Chr 14 | 73959704 | (T)16 | RIOX1 | Non-standard | <0.0001 | <0.0001 |
| 38 | MSI-086 | Chr 14 | 74054709 | (T)15 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 39 | MSI-092 | Chr 16 | 11051908 | (A)16 | CLEC16A | Non-standard | <0.0001 | <0.0001 |
| 40 | MSI-093 | Chr 16 | 22650153 | (A)16 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 41 | MSI-094 | Chr 17 | 13981240 | (T)21 | COX10 | Non-standard | <0.0001 | <0.0001 |
| 42 | MSI-095 | Chr 17 | 19314918 | (T)18 | RNF112 | Non-standard | <0.0001 | <0.0001 |
| 43 | MSI-096 | Chr 17 | 27081793 | (T)16 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 44 | MSI-097 | Chr 17 | 52991732 | (A)16 | TOM1L1 | Non-standard | <0.0001 | <0.0001 |
| 45 | MSI-100 | Chr 18 | 511981 | (T)14 | LINC01925 | Non-standard | <0.0001 | <0.0001 |
| 46 | MSI-101 | Chr 18 | 649880 | (T)15 | CLUL1 | Non-standard | <0.0001 | <0.0001 |
| 47 | MSI-104 | Chr 19 | 14104689 | (T)14 | RFX1 | Non-standard | <0.0001 | <0.0001 |
| 48 | MSI-110 | Chr 22 | 26352179 | (T)16 | MYO18B | Non-standard | <0.0001 | <0.0001 |
| 49 | MSI-111 | Chr 22 | 29696469 | (T)16 | EWSR1 | Non-standard | <0.0001 | <0.0001 |
| 50 | MSI-112 | Chr 22 | 37401595 | (T)15 | CIMIP4 | Non-standard | <0.0001 | <0.0001 |
| 51 | MSI-114 | Chr X | 101409255 | (T)16 | BEX5 | Non-standard | <0.0001 | <0.0001 |

**Establishing robust MSI classification criteria using Model-Based Thresholds**

A model-based clustering algorithm was applied to the MSI scores of each of 51 loci to define locus-specific cut-off values. For each locus, the score value segregating the computed clusters was determined independently in both CRC and EC subsets. However, in some loci the cluster boundaries did not align well with the observed distribution of case scores, particularly in the

presence of borderline values that did not optimally discriminate MSI and MSS groups. To address this, we complemented the clustering output with visual inspection of score distributions across all cases and adjusted the threshold when necessary to maximise between-group separation while minimising within-group variance. These values were compared between the two cohorts to establish a consensus threshold to be used as cut-off for classification of instability. An example of determination of threshold value for BAT-26 is shown in figure 1.

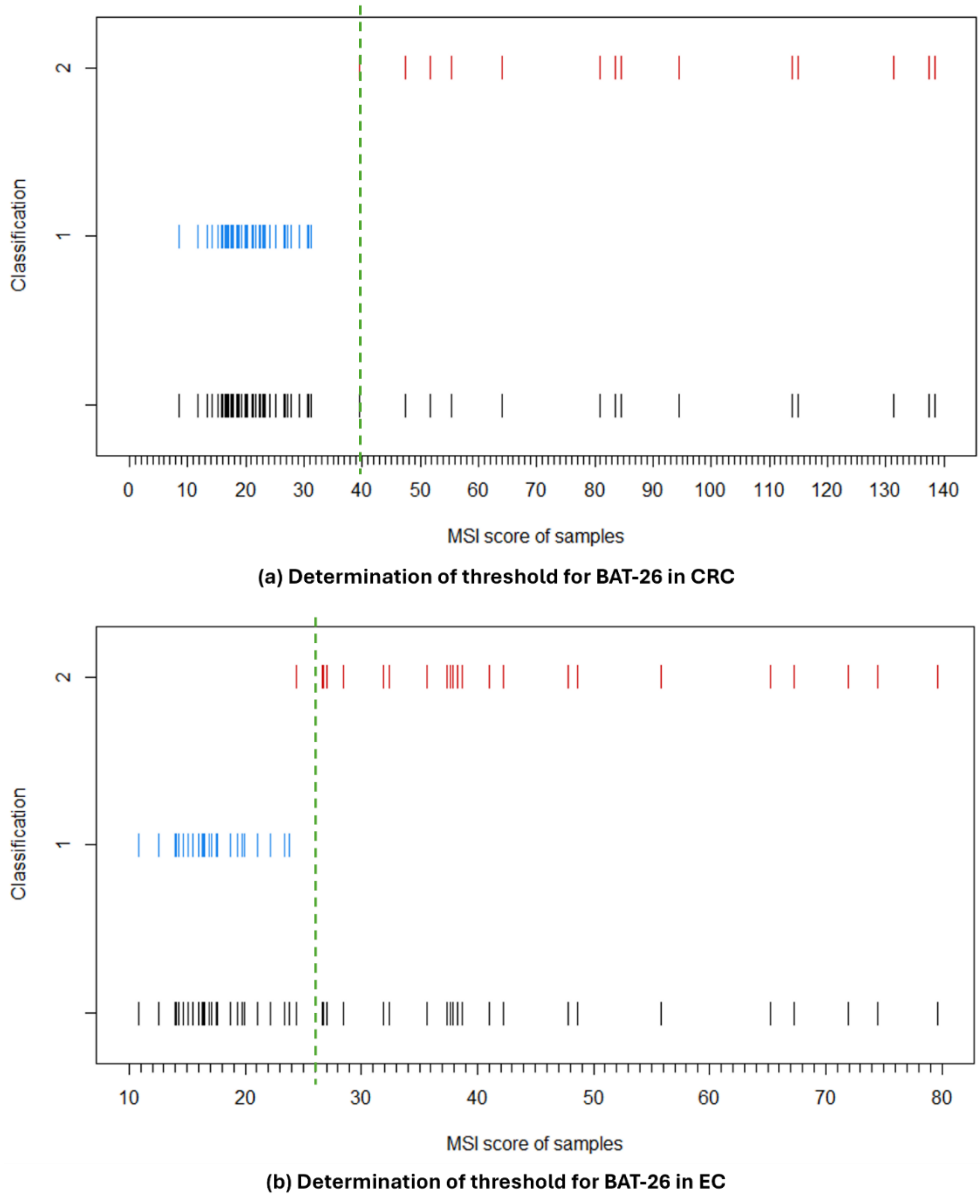



**Figure 9: Identification of threshold value for BAT-26.** Model-based clustering algorithm was applied independently to MSI scores of concordant cases across (a) CRC (n=60) and (b) EC (n=51) to determine the cut-off value for each locus in both cohorts. The x-axis represents individual samples ranked by MSI score, while the y-axis indicates the clustering-derived classification groups. Cluster 1 (y=1, blue) corresponds to MSS cases, while Cluster 2 (y=2, red) comprises samples with higher scores, aligned with MSI status. The dotted green line denotes the optimised threshold established to distinguish MSI from MSS cases as (a) 40 for CRC and (b) 26 for EC.

Notably, the threshold values derived for both CRCs and ECs cohorts were found to be identical for all 51 informative loci, with the sole minor exception of BAT-26 (figure 2). This remarkable consistency is particularly significant given the known challenges in accurately defining MSI

status in ECs, suggesting that these identified loci and their associated thresholds are robust and broadly applicable across these different tumour types. These locus-specific thresholds were then applied to the entire CRCs and ECs cohorts to revise and refine MSI classification based on NGS-derived scores. This approach enabled a more tailored and statistically optimised distinction between MSI and MSS cases.

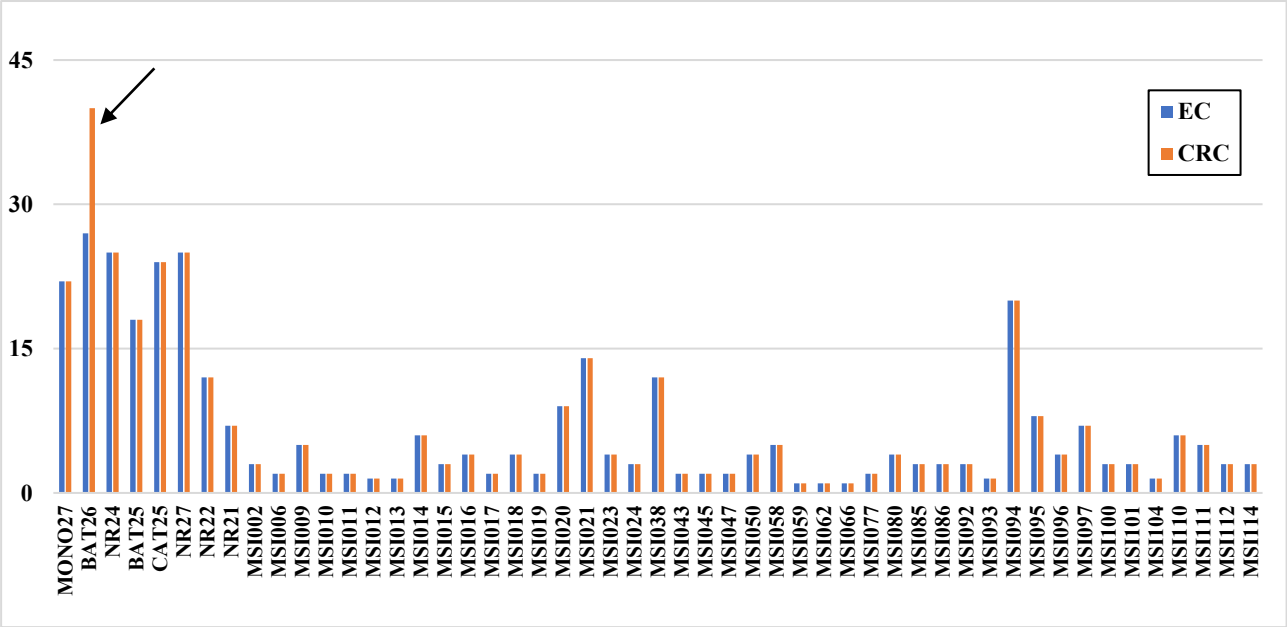



**Figure 10: Comparison of optimal cut-off scores for 51 informative microsatellite loci between ECs and CRCs.** The bar chart displays the locus-specific threshold scores derived using model-based clustering for distinguishing MSI from MSS in ECs (blue) and CRCs (orange). Each pair of bars represents the optimised cut-off value at a given microsatellite locus. For the majority of loci, the cut-off values were consistent, except for BAT-26 that showed a markedly higher cut-off score in CRC compared to EC (arrow).

Further, to improve the overall accuracy of MSI classification, we established a refined, panel-wide criterion. By employing the model-based clustering algorithm on concordant MSI and MSS groups of CRC and EC cohorts (supplementary table 1), we estimated the minimum number of unstable loci required for MSI classification. This analysis suggested a threshold of >5 loci for CRC and >6 loci for EC to classify a case as MSI (figure 3). For consistency and optimised performance across tumour types, we adopted the threshold of >5 loci for MSI obtained from CRC group. Instead of conventional three-tier classification (MSI-H, MSI-L and MSS), we adopted a clinically useful binary classification approach, segregating MSI-H as MSI and grouping MSI-L together with MSS. Therefore, a sample was classified as MSI if instability (defined by refined locus-specific thresholds) was observed in more than five (>5) of the 51 selected loci and as MSS if less than or equal to five (≤5) loci were found unstable. This comprehensive criterion incorporated both the 8 standard MNR markers and the 43 additional informative loci identified through our clustering and statistical analysis.

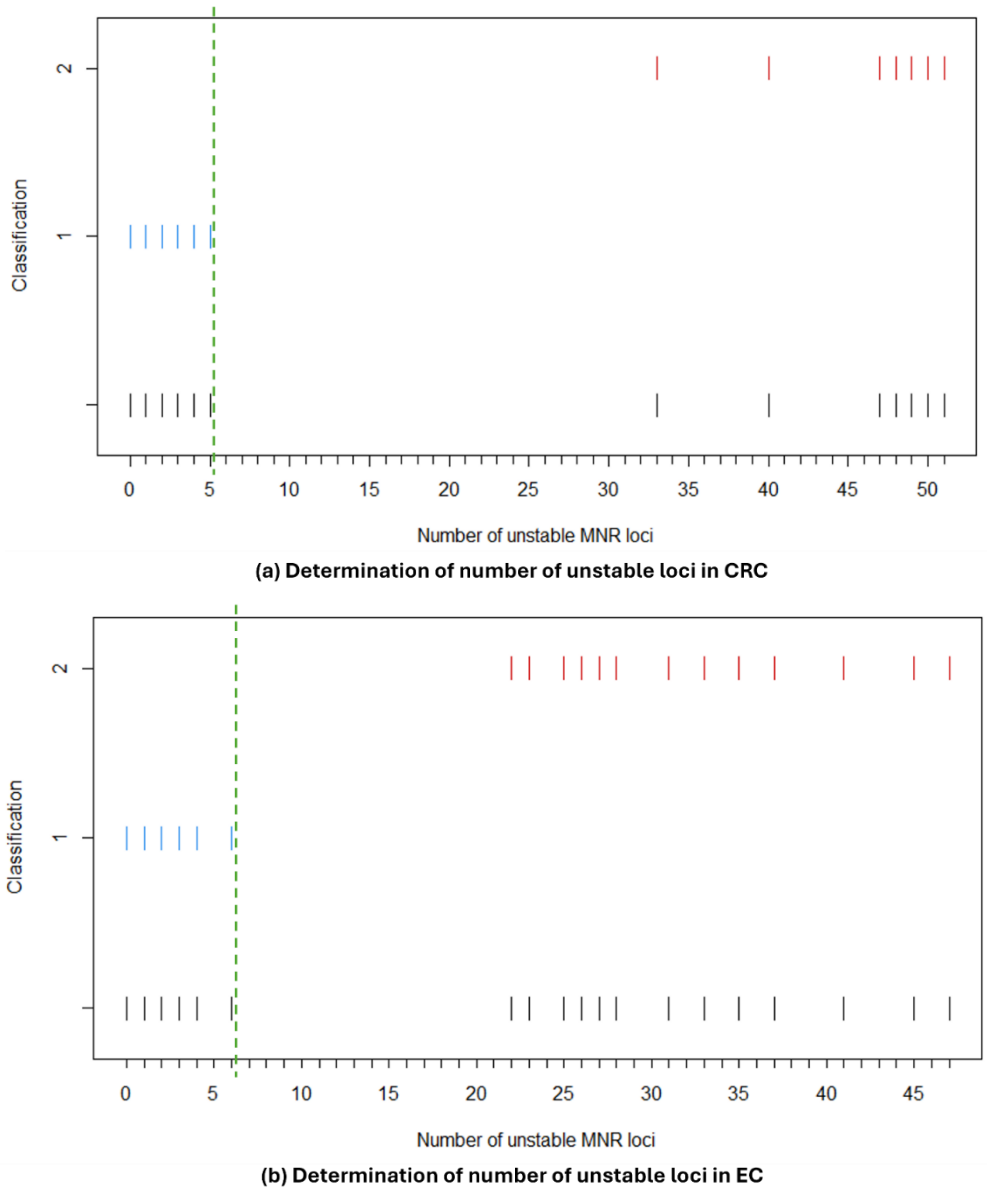
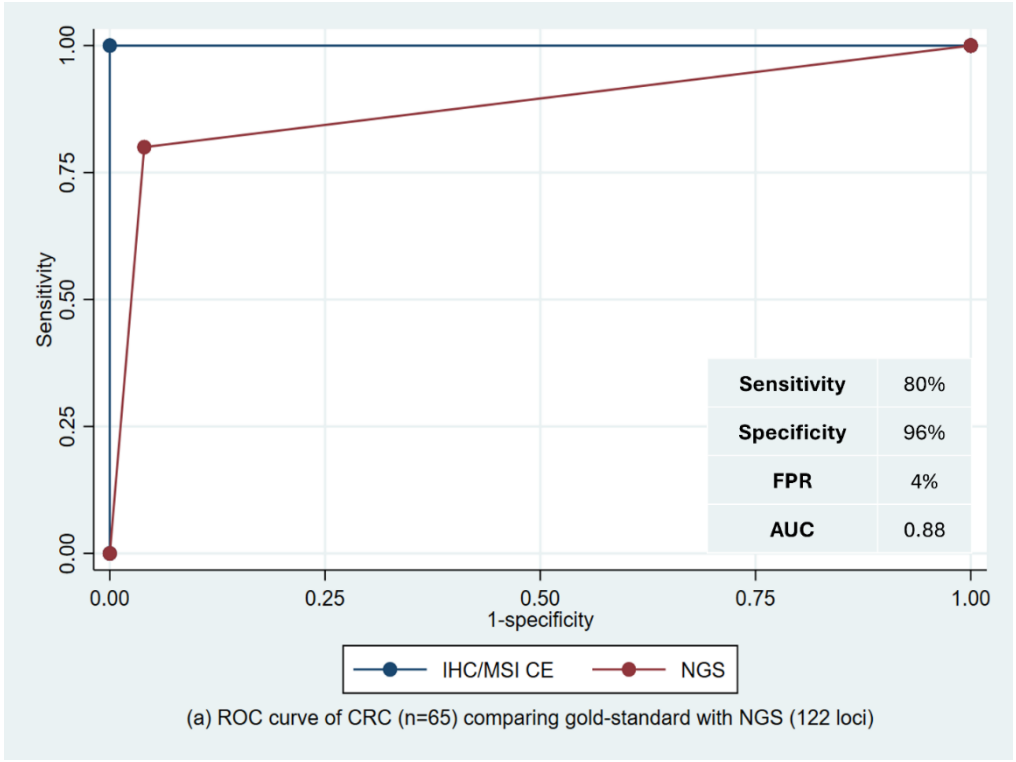
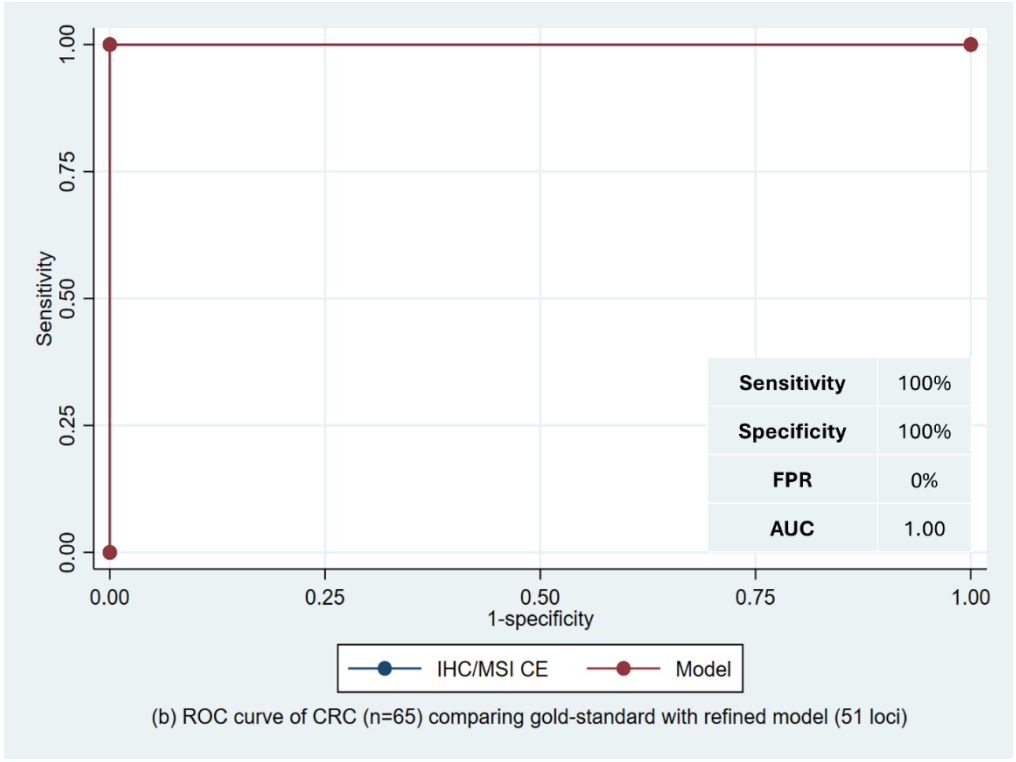



**Figure 11: Estimation of number of unstable loci to distinguish MSI and MSS groups.** Model-based clustering was applied independently to concordant cases of (a) CRC (n=60) and (b) EC (n=51) to estimate threshold for number of unstable loci that separate MSI from MSS groups. Each vertical bar represents the number of unstable loci observed, bars are condensed as multiple cases shared identical counts. The vertical dotted green line denotes the minimum number of unstable loci required to classify a case as MSI, identified as (a) >5 loci for CRC and (b) >6 loci for EC. For consistency across tumour types and optimised classification performance, the threshold of >5 loci derived from CRC was adopted.

With this revised classification model, we re-assessed our cohorts. Among 65 CRCs, we obtained 15 MSI and 50 MSS cases, showing 100% concordance compared with gold standard methods. For 66 ECs, we obtained 26 MSI and 40 MSS cases, with an 86.4% concordance (57 cases). Comparing MSI calls based on our revised classification model with gold standard methods (IHC/MSI CE), we observed sensitivity and specificity of 100% and AUC 1.00 for CRC samples and a 79% sensitivity, 92% specificity, FPR 8% and AUC 0.86 for EC samples (figure 4).




(a) ROC curve of CRC (n=65) comparing gold-standard with NGS (122 loci)




(b) ROC curve of CRC (n=65) comparing gold-standard with refined model (51 loci)

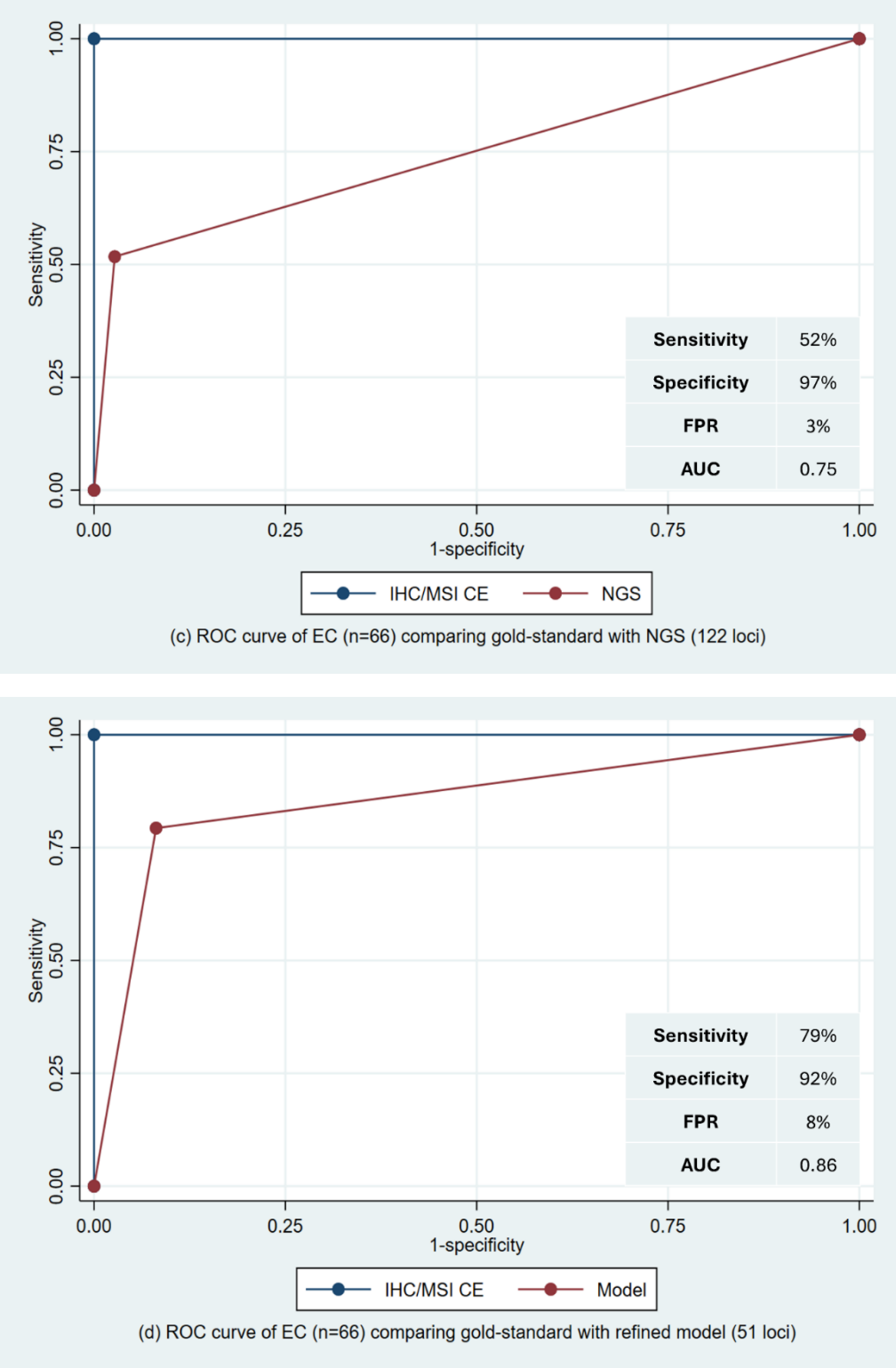



**Figure 12: ROC curve analysis of MSI evaluation.** Comparison of results from gold standard methods (IHC/MSI CE) with proprietary NGS panel of 122 loci (Myriapod® NGS Data Analysis Software v.5.0.11) and refined model of 51 loci. (a) ROC curve for CRCs (n=65) using the proprietary pipeline and (b) using the refined model; (c) ROC curve for ECs (n=66) using the proprietary pipeline and (d) using the refined model. Improved sensitivity in both tumour groups reflect the discriminative power of the data-driven approach in classification of MSI and MSS cases.

In CRCs, the MSI group consistently showed a significantly higher number of unstable loci (mean ± SD = 42.1±12.6) compared to the MSS group (mean ± SD = 0.9±1.3, $p<0.0001$) (figure 5a). Similarly, in ECs, MSI tumours exhibited significantly more unstable loci (mean ± SD =

27.5±12.6) than MSS tumours (mean ± SD = 0.9±1.0, p<0.0001) (figure 5b).

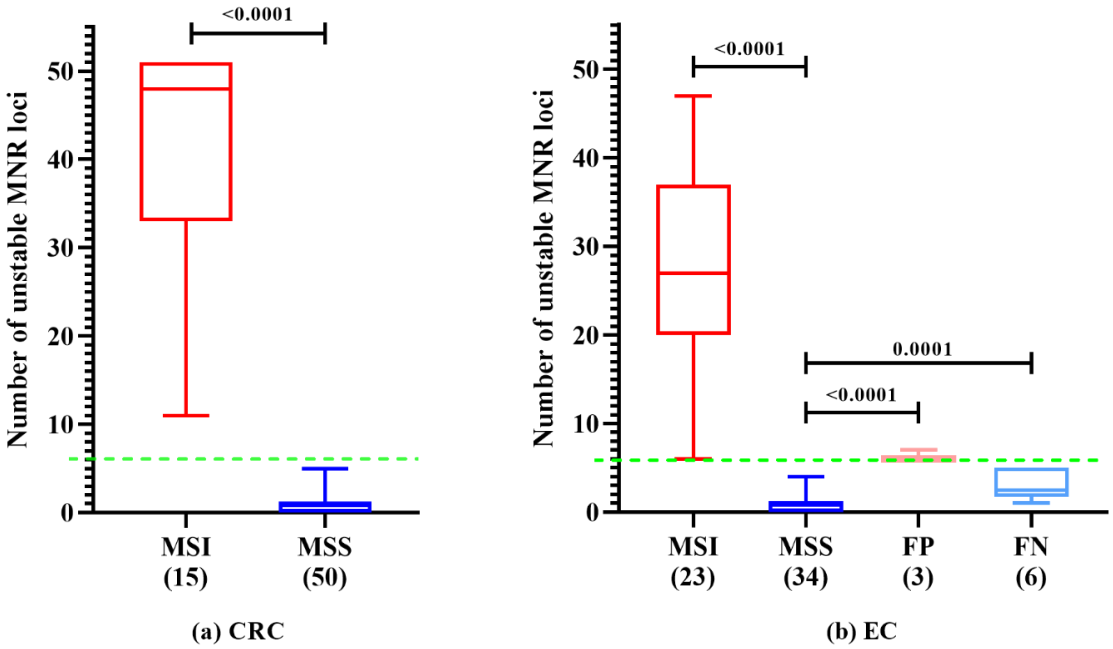



**Figure 13: Quantitative comparison of instability across CRCs and ECs.** Number of unstable mononucleotide repeat (MNR) loci in (a) CRCs (n=65) and (b) ECs (n=66) assessed by model-based clustering. The dotted green line (at y=6) represents the estimated cut-off of unstable loci for evaluating a case as either MSS (≤5) or MSI (>6) established from concordant cases of CRC cohort.

### Positivity rate of mutations in MSI and MSS groups based on the new model

To further validate our refined MSI classification and investigate its clinical implications, we assessed the rates of pathogenic/likely pathogenic mutations in tumours categorised as MSI or MSS by our novel method. We also thoroughly examined discordant cases, comparing their mutational profiles and unstable loci counts to gain deeper insights into potential discrepancies with gold standard methods, including IHC and fragment analysis by CE. The details of the mutations along with the MSI status and number of unstable loci determined using model in CRCs and ECs cohorts are given in supplementary table 4.

CRC classified as MSI by our model showed a significantly higher mean number of mutations (mean ± SD = 3.3 ± 1.6) compared to MSS tumours (mean ± SD = 1.7 ± 0.9, p < 0.0001) (figure 6a). In line with the molecular classification of CRCs [15], analysis of pathogenic gene mutations in CRC revealed distinct mutational profiles: *BRAF, PIK3CA* and *PTEN* mutations were significantly enriched in MSI group, while *TP53* mutations showed significant predominance in MSS tumours (figure 6b).

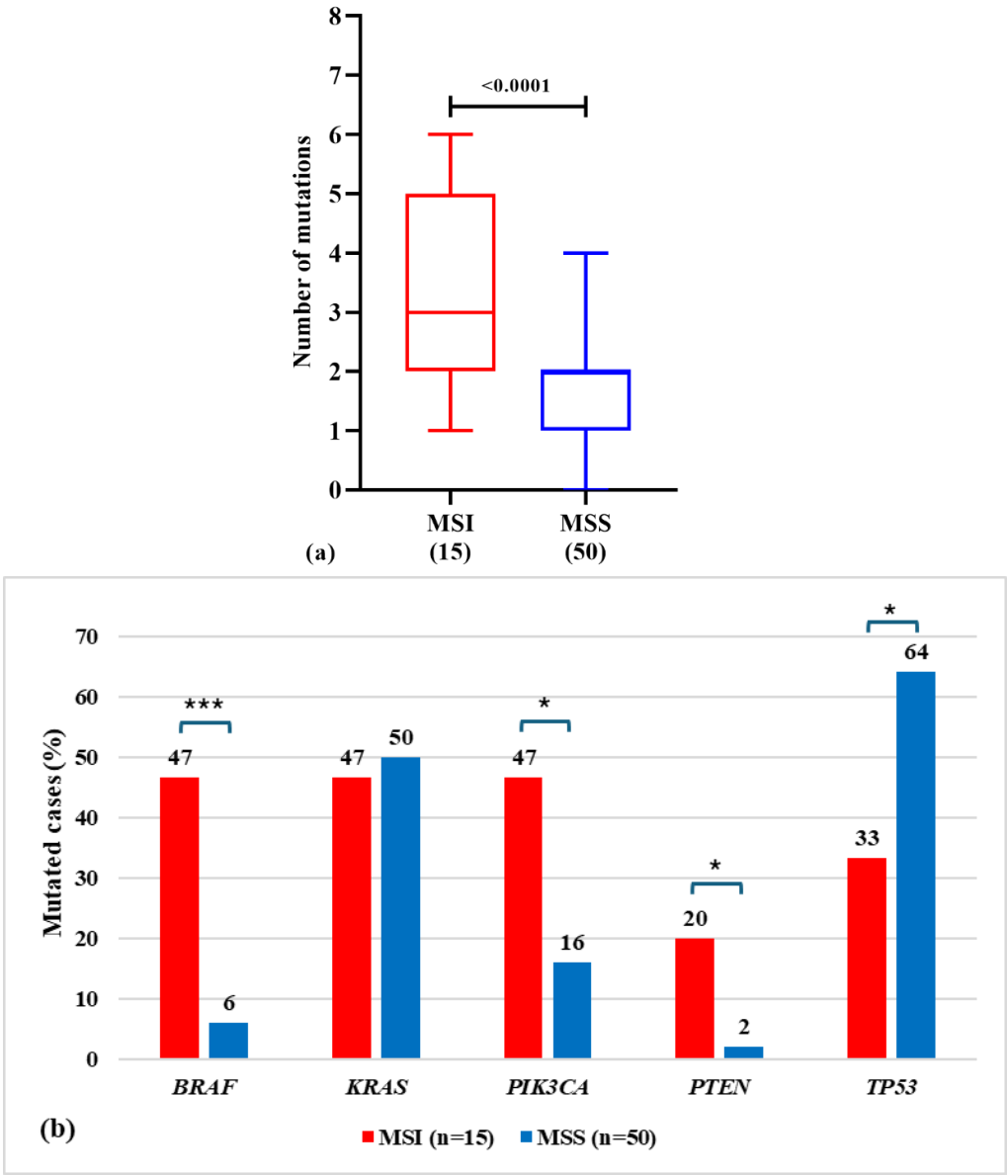



**Figure 14: Distribution of pathogenic mutations in CRCs (n=65).** (a) Number of pathogenic mutations in CRCs showing significant difference between MSI and MSS groups. P-value was calculated using a two-tailed *t-test* comparing MSI with MSS. (b) Distinct mutational pattern was observed between MSI and MSS groups, with *BRAF*, *PIK3CA* and *PTEN* variants significantly enriched in MSI tumours and *TP53* variants significant in MSS tumours. *KRAS* was frequently and equally mutated in both groups. Two-tailed chi-square analysis was performed to calculate the p-values (* $p<0.05$; *** $p<0.001$).

As reported in the literature [16], the mutation rate in EC was high in both MSI and *POLE*-mutated cases (figure 7a). Consequently, the number of mutations is not a reliable parameter to differentiate MSI from MSS status in EC, given the confounding hypermutation caused by *POLE* mutations. Five cases (4 MSS and 1 MSI) carried pathogenic *POLE* variants, while two cases (1 MSS and 1 MSI) had variants of uncertain significance (VUS).

Based on the mutation profiles, MSI EC cases showed a significantly higher frequency of *KRAS* mutations. Specifically, *KRAS* mutations were present in 32% of MSI tumours while in 12% of

MSS tumours. In contrast, although not significant, *TP53* mutations were slightly more frequent in MSS EC tumours (26% in MSS versus 12% in MSI group), as shown in figure 7b.

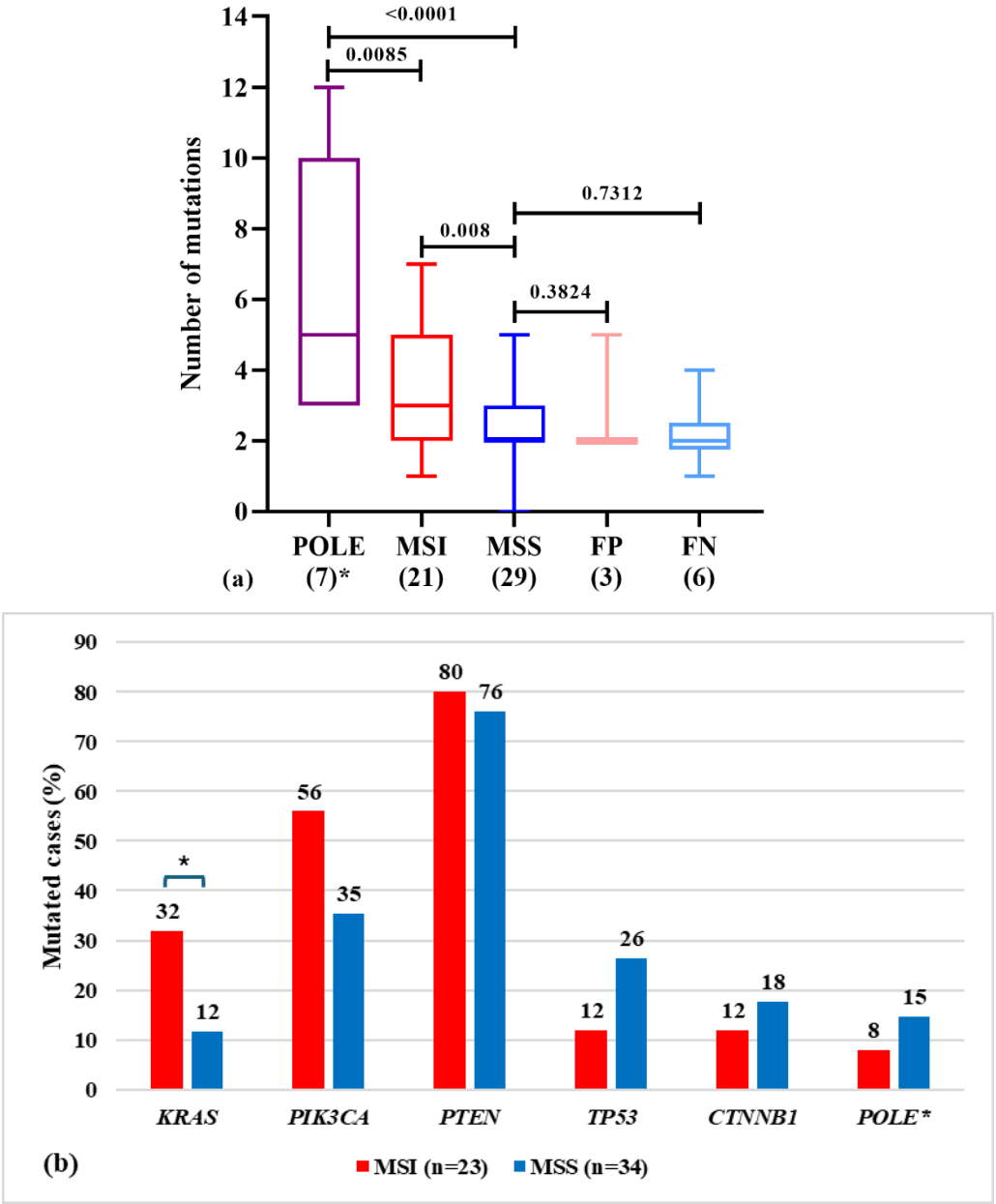



**Figure 15: Distribution of pathogenic mutations across commonly altered genes in ECs.** (a) Number of pathogenic mutations in ECs showing significant difference between MSI, MSS and discordant cases. *POLE* ultra-mutated tumours *(* includes five cases with pathogenic mutations and two with VUS)* are shown separately with elevated mutation counts irrespective of MSI classification. P-values were calculated using a two-tailed *t-test* by comparing each group with MSS taken as reference. For *POLE* group, p-value was compared with both MSI and MSS groups. (b) MSI tumours show significant variants in *KRAS* gene, while both MSI and MSS tumours had frequent *PIK3CA* and *PTEN* mutations. Two-tailed chi-square analysis was performed to calculate the p-value (* $p<0.05$).

For EC discordant cases, the three FP cases (MSI by our model, pMMR by IHC) showed borderline instability (mean ± SD = 6.3±0.6) showing significance compared to MSS group ($p<0.0001$) (figure 5b). Conversely, the six FN cases (dMMR by IHC and MSI test, but MSS by our model) exhibited intermediate instability (mean ± SD = 3.0±1.7) showing significance with MSS cases ($p=0.0001$) (figure 5b). Upon re-evaluation of electropherograms for these FN cases,

minimal peak shifts were observed which were challenging to be assessed (supplementary figure 5). These minimal drifts in peaks suggest that these cases might involve subtle, low-level or regional MSI events that fall below the detection thresholds of the NGS method (table 2). The findings also highlight the challenges in classifying MSI status in EC even with gold standard methods and favour the application of a comprehensive approach, utilising multiple techniques and focusing the nuances of genetic profiles, to accurately identify these specific tumour subtypes.

**Table 5: Molecular and genomic characteristics of discordant cases.** The table summarises the molecular features of discordant cases of ECs (n=9), grouped based on the type of discordance. Reference classification from IHC or CE-based MSI is compared with model-based prediction (based on the number of unstable loci among the 51 informative loci panel; #/51). Pathogenic gene mutations are annotated with their genomic and protein changes, along with variant allele frequencies (AF %).

| Type | Case ID | MMR-IHC / CE-MSI Status | Model Prediction | Positive Loci (#/51) | Mutated Genes | Mutations | AF (%) |
|---|---|---|---|---|---|---|---|
| **False positive (n=3)** | E21 | NA / MSS | MSI | 6 | *KRAS*<br>*PIK3CA* | c.38G>A, p.(Gly13Asp)<br>c.311C>T, p.(Pro104Leu) | 8.30<br>7.10 |
| | E49 | pMMR / MSS | MSI | 7 | *CTNNB1*<br>*PIK3CA*<br>*PIK3CA*<br>*PIK3CA*<br>*PTEN* | c.110C>T, p.(Ser37Phe)<br>c.277C>T, p.(Arg93Trp)<br>c.3061T>C, p.(Tyr1021His)<br>c.3145G>A, p.(Gly1049Ser)<br>c.585del, p.(His196ThrfsTer3) | 9.30<br>10.70<br>10.40<br>7.60<br>23.60 |
| | E59 | pMMR / MSS | MSI | 6 | *PTEN*<br>*PTEN* | c.697C>T, p.(Arg233Ter)<br>c.741_742insAA, p.(Pro248AsnfsTer9) | 13.90<br>14.10 |
| **False negative (n=6)** | E7 | MLH1- PMS2- / MSI-H | MSS | 2 | *PTEN*<br>*RB1* | c.741dup, p.(Pro248ThrfsTer5)<br>c.62dup, p.(Ala22GlyfsTer9) | 13.40<br>8.60 |
| | E10 | MLH1- PMS2- / MSI-H | MSS | 1 | *PTEN* | c.388C>G, p.(Arg130Gly) | 15.60 |
| | E24 | MLH1- PMS2- / MSI-H | MSS | 3 | *KRAS*<br>*PTEN* | c.35G>T, p.(Gly12Val)<br>c.955dup, p.(Thr319AsnfsTer6) | 50.30<br>26.30 |
| | E42 | MLH1- PMS2- / MSI-H | MSS | 5 | *KRAS*<br>*PIK3CA*<br>*PTEN*<br>*PTEN* | c.35G>T, p.(Gly12Val)<br>c.1634A>C, p.(Glu545Ala)<br>c.388C>G, p.(Arg130Gly)<br>c.968del, p.(Asn323MetfsTer21) | 20.70<br>19.10<br>27.00<br>25.40 |
| | E50 | MLH1- PMS2- / MSI-H | MSS | 2 | *PTEN*<br>*PTEN* | c.31A>T, p.(Arg11Ter)<br>c.1011_1023del, p.(Phe337LeufsTer3) | 24.70<br>12.20 |
| | E53 | NA / MSI-H | MSS | 5 | *KRAS*<br>*PIK3CA* | c.35G>A, p.(Gly12Asp)<br>c.3140A>G, p.(His1047Arg) | 9.50<br>13.00 |

**Discussion**

In this study, we aimed to optimise and validate a novel robust NGS-based approach for MSI detection, exploiting the widespread availability of NGS technology in many laboratories. Our data-driven methodology, leveraging an extended panel of 122 microsatellite loci and advanced statistical analyses, successfully identified a refined set of 51 informative markers and established optimised classification thresholds. Considering the analysed sample cohort, this approach showed a significant concordance with results obtained from gold standard methods, demonstrating the important positive impact of including additional MNRs, beyond the clinically approved markers, for the determination of MSI status by NGS in CRC and EC, saving time and tumour slides.

To further validate our refined MSI classification and investigate its clinical implications, we assessed the rates of pathogenic/likely pathogenic mutations in tumours categorised as MSI or MSS by our novel method. The discordant cases, in particular, proved to be peculiar and insidious, often exhibiting a borderline MSI phenotype which complicates their precise classification even with gold standard methodologies. Globally, our findings underscore the intrinsic challenges in accurately identifying these tumour subtypes and favour the application of a comprehensive methodological approach. We also thoroughly examined these discordant cases, comparing their mutational profiles and unstable loci counts to gain deeper insights into potential discrepancies with gold standard methods, including IHC and fragment analysis by CE.

Currently, despite various NGS algorithms for MSI detection exist, including mSINGS [2], MSIsensor [3], MSIseq [4], MANTIS [5], mSILICO [6], MSIdetect [7], MSIPeak [8], MSIDRL [9], the lack of consensus on a standardised panel of markers and the variability across different quantification strategies remain a significant challenge [17]. Our integrated pipeline allowed us to identify 51 highly informative microsatellite loci that effectively distinguish MSI from MSS cases. Crucially, this refined panel includes not only the eight conventional markers but also 43 novel loci, which altogether contribute to an enhanced ability to detect instability. The remarkable consistency of the derived locus-specific thresholds across both CRC and EC cohorts, with only a minor exception for BAT-26, underscores the robustness and broad applicability of our optimised criteria across different tumour types. This consistency is particularly noteworthy given the known challenges in accurately defining MSI status in EC and represents a significant step towards the standardisation of NGS-based MSI assessment [18-22]. The clinical utility of our NGS-based MSI classification is evidenced by the high concordance and performance in comparison with results obtained using gold standard methods on the sample cohort. For CRCs, we achieved high diagnostic accuracy with substantial reduction of FP and FN cases compared to the 8 conventional

MNR microsatellite loci classification. Even more critically, for ECs, where MSI detection can be particularly challenging, our model demonstrated a marked improvement in sensitivity and overall performance. This accuracy is crucial for both Lynch syndrome screening and guiding immunotherapy decisions, especially in cases where traditional IHC or CE-based fragment analysis yield equivocal or low-burden MSI results. The ability to analyse a larger number of markers concurrently also addresses the limitations of traditional methods in detecting subtle or regional instability [23, 24].

Our in-depth analysis of discordant cases further elucidated the complexities of MSI assessment and highlighted the unique insights provided by our NGS integrated pipeline. The three FP EC cases, despite being classified as MSI by the model, showed borderline instability and mutational profiles similar to MSS tumours. This suggests that some tumours may exhibit a degree of instability without the corresponding hypermutation typically seen in MSI tumours, also reported in a case study by Bielska et al. [25]. Conversely, the six FN EC cases demonstrated an intermediate level of instability and few mutations, even though they were classified as MSI by standard methods. Globally, these data suggest that only a comprehensive approach, integrating multiple techniques and considering the genetic complexity of tumours, is necessary to accurately identify these challenging cases. Beyond MSI classification, the simultaneous assessment of mutational profiling provided by the Myriapod® NGS cancer probe PLUS panel is extremely valuable, serving a dual purpose: it offers biological validation of the MSI status and enables a comprehensive molecular characterisation for clinical decision-making.

The observed differences in mutation rates and the involvement of specific gene (e.g., enrichment of *BRAF*, *PIK3CA*, *PTEN* mutations in MSI CRC [26] and *KRAS* mutations in MSI EC [27] serve to biologically validate our MSI classification by corroborating established oncogenic associations.

Furthermore, the mutational data is crucial for patient stratification. The detection of *POLE* mutations in a subset of EC cases, characterised by an ultra-mutated phenotype independent of MSI status, demonstrates the panel's capability to identify distinct molecular subtypes that might otherwise be missed. This integrated approach not only improves MSI detection but also provides critical information for patient stratification and therapeutic guidance, particularly in the context of precision oncology.

Despite the significant advancements, our study has certain limitations. Although the cohort of 131 FFPE samples is substantial, larger cohorts from multiple institutions would allow to address

different pre-analytical preparations, thereby further strengthening the generalisability of our findings. Additionally, further studies are needed to prospectively evaluate the clinical utility and impact of our optimised pipeline on patient management and outcomes, especially for the discordant cases where the clinical relevance of the detected instability levels warrants further investigation.

In conclusion, our data-driven NGS approach represents a significant step forward in MSI assessment for CRCs and ECs. By identifying a most comprehensive set of informative microsatellite loci and establishing robust, standardised classification thresholds, we have demonstrated a clear positive impact in MSI detection accuracy with a high concordance with results obtained using gold standard methods. Our approach demonstrates also the positive contribution to diagnostic performance using NGS panels able to provide data from a more comprehensive set of microsatellite loci combining this with SNV detection. This integrated molecular profiling approach not only enhances the accuracy of MSI detection but also provides crucial insights into the broader mutational landscape, facilitating more precise molecular diagnoses and guiding tailored therapeutic strategies in cancer patients.

**References**

1. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, Miller R, Riaz N, Douillard JY, Andre F, Scarpa A: **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.** *Ann Oncol* 2019, **30:**1232-1243.

2. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC: **Microsatellite instability detection by next generation sequencing.** *Clin Chem* 2014, **60:**1192-1199.

3. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L: **MSIsensor: microsatellite instability detection using paired tumor-normal sequence data.** *Bioinformatics* 2014, **30:**1015-1016.

4. Huang MN, McPherson JR, Cutcutache I, Teh BT, Tan P, Rozen SG: **MSIseq: Software for Assessing Microsatellite Instability from Catalogs of Somatic Mutations.** *Sci Rep* 2015, **5:**13321.

5. Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, Roychowdhury S: **Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS.** *Oncotarget* 2017, **8:**7452-7463.

6. Lee Y, Lee JA, Park HE, Han H, Kim Y, Bae JM, Kim JH, Cho NY, Kim HP, Kim TY, Kang GH: **Targeted next-generation sequencing-based detection of microsatellite instability in colorectal carcinomas.** *PLoS One* 2021, **16:**e0246356.

7. Marques AC, Ferraro-Peyret C, Michaud F, Song L, Smith E, Fabre G, Willig A, Wong MML, Xing X, Chong C, et al: **Improved NGS-based detection of microsatellite instability using tumor-only data.** *Front Oncol* 2022, **12:**969238.

8. Zhou B, Wang Y, Ding L, Tian X, Sun W, Zhang W, Liu YH: **A novel algorithm for the detection of microsatellite instability in endometrial cancer using next-generation sequencing data.** *Oncol Lett* 2025, **29:**86.

9. Shang Y, Yang L, Hu P, Zhao S, Xu J, Zhao Y, Ren X, Zhang D, He Q, Liu X: **Applying next-generation sequencing to detect microsatellite instability in pan-cancer patients: a retrospective study of 35,563 Chinese cases.** *NPJ Precis Oncol* 2025, **9:**303.

10. Cheng J, He J, Wang S, Zhao Z, Yan H, Guan Q, Li J, Guo Z, Ao L: **Biased Influences of Low Tumor Purity on Mutation Detection in Cancer.** *Front Mol Biosci* 2020, **7:**533196.

11. Umar A, Boland CR, Terdiman JP, Syngal S, de la Chapelle A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, et al: **Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability.** *J Natl Cancer Inst* 2004, **96:**261-268.

12. Murphy KM, Zhang S, Geiger T, Hafez MJ, Bacher J, Berg KD, Eshleman JR: **Comparison of the microsatellite instability analysis system and the Bethesda panel for the determination of microsatellite instability in colorectal cancers.** *J Mol Diagn* 2006, **8:**305-311.

13. Bianchi F, Galizia E, Catalani R, Belvederesi L, Ferretti C, Corradini F, Cellerino R: **CAT25 is a mononucleotide marker to identify HNPCC patients.** *J Mol Diagn* 2009, **11:**248-252.

14. Fraley C, Raftery AE: **Model-Based Clustering, Discriminant Analysis, and Density Estimation.** *Journal of the American Statistical Association* 2002, **97:**611--631.

15. Network CGA: **Comprehensive molecular characterization of human colon and rectal cancer.** *Nature* 2012, **487:**330-337.

16. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, et al: **Integrated genomic characterization of endometrial carcinoma.** *Nature* 2013, **497:**67-73.

17. Škerl P, Vogrič V, Stegel V, Šetrajčič Dragoš V, Blatnik O, Klančar G, Novaković S: **Real-World Evaluation of Microsatellite Instability Detection via Targeted NGS Panels in Routine Molecular Diagnostics.** *Int J Mol Sci* 2025, **26**.

18. Kanopiene D, Vidugiriene J, Valuckas KP, Smailyte G, Uleckiene S, Bacher J: **Endometrial cancer and microsatellite instability status.** *Open Med (Wars)* 2015, **10:**70-76.

19. Goodfellow PJ, Billingsley CC, Lankes HA, Ali S, Cohn DE, Broaddus RJ, Ramirez N, Pritchard CC, Hampel H, Chassen AS, et al: **Combined Microsatellite Instability, MLH1 Methylation Analysis, and Immunohistochemistry for Lynch Syndrome Screening in Endometrial Cancers From GOG210: An NRG Oncology and Gynecologic Oncology Group Study.** *J Clin Oncol* 2015, **33:**4301-4308.

20. Stelloo E, Jansen AML, Osse EM, Nout RA, Creutzberg CL, Ruano D, Church DN, Morreau H, Smit VTHB, van Wezel T, Bosse T: **Practical guidance for mismatch repair-deficiency testing in endometrial cancer.** *Ann Oncol* 2017, **28:**96-102.

21. Raffone A, Travaglino A, Cerbone M, Gencarelli A, Mollo A, Insabato L, Zullo F: **Diagnostic Accuracy of Immunohistochemistry for Mismatch Repair Proteins as Surrogate of Microsatellite Instability Molecular Testing in Endometrial Cancer.** *Pathol Oncol Res* 2020, **26:**1417-1427.

22. Riedinger CJ, Esnakula A, Haight PJ, Suarez AA, Chen W, Gillespie J, Villacres A, Chassen A, Cohn DE, Goodfellow PJ, Cosgrove CM: **Characterization of mismatch-repair/microsatellite instability-discordant endometrial cancers.** *Cancer* 2024, **130:**385-399.

23. Wang C, Zhang L, Vakiani E, Shia J: **Detecting mismatch repair deficiency in solid neoplasms: immunohistochemistry, microsatellite instability, or both?** *Mod Pathol* 2022, **35:**1515-1528.

24. Rüschoff J, Schildhaus HU, Rüschoff JH, Jöhrens K, Bocker Edmonston T, Dietmaier W, Bläker H, Baretton G, Horst D, Dietel M, et al: **Testing for deficient mismatch repair and microsatellite instability : A focused update.** *Pathologie (Heidelb)* 2023, **44:**61-70.

25. Bielska AA, Chatila WK, Walch H, Schultz N, Stadler ZK, Shia J, Reidy-Lagunes D, Yaeger R: **Tumor Mutational Burden and Mismatch Repair Deficiency Discordance**

**as a Mechanism of Immunotherapy Resistance.** *J Natl Compr Canc Netw* 2021, **19:**130-133.

26. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, et al: **CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer.** *Nat Genet* 2006, **38:**787-793.

27. Kilowski KA, Dietrich MF, Xiu J, Baca Y, Hinton A, Ahmad S, Herzog TJ, Thaker P, Holloway RW: **KRAS mutations in endometrial cancers: Possible prognostic and treatment implications.** *Gynecol Oncol* 2024, **191:**299-306.

## GENERAL DISCUSSION

This thesis integrates two independent yet inter-related studies that together advance the understanding of molecular and diagnostic determinants of EC. Our work was focused on the interaction between genetic and epigenetic features and translating these insights into improved diagnostic and prognostic accuracy.

The first study was designed to interrogate the interplay between somatic driver mutations, global DNA methylation and MSI status, specifically focusing on its correlation with *MLH1* methylation, in EEC. We combined an *in-silico* analysis of public datasets with an institutional cohort characterised by NGS for mutation profiling, MMR-IHC, MSI evaluation by PCR-based CE, MS-MLPA for *MLH1* methylation and pyrosequencing for LINE-1. One of the main findings of our study included that the *KRAS* activating mutations are associated with MSI in EEC, in particular with an increased likelihood of *MLH1* promoter hypermethylation. A significant association was observed in the *in-silico* cohort and a consistent correlative trend was present in our VA cohort. The frequencies and patterns we observed are consistent with prior reports that *MLH1* hypermethylation underlies a large proportion of sporadic MSI-positive ECs [72, 73]. Global methylation assessment, using LINE-1, in our EEC cohort demonstrated significant hypermethylation in *KRAS*-mutated tumours compared with *KRAS*-wildtype cases (chapter 3, figure 1). Crucially, the combined observation of *MLH1* hypermethylation and high LINE-1 methylation levels associated with *KRAS* and *ARID1A* mutations represents novel findings derived from this work.

Although some studies reported the association of *KRAS* activation with genome-wide hypomethylation and chromosomal instability in aggressive cancers [74, 75], our findings instead support that *KRAS*-driven tumorigenesis in EEC is linked to a relatively hypermethylated phenotype, consistent with the reports that epigenetic remodeling in CRC is tissue-specific [76, 77]. In addition, we found the co-occurrence of *KRAS* and *ARID1A* mutations in EEC and both are reported to be linked with CIMP in CRC [24, 35], suggesting their role in DNA methylation.

The second study was focused on assessing and optimising an NGS-based MSI detection panel targeting 122 loci on CRC and EC cases. With the aim to refine a smaller subset of microsatellite loci, we found that 51 loci out of the 122 retained discriminatory performance for MSI status when combined with model-based clustering and threshold optimisation. By comparing NGS-derived MSI scores with gold standard (IHC and CE) methods, this study proposed a standardised and accurate workflow, thus, allowing robust MSI classification across the two cohorts and improving

assay sensitivity and specificity (chapter 4, figure 4).

Considering the integration of two studies in relevance to the mutation profiles and MSI testing, the use of NGS can identify the multiple molecular features in a single assay rapidly and precisely [78-80]. It can help to understand how oncogenic signalling, methylation reprogramming and genomic instability collectively shape EC pathogenesis and its clinical assessment. Hence, it could not only enhance the diagnostic accuracy but also identify the distinct molecular subsets as future targets of clinical intervention.

*KRAS* is widely characterised as an oncogenic driver that activates downstream RAS/MAPK pathway, promoting proliferation and survival across multiple tumour types [81-83]. However, growing evidence indicates that *KRAS* activation is not limited to signaling disruption, it can be associated with distinct alterations in DNA methylation landscapes, which in turn shape transcriptional activity and/or chromatin remodeling [84]. Experimental studies in *KRAS*-mutant pancreatic and lung cell lines show that *KRAS* mutations alter CpG methylation at specific loci, suggesting a contribution of *KRAS* signaling to epigenetic regulation [75]. These changes may include promoter methylation at specific genes like *MLH1* and broader alterations in the methylation patterns across the genome [75].

The potential mechanisms describing association of *KRAS* and *MLH1* methylation could involve the influence of RAS/MAPK signalling on DNA methyltransferase (DNMT) expression [75, 85], localisation of chromatin modifiers [84, 86] and metabolic disturbance (e.g., chronic oestrogen signalling, obesity-related inflammation [87, 88]) predisposing cells both to *KRAS*-activating events and increasing vulnerability to promoter hypermethylation of tumour suppressor or DNA repair (such as *MLH1*) genes.

In CRC, a complex interplay among *KRAS*, *BRAF* and *MLH1* promoter methylation exists, where *MLH1* hypermethylation and CIMP are more tightly linked to $BRAF^{V600E}$-positive serrated pathway tumours [89], while *KRAS* mutations often exhibit alternative pathways as Farchoukh *et al*. reported the link between *KRAS*-mutations and *MLH1*-hypermethylation in *BRAF*-wildtype tumours to be associated with conventional tubular adenomas [90].

In contrast, given that *BRAF* mutation is generally absent in EEC, our analysis focused specifically on *KRAS* and EEC, showing associations between hotspot *KRAS* mutations (frequently in codons 12/13) and *MLH1* promoter methylation in *in-silico* analysis (odds ratio, OR=2.8) and a supporting trend in our institutional cohort (OR=2.2). This indicates that *KRAS* activation can co-exist and potentially be related to the epigenetic inactivation of *MLH1* in at least a subset of EECs, as

reported by Cohen *et al*. [91]. The observed discrepancies in *KRAS-MLH1* relationship between CRC and EC highlight tissue-specific associations. In view of our results and the evidence that *KRAS* mutations are present in 10-30% endometrial hyperplasia and are involved in the early endometrial carcinogenesis [20, 21], it could be inferred that *KRAS* activation may play a role not only in initiating tumorigenic transformation but also in promoting early epigenetic alterations such as *MLH1* hypermethylation. Future studies should investigate whether specific driver alterations in EEC, particularly the *KRAS-MLH1* association observed here, define a tumorigenic pathway analogous to the "serrated pathway" reconstructed for intestinal polyps.

*MLH1* promoter hypermethylation is the predominant epigenetic mechanism accounting for MSI/dMMR phenotype in sporadic EC. Classical studies have consistently demonstrated that a high proportion of MSI-positive ECs harbour *MLH1* promoter methylation [92, 93], supporting that *MLH1* methylation is enriched in more than 90% MLH1-deficient cases in sporadic presentations [94]. This incidence has direct implications for diagnostics: the presence of *MLH1* promoter methylation in an *MLH1*-deficient tumour favours a sporadic etiology (epigenetic silencing) over Lynch syndrome, whereas absence of *MLH1* methylation in an MLH1-deficient tumour raises suspicion for germline *MLH1* mutation (Lynch syndrome) and indicates the need for germline testing [95]. We followed this rationale by combining MMR-IHC, MS-MLPA for *MLH1* and PCR-based MSI testing to evaluate the sporadic nature of EECs in our institutional cohort.

Not all *MLH1*-methylated MSI tumours are identical; their clinical behaviour, mutational spectrum (*BRAF* or *KRAS* mutations), immune infiltration and clinical outcomes can vary [96]. Some *MLH1*-methylated ECs have prominent immune cell infiltration and high mutation rate, consistent with neoantigen-driven immune recognition, which partly explains the high response rates of MSI tumours to PD-1/PD-L1 blockade [96, 97]. In our analysis, *MLH1*-methylated MSI EECs formed a distinct cluster, suggesting shared molecular features. The presence of *KRAS* mutations in this context, either independently or synergistically with *ARID1A* loss, may promote *MLH1*-hypermethylation, possibly through sustained MAPK-dependent chromatin reconfiguration [91, 98].

The observed significant co-occurrence of *KRAS* and *ARID1A* mutations in the cBioPortal cohort strengthens a potential interactive role of these alterations as oncogenic and epigenetic modulators in EEC. The prevalence of MSI as well as *MLH1* methylation in more than 50% of these concurrently mutated cases highlights a strong association between dual mutations and dMMR due to epigenetic deregulation. Collectively, it supports the hypothesis that *KRAS* activation and

*ARID1A* loss may cooperate not only in driving oncogenic signalling [99-101] but also in shaping the epigenetic landscape, potentially facilitating a hypermethylator phenotype and influencing tumour biology in EC [24].

LINE-1 methylation, assessed by pyrosequencing of consensus LINE-1 CpG sites, is a validated surrogate for global DNA methylation and has been used across tumour types [102]. LINE-1 hypomethylation is associated with chromosomal instability, aneuploidy and worse prognosis in several cancers; conversely, MSI tumours, which arise through a hypermethylation mechanism rather than chromosomal instability, often display relatively preserved LINE-1 methylation compared with MSS tumours [103].

The literature reports variable relationships between *KRAS* mutations and LINE-1 methylation. Tang *et al*. found no association between LINE-1 methylation and *KRAS* status [104], while others reported a link between LINE-1 hypomethylation and *KRAS* mutations [74, 75]. Meta-analyses in CRC have suggested that LINE-1 hypomethylation is more common in non-MSI tumours, supporting the model where global hypomethylation and chromosomal instability are distinct from the MSI hypermutator pathway [103]. In our EEC cohort, LINE-1 methylation presented complementary, orthogonal insights to *MLH1* promoter methylation, suggesting that *KRAS* mutations are underlying cause of locus-specific as well as global hypermethylation. However, LINE-1 is not a substitute for MSI testing, rather it complements it by highlighting the global epigenetic state and potential mechanisms of genomic instability. In translational terms, integrating LINE-1 methylation analysis into molecular classification enriches the biological narrative behind each tumour and may improve risk stratification.

The traditional diagnostic algorithm for MSI and dMMR has relied on two orthogonal approaches: 1) IHC, which is rapid, low-cost and functional assessment of MMR proteins (MLH1, MSH2, MSH6, PMS2), but it cannot distinguish germline from epigenetic causes which requires *MLH1* promoter methylation testing by techniques like MS-MLPA [105]; 2) PCR-based MSI testing with CE, which employs panels such as the Bethesda markers or mononucleotide repeat markers to assess length instability at canonical microsatellite loci and it correlates well with IHC [106, 107]. Although both conventional methods remain the gold standard in clinical practice, the growing reliance on comprehensive genomic profiling for therapeutic decision-making highlights the value of integrating MSI assessment into NGS workflows, thereby consolidating testing, preserving tissue and providing clinicians with a unified report [108].

Many computational tools infer MSI status from targeted or exome NGS data. Their approaches

differ, some compare tumour versus matched normal (MSIsensor [109]), others perform tumour-only analyses (MSIsensor-pro [110]) and some rely on panels of informative mononucleotide repeats captured in targeted sequencing (mSINGS [111], MANTIS [112], MSIdetect [70]). However, panel design, algorithm and computational thresholds substantially influence results [113].

Given the diversity of available NGS panels, a key practical approach is the selection of such a panel that includes informative, monomorphic loci that are reliably covered by the panel and are stable in MSS samples thus reducing false positives and improves sensitivity [71, 111]. Our work starting from a broad 122-loci panel and deriving a 51-loci subset driven by model-based clustering of data directly addresses this principle by empirically selecting loci with the most discriminatory power across CRC and EC cohorts.

Many algorithmic validations have focused on CRC, where MSI is well characterised [114, 115]. However, performance can vary by tumour type because the distribution of instability differs between CRC and EC [116, 117]. Therefore, cross-tumour validation is essential to ensure that a reduced set of loci and algorithmic threshold optimised in one tumour type generalises to others. Our approach on optimising thresholds with model-based clustering (*mclust*) and assessing performance separately within CRC and EC concordant cases provided confidence that the reduced and refined subset is broadly informative. This approach mirrors recent literature emphasizing calibration of NGS MSI methods in different tumour types [71].

In our comparison between gold standard and NGS-MSI, concordant MSI cases typically showed MLH1 protein loss and high mutation rate. These cases validate the integrated pipeline and confirm that NGS MSI scoring with the optimised locus set and thresholds reproduces traditional gold standard calls with high sensitivity and specificity, consistent with multiple independent studies showing high concordance between NGS-based methods and IHC/CE [70, 71, 108, 114, 115, 118, 119]. Moreover, we obtained a complete concordance for CRC cases with no discordance observed.

Discordant cases often reside at the interface between MSI and MSS categories and require careful interpretation for optimal inference [120]. In our data, we observed 9 discordant cases among EC dataset (chapter 4, table 2) with 3 FP and 6 FN cases. The FP cases showed a borderline instability while an intermediate number of unstable loci were observed in FN cases. Both the groups of discordant cases had a low mutation frequency similar to MSS tumours making it challenging to reach a conclusive interpretation. In practical terms, discordant cases emphasize the importance of

multimodal testing to integrate molecular features and to better understand the underlying genetic complexity of such tumours.

While molecular classification currently has the strongest impact on adjuvant and systemic therapy decisions, there is increasing interest in using molecular features to guide surgical planning and postoperative surveillance [15]. The therapeutic importance of MSI analysis is underscored by the approval of pembrolizumab for MSI/dMMR solid tumours that established a biomarker-driven approach [25]. Specifically for ECs, subsequent approvals and clinical trials, including dostarlimab for dMMR EC and use of pembrolizumab in the recurrence cases, expanded therapeutic options for patients [25, 121, 122]. The ability to identify MSI EC with high reliability therefore has immediate consequences for treatment selection, trial eligibility and prognosis.

On the other hand, the frequent occurrence of *KRAS* mutations in EEC and development of *KRAS* inhibitors (sotorasib, adagrasib and divarasib) may represent promising future therapeutic option [59-61]. The use of *KRAS* inhibitors, particularly in combination with ICIs for *KRAS*-mutated/MSI tumours is under investigation and could improve the efficacy in tumours resistance to *KRAS* monotherapy [62]. Integration of *KRAS*-targeted therapies into molecularly guided treatment procedures could therefore offer a precision oncology approach for a biologically defined subset of EC patients.

MSI ECs frequently display distinctive clinical behaviours, with outcomes influenced by tumour stage, patient comorbidities and the mutational landscape [96]. As observed in CRC, the co-occurrence of MSI with specific driver mutations, such as, *KRAS, NRAS* or *BRAF,* can alter the expected prognosis compared with canonical MSI tumours [123, 124]. This underscores the necessity for integrated reporting that considers both mutation profiles and MSI status. In line with this, the subset of VA EECs exhibiting concurrent *KRAS* and *ARID1A* mutations alongside MSI, as characterised in the first manuscript, suggests further investigation to determine whether these tumours possess distinct prognostic features or differential responsiveness to ICIs. The findings from the second manuscript complement this clinical perspective by optimising NGS-based MSI detection within a single workflow that simultaneously provides mutational profiling, thereby facilitating a more comprehensive molecular assessment.

From both laboratory operational and translational perspectives, multiple advantages emerge from an integrated approach. A targeted NGS panel capable of simultaneously providing somatic mutation calls (e.g., *PTEN, KRAS, PIK3CA, ARID1A, CTNNB1, TP53*) alongside MSI scoring minimises tissue consumption and reduces turnaround time. Combining mutation profiling, TMB

and MSI assessment within a single workflow delivers immediate clinical utility, offering both diagnostic and therapeutic insights. Furthermore, the integrated analysis improves sensitivity in borderline MSI cases, decreasing the likelihood of ambiguous results that would otherwise require repeated testing.

**Limitations**

With respect to the experimental and analytical workflow, several caveats can be recognised that may influence the interpretation of findings. The potential factors requiring careful consideration include inter-observer variability in IHC interpretation, potential DNA quality differences due to extraction from FFPE samples [125, 126], assay-specific variability in methylation quantification by pyrosequencing including bisulphite conversion [127], subjective interpretation of electropherograms in MSI testing, limited probe coverage in MS-MLPA assays and technical variability introduced during NGS library preparation. All of which may collectively influence the accuracy and reproducibility of the results.

While our combined analysis of two datasets (cBioPortal and institutional VA cohort) in first study provides complementary strengths, the institutional cohort size limits statistical power for some subgroup analyses like interactions among *KRAS*, *ARID1A*, LINE-1 methylation and MSI status. Larger multi-centre cohorts, ideally with prospective sample increase would reduce uncertainty and increase the statistical significance.

In second study, the refined 51-loci subset and thresholds were optimised using CRC and EC cohorts and are implemented within the proprietary algorithms of MSI-calling pipelines of Myriapod® NGS Data Analysis Software (Diatech Pharmacogenetics). Tool performance can vary with panel design, read length and platform-specific algorithms [113]. Thus, analytical validation is necessary prior to clinical implementation.

**Conclusions**

This thesis advances the understanding of how oncogenic mutations, notably *KRAS*, intersect with epigenetic alterations (focal *MLH1* promoter methylation and global LINE-1 methylation) to shape MSI phenotypes in EEC and demonstrates a logical, validated approach for consolidating MSI detection into targeted NGS workflows. The empiric finding that *KRAS* activation correlates with *MLH1* methylation and MSI is biologically plausible and consistent with a growing literature that emphasizes tissue-specific relationships between driver mutations and the methylome. While causative connection of this finding remains to be proven, the association calls for a mechanistic follow-up and suggests that co-mutational context would improve clinical interpretation and

therapeutic strategies.

## BIBLIOGRAPHY


1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F: **Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries.** *CA Cancer J Clin* 2021, **71:**209-249.

2. AIOM, AIRTUM: *I numeri del cancro in Italia - 2024.* Intermedia Editore; 2024.

3. Crosbie EJ, Kitson SJ, McAlpine JN, Mukhopadhyay A, Powell ME, Singh N: **Endometrial cancer.** *Lancet* 2022, **399:**1412-1428.

4. Lee NK, Cheung MK, Shin JY, Husain A, Teng NN, Berek JS, Kapp DS, Osann K, Chan JK: **Prognostic factors for uterine cancer in reproductive-aged women.** *Obstet Gynecol* 2007, **109:**655-662.

5. Hamilton CA, Pothuri B, Arend RC, Backes FJ, Gehrig PA, Soliman PT, Thompson JS, Urban RR, Burke WM: **Endometrial cancer: A society of gynecologic oncology evidence-based review and recommendations.** *Gynecol Oncol* 2021, **160:**817-826.

6. Bokhman JV: **Two pathogenetic types of endometrial carcinoma.** *Gynecol Oncol* 1983, **15:**10-17.

7. WHO: *WHO Classification of Tumours: Female Genital Tumours.* 5 edn: International Agency for Research on Cancer; 2020.

8. Mutter GL: **Endometrial intraepithelial neoplasia (EIN): will it bring order to chaos? The Endometrial Collaborative Group.** *Gynecol Oncol* 2000, **76:**287-290.

9. Moreno-Moreno E, Caniego-Casas T, Carretero-Barrio I, Cortés A, Muriel A, Domínguez-Rullán JA, Martín-Gromaz C, Moreno-Bueno G, Matías-Guiu X, Palacios J, Pérez-Mies B: **Histologic and Molecular Type Changes in Endometrial Cancer Recurrences in Comparison With Their Corresponding Primary Tumors.** *Am J Surg Pathol* 2024, **48:**1580-1587.

10. Berek JS, Matias-Guiu X, Creutzberg C, Fotopoulou C, Gaffney D, Kehoe S, Lindemann K, Mutch D, Concin N, Endometrial Cancer Staging Subcommittee FIGOWsCC: **FIGO staging of endometrial cancer: 2023.** *Int J Gynaecol Obstet* 2023, **162:**383-394.

11. Galant N, Krawczyk P, Monist M, Obara A, Gajek Ł, Grenda A, Nicoś M, Kalinka E, Milanowski J: **Molecular Classification of Endometrial Cancer and Its Impact on Therapy Selection.** *Int J Mol Sci* 2024, **25**.

12. Anca-Stanciu MB, Manu A, Olinca MV, Coroleucă C, Comandașu DE, Coroleuca CA, Maier C, Bratila E: **Comprehensive Review of Endometrial Cancer: New Molecular and FIGO Classification and Recent Treatment Changes.** *J Clin Med* 2025, **14**.

13. Soslow RA, Tornos C, Park KJ, Malpica A, Matias-Guiu X, Oliva E, Parkash V, Carlson J, McCluggage WG, Gilks CB: **Endometrial Carcinoma Diagnosis: Use of FIGO Grading and Genomic Subcategories in Clinical Practice: Recommendations of the International Society of Gynecological Pathologists.** *Int J Gynecol Pathol* 2019, **38 Suppl 1:**S64-S74.

14. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, Robertson AG, Pashtan I, Shen R, Benz CC, et al: **Integrated genomic characterization of endometrial carcinoma.** *Nature* 2013, **497:**67-73.

15. León-Castillo A, de Boer SM, Powell ME, Mileshkin LR, Mackay HJ, Leary A, Nijman HW, Singh N, Pollock PM, Bessette P, et al: **Molecular Classification of the PORTEC-3 Trial for High-Risk Endometrial Cancer: Impact on Prognosis and Benefit From Adjuvant Therapy.** *J Clin Oncol* 2020, **38:**3388-3397.

16. Njoku K, Barr CE, Crosbie EJ: **Current and Emerging Prognostic Biomarkers in Endometrial Cancer.** *Front Oncol* 2022, **12:**890908.

17. Amant F, Moerman P, Neven P, Timmerman D, Van Limbergen E, Vergote I: **Endometrial cancer.** *Lancet* 2005, **366:**491-505.

18. Makker V, MacKay H, Ray-Coquard I, Levine DA, Westin SN, Aoki D, Oaknin A: **Endometrial cancer.** *Nat Rev Dis Primers* 2021, **7:**88.

19. Mahdy H, Vadakekut ES, Crotzer D: *Endometrial Cancer.* Treasure Island (FL): StatPearls Publishing; 2025.

20. Chou AJ, Bing RS, Ding DC: **Endometrial Atypical Hyperplasia and Risk of Endometrial Cancer.** *Diagnostics (Basel)* 2024, **14**.

21. Sideris M, Emin EI, Abdullah Z, Hanrahan J, Stefatou KM, Sevas V, Emin E, Hollingworth T, Odejinmi F, Papagrigoriadis S, et al: **The Role of KRAS in Endometrial Cancer: A Mini-Review.** *Anticancer Res* 2019, **39:**533-539.

22. Bosse T, ter Haar NT, Seeber LM, v Diest PJ, Hes FJ, Vasen HF, Nout RA, Creutzberg CL, Morreau H, Smit VT: **Loss of ARID1A expression and its relationship with PI3K-Akt pathway alterations, TP53 and microsatellite instability in endometrial cancer.** *Mod Pathol* 2013, **26:**1525-1535.

23. Toumpeki C, Liberis A, Tsirkas I, Tsirka T, Kalagasidou S, Inagamova L, Anthoulaki X, Tsatsaris G, Kontomanolis EN: **The Role of ARID1A in Endometrial Cancer and the Molecular Pathways Associated With Pathogenesis and Cancer Progression.** *In Vivo* 2019, **33:**659-667.

24. Yamada H, Takeshima H, Fujiki R, Yamashita S, Sekine S, Ando T, Hattori N, Okabe A, Yoshikawa T, Obama K, et al: **ARID1A loss-of-function induces CpG island methylator phenotype.** *Cancer Lett* 2022, **532:**215587.

25. De Tommasi O, Marchetti M, Tripepi M, Bigardi S, Incognito GG, Tuninetti V, Facchetti E, Tasca G, Noventa M, Saccardi C, et al: **PD-1 and PD-L1 Expression in Endometrial Cancer: A Systematic Review of the Literature.** *J Clin Med* 2025, **14**.

26. Urick ME, Bell DW: **Clinical actionability of molecular targets in endometrial cancer.** *Nat Rev Cancer* 2019, **19:**510-521.

27. Kurnit KC, Kim GN, Fellman BM, Urbauer DL, Mills GB, Zhang W, Broaddus RR: **CTNNB1 (beta-catenin) mutation identifies low grade, early stage endometrial cancer patients at increased risk of recurrence.** *Mod Pathol* 2017, **30:**1032-1041.

28. Jeong JW, Lee HS, Franco HL, Broaddus RR, Taketo MM, Tsai SY, Lydon JP, DeMayo FJ: **beta-catenin mediates glandular formation and dysregulation of beta-catenin induces hyperplasia formation in the murine uterus.** *Oncogene* 2009, **28:**31-40.

29. Ruz-Caracuel I, López-Janeiro Á, Heredia-Soto V, Ramón-Patino JL, Yébenes L, Berjón A, Hernández A, Gallego A, Ruiz P, Redondo A, et al: **Clinicopathological features and prognostic significance of CTNNB1 mutation in low-grade, early-stage endometrial endometrioid carcinoma.** *Virchows Arch* 2021, **479:**1167-1176.

30. Stampoliou A, Arapantoni-Dadioti P, Pavlakis K: **Epigenetic mechanisms in endometrial cancer.** *J BUON* 2016, **21:**301-306.

31. Pan Y, Liu G, Zhou F, Su B, Li Y: **DNA methylation profiles in cancer diagnosis and therapeutics.** *Clin Exp Med* 2018, **18:**1-14.

32. Kang S, Kim JW, Kang GH, Lee S, Park NH, Song YS, Park SY, Kang SB, Lee HP: **Comparison of DNA hypermethylation patterns in different types of uterine cancer: cervical squamous cell carcinoma, cervical adenocarcinoma and endometrial adenocarcinoma.** *Int J Cancer* 2006, **118:**2168-2171.

33. Shibata D, Mori Y, Cai K, Zhang L, Yin J, Elahi A, Hamelin R, Wong YF, Lo WK, Chung TK, et al: **RAB32 hypermethylation and microsatellite instability in gastric and endometrial adenocarcinomas.** *Int J Cancer* 2006, **119:**801-806.

34. Kilowski KA, Dietrich MF, Xiu J, Baca Y, Hinton A, Ahmad S, Herzog TJ, Thaker P, Holloway RW: **KRAS mutations in endometrial cancers: Possible prognostic and treatment implications.** *Gynecol Oncol* 2024, **191:**299-306.

35. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, et al: **CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer.** *Nat Genet* 2006, **38:**787-793.

36. Chen K, Yuan J, Sia Y, Chen Z: **Mechanism of action of the SWI/SNF family complexes.** *Nucleus* 2023, **14:**2165604.

37. Cajuso T, Hänninen UA, Kondelin J, Gylfe AE, Tanskanen T, Katainen R, Pitkänen E, Ristolainen H, Kaasinen E, Taipale M, et al: **Exome sequencing reveals frequent inactivating mutations in ARID1A, ARID1B, ARID2 and ARID4A in microsatellite unstable colorectal cancer.** *Int J Cancer* 2014, **135:**611-623.

38. Bartosch C, Lopes JM, Jerónimo C: **Epigenetics in endometrial carcinogenesis - part 2: histone modifications, chromatin remodeling and noncoding RNAs.** *Epigenomics* 2017, **9:**873-892.

39. Balch C, Matei DE, Huang TH, Nephew KP: **Role of epigenomics in ovarian and endometrial cancers.** *Epigenomics* 2010, **2:**419-447.

40. Janzen DM, Rosales MA, Paik DY, Lee DS, Smith DA, Witte ON, Iruela-Arispe ML, Memarzadeh S: **Progesterone receptor signaling in the microenvironment of endometrial cancer influences its response to hormonal therapy.** *Cancer Res* 2013, **73:**4697-4710.

41. Bartosch C, Monteiro-Reis S, Vieira R, Pereira A, Rodrigues M, Jerónimo C, Lopes JM: **Endometrial Endometrioid Carcinoma Metastases Show Decreased ER-Alpha and PR-A Expression Compared to Matched Primary Tumors.** *PLoS One* 2015, **10:**e0134969.

42. Beck CR, Garcia-Perez JL, Badge RM, Moran JV: **LINE-1 elements in structural variation and disease.** *Annu Rev Genomics Hum Genet* 2011, **12:**187-215.

43. Lavie L, Maldener E, Brouha B, Meese EU, Mayer J: **The human L1 promoter: variable transcription initiation sites and a major impact of upstream flanking sequence on promoter activity.** *Genome Res* 2004, **14:**2253-2260.

44. Hermant C, Torres-Padilla ME: **TFs for TEs: the transcription factor repertoire of mammalian transposable elements.** *Genes Dev* 2021, **35:**22-39.

45. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, Santamarina M, Ju YS, Temes J, Garcia-Souto D, et al: **Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition.** *Nat Genet* 2020, **52:**306-319.

46. Hancks DC, Kazazian HH: **Roles for retrotransposon insertions in human disease.** *Mob DNA* 2016, **7:**9.

47. Burns KH: **Transposable elements in cancer.** *Nat Rev Cancer* 2017, **17:**415-424.

48. Pavicic W, Joensuu EI, Nieminen T, Peltomäki P: **LINE-1 hypomethylation in familial and sporadic cancer.** *J Mol Med (Berl)* 2012, **90:**827-835.

49. Concin N, Matias-Guiu X, Vergote I, Cibula D, Mirza MR, Marnitz S, Ledermann J, Bosse T, Chargari C, Fagotti A, et al: **ESGO/ESTRO/ESP guidelines for the management of patients with endometrial carcinoma.** *Int J Gynecol Cancer* 2021, **31:**12-39.

50. Floyd JL, Campbell S, Rauh-Hain JA, Woodard T: **Fertility preservation in women with early-stage gynecologic cancer: optimizing oncologic and reproductive outcomes.** *Int J Gynecol Cancer* 2021, **31:**345-351.

51. Bellone S, Bignotti E, Lonardi S, Ferrari F, Centritto F, Masserdotti A, Pettinella F, Black J, Menderes G, Altwerger G, et al: **Polymerase ε (POLE) ultra-mutation in uterine tumors correlates with T lymphocyte infiltration and increased resistance to platinum-based chemotherapy in vitro.** *Gynecol Oncol* 2017, **144:**146-152.

52. Fader AN, Roque DM, Siegel E, Buza N, Hui P, Abdelghany O, Chambers SK, Secord AA, Havrilesky L, O'Malley DM, et al: **Randomized Phase II Trial of Carboplatin-Paclitaxel Versus Carboplatin-Paclitaxel-Trastuzumab in Uterine Serous Carcinomas That Overexpress Human Epidermal Growth Factor Receptor 2/neu.** *J Clin Oncol* 2018, **36:**2044-2051.

53. Lindemann K, Smogeli E, Småstuen MC, Bruheim K, Trovik J, Nordberg T, Kristensen GB, Werner HMJ, Nakken E: **Salvage Radiation for Pelvic Relapse after Surgically Treated Endometrial Cancer.** *Cancers (Basel)* 2021, **13**.

54. Timar J, Kashofer K: **Molecular epidemiology and diagnostics of KRAS mutations in human cancer.** *Cancer Metastasis Rev* 2020, **39:**1029-1038.

55. Wang H, Chi L, Yu F, Dai H, Gao C, Si X, Wang Z, Liu L, Zheng J, Shan L, et al: **Annual review of KRAS inhibitors in 2022.** *Eur J Med Chem* 2023, **249:**115124.

56. Huang L, Guo Z, Wang F, Fu L: **KRAS mutation: from undruggable to druggable in cancer.** *Signal Transduct Target Ther* 2021, **6:**386.

57. Liu Z, Lenz HJ, Yu J, Zhang L: **Differential Response and Resistance to KRAS-Targeted Therapy.** *Mol Carcinog* 2025, **64:**1135-1148.

58. Ostrem JM, Peters U, Sos ML, Wells JA, Shokat KM: **K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.** *Nature* 2013, **503:**548-551.

59. Parikh K, Banna G, Liu SV, Friedlaender A, Desai A, Subbiah V, Addeo A: **Drugging KRAS: current perspectives and state-of-art review.** *J Hematol Oncol* 2022, **15:**152.

60. Dang S, Zhang S, Zhao J, Li W: **Efficacy and toxicity of KRAS.** *World J Surg Oncol* 2024, **22:**182.

61. Sacher AG, Miller WH, Patel MR, Paz-Ares L, Santoro A, Ahn MJ, Dziadziuszko R, Freres P, Luo J, Bowyer S, et al: **Single-Agent Divarasib in Patients With.** *J Clin Oncol* 2025, **43:**3249-3253.

62. Miyashita H, Kato S, Hong DS: **KRAS G12C inhibitor combination therapies: current evidence and challenge.** *Front Oncol* 2024, **14:**1380584.

63. D'Alessio-Sands L, Gaynier J, Michel-Milian V, Agbowuro AA, Brackett CM: **Current Strategies and Future Dimensions in the Development of KRAS Inhibitors for Targeted Anticancer Therapy.** *Drug Dev Res* 2025, **86:**e70042.

64. Stelloo E, Jansen AML, Osse EM, Nout RA, Creutzberg CL, Ruano D, Church DN, Morreau H, Smit VTHB, van Wezel T, Bosse T: **Practical guidance for mismatch repair-deficiency testing in endometrial cancer.** *Ann Oncol* 2017, **28:**96-102.

65. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, Miller R, Riaz N, Douillard JY, Andre F, Scarpa A: **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.** *Ann Oncol* 2019, **30:**1232-1243.

66. Bostan IS, Mihaila M, Roman V, Radu N, Neagu MT, Bostan M, Mehedintu C: **Landscape of Endometrial Cancer: Molecular Mechanisms, Biomarkers, and Target Therapy.** *Cancers (Basel)* 2024, **16**.

67. Kommoss S, McConechy MK, Kommoss F, Leung S, Bunz A, Magrill J, Britton H, Grevenkamp F, Karnezis A, Yang W, et al: **Final validation of the ProMisE molecular classifier for endometrial carcinoma in a large population-based case series.** *Ann Oncol* 2018, **29:**1180-1188.

68. Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, Chung HC, Kindler HL, Lopez-Martin JA, Miller WH, et al: **Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study.** *Lancet Oncol* 2020, **21:**1353-1365.

69. O'Malley DM, Bariani GM, Cassier PA, Marabelle A, Hansen AR, De Jesus Acosta A, Miller WH, Safra T, Italiano A, Mileshkin L, et al: **Pembrolizumab in Patients With Microsatellite Instability-High Advanced Endometrial Cancer: Results From the KEYNOTE-158 Study.** *J Clin Oncol* 2022, **40:**752-761.

70. Marques AC, Ferraro-Peyret C, Michaud F, Song L, Smith E, Fabre G, Willig A, Wong MML, Xing X, Chong C, et al: **Improved NGS-based detection of microsatellite instability using tumor-only data.** *Front Oncol* 2022, **12:**969238.

71. Škerl P, Vogrič V, Stegel V, Šetrajčič Dragoš V, Blatnik O, Klančar G, Novaković S: **Real-World Evaluation of Microsatellite Instability Detection via Targeted NGS Panels in Routine Molecular Diagnostics.** *Int J Mol Sci* 2025, **26**.

72. Addante F, d'Amati A, Santoro A, Angelico G, Inzani F, Arciuolo D, Travaglino A, Raffone A, D'Alessandris N, Scaglione G, et al: **Mismatch Repair Deficiency as a Predictive and Prognostic Biomarker in Endometrial Cancer: A Review on Immunohistochemistry Staining Patterns and Clinical Implications.** *Int J Mol Sci* 2024, **25**.

73. Liu C, Ping H, Yao M, Li X, Li Q, Hu R, Xu Y, Meng K, Gao F: **Traditional and New Views on MSI-H/dMMR Endometrial Cancer.** *Biomolecules* 2025, **15**.

74. Ugai S, Yao Q, Takashima Y, Zhong Y, Matsuda K, Kawamura H, Imamura Y, Okadome K, Mima K, Arima K, et al: **Clinicopathological, molecular, and prognostic features of colorectal carcinomas with KRAS c.34G>T (p.G12C) mutation.** *Cancer Sci* 2024, **115:**3455-3465.

75. Tew BY, Durand JK, Bryant KL, Hayes TK, Peng S, Tran NL, Gooden GC, Buckley DN, Der CJ, Baldwin AS, Salhia B: **Genome-wide DNA methylation analysis of KRAS mutant cell lines.** *Sci Rep* 2020, **10:**10149.

76. Nagasaka T, Koi M, Kloor M, Gebert J, Vilkin A, Nishida N, Shin SK, Sasamoto H, Tanaka N, Matsubara N, et al: **Mutations in both KRAS and BRAF may contribute to the methylator phenotype in colon cancer.** *Gastroenterology* 2008, **134:**1950-1960, 1960.e1951.

77. Serra RW, Fang M, Park SM, Hutchinson L, Green MR: **A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype.** *Elife* 2014, **3:**e02313.

78. Meintani A, Ozdogan M, Touroutoglou N, Papazisis K, Boukovinas I, Bilir C, Giassas S, Sualp T, Lacin S, Dan Corneliu J, et al: **Comprehensive Evaluation of a 1021-Gene Panel in FFPE and Liquid Biopsy for Analytical and Clinical Use.** *Int J Mol Sci* 2025, **26**.

79. Jennings LJ, Arcila ME, Corless C, Kamel-Reid S, Lubin IM, Pfeifer J, Temple-Smolkin RL, Voelkerding KV, Nikiforova MN: **Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists.** *J Mol Diagn* 2017, **19:**341-365.

80. Mosele MF, Westphalen CB, Stenzinger A, Barlesi F, Bayle A, Bièche I, Bonastre J, Castro E, Dienstmann R, Krämer A, et al: **Recommendations for the use of next-generation sequencing (NGS) for patients with advanced cancer in 2024: a report from the ESMO Precision Medicine Working Group.** *Ann Oncol* 2024, **35:**588-606.

81. Prior IA, Hood FE, Hartley JL: **The Frequency of Ras Mutations in Cancer.** *Cancer Res* 2020, **80:**2969-2974.

82. Moore AR, Rosenberg SC, McCormick F, Malek S: **RAS-targeted therapies: is the undruggable drugged?** *Nat Rev Drug Discov* 2020, **19:**533-552.

83. Hobbs GA, Der CJ, Rossman KL: **RAS isoforms and mutations in cancer at a glance.** *J Cell Sci* 2016, **129:**1287-1292.

84. Kadur Lakshminarasimha Murthy P, Xi R, Arguijo D, Everitt JI, Kocak DD, Kobayashi Y, Bozec A, Vicent S, Ding S, Crawford GE, et al: **Epigenetic basis of oncogenic-Kras-mediated epithelial-cellular proliferation and plasticity.** *Dev Cell* 2022, **57:**310-328.e319.

85. Chen Z, Zhang L, Yang Y, Liu H, Kang X, Nie Y, Fan D: **DNMT1 expression partially dictates 5-Azacytidine sensitivity and correlates with RAS/MEK/ERK activity in gastric cancer cells.** *Epigenetics* 2023, **18:**2254976.

86. Mathison AJ, Kerketta R, de Assuncao TM, Leverence E, Zeighami A, Urrutia G, Stodola TJ, di Magliano MP, Iovanna JL, Zimmermann MT, et al: **Kras.** *Genome Biol* 2021, **22:**289.

87. Sultana A, Rana S: **Mechanisms underlying obesity-malignancy connection: a systematic narrative review.** *J Physiol Biochem* 2025, **81:**403-439.

88. Ring KL, Yates MS, Schmandt R, Onstad M, Zhang Q, Celestino J, Kwan SY, Lu KH: **Endometrial Cancers With Activating KRas Mutations Have Activated Estrogen Signaling and Paradoxical Response to MEK Inhibition.** *Int J Gynecol Cancer* 2017, **27:**854-862.

89. Kim JH, Bae JM, Cho NY, Kang GH: **Distinct features between MLH1-methylated and unmethylated colorectal carcinomas with the CpG island methylator phenotype: implications in the serrated neoplasia pathway.** *Oncotarget* 2016, **7:**14095-14111.

90. Farchoukh L, Kuan SF, Dudley B, Brand R, Nikiforova M, Pai RK: **MLH1-deficient Colorectal Carcinoma With Wild-type BRAF and MLH1 Promoter Hypermethylation Harbor KRAS Mutations and Arise From Conventional Adenomas.** *Am J Surg Pathol* 2016, **40:**1390-1399.

91. Cohen J, R W, BY K, C W: **A KRAS hot spot mutation correlates with MLH1 methylation in endometrial carcinomas with microsatellite instability: A potential triage tool for Lynch syndrome evaluation.** *Gynecologic Oncology* 2014, **133:**79-80.

92. Esteller M, Levine R, Baylin SB, Ellenson LH, Herman JG: **MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas.** *Oncogene* 1998, **17:**2413-2417.

93. Simpkins SB, Bocker T, Swisher EM, Mutch DG, Gersell DJ, Kovatich AJ, Palazzo JP, Fishel R, Goodfellow PJ: **MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers.** *Hum Mol Genet* 1999, **8:**661-666.

94. Ellenson LH: **hMLH1 promoter hypermethylation in microsatellite instability-positive endometrial carcinoma. Cause or consequence?** *Am J Pathol* 1999, **155:**1399-1402.

95. Plotkin A, Olkhov-Mitsel E, Nofech-Mozes S: **Methylation Testing as an Integral Component of Universal Endometrial Cancer Screening-A Critical Appraisal.** *Cancers (Basel)* 2023, **15**.

96. Manning-Geist BL, Liu YL, Devereaux KA, Paula ADC, Zhou QC, Ma W, Selenica P, Ceyhan-Birsoy O, Moukarzel LA, Hoang T, et al: **Microsatellite Instability-High Endometrial Cancers with MLH1 Promoter Hypermethylation Have Distinct Molecular and Clinical Profiles.** *Clin Cancer Res* 2022, **28:**4302-4311.

97. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, Skora AD, Luber BS, Azad NS, Laheru D, et al: **PD-1 Blockade in Tumors with Mismatch-Repair Deficiency.** *N Engl J Med* 2015, **372:**2509-2520.

98. Levine AJ, Phipps AI, Baron JA, Buchanan DD, Ahnen DJ, Cohen SA, Lindor NM, Newcomb PA, Rosty C, Haile RW, et al: **Clinicopathologic Risk Factor Distributions for MLH1 Promoter Region Methylation in CIMP-Positive Tumors.** *Cancer Epidemiol Biomarkers Prev* 2016, **25:**68-75.

99. Livshits G, Alonso-Curbelo D, Morris JP, Koche R, Saborowski M, Wilkinson JE, Lowe SW: **Arid1a restrains Kras-dependent changes in acinar cell identity.** *Elife* 2018, **7**.

100. Guo B, Friedland SC, Alexander W, Myers JA, Wang W, O'Dell MR, Getman M, Whitney-Miller CL, Agostini-Vulaj D, Huber AR, et al: **Arid1a mutation suppresses TGF-β signaling and induces cholangiocarcinoma.** *Cell Rep* 2022, **40:**111253.

101. Sen M, Wang X, Hamdan FH, Rapp J, Eggert J, Kosinsky RL, Wegwitz F, Kutschat AP, Younesi FS, Gaedcke J, et al: **ARID1A facilitates KRAS signaling-regulated enhancer activity in an AP1-dependent manner in colorectal cancer cells.** *Clin Epigenetics* 2019, **11:**92.

102. Joyce BT, Gao T, Zheng Y, Liu L, Zhang W, Dai Q, Shrubsole MJ, Hibler EA, Cristofanilli M, Zhang H, et al: **Prospective changes in global DNA methylation and cancer incidence and mortality.** *Br J Cancer* 2016, **115:**465-472.

103. Ogino S, Kawasaki T, Nosho K, Ohnishi M, Suemoto Y, Kirkner GJ, Fuchs CS: **LINE-1 hypomethylation is inversely associated with microsatellite instability and CpG island methylator phenotype in colorectal cancer.** *Int J Cancer* 2008, **122:**2767-2773.

104. Tang JT, Wang ZH, Fang JY: **Assessing the potential value of long interspersed element-1 hypomethylation in colorectal cancer: evidence from retrospective studies.** *Onco Targets Ther* 2015, **8:**3265-3276.

105. Silva FFVE, Ballini A, Caponio VCA, Pérez-Sayáns M, Cortés MG, Rojo-Álvarez LI, García-García A, Suaréz-Peñaranda JM, Di Domenico M, Padín-Iruegas ME: **Insights into MLH1 Methylation in Endometrial Adenocarcinoma through Pyrosequencing Analysis: A Retrospective Observational Study.** *Cancers (Basel)* 2024, **16**.

106. Wang C, Kuang W, Zeng J, Ren Y, Liu Q, Sun H, Feng M, Liang D: **A retrospective study of consistency between immunohistochemistry and polymerase chain reaction of microsatellite instability in endometrial cancer.** *PeerJ* 2023, **11:**e15920.

107. Huang W, Ho CL, Lee CT, Chen WL, Yang SC, Chow NH, Chen YL: **High concordance rate of capillary electrophoresis workflow for microsatellite instability analysis and mismatch repair (MMR) immunostaining in colorectal carcinoma.** *PLoS One* 2023, **18:**e0284227.

108. Bonneville R, Krook MA, Chen HZ, Smith A, Samorodnitsky E, Wing MR, Reeser JW, Roychowdhury S: **Detection of Microsatellite Instability Biomarkers via Next-Generation Sequencing.** *Methods Mol Biol* 2020, **2055:**119-132.

109. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, Wendl MC, Ding L: **MSIsensor: microsatellite instability detection using paired tumor-normal sequence data.** *Bioinformatics* 2014, **30:**1015-1016.

110. Jia P, Yang X, Guo L, Liu B, Lin J, Liang H, Sun J, Zhang C, Ye K: **MSIsensor-pro: Fast, Accurate, and Matched-normal-sample-free Detection of Microsatellite Instability.** *Genomics Proteomics Bioinformatics* 2020, **18:**65-71.

111. Salipante SJ, Scroggins SM, Hampel HL, Turner EH, Pritchard CC: **Microsatellite instability detection by next generation sequencing.** *Clin Chem* 2014, **60:**1192-1199.

112. Kautto EA, Bonneville R, Miya J, Yu L, Krook MA, Reeser JW, Roychowdhury S: **Performance evaluation for rapid detection of pan-cancer microsatellite instability with MANTIS.** *Oncotarget* 2017, **8:**7452-7463.

113. Anthony H, Seoighe C: **Performance assessment of computational tools to detect microsatellite instability.** *Brief Bioinform* 2024, **25**.

114. Xiao J, Li W, Huang Y, Huang M, Li S, Zhai X, Zhao J, Gao C, Xie W, Qin H, et al: **A next-generation sequencing-based strategy combining microsatellite instability and tumor mutation burden for comprehensive molecular diagnosis of advanced colorectal cancer.** *BMC Cancer* 2021, **21:**282.

115. Ratovomanana T, Cohen R, Svrcek M, Renaud F, Cervera P, Siret A, Letourneur Q, Buhard O, Bourgoin P, Guillerm E, et al: **Performance of Next-Generation Sequencing for the Detection of Microsatellite Instability in Colorectal Cancer With Deficient DNA Mismatch Repair.** *Gastroenterology* 2021, **161:**814-826.e817.

116. Kavun A, Veselovsky E, Lebedeva A, Belova E, Kuznetsova O, Yakushina V, Grigoreva T, Mileyko V, Fedyanin M, Ivanov M: **Microsatellite Instability: A Review of Molecular Epidemiology and Implications for Immune Checkpoint Inhibitor Therapy.** *Cancers (Basel)* 2023, **15**.

117. Nádorvári ML, Lotz G, Kulka J, Kiss A, Tímár J: **Microsatellite instability and mismatch repair protein deficiency: equal predictive markers?** *Pathol Oncol Res* 2024, **30:**1611719.

118. Kang SY, Kim DG, Ahn S, Ha SY, Jang KT, Kim KM: **Comparative analysis of microsatellite instability by next-generation sequencing, MSI PCR and MMR immunohistochemistry in 1942 solid cancers.** *Pathol Res Pract* 2022, **233:**153874.

119. Ali-Fehmi R, Krause HB, Morris RT, Wallbillich JJ, Corey L, Bandyopadhyay S, Kheil M, Elbashir L, Zaiem F, Quddus MR, et al: **Analysis of Concordance Between Next-Generation Sequencing Assessment of Microsatellite Instability and**

**Immunohistochemistry-Mismatch Repair From Solid Tumors.** *JCO Precis Oncol* 2024, **8:**e2300648.

120. Riedinger CJ, Esnakula A, Haight PJ, Suarez AA, Chen W, Gillespie J, Villacres A, Chassen A, Cohn DE, Goodfellow PJ, Cosgrove CM: **Characterization of mismatch-repair/microsatellite instability-discordant endometrial cancers.** *Cancer* 2024, **130:**385-399.

121. Mirza MR, Chase DM, Slomovitz BM, dePont Christensen R, Novák Z, Black D, Gilbert L, Sharma S, Valabrega G, Landrum LM, et al: **Dostarlimab for Primary Advanced or Recurrent Endometrial Cancer.** *N Engl J Med* 2023, **388:**2145-2158.

122. Hussain HU, Burney MH, Rehan ST, Hasan MM: **Dostarlimab: A breakthrough in the field of oncology.** *Ann Med Surg (Lond)* 2022, **80:**104046.

123. Hu J, Yan WY, Xie L, Cheng L, Yang M, Li L, Shi J, Liu BR, Qian XP: **Coexistence of MSI with KRAS mutation is associated with worse prognosis in colorectal cancer.** *Medicine (Baltimore)* 2016, **95:**e5649.

124. Orlandi E, Giuffrida M, Trubini S, Luzietti E, Ambroggi M, Anselmi E, Capelli P, Romboli A: **Unraveling the Interplay of KRAS, NRAS, BRAF, and Micro-Satellite Instability in Non-Metastatic Colon Cancer: A Systematic Review.** *Diagnostics (Basel)* 2024, **14**.

125. Bonnet E, Moutet ML, Baulard C, Bacq-Daian D, Sandron F, Mesrob L, Fin B, Delépine M, Palomares MA, Jubin C, et al: **Performance comparison of three DNA extraction kits on human whole-exome data from formalin-fixed paraffin-embedded normal and tumor samples.** *PLoS One* 2018, **13:**e0195471.

126. Steiert TA, Parra G, Gut M, Arnold N, Trotta JR, Tonda R, Moussy A, Gerber Z, Abuja PM, Zatloukal K, et al: **A critical spotlight on the paradigms of FFPE-DNA sequencing.** *Nucleic Acids Res* 2023, **51:**7143-7162.

127. Roessler J, Lehmann U: **Quantitative DNA Methylation Analysis by Pyrosequencing®.** *Methods Mol Biol* 2015, **1315:**175-188.

# APPENDIX A

**Supplementary table 1: Archer® VariantPlex® Solid Tumour NGS panel.**

| Serial | Gene Name | Transcript | Exons |
|---|---|---|---|
| 1 | ABL1 | NM_005157 | 4-7 |
| 2 | AKT1 | NM_005163 | 3,6 |
| 3 | ALK | NM_004304 | 21,22,23,25 |
| 4 | APC | NM_000038 | 16 |
| 5 | ARID1A | NM_006015.6 | 1-20 |
| 6 | ATM | NM_000051 | 8,9,12,17,26,34,35,36,39,50,54,55,56,59,61,63 |
| 7 | AURKA | NM_003600 | 2,5,6,7,8 |
| 8 | BRAF | NM_004333 | 11,15 |
| 9 | CCND1 | NM_053056 | na |
| 10 | CCNE1 | NM_001238 | 3,4,5,6,7,8,10,12 |
| 11 | CDH1 | NM_004360 | 1-16 |
| 12 | CDK4 | NM_000075 | 2-7 |
| 13 | CDKN2A | NM_000077 | 1-3 |
| 14 | CSF1R | NM_005211 | 7,22 |
| 15 | CTNNB1 | NM_001904 | 3 |
| 16 | DDR2 | NM_006182 | 12-18 |
| 17 | EGFR | NM_005228 | 3,7,15,18,19,20,21 |
| 18 | ERBB2 | NM_004448 | 10,19,20,21,24 |
| 19 | ERBB3 | NM_001982 | 2,3,7,8 |
| 20 | ERBB4 | NM_005235 | 3,4,6,7,8,9,15,23 |
| 21 | ESR1 | NM_000125 | 8 |
| 22 | EZH2 | NM_004456 | 16 |
| 23 | FBXW7 | NM_018315 | 1-11 |
| 24 | FGFR1 | NM_015850 | 4,7,8,13,15,17 |
| 25 | FGFR2 | NM_000141 | 7,9,12,14 |
| 26 | FGFR3 | NM_000142 | 7,8,9,14,15,16,18 |
| 27 | FLT3 | NM_004119 | 11,14,16,20 |
| 28 | FOXL2 | NM_023067 | 1 |
| 29 | GNA11 | NM_002067 | 5 |
| 30 | GNAQ | NM_002072 | 4,5 |
| 31 | GNAS | NM_000516 | 6-9 |
| 32 | H3F3A | NM_002107 | 2 |
| 33 | HNF1A | NM_000545 | 3,4 |
| 34 | HRAS | NM_005343 | 2,3 |
| 35 | IDH1 | NM_005896 | 3,4 |
| 36 | IDH2 | NM_002168 | 4 |
| 37 | JAK2 | NM_004972 | 11,13,14,16,19 |
| 38 | JAK3 | NM_000215 | 4,13,16 |
| 39 | KDR | NM_002253 | 6,7,11,19,21,26,27,30 |
| 40 | KIT | NM_000222 | 2,8,9,10,11,13,14,15,17,18 |
| 41 | KRAS | NM_004985 | 2-5 |

| | | | |
|---|---|---|---|
| 42 | MAP2K1 | NM_002755 | 2,3 |
| 43 | MDM2 | NM_002392 | 2,3,4,6,8,10 |
| 44 | MET | NM_000245 | 2,11,14,16,19,21 |
| 45 | MLH1 | NM_000249 | 12 |
| 46 | MPL | NM_005373 | 10 |
| 47 | MYC | NM_002467 | na |
| 48 | MYCN | NM_005378 | na |
| 49 | NOTCH1 | NM_017617 | 25,26,27,34 |
| 50 | NPM1 | NM_002520 | 11 |
| 51 | NRAS | NM_002524 | 2-5 |
| 52 | PDGFRA | NM_006206 | 12,14,15,18,23 |
| 53 | PIK3CA | NM_006218 | 2,5,7,8,10,14,19,21 |
| 54 | PIK3R1 | NM_181504 | 1-10 |
| 55 | POLD1 | NM_002691.3 | 1-27 |
| 56 | POLE | NM_006231.4 | 1-49 |
| 57 | PTEN | NM_000314 | 1-9 |
| 58 | PTPN11 | NM_002834 | 3,13 |
| 59 | RB1 | NM_000321 | 4,6,10,11,14,17,18,20,21,22 |
| 60 | RET | NM_020630 | 10-16 |
| 61 | RHOA | NM_001664 | 2,3 |
| 62 | ROS1 | NM_002944 | 38 |
| 63 | SMAD4 | NM_005359 | 2-12 |
| 64 | SMARCB1 | NM_003073 | 2,4,5,9 |
| 65 | SMO | NM_005631 | 3,5,6,9,11 |
| 66 | SRC | NM_005417 | 14 |
| 67 | STK11 | NM_000455 | 1-9 |
| 68 | TERT | NM_198253 | Promoter, 1 |
| 69 | TP53 | NM_000546 | 1-11 |
| 70 | VHL | NM_000551 | 1-3 |

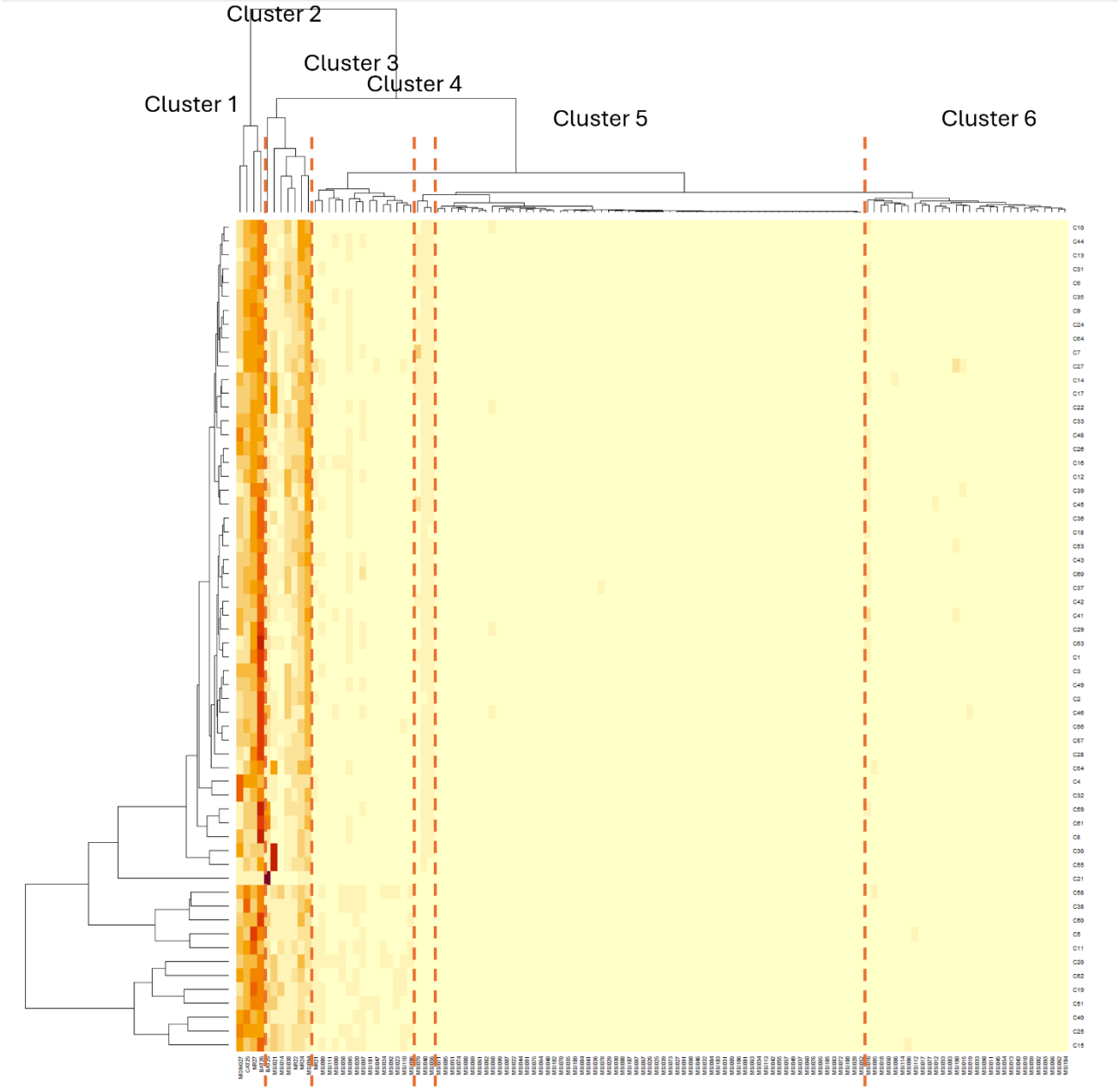
**Supplementary table 2: SALSA MS-MLPA ME011 probes.**

| Gene/exon | Length (nt) | SALSA MLPA probe | Ligation site | Sequence | Distance to next probe |
|---|---|---|---|---|---|
| ***EPCAM*** exon 8 | 197 | 11983-L30436 | 1065-1066 | GTTATTTCCAGA-AAGAAGAGAATG | 1.5 kb |
| ***EPCAM*** exon 9 | 310 | 13131-L03603 | 1189-1190 | AAATGGACACAA-ATTACAAATGTG | 0.1 kb |
| ***EPCAM*** exon 9 | 220 | 22424-L31926 | 1320-1319 reverse | GGTCAAATTTCA-AGATTGGTAAAG | 16.1 kb to *MSH2* |
| ***MSH2*** upstream | 184 | 06227-L07711 | 233 nt before exon 1 reverse | GAAACCCGCAGACGCGCATCCT-TAGTAGAGCTCCTTTCTGTGTTTACTCAGCT GCAAGGCTTG | 0.1 kb |
| ***MSH2*** upstream | 352 | 02735-L02162 | 156 nt before exon 1 | CAGTAGCTAAAGTCACCAGCGTGCGCGGGA AGCTGGGCCGCGTCTGCTTATGATTGGTTGC CGC | 0.2 kb |
| ***MSH2*** exon 1 | 254 | 06226-L31127 | 5-6 | GGAAACAGCTTAGTGGGTGTGGGGTCGCGC ATT-TTCTTCAACCAGGAGGTGAGGAGGTTTCGA CATGGCGGTG | 379.8 kb to *MSH6* |
| ***MSH6*** upstream | 148 | 21588-L30979 | 228 nt before exon 1 | CGGCCAGCCCCGCGGCGTGAGGGAAGGGG AGCTCAGCAGTTCCCCGCGCGGGGCC | 0.2 kb |
| ***MSH6*** upstream | 213 | 06230-L29780 | 37 nt before exon 1 | CGGCGAGGCGCCTGTTGATTGGCCACTGGG GCCCGGGTTCCTCCGGCGGAGCGCGCCT | 0.1 kb |
| ***MSH6*** exon 1 | 167 | 06228-L30148 | 41-40 reverse | CAACCGTTCTGTCGGACGGAGCTCCTAAAA-GCACCGCATCTACCGCGCGGCTCCTGCTGG CGGGAAATCTG | - |
| ***MLH1*** upstream | 247 | 07187-L26307 | 628 nt before exon 1 (Deng **A**) | CGTCCGCCACATACCGCTCGTAGTAT-TCGTGCTCAGCCTCGTAGTGGCGCCTGACG TCGCGTT | 0.3 kb |
| ***MLH1*** upstream | 278 | 06221-L31128 | 352 nt before exon 1 reverse (Deng **B**) | TGTGCCTCTGCTGAGGTGATCTGGCGCAGA-GCGGAGGAGGTGCTTGGCGCTTCTCAGGCT CCTCCTCT | 0.1 kb |
| ***MLH1*** upstream | 201 | 06222-L26305 | 215 nt before exon 1 (Deng **C**) | CAATAGGAAGAGCGGACAGCGATCTCTAAC GCGCAA-GCGCATATCCTTCTAGGTAGCGGGCAGTAG CCGCTTCAGG | 0.2 kb |
| ***MLH1*** exon 1 | 289 | 22572-L31773 | 25-24 reverse; **SNP-specific** | GGCTCTTCTGGT-GCCAAAATGTCG | 0 kb |
| ***MLH1*** exon 1 | 172 | 01686-L28585 | 18-19 (Deng **D**) | CTTCCGTTGAGCATCTAGACGTTTCCTTGGC TCT-TCTGGCGCCAAAATGTCGTTCGTGGCAGGG GTTATTC | 0.2 kb |
| ***MLH1*** intron 1 | 297 | 02258-L30977 | 93 nt after exon 1 | CGGACACGCCTCTTTGCCCGGGCAGAGGCA TGTACAGCGCATGCCCACAACGGCGGAGGC C | - |

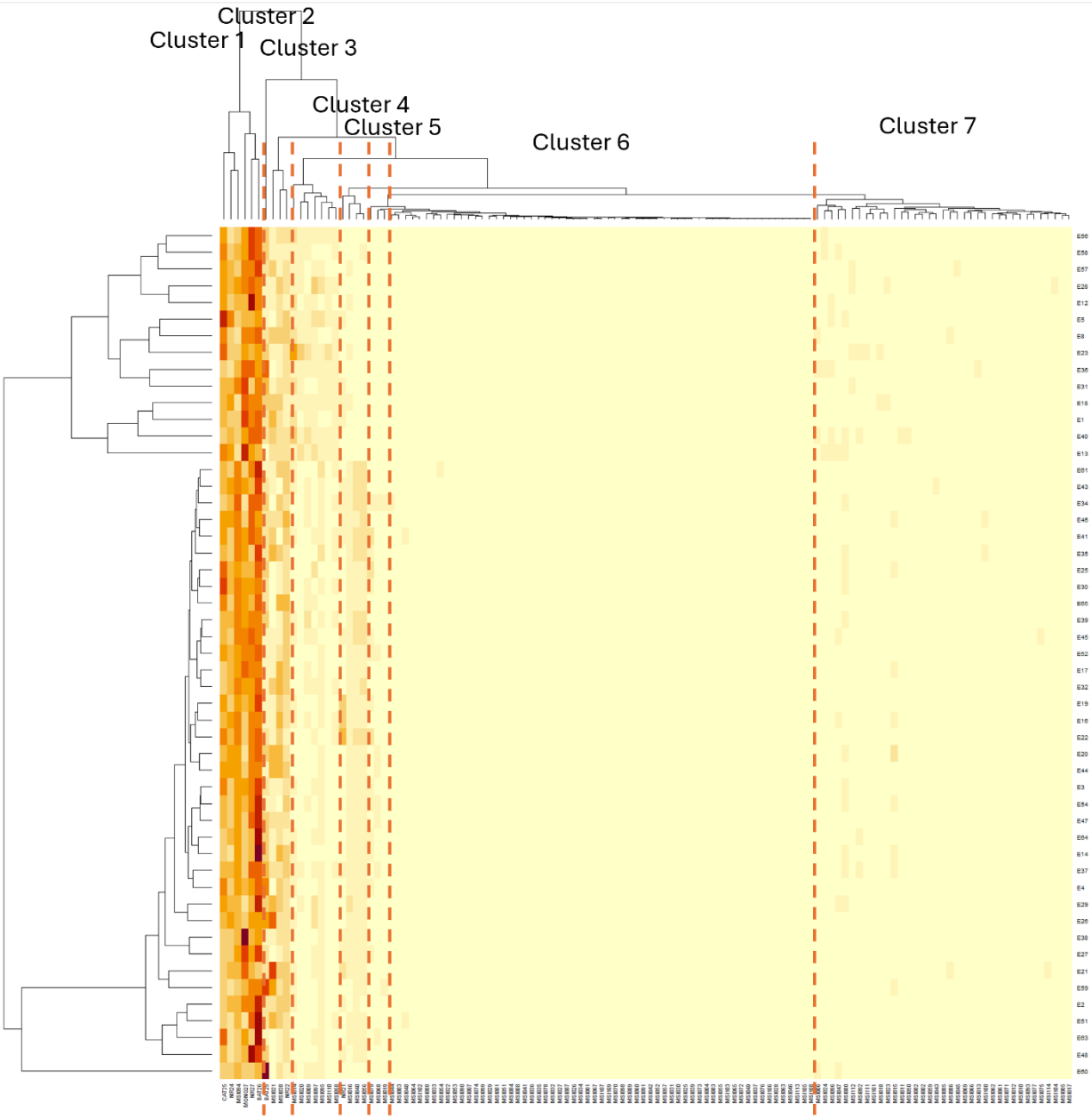
| | | | | | |
|---|---|---|---|---|---|
| ***PMS2*** intron 1 | 154 | 11966-L13112 | 20 nt after exon 1 | GCTCGAGGTGAGCGGGGCTCGCAGTCTTCC GGTGTCCCCTCTCGCGCGCCCTCTTTGAGAC | 0.1 kb |
| ***PMS2*** upstream | 142 | 07935-L16571 | 32 nt before exon 1 | GCCAATGGGAGTTCAGGAGGCGGAGCGCCT GTGGGAGCCCTGGAGGGAACTTTCCCAGT | 0.2 kb |
| ***PMS2*** upstream | 338 | 07934-L16147 | 276 nt before exon 1 | GGCAGAACCAAAGCAAAAGGGGGTAGCGC GTGCCAAAG-GCCAACGCTCAGAAACCGTCAGAGGTCACG ACGGAGAC | - |

*The selected GCGC sites are highlighted in grey. MS-MLPA, Methylation-Specific Multiplex Ligation-dependent Probe Amplification.*

# APPENDIX B

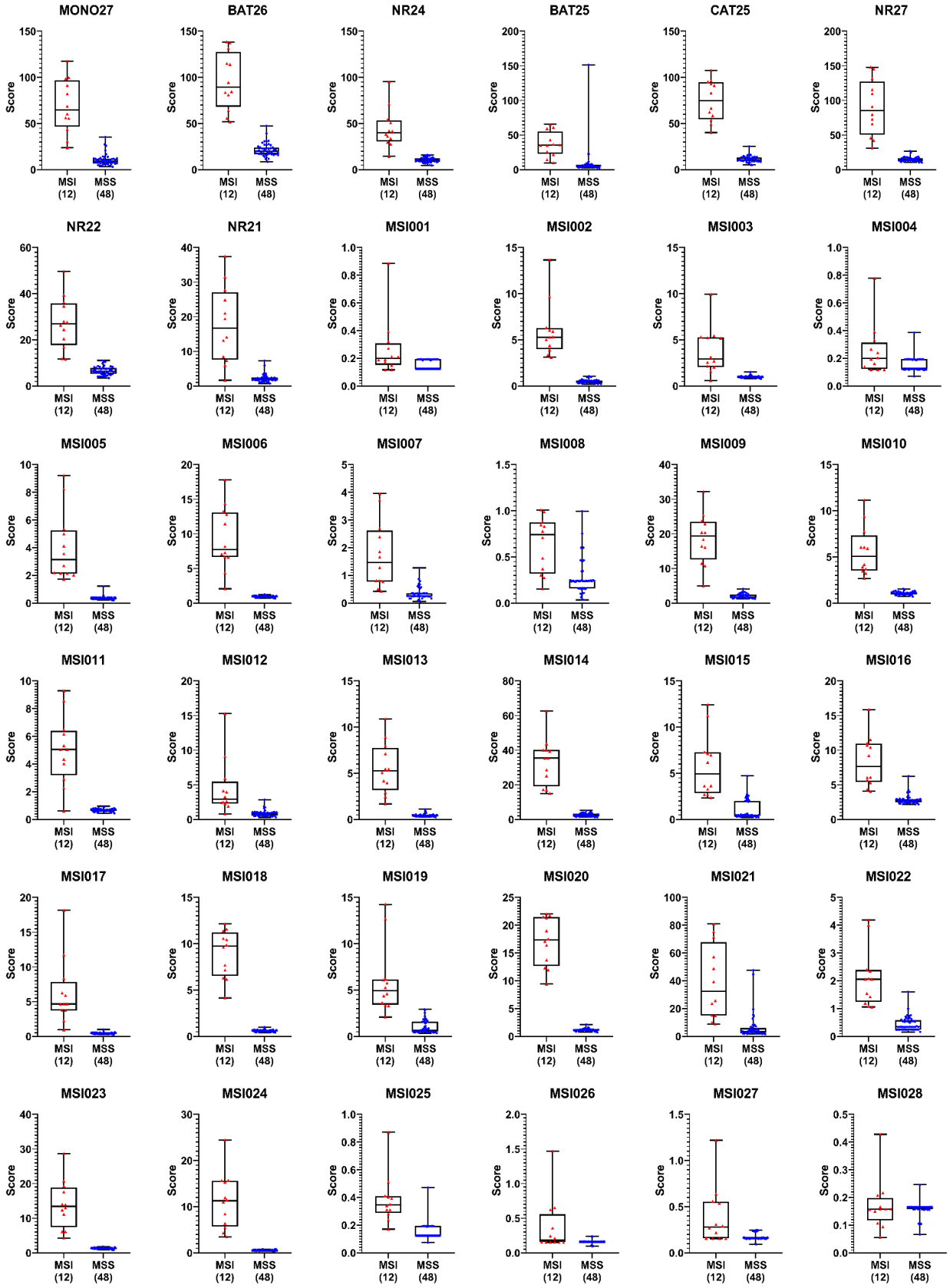


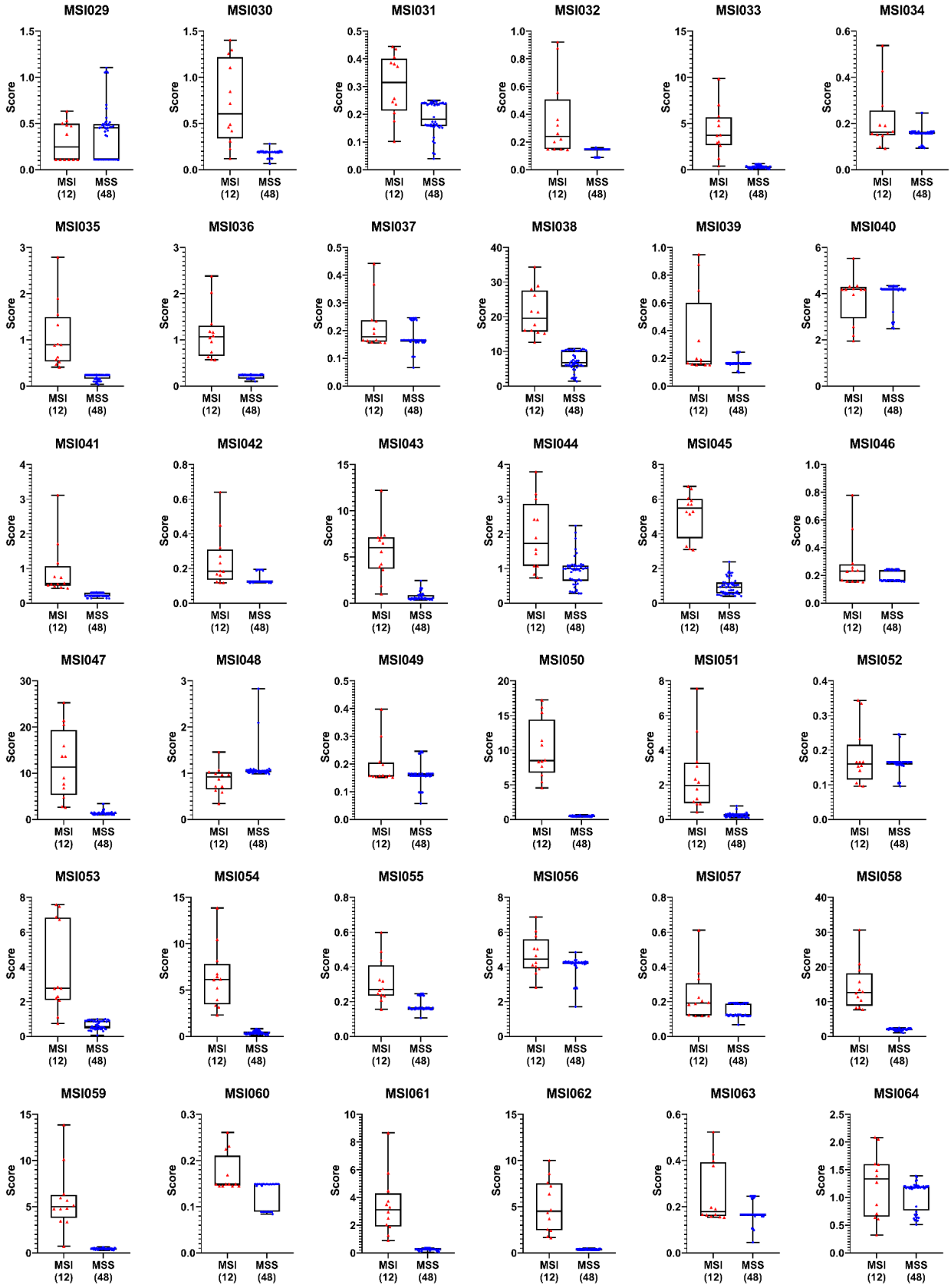

(a) Heatmap of 122 loci across concordant CRCs (n=60)

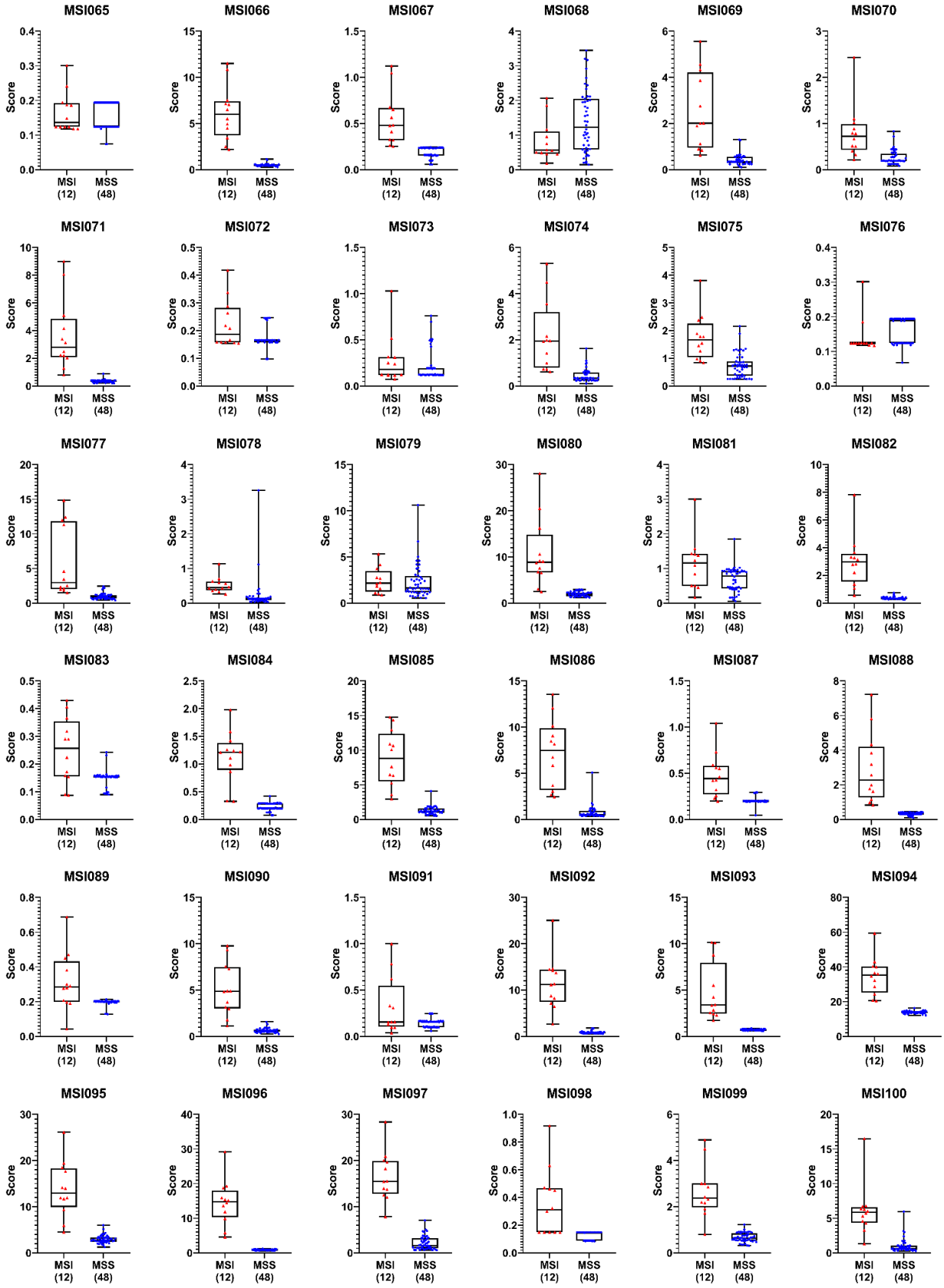


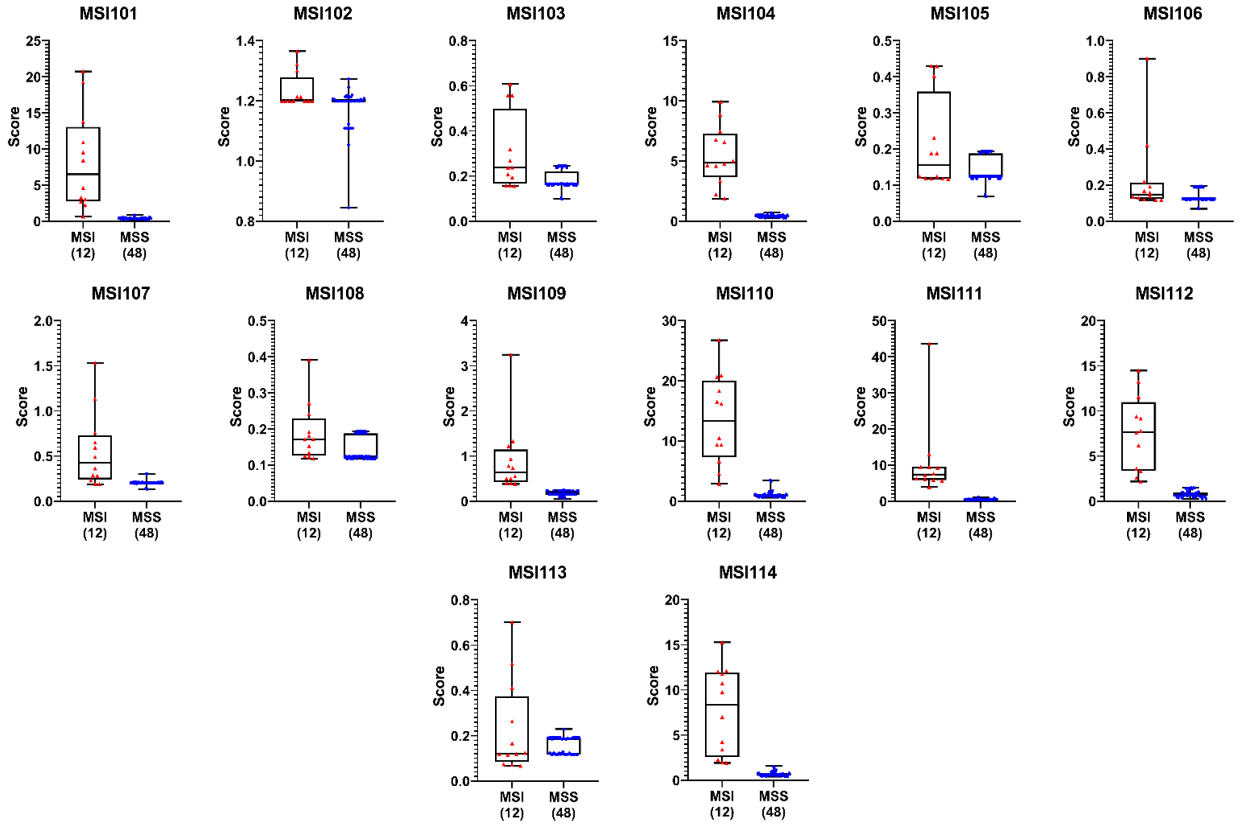

(b) Heatmap of 122 loci across concordant ECs (n=51)

**Supplementary figure 1: Unsupervised hierarchical clustering of microsatellite loci in CRC and EC.** (a) Hierarchal clustering of 122 loci across MSI and MSS cases of CRC (n=60) concordant between MMR (IHC/MSI CE) and NGS testing resulted in six distinct clusters. Clusters 1, 2, 3, 4 and 6 were selected as they contained most informative loci for distinguishing MSI from MSS tumours. (b)Whereas clustering of 122 loci across concordant MSI and MSS cases EC (n=51) yielded seven clusters. Clusters 1, 2, 3, 4, 5, and 7 were retained for selection of informative loci.

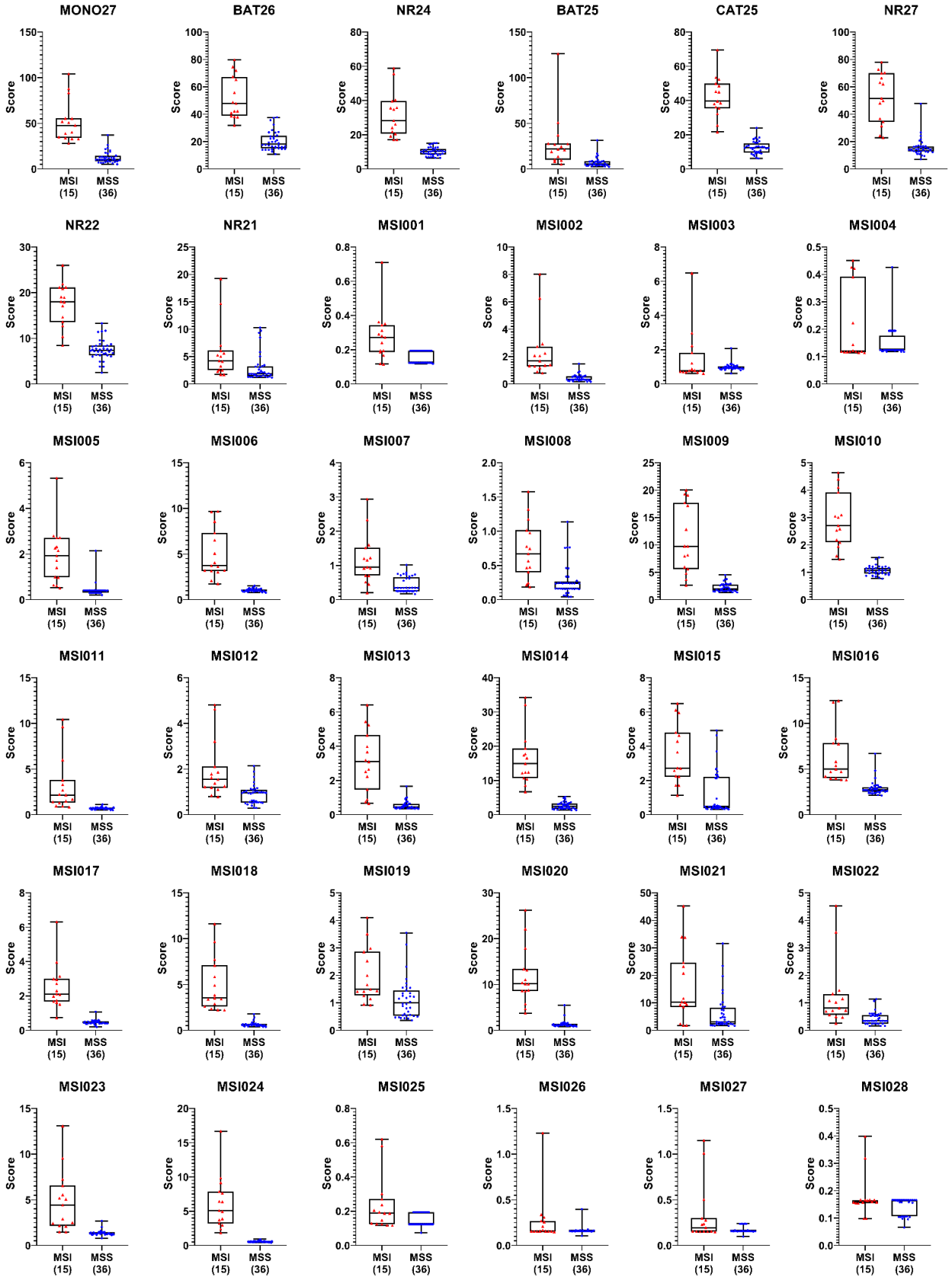

MONO27
BAT26
NR24
BAT25
CAT25
NR27
NR22
NR21
MSI001
MSI002
MSI003
MSI004
MSI005
MSI006
MSI007
MSI008
MSI009
MSI010
MSI011
MSI012
MSI013
MSI014
MSI015
MSI016
MSI017
MSI018
MSI019
MSI020
MSI021
MSI022
MSI023
MSI024
MSI025
MSI026
MSI027
MSI028
Score
MSI (12)
MSS (48)

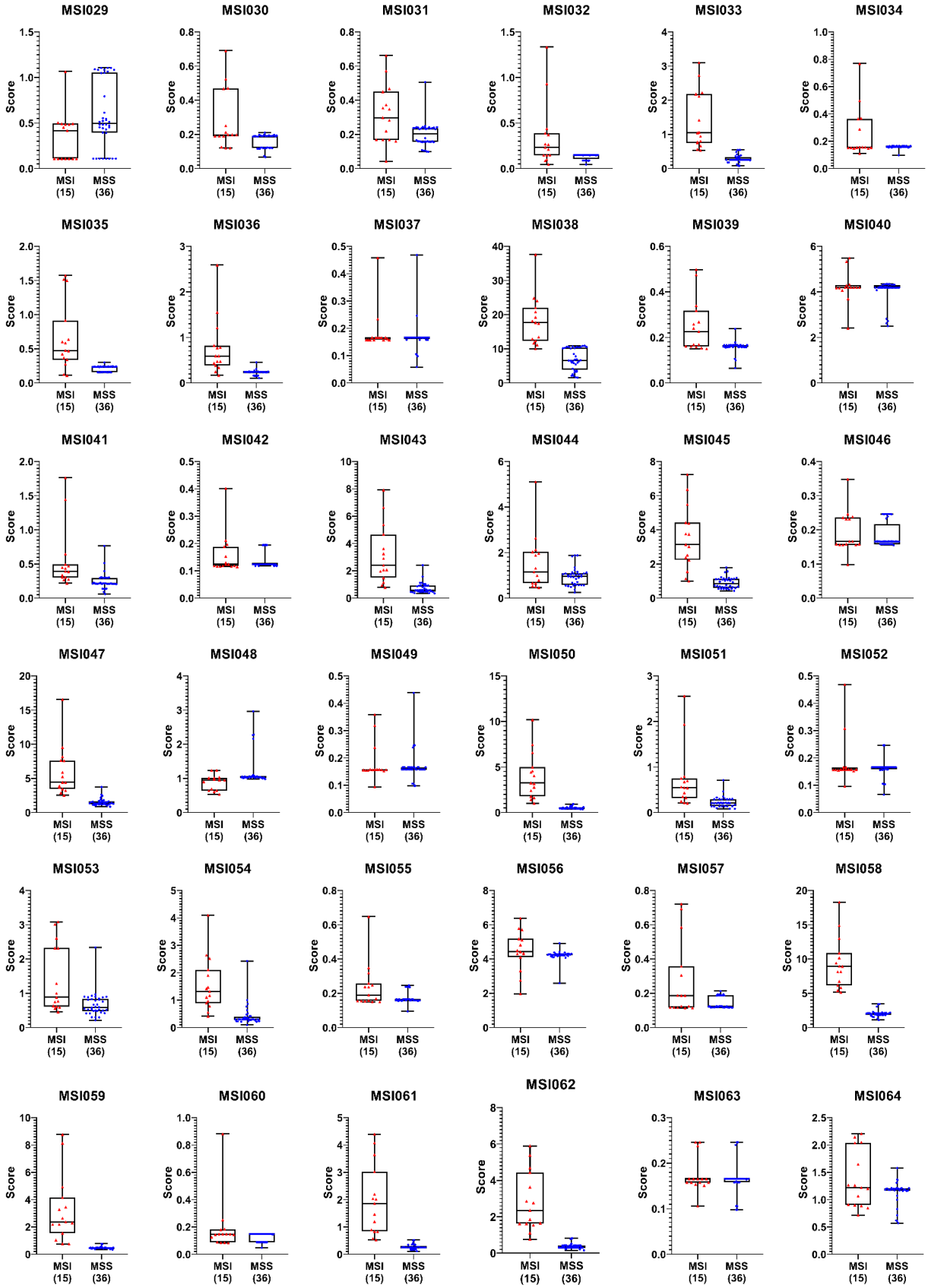

MSI029
MSI030
MSI031
MSI032
MSI033
MSI034
MSI035
MSI036
MSI037
MSI038
MSI039
MSI040
MSI041
MSI042
MSI043
MSI044
MSI045
MSI046
MSI047
MSI048
MSI049
MSI050
MSI051
MSI052
MSI053
MSI054
MSI055
MSI056
MSI057
MSI058
MSI059
MSI060
MSI061
MSI062
MSI063
MSI064
Score
MSI (12)
MSS (48)

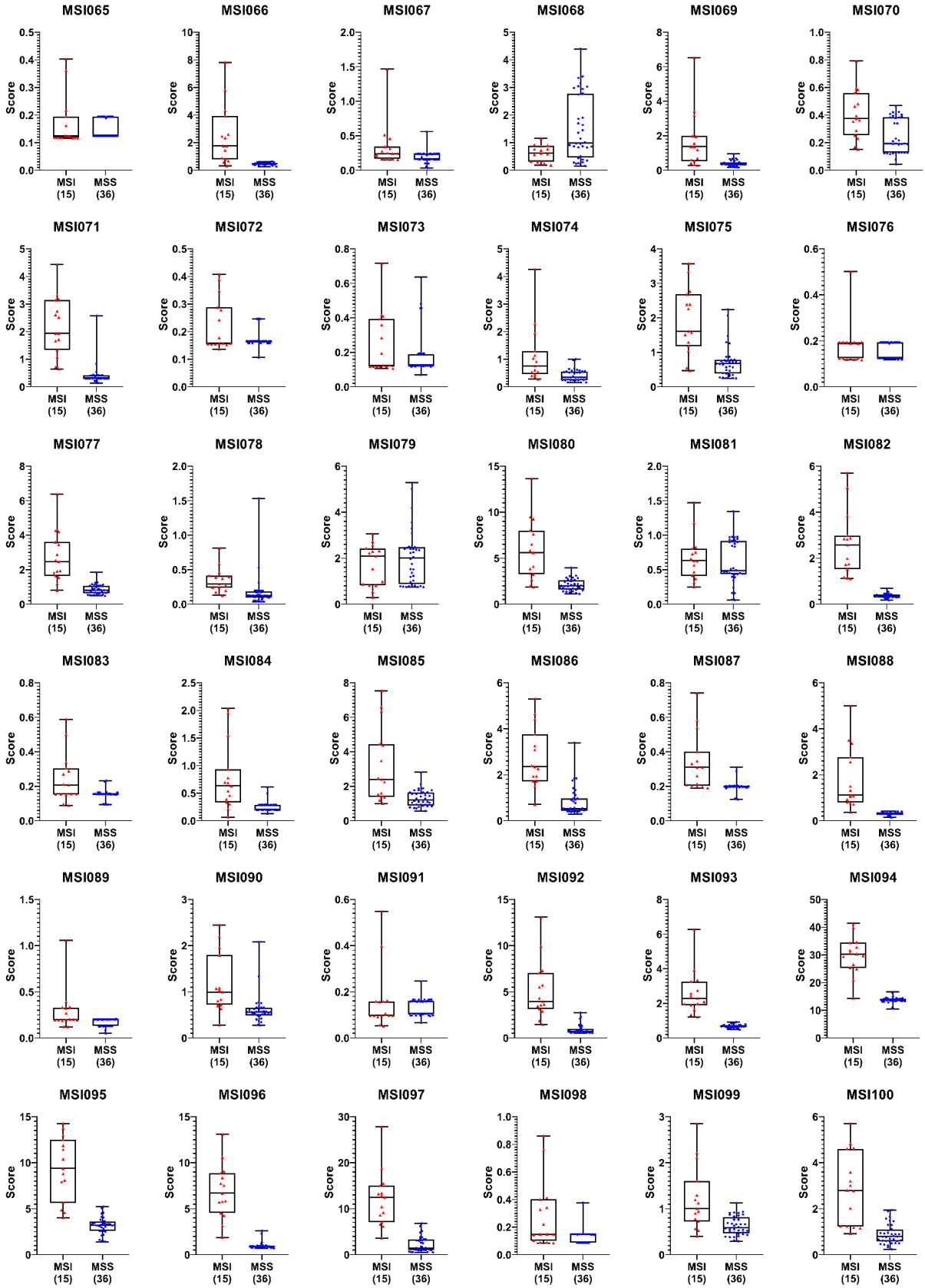

MSI065
MSI066
MSI067
MSI068
MSI069
MSI070
MSI071
MSI072
MSI073
MSI074
MSI075
MSI076
MSI077
MSI078
MSI079
MSI080
MSI081
MSI082
MSI083
MSI084
MSI085
MSI086
MSI087
MSI088
MSI089
MSI090
MSI091
MSI092
MSI093
MSI094
MSI095
MSI096
MSI097
MSI098
MSI099
MSI100
Score
MSI (12)
MSS (48)

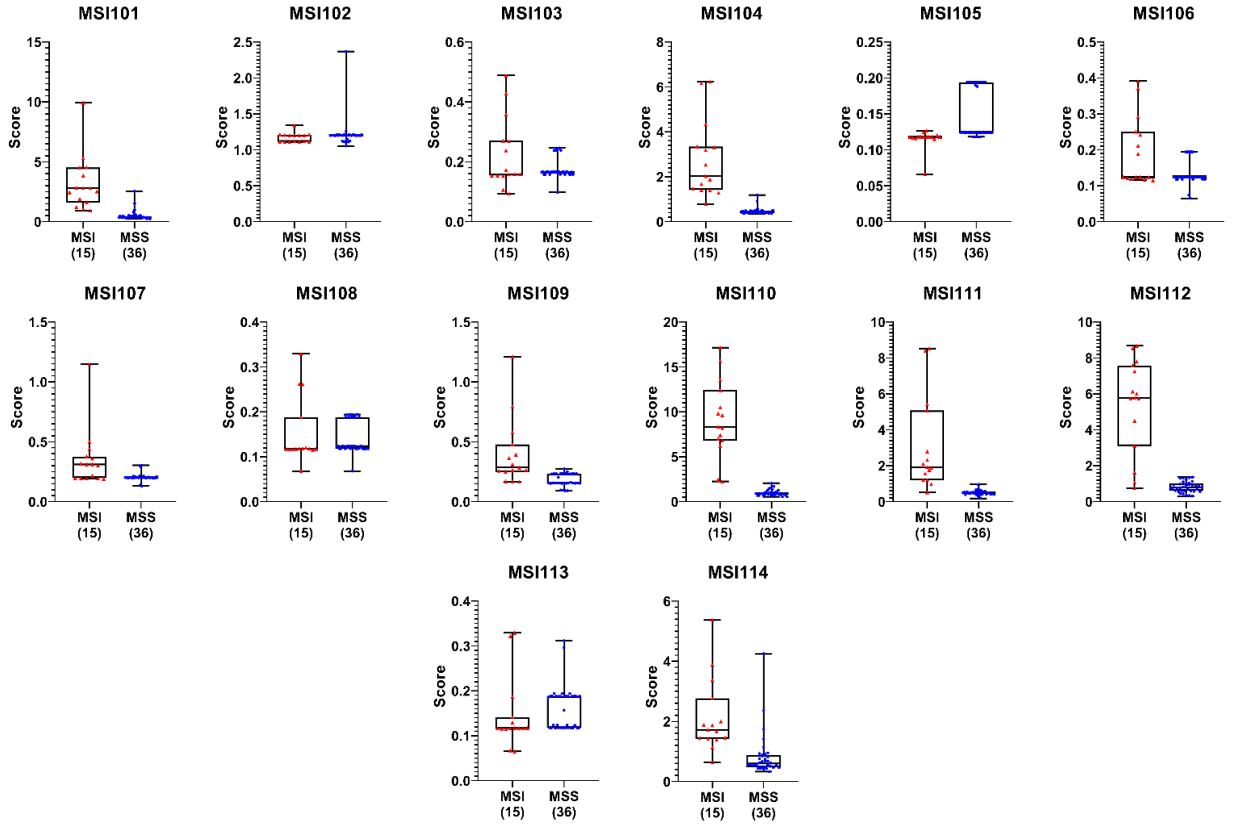



**Supplementary figure 2: Boxplots show instability scores of 122 loci in CRC cohort.** A two-sample t-test was performed separately on concordant CRC (n=60) cases to compare MSI versus MSS groups. The analysis identified 51 loci with significant differences ($p < 0.05$).

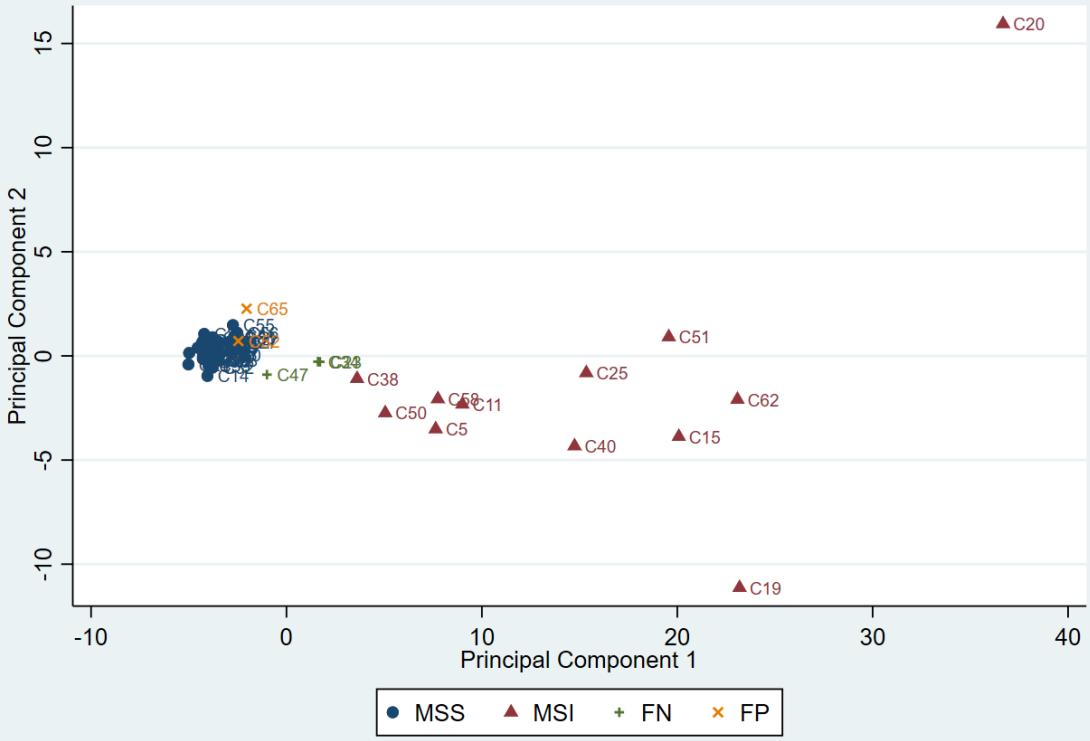

MONO27
BAT26
NR24
BAT25
CAT25
NR27
NR22
NR21
MSI001
MSI002
MSI003
MSI004
MSI005
MSI006
MSI007
MSI008
MSI009
MSI010
MSI011
MSI012
MSI013
MSI014
MSI015
MSI016
MSI017
MSI018
MSI019
MSI020
MSI021
MSI022
MSI023
MSI024
MSI025
MSI026
MSI027
MSI028
Score
MSI (15)
MSS (36)


MSI029
MSI030
MSI031
MSI032
MSI033
MSI034
MSI035
MSI036
MSI037
MSI038
MSI039
MSI040
MSI041
MSI042
MSI043
MSI044
MSI045
MSI046
MSI047
MSI048
MSI049
MSI050
MSI051
MSI052
MSI053
MSI054
MSI055
MSI056
MSI057
MSI058
MSI059
MSI060
MSI061
MSI062
MSI063
MSI064
Score
MSI (15)
MSS (36)

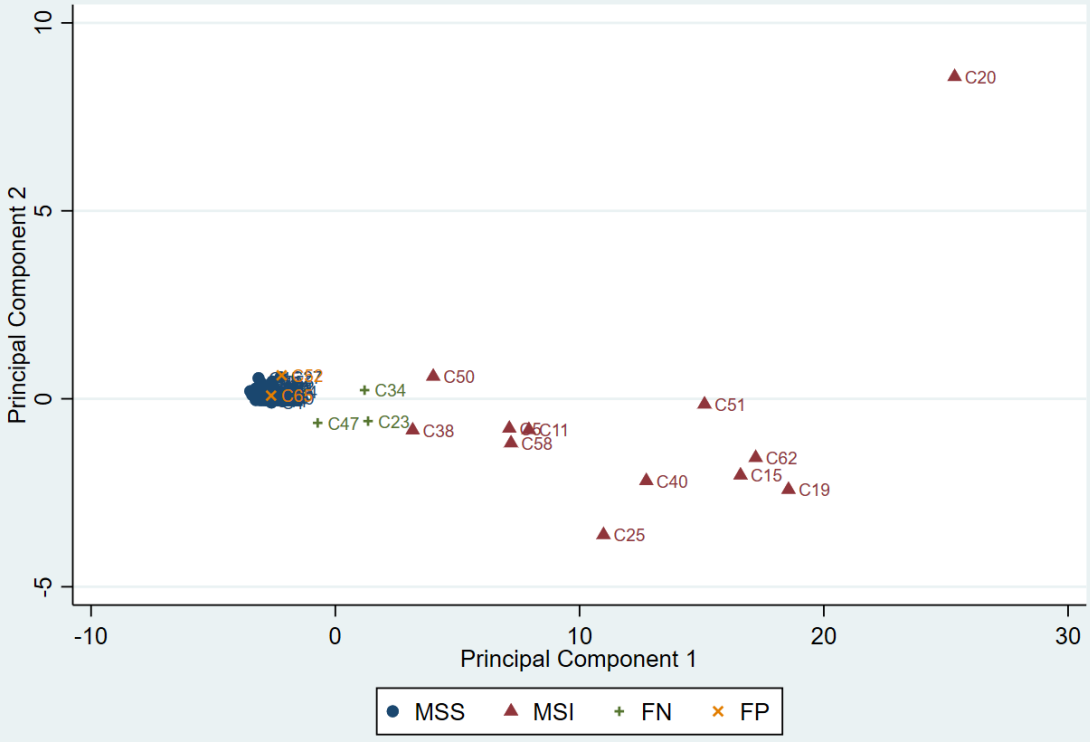

MSI065
MSI066
MSI067
MSI068
MSI069
MSI070
MSI071
MSI072
MSI073
MSI074
MSI075
MSI076
MSI077
MSI078
MSI079
MSI080
MSI081
MSI082
MSI083
MSI084
MSI085
MSI086
MSI087
MSI088
MSI089
MSI090
MSI091
MSI092
MSI093
MSI094
MSI095
MSI096
MSI097
MSI098
MSI099
MSI100
Score
MSI (15)
MSS (36)



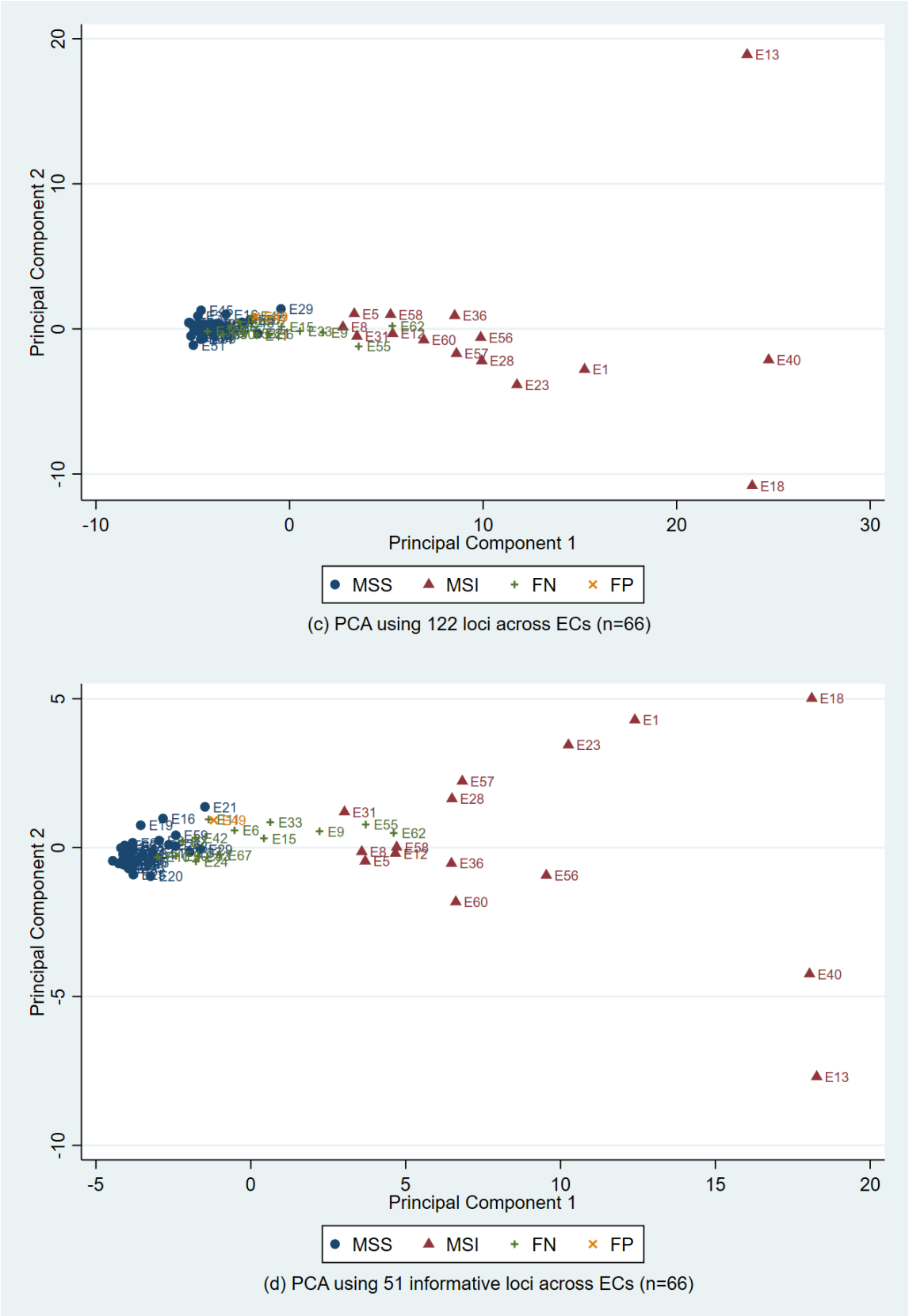

**Supplementary figure 3: Boxplots show instability scores of 122 loci in EC cohort.** A two-sample t-test was performed separately on concordant EC (n=51) cases to compare MSI versus MSS groups. This analysis identified 51 loci with significant differences ($p < 0.05$).



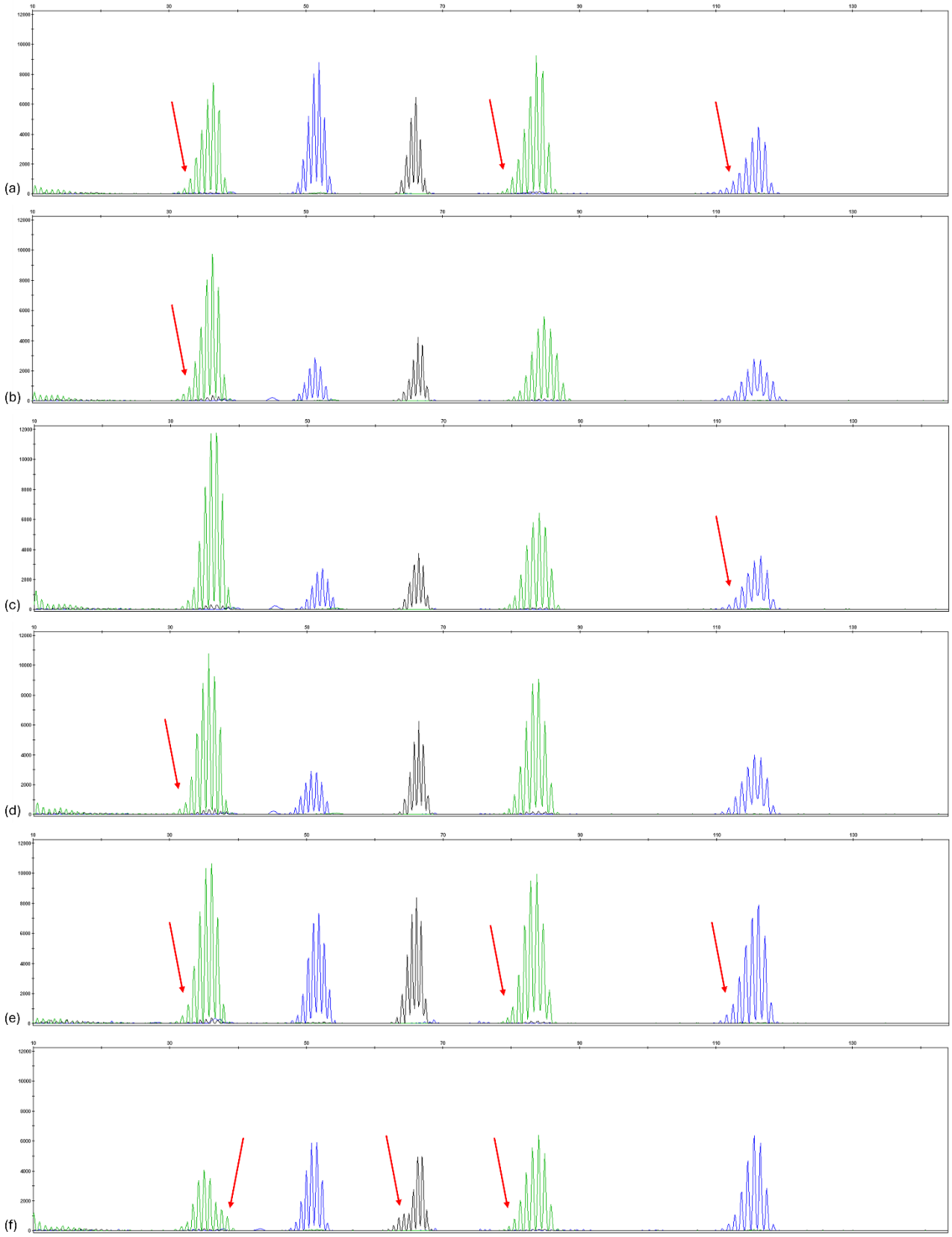

(a) PCA using 122 loci across CRCs (n=65)




(b) PCA using 51 informative loci across CRCs (n=65)




**Supplementary figure 4: Principal component analysis (PCA) of CRC and EC cohorts.** PCA was performed on the full set of 122 microsatellite loci and the refined subset of 51 informative loci to compare clustering patterns. The distribution of cases in CRC (n=65) with (a) 122 loci and (b) 51 loci was highly consistent. Similar results were obtained for EC (n=66) with (c) 122 loci and (d) 51 loci indicating that the informative loci subset retains the discriminatory power of the complete panel. The samples are labelled based on the concordance between MMR (IHC/MSI CE) and NGS testing.




**Supplementary figure 5: Electropherograms of discordant false negative cases of EC (n=6).** The fragment analysis of five standard mononucleotide repeat (MNR) loci using capillary electrophoresis of 6 FN cases in EC: (a) E7, (b) E10, (c) E24, (d) E42, (e) E50 and (f) E53, showed only minimal peak shifts. The subtle peak changes in electropherograms make the interpretation challenging, highlighting the borderline nature of MSI events in these tumours.

**Supplementary table 1A: Comparison of MSI Status by IHC, PCR (CE), and NGS in CRC samples (N=65)**

| Serial | ID | MSI CE | MMR IHC | NG101 | NG101 Unstable Loci (N=8) | Status |
|---|---|---|---|---|---|---|
| 5 | C5 | MSI | / | MSI-H | 7 | concordant |
| 11 | C11 | MSI | / | MSI-H | 6 | concordant |
| 15 | C15 | MSI | / | MSI-H | 8 | concordant |
| 19 | C19 | MSI | / | MSI-H | 8 | concordant |
| 20 | C20 | MSI | / | MSI-H | 8 | concordant |
| 25 | C25 | MSI | / | MSI-H | 7 | concordant |
| 38 | C38 | MSI | / | MSI-H | 2 | concordant |
| 40 | C40 | MSI | / | MSI-H | 8 | concordant |
| 50 | C50 | MSI | / | MSI-H | 4 | concordant |
| 51 | C51 | MSI | / | MSI-H | 8 | concordant |
| 58 | C58 | MSI | / | MSI-H | 3 | concordant |
| 62 | C62 | MSI | / | MSI-H | 8 | concordant |
| 1 | C1 | MSS | / | MSS | - | concordant |
| 2 | C2 | MSS | / | MSS | - | concordant |
| 3 | C3 | MSS | / | MSS | - | concordant |
| 6 | C6 | MSS | / | MSS | - | concordant |
| 7 | C7 | MSS | / | MSS | - | concordant |
| 8 | C8 | MSS | / | MSS | - | concordant |
| 9 | C9 | MSS | / | MSS | - | concordant |
| 10 | C10 | MSS | / | MSS | - | concordant |
| 12 | C12 | MSS | / | MSS | - | concordant |
| 13 | C13 | MSS | / | MSS | - | concordant |
| 14 | C14 | MSS | / | MSS | - | concordant |
| 16 | C16 | MSS | / | MSS | - | concordant |
| 17 | C17 | MSS | / | MSS | - | concordant |
| 18 | C18 | MSS | / | MSS | - | concordant |
| 21 | C21 | MSS (BAT-25) | / | MSI-L | BAT-25 | concordant |
| 22 | C22 | MSS | / | MSS | - | concordant |
| 24 | C24 | MSS | / | MSS | - | concordant |
| 26 | C26 | MSS | / | MSS | - | concordant |
| 28 | C28 | MSS | / | MSS | - | concordant |
| 29 | C29 | MSS | / | MSS | - | concordant |
| 30 | C30 | MSS | / | MSS | - | concordant |
| 31 | C31 | MSS | / | MSS | - | concordant |
| 32 | C32 | MSS | / | MSS | - | concordant |
| 33 | C33 | MSS | / | MSS | - | concordant |
| 35 | C35 | MSS | / | MSS | - | concordant |
| 36 | C36 | MSS | / | MSS | - | concordant |
| 37 | C37 | MSS | / | MSS | - | concordant |
| 39 | C39 | MSS | / | MSS | - | concordant |
| 41 | C41 | MSS | / | MSS | - | concordant |
| 42 | C42 | MSS | / | MSS | - | concordant |
| 43 | C43 | MSS | / | MSS | - | concordant |

| | | | | | | |
|---|---|---|---|---|---|---|
| 44 | C44 | MSS | / | MSS | - | concordant |
| 45 | C45 | MSS | / | MSS | - | concordant |
| 46 | C46 | MSS | / | MSS | - | concordant |
| 48 | C48 | MSS | / | MSS | - | concordant |
| 49 | C49 | MSS | / | MSS | - | concordant |
| 53 | C53 | / | pMMR | MSS | - | concordant |
| 56 | C56 | / | pMMR | MSS | - | concordant |
| 57 | C57 | / | pMMR | MSS | - | concordant |
| 59 | C59 | / | pMMR | MSI-L | BAT-25 | concordant |
| 61 | C61 | / | pMMR | MSI-L | BAT-25 | concordant |
| 63 | C63 | / | pMMR | MSS | - | concordant |
| 55 | C55 | MSS | pMMR | MSS | - | concordant |
| 60 | C60 | MSS | pMMR | MSS | - | concordant |
| 64 | C64 | MSS | pMMR | MSS | - | concordant |
| 4 | C4 | MSS | / | MSS | - | concordant |
| 27 | C27 | MSS | / | MSS | - | concordant |
| 54 | C54 | / | pMMR | MSS | - | concordant |
| **34** | **C34** | **MSI-H** | **/** | **MSS** | **-** | **FN** |
| **47** | **C47** | **MSI-H** | **/** | **MSS** | **-** | **FN** |
| **23** | **C23** | **MSI-H** | **MLH1-, PMS2-** | **MSI-L** | **CAT-25** | **FN** |
| **52** | **C52** | **/** | **pMMR** | **MSI-H** | **NR-21, NR-24** | **FP** |
| **65** | **C65** | **MSS** | **pMMR** | **MSI-H** | **CAT-25, NR-24** | **FP** |

**Supplementary table 1B: Comparison of MSI Status by IHC, PCR (CE), and NGS in EC samples (N=66)**

| Serial | ID | MSI CE | MMR IHC | NG101 | NG101 Unstable Loci (N=8) | Status |
|---|---|---|---|---|---|---|
| 1 | E1 | / | MLH1-, PMS2- | MSI-H | 7 | concordant |
| 5 | E5 | / | MLH1-, PMS2- | MSI-H | 3 | concordant |
| 8 | E8 | / | MLH1-, PMS2- | MSI-H | 2 | concordant |
| 12 | E12 | / | MLH1-, PMS2- | MSI-H | 2 | concordant |
| 13 | E13 | / | MLH1-, PMS2- | MSI-H | 5 | concordant |
| 28 | E28 | / | MLH1-, PMS2- | MSI-H | 2 | concordant |
| 36 | E36 | / | MLH1-, PMS2- | MSI-H | 4 | concordant |
| 40 | E40 | / | MSH2-, MSH6- | MSI-H | 5 | concordant |
| 56 | E56 | / | dMMR | MSI-H | 6 | concordant |
| 57 | E57 | / | dMMR | MSI-H | 4 | concordant |
| 58 | E58 | / | MLH1-, PMS2- | MSI-H | 3 | concordant |
| 60 | E60 | / | MLH1-, PMS2- | MSI-H | 3 | concordant |
| 18 | E18 | MSI-H | MSH2-, MSH6- | MSI-H | 8 | concordant |
| 23 | E23 | MSI-H | pMMR | MSI-H | 2 | concordant |
| 31 | E31 | MSI-H | MSH6- | MSI-H | 3 | concordant |
| 3 | E3 | / | pMMR | MSS | - | concordant |
| 4 | E4 | / | pMMR | MSS | - | concordant |
| 14 | E14 | / | pMMR | MSS | - | concordant |
| 17 | E17 | / | pMMR | MSS | - | concordant |
| 19 | E19 | MSS | / | MSS | - | concordant |
| 20 | E20 | MSS | / | MSS | - | concordant |
| 22 | E22 | / | pMMR | MSS | - | concordant |
| 25 | E25 | / | pMMR | MSS | - | concordant |
| 26 | E26 | MSS | / | MSS | - | concordant |
| 30 | E30 | / | pMMR | MSS | - | concordant |
| 32 | E32 | / | pMMR | MSS | - | concordant |
| 34 | E34 | / | pMMR | MSS | - | concordant |
| 35 | E35 | / | pMMR | MSS | - | concordant |
| 38 | E38 | / | pMMR | MSS | - | concordant |
| 39 | E39 | / | pMMR | MSS | - | concordant |
| 41 | E41 | / | pMMR | MSS | - | concordant |
| 43 | E43 | / | pMMR | MSS | - | concordant |
| 44 | E44 | / | pMMR | MSS | - | concordant |
| 45 | E45 | / | pMMR | MSS | - | concordant |
| 46 | E46 | / | pMMR | MSS | - | concordant |
| 47 | E47 | / | pMMR | MSS | - | concordant |
| 52 | E52 | / | pMMR | MSS | - | concordant |
| 54 | E54 | / | pMMR | MSS | - | concordant |
| 61 | E61 | / | pMMR | MSS | - | concordant |
| 63 | E63 | / | pMMR | MSS | - | concordant |
| 65 | E65 | / | pMMR | MSS | - | concordant |
| 16 | E16 | MSS | / | MSS | - | concordant |
| 21 | E21 | MSS | / | MSS | - | concordant |
| 27 | E27 | MSS | pMMR | MSS | - | concordant |

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | E2 | MSS | pMMR | MSS | - | concordant |
| 29 | E29 | MSS | MSH2 - | MSS | - | concordant |
| 37 | E37 | MSS | MLH1-, PMS2- | MSS | - | concordant |
| 48 | E48 | MSS | pMMR | MSS | - | concordant |
| 51 | E51 | MSS | pMMR | MSS | - | concordant |
| 64 | E64 | MSS | MLH1- PMS2- focale | MSS | - | concordant |
| 59 | E59 | MSS | pMMR | MSI-L | BAT-25 | concordant |
| **9** | **E9** | **/** | **MLH1-, PMS2-** | **MSS** | **-** | **FN** |
| **15** | **E15** | **/** | **MLH1-, PMS2-** | **MSS** | **-** | **FN** |
| **33** | **E33** | **/** | **MLH1-, PMS2-** | **MSS** | **-** | **FN** |
| **62** | **E62** | **MSI** | **/** | **MSI-L** | **BAT-25** | **FN** |
| **53** | **E53** | **MSI-H** | **/** | **MSI-L** | **CAT-25** | **FN** |
| **7** | **E7** | **MSI-H** | **MLH1-, PMS2-** | **MSS** | **-** | **FN** |
| **10** | **E10** | **MSI-H** | **MLH1-, PMS2-** | **MSS** | **-** | **FN** |
| **55** | **E55** | **MSI-H** | **pMMR** | **MSI-L** | **NR-22** | **FN** |
| **6** | **E6** | **MSI-H** | **MLH1-, PMS2-** | **MSI-L** | **CAT-25** | **FN** |
| **24** | **E24** | **MSI-H** | **MLH1-, PMS2-** | **MSS** | **-** | **FN** |
| **50** | **E50** | **MSI-H** | **MLH1-, PMS2-** | **MSS** | **-** | **FN** |
| **11** | **E11** | **MSI-H** | **pMMR** | **MSS** | **-** | **FN** |
| **42** | **E42** | **MSI-H** | **MLH1-, PMS2-** | **MSS** | **-** | **FN** |
| **66** | **E67** | **MSI** | **MLH1-** | **MSS** | **-** | **FN** |
| **49** | **E49** | **MSS** | **pMMR** | **MSI-H** | **NR-21, NR-27** | **FP** |

**Supplementary table 2A: List of genes in NGS panel.**

| | Gene Name | SNVs/InDel | Exons | CNV | Fusion | Transcript |
|---|---|---|---|---|---|---|
| **1** | **AKT1** | X | 4, 5 | | | NM_001014431.2 |
| **2** | **ALK** | X | All (1-29) | | X | NM_004304.5 |
| **3** | **AR** | X | All (1-8) | | | NM_001011645.3 |
| **4** | **BRAF** | X | 11, 12, 15, 18 | | | NM_004333.4 |
| **5** | **CDK4** | X | 2 | | | NM_000075.2 |
| **6** | **CDKN2A** | X | All (1-2) | | | Più trascritti |
| **7** | **CTNNB1** | X | 3 | | | NM_001904.4 |
| **8** | **DDR2** | X | 5, 8, 13, 14, 15, 16, 17, 18 | | | NM_006182.4 |
| **9** | **EGFR** | X | All (1-28) | X | | NM_005228.5 |
| **10** | **ERBB2** | X | All (1-27) | X | | NM_004448.4 |
| **11** | **ERBB3** | X | All (1-28) | | | NM_001982.4 |
| **12** | **ERBB4** | X | 10, 12 | | | NM_005235.3 |
| **13** | **ESR1** | X | 5, 8 | | | NM_001122740.2 |
| **14** | **FGFR1** | X | All (2-18) | | X | NM_001174067.2 |
| **15** | **FGFR2** | X | All (2-18) | | X | NM_000141.5 |
| **16** | **FGFR3** | X | All (2-18) | | X | NM_000142.5 |
| **17** | **FGFR4** | X | 3, 6, 9, 10, 12, 13, 15, 16 | | | NM_213647.3 |
| **18** | **GNA11** | X | 4, 5 | | | NM_002067.5 |
| **19** | **GNAQ** | X | 4, 5 | | | NM_002072.5 |
| **20** | **GNAS** | X | 8 | | | NM_016592 |
| **21** | **HRAS** | X | 2, 3, 4 | | | NM_005343.4 |
| **22** | **IDH1** | X | 4 | | | NM_005896.4 |
| **23** | **IDH2** | X | 4 | | | NM_002168.4 |
| **24** | **KEAP1** | X | All (2-6) | | | NM_203500.2 |
| **25** | **KIT** | X | 8, 9, 10, 11, 13, 14, 17, 18 | | | NM_000222.3 |
| **26** | **KRAS** | X | 2, 3, 4 | | | NM_004985.5 |
| **27** | **MAP2K1 (MEK1)** | X | 2, 3, 6 | | | NM_002755.3 |
| **28** | **MET** | X | 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21 | X | X | NM_000245.4 |
| **29** | **MTOR** | X | All (2-58) | | | NM_004958.4 |
| **30** | **NF1** | X | All (1-58) | | | NM_001042492.3 |
| **31** | **NRAS** | X | 2, 3, 4 | | | NM_002524.5 |
| **32** | **NRG1** | | All (1-13) | | X | NM_013956.5 |
| **33** | **NTRK1** | X | All (1-17) | | X | NM_001007792.1 |
| **34** | **NTRK2** | X | All (2-19) | | X | NM_006180.5 |
| **35** | **NTRK3** | X | All (3-20) | | X | NM_001012338.3 |
| **36** | **PDGFRA** | X | 12, 14, 18 | | | NM_006206.6 |
| **37** | **PIK3CA** | X | 2, 3, 5, 6, 8, 9, 10, 14, 21 | | | NM_006218.4 |
| **38** | **POLE** | X | 1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, | | | NM_006231.4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | 19 | | | |
| **39** | **PPARG** | | All (3-8) | | X | NM_138712.5 |
| **40** | **PTEN** | X | 1, 2, 3, 4, 5, 6, 7, 8 | | | NM_000314.8 |
| **41** | **RAF1** | X | All (2-17) | | | NM_001354689.3 |
| **42** | **RB1** | X | All (1-27) | | | NM_000321.3 |
| **43** | **RET** | X | All (1-20) | | X | NM_020975.6 |
| **44** | **ROS1** | X | All (1-43) | | X | NM_002944.3 |
| **45** | **SMAD4** | X | All (2-12) | | | NM_005359.6 |
| **46** | **SMO** | X | All (1-12) | | | NM_005631.5 |
| **47** | **STK11** | X | All (1-9) | | | NM_000455.5 |
| **48** | **TERT** | X | Promoter, 1, 8, 9, 13 | | | NM_198253.3 |
| **49** | **TP53** | X | All (2-11) | | | NM_000546.6 |
| **50** | **TSC1** | X | All (3-23) | | | NM_000368.5 |

**Supplementary table 2B: List of MSI loci in NGS panel.**

| MSI Primary Loci | | | | | |
|---|---|---|---|---|---|
| **Locus Name** | **chr** | **start** | **end** | **Nucleotide** | **Repetitions** |
| BAT25 | chr4 | 55598212 | 55598236 | T | 25 |
| BAT26 | chr2 | 47641560 | 47641586 | T | 26 |
| CAT25 | chr7 | 143003343 | 143003367 | T | 25 |
| MONO27 | chr2 | 39573063 | 39573089 | T | 27 |
| NR21 | chr14 | 23652347 | 23652367 | T | 21 |
| NR22 | chr11 | 125490766 | 125490786 | T | 22 |
| NR24 | chr2 | 95849362 | 95849384 | T | 24 |
| NR27 | chr11 | 102193509 | 102193534 | T | 27 |

| MSI Secondary Loci | | | | | |
|---|---|---|---|---|---|
| **Locus Name** | **chr** | **start** | **end** | **Nucleotide** | **Repetitions** |
| MSI-001 | chr1 | 6257785 | 6257792 | T | 8 |
| MSI-002 | chr1 | 16572709 | 16572722 | T | 14 |
| MSI-003 | chr1 | 25685999 | 25686012 | A | 14 |
| MSI-004 | chr1 | 27621108 | 27621115 | G | 8 |
| MSI-005 | chr1 | 28618053 | 28618066 | A | 14 |
| MSI-006 | chr1 | 33402335 | 33402351 | A | 17 |
| MSI-007 | chr1 | 47325300 | 47325312 | A | 13 |
| MSI-008 | chr1 | 89449509 | 89449519 | T | 11 |
| MSI-009 | chr1 | 145002892 | 145002910 | A | 19 |
| MSI-010 | chr1 | 145260856 | 145260871 | T | 16 |
| MSI-011 | chr1 | 149900986 | 149901001 | A | 16 |
| MSI-012 | chr1 | 159290623 | 159290636 | T | 14 |
| MSI-013 | chr1 | 165793509 | 165793523 | T | 15 |
| MSI-014 | chr1 | 170733810 | 170733831 | T | 22 |
| MSI-015 | chr1 | 180075942 | 180075955 | T | 14 |
| MSI-016 | chr1 | 201754411 | 201754427 | T | 17 |
| MSI-017 | chr1 | 204084143 | 204084157 | T | 15 |
| MSI-018 | chr1 | 220318426 | 220318440 | A | 15 |
| MSI-019 | chr1 | 227618722 | 227618736 | A | 15 |
| MSI-020 | chr1 | 235486356 | 235486372 | T | 17 |
| MSI-021 | chr2 | 39564894 | 39564921 | T | 28 |
| MSI-022 | chr2 | 48032741 | 48032753 | T | 13 |
| MSI-023 | chr2 | 62063094 | 62063110 | A | 17 |
| MSI-024 | chr2 | 97309469 | 97309484 | T | 16 |
| MSI-025 | chr2 | 148683686 | 148683693 | A | 8 |
| MSI-026 | chr2 | 165551296 | 165551304 | A | 9 |
| MSI-027 | chr2 | 203922058 | 203922066 | A | 9 |
| MSI-028 | chr2 | 207174428 | 207174436 | A | 9 |
| MSI-029 | chr2 | 242743805 | 242743810 | C | 6 |
| MSI-030 | chr3 | 30691872 | 30691881 | A | 10 |
| MSI-031 | chr3 | 37030781 | 37030791 | T | 11 |
| MSI-032 | chr3 | 51417604 | 51417610 | C | 7 |

| | | | | | |
|---|---|---|---|---|---|
| MSI-033 | chr3 | 71008342 | 71008354 | T | 13 |
| MSI-034 | chr3 | 100039736 | 100039744 | A | 9 |
| MSI-035 | chr3 | 113377482 | 113377492 | T | 11 |
| MSI-036 | chr3 | 130733047 | 130733057 | T | 11 |
| MSI-037 | chr3 | 157081227 | 157081235 | T | 9 |
| MSI-038 | chr4 | 39501723 | 39501740 | A | 18 |
| MSI-039 | chr4 | 83785565 | 83785573 | T | 9 |
| MSI-040 | chr5 | 42495180 | 42495189 | C | 10 |
| MSI-041 | chr5 | 67584513 | 67584524 | T | 12 |
| MSI-042 | chr5 | 79970915 | 79970922 | A | 8 |
| MSI-043 | chr5 | 172421761 | 172421775 | T | 15 |
| MSI-044 | chr6 | 11714640 | 11714653 | A | 14 |
| MSI-045 | chr6 | 43021977 | 43021988 | G | 12 |
| MSI-046 | chr6 | 100382358 | 100382366 | A | 9 |
| MSI-047 | chr6 | 142691951 | 142691967 | T | 17 |
| MSI-048 | chr7 | 17812489 | 17812496 | C | 8 |
| MSI-049 | chr7 | 77423460 | 77423468 | T | 9 |
| MSI-050 | chr8 | 17665798 | 17665812 | T | 15 |
| MSI-051 | chr8 | 23712067 | 23712078 | T | 12 |
| MSI-052 | chr8 | 37791834 | 37791842 | T | 9 |
| MSI-053 | chr8 | 79629739 | 79629752 | A | 14 |
| MSI-054 | chr8 | 134485099 | 134485112 | A | 14 |
| MSI-055 | chr9 | 33675365 | 33675373 | A | 9 |
| MSI-056 | chr9 | 121898658 | 121898668 | G | 11 |
| MSI-057 | chr9 | 136918529 | 136918536 | G | 8 |
| MSI-058 | chr10 | 12381073 | 12381088 | A | 16 |
| MSI-059 | chr10 | 12424938 | 12424952 | T | 15 |
| MSI-060 | chr10 | 29760116 | 29760122 | C | 7 |
| MSI-061 | chr10 | 75630925 | 75630937 | A | 13 |
| MSI-062 | chr10 | 79635929 | 79635942 | A | 14 |
| MSI-063 | chr10 | 97918856 | 97918864 | A | 9 |
| MSI-064 | chr11 | 1490093 | 1490102 | G | 10 |
| MSI-065 | chr11 | 62649529 | 62649536 | A | 8 |
| MSI-066 | chr11 | 62703438 | 62703452 | T | 15 |
| MSI-067 | chr11 | 63149671 | 63149681 | A | 11 |
| MSI-068 | chr11 | 93170910 | 93170918 | C | 9 |
| MSI-069 | chr11 | 94151548 | 94151560 | T | 13 |
| MSI-070 | chr11 | 112832277 | 112832286 | G | 10 |
| MSI-071 | chr11 | 115047033 | 115047046 | T | 14 |
| MSI-072 | chr11 | 118220583 | 118220591 | A | 9 |
| MSI-073 | chr11 | 119545225 | 119545232 | G | 8 |
| MSI-074 | chr11 | 125763611 | 125763623 | T | 13 |
| MSI-075 | chr12 | 54405180 | 54405190 | C | 11 |
| MSI-076 | chr12 | 55759486 | 55759493 | T | 8 |
| MSI-077 | chr12 | 85285921 | 85285937 | A | 17 |
| MSI-078 | chr12 | 122242658 | 122242665 | C | 8 |

| | | | | | |
|---|---|---|---|---|---|
| MSI-079 | chr12 | 130785787 | 130785796 | G | 10 |
| MSI-080 | chr13 | 31722621 | 31722637 | A | 17 |
| MSI-081 | chr13 | 58299435 | 58299445 | A | 11 |
| MSI-082 | chr13 | 113965003 | 113965016 | T | 14 |
| MSI-083 | chr14 | 24040436 | 24040444 | C | 9 |
| MSI-084 | chr14 | 53513440 | 53513451 | A | 12 |
| MSI-085 | chr14 | 73959704 | 73959719 | T | 16 |
| MSI-086 | chr14 | 74054709 | 74054723 | T | 15 |
| MSI-087 | chr14 | 93708031 | 93708040 | A | 10 |
| MSI-088 | chr14 | 102821856 | 102821869 | A | 14 |
| MSI-089 | chr15 | 34157542 | 34157551 | A | 10 |
| MSI-090 | chr15 | 45897772 | 45897785 | T | 14 |
| MSI-091 | chr16 | 10867203 | 10867211 | A | 9 |
| MSI-092 | chr16 | 11051908 | 11051923 | A | 16 |
| MSI-093 | chr16 | 22650153 | 22650168 | A | 16 |
| MSI-094 | chr17 | 13981240 | 13981260 | T | 21 |
| MSI-095 | chr17 | 19314918 | 19314935 | T | 18 |
| MSI-096 | chr17 | 27081793 | 27081808 | T | 16 |
| MSI-097 | chr17 | 52991732 | 52991747 | A | 16 |
| MSI-098 | chr17 | 56435161 | 56435167 | C | 7 |
| MSI-099 | chr17 | 74688718 | 74688732 | A | 15 |
| MSI-100 | chr18 | 511981 | 511994 | T | 14 |
| MSI-101 | chr18 | 649880 | 649894 | T | 15 |
| MSI-102 | chr18 | 52445092 | 52445099 | C | 8 |
| MSI-103 | chr18 | 57013194 | 57013202 | T | 9 |
| MSI-104 | chr19 | 14104689 | 14104702 | T | 14 |
| MSI-105 | chr19 | 49458971 | 49458978 | G | 8 |
| MSI-106 | chr19 | 49850473 | 49850480 | G | 8 |
| MSI-107 | chr20 | 46286321 | 46286330 | A | 10 |
| MSI-108 | chr20 | 47858504 | 47858511 | A | 8 |
| MSI-109 | chr20 | 61536692 | 61536702 | A | 11 |
| MSI-110 | chr22 | 26352179 | 26352194 | T | 16 |
| MSI-111 | chr22 | 29696469 | 29696484 | T | 16 |
| MSI-112 | chr22 | 37401595 | 37401609 | T | 15 |
| MSI-113 | chrX | 37312611 | 37312618 | C | 8 |
| MSI-114 | chrX | 101409255 | 101409270 | T | 16 |

**Supplementary table 3: Results of *t*-test performed for complete set of 122 loci across CRC and EC.**

| Serial | Locus Name | Chromosome | Start Position (hg19) | Reference Repeat Motif and Length | Gene Association | Locus Category | P-value (CRC) | P-value (EC) |
|---|---|---|---|---|---|---|---|---|
| 1 | MONO-27 | Chr 2 | 39573063 | (T)27 | MAP4K3 | Standard | <0.0001 | <0.0001 |
| 2 | BAT-26 | Chr 2 | 47641560 | (T)26 | MSH2 | Standard | <0.0001 | <0.0001 |
| 3 | NR-24 | Chr 2 | 95849362 | (T)24 | ZNF2 | Standard | <0.0001 | <0.0001 |
| 4 | BAT-25 | Chr 4 | 55598212 | (T)25 | KIT | Standard | 0.0001 | 0.0001 |
| 5 | CAT-25 | Chr 7 | 143003343 | (T)25 | CASP2 | Standard | <0.0001 | <0.0001 |
| 6 | NR-27 | Chr 11 | 102193509 | (T)27 | BIRC3 | Standard | <0.0001 | <0.0001 |
| 7 | NR-22 | Chr 11 | 125490766 | (T)22 | STT3A | Standard | <0.0001 | <0.0001 |
| 8 | NR-21 | Chr 14 | 23652347 | (T)21 | SLC7A8 | Standard | <0.0001 | 0.013 |
| 9 | MSI-001 | Chr 1 | 6257785 | (T)8 | RPL22 | Non-standard | 0.0007 | <0.0001 |
| 10 | MSI-002 | Chr 1 | 16572709 | (T)14 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 11 | MSI-003 | Chr 1 | 25685999 | (A)14 | TMEM50A | Non-standard | <0.0001 | 0.0631 |
| 12 | MSI-004 | Chr 1 | 27621108 | (G)8 | WDTC1 | Non-standard | 0.0017 | 0.0292 |
| 13 | MSI-005 | Chr 1 | 28618053 | (A)14 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 14 | MSI-006 | Chr 1 | 33402335 | (A)17 | RNF19B | Non-standard | <0.0001 | <0.0001 |
| 15 | MSI-007 | Chr 1 | 47325300 | (A)13 | CYP4Z2P | Non-standard | <0.0001 | <0.0001 |
| 16 | MSI-008 | Chr 1 | 89449509 | (T)11 | RBMXL1 | Non-standard | <0.0001 | <0.0001 |
| 17 | MSI-009 | Chr 1 | 145002892 | (A)19 | PDE4DIP | Non-standard | <0.0001 | <0.0001 |
| 18 | MSI-010 | Chr 1 | 145260856 | (T)16 | NOTCH2NLR | Non-standard | <0.0001 | <0.0001 |
| 19 | MSI-011 | Chr 1 | 149900986 | (A)16 | MTMR11 | Non-standard | <0.0001 | <0.0001 |
| 20 | MSI-012 | Chr 1 | 159290623 | (T)14 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 21 | MSI-013 | Chr 1 | 165793509 | (T)15 | TMCO1 | Non-standard | <0.0001 | <0.0001 |
| 22 | MSI-014 | Chr 1 | 170733810 | (T)22 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 23 | MSI-015 | Chr 1 | 180075942 | (T)14 | CEP350 | Non-standard | <0.0001 | <0.0001 |
| 24 | MSI-016 | Chr 1 | 201754411 | (T)17 | NAV1 | Non-standard | <0.0001 | <0.0001 |
| 25 | MSI-017 | Chr 1 | 204084143 | (T)15 | SOX13 | Non-standard | <0.0001 | <0.0001 |
| 26 | MSI-018 | Chr 1 | 220318426 | (A)15 | IARS2 | Non-standard | <0.0001 | <0.0001 |
| 27 | MSI-019 | Chr 1 | 227618722 | (A)15 | LINC01641 | Non-standard | <0.0001 | 0.0014 |
| 28 | MSI-020 | Chr 1 | 235486356 | (T)17 | ARID4B | Non-standard | <0.0001 | <0.0001 |
| 29 | MSI-021 | Chr 2 | 39564894 | (T)28 | MAP4K3 | Non-standard | <0.0001 | 0.0007 |
| 30 | MSI-022 | Chr 2 | 48032741 | (T)13 | MSH6 | Non-standard | <0.0001 | 0.0002 |
| 31 | MSI-023 | Chr 2 | 62063094 | (A)17 | FAM161A | Non-standard | <0.0001 | <0.0001 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 32 | MSI-024 | Chr 2 | 97309469 | (T)16 | FER1L5 | Non-standard | <0.0001 | <0.0001 |
| 33 | MSI-025 | Chr 2 | 148683686 | (A)8 | ACVR2A | Non-standard | <0.0001 | 0.003 |
| 34 | MSI-026 | Chr 2 | 165551296 | (A)9 | COBLL1 | Non-standard | 0.0002 | 0.0296 |
| 35 | MSI-027 | Chr 2 | 203922058 | (A)9 | NBEAL1 | Non-standard | <0.0001 | 0.0046 |
| 36 | MSI-028 | Chr 2 | 207174428 | (A)9 | ZDBF2 | Non-standard | 0.3442 | 0.0218 |
| 37 | MSI-029 | Chr 2 | 242743805 | (C)6 | No direct gene | Non-standard | 0.3163 | 0.029 |
| 38 | MSI-030 | Chr 3 | 30691872 | (A)10 | TGFBR2 | Non-standard | <0.0001 | 0.0004 |
| 39 | MSI-031 | Chr 3 | 37030781 | (T)11 | EPM2AIP1 | Non-standard | <0.0001 | 0.0011 |
| 40 | MSI-032 | Chr 3 | 51417604 | (C)7 | DOCK3 | Non-standard | <0.0001 | 0.0008 |
| 41 | MSI-033 | Chr 3 | 71008342 | (T)13 | FOXP1 | Non-standard | <0.0001 | <0.0001 |
| 42 | MSI-034 | Chr 3 | 100039736 | (A)9 | TBC1D23 | Non-standard | 0.0031 | 0.0031 |
| 43 | MSI-035 | Chr 3 | 113377482 | (T)11 | USF3 | Non-standard | <0.0001 | <0.0001 |
| 44 | MSI-036 | Chr 3 | 130733047 | (T)11 | ASTE1 | Non-standard | <0.0001 | <0.0001 |
| 45 | MSI-037 | Chr 3 | 157081227 | (T)9 | VEPH1 | Non-standard | 0.0103 | 0.3411 |
| 46 | MSI-038 | Chr 4 | 39501723 | (A)18 | UGDH | Non-standard | <0.0001 | <0.0001 |
| 47 | MSI-039 | Chr 4 | 83785565 | (T)9 | SEC31A | Non-standard | 0.0001 | <0.0001 |
| 48 | MSI-040 | Chr 5 | 42495180 | (C)10 | GHR | Non-standard | 0.4094 | 0.3925 |
| 49 | MSI-041 | Chr 5 | 67584513 | (T)12 | PIK3R1 | Non-standard | <0.0001 | 0.0004 |
| 50 | MSI-042 | Chr 5 | 79970915 | (A)8 | MSH3 | Non-standard | <0.0001 | 0.1404 |
| 51 | MSI-043 | Chr 5 | 172421761 | (T)15 | ATP6V0E1 | Non-standard | <0.0001 | <0.0001 |
| 52 | MSI-044 | Chr 6 | 11714640 | (A)14 | ASTRP | Non-standard | <0.0001 | 0.0117 |
| 53 | MSI-045 | Chr 6 | 43021977 | (G)12 | MRPL2 | Non-standard | <0.0001 | <0.0001 |
| 54 | MSI-046 | Chr 6 | 100382358 | (A)9 | MCHR2 | Non-standard | 0.0036 | 0.4494 |
| 55 | MSI-047 | Chr 6 | 142691951 | (T)17 | ADGRG6 | Non-standard | <0.0001 | <0.0001 |
| 56 | MSI-048 | Chr 7 | 17812489 | (C)8 | No direct gene | Non-standard | 0.016 | 0.0207 |
| 57 | MSI-049 | Chr 7 | 77423460 | (T)9 | TMEM60 | Non-standard | 0.0286 | 0.4984 |
| 58 | MSI-050 | Chr 8 | 17665798 | (T)15 | MTUS1-DT | Non-standard | <0.0001 | <0.0001 |
| 59 | MSI-051 | Chr 8 | 23712067 | (T)12 | STC1 | Non-standard | <0.0001 | <0.0001 |
| 60 | MSI-052 | Chr 8 | 37791834 | (T)9 | GOT1L1 | Non-standard | 0.1941 | 0.1635 |
| 61 | MSI-053 | Chr 8 | 79629739 | (A)14 | ZC2HC1A | Non-standard | <0.0001 | 0.0003 |
| 62 | MSI-054 | Chr 8 | 134485099 | (A)14 | ST3GAL1 | Non-standard | <0.0001 | <0.0001 |
| 63 | MSI-055 | Chr 9 | 33675365 | (A)9 | PTENP1 | Non-standard | <0.0001 | 0.0058 |
| 64 | MSI-056 | Chr 9 | 121898658 | (G)11 | No direct gene | Non-standard | 0.0085 | 0.2465 |
| 65 | MSI-057 | Chr 9 | 136918529 | (G)8 | BRD3 | Non-standard | 0.0008 | 0.0009 |
| 66 | MSI-058 | Chr 10 | 12381073 | (A)16 | No direct gene | Non-standard | <0.0001 | <0.0001 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 67 | MSI-059 | Chr 10 | 12424938 | (T)15 | CAMK1D | Non-standard | <0.0001 | <0.0001 |
| 68 | MSI-060 | Chr 10 | 29760116 | (C)7 | SVIL | Non-standard | <0.0001 | 0.0673 |
| 69 | MSI-061 | Chr 10 | 75630925 | (A)13 | CAMK2G | Non-standard | <0.0001 | <0.0001 |
| 70 | MSI-062 | Chr 10 | 79635929 | (A)14 | DLG5 | Non-standard | <0.0001 | <0.0001 |
| 71 | MSI-063 | Chr 10 | 97918856 | (A)9 | ZNF518A | Non-standard | 0.0003 | 0.3689 |
| 72 | MSI-064 | Chr 11 | 1490093 | (G)10 | No direct gene | Non-standard | 0.1438 | 0.0313 |
| 73 | MSI-065 | Chr 11 | 62649529 | (A)8 | SLC3A2 | Non-standard | 0.4562 | 0.3153 |
| 74 | MSI-066 | Chr 11 | 62703438 | (T)15 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 75 | MSI-067 | Chr 11 | 63149671 | (A)11 | SLC22A9 | Non-standard | <0.0001 | 0.011 |
| 76 | MSI-068 | Chr 11 | 93170910 | (C)9 | DEUP1 | Non-standard | 0.0468 | 0.0059 |
| 77 | MSI-069 | Chr 11 | 94151548 | (T)13 | MRE11 | Non-standard | <0.0001 | <0.0001 |
| 78 | MSI-070 | Chr 11 | 112832277 | (G)10 | NCAM1 | Non-standard | <0.0001 | 0.0006 |
| 79 | MSI-071 | Chr 11 | 115047033 | (T)14 | CADM1 | Non-standard | <0.0001 | <0.0001 |
| 80 | MSI-072 | Chr 11 | 118220583 | (A)9 | CD3G | Non-standard | 0.0042 | 0.0064 |
| 81 | MSI-073 | Chr 11 | 119545225 | (G)8 | NECTIN1 | Non-standard | 0.3102 | 0.1147 |
| 82 | MSI-074 | Chr 11 | 125763611 | (T)13 | PUS3 | Non-standard | <0.0001 | 0.0004 |
| 83 | MSI-075 | Chr 12 | 54405180 | (C)11 | HOXC8 | Non-standard | <0.0001 | <0.0001 |
| 84 | MSI-076 | Chr 12 | 55759486 | (T)8 | OR6C75 | Non-standard | 0.2426 | 0.17 |
| 85 | MSI-077 | Chr 12 | 85285921 | (A)17 | SLC6A15 | Non-standard | <0.0001 | <0.0001 |
| 86 | MSI-078 | Chr 12 | 122242658 | (C)8 | SETD1B | Non-standard | 0.0367 | 0.0202 |
| 87 | MSI-079 | Chr 12 | 130785787 | (G)10 | No direct gene | Non-standard | 0.8397 | 0.3751 |
| 88 | MSI-080 | Chr 13 | 31722621 | (A)17 | HSPH1 | Non-standard | <0.0001 | <0.0001 |
| 89 | MSI-081 | Chr 13 | 58299435 | (A)11 | PCDH17 | Non-standard | 0.0028 | 0.5704 |
| 90 | MSI-082 | Chr 13 | 113965003 | (T)14 | LAMP1 | Non-standard | <0.0001 | <0.0001 |
| 91 | MSI-083 | Chr 14 | 24040436 | (C)9 | JPH4 | Non-standard | <0.0001 | 0.0006 |
| 92 | MSI-084 | Chr 14 | 53513440 | (A)12 | DDHD1 | Non-standard | <0.0001 | <0.0001 |
| 93 | MSI-085 | Chr 14 | 73959704 | (T)16 | RIOX1 | Non-standard | <0.0001 | <0.0001 |
| 94 | MSI-086 | Chr 14 | 74054709 | (T)15 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 95 | MSI-087 | Chr 14 | 93708031 | (A)10 | BTBD7 | Non-standard | <0.0001 | <0.0001 |
| 96 | MSI-088 | Chr 14 | 102821856 | (A)14 | CINP | Non-standard | <0.0001 | <0.0001 |
| 97 | MSI-089 | Chr 15 | 34157542 | (A)10 | RYR3 | Non-standard | <0.0001 | 0.0033 |
| 98 | MSI-090 | Chr 15 | 45897772 | (T)14 | BLOC1S6 | Non-standard | <0.0001 | 0.0002 |
| 99 | MSI-091 | Chr 16 | 10867203 | (A)9 | TVP23A | Non-standard | 0.0002 | 0.2842 |
| 100 | MSI-092 | Chr 16 | 11051908 | (A)16 | CLEC16A | Non-standard | <0.0001 | <0.0001 |
| 101 | MSI-093 | Chr 16 | 22650153 | (A)16 | No direct gene | Non-standard | <0.0001 | <0.0001 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 102 | MSI-094 | Chr 17 | 13981240 | (T)21 | COX10 | Non-standard | <0.0001 | <0.0001 |
| 103 | MSI-095 | Chr 17 | 19314918 | (T)18 | RNF112 | Non-standard | <0.0001 | <0.0001 |
| 104 | MSI-096 | Chr 17 | 27081793 | (T)16 | No direct gene | Non-standard | <0.0001 | <0.0001 |
| 105 | MSI-097 | Chr 17 | 52991732 | (A)16 | TOM1L1 | Non-standard | <0.0001 | <0.0001 |
| 106 | MSI-098 | Chr 17 | 56435161 | (C)7 | RNF43 | Non-standard | <0.0001 | 0.0008 |
| 107 | MSI-099 | Chr 17 | 74688718 | (A)15 | MXRA7 | Non-standard | <0.0001 | <0.0001 |
| 108 | MSI-100 | Chr 18 | 511981 | (T)14 | LINC01925 | Non-standard | <0.0001 | <0.0001 |
| 109 | MSI-101 | Chr 18 | 649880 | (T)15 | CLUL1 | Non-standard | <0.0001 | <0.0001 |
| 110 | MSI-102 | Chr 18 | 52445092 | (C)8 | No direct gene | Non-standard | 0.0093 | 0.3932 |
| 111 | MSI-103 | Chr 18 | 57013194 | (T)9 | LMAN1 | Non-standard | <0.0001 | 0.0328 |
| 112 | MSI-104 | Chr 19 | 14104689 | (T)14 | RFX1 | Non-standard | <0.0001 | <0.0001 |
| 113 | MSI-105 | Chr 19 | 49458971 | (G)8 | BAX | Non-standard | 0.0005 | 0.0008 |
| 114 | MSI-106 | Chr 19 | 49850473 | (G)8 | TEAD2 | Non-standard | 0.0043 | 0.0014 |
| 115 | MSI-107 | Chr 20 | 46286321 | (A)10 | SULF2 | Non-standard | <0.0001 | 0.0007 |
| 116 | MSI-108 | Chr 20 | 47858504 | (A)8 | DDX27 | Non-standard | 0.0011 | 0.4551 |
| 117 | MSI-109 | Chr 20 | 61536692 | (A)11 | DIDO1 | Non-standard | <0.0001 | <0.0001 |
| 118 | MSI-110 | Chr 22 | 26352179 | (T)16 | MYO18B | Non-standard | <0.0001 | <0.0001 |
| 119 | MSI-111 | Chr 22 | 29696469 | (T)16 | EWSR1 | Non-standard | <0.0001 | <0.0001 |
| 120 | MSI-112 | Chr 22 | 37401595 | (T)15 | CIMIP4 | Non-standard | <0.0001 | <0.0001 |
| 121 | MSI-113 | Chr X | 37312611 | (C)8 | PPRG1 | Non-standard | 0.0259 | 0.3038 |
| 122 | MSI-114 | Chr X | 101409255 | (T)16 | BEX5 | Non-standard | <0.0001 | <0.0001 |

**Supplementary table 4A: MMR status (IHC/MSI-CE), number of unstable loci determined using model and number of mutations in CRC.**

| Sr | Case ID | MMR (IHC/MSI CE) | Model | Model-Estimated Unstable Loci (N=51) | concordance | KRAS | BRAF | NRAS | ERBB2 | OTHERS | VUS | Pathogenic (#) | VUS (#) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | C5 | dMMR | MSI | 48 | concordant | WT | c.1799T>A, p.(Val600Glu), 14.1% | WT | WT | PTEN c.800del, p.(Lys267fs), 15.1% PTEN c.968del, p.(Asn323fs), 15.9% | AR, NTRK1, NTRK3, RAF1, SMO, STK11 | 3 | 6 |
| 11 | C11 | dMMR | MSI | 47 | concordant | c.35G>A, p.(Gly12Asp), 22.6% | WT | WT | c.2033G>A, p.(Arg678Gln), 24.2% | PIK3CA c.1258T>C, p.(Cys420Arg), 20.9% STK11 c.157del, p.(Asp53fs), 21.8% CTNNB1 c.121A>G, p.(Thr41Ala), 49.0% | - | 5 | 0 |
| 15 | C15 | dMMR | MSI | 51 | concordant | c.182A>G, p.(Gln61Arg), 20.50% | WT | WT | WT | TP53 c.140del, p.(Pro47ArgfsTer76), 18% GNAS c.601C>T, p.(Arg201Cys), 16.70% AR c.2546dup, p.(Asn849LysfsTer32), 34.10% | NRG1 | 4 | 1 |
| 19 | C19 | dMMR | MSI | 51 | concordant | WT | WT | WT | WT | TP53 c.743G>A, p.(Arg248Gln), 28.70% | - | 1 | 0 |
| 20 | C20 | dMMR | MSI | 51 | concordant | c.38G>A, p.(Gly13Asp), 37.80% | WT | WT | WT | TP53 c.524G>A, p.(Arg175His), 29.30% TP53 c.472C>T, p.(Arg158Cys), 6.60% TP53 c.733G>A, p.(Gly245Ser), 8.90% RB1 c.1959del, p.(Val654fs), 11.60% GNAS c.602G>A, p.(Arg201His), 37.10% | RET | 6 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | C25 | dMMR | MSI | 49 | concordant | WT | c.1799T>A, p.(Val600Glu), 25.80% | WT | WT | - | EGFR 2X | 1 | 2 |
| 34 | C34 | dMMR | MSI | 23 | concordant | c.436G>A, p.(Ala146Thr), 12.70% | WT | WT | WT | PIK3CA c.3140A>G, p.(His1047Arg), ROS1 c.1267del, p.Ile423LeufsTer14), 13.90% MET c.4169_4171del, p.Ser1390del), 12.30% NTRK3 c.895dup, p.( Leu299ProfsTer16), 12.60% | ERBB2 | 5 | 1 |
| 38 | C38 | dMMR | MSI | 33 | concordant | c.38G>A, p.(Gly13Asp), 13.40% | WT | WT | WT | CTNNB1 c.121A>G, p.(Thr41Ala), 28.10% | - | 2 | 0 |
| 40 | C40 | dMMR | MSI | 51 | concordant | c.35G>A, p.(Gly12Asp), 11.20% | WT | WT | WT | PIK3CA c.263G>A, p.(Arg88Gln), 13.20% PIK3CA c.3139C>T, p.(His1047Tyr), 14.10% PTEN c.286C>T, p.Pro96Ser), 10.80% | - | 4 | 0 |
| 47 | C47 | dMMR | MSI | 11 | concordant | c.34G>T, p.(Gly12Cys), 34.70% | WT | WT | WT | TP53 c.488A>G, p.(Tyr163Cys), 54.90% | - | 2 | 0 |
| 50 | C50 | dMMR | MSI | 40 | concordant | WT | c.1799T>A, p.(Val600Glu), 22.30% | WT | WT | PIK3CA c.1633G>A, p.(Glu545Lys), 18.00% PIK3CA c.2173G>A, p.(Asp725Asn), 17.90% | - | 3 | 0 |
| 51 | C51 | dMMR | MSI | 50 | concordant | WT | c.1799T>A, p.(Val600Glu), 32.3% | WT | WT | MTOR | AR, DDR2, EGFR, RET | 2 | 4 |
| 58 | C58 | dMMR | MSI | 48 | concordant | WT | c.1799T>A, p.(Val600Glu), 14.2% | WT | WT | PIK3CA, PTEN, TP53 | AR, MET, TP53 | 4 | 3 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 62 | C62 | dMMR | MSI | 51 | concordant | WT | c.1799T>A, p.(Val600Glu), 18.4% | WT | WT | NF1, PIK3CA (3) | NTRK1 | 5 | 1 |
| 23 | C23 | dMMR* | MSI | 28 | concordant | WT | c.1799T>A, p.(Val600Glu), 6.80% | WT | WT | PIK3CA c.1359_1361del, p.(Glu453del), 6.20% | - | 2 | 0 |
| 1 | C1 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | TP53 c.586C>T, p.(Arg196Ter), 27.3% CDKN2A c.99_109del, p.(Glu33fs), 19.9% | - | 2 | 0 |
| 2 | C2 | pMMR | MSS | 0 | concordant | c.34G>T, p.(Gly12Cys), 74.9% | WT | WT | c.2030G>T, p.(Arg677Leu), 65.4% | ERBB3 c.695C>T, p.(Ala232Val), 42.0% SMAD4 c.725C>G, p.(Ser242Ter), 62.8% | - | 4 | 0 |
| 3 | C3 | pMMR | MSS | 1 | concordant | c.35G>C, p.(Gly12Ala), 19.8% | WT | WT | WT | TP53 c.455dup, p.(Pro153fs), 27.3% | NTRK3 | 2 | 1 |
| 4 | C4 | pMMR | MSS | 5 | concordant | WT | WT | WT | WT | TP53 c.339_340deL, p.(Leu114fs), 6.8% TP53 c.272G>A, p.(Trp91Ter), 20.4% | ERBB2, NTRK3 | 2 | 2 |
| 6 | C6 | pMMR | MSS | 0 | concordant | c.37G>T, p.(Gly13Cys), 23.8% | WT | WT | WT | TP53 c.1024C>T, p.(Arg342Ter), 20.2% | - | 2 | 0 |
| 7 | C7 | pMMR | MSS | 0 | concordant | c.35G>T, p.(Gly12Val), 9.7% | WT | WT | WT | PIK3CA c.1633G>A, p.(Glu545Lys), 18.2% TP53 c.733G>A, p.(Gly245Ser), 14.7% | - | 3 | 0 |
| 8 | C8 | pMMR | MSS | 2 | concordant | WT | c.1799T>A, p.(Val600Glu), 24.4% | WT | WT | GNAS c.602G>A, p.(Arg201His), 35.9% | - | 2 | 0 |
| 9 | C9 | pMMR | MSS | 0 | concordant | c.38G>A, p.(Gly13Asp), 12.6% | WT | WT | WT | TP53 c.524G>A, p.(Arg175His), 12.3% | - | 2 | 0 |
| 10 | C10 | pMMR | MSS | 0 | concordant | c.35G>A, p.(Gly12Asp), 25.0% | WT | WT | WT | PIK3CA c.1357G>A, p.(Glu453Lys), 14.50% | ALK | 2 | 1 |
| 12 | C12 | pMMR | MSS | 0 | concordant | c.35G>A, | WT | WT | WT | - | - | 1 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | p.(Gly12Asp), 36.8% | | | | | | | |
| 13 | C13 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | TP53 c.404G>A, p.(Cys135Tyr), 11.20% | - | 1 | 0 |
| 14 | C14 | pMMR | MSS | 1 | concordant | WT | WT | WT | WT | TP53 c.586C>T, p.(Arg196Ter), 40.60% SMAD4 c.1051G>C, p.(Asp351His), 31.50% | - | 2 | 0 |
| 16 | C16 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | TP53 c.637C>T, p.(Arg213Ter), 43.60% | ERBB2, MTOR | 1 | 2 |
| 17 | C17 | pMMR | MSS | 1 | concordant | WT | WT | WT | WT | TP53 c.646G>A, p.(Val216Met), 88.70% | - | 1 | 0 |
| 18 | C18 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | SMAD4 c.1096C>T, p.(Gln366Ter), 37.50% | - | 1 | 0 |
| 21 | C21 | pMMR | MSS | 1 | concordant | c.35G>A, p.(Gly12Asp), 18.30% | WT | WT | WT | PIK3CA c.1633G>A, p.(Glu545Lys), 19.10% | - | 2 | 0 |
| 22 | C22 | pMMR | MSS | 1 | concordant | WT | WT | WT | WT | - | - | 0 | 0 |
| 24 | C24 | pMMR | MSS | 0 | concordant | WT | c.1799T>A, p.(Val600Glu), 20.30% | WT | WT | TP53 c.403del, p.(Cys135AlafsTer35), 19.80% | SMAD4 | 2 | 1 |
| 26 | C26 | pMMR | MSS | 0 | concordant | c.35G>A, p.(Gly12Asp), 24.20% | WT | WT | WT | TP53 c.818G>A, p.(Arg273His), 28.00% | - | 2 | 0 |
| 27 | C27 | pMMR | MSS | 4 | concordant | c.40G>A, p.(Val14Ile), 22.50% c.99T>G, p.(Asp33Glu), 21.40% | WT | WT | WT | IDH1 c.394C>T, p.(Arg132Cys), 26.00% PIK3CA c.3140A>G, p.(His1047Arg), 17.60% | - | 4 | 0 |
| 28 | C28 | pMMR | MSS | 1 | concordant | c.35G>A, p.(Gly12Asp), 14.80% | WT | WT | WT | - | - | 1 | 0 |
| 29 | C29 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | PIK3CA c.1624G>A, p.(Glu542Lys), 14.50% | - | 2 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | TP53 c.751A>T, p.(Ile251Phe), 15.80% | | | |
| 30 | C30 | pMMR | MSS | 2 | concordant | c.180_181delinsAA, p.(Gln61Lys), 27.20% | WT | WT | WT | SMAD4 c.1493T>G, p.(Leu498Ter), 12.70% | - | 2 | 0 |
| 31 | C31 | pMMR | MSS | 1 | concordant | WT | WT | WT | WT | TP53 c.586C>T, p.(Arg196Ter), 36.70% | - | 1 | 0 |
| 32 | C32 | pMMR | MSS | 1 | concordant | WT | WT | WT | WT | TP53 c.743G>A, p.(Arg248Gln), 10.90% TP53 c.731G>A, p.(Gly244Asp), 34.70% | SMAD4 | 2 | 1 |
| 33 | C33 | pMMR | MSS | 1 | concordant | WT | WT | WT | WT | TP53 c.623A>T, p.(Asp208Val), 69.10% | - | 1 | 0 |
| 35 | C35 | pMMR | MSS | 0 | concordant | c.38G>A, p.(Gly13Asp), 33.10% | WT | WT | WT | TP53 c.743G>A, p.(Arg248Gln), 43.00% | EGFR, MTOR | 2 | 2 |
| 36 | C36 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | - | ALK, ROS1 | 0 | 2 |
| 37 | C37 | pMMR | MSS | 0 | concordant | c.34G>T, p.(Gly12Cys), 35.40% | WT | WT | WT | - | SMAD4 | 1 | 1 |
| 39 | C39 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | TP53 c.261_270del, p.(Ala88GlyfsTer32), 39.20% | RET, ROS1 | 1 | 2 |
| 41 | C41 | pMMR | MSS | 1 | concordant | c.68T>G, p.(Leu23Arg), 77.70% | WT | WT | WT | TP53 c.743G>A, p.(Arg248Gln), 46.50% | - | 2 | 0 |
| 42 | C42 | pMMR | MSS | 1 | concordant | WT | WT | c.34G>T, p.(Gly12Cys), 32.50% | WT | TP53 c.178_179del, p.(Pro60ArgfsTer2), 43.30% | - | 2 | 0 |
| 43 | C43 | pMMR | MSS | 0 | concordant | c.38G>A, p.(Gly13Asp), 3.10% | WT | WT | WT | TERT promotore), c.-124C>T), 9.70% | - | 2 | 0 |
| 44 | C44 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | TP53 c.880G>T, p.(Glu294Ter), 5.50% | - | 1 | 0 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 45 | C45 | pMMR | MSS | 1 | concordant | c.38G>A, p.(Gly13Asp), 48.80% | WT | WT | WT | TP53 c.818G>A, p.(Arg273His), 68.60% | - | 2 | 0 |
| 46 | C46 | pMMR | MSS | 1 | concordant | WT | WT | WT | WT | TP53 c.747G>T, p.(Arg249Ser), 77.10% | - | 1 | 0 |
| 48 | C48 | pMMR | MSS | 0 | concordant | c.34G>T, p.(Gly12Cys), 24.60% | WT | WT | WT | TP53 c.375+1G>C, 53.10% | - | 2 | 0 |
| 49 | C49 | pMMR | MSS | 0 | concordant | c.38G>A, p.(Gly13Asp), 33.60% | WT | WT | WT | TP53 c.563A>T, p.(His179Leu), 62.70% | - | 2 | 0 |
| 52 | C52 | pMMR | MSS | 3 | concordant | c.35G>A, p.(Gly12Asp), 17.9% | WT | WT | WT | - | IDH1 | 1 | 1 |
| 53 | C53 | pMMR | MSS | 1 | concordant | c.35G>A, p.(Gly12Asp), 47.4% | WT | WT | WT | PIK3CA | SMAD4, TP53 | 2 | 2 |
| 54 | C54 | pMMR | MSS | 5 | concordant | c.34G>T, p.(Gly12Cys), 36.6% | WT | WT | WT | TP53 | - | 2 | 0 |
| 56 | C56 | pMMR | MSS | 2 | concordant | WT | WT | WT | WT | PIK3CA, PTEN, TP53 | - | 3 | 0 |
| 57 | C57 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | TP53 | - | 1 | 0 |
| 59 | C59 | pMMR | MSS | 2 | concordant | c.35G>A, p.(Gly12Asp), 20.8% | WT | WT | WT | PIK3CA, TP53 | - | 3 | 0 |
| 61 | C61 | pMMR | MSS | 1 | concordant | c.35G>T, p.(Gly12Val), 20% | WT | WT | WT | TP53 | - | 2 | 0 |
| 63 | C63 | pMMR | MSS | 0 | concordant | WT | WT | WT | WT | - | ALK | 0 | 1 |
| 55 | C55 | pMMR* | MSS | 3 | concordant | c.35G>T, p.(Gly12Val), 20.1% | WT | WT | WT | TP53 | RET | 2 | 1 |
| 60 | C60 | pMMR* | MSS | 1 | concordant | WT | c.1799T>A, p.(Val600Glu), | WT | WT | MTOR, TP53 | ALK | 3 | 1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | 22.2% | | | | | | |
| 64 | C64 | pMMR* | MSS | 1 | concordant | WT | WT | WT | WT | - | AKT1 | 0 | 1 |
| 65 | C65 | pMMR* | MSS | 2 | concordant | WT | WT | WT | WT | - | ALK, PTEN | 0 | 2 |

**Supplementary table 4B: MMR status (IHC/MSI-CE), number of unstable loci determined using model and number of mutations in EC.**

| Sr | Case ID | MMR (IHC/MSI CE) | Model | Model-Estimated Unstable Loci (N=51) | concordance | POLE | TP53 | CTNNB1 | OTHERS | VUS | Pathogenic (#) | VUS (#) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | E1 | dMMR | MSI | 45 | concordant | WT | WT | WT | PIK3CA c.1624G>A, p.(Glu542Lys), 38.7% PTEN c.968dup, p.(Asn323fs), 11.6% TSC1 c.1257del, p.(Arg420fs), 17.6% | POLE c.1109C>T, p.(Pro370Leu), 5% KEAP1, NF1, NTRK2, PTEN | 3 | 5 |
| 5 | E5 | dMMR | MSI | 23 | concordant | WT | WT | WT | KRAS c.35G>A, p.(Gly12Asp), 18.5% TP53 c.216del, p.(Val73fs), 18.3% PTEN c.68_69del, p.(Leu23fs), 42.1% | EGFR | 3 | 1 |
| 8 | E8 | dMMR | MSI | 25 | concordant | WT | WT | WT | PIK3CA c.112C>A, p.(Arg38Ser), 19.20% | | 1 | 0 |
| 9 | E9 | dMMR | MSI | 20 | concordant | WT | WT | WT | PTEN c.425G>C, p.(Arg142Pro), 76.40% RB1 c.219_220dup, p.(Ala74fs), 56.30% | ERBB3 | 2 | 1 |
| 12 | E12 | dMMR | MSI | 28 | concordant | WT | WT | WT | PIK3CA c.263G>A, p.(Arg88Gln), 10.00% PTEN c.388C>G, p.(Arg130Gly), 54.60% NF1 c.2033dup, p.(Ile679AspfsTer21), 30.20% NF1 c.4436del, p.(Phe1479SerfsTer4), 29.90% | | 4 | 0 |
| 13 | E13 | dMMR | MSI | 45 | concordant | WT | WT | WT | KRAS c.35G>T, p.(Gly12Val), 36.90% PIK3CA c.263G>A, p.(Arg88Gln), 38.30% PTEN c.170del, | TERT | 4 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | p.(Leu57TrpfsTer42), 32.40%<br>PTEN c.209+4_209+7del), 29.60% | | | |
| 15 | E15 | dMMR | MSI | 15 | concordant | WT | WT | WT | PTEN c.82del, p.(Ile28PhefsTer26), 43.00% | | 1 | 0 |
| 28 | E28 | dMMR | MSI | 31 | concordant | WT | WT | WT | AR c.1208C>T, p.(Ala403Val), 7.20%<br>ESR1 c.1607T>C, p.(Leu536Pro), 37.40%<br>FGFR2 c.1172T>G, p.(Met391Arg), 38.00%<br>PIK3CA c.328_330del)), p.Glu110del), 32.60%<br>PIK3CA c.1033A>G, p.(Asn345Asp), 40.10%<br>PTEN c.388C>G, p.(Arg130Gly), 70.80% | | 6 | 0 |
| 33 | E33 | dMMR | MSI | 15 | concordant | WT | WT | WT | PTEN c.448G>T, p.(Glu150Ter), 25.80%<br>PTEN c.918_933del, p.(Ile306MetfsTer6), 5.50%<br>NF1 c.2033del, p.(Pro678ArgfsTer10), 6.30%<br>NF1 c.1885G>A, p.(Gly629Arg), 5.00%<br>ESR1 c.1607T>C, p.(Leu536Pro), 27.20% | | 5 | 0 |
| 36 | E36 | dMMR | MSI | 27 | concordant | WT | WT | c.133_135del, p.(Ser45del), 18.50% | PIK3CA c.3140A>G, p.(His1047Arg), 28.30%<br>PTEN c.800del, p.(Lys267ArgfsTer9), | | 4 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 31.90%<br>PTEN c.955_958del, p.(Thr319Ter), 21.30% | | | |
| 40 | E40 | dMMR | MSI | 47 | concordant | WT | WT | WT | KRAS c.38G>A, p.(Gly13Asp), 30.20%<br>PIK3CA c.263G>A, p.(Arg88Gln), 9.0%<br>RET c.229C>T, p.(Arg77Cys), 32.90%<br>PTEN c.97_99del, p.(Ile33del), 29.70%<br>PTEN c.800del, p.(Lys267ArgfsTer9), 33.0%<br>NF1 c.2033dup, p.(Ile679AspfsTer21), 26.90%<br>NF1 c.5786_5787del, p.(Glu1929ValfsTer6), 23.40% | | 7 | 0 |
| 56 | E56 | dMMR | MSI | 41 | concordant | WT | WT | WT | KRAS c.35G>C, p.(Gly12Ala), 36.70%<br>PTEN c.922dup, p.(Arg308ProfsTer4), 77.10% | | 2 | 0 |
| 57 | E57 | dMMR | MSI | 35 | concordant | WT | WT | c.133_135del, p.(Ser45del), 27.0% | PIK3CA c.113G>A, p.(Arg38His), 28.40%<br>PTEN c.757_758del, p.(Ile253GlnfsTer44), 62.30% | ERBB2, ERBB3 | 3 | 2 |
| 58 | E58 | dMMR | MSI | 26 | concordant | WT | WT | WT | ERBB3 c.310G>A, p.(Val104Met), 26.80%<br>KRAS c.38G>A, p.(Gly13Asp), 24.50%<br>PIK3CA c.263G>A, p.(Arg88Gln), 27.30%<br>PTEN c.800del, p.(Lys267ArgfsTer9), 24.70% | SMAD4 | 6 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | PTEN c.968del, p.(Asn323MetfsTer21 26.70% RB1 c.1961-1G>C), sito di splicing), 24.90% | | | |
| 60 | E60 | dMMR | MSI | 33 | concordant | WT | WT | WT | PIK3CA), c.3140A>G, p.(His1047Arg), 39.90% PTEN c.800del, p.(Lys267ArgfsTer9), 41.00% | | 2 | 0 |
| 62 | E62 | dMMR | MSI | 30 | concordant | c.1372T>C, p.(Tyr458His), 31.8% | c.817C>T, p.(Arg273Cys), 34.9% | WT | ERBB3, FGFR1, FGFR3, PIK3CA, PTEN (2), RAF1 | ALK, HRAS, MET (2), MTOR (2), NF1 (2), NTRK2, TP53 (2) | 9 | 11 |
| 6 | E6 | dMMR* | MSI | 7 | concordant | WT | WT | WT | KRAS c.35G>C, p.(Gly12Ala), 18.8% TSC1 c.474del, p.(Phe158fs), 7.1% | NF1 | 2 | 1 |
| 11 | E11 | dMMR* | MSI | 8 | concordant | WT | WT | c.110C>T, p.(Ser37Phe), 6.80% | KRAS c.175G>A, p.(Ala59Thr), 8.40% PIK3CA c.317G>T, p.(Gly106Val), 20.70% PTEN c.389G>T, p.(Arg130Pro), 11.70% PTEN c.955_958del, p.(Thr319Ter), 26.80% | | 5 | 0 |
| 18 | E18 | dMMR* | MSI | 47 | concordant | WT | WT | WT | PTEN c.270del, p.(Phe90LeufsTer9), 25.50% PTEN c.800del, p.(Lys267ArgfsTer9), 25.20% ERBB3 c.310G>A, p.(Val104Met), 26.90% | | 3 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 23 | E23 | dMMR* | MSI | 37 | concordant | WT | WT | WT | KRAS c.35G>A, p.(Gly12Asp), 20.20% PTEN c.39_40del, p.(Arg14GlufsTer29), 19.20% PTEN c.388C>G, p.(Arg130Gly), 17.60% PIK3CA c.328_330del, p.(Glu110del), 21.40% | NF1 | 4 | 1 |
| 31 | E31 | dMMR* | MSI | 22 | concordant | WT | c.524G>A, p.(Arg175His), 21.20% | WT | NF1 c.3827G>A, p.(Arg1276Gln), 17.40% PIK3CA c.263G>A, p.(Arg88Gln), 28.90% PTEN c.212G>A, p.(Cys71Tyr), 22.60% PTEN c.517C>T, p.(Arg173Cys), 11.60% | | 5 | 0 |
| 55 | E55 | dMMR* | MSI | 20 | concordant | WT | WT | WT | NF1 c.6772C>T, p.(Arg2258Ter), 20.70% PTEN), c.388C>T, p.(Arg130Ter), 39.90% | | 2 | 0 |
| 66 | E67 | dMMR* | MSI | 6 | concordant | WT | WT | WT | BRAF c.2131C>T, p.(Leu711Phe), 4.70% PIK3CA c.112C>T, p.(Arg38Cys), 9.60% | ALK, FGFR2, MTOR, PIK3CA, ROS1, STK11 | 2 | 6 |
| 3 | E3 | pMMR | MSS | 0 | concordant | WT | WT | WT | PTEN c.968dup, p.(Asn323fs), 38.1% TERT promoter), c.-124C>T), 45.9% | ROS1 | 2 | 1 |
| 4 | E4 | pMMR | MSS | 0 | concordant | WT | WT | WT | | | 0 | 0 |
| 14 | E14 | pMMR | MSS | 1 | concordant | WT | WT | c.98C>A, p.(Ser33Tyr), 33.10% | PTEN c.697C>T, p.(Arg233Ter), 33.20% | PTEN | 2 | 1 |
| 16 | E16 | pMMR | MSS | 3 | concordant | WT | WT | WT | KRAS c.35G>T, p.(Gly12Val), 43.60% PTEN c.968dup, p.(Asn323LysfsTer2), | TERT PROMOTORE, PDGFRA | 2 | 2 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 39.70% | | | |
| 17 | E17 | pMMR | MSS | 0 | concordant | WT | c.713G>T, p.(Cys238Phe), 29.20% | WT | NRAS c.182A>G, p.(Gln61Arg), 24.10% CDKN2A c.389_410del, p.(Leu130ProfsTer9), 27.90% PTEN c.44_56del, p.(Arg15MetfsTer5), 17.80% PTEN c.722T>C, p.(Phe241Ser), 18.20% | | 5 | 0 |
| 19 | E19 | pMMR | MSS | 1 | concordant | WT | c.723del, p.(Cys242AlafsTer5), 67.4% | WT | FGFR4 c.310C>T, p.(Arg104Ter), 42.5% | FGFR4 | 2 | 1 |
| 20 | E20 | pMMR | MSS | 2 | concordant | WT | c.844C>T, p.(Arg282Trp), 57.30% | WT | KRAS c.35G>C, p.(Gly12Ala), 51.70% PTEN c.113del, p.(Pro38LeufsTer16), 68.20% | | 3 | 0 |
| 22 | E22 | pMMR | MSS | 1 | concordant | WT | WT | WT | PTEN c.388C>G, p.(Arg130Gly), 54.00% PIK3CA c.311C>T, p.(Pro104Leu), 20.30% PIK3CA c.1030G>A, p.(Val344Met), 21.90% ERBB2 c.2033G>A, p.(Arg678Gln), 24.00% | ESR1, c.1610A>C, p.(Tyr537Ser), 34.0%, MTOR | 4 | 2 |
| 25 | E25 | pMMR | MSS | 0 | concordant | WT | WT | c.121A>G, p.(Thr41Ala), 5.80% | PTEN c.545T>A, p.(Leu182Ter), 30.30% | NRG1, TSC1 | 2 | 2 |
| 26 | E26 | pMMR | MSS | 2 | concordant | WT | WT | WT | PIK3CA c.1034A>C, p.(Asn345Thr), 10.60% PTEN c.389G>T, p.(Arg130Leu), 9.00% PTEN c.979_982del, p.(Lys327GlnfsTer16), 5.30% | | 4 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | FGFR2 c.755C>G, p.(Ser252Trp), 12.90% | | | |
| 30 | E30 | pMMR | MSS | 0 | concordant | WT | WT | WT | PIK3CA c.248T>C, p.(Phe83Ser), 7.20%<br>PIK3CA c.263G>A, p.(Arg88Gln), 7.20%<br>PTEN c.394_412del, p.(Gly132IlefsTer9), 6.30% | | 3 | 0 |
| 32 | E32 | pMMR | MSS | 0 | concordant | WT | WT | c.100G>A, p.(Gly34Arg), 19.70% | PTEN c.723del, p.(Phe241LeufsTer15), 33.10%<br>PTEN c.875del, p.(Asn292MetfsTer15), 16.30% | | 3 | 0 |
| 34 | E34 | pMMR | MSS | 1 | concordant | WT | WT | c.98C>T, p.(Ser33Phe), 21.30% | PIK3CA c.263G>A, p.(Arg88Gln), 9.40%<br>PTEN c.703G>T, p.(Glu235Ter), 52.40% | | 3 | 0 |
| 35 | E35 | pMMR | MSS | 0 | concordant | WT | WT | WT | PIK3CA c.3140A>G, p.(His1047Arg), 18.60%<br>PTEN c.389G>A, p.(Arg130Gln), 20.70% | | 2 | 0 |
| 38 | E38 | pMMR | MSS | 1 | concordant | WT | c.743G>A, p.(Arg248Gln), 23.20%<br>c.524G>A, p.(Arg175His), 23.30% | WT | PIK3CA c.3127A>G, p.(Met1043Val), 24.70% | | 3 | 0 |
| 39 | E39 | pMMR | MSS | 0 | concordant | WT | c.288_291del, p.(Val97LeufsTer25), 20.80% | WT | | | 1 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 41 | E41 | pMMR | MSS | 0 | concordant | WT | WT | c.110C>T, p.(Ser37Phe), 10.30% | PTEN c.611del, p.(Pro204GlnfsTer17), 16.50%<br>PTEN c.801+1G>A), variante di splicing), 8.70% | | 3 | 0 |
| 43 | E43 | pMMR | MSS | 1 | concordant | WT | WT | WT | PTEN c.493-2A>G), variante di splicing), 40.40% | | 1 | 0 |
| 44 | E44 | pMMR | MSS | 0 | concordant | WT | c.731G>A, p.(Gly244Asp), 17.20% | WT | PTEN c.388C>G, p.(Arg130Gly), 17.70%<br>PTEN c.1026+2T>G, variante di splicing), 16.30% | | 3 | 0 |
| 45 | E45 | pMMR | MSS | 1 | concordant | WT | WT | WT | KRAS c.35G>C, p.(Gly12Ala), 30.90%<br>PTEN c.214_224del, p.(Ala72LeufsTer2), 36.80% | | 2 | 0 |
| 46 | E46 | pMMR | MSS | 0 | concordant | WT | WT | c.94G>T, p.(Asp32Tyr), 6.50%<br>c.101G>T, p.(Gly34Val), 14.20% | PTEN c.1026+2T>G), variante di splicing), 66.80%<br>FGFR2 c.755C>G, p.(Ser252Trp), 41.40% | | 4 | 0 |
| 47 | E47 | pMMR | MSS | 0 | concordant | WT | WT | WT | ERBB2 c.2033G>A , p.(Arg678Gln), 21.70%<br>PTEN c.388C>G, p.(Arg130Gly), 18.70%<br>PTEN c.626del, p.(Gly209GlufsTer12), 17.90% | | 3 | 0 |
| 52 | E52 | pMMR | MSS | 1 | concordant | WT | WT | WT | FGFR2 c.755C>G, p.(Ser252Trp), 26.90%<br>MTOR c.5185del, p.(Gln1729SerfsTer24), 30.90%<br>PTEN c.584_600del, P.Phe195Ter), 52.0% | | 3 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 54 | E54 | pMMR | MSS | 0 | concordant | WT | WT | WT | PIK3CA c.1633G>A, p.(Glu545Lys), 7.20% | | 1 | 0 |
| 61 | E61 | pMMR | MSS | 0 | concordant | c.1231G>T, p.Val411Leu), 25.20% | WT | WT | PIK3CA c.263G>A, p.(Arg88Gln), 22.30% PTEN c.389G>A, p.(Arg130Gln), 25.10% | VUS in 7 geni | 3 | 7 |
| 63 | E63 | pMMR | MSS | 1 | concordant | c.1231G>T, p.(Val411Leu), 1.7% | c.637C>T, p.(Arg213Ter), 2.1% | WT | PTEN | NF1, RAF1 | 3 | 2 |
| 65 | E65 | pMMR | MSS | 0 | concordant | WT | WT | WT | NF1, PTEN | | 2 | 0 |
| 2 | E2 | pMMR* | MSS | 1 | concordant | WT | WT | WT | PTEN c.842del, p.(Pro281fs), 23.0% FGFR2 c.755C>G, p.(Ser252Trp), 9.3% | | 2 | 0 |
| 27 | E27 | pMMR* | MSS | 1 | concordant | WT | WT | WT | | | 0 | 0 |
| 29 | E29 | pMMR* | MSS | 4 | concordant | c.1231G>T, p.(Val411Leu), 43.20% | c.536A>G, p.(His179Arg), 46.50% c.976G>T, p.(Glu326Ter), 40.60% | WT | AKT1 c.49G>A, p.(Glu17Lys), 44.10% AR c.2494C>T, p.(Arg832Ter), 46.40% BRAF c.1405G>A, p.(Gly469Arg), 45.0% FGFR3 c.1196G>A, p.(Arg399His), 45.80% MTOR c.664C>A, p.(Ser2215Tyr), 42.0% PIK3CA c.278G>A, p.(Arg93Gln), 38.30% RB1 c.2023G>T, p.(Glu675Ter), 41.80% | 29 geni | 10 | 29 |
| 37 | E37 | pMMR* | MSS | 2 | concordant | WT | WT | WT | KRAS c.35G>A, p.(Gly12Asp), 5.90% PTEN c.388C>G, p.(Arg130Gly), 12.30% | | 2 | 0 |
| 48 | E48 | pMMR* | MSS | 2 | concordant | c.1231G>T, p.(Val411Leu), 12.10% | WT | WT | IDH1 c.394C>T, p.(Arg132Cys), 18.0% PIK3CA c.263G>A, p.(Arg88Gln), 19.0% PIK3CA c.3073A>G, p.(Thr1025Ala), | ESR1: c.1186A>G, p.(Met396Val), 49.90% numero elevato di varianti VUS | 5 | 1 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | 19.80% PTEN), c.389G>A, p.(Arg130Gln), 39.50% | | | |
| 51 | E51 | pMMR* | MSS | 3 | concordant | WT | WT | WT | PIK3CA c.1624G>A, p.(Glu542Lys), 30.90% | | 1 | 0 |
| 64 | E64 | pMMR* | MSS | 0 | concordant | WT | c.637C>T, p.(Arg213Ter), 16.3% + c.638G>A, p.(Arg213Gln), 9.3% | WT | MTOR, NF1 (2), PIK3CA (2), PTEN (3), RB1, SMAD4 | POLE c.1331T>A, p.(Met444Lys), 15.6%, CDKN2A (2), ERBB2, NF1, RB1, RET | 12 | 7 |
| 53 | E53 | dMMR | MSS | 5 | FN | WT | WT | WT | KRAS c.35G>A, p.(Gly12Asp), 9.50% PIK3CA c.3140A>G, p.(His1047Arg), 13.00% | | 2 | 0 |
| 7 | E7 | dMMR* | MSS | 2 | FN | WT | WT | WT | PTEN c.741dup, p.(Pro248fs), 13.40% RB1 c.62dup, p.(Ala22fs), 8.60% | | 2 | 0 |
| 10 | E10 | dMMR* | MSS | 1 | FN | WT | WT | WT | PTEN c.388C>G, p.(Arg130Gly), 15.60% | AR | 1 | 1 |
| 24 | E24 | dMMR* | MSS | 3 | FN | WT | WT | WT | PTEN c.955dup, p.(Thr319AsnfsTer6), 26.30% KRAS c.35G>T, p.(Gly12Val), 50.30% | | 2 | 0 |
| 42 | E42 | dMMR* | MSS | 5 | FN | WT | WT | WT | KRAS c.35G>T, p.(Gly12Val), 20.70% PIK3CA c.1634A>C, p.(Glu545Ala), 19.10% PTEN c.388C>G, p.(Arg130Gly), 27.0% PTEN c.968del, p.(Asn323MetfsTer21), 25.40% | | 4 | 0 |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 50 | E50 | dMMR* | MSS | 2 | FN | WT | WT | WT | PTEN c.31A>T , p.(Arg11Ter), 24.70% PTEN c.1011_1023del, p.(Phe337LeufsTer3), 12.20% | | 2 | 0 |
| 21 | E21 | pMMR | MSI | 6 | FP | WT | WT | WT | KRAS c.38G>A, p.(Gly13Asp), 8.30% PIK3CA c.311C>T, p.(Pro104Leu), 7.10% | | 2 | 0 |
| 49 | E49 | pMMR* | MSI | 7 | FP | WT | WT | c.110C>T, p.(Ser37Phe), 9.30% | PIK3CA c.277C>T, p.(Arg93Trp), 10.70% PIK3CA c.3061T>C, p.(Tyr1021His), 10.40% PIK3CA c.3145G>A, p.(Gly1049Ser), 7.60% PTEN c.585del, p.(His196ThrfsTer3), 23.60% | TSC1 | 5 | 1 |
| 59 | E59 | pMMR* | MSI | 6 | FP | WT | WT | WT | PTEN c.697C>T, p.(Arg233Ter), 13.90% PTEN c.741_742insAA, p.(Pro248AsnfsTer9), 14.10% | RET | 2 | 1 |

*Blank for Acknowledgements*