Distinct molecular profiles of sporadic early onset colorectal cancer: a population-based cohort and systematic review

Ashleigh C. Hamilton, Finian J. Bannon, Philip D. Dunne, Jacqueline James, Stephen McQuaid, Ronan T. Gray, Manuel Salto-Tellez, Chris R. Cardwell, Maurice B. Loughrey, Helen G. Coleman

PII: S2772-5723(22)00192-3
DOI: https://doi.org/10.1016/j.gastha.2022.11.005
Reference: GASTHA 214

To appear in: Gastro Hep Advances

Received Date: 27 July 2022
Revised Date: 3 October 2022
Accepted Date: 2 November 2022


This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2022 Published by Elsevier, Inc on behalf of the AGA Institute
Title: Distinct molecular profiles of sporadic early onset colorectal cancer: a population-based cohort and systematic review

Short title: Early onset colorectal cancer: tumour profile

Authors: Ashleigh C Hamilton,¹ Finian J Bannon,¹ Philip D Dunne,³,⁷ Jacqueline James,³,⁴,⁵ Stephen McQuaid,³,⁴ Ronan T. Gray,¹,⁶ Manuel Salto-Tellez,³,⁵ Chris R. Cardwell,¹ Maurice B Loughrey,¹,²,³ Helen G Coleman¹,³

Affiliations:
¹Centre for Public Health, Queen’s University Belfast, Northern Ireland, UK
²Department of Cellular Pathology, Belfast Health and Social Care Trust, Northern Ireland, UK
³Patrick G. Johnston Centre for Cancer Research, Queen’s University Belfast, Northern Ireland, UK
⁴Northern Ireland Biobank, Belfast, Northern Ireland, UK
⁵Precision Medicine Centre of Excellence, Queen’s University Belfast, Northern Ireland, UK
⁶South Eastern Health and Social Care Trust, Northern Ireland, UK
⁷CRUK Beatson Institute, Glasgow, UK

Corresponding author: Dr Ashleigh Hamilton, Centre for Public Health, Queen’s University Belfast, ICS-B Building, Royal Victoria Hospital Site, Grosvenor Rd, Belfast, Northern Ireland, BT12 6BA.
Email: ahamilton17@qub.ac.uk
Phone number: +447738549888
Disclosures: All authors declare no conflict of interest.

Data transparency statement: Applications for secondary use of data generated and presented in this manuscript, or materials described, as part of the Epi700 cohort creation, can be made by contacting the Northern Ireland Biobank, email: nibiobank@qub.ac.uk.

Word Count: 6,189 (inclusive of main text, references, and table/figure legends)

Grant support: (including departmental or institutional funding when no extramural funding was received): Ashleigh Hamilton is funded by a 2019 Health and Social Care (HSC) Research and Development Doctoral Fellowship (EAT/15494/18) from the R&D Division of the Public Health Agency in Northern Ireland. Helen Coleman is funded by a Cancer Research UK Career Establishment Award (Reference:C37703/A25820). The resource creation and data analysed for the Epi700 cohort analysis was supported by a HSC Research and Development Division of the Public Health Agency Doctoral Fellowship (EAT/4905/13 – Ronan Gray), a Cancer Research UK Research Bursary (C50104/A17592 – Ronan Gray), and a CRUK Population Health Postdoctoral Fellowship (C37703/A15333 – Helen Coleman).

Abbreviations:

Chromosomal instability: CIN
CpG Island methylator phenotype: CIMP
Colorectal cancer: CRC
Confidence intervals: CI
Deficient mismatch repair: dMMR
Early onset colorectal cancer: EOCRC
Eastern Cooperative Oncology Group: ECOG
Inflammatory bowel disease: IBD
Late onset colorectal cancer: LOCRC
Microsatellite instability-high: MSI-H
Microsatellite stable: MSS
Proficient mismatch repair: pMMR

**Acknowledgements:** The authors would like to acknowledge the contributions of staff members who assisted with the creation of the Epi700 cohort data. The samples used in this research were received from the Northern Ireland Biobank\(^{(1)}\) which is funded by Health and Social Care Research and Development Division of the Public Health Agency in Northern Ireland and Cancer Research UK (via the former Belfast CRUK Centre and the Northern Ireland Experimental Cancer Medicine Centre); additional support was received from the Friends of the Cancer Centre. The Northern Ireland Molecular Pathology Laboratory, which was responsible for creating resources for the Northern Ireland Biobank, has received funding from Cancer Research UK, the Friends of the Cancer Centre and the Sean Crummey Foundation. Clinical data collected and analysed was facilitated by the Northern Ireland Cancer Registry, which is funded by the Public Health Agency, Northern Ireland.
**Author contributions for systematic review:**

ACH: conceptualisation, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, software, visualisation, writing – original draft

FJB: conceptualisation, methodology, writing - review and editing

PD: conceptualisation, methodology, visualisation, writing – review and editing

CC: data curation, formal analysis, methodology, software, writing – review and editing

MBL: conceptualisation, methodology, investigation, validation, supervision, visualisation, writing – original draft, writing – review and editing

HGC: conceptualisation, methodology, formal analysis, funding acquisition, investigation, project administration, software, supervision, validation, visualisation, writing – original draft, writing – review and editing

**Author contributions for Epi700 study:**

ACH: conceptualisation, formal analysis, investigation, methodology, software, visualisation, writing – original draft

FJB: methodology, formal analysis, supervision, visualisation, writing – original draft, writing – review and editing

JJ: conceptualisation, data curation, funding acquisition, formal analysis, investigation, methodology, project administration, resources, software, supervision, writing - review and editing
SM: conceptualisation, data curation, funding acquisition, formal analysis, investigation, methodology, project administration, resources, software, supervision, writing - review and editing

RG: conceptualisation, data curation, funding acquisition, formal analysis, investigation, methodology, project administration, resources, software, writing - review and editing

MS: conceptualisation, data curation, funding acquisition, formal analysis, investigation, methodology, project administration, resources, software, supervision, writing - review and editing

MBL: conceptualisation, data curation, funding acquisition, formal analysis, investigation, methodology, project administration, resources, software, supervision, writing – original draft, writing - review and editing

HGC: conceptualisation, data curation, funding acquisition, formal analysis, investigation, methodology, project administration, resources, software, supervision, writing – original draft, writing - review and editing

**Ethical Statement:** The corresponding author, on behalf of all authors, jointly and severally, certifies that their institution has approved the protocol for any investigation involving humans or animals and that all experimentation was conducted in conformity with ethical and humane principles of research.
Abstract

Background and aims:

The observed increase in incidence of early onset colorectal cancer (EOCRC) is being driven by sporadic cases, but the molecular characteristics of these tumours are not fully understood. Our objective was to investigate the prevalence of microsatellite instability (MSI) and selected mutations in sporadic EOCRC, and their association with survival.

Methods:

Firstly, we compared the prevalence of molecular characteristics and survival within a population-based cohort study of 652 Stage II and III colon cancer patients in Northern Ireland, comparing sporadic early onset (<50 years, n=35) with older (60-69 years, n=179) patients. Secondly, a systematic review for studies reporting the prevalence of MSI, mismatch repair deficiency (dMMR) or \textit{BRAF}, \textit{KRAS}, \textit{NRAS}, \textit{PIK3CA} and \textit{TP53} mutations in sporadic EOCRC was conducted. Meta-analysis was performed to calculate pooled estimates of the prevalence of molecular features in sporadic EOCRC.

Results:

Firstly, within the cohort study, EOCRC patients did not have a significantly increased risk of CRC-specific death (adjusted HR 1.20; 95%CI 0.61-2.39) compared with 60-69 year olds. Secondly, 32 studies were included in the systematic review. Pooled analysis estimated a prevalence of 10% (95%CI 7-14%) for MSI-H/dMMR in sporadic EOCRC. \textit{BRAF} and \textit{KRAS} mutations had a prevalence of 1% (95%CI 0-3%) and 32% (95%CI 23-40%), respectively.

Conclusion:
The molecular characteristics of sporadic EOCRC differ from those of older adults, particularly regarding reduced prevalence of BRAF mutations. Ten percent of sporadic EOCRC display MSI-H/dMMR. Further studies are needed to address survival in sporadic EOCRC and whether molecular profiles influence EOCRC outcomes in this patient group.

Keywords: colorectal cancer; early onset; microsatellite instability; mismatch repair; mutations
Introduction

An increase in the incidence of colorectal cancer (CRC) in adults under the age of 50 years, known as early onset colorectal cancer (EOCRC), has been observed in high-income countries.\(^{(2-5)}\) Studies have suggested that the majority of EOCRC is sporadic in nature, with cases associated with identified germline mutations accounting for up to 35%.\(^{(6, 7)}\) Not all previous studies have separated sporadic from hereditary cases of EOCRC however, and while the molecular pathogenesis of CRC due to inherited conditions such as Lynch syndrome is well defined, the pathways that lead to development of sporadic EOCRC remain incompletely understood.

CRC is a molecularly heterogeneous disease resulting from stepwise accumulation of mutations in key oncogenes and tumour suppressor genes leading to development of malignancy via a number of pathways, namely the chromosomal instability pathway (CIN), microsatellite instability (MSI) pathway and the serrated pathway. Each pathway displays multiple characteristic gene mutations and epigenetic changes. CRCs developing via the CIN pathway are associated with mutations in APC as an early event, with subsequent mutations in RAS, RAF, PIK3CA, SMAD4 and/or TP53 genes, among others.\(^{(8)}\) KRAS and NRAS mutation status are used in the clinical setting to inform systemic treatment options.\(^{(9)}\) CRCs arising through the MSI pathway display deficient mismatch repair (dMMR), which is synonymous with MSI,\(^{(10)}\) resulting from uncorrected errors during DNA replication.\(^{(11)}\) Lynch syndrome, a hereditary condition predisposing to the development of several cancers, results in microsatellite-instability-high (MSI-H) CRC. However, due to inclusion of both hereditary and sporadic cases of EOCRC in many studies to date, it is unclear how frequently sporadic MSI-H tumours occur in EOCRC. CRCs arising via the serrated pathway are MSI-H, associated with an
increased prevalence of \textit{BRAF} mutations (which is a distinguishing feature), and have high levels of CpG island methylation, known as the CpG Island methylator phenotype (CIMP).\textsuperscript{(12)}

A 2019 report reviewed 37 studies with regard to prognosis of EOCRC compared to late onset CRC (LOCRC), and found conflicting results for poorer, similar or better prognosis in younger patients.\textsuperscript{(13)} It is possible that survival differences between EOCRC and older patients may reflect different molecular profiles of tumours occurring in these patients. However, to our knowledge the evidence for molecular profiles of sporadic EOCRC tumours has not been systematically collated.

In the present study, we analysed a population-based cohort of patients with Stage II and III colon cancer, to investigate molecular characteristics in sporadic CRC and survival outcomes according to age categories. We also undertook a systematic review and meta-analysis of the prevalence of MSI status and selected tumour mutations in sporadic EOCRC.
Methods

Population-based cohort study

Patient population

The study cohort (known as Epi700) was established as previously described. In summary, 661 stage II and III colon cancer patients diagnosed in two healthcare trusts in Northern Ireland from 2004 to 2008, for whom resection specimens were available to be retrieved from the Northern Ireland Biobank, were identified using the Northern Ireland Cancer Registry. Patients were followed up for recurrence and cause of death to 31st December 2013.

Tumour pathology characteristics and clinical data collection

Where tumour pathology characteristics were not readily available from routinely extracted Cancer Registry information, further pathology details, for example tumour differentiation, were retrieved by manual review of pathology reports.

Clinical variables used in this study including family history of CRC, oncological treatments, Eastern Cooperative Oncology Group (ECOG) performance status, lifestyle information (including smoking and alcohol) and comorbidities were extracted from the Northern Ireland Clinical Oncology Information System, a prospective electronic record of patient management.

Tumour molecular analysis

Following tumour annotation and macrodissection, DNA was extracted according to the manufacturer’s instructions from 5µm sections of representative whole tumour blocks using
the Maxwell 16 instrument (Promega, Southampton, UK) and Promega DNA extraction kit (Promega, Southampton, UK).

MSI analysis was performed within the Northern Ireland Molecular Pathology Laboratory, using the MSI Analysis System, version 1.2 kit (Promega, Southampton, UK) for five mononucleotide repeat markers (BAT-25, BAT-26, NR-21, NR-24 and MONO-27). PCR products were separated by capillary electrophoresis using an ABI 3500 Genetic Analyzer (Fisher Scientific - UK Ltd, Loughborough, UK). The output data was analysed using GeneMapper® v4.1 (Fisher Scientific – UK Ltd, Loughborough, UK) to determine MSI status.\(^{10}\)

Tumour samples were analysed for mutational status for established CRC markers. This included a ColoCarta panel of KRAS, NRAS, BRAF, CMET and PIK3CA using a validated mass spectrometry-based targeted screening panel of 32 somatic mutations in 6 genes. (Agena Bioscience, Hamburg, Germany). Samples were shipped via the Genomics Core Technology Unit (Queen’s University Belfast) and the assays performed by Agena Custom Services Laboratory (Hamburg, Germany).

Statistical analysis

Statistical analysis was performed using Stata 16 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC). Chi-squared tests were used to compare descriptive and molecular characteristics across age categories. Survival analysis was performed using the Cox proportional hazards model to calculate hazard ratios and 95% confidence intervals (CI). The multivariable model for CRC-specific survival included sex, family history of CRC, stage (II / III), grade/differentiation, adjuvant chemotherapy receipt,
ECOG performance status, alcohol, smoking, inflammatory bowel disease (IBD) and emergency surgery. Results from this study were included in the subsequent systematic review and meta-analysis.

Systematic review and meta-analysis

This study was reported according to the Meta-analysis of Observation Studies in Epidemiology (MOOSE) checklist.\(^\text{[16]}\) The review protocol was registered on PROSPERO (CRD42021232567).

Study population

The population of interest was patients with sporadic EOCRC, defined as adults aged less than 50 years at their incident CRC diagnosis who had no identified inherited genetic syndrome which predisposes to CRC. Studies were included if sporadic cases were separated from hereditary cases by the authors or if the article contained information to enable distinction of hereditary from sporadic cases, such as results of genetic testing or family history.

Outcome

The primary outcome was to estimate the prevalence of MSI-H/dMMR status, \textit{KRAS}, \textit{NRAS}, \textit{BRAF}, \textit{PIK3CA} and \textit{TP53} mutations in EOCRC. The protocol specified a secondary outcome investigating the influence of molecular profile on survival in EOCRC, but insufficient data was available in potentially eligible articles and so we restricted the reporting of the review to the prevalence of molecular features as outlined.

Search strategy
The electronic databases Ovid Medline, Embase and Web of Science were systematically searched from 2000 to 12th April 2021. The full search terms are available in Supplementary Appendix 1. Observational studies, descriptive studies, case series and interventional studies were eligible for inclusion. All stages of CRC were included, with a focus on colorectal adenocarcinoma as the primary histology, and no language restrictions were imposed. Review articles, editorials, comments, abstract or conference proceedings, individual case studies, or case series with less than ten patients were excluded.

Articles from the search were imported into Covidence and duplicates were removed. Titles and abstracts were reviewed by two authors independently (AH and HC). Full texts were read by two authors (AH reviewed all articles; HC and ML reviewed independent subsets). Any discrepancies were resolved by discussion amongst the two reviewers of the text, with the third reviewer involved if required.

**Data extraction**

Data was extracted by AH and verified by HC. Data extracted related to study location, number of sporadic EOCRC cases, definition of sporadic cases, mutation testing, MSI or dMMR testing methods and the prevalence of each molecular characteristic in the study population, along with sex and anatomical tumour location if available. Where mutation testing had resulted in an unknown or ambiguous result, our approach was to exclude these cases from the analysis. Where a study contained data for both MSI status and MMR proteins, the MSI status results were used. Details of molecular testing undertaken in each study are shown in Supplementary Table 4. Immunohistochemistry for the p53 protein has been used as a surrogate for TP53 mutation testing,\(^{(17)}\) where reported in studies. In a change to the study
protocol, the Joanna Briggs Institute Checklist for Prevalence Studies was used for quality assessment.

**Statistical analysis**

Stata 16 was used to perform meta-analysis to produce pooled estimates of prevalence and 95% CI. The Freeman-Tukey Arcsine Transformation method was used to calculate estimates and standard errors which were back transformed to calculate a pooled prevalence.\(^{(18)}\) The logistic-normal random effects model was also carried out to ensure the pooled estimates and 95% CI were similar. Sensitivity analysis was performed, and subgroup analysis was carried out by sex and anatomical tumour location. Heterogeneity between studies was determined using the \(I^2\) statistic.\(^{(19)}\) Publication bias was assessed using funnel plots.
Results

Population-based cohort study

The number of patients in the cohort for whom a surgical resection specimen was retrieved for molecular analyses was 661. Nine patients had a known hereditary cancer syndrome: Lynch syndrome (n=6), familial adenomatous polyposis (n=1) and other familial syndrome (n=2). These patients were excluded from the analyses, leaving a cohort of 652 patients with presumed sporadic CRC.

Patients under the age of 50 (n=35) comprised 5.4% of our cohort of all stage II and III sporadic colon adenocarcinoma patients within the jurisdiction of the Northern Ireland Biobank. The demographic and clinical characteristics of the overall cohort are summarised in Supplementary Table 1.

The distribution of mutations and molecular features by age category is shown in Figure 1. EOCRC patients did not have any \textit{BRAF} or \textit{NRAS}-mutant tumours, and the proportions of these features across age categories was significant (p<0.01 and p=0.01 respectively). EOCRC had the highest proportion of MSI-H tumours (25.7%) and \textit{PIK3CA} mutations (25.7%) of all the age groups, but this did not reach statistical significance.

The results of survival analyses are shown in Table 1. Compared with 60-69 year old patients, EOCRC did not have a significantly increased risk of CRC death in stage II/III disease (adjusted HR 1.20, 95%CI 0.61-2.39). Compared to patients with microsatellite stable (MSS) tumours, patients with MSI-H tumours had a significantly reduced risk of CRC death (unadjusted HR 0.66, 95%CI 0.45-0.97) in stage II/III disease. In multivariable analysis, patients with MSI-H
tumours had a reduced risk of CRC death, but this was not statistically significant (HR 0.71, 95%CI 0.47-1.09). Subgroup analyses of survival by MSI status are shown in Figure 2. EOCRC patients with MSI-H tumours did not have a significantly decreased risk of CRC death (adjusted HR 0.65 95% CI 0.07-6.32). EOCRC patients with MSS tumours did not have a significantly increased risk of CRC death compared to 60-69 year olds (adjusted HR 1.58, 95%CI 0.71-3.51).

**Systematic review and meta-analysis**

The search strategy identified 2415 studies, and 1088 duplicates were removed by Covidence leaving 1327 articles for screening. Following title and abstract screening, 170 studies were eligible for full text review. Following full text review, 140 studies were excluded for the reasons outlined in Figure 3. Thirty-two articles were included in the review (31 resulting from our search strategy and 1 resulting from our population-based cohort study described above, referred to herein as *Hamilton et al*, 2022). The characteristics of the included studies are summarised in Supplementary Table 2. The rationale for determination of sporadic cases in each study is shown in Supplementary Table 3.

**Microsatellite instability / deficient mismatch repair**

Twenty-six studies reported data on tumour MSI status,(20-44) including our population-based cohort study. Fifteen studies reported data on expression of MMR proteins, indicating MMR status.(21, 23, 24, 26, 27, 33, 37, 41, 43, 45-50) A combined meta-analysis was performed of studies that had MSI and/or MMR results and this is shown in Figure 4A. Pooled analysis revealed a prevalence of 10% (95%CI 7-14%) of MSI-H/dMMR in presumed sporadic EOCRC. Observed heterogeneity was high (I² 85.73%, p<0.01). Separate meta-analyses were carried out for MSI-H and dMMR tumours (Supplementary Figures 1 and 2).
**BRAF mutations**

Nine studies reported data on BRAF mutations,\(^{(22, 27, 31, 33, 37, 43, 44, 48)}\) including our population-based cohort study. Pooled analysis revealed a prevalence of 1% (95%CI 0-3%) of BRAF mutations in sporadic EOCRC (Figure 4B). Observed heterogeneity was moderate (\(I^2 36.78\%\), \(p=0.12\)).

**KRAS mutations**

Thirteen studies reported data on KRAS mutations,\(^{(22, 24, 27, 31-33, 37, 38, 40, 43, 44, 48)}\) including our population-based cohort study. Pooled analysis revealed a prevalence of 32% (95%CI 23-40%) of KRAS mutations in sporadic EOCRC (Figure 4C). Observed heterogeneity was high (\(I^2 84.52\%\), \(p<0.01\)).

**NRAS mutations**

Four studies reported data on NRAS mutations,\(^{(43, 44, 48)}\) including our population-based cohort study. Pooled analysis revealed a prevalence of 3% (95%CI 1-4%) of NRAS mutations in sporadic EOCRC (Figure 4D). Observed heterogeneity was low (\(I^2 0.00\%\), \(p=0.42\)).

**PIK3CA mutations**

Five studies reported data on PIK3CA mutations,\(^{(22, 31, 44, 48)}\) including our population-based cohort study. Pooled analysis revealed a prevalence of 14% (95%CI 5-25%) of PIK3CA mutations in sporadic EOCRC (Figure 4E). Observed heterogeneity was high (\(I^2 83.31\%\), \(p<0.01\)).
**TP53 mutations**

Seven studies reported data on TP53 mutations.\(^{22-24, 31, 38, 40, 44}\) Pooled analysis revealed a prevalence of 64% (95%CI 56-71%) of TP53 mutations in sporadic EOCRC (Figure 4F). Observed heterogeneity was moderate (I\(^2\) 57.80%, p=0.03).

**Sensitivity and subgroup analyses**

Sensitivity analyses were performed by excluding one study at a time for each of the mutations and MSI/MMR status, to assess the robustness of the results. Results are shown in Supplementary Table 5. Excluding some studies had a marginal effect on heterogeneity for BRAF, PIK3CA, TP53 and NRAS mutations, but the prevalence of these mutations remained similar.

Subgroup analyses were undertaken by sex for MSI/MMR status, KRAS, BRAF and PIK3CA mutations (Table 2). There was insufficient data to undertake subgroup analyses for TP53 and NRAS mutations. The prevalence of MSI-H/dMMR, BRAF and PIK3CA mutations were similar in males and females. The prevalence of KRAS mutations was slightly higher in females (35%) than males (27%).

Subgroup analyses were undertaken by tumour location (Table 2). Analyses for colon were performed for MSI/MMR status, KRAS, BRAF and PIK3CA mutations, and for rectum by MSI/MMR status. There were insufficient data for TP53 and NRAS mutations for any subgroup analysis by tumour location, or for KRAS, BRAF and PIK3CA mutations for rectal cancer. Results showed that the prevalence of MSI-H/dMMR was higher in the colon compared to the rectum (16% v 6%) and in the right colon compared to the left colon (32% v 3%). The prevalence of
KRAS and BRAF mutations were similar in right and left colon, while PIK3CA mutations showed a higher prevalence in the right colon compared to the left colon (18% v 1%).

Publication bias

Publication bias was assessed using funnel plots where we plotted the proportion against the study size (Supplementary Figure 3). No evidence of publication bias was detected.

Quality assessment

Quality assessment was done using the Joanna Briggs Institute Checklist for prevalence studies, and this is shown in Supplementary Table 6. No studies were excluded based on quality assessment.
Discussion

To our knowledge this is the first systematic review investigating the prevalence of MSI-H/dMMR status and somatic mutations in sporadic EOCRC. The importance of distinguishing between sporadic and hereditary EOCRC is becoming increasingly recognised, with the molecular pathogenesis, treatment response and outcomes of sporadic EOCRC less understood than hereditary cases.

MSI/MMR status

Our systematic review shows that MSI-H/dMMR has a prevalence of 10% in sporadic EOCRC. These results were consistent across analyses for MSI-H tumours and dMMR tumours, suggesting that MSI and MMR status are highly correlated and that our findings are robust. Other studies have shown near-perfect concordance between immunohistochemistry testing for MMR proteins and MSI testing.[10]

MSI-H/dMMR CRCs are encountered in two clinical settings, representing the phenotypic convergence of two clinically distinct pathogeneses. Firstly, such tumours are the hallmark of CRC arising in the context of Lynch syndrome, resulting from a germline mutation in one of the MMR genes, most commonly MLH1 or MSH2, and young age at cancer diagnosis is regarded as an indicator for a possible hereditary cause of the disease. Our results show that 10% of MSI-H/dMMR tumours in EOCRC do not arise from Lynch syndrome. However, given the historical case series in reported studies it is possible that some cases of seemingly sporadic EOCRC in our review may have undiagnosed Lynch syndrome or Lynch-like syndrome.[51-53]
Secondly, MSI-H tumours comprise a proportion of sporadic CRC and are these are considerably more common than Lynch syndrome-related MSI-H CRCs. Sporadic MSI-H CRCs arise via the serrated neoplasia pathway and are strongly associated with BRAF mutations, older age, right sided tumour location and high levels of CIMP.\(^{(54)}\) Given the extremely low prevalence of BRAF mutations in EOCRC, it is unlikely the serrated pathway is the mechanism by which MSI-H tumours develop in younger patients. Similar to Lynch syndrome, results from subgroup analysis suggest that sporadic MSI-H EOCRCs also have a predilection for the right colon.

CIMP-high CRCs are associated with older aged patients, female sex, proximal tumour location, MSI-H status and somatic BRAF mutation.\(^{(55)}\) Evidence regarding CIMP in EOCRC is sparse, but CIMP-high tumours appear to be less prevalent in younger patients with CRC.\(^{(56)}\) However, future studies are required to elucidate the role of CIMP in sporadic EOCRC.

**BRAF mutations**

BRAF mutations occur in approximately 8% of all CRC, the vast majority being V600E mutations.\(^{(57)}\) Our results show that the prevalence of BRAF mutations in sporadic EOCRC is 1%, which is a lower proportion compared to LOCRC, and indicates this is a rare mutation in younger adults. Given the rarity of BRAF mutations in EOCRC, our results also suggest the association of BRAF mutations with MSI-H tumours seen in CRC in older age groups\(^{(58)}\) does not apply in these younger patients.

BRAF mutations are a negative prognostic marker, with worse survival outcomes being reported in metastatic disease.\(^{(59)}\) However, our results suggest a low prevalence of BRAF
mutations in EOCRC does not necessarily translate into better survival in this group, and the reasons for this are unclear. In 2021 NICE approved the use of encorafenib, a \textit{BRAF} inhibitor, in \textit{BRAF} V600E-mutation positive metastatic CRC in the UK.\cite{60} However, given the rarity of \textit{BRAF} mutations in EOCRC it is likely only a small proportion of young patients will be able to avail of this treatment, and optimal treatment strategies for EOCRC remain to be determined.

\textit{RAS} mutations

Our results show that the prevalence of \textit{KRAS} mutations in sporadic EOCRC is 32%. A large Memorial Sloan Kettering Cancer Centre study published after our literature search reported a prevalence of 42.5% for \textit{KRAS} mutations in sporadic EOCRC.\cite{61} Together with our meta-analysis, these results are comparable to a systematic review investigating \textit{KRAS} mutations in metastatic CRC, which reported a pooled prevalence of 35.9%.\cite{62} This suggests that the prevalence of \textit{KRAS} mutations is broadly similar in EOCRC and LOCRC.

Knowledge regarding \textit{NRAS} mutations in CRC is limited due to its low frequency. A systematic review found a prevalence of 4.1% (95% CI 3.5%-4.8%) of \textit{NRAS} mutations in metastatic CRC in all ages.\cite{62} Our results show that the prevalence of \textit{NRAS} mutations in tumours in sporadic EOCRC is 3%, suggesting the prevalence of \textit{NRAS} mutations is similar in young and older patients.

\textit{PIK3CA} mutations

A 2020 systematic review reported \textit{PIK3CA} mutations had a prevalence of 12.9% in CRC in patients of all ages.\cite{63} Our results show that the prevalence of \textit{PIK3CA} mutations in tumours in sporadic EOCRC is 14%, suggesting that the proportion of \textit{PIK3CA} mutations is similar in
EOCRC and LOCRC. However, knowledge regarding PIK3CA mutations in EOCRC remains limited, as shown by the small number of studies in our meta-analysis. PIK3CA mutations currently have no clinical role as predictive or prognostic biomarkers, with a previous systematic review and meta-analysis finding no significant association between PIK3CA mutation status and survival outcomes.(64)

**TP53 mutations**

Mutation of TP53 is a late event in the stepwise development of CRC, most commonly via the CIN pathway.(8) TP53 mutations have been shown to be present in up to 60% of CRCs.(65) Our results show that the prevalence of TP53 mutations in sporadic EOCRC is 64%, the highest prevalence of any mutation in this study. Similar findings were observed in a whole-exome sequencing study which found TP53 was the most common mutation in EOCRC,(66) with subsequent targeted deep sequencing (n=833) showing a higher frequency of TP53 mutation in EOCRC compared to LOCRC (80% v 72%, Fisher’s exact p=0.03). TP53 is currently is not used as a prognostic or predictive biomarker in clinical practice, and more research is required into its clinical implications.

**Survival**

Results from our population-based cohort study indicate that stage II and III sporadic EOCRC does not have a significantly worse survival compared to patients aged 60-69 years, but indicate there may be an aggressive subset within this young age group, driven by MSS tumours. CRCs displaying MSI have a better prognosis in early stage disease, with improved 5 year overall survival(67) and 5 year recurrence free survival,(68) but less expected benefit from adjuvant chemotherapy.(67, 69) Results from our population-based cohort study support the
conclusion that non-metastatic MSI-H tumours in sporadic EOCRC patients also carry a better prognosis than MSS tumours, although sample sizes were limited and so results were not statistically significant. We are unable to draw any conclusions about survival in metastatic sporadic EOCRC from this study.

Studies have shown that EOCRC patients have a more advanced stage at presentation than older patients, (70) which could be due to more aggressive biology or a delay in diagnosis. Delayed diagnosis may be caused by a number of factors, including failure of younger patients to seek healthcare, a delay in referral by healthcare professionals, or the exclusion of younger individuals from bowel cancer screening programmes.

Further research is urgently needed on outcomes for patients with EOCRC and more specifically on the impact of tumour molecular profile on survival, particularly how this varies by stage of disease.

Strengths and limitations

Our study has a number of strengths. To our knowledge, this the first systematic review that determines prevalence of key mutations and MSI-H/dMMR in sporadic EOCRC. Quality assessment was undertaken using the Joanna Briggs Institute Checklist for Prevalence Studies, which was felt to be rigorous in a recent systematic review. (71) Sensitivity analyses demonstrated largely stable heterogeneity, particularly for MSI-H/dMMR tumours.

One weakness of the systematic review is that despite attempts to ensure that only sporadic EOCRC cases were included, there may be undiagnosed hereditary cases in our review.
However, our methodology was rigorous and studies were excluded if the information provided allowed us to confidently separate, as far as possible, sporadic from hereditary cases. For meta-analyses we have used the Freeman-Tukey Arcsine Transformation method. A weakness of this method is that it breaks down with extremely sparse data.\(^{72}\) To ensure the accuracy of our results, we also carried out meta-analyses using the logistic-normal random-effects model, which showed similar pooled proportions and 95% CI. In addition, we were unable to undertake subgroup analysis by stage or race/ethnicity.

The studied molecular characteristics of CRC vary with stage of disease. For example, MSI-H occurs in approximately 15-20% of stage II and III CRC, but is less common in metastatic CRC, occurring in approximately 4% of cases.\(^{73, 74}\) BRAF mutations are associated with an advanced stage of disease.\(^{75}\) Insufficient information was available to undertake subgroup analyses by stage in our study, but variation in molecular profile by stage may account for some of the observed differences between studies. This is an important issue to address in future studies.

Studies investigating the molecular profile of rectal cancer are lacking and, within this subgroup, we were only able to undertake analysis for MSI/MMR status as the number of studies describing the selected mutations in rectal cancers was insufficient. Further research is needed into how rectal cancer differs from colon cancer in terms of mutational profile. In addition, newly discovered germline mutations in genes such as POLE and POLD1 will be contributing to a small percentage of EOCRC.\(^{76}\) However, whilst POLE/POLD1-mutated CRC shares some features with MSI-H CRCs (such as a high tumour mutation burden), they are MSS tumours and are unlikely to account for any of the sporadic MSI-H EOCRC cases in our meta-analysis.\(^{77}\)
Conclusion

In conclusion, this systematic review addresses a research gap regarding sporadic EOCRC, and provides evidence of differing molecular profiles in younger patients with CRC compared to older patients. Approximately 10% of seemingly sporadic EOCRC are MSI-H, and \textit{BRAF} mutations are a rare event in these tumours, having a much lower prevalence than in LOCRC. \textit{KRAS}, \textit{NRAS}, \textit{PIK3CA} and \textit{TP53} mutations have a similar prevalence to LOCRC. The molecular pathogenesis of sporadic EOCRC remains unclear, with the serrated neoplasia pathway unlikely to play a major role. EOCRC patients were not at increased risk of cancer-specific death compared with older patients in our population-based cohort, but further studies are needed to address whether molecular profiles differentially influence EOCRC patient outcomes in this patient group.
Figure legends

Figure 1. Distribution of molecular characteristics by age category in sporadic stage II and III colon cancer

Figure 2. Risk of colorectal cancer-specific death according to age categories in stage II and III colon cancer cases, by microsatellite instability status

Figure 3. Flow chart of the selection of articles included in the review

Figure 4A. Forest plot illustrating meta-analysis of the prevalence of microsatellite instability-high/deficient mismatch repair tumours in sporadic early onset colorectal cancer

Figure 4B. Forest plot illustrating meta-analysis of the prevalence of \textit{BRAF} mutations in sporadic early onset colorectal cancer

Figure 4C. Forest plot illustrating meta-analysis of the prevalence of \textit{KRAS} mutations in sporadic early onset colorectal cancer

Figure 4D. Forest plot illustrating meta-analysis of the prevalence of \textit{NRAS} mutations in sporadic early onset colorectal cancer

Figure 4E. Forest plot illustrating meta-analysis of the prevalence of \textit{PIK3CA} mutations in sporadic early onset colorectal cancer

Figure 4F. Forest plot illustrating meta-analysis of the prevalence of \textit{TP53} mutations in sporadic early onset colorectal cancer

Table 1. Survival analysis by age category in sporadic stage II and III colon cancer

Table 2. Subgroup analysis by sex and tumour location
References


60. NICE. Encorafenib plus cetuximab for previously treated BRAF V600E mutation-positive metastatic colorectal cancer 2021 [Available from: https://www.nice.org.uk/guidance/ta668].


<table>
<thead>
<tr>
<th>Age (Years)</th>
<th>No of CRC deaths/patients</th>
<th>CRC Death</th>
<th>Overall Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No of deaths/CRC patients</td>
<td>Unadjusted HR (95%CI)</td>
</tr>
<tr>
<td>&lt;50</td>
<td>11/34</td>
<td>12/35</td>
<td>1.14 (0.59-2.18)</td>
</tr>
<tr>
<td>50-59</td>
<td>16/58</td>
<td>19/61</td>
<td>0.83 (0.47-1.45)</td>
</tr>
<tr>
<td>60-69</td>
<td>52/170</td>
<td>61/179</td>
<td>1 (reference)</td>
</tr>
<tr>
<td>70-79</td>
<td>70/193</td>
<td>115/238</td>
<td>1.04 (0.73-1.49)</td>
</tr>
<tr>
<td>≥80</td>
<td>61/104</td>
<td>96/139</td>
<td>1.96 (1.35-2.84)</td>
</tr>
</tbody>
</table>

* multivariable model adjusted HR adjusted for sex, adjuvant chemotherapy receipt, stage, tumour differentiation, family history of CRC and ECOG performance status, alcohol, smoking, inflammatory bowel disease and emergency surgery

** multivariable model adjusted for all variables in footnote* and Charlson comorbidity score
Table 2. Subgroup analysis by sex and tumour location

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Studies included</th>
<th>Pooled analysis (95%CI)</th>
<th>I² (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High/dMMR</td>
<td>9</td>
<td>0.16 (0.09-0.24)</td>
<td>55.53</td>
<td>0.02</td>
</tr>
<tr>
<td>BRAF</td>
<td>3</td>
<td>0.02 (0.00-0.09)</td>
<td>27.43</td>
<td>0.25</td>
</tr>
<tr>
<td>KRAS</td>
<td>4</td>
<td>0.27 (0.16-0.38)</td>
<td>0.00</td>
<td>0.78</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>3</td>
<td>0.11 (0.00-0.35)</td>
<td>79.80</td>
<td>0.01</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High/dMMR</td>
<td>9</td>
<td>0.14 (0.07-0.21)</td>
<td>46.86</td>
<td>0.06</td>
</tr>
<tr>
<td>BRAF</td>
<td>3</td>
<td>0.02 (0.00-0.13)</td>
<td>55.04</td>
<td>0.11</td>
</tr>
<tr>
<td>KRAS</td>
<td>4</td>
<td>0.35 (0.23-0.48)</td>
<td>0.00</td>
<td>0.54</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>3</td>
<td>0.13 (0.00-0.40)</td>
<td>81.83</td>
<td>0.00</td>
</tr>
<tr>
<td>All colon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High/dMMR</td>
<td>7</td>
<td>0.16 (0.06-0.29)</td>
<td>74.74</td>
<td>0.00</td>
</tr>
<tr>
<td>BRAF</td>
<td>3</td>
<td>0.04 (0.00-0.14)</td>
<td>58.32</td>
<td>0.09</td>
</tr>
<tr>
<td>KRAS</td>
<td>3</td>
<td>0.34 (0.24-0.45)</td>
<td>0.00</td>
<td>0.72</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>3</td>
<td>0.12 (0.00-0.37)</td>
<td>85.80</td>
<td>0.00</td>
</tr>
<tr>
<td>Right colon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High/dMMR</td>
<td>6</td>
<td>0.32 (0.19-0.46)</td>
<td>43.95</td>
<td>0.11</td>
</tr>
<tr>
<td>BRAF</td>
<td>3</td>
<td>0.02 (0.00-0.14)</td>
<td>54.22</td>
<td>0.11</td>
</tr>
<tr>
<td>KRAS</td>
<td>3</td>
<td>0.35 (0.22-0.50)</td>
<td>0.00</td>
<td>0.84</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>3</td>
<td>0.18 (0.01-0.46)</td>
<td>77.75</td>
<td>0.01</td>
</tr>
<tr>
<td>Left colon</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High/dMMR</td>
<td>6</td>
<td>0.03 (0.00-0.12)</td>
<td>47.82</td>
<td>0.09</td>
</tr>
<tr>
<td>BRAF</td>
<td>3</td>
<td>0.04 (0.00-0.17)</td>
<td>17.24</td>
<td>0.30</td>
</tr>
<tr>
<td>KRAS</td>
<td>3</td>
<td>0.33 (0.17-0.51)</td>
<td>0.00</td>
<td>0.76</td>
</tr>
<tr>
<td>PIK3CA</td>
<td>3</td>
<td>0.01 (0.00-0.10)</td>
<td>0.00</td>
<td>0.50</td>
</tr>
<tr>
<td>Rectum</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High/dMMR</td>
<td>6</td>
<td>0.06 (0.01-0.13)</td>
<td>12.59</td>
<td>0.33</td>
</tr>
</tbody>
</table>
CI: confidence intervals; HR: hazard ratio; MSI: microsatellite instability

** adjusted for sex, adjuvant chemotherapy receipt, stage, tumour differentiation, family history of CRC, ECOG performance status, smoking, alcohol, inflammatory bowel disease and emergency surgery

<table>
<thead>
<tr>
<th>Microsatellite Instability Status</th>
<th>Adjusted HR** (95%CI)</th>
<th>CRC deaths/patients</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>&lt;50 years old</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High</td>
<td>0.65 (0.07-6.32)</td>
<td>1/8</td>
</tr>
<tr>
<td>Non MSI-High</td>
<td>1.58 (0.71-3.51)</td>
<td>9/21</td>
</tr>
<tr>
<td><strong>50-59 years old</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High</td>
<td>0.61 (0.06-6.29)</td>
<td>1/7</td>
</tr>
<tr>
<td>Non MSI-High</td>
<td>0.99 (0.52-1.86)</td>
<td>14/46</td>
</tr>
<tr>
<td><strong>60-69 years old</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High</td>
<td>1.0 (reference)</td>
<td>9/31</td>
</tr>
<tr>
<td>Non MSI-High</td>
<td>1.0 (reference)</td>
<td>39/128</td>
</tr>
<tr>
<td><strong>70-79 years old</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High</td>
<td>0.38 (0.12-1.23)</td>
<td>9/46</td>
</tr>
<tr>
<td>Non MSI-High</td>
<td>1.04 (0.67-1.62)</td>
<td>52/133</td>
</tr>
<tr>
<td><strong>≥80 years old</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSI-High</td>
<td>0.79 (0.19-3.26)</td>
<td>11/21</td>
</tr>
<tr>
<td>Non MSI-High</td>
<td>1.58 (0.98-2.55)</td>
<td>45/74</td>
</tr>
</tbody>
</table>
Records identified through database search n=2415

Duplicates removed n=1088

Records screened n=1327

Studies excluded (n=140)
   Reasons:
   - Wrong patient population (n=60)
   - Abstract only (n=51)
   - Wrong intervention (n=9)
   - Wrong study design (n=9)
   - Wrong outcomes (n=7)
   - Authors did not respond to request for further information (n=1)
   - Clinical news report (n=1)
   - Duplicate (n=2)

Studies included n=31

Included Epi700 study n=1

Studies included n=32

Full text studies assessed for eligibility n=171
Studies Ak et al through to Hamilton et al have presented microsatellite instability-high data, determined by PCR; studies Aitchison et al through to Suzuki et al have used mismatch repair immunohistochemistry. Details of molecular testing for each study are found in Supplementary Table 4.

ES: effect size (equivalent to proportion)
BRAF

<table>
<thead>
<tr>
<th>Study</th>
<th>ES (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berg et al, 2010</td>
<td>0.07 (0.02, 0.18)</td>
<td>10.34</td>
</tr>
<tr>
<td>Goel et al, 2010</td>
<td>0.00 (0.00, 0.05)</td>
<td>14.52</td>
</tr>
<tr>
<td>Jiang et al, 2020</td>
<td>0.05 (0.01, 0.17)</td>
<td>9.67</td>
</tr>
<tr>
<td>Kirzin et al, 2014</td>
<td>0.00 (0.00, 0.09)</td>
<td>9.50</td>
</tr>
<tr>
<td>Magnani et al, 2015</td>
<td>0.00 (0.00, 0.13)</td>
<td>6.79</td>
</tr>
<tr>
<td>Pilozzi et al, 2015</td>
<td>0.00 (0.00, 0.15)</td>
<td>6.14</td>
</tr>
<tr>
<td>Watson et al, 2016</td>
<td>0.00 (0.00, 0.19)</td>
<td>4.72</td>
</tr>
<tr>
<td>Willauer et al, 2019</td>
<td>0.04 (0.03, 0.06)</td>
<td>28.52</td>
</tr>
<tr>
<td>Hamilton et al, 2022</td>
<td>0.00 (0.00, 0.10)</td>
<td>8.78</td>
</tr>
<tr>
<td>Overall (*$^2 = 36.78%$, p = 0.12)</td>
<td>0.01 (0.00, 0.03)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Proportion

Overall
TP53

<table>
<thead>
<tr>
<th>Study</th>
<th>ES (95% CI)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berg et al, 2010</td>
<td>0.64 (0.50, 0.77)</td>
<td>14.37</td>
</tr>
<tr>
<td>Dieumegard et al, 2000</td>
<td>0.93 (0.69, 0.99)</td>
<td>6.86</td>
</tr>
<tr>
<td>Fernebro et al, 2002</td>
<td>0.60 (0.39, 0.78)</td>
<td>8.85</td>
</tr>
<tr>
<td>Kirzin et al, 2014</td>
<td>0.44 (0.29, 0.59)</td>
<td>13.33</td>
</tr>
<tr>
<td>Raman et al, 2014</td>
<td>0.67 (0.57, 0.76)</td>
<td>19.57</td>
</tr>
<tr>
<td>Soliman et al, 2001</td>
<td>0.57 (0.37, 0.74)</td>
<td>9.71</td>
</tr>
<tr>
<td>Willauer et al, 2019</td>
<td>0.65 (0.61, 0.69)</td>
<td>27.31</td>
</tr>
<tr>
<td>Overall (I^2 = 57.80%, p = 0.03)</td>
<td>0.64 (0.56, 0.71)</td>
<td>100.00</td>
</tr>
</tbody>
</table>