JYU DISSERTATIONS 821

Tero Sievänen

The Associations of Circulating MicroRNAs and Lifestyle Habits with Cancer Risk in Lynch Syndrome

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF SPORT AND
HEALTH SCIENCES

JYU DISSERTATIONS 821

Tero Sievänen

The Associations of Circulating MicroRNAs and Lifestyle Habits with Cancer Risk in Lynch Syndrome

Esitetään Jyväskylän yliopiston liikuntatieteellisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi päärakennuksen salissa C4
syyskuun 20. päivänä 2024 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Sport and Health Sciences of the University of Jyväskylä,
in Main Building, hall C4, on September 20, 2024, at 12 o'clock.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2024

Editors
Anne Viljanen
Faculty of Sport and Health Sciences, University of Jyväskylä
Päivi Vuorio
Open Science Centre, University of Jyväskylä

ISBN 978-952-86-0284-2
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-952-86-0284-2

ABSTRACT

Sievänen, Tero
The associations of circulating microRNAs and lifestyle habits with cancer risk in Lynch syndrome
Jyväskylä: University of Jyväskylä, 2024, 102 p. + original articles
(JYU Dissertations
ISSN 2489-9003; 821)
ISBN 978-952-86-0284-2 (PDF)

Lynch syndrome (LS) is the most common hereditary cancer syndrome. This thesis explored the associations between circulating microRNAs (c-miRs), lifestyle habits, and the incidence of LS cancer. By utilizing high-throughput sequencing and bioinformatic approaches, the aims of this thesis were to characterize the serum-based c-miR landscape of cancer-free LS carriers to inspect whether any of those c-miRs are potential indicators of upcoming colorectal cancer (CRC), and to determine whether they are associated with modifiable CRC risk factors, such as body mass index and physical activity. Furthermore, this thesis applied retrospective lifestyle questionnaire data to investigate whether longitudinal body weight gain and physical activity are associated with LS cancer risk. It was observed that cancer-free LS carriers (n = 101) displayed aberrant serum c-miR expression compared to the control group (n = 37), but not when compared to sporadic CRC patients (n = 24). A panel composed of these aberrantly expressed c-miRs, including hsa-miR-10b-5p, hsa-miR-19b-3p, hsa-miR-27b-3p, hsa-miR-200a-3p, and hsa-miR-3615, predicted CRC incidence in a prospective analysis. These findings indicated that c-miR profile mirrors early-stage carcinogenesis and may have risk stratification potential during surveillance. The CRC predictive c-miRs did not correlate with either body mass index or physical activity, suggesting that they are associated with LS CRC risk independently of lifestyle habits. However, in the retrospective analysis (n = 465), adulthood weight gain was seen as a cancer risk factor for males, whereas near-term weight was a protective factor for females. Longitudinal physical activity was associated with a decreased overall cancer risk in male LS carriers. Further research is required to validate these findings and to elucidate the complex factors underlying lifestyle and LS cancer.

Keywords: colorectal cancer, microRNA, lifestyle, Lynch syndrome

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Sievänen, Tero
Verenkierron mikro-RNA:n, elämäntapojen ja syöpäriskin yhteydet Lynchin oireyhtymässä
Jyväskylä: Jyväskylän yliopisto, 2024, 102 s. + alkuperäiset artikkelit
(JYU Dissertations
ISSN 2489-9003; 821)
ISBN 978-952-86-0284-2 (PDF)

Lynchin oireyhtymä (LS) on perinnöllinen syöpäalttiusoireyhtymä, joka altistaa yksilön useille syöville. Tämä väitöskirjatutkimus seuloi sekvensointimenetelmien sekä bioinformatiikan avulla syöpävapaiden suomalaisten LS-kantajien (n = 101) verenkierron mikro-RNA (c-miR) profiileita vertaamalla niitä ei-perinnöllisten suolistosyöpäpotilaiden (n = 24) sekä terveiden verrokkien (n = 37) vastaaviin profiileihin. Pitkittäisasetelmassa tutkittiin, voitiinko tunnistettujen c-miR:n avulla ennustaa suolistosyöpään sairastumista, ja ovatko c-miR:t yhteydessä kehon painoon ja fyysiseen aktiivisuuteen. Lisäksi tehtiin retrospektiivinen elämäntapakysely (n = 465), jolla selvitettiin, olivatko kehon paino tai fyysinen aktiivisuus yhteydessä syöpäriskiin. Syöpävapaiden LS-kantajien c-miR-profiilien havaittiin poikkeavan merkittävästi terveistä verrokeista, mutta ei suolistosyöpäpotilaista. Näistä viisi miR:ta, hsa-miR-10b-5p, hsa-miR-19b-3p, hsa-miR-27b-3p, hsa-miR-200a-3p ja hsa-miR-3615, muodostivat riskiennustepaneelin, joka ennusti suolistosyöpään sairastumista neljän vuoden seurannan aikana, mutta ei ollut yhteydessä elämäntapatekijöihin. Retrospektiivinen analyysi osoitti, että aikuisiän painonnousu lisää syöpäriskiä miehillä, kun taas lyhyen aikavälin painonnousun huomattiin alentavan naisten suolistosyöpäriskiä. Lisäksi havaittiin, että fyysinen aktiivisuus voi suojata erityisesti miespuolisia LS-kantajia syöviltä. Saadut tulokset osoittivat, että terveiden LS-kantajien c-miR-profiilin muutokset kuvastavat aikaisen vaiheen suolistosyövän kehittymistä ja ennustavat siihen sairastumista. Näin ollen c-miR-profiilit voivat olla potentiaalisia merkkiaineita, jotka voitaisiin yhdistää olemassa oleviin seulontatyökaluihin korkean syöpäriskin potilaiden ohjaamiseksi intensiivisempään seurantaan. Tutkimuksessa havaittiin myös, että painonhallinta ja fyysinen aktiivisuus voivat suojata erityisesti LS-kantajamiehiä syöviltä. Lisätutkimusta aiheesta tarvitaan saatujen tulosten vahvistamiseksi.

Asiasanat: elämäntavat, Lynchin oireyhtymä, mikro-RNA, suolistosyöpä

Author

Tero Sievänen, MSc
Gerontology Research Center
Faculty of Sport and Health Sciences
University of Jyväskylä
Finland
tero.o.sievanen@jyu.fi
ORCID: 0000-0002-9660-4559

Supervisors

Associate Professor Eija Laakkonen, PhD
Gerontology Research Center
Faculty of Sport and Health Sciences
University of Jyväskylä
Finland

Associate Professor Toni Seppälä, MD, PhD
Faculty of Medicine and Health Technology
University of Tampere
Finland

Associate Professor Elina Sillanpää, PhD
Gerontology Research Center
Faculty of Sport and Health Sciences
University of Jyväskylä
Wellbeing Services County of Central Finland
Finland

Reviewers

Associate Professor Valerio Izzi, PhD
Faculty of Biochemistry and Molecular Medicine
University of Oulu
Finland

Professor Mef Nilbert, PhD
Institute of Clinical Sciences
Division of Oncology and Pathology
University of Lund
Sweden

Opponent

Professor Arto Mannermaa, PhD
Institute of Clinical Medicine
Faculty of Medicine
University of Eastern Finland
Finland

ACKNOWLEDGEMENTS

This thesis is the 52nd extension of the long continuum of the Finnish Lynch syndrome research tradition conducted at the Faculty of Sport and Health Sciences at the University of Jyväskylä between 2019 and 2024. While the initial articles encompassed various cancer types within the Lynch syndrome cohort, this thesis mainly focuses on colorectal cancer. The decision was driven by the recognition that colorectal cancer is the predominant cancer associated with Lynch syndrome. Given its prevalence in our sample, I believe that focusing on colorectal cancer yielded results that could be more readily extrapolated to other conditions associated with Lynch syndrome. In addition, the results of this thesis and other relevant studies could be compared in a more straightforward manner.

Here, I would like to extend my gratitude to those involved in this project, and especially to those who made the undertaking possible in the first place.

First, I would like to thank my supervisors, Eija Laakkonen, Toni Seppälä, and Elina Sillanpää, for their encouragement and guidance along the way. Eija, without your belief in me in 2017 when I applied for a position at the MonEx project, I probably would not have done this thesis. Your broad interdisciplinary scientific knowledge, ranging from genetics to non-coding RNAs and beyond, your insight into human character, and your warm personality have truly made me a proper scientist. You will always have my deepest gratitude. Toni, it has been a great pleasure to have you involved and to draw inspiration and ideas from your immense insight into cancer research and career development. Thank you very much. Elina, your vast knowledge on gerontology and biological aging, and our inspiring discussions about the world of science and beyond have inspired me along the way and taught me several skills I hold dear for pursuing a career in science—determination, diligence, and curiosity. A wholehearted thank you to you, Elina. I would also like to express my gratitude to all my supervisors for introducing me to their colleagues and contacts, which has given me an opportunity to travel and present my work all around the world.

Second, I wish to thank all the coauthors of the journal articles. It has been a great pleasure to draw inspiration from the expertise you all possess in various research disciplines. Jukka-Pekka, thank you for all the inspiring lectures on Lynch syndrome, and for the ever-warm welcome to the Finnish Lynch syndrome research community. Kirsi, thank you for your valuable help with data management. I would also like to thank the whole Molecular Pathology Department at Nova Hospital for their valuable help, especially Laura, Maarit, and Teijo. Tiina, Tia-Marje, and Matti, thank you for your endless patience and help with the data analyses and writing. Finally, I want to thank Timo and Juha K. for helping me with statistics.

I also extend my warm appreciation to the reviewers of my thesis, Prof. Valerio Izzi and Prof. Mef Nilbert, for dedicating their time to familiarizing themselves with my work and providing invaluable feedback. I am grateful to Prof. Arto Mannermaa for graciously agreeing to serve as the opponent in the public defense of this dissertation.

This thesis has been supported by the Faculty of Sport and Health Sciences, University of Jyväskylä, Nova Hospital Foundation, and the Finnish Cancer Foundation. My deepest gratitude to all the funders who made this possible.

Without my lovely colleagues and friends, there would be no thesis. Emilia, a heartfelt thank you for your support and kindhearted company from the freshmen years to this doctorate. Sira, Niko, Jari, Hanna-Kaarina, Anna, Elina, Sakari, Johanna, Emmi, and the rest of the Gerontology Research Center—I offer my sincere gratitude for everything, inside and outside the lab. I also want to thank all the other members of the faculty who helped me in many ways during this project—especially our lab personnel, Mervi and Hanne. You have been very valuable to my research. Thank you. I would like to extend my gratitude to Juha H. for being an encouraging and inspiring thesis committee member. Of course, I wish to express my sincerest gratitude to all my friends outside of science.

I would like to thank my mom, Sirpa, for your endless support and love throughout my life. You have taught me to always pursue my dreams and have given me the tools to reach for the stars. There are really no words that could describe how thankful I am about everything.

Finally, I wish to devote my sincerest and deepest gratitude to my lovely spouse, Anna, for accompanying me on this journey. I'm, again, lacking the words to express what your endless love, support, and encouragement have meant to me. If I ever doubted myself, you didn't. Thank you from the very bottom of my heart.

Jyväskylä 31.1.2024
Tero Sievänen

ORIGINAL PUBLICATIONS AND AUTHOR CONTRIBUTIONS

This thesis is based on the following three original publications, which will be referred to by their Roman numbers:

I. Sievänen T, Korhonen T-M, Jokela T, Ahtiainen M, Lahtinen L, Kuopio T, Lepistö A, Sillanpää E, Mecklin J-P, Seppälä TT, Laakkonen EK. 2023. Systemic circulating microRNA landscape in Lynch syndrome. *International Journal of Cancer*, *152*(5), 932–944. https://doi.org/10.1002/ijc.34338

II. Sievänen T, Jokela T, Hyvärinen M, Korhonen T-M, Pylvänäinen K, Mecklin J-P, Karvanen J, Sillanpää E, Seppälä TT, Laakkonen EK. 2024. Circulating miRNA signature predicts cancer incidence in Lynch syndrome—A pilot study. *Cancer Prevention Research*, *17*(6), 243-254. https://doi.org/10.1158/1940-6207.CAPR-23-0368

III. Sievänen T, Törmäkangas T, Laakkonen EK, Mecklin J-P, Pylvänäinen K, Seppälä TT, Peltomäki P, Sipilä S, Sillanpää E. Body weight, physical activity, and risk of cancer in Lynch syndrome. 2021. *Cancers*, *13*(8), 1849. https://doi.org/10.3390/cancers13081849

I was the single first author of Studies I and II where I drafted the manuscripts and made revisions based on the feedback from the coauthors and reviewers. For Study I, I conducted all the laboratory work, as well as the majority of bioinformatics and statistical analyses, with the help of T-M. Korhonen. Regarding Study II, I participated in designing the risk prediction model with J. Karvanen. I conducted all the bioinformatic analyses and the majority of the statistical analyses with the help of M. Hyvärinen. In Study III, I shared first authorship with T. Törmäkangas. I drafted and revised the manuscript and was responsible for all the descriptive statistical analyses, while Törmäkangas conducted the longitudinal modeling. All studies were designed in collaboration with my supervisors, E. Laakkonen, T. Seppälä, and E. Sillanpää.

FIGURES

TABLES

ABBREVIATIONS

AKT	Protein kinase 1
APC	Adenomatous polyposis coli
BMI	Body mass index
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BRCA	BRCA1/BRCA2 DNA repair associated
CDKN1A	Cyclin-dependent kinase inhibitor 1A
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CI	Confidence interval
c-miR	Circulating microRNA
CRC	Colorectal cancer
CREB	cAMP responsive element binding protein
CTNNB1	Catenin beta 1
DGCR8	DGCR8 microprocessor complex subunit
dMMR	Deficient mismatch repair
EGFR	Epidermal growth factor receptor
ERK	Extracellular signal-regulated kinase
ERMA	Estrogenic Regulation of Muscle Apoptosis study
FDR	False discovery rate
FOXO	Forkhead box O
HNPCC	Hereditary non-polyposis colorectal cancer
HR	Hazard ratio
hsa	Homo sapiens
KIT	KIT proto-oncogene, receptor tyrosine kinase
KRAS	KRAS proto-oncogene, GTPase
Lasso	Least absolute shrinkage and selection operator
LS	Lynch syndrome
LSRFi	Finnish Lynch Syndrome Research Registry
MAPK	Mitogen activated protein kinase
MET	Metabolic equivalent task
miR	MicroRNA
MLH1	mutL homolog 1
MMR	Mismatch repair
mRNA	Messenger RNA
MSH2	mutS homolog 2
MSH6	mutS homolog 6
MSI	Microsatellite instability
MSS	Microsatellite stability
MYC	MYC proto-oncogene
NF-Kb	Nuclear factor kappa B
NRAS	NRAS proto-oncogene
p53	Tumor protein p53
PACT	Protein activator of interferon induced protein kinase EIF2AK2
PIK3	Phosphatidylinositol-4,5-bisphosphate 3-kinase

PLSD	Prospective Lynch Syndrome Database
PMS2	PMS1 homolog 2
RAS	Rat sarcoma
RISC	RNA-induced silencing complex
SMAD	SMAD family member
STAT3	Signal transducer and activator of transcription 3
TGF/B	Transforming growth factor/ beta
TP53	Tumor protein p53
TRBP	TARBP2 subunit of RISC loading complex
TSG	Tumor suppressor gene
WNT	Proto-oncogene int-1 homolog

CONTENTS

1 INTRODUCTION

"No disease to which the human species is subject carries with it so formidable an appearance, or is productive of such dreadful consequences, as that which is called cancer. It has ever been the reproach of the medical art, and the most learned and experienced of the profession have employed their time and attention to but little purpose towards perfecting its cure."

Still, almost two and a half centuries after this description by Dr. Robert White—whose work in the late 1700s represents one of the earliest scientific reports on cancer (White, 1784)—cancer remains one of the most daunting medical challenges of the modern era. While the 19th century witnessed a surge in surgical interventions, the subsequent century ushered in a transformative era in cancer research by demonstrating the link between tobacco use and lung cancer, as well as the development of chemotherapy and radiation therapy as viable cancer treatments (Greenivald & Dunn, 2009). However, despite progress and improved understanding, the global burden of cancer continues to increase, placing substantial strain on individuals, families, communities, and health systems, both physically and financially.

Cancer is a general term for a large group of diseases whose causes, characteristics, and occurrences can vary greatly. To date, cancer is one of the leading causes of death globally (Sung et al., 2021). Regarding the global cancer burden, colorectal cancer (CRC) accounts for approximately 10% of global cancer cases and deaths annually, with rising incidence rates (Siegel et al., 2021; Sung et al., 2021). In Finland, 17% of diagnosed females and 27% of diagnosed males died of CRC in 2021 (Seppä et al., 2023). The majority of CRCs stem from polyps, evolving over an estimated 10 to 15 years from aberrant crypts to neoplastic polyps and eventually to CRC (Dekker et al., 2019). As with all cancers, the prognosis of CRC is better when the carcinoma is detected early (Seppä et al., 2023). Environmental factors, such as obesity and increased sedentary behavior, are known risk factors for several cancers (Sung et al., 2019). This growing inactivity manifests as an increased number of incident cancers (Siegel et al., 2021; Sung et al., 2019), although enhanced diagnostics via nationwide screening

initiatives, as well as the fact that the average human lifespan has risen drastically over the last century, also contribute to the number of detected cancers.

Lynch syndrome (LS) is the most common hereditary cancer syndrome, with an estimated prevalence of 1:279 (Haraldsdottir et al., 2017). LS is caused by pathogenic variants in the DNA mismatch repair genes *MLH1*, *MSH2*, *MSH6*, and *PMS2*, causing mutation accumulation and an increased risk of multiple cancers, especially CRC (Lynch et al., 2015). Therefore, LS carriers are offered lifelong surveillance and frequent cancer screening via colonoscopy with polypectomy as the standard of care. However, all pathogenic gene variants possess different clinical presentations, risk estimates, and molecular features, and thus LS presents a collection of diseases that should be treated exclusively (Møller et al., 2023). For example, CRC risk estimates for *MLH1* carriers are substantially higher than those of *PMS2* carriers (Dominguez-Valentin et al., 2020), and pathogenic *MLH1*-driven CRC displays several unique features, from natural history to clinical presentation, making screening and prevention procedures unevenly effective (Ahadova et al., 2021; Seppälä et al., 2023). Thus, there is an unmet need for more enhanced risk stratification to find patients who are at enhanced risk of developing CRC and who would most benefit from the screenings.

Recently, there has been a growing emphasis on personalized medicine and targeted therapies that seek to enhance treatment efficacy and improve outcomes by tailoring interventions to the unique and individualized factors inherent in each cancer patient (Ignatiadis et al., 2021; Mauri et al., 2022). MicroRNAs (miRs) are small non-coding RNAs that orchestrate several core processes of carcinogenesis and physiological responses through the regulation of gene expression (Mori et al., 2019). MiRs are stable in circulation and easily collected, thus showing considerable potential as minimally invasive cancer biomarkers (Francavilla et al., 2020; Jung et al., 2020; Sapp et al., 2017). Importantly, miRs have been shown to predict CRC incidence several years prior to diagnosis (Raut et al., 2021; Wikberg et al., 2018). Therefore, they might possess risk stratification potential that could be applied to identify LS carriers who are at increased risk of developing CRC in the near future. To date, the association between miRs and LS has not been extensively studied.

Although the practice of medicine is progressing toward more personalized treatment approaches, the avoidance of excess body weight and increased physical activity have been recurrently proven to be an effective "one-size-fits-all" approach for cancer prevention (de Rezende et al., 2018; Dixon, 2010; Friedenreich et al., 2010; Kitahara et al., 2013; Lee et al., 2012; Mctiernan et al., 2019; Papadimitriou et al., 2020; Sung et al., 2019). Excess body weight increases cancer risk through several biological mechanisms, including steroid hormone signaling, chronic inflammation, and insulin resistance (Bull et al., 2020). On the contrary, physical activity has been suggested to lower cancer risk by reducing body adiposity and improving the immune system and insulin sensitivity (Friedenreich et al., 2021). Interestingly, these factors represent modifiable lifestyle behaviors that influence the expression of several miRs, which are also

altered in CRC (Dufresne et al., 2018; Mullany & Slattery, 2019; Sapp et al., 2017; Slattery, Herrick, et al., 2017). Therefore, the exploration of the extent to which lifestyle factors are associated with miR expression could provide valuable insights into the mechanisms through which exercise exerts beneficial effects on health and may prevent cancer.

This thesis took an innovative and exploratory approach to studying cancer-free LS carriers, their miR profiles, and their lifestyle habits by combining an extensive literature review with three original peer-reviewed articles that cover various features of LS cancers, circulating miRs, and lifestyle habits.

2 LITERATURE REVIEW

2.1 Cancer overview

Cancer is the general term for a large group of diseases whose causes, characteristics, and occurrences can vary greatly. It is one of the primary causes of death in both developed and developing countries around the world (Sung et al., 2019). The most prevalent cancers include breast cancer, lung cancer, and CRC, and when combined, these malignancies account for more than 30% of global cancer mortality (Sung et al., 2019). Of these, CRC especially shows increasing annual incidence rates in Western populations, thus highlighting the need for improvement in all areas related to cancer medicine, such as screening, detection, treatment, and prevention. The Finnish Cancer Foundation reported 31,543 newly diagnosed cancer cases and 13,355 cancer-related deaths in Finland in 2021 (Seppä et al., 2023). Among these numbers, the most common cancer among females was breast cancer, followed by CRC and lung cancer. In males, prostate cancer, CRC, and lung cancer were the leading cancer types. Breast and prostate cancers exhibited five-year survival rates of over 90%, whereas CRC had a five-year survival rate exceeding 70% for both sexes. Among females, breast and lung cancer had the highest mortality rates, while lung and prostate cancer were the most fatal in males. The number of all cancer incidences has risen annually by approximately 0.8% in both sexes since 1990. In contrast, mortality rates have decreased annually by approximately 0.5% in females and by approximately 1.1% in males in Finland (Seppä et al., 2023).

In their groundbreaking publication in 2000, Douglas Hanahan and Robert Weinberg proposed six essential hallmarks of cancer that serve as a comprehensive framework for understanding the complex nature of cancer development and progression (Hanahan & Weinberg, 2000). Since then, their framework has been extended to cover several other hallmark properties that describe the fundamental characteristics that distinguish cancer cells from

normal cells and highlight the various cellular and molecular mechanisms that contribute to tumor formation and growth (Hanahan, 2022; Hanahan & Weinberg, 2011). These hallmark traits enable cancer cells to sustain proliferative signaling (which drives uncontrolled proliferation), evade growth suppressors that would otherwise inhibit such proliferation, and resist cell death that promotes their survival (Glaviano et al., 2023; Harris & Levine, 2005). In addition, cancer cells maintain replicative immortality, which is a pivotal factor that provides them with an evolutionary growth advantage through increased fitness. The rapid proliferation of cancer cells results in genomic instability, which is characterized by numerous mutations in their genome. Furthermore, cancer cells sustain tumor-promoting inflammation, which attracts bioactive molecules vital for growth signaling and evasion of cell death by apoptosis or the immune system (Hanahan & Weinberg, 2011; Taniguchi & Karin, 2018; H.Q. Wang et al., 2022). In addition to these enabling hallmarks, cancer cells induce angiogenesis to secure a steady supply of nutrients and oxygen and to support rapid growth. They also undergo metabolic reprogramming through the deregulation of cellular energetics to further augment tumor formation. By avoiding immune destruction, cancer cells persist and grow despite the body's defense mechanisms. After tumor formation, cancer cells may acquire the ability to invade surrounding tissues and metastasize to distant organs, facilitating the dissemination of the disease to secondary sites in the body (Hanahan & Weinberg, 2011).

Most sporadic cancers may take between several years and decades to develop depending on the cancer type (Hanahan & Weinberg, 2000), but all cancers are thought to share a common general pathogenesis despite the origin of emergence (Stratton et al., 2009). Cancer development (carcinogenesis) characterizes the genetic and epigenetic changes within normal cells that drive uncontrolled cell proliferation. This transformation results in the emergence of malignant cancer cells, which have the capacity to infiltrate tissue boundaries and metastasize to distant tissues and organs (Hanahan & Weinberg, 2000). Because this process is intricately associated with the regenerative capability of multicellular organisms, where cellular growth manifests as the ability to multiply through division—an indicator of biological fitness—carcinogenesis follows the principles of Darwinian evolution through natural selection (Campisi, 2013; Stratton et al., 2009). Carcinogenesis involves two basic mechanisms: the ongoing accumulation of heritable genetic changes in individual cells caused by random mutations, followed by natural selection operating on the resulting diversity of traits (Stratton et al., 2009). The selection process either eliminates cells carrying deleterious mutations or fosters those bearing alterations that confer enhanced proliferative and survival capacities, enabling them to outcompete their neighboring cells (Stratton et al., 2009). Cells that escape from regular cell growth control and acquire evolutionary advantage are removed by genes that repair DNA damage and mutations during replication, or by apoptosis (Hanahan & Weinberg, 2000). However, occasionally these cellular defense

mechanisms fail, which then drives the malignant transformation of normal cells, resulting in cancer.

Mutations in the genome that drive carcinogenesis accumulate over the life course, and most of them are acquired at the pre-cancerous stage (Stratton et al., 2009; Vogelstein et al., 2013). Throughout life, DNA is under a consecutive mutagenic burden that is caused either by external factors, such as lifestyle, dietary habits, and ultraviolet radiation, or by intrinsic factors, such as radical oxygen species and deficient DNA mismatch repair (dMMR) machinery (Stratton et al., 2009). These acquired somatic mutations are categorized as driver and passenger mutations based on their roles in carcinogenesis (Gerstung et al., 2020; Vogelstein et al., 2013; Vogelstein & Kinzler, 2004). Driver mutations occur in proto-oncogenes, which are normal genes that regulate cell growth and differentiation, and thus provide an evolutionary growth advantage for the cells. These cells are positively selected during cancer evolution (Stratton et al., 2009). Passenger mutations do not promote cancer directly. They occur randomly and have neutral effects on gene function. They likely exist in a cancer cell's ancestor by chance when the cell acquires a driver mutation, without actively contributing to carcinogenesis (Vogelstein et al., 2013).

The Catalogue of Somatic Mutations in Cancer (Tate et al., 2019) is a curated database that categorizes genes involved in carcinogenesis. Typically, these genes fall into two main categories: oncogenes and tumor suppressor genes (TSG). Oncogenes, originally proto-oncogenes, undergo gain-of-function mutations, which lead to overexpression and uncontrolled cell proliferation, thus promoting cancer development (Vogelstein et al., 2013). In contrast, TSGs regulate and limit cell growth, division, DNA repair, and apoptosis. Loss-of-function mutations inactivate TSGs, which further promote cancer (Vogelstein et al., 2013). Mutations in oncogenes are typically dominant, whereas in TSGs, they are commonly recessive and require a second hit for inactivation, following the model introduced by Knudson in 1971 (Knudson, 1971; Stratton et al., 2009). In addition, mutations in the MMR genes cause the accumulation of single nucleotide changes and small insertions and deletions in the genome that result in microsatellite instability (MSI). MSI promotes cancer by altering TSG function (Jeggo et al., 2016). A list of common oncogenes and TSGs, as well as MMR genes relevant to this thesis, is provided in Table 1.

TABLE 1 Genes commonly associated with cancers.

Gene*	Type	Pathway*
CREB	Oncogene	NF-kB, cell proliferation
NRAS	Oncogene	RAS/MAPK, cell proliferation
EGFR	Oncogene	EGFR-signaling, cell proliferation
BRAF	Oncogene	MAPK/ERK, cell proliferation
STAT3	Oncogene	MAPK, cell proliferation
FOXO	Tumor suppressor gene	PIK3/AKT/, cell fate
PIK3	Tumor suppressor gene	PIK3/AKT/mTOR, cell growth
APC	Tumor suppressor gene	WNT-signaling, cell proliferation
TP53	Tumor suppressor gene	p53 pathway, cell cycle
SMAD	Tumor suppressor gene	TGF-B-signaling, cell growth
TGF	Tumor suppressor gene	TGF-B-signaling, cell growth
CDKN1A	Tumor suppressor gene	p53 pathway, cell cycle
CDKN2A	Tumor suppressor gene	p53 pathway, cell cycle
MLH1	DNA mismatch repair	DNA damage control
MSH2	DNA mismatch repair	DNA damage control
MSH6	DNA mismatch repair	DNA damage control
PMS2	DNA mismatch repair	DNA damage control

*Gene and pathway names are detailed in the abbreviations section.

2.2 Lynch syndrome

American pathologist Aldred S. Warthin first recognized the hereditary predisposition to gastrointestinal cancers in his studies of "Family G" in the late 19th century (Lynch et al., 2015). Based on the groundwork of Warthin, later research conducted by Henry T. Lynch, also an American physician, on "Family N" in the mid-1900s established that this condition follows an autosomal dominant inheritance pattern and includes cancers of the endometrium. Lynch and his colleagues then denoted the condition as "Cancer Family Syndrome" in 1971 to describe the familial cluster of the cancers they observed. However, again in 1984, Lynch et al. transformed "Cancer Family Syndrome" to "Hereditary Non-Polyposis Colorectal Cancer" (HNPCC) to distinguish it from other familial CRC syndromes. As the understanding of HNPCC broadened to encompass extraintestinal cancers, and after the discovery of the link between HNPCC and the four MMR genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*) in the 1990s, HNPCC was subsequently renamed "Lynch Syndrome" in 2003 after Henry T. Lynch (Aaltonen et al., 1993; Lynch et al., 2015; Peltomaki et al., 1993).

2.2.1 Pathogenesis, cancer risk, and identification of Lynch syndrome

LS is caused by inherited pathogenic germline variants in one of the four MMR genes, *MLH1*, *MSH2*, *MSH6*, and *PMS2* (Lynch et al., 2015). MMR genes are largely responsible for repairing DNA mismatch errors, such as erroneous single-base substitutions, insertions, deletions, and certain forms of DNA damage that

result from environmental factors, cellular processes, and/or processes that occur naturally during DNA replication (Peltomäki et al., 2023). Of the MMR proteins, *MSH2* (or *MSH3*) and *MSH6* form a complex (hMutSa) that is responsible for the single-base mismatch and/or insertion—deletion recognition (Peltomäki et al., 2023). *MLH1* and *PMS2* are needed to form the MMR protein complex (hMutL), which coordinates the interplay between the mismatch recognition complex and other proteins necessary for MMR (Peltomäki et al., 2023).

Along with normally functioning wild-type MMR allele, LS carriers carry an inherited deficient MMR allele that is present in every cell. LS cancers arise from the somatic loss of function in the remaining wild-type allele of the affected MMR gene, in accordance with Knudson's "two-hit hypothesis" (Knudson, 1971). When a defective allele is inherited in the germline, a single somatic mutation leads to inactivation. Consequently, carriers tend to develop the disease at a younger age. The loss of wild-type allele then results in the dMMR genotype that is incapable of producing functional MMR proteins, and thus unable to recognize or repair mismatches during DNA replication and/or recombination (Lynch et al., 2015; Peltomäki et al., 2023; Seppälä et al., 2023). As a consequence, mismatch mutations accumulate and are passed on to subsequent cell populations. Thus, the accumulation of these mismatches along the cell genome predisposes LS carriers to an excessive mutational burden that may result in vast amounts of genetic length changes within microsatellites (Aaltonen et al., 1993), which are short repetitive sequences of DNA found in the non-coding regions of DNA. Eventually, when MSI mutations hit the protein-coding regions residing in proto-oncogenes or TSGs, they may possess severe deleterious effects that ultimately initiate carcinogenesis and often result in hypermutated tumor genotypes (Ahadova et al., 2018).

LS is identified as the most common hereditary cancer predisposition syndrome, with a global prevalence estimate ranging from 1:125 to 1:279 (Haraldsdottir et al., 2017; Win et al., 2017). Due to the dMMR genotype, LS carriers are predisposed to an increased lifetime risk of multiple gastrointestinal and extraintestinal cancers (Dominguez-Valentin et al., 2020; Møller et al., 2017a). The most common clinical manifestations of LS cancer risk are CRC and endometrial cancer, which are detected in all pathogenic variant carriers (Dominguez-Valentin et al., 2020). In addition, the LS cancer spectrum includes cancers of the ovaries, duodenum and small bowel, biliary tract, pancreas, gastric, upper urothelial, bladder, prostate, skin, and brain (Dominguez-Valentin et al., 2020). Pathogenic *MLH1* and *MSH2* are high-penetrance genes for CRC occurrence, as well as for endometrial and ovarian cancer, whereas pathogenic *MSH6* is of moderate penetrance and pathogenic *PMS2* is of low penetrance (Dominguez-Valentin et al., 2020). Pathogenic *MLH1* and *MSH2* carriers are associated with early-onset CRCs with a median age of 45–50 years (Dominguez-Valentin et al., 2020, 2023). Commonly, CRC is not detected in pathogenic *MSH6* carriers before the age of 35, while in pathogenic *PMS2* carriers, the risk of CRC before the age of 50 is very low (Dominguez-Valentin et al., 2020). LS CRCs are

also more commonly located in the proximal colon (Ahadova et al., 2018; Fitzgibbons et al., 1987). Pathogenic *MSH6* variants predominantly elevate endometrial cancer risk (Dominguez-Valentin et al., 2020). Prostate and urinary tract cancers predominantly occur in pathogenic *MSH2* carriers, whereas upper gastrointestinal tract cancers predominantly occur in pathogenic *MLH1* carriers (Møller et al., 2018).

However, the risk of LS cancers varies substantially not only based on gene variant but also the cancer history, age, and sex of an individual. Therefore, LS is nowadays warranted not to be treated as one general entity but as four distinct conditions (Møller et al., 2023). The lifetime risk of LS cancer types, according to the Prospective Lynch Syndrome Database (PLSD, http://www.plsd.eu/, visited 10/2023 (Møller, 2020)), from age 25 to age 75 is presented in Table 2. PLSD was established as an international collaborative effort to pool registry data from specialized LS centers from 25 countries around the world. Currently, PLSD consists of prospective and clinical data from 8,500 LS carriers with 71,713 follow-up years (Møller et al., 2023), thus making it the largest database of identified LS carriers. According to PLSD, pathogenic *MLH1* and *MSH2* showed the highest overall cancer risk when compared to pathogenic *MSH6* and *PMS2* variant carriers.

TABLE 2 Cumulative cancer risk with Lynch syndrome up to age 75.

Organ	***path_MLH1***		***path_MSH2***		***path_MSH6***		***path_PMS2***
	Male	Female	Male	Female	Male	Female	Both
Any cancer	71.4%[a] [13.9-81.3][b]	81.0% [11.1-88.4]	75.2% [7.7-85.7]	84.3% [9.7-91.0]	41.7% [1.7-67.1]	61.8% [0.5-78.7]	34.1% [0-59.6]
Colorectal cancer	57.1% [22.8-67.9]	48.3% [8.2-57.4]	51.4% [5.8-65.0]	46.6% [3.8-55.4]	18.2% [1.7-43.2]	20.3% [0.4-40.5]	10.4% [0-40.8]

[a] The average risk % at age 75.
[b] The risk % from ages 40 to 75.
Path = pathogenic.

International clinical guidelines have been developed to aid in the identification of individuals and families with LS (Table 3). The first standardized clinical criteria for this purpose were the Amsterdam I criteria, which were established by the International Collaborative Group of HNPCC in Amsterdam in 1990 (Vasen et al., 1991). These criteria focused on a strong family history of CRC at a young age of onset but were later updated to Amsterdam II criteria in 1999 (Vasen et al., 1999) and revised Bethesda guidelines in 2004 (Umar et al., 2004) to also account for extra-colonic cancers and molecular testing, respectively. Currently, the definitive diagnosis of LS requires molecular genetic testing of a pathogenic or likely pathogenic heterozygous variant in any of the four MMR genes (Peltomäki et al., 2023; Seppälä et al., 2023). Universal tumor screening is conducted with immunohistochemical staining by antigens of the four MMR

proteins to identify the lack of expression of these proteins. PCR-based or sequencing-based tests are used to detect MSI in LS tumors. In Finland, immunohistochemical staining is performed on all CRC and endometrial cancer cases to identify genes for subsequent mutational analysis (Kansikas et al., 2011; Peltomäki et al., 2023).

TABLE 3 Clinical guidelines for Lynch syndrome identification.

Amsterdam I criteria (Vasen et al., 1991)
1. At least three relatives with histologically verified CRC **2.** One is a first-degree relative of the other two **3.** At least two successive generations are affected **4.** At least one of the relatives with CRC is diagnosed at <50 years of age **5.** Familial adenomatous polyposis has been excluded
Amsterdam II criteria (Vasen et al., 1999)
1. At least three relatives with an LS-associated cancer (CRC and cancers of the endometrium, stomach, ovary, ureter or renal pelvis, brain, small bowel, hepatobiliary tract, and skin) **2.** One is a first-degree relative of the other two **3.** At least two successive generations are affected **4.** At least one of the LS-associated cancers should be diagnosed at <50 years of age **5.** Familial adenomatous polyposis should be excluded in any CRC cases **6.** Tumors should be verified by pathology whenever possible
Revised Bethesda guidelines (Umar et al., 2004)
1. CRC diagnosed in a patient who is <50 years of age **2.** Presence of synchronous or metachronous CRCs or other LS-associated tumor, regardless of age **3.** CRC with high MSI histology diagnosed <60 years of age **4.** CRC diagnosed in one or more first-degree relatives with LS-associated tumor, one of the cancers diagnosed <50 years of age **5.** CRC diagnosed in two or more first- or second-degree relatives with LS-associated tumors, regardless of age

CRC = colorectal cancer; LS = Lynch syndrome; MSI = microsatellite instability.

2.2.2 Colorectal cancer and colonoscopy screening in Lynch syndrome

CRC is the hallmark cancer of LS. The suggested pathway models of LS CRC development are presented in Figure 1. According to the first pathway model, LS CRC may develop through pre-formed polyps or adenomas that are MMR proficient and become dMMR by a secondary inactivation of MMR. As LS tumor formation is commonly initiated by dMMR, it has been proposed that dMMR may not increase the adenoma initiation rate, but rather accelerate the progression of these pre-formed MMR-proficient adenomas into carcinoma (Ahadova et al., 2018). This pathway is commonly associated with *MSH6* and *PMS2* carriers who have been shown to display microsatellite stability (MSS) in low-grade adenomas (Ahadova et al., 2018, 2021). In support of this, Engel et al. reported that *MSH6* mutation carriers are associated with low frequencies of *CTNNB1* mutations and high frequencies of *APC* mutations, suggesting that the

onset of MMR deficiency occurs only after adenoma formation in these carriers (Engel et al., 2020). In addition, in *PMS2*-deficient cancers and adenomas, it has been shown that *KRAS* mutations take place earlier during development than dMMR, which may indicate that MMR deficiency does not drive carcinogenesis in *PMS2*-associated CRC (Seppälä et al., 2023; ten Broeke et al., 2015; ten Broeke, van Bavel et al., 2018).

Regarding the second and third pathways, normal-looking colorectal mucosa of LS carriers has been reported to contain dMMR niches, which may give rise to cancers that develop through adenomas or polyps following *APC* inactivation or without preexisting lesions via the activation of *CTNNB1* (Ahadova et al., 2016, 2018, 2021; Engel et al., 2020). Adenomas and polyps are more frequently detected in *MSH2*-deficient CRCs, which almost always include somatic *APC* variants (Engel et al., 2020), and thus follow the second pathway model. Bohaumilitzky et al. reported that the normal mucosa of cancer-free LS carriers contains higher amounts of CD3 and CD8 immune cells compared with LS cancer patients and MMR-proficient cancer patients (Bohaumilitzky et al., 2022). In the same article, they also reported a correlation between the time to CRC development and the relative abundance of immune cells (Bohaumilitzky et al., 2022). In contrast, *MLH1*-deficient CRCs are commonly *CTNNB1*-mutated but not *APC*-mutated, and thus display far fewer adenomas, as suggested by the third pathway model (Ahadova et al., 2018; Engel et al., 2020). Interestingly, these *CTNNB1*-mutated CRCs may develop through flat precursor lesions, making them difficult to detect in regular colonoscopies, even with short screening intervals (Ahadova et al., 2021). In general, dMMR-driven CRC can develop at an accelerated rate, often taking only one to three years, whereas sporadic CRC often takes 10 to 15 years to develop. This rapid CRC development is also commonly associated with the third pathway model associated with *MLH1*-carriers (Ahadova et al., 2018), thus suggesting that the *MLH1* variant is the most dangerous of the four pathogenic variants.

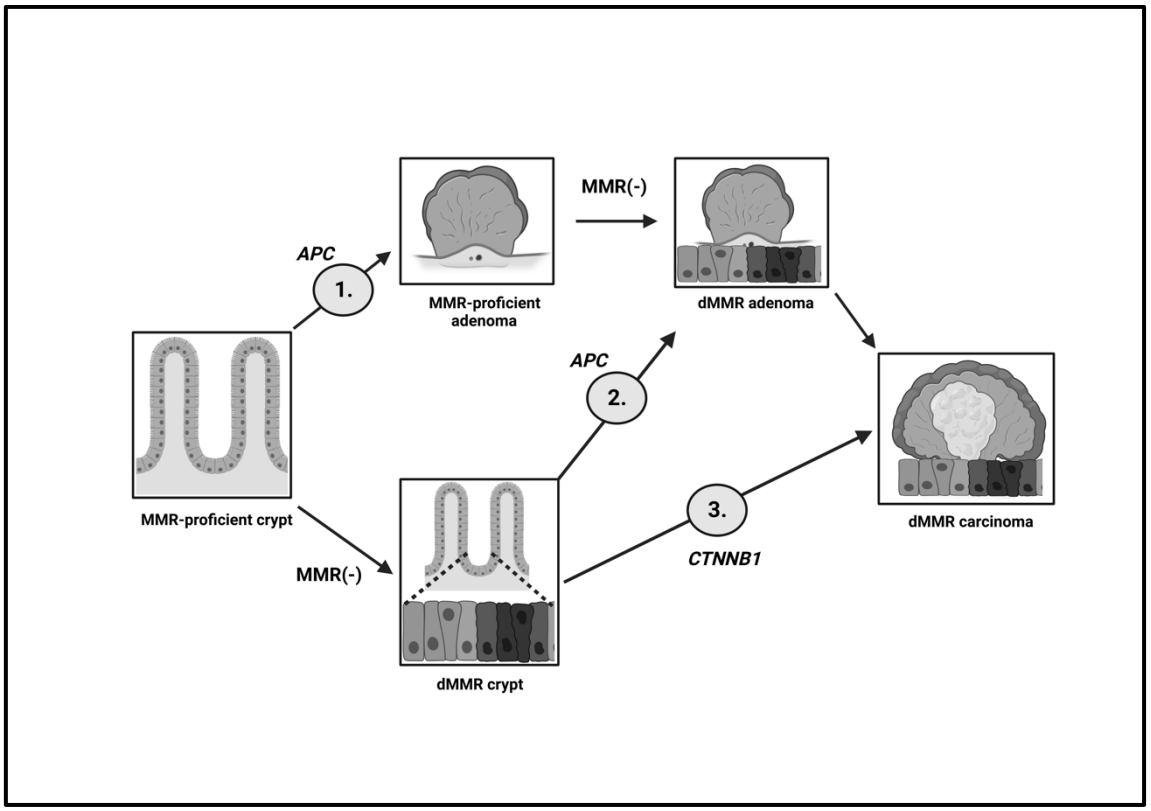


FIGURE 1 The three-pathway model of Lynch syndrome colorectal cancer carcinogenesis adapted from Peltomäki et al. (Peltomäki et al., 2023). APC = Adenomatous polyposis coli; CTNNB1 = Catenin beta 1; dMMR = deficient mismatch repair.

As most CRCs develop from polyps that can be detected and removed during colonoscopy, colonoscopy screening may prevent CRC. Regular endoscopic surveillance by colonoscopy with polypectomy is the standard of care for detecting incident CRCs in cancer-free LS carriers. In LS, surveillance is initiated at age 25 for carriers of *MLH1* and *MSH2,* and at age 35 for carriers of *MSH6* and *PMS2* (Dominguez-Valentin et al., 2020; Seppälä et al., 2023; ten Broeke, van der Klift, et al., 2018). However, recent epidemiological studies have shown that the incidence of CRC remains high even under colonoscopic surveillance (Ahadova et al., 2018), but the overall survival after prospectively observed incident CRCs is very good, reaching 90% after 10 years (Dominguez-Valentin et al., 2019, 2020; Engel et al., 2020; Seppälä et al., 2023).

Optimal colonoscopy screening intervals are under debate. The intervals vary widely, ranging from one year in Germany to one to two years in the Netherlands and three years in Finland (Engel et al., 2018; Järvinen et al., 2000; Vasen et al., 2010). Despite the observed rapid adenoma-to-carcinoma progression in LS-associated CRCs, colonoscopy intervals shorter than three years have not proven effective in decreasing CRC incidence (Dominguez-Valentin et al., 2019, 2020, 2023; Møller et al., 2017b, 2018, 2023; Seppälä et al., 2021). Regarding *MLH1* carriers, carcinomas that develop through pathway

model 3 (Figure 1), where precursor lesions might be flat or skipped totally, rapidly developed carcinomas might not even be detectable by colonoscopy (Ahadova et al., 2018, 2021). In addition, the early age of onset and proximal location of LS CRCs might partly explain the low efficacy of CRC prevention by colonoscopy (Ahadova et al., 2021). This is supported by studies from the general population that have reported a higher risk reduction for distal CRCs (71%–75%) than for proximal CRCs (42%–65%) and that the preventive effect of colonoscopy on proximal CRC was effective only for patients with a later age of onset (older than 60) (Brenner et al., 2014; Doubeni et al., 2018; Samadder et al., 2016). A risk reduction of over 60% on average has been reported for incident sporadic CRCs after a 10-year follow-up, although the numbers show high variation, ranging from 18%–77% (Brenner et al., 2014; Bretthauer et al., 2022; Samadder et al., 2016).

Although CRC mortality rates are generally 20%–30% lower in locations where national screening programs exist (Hull et al., 2020), it appears rather explicit that the implementation of intense colonoscopy screening programs is not uniformly effective among cancer-free LS carriers. Despite the success of current programs, CRC causes a million deaths annually, thus highlighting that this "one-size-fits-all" approach is not effective nowadays (Francavilla et al., 2020; Mauri et al., 2022; Pardini et al., 2023; Sur et al., 2022). Therefore, there is a need to develop methods that could enhance the accuracy of patient selection criteria for risk-based screening programs. By identifying the patients who would mostly benefit from the screenings, the discomfort perceived by the patients, as well as the monetary burden of colonoscopies due to the number of examinations, could be significantly reduced.

2.3 Lifestyle habits and cancer risk

2.3.1 Body weight and physical activity as cancer risk modifiers

Excessive body weight, which may result in developing obesity (body mass index (BMI) > 30 kg/m^2), originates from an imbalance between energy intake and energy expenditure (Friedenreich et al., 2021). Obesity stands as a significant global public health concern, with its prevalence rising steeply over the past four decades (Di Cesare et al., 2016). Among females, it has more than doubled, while among males, it has tripled (Stevens et al., 2012). The combined number of overweight and obese individuals surged from around 800 million in 1980 to two billion in 2013 (M. Ng et al., 2014), showing no decreasing trends (Clinton et al., 2020). Strong evidence indicates that lifestyle habits, such as reduced body adiposity and increased physical activity, are associated with decreased cancer risk (de Rezende et al., 2018; Kyrgiou et al., 2017; Moore et al., 2016). Excess body weight has been suggested to increase cancer risk by several biological mechanisms, including obesity-induced effects on steroid hormone signaling and metabolic activity that promotes inflammation and insulin resistance (Bull et al., 2020). On the contrary, physical activity has been reported to mitigate cancer risk

and progression through multiple biological mechanisms, such as reductions in chronic inflammation, regulation of metabolic factors, changes in insulin resistance, enhanced immune system function, and altering the levels of adipokines (Friedenreich et al., 2010, 2021; Mctiernan et al., 2019).

By pooling data from 204 meta-analyses, Kyrgiou et al. reported that an increase in BMI was associated with a higher risk of developing colon and rectal cancer in males, endometrial cancer in premenopausal female, and esophageal adenocarcinoma, biliary tract system, and pancreatic cancer in both sexes (Kyrgiou et al., 2017). Weight gain and waist-to-hip circumference ratio were associated with higher risks of postmenopausal breast cancer in females who had never used hormone replacement therapy and endometrial cancer, respectively (Kyrgiou et al., 2017). The same study also reported that every 5 kg/m^2 increase during adulthood increased CRC risk by 9% in males, whereas the same amount of increment in weight was seen to be associated with 11% increase in post-menopausal breast cancer risk and 56% increase in biliary track system cancer risk in females (Kyrgiou et al., 2017). Similar results were seen in a systematic review by Renehan et al. that reported men to have a higher risk for CRCs than females, and postmenopausal females to have an increased risk of breast cancer per 5 kg/m^2 increase in BMI (Renehan et al., 2015). The insulin signaling pathway could potentially explain the link between obesity and CRC in men (Friedenreich et al., 2021; Renehan et al., 2008). Higher levels of circulating insulin, triggered by excess adiposity, are more pronounced in men due to their greater tendency toward abdominal fat accumulation compared to females (Geer & Shen, 2009). Additionally, both endogenous and exogenous estrogens have been linked to protective effects against CRC in females, possibly contributing to the stronger association between adiposity and CRC in men (Murphy et al., 2015).

Most studies assessing the associations between body weight and cancer risk in LS have focused on CRC and endometrial cancer. These studies suggest that like sporadic CRC, LS CRC risk is also modified by body weight (Coletta et al., 2019), although the quality of evidence and the parametrization of outcome variables vary. Campbell et al. reported that obesity is associated with an increased risk of CRC of over 90% in males who met the revised Bethesda or Amsterdam criteria for LS (Campbell, Cotterchio, et al., 2007). They also suggested that insulin resistance, which has been reported to have a mechanistic link to dMMR (Duval & Hamelin, 2002), could underlie these observations (Campbell, Cotterchio, et al., 2007). Similarly, Botma et al. showed that every 5 kg increase in body weight was associated with an eightfold increased risk of developing incident adenomatous polyps in overweight LS males without cancer history but not in males who had had cancers previously (Botma et al., 2010). Furthermore, Win et al. reported a 30% increment in CRC risk for every 5 kg/m^2 in early adulthood but found no difference between sexes (Win et al., 2011). In MMR-variant stratified analysis, this risk was increased by 36% in *MLH1/PMS2* carriers and 26% in *MSH2/MSH6* carriers (Win et al., 2011). The authors suggested that the observed inter-variant variation in CRC risk could be explained by differences in MMR complex structures. The same group found no

associations between early adulthood BMI and endometrial cancer risk (Win et al., 2011). Movahedi et al. reported a twice as high CRC and overall cancer risk for obese LS carriers, especially in pathogenic *MLH1* carriers, although the risk was abrogated in those taking aspirin (Movahedi et al., 2015). Furthermore, Lazzeroni et al. found a twofold increase in CRC risk for every 5 kg/m^2 increase in obese males compared to non-obese males, as well as a 49% CRC risk increase in *MLH1* carriers, but no associations were found in females (Lazzeroni et al., 2021). Similar findings were made in a recent meta-analysis that reported an association with obesity and an over twofold increase in CRC risk, but no association with endometrial cancer was seen (Power et al., 2024), as previously suggested (Coletta et al., 2019).

Physical activity, defined as any bodily movement generated by skeletal muscles that consumes energy, is commonly measured as the summation of the type of activity, the volume of activity, and the timeframe in which the activity occurred (Friedenreich et al., 2021). Moore et al. pooled questionnaire data from 10 prospective studies and reported that a higher level of leisure-time physical activity (over six metabolic equivalent tasks (MET) hours per week) compared to a lower-level activity (over three MET hours per week) (90th vs. 10th percentiles) could decrease the risk of CRC by 13%–16%, esophageal cancer by 42%, endometrial cancer by 21%, and breast cancer by 10% (Moore et al., 2016). Similarly, an umbrella review of systematic reviews and meta-analyses composed of 22 anatomical sites and over 700,000 cancer cases concluded that physical activity was associated with a lower risk of several cancers, including CRC, breast, endometrial, esophageal, and pancreatic cancers, but supported strong evidence only for CRC and breast cancer (de Rezende et al., 2018). Similar findings were also reported by the US Physical Activity Guidelines Advisory Committee (DiPietro et al., 2019; Mctiernan et al., 2019).

At the onset of this thesis, it was found that few studies assessed the role of physical activity in LS CRC prevention. Kamiza et al. were the first to assess whether regular physical activity is associated with LS CRC risk (Kamiza et al., 2015). Their analysis showed that conducting physical activity (jogging more than 16 km/week, swimming more than 3.2 km/week, or participating in other activities for more than 5 h/week) reduced the risk of CRC by 38% in *MLH1* and *MSH2* carriers (Kamiza et al., 2015). Dashti et al. reported that vigorous leisure-time physical activity (> 35 MET-h/week) compared to low levels of physical activity (< 3.5 MET-h/week) decreased LS CRC risk by 23% when assessed with self-reported questionnaires (Dashti et al., 2018). However, they did not stratify for sex or MMR variant, although 56% of the study participants were females, and most of the participants were *MSH2* carriers (49.1%) (Dashti et al., 2018). The authors of both studies suggested that the same biological mechanisms could be postulated to play a role in modifying LS cancer risk, as suggested in studies with the general population (Dashti et al., 2018; Kamiza et al., 2015). A recent study by Deng et al. suggested that exercise training may be beneficial in CRC prevention through a reduction in colonic inflammation (Deng et al., 2023). Hence, lifestyle

recommendations concerning weight management and physical activity could also be relevant for cancer prevention among LS carriers.

2.3.2 Other lifestyle factors as cancer risk modifiers

There is convincing evidence derived from the general population that dietary factors, alcohol consumption, and tobacco use modify the risk of several cancers (Clinton et al., 2020; Key et al., 2020; Veettil et al., 2021; Zhao et al., 2023). For example, an umbrella review of 45 meta-analyses reported that alcohol intake (4 drinks/day) and high intake of red meat were associated with higher CRC risk, whereas higher intake of dietary fiber, dietary calcium, and yogurt were associated with a lower risk of CRC (Veettil et al., 2021). Similar findings have been reported by the World Cancer Research Fund and the American Institute for Cancer Research, which recommend eating a diet rich in whole grains, vegetables, fruit, and beans, and limited consumption of alcohol, processed red meats, and sugar-sweetened drinks (Clinton et al., 2020).

Of the dietary factors, resistant starch in particular has been studied in LS (Burn et al., 2008, 2011; Mathers et al., 2012, 2022; Movahedi et al., 2015). Burn et al. reported that a daily dose of 30 g of resistant starch had no effect on LS CRC or advanced adenomas (Burn et al., 2008). The latest study by Mathers et al. reported a daily dose of 30 g of resistant starch to reduce the risk of non-colorectal cancers by almost 50%, but they did not observe any effect on CRC (Mathers et al., 2022). Calcium and multivitamin supplements have been shown to reduce LS CRC risk (Coletta et al., 2019), but smoking and alcohol intake were not associated with increased CRC risk in a recent meta-analysis (Power et al., 2024). In contrast, Dashti et al. reported that alcohol consumption was associated with increased CRC risk (Dashti et al., 2017).

Additionally, in a randomized placebo-controlled study, acetylsalicylic acid (aspirin, an anti-inflammatory drug) of 600 mg per day for two to four years was shown to reduce the incidence of CRC in LS carriers by half, reflecting the trends observed in general population-based studies (Burn et al., 2020). The preventive effect of aspirin on CRC was observed four years after the therapy, and the incidence reduction was maintained for 10 to 20 years in follow-up (Yurgelun & Chan, 2020). The significant variation in the cancer risk between sexes and different pathogenic MMR-variant carriers highlights the potential role of lifestyle habits as LS cancer risk modifiers (Win et al., 2021). The identification of potentially protective or harmful and avoidable risk factors creates opportunities for LS carriers to reduce their life-long risks of multiple malignancies (Win et al., 2012). In general, the modifying effect of risk factors could be useful for cancer risk prediction and individual treatment plans.

2.4 MicroRNAs

2.4.1 MicroRNA biogenesis and gene regulation

Unlike genetic changes that result in permanent changes in the DNA sequence, epigenetic changes affect gene expression (Jung et al., 2020). Epigenetic changes, such as DNA methylation, histone modification, and gene regulation of non-coding RNAs, possess central pathophysiological roles in the initiation and progression of several cancers (Esteller, 2011; Hanahan, 2022; Jung et al., 2020). Of the non-coding RNAs, miRs are the most studied in cancers. MiRs are small (typically ~22 nt in size) regulatory non-coding RNA molecules that exhibit a high degree of conservation across evolutionary scales, and their diversity and abundance correlate with organismal complexity (Mori et al., 2019). Illustratively, the human genome encompasses approximately 2,600 distinct miR species, whereas, for example, the mouse (*Mus musculus*) repertoire comprises around 1,500 different miRs (miRbase, v. 22) (Griffiths-Jones et al., 2006). The genomic encoding of miRs occurs within inter- and intra-genic regions and is often characterized by clustering and co-transcription, thereby augmenting their regulatory impact (Bracken et al., 2016; Roush & Slack, 2008). Generally, miR regulation occurs at the post-transcriptional level, overseeing the translation of over 60% of protein-coding genes (Jung et al., 2020), either by messenger-RNA (mRNA) cleavage, mRNA destabilization, or inhibition of translation (He & Hannon, 2004). While many miRNAs exhibit ubiquitous expression, others display high tissue specificity (Lagos-Quintana et al., 2002; Ludwig et al., 2016). Similar to mRNAs, miR expression profiles can serve as distinctive signatures indicating cell identity or state (Mori et al., 2019).

The synthesis of miRs is a complex process that is regulated at multiple levels (Ha & Kim, 2014). This biogenesis is initiated in the nucleus and further processed in the cytoplasm (Ha & Kim, 2014). Briefly, within the nucleus, miR genes are transcribed into ~1 kb stem-loop pri-miR transcripts by RNA polymerase II and III in association with several transcription factors, and subsequently cleaved by enzyme Drosha (RNase III) into a pre-miR hairpin structure of ~65 bp (Cai et al., 2004). The pre-miR hairpins are further processed by the microprocessor complex composed of RNase III endonuclease Drosha, as well as its cofactor DGCR8 (Lee et al., 2003). After being processed by Drosha and DGCR8, the pre-miRs are transported from the nucleus into the cytosol by the Exportin-5-RanGTP-binding complex (Lund et al., 2004). Within the cytosol, pre-miRs are further processed by type III endoribonuclease Dicer in association with RNA-binding proteins TRBP and PACT, producing double-stranded miR duplexes consisting of guide and passenger strands (Bernstein et al., 2001). These duplexes are loaded into the RNA-induced pre-silencing complex (pre-RISC) (Winter & Diederichs, 2011). The guide miR strand (~22 nt) and argonaute proteins (1–4) form the mature RISC after separation from the less abundant passenger miR strand, which is commonly discarded (Ha & Kim, 2014; Huntzinger & Izaurralde, 2011). The guide miR in the RISC includes a "seed

region" at its 5′ tail (residues 2–7), which leads the RISC to occupy the right position on the target mRNA (Doench & Sharp, 2004).

Typically, miRs negatively regulate gene expression by accelerating the deadenylation and degradation of target mRNAs (Mori et al., 2019; O'Brien et al., 2018). The argonaute proteins bind different classes of small non-coding RNAs and function as effectors that recruit other factors essential for translational repression and mRNA decay (Ha & Kim, 2014). The mature miR can complementarily bind to 3′ untranslated region of the target mRNA transcript via seed sequence, defined as the first 2–8 nucleotides of the 5′ end of miRs, and thus initiate degradation by guiding the miR silencing complex to its target (Ha & Kim, 2014). However, complementary binding can be partly imperfect, which leads to silencing of the mRNA transcript instead of degradation. Imperfect binding also establishes a comprehensive regulatory machinery that enables individual miRs to regulate up to several hundred mRNA transcripts and share multiple mRNA targets (Esquela-Kerscher & Slack, 2006). The binding of only a single type of miR to a target mRNA results in a relatively modest reduction in target expression. However, in instances where several miRs target multiple components of a regulatory system, the net effect may be substantial (Bracken et al., 2016).

2.4.2 Circulating microRNAs

MiRs are produced by every cell type in the body and thus can be found in a stable form in virtually all body fluids, including blood (Chen et al., 2008; Mori et al., 2019; Weber et al., 2010). They play a pivotal role in cell and tissue communication, which is facilitated by their export and import via extracellular vesicle trafficking and protein carriers like argonaute proteins (Valadi et al., 2007). These miRs are referred to as circulating miRs (c-miRs), which can prompt downstream effects upon uptake by target recipient cells by regulating the translation of complementary mRNAs (Figure 2). Several extracellular miR transport routes are identified: active transport via extracellular vesicles and transportation within protein–miR complexes, such as argonaute proteins and lipoproteins. Additionally, miR release can occur from damaged or senescent cells (Mori et al., 2019). The intercellular transport of miRs and subsequent functional regulation of gene expression in recipient cells is a well-supported mechanism of cell-to-cell communication involving a variety of cell types and transport methods (Sapp et al., 2017). However, the precise mechanisms by which the packaging occurs are not known, although a proportion of c-miRs are suggested to be derived from leukocytes and endothelial cells, as well as from organs exposed to high blood flow (Aoi, 2015; Pritchard et al., 2012).

Changes in miR expression in different tissues are potentially reflected in blood circulation (Skog et al., 2008; Waters et al., 2012). Thus, there has been extensive exploration of miRs as promising candidates for the development of less invasive biomarkers. A robust biomarker exhibits several characteristics associated with miRs, such as specificity, sensitivity, and stability, and is coupled with the advantage of being obtainable in a relatively non-invasive manner (Mori

et al., 2019). The collection of biomarkers from non-solid biological tissues is called a liquid biopsy. In contrast to traditional tissue biopsy, liquid biopsy techniques are generally non-invasive or minimally invasive, thus offering a means to assess the health or disease status of organs and sites that are challenging to access directly (Toden & Goel, 2022). Liquid biopsy also facilitates easier and more frequent sampling over the course of the disease, offering an opportunity to use c-miRs for real-time monitoring of cancer treatment responses and disease progression. Nowadays, profiling of global c-miR expression has become prevalent, and miR expression can be correlated with cancer type, stage, and other clinical variables (Francavilla et al., 2020). Therefore, aberrantly expressed miRs have been linked with diagnostic, predictive, and prognostic potential in the molecular profiling and early detection of cancers.

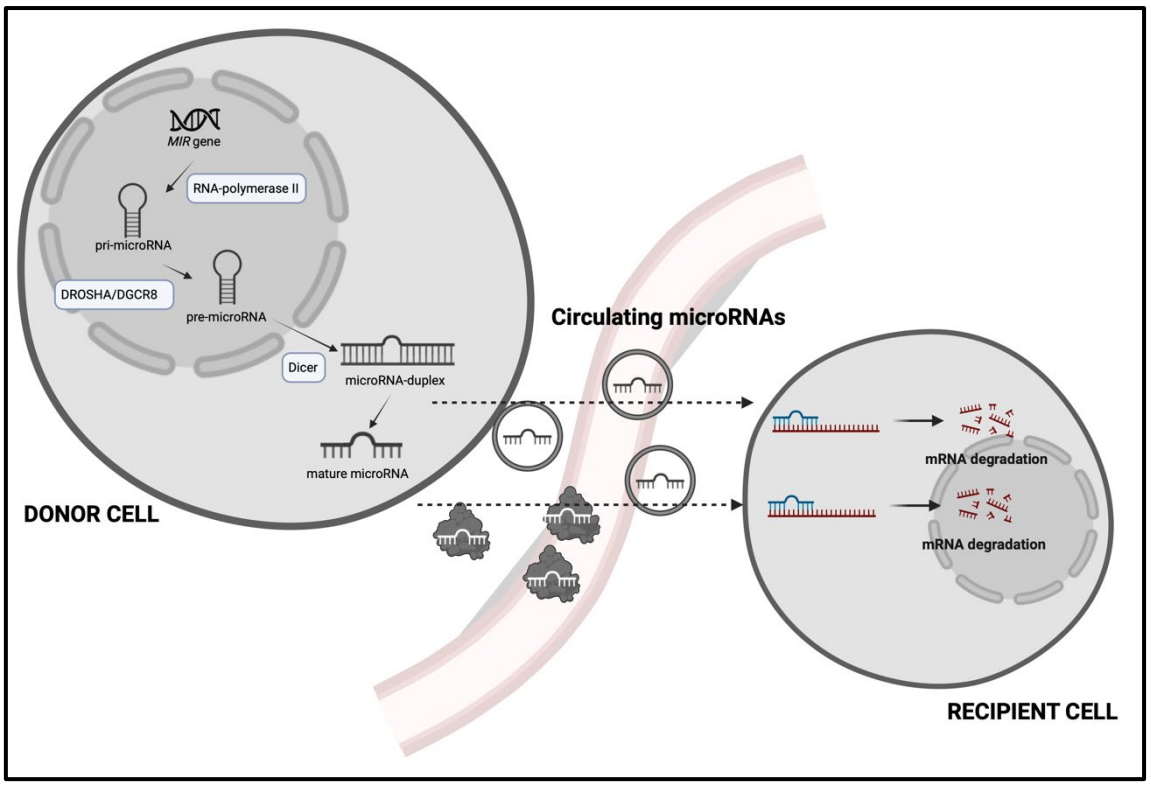


FIGURE 2 Circulating microRNAs.

2.4.3 MicroRNAs in sporadic and hereditary cancer

Numerous studies have revealed distinct miR expression patterns between cancer tissues and adjacent normal tissues (Ma et al., 2012). Consequently, miRs can exhibit upregulation or downregulation in tumor tissues, with a higher prevalence of overexpression observed in cancer (Goodall & Wickramasinghe, 2021; Jung et al., 2020). In general, miRs in cancer can be categorized into oncogenic and tumor-suppressive miRs, although their role may be altered based on cancer type and disease stage. MiRs located in genomic regions amplified in cancers (such as miR-10b, miR-17-92 cluster, miR-20a, and miR-155) function as

oncogenes, whereas miRs located in portions of chromosomes deleted in cancers (such as the miR-15a–miR-16-1 cluster, let-7, miR-34, and miR-200 families) function as tumor suppressors (Calin & Croce, 2006; Otmani & Lewalle, 2021). Thus, miRs that focus on inhibiting the negative regulators of oncogenic pathways may exhibit oncogenic characteristics when their regulation is disrupted and, conversely, may demonstrate tumor-suppressive properties when targeting positive regulators.

The primary factor contributing to alterations in miR functionality within cancer cells is abnormal gene expression. This abnormality is defined by atypical expression levels of miR sequences in contrast to the corresponding normal tissues (Peng & Croce, 2016). As an example, miRs such as miR-34 and miR-200 are upregulated by the TSG *TP53*, which is frequently deleted, mutated, and/or inactivated in many cancer types (He et al., 2007; Kim et al., 2011). Therefore, the suppression of these tumor-suppressive miRs by *TP53* inactivation promotes cancer development. Conversely, miR-17 and miR-20a, among others, are induced by *MYC*, an oncogene that undergoes hyperactivation and/or overexpression in various types of cancer (O'Donnell et al., 2005). As miRs have the capacity to target multiple mRNAs concurrently, the oncogenic impact of each miR probably arises from the suppression of various targets. This intricate regulation encompasses the targeting of numerous genes by individual miRs, the mutual targeting of specific genes by multiple miRs, and the subsequent downstream effects facilitated by the miR-induced regulation of transcription factors (Barabási et al., 2011; Bracken et al., 2016; Ooi et al., 2011). These complex interactions give rise to intricate networks of miRs and their target genes, where nodes (Bracken et al., 2016)—particularly those with unusually high connections (hubs)—emerge as pivotal sites of signaling convergence. Identifying such hubs is valuable and provides explanatory insights into network behavior and potential clinical applications (Barabási et al., 2011).

A plethora of miRs have been linked to various facets of cancer, including diagnosis, prognosis, therapeutic response, and prediction (Dhawan et al., 2018; Mullany & Slattery, 2019). For instance, elevated expression of miR-10b correlates with larger tumor size, increased invasion, metastasis, advanced stage, and poorer survival (Sheedy & Medarova, 2018). On the other hand, reduced expression of the miR-200 family, such as miR-200a and miR-141-3p, is associated with diminished survival, exerting its influence through the regulation of genes related to the epithelial-to-mesenchymal transition (Høye et al., 2022; Pichler et al., 2014). Moreover, the discrimination of patients with early colorectal neoplasms from healthy controls is possible through serum-based miR-21, miR-29a, and miR-125b (Yamada et al., 2015). In most of these prior studies, the analysis of miR signatures has been limited to patients who have already received a cancer diagnosis, making it challenging to ascertain their potential utility in risk stratification. Wikberg et al. observed that major changes in miR patterns occur mainly three years prior to sporadic CRC diagnosis by showing a temporal pattern of increase in miR-21-5p expression using pre- and post-diagnostic plasma samples (Wikberg et al., 2018). Interestingly, a study by Raut

et al. showed that altered expression of let-7g, miR-19a, miR-23a, miR-92a, miR-144, miR-21, and miR-27a could predict sporadic CRC incidence several years prior to diagnosis (Raut et al., 2021). They reported that a risk sum score of these seven c-miRs was highly predictive for sporadic CRC risk in a prospective cohort with a follow-up time of up to 14 years and a median follow-up of 6.8 years.

Although various studies have shown that c-miR expression patterns change with carcinogenesis in various sporadic cancers, the role of miRs in LS has remained understudied. MiRs can be used to distinguish cancer tissues from normal tissues, and they are suitable for risk prediction purposes. In 2011, Balaguer et al. showed that miRs can be used in tumor classification and the discrimination of sporadic and hereditary tumors with MSI (Balaguer et al., 2011), thus highlighting the potential role of miRs as LS biomarkers. Valeri et al., Liccardo et al., and Zhou et al. postulated that miRs could have functional roles in LS carcinogenesis, for example, by targeting MMR genes (Liccardo et al., 2021; Valeri et al., 2010) and various tumor-suppressor genes (Zhou et al., 2016). However, these studies, along with other reports, have assessed miR functions in the colorectum and CRC tissues and cells, as well as with microarray data (Balaguer et al., 2010; Pavicic et al., 2011) but not in circulation, which possesses the highest potential for liquid biopsy-based early diagnostics. Thus, given the potential of miRs in CRC risk stratification as a minimally invasive screening tool, the characterization of the LS miR landscape could help establish a risk prediction signature that complements current screening strategies and early diagnostics of LS CRCs. MiRs that have been associated with LS and are relevant to this thesis are listed in Table 4.

TABLE 4 MicroRNAs associated with Lynch syndrome.

microRNA	**Expression**	**Reference**
hsa-miR-21	Upregulated	Valeri et al. 2010
hsa-miR-23b	Upregulated	Moreno et al. 2019
hsa-miR-24	Upregulated	Moreno et al. 2019
hsa-miR-27b	Upregulated	Moreno et al. 2019
hsa-miR-125b	Upregulated	Balaguer et al. 2011
hsa-miR-137	Up-/ downregulated	Balaguer et al. 2010
hsa-miR-155	Upregulated	Valeri et al. 2010
hsa-miR-192	Upregulated	Balaguer et al. 2011
hsa-miR-320	Downregulated	Moreno et al. 2019
hsa-miR-362	Upregulated	Balaguer et al. 2011
hsa-miR-486	Upregulated	Balaguer et al. 2011
hsa-miR-520e	Downregulated	Zhou et al. 2016
hsa-miR-590	Upregulated	Zhou et al. 2016
hsa-miR-622	Downregulated	Balaguer et al. 2011
hsa-miR-1238	Downregulated	Balaguer et al. 2011

2.4.4 MicroRNAs, cancer, and lifestyle habits

There is a growing interest in studying how exercise benefits health and prevents disease at the molecular level. It is well established that miRs play a critical role in the regulation of the core mechanisms of carcinogenesis and stress responses and are involved in most physiological processes (Mori et al., 2019; Sapp et al., 2017). Thus, understanding how lifestyle habits influence miR expression could provide valuable insights into the mechanisms through which exercise has beneficial effects on health and may prevent cancer.

Studies that have shown an association between body adiposity and/or physical activity and c-miRs have been extensively reviewed by Iacomino & Siani (2017) and Dufresne et al (2018). Many of the reviewed c-miRs, such as the miRs of the let-7 family (Esquela-Kerscher & Slack, 2006; Roush & Slack, 2008), miR-17/19 cluster (Gits et al., 2013), miR-21 (Yamada et al., 2015), miR-125b (Ortega et al., 2013; Yamada et al., 2015), miR-126 (Yamaguchi et al., 2014), miR-206 (Parasramka et al., 2012), and miR-221/miR-222 cluster (Gits et al., 2013), have been shown to be associated with the risk of developing various types of cancer. Ortega et al. found that morbidly obese patients had elevated levels of circulating miR-140, miR-142, and miR-222, while levels of miR-532, miR-125b, miR-130b, miR-221, miR-15a, miR-423, and miR-520c were reduced (Ortega et al., 2013). They also reported a significant decrease in circulating miR-140, miR-122, miR-193a, and miR-16-1, and an increase in miR-221 and miR-199a after weight-loss surgery (Ortega et al., 2013). Alterations in circulating miR-23a, miR-27a, miR-130, miR-195, miR-197, miR-320a, and miR-509-5p have been associated with metabolic syndrome (Deiuliis, 2016; Karolina et al., 2012). Furthermore, miR-10b and miR-200a are associated with elevated plasma total cholesterol levels, disrupted lipid metabolism, and obesity, and their expression levels are altered in cancer (Mens et al., 2020; Ortega et al., 2013; Ruiz-Roso et al., 2020). In general, miRs have been shown not only to play a role in elevated plasma total cholesterol levels, but their dysregulation also contributes to disrupted metabolism, which is a hallmark of obesity and cancer (Abozaid et al., 2022; Chadid et al., 2018; Hanahan & Weinberg, 2011; Heyn et al., 2020; Iacomino & Siani, 2017; Otsuka et al., 2023; Renehan et al., 2008).

Acute and chronic physical activity represents a lifestyle behavior that influences the expression of several c-miRs (Baggish et al., 2011; Bye et al., 2013; Nielsen et al., 2014; Sapp et al., 2017), including some of which have also been associated with cancer (Dufresne et al., 2018). For example, miR-221 and miR-222 may function as either tumor-suppressive or oncogenic c-miRs. In gastrointestinal stromal tumors, they act prophylactically by suppressing the *KIT* receptor, which activates cancer-promoting pathways, such as *STAT3*, *PI3K*, and the *MAPK* cascade (Gits et al., 2013). Thus, the modulation of these miRs through physical activity may reduce the risk of cancer by inhibiting *KIT* activation and downstream pathways. Physical activity has also been shown to affect the expression of miR-133 (Nielsen et al., 2014), which has been recognized as a tumor suppressor through targeting of oncogene *EGFR* in CRC (Dong et al., 2013) and breast cancer (Cui et al., 2013), among others (Dufresne et al., 2018).

Moreover, let-7 is a tumor suppressor miR whose expression is decreased in cancers (Roush & Slack, 2008), and reported to be modulated by exercise (Bye et al., 2013; Nielsen et al., 2014).

Interestingly, studies that have looked at the interaction between physical activity and BMI show the two to interact, with people at greatest risk of cancer being those with a large BMI who do not participate in vigorous physical activity (Shaw et al., 2018). Thus, identifying c-miRs that function at the intersection of physical activity, BMI, and cancer could serve as potential therapeutic targets in the future as well as potential candidates for real-time monitoring of, for example, exercise interventions.

3 AIMS OF THE STUDY

The aims of this thesis were to characterize the serum-based c-miR landscape of cancer-free LS carriers, to inspect whether any of those c-miRs are potential indicators of upcoming CRC, and to determine whether they are associated with modifiable cancer risk factors, such as body weight and physical activity. Furthermore, this thesis applied retrospective lifestyle questionnaire data to investigate whether longitudinal body weight gain and physical activity are associated with LS cancer risk. To address these critical aspects, the following questions were answered:

1. Do cancer-free LS carriers display differential systemic c-miR expression compared to the healthy non-carrier group and sporadic CRC patients? (Study I)
2. Can systemic c-miRs predict LS cancer incidence during a four-year prospective surveillance period? (Study II)
3. Are lifestyle habits, such as body weight gain and physical activity, associated with c-miRs and LS cancer risk? (Study II & III)

4 MATERIALS AND METHODS

4.1 Study designs and participants

This thesis was based on cross-sectional and longitudinal human study designs (Table 5).

The **LS miR study** (Study I, https://doi.org/10.17011/jyx/dataset/93204) characterized cross-sectionally the systemic serum c-miR profiles of LS carriers, sporadic CRC patients, and healthy non-carrier controls (controls) who were assigned to independent discovery and cancer cohorts. The discovery cohort (n = 118) was composed of 81 cancer-free LS carriers and 37 controls whose c-miR profiles were sequenced. The cancer cohort (n = 37) was composed of 13 LS carriers who had cancer and 24 sporadic CRC patients whose c-miR profiles were sequenced.

The **LS biomarker study** (Study II) investigated longitudinally whether pre-diagnostic c-miR profiles can predict cancer incidence during a prospective surveillance period of approximately four years (2018–2022) and whether the predictive signature is associated with BMI and physical activity. The study cohort (n = 138) consisted of 77 cancer-free LS carriers and 37 controls derived from the LS miR study, as well as 24 newly collected cancer-free LS samples.

The **LS lifestyle study** (Study III) retrospectively examined whether longitudinal physical activity and body weight gain of LS carriers during adulthood from the age of 20 until 2016 or 2020 was associated with cancer incidence. Questionnaires for anthropometric, socioeconomic, and life style habit data collection were sent to 1038 adult LS carriers whose addresses were available in the Finnish Lynch Syndrome Research Registry (LSRFi) in December 2016 and July 2020. Of them, 480 (response rate 46.2%) returned the questionnaire. However, 15 participants did not carry the pathogenic MMR variant; therefore, they did not fulfill the eligibility criteria and were excluded from the study. Thus, the final study cohort consisted of 465 LS carriers.

All LS carriers in this study were registered with the LSRFi and provided consent for research-related contacts. Age, sex, MMR-variant status, family cancer history, and all cancer diagnoses with the cancer type and date of each diagnosis were confirmed from hospital medical records and national cancer registries upon recording in the LSRFi. The LSRFi is a nationwide research registry (est. 1982), operating in Jyväskylä and Helsinki, which organizes surveillance and cancer prevention for LS families. The registry consists of clinical and family history data of over 400 LS families and over 1800 LS carriers under frequent surveillance. Individuals were identified in the registry before genetic testing became available, based on Amsterdam and Bethesda clinical criteria (Umar et al., 2004; Vasen et al., 1999), and subsequently through cascade testing of the families and universal testing of tumors. Adult members of the LSRFi with confirmed pathogenic MMR variants (classes 4 and 5, according to InSIGHT criteria (Spier et al., 2023) were eligible for the study.

Sporadic CRC patients were enrolled at the time of their initial appointment for surgery at the surgical clinic of the local tertiary center responsible for the management of rectal cancer in Helsinki University Central Hospital, Unit of Rectal Surgery, Helsinki, Finland.

Control samples were acquired from the Biobank of Eastern Finland, Kuopio, Finland (n = 27), in 2020, or from the Estrogenic Regulation of Muscle Apoptosis (ERMA) study (n = 10) consisting of healthy females ages 47 to 55. Persons with no cancers, blood disorders, acute or chronic infectious diseases, rheumatoid arthritis, or known *BRCA* or MMR gene germline mutations were eligible for the control group.

TABLE 5 Study designs.

	Study I	**Study II**	**Study III**
Design	Cross-sectional	Longitudinal, prospective	Longitudinal, retrospective
N	155	138	465
Lynch syndrome	94	101	465
Cancer-free	81	101	242
Cancer	13	-	223
Sporadic CRC	24	-	-
Control	37	37	-
Data collection	Small RNA seq	Small RNA seq, questionnaire	Questionnaire
Data type	Measured	Measured, self-reported	Self-reported
Main methods	DESeq2(Love et al., 2014)	DESeq2, Lasso-Cox(Tibshirani, 1997)	Cox regression(Therneau & Grambsch, 2000)

CRC = colorectal cancer; Lasso = least absolute shrinkage and selection operator.

4.2 Ethics

All the studies in this thesis were conducted in accordance with the guidelines of the Declaration of Helsinki and approved by the Central Finland Healthcare District Ethics Committee (KSSHP D# 1U/2018 and 1/2019 and KSSHP 3/2016) and the Helsinki and Uusimaa Healthcare District Ethics Committee (HUS/155/2021). Written informed consent and permission to use and publish data for scientific purposes were obtained from all study participants.

4.3 Measurements

4.3.1 Serum sample collection and small RNA extraction

Blood sampling of LS carriers was performed at their regular colonoscopy surveillance appointments at Helsinki University Central Hospital in Helsinki and Central Finland Central Hospital in Jyväskylä, Finland. Blood sampling of sporadic CRC patients was performed at the time of their initial appointment for surgery at the surgical clinic of the local tertiary center responsible for the management of rectal cancer in Helsinki University Central Hospital, Unit of Rectal Surgery, Helsinki, Finland. ERMA samples were collected at the University of Jyväskylä in Jyväskylä, Finland. Venous blood samples of LS carriers, sporadic CRC patients, and ERMA controls were drawn at a fasted state. The duration of fasting was not reported for the samples obtained from the Biobank of Eastern Finland. Samples were taken from the antecubital vein into standard serum tubes (Greiner). To separate serum, the whole blood samples were allowed to clot for 30 minutes at room temperature, centrifuged at 1,800 *g* for 10 minutes, and aliquoted and stored at -80°C whenever necessary.

All c-miR isolations from blood serum were carried out using an affinity column-based miRNeasy Serum/Plasma Advanced Kit (Qiagen) according to the manufacturer's instructions. Briefly, 0.5 mL of thawed serum was used to isolate all c-miRs. Cel-miR-39 miR mimic (Qiagen) was added to each sample to serve as a spike-in control for monitoring miR purification and amplification. Phase separation centrifugation was executed at 12,000 *g* for 3 minutes at room temperature (Heraeus, Biofuge Pico, and Fresco 17, ThermoFisher), and the rest of the centrifugations were performed at 16,000 *g* whenever a range of 8,000–20,000 *g* was recommended. Circulating miRs were eluted to nuclease-free water and stored at -80°C whenever necessary.

RNA quality and recovery of spike-in Cel-miR-39 were checked by RT-qPCR (CFX96 and 384™ Real-Time PCR Detection System, Bio-Rad) prior to library preparation according to the manufacturer's protocol. After small RNA isolation, cDNA was synthetized using a miRCURY®LNA® RT kit (Qiagen). cDNA synthesis was done from 12 µl of non-diluted template RNA, and the PCR protocol was carried out in a standard thermocycler (Eppendorf). Transcript

levels were measured using miRCURY LNA™ miRNA PCR assays (Qiagen) and miRCURY LNA SYBR® Green kit (Qiagen). One µl of 1:2 diluted cDNA was used per well, and the samples were run as triplicates. The RT-qPCR protocol was as follows: 95°C (2 min, activation), 95°C (10 s), 56°C (60 s) with 40 cycles. Fold expression was calculated using the formula 2(-ΔΔCt), where ΔCt(sample)-ΔCt (mean Ct from all samples), ΔCt is Ct (c-miR of interest)-Ct (mean Ct from the group) and Ct is the cycle at which the detection threshold is crossed. Samples with Ct values of over 35 were excluded from the analysis.

4.3.2 Library preparation and small RNA sequencing

Small RNA sequencing libraries were prepared with a QIAseq miRNA Library Preparation Kit (Qiagen) according to the manufacturer's instructions using multiplexing adapters. Briefly, the small RNA fractions were first ligated to sequencing adapters from both 5′ and 3′ ends, reverse transcribed into cDNA using UMI-assigning primers, and purified using magnetic beads. A universal indexing sequence was also added in the reverse transcription step to allow the samples to be distinguished from each other. The libraries were then amplified with a standard thermocycler (Eppendorf), purified, and eluted into nuclease-free water. The libraries were stored at -20°C if used within two weeks and at -80°C otherwise.

The sequencing library concentrations were measured with a Qubit fluorometer (Invitrogen), and quantified, diluted, and pooled into a single mixture in equal amounts (1.8 pM per sample) prior to sequencing. Sequencing of the small RNA libraries was done with NextSeq 500 (Illumina) using the NextSeq 500/550 High Output Kit v. 2.5 with 75 cycles (Illumina) to produce 75-base pair single-end reads with aimed mean sequencing depth of > 5 M reads per sample, as recommended by the manufacturer (Qiagen). The samples used in this thesis were sequenced in four distinct batches. Quality assessment of the small RNA libraries prior to sequencing was completed with TapeStation 4200 (Agilent).

4.3.3 Body anthropometrics

Body anthropometrics were self-reported. The participants were asked to report their height and to measure their body weight before breakfast and without clothes. If weight could not be measured, participants were asked to fill in their weight history and the last known weight measurement. BMI was calculated as weight in kilograms divided by the squared height in meters (kg/m^2). Weight was measured by a clinician during the participant's regular colonoscopy surveillance visit. Whether the participant did not report or recall their weight, and if the weight was not measured by a clinician, the missing weight data was imputed by using multiple imputation (van Buuren & Groothuis-Oudshoorn, 2011). To measure body weight history during adulthood, participants recalled their body weight in kilograms at ages 20, 30, 40, 50, 60, and up to 70+ years.

4.3.4 Physical activity assessment

Physical activity was assessed using a self-reported questionnaire. The questionnaire included seven questions about the frequency, intensity, and duration of leisure-time physical activity and commuting activity. Based on the responses, the MET hours per day for leisure-time physical activity were calculated. Missing data were handled using multiple imputation (van Buuren & Groothuis-Oudshoorn, 2011).

Physical activity during adulthood was assessed via four-option scale questions through which participants recalled the level of regular physical activity they had at different adult age ranges throughout their lives. Participants reported their past physical activity at age ranges of 20–29, 30–39, 40–49, 50–59, 60–75, and 75+ years up to their current age period at the time of measurement. The four response options for each age period were as follows: (1) no regular physical activity, (2) regular independent leisure-time physical activity (all non-organized occupational or leisure-time physical activity, i.e., commuting, school/work activities), (3) regular goal-oriented competitive sport and training related to that sport, and (4) other regular supervised physical activity (physical activity that was organized in a sport club, etc., but was not related to competitive sports participation). These four categories were re-categorized into low activity (options 1 and 2) and high activity (options 3 and 4) used for modeling. The same re-categorization was applied to each age range.

4.4 Statistical analyses

Means and standard deviations were used as descriptive statistics for continuous measurements, frequencies, and percentages of categorical data. Pearson and Spearman methods were used to inspect correlations between parametric and non-parametric continuous variables, respectively. Levene's test was used to inspect homoscedasticity, while the Shapiro–Wilk test was used for testing normality. P-values and/or false discovery rate (FDR) < 0.05 were considered significant in all analyses. All statistical analyses and data visualizations were performed in the R-programming environment (R Core Team, 2022) (v.4.2.2) using RStudio and in-house R-scripts.

4.4.1 Longitudinal and near-term cancer risk

Cox's proportional hazards model with time-dependent covariates was used to examine the association between body weight, physical activity, and cancer incidence in longitudinal and near-term settings. Age served as the time variable, capturing cancer status at follow-up completion. The follow-up period spanned from study entry at age 20 until cancer diagnosis (event) or remaining cancer-free (censored). To account for time-dependent weight and physical activity measurements, a counting process approach was employed to analyze the

relative risk of cancer associated with these exposures. The data were divided into 10-year intervals, with each interval characterized by corresponding weight and physical activity measurements. Separate models were constructed for males and females, as well as for any type of cancer or CRC. Nested random effects were incorporated to address the family structure. Hazard ratio (HR) and 95% confidence interval (CI) were reported from both unadjusted and adjusted models, accounting for relevant factors, such as affected MMR gene variant, height, education, smoking, alcohol use, and non-steroidal anti-inflammatory medication.

4.4.2 Construction and validation of cancer risk prediction model

Least absolute shrinkage and selection operator (Lasso) regularized Cox regression was used to find predictor c-miRs from the pool of identified LS-associated differentially expressed c-miRs using the entire study sample. The optimal value for the Lasso regularization parameter lambda was chosen with 10-fold cross-validation. The expression levels of the Lasso-obtained c-miRs were used to compute an individual risk sum score (linear predictor) for all participants using the following formula:

$$\text{Risk sum score} = \text{Expr}(\text{miR}_A) * \beta(\text{miR}_A) + \text{Expr}(\text{miR}_B) * \beta(\text{miR}_B) \ldots,$$

Expr(miR) represents the normalized and variance stabilized c-miR expression and β(miR) indicates the regression coefficient in the Lasso Cox regression model. By using univariate and multivariate Cox regression models, the c-miR risk sum score was then applied to predict the risk of cancer incidence. The entire study sample (n = 101) was used to fit the risk prediction model. The predictive performance of the risk prediction model was validated with fivefold cross-validation, and the model concordance was evaluated with Harrel's C-index (Harrell et al., 1982) (scale 0.5–1.0), where 0.5 indicates poor performance and 1.0 indicates excellent performance.

The surveillance time used for risk prediction was determined from the time-point of initial serum sampling (2018–2020) until the latest update of the LSRFi (11/2022). The response variable in the risk prediction model was the age at the time of cancer diagnosis (event) or the age at the final update of the LSRFi (censoring). HR and 95% CI of the c-miR risk sum score were estimated for the unadjusted model as well as for the sex-adjusted model. The proportional hazards assumption was tested using Schoenfeld residuals. The "glmnet" R-package (Friedman et al., 2010) was used for the cross-validation procedure, as well as for the Lasso-regularized Cox regression. The "survival" R-package (Therneau & Grambsch, 2000) was used for Cox regression modeling.

4.4.3 Missing data

There were no missing c-miR data. Regarding Study II, missing lifestyle data (physical activity: 30.9% and BMI: 4.5%) occurred due to incomplete questionnaire responses. Missing data were assumed to occur at random, and

multiple imputation with 50 iterations was used to create and analyze 50 multiply imputed datasets using the "mice" R-package (van Buuren & Groothuis-Oudshoorn, 2011) with default settings. All lifestyle variables, as well as sex, age, pathogenic MMR variant, cancer status, and c-miR expression, were used for imputation of each lifestyle variable, and results were pooled using the "pool" function in mice.

4.5 Bioinformatic analyses

4.5.1 Sequencing data preprocessing

Sequencing output data were converted to FASTQ format using bcl2fastq software (v.2.20, Illumina). The QIAseq sequencing adapters were trimmed from the 3′ end of the reads with FastX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) using default parameters with minimum alignment length -M 19. Only clipped reads >20 bp in length were selected for downstream analysis. After adapter clipping, the reads were trimmed to 22 bp to enrich miR-sequences and then quality filtered with FastX-toolkit. Only high-quality reads (Phred score > 25) were selected for alignment to the reference genome. Before alignment, all four sample lanes were merged to obtain the overall sample read count and to ensure better mapping quality. Samples that had < 1 M reads were excluded from the analyses. Subsequently, the preprocessed reads were mapped to human mature miRnome (miRbase v.22) (Griffiths-Jones et al., 2006) with the Bowtie (Langmead et al., 2009) alignment tool for single-end data with v-mode and best strata parameters. Only uniquely mapped miR reads were selected for differential expression analysis. All the steps in the preprocess pipeline were conducted with the Puhti supercomputer cluster (CSC, Finland) using in-house shell-scripts and algorithms. FastQC was used for quality controls (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

4.5.2 Differential gene expression analysis

Differential expression analyses from raw c-miR counts were based on statistical procedures of EdgeR (Robinson et al., 2009) and DESeq2 (Love et al., 2014) packages and conducted in R-studio (v. 4.1.2) (R Core Team, 2022). Briefly, the analyses were performed on c-miR raw read count matrices after the low expressed genes were filtered out, normalized with the median of ratios method, and variance stabilized in DESeq2. Circulating miRs that had more than 1 count per million in 70% of the samples in a group analysis were selected for the analysis. Filtered and normalized c-miR counts were used to set up a design matrix in DESeq2 that was adjusted for sex and batch. The Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) in DESeq2 was used to correct for multiple testing.

4.5.3 Target gene prediction and pathway analysis

Putative c-miR-target gene prediction was performed using the mirWalk (Sticht et al., 2018) tool, which utilizes a random forest-based approach, an ensemble learning method based on multiple decision trees, to predict target genes. Only the predicted c-miR-target genes targeting 3′ untranslated region with experimental validation from miRTarBase (Hsu et al., 2011), which were included and verified in mirDB (X. Wang, 2008) and TargetScan (Agarwal et al., 2015) databases, were selected for downstream over-presentation analysis. Over-presentation of gene ontology (Ashburner et al., 2000), biological processes, Kyoto Encyclopedia of Genes and Genomes (Kanehisa & Goto, 2000), and Reactome pathways were conducted with miRWalk, Reactome (Fabregat et al., 2016; Joshi-Tope et al., 2005), and Search Tool for Retrieval of Interacting Genes/Proteins (Snel et al., 2000; Szklarczyk et al., 2021). These tools provide a standard over-presentation analysis based on hypergeometric tests. The Catalogue of Somatic Mutations in Cancer (Tate et al., 2019) project database was used for the predicted target gene investigation.

5 RESULTS

5.1 Participant characteristics

Study participant characteristics are described in Table 6. Most of the LS carriers in all the studies had the pathogenic *MLH1* variant (> 60%), followed by *MSH2*, *MSH6*, and *PMS2*. The mean ages of the participants across all three studies were 53.5 years and 55.5 years for the cancer-free LS carriers and 62.0 years for those who had or developed CRC, including sporadic CRC and LS CRC. Regarding Studies II and III, those who developed cancer had pathogenic *MLH1* variants, a history of previous cancers, and a higher BMI compared to their cancer-free counterparts. In Study II, the mean surveillance time was 1.5 years for those who developed CRC and 3.5 years for those who remained cancer-free. In Study II, those LS carriers who developed cancer had higher physical activity levels than those whose remained cancer-free, whereas in Study III cancer-free LS carriers reported higher physical activity. In Study III, the average surveillance times were 27.4 years for those who developed cancer, and 29.6 for those who did not. No loss to follow-up occurred in either Studies II or III.

TABLE 6 The study participants' characteristics.

Variable	Study I			Study II		Study III	
	Cancer-free LS	Sporadic CRC	Control	Cancer-free LS	LS CRC	Cancer-free LS	LS cancer
N	81	24	37	92	9	242	223
Age, years (SD)	59.5 (10.7)	69.8 (9.9)	54.9 (10.7)	57.5 (11.4)	51.4 (10.9)	49.6 (14.0)	64.8 (10.4)
Sex, N (%)							
Male	40 (49.4)	10 (41.6)	18 (48.6)	44	6 (66.7)	114 (47.1)	101 (45.3)
Female	41 (50.6)	14 (58.4)	19 (51.4)	48	3 (33.3)	128 (52.9)	122 (54.7)
MMR status, N (%)							
MLH1	50 (61.7)	-	-	60 (65.2)	8 (88.9)	168 (69.4)	149 (66.8)
MSH2	17 (21.0)	-	-	17 (18.5)	1 (11.1)	38 (15.7)	43 (19.3)
MSH6	12 (14.8)	-	-	13 (14.1)	-	36 (14.9)	29 (13.0)
PMS2	2 (2.5)	-	-	2 (2.2)	-	-	2 (0.9)
Cancer history, N (%)							
Yes	42 (51.9)	-	-	44	6 (66.7)	0 (0)	134 (60.1)
No	39 (48.1)	-	-	48	3 (33.3)	242 (100)	89 (39.9)
BMI, kg/m² (SD)[A]	27.3 (5.7)	27.6 (6.3)	28.0 (6.2)	27.7 (6.1)	30.4 (3.3)	26.6 (4.8)	27.2 (5.9)
Physical activity, MET-h/day (SD)[B]	-	-	-	3.7 (3.6)	7.4 (4.5)	4.8 (4.3)	3.7 (4.0)
Surveillance time, years (SD)	-	-	-	3.5 (0.6)	1.5 (1.2)	29.6 (14.0)	27.4 (10.4)

[A] Missing BMI data of LS carriers: Study I = 12; Study II = 5; Study III = 1.
[B] Missing physical activity data of LS carriers: Study II = 34; Study III = 3.
BMI = body mass index; CRC = colorectal cancer; LS = Lynch syndrome; MET = metabolic equivalent task; MMR = mismatch repair; SD = standard deviation.

5.2 The systemic circulating microRNA landscape of cancer-free Lynch syndrome carriers

A small RNA sequencing experiment was performed to inspect the systemic c-miR signatures in the study cohorts and to characterize global expression patterns. In total, four separate sequencing runs were performed that resulted in 1349 c-miRs common to cancer-free LS carriers, sporadic CRC patients, and controls with an average depth of 3.2 M reads per sample and over 70% alignment rate. After the removal of c-miRs with low and/or zero raw reads count, 228 c-miRs were identified and used in the downstream analyses.

Hsa-miR-206 was observed as downregulated in male LS carriers and thus the models were adjusted with sex. Even though no difference in c-miR expression among the different MMR variants or among those with or without previous cancers was observed, models were adjusted with the MMR variant as recommended by PLSD studies (Møller et al., 2023). It was discovered that 37 out of 228 c-miRs displayed differential expression in cancer-free LS carriers when compared to the controls (Figure 3A). However, no differential expression was

observed when the c-miR profiles of cancer-free LS carriers were compared to sporadic CRC patients (Figure 3B).

Of the differentially expressed c-miRs, 14 were upregulated and 23 were downregulated in cancer-free LS carriers with log2 fold changes varying between -1.56 and 0.94. Hsa-miR-155-5p, hsa-let-7c-5p, and hsa-let-7e-5p had the most significant upregulation within cancer-free LS carriers, whereas hsa-miR-320a-3p was the most significantly downregulated, followed by hsa-miR-15a-5p, hsa-miR-186-5p, and hsa-miR-3615, respectively (Table 7). Regarding the cancer-free LS carriers and sporadic CRC patients, hsa-miR-10a-5p was the most significantly differentially expressed c-miR between the groups, although it did not reach the FDR < 0.05 level.

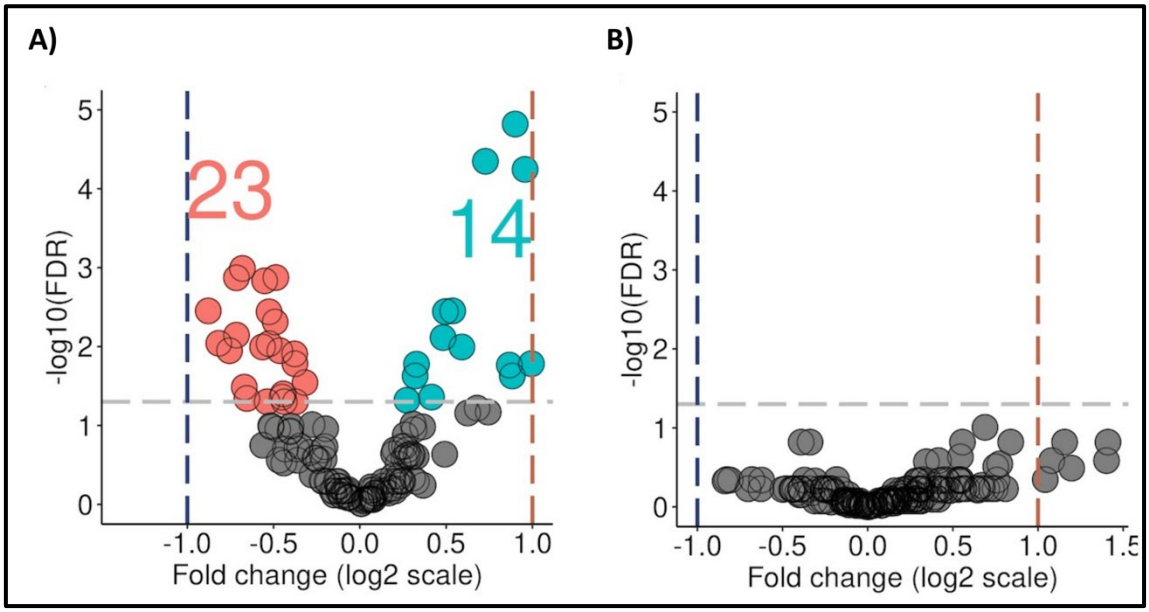


FIGURE 3 Differentially expressed circulating microRNAs (c-miRs) between cancer-free Lynch syndrome (LS) carriers and controls (A) and cancer-free LS carriers and sporadic colorectal cancer patients (B). Red = downregulated c-miRs, blue = upregulated c-miRs, and gray = not differentially expressed c-miRs. Red dashed line = log2 fold change of 1, blue dashed line = log2 fold change of -1, and gray dashed line = -log10 false discovery rate (FDR) significant at <0.05 level.

TABLE 7 Differentially expressed circulating microRNAs between the study groups.

Cancer-free LS vs. Controls			**Cancer-free LS vs. Sporadic CRC patients**		
c-miR	log2FC	FDR	c-miR	log2FC	FDR
hsa-miR-155-5p	0.886	**<0.001**	hsa-miR-10a-5p	0.689	0.098
hsa-let-7c-5p	0.698	**<0.001**	hsa-miR-1180-3p	1.156	0.151
hsa-let-7e-5p	0.938	**<0.001**	hsa-miR-126-3p	-0.391	0.151
hsa-miR-320a-3p	-0.747	**0.001**	hsa-miR-148b-3p	-0.339	0.151
hsa-miR-15a-5p	-0.666	**0.001**	hsa-miR-196a-5p	1.410	0.151
hsa-miR-186-5p	-0.560	**0.001**	hsa-miR-320a-3p	0.556	0.151
hsa-let-7a-5p	0.540	**0.002**	hsa-miR-320b	0.839	0.151
hsa-miR-185-5p	-0.474	**0.002**	hsa-miR-486-5p	0.544	0.233
hsa-miR-3615	-1.562	**0.002**	hsa-miR-200b-3p	1.402	0.259
hsa-miR-3613-5p	-0.876	**0.004**	hsa-miR-223-3p	0.415	0.259
hsa-miR-22-3p	-0.536	**0.004**	hsa-miR-320c	1.080	0.259
hsa-miR-339-5p	-0.879	**0.004**	hsa-miR-185-5p	0.344	0.268
hsa-miR-19b-3p	-0.482	**0.006**	hsa-miR-483-5p	0.776	0.285
hsa-miR-15b-5p	-0.540	**0.006**	hsa-miR-222-3p	0.752	0.310
hsa-miR-451a	-0.727	**0.006**	hsa-miR-2110	1.197	0.323
hsa-miR-484	-0.798	**0.009**	hsa-let-7d-3p	0.460	0.451
hsa-let-7f-5p	0.338	**0.012**	hsa-miR-11400	-0.827	0.451
hsa-miR-10b-5p	0.446	**0.012**	hsa-miR-134-5p	-0.679	0.451
hsa-miR-25-3p	-0.406	**0.012**	hsa-miR-193a-5p	0.534	0.451
hsa-miR-141-3p	0.914	**0.012**	hsa-miR-25-3p	0.323	0.451
hsa-miR-27a-3p	-0.386	**0.014**			
hsa-miR-32-5p	-0.534	**0.019**			
hsa-miR-107	-0.431	**0.020**			
hsa-miR-23a-3p	-0.492	**0.024**			
hsa-miR-125a-5p	0.405	**0.026**			
hsa-miR-221-3p	-0.339	**0.031**			
hsa-miR-486-5p	-0.499	**0.031**			
hsa-miR-126-3p	0.322	**0.031**			
hsa-miR-424-5p	-0.693	**0.031**			
hsa-miR-92a-3p	-0.409	**0.036**			
hsa-let-7i-5p	0.278	**0.038**			
hsa-miR-200a-3p	0.835	**0.038**			
hsa-miR-222-3p	-0.676	**0.047**			
hsa-miR-125b-5p	0.474	**0.047**			
hsa-miR-27b-3p	0.417	**0.047**			
hsa-miR-15b-3p	-0.823	**0.047**			
hsa-miR-206	0.869	**0.049**			

N cancer-free LS = 101; *N* controls = 37; *N* sporadic CRC patients = 24. c-miR = circulating microRNA; CRC = colorectal cancer; FDR = false discovery rate; log2FC = log base 2 fold change; LS = Lynch syndrome. FDR significant at <0.05 level.

5.3 Pre-diagnostic circulating microRNA signature in Lynch syndrome colorectal cancer risk stratification

During the four years of prospective surveillance, nine out of 101 cancer-free LS carriers developed CRC within a mean surveillance time of 1.5 years (0.1–3.3 years) (Table 6). Out of the 37 c-miRs, Lasso selected hsa-miR-10b-5p, hsa-miR-19b-3p, hsa-miR-200a-3p, hsa-miR-27b-3p, and hsa-miR-3615 as the best predictors that separated LS carriers who developed CRC from those who remained cancer-free during the surveillance. It was observed that the baseline expression of all these c-miRs, except for hsa-miR-27b-3p, was higher in those LS carriers who developed cancer than in those who did not, although it did not reach statistical significance at the $p < 0.05$ level (Figure 4A–E). Of the five c-miRs, hsa-miR-10b-5p, hsa-miR-200a-3p and hsa-miR-3615 were independently associated with an increased cancer risk, and the model showed excellent concordance (C-index = 0.94) (Table 8). Furthermore, hsa-miR-10b-5p, hsa-miR-27b-3p, and hsa-miR-200a-3p were upregulated in cancer-free LS carriers when compared to controls, whereas hsa-miR-19b-3p and hsa-miR-3615 were downregulated in cancer-free LS carriers compared to controls (bolded in Table 7).

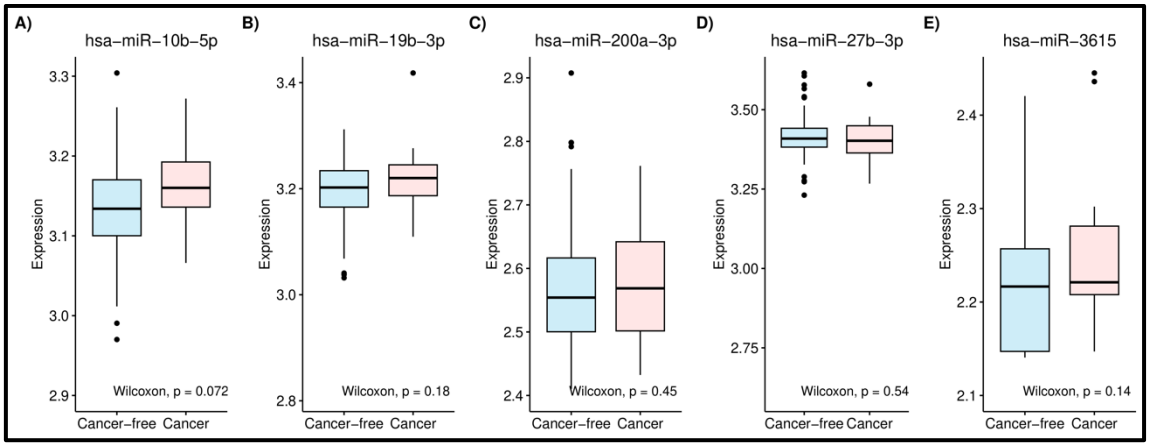


FIGURE 4 Expression differences of hsa-miR-10b-5p (A), hsa-miR-19b-3p (B), hsa-miR-200a-3p (C), hsa-miR-27b-3p (D), and hsa-miR-3615 (E) between Lynch syndrome carriers who got colorectal cancer and cancer-free Lynch syndrome carriers. The expression values on the Y-axis are presented as normalized and variance stabilized circulating microRNA counts. P-value significant at <0.05 level.

TABLE 8 Cox regression model fits of circulating microRNAs in the full sample.

c-miR	HR	95% CI	β	P-value
hsa-miR-10b-5p	22.78	2.672-194.220	3.130	**0.004**
hsa-miR-19b-3p	7.29	0.663-80.136	1.996	0.104
hsa-miR-200a-3p	6.69	1.216-36.792	1.901	**0.029**
hsa-miR-27b-3p	0.22	0.034-1.411	-1.521	0.102
hsa-miR-3615	21.29	1.007-449.880	3.058	**0.049**

C-index= 0.936 (se = 0.02)
Likelihood ratio test = 22.91 on 5 *df*, *p* = 4e-04
Wald test = 16.32 on 5 *df*, *p* = 0.006
Score (log rank) test = 22.42 on 5 *df*, *p* = 4e-04
β = regression coefficient; HR = hazard ratio; 95% CI = 95% confidence interval; C-index = Harrel's concordance (C) index. P-value significant at <0.05 level.

It was observed that a risk sum score composed of the five Lasso-obtained c-miRs was significantly associated with an increased risk of cancer incidence (HR 2.72; 95% CI 1.67–4.42; β = 1.00; C-index = 0.94) also after adjusting for sex and MMR variant (HR 2.54; 95% CI 1.57–4.13; β = 9.34; C-index = 0.92) (Table 9). The mean c-miR risk sum score was higher in those LS carriers who developed cancer compared to those who did not ($p < 0.001$) (Table 9). An internal validation with fivefold cross-validation of this risk prediction model resulted in an average C-index of 0.84 (0.60–1.00), thus presenting good concordance (Table 9).

TABLE 9 The association of circulating microRNA risk sum score and colorectal cancer incidence.

Model	Mean (cancer-free)	Mean (CRC)	P-value	HR	95% CI	β	P-value	C-index
Risk sum score (unadjusted)	54.99	57.73	<0.001	2.72	1.67-4.42	1.000	**<0.001**	0.94 (se 0.022)
Risk sum score (adjusted)	-	-	-	2.54	1.57-4.13	0.934	**<0.001**	0.92 (se 0.021)

5-fold cross-validation mean C-index = 0.84 (0.6-1.0).
β = regression coefficient; C-index = Harrel's concordance index; CRC = colorectal cancer; HR = hazard ratio; 95% CI = 95% confidence interval; se = standard error. P-value significant at <0.05 level.

Hsa-miR-200a-3p correlated with hsa-miR-10b-5p (rho = 0.29; $p < 0.01$) and hsa-miR-27b-3p (rho = 0.23; $p < 0.01$) whereas hsa-miR-27b-3p correlated negatively with hsa-miR-19b-3p (rho = -0.20, $p < 0.05$) and hsa-miR-3615 (rho = -0.43, $p < 0.001$) thus displaying correlation and expression concordance (Figure 5A). It was discovered that these five c-miRs targeted 180 distinct genes with experimental validation (Studies I and II). These genes were observed to form high confident interacting gene hubs ($p < 0.01$) around common oncogenes and TSGs such as *TP53*, *CDKN1A*, *CDKN2A*, *SMAD2*, *FOXO1*, *EGFR*, and *CREB1* genes (Figure 5B and Table 1). A pathway analysis conducted on those experimentally confirmed

target genes showed that these genes were significantly enriched ($p < 0.05$) in several cancer-associated pathways and hallmark properties, including regulation of apoptosis, p53-/FOXO-/AKT-/TGFB-signaling and cellular senescence, among others (Figure 5C).

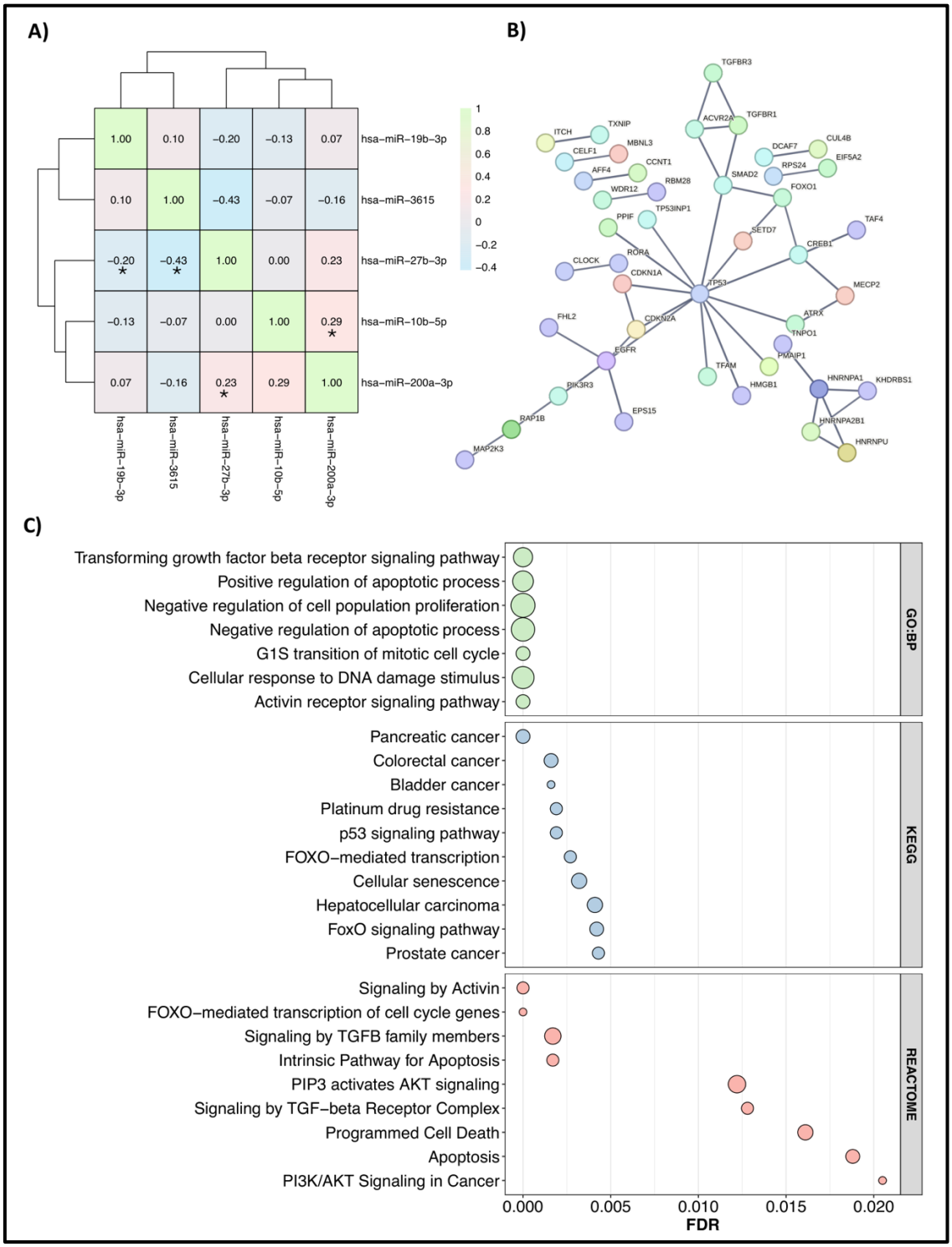


FIGURE 5 Correlation heatmap of hsa-miR-10b-5p, hsa-miR-19b-3p, hsa-miR-200a-3p, hsa-miR-27b-3p, and hsa-miR-3615 (A), and the most important gene nodes (B) and pathway analysis (C) of their target genes. FDR = false discovery rate; GO:BP = Gene ontology: Biological process; KEGG = Kyoto encyclopedia of Genes and Genomes. P-value significant at <0.05 level (*).

5.4 Lifestyle habits and Lynch syndrome cancer risk

The c-miR risk sum score did not show any correlation to physical activity (r = 0.173; p = 0.121) or BMI (r = 0.167; p = 0.101) (Table 10).

TABLE 10 Correlations between circulating microRNAs, body mass index, and physical activity.

Variable	r	95% CI	P-value
Physical activity	0.173	[-0.05, 0.39]	0.121
BMI	0.167	[-0.03, 0.36]	0.101

BMI = body mass index; r = Pearson correlation coefficient; 95% CI = 95% confidence interval. P-value significant at <0.05 level.

The associations between adult life body weight and risk of cancer are presented in Table 11. For consistency, all results in the text regarding body weight and cancer risk are presented only from models adjusted for height, MMR gene, education, alcohol consumption, smoking status, and the use of anti-inflammatory drugs. Longitudinal weight gain increased CRC risk by 3% in males (HR 1.03; 95% CI 1.01–1.05) but not in females. Regarding overall cancer risk, longitudinal weight gain increased cancer risk by 2% in males (HR 1.02, 95% CI 1.00–1.04) but no associations were observed in females. Among females, near-term weight gain within the 10-year interval before CRC diagnosis was associated with a 4% decreased risk of CRC (HR 0.96; 95% CI 0.92–0.99). No associations between near-term weight gain and CRC risk were observed in males.

The associations between physical activity and cancer risks are presented in Table 12. Longitudinal physical activity was observed to lower the overall cancer risk by 63% (HR 0.37; 95% CI 0.15–0.98) in males. No longitudinal associations between physical activity and cancer risk was observed in females. In the near term, physical activity was not associated with overall cancer risk in either males or females. There was no significant association between physical activity and the risk of CRC.

TABLE 11 Associations of longitudinal and near-term body weight change and Lynch syndrome cancer risk.

	Unadjusted model				Adjusted model			
Males	Events	Observations	HR (95% CI)	*P*	Events	Observations	HR (95% CI)	*P*
CRC								
Longitudinal change	60	610	1.02 (1.00–1.03)	0.023	57	579	1.03 (1.01–1.05)	**0.004**
Near-term change	60	185	1.00 (0.98–1.01)	0.695	57	174	1.00 (0.98–1.02)	0.861
All cancers								
Longitudinal change	77	610	1.02 (1.00–1.03)	0.048	74	579	1.02 (1.00–1.04)	**0.022**
Near-term change	77	185	0.99 (0.97–1.00)	0.345	74	174	0.99 (0.97–1.01)	0.424
Females								
CRC								
Longitudinal change	50	758	0.99 (0.96–1.01)	0.258	48	720	0.99 (0.96–1.02)	0.454
Near-term change	50	221	0.98 (0.95–1.00)	0.106	48	209	0.96 (0.92–0.99)	**0.015**
All cancers								
Longitudinal change	95	758	0.99 (0.97–1.00)	0.290	91	720	1.00 (0.98–1.02)	0.887
Near-term change	95	221	0.98 (0.97–1.00)	0.059	91	209	0.98 (0.96–1.00)	0.059

Model adjusted with mismatch repair gene variant, height, education, smoking, alcohol use, and non-steroidal anti-inflammatory medication. HR = hazard ratio; 95% CI = 95% confidence interval; CRC = colorectal cancer. P-value significant at <0.05 level.

TABLE 12 Associations of longitudinal and near-term physical activity change and Lynch syndrome cancer risk.

	Unadjusted model				Adjusted model			
Males	Events	Observations	HR (95% CI)	*P*	Events	Observations	HR (95% CI)	*P*
CRC								
Longitudinal change	73	683	0.57 (0.25-1.33)	0.194	68	648	0.52 (0.20–1.36)	0.181
Near-term change	73	205	0.93 (0.39–2.24)	0.874	68	192	0.99 (0.36–2.73)	0.938
All cancers								
Longitudinal change	91	683	0.44 (0.19–1.04)	0.063	86	648	0.37 (0.15–0.98)	**0.044**
Near-term change	91	205	0.69 (0.29–1.64)	0.403	86	192	0.74 (0.27–2.01)	0.557
Females								
CRC								
Longitudinal change	59	823	1.16 (0.65–2.10)	0.612	57	789	1.28 (0.65–2.52)	0.471
Near-term change	59	238	0.92 (0.49–1.72)	0.797	57	227	0.99 (0.48–2.02)	0.973
All cancers								
Longitudinal change	110	823	1.31 (0.86–1.97)	0.206	107	789	1.26 (0.79–2.00)	0.341
Near-term change	110	238	1.42 (0.89–2.25)	0.138	107	227	1.34 (0.80–2.23)	0.268

Model adjusted with mismatch repair gene variant, height, education, smoking, alcohol use, and non-steroidal anti-inflammatory medication. HR = hazard ratio; 95% CI = 95% confidence interval; CRC = colorectal cancer. P-value significant at <0.05 level.

6 DISCUSSION

This thesis explored the associations between c-miRs, lifestyle habits and LS CRC incidence. The aims were to characterize the serum-based c-miR landscape of cancer-free LS carriers, to inspect whether any of those c-miRs are potential indications of upcoming CRC, and to determine whether they are associated with modifiable cancer risk factors, such as BMI and physical activity. Furthermore, this thesis retrospectively examined whether weight gain and physical activity are associated with cancer risk in the Finnish LS population.

It was observed that cancer-free LS carriers displayed aberrant serum c-miR expression compared to the healthy non-carriers, but no differential expression was seen between the cancer-free LS carriers and sporadic CRC patients. A panel composed of these differentially expressed c-miRs, including hsa-miR-10b-5p, hsa-miR-19b-3p, hsa-miR-27b-3p, hsa-miR-200a-3p, and hsa-miR-3615, was associated with an increased risk of developing CRC in the near future, thus indicating risk stratification potential during surveillance. The CRC risk predictive c-miR panel did not correlate either with BMI or physical activity, thus suggesting that they are associated with LS CRC risk independently of lifestyle within this population. However, longitudinal weight gain was seen as a cancer risk factor for males and, in the near term, as a protective factor for females in the retrospective analysis. Moreover, longitudinal physical activity was associated with a significant reduction in LS cancer risk in males.

6.1 Circulating microRNAs and Lynch syndrome

In recent years, the expanded accessibility of gene expression data and advancements in methodologies that profile miR targets *en masse* have significantly enriched the comprehension of miR functions, as well as the origins and repercussions of miR dysregulation. Consequently, there is growing interest in applying c-miRs as cancer biomarkers due to their stability and easy collectability, which allows continuous and minimally invasive testing of an

individual. The LS cohort provides an ideal population for such biomarker harvesting due to the well-predicted cancer risk of persons with strong genetic predisposition who are under frequent surveillance and are most likely to develop cancers within short intervals. This thesis presents the first serum-based c-miRs characterized in the Finnish LS population that have cancer biomarker potential.

Balaguer et al. studied miRs extracted from tumors of LS carriers and sporadic CRC patients with verified MSI and normal tissue samples (Balaguer et al., 2010, 2011). They used a set of more than 700 miR-probes with microarray analysis and detected hundreds of differentially expressed miRs among the tissue samples, showing that LS tumors can be separated from sporadic tumors with MSI, as well as that suspected LS samples discern from confirmed LS samples. They detected several differentially expressed miRs with diagnostic potential in LS, including hsa-miR-125b-5p, hsa-miR-137, hsa-miR-622, hsa-miR-192, and hsa-miR-1238. Alternatively, Zhou et al. demonstrated that hsa-miR-137, hsa-miR-520e, and hsa-miR-590-3p are indications of LS using a subset of the same LS tumor samples and normal tissue samples as Balaguer et al. (Zhou et al., 2016).

The reason for the difference in differentially expressed miR numbers between our study and the previous studies is likely explained by the study setting, used specimen type, and methodology. As a general note, since the individuals included in our study were cancer-free at the time of sample collection, their observed response may not demonstrate the same robustness as in cancer patients with diagnosed pathology at specific tissue sites. Moreover, it is worth noting that the c-miRs analyzed in this thesis were obtained from blood samples. We and others have observed that the c-miR abundancies are typically lower in serum and plasma when compared to those from tissue samples (Fehlmann et al., 2016). In addition, the differing detection procedures (microarray vs. small RNA sequencing) may have affected the magnitude of the observed c-miR expression (Chatterjee et al., 2015; Git et al., 2010).

6.1.1 Upregulated circulating microRNAs in Lynch syndrome

The most significant differentially expressed c-miR in the cancer-free LS group was oncogenic hsa-miR-155-5p. Hsa-miR-155-5p is a well-established CRC miR that is commonly more frequent in MSI than in MSS CRC (Earle et al., 2010; Lanza et al., 2007; E.K.O. Ng et al., 2009). Valeri et al. reported that hsa-miR-155-5p targets several MMR genes and that upregulation of hsa-miR-155-5p downregulates *MLH1* and *MSH2* in CRC cell lines (Valeri et al., 2010). Furthermore, Svrcek et al. observed that miR-155 upregulation was associated with sporadic MSI CRCs in patients with inflammatory bowel disease and that it correlated with distant non-cancerous mucosa (Svrcek et al., 2013). The authors suggested that pre-cancerous miR-155 upregulation may promote the proliferation of dMMR clones in the colonic mucosa of inflammatory bowel disease patients that are yet to undergo malignant transformation due to MSI. In alignment with those studies, the results of this thesis showed a modest

upregulation (log2 fold change < 1.00) of hsa-miR-155-5p. The gene enrichment analysis found TSG *SMAD2* to be a key target gene, as seen in previous reports (Fleming et al., 2013; Louafi et al., 2010), which implies that hsa-miR-155-5p might have a role in the modulation of LS carcinogenesis in this cohort.

The let-7 miR family is commonly regarded as tumor suppressors through direct negative regulation of *RAS* oncogenes, such as *KRAS* and *NRAS* (Esquela-Kerscher & Slack, 2006; Roush & Slack, 2008). Ras proteins, as membrane-associated GTPase signaling proteins associated with 10%–30% of human cancers, play a crucial role in controlling cellular growth and differentiation (Prior et al., 2020). Activating mutations in RAS genes drive cellular transformation. Thus, miRs that regulate the expression of these oncogenic proteins are expected to curb cellular proliferation rates. In this thesis, the let-7 family c-miRs were observed to be significantly upregulated, which could hint at increased oncogenic stress in cancer-free LS carriers. A total of more than 120 experimentally verified target genes, which were enriched in several cancer-associated pathways, were found for these c-miRs (Study I). Of the most enriched target genes in the study, the oncogene *NRAS* was targeted by hsa-let-7a-5p and hsa-let-7c-5p, whereas TSG *CDKN1A* was targeted by hsa-let-7e-5p (Study I). Low expression of hsa-let-7i in primary CRC tumors has been reported to be associated with worsened prognosis and distant metastasis (Hur et al., 2015). These results indicate that the let-7 family c-miRs are primarily tumor suppressive in the studied LS cohort.

Furthermore, hsa-miR-10b-5p, hsa-miR-27b-3p, hsa-miR-125a-5p, hsa-miR-125b-5p, hsa-miR-126-3p, hsa-miR-141-3p, hsa-miR-200a-3p, and hsa-miR-206 were also observed to be upregulated in this thesis. In support of our findings, hsa-miR-10b-5p, hsa-miR-200a-3p, hsa-miR-27b-3p, hsa-miR-141-3p, and hsa-miR-125b are well established CRC miRs whose upregulation is associated with larger tumor size (Sheedy & Medarova, 2018), advanced epithelial-to-mesenchymal transition in sporadic CRC (Pichler et al., 2014), advanced stage of sporadic and LS CRC, later age of onset (Moreno et al., 2019; Zhang et al., 2018), and poorer survival (Hur et al., 2015). Although not found in this thesis, hsa-miR-23b and hsa-miR-24a, which belong to the same cluster as hsa-miR-27b-3p found in this thesis, have previously been shown to display upregulation in primary LS CRC tumors (Moreno et al., 2019). Interestingly, miR-125b is linked to non-advanced neoplasms, such as tubular adenomas, and smaller tumor sizes, suggesting its likely involvement in an earlier stage of colorectal carcinogenesis (Yamada et al., 2015). Given that nine out of the 101 cancer-free LS carriers in our cohort developed CRC, this observation supports the findings of this thesis, even if hsa-miR-125b-5p was not included in the CRC risk prediction model. However, hsa-miR-125b-5p was a cancer predictive c-miR when the analysis was not stratified to CRC but considered all cancer cases, further supporting its potential role in the early development of LS CRC. This miR has also been detected as upregulated in earlier LS studies, thus providing evidence of potential early carcinogenesis in our cohort (Balaguer et al., 2011).

Yamaguchi et al. showed that hsa-miR-126 is downregulated in several different CRC types, including sporadic MSI and MSS tumors, as well as LS and familial adenomatous polyposis-derived tumors (Yamaguchi et al., 2014). They suggested that downregulation of this miR induces angiogenesis at the early stage of CRC carcinogenesis (Yamaguchi et al., 2014). Within this concept, by displaying upregulation, it appears that hsa-miR-126-3p might play an opposing role in LS. A previous study by Parasramka et al. reported that the forced upregulation of miR-206 significantly increased HCT116 cell proliferation (Parasramka et al., 2012). This finding aligns with our results, since the HCT116 cell line is commonly used as a proxy for LS due to homozygous nonsense mutation in *MLH1*, which introduces high MSI, a hallmark of LS, to the cell line.

In total, the magnitude of the observed c-miR upregulation was mild, which could be explained by the cancer-free status of the cohort. On the other hand, post-transcriptional gene regulation by miRs operates through mechanisms involving translational inhibition and mRNA destabilization, which results in reduced mRNA levels (Baek et al., 2008; Guo et al., 2010). Studies have shown that mRNA destabilization emerges as the primary effect of miRs when substantial target repression occurs (Eichhorn et al., 2014). However, in the majority of instances, the extent of miR-induced mRNA depletion is modest (Selbach et al., 2008), which is somewhat unexpected considering the broad roles that miRs play in various biological processes and pathologies. If targets are enriched for genes whose products participate in common signaling pathways, as indicated in the gene enrichment analyses (Study I), then the cumulative effect of typically modest interactions—often coupled with a small subset of strongly regulated target genes (Study I)—can elicit a stronger response than could be achieved through the direct regulation of any single gene in isolation. The simultaneous targeting of multiple genes may also facilitate more specific fine-tuning through the regulation of distinct subnetworks (Bracken et al., 2015, 2016). Thus, even mild c-miR upregulation could have a significant impact on LS carcinogenesis, as strongly indicated by the CRC risk prediction model.

6.1.2 Downregulated circulating microRNAs in Lynch syndrome

A majority of c-miRs introduced in this thesis were downregulated in cancer-free LS carriers. Among these c-miRs, the most significantly downregulated c-miR in the LS population was hsa-miR-320a-3p, followed by hsa-miR-15a-5p, hsa-miR-186-5p, hsa-miR-185-5p, and hsa-miR-3615, respectively. In support of our findings, hsa-miR-320a-3p has also been previously reported to display downregulation in LS primary CRC tumors and sporadic CRC primary tumors when compared to controls (Moreno et al., 2019). The downregulation of this miR has also been shown to correlate with advanced tumor stage and non-invasiveness (Hur et al., 2015; Moreno et al., 2019). A downregulation of hsa-miR-15a-5p is associated with better CRC survival (Mullany et al., 2018). In contrast to our findings, the upregulation of hsa-miR-185-5p in tissues and hsa-miR-186-5p in saliva has been linked to sporadic advanced adenomas and early-stage CRC (Shi et al., 2023), as well as late-stage CRC (Rapado-González et al., 2019). In

alignment with this thesis, hsa-miR-3615 has previously been reported to display downregulation in MSI colorectal tumors when compared to their MSS counterparts (Slattery, Lee, et al., 2017).

Of note, hsa-miR-15b-5p, hsa-miR-19b-3p, hsa-miR-23a-3p, hsa-miR-27a-3p, hsa-miR-32-5p, and hsa-miR-92a-3p were also observed to be downregulated. Interestingly, a previous study has shown that hsa-miR-15b-5p and hsa-miR-19b-3p are upregulated in the serum of patients with sporadic advanced adenomas or CRC and possess a high discriminative capacity between CRC patients and controls (Marcuello et al., 2019). Hsa-miR-32-5p has been established as a prognostic miR with higher expression linked to worsened survival in The Cancer Genome Atlas (Weinstein et al., 2013; Yang et al., 2019). Raut et al. reported that a risk score based on plasma derived, including but not limited to, hsa-miR-23a-3p, hsa-miR-27a-3p, and hsa-miR-92a-3p, predicted sporadic CRC incidence in a prospective cohort with a median follow-up time of 6.8 years (Raut et al., 2021). Also, Vychytilova-Faltejskova et al. reported that the upregulation of hsa-miR-23a-3p could be involved in the early steps of sporadic CRC carcinogenesis, and a predictive model including hsa-miR-23a-3p and hsa-miR-27a-3p could have early detection and prognosis prediction potential with high specificity and sensitivity (Vychytilova-Faltejskova et al., 2016). Previous studies have also shown that miR-23a-3p and miR-27a-3p function together within a cluster, and elevated levels of these miRs were observed in *APC* mutant/dMMR invasive adenocarcinomas (Jahid et al., 2012). Additionally, the same authors showed that miR-23a was upregulated in early-stage CRCs, while miR-27a was overexpressed in advanced-stage CRCs. In our study, hsa-miR-27a-3p targeted *KRAS*, among other genes. In contrast to these previous studies, the majority of those c-miRs were downregulated in cancer-free LS carriers. This observation could hint at the ongoing suppression of early-stage carcinogenesis, which was also supported by the gene enrichment analyses conducted in Study I of this thesis.

However, it is important to recognize that there are differences between sporadic CRC and LS CRC carcinogenesis, primarily attributed to dMMR, which may have altered the observed c-miR responses (Balaguer et al., 2011; Moreno et al., 2019). LS cancers are hypermutated with a 100–1000-fold increase in microsatellite mutation rate due to high MSI (Bohaumilitzky et al., 2022; Parsons et al., 1993). Such cancers are known to provoke the immune system via consecutive neoantigen presentation and have been shown to be highly immunogenic, as well as associated with strong immune cell infiltration (Pastor & Schlom, 2021; Seth et al., 2018). In addition, studies have shown that tumor microenvironment infiltrating immune cells are present and highly reactive to neoantigens in LS CRC, but not in cancers that have no MMR deficiency (Kloor & Von Knebel Doeberitz, 2016; Roudko et al., 2021). Thus, immunogenic LS CRCs generally possess a better prognosis, and the patients display enhanced survival over sporadic CRC patients (Stigliano et al., 2008). Since c-miRs may partly originate from immune cells (Matsuzaki et al., 2023; Pritchard et al., 2012), it is tempting to speculate that the observed global dysregulation of the c-miR

landscape in LS could be due to the dMMR-driven enhanced immunogenicity. However, even if we did not have the opportunity to characterize the circulating immune cell landscape of our cohort, it has been reported that normal colonic mucosa of cancer-free LS carriers shows enhanced immune cell infiltration compared to controls (Bohaumilitzky et al., 2022). Whether such alterations mirror circulation and/or c-miR responses remains to be elucidated in future work.

6.1.3 Circulating microRNAs and colorectal cancer risk in Lynch syndrome

Several cancer risk models have previously been developed for sporadic CRC, primarily for prognostic (Yang et al., 2019; Zhan et al., 2021) and diagnostic (Pardini et al., 2023) purposes but also for classifying consensus molecular subtypes (Adam et al., 2022) and predicting incident CRC (Raut et al., 2021). Hsa-miR-10b-5p was the only overlapping miR between those models and ours (Yang et al., 2019), whereas isoforms of hsa-miR-19b-3p (19a) and hsa-miR-27b-3p (27a) were also identified by Raut et al. in their risk prediction model (Raut et al., 2021). Other miRs that were not included in our model but were part of the aforementioned models and found to be dysregulated in LS in this thesis included hsa-miR-141-3p, hsa-miR-23a-3p, hsa-miR-92a-3p, and hsa-miR-155-5p.

To our knowledge, this thesis introduced the first high-concordance serum-based c-miR CRC risk prediction model that was able to separate cancer-free LS carriers from those who developed CRC during a prospective surveillance. Importantly, the cross-validation of the risk prediction model showed that the five c-miR risk sum score has potential for LS patient risk stratification across diverse cohorts with varying numbers of events and surveillance times. This observation is valuable because the variation in individual cancer risk is high among LS carriers, and the implementation of intense screening programs is not uniformly effective.

Of the five c-miRs applied in our model, all except for hsa-miR-27b-3p displayed higher expression in LS carriers who developed CRC compared to those who did not. Furthermore, hsa-miR-200-3p displayed a positive correlation with hsa-miR-10b-5p and hsa-miR-27b-3p, while hsa-miR-27b-3p exhibited a negative correlation with hsa-miR-19b-3p and hsa-miR-3615. The target gene analysis of these c-miRs showed them to target well-established TSGs, such as *TP53* (hsa-miR-19b-3p), *TP53INP1* (hsa-miR-27b-3p), *CDKN1/2A* (hsa-miR-10b-5p), *SMAD2* (hsa-miR-27b-3p and hsa-miR-200a-3p), *FOXO1* (hsa-miR-27b-3p) and *TGFBR1/3* (hsa-miR-27b-3p), as well as oncogenes *CREB1* (hsa-miR-10b-5p), *EGFR* (hsa-miR-200a-3p), and *MAP2K3* (hsa-miR-19b-3p) (Sondka et al., 2018; Tate et al., 2019). These genes formed significantly interconnected hubs, which further indicated a similar role and biological connection among them. In addition, all of these genes were enriched in several cancer-relevant biological pathways. Even though mechanistic experiments on c-miR–mRNA interactions were not conducted, these findings suggest co-regulation among these c-miRs.

The CRC risk prediction model outlined in this thesis demonstrates credibility, although some peculiarities remain. Hsa-miR-19b-3p and hsa-miR-

27b-3p were not independently associated with CRC in our model. Interestingly, hsa-miR-27b-3p was actually associated with a lower risk of CRC incidence, which appeared counterintuitive. To speculate, this protective effect could be observed because the model considers the overall expression pattern and regulatory interactions of multiple miRs. Thus, even if hsa-miR-27b-3p likely promotes carcinogenesis by targeting the abovementioned TSGs, its impact on the overall cancer phenotype may be attenuated by other factors considered in the risk prediction model. For example, the negative correlation between hsa-miR-27b-3p, hsa-miR-19b-2p, and hsa-miR-3615—coupled with the association of miR-19b with increased cancer risk in the model and its dual targeting of TSGs and oncogenes—indicates a potential antagonistic relationship between these c-miRs in cancer development. It should also be noted that the dMMR phenotype may introduce anomalies to the c-miR response.

Collectively, based on our cross-sectional and longitudinal analyses, along with bioinformatic investigations, the c-miRs identified in this thesis emerge as potential indicators of impending LS CRC. Consistent with findings from prior studies involving sporadic CRC patients and demonstrated by our risk prediction model, c-miRs show promise for patient risk stratification in LS. However, given the exploratory nature of our results and the limited testing of the risk prediction model in a small cohort composed mainly of *MLH1* carriers, further validation with larger and more diverse cohorts is imperative. Even if the clinical utility of c-miRs remains to be elucidated in the future, this thesis provides a solid background for further investigation into their potential roles in CRC risk assessment and patient stratification.

6.1.4 Circulating microRNAs and lifestyle habits

The risk of various LS cancers is significantly elevated by sedentary behavior and excess body weight, while physical activity and maintaining a healthy body weight have been shown to mitigate these risks (Coletta et al., 2019; Dashti et al., 2018; Power et al., 2024). As discussed in the literature review, previous studies have suggested that miRs may mediate the mechanisms through which exercise has beneficial effects on health and may prevent cancer (Dufresne et al., 2018; Garai et al., 2021). This thesis observed that the CRC-predicting c-miR signature showed no association with body weight or physical activity.

This finding deviates slightly from our previous results. In Study II, when we modeled the overall LS cancer risk, including CRC, using a c-miR signature largely similar to the one introduced in this thesis, we observed a clear correlation between the risk sum score and BMI. Those results indicated potential links between c-miRs, lifestyle, and cancer through metabolic dysregulation occurring via *p53*, *FOXO*, and cellular senescence pathways. The signature for overall cancer risk in that study featured hsa-miR-125b and hsa-miR-3613, replacing hsa-miR-27b-3p and hsa-miR-19b-3p, respectively. Hence, it appears that hsa-miR-27b-3p and hsa-miR-19b-3p may contribute differently to overall cancer risk compared to their roles specifically within the context of CRC. Regarding the other c-miRs in our CRC model, hsa-miR-10b-5p and hsa-miR-200a-3p have

previously been linked with increased levels of plasma total cholesterol and dysregulated lipid metabolism (Mens et al., 2020; Ruiz-Roso et al., 2020). Alternatively, we observed hsa-miR-27b-3p to target *FOXO1*, a transcription factor known to regulate immune response and inflammation (D'Onofrio et al., 2023). To speculate, this interaction could have a modulating effect on immune response (D'Onofrio et al., 2023; Dufresne et al., 2018). Since LS cancers are highly immunogenic, the potential immunosuppressive effect of hsa-miR-27b and its potential antagonistic relationship with hsa-miR-19b might have a confounding effect. However, without further research, this suggestion remains purely speculative.

It should be noted that the average time period between lifestyle data and blood sample collection was two years, which could have influenced the findings presented in this thesis. However, this timeframe had no effect when we modeled the associations with overall cancer risk. Presumably, a more systematic approach, including matching collection time points and time series data, to model this potential interaction could be beneficial. Thus, further studies are needed to determine whether c-miRs could modulate the beneficial effects of lifestyle on the risk of LS CRC.

6.2 Lifestyle habits and Lynch syndrome cancer risk

The latest recommendations from the World Cancer Research Fund Continuous Update Project Expert Report advocate for the reduction of excess body weight, the increase in physical activity, and the minimization of alcohol and tobacco consumption to lower the risk of CRC (Clinton et al., 2020). While these lifestyle risk factors are known to be associated with CRC in the general population, their association with the development of CRC in LS has remained less well characterized. This thesis found that an overall increase in total body weight throughout the lifespan slightly elevated the risk of CRC in males. Regarding females, near-term weight increases are associated with a decreased risk of CRC.

Body weight typically accumulates during an adult's lifespan. However, aging is associated not only with increased body weight due to fat accumulation but also with changes in body composition (Sillanpää et al., 2014). Starting around age 30, muscle mass tends to decline, accelerating after 50, especially in females due to menopause (Juppi et al., 2020). Typically, males are predisposed to androgen-type fat accumulation throughout their lives (Szulc et al., 2017). Conversely, females tend to have gynoid-type fat distribution at pre-menopause until transitioning to androgen-type at post-menopause (Kirchengast et al., 1997). These age-related weight trends differ between sexes, which could potentially explain our varied findings. The link between weight gain and CRC risk may also be influenced by hormonal factors (Campbell, Newcomb, et al., 2007; Friedenreich et al., 2021), particularly estrogen levels from adipose tissue (Nelson & Bulun, 2001), which may provide some protection through the anti-inflammatory action of estrogens (Straub, 2007). A recent meta-analysis found

that obesity may be less of a risk factor in LS–associated endometrial cancer, and the use of hormonal contraceptives was associated with a decreased risk of endometrial cancer (Power et al., 2024). However, the specific role of estrogen in cancer risk remains speculative, as this study did not explore hormone therapy or measured estrogen levels.

In alignment with the findings of this thesis, a recent meta-analysis composed of seven eligible studies considering lifestyle habits and LS cancer risk found that obesity and lack of physical activity were associated with significantly increased CRC risk (Power et al., 2024). Similar results have been published previously. In 2019, a qualitative evidence synthesis without meta-analysis concluded that excess body weight is a significant risk factor for LS CRC, especially in males (Coletta et al., 2019). Interestingly, when examining overall cancer risk, it was observed in this thesis that males who consistently engaged in more intensive physical activities throughout adulthood had a significantly reduced cancer risk. However, when only CRC was considered the endpoint, no association was found in either males or females in this thesis.

Physical activity may exert its beneficial effect partly by reducing excess body adiposity, which in turn enhances metabolic function and reduces chronic low-level inflammation (Friedenreich et al., 2021). Jokela et al. found that the LS cohort shows similarity with the sporadic CRC cohort regarding inflammation marker GlycA signatures, thus hinting at increased inflammation in LS already in a cancer-free state (Jokela et al., 2024). A recent study by Deng et al. composed of 21 LS patients showed that high-intensity physical activity reduced inflammation markers in the colon and blood and increased immune infiltration in the colonic mucosa (Deng et al., 2023), which has also been observed previously by Bohaumilitzky et al. (2022). Importantly, non-steroidal anti-inflammatory drug usage may abrogate the increased cancer risk, especially in obese individuals (Burn et al., 2020; Movahedi et al., 2015), thus providing more evidence that inflammation has a key role in LS CRC prevention. Thus, it appears that physical activity could positively modulate LS cancer risk by further increasing immunogenicity and reducing inflammation. Evidence from the general population also endorses the immune-enhancing effect of physical activity in reducing the risk of cancers, including CRC (Friedenreich et al., 2010).

Overall, our findings underscore the significance of maintaining a healthy weight and engaging in physical activity throughout life for cancer prevention, particularly among male LS carriers. Given the limited number of participants in our study, the observed association between body weight and heightened cancer risk suggests that the impact of these modifiable behavioral risk factors might be accentuated in LS carriers due to their strong genetic predisposition. Knowledge of modifiable risk factors is an important determinant of adherence to lifestyle recommendations (Hoedjes et al., 2023), indicating that it could be advisable to monitor these modifiable risk factors during routine healthcare visits. Adopting an optimal lifestyle may mitigate the significant genetic predisposition to cancer and contribute to cancer prevention. Additionally, online tools are available for calculating and illustrating individual cancer risk, such as PLSD (Møller, 2020)

and MyLynch (Knapp et al., 2023). These tools could be utilized to enhance motivation for maintaining a healthy lifestyle or making lifestyle changes.

6.3 Critical considerations

This thesis was conducted according to the best available resources. However, like many exploratory and pilot studies, this thesis has potential pitfalls. In general, it should be acknowledged that since the study population was comprised mainly of *MLH1* carriers, the results of this thesis might have limited generalizability to other pathogenic MMR variant carriers. Conversely, since LS is a collection of four distinct diseases, by focusing mainly on *MLH1*, we were able to limit the possible confounding effects originating from the other variants.

Regarding the methodology, there are issues related to c-miR research. A common issue with c-miRs is the identification of their primary and target locations. Therefore, it can only be speculated, for example, from which cell types or tissues the observed c-miRs are derived, which introduces a certain degree of uncertainty over the interpretations of the observations. Furthermore, serum may contain a different c-miR spectrum due to the coagulation process when compared to plasma (K. Wang et al., 2012), and different c-miR carriers may alter in their content (Karvinen et al., 2023). In this thesis, the different c-miR carrier profiles were not characterized. However, not requiring the separation of c-miR carriers to detect a signal could enhance the robustness of our findings in terms of biomarker discovery. Regarding the sequencing methodology, there is no gold standard that sequencing depth should be aimed at when assessing the differential expression of c-miRs. In this study, the aimed mean sequencing depth was 5 M reads per sample, but the achieved mean sequencing depth was 3.2 M reads due to underclustering issues in two out of the four sequencing runs. The underclustering (raw cluster density < 170 K/mm^2) might have masked potential LS-associated c-miRs with low expression.

Another potential pitfall of our study is the small sample, which hindered the validation of the CRC risk prediction model. Despite our best efforts to look for an external validation dataset, we unfortunately did not find a suitable candidate dataset nor had the opportunity to increase our sample size, and thus were unable to validate the Lasso model. Even though several miR datasets are available, for example, through The Cancer Genome Atlas, none of the existing studies had a similar study design as ours and considered sporadic CRC and not LS. Therefore, there is a clear need for more extensive external and internal validation of our findings. Moreover, since we had a small sample of sporadic CRC cases, we cannot exclude the possibility of differing c-miR landscapes between them and cancer-free LS carriers.

In the retrospective lifestyle habit analysis, body weight and physical activity were evaluated using self-recall instruments. Despite a recent study indicating that cross-sectional self-reported measurements of BMI closely aligned with recent direct measurements (Davies et al., 2020), the recall of weight in the

more distant past has lower reliability (Dahl & Reynolds, 2013), and for some of the older participants in this thesis, the recall time was several decades. Furthermore, sex-based discrepancies may exist, with females tending to underestimate their weight and males tending to overestimate it (Tuomela et al., 2019). Recalling physical activity from the distant past has also been shown to exhibit moderate reproducibility, yet this is inadequate at the individual level (Smith et al., 2013). Thus, recall bias might have influenced the risk estimates in this analysis.

A major strength of this study is that we were able to conduct MMR, sex, and CRC stratified analyses with pre-diagnostic LS samples. This approach allowed for a detailed cross-sectional characterization of the c-miR landscape of cancer-free LS carriers, which is crucial when mining potential early-detection biomarkers. We were also able to conduct a prospective analysis that tested minimally invasive patient risk stratification in a high-risk cohort. Instead of an *a priori* chosen gene panel, we conducted a systemic-level investigation, which provided a more comprehensive view of how already identified c-miRs and their putative target genes contribute to distorted biological networks in cancer without potential selection bias. This approach also enabled the discovery of new putative LS CRC-associated c-miRs, as introduced in our risk prediction model. We used robust and up-to-date methodology to interrogate and analyze the c-miR signatures and their associations with LS cancer risk and lifestyle habits. This is particularly important since the Cox proportional hazards regression analysis—widely recognized as the predominant method for modeling covariate information in survival analysis—may encounter limitations when applied to high-dimensional datasets with a low sample size-to-variables ratio. Instead, Lasso, which was used in our risk prediction model development, was introduced to eliminate this limitation (Tibshirani, 1996, 1997). Regarding our cohort, since the study subjects had undergone comprehensive screenings of LS-predisposing mutations, with ascertainment utilizing Amsterdam (Vasen et al., 1999) and Bethesda clinical criteria (Umar et al., 2004) and cascade testing, there were no potentially confounding effects from other potential hereditary CRC syndromes. Finally, to enhance reproducibility, all original articles were published under the gold open access model with comprehensive supplementary material, including the code supplementary files and sequencing data. Regarding the development and validation of the risk prediction model, we followed transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (Collins et al., 2015).

6.4 Future perspectives

Precision oncology is an approach to cancer treatment that involves tailoring medical care to individual patients based on their unique genetic, molecular, and clinical characteristics. This approach aims to optimize treatment outcomes by selecting therapies that are most likely to be effective for a particular patient

while minimizing potential side effects. Regarding LS, precision oncology is ever more needed, as the current evidence strongly shows that the individual cancer risk of a carrier relies heavily on which pathogenic variant they carry, as well as on the sex, age, and cancer history of an individual. Thus, a "one-size-fits-all" approach is no longer justified in treating LS malignancies.

Fortunately, advances in genomic profiling technologies, such as high-throughput sequencing, have enabled the comprehensive profiling of a patient's tumor tissue or blood sample to identify specific genetic alterations, mutations, or other biomarkers associated with their cancer. Once these genomic alterations or biomarkers are identified, targeted therapies can be selected to specifically inhibit the activity of proteins or pathways that drive tumor growth or survival. These targeted therapies may include, for example, immune checkpoint inhibitors, such as PD-1 blockers, which are already demonstrated to be highly efficient in dMMR-driven cancers (Cercek et al., 2022). In addition to guiding treatment selection, precision oncology aims to identify predictive biomarkers that can help clinicians anticipate how a patient will respond to a particular treatment or find those at increased risk of developing cancer. For such purposes, miRs have been under intense investigation for over two decades, but only in sporadic cancer patients. This thesis was the first to investigate the biomarker potential of c-miRs in the context of LS patient risk stratification. The next step in the process would be to investigate whether there are correlations in the miR content between blood samples and tissue samples in a more diverse and larger LS cohort.

The practice of medicine relies on clinical trials, which are at the core of finding suitable treatments and potential biomarkers for clinical use. A search term of "cancer" AND "microRNA" OR "miRNA" from clinicaltrials.gov resulted in 400 hits in total. Interestingly, among these trials are c-miRs, also introduced in this thesis. For instance, a clinical trial (NCT01849952) is currently underway to evaluate miR-10b expression in patients with various subtypes of brain cancer. Additionally, a preclinical trial involving the miR-10b inhibitor drug TTX-MC138 has recently begun (NCT05908773). Furthermore, the correlation between miR-141 expression and radiation resistance has been investigated in prostate cancer patients (NCT02391051). It appears evident that the potential of miR diagnostics continues to be explored extensively in cancer research.

A more comprehensive understanding of how modifiable cancer risk factors are associated with the unique molecular characteristics of an individual is needed. The evidence discussed in this thesis shows rather clearly that optimal lifestyle habits regarding body weight and amount of physical activity reduce LS cancer risk. Therefore, lifestyle interventions could help reduce the LS cancer burden. In the context of precision oncology, monitoring how the patient's risk-associated molecular profile responds to the treatment during intervention would provide valuable information regarding the efficacy of the chosen therapy, the potential development of resistance mechanisms, and the need for adjustments in treatment strategies to optimize outcomes and minimize adverse

effects. This dynamic approach to treatment monitoring and adaptation is highly valuable for achieving personalized and effective cancer care. Based on the current literature and the results introduced in this thesis, c-miRs, alongside other markers that complement current screening and detection strategies, could have potential for such purposes, but more research is required. Given the significant role of the immune system in LS carcinogenesis, a logical next step in exploring the potential involvement of c-miRs in the interplay between lifestyle factors and LS cancer risk is to broaden our understanding of the immune cell landscape in LS.

As a closing remark, with the continuing expansion of data quantity and computational capabilities, artificial intelligence and machine learning techniques are expected to revolutionize the field of oncology, as suggested by developments in the field of digital pathology (Unger & Kather, 2024). These technologies will enable the mining of vast genomic databases to identify candidate molecules with predictive, prognostic, and diagnostic potential for the treatment of various sporadic and hereditary cancers in the near future. This approach holds immense promise for advancing personalized cancer care by harnessing big data into computational models tailored to individual risk assessment.

7 MAIN FINDINGS AND CONCLUSIONS

The main findings of this thesis are as follows:

1. Cancer-free LS carriers displayed aberrant serum c-miR expression compared to the healthy control group, but no differential expression was observed between the cancer-free LS carriers and sporadic CRC patients. Through targeting enriched and validated oncogenes and TSGs, this c-miR landscape has the potential to track early-stage carcinogenesis in LS.

2. A risk sum score composed of differentially expressed c-miRs, including hsa-miR-10b-5p, hsa-miR-19b-3p, hsa-miR-27b-3p, hsa-miR-200a-3p, and hsa-miR-3615, was associated with increased risk of developing CRC during a four-year surveillance in a high-concordance and cross-validated risk prediction model. This finding highlights the risk stratification potential of c-miRs during surveillance. However, the c-miR risk sum score was not associated with BMI and physical activity.

3. Longitudinal weight gain was associated with increased overall cancer risk and CRC risk in male LS carriers. Near-term weight gain lowered the risk of CRC in female LS carriers. Physical activity was associated with decreased overall cancer risk in male LS carriers, but no association was observed with CRC risk. These results emphasize the importance of weight maintenance and physical activity throughout the lifespan in cancer prevention, especially in male LS carriers.

YHTEENVETO (SUMMARY IN FINNISH)

Lynchin oireyhtymä (LS) on perinnöllinen syöpäalttiusoireyhtymä, joka altistaa yksilön useille syöville. LS johtuu ituradan geenimutaatiosta, joka vaikuttaa DNA-kopiointikorjausmekanismeihin lisäten oireyhtymän kantajien perimän mutaatiotaakkaa. Lisääntynyt mutaatiotaakka altistaa LS-kantajat suuremmalle riskille saada syöpiä nuorella iällä. Aiemmat tutkimukset viittaavat, että LS-syöpäriskiä voidaan kuitenkin alentaa terveellisillä elämäntavoilla. LS-suolistosyöpäennustetta voidaan myös parantaa merkittävästi säännöllisen kolonoskopiaseurannan avulla. Tämän vuoksi oireyhtymän kantajat voivat käydä seulonnoissa koko elämänsä ajan. Nykyisen tutkimustiedon valossa kolonoskopiaseuranta ei kuitenkaan merkittävästi ehkäise suolistosyöpää. Tämän vuoksi seurantamenetelmiä tulisi kehittää muun muassa seulomalla yksilöllisiä merkkimolekyylejä, joiden avulla korkean riskin potilaita voitaisiin ohjata seulontoihin tehokkaammin.

Verenkierron mikro-RNAt (c-miR) ovat pieniä molekyylejä, jotka säätelevät geenien ilmentymistä niiden kohdekudoksissa muun muassa vasteena sairauksiin ja fyysiseen aktiivisuuteen. Ne ovat myös tärkeä osa kudosten ja solujen välistä viestintää, minkä vuoksi verenkierron c-miR-pitoisuudet voivat heijastaa kudoksissa tapahtuvia muutoksia. Tutkimukset osoittavat niiden pystyvän erottelemaan useita syöpätyyppejä toisistaan ja toisaalta ennustavan syövän kehittymistä ei-perinnöllisissä kohorteissa. MiR-tutkimus on kuitenkin ollut erittäin vähäistä LS-kohorteissa huolimatta osoitetusta syöpähoitopotentiaalista. MiR:t ovat kiinnostava kohde tutkittaessa syövän sekä riskitekijöiden molekulaarisia yhteyksiä, koska miR:t toimivat sekä sairauksien diagnostiikassa että potentiaalisina liikuntavaikutusten välittäjinä.

Tässä väitöskirjatutkimuksessa seulottiin sekvensointimenetelmien sekä bioinformatiikan avulla syöpävapaiden suomalaisten LS-kantajien (n = 101) c-miR-profiilit vertaamalla niitä ei-perinnöllisten suolistosyöpäpotilaiden (n = 24) sekä terveiden verrokkien (n = 37) vastaaviin profiileihin. Pitkittäisasetelmassa tutkittiin, voitiinko tunnistettujen c-miR:n avulla ennustaa suolistosyöpään sairastumista. Lisäksi tehtiin retrospektiivinen elämäntapakysely (n = 465), jolla selvitettiin, olivatko kehon paino tai fyysinen aktiivisuus yhteydessä LS-syöpäriskiin. Näiden elämäntapatekijöiden yhteyttä suolistosyöpää ennustaviin c-miR-profiileihin kartoitettiin myös.

Tässä väitöskirjatutkimuksessa havaittiin syöpävapaiden LS-kantajien c-miR-profiilien poikkeavan merkittävästi terveistä verrokeista muttei ei-perinnöllisistä suolistosyöpäpotilaista. Näistä viisi miR:ta, hsa-miR-10b-5p, hsa-miR-19b-3p, hsa-miR-27b-3p, hsa-miR-200a-3p ja hsa-miR-3615, muodostivat riskiennustepaneelin, joka ennusti suolistosyöpään sairastumista seurannan aikana mutta ei ollut yhteydessä elämäntapatekijöihin. Lisäksi tämä väitöskirjatutkimus osoitti, että aikuisiän painonnousu lisää LS-kantajamiesten suolistosyöpäriskiä, kun taas lyhyen aikavälin painonnousun huomattiin alentavan LS-kantajanaisten suolistosyöpäriskiä. Korkean intensiteetin fyysinen aktiivisuus aikuisiällä alensi

merkittävästi yleistä syöpäriskiä LS-kantajamiehillä, mutta fyysisen aktiivisuuden ja suolistosyöpäriskin välillä ei havaittu yhteyttä.

Merkittävimmät tulokset osoittivat, että syöpävapaiden LS-kantajien c-miR-profiilin muutokset kuvastavat aikaisen vaiheen suolistosyövän kehittymistä ja ennustavat siihen sairastumista. Näin ollen c-miR-profiilit voivat tarjota potentiaalisen merkkimolekyylijoukon, jonka avulla voitaisiin ohjata korkean riskin potilaita tehokkaammin esimerkiksi kolonoskopiaseurantaan yhdistämällä c-miR-tieto olemassa oleviin seulontatyökaluihin. Lisätutkimusta aiheesta kuitenkin tarvitaan tuloksien vahvistamiseksi. Tämä väitöskirjatutkimus vahvisti myös aiempia havaintoja elämäntapojen ja syöpäriskin yhteyksistä osoittamalla, että terveelliset elämäntavat voivat suojata suomalaisia LS-kantajia syöviltä. Elämäntapojen ja suolistosyöpäriskin välisten molekulaaristen yhteyksien laajempi ymmärtäminen kuitenkin vaatii jatkotutkimuksia.

Tämän yhteenvedon kirjoitushetkellä on käynnissä useita miR-profiileja hyödyntäviä kliinisiä kokeita, mikä viittaa miR-pohjaisen diagnostiikan herättävän kiinnostusta syöpätutkimuksen parissa. Tietokoneiden kasvava laskentateho ja tekoälyn kehitys ennakoivat lupaavaa tulevaisuutta syöpätutkimukselle mahdollistamalla massiivisten molekulaaristen aineistojen tehokkaan hyödyntämisen merkkimolekyylien seulonnassa. Kaiken kaikkiaan tämä väitöstutkimus tuotti merkittävää uutta tietoa, kuinka c-miR-profiileja voidaan mahdollisesti hyödyntää suolistosyöpien hoidoissa, ja miten elämäntavat vaikuttavat syöpäriskiin geneettisesti yhtenäisessä suomalaisessa LS-kohortissa.

REFERENCES

Aaltonen, L., Peltomäki, P., Leach, F. S., Sistonen, P., Pylkkänen, L., Mecklin, J.-P., Järvinen, H., Powell, S. M., Jen, J., Hamilton, S. R., Petersen, G. M., Kinzler, K. W., Vogelstein, B., & de la Chapelle, A. (1993). Clues to the pathogenesis of familial colorectal cancer. *Science*, *260*, 812–816.

Abozaid, Y. J., Zhang, X., Mens, M. M. J., Ahmadizar, F., Limpens, M., Ikram, M. A., Rivadeneira, F., Voortman, T., Kavousi, M., & Ghanbari, M. (2022). Plasma circulating microRNAs associated with obesity, body fat distribution, and fat mass: the Rotterdam Study. *International Journal of Obesity*, *46*(12), 2137–2144. https://doi.org/10.1038/s41366-022-01227-8

Adam, R. S., Poel, D., Ferreira Moreno, L., Spronck, J. M. A., de Back, T. R., Torang, A., Gomez Barila, P. M., ten Hoorn, S., Markowetz, F., Wang, X., Verheul, H. M. W., Buffart, T. E., & Vermeulen, L. (2022). Development of a miRNA-based classifier for detection of colorectal cancer molecular subtypes. *Molecular Oncology*, *16*(14), 2693–2709. https://doi.org/10.1002/1878-0261.13210

Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife,* 4:e05005. https://doi.org/10.7554/eLife.05005.001

Ahadova, A., Gallon, R., Gebert, J., Ballhausen, A., Endris, V., Kirchner, M., Stenzinger, A., Burn, J., von Knebel Doeberitz, M., Bläker, H., & Kloor, M. (2018). Three molecular pathways model colorectal carcinogenesis in Lynch syndrome. *International Journal of Cancer*, *143*(1), 139–150. https://doi.org/10.1002/ijc.31300

Ahadova, A., Seppälä, T. T., Engel, C., Gallon, R., Burn, J., Holinski-Feder, E., Steinke-Lange, V., Möslein, G., Nielsen, M., ten Broeke, S. W., Laghi, L., Dominguez-Valentin, M., Capella, G., Macrae, F., Scott, R., Hüneburg, R., Nattermann, J., Hoffmeister, M., Brenner, H., … Kloor, M. (2021). The "unnatural" history of colorectal cancer in Lynch syndrome: Lessons from colonoscopy surveillance. *International Journal of Cancer,* 148(4), 800–811. https://doi.org/10.1002/ijc.33224

Ahadova, A., von Knebel Doeberitz, M., Bläker, H., & Kloor, M. (2016). CTNNB1-mutant colorectal carcinomas with immediate invasive growth: a model of interval cancers in Lynch syndrome. *Familial Cancer*, *15*(4), 579–586. https://doi.org/10.1007/s10689-016-9899-z

Aoi, W. (2015). Frontier impact of microRNAs in skeletal muscle research: A future perspective. *Frontiers in Physiology*, *5*, 495. https://doi.org/10.3389/fphys.2014.00495

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics, 25(1), 25–29.* https://doi.org/10.1038/75556

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., & Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature*, *455*(7209), 64–71. https://doi.org/10.1038/nature07242

Baggish, A. L., Hale, A., Weiner, R. B., Lewis, G. D., Systrom, D., Wang, F., Wang, T. J., & Chan, S. Y. (2011). Dynamic regulation of circulating microRNA during acute exhaustive exercise and sustained aerobic exercise training. *Journal of Physiology*, *589*(16), 3983–3994. https://doi.org/10.1113/jphysiol.2011.213363

Balaguer, F., Link, A., Lozano, J. J., Cuatrecasas, M., Nagasaka, T., Boland, C. R., & Goel, A. (2010). Epigenetic silencing of miR-137 is an early event in colorectal carcinogenesis. *Cancer Research*, *70*(16), 6609–6618. https://doi.org/10.1158/0008-5472.CAN-10-0622

Balaguer, F., Moreira, L., Lozano, J. J., Link, A., Ramirez, G., Shen, Y., Cuatrecasas, M., Arnold, M., Meltzer, S. J., Syngal, S., Stoffel, E., Jover, R., Llor, X., Castells, A., Boland, C. R., Gironella, M., & Goel, A. (2011). Colorectal cancers with microsatellite instability display unique miRNA profiles. *Clinical Cancer Research*, *17*(19), 6239–6249. https://doi.org/10.1158/1078-0432.CCR-11-1424

Barabási, A. L., Gulbahce, N., & Loscalzo, J. (2011). Network medicine: A network-based approach to human disease. *Nature Reviews Genetics, 12*(1), 56–68. https://doi.org/10.1038/nrg2918

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. http://www.jstor.org/stable/2346101

Bernstein, E., Caudy, A. A., Hammond, S. M., & Hannon, G. J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, *409* (6818), 363–366. https://doi.org/10.1038/35053110

Bohaumilitzky, L., Kluck, K., Hüneburg, R., Gallon, R., Nattermann, J., Kirchner, M., Kristiansen, G., Hommerding, O., Pfuderer, P. L., Wagner, L., Echterdiek, F., Kösegi, S., Müller, N., Fischer, K., Nelius, N., Hartog, B., Borthwick, G., Busch, E., Haag, G. M., … Ahadova, A. (2022). The different immune profiles of normal colonic mucosa in cancer-free Lynch syndrome carriers and Lynch syndrome colorectal cancer patients. *Gastroenterology*, *162*(3), 907-919.e10. https://doi.org/10.1053/j.gastro.2021.11.029

Botma, A., Nagengast, F. M., Braem, M. G. M., Hendriks, J. C. M., Kleibeuker, J. H., Vasen, H. F. A., & Kampman, E. (2010). Body mass index increases risk of colorectal adenomas in men with lynch syndrome: The GEOLynch cohort study. *Journal of Clinical Oncology*, *28*(28), 4346–4353. https://doi.org/10.1200/JCO.2010.28.0453

Bracken, C. P., Khew-Goodall, Y., & Goodall, G. J. (2015). Network-based approaches to understand the roles of miR-200 and other microRNAs in cancer. *Cancer Research, 75*(13), 2594–2599. https://doi.org/10.1158/0008-5472.CAN-15-0287

Bracken, C. P., Scott, H. S., & Goodall, G. J. (2016). A network-biology perspective of microRNA function and dysfunction in cancer. *Nature Reviews Genetics, 17*(12), 719–732. https://doi.org/10.1038/nrg.2016.134

Brenner, H., Stock, C., & Hoffmeister, M. (2014). Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: Systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ*, *348*, g2467. https://doi.org/10.1136/bmj.g2467

Bretthauer, M., Løberg, M., Wieszczy, P., Kalager, M., Emilsson, L., Garborg, K., Rupinski, M., Dekker, E., Spaander, M., Bugajski, M., Holme, Ø., Zauber, A. G., Pilonis, N. D., Mroz, A., Kuipers, E. J., Shi, J., Hernán, M. A., Adami, H.-O., Regula, J., … Kaminski, M. F. (2022). Effect of colonoscopy screening on risks of colorectal cancer and related death. *New England Journal of Medicine*, *387*(17), 1547–1556. https://doi.org/10.1056/nejmoa2208375

Bull, C. J., Bell, J. A., Murphy, N., Sanderson, E., Davey Smith, G., Timpson, N. J., Banbury, B. L., Albanes, D., Berndt, S. I., Bézieau, S., Bishop, D. T., Brenner, H., Buchanan, D. D., Burnett-Hartman, A., Casey, G., Castellví-Bel, S., Chan, A. T., Chang-Claude, J., Cross, A. J., … Gunter, M. J. (2020). Adiposity, metabolites, and colorectal cancer risk: Mendelian randomization study. *BMC Medicine*, *18*(1), 396. https://doi.org/10.1186/s12916-020-01855-9

Burn, J., Bishop, D. T., Chapman, P. D., Elliott, F., Bertario, L., Dunlop, M. G., Eccles, D., Ellis, A., Evans, D. G., Fodde, R., Maher, E. R., Möslein, G., Vasen, H. F. A., Coaker, J., Phillips, R. K. S., Bülow, S., & Mathers, J. C. (2011). A randomized placebo-controlled prevention trial of aspirin and/or resistant starch in young people with familial adenomatous polyposis. *Cancer Prevention Research*, *4*(5), 655–665. https://doi.org/10.1158/1940-6207.CAPR-11-0106

Burn, J., Bishop, D. T., Mecklin, J.-P., Macrae, F., Möslein, G., Olschwang, S., Bisgaard, M.-L., Ramesar, R., Eccles, D., Maher, E. R., Bertario, L., Jarvinen, H. J., Lindblom, A., Gareth, D., Lubinski, J., Morrison, P. J., Ho, J. W. C., Vasen, H. F. A., Side, L., … Mathers, J. C. (2008). Effect of aspirin or resistant starch on colorectal neoplasia in the Lynch syndrome. *New England Journal of Medicine*, *359*(24), 2567–2578. https://doi.org/10.1056/NEJMoa0801297

Burn, J., Sheth, H., Elliott, F., Reed, L., Macrae, F., Mecklin, J. P., Möslein, G., McRonald, F. E., Bertario, L., Evans, D. G., Gerdes, A. M., Ho, J. W. C., Lindblom, A., Morrison, P. J., Rashbass, J., Ramesar, R., Seppälä, T., Thomas, H. J. W., Pylvänäinen, K., … Side, L. (2020). Cancer prevention with aspirin in hereditary colorectal cancer (Lynch syndrome), 10-year follow-up and registry-based 20-year data in the CAPP2 study: A double-blind, randomised, placebo-controlled trial. *The Lancet*, *395*(10240), 1855–1863. https://doi.org/10.1016/S0140-6736(20)30366-4

Bye, A., Røsjø, H., Aspenes, S. T., Condorelli, G., Omland, T., & Wisløff, U. (2013). Circulating microRNAs and aerobic Fitness - The HUNT-study. *PLoS ONE*, *8*(2), e57496. https://doi.org/10.1371/journal.pone.0057496

Cai, X., Hagedorn, C. H., & Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, *10*(12), 1957–1966. https://doi.org/10.1261/rna.7135204

Calin, G. A., & Croce, C. M. (2006). MicroRNA signatures in human cancers. *Nature Reviews Cancer, 6*(11), 857–866. https://doi.org/10.1038/nrc1997

Campbell, P. T., Cotterchio, M., Dicks, E., Parfrey, P., Gallinger, S., & McLaughlin, J. R. (2007). Excess body weight and colorectal cancer risk in Canada: Associations in subgroups of clinically defined familial risk of cancer. *Cancer Epidemiology Biomarkers and Prevention*, *16*(9), 1735–1744. https://doi.org/10.1158/1055-9965.EPI-06-1059

Campbell, P. T., Newcomb, P., Gallinger, S., Cotterchio, M., & McLaughlin, J. R. (2007). Exogenous hormones and colorectal cancer risk in Canada: Associations stratified by clinically defined familial risk of cancer. *Cancer Causes and Control*, *18*(7), 723–733. https://doi.org/10.1007/s10552-007-9015-7

Campisi, J. (2013). Aging, cellular senescence, and cancer. *Annual Review of Physiology 75*, 685–705. https://doi.org/10.1146/annurev-physiol-030212-183653

Cercek, A., Lumish, M., Sinopoli, J., Weiss, J., Shia, J., Lamendola-Essel, M., El Dika, I. H., Segal, N., Shcherba, M., Sugarman, R., Stadler, Z., Yaeger, R., Smith, J. J., Rousseau, B., Argiles, G., Patel, M., Desai, A., Saltz, L. B., Widmar, M., … Diaz, L. A. (2022). PD-1 Blockade in mismatch repair–deficient, locally advanced rectal cancer. *New England Journal of Medicine*, *386*(25), 2363–2376. https://doi.org/10.1056/nejmoa2201445

Chadid, S., Singer, M. R., Kreger, B. E., Bradlee, M. L., & Moore, L. L. (2018). Midlife weight gain is a risk factor for obesity-related cancer. *British Journal of Cancer*, *118*(12), 1665–1671. https://doi.org/10.1038/s41416-018-0106-x

Chatterjee, A., Leichter, A. L., Fan, V., Tsai, P., Purcell, R. V., Sullivan, M. J., & Eccles, M. R. (2015). A cross comparison of technologies for the detection of microRNAs in clinical FFPE samples of hepatoblastoma patients. *Scientific Reports*, *5*, 10438. https://doi.org/10.1038/srep10438

Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., Guo, J., Zhang, Y., Chen, J., Guo, X., Li, Q., Li, X., Wang, W., Zhang, Y., Wang, J., Jiang, X., Xiang, Y., Xu, C., Zheng, P., … Zhang, C. Y. (2008). Characterization of microRNAs in serum: A novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Research*, *18*(10), 997–1006. https://doi.org/10.1038/cr.2008.282

Clinton, S. K., Giovannucci, E. L., & Hursting, S. D. (2020). The World Cancer Research Fund/American Institute for Cancer Research Third Expert Report on Diet, Nutrition, Physical Activity, and Cancer: Impact and Future Directions. *Journal of Nutrition, 150*(4), 663–671. https://doi.org/10.1093/jn/nxz268

Coletta, A. M., Peterson, S. K., Gatus, L. A., Krause, K. J., Schembre, S. M., Gilchrist, S. C., Pande, M., Vilar, E., You, Y. N., Rodriguez-Bigas, M. A., Strong, L. L., Lynch, P. M., Lu, K. H., & Basen-Engquist, K. (2019). Energy balance related lifestyle factors and risk of endometrial and colorectal cancer among individuals with lynch syndrome: A systematic review. *Familial Cancer*, *18*(4), 399–420. https://doi.org/10.1007/s10689-019-00135-7

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *BMJ*, *350*, g7594. https://doi.org/10.1136/bmj.g7594

Cui, W., Zhang, S., Shan, C., Zhou, L., & Zhou, Z. (2013). MicroRNA-133a regulates the cell cycle and proliferation of breast cancer cells by targeting epidermal growth factor receptor through the EGFR/Akt signaling pathway. *FEBS Journal*, *280*(16), 3962–3974. https://doi.org/10.1111/febs.12398

Dahl, A. K., & Reynolds, C. A. (2013). Accuracy of recalled body weight - A study with 20-years of follow-up. *Obesity*, *21*(6), 1293–1298. https://doi.org/10.1002/oby.20299

Dashti, S. G., Buchanan, D. D., Jayasekara, H., Ouakrim, D. A., Clendenning, M., Rosty, C., Winship, I. M., MacRae, F. A., Giles, G. G., Parry, S., Casey, G., Haile, R. W., Gallinger, S., Le Marchand, L., Thibodeau, S. N., Lindor, N. M., Newcomb, P. A., Potter, J. D., Baron, J. A., … Win, A. K. (2017). Alcohol consumption and the risk of colorectal cancer for mismatch repair gene mutation carriers. *Cancer Epidemiology Biomarkers and Prevention*, *26*(3), 366–375. https://doi.org/10.1158/1055-9965.EPI-16-0496

Dashti, S. G., Win, A. K., Hardikar, S. S., Glombicki, S. E., Mallenahalli, S., Thirumurthi, S., Peterson, S. K., You, Y. N., Buchanan, D. D., Figueiredo, J. C., Campbell, P. T., Gallinger, S., Newcomb, P. A., Potter, J. D., Lindor, N. M., Le Marchand, L., Haile, R. W., Hopper, J. L., Jenkins, M. A., … Pande, M. (2018). Physical activity and the risk of colorectal cancer in Lynch syndrome. *International Journal of Cancer*, *143*(9), 2250–2260. https://doi.org/10.1002/ijc.31611

Davies, A., Wellard-Cole, L., Rangan, A., & Allman-Farinelli, M. (2020). Validity of self-reported weight and height for BMI classification: A cross-sectional study among young adults. *Nutrition*, *71*, 110622. https://doi.org/10.1016/j.nut.2019.110622

de Rezende, L. F. M., de Sá, T. H., Markozannes, G., Rey-López, J. P., Lee, I. M., Tsilidis, K. K., Ioannidis, J. P. A., & Eluf-Neto, J. (2018). Physical activity and cancer: An umbrella review of the literature including 22 major anatomical sites and 770 000 cancer cases. *British Journal of Sports Medicine*, *52*(13), 826–833. https://doi.org/10.1136/bjsports-2017-098391

Deiuliis, J. A. (2016). MicroRNAs as regulators of metabolic disease: Pathophysiologic significance and emerging role as biomarkers and therapeutics. *International Journal of Obesity*, *40*(1), 88–101. https://doi.org/10.1038/ijo.2015.170

Dekker, E., Tanis, P. J., Vleugels, J. L. A., Kasi, P. M., & Wallace, M. B. (2019). Colorectal cancer. *The Lancet*, *394*(10207), 1467–1480. https://doi.org/10.1016/S0140-6736(19)32319-0

Deng, N., Reyes-Uribe, L., Fahrmann, J. F., Thoman, W. S., Munsell, M. F., Dennison, J. B., Murage, E., Wu, R., Hawk, E. T., Thirumurthi, S., Lynch, P. M., Dieli-Conwright, C. M., Lazar, A. J., Jindal, S., Chu, K., Chelvanambi, M., Basen-Engquist, K., Li, Y., Wargo, J. A., … Vilar, E. (2023). Exercise training reduces the inflammatory response and promotes intestinal mucosa-associated immunity in Lynch syndrome. *Clinical Cancer Research*, *29*(21), 4361–4372. https://doi.org/10.1158/1078-0432.CCR-23-0088

Dhawan, A., Scott, J. G., Harris, A. L., & Buffa, F. M. (2018). Pan-cancer characterisation of microRNA across cancer hallmarks reveals microRNA-mediated downregulation of tumour suppressors. *Nature Communications*, *9*(1), 5228. https://doi.org/10.1038/s41467-018-07657-1

Di Cesare, M., Bentham, J., Stevens, G. A., Zhou, B., Danaei, G., Lu, Y., Bixby, H., Cowan, M. J., Riley, L. M., Hajifathalian, K., Fortunato, L., Taddei, C., Bennett, J. E., Ikeda, N., Khang, Y. H., Kyobutungi, C., Laxmaiah, A., Li, Y., Lin, H. H., … Cisneros, J. Z. (2016). Trends in adult body-mass index in 200 countries from 1975 to 2014: A pooled analysis of 1698 population-based measurement studies with 19.2 million participants. *The Lancet*, *387*(10026), 1377–1396. https://doi.org/10.1016/S0140-6736(16)30054-X

DiPietro, L., Buchner, D. M., Marquez, D. X., Pate, R. R., Pescatello, L. S., & Whitt-Glover, M. C. (2019). New scientific basis for the 2018 U.S. Physical Activity Guidelines. *Journal of Sport and Health Science*, *8*(3), 197–200. https://doi.org/10.1016/j.jshs.2019.03.007

Dixon, J. B. (2010). The effect of obesity on health outcomes. *Molecular and Cellular Endocrinology*, *316*(2), 104–108. https://doi.org/10.1016/j.mce.2009.07.008

Doench, J. G., & Sharp, P. A. (2004). Specificity of microRNA target selection in translational repression. *Genes and Development*, *18*(5), 504–511. https://doi.org/10.1101/gad.1184404

Dominguez-Valentin, M., Haupt, S., Seppälä, T. T., Sampson, J. R., Sunde, L., Bernstein, I., Jenkins, M. A., Engel, C., Aretz, S., Nielsen, M., Capella, G., Balaguer, F., Evans, D. G., Burn, J., Holinski-Feder, E., Bertario, L., Bonanni, B., Lindblom, A., Levi, Z., … Moller, P. (2023). Mortality by age, gene and gender in carriers of pathogenic mismatch repair gene variants receiving surveillance for early cancer diagnosis and treatment: A report from the prospective Lynch syndrome database. *EClinicalMedicine*, *58*, 101909. https://doi.org/10.1016/j.eclinm.2023.101909

Dominguez-Valentin, M., Sampson, J. R., Seppälä, T. T., Broeke, S. W., Plazzer, J.-P., Nakken, S., Engel, C., Aretz, S., Jenkins, M. A., Sunde, L., Bernstein, I., Capella, G., Balaguer, F., Thomas, H., Gareth Evans, D., Burn, J., Greenblatt, M., Hovig, E., Vos tot Nederveen Cappel, W. H., … Møller, P. (2020). Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome

Database. *Genetics in Medicine*, *22*(1), 15–25. https://doi.org/10.1038/s41436-019-0596-9

Dominguez-Valentin, M., Seppälä, T. T., Sampson, J. R., MacRae, F., Winship, I., Evans, D. G., Scott, R. J., Burn, J., Möslein, G., Bernstein, I., Pylvänäinen, K., Renkonen-Sinisalo, L., Lepistö, A., Lindblom, A., Plazzer, J. P., Tjandra, D., Thomas, H., Green, K., Lalloo, F., … Møller, P. (2019). Survival by colon cancer stage and screening interval in Lynch syndrome: A prospective Lynch syndrome database report. *Hereditary Cancer in Clinical Practice*, *17*, 28. https://doi.org/10.1186/s13053-019-0127-3

Dong, Y., Zhao, J., Wu, C. W., Zhang, L., Liu, X., Kang, W., Leung, W. W., Zhang, N., Chan, F. K. L., Sung, J. J. Y., Ng, S. S. M., & Yu, J. (2013). Tumor suppressor functions of miR-133a in colorectal cancer. *Molecular Cancer Research*, *11*(9), 1051–1060. https://doi.org/10.1158/1541-7786.MCR-13-0061

D'Onofrio, N., Prattichizzo, F., Martino, E., Anastasio, C., Mele, L., La Grotta, R., Sardu, C., Ceriello, A., Marfella, R., Paolisso, G., & Balestrieri, M. L. (2023). MiR-27b attenuates mitochondrial oxidative stress and inflammation in endothelial cells. *Redox Biology*, *62*, 102681. https://doi.org/10.1016/j.redox.2023.102681

Doubeni, C. A., Corley, D. A., Quinn, V. P., Jensen, C. D., Zauber, A. G., Goodman, M., Johnson, J. R., Mehta, S. J., Becerra, T. A., Zhao, W. K., Schottinger, J., Doria-Rose, V. P., Levin, T. R., Weiss, N. S., & Fletcher, R. H. (2018). Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: A large community-based study. *Gut*, *67*(2), 291–298. https://doi.org/10.1136/gutjnl-2016-312712

Dufresne, S., Rébillard, A., Muti, P., Friedenreich, C. M., & Brenner, D. R. (2018). A review of physical activity and circulating miRNA expression: Implications in cancer risk and progression. *Cancer Epidemiology Biomarkers and Prevention*, *27*(1), 11–24. https://doi.org/10.1158/1055-9965.EPI-16-0969

Duval, A., & Hamelin, R. (2002). Mutations at coding repeat sequences in mismatch repair-deficient human cancers: Toward a new concept of target genes for instability. *Cancer Research*, *62*(9), 2447–2454.

Earle, J. S. L., Luthra, R., Romans, A., Abraham, R., Ensor, J., Yao, H., & Hamilton, S. R. (2010). Association of MicroRNA expression with microsatellite instability status in colorectal adenocarcinoma. *Journal of Molecular Diagnostics*, *12*(4), 433–440. https://doi.org/10.2353/jmoldx.2010.090154

Eichhorn, S. W., Guo, H., McGeary, S. E., Rodriguez-Mias, R. A., Shin, C., Baek, D., Hsu, S. hao, Ghoshal, K., Villén, J., & Bartel, D. P. (2014). MRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular Cell*, *56*(1), 104–115. https://doi.org/10.1016/j.molcel.2014.08.028

Engel, C., Ahadova, A., Seppälä, T. T., Aretz, S., Bigirwamungu-Bargeman, M., Bläker, H., Bucksch, K., Büttner, R., de Vos tot Nederveen Cappel, W. T.,

Endris, V., Holinski-Feder, E., Holzapfel, S., Hüneburg, R., Jacobs, M. A. J. M., Koornstra, J. J., Langers, A. M., Lepistö, A., Morak, M., Möslein, G., … Vasen, H. F. (2020). Associations of pathogenic variants in MLH1, MSH2, and MSH6 with risk of colorectal adenomas and tumors and with somatic mutations in patients with Lynch syndrome. *Gastroenterology*, *158*(5), 1326–1333. https://doi.org/10.1053/j.gastro.2019.12.032

Engel, C., Vasen, H. F., Seppälä, T., Aretz, S., Bigirwamungu-Bargeman, M., de Boer, S. Y., Bucksch, K., Büttner, R., Holinski-Feder, E., Holzapfel, S., Hüneburg, R., Jacobs, M. A. J. M., Järvinen, H., Kloor, M., von Knebel Doeberitz, M., Koornstra, J. J., van Kouwen, M., Langers, A. M., van de Meeberg, P. C., … Loeffler, M. (2018). No difference in colorectal cancer incidence or stage at detection by colonoscopy among 3 countries with different Lynch syndrome surveillance policies. *Gastroenterology*, *155*(5), 1400-1409.e2. https://doi.org/10.1053/j.gastro.2018.07.030

Esquela-Kerscher, A., & Slack, F. J. (2006). Oncomirs - MicroRNAs with a role in cancer. *Nature Reviews Cancer, 6*(4), 259–269. https://doi.org/10.1038/nrc1840

Esteller, M. (2011). Non-coding RNAs in human disease. *Nature Reviews Genetics*, *12*(12), 861–874. https://doi.org/10.1038/nrg3074

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., … D'Eustachio, P. (2016). The reactome pathway knowledgebase. *Nucleic Acids Research*, *44*(D1), D481–D487. https://doi.org/10.1093/nar/gkv1351

Fehlmann, T., Ludwig, N., Backes, C., Meese, E., & Keller, A. (2016). Distribution of microRNA biomarker candidates in solid tissues and body fluids. *RNA Biology, 13*(11), 1084–1088. https://doi.org/10.1080/15476286.2016.1234658

Fitzgibbons, R. J., Lynch, H. T., Stanislav, G. V, Watson, P. A., Lanspa, S. J., Marcus, J. N., Smyrk, T., Kriegler, M. D., & Lynch, J. F. (1987). Recognition and treatment of patients with hereditary nonpolyposis colon cancer (Lynch syndromes I and II). *Annals of Surgery*, *206*(3), 289–295.

Fleming, N. I., Jorissen, R. N., Mouradov, D., Christie, M., Sakthianandeswaren, A., Palmieri, M., Day, F., Li, S., Tsui, C., Lipton, L., Desai, J., Jones, I. T., McLaughlin, S., Ward, R. L., Hawkins, N. J., Ruszkiewicz, A. R., Moore, J., Zhu, H. J., Mariadason, J. M., … Sieber, O. M. (2013). SMAD2, SMAD3 and SMAD4 mutations in colorectal cancer. *Cancer Research*, *73*(2), 725–735. https://doi.org/10.1158/0008-5472.CAN-12-2706

Francavilla, A., Turoczi, S., Tarallo, S., Vodicka, P., Pardini, B., & Naccarati, A. (2020). Exosomal microRNAs and other non-coding RNAs as colorectal cancer biomarkers: A review. *Mutagenesis, 35*(3), 243–260. https://doi.org/10.1093/mutage/gez038

Friedenreich, C. M., Neilson, H. K., & Lynch, B. M. (2010). State of the epidemiological evidence on physical activity and cancer prevention.

European Journal of Cancer, *46*(14), 2593–2604. https://doi.org/10.1016/j.ejca.2010.07.028

Friedenreich, C. M., Ryder-Burbidge, C., & McNeil, J. (2021). Physical activity, obesity and sedentary behavior in cancer etiology: Epidemiologic evidence and biologic mechanisms. *Molecular Oncology, 15*(3), 790–800. https://doi.org/10.1002/1878-0261.12772

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *JSS Journal of Statistical Software, 33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Garai, K., Adam, Z., Herczeg, R., Banfai, K., Gyebrovszki, A., Gyenesei, A., Pongracz, J. E., Wilhelm, M., & Kvell, K. (2021). Physical activity as a preventive lifestyle intervention acts through specific exosomal miRNA Species—Evidence from human short- and long-term pilot studies. *Frontiers in Physiology*, *12,* 658218. https://doi.org/10.3389/fphys.2021.658218

Geer, E. B., & Shen, W. (2009). Gender differences in insulin resistance, body composition, and energy balance. *Gender Medicine*, *6*, 60–75. https://doi.org/10.1016/j.genm.2009.02.002

Gerstung, M., Jolly, C., Leshchiner, I., Dentro, S. C., Gonzalez, S., Rosebrock, D., Mitchell, T. J., Rubanova, Y., Anur, P., Yu, K., Tarabichi, M., Deshwar, A., Wintersinger, J., Kleinheinz, K., Vázquez-García, I., Haase, K., Jerman, L., Sengupta, S., Macintyre, G., … Van Loo, P. (2020). The evolutionary history of 2,658 cancers. *Nature*, *578*(7793), 122–128. https://doi.org/10.1038/s41586-019-1907-7

Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., & Caldas, C. (2010). Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA*, *16*(5), 991–1006. https://doi.org/10.1261/rna.1947110

Gits, C. M. M., Van Kuijk, P. F., Jonkers, M. B. E., Boersma, A. W. M., Van Ijcken, W. F., Wozniak, A., Sciot, R., Rutkowski, P., Schöffski, P., Taguchi, T., Mathijssen, R. H. J., Verweij, J., Sleijfer, S., Debiec-Rychter, M., & Wiemer, E. A. C. (2013). MiR-17-92 and miR-221/222 cluster members target KIT and ETV1 in human gastrointestinal stromal tumours. *British Journal of Cancer*, *109*(6), 1625–1635. https://doi.org/10.1038/bjc.2013.483

Glaviano, A., Foo, A. S. C., Lam, H. Y., Yap, K. C. H., Jacot, W., Jones, R. H., Eng, H., Nair, M. G., Makvandi, P., Geoerger, B., Kulke, M. H., Baird, R. D., Prabhu, J. S., Carbone, D., Pecoraro, C., Teh, D. B. L., Sethi, G., Cavalieri, V., Lin, K. H., … Kumar, A. P. (2023). PI3K/AKT/mTOR signaling transduction pathway and targeted therapies in cancer. *Molecular Cancer, 22*(1), 138. https://doi.org/10.1186/s12943-023-01827-6

Goodall, G. J., & Wickramasinghe, V. O. (2021). RNA in cancer. *Nature Reviews Cancer, 21*(1), 22–36. https://doi.org/10.1038/s41568-020-00306-0

Greenivald, P., & Dunn, B. K. (2009). Landmarks in the history of cancer epidemiology. *Cancer Research*, *69*(6), 2151–2162. https://doi.org/10.1158/0008-5472.CAN-09-0416
Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., & Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research*, *34*, D140–D144. https://doi.org/10.1093/nar/gkj112
Guo, H., Ingolia, N. T., Weissman, J. S., & Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, *466*(7308), 835–840. https://doi.org/10.1038/nature09267
Ha, M., & Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology, 15*(8), 509–524. https://doi.org/10.1038/nrm3838
Hanahan, D. (2022). Hallmarks of cancer: New dimensions. *Cancer Discovery, 12* (1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059
Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*, 57–70. https://doi.org/10.1016/s0092-8674(00)81683-9
Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell, 144*(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013
Haraldsdottir, S., Rafnar, T., Frankel, W. L., Einarsdottir, S., Sigurdsson, A., Hampel, H., Snaebjornsson, P., Masson, G., Weng, D., Arngrimsson, R., Kehr, B., Yilmaz, A., Haraldsson, S., Sulem, P., Stefansson, T., Shields, P. G., Sigurdsson, F., Bekaii-Saab, T., Moller, P. H., … Stefansson, K. (2017). Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nature Communications*, *8,* 14755. https://doi.org/10.1038/ncomms14755
Harrell, F. E. J., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *JAMA*, *247*(18), 2543–2546.
Harris, S. L., & Levine, A. J. (2005). The p53 pathway: Positive and negative feedback loops. *Oncogene, 24*(17), 2899–2908. https://doi.org/10.1038/sj.onc.1208615
He, L., & Hannon, G. J. (2004). MicroRNAs: Small RNAs with a big role in gene regulation. *Nature Reviews Genetics, 5*(7), 522–531. https://doi.org/10.1038/nrg1379
He, L., He, X., Lim, L. P., De Stanchina, E., Xuan, Z., Liang, Y., Xue, W., Zender, L., Magnus, J., Ridzon, D., Jackson, A. L., Linsley, P. S., Chen, C., Lowe, S. W., Cleary, M. A., & Hannon, G. J. (2007). A microRNA component of the p53 tumour suppressor network. *Nature*, *447*(7148), 1130–1134. https://doi.org/10.1038/nature05939
Heyn, G. S., Corrêa, L. H., & Magalhães, K. G. (2020). The impact of adipose tissue–derived miRNAs in metabolic syndrome, obesity, and cancer. *Frontiers in Endocrinology, 11,* 563816. https://doi.org/10.3389/fendo.2020.563816
Hoedjes, M., Vrieling, A., de Brauwer, L., Visser, A., Gómez García, E., Hoogerbrugge, N., & Kampman, E. (2023). Determinants of adherence to recommendations on physical activity, red and processed meat intake, and

body weight among lynch syndrome patients. *Familial Cancer*, *22*(2), 155–166. https://doi.org/10.1007/s10689-022-00315-y

Høye, E., Fromm, B., Böttger, P. H. M., Domanska, D., Torgunrud, A., Lund-Andersen, C., Abrahamsen, T. W., Fretland, Å. A., Dagenborg, V. J., Lorenz, S., Edwin, B., Hovig, E., & Flatmark, K. (2022). A comprehensive framework for analysis of microRNA sequencing data in metastatic colorectal cancer. *NAR Cancer*, *4*(1), zcab051. https://doi.org/10.1093/narcan/zcab051

Hsu, S. Da, Lin, F. M., Wu, W. Y., Liang, C., Huang, W. C., Chan, W. L., Tsai, W. T., Chen, G. Z., Lee, C. J., Chiu, C. M., Chien, C. H., Wu, M. C., Huang, C. Y., Tsou, A. P., & Huang, H. Da. (2011). MiRTarBase: A database curates experimentally validated microRNA-target interactions. *Nucleic Acids Research*, *39*, D163–D169. https://doi.org/10.1093/nar/gkq1107

Hull, M. A., Rees, C. J., Sharp, L., & Koo, S. (2020). A risk-stratified approach to colorectal cancer prevention and diagnosis. *Nature Reviews Gastroenterology and Hepatology*, *17*(12), 773–780. https://doi.org/10.1038/s41575-020-00368-3

Huntzinger, E., & Izaurralde, E. (2011). Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. *Nature Reviews Genetics*, *12*(2), 99–110. https://doi.org/10.1038/nrg2936

Hur, K., Toiyama, Y., Schetter, A. J., Okugawa, Y., Harris, C. C., Boland, C. R., & Goel, A. (2015). Identification of a metastasis-specific microRNA signature in human colorectal cancer. *Journal of the National Cancer Institute*, *107*(3), dju492. https://doi.org/10.1093/jnci/dju492

Iacomino, G., & Siani, A. (2017). Role of microRNAs in obesity and obesity-related diseases. *Genes and Nutrition*, *12*(1), 23. https://doi.org/10.1186/s12263-017-0577-z

Ignatiadis, M., Sledge, G. W., & Jeffrey, S. S. (2021). Liquid biopsy enters the clinic — implementation issues and future challenges. *Nature Reviews Clinical Oncology*, *18*(5), 297–312. https://doi.org/10.1038/s41571-020-00457-x

Jahid, S., Sun, J., Edwards, R. A., Dizon, D., Panarelli, N. C., Milsom, J. W., Sikandar, S. S., Gümüş, Z. H., & Lipkin, S. M. (2012). miR-23a promotes the transition from indolent to invasive colorectal cancer. *Cancer Discovery*, *2*(6), 540–553. https://doi.org/10.1158/2159-8290.CD-11-0267

Järvinen, H. J., Aarnio, M., Mustonen, H., Aktan-Collan, K., Aaltonen, L. A., Peltomäki, P., De la Chapelle, A., & Mecklin, J.-P. (2000). Controlled 15-year trial on screening for colorectal cancer in families with hereditary nonpolyposis colorectal cancer. *Gastroenterology*, *118*(5), 829–834. https://doi.org/10.1016/s0016-5085(00)70168-5

Jeggo, P. A., Pearl, L. H., & Carr, A. M. (2016). DNA repair, genome stability and cancer: A historical perspective. *Nature Reviews Cancer*, *16*(1), 35–42. https://doi.org/10.1038/nrc.2015.4

Jokela, T. A., Karppinen, J. E., Kärkkäinen, M., Mecklin, J.-P., Walker, S., Seppälä, T. T., & Laakkonen, E. K. (2024). Circulating metabolome

landscape in Lynch syndrome. *Cancer & Metabolism*, *12*(1), 4. https://doi.org/10.1186/s40170-024-00331-9

Joshi-Tope, G., Gillespie, M., Vastrik, I., D'Eustachio, P., Schmidt, E., de Bono, B., Jassal, B., Gopinath, G. R., Wu, G. R., Matthews, L., Lewis, S., Birney, E., & Stein, L. (2005). Reactome: A knowledgebase of biological pathways. *Nucleic Acids Research*, *33*, D428–D432. https://doi.org/10.1093/nar/gki072

Jung, G., Hernández-Illán, E., Moreira, L., Balaguer, F., & Goel, A. (2020). Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nature Reviews Gastroenterology and Hepatology*, *17*(2), 111–130. https://doi.org/10.1038/s41575-019-0230-y

Juppi, H. K., Sipilä, S., Cronin, N. J., Karvinen, S., Karppinen, J. E., Tammelin, T. H., Aukee, P., Kovanen, V., Kujala, U. M., & Laakkonen, E. K. (2020). Role of menopausal transition and physical activity in loss of lean and muscle mass: A follow-up study in middle-aged Finnish women. *Journal of Clinical Medicine*, *9*(5). https://doi.org/10.3390/jcm9051588

Kamiza, A. B., Hsieh, L. L., Tang, R., Chien, H. T., Lai, C. H., Chiu, L. L., Lo, T. P., Hung, K. Y., Wang, C. Y., You, J. F., Hsiung, C. A., & Yeh, C. C. (2015). Risk factors associated with colorectal cancer in a subset of patients with mutations in MLH1 and MSH2 in Taiwan fulfilling the Amsterdam II criteria for Lynch syndrome. *PLoS ONE*, *10*(6), e0130018. https://doi.org/10.1371/journal.pone.0130018

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kansikas, M., Kariola, R., & Nyström, M. (2011). Verification of the three-step model in assessing the pathogenicity of mismatch repair gene variants. *Human Mutation*, *32*(1), 107–115. https://doi.org/10.1002/humu.21409

Karolina, D. S., Tavintharan, S., Armugam, A., Sepramaniam, S., Pek, S. L. T., Wong, M. T. K., Lim, S. C., Sum, C. F., & Jeyaseelan, K. (2012). Circulating miRNA profiles in patients with metabolic syndrome. *Journal of Clinical Endocrinology and Metabolism*, *97*(12), E2271–E2276. https://doi.org/10.1210/jc.2012-1996

Karvinen, S., Korhonen, T. M., Sievänen, T., Karppinen, J. E., Juppi, H. K., Jakoaho, V., Kujala, U. M., Laukkanen, J. A., Lehti, M., & Laakkonen, E. K. (2023). Extracellular vesicles and high-density lipoproteins: Exercise and oestrogen-responsive small RNA carriers. *Journal of Extracellular Vesicles*, *12*(2), e12308. https://doi.org/10.1002/jev2.12308

Key, T. J., Bradbury, K. E., Perez-Cornago, A., Sinha, R., Tsilidis, K. K., & Tsugane, S. (2020). Diet, nutrition, and cancer risk: What do we know and what is the way forward? *BMJ*, *368*, m511. https://doi.org/10.1136/bmj.m511

Kim, T., Veronese, A., Pichiorri, F., Lee, T. J., Jeon, Y. J., Volinia, S., Pineau, P., Marchio, A., Palatini, J., Suh, S. S., Alder, H., Liu, C. G., Dejean, A., & Croce, C. M. (2011). p53 regulates epithelial-mesenchymal transition

through microRNAs targeting ZEB1 and ZEB2. *Journal of Experimental Medicine*, *208*(5), 875–883. https://doi.org/10.1084/jem.20110235

Kirchengast, S., Gruber, D., Sator, M., Hartmann, B., Knogler, W., & Huber, J. (1997). Menopause-associated differences in female fat patterning estimated by dual-energy X-ray absorptiometry. *Annals of human biology*, *24*(1), 45–54. https://doi.org/10.1080/03014469700004762

Kitahara, C. M., Berndt, S. I., Berrington de González, A., Coleman, H. G., Schoen, R. E., Hayes, R. B., & Huang, W. Y. (2013). Prospective investigation of body mass index, colorectal adenoma, and colorectal cancer in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *Journal of Clinical Oncology*, *31*(19), 2450–2459. https://doi.org/10.1200/JCO.2012.48.4691

Kloor, M., & Von Knebel Doeberitz, M. (2016). The immune biology of microsatellite-unstable cancer. *Trends in Cancer, 2*(3), 121–133. https://doi.org/10.1016/j.trecan.2016.02.004

Knapp, S. T., Revette, A., Underhill-Blazey, M., Stopfer, J. E., Ukaegbu, C. I., Poulin, C., Parenteau, M., Syngal, S., Bae, E., Bickmore, T., Hampel, H., Idos, G. E., Parmigiani, G., Yurgelun, M. B., & Braun, D. (2023). MyLynch: A patient-facing clinical decision support tool for genetically-guided personalized medicine in Lynch syndrome. *Cancers*, *15*(2), 391. https://doi.org/10.3390/cancers15020391

Knudson, A. G. (1971). Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America, 68*(4), 820–823. https://doi.org/10.1073/pnas.68.4.820

Kyrgiou, M., Kalliala, I., Markozannes, G., Gunter, M. J., Paraskevaidis, E., Gabra, H., Martin-Hirsch, P., & Tsilidis, K. K. (2017). Adiposity and cancer at major anatomical sites: Umbrella review of the literature. *BMJ, 356,* j477. https://doi.org/10.1136/bmj.j477

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., & Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. *Current Biology, 12*(9), 735–739. https://doi.org/10.1016/s0960-9822(02)00809-6

Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*(3), R25. https://doi.org/10.1186/gb-2009-10-3-r25

Lanza, G., Ferracin, M., Gafà, R., Veronese, A., Spizzo, R., Pichiorri, F., Liu, C. G., Calin, G. A., Croce, C. M., & Negrini, M. (2007). mRNA/microRNA gene expression profile in microsatellite unstable colorectal cancer. *Molecular Cancer*, *6,* 54. https://doi.org/10.1186/1476-4598-6-54

Lazzeroni, M., Bellerba, F., Calvello, M., Macrae, F., Win, A. K., Jenkins, M., Serrano, D., Marabelli, M., Cagnacci, S., Tolva, G., Macis, D., Raimondi, S., Mazzarella, L., Chiocca, S., Caini, S., Bertario, L., Bonanni, B., & Gandini, S. (2021). A meta-analysis of obesity and risk of colorectal cancer in patients with lynch syndrome: The impact of sex and genetics. *Nutrients*, *13*(5), 1736. https://doi.org/10.3390/nu13051736

Lee, I. M., Shiroma, E. J., Lobelo, F., Puska, P., Blair, S. N., Katzmarzyk, P. T., Alkandari, J. R., Andersen, L. B., Bauman, A. E., Brownson, R. C., Bull, F. C., Craig, C. L., Ekelund, U., Goenka, S., Guthold, R., Hallal, P. C., Haskell, W. L., Heath, G. W., Inoue, S., … Wells, J. C. (2012). Effect of physical inactivity on major non-communicable diseases worldwide: An analysis of burden of disease and life expectancy. *The Lancet*, *380*(9838), 219–229. https://doi.org/10.1016/S0140-6736(12)61031-9

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Rå Dmark, O., Kim, S., & Kim, V. N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature*, *425*(6956), 415–419. https://doi.org/10.1038/nature01957

Liccardo, R., Sessa, R., Trombetti, S., Rosa, M. De, Izzo, P., Grosso, M., & Duraturo, F. (2021). Mir-137 targets the 3′ untranslated region of msh2: Potential implications in lynch syndrome-related colorectal cancer. *Cancers*, *13*(18), 4662. https://doi.org/10.3390/cancers13184662

Louafi, F., Martinez-Nunez, R. T., & Sanchez-Elsner, T. (2010). MicroRNA-155 targets SMAD2 and modulates the response of macrophages to transforming growth factor-β. *Journal of Biological Chemistry*, *285*(53), 41328–41336. https://doi.org/10.1074/jbc.M110.146852

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., Rheinheimer, S., Meder, B., Stähler, C., Meese, E., & Keller, A. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Research*, *44*(8), 3865–3877. https://doi.org/10.1093/nar/gkw116

Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E., & Kutay, U. (2004). Nuclear export of microRNA precursors. *Science*, *303*(5654), 95–98. https://doi.org/10.1126/science.1090599

Lynch, H. T., Snyder, C. L., Shaw, T. G., Heinen, C. D., & Hitchins, M. P. (2015). Milestones of Lynch syndrome: 1895-2015. *Nature Reviews Cancer*, *15*(3), 181–194. https://doi.org/10.1038/nrc3878

Ma, R., Jiang, T., & Kang, X. (2012). Circulating microRNAs in cancer: Origin, function and application. *Journal of Experimental & Clinical Cancer Research*, *31*(1), 38. https://doi.org/10.1186/1756-9966-31-38

Marcuello, M., Duran-Sanchon, S., Moreno, L., Lozano, J. J., Bujanda, L., Castells, A., & Gironella, M. (2019). Analysis of a 6-mirna signature in serum from colorectal cancer screening participants as non-invasive biomarkers for advanced adenoma and colorectal cancer detection. *Cancers*, *11*(10), 1542. https://doi.org/10.3390/cancers11101542

Mathers, J. C., Elliott, F., Macrae, F., Mecklin, J. P., Möslein, G., McRonald, F. E., Bertario, L., Evans, D. G., Gerdes, A. M., Ho, J. W. C., Lindblom, A., Morrison, P. J., Rashbass, J., Ramesar, R. S., Seppälä, T. T., Thomas, H. J. W., Sheth, H. J., Pylvänäinen, K., Reed, L., … Bacon, A. (2022). Cancer prevention with resistant starch in Lynch syndrome patients in the CAPP2-

randomized placebo controlled trial: Planned 10-year follow-up. *Cancer Prevention Research*, *15*(9), 623–634. https://doi.org/10.1158/1940-6207.CAPR-22-0044

Mathers, J. C., Movahedi, M., Macrae, F., Mecklin, J. P., Moeslein, G., Olschwang, S., Eccles, D., Evans, G., Maher, E. R., Bertario, L., Bisgaard, M. L., Dunlop, M., Ho, J. W. C., Hodgson, S., Lindblom, A., Lubinski, J., Morrison, P. J., Murday, V., Ramesar, R., … Burn, J. (2012). Long-term effect of resistant starch on cancer risk in carriers of hereditary colorectal cancer: An analysis from the CAPP2 randomised controlled trial. *The Lancet Oncology*, *13*(12), 1242–1249. https://doi.org/10.1016/S1470-2045(12)70475-8

Matsuzaki, J., Kato, K., Oono, K., Tsuchiya, N., Sudo, K., Shimomura, A., Tamura, K., Shiino, S., Kinoshita, T., Daiko, H., Wada, T., Katai, H., Ochiai, H., Kanemitsu, Y., Takamaru, H., Abe, S., Saito, Y., Boku, N., Kondo, S., … Takashima, H. (2023). Prediction of tissue-of-origin of early stage cancers using serum miRNomes. *JNCI Cancer Spectrum*, *7*(1), pkac080. https://doi.org/10.1093/jncics/pkac080

Mauri, G., Vitiello, P. P., Sogari, A., Crisafulli, G., Sartore-Bianchi, A., Marsoni, S., Siena, S., & Bardelli, A. (2022). Liquid biopsies to monitor and direct cancer treatment in colorectal cancer. *British Journal of Cancer, 127*(3), 394–407. https://doi.org/10.1038/s41416-022-01769-8

Mctiernan, A., Friedenreich, C. M., Katzmarzyk, P. T., Powell, K. E., Macko, R., Buchner, D., Pescatello, L. S., Bloodgood, B., Tennant, B., Vaux-Bjerke, A., George, S. M., Troiano, R. P., & Piercy, K. L. (2019). Physical activity in cancer prevention and survival: A systematic review. *Medicine and Science in Sports and Exercise, 51*(6), 1252–1261. https://doi.org/10.1249/MSS.0000000000001937

Mens, M. M. J., Maas, S. C. E., Klap, J., Weverling, G. J., Klatser, P., Brakenhoff, J. P. J., van Meurs, J. B. J., Uitterlinden, A. G., Ikram, M. A., Kavousi, M., & Ghanbari, M. (2020). Multi-omics analysis reveals microRNAs associated with cardiometabolic traits. *Frontiers in Genetics*, *11*, 110. https://doi.org/10.3389/fgene.2020.00110

Møller, P. (2020). The Prospective Lynch Syndrome Database reports enable evidence-based personal precision health care. *Hereditary Cancer in Clinical Practice*, *18*(1), 1–7. https://doi.org/10.1186/s13053-020-0138-0

Møller, P., Seppälä, T., Bernstein, I., Holinski-Feder, E., Sala, P., Evans, D. G., Lindblom, A., MacRae, F., Blanco, I., Sijmons, R., Jeffries, J., Vasen, H., Burn, J., Nakken, S., Hovig, E., Rødland, E. A., Tharmaratnam, K., De Vos Tot Nederveen Cappel, W. H., Hill, J., … Capella, G. (2017a). Incidence of and survival after subsequent cancers in carriers of pathogenic MMR variants with previous cancer: A report from the prospective Lynch syndrome database. *Gut*, *66*(9), 1657–1664. https://doi.org/10.1136/gutjnl-2016-311403

Møller, P., Seppälä, T., Bernstein, I., Holinski-Feder, E., Sala, P., Evans, D. G., Lindblom, A., Macrae, F., Blanco, I., Sijmons, R., Jeffries, J., Vasen, H., Burn,

J., Nakken, S., Hovig, E., Rødland, E. A., Tharmaratnam, K., De Vos Tot Nederveen Cappel, W. H., Hill, J., … Möslein, G. (2017b). Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: First report from the prospective Lynch syndrome database. *Gut*, *66*(3), 464–472. https://doi.org/10.1136/gutjnl-2015-309675

Møller, P., Seppälä, T. T., Ahadova, A., Crosbie, E. J., Holinski-Feder, E., Scott, R., Haupt, S., Möslein, G., Winship, I., Broeke, S. W. B.-T., Kohut, K. E., Ryan, N., Bauerfeind, P., Thomas, L. E., Evans, D. G., Aretz, S., Sijmons, R. H., Half, E., Heinimann, K., … Prospective Lynch Syndrome Database (www.plsd.eu) and The European Hereditary Tumour Group (www.ehtg.org). (2023). Dominantly inherited micro-satellite instable cancer - the four Lynch syndromes - an EHTG, PLSD position statement. *Hereditary Cancer in Clinical Practice*, *21*(1), 19. https://doi.org/10.1186/s13053-023-00263-3

Møller, P., Seppälä, T. T., Bernstein, I., Holinski-Feder, E., Sala, P., Evans, D. G., Lindblom, A., Macrae, F., Blanco, I., Sijmons, R. H., Jeffries, J., Vasen, H. F. A., Burn, J., Nakken, S., Hovig, E., Rødland, E. A., Tharmaratnam, K., De Vos Tot Nederveen Cappel, W. H., Hill, J., … Capella, G. (2018). Cancer risk and survival in path-MMR carriers by gene and gender up to 75 years of age: A report from the Prospective Lynch Syndrome Database. *Gut*, *67*(7), 1306–1316. https://doi.org/10.1136/gutjnl-2017-314057

Moore, S. C., Lee, I. M., Weiderpass, E., Campbell, P. T., Sampson, J. N., Kitahara, C. M., Keadle, S. K., Arem, H., De Gonzalez, A. B., Hartge, P., Adami, H. O., Blair, C. K., Borch, K. B., Boyd, E., Check, D. P., Fournier, A., Freedman, N. D., Gunter, M., Johannson, M., … Patel, A. V. (2016). Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. *JAMA Internal Medicine*, *176*(6), 816–825. https://doi.org/10.1001/jamainternmed.2016.1548

Moreno, E. C., Pascual, A., Prieto-Cuadra, D., Laza, V. F., Molina-Cerrillo, J., Ramos-Muñoz, M. E., Rodríguez-Serrano, E. M., Soto, J. L., Carrato, A., García-Bermejo, M. L., & Guillén-Ponce, C. (2019). Novel molecular characterization of colorectal primary tumors based on mirnas. *Cancers*, *11*(3), 346. https://doi.org/10.3390/cancers11030346

Mori, M. A., Ludwig, R. G., Garcia-Martin, R., Brandão, B. B., & Kahn, C. R. (2019). Extracellular miRNAs: From biomarkers to mediators of physiology and disease. *Cell Metabolism, 30*(4), 656–673. https://doi.org/10.1016/j.cmet.2019.07.011

Movahedi, M., Bishop, D. T., Macrae, F., Mecklin, J. P., Moeslein, G., Olschwang, S., Eccles, D., Evans, D. G., Maher, E. R., Bertario, L., Bisgaard, M. L., Dunlop, M. G., Ho, J. W. C., Hodgson, S. V., Lindblom, A., Lubinski, J., Morrison, P. J., Murday, V., Ramesar, R. S., … Mathers, J. C. (2015). Obesity, aspirin, and risk of colorectal cancer in carriers of hereditary colorectal cancer: A prospective investigation in the CAPP2 study. *Journal*

of Clinical Oncology, *33*(31), 3591–3597. https://doi.org/10.1200/JCO.2014.58.9952

Mullany, L. E., Herrick, J. S., Sakoda, L. C., Samowitz, W., Stevens, J. R., Wolff, R. K., & Slattery, M. L. (2018). miRNA involvement in cell cycle regulation in colorectal cancer cases. *Genes & Cancer, 9*(1-2), 53–65. https://doi.org/10.18632/genesandcancer.167

Mullany, L. E., & Slattery, M. L. (2019). The functional role of miRNAs and colorectal cancer: Insights from a large population-based study. *Cancer Biology and Medicine, 16*(2), 211–219. https://doi.org/10.20892/j.issn.2095-3941.2018.0514

Murphy, N., Strickler, H. D., Stanczyk, F. Z., Xue, X., Wassertheil-Smoller, S., Rohan, T. E., Ho, G. Y. F., Anderson, G. L., Potter, J. D., & Gunter, M. J. (2015). A prospective evaluation of endogenous sex hormone levels and colorectal cancer risk in postmenopausal women. *Journal of the National Cancer Institute*, *107*(10), djv210. https://doi.org/10.1093/jnci/djv210

Nelson, L. R., & Bulun, S. E. (2001). Estrogen production and action. *Journal of the American Academy of Dermatology*, *45*(3), S116–S124. https://doi.org/10.1067/mjd.2001.117432

Ng, E. K. O., Chong, W. W. S., Jin, H., Lam, E. K. Y., Shin, V. Y., Yu, J., Poon, T. C. W., Ng, S. S. M., & Sung, J. J. Y. (2009). Differential expression of microRNAs in plasma of patients with colorectal cancer: A potential marker for colorectal cancer screening. *Gut*, *58*(10), 1375–1381. https://doi.org/10.1136/gut.2008.167817

Ng, M., Fleming, T., Robinson, M., Thomson, B., Graetz, N., Margono, C., Mullany, E. C., Biryukov, S., Abbafati, C., Abera, S. F., Abraham, J. P., Abu-Rmeileh, N. M. E., Achoki, T., Albuhairan, F. S., Alemu, Z. A., Alfonso, R., Ali, M. K., Ali, R., Guzman, N. A., … Gakidou, E. (2014). Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: A systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*, *384*(9945), 766–781. https://doi.org/10.1016/S0140-6736(14)60460-8

Nielsen, S., Åkerström, T., Rinnov, A., Yfanti, C., Scheele, C., Pedersen, B. K., & Laye, M. J. (2014). The miRNA plasma signature in response to acute aerobic exercise and endurance training. *PLoS ONE*, *9*(2), e87308. https://doi.org/10.1371/journal.pone.0087308

O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in Endocrinology, 9,* 402. https://doi.org/10.3389/fendo.2018.00402

O'Donnell, K. A., Wentzel, E. A., Zeller, K. I., Dang, C. V., & Mendell, J. T. (2005). c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, *435*(7043), 839–843. https://doi.org/10.1038/nature03677

Ooi, C. H., Oh, H. K., Wang, H. Z. A., Tan, A. L. K., Wu, J., Lee, M., Rha, S. Y., Chung, H. C., Virshup, D. M., & Tan, P. (2011). A densely interconnected genome-wide network of micrornas and oncogenic pathways revealed

using gene expression signatures. *PLoS Genetics*, *7*(12), e1002415. https://doi.org/10.1371/journal.pgen.1002415

Ortega, F. J., Mercader, J. M., Catalán, V., Moreno-Navarrete, J. M., Pueyo, N., Sabater, M., Gómez-Ambrosi, J., Anglada, R., Fernández-Formoso, J. A., Ricart, W., Frühbeck, G., & Fernández-Real, J. M. (2013). Targeting the circulating microRNA signature of obesity. *Clinical Chemistry*, *59*(5), 781–792. https://doi.org/10.1373/clinchem.2012.195776

Otmani, K., & Lewalle, P. (2021). Tumor suppressor miRNA in cancer cells and the tumor microenvironment: Mechanism of deregulation and clinical implications. *Frontiers in Oncology, 11,* 708765. https://doi.org/10.3389/fonc.2021.708765

Otsuka, K., Nishiyama, H., Kuriki, D., Kawada, N., & Ochiya, T. (2023). Connecting the dots in the associations between diet, obesity, cancer, and microRNAs. *Seminars in Cancer Biology, 93,* 52–69. https://doi.org/10.1016/j.semcancer.2023.05.001

Papadimitriou, N., Dimou, N., Tsilidis, K. K., Banbury, B., Martin, R. M., Lewis, S. J., Kazmi, N., Robinson, T. M., Albanes, D., Aleksandrova, K., Berndt, S. I., Timothy Bishop, D., Brenner, H., Buchanan, D. D., Bueno-de-Mesquita, B., Campbell, P. T., Castellví-Bel, S., Chan, A. T., Chang-Claude, J., … Murphy, N. (2020). Physical activity and risks of breast and colorectal cancer: A Mendelian randomisation analysis. *Nature Communications*, *11*(1), 597. https://doi.org/10.1038/s41467-020-14389-8

Parasramka, M. A., Dashwood, W. M., Wang, R., Saeed, H. H., Williams, D. E., Ho, E., & Dashwood, R. H. (2012). A role for low-abundance miRNAs in colon cancer: The miR-206/Krüppel-like factor 4 (KLF4) axis. *Clinical Epigenetics*, *4*(1), 16. https://doi.org/10.1186/1868-7083-4-16

Pardini, B., Ferrero, G., Tarallo, S., Gallo, G., Francavilla, A., Licheri, N., Trompetto, M., Clerico, G., Senore, C., Peyre, S., Vymetalkova, V., Vodickova, L., Liska, V., Vycital, O., Levy, M., Macinga, P., Hucl, T., Budinska, E., Vodicka, P., … Naccarati, A. (2023). A fecal microRNA signature by small RNA sequencing accurately distinguishes colorectal cancers: Results from a multicenter study. *Gastroenterology*, *165*(3), 582-599.e8. https://doi.org/10.1053/j.gastro.2023.05.037

Parsons, R., Li, G.-M., Longley, M. J., Fang, W.-H., Papadopoulos, N., Jen, J., De La Chapelle, A., Kinzler, K. W., Vogelstein, B., & Modricht, P. (1993). Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell, 75* (6), 1227–1236. https://doi.org/10.1016/0092-8674(93)90331-j

Pastor, D. M., & Schlom, J. (2021). Immunology of Lynch syndrome. *Current Oncology Reports*, *23*(8), 96. https://doi.org/10.1007/s11912-021-01085-z

Pavicic, W., Perkiö, E., Kaur, S., & Peltomäki, P. (2011). Altered methylation at microRNA-associated CpG islands in hereditary and sporadic carcinomas: A methylation- specific multiplex ligation-dependent probe amplification (MS-MLPA)-based approach. *Molecular Medicine*, *17*(7–8), 726–735. https://doi.org/10.2119/molmed.2010.00239

Peltomaki, P., Aaltonen, L. A., Sistonen, P., Pylkkanen, L., Mecklin, J.-P., Jarvinen, H., Green, J. S., Jass, J. R., Weber, J. L., Leach, F. S., Petersen, G. M., & Hamilton, S. R. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. *Science, 260*(5109), 810–812. https://doi.org/10.1126/science.8484120

Peltomäki, P., Nyström, M., Mecklin, J. P., & Seppälä, T. T. (2023). Lynch syndrome genetics and clinical implications. *Gastroenterology*, *164*(5), 783–799. https://doi.org/10.1053/j.gastro.2022.08.058

Peng, Y., & Croce, C. M. (2016). The role of microRNAs in human cancer. *Signal Transduction and Targeted Therapy, 1,* 15004. https://doi.org/10.1038/sigtrans.2015.4

Pichler, M., Ress, A. L., Winter, E., Stiegelbauer, V., Karbiener, M., Schwarzenbacher, D., Scheideler, M., Ivan, C., Jahn, S. W., Kiesslich, T., Gerger, A., Bauernhofer, T., Calin, G. A., & Hoefler, G. (2014). MiR-200a regulates epithelial to mesenchymal transition-related gene expression and determines prognosis in colorectal cancer patients. *British Journal of Cancer*, *110*(6), 1614–1621. https://doi.org/10.1038/bjc.2014.51

Power, R. F., Doherty, D. E., Parker, I., Gallagher, D. J., Lowery, M. A., & Cadoo, K. A. (2024). Modifiable risk factors and risk of colorectal and endometrial cancers in Lynch syndrome: A systematic review and meta-analysis. *JCO Precision Oncology*, *8*, e2300196. https://doi.org/10.1200/PO.23.00196

Prior, I. A., Hood, F. E., & Hartley, J. L. (2020). The frequency of ras mutations in cancer. *Cancer Research*, *80*(14), 2669–2974. https://doi.org/10.1158/0008-5472.CAN-19-3682

Pritchard, C. C., Kroh, E., Wood, B., Arroyo, J. D., Dougherty, K. J., Miyaji, M. M., Tait, J. F., & Tewari, M. (2012). Blood cell origin of circulating microRNAs: A cautionary note for cancer biomarker studies. *Cancer Prevention Research*, *5*(3), 492–497. https://doi.org/10.1158/1940-6207.CAPR-11-0370

R Core Team. (2022). R: A language and environment for statistical computing (4.2.2). R Foundation for Statistical Computing.

Rapado-González, Ó., Majem, B., Álvarez-Castro, A., Díaz-Peña, R., Abalo, A., Suárez-Cabrera, L., Gil-Moreno, A., Santamaría, A., López-López, R., Muinelo-Romay, L., & Suarez-Cunqueiro, M. M. (2019). A novel saliva-based mirna signature for colorectal cancer diagnosis. *Journal of Clinical Medicine*, *8*(12), 2029. https://doi.org/10.3390/jcm8122029

Raut, J. R., Schöttker, B., Holleczek, B., Guo, F., Bhardwaj, M., Miah, K., Schrotz-King, P., & Brenner, H. (2021). A microRNA panel compared to environmental and polygenic scores for colorectal cancer risk prediction. *Nature Communications*, *12*(1), 4811. https://doi.org/10.1038/s41467-021-25067-8

Renehan, A. G., Roberts, D. L., & Dive, C. (2008). Obesity and cancer: Pathophysiological and biological mechanisms. *Archives of Physiology and Biochemistry, 114*(1), 71–83. https://doi.org/10.1080/13813450801954303

Renehan, A. G., Zwahlen, M., & Egger, M. (2015). Adiposity and cancer risk: New mechanistic insights from epidemiology. *Nature Reviews Cancer, 15*(8), 484–498. https://doi.org/10.1038/nrc3967

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2009). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Roudko, V., Cimen Bozkus, C., Greenbaum, B., Lucas, A., Samstein, R., & Bhardwaj, N. (2021). Lynch syndrome and MSI-H cancers: From mechanisms to "off-the-shelf" cancer vaccines. *Frontiers in Immunology, 12, 757804.* https://doi.org/10.3389/fimmu.2021.757804

Roush, S., & Slack, F. J. (2008). The let-7 family of microRNAs. *Trends in Cell Biology, 18*(10), 505–516. https://doi.org/10.1016/j.tcb.2008.07.007

Ruiz-Roso, M. B., Gil-Zamorano, J., López de las Hazas, M. C., Tomé-Carneiro, J., Crespo, M. C., Latasa, M. J., Briand, O., Sánchez-López, D., Ortiz, A. I., Visioli, F., Martínez, J. A., & Dávalos, A. (2020). Intestinal lipid metabolism genes regulated by miRNAs. *Frontiers in Genetics*, *11,* 707. https://doi.org/10.3389/fgene.2020.00707

Samadder, N. J., Curtin, K., Pappas, L., Boucher, K., Mineau, G. P., Smith, K., Fraser, A., Wan, Y., Provenzale, D., Kinney, A. Y., Ulrich, C., & Burt, R. W. (2016). Risk of incident colorectal cancer and death after colonoscopy: A population-based study in Utah. *Clinical Gastroenterology and Hepatology*, *14*(2), 279-286.e2. https://doi.org/10.1016/j.cgh.2015.08.033

Sapp, R. M., Shill, D. D., Roth, S. M., & Hagberg, J. M. (2017). Circulating microRNAs in acute and chronic exercise: More than mere biomarkers. *Journal of Applied Physiology, 122*(3), 702–717. https://doi.org/10.1152/japplphysiol.00982.2016

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., & Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, *455*(7209), 58–63. https://doi.org/10.1038/nature07228

Seppä, K., Tanskanen, T., Heikkinen, S., Malila, N., & Pitkäniemi, J. (2023). Syöpä 2021. Tilastoraportti Suomen syöpätilanteesta. *Suomen Syöpäyhdistys*. https://syoparekisteri.fi/assets/files/2023/05/Syopa_2021_final_31052023.pdf

Seppälä, T. T., Burkhart, R. A., & Katona, B. W. (2023). Hereditary colorectal, gastric, and pancreatic cancer: Comprehensive review. *BJS Open, 7*(3), zrad023. https://doi.org/10.1093/bjsopen/zrad023

Seppälä, T. T., Dominguez-Valentin, M., Sampson, J. R., & Møller, P. (2021). Prospective observational data informs understanding and future management of Lynch syndrome: Insights from the Prospective Lynch Syndrome Database (PLSD). *Familial Cancer, 20*(1), 35–39. https://doi.org/10.1007/s10689-020-00193-2

Seth, S., Ager, A., Arends, M. J., & Frayling, I. M. (2018). Lynch syndrome – cancer pathways, heterogeneity and immune escape. *Journal of Pathology*, *246*(2), 129–133. https://doi.org/10.1002/path.5139

Shaw, E., Farris, M. S., Stone, C. R., Derksen, J. W. G., Johnson, R., Hilsden, R. J., Friedenreich, C. M., & Brenner, D. R. (2018). Effects of physical activity on colorectal cancer risk among family history and body mass index subgroups: A systematic review and meta-analysis. *BMC Cancer*, *18*(1), 71. https://doi.org/10.1186/s12885-017-3970-5

Sheedy, P., & Medarova, Z. (2018). The fundamental role of miR-10b in metastatic cancer. *American Journal of Cancer Research, 8*(9), 1674–1688.

Shi, Y. J., Fang, Y. X., Tian, T. G., Chen, W. P., Sun, Q., Guo, F. Q., Gong, P. Q., Li, C. M., Wang, H., Hu, Z. Q., & Li, X. X. (2023). Discovery of extracellular vesicle-delivered miR-185-5p in the plasma of patients as an indicator for advanced adenoma and colorectal cancer. *Journal of Translational Medicine*, *21*(1), 421. https://doi.org/10.1186/s12967-023-04249-6

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer statistics, 2021. *CA: A Cancer Journal for Clinicians*, *71*(1), 7–33. https://doi.org/10.3322/caac.21654

Sillanpää, E., Cheng, S., Häkkinen, K., Finni, T., Walker, S., Pesola, A., Ahtiainen, J., Stenroth, L., Selänne, H., & Sipilä, S. (2014). Body composition in 18- to 88-year-old adults - Comparison of multifrequency bioimpedance and dual-energy X-ray absorptiometry. *Obesity*, *22*(1), 101–109. https://doi.org/10.1002/oby.20583

Skog, J., Würdinger, T., van Rijn, S., Meijer, D. H., Gainche, L., Curry, W. T., Carter, B. S., Krichevsky, A. M., & Breakefield, X. O. (2008). Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nature Cell Biology*, *10*(12), 1470–1476. https://doi.org/10.1038/ncb1800

Slattery, M. L., Herrick, J. S., Mullany, L. E., Stevens, J. R., & Wolff, R. K. (2017). Diet and lifestyle factors associated with miRNA expression in colorectal tissue. *Pharmacogenomics and Personalized Medicine*, *10*, 1–16. https://doi.org/10.2147/PGPM.S117796

Slattery, M. L., Lee, F. Y., Pellatt, A. J., Mullany, L. E., Stevens, J. R., Samowitz, W. S., Wolff, R. K., & Herrick, J. S. (2017). Infrequently expressed miRNAs in colorectal cancer tissue and tumor molecular phenotype. *Modern Pathology*, *30*(8), 1152–1169. https://doi.org/10.1038/modpathol.2017.38

Smith, A. W., Cronin, K. A., Bowles, H., Willis, G., Jacobs, D. R., Ballard-Barbash, R., & Troiano, R. P. (2013). Reproducibility of physical activity recall over fifteen years: Longitudinal evidence from the CARDIA study. *BMC Public Health*, *13*(1), 180. https://doi.org/10.1186/1471-2458-13-180

Snel, B., Lehmann, G., Bork, P., & Huynen, M. A. (2000). String: A web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Research*, *28*(18), 3442–3444. https://doi.org/10.1093/nar/28.18.3442

Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., & Forbes, S. A. (2018). The COSMIC Cancer Gene Census: Describing genetic dysfunction across all human cancers. *Nature Reviews Cancer, 18*(11), 696–705. https://doi.org/10.1038/s41568-018-0060-1

Spier, I., Yin, X., Richardson, M., Pineda, M., Laner, A., Ritter, D., Boyle, J., Mur, P., Hansen, T. v. O., Shi, X., Mahmood, K., Plazzer, J. P., Ognedal, E., Nordling, M., Farrington, S. M., Yamamoto, G., Baert-Desurmont, S., Martins, A., Borras, E., … Aretz, S. (2023). Gene-specific ACMG/AMP classification criteria for germline APC variants: Recommendations from the ClinGen InSiGHT Hereditary Colorectal Cancer/Polyposis Variant Curation Expert Panel. *Genetics in Medicine, 26*(2), 100992. https://doi.org/10.1016/j.gim.2023.100992

Stevens, G. A., Singh, G. M., Lu, Y., Danaei, G., Lin, J. K., Finucane, M. M., Bahalim, A. N., McIntire, R. K., Gutierrez, H. R., Cowan, M., Paciorek, C. J., Farzadfar, F., Riley, L., & Ezzati, M. (2012). National, regional, and global trends in adult overweight and obesity prevalences. *Population Health Metrics*, *10*(1), 22. https://doi.org/10.1186/1478-7954-10-22

Sticht, C., De La Torre, C., Parveen, A., & Gretz, N. (2018). Mirwalk: An online resource for prediction of microrna binding sites. *PLoS ONE*, *13*(10), e0206239. https://doi.org/10.1371/journal.pone.0206239

Stigliano, V., Assisi, D., Cosimelli, M., Palmirotta, R., Giannarelli, D., Mottolese, M., Mete, L. S., Mancini, R., & Casale, V. (2008). Survival of hereditary non-polyposis colorectal cancer patients compared with sporadic colorectal cancer patients. *Journal of Experimental and Clinical Cancer Research*, *27*(1), 39. https://doi.org/10.1186/1756-9966-27-39

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature, 458*(7239), 719–724. https://doi.org/10.1038/nature07943

Straub, R. H. (2007). The complex role of estrogens in inflammation. *Endocrine Reviews, 28*(5), 521–574. https://doi.org/10.1210/er.2007-0001

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, *71*(3), 209–249. https://doi.org/10.3322/caac.21660

Sung, H., Siegel, R. L., Torre, L. A., Pearson-Stuttard, J., Islami, F., Fedewa, S. A., Goding Sauer, A., Shuval, K., Gapstur, S. M., Jacobs, E. J., Giovannucci, E. L., & Jemal, A. (2019). Global patterns in excess body weight and the associated cancer burden. *CA: A Cancer Journal for Clinicians*, *69*(2), 88–112. https://doi.org/10.3322/caac.21499

Sur, D., Advani, S., & Braithwaite, D. (2022). MicroRNA panels as diagnostic biomarkers for colorectal cancer: A systematic review and meta-analysis. *Frontiers in Medicine, 9,* 915226. https://doi.org/10.3389/fmed.2022.915226

Svrcek, M., El-Murr, N., Wanherdrick, K., Dumont, S., Beaugerie, L., Cosnes, J., Colombel, J. F., Tiret, E., Fléjou, J. F., Lesuffleur, T., & Duval, A. (2013). Overexpression of microRNAs-155 and 21 targeting mismatch repair

proteins in inflammatory bowel diseases. *Carcinogenesis*, *34*(4), 828–834. https://doi.org/10.1093/carcin/bgs408

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074

Szulc, P., Duboeuf, F., & Chapurlat, R. (2017). Age-related changes in fat mass and distribution in men—the cross-sectional STRAMBO study. *Journal of Clinical Densitometry*, *20*(4), 472–479. https://doi.org/10.1016/j.jocd.2016.08.003

Taniguchi, K., & Karin, M. (2018). NF-B, inflammation, immunity and cancer: Coming of age. *Nature Reviews Immunology, 18*(5), 309–324. https://doi.org/10.1038/nri.2017.142

Tate, J. G., Bamford, S., Jubb, H. C., Sondka, Z., Beare, D. M., Bindal, N., Boutselakis, H., Cole, C. G., Creatore, C., Dawson, E., Fish, P., Harsha, B., Hathaway, C., Jupe, S. C., Kok, C. Y., Noble, K., Ponting, L., Ramshaw, C. C., Rye, C. E., … Forbes, S. A. (2019). COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*, *47*(D1), D941–D947. https://doi.org/10.1093/nar/gky1015

ten Broeke, S. W., Brohet, R. M., Tops, C. M., Van Der Klift, H. M., Velthuizen, M. E., Bernstein, I., Munar, G. C., Garcia, E. G., Hoogerbrugge, N., Letteboer, T. G. W., Menko, F. H., Lindblom, A., Mensenkamp, A. R., Moller, P., Van Os, T. A., Rahner, N., Redeker, B. J. W., Sijmons, R. H., Spruijt, L., … Wijnen, J. T. (2015). Lynch syndrome caused by germline PMS2 mutations: Delineating the cancer risk. *Journal of Clinical Oncology*, *33*(4), 319–325. https://doi.org/10.1200/JCO.2014.57.8088

ten Broeke, S. W., van Bavel, T. C., Jansen, A. M. L., Gómez-García, E., Hes, F. J., van Hest, L. P., Letteboer, T. G. W., Olderode-Berends, M. J. W., Ruano, D., Spruijt, L., Suerink, M., Tops, C. M., van Eijk, R., Morreau, H., van Wezel, T., & Nielsen, M. (2018). Molecular background of colorectal tumors from patients with Lynch syndrome associated germline variants in PMS2. *Gastroenterology*, *155*(3), 844–851. https://doi.org/10.1053/j.gastro.2018.05.020

ten Broeke, S. W., van der Klift, H. M., Tops, C. M., Aretz, S., Bernstein, I., Buchanan, D. D., de la Chapelle, A., Capella, G., Clendenning, M., Engel, C., Gallinger, S., Gomez Garcia, E., Figueiredo, J. C., Haile, R., Hampel, H. L., van Hest, L., Hopper, J. L., Hoogerbrugge, N., von Knebel Doeberitz, M., … Ko Win, A. (2018). Cancer risks for PMS2-associated Lynch syndrome. *Journal of Clinical Oncology*, *36*(29), 2961–2968. https://doi.org/10.1200/JCO

Therneau, T. M., & Grambsch, P. M. (2000). *The Cox Model BT - Modeling Survival Data: Extending the Cox Model*, pp. 39–77. Springer. https://doi.org/10.1007/978-1-4757-3294-8_3

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*(1), 267–288. http://www.jstor.org/stable/2346178

Tibshirani, R. (1997). The lasso method for variable selection in the Cox model. *Statistics in Medicine*, *16*(4), 385–395. https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3

Toden, S., & Goel, A. (2022). Non-coding RNAs as liquid biopsy biomarkers in cancer. *British Journal of Cancer*, *126*(3), 351–360. https://doi.org/10.1038/s41416-021-01672-8

Tuomela, J., Kaprio, J., Sipilä, P. N., Silventoinen, K., Wang, X., Ollikainen, M., & Piirtola, M. (2019). Accuracy of self-reported anthropometric measures — Findings from the Finnish Twin Study. *Obesity Research and Clinical Practice*, *13*(6), 522–528. https://doi.org/10.1016/j.orcp.2019.10.006

Umar, A., Boland, C. R., Terdiman, J. P., Syngal, S., de la Chapelle, A., Rüschoff, J., Fishel, R., Lindor, N. M., Burgart, L. J., Hamelin, R., Hamilton, S. R., Hiatt, R. A., Jass, J., Lindblom, A., Lynch, H. T., Peltomaki, P., Ramsey, S. D., Rodriguez-Bigas, M. A., Vasen, H. F. A., … Srivastava, S. (2004). Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *Journal of the National Cancer Institute*, *96*(4), 261–268. https://doi.org/10.1093/jnci/djh034

Unger, M., & Kather, J. N. (2024). A systematic analysis of deep learning in genomics and histopathology for precision oncology. *BMC Medical Genomics*, *17*(1), 48. https://doi.org/10.1186/s12920-024-01796-9

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J. J., & Lötvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nature Cell Biology*, *9*(6), 654–659. https://doi.org/10.1038/ncb1596

Valeri, N., Gasparini, P., Fabbri, M., Braconi, C., Veronese, A., Lovat, F., Adair, B., Vannini, I., Fanini, F., Bottoni, A., Costinean, S., Sandhu, S. K., Nuovo, G. J., Alder, H., Gafa, R., Calore, F., Ferracin, M., Lanza, G., Volinia, S., … Croce, C. M. (2010). Modulation of mismatch repair and genomic stability by miR-155. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(15), 6982–6987. https://doi.org/10.1073/pnas.1002472107

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*(3 SE-Articles), 1–67. https://doi.org/10.18637/jss.v045.i03

Vasen, H. F. A., Abdirahman, M., Brohet, R., Langers, A. M. J., Kleibeuker, J. H., van Kouwen, M., Koornstra, J. J., Boot, H., Cats, A., Dekker, E., Sanduleanu, S., Poley, J. W., Hardwick, J. C. H., de Vos tot Nederveen Cappel, W. H., van der Meulen-de Jong, A. E., Tan, T. G., Jacobs, M. A. J. M., Mohamed, F. L. A., de Boer, S. Y., … Nagengast, F. M. (2010). One to 2-year surveillance intervals reduce risk of colorectal cancer in families with Lynch syndrome. *Gastroenterology*, *138*(7), 2300–2306. https://doi.org/10.1053/j.gastro.2010.02.053

Vasen, H. F. A., Mecklin, J.-P., Meera Khan, P., & Lynch, H. T. (1991). The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). *Diseases of the Colon and Rectum*, *34*(5), 424–425. https://doi.org/10.1007/BF02053699

Vasen, H. F. A., Watson, P., Mecklin, J. P., & Lynch, H. T. (1999). New clinical criteria for hereditary nonpolyposis colorectal definition of HNPCC. *Gastroenterology*, *116*, 1453–1456. https://doi.org/10.1016/s0016-5085(99)70510-x

Veettil, S. K., Wong, T. Y., Loo, Y. S., Playdon, M. C., Lai, N. M., Giovannucci, E. L., & Chaiyakunapruk, N. (2021). Role of diet in colorectal cancer incidence: Umbrella review of meta-analyses of prospective observational studies. *JAMA Network Open, 4*(2), e2037341. https://doi.org/10.1001/jamanetworkopen.2020.37341

Vogelstein, B., & Kinzler, K. W. (2004). Cancer genes and the pathways they control. *Nature Medicine, 10*(8), 789–799. https://doi.org/10.1038/nm1087

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*, *340*(6127), 1546–1558. https://doi.org/10.1126/science.1235122

Vychytilova-Faltejskova, P., Radova, L., Sachlova, M., Kosarova, Z., Slaba, K., Fabian, P., Grolich, T., Prochazka, V., Kala, Z., Svoboda, M., Kiss, I., Vyzula, R., & Slaby, O. (2016). Serum-based microRNA signatures in early diagnosis and prognosis prediction of colon cancer. *Carcinogenesis*, *37*(10), 941–950. https://doi.org/10.1093/carcin/bgw078

Wang, H. Q., Man, Q. W., Huo, F. Y., Gao, X., Lin, H., Li, S. R., Wang, J., Su, F. C., Cai, L., Shi, Y., Liu, B., & Bu, L. L. (2022). STAT3 pathway in cancers: Past, present, and future. *MedComm, 3*(2), e124. https://doi.org/10.1002/mco2.124

Wang, K., Yuan, Y., Cho, J. H., McClarty, S., Baxter, D., & Galas, D. J. (2012). Comparing the MicroRNA spectrum between serum and plasma. *PLoS ONE*, *7*(7), e41561. https://doi.org/10.1371/journal.pone.0041561

Wang, X. (2008). miRDB: A microRNA target prediction and functional annotation database with a wiki interface. *RNA*, *14*(6), 1012–1017. https://doi.org/10.1261/rna.965408

Waters, P. S., McDermott, A. M., Wall, D., Heneghan, H. M., Miller, N., Newell, J., Kerin, M. J., & Dwyer, R. M. (2012). Relationship between circulating and tissue microRNAs in a murine model of breast cancer. *PLoS ONE*, *7*(11), e50459. https://doi.org/10.1371/journal.pone.0050459

Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., Huang, K. H., Lee, M. J., Galas, D. J., & Wang, K. (2010). The microRNA spectrum in 12 body fluids. *Clinical Chemistry*, *56*(11), 1733–1741. https://doi.org/10.1373/clinchem.2010.147405

Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Sander, C., Stuart, J. M., Chang, K., Creighton, C. J., Davis, C., Donehower, L., Drummond, J., Wheeler, D., Ally, A., Balasundaram, M., Birol, I., Butterfield, Y. S. N., Chu, A., … Kling, T. (2013). The cancer

genome atlas pan-cancer analysis project. *Nature Genetics*, *45*(10), 1113–1120. https://doi.org/10.1038/ng.2764

White, R. (1784). Remarks on the nature and treatment of cancers. *The London Medical Journal*, *5*(1), 70–75.

Wikberg, M. L., Myte, R., Palmqvist, R., van Guelpen, B., & Ljuslinder, I. (2018). Plasma miRNA can detect colorectal cancer, but how early? *Cancer Medicine*, *7*(5), 1697–1705. https://doi.org/10.1002/cam4.1398

Win, A. K., Dowty, J. G., Antill, Y. C., English, D. R., Baron, J. A., Young, J. P., Giles, G. G., Southey, M. C., Winship, I., Lipton, L., Parry, S., Thibodeau, S. N., Haile, R. W., Gallinger, S., Le Marchand, L., Lindor, N. M., Newcomb, P. A., Hopper, J. L., & Jenkins, M. A. (2011). Body mass index in early adulthood and endometrial cancer risk for mismatch repair gene mutation carriers. *Obstetrics and Gynecology*, *117*(4), 899–905. https://doi.org/10.1097/AOG.0b013e3182110ea3

Win, A. K., Dowty, J. G., Reece, J. C., Lee, G., Templeton, A. S., Plazzer, J. P., Buchanan, D. D., Akagi, K., Aksoy, S., Alonso, A., Alvarez, K., Amor, D. J., Ankathil, R., Aretz, S., Arnold, J. L., Aronson, M., Austin, R., Backman, A. S., Bajwa-ten Broeke, S. W., … Jenkins, M. A. (2021). Variation in the risk of colorectal cancer in families with Lynch syndrome: A retrospective cohort study. *The Lancet Oncology*, *22*(7), 1014–1022. https://doi.org/10.1016/S1470-2045(21)00189-3

Win, A. K., Jenkins, M. A., Dowty, J. G., Antoniou, A. C., Lee, A., Giles, G. G., Buchanan, D. D., Clendenning, M., Rosty, C., Ahnen, D. J., Thibodeau, S. N., Casey, G., Gallinger, S., Le Marchand, L., Haile, R. W., Potter, J. D., Zheng, Y., Lindor, N. M., Newcomb, P. A., … MacInnis, R. J. (2017). Prevalence and penetrance of major genes and polygenes for colorectal cancer. *Cancer Epidemiology Biomarkers and Prevention*, *26*(3), 404–412. https://doi.org/10.1158/1055-9965.EPI-16-0693

Win, A. K., MacInnis, R. J., Hopper, J. L., & Jenkins, M. A. (2012). Risk prediction models for colorectal cancer: A review. *Cancer Epidemiology Biomarkers and Prevention*, *21*(3), 398–410. https://doi.org/10.1158/1055-9965.EPI-11-0771

Winter, J., & Diederichs, S. (2011). Argonaute proteins regulate microRNA stability: Increased microRNA abundance by Argonaute proteins is due to microRNA stabilization. *RNA Biology*, *8*(6), 1149–1157. https://doi.org/10.4161/rna.8.6.17665

Yamada, A., Horimatsu, T., Okugawa, Y., Nishida, N., Honjo, H., Ida, H., Kou, T., Kusaka, T., Sasaki, Y., Yagi, M., Higurashi, T., Yukawa, N., Amanuma, Y., Kikuchi, O., Muto, M., Ueno, Y., Nakajima, A., Chiba, T., Boland, C. R., & Goel, A. (2015). Serum MIR-21, MIR-29a, and MIR-125b are promising biomarkers for the early detection of colorectal neoplasia. *Clinical Cancer Research*, *21*(18), 4234–4242. https://doi.org/10.1158/1078-0432.CCR-14-2793

Yamaguchi, T., Iijima, T., Wakaume, R., Takahashi, K., Matsumoto, H., Nakano, D., Nakayama, Y., Mori, T., Horiguchi, S., & Miyaki, M. (2014).

Underexpression of miR-126 and miR-20b in hereditary and nonhereditary colorectal tumors. *Oncology*, *87*(1), 58–66. https://doi.org/10.1159/000363303

Yang, G., Zhang, Y., & Yang, J. (2019). A Five-microRNA signature as prognostic biomarker in colorectal cancer by bioinformatics analysis. *Frontiers in Oncology*, *9,* 1207. https://doi.org/10.3389/fonc.2019.01207

Yurgelun, M. B., & Chan, A. T. (2020). Aspirin for Lynch syndrome: A legacy of prevention. *The Lancet, 395*(10240), 1818–1820. https://doi.org/10.1016/S0140-6736(20)31298-8

Zhan, Y., Zhang, R., Li, C., Xu, X., Zhu, K., Yang, Z., Zheng, J., & Guo, Y. (2021). A microRNA-clinical prognosis model to predict the overall survival for kidney renal clear cell carcinoma. *Cancer Medicine*, *10*(17), 6128–6139. https://doi.org/10.1002/cam4.4148

Zhang, J., Raju, G. S., Chang, D. W., Lin, S. H., Chen, Z., & Wu, X. (2018). Global and targeted circulating microRNA profiling of colorectal adenoma and colorectal cancer. *Cancer*, *124*(4), 785–796. https://doi.org/10.1002/cncr.31062

Zhao, J., Xu, L., Sun, J., Song, M., Wang, L., Yuan, S., Zhu, Y., Wan, Z., Larsson, S., Tsilidis, K., Dunlop, M., Campbell, H., Rudan, I., Song, P., Theodoratou, E., Ding, K., & Li, X. (2023). Global trends in incidence, death, burden and risk factors of early-onset cancer from 1990 to 2019. *BMJ Oncology*, *2*(1), e000049. https://doi.org/10.1136/bmjonc-2023-000049

Zhou, C., Li, J., Li, J., Wan, Y., Li, T., Ma, P., Wang, Y., & Sang, H. (2016). Hsa-miR-137, hsa-miR-520e and hsa-miR-590-3p perform crucial roles in lynch syndrome. *Oncology Letters*, *12*(3), 2011–2017. https://doi.org/10.3892/ol.2016.4816

ORIGINAL PAPERS

Ia

SYSTEMIC CIRCULATING MICRORNA LANDSCAPE IN LYNCH SYNDROME

by

Sievänen T, Korhonen T-M, Jokela T, Ahtiainen M, Lahtinen L, Kuopio T, Lepistö A, Sillanpää E, Mecklin J-P, Seppälä TT, Laakkonen EK, 2022

International Journal of Cancer, *152*(5):932-944

https://doi.org/10.1002/ijc.34338

Received: 26 April 2022 | Revised: 2 September 2022 | Accepted: 7 October 2022

DOI: 10.1002/ijc.34338

CANCER GENETICS AND EPIGENETICS

IJC INTERNATIONAL JOURNAL of CANCER

Systemic circulating microRNA landscape in Lynch syndrome

Tero Sievänen[1] | Tia-Marje Korhonen[1] | Tiina Jokela[1] | Maarit Ahtiainen[2] | Laura Lahtinen[3] | Teijo Kuopio[3,4] | Anna Lepistö[5] | Elina Sillanpää[1,6] | Jukka-Pekka Mecklin[7,8] | Toni T. Seppälä[5,9] | Eija K. Laakkonen[1]

[1]Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

[2]Department of Education and Research, Central Finland Health Care District, Jyväskylä, Finland

[3]Department of Pathology, Central Finland Health Care District, Jyväskylä, Finland

[4]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

[5]Department of Surgery, Abdominal Center, Helsinki University Hospital, Helsinki, Finland

[6]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland

[7]Department of Surgery, Central Finland Health Care District, Jyväskylä, Finland

[8]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland

[9]Applied Tumor Genomics Research Program, University of Helsinki, Helsinki, Finland

Correspondence
Tero Sievänen and Eija K. Laakkonen, Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, P.O. Box 35 (VIV), 40014, Jyväskylä, Finland.
Email: tero.o.sievanen@jyu.fi (T. S.) and eija.k.laakkonen@jyu.fi (E. K. L.)

Funding information
Emil Aaltosen Säätiö; Finnish Cancer Foundation; HUS State Research Funds, Grant/Award Numbers: TYH2021123, TYH2022323; iCan Digital Precision Medicine Flagship; Jane ja Aatos Erkon Säätiö; KYS State Research Funds; Päivikki ja Sakari Sohlbergin

Abstract

Circulating microRNAs (c-miRs) are small noncoding RNA molecules that migrate throughout the body and regulate gene expression. Global c-miR expression patterns (c-miRnomes) change with sporadic carcinogenesis and have predictive potential in early detection of cancers. However, there are no studies that have assessed whether c-miRnomes display similar potential in carriers of inherited pathogenic mismatch-repair gene variants (*path_MMR*), known as Lynch syndrome (LS), who are predisposed to highly increased cancer risk. Using high-throughput sequencing and bioinformatic approaches, we conducted an exploratory analysis to characterize systemic c-miRnomes of *path_MMR* carriers, sporadic rectal cancer patients and non-LS controls. We showed for the first time that cancer-free *path_MMR* carriers have a systemic c-miRnome of 40 differentially expressed c-miRs that can distinguish them from non-LS controls. The systemic c-miRnome of cancer-free *path_MMR* carriers also resembles the systemic c-miRnomes of cancer patients with or without *path_MMR*. Our pathway analysis linked the found differentially expressed c-miRs to carcinogenesis. A total of 508 putative target genes were identified for 32 out of 40 differentially expressed c-miRs, and 238 of them were enriched in cancer-related pathways. The most enriched c-miR-target genes include well-known oncogenes and tumor suppressor genes such as *BCL2*, *AKT3*, *PIK3CA*, *KRAS*, *NRAS*, *CDKN1A* and *PIK3R1*. Taken together, our findings suggest that LS and sporadic carcinogenesis share common biological pathways and alterations in these pathways can produce a c-miR signature which can track potential oncogenic stress in cancer-free *path_MMR* carriers. Therefore, c-miRs hold potential in monitoring the LS risk stratification patterns during clinical surveillance or cancer management.

KEYWORDS
bioinformatics, hereditary cancer, Lynch syndrome, microRNA, next generation sequencing

Abbreviations: BP, biological process; c-miR, circulating microRNA; c-miRnome, global circulating miR expression profile; COSMIC-CGC, Cancer Gene Census of the Catalogue of Somatic Mutations in Cancer; DE, differential expression; dMMR, deficient MMR; ERMA, estrogenic regulation of muscle apoptosis; FDR, false discovery rate; GO, Gene Ontology; GSEA, gene set enrichment analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LS, Lynch syndrome; miR, microRNA; *path_MMR*, pathogenic mismatch-repair gene variant; t-SNE, T-distributed stochastic neighbor embedding.

Toni T. Seppälä and Eija K. Laakkonen contributed equally to our study.

Säätiö; Relander Foundation; Sigrid Juséliuksen Säätiö; Suomen Lääketieteen Säätiö; Terveyden Tutkimuksen Toimikunta, Grant/Award Numbers: 338657, 341750

What's new?

Systemic circulating microRNA expression patterns (c-miRnomes) are altered during sporadic carcinogenesis and they have predictive potential in early cancer detection. However, their potential in carriers of inherited pathogenic mismatch-repair gene variants associated with Lynch syndrome remains understudied. Using high-throughput sequencing and bioinformatics, the authors show that Lynch syndrome and sporadic carcinogenesis share common biological pathways. Alterations in these pathways produce a c-miRnome signature that could help track oncogenic stress in cancer-free Lynch syndrome carriers. The findings suggest that systemic c-miRnomes could potentially facilitate the monitoring of Lynch syndrome carriers that require more intensive surveillance or clinical management.

1 | INTRODUCTION

Lynch syndrome (LS) is an inherited cancer predisposition syndrome caused by pathogenic gene variants in DNA-mismatch repair (*path_MMR*) genes *MLH1*, *MSH2*, *MSH6* or *PMS2*.[1] By genetic or epigenetic silencing, deficient MMR (dMMR) significantly increases cellular mutation rates thus predisposing *path_MMR* carriers to increased cancer risk and excessive cancer occurrence.[1,2] Colorectal cancer is a traditional hallmark cancer of LS that is commonly cured by surveillance, followed by modern surgical and oncological management, with over 90% 10-year overall survival.[2,3] Despite the good recovery rate in first cancers, the persons at risk will develop frequently more lethal cancers still at relatively young age.[4] This highlights the need for an improved molecular assessment and identification of which patients would require more intensive surveillance or clinical management.

MicroRNAs (miRs) are small (18-25 nucleotides) noncoding RNA-molecules that regulate gene expression by translational repression.[5] MiRs play a role in regulation of >30% of the human genes controlling critical biological processes such as cell proliferation, cell differentiation, and apoptosis.[5-7] In cancers, miRs can be regarded as tumor suppressive or oncogenic, thus resulting in down-regulation or upregulation of the affected target genes, respectively.[7] Compared to tissue-based miRs, circulating-miRs (c-miRs) migrate throughout the body within various body fluids and are part of active intertissue crosstalk.[8,9] Nowadays, profiling of the global c-miR expression levels (c-miRnome) has become prevalent and miR expression can be correlated with cancer type, stage, and other clinical variables.[10-13] Therefore, aberrantly expressed miRs could have diagnostic, predictive, and prognostic potential in molecular profiling and early detection of cancers.

LS cohort provides an ideal population for biomarker mining due to well-predicted cancer risk of persons under frequent surveillance. The role of miRs in LS have remained understudied even if various studies have shown that c-miR expression patterns change with carcinogenesis in various sporadic cancers. Balaguer et al have shown that miRs can be used in tumor classification and discrimination of sporadic and hereditary tumors with microsatellite instability,[14] thus highlighting the potential role of miRs as LS biomarkers. In support, Valeri et al, Liccardo et al and Zhou et al postulated that miRs could have functional roles in LS carcinogenesis, for example, by targeting MMR-proteins[15,16] and various tumor-suppressor genes.[17] However, these studies along with other reports have assessed miR functions in the colorectum and colorectal cancer tissues and cells as well as with microarray data in silico[14-19] but not in circulation.

Instead of using a targeted panel of a priori chosen c-miRs, it is beneficial to characterize the systemic c-miRnome of *path_MMR* carriers. This "omics-approach" provides a more comprehensive view of how c-miRs could contribute to LS pathogenesis, and plausibly pave way for future use of c-miRs in risk stratification and early detection of LS cancers. Our exploratory study compared the systemic c-miRnome of cancer-free *path_MMR* carriers with c-miRnomes of non-LS controls (discovery cohort), sporadic rectal cancer patients and *path_MMR* carriers with cancer (cancer cohort) using high-throughput sequencing and bioinformatic approaches.

2 | MATERIALS AND METHODS

2.1 | Study subjects

Our study consisted of independent discovery and cancer cohorts. The discovery cohort (n = 118) was composed of 81 currently cancer-free (healthy) Finnish *path_MMR* carriers and 37 non-LS controls whose c-miRnomes were sequenced. The cancer cohort (n = 37) was composed of 13 *path_MMR* carriers who currently had cancer and 24 sporadic rectal cancer patients whose c-miRnomes were sequenced.

All *path_MMR* carriers were enrolled in the study and blood sampling was performed at their regular colonoscopy surveillance appointments at Helsinki University Central Hospital in Helsinki and Central Finland Central Hospital in Jyväskylä, Finland. They were also registered participants in the nationwide Finnish Lynch Syndrome Research Registry (LSRFi, www.lynchsyndrooma.fi, accessed 05/2021). The families and individuals were identified in the registry based on clinical criteria (Amsterdam and Bethesda criteria)[20,21] and subsequently through cascade testing of the families and universal testing of tumors. Adult members of LSRFi with confirmed *path_MMR* variants (classes 4 and 5 by InSiGHT criteria)[22] were eligible for the study.

Sporadic rectal cancer patients were enrolled, and blood sampling performed at the time of their initial appointment for surgery at surgical clinic at the local tertiary center responsible for management of

rectal cancer in the Southern Finland area (Helsinki University Central Hospital, unit of rectal surgery, Helsinki, Finland).

Non-LS control samples were acquired from Biobank of Eastern Finland, Kuopio, Finland (n = 27) in 2020 or were part of the Estrogenic Regulation of Muscle Apoptosis (ERMA) cohort (n = 10) consisting of healthy 47-55-years old women.[23] ERMA samples were collected at University of Jyväskylä in Jyväskylä, Finland. Persons with no cancers, blood disorders, acute or chronic infectious diseases, rheumatoid arthritis and known *BRCA* or MMR-gene germline mutations were eligible for the non-LS control group. Ethnicity throughout the study population was widely white Caucasian.

2.2 | Sample collection

Path_MMR carriers' and sporadic rectal cancer patients' venous blood samples were drawn after surveillance colonoscopy visits at fasted state. All ERMA participants fasted overnight before blood sampling. The duration of fasting is not reported for the samples obtained through biobank (n = 27). Samples were taken from antecubital vein to standard serum tubes (455 092, Greiner). To separate serum, the whole blood samples were allowed to clot for 30 minutes at room temperature, centrifuged at 1800*g* for 10 min and aliquoted.

2.3 | Small-RNA isolation and quality evaluation

c-miR isolations from blood serum were carried out using affinity column-based miRNeasy Serum/Plasma Advanced Kit (217204, Qiagen) according to the manufacturer's instructions. Briefly, 0.5 mL of thawed serum was used to isolate miRs. All the required solutions were added in amounts recommended by the manufacturer. Cel-miR-39 miR mimic (MS00019789, Qiagen) was added to each sample to serve as a spike-in control for monitoring the miR purification and amplification. Phase separation centrifugation was executed in 12 000*g* for 3 min at room temperature (Heraeus, Biofuge Pico and Fresco 17, ThermoFisher) and rest of the centrifugations were performed at 16000*g* whenever a range of 8000-20000*g* was recommended. C-miRs were eluted to nuclease-free water. Prior to the library preparation, RNA quality and recovery were checked by RT-qPCR (CFX96-RT-qPCR, Bio-Rad) according to manufacturer's protocol (MiScript Primer assays and II RT kit for cDNA synthesis and MiScript SYBR Green PCR Kit for RT-qPCR, 218 161, Qiagen) from which the recovery of cel-miR-39 spike-in control was confirmed.

2.4 | Small-RNA library preparation and sequencing

Small-RNA Library preparations were executed with QIAseq miRNA Library Preparation Kit (1103679, Qiagen) according to the manufacturer's instructions using multiplexing adapters. Briefly, the small RNA fractions were first ligated to sequencing adapters from both 5′ and 3′ ends, reverse transcribed into cDNA using UMI-assigning primers and purified using magnetic beads. A universal indexing sequence was also added in the reverse transcription step, thus allowing samples to be distinguished from each other. The samples were then amplified with standard thermocycler (Eppendorf), purified, and eluted into nuclease-free water. Quality assessment of the libraries was completed with TapeStation 4200 (Agilent). The library sample concentrations were measured with Qubit fluorometer (Invitrogen), quantified, diluted, and pooled into a single mixture in equal amounts (1.8 pM per sample) prior to sequencing. Sequencing of the small-RNA libraries were done with NextSeq 500 (Illumina) using NextSeq 500/550 High Output Kit v. 2.5 with 75 cycles (15057934, Illumina) to produce 75-base pair single-end reads with aimed mean sequencing depth of >5 M reads per sample as recommended by the manufacturer (Qiagen).

2.5 | Raw data processing and alignment

Sequencing output data was converted to FASTQ-format using bcl2fastq software (v.2.20, Illumina, USA). FastQC was used for quality controls.[24] The QIAseq sequencing adapters were trimmed from the 3′ end of the reads with FastX-toolkit[25] using default parameters with minimum alignment length-M 19. Only clipped reads >20 bp in length were selected for downstream analysis. After adapter clipping, the reads were trimmed to 22 bp to enrich miR-sequences and then quality filtered with FastX-toolkit. Only high-quality reads (Phred score >25) were selected for alignment to reference genome. Before alignment, all the four sample lanes were merged to obtain the overall sample read count and to ensure better mapping quality. Samples that had <1 M reads were excluded from the analyses. Subsequently, the preprocessed reads were mapped to human mature miR-genome (miRbase v.22)[26] with Bowtie alignment tool for single end data with v-mode and best strata parameters.[27] Only uniquely mapped miR-reads were selected for differential expression (DE) analysis.

2.6 | Differential expression analysis

DE analyses from raw c-miR counts were based on statistical procedures of EdgeR[28] and DESeq2[29] packages and conducted in R-studio (v. 3.6.3)[30] (Supplementary file S3). Briefly, DE analyses were performed on c-miR raw read count matrices after the low expressed genes were filtered out, normalized with the median of ratios method and variance stabilized in DESeq2. C-miRs that had more than 1 count per million in 70% of the samples in a group were selected for DE analyses. Filtered and normalized c-miR counts were used to set up a design matrix in DESeq2 that adjusted for sex and potential batch effect. Benjamini-Hochberg procedure in DESeq2 was used to correct for multiple testing. C-miRs that had a false discovery rate (FDR) <0.05 were considered DE.

2.7 | Dimension reduction analysis

Dimension reduction of the DESeq2-normalized data was conducted using the t-distributed stochastic neighbor embedding (t-SNE) method, which is a nonlinear and unsupervised technique to simplify high dimensional data for visualization in low-dimensional space.[31] t-SNE analysis was performed to identify and visualize possible clustering of subpopulations within the dataset. Rtsne package in R-studio was used with output dimensionality set to 2, perplexity set to 35 and theta set to 0.5.

2.8 | Target gene prediction and pathway analysis

Putative miR-target gene prediction was performed using mirWalk tool that utilizes a random-forest-based approach, an ensemble learning method based on multiple decision trees, to predict target genes.[32,33] Only the predicted miR-target genes targeting 3′ untranslated region with experimental validation from miRTarBase[34] and which were included and verified in mirDB[35] and TargetScan[36] databases were selected for downstream gene set enrichment analysis (GSEA).[37] GSEA of gene ontology biological processes (GO:BP) and Kyoto Encyclopedia of Genes and Genomes (KEGG)[38] pathways were also conducted with mirWalk. MirWalk provides a standard enrichment analysis based on hypergeometric tests. GO and KEGG terms with FDR-corrected *P*-values of <.05 were considered enriched. Cancer Gene Census of the Catalog of Somatic Mutations in Cancer (COSMIC-CGC)[39] project database were used for target gene investigation.

2.9 | Statistical analysis

Data regarding study subjects are presented using means and standard deviations. DE-analyses were based on statistical procedures of DESeq2 package accounting for normalization and exclusion of outliers. Mann-Whitney *U*-test and Kruskal-Wallis-test was used in the validation analysis and cell line experiment (Supplementary file S1, Supplementary materials and methods), respectively. Pearson correlation was used to compare gene fold correlation between the discovery and validation cohorts (Supplementary file S1, Supplementary materials and methods). In all analyses, *P*-value, or FDR <.05 were considered to indicate statistical significance.

3 | RESULTS

3.1 | A pool of 228 c-miRs is shared between the discovery and cancer cohorts

Descriptive characteristics of study subjects in the discovery cohort and cancer cohort are presented in Table 1.

Human genome encodes approximately 2600 mature miRs (miRbase, v.22).[26] To inspect the systemic c-miR content in the discovery and cancer cohorts, we performed small-RNA sequencing experiment to characterize the serum c-miRnomes. We identified a total of 1349 distinct c-miRs in three separate sequencing runs with an average sequencing depth of 3.2 M reads per sample (Supplementary file S1, Supplementary materials and methods and Supplementary file S1, Table S1 and Supplementary file S2, Table S1). After processing of raw data and filtering of low expressed c-miRs, 228 c-miRs common to both cohorts were identified (Supplementary file S1, Figure S1 and Supplementary file S2, Table S2).

The most highly expressed c-miRs among *path_MMR* carriers with or without cancer were hsa-let-7a-5p, hsa-let-7b-5p, hsa-miR-122-5p, hsa-miR-16-5p and hsa-mir-223-3p (Supplementary file S1, Figure S2). The most abundant c-miRs in non-LS control group were the same as in *path_MMR* carriers with or without cancer (Supplementary file S1, Figure S3). Among sporadic rectal cancer group, the top c-miRs were otherwise the same except hsa-miR-451a replaced hsa-miR-122-5p (Supplementary file S1, Figure S4). All these top c-miRs in total accounted for approximately 50% of all c-miR counts in all cohorts, thus displaying major overrepresentation that could have possibly affected the c-miR pool size. In summary, our sequencing analysis provided moderate coverage of c-miRnomes in LS.

3.2 | Healthy *path_MMR* carriers have a c-miRnome that differs from non-LS controls but resembles the c-miRnomes of patients with sporadic or hereditary cancer

The phenotype and cancer risk spectrum vary within LS cohort, for example, due to *path_MMR* variant and sex.[1] As our discovery cohort consisted of males and females with all *path_MMR* variants included, we first explored whether these traits influenced c-miR expression in healthy *path_MMR* carriers. We used the pool of identified 228 c-miRs to form the count matrix for all DE-analyses (Supplementary file S3). Hsa-miR-206 and hsa-miR-223-5p were observed downregulated in males compared to females and thus sex was added as a covariate to further analyses (Supplementary file S3). We did not find DE c-miRs when *path_MMR* variants were compared to each other or when *path_MLH1* carriers were compared to all other *path_MMR* variants combined (Supplementary file S3). These results show that different *path_MMR* variants do not cause heterogeneity that would generate a recognizable c-miR profile, thus suggesting a shared systemic response common to all *path_MMR* variants. Furthermore, we also tested if the c-miR expression profile is altered in persons who had had cancer or multiple cancers previously, but we did not find significant differences (Supplementary file S3).

Alterations in the immune cell abundance of normal colorectal mucosa in cancer-free *path_MMR* carriers separate them from those with cancer.[40] To see whether we can identify a LS-specific c-miR signature, our primary objective was to characterize systemic c-miRnome of healthy *path_MMR* carriers, which has not been done previously. We thus performed DE-analysis within the discovery cohort and RT-

TABLE 1 Descriptive characteristics of study subjects in the discovery cohort and cancer cohort

Variable	Discovery cohort		Cancer cohort	
	Path_MMR, healthy	non-LS, healthy	*Path_MMR*, cancer	Sporadic rectal cancer patients
N	81	37	13	24
Sex (N [%])				
Male	40 (49.4)	18 (48.6)	10 (76.9)	10 (41.6)
Female	41 (50.6)	19 (51.4)	3 (23.1)	14 (58.4)
Age, years (mean ± SD)	59.5 (10.7)	54.9 (10.7)	60.7 (15.3)	69.8 (9.9)
Body mass index, kg/m^2 (mean ± SD)[a]	27.3 (5.7)	28.0 (6.2)	28.2 (3.4)	27.6 (6.3)
Path_MMR (N [%])				
MLH1	50 (61.7)	–	8 (61.5)	–
MSH2	17 (21.0)	–	2 (15.4)	–
MSH6	12 (14.8)	–	3 (23.1)	–
PMS2	2 (2.5)	–	0 (0.0)	–
Previous cancers (N [%])				
Yes	42 (51.9)	–	10 (76.9)	–
No	39 (48.1)	–	3 (23.1)	–
Cancer type (N [%])				
Colorectal cancer	–	–	5 (38.5)	–
Prostate cancer	–	–	3 (23.0)	–
Other cancer[b]	–	–	5 (38.5)	–
Rectal cancer	–	–	–	24 (100.0)

[a]Missing data: Discovery cohort, n = 12 in *path_MMR* carriers; Cancer cohort, n = 3 in *path_MMR* carriers.
[b]Other cancer include esophageal cancer, n = 1; spinocellular cancer, n = 1; glioblastoma, n = 1; gastric cancer, n = 1 and thymic cancer, n = 1.

qPCR validation analysis within similar but independent validation cohort (Supplementary file S1, Supplementary materials and methods) to compare healthy *path_MMR* carriers to healthy non-LS controls (Supplementary file S1, Figure S5). In DE-analysis, we found 40 out of 228 c-miRs to display aberrant expression in healthy *path_MMR* carriers (Table 2). Of them, 15 were upregulated and 25 downregulated in *path_MMR* carriers compared to non-LS controls, but the fold changes remained low varying from minimum of −0.88 to maximum of 1.25 (Figure 1A). Hsa-miR-155-5p, hsa-let-7c-5p and -let-7 e-5p and -122b-3p had the most significant upregulation within healthy *path_MMR* carriers (Table 2). Of the downregulated c-miRs, hsa-miR-15a-5p was the most significantly downregulated followed by hsa-miR-185-5p, -320a-3p and -186-5p, respectively (Table 2). Overall, aberrant expression of multiple c-miRs in healthy *path_MMR* carriers might indicate that some systemic alterations in c-miR-mediated regulation of biological pathways associated with dMMR may be ongoing even at cancer-free state in *path_MMR* carriers.

To understand this phenomenon further, we explored whether the *path_MMR* carriers who currently have cancer also display unique c-miR expression. By using tumor samples, Balaguer et al have shown that miR expression can distinguish LS tumors from sporadic tumors with microsatellite instability.[14] To test if we can similarly reveal differences in c-miRs, we first inspected c-miRnomes within the cancer cohort but did not find any differences (Figure 1B and Supplementary file S1, Table S2), thus suggesting a mutual c-miR response among the cancer types. Furthermore, our second analysis scheme comparing healthy *path_MMR* carriers to sporadic rectal cancer patients (Figure 1C and Table 2), our third analysis scheme comparing healthy *path_MMR* carriers to *path_MMR* carriers with cancer (Figure 1D and Table 2) and our fourth analysis scheme comparing *path_MMR* carriers with cancer to healthy non-LS controls (Figure 1E and Supplementary file S1, Table S2) were also unable to detect DE c-miRs. These observations imply that c-miRnomes within our dataset cannot discern healthy *path_MMR* carriers from cancer patients with or without dMMR.

Several DE c-miRs have been implicated to sporadic cancer progression.[41,42] To study this in our dataset, we compared sporadic rectal cancer patients to non-LS controls. We found that hsa-miR-200a-3p, -10a-5p, -196a-5p and -200c-3p were significantly upregulated in sporadic rectal cancer patients differentiating them from non-LS controls (Figure 1F and Table 2). All of these c-miRs have earlier been shown to associate with colorectal cancer, and of them, hsa-miR-200a-3p was also significantly upregulated in healthy *path_MMR* carriers compared to non-LS controls with fold change of 0.88. In this analysis scheme, the fold change in hsa-miR-

TABLE 2 DE and non-DE c-miRs within and between the discovery and cancer cohorts

Healthy *path_MMR* vs non-LS control			Sporadic rectal cancer patients vs healthy *path_MMR*			Healthy *path_MMR* vs *path_MMR* with cancer			Sporadic rectal cancer patients vs non-LS control		
c-miR	log2FC	FDR	c-miR	log2FC	FDR	c-miR	log2FC	FDR	c-miR	log2FC	FDR
hsa-miR-155-5p	0.905	**<0.001**	hsa-miR-10a-5p	0.700	0.088	hsa-miR-127-3p	1.548	0.277	hsa-miR-200a-3p	1.755	**<0.001**
hsa-let-7c-5p	0.729	**<0.001**	hsa-miR-1180-3p	1.147	0.155	hsa-let-7b-5p	−0.250	0.991	hsa-miR-10a-5p	0.981	**0.003**
hsa-let-7 e-5p	0.955	**<0.001**	hsa-miR-126-3p	−0.395	0.155	hsa-let-7c-5p	0.191	0.991	hsa-miR-196a-5p	1.813	**0.028**
hsa-miR-122b-3p	1.252	**0.001**	hsa-miR-148b-3p	−0.336	0.155	hsa-let-7d-3p	−0.146	0.991	hsa-miR-200c-3p	1.133	**0.028**
hsa-miR-15a-5p	−0.677	**0.001**	hsa-miR-196a-5p	1.414	0.155	hsa-let-7d-5p	−0.325	0.991			
hsa-miR-185-5p	−0.483	**0.001**	hsa-miR-320a-3p	0.557	0.155	hsa-let-7 e-5p	0.331	0.991			
hsa-miR-320a-3p	−0.709	**0.001**	hsa-miR-320b	0.845	0.155	hsa-let-7f-5p	0.174	0.991			
hsa-miR-186-5p	−0.548	**0.002**	hsa-miR-486-5p	0.542	0.243	hsa-let-7i-5p	0.067	0.991			
hsa-let-7a-5p	0.535	**0.003**	hsa-miR-320c	1.097	0.262	hsa-miR-100-5p	−0.453	0.991			
hsa-miR-10b-5p	0.500	**0.003**	hsa-miR-185-5p	0.344	0.288	hsa-miR-101-3p	0.171	0.991			
hsa-miR-3613-5p	−0.880	**0.003**	hsa-miR-223-3p	0.413	0.288	hsa-miR-103a-3p	0.246	0.991			
hsa-miR-22-3p	−0.522	**0.004**	hsa-miR-483-5p	0.774	0.319	hsa-miR-103b	0.239	0.991			
hsa-miR-19b-3p	−0.490	**0.005**	hsa-miR-2110	1.198	0.342	hsa-miR-106b-3p	−0.216	0.991			
hsa-miR-125a-5p	0.490	**0.007**	hsa-miR-222-3p	0.750	0.342	hsa-miR-106b-5p	0.535	0.991			
hsa-miR-451a	−0.714	**0.007**	hsa-miR-486-3p	0.475	0.462	hsa-miR-107	0.243	0.991			
hsa-miR-125b-5p	0.600	**0.009**	hsa-let-7d-3p	0.462	0.462	hsa-miR-10a-5p	−0.194	0.991			
hsa-miR-15b-5p	−0.525	**0.009**	hsa-miR-11 400	−0.824	0.462	hsa-miR-10b-5p	−0.170	0.991			
hsa-miR-32-5p	−0.564	**0.009**	hsa-miR-134-5p	−0.670	0.462	hsa-miR-11 400	0.549	0.991			
hsa-miR-339-5p	−0.806	**0.009**	hsa-miR-193a-5p	0.532	0.462	hsa-miR-1180-3p	−0.864	0.991			
hsa-miR-107	−0.464	**0.012**	hsa-miR-196b-5p	0.447	0.462	hsa-miR-1255b-5p	0.373	0.991			
hsa-miR-484	−0.748	**0.012**									
hsa-let-7f-5p	0.328	**0.015**									
hsa-miR-206	0.994	**0.015**									
hsa-miR-25-3p	−0.375	**0.015**									
hsa-miR-27a-3p	−0.373	**0.015**									
hsa-miR-486-3p	−0.565	**0.015**									
hsa-miR-141-3p	0.874	**0.016**									
hsa-miR-3074-5p	−0.537	**0.020**									
hsa-miR-126-3p	0.328	**0.021**									
hsa-miR-200a-3p	0.884	**0.021**									
hsa-miR-221-3p	−0.312	**0.033**									
hsa-miR-424-5p	−0.662	**0.034**									
hsa-let-7i-5p	0.275	**0.040**									
hsa-miR-23a-3p	−0.437	**0.040**									
hsa-miR-27b-3p	0.420	**0.040**									
hsa-miR-486-5p	−0.447	**0.040**									
hsa-miR-19a-3p	−0.441	**0.046**									
hsa-miR-222-3p	−0.647	**0.046**									
hsa-miR-363-3p	−0.537	**0.049**									
hsa-miR-92a-3p	−0.370	**0.050**									

Note: *N*, healthy *path_MMR* = 81; *N*, *path_MMR* with cancer = 13; *N*, sporadic rectal cancer patients = 24; *N*, non-LS controls = 37. FDR <0.05 highlighted with bold.
Abbreviations: c-miR, circulating microRNA; FDR, false discovery rate; log2FC, logarithmic2 fold change.

200a-3p was 1.76, indicating significantly higher expression compared to the healthy non-LS controls (Table 2).

Taken together, our findings imply that healthy *path_MMR* carriers have a systemic c-miRnome that separates them from healthy non-LS persons but resemble the c-miRnome of cancer patients with or without dMMR. Thus, these findings suggest that sporadic and dMMR-directed carcinogenesis share common miR-targeted biological pathways where potential alterations may produce a detectable c-miR signature in the healthy *path_MMR* carriers.

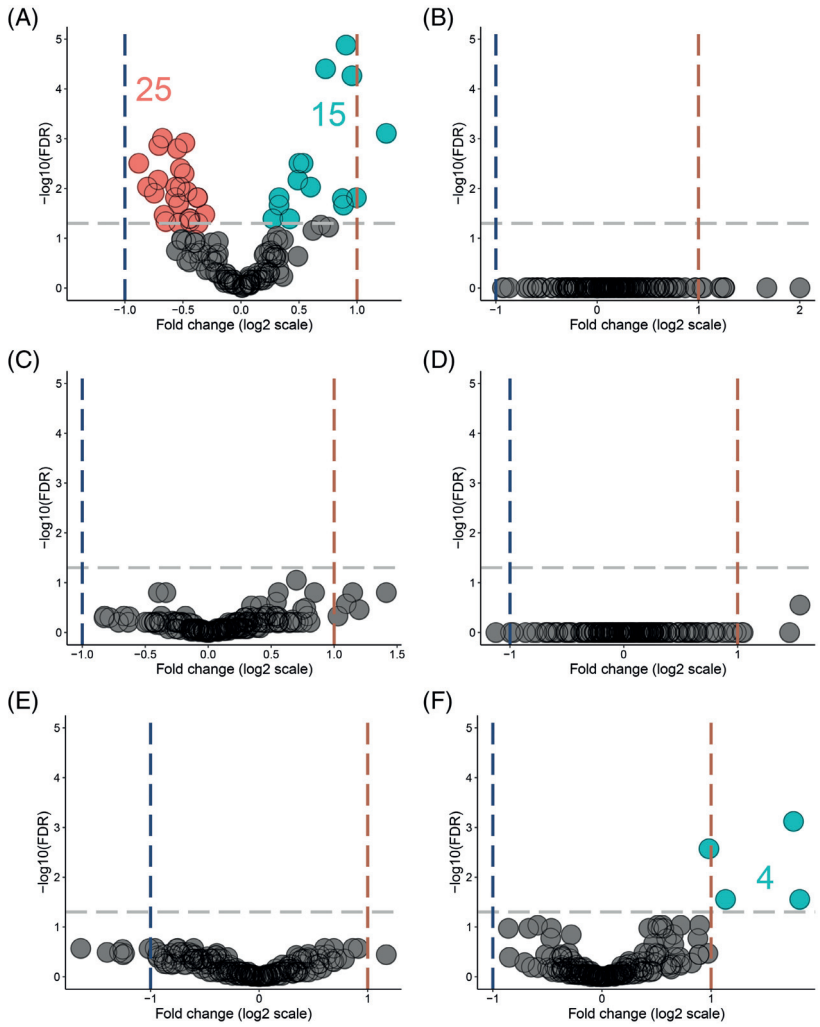


FIGURE 1 Healthy *path_MMR* carriers have a c-miRnome that differ from non-LS controls but resembles the c-miRnomes of patients with sporadic or hereditary cancer. (A) DE c-miRs in healthy *path_MMR* carriers vs non-LS controls. (B) DE c-miRs in sporadic rectal cancer patients vs *path_MMR* carriers with cancer. (C) DE c-miRs in healthy *path_MMR* carriers vs sporadic rectal cancer patients. (D) DE c-miRs in healthy *path_MMR* carriers vs *path_MMR* carriers with cancer. (E) DE c-miRs in *path_MMR* carriers with cancer vs non-LS controls. (F) DE c-miRs in sporadic rectal cancers patients vs non-LS controls. Blue dash lines indicate negative fold change of expression, red dash line indicate positive fold change of expression and gray dash line indicate FDR <0.05. Downregulated c-miRs are highlighted in red, upregulated c-miRs are highlighted in cyan and nonsignificantly expressed c-miRs are highlighted in gray. Dots represents c-miRs. c-miR, circulating microRNA; FDR, false discovery rate

3.3 | Dimension reduction analysis of multiple traits was unable to discern *path_MMR* carriers from sporadic rectal cancer patients

We did not identify DE c-miRs between *path_MMR* carriers and sporadic rectal cancer patients. Therefore, by using the expression data of all 228 c-miRs shared between the discovery and cancer cohorts, we performed a dimension reduction analysis with t-SNE to identify possible subpopulations within *path_MMR* carriers and sporadic rectal cancer patients. First, we investigated if phenotypic traits such as being *path_MMR* carrier, current cancer status, cancer history or *path_MMR* variant type, would reveal clustering of samples, but did not find any clear patterns (Figure 2A-D). We also investigated if age or BMI would be the discerning traits, but they also failed to reveal any clustering (Figure 2E-G). Finally, sex and the sequencing batch did not form clusters within our dataset (Figure 2H,I). Taken together, the t-SNE analysis supported the DE c-miR findings and was not able to differentiate *path_MMR* carriers from sporadic cancer patients, which may be an indicative of shared c-miR-mediated regulation as seen in the DE-analyses.

3.4 | Pathway analysis revealed putative c-miR-target genes that are linked to biological processes and pathways associated with cancer

To further evaluate our hypothesis that healthy *path_MMR* carriers might have a c-miRnome that resembles the c-miRnome of cancer cohort due to shared miR-targeted biological pathways, we next investigated what are the target genes of the observed DE c-miRs. We also inspected what biological processes and pathways these target genes associate with. With mirWalk, we used random-forest-based approach to predict the target genes using databases with experimental validation and high confidence of reported miR-target gene interactions. MirWalk identified a total of 1731 miR-target gene interactions with 508 distinct putative target genes for 32 out of

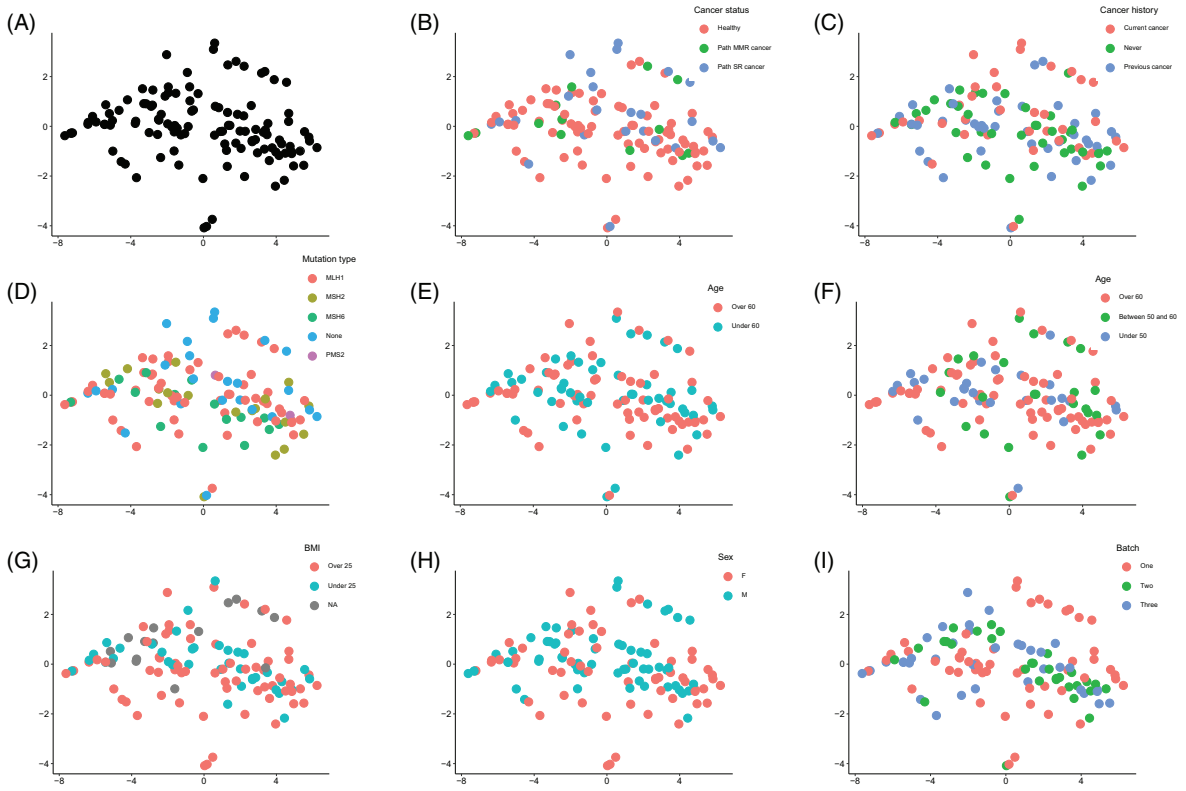


FIGURE 2 Dimension reduction analysis of multiple traits was unable to discern *path_MMR* carriers from sporadic rectal cancer patients. (A) *Path_MMR* carriers and sporadic rectal cancer patients. (B) Cancer status. Healthy, cancer-free *path_MMR* carriers; *path_MMR* cancer, *path_MMR* carriers with cancer; path SR cancer, sporadic rectal cancer patients. (C) Cancer history. Current cancer, has cancer currently; Never, currently healthy, never had cancers; Previous cancer, currently healthy, had had cancer or multiple cancers; (D) *path_MMR* variant. (E) Dichotomous age. Over 60, persons >60-years of age; Under 60, persons <60-years of age. (F) Nondichotomous age. Over 60, persons >60-years of age; Between 50 and 60, persons between 50 and 60-years of age; Under 50, persons <50-years of age. (G) BMI. Over 25, persons with BMI > 25; Under 25, persons with BMI < 25; NA, no reported BMI. (H) Sex. M, males; F, females. (I) Batch effect of three separate sequencing runs in running order. All t-SNE plots are 2D constructions. Dots represent study subjects. BMI, body mass index

40 observed DE c-miRs from discovery cohort analysis (Supplementary file S2, Tables S3 and S4).

We then performed mirWalk-GSEA analysis on the 508 predicted target genes to explore what functional roles the DE c-miRs might possess. The GSEA analysis revealed 195 distinct significantly enriched biological processes (Supplementary file S2, Tables S5 and S6). To identify the key biological processes, we then narrowed the given output list based on FDR and the number of involved target genes to the top 30 most significantly enriched biological processes (Supplementary file S2, Table S7). Most of the discovered biological processes were linked to apoptosis, regulation of transcription, cell cycle, cell proliferation, DNA damage and signal transduction (Figure 3A). We then conducted a small-scale cell line experiment to investigate how c-miR over- and underexpression affect Human colorectal cell line (HCT116) viability (Supplementary file S1, Supplementary materials and methods). We chose hsa-miR-122b and -451a as representatives of over- and underexpressed miRs found in healthy *path_MMR* carriers vs non-LS control comparisons. HCT116 cell line was chosen to mimic LS colorectal cancer. The cell line experiment hinted that overexpression of hsa-miR-122b could reduce cell viability via increased apoptosis whereas underexpression of hsa-miR-451a also resulted in reduced viability but did not induce apoptosis of HCT116 cells (Supplementary file S1, Figure S6). We observed considerable overlap between the identified pathways since 208 out of 508 identified distinct c-miR-target genes contributed to the top biological processes (Supplementary file S2, Table S8). *TGFBR1*, *CDKN1A*, *IGF1*, *TRAF6* and *BCL2* genes were present in most of the observed biological processes along with several other genes (Supplementary file S2, Table S8). The performed in silico target analysis showed that *TGFBR1* was targeted by hsa-miR-27b-3p, *CDKN1A* and *IGF1* were targets of hsa-let-7 e-5p, *TRAF6* was targeted by hsa-miR-125a-5p and *BCL2* was targeted by hsa-miR-125b-5p and hsa-miR-15b-5p (Supplementary file S2, Table S3). Of these c-miRs, all

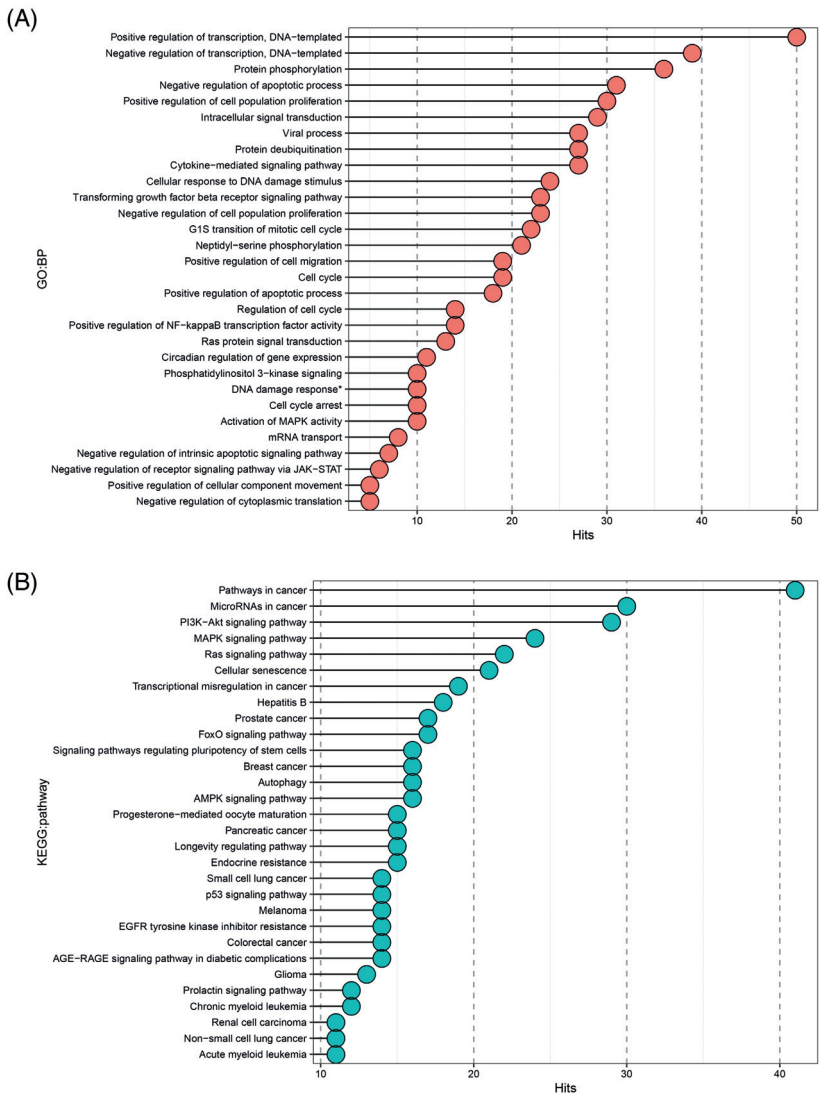


FIGURE 3 Pathway analysis revealed putative c-miR-target genes that are linked to biological processes and pathways associated with cancer. (A) Top 30 most enriched biological processes annotated to the identified target genes of 32 out of 40 DE c-miRs found in healthy *path_MMR* carriers. FDR, false discovery rate; GO:BP, Gene Ontology: biological process; Hits, number of target genes annotated to the biological process. *Signal transduction by p53 class mediator resulting in cell cycle arrest. (B) Top 30 most enriched KEGG pathways annotated to the identified target genes of 32 out of 40 DE c-miRs found in healthy *path_MMR* carriers. c-miR, circulating microRNA; FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes pathway; Hits, number of target genes annotated to the pathway

except hsa-miR-15b-5p were upregulated in *path_MMR* compared to controls (Table 3).

Next, we explored how the c-miR-target genes interact with KEGG pathways. GSEA analysis of the same gene set discovered 88 significantly enriched KEGG biological pathways (Supplementary file S2, Tables S9 and S10). Again, to focus on the possible key pathways, we narrowed the output list to the top 30 of the most significant pathways based on similar parameters than in the previous analysis (Supplementary file S2, Table S11). A great majority of the discovered pathways linked to cancer, cancer signaling and cell aging (Figure 3B). Of the 508 predicted target genes, 113 were involved in the discovered top KEGG pathways (Supplementary file S2, Table S12). *AKT3*, *PIK3R1* and *PIK3CA* genes were involved in 27 out of 30 KEGG pathways, whereas *KRAS* had 26 and *NRAS* had 24 hits, respectively (Supplementary file S2, Table S12). *AKT3* was targeted by hsa-miR-15b-5p, *PIK3R1* was targeted by hsa-miR-107 and hsa-miR-486-5p, *PIK3CA* was targeted by hsa-miR-19a-3p, *KRAS* was targeted by hsa-miR-27a-3p and *NRAS* was a target of hsa-let-7a and -7c-5p (Supplementary file S2, Table S3). Of these c-miRs, all except hsa-let7a and -7c, were downregulated in *path_MMR* compared to controls (Table 3).

As these key target genes were interacting in the majority of the identified cancer-associated biological processes and pathways, we then explored and validated their potential carcinogenic roles. We submitted the gene set to COSMIC-CGC database and found that *BCL2*, *AKT3*, *PIK3CA*, *KRAS* and *NRAS* possess oncogenic functions, whereas *CDKN1A* is a potential oncogene or tumor suppressor gene and *PIK3R1* functions as a tumor suppressor gene (Table 3). All these genes have well-documented roles in multiple tumor types, including colorectal cancer, and with most having functions in hallmarks of cancer.[43] Of the target gene set, *TGFBR1*, *IGF1* nor *TRAF6* were not included in COSMIC-CGC database. These results support our

hypothesis that the observed resemblance of the c-miRnomes between *path_MMR* carriers and sporadic rectal cancer patients can be due to shared biological processes and pathways that include well-known oncogenes and tumor-suppressor genes.

Taken together, our in silico analysis shows that the c-miRs in hsa-let-7 family, as well as hsa-miR-15b-5p, hsa-miR-19a-3p, hsa-miR-27a-3p and -27b-3p, hsa-miR-107, hsa-miR-125b-5p and hsa-miR-486-5p could target genes that are ubiquitous in cancer-associated biological processes and pathways. These findings imply that the altered c-miRnome expression pattern of cancer-free *path_MMR* carriers may hold predictive value by tracking potential oncogenic stress caused by dMMR-driven distortions.

4 | DISCUSSION

Our study pioneered in characterizing the systemic c-miRnomes of *path_MMR* carriers. By utilizing high throughput sequencing, a total of 228 distinct c-miRs common to all study subjects were detected. Of these, we showed healthy *path_MMR* carriers to have an exclusive c-miRnome of 40 DE c-miRs that differs from non-LS-controls, but that does not differ from the c-miRnome of cancer patients with or without dMMR. Our c-miR expression analysis combined with in silico tools revealed that the observed resemblance in the c-miRnomes is possibly caused by distortions in several biological networks that are governed by well-known oncogenes and tumor suppressor genes, thus suggesting that c-miRnome could be used to track potential oncogenic stress at cancer-free state.

There is a growing interest in exploiting miRs as cancer biomarkers. Balaguer et al studied miRs that were extracted from tumors of *path_MMR* carriers and sporadic colorectal cancer patients with verified microsatellite instability and normal tissue samples.[14,18] They used a set of >700 miR-probes with microarray analysis and detected hundreds of DE miRs among the tissue samples, showing that LS tumors can be separated from sporadic tumors with microsatellite instability, as well as that suspected LS samples discern from confirmed LS samples. Aligned with their study, we also showed that different *path_MMR* variants do not display unique c-miR expression thus implying a shared systemic response. However, we could not pinpoint DE c-miRs that would distinguish *path_MMR* carriers from sporadic cancer patients although we did, as well as in numerous other studies, detect a c-miR signature unique to sporadic cancer patients when compared to healthy non-LS controls. The observation that *path_MMR* carriers do not differ from sporadic cancer patients in their c-miRnome was also supported by our t-SNE analysis that did not reveal any clustering within our dataset based on several variables. The reason behind the substantial difference in DE c-miR numbers between our and the study by Balaguer et al is likely explained by the study setting, used specimen type and methodology. In our study, the DE c-miRs were sequenced from the circulation of cancer-free persons where such a robust c-miR signature is not presumably detected when compared to miRs at the site of pathology.

Furthermore, Balaguer et al detected several DE miRs with diagnostic potential in LS, including hsa-miR-125b-5p, -137, -622, -192 and -1238, whereas Zhou et al displayed that hsa-miR-137, -520 e and -590-3p are indicatives of LS by using a subset of *path_MMR* cancer tumor samples and normal tissue samples from the study by Balaguer et al.[17] We did not find significant overlapping of DE miR content between our c-miRs and tumor-miRs from those studies, except for hsa-miR-125b-5p, that was also identified by Balaguer et al. Aberrant expression of hsa-miR-125b-5p has been reported for multitude of cancer types and it has been implied to serve as a circulating cancer biomarker by targeting apoptosis-regulating oncogene *BCL2*.[44]

The most significant DE c-miR in our setting was hsa-miR-155-5p, followed by hsa-let-7c-5p and -7 e-5p, -122b-3p and 15a-5p, which all except hsa-miR-15a-5p were upregulated in healthy *path_MMR* carriers. Valeri et al demonstrated that hsa-miR-155-5p targets several MMR-genes and that overexpression of hsa-miR155-5p downregulates *MLH1* and *MSH2* in colorectal cancer cell lines.[15] Within this concept, our DE findings also support the role of hsa-miR-155p modulation in LS pathogenesis even though the performed in silico analysis could not identify MMR-genes as targets of hsa-miR-155-5p. miRs in hsa-let-7 family have been suggested to increase colorectal cancer risk in *path_MMR* carriers with proficient MMR by lowering the expression of *TGFBR1* haplotype.[45] We found hsa-let-7 family to target *TGFBR3* and hsa-miR-27b-3pto target *TGFBR1*. We did not find experimentally verified target genes for hsa-miR-122b-3p. However, we could see that overexpression of hsa-miR-122b might result in reduced cell viability, plausibly due to increased apoptosis. Previous studies have linked hsa-miR-15a-5p to sporadic endometrial cancer[46] and colorectal cancer,[47] both being hallmark cancers of LS. In our analysis, hsa-miR-15a-5p was seen to target several genes, including known oncogenes and tumor suppressor genes such as *CCND1*, *CDK6* and *DICER1*, thus suggesting biomarker potential also in LS.

MiRs have critical functions across various biological processes and pathways involved in carcinogenesis. We found 508 putative target genes for 32 out of 40 observed DE c-miRs that associate with several pathways common to cancer. In addition to above mentioned c-miRs, we also identified several other c-miRs that could be key regulators in dMMR-driven carcinogenesis. The performed in silico analysis indicated that all these c-miRs target several well-known oncogenes and tumor suppressor genes such as *KRAS*, *NRAS*, *PIK3RI*, and *PIK3CA*, that were significantly enriched in our pathway analysis. Supported by our DE-analysis, the observation that these identified DE c-miRs target known oncogenes and tumor suppressors, could indicate upregulation of the oncogenes and consequently downregulation of the tumor suppressor genes. However, since we studied cell-free c-miRs without possibility to investigate expression levels of their putative target genes, this suggestion remains hypothetical. Unfortunately, c-miRs are not easily tracked where tracking of c-miRs would provide us clues to what tissues they will be affecting and where to seek further signs of cancer development. Matching pairwise tissue samples to observed c-miRs could help elucidate these issues but we

TABLE 3 Key target genes of DE c-miRs in healthy *path_MMR* carriers compared to non-LS controls

Key target gene	Gene name	Hits	COSMIC-CGC Role in cancer	c-miR
GO:BP				
TGFBR1	Transforming growth factor-beta receptor type 1	10	NA	hsa-miR-27b-3p ↑
CDKN1A	Cyclin dependent kinase inhibitor 1A	8	Oncogene, tumor suppressor gene	hsa-let-7 e-5p ↑
IGF1	Insulin growth factor 1	7	NA	hsa-let-7 e-5p ↑
TRAF6	TNF receptor-associated factor 6	7	NA	hsa-miR-125a-5p ↑
BCL2	B-cell CLL/lymphoma 2	6	Oncogene, fusion	hsa-miR-125b-5p ↑ hsa-miR-15b-5p ↓
KEGG				
AKT3	V-akt murine thymoma viral oncogene homolog 3	27	Oncogene	hsa-miR-15b-5p ↓
PIK3R1	Phosphoinositide-3-kinase. Regulatory subunit 1 (alpha)	27	Tumor suppressor gene	hsa-miR-107 ↓ hsa-miR-486-5p ↓
PIK3CA	Phosphoinositide-3-kinase. Catalytic. alpha polypeptide	27	Oncogene	hsa-miR-19a-3p ↓
KRAS	KRAS Proto-Oncogene, GTPase	26	Oncogene	hsa-miR-27a-3p ↓
NRAS	NRAS Proto-Oncogene, GTPase	24	Oncogene	hsa-let-7a-5p ↑ hsa-let-7c-5p ↑

Note: Arrows indicate up- (↑) or downregulation (↓) of c-miR in DE-analysis. Hits indicate the number of top GO:BP or KEGG-pathways where the gene is present.

Abbreviations: c-miR, circulating microRNA; COSMIC-CGC, The Catalogue of Somatic Mutations in Cancer and Cancer Gene Census database; GO:BP, Gene Ontology:biological process; KEGG, Kyoto Encyclopedia of Genes and Genomes pathway.

had no possibility to do so. Nevertheless, our exploratory findings indicate that *path_MMR* carriers display oncogenic stress even when they are cancer-free, but more studies are needed to verify our results and to show if they have true power as a biomarkers of early cancer development. A future goal is to determine whether the longitudinal change or development of c-miRnomes appears in conjunction with cancer incidence and treatment. The biological basis for aberrant c-miR expression between *path_MMR* carriers and non-LS controls remains a clinical question to be elucidated also in the future work.

A major strength of our study is that the study subjects had undergone comprehensive screenings of LS-predisposing mutations, with ascertainment utilizing Amsterdam and Bethesda clinical criteria and cascade testing. Also, instead of a priori chosen gene panel, we conducted a systemic level investigation of c-miRnome, which provides a more comprehensive view of how already identified c-miRs and putative target genes contribute to distorted biological networks in sporadic and hereditary cancer. For example, our findings allow construction of c-miRnome-target gene collection to be explored for potentially distorted biological networks associated with dMMR. Also, it can be used for establishing candidate hypotheses to drive further research and for further exploratory c-miR analyses of potential contributing gene clusters not previously discovered. Finally, the bioinformatic analyses in our study were performed in precise detail according to the latest knowledge using state-of-the-art tools and algorithms.

Our study has potential pitfalls. Although largest to date, the study sample was relatively small especially in the cancer cohort, which could have reduced the statistical power of DE-analyses. Regarding the methodology, there is no conclusive rule which sequencing depth should be aimed at when assessing DE of c-miRs. In our study, the aimed mean sequencing depth was 5 M reads per sample, but the achieved mean sequencing depth was 3.2 M reads due to underclustering issues in sequencing. The underclustering might have affected c-miR detection by favoring highly expressed c-miRs and thus resulting in overpresentation of these c-miRs and underpresentation or masking of c-miRs with low expression and potential cancer- or dMMR-relevant functions. A common issue with c-miRs is the identification of their primary and target locations, and alike in many other studies, we did not track the observed c-miRs to certain locations, which introduce a certain degree of uncertainty over the interpretations of the observations. Unfortunately, our efforts to validate DE findings with RT-qPCR were not completely successful when using an independent validation cohort, although we observed a trend of parallel expression in both cohorts in eight out of nine validation c-miRs. Overall considerable variation in c-miR expression levels were detected with both methods and cohorts, which could explain why significant differences between groups in the smaller validation cohort were not detected. Furthermore, we cannot completely exclude the possibility that varying ascertainment site for sample collection may have increased between sample variation and could thereby have affected our analyses.

To conclude, our exploratory study was the first to characterize the systemic c-miRnomes of *path_MMR* carriers. We showed that systemic c-miRnome can be used to track potential oncogenic stress in cancer-free *path_MMR* carriers thus paving way for the future investigation of c-miRs in monitoring the risk stratification patterns during the risk-reducing clinical surveillance and possible cancer

management. Our study also produced novel insight that allows construction of a c-miRnome-target gene collection to be explored for potentially distorted biological networks and c-miRnome-target gene interactions in LS.

AUTHOR CONTRIBUTIONS

Tero Sievänen: Formal analysis, investigation, methodology, software, validation, visualization, writing original draft, writing-review and -editing. **Tia-Marje Korhonen**: Formal analysis, methodology, software, visualization, writing-review and -editing. **Tiina Jokela**: Methodology, software, validation, visualization. **Maarit Ahtiainen**: Methodology, writing-review and -editing. **Laura Lahtinen**: Methodology, writing-review and -editing. **Teijo Kuopio**: Methodology, writing-review and -editing. **Anna Lepistö**: Writing-review and -editing. **E. Sillanpää**: Conceptualization, supervision, writing-review and -editing. **Jukka-Pekka Mecklin**: Conceptualization, resources, writing-review and -editing. **Toni T. Seppälä**: Conceptualization, funding acquisition, resources, supervision, writing-review and -editing. **Eija K. Laakkonen**: Conceptualization, data curation, funding acquisition, project administration, resources, supervision, writing-review and -editing. The work reported in the paper has been performed by the authors, unless clearly specified in the text.

ACKNOWLEDGEMENTS

We acknowledge the support from Biosciences team of IT Center for Science Finland (CSC) for providing HPC-resources for our data analytics. We would also like to thank research assistant Minta Kärkkäinen for her valuable help with the cell line experiments.

FUNDING INFORMATION

Eija K. Laakkonen was supported by the Päivikki and Sakari Sohlberg Foundation. Elina Sillanpää was supported by the Academy of Finland research fellowship (grant number: 341750). Toni T. Seppälä was supported by Finnish Medical Foundation, Emil Aaltonen Foundation, Jane and Aatos Erkko Foundation, Sigrid Juselius Foundation, Finnish Cancer Foundation, Relander Foundation, Academy of Finland (grant number: 338657), HUS State Research Funds (TYH2021123 and TYH2022323) and iCAN Digital Precision Medicine Flagship. Jukka-Pekka Mecklin was supported by Jane and Aatos Erkko Foundation, Finnish Cancer Foundation and KYS State Research Funds.

CONFLICT OF INTEREST

Toni T. Seppälä declares being CEO and co-owner of HealthFund Finland and Consultation fees from Boehringer Ingelheim and Amgen Finland. The other authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The datasets supporting the conclusions of this article are available in the Gene Expression Omnibus (GSE198834). Other data that support the findings of our study are available from the corresponding author upon request. Step-by-step analysis protocols can be accessed via GitHub repository (https://zenodo.org/badge/latestdoi/467491700).

ETHICS STATEMENT

Informed consent was obtained from all participants, and the Helsinki and Uusimaa Health Care District (HUS/155/2021) and Central Finland Health Care District Ethics Committee (KSSHP D# 1U/2018 and 1/2019 and KSSHP 3/2016) approved the study protocol. The study was conducted according to the guidelines of the Declaration of Helsinki.

ORCID

Tero Sievänen https://orcid.org/0000-0002-9660-4559

TWITTER

Tero Sievänen @SievanenTero
Eija K. Laakkonen @eija_laakkonen

REFERENCES

1. Dominguez-Valentin M, Sampson JR, Seppälä TT, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the prospective Lynch syndrome database. *Genet Med*. 2020;22:15-25.
2. Møller P, Seppälä T, Bernstein I, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. *Gut*. 2017;66:464-472.
3. Renkonen-Sinisalo L, Seppälä TT, Järvinen HJ, Mecklin JP. Subtotal colectomy for colon cancer reduces the need for subsequent surgery in Lynch syndrome. *Dis Colon Rectum*. 2017;60:792-799.
4. Møller P, Seppälä T, Bernstein I, et al. Incidence of and survival after subsequent cancers in carriers of pathogenic MMR variants with previous cancer: a report from the prospective Lynch syndrome database. *Gut*. 2017;66:1657-1664.
5. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006;6:857-866.
6. Jung G, Hernández-Illán E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. *Nat Rev Gastroenterol Hepatol*. 2020;17:111-130.
7. Goodall GJ, Wickramasinghe VO. RNA in cancer. *Nat Rev Cancer*. 2021;21:22-36.
8. Chen X, Ba Y, Ma L, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res*. 2008;18:997-1006.
9. Mori MA, Ludwig RG, Garcia-Martin R, Brandão BB, Kahn CR. Extracellular miRNAs: from biomarkers to mediators of physiology and disease. *Cell Metab*. 2019;30:656-673.
10. Francavilla A, Turoczi S, Tarallo S, Vodicka P, Pardini B, Naccarati A. Exosomal microRNAs and other non-coding RNAs as colorectal cancer biomarkers: a review. *Mutagenesis*. 2020;35:243-260.
11. Muinelo-Romay L, Casas-Arozamena C, Abal M. Liquid biopsy in endometrial cancer: new opportunities for personalized oncology. *Int J Mol Sci*. 2018;19:19.
12. He Y, Deng F, Yang S, et al. Exosomal microRNA: a novel biomarker for breast cancer. *Biomark Med*. 2018;12:177-188.
13. Hu C, Meiners S, Lukas C, Stathopoulos GT, Chen J. Role of exosomal microRNAs in lung cancer biology and clinical applications. *Cell Prolif*. 2020;53:e12828.
14. Balaguer F, Moreira L, Lozano JJ, et al. Colorectal cancers with microsatellite instability display unique miRNA profiles. *Clin Cancer Res*. 2011;17:6239-6249.
15. Valeri N, Gasparini P, Fabbri M, et al. Modulation of mismatch repair and genomic stability by miR-155. *Proc Natl Acad Sci*. 2010;107:6982-6987.

16. Liccardo R, Sessa R, Trombetti S, et al. Mir-137 targets the 3′ untranslated region of msh2: potential implications in lynch syndrome-related colorectal cancer. *Cancer*. 2021;13:1-12.
17. Zhou C, Li J, Li J, et al. Hsa-miR-137, hsa-miR-520 e and hsa-miR-590-3p perform crucial roles in lynch syndrome. *Oncol Lett*. 2016;12:2011-2017.
18. Balaguer F, Link A, Lozano JJ, et al. Epigenetic silencing of miR-137 is an early event in colorectal carcinogenesis. *Cancer Res*. 2010;70:6609-6618.
19. Pavicic W, Perkiö E, Kaur S, Peltomäki P. Altered methylation at microRNA-associated CpG islands in hereditary and sporadic carcinomas: a methylation- specific multiplex ligation-dependent probe amplification (MS-MLPA)-based approach. *Mol Med*. 2011;17:726-735.
20. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst*. 2004;96:261-268.
21. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the international collaborative group on HNPCC. *Gastroenterology*. 1999;116:1453-1456.
22. Thompson BA, Spurdle AB, Plazzer J-P, et al. Application of a 5-tiered scheme for standardized classification of 2,360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat Genet*. 2014;46:107-115.
23. Kovanen V, Aukee P, Kokko K, et al. Design and protocol of estrogenic regulation of muscle apoptosis (ERMA) study with 47 to 55-year-old women's cohort: novel results show menopause-related differences in blood count. *Menopause*. 2018;25:1020-1032.
24. Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data* [Online]. 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed April 26, 2022.
25. Hannon GJ. *FASTX-Toolkit*. 2010. http://hannonlab.cshl.edu/fastx_toolkit. Accessed April 26, 2022.
26. Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*. 2006;34:140-144.
27. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2009;26:139-140.
29. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:1-21.
30. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Version 4.0.5. Vienna, Austria: R Foundation for Statistical Computing; 2021. https://www.R-project.org/. Accessed April 26, 2022.
31. van der Maaten L, Hinton G. Visualizing data using t-SNE Laurens. *J Mach Learn Res*. 2008;9:2579-2605.
32. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: an online resource for prediction of microRNA binding sites. *PLoS One*. 2018;13:e0206239.
33. Ding J, Li X, Hu H. TarPmiR: a new approach for microRNA target site prediction. *Bioinformatics*. 2016;32:2768-2775.
34. Huang HY, Lin YCD, Li J, et al. MiRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res*. 2020;48:D148-D154.
35. Chen Y, Wang X. miRDB: an online database for prediction of functional microRNA targets. *Nucleic Acids Res*. 2020;48:D127-D131.
36. McGeary SE, Lin KS, Shi CY, et al. The biochemical basis of microRNA targeting efficacy. *Science*. 2019;366:6472.
37. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci*. 2005;102:15545-15550.
38. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res*. 2000;28:27-30.
39. Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA. The COSMIC cancer gene census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer*. 2018;18:696-705.
40. Bohaumilitzky L, Kluck K, Hüneburg R, et al. The different immune profiles of normal colonic mucosa in cancer-free Lynch syndrome carriers and Lynch syndrome colorectal cancer patients. *Gastroenterology*. 2022;162:907-919.e10.
41. Saberinia A, Alinezhad A, Jafari F, Soltany S, Akhavan SR. Oncogenic miRNAs and target therapies in colorectal cancer. *Clin Chim Acta*. 2020;508:77-91.
42. Carter JV, Galbraith NJ, Yang D, Burton JF, Walker SP, Galandiuk S. Blood-based microRNAs as biomarkers for the diagnosis of colorectal cancer: a systematic review and meta-analysis. *Br J Cancer*. 2017;116:762-774.
43. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144:646-674.
44. Wang Y, Zeng G, Jiang Y. The emerging roles of miR-125b in cancers. *Cancer Manag Res*. 2020;12:1079-1088.
45. Xicola RM, Bontu S, Doyle BJ, et al. Association of a let-7 miRNA binding region of TGFBR1 with hereditary mismatch repair proficient colorectal cancer (MSS HNPCC). *Carcinogenesis*. 2016;37:751-758.
46. Zhou L, Wang W, Wang F, et al. Plasma-derived exosomal miR-15a-5p as a promising diagnostic biomarker for early detection of endometrial carcinoma. *Mol Cancer*. 2021;20:57.
47. Li Z, Zhu Z, Wang Y, et al. hsa-miR-15a-5p inhibits colon cell carcinoma via targeting CCND1. *Mol Med Rep*. 2021;24:735.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Sievänen T, Korhonen T-M, Jokela T, et al. Systemic circulating microRNA landscape in Lynch syndrome. *Int J Cancer*. 2023;152(5):932-944. doi:10.1002/ijc.34338

Ib

CORRECTION TO "SYSTEMIC CIRCULATING MICRORNA LANDSCAPE IN LYNCH SYNDROME"

by

Sievänen T, Korhonen T-M, Jokela T, Ahtiainen M, Lahtinen L, Kuopio T, Lepistö A, Sillanpää E, Mecklin J-P, Seppälä TT, Laakkonen EK, 2023

International Journal of Cancer, *152*(5):932-944

https://doi.org/10.1002/ijc.34828

DOI: 10.1002/ijc.34828

ERRATUM

IJC INTERNATIONAL JOURNAL of CANCER

Correction to "Systemic circulating microRNA landscape in Lynch syndrome"

Sievänen T, Korhonen T-M, Jokela T, et al. Systemic circulating microRNA landscape in Lynch syndrome. *Int J Cancer*. 2023; 152(5): 932–944. doi:10.1002/ijc.34338

In the paper by Sievänen et al. 2023, the authors discovered an error in the analytical code for the study, which affected Table 2, Figure 1, Figure 3 legend, supporting information and some sentences in the Abstract, Methods, Results and Discussion.

This work originally used the *bowtie* aligner with default options for reference strand selection to analyze sequencing data, which was later discovered by the authors to align the microRNA (miR) reads in some instances to the reverse complements of the miRs present in the miRbase. Despite a slight change in the results, i.e., hsa-miR-122b-3p is no longer identified in our dataset, the applied reanalysis did not change the conclusions of this article.

The corrected Table 2 is listed below:

TABLE 2 DE and non-DE c-miRs within and between the discovery and cancer cohorts.

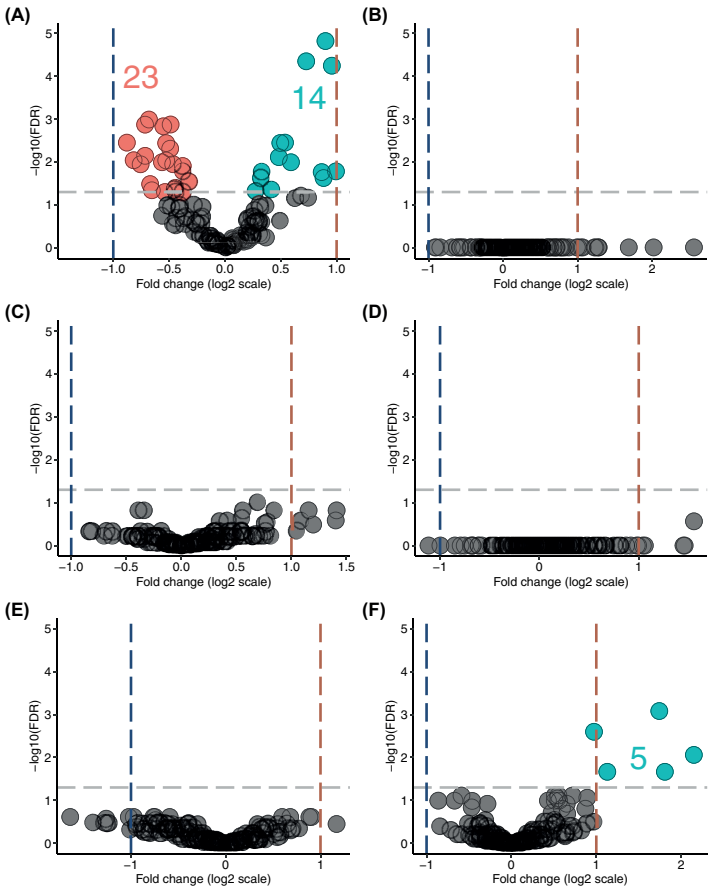
Healthy *path_MMR* vs non-LS control			Sporadic rectal cancer patients vs healthy *path_MMR*			Healthy *path_MMR* vs *path_MMR* with cancer			Sporadic rectal cancer patients vs non-LS control		
c-miR	log2FC	FDR	c-miR	log2FC	FDR	c-miR	log2FC	FDR	c-miR	log2FC	FDR
hsa-miR-155-5p	0.900	**<0.001**	hsa-miR-10a-5p	0.689	0.098	hsa-miR-127-3p	1.556	0.272	hsa-miR-200a-3p	1.742	**<0.001**
hsa-let-7c-5p	0.728	**<0.001**	hsa-miR-1180-3p	1.156	0.151	hsa-let-7a-5p	−0.014	0.997	hsa-miR-10a-5p	0.973	**0.003**
hsa-let-7e-5p	0.957	**<0.001**	hsa-miR-126-3p	−0.391	0.151	hsa-let-7b-5p	−0.240	0.997	hsa-miR-200b-5p	2.149	**0.009**
hsa-miR-15a-5p	−0.680	**0.001**	hsa-miR-148b-3p	−0.339	0.151	hsa-let-7c-5p	0.201	0.997	hsa-miR-200c-3p	1.123	**0.022**
hsa-miR-185-5p	−0.487	**0.001**	hsa-miR-196a-5p	1.410	0.151	hsa-let-7d-3p	−0.147	0.997	has-miR-196a-5p	1.807	**0.022**
hsa-miR-320a-3p	−0.715	**0.001**	hsa-miR-320a-3p	0.556	0.151	hsa-let-7d-5p	−0.323	0.997			
hsa-miR-186-5p	−0.552	**0.001**	hsa-miR-320b	0.839	0.151	hsa-let-7e-5p	0.335	0.997			
hsa-let-7a-5p	0.538	**0.004**	hsa-miR-486-5p	0.544	0.233	hsa-let-7f-5p	0.180	0.997			
hsa-miR-3613-5p	−0.878	**0.004**	hsa-miR-200b-3p	1.402	0.259	hsa-let-7 g-5p	0.011	0.997			
hsa-miR-10b-5p	0.496	**0.004**	hsa-miR-223-3p	0.415	0.259	hsa-let-7i-5p	0.072	0.997			
hsa-miR-22-3p	−0.526	**0.004**	hsa-miR-320c	1.080	0.259	hsa-miR-101-3p	0.183	0.997			
hsa-miR-19b-3p	−0.491	**0.005**	hsa-miR-185-5p	0.344	0.268	hsa-miR-103a-3p	0.256	0.997			
hsa-miR-451a	−0.715	**0.007**	hsa-miR-483-5p	0.776	0.285	hsa-miR-106b-3p	−0.207	0.997			
hsa-miR-125a-5p	0.482	**0.008**	hsa-miR-222-3p	0.752	0.310	hsa-miR-106b-5p	0.542	0.997			
hsa-miR-339-5p	−0.818	**0.009**	hsa-miR-2110	1.197	0.323	hsa-miR-107	0.180	0.997			
hsa-miR-15b-5p	−0.526	**0.009**	hsa-let-7d-3p	0.460	0.451	hsa-miR-10a-5p	−0.190	0.997			
hsa-miR-125b-5p	0.591	**0.010**	hsa-miR-11,400	−0.827	0.451	hsa-miR-10b-5p	−0.166	0.997			
hsa-miR-32-5p	−0.564	**0.010**	hsa-miR-134-5p	−0.679	0.451	hsa-miR-11,400	0.548	0.997			
hsa-miR-107	−0.464	**0.011**	hsa-miR-193a-5p	0.534	0.451	hsa-miR-1180-3p	−0.851	0.997			
hsa-miR-484	−0.755	**0.011**	hsa-miR-25-3p	0.323	0.451	hsa-miR-122-5p	0.032	0.997			
hsa-miR-27a-3p	−0.378	**0.012**									
hsa-miR-206	0.994	**0.016**									
hsa-miR-25-3p	−0.376	**0.016**									
hsa-let-7f-5p	0.328	**0.017**									

TABLE 2 (Continued)

Healthy *path_MMR* vs non-LS control			Sporadic rectal cancer patients vs healthy *path_MMR*			Healthy *path_MMR* vs *path_MMR* with cancer			Sporadic rectal cancer patients vs non-LS control		
c-miR	log2FC	FDR	c-miR	log2FC	FDR	c-miR	log2FC	FDR	c-miR	log2FC	FDR
hsa-miR-141-3p	0.866	**0.017**									
hsa-miR-126-3p	0.320	**0.024**									
hsa-miR-200a-3p	0.883	**0.024**									
hsa-miR-221-3p	−0.318	**0.029**									
hsa-miR-424-5p	−0.670	**0.033**									
hsa-miR-23a-3p	−0.444	**0.040**									
hsa-miR-27b-3p	0.415	**0.044**									
hsa-miR-486-5p	−0.451	**0.044**									
hsa-miR-222-3p	−0.655	**0.045**									
hsa-let-7i-5p	0.273	**0.048**									
hsa-miR-19a-3p	−0.438	**0.050**									
hsa-miR-363-3p	−0.537	**0.050**									
hsa-miR-92a-3p	−0.373	**0.050**									

N, healthy *path_MMR* = 81; *N*, *path_MMR* with cancer = 13; *N*, sporadic rectal cancer patients = 24; *N*, non-LS controls = 37.
FDR <0.05 highlighted with bold. FDR = false discovery rate; log2FC = logarithmic2 fold change; c-miR = circulating microRNA.

The corrected Figure 1 is shown below:



The value 40 should be changed to 37 in Figure 3 legend. The corrected legend is below:

FIGURE 3 Pathway analysis revealed putative c-miR-target genes that are linked to biological processes and pathways associated with cancer. (A) Top 30 most enriched biological processes annotated to the identified target genes of 32 out of **37** DE c-miRs found in healthy path_MMR carriers. FDR, false discovery rate; GO: BP, Gene Ontology: biological process; Hits, number of target genes annotated to the biological process. *Signal transduction by p53 class mediator resulting in cell cycle arrest. (B) Top 30 most enriched KEGG pathways annotated to the identified target genes of 32 out of **37** DE c-miRs found in healthy path_MMR carriers. c-miR, circulating microRNA; FDR, false discovery rate; KEGG, Kyoto Encyclopedia of Genes and Genomes pathway; Hits, number of target genes annotated to the pathway.

In the **abstract**, the number of DE miRs identified in cancer-free *path_MMR* carriers compared to the non-LS controls should be from 40 to 37.

In the **Methods** section paragraph 2.5 the corrected sentences should read:

Subsequently, the pre-processed reads were mapped to human mature miR-genome (miRbase v.22)[26] with Bowtie alignment tool for single-end data with v-mode and best strata parameters (-v 2 -k 1 - - best - - norc).[27] The one best mapping for each miR-read was selected for differential expression (DE) analysis.

In the **Results** section paragraph 3.2, the correct sentences should read:

- In DE-analysis, we found **37** out of 228 c-miRs to display aberrant expression in healthy *path_MMR* carriers (Table 2). Of them, **14** were upregulated and **23** downregulated in *path_MMR* carriers compared to non-LS controls, but the fold changes remained low varying from minimum of −0.88 to maximum of **0.99** (Figure 1A). Hsa-miR-155-5p, hsa-let-7c-5p and -**let-7e-5p** had the most significant upregulation within healthy *path_MMR* carriers (Table 2).
- We found that hsa-miR-200a-3p, -10a-5p, -196a-5p, **-200b-3p** and -200c-3p were significantly upregulated in sporadic rectal cancer patients differentiating them from non-LS controls (Figure 1F and Table 2).
- In this analysis scheme, the fold change in hsa-miR-200a-3p was **1.74**, indicating significantly higher expression compared to the healthy non-LS controls (Table 2).

In the **Results** section paragraph 3.4, the corrected sentences should read:

- MirWalk identified a total of 1731 miR-target gene interactions with 508 distinct putative target genes for 32 out of **37** observed DE c-miRs from discovery cohort analysis (Supplementary file S2, Tables S3 and S4).
- We chose **has-miR-451a** as representative DE miRs found in healthy *path_MMR* carriers vs non-LS control comparisons. HCT116 cell line was chosen to mimic LS colorectal cancer. The cell line experiment hinted that **underexpression** of hsa-miR-451a results in reduced **cell** viability but did not induce apoptosis of HCT116 cells (Supplementary file S1, Figure S6).

In the **Discussion** section, the correct sentences should read:

- Of these, we showed healthy *path_MMR* carriers to have an exclusive c-miRnome of **37** DE c-miRs that differs from non-LS-controls, but that does not differ from the c-miRnome of cancer patients with or without dMMR.
- The most significant DE c-miR in our setting was hsa-miR-155-5p, followed by hsa-let-7c-5p and -7e-5p and hsa-miR-15a-5p, which all except hsa-miR-15a-5p were upregulated in healthy *path_MMR* carriers.
- We found 508 putative target genes for 32 out of **37** observed DE c-miRs that associate with several pathways common to cancer.

In the Discussions section, this sentence: "We did not find experimentally verified target genes for hsa-miR-122b-3p. However, we could see that overexpression of hsa-miR-122b might result in reduced cell viability, plausibly due to increased apoptosis." was removed.

The Supporting Information has been corrected in the online version of the article.

We apologize for these errors.

II

CIRCULATING MIRNA SIGNATURE PREDICTS CANCER INCIDENCE IN LYNCH SYNDROME – A PILOT STUDY

by

Sievänen T, Jokela T, Hyvärinen M, Korhonen T-M, Pylvänäinen K, Mecklin J-P, Karvanen J, Sillanpää E, Seppälä TT, Laakkonen EK, 2024

Cancer Prevention Research, *17*(6): 243-254

https://doi.org/10.1158/1940-6207.CAPR-23-0368

Circulating miRNA Signature Predicts Cancer Incidence in Lynch Syndrome—A Pilot Study

Tero Sievänen[1], Tiina Jokela[1], Matti Hyvärinen[1], Tia-Marje Korhonen[1], Kirsi Pylvänäinen[2], Jukka-Pekka Mecklin[2,3], Juha Karvanen[4], Elina Sillanpää[1,2], Toni T. Seppälä[5,6,7,8], and Eija K. Laakkonen[1]

ABSTRACT

Lynch syndrome (LS) is the most common autosomal dominant cancer syndrome and is characterized by high genetic cancer risk modified by lifestyle factors. This study explored whether a circulating miRNA (c-miR) signature predicts LS cancer incidence within a 4-year prospective surveillance period. To gain insight how lifestyle behavior could affect LS cancer risk, we investigated whether the cancer-predicting c-miR signature correlates with known risk-reducing factors such as physical activity, body mass index (BMI), dietary fiber, or NSAID usage. The study included 110 c-miR samples from LS carriers, 18 of whom were diagnosed with cancer during a 4-year prospective surveillance period. Lasso regression was utilized to find c-miRs associated with cancer risk. Individual risk sum derived from the chosen c-miRs was used to develop a model to predict LS cancer incidence. This model was validated using 5-fold cross-validation. Correlation and pathway analyses were applied to inspect biological functions of c-miRs. Pearson correlation was used to examine the associations of c-miR risk sum and lifestyle factors. hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615 were identified as cancer predictors by Lasso, and their risk sum score associated with higher likelihood of cancer incidence (HR 2.72, 95% confidence interval: 1.64–4.52, C-index = 0.72). In cross-validation, the model indicated good concordance with the average C-index of 0.75 (0.6–1.0). Coregulated hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p targeted genes involved in cancer-associated biological pathways. The c-miR risk sum score correlated with BMI ($r = 0.23$, $P < 0.01$). In summary, BMI-associated c-miRs predict LS cancer incidence within 4 years, although further validation is required.

Prevention Relevance: The development of cancer risk prediction models is key to improving the survival of patients with LS. This pilot study describes a serum miRNA signature–based risk prediction model that predicts LS cancer incidence within 4 years, although further validation is required.

Introduction

Lynch syndrome (LS) is the most common inherited cancer predisposition syndrome, with an estimated prevalence of 1:300 (1, 2). Distinct LS phenotypes are caused by germline mutations in DNA mismatch repair (MMR) genes *MLH1*, *MSH2*, *MSH6*, and *PMS2* (2). The impaired MMR manifests as an increased risk of multiple cancers, and depending on the cancer type, the risk is modified by lifestyle factors such as physical activity, body weight, consumption of dietary resistant starch, and NSAID usage (2–8). LS cancer spectrum includes various cancer types, colorectal cancer and endometrial cancers being most common (6). As the cancer risk varies greatly among pathogenic MMR variant carriers (6), it is pivotal to innovate risk stratification biomarkers that could be used to identify LS carriers who may develop cancer in the near future.

Circulating miRNAs (c-miR) are short, noncoding RNA molecules that function as intercellular messengers by migrating throughout the body (9). They play a crucial role in cancer biology by regulating core cellular processes, such as proliferation and apoptosis, through the suppression of target gene translation (10). Multiple studies have reported c-miRs as potential biomarkers for various sporadic cancers (11–15) by demonstrating differential expression (DE) between the c-miR signatures of patients with cancer and healthy controls. In most of these prior studies, the analysis of c-miR signatures has been limited to patients who have already received a colorectal cancer diagnosis, making it challenging to ascertain their

[1]Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland. [2]The wellbeing services county of Central Finland, Jyväskylä, Finland. [3]Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä, Finland. [4]Department of Mathematics and Statistics, University of Jyväskylä, Jyväskylä, Finland. [5]Applied Tumor Genomics Research Program, University of Helsinki, Helsinki, Finland. [6]Department of Abdominal Surgery, Helsinki University Hospital and University of Helsinki, Helsinki, Finland. [7]Department of Gastroenterology and Alimentary Tract Surgery and TAYS Cancer Centre, Tampere University Hospital, Tampere, Finland. [8]Faculty of Medicine and Health Technology, University of Tampere, Tampere, Finland.

T.T. Seppälä and E.K. Laakkonen contributed equally to this article.

Corresponding Authors: Tero Sievänen, Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, Jyväskylä 40700, Finland. E-mail: tero.o.sievanen@jyu.fi; and Eija K. Laakkonen, eija.k.laakkonen@jyu.fi

Cancer Prev Res 2024;17:243–54

doi: 10.1158/1940-6207.CAPR-23-0368

potential utility in risk stratification. Interestingly, a recent study by Raut and colleagues showed that altered c-miR expression could predict sporadic colorectal cancer incidence several years prior the diagnosis (16). However, it has remained unclear whether this observation extends to LS.

In LS, the risk of various cancers is significantly elevated by sedentary behavior and excess body weight, while physical activity, maintaining a healthy body weight, and the consumption of dietary resistant starch and NSAIDs have been shown to mitigate these risks (3, 7, 8). Although it is well acknowledged that adopting an optimal lifestyle can reduce cancer incidence, the underlying molecular mechanisms remain less elucidated. c-miRs, due to their capacity to modulate pathophysiologic responses to changing lifestyle behaviors (9) and their ability to exhibit DE profiles between sedentary and physically active individuals (17), offer potential insights into how lifestyle behaviors influence LS-associated cancer risk.

We were first to report that the c-miR signature of cancer-free LS carriers is associated with carcinogenesis by displaying aberrant expression compared with healthy population but similar expression when compared with patients with sporadic rectal cancer (18). To build on that, the primary aim of this study was to investigate whether c-miRs can be used in LS cancer risk prediction during a 4-year prospective surveillance period. Considering the modulatory role of c-miRs in lifestyle habits, our secondary aim was to explore whether any of the LS cancer predictive c-miRs are associated with physical activity, body weight, dietary fiber, or NSAID usage.

Materials and Methods

The study flow chart and general outline is detailed in **Fig. 1**.

Patients and sample collection

The clinical data of our study were derived from the nationwide Finnish Lynch Syndrome Research Registry (LSRFi, www.lynchsyndrooma.fi, accessed November 2022). Age, sex, MMR mutation status, family cancer history, and all cancer diagnoses with the cancer type and date of each diagnosis were confirmed from hospital medical records and national cancer registries upon recording in the LSRFi. To date, LSRFi includes 1,800 LS carriers from 400 families and contains clinicopathologic information on all cancers of the registered individuals. In the current study, we reviewed baseline medical records of Finnish cancer-free LS carriers whose c-miR expression profile was characterized ($n = 110$). Ethnicity throughout the study population was White Caucasian.

LS carriers were enrolled in the study, and whole blood was collected at their regular colonoscopy surveillance appointments at Helsinki University Central Hospital in Helsinki and Central Finland Central Hospital in Jyväskylä, Finland. Non-LS control samples were acquired from Biobank of Eastern Finland, Kuopio, and a previously studied Estrogenic Regulation of Muscle Apoptosis cohort consisting of healthy 47–55 years old women. To separate serum, the whole blood samples were allowed to clot for 30 minutes at room temperature, centrifuged at 1,800 × *g* for 10 minutes and aliquoted. Methods of sample collection, preanalytic preparation, c-miR extraction, library preparation, and sequencing have been described previously in detail (18).

Data collection and ethical issues

High-throughput c-miR expression data of cancer-free LS carriers ($n = 86$) as well as of healthy non-carrier control samples ($n = 37$) were generated as described earlier (18). Briefly, c-miRs were extracted using affinity column-based approach (miRNeasy Serum/Plasma advanced kit, Qiagen), ligated to sequencing adapters from both 5′ and 3′ end, reverse transcribed into cDNA using unique molecular identifier (UMI)-assigning primers, and purified with magnetic beads (Qiaseq miRNA Library preparation kit, Qiagen). Sequencing of the c-miR libraries were done with NextSeq 500 (Illumina) using NextSeq 500/550 High Output Kit v. 2.5 with 75 cycles aiming for depth of 5M reads per sample. Quality controls throughout the RNA isolation, library preparation, and sequencing protocols were conducted with qRT-PCR (Bio-Rad), TapeStation 4200 (Agilent) and Qubit fluorometer (Invitrogen), respectively. To increase the cohort size, we performed small RNA sequencing (RNA-seq) experiment on additional 24 LS carriers using the same analysis pipeline as described (18). Thus, the current study composed of 110 cancer-free LS carriers who are registered in the LSRFi and 37 healthy non-carrier control samples. Healthy non-carrier control samples were included only in the DE analysis to confirm previously reported LS-associated c-miR signature.

The corresponding lifestyle data of the cancer-free LS carriers in the current study were collected as described previously in detail (3). Briefly, questionnaires for anthropometric, socioeconomic, and lifestyle data collection were sent to adult Finnish LS carriers whose contact information was available in LSRFi in 2017 and 2020. Alongside with the lifestyle data collection, dietary habits data of the same persons were collected by a validated semiquantitative food frequency questionnaire (19). The average time period between the questionnaires' data collection and blood sampling was 2.0 (0.3−3.9) years. A written informed consent was obtained from all participants, and the Helsinki and Uusimaa Health Care District (HUS/155/2021) and Central Finland Health Care District Ethics Committee (KSSHP D# 1U/2018 and 1/2019 and KSSHP 3/2016) approved the study protocol. The study was conducted according to the guidelines of the Declaration of Helsinki.

Missing data

There were no missing c-miR data. Missing lifestyle and dietary data [physical activity: 30.9%; body mass index (BMI): 4.5%; dietary fiber intake: 29.0% and NSAID usage: 29.0%; Supplementary Table S1] occurred due to incomplete questionnaire responses. Missing data were assumed to occur at random and multiple imputation with 50 iterations was used to create and analyze 50 multiply imputed datasets using mice"

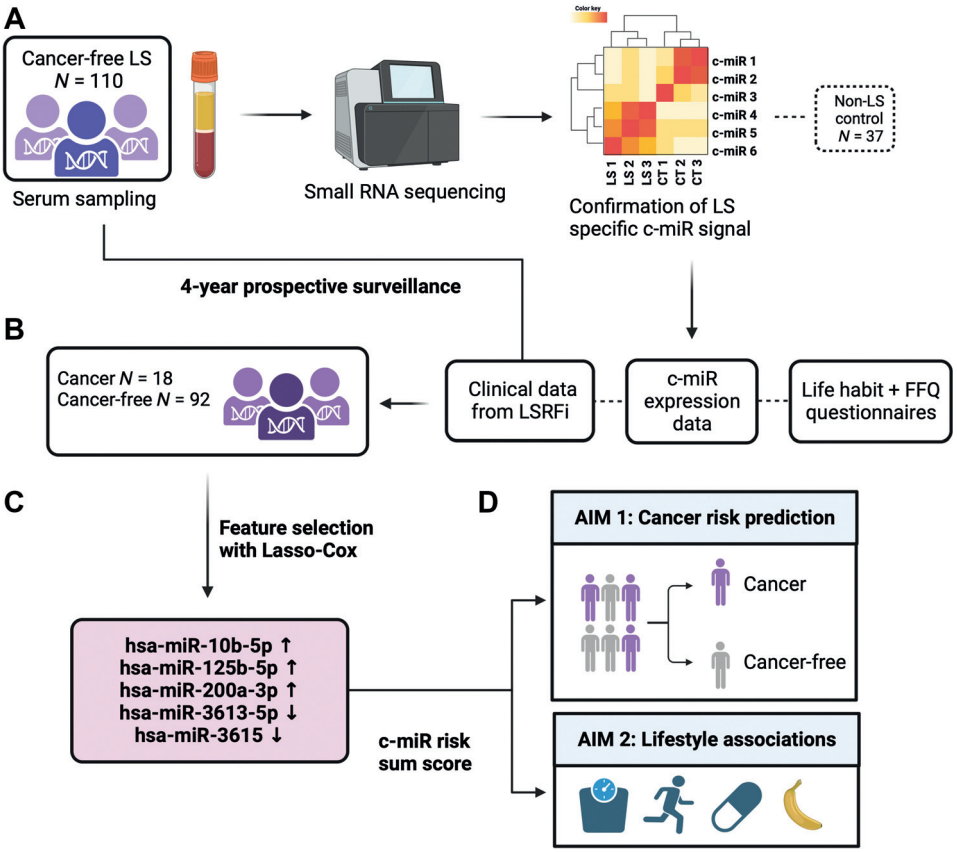


Figure 1.

The study flow chart and general outline. **A,** Serum samples of 110 cancer-free LS carriers and 37 non-carrier controls were sequenced to confirm previously identified LS specific c-miR signature. **B,** LS clinical data were derived from LSRFi to assess the cancer status of cancer-free LS carriers after 4 years of prospective surveillance. Of the 110 cancer-free LS carriers, 18 had developed cancer during the surveillance period. Then, c-miR expression data were matched with the corresponding clinical and lifestyle data to investigate whether the c-miR signature can predict LS cancer risk during the surveillance. Lifestyle data were collected 2017 or 2020 with a questionnaire. Blood sample was taken at regular colonoscopy visit between 2018 and 2020. The average time period between lifestyle data collection and blood sample was 2.0 (0.3–3.9) years. **C,** Lasso-regularized Cox regression was used to select the most important predictor c-miRs from the entire cohort. Lasso-obtained c-miRs were used to compute c-miR risk sum score. Arrows indicate upregulation (↑) and downregulation (↓) of the c-miR in LS carriers when compared with healthy non-carriers. **D,** c-miR risk sum score was used in LS cancer risk prediction with 5-fold cross-validation and to inspect associations with lifestyle data. c-miR = circulating miRNA; Lasso = Least absolute shrinkage and selection method; LS = Lynch syndrome; LSRFi = Finnish Lynch Syndrome Research Registry. This figure was created with BioRender.com.

R-package (20) with default settings. All lifestyle variables as well as sex, age, pathogenic MMR-variant, cancer status, and c-miR expression were used for imputation of each lifestyle variable, and results were pooled using "pool" function in "mice".

DE analysis

DE analysis between cancer-free LS carriers and healthy non-carrier controls was performed with "DESeq2" R-package (ref. 21; RRID:SCR_000154) using raw c-miR counts (Supplementary Materials and Methods S1). Sex and sequencing batch were added to the DE analysis design formula to account for their potential confounding effect. Normalization and variance stabilization transformations were done with DESeq2 by applying median of ratios method (21) and "rlog" function, respectively. Low count c-miRs were filtered prior to DE analysis. Filtering was done with "filterByExpr" function in "edgeR" R-package (22) that excluded c-miRs with <1 count per million in 70% of samples. Benjamini–Hochberg procedure with FDR 0.05 was used to correct for multiple testing. Hierarchical clustering based on Euclidean distances and the "complete" method was applied to verify DE findings. "hclust" function in "stats" base R-package was used for the hierarchical clustering analysis.

Covariates

c-miRs

c-miR expression data were derived from small RNA-seq experiments and measured as counts relative to sample library size where counts represent molecules in blood serum. DESeq2

normalized and variance stabilized c-miR counts were used for all analyses.

Physical activity

Physical activity was assessed by a self-reported questionnaire. The questionnaire included four questions about the frequency, intensity, and duration of leisure time physical activity and commuting activity. On the basis of the responses, the metabolic equivalent task hours per day for leisure time physical activity was calculated.

BMI

Body weight and height were measured by the clinician during the study subjects' regular colonoscopy appointment. If body weight and height information were missing, we used the last known self-measured weight and height measurement. BMI was calculated as weight in kilograms divided by the height squared in meters (kg/m^2) according to World Health Organization guidelines.

Dietary fiber

Dietary fiber including resistant starch amount was derived from self-reported food frequency questionnaire and assessed as grams per day.

NSAID usage

Study subjects self-reported whether (yes/no) they used NSAIDs, such as acetylsalicylic acid, ibuprofen or ketoprofen products frequently.

Construction and validation of the LS cancer risk prediction model

Least absolute shrinkage and selection method (Lasso; ref. 23) regularized Cox regression was used to find predictor c-miRs from the pool of identified LS-associated DE c-miRs using the entire study sample. Optimal value for the Lasso regularization parameter lambda was chosen with 10-fold cross-validation. The expression levels of the Lasso-obtained c-miRs were used to compute an individual risk sum score (linear predictor) for all the participants by using formula:

$$\text{Risk sum score} = \text{Expr}(\text{miR}_A) * \beta(\text{miR}_A) + \text{Expr}(\text{miR}_B) * \beta(\text{miR}_B) \ldots,$$

where Expr(miR) represents the normalized and variance stabilized c-miR expression and β(miR) indicates the regression coefficient in Lasso-Cox regression model (16). By using univariate and multivariate Cox regression models, the c-miR risk sum score was then applied to predict the risk of cancer incidence. We used the entire study sample ($n = 110$) for fitting the risk prediction model. The predictive performance of the risk prediction model was validated with 5-fold cross-validation and the model concordance evaluated with Harrel C-index (scale 0.5–1.0) where 0.5 indicates poor performance and 1.0 indicates excellent performance (ref. 24; Supplementary Materials and Methods S1).

The surveillance time used for risk prediction was determined from the timepoint of initial serum sampling (2018–2020) until the latest update of LSRFi (November 2022). The response variable in the risk prediction model was the age at the time of cancer diagnosis (event) or the age at the final update date of LSRFi (censoring). HR and 95% confidence intervals (CI) of the c-miR risk sum score were estimated for unadjusted model as well as for sex and MMR-variant adjusted model. Proportional hazards assumption was tested using Schoenfeld residuals (Supplementary Fig. S1). Regarding the risk prediction model development and validation, we followed Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) reporting checklist (25). We used "glmnet" R-package (26) for the cross-validation procedure as well as for Lasso-regularized Cox regression. "survival" R-package was used for Cox regression modeling (27).

Pathway analysis

We identified potential targets genes of the Lasso-obtained c-miRs from miRTarBase (28) by using miRWalk online tool (29). We considered only the genes with experimental validation in MiRTarBase (28) to exclude low evidence targets. The obtained target gene list was applied to overpresentation analysis with hypergeometric tests using Search Tool for Retrieval of Interacting Genes/Proteins (STRING; ref. 30) and Reactome (31) databases.

Statistical analysis

All statistical analyses were performed in R-programming environment (v.4.2.2) using RStudio and in-house R-scripts. Levene test was used to inspect homoscedasticity. Study subject characteristics are presented as means and SDs for continuous variables and as number of study subjects and percentages for categorical variables. Regarding **Table 1**, Welch two-sample *t* test was used for continuous variables whereas χ^2 test was applied for categorical variables. Because of skewed nature of RNA-seq data, Spearman method was applied to inspect correlations between the Lasso-obtained c-miRs. Pearson method was applied to examine correlations between the multiple imputed lifestyle data and c-miR risk sum score.

Data availability

The sequence data generated in this study are publicly available in Sequence Read Archive (SRA) at PRJNA1088397.

Results

Study subject characteristics

The study subjects' clinical characteristics are described in **Table 1**. Most had a pathogenic *MLH1* germline variant (67.3%) followed by *MSH2* (17.3%), *MSH6* (13.6), and *PMS2* (1.8), respectively. Of the 110 study subjects, 18 (13 males and 5 females) developed cancer during the prospective surveillance. The mean surveillance time for those who developed cancer was 1.3 years whereas for those who remained cancer-free it

Table 1. Study subject characteristics.

Parameter	Total cohort	Cancer during surveillance	Cancer-free after surveillance	*P*-value
N (%)	110	18 (16.4)	92 (83.6)	
Sex, *N* (%)				0.071
Male	57 (51.8)	13 (72.2)	44 (47.8)	
Female	53 (48.2)	5 (27.8)	48 (52.2)	
MMR status, *N* (%)				0.777
MLH1	74 (67.3)	14 (77.8)	60 (65.2)	
MSH2	19 (17.3)	2 (11.1)	17 (18.5)	
MSH6	15 (13.6)	2 (11.1)	13 (14.1)	
PMS2	2 (1.8)	—	2 (2.2)	
Physical activity, MET/hours/day (SD)[a]	4.4 (± 4.5)	7.6 (± 7.2)	3.7 (± 3.6)	0.094
BMI, kg/m^2 (SD)[a]	27.8 (± 5.8)	27.9 (± 4.4)	27.7 (± 6.1)	0.875
Dietary fiber intake, g/day[a] (SD)	23.4 (± 10.0)	21.1 (± 9.8)	23.9 (± 10.0)	0.379
NSAID usage, *N* (%)[a]				0.736
Yes	26 (33.3)	3 (25.0)	23 (34.8)	
No	52 (66.7)	9 (75.0)	43 (65.2)	
Age at the start of surveillance, years[a] (SD)	57.5 (± 11.8)	57.6 (± 14.3)	57.7 (± 11.4)	0.967
Age at the end of surveillance, years[a] (SD)	60.7 (± 12.0)	58.9 (± 14.4)	61.0 (± 11.5)	0.575
Surveillance time, years[a] (SD)	3.1 (± 1.1)	1.3 (± 1.1)	3.5 (0.6)	<**0.001**
Cancer history, *N* (%)				0.636
Yes	54 (49.1)	10 (55.6)	44 (47.8)	
No	56 (50.9)	8 (44.4)	48 (52.2)	
Cancer, *N* (%)	18 (16.4)	18 (16.4)	—	
CRC	9 (50.0)	9 (50.0)	—	
Other[b]	9 (50.0)	9 (50.0)	—	

Abbreviations: BMI: body mass index; CRC: colorectal cancer; MET: metabolic equivalent task; MMR: mismatch-repair gene; NSAID: non-steroidal anti-inflammatory drug; SD: standard deviation.

[a]Missing values, total cohort: Physical activity, $n = 34$; BMI, $n = 5$; dietary fiber intake, $n = 32$; NSAID usage, $n = 32$. Missing values, cancer: Physical activity, $n = 6$; dietary fiber intake, $n = 6$; NSAID usage, $n = 6$. Missing values, cancer-free: Physical activity, $n = 28$; BMI, $n = 5$; dietary fiber intake, $n = 26$; NSAID usage, $n = 26$.

[b]Other cancers included bladder cancer ($n = 1$), breast cancer ($n = 1$), esophageal cancer ($n = 1$), glioma ($n = 1$) gastric cancer ($n = 1$), prostate cancer ($n = 3$), and spinocellular cancer ($n = 1$).

was 3.5 years. Half of the diagnosed cancers were colorectal cancers and the other half consisted of several other cancer types (Supplementary Table S2). No loss to follow-up occurred.

Confirmation of LS-associated c-miR signature

We have previously identified a c-miR signature that distinguished LS carriers from healthy non-carrier population. However, as the LS cohort used in the present study included 24 new cases, we reprocessed the data to seek for more LS-associated c-miRs and to verify our previous finding. DE analysis resulted in 37 DE c-miRs between cancer-free LS carriers and healthy non-carrier controls (**Fig. 2A**; Supplementary Table S3). We found 14 upregulated DE c-miRs and 23 downregulated DE c-miRs (**Fig. 2B**). These 37 DE c-miRs were confirmed as LS-associated and thus chosen for the downstream analyses.

The expression levels of hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615 are associated with increased risk of cancer incidence

Several multi c-miR panels have been reported to have predictive or prognostic value in sporadic cancer risk assessment. Thus, we wanted to investigate whether the expression of any of the LS-associated DE c-miRs showed potential in LS cancer risk prediction during the prospective surveillance. Out of the 37 DE c-miRs, Lasso selected hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615 as the best predictors that separated those LS carriers who developed cancer from those who remained cancer-free during the surveillance (**Fig. 3A**). The expression of all these c-miRs was higher in those LS carriers who developed cancer during the surveillance compared with those LS carriers who remained cancer-free, although only hsa-miR-3613-5p displayed statistical significance (**Fig. 3B**). Of them, only hsa-miR-10b-5p was independently associated with an increased cancer risk (HR 6.58, 95% CI: 1.43–30.21, $\beta = 1.88$; Supplementary Table S4). The full model showed good concordance (C-index = 0.72; Supplementary Table S4).

Because efficient miR-based biological regulation relies on additive effects of multiple miRs (32), we wanted to investigate the pooled performance of the selected c-miRs on predicting LS cancer risk. We observed that c-miR risk sum score was significantly associated with increased risk of cancer incidence (HR 2.72, 95% CI: 1.64–4.52, $\beta = 1.00$, C-index = 0.72) also after adjusting for sex and MMR-variant (HR 2.71, 95% CI: 1.62–4.52, $\beta = 1.00$, C-index = 0.77; **Fig. 3C**). A 5-fold cross-validation of this risk prediction model resulted in average

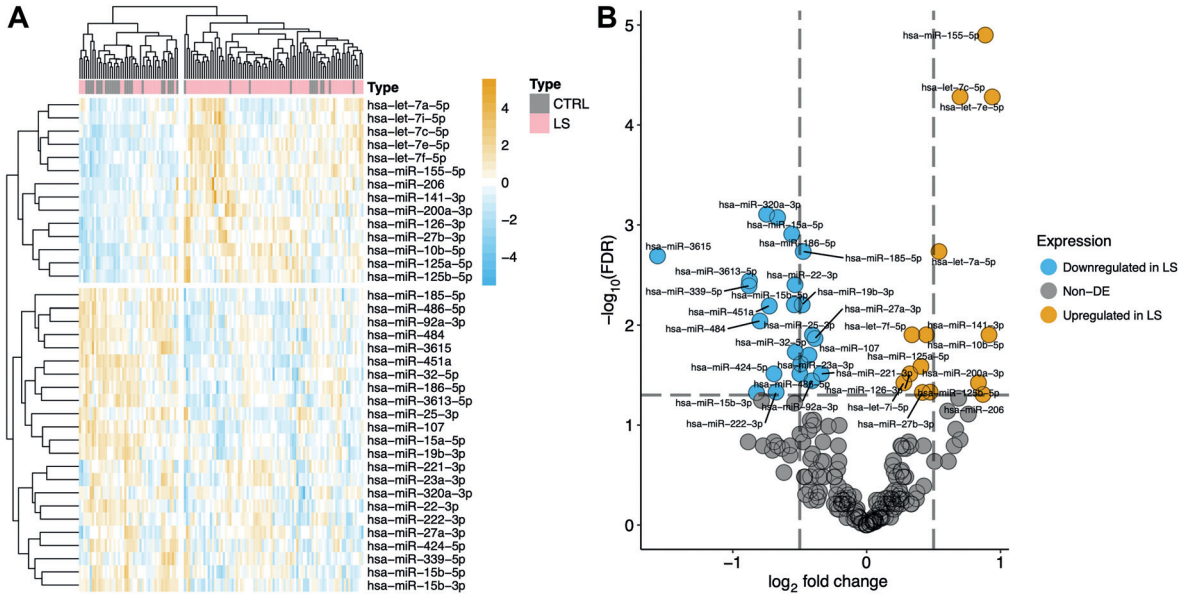


Figure 2.

Confirmation of LS-associated c-miR signature. **A,** Heat map with hierarchical clustering of DE c-miRs ($n = 37$) that separated cancer-free LS carriers and non-carrier controls. Orange color indicates c-miR upregulation in LS group whereas blue color indicates c-miR downregulation in LS group. The scale represents normalized and variance stabilized c-miR counts. LS samples are annotated with pink color and non-carrier controls with gray. **B,** Volcano plot of DE c-miRs that separated cancer-free LS carriers and non-carrier controls. Only the upregulated (orange) and downregulated (blue) c-miRs in LS group are annotated. Gray dots represent non-DE c-miRs. *Y*-axis indicates $-\log_{10}$ FDR whereas *X*-axis indicate $\log_2$ fold change of c-miR expression.

C-index of 0.75 (0.60–1.00; **Fig. 3D**) thus presenting good concordance (Supplementary Table S5). The mean c-miR risk sum score was higher in those LS carriers who developed cancer (mean = 44.0) compared with those who did not (mean = 43.1; $P < 0.01$).

We also conducted two sensitivity analyses that included either *MLH1* carriers ($N = 74$ of whom 14 developed cancer) or colorectal cancer cases ($N = 101$ of whom 9 developed colorectal cancer; Supplementary Tables S6–S11). Lasso selected hsa-let-7e-5p, hsa-miR-10b-5p, and hsa-miR-3613-5p as the best predictors to separate those who developed cancer from those who did not in the *MLH1* subgroup. Regarding the colorectal cancer cases, hsa-miR-10b-5p, hsa-miR-19b-3p, hsa-miR-200a-3p, hsa-miR-27b-3p, and hsa-miR-3615 were selected as the best predictors. Although a risk sum score in both sensitivity analyses was independently associated with increased cancer incidence after adjusting, an enhanced risk prediction performance was seen only in colorectal cancer–stratified model (C-index = 0.84) but not in *MLH1* model (C-index = 0.56) when compared with the unstratified model.

Taken together, risk prediction models composed of hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615 could classify between those LS carriers who developed cancer during the surveillance period and those who did not, also when stratified for *MLH1* or colorectal cancer. Higher prediagnostic expression levels of these c-miRs are associated with increased risk of cancer incidence.

Pathway analysis links coregulated hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p to cell cycle regulation, programmed cell death, cellular senescence, and transcriptional regulation

The targeting of multiple genes within a specific pathway, as well as the additive effects of coregulated c-miR clusters, are key elements of effective c-miR regulation (32). First, we conducted a correlation analysis to inspect whether the Lasso-obtained c-miRs present possible coregulation. Hsa-miR-10b-5p correlated with hsa-miR-200a-3p (rho = 0.28, $P < 0.01$) and with hsa-miR-125b-5p (rho = 0.29, $P < 0.01$), hsa-miR-200a-3p correlated with hsa-miR-125b (rho = 0.41, $P < 0.001$) whereas hsa-miR-3613-5p correlated only with hsa-miR-3615 (rho = 0.31, $P < 0.01$) thus displaying correlation and expression concordance (**Fig. 4A**). hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p were upregulated in LS whereas hsa-miR-3613-5p and hsa-miR-3615 were downregulated when compared with the healthy non-carrier controls, respectively (**Fig. 2B**).

To gain insight on relevant biological processes of hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615, we first predicted their putative target genes using miRWalk. We found 128 unique target genes for all the c-miRs expect for hsa-miR-3613-5p (Supplementary Table S12). The most important gene nodes are presented in **Fig. 4B**. These nodes had significant interactions among each other ($P < 0.001$) which provided support for biological connection. Of them, *BCL2*, *EGFR*, *CDKN1A*, *CDKNA2A*, *STAT3*, *SMAD2*, *CREB1*, *ETS1*, and *CD44* had the most

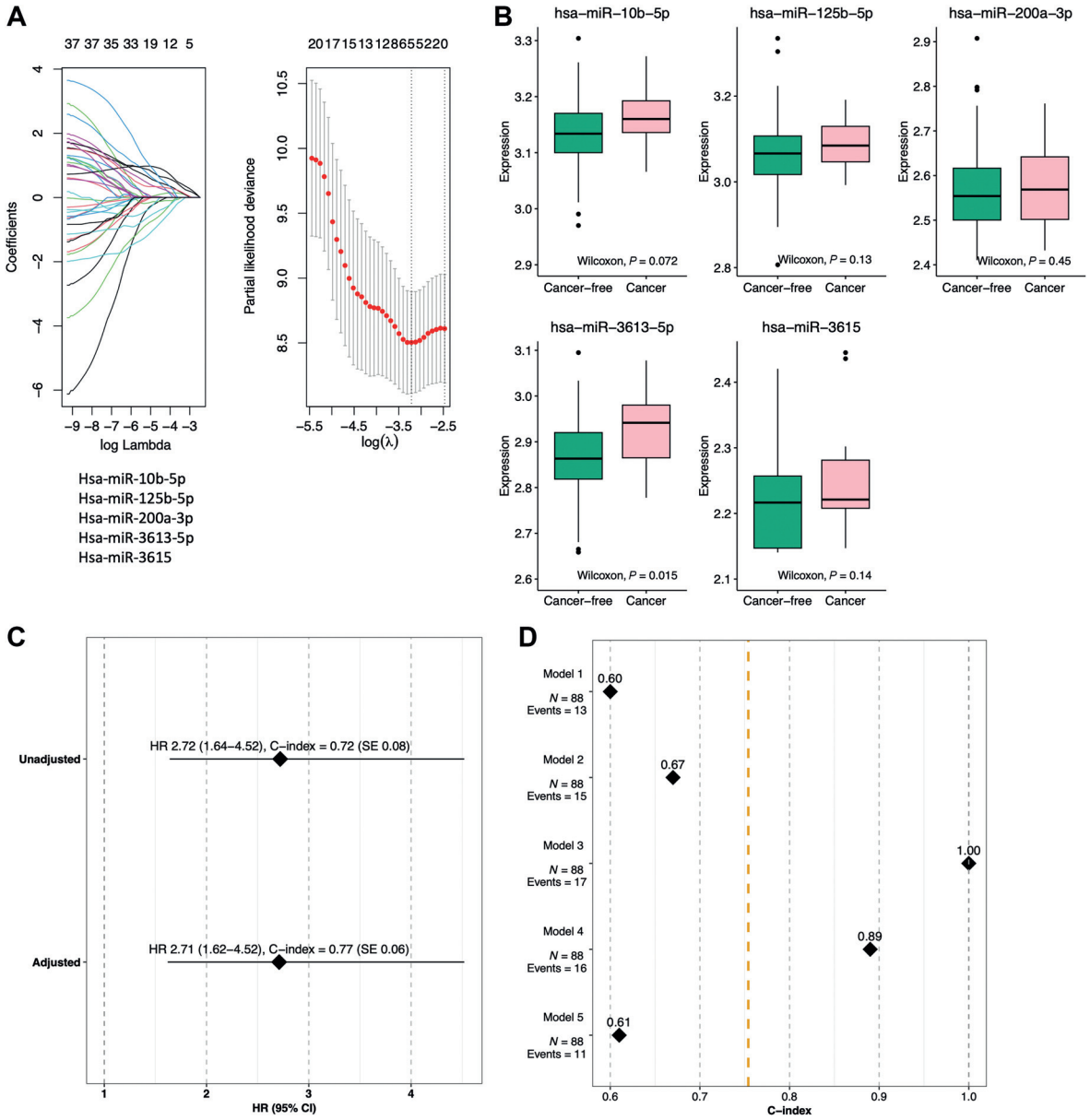


Figure 3.

hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615 are associated with increased risk of cancer incidence. **A,** Left panel presents a Lasso feature selection graph where every colored line indicate one of the 37 DE c-miRs found between cancer-free LS carriers and non-carrier controls. Regression coefficient is presented as a function of the tuning parameter lambda. Right panel shows the partial likelihood deviance as a function of lambda. The area between the dashed lines presents the optimal lambda value ($\lambda = -2.5$) after 10-fold cross-validation. **B,** Boxplots present expression differences of the Lasso-obtained c-miRs between LS carriers who got cancer (pink) and cancer-free LS carriers (green). All of the Lasso-obtained c-miRs were upregulated in those who developed cancer during the surveillance. The expression values on the *Y*-axis are presented as normalized and variance stabilized c-miR counts. **C,** Unadjusted and sex and MMR-variant-adjusted LS cancer risk prediction models. HRs, 95% CIs, and model C-indices are shown. **D,** A total of 5-fold cross-validated LS cancer risk prediction model. Number of samples and events are shown for the training folds (80% of data) whereas C-indices are shown for the validation fold (20% of data). Orange color indicates the mean C-index (0.75) across all folds.

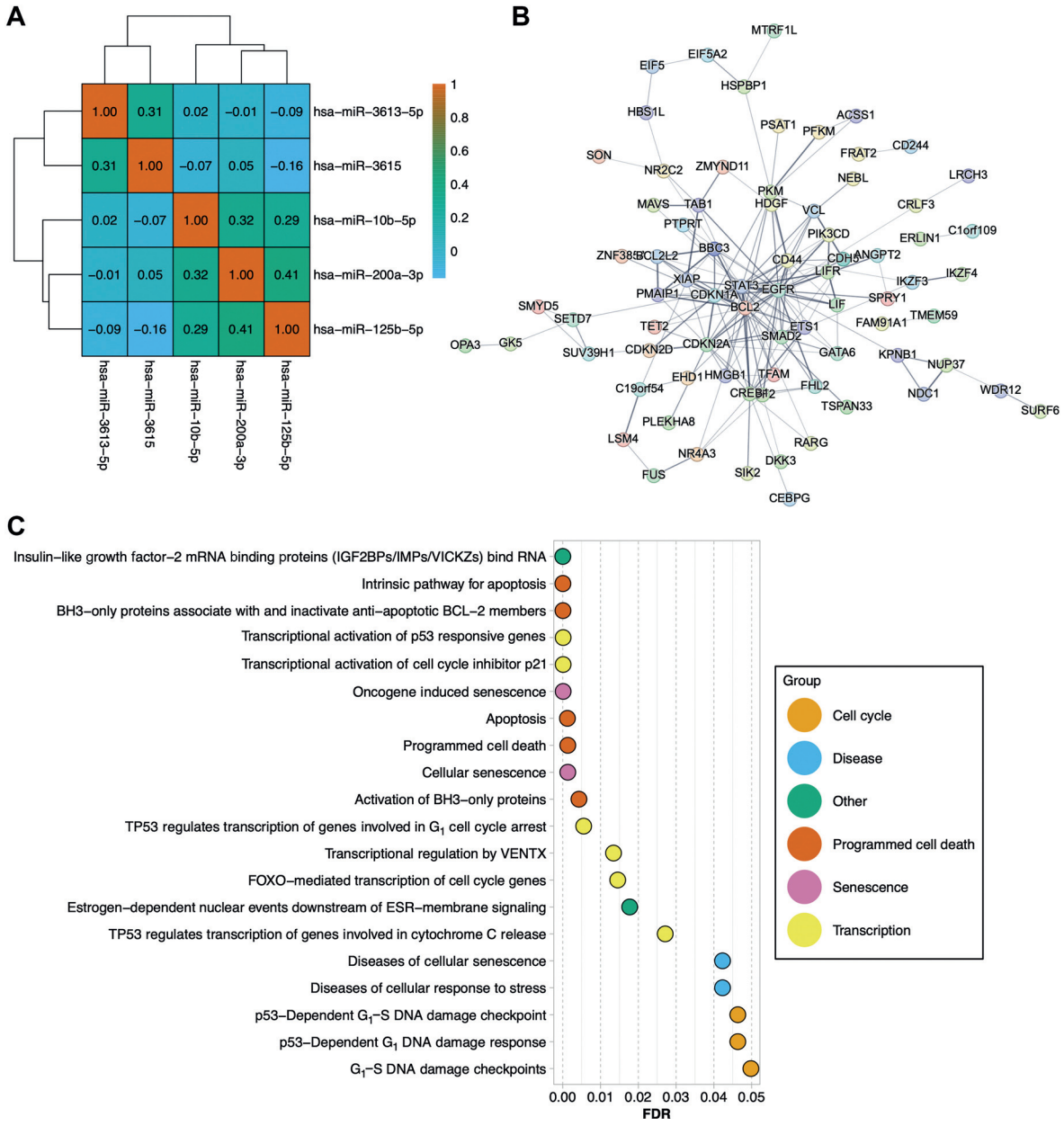


Figure 4.

Pathway analysis linked coregulated hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p to cell cycle regulation, programmed cell death, cellular senescence, and transcriptional regulation. **A,** Heat map of correlations among the Lasso-obtained c-miRs hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615. hsa-miR-10b-5p correlated with hsa-miR-125b-5p ($P < 0.01$) and hsa-miR-200a-3p ($P < 0.01$), hsa-miR-125b-5p correlated with hsa-miR-200a-3p ($P < 0.01$) whereas hsa-miR-3613-5p correlated with hsa-miR-3615 ($P < 0.01$). $P < 0.05$ was considered significant. The scale represents the magnitude of correlation. Blue indicates low correlation and orange indicate high correlation. **B,** The most important gene nodes included *EGFR, CDKN1A*, *CDKNA2A*, *STAT3*, *SMAD2, CREB1, ETS1*, and *CD44* and were observed to have significant interactions ($P < 0.001$) with each other. Edge thickness indicates the strength of data support between the nodes. **C,** Pathway analysis of 86 experimentally verified target genes of hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p were significantly enriched (FDR < 0.05) in several pathways linked to cell cycle regulation, programmed cell death, cellular senescence, and transcriptional regulation.

interactions. hsa-miR-10b-5p targeted tumor suppressor genes *CDKN1A, CDKNA2A*, and *CREB1*, hsa-miR-125b-5p targeted oncogenes *BCL2* and *STAT3*, proto-oncogene *ETS1* as well as *CD44*, hsa-miR-200a-3p targeted oncogene *EGFR* and tumor suppressor gene *SMAD2*, that further supported possible coregulation of these c-miRs. The complete gene node map is presented in Supplementary Fig. S2.

Next, we conducted a pathway analysis on the experimentally confirmed c-miR target genes reported in MiRTarBase (28). A total of 86 out of 128 of the found target genes were significantly enriched in several pathways related to cell cycle regulation, programmed cell death, cellular senescence as well as transcriptional regulation (**Fig. 4C**). The observed pathways, such as those linked to DNA damage response and programmed cell death, are also in line with the acknowledged biology of cancers. These pathways along with cellular senescence pathways were targeted by coregulative and upregulated hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p (Supplementary Table S13). In summary, hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p showed potential coregulation by displaying reciprocal correlation and by targeting genes involved in several biological pathways relevant to cancers.

c-miR risk sum score correlates with BMI

c-miRs modulate multisystemic adaptations in the human body in response to lifestyle behavior. Therefore, we investigated whether the five c-miR risk sum score was associated with lifestyle factors that are reported to reduce LS cancer risk, or age which is a significant cancer risk factor in LS. Of the chosen lifestyle factors, only BMI showed significant correlation with the c-miR risk sum score (**Table 2**). Using the multiple imputed datasets did not show significant differences to a complete-case analysis (Supplementary Table S14). These findings indicate that the expression levels of hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a, hsa-miR-3613-5p, and hsa-miR-3615 might be affected by BMI thus suggesting potential link between lifestyle, c-miRs and LS cancer risk.

Table 2. Pearson correlations of c-miR risk sum score and physical activity, BMI, dietary fiber consumption, NSAID usage, and age.

	r	95% CI	*P*-value
Physical activity	0.03	[−0.19, 0.26]	0.76
BMI	0.23	[0.04–0.43]	**0.01**
Dietary fiber intake	0.04	[−0.18, 0.26]	0.71
NSAID usage	−0.03	[−0.25, 0.18]	0.75
Age	−0.14	[−0.33, −0.05]	0.14

Note: Lifestyle data were collected 2017 or 2020 with a questionnaire. Blood sample was taken at regular colonoscopy visit. The average time-period between lifestyle data collection and blood sample was 2.0 (0.3–3.9) years. *P*-value significant at 0.05 level.
Abbreviations: r = Pearson correlation coefficient; 95% CI = 95% confidence interval; BMI = body mass index; NSAID = non-steroidal anti-inflammatory drug usage.

Discussion

Our pilot study was the first to assess whether a c-miR expression signature could be used in LS cancer risk prediction during a 4-year prospective surveillance period. We also investigated whether this signature associates with lifestyle factors and age. Using Lasso regression and bioinformatics approaches, we showed that a risk sum score composed of hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615 associates with an increased risk of LS cancer incidence. We also observed that this c-miR risk sum score correlates positively with BMI.

Identifying reliable biomarkers has the potential to aid in risk stratification of high-risk patients (33). Integrating these biomarkers with clinicopathologic factors could enhance the accuracy of patient selection criteria for risk-based screening programs. In the current study, Lasso-Cox model successfully separated LS carriers who developed cancer from those who did not by using a c-miR signature. Our finding suggests that c-miR expression can classify high-risk cases in LS population, also when stratified for *MLH1*-variant or colorectal cancer, but further validation is required. This observation is valuable because the variation of cancer risk is high among LS carriers (34), and the implementation of intense screening programs is not uniformly effective (35). Therefore, a more nuanced approach is needed to identify those patients who are most likely to benefit from the screenings.

Cross-validations of the risk prediction models showed that c-miR risk sum scores have risk prediction potential also in randomly generated subsets with varying surveillance time and number of events. This finding is supported by previous research. For example, hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p, that were upregulated in those who developed cancer within the LS cohort, are well-recognized sporadic colorectal cancer miRs with multiple roles and reported biomarker potential (13, 14, 36–38). hsa-miR-3613-5p has been established as a colorectal cancer miR (39) whereas hsa-miR-3615 has been previously reported to display downregulation in microsatellite unstable colorectal tumors, which are hallmark tumors of LS, when compared with their microsatellite stable counterparts (40).

Furthermore, these five c-miRs displayed correlation as well as higher expression in those LS carriers who developed cancer compared those who did not, thus suggesting potential coregulation and biological connection. In support, we found that four out of the five c-miRs (hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, and hsa-miR-3615) have been experimentally shown to target several well-established oncogenes and tumor suppressor genes, including *BCL2*, *EGFR*, *CDKN1A*, *CDKN2A*, *CREB1*, *STAT3*, and *SMAD2* (41). Also, these genes formed interconnected nodes, which indicates similar role and biological connection among them and provide more support for the suggested coregulation of these c-miRs. All of these genes are part of cancer-relevant biological pathways, such as those in apoptosis, DNA damage, and cellular senescence (42).

Wikberg and colleagues observed that major changes of miR patterns occur mainly 3 years prior to sporadic colorectal cancer diagnosis by showing a temporal pattern of increase in miR-21-5p expression by using prediagnostic and postdiagnostic plasma samples (43). Raut and colleagues reported that a risk sum score of seven c-miRs was highly predictive for sporadic colorectal cancer risk in a prospective cohort with a follow-up time up to 14 years and median follow-up of 6.8 years (16). However, the c-miR signature we identified did not include any of the miRs observed by Raut and colleagues In contrast to sporadic colorectal cancer that develops commonly in 10–15 years, the development of LS colorectal cancer is significantly accelerated, often taking only 1 to 3 years to progress to carcinoma with or without pre-existing adenoma (44), which may explain the discrepancies between our study and the study by Raut and colleagues. As LS carriers in our study were diagnosed with cancer in 1.3 years on average from the serum sampling, it is possible that the observed c-miR signature originates from tumors. However, it is also possible that the observed c-miR levels may reflect risk rather than tumor presence because our sample was not limited to colorectal cancers. Nonetheless, these studies as well as our bioinformatics analyses show promising results for using c-miRs in LS cancer risk prediction.

Interestingly, we found a positive correlation between the c-miR risk sum and BMI suggesting a potential link between excess body weight, c-miRs, and cancer risk. In support to our findings, hsa-miR-10b-5p, hsa-miR-125b-5p, and hsa-miR-200a-3p have been previously linked with increased levels of plasma total cholesterol, dysregulated lipid metabolism, and overweight/obesity in general (45–47). Mens and colleagues, reported that upregulation of hsa-miR-10b-5p and hsa-miR-125b-5p associate with increased total cholesterol (45). Conversely, Ortega and colleagues reported a positive correlation between decreased levels of hsa-miR-125b and BMI after surgery-induced weight loss in obese patients (46). Ruiz-Roso and colleagues showed upregulated miR-200a to regulate lipid metabolism–related genes in a mouse model (47), although we did not find hsa-miR-200a-3p to target those genes. Moreover, Dogan and colleagues reported 1,558 miR-target gene interactions in obesity, including miR-125b, that were also detected in multiple cancer types. They also showed that metabolism and growth signaling pathways are shared by obesity and obesity-related cancer (48). Of the pathways reported by Dogan and colleagues, p53-signaling pathway was also identified in our study as a key pathway targeted by the c-miRs of the risk sum score. In addition, cellular senescence and FOXO pathways emerged in our analysis. These pathways have been reported to associate with cancer metabolism and obesity via alteration of energy metabolism and adipose tissue (49, 50). However, it is important to note that c-miRs have multifaceted roles in metabolism, and their profiles change with disease progression. Without mechanistic studies, it is challenging to exclude the potential confounding effects of disease and genetics in our findings. As metabolomic abnormality is an acknowledged cancer hallmark (42), these c-miRs could be promising targets to study when assessing the interactions of metabolic dysregulation and cancer.

A major strength of our study is that we were able to conduct an analysis using prediagnostic samples from a high-risk cohort under frequent surveillance. We also used robust methodology to interrogate c-miR signatures and their associations with LS cancer risk. All of the analyses were conducted carefully with state-of-the-art methods and tools. By utilizing Lasso regression, we were able to choose the most promising c-miRs and integrate them along with the surveillance time into well-established tool used for risk prediction, thus allowing comprehensive biomarker signature investigation. Missing values were handled with multiple imputation that is reported to have negligible bias when missingness occurs randomly (51). Finally, we followed TRIPOD checklist to enhance transparency in our risk prediction model development and validation as well as to improve reproducibility of these results.

As in many pilot experiments, the potential pitfall of our study is the small sample and effect size. Despite our best efforts to look for an external validation dataset, we unfortunately did not find a suitable candidate dataset nor had the opportunity to increase our sample size. For these reasons, we could not validate our predictor selection model. Because the majority of LS carriers are not most likely identified (44), and due to lack of resources, it is difficult to gather enough samples as well as it is costly to obtain enough small RNA-seq data for a more comprehensive investigation. An international collaboration study would be beneficial for such purposes. We also acknowledge that because the study population was comprised mainly of *MLH1* carriers, our results might have limited generalizability to other pathogenic MMR variant carriers. Finally, the average time period of 2.0 years between the lifestyle questionnaire data collection and blood sampling is also a potential limitation of this study.

To conclude, we report that a risk sum score composed of hsa-miR-10b-5p, hsa-miR-125b-5p, hsa-miR-200a-3p, hsa-miR-3613-5p, and hsa-miR-3615 has potential in LS cancer risk prediction, and thus may serve as a stratification biomarker signature for finding LS carriers at increased cancer risk in the future. However, more experiments with larger sample size are needed to confirm our findings. The molecular mechanisms underlying the associations of body weight, LS cancer risk and c-miRs remain to be elucidated in future studies.

Authors' Disclosures

T. Jokela reports grants from European Comission Union Marie Skłodowska-Curie Individual Fellowships during the conduct of the study. T.T. Seppälä reports grants from Finnish Medical Foundation, Emil Aaltonen Foundation, Jane and Aatos Erkko Foundation, Relander Foundation, and Cancer Foundation Finland during the conduct of the study; personal fees from Amgen Finland, personal fees and other support from LS CancerDiag, grants from Academy of Finland, and other support from Healthfund Finland outside the submitted work. E.K. Laakkonen reports grants from Päivikki and Sakari Sohlberg Foundation during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

T. Sievänen: Conceptualization, software, formal analysis, investigation, visualization, methodology, writing–original draft, writing–review and editing. **T. Jokela:** Software, investigation, writing–review and editing. **M. Hyvärinen:** Software, formal analysis, methodology, writing–review and editing. **T.-M. Korhonen:** Software, writing–review and editing. **K. Pylvänäinen:** Data curation, writing–review and editing. **J.-P. Mecklin:** Resources, data curation, funding acquisition, project administration, writing–review and editing. **J. Karvanen:** Conceptualization, software, methodology, writing–review and editing. **E. Sillanpää:** Conceptualization, supervision, writing–review and editing. **T.T. Seppälä:** Conceptualization, resources, data curation, supervision, funding acquisition, project administration, writing–review and editing. **E.K. Laakkonen:** Conceptualization, resources, data curation, supervision, funding acquisition, project administration, writing–review and editing.

Acknowledgments

We would like to extend our gratitude to prof. Sarianna Sipilä (Gerontology Research Center and Faculty of Sport and Health Sciences, University of Jyväskylä, Finland) for the lifestyle data acquisition. E.K. Laakkonen was supported by grants from Päivikki and Sakari Sohlberg Foundation. T. Jokela was supported by European Commission Union Marie Skłodowska-Curie Individual Fellowships (grant number: H2020-MSCA-IF-2020 #101026706). T.T. Seppälä was supported by funding from the Academy of Finland and iCAN Precision Medicine Flagship of Academy of Finland, and research grants by Jane and Aatos Erkko Foundation, Finnish Medical Foundation, Sigrid Juselius Foundation, Emil Aaltonen Foundation, Cancer Foundation Finland, Relander Foundation, and state research funding from the Finnish Goverment, which is allocated as competed grants through employing institutions to researchers within their university hospital co-operation area (Tampere University Hospital/Helsinki University Hospital).

Note

Supplementary data for this article are available at Cancer Prevention Research Online (http://cancerprevres.aacrjournals.org/).

Received September 6, 2023; revised January 3, 2024; accepted March 27, 2024; published first March 28, 2024.

References

1. Win AK, Jenkins MA, Dowty JG, Antoniou AC, Lee A, Giles GG, et al. Prevalence and penetrance of major genes and polygenes for colorectal cancer. Cancer Epidemiol Biomarkers Prev 2017;26:404–12.
2. Peltomäki P, Nyström M, Mecklin J-P, Seppälä TT. Lynch syndrome genetics and clinical implications. Gastroenterology 2023;5:783–99.
3. Sievänen T, Törmäkangas T, Laakkonen EK, Mecklin JP, Pylvänäinen K, Seppälä TT, et al. Body weight, physical activity, and risk of cancer in lynch syndrome. Cancers 2021;13:1849.
4. Jamizadeh N, Walton Bernstedt S, Haxhijaj A, Andreasson A, Björk J, Forsberg A, et al. Endoscopic surveillance of Lynch syndrome at a highly specialized center in Sweden: an observational study of interval colorectal cancer and individual risk factors. Front Oncol 2023;13:1127707.
5. Dashti SG, Win AK, Hardikar SS, Glombicki SE, Mallenahalli S, Thirumurthi S, et al. Physical activity and the risk of colorectal cancer in Lynch syndrome. Int J Cancer 2018;143:2250–60.
6. Dominguez-Valentin M, Sampson JR, Seppälä TT, ten Broeke SW, Plazzer JP, Nakken S, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the prospective lynch syndrome database. Genet Med 2020;22:15–25.
7. Burn J, Sheth H, Elliott F, Reed L, Macrae F, Mecklin J-P, et al. Cancer prevention with aspirin in hereditary colorectal cancer (Lynch syndrome), 10-year follow-up and registry-based 20-year data in the CAPP2 study: a double-blind, randomised, placebo-controlled trial. Lancet 2020;395:1855–63.
8. Mathers JC, Elliott F, Macrae F, Mecklin J-P, Möslein G, McRonald FE, et al. Cancer prevention with resistant starch in lynch syndrome patients in the CAPP2-randomized placebo controlled trial: planned 10-year follow-up. Cancer Prev Res 2022;15:623–34.
9. Mori MA, Ludwig RG, Garcia-Martin R, Brandão BB, Kahn CR. Extracellular miRNAs: from biomarkers to mediators of physiology and disease. Cell Metab 2019;30:656–73.
10. Goodall GJ, Wickramasinghe VO. RNA in cancer. Nat Rev Cancer 2021;21:22–36.
11. Jung G, Hernández-Illán E, Moreira L, Balaguer F, Goel A. Epigenetics of colorectal cancer: biomarker and therapeutic potential. Nat Rev Gastroenterol Hepatol 2020;17:111–30.
12. Shah R, Patel T, Freedman JE. Circulating extracellular vesicles in human disease. N Engl J Med 2018;379:958–66.
13. Yang G, Zhang Y, Yang J. A five-microRNA signature as prognostic biomarker in colorectal cancer by bioinformatics analysis. Front Oncol 2019;9:1207.
14. Zhang J, Song W, Chen Z, Wei J, Liao Y, Lei J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. Lancet Oncol 2013;14:1295–306.
15. Adam RS, Poel D, Ferreira Moreno L, Spronck JMA, de Back TR, Torang A, et al. Development of a miRNA-based classifier for detection of colorectal cancer molecular subtypes. Mol Oncol 2022;16:2693–709.
16. Raut JR, Schöttker B, Holleczek B, Guo F, Bhardwaj M, Miah K, et al. A microRNA panel compared to environmental and polygenic scores for colorectal cancer risk prediction. Nat Commun 2021;12:4811.
17. Bye A, Røsjø H, Aspenes ST, Condorelli G, Omland T, Wisløff U. Circulating MicroRNAs and aerobic fitness - the HUNT-study. PLoS One 2013;8:e57496.
18. Sievänen T, Korhonen T-M, Jokela T, Ahtiainen M, Lahtinen L, Kuopio T, et al. Systemic circulating microRNA landscape in Lynch syndrome. Int J Cancer 2023;152:932–44.
19. Kaartinen NE, Tapanainen H, Valsta LM, Similä ME, Reinivuo H, Korhonen T, et al. Relative validity of a FFQ in measuring carbohydrate fractions, dietary glycaemic index and load: exploring the effects of subject characteristics. Br J Nutr 2012;107:1367–75.
20. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45:1–67.
21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 2014;15:550.
22. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2009;26:139–40.
23. Tibshirani R. The Lasso method for variable selection in the Cox model. Stat Med 1997;16:385–95.
24. Harrell FEJ, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. JAMA 1982;247:2543–6.
25. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMC Med 2015;13:1–10.

26. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33:1–22.
27. Therneau TM, Grambsch PM. The Cox model BT - modeling survival data: extending the Cox model. In: Therneau TM, Grambsch PM, editors. New York (NY): Springer New York; 2000. p. 39–77.
28. Huang HY, Lin YCD, Li J, Huang KY, Shrestha S, Hong HC, et al. MiRTarBase 2020: updates to the experimentally validated microRNA-target interaction database. Nucleic Acids Res 2020;48:D148–54.
29. Sticht C, De La Torre C, Parveen A, Gretz N. miRWalk: an online resource for prediction of microRNA binding sites. PLoS One 2018;13:e0206239.
30. Snel B, Lehmann G, Bork P, Huynen MA. String: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. Nucleic Acids Res 2000;28:3442–4.
31. Wu G, Haw R. Functional interaction network construction and analysis for disease discovery. Methods Mol Biol 2017;1558:235–53.
32. Bracken CP, Scott HS, Goodall GJ. A network-biology perspective of microRNA function and dysfunction in cancer. Nat Rev Genet 2016;17:719–32.
33. Sur D, Advani S, Braithwaite D. MicroRNA panels as diagnostic biomarkers for colorectal cancer: a systematic review and meta-analysis. Front Med 2022;9:915226.
34. International Mismatch Repair Consortium. Variation in the risk of colorectal cancer in families with Lynch syndrome: a retrospective cohort study. Lancet Oncol 2021;22:1014–22.
35. Møller P, Seppälä T, Bernstein I, Holinski-Feder E, Sala P, Evans DG, et al. Cancer incidence and survival in Lynch syndrome patients receiving colonoscopic and gynaecological surveillance: first report from the prospective Lynch syndrome database. Gut 2017;66:464–72.
36. Sheedy P, Medarova Z. The fundamental role of miR-10b in metastatic cancer. Am J Cancer Res 2018;8:1674–88.
37. Pichler M, Ress AL, Winter E, Stiegelbauer V, Karbiener M, Schwarzenbacher D, et al. MiR-200a regulates epithelial to mesenchymal transition-related gene expression and determines prognosis in colorectal cancer patients. Br J Cancer 2014;110:1614–21.
38. Yamada A, Horimatsu T, Okugawa Y, Nishida N, Honjo H, Ida H, et al. Serum MIR-21, MIR-29a, and MIR-125b are promising biomarkers for the early detection of colorectal neoplasia. Clin Cancer Res 2015;21:4234–42.
39. Xiang F, Xu X. CirRNA F-circEA-2a suppresses the role of miR-3613–3p in colorectal cancer by direct sponging and predicts poor survival. Cancer Manag Res 2022;14:1825–33.
40. Slattery ML, Lee FY, Pellatt AJ, Mullany LE, Stevens JR, Samowitz WS, et al. Infrequently expressed miRNAs in colorectal cancer tissue and tumor molecular phenotype. Mod Pathol 2017;30:1152–69.
41. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res 2019;47:D941–7.
42. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov 2022;12:31–46.
43. Wikberg ML, Myte R, Palmqvist R, van Guelpen B, Ljuslinder I. Plasma miRNA can detect colorectal cancer, but how early? Cancer Med 2018;7:1697–705.
44. Seppälä TT, Burkhart RA, Katona BW. Hereditary colorectal, gastric, and pancreatic cancer: comprehensive review. BJS Open 2023;7:zrad023.
45. Mens MMJ, Maas SCE, Klap J, Weverling GJ, Klatser P, Brakenhoff JPJ, et al. Multi-omics analysis reveals MicroRNAs associated with cardiometabolic traits. Front Genet 2020;11:110.
46. Ortega FJ, Mercader JM, Catalán V, Moreno-Navarrete JM, Pueyo N, Sabater M, et al. Targeting the circulating microRNA signature of obesity. Clin Chem 2013;59:781–92.
47. Ruiz-Roso MB, Gil-Zamorano J, López de las Hazas MC, Tomé-Carneiro J, Crespo MC, Latasa MJ, et al. Intestinal lipid metabolism genes regulated by miRNAs. Front Genet 2020;11:707.
48. Dogan H, Shu J, Hakguder Z, Xu Z, Cui J. Elucidation of molecular links between obesity and cancer through microRNA regulation. BMC Med Genet 2020;13:161.
49. Smith U, Li Q, Rydén M, Spalding KL. Cellular senescence and its role in white adipose tissue. Int J Obes 2021;45:934–43.
50. Yadav RK, Chauhan AS, Zhuang L, Gan B. FoxO transcription factors in cancer metabolism. Semin Cancer Biol 2018;50:65–76.
51. White IR, Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. Stat Med 2010;29:2920–31.

III

BODY WEIGHT, PHYSICAL ACTIVITY, AND RISK OF CANCER IN LYNCH SYNDROME

by

Sievänen T, Törmäkangas T, Laakkonen EK, Mecklin J-P, Pylvänäinen K, Seppälä TT, Peltomäki P, Sipilä S, Sillanpää E, 2021

Cancers, *13*(8):1849

https://doi.org/10.1158/1940-6207.CAPR-23-0368

cancers

MDPI

Article

Body Weight, Physical Activity, and Risk of Cancer in Lynch Syndrome

Tero Sievänen [1,*,†], Timo Törmäkangas [1,†], Eija K. Laakkonen [1], Jukka-Pekka Mecklin [2,3], Kirsi Pylvänäinen [4], Toni T. Seppälä [5,6], Päivi Peltomäki [7], Sarianna Sipilä [1] and Elina Sillanpää [1,8]

1 Gerontology Research Centre and Faculty of Sport and Health Sciences, University of Jyväskylä, P.O. Box 35 (VIV), 40014 Jyväskylä, Finland; timo.tormakangas@jyu.fi (T.T.); eija.k.laakkonen@jyu.fi (E.K.L.); sarianna.sipila@jyu.fi (S.S.); elina.sillanpaa@jyu.fi (E.S.)
2 Department of Surgery, Central Finland Health Care District, 40620 Jyväskylä, Finland; jukka-pekka.mecklin@ksshp.fi
3 Faculty of Sport and Health Sciences, University of Jyväskylä, 40014 Jyväskylä, Finland
4 Department of Education, Central Finland Health Care District, 40620 Jyväskylä, Finland; kirsi.pylvanainen@ksshp.fi
5 Department of Surgical Oncology, Johns Hopkins University, Baltimore, MD 21218, USA; toni.t.seppala@hus.fi
6 Department of Surgery, Helsinki University Hospital, University of Helsinki, 00100 Helsinki, Finland
7 Department of Medical and Clinical Genetics, University of Helsinki, 00100 Helsinki, Finland; paivi.peltomaki@helsinki.fi
8 Institute for Molecular Medicine Finland, University of Helsinki, 00100 Helsinki, Finland
* Correspondence: tero.o.sievanen@jyu.fi
† These authors contributed equally to this work.

Citation: Sievänen, T.; Törmäkangas, T.; Laakkonen, E.K.; Mecklin, J.-P.; Pylvänäinen, K.; Seppälä, T.T.; Peltomäki, P.; Sipilä, S.; Sillanpää, E. Body Weight, Physical Activity, and Risk of Cancer in Lynch Syndrome. *Cancers* **2021**, *13*, 1849. https://doi.org/10.3390/cancers13081849

Academic Editors: Luca Roncucci and Nathan Berger

Received: 9 February 2021
Accepted: 7 April 2021
Published: 13 April 2021

Simple Summary: Lifestyle modifies cancer risk in the general public. How lifestyle modifies cancer risk in individuals carrying the inherited pathogenic gene variants in DNA mismatch repair genes (Lynch syndrome) remains understudied. We conducted a retrospective study with cancer register data to investigate associations between body weight, physical activity, and cancer risk among Finnish Lynch syndrome carriers (n = 465, 54% women). The results of our study indicated that longitudinal weight gain increases cancer risk, whereas being highly physically active during adulthood could decrease cancer risk in men. Further, women were observed to be less prone to lifestyle-related risk factors than men. The results emphasize the role of weight maintenance and high-intensity physical activity throughout the lifespan, especially in men with Lynch syndrome.

Abstract: Lynch syndrome (LS) increases cancer risk. There is considerable individual variation in LS cancer occurrence, which may be moderated by lifestyle factors, such as body weight and physical activity (PA). The potential associations of lifestyle and cancer risk in LS are understudied. We conducted a retrospective study with cancer register data to investigate associations between body weight, PA, and cancer risk among Finnish LS carriers. The participants (n = 465, 54% women) self-reported their adulthood body weight and PA at 10-year intervals. Overall cancer risk and colorectal cancer (CRC) risk was analyzed separately for men and women with respect to longitudinal and near-term changes in body weight and PA using extended Cox regression models. The longitudinal weight change was associated with an increased risk of all cancers (HR 1.02, 95% CI 1.00–1.04) and CRC (HR 1.03, 1.01–1.05) in men. The near-term weight change was associated with a lower CRC risk in women (HR 0.96, 0.92–0.99). Furthermore, 77.6% of the participants retained their PA category over time. Men in the high-activity group had a reduced longitudinal cancer risk of 63% (HR 0.37, 0.15–0.98) compared to men in the low-activity group. PA in adulthood was not associated with cancer risk among women. These results emphasize the role of weight maintenance and high-intensity PA throughout the lifespan in cancer prevention, particularly in men with LS.

Keywords: epidemiology; hereditary non-polyposis colorectal cancer; lifestyle

1. Introduction

Colorectal cancer (CRC) is the third most commonly diagnosed cancer in Europe, with an estimated 499,667 new cases in 2018, and is the second most common cause of cancer mortality [1]. Approximately 3–5% of all CRC cases may be due to hereditary cancer syndrome, also known as Lynch syndrome (LS) [2], which is caused by pathogenic germline variants in DNA mismatch repair genes (*path_MMR*): *MLH1*, *MSH2*, *MSH6*, or *PMS2* [3]. Individuals with *path_MMR* variants are at considerably greater risk of developing CRC (40–80%), endometrial cancer (40–60%) and various other cancers compared to the general population [4,5]. However, cancer risk is highly variable among different mutation carriers and thus far it is not known why not all *path_MMR* carriers develop cancer, whereas others develop cancer at a young age and/or suffer from multiple different types of cancers during their lifespan.

There is strong evidence derived from the general population that increased physical activity and reduced body adiposity are associated with decreased cancer risk [6–12]. Recent research suggests that LS CRC risk can also be moderated by these lifestyle factors [13–17], but the number of studies that investigate the associations between lifestyle and LS cancer risk are scarce. Currently, there are only two studies that we are aware of that have assessed the association of physical activity and LS cancer risk [13,14] and they have not taken into consideration the fact that this potential association may vary during a lifespan. In addition, the evidence regarding the associations of cancer risk and obesity and overweight is contradictory among women with LS [15], and thus, the risk analyses should be performed separately for both sexes. It is of great importance to identify modifiable behavioural factors of LS cancer risk to motivate variant carriers to change suboptimal conduct or maintain a healthy lifestyle. Lifestyle modification could efficiently aid in reducing individual cancer risk despite a strong genetic predisposition.

In this study, we hypothesized that the interplay between genetic factors and lifestyle is associated with variable cancer risk in a distinct high-risk population. Two founder mutations in *MLH1* are found in a major proportion of Finnish LS families [18–20], which offers a possibility for investigating lifestyle factors as a modifier of cancer risk in a relatively homogenous LS population and hence may limit the influence of genetic discrepancies. The aim of this retrospective study with longitudinal lifestyle and cancer register data was to investigate associations between body weight and physical activity on CRC and overall cancer risk among adult Finnish *path_MMR* men and women. We modeled the recalled levels of weight and physical activity as time-dependent variables in the relative risk model.

2. Results

2.1. Descriptive Statistics

Descriptive data on *path_MMR* variants, cancer history, and lifestyle and socioeconomic characteristics are presented in Table 1. Of the 465 participants, 215 (46.2%) were men and 250 (53.8%) were women. The mean age at the time of data collection was 56.4 years vs. 57.4 years, respectively.

Table 1. Descriptive demographic, genetic, cancer history, socioeconomic, and lifestyle-related characteristics of the study population.

Background Variable	Men		Women	
	Cancer	Healthy	Cancer	Healthy
	***N* = 101**	***N* = 114**	***N* = 122**	***N* = 128**
Age at data collection (years (SD))	64.2 (11.2)	49.5 (14.0)	65.4 (9.6)	49.7 (14.0)
Age at first cancer (years (SD))	45.9 (10.4)		48.6 (29–79)	
Age at first CRC (years (SD))	45.6 (11.1)		48.4 (29–79)	
Cancers diagnosed (*n* (%))				
CRC	89 (88.1)		77 (63.1)	
Endometrial cancer	17 (16.8)		57 (46.7)	

Table 1. *Cont.*

Background Variable	Men		Women	
	Cancer	Healthy	Cancer	Healthy
	***N* = 101**	***N* = 114**	***N* = 122**	***N* = 128**
Other cancers [a]			30 (24.6)	
MMR gene affected (*n* (%))				
MLH1	50 (49.5)	56 (49.1)	50 (41.0)	64 (50.0)
MLH1 other than exon 16 deletion	24 (23.8)	26 (22.8)	35 (28.7)	34 (26.6)
MSH2	15 (14.9)	22 (19.3)	26 (21.3)	13 (10.2)
MSH6	10 (9.9)	10 (8.8)	11 (9.0)	17 (13.3)
PMS2	2 (2.0)			
Socioeconomic characteristics				
Education (*n* (%))				
Basic education	20 (19.8)	13 (11.4)	36 (29.5)	16 (12.5)
Upper secondary degrees	46 (45.6)	60 (52.6)	51 (41.8)	63 (49.4)
Polytechnic degree	13 (12.9)	18 (15.8)	11 (9.0)	19 (14.8)
University degree	22 (21.8)	23 (20.2)	24 (19.7)	30 (23.4)
Occupational status (*n* (%)) [#]				
Worker/employee	40 (39.6)	76 (66.7)	47 (39.2)	92 (72.4)
Retired	45 (44.6)	18 (15.8)	56 (46.7)	22 (17.3)
Other [b]	16 (15.8)	20 (17.5)	17 (14.2)	13 (10.3)
Marital status (*n* (%))				
Living alone	20 (19.8)	22 (19.3)	40 (32.8)	32 (25.0)
Married/cohabitation	81 (80.2)	92 (80.7)	82 (67.2)	96 (75.0)
Perceived health and physical fitness				
Self-rated health (*n* (%))				
Poor	16 (15.8)	13 (11.4)	15 (12.3)	19 (14.8)
Average	39 (38.6)	27 (23.7)	49 (40.2)	32 (25.0)
Good	46 (45.5)	74 (64.9)	57 (46.7)	77 (60.2)
Self-rated physical fitness (*n* (%))				
Poor	22 (21.8)	18 (15.8)	16 (13.1)	21 (16.4)
Average	39 (38.6)	29 (25.4)	50 (41.0)	40 (31.3)
Good	40 (39.6)	67 (58.8)	56 (45.9)	67 (52.3)
Lifestyle variables				
Alcohol consumption (portions/week (SD)) [#]	4.8 (6.5)	4.8 (5.3)	2.3 (3.5)	3.0 (4.3)
Smoking status (*n* (%)) [#]				
Never	43 (42.6)	45 (39.5)	61 (50.0)	64 (50.0)
Former	47 (46.5)	48 (42.1)	52 (42.6)	46 (35.9)
Current	11 (10.9)	21 (18.4)	8 (6.6)	17 (13.3)
Use of anti-inflammatory drugs (*n* (%)) [#]				
No	83 (85.6)	89 (78.1)	88 (72.7)	84 (65.6)
Yes	14 (14.4)	25 (21.9)	33 (27.3)	44 (34.4)
Current physical activity (*n* (%)) [#]				
Low	28 (28.0)	23 (20.2)	23 (18.9)	21 (16.4)
Medium	32 (32.0)	24 (21.1)	44 (36.1)	47 (36.7)
High	40 (40.0)	67 (58.8)	55 (45.1)	60 (46.9)
BMI (kg/m^2 (SD)) [#]	27.2 (5.3)	26.6 (4.2)	27.1 (5.8)	27.6 (11.6)
BMI categories (*n* (%)) [#]				
Underweight	2 (2.0)	2 (1.8)	1 (0.8)	2 (1.6)
Normal weight	40 (40.0)	38 (33.3)	48 (40.7)	56 (44.1)
Overweight	34 (34.0)	57 (50.0)	39 (33.1)	34 (26.8)
Obese	24 (24.0)	17 (14.9)	30 (25.4)	35 (27.6)
Waist circumference (cm (SD)) [#]	100.5 (14.4)	97.8 (11.0)	90.6 (14.4)	88.6 (14.5)

[#] Missing data: occupational status *n* = 3, alcohol consumption *n* = 3, smoking status *n* = 2, anti-inflammatory drugs *n* = 5, BMI and BMI categories *n* = 6, current physical activity *n* = 1, waist circumference *n* = 8. [a] other cancers included breast cancer, ovarian cancer, prostate cancer and skin cancers; [b] other included students, unemployed, and persons on parental leave. BMI = body mass index; CRC = colorectal cancer.

Almost half of the men (47.0%) and women (49.0%) had had one or more cancers prior to lifestyle data collection. The most common cancer was CRC, with a higher prevalence in men (88.1%) compared to women (63.1%). The mean age at the first cancer incidence was 45.9 years in men and 48.6 years in women, and 45.6 years and 48.4 years at the age of the first CRC, respectively. MMR-gene variant frequencies for the entire study population were 47.3% for *MLH1*, 25.6% for *MLH1* other than ex 16, 16.3% for *MSH2*, 10.3% for *MSH6*, and 0.4% for *PMS2*.

Healthy participants were more often working, whereas participants with cancer were more often retired. Women with cancer had a lower level of education and they were more commonly living alone compared to their healthy counterparts. Self-rated health status and fitness, as well as current physical activity level, was in general better among participants who had not had cancer. In addition, other lifestyle variables were moderately similar among participants with and without cancer.

Table 2 describes the cumulative cancer event history of the entire study population during retrospective follow-up separately for both sexes. Among both men and women, most of the cancer events occurred from the age of 40 years to 70 years.

Table 2. Summary of individuals at risk for cancer, events, and censorings occurring at the beginning of 10-year periods for men and women.

Sex	Period	At Risk	All Cancers		CRC	
			Events	Censored	Events	Censored
Men						
	(20, 30)	215	0	0	0	0
	(30, 40)	199	5	11	5	11
	(40, 50)	156	30	29	25	34
	(50, 60)	89	65	61	55	71
	(60, 70)	48	83	84	69	98
	(70, 77)	8	98	109	80	127
	77	0	98	117	80	135
Women						
	(20, 30)	250	0	0	0	0
	(30, 40)	242	1	7	1	7
	(40, 50)	190	25	35	19	41
	(50, 60)	117	67	66	40	93
	(60, 70)	55	100	95	55	140
	(70, 80)	14	117	119	63	173
	(80, 85)	1	120	129	65	184
	85	0	120	130	65	185

Period: A 10-year interval (years). At risk: participants at risk (*n*) at each 10-year interval. Events: cumulative *n* of occurred cancer events. Censored: cumulative *n* of censored participants. CRC = colorectal cancer. Note: for time interval lower limit includes the indicated value and the value for the upper limit is excluded.

2.2. Body Weight History

Table 3 describes the changes in mean body weight during the retrospective follow-up. The mean body weight increased throughout the lifespan in both sexes. Furthermore, the mean individual change in body weight was positive in each 10-year interval, both among men and women during the adult lifespan. From the age of 40 years onwards, average individual weight increased with respect to recalled weight at the age of 20 years and ranged between 8 kg and 12 kg per year for men and between 8 kg and 13 kg for women.

Table 3. Means and standard deviations for weight measurements (kg) and individual weight change in 10-year periods among participants with Lynch syndrome.

Sex	*N*	Period	At Risk, Start of Period		Individual Change, Near-Term		Individual Change, Longitudinal	
			Mean	SD	Mean	SD	Mean	SD
Men								
	215	(20, 30)	72.65	11.93	-	-	-	-
	199	(30, 40)	77.06	12.68	5.06	5.99	5.06	5.99
	156	(40, 50)	79.56	14.91	3.76	7.05	8.23	10.29
	89	(50, 60)	79.67	12.43	4.03	6.73	10.61	10.52
	48	(60, 70)	77.05	11.76	0.72	4.10	8.22	9.25
	8	(70, 77)	83.40	15.37	4.60	7.09	12.20	12.28
Women								
	250	(20, 30)	59.16	11.28	-	-	-	-
	242	(30, 40)	62.88	13.48	3.96	6.59	3.96	6.59
	190	(40, 50)	66.30	14.78	3.98	5.86	7.53	8.72
	117	(50, 60)	68.30	13.65	2.83	6.11	9.33	10.29
	55	(60, 70)	68.98	11.58	2.85	4.41	9.39	9.24
	14	(70, 80)	73.91	12.84	4.73	9.02	13.18	11.97
	1	(80, 85)	79.00	- [a]	0.00	- [a]	20.00	- [a]

[a] The variation is estimable only for one case. Period: A 10-year interval (years). Longitudinal change: change in body weight (kg) relative to the body weight at the age of 20 years. Near-term change: change in body weight (kg) relative to the body weight at the previous 10-year interval before diagnosis or censoring. Mean (kg). SD = standard deviation (kg). Note: for time interval lower limit includes the indicated value and the value for the upper limit is excluded.

Table 4 presents the associations between adult life body weight and cancer risk. For consistency, all results in the text regarding body weight and cancer risk are presented only from models adjusted for height, MMR gene, education, alcohol consumption, smoking status, and the use of anti-inflammatory drugs.

Table 4. Hazard ratios for associations of body weight and cancer risk in participants with Lynch syndrome.

Setting	Unadjusted Model				Adjusted Model *			
	Cancer Events	Observations	HR (95% CI)	*p*-Value	Cancer Events	Observations	HR (95% CI)	*p*-Value
Men								
All cancers								
Longitudinal change	77	610	1.02 (1.00–1.03)	0.048	74	579	1.02 (1.00–1.04)	0.022
Near-term change	77	185	0.99 (0.97–1.00)	0.345	74	174	0.99 (0.97–1.01)	0.424
Colorectal cancer								
Longitudinal change	60	610	1.02 (1.00–1.03)	0.023	57	579	1.03 (1.01–1.05)	0.004
Near-term change	60	185	1.00 (0.98–1.01)	0.695	57	174	1.00 (0.98–1.02)	0.861
Women								
All cancers								
Longitudinal change	95	758	0.99 (0.97–1.00)	0.290	91	720	1.00 (0.98–1.02)	0.887
Near-term change	95	221	0.98 (0.97–1.00)	0.059	91	209	0.98 (0.96–1.00)	0.059
Colorectal cancer								
Longitudinal change	50	758	0.99 (0.96–1.01)	0.258	48	720	0.99 (0.96–1.02)	0.454
Near-term change	50	221	0.98 (0.95–1.00)	0.106	48	209	0.96 (0.92–0.99)	0.015

* Model adjusted for height, MMR-gene, education, smoking, alcohol consumption, and use of anti-inflammatory drugs. Longitudinal change: longitudinal change in body weight from the age of 20 years until the first cancer diagnosis or censoring. Near-term change: age-stratified change in body weight relative to the body weight at the previous 10-year interval before diagnosis or censoring. *p*-values statistically significant at <0.05 level. Statistically significant hazard ratios are highlighted in bold. Cancer events: number of occurred cancers. Observations: number of observations across each 10-year interval. HR = hazard ratio. CI = confidence interval.

2.2.1. Risk of All Cancers

A change in longitudinal weight throughout the lifespan, calculated per one-kg weight increase, was associated with a 2% increased risk of cancers in men (HR 1.02,

95% CI 1.00–1.03), whereas no such association was observed in women. Moreover, near-term weight change had no impact on the risk of all cancers in either sex.

2.2.2. Risk of CRC

Longitudinal weight gain increased the CRC risk by 3% in men only (HR 1.03, 95% CI 1.01–1.05). Among women, near-term weight gain within the 10-year interval before cancer diagnosis was associated with a 4% decreased risk of CRC (HR 0.96, 95% CI 0.92–0.99). No associations between near-term weight gain and CRC risk were observed in men.

2.3. Physical Activity during Adulthood

Table 5 describes changes in the intensity of physical activity during the retrospective follow-up. In both sexes, a great majority retained their activity category over time. However, when the category changed, it was more common among men to move from the organized physical activity participation category to the lower activity category.

Table 5. Frequencies and cross-period tables for participation in organized physical activity in 10-year periods among participants with Lynch syndrome.

Sex	Period	Physical Activity	*N* (%)	Near-Term Change		Longitudinal Change	
				Low Activity	High Activity	Low Activity	High Activity
Men	(20, 30)	Low activity	155 (76)	-	-	-	-
		High activity	50 (24)	-	-	-	-
	(30, 40)	Low activity	153 (81)	131 (94)	22 (44)	131 (94)	22 (44)
		High activity	37 (19)	9 (6)	28 (54)	9 (6)	28 (56)
	(40, 50)	Low activity	129 (86)	121 (98)	8 (30)	104 (95)	25 (62)
		High activity	21 (14)	2 (2)	19 (70)	6 (5)	15 (38)
	(50, 60)	Low activity	77 (91)	70 (100)	7 (47)	59 (97)	18 (75)
		High activity	8 (9)	0 (0)	8 (53)	2 (3)	6 (25)
	(60, 70)	Low activity	43 (96)	42 (100)	1 (33)	31 (100)	12 (86)
		High activity	2 (4)	0 (0)	2 (67)	0 (0)	2 (14)
	(70, 77)	Low activity	8 (100)	8 (100)	0 (-)	6 (100)	2 (100)
		High activity	0 (0)	0 (0)	0 (-)	0 (0)	0 (0)
Women	(20, 30)	Low activity	180 (76)	-	-	-	-
		High activity	58 (24)	-	-	-	-
	(30, 40)	Low activity	166 (72)	151 (87)	15 (26)	151 (87)	15 (26)
		High activity	64 (28)	22 (13)	42 (74)	22 (13)	42 (74)
	(40, 50)	Low activity	121 (68)	114 (87)	7 (15)	111 (92)	31 (53)
		High activity	58 (32)	17 (13)	41 (85)	10 (8)	27 (47)
	(50, 60)	Low activity	86 (77)	75 (96)	11 (32)	82 (95)	13 (50)
		High activity	26 (23)	3 (4)	23 (68)	4 (5)	13 (50)
	(60, 70)	Low activity	40 (78)	38 (97)	2 (17)	38 (95)	5 (45)
		High activity	11 (22)	1 (3)	10 (83)	2 (5)	6 (55)
	(70, 77)	Low activity	10 (83)	10 (100)	0 (0)	10 (100)	1 (50)
		High activity	2 (17)	0 (0)	2 (100)	0 (0)	1 (50)

Period: A 10-year interval (years). Physical activity: physical activity group at the beginning of the 10-year interval. Longitudinal change: change in physical activity group (*n*) relative to the physical activity group at the age of 20 years. Near-term change: change in physical activity group (*n*) relative to the physical activity group at the previous 10-year interval. Note: for time interval lower limit includes the indicated value and the value for the upper limit is excluded.

The associations between physical activity and cancer risks are presented in Table 6. For consistency, all results in the text regarding physical activity and cancer risk are presented only from adjusted models.

Table 6. Hazard ratios for associations of physical activity and cancer risk in participants with Lynch syndrome.

Setting	Unadjusted Model				Adjusted Model *			
	Cancer Events	Observations	HR (95% CI)	*p*-Value	Cancer Events	Observations	HR (95% CI)	*p*-Value
Men								
All cancers								
Longitudinal change	91	683	0.44 (0.19–1.04)	0.063	86	648	0.37 (0.15–0.98)	0.044
Near-term change	91	205	0.69 (0.29–1.64)	0.403	86	192	0.74 (0.27–2.01)	0.557
Colorectal cancer								
Longitudinal change	73	683	0.57 (0.25–1.33)	0.194	68	648	0.52 (0.20–1.36)	0.181
Near-term change	73	205	0.93 (0.39–2.24)	0.874	68	192	0.99 (0.36–2.73)	0.983
Women								
All cancers								
Longitudinal change	110	823	1.31 (0.86–1.97)	0.206	107	789	1.26 (0.79–2.00)	0.341
Near-term change	110	238	1.42 (0.89–2.25)	0.138	107	227	1.34 (0.80–2.23)	0.268
Colorectal cancer								
Longitudinal change	59	823	1.16 (0.65–2.10)	0.612	57	789	1.28 (0.65–2.52)	0.471
Near-term change	59	238	0.92 (0.49–1.72)	0.797	57	227	0.99 (0.48–2.02)	0.973

* Model adjusted for MMR-gene, education, smoking, alcohol consumption, and use of anti-inflammatory drugs. In all analyses, the reference group was *Low activity.* Longitudinal change: longitudinal change in physical activity from the age of 20 years until the first cancer diagnosis or censoring. Near-term change: age-stratified change in physical activity level relative to the physical activity level at the previous 10-year interval before diagnosis or censoring. *p*-values statistically significant at the <0.05 level. Statistically significant hazard ratios are highlighted in bold. Cancer events: number of occurred cancers. Observations: number of observations across each 10-year interval. HR = hazard ratio. CI = confidence interval.

2.3.1. Risk of All Cancers

Men in the high activity group were found to have a reduced longitudinal cancer risk of 63% (HR 0.37, 95% CI 0.15–0.98) compared to men in the low activity group. There were no longitudinal associations between physical activity and cancer risk observed in women. In the near-term, participating in physical activity had no impact on cancer risk in either men or women.

2.3.2. Risk of CRC

There was no association between physical activity and the risk of CRC.

3. Discussion

We conducted a retrospective study with longitudinal data collection and cancer register data among Finnish *path_MMR* carriers to elucidate the associations between changes in adult body weight, physical activity, and cancer risk. Our results suggest that associations between lifestyle and cancer risk differ between men and women and may vary during the course of life. We found that an overall increase in total body weight throughout the lifespan slightly elevated the risk of cancers, including CRC, in men. We also observed that men who continued to participate in more intensive physical activities over their adult life were at lower risk of all cancers.

In Western societies, body weight typically accumulates during the adult lifespan and growing levels of obesity predispose individuals to multiple health complications. Obesity is acknowledged as one of the most important risk factors of non-communicable diseases [9,21,22]. It is well established in the general population that excess body weight and adiposity, particularly in overweight individuals, is an important risk factor for several cancers [8] and the risk could be reduced by lowering excess body mass [23]. In the population of *path_MMR* carriers of the current study, participants increased their body weight during their adult years, and 61% of men and 55% of women were overweight or obese. Obesity and overweight may be more harmful for men, as we found that a trend of body weight accumulation during adult years was associated with an increased cancer risk in men but not in women. This is in agreement with other reports which have investigated weight accumulation in relation to the risk of LS cancer [17,24].

Unfortunately, in most of the studies, including ours, a lifelong change in body composition could only be determined by changes in body weight. However, aging is not

only associated with an increase in body weight that is related to fat accumulation but also to changes in body composition [25]. Beginning from the age of 30 years, muscle mass tends to decrease and the decline accelerates after the age of 50 years, particularly among women due to menopause [26]. Concurrently, the amount and distribution of body fat may change, thereby resulting in the accumulation of fat—particularly in visceral areas—which increases cancer risk through several already identified biological pathways. The best-characterized association is between abdominal obesity and disturbed insulin metabolism, which may influence cancer risk through cell proliferation and apoptosis [10,27]. Age-related trends in body weight accumulation are different between men and women, which may explain our dissimilar findings regarding weight accumulation and cancer risk. In general, men are more prone to increased android-type fat distribution—that is, abdominal fat distribution—throughout their lifespan [28]. In contrast, women tend to be more prone to gynoid-type fat distribution during their premenopausal years and then shift to androgen-type fat accumulation in their postmenopausal years [29]. In our study, the weight accumulation was lower among women, but we cannot exclude the potential confounding role of menopause on the association between body weight and cancer risk as our data did not include information regarding the menopausal status of the female participants. Based on the population averages of menopause age being between 50 years and 53 years, we can estimate that 67% of women participating in this study were post-menopausal, but the considerable individual variation observed makes this estimate imprecise [30]. Nevertheless, visceral adiposity has been associated with higher cancer risk in both sexes, although in women the risk estimates appear to differ between pre- and post-menopausal women [31].

Intriguingly, we also found that near-term weight increase had a CRC-protective effect in women. To speculate, it is possible that hormonal factors might have influenced the risk estimates. The primary source of endogenous estrogen in post-menopausal women has been suggested to be adipose tissue [32,33]. Therefore, higher adiposity could maintain a higher systemic estrogen level, which in turn may provide some protection through (for instance) the anti-inflammatory action of estrogens [34]. However, this is highly speculative. There is evidence that exogenous estrogen use can be cancer protective, neutral or to increase the risk of different cancers [35–37]. In the current study, we did not investigate the potential role of hormone therapy, nor did we have the ability to measure estrogen levels; thus, we cannot exclude the role of systemic estrogen level. Therefore, as the great majority of CRC-diagnosed women in our study were over 60 years of age, there remains a possibility that the plausible protective effects of estrogen derived from adipose tissue might have masked the effect of weight gain on colorectal cancer risk.

Our results suggest that performing more vigorous guided physical activity exhibit a cancer-preventive effect in men, as those who continued at higher levels of physical activity were at a 63% lower cancer risk when compared to less active men performing non-guided physical activity. To date, we are aware of only two previous studies that have assessed the impact of physical activity on LS cancer risk [13,14]. Both designs were retrospective like our study but did not assess risk estimates separately for men and women. Kamiza et al. (2015) [13] reported that among 301 Taiwanese individuals (51.8% women) carrying *path_MLH1* and *path_MSH2* regular vigorous leisure time physical activity over a year prior to cancer diagnosis decreased CRC risk by 38% when compared to those who did not indulge in any such activity. Although we did not observe associations between physical activity and cancer risk in the near-term like Kamiza et al. (2015) [13], similarly to their findings, our results also suggested that performing vigorous physical activity could reduce cancer risk.

Further, the study by Dashti et al. (2018) [14] comprised 2042 *path_MMR* carriers (57% women). As in our study, they modeled longitudinal and near-term changes in physical activity separately. Unlike our study, Dashti et al. (2018) [14] did not find an association between physical activity and cancer risk when assessed over several age periods, even though a trend of lowering the risk of CRC was observed. In addition,

they used MET-h/week to assess the amount of physical activity, but we did not do so. Near-term cancer risk assessment, which was used in the current study, may be particularly effective for identifying specific suboptimal lifestyle changes that could be associated with carcinogenesis and precede cancer occurrence. Dashti et al. (2018) [14] found higher levels of near-term physical activity (>35 MET-h/week) to be protective against such a risk, whereas we did not find an association between near-term physical activity and cancer risk (any cancer or CRC).

We did not find longitudinal or near-term evidence linking physical activity to cancer risk in *path_MMR* women, which may be due to differences in physical activity behavior between men and women, including the intensity and type of physical activity, as well as the timing of physical activity exposure in life. For example, throughout the follow-up, female participants mainly reported participating more frequently in guided leisure time physical activity than men, who reported performing competitive sports more often (Table S4). Overall, our results highlight the potential role of physical activity in cancer prevention among *path_MMR* carrier men, as already advocated in clinical guidelines for LS [38]. Although most *path_MMR* carriers suffer from cancer at some point of their life, this is an important finding.

Various reports in the extant literature have suggested several mechanisms that link physical activity with a reduced cancer risk [11]. For example, physical activity produces multiple beneficial changes in cardiorespiratory systems [39], and being physically active also helps with weight control, as well as with reducing excess adiposity [12]. These combined effects might have a beneficial impact on biological mechanisms that interact directly or indirectly with cancer, such as improved insulin sensitivity and reduced chronic low-level inflammation, which is also linked with favorable immunomodulation [40,41]. However, the existing evidence originates from sporadic cancer patients who could differ from *path_MMR* carriers with respect to disease mechanisms, carcinogenesis, and biological regulation.

As described previously, the association of decreased cancer risk and healthy lifestyle in the general population has been observed in large population-based studies. Since the influence of healthy behaviour on decreased cancer risk was observed in our study with a limited number of participants, it could be possible that the effect of the modifiable behavioral risk factors—physical activity and body weight—is emphasised in *path_MMR* carriers due to their strong genetic predisposition to cancers. Therefore, it is important to follow these modifiable risk factors among *path_MMR* carriers during their regular healthcare visits. An optimal lifestyle could partially compensate for the strong genetic predisposition to cancers and thus help in cancer prevention. Nowadays, individual cancer risk can be calculated and demonstrated via online tools (www.plsd.eu, accessed 9 January 2021), which can be used to improve motivational support for healthy lifestyle maintenance or lifestyle changes.

A major strength of the current study is that the study cohort comprised participants who had undergone comprehensive screenings of LS-predisposing mutations, with ascertainment utilizing Amsterdam and Bethesda clinical criteria and cascade testing, and those who had been offered colonoscopy surveillance at 2–3-year intervals. Our body weight and physical activity data collection encompassed the entire adult lifespan and was carefully analyzed by considering potential time-varying risks, sex differences, and potential confounders regarding cancer incidence. The observation period was initiated from the age of 20 years, instead of birth, to avoid the detection of changes in body weight which were merely due to natural growth and maturation. In doing so, we were also able to exclude the time-period when cancer incidence tends to be extremely low even among *path_MMR* carriers. We also chose to model the change in cancer risk in the near-term setting (during the age-period of cancer or censoring) as it could be more accurate than longitudinal change, which could be influenced by poor recall. We also used time-dependent covariate values, which allow us to account for changes in predictor values over time.

However, there are also several limitations. Weight and physical activity were assessed using self-recall instruments. Even though a recent study found that cross-sectional self-reported measurements of BMI were reasonably close to recent direct measurements [42], the recall of weight in the more distant past has lower reliability [43] and for some of our older participants, the recall time was several decades. Moreover, there might also be sex-based discrepancies, as women tend to underestimate their weight and men tend to overestimate it [44]. Finally, Smith et al. (2013) [45] found the recall of physical activity of the distant past to be moderately reproducible, but poor at the individual level. Taken together, we cannot exclude the possibility that recall bias might have influenced the risk estimates.

4. Materials and Methods

4.1. Study Sample

The study cohort included those carriers of *path_MMR* who were registered in the Finnish Lynch Syndrome Research Registry (LSRFi; www.lynchsyndrooma.fi, accessed 16 June 2020) and provided consent for research-related contacts. LSRFi is a nation-wide research registry (est. 1982) operating in Jyväskylä and Helsinki that organizes surveillance and cancer prevention for LS families. Currently, the registry consists of clinical and family history data of over 300 LS families and over 1700 pathogenic variant carriers under frequent surveillance. Individuals were identified in the registry before the genetic testing became available, based on clinical criteria (Amsterdam and Bethesda criteria) [46,47], and subsequently through cascade testing of the families and universal testing of tumors. Adult members of LSRFi with confirmed *path_MMR* variants (classes 4 and 5 by InSiGHT criteria) [48] were eligible for the study.

4.2. Cancer Register Data

Age, sex, and all cancer diagnoses with the cancer type and date of each diagnosis, mutation status, and family cancer history were confirmed from hospital medical records and national cancer registries upon recording in the LSRFi. With regard to analyses, participants in the cancer group were required to have at least one past cancer diagnosis in the medical registries although he/she could have been healthy at the time of data collection. The healthy group included only *path_MMR* carriers who had remained cancer free until data collection.

4.3. Questionnaire Data Collection

Questionnaires for anthropometric, socioeconomic, and lifestyle data collection were sent to 1038 adult *path_MMR* carriers whose addresses were available in LSRFi in December 2016 and July 2020. Of them, 480 (response rate 46.2%) returned the questionnaire. However, 15 participants did not carry the *path_MMR* variant and therefore they did not fulfil the eligibility criteria and were excluded from the study. Then, the final study sample included 465 participants.

4.4. Descriptive Variables

4.4.1. Socioeconomic Characteristics

The education level was categorized according to the Finnish schooling system into four categories: basic education (including elementary or comprehensive school), upper secondary education (including vocational school and high school level degrees), polytechnic degree, and university degree. Occupational status included the categories of worker or employee, retiree (including both disability and old age pensioners), and other (including students, unemployed people, and people on parental leave). Marital status was categorized as living alone or married/cohabitating.

4.4.2. Perceived Health and Physical Fitness

Self-rated health and physical fitness were collected using standard five-scale questions. Due to the low number of responses in a few categories, answers were re-categorized into poor (also including very poor), average, and good (also including very good) for statistical analysis.

4.4.3. Lifestyle Variables

The level of alcohol consumption was identified by two questions assessing the frequency of alcohol use and the number of alcohol portions consumed per occasion. One portion refers to 10–14 g alcohol, which one may obtain, for example, from a single serving of 0.33 liters of beer or a similar light alcoholic beverage, from 12 cL of wine, or from 4 cL of spirits. Furthermore, the subjects' smoking status was defined as never smoker if they reported being non-smokers and having never been a smoker, or smoked <100 cigarettes during their entire life; as former smoker if they reported currently being non-smokers but were regular smokers in the past; or as current smoker if they reported being current and regular smokers. Participants were also asked whether or not they used any anti-inflammatory drugs regularly during the surveyed time period (yes/no). Current leisure-time physical activity was assessed via the seven-option scale question [49,50]. The scale options were re-categorized into low (light walking and outdoor activities 1–2 times per week), medium (some light walking and outdoor activities several times a week, or engaging in brisk physical activity 1–2 times per week causing some shortness of breath and perspiration), and high (brisk physical activity 3–5 times a week causing some shortness of breath and perspiration or fitness training several times a week causing heavy perspiration and being out-of-breath during exercise or playing competitive sports and maintaining regular fitness).

4.4.4. Anthropometrics

Anthropometrics were self-measured. Participants were asked to report their height and to measure their body weight before breakfast and without clothes. If weight could not be measured, participants were asked to fill in the last weight measurement known. BMI was calculated as weight in kilograms divided by the height squared in meters (kg/m^2). Participants were categorized into four BMI groups—underweight (BMI < 18.5), normal weight (BMI 18.5–24.9), overweight (BMI 25–29.9), and obese (BMI $\geq$ 30)—according to the WHO classifications. Measuring tape was sent with the questionnaire to examine the waist circumference, along with written instructions. Waist circumference was measured without clothing in a standing position, 2 cm above the umbilicus.

4.5. Outcome Variables for Life-Long Exposures

To measure body weight history during adulthood, participants recalled their body weight in kilograms at age 20, 30, 40, 50, 60, and up to 70+ years.

Physical activity during adulthood was assessed via four-option scale questions through which participants recalled the level of regular physical activity they had at different adult age ranges throughout their lives [51]. Participants reported their past physical activity at age ranges 20–29, 30–39, 40–49, 50–59, 60–75, and 75+ years up to their current age period at the time of measurement. The four response options for each age-period were (1) no regular physical activity, (2) regular independent leisure-time physical activity (all non-organized occupational or leisure-time physical activity, i.e., commuting, school/work activities), (3) regular goal-oriented competitive sport and training related to that sport, and (4) other regular supervised physical activity (physical activity that was organized in a sport club, etc., but was not related to competitive sports participation). These four categories were re-categorized into low activity (options 1 and 2) and high activity (options 3 and 4). The same re-categorization was applied to each age range.

4.6. Statistical Analysis

Means and standard deviations were used as descriptive statistics for continuous measurements, frequencies, and percentages of categorical data. The proportional hazards model, extended for time-dependent covariates, was used in modeling the association of weight and physical activity on cancer incidence in (1) longitudinal and (2) near-term settings. We used age as the time variable and determined cancer status at the end of follow-up. Thus, follow-up time extended from study entry at the age of 20 years to exit-age due to cancer diagnosis (event) or remaining free of cancer (censored). As we used time-dependent measurements of weight and physical activity, we utilized the counting process approach [52] for the analysis of the relative risk of cancer related to these exposures. Data were divided into 10-year intervals and each interval was represented by the weight and physical activity measurements of that interval. Due to sex-related differences, we reported models separately for men and women and we also constructed separate models for cancers of any type and CRC; we also used a joint weight/physical activity model to examine the possible interaction between these two exposures. Further details of the models, data management, and model diagnostics can be found in the Supplementary Materials. We reported hazard ratios from the crude unadjusted model, as well as a model adjusted for the affected MMR-gene variant, height, education, smoking, alcohol use, and the use of anti-inflammatory medication. Nested random effects were used to adjust for individuals within the family structure.

5. Conclusions

To conclude, our results suggest that men with *path_MMR* were particularly susceptible to lifestyle exposures that may be either protective or hazard increasing. We found that weight gain in adulthood increased the risk of cancer in men, whereas participating in more intense physical activity across the lifespan may have a cancer-preventive effect. According to our results, women were not as prone to lifestyle-related risk factors. The sex-based difference in the associations could be explained by differences in weight gain, which was smaller in women, and by sex-related factors modifying body composition over time. Taken together, our results emphasize the importance of weight maintenance and high-intensity physical activity throughout the lifespan in cancer prevention in men with *path_MMR*. The results of our study could be used in developing a cancer risk quantification methodology based on the consideration of various risk factors that are modifiable by the individuals themselves.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/cancers13081849/s1, Text S1: Statistical modeling of cancer incidence, Figure S1: Modeling time-dependent predictors and effects, Figure S2: Treatment and observation of missing data patterns, Figure S3: Individual and average trajectories of recalled participation in organized physical activity based on the original four response categories. We used jittering to visually highlight variation activity patterning of the activity measures (i.e., a random uniform-distributed value in the range [−0.4,0.4] was added to each activity value), Figure S4: Individual and average recalled weight trajectories for men and women, Table S1: Tests of the proportionality of hazards based on scaled Schoenfeld residuals for models utilizing longitudinal trajectories of weight and physical activity as time-dependent predictors, Table S2: Tests of the proportionality of hazards based on scaled Schoenfeld residuals for models utilizing last measurement interval weight and physical activity as predictors, Table S3: Summary of hazard ratios (HR) and 95% confidence intervals from various models, Table S4: Summary of the participants' physical activity categories throughout the retrospective follow-up.

Author Contributions: Conceptualization, E.K.L., E.S., T.T. and S.S.; methodology, E.K.L., E.S., S.S. and T.T.; software, T.T.; validation, E.K.L., E.S., S.S., T.S. and T.T.; formal analysis, E.S., T.S. and T.T.; investigation, E.K.L., E.S., S.S., T.S. and T.T.; resources, E.K.L., E.S., J.-P.M., K.P., P.P., T.S., T.T.S., T.T. and S.S.; data curation, E.K.L., E.S., K.P., T.S. and S.S.; writing—original draft preparation, E.S., E.K.L.,T.S. and T.T.; writing—review and editing, E.K.L., E.S., J.-P.M., K.P., P.P., T.S., T.T.S., T.T. and S.S.; visualization, T.S. and T.T.; supervision, E.K.L., E.S. and T.T.S.; project administration, E.K.L.,

J.-P.M., T.T.S. and S.S.; funding acquisition, J.-P.M., P.P. and T.T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Cancer Society Finland, iCAN Flagship of the Finnish Academy, Emil Aaltonen Foundation, Finnish Medical Foundation, Sigrid Juselius Foundation, Instrumentarium Science Foundation, and Jane and Aatos Erkko Foundation. There are no unique identifiers, or grant numbers, associated with the granted funds.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Central Finland Health Care District (KSSHP 3/2016).

Informed Consent Statement: Informed consent was obtained from all participants, and the Central Finland Health Care District Ethics Committee (KSSHP 3/2016) approved the study protocol.

Data Availability Statement: The data are not publicly available due to privacy or ethical restrictions and EU legislation. However, application for the clinical datasets can be made via The Finnish Lynch Syndrome Research Registry and for the datasets obtained through questionnaire survey via contacting the corresponding author.

Acknowledgments: We thank Susmita Chakraborty for her help in data management.

Conflicts of Interest: The authors declare that they have no conflict of interest.

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **2018**, *68*, 394–424. [CrossRef]
2. Haraldsdottir, S.; Rafnar, T.; Frankel, W.L.; Einarsdottir, S.; Sigurdsson, A.; Hampel, H.; Snaebjornsson, P.; Masson, G.; Weng, D.; Arngrimsson, R.; et al. Comprehensive population-wide analysis of Lynch syndrome in Iceland reveals founder mutations in MSH6 and PMS2. *Nat. Commun.* **2017**, *8*, 1–11. [CrossRef]
3. Hampel, H.; Frankel, W.L.; Martin, E.; Arnold, M.; Khanduja, K.; Kuebler, P.; Nakagawa, H.; Sotamaa, K.; Prior, T.W.; Westman, J.; et al. Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). *N. Engl. J. Med.* **2005**, *352*, 1851–1860. [CrossRef]
4. Dominguez-Valentin, M.; Sampson, J.R.; Seppälä, T.T.; ten Broeke, S.W.; Plazzer, J.P.; Nakken, S.; Engel, C.; Aretz, S.; Jenkins, M.A.; Sunde, L.; et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: Findings from the Prospective Lynch Syndrome Database. *Genet. Med.* **2020**, *22*, 15–25. [CrossRef]
5. Møller, P.; Seppälä, T.T.; Bernstein, I.; Holinski-Feder, E.; Sala, P.; Evans, D.G.; Lindblom, A.; Macrae, F.; Blanco, I.; Sijmons, R.H.; et al. Cancer risk and survival in path-MMR carriers by gene and gender up to 75 years of age: A report from the Prospective Lynch Syndrome Database. *Gut* **2018**, *67*, 1306–1316. [CrossRef]
6. Keum, N.; Greenwood, D.C.; Lee, D.H.; Kim, R.; Aune, D.; Ju, W.; Hu, F.B.; Giovannucci, E.L. Adult weight gain and adiposity-related cancers: A dose-response meta-analysis of prospective observational studies. *J. Natl. Cancer Inst.* **2015**, *107*, 1–14. [CrossRef]
7. Rezende, L.F.M.d.; Sá, T.H.d.; Markozannes, G.; Rey-López, J.P.; Lee, I.M.; Tsilidis, K.K.; Ioannidis, J.P.A.; Eluf-Neto, J. Physical activity and cancer: An umbrella review of the literature including 22 major anatomical sites and 770,000 cancer cases. *Br. J. Sports Med.* **2018**, *52*, 826–833. [CrossRef]
8. Kyrgiou, M.; Kalliala, I.; Markozannes, G.; Gunter, M.J.; Paraskevaidis, E.; Gabra, H.; Martin-Hirsch, P.; Tsilidis, K.K. Adiposity and cancer at major anatomical sites: Umbrella review of the literature. *BMJ* **2017**, *356*, 1–10. [CrossRef]
9. MacMahon, S.; Baigent, C.; Duffy, S.; Rodgers, A.; Tominaga, S.; Chambless, L.; De Backer, G.; De Bacquer, D.; Kornitzer, M.; Whincup, P.; et al. Body-mass index and cause-specific mortality in 900,000 adults: Collaborative analyses of 57 prospective studies. *Lancet* **2009**, *373*, 1083–1096. [CrossRef]
10. John, B.J.; Irukulla, S.; Abulafi, A.M.; Kumar, D.; Mendall, M.A. Systematic review: Adipose tissue, obesity and gastrointestinal diseases. *Aliment. Pharmacol. Ther.* **2006**, *23*, 1511–1523. [CrossRef]
11. Friedenreich, C.M.; Neilson, H.K.; Lynch, B.M. State of the epidemiological evidence on physical activity and cancer prevention. *Eur. J. Cancer* **2010**, *46*, 2593–2604. [CrossRef] [PubMed]
12. Moore, S.C.; Lee, I.M.; Weiderpass, E.; Campbell, P.T.; Sampson, J.N.; Kitahara, C.M.; Keadle, S.K.; Arem, H.; De Gonzalez, A.B.; Hartge, P.; et al. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. *JAMA Intern. Med.* **2016**, *176*, 816–825. [CrossRef]
13. Kamiza, A.B.; Hsieh, L.L.; Tang, R.; Chien, H.T.; Lai, C.H.; Chiu, L.L.; Lo, T.P.; Hung, K.Y.; Wang, C.Y.; You, J.F.; et al. Risk factors associated with colorectal cancer in a subset of patients with mutations in MLH1 and MSH2 in Taiwan fulfilling the Amsterdam II criteria for Lynch syndrome. *PLoS ONE* **2015**, *10*, e0130018. [CrossRef] [PubMed]

14. Dashti, S.G.; Win, A.K.; Hardikar, S.S.; Glombicki, S.E.; Mallenahalli, S.; Thirumurthi, S.; Peterson, S.K.; You, Y.N.; Buchanan, D.D.; Figueiredo, J.C.; et al. Physical activity and the risk of colorectal cancer in Lynch syndrome. *Int. J. Cancer* **2018**, *143*, 2250–2260. [CrossRef] [PubMed]
15. Coletta, A.M.; Peterson, S.K.; Gatus, L.A.; Krause, K.J.; Schembre, S.M.; Gilchrist, S.C.; Pande, M.; Vilar, E.; You, Y.N.; Rodriguez-Bigas, M.A.; et al. Energy balance related lifestyle factors and risk of endometrial and colorectal cancer among individuals with lynch syndrome: A systematic review. *Fam. Cancer* **2019**, *18*, 399–420. [CrossRef]
16. Win, A.K.; Dowty, J.G.; English, D.R.; Campbell, P.T.; Young, J.P.; Winship, I.; MacRae, F.A.; Lipton, L.; Parry, S.; Young, G.P.; et al. Body mass index in early adulthood and colorectal cancer risk for carriers and non-carriers of germline mutations in DNA mismatch repair genes. *Br. J. Cancer* **2011**, *105*, 162–169. [CrossRef]
17. Campbell, P.T.; Cotterchio, M.; Dicks, E.; Parfrey, P.; Gallinger, S.; McLaughlin, J.R. Excess body weight and colorectal cancer risk in Canada: Associations in subgroups of clinically defined familial risk of cancer. *Cancer Epidemiol. Biomark. Prev.* **2007**, *16*, 1735–1744. [CrossRef]
18. Nyström-Lahti, M.; Kristo, P.; Nicolaides, N.C.; Chang, S.Y.; Aaltonen, L.A.; Moisio, A.L.; Järvinen, H.J.; Mecklin, J.P.; Kinzler, K.W.; Vogelstein, B.; et al. Founding mutations and Alu-mediated recombination in hereditary colon cancer. *Nat. Med.* **1995**, *1*, 1203–1206. [CrossRef]
19. Moisio, A.L.; Sistonen, P.; Weissenbach, J.; De La Chapelle, A.; Peltomäki, P. Age and origin of two common MLH1 mutations predisposing to hereditary colon cancer. *Am. J. Hum. Genet.* **1996**, *59*, 1243–1251.
20. Pylvänäinen, K.; Lehtinen, T.; Kellokumpu, I.; Järvinen, H.; Mecklin, J.P. Causes of death of mutation carriers in Finnish Lynch syndrome families. *Fam. Cancer* **2012**, *11*, 467–471. [CrossRef]
21. Stevens, G.A.; Singh, G.M.; Lu, Y.; Danaei, G.; Lin, J.K.; Finucane, M.M.; Bahalim, A.N.; McIntire, R.K.; Gutierrez, H.R.; Cowan, M.; et al. National, regional, and global trends in adult overweight and obesity prevalences. *Popul. Health Metr.* **2012**, *10*, 1. [CrossRef]
22. Bentham, J.; Di Cesare, M.; Bilano, V.; Bixby, H.; Zhou, B.; Stevens, G.A.; Riley, L.M.; Taddei, C.; Hajifathalian, K.; Lu, Y.; et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: A pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet* **2017**, *390*, 2627–2642. [CrossRef]
23. Lauby-Secretan, B.; Scoccianti, C.; Loomis, D. Body Fatness and Cancer—Viewpoint of the IARC Working Group. *N. Engl. J. Med.* **2016**, *375*, 794–798. [CrossRef]
24. Botma, A.; Nagengast, F.M.; Braem, M.G.M.; Hendriks, J.C.M.; Kleibeuker, J.H.; Vasen, H.F.A.; Kampman, E. Body mass index increases risk of colorectal adenomas in men with lynch syndrome: The GEOLynch cohort study. *J. Clin. Oncol.* **2010**, *28*, 4346–4353. [CrossRef] [PubMed]
25. Sillanpää, E.; Cheng, S.; Häkkinen, K.; Finni, T.; Walker, S.; Pesola, A.; Ahtiainen, J.; Stenroth, L.; Selänne, H.; Sipilä, S. Body composition in 18- to 88-year-old adults—Comparison of multifrequency bioimpedance and dual-energy X-ray absorptiometry. *Obesity* **2014**, *22*, 101–109. [CrossRef] [PubMed]
26. Juppi, H.-K.; Sipilä, S.; Cronin, N.J.; Karvinen, S.; Karppinen, J.E.; Tammelin, T.H.; Aukee, P.; Kovanen, V.; Kujala, U.M.; Laakkonen, E.K. Role of Menopausal Transition and Physical Activity in Loss of Lean and Muscle Mass: A Follow-Up Study in Middle-Aged Finnish Women. *J. Clin. Med.* **2020**, *9*, 1588. [CrossRef] [PubMed]
27. Gunter, M.J.; Leitzmann, M.F. Obesity and colorectal cancer: Epidemiology, mechanisms and candidate genes. *J. Nutr. Biochem.* **2006**, *17*, 145–156. [CrossRef]
28. Szulc, P.; Duboeuf, F.; Chapurlat, R. Age-Related Changes in Fat Mass and Distribution in Men-the Cross-Sectional STRAMBO Study. *J. Clin. Densitom.* **2017**, *20*, 472–479. [CrossRef]
29. Kirchengast, S.; Gruber, D.; Sator, M.; Hartmann, B.; Knogler, W.; Huber, J. Menopause-associated differences in female fat patterning estimated by dual-energy X-ray absorptiometry. *Ann. Hum. Biol.* **1997**, *24*, 45–54. [CrossRef]
30. Dratva, J.; Gómez Real, F.; Schindler, C.; Ackermann-Liebrich, U.; Gerbase, M.W.; Probst-Hensch, N.M.; Svanes, C.; Omenaas, E.R.; Neukirch, F.; Wjst, M.; et al. Is age at menopause increasing across Europe? Results on age at menopause and determinants from two population-based studies. *Menopause* **2009**, *16*, 385–394. [CrossRef]
31. Silveira, E.A.; Kliemann, N.; Noll, M.; Sarrafzadegan, N.; de Oliveira, C. Visceral obesity and incident cancer and cardiovascular disease: An integrative review of the epidemiological evidence. *Obes. Rev.* **2021**, *22*, e13088. [CrossRef] [PubMed]
32. Simpson, E.R.; Bulun, S.E.; Nichols, J.E.; Zhao, Y. Estrogen biosynthesis in adipose tissue: Regulation by paracrine and autocrine mechanisms. *J. Endocrinol.* **1996**, *150*, S51–S57.
33. Nelson, L.R.; Bulun, S.E. Estrogen production and action. *J. Am. Acad. Dermatol.* **2001**, *45*, S116–S124. [CrossRef] [PubMed]
34. Straub, R.H. The complex role of estrogens in inflammation. *Endocr. Rev.* **2007**, *28*, 521–574. [CrossRef]
35. Campbell, P.T.; Newcomb, P.; Gallinger, S.; Cotterchio, M.; McLaughlin, J.R. Exogenous hormones and colorectal cancer risk in Canada: Associations stratified by clinically defined familial risk of cancer. *Cancer Causes Control* **2007**, *18*, 723–733. [CrossRef]
36. La Vecchia, C.; Gallus, S.; Fernandez, E. Hormone replacement therapy and colorectal cancer: An update. *J. Br. Menopause Soc.* **2005**, *11*, 166–172. [CrossRef]
37. Liang, J.; Shang, Y. Estrogen and cancer. *Annu. Rev. Physiol.* **2013**, *75*, 225–240. [CrossRef]
38. Seppälä, T.T.; Latchford, A.; Negoi, I.; Sampaio Soares, A.; Jimenez-Rodriguez, R.; Sánchez-Guillén, L.; Evans, D.G.; Ryan, N.; Crosbie, E.J.; Dominguez-Valentin, M.; et al. European guidelines from the EHTG and ESCP for Lynch syndrome: An updated third edition of the Mallorca guidelines based on gene and gender. *Br. J. Surg.* **2020**. [CrossRef]

39. Lavie, C.J.; Ozemek, C.; Carbone, S.; Katzmarzyk, P.T.; Blair, S.N. Sedentary Behavior, Exercise, and Cardiovascular Health. *Circ. Res.* **2019**, *124*, 799–815. [CrossRef] [PubMed]
40. Friedenreich, C.M.; Shaw, E.; Neilson, H.K.; Brenner, D.R. Epidemiology and biology of physical activity and cancer recurrence. *J. Mol. Med.* **2017**, *95*, 1029–1041. [CrossRef]
41. Avgerinos, K.I.; Spyrou, N.; Mantzoros, C.S.; Dalamaga, M. Obesity and cancer risk: Emerging biological mechanisms and perspectives. *Metabolism* **2019**, *92*, 121–135. [CrossRef] [PubMed]
42. Davies, A.; Wellard-Cole, L.; Rangan, A.; Allman-Farinelli, M. Validity of self-reported weight and height for BMI classification: A cross-sectional study among young adults. *Nutrition* **2020**, *71*, 110622. [CrossRef]
43. Dahl, A.K.; Reynolds, C.A. Accuracy of recalled body weight—A study with 20-years of follow-up. *Obesity* **2013**, *21*, 1293–1298. [CrossRef] [PubMed]
44. Tuomela, J.; Kaprio, J.; Sipilä, P.N.; Silventoinen, K.; Wang, X.; Ollikainen, M.; Piirtola, M. Accuracy of self-reported anthropometric measures—Findings from the Finnish Twin Study. *Obes. Res. Clin. Pract.* **2019**, *13*, 522–528. [CrossRef]
45. Smith, A.W.; Cronin, K.A.; Bowles, H.; Willis, G.; Jacobs, D.R.J.; Ballard-Barbash, R.; Troiano, R.P. Reproducibility of physical activity recall over fifteen years: Longitudinal evidence from the CARDIA study. *BMC Public Health* **2013**, *13*, 180. [CrossRef] [PubMed]
46. Vasen, H.F.; Watson, P.; Mecklin, J.P.; Lynch, H.T. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. *Gastroenterology* **1999**, *116*, 1453–1456. [CrossRef]
47. Umar, A.; Boland, C.R.; Terdiman, J.P.; Syngal, S.; de la Chapelle, A.; Rüschoff, J.; Fishel, R.; Lindor, N.M.; Burgart, L.J.; 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. [CrossRef]
48. Thompson, B.A.; Spurdle, A.B.; Plazzer, J.-P.; Greenblatt, M.S.; Akagi, K.; Al-Mulla, F.; Bapat, B.; Bernstein, I.; Capellá, G.; den Dunnen, J.T.; et al. Application of a 5-tiered scheme for standardized classification of 2360 unique mismatch repair gene variants in the InSiGHT locus-specific database. *Nat. Genet.* **2014**, *46*, 107–115. [CrossRef] [PubMed]
49. Hirvensalo, M.; Lampinen, P.; Rantanen, T. Physical Exercise in Old Age: An Eight-Year Follow-Up Study on Involvement, Motives, and Obstacles Among Persons Age 65–84. *J. Aging Phys. Act.* **1998**, *6*, 157–168. [CrossRef]
50. Hyvärinen, M.; Sipilä, S.; Kulmala, J.; Hakonen, H.; Tammelin, T.H.; Kujala, U.M.; Kovanen, V.; Laakkonen, E.K. Validity and reliability of a single question for leisure-time physical activity assessment in middle-aged women. *J. Aging Phys. Act.* **2020**, *28*, 231–241. [CrossRef]
51. Hirvensalo, M.; Lintunen, T.; Rantanen, T. The continuity of physical activity—A retrospective and prospective study among older people. *Scand. J. Med. Sci. Sports* **2000**, *10*, 37–41. [CrossRef] [PubMed]
52. Therneau, T.M.; Grambsch, P.M. *Modeling Survival Data: Extending the Cox Model*; Springer: New York, NY, USA, 2000; ISBN 9781441931610.