



Faculty of Health Sciences
UNIVERSITY OF CAPE TOWN
DIVISION OF HUMAN GENETICS
MSc. Med (by dissertation) in Human Genetics

Elucidation of the Pathways Underlying the Spectrum of Cancers Affecting Different Tissues in Lynch Syndrome

May Krause

Student number: KRSMAY002

PS Number: 1684754

KRSMAY002@myuct.ac.za

SUBMITTED TO THE UNIVERSITY OF CAPE TOWN
In fulfilment of the requirements for this degree

Supervisor: Prof. Raj Ramesar – raj.ramersar@uct.ac.za, Department of Pathology,
University of Cape Town

Co-supervisor: Dr. Sinkala Musalula – musalula.sinkala@uct.ac.za, Department of
Integrative Biomedical Sciences, University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Declaration

I, *May Krause*, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature:

Signed by candidate

Date: 13 November 2024

Table of Contents:

Acknowledgements	i
Abbreviations	ii
List of Figures.....	iv
List of Tables.....	v
Thesis Organisation.....	vi
Abstract.....	vii
Chapter 1: Introduction.....	1
1.1 Introduction to Lynch Syndrome	1
1.1.1 Historical Background	2
1.2 Genetic Basis of LS	3
1.2.1 DNA MMR Pathway	4
1.2.2 Two-Hit Paradigm and Haploinsufficiency	6
1.2.3 Epigenetic Silencing of MMR Genes	6
1.2.4 MSI and Its Role in Tumorigenesis within LS	7
1.2.5 <i>BRAF</i> and <i>KRAS</i> Mutations in LS.....	9
1.2.6 Key Molecular Pathways in Cancer.....	11
1.3 Clinical Features and Diagnostic Criteria	12
1.3.1 Amsterdam Criteria and Revised Bethesda Guidelines	12
1.3.2 Predictive Models for LS Identification.....	14
1.3.3 Clinical Manifestations and Spectrum of Tumours.....	16
1.3.4 LS-Associated Cancers	17
1.4 Conclusion	23
1.5 Aims and Objectives	23
Chapter 2: Elucidation of Dysregulated Pathways in Sporadic and Lynch Syndrome-associated Colorectal Cancer – Implications for Treatment	25
2.1 Introduction.....	25
2.2 Materials and methods.....	26
2.2.1 Data Acquisition.....	26

2.2.2	Pathogenic Variant Identification	26
2.2.3	Classification of LLS- and LSp-CRC	26
2.2.4	Mutational Signature Analysis	28
2.2.5	DGE and Pathway Enrichment Analysis	28
2.2.6	Prediction of Regulator Kinases and TFs.....	29
2.2.7	Mutation Analyses	29
2.2.8	Drug Sensitivity and Dependency Analysis	29
2.2.9	Statistical Analyses	31
2.2.10	Data, Code and Supplementary Files Availability	31
2.3	Results	31
2.3.1	Dataset Assembly and Initial Classification.....	31
2.3.2	Identifying LLS- and LSp-CRC Samples	32
2.3.3	Clinical and Molecular Characteristics of CRC Subtypes.....	34
2.3.4	The Pathogenic Mutational Landscape of CRC Subtypes	38
2.3.5	Top Mutated Genes	41
2.3.6	Differential Gene Expression Analysis	43
2.3.7	Pathway Enrichment Analysis.....	44
2.3.8	Kinase and Transcription Factor Analysis	48
2.3.9	Validation using Known LS Samples.....	49
2.3.10	Drug Sensitivity and Dependency Analysis	51
2.4	Discussion	57
2.5	Conclusion	62

Chapter 3: Multiomic Analysis of Endometrial Lynch Syndrome-Associated Cancer and Investigating the Inclusion of Breast Cancer in the Lynch Syndrome Spectrum..... 63

3.1	Introduction	63
3.2	Materials and methods.....	64
3.2.1	Data Acquisition.....	64

3.2.2	Classification of LLS and LSp Samples	64
3.2.3	DGE and Pathway Enrichment Analysis	64
3.2.4	Predication of Regulatory Kinases and TFs	65
3.2.5	Mutation Analyses	65
3.2.6	Statistical Analyses	65
3.2.7	Statistical Analyses	66
3.2.8	Data, Code and Supplementary Files Availability	67
3.3	Results	67
3.3.1	Dataset Assembly and Initial Classification.....	67
3.3.2	Clinical and Molecular Characteristics of Cancer Subtypes	68
3.3.3	The Mutational Landscape of EC and BC Subtypes.....	75
3.3.4	Top Mutated Genes.....	80
3.3.5	Differential Gene Expression Analysis	81
3.3.6	Pathway Enrichment Analysis.....	82
3.3.7	Kinase and TF Analysis	84
3.3.8	Drug Sensitivity and Dependency Analysis	87
3.4	Discussion	92
3.5	Conclusion	96
Chapter 4: Concluding Remarks		97
4.1	General Discussion.....	97
4.2	Study Limitations and Future Directions	99
References.....		102
Appendices		115
Appendix A: Supplementary Figures		115
Appendix B: Supplementary File Description		120
Appendix C.....		122

Acknowledgements

I am profoundly thankful to parents for all the opportunities they have provided for me, for all their support and encouragement along the way. Thank you for helping me chase my dreams.

My friends, family and partner Dillon whose unwavering encouragement, patience, and belief in my abilities have been my pillars of strength. Your continuous support has fuelled my determination to reach this milestone.

I would like to express my gratitude to Prof Ramesar for providing me with the invaluable opportunity to pursue my research. I am deeply appreciative of the University of Cape Town and its generous donors for the partial scholarship for my first year of Masters. Additionally, I am extremely grateful for The Oppenheimer Memorial Trust for providing me with my scholarship for my final year that eased the financial burden of my education. Their commitment to advancing education has made it possible for me to pursue my academic aspirations.

A special word of appreciation goes to Dr Musalula Sinkala, whose unwavering support provided me with the motivation to overcome challenges and stay focused on my goals. Your belief in my potential has been a constant source of inspiration.

In addition, I extend my gratitude to all the individuals, colleagues, and mentors who have contributed to my academic and personal growth along this journey. Your insights and guidance have been invaluable.

Lastly, I would like to acknowledge the countless individuals who have played a role, however small, in my academic and personal development. Your contributions have not gone unnoticed and are deeply appreciated. This thesis is a culmination of the collective support and encouragement I have received, and I am truly grateful for each and every person who has been a part of this remarkable journey.

Abbreviations

AI	Artificial intelligence
BC	Breast cancer
BTC	Biliary tract cancer
cAMP	Cyclic adenosine monophosphate
CCLE	Cancer Cell Line Encyclopaedia
ChEA	ChIP-seq experiments
CIMP	CpG island methylator phenotype
CMMR-D	Constitutional MMR deficiency
COSMIC	Catalogue of Somatic Mutations in Cancer
CRC	Colorectal cancer
DEG	Differentially expressed gene
DGE	Differential gene expression
dMMR	Deficient MMR
EC	Endometrial cancer
ECM	Extracellular matrix
<i>EPCAM</i>	Epithelial cellular adhesion molecule
EXO1	Exonuclease 1
FAP	Familial adenomatous polyposis
FDR	False discovery rate
GC	Gastric cancer
GDSC	Genomics of Drug Sensitivity in Cancer
GEO	Gene Expression Omnibus
GSEA	Gene set enrichment analysis
hg38	Human genome reference assembly 38
HGG	High-grade glioma
HNPCC	Hereditary nonpolyposis colorectal cancer
ICG-HNPCC	International Collaborative Group on HNPCC
IHC	Immunohistochemistry
IL-17	Interleukin-17
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LLS	Likely lynch syndrome
LS	Lynch syndrome
LSp	Likely sporadic
MAF	Mutation Annotation Format
MAP	Mitogen-activated protein
<i>MLH1</i>	MutL homolog 1
MMR	Mismatch repair
mRMR	Maximum relevance-minimum redundancy
mRNA	messenger RNA
<i>MSH2</i>	MutS homolog 2
<i>MSH6</i>	MutS homolog 6
MSI	Microsatellite instability
MSI-H	Microsatellite instability high
MSI-L	Low-degree MSI
MSS	Microsatellite stable

MutL α	MLH1-PMS2
MutS α	MSH2-MSH6
MutS β	MSH2-MSH3
NCA	Neighbourhood Component Analysis
NGS	Next-generation sequencing
OC	Ovarian cancer
OG	Oncogene
PC	Pancreatic cancer
PCA	Principal component analysis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PI3K-Akt	Phosphatidylinositol 3'-kinase-protein kinase B
PLSD	Prospective Lynch Syndrome Database
<i>PMS2</i>	PMS1 homolog 2
Pol δ	Polymerase δ
PPC	Primary peritoneal cancer
RFC	Replication factor C
RNA-seq	RNA sequencing
RTK	Receptor tyrosine kinase
SMOTE	Synthetic Minority Over-sampling Technique
TCGA	The Cancer Genome Atlas
TGF-beta	Transforming Growth Factor-beta
TMB	Tumour mutation burden
TS	Turcot syndrome
TSG	Tumour suppressor gene
UTUC	Upper tract urothelial carcinoma
WES	Whole exome sequencing
WT	Wild-type
X2K	<i>Expression2Kinases</i>

List of Figures

Figure 1. The Eukaryotic Mismatch Repair (MMR) System.	5
Figure 2. Visualisation of Subtypes in CRC Samples.	33
Figure 3. Mechanism of <i>MLH1</i> Hypermethylation or Mutation Leading to Deficient Mismatch Repair (dMMR) and Genomic Instability.	33
Figure 4. Comparison of Microsatellite Instability (MSI) and Tumour Mutation Burden (TMB) Between Likely Lynch Syndrome (LLS) and Likely Sporadic (LSp) Colorectal Cancer (CRC) Samples.	37
Figure 5. An oncoplot of the top pathogenic mutated tumour suppressor and oncogenes in the likely Lynch Syndrome colorectal cancer samples (n=12).	40
Figure 6. An oncoplot of the top pathogenic mutated tumour suppressor and oncogenes in the Likely Sporadic colorectal cancer samples (n=313).	41
Figure 7. Biological processes and pathways in the Likely Lynch Syndrome Colorectal Cancer subtype (n=12).	42
Figure 8. Top 50 mutant genes in colorectal cancer (CRC) subtypes.	43
Figure 9. Pathway enrichment analyses: KEGG pathway enrichment results highlight some significant pathways that are differentially altered between the colorectal cancer subtypes.	44
Figure 10. A comparison of PI3K-Akt signalling pathway activity between Likely Lynch Syndrome (LLS) and Likely Sporadic (LSp) Colorectal Cancer (CRC) subtypes based on mRNA expression data.	47
Figure 11. Predicted Regulatory Network and Key Protein Kinases in Likely Lynch Syndrome Colorectal Cancer (LLS-CRC).	48
Figure 12. Predicted Transcriptional Regulatory Network in Likely Lynch Syndrome Colorectal Cancer (LLS-CRC).	49
Figure 13. Relationship between Achilles gene dependence scores or mutations of PI3K-Akt pathway genes and the Responses of the cell line to PI3K-Akt pathway inhibitors.	54
Figure 14. L100 Responses of the Four Cell Lines to Inhibitors Alpelisib and OSI-027.	56
Figure 15. Comparison of Microsatellite Instability (MSI) and Tumour Mutation Burden (TMB) Between likely Lynch Syndrome (LLS) and likely Sporadic (LSp) Endometrial Cancer (EC) and Breast Cancer (BC) Samples.	71
Figure 16. An oncoplot of the top mutated tumour suppressor and oncogenes in the (A) likely Lynch Syndrome (LLS) endometrial cancer and (B) LLS breast cancer samples.	77
Figure 17. An oncoplot of the top mutated tumour suppressor and oncogenes in (A) likely sporadic (LSp) endometrial cancer and (B) LSp breast cancer samples.	79
Figure 18. Signalling Pathways in Cancer Subtypes.	81
Figure 19. Pathway enrichment analyses: KEGG pathway enrichment results highlight some significant pathways that are differentially altered between the (A) endometrial and (B) breast cancer subtypes.	82
Figure 20. Predicted Regulatory Network and Key Protein Kinases in Likely Lynch Syndrome (LLS) Endometrial (EC) and Breast Cancer (BC).	85
Figure 21. Predicted Transcriptional Regulatory Network in Likely Lynch (LLS) Endometrial (EC) and Breast Cancer (BC).	87
Figure 22. Relationship between Achilles gene dependence scores or mutations of TGF-Beta-related pathway genes and the Responses of the cell line to TGF-Beta-related pathway inhibitors.	90
Figure S1. Determination of Optimal Clustering and Visualisation of Subtypes in Colorectal Cancer (CRC) Samples.	115
Figure S2. An oncoplot of the top pathogenic mutated tumour suppressor and oncogenes across all colorectal cancer samples.	115
Figure S3. Differential Gene Expression Analysis of Likely Lynch Syndrome (LLS) and Sporadic (LSp) Colorectal Cancer (CRC) Subtypes.	116
Figure S4. Relationship between Achilles gene dependence scores or mutations of Wnt pathway genes and the Responses of the cell line to Wnt pathway inhibitors.	117
Figure S5. Differential Gene Expression Analysis of Likely Lynch Syndrome (LLS) and Sporadic (LSp) Endometrial (EC) and Breast Cancer (BC) Subtypes.	118
Figure S6. Relationship between Achilles gene dependence scores or mutations of cAMP-related pathway genes and the Responses of the cell line to cAMP-related pathway inhibitors.	119

List of Tables

Table 1. Amsterdam Criteria for Identifying Individuals with LS	13
Table 2. Bethesda Guidelines for Identifying Individuals with LS	14
Table 3. Detailed Distribution of Mismatch Repair Gene Variants in the 12 Likely Lynch Syndrome Samples	35
Table 4. Clinical and Molecular Features of the Two Colorectal Cancer Groups	38
Table 5. Detailed Distribution of Mismatch Repair Gene Variants in the 13 and 14 Likely Lynch Syndrome Endometrial and Breast Cancer Samples, Respectively	69
Table 6. Clinical and Molecular Features of the Two Endometrial Cancer Groups.....	73
Table 7. Clinical and Molecular Features of the Two Breast Cancer Groups.....	74
Table C1. Key Resources used in this Thesis	122

Thesis Organisation

This thesis is organised into four chapters:

- **Chapter One** provides a general background on Lynch Syndrome, covering its molecular pathology and extracolonic manifestations. It also introduces the aims and objectives of this study.
- **Chapters Two and Three** are dedicated to experimental analyses:
 - **Chapter Two** focuses on analysing publicly available genomic and transcriptomic data from colorectal cancer (CRC) patients. The chapter identifies distinct signalling and metabolic pathways associated with Lynch Syndrome (LS)-associated CRC and sporadic CRC and potential therapeutic interventions.
 - **Chapter Three** presents a comprehensive multiomic analysis of publicly available somatic data from endometrial cancer (EC) and breast cancer (BC) cases. The chapter aims to highlight signalling pathways, and alterations unique to LS-associated EC, potentially supporting the inclusion of BC within the LS cancer spectrum.
- **Chapter Four** provides an overview of the project and includes potential research prospects

Abstract

Lynch Syndrome (LS) is a hereditary disorder that predisposes individuals to an increased risk of several cancers, particularly colorectal cancer (CRC) and endometrial cancer (EC). LS arises from constitutional (“germline”) pathogenic variants affecting one of four DNA mismatch repair (MMR) genes. When coupled with somatic mutations (“second hits”) in tumours affecting the same gene, this results in deficient MMR. However, the mechanisms by which MMR deficiency contribute to cancer development remain unclear.

This thesis aims to improve LS ascertainment by investigating driver genes and dysregulated pathways in CRC, EC and breast cancer (BC) and to address whether BC may be part of the LS cancer spectrum. Furthermore, it aims to identify subtype-specific drugs targeting these pathways. Using somatic data from cBioPortal (<https://www.cbioportal.org>), samples were triaged as likely LS (LLS) or likely sporadic (LSp), with computational analyses providing molecular insights into the subtypes of LS-associated cancers. Key analyses included differential expression, pathway enrichment, kinase activity, mutational signatures, and drug sensitivity assessments to characterise cancer subtypes.

The study identified PI3-Akt and Wnt signalling as dysregulated in LLS-CRC and LSp-EC, respectively. Similar analyses in EC and BC revealed that the TGF-beta and cAMP signalling pathways were notably dysregulated in LLS and LSp subtypes, respectively, for both EC and BC, supporting the inclusion of BC within the LS spectrum. Drug sensitivity analysis showed that cell lines with higher dependency on key pathways exhibited increased sensitivity to corresponding inhibitors, suggesting potential therapeutic targets in LS-associated cancers. This study elucidates key driver genes, dysregulated pathways and drug sensitivities in LS-associated cancers, offering insights into the molecular mechanisms driving these cancers. Variations in pathway enrichment and drug responses between LLS and LSp cases across cancer types suggest new therapeutic avenues and underscore the importance of personalised treatment strategies based on cancer subtype, genetic profile, and pathway dependencies.

Keywords: Lynch Syndrome; Cancer pathways, Bioinformatics

Chapter 1: Introduction

1.1 Introduction to Lynch Syndrome

Lynch syndrome (LS), previously known as hereditary nonpolyposis colorectal cancer (HNPCC), is a hereditary disorder that significantly increases the risk of developing multiple types of cancer, including colonic, rectal, endometrial, small intestinal, ovarian, hepatobiliary and urothelial cancers¹. LS is the most common hereditary colorectal cancer (CRC) syndrome, with LS accounting for approximately 3–5% of all CRC cases and up to 3% of endometrial cancer (EC) cases^{2,3}. This autosomal dominant condition is characterised by early cancer onset and arises from constitutional pathogenic variants affecting one of four DNA mismatch repair (MMR) genes: MutL homolog 1 (*MLH1*), MutS homolog 2 (*MSH2*), MutS homolog 6 (*MSH6*), and PMS1 homolog 2 (*PMS2*). Additionally, included are deletions in the 3' region of epithelial cellular adhesion molecule (*EPCAM*), leading to a loss of *MSH2* expression^{4,5}.

The MMR system plays a crucial role in maintaining genetic stability by identifying and repairing errors such as mismatches and small insertions/deletions during DNA replication⁶. In individuals with LS, defects in the MMR system lead to the accumulation of mutations, thereby increasing the risk of cancer development⁵. These defects arise after a second hit inactivates the remaining wild-type (WT) allele of the MMR gene, resulting in a complete loss of MMR function. Microsatellite instability (MSI) is a hallmark of LS-associated cancers and is characterised by variations in the length of short tandem nucleotide repeats known as microsatellites⁷. While MSI is predominantly associated with LS in younger patients, it is also observed in sporadic cancers, particularly in older individuals, where it arises through mechanisms such as *MLH1* promoter hypermethylation. In LS, MSI reflects the impact of MMR deficiency (dMMR) and contributes to an elevated risk of CRC and EC, with these cancers typically occurring at younger ages compared to sporadic cases⁷.

With an estimated global incidence of 1 in 100 to 1 in 180 individuals, LS represents a significant public health challenge⁸. LS exhibits a high penetrance, with affected individuals having a 50% chance of passing the variant to their offspring⁹. Understanding the genetic basis of LS and its association with MMR gene variants is crucial for the accurate diagnosis and

effective management of affected individuals. Genetic testing enables the presymptomatic identification of high-risk individuals, allowing for tailored screening protocols and preventive measures. By recognising the hereditary nature of LS and its diverse cancer risks, healthcare providers can implement strategies to mitigate the impact of this condition on individuals and their families. Early detection and intervention can substantially affect the prognosis and quality of life of individuals with LS and can contribute to reducing the overall burden of LS-related cancers in the population.

The purpose of this study was to investigate the molecular underpinning of LS-associated CRC and EC by analysing the genomic and transcriptomic data available for likely LS and sporadic cancer cases. This study aims to identify pathways underlying the LS form of CRC, and potential therapeutic targets that could improve the management and clinical outcomes of LS-associated cancers. Additionally, this study aimed to explore whether breast cancer (BC) shares molecular characteristics with LS CRC and EC patients, potentially supporting its inclusion in the LS cancer spectrum.

1.1.1 Historical Background

The first documentation of what would later be recognised as LS occurred in 1913 by Dr. Aldred Scott Warthin, a pathologist from the University of Michigan, who observed a pattern of endometrial and gastric cancers (GC) in a family known as “Family G”¹⁰. This case laid the groundwork for identifying similar familial cancer characteristics. In 1966, Dr. Henry Lynch, an American physician, detailed the cancer profiles of two large Midwestern families with members affected by colon, gastric, and endometrial cancers⁹. He proposed that these cancers could be attributed to an autosomal dominant “cancer family syndrome”¹¹. In 1984, the term “Lynch syndrome” was coined to differentiate between families primarily affected by CRC (Lynch I) and those with additional non-colonic tumours (Lynch II)¹². The term “Hereditary nonpolyposis colorectal cancer” was introduced to contrast these cases from familial adenomatous polyposis (FAP), a more common autosomal dominant CRC syndrome characterised by numerous colorectal adenomatous polyps⁹.

For many years after Warthin's initial description of Family G, LS was primarily defined based on its clinical features ¹⁰. However, in 1993, genome-wide linkage analysis identified loci on chromosomes 2p and 3p as significant in families with a history of autosomal dominant CRC. The tumours associated with these loci exhibited distinctive histopathological features, most notably MSI, indicating errors in DNA MMR as a novel pathway in CRC development¹⁰. Subsequent research has established a link between MSI in CRC tumours and defects in DNA MMR ¹³. This finding led to the identification of four key MMR genes commonly associated with LS: *MLH1*, *MSH2*, *MSH6*, and *PMS2*¹⁴.

As research expanded, the association between LS and other cancers became apparent, particularly gynaecological cancers such as endometrial and ovarian cancers, alongside CRC. The recognition of these broader cancer risks has shaped the current understanding of LS as a multi-cancer syndrome, influencing clinical guidelines for diagnosis, screening and management¹⁵.

1.2 Genetic Basis of LS

The genetic landscape of LS is complex, encompassing both mutations and/or a spectrum of epigenetic alterations. Central to this condition are pathogenic germline variants in MMR genes, which play a critical role in correcting DNA replication errors and maintaining genomic stability ⁶. The loss of function in any of these genes results in dMMR, leading to the accumulation of mutations and the development of MSI, hallmark features of LS-associated cancers⁷. In addition to variants in core MMR genes, epigenetic changes, such as methylation, can silence these genes, preventing them from functioning properly. This silencing causes dMMR and MSI, which are hallmark features of LS-associated cancers. However, in such cases, the dysfunction arises from epigenetic changes rather than inherited genetic variants, making the tumour appear similar to LS even though it is not caused by a germline variant¹⁶. Additionally, the involvement of other genetic factors, including but not limited to mutations in *BRAF* and *KRAS*, adds complexity to the genetic profile of CRC¹⁷. These mutations can interact with or operate independently of the MMR system, influencing tumour development and progression in distinct ways. Such genetic variability contributes to the heterogeneity observed in CRC, affecting its molecular characteristics, prognosis, and response to

treatment. The interplay of these genetic and epigenetic elements not only elucidates the pathophysiology of LS but also informs the development of targeted screening and management strategies for those affected by this condition.

1.2.1 DNA MMR Pathway

The DNA MMR pathway is a critical mechanism that maintains genomic stability by correcting errors that occur during DNA replication⁶. The MMR system functions through a series of coordinated steps involving recognition, excision and repair of these error. The process begins with recognition of mismatched bases by MutS protein complexes (Figure 1). In eukaryotic cells, these complexes consist of either MutS α (a heterodimer of MSH2 and MSH6) or MutS β (a heterodimer of MSH2 and MSH3). MutS α primarily recognises single-base mismatches and dinucleotide insertion-deletion loops, while MutS β detects larger insertion-deletion loops¹⁸. After the detection of a mismatch, the MutS complex binds to the DNA at the site of the error.

Following mismatch recognition, the MutS complex recruits the MutL complex to the site of the lesion. The most prominent MutL complex in humans is MutL α , which is composed of MLH1 and PMS2. The MutL complex acts as a scaffold, facilitating the interaction of additional proteins that are necessary for excising the erroneous DNA. These proteins include Exonuclease 1 (EXO1), which removes the mismatched segment of the newly synthesised DNA strand, creating a gap. Once the erroneous DNA has been excised, the gap is filled in by the replicative DNA polymerase, Polymerase δ (Pol δ). This enzyme synthesises the correct DNA sequence using the complementary strand as a template. Finally, the newly synthesised DNA strand is ligated by DNA Ligase 1, restoring the integrity of the DNA and completing the repair process¹⁹.

The efficiency of the MMR system is critical for accurate genetic information transfer, enabling a mutation rate of less than 2×10^{-10} base substitutions per cell division²⁰. This low mutation rate is vital for preventing the development of malignancies and underlines the importance of the MMR pathway in cellular integrity and cancer prevention in LS²⁰.

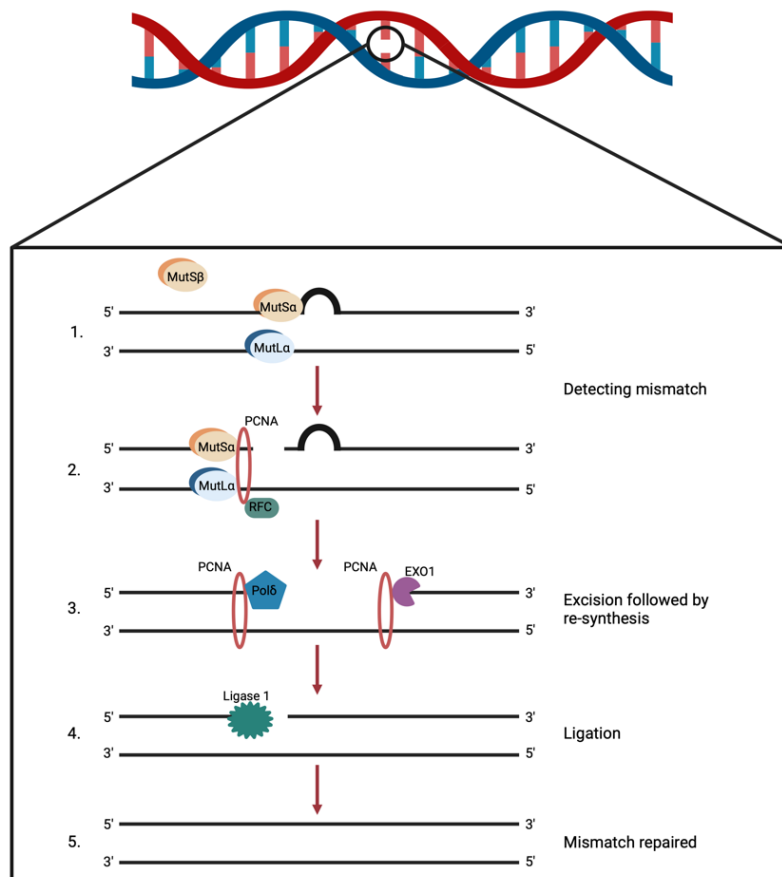


Figure 1. The Eukaryotic Mismatch Repair (MMR) System. (1) Mismatch Detection: The MMR process begins when the MutS heterodimers MSH2-MSH6 (MutSa) or MSH2-MSH3 (MutSβ) detect a DNA mismatch. The complex is bound by the primary MutL heterodimer, MLH1-PMS2 (MutLa), which is pivotal in coordinating the subsequent steps of repair. **(2) Recruitment of Repair Factors:** Proliferating cell nuclear antigen (PCNA) and replication factor C (RFC) are recruited to the site of the mismatch. PCNA functions as a sliding clamp that stabilises DNA polymerase, whereas RFC acts as a clamp loader. **(3) Excision and Re-synthesis:** The endonuclease activity of PMS2 introduces a single-strand break at the mismatch site. Exonuclease 1 (EXO1) is recruited to remove fragments containing the mismatch. DNA polymerase δ (Pol δ) is responsible for synthesising the new, correct DNA fragment to replace the excised section. **(4) DNA Ligation:** DNA ligase 1 performs the ligation step, sealing the nicks and restoring the continuous phosphodiester backbone of the DNA. **(5) Repair Completion:** The DNA mismatch repair is completed and the integrity of the DNA is restored.

In LS, germline pathogenic variants in one or more MMR genes compromise MMR system integrity. This disruption undermines the essential function of the system in correcting errors in DNA replication, leading to an unchecked accumulation of mutations. This process is significantly correlated with the pathogenesis of tumours²¹. Such accumulation of genetic aberrations can precipitate the onset of various cancers characteristic of LS, underscoring the critical role of functional MMR genes in maintaining cellular genomic stability. Additionally, using a mouse model, Yates *et al.* have suggested that LS-associated ECs exhibit unique

metabolic changes, such as decreased mitochondrial function, tied to germline pathogenic variants in *MSH2*²². This highlights the broader systemic impact of MMR deficiencies beyond an increased mutation rate.

1.2.2 Two-Hit Paradigm and Haploinsufficiency

LS demonstrates an autosomal dominant inheritance pattern, with carriers of a pathogenic constitutional variant of one of the four MMR genes and/or *EPCAM* facing a 50% likelihood of passing this variant to their offspring²³. Although carrying such a variant correlates with increased cancer risk, full penetrance is seldom observed²⁴. The complete impairment of MMR functionality necessitates a 'second hit' – a mutation that inactivates the remaining functional allele of the gene, leading to dMMR. This dMMR induces a mutator phenotype characterised by MSI and a globally elevated mutation rate in the cell genome, which significantly increases cancer susceptibility in individuals with LS²⁰. These observations are consistent with Knudson's two-hit hypothesis, which states that two mutations are often required for the manifestation of certain genetic disorders²⁴.

Moreover, the presence of a WT allele may not suffice to maintain normal gene function because of haploinsufficiency, where a single functional copy of a gene does not produce an adequate quantity of its product for normal function²⁴. Evidence from studies involving heterozygous *MLH1* and *MSH2* transgenic mice indicate reduced MMR protein expression and an increase in genomic frameshift deletions within the colon, corroborating the biological impact of haploinsufficiency^{25,26}. Collectively, the two-hit paradigm and haploinsufficiency provide a conceptual framework for elucidating the genetic underpinnings of LS and the aetiology of associated malignancies.

1.2.3 Epigenetic Silencing of MMR Genes

Epigenetic silencing is a phenomenon that plays a pivotal role in the regulation of gene expression, with DNA methylation being the primary epigenetic modification that can lead to silencing of gene activity²⁷. This modification involves the addition of a methyl group to the cytosine base within CpG islands, which are often located near the promoter regions of genes²⁸. In cancer cells, epigenetic changes, such as hypomethylation or hypermethylation,

are common²⁹. Global hypomethylation (decreased methylation) can result in chromosomal instability, whereas hypermethylation (increased methylation), particularly of tumour suppressor genes, specifically *MLH1*, significantly contributes to their silencing and, therefore, to malignancy³⁰.

The occurrence of promoter hypermethylation leading to gene expression silencing has been extensively documented in sporadic CRC, where patients lack a family history of malignancy and do not fulfil the diagnostic criteria for LS ^{30,31}. This epigenetic mechanism plays a significant role in disrupting the functionality of the MMR system, contributing to the accumulation of errors during DNA replication and thus promoting tumour development marked by MSI. A common epigenetic event in sporadic CRC involves inactivation of *MLH1*, a phenomenon that is closely associated with a subset of CRCs exhibiting MSI and characterised by the CpG island methylator phenotype (CIMP) ³². Approximately 15% of sporadic CRC cases with MSI, which also show CIMP, are characterised by biallelic *MLH1* inactivation due to promoter methylation ³². This process of methylation-driven suppression of gene expression effectively results in hallmark LS features, even though the cancer is sporadic in origin.

1.2.4 MSI and Its Role in Tumorigenesis within LS

Microsatellites, short tandem repeat sequences of DNA (<5 base pairs in length), are abundantly dispersed throughout the eukaryotic genome, appearing approximately every 50 kilobases ³³. These sequences are susceptible to slippage by the DNA polymerase during replication, leading to potential insertions or deletions²⁰. Normally, the MMR system efficiently corrects such errors. However, pathogenic variants in MMR genes compromise this repair mechanism in LS. Consequently, microsatellites exhibit alterations in length, a phenomenon referred to as MSI, which is a defining feature of tumours associated with LS ¹³. The degree of MSI varies depending on the affected MMR protein and the nature of its mutation ¹³. Notably, deficiencies in *MLH1* or *MSH2*, as well as loss of *PMS2*, are linked to high-frequency MSI (MSI-H), affecting both mononucleotide and short tandem repeats ^{34,35}. In contrast, pathogenic variants in *MSH6* tend to result in a lower degree of MSI (MSI-L), predominantly affecting mononucleotide repeats²⁰.

Furthermore, the occurrence of MSI is not limited to genetic variants but can also be triggered by epigenetic changes such as *MLH1* promoter hypermethylation, adding another layer of complexity to MMR dysfunction³⁶. This deficiency in the MMR pathway leads to the accumulation of mutations in genes crucial for cell growth regulation and DNA repair, ultimately initiating tumorigenesis. Thus, the link between MSI and tumorigenesis, particularly the development of LS-associated cancers, is critical. MSI, resulting from dMMR, accelerates carcinogenesis, with the rapid progression from adenoma to CRC occurring within 2–3 years in LS, in stark contrast to the 6–10 years observed in sporadic cases²⁰.

1.2.4.1 Artificial Intelligence based MSI Detection

Traditional methods for MSI detection include polymerase chain reaction (PCR) and immunohistochemistry (IHC). In addition to these, bioinformatic methods, such as MSIsensor and other tools designed for analysing sequencing data, have emerged as valuable alternatives for detecting MSI status. However, both traditional and sequencing-based methods can be labour-intensive, require high-quality DNA, and may lead to ambiguous results in certain cases³⁷.

Artificial intelligence (AI)-based approaches, particularly those leveraging deep learning and machine learning algorithms, have emerged as powerful tools for the automated and accurate detection of MSI status directly from histopathology slides³⁸. These computational methods can be used to analyse complex patterns in histological images that are often imperceptible to the human eye. Echle *et al.* showed that through the integration of AI in MSI detection for CRC, approximately 53% of specimens were accurately classified as microsatellite stable (MSS) and MMR proficient³⁹. Additionally, the AI tool *MSIntuit* has been used to accurately detect MSI in patients with CRC with a high sensitivity of 96-98%⁴⁰. This makes it a potentially valuable pre-screening tool for alleviating the clinical burden of MSI testing by identifying and excluding MSS samples from further analysis. This is complemented by the work of Kather *et al.*, who created a deep-learning model for MSI status classification from histopathology images, offering a significant leap in diagnostic speed and scalability³⁸.

While these AI-driven methods represent a major cost-saving opportunity for MSI testing, they still require tissue samples, as they rely on histology slides for analysis. Thus, they are not non-invasive but do promise a shift towards faster and more cost-effective diagnostic workflows. By streamlining MSI detection, AI-based approaches enable more tailored treatments and offer new insights into cancer biology.

1.2.5 BRAF and KRAS Mutations in LS

KRAS and *BRAF* are oncogenes that are commonly mutated in various types of cancer, including CRC. *BRAF* mutations are commonly found in sporadic CRC cases, particularly the *BRAF*^{V600E} mutation⁴¹. *KRAS* and *BRAF* mutations are less prevalent in LS-associated CRC. *BRAF*^{V600E} mutations are often associated with MSS colorectal tumours and have been linked to a worse prognosis in sporadic CRC cases⁴¹.

1.2.5.1 BRAF^{V600E} Mutation in CRC

The *BRAF* gene codes for a protein involved in cell signalling pathways⁴². The BRAF protein is an essential component of the mitogen-activated protein (MAP) kinase pathway, which plays a critical role in normal cell growth, apoptosis and proliferation⁴². The most common *BRAF* mutation, *BRAF*^{V600E}, is strongly associated with *MLH1* promotor hypermethylation, which leads to the epigenetic silencing of *MLH1*⁴³. As reported in previously published studies, this mutation is frequently seen in sporadic CRCs but not in tumours of LS individuals⁴⁴. Therefore, the detection of this mutation is particularly useful in cases where dMMR is identified by IHC, as it can help distinguish sporadic CRC from heritable CRC. Domingo *et al.* corroborated this, suggesting that *BRAF* screening be utilised in conjunction with dMMR IHC testing to determine whether an individual's CRC is inherited or sporadic⁴⁵. In addition, significant adverse effects of this mutation have been commonly reported in the scientific literature. In a study, by Ogino *et al.*, a large patient cohort was utilised to showcase these adverse effects⁴⁶. This was corroborated in the PETACC-3 study, which showed a significantly worse overall survival of patients with the *BRAF*^{V600E} mutation compared to those with WT gene status⁴⁷. French *et al.* further investigated the combined effect of the mutation and MSI status⁴⁸. The authors concluded that those with MSI cancers and *BRAF*^{V600E} WT gene status had significantly improved outcomes when compared to those with MSI cancers and the *BRAF*^{V600E} mutation⁴⁸.

Coupled with this, Kakar *et al.* reported that patients with MSS and the $BRAF^{V600E}$ mutation displayed a significantly poorer clinical outcome⁴⁹. The presence of the $BRAF^{V600E}$ mutation is therefore indicative of sporadic CRC and is strongly associated with high levels of MSI in sporadic CRCs where *MLH1* is epigenetically silenced. In contrast, in MSS CRCs, the $BRAF^{V600E}$ mutation is linked to lower life expectancy and poorer clinical outcomes. Conversely, the absence of the *BRAF* mutation may indicate LS, especially when dMMR results from a germline mutation affecting an MMR gene. However, it is important to note that dMMR can also result from other mechanisms, such as somatic mutations or epigenetic silencing of MMR genes unrelated to LS.

1.2.5.2 *KRAS* Mutations and Their Association with LS

KRAS mutations, while prevalent in sporadic CRC, are not typically characteristic of LS. The *KRAS* gene, a member of the Ras family, codes for the protein K-Ras⁵⁰. K-Ras forms part of the RAS/MAPK signalling pathway, which is integral to the regulation of cell growth, differentiation, apoptosis, and proliferation⁵⁰. *RAS* is one of the most commonly mutated genes in cancer, as *RAS* mutations are found in approximately 30% of all human cancers⁵¹. These mutations impair the GTPase activity of *KRAS* and prevent the promotion of GTP hydrolysis⁵². Consequently, there is an accumulation of the active form of K-Ras proteins. Additionally, these mutations lead to the activation of downstream pro-proliferative signalling pathways⁵². To demonstrate the severity of this mutation, Tan and Du reported that mutant *KRAS* is found in approximately 35-45% of CRCs, with codons 12 and 13 being hotspots⁵¹. The most common *KRAS* mutations were found to be G12D, G12V, and G13D, as reported by Neumann *et al.*⁵³. The frequency of occurrence of these point mutations was reported to be 14.1% for G12D, 8.5% for G12V, and 7.3% for G13D⁵³. Zlobec *et al.*'s findings confirmed this with G12D, G12V, and G13D occurring in 13.4%, 6.4%, and 6.2% of cases, respectively⁵⁴. Furthermore, patients with $KRAS^{G12D}$ mutations have a poorer prognosis than those with G13D mutations⁵⁴. This indicates that G12D is a more promising prognostic and predictive factor than G13D. Ultimately, *KRAS* mutations are more frequent in sporadic CRCs than in LS-related CRCs.

Hence, investigating *BRAF* and *KRAS* mutations can help assess the hereditary nature of the disease and their impact. Specifically, in cases where dMMR results from a germline mutation, *BRAF* typically remains unmutated, suggesting of LS. Conversely, if dMMR is a result of promoter hypermethylation of *MLH1*, *BRAF* mutations are usually present, indicating sporadic cancer⁴¹. Identification of these mutations can be useful in diagnosing and determining the prognosis of patients as well as in selecting the most appropriate treatment options. In addition, studies have shown that the presence of *BRAF*^{V600E} and *KRAS* mutations are associated with poor survival and an increased risk of recurrence^{55,56}. Therefore, understanding the role of *BRAF* and *KRAS* mutations in cancer development and progression is crucial for improving patient outcomes.

1.2.6 Key Molecular Pathways in Cancer

In the intricate landscape of colorectal and endometrial cancers, distinct molecular pathways play pivotal roles in tumour development and progression. Understanding these pathways is essential for tailoring effective therapeutic strategies.

In LS-associated CRC, the phosphatidylinositol 3'-kinase-protein kinase B (PI3K-Akt) pathway is frequently dysregulated. This pathway is crucial for regulating cell growth, proliferation, and survival. Mutations in genes such as *PIK3CA* and *PTEN* lead to continuous activation of the PI3K-Akt pathway, promoting tumorigenesis in LS patients⁵⁷. A study by Heriyanto *et al.* found a significant association between LS-related CRC and mutations in the *PIK3CA* gene, highlighting the pathway's role in cancer development within this population⁵⁷.

Conversely, the Wnt signalling pathway is predominantly implicated in sporadic CRC cases⁵⁹. Mutations in the *APC* gene or components like *β-catenin* lead to aberrant activation of Wnt signalling, resulting in uncontrolled cell proliferation and tumour formation. This pathway's activation is a hallmark of sporadic CRC, distinguishing it from LS-associated CRC⁵⁹. In sporadic BCs and ECs, the Transforming Growth Factor-beta (TGF-beta) signalling pathway plays a dual role, acting as both a tumour suppressor and promoter depending on the context⁶¹. In early stages, TGF-beta inhibits cell proliferation, but in advanced stages, it can promote tumour progression and metastasis. Mutations or alterations in components of the

TGF-beta pathway are commonly observed in these sporadic cancers, contributing to their development and progression⁶¹. For instance, mutations in the *TGFBR2* gene have been identified in a subset of sporadic ECs, implicating the pathway in tumorigenesis.

Understanding these distinct molecular pathways is vital for developing targeted therapies and improving patient outcomes in both sporadic and hereditary cancer contexts.

1.3 Clinical Features and Diagnostic Criteria

Identifying individuals at risk for LS involves a combination of clinical criteria, predictive models, and tumour screening techniques. Despite advancements, a significant number of patients with LS remain undiagnosed. Various criteria and predictive models have been developed to improve the identification of individuals at elevated risk of LS and hereditary CRCs.

1.3.1 Amsterdam Criteria and Revised Bethesda Guidelines

The Amsterdam Criteria and Revised Bethesda Guidelines are established historical diagnostic criteria used to identify individuals who may have LS⁵⁸. These criteria can help healthcare professionals to determine when genetic testing should be considered. Initially established in 1990 by the International Collaborative Group on HNPCC (ICG-HNPCC), the Amsterdam Criteria were designed to identify LS based on the familial incidence of CRC. In 1999, these criteria were expanded (Modified Amsterdam Criteria) to include non-CRC LS-associated cancers, addressing the need for broader, more inclusive diagnostic criteria (Table 1).

Despite their widespread use, the Amsterdam Criteria prioritises specificity (98%) over sensitivity (22-42%), leading to a low sensitivity rate and the exclusion of more than 50% of LS-affected families^{17,60}. The Revised Bethesda Guidelines emerged as a response, offering a more inclusive set of criteria that considers a wider range of clinical and histopathological features, aiming to enhance the sensitivity of LS detection⁶².

Table 1. Amsterdam Criteria for Identifying Individuals with LS

Amsterdam Criteria
1. Three or more relatives with confirmed CRC, one being a first-degree relative of the other two
2. At least two successive generations affected
3. At least one affected individual diagnosed before age 50
Modified Amsterdam Criteria
1. Three or more relatives with LS-associated cancer, one being a first-degree relative of the other two
2. At least two successive generations affected
3. At least one affected individual diagnosed before age 50

Abbreviations: CRC, colorectal cancer; LS, Lynch Syndrome

The Bethesda Guidelines, introduced in 1997 and revised in 2004, recommend MSI testing for CRC cases meeting specific criteria to improve LS detection (Table 2) ⁶³. These guidelines incorporate clinical and pathological data, including the assessment of tumour-based characteristics, such as MSI and IHC for MMR proteins, along with family history, to enhance diagnostic sensitivity ¹⁴.

Table 2. Bethesda Guidelines for Identifying Individuals with LS

Original Bethesda Guidelines	
1.	Individuals with two LS-related cancers, including synchronous and metachronous CRCs cancers or associated extracolonic cancers
2.	Individuals with CRC and a first-degree relative with CRC and/or LS-related extracolonic cancer and/or colorectal adenoma, one of the cancers diagnosed at age < 45 years, and the adenoma diagnosed at age <40 years
3.	Individuals with CRC or endometrial cancer diagnosed at age <45 years
4.	Individuals with right-sided CRC with an undifferentiated pattern (solid/cribriform) on histopathology diagnosed at age < 45 years
5.	Individuals with signet ring cell type CRC diagnosed at age < 45 years
6.	Individuals with adenomas diagnosed at age < 40 years
Revised Bethesda Guidelines	
1.	CRC diagnosed in an individual under 50 years of age
2.	Presence of synchronous, metachronous colorectal or other LS-associated cancers regardless of age
3.	CRC with MSI-H histology in an individual under 60 years of age
4.	CRC in an individual with one or more first-degree relatives with a LS-associated cancer, with at least one of the cancers diagnosed under 50 years of age
5.	CRC diagnosed in two or more first- or second-degree relatives with LS-related cancers, regardless of age

Abbreviations: LS, Lynch Syndrome; CRC, colorectal cancer; MSI-H, microsatellite high

The evolution from the Amsterdam to the Revised Bethesda Guidelines reflects a shift towards increased sensitivity (82-95%), albeit at the cost of reduced specificity (77-93%)⁶⁴. Despite advancements, challenges remain in accurately identifying LS carriers, with a notable number of patients remaining undetected⁶⁵. Moreover, the effectiveness of these criteria is compromised by the possibility of missing crucial information, such as family history. Consequently, there was a recognised need for a LS detection method that combines both high sensitivity and specificity, as existing clinical criteria fail to identify a significant portion of LS carriers⁶⁶.

1.3.2 Predictive Models for LS Identification

To enhance the identification of LS in patients with CRC, numerous prediction models have been developed that integrate clinical and family history data to estimate the likelihood of

carrying an LS-related variant in MMR genes. These tools vary in their predictive algorithms and patient demographics for which they were designed and validated.

MMRpredict and *MMRpro*, developed in 2006, are web-based platforms that utilise multivariate logistic regression to assess the correlation between genetic test results and factors such as family history and clinical traits⁶⁷. *MMRpredict* provides risk assessments by evaluating sex, age at CRC diagnosis, tumour location, and history of CRC and EC⁶². However, it is known to overestimate and underestimate probabilities, categorising them at extremes of 100% or 0%⁶⁸. It specifically gauges the risk of pathogenic variants of *MLH1* and *MSH2*. Conversely, *MMRpro* employs Bayesian analysis to include a broader spectrum of cancer history and tumour characteristics, such as MSI and IHC results, offering risk evaluations for *MLH1*, *MSH2*, and *MSH6*, similar to the *PREMM*_{1,2,6} model⁶⁹.

Historically, these models did not account for *PMS2* variants, which represent 15% of MMR gene variants⁷⁰. The introduction of *PREMM*₅ aimed to fill this gap by estimating the risk across all five LS-associated genes, thereby providing a more comprehensive risk analysis⁷¹. Based on a large cohort and external validation, *PREMM*₅ considers the personal and familial history of CRC, EC, and other LS-related cancers, significantly broadening the scope of risk assessment by including *PMS2* and *EPCAM* variants⁶⁸.

A cut-off score of $\geq 5\%$ is used for these models to recommend germline genetic testing⁷². Originally developed with a focus on CRC, these models exhibit limited efficacy in differentiating variant carriers from non-carriers in patients with EC at the 5% threshold⁷³. The availability of these models to the public marks progress; however, extending their applicability to a wider array of LS-associated cancers is essential for a more thorough family history evaluation and better identification of individuals who could benefit from LS-specific germline genetic testing.

PREMMplus, introduced in 2017, signifies a pivotal enhancement in predictive modelling by incorporating a broader spectrum of LS-associated genes and considering extracolonic cancers⁷¹. This model, aimed at providing an expansive risk evaluation, acknowledges the

intricate genetic landscape of LS. *PREMMplus*, which includes 19 cancer susceptibility genes, has been validated for its high sensitivity for predicting germline pathogenic variants. Although it encompasses genes of moderate penetrance, leading to a somewhat lower specificity, *PREMMplus* represents a promising approach for more inclusive and accurate LS risk assessment.

1.3.3 Clinical Manifestations and Spectrum of Tumours

Studies have consistently highlighted the elevated risk that LS patients face various additional cancers, including those of the stomach, ovary, pancreas, and several other organs, compared to individuals without LS pathogenic variants^{74–76}. Dominguez-Valentin *et al.* identified a specific association between the *MSH2* genotype and an increased risk for urothelial and prostate cancers.⁷⁷ Tumours in individuals with pathogenic variants in MMR genes often lack MMR proteins, and the frequency of MSI-H varies across tumour types. For instance, MSI-H is detected in 80-100% of stomach, ovarian, colon, and ureter tumours, around 50% in bladder, endometrium, and kidney tumours, and less frequently in breast tumours, with an occurrence of approximately 35%²⁴. In contrast, brain tumours associated with pathogenic MMR variants rarely, if ever, exhibit MSI-H²⁴.

The variability in MSI frequencies across different tumour types in LS patients may be attributable to clonal heterogeneity, which refers to the existence of diverse subpopulations within a single tumour, each with its genetic profile⁷⁸. Latham *et al.* conducted a study on over 15,000 tumours across a spectrum of cancer types for MSI presence, analysing matched constitutional DNA for MMR genes and *EPCAM* alterations, irrespective of MSI status⁷⁹. The study revealed that 16.3%, 1.9%, and 0.3% of the patients with MSI-H, MSI-indeterminate, and MSS tumours, respectively, had pathogenic variants⁷⁹. Notably, half of the LS patients with MSI-H (plus indeterminate) tumours had cancers outside the typical CRC or EC spectrum, including tumour types not typically linked to LS, such as mesothelioma, melanoma, soft tissue sarcoma, and prostate cancer⁷⁹. Additionally, 36% of patients with LS presented with MSS tumours, predominantly non-colorectal or ECs⁷⁹.

The occurrence of MSS tumours in patients with LS could be attributed to the reduced penetrance of MMR gene variants or low tumour purity, which might mask the presence of

MSI ⁸⁰. Alternatively, MSS tumours might be from non-MMR functions of MMR proteins, where the specific mutated MMR gene may not be involved in the development of the tumour²³. Moreover, pathogenic variants in *MSH6*, which are less commonly associated with CRC and may not exhibit MSI, underscore the genetic complexity and variability in tumour manifestations among patients with LS ⁸¹.

1.3.4 LS-Associated Cancers

1.3.4.1 Colorectal Cancer

CRC encompasses cancers arising from the colon and rectum; however, which are often grouped due to the many features that they share ⁸². However, it is important to note that their distinct embryological origins and biological characteristics. Colonic cancers primarily originate from the midgut, while rectal cancers develop from the hindgut, resulting in differences in tumour biology and molecular features ⁸². This distinction is particularly relevant in the context of dMMR. dMMR rectal tumours are rare and almost exclusively associated with LS, reflecting their hereditary nature. In contrast, dMMR can occur sporadically in colonic tumours through mechanisms such as biallelic *MLH1* promoter hypermethylation, which involves simultaneous mechanisms of both alleles ³⁰. These mechanistic differences highlight the diverse underlying dMMR in CRC and underscore the need to consider anatomical and molecular distinctions in CRC studies.

CRC is the most common type of cancer associated with LS. Individuals with LS have a significantly increased risk of developing CRC compared to the general population. In LS individuals, the risk of CRC is approximately 80% for men and 40-60% for women compared to a 4-5% risk in the general population ¹⁵. These cancers often occur at an earlier age and have a propensity for the right side of the colon.

CRC records high levels of inter-patient variability and spatial heterogeneity therefore, targeting more than one abnormality is required to inhibit the growth and survival of cancer cells ^{83,84}. In 2020, it was estimated that there were approximately 0.9 million CRC-related deaths worldwide ⁸⁵. Not only is CRC the third most common malignancy but it is also the fourth most frequent neoplastic cause of death in South Africa ⁸⁶. Moreover, The Centres for

Disease Control and Prevention (United States of America) reported that the CRC incidence is 40 per 100 000 people, with greater incidence rates in developed countries than in African and South Asian countries ⁸⁵.

CRC can be broadly divided into hereditary and sporadic types. Hereditary forms, such as those associated with LS and FAP, constitute approximately 10% of all CRC cases, while the remaining 90% are sporadic⁹. The development of sporadic CRC is thought to arise from a combination of chromosomal instability, MSI, and the cumulative effects of genetic variants, which together increase the likelihood of tumour formation ⁸⁷. Identifying the genetic and epigenetic differences between these cancer types is essential for improving diagnostic accuracy and tailoring therapeutic interventions.

1.3.4.2 Gynaecological Cancers

EC, a malignancy affecting the lining of the uterus, is another frequently observed cancer in individuals with LS. It is estimated that around 5% of all EC cases are hereditary, with LS accounting for the majority ⁸⁸. Women with LS have a notably high risk of developing EC, ranging between 40–60%, a figure that equals or even surpasses their risk for CRC. This is a stark contrast to the general population, where the risk of EC is only about 3% ¹⁵. Moreover, EC tends to manifest at a younger age in LS carriers, typically between 47 and 55 years, compared to the average onset at 63 years in sporadic cases²⁰.

LS not only increases the risk for CRC and EC but also significantly increases the likelihood of ovarian, fallopian tube, and primary peritoneal cancers (PPC) in women, particularly for those with pathogenic variants in *MSH2* and *MLH1* ⁸⁹. PPC originates in the peritoneal lining and is closely related to epithelial OC, sharing similar pathological features and risk factors, especially in the context of LS. The lifetime risk of ovarian cancer (OC) for LS patients is estimated to be 12-15% ⁸⁹. The likelihood and incidence of EC and OC in LS varies significantly depending on the specific MMR gene pathogenic variant. According to the Prospective LS Database, by age 75, the incidence rates for EC and OC vary among gene carriers: approximately 37% for EC and 11% for OC among *MLH1* carriers; 49% for EC and 17% for OC in *MSH2* carriers; and 41% for EC and 11% for OC in *MSH6* carriers ⁹⁰. The endometrioid

subtype of OC, in particular, is often associated with *MLH1* gene pathogenic variants⁹¹. Moreover, it is relatively common for individuals with LS to receive concurrent diagnoses of endometrioid EC and endometrioid OC²⁰.

Interestingly, pathogenic variants in the *MSH6* gene are associated with a milder clinical presentation, with a later onset of both CRC and EC⁸⁸. While the exact lifetime risk for women with pathogenic variants in *PMS2* remains undetermined, the available evidence indicates that they tend to exhibit a milder clinical phenotype compared to those with pathogenic variants in *MLH1* and *MSH2*⁸⁸.

1.3.4.3 Breast Cancer

BC, a malignancy originating from the epithelial cells of breast tissue, remains the leading cause of cancer-related death among women globally⁹². Research has begun to explore the potential inclusion of BC within the spectrum of LS-associated cancers, though findings remain mixed and somewhat controversial⁹³.

Elevated age-specific incidence rate ratios of BC have been reported in LS patients through cohort studies, underscoring a possible significant correlation^{94,95}. da Silva *et al.* reinforced this possibility, revealing a higher prevalence of BC among LS populations alongside an earlier age of onset when compared to the general population. Additionally, this study noted that the incidence rate of BC in LS families is considerably higher than in the general population, with a median age of diagnosis almost 20 years earlier⁹⁶.

Investigations into the molecular characteristics of BC in LS patients, including assessments of MSI, IHC, and MMR gene deficiencies, indicate that LS-associated BCs are more likely to exhibit MSI and loss of MMR proteins, distinguishing them from sporadic BC cases⁹³. Despite these findings, other studies challenge the association between LS and an elevated risk of BC, arguing against the need for increased surveillance for LS patients⁹⁷. This contention points to the complexity of understanding the risk for BC within the LS context and highlights the necessity for further research.

Pathogenic variants in the *MSH6* and *PMS2* genes have been implicated in an increased risk of BC⁹³. A case-control study by Couch *et al.* found that *MSH6* pathogenic variants are significantly associated with an elevated risk for BC⁹⁷. Additionally, Sheehan *et al.* revealed that the prevalence of BC is notably higher in carriers of *PMS2* pathogenic variants within the cohort, compared to carriers of other mutation types⁹³.

1.3.4.4 Hepatobiliary Cancer

The hepatobiliary tract, comprising the pancreas, biliary system and liver, is increasingly recognised for its susceptibility to cancers within the context of LS. This is underscored by recent studies highlighting the distinct risk profile and clinical manifestations of pancreatic (PC) and biliary tract cancers (BTC) among LS patients, distinguishing them from sporadic cases⁹⁸. Notably, research from the Prospective Lynch Syndrome Database reveals that by the age of 75 years, the risk for PC is 6.2%, 0.5% and 1.4% and for BTC is 3.7%, 1.7% and 0% respectively for patients with pathogenic variants in *MLH1*, *MSH2* and *MSH6*⁹⁸. Interestingly, Zalevskaja *et al.* found pathogenic variants in *PMS2* not to be significant in PC and BTC patients with LS⁹⁸. This finding was corroborated by Møller *et al.*, who reported no PC or BTC among 124 carriers of the *PMS2* pathogenic variant⁹⁰. While the incidence of sporadic PC and BTC typically increases with age, with the median age at diagnosis around 70 years, LS-associated cases present at a younger age⁹⁹. This pattern aligns with the early onset characteristic of all cancers linked to LS.

1.3.4.5 Gastric Cancer

GC, also known as stomach cancer, is a malignant tumour that develops from the cells lining the inner surface of the stomach. The relationship between GC and LS has been a topic of debate, with the earliest recognitions of this connection observed in Family G, where both gastric and CRC were prevalent¹⁰. However, the reported incidence of GC in LS patients varies, ranging from 6 to 13%, with geographical and ethnic variations contributing to this diversity^{100,101}. The types of GCs in LS predominantly fall into the intestinal category, accounting for 62-79% of cases, while a smaller proportion of cases are diffuse or poorly differentiated, making up 23-32% of cases¹⁰².

A study by Kim *et al.* that involved 3,828 carriers of pathogenic variants found that 1.1% had a history of GC, and 9.1% had one or more first- or second-degree relatives with the disease. The study highlighted several significant risk factors for GC in this population, including male sex, older age, pathogenic variants in *MLH1* or *MSH2*, and having a first-degree relative with GC¹⁰³. Further, Cho *et al.*, studied 118 LS patients, 26 were diagnosed with gastric neoplasms by age 70 years. Remarkably, half of these patients developed multiple gastric neoplasms, underscoring the tendency towards multiple tumour formations in LS-associated GC. The vast majority of these tumours were associated with intestinal metaplasia and were predominantly dMMR, further emphasizing the critical role of LS in the development of these gastric tumours¹⁰⁴.

1.3.4.6 Small Intestinal Cancers

Small intestinal cancer, also known as small bowel cancer, refers to epithelial malignancies that originate in the small intestine. Although rare, individuals with LS face a lifetime risk of 4% for developing this cancer, a risk significantly higher than the 0.3% risk observed in the general population^{105,106}. This stark contrast underscores the increased susceptibility of LS carriers to small bowel cancer, emphasising the significance of this condition.

As expected in LS-associated cancers, Park *et al.* found that patients with LS develop small bowel cancer at an early age. Specifically, this study revealed that more than 10% of the LS patients were diagnosed with small intestinal cancer before the age of 30 years¹⁰⁷. This early onset contrasts with the typical age of cancer development in the general population, which is commonly between ages of 60 and 70 years¹⁰⁸.

An analysis by Koornstra *et al.* of 199 small intestinal tumours in LS patients showed that 43% were located in the duodenum, 37% in the jejunum, and 20% in the ileum, which is a distribution pattern somewhat distinct from sporadic cases that are more frequently found in the duodenum. Furthermore, the incidence of small bowel cancer appeared consistent among carriers of *MLH1* and *MSH2* pathogenic variants and was also present in those with *MSH6* and *PMS2* pathogenic variants, although to a lesser extent¹⁰⁸.

1.3.4.7 Uroepithelial Tract

The upper uroepithelial tract, also known as the upper urinary tract, refers to the parts of the urinary system that are located more proximally, specifically the renal pelvis and the ureters. The upper uroepithelial tract is notably susceptible to cancers in individuals with LS, particularly upper tract urothelial carcinoma (UTUC)¹⁰⁹. LS patients face a lifetime UTUC risk of up to 20%, with those harbouring *MSH2* pathogenic variants at the highest risk^{109,110}. Characteristically, UTUC in LS patients manifests and is often without the usual risk factors seen in the general population¹⁰⁹.

Møller et al. found that the incidences by age 75 years for urinary tract cancers are 8% for *MLH1* variants, 25% for *MSH2*, and 11% for *MSH6*, indicating a significantly elevated risk for *MSH2* variant carriers⁹⁰. UTUC accounts for 5-10% of all urothelial carcinoma cases but is distinguished by its aggressive nature at diagnosis, with approximately 60% of cases being invasive and 7% metastasized upon initial detection¹¹⁰.

1.3.4.8 Brain Cancer

Individuals with LS face an increased risk of brain tumours, including glioblastomas, astrocytomas, and oligodendrogliomas¹¹¹. One specific manifestation of LS in some patients is Turcot syndrome (TS), characterized by the co-occurrence of multiple colorectal adenomas and brain cancer, particularly high-grade gliomas (HGG), which are seen four times more frequently in LS patients compared to the general population^{111,112}.

TS can be categorized into two types based on mutations: Type I, related to MMR mutations leading to gliomas, and Type II, associated with *APC* mutations resulting in medulloblastomas as part of FAP¹¹¹. Furthermore, constitutional MMR deficiency (CMMR-D) syndrome, characterized by biallelic germline MMR variants, presents a distinct clinical profile involving haematological malignancies, central nervous system tumours (including brain tumours), LS spectrum tumours, and multiple intestinal polyps¹¹³.

Additionally, glioblastoma development has been linked to Muir–Torre syndrome, a variant of LS primarily associated with *MSH2* pathogenic variants and characterized by sebaceous

skin tumours¹¹¹. The occurrence of brain tumours alongside CRC, under the umbrella of TS, highlights the complex interplay between classical LS, CMMR-D syndrome, and Muir–Torre syndrome in predisposing individuals to a diverse spectrum of malignancies, including notably increased risks for brain tumours.

1.4 Conclusion

The importance of research into LS lies in the potential to provide a more comprehensive understanding of the genetic underpinnings of this condition, which could lead to the development of new treatment and intervention strategies. Identifying the driver genes and mutations involved in each cancer type associated with LS may improve understanding of the disease and help develop more effective screening, prevention, and treatment strategies for individuals with this condition. Moreover, with LS affecting approximately 1 in 660-2,000 individuals in the general population, understanding its underlying mechanisms is of utmost importance to public health. Early detection and management of LS are critical to reducing morbidity and mortality associated with this condition¹¹⁴. Identifying the disrupted or expressed pathways that lead to LS-associated cancers may help us better understand the disease, and in turn, develop more effective treatment and prevention strategies. Additionally, LS-associated cancers are often aggressive and difficult to treat, highlighting the need for targeted therapies¹¹⁴. Investigating LS genetics may pave the way for the development of targeted therapies that could improve cancer management by identifying new therapeutic targets. Such therapies could potentially reduce the burden of cancer associated with LS and improve patient outcomes¹¹⁴. Ultimately, this research has the potential to reduce morbidity and mortality associated with LS and improve patient outcomes.

1.5 Aims and Objectives

This study aims to improve the ascertainment of LS by investigating pathways associated with cancers within the LS tumour spectrum. Using real-world data from the open-source platform cBioPortal (<https://www.cbioportal.org>), this research will leverage mutational and messenger RNA (mRNA) data from triaged likely LS (LLS) and likely sporadic (LSp) cases. Focusing on CRC, EC and BC, the study aims to uncover driver genes and dysregulated

pathways specifically associated with LS, identifying genetic alterations that may contribute to cancer development and progression.

The ultimate goal is to deepen understanding of the genetic basis of LS-associated cancers and to identify potential targets for future treatment and interventions. Additionally, this study aims to explore the potential inclusion of BC in the LS tumour spectrum by examining pathway dysregulation in LLS-BC and LSp-BC and comparing it with dysregulation patterns observed in LLS-EC. Through these analyses, the study aims to provide new insights into subtype-specific pathways that may guide personalised therapeutic strategies for LS-associated cancers.

Study Objectives

- To triage LS-associated cancer samples into LLS and LSp subtypes
- To identify pathways dysregulated in LLS and LSp subtypes of LS-associated cancers
- To examine driver genes and genetic alterations contributing to tumour development and progression in LLS and LSp subtypes of LS-associated cancers
- To investigate potential therapeutic drugs for LLS and LSp subtypes of LS-associated cancers based on identified pathways
- To explore differences in clinical and molecular characteristics between LLS and LSp subtypes of LS-associated cancers
- To evaluate the potential inclusion of BC within the LS tumours spectrum

Chapter 2: Elucidation of Dysregulated Pathways in Sporadic and Lynch Syndrome-associated Colorectal Cancer – Implications for Treatment

2.1 Introduction

The current study investigates molecular distinctions between LS-associated CRC and sporadic CRC through an in-depth multiomic analysis. Despite extensive research, the precise cellular pathways through which MMR pathogenic variants contribute to LS-associated cancer development remain unclear. Although LS tumours exhibit distinct genetic and molecular characteristics, the link between these and dysregulation of specific signalling and metabolic pathways has not been fully understood. Elucidating these mechanisms is essential for advancing the diagnosis, management, and treatment of LS-related cancers.

The advent of next-generation sequencing (NGS) technologies and the availability of large-scale datasets, such as those from The Cancer Genome Atlas (TCGA) and cBioPortal, have greatly improved the ability to characterise the molecular landscape of CRC and a wider range of other cancers^{115,116}. These comprehensive datasets in addition to demographic, clinical and pathology information, also provide extensive genomic, transcriptomic, and proteomic data, that can be used to explore the underlying mechanisms of tumorigenesis in both LS and sporadic CRC. Previous studies have leveraged these resources to identify key mutations involved in CRC¹¹⁷. However, integrating these diverse data types to achieve a comprehensive understanding of the distinct CRC subtypes remains a challenge, largely because of the technical difficulties in harmonising data across different platforms, the computational complexity of managing such large datasets, and the analytical hurdles in drawing meaningful biological insights from multi-omic correlations.

While a few previous studies have sought to characterise the genomic and mutational landscape of CRC, few studies have focused on the dysregulated pathways in LS-associated CRC¹¹⁸. The present study aims to address the critical knowledge gap by investigating the distinct molecular pathways disrupted in LS-associated and sporadic CRC. By integrating genomic and transcriptomic data from CRC patients, the current work sought to identify specific signalling and metabolic pathways that distinguish each CRC subtype.

This particular integrative multi-omic analysis advances the understanding of the distinct molecular characteristics of LS and sporadic CRC, paving the way for the development of targeted therapeutic strategies such as pathway-specific inhibitors or personalised treatment regimens tailored to CRC subtypes. By filling existing knowledge gaps, this study contributes to the broader goal of personalising cancer treatment and improving patient outcomes in LS.

2.2 Materials and methods

2.2.1 Data Acquisition

Data from the Pan Cancer Atlas project of TCGA was accessed via the cBioPortal platform^{116,119–121}. This dataset included clinical data for 594 CRC patients, WES data for 534 patients, and transcriptomic data for 592 patients, obtained through RNA-seq. Due to sample availability and assay failures, not all data types were available for all patients. All mutations analysed in this study are somatic, identified from tumour samples. No germline data is available for these patients in the TCGA.

2.2.2 Pathogenic Variant Identification

Several pathogenicity predictors were employed to assess the pathogenicity of genetic variants, including InterVar, ClinPred, ClinVar, and MutPred. Variants classified as "pathogenic" or "likely pathogenic" by InterVar or ClinVar, designated as "D" (deleterious) by ClinPred, or assigned a MutPred score ≥ 0.68 (indicating a 10% false discovery rate) were retained for downstream analyses.

2.2.3 Classification of LLS- and LSp-CRC

Samples were classified based on a multi-step approach.

2.2.3.1 Initial Triaging: Samples with a pathogenic variant in at least one MMR gene and a diagnosis age of ≤ 60 years were classified as LLS. Samples without MMR variants and a diagnosis age > 60 years were classified as LSp. Samples lacking mRNA expression data were excluded from further analyses.

2.2.3.2 Validation with mRNA Expression Data and Cluster

Analysis: supervised hierarchical clustering was applied to RNA-seq data to refine classifications. Genes exhibiting minimal variance were excluded before clustering. To address class imbalance, the SMOTE algorithm was used, ensuring adequate representation of the minority class (LLS). SMOTE works by generating synthetic data points for the minority class by interpolating between existing samples, thus creating a more balanced dataset without simply duplicating samples. This approach helps prevent the clustering algorithm from being biased towards the majority class (LSp) and improves the ability to detect meaningful clusters across both groups¹²². Feature selection was conducted using the maximum relevance-minimum redundancy (mRMR) method, followed by Neighbourhood Component Analysis (NCA)¹²³. NCA was performed in Python 3, as R Studio did not have the necessary package support for this method. K-means clustering was applied to identify patterns within the data. K-means clustering is an unsupervised learning algorithm that groups data into K distinct clusters based on their similarity¹²⁴. In the analysis, silhouette scores were used to determine the optimal number of clusters (K=2), which corresponded to the two primary subtypes: LLS and LSp. K-means clustering was selected due to its simplicity and effectiveness in identifying distinct molecular subtypes within large, high-dimensional datasets such as transcriptomic data used in this study¹²⁴. Principal component analysis (PCA) plots were generated to visualise clustering patterns and highlight the relationship between genetic profiles and clinical classifications. Samples clustering in different groups compared to the originally triaged groups were manually reviewed by looking at clinical, mutational, RNA and methylation data. If there was enough evidence to prove that the sample was not LLS or not LSp it was removed from further analyses. Subtypes were compared to those categorised based on MMR variant status and age of diagnosis.

2.2.4 Mutational Signature Analysis

A mutational signature analysis was conducted using simple nucleotide variation data from TCGA Pan Cancer Atlas dataset. Mutation Annotation Format (MAF) files were processed to extract mutational signatures based on trinucleotide mutation contexts mapped to the human genome reference assembly 38 (hg38). To identify the mutational signatures, Non-negative Matrix Factorisation (NMF) was applied to decompose the mutational matrices and extract de novo mutational signatures from the data¹²⁵. These signatures were then compared to the known mutational signatures catalogued in the COSMIC using cosine similarity. In particular, signatures associated with dMMR were focused on, including COSMIC signatures 6, 15, 20, and 26, which are linked to MSI and hypermutated phenotypes in CRC. For downstream analysis, samples with significant contributions (>10%) from dMMR-associated signatures were flagged. These flagged samples were prioritised for further validation and stratification based on their classification as LLS-CRC or LSp-CRC. LLS samples not showing significant contributions from dMMR signatures were re-evaluated, and, if necessary, excluded from subsequent analyses. Conversely, LSp-CRC samples exhibiting high contributions from dMMR signatures were further investigated to assess whether they should be removed, particularly when epigenetic alterations such as *MLH1* hypermethylation were observed.

2.2.5 DGE and Pathway Enrichment Analysis

To investigate DGE between the CRC subtypes, RNA-seq data was utilised. In MATLAB, differential expression was tested between the subtypes using a negative binomial model, with local regression applied to adjust for variance between the groups. Mean expression levels for each gene in both subtypes were calculated, and the log2 fold changes were computed to quantify the relative differences in gene expression. *p*-Values from the differential expression analysis were adjusted for multiple comparisons using the Benjamini-Hochberg false discovery rate (FDR) method, with genes showing an adjusted *p*-value (*q*-value) of less than 0.05 considered differentially expressed. GSEA was performed to identify pathways enriched among the DEGs, using the KEGG database as a reference. Enrichment analysis was conducted through the *Enrichr* platform, and pathways with *q*-values < 0.05

were deemed significantly enriched¹²⁶. This analysis revealed key dysregulated pathways in LS-CRC, including the PI3K-Akt, calcium, cAMP, and IL-17 signalling pathways. Furthermore, gene expression levels were mapped onto template WikiPathways for the PI3K-Akt signalling pathway using the PathVisio3 software¹²⁷. Z-score normalised expression data were categorised into three levels: (1) low for z-scores below -0.5; (2) no change for z-scores between -0.5 to 0.5; and (3) high for z-scores above 0.5. This scaling was consistently applied to capture gene expression variations across entire pathways.

2.2.6 Prediction of Regulator Kinases and TFs

Upstream regulatory kinases and TFs likely influencing the observed differences in gene expression between the subtypes were computationally predicted using X2K¹²⁸. X2K employs a reverse-engineering, network-based approach to predict upstream regulatory kinases based on existing knowledge.

2.2.7 Mutation Analyses

The spectrum of genomic mutations across CRC subtypes were evaluated by focusing on OGs and TSGs. A comprehensive list of OGs and TSGs was obtained from OncoKB, comprising 740 OGs and TSGs¹²⁹. This list was used to extract mutations expected to have an impact on CRC oncogenesis. Specifically, pathogenic variants in known OGs and TSGs were filtered for. Using the processed data, frequently altered genes within each subtype were identified. The frequency of mutated genes was calculated for both subtypes, and Fisher's exact test was applied to identify statistically significant differences. To control for multiple comparisons, *p*-values were adjusted using Benjamini-Hochberg correction. All significant genes were found to be enriched in the LLS-CRC subtype.

2.2.8 Drug Sensitivity and Dependency Analysis

Drug sensitivity and dependency analysis was performed using data from the GDSC database¹³⁰. Dose-response data for 344 cancer cell lines treated with PI3K-Akt pathway inhibitors (Wnt pathway inhibitors for the LSp-CRC drug analyses) were retrieved. Cell lines were characterised into high- and low-dependency groups based on CRISPR-derived dependency scores for PI3K-Akt pathway genes. To assess drug efficacy, the mean IC50 values

for each inhibitor between the high- and low-dependency groups were compared. The relationship between gene dependencies and drug sensitivity was further evaluated using a linear regression model. Significant differences in drug sensitivity between the groups were identified using a Student's t-test, with p -values adjusted for multiple comparisons.

For each pathway gene, the cell lines were stratified into two groups: those with a high CRISPR-determined dependence on the gene (e.g., a dependence score < -0.5) and those with a low dependence on that gene (e.g., dependence score > 0.5). The IC₅₀ values for each of the 51 PI3K-Akt pathway inhibitors were then compared between the two groups. Additionally, the cell lines were classified based on the mutation status of specific PI3K-Akt pathway genes. For example, cell lines with a mutation in a particular gene were compared to those without, and IC₅₀ values for each of the 51 PI3K-Akt pathway inhibitors were analysed between mutant and non-mutant groups.

Next, the number of PI3K-Akt pathway genes that were either “common essential” or “strongly selective” across each cell line were quantified. This yielded an absolute count of PI3K-Akt pathway genes that each cell line is most dependent on. It was hypothesised that cancer cell lines with greater dependency on PI3K-Akt pathway genes would show enhanced sensitivity to PI3K-Akt pathway inhibitors. To test this, the cell lines were divided into two categories: (1) those with more than the median number of “common essential” or “strongly selective” PI3K-Akt pathway genes (indicating higher PI3K-Akt pathway dependency), and (2) those with fewer than the median number of such genes (indicating lower PI3K-Akt pathway dependency). The mean IC₅₀ values for the 51 PI3K-Akt pathway inhibitors were then compared between these two groups.

To evaluate the association between PI3K-Akt pathway gene dependencies and mRNA transcription profiles following PI3K-Akt pathway inhibitor treatment, mRNA transcription responses profiled by the LINCS project were obtained for six cancer cell lines after small molecular inhibitor perturbations. The dataset included information on the names and concentrations of anti-cancer drugs used to treat six cell lines, as well as the mRNA transcription responses following drug treatment. Using MATLAB, the dataset was filtered to

include only cancer cell lines with corresponding dose-response profiles from the GDSC and CCLE databases for the PI3K-Akt pathway inhibitors represented in the LINCS dataset. This analysis revealed that only the MCF7, A375, YAP and A549 cell lines, when treated with PI3K-Akt inhibitors OSI-027 and alpelisib, were common across the databases. Consequently, the mRNA transcription signatures associated with PI3K-Akt pathway activity in these cell lines following treatment with OSI-027 and alpelisib were assessed.

2.2.9 Statistical Analyses

All statistical analyses were performed using R Studio, unless stated otherwise. Docker was used to encapsulate the computational environment and ensure the reproducibility of all statistical analyses. The 'animates/anima:4.2.2' Docker image was employed to create a container for RStudio, facilitating consistency in data processing and analysis. Continuous variables were summarised using the median and interquartile range, with comparisons made using Wilcoxon rank-sum test. Categorical variables were summarised using count and percentage, with comparisons made using Fisher's exact test. All statistical tests were two-sided, and p -values were corrected for multiple comparisons according to the Benjamini-Hochberg method. Results with $p < 0.05$ or $q < 0.05$ (adjusted p -value) considered significant.

2.2.10 Data, Code and Supplementary Files Availability

All datasets used in this study were obtained from publicly accessible sources. Detailed instructions for accessing these datasets are available via the respective platforms (Appendix C, Table C1) The code used for all analyses is publicly available on GitHub (<https://github.com/may-julia/Elucidation-of-Dysregulated-Pathways-in-Sporadic-and-Lynch-Syndrome-Colorectal-Cancer.git>). Supplementary files can be found online at (<https://doi.org/10.5281/zenodo.14228072>). The supplementary file descriptions are provided in appendix B

2.3 Results

2.3.1 Dataset Assembly and Initial Classification

A comprehensive CRC dataset from the TCGA Pan-Cancer Atlas project was compiled and accessed via cBioPortal, consisting of clinical information for 326 samples, along with

corresponding transcription data from RNA sequencing (RNA-seq) and genomic somatic mutational data derived from whole-exome sequencing (WES)^{116,119–121}. The aim was to apply unsupervised clustering techniques, such as K-means clustering, in combination with integrative pathway and network analyses, to classify CRCs into distinct subtypes and identify potential oncogenic mechanisms within each subtype.

2.3.2 Identifying LLS- and LSp-CRC Samples

To classify LLS- and LSp-CRC samples, a multi-step approach was adopted that integrated mutational, clinical and transcriptomic data. Initially, samples were classified based on the presence of pathogenic MMR gene variants and patient age of diagnosis. Samples were classified as LLS if they harboured at least one pathogenic variant affecting an MMR gene and the patient was diagnosed at or below 60 years of age, a threshold commonly used in clinical practice to differentiate hereditary from sporadic cases of CRC, although it is not entirely reliable. Conversely, samples without pathogenic MMR variants and a diagnosis age greater than 60 years were classified as LSp.

To validate the classifications, unsupervised K-means clustering was applied to mRNA expression data derived from RNA-seq experiments. Clustering was performed after filtering out low-expression and low-variance genes. To determine the optimal number of clusters ($K = 2$) silhouette analysis was used, which demonstrated that the majority of samples grouped into two distinct molecular subtypes corresponding to the original LLS and LSp classification, with high concordance (Appendix A, Figure S1). To address the imbalanced representation of the two subtypes (LLS $n = 13$ vs. LSp $n = 313$), the Synthetic Minority Over-sampling Technique (SMOTE) was applied. SMOTE generates synthetic samples by interpolating between existing minority class samples, ensuring a balanced dataset during clustering¹²².

Upon clustering, 13 samples that were initially classified as LLS were grouped into one cluster, while the majority of LSp samples grouped into the second cluster. Two samples, originally classified as LSp based on mutational data, clustered with the LLS group (Figure 2A and 2B). Further investigation revealed that these two samples exhibited somatic *MLH1* hypermethylation and reduced *MLH1* expression, which often leads to MSI similar to that

seen in LS-associated cancers. However, because this deficiency arose from epigenetic changes (*MLH1* hypermethylation) rather than a pathogenic variant affecting an MMR gene, their classification was retained as LSp to maintain the genetic distinction between subtypes (Figure 3).

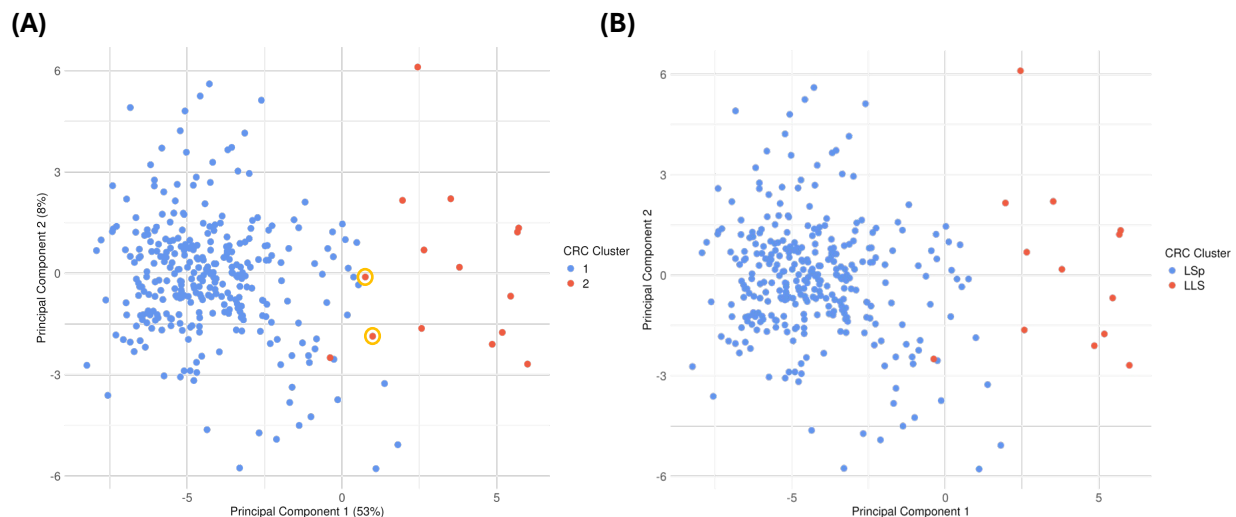


Figure 2. Visualisation of Subtypes in CRC Samples. (A) Principal Component Analysis (PCA) plot illustrating the two clusters of 326 CRC tumours identified using K-means clustering, where Cluster 1 is shown in blue, and Cluster 2 in red. These clusters represent distinct molecular subtypes. The two samples with *MLH1* hypermethylation that were initially classified as likely sporadic (LSp) but clustered with likely Lynch Syndrome (LLS) are highlighted with yellow circles (B) PCA plot highlighting the initial classification of CRC samples LLS in red and LSp in blue. Classification is based on mismatch repair gene variants and age at diagnosis.

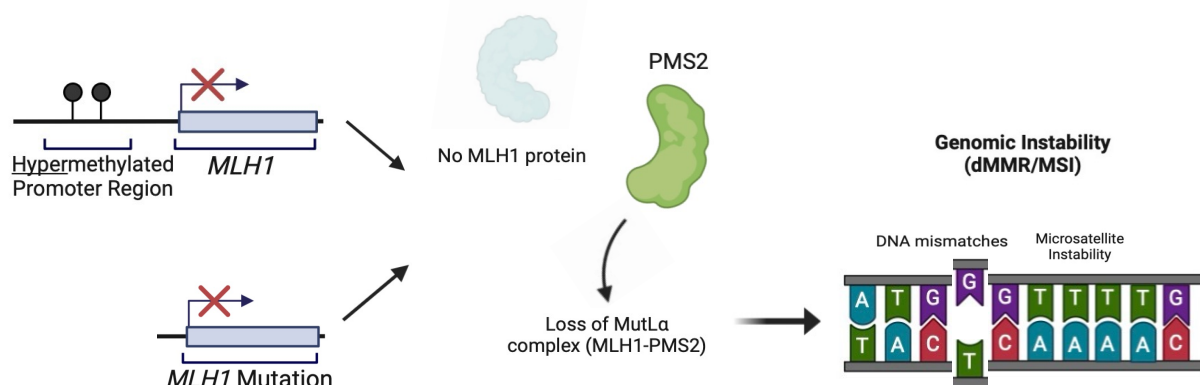


Figure 3. Mechanism of *MLH1* Hypermethylation or Mutation Leading to Deficient Mismatch Repair (dMMR) and Genomic Instability. The pathway leading to genomic instability (dMMR/MSI) associated with *MLH1* hypermethylation or mutation. The top panel shows how biallelic hypermethylation of the *MLH1* promoter region results in the absence of MLH1 protein expression, preventing the formation of the MutLα complex (MLH1-PMS2). The lower panel shows where a somatic pathogenic *MLH1* added to an inherited constitutional pathogenic variant similarly results in the loss of the MutLα complex. In both cases, the absence of the functional MutLα complex leads to DNA dMMR, causing genomic instability manifested as DNA mismatches and/or microsatellite instability.

In addition, the contribution of dMMR signatures was analysed using mutational signature analysis. Samples exhibiting significant contributions (>10%) from dMMR-associated signatures – specifically Catalogue of Somatic Mutations in Cancer (COSMIC) signatures 6, 15, 20, and 26, which are characterised by high mutation rates and MSI – were flagged for further review ¹³¹. This approach helped refine the classification. For instance, one sample initially classified as LLS based on age and mutation data did not exhibit dMMR signatures but instead carried polymerase proofreading mutations in the *POLE* gene, associated with a distinct mutational process and a higher mutation burden unrelated to dMMR. Consequently, this sample was excluded from the final LLS classification, resulting in a refined cohort of 12 LLS samples: eight right-sided colonic tumours and four rectal tumours.

Epigenetic alterations were also examined, particularly *MLH1* hypermethylation, which can silence gene expression causing dMMR, in samples with ambiguous classification. Samples with hypermethylation of *MLH1* but lacking pathogenic MMR variants were treated as LSp cases. This multilayered validation ensured that the final classification was robust, capturing the distinct characteristics of LLS and LSp CRC.

2.3.3 Clinical and Molecular Characteristics of CRC Subtypes

To better understand the distinct clinical and molecular characteristics of LLS and LSp CRC, key clinical features and mutational profiles were analysed. First, the frequency of pathogenic variants in the MMR genes across the 12 LLS samples was examined. Mutations were classified as pathogenic using a combination of tools and criteria, including annotations from Intervar, ClinPred, Clinvar, and MutPred. The analysis revealed that *MSH2* variants were the most prevalent, accounting for six (29%) of the 21 detected pathogenic MMR variants, which aligns with the known role of *MSH2* in LS. *PMS2*, *MLH1* and *MSH6* each harboured five pathogenic variants (24% of 21 variants). No pathogenic alterations in *EPCAM* were found (Table 3). Several samples harboured variants in more than one MMR gene, which is not uncommon in dMMR tumours given their high mutational burden. Importantly, the majority of samples with multiple MMR variants included a secondary *MSH6* variant, a phenomenon that is not uncommon in dMMR tumours. These secondary mutations likely arise as a consequence of the defective MMR system, which fails to correct replication errors, leading

to additional variants in other MMR genes. For instance, sample TCGA-CA-6717 exhibited variants in *MSH2*, *PMS2*, and *MSH6*, while sample TCGA-AG-A002 showed pathogenic variants in *MSH2*, *MLH1*, and *MSH6*. This highlights the complex interplay of primary and secondary variants in driving tumorigenesis in LS-associated cancers.

In contrast, the LSp samples predominantly exhibited variants in the *APC* and *TP53* genes, which are recognised as key molecular drivers in sporadic CRC. This finding highlights the distinct genetic mechanisms underlying hereditary and sporadic CRC. While pathogenic MMR gene variants are central to the pathogenesis of LS-associated CRC, their primary role is causing dMMR, which facilitates the accumulation of secondary mutations that act as drivers of tumorigenesis. These results reflect the genetic diversity within LLS tumours and underscore the critical role of dMMR in enabling the development of LS-associated CRC.

Table 3. Detailed Distribution of Mismatch Repair Gene Variants in the 12 Likely Lynch Syndrome Samples

Sample ID	Location	Side (if colonic)	MMR Gene Variants	Variant Type	MSI Status
TCGA-5M-AAT6	Colon	Right	<i>MSH2</i>	SNV	MSI
TCGA-AA-A01R	Colon	Right	<i>PMS2</i> , <i>MSH6</i>	SNV	MSI
TCGA-CA-6717	Colon	Right	<i>MSH2</i> , <i>PMS2</i> , <i>MSH6</i>	SNV	MSS
TCGA-CM-6674	Colon	Right	<i>MLH1</i>	Deletion	MSI
TCGA-D5-6530	Colon	Right	<i>MLH1</i>	SNV	MSI
TCGA-F4-6856	Colon	Right	<i>MLH1</i>	Deletion	MSI
TCGA-G4-6588	Colon	Right	<i>PMS2</i>	SNV	MSI
TCGA-WS-AB45	Colon	Right	<i>MSH6</i>	SNV	MSI
TCGA-AG-3892	Rectum	-	<i>MSH2</i> , <i>MSH6</i>	SNV	MSI
TCGA-AG-A002	Rectum	-	<i>MSH2</i> , <i>MLH1</i> , <i>MSH6</i>	SNV	MSS
TCGA-EI-6917	Rectum	-	<i>MSH2</i> , <i>MLH1</i> , <i>PMS2</i>	SNV	MSS
TCGA-F5-6814	Rectum	-	<i>MSH2</i> , <i>PMS2</i>	SNV	MSS

Abbreviations: MMR, mismatch repair; MSI, microsatellite instability; MMS, microsatellite stable; SNV, single nucleotide variant

The MSI status and tumour mutation burden (TMB) between LLS and LSp samples were compared (Figure 4A and 4B). The analysis showed that the majority of LLS samples (eight of 12; 67%) were classified with MSI, reflecting the defective DNA repair mechanisms

characteristic of LS. MSI status is a critical biomarker in CRC, influencing both prognosis and therapeutic strategies, particularly in determining eligibility for immunotherapy ¹³². In contrast, 87% of LSp samples (272 of 313; 87%) were MSS. Additionally, LLS samples exhibited a significantly higher TMB compared to LSp samples, consistent with their hypermutated phenotype. A higher TMB is often associated with increased neoantigen load, which can enhance the immunogenicity of the tumour and potentially make it more responsive to immunotherapies ¹³³. These findings further differentiate the molecular profiles of LLS and LSp CRCs, emphasizing the distinct oncogenic processes driving each subtype.

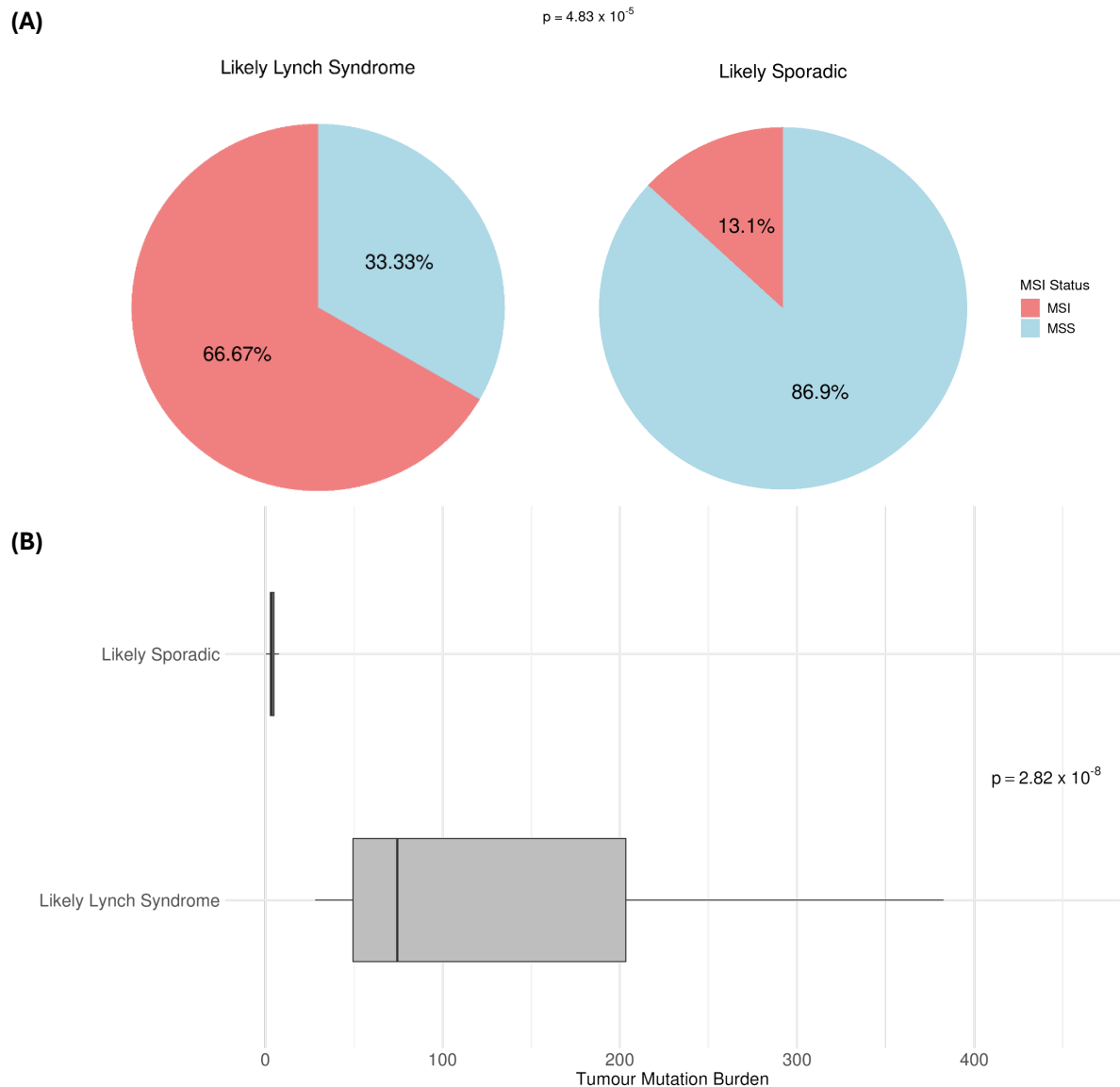


Figure 4. Comparison of Microsatellite Instability (MSI) and Tumour Mutation Burden (TMB) Between Likely Lynch Syndrome (LLS) and Likely Sporadic (LSp) Colorectal Cancer (CRC) Samples. (A) Pie charts displaying the distribution of MSI and microsatellite stable (MSS) status among LLS (n=12) and LSp (n=313) samples. The p -value was from Fishers test **(B)** Box and whisker plots comparing the TMB between the two subtypes, highlighting the higher mutation burden in LLS samples. The p -value was from the Wilcoxon rank-sum test.

Further analysis of clinical data revealed that LLS tumours were more frequently located on the right side of the colon, whereas LSp tumours occurred predominantly on the left. Although, this difference was not statistically significant, right-sided colorectal tumours are generally associated with different molecular profiles and often have a poorer prognosis compared to left-sided tumours, highlighting the potential clinical relevance of tumour location in CRC. Additionally, no significant associations were observed between sex or AJCC pathological tumour stage and the classification of CRC cases ad LLS or LSp (Table 4).

Table 4. Clinical and Molecular Features of the Two Colorectal Cancer Groups

Characteristic	LLS CRC (n=12)	LSp CRC (n=313)	p-Value
Sex			0.56
Male	8 (66.7%)	170 (54.3%)	
Female	4 (33.3%)	143 (45.7%)	
<i>BRAF</i> ^{V600E} Mutation			1
Wild type	11 (91.7%)	284 (90.7%)	
Mutated	1 (8.3%)	29 (9.3%)	
Pathogenic <i>KRAS</i> mutation			0.56
Wild type	8 (66.7%)	175 (55.7%)	
Mutated	4 (33.3%)	138 (44.3%)	
Microsatellite Status			4.84 x 10 ⁻⁰⁵ *
MSI	8 (66.7%)	41 (13.1%)	
MSS	4 (33.3%)	272 (86.9%)	
AJCC pathologic tumour stage			0.73
I	4 (33.3%)	63 (20.5%)	
II	5 (41.7%)	122 (39.6%)	
III	2 (16.7%)	84 (27.3%)	
IV	1 (8.3%)	39 (12.7%)	
Pathogenic variant count, median, IQR	965; 700-3095	52; 40–69	2.55 x 10 ⁻⁰⁸ *
TMB, median, IQR	83.4; 54-251	3.7; 3-5	2.82 x 10 ⁻⁰⁸ *
Primary Tumour Site			0.24
Right	8 (66.7%)	123 (39.3%)	
Left	4 (33.3%)	173 (55.3%)	
Unknown	0 (0%)	17 (5.4%)	

Abbreviations: LLS, likely Lynch Syndrome; LSp, likely sporadic; CRC, colorectal cancer; MSI, microsatellite instability; MSS, microsatellite stable; IQR, interquartile range; TMB, tumour mutation burden. *p*-Values for continuous variables and categorical variables were from the Wilcoxon rank-sum and Fishers tests.

*Significant

2.3.4 The Pathogenic Mutational Landscape of CRC Subtypes

The somatic DNA from all 325 CRC samples was analysed for alterations in 14,136 genes using WES data. As anticipated, the median number of pathogenic variants per sample was significantly higher in LLS-CRC compared to LSp-CRC (**965**; IQR: 700-3095 vs **52**; IQR: 40–69; $p = 2.6 \times 10^{-8}$), suggesting that the hypermutated phenotype in LLS-CRC may contribute to increased tumorigenesis (and potential resistance to therapies).

A total of 23,973 pathogenic somatic mutations were identified across all LLS samples, involving 9,418 different genes (Supplementary File 1). The median number of somatically altered genes that contained pathogenic variants, per LLS sample, was **908** (IQR: 655-2,450). In contrast, LSp samples displayed a lower median number of genes that had pathogenic variants at **51** per sample (IQR: 39–68) (Supplementary File 1). Moreover, the median number of non-synonymous somatic mutations per megabase (TMB) in LLS-CRC samples was **83**/Mb, compared to **3.7**/Mb in LSp-CRC samples, reflecting a significant difference in TMB. All LLS samples were hypermutated (defined as >10 non-synonymous somatic mutations per Mb). Interestingly, a hypermutated subset was identified within the LSp-CRC group (33/313; 10.5%), of which 30 samples were MSI-H, with all but one exhibiting *MLH1* hypermethylation. These results indicate that this subset may represent a distinct group of dMMR driven by epigenetic mechanisms, rather than belonging strictly to the LSp subtype. The hypermutated nature and MSI-H status of these tumours suggests potential overlaps in treatment strategies with LLS cases.

Further evaluation of genomic alterations in the CRC subtypes focused on pathogenic variants in tumour suppressor genes (TSG) and oncogenes (OGs). Across all CRC subtypes, and consistent with previous reports, the most commonly altered genes were *APC*, *KRAS* and *TP53* (Appendix A, Figure S2). In LSp-CRC samples, *APC* mutations were observed in 44% of cases. These samples exhibited the lowest transcript levels of *APC*, suggesting reduced *APC* function and its contribution to CRC progression. In LSp-CRC, *APC* mutations typically arise early, initiating tumorigenesis. In contrast, in LLS patients, MMR defects likely drive later-stage *APC* mutations, accelerating the shift from early dysplasia to invasive cancer¹³⁴.

In the LLS samples, the most frequently mutated OGs and TSGs were *APC* (present in nine of 12 samples; 75%), followed by *ATM*, *LRP1B* and *NRF* (each present in eight of 12 samples; 67%) (Figure 5). The observed frequency of *APC* mutations is higher than typically reported in LS-associated CRC, where *APC* mutations are often detected in approximately 20-30% of cases¹³⁴. This discrepancy may reflect cohort-specific factors, such as variations in chromosomal instability or methodological differences in mutation detection. *TP53*

mutations were less common, occurring in two of 12 samples (17%), with no significant difference in *TP53* mutation frequency between the LLS and LSp groups. This finding suggests that *TP53* mutations may be less critical in differentiating between these subtypes.

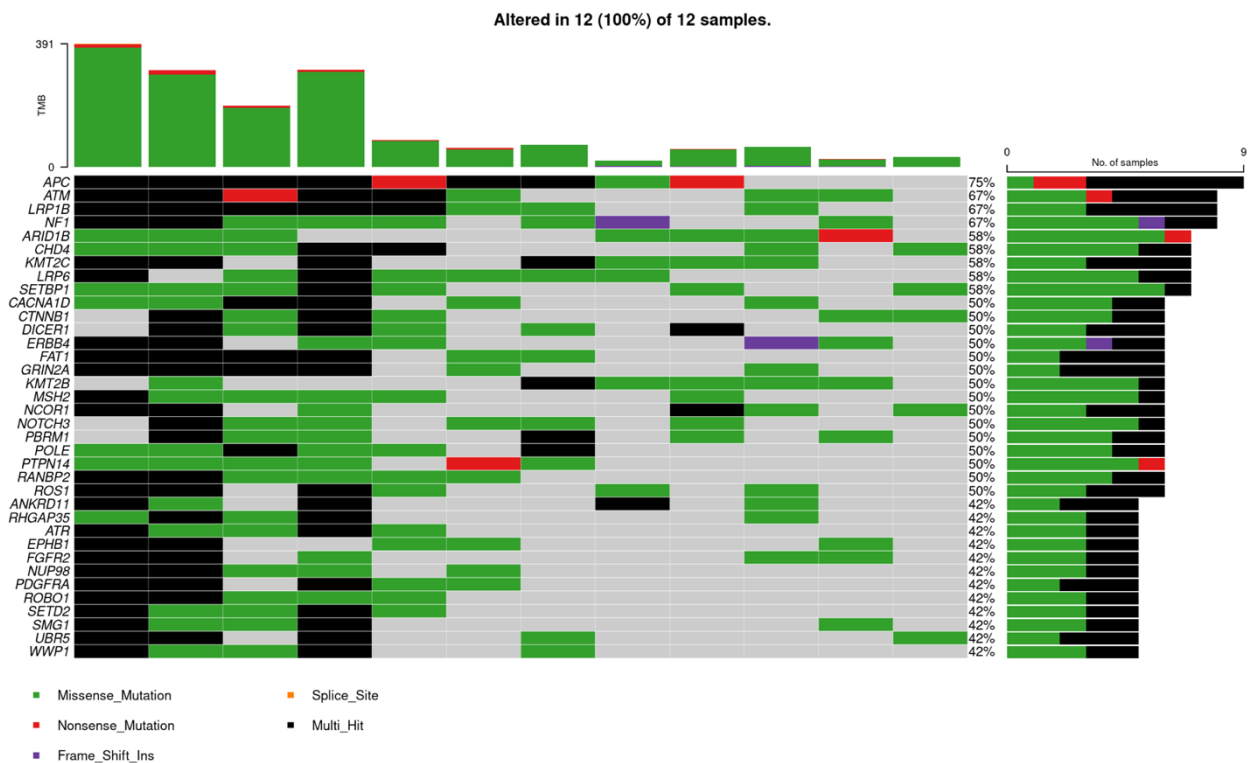
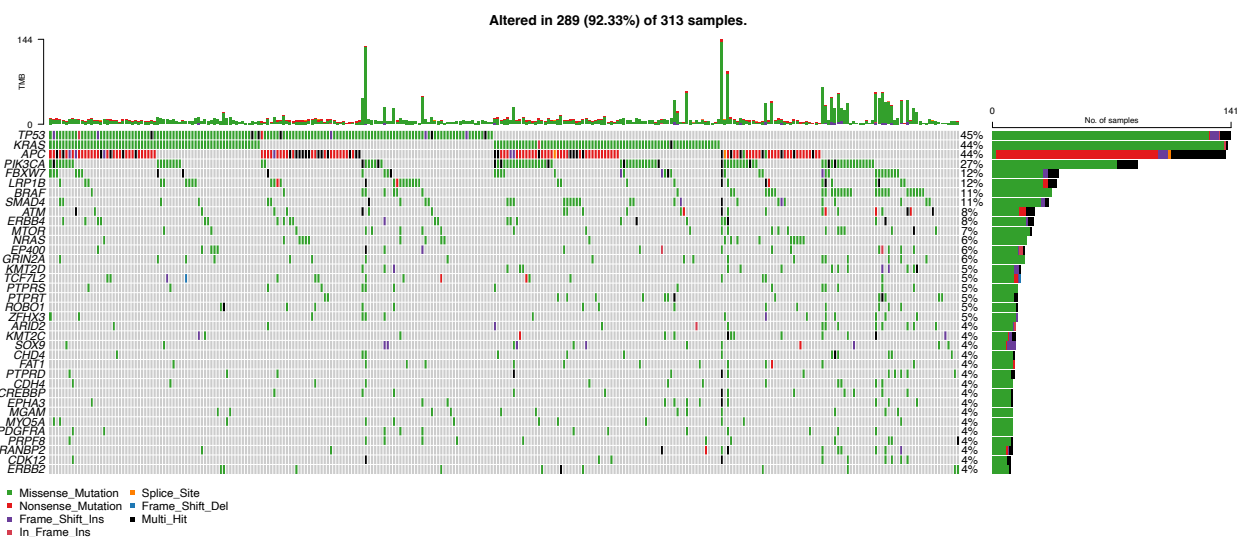


Figure 5. An oncoplot of the top pathogenic mutated tumour suppressor and oncogenes in the likely Lynch Syndrome colorectal cancer samples (n=12).

To further understand the functional consequences of the mutations identified in the oncoplot (Figure 5), the mutation status of the top genes was correlated with their corresponding average gene expression z-scores in LLS samples. Among the most frequently mutated genes, the TSGs, *ATM*, *LRP1B* and *NF1*, exhibited reduced expression levels. In LLS samples, *ATM* had an average expression z-score of -0.21, reflecting mild downregulation. Similarly, *NF1* showed moderate downregulation with an average z-score of -0.26. *LRP1B* exhibited the most pronounced reduction in expression, with a z-score of -0.64, indicating significant transcriptional downregulation.

The reduced expression levels observed in these mutated genes suggest functional consequences consistent with tumour suppressive roles, rather than representing random somatic mutations. Notably, the pronounced downregulation of *LRP1B* in LLS samples may

point to its role in tumorigenesis, warranting further investigation. *LRP1B* is one of the most frequently altered genes in cancer and has been associated higher TMB in *LRP1B*-mutated CRC patients¹³⁵. Given its known role in tumour suppression, *LRP1B* downregulation could significantly contribute to pathogenesis in the context of the mutations studied. Additionally, the role of *NF1* in regulating the RAS/MAPK pathway suggests that even mild downregulation could have functional consequences¹³⁶. Together, these findings highlight the likely contributions of these genes in CRC progression in the context of LLS.



2.3.5 Top Mutated Genes

identified, including 14 involved in the PI3K-Akt signalling pathway ($q = 3.1 \times 10^{-8}$) (Figure 7). This pathway regulates key cellular processes such as growth, proliferation, and survival, and its dysregulation in LLS tumours may suggest a potential role in tumorigenesis¹³⁸.

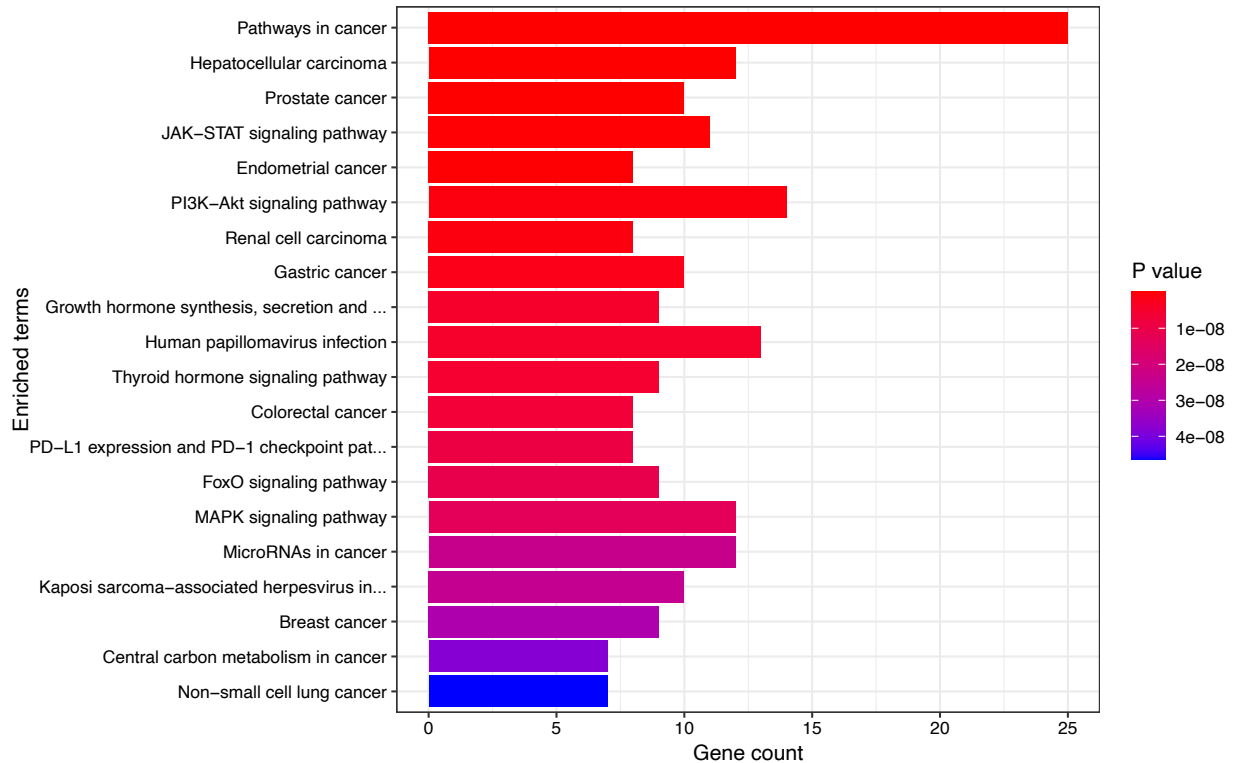


Figure 7. Biological processes and pathways in the Likely Lynch Syndrome Colorectal Cancer subtype (n=12). This figure shows the involvement of the top mutated genes in key signalling pathways.

Figure 8 displays the top 50 mutated genes in LLS samples (blue bars) compared with the LSp samples (orange bars). Among these, genes such as *ATM*, *LRP1B*, and *NF1* show significantly higher mutation frequencies (and reduced transcription) in LLS-CRC, as mentioned in Section 2.2.4. This figure highlights the relative enrichment of these mutant genes in the LLS subtype, emphasising the distinct molecular drivers between the two subtypes. Additionally, eight of these top mutated genes, including *ATR*, *KMT2C*, and *NCOR1*, contain mononucleotide repeat regions. This indicates that the higher mutation frequency in LLS-CRC may be partially driven by the vulnerability of these repeat regions to mutagenesis, further contributing to the molecular distinctions between LLS and LSp subtypes. Further, *ATR* showed reduced transcription, with an expression z-score of -0.54 in LLS samples, indicating a potential loss of function. Interestingly, *NCOR1* exhibited increased transcription, with an expression z-score of 0.87, suggesting a potential gain-of-function or alternative regulatory role in LLS-CRC.

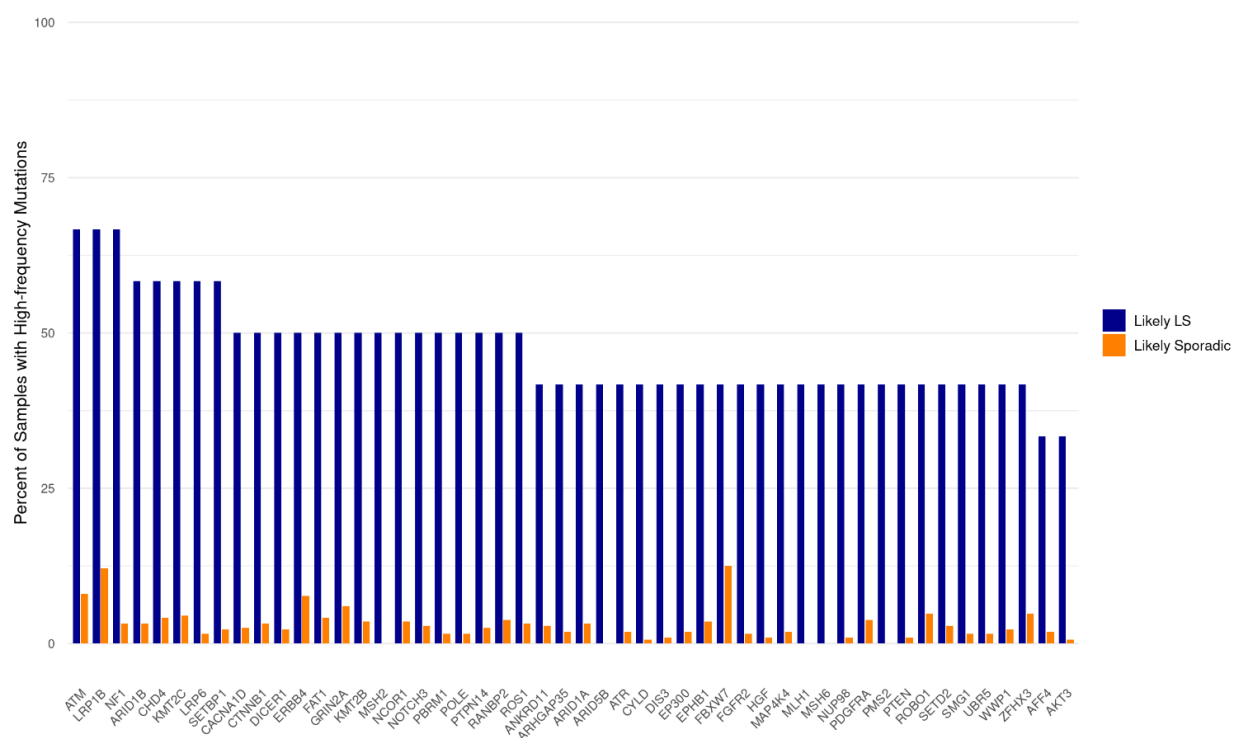


Figure 8. Top 50 mutant genes in colorectal cancer (CRC) subtypes. This figure compares the involvement of the top 50 likely Lynch Syndrome (LLS) CRC associated genes (x-axis) in the LLS samples (n=12, dark blue bars), compared to the Likely Sporadic (LSp) samples (n=313, orange bars).

Overall, pathogenic variants in PI3K-Akt pathway genes were found in 83% of LLS tumours, whereas only 20% of LSp tumours harboured such mutations. These mutations were identified as significantly different based on statistical analysis using Fisher's exact test, with results adjusted for multiple comparisons to ensure robustness (Supplementary File 2). These findings underscore the distinct molecular drivers in LLS-CRC and highlight the PI3K-Akt pathway as a key oncogenic mechanism in this subtype.

2.3.6 Differential Gene Expression Analysis

To investigate the gene expression profiles between the different CRC subtypes, a differential gene expression (DGE) analysis was performed. This analysis identified several genes that were significantly differentially expressed between LLS and LSp-CRC samples (Appendix A, Figure S3A and S3B; Supplementary File 3).

2.3.7 Pathway Enrichment Analysis

To gain insights into the biological processes and pathways associated with the differentially expressed genes, a gene set enrichment analysis (GSEA) was conducted. *Enrichr* was used to compare Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway classifications of genes that were upregulated in tumours from each CRC subtype (Figure 9)^{126,139}. The analysis identified several key pathways consistently altered across CRC subtypes, though to varying degrees (Supplementary File 4).

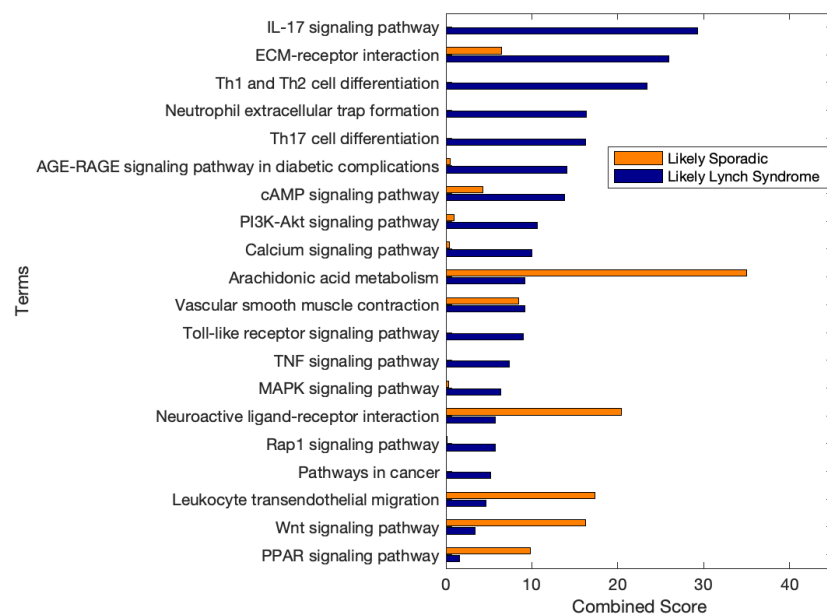


Figure 9. Pathway enrichment analyses: KEGG pathway enrichment results highlight some significant pathways that are differentially altered between the colorectal cancer subtypes. Refer to Supplementary File 4 for a complete list of KEGG pathways enriched based on the mRNA transcript levels.

In LLS-CRC tumours, the interleukin-17 (IL-17), cyclic adenosine monophosphate (cAMP), PI3K-Akt, and calcium-signalling pathways were significantly upregulated (q -values < 0.05). Notably, calcium channel genes such as *Orai1*, *TRPV6*, and *TRPM8* were overexpressed, which are associated with increased calcium influx that promotes cell migration and proliferation, potentially contributing to the enhanced invasiveness observed in LLS-CRC tumours^{140,141}. This finding suggests that targeting calcium signalling may offer therapeutic potential for limiting tumour spread in LLS-CRC. Similarly, the cAMP-signalling pathway, which plays a critical role in regulating cell proliferation, differentiation, and apoptosis, was notably upregulated in LLS-CRC tumours¹⁴². In addition, the IL-17 signalling pathway, showed

significant upregulation in LLS-CRC, suggesting that inflammatory responses may play a role in LLS-CRC.

While it may initially support immune surveillance and neoantigen recognition, chronic IL-17-driven inflammation could tip the balance toward autoimmunity or even promote tumour progression under certain conditions. This upregulation could be either a characteristic feature of LLS-CRC or a factor influencing disease progression in this subtype¹⁴³¹⁴⁴.

Consistent with the earlier mutational analyses, the PI3K-Akt pathway was significantly upregulated in LLS-CRC compared to LSp-CRC tumours (q -value < 0.05). Several OGs involved in angiogenesis and extracellular matrix remodelling, such as *FLT4*, *COL1A1*, *CSF3R*, *NTRK2*, *KIT* and *IL7R*, were also overexpressed in LLS-CRC samples. This suggests a potential link between these processes and tumour progression. However, it is important to note that the overexpression of OGs may have distinct roles compared to TSGs and could contribute differently to tumorigenesis across various cancer subtypes.

Conversely, LSp-CRC tumours exhibited significant upregulation of genes involved in the Wnt signalling pathway (p -value = 0.002) when compared to LLS-CRC. Key Wnt pathway activators, such as *WNT7A* (a ligand that initiates Wnt signalling by binding to Frizzled receptors), *FZD10* (a Frizzled receptor that propagates the Wnt signal), and *RSPO2* and *RSPO4* (potent enhancers of Wnt signalling), showed greater upregulation in LSp-CRC. Despite the presence of upregulated inhibitors such as *RNF43*, *WIF1*, *DKK4*, *NKD1*, and *NKD2*, which act to suppress the pathway, the significant upregulation of these key activators suggests that Wnt signalling remains more active in LSp-CRC compared to LLS-CRC. This differential upregulation highlights the distinct molecular mechanisms driving tumour progression in LSp versus LLS-CRC¹⁴⁵.

Given the control role of the PI3K-Akt pathway, its activation across CRC subtypes was further examined by mapping mRNA expression levels onto the pathway components (Figure 10). This analysis demonstrated consistently higher mRNA expression levels of key genes in the PI3K-Akt pathway in LLS tumours, suggesting a more pronounced activation of this pathway in LLS compared to LSp CRC subtypes. This enhanced pathway activation could be a major

driver of oncogenesis in LLS tumours, highlighting the PI3K-Akt pathway as a promising target for therapeutic intervention.

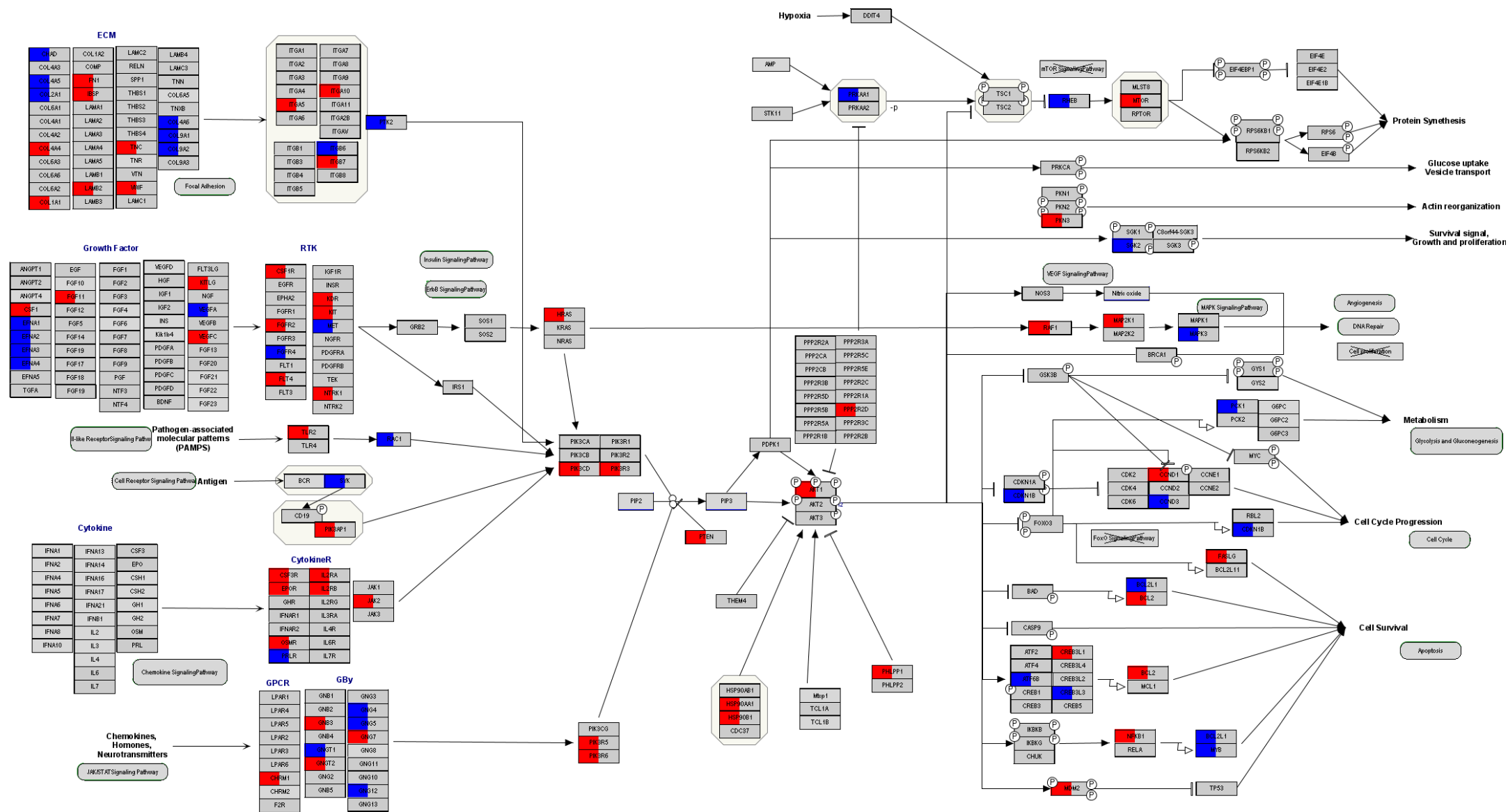


Figure 10. A comparison of PI3K-Akt signalling pathway activity between Likely Lynch Syndrome (LLS) and Likely Sporadic (LSp) Colorectal Cancer (CRC) subtypes based on mRNA expression data. Node denotes proteins – left sections = LLS-CRC and right section = LSp tumours. Nodes are coloured based on overall subtype mRNA-expression z-score (blue = low, grey = no change, and red = high)

2.3.8 Kinase and Transcription Factor Analysis

Many kinases have been implicated in carcinogenesis. To identify the upstream regulatory kinases and transcription factors driving the differential gene expression patterns observed between LLS and LSp-CRC, an *Expression2Kinases* (X2K) analysis was conducted (Figure 11)¹²⁸. This analysis identified several key kinases and transcription factors with strong regulatory roles in these CRC subtypes.

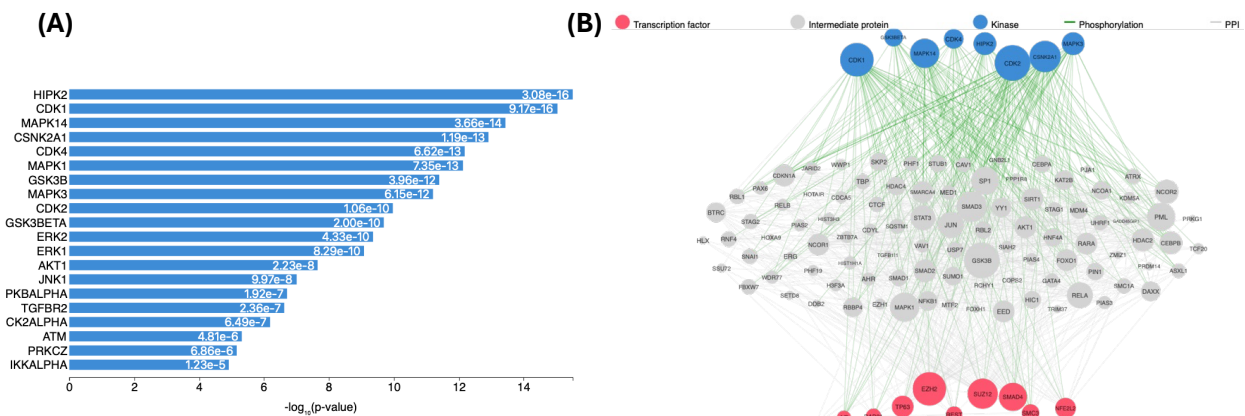


Figure 11. Predicted Regulatory Network and Key Protein Kinases in Likely Lynch Syndrome Colorectal Cancer (LLS-CRC). (A) A ranked list of the top predicted protein kinases likely to regulate the observed gene expression differences. (B) An X2K network illustrating the inferred upstream regulatory network predicted to regulate LLS-CRC, highlighting key kinases and transcription factors. Pink nodes represent the top transcription factors predicted to regulate the expression of LLS-CRC, and blue nodes indicate the top predicted protein kinases known to phosphorylate the proteins within the expanded subnetwork. Green edges represent kinases-substrate phosphorylation interactions, while grey edges depict protein-protein interactions.

Among the top 10 kinases identified, several OGs emerged, including *CDK4*, *MAPK1*, *GSK3B*, *MAPK3* and *GSK3BETA*, all of which are involved in the PI3K-Akt pathway. The overexpression of these kinases suggests hyperactivation of the PI3K-Akt pathway. Additionally, *CDK1*, *CDK2*, and *ATM* were prominent, indicating their involvement in cell cycle regulation and DNA damage response, potentially facilitating the unchecked growth characteristics of LLS-CRC.

In addition to kinases, several transcription factors were identified as potential regulators of gene expression differences between LLS and LSp-CRC. Notably, *SUZ12*, *EZH2*, and *REST* were among the top transcription factors identified, each of which plays an established role in cancer progression (Figure 12)¹⁴⁶. *SUZ12* and *EZH2*, components of the polycomb repressive complex 2 (PRC2), are known to silence TSGs, while *REST* is involved in repressing neuronal

genes and promoting cell survival^{147,148}. The significant overexpression of *SUZ12*, *EZH2* and *REST* in LLS-CRC suggests their involvement in driving tumorigenesis through interactions with the PI3K-Akt pathway.

Interestingly, *SUZ12* is located 40kb telomeric to *NF1*, a TSG found to be frequently mutated in the LLS samples (as shown in Figure 8)¹³⁶. Given the proximity and functional relationship between these two genes, loss of heterozygosity events affecting both *SUZ12* and *NF1* may contribute to tumour progression, This interplay suggests that *SUZ12* and *NF1*, through their cooperative roles, could influence the dysregulated pathways driving LLS-CRC tumorigenesis.

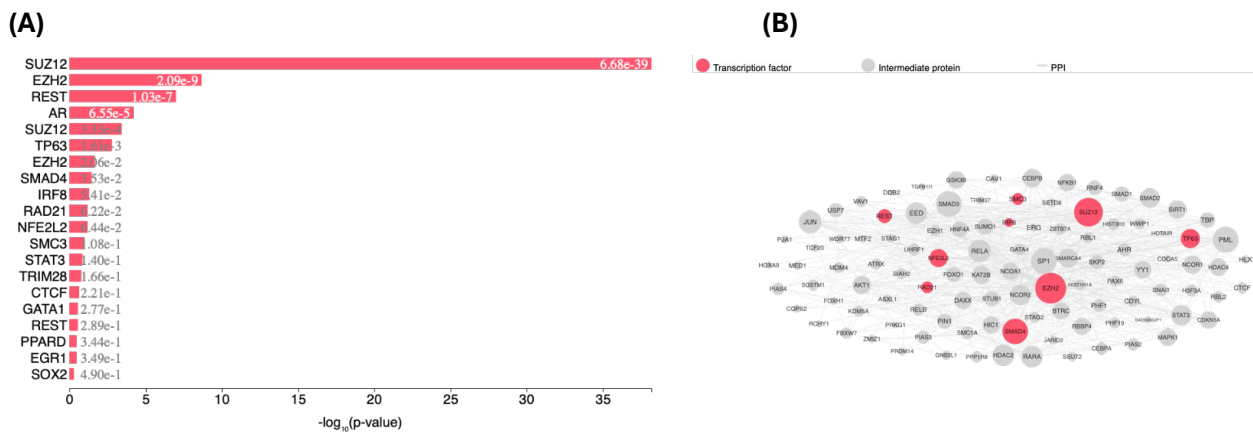


Figure 12. Predicted Transcriptional Regulatory Network in Likely Lynch Syndrome Colorectal Cancer (LLS-CRC). (A) Ranked list of the top predicted transcription factors regulating LLS-CRC, as determined by ChIP-seq experiments (ChEA). (B) Ball-and-stick diagram of the subnetwork of connected transcription factors and their interacting proteins. Pink nodes represent transcription factors, while grey nodes indicate proteins that connect them. Node sizes are proportional to their degree of connectivity.

Collectively, these kinases and transcription factors regulate key molecular networks underlying LLS-CRC and present potential therapeutic targets.

2.3.9 Validation using Known LS Samples

To validate the role of the PI3K-Akt signalling pathway in LS-associated CRC, a DGE analysis was conducted using RNA-seq data from the Gene Expression Omnibus (GEO) series GSE224707¹⁴⁹. This dataset included samples from known LS patients, categorised into three stages: precancer (n=16), advanced precancer (n=7), and cancer (n=9). Samples lacking mRNA data were excluded. The aim was to evaluate the molecular changes across these stages and

determine whether the PI3K-Akt pathway plays a pivotal role in driving the transition from precancerous lesions to invasive cancer.

Consistent with the findings from the TCGA analysis, *MSH2* variants were the most prevalent in the GEO dataset, accounting for 13 (41%) of the 32 MMR variants. Additionally, MSI-H status was observed in 81% of precancer samples, while 57% of advanced precancer samples and 11% of cancer samples were MSS, reflecting a clear shift as lesions progress towards malignancy.

The comparison of the precancerous and advanced precancerous stages revealed 652 upregulated differentially expressed genes (DEGs) in the precancerous stage and 555 in the advanced precancer stage. Pathways enriched in the precancerous stage were largely associated with immune responses, metabolic processes, and cell adhesion (Supplementary File 5). Notably, there was significant activation of NF-kappa B signalling, chemokine signalling, and Th1/Th2/Th17 cell differentiation pathways, consistent with the findings described in Section 2.2.7. These findings suggest that early LS-associated CRC lesions are characterized by an active immune environment, indicating the body's initial attempts to counteract abnormal cell growth. Equally, this immunogenic/hyperinflammatory response may be triggered by chronic neopeptide or neoantigen exposure, which is characteristic of the LS state. In the advanced precancerous stage, there was a sharp decline in pathway enrichment, with only olfactory transduction being significantly upregulated. The reduction of immune-related pathways at this stage suggests a transition from an immune-reactive environment to a more subdued state, potentially reflecting immune evasion, where precancerous lesions evade immune detection and control¹⁵⁰.

A striking increase in the number of upregulated DEGs was observed when comparing advanced precancerous lesions to cancerous tissue. In the advanced precancer stage, 483 genes were upregulated, but enrichment was limited to the olfactory transduction pathway, which is not typically associated with cancer (Supplementary File 5). In contrast, 1,172 upregulated DEGs were identified in cancerous lesions, with significant enrichment in multiple cancer-related pathways, including PI3K-Akt signalling, extracellular matrix (ECM)

receptor interaction, focal adhesion, and proteoglycans in cancer. The activation of the PI3K-Akt signalling pathway exclusively in the cancer stage strongly supports its pivotal role in driving LS-associated CRC progression. Key genes such as *PIK3R5* (a regulatory subunit of PI3K) and *AKT3* (a key kinase in the pathway) were upregulated in cancer compared to advanced precancerous lesions, suggesting increased PI3K-Akt activity as a critical promoter of cancer cell survival and proliferation. Additionally, receptor tyrosine kinases such as *PDGFRA*, *PDGFRB*, and *FGFR1*, which activate PI3K-Akt signalling, were also upregulated, highlighting enhanced signalling through growth factor receptors during cancer progression.

In cancerous lesions, genes associated with the ECM, such as *COL1A1*, *COL1A2*, *COL4A1*, and *ITGA5*, were significantly upregulated, emphasizing the role of the PI3K-Akt pathway in tumour invasion and interaction with the tumour microenvironment. Integrins, including *ITGB5* and *ITGA11*, which mediate cell-ECM interactions and are known to activate PI3K-Akt signalling, were also highly expressed in cancer, further reinforcing the importance of this pathway in tumour growth and likely facilitating metastasis.

2.3.10 Drug Sensitivity and Dependency Analysis

To explore the potential therapeutic implications, the relationship between PI3K-Akt pathway dependency and drug sensitivity was examined using the Genomics of Drug Sensitivity in Cancer (GDSC) database¹³⁰. Dose-response data for 344 cancer cell lines treated with 51 PI3K-Akt pathway inhibitors were analysed.

For each PI3K-Akt pathway gene, cancer cell lines were categorised into high or low-dependency groups based on their CRISPR-derived dependency scores (Supplementary File 6). Sensitivity to PI3K-Akt pathway inhibitors varied according to cell line dependency on specific PI3K-Akt pathway genes. Specifically, high dependency scores of 45 genes were linked to significantly increased sensitivity to these inhibitors, while low dependency scores of 34 genes were associated with decreased sensitivity. Interestingly, 35 other genes exhibited mixed responses; i.e. significantly increased sensitivity to some of the inhibitors and significantly decreased sensitivity to others, indicating complex gene-drug interactions. Cell lines with higher dependency on genes such as *EIF4E2*, *GNG5*, and *MCL1* showed significantly

increased sensitivity to 67% of the PI3K-Akt inhibitors. Conversely, dependence on *RXRA* and *KRAS* correlated with reduced sensitivity to 38% and 33% of inhibitors, respectively.

Further, grouping cell lines based on the presence of mutations in the PI3K-Akt pathway gene revealed that mutation status alone was not a reliable predictor of drug sensitivity (Supplementary File 6). Only mutations in *PKN1* were associated with significantly increased sensitivity, suggesting that functional dependency, as measured by CRISPR score, is a more accurate indicator of drug response than mutation status.

Overall, 926 instances were identified where high dependency or mutations in PI3K-Akt pathway genes significantly influenced drug sensitivity (Figure 13A and 13B). High dependency on PI3K-Akt pathway genes was associated with increased sensitivity to drugs such as pictilisib, capivasertib and apitolisib in 636 instances and resistance in 289 instances. In contrast, mutation status influenced sensitivity in only one instance to uprosertib, emphasising the greater predictive power of functional dependency in determining drug response. These findings suggest that targeting specific gene dependencies could optimise treatment strategies for LS-CRC patients.

Cell lines were further stratified into two groups (higher and lower CRISPR-derived dependency) based on their overall PI3K-Akt pathway dependency and compared drug IC50 values between these groups for the 51 inhibitors. Cell lines with higher dependency were significantly more sensitive to all PI3K-Akt inhibitors, with taselisib ($t = 8.5$, $p = 1.4 \times 10^{-15}$), pictilisib ($t = 8.1$, $p = 2.2 \times 10^{-15}$), and alpelisib ($t = 7.6$, $p = 4.7 \times 10^{-13}$) showing the most significant differences in dose-responses between the groups (Figure 13C). These findings suggest that these inhibitors could be particularly effective in treating PI3K-Akt-dependent LLS cancers and may inform future patient stratification strategies based on pathway dependency.

To gain further insights into the mechanisms underlying drug sensitivity, mRNA expression profiles from the LINCS dataset were analysed, focusing on transcriptional changes in

response to PI3K-Akt pathway inhibitors in four cell lines (MCF7, A549, A375, and YAPC). Despite differences in drug sensitivity, transcriptional responses were consistent across cell lines, indicating that while gene expression changes may be uniform, the degree of pathway dependency critically determines drug efficacy. Therefore, targeting these dependencies with precision therapies could optimise outcomes in cancers characterised by high PI3K-Akt pathway activity.

These findings have important therapeutic implications for LS-associated CRC. The high PI3K-Akt pathway activity suggests that inhibitors like taselisib, pictilisib, and alpelisib could provide significant clinical benefits, particularly for patients whose tumours show high dependency on this pathway.

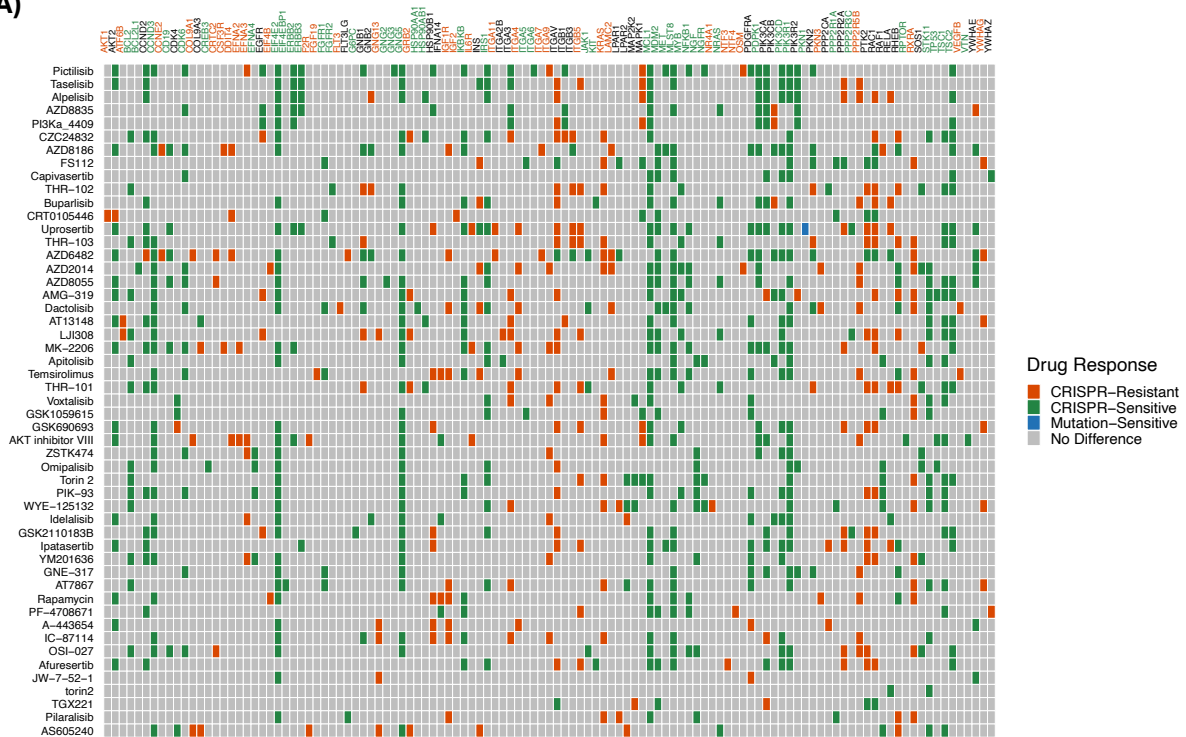
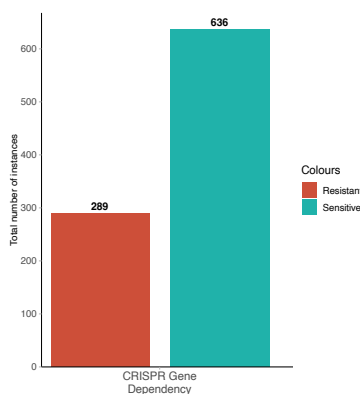
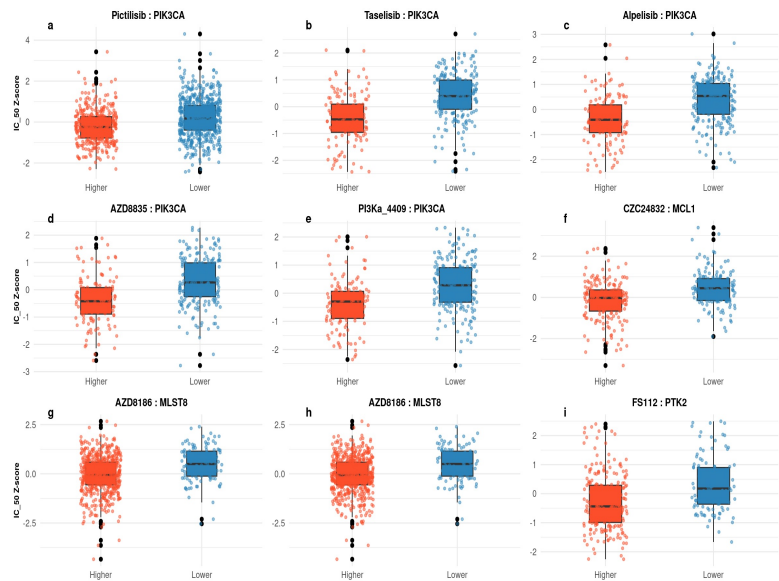
(A)**(B)****(C)**

Figure 13. Relationship between Achille gene dependence scores or mutations of PI3K-Akt pathway genes and the Responses of the cell line to PI3K-Akt pathway inhibitors. (A) A clustered heatmap showing the relationship between gene dependencies or mutations and drug responses across cancer cell lines. The marks on the heatmap are coloured based on how a high dependence on, or mutations in, the gene along each column affect the efficacy of the drug given along each row: with orange for gene dependence denoting significantly increased resistance, green (for gene dependence) and blue (for gene mutations) denoting significantly (10% false discovery rate) increased sensitivity, grey for no statistically significant difference between cell line with a higher and lower dependence on the gene. The gene names (column labels) are coloured based on the overall calculated effect that high dependence on the gene has on the efficacy of the drug given along rows. Green, all the cell lines are significantly more sensitive to all the PI3K-Akt pathway inhibitors, orange, all the cell lines are significantly more resistant to all the PI3K-Akt pathway inhibitors and black is a mixed response to PI3K-Akt pathway inhibitors. **(B)** A bar graph showing the total number of instances for the resistant (red) and sensitive (blue) CRISPR gene dependencies **(C)** Comparison of the dose-response profiles between the higher and lower dependency groups of the GDSC cancer cell lines for the top nine inhibitors. The boxplots show the logarithm-

transformed mean IC50 values of the cancer cell lines. On each box, the central black line indicates the median, and the bottom edge represents the 25th percentiles, while the top edge of the box represents the 75th percentiles. The whiskers extend to the most extreme data points not considering outliers, which are plotted individually in bold, black circles.

Given the observed correlation between cell line responses to PI3K-Akt pathway inhibitors and their dependency on specific PI3K-Akt pathway genes, it was hypothesised that transcriptional changes induced by these inhibitors would mirror the underlying cellular adaptations. To investigate this, the LINCS dataset was utilised, which provides comprehensive mRNA transcription profiles for approximately 1,000 genes in various cell lines after exposure to seven distinct PI3K-Akt pathway inhibitors. This analysis focused on four cell lines (MCF7, A549, A375, and YAPC) and two PI3K-Akt inhibitors, alpelisib and OSI-027, both of which were common across the LINCS, GDSC, and Cancer Cell Line Encyclopaedia (CCLE) datasets¹⁵¹. Alpelisib specifically targets *PIK3CA*, a catalytic subunit of PI3K, while OSI-027 acts as a dual inhibitor of *mTORC1* and *mTORC2*, critical downstream components of the PI3K-Akt pathway. By comparing the transcriptional responses across these cell lines, we aimed to elucidate how differential gene expression patterns reflect the mechanistic efforts of pathway inhibition and the cellular dependencies that drive sensitivity or resistance to these drugs.

Dose-response profiles were obtained from the GDSC demonstrating varying degrees of sensitivity to PI3K-Akt pathway inhibitors across the tested cell lines. MCF7 was the most sensitive to alpelisib (IC50 = 9.5 μ m), followed by A549 (IC50 = 16.2 μ m), YAPC (IC50 = 30.6 μ m), and A375 (IC50 = 30.6 μ m). Interestingly, all cell lines exhibited heightened sensitivity to dactolisib, with A549 showing the greatest sensitivity (IC50 = 0.01 μ m) and YAPC the least (IC50 = 0.42 μ m). Sensitivity to OSI-027 was generally lower across all lines, with YAPC being the most responsive (IC50 = 60.2 μ m) and MCF7 showing the least sensitivity (IC50 = 241 μ m). The analysis also revealed that A549 and MCF7 (IC50 = 11.0 μ m) were more sensitive to afuresertib compared to YAPC (IC50 = 149 μ m), suggesting differential responses to inhibitors across the cell lines.

Transcriptional responses to alpelisib and OSI-027 were strongly positively correlated across the cell lines, with correlation coefficients of $R = 0.77$ for MCF7, $R = 0.85$ for A549, $R = 0.77$ for A375, and $R = 0.98$ for YAPC ($p < 1 \times 10^{-300}$ for all comparisons; Figure 14). This strong correlation suggests consistent transcriptional shifts in response to these inhibitors, despite

differences in drug sensitivity, indicating that while gene expression changes are uniform, the degree of dependency on the PI3K-Akt pathway may drive the observed variation in sensitivity.

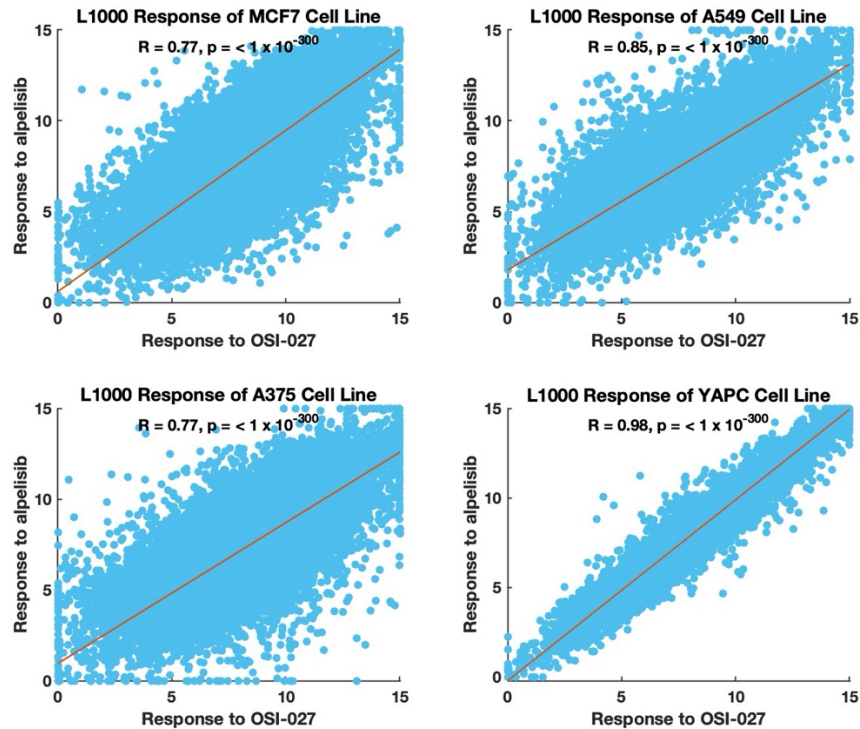


Figure 14. L100 Responses of the Four Cell Lines to Inhibitors Alpelisib and OSI-027.

Building upon the identification of PI3K-Akt pathway dependencies in LLS-CRC, the analysis was expanded to investigate specific vulnerabilities within the Wnt signalling pathway in LSp-CRC. Recognising the role of this pathway in the progression of LSp-CRC, we aimed to assess the effectiveness of targeted inhibitors as potential therapeutic interventions. Dose-response data for eight Wnt pathway inhibitors across multiple cancer cell lines was analysed, stratified by their CRISPR-derived dependency scores for Wnt pathway genes (Supplementary File 7). The results revealed varied impacts of Wnt pathway gene dependency on drug sensitivity. Notably, 16 genes were associated with increased sensitivity to Wnt pathway inhibitors, while 25 genes were correlated with decreased sensitivity (Appendix A, Figure S4A). Among these, cell lines with high dependency on genes such as *GNG5* and *HIST1H2AJ* exhibited increased sensitivity to 71% of Wnt inhibitors, indicating that targeting these dependencies could offer therapeutic benefits. On the other hand, cell lines with high dependency

on *RAC1* demonstrated reduced sensitivity to 50% of inhibitors, suggesting potential resistance mechanisms.

Mutation status alone proved to be a less reliable predictor of drug response, with only *VANGL2* mutations showing significant associations with sensitivity to treatment. This highlights the possibility that functional dependency, rather than genetic mutation status, may serve as a more accurate indicator of drug efficacy in sporadic CRC. The analysis identified 78 cases where either high dependence or mutations in Wnt pathway genes had a significant impact on drug sensitivity. Specifically, cell lines with greater dependency on Wnt pathway genes demonstrated heightened sensitivity in 30 cases and resistance in 47 cases (Appendix A, Figure S4B). These results further underscore the crucial role of pathway dependency in influencing drug responses. These insights point to the therapeutic potential of targeting the Wnt pathway in sporadic CRC, especially with inhibitors such as XAV939 and AZD7969, which may prove beneficial for patients with Wnt-dependent tumours.

Cell lines were further stratified into two groups (higher and lower CRISPR-derived dependency) based on their overall Wnt pathway dependency and compared drug IC50 values between these groups for the 51 inhibitors. Cell lines with higher dependency were significantly more sensitive to all Wnt inhibitors, with IWP-2 ($t = 4.2$, $p = 4.8 \times 10^{-05}$), LGK974 ($t = 3.5$, $p = 7.1 \times 10^{-04}$), and WIKI4 ($t = 4.9$, $p = 8.2 \times 10^{-04}$) showing the most significant differences in dose-responses between the groups (Appendix A; Figure S4C). These findings suggest that these inhibitors could be particularly effective in treating Wnt dependent LSp cancers and may inform future patient stratification strategies based on pathway dependency.

2.4 Discussion

This study significantly advances the understanding of the distinct molecular mechanisms driving LS-associated and sporadic CRC by identifying key dysregulated pathways, particularly the PI3K-Akt signalling pathway in LS-CRC and the Wnt signalling pathway in sporadic CRC. These findings not only elucidate the complex biological processes that underlie these two CRC subtypes but also open up promising therapeutic opportunities. By characterising the

molecular heterogeneity between LS and sporadic CRC, this work lays the foundation for the development of more personalised treatment strategies tailored to the unique genetic and molecular profiles of each subtype.

One of the key findings of this study is the identification of aberrant PI3K-Akt signalling as a major oncogenic driver of LS-CRC. The validation using RNA-seq data from the GEO series GSE22470726 further supports this conclusion. While immune-related pathways dominated in the early, precancerous stages, the exclusive activation of the PI3K-Akt pathway in the cancer stage suggests it plays a critical role in the transition from precancerous lesions to (invasive) malignancy. The PI3K-Akt pathway plays a crucial role in regulating cellular functions such as metabolism, proliferation, survival, and growth, and its hyperactivation in LS-CRC suggests a potential oncogenic mechanism specific to this subtype. Notably, this hyperactivation also correlates with the high TMB observed in dMMR tumours. Key mutant genes associated with this pathway, *FLT4*, *CSF3R*, *KIT*, and *IL7R*, were found to be significantly overexpressed, further implicating PI3K-Akt signalling in the pathogenesis of LS-CRC. Given these findings, PI3K-Akt pathway inhibitors, including tasisib, pictilisib, and alpelisib, emerge as promising therapeutic options. The dose-response data, retrieved from the GDSC database, shows that cell lines highly dependent on the PI3K-Akt pathway exhibit increased sensitivity to these inhibitors, supporting the clinical potential of targeting this pathway in LS-CRC patients.

In contrast, LSp-CRC was primarily driven by dysregulation in the Wnt signalling pathway. Mutations in the *APC* gene, a well-established key driver of colorectal tumorigenesis, played a central role. These findings revealed a significant enrichment of mutations in other Wnt pathway components, such as *RNF43* and *WIF1*, both of which function as tumour suppressors involved in regulating Wnt signalling. The pivotal role of the Wnt pathway in LSp-CRC is further emphasised by its influence on cellular differentiation, migration, and proliferation. This study confirmed the enrichment of mutations in Wnt pathway components in LSp-CRC, further underscoring its importance in this subtype. Targeted inhibitors of the Wnt pathway, such as XAV939 and AZD7969, have shown potential as therapeutic options for Wnt-dependent cancers. The analysis revealed that cell lines with a high dependency on Wnt

pathway genes exhibited varying sensitivities to these inhibitors. In particular, cell lines with mutations in *RNF43* and *WIF1* demonstrated increased sensitivity to these drugs, highlighting the possibility of precision-targeted therapies for LSp-CRC patients.

Additionally, this study identified a hypermutated subset within the LSp-CRC cohort, accounting for approximately 10.5% of the samples, characterized by high levels of MSI. This hypermutated phenotype, primarily driven by epigenetic alterations such as *MLH1* hypermethylation, displays a high TMB that superficially resembles the molecular profile seen in LS-CRC. However, while these similarities may suggest a potential overlap in terms of characteristics between hypermutated LSp and LLS-CRC, the analysis of signalling pathways reveals distinct differences. Specifically, the PI3K-Akt signalling pathway was still significantly more upregulated in LLS-CRC compared to these hypermutated LSp-CRC samples. This suggests that despite a similar TMB, the underlying biological pathways are divergent, indicating that LS-CRC may not be as closely related to hypermutated sporadic cases as TMB levels alone might imply. This distinction highlights the importance of exploring targeted therapies that address specific pathway dysregulations, offering more tailored treatment options for CRC patients with these distinct molecular features.

These findings also emphasise the significance of tumour location in the molecular characterisation of CRC subtypes. LLS-CRC tumours were more commonly located on the right side of the colon, while LSp-CR tumours occurred predominantly on the left. Although this difference was not statistically significant, it aligns with existing literature suggesting distinct molecular and clinical profiles based on tumour location. Right-sided tumours, particularly those associated with MSI-H, are generally associated with better prognoses and may respond differently to specific therapies. These observations further underscore the importance of personalised treatment strategies informed by molecular profiling.

The differential gene expression analysis revealed several critical transcription factors and kinases that may serve as regulatory drivers in the distinct molecular landscapes of LLS-CRC and LSp-CRC. In LLS-CRC, the overexpression of transcription factors such as *SUZ12*, *EZH2*, and *REST* suggests their involvement in the hyperactivation of the PI3K-Akt pathway.

Furthermore, the X2K analysis identified key kinases including *CDK4*, *MAPK1*, and *GSK3B*, which are likely to regulate the gene expression patterns observed in LLS-CRC, underscoring the complexity of the regulatory networks driving oncogenesis in this subtype¹⁵². These findings point to potential combinatorial treatment strategies, such as pairing PI3K-Akt inhibitors with drugs targeting *CDK4* or *GSK3B*. Such combinations could enhance therapeutic efficacy in LLS-CRC by simultaneously disrupting multiple key nodes within the same signalling network. This approach is in line with the emerging trend towards combination therapies in oncology, designed to overcome resistance mechanisms and improve treatment outcomes.

The immune pathways enriched in the precancerous stage, such as NF-kappa B signalling and Th1/Th2/Th17 differentiation, suggest that inflammation plays a pivotal role early in disease development or containment. However, by the advanced precancerous stage, these immune responses diminish, potentially allowing cells to escape immune surveillance. This reduction in immune activity may mark a critical point where lesions are no longer effectively controlled by the immune system, facilitating their progression toward malignancy¹⁵³. Moreover, the significant involvement of ECM-receptor interaction and focal adhesion pathways in the cancer stage highlights the importance of tumour-stroma interactions in the invasive and metastatic potential of LS-associated CRC. This finding is consistent with the "seed and soil" hypothesis of cancer metastasis, where cancer cells (the seed) require a supportive tumour microenvironment (the soil) to invade and metastasize effectively¹⁵⁴.

The investigation into drug sensitivity and pathway dependency highlighted the importance of functional dependency, rather than solely mutation status, in predicting drug response. For instance, while mutation status in Wnt pathway genes was not a reliable predictor of sensitivity to Wnt pathway inhibitors, functional dependency scores derived from CRISPR knockout screens were far more predictive. Cell lines with higher dependency on Wnt pathway genes exhibited increased sensitivity to Wnt inhibitors in some cases, while in others, high dependency was associated with resistance. This finding underscores the complexity of tumour biology and the need to move beyond simple mutation profiling to more nuanced assessments of pathway dependencies when designing treatment strategies. It also

emphasises the importance of access to comprehensive WES, whole genome sequencing and transcriptomic sequencing data for studies of this nature.

However, the study has several limitations that should be acknowledged. First, the reliance on publicly available datasets, while providing a wealth of genomic and transcriptomic information, may introduce biases related to sample selection. Moreover, the relatively small number of LLS samples limits the generalisability of these findings. Validation in larger and more diverse cohorts is necessary to confirm the clinical relevance of the PI3K-Akt pathway in LS-CRC. Additionally, while key dysregulated pathways were identified, further experimental work is needed to elucidate the mechanistic roles these pathways play in CRC development.

Another limitation lies in the approach used to classify LLS cases. Samples were classified based on the presence of pathogenic variants in MMR genes and an age threshold of 60 years for diagnosis. While this approach leverages available data, it does not incorporate MSI status or germline testing, which are standard diagnostic criteria for LS. This may have resulted in the inclusion of cases with somatic MMR variants or epigenetic silencing of MMR genes, which exhibit hallmark LS features but are not hereditary. While steps were taken to address this, such as excluding specific cases with *MLH1* hypermethylation, the possibility of misclassification cannot be entirely ruled out and may have influenced the observed pathway dysregulation.

Overall, the results of this study highlight the distinct molecular landscapes of LS-CRC and LSp-CRC and suggest a range of potential therapeutic strategies tailored to the unique characteristics of each subtype. The identification of critical pathways, such as PI3K-Akt in LS-CRC and Wnt in LSp-CRC, provides a strong rationale for the development of targeted therapies. Furthermore, the findings underscore the potential benefits of combination therapies that target both primary oncogenic drivers and key regulatory elements within these pathways. Moving forward, studies should focus on validating the results in independent datasets and exploring the functional consequences of these dysregulated pathways in *in vitro* and *in vivo* models. Investigating the role of PI3K-Akt inhibitors in LS-CRC

cell lines could provide direct evidence for the therapeutic potential of targeting this pathway. Furthermore, exploring the interplay between other pathways, such as immune-related pathways, could uncover additional therapeutic targets.

2.5 Conclusion

This integrative multi-omic analysis has illuminated distinct molecular drivers of oncogenesis in LS-CRC and LSp-CRC. Specifically, the PI3K-Akt signalling pathway emerged as a key dysregulated pathway in LS-CRC, while Wnt signalling was predominantly implicated in LSp-CRC. These findings underscore the critical role of understanding the biological basis of disease with regards to emerging targeted therapies that address the unique molecular profiles of each CRC subtype. Moreover, the study highlights the therapeutic potential of PI3K-Akt pathway inhibitors, such as taselisib, pictilisib, and alpelisib, particularly for LS-CRC patients with high dependency on this pathway. Similarly, targeting the Wnt signalling pathway in Wnt-dependent LSp-CRC tumours, using inhibitors like XAV939 and AZD7969, presents a promising therapeutic avenue.

This study also underscores the importance of precision medicine in CRC treatment. By integrating functional dependency data with pathway analyses, we can better guide therapeutic decisions and optimize patient outcomes. The ability to tailor treatment strategies to the molecular characteristics of CRC subtypes has the potential to significantly improve prognosis and treatment efficacy for CRC patients.

Chapter 3: Multiomic Analysis of Endometrial Lynch Syndrome-Associated Cancer and Investigating the Inclusion of Breast Cancer in the Lynch Syndrome Spectrum

3.1 Introduction

EC is the most common extracolonic sentinel cancer in LS, with women having an approximately 40%-60% lifetime risk of developing EC as their first malignancy⁶. Key clinicopathologic features of LS-associated EC include early age of onset, lower body mass index, endometrioid histology and frequent involvement of the lower uterine segment⁶. A study by Marques-de-Sá et al. reported that 48% of individuals with LS-associated EC harboured *MSH2* variants, while 32%, 15%, and 5% carried variants in *MLH1*, *MSH6* and *PMS2*, respectively¹⁵⁵. Although LS is primarily linked to CRC and EC, the inclusion of BC in the LS spectrum remains controversial and insufficiently understood.

Several studies have observed an increased incidence of BC among individuals with LS variants, though the molecular mechanisms connecting BS to LS remain poorly defined^{93,156}. Given that LS is driven by MMR gene defects, it is uncertain whether BC in LS patients shares molecular characteristics with other LS-associated cancers, such as EC. Investigating potential overlaps in molecular pathways between these cancers could enhance our understanding of LS-associated tumorigenesis and broaden the recognised cancer spectrum of LS

Recent advances in multiomic profiling, which integrates genomic, transcriptomic, and metabolic data, presents a valuable opportunity to uncover the molecular pathways driving cancer development in LS. Multiomic approaches have already proven valuable in identifying dysregulated pathways in LS-associated CRC, particularly highlighting the PI3K-Akt signalling pathway as a critical oncogenic driver, as seen in Chapter 2 of this dissertation. However, the application of such integrative analyses to LS-associated EC and BCs has been limited, leaving gaps in our understanding of their molecular underpinnings.

In this study, a comprehensive multiomic analysis of publicly available genomic and mutational datasets was conducted to elucidate the progression model in LS-associated EC. Further, the analysis explores whether BC in LS patients exhibits similar or distinct molecular

characteristics compared to LS-associated EC, thus assessing the potential inclusion of BC within the LS cancer spectrum. This study aims to identify shared and/or cancer-specific dysregulated pathways that could help refine the molecular scope of LS.

3.2 Materials and methods

3.2.1 Data Acquisition

Data from the Pan Cancer Atlas project of TCGA was accessed via the cBioPortal platform. This dataset included clinical data, WES data, and transcriptomic data, obtained through RNA-seq. Due to sample availability and assay failures, not all data types were available for all patients. All mutations analysed in this study are somatic, identified from tumour samples. No germline data is available for these patients in the TCGA.

3.2.2 Classification of LLS and LSp Samples

Samples were classified based on the presence of pathogenic variants in MMR genes. Pathogenicity was assessed using the *Franklin* online tool. Samples harbouring at least one pathogenic MMR gene variant were categorised as LLS, while those without such variants were classified as LSp. Notably, none of the LLS-BC samples harboured *BRCA* mutations. For the EC dataset, both genetic and age criteria were employed. Samples were classified as LLS if they contained a pathogenic MMR variant and if the patient's age at diagnosis was 60 years or younger. Samples were classified as LSp if they lacked pathogenic MMR variants and the patient's age at diagnosis was over 60 years. This age criterion aligns with recommendations for identifying LS in EC patients¹⁵⁷. In contrast, no such age-based guidelines exist for BC, as its association with LS remains inconclusive.

3.2.3 DGE and Pathway Enrichment Analysis

To investigate DGE between the EC and BC subtypes, RNA-seq data was utilised. In MATLAB, differential expression was tested between the subtypes using a negative binomial model, with local regression applied to adjust for variance between the groups. Mean expression levels for each gene in both subtypes were calculated, and the log2 fold changes were computed to quantify the relative differences in gene expression. *p*-Values from the differential expression analysis were adjusted for multiple comparisons using the Benjamini-

Hochberg FDR method, with genes showing an adjusted p -value (q -value) of less than 0.05 considered differentially expressed. GSEA was performed to identify pathways enriched among the DEGs, using the KEGG database as a reference¹³⁹. Enrichment analysis was conducted through the *Enrichr* platform, and pathways with q -values < 0.05 were deemed significantly enriched¹²⁶.

3.2.4 Predication of Regulatory Kinases and TFs

Upstream regulatory kinases and TFs likely influencing the observed differences in gene expression between the subtypes were computationally predicted using X2K¹²⁸. X2K employs a reverse-engineering, network-based approach to predict upstream regulatory kinases based on existing knowledge.

3.2.5 Mutation Analyses

The spectrum of genomic mutations across EC and BC subtypes were evaluated by focusing on OGs and TSGs. A comprehensive list of OGs and TSGs was obtained from OncoKB, comprising 740 OGs and TSGs¹²⁹. This list was used to extract mutations expected to have an impact on EC and BC oncogenesis. Specifically, mutations in known OGs and TSGs were filtered for. Using the processed data, frequently altered genes within each subtype were identified. The frequency of mutated genes was calculated for both subtypes for each cancer type, and Fisher's exact test was applied to identify statistically significant differences. To control for multiple comparisons, p -values were adjusted using Benjamini-Hochberg correction. All significant genes were found to be enriched in the LLS-EC and LLS-BC subtypes.

3.2.6 Statistical Analyses

Drug sensitivity and dependency analysis was performed using data from the GDSC database. Dose-response data for cancer cell lines treated with TGF-beta-related pathway inhibitors (and cAMP-related pathway inhibitors for LSp subtypes) were retrieved. Cell lines were characterised into high- and low-dependency groups based on CRISPR-derived dependency scores for TGF-beta-related pathway genes. To assess drug efficacy, the mean IC50 values for each inhibitor between the high- and low-dependency groups were compared. The relationship between gene dependencies and drug sensitivity was further evaluated using a

linear regression model. Significant differences in drug sensitivity between the groups were identified using a Student's t-test, with *p*-values adjusted for multiple comparisons.

For each pathway gene, the cell lines were stratified into two groups: those with a high CRISPR-determined dependence on the gene (e.g., a dependence score < -0.5) and those with a low dependence on that gene (e.g., dependence score > 0.5). The IC₅₀ values for each of the TGF-beta-related pathway inhibitors were then compared between the two groups. Additionally, the cell lines were classified based on the mutation status of specific TGF-beta-related pathway genes. For example, cell lines with a mutation in a particular gene were compared to those without, and IC₅₀ values for each of the TGF-beta-related pathway inhibitors were analysed between mutant and non-mutant groups.

Next, the number of TGF-beta-related pathway genes that were either “common essential” or “strongly selective” across each cell line were quantified. This yielded an absolute count of TGF-beta-related pathway genes that each cell line is most dependent on. It was hypothesised that cancer cell lines with greater dependency on TGF-beta-related pathway genes would show enhanced sensitivity to TGF-beta-related pathway inhibitors. To test this, the cell lines were divided into two categories: (1) those with more than the median number of “common essential” or “strongly selective” TGF-beta-related pathway genes (indicating higher TGF-beta-related pathway dependency), and (2) those with fewer than the median number of such genes (indicating lower TGF-beta-related pathway dependency). The mean IC₅₀ values for the TGF-beta-related pathway inhibitors were then compared between these two groups.

3.2.7 Statistical Analyses

All statistical analyses were performed using R Studio, unless stated otherwise. Docker was used to encapsulate the computational environment and ensure the reproducibility of all statistical analyses. The 'animates/anima:4.2.2' Docker image was employed to create a container for RStudio, facilitating consistency in data processing and analysis. Continuous variables were summarised using the median and interquartile range, with comparisons made using Wilcoxon rank-sum test. Categorical variables were summarised using count and percentage, with comparisons made using Fisher's exact test. All statistical tests were two-

sided, and p -values were corrected for multiple comparisons according to the Benjamini-Hochberg method. Results with $p < 0.05$ or $q < 0.05$ (adjusted p -value) considered significant.

3.2.8 Data, Code and Supplementary Files Availability

All datasets used in this study were obtained from publicly accessible sources. Detailed instructions for accessing these datasets are available via the respective platforms (Appendix C, Table C1). The code used for all analyses is publicly available on GitHub (<https://github.com/may-julia/Analysis-of-Endometrial-LS-Associated-Cancer-and-Investigating-the-Inclusion-of-BC-in-LS-Spectrum-.git>). Supplementary files can be found at (<https://doi.org/10.5281/zenodo.14228072>). The supplementary file descriptions are provided in appendix B.

3.3 Results

3.3.1 Dataset Assembly and Initial Classification

Comprehensive datasets for both EC and BC were compiled using data from the TCGA Pan-Cancer Atlas project, accessed via cBioPortal^{115,116}. Each dataset included clinical information, RNA-seq transcription data, and genomic mutational data derived from WES. The EC dataset comprised 292 samples, while the BC dataset contained 1,007 samples.

To differentiate LLS from LSp cases within the BC dataset, samples were classified based on the presence of pathogenic variants in MMR genes. Samples harbouring at least one pathogenic MMR gene variant were categorised as LLS, while those without such variants were classified as LSp. Notably, none of the LLS-BC samples harboured *BRCA* mutations. For the EC dataset, both genetic and age criteria were applied. Samples were classified as LLS if they contained a pathogenic MMR variant and if the patient's age at diagnosis was 60 years or younger. Samples were classified as LSp if they lacked pathogenic MMR variants and the patient's age at diagnosis was over 60 years. This age criterion aligns with recommendations for identifying LS in EC patients¹⁵⁷. In contrast, no such age-based guidelines exist for BC, as its association with LS remains inconclusive.

Applying these criteria, 14 LLS and 993 LSp cases were identified in the BC dataset. In the EC dataset, 13 samples were classified as LLS and 279 as LSp. The classification approach was dictated by the nature of the datasets used, which did not include IHC results or consistent MSI testing data. Starting with MMR gene variants allowed the study to utilise the available genomic information effectively and focus on molecularly define cases for downstream analyses. Additionally, this strategy ensured comparability across datasets for EC and BC, where the availability of MSI testing or IHC results might have differed.

3.3.2 Clinical and Molecular Characteristics of Cancer Subtypes

To further explore the distinct clinical and molecular characteristics of LLS and LSp cases in each cancer type, key clinical features and mutational profiles were analysed. Since the initial progression in LS-associated cancers begins with MMR gene variants, which lead to deficient DNA repair and an accumulation of mutations across the genome, the analysis focused on the frequency of pathogenic variants in MMR genes across the LLS samples. In LLS-EC cases, *MSH2* variants were the most frequent, accounting for seven (39%) of the 18 detected pathogenic MMR variants. This finding aligns with data from the Prospective Lynch Syndrome Database (PLSD) and the study by Marques-de-Sá et al., which reported that *MSH2* is the most common variant in LS-associated EC, with a cumulative risk of 49%^{77,155}. Additionally, *MLH1* variants were identified in five cases (28%), *MSH6* variants in four cases (22%) and *PMS2* variants in two cases (11%). No *EPCAM* variants were found (Table 5). In contrast, in the LLS-BC samples, we identified five variants in *MSH6* and *PMS2* (33%), four in *MLH1* (27%), and one in *MSH2* (6.7%). No *EPCAM* variants were detected.

Table 5. Detailed Distribution of Mismatch Repair Gene Variants in the 13 and 14 Likely Lynch Syndrome Endometrial and Breast Cancer Samples, Respectively

Sample ID	Cancer Type	MMR Gene Variant	Variant Type	MSI Status
TCGA-A8-A07U	Breast	<i>MSH6</i>	SNV	MSS
TCGA-A8-A0A2	Breast	<i>PMS2</i>	SNV	MSS
TCGA-AC-A23C	Breast	<i>PMS2</i>	SNV	MSS
TCGA-AC-A23H	Breast	<i>MSH6</i>	SNV	MSS
TCGA-AN-A046	Breast	<i>MLH1, MSH6</i>	SNV	MSS
TCGA-AN-A0AK	Breast	<i>MLH1</i>	SNV	MSI
TCGA-AO-A03M	Breast	<i>PMS2</i>	SNV	MSS
TCGA-B6-A0IK	Breast	<i>PMS2</i>	SNV	MSS
TCGA-BH-A0HA	Breast	<i>MLH1</i>	SNV	MSS
TCGA-D8-A1J8	Breast	<i>MSH6</i>	SNV	MSS
TCGA-D8-A1XK	Breast	<i>MLH1</i>	SNV	MSI
TCGA-D8-A1Y1	Breast	<i>MSH2</i>	SNV	MSS
TCGA-EW-A2FV	Breast	<i>PMS2</i>	DEL	MSS
TCGA-OL-A5RU	Breast	<i>MSH6</i>	SNV	MSS
TCGA-AJ-A3OJ	Endometrial	<i>MSH6</i>	DEL	MSI
TCGA-AJ-A8CT	Endometrial	<i>MSH2</i>	SNV	MSS
TCGA-AP-A1DM	Endometrial	<i>PMS2</i>	DEL	MSI
TCGA-AX-A1CE	Endometrial	<i>MSH2, MSH6</i>	SNV (<i>MSH2, MSH6</i>); Insertion (<i>MSH6</i>)	MSI
TCGA-AX-A2HA	Endometrial	<i>MSH2</i>	Deletion, SNV	MSI
TCGA-B5-A1MW	Endometrial	<i>PMS2</i>	SNV	MSI
TCGA-B5-A1MX	Endometrial	<i>MLH1</i>	SNV	MSI
TCGA-D1-A174	Endometrial	<i>MSH2</i>	Deletion, Insertion	MSI
TCGA-DI-A1BU	Endometrial	<i>MLH1</i>	SNV	MSI
TCGA-EC-A24G	Endometrial	<i>MSH6</i>	SNV	MSS
TCGA-EO-A22T	Endometrial	<i>MLH1</i>	DEL	MSI
TCGA-EY-A215	Endometrial	<i>MLH1, MSH2</i>	SNV (<i>MLH1</i>), Insertion (<i>MSH2</i>)	MSI
TCGA-EY-A2OM	Endometrial	<i>MLH1</i>	SNV	MSI

Abbreviations: MMR, mismatch repair; MSI, microsatellite instability; MSS, microsatellite stable; SNV, single nucleotide variant

The MSI status and TMB were compared between LLS and LSp samples to explore the impact of MMR deficiency on mutation accumulation and MSI across EC and BCs (Tables 6 and 7). The analysis revealed that the majority of LLS-EC samples (11 of 13; 85%) exhibited MSI,

indicative of defective DNA repair mechanisms that are characteristic of LS and central to the progression model in LS-associated tumours (Figure 15A). MSI, as a hallmark of LS-associated cancers, underscores how MMR deficiencies drive genomic instability, enabling the rapid accumulation of mutations that fuel carcinogenesis in LLS-EC. In contrast, 990 out of 993 (99.7%) LSp-EC samples were microsatellite stable (MSS), highlighting MSI as a distinguishing feature in LS-driven cancers.

In LLS-BC samples, although the majority were MSS (86%), a notable proportion (14%) were MSI. By definition, LLS tumours are expected to be MSI due to dMMR; thus, the MSS classification observed here may reflect alternative mechanisms, such as low tumour purity, somatic rescue of MMR gene function, or tissue-specific factors that influence MSI expression¹⁵⁸. The distinction is particularly relevant for understanding the unique biology of BCs within the LS spectrum. Additionally, the significant difference in MSI presence between LLS-BC and LSp-BC samples underscores molecular distinctions that align LLS-BC cases more closely with LS-associated cancers than with sporadic BC (Figure 15B). These potentially support the hypothesis that a subset of BCs may share mechanisms with other LS-associated cancers, warranting further investigation into the inclusion of BCs in the LS spectrum.

The TMB analysis further highlights the progression from MMR variants to high mutation burdens in LS-associated cancers. The median TMB in LLS-EC samples was **33/Mb**, compared to **2.1/Mb** in LSp-EC samples, indicating a significant difference in mutation burden (Figure 15C). All the LLS-EC samples met the hypermutated threshold (>10 non-synonymous somatic mutations per Mb), underscoring the mutation-driven progression pathway typical of LS-associated tumours. Similarly, LLS-BC samples exhibited a median TMB of **17/Mb**, significantly higher than the **1.4/Mb** observed in LSp-BC samples (Figure 15D). Among the 14 LLS-BC samples, eight (57%) were classified as hypermutated. This elevated mutation burden suggests that MMR deficiencies contribute to a mutation accumulation model in LS-associated BC, paralleling the hypermutation seen in EC and supporting a shared pathway of carcinogenesis.

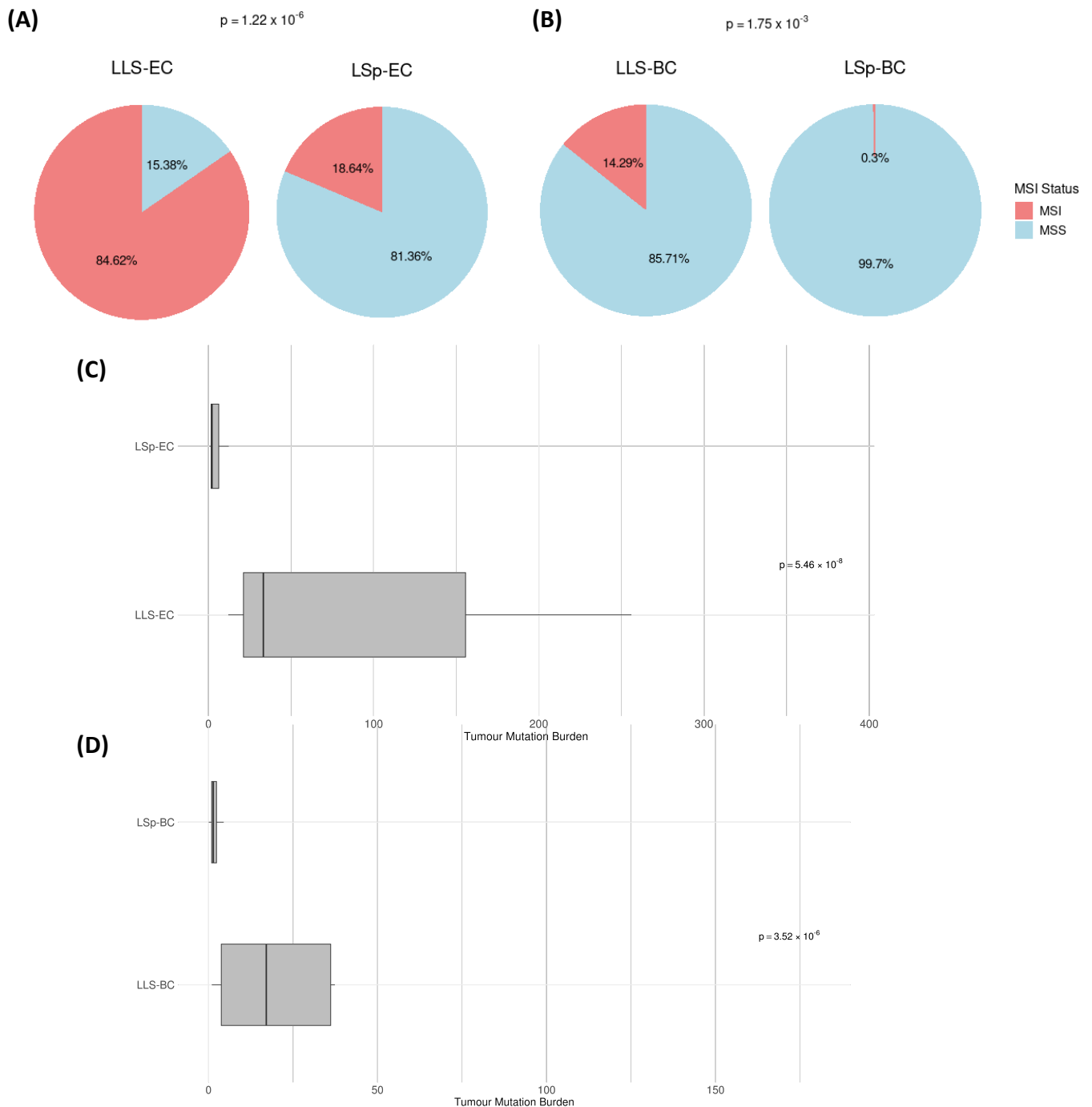


Figure 15. Comparison of Microsatellite Instability (MSI) and Tumour Mutation Burden (TMB) Between likely Lynch Syndrome (LLS) and likely Sporadic (LSp) Endometrial Cancer (EC) and Breast Cancer (BC) Samples. Pie charts displaying the distribution of MSI and microsatellite stable (MSS) status among LLS and LSp in **(A)** EC and **(B)** BC samples. The p -value was from Fishers test. Box and whisker plots comparing the TMB between the two subtypes, highlighting the higher mutation burden in LLS samples in both **(C)** EC and **(D)** BC samples. The p -value was from the Wilcoxon rank-sum test.

Clinical comparisons of EC revealed notable differences between LLS and LSp groups, with uterine endometrioid carcinoma comprising 92% of LLS-EC cases versus 65% in LSp-EC ($p = 0.02$). In contrast, uterine serous carcinoma and uterine papillary serous carcinoma were

more prevalent in LSp samples, reinforcing the distinct histological presentation of LS-associated EC. Other clinical features, such as neoplasm histologic grade, did not differ significantly between the LLS and LSp EC subtypes (Table 6). To further explore this, grades one and two were re-evaluated as low grade and grade three as high grade, in accordance with current recommendations for a binary grading system. Although the majority of LLS-EC cases were high grade, consistent with the histological features of MSI in CRC, the difference in grade distribution between LLS and LSp EC subtypes remained statistically insignificant ($p = 0.23$). For BC, clinical features, including AJCC pathological tumour stage, did not show significant associations with LS status, suggesting that molecular characteristics, rather than clinical staging, may be more indicative of LS-driven pathways in BC (Table 7).

Table 6. Clinical and Molecular Features of the Two Endometrial Cancer Groups

Characteristic	LLS EC (n=13)	LSp EC (n=279)	p-Value
Diagnosis age, median, IQR	55 ; 54 – 58	69 ; 64 – 75	1.07 x 10 ⁻⁰⁹ *
Pathogenic <i>BRAF</i> mutation			0.17
Wild type	12 (92.3%)	276 (98.9%)	
Mutated	1 (7.7%)	3 (1.1%)	
Pathogenic <i>KRAS</i> mutation			0.41
Wild type	10 (76.9%)	240 (86.0%)	
Mutated	3 (23.1%)	39 (14.0%)	
Microsatellite Status			1.22×10 ⁻⁰⁶ *
MSI	11 (84.6%)	52 (18.6%)	
MSS	2 (15.4%)	227 (81.4%)	
Patient Weight, median, IQR	65 ; 58 - 93	83 ; 66 - 98	0.25
Mutation count, median, IQR	1841; 1221 – 8405	101 ; 76 – 315	3.43 x 10 ⁻⁰⁸ *
TMB, median, IQR	33.2 ; 21 – 156	2.1; 2 – 6	5.46×10 ⁻⁰⁸ *
Cancer Type			0.02 *
Endometrial Endometrioid Carcinoma	12 (92.3%)	181 (64.9%)	
Uterine Mixed Endometrial Carcinoma	1 (7.7%)	11 (3.9%)	
Endometrial Serous Carcinoma/Uterine Papillary Serous Carcinoma	0 (0%)	87 (31.2%)	
Neoplasm Histologic Grade			0.23
G1	2 (15.4%)	42 (15.1%)	
G2	0 (0%)	58 (20.8%)	
G3	11 (84.6%)	179 (64.2%)	

Abbreviations: LLS, likely Lynch Syndrome; LSp, likely sporadic; EC, endometrial cancer; MSI, microsatellite instability; MMS, microsatellite stable; IQR, interquartile range; TMB, tumour mutation burden.

*Significant

Table 7. Clinical and Molecular Features of the Two Breast Cancer Groups

Characteristic	LLS BC (n=14)	LSp BC (n=993)	p-Value
Diagnosis age, median, IQR	65; 57 – 74	58; 48 – 67	0.20
Sex			1
Male	0 (0%)	11 (1.1%)	
Female	14 (100%)	982 (98.9%)	
Pathogenic <i>BRAF</i> mutation			1
Wild type	14 (100%)	992 (99.9%)	
Mutated	0 (0%)	1 (0.1%)	
Pathogenic <i>KRAS</i> mutation			1
Wild type	13 (92.9%)	988 (99.5%)	
Mutated	1 (7.1%)	5 (0.5%)	
Microsatellite Status			1.75 x 10 ⁻⁰³ *
MSI	2 (14.3%)	3 (0.3%)	
MSS	12 (85.7%)	990 (99.7%)	
AJCC pathologic tumour stage			0.92
I	2 (14.3%)	168 (16.9%)	
II	8 (57.1%)	574 (57.8%)	
III	4 (28.6%)	214 (21.6%)	
IV	0 (0%)	18 (1.8%)	
Unknown	0 (0%)	19 (1.9%)	
Mutation count, median, IQR	758; 167 – 1,668	60 ; 40-102	1.42 x 10 ⁻⁰⁶ *
TMB, median, IQR	17; 4 – 36	1; 1 – 2	3.52 x 10 ⁻⁰⁶ *
Subtype			0.92
Basal-Like	2 (14.3%)	169 (17.0%)	
HER2-Enriched	2 (14.3%)	76 (7.7%)	
Luminal A	7 (50.0%)	492 (49.6%)	
Luminal B	3 (21.4%)	194 (19.5%)	
Normal-Like	0 (0%)	36 (3.6%)	
Unknown	0 (0%)	26 (2.6%)	

Abbreviations: LLS, likely Lynch Syndrome; LSp, likely sporadic; BC, breast cancer; MSI, microsatellite instability; MMS, microsatellite stable; IQR, interquartile range; TMB, tumour mutation burden.

*Significant

3.3.3 The Mutational Landscape of EC and BC Subtypes

To further understand the progression from MMR variants to malignancy, somatic DNA alterations from all 292 EC and 1,007 BC samples were analysed, using WES data. Consistent with an accelerated mutation accumulation model in LS, the median number of somatic mutations per sample was significantly higher in LLS samples compared to LSp samples in both cancer types. In EC, LLS samples had a median of **1,841** mutations per sample (IQR:1221 – 8405), markedly higher than **101** mutations (IQR: 76–315; $p = 3.4 \times 10^{-08}$) observed in LSp samples. A similar trend was observed in BC, where LLS samples showed a median of **758** mutations per sample (IQR:167 – 1,668) compared to **60** mutations (IQR: 40–102; $p = 1.4 \times 10^{-06}$) in LSp-BC. These findings emphasise the extent of genomic instability in LLS samples, aligning with the high mutation burden driven by MMR deficiencies.

In LLS-EC, a total of 67,658 somatic mutations were identified, involving 17,258 distinct genes. The median number of somatically altered genes per LLS-EC sample was **1,696** (IQR: 1,163–6,212), compared to a significantly lower median of **98** altered genes per sample (IQR: 75–303) in LSp-EC. This substantial difference in the number of altered genes highlights how MMR deficiency in LLS-EC drives extensive mutational events, increasing the likelihood of mutations in key OGs and TSGs that facilitate cancer progression. In LLS-BC, **24,927** somatic mutations were identified across all samples, affecting **11,613** different genes. LLS-BC samples had a median of **702** altered genes (IQR: 162–1,529), which was also significantly higher than the **59** altered genes per sample (IQR: 39–101) observed in LSp-BC. While most of these mutations are likely passengers with no direct functional impact, the extensive genomic instability they reflect increases the likelihood of driver mutations emerging, which can facilitate malignancy. Despite the majority of LLS-BC samples being MSS, the elevated mutational load observed indicates persistent genomic instability, underscoring the potential for tumour progression through key mutational events.

To characterise the types of genomic alterations driving cancer progression, we further focused on mutations in TSG and OGs, as these mutations directly influence cellular pathways involved in tumorigenesis. In LLS-BC, the most frequently altered genes were *PIK3CA* (present in 10 of the 14 samples; (71%) followed by *CACNA1D* and *KMT2A* (both present in six of the

samples; 43%) (Figure 16B). These mutations point to critical pathways involved in cell signalling, proliferation and epigenetic regulation. *PIK3CA* mutations, for instance, activate the PI3K-Akt pathway, a common driver of oncogenesis that facilitates cellular proliferation and survival linking mutation accumulation to pathway dysregulation in LLS-BC¹⁵⁹.

In LLS-EC, the most frequently mutated genes included *PTEN* (present in all 13 samples; 100%), followed by *ARID1A* (present in 11 samples; 85%), and *KMT2B* (present in nine of 13 samples; 69%) (Figure 16A). *PTEN*, a critical TSG, regulates cell growth and inhibits uncontrolled proliferation through PI3K-Akt pathway suppression¹⁶⁰. While mutations in *PTEN* and *ARID1A* were highly prevalent in LLS-EC, it is important to note that these genes are also frequently mutated in endometrial endometrioid adenocarcinomas, regardless of MMR status. Thus their presence in LLS-EC may reflect a background mutational process common to endometrioid tumours rather than being specific to LS-associated EC. However, in the context of dMMR, the widespread genomic instability in LLS-EC may exacerbate the functional consequences of these mutations, disrupting key oncogenic pathways and contributing to tumour progression.

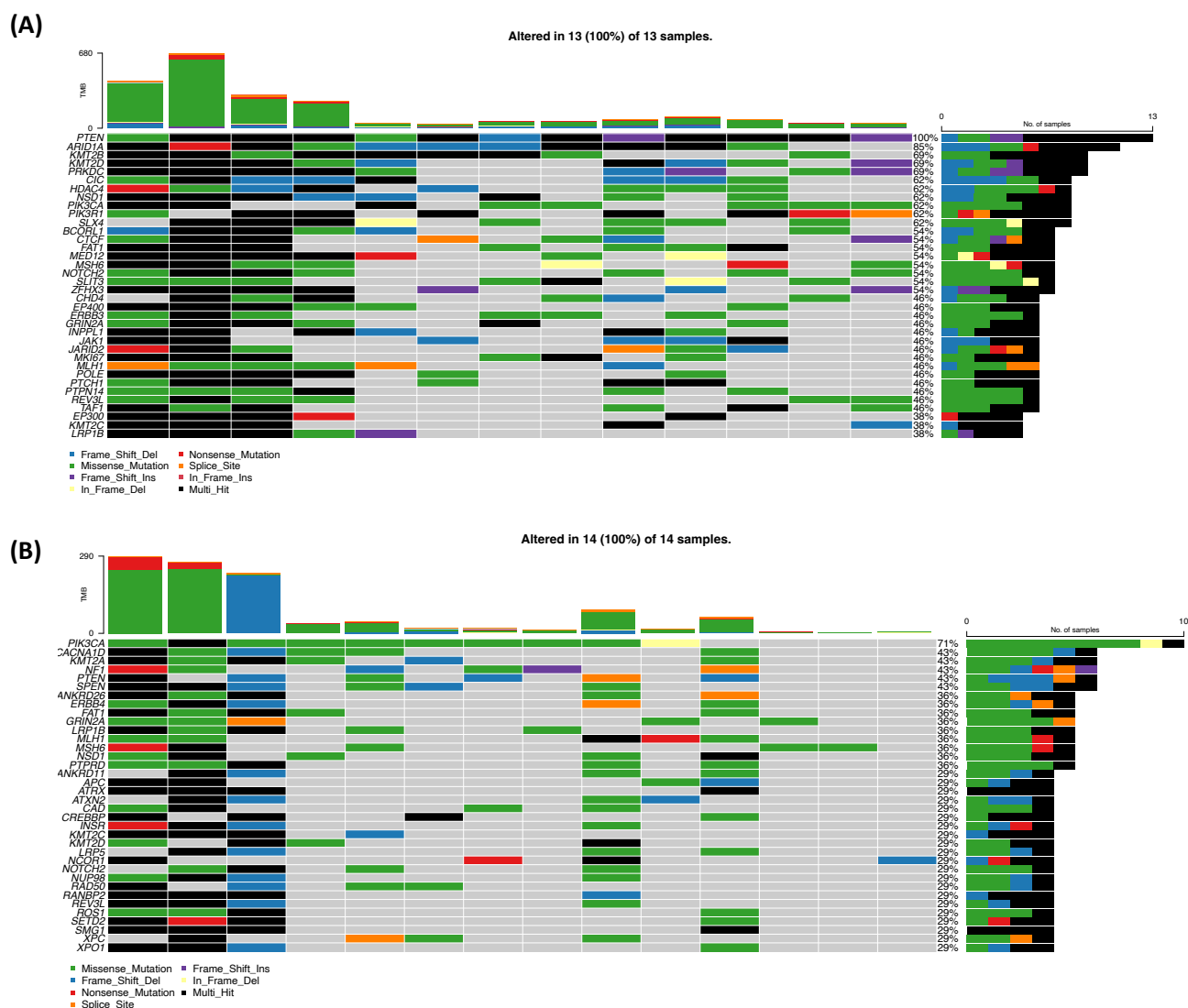


Figure 16. An oncoPrint of the top mutated tumour suppressor and oncogenes in the (A) likely Lynch Syndrome (LLS) endometrial cancer and (B) LLS breast cancer samples.

To further understand the functional consequences of mutations identified in the oncoPrint (Figure 16), we correlated the mutation status of the top genes with their corresponding average gene expression z-scores in LLS samples. In EC, the genes with the highest mutation frequency were the TSGs, *PTEN* and *ARID1A*. In the LLS-EC samples, *PTEN* had an average expression z-score of -0.23, indicating reduced transcription. For *ARID1A*, the average z-score in LLS-EC samples was 0.14, whereas LSp-EC samples exhibited a slightly lower expression level, with a z-score of -0.004. These reduced expression levels in the mutated genes suggest functional roles in tumour suppression, reinforcing their likely involvement in driving cancer progression in LLS-EC samples, when mutated.

In LLS-BC, the genes with the highest mutation frequency were the OGs, *PIK3CA* and *CACNA1D*. In the LLS-BC samples, *PIK3CA* had an average expression z-score of -0.22, indicating reduced transcription. Similarly, *CACNA1D* showed a significantly reduced expression level, with a z-score of -0.53, suggesting downregulation associated with these mutations. The constant downregulation observed in these mutated genes highlights their likely functional roles in tumour suppression, further supporting their involvement in driving cancer progression in LLS-BC samples.

In contrast, the mutation profiles in LSp samples reveal distinct molecular pathways, highlighting differences in the underlying mechanisms driving cancer in sporadic cases versus LS-associated cases. In LSp-EC, the most frequently mutated genes *PTEN* (52%), *TP53* (47%), and *PIK3CA* (40%) (Figure 17A). For LSp-BC, the most common mutations were found in *TP53* (34%), *PIK3CA* (34%), and *CDH1* (13%) (Figure 17B). The prominence of *TP53* mutations in LSp samples suggests that sporadic cancers may progress through mechanisms involving genomic stability checkpoints, distinct from the MMR-driven instability seen in LS¹⁶¹.

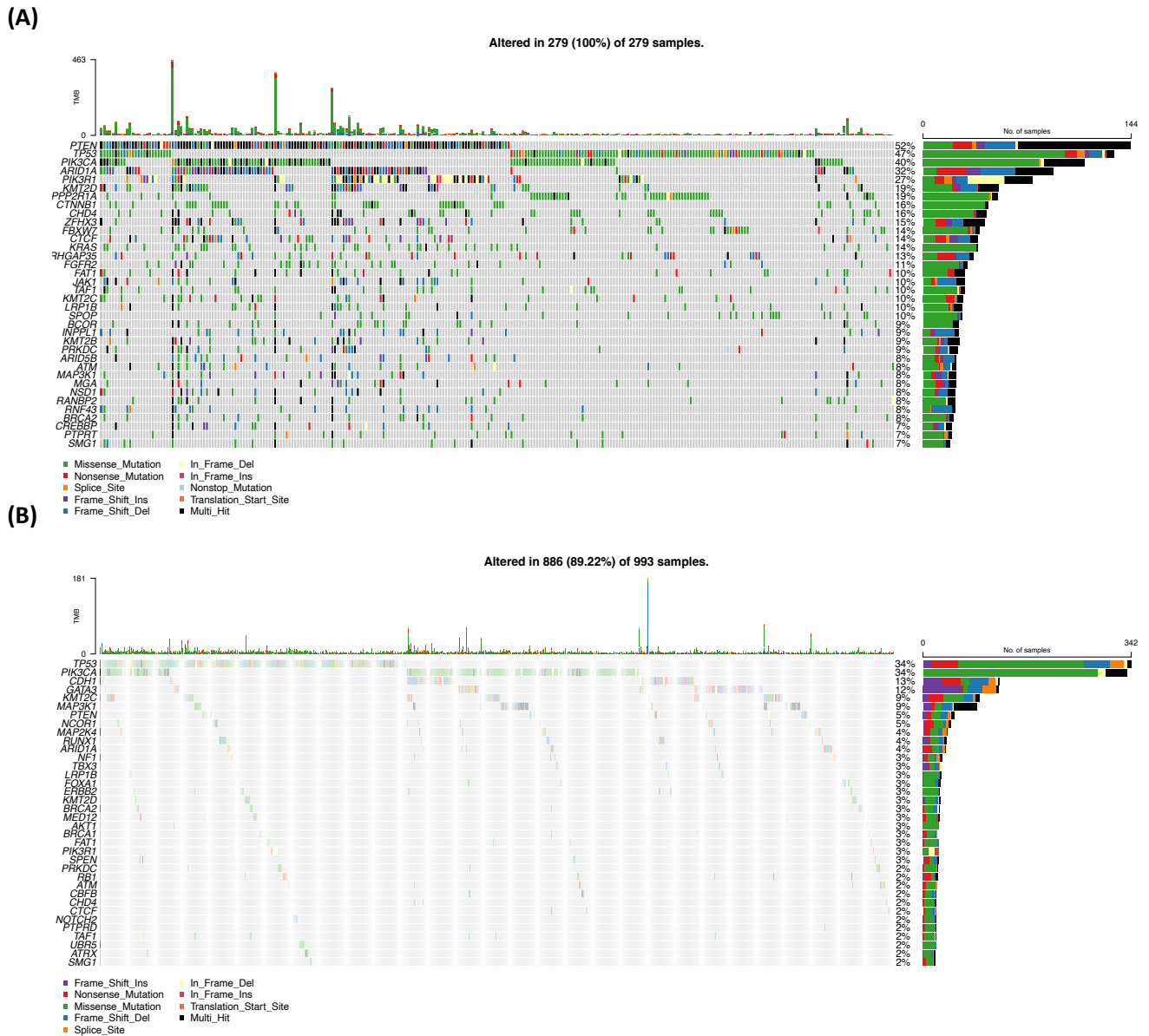


Figure 17. An oncoplot of the top mutated tumour suppressor and oncogenes in (A) likely sporadic (LSp) endometrial cancer and (B) LSp breast cancer samples.

In LLS cancers, MMR deficiencies lead to widespread genomic instability, facilitating mutations in oncogenes and tumour suppressor genes that activate pathways like PI3K-Akt and calcium signalling, driving rapid progression to malignancy. In contrast, LSp cancers follow a pathway more reliant on specific mutations in *TP53* and other stability regulators, reflecting a less extensive mutation-driven model. This divergence in pathways has important implications for targeted therapies, as LS-associated cancers may respond differently to treatments targeting pathway dysregulation linked to MMR deficiency.

3.3.4 Top Mutated Genes

Using Fisher's exact test, TSGs and OGs with significantly different somatic mutation frequencies between the EC and BC subtypes were identified. Notably, all genes showing significant differences in mutation frequency were enriched in the LLS subtypes, highlighting the unique mutational landscape driven by MMR deficiencies in LS-associated cancers. Genes with mutations in approximately one-third or more of LLS cases were considered potential cancer drivers, as their frequent alteration suggests a direct role in driving LS-associated tumorigenesis (Supplementary File 8).

In LLS-EC, 188 "top mutant" genes were identified, including 28 genes involved in the PI3K-Akt signalling pathway ($q = 2.9 \times 10^{-16}$) and 21 in MAPK signalling pathway ($q = 9.5 \times 10^{-12}$) (Figure 18A). The PI3K-Akt pathway is essential for cellular growth, proliferation, and survival, and its dysregulation in LLS tumours suggests that it plays a pivotal role in the progression from MMR deficiency to malignancy¹³⁸. Frequent alterations in PI3K-Akt pathway genes in LLS-EC support a model where MMR variants enable additional mutations in downstream signalling pathways, driving oncogenic transformation and enhancing cell proliferation. The MAPK pathway, similarly involved in regulating cell division and differentiation, further supports this model by providing multiple routes through which mutational dysregulation can accelerate cancer progression in LLS-EC¹⁶².

In LLS-BC, 22 "top mutant" genes were identified, though with a different pathway profile than observed in LLS-EC (Figure 18B). Among these, four genes were implicated in the cAMP signalling pathway ($q = 0.002$) and three were involved in the calcium signalling pathway ($q = 0.02$). While these pathways regulate cellular processes such as signalling and metabolism, their dysregulation in LLS-BC suggests they may support cancer progression in breast tissue by promoting cell proliferation and survival^{140,142}. The presence of mutations in these pathways, even with relatively lower numbers of "top mutant" genes, underscore the role of MMR-driven mutational profiles in activating oncogenic pathways specific to LLS-BC.

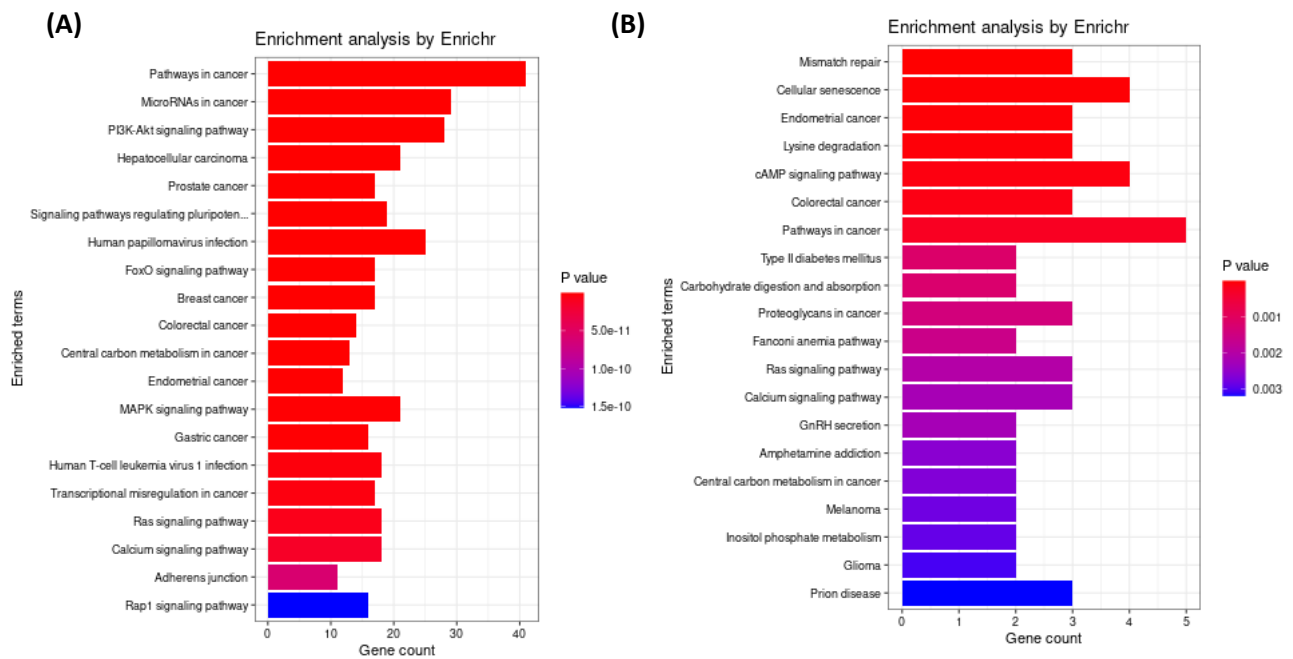


Figure 18. Signalling Pathways in Cancer Subtypes. This figure shows the involvement of the top mutated genes in key signalling pathways in **(A)** likely Lynch Syndrome (LLS) endometrial cancer and **(B)** LLS breast cancer samples.

The differences in pathway enrichment between LLS-EC and LLS-BC highlight how MMR deficient influences distinct molecular pathways in each cancer type. In LLS-EC the prominent dysregulation of PI3K-Akt and MAPK pathways suggests that these signalling mechanisms are central to the aggressive progression from benign to malignant states, driven by MMR deficiencies. In LLS-BC, the enrichment of mutations in the cAMP and calcium signalling pathways underscores an alternative pathway-driven model, where MMR deficiencies contribute to BC progression through unique molecular routes.

3.3.5 Differential Gene Expression Analysis

To investigate the gene expression profiles between the different EC and BC subtypes, a DGE analysis was performed. This analysis identified several genes that were significantly differentially expressed between LLS and LSp-EC; as well as between LLS and LSp-BC samples (Appendix A, Figure S5; Supplementary File 9). These DEGs provide insights into the underlying biological processes and pathways altered in LS-driven cancers, suggesting unique mechanisms that may drive progression from MMR deficient to malignancy in each subtype.

3.3.6 Pathway Enrichment Analysis

To gain insights into the biological processes and pathways associated with the DEGs, we conducted a GSEA using *Enrichr* to compare KEGG pathway classifications of genes upregulated in each EC and BC subtype (Figure 19A and 19B; Supplementary File 10)¹²⁶. This analysis highlighted several key pathways that were consistently altered between the subtypes, revealing distinct molecular pathways and supporting evidence for divergent tumorigenic mechanisms in LLS versus LSp cancers.

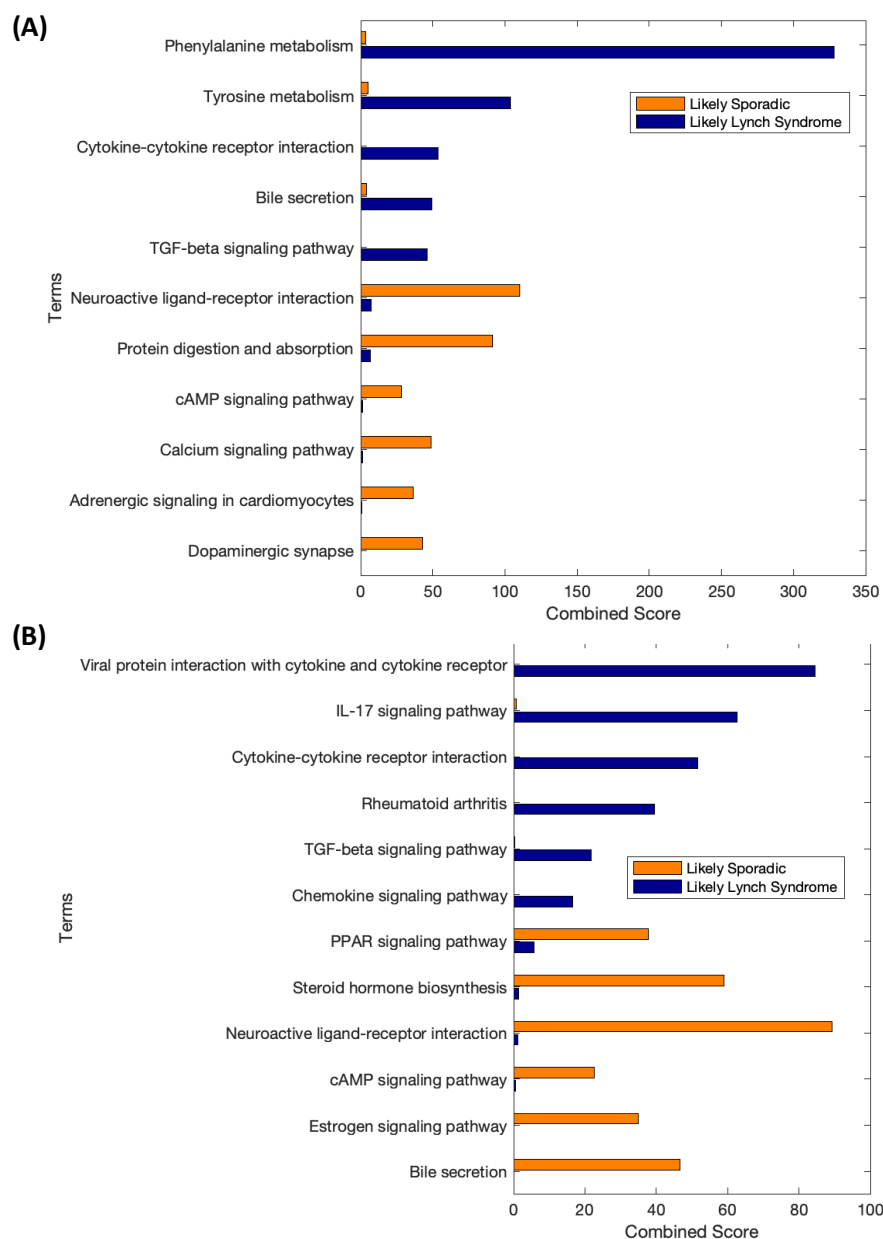


Figure 19. Pathway enrichment analyses: KEGG pathway enrichment results highlight some significant pathways that are differentially altered between the **(A)** endometrial and **(B)** breast cancer subtypes. Refer to Supplementary File 10 for a complete list of KEGG pathways enriched based on the mRNA transcript levels.

In both LLS-EC and LLS-BC tumours, the TGF-beta signalling pathway was significantly upregulated (p -values < 0.05), although q -values were not significant. The TGF-beta pathway regulates cell growth, proliferation, and apoptosis, and its upregulation in LS-associated cancers suggests that it plays a pivotal role in cancer progression from MMR deficiency to malignancy¹⁶³. In LLS-EC, upregulated genes included *GREM1* (average z-score: 0.47 vs. -0.01 in LSp-EC) and *IFNG* (average z-score: 0.65 vs. -0.02 in LSp-EC), while in LLS-BC, *BMP4* (average z-score: 0.50 vs. -0.01 in LSp-BC) and *IFNG* (average z-score: 0.69 vs. -0.01 in LSp-BC) were upregulated. The involvement of the TGF-beta signalling across both cancer types in LLS suggests that it may be a common oncogenic driver in LS-associated tumours, facilitating cellular changes that enable malignancy.

Furthermore, in LLS-BC, we observed significant upregulation in the IL-17 signalling pathway. The IL-17 pathway is known to be a mediator of inflammation and tumour progression, underscoring the role of inflammatory responses in driving tumorigenesis in LLS-BC¹⁴³. This pathway's activation highlights the potential importance of immune modulation in LS-driven breast cancer, potentially creating an environment conducive to malignancy.

In contrast, LSp-EC and LSp-BC tumours both showed significant alterations in the cAMP signalling pathway (q -values < 0.05). The cAMP pathway plays a critical role in regulating cell proliferation, differentiation, and apoptosis, and its activation in sporadic cancers suggests an alternative mechanism of tumour progression compared to LS-driven cancers¹⁴². Additionally, the calcium signalling pathway was significantly upregulated in LSp-EC. This pathway regulates various cellular processes, such as contraction and neurotransmission, and its activation in LSp-EC highlights distinct, sporadic progression pathways divergent from those driven by MMR deficiencies¹⁴¹.

These findings emphasize the divergent molecular pathways between LLS and LSp cancers, where MMR deficiency in LLS cancers predominantly drives oncogenic pathways like TGF-beta and IL-17, facilitating tumour-promoting inflammation and cellular dysregulation. Conversely, LSp cancers rely more on metabolic and cell signalling pathways such as cAMP, calcium, and PPAR, suggesting that sporadic cancers may follow an alternative progression model less

dependent on genomic instability. These pathway differences have potential implications for targeted therapies, as LS-driven tumours may respond better to treatments targeting inflammation and signalling dysregulation, while LSp cancers may benefit from approaches targeting metabolic processes.

3.3.7 Kinase and TF Analysis

Many kinases have been implicated in carcinogenesis, and to identify the upstream regulatory kinases and transcription factors that may be driving the differential gene expression patterns observed between LLS and LSp-CRC, we employed X2K analysis (Figure 20)¹²⁸. The analysis identified several key kinases and transcription factors with strong regulatory roles in these LLS-EC and LLS-BC subtypes.

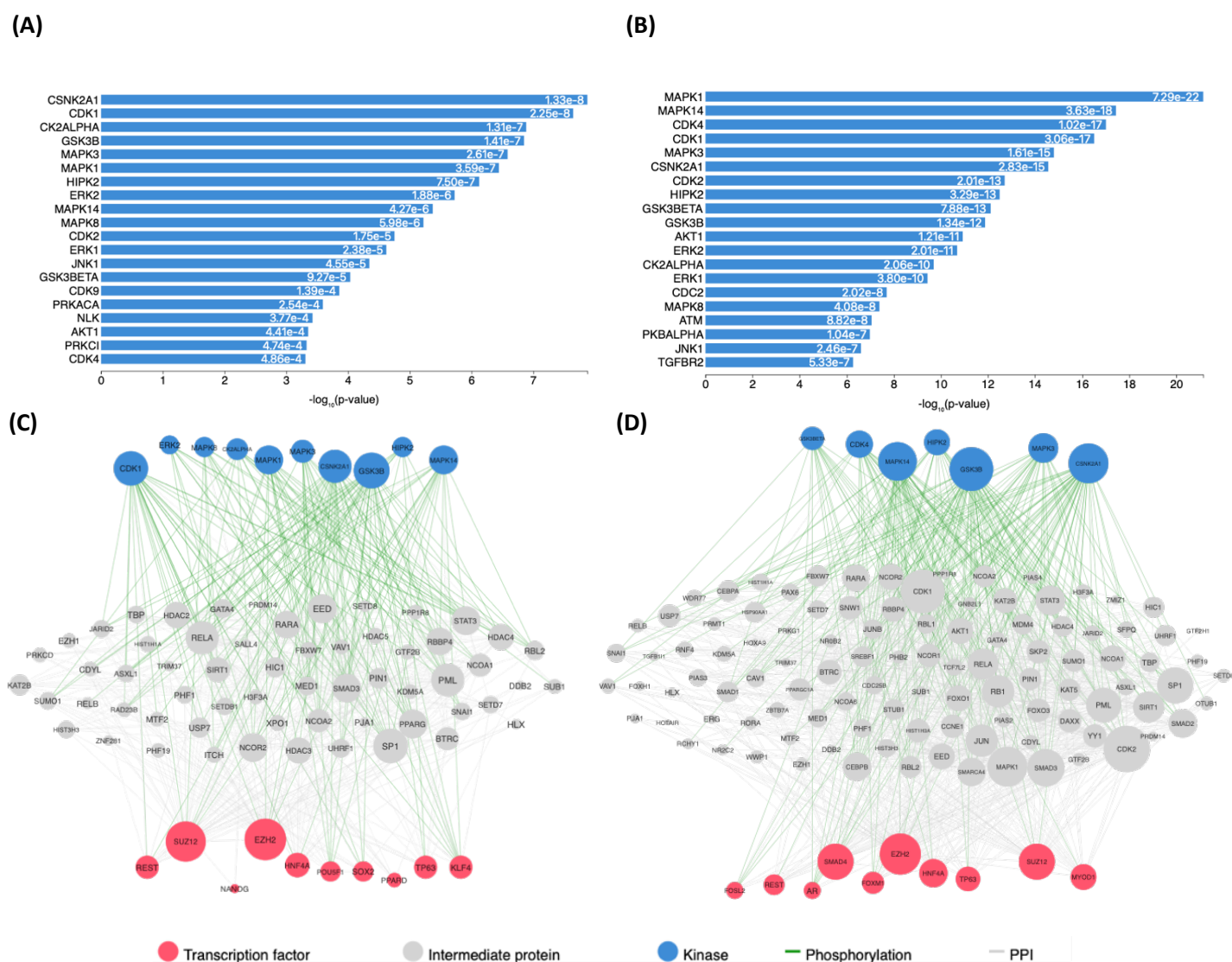


Figure 20. Predicted Regulatory Network and Key Protein Kinases in Likely Lynch Syndrome (LLS) Endometrial (EC) and Breast Cancer (BC). A ranked list of the top predicted protein kinases likely to regulate the observed gene expression differences in **(A)** LLS-EC and **(B)** LLS-BC. An X2K network illustrating the inferred upstream regulatory network predicted to regulate **(C)** LLS-EC and **(D)** LLS-BC, highlighting key kinases and transcription factors. Pink nodes represent the top transcription factors predicted to regulate the expression of the LLS cancers, and blue nodes indicate the top predicted protein kinases known to phosphorylate the proteins within the expanded subnetwork. Green edges represent kinases-substrate phosphorylation interactions, while grey edges depict protein-protein interactions.

In LLS-EC, among the top 10 kinases identified, three OGs were observed: *GSK3*, *MAPK3* and *MAPK1*. In LLS-BC, four OGs emerged among the top kinases: *MAPK1*, *CDK4*, *MAPK3* and *GSK3B*. Both *MAPK3* and *MAPK1* are involved in the TGF-beta signalling pathway, underscoring the pathway's potential role in promoting tumour progression across these LS-

associated cancers. In LLS-BC, *CDK4* and *GSK3B* were also prominent, suggesting their involvement in cell cycle regulation, which may contribute to the unchecked growth characteristics of LLS-BC.

In addition to kinases, several transcription factors were identified as potential regulators of gene expression differences between LLS and LSp subtypes in both EC and BC. Notably, in LLS-EC, *SUZ12*, *EZH2*, and *REST* were among the top transcription factors identified, each playing an established role in cancer progression (Figure 21)¹⁴⁶. *SUZ12* and *EZH2*, components of the PRC2, are known to silence TSGs through histone methylation, thereby promoting a repressive chromatin state that facilitates tumour growth. *REST*, on the other hand, represses neuronal genes and enhances cell survival in non-neuronal cells, actions that contribute to tumour initiation and resistance to apoptosis^{147,148}. The significant overexpression of *SUZ12*, *EZH2* and *REST* in LLS-EC suggests their involvement in driving tumorigenesis through interactions with the TGF-beta pathway, which plays a crucial role in cell differentiation, proliferation and immune modulation. In LLS-BC, as in LLS-EC, *SUZ12* and *REST* were prominent among the top transcription factors, indicating that these transcriptional regulators may contribute to oncogenesis in a similar manner across both cancer types. Additionally, *SMAD4* was identified in LLS-BC, a key mediator of the canonical TGF-beta signalling pathway.

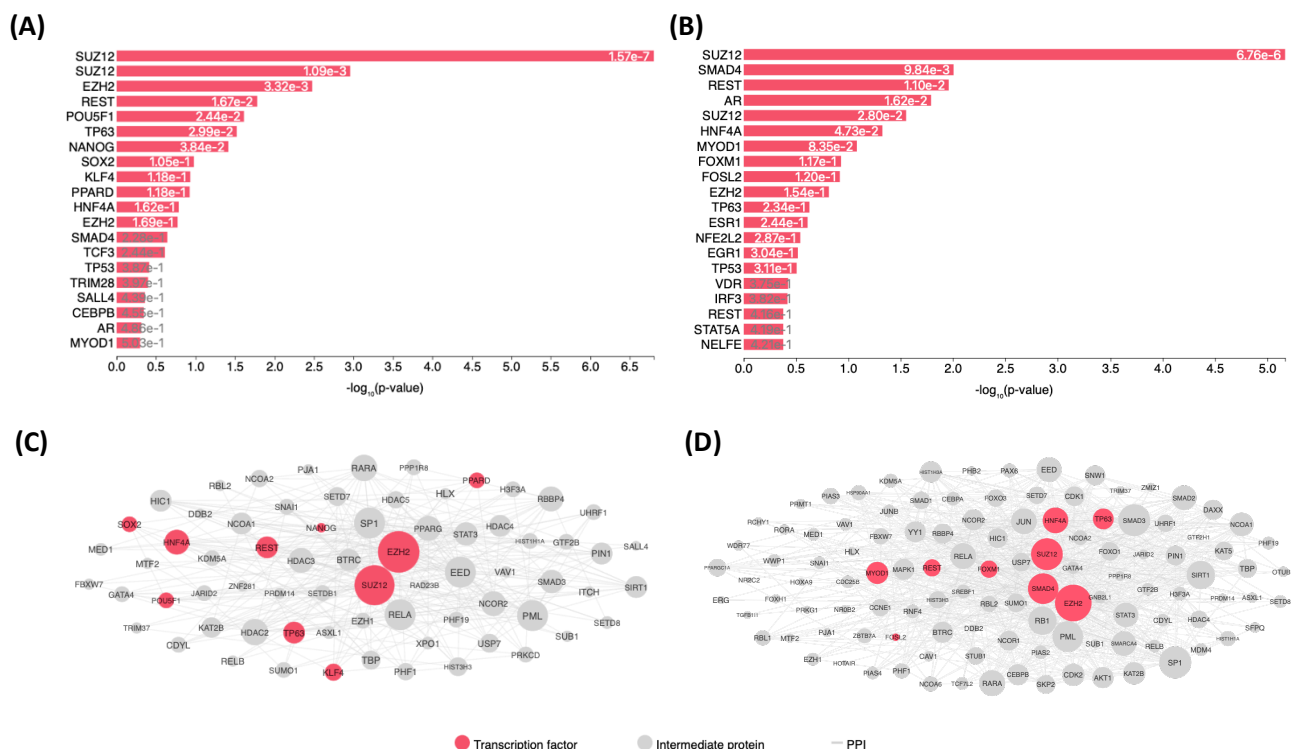


Figure 21. Predicted Transcriptional Regulatory Network in Likely Lynch (LLS) Endometrial (EC) and Breast Cancer (BC). Ranked list of the top predicted transcription factors regulating (A) LLS-EC and (B) LLS-BC, as determined by ChIP-seq experiments (ChEA). Ball-and-stick diagrams of the subnetwork of connected transcription factors and their interacting proteins in (C) LLS-EC and (D) LLS-BC. Pink nodes represent transcription factors, while grey nodes indicate proteins that connect them. Node sizes are proportional to their degree of connectivity.

3.3.8 Drug Sensitivity and Dependency Analysis

To explore the potential therapeutic implications, this study aimed to examine the relationship between TGF-beta signalling pathway dependency and drug sensitivity using data from the GDSC database¹³⁰. However, the GDSC lacks data specific to the TGF-beta pathway, the analysis instead focused on pathways with known interactions with TGF-beta signalling: receptor tyrosine kinase (RTK) signalling, ERK MAPK signalling, cell cycle regulation and apoptosis.

RTLs can initiate signalling cascades that interact with or modulate TGF-beta pathways. In the tumour microenvironment, RTK and TGF-beta signalling often engage in crosstalk, influencing processes such as cell growth and differentiation¹⁶⁴. Similarly, the ERK MAPK pathway intersects with TGF-beta signalling in cancer progression, particularly in epithelial-to-mesenchymal transition, which is relevant to LLS-EC and LLS-BC¹⁶⁵.

TGF-beta also plays a roles in cell cycle regulation by influencing growth arrest and progression, suggesting that drugs targeting cell cycle pathways may indirectly impact TGF-beta pathway activity¹⁶⁶. Additionally, TGF-beta signalling can promote either cell survival or apoptosis depending on the cellular context, making drugs affecting apoptosis regulation potentially relevant due TGF-beta's impact on cancer cell fate. Therefore, response data from 357 cancer cell lines treated with 109 drugs targeting these pathways were analysed to identify drugs that may influence TGF-beta-related activities indirectly through these intersecting pathways.

For each gene involved in TGF-beta-related pathways, cancer cell lines were categorised into high or low-dependency groups based on their CRISPR-derived dependency scores (Supplementary File 11). It was observed that the sensitivity of cell lines to the genes varied according to their dependency on specific TGF-beta-related pathway genes. Specifically, high dependency scores of two genes (*RHOA* and *SP1*) were linked to significantly increased sensitivity to these inhibitors, while low dependency scores of six genes were associated with decreased sensitivity. Interestingly, 14 other genes exhibited mixed responses; i.e. significantly increased sensitivity to some of the inhibitors and significantly decreased sensitivity to others, indicating complex gene-drug interactions. Cell lines with higher dependency on genes such as *SP1* and *RHOA* showed significantly increased sensitivity to 64% and 56% of the TGF-beta-related inhibitors, respectively. Conversely, dependence on *HJV* and *RBX1* correlated with reduced sensitivity to 51% and 43% of inhibitors, respectively.

Further, grouping cell lines based on the presence of mutations in the TGF-beta-related pathway genes revealed that high dependency scores of three genes (*MYC*, *AMHR2* and *SMURF1*) were linked to significantly increased sensitivity to these inhibitors, while low dependency scores of 26 genes were associated with decreased sensitivity (Supplementary File 11). One other gene (*LTBP1*) exhibited mixed responses. Cell lines with higher dependency *AMHR2* showed significantly increased sensitivity to 50% of the TGF-beta-related inhibitors. Conversely, dependence on *SMAD4* and *ID4* correlated with reduced sensitivity to 43% and 36% of inhibitors, respectively.

Overall, we identified 319 instances where high dependency or mutations in TGF-beta-related pathway genes significantly influenced drug sensitivity (Figure 22A and 22B). High dependency on TGF-beta-related pathway genes was associated with increased sensitivity to drugs such as selumetinib, refametinib and trametinib in 127 instances and resistance in 148 instances. In contrast, mutation status was associated with increased sensitivity to drugs such as selumetinib, refametinib and sepantronium bromide in four instances and resistance in 40 instances. This suggests that there is a greater predictive power of functional dependency in determining drug response. These findings suggest that targeting specific gene dependencies could optimise treatment strategies for LS-BC and LS-BC patients.

Cell lines were further stratified into two groups (higher and lower CRISPR-derived dependency) based on their overall TGF-beta-related pathway dependency and compared drug IC₅₀ values between these groups for the 51 inhibitors. Cell lines with higher dependency were significantly more sensitive to all TGF-beta-related inhibitors, with selumetinib ($t = 10.7$, $p = 0.8 \times 10^{-20}$), refametinib ($t = 10.4$, $p = 4.4 \times 10^{-18}$), and trametinib ($t = 9.4$, $p = 3.3 \times 10^{-14}$) showing the most significant differences in dose-responses between the groups (Figure 22C). These findings suggest that these inhibitors could be particularly effective in treating TGF-beta-related-dependent LLS cancers and may inform future patient stratification strategies based on pathway dependency.

These results have important therapeutic implications for LS-associated EC and BC. The high TGF-beta pathway activity observed suggests that inhibitors like refametinib and selumetinib could offer significant clinical benefits, particularly for patients whose tumours show high dependency on this pathway.

along rows. Green, all the cell lines are significantly more sensitive to all the TGF-beta-related pathway inhibitors, orange, all the cell lines are significantly more resistant to all the TGF-beta-related pathway inhibitors and black is a mixed response to TGF-beta-related pathway inhibitors. **(B)** A bar graph showing the total number of instances for the resistant (red) and sensitive (blue) CRISPR gene dependencies **(C)** Comparison of the dose-response profiles between the higher and lower dependency groups of the GDSC cancer cell lines for the top nine inhibitors. The boxplots show the logarithm-transformed mean IC50 values of the cancer cell lines. On each box, the central black line indicates the median, and the bottom edge represents the 25th percentiles, while the top edge of the box represents the 75th percentiles. The whiskers extend to the most extreme data points not considering outliers, which are plotted individually in bold, black circles.

Building upon the identification of TGF-beta pathway dependencies in LLS-EC and LLS-BC, the analysis was expanded to investigate specific vulnerabilities within the cAMP signalling pathway in LSp-EC and BC. Recognising the role of this pathway in the progression of LSp-EC and BC, this analysis aimed to assess the effectiveness of targeted inhibitors as potential therapeutic interventions. However, as the GDSC lacks data specific to the cAMP pathway, the analysis instead focused on pathways with known interactions with cAMP signalling, including hormone-related and metabolism pathways.

Dose-response data for cAMP-related pathway inhibitors across multiple cancer cell lines was analysed, stratified by their CRISPR-derived dependency scores for cAMP-related pathway genes (Supplementary File 12). The results revealed varied impacts of cAMP-related pathway gene dependency on drug sensitivity. Notably, 19 genes were associated with increased sensitivity to cAMP-related pathway inhibitors, while additionally 19 genes were correlated with decreased sensitivity (Appendix A; Figure S6A). Interestingly, seven other genes exhibited mixed responses. Cell lines with high dependency on genes such as *ATP1B3* exhibited increased sensitivity to 60% of cAMP-related inhibitors, indicating that targeting these dependencies could offer therapeutic benefits. On the other hand, cell lines with high dependency on *RAC1* demonstrated reduced sensitivity to 67% of inhibitors, suggesting potential resistance mechanisms.

Mutation status alone proved to not be a reliable predictor of drug response, with no mutations showing significant associations with sensitivity to treatment. This highlights the possibility that functional dependency, rather than genetic mutation status, may serve as a more accurate indicator of drug efficacy in sporadic EC and BC. Our analysis identified 116 cases where either high dependence or mutations in cAMP-related pathway genes had a significant impact on drug sensitivity. Specifically, cell lines with greater dependency on

cAMP-related pathway genes demonstrated heightened sensitivity in 49 cases and resistance in 67 cases (Appendix A; Figure S6B). These results further underscore the crucial role of pathway dependency in influencing drug responses. These insights point to the therapeutic potential of targeting the cAMP-related pathway in sporadic EC and BC, especially with inhibitors such as Tamoxifen, N-acetyl cysteine and Bicalutamide, which may prove beneficial for patients with cAMP-related-dependent tumours.

Cell lines were further stratified into two groups (higher and lower CRISPR-derived dependency) based on their overall cAMP-related pathway dependency and compared drug IC50 values between these groups for the 51 inhibitors. Cell lines with higher dependency were significantly more sensitive to all cAMP-related inhibitors, with daporinad ($t = 4.5$, $p = 1.3 \times 10^{-05}$), AGI-6780 ($t = 4.2$, $p = 2.9 \times 10^{-05}$), and fulvestrant ($t = 4.4$, $p = 3.6 \times 10^{-05}$) showing the most significant differences in dose-responses between the groups (Appendix A; Figure S6C). These findings suggest that these inhibitors could be particularly effective in treating cAMP-related dependent LSp cancers and may inform future patient stratification strategies based on pathway dependency.

3.4 Discussion

This study explores mechanisms in LLS-associated EC and BC, focusing on mutation accumulation, pathway dysregulation, and DGE profiles. By comparing LS-associated cancers to their sporadic counterparts, distinct molecular and genomic patterns were identified, providing critical insights into how dMMR drives tumorigenesis in LS-associated cancers and exploring the potential inclusion of BC within the LS cancer spectrum.

A key finding of this study is the markedly higher TMB and MSI observed in LS-associated cancers compared to their sporadic counterparts, consistent with prior findings in LS-driven CRC¹⁶⁷. In LLS-EC, the high TMB and MSI in 85% of cases underscore the significant impact of dMMR in driving genomic instability, which accelerated mutation accumulation. Similarly, in LLS-BC, while the majority of samples were MSS – a finding that warrants further investigation – a notable subset exhibited MSI, aligning more closely with the expected molecular profile of LS-associated tumours. The MSS status observed in the majority of LLS-BC samples may

reflect alternative mechanisms, such as somatic rescue of MMR gene function, low tumour purity, or tissue-specific factors that suppress the manifestation of MSI¹⁵⁸. These findings highlight the complexity of LLS-BC classification and suggest that further research is needed to refine the molecular definition of LS-associated BC.

The increased mutation burden in LLS-EC and LLS-BC aligns with an accelerated progression model in LS-associated cancers, where dMMR leads to the rapid accumulation of somatic mutations, including critical alterations in OGs and TSGs¹⁶⁸. This contrasts with the progression model in sporadic cancers, which relies more heavily on mutations in stability-regulating genes like *TP53*. These differences have potential clinical implications, as LS-associated cancers may benefit from treatments targeting MSI-driven genomic instability, such as immune checkpoint inhibitors.

Pathway enrichment analyses, using mutation data, revealed key differences in oncogenic pathways between LLS and LSp cancers, highlighting mechanisms through which LS-associated cancers progress. In LLS-EC, frequent mutations in *PTEN*, *ARID1A*, and *KMT2B* suggest tumorigenic mechanisms centred on cell cycle control and chromatin remodelling. Similarly, LLS-BC showed frequent alterations in *PIK3CA*, *CACNA1D*, and *KMT2A* implicating pathways related to cell signalling and epigenetic regulation¹⁶⁹. The PI3K-Akt pathway was particularly dysregulated in LLS-EC, with 28 “top mutant” genes associated with this pathway. This role of this pathway in regulating cell growth, proliferation, and survival underscores its importance as an initiation/progression mechanism in LS-associated EC, where dMMR-driven mutations facilitate pathway activation^{138,170}.

However from the GSEA, TGF-beta signalling emerged as a major oncogenic driver in both LLS-EC and LLS-BC, though the composition of upregulated genes differed. In LLS-EC, genes such as *GREM1* and *BMP2* were upregulated, while *BMP4* and *NOG* were prominent in LLS-BC. TGF-beta signalling promotes tumour progression by enhancing cell proliferation, invasion, and metastasis, while simultaneously suppressing antitumour immune responses, fostering an immunosuppressive tumour microenvironment¹⁶⁶. This dual role highlights its potential as a therapeutic target in LS-associated cancers.

Drugs like refametinib and selumetinib emerged as promising therapeutic options by targeting the downstream ERK/MAPK signalling cascade, which often interacts with TGF-beta to promote tumour growth and immune evasion¹⁷¹. These drugs may inhibit tumour progression and restore immune responsiveness, enhancing the efficacy of existing treatment strategies. The role of TGF-beta signalling as a shared driver in LS-associated cancers underscores the importance of pathway-specific therapies to address the challenges posed by these tumours.

Conversely, LSp cancers showed reliance on pathways such as cAMP and calcium signalling, which play roles in cell signalling and metabolism^{141,142}. The upregulation of the cAMP signalling pathway in LSp-EC and LSp-BC suggests that sporadic cancers may depend more on metabolic adaptations and signalling dynamics, rather than the genomic instability that characterises LS-associated tumours. This reliance on metabolic pathways highlights an alternative model of tumour progression, potentially offering targeted therapy options distinct from those suited for LS-driven cancers. Hormone-related inhibitors such as tamoxifen and fulvestrant disrupt tumour-promoting signalling in hormone-sensitive cancers like EC and BC¹⁷². Metabolic inhibitors such as daporinad, a nicotinamide phosphoribosyltransferase inhibitor, and AGI-6780, an isocitrate dehydrogenase 2 inhibitor, target metabolic vulnerabilities critical for cancer cell survival^{173,174}. These drugs exemplify how targeted approaches can exploit metabolic and signalling dependencies in cancers.

Additionally, the analysis identified critical transcription factors and kinases as regulatory drivers of the distinct molecular landscapes in LLS-EC and LLS-BC. In LLS-EC, transcription factors such as *SUZ12*, *EZH2*, and *REST*, along with kinases like *GSK3*, *MAPK1*, and *MAPK3*, were prominent regulators, interacting with TGF-beta signalling to drive tumour progression. In LLS-EC, transcription factors like *SUZ12* and *REST* were similarly prominent, while kinases included *MAPK1*, *CDK4*, *MAPK3* and *GSK3B*. *SMAD4*, a mediator of the canonical TGF-beta signalling pathway, was also identified in LLS-BC¹⁷⁵. These findings suggest that combining TGF-beta inhibitors with drugs targeting these regulatory nodes could enhance therapeutic efficacy in LS-associated cancers.

The distinct molecular characteristics identified in LS-associated EC and BC have important clinical implications. The high TMB, MSI, and unique mutational profiles in LS-driven cancers may serve as biomarkers for identifying patients with LS, especially in cancers like BC where its inclusion in the LS spectrum is still debated. Furthermore, the differential reliance on pathways like PI3K-Akt, TGF-beta, and immune signalling in LS-associated cancers suggests targeted therapeutic strategies that exploit these LS-specific vulnerabilities.

This study provides evidence supporting the inclusion of BC within the LS spectrum. The MSI observed in a subset of LLS-BC samples and the shared dysregulated pathways between LLS-BC and LSp-BC highlight potential common mechanisms of tumorigenesis. These findings suggest that BC in LS patients may follow similar progression models as LS-associated EC, warranting its consideration within the LS tumour spectrum.

While these findings provide novel insights, the study is not without limitations. First, the relatively small sample sizes for LLS-associated EC and BC may limit the generalisability of the findings. Expanding the sample size across cancer types could enhance statistical power and provide more comprehensive insights. Second, the reliance on computational analyses without experimental validation restricts the clinical applicability of the findings. Functional assays and *in vivo* studies are essential to validate the roles of identified driver genes and pathways. Lastly, the absence of direct pathway-specific data, such as for the cAMP and TGF-beta pathways, restricted the evaluation of therapeutic targets.

Future research should explore emerging technologies such as single-cell transcriptomics and spatial transcriptomics to investigate tumour microenvironment interactions in LS-associated cancers. Additionally, expanding the analysis to other cancers within the LS spectrum could provide a more comprehensive understanding of LS-driven tumorigenesis. The findings of this study emphasise the need for personalised therapeutic strategies based on tumour subtype, genetic profile, and pathway dependencies to optimise patient outcomes in LS-associated cancers.

3.5 Conclusion

This study provides a comprehensive overview of the molecular mechanisms distinguishing LS-associated from sporadic EC and BCs, elucidating the progression from dMMR to malignancy in the LS context. The findings support the potential inclusion of BC within the LS spectrum, as LLS-BC exhibits pathway dysregulation and mutation accumulation patterns consistent with established LS cancers. Notably, the identification of the TGF-beta pathway as a critical driver in LS-associated cancers highlights the value of targeted therapeutic approaches that address LS-specific molecular alterations, potentially improving treatment outcomes for patients with LS-driven cancers. In contrast, cAMP signalling emerged as a key dysregulated pathway in LSp EC and BC, suggesting distinct progression mechanisms in sporadic cancers. Furthermore, the study underscores the potential pathway-specific drugs tailored for LS and sporadic patients with high dependency on their respective pathways, paving the way for more personalised and effective therapeutic strategies.

Chapter 4: Concluding Remarks

The aim of this study was to elucidate the underlying mechanism and pathways contributing to LS-associated cancers, including CRC and EC. In addition, it aimed to evaluate whether BC could be considered part of the LS spectrum of cancers.

4.1 General Discussion

This thesis has explored the molecular mechanisms underpinning LS-associated cancers, focusing on CRC, EC and BC. By integrating multi-omic data and conducting pathway and mutational analyses, this work provides new insights into how dMMR drives tumorigenesis and highlights distinct differences between LS-associated and sporadic cancers.

The findings presented here confirm the central role of genomic instability in LS-associated cancers. Elevated TMB and MSI were hallmarks of LS-driven CRC and BC. While MSI was less frequent in LLS-BC, its presence in a subset of cases, coupled with a significant difference in MSI status compared to that of LSp-BC, strengthens the argument for the inclusion of BC within the LS tumour spectrum. This data reinforces the view that dMMR sets the stage for tumorigenesis by enabling rapid mutation accumulation.

Key pathways have been identified as drivers of LS-associated cancers. The PI3K-Akt signalling pathway emerged as a major oncogenic driver in LLS-CRC, while TGF-beta signalling was a dominant feature in both LLS-EC and LLS-BC. Conversely, in sporadic cancers, LSp-CRC exhibited aberrant Wnt signalling pathway, while LSp-EC and LSp-BC exhibited reliance on the cAMP signalling pathway, highlighting fundamental differences in tumour initiation and progression between LLS-associated and LSp cancers.

The differences in pathway activity between LLS-CRC and LLS-EC and LLS-BC may arise from the unique biology and environment of each tissue type. CRC develops in the gut, a dynamic microenvironment shaped by the microbiome, diet, and high rates of cell turnover¹⁷⁶. These factors create selective pressures that favour activation of pathways like PI3K-Akt, which supports cell proliferation and survival under the intense metabolic demands and inflammatory conditions of the gastrointestinal tract¹⁷⁷. In contrast, the tissues affected by

ECs and BCs are hormonally regulated, with growth often influenced by oestrogen and progesterone. This shared hormonal sensitivity likely explains why LLS-EC and LLS-BC both show upregulation of the TGF-beta pathway, a key driver of processes like epithelial-mesenchymal transition, immune evasion and metastasis^{178,179}.

Although the underlying defects in all LS cancers stems from dMMR and MSI, the way these molecular disruptions manifest varies depending on the tissue. In colorectal tissue, chronic exposure to dietary carcinogens, bile acids, and gut microbiota metabolites may drive mutations and selectively activate pathways like PI3K-Akt, which is critical for overcoming the pro-apoptotic environment of the colon¹⁸⁰. By contrast, in endometrial and breast tissues, the tumour microenvironment and systemic hormonal signal create conditions that favour TGF-beta signalling, which supports stromal-epithelial interaction and facilitates tumour progression^{178,179}.

The similarity in pathway activity between LLS-EC and LLS-BC may be rooted in their shared dependence on hormonal and stromal interactions. Both tissues are part of the reproductive axis and respond to similar endocrine signals. The role of the TGF-beta pathway in regulating immune response, extracellular matrix remodelling, and metastatic potential makes it particularly relevant to these cancers¹⁶³. Additionally, these tissues exhibit a unique balance between epithelial and stromal contributions, further tying their molecular profiles together.

These distinctions have significant implications for treatment strategies. The study also revealed potential therapeutic opportunities. For LLS-CRC, targeting the PI3K-Akt pathway with specific inhibitors (e.g. alpelisib) could address the reliance of the tumour on this pathway for survival and growth. Meanwhile, therapies for LLS-EC and LLS-BC might be more effective if they focus on blocking TGF-beta signalling or interfering with stromal-epithelial interactions. Ultimately, the differences in pathway activity highlight how the unique biology of each tissue influences tumour development, while the overlap between LLS-EC and LLS-BC reveals shared vulnerabilities that could guide targeted therapies. These insights not only deepen our understanding of LS-associated cancers but also underscore the importance of tailoring treatment to the specific context of each cancer type.

4.2 Study Limitations and Future Directions

This study highlights the molecular distinctions between LS and sporadic cancers, providing important insights into tumour initiation, progression and therapeutic opportunities. However, several limitations must be acknowledged. A primary limitation is the reliance on publicly available datasets, which, while offering extensive genomic and transcriptomic data, may introduce biases due to sample selection and cohort representation. The relatively small number of LLS cases, reduces the statistical power and limits the generalisability of the findings. At best the present study ought to be considered exploratory; expanding the study to include larger and more diverse cohorts is essential to validate the observed molecular distinctions and enhance the robustness of the conclusions.

One key limitation of this study is the absence of germline data for the patients in the TCGA datasets. As a result, the potential hereditary contributions of specific mutations to tumour development could not be assessed, limiting the full understanding of the genetic landscape of LS. The reliance on somatic mutations for classification and analysis may lead to the misclassification of samples. Consequently, this limitation could affect the interpretation of LS tumour characteristics, as the findings predominantly focus on somatic mutations in tumour tissue. Furthermore, the classification of tumours as LLS or LSp was based on the presence of somatic mutations and age of diagnosis. MSI-H tumours, which could belong to either group due to the lack of germline verification, may have been misclassified, further confounding the analysis. This is particularly significant as somatic mutations may not fully capture the complex interplay between germline mutations and epigenetic changes in tumour development.

In LLS subtypes, the identification of the PI3K-Akt and TGF-beta pathways as critical oncogenic drivers suggests promising therapeutic targets, but further research is needed to establish their functional roles and the validity of these findings. Mechanistic studies using *in vitro* and *in vivo* models could directly confirm the involvement of these pathways in tumorigenesis and provide evidence for their therapeutic potential. Further investigation is needed regarding the pathway interactions with other regulatory pathways. Functional experiments exploring

the effects of pathway-specific inhibitors will be crucial in validating these findings and advancing therapeutic strategies.

The inclusion of BC within the LS spectrum remains a subject of debate. While MSI was less frequent in LLS-BC compared to LS-CRC or LS-EC, its presence in a subset of cases and the significant differences in MSI status between LLS-BC and LSp-BC support the potential inclusion of BC within the spectrum of LS-associated cancers. Larger cohort studies and integrative multiomic approaches, incorporating proteomic and epigenomic data, are necessary to provide more conclusive evidence. Furthermore, experimental validation of pathways such as TGF-beta signalling in LS-associated BC and EC could elucidate their roles in tumour progression and inform the development of targeted therapies.

Another limitation is the reliance on computational methods to infer pathway dependencies and drug sensitivities. While these approaches offer powerful insights, experimental validation is critical to confirm the clinical relevance of these findings. Preclinical studies using cell lines and patient-derived xenograft models could assess the efficacy of inhibitors targeting key pathways like PI3K-Akt and TGF-beta in LS-CRC, LS-EC, and LS-BC. Additionally, investigating the potential of combination therapies that target both primary oncogenic drivers and immune-modulatory pathways could enhance therapeutic outcomes. The study also underscores the need for deeper exploration of immune-related pathways and their interplay with oncogenic drivers in LS-associated cancers to identify additional therapeutic opportunities.

Future efforts should focus on validating the findings in independent datasets and, perhaps, in diverse populations to ensure their applicability across different patient groups. Functional studies should prioritise understanding the mechanistic roles of key dysregulated pathways, such as PI3K-Akt in LS-CRC and TGF-beta in LS-EC and LS-BC. Integrating proteomic, epigenomic, and metabolomic data with the current genomic and transcriptomic analyses could provide a more comprehensive understanding of the molecular mechanisms driving tumour progression in LS-associated and sporadic cancers. Finally, drug development efforts should include testing pathway-specific inhibitors in preclinical models and exploring the

efficacy of combination therapies tailored to unique molecular profiles, potentially targeting different components of various pathways within each cancer subtype. Addressing these limitations and pursuing these future directions will strengthen the foundation for precision medicine approaches that improve outcomes for patients with LS-associated and sporadic cancers.

References

1. Georgiou, D., Monje-Garcia, L., Miles, T., Monahan & Neil, K., and Ryan, A.J.A.J. (2023). A Focused Clinical Review of Lynch Syndrome. *Cancer Management and Research* 15, 67–85.
2. Sinicrope, F.A. (2018). Lynch Syndrome—Associated Colorectal Cancer. *New England Journal of Medicine* 379, 764–773.
3. Ryan, N.A.J., Glaire, M.A., Blake, D., Cabrera-Dandy, M., Evans, D.G., and Crosbie, E.J. (2019). The proportion of endometrial cancers associated with Lynch syndrome: a systematic review of the literature and meta-analysis. *Genetics in Medicine* 21, 2167–2180.
4. Idos, G., and Valle, L. (2021). Lynch Syndrome. *GeneReviews* 12, 1-42.
5. Bhattacharya, P., and McHugh, T.W. (2022). Lynch Syndrome. *Encyclopedia of Gastroenterology* 4, 490–494.
6. Zhao, S., Chen, L., Zang, Y., Liu, W., Liu, S., Teng, F., Xue, F., and Wang, Y. (2022). Endometrial cancer in Lynch syndrome. *International Journal of Cancer* 150, 7–17.
7. Hause, R.J., Pritchard, C.C., Shendure, J., and Salipante, S.J. (2016). Classification and characterization of microsatellite instability across 18 cancer types. *Nature Medicine* 22, 1342–1350.
8. Cerretelli, G., Ager, A., Arends, M.J., and Frayling, I.M. (2020). Molecular pathology of Lynch syndrome. *The Journal of Pathology* 250, 518–531.
9. Jaspersion, K.W., Tuohy, T.M., Neklason, D.W., and Burt, R.W. (2010). Hereditary and Familial Colon Cancer. *Gastroenterology* 138, 2044-2058.
10. Kastrinos, F., and Stoffel, E.M. (2014). History, Genetics, and Strategies for Cancer Prevention in Lynch Syndrome. *Clinical Gastroenterology and Hepatology* 12, 715–727.
11. Boland, C.R., and Lynch, H.T. (2013). The History of Lynch Syndrome. *Familial Cancer* 12, 145-157.
12. Lynch, H.T., Lanspa, S., Smyrk, T., Boman, B., Watson, P., and Lynch, J. (1991). Hereditary nonpolyposis colorectal cancer (Lynch syndromes I & II). Genetics, pathology, natural history, and cancer control, Part I. *Cancer Genetics and Cytogenetics* 53, 143–160.
13. Boland, C.R., and Goel, A. (2010). Microsatellite Instability in Colorectal Cancer. *Gastroenterology* 138, 2073-2087.
14. Wolf, A.I., Buchanan, A.H., and Farkas, L.M. (2013). Historical review of Lynch syndrome. *Journal of Coloproctology* 33, 95–110.
15. Schmeler, K.M., and Lu, K.H. (2008). Gynecologic cancers associated with Lynch syndrome/HNPCC. *Clinical and Translational Oncology* 10, 313–317.
16. Lee, M.P. (2019). Understanding cancer through the lens of epigenetic inheritance, allele-specific gene expression, and high-throughput technology. *Frontiers in Oncology* 9, 1–8.
17. Leclerc, J., Vermaut, C., and Buisine, M.P. (2021). Diagnosis of Lynch Syndrome and Strategies to Distinguish Lynch-Related Tumors from Sporadic MSI/dMMR Tumors. *Cancers (Basel)* 13, 1–33.
18. Kolodner, R.D. (2004). DNA Mismatch Repair Defects and Cancer. *Encyclopedia of Biological Chemistry*, 675–681.
19. Marinus, M.G. (2012). DNA Mismatch Repair. *EcoSal Plus* 5, 1–62.

20. Gambini, D., Ferrero, S., and Kuhn, E. (2022). Lynch Syndrome: From Carcinogenesis to Prevention Interventions. *Cancers* *14*, 4102-4123.
21. Pećina-Šlaus, N., Kafka, A., Salamon, I., and Bukovac, A. (2020). Mismatch Repair Pathway, Genome Stability and Cancer. *Frontiers in Molecular Biosciences* *7*, 535–547.
22. Yates, M.S. (2024). Abstract IA013: Unique metabolic changes in Lynch syndrome-associated endometrial cancer. *Clinical Cancer Research* *30*, 13–13.
23. Peltomäki, P., Nyström, M., Mecklin, J.-P., and Seppälä, T.T. (2023). Lynch Syndrome Genetics and Clinical Implications. *Gastroenterology* *164*, 783-700.
24. Peltomäki, P. (2016). Update on Lynch syndrome genomics. *Familial Cancer* *15*, 385-393.
25. Zhang, S., Lloyd, R., Bowden, G., Glickman, B.W., and De Boer, J.G. (2002). Msh2 deficiency increases the mutation frequency in all parts of the mouse colon. *Environmental and Molecular Mutagenesis* *40*, 243–250.
26. Shrestha, K.S., Aska, E.M., Tuominen, M.M., and Kauppi, L. (2021). Tissue-specific reduction in MLH1 expression induces microsatellite instability in intestine of Mlh1+/- mice. *DNA Repair (Amst)* *106*, 1-10.
27. Zhao, Q., Zong, H., Zhu, P., Su, C., Tang, W., Chen, Z., and Jin, S. (2024). Crosstalk between colorectal CSCs and immune cells in tumorigenesis, and strategies for targeting colorectal CSCs. *Experimental Hematology & Oncology* *13*, 1–31.
28. Lee, J.E., and Kim, M.Y. (2022). Cancer epigenetics: Past, present and future. *Seminars in Cancer Biology* *83*, 4–14.
29. Clark, S.J., and Molloy, P.L. (2022). Early Insights into Cancer Epigenetics: Gene Promoter Hypermethylation Emerges as a Potential Biomarker for Cancer Detection. *Cancer Research* *82*, 1461–1463.
30. Fleisher, A.S., Esteller, M., Tamura, G., Rashid, A., Stine, O.C., Yin, J., Zou, T.T., Abraham, J.M., Kong, D., Nishizuka, S., et al. (2001). Hypermethylation of the hMLH1 gene promoter is associated with microsatellite instability in early human gastric neoplasia. *Oncogene* *20*, 329–335.
31. Pasanen, A., Loukovaara, M., and Bützow, R. (2020). Clinicopathological significance of deficient DNA mismatch repair and MLH1 promoter methylation in endometrioid endometrial carcinoma. *Modern Pathology* *33*, 1443–1452.
32. Peltomäki, P. (2014). Epigenetic mechanisms in the pathogenesis of Lynch syndrome. *Clinical Genetics* *85*, 403–412.
33. Liao, X., Zhu, W., Zhou, J., Li, H., Xu, X., Zhang, B., and Gao, X. (2023). Repetitive DNA sequence detection and its role in the human genome. *Communications Biology* *6*, 1–21.
34. Lanza, G., Gafà, R., Maestri, I., Santini, A., Matteuzzi, M., and Cavazzini, L. (2002). Immunohistochemical Pattern of MLH1/MSH2 Expression Is Related to Clinical and Pathological Features in Colorectal Adenocarcinomas with Microsatellite Instability. *Modern Pathology* *15*, 741–749.
35. Hendriks, Y.M.C., Jagmohan-Changur, S., Van Der Klift, H.M., Morreau, H., Van Puijenbroek, M., Tops, C., Van Os, T., Wagner, A., Ausems, M.G.F.M., Gomez, E., et al. (2006). Heterozygous Mutations in PMS2 Cause Hereditary Nonpolyposis Colorectal Carcinoma (Lynch Syndrome). *Gastroenterology* *130*, 312–322.
36. Harada, S., and Morlote, D. (2020). Molecular Pathology of Colorectal Cancer. *Advances in Anatomic Pathology* *27*, 20–26.

37. Styk, J., Pös, Z., Pös, O., Radvanszky, J., Turnova, E.H., Buglyó, G., Klimova, D., Budis, J., Repiska, V., Nagy, B., et al. (2023). Microsatellite instability assessment is instrumental for Predictive, Preventive and Personalised Medicine: status quo and outlook. *EPMA Journal* 14, 143–165.
38. Kather, J.N., Pearson, A.T., Halama, N., Jäger, D., Krause, J., Loosen, S.H., Marx, A., Boor, P., Tacke, F., Neumann, U.P., et al. (2019). Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nature Medicine* 25, 1054–1056.
39. Echle, A., Ghaffari Laleh, N., Quirke, P., Grabsch, H.I., Muti, H.S., Saldanha, O.L., Brockmoeller, S.F., van den Brandt, P.A., Hutchins, G.G.A., Richman, S.D., et al. (2022). Artificial intelligence for detection of microsatellite instability in colorectal cancer-a multicentric analysis of a pre-screening tool for clinical application. *ESMO Open* 7, 1–12.
40. Saillard, C., Dubois, R., Tchita, O., Loiseau, N., Garcia, T., Adriansen, A., Carpentier, S., Reyre, J., Enea, D., von Loga, K., et al. (2023). Validation of MSIIntuit as an AI-based pre-screening tool for MSI detection from colorectal cancer histology slides. *Nature Communications* 14, 1–11.
41. Fischer, J., Walker, L.C., Robinson, B.A., Frizelle, F.A., Church, J.M., and Eglinton, T.W. (2019). Clinical implications of the genetics of sporadic colorectal cancer. *ANZ Journal of Surgery* 89, 1224–1229.
42. Roskoski, R. (2018). Targeting oncogenic Raf protein-serine/threonine kinases in human cancers. *Pharmacological Research* 135, 239–258.
43. Lu, Y., Wajapeyee, N., Turker, M.S., and Glazer, P.M. (2014). Silencing of the DNA mismatch repair gene, MLH1, induced by hypoxic stress in a pathway dependent on the histone demethylase, LSD1. *Cell Reports* 8, 501–513.
44. Deng, G., Bell, I., Crawley, S., Gum, J., Terdiman, J.P., Allen, B.A., Truta, B., Sleisenger, M.H., and Kim, Y.S. (2004). BRAF Mutation Is Frequently Present in Sporadic Colorectal Cancer with Methylated hMLH1, but Not in Hereditary Nonpolyposis Colorectal Cancer. *Clinical Cancer Research* 10, 191–195.
45. Domingo, E., Laiho, P., Ollikainen, M., Pinto, M., Wang, L., French, A.J., Westra, J., Frebourg, T., Espín, E., Armengol, M., et al. (2004). BRAF screening as a low-cost effective strategy for simplifying HNPCC genetic testing. *Journal of Medical Genetics* 41, 664–668.
46. Ogino, S., Nosho, K., Kirkner, G.J., Kawasaki, T., Meyerhardt, J.A., Loda, M., Giovannucci, E.L., and Fuchs, C.S. (2009). CpG island methylator phenotype, microsatellite instability, BRAF mutation and clinical outcome in colon cancer. *Gut* 58, 90–96.
47. Roth, A.D., Tejpar, S., Delorenzi, M., Yan, P., Fiocca, R., Klingbiel, D., Dietrich, D., Biesmans, B., Bodoky, G., Barone, C., et al. (2010). Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60-00 trial. *Journal of Clinical Oncology* 28, 466–474.
48. French, A.J., Sargent, D.J., Burgart, L.J., Foster, N.R., Kabat, B.F., Goldberg, R., Shepherd, L., Windschitl, H.E., and Thibodeau, S.N. (2008). Prognostic Significance of Defective Mismatch Repair and BRAF-V600E in Patients with Colon Cancer. *Clinical Cancer Research* 14, 3408–3415.
49. Kakar, S., Deng, C., Sahai, V., Matsuzaki, K., Tanaka, H., Miura, S., and Kim, Y.S. (2008). Clinicopathologic characteristics, CpG island methylator phenotype, and BRAF

- mutations in microsatellite-stable colorectal cancers without chromosomal instability. *Archives of Pathology and Laboratory Medicine* **132**, 958–964.
50. Hajdúch, M., Jančík, S., Drábek, J., and Radzioch, D. (2010). Clinical relevance of KRAS in human cancers. *Journal of Biomedicine and Biotechnology* **2010**, 1-13.
 51. Tan, C., and Du, X. (2012). KRAS mutation testing in metastatic colorectal cancer. *World Journal of Gastroenterology* **18**, 5171-5180.
 52. Zhu, C., Guan, X., Zhang, X., Luan, X., Song, Z., Cheng, X., Zhang, W., and Qin, J.J. (2022). Targeting KRAS mutant cancers: from druggable therapy to drug resistance. *Molecular Cancer* **21**, 1–19.
 53. Neumann, J., Zeindl-Eberhart, E., Kirchner, T., and Jung, A. (2009). Frequency and type of KRAS mutations in routine diagnostic analysis of metastatic colorectal cancer. *Pathology Research and Practice* **205**, 858–862.
 54. Zlobec, I., Kovac, M., Erzberger, P., Molinari, F., Bihl, M.P., Ruffle, A., Foerster, A., Frattini, M., Terracciano, L., Heinimann, K., et al. (2010). Combined analysis of specific KRAS mutation, BRAF and microsatellite instability identifies prognostic subgroups of sporadic and hereditary colorectal cancer. *International Journal of Cancer* **127**, 2569–2575.
 55. Nakayama, I., Hirota, T., and Shinozaki, E. (2020). BRAF Mutation in Colorectal Cancers: From Prognostic Marker to Targetable Mutation. *Cancers (Basel)* **12**, 1–29.
 56. Yokota, T., Ura, T., Shibata, N., Takahari, D., Shitara, K., Nomura, M., Kondo, C., Mizota, A., Utsunomiya, S., Muro, K., et al. (2011). BRAF mutation is a powerful prognostic factor in advanced and recurrent colorectal cancer. *British Journal of Cancer* **104**, 856–862.
 57. Heriyanto, D.S., Yoshuantari, N., Akbariani, G., Lau, V., Hanini, H., Hidayati, Z., Arief, M.Z., Gunawan, A.N., Ridwanuloh, A.M., Kusharyoto, W., et al. (2024). High Probability of Lynch Syndrome among colorectal cancer patients in Indonesia is associated with higher occurrence of KRAS and PIK3CA mutations. *medRxiv*, 1–22.
 58. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* **17**, 405–423.
 59. Zhao, H., Ming, T., Tang, S., Ren, S., Yang, H., Liu, M., Tao, Q., and Xu, H. (2022). Wnt signaling in colorectal cancer: pathogenic role and therapeutic target. *Molecular Cancer* **21**, 1–34.
 60. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* **17**, 405–423.
 61. Palomaki, G.E., McClain, M.R., Melillo, S., Hampel, H.L., and Thibodeau, S.N. (2009). EGAPP supplementary evidence review: DNA testing strategies aimed at reducing morbidity and mortality from Lynch syndrome. *Genetics in Medicine* **11**, 42-65.
 62. Barnetson, R.A., Tenesa, A., Farrington, S.M., Nicholl, I.D., Cetnarskyj, R., Porteous, M.E., Campbell, H., and Dunlop, M.G. (2006). Identification and Survival of Carriers of Mutations in DNA Mismatch-Repair Genes in Colon Cancer. *New England Journal of Medicine* **354**, 2751–2763.

63. 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. (2004). Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. *Journal of the National Cancer Institute* 96, 261-268.
64. Syngal, S., Brand, R.E., Church, J.M., Giardiello, F.M., Hampel, H.L., and Burt, R.W. (2015). ACG clinical guideline: Genetic testing and management of hereditary gastrointestinal cancer syndromes. *The American Journal of Gastroenterology* 110, 223–262.
65. Trujillo-Rojas, M.A., Ayala-Madrigal, M. de la L., Gutiérrez-Angulo, M., González-Mercado, A., and Moreno-Ortiz, J.M. (2023). Diagnosis of patients with Lynch syndrome lacking the Amsterdam II or Bethesda criteria. *Hereditary Cancer in Clinical Practice* 21, 1–8.
66. Gupta, S., Provenzale, D., Llor, X., Halverson, A.L., Grady, W., Chung, D.C., Haraldsdottir, S., Markowitz, A.J., Slavin, T.P., Hampel, H., et al. (2019). NCCN Guidelines Insights: Genetic/Familial High-Risk Assessment: Colorectal, Version 2.2019: Featured Updates to the NCCN Guidelines. *Journal of the National Comprehensive Cancer Network* 17, 1032–1041.
67. Shyr, C., Blackford, A.L., Huang, T., Ke, J., Ouardaoui, N., Trippa, L., Syngal, S., Ukaegbu, C., Uno, H., Nafa, K., et al. (2022). A validation of models for prediction of pathogenic variants in mismatch repair genes. *Genetics in Medicine* 24, 2155–2166.
68. Pouchet, C.J., Wong, N., Chong, G., Sheehan, M.J., Schneider, G., Rosen-Sheidley, B., Foulkes, W., and Tischkowitz, M. (2009). A comparison of models used to predict MLH1, MSH2 and MSH6 mutation carriers. *Annals of Oncology* 20, 681–688.
69. Chen, S., Wang, W., Lee, S., Nafa, K., Lee, J., Romans, K., Watson, P., Gruber, S.B., Euhus, D., Kinzler, K.W., et al. (2006). Prediction of Germline Mutations and Cancer Risk in the Lynch Syndrome. *JAMA* 296, 1479–1487.
70. Goverde, A., Spaander, M.C.W., Nieboer, D., van den Ouweland, A.M.W., Dinjens, W.N.M., Dubbink, H.J., Tops, C.J., ten Broeke, S.W., Bruno, M.J., Hofstra, R.M.W., et al. (2018). Evaluation of current prediction models for Lynch syndrome: updating the PREMM5 model to identify PMS2 mutation carriers. *Familial Cancer* 17, 361-370.
71. Kastrinos, F., Uno, H., Ukaegbu, C., Alvero, C., McFarland, A., Yurgelun, M.B., Kulke, M.H., Schrag, D., Meyerhardt, J.A., Fuchs, C.S., et al. (2017). Development and Validation of the PREMM5 Model for Comprehensive Risk Assessment of Lynch Syndrome. *Journal of Clinical Oncology* 35, 2165-2172.
72. Muller, C., Matthews, L., Kupfer, S.S., and Weiss, J.M. (2019). Effective Identification of Lynch Syndrome in Gastroenterology Practice. *Current Treatment Options in Gastroenterology* 17, 666-680.
73. Mercado, R.C., Hampel, H., Kastrinos, F., Steyerberg, E., Balmana, J., Stoffel, E., Cohn, D.E., Backes, F.J., Hopper, J.L., Jenkins, M.A., et al. (2012). Performance of PREMM(1,2,6), MMRpredict, and MMRpro in detecting Lynch syndrome among endometrial cancer cases. *Genetics in Medicine* 14, 670–680.
74. Engel, C., Loeffler, M., Steinke, V., Rahner, N., Holinski-Feder, E., Dietmaier, W., Schackert, H.K., Goergens, H., Von Knebel Doeberitz, M., Goecke, T.O., et al. (2012). Risks of less common cancers in proven mutation carriers with lynch syndrome. *Journal of Clinical Oncology* 30, 4409–4415.
75. Win, A.K., Young, J.P., Lindor, N.M., Tucker, K.M., Ahnen, D.J., Young, G.P., Buchanan, D.D., Clendenning, M., Giles, G.G., Winship, I., et al. (2012). Colorectal and other

- cancer risks for carriers and noncarriers from families with a DNA mismatch repair gene mutation: A prospective cohort study. *Journal of Clinical Oncology* 30, 958–964.
76. Samadder, N.J., Smith, K.R., Wong, J., Thomas, A., Hanson, H., Boucher, K., Kopituch, C., Cannon-Albright, L.A., Burt, R.W., and Curtin, K. (2017). Cancer Risk in Families Fulfilling the Amsterdam Criteria for Lynch Syndrome. *JAMA Oncology* 3, 1697–1701.
 77. 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. (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, 15–25.
 78. Diaz-Cano, S.J. (2012). Tumor Heterogeneity: Mechanisms and Bases for a Reliable Application of Molecular Marker Design. *International Journal of Molecular Sciences* 13, 1951–2011.
 79. Latham, A., Srinivasan, P., Kemel, Y., Shia, J., Bandlamudi, C., Mandelker, D., Middha, S., Hechtman, J., Zehir, A., Dubard-Gault, M., et al. (2019). Microsatellite instability is associated with the presence of Lynch syndrome pan-cancer. *Journal of Clinical Oncology* 37, 286–295.
 80. Liccardo, R., De Rosa, M., Izzo, P., and Duraturo, F. (2017). Novel Implications in Molecular Diagnosis of Lynch Syndrome. *Gastroenterology Research and Practice* 2017, 1–12.
 81. Buttin, B.M., Powell, M.A., Mutch, D.G., Babb, S.A., Huettner, P.C., Edmonston, T.B., Herzog, T.J., Rader, J.S., Gibb, R.K., Whelan, A.J., et al. (2004). Penetrance and Expressivity of MSH6 Germline Mutations in Seven Kindreds Not Ascertained by Family History. *The American Journal of Human Genetics* 74, 1262–1269.
 82. Hagggar, F.A., and Boushey, R.P. (2009). Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors. *Clinics in Colon and Rectal Surgery* 22, 191–197.
 83. Janku, F. (2014). Tumor heterogeneity in the clinic: is it a real problem? *Therapeutic Advances in Medical Oncology* 6, 43–51.
 84. Molinari, C., Marisi, G., Passardi, A., Matteucci, L., De Maio, G., and Ulivi, P. (2018). Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? *International Journal of Molecular Sciences* 19, 1–18.
 85. Xi, Y., and Xu, P. (2021). Global colorectal cancer burden in 2020 and projections to 2040. *Translational Oncology* 14, 1–7.
 86. Brand, M., Gaylard, P., and Ramos, J. (2018). Colorectal cancer in South Africa: An assessment of disease presentation, treatment pathways and 5-year survival. *South African Medical Journal* 108, 118–122.
 87. Stigliano, V., Sanchez-Mete, L., Martayan, A., and Anti, M. (2014). Early-onset colorectal cancer: A sporadic or inherited disease? *World Journal of Gastroenterology* 20, 12420–12420.
 88. Bounous, V.E., Robba, E., Perotto, S., Pasini, B., Cont, N.T., Ricci, M.T., Ditto, A., Vitellaro, M., Raspagliesi, F., and Biglia, N. (2022). Gynecological Cancers in Lynch Syndrome: A Comparison of the Histological Features with Sporadic Cases of the General Population. *Journal of Clinical Medicine* 11, 3689–3701.
 89. Watson, P., Vasen, H.F.A., Mecklin, J.P., Bernstein, I., Aarnio, M., Järvinen, H.J., Myrhoj, T., Sunde, L., Wijnen, J.T., and Lynch, H.T. (2008). The Risk of Extra-colonic, Extra-endometrial Cancer in the Lynch Syndrome. *International journal of cancer. Journal international du cancer* 123, 444–449.

90. 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. (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, 1306–1316.
91. Fraune, C., Rosebrock, J., Simon, R., Hube-Magg, C., Makrypidi-Fraune, G., Kluth, M., Büscheck, F., Höflmayer, D., Schmalfeldt, B., Müller, V., et al. (2020). High homogeneity of MMR deficiency in ovarian cancer. *Gynecologic Oncology* 156, 669–675.
92. Smolarz, B., Zadrożna Nowak, A., and Romanowicz, H. (2022). Breast Cancer—Epidemiology, Classification, Pathogenesis and Treatment (Review of Literature). *Cancers* 14, 106–109.
93. Sheehan, M., Heald, B., Yanda, C., Kelly, E.D., Grobmyer, S., Eng, C., Kalady, M., and Pederson, H. (2020). Investigating the Link between Lynch Syndrome and Breast Cancer. *European Journal of Breast Health* 16, 106-109.
94. Harkness, E.F., Barrow, E., Newton, K., Green, K., Clancy, T., Lalloo, F., Hill, J., and Gareth Evans, D. (2015). Lynch syndrome caused by MLH1 mutations is associated with an increased risk of breast cancer: a cohort study. *Journal of Medical Genetics* 52, 553–556.
95. Therkildsen, C., Ladelund, S., Smith-Hansen, L., Lindberg, L.J., and Nilbert, M. (2017). Towards gene- and gender-based risk estimates in Lynch syndrome; age-specific incidences for 13 extra-colorectal cancer types. *British Journal of Cancer* 117, 1702–1710.
96. Da Silva, F.C., De Oliveira, L.P., Santos, É.M., Nakagawa, W.T., Aguiar Junior, S., Valentin, M.D., Rossi, B.M., and De Oliveira Ferreira, F. (2010). Frequency of extracolonic tumors in Brazilian families with Lynch syndrome: analysis of a hereditary colorectal cancer institutional registry. *Familial Cancer* 9, 563–570.
97. Couch, F.J., Shimelis, H., Hu, C., Hart, S.N., Polley, E.C., Na, J., Hallberg, E., Moore, R., Thomas, A., Lilyquist, J., et al. (2017). Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncology* 3, 1190–1196.
98. Zalevska, K., Mecklin, J.P., and Seppälä, T.T. (2023). Clinical characteristics of pancreatic and biliary tract cancers in Lynch syndrome: A retrospective analysis from the Finnish National Lynch Syndrome Research Registry. *Frontiers in Oncology* 13, 1–9.
99. Wang, H., Liu, J., Xia, G., Lei, S., Huang, X., and Huang, X. (2020). Survival of pancreatic cancer patients is negatively correlated with age at diagnosis: a population-based retrospective study. *Scientific Reports* 10, 1–9.
100. Shah, S.C., McKinley, M., Gupta, S., Peek, R.M., Martinez, M.E., and Gomez, S.L. (2020). Population-based Analysis of Differences in Gastric Cancer Incidence Among Races and Ethnicities in Individuals Age 50 Years and Older. *Gastroenterology* 159, 1705-1714.
101. Ladigan-Badura, S., Vangala, D.B., Engel, C., Bucksch, K., Hueneburg, R., Perne, C., Nattermann, J., Steinke-Lange, V., Rahner, N., Schackert, H.K., et al. (2021). Value of upper gastrointestinal endoscopy for gastric cancer surveillance in patients with Lynch syndrome. *International Journal of Cancer* 148, 106–114.
102. Boland, C.R., Yurgelun, M.B., Mraz, K.A., and Boland, P.M. (2022). Managing gastric cancer risk in lynch syndrome: controversies and recommendations. *Familial Cancer* 21, 75–78.

103. Kim, J., Braun, D., Ukaegbu, C., Dhingra, T.G., Kastrinos, F., Parmigiani, G., Syngal, S., and Yurgelun, M.B. (2020). Clinical Factors Associated With Gastric Cancer in Individuals with Lynch Syndrome. *Clinical Gastroenterology Hepatology* 18, 830–837.
104. Cho, H., Yamada, M., Sekine, S., Tanabe, N., Ushima, M., Hirata, M., Ogawa, G., Gotoh, M., Yoshida, T., Yoshikawa, T., et al. (2021). Gastric cancer is highly prevalent in Lynch syndrome patients with atrophic gastritis. *Gastric Cancer* 24, 283–291.
105. Win, A.K., Lindor, N.M., Young, J.P., MacRae, F.A., Young, G.P., Williamson, E., Parry, S., Goldblatt, J., Lipton, L., Winship, I., et al. (2012). Risks of Primary Extracolonic Cancers Following Colorectal Cancer in Lynch Syndrome. *Journal of the National Cancer Institute* 104, 1363–1372.
106. Barsouk, A., Rawla, P., Barsouk, A., and Thandra, K.C. (2019). Epidemiology of Cancers of the Small Intestine: Trends, Risk Factors, and Prevention. *Medical Sciences* 7, 1–13.
107. Park, J.G., Kim, D.W., Hong, C.W., Nam, B.H., Shin, Y.K., Hong, S.H., Kim, I.J., Lim, S.B., Aronson, M., Bisgaard, M.L., et al. (2006). Germ line mutations of mismatch repair genes in hereditary nonpolyposis colorectal cancer patients with small bowel cancer: International Society for Gastrointestinal Hereditary Tumours Collaborative Study. *Clinical cancer research* 12, 3389–3393.
108. Koornstra, J.J., Kleibeuker, J.H., and Vasen, H.F. (2008). Small-bowel cancer in Lynch syndrome: is it time for surveillance? *Lancet Oncology* 9, 901–905.
109. Harper, H.L., McKenney, J.K., Heald, B., Stephenson, A., Campbell, S.C., Plesec, T., and Magi-Galluzzi, C. (2016). Upper tract urothelial carcinomas: frequency of association with mismatch repair protein loss and lynch syndrome. *Modern Pathology* 30, 146–156.
110. Lindner, A.K., Schachtner, G., Tulchiner, G., Thurnher, M., Untergasser, G., Obrist, P., Pipp, I., Steinkohl, F., Horninger, W., Culig, Z., et al. (2021). Lynch Syndrome: Its Impact on Urothelial Carcinoma. *International Journal of Molecular Sciences* 22, 1–21.
111. Therkildsen, C., Ladelund, S., Rambech, E., Persson, A., Petersen, A., and Nilbert, M. (2015). Glioblastomas, astrocytomas and oligodendrogliomas linked to Lynch syndrome. *European Journal of Neurology* 22, 717–724.
112. Kim, H., Lim, K.Y., Park, J.W., Kang, J., Won, J.K., Lee, K., Shim, Y., Park, C.K., Kim, S.K., Choi, S.H., et al. (2021). Sporadic and Lynch syndrome-associated mismatch repair-deficient brain tumors. *Laboratory Investigation* 102, 160–171.
113. Wimmer, katharina, Kratz, C.P., Vasen, H.F.A., Caron, O., Colas, C., Entz-Werle, N., Gerdes, A.M., Goldberg, Y., Ilencikova, D., Muleris, M., et al. (2014). Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium “care for CMMRD” (C4CMMRD). *Journal of Medical Genetics* 51, 355–365.
114. Lynch, H.T., Lynch, P.M., Lanspa, S.J., Snyder, C.L., Lynch, J.F., and Boland, C.R. (2009). Review of the Lynch syndrome: history, molecular genetics, screening, differential diagnosis, and medicolegal ramifications. *Clinical Genetics* 76, 1–18.
115. 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., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* 45, 1113–1120.
116. Cerami, E., Gao, J., Dogrusoz, U., Gross, B.E., Sumer, S.O., Aksoy, B.A., Jacobsen, A., Byrne, C.J., Heuer, M.L., Larsson, E., et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discovery* 2, 401–404.

117. Li, H., Sun, L., Zhuang, Y., Tian, C., Yan, F., Zhang, Z., Hu, Y., and Liu, P. (2022). Molecular mechanisms and differences in lynch syndrome developing into colorectal cancer and endometrial cancer based on gene expression, methylation, and mutation analysis. *Cancer Causes & Control* 33, 489-501.
118. Liu, Z., Yang, C., Li, X., Luo, W., Roy, B., Xiong, T., Zhang, X., Yang, H., Wang, J., Ye, Z., et al. (2018). The landscape of somatic mutation in sporadic Chinese colorectal cancer. *Oncotarget* 9, 27412-27422.
119. Hoadley, K.A., Yau, C., Hinoue, T., Wolf, D.M., Lazar, A.J., Drill, E., Shen, R., Taylor, A.M., Cherniack, A.D., Thorsson, V., et al. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 173, 291-304.
120. Gao, J., Aksoy, B.A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S.O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signaling* 6, 1-34.
121. de Bruijn, I., Kundra, R., Mastrogiacomio, B., Tran, T.N., Sikina, L., Mazor, T., Li, X., Ochoa, A., Zhao, G., Lai, B., et al. (2023). Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Research* 83, 3861–3867.
122. Elreedy, D., and Atiya, A.F. (2019). A Comprehensive Analysis of Synthetic Minority Oversampling Technique (SMOTE) for handling class imbalance. *Information Sciences* 505, 32–64.
123. De Jay, N., Papillon-Cavanagh, S., Olsen, C., El-Hachem, N., Bontempi, G., and Haibe-Kains, B. (2013). mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics* 29, 2365–2368.
124. Ikotun, A.M., Ezugwu, A.E., Abualigah, L., Abuhaija, B., and Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences* 622, 178–210.
125. Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 11, 1–9.
126. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* 14, 128–128.
127. Kutmon, M., van Iersel, M.P., Bohler, A., Kelder, T., Nunes, N., Pico, A.R., and Evelo, C.T. (2015). PathVisio 3: An Extendable Pathway Analysis Toolbox. *PLoS Computational Biology* 11, 1-13.
128. Clarke, D.J.B., Kuleshov, M. V., Schilder, B.M., Torre, D., Duffy, M.E., Keenan, A.B., Lachmann, A., Feldmann, A.S., Gundersen, G.W., Silverstein, M.C., et al. (2018). eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Research* 46, 171–179.
129. Chakravarty, D., Gao, J., Phillips, S., Kundra, R., Zhang, H., Wang, J., Rudolph, J.E., Yaeger, R., Soumerai, T., Nissan, M.H., et al. (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO Precision Oncology* 17, 1–16.
130. Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research* 41, 955–961.
131. Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R., et al. (2004). The COSMIC (Catalogue of

- Somatic Mutations in Cancer) database and website. *British Journal of Cancer* **91**, 355–358.
132. Taieb, J., Svrcek, M., Cohen, R., Basile, D., Tougeron, D., and Phelip, J.M. (2022). Deficient mismatch repair/microsatellite unstable colorectal cancer: Diagnosis, prognosis and treatment. *European Journal of Cancer* **175**, 136–157.
 133. Büttner, R., Longshore, J.W., López-Ríos, F., Merkelbach-Bruse, S., Normanno, N., Rouleau, E., and Penault-Llorca, F. (2019). Implementing TMB measurement in clinical practice: considerations on assay requirements. *ESMO Open* **4**, 1-12.
 134. Aghabozorgi, A.S., Bahreyni, A., Soleimani, A., Bahrami, A., Khazaei, M., Ferns, G.A., Avan, A., and Hassanian, S.M. (2019). Role of adenomatous polyposis coli (APC) gene mutations in the pathogenesis of colorectal cancer; current status and perspectives. *Biochimie* **157**, 64–71.
 135. Liu, R., Duan, Q., and Zhang, Q. (2021). Analysis of LRP1B as a potential biomarker for colorectal cancer immunotherapy. *Journal of Clinical Oncology* **39**, 1–8.
 136. Philpott, C., Tovell, H., Frayling, I.M., Cooper, D.N., and Upadhyaya, M. (2017). The NF1 somatic mutational landscape in sporadic human cancers. *Human Genomics* **11**, 1-19.
 137. Rachmawati, M., Yulianti, H., Hernowo, B.S., Suryanti, S., Wijaya, I., Rahadiani, N., Heriyanto, D.S., and Irianiwati, I. (2019). The Correlation of KRAS Gene Expression and P53 Immunoexpression in Colorectal Adenocarcinoma. *Open Access Macedonian Journal of Medical Sciences* **7**, 1940-1945.
 138. He, Y., Sun, M.M., Zhang, G.G., Yang, J., Chen, K.S., Xu, W.W., and Li, B. (2021). Targeting PI3K/Akt signal transduction for cancer therapy. *Signal Transduction and Target Therapy* **6**, 1-17.
 139. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research* **45**, 353–361.
 140. Monteith, G.R., Prevarskaya, N., and Roberts-Thomson, S.J. (2017). The calcium-cancer signalling nexus. *Nature Reviews Cancer* **17**, 367–381.
 141. Wang, W., Yu, S., Huang, S., Deng, R., Ding, Y., Wu, Y., Li, X., Wang, A., Wang, S., Chen, W., et al. (2019). A complex role for calcium signaling in colorectal cancer development and progression. *Molecular Cancer Research* **17**, 2145–2153.
 142. Bolger, G.B. (2022). The cAMP-signaling cancers: Clinically-divergent disorders with a common central pathway. *Front Endocrinology* **13**, 1-16.
 143. Razi, S., Baradaran Noveiry, B., Keshavarz-Fathi, M., and Rezaei, N. (2019). IL-17 and colorectal cancer: From carcinogenesis to treatment. *Cytokine* **116**, 7–12.
 144. Wang, K., Kim, M.K., DiCaro, G., Wong, J., Shalapour, S., Wan, J., Zhang, W., Zhong, Z., Sanchez-Lopez, E., Wu, L.W., et al. (2014). Interleukin-17 receptor A signaling in transformed enterocytes promotes early colorectal tumorigenesis. *Immunity* **41**, 1052-1063.
 145. Liu, J., Xiao, Q., Xiao, J., Niu, C., Li, Y., Zhang, X., Zhou, Z., Shu, G., and Yin, G. (2022). Wnt/ β -catenin signalling: function, biological mechanisms, and therapeutic opportunities. *Signal Transduction and Targeted Therapy* **7**, 1–23.
 146. Guo, L., and Zhang, C. (2023). SUZ12 promotes the malignant behavior of gastric cancer cells by CDKs. *Journal of Biochemical and Molecular Toxicology* **37**, 1–9.
 147. Svensson, C., Ceder, J., Iglesias-Gato, D., Chuan, Y.C., Pang, S.T., Bjartell, A., Martinez, R.M., Bott, L., Helczynski, L., Ulmert, D., et al. (2014). REST mediates androgen

- receptor actions on gene repression and predicts early recurrence of prostate cancer. *Nucleic Acids Research* 42, 999-1015.
148. Erokhin, M., Chetverina, O., Györfy, B., Tatarskiy, V. V., Mogila, V., Shtil, A.A., Roninson, I.B., Moreaux, J., Georgiev, P., Cavalli, G., et al. (2021). Clinical correlations of polycomb repressive complex 2 in different tumor types. *Cancers (Basel)* 13, 1-26.
 149. Bolivar, A.M., Duzagac, F., Deng, N., Reyes-Urbe, L., Chang, K., Wu, W., Bowen, C.M., Taggart, M.W., Thirumurthi, S., Lynch, P.M., et al. (2024). Genomic Landscape of Lynch Syndrome Colorectal Neoplasia Identifies Shared Mutated Neoantigens for Immunoprevention. *Gastroenterology* 166, 787-801.
 150. Galassi, C., Chan, T.A., Vitale, I., and Galluzzi, L. (2024). The hallmarks of cancer immune evasion. *Cancer Cell* 42, 1825–1863.
 151. Wang, L., Li, X., Zhang, L., and Gao, Q. (2017). Improved anticancer drug response prediction in cell lines using matrix factorization with similarity regularization. *BMC Cancer* 17, 1-12.
 152. Chen, E.Y., Xu, H., Gordonov, S., Lim, M.P., Perkins, M.H., and Ma'ayan, A. (2012). Expression2Kinases: mRNA profiling linked to multiple upstream regulatory layers. *Bioinformatics* 28, 105–111.
 153. Gonzalez, H., Hagerling, C., and Werb, Z. (2018). Roles of the immune system in cancer: from tumor initiation to metastatic progression. *Genes and Development* 32, 1267-1284.
 154. Langley, R.R., and Fidler, I.J. (2011). The seed and soil hypothesis revisited - the role of tumor-stroma interactions in metastasis to different organs. *International journal of cancer. Journal international du cancer* 128, 2527-2535.
 155. Marques-de-Sá, I., Castro, R., Pita, I., Dinis-Ribeiro, M., and Brandão, C. (2020). Cancer-risk by family history and mismatch-repair mutation in Lynch syndrome. *Scand Journal of Gastroenterology* 55, 701–705.
 156. Akbar, F., Siddiqui, Z., Waheed, M.T., Ehsan, L., Minhas, K., Sattar, A., and Kirmani, S. (2022). eP039: Does Lynch syndrome cause predisposition to breast cancer? Experience from a hereditary breast cancer clinic in Pakistan. *Genetics in Medicine* 24, 26–27.
 157. Sgo. (2014). Screening for Lynch Syndrome in Endometrial Cancer (Society of Gynecologic Oncology).
 158. Liu, Y., Chen, L., Zhang, S., Shu, Y., Qi, Q., Zhu, M., Peng, Y., and Ling, Y. (2020). Somatic mutations in genes associated with mismatch repair predict survival in patients with metastatic cancer receiving immune checkpoint inhibitors. *Oncology Letters* 20, 1-7.
 159. Liao, X., Morikawa, T., Lochhead, P., Imamura, Y., Kuchiba, A., Yamauchi, M., Nosh, K., Qian, Z.R., Nishihara, R., Meyerhardt, J.A., et al. (2012). Prognostic role of PIK3CA mutation in colorectal cancer: Cohort study and literature review. *Clinical Cancer Research* 18, 2257–2268.
 160. Velho, S., Oliveira, C., Ferreira, A., Ferreira, A.C., Suriano, G., Schwartz, S., Duval, A., Carneiro, F., Machado, J.C., Hamelin, R., et al. (2005). The prevalence of PIK3CA mutations in gastric and colon cancer. *European Journal of Cancer* 41, 1649–1654.
 161. Russo, A., Bazan, V., Iacopetta, B., Kerr, D., Soussi, T., and Gebbia, N. (2005). The TP53 colorectal cancer international collaborative study on the prognostic and predictive significance of p53 mutation: Influence of tumor site, type of mutation, and adjuvant treatment. *Journal of Clinical Oncology* 23, 7518–7528.

162. Moustardas, P., Aberdam, D., and Lagali, N. (2023). MAPK Pathways in Ocular Pathophysiology: Potential Therapeutic Drugs and Challenges. *Cells* 12, 1-27.
163. Deng, Z., Fan, T., Xiao, C., Tian, H., Zheng, Y., Li, C., and He, J. (2024). TGF- β signaling in health, disease and therapeutics. *Signal Transduction and Targeted Therapy* 9, 1–40.
164. Hubbard, S.R., and Miller, W.T. (2007). Receptor tyrosine kinases: mechanisms of activation and signaling. *Current Opinion in Cell Biology* 19, 117-123.
165. Lee, M.K., Pardoux, C., Hall, M.C., Lee, P.S., Warburton, D., Qing, J., Smith, S.M., and Derynck, R. (2007). TGF- β activates Erk MAP kinase signalling through direct phosphorylation of ShcA. *EMBO Journal* 26, 3957–3967.
166. Zhang, Y., Alexander, P.B., and Wang, X.F. (2017). TGF- β Family Signaling in the Control of Cell Proliferation and Survival. *Cold Spring Harbor Perspectives in Biology* 9, 1-22.
167. Sun, P., Shen, Y., Wang, T., He, Y., Zhang, Y., Tian, W., Yang, B., and Hu, Y. (2021). Distinct clinical and genetic mutation characteristics in sporadic and Lynch syndrome-associated endometrial cancer in a Chinese population. *Cancer Epidemiology* 73, 1-7.
168. Møller, P., Haupt, S., Ahadova, A., Kloor, M., Sampson, J.R., Sunde, L., Seppälä, T., Burn, J., Bernstein, I., Capella, G., et al. (2024). Incidences of colorectal adenomas and cancers under colonoscopy surveillance suggest an accelerated “Big Bang” pathway to CRC in three of the four Lynch syndromes. *Hereditary Cancer in Clinical Practice* 22, 1-12.
169. Momeni-Boroujeni, A., Vanderbilt, C., Yousefi, E., Abu-Rustum, N.R., Aghajanian, C., Soslow, R.A., Ellenson, L.H., Weigelt, B., and Murali, R. (2023). Landscape of Chromatin Remodeling Gene Alterations in Endometrial Carcinoma. *Gynaecological Oncology* 172, 54-64.
170. Danielsen, S.A., Eide, P.W., Nesbakken, A., Guren, T., Leithe, E., and Lothe, R.A. (2015). Portrait of the PI3K/AKT pathway in colorectal cancer. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1855, 104–121.
171. Fernández, M.L., DiMattia, G.E., Dawson, A., Bamford, S., Anderson, S., Hennessy, B.T., Anglesio, M.S., Shepherd, T.G., Salamanca, C., Hoenisch, J., et al. (2016). Differences in MEK inhibitor efficacy in molecularly characterized low-grade serous ovarian cancer cell lines. *American Journal of Cancer Research* 6, 2235-2251.
172. Nathan, M.R., and Schmid, P. (2017). A Review of Fulvestrant in Breast Cancer. *Oncological Therapy* 5, 17-29.
173. Wang, F., Travins, J., DeLaBarre, B., Penard-Lacronique, V., Schalm, S., Hansen, E., Straley, K., Kernysky, A., Liu, W., Gliser, C., et al. (2013). Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* 340, 622–626.
174. Park, M., Lee, B.I., Choi, J., Park, Y., Park, S.J., Lim, J.H., Lee, J., and Shin, Y.G. (2022). Quantitative Analysis of Daporinad (FK866) and Its In Vitro and In Vivo Metabolite Identification Using Liquid Chromatography-Quadrupole-Time-of-Flight Mass Spectrometry. *Molecules* 27, 1-16.
175. Wan, R., Feng, J., and Tang, L. (2021). Consequences of Mutations and Abnormal Expression of SMAD4 in Tumors and T Cells. *OncoTargets and Therapy* 14, 2531-2040.
176. Peddareddigari, V.G., Wang, D., and Dubois, R.N. (2010). The Tumor Microenvironment in Colorectal Carcinogenesis. *Cancer Microenvironment* 3, 149-166.

177. Stefani, C., Miricescu, D., Stanescu-Spinu, I.I., Nica, R.I., Greabu, M., Totan, A.R., and Jinga, M. (2021). Growth Factors, PI3K/AKT/mTOR and MAPK Signaling Pathways in Colorectal Cancer Pathogenesis: Where Are We Now? *International Journal of Molecular Sciences* *22*, 1-24.
178. Miki, Y. (2020). New Insights into Breast and Endometrial Cancers. *Cancers (Basel)* *12*, 1–6.
179. Band, A.M., and Laiho, M. (2011). Crosstalk of TGF- β and estrogen receptor signaling in breast cancer. *Journal of Mammary Gland Biology and Neoplasia* *16*, 109–115.
180. Kuru-Yaşar, R., and Üstün-Aytekin, Ö. (2024). The Crucial Roles of Diet, Microbiota, and Postbiotics in Colorectal Cancer. *Current Nutrition Reports* *13*, 126–151.
181. Nguyen, D.T., Mathias, S., Bologa, C., Brunak, S., Fernandez, N., Gaulton, A., Hersey, A., Holmes, J., Jensen, L.J., Karlsson, A., et al. (2017). Pharos: Collating protein information to shed light on the druggable genome. *Nucleic Acids Research* *45*, 995–1002.

Appendices

Appendix A: Supplementary Figures

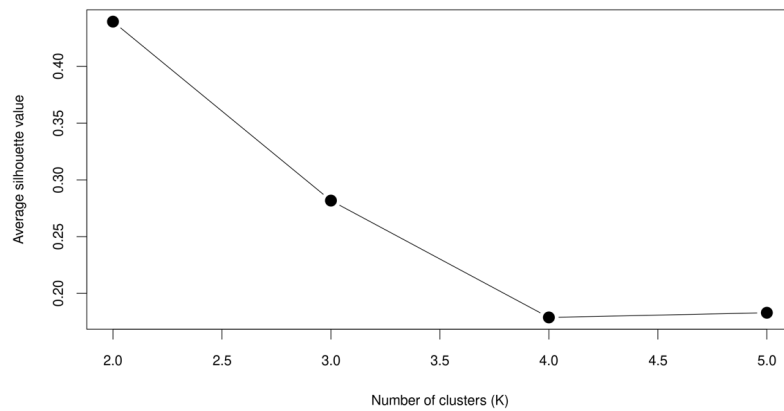


Figure S1. Determination of Optimal Clustering and Visualisation of Subtypes in Colorectal Cancer (CRC) Samples. Silhouette plot used to determine the optimal number of clusters (K) based on mRNA expression data. The highest silhouette score indicated the optimal K (K=2).

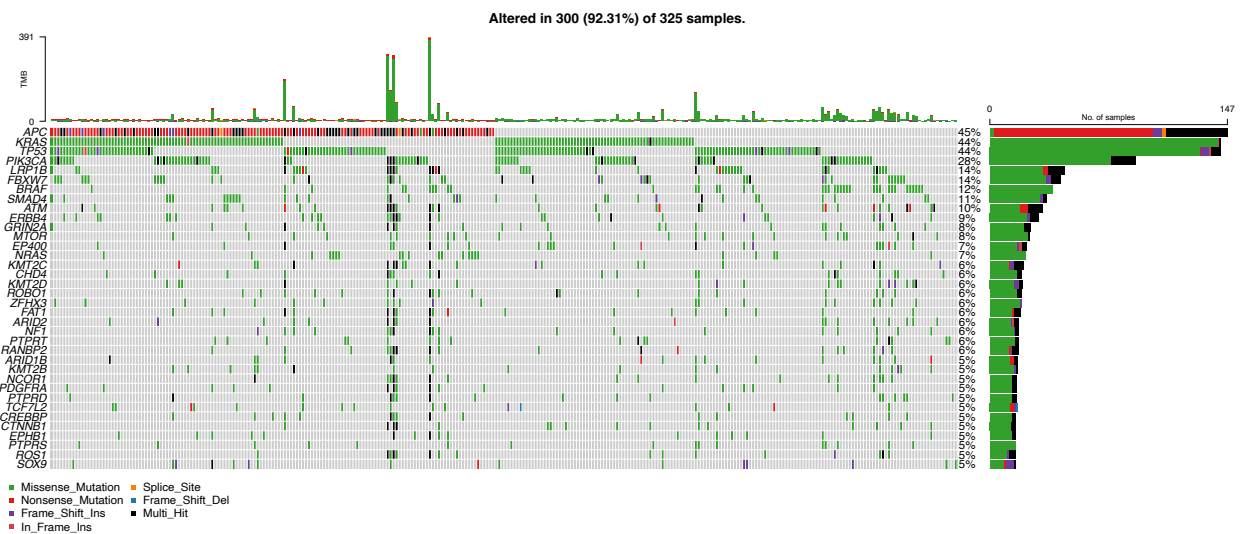


Figure S2. An oncoprint of the top pathogenic mutated tumour suppressor and oncogenes across all colorectal cancer samples.

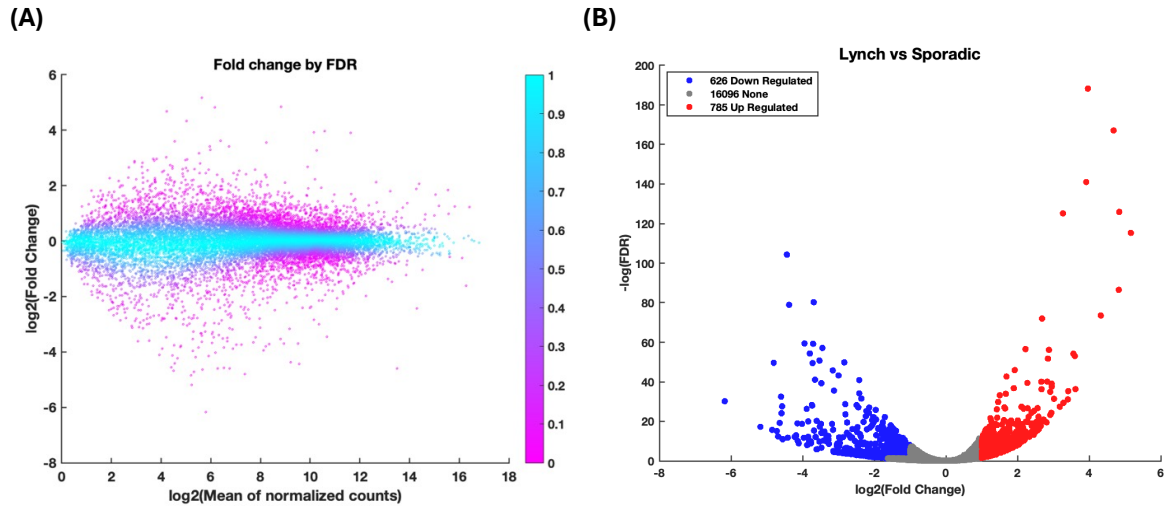


Figure S3. Differential Gene Expression Analysis of Likely Lynch Syndrome (LLS) and Sporadic (LSp) Colorectal Cancer (CRC) Subtypes (A) Fold changes of differentially expressed genes, where a positive log2 fold change (log2FC) indicates upregulated in LLS-CRC tumours, and a negative log2FC indicates upregulation in LSp-CRC tumours (B) A volcano plot visualising the differentially expressed genes between LLS and LSp CRC samples. Blue dots represent downregulated genes (upregulated in LSp), red dots indicate upregulated genes (in LLS), and grey dots represent nonsignificant changes.

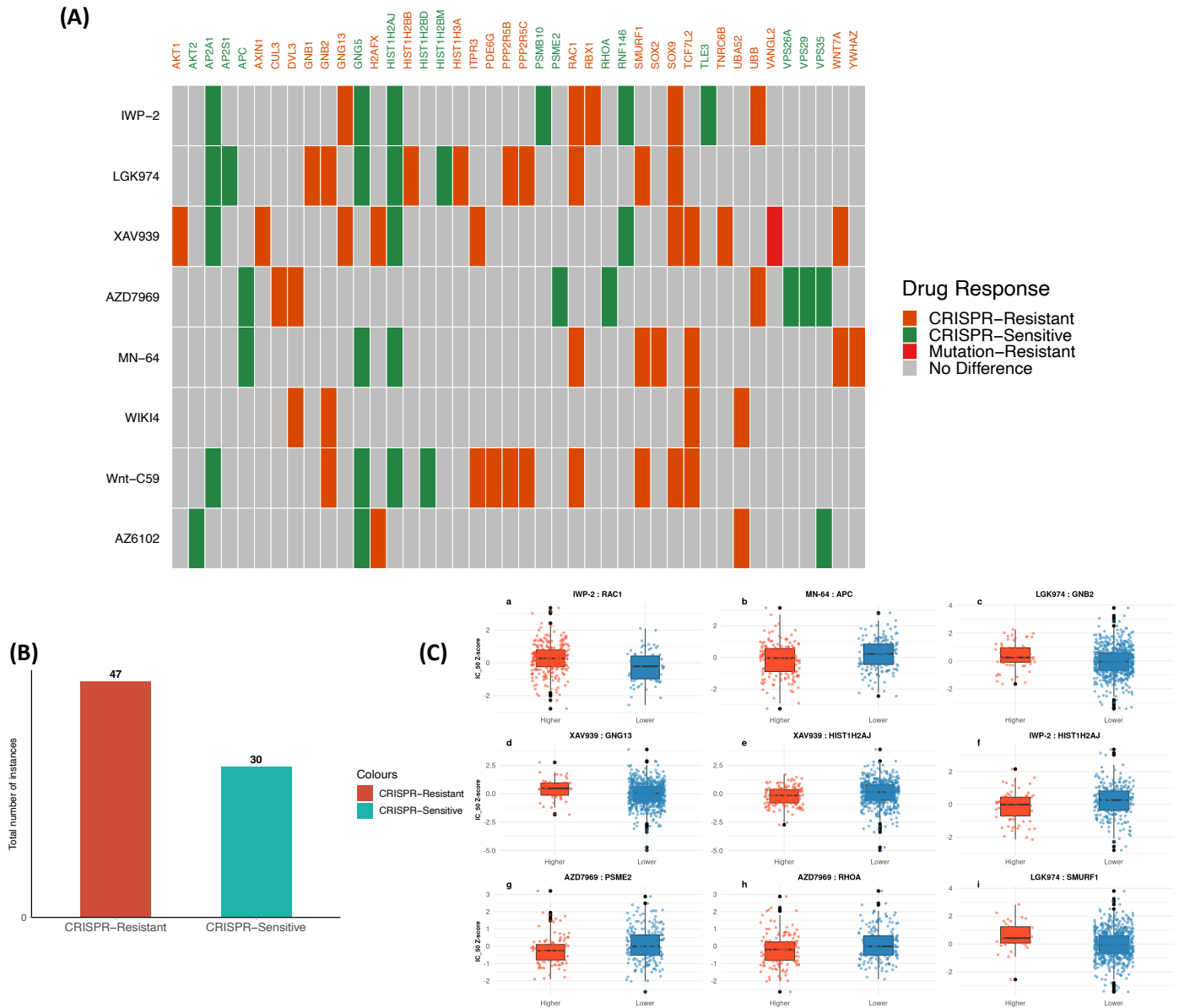


Figure S4. Relationship between Achille gene dependence scores or mutations of Wnt pathway genes and the Responses of the cell line to Wnt pathway inhibitors. (A) A clustered heatmap showing the relationship between gene dependencies or mutations and drug responses across cancer cell lines. The marks on the heatmap are coloured based on how a high dependence on, or mutations in, the gene along each column affect the efficacy of the drug given along each row: with orange for gene dependence denoting significantly increased resistance, green (for gene dependence) and red (for gene mutations) denoting significantly (10% false discovery rate) reduced sensitivity, grey for no statistically significant difference between cell line with a higher and lower dependence on the gene. The gene names (column labels) are coloured based on the overall calculated effect that high dependence on the gene has on the efficacy of the drug given along rows. Green, all the cell lines are significantly more sensitive to all the Wnt pathway inhibitors, orange, all the cell lines are significantly more resistant to all the Wnt pathway inhibitors. **(B)** A bar graph showing the total number of instances for the resistant (red) and sensitive (blue) CRISPR gene dependencies. **(C)** Comparison of the dose-response profiles between the higher and lower dependency groups of the GDSC cancer cell lines for the top nine inhibitors. The boxplots show the logarithm-transformed mean IC₅₀ values of the cancer cell lines. On each box, the central black line indicates the median, and the bottom edge represents the 25th percentiles, while the top edge of the box represents the 75th percentiles. The whiskers extend to the most extreme data points not considering outliers, which are plotted individually in bold, black circles.

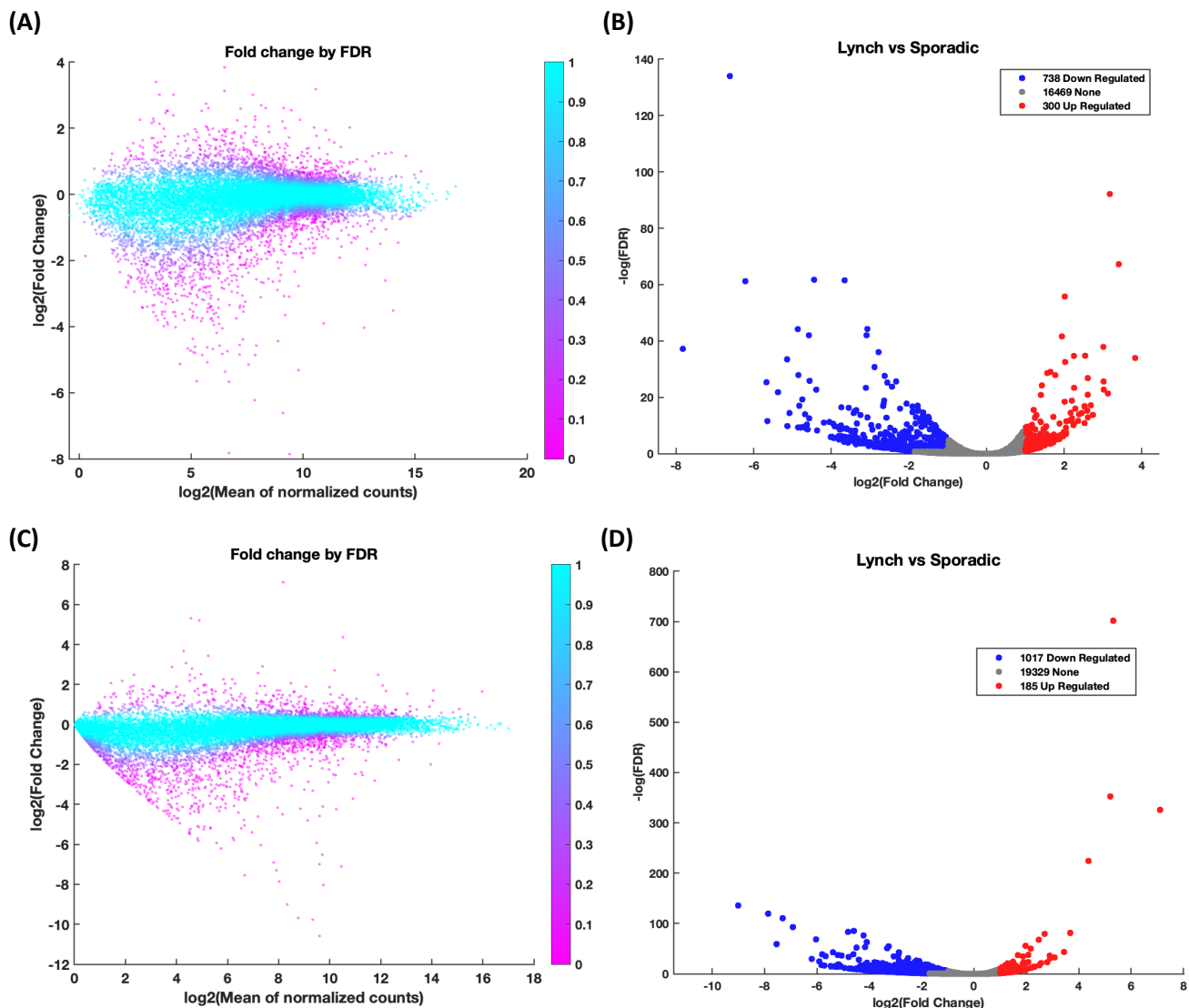


Figure S5. Differential Gene Expression Analysis of Likely Lynch Syndrome (LLS) and Sporadic (LSp) Endometrial (EC) and Breast Cancer (BC) Subtypes (A) EC fold changes of differentially expressed genes, where a positive $\log_2$ fold change ($\log_2\text{FC}$) indicates upregulated in LLS-EC tumours, and a negative $\log_2\text{FC}$ indicates upregulation in LSp-EC tumours (B) EC volcano plot visualising the differentially expressed genes between LLS and LSp EC samples. Blue dots represent downregulated genes (upregulated in LSp), red dots indicate upregulated genes (in LLS), and grey dots represent nonsignificant changes. (C) BC fold changes of differentially expressed genes, where a positive $\log_2$ fold change ($\log_2\text{FC}$) indicates upregulated in LLS-BC tumours, and a negative $\log_2\text{FC}$ indicates upregulation in LSp-BC tumours (B) BC volcano plot visualising the differentially expressed genes between LLS and LSp BC samples. Blue dots represent downregulated genes (upregulated in LSp), red dots indicate upregulated genes (in LLS), and grey dots represent nonsignificant changes.

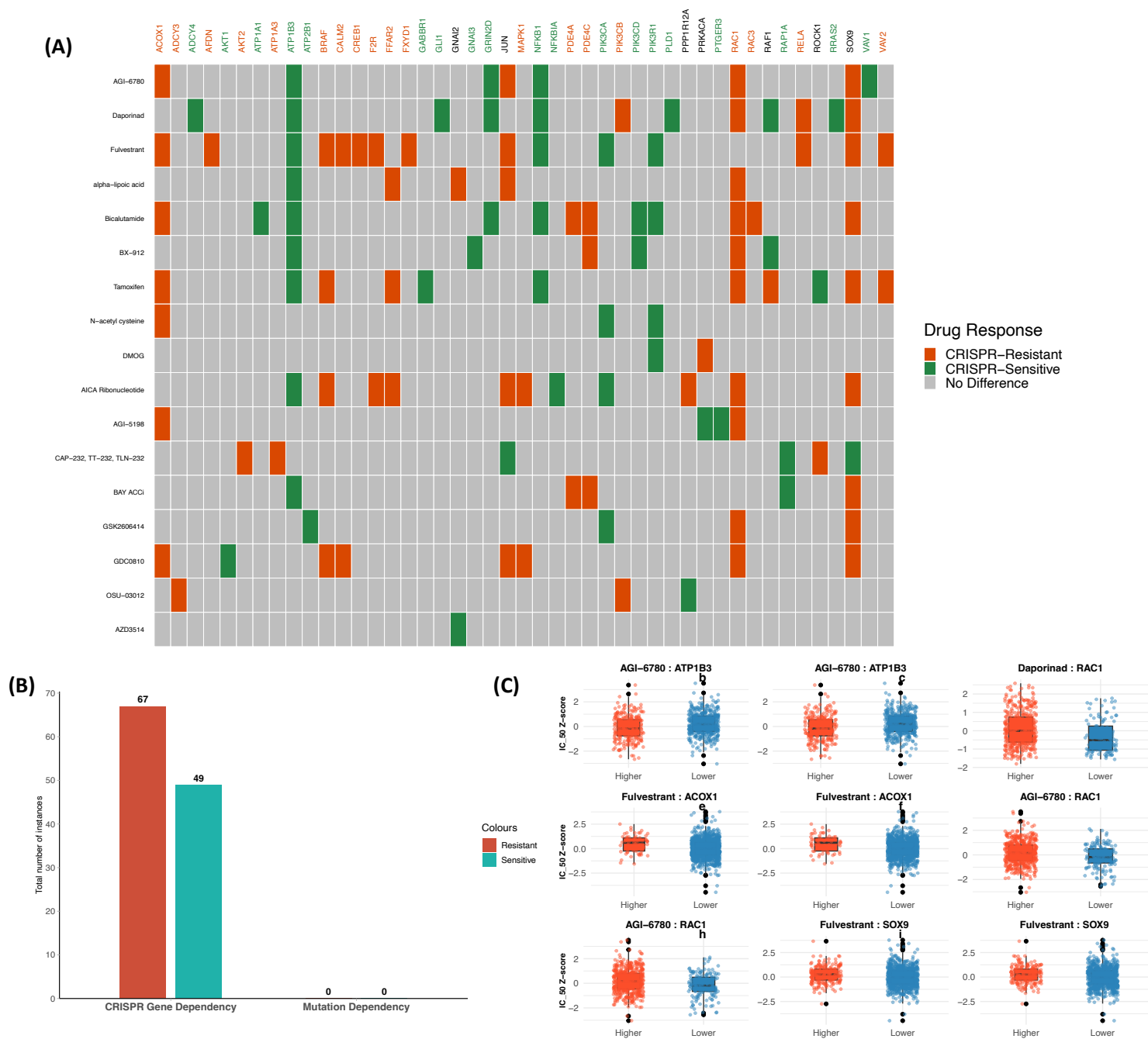


Figure S6. Relationship between Achilles gene dependence scores or mutations of cAMP-related pathway genes and the Responses of the cell line to cAMP-related pathway inhibitors. (A) A clustered heatmap showing the relationship between gene dependencies or mutations and drug responses across cancer cell lines. The marks on the heatmap are coloured based on how a high dependence on, or mutations in, the gene along each column affect the efficacy of the drug given along each row: with orange for gene dependence denoting significantly increased resistance, green (for gene dependence) and red (for gene mutations) denoting significantly (10% false discovery rate) reduced sensitivity, grey for no statistically significant difference between cell line with a higher and lower dependence on the gene. The gene names (column labels) are coloured based on the overall calculated effect that high dependence on the gene has on the efficacy of the drug given along rows. Green, all the cell lines are significantly more sensitive to all the Wnt pathway inhibitors, orange, all the cell lines are significantly more resistant to all the Wnt pathway inhibitors. **(B)** A bar graph showing the total number of instances for the resistant (red) and sensitive (blue) CRISPR gene dependencies. **(C)** Comparison of the dose-response profiles between the higher and lower dependency groups of the GDSC cancer cell lines for the top nine inhibitors. The boxplots show the logarithm-transformed mean IC50 values of the cancer cell lines.

On each box, the central black line indicates the median, and the bottom edge represents the 25th percentiles, while the top edge of the box represents the 75th percentiles. The whiskers extend to the most extreme data points not considering outliers, which are plotted individually in bold, black circles.

Appendix B: Supplementary File Description

Supplementary File 1: Pathogenic somatic mutations across all likely Lynch Syndrome and likely sporadic colorectal cancer subtypes. The dataset provides detailed information on each mutation, including the gene name, type of mutation and chromosomal location.

Supplementary File 2: The results of Fisher's exact test used to identify tumour suppressor genes and oncogenes with significantly different somatic mutation frequencies between colorectal cancer subtypes. All genes were significantly enriched in likely Lynch Syndrome (LLS) samples. Genes present in approximately one-third or more of LLS cases were classified as potential cancer drivers.

Supplementary File 3: Results of the differential gene expression analysis conducted to compare gene expression profiles between likely Lynch Syndrome (LLS) and likely sporadic (LSp) colorectal cancer (CRC) subtypes. LLS-CRC Upregulated Genes: list of upregulated genes in the LLS-CRC samples compared to the LSp-CRC samples. LSp-CRC Upregulated Genes: list of upregulated genes in the LSp-CRC samples compared to the LLS-CRC samples.

Supplementary File 4: Results of the enrichment analyses of the KEGG pathway for the colorectal cancer subtypes obtained using the differentially expressed genes.

Supplementary File 5: Results of the enrichment analyses for the GEO validation dataset of the KEGG pathway results for the subtypes of colorectal cancer obtained using the differentially expressed genes.

Supplementary File 6: Results of the analyses exploring the relationship between pathway dependency and drug sensitivity for PI3K-Akt signalling inhibitors in likely Lynch Syndrome colorectal cancer.

Supplementary File 7: Results of the analyses exploring the relationship between pathway dependency and drug sensitivity for Wnt signalling inhibitors in likely sporadic colorectal cancer.

Supplementary File 8: The results of the Fisher's exact tests used to identify tumour suppressor genes and oncogenes with significantly different somatic mutation frequencies between endometrial and breast cancer subtypes. All genes were significantly enriched in the likely Lynch Syndrome (LLS) subtypes. Genes present in approximately one-third or more of LLS cases were classified as potential cancer drivers.

Supplementary File 9: Results of the differential gene expression (DGE) analysis conducted to compare gene expression profiles between likely Lynch Syndrome (LLS) and likely sporadic (LSp) endometrial cancer (EC) subtypes. Included are the results of the DGE analysis conducted between the LLS and LSp breast cancer (BC) subtypes. LLS-EC Upregulated Genes: list of upregulated genes in the LLS-EC samples compared to the LSp-EC samples. LSp-EC Upregulated Genes: list of upregulated genes in the LSp- EC samples compared to the LLS- EC samples. LLS-BC Upregulated Genes: list of upregulated genes in the LLS-BC samples compared to the LSp-BC samples. LSp-BC Upregulated Genes: list of upregulated genes in the LSp-BC samples compared to the LLS-BC samples.

Supplementary File 10: Results of the enrichment analyses of the KEGG pathway results for the subtypes of endometrial and breast cancer obtained using the differentially expressed genes.

Supplementary File 11: Results of the analyses exploring the relationship between pathway dependency and drug sensitivity for TGF-beta-related signalling inhibitors in likely Lynch Syndrome endometrial and breast cancer.

Supplementary File 12: Results of the analyses exploring the relationship between pathway dependency and drug sensitivity for cAMP-related signalling inhibitors in likely sporadic endometrial and breast cancer.

Appendix C

Table C1. Key Resources used in this Thesis

Resource	Paper	Source
Software and Algorithms		
R	N/A	https://www.r-project.org
Python	N/A	https://www.python.org
MATLAB	N/A	https://www.mathworks.com/products/matlab.html
PathVisio3	127	https://www.pathvisio.org/
Expression2Kinases	128	http://www.maayanlab.net/X2K/
Enrichr	115	http://amp.pharm.mssm.edu/Enrichr/
Franklin	N/A	http://franklin.genoox.com
Big Data and Biological Knowledge Resources		
cBioPortal	120,121	https://www.cbioportal.org/
The Cancer Genome Atlas Data Portal	115	https://portal.gdc.cancer.gov/
Genomics of Drug Sensitivity in Cancer Database	130	https://www.cancerrxgene.org/
KEGG Pathways Database	139	https://www.genome.jp/kegg/pathway.html
Cancer Cell Line Encyclopaedia	151	https://portals.broadinstitute.org/ccle
Pharos Knowledgebase	181	https://pharos.nih.gov/
OncoKB	129	https://www.oncokb.org