



HAL
open science

Characterization of the Immune Microenvironment in Colorectal Precancerous Lesions

Erwan Morgand

► **To cite this version:**

Erwan Morgand. Characterization of the Immune Microenvironment in Colorectal Precancerous Lesions. Santé publique et épidémiologie. Sorbonne Université, 2022. English. NNT : 2022SORUS206 . tel-04124400

HAL Id: tel-04124400

<https://theses.hal.science/tel-04124400>

Submitted on 10 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



SORBONNE UNIVERSITÉ

École doctorale **Physiologie, Physiopathologie et Thérapeutique**

Centre de Recherche des Cordeliers - Equipe Integrative Cancer Immunology

Ligue Nationale contre le Cancer

Characterization of the Immune Microenvironment in Colorectal Precancerous Lesions

Par Erwan Morgand

Thèse de Doctorat d'Immunologie

Dirigée par Jérôme GALON

Présentée et soutenue publiquement le 9 juin 2022

Devant un jury composé de :

Stéphanie GRAFF-DUBOIS	Sorbonne Université	Présidente
Daniel OLIVE	Aix Marseille Université	Rapporteur
Pierre LAURENT-PUIG	Université Paris Cité	Rapporteur
Armelle BLONDEL	Université Paris Cité	Examinatrice
Michel DUCREUX	Université Paris Saclay	Examinateur
Jérôme GALON	Sorbonne Université	Directeur

Acknowledgement

Abstract

Title : Characterization of the Immune Microenvironment in Colorectal Precancerous Lesions

Most solid cancers develop from benign lesions, called polyps in colorectal cancer (CRC). Polyps can emerge from two main pathways which are associated with very different mutational and molecular profiles and will both lead to cancer. Among average-risk individuals, some patients will still develop an abnormally high number of polyps. This high polyp rate has been described as a risk factor for polyp or CRC recurrence. Identifying parameters associated with these high-risk patients could lead to a better understanding of the mechanisms underlying this high rate of polyp development and implement a more appropriate follow-up of the patient.

Among the factors shaping the carcinogenesis process, the immune system present in the tumor microenvironment (TME) was shown to be associated with the clinical outcome of cancer patients, in terms of survival and therapeutic responses. Characterization of immune TME parameters such as the density of immune cell types, their localization, activation status, and potential organization into tertiary lymphoid structures (TLSs) led to a better understanding of mechanisms regulating interactions between the immune and tumor compartment in the TME.

Dynamics between the immune response and cancer can be resumed in three phases: Elimination, where immune cells will be able to recognize and eliminate malignant cells, Equilibrium, where tumor growth will be nullified by the immune response, and Escape, where tumor cells will be able to escape immune response, leading to tumor growth and carcinogenesis development.

As cancers are usually detected and surgically removed at advanced stages, immune TME in precancerous lesions is understudied compared to invasive stages. However, several studies aimed to fill this gap, demonstrating in various types of cancer (including CRC) that even the immune system is able to recognize and eliminate tumor cells even at the earlier stages of carcinogenesis, and that mechanisms leading to immune escape are also present in premalignant states.

We took advantage of a cohort of 26 average-risk patients (and a total of 131 premalignant lesions) who presented a heterogeneous number and type of polyps to associate immune TME parameters with the polyp development rate. Performing multiplex immunohistochemistry, mRNA sequencing, and whole exome sequencing, we developed an integrative approach, demonstrating strong differences in immune profile between the two premalignant colorectal pathways. We then showed that patients with a low polyp development rate were associated with a higher presence of mature TLS. Patients with a low polyp development rate also presented higher levels of immune infiltrate in general, as well as more PD-L1 expressing cells. Our data suggest that a higher rate of polyp development could be associated with a defect of the local immune response, potentially paving the way for immunotherapy management of these patients.

Contents

Acknowledgement	1
Abstract	2
Contents	7
List of Figures	11
List of Tables	12
Notations	13
1 Introduction	15
1.1 Introduction to Immunology	16
1.1.1 History of Immunology	16
1.1.2 Cells of the Immune System	18
1.1.2.1 Hematopoiesis	18
1.1.2.2 Myeloid Lineage	18
1.1.2.3 Lymphoid Lineage	18
1.1.2.4 Intercellular communication	18
1.1.3 Structure of the Immune System	20
1.1.4 Innate Immunity	21
1.1.4.1 Humoral innate immunity via the complement system	22
1.1.4.1.1 Activation of the complement system	22
1.1.4.1.2 Effector functions of the complement system	22
1.1.4.2 Cellular innate immunity	22
1.1.4.2.1 Macrophages	22
1.1.4.2.2 Granulocytes	23
1.1.4.2.3 NK cells	23
1.1.4.2.4 Dendritic cells	23
1.1.4.3 Initiation of the acquired immune response by antigen presenting cells	24
1.1.4.3.1 Peptide presentation by MHC molecules	24
1.1.4.3.2 Migration of antigen presenting cells toward lymphoid or-	
gans leads to acquired immunity priming	24
1.1.5 Acquired Immunity	25
1.1.5.1 T lymphocytes	26

CONTENTS

1.1.5.1.1	TCR structure and diversity	26
1.1.5.1.2	Thymic T cell development	26
1.1.5.1.3	T lymphocyte circulation and activation	26
1.1.5.1.4	CD4 ⁺ T cells	27
1.1.5.1.5	CD8 ⁺ T cells	27
1.1.5.1.6	Memory T cells	28
1.1.5.1.7	Immune Checkpoints and T cell exhaustion	28
1.1.5.2	B lymphocytes	30
1.1.5.2.1	BCR structure and B cell development	30
1.1.5.2.2	B cell activation and differentiation	30
1.1.5.2.3	Antibody mediated immunity	31
1.1.6	Physiological Colorectal Immunity	33
1.2	Colorectal Cancer	35
1.2.1	Cancer Definition	35
1.2.1.1	Hallmarks of tumor cell proliferation	37
1.2.1.2	Other hallmarks of cancer	37
1.2.2	Epidemiology of Colorectal Cancer	38
1.2.2.1	Incidence	38
1.2.2.2	Risk factors	38
1.2.3	Early Carcinogenesis: From Healthy Colon to Carcinoma <i>in Situ</i>	38
1.2.3.1	Colorectal premalignant pathways	39
1.2.3.2	The Adenomatous Pathway	40
1.2.3.3	The Serrated Pathway	40
1.2.3.4	Epidemiology of colorectal polyp in the average risk population	42
1.2.3.5	Hereditary colorectal cancer syndromes	42
1.2.3.5.1	Lynch	42
1.2.3.5.2	Familial Adenomatous Polyposis	43
1.2.4	Late Carcinogenesis	43
1.2.5	Screening for Colorectal Cancer	43
1.2.6	Colorectal Cancer Classification	45
1.2.6.1	The TNM classification system	45
1.2.6.2	Molecular	45
1.2.6.2.1	Microsatellite Instability	45
1.2.6.2.2	Consensus Molecular Subtypes	46
1.2.6.3	Immunoscore	47
1.2.7	Treatment of Colorectal Cancer	48
1.3	Tumor immunity in colorectal cancer	50
1.3.1	History of Tumor Immunology	50
1.3.1.1	Cancer immunosurveillance and immunoediting	50

CONTENTS

1.3.1.2	Immune contexture	51
1.3.1.3	Cancer immunotherapy	51
1.3.1.4	Precancerous lesions and tumor immunology	53
1.3.2	The cancer-immunity cycle: from release of tumor antigens to cytotoxicity . .	54
1.3.2.1	Tumor associated antigens	54
1.3.2.2	The cancer-immunity cycle	55
1.3.2.3	Tumor immune escape	56
1.3.3	Colorectal immune tumor microenvironment and immune contexture	56
1.3.3.1	Macrophages	56
1.3.3.2	Neutrophils	57
1.3.3.3	Myeloid-derived suppressor cells	58
1.3.3.4	Mast cells	58
1.3.3.5	NK cells	58
1.3.3.6	Dendritic Cells	58
1.3.3.7	T cells	59
1.3.3.7.1	CD4 ⁺ T cells	59
1.3.3.7.2	Cytotoxic CD8 ⁺ T cells	59
1.3.3.7.3	T cell exhaustion	60
1.3.3.8	B cells and plasma cells	60
1.3.3.9	Tertiary lymphoid structures	61
1.3.3.10	Immune parameters as predictive factors	61
1.3.3.11	iTME in precancerous polyps	62
2	Objectives	64
2.1	Context and underlying hypotheses	64
2.2	Objective and methodological tools	64
3	Research Methodology	66
3.1	Experimental Model and annotations	66
3.1.1	Cohort constitution	66
3.1.2	Cohort annotation	66
3.2	RNA sequencing and whole exome sequencing	67
3.2.1	RNA and DNA extraction	67
3.2.2	RNA sequencing and alignment	67
3.2.3	Differential expression analysis	67
3.2.4	CMS classification	67
3.2.5	Consensus ^{TME} scoring	67
3.2.6	Whole exome sequencing and alignment	68
3.2.7	Oncoplot generation	68
3.3	Multiplex immunohistochemistry	68

CONTENTS

3.3.1	Slide preprocessing	68
3.3.2	Multispectral IHC	68
3.3.2.1	Staining	68
3.3.2.2	Image acquisition	69
3.3.2.3	Image analysis	69
3.3.3	Brightplex IHC	69
3.3.3.1	Staining	69
3.3.3.2	Image analysis	69
3.3.4	Linear Mixed Models	69
3.3.5	Visualization and Statistical analysis	70
4	Results	71
4.1	Cohort presentation and molecular characterization	71
4.1.1	Cohort presentation	71
4.1.2	The landscape of somatic alterations between colorectal premalignant pathways	73
4.1.3	CMS classification between colorectal premalignant pathways	75
4.2	Immunohistochemistry assays and associated analysis	76
4.2.1	Lymphoid panel analysis workflow	77
4.2.2	T cell exhaustion panel analysis workflow	80
4.2.3	Immunohistochemistry analysis	81
4.3	Colorectal premalignant pathways harbor distinct iTME	83
4.3.1	Interpathway analysis	83
4.3.2	Intrapathway analysis	89
4.4	Polyp development rate is associated with variation of iTME of premalignant lesions	92
4.4.1	Molecular characterization of frequency groups	92
4.4.2	First glance at the iTME between frequency group	92
4.4.3	Polyp frequency and iTME in the serrated pathway	94
4.4.4	Polyp frequency and iTME in the adenomatous pathway	99
4.4.5	Influence of time and recurrence on the iTME in the adenomatous pathway	104
4.5	Development of a linear mixed model describing variable associated with iTME variations	108
5	Discussion and Conclusion	111
5.1	Discussion	111
5.1.1	Molecular characterization of the cohort	111
5.1.2	iTME characterization	112
5.1.3	Interpathway characterization of the iTME	113
5.1.4	Intrapathway characterization of the iTME	114
5.1.5	Frequency groups	114
5.1.6	Modeling	116

CONTENTS

5.2 Conclusion	117
5.3 Perspectives	118
References	134
A Appendix	135

List of Figures

1.1	The two early theories of immunity	17
1.2	Simplified schematic representation of hematopoiesis. Adapted from [12]	19
1.3	Primary and secondary lymphoid organs: Naive lymphocytes generated in BM and thymus will encounter and respond to antigen in the peripheral lymphoid organs. From [16]	21
1.4	Activating and inhibitory checkpoints presented on the surface of T cells (right) and APC (left) From [34]	29
1.5	Colorectal anatomy	34
1.6	Hallmarks of cancer. From [54]	36
1.7	Risk factors of colorectal cancer. From [48]	39
1.8	Precancerous pathway of colorectal cancer	41
1.9	Immunoscore as a prognostic marker in CRC	48
1.10	effects of the immune infiltrate on the prognosis of patients with cancer. From [5]	52
1.11	The Cancer-Immunity Cycle and its regulators. From [140]	55
1.12	The immune tumor microenvironment. From [5]	57
1.13	The composition and function of tertiary lymphoid structures in cancer. from [192]	62
2.1	Integrative analysis of colorectal iTME	65
4.1	Graphical overview of the cohort. Each patient is represented by an horizontal line. Each time point, noted tx, can present several lesions of various type. Recurrence event that happened after the inclusion and are not included in the study samples are represented in dotted lines	71
4.2	Establishment of a risk variable using the frequency of detected polyps. A. Patients were divided into three frequency group based on their number of polyp developed per year. B. Frequency group description C. Premalignant pathway distribution between frequency groups	73
4.3	Mutational profile of polyps. A. Oncoplot displaying somatic mutations present in $\geq 10\%$ of samples. (Top) Mutation burden represented by bar plot. (Right) Percentage of mutations within samples. B. Proportion of mutated oncogene per precancerous pathway and cancer C. TMB levels between various lesions groups. Mann Whitney U test, *p<0.05	74

LIST OF FIGURES

4.4	Distributions of consensus molecular subtypes (CMS) in the cohort	76
4.5	Lymphoid panel presentation. A. Lymphoid panel markers. B. Unmixed signal of lymphoid panel markers. C. Multichannel image of a TLS stained by the lymphoid panel	78
4.6	Definition of ROIs and TLS using an AI-based classifier. A. AI-based tissue classifier was trained on DAPI, CD3 and CD20 signals to identify stromal, epithelial and TLS compartments. B. Definitions of 3 ROIs using pathologist annotations and tissue classifier.	79
4.7	TCE panel presentation. A. TCE panel markers. B. Chromogenic images (top) and extracted signal (bottom) after deconvolution. C. Multichannel image after fusion of extracted signals from the TCE panel.	80
4.8	Definition of TLS types	81
4.9	Data extracted from the IHC analysis A. Selected phenotypes from the lymphoid and TCE panels. B. Heatmap representing two-dimensional UHC by IHC phenotype and sample ROIs. Row splitting is based on the UHC dendrogram C. Proportion of TLS types within the whole cohort.	82
4.10	Colorectal precancerous pathways exhibit distinct gene expression profiles. tSNE analysis and visualization of gene expression profiles of various colorectal lesions . . .	84
4.11	Colorectal precancerous pathways exhibit distinct immune signature patterns Heatmap representing two-dimensional UHC by Consensus ^{TME} scores and precancerous lesions. Row and column splitting are based on the UHC dendrogram	85
4.12	Colorectal precancerous pathways exhibit distinct iTME pattern. A. PCA analysis and representation of samples per pathway based on their IHC phenotype densities in CT, AE and IM. B. TLS profiles between the two pathways. Mann Whitney U test	85
4.13	Adenomatous polyps present a more infiltrated iTME at the CT A. Heatmap representing two-dimensional UHC by IHC phenotype densities in CT and sample. Row splitting is based on the UHC dendrogram. B. Densities of IHC phenotypes in the CT of serrated and adenomatous pathways. Mann Whitney U test	86
4.14	Adenomatous polyps present a more infiltrated iTME at the IM A. Heatmap representing two-dimensional UHC by IHC phenotype densities in IM and samples. Row splitting is based on the UHC dendrogram. B. Densities of IHC phenotypes in the IM of serrated and adenomatous pathways. Mann Whitney U test	87
4.15	Total cell density explains differences between Consensus^{TME} scores and IHC densities in premalignant pathways. A. CD8 ⁺ and CD4 ⁺ T cells Consensus ^{TME} score and CT densities in serrated and adenomatous pathways B. (top) H&E representative image of serrated and adenomatous polyps (bottom) Total cell density in CT of serrated and adenomatous pathways. C. CD8 ⁺ and CD4 ⁺ T cells as proportion of total cells in CT of serrated and adenomatous pathways.	88

4.16 **Serrated lesions present an homogeneous iTME profile** A. Densities of IHC phenotypes in HPs and SSLs. B. Densities of IHC phenotypes in non advanced and advanced SSLs. Mann Whitney U test 90

4.17 **Adenomatous lesions present an homogeneous iTME profile** A. Densities of IHC phenotypes in LG APs and HG APs. B. Densities of IHC phenotypes in non advanced and advanced APs. Mann Whitney U test 91

4.18 **Frequency groups present similar molecular profiles** A. Premalignant pathway distribution between frequency groups B. Proportion of mutated oncogene per frequency group and pathway C. TMB levels between frequency groups D. Distributions of consensus molecular subtypes (CMS) per frequency group and pathway 93

4.19 **Frequency groups present specific iTME profile and TLS maturation** A. PCA analysis and representation of samples per frequency group based on their IHC phenotype densities in CT, AE and IM. B. TLS profiles between frequency groups C. TLS profiles between frequency groups for serrated lesions. D. TLS profiles between frequency groups for adenomatous lesions. Mann Whitney U test, *p<0.05; **p<0.01 94

4.20 **F1 serrated polyps present a more infiltrated iTME at the CT** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in CT and serrated sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Densities of IHC phenotypes in the CT of serrated samples according to frequency groups. Mann Whitney U test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 95

4.21 **F1 serrated polyps present a more infiltrated iTME at the IM** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in IM and serrated sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Densities of IHC phenotypes in the IM of serrated samples according to frequency groups. Mann Whitney U test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 97

4.22 **F1 serrated polyps present higher levels of immune signatures** A. Heatmap representing two-dimensional UHC by Consensus^{TME} score and immune signatures, and serrated sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Consensus^{TME} score and immune signatures in serrated samples according to frequency groups. Mann Whitney U test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 98

4.23 **Frequency group adenomatous polyps present a distinct iTME** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in CT and adenomatous sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Densities of IHC phenotypes in the CT of adenomatous samples according to frequency groups. Mann Whitney U test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 100

LIST OF FIGURES

4.24 **F1 adenomatous polyps present a more infiltrated iTME at the IM** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in IM and adenomatous sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Densities of IHC phenotypes in the IM of adenomatous samples according to frequency groups. Mann Whitney U test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 101

4.25 **F1 adenomatous polyps present higher levels of immune signatures** A. Heatmap representing two-dimensional UHC by Consensus^{TME} score and immune signatures, and adenomatous sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Consensus^{TME} score and immune signatures in adenomatous samples according to frequency groups. Mann Whitney U test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 103

4.26 **Variation of the iTME of adenomatous lesions regarding frequency group are independent of time points** A. Densities of IHC phenotypes in the CT of t0 adenomatous samples according to frequency groups. B. Densities of IHC phenotypes in the IM of t0 adenomatous samples according to frequency groups. Mann Whitney U test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 106

4.27 **Variation of the iTME of adenomatous lesions regarding frequency group are independent of time points** A. Densities of IHC phenotypes in the CT of adenomatous samples according to frequency groups and time point. B. Densities of IHC phenotypes in the IM of adenomatous samples according to frequency groups and time points. Mann Whitney U test, *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001 107

5.1 **Serrated lesions present higher intraepithelial CTLs densities** 113

5.2 **Veen diagram of differentially expressed genes of the immunome between frequency group in each colorectal premalignant pathway** 118

5.3 **F1 lesions are enriched in non coding RNAs.** The biotype of differentially expressed RNAs (p-value > 0.05) between frequency groups was assessed while pooling F2 and F3 lesions together 119

A.1 Heatmap of all ROI from all samples 137

A.2 Heatmap of the serrated per polyp frequency 138

A.3 Heatmap of the adenomatous per polyp frequency 139

List of Tables

1.1	Phases of the immune response. From [12]	32
1.2	Molecular Subtypes of Colorectal Cancer. From [94]	46
3.1	Patients characteristics	66
3.2	Antibodies used for the lymphoid panel	68
4.1	Experimental design and sample size	72
4.2	Summary of the characterization of the iTME between polyp frequency groups	104
4.3	Linear mixed modeling of CT cell densities	109
4.4	Linear mixed modeling of IM cell densities	110
A.1	List of DEGs from the immunome between frequency group. In each precancerous pathway, genes from the immunome were extracted from a differential gene expression analysis between frequency group (using the thresholds $fc = 0.5$ and $p\text{-value} = 0.05$). Genes common to both pathway were identified	135
A.2	Gene list for Consensus ^{TME} score and TLS signatures	136

Notations

ADCC	Antigen-dependent cell-mediated cytotoxicity
AE	Adjacent epithelium
AP	Adenomatous polyps
APC	Antigen presenting cell
BCT	B cell receptor
cDC	conventional dendritic cell
CIN	Chromosomal instability
CRC	Colorectal cancer
CT	Center of tumor
CTL	Cytotoxic T lymphocyte
CXCL	Chemokine ligand
CXCR	Chemokine receptor
DAMP	Danger-associated molecular patterns
DC	Dendritic cells
DEG	Differentially expressed genes
GALT	Gut-associated lymphoid tissues
GC	Germinal center
HG AP	High-grade adenomatous polyp
HLA	Human leukocyte antigen
HP	Hyperplastic polyp
IFN	Interferon
IL	Interleukin
ILF	Isolated lymphoid follicles
IM	Invasive margin
iTME	Tumor immune microenvironment
LG AP	Low-grade adenomatous polyps
MDSC	Myeloid-derived suppressor cell
MHC	Major histocompatibility complex
MSI	Microsatellite instability
NETs	Neutrophil extracellular traps
NK cells	Natural killer cells

LIST OF TABLES

PAMP	Pathogen-associated molecule patterns
pDC	Plasmacytoid dendritic cells
PRR	Pathogen recognition receptors
RNAseq	RNA sequencing
SILT	Solitary intestinal lymphoid tissues
SLO	Secondary lymphoid organ
SSL	Sessile serrated lesions
TAA	Tumor associated antigen
TCR	T cell receptor
Tfh	T follicular helper
TGF	Transforming growth factor
Th	T helper
TIBL	Tumor infiltrating B lymphocyte
TIL	Tumor infiltrating lymphocyte
TITL	Tumor infiltrating T lymphocyte
TLS	Tertiary lymphoid structure
TME	Tumor microenvironment
TNF	Tumor necrosis factor
Treg	Regulatory T cells
UHC	Unsupervised hierarchical clustering
VEGF	Vascular endothelial growth factor
WES	Whole Exome Sequencing

1 Introduction

Today, cancer is commonly viewed and analyzed as an evolving ecosystem whose fate depends on intrinsic and extrinsic factors [1]. Indeed, cancer cells are in perpetual interactions with various components of the tumor microenvironment (TME), and this constant crosstalk plays an important role in tumor development.

The idea that the immune system could shape carcinogenesis was proposed as early as the 1950s, when F. Macfarlane Burnet proposed the concept of cancer immunosurveillance. However, this theory was met with skepticism due to the lack of direct scientific demonstration [2]. In the early 2000s, however, Robert D. Schreiber and his colleagues discovered that mice deficient in both adaptive and innate immunity spontaneously develop tumors and that cancers can escape immune surveillance by modulating their immunogenicity, a mechanism called “immunoediting” [3].

Since then, interest in the field of tumor immunology has increased, in part due to the remarkable results of novel immunotherapeutic approaches, especially immune checkpoint blockade. In parallel, the immune microenvironment of cancer has been studied to identify markers that could be associated with the clinical outcome of cancer patients, in terms of survival and therapeutic responses [4]. The characterization of the tumor immune microenvironment (iTME) and the so-called ‘immune contexture’ (analysis of type, location, and density of tumor-infiltrating immune cells in association with patient prognostic) resulted in the identification of immune cell types that can be associated with either good or bad patient prognostic [5]. For example, studies have shown that lymphocytic infiltration within tumor or peritumoral tissue is a favorable prognostic factor in colorectal cancer [6] and a wide range of cancer types. The development of such prognostic factors can lead to a more personalized and suitable treatment for patients. However, despite recent advances, the multiplicity of mechanisms by which tumors can escape immunotherapies make most treatment only efficient in a small subset of patients [4].

Most solid tumors are growing from preexisting benign lesions, called precancerous or premalignant lesions. If tumor immune contexture is thoroughly studied in a wide range of cancer types, little is known about the state of the immune system in precancerous lesions. This knowledge gap can be mainly explained by the fact that most studies focus on invasive carcinoma and that cancer is often diagnosed at an advanced stage, leaving precancerous lesions rarely detected and even more rarely surgically removed. Colorectal premalignant lesions, among others, can be considered good models of early carcinogenesis because they can be accessed without heavy surgical intervention and are, in most developed countries, screened and surgically removed to prevent cancer development. Studying this kind of lesions could provide valuable information to understand the mechanisms that take place during early carcinogenesis. A first study published in *Nature* [7] by our team in 2019 focused on 9 stages of lung precancerous lesions and demonstrated that the immune

1.1. INTRODUCTION TO IMMUNOLOGY

system was responding to these lesions and that some of them were able to escape this immune response, leading to cancer development. Moreover, while current precancerous treatment focuses on surgical removal, characterization of the immune contexture could lead to a better patient stratification, avoiding unnecessary surgical intervention and its consequences. This approach could also result in immunological treatment preventing the development of secondary premalignant lesions and their progression toward cancer, especially in high-risk patients. Tackling carcinogenesis using immunotherapy might be easier at early stages than in the more challenging mature TME [8]. This study will focus on characterizing the immune contexture in colorectal premalignant lesions, and try to identify immune parameters associated with patients developing an abnormally high number of polyp.

In order to properly grasp the stakes and challenges of this subject, it is necessary to first present the mechanisms underlying the immune response on one side, and colorectal carcinogenesis on the other. Therefore, I will start by introducing the human immune response, then I will follow by an overview of the colorectal carcinogenesis process and finally, I will expose the state of the art concerning the immune microenvironment of colorectal cancer.

1.1 Introduction to Immunology

1.1.1 History of Immunology

Empirical applications of immunology were present since the antiquity, as illustrated by the historian Thucydides relating that, during an outbreak of plague in Athene (V^e century BC), only people who already contracted this disease could take care of the sick ones. More recently, in 1776, the English doctor Edward Jenner developed the first vaccination based on an observation: cowboys were resistant to smallpox thanks to previous exposition to a benign disease close to smallpox, called vaccinia and contracted in contact with cows. However, as these examples were only based on empirical experience, the field of Immunology is considered to have really arisen in the end of the nineteenth century, out of the field of medical microbiology [9].

From this field came the concept that infectious diseases are caused by microorganisms, a theory demonstrated by the work of Robert Koch (1843–1910) on the tuberculosis bacillus and Louis Pasteur (1822–1895) on the avian cholera. Louis Pasteur then went on to demonstrate empirically and rationally the vaccine results obtained by Edward Jenner and developed a vaccine against rabies.

The main question raised by these discoveries was regarding the potential existence of a defense mechanism within the host of the disease and the form that this mechanism could take. This question was answered at the same time in two different places, as two of future main immunology branches were being discovered. On the one hand, Elias Metchnikoff (1845–1916) created the cellular theory of immunology and paved the way for the branch of innate immunity by discovering the existence

1.1. INTRODUCTION TO IMMUNOLOGY

of phagocytic cells, able to engulf and destroy invading pathogens. On the other hand, Emil von Behring (1854-1917) and Paul Ehrlich (1854-1915) gave birth to the humoral theory of immunity by identifying antibodies capable of neutralizing microbial toxins, thus laying the basis for the acquired immunity branch [10].

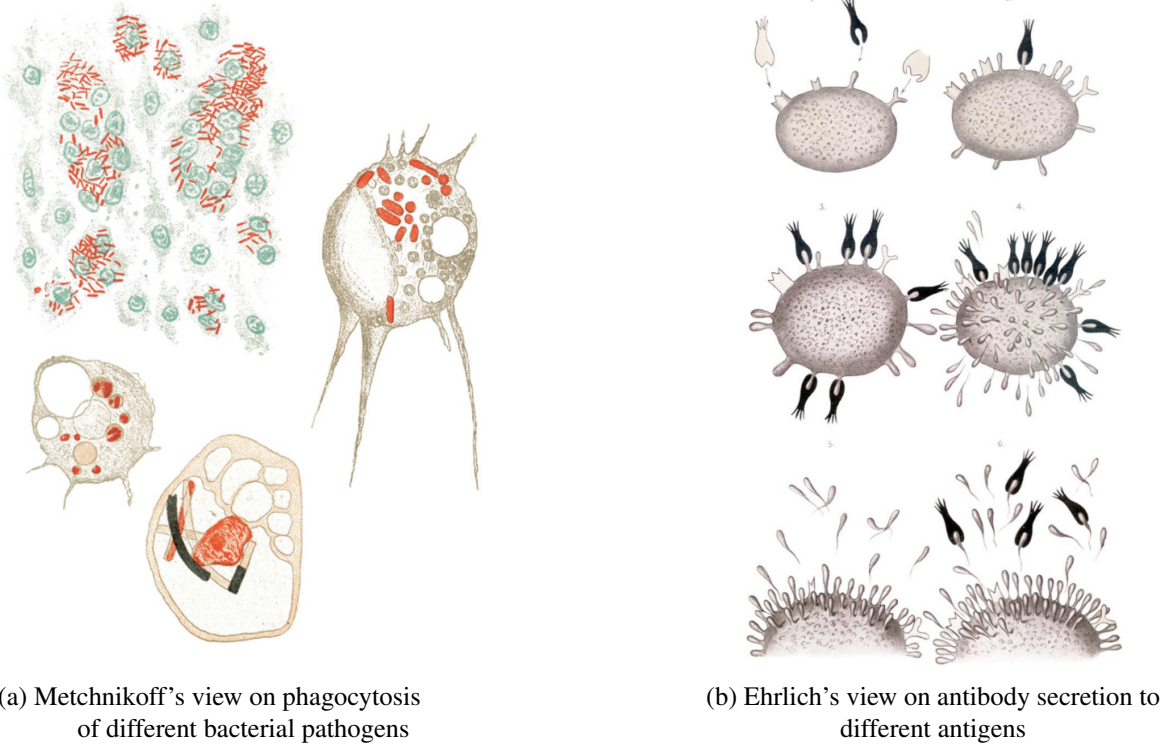


Figure 1.1: The two early theories of immunity

For some time, these two theories were considered incompatible and led to some controversy, until concepts of complementarity emerged, leading to the immunology as we know today. The two main branches of the immune response, the innate and adaptive immunity, present complementary properties.

To effectively protect the individual against disease, the main functions of the immune system can be divided into 4 steps:

- Immune recognition, that will detect the presence of an infection. This can be done at the time of infection by cells of the innate immune system, which can provide an immediate immune response, and later by the cells of acquired immunity.
- Immune effector functions, aiming to contain and, if possible, eliminate the infection.
- Control of the immune response to avoid tissue damage related to the immune system. This implies the presence of immune regulation mechanisms, often performed by immune cells themselves. Failure leads to conditions such as allergies or autoimmunity.

1.1. INTRODUCTION TO IMMUNOLOGY

- Immunological memory, a specificity of the acquired immune system, which will be able to protect the individual from reinfection by the same pathogen for a long period of time.

These various roles, in association with the variety of potential pathogens to recognize and eliminate, require a high diversity of immune cell types and functions, emerging from a common precursor through the hematopoiesis process.

1.1.2 Cells of the Immune System

1.1.2.1 Hematopoiesis

Hematopoiesis is the process responsible for generation and maintenance of the blood cellular elements. These include erythrocytes, platelet and leukocytes, which represent all immune cells. These cells derive from a common progenitor, the hematopoietic stem cells (HSC), primary found in the bone marrow (BM) and characterized by their ability to self-renew and differentiate into numerous effector cells, as seen in Figure 1.2 [11].

Hematopoietic cells can be divided into two lineages, each having its own specific precursor: The myeloid and lymphoid lineage. Following reception of the differentiation signal, HSCs will differentiate into a multipotent progenitor, losing their self-renewal potency. This progenitor will itself differentiate either into myeloid progenitor cell or lymphoid progenitor cell [13].

1.1.2.2 Myeloid Lineage

Three major functions are associated with the myeloid lineage: oxygen transport, blood clotting, and innate immunity. All cells responsible for these functions derive directly or indirectly from a common myeloid progenitor. This progenitor can differentiate into erythrocytes, responsible for oxygen transport throughout the body, megakaryocytes that will produce platelets, responsible for coagulation, and cells of the myeloid immune lineage: dendritic cells, eosinophils, basophils, neutrophils, and monocytes. The latter will themselves differentiate into dendritic cells and macrophages that will play a critical role in innate immunity.

1.1.2.3 Lymphoid Lineage

The lymphoid lineage arises from a common lymphoid progenitor and only includes immune cells. Among these lymphoid cells, T and B cells are the main actors of acquired immunity, while natural killer cells (NK cells) and dendritic cells (DCs) are part of the innate immune response.

1.1.2.4 Intercellular communication

In order to coordinate an efficient response, immune cells communicate by secreting molecules, called cytokines, that will regulate most immune functions. For example, cytokines are responsible for hematopoiesis homeostasis or regulation of circulating lymphocytes. The complexity of this

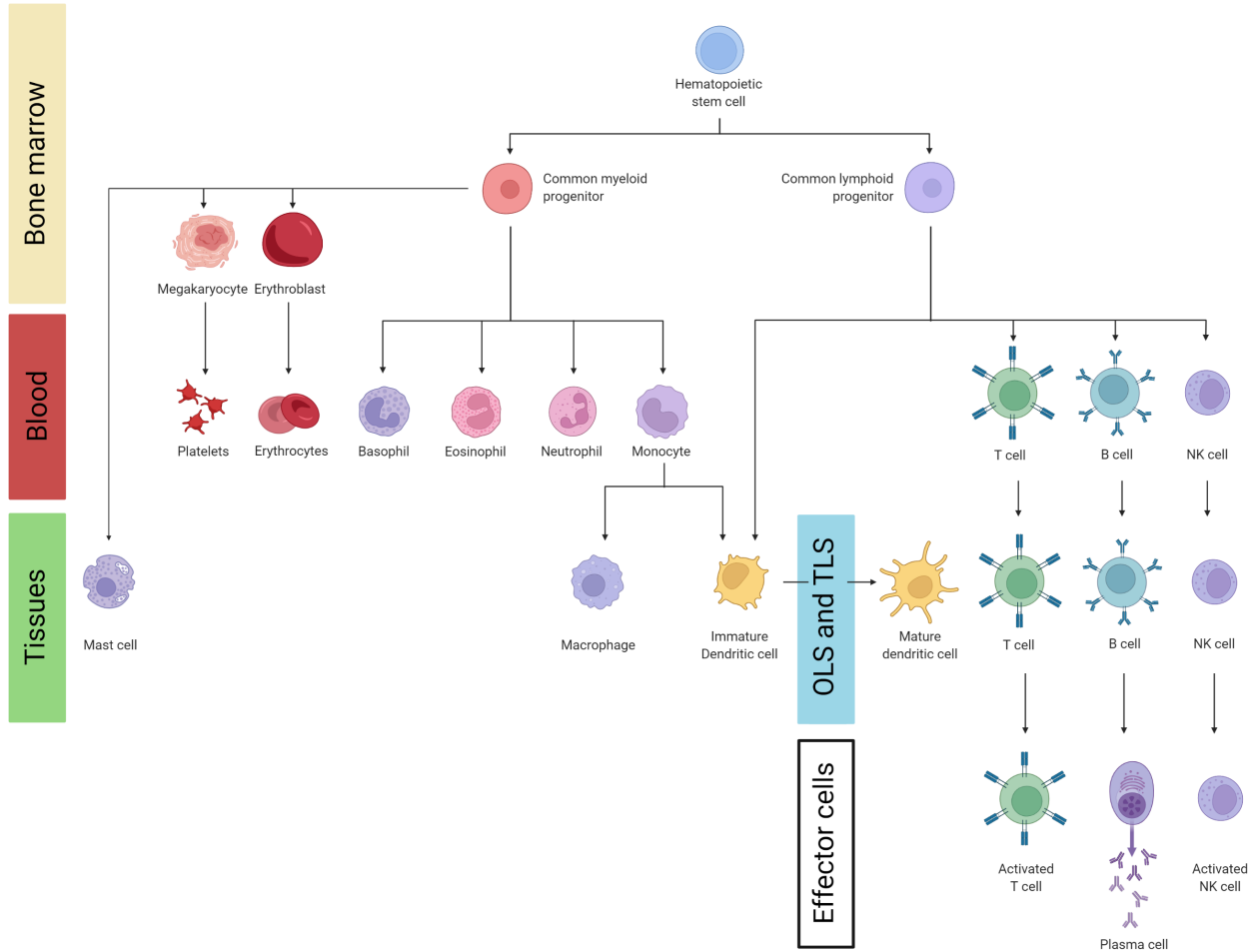


Figure 1.2: Simplified schematic representation of hematopoiesis.
Adapted from [12]

communication is due to the fact that numerous cell types can secrete the same cytokine under various conditions, and this cytokine will have different effects depending on the receptor cell and immune context. Non immune cells can also secrete or respond to cytokines, participating in immune response modulation [14]. Cytokines can be classified into different groups:

- Interleukins (IL) are a class of cytokines whose role is to modulate the development and function of other immune cells. Some example are IL-2, secreted by T cells, DCs, NK cells and mast cells, and responsible for proliferation of effector lymphocytes and regulatory T cells (T_{reg}) and IL-10, secreted by most immune cell type and able to drive immunosuppression and anti-inflammatory process in the majority of these cells.
- Chemokines play the role of chemoattractant, leading to recruitment of other immune cells. For instance, chemokine ligand 13 (CXCL13), expressed in secondary and tertiary lymphoid organs, will lead to recruitment of lymphocytes expressing its receptor, chemokine receptor 5 (CXCR5), to lymph nodes and B cell follicles

1.1. INTRODUCTION TO IMMUNOLOGY

- Interferons (IFN) are critical in the antiviral response, promoting cytotoxicity, T_{h1} orientation and upregulation of antigen presentation among other.
- Transforming growth factor (TGF) are implied in anti-inflammatory effects, induction of T helper 17 cells (T_{h17}), T_{reg} and immune tolerance.
- Tumor necrosis factor (TNF) are promoting an inflammatory response upon secretion by macrophages
- Colony-stimulating factors are the main regulator of hematopoiesis

Through this intercellular communication, the immune system can maintain homeostasis under physiological conditions, respond to pathogen invasion, and reduce inflammation to avoid immune-related damage once the infection has resolved. Another key characteristic of the immune system that allows the maintenance of homeostasis and response to infections is its structure, from local immunity to lymphoid organs.

1.1.3 Structure of the Immune System

Organs responsible for T and B cells generation are called primary lymphoid organs (BM and thymus). After leaving this compartment, immune cell will circulate through the circulatory system, from where they will replenish peripheral immune site and perform systemic immunosurveillance (Figure 1.3). From the blood, lymphocytes can either go to a potential entry site of pathogens such as mucosal epitheliums, or to secondary lymphoid organs (SLOs), where the adaptive immune response is orchestrated. Another circulating compartment, the lymphatic system, will drain peripheral tissues, allowing immune cells to migrate toward secondary lymphoid organs, leading to the presentation of pathogenic patterns detected at the tissue level to be presented to naive lymphocytes, leading to their selection, proliferation and differentiation. The lymphatic system also allows immune cells from tissue and SLO to recirculate toward the blood [15].

SLOs can be encapsulated organs, such as lymphatic nodes, mesenteric nodes and the spleen, or networks of lymphoid cell aggregates and tissue that is distributed into specific tissues. Example of the latter are mucosa associated lymphoid tissue (MALT), distributed in submucosal and mucosal layers in guts (e.g. Peyer's patches), bronchi or the nasal tissue. SLO are the main orchestrator of adaptive immune response and tolerance, and are compartment-specialized: the spleen for the blood, the nodes for the tissues and the MALT for the mucosal areas.

In addition to SLO, tertiary lymphoid structures (TLSs) can also be observed, especially in the case of chronic infections and cancer. Like SLOs, they can include germinal centers (critical for B cell response), drain immune cells from nearby tissues, and are the siege of lymphocyte activation and proliferation. They are induced in cases of chronic inflammation and lead to the formation of an adaptive immune center near the infection site, allowing for an optimal immune response [17].

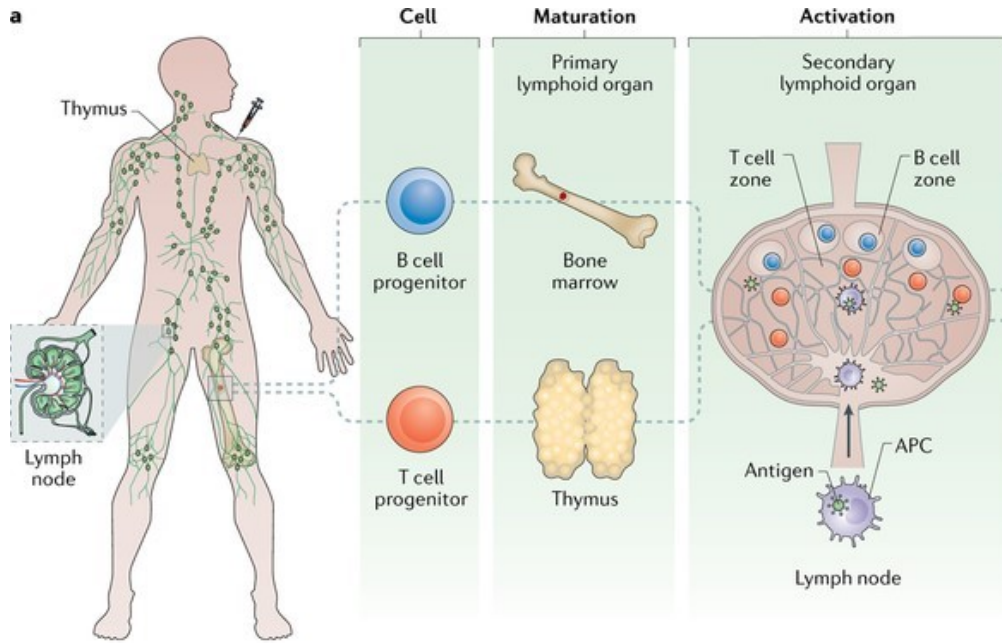


Figure 1.3: **Primary and secondary lymphoid organs:** Naive lymphocytes generated in BM and thymus will encounter and respond to antigen in the peripheral lymphoid organs. From [16]

1.1.4 Innate Immunity

As presented earlier, one of the main functions of the immune system is to contain potential pathogenic infection and tissue damage to their primary site. This implies a fast and efficient response, which is primarily enforced by the innate arm of immunity. Following tissue damage and/or invasion of micro-organism, one of the first responses to limit damage and promote healing is the coagulation system. In addition to limiting pathogen dissemination, this system interacts directly with the immune system, supporting pathogen elimination through innate immunity activation.

Cells of the innate immunity are the first line of response in case of breach of the physical barriers and are able to mount an immediate immune response. Their efficiency is based on the recognition of molecular pattern common to many pathogens. These patterns are then called pathogen-associated molecule patterns (PAMPs), such as bacterial cell wall proteins, viral nucleic acids or mannans from fungi cell walls. Being specific to micro-organisms and constant within micro-organism from a same taxon, PAMPs are ideal targets to discriminate between self and non self. Another signal able to activate innate immune cells are danger-associated molecular patterns (DAMPs), which are restricted to the intracellular compartment in physiological conditions but will get exposed in the extracellular compartment upon situation such as tissue damage, leading to immune activation [18].

Both DAMPs and PAMPs will be detected by pathogen recognition receptors (PRRs) expressed by myeloid cells. PRRs give the innate immune system its ability to distinguish self from non-self, triggering the innate immune response and initiating the adaptive immune response in a second time. It is important to note that the adaptive immune response can in return trigger innate

1.1. INTRODUCTION TO IMMUNOLOGY

immune response, for instance with the production of antibodies being able to trigger phagocytosis from neutrophils and macrophages. The generic recognition patterns involved in the innate immune activation give this immune arm its main properties: large spectrum recognition, fast activation and response.

1.1.4.1 Humoral innate immunity via the complement system

1.1.4.1.1 Activation of the complement system

The complement system is made up of more than 30 plasma proteins, which work as an activation cascade through conformational changes and proteolysis. During infection, the complement system can be activated through three different possible pathways of complement activation: the classical pathway is activated by antibody-coated pathogens and pathogen patterns, the alternative pathway by spontaneous hydrolysis of the C3 protein (and thus maintaining a permanent low level of activation) and the lectin pathway by the presence of lectin-type proteins on the surface of pathogens.

Each complement pathway will end up triggering a proteolysis cascade, amplifying tenfold the response and resulting in antipathogenic functions through the cleavage of the C3 protein into CD3a, a peptide that induces inflammation, and CD3b, the main effector molecule of the complement cascade [19].

1.1.4.1.2 Effector functions of the complement system

Following C3 proteolysis, several mechanisms can lead to pathogen clearance.

First, C3a, in interaction with C5a (another intermediate of the complement cascade), will promote inflammation and recruitment of phagocytic cells to the site of infection. These molecules, called anaphylatoxins, can also promote T cell survival and the recruitment of lymphocytes and neutrophils.

The C3b molecule will coat the pathogen in a process called opsonization, leading to phagocytosis by innate immune cells with receptors for C3b.

Finally, once the complement cascade is completed, several of its components will assemble together, forming a membrane-attack complex, which will disrupt the pathogen membrane, leading to cell lysis.

1.1.4.2 Cellular innate immunity

1.1.4.2.1 Macrophages

Macrophages are tissue resident cells, either of prenatal origin or derived from monocytes, to replenish the macrophage pool or in response to inflammation [20]. They are responsible for tissue homeostasis through phagocytosis of apoptotic bodies under physiological conditions and are generally the first responders in case of pathogenic infection. Macrophages are found in most tissues and especially connective tissue, such as the submucosal layer of the gastrointestinal tract. Following PRRs activation, macrophages will eliminate pathogens or infected cells through phagocytosis, as

1.1. INTRODUCTION TO IMMUNOLOGY

well as secrete cytokines and chemokines, leading to tissue inflammation and chemoattraction of circulating immune cells such as monocytes, neutrophils or T cells subtypes.

1.1.4.2.2 Granulocytes

Neutrophils, with eosinophil, basophils and mast cells, are part of the family of innate immune cells called granulocytes, due to the presence of specific granules in their cytoplasm. Although all cells in this family are capable of phagocytosis, neutrophils are the most potent.

Together with macrophages, neutrophils play a critical role in innate immunity thanks to their ability to recognize, engulf and destroy microorganisms harboring PAMPs. While representing the majority of circulating immune cells, they are not infiltrated in healthy tissue, but recruited under inflammatory conditions [21]. They make up the first wave of cells that cross the blood vessel wall to enter an inflamed tissue. In addition to their phagocytic capacities, neutrophils can also extrude their nuclear chromatin into the extracellular space, leading to cell death and the formation of a fibril matrix called neutrophil extracellular traps (NETs). NETs are able to trap microorganisms, facilitating phagocytosis by other immune cells.

Basophils have mainly anti-parasitic functions, and both basophils and eosinophils are circulating in the blood, while mast cells enforce inflammatory response and anti-microbial response within the tissues. Eosinophils, basophils, and mast cells also play a role in the allergic inflammatory response [22].

1.1.4.2.3 NK cells

Although deriving from the common lymphoid progenitor, NK cells are involved in the innate immune response, which they achieve through cytotoxic activity or cytokinic secretion [23].

Although lacking antigen specificity, NK cells are capable of detecting the presence of class I MHC molecules at the cell surface, which will inhibit their activation. When in contact with cells that do not present any class I MHC molecule, such as foreign cells or virus invaded cells that have downregulated MHC I expression to avoid viral antigen recognition, NK cells will trigger their effector function, leading to cytotoxicity and cytokine production. Cytotoxicity is performed by secreting cytolytic molecules such as granzymes and perforin, leading to infected cell lysis.

NK cell activation is enhanced by interferon- and macrophage-derived cytokines, as well as fixation of the constant antibody fraction, the FC fraction, as they are carrying FC receptors.

1.1.4.2.4 Dendritic cells

DCs can emerge from both the myeloid or the lymphoid pathway. They can be found in blood or tissues in their immature form, where they will be able to recognize and phagocyte microorganisms [24]. Upon recognition and phagocytosis of the pathogen, DCs will mature and migrate to SLOs and TLSs. Mature DCs will lose most of their phagocytic potential while increasing cytokine production and costimulation molecule expression. This changes toward a professional antigen presenting cell (APC) phenotype make them very efficient at triggering acquired immunity. DCs are a very hetero-

geneous population and can be divided into various subsets: type 1 and 2 conventional DCs (cDC1 and 2), plasmacytoid DCs (pDCs) and monocyte-derived DCs [25].

1.1.4.3 Initiation of the acquired immune response by antigen presenting cells

1.1.4.3.1 Peptide presentation by MHC molecules

While innate immunity rely on the recognition of generic pathogenic patterns, one of the main specificity of the acquired immunity is the recognition of specific peptide, i.e. amino acid sequence from a broke down protein. The detection of peptides not encountered under physiological conditions, such as pathogenic peptides or peptides associated with infected cells, will activate acquired immunity.

One of the main mechanisms enabling this activation is the presentation by most nucleated cells of endogenous peptide coming from their intracellular proteasome activity. Under physiological conditions, unwanted proteins (e.g. wrongly folded or excluded through turnover) are directed toward a degradation pathway implying the proteasome complex, generating peptides representative of the proteome of the cell. These peptides will then be presented at the cell surface by the class I major histocompatibility complex (MHC), an ensemble of molecules coded by the human leukocyte antigen (HLA) genes A, B and C in human. As they are coming from cytoplasmic protein degradation, endogenous peptide presented on class I MHC will reflect the internal state of the cell. For instance, upon intracellular infection, viral or bacterial peptide can be presented at the cell surface, which will possibly trigger acquired immune response. The main outcome of endogenous peptide presentation is to trigger cytotoxicity against the presenting cells, removing infected or cancerous cells [26].

Although class I MHC proteins are expressed on most cells (including immune cells), another pathway of peptide presentation is specifically associated with specific cells of the immune system, APCs. This pathway involve the class II MHC proteins, and leads to the presentations of extracellular peptides through endocytosis (including phagocytosis). Class II MHC molecules are coded by the HLA genes DP, DR and DQ and are especially expressed in B cells, DCs and macrophages. In tissues, the main role of exogenous peptide presentation is the activation of helper T cells. For instance, presentation of class II MHC associated peptide by B cells or macrophages will lead to their activation and function through helper T cells. However, DCs will accomplish their main function as APC within lymphoid tissue, where they will prime the acquired immune response.

1.1.4.3.2 Migration of antigen presenting cells toward lymphoid organs leads to acquired immunity priming

In addition to the aforementioned effects, inflammation will also increases the flow of lymph carrying microbes and APCs from the inflamed tissue to nearby lymphoid tissues. This will INCREASE the chance that mature DCs will be able to activate acquired immunity once in SLOs or TLSs.

In consequence, these lymphoid organs will become a place where lymphocytes will have a high probability of encountering antigens associated with pathogenic infections. In addition to presenting

antigens drained from the surrounding, DCs also carry a costimulatory signal necessary for T cell activation, making them extremely efficient in priming T cells, leading to activation of acquired immunity.

1.1.5 Acquired Immunity

In contrast to innate immunity, based on broad patterns either generic to a pathogenic class or tissue damage, the acquired immunity system relies on very specific recognition of pathogenic peptides by receptors carried by both B and T lymphocytes. This high specificity implies an enormous diversity of receptors, called immunoreceptors. In humans, the number of unique lymphocyte receptors is estimated to exceed 100 billion clones [27]. As the limited size of the genome does not allow one gene for each immunoreceptor, it has been hypothesized and demonstrated that these receptors rely on a set of genes that will undergo random recombination, leading to somatic diversification. The great variability of these receptors gives acquired immunity the ability to face all possible antigens.

Immunoreceptors expressed by T and B lymphocytes are lineage specific, but also share many common characteristics. They are called TCR and BCR (T and B cells receptors, respectively), and can be divided into a variable portion, which will recognize the antigen and is specific to its lymphocyte, and the constant portion. Immunoreceptors are composed of 2 different chains which increase their diversity. The overall diversity of TCR and BCR from an individual is called its T or B repertoire, respectively. Antigens are molecules specifically recognized by these immunoreceptors, and the capacity of these antigens to trigger an immune response is called immunogenicity.

Mirroring their common immunoreceptor properties, the development of B and T lymphocytes will share similar steps toward differentiation into effector and memory cells. Early in their development, T and B cells undergo the V(D)J recombination, a random rearrangement of several DNA loci that will form a unique pair of chains constituting the TCR and BCR receptors. Each lymphocyte thus expresses a unique receptor and can recognize a specific antigen. The recognition of this antigen will trigger signal transduction, which intensity will depend on the strength of the interaction [28].

As immunoreceptors are randomly generated, two critical selection steps, taking place in primary lymphoid organs, are required for the acquired immunity to be functional and safe for the host. First, we have positive selection, where the immunoreceptor will be tested for its ability to functionally recognize a peptide. Without the signal associated with this recognition, lymphocytes will undergo "death by neglect", eliminating useless lymphocytes from the following steps. Then, negative selection will eliminate most of self-reactive lymphocytes. As immunoreceptors are generated in a stochastic way, they are not inherently specific to non-self antigens and can recognize self molecules. In order to avoid autoimmune reactions, where the immune system will turn against host cells, mechanisms leading to self tolerance are needed [29].

Once the selection steps are cleared, naive (or antigen inexperienced) lymphocytes will be on

1.1. INTRODUCTION TO IMMUNOLOGY

the hunt for their putative cognate antigen. This step will mainly take place in SLOs and TLSs, where the presence of professional APCs will greatly increase the chance that naive lymphocytes will meet their specific peptide. Recognition will lead to positive selection and clonal amplification of the lymphocyte, and differentiation from the naive state to effector or memory.

While sharing numerous similarities highlighted above, T and B cells also have very specific characteristics that distinguish these two lineages.

1.1.5.1 T lymphocytes

1.1.5.1.1 TCR structure and diversity

The TCR is a heterodimer, made up of 2 variable chains. Most T cells bear a TCR composed of an α and a β chain, while a scarcer subtype is presenting $\gamma\delta$ TCR and plays a role in both innate and adaptive immunity. We will focus on the T lymphocytes associated with $\alpha\beta$ TCR. TCRs are associated with invariant chains that will be involved in signal transduction following peptide recognition, among them the CD3 complex, a commonly used marker to identify T cells. TCRs are also associated with a coreceptor, CD8 and CD4, depending on the functional differentiation of the lymphocyte [30].

The TCR is only able recognize antigenic peptides presented by MHC molecules, and the coreceptor will dictate which MHC can the TCR interact with. Because the TCR will be specific for a given peptide following the somatic recombination of its two chains, recognition of the TCR will require both the right peptide presented on the right MHC.

1.1.5.1.2 Thymic T cell development

Following the migration of T cells progenitor to the thymus, positive and negative selection steps will take place in this specialized organ. Thymic epithelial cells are the key component of the thymus and are located in two distinct morphological zones, the cortex and the medulla.

Medullary thymic epithelial cells are responsible for positive and negative selection [29]. This process is based on the ectopic expression of proteins that are otherwise restricted to differentiated organs in the periphery by both MHC classes. While T cells unable to interact with presented peptides will go through death by neglect, those that react too strongly to those self-antigens are eliminated, thereby preventing the onset of peripheral autoimmunity. The nature (CD4 or CD8) of the coreceptor is also established during this process.

1.1.5.1.3 T lymphocyte circulation and activation

Once mature and engaged in the CD4 or CD8 differentiation pathway, naive T lymphocytes will be released in the circulation, migrating towards SLOs where they will enter in contact with APCs, mainly DCs, presenting potential pathogenic peptides and costimulatory signals (proteins

1.1. INTRODUCTION TO IMMUNOLOGY

and cytokines). T cells that will not encounter their cognate peptide will exit the lymph node, recirculating in the blood toward other lymphoid tissues. Following peptide recognition and in the presence of the right costimulatory signals, T lymphocytes will be activated, leading to clonal proliferation and differentiation into effector or memory phenotype. Lack of a costimulatory signal (for instance, binding of the T cell co-receptor CD28 to its ligand CD80 or CD86, expressed on APCs) is often associated with self-antigen recognition and will lead to anergy or apoptosis of the lymphocyte, preserving peripheral tolerance.

While these previously described steps are similar between the 2 main lineage of T lymphocytes, the function of CD4⁺ T cells and CD8⁺ T cells are very distinct, as presented below.

1.1.5.1.4 CD4⁺ T cells

CD4⁺ T cells will recognize peptides presented by the class II MHC, and will in turn produce cytokines and present stimulatory proteins, supporting the establishment of an efficient and adapted immune response, earning themselves the name of T helper cells. Depending on the origin of the peptide, CD4⁺ T cells will differentiate into T helper cells 1, 2, or 17 (T_h1, 2, and 17, respectively) [31].

Intracellular pathogen peptides such as viruses and some bacteria will lead to differentiation into T_h1, which will promote cellular immunity, secreting IFN γ , TNF α and IL2, activating macrophages, NK cells, and CD8 T cells.

Presentation of peptides from extracellular pathogens such as some bacteria and parasites will promote humoral immunity via differentiation of CD4⁺ T cells into T_h2. T_h2 will cooperate with B cells as well as basophils, eosinophils or mast cells while producing cytokines such as IL4, 5, 6 or 13 [12].

T_h17 owe their name to their high production of IL-17A and IL-17F. They are induced in response to extracellular bacteria and fungi and will boost the neutrophil response to eliminate these pathogens.

Another subset of T helper cells, the T follicular helper (T_{fh}), is found mainly in SLO and TLS. Its main function is to assist B cells in generating class-switched immunoglobulins. In consequence, they contribute to the immune response against most pathogenic classes. T_{fh} can be identified with the expression of CXCR5 and PD1 in association with other classical CD4⁺ T cells markers (e.g. CD3 and CD4).

CD4⁺ T cells can also differentiate into T_{reg}, which have a major role in limiting the runaway immune response, avoiding host-related immune damage. Notably, they are associated with the FOXP3 transcription factor, and secrete the inhibitory cytokines IL10 and TGF- β .

1.1.5.1.5 CD8⁺ T cells

On the other hand, CD8⁺ T cells recognize peptides presented by the class I MHC molecule, and are a direct effector of the immune system, as they can perform a cytotoxic action by secreting cytolytic molecules such as granzymes and perforin, destroying the targeted cell. Cytotoxic CD8⁺

1.1. INTRODUCTION TO IMMUNOLOGY

T cells are referred to as cytotoxic T lymphocytes (CTLs). As class I MHC molecule are presented by all nucleated cells, CD8⁺ T cells do not require professional APC to detect their cognate peptide. However, the presence of costimulatory molecules presented by APC and secreted by activated T_{h1} cells will ensure an optimal and long-lasting cytotoxic response.

Professional APCs, especially DCs, are able to perform cross-presentation, resulting in the presentation of exogenous peptide on class I MHC molecule. This phenomenon is especially useful for antiviral immunity, where viruses can inhibit antigen presentation in infected cells. Phagocytosis of these infected cells by professional APCs will still lead to cross-priming of CD8⁺ T cells against viral peptides, allowing for an efficient cytotoxic response [32].

1.1.5.1.6 Memory T cells

The functions described above are all effector T cell functions. While the majority of T cells are heading toward an effector phenotype upon activation, a fraction will differentiate into memory T cells, facilitating immunosurveillance and recall responses to reinfection. The main features of these long-lived cells are low or none effector function, but strong proliferative and activating capacity when re-encountering the antigen.

Memory T cells can specialized into a tissue-resident subset, leading to improved local immunity in most organs [33]. These cells will retain migration potential but only within their respective tissue, and at a slow pace, enabling them to scan their surroundings for recrudescence or reinfecting pathogens. The local memory response enables a rapid first response before the recruitment of memory T cells from the circulation.

1.1.5.1.7 Immune Checkpoints and T cell exhaustion

The majority of immune functions presented above are part of an activation cascade that must be kept under control to avoid its propagation, which could lead to autoimmunity and tissue damage. In cases of acute inflammation, the source of immune activation is quickly resolved, bringing the level of inflammation under control. However, in the case of chronic pathology leading to chronic antigen stimulation of adaptive immune cells, mechanisms limiting immune overactivation are needed.

In T cells, the amplitude and quality of the response are dictated by a balance of co-stimulatory and inhibitory signals transmitted via proteins called immune checkpoints and presented in Figure 1.4 [34]. Typically, these signals are triggered following the interaction of a T cell membranous receptor with its ligand, often carried by an APC. For instance, the previously described co-receptor CD28 binding to CD80 or CD86 will enable T cell activation during antigen recognition. Among inhibitory immune checkpoints, PD1, TIM3 and LAG3 are among the most studied. Recognition of their ligand (such as PDL1 for PD1) will lead to a dysfunctional state, often referred to as exhaustion.

Exhausted T cells are characterized by loss of effector function, such as decreased cytotoxic capacity for CTLs or decreased capacity to form memory T cells. While being counter intuitive, the presence of exhausted T cell has been proven essential to maintain an immune response in the case

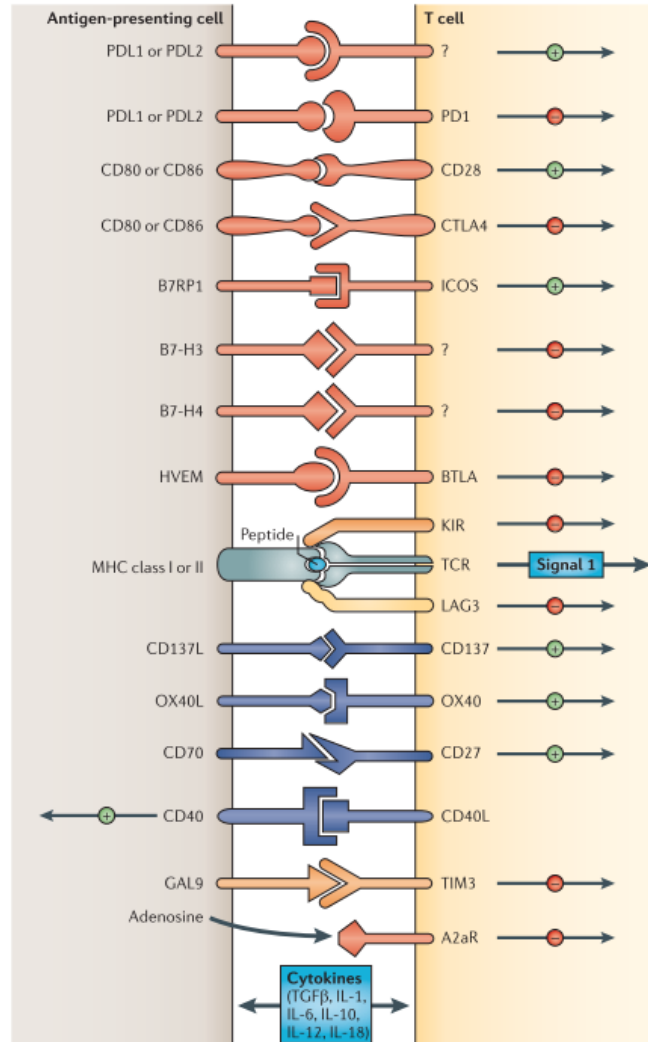


Figure 1.4: Activating and inhibitory checkpoints presented on the surface of T cells (right) and APC (left)
From [34]

of chronic pathologies, as chronic TCR stimulation of classical (i.e. non exhausted) effector T cells will trigger apoptosis (a mechanism helping to achieve self-tolerance in physiological conditions).

In the CD8 T cell subset, a precursor population, expressing PD1 and the transcriptional factor TCF1 (and requiring the transcriptional factor TOX), has been shown to be crucial in the generation of terminally exhausted T cells such as the TCF1⁺PD1⁺TIM3⁺ subset [35]. TOX expression being regulated by the length of TCR stimulation, antigen presentation and viral load, this subset is only observed in chronic situations [36]. TCF1 also appears to be essential in the formation and function of central memory CD8⁺ T cells after resolution of acute infections, the generation of T_{fh} and the maintenance of T_{reg} regulatory functions [37].

The TCF1⁺PD1⁺CD8⁺ T cell subset, precursor of exhausted T cells, persist and function outside secondary lymphoid organs, in tumor tissues for instance [38], and as such, can be compared with tissue-resident memory T cells. This subset, while lacking effector function, will differentiate into

1.1. INTRODUCTION TO IMMUNOLOGY

terminally exhausted CTLs that can contribute to viral or tumor control through diminished but still present effector functions.

In conclusion, T cell exhaustion should be seen as a process that prevents immune-mediated pathology while retaining some level of effector function, keeping chronic disease under control [39]. T cell exhaustion constitutes an essential physiological adaptation to chronic infections rather than a functional impairment of the immune system.

1.1.5.2 B lymphocytes

1.1.5.2.1 BCR structure and B cell development

B cells are the other arm of the adaptive immune response. In contrast to T cells, they undergo their first rounds of differentiation and maturation within the BM, from VDJ recombination to negative selection, eliminating the majority of autoreactive B lymphocytes. Their immunoreceptors are called immunoglobulins (Ig), and are formed by a pair of heterodimers linked by a disulfide bond. Each heterodimer is composed of a light and a heavy chain, associated to create the antigen binding site, which is highly variable, and a constant region called Fc fragment. The variable region is the result of somatic recombination defining the antigen binding site and allowing each B cell clone to be highly specific and unique. Ig can take a membranous or a soluble form, called respectively BCR and antibody. Ig are also characterized by several constant region among 9 (IgD, IgM, IgG1, IgG2, IgG3, IgG4, IgA1, IgA2, IgE), called isotype and associated to various effective functions. In contrast with TCR, the BCR can recognize its target without requiring MHC-associated presentation and can bind antigens that are not necessarily proteins. However, activation of B cells requires the presence of T helper cell, helping to preserve peripheral tolerance by avoiding autoreactive B cells from turning into effector cells.

Similarly to T cells, once mature, naive B cells will recirculate through blood toward SLOs and TLSs, where they will possibly recognize their cognate antigen, leading to activation and differentiation into memory or antibody producing cells, called plasma cells.

1.1.5.2.2 B cell activation and differentiation

Naive B cells express IgM and IgD isotypes. Upon BCR stimulation, B cells will upregulate class II MHC molecules, as well as the costimulatory protein CD80/CD86. Internalization of BCR / antigen complexes and presentation of the same peptide on class II MHC molecules will lead to auxiliary CD4 T cell activation [40]. The activated T helper (T_h) cell will then in return express CD40L, activating B cells through CD40 stimulation while CD80 / CD86 activates T_h cells through the CD28 / CD80 or CD28 / CD86 interaction. Finally, auxiliary T_h cells will produce cytokine while the B cell will express its receptors, leading to B cell proliferation and differentiation.

In lymphoid organs, some B and T cell couples (specific for the same antigen) will migrate to primary follicles and proliferate to form a germinal center (GC). In this GC, proliferating B

1.1. INTRODUCTION TO IMMUNOLOGY

cells will undergo somatic hypermutation, Ig switch and functional differentiation. Somatic hypermutation will affect the variable portion of the BCR, potentially increasing the binding affinity to their cognate antigens. B cells carrying newly formed BCR not competent for antigen recognition or with autoreactive properties will be counterselected leading to their apoptosis. B cells with increased affinity will undergo Ig switch, exiting from GC reaction producing different isotypes such as IgA, IgG or IgE depending on the immune context. Lastly a pool of memory B cell and long-lived plasma cells will be generated from GC with high affinity BCR and the ability to be rapidly activated in case of re-exposure to pathogenic antigens [41].

Memory cells circulate in the blood and tissues, are long-lived and present surface Ig and class II MHC, and still display potential for proliferation, somatic hypermutation and isotype switch. Their presence allows for a fast adaptive immune reaction in the event of a new encounter with the same pathogen. On the other hand, long-lived plasma cells are terminally differentiated cells which will reside in niches in the BM and mucosal tissues. Their main function is to produce large amount of soluble antibodies for long term protection of the organism [42].

1.1.5.2.3 Antibody mediated immunity

The binding of soluble antibodies to their target can have various effects:

- Neutralization of bacterial toxins which action rely on their capacity to enter host cells. By binding to the toxin while in the extracellular space, IgG (within the extracellular fluid) and IgA (at the mucosal surface) antibodies will block their binding site to the cell surface receptor, making them unable to enter the cell, thus ineffective.
- Neutralization of viral and bacterial infections. By the same mechanism, i.e. blocking the binding site to cell surface receptor, antibodies can prevent the entry of viruses and bacteria into their target cell or nullify bacteria whose pathogenic potential relies on their capacity to remain attached to the cell surface as extracellular pathogens. IgG and IgA are again the main effector of this antibody mediated effect.
- Complement activation through C1q. As presented earlier, the complement system can be activated through various pathways. In the classical pathway, the C1q protein will bind to the constant region (called Fc domain) of antibodies (IgG or IgM) attached at the surface of a pathogen, triggering the complement cascade and leading to increased inflammation, membrane attack complex formation and pathogen opsonization and phagocytosis.
- Formation of immune complex leading to phagocytosis. Immune complexes are the combination of antigens and their cognate antibody, which can be carried by toxins, debris from dead host cells and microorganisms. Phagocytic cells carry an Fc receptor that will enable them to detect these immune complexes and engulf them, leading to pathogen destruction. Fc receptor are not activated by the binding of antibodies.

1.1. INTRODUCTION TO IMMUNOLOGY

- Antibody dependant cell cytotoxicity (ADCC). Cells undergoing intracellular infection usually present non-self-antigens bound to class I MHC molecule, triggering CD8⁺ T cell cytotoxicity. Infected cells can also present foreign proteins at their surface (i.e. viral protein resulting from the fusion of the viral membrane with cell membrane), that can be recognize by antibodies. Host cells carrying antibodies will trigger the FC receptor on the surface of NK cells, leading to recognition and destruction of the cell in a process called ADCC.

In summary, the various components of the innate and adaptive immune response present specificities that allow the immune system as a whole to mount an efficient and timely immune against the majority of pathogens, as presented in Table 1.1.

Table 1.1: Phases of the immune response. From [12]

Response	Mechanisms	Time after infection	Duration of response
Innate immune response	Inflammation, complement activation, phagocytosis, and destruction of pathogen	Minutes	Days
	Interaction between antigen-presenting dendritic cells and antigen-specific T cells: recognition of antigen, adhesion, costimulation, T-cell proliferation and differentiation	Hours	Days
Adaptive immune response	Activation of antigen-specific B cells	Hours	Days
	Formation of effector and memory T cells	Days	Weeks
	Interaction of T cells with B cells, formation of germinal centers. Formation of effector B cells (plasma cells) and memory B cells. Production of antibody	Days	Weeks
	Emigration of effector lymphocytes from peripheral lymphoid organs	Days	Weeks
	Elimination of pathogen by effector cells and antibody	Days	Weeks
Immunological memory	Maintenance of memory B cells and T cells and high serum or mucosal antibody levels. Protection against reinfection	Days to weeks	Up to lifelong

1.1.6 Physiological Colorectal Immunity

The necessity for an efficient but controlled immune response is especially true in colorectal immunity, where intestinal microbial colonization and food-derived antigens define an especially challenging environment. Consequently, the intestine presents a very developed immune system that will contribute to maintaining a homeostasis state with immune tolerance to food antigens and commensal microbiota, while retaining the capacity to fight off infections and harmful pathogens.

The first level of defense at the intestine level is the epithelial barrier which, in addition to playing its role as a physical barrier, sends signals to the mucosal immune system, such as cytokines and chemokines. Intraepithelial lymphocytes, which can be $\alpha\beta$ or $\gamma\delta$ T cells, also play a critical role in first line of defense, protecting the epithelial barrier from injury and regulating inflammation through IFN- γ production. DCs are also a major player in keeping the gut microbiota in check. They are able to pass antigen from the lumen to the lymphoid structures presented below, promoting the differentiation of T cells into T_{reg} or T_h cells, limiting local inflammation at steady state or eliminating pathogens during the host defense reaction, respectively [43].

Under the epithelial cells is the lamina propria, rich in B cells, T cells and macrophages. In response to signals from the epithelium, T cells will promote inflammatory and anti-inflammatory responses. SLO and TLS can also be found at the lamina propria level, which will generate B and plasma cells, whose IgA production will be critical for gut barrier protection. Intestinal epithelium and lamina propria are sites where antigen-experienced lymphocytes accumulate and can persist in the long term as committed effector or regulatory cells.

Two main SLOs are present in the gut, mesenteric lymph nodes, and gut-associated lymphoid tissues (GALT). The latter are directly localized at the mucosa level as seen in Figure 1.5a, enabling the priming and differentiation of adaptive immune cells directly at the effector site, promoting barrier integrity and protective immunity [44]. GALT include multifollicular lymphoid organs, such as Peyer's patches in the small intestine or the more seldom colonic patches in the colon, as well as multiple solitary intestinal lymphoid tissues (SILT). Unlike patches, SILT development is initiated during the early postnatal phase and is partly regulated by the intestinal microbiota. SILT can take several shapes, from small patches of T and dendritic cells, called cryptopatches to large mature isolated lymphoid follicles with a germinal center called isolated lymphoid follicles (ILF) [45]. ILF are present in both the small and large intestines, for an estimated total of 30 000 ILF in the human intestine.

GALT are separated from the intestinal lumen by a specialized follicle-associated epithelium, which play a key role in transferring luminal antigens into the GALT through transcytosis [46]. Although mouse models have shown a critical role for GALT in protecting the intestine against mucosal pathogens, no clear demonstration has been performed in humans. It has been shown that intestinal inflammation can drive the expansion of intestinal lymphoid structures, and this new lymphoid organs seem to be TLSs, contrary to SILT [47].

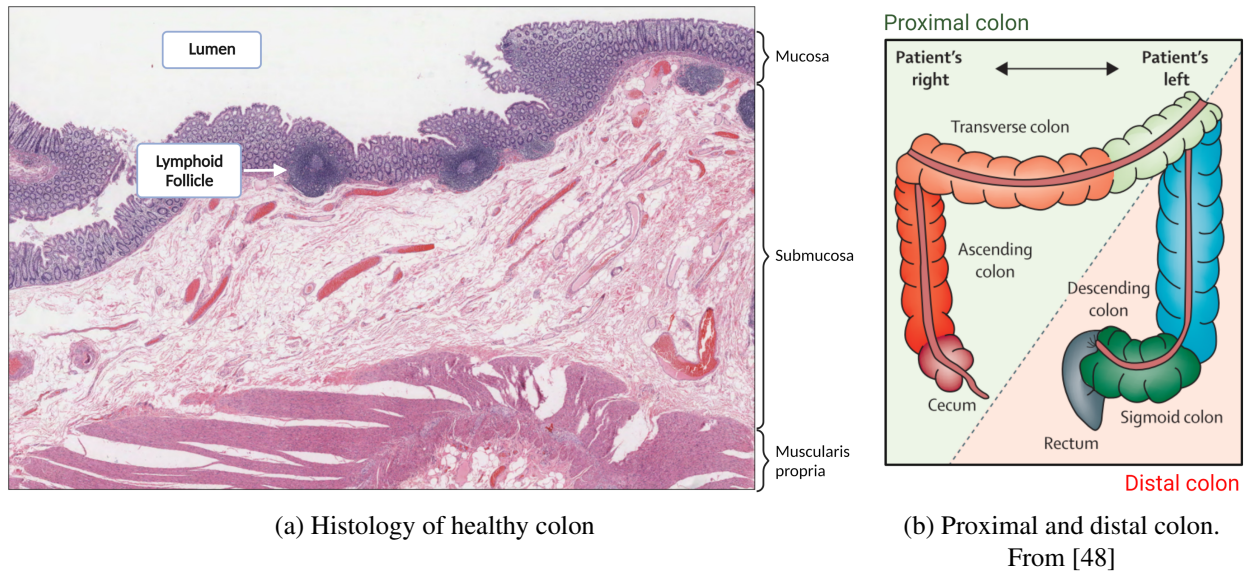


Figure 1.5: Colorectal anatomy

While the distribution of the multifollicular lymphoid organs of GALT highlights a clear difference between the small and large intestine, major differences also exist within the large intestine, which can be divided into two compartments, as seen in Figure 1.5b: the right or proximal colon, from the cecum to a point approximately half to two-thirds of the way along the transverse colon, and the left or distal colon, which includes the remaining large intestine to the rectum. This differentiation finds its basis in the embryonic development as distal and proximal colon present a different embryonic origin, leading to clear different gene expression profiles [49]. The microbiota of these two regions is also distinct, with more diversity in the proximal colon.

When it comes to the immune microenvironment between these two compartments, it has been shown that they harbor different types of ILF [46], as well as different overall immune profiles, which is not surprising given the symbiotic relationship between colon immunity and the microbiome [50]. Notably, T_H17 and follicular B cells proportions are decreasing when going from the cecum to the sigmoid colon, while activated CD4 T cells and IgA plasma cells are increasing.

In conclusion, in physiological conditions, the colon mucosal immune system is composed of the epithelial barrier and its associated immune cells and lymphoid structures. These structures include mesenteric lymph nodes, which are at a distance from the mucosa, and local lymphoid structures, directly between the epithelial barrier and the lamina propria. This very developed and structured immune system is crucial to keep the gut microbiota under control. In healthy conditions, commensal microbiota is regulating the maturation of the mucosal immune system while being under the surveillance of the latter, leading to the well-known symbiotic situation. Deregulation of this equilibrium, coming from microbial pathogens or abnormal inflammation, will often result in intestinal disease [51].

1.2 Colorectal Cancer

As demonstrated in the previous section, the immune system is crucial in identifying and eliminating pathogenic agents based on recognition of self, maintaining integrity and health of the body. However, in the case of cancer, the establishment of the immune response is not as straightforward. Indeed, tumor cells originate from healthy cells and, as such, are very similar to the latter at first. They will slowly become more and more estranged during carcinogenesis. While it is thought that the immune system is capable of early elimination of the vast majority of sporadic cancer, some will still escape immune surveillance and develop into full blown cancer through numerous mechanisms. In order to better understand how cancer can escape the immune response, we have first to define cancer, focusing on the case of colorectal cancer. We will also present the main of colorectal carcinogenesis, from healthy colon to metastatic cancer. Finally, we will describe how CRC is currently detected, classified and treated.

1.2.1 Cancer Definition

Cancer can be defined as a disease caused by an uncontrolled division of abnormal cells in a part of the body. The basic concept underlying this abnormal growth that will lead to tumor (or neoplasia) development is that each cell cycle is associated with a sporadic mutation rate. Through normal cell division, mutations that can give an advantage to a given cell can appear, causing this cell and its clonal progeny to get an edge compared to its neighbors. Tumor development can then be seen as an evolutionary process [1]. Mirroring selective pressure of Darwin's evolution theory, these abnormal cells with a developmental edge will slowly outgrow their neighbours, forming a mass of mutated cells. These cells, which proliferate at an abnormally high rate, can continue to accumulate mutation. The selective pressure will then take place within the tumor, where some subclonal population will be fitter for their environment, getting an edge on other tumor cells. This phenomenon will lead to tumor heterogeneity, where several subclonal populations cohabit under the laws of natural selection. As a consequence, carcinogenesis will lead to a tumor adapted to the constraints of its environment, such as access to nutrients and surveillance by the immune system.

Along with the study of cancer properties, development, and interactions with its host, several characteristics necessary for carcinogenesis have been identified, helping to better understand cancer and unveiling potential therapeutic approaches. Hallmarks of cancer have been introduced in 2000 [52] and refined since [53, 54] following advanced in the oncology field. The idea behind these hallmarks is that for a normal cell to progress to a neoplastic disease, it acquires several traits in a multistep process accompanying carcinogenesis presented in Figure 1.6.

While the first Hallmark edition was very tumor-centric, it is now clear tumor were not developing isolated from other cells of the body, but in constant interaction with other components present at the tumor site such as blood vessel, necessary for nutrient and oxygen supply, fibroblasts or immune cells. As a consequence, updated hallmarks integrate parameters associated with the tumor

1.2. COLORECTAL CANCER

microenvironment (TME), taking into account these new parameters.

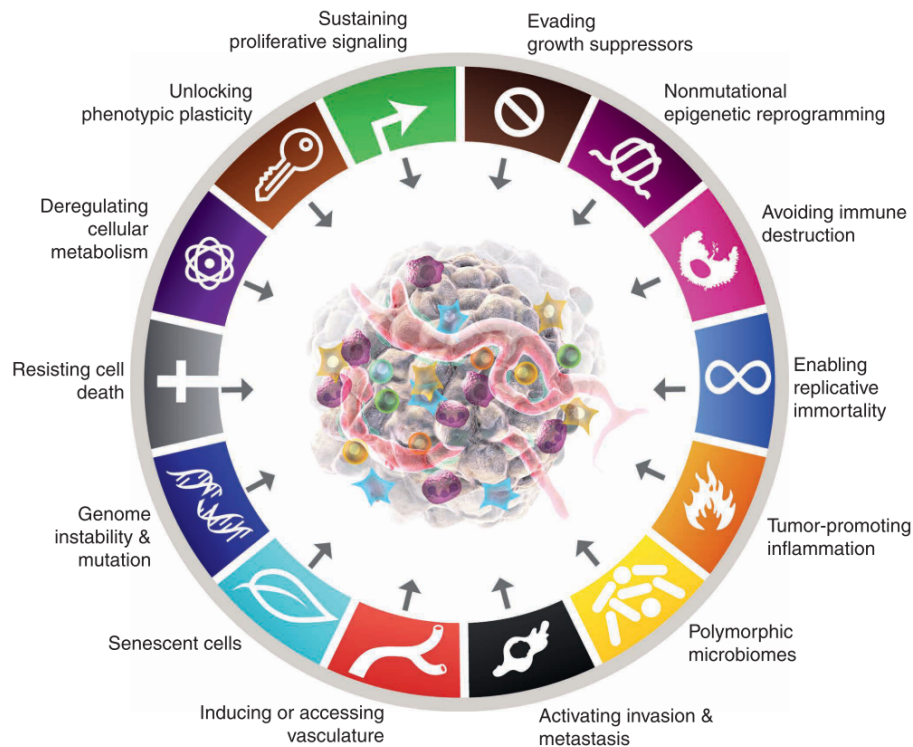


Figure 1.6: Hallmarks of cancer. From [54]

These hallmarks are acquired through mutations affecting either oncogenes (gain of function mutation) or tumor suppressor genes (loss of function mutation), called driver mutations, that will lead to the tumor acquiring critical properties for its survival and development. In consequence, the cancer hallmark *genome instability and mutation* is presented as an enabling hallmark, leading to the accumulation of potentially driver mutations. In normal cells, numerous safeguard mechanisms are present to keep mutation occurrence at a very low rate. Several driver mutation such as loss of function in a gene coding a component of genomic maintenance machinery or a mechanism of DNA repair can lead to faster mutation accumulation and boost carcinogenesis. For example, mutations in the *adenomatous polyposis coli (APC)* gene, involved in chromosome segregation during mitosis, can lead to a genome instability called chromosomal instability (CIN) [55]. Another example is a defect in the mismatch repair system, one of many DNA repair mechanisms, leading to genomic instability called microsatellite instability (MSI), which is frequently observed in colorectal, endometrial, and gastric adenocarcinomas [56].

As one of the main characteristic of cancer is its capacity to proliferate exponentially, another fundamental trait of driver mutations is to allow tumor cells to proliferate in an uncontrolled way, leading to tumor growth and accumulation of mutations.

1.2. COLORECTAL CANCER

1.2.1.1 Hallmarks of tumor cell proliferation

This aspect encompasses several compatible traits. The first one is the presence of *sustained proliferative signaling*, making tumor proliferation independent of extrinsic growth factors. This feature can be achieved through production of growth factors by the tumor cell itself, or induction of this production by cells present within the TME, leading to autocrine or paracrine proliferative stimulation, respectively. Upregulation of the growth factor receptor level or constitutive activation of the growth signaling pathway can also lead to this result. For instance, mutation of oncogenes *BRAF* or *KRAS* can lead to constitutive activation of the *MAPK* pathway, which is responsible for transmitting extracellular proliferative signals to the nucleus of receptive cells [57]. The PI3K-mTOR pathway interacts with the MAPK pathway to regulate cell survival, differentiation, proliferation, metabolism, and motility. Somatic mutations inducing activation of PI3K (e.g. mutation in *PI3KCA*) are known to potentially induce neoplasia and are found in numerous cancers [58].

In the same vein as the hallmark *sustained proliferative signaling*, tumor cells can also favor cancer growth by *evading growth suppressor signals* by losing function in a proliferation suppressor such as TP53 or RB proteins, which are central regulators, controlling the balance between proliferation or senescence and apoptotic programs. Loss of function within the contact inhibition or TGF- β pathway can also lead to evasion of growth suppressor signals.

Resisting cell death and *enabling replicative immortality*, two other critical hallmarks of carcinogenesis leading to overproliferation. The most common mutation associated with evading apoptosis is the loss of the p53 protein, but similar results can be obtained through increasing expression of antiapoptotic regulators such as Bcl-2 or survival signals, as well as downregulating proapoptotic factors. Besides evading apoptosis, replicative immortality is often associated with increased telomerase activity, avoiding the telomere shortening that will lead to cell death in normal cell proliferation.

1.2.1.2 Other hallmarks of cancer

In order to sustain their growth, tumors require an increase in nutrient and oxygen intake. The most commonly observed mechanism that will address this need is the *induction of angiogenesis*, often through the secretion of vascular endothelial growth factor (VEGF) which will lead to neovascular development and increase the delivery of nutrients and oxygen to the tumor site.

This neovascularization will also promote tumor spread throughout the body, leading to the metastatic stage. In order to go through the *activate invasion and metastasis* stage, tumor cells have to acquire several qualities leading to invasion of nearby tissue compartments, intravasation and transit in nearby blood or lymph vessels, and extravasation into the parenchyma of distant tissues, forming new colonies.

The two hallmarks of cancer involving the immune system, *avoiding immune destruction* and *tumor-promoting inflammation* will be further discussed in the section 1.3

1.2.2 Epidemiology of Colorectal Cancer

1.2.2.1 Incidence

Colorectal cancer (CRC) is the third most common cancer worldwide and the second most deadly with almost 2 million new cases and 670 000 associated deaths in 2018 [59]. In Europe in 2020, with nearly 520 000 new cases and 245 000 deaths, CRC was considered to be the second most common in both cases and death [60]. CRC accounts for approximately 10% of all diagnosed cancers and cancer-related deaths worldwide. It is worth noting that its incidence and mortality rate are 25% lower in women than men [48], and CRC incidence is lower in men under 50, but a concerning rise in this age class has been observed recently. The rates of CRC are the highest in the most developed countries, where the trend is toward stabilization or decrease in incidence, while an increase is observed in developing countries. Screening strategies and risk factors can account for most of these discrepancies.

In France, the overall 5-year survival rate for CRC is 64% for women and 62% for men [61]. However, this number drastically varies depending on cancer stage: if the cancer is diagnosed at a localized stage, the survival rate is 91%, if the cancer has spread to surrounding tissues or organs and/or the regional lymph nodes, the 5-year survival rate is 72% and colon cancer has spread to distant parts of the body, the 5-year survival rate is 14%. These numbers strongly highlight the need for an efficient screening system for CRC, with the aim of detecting the disease in its earlier stages.

1.2.2.2 Risk factors

It is estimated that up to 20% of European CRC cases can be attributed to the failure to follow various guidelines for a healthy lifestyle (no smoking, low alcohol intake, healthy weight [60]). These recommendations attempt to address environmental risk factors for CRC. These numerous modifiable risk factors are presented in Figure 1.7, together with non-modifiable ones such as male sex or advanced age, both associated with an increased risk of disease.

Family history is also an important risk factor, which is said to account for 10-20% of all CRC patients. A subgroup of these patients is affected by an hereditary CRC syndrome that encapsulates several diseases described in Section 1.2.3 and accounts for 5–7% of all CRC cases. Patients who have a history of inflammatory bowel disease or colorectal cancer or adenomas also have an increased risk of developing CRC.

1.2.3 Early Carcinogenesis: From Healthy Colon to Carcinoma *in Situ*

Most solid tumors are growing from preexisting benign (i.e. non invasive) lesions, called precancerous or premalignant lesions. As cancer is often diagnosed at an advanced stage, leaving precancerous lesions rarely detected and even more rarely surgically removed, fewer is known about the pre-invasive stages, where the tumor is still restricted to the mucosa. A possible approach to

1.2. COLORECTAL CANCER

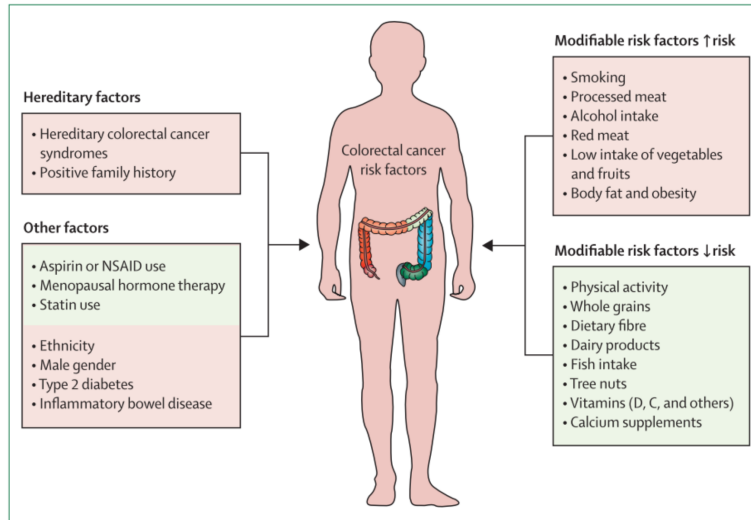


Figure 1.7: Risk factors of colorectal cancer. From [48]

better understand early carcinogenesis is to use sequencing data from invasive tumor samples and try to recreate the evolutionary history of the lesion based on its accumulated mutations [62]. This approach revealed that if late tumor stages present an increased diversification of driver genes and genomic instability, driver mutations often precede diagnosis by many years, if not decades. The presence of driver mutations such as *TP53* and *KRAS* was found in the precancerous events of the majority of cancers, and in CRC, *APC* was shown to have the most chances to appear early, followed by *KRAS* and *TP53*. However, these results are based on inference, and need to be validated through empirical observation when possible.

Colorectal, head-and-neck and cervical premalignant lesions, among a few others, can be seen as good models for precancerous studies because they can be accessed without heavy surgical intervention and are, in most developed countries, screened for and surgically removed in order to prevent cancer development. Studying this type of lesions could provide valuable information to understand the mechanisms that occur during early carcinogenesis [8].

1.2.3.1 Colorectal premalignant pathways

A study characterizing the landscape of somatic mutation in normal colorectal epithelial cells [63] highlighted that even in healthy colorectal mucosal crypts, potential driver mutations were present even if at a very low rate (<1%). This demonstrates that premalignant lesions and CRC are the scarce result of an omnipresent process first taking place in normal colorectal mucosa.

Among the various dysplasia, or polyps, that emerge in colorectal tissues, two lesions are responsible for the vast majority of CRC: conventional adenomatous polyps (APs), emerging from the adenomatous pathway, while sessile serrated lesions (SSLs) that arise from the serrated pathway. Each pathway is associated with distinct histological features, mutation sequence, molecular landscape (MSI, CIN...) and origin.

1.2. COLORECTAL CANCER

1.2.3.2 The Adenomatous Pathway

Being responsible for 70 to 80% of CRC and used during the last 30 years as a model to understand CRC pathogenesis, the adenomatous pathway is often presented as the conventional pathway. Macroscopically, most adenomas are prominent into the colorectal lumen, either sessile with broad attachment or on a stalk [64]. Histologically, adenomas are defined by the presence of dysplasia (abnormal proliferation) of various grade, that will enable pathologist to classify APs into low and high-grade adenomatous polyps (LG AP and HG AP, respectively). Their size (below or above 10 mm) and histology (tubular or villous, the latter being associated with an increased risk of CRC) are associated with a risk of progression to adenocarcinoma [65]. Underlying this abnormal proliferation is the accumulation of mutations, with several genes clearly associated with this pathway: *APC*, *KRAS* and *TP53*.

The first model suggested [66] proposed a very clear sequence, the first step being a mutation in the tumor suppressor gene *APC* [67], resulting in a CIN phenotype and activation of the WNT pathway (where APC plays a critical role). WNT is an essential pathway in normal intestinal function, especially implicated in the self-renewal functions of epithelial stem cells located at the base of intestinal crypts [68]. Overactivation of WNT leads to intestinal hyperplasia, and CIN induce genome instability, paving the way for carcinogenesis. In this model, the mutation *APC* is followed by gain of function mutations in the *KRAS* oncogene and loss of *TP53*, another tumor suppressor gene. However, it is now known that if these mutations are overrepresented in CRC, the adenomatous pathway can also progress without the full complement of driver mutations [69] and numerous other oncogenes and tumor suppressor genes can be driving adenomatous carcinogenesis. For example, mutations in *PIK3CA*, common in CRC [70], were found to be associated with the adenomatous pathway but in a subset of APs non mutated for *KRAS* [71].

Along the accumulation of mutations, the neoplasia will gain several properties, leading to APs size increase, apparition to severe dysplasia associated with loss of tissue structure, up to the carcinoma *in situ* stage, where the invasion is still limited to the colorectal mucosa, as seen in Figure 1.8. Next steps will lead to tissue invasion, entering the infiltrating adenocarcinoma stage. The adenomatous pathway is often associated with the microsatellite stable (MSS, as opposed to MSI) and CIN phenotypes, and so are CRC resulting from this pathway. ADs are more often found in the distal (or left) colon and develop from WNT-driven expansion of stem cells situated at the base of intestinal crypts.

1.2.3.3 The Serrated Pathway

An alternative route to CRC is the serrated pathway, which accounts for 20 to 30% of total CRC cases, and has been characterized more recently. Unlike APs, serrated polyps do not protrude from the colorectal lumen, but rather flat to sessile, making them easier to miss during screening [64].

1.2. COLORECTAL CANCER

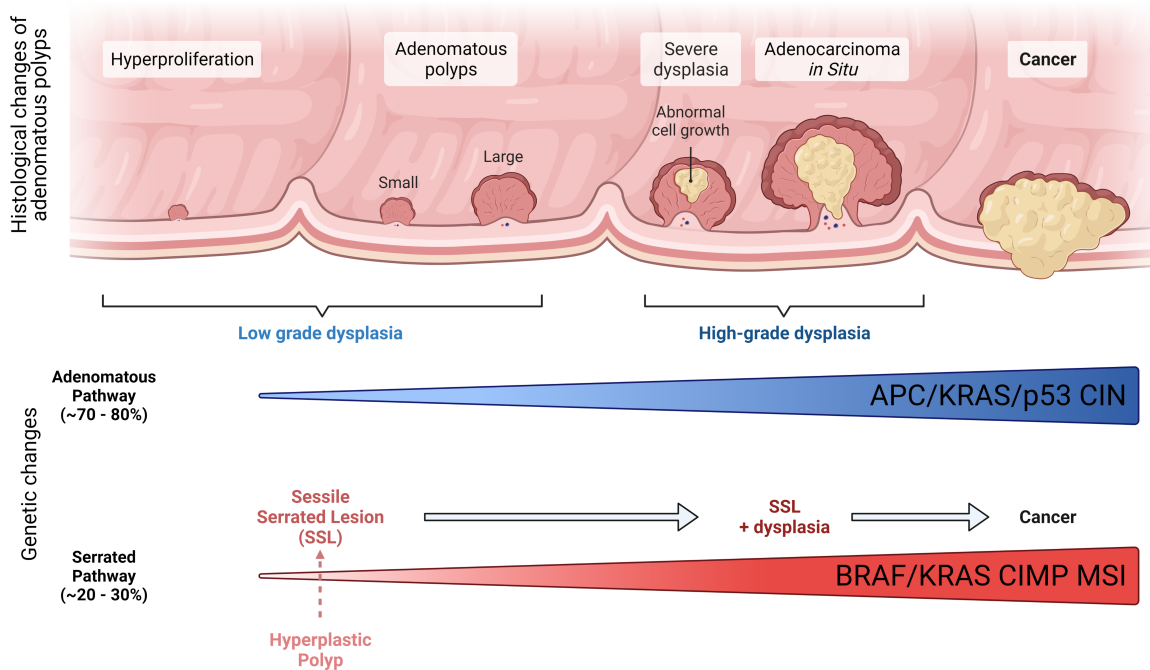


Figure 1.8: Precancerous pathway of colorectal cancer

Polyps from the serrated pathway can be divided into 3 categories: Hyperplastic polyps (HP), Traditional serrated polyps (TSA) and sessile serrated lesions (SSL), previously called sessile serrated adenomas or sessile serrated polyps. As HP does not have malignant potential and TSA are very rare, we will mainly focus on SSLs. It should be noted that some MVHPs (a subgroup of HP) may progress to SSLs and thus gain malignant potential. SSL can present the MSI or MSS phenotype, and only few of them (4-8%) present dysplasia [72].

Mutations in the *BRAF* oncogene are a major and early step in SSLs carcinogenesis, leading to increased proliferation and resistance to apoptosis. Gain of function mutation in *BRAF* leads to widespread methylation of CpG islands, called the CpG island methylator phenotype (CIMP) [73]. These epigenetic changes induce silencing of numerous genes, among which tumor suppressor genes. For instance, *MLH1* overmethylation leads to loss of the protein MLH1, which is part of the mismatch repair machinery, resulting in MSI SSLs.

Similarly to the adenomatous pathway, SSL progression is associated with activation of the WNT signaling pathway. However, unlike ADs where *APC* mutations are found in more than 90% of cases, only a small proportion (<10%) of SSLs harbor a mutation in this gene and WNT overactivation is associated with other causes.

In conclusion, CRC developed from SSLs are most of the time *BRAF* mutated and present the CIMP phenotype, while only part of these serrated-derived tumors are MSI. While ADs have been shown to develop from stem cells at the basis of epithelial crypts, a recent publication hypothesized that the origin of SSLs is very different: epithelium damage, which could be microbiota-related,

1.2. COLORECTAL CANCER

would activate a regenerative program leading to metaplasia, a process by which differentiated cells transition into cell types non-native to the tissue. This regenerative program would select survival / proliferative pathways, such as activating BRAF mutations, initiating serrated neoplasia [74].

Also in opposition to ADs, SSLs are more localized to the proximal (or right) colon.

1.2.3.4 Epidemiology of colorectal polyp in the average risk population

A European study including 12000 average-risk individuals aged 55-64 years who underwent colonoscopy found polyp(s) in 48% of them [75]. Another study, on a Japanese population with a median age of 48, found polyps in 20% of 22 395 individuals [76]. These discrepancies in the incidence of neoplasms emphasize the importance of risk factors such as age, environmental factors, and genetics.

Although most studies do not quantify the number of polyps per patient, some focused on using this number as an indicator of the quality of clinical colonoscopy or a predictive marker of recurrence. A study found polyps in 58% of 1937 individuals (median age 64) and reported their polyp number. For polyp-positive patients, the mean number of detected neoplasms was 2.58. Specifically, 43% had only one polyp, 72% less than 4, and 14% 5 or more [77]. Studies investigating the potential of the number of polyps in the first colonoscopy as a predictor of recurrence demonstrated that these patients with a large number of polyps 3 + (or 5 +) had a higher risk of harboring polyps [78] (or advanced polyps [79], respectively) during the next colonoscopy. A French study provided additional evidence that the number of initial colorectal polyps is useful in predicting the risk of polyp recurrence [80].

As all of these studies excluded patients with risk factors such as known hereditary syndromes, advanced CRC, presence of inflammatory bowel disease or history of colorectal resection, this implies that patients at high risk of developing polyps can be found in the average-risk population, for reasons not fully understood at the moment.

1.2.3.5 Hereditary colorectal cancer syndromes

While the majority (90-95%) of CRC are sporadic in origin, approximately 5% to 7% of cases are associated with germline mutations that confer an inherited predisposition to cancer. The diagnosis of these diseases is very important as it provides the patient with an optimal surveillance strategy to prevent colorectal cancer and appropriate surveillance advice for relatives at risk. The two main syndromes, Lynch and FAP, will be presented below. Other less common syndromes exist, such as MYH-associated Polyposis or Hamartomatous Polyposis. [81].

1.2.3.5.1 Lynch

The Lynch syndrome is caused by an inherited dysfunction of the mismatch repair system, leading to the development of numerous MSI adenomatous polyps, which share the appearance of a sporadic adenoma. Although not as obvious as the FAP syndrome under physical examination, the fact that

1.2. COLORECTAL CANCER

this syndrome is the only known source of MSI APs (sporadic APs are always MSS) makes it a good marker for its diagnostic. Because the timeline of carcinogenesis is greatly accelerated in these patients, the known lynch patient will undergo colonoscopy every 2 years from 20 to 25 years of age to prevent the development of CRC. Lynch syndrome accounts for 3 to 5% of CRC cases, and its carriers have a 50 to 80% chance of developing CRC over their lifetime. The incidence of Lynch syndrome is estimated to be around 1 per 300 [82].

1.2.3.5.2 Familial Adenomatous Polyposis

This syndrome is the second most common inherited CRC syndrome after Lynch, and is characterized by the development of hundreds to thousands of colorectal adenomatous polyps, due to a germline mutation in *APC*. Patients with this syndrome will develop CRC with a probability of 100% if the colon is not removed. Its incidence is estimated between 1 per 7000 and 1 per 30 000 newborns, accounting for approximately 1% of CRC [64].

1.2.4 Late Carcinogenesis

If not removed, colorectal polyps will keep proliferating and accumulating mutation, resulting in invasion of the submucosa, other compartments and beyond the colorectal wall: this is the invasive carcinoma stage. If left unchecked, cancerous cells will continue to develop, selecting the fittest subclonal population, and may develop the capacity to perform intravasation into the lymph or blood circulation. This will lead to lymph node invasion, which can occur as soon as the carcinoma *in Situ* stage, and finally metastasis, which is considered responsible for 90% of cancer mortality [83].

1.2.5 Screening for Colorectal Cancer

As most deaths from colorectal cancer are associated with metastasis, it is critical to detect and treat this disease as early as possible. However, while some symptoms such as rectal bleeding and anaemia can be associated with premalignant or early colorectal tumors, this disease early stage are mainly asymptomatic [48]. Although the onset of one of these symptoms should lead to a medical check-up, spontaneous screening remains the main tool to effectively fight CRC.

The slow development of CRC (10-15 years), in combination with the non surgical nature of the screening, allows for detection and possible direct removal of precancerous lesions or even early invasive carcinoma stages. Efficient screening programs are by far the most efficient when it comes to reducing CRC incidence, recurrence, mortality and cost of treatment.

The frequency of screening will be adapted after the patient's symptoms, history of polyps, and his family. For example, patients with serrated polyps, particularly SSLs and TSAs, have an increased risk of advanced synchronous and metachronous neoplasia [72] and patients with hereditary polyp syndrome will undergo a more intensive follow-up.

1.2. COLORECTAL CANCER

When it comes to CRC, several screening methods exist. The choice of the appropriate method is based on several parameters: its sensitivity (rate of true positive), its specificity (rate of true negative) and its acceptability from the individual perspective (cost and invasiveness). This last point is critical as, when screening for an asymptomatic disease, cooperation is necessary to develop an efficient screening program [84].

Three main screening tools are currently represented in health guidelines:

- Faecal immunochemical test: aiming to detect potential CRC marker such as blood or molecular marker, completely non-invasive and currently the most widely used test. It is however greatly less sensitive than the other two screening methods.
- Colonoscopy: Although being the most invasive method, its high accuracy (which account for both sensitivity and specificity), and potential to perform biopsies and direct surgical removal of tumors up to early stage of CRC often make colonoscopy the method of choice for cases of elevated risk and patients for whom other screening tests led to suspicion of CRC. In cases of incomplete or inadequate colonoscopy, imaging can be used (e.g. MRI is often used for rectal cancer staging).
- Flexible sigmoidoscopy: Invasive but in a lesser extent than colonoscopy as this visualisation method only screen the sigmoid colon, presented in 1.5b. However, this specificity makes flexible sigmoidoscopy less efficient than colonoscopy for preventing mortality from cancer of the proximal colon [85].

For screening of the general population, non-invasive methods are widely used, often leading to colonoscopy if positive.

After detection and removal of a suspected polyp, several other parameters can be evaluated, guiding the diagnosis and follow-up of the patient. For example, in SSL, high size ($> 10\text{mm}$) and the presence of dysplasia are associated with a higher risk of developing CRC and should lead to closer follow-up. The same can be said in APs for the presence of high-grade dysplasia, size $> 10\text{ mm}$ and villous histology [86]. Immunohistochemistry (IHC) tests can also be performed to assess molecular status such as MSI status (staining for mismatch repair protein in this case). Polyp characteristics (number, histology, location) have also been associated with proximal colon cancer recurrence after polypectomy, highlighting the importance of a precise characterization of these precancerous lesions [87].

These different parameters allow for polyp classification (pathway, size, hereditary syndrome, etc.), leading to an optimized patient care and decrease of CRC incidence and mortality. Similarly, it is critical to use a precise and personalized classification for invasive carcinoma, as the outcome of CRC is highly dependent on the stage of the disease and its molecular characteristics.

1.2.6 Colorectal Cancer Classification

1.2.6.1 The TNM classification system

The consensus classification system for the prognosis of patients with resectable colorectal cancer is the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC) TNM classification system. This system is based on histopathological parameters accounting for the extent of spread of cancer [88, 89]. Developed around 1950 in France, following the observation that survival rates were higher in patient with localized disease, in opposition with patients with metastatic spread [90]. In constant evolution since then, the AJCC / UCC-TNM system is based on the evaluation of three components: first T for the grade of the primary tumor, with T0 being associated with no evidence of the primary tumor, Tis being for carcinoma *in situ*, and T1 to T4 for the size and invasive extent of the invasive carcinoma. For example, invasion through the muscularis mucosa into the submucosa constitutes a T1 disease. The next component of TNM is N, for lymph node metastasis, with N0 for the absence of such and N1 to N3 describing the extent of the invasion. Finally, the parameter M is for the absence or presence of metastasis, T0 and T1, respectively. Subdivisions of some main categories are available for those who need greater specificity, and optional descriptors such as lymphatic or venous invasion can also be used.

The stage classification (form 0 to IV) of CRC stems from the TNM classification, with stage 0 being carcinoma *in situ*, stage I being T1 or 2, stage II T3 or T4. Stages 0 to II are negative for lymph node invasion and metastasis. Stage III matches primary tumor with lymph node invasion and no metastasis, while stage IV include all primary tumor grade with presence of metastasis.

Although powerful, the AJCC/UICC-TNM classification fails to provide complete prognostic information, as shown by the high heterogeneity of outcomes within a given TNM stage. An obvious reason for this is that the TNM classification does not take into account the biological pathways or processes activated within the tumor, and thus lacks biological relevance and predictive power regarding treatment targeting tumor biology.

1.2.6.2 Molecular

Given the limitations of the TNM classification, several other approaches have been developed. Taking advantage of the numerous public multiomics datasets such as the Cancer Genome Atlas (TCGA), molecular classifications based on gene expression data have emerged for most cancer types, highlighting cancer subset associated with specific molecular phenotypes or pathway activation [91]. In colorectal cancer, several molecular classifications emerged, with the Microsatellite Instability (MSI) and the Consensus Molecular Subtypes (CMS) classification being among the most relevant today.

1.2.6.2.1 Microsatellite Instability

As presented earlier, MSI account for 15% of sporadic CRC, and is associated with MMR deficiency and hypermutated phenotype. MSI status is assessed through DNA analysis, looking for 5

1.2. COLORECTAL CANCER

microsatellite markers. Patient negative for all markers are presented as MSS, patient with only 1 positive marker are MSI-low and patient with 2 or more positive marker are MSI-high [92].

Evaluation of MSI status has been shown to be important in the prognostic of the patient, as MSI patients present a higher survival than MSS patients and fewer chances of developing metastases [93]. Moreover, these tumors display specific characteristics, such as a high tumor mutation burden, than can be taken in account when choosing the suitable therapy.

1.2.6.2.2 Consensus Molecular Subtypes

The CMS classification, developed to harmonize several gene expression-based CRC classifications, consists of 4 molecular subtypes with clear distinct characteristics as presented in Table 1.2 [94].

Table 1.2: Molecular Subtypes of Colorectal Cancer. From [94]

	CMS1 MSI Immune	CMS2 Canonical	CMS3 Metabolic	CMS4 Mesenchymal
	14%	37%	13%	23%
Molecular phenotype	MSI, CIMP high, hypermutation	CIN high	Mixed MSI status, CIN low, CIMP low	CIN high
Driver mutations	<i>BRAF</i>	<i>APC</i>	<i>KRAS APC</i>	<i>APC</i>
Activated pathway	Immune infiltration and activation	WNT and MYC activation	Metabolic deregulation	Stromal infiltration, TGF- β activation, angiogenesis
Associated prognosis	Worse survival after relapse	Best survival after relapse		Worse relapse-free and overall survival

CIMP, CpG island methylator phenotype; MSI, microsatellite instability; CIN, chromosomal instability

From a phenotype point of view, the CMS1 subset is associated with MSI and CIMP tumors, while the CMS2, 3 and 4 subtypes are associated with the CIN phenotype. In terms of mutation, the CMS1 group is associated with a mutation of *BRAF*, while CMS 2 to 4 carry a mutation of *APC* and *KRAS*, and CMS3 is particularly enriched in *KRAS*. Concerning the cellular process associated with each subtype, CMS1 tumors are associated with immune infiltration, especially composed of Th1 and CTLs, and display a high activation of immune evasion pathways. The CMS2 pathway is correlated with epithelial differentiation, and activation of WNT and MYC pathways, which are both implicated in CRC carcinogenesis. CMS3 tumors are enriched for multiple metabolic signatures, consistent with the *KRAS* mutation being known to induce metabolic adaptation. Finally, the CMS4 subtype is associated with an up-regulation of genes associated with stromal infiltration, epithelial-to-mesenchymal transition, TGF- β signaling, angiogenesis, and complement-mediated inflammatory system.

As presented earlier, the deep understanding of the biology associated with each of these sub-

1.2. COLORECTAL CANCER

sets makes this classification much more informative and thus powerful than the TNM classification alone. This clinical relevance is highlighted by CMS4 tumors being associated with bad survival, while CMS1 and CMS2 present the worst and best survival after relapse, respectively.

In previous years, most classification attempts in oncology have been challenged by the concept of intratumor heterogeneity (ITH) which is increasingly supported by the accumulation of evidence. The CMS system is not spared and has been updated to take ITH into account [95]. It is now accepted that CMS ITH is frequent and is critical to assess in order to get a better picture of tumor biology, leading to more personalized medicine, and precise diagnostic and predictive power.

AP

CMS classification has been shown to be relevant for immunotherapy decision making [96]. This is not surprising given the immune parameters associated with the different subsets, which highlights the potential impact of the immune system on the classification, prognosis, and prediction of the response to treatment for CRC. However, most accepted cancer classification systems up to now do not take into account immune parameters and spatial parameters such as tumor heterogeneity.

1.2.6.3 Immunoscore

A potential classification system that accounts for the immune parameters of the tumor is the Immunoscore. The founding event of this scoring system was the observation in CRC that several characteristics of the immune tumor microenvironment (such as type, density and location of immune cells within the tumor site) could predict survival of the patient [6]. This predictive power has been shown to be more accurate than the TNM system in CRC for the first time in any type of cancer [97].

According to numerous reports that demonstrate the prognostic value of CD8⁺ T cells in various types of cancer [5], the Immunoscore quantifies the density of CD3⁺ and CD8⁺ cells both in the center of the tumor (CT) and at its invasive margin (IM) as shown in Figure 1.9a. Both marker are deemed low or high in both region of interests (ROIs), leading to a score going from 0 (low CD3 and CD8 infiltrate in both ROIs) to 4 (high CD3 and CD8 infiltrate in both ROIs).

This classification has been shown to predict clinical outcome in early [100] and advanced CRC cases [101]. The performance of this classification system, in association with the emergence of image analysis software able to quantify the cells of interest within tumors, led to the development of an international validation of the Immunoscore assay [99]. This approach demonstrated that the Immunoscore outperformed the TNM classification to predict the risk of relapse for stage I-III colon cancer, regardless of the MSI status as seen in Figure 1.9b [102], leading to the conclusion that an immune component was needed in addition to the classical TNM classification. The World Health Organization (WHO) followed these recommendations, and introduced the immune component, citing the Immunoscore, as an essential and desirable criteria for colorectal cancer diagnosis [103].

1.2. COLORECTAL CANCER

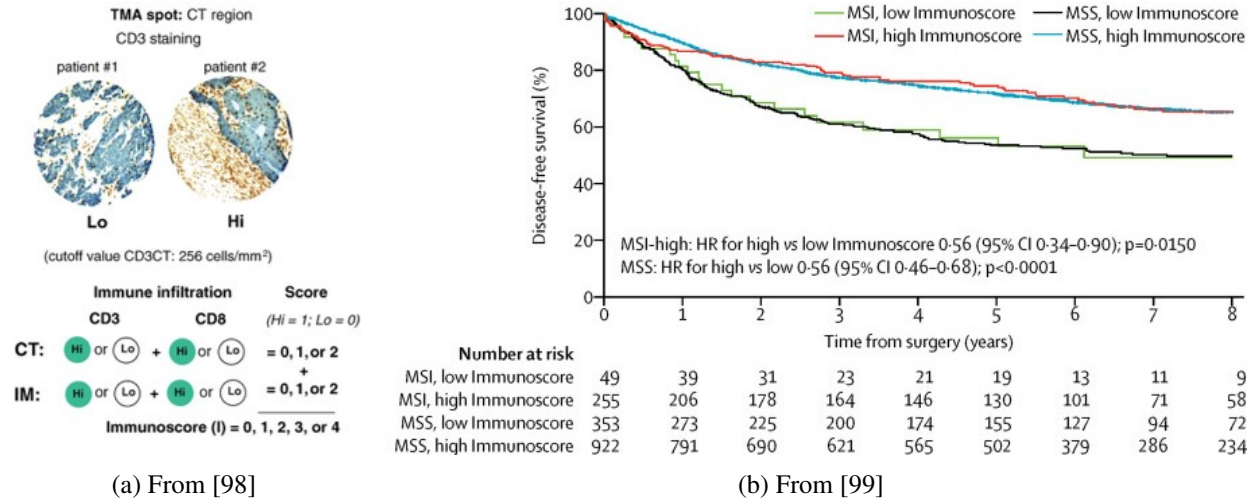


Figure 1.9: Immunoscoring as a prognostic marker in CRC

Since then, Immunoscoring has also made its appearance within Asian and ESMO guidelines [104].

1.2.7 Treatment of Colorectal Cancer

When it comes to CRC treatment, surgical removal is the most common approach. This can be performed through a simple endoscopic intervention for early CRC, such as polyps or some small T1 tumors. Such intervention is often safer, less expensive, and less invasive for the patient, but may require additional intervention to remove potential lymph node metastasis [105].

If endoscopic mucosal resection is deemed not sufficient, classical surgery will be performed in most cases, especially in case of colon cancer. Indeed, rectal surgery is more complicated due to anatomical parameters, which often leads to less invasive approaches.

Radiotherapy, often complemented with chemotherapy (using fluoropyrimidine as radiation sensitiser), is often used as a preoperative treatment. This approach has been shown to reduce the risk of recurrence following surgery [106], and is used for intermediate to high-risk cancers, as determined through MRI. The efficiency of this treatment, sometimes leading to complete clinical response, has led to the development of the rectal watch-and-wait strategy, which avoids the surgery process unless tumor regrowth is detected [107].

Concerning metastatic disease (stage IV), several options are available, both for local and systemic treatment. Locally, low-risk surgery is performed in some organs, such as the liver. However, radiofrequency or microwave ablation remains the preferred option in most cases.

Concerning systemic treatment, fluoropyrimidine-based chemotherapy is used in intermediate CRC (stage III and some stage II), and is often complemented with oxaliplatin. For metastatic patients, systemic therapy is dependant on predictive markers, patient or disease related. For example, the decision between curative and palliative treatment will greatly impact the designated therapy.

1.2. COLORECTAL CANCER

Beyond the generalist therapies presented above, more personalized approaches can be applied, especially in the case of metastatic patients. These therapies often act through modification of the tumor microenvironment or molecular pathway specific to some CRC subsets and must be tailored to the patient.

For instance, anti-VEGF agent will target angiogenesis, depriving the tumor from a appropriate nutrient and oxygen intake. This was the first biologic agent to be proven beneficial for patient metastatic colorectal cancer [108]. Anti-EGFR, a growth factor receptor, is only beneficial for left-sided colorectal cancer, and its efficiency is also dependant on *KRAS* and *BRAF* mutation status, highlighting the need for tumor characterization in personalized medicine [49].

Another category of therapy, called immunotherapies, are aiming to achieve antitumor effects by boosting the immune response directed toward cancer. The main immunotherapy drugs are monoclonal antibodies that block the immune checkpoint molecule, such as anti-PD1 or anti-PDL1, resulting in the reinvigoration of the adaptive immune response and the antitumor effect. This approach has been proven to be particularly effective in MSI tumors, where the Food and Drug Administration approves PD1 blockade (nivolumab or pembrolizumab) [109]. Immunotherapy response can also be influenced by the mutation status of the tumor.

Overall, while more and more promising therapies are emerging, only a subset of patients respond to these therapies. This highlights the need for biomarkers that help predict the response to therapy. In addition to driver mutation status, the previously presented CRC classification can also be used as a predictive marker in addition to a prognostic marker. For example, MSI status, in addition to being critical for response to immunotherapy response, is also associated with stage II patients who do not benefit from adjuvant chemotherapy [110]. The Immunoscore is also being tested as a potential predictive factor for neoadjuvant therapies, using biopsy based Immunoscore [111].

It is also interesting to note that, beyond the Immunoscore, pure immune parameters such as the expression of PDL1, the density of T_{h2} cells and cytotoxic T cells are predictive of the response to immunotherapy, highlighting the need for a deeper characterization of the TME, and especially its immune component, called the immune tumor microenvironment or iTME [112].

One can also notice that, in the case of advanced CRC, which is the most current situation upon diagnosis, most of these therapies will fail to achieve disease-free survival. This can be mainly attributed to the numerous escape mechanisms that the tumor accumulated along its development and selection of clones able to thrive in a very hostile microenvironment (immune surveillance, nutrient and oxygen deprivation, etc.). To avoid these pitfalls, preventing CRC to get to an advanced stage appears as a solution, for which several approaches can be considered. Prophylactic treatments, targeting carcinogenesis as a whole, especially in the form of vaccines [113]. For the colon, a vaccine based on MUC1, a glycoprotein abnormally expressed in most CRC and a majority of human

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

cancers, has been tested and shown to be capable to induce an immune response. Vaccination of patients presenting precancerous lesions, also shown to abnormally express MUC1, induced high levels of anti-MUC1 antibodies and long lasting immune memory [114] in 44% of patients. Although these results are promising and highlight the potential of vaccines in CRC, the majority of patients failed to respond, which was associated with the presence of immunosuppressive mechanisms. These results led to the first clinical trial of a prophylactic cancer vaccine clinical trial based on a non-viral antigen, targeting healthy individuals at high-risk of colon cancer based on their polyp history [115]. Another approach to prevent CRC from reaching an advanced stage is to detect and surgically remove colorectal lesions at an earlier stage by developing efficient screening strategies and improving the understanding of premalignant colorectal lesions.

In this chapter, we addressed several times the use of immune parameters for classification or therapy of cancer patients (Immunoscore, prophylactic vaccine, immunotherapies), highlighting the crucial role of the immune system in colorectal carcinogenesis. This subject will be the focus of the final chapter.

1.3 Tumor immunity in colorectal cancer

1.3.1 History of Tumor Immunology

1.3.1.1 Cancer immunosurveillance and immunoediting

Three primary roles can be attributed to the immune system in the prevention of tumors. First, by eliminating viral infections, the immune system limits the occurrence of virus-induced tumors, such as tumors caused by Epstein-Barr virus, human papilloma virus, or hepatitis B virus. Second, it prevents the onset of an inflammatory environment, which is favorable for the occurrence of cancer, through the elimination of pathogens and the resolution of inflammation. Finally, the immune system is capable of specifically recognizing and eliminating tumor cells expressing tumor-specific antigens [116]. As the first two points occur before the onset of carcinogenesis, only the last point will be addressed in this chapter.

As discussed above, the interactions of cancer cells with the tumor microenvironment(TME) is a critical parameter that influences carcinogenesis. It has been accepted, or are least suspected, since the early 1900s that the immune tumor microenvironment (iTME) can shape carcinogenesis. Paul Ehrlich, one of the fathers of adaptive immunity, hypothesized that host defense could prevent tumor development [117]. Later, around the 1920s, the presence of lymphocytes within the gastric carcinoma TME was associated with better patient survival [118]. The theory of immunosurveillance, which states that the immune system is capable of recognizing and destroying neoplastic cells, was introduced around 1960 [2] following a better understanding of the immune system and the proof of the existence of tumor antigens that would allow the immune response to discriminate between health and tumor cells. However, the field of tumor immunology remained secondary until

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

the demonstration of immunosurveillance around 2000 [119, 120], with the emergence of mice models of immunodeficiency [121].

The role of the immune system during cancer development is now considered as a dynamic process called immunoediting [3, 121], which can be divided into three phases: the first phase, the elimination phase, during which immunosurveillance prevents cancer development. Then the equilibrium phase, where the destruction of tumor cells is incomplete but cancer development is still controlled by the immune system. During this phase, tumor cells continue to proliferate and accumulate mutations, while immune pressure favors clones with attenuated tumor antigenicity. The results of this immune selective pressure have recently been demonstrated by screening mouse tumor models developed with and without adaptive immune selective pressure, using CRISPR technology [122]. This study highlighted that tumors of immunocompromised mice harbored fewer loss-of-function mutations in tumor suppressor genes in multiple cancer models. These clones will be able to escape the immune response, leading to tumor growth and the tumor escape phase, considered a failure of the immune system to control tumor growth. This phase is often associated with the onset of an immunosuppressive microenvironment and / or the inability of immune cells to recognize cancerous cells.

1.3.1.2 Immune contexture

However, while it was initially believed that during the escape phase the immune system had lost and therefore was useless for the control of carcinogenesis, it has been shown many times that the immune system can still provide useful information on the prognosis of patients and the outcome of the disease. The concept of 'immune contexture' encompasses this notion and describes the potential of the quantity and quality of the iTME to predict and impact cancer patient survival [123]. This concept followed the demonstration that the density and location of memory and cytotoxic T cells within the tumor were superior predictors of outcome than the classification AJCC / UIC-TNM in CRC [6], which was a milestone in the field of cancer immunology.

The immune contexture is defined by the nature, location, density, and functional orientation of the iTME. Although first defined in CRC, its concept has been applied to numerous solid cancers, where the presence of various types of immune cells has been associated with prognosis and classification power, as seen in Figure 1.10. The development of such prognostic factors can lead to more personalized and efficient patient care and also highlights the potential of therapies aiming to restore the immune capacity to control tumor growth, called immunotherapies.

1.3.1.3 Cancer immunotherapy

The history of immunotherapies also started very early, in 1891, with the bone sarcoma surgeon William Bradley Coley stimulating the immune system through the injection of a mix of bacteria

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

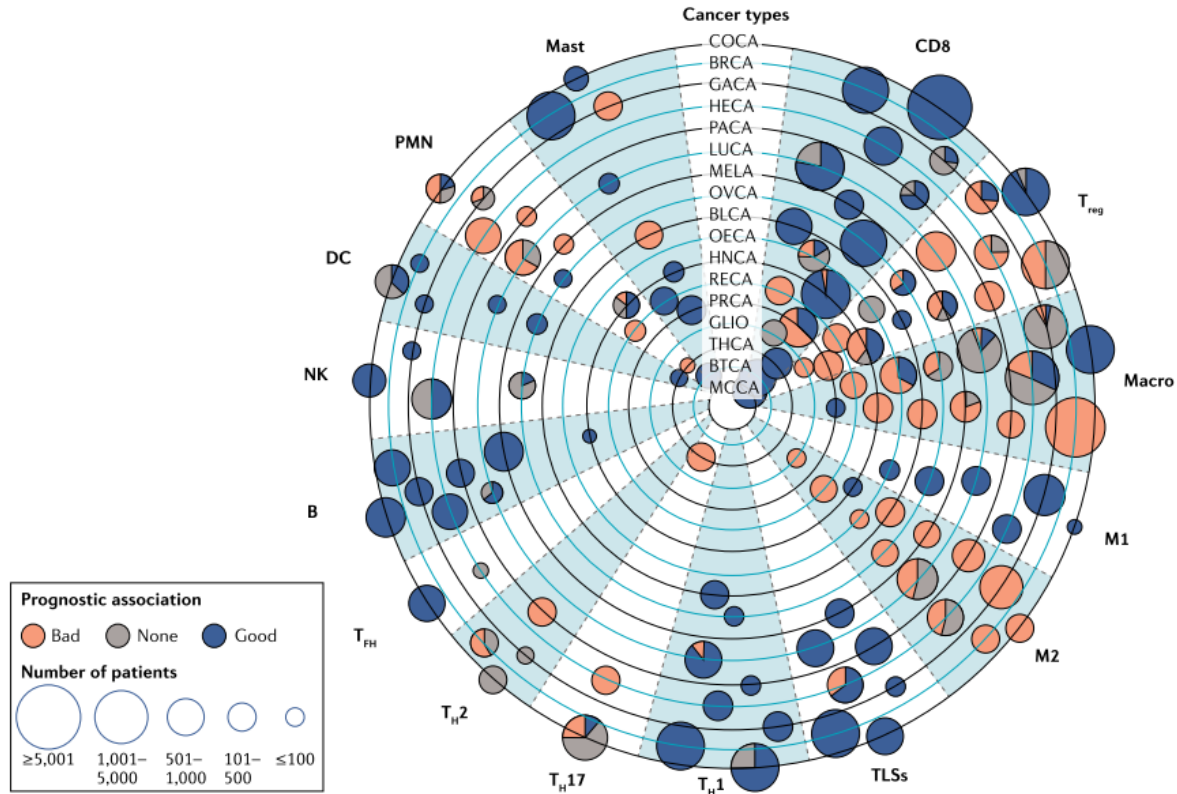


Figure 1.10: effects of the immune infiltrate on the prognosis of patients with cancer. From [5]

and bacterial products [124]. While it yielded promising results, this approach was quickly abandoned in favor of radio and chemotherapies. Interestingly, the efficiency of these treatments is now partially attributed to their ability to activate the immune response [125]. The lack of interest in immunotherapies until the 2000s can also be attributed to the fact that, according to the immunoediting theory, emerging tumors must be immunoedited or immunoselected to escape immune control. Invasive tumor would then have to be highly resistant to immune elimination [126], making immunotherapy unsuitable for patient care.

While not receiving their current recognition, immunotherapies emerged towards the end of the XX^e century with the administration of soluble immune mediators such as IL-2 and TGF- β to cancer patients, as well as adoptive transfer of expanded tumor infiltrating lymphocytes (TILs) to patients with metastatic disease. Immunotherapies can even be traced back to 1960 with the first hematopoietic stem cell transplantation in the form of BM infusion [127].

As the biology of immune checkpoints and their ability to modulate T cell activation was better understood [128, 129] their potential as an immunotherapy target appeared. The demonstration of the long-lasting immune protection conferred by anti-CTLA-4 antibodies in mice [130] was the first step towards numerous successful immune checkpoint blockade immunotherapies. However, while this kind of immunotherapy produced very promising results, only a small proportion of

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

patients is responding, calling for the combination with other therapies and the need for markers that help predict the patient's response to treatment. Classification including immune parameters, such as Immunoscore or other parameters of immune contexture, seems especially promising as immunotherapy predictive factors.

Combined, demonstrations of immunosurveillance theory, the importance of immune contexture, and the possibility of reestablishing immune control using immune checkpoint antibodies led to the explosion of the field of cancer immunotherapy. In particular, antibodies against CTLA-4 and PD1 have been approved for multiple types of cancer and led to the attribution of the 2018 Nobel Prize in Physiology or Medicine to the immunologists James Allison and Tasuku Honjo.

1.3.1.4 Precancerous lesions and tumor immunology

In the light of the immunoediting theory, it appears that most invasive cancers have achieved immune escape to some extent, a necessary condition for tumor growth. The results in the immunodeficient mice model, which were more susceptible to carcinogen-induced and spontaneous primary tumor formation compared to their wild-type counterpart [119, 121], support the idea that at least immune surveillance and equilibrium are already taking place in the precancerous stages. Observations that the adaptive immune response within tumors is strongest in the early stage of carcinoma also highlight the importance of the immune mechanisms underlying early carcinogenesis [131, 97].

Furthermore, the transition from immune surveillance to escape is most likely reflected by changes in immune contexture, as most escape mechanisms involve modulation of the immune response. To better understand the mechanisms underlying the early steps of carcinogenesis, including immunoediting theory, the characterization of the immune contexture in precancerous lesions appears to be critical. However, compared to invasive carcinoma, where tumor immune contexture is thoroughly studied in a wide range of cancer types, little is known about the state of the immune system in precancerous lesions. This knowledge gap can be explained mainly by the fact that the main focus is on invasive tumor tissue, responsible for most patient deaths, and that cancer is often diagnosed at an advanced stage, leaving precancerous lesions rarely detected and even more rarely surgically removed.

In recent years, however, more and more studies have pursued the goal of filling this gap, leading to characterization of the immune response in several precancerous types. A study published in *Nature* [7] by our team in 2019 focused on 9 stages of precancerous lung lesions and demonstrated immune activation even at early stages, highlighting the presence of early immunosurveillance, and that advanced lung precancerous lesions were able to escape this immune response, leading to cancer development. Another study, in cervical precancerous lesions, also demonstrated immune surveillance even in low-grade precancerous stages, while high-grade lesions harbored hallmarks of immune escape such as immunosuppressive molecules and upregulation and immune checkpoints [132]. Study of preneoplastic / early liver lesions highlighted once more immune activation even at very early

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

stages of carcinogenesis [133]. Interestingly, this study also highlighted the presence of TLS, for the most part immature, in one quarter of lesions. These lesions were associated with higher activation of the adaptive system, as well as higher levels of immunosuppression and immune exhaustion.

Overall, the data to date show that the immunoediting process is active during precancerous stages, and with immune surveillance at very early stage and immune escape mechanisms developing later, favoring the transition toward invasive carcinoma. In colorectal tissue, several recent studies also characterized the immune contexture in precancerous lesions. Although most used either transcriptomic or immunohistochemistry, making them rather unidimensional, a few recent studies chose a more integrative approach, merging the advantages of different techniques and exploiting the power of single-cell sequencing. These new studies push the characterization of the TME of colorectal polyp to a new level, helping to better understand its mechanisms. However, while the first of these integrative studies was published in *Cell* [74], the other one is still in the preprint stage [134].

To draw a clear picture of the immune response during early carcinogenesis, I will first expose the main concepts underlying the general cancer immunity cycle, and then present current knowledge concerning the immune contexture in CRC and in precancerous lesions.

1.3.2 The cancer-immunity cycle: from release of tumor antigens to cytotoxicity

1.3.2.1 Tumor associated antigens

The necessary condition for the immune system to be able to recognize and eliminate tumor cells is the detection of tumor cells as non-self or abnormal cells. This process is based on the presence of tumor-associated antigens (TAA) in the TME. TAAs can arise from two different origins. The first is the presence of somatic mutations in the cancer cell genome, leading to the transcription and translation of proteins that carry aberrant sequences, leading to the generation of neoantigens [135]. The second type of TAAs comes from aberrant expression of germline sequences. This type includes oncogenic viral antigens, peptides arising from transcription and traduction of a normally silent part of the genome, such as testis antigens (usually restricted to the testes) and noncoding portions of the genome, and overexpression of normally presented self-antigens, able to elicit an immune response above a given threshold [136].

As the two classes of TAAs are neither encountered under physiological conditions nor expressed by thymic epithelial cells, they are both capable of eliciting an immune response if detected by the immune system, and it has been shown that these TAAs can be recognized by CD8 + T cells in cancer patients when associated with class I MHC [137].

High TAAs levels are often associated with high immunogenicity and tumor infiltrating T lymphocytes (TITLs) levels [138], and can be associated with better survival and response to therapy. High somatic tumor mutational burden (TMB) is associated with a longer survival in patients receiving immune checkpoint inhibitors in several types of cancer [139].

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

For these TAAs to trigger an efficient immune response leading to effective killing of cancer cells, a series of events, called the Cancer-Immunity Cycle (Figure 1.11) must be performed [140].

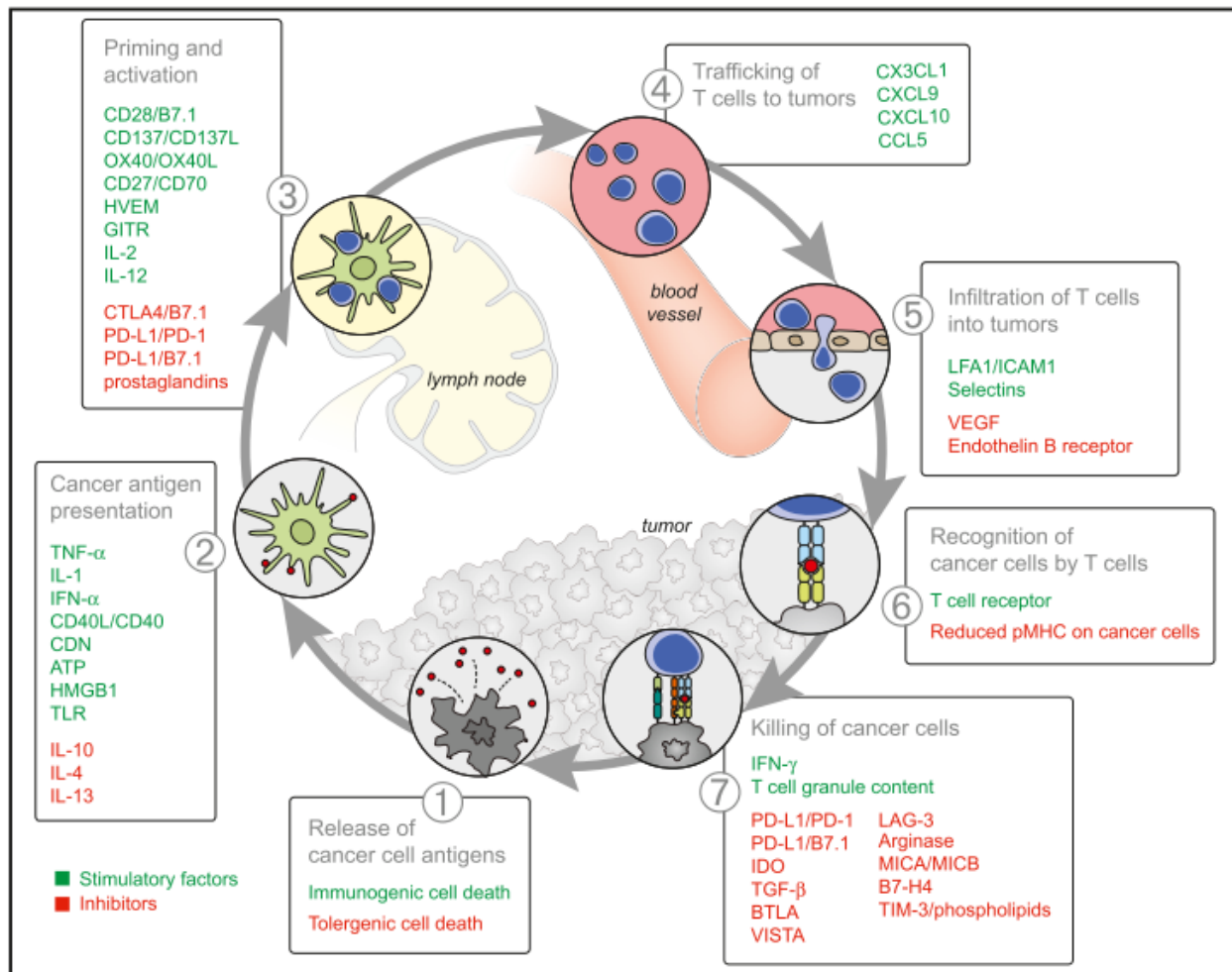


Figure 1.11: The Cancer-Immunity Cycle and its regulators. From [140]

1.3.2.2 The cancer-immunity cycle

The first step is the capture and processing of TAAs by APCs (mainly DCs) present in the TME. Dying tumor cells are the main source of these TAAs, as in addition to releasing them into the TME, they provide pro-inflammatory cytokines and factors that will participate in the priming of the immune response.

As presented in Section 1.1, APC will then migrate to SLO or TLS and trigger tumor-specific T cell priming. Activated T cells will, in turn, travel to the tumor site. On tumor cells, the presentation of TAAs by class I MHC will trigger the cytotoxic function of CTLs, leading to cancer cell death and allowing the cancer-immunity cycle to expand.

Numerous stimulatory and inhibitory factors of the immune response, presented in Figure 1.11,

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

are responsible for modulation of this cycle, and their balance can lead to the 3 phases of the immunoeediting theory: elimination, equilibrium, and escape. Although stimulatory signals are associated with immunogenic cell death (such as necrosis) and the classical immune activation pathway, the TME can develop escape mechanisms to escape destruction by the immune system.

1.3.2.3 Tumor immune escape

Mechanisms associated with immune escape and inhibition of the cancer-immunity cycle are multiple. Favoring tolerogenic cell death will limit the priming of APC and the presentation of TAAs in lymphoid organs. Establishing an immunosuppressive TME, often associated with the presence of T_{reg} and IL-10, will lead to the same result. Overexpression of inhibitory immune checkpoint by T cells, in combination with the presence of their ligand within the lymphoid organ or the TME level, will inhibit the priming of T cells or effector function (i.e., cytotoxicity), respectively. Tumor cells can also evade cytotoxicity by downregulating the level of peptide / CMH complex at their surface, evading recognition by CTLs [141].

Another mechanisms leading to immune escape is the exclusion of immune component, especially T cells, from the TME, leading to the phenotype "cold", as opposed to "hot" tumors, who display high level of T cell infiltration. As the presence of TITLs is often associated with better survival and improved immunotherapy responses, cold tumors are often of bad prognostic [4]

1.3.3 Colorectal immune tumor microenvironment and immune contexture

As introduced previously, various components shape the TME, among which are fibroblasts, blood vessels, and the immune system. As presented in Figure 1.12, the iTME can be characterized by numerous factor, including immune cell types, their location (at the CT or its periphery), their activation / exhaustion status and the presence of organized immune structures, TLSs. Macrophages often account for a high proportion of immune cells in the iTME, but lymphocytes are also often present at a significant density [96, 131].

Correlating these various parameters with patient survival will lead to the definition of a positive or negative immune contexture. As we will see in this chapter, capturing the functional orientation of these immune cells is particularly critical as most immune cells can perform anti or protumoral actions, depending on their activation status.

1.3.3.1 Macrophages

Tumor-associated macrophages (TAMs) can present very different functional orientations. Although reality is more of a continuum between these two states, the two extreme archetypes of this spectrum are called M1 and M2 [142]. The M1 subset is associated with proinflammatories properties, such as secreting IL-12 and promotion of antitumor T_H1 response. On the other hand, M2 TAMs are anti-inflammatory, secreting IL-10 and TGF- β and leading to the establishment of a protumoral TME. As seen in Figure 1.10, the presence of M1-TAMs is associated with good prognosis,

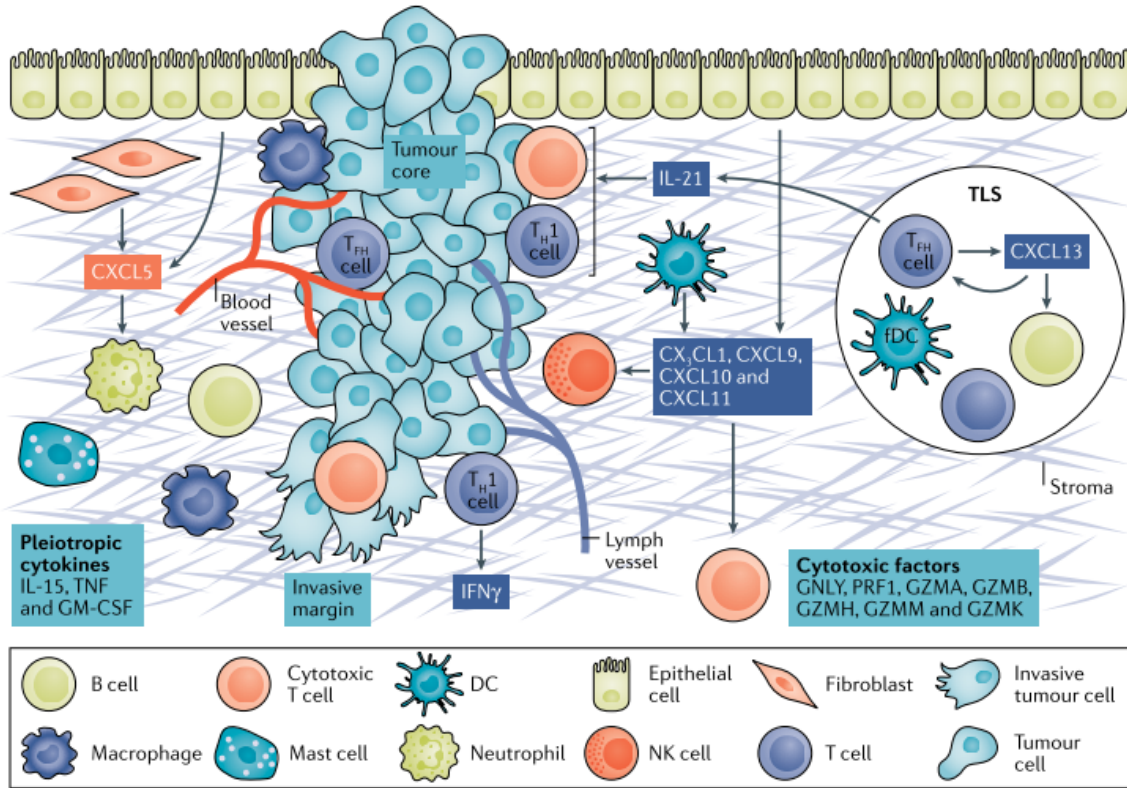


Figure 1.12: The immune tumor microenvironment. From [5]

while the opposite is true for M2-TAMs. We can also notice that the presence of TAMs without functional discrimination is associated with bad prognosis in most cancer.

However, in CRC, the high density of TAMs has been shown to be associated with a favorable prognosis [143]. This highlights the importance of phenotypically characterized TAM infiltration, as a high proportion of M2 TAMs (characterized by the CD163 / CD68 ratio) in CRC have been associated with poor prognosis. The presence of M2 TAMs is also associated with aggressive CRC phenotypes, such as epithelial mesenchymal transition, advanced TNM stage, and tumor invasion. The M2/M1 ratio is higher in IM versus CT [144].

The origin of macrophages, either monocyte-derived or tissue resident, also appears to be determinant in their function within the TME [145], but remain understudied.

1.3.3.2 Neutrophils

Tumor associated neutrophils, or TANs, are recruited from the circulation under inflammatory conditions and have been shown to have anti- and pro-tumoral activity, but are generally negative for the patient. For instance, neutrophil-derived NETs have been shown to cover tumor cells, preventing cytotoxicity from NK cells and CTLs [146], but TANs have also been demonstrated to selectively kill cancer cells and attenuate tumorigenesis by secreting neutrophil elastase [147]. The presence of TANs is generally associated with a poor prognosis [148], but is variable depending on the cancer

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

type and neutrophil markers. While a pan cancer study using bulk deconvolution analyses of the tumour transcriptome showed that the presence of TANs was associated with bad prognosis in CRC [149], the opposite has also been demonstrated using immunohistochemistry [150]

1.3.3.3 Myeloid-derived suppressor cells

Myeloid-derived suppressor cells (MDSCs) are a very heterogeneous subset of myeloid cells recruited to the TME. They can be derived from circulating monocytes or immature neutrophils and are found only under pathogenic conditions[151]. They are characterized by protumoral activity, supporting cancer progression and metastatic spread through suppression of the antitumor immune response [152]. MDSCs are capable of increasing T_{reg} proliferation by secreting indoleamine 2,3-dioxygenase 1 (IDO1), and have been shown to increase $CD8^+$ T cells apoptosis *in vitro* [153].

However, their plasticity and heterogeneity make it difficult to clearly identify their presence in the TME without a wide combination of markers [154], and their prognostic power is more often assessed using circulating MDSC, which has been associated with shorter survival in multiple types of tumor [5].

1.3.3.4 Mast cells

Although often overlooked in tumor immunology as canonically associated with allergic disorders, mast cells have been associated with protumoral and antitumoral functions in several types of cancer [155]. While being associated with bad prognosis in bladder and gastric cancer, their presence in the TME is a good prognosis in other types of solid cancer. Characterization of their localization and (e.g., peritumoral vs. intratumoral) and functional orientation could help to better understand their protumoral and antitumoral functions within the TME.

In CRC, mast cells have been shown to be present at lower densities compared to normal colon tissue and were associated with longer patient survival [156].

1.3.3.5 NK cells

NK are critical effector cells in antitumor immunity, able to recognize and eliminate neoplastic cells. While being major actor of innate immunity, they have also been shown to favor the recruitment of a DC subset with high T cell activation potential, leading to priming of the adaptive immune response [157]. Although rarely characterized in iTME, its presence has been associated with good prognosis in the majority of cases, including CRC [158].

1.3.3.6 Dendritic Cells

Due to their professional APC functions, DCs are central actors in the activation and differentiation of T cells. Overall, DCs have been associated with a good prognosis, including in CRC [159]. However, their prognostic impact is dependent on their subset (pDCs, cDC1 and 2 or monocyte-derived DCs) and functional characterization [25]. For example, dysfunctional pDCs have been

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

shown to favor T_{reg} expansion and have been associated with poor prognosis in breast cancer [160], highlighting plasticity of these populations. The cDC1 subset is particularly effective in the capture of dead tumor cell fragments and in the transport of tumor antigens to tumor-draining lymph nodes where they form the main DC subtype responsible for the cross-priming of $CD8^+$ T cells against tumors. However, this subset is rarely found in the TME, which could constitute an obstacle to antitumor immunity. [161].

1.3.3.7 T cells

T cells in general have been associated with favorable prognosis as shown by the inclusion of the CD3 marker in the Immunoscore [6, 102]. However, the numerous subsets and functional states present in the T cell lineage require further phenotypical and functional characterization. Distinguishing $CD8^+$ from $CD4^+$ T cells, and assessing their activation / exhaustion status is critical when trying to decipher the T cells mechanisms of the iTME.

1.3.3.7.1 $CD4^+$ T cells

Including numerous subset (T_h1 , T_h2 , T_h17 , T_{fh} and T_{reg}) and associated functions, $CD4^+$ T cells of the TME need to be further defined to be accurately associated with a function and a prognosis. While T_h1 cells and their associated cytokines (e.g., $IFN-\gamma$) have been shown to be associated with a good prognosis in most cancers, the prognostic value of other subsets of $CD4^+$ T cells is subject to variation depending on cancer type, stage, and TME [162].

In CRC, a study found that while the presence of T_h1 is associated with improved survival, the opposite is true for T_h17 and no prediction of the prognosis was associated with the presence of T_h2 or T_{reg} [163]. However, another article did not associated T_h17 with prognostic potential. However, when estimating the spatial information of epithelial and stromal compartment, intraepithelial T_h17 were found to correlate with improved survival [164]. Discrepancies between results can be explained by the size and characteristics of the cohort (e.g., tumor stage and molecular subtype) of the T_h17 subset, but also demonstrates the importance of the spatial component in the definition of immune contexture. Similarly, T_{reg} have also been associated with good prognosis in CRC [165]. T_{fh} , often associated with TLS presence, B cell maturation and memory functions, has been associated with improved survival in CRC patients [131].

1.3.3.7.2 Cytotoxic $CD8^+$ T cells

As already presented in 1.3.2, CTLs are major actors of the cancer-immunity cycle, and thus central for an efficient antitumor immune response. The association of CTLs presence in the TME with improved patient survival was central in the establishment of the concept of immune contexture and in the development of the cancer immune field [6]. Since, CTLs have been assessed as a prognostic and predictive factor in numerous cancer types, and a meta-analysis of the literature demonstrated that CTLs were the immune cells with the strongest positive impact on patients' survival [5].

Prostate cancer is one of the few exceptions where a high density of CTLs is associated with

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

shorter survival. Interestingly, it has been shown that in prostate cancer, high CTLs infiltration was associated with expression of the checkpoint protein PDL1 by tumor cells, which may partially explain this result [166]. This observation highlights the need for a quantification of immune activation and exhaustion status in the TME.

1.3.3.7.3 T cell exhaustion

As chronic conditions, cancers constitute a favorable environment for the development of T cell exhaustion. Inhibitory immune checkpoints such as CTLA-4, PD1, TIM3 or LAG3 are indeed frequently associated with the iTME, highlighting the chronic stimulation of TCR in the TME. Mouse model found that T cell dysfunction was induced early during carcinogenesis, and progressed with the disease [167, 168]. In kidney cancer, progenitor of exhausted T cells TCF1⁺ CD8⁺ TITLs, have been shown to form intratumoral niche, rich in APC, able to differentiate into terminally differentiated exhausted T cells (who loose TCF1 expression and increase their levels of checkpoint molecules) [169]. This study demonstrated that patient whose disease is progressing are lacking these niches.

Although T cell exhaustion is natural in such a chronic disease, most cancer are also diverting these mechanisms, favoring immune escape. In particular, overexpression of the immune checkpoint ligand, such as PDL1 and CD80, is often found in tumors, inhibiting effector function and memory cell formation. The expression of PDL1 has been observed in various types of cancer and is most of the time associated with poor prognosis [170]. In some cancers, such as CRC and melanoma, the prognostic value of PDL1 is still not clear. The expression of PDL1 is associated with improved clinical results in breast cancer and Merkel cell cancer. It has been shown that marker of T cell exhaustion are present even during early carcinogenesis, but that these exhausted phenotypes can be rescued, which is not the case in invasive cancer [171]

In CRC, PD1 expression by TITLs have been shown to be associated with PDL1 and MSI tumors had higher levels of these protein when compared to MSS [172]. Interestingly, it was found that PD1 was a good prognosis marker only in tumors with low-level PDL1 expression. In high-level PDL1 tumors, PD1 was associated significantly worse recurrence free survival.

1.3.3.8 B cells and plasma cells

While the mechanisms of TITLs has been demonstrated for numerous subsets, antitumor humoral adaptive immunity is less defined. While tumor infiltrated B lymphocytes (TIBLs) account for a significant part of the immune cells found in the TME, their function are still subject of debate [173]. For antitumoral effects, a murine model of lung adenocarcinoma demonstrated that TIBLs could promote differentiation of tumor specific T_{fh} through antigen presentation, leading to enhance CTL activity [174]. In ovarian cancer, B cells specific for surface autoantigens expressed on tumor cells were shown to undergo somatic hypermutation, leading to coating of tumor cells by IgGs and possibly supporting ADCC and activation of phagocytic cells through interaction with Fc receptors [175]. The protumoral effects of TIBLs were described in squamous cell carcinoma, where IgG

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

immune complexes would activate myeloid cells through their FC receptor [176].

Mirroring this duality, studies found that B cells and plasma cells are correlated with a good or bad prognosis, depending on the type of cancer and the study. Mostly, TIBLs were associated with good prognosis in numerous cancer types, while plasma cells have been less studied but also appear to be favorable for patient outcome [177].

In CRC, B and plasma cells have been shown to decrease with cancer stage, and high densities of B cells were associated with improved survival after adjusting for age, TNM stage, differentiation grade and vascular invasion [178]. Edin *et al.* confirmed these results by associating CD20+ B cells with the presence of CTLs within the TME, as well as lower cancer stage and improved survival [179]. In advanced and metastatic CRC, an increase in the regulatory phenotype of B cells has been observed, reflecting a possible immune escape [180]. Metastasis have also been associated with decreased CD20 density when compared to primary tumors [181].

1.3.3.9 Tertiary lymphoid structures

In cancer, TLSs are strongly associated with B cell infiltration levels, which is often low when no TLS is present [182]. While both early and mature TLS appear to be the site of antigens presentation to B cells, the selection of B cells with high-affinity BCRs only seems to happen in the GC of mature TLSs, where activated B cells are expressing Ki67, suggesting of their proliferation [183] (Figure 1.13. In renal cell carcinoma, mature TLSs have been shown to generate and propagate plasma cells producing antitumor antibodies [184]. TLSs are also a privileged site of priming of the T cell response through antigen presentation by professional APCs, and their presence correlates with the densities of T cells in cancer [185].

In addition to the fact that B cells, plasma cells and T_{fh} have all been associated with good prognosis, the CXCL13, central in TLS formation and organization has been shown to correlate with improved survival in CRC [131] and melanoma [186]. Altogether, these results suggest a positive impact of the presence of TLSs on antitumor immunity and patient survival, which has been proven multiple times across several cancer type [17, 187].

In CRC, a subset of patient called Crohn Like Reaction patients present a iTME rich in TLS, and a better survival when compared to CRC patient without TLSs [188]. This association of TLSs with improved survival has been confirmed at early stages of CRC [189]. CRC TLSs have been shown to correlate with low macrophages densities and high B cell levels [190]. Differences in colorectal TLS structure level (organized or not) might be due to differences in the intestinal microbiota, as some bacteria are known to induce the differentiation of CD4 T cells towards T_{fh} [191].

1.3.3.10 Immune parameters as predictive factors

In addition to being associated with better survival, the iTME parameters can also be used as predictive factors, describing chances of response to a specific anticancer treatment. In particular, TITLs levels and activation status has been shown to be efficient predictive markers. For example,

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

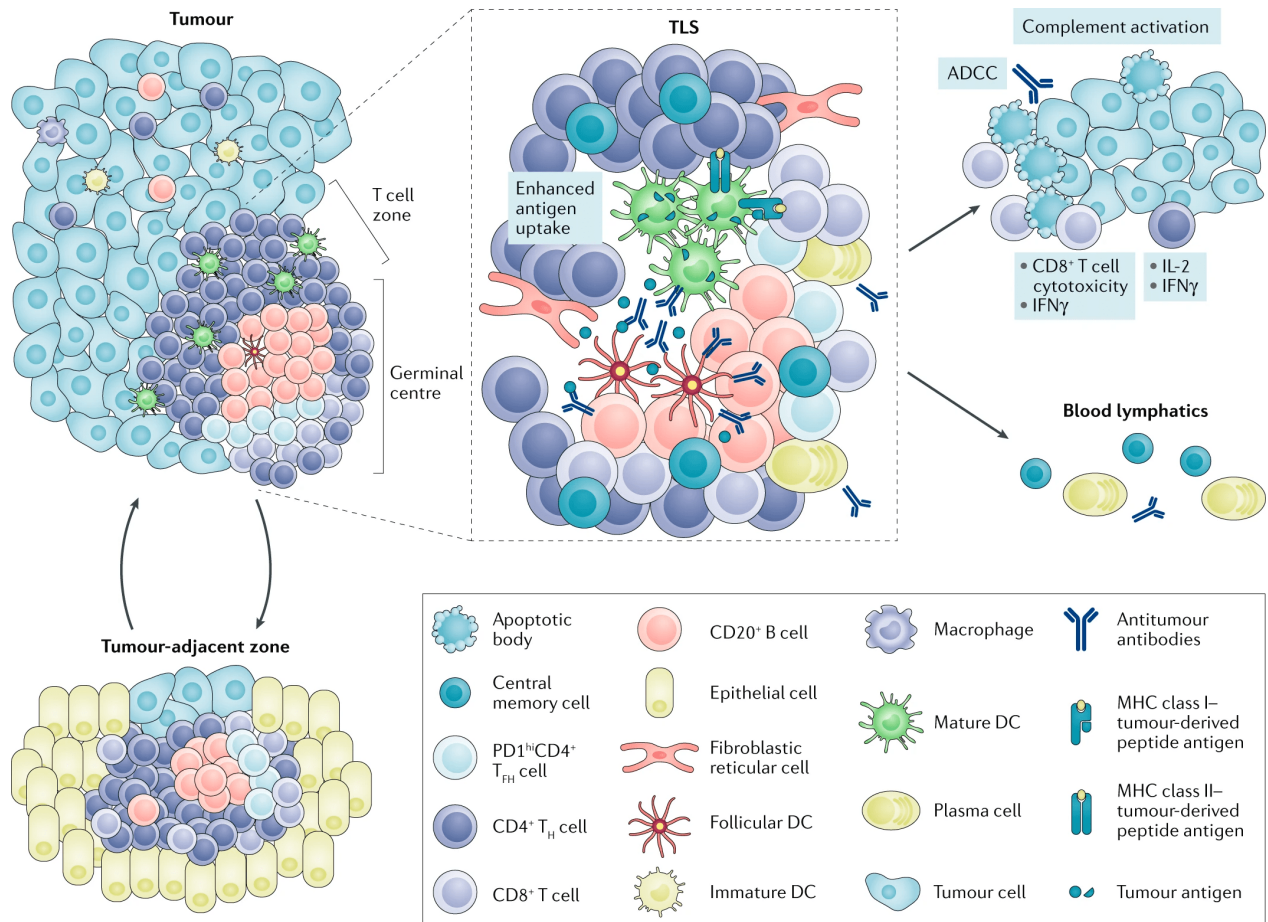


Figure 1.13: The composition and function of tertiary lymphoid structures in cancer. from [192]

responders to anti-PD1 therapy have been shown to harbor higher levels of CD8 and PDL1 when compared to non-responder [193]. PDL1 has also been shown to correlate with TITLs and response to neoadjuvant chemotherapy in breast cancer [194], highlighting the potential role of the iTME in response to non-immune based therapies. The presence of TCF1⁺ CD8⁺ TITLs was shown to correlate with improved survival responses to ICB in patients with melanoma [195].

TMB, another prognostic and predictive factor [102], has also been associated with TITL densities, specifically for CTLs [196].

1.3.3.11 iTME in precancerous polyps

While most studies focus on invasive CRC, along the year, several studies were aimed at characterizing iTME within premalignant colorectal lesions. Adenomatous lesions are the main target of these polyps, as they are more frequent and have been studied for a longer time.

As early as 1999, Rubio *et al.* found higher levels of intraepithelial TILs in HG APs when compared to LG APs using simple H&E staining [197]. Technology has since come a long way, and Chen *et al.*, using single cell RNAseq, compared the two colorectal premalignant pathways and

1.3. TUMOR IMMUNITY IN COLORECTAL CANCER

found a similar level of myeloid cells, B cells, plasma cells, mast cells, CD4⁺ T cells, and T_{reg}, while serrated samples presented higher intraepithelial levels of T cells, CD8⁺ T cells and NK cells [74].

In general, no consensus has been reached concerning the variation of the iTME between the serrated and adenomatous pathways and within each pathway along carcinogenesis. One of the obstacles that prevents homogeneous data is that studies classify APs based on different parameters, such as dysplasia grade [134], histological classification [198], or polyp size and shape [199].

In adenomatous lesions, mast cell densities, T_h17, NK cells appeared lower in tubulovillous adenomas when compared to villous ones, suggesting emergence of an immunosuppressive TME. Study of dysplasia only found mast cells to increase from LG to HG [198].

Along the adenocarcinoma sequence (from LG AP to HG AP to invasive carcinoma), Freitas *et al.* found that densities of T cells, CD8⁺ T cells, CD4⁺ T cells, T_{reg} and NK cells are decreasing as well as the expression of PDL1, while the burden of tumor mutations and MHC I are increasing, suggesting the progressive establishment of an immune cold microenvironment. [200]. Other studies found that immune cell densities decrease along progression toward carcinoma, for instance demonstrating that T_h1 levels increased from healthy colon to APs, and then decreased to CRC [201]. Several studies also found that this progression was accompanied by an increase of T_{reg}, promoting an immunosuppressive TME [202, 134].

Concerning the exhaustion status of TITLs in the iTME, Becket *et al.* showed in their preprint that exhaustion levels, characterized by scores for genes of exhausted T cell markers, including *BATF*, *CTLA4*, *PDCD1* (coding for PD1) and *TOX*, did not increase between adenomatous precancerous stage, but were higher in invasive carcinoma [134].

In the serrated pathway, increased TITLs levels and PD1/PDL1 expression were shown to correlate with sequential progression of SSLs, from no dysplasia to high-grade dysplasia and cancer [203, 204].

2 Objectives

2.1 Context and underlying hypotheses

As presented in section 1.2.3, among average-risk individuals, some patients will still develop an abnormally high number of polyps, which has been described as a risk factor for the recurrence of polyps or CRC [77, 78]. However, causes responsible for this high rate of polyp development are unknown in the majority of cases.

Among the factors that shape the carcinogenesis process, iTME was shown to be associated with the clinical outcome of cancer patients, in terms of survival and therapeutic responses. In particular, characterization of the immune contexture in most cancers has unveiled numerous mechanisms associated with anti or protumoral properties.

Characterization of the iTME in colorectal premalignant lesions of average risk patients developing an abnormally high number of polyps could lead to a better understanding of the process responsible for this high rate.

2.2 Objective and methodological tools

This work aim to establish a relationship between iTME parameters and rate of polyp development in average-risk individuals. To do so, we applied an integrative approach to a retrospective cohort. Access to the patient's clinical history gave us the opportunity to select a group of patient presenting a normal profile and a group of patients with a rich history of polyps despite no apparent risk factor. Several tools were used to characterize the iTME of these polyps.

As previously presented, the iTME is defined by several parameters such as immune cell type, density, localization, and quality of response. Several techniques are capable of deciphering its component, each presenting specific pros and cons. Multiplex IHC was our method of choice as, contrary to other classical methods, IHC provide spatial information. This technique was complemented by RNA sequencing (RNAseq), which provides the advantage of being an untargeted approach, which means that no target choices are needed and all aspects of the transcriptome can be studied *a posteriori*. Using the gene expression-based TME deconvolution tool [205], RNAseq can also estimate the relative abundance of immune cell types using the gene signature specific to cell types of the TME. Finally, we performed whole exome sequencing (WES) to obtain access to the mutational profile of our lesions, which is critical to assessing the carcinogenesis process (Figure 2.1).

To assess the possible association between the iTME and polyp frequency, we first needed to properly characterize our samples, both from a molecular and iTME perspective. Indeed, as colorectal premalignant lesions can come from two very different pathways and present numerous character-

2.2. OBJECTIVE AND METHODOLOGICAL TOOLS

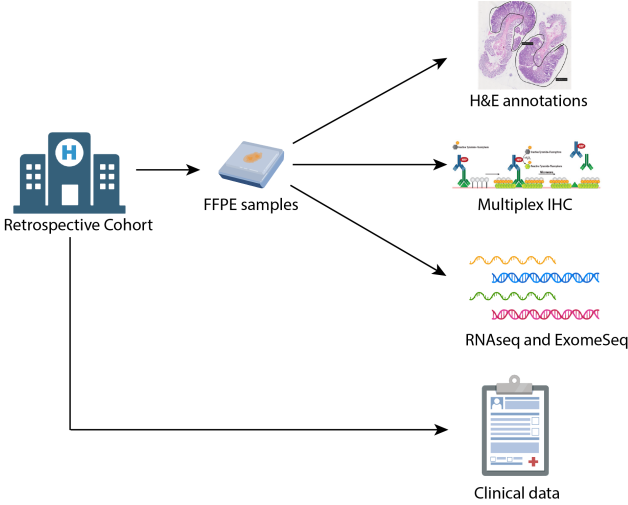


Figure 2.1: Integrative analysis of colorectal iTME

istics, it appeared critical to first analyze the iTME parameters specific to these variables, which would allow us to subsequently identify the variation of iTME specific to the frequency of the polyp.

3 Research Methodology

3.1 Experimental Model and annotations

3.1.1 Cohort constitution

The cohort was selected from adults who underwent surveillance colonoscopy or surgery for polyp resection at the Cliniques Universitaires Saint-Luc in Bruxelles. As presented in Table 3.1, 26 patients were selected. Each patient has been surgically treated for colorectal precancerous lesions and/or cancer one to seven times. Each time, between one and 12 lesions have been removed for a total of 140 lesions. For each lesion, eight fixed-formalin paraffin embedded tissue slides were collected (4x4 μ m for IHC and 4x10 μ m for DNA and RNA extraction) For annotation purpose, a scan of hematoxylin & eosin (H&E) staining was also collected. When available, two slide of healthy tissue were also selected for germline analysis (WES).

Table 3.1: Patients characteristics

	Patients ($n = 26$)
Age, mean (SD)	64.6 (9.95)
Sex, $n(\%)$	
Female	9 (34.6)
Male	17 (65.4)
Body mass index, mean (SD)	26.2 (5.30)
Follow-up (years), mean (SD)	4.98 (2.23)
Lesion number, mean (SD)	5.81 (4.73)

3.1.2 Cohort annotation

Colorectal lesions type, grade and area were annotated on an H&E staining with the help of a pathologist. In the case of adjacent lesions of various grades (e.g., some high-grade dysplasia next to low-grade dysplasia), the whole lesion was counted as one and graded as the most advanced lesion for DNA and RNA extraction. For IHC analysis, adjacent lesions were annotated and analyzed separately.

3.2 RNA sequencing and whole exome sequencing

3.2.1 RNA and DNA extraction

RNA and DNA were isolated from FFPE tissue slides using NucleoSpin totalRNA FFPE XS or NucleoSpin DNA FFPE XS kits (MACHEREY-NAGEL), respectively, and according to the manufacturer's instructions. RNA concentration was estimated using a spectrophotometer NanoDrop 2000 (ThermoFisher Scientific). DNA concentration was assessed using the Qubit 4 Fluorometer (ThermoFisher Scientific).

3.2.2 RNA sequencing and alignment

RNA-sequencing was performed using QuantSeq 3' RNA-755 Seq Library Prep Kit FWD for Illumina (75 single-end) with a read depth of 8M (Sidra Medicine, Doha, Qatar). Single samples were sequenced across four lanes, and the resulting FASTQ files were merged by sample. Quality trimming is performed to remove adapter sequences and polyA tails. Then trimmed reads were aligned to human genome GRCh37/hg37 (Genome Reference Consortium Human Build 37) using STAR 2.6.1d. FeatureCounts v2.0.0 was used to generate the raw counts. Raw expression data were normalized to size factor effects using R package DESeq2 [206], and patient presenting particularly low total normalized count were removed as quality control. Subsequent analysis were performed using R.

3.2.3 Differential expression analysis

Differential expression analysis between frequency groups was performed using R package DESeq2.

3.2.4 CMS classification

CMS classification was carried out on normalized counts using the R package CMScaller [207], using the 'Nearest Template Prediction' algorithm to predict which CMS class described the sample best. Sample that could not be associated with a CMS class were annotated as "Indeterminate".

3.2.5 Consensus^{TME} scoring

Consensus^{TME} was carried out on normalized counts using the R package ConsensusTME [205]. Each score is determined based on the consensus gene sets presented in the Appendix. Gene sets are then used within a single-sample Gene Set Enrichment Analysis framework to provide normalized enrichment scores for each of the cell types representing the relative abundance of cell types across multiple samples.

3.3. MULTIPLEX IMMUNOHISTOCHEMISTRY

3.2.6 Whole exome sequencing and alignment

Standard Whole exome sequencing (WES) was performed (Sidra Medicine, Doha, Qatar). WES reads were aligned to the human genome GRCh38/hg38 (Genome Reference Consortium Human Build 38) using BWA aligner. Duplicated reads were removed using GATK preprocessing. SNV were called using Mutect2 [208]. Several filtering steps were then applied to generate MAF files containing single nucleotide variant per sample. Somalier analysis was performed to rate the relatedness of samples and check the proper germline sample - somatic sample association.

3.2.7 Oncoplot generation

Oncoplot and total mutation burden were extracted from the MAF file using the R package maftools [209].

3.3 Multiplex immunohistochemistry

3.3.1 Slide preprocessing

For deparaffinization, formalin-fixed and paraffin-embedded (FFPE) slides of tonsil and colorectal cancer (CRC) were incubated at 56°C for four hours, then soaked in Clearene, rehydrated by immersion in a graded ethanol series (100%, 90%, 70%, 50% and distilled water) and finally fixed in 10% Neutral buffered formalin.

3.3.2 Multispectral IHC

3.3.2.1 Staining

For the lymphoid panel, multispectral IHC staining were performed using the Opal-TSA technology (Opal Polaris 7 Color IHC Detection Kits, Akoya Biosciences). Staining was performed using the autostainer Bond RX (Leica Biosystems) following protocols recommended by Akoya Biosciences. Slides were stained using antibodies and conditions presented in Table 3.2, manually counterstained with spectral DAPI (Akoya Biosciences) and mounted using the mounting medium ProLong™ Diamond Antifade Mountant (Thermofisher Scientific).

Table 3.2: Antibodies used for the lymphoid panel

Marker	Specie	Clone	Company	Reference	Dilution	AR
CD4	Rabbit	EPR6855	Abcam	ab133616	1/500	pH9
CD8	Mouse		HalioDx	HD-FG-000019	1	pH9
CD3	Rabbit		HalioDx	HD-FG-000013	1	pH9
MUM1	Mouse	MUM1p	Dako	M725929	1/100	pH9
FoxP3	Rabbit	D2W8E	CST	98377S	1/100	pH6
CD20	Mouse	L26	Dako	M0755	1/500	pH9

Note: AR = Antigen retrieval

3.3. MULTIPLEX IMMUNOHISTOCHEMISTRY

3.3.2.2 Image acquisition

Stained slides were scanned with the 20X lens of the PhenoImager HT (Akoya Biosciences) under fluorescent conditions, providing whole tissue RGB images. Areas of interest were then selected using Phenochart (Akoya Biosciences) software and extracted. Specific signal from each fluorochrome was extracted from the original image, creating a 7-channel image which will be further analyzed using the DP software HALO (Indicalabs).

3.3.2.3 Image analysis

Using the ai-based DP software HALO (Indicalabs), a tissue classifier taking in account DAPI, CD3 and CD20 signals was trained to detect epithelial and stromal compartments, as well as TLSs presenting an area $> 25000 \mu\text{m}^2$.

Pathologist annotations were then reported to the tissue and a 1mm wide margin was drawn around it. Subsequently, this margin was divided into stromal and epithelial compartments using our tissue classifier. For each compartment, cells were detected on the basis of the DAPI signal using an AI-trained nuclei classifier. Finally, a positivity threshold was set for each marker, attributing a phenotype to each cell. Cell densities were exported and processed with the software R.

3.3.3 Brightplex IHC

3.3.3.1 Staining

For the T cell exhaustion panel, FFPE slides were stained using the Brightplex technology (Veracyte). Briefly, slides were stained with ImmPACT AMEC Red Substrate (Vector Laboratories) using the Bond RX following a Brightplex specific protocole. Slide were then mounted using an aqueous mounting medium (Vectamount, Vector). Images were acquired using the scanner Nanozoomer 2.1 (Hamamatsu) before unmounting the slides using a 56 ° C water bath. Staining is then removed using ethanol and the antibody is stripped using a denaturing solution and a heating step. The tissue slide can then undergo another round of this staining, using the next antibody. Staining conditions are subject to a confidentiality agreement with Veracyte.

3.3.3.2 Image analysis

Once images for all markers of interest have been acquired, their specific signal is extracted and merged into a multichannel image using the DP software HALO. Annotations from the lymphoid panel analysis were reported and the same pipeline was applied, leading to cell density extraction.

3.3.4 Linear Mixed Models

Linear mixed model analysis was carried out using the R package lme4 [210]. Random effects (individual, sex, and BMI levels) were chosen according to the Akaike Information Criterion.

3.3.5 Visualization and Statistical analysis

Unsupervised Hierarchical clustering and heatmaps were generated using the R package ComplexHeatmap. PCAs were performed and plotted using the FactoMineR R package. Unless stated otherwise, Mann-Whitney U test were performed using the software R.

4 Results

To identify immune features of the TME associated with higher rate of precancerous colorectal lesions development, the cohort selected to address this question will first be presented and characterized from a molecular point of view. Analysis of multiplex IHC assays will then be detailed, followed by the characterization of the iTME in the two main colorectal carcinogenesis pathways (serrated or adenomatous). The association of iTME with precancerous pathways stage will be discussed. Finally, the differences in iTME between patients presenting high and low polyp frequency will be highlighted, followed by the construction of a mixed linear model of the iTME in an attempt to validate our findings independently of individual-level variability.

4.1 Cohort presentation and molecular characterization

4.1.1 Cohort presentation

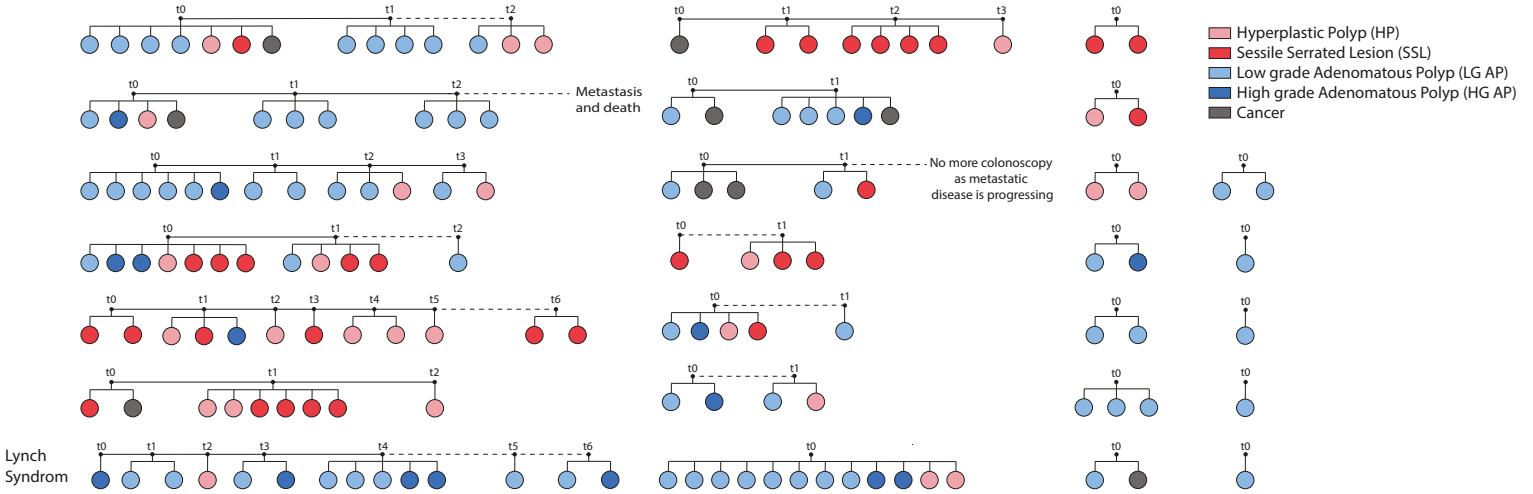


Figure 4.1: **Graphical overview of the cohort.** Each patient is represented by an horizontal line. Each time point, noted tx, can present several lesions of various type. Recurrence event that happened after the inclusion and are not included in the study samples are represented in dotted lines

To evaluate variations of the iTME between patients at low or high risk of developing precancerous colorectal lesions, a retrospective cohort was selected that included a similar proportion of the two classes. Polyps were histologically classified by a pathologist. As shown in Figure 4.1, our cohort is made up of 26 patients for a total of 140 lesions, including serrated polyps (hyperplastic polyp and sessile serrated lesions or HP and SSL, respectively), adenomatous polyps (low-grade and high-grade called LG AP and HG AP, respectively) and cancerous lesions. Each patient presents

4.1. COHORT PRESENTATION AND MOLECULAR CHARACTERIZATION

a total of 1 to 14 lesions and 1 to 7 time points, reflecting events of surgical treatment. Only one patient was known to present a factor responsible for a high rate of polyp development, the Lynch syndrome.

As RNAseq and whole-exome sequencing (WES) assays require a certain amount of extracted nucleic acid to be performed, not all lesions were included for these techniques. Of the 140 lesions, 89 and 51 yielded enough RNA and DNA to pass quality control, respectively (Table 4.1). In addition to the type of lesion (including HP, SSL, LG AP and HG AP), advanced status was also assessed, reflecting the potential of precancerous lesions to transform into invasive CRC if left untreated. Advanced serrated polyps are SSL with size > 10mm, while advanced adenomatous polyps with at least one of the following features: HG AP, lesions of size > 10mm or lesion with a histological villous component.

Primary Cohort (26 patients, 140 lesions)					
	<i>Serrated Polyps</i>		<i>Adenomatous Polyps</i>		<i>cancer</i>
Immunohistochemistry	<i>n</i> = 46		<i>n</i> = 85		<i>n</i> = 9
<i>n</i> = 140	19 HP	27 SSL	66 LG AP	17 HG AP	
	40 non-advanced	6 advanced	47 non-advanced	36 advanced	
RNA sequencing	<i>n</i> = 33		<i>n</i> = 48		<i>n</i> = 8
<i>n</i> = 89	11 HP	22 SSL	35 LG AP	13 HG AP	
	28 non-advanced	5 advanced	23 non-advanced	25 advanced	
Whole exome sequencing	<i>n</i> = 12		<i>n</i> = 30		<i>n</i> = 9
<i>n</i> = 51	0 HP	12 SSL	21 LG AP	9 HG AP	
	7 non-advanced	5 advanced	11 non-advanced	19 advanced	

Table 4.1: **Experimental design and sample size**

At first, low-risk patients were characterized as having a low number of lesions (1-4 per patient) and no recurrent event (only one time point), while high-risk patients had a high number of polyps (5 or more). Although recurrence status was first considered as the variable to separate the two risk groups, post-inclusion recurrence events (represented as a dotted line in Figure 4.1) in patients presenting a low number of polyps and the presence of a non-recurrent patient with 12 lesions highlighted the need for a more suitable variable.

The number of lesion per year of follow-up was chosen as a more reliable criterion to separate patients by risk groups 4.2A. This criterion led to the establishment of 3 frequency groups, the characteristics of which are shown in Figure 4.2B. Patients in the F1 group develop less than 0.6 polyps per year and do not have recurrent events despite a mean follow-up of more than 5 years. The F2 group includes patients who develop between 0.88 and 2.21 polyps per year and present recurrent lesions despite a mean follow-up duration very close to F1 (5.52 and 5.17 years, respectively). The only F2 patient without recurrence had the shortest follow-up time of F1 and F2, at 2 years and 3 months. Finally, the F3 group presented 3 patients associated with a very high number of polyps

4.1. COHORT PRESENTATION AND MOLECULAR CHARACTERIZATION

per year (ranging from 5.78 to 6) in combination with 3 of the 4 shortest follow-up times, giving a mean follow-up time for F3 of only 2.14 years.

No correlation was observed between precancerous pathways and frequency groups, as the proportions of both pathways were similar between risk groups (Figure 4.2C).

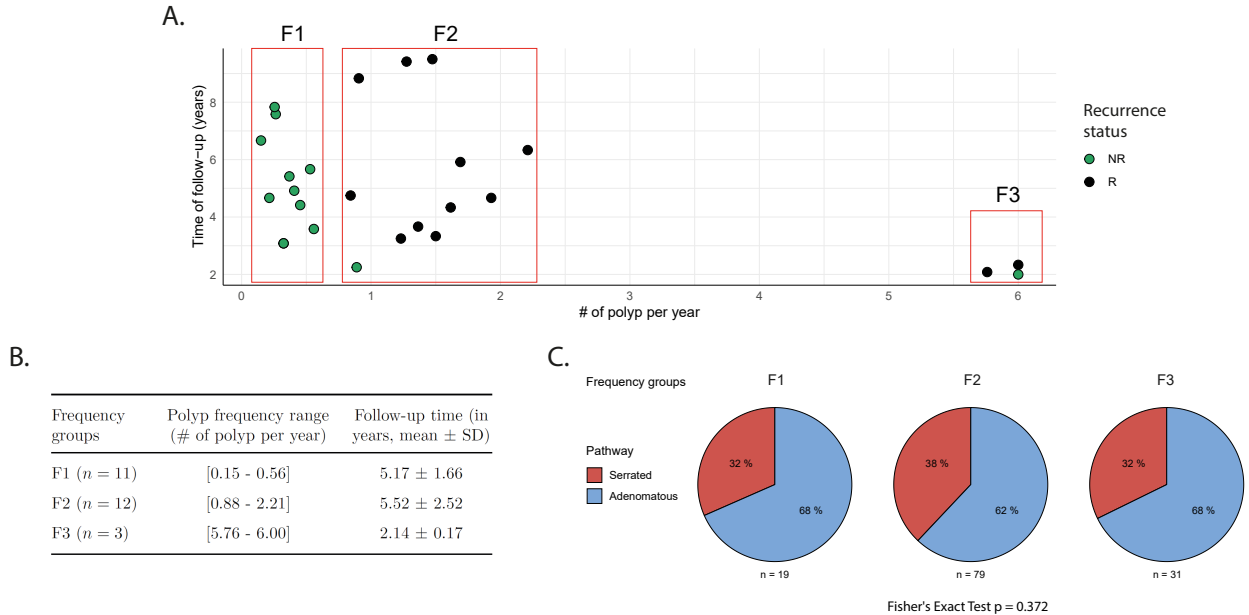


Figure 4.2: **Establishment of a risk variable using the frequency of detected polyps.** A. Patients were divided into three frequency group based on their number of polyp developed per year. B. Frequency group description C. Premalignant pathway distribution between frequency groups

4.1.2 The landscape of somatic alterations between colorectal premalignant pathways

We characterized the mutational profiles of SSLs, APs, and cancer by performing WES and calling for somatic mutations (Figure 4.3A). None of the HP samples yielded enough DNA to perform WES. As expected, specific driver mutations were clearly associated with precancerous pathways (Figure 4.3B). *BRAF* mutations were observed in 92% of SSLs while being absent of APs samples. In contrast, *APC* and *KRAS* mutations were absent in SSL and were present in 73% and 37% of APs, respectively. Driver mutations known to be associated with advanced carcinogenesis, *p53* and *PI3KCA*, were rare in precancerous samples (none in SSLs and, respectively, 7% and 3% in APs) but observed in a third of cancerous lesions. Cancerous lesions also appeared to be enriched in oncogenes of the adenomatous pathway, with 44% and 22% cancer presenting *APC* and *KRAS* mutation, respectively, and only 1 of 9 samples carrying a *BRAF* mutation.

4.1. COHORT PRESENTATION AND MOLECULAR CHARACTERIZATION

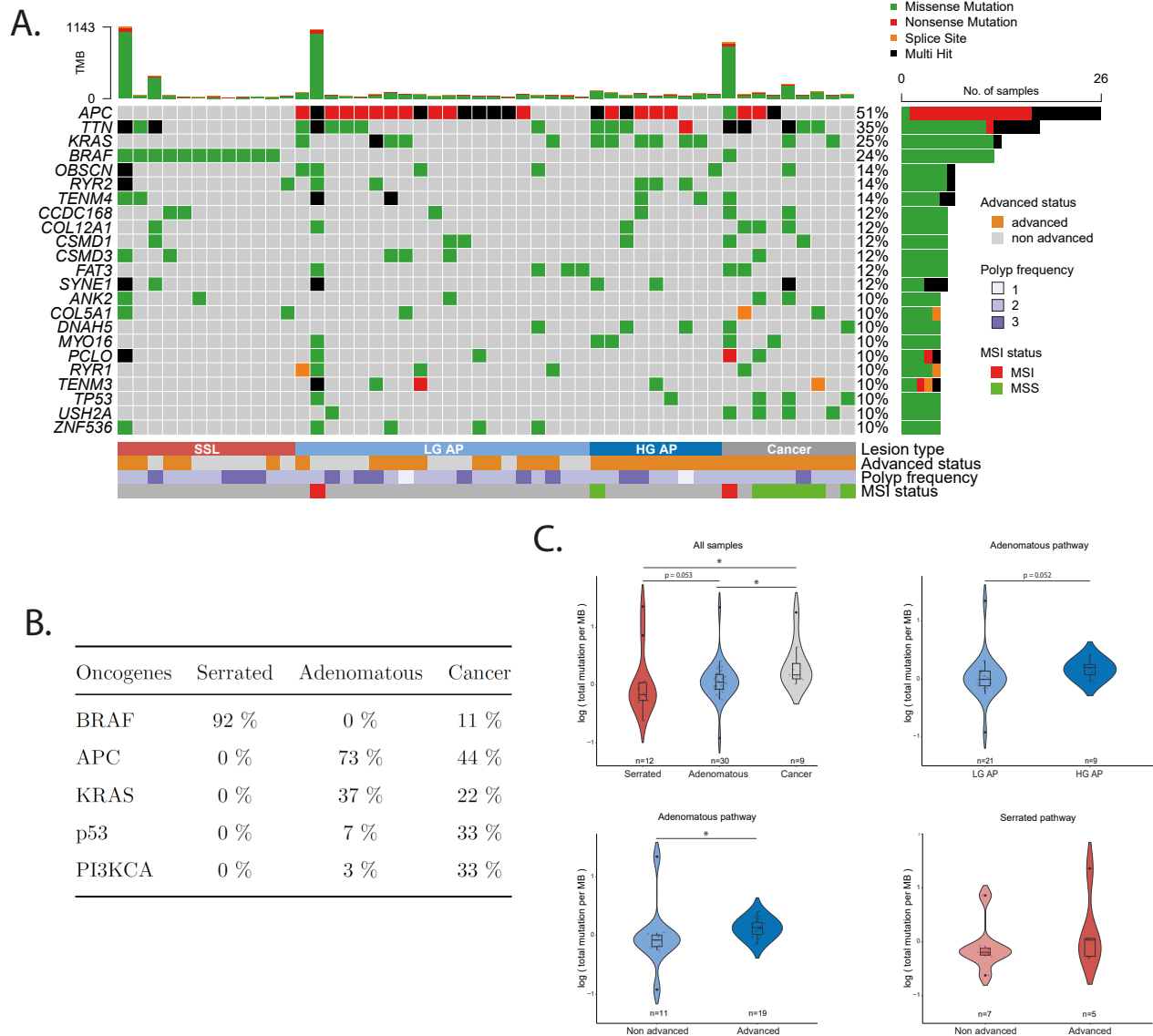


Figure 4.3: **Mutational profile of polyps.** A. Oncoplot displaying somatic mutations present in $\geq 10\%$ of samples. (Top) Mutation burden represented by bar plot. (Right) Percentage of mutations within samples. B. Proportion of mutated oncogene per precancerous pathway and cancer C. TMB levels between various lesions groups. Mann Whitney U test, $*p < 0.05$

4.1. COHORT PRESENTATION AND MOLECULAR CHARACTERIZATION

The total mutation burden (TMB) per sample was also assessed (top of Figure 4.3A), and three samples (1 SSL, 1 LG AP and 1 cancer) were associated with a particularly high TMB, a potential sign of hypermutated phenotype. Unsurprisingly, in high TMB samples, the cancer sample was the only cancer presenting a MSI phenotype, the AP sample is the only sample from the Lynch patient that underwent WES, and the SSL sample is the only SSL in the cohort that exhibited dysplastic characteristics, which has been associated with a higher risk of MSI phenotype [103].

Comparison of TMB between our different lesion groups showed that cancer samples carried more mutations than samples from both precancerous pathways, and the Mann Whitney U test between the serrated and adenomatous pathways produced a p-value of 0.053 (Figure 4.3C), indicating a trend for the adenomatous pathway to be more mutated. Advanced APs carried more mutation than their non-advanced counterparts, and the same tendency was observed between HG APs and LG APs, respectively (Mann Whitney U p-value = 0.052). These results suggested that mutations accumulate during the progression of adenomatous carcinogenesis toward cancer. On the other hand, no difference appeared between non-advanced and advanced SSLs.

4.1.3 CMS classification between colorectal premalignant pathways

CMS status was extracted from RNAseq data, associating each sample with a CMS subtype or indeterminate status. As expected, the serrated pathway was strongly associated with CMS1 (42.4%), a subtype associated with MSI and hypermutation, but also with CMS4 (36.4%). CMS3 only represented 9.1% of samples, and no CMS2 serrated polyp was observed. The indeterminate status represented 12.1% of samples. Interestingly, HPs and SSLs had different CMS profiles, with HPs clearly dominated by CMS4 (54.5%) and SSLs by CMS1 (50%). Proportions of other subtypes were comparable between the two serrated subgroups 4.4. However the general CMS profiles were not statistically distinct (Fisher's Exact Test $p = 0.51$).

APs presented a CMS distribution different from serrated polyps (Fisher's Exact Test $p = 9.21 \times 10^{-9}$), CMS2 (associated with dysregulation of the WNT pathway) being the most represented class (31.3%) after the indeterminate status (33.3%) and followed by CMS3. LG AP and HG AP presented very similar profiles, suggesting a homogeneous CMS distribution among the adenomatous pathway.

4.2. IMMUNOHISTOCHEMISTRY ASSAYS AND ASSOCIATED ANALYSIS

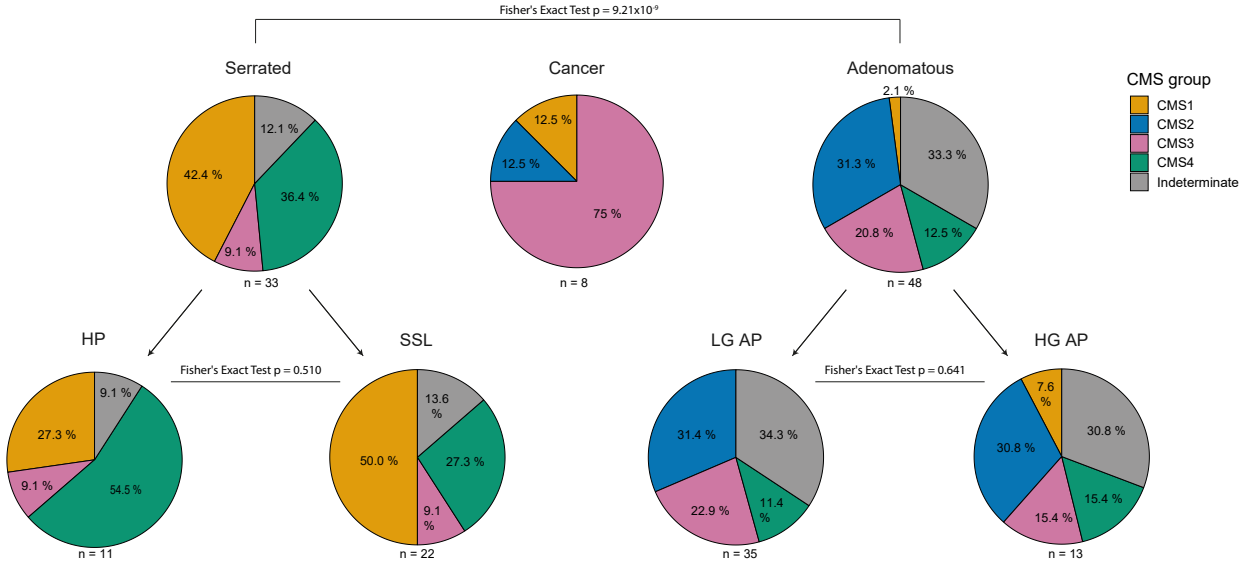


Figure 4.4: Distributions of consensus molecular subtypes (CMS) in the cohort

4.2 Immunohistochemistry assays and associated analysis

IHC is one of the most used methods for studying the presence of an immune cell type in a tissue section. Classic IHC, also called chromogenic IHC, is based on the detection of a given antigen by a specific antibody, which will enable the fixation of a secondary antibody carrying an enzyme, which will catalyze the precipitation of a chemical compound, providing information about the location of the targeted antigen. However, with this approach, tissue analysis is mostly limited to two to three molecules and is based on visual perception, leading to a qualitative analysis more than a quantitative one and high interobserver variability.

To overcome these several limitations, multiplex IHC can be used in combination with digital pathology (DP) [211]. Multiplex IHC encompasses several IHC assays that allow the quantification and localization of a larger number of markers in a tissue section, limiting the number of tissue slides used while increasing the potential for colocalization analysis. DP is defined as a pathological diagnosis transmitted over a distance, together with digital images that allow the diagnosis. Once the sample is digitized, most of the previously described hindrances of classic histopathology can be overcome thanks to the use of DP software.

Most of these software allow for quantitative analysis, for example giving access to the number of cells present in a specific histological compartment or the number of cells positive for a specific marker when looking at IHC staining [212]. Moreover, DP also allows for automated analysis, increasing the reproducibility of the results and reducing the analysis time. Even for analysis where the pathologist's assessment is required (such as tumor classification), DP can help pathologists to screen sample and isolate areas to analyze or assess samples using established scores. Finally, the incorporation of artificial intelligence and machine learning in DP software can increase these benefits tenfold, giving the possibility to assess complex patterns, classify tissue in a more reproducible

4.2. IMMUNOHISTOCHEMISTRY ASSAYS AND ASSOCIATED ANALYSIS

fashion, and detect subtle variations within samples.

To characterize the iTME in samples from the cohort, two different multiplex IHC assays were applied and analyzed using an AI-based digital pathology software. Each assay has specific properties that will be presented in this chapter, in parallel with the associated panels and analysis.

As most precancerous studies highlighted the presence of immunosurveillance in the early stages of premalignant disease and immune escape in more advanced lesions [7, 133, 74], the first IHC panel, called the lymphoid panel, was designed to quantify the presence of the adaptive immune system in the iTME. The other panel was designed to quantify the exhaustion status of TITLs, and referred to as the T-cell exhaustion panel, or TCE panel.

4.2.1 Lymphoid panel analysis workflow

As presented in Figure 4.5A, the lymphoid panel is a combination of markers that allows phenotyping of the main T and B lymphocyte populations in the iTME: CD4⁺ and CD8⁺ T cells, regulatory T cells, B cells and plasma cells. This panel uses multispectral technology, allowing high-throughput staining and image processing of six markers associated with a DAPI staining for nuclei detection. The multispectral technology is based on the use of a multispectral microscope, which is capable of acquiring a wide emission spectrum from a stained slide. Deconvolution of this spectral image using a spectral library containing the emission spectrum of each fluoro-chrome used will provide a specific signal for each stained marker, as shown in Figure 4.5B. This approach also greatly limits the autofluorescence-related issues encountered in classical fluorescence microscopy [213]. Staining of the cohort with the lymphoid panel revealed the presence of numerous lymphoid structures, as shown in Figure 4.5C.

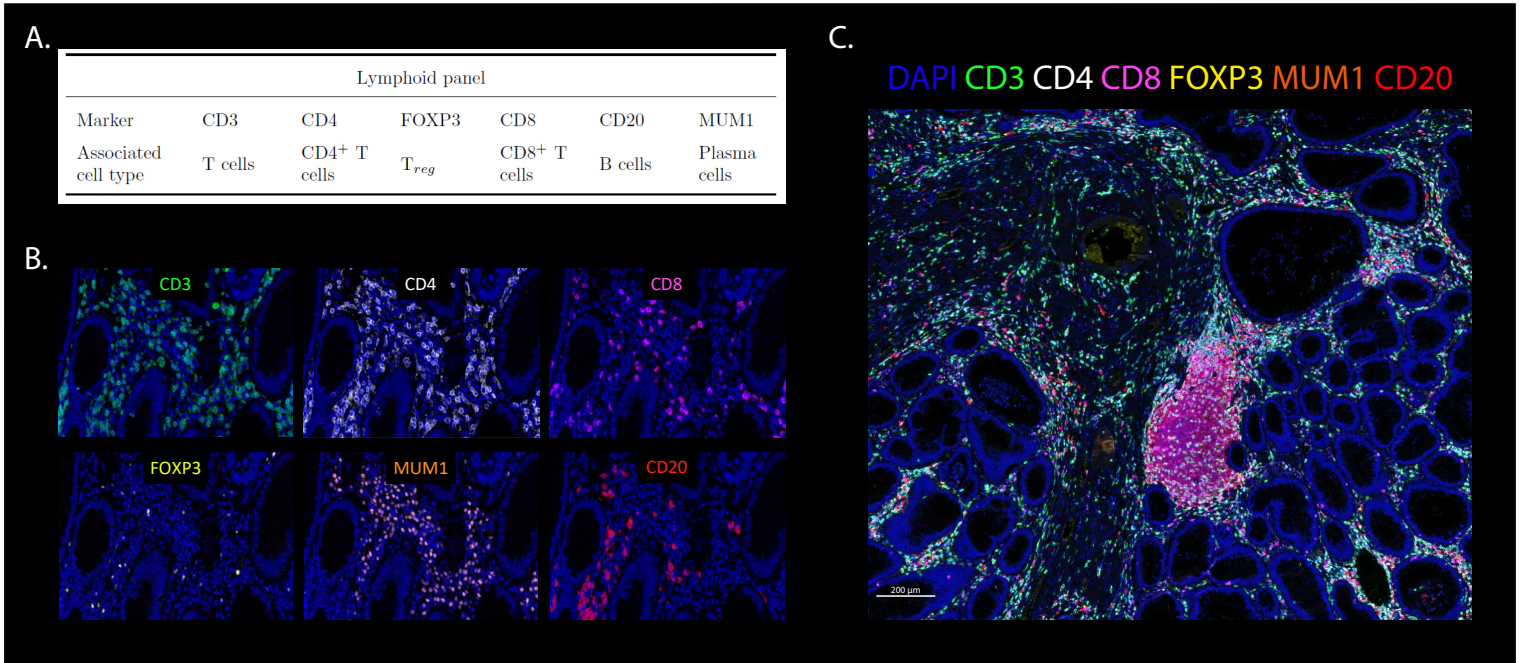


Figure 4.5: **Lymphoid panel presentation.** A. Lymphoid panel markers. B. Unmixed signal of lymphoid panel markers. C. Multichannel image of a TLS stained by the lymphoid panel

Following their acquisition and deconvolution, images from the lymphoid panel were analyzed using an AI-based digital pathology software. A tissue-classifying algorithm was developed and applied to the cohort, enabling the software to classify our tissue into three compartments: stromal, epithelial, and TLS (Figure 4.6)A.

As presented in Figure 4.6B, pathologist lesion annotations were then reported on the fluorescence image and labeled center of tumor (CT). Then a 1mm wide margin was drawn around the CT and divided into stromal and epithelial compartment using the previously described tissue classifier, creating the adjacent epithelium (AE) and invasive margin (IM) ROIs, respectively.

Of note, all detected TLSs were reintegrated into their respective ROIs for the cell density analysis that will be briefly described. For each ROI, AI-based nuclei detection was performed using DAPI signal. Then, for each marker, a detection threshold was established, discriminating between positivity and negativity for this given marker for every detected cell. Finally, each cell was assigned a phenotype based on its marker combination (e.g., CD3⁺CD4⁺FOXP3⁺ cells were associated with the T_{reg} phenotype).

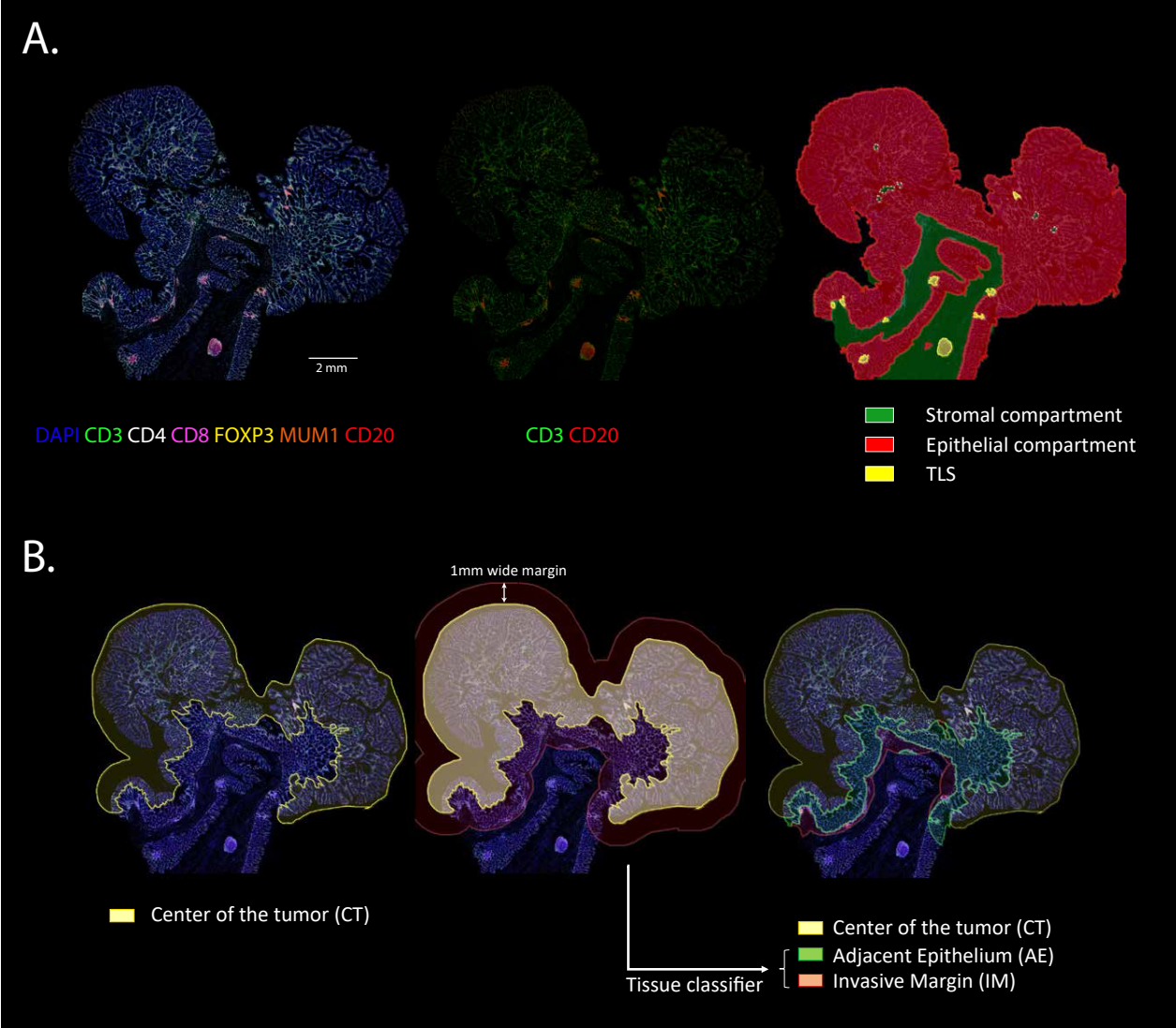


Figure 4.6: **Definition of ROIs and TLS using an AI-based classifier.** A. AI-based tissue classifier was trained on DAPI, CD3 and CD20 signals to identify stromal, epithelial and TLS compartments. B. Definitions of 3 ROIs using pathologist annotations and tissue classifier.

4.2.2 T cell exhaustion panel analysis workflow

As presented in 4.7A, the TCE panel is a combination of eight markers that allow for the quantification of the exhaustion status of T cells. For this panel, all markers except cytokeratin (CK) can be found on a single cell, making multispectral IHC obsolete as this technology is considered to not adequately deal with the colocalization of more than 4 markers. The Brightplex[®] technology, which does not present this limitation, was thus chosen for this panel.

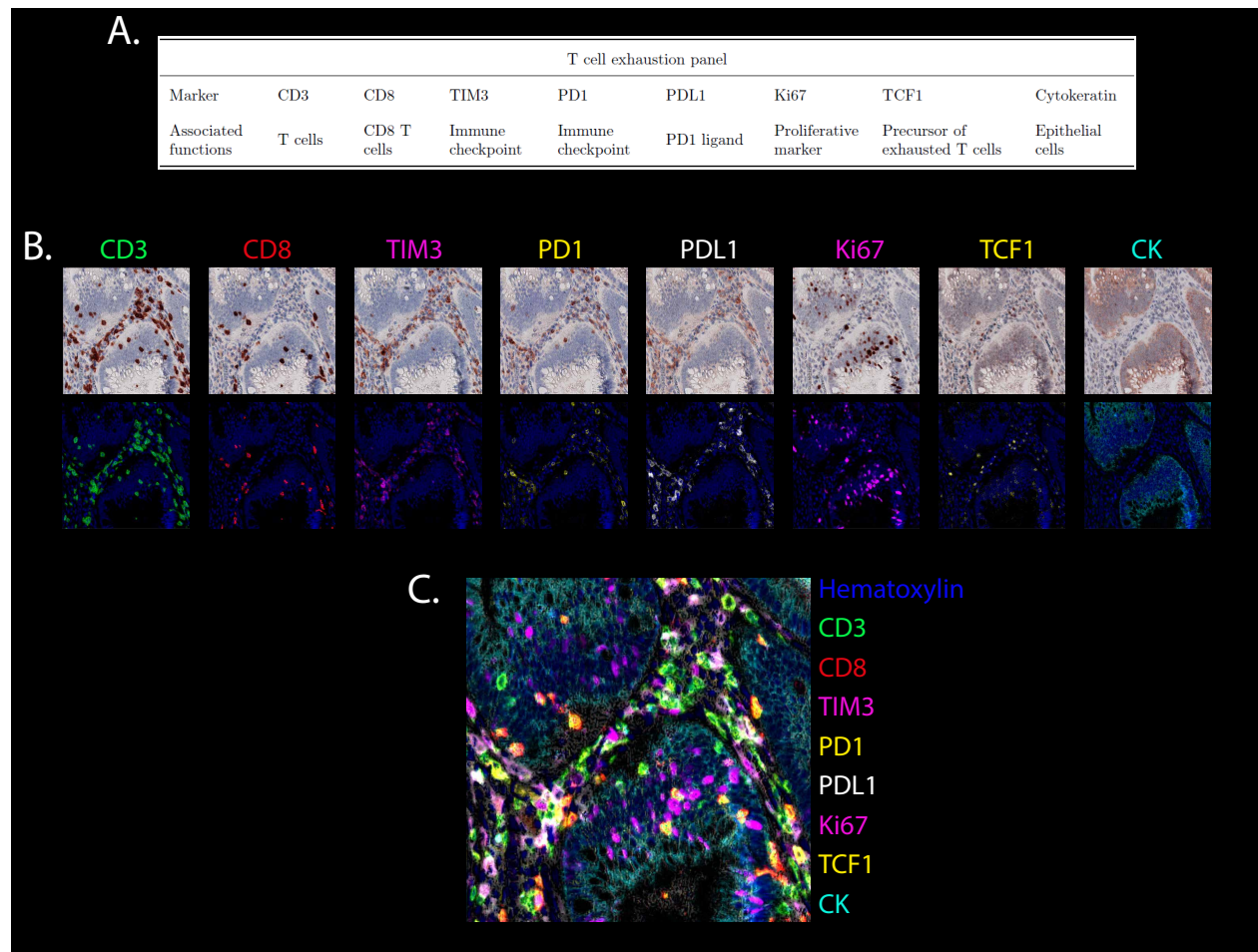


Figure 4.7: **TCE panel presentation.** A. TCE panel markers. B. Chromogenic images (top) and extracted signal (bottom) after deconvolution. C. Multichannel image after fusion of extracted signals from the TCE panel.

Brightplex[®] technology is based on an iterative staining approach. The slide is stained following a basic IHC workflow, leading to a classic chromogenic staining. After brightfield acquisition of this image, staining is removed and the antibody is stripped, allowing for another round of staining using the next antibody. Once images were acquired for all markers of interest, as shown in the upper panel of 4.7B, specific signals were extracted using our DP software deconvolution algorithm (4.7B, bottom). After a synchronization step, all signals can be merged into one multi-channel image

4.2. IMMUNOHISTOCHEMISTRY ASSAYS AND ASSOCIATED ANALYSIS

(4.7C). This multi-channel image can then be treated similarly to the fluorescence image from the lymphoid panel, leading to quantification of cell density for our various phenotypes in CT, AE, and IM. As this panel does not include B cell markers, TLS annotations were reported from the analysis of the lymphoid panel.

4.2.3 Immunohistochemistry analysis

In addition to the cell density analysis, the tissue-classifying algorithm presented in 4.2.1 allowed the quantification of several parameters associated with TLS. First, for each lesion, the percentage of the surface occupied by TLS (called TLS surface) was assessed by dividing the surface of the TLS compartment by the total area of analyzed tissue.

Each TLS was then classified into various categories based on its pattern of CD3, CD20, PD1 and Ki67 staining. As presented in Figure 4.8, TLS without a clear B cell zone were classified as unstructured and intermediate, depending on their B cell density. TLSs with a well-defined B cell zone but no germinal center (characterized by a cluster of CD20⁺Ki67⁺ proliferating B cells surrounded by a crown of CD3⁺PD1⁺ T_{fh}) were defined as structured. Finally, TLSs presenting a B cell zone and a germinal center were classified as mature.

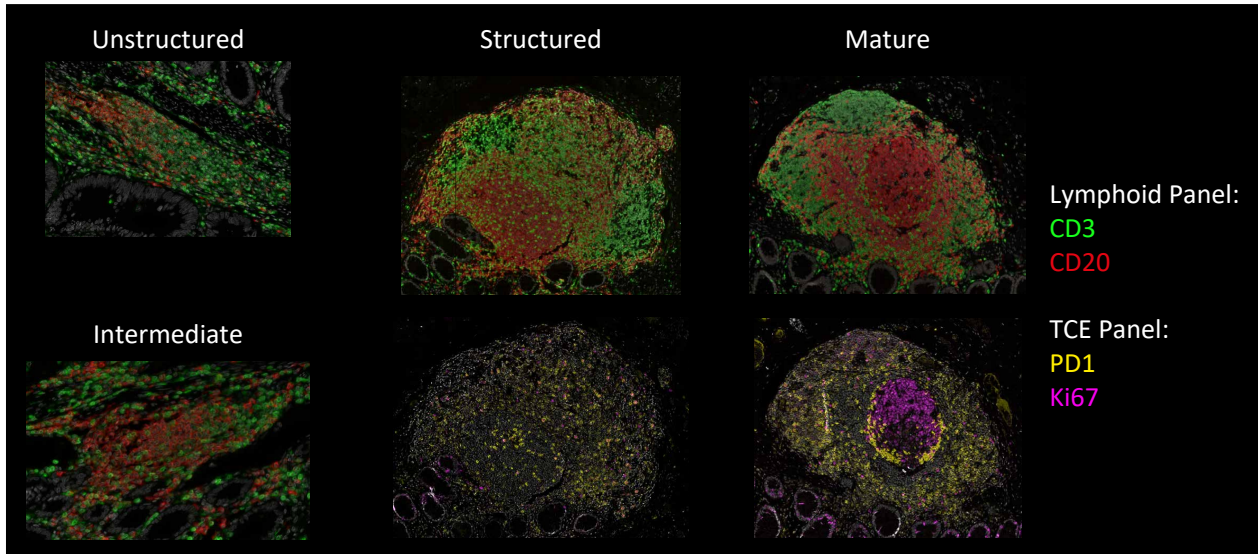


Figure 4.8: **Definition of TLS types**

Taking advantage of these two multiplex IHC panels and their cell density analysis presented above, the densities of numerous cell phenotypes were quantified in CT, AE and IM. Only biologically relevant phenotypes were extracted from the two panel analysis (e.g. CD3⁺CD20⁺ or CD3⁺CK⁺ were not included). The selected phenotypes are presented in 4.9A. Due to experimental issues, CD8 staining of the lymphoid panel was defective in approximately 20% of samples. Consequently, T cells and CD8⁺ densities were extracted from the analysis of the TCE panel.

4.2. IMMUNOHISTOCHEMISTRY ASSAYS AND ASSOCIATED ANALYSIS

To better characterize the various ROIs isolated during the IHC analysis, densities of all selected IHC phenotypes and TLS surface were taken in account to perform unsupervised hierarchical clustering (UHC) analysis of ROIs for all lesions of the cohort.

A clear clustering of IM ROIs was observed, while CT and AE appeared to present closer features (4.9B). Unsurprisingly, row cluster 1, mainly composed of epithelial cells, was the clearest determinant for segregation between ROIs. Most IM compartments display particularly low densities of epithelial cells compared to the epithelial compartments CT and AE. Interestingly, plasma cells appeared to follow the same trends as epithelial cells, being particularly low in the IM compartments.

UHC analysis separated IM into two clusters, one presenting high levels of TILs, while the other seemed to be in general poorly infiltrated.

Although AE and CT presented very close profiles, plasma cell densities appeared particularly high in AE, while CT presented higher levels of epithelial cells (CK cells) and proliferating epithelial cells (CK⁺Ki67⁺ cells) which might be due to the presence of dysplasia in CT only. As AE and CT have similar profiles and for clarity reasons, only CT was selected for further analysis.

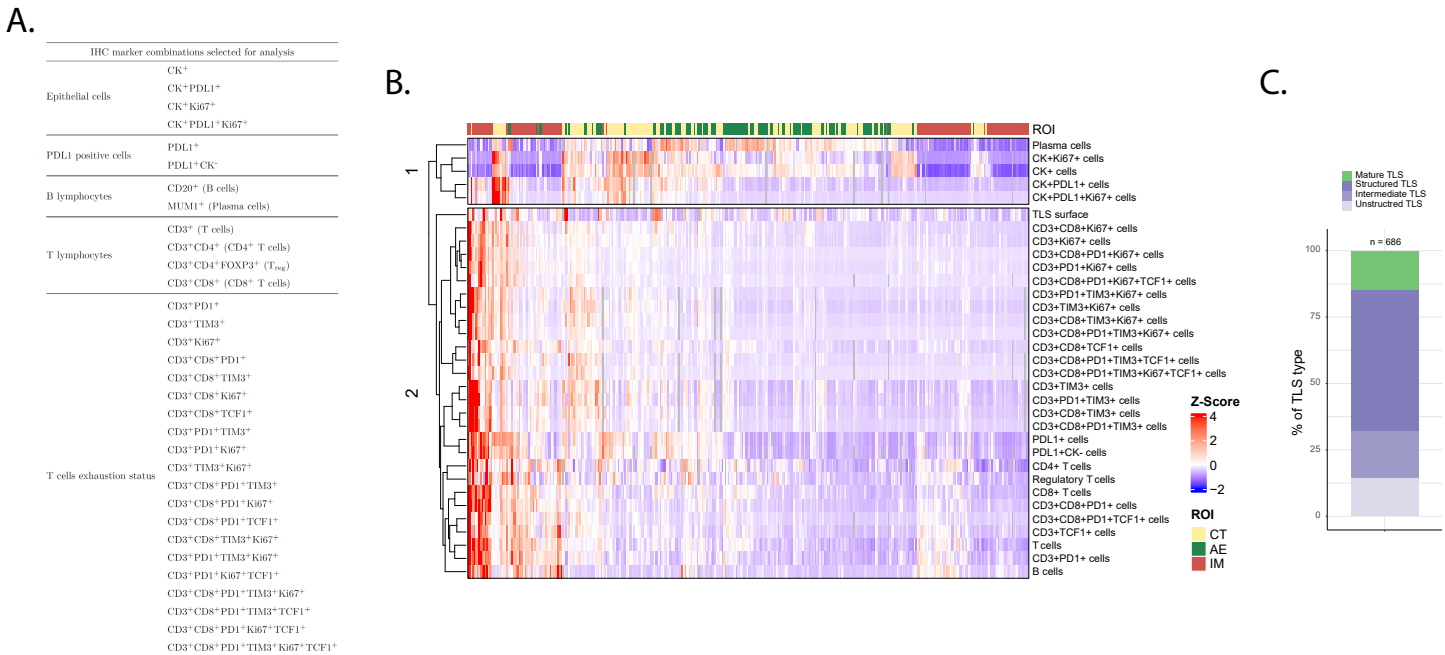


Figure 4.9: Data extracted from the IHC analysis A. Selected phenotypes from the lymphoid and TCE panels. B. Heatmap representing two-dimensional UHC by IHC phenotype and sample ROIs. Row splitting is based on the UHC dendrogram C. Proportion of TLS types within the whole cohort.

A total of 686 TLS were detected using our analysis pipeline, among which 14.7% were mature TLS. Quantified immature TLS are mostly structured (53.4%), with only 17.6% and 14.7% of intermediate and unstructured TLS, respectively (4.9C).

4.3 Colorectal premalignant pathways harbor distinct iTME

4.3.1 Interpathway analysis

Along with samples from our cohort, three other colorectal cohorts of our team (composed of primary and metastatic CRC), underwent RNA sequencing. To visualize the continuum of the whole carcinogenesis process, a tSNE analysis (dimensionality reduction technique) was performed on the total gene expression data of these samples, and the results are presented in Figure 4.10. This visualization method suggested that gene expression profiles follow a stereotyped progression from early carcinogenesis to metastatic CRC.

Four main clusters were identified: APs, serrated polyps, primary CRC, and metastatic samples. The progress of carcinogenesis appeared to go from the bottom left to the top right of the graph, and within the adenomatous cluster, advanced polyps were closer to the cancer samples than non-advanced ones.

Primary tumors from our cohort were observed to cluster with primary CRC from other cohorts, while the three non adenomatous samples clustering with APs presented specific profiles. These three outliers included the only SSL with dysplasia, the only serrated sample from the Lynch syndrome patient, and a primary CRC presenting a large precancerous area.

Although both precancerous groups presented gene expression profiles specific to early carcinogenesis, their clear segregation also suggested pathway-specific transcriptomic profiles.

To investigate whether these pathway specific gene expression profiles were also observed at the iTME level, we relied on a gene-expression-based TME deconvolution tool, the Consensus^{TME} algorithm [205]. Using a specific set of gene signature for a given cell type, each sample was assigned scores representing the relative abundance of this cell type within the TME. UHC analysis was performed on the basis of these various TME scores and plotted in 4.11. Signatures are presented in A.2

This analysis identified two main clusters, a hot cluster scoring high for all TME cells, and a cold one, presenting low scores for the majority of Consensus^{TME} scores. The serrated pathway represented 63.7% of the hot cluster while only accounting for 13.5% of the cold cluster samples, suggesting a more infiltrated TME in serrated polyps. Within the cold group, LG APs appeared to be associated with the lowest infiltration levels.

We then evaluated iTME within precancerous lesion using cell densities extracted from IHC analysis. After performing a principal component analysis (PCA) based on all phenotypes and ROIs, we projected our samples into the first two principal components of the PCA, revealing a clear serrated subset in the lower left quadrant, while the adenomatous samples appeared more heterogeneously dispersed (Figure 4.12A). In accordance with the Consensus^{TME} results, these results suggested a distinct iTME between adenomatous and serrated pathways.

4.3. COLORECTAL PREMALIGNANT PATHWAYS HARBOR DISTINCT ITME

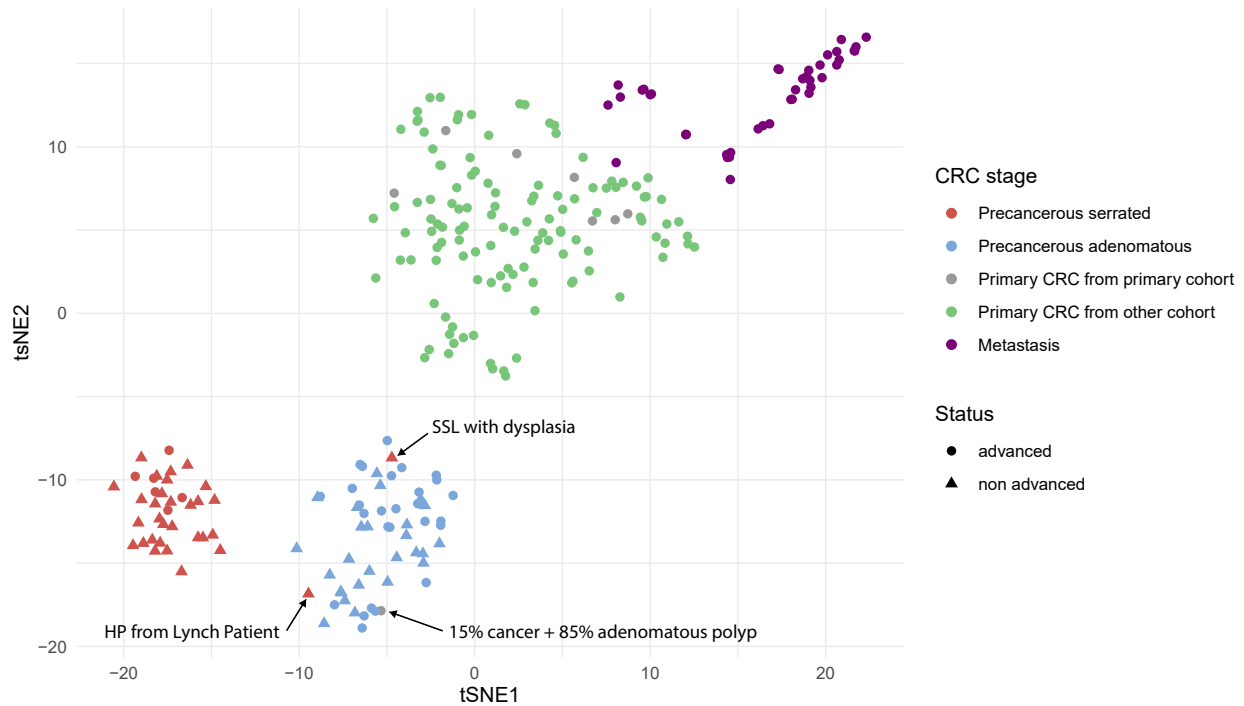


Figure 4.10: **Colorectal precancerous pathways exhibit distinct gene expression profiles.** tSNE analysis and visualization of gene expression profiles of various colorectal lesions

TLS analysis per pathway showed that the TLS surface was higher in serrated precancerous lesions and that the proportion of TLS types was similar between the two pathways (Figure 4.12B).

Mirroring PCA visualization, UHC analysis and plotting of CT cell densities of precancerous lesions identified a clear serrated cluster, while adenomatous samples were more dispersed (Figure 4.13A). Surprisingly, serrated lesions seemed to be associated with low levels for most cell types, except plasma cells.

These observations were statistically validated and presented in Figure 4.13B. Only $CD8^+$ T cells, $CD3^+TIM3^+$ cells and $CD3^+CD8^+TIM3^+$ cells presented similar CT densities between the two pathways, while CT plasma cell densities were higher in serrated polyps. All other phenotypes of B and T cells were more infiltrated in APs compared to serrated lesions. APs also had higher densities of $PDL1^+$ cells, $PDL1^+$ epithelial cells and proliferating epithelial cells ($Ki67^+CK^+$).

The same analysis was then performed on IM cell densities. Although the UHC analysis did not reveal a clear serrated cluster, serrated samples again appeared to be associated with lower general infiltration levels (Figure 4.14A). Statistical analysis showed that $PDL1^+$ cells and all immune cell types (except B cells) were found at higher densities in APs (Figure 4.14B).

4.3. COLORECTAL PREMALIGNANT PATHWAYS HARBOR DISTINCT ITME

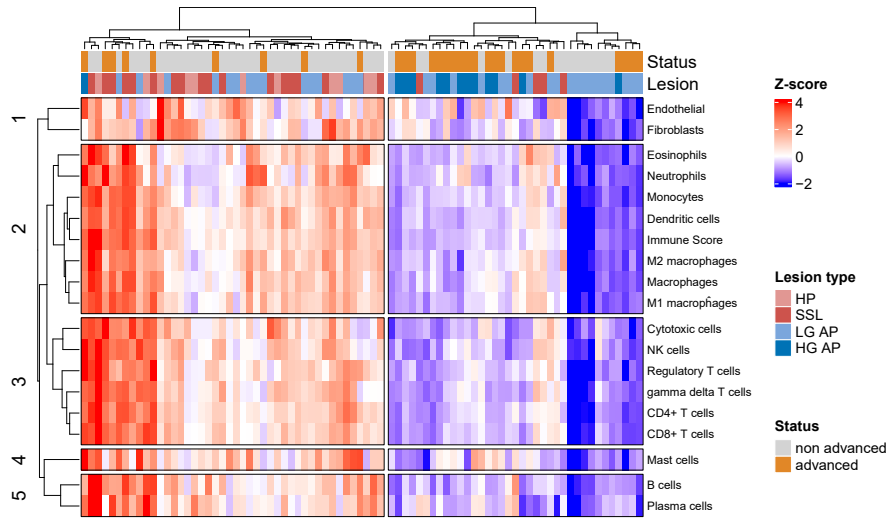


Figure 4.11: **Colorectal precancerous pathways exhibit distinct immune signature patterns** Heatmap representing two-dimensional UHC by Consensus^{TME} scores and precancerous lesions. Row and column splitting are based on the UHC dendrogram

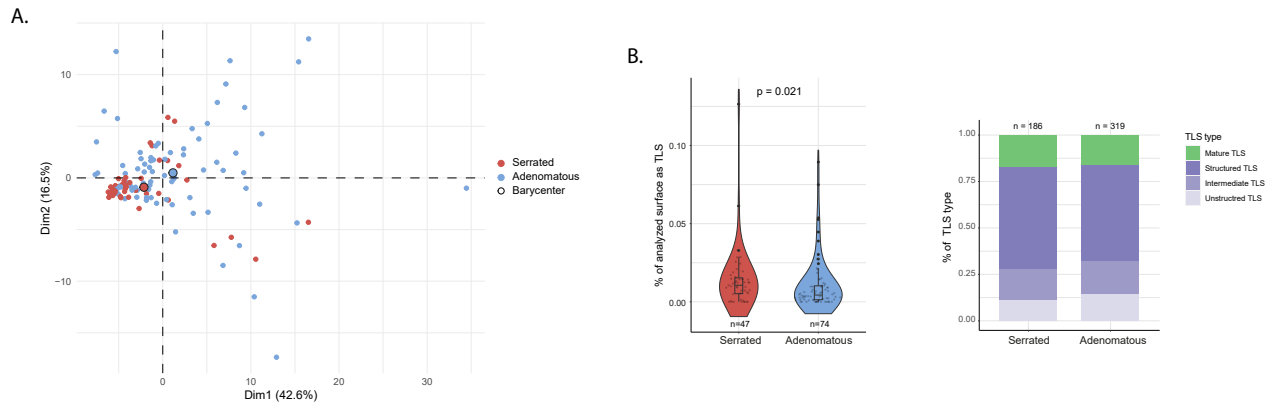


Figure 4.12: **Colorectal precancerous pathways exhibit distinct iTME pattern.** A. PCA analysis and representation of samples per pathway based on their IHC phenotype densities in CT, AE and IM. B. TLS profiles between the two pathways. Mann Whitney U test

4.3. COLORECTAL PREMALIGNANT PATHWAYS HARBOR DISTINCT ITME

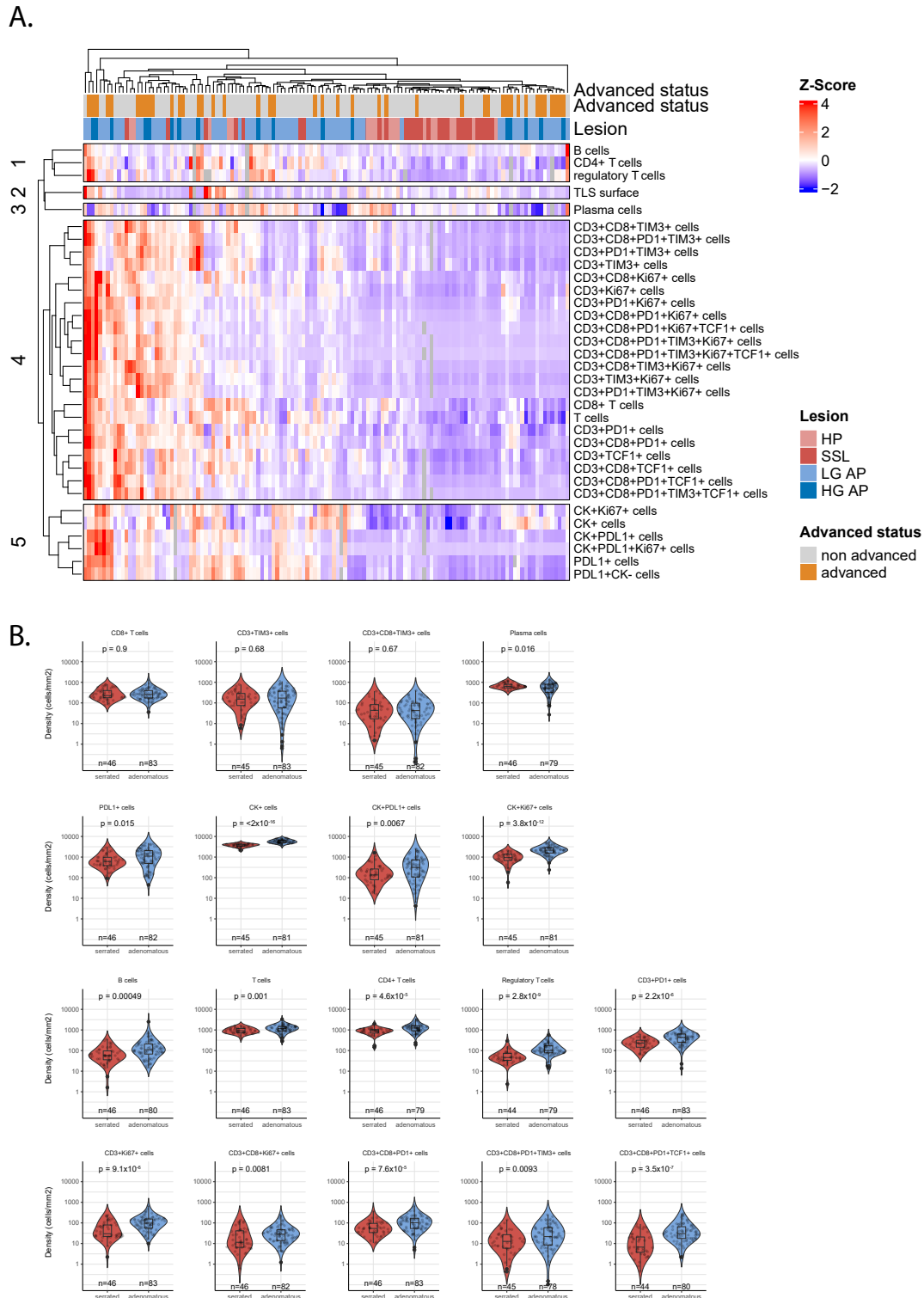


Figure 4.13: **Adenomatous polyps present a more infiltrated iTME at the CT** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in CT and sample. Row splitting is based on the UHC dendrogram. B. Densities of IHC phenotypes in the CT of serrated and adenomatous pathways. Mann Whitney U test

4.3. COLORECTAL PREMALIGNANT PATHWAYS HARBOR DISTINCT ITME

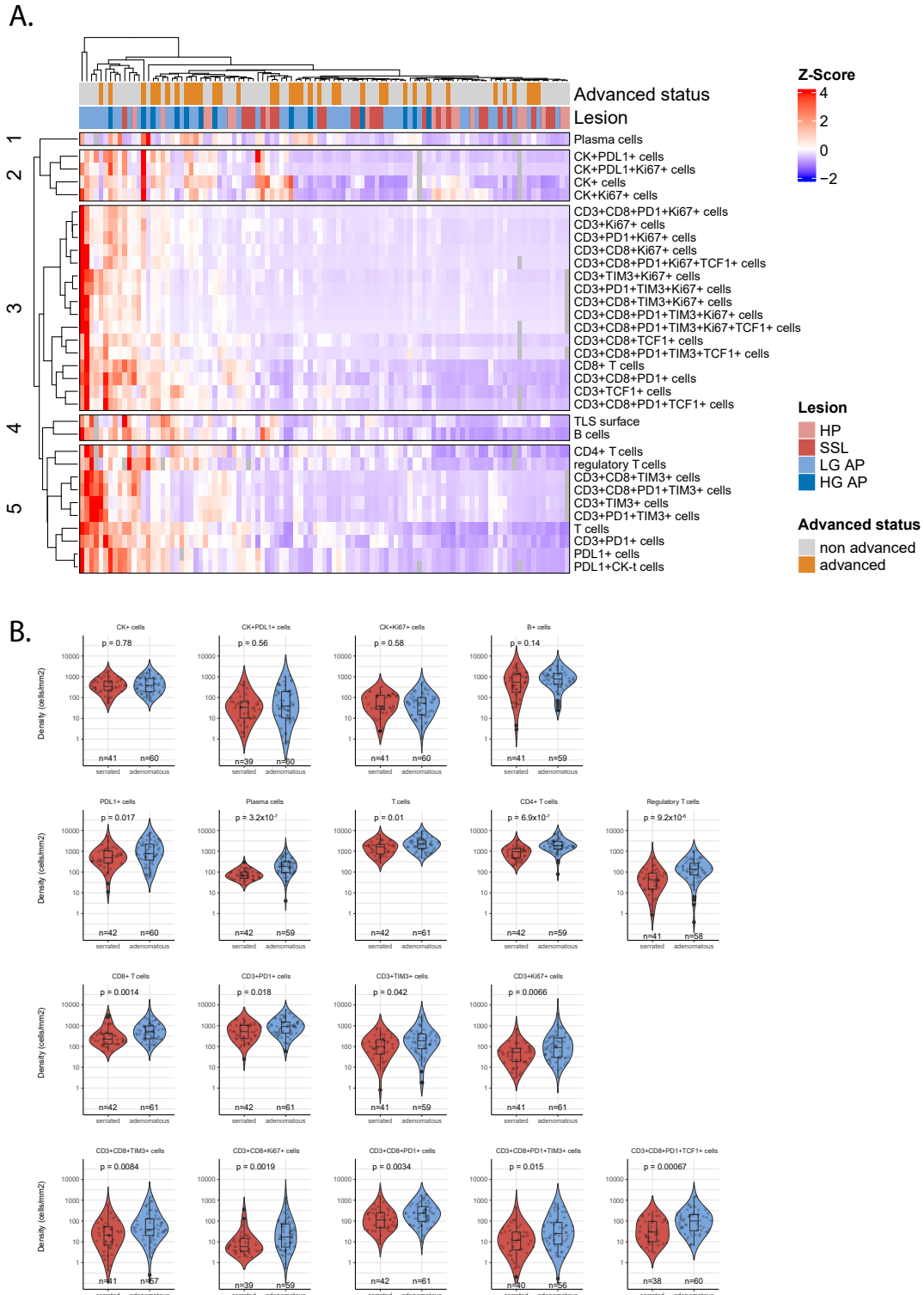


Figure 4.14: **Adenomatous polyps present a more infiltrated iTME at the IM** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in IM and samples. Row splitting is based on the UHC dendrogram. B. Densities of IHC phenotypes in the IM of serrated and adenomatous pathways. Mann Whitney U test

4.3. COLORECTAL PREMALIGNANT PATHWAYS HARBOR DISTINCT ITME

Analysis of Consensus^{TME} scores and IHC densities presented adverse conclusions, with Consensus^{TME} scores presenting the serrated iTME as more infiltrated than the adenomatous one, while the opposite result came from the IHC analysis (Figure 4.15A).

We hypothesized that observed discrepancies could be due to TME scores being a relative abundance score, whereas IHC densities are absolute parameters. Indeed, TME scores are defined as "normalized enrichment scores for each of the cell types representing the relative abundance of cell types across multiple samples" [205]. This hypothesis was primarily based on the idea that serrated and adenomatous have a very different histological structure and total cell density. As serrated lesions very rarely present dysplasia, their epithelial cell density is expected to be lower than their adenomatous counterparts. In general, as presented at the top of Figure 4.15B, serrated and adenomatous polyps are histologically very distinct, and the total cell density in APs appears much higher than in serrated polyps.

We confirmed this idea by extracting the CT total cell density from the TCE panel analysis, resulting in a mean total cell density of 9495 cells/mm² in APs against 7113 cells/mm² in serrated polyps (Figure 4.15B, bottom).

A relative abundance score (i.e., proportion of total cells) was extracted from IHC densities by dividing the phenotype densities by the total cell density. Analysis of these percentages demonstrated that serrated polyps presented the same proportion of CD4⁺ T cells and a lower proportion of CD8⁺ T cells than APs.

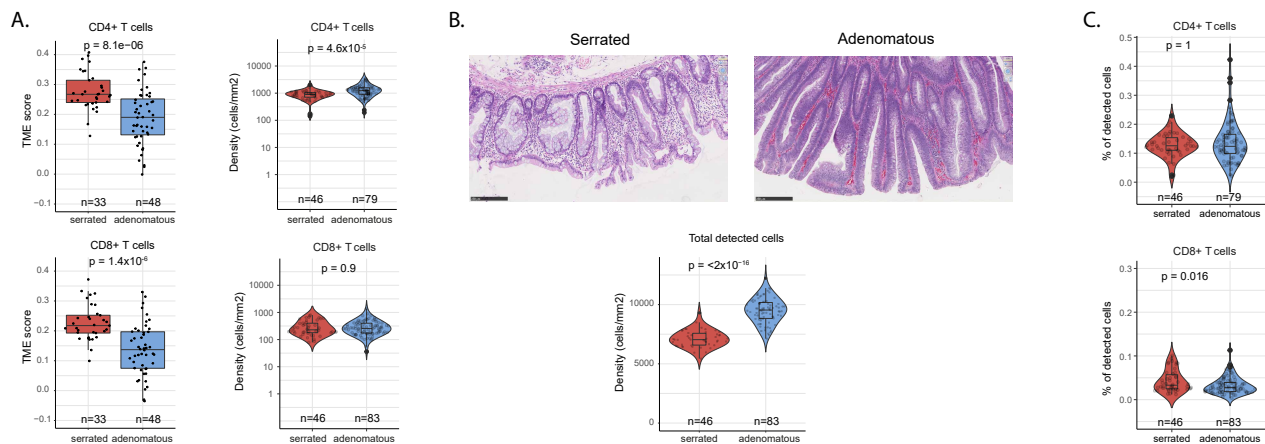


Figure 4.15: Total cell density explains differences between Consensus^{TME} scores and IHC densities in premalignant pathways. A. CD8⁺ and CD4⁺ T cells Consensus^{TME} score and CT densities in serrated and adenomatous pathways B. (top) H&E representative image of serrated and adenomatous polyps (bottom) Total cell density in CT of serrated and adenomatous pathways. C. CD8⁺ and CD4⁺ T cells as proportion of total cells in CT of serrated and adenomatous pathways.

Therefore, the discrepancies in the estimation of TME between IHC densities and Consensus^{TME} can be at least partially explained by histological differences between the two premalignant colorectal

4.3. COLORECTAL PREMALIGNANT PATHWAYS HARBOR DISTINCT iTME

pathways. It appears that a group of samples presenting such a difference should not be compared using gene-expression-based TME deconvolution tools.

4.3.2 Intrapathway analysis

After comparing the iTME between serrated and adenomatous lesions, we focused on the potential variation of the iTME within each pathway. Within the adenomatous pathway, IHC densities were compared between grades (LG AP vs. HG AP) and advanced status. In serrated lesions, the grouping was based on the type of lesion (HP vs. SSL) and the advanced status of SSLs.

In the serrated pathway, the CT and IM densities of the main cell types (PDL1⁺ cells, B and plasma cells, CD4, CD8 and regulatory T cells) were not affected by the type of lesion or the advanced status of SSLs (Figure 4.16). However, when comparing HP with SSL, HP had higher CT densities of exhausted T cell precursors, characterized by the CD3⁺CD8⁺PD1⁺TCF1⁺ phenotype. When focusing on the advanced status of SSLs, it was noted that CT of non-advanced SSL presented more proliferating epithelial cells (CK⁺Ki67⁺). IM of non-advanced SSLs also displayed higher densities of T cell subtypes expressing the immune checkpoint TIM3 (CD3⁺TIM3⁺, CD3⁺CD8⁺TIM3⁺ and CD3⁺CD8⁺PD1⁺TIM3⁺ phenotypes), suggesting higher exhaustion levels in non-advanced SSLs.

In the adenomatous pathway, LG AP and HG AP presented similar densities for all phenotypes in both CT and IM. Densities for the main cell populations are presented in Figure 4.17A. When grouping by advanced status, plasma cells and T_{reg} were observed to have lower and higher densities, respectively, in the CT of advanced APs, while other cell densities were similar in both CT and IM (Figure 4.17B).

Overall, iTME did not appear to fluctuate much during adenomatous lesion development, except CT plasma cells and T_{reg} that seemed to correlate with advanced status.

In conclusion, within colorectal premalignant pathways, the iTME does not seem to be shaped by the lesion type, grade or advanced status. However, when looking at iTME profile within each pathway, a lot of heterogeneity can be observed. This is especially true in the IM compartment of both adenomatous and serrated polyps. (4.13A).

4.3. COLORECTAL PREMALIGNANT PATHWAYS HARBOR DISTINCT ITME

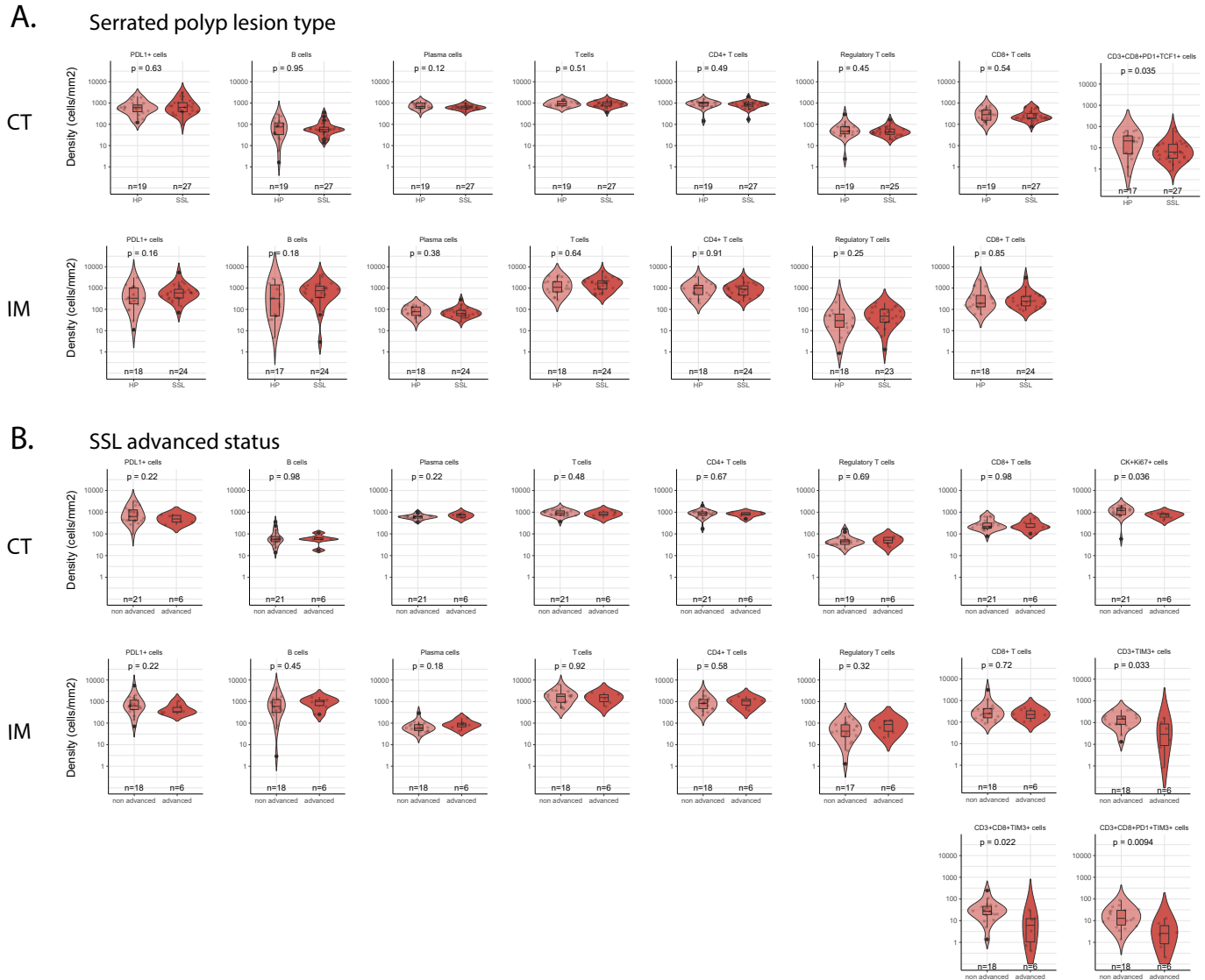
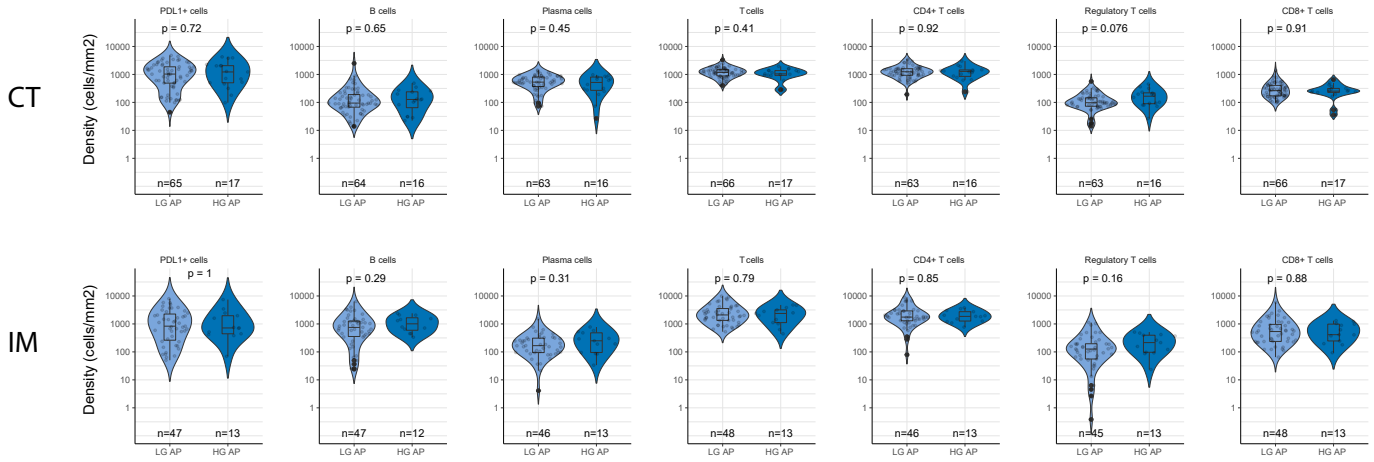


Figure 4.16: **Serrated lesions present an homogeneous iTME profile** A. Densities of IHC phenotypes in HPs and SSLs. B. Densities of IHC phenotypes in non advanced and advanced SSLs. Mann Whitney U test

4.3. COLORECTAL PREMALIGNANT PATHWAYS HARBOR DISTINCT iTME

A. Adenomatous polyp grade



B. Adenomatous polyp advanced status

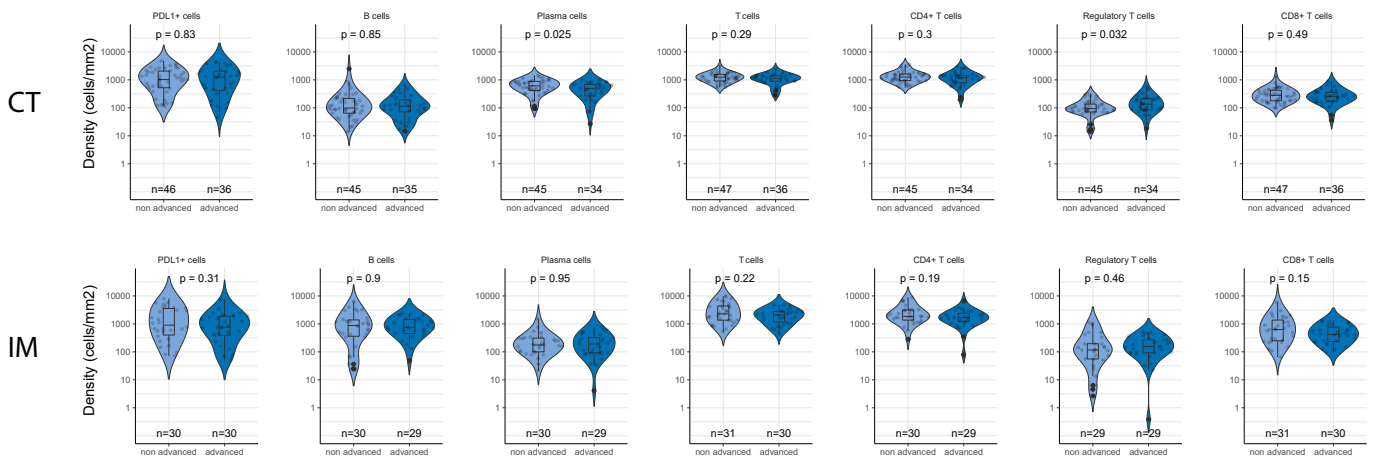


Figure 4.17: Adenomatous lesions present an homogeneous iTME profile A. Densities of IHC phenotypes in LG APs and HG APs. B. Densities of IHC phenotypes in non advanced and advanced APs. Mann Whitney U test

4.4 Polyp development rate is associated with variation of iTME of pre-malignant lesions

In order to assess if intrapathway variability can be associated with the patient risk to develop polyps, we performed all previously described analysis in regards of frequency groups F1, F2 and F3, starting by molecular characterization.

4.4.1 Molecular characterization of frequency groups

As previously presented, frequency groups were not associated with a specific pathway (Figure 4.18A).

When looking at driver mutations between groups (Figure 4.18B), it was first noticed that only two F1 samples were sequenced. Both F1 samples are adenomatous lesions carrying *KRAS* mutation, and one also presented an *APC* mutation. As serrated samples presented a very homogeneous profiles with all samples but one presenting *BRAF* mutation, F2 and F3 serrated samples driver mutations profiles were unsurprisingly very similar. Finally, F2 and F3 APs also presented similar profiles, with 74% and 78% of samples carrying a *APC* mutation, respectively. 32% of F2 APs and 33% of F3 APs harbored a mutated *KRAS*, while only F2 samples presented *p53* and *PI3KCA* mutations (11% and 5%, respectively). Furthermore, no significant differences were found when comparing TMB between frequency groups (Figure 4.18C). Overall, these results suggest that frequency groups are not associated with a particular mutational profile.

Finally, we looked at the CMS subtype per frequency group within each pathway (Figure 4.18D). Once again, the F1 group was underrepresented with only two serrated samples (both were Indeterminate) and six adenomatous samples (three indeterminate, two CMS2 and one CMS1). Overall, for both pathways, no significant difference was found between F2 and F3 CMS distribution, suggesting that frequency groups are not associated with a particular CMS profile.

4.4.2 First glance at the iTME between frequency group

We assessed the general aspect of the iTME between frequency groups by looking at the principal component analysis results presented in Figure 4.12A, this time splitting samples by frequency group (Figure 4.19A). It can first be noted that the number of sample per frequency group is disparate, with F2 representing 58.7% (71 of 121) of samples processed in IHC, while F1 only account for 15.7% of them (19 of 121). Barycenters of all groups were distinct, suggesting different iTME profiles depending on the frequency group, and F1 lesions appears particularly separated from others.

As presented in the left panel of Figure 4.19B, TLS surface was shown to be greater in F1 polyps when compared to F2 and F3 groups. TLS structure and maturation status also appeared to differ according to the risk groups of the patients. F1 samples presented the highest proportion of mature

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

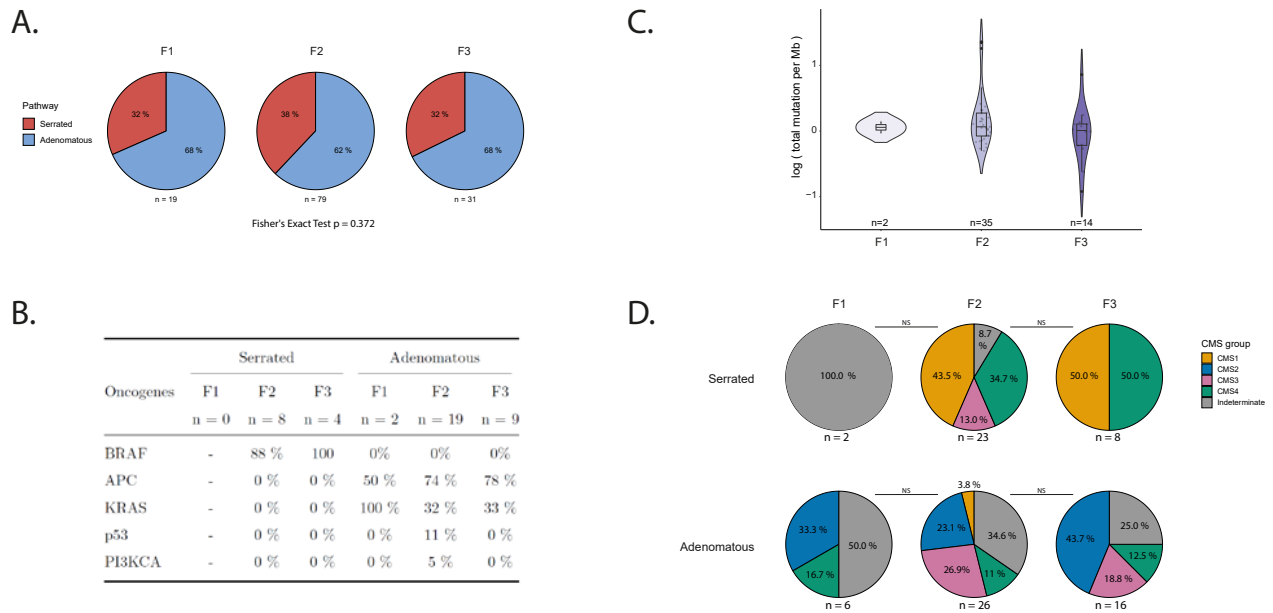


Figure 4.18: **Frequency groups present similar molecular profiles** A. Premalignant pathway distribution between frequency groups B. Proportion of mutated oncogene per frequency group and pathway C. TMB levels between frequency groups D. Distributions of consensus molecular subtypes (CMS) per frequency group and pathway

TLS (32.3%), followed by F2 and F3 polyps, with 13.8% and 5.43%, respectively (right panel of Figure 4.19B). Structured TLS proportion followed the opposite trend, increasing from F1 to F2 to F3, and differences in the proportion of the TLS subtypes between frequency groups were validated by chi-squared test ($p = 3.255 \times 10^{-8}$).

In the serrated pathway, the surface of TLS did not appear to vary with frequency group, and although the serrated samples of F3 seem to present a lower percentage of mature TLS, the chi-square test did not confirmed it (Figure 4.19C). Observations in adenomatous lesions were similar to those presented for all polyps in Figure 4.19B, with TLS surface greater in F1 compared to F2 and F3, and a proportion of mature TLS decreasing from F1 to F3 (chi-square test $p = 3.107 \times 10^{-9}$, Figure 4.19D).

To better isolate effects related to the frequency groups from interpathway variability, each pathway was analyzed on its own, avoiding bias due to interpathway variability (which is especially true for iTME analysis through RNAseq data).

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

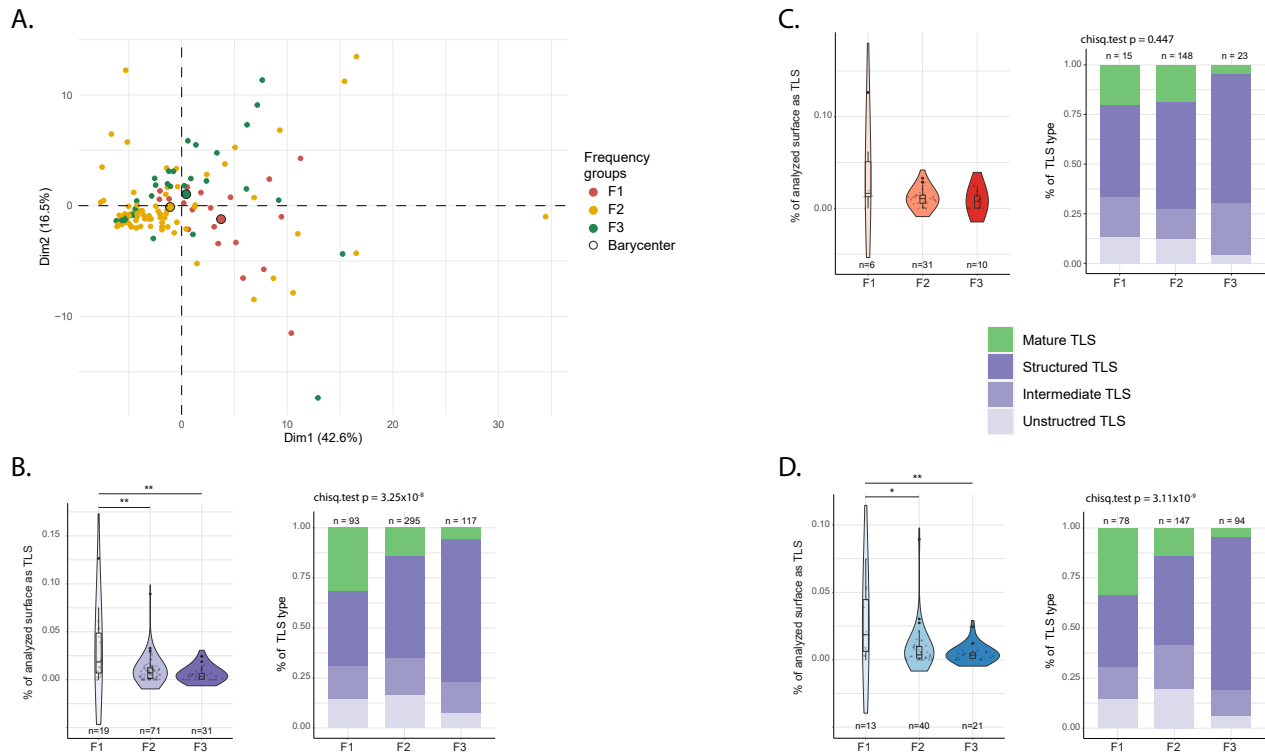


Figure 4.19: **Frequency groups present specific iTME profile and TLS maturation** A. PCA analysis and representation of samples per frequency group based on their IHC phenotype densities in CT, AE and IM. B. TLS profiles between frequency groups C. TLS profiles between frequency groups for serrated lesions. D. TLS profiles between frequency groups for adenomatous lesions. Mann Whitney U test, * $p < 0.05$; ** $p < 0.01$

4.4.3 Polyp frequency and iTME in the serrated pathway

The impact of polyp frequency on the CT iTME of serrated lesions was first assessed and presented in Figure 4.20. UHC analysis highlighted that the serrated F1 group presented a quite homogeneous profile compared to F2 and F3, which both present highly and poorly infiltrated samples (Figure 4.20A). The F1 samples showed a high density of PDL1 expressing cells and epithelial cells, as well as medium T cell T and B cell infiltration, while F2 and F3 mostly revealed low levels of PDL1⁺ cells and very heterogeneous levels of lymphocyte infiltration.

Statistical analysis validated a higher density of PDL1⁺ cells and epithelial cells in the CT of F1 serrated samples when compared to F2 and F3 groups (Figure 4.20B). This observation was also true for subsets of Ki67⁺ or PDL1⁺ epithelial cells. Compared to F2 lesions, F1 samples also presented higher levels of T cells, CD8⁺ T cells, proliferating T cells, exhausted T cells precursors, and CD8 T cells subsets expressing TIM3, Ki67 or PD1 and TIM3. While the same trend was observed between F1 and F3 lesions, CD8 T cells densities were significantly higher in F1. Higher levels of plasma cells were also found in F1 and F2 samples when compared to F3.

The same analysis was performed on IM ROIs, and when compared to F2, F1 lesions were associated with higher levels for all quantified phenotypes except epithelial cells, plasma cells, CD4⁺

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

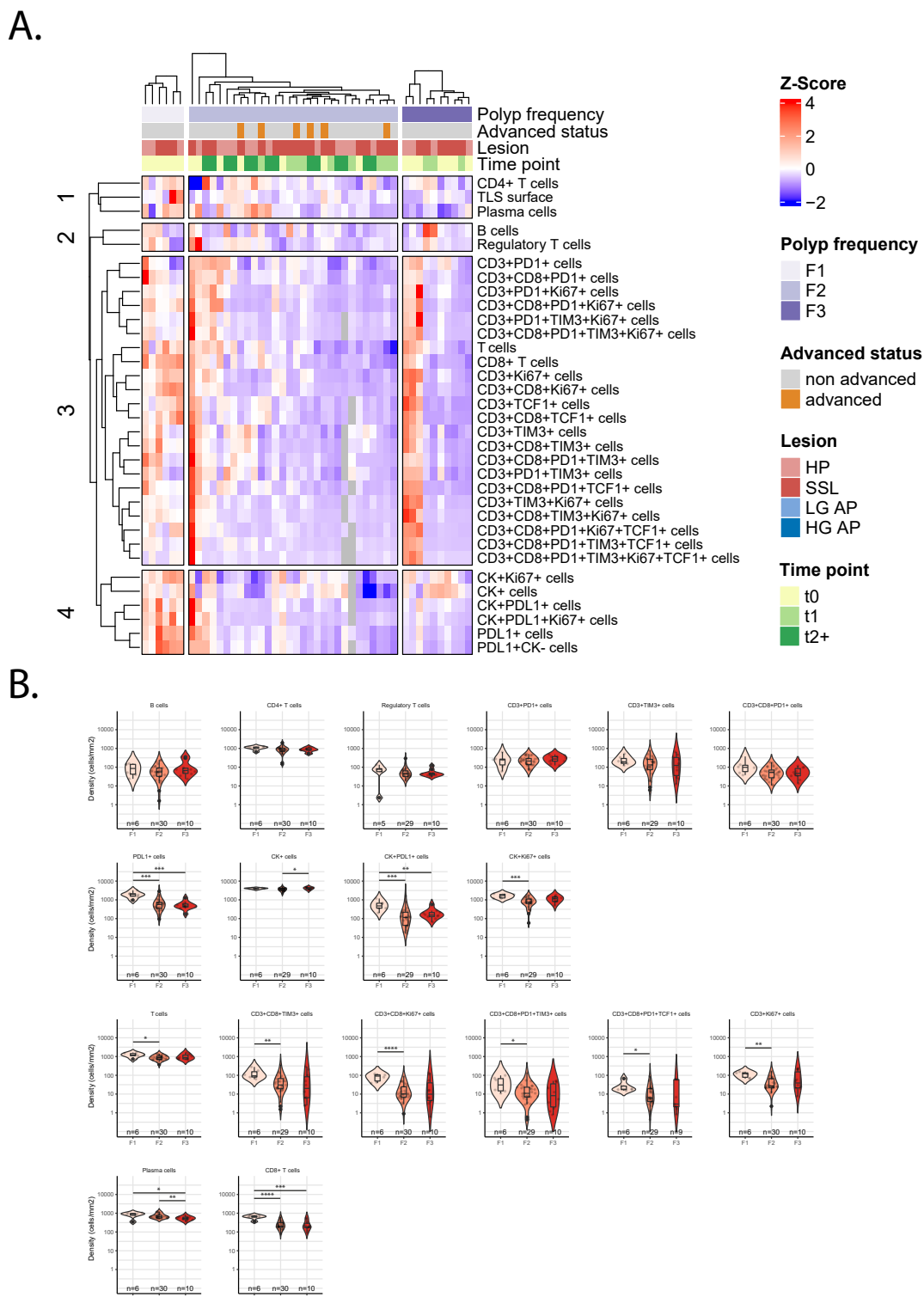


Figure 4.20: **F1 serrated polyps present a more infiltrated iTME at the CT** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in CT and serrated sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Densities of IHC phenotypes in the CT of serrated samples according to frequency groups. Mann Whitney U test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

cells and T_{reg} . (Figure 4.21). Once again, the same trend was observed between F1 and F3 but no significance was found for B cells, $CD8^+TIM3^+$ cells and $CD8^+Ki67^+$ cells.

In conclusion, in the serrated pathway both CT and IM of F1 lesions presented higher or equal densities for most phenotypes when compared to F2 and F3 groups, suggesting an overall more developed iTME. F1 lesions presented particularly high levels of PDL1 expressing cells as well as T cells and $CD8^+$ T cells and their associated exhausted phenotypes. $CD4^+$ T cells and T_{reg} did not appear to be affected by frequency groups.

We then analyzed the Consensus^{TME} scores of serrated lesions between frequency groups, to which we added *CXCL13* gene expression and two TLS signatures: a 12-chemokine signature associated with TLS presence in CRC (*CCL2*, *CCL3*, *CCL4*, *CCL5*, *CCL8*, *CCL18*, *CCL19*, *CCL21*, *CXCL9*, *CXCL10*, *CXCL11*, and *CXCL13*)[214], and a signature derived from a compendium of TLS-hallmark genes (*CCL19*, *CCL21*, *CXCL13*, *CCR7*, *CXCR5*, *SELL* and *LAMP3*) [17, 215, 214]. Expression of genes associated with T cell exhaustion (*CD274*, *PDCD1*, *CTLA4*, *HAVCR2*, *LAG3*, *TOX* and *TCF7*) was also integrated to the analysis presented in Figure 4.22. Of note, *CD274*, *PDCD1*, *HAVCR2* and *TCF7* are coding for the proteins PDL1, PD1, TIM and TCF1, respectively.

UHC analysis and plotting revealed that the two F1 serrated samples presented a relatively low score for all Consensus^{TME} scores, signatures and exhaustion related genes excepted *CTLA4*, *CD274* and *LAG3*. Once again, all scores and signatures appeared to vary similarly, with most samples being either enriched for all or for none, defining hot (enriched) and cold samples. F2 samples presented an heterogeneous profile, with approximately a third of hot samples, a third of intermediate samples and a third of cold samples. On the other hand, F3 samples appeared to have an overall lower score for the inspected parameters, except *PDCD1* and *HAVCR2*

Unfortunately, for statistical analysis, the low number of sequenced serrated samples from the F1 group compels us to focus on the differences between F2 and F3 groups only. For lymphocyte Consensus^{TME} scores, no statistical difference was observed between the two groups. For myeloid scores however, F2 samples presented higher levels of neutrophils, macrophages, M1, and monocytes when compared to F3 lesions. Finally, for TLS signatures and gene expression levels, only *TCF7* presented a statistical difference, being expressed at higher levels in serrated samples from F2 patients.

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

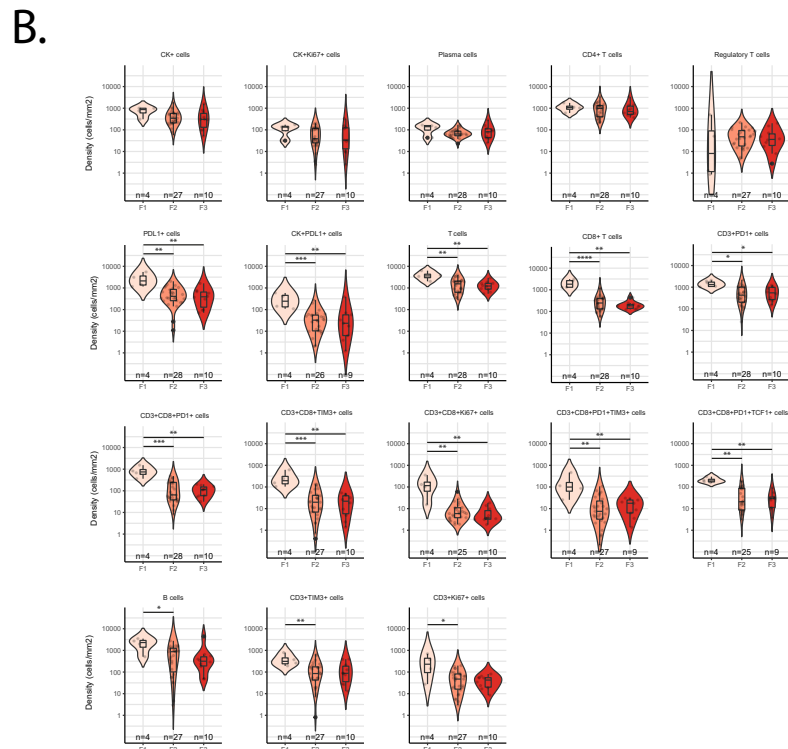
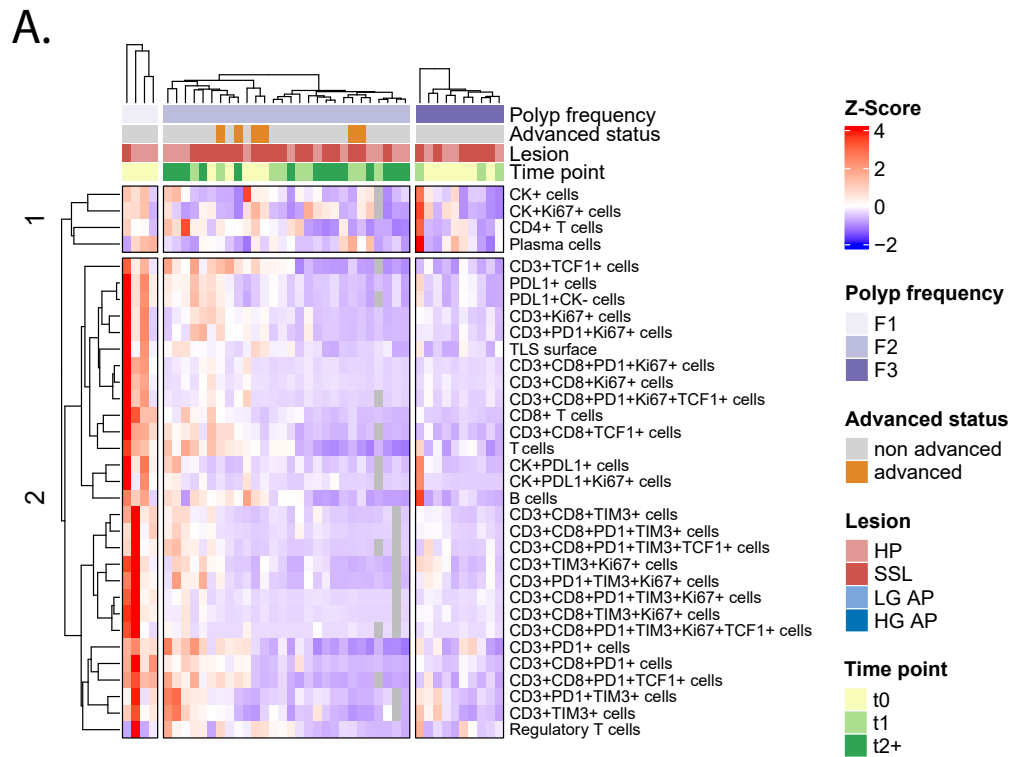


Figure 4.21: **F1 serrated polyps present a more infiltrated iTME at the IM** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in IM and serrated sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Densities of IHC phenotypes in the IM of serrated samples according to frequency groups. Mann Whitney U test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

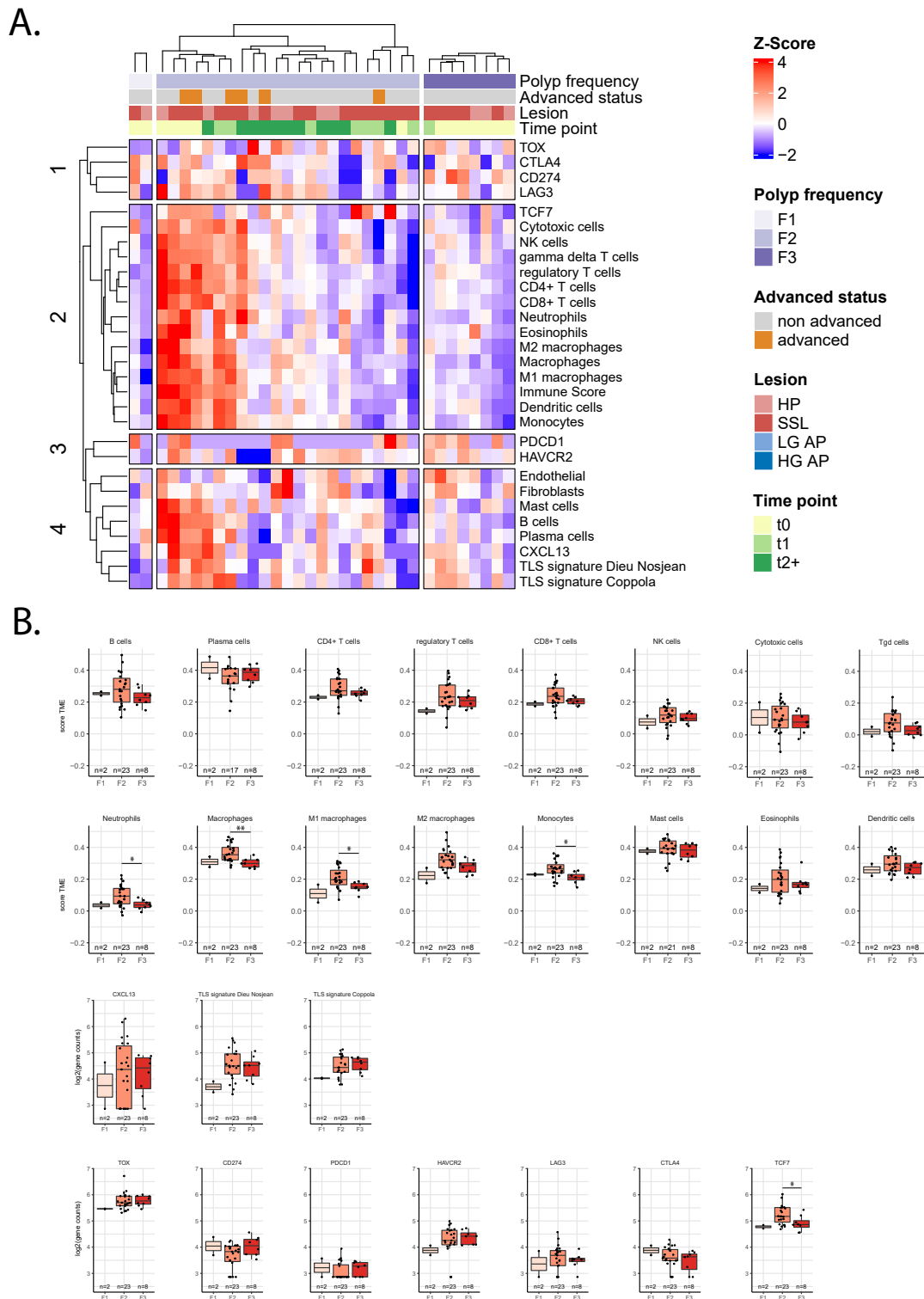


Figure 4.22: **F1 serrated polyps present higher levels of immune signatures** A. Heatmap representing two-dimensional UHC by Consensus^{TME} score and immune signatures, and serrated sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Consensus^{TME} score and immune signatures in serrated samples according to frequency groups. Mann Whitney U test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$

4.4.4 Polyp frequency and iTME in the adenomatous pathway

UHC analysis of adenomatous CT per frequency groups suggested high levels of PDL1 expressing cells (epithelial or not) in F1 samples when compared to F2 and F3. All 3 groups presented heterogeneous levels of other immune cell phenotype, with the F2 frequency group appearing to be less infiltrated in general except four samples that appeared particularly rich in immune cells. Of note, three of these four samples are coming from the only F2 patient who do not present any recurrent event (4.23A).

Statistical analysis demonstrated that the CT of F1 samples presented higher levels of PDL1⁺ cells, PDL1⁺ epithelial cells and proliferating epithelial cells. Concerning immune cells, while the main immune cell types do not appear to vary following the frequency grouping, F2 samples presented lower levels of CD3⁺Ki67⁺, CD3⁺CD8⁺Ki67⁺, CD3⁺CD8⁺TIM3⁺, CD3⁺CD8⁺PD1⁺TIM3⁺ cells than both F1 and F3, and lower densities of CD3⁺CD8⁺TIM3⁺ and CD3⁺CD8⁺PD1⁺TCF1⁺ cells than F3 samples (4.23B).

In conclusion, CT of adenomatous F1 samples seem to present particularly high PDL1 levels while F2 samples present low densities of exhausted phenotypes (T cells subsets expressing TIM3), T exhausted precursors, and proliferating T cells

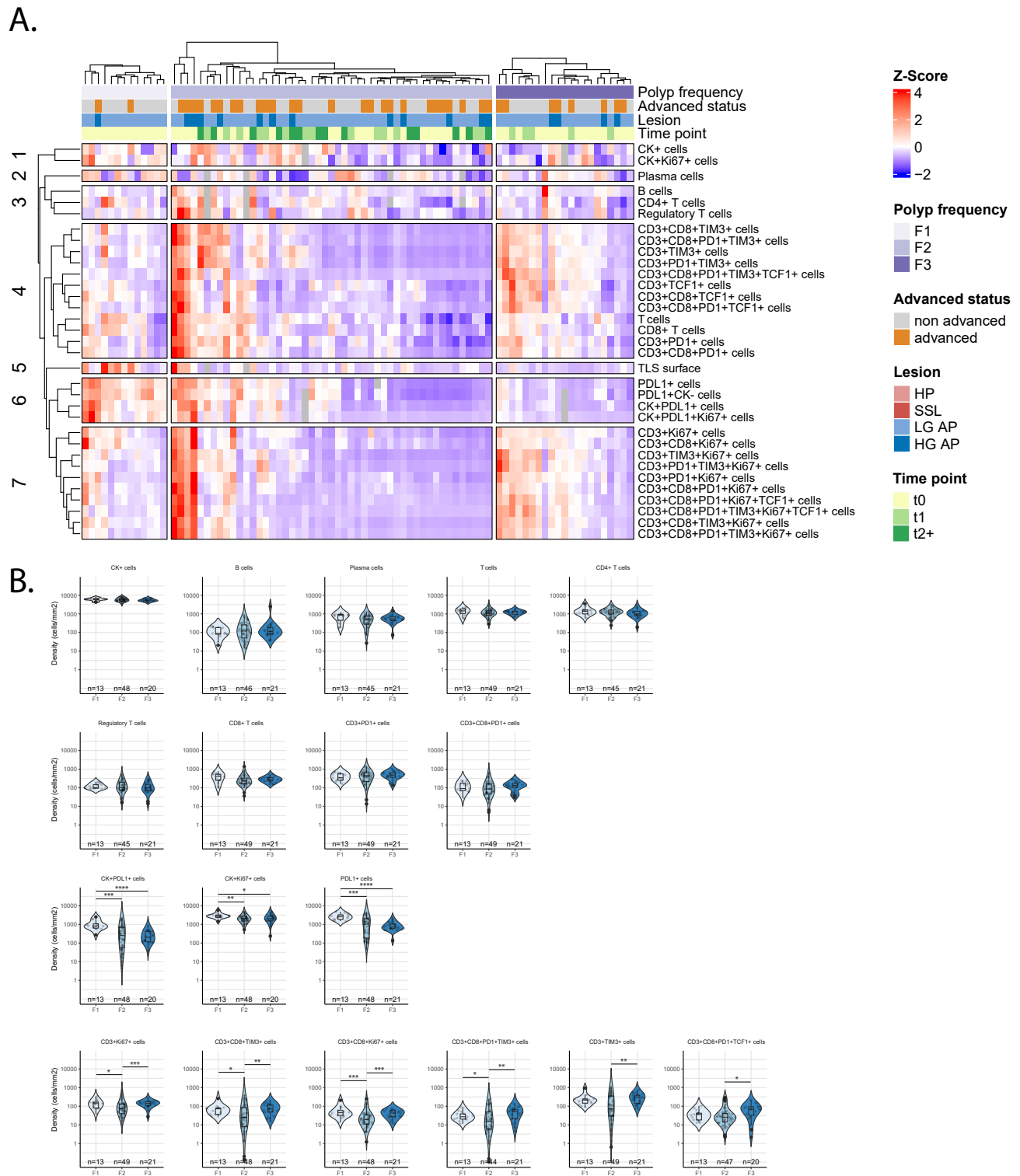
In adenomatous IM, heatmap representation of the UHC analysis exhibited a F1 group presenting high densities for all phenotypes excepted plasma cells when compared to F2 and F3. Interestingly, CD4⁺ phenotypes, TIM3⁺ T cells, Ki67⁺ T cells and epithelial cells are clustering on their own (row cluster 1, 2, 4 and 7), while PDL1⁺ cells, TLS surface, B cells, T cells and CD3⁺PD1⁺ cells cluster together (4.24A).

Phenotype based analysis revealed that epithelial cells, plasma cells, CD3⁺PD1⁺ and CD3⁺CD8⁺PD1⁺TCF1⁺ cells densities in IM did not fluctuate between frequency groups. The densities of T cells and CD8⁺ T cells were higher in F1 compared to F2 and F3 as well as in F2 compared to F3. IM of F1 group also presented higher densities of PDL1⁺ cells, B cells, proliferating T cells and CD8⁺ T cells than F2 and F3 IM. F3 lesions were shown to have lower densities of CD4⁺ cells than F1 and F2 groups. F1 margins also had higher densities of T_{reg} and CD3⁺CD8⁺PD1⁺ cells when compared to F2 and F3, respectively. Finally, TIM3 expressing T cells were particularly low in the IM of F2 lesions: CD3⁺TIM3⁺, CD3⁺CD8⁺TIM3⁺ and CD3⁺CD8⁺PD1⁺TIM3⁺ phenotypes were present at higher densities in both F1 and F3 when compared to F2. CD3⁺CD8⁺TIM3⁺ cells were also found at higher level in F1 IM when compared to F3 (4.24B).

In conclusion, similar to CT, the IM of adenomatous F1 lesions is associated with high density of PDL1 expressing cells, and F2 lesions were shown to have low levels of exhausted T cells. However, the IM of adenomatous F1 also presented higher densities for most immune cell types, highlighting a generally more dense and proliferating iTME that matches their TLS profiles.

For the analysis of Consensus^{TME} scores, TLS signature, and exhaustion-related genes, after UHC analysis and plotting, it was noted that TLS signatures and *CD274* expression presented

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS



4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

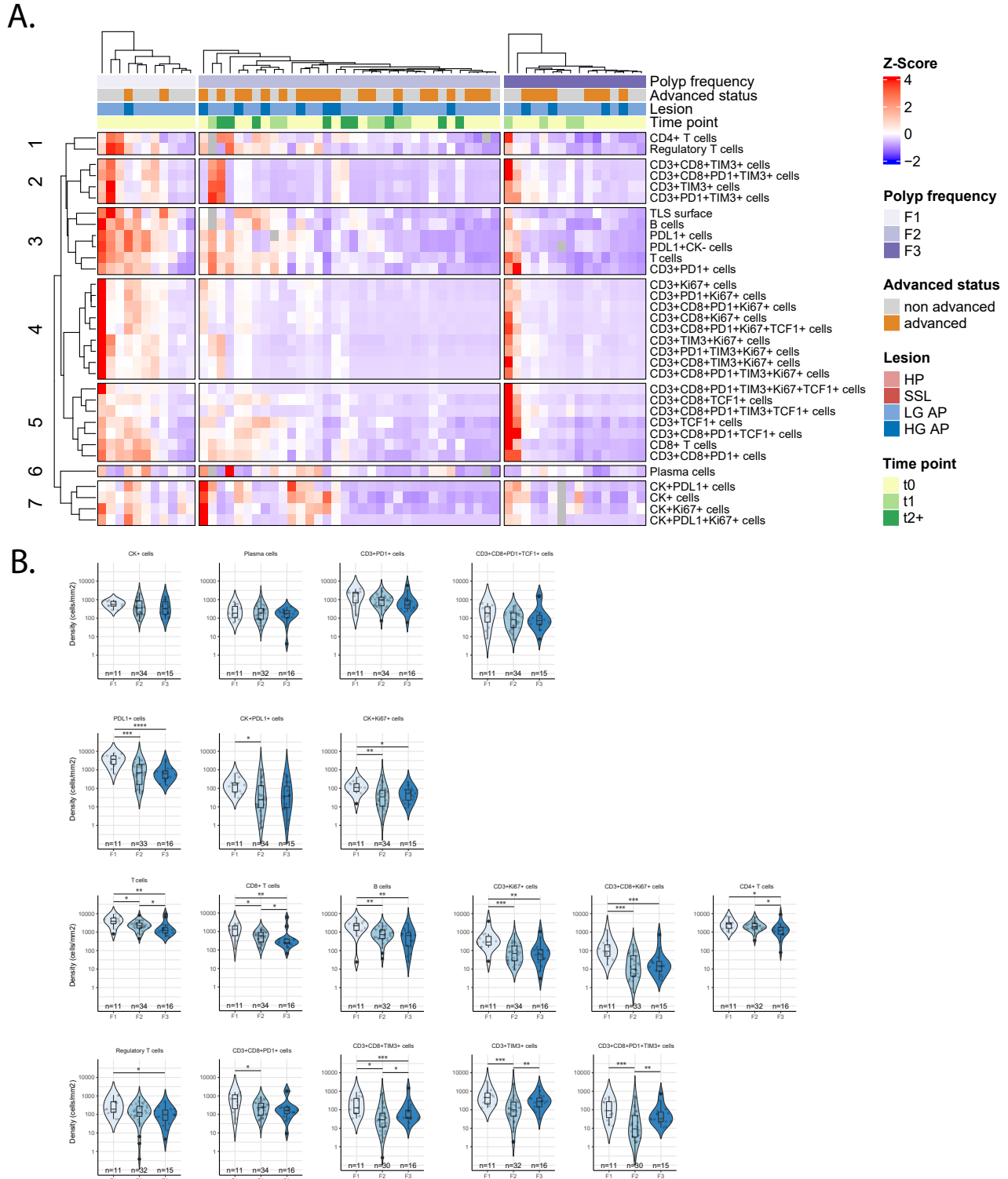


Figure 4.24: **F1 adenomatous polyps present a more infiltrated iTME at the IM** A. Heatmap representing two-dimensional UHC by IHC phenotype densities in IM and adenomatous sample. Row splitting is based on the UHC dendrogram, column splitting is based on frequency groups. B. Densities of IHC phenotypes in the IM of adenomatous samples according to frequency groups. Mann Whitney U test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF TIME OF PREMALIGNANT LESIONS

the same profile as Consensus^{TME} populations: F1 adenomatous samples had high score for all populations, while F2 and F3 samples were more heterogeneous and a fraction of F3 samples present the lowest levels for most populations. Remaining genes associated with exhaustion presented more heterogeneous profiles and seemed to score higher in F2 and F3 groups compared to F1 Figure (4.25A).

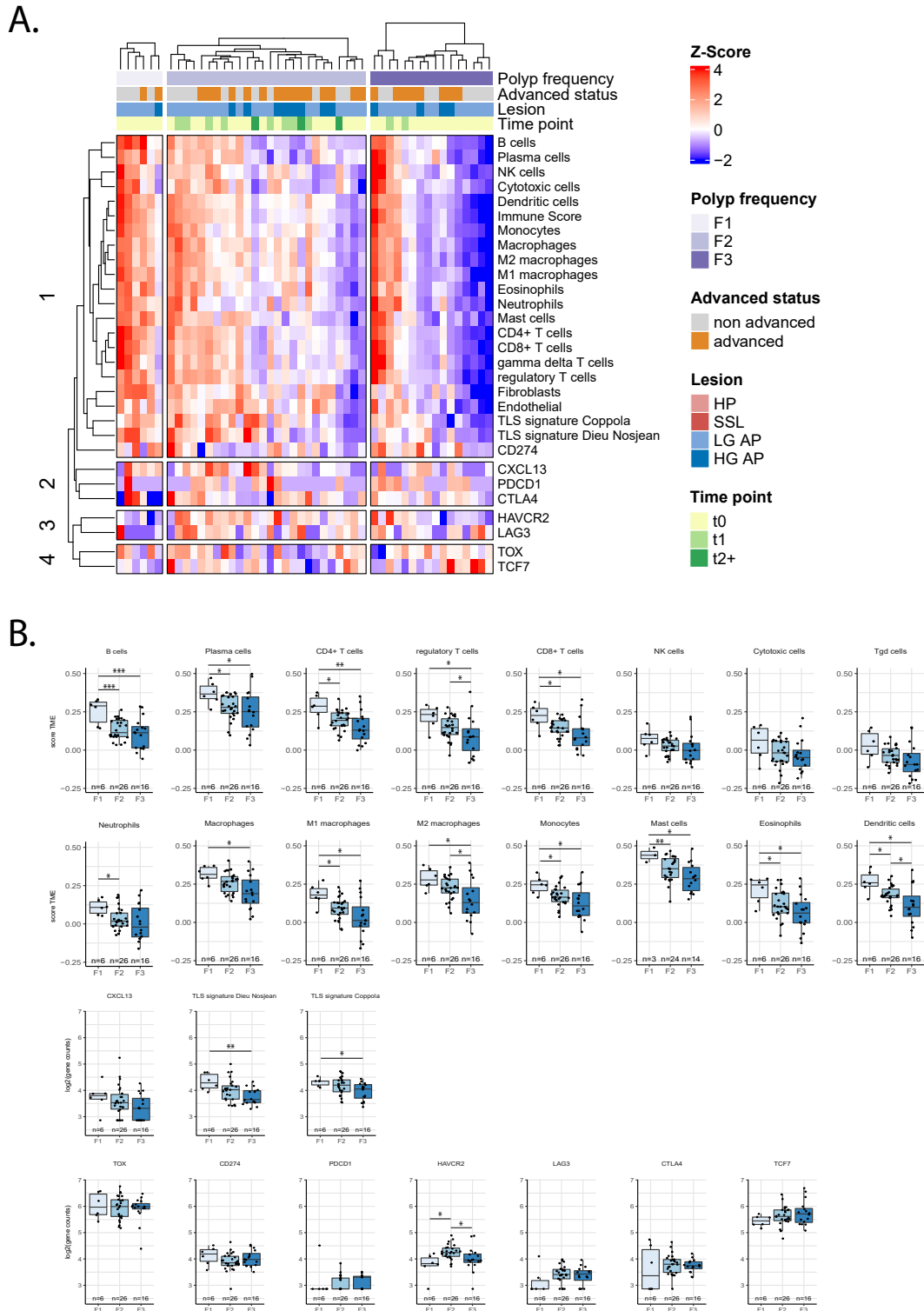
Plotting each cell type separately highlighted a trend for a abundance score decreasing from F1 to F2 to F3 (Figure 4.25B) for all of them. Statistical analysis showed that adenomatous F1 samples presented higher scores for B cells, plasma cells, CD4⁺ T cells and CD8⁺ T cells when compared to F2 and F3. Score for T_{reg} was lower in F3 samples relatively to F1 and F2.

For myeloid populations, neutrophil score was increased in F1 versus F2, macrophages had a higher score in F1 compared to F3. Scores for M1, monocytes, mast cells and eosinophils were higher in F1 F2 in comparison to both F2 and F3. M2 score was significantly lower when compared to F1 and F2. The score of dendritic cells decreased significantly from F1 to F2 to F3.

A trend for both TLS signatures and *CXCL13* expression decreasing while the polyp frequency increased was observed, and TLS signatures were statistically higher in F1 compared to F3.

Finally, excepted for *HAVRC2*, no clear tendency came out of the analysis of exhaustion-related gene expression. *HAVRC2*, which code for TIM3 was observed to be higher in F2 samples when compared to F1 and F3.

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS



4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

A summary of observations performed when comparing frequency groups within adenomatous and serrated pathways is presented in Table 4.2

Table 4.2: Summary of the characterization of the iTME between polyp frequency groups

	Serrated pathway	Adenomatous pathway
TLS	Proportion of mature TLSs decrease in F3 vs. F1 and F2	Surface occupied by TLSs and proportion of mature TLS decrease from F1 to F2 to F3
CT (IHC)	Higher levels of PDL1 ⁺ cells, T cells, proliferating T cells, CD8 T cells, CD3 ⁺ CD8 ⁺ TIM3 ⁺ cells and precursor of exhausted T cells in F1. Less plasma cells in F3	More PDL1 ⁺ cells in F1. Less proliferating T cells, proliferating CD8 ⁺ T cells, exhausted T cells and precursor of exhausted T cells in F2
IM (IHC)	Most phenotypes densities except CD4 ⁺ T cells, T _{reg} and plasma cells are higher in F1 vs. F2 + F3	Most phenotypes densities except plasma cells and PD1 ⁺ subsets are higher in F1 vs. F2 + F3 T cells, CD8 T, and CD4 ⁺ T cells densities are higher in F2 vs. F3. More TIM3 ⁺ phenotypes in F3 vs. F2
RNAseq	No conclusion for F1 Higher neutrophils, macrophage and monocyte scores in F2 vs F3 Higher TCF7 expression in F2 vs F3	All immune cell scores and TLS signatures seem to decrease from F1 to F2 to F3. TIM3 expression is higher in F2
	F1 samples present an higher density for most immune cells and PDL1 ⁺ cells F1 immune infiltration is strong in both CT and IM	F1 samples present an higher density for most immune cells and PDL1 ⁺ cells F1 immune infiltration is strong in IM and seem TLS related

4.4.5 Influence of time and recurrence on the iTME in the adenomatous pathway

By looking at the heatmaps of the previous section, it can be noted that frequency groups present very different distribution of time points: as they are non-recurrent, F1 patients only present t0 samples, while F3 patients have t0 and t1 samples, and samples from F2 are dominated by t1, t2 and more (t2+) samples. It could then be hypothesized that the differences observed between frequency groups are mainly due to time point discrepancies, implying that t0 samples from F1 F2 and F3 groups would present similar iTME profiles, while samples within a frequency group would differ according to their time point.

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

In order to decipher whether the variation of iTME associated with frequency group were due to their time point profile, we analyzed immune infiltration profiles between t0 samples of different frequency groups and then compared immune infiltration between t0 and recurrent (t1+) samples within frequency groups. As serrated samples presented low number of F1 t0 (n=6), F1 t0 (n=6) and F2 t1+ (n=3), we decided to focus on the adenomatous pathway for this analysis.

In adenomatous CT, similarly to the analysis that included all time points (Figure 4.23), no significant differences were observed between frequency groups for epithelial cells, B cells, plasma cells, CD4⁺ cells, T_{reg}, CD8⁺ cells, CD3⁺Ki67⁺ cells and CD3⁺CD8⁺Ki67⁺ cells. However, while the previous analysis found statistical differences for CD3⁺Ki67⁺ cells, CD3⁺CD8⁺Ki67⁺ cells, CD3⁺CD8⁺PD1⁺TIM3⁺ cells and CD3⁺CD8⁺PD1⁺TCF1⁺ cells, none was found when looking at t0 only. However, the noticeable trends remained the same between the two analysis. Finally, similar to the results for all time points, we observed a higher density for PDL1⁺ cells, PDL1⁺ epithelial cells and Ki67⁺ epithelial cells in F1 lesions compared to F2 and F3, while F1 and F3 samples presented higher densities of CD3⁺TIM3⁺ cells and CD3⁺CD8⁺TIM3⁺ cells (Figure 4.26A).

In adenomatous IM, phenotype analysis revealed that epithelial cells, plasma cells, CD3⁺PD1⁺ and CD3⁺CD8⁺PD1⁺TCF1⁺ cells densities in IM did not fluctuate between groups. Densities were higher in F1 compared to F2 and F3 for PDL1⁺ cells, T cells, CD8⁺ T cells, B cells, proliferating CD8⁺ T cells and CD3⁺CD8⁺TIM3⁺ cells. F1 margins also had higher densities of proliferating T cells and CD3⁺CD8⁺PD1⁺ cells T_{reg} when compared to F2 lesions. F3 lesions were shown to have lower densities of CD4⁺ cells than both F1 and F2 groups and of T_{reg} when compared to F1 only. Finally, TIM3 expressing T cells were particularly low in the IM of F2 lesions: CD3⁺TIM3⁺ cells and CD3⁺CD8⁺PD1⁺TIM3⁺ phenotypes were present at higher densities in both F1 and F3 when compared to F2 (Figure 4.26B).

Most of these observations were also true for the analysis that included all time points (Figure 4.24, suggesting a minor impact of time points on the differences observed between frequency groups.

Finally, for F2 and F3 frequency groups, we compared the immune densities between t0 and t1+ in both CT and IM of adenomatous lesions. Overall, cell densities appeared to be very similar between time point of both F1 and F2. In CT, the t1+ samples of the F2 group presented higher densities of epithelial cells, and lower densities of T_{reg} and plasma cells than their t0 counterparts (Figure 4.27A). As presented in Figure 4.27B, no differences were observed between t0 and t1+ within the IM compartment of adenomatous lesions of F2 and F3 groups.

Collectively, these results suggest that the iTME does not vary significantly with time and lesion recurrence, invalidating the hypothesis that the differences between frequency groups are due to the time point parameter.

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

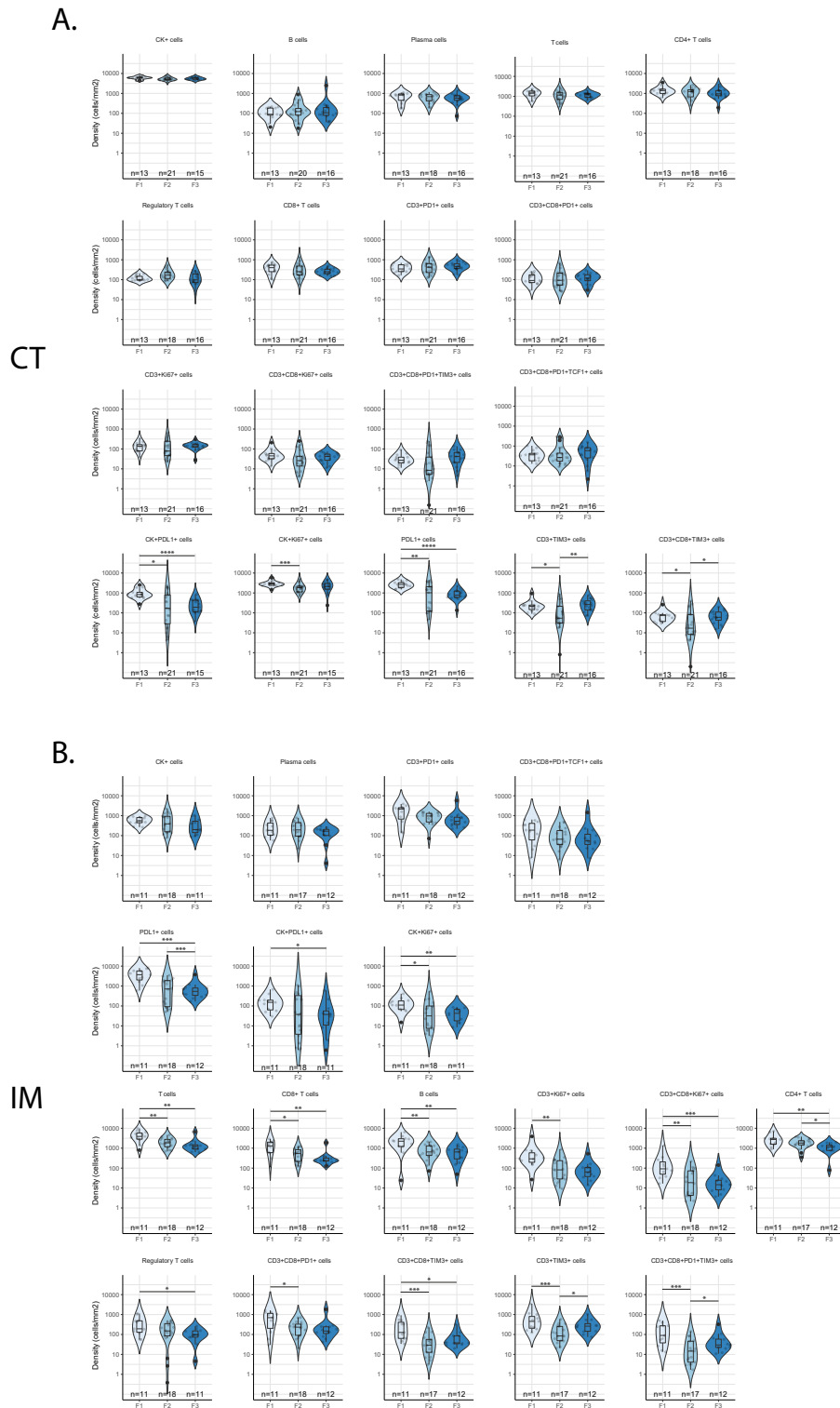


Figure 4.26: Variation of the iTME of adenomatous lesions regarding frequency group are independent of time points A. Densities of IHC phenotypes in the CT of t0 adenomatous samples according to frequency groups. B. Densities of IHC phenotypes in the IM of t0 adenomatous samples according to frequency groups. Mann Whitney U test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$

4.4. POLYP DEVELOPMENT RATE IS ASSOCIATED WITH VARIATION OF ITME OF PREMALIGNANT LESIONS

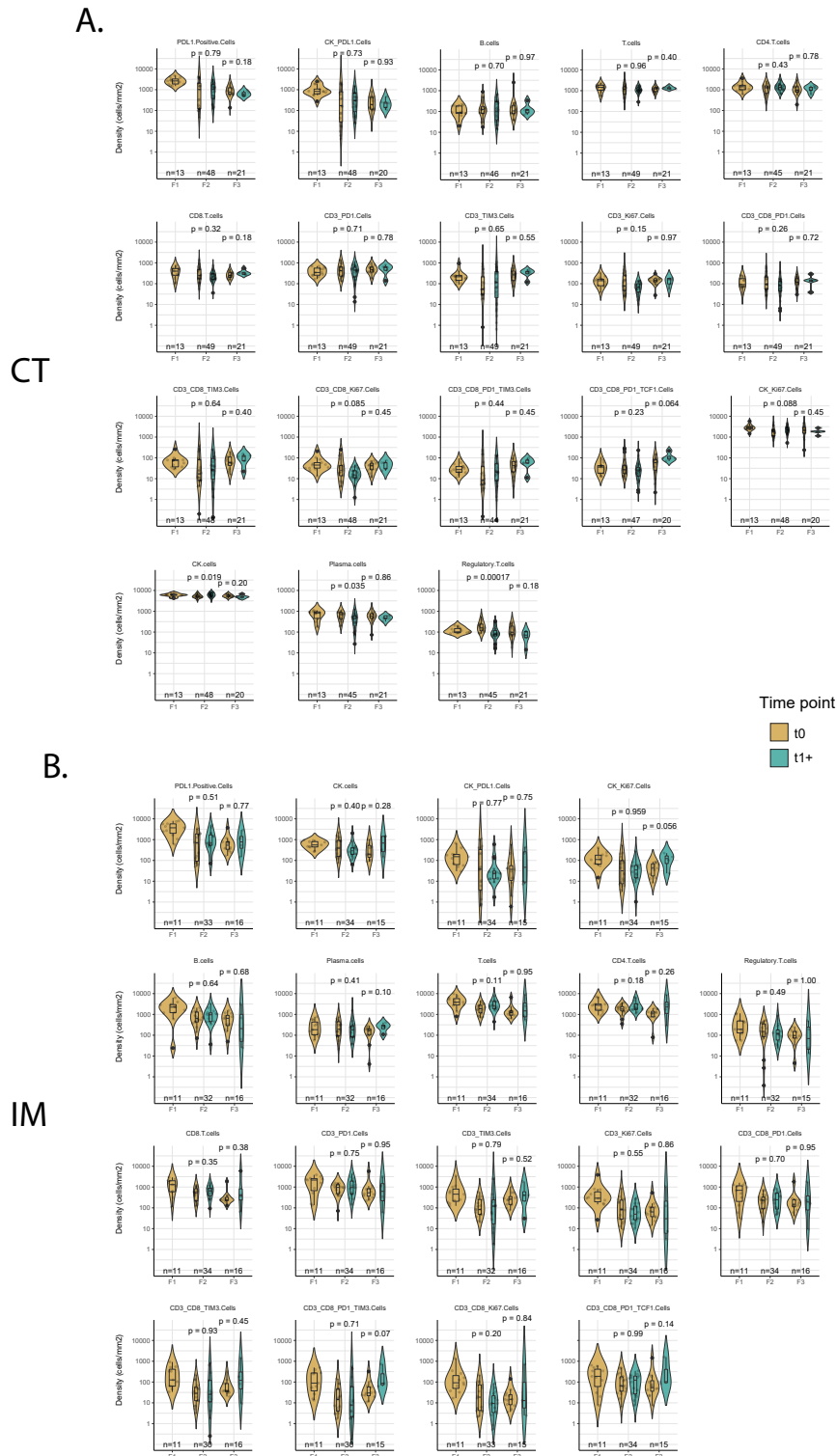


Figure 4.27: Variation of the iTME of adenomatous lesions regarding frequency group are independent of time points A. Densities of IHC phenotypes in the CT of adenomatous samples according to frequency groups and time point. B. Densities of IHC phenotypes in the IM of adenomatous samples according to frequency groups and time points. Mann Whitney U test, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$

4.5 Development of a linear mixed model describing variable associated with iTME variations

Until now, in order to isolate the effect of the parameters studied (advanced status, frequency group, time point) on the iTME components, we had to reduce our sample pool, e.g., focusing on serrated or adenomatous lesion only. Furthermore, we did not account for individual-level variability despite having up to 14 samples from a single patient.

Linear mixed effect models enable the evaluation of an entire set of data for subgroup differences without requiring to split the data into subgroups, while allowing the separation of fixed effects from random effects [210].

Fixed effects, or explanatory variables, are variables that are expected to have an effect on the dependent / response variable. In our case, we expect premalignant pathways and frequency groups to affect cell densities in the TME. We can also wonder if the advanced status and time point are indeed negligible to describe iTME. In the model, the term "*" will describe the terms that interact with each other. In our case, the pathway and advanced status are intrinsically associated, and time point profiles are greatly dependent of the frequency group.

Random effects are usually grouping factors for which we are trying to control. In our case, as for most studies with multiple samples from the same patient, the main random effect is the patient's ID. BMI levels (Normal, overweight, obese and severely obese) and sex are other factors that we can try to control. Using the Akaike information criterion (a mathematical method for evaluating how well a model fits the data) to compare models with several combinations of our random effects, it appears that the model including patient ID, BMI level and sex was the most adapted to our data. In the model, randoms effects will be noted as (1|Effect).

This gives us the final model:

$$\text{Cell.Pop} = \text{Pathway*Advanced} + \text{Frequency_group*Time_point} + (1|\text{Patient}) + (1|\text{BMI levels}) + (1|\text{Sexe})$$

Were "Cell.Pop" was replaced by the immune phenotypes of interest. We applied this model to immune densities extracted from IHC for both the CT and IM compartments. The main types of immune cells, as well as CD8⁺ T cells positive for PD1 or TIM3 were tested.

As presented in Table 4.3, we can observe that for each cell type, the model gives back a "Constant" value, which reflect the mean density for all variables in their default settings: Pathway = serrated, Status = non-advanced, Frequency = F1, Time point = t0. The p-value of the constant reflect if its significantly different from 0.

In CT (Table 4.3), we can note than the constants for B cells, T_{reg} and CD3⁺TIM3⁺PD1⁺ are not significantly different from 0, reflecting a low density of these populations. Similarly to what was observed in 4.3, the model predicted higher densities of B cells, T cells, CD4⁺ T cells, and T_{reg} in the adenomatous pathway, while CD8⁺ T cells and CD3⁺CD8⁺TIM3⁺ cells are not significantly different. No significant differences were found in our model for PDL1⁺, plasma cells

4.5. DEVELOPMENT OF A LINEAR MIXED MODEL DESCRIBING VARIABLE ASSOCIATED WITH TIME VARIATIONS

and CD3⁺CD8⁺PD1⁺ cells while they were observed earlier, but a trend for increased density was still observed in the adenomatous pathway.

The advanced status and the time point were not found to be determinants of the predicted CT densities of selected immune phenotypes, except plasma cells that appear to increase in t2+ samples.

Finally, for the impact of frequency groups on CT immune infiltrate, predicted densities of PDL1⁺ cells were the only one significantly lower for F2 and F3 parameters.

Table 4.3: Linear mixed modeling of CT cell densities

	Cell densities (mm ²)								
	PDL1 ⁺ cells	B cells	Plasma cells	T cells	CD4 ⁺ T cells	Regulatory T cells	CD8 ⁺ T cells	CD3 ⁺ CD8 ⁺ PD1 ⁺ cells	CD3 ⁺ CD8 ⁺ TIM3 ⁺ cells
Pathway: Adenomatous	233.0 (183.9)	186.2** (59.3)	-17.8 (56.5)	280.8** (102.7)	541.2*** (135.4)	64.9*** (19.3)	-23.6 (39.6)	28.7 (21.0)	-31.6 (18.8)
Status: Advanced	-229.5 (286.1)	11.7 (110.5)	9.8 (114.6)	-165.5 (166.1)	-70.8 (229.8)	2.6 (31.6)	-84.5 (61.6)	-32.6 (33.0)	-35.4 (31.1)
Frequency: F2	-1180.4** (424.5)	102.0 (87.2)	-16.1 (93.0)	-80.1 (195.7)	-223.6 (216.5)	69.6 (36.2)	-88.6 (91.3)	53.8 (49.5)	-6.6 (33.6)
Frequency: F3	-1594.8** (610.7)	135.5 (95.5)	-70.2 (83.3)	-172.4 (267.7)	-250.7 (268.1)	-11.7 (43.4)	-186.7 (131.4)	-42.5 (64.1)	-12.6 (43.2)
Time point: t1	-348.5 (272.7)	-93.9 (104.8)	-86.5 (104.3)	-39.2 (158.6)	-131.0 (219.6)	-43.7 (30.3)	16.8 (58.8)	12.4 (31.5)	8.4 (29.8)
Time point: t2+	210.7 (210.7)	-10.1 (74.4)	-177.7* (76.0)	-46.5 (120.5)	317.3 (164.3)	-30.0 (22.8)	-48.6 (45.4)	-20.8 (24.2)	-12.1 (22.3)
Adenomatous:Advanced	426.4 (309.9)	-110.6 (121.4)	-229.2 (127.4)	-52.0 (180.1)	-88.3 (251.4)	20.5 (34.5)	22.5 (66.7)	42.1 (35.8)	35.9 (33.8)
Frequency:Time.Point	672.9* (338.9)	13.3 (128.0)	-215.4 (129.1)	48.3 (196.0)	153.6 (271.9)	-0.4 (37.6)	-33.9 (73.0)	-4.2 (39.1)	-6.9 (36.5)
Constant	2258.3*** (328.0)	-7.7 (72.3)	793.9*** (96.3)	1205.6*** (155.4)	1028.8*** (176.1)	35.8 (35.2)	495.2*** (70.6)	69.7 (52.4)	103.4** (34.9)
Observations	128	127	125	129	125	127	129	129	128

Note: Constant is the density value for pathway = serrated, status = non-advanced, frequency = F1 and time point = 0.
Below each value is, in brackets, the SD.

*p<0.05; **p<0.01; ***p<0.001

The results of the mixed linear model on IM densities are presented in Table 4.4. It can be noted that only plasma cells present a constant that is not significantly superior to 0. According to the model, the densities of PDL1⁺ cells, plasma cells, T cells, CD4⁺ T cells, T_{reg} and CD3⁺CD8⁺TIM3⁺ cells are significantly higher in the IM of adenomatous samples, which again matches our IHC observations. Similarly to our findings, B cells were found to have similar levels between the two pathways. CD8⁺ T cells and CD3⁺CD8⁺PD1⁺ cells are not significantly increased in adenomatous samples, but still present such a trend.

Finally, almost all predicted densities of IM cells (except plasma cells) were lower for F2 and F3 levels compared to F1.

Overall, results from the linear mixed model strengthen most observations from previous sections,

4.5. DEVELOPMENT OF A LINEAR MIXED MODEL DESCRIBING VARIABLE ASSOCIATED WITH TIME VARIATIONS

Table 4.4: Linear mixed modeling of IM cell densities

	Cell densities (/mm ²)								
	PDL1 ⁺ cells	B cells	Plasma cells	T cells	CD4 ⁺ T cells	Regulatory T cells	CD8 ⁺ T cells	CD3 ⁺ CD8 ⁺ PD1 ⁺ cells	CD3 ⁺ CD8 ⁺ TIM3 ⁺ cells
Adenomatous	893.4** (316.4)	195.8 (243.6)	186.8*** (49.5)	1197.3** (425.9)	1586.5*** (347.2)	126.9** (46.9)	335.9 (185.2)	156.9 (95.6)	102.8* (51.8)
Advanced	-300.1 (436.3)	129.5 (436.1)	-0.9 (88.6)	-212.4 (773.8)	211.4 (622.4)	33.5 (85.0)	-143.8 (332.2)	-64.8 (155.9)	-6.6 (92.2)
Frequency F2	-2972.3*** (718.2)	-1623.4*** (389.5)	-60.0 (78.6)	-2077.2** (638.9)	-841.6 (545.4)	-155.7* (67.4)	-878.2** (267.6)	-434.6** (134.2)	-214.7** (77.1)
Frequency F3	-2152.2* (866.8)	-1490.8*** (352.1)	-74.6 (71.3)	-2406.5*** (606.5)	-1068.3* (498.4)	-228.2*** (65.5)	-1143.2*** (266.5)	-466.6** (144.8)	-199.0** (74.9)
Time point t1	88.7 (465.0)	730.2 (419.9)	92.1 (85.3)	1023.4 (747.8)	1619.8** (599.6)	70.1 (82.2)	731.5* (327.2)	174.9 (161.2)	198.6* (90.5)
Time point t2+	-80.8 (346.7)	-335.1 (315.8)	-14.6 (64.4)	-102.7 (557.4)	445.3 (451.8)	-18.8 (61.0)	-86.2 (240.3)	-17.9 (115.4)	39.8 (67.2)
Adenomatous:Advanced	-165.0 (488.0)	-190.5 (495.2)	-9.6 (100.7)	-447.7 (880.4)	-632.3 (707.8)	-31.6 (96.7)	-122.8 (377.4)	-39.8 (175.9)	-50.1 (104.7)
Frequency:Time.point	378.0 (574.5)	-1083.8* (529.2)	-153.0 (107.6)	-725.8 (930.1)	-1760.7* (756.0)	-107.0 (102.9)	-828.5* (404.7)	-166.2 (195.9)	-175.6 (112.1)
Constant	3416.6*** (718.5)	2381.3*** (384.4)	139.3 (75.9)	3626.4*** (557.7)	1660.7** (511.8)	229.1*** (57.1)	1344.5*** (226.5)	625.6*** (122.1)	224.3*** (67.2)
Observations	102	102	101	103	101	102	103	103	102

Note: Constant is the density value for pathway = serrated, status = non-advanced, frequency = F1 and time point = 0.
Below each value is, in brackets, the SD.

*p<0.05; **p<0.01; ***p<0.001

eliminating individual-level variability.

5 Discussion and Conclusion

5.1 Discussion

To characterize iTME in association with the risk of precancerous lesions, we applied an integrative approach to a cohort that included 26 patients of different risk levels, for a total of 140 lesions. Taking advantage of WES, RNAseq and multiplex IHC assays, we first characterized the cohort from a molecular point of view and then described the iTME observed in premalignant lesions. Variations in iTME were analyzed in the light of inter and intrapathway specificity, and patient polyps frequency in order to assess which parameters were responsible for the heterogeneity of the immune infiltrate within these samples.

This study suggests that the adenomatous polyps presented a more developed immune infiltrate compared to their serrated counterparts. Moreover, this analysis outlined that patients with a low frequency of polyps harbor more TLS, which appear to be more mature, and a more infiltrated iTME when compared to patients developing polyps at a higher rate.

5.1.1 Molecular characterization of the cohort

As reported in section 4.1.2, following mutation calling, we detected BRAF mutations in 92% of serrated lesions and *APC* and *KRAS* mutations in respectively 73% and 37% of adenomatous samples. None of these driver mutations is found in the other pathway. Such a dichotomy was expected [203, 74], even *KRAS* mutations have been described in a small proportion of SSL [72].

More surprisingly, the observation was made that SSLs (associated with hypermutation and MSI) were not enriched in hypermutated phenotype compared to APs and even had a lower TMB. However, this observation was consistent with a study by Chen *et al.*, which found that SSLs without dysplasia have not yet acquired their hypermutated phenotype and that APs carry a higher TMB [74]. Similarly to our results, they also demonstrated that TMB increases from non-advanced to advanced APs.

Three of our samples presented a particularly high TMB: The only MSI cancer sample, the only AP from the Lynch syndrome patient, and the only SSL presenting dysplasia. This last result is consistent with the literature, which found that non-dysplastic SSLs are very rarely hypermutated, and the proportion of MSI SSL increases with the grade of dysplasia [203, 204]. Our CRC samples were dominated by oncogenes associated with the adenomatous pathway, consistent with the idea that the majority of CRC arise from APs, and presented higher levels of oncogenes associated with advanced carcinogenesis when compared to precancerous lesions (*p35* and *PI3KCA*).

5.1. DISCUSSION

Taking advantage of RNAseq data, the CMS profile of our samples was assessed (section 4.1.3). We found CMS1 and CMS4 to be particularly associated with SSLs and HPs, respectively, while APs were dominated by CMS2 and presented 20.8% and 12.5% of CMS3 and 4, respectively. A study analyzing several public datasets found similar results for APs and showed that both SSLs and HPs were dominated by CMS1 subtypes, suggesting that CMS3 and CMS4 appear only later in colorectal carcinogenesis [216]. Chen *et al.* found that APs score higher in CMS2 and SSLs are enriched in CMS1 and CMS3.

Thus, the observed CMS profiles appear to be quite coherent with the literature for APs, while the serrated result appears more heterogeneous. Surprisingly, another study found the CMS3 subtype to represent 73% of APs [217]. Discrepancies in results could partially be explained by the type of sequencing (RNAseq or microarray) and the choice of algorithm used for CMS scoring. However, this important variability probably also probably reflects cohort-specific differences. The high indeterminate rate observed in the adenomatous pathway could be explained by tumor heterogeneity, but it would be surprising to have such a high heterogeneity early in carcinogenesis [95]

Notably, six of our eight CRC were CMS3, and the other two were CMS1 and 2. In the literature, primary CRCs are mainly CMS2 (37%), followed by CMS4(23%), CMS1 (14%) and CSM3 (13%) [94].

5.1.2 iTME characterization

We then proceeded to characterize the iTME of precancerous lesions, mainly using multiplex IHC and Consensus^{TME} scoring. Given the study design, we could not use TLS maturation specific markers such as CD23 or BCL6 [187]. However, Ki67 is still a strong marker of the germinal center, and we observed that its combination with the staining of CD20, CD3, and PD1 makes for a quite potent classification marker (section 4.2) . The characterization of T_{reg} using FOXP3 in combination with CD3 and CD4 also needs to be nuanced, as FOXP3^{low} T_{reg} have been characterized in CRC, presenting an opposite function (proinflammatory) and prognostic value (good) when compared to traditional T_{reg} [218].

When comparing measured immune densities to the literature, we found our data in the range of previous studies: In LG AP, Freitas *et al.* [200] observed densities of 1400, 450, and 220 cells / mm² for T cells, CD8⁺ T cells and T_{reg}, respectively, without discriminating for CT or IM. On our side, we observed median densities of 1180 cells / mm² and 2097 cells / mm² for T cells in CT and IM, 319 cells / mm² and 534 cells / mm² for CD8⁺ T cells and 123 cells / mm² and 198 cells / mm² for T_{reg}.

All types of immune cells, except plasma cells, were found to have higher densities in IM compared to CT and AE. These results are in line with the observation that B cells were more present at the margin of TLS + tumors, while plasma cells, potentially emerging from this TLS, are found within the CT [184].

Interestingly, the study revealed that of 140 lesions, only 7 did not present TLS either at the

5.1. DISCUSSION

lesion or its surrounding, and around 15% of this TLS presented a germinal center. This result is quite surprising when compared to the situation in hepatic premalignant lesions, where only 24% of the lesions presented TLS, which were all immature [133]. Unfortunately, little is known about TLS in colorectal polyps. One potential explanation for this important presence of follicular structure is the presence of strong preexisting immunity in healthy colon, which already present TLS-like structures in GALTs. There is indeed a non-null probability that observed TLS are anterior to the apparition of the lesion [219].

5.1.3 Interpathway characterization of the iTME

As presented in section 4.3.1, when comparing the iTME of serrated and adenomatous polyps using multiplex IHC data, the most striking observation was that the two pathways presented distinct iTME profiles. APs presented an overall more infiltrated CT and IM, while serrated polyps had more surface occupied by TLSs. The only cell type presenting higher density in serrated lesions was plasma cells of the CT compartment, which could be explained by the higher presence of TLS.

It could be expected that serrated lesions, being mainly CMS1 and precursor to MSI CRC, would present a higher immune infiltrate than APs [95]. However, since these lesions do not yet have a hypermutated phenotype, the main parameter responsible for this higher infiltration in CRC (high TMB and immunogenicity) is missing [196].

Few studies compared immune infiltrate between the two pathways. The ones who did found higher TITLs intraepithelial densities when looking at T cells and CD8 T cells [74] or lymphocytes in general, assessed via H&E staining [203]. No differences were found for CD4⁺ T cells, T_h17 and mast cells [198]. When looking at intraepithelial densities between pathways using the CK staining, we also observed similar densities of T cells and higher densities of CD8⁺ T cells in serrated lesions, as presented in Figure 5.1

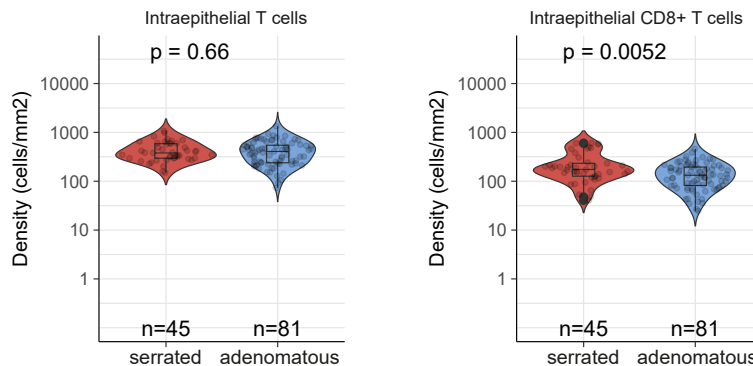


Figure 5.1: Serrated lesions present higher intraepithelial CTLs densities

We also observed more proliferating epithelial cells in APs when compared to serrated lesions, which is coherent regarding the presence of dysplasia in AP only, and the fact that APs epithelial cells have been shown to present more stem cell signature when compared to their serrated counter-

5.1. DISCUSSION

parts.

Surprisingly, the estimation of the iTME based on RNAseq data provided the opposite result, suggesting a more developed iTME in serrated lesions. The opposite results between IHC and RNAseq were explained by the fact that the Consensus^{TME} provides a relative abundance score. We showed that, due to the presence of dysplasia in APs only, these lesions have a higher overall cell density, which could explain the discrepancies of the results between IHC and RNAseq. Other deconvolution techniques, such as the Cybersort algorithm, also present an absolute score, and should be tested to verify this hypothesis [220].

5.1.4 Intrapathway characterization of the iTME

Working within APs or serrated polyps only (section 4.3.2, it was observed that iTME presented a very homogeneous profile within each pathway. Although SSLs have premalignant potential, they still appear to have similar iTME compared to HPs: the only difference was a higher density of exhausted T cell precursor in the CT of HPs. The fact that HP are still a heterogeneous category and that microvesicular HPs are suspected to be precursors to HPs could explain this surprising observation [72, 74].

Surprisingly, advanced SSLs also revealed lower levels of TIM3⁺ T lymphocytes than non-advanced ones, suggesting lower levels of T cell exhaustion. These results are not in line with those of Acosta-Gonzalez *et al.*, who found that as SSLs progress, TITLs and PD1 / PDL1 level increase [204]. However, these comparisons were made between SSLs presenting various levels of dysplasia, while we only relied on the size of the lesion to classify SSLs as advanced. Moreover, for all interpretations within the serrated pathway, the results must be nuanced in light of the sample size. For example, we only have 6 advanced SSLs.

Within the adenomatous pathway, we found advanced APs to present lower levels of plasma cells and higher levels of T_{reg} when compared to non-advanced. All other phenotypes presented similar densities, in line with results obtained by Becker *et al.* (preprint), who demonstrated that while T_{reg} levels increased with polyp progression, exhaustion levels only increased during invasive stages [134]. However, these results are not in line with the study of Freitas *et al.*, that presented a strong decrease in T cells, CD4⁺ T cells and CD8⁺ T cells from LG AP to HG AP [200]. Rubio *et al.*, on the other hand, found higher levels of intraepithelials TILs in HG AP when compared to LG AP [197]

5.1.5 Frequency groups

In order to study the iTME in patient presenting various risk of developing polyps, we created a risk variable based on the frequency of polyps developed per year in each patient. It is important to note that the F1 group is representative of the majority of average-risk patients, and have been

5.1. DISCUSSION

selected as such on purpose. Studies on the general population showed that 72% of the patients were treated for 3 polyps less for their first colonoscopy [77], and that patients presenting high (more than 3) number of polyps also had higher chances of recurrence [78, 221, 80]. F1 patient presented one to three polyp, and no recurrence event. As such, they can be considered normal or low-risk patients. F2 patients have a higher number of polyp per year and present recurrent event: they can be seen as high-risk patients (relatively to the average-risk patient pool). The only non-recurrent patient in F2 present a very short follow-up period and a cancer. Finally, F3 patients present a high number of polyps despite a short follow-up period, making them at even higher risk. However, they could regress to F2 after a longer follow-up period.

Based on the differences between colorectal premalignant pathways observed earlier, we focused our frequency analysis in each pathway separately, losing sample size and statistical power but avoiding mixing effects. As lesion type, grade and advanced status did not appear to significantly affect the iTME, we focused on each pathway as a whole. Both pathways were present in a similar proportion between risk groups (Section 4.4.1) and we showed that the risk level is not associated with a particular premalignant pathway, nor with a specific mutational or CMS profile.

The most striking observation that emerged from this analysis was that in both pathways, a higher density of the majority of immune cells types and PDL1⁺ cells was observed in F1 samples when compared to F2 and F3. This observation suggests that low-frequency patients present a more developed local immunity, which could explain why these patients develop polyps at a lower rate than F2 and F3 patients. Higher levels of exhaustion markers such as PDL1, PD1, TIM3 and exhausted T cell precursors also suggest a more exhausted iTME.

More specifically, in the serrated pathway (section 4.4.3), a trend for a diminution of the proportion of mature TLS was observed in F3 lesions compared to F1 and F2.

The CT of serrated polyps was overall more infiltrated in F1 lesions when compared to F2 and F3 lesions, which presented similar iTME profiles. Among the main types of immune cells, only B cells, CD4⁺ T cells, and T_{reg} were not higher in F1. Mirroring the low frequency of mature TLS in F3 samples, plasma cells were only lower in F3 while they had similar levels between F1 and F2.

Similar observations were done in IM, were all phenotypes except plasma cells, CD4⁺ T cells, and T_{reg} were increased in F1 compared to F2 and F3. These results highlight the segregation of B cells to IM and plasma cells to CT.

RNaseq analysis demonstrated higher myeloid infiltration in F3 samples compared to F2, but no results can be attributed to F1 as only two serrated samples were available. Although the F2 and F3 groups seem equivalent with respect to most immune cell types, TLSs, plasma cells, and myeloid cells might differentiate these two frequency groups.

In adenomatous samples (section 4.4.4), TLSs were shown to vary even more between frequency

5.1. DISCUSSION

groups, as the TLS surface and TLS maturation levels gradually decreased from F1 to F2 to F3. Surprisingly, no difference of plasma cell density was observed between the frequency groups.

In CT, the only cell populations present at higher density in F1 compared to F2 and F3 were PDL1⁺ cells, PDL1⁺ epithelial cells and proliferating epithelial cells. On the other hand, in IM, the majority of immune cell types except plasma cells were found to have a higher density in F1.

Unlike serrated samples, adenomatous F2 and F3 also present distinct profiles: F3 samples present higher levels of T cells subsets expressing TIM3 in both CT and IM, while F2 samples harbor higher T cells, CD8⁺ T cells and CD4⁺ T cells in IM.

Finally, the RNAseq-based analysis of iTME presented scores gradually decreasing from F1 to F2 to F3 for most immune cell types, lymphoid or myeloid. TLS signatures were also higher in F1 compared to T3. Surprisingly, the only gene marker of exhaustion that showed significant differential expression was *HAVCR2*, coding for TIM3, which was expressed at a higher level in F2 compared to F1 and F3. As numerous cell types (including DCs, macrophages, MDSCs and NK cells) express TIM3, this could explain the disparities between IHC and RNAseq results [222].

Overall, while F1 samples of both pathway present higher immune densities, the differences within the serrated iTME appear more diffuse, appearing both in CT and IM. On the other hand, variations of the iTME from F1 to F2 and F3 in adenomatous samples seem more IM centered, with higher TLS density and maturation along with higher immune densities in IM. F2 and F3 serrated samples present quite similar profile, while adenomatous F2 samples appear to be more infiltrated and less exhausted than F3.

For several observations, only the differences between F1 and F2 were significant, while F2 and F3 presented the same profile. This is probably due to the sample size, since the F2 group represents 59% of total samples. Discrepancies between RNAseq and IHC could be associated to the fact that not all samples are present in RNAseq.

In order to validate these findings, we first demonstrated that these results were independent of time point disparities between frequency groups (section 4.4.5): in the adenomatous pathway, taking only t0 samples from each group, we observed the same trends as previously presented. The loss of significance for some observations can mainly be attributed to decreased sample size. This observation was validated by the fact that very similar levels of immune densities were observed between the time points t0 and t1 + in the frequency groups of the adenomatous pathway.

5.1.6 Modeling

Finally, to consolidate our observations, we tackled the issue of individual-level variability. When looking at the heatmaps in the Appendix, which are the same as the one presented earlier but with more annotations (information of the patient, time point, lesion, frequency, synchronous or metachronous tumor), we can notice some patient-specific patterns. For example, in Figure A.2, we can notice that one patient is overrepresented in F3 and present a particularly cold iTME profile in CT. Similarly, in Figure A.3, a F2 patient present a very hot profile for several samples. Of note, this patient appear to be the only non recurrent patient included in F2.

5.2. CONCLUSION

In order to prevent individual-level bias, we developed a linear mixed model, often used in the case of study with multiple sample per individual [223, 224, 101]. This model validated the major described trends, such as F1 patients and the adenomatous pathway being overall more infiltrated, especially in the IM (section 4.5).

Observations between frequency groups that were specific to one pathway were more complicated to validate, as the model does not separate samples per pathway when looking at the impact of frequency group. A solution could be to add levels of frequency group per pathway, but giving too many parameters can lead to overfitting: creating a model that will end up being super tailored to this specific dataset and its variance, but not necessarily representative of the generalized process.

5.2 Conclusion

In summary, we demonstrated through integrative analysis that the two main colorectal pre-malignant pathway present two very distinct iTME, with the adenomatous pathway presenting an overall immune infiltration while serrated lesion were associated with more TLS surface and CT plasma cells. We completed these observation by showing that little variation of the iTME was observed between HPs and SSLs as well as LG AP and HG AP. Similarly, advanced status was not associated with important changes in iTME, except for T_{reg} and plasma cells in adenomatous lesions which were increased and decreased in advanced APs, respectively.

We then provided, to our knowledge, the first evidence that patients developing sporadic polyps at a high rate presented a less infiltrated iTME when compared to normal patients. The two pathways presented different profiles, with serrated polyps of low frequency patient presenting a strong immune infiltration disseminated between CT and IM, while APs of low frequency patient presented more mature TLS and IM infiltration than their high frequency counterparts.

We validated these observations by demonstrating that time point and recurrence did not impact these results, and by confirming that our observation were independent of intra-individual variables through the construction of a linear mixed model.

This study was limited by sample number, particularly after having to split samples by pathway and lesion type to avoid confounding effects. The small size of our samples also greatly reduced the number of samples available for RNAseq and WES, resulting in further loss of statistical power and sample variability. For these techniques, a bias was also induced, since only the larger lesions, associated with a higher rate of progression toward invasive carcinoma, were included. Another important limitation is the access to FFPE samples only, which limits experimental possibilities. Fresh samples could provide access to flow cytometry data or single-cell RNAseq, providing better resolution of the iTME.

The low number and high variability of cancer present in this cohort pushed us to exclude these samples, but a study involving more cancer samples could present important results regarding the

5.3. PERSPECTIVES

evolution of the iTME through carcinogenesis. This kind of approach has already been elegantly performed by Chen *et al.* In our cohort, the effect of synchronous or metachronous CRC on polyp iTME could also be assessed. A first and superficial analysis did not reveal any pattern associated with this parameter (Cf. heatmaps in Appendix)

5.3 Perspectives

Future research is needed to validate these results, in a larger cohort, but also by performing a deeper characterization of our cohort. As most of the results came in late during my PhD, several possible analyses still can and should be performed. For modeling, a model automatically selecting the variables of interest could help to better isolate the parameters that shape the iTME in our cohort. Bayesian projective prediction or LASSO model could both help answer this question.

Concerning RNAseq data, only immune scores and TLS signatures data were presented.

From the Consensus^{TME} analysis, evidence emerged that low frequency patients also presented a distinct myeloid iTME. An IHC panel that characterizes the main myeloid immune cells could be developed and applied to our cohort to better describe this part of the immune response.

Differential expression analysis has also been performed, but so far no clear immune effect has been isolated. By comparing differentially expressed genes (DEGs) between F1, F2, and F3 samples within each pathway, several genes were identified as part of the immunome, a set of genes that characterize immune populations in CRC [131]. Only 19 genes were common to both pathways, while 44 DEGs were specific to serrated lesions, and 39 to APs (Figure 5.2). Further analysis is required to better characterize transcriptomic modulation between risk groups. DEGs are presented in Table A.1.

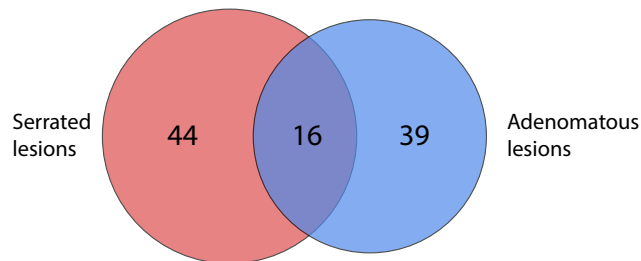


Figure 5.2: **Venn diagram of differentially expressed genes of the immunome between frequency group in each colorectal premalignant pathway**

Other parameters such as localization of the lesions within the colon could be taken into account, as it has been shown to greatly impact immune response [198].

Furthermore, WES should be used to assess the immunogenicity and CIN status of our samples. Differences in immunogenicity could explain the stronger immune response observed in low-risk patients. Interestingly, we observed that overrepresented transcripts in F1 patients were enriched in

5.3. PERSPECTIVES

noncoding RNAs, such as long noncoding RNAs (lncRNA) or pseudogenes (Figure 5.3 . lncRNAs are known to play an important role in carcinogenesis [225, 226], while pseudogenes and noncoding RNA in general are known for being particularly immunogenic and able to elicit immune response in CRC [227, 228, 229].

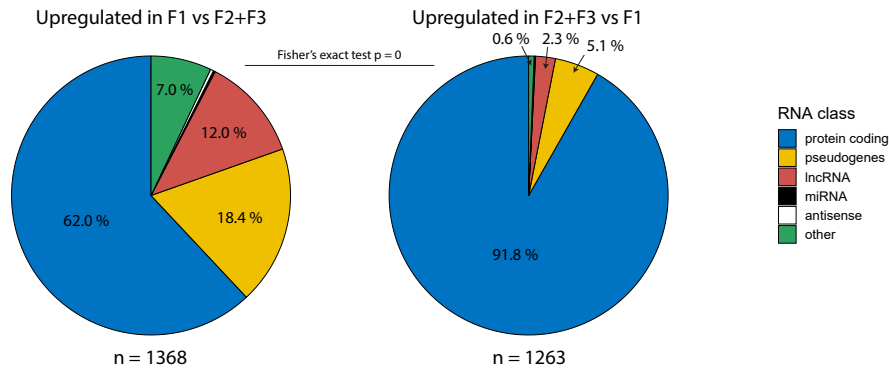


Figure 5.3: **F1 lesions are enriched in non coding RNAs.** The biotype of differentially expressed RNAs (p-value > 0.05) between frequency groups was assessed while pooling F2 and F3 lesions together

Finally, a new study aiming to assess the origin of the stronger immune response in low frequency patients could be very interesting. This strong immunity could develop along polyp growth (in relation to the immunogenicity of the lesion, for example) or be anterior to carcinogenesis, suggesting that the high frequency patient presents a default in their physiological colorectal immunity. Access to healthy samples and normal colon samples from low and high frequency patients could help provide a response to these questions. Mice models could also be used for a more mechanistic approach, but as the target population is average-risk patients, mice models selected for their polyp development capacity could probably not be adapted.

References

- [1] D.P. Tabassum and K. Polyak. Tumorigenesis: It takes a village. *Nature Reviews Cancer* [online]. 15.8 (2015), 473–483.
- [2] M. Burnet. Cancer-A Biological Approach I. The Processes Of Control. *British Medical Journal* [online]. 1.5022 (Apr. 1957), 779–786.
- [3] G.P. Dunn et al. *Cancer immunoediting: From immunosurveillance to tumor escape*. 2002.
- [4] J. Galon and D. Bruni. *Approaches to treat immune hot, altered and cold tumours with combination immunotherapies*. Mar. 2019.
- [5] D. Bruni, H.K. Angell, and J. Galon. The immune contexture and Immunoscore in cancer prognosis and therapeutic efficacy. *Nature Reviews Cancer* [online]. 20.11 (Aug. 2020), 662–680.
- [6] J. Galon et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* [online]. 313.5795 (Sept. 2006), 1960–1964.
- [7] C. Mascaux et al. Immune evasion before tumour invasion in early lung squamous carcinogenesis. *Nature* [online]. 571.7766 (2019), 570–575.
- [8] M.R.I. Young. Redirecting the focus of cancer immunotherapy to premalignant conditions. *Cancer Letters* [online]. 391 (2017), 83–88.
- [9] S.H. Kaufmann. Immunology’s coming of age. *Frontiers in Immunology* [online]. 10.APR (Apr. 2019), 1–13.
- [10] S.H. Kaufmann. Immunology’s foundation: The 100-year anniversary of the Nobel Prize to Paul Ehrlich and Elie Metchnikoff. *Nature Immunology* [online]. 9.7 (2008), 705–712.
- [11] S. Pinho and P.S. Frenette. Haematopoietic stem cell activity and interactions with the niche. *Nature Reviews Molecular Cell Biology* [online]. 20.5 (2019), 303–320.
- [12] K. Murphy et al. *Janeway’s Immunobiology* 9th edition (2016).
- [13] A.Y. Lai and M. Kondo. T and B lymphocyte differentiation from hematopoietic stem cell. *Seminars in Immunology* [online]. 20.4 (Aug. 2008), 207–212.
- [14] J.W. Griffith, C.L. Sokol, and A.D. Luster. Chemokines and chemokine receptors: Positioning cells for host defense and immunity. *Annual Review of Immunology* [online]. 32 (2014), 659–702.
- [15] T. Boehm and C.C. Bleul. The evolutionary history of lymphoid organs. *Nature Immunology* [online]. 8.2 (Feb. 2007), 131–135.
- [16] S. Kim et al. Multiscale engineering of immune cells and lymphoid organs. *Nature Reviews Materials* [online]. 4.6 (2019), 355–378.

REFERENCES

- [17] M.C. Dieu-Nosjean et al. Tertiary lymphoid structures in cancer and beyond. *Trends in Immunology* [online]. 35.11 (2014), 571–580.
- [18] T. Pradeu and E.L. Cooper. The danger theory: 20 years later. *Frontiers in Immunology* [online]. 3.SEP (2012).
- [19] N.S. Merle et al. Complement system part I - molecular mechanisms of activation and regulation. *Frontiers in Immunology* [online]. 6.JUN (2015), 1–30.
- [20] S. Epelman, K.J. Lavine, and G.J. Randolph. Origin and Functions of Tissue Macrophages. *Immunity* [online]. 41.1 (July 2014), 21–35.
- [21] E. Kolaczkowska and P. Kubes. Neutrophil recruitment and function in health and inflammation. *Nature Reviews Immunology* [online]. 13.3 (Mar. 2013), 159–175.
- [22] J.J. Costa, P.F. Weller, and S.J. Galli. The cells of the allergic response: Mast cells, basophils, and eosinophils. *Journal of the American Medical Association* [online]. 278.22 (Dec. 1997), 1815–1822.
- [23] C. Chester, K. Fritsch, and H.E. Kohrt. Natural killer cell immunomodulation: Targeting activating, inhibitory, and co-stimulatory receptor signaling for cancer immunotherapy. *Frontiers in Immunology* [online]. 6.DEC (Dec. 2015).
- [24] T. Worbs, S.I. Hammerschmidt, and R. Förster. Dendritic cell migration in health and disease. *Nature Reviews Immunology* [online]. 17.1 (2017), 30–48.
- [25] S.K. Wculek et al. Dendritic cells in cancer immunology and immunotherapy. *Nature Reviews Immunology* [online]. 20.1 (2020), 7–24.
- [26] P.A. Reche and E.L. Reinherz. Sequence variability analysis of human class I and class II MHC molecules: Functional and structural correlates of amino acid polymorphisms. *Journal of Molecular Biology* [online]. 331.3 (Aug. 2003), 623–641.
- [27] T. Mora and A.M. Walczak. How many different clonotypes do immune repertoires contain? *Current Opinion in Systems Biology* [online]. 18 (Dec. 2019), 104–110.
- [28] S. Tonegawa et al. Evidence for somatic generation of antibody diversity. *Proceedings of the National Academy of Sciences of the United States of America* [online]. 71.10 (Oct. 1974), 4027–4031.
- [29] L. Klein et al. Positive and negative selection of the T cell repertoire: What thymocytes see (and don't see). *Nature Reviews Immunology* [online]. 14.6 (2014), 377–391.
- [30] X. Xu, H. Li, and C. Xu. Structural understanding of T cell receptor triggering. *Cellular and Molecular Immunology* [online]. 17.3 (Mar. 2020), 193–202.
- [31] J. Zhu and W.E. Paul. CD4 T cells: Fates, functions, and faults. *Blood* [online]. 112.5 (Sept. 2008), 1557–1569.
- [32] A. Brewitz et al. CD8+ T Cells Orchestrate pDC-XCR1+ Dendritic Cell Spatial and Functional Cooperativity to Optimize Priming. *Immunity* [online]. 46.2 (Feb. 2017), 205–219.
- [33] S.N. Mueller and L.K. Mackay. Tissue-resident memory T cells: Local specialists in immune defence. *Nature Reviews Immunology* [online]. 16.2 (2016), 79–89.

REFERENCES

- [34] D.M. Pardoll. The blockade of immune checkpoints in cancer immunotherapy. *Nature Reviews Cancer* [online]. 12.4 (2012), 252–264.
- [35] T.H. Mann and S.M. Kaech. Tick-TOX, it's time for T cell exhaustion. *Nature Immunology* [online]. 20.9 (2019), 1092–1094.
- [36] O. Khan et al. TOX transcriptionally and epigenetically programs CD8+ T cell exhaustion. *Nature* [online]. 571.7764 (2019), 211–218.
- [37] X. Zhao, Q. Shan, and H.H. Xue. TCF1 in T cell immunity: a broadened frontier. *Nature Reviews Immunology* [online]. 22.3 (2022), 147–157.
- [38] I. Siddiqui et al. Intratumoral Tcf1 + PD-1 + CD8 + T Cells with Stem-like Properties Promote Tumor Control in Response to Vaccination and Checkpoint Blockade Immunotherapy. *Immunity* [online]. 50.1 (2019), 195–211.
- [39] A. Kallies, D. Zehn, and D.T. Utzschneider. *Precursor exhausted T cells: key to successful immunotherapy?* 2020.
- [40] F. Forquet et al. Presentation of antigens internalized through the B cell receptor requires newly synthesized MHC class II molecules. *Journal of immunology (Baltimore, Md. : 1950)* [online]. 162.6 (Mar. 1999), 3408–16.
- [41] R.H. Carter. B cells in health and disease. *Mayo Clinic Proceedings* [online]. 81.3 (Mar. 2006), 377–384.
- [42] T. Kurosaki, K. Kometani, and W. Ise. Memory B cells. *Nature Reviews Immunology* [online]. 15.3 (Mar. 2015), 149–159.
- [43] M. Rescigno et al. Dendritic cells express tight junction proteins and penetrate gut epithelial monolayers to sample bacteria. *Nature Immunology* [online]. 2.4 (Apr. 2001), 361–367.
- [44] U.M. Mörbe et al. Human gut-associated lymphoid tissues (GALT); diversity, structure, and function. *Mucosal Immunology* [online]. 14.4 (2021), 793–802.
- [45] M. Buettner and M. Lochner. Development and function of secondary and tertiary lymphoid organs in the small intestine and the colon. *Frontiers in Immunology* [online]. 7.SEP (2016), 1–11.
- [46] T.M. Fenton et al. Immune Profiling of Human Gut-Associated Lymphoid Tissue Identifies a Role for Isolated Lymphoid Follicles in Priming of Region-Specific Immunity. *Immunity* [online]. 52.3 (2020), 557–570.
- [47] E. Kaiserling. Newly-formed lymph nodes in the submucosa in chronic inflammatory bowel disease. *Lymphology* [online]. 34.1 (Mar. 2001), 22–29.
- [48] E. Dekker et al. Colorectal cancer. *The Lancet* [online]. 394.10207 (Oct. 2019), 1467–1480.
- [49] D. Arnold et al. Prognostic and predictive value of primary tumour side in patients with RAS wild-type metastatic colorectal cancer treated with chemotherapy and EGFR directed antibodies in six randomized trials. *Annals of Oncology* [online]. 28.8 (2017), 1713–1729.

REFERENCES

- [50] K.R. James et al. Distinct microbial and immune niches of the human colon. *Nature Immunology* [online]. 21.3 (2020), 343–353.
- [51] N. Shi et al. Interaction between the gut microbiome and mucosal immune system. *Military Medical Research* [online]. 4.1 (2017).
- [52] D. Hanahan and R.A. Weinberg. The hallmarks of cancer. *Cell* [online]. 100.1 (Jan. 2000), 57–70.
- [53] D. Hanahan and R.A. Weinberg. Hallmarks of cancer: The next generation. *Cell* [online]. 144.5 (2011), 646–674.
- [54] D. Hanahan. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* [online]. 12.1 (2022), 31–46.
- [55] L. Zhang and J.W. Shay. Multiple Roles of APC and its Therapeutic Implications in Colorectal Cancer. *Journal of the National Cancer Institute* [online]. 109.8 (2017).
- [56] R. Bonneville et al. Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precision Oncology* [online]. 1 (2017), 1–15.
- [57] M. Dankner et al. *Classifying BRAF alterations in cancer: New rational therapeutic strategies for actionable mutations*. 2018.
- [58] Y. Samuels et al. High Frequency of Mutations of the PIK3CA Gene in Human Cancers. *Science* [online]. 304.5670 (Apr. 2004), 554.
- [59] F. Bray et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians* [online]. 68.6 (2018), 394–424.
- [60] R. Cardoso et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. *The Lancet Oncology* [online]. 22.7 (July 2021), 1002–1013.
- [61] A. Cowppli-Bony et al. Descriptive epidemiology of cancer in metropolitan France: Incidence, survival and prevalence. *Bulletin du Cancer* [online]. 106.7-8 (July 2019), 617–634.
- [62] M. Gerstung et al. The evolutionary history of 2,658 cancers. *Nature* [online]. 578.7793 (2020), 122–128.
- [63] H. Lee-Six et al. The landscape of somatic mutation in normal colorectal epithelial cells. *Nature* [online]. 574.7779 (Oct. 2019), 532–537.
- [64] I.D. Nagtegaal et al. The 2019 WHO classification of tumours of the digestive system. *Histopathology* [online]. 76.2 (2020), 182–188.
- [65] A.H. Calderwood, K.E. Lasser, and H.K. Roy. Colon Adenoma Features and Their Impact on Risk of Future Advanced Adenomas and Colorectal Cancer. *World Journal of Gastrointestinal Oncology* [online]. 8.12 (2016), 826–834.
- [66] E.R. Fearon and B. Vogelstein. A genetic model for colorectal tumorigenesis. *Cell* [online]. 61.5 (1990), 759–767.

REFERENCES

- [67] S.M. Powell et al. APC mutations occur early during colorectal tumorigenesis. *Nature* [online]. 359.6392 (Sept. 1992), 235–237.
- [68] E.M. Schatoff, B.I. Leach, and L.E. Dow. WNT Signaling and Colorectal Cancer. *Current Colorectal Cancer Reports* [online]. 13.2 (2017), 101–110.
- [69] R.D. Burk et al. Integrated genomic and molecular characterization of cervical cancer. *Nature* [online]. 543.7645 (2017), 378–384.
- [70] D. Papadatos-Pastos et al. The role of the PI3K pathway in colorectal cancer. *Critical Reviews in Oncology/Hematology* [online]. 94.1 (Apr. 2015), 18–30.
- [71] H. Lee et al. PIK3CA amplification is common in left side-tubular adenomas but uncommon sessile serrated adenomas exclusively with KRAS mutation. *International Journal of Medical Sciences* [online]. 12.4 (2015), 349–353.
- [72] S.D. Crockett and I.D. Nagtegaal. Terminology, Molecular Features, Epidemiology, and Management of Serrated Colorectal Neoplasia. *Gastroenterology* [online]. 157.4 (2019), 949–966.
- [73] D.J. Weisenberger et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nature Genetics* [online]. 38.7 (2006), 787–793.
- [74] B. Chen et al. Differential pre-malignant programs and microenvironment chart distinct paths to malignancy in human colorectal polyps. *Cell* [online]. 184.26 (Dec. 2021), 6262–6280.
- [75] M. Øines et al. Epidemiology and risk factors of colorectal polyps. *Best Practice and Research: Clinical Gastroenterology* [online]. 31.4 (2017), 419–424.
- [76] Y. Yamaji et al. Incidence and recurrence rates of colorectal adenomas estimated by annually repeated colonoscopies on asymptomatic Japanese. *Gut* [online]. 53.4 (Apr. 2004), 568–572.
- [77] T. Amano et al. Number of polyps detected is a useful indicator of quality of clinical colonoscopy. *Endoscopy International Open* [online]. 06.07 (2018), E878–E884.
- [78] K.C. Noshirwani et al. Adenoma size and number are predictive of adenoma recurrence: Implications for surveillance colonoscopy. *Gastrointestinal Endoscopy* [online]. 51.4 I (Apr. 2000), 433–437.
- [79] H. Yoon et al. Total polyp number may be more important than size and histology of polyps for prediction of metachronous high-risk colorectal neoplasms. *BMC Gastroenterology* [online]. 22.1 (2022).
- [80] J.F. Viel et al. Predictors of Colorectal Polyp Recurrence after the First Polypectomy in Private Practice Settings: A Cohort Study. *PLoS ONE* [online]. 7.12 (2012).
- [81] F. Kastrinos and S. Syngal. Inherited colorectal cancer syndromes. *Cancer Journal* [online]. 17.6 (Nov. 2011), 405–415.
- [82] K. Chang et al. Immune profiling of premalignant lesions in patients with Lynch syndrome. *JAMA Oncology* [online]. 4.8 (Aug. 2018), 1085–1092.

REFERENCES

- [83] C.L. Chaffer and R.A. Weinberg. A perspective on cancer cell metastasis. *Science* [online]. 331.6024 (Mar. 2011), 1559–1564.
- [84] K. Simon. *Colorectal cancer development and advances in screening*. July 2016.
- [85] H. Brenner, C. Stock, and M. Hoffmeister. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: Systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ (Online)* [online]. 348.apr09 1 (Apr. 2014), g2467–g2467.
- [86] X. He et al. Long-term Risk of Colorectal Cancer After Removal of Conventional Adenomas and Serrated Polyps. *Gastroenterology* [online]. 158.4 (2020), 852–861.
- [87] R. Harewood et al. Adenoma characteristics associated with post-polypectomy proximal colon cancer incidence: a retrospective cohort study. *British Journal of Cancer* [online] (2022).
- [88] G.Y. Locker et al. *ASCO 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer*. 2006.
- [89] M.B. Amin et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more “personalized” approach to cancer staging. *CA: A Cancer Journal for Clinicians* [online]. 67.2 (Mar. 2017), 93–99.
- [90] Denoix PF. Tumor, node and metastasis (TNM). *Bull Inst Nat Hyg*. 1.6 (1944), 1–69.
- [91] K.A. Hoadley et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* [online]. 158.4 (Aug. 2014), 929–944.
- [92] C.R. Boland et al. A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: Development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Research*. 58.22 (Nov. 1998), 5248–5257.
- [93] R. Gryfe et al. Tumor Microsatellite Instability and Clinical Outcome in Young Patients with Colorectal Cancer. *New England Journal of Medicine* [online]. 342.2 (Jan. 2000), 69–77.
- [94] J. Guinney et al. The consensus molecular subtypes of colorectal cancer. *Nature Medicine* [online]. 21.11 (Nov. 2015), 1350–1356.
- [95] L. Marisa et al. Intratumor CMS heterogeneity impacts patient prognosis in localized colon cancer. *Clinical Cancer Research* [online]. 27.17 (Sept. 2021), 4768–4780.
- [96] E. Becht et al. Immune and stromal classification of Colorectal cancer is associated with molecular subtypes and relevant for precision immunotherapy. *Clinical Cancer Research* [online]. 22.16 (2016), 4057–4066.
- [97] B. Mlecnik et al. Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *Journal of Clinical Oncology* [online]. 29.6 (Feb. 2011), 610–618.
- [98] M.G. Anitei et al. Prognostic and predictive values of the immunoscore in patients with rectal cancer. *Clinical Cancer Research* [online]. 20.7 (2014), 1891–1899.

REFERENCES

- [99] F. Pagès et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *The Lancet* [online]. 391.10135 (2018), 2128–2139.
- [100] F. Pagès et al. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *Journal of Clinical Oncology* [online]. 27.35 (2009), 5944–5951.
- [101] B. Mlecnik et al. Comprehensive intrametastatic immune quantification and major impact of immunoscore on survival. *Journal of the National Cancer Institute* [online]. 110.1 (2018), 97–108.
- [102] B. Mlecnik et al. Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity* [online]. 44.3 (2016), 698–711.
- [103] J.I. Quezada-Marín et al. Gastrointestinal tissue-based molecular biomarkers: a practical categorisation based on the 2019 World Health Organization classification of epithelial digestive tumours. *Histopathology* [online]. 77.3 (Sept. 2020), 340–350.
- [104] G. Argilés et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology* [online]. 31.10 (Oct. 2020), 1291–1305.
- [105] M. Ferlitsch et al. Colorectal polypectomy and endoscopic mucosal resection (EMR): European Society of Gastrointestinal Endoscopy (ESGE) Clinical Guideline. *Endoscopy* [online]. 49.3 (2017), 270–297.
- [106] B. Ma et al. What has preoperative radio(chemo)therapy brought to localized rectal cancer patients in terms of perioperative and long-term outcomes over the past decades? A systematic review and meta-analysis based on 41,121 patients. *International Journal of Cancer* [online]. 141.5 (Sept. 2017), 1052–1065.
- [107] M.J. van der Valk et al. Long-term outcomes of clinical complete responders after neoadjuvant treatment for rectal cancer in the International Watch & Wait Database (IWWD): an international multi-centre registry study. *The Lancet* [online]. 391.10139 (2018), 2537–2545.
- [108] L.B. Saltz et al. Bevacizumab in combination with oxaliplatin-based chemotherapy as first-line therapy in metastatic colorectal cancer: A randomized phase III study. *Journal of Clinical Oncology* [online]. 26.12 (2008), 2013–2019.
- [109] D.T. Le et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *New England Journal of Medicine* [online]. 372.26 (June 2015), 2509–2520.
- [110] B.M. Meyers et al. Adjuvant Chemotherapy for Stage II and III Colon Cancer Following Complete Resection: A Cancer Care Ontario Systematic Review. *Clinical Oncology* [online]. 29.7 (2017), 459–465.
- [111] C.E. Sissy et al. Therapeutic implications of the immunoscore in patients with colorectal cancer. *Cancers* [online]. 13.6 (2021), 1–8.
- [112] J. Galon et al. The Continuum of Cancer Immunosurveillance: Prognostic, Predictive, and Mechanistic Signatures. *Immunity* [online]. 39.1 (July 2013), 11–26.

REFERENCES

- [113] A. Buqué et al. Immunoprophylactic and immunotherapeutic control of hormone receptor-positive breast cancer. *Nature Communications* [online]. *11.1* (2020).
- [114] T. Kimura et al. MUC1 vaccine for individuals with advanced adenoma of the colon: A cancer immunoprevention feasibility study. *Cancer Prevention Research* [online]. *6.1* (2013), 18–26.
- [115] J.J. Lohmueller et al. Antibodies elicited by the first non-viral prophylactic cancer vaccine show tumor-specificity and immunotherapeutic potential. *Scientific Reports* [online]. *6* (2016).
- [116] M.D. Vesely et al. Natural innate and adaptive immunity to cancer. *Annual Review of Immunology* [online]. *29* (2011), 235–271.
- [117] P. Ehrlich. Über den jetzigen Stand der Chemotherapie. *Berichte der deutschen chemischen Gesellschaft* [online]. *42.1* (Jan. 1909), 17–47.
- [118] W. MacCarty and A. Mahle. Relation of differentiation and lymphocytic infiltration to postoperative longevity in gastric carcinoma. *J Lab Clin Med.* *6* (1921), 473–480.
- [119] D.H. Kaplan et al. Demonstration of an interferon γ -dependent tumor surveillance system in immunocompetent mice. *Proceedings of the National Academy of Sciences of the United States of America* [online]. *95.13* (1998), 7556–7561.
- [120] V. Shankaran et al. IFN γ , and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* [online]. *410.6832* (Apr. 2001), 1107–1111.
- [121] R.D. Schreiber, L.J. Old, and M.J. Smyth. Cancer immunoediting: Integrating immunity's roles in cancer suppression and promotion. *Science* [online]. *331.6024* (2011), 1565–1570.
- [122] T.D. Martin et al. The adaptive immune system is a major driver of selection for tumor suppressor gene inactivation. *Science* [online]. *373.6561* (Sept. 2021), 1327–1335.
- [123] J. Galon, W.H. Fridman, and F. Pages. The adaptive immunologic microenvironment in colorectal cancer: A novel perspective. *Cancer Research* [online]. *67.5* (Mar. 2007), 1883–1886.
- [124] W.B. COLEV. I. The Treatment of Malignant Tumors by Repeated Inoculations of Erysipelas. *Annals of Surgery* [online]. *18* (July 1893), 68.
- [125] L. Galluzzi et al. *Immunological Effects of Conventional Chemotherapy and Targeted Anticancer Agents*. 2015.
- [126] C.R. Parish. Cancer immunotherapy: The past, the present and the future. *Immunology and Cell Biology* [online]. *81.2* (2003), 106–113.
- [127] J. Galon and D. Bruni. *Tumor Immunology and Tumor Evolution: Intertwined Histories*. 2020.
- [128] J.F. Brunet et al. A new member of the immunoglobulin superfamily-CTLA-4. *Nature* [online]. *328.6127* (1988), 267–270.
- [129] F.A. Harding et al. CD28-mediated signalling co-stimulates murine T cells and prevents induction of anergy in T-cell clones. *Nature* [online]. *356.6370* (Apr. 1992), 607–609.
- [130] D.R. Leach, M.F. Krummel, and J.P. Allison. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* [online]. *271.5256* (Mar. 1996), 1734–1736.

REFERENCES

- [131] G. Bindea et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* [online]. 39.4 (2013), 782–795.
- [132] Y. Wang et al. The immune landscape during the tumorigenesis of cervical cancer. *Cancer Medicine* [online]. 10.7 (2021), 2380–2395.
- [133] M. Meylan et al. Early Hepatic Lesions Display Immature Tertiary Lymphoid Structures and Show Elevated Expression of Immune Inhibitory and Immunosuppressive Molecules. *Clinical Cancer Research* [online]. 26.16 (2020), 4381–4389.
- [134] W.R. Becker et al. Single-cell analyses reveal a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *bioRxiv* [online] (2021), 2021.03.24.436532.
- [135] T.N. Schumacher and R.D. Schreiber. Neoantigens in cancer immunotherapy. *Science* [online]. 348.6230 (2015), 69–74.
- [136] C.C. Smith et al. Alternative tumour-specific antigens. *Nature Reviews Cancer* [online]. 19.8 (2019), 465–478.
- [137] T. Boon et al. Tumor antigens recognized by T lymphocytes. *Annual Review of Immunology* [online]. 12 (1994), 337–365.
- [138] T.N. Schumacher and R.D. Schreiber. Neoantigens in cancer immunotherapy. *Science* [online]. 348.6230 (Apr. 2015), 69–74.
- [139] R.M. Samstein et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nature Genetics* [online]. 51.2 (2019), 202–206.
- [140] D.S. Chen and I. Mellman. Oncology meets immunology: The cancer-immunity cycle. *Immunity* [online]. 39.1 (2013), 1–10.
- [141] M.L. Burr et al. An Evolutionarily Conserved Function of Polycomb Silences the MHC Class I Antigen Presentation Pathway and Enables Immune Evasion in Cancer. *Cancer Cell* [online]. 36.4 (2019), 385–401.
- [142] S. Aras and M. Raza Zaidi. TAMEless traitors: Macrophages in cancer progression and metastasis. *British Journal of Cancer* [online]. 117.11 (2017), 1583–1591.
- [143] M. Kinouchi et al. Infiltration of CD40-positive tumor-associated macrophages indicates a favorable prognosis in colorectal cancer patients. *Hepato-Gastroenterology* [online]. 60.122 (2013), 83–88.
- [144] C. Yang et al. Elevated CD163+/CD68+ Ratio at Tumor Invasive Front is Closely Associated with Aggressive Phenotype and Poor Prognosis in Colorectal Cancer. *International journal of biological sciences* [online]. 15.5 (2019), 984–998.
- [145] R. Nalio Ramos et al. Tissue-resident FOLR2+ macrophages associate with CD8+ T cell infiltration in human breast cancer. *Cell* [online]. 185.7 (Mar. 2022), 1189–1207.
- [146] Á. Teijeira et al. CXCR1 and CXCR2 Chemokine Receptor Agonists Produced by Tumors Induce Neutrophil Extracellular Traps that Interfere with Immune Cytotoxicity. *Immunity* [online]. 52.5 (Apr. 2020), 856–871.

REFERENCES

- [147] C. Cui et al. Neutrophil elastase selectively kills cancer cells and attenuates tumorigenesis. *Cell* [online]. 184.12 (2021), 3163–3177.
- [148] M. Shen et al. Tumor-associated neutrophils as a new prognostic factor in cancer: A systematic review and meta-analysis. *PLoS ONE* [online]. 9.6 (2014), e98259.
- [149] A.J. Gentles et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nature Medicine* [online]. 21.8 (2015), 938–945.
- [150] M.R. Galdiero et al. Occurrence and significance of tumor-Associated neutrophils in patients with colorectal cancer. *International Journal of Cancer* [online]. 139.2 (2016), 446–456.
- [151] F. Veglia, M. Perego, and D. Gabrilovich. Myeloid-derived suppressor cells coming of age review-article. *Nature Immunology* [online]. 19.2 (2018), 108–119.
- [152] J. Yu et al. Myeloid-Derived Suppressor Cells Suppress Antitumor Immune Responses through IDO Expression and Correlate with Lymph Node Metastasis in Patients with Breast Cancer. *The Journal of Immunology* [online]. 190.7 (2013), 3783–3797.
- [153] Y. Chen et al. Functional analysis of CD14+HLA-DR-/low myeloid-derived suppressor cells in patients with lung squamous cell carcinoma. *Oncology Letters* [online]. 14.1 (July 2017), 349–354.
- [154] V. Bronte et al. *Recommendations for myeloid-derived suppressor cell nomenclature and characterization standards*. 2016. arXiv: 0402594v3 [cond-mat].
- [155] G. Varricchi et al. Are mast cells MASTers in cancer? *Frontiers in Immunology* [online]. 8.APR (2017), 1.
- [156] L. Mehdawi et al. High tumor mast cell density is associated with longer survival of colon cancer patients. *Acta Oncologica* [online]. 55.12 (2016), 1434–1442.
- [157] J.P. Böttcher et al. NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control. *Cell* [online]. 172.5 (2018), 1022–1037.
- [158] S. Coca et al. The prognostic significance of intratumoral natural killer cells in patients with colorectal carcinoma. *Cancer* [online]. 79.12 (June 1997), 2320–2328.
- [159] M.H. Sandel et al. Prognostic value of tumor-infiltrating dendritic cells in colorectal cancer: Role of maturation status and intratumoral localization. *Clinical Cancer Research* [online]. 11.7 (Apr. 2005), 2576–2582.
- [160] V. Sisirak et al. Impaired IFN- α production by plasmacytoid dendritic cells favors regulatory T-cell expansion that may contribute to breast cancer progression. *Cancer research* [online]. 72.20 (Oct. 2012), 5188–97.
- [161] E.W. Roberts et al. Critical Role for CD103+/CD141+ Dendritic Cells Bearing CCR7 for Tumor Antigen Trafficking and Priming of T Cell Immunity in Melanoma. *Cancer Cell* [online]. 30.2 (Aug. 2016), 324–336.
- [162] J. Galon et al. Towards the introduction of the 'Immunoscore' in the classification of malignant tumours. *Journal of Pathology* [online]. 232.2 (Jan. 2014), 199–209.

REFERENCES

- [163] M. Tosolini et al. Clinical impact of different classes of infiltrating T cytotoxic and helper cells (Th1, Th2, Treg, Th17) in patients with colorectal cancer. *Cancer Research* [online]. 71.4 (2011), 1263–1271.
- [164] F. Amicarella et al. Dual role of tumour-infiltrating T helper 17 cells in human colorectal cancer. *Gut* [online]. 66.4 (2017), 692–704.
- [165] P. Salama et al. Tumor-infiltrating FOXP3+ T regulatory cells show strong prognostic significance in colorectal cancer. *Journal of Clinical Oncology* [online]. 27.2 (2009), 186–192.
- [166] F. Petitprez et al. Transcriptomic analysis of the tumor microenvironment to guide prognosis and immunotherapies. *Cancer immunology, immunotherapy : CII* [online]. 67.6 (2018), 981–988.
- [167] A. Schietinger et al. Tumor-Specific T Cell Dysfunction Is a Dynamic Antigen-Driven Differentiation Program Initiated Early during Tumorigenesis. *Immunity* [online]. 45.2 (Aug. 2016), 389–401.
- [168] M. Philip and A. Schietinger. CD8+ T cell differentiation and dysfunction in cancer. *Nature Reviews Immunology* [online]. 22.4 (July 2022), 209–223.
- [169] C.S. Jansen et al. An intra-tumoral niche maintains and differentiates stem-like CD8 T cells. *Nature* [online]. 576.7787 (2019), 465–470.
- [170] X. Wang et al. PD-L1 expression in human cancers and its association with clinical outcomes. *OncoTargets and Therapy* [online]. 9 (2016), 5023–5039.
- [171] M. Philip et al. Chromatin states define tumour-specific T cell dysfunction and reprogramming. *Nature* [online]. 545.7655 (2017), 452–456.
- [172] L.H. Lee et al. Patterns and prognostic relevance of PD-1 and PD-L1 expression in colorectal carcinoma. *Modern Pathology* [online]. 29.11 (2016), 1433–1442.
- [173] G.J. Yuen, E. Demissie, and S. Pillai. B Lymphocytes and Cancer: A Love–Hate Relationship. *Trends in Cancer* [online]. 2.12 (2016), 747–757.
- [174] C. Cui et al. Neoantigen-driven B cell and CD4 T follicular helper cell collaboration promotes anti-tumor CD8 T cell responses. *Cell* [online]. 184.25 (2021), 6101–6118.
- [175] R.D. Mazor et al. Tumor-reactive antibodies evolve from non-binding and autoreactive precursors. *Cell* [online]. 185.7 (2022), 1208–1222.
- [176] N.I. Affara et al. B cells regulate macrophage phenotype and response to chemotherapy in squamous carcinomas. *Cancer Cell* [online]. 25.6 (June 2014), 809–821.
- [177] W.H. Fridman et al. B cells and cancer: To B or not to B? *Journal of Experimental Medicine* [online]. 218.1 (2021).
- [178] J. Berntsson et al. Prognostic impact of tumour-infiltrating B cells and plasma cells in colorectal cancer. *International Journal of Cancer* [online]. 139.5 (Sept. 2016), 1129–1139.
- [179] S. Edin et al. The Prognostic Importance of CD20+ B lymphocytes in Colorectal Cancer and the Relation to Other Immune Cell subsets. *Scientific Reports* [online]. 9.1 (Dec. 2019), 19997.

REFERENCES

- [180] A. Shimabukuro-Vornhagen et al. Characterization of tumor-associated B-cell subsets in patients with colorectal cancer. *Oncotarget* [online]. 5.13 (2014), 4651–4664.
- [181] M. Van den Eynde et al. The Link between the Multiverse of Immune Microenvironments in Metastases and the Survival of Colorectal Cancer Patients. *Cancer Cell* [online]. 34.6 (2018), 1012–1026.
- [182] F. Petitprez et al. B cells are associated with survival and immunotherapy response in sarcoma. *Nature* [online]. 577.7791 (Jan. 2020), 556–560.
- [183] C. Germain et al. Presence of B cells in tertiary lymphoid structures is associated with a protective immunity in patients with lung cancer. *American Journal of Respiratory and Critical Care Medicine* [online]. 189.7 (2014), 832–844.
- [184] M. Meylan et al. Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. *Immunity* [online]. 55.3 (Mar. 2022), 527–541.
- [185] J. Goc et al. Dendritic cells in tumor-associated tertiary lymphoid structures signal a th1 cytotoxic immune contexture and license the positive prognostic value of infiltrating CD8+ t cells. *Cancer Research* [online]. 74.3 (2014), 705–715.
- [186] B.A. Helmink et al. B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* [online]. 577.7791 (2020), 549–555.
- [187] W.H. Fridman et al. B cells and tertiary lymphoid structures as determinants of tumour immune contexture and clinical outcome. *Nature reviews. Clinical oncology* [online] (Apr. 2022).
- [188] A. Maoz, M. Dennis, and J.K. Greenson. The Crohn’s-like lymphoid reaction to colorectal cancer-tertiary lymphoid structures with immunologic and potentially therapeutic relevance in colorectal cancer. *Frontiers in Immunology* [online]. 10.AUG (2019), 1884.
- [189] G. Di Caro et al. Occurrence of tertiary lymphoid tissue is associated with T-cell infiltration and predicts better prognosis in early-stage colorectal cancers. *Clinical Cancer Research* [online]. 20.8 (Apr. 2014), 2147–2158.
- [190] C.M. Schürch et al. Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* [online]. 182.5 (2020), 1341–1359.
- [191] A.E. Overacre-Delgoffe et al. Microbiota-specific T follicular helper cells drive tertiary lymphoid structures and anti-tumor immunity against colorectal cancer. *Immunity* [online]. 54.12 (Nov. 2021), 2812–2824.
- [192] C. Sautès-Fridman et al. Tertiary lymphoid structures in the era of cancer immunotherapy. *Nature Reviews Cancer* [online]. 19.6 (June 2019), 307–325.
- [193] P.C. Tumeh et al. PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* [online]. 515.7528 (2014), 568–571.
- [194] H. Wimberly et al. PD-L1 expression correlates with tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy in breast cancer. *Cancer Immunology Research* [online]. 3.4 (Apr. 2015), 326–332.

REFERENCES

- [195] M. Sade-Feldman et al. Defining T Cell States Associated with Response to Checkpoint Immunotherapy in Melanoma. *Cell* [online]. 175.4 (2018), 998–1013.
- [196] P. Maby et al. Correlation between density of CD8+ T-cell infiltrate in microsatellite unstable colorectal cancers and frameshift mutations: A rationale for personalized immunotherapy. *Cancer Research* [online]. 75.17 (2015), 3446–3455.
- [197] C.A. Rubio, B. Jacobsson, and E. Castaños-Velez. Cytotoxic intraepithelial lymphocytes in colorectal polyps and carcinomas. *Anticancer Research* [online]. 19.4 B (1999), 3221–3227.
- [198] K. Wallace et al. Immune responses vary in preinvasive colorectal lesions by tumor location and histology. *Cancer Prevention Research* [online]. 14.9 (Sept. 2021), 885–892.
- [199] A. Maglietta et al. The immune landscapes of polypoid and nonpolypoid precancerous colorectal lesions. *PLoS ONE* [online]. 11.7 (2016), 1–18.
- [200] J.A. Freitas et al. The adaptive immune landscape of the colorectal adenoma–carcinoma sequence. *International Journal of Molecular Sciences* [online]. 22.18 (2021), 1–12.
- [201] G. Cui et al. Reduced expression of microenvironmental Th1 cytokines accompanies adenomas–carcinomas sequence of colorectum. *Cancer Immunology, Immunotherapy* [online]. 56.7 (2007), 985–995.
- [202] T.J. Jang. Progressive increase of regulatory T cells and decrease of CD8+ T cells and CD8+ T cells/regulatory T cells ratio during colorectal cancer development. *Korean Journal of Pathology* [online]. 47.5 (2013), 443–451.
- [203] T.T. Rau et al. Inflammatory response in serrated precursor lesions of the colon classified according to WHO entities, clinical parameters and phenotype-genotype correlation. *Journal of Pathology: Clinical Research* [online]. 2.2 (2016), 113–124.
- [204] G. Acosta-Gonzalez et al. Immune environment in serrated lesions of the colon: intraepithelial lymphocyte density, PD-1, and PD-L1 expression correlate with serrated neoplasia pathway progression. *Human Pathology* [online]. 83 (Jan. 2019), 115–123.
- [205] A. Jimenez-Sanchez, O. Cast, and M.L. Miller. Comprehensive benchmarking and integration of tumor microenvironment cell estimation methods. *Cancer Research* [online]. 79.24 (2019), 6238–6246.
- [206] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome Biology* [online]. 11.10 (2010).
- [207] P.W. Eide et al. CMScaller: An R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Scientific Reports* [online]. 7.1 (2017).
- [208] G.A. Van der Auwera et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics* [online]. 43.SUPL.43 (2013), 1–11.

REFERENCES

- [209] A. Mayakonda et al. Maftools: Efficient and comprehensive analysis of somatic variants in cancer. *Genome Research* [online]. 28.11 (2018), 1747–1756.
- [210] D. Bates et al. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* [online]. 67.1 (2015).
- [211] W.C.C. Tan et al. Overview of multiplex immunohistochemistry/immunofluorescence techniques in the era of cancer immunotherapy. *Cancer Communications* [online]. 40.4 (2020), 135–153.
- [212] J.D. Pallua et al. The future of pathology is digital. *Pathology Research and Practice* [online]. 216.9 (Sept. 2020).
- [213] R.M. Levenson and J.R. Mansfield. Multispectral imaging in biology and medicine: Slices of life. *Cytometry Part A* [online]. 69.8 (2006), 748–758.
- [214] D. Coppola et al. Unique ectopic lymph node-like structures present in human primary colorectal carcinoma are identified by immune gene array profiling. *American Journal of Pathology* [online]. 179.1 (2011), 37–45.
- [215] R. Cabrita et al. Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* [online]. 577.7791 (Jan. 2020), 561–565.
- [216] K. Chang et al. Colorectal premalignancy is associated with consensus molecular subtypes 1 and 2. *Annals of Oncology* [online]. 29.10 (2018), 2061–2067.
- [217] M.A. Komor et al. Consensus molecular subtype classification of colorectal adenomas. *Journal of Pathology* [online]. 246.3 (2018), 266–276.
- [218] T. Saito et al. Two FOXP3 + CD4 + T cell subpopulations distinctly control the prognosis of colorectal cancers. *Nature Medicine* [online]. 22.6 (2016), 679–684.
- [219] F. Bergomas et al. Tertiary intratumor lymphoid tissue in colo-rectal cancer. *Cancers* [online]. 4.1 (2012), 1–10.
- [220] B. Chen et al. Profiling tumor infiltrating immune cells with CIBERSORT. *Methods in Molecular Biology* [online]. 1711 (2018), 243–259.
- [221] P.S. Yoon et al. Advances in Modeling the Immune Microenvironment of Colorectal Cancer. *Frontiers in Immunology* [online]. 11 (2021).
- [222] Y. Wolf, A.C. Anderson, and V.K. Kuchroo. TIM3 comes of age as an inhibitory receptor. *Nature Reviews Immunology* [online]. 20.3 (2020), 173–185.
- [223] L.J. Edwards. Modern statistical techniques for the analysis of longitudinal data in biomedical research. *Pediatric Pulmonology* [online]. 30.4 (Oct. 2000), 330–344.
- [224] J.P. Bernardes et al. Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. *Immunity* [online]. 53.6 (2020), 1296–1314.
- [225] S.J. Liu et al. Long noncoding RNAs in cancer metastasis. *Nature Reviews Cancer* [online]. 21.7 (July 2021), 446–460.

REFERENCES

- [226] S. Chen and X. Shen. Long noncoding RNAs: functions and mechanisms in colon cancer. *Molecular Cancer* [online]. 19.1 (2020).
- [227] J. Cleyle et al. Immunopeptidomic analyses of colorectal cancers with and without microsatellite instability. *Molecular & Cellular Proteomics* [online] (2022), 100228.
- [228] R. Xiang et al. Increased expression of peptides from non-coding genes in cancer proteomics datasets suggests potential tumor neoantigens. *Communications Biology* [online]. 4.1 (Dec. 2021), 496.
- [229] C.M. Laumont et al. Noncoding regions are the main source of targetable tumor-specific antigens. *Science Translational Medicine* [online]. 10.470 (Dec. 2018), 1–12.

A Appendix

Table A.1: **List of DEGs from the immunome between frequency group.** In each precancerous pathway, genes from the immunome were extracted from a differential gene expression analysis between frequency group (using the thresholds $f_c = 0.5$ and $p\text{-value} = 0.05$). Genes common to both pathway were identified

Serrated specific DEGs	Adenomatous specific DEGs	Common DEGs
ANK1	ABT1	ATF7IP
APOE	ADARB1	BCL2
AQP3	AKT3	CD96
BLVRB	ATG7	DDX17
CASP8	ATM	EWSR1
CD19	BACH2	GADD45A
CD4	BIRC5	GOLGA8A
CENPF	CD2	HES1
CHI3L2	CD3E	HMGB1
CHIT1	CDC14A	HPGDS
CRISPLD2	CHI3L1	LILRB2
CSF3R	CMAHP	MS4A2
DNAJB1	COLEC12	PCM1
DOCK9	CPA3	PGAM5
FRYL	CR2	SNRPN
GNAS	FABP4	TNFRSF14
GPC4	GALC	
HDC	IGKC	
IGFBP5	KLF12	
IL21R	KLRB1	
INPP4B	LDLRAD4	
MAOB	LRP8	
MAPRE3	MAF	
MEF2C	MICAL2	
MICAL3	MS4A1	
NFATC4	MS4A6A	
NR2C2	MSR1	
NUDT9	PDXK	
PPP2R5C	PTGDS	
PRKCQ	PTPN13	
RIPK3	RAI14	
RPA1	RORA	
SEC24C	SLC7A6	
SH3TC1	SPN	
SLC15A2	STX16	
SLC16A7	TACSTD2	
SLC26A6	TCL1A	
SLC30A5	TLR4	
TAL1	TRAF3IP3	
TMC6		
TPSAB1		
TRAPPC9		
TSC22D3		
ZFP36L2		

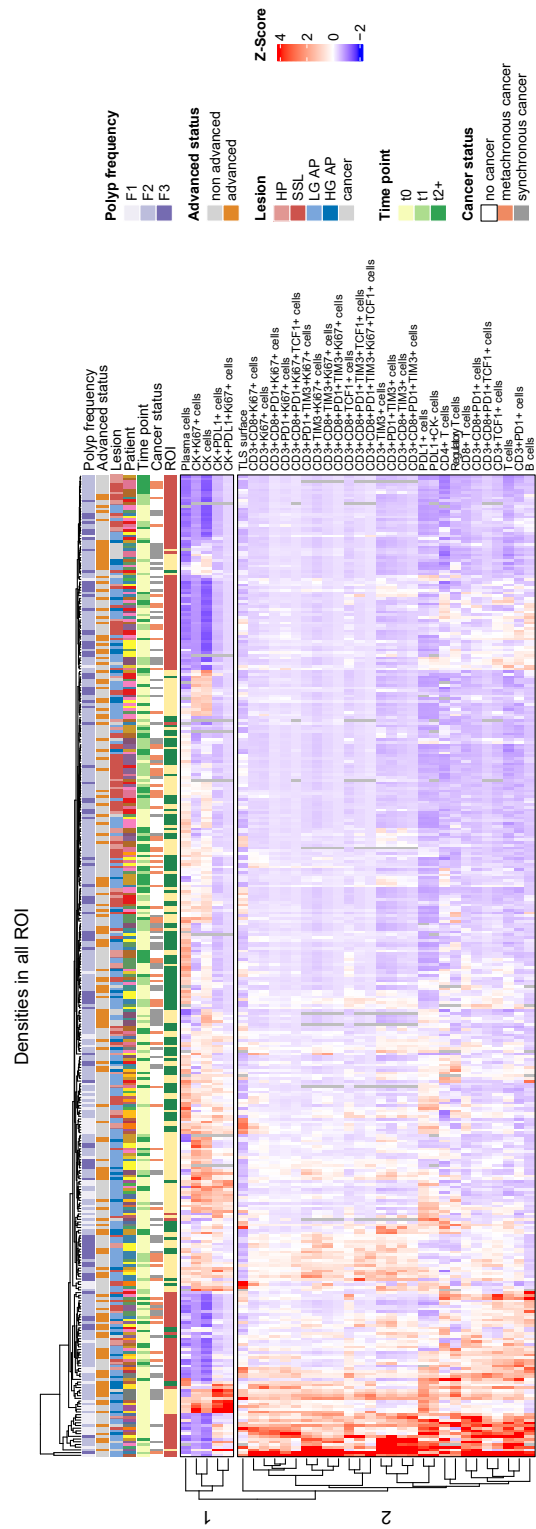
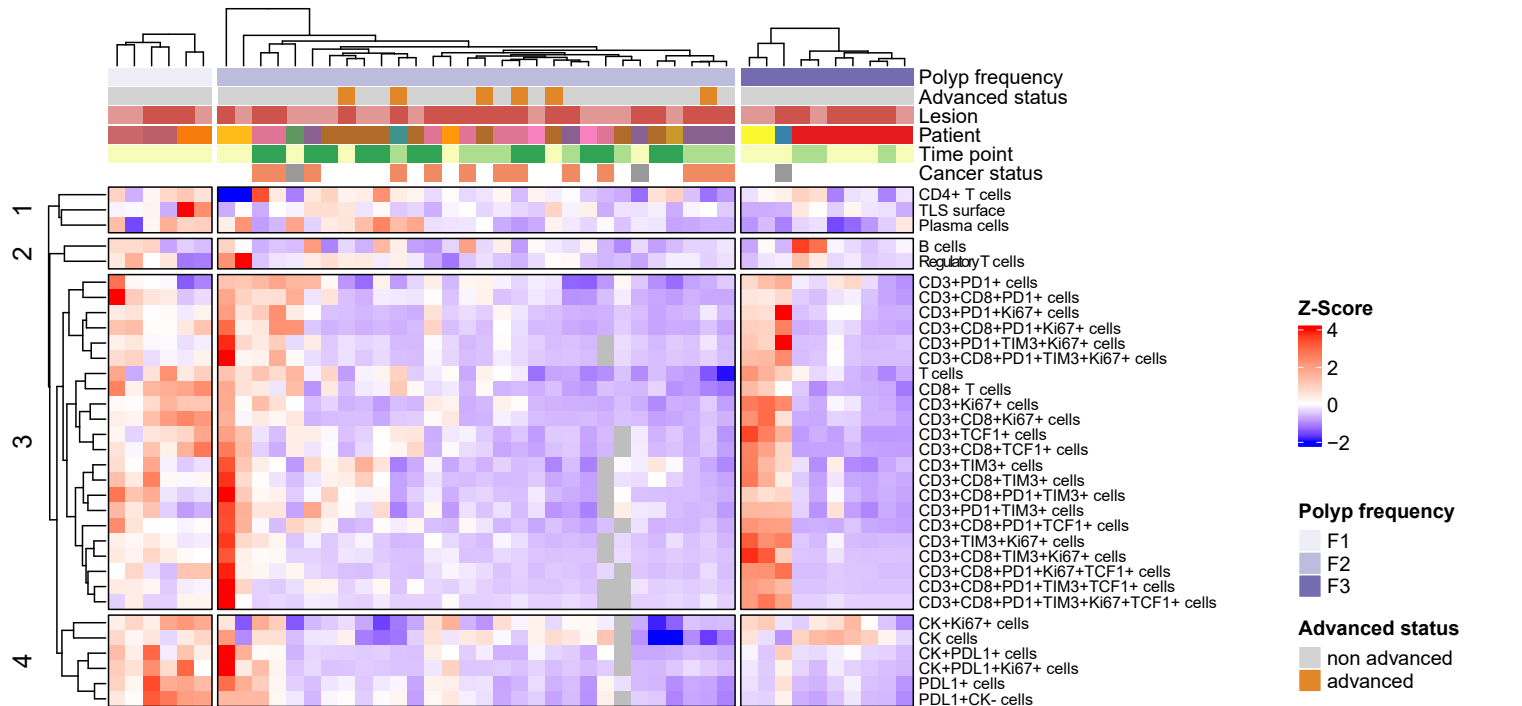


Figure A.1: Heatmap of all ROI from all samples

Densities in serrated CT



Densities in serrated IM

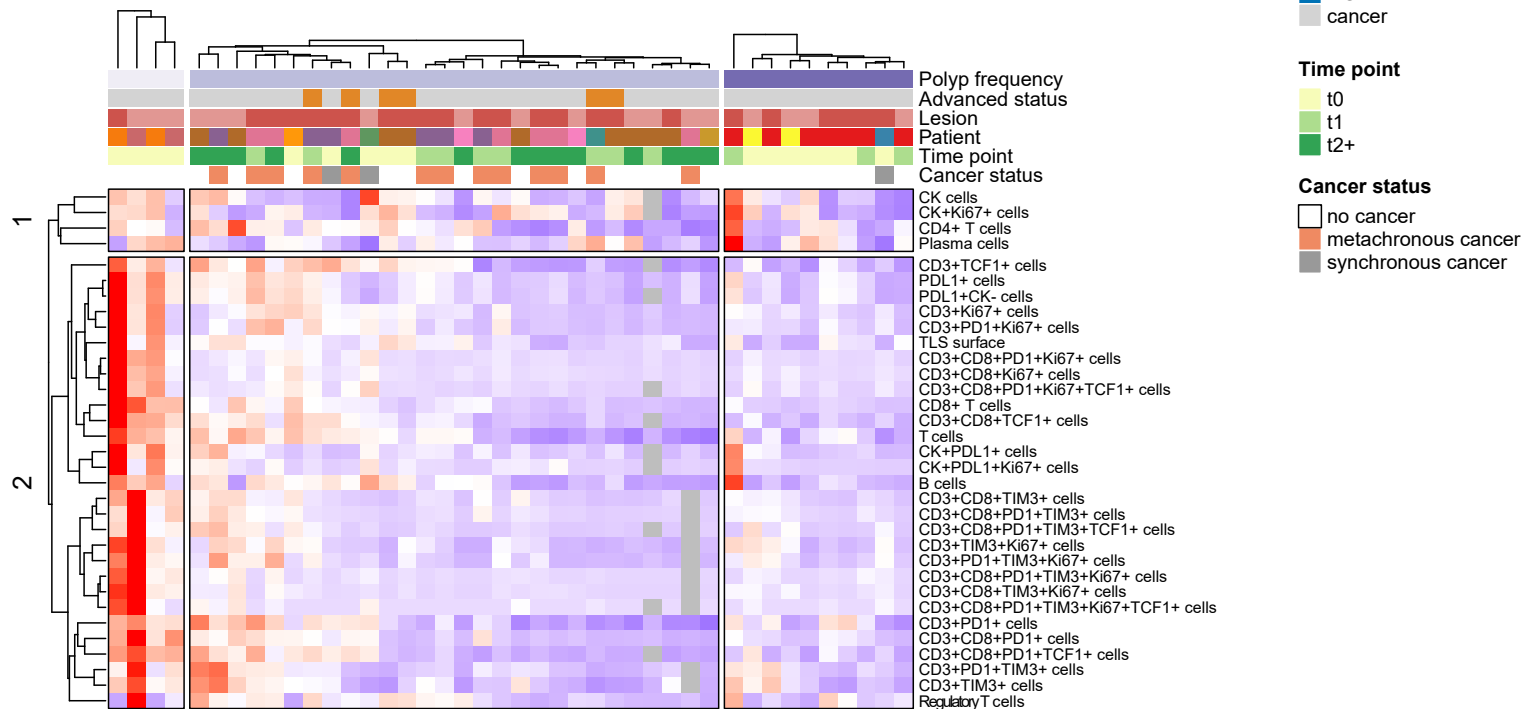


Figure A.2: Heatmap of the serrated per polyp frequency

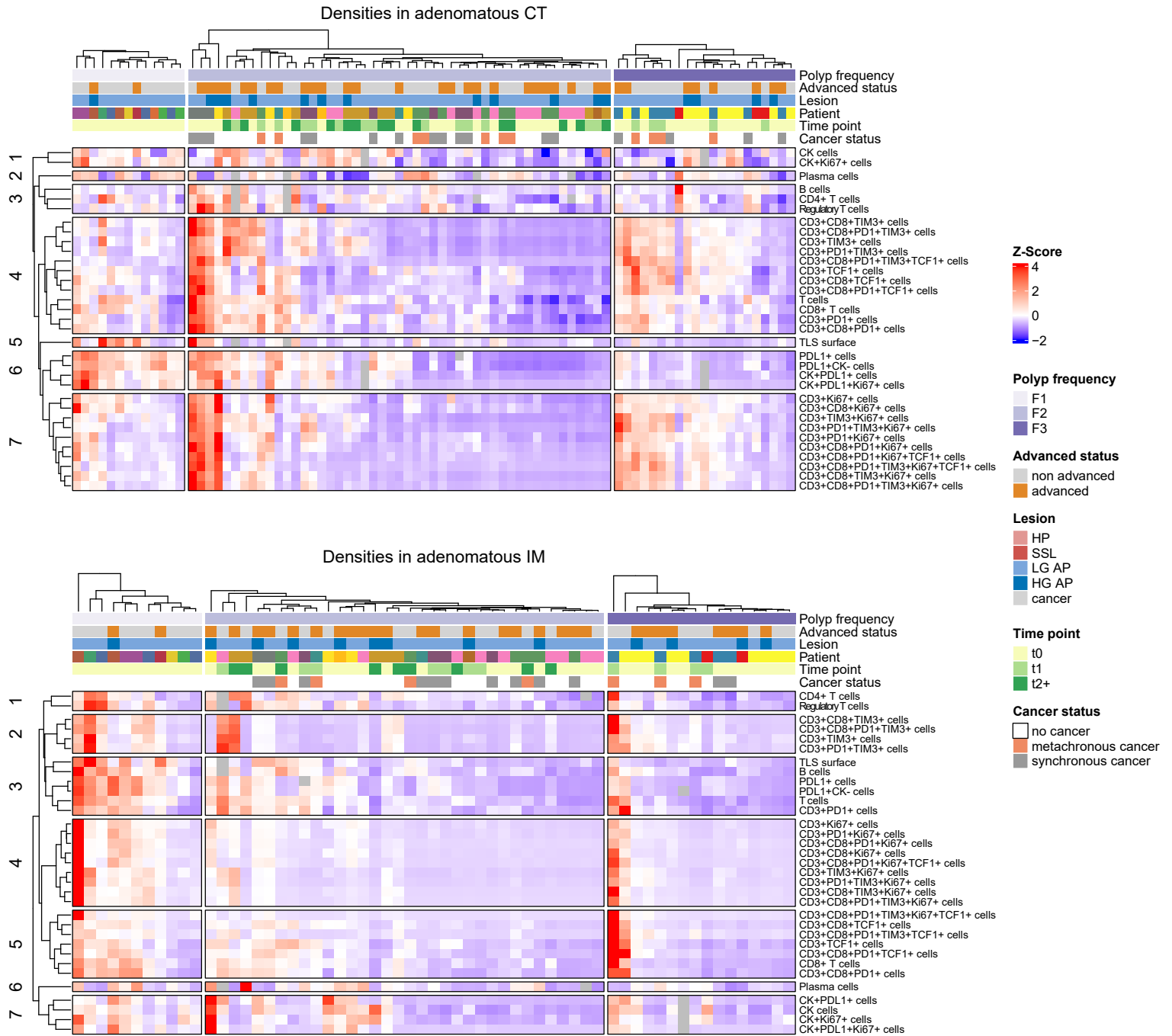


Figure A.3: Heatmap of the adenomatous per polyp frequency