

**Inaugural dissertation
for
obtaining the doctoral degree
of the
Combined Faculty of Mathematics, Engineering
and Natural Sciences
of the
Ruprecht - Karls - University
Heidelberg**

**Presented by
M.Sc. Nives Cecere
Born in: Viterbo, Italy
Oral examination: 16-12-24**

**Bovine Meat and Milk Factor (BMMF) Protein Expression
Coincides with Peritumor Alveolar Macrophages in Lung
Cancer Tissues**

**Referees: Prof. Dr. Dr. h.c. Ralf Bartenschlager
Prof. Dr. Mathias Heikenwälder**

Zusammenfassung

Obwohl bis zu 90% aller Lungenkrebsfälle mit dem Rauchen in Verbindung gebracht werden, gelten auch andere Umwelt- und Ernährungsfaktoren sowie chronische Entzündungen als Krebsrisikofaktoren. Bei Metzgern und Arbeitern in Schlachtereien und anderen Fleischverarbeitungen konnte eine erhöhte Krebsinzidenz festgestellt werden; umgekehrt wurde eine geringere Inzidenz bei Personen mit Laktoseintoleranz oder bei regelmäßiger Einnahme nicht-steroidaler entzündungshemmender Medikamente beschrieben. Die globale Epidemiologie deutet auf einen Zusammenhang zwischen dem Auftreten von Lungenkrebs und dem Verzehr von Rinderprodukten hin. Die Suche nach infektiösen Risikofaktoren führte zur Isolierung von plasmidähnlichen DNA-Molekülen, den sogenannten Bovine Meat and Milk Factors (BMMF). BMMF wurden auf Grund einer erhöhten Expression eines konservierten BMMF-encodierten Proteins (Rep) in Darmkrebsgeweben gegenüber gesunder Kontrollen als mögliche Krebsrisikofaktoren beschrieben. Obwohl vesikuläre BMMF-Strukturen bereits in einigen wenigen Lungenkrebsgeweben durch Immunolektronenmikroskopie unterstützt durch Immunhistochemie (IHC) identifiziert wurden, wurde das Vorhandensein von BMMF in Lungenkrebs bisher noch nicht systematisch untersucht. In dieser Studie wurde die Rep-Expression in assoziierten Tumor- und Peritumorgeweben von Lungenkrebspatienten sowie krebsfreien Personen (Kontrollgruppe) mittels IHC, Co-Immunfluoreszenzmikroskopie (IF) und Immunoblot auf der Basis monoklonaler anti-Rep-Antikörper und Massenspektroskopie quantifiziert. Darüber hinaus wurde ein öffentlich zugänglicher scRNA-seq-Datensatz zur Detektion möglicher BMMF-positiver Zellen ausgewertet. Unsere Arbeiten zeigten, dass mit 5 von 8 getesteten Antikörpern eine spezifische und reproduzierbare IHC Gewebefärbung möglich war. In allen getesteten Peritumorgeweben (n=42) wurde eine positive Färbung festgestellt, während Tumor- und Kontrollgewebe weniger Gewebefärbung aufwiesen (12 von 37 positiv getesteten Tumorgeweben).

In Peritumorgeweben wurde mittels IHC und IF in Alveolarmakrophagen eine kongruente Färbung von Rep+CD68+-Zellen beobachtet. Die IF-Dreifachfärbung mit Rep, CD68 und CD163 (Marker für M2-ähnliche Makrophagen) ergab eine kongruente Färbung von Rep in CD68+CD163+ Alveolarmakrophagen. Eine zellbasierte Quantifizierung zeigte eine signifikant erhöhte Anzahl von Rep+CD68+CD163+ Zellen

im Peritumor im Vergleich zu Tumor- und Kontrollgewebe. Die BMMF-Färbung war nicht mit der Intensität des Rauchens assoziiert, sondern zeigte Tendenzen einer umgekehrten Korrelation. Das Vorhandensein von Rep in Lungenperitumorgewebe wurde durch den Nachweis von Banden durch Immunoblotting/Präzipitation von Gewebelysaten und BMMF (Rep-spezifischen) Peptiden, die durch Massenspektroskopie validiert wurden, bestätigt. Ein Screening nach BMMF-Reads in scRNA-seq-Daten von Lungenadenokarzinompatienten ergab BMMF+-Zellen in 83% der Patienten (Gesamtzahl der Patienten n=30), hauptsächlich in Makrophagen und T-Zellen. Die Färbung mit BMMF-Antikörpern ergab eine erhöhte Rep-Expression in CD68+(CD163+)-Makrophagen von Lungenperitumorgewebe im Vergleich zur Kontrollgruppe, was auf die Verwendung als Krebs-Biomarker und einen möglichen kausalen Beitrag hindeutet.

Die Identifizierung von Rep+-Makrophagen in scRNA-seq, ergänzt durch weitere scRNA-Daten, könnte es ermöglichen, gezielt BMMF-induzierte Expressionsänderungen zu untersuchen. Dies ist von besonderem Interesse, da Alveolarmakrophagen als treibende Kraft bei der frühen Tumorentstehung und als Ziel für therapeutische Maßnahmen beschrieben wurden.

Summary

Although up to 90% of lung cancer cases are linked to smoking, also exposure to other environmental and dietary factors and chronic inflammation are regarded as cancer risk factors. Increased cancer incidence was found for butchers and meat workers; conversely, decreased incidence was reported for people with lactose intolerance or regularly taking non-steroidal anti-inflammatory drugs. Global epidemiology points towards an association between lung cancer incidence and consumption of bovine products. A search for infectious risk factors led to the isolation of plasmid-like DNA molecules termed Bovine Meat and Milk Factors (BMMF), which have been described as cancer risk factors acting via inflammation-linked indirect carcinogenesis upon experimental detection of increased expression levels of a conserved BMMF protein (Rep) in tissues of colorectal cancer patients when compared to healthy controls by immunohistochemistry (IHC). Although vesicular BMMF structures were identified in lung cancer tissues by immune electron microscopy supported by IHC, the presence of BMMF was not systematically tested in lung cancer. In this study, Rep expression was quantified in paired tumour and peritumour tissues from lung cancer patients as well as cancer-free individuals (control) by IHC, co-immunofluorescence (IF) microscopy and immunoblot based on monoclonal anti-Rep antibodies and mass spectroscopy. In addition, publicly available scRNA-seq dataset was assessed to monitor BMMF⁺ cells. We observed specific and reproducible IHC staining with 5 out of 8 antibodies tested. We obtained positive staining in all peritumor tissues tested (n=42) while tumor and control tissues showed less staining (12 out of 37 tumor tissues tested positive).

In peritumour tissues, a congruent staining of Rep⁺CD68⁺ cells was observed in IHC and IF in alveolar macrophages. IF triple staining with Rep, CD68 and CD163 (marker of M2-like macrophages) revealed congruent staining of Rep in CD68⁺CD163⁺ alveolar macrophages. Cell-based quantification indicated a significantly increased number of Rep⁺CD68⁺CD163⁺ cells in peritumour when compared to tumor and control tissues. BMMF staining was not associated with the intensity of smoking, but showed trends of an inverse correlation. The presence of Rep in lung peritumour tissues was supported by detection of bands by immunoblotting/-precipitation of tissue lysates and BMMF (Rep-specific) peptides validated by mass spectroscopy. Screening for BMMF reads in scRNA-seq data from lung adenocarcinoma patients revealed BMMF⁺ cells in 83% of patients

(total number of patients n=30), mostly in macrophages and T-cells. In conclusion, BMMF antibody staining revealed Rep expression in CD68⁺(CD163⁺) macrophages of lung peritumour tissues and was enriched in case vs. control suggesting use as a cancer biomarker and a possible causal contribution. The identification of Rep⁺ macrophages in scRNA-seq, with the addition of further scRNA data, might enable to study expression changes induced by BMMF. This is of particular interest as alveolar macrophages have been described as drivers of early tumorigenesis and target for therapeutic intervention.

Table of contents

Abbreviations.....	4
1. Introduction.....	7
1.1. Bovine Meat and Milk Factors (BMMF)	7
1.1.1. Identification and source of BMMF	7
1.1.2. Genomic features and bioactivity of BMMF.....	9
1.1.3. BMMF expression in CRC tissues and indirect carcinogenesis.....	12
1.1.4. Case vs. control analysis in CRC and implications in prognosis	16
1.1.5. Pleomorphic vesicular BMMF structures detected by iEM in LUAD, PDAC and CRC	17
1.1.6. Indications of BMMF involvement in lung cancer development.....	18
1.2. Lung cancer	20
1.2.1. Macrophages in pulmonary microenvironment.....	20
1.2.2. Lung cancer epidemiology	25
1.2.3. Risk factors for lung cancer development	27
1.2.4. ROS and inflammation in lung cancer development.....	30
1.2.5. Pathology of lung cancer	33
1.2.6. Advances and limitations in lung cancer screening and therapy.....	36
2. Aims of the Study	40
3. Materials and Methods.....	42
3.1. Materials.....	42
3.1.1. Cohort description	42
3.1.2. Lists of reagents, buffers, devices and software.....	46
3.2. Methods.....	51
3.2.1. FFPE sections preparation	51
3.2.2. Immunohistochemistry staining	52
3.2.2.1. DAB positive area quantification	52
3.2.2.2. Cell-based quantification	52
3.2.2.3. Analysis of smoking status and clinical data of patients in correlation to Rep expression.....	53
3.2.3. Immunoblot	53

3.2.3.1. Tissue lysate preparation	53
3.2.3.2. Bradford protein assay	53
3.2.3.3. SDS-PAGE and western blot	54
3.2.3.4. Immunodetection	54
3.2.4. Immunoprecipitation	55
3.2.5. Mass spectrometry (proteomics DKFZ core facility).....	55
3.2.6. <i>In situ</i> hybridization.....	56
3.2.7. Quantitative PCR.....	57
3.2.7.1. RNA extraction	57
3.2.7.2. cDNA preparation.....	57
3.2.7.3. RT-qPCR	57
3.2.8. Fluorescence-Activated Cell Sorting (FACS)	58
3.2.8.1. Fresh tissues dissociation.....	58
3.2.8.2. Fluorescent antibody labeling	58
3.2.8.3. Sorting (DKFZ Flow Cytometry – FACS Facility)	59
3.2.9. RNA sequencing.....	59
3.2.9.1. RNA extraction from sorted cells	59
3.2.9.2. Bulk RNA-seq (DKFZ NGS Core Facility)	60
3.2.9.3. Analysis of Bulk RNA-seq results.....	60
3.2.10. <i>In silico</i> analysis	61
3.2.10.1. Pan-Cancer Analysis of Whole Genomes (PCAWG) screening	61
3.2.10.2. Single Cell RNA Sequencing (scRNA-seq)	61
4. Results.....	63
4.1. Cytoplasmic staining of specific lung cells detected by IHC	63
4.2. BMMF Rep staining localizes primarily in peritumor tissues	64
4.3. Rep expression is predominantly observed in tissue parts with CD68 ⁺ (alveolar) macrophages.....	69
4.4. Increased Rep expression in peritumor compared to tumor and control tissues..	70
4.5. Rep expression correlates with the detection of (M2-like) macrophages	71
4.6. BMMF expression is evident in low dose smokers	76
4.7. Characterization of tissue lysates by immunoblotting	79

4.8. Immunoprecipitation with tissue lysates results in increased WB band detection	79
4.9. Validation of BMMF Rep peptide through mass spectrometry analysis	80
4.10. BMMF reads detected by <i>in silico</i> profiling	81
4.11. BMMF reads detected in publicly available scRNA-seq dataset.....	86
5. Discussion.....	89
5.1. Rep expression is associated with alveolar macrophages	89
5.2. Rep expression is increased in case versus control.....	93
5.3. Are Rep ⁺ macrophages biomarkers for lung cancer?.....	93
5.4. Evaluation of possible association of BMMF expression with sex, age and pathology of the tumor	95
5.5. Heterogeneity of Rep expression in low-dose smokers	97
5.6. Detection of BMMF hits by immunodetection and mass spectroscopy and by <i>in silico</i> BMMF profiling	98
5.7. Trend for association of Rep ⁺ cells with “M2-like” macrophages in scRNA-seq	103
5.8. Limitations of the study	104
6. Conclusions and outlook.....	107
7. Supporting information.....	112
7.1. Figures.....	112
7.2. Tables	121
8. Literature.....	128
Acknowledgments	140

List of pictures

Figure 1. Representative Bovine Meat and ilk Factor (BMMF) isolates grouped according to their sequence homology into BMMF1 and BMMF2.	10
Figure 2. Localization of anti-BMMF1 Rep antibody epitopes together with clinical information of BMMF detection by IHC in CRC.....	15
Figure 3. Ultrastructural representation of BMMF in tissues of CRC, LUAD and PDAC patients stained with anti-Rep AB3.....	18
Figure 4. Heterogeneity of lung macrophages.....	22
Figure 5. Estimated age-standardized incidence rates (world) in 2020, lung, both sexes, all ages.	26
Figure 6. Main causes of cancer mortality across OECD (Organization for Economic Co-operation and Development) countries, by sex, 2017.	27
Figure 7. Pathological differentiation of lung cancer.	34
Figure 8. Histological profile of carcinomas of the lung and pleura.	36
Figure 9. Immunohistochemical detection of BMMF Rep in peritumor tissues.....	64
Figure 10. Immunohistochemical detection of BMMF Rep and CD68 in peritumor and tumor tissues.....	67
Figure 11. Blocking of anti-Rep antibody staining by pre-incubation with Rep in DAB IHC.	68
Figure 12. Immunohistochemical detection of BMMF Rep and CD68 in peritumor tissues.	70
Figure 13. Quantification of Rep ⁺ and CD68 ⁺ cells based on co-immunofluorescence microscopy.	72
Figure 14. Co-immunodetection of Rep, anti-CD68 and anti-CD163 expression.	74
Figure 15. Quantification of Rep ⁺ , CD68 ⁺ and CD163 ⁺ cells based on co-immunofluorescence microscopy and correlation of Rep ⁺ cells with M1 or M2-like macrophages.	75
Figure 16. Rep expression in non-smokers, ex-smokers and smokers.	78

Figure 17. Detection of BMMF protein by immunoblot and mass spectrometry analysis.....	81
Figure 18. Detection of BMMF1 based on RT-qPCR.....	84
Figure 19. Gene expression levels for M1- and M2-like genes.....	85
Figure 20. Histologic detection of BMMF RNA by in situ hybridization in peritumor of lung cancer patients	86
Figure 21. BMMF reads detected in publicly available scRNA-seq dataset.	88

List of tables

Table 1. Macrophages polarization markers in humans and mice.....	23
Table 2. Established and putative risk factors for lung cancer.	30
Table 3. Overview of cohorts and methods used for the analysis.	43
Table 4. Clinical information of patients included in the analysis	44
Table 5. Antibodies.....	46
Table 6. Probes and primers	47
Table 7. Chemicals, enzymes, kits, equipment.....	48
Table 8. Devices	50
Table 9. Softwares	51
Table 10. Overview of IHC DAB staining results.....	66
Table 11. <i>In silico</i> BMMF screening in lung cancer sequencing archives...83	
Table 12. BMMF hits in sorted cells	85

Abbreviations

°C	Degree Celsius
μ	Micro
4DCT	Four-Dimensional Computed Tomography
Aa	Amino Acids
AB	Antibody
AM	Alveolar Macrophage
ANOVA	Analysis of Variance
BAL	Bronchial Lavage
BLAST	Basic Local Alignment Search Tool
BMMF	Bovine Meat and Milk Factor
Bp	Base Pair
BSA	Bovine Serum Albumin
BWA	Burrows-Wheeler Aligner
C1MI	Cow Milk Isolate BMMF group 1
C2MI	Cow Milk Isolate BMMF group 2
COPD	Chronic Obstructive Pulmonary Disease
CRC	Colorectal Cancer
CT	Computed Tomography
CT	Cycle Threshold
CXR	Chest X-rays
DAB	Diaminobenzidine
DAPI	Diamidino Phenylindole
DKFZ	Deutsche Krebsforschungszentrum
DNA	Deoxyribonucleic Acid
D-ViSioN	Detection of Integrated Viral Sequences by Singletons
EBC	Exhaled Breath Condensate
EM	Electron Microscopy
FACS	Fluorescence-Activated Cell Sorting
FFPE	Formalin-Fixed Paraffin-Embedded
H1MSB	See “MSBI”
HIV	Human Immunodeficiency Virus

HRP	Horseradish Peroxidase
ICH	Immunohistochemistry
IEM	Immunolectron Microscopy
IF	Immunofluorescence
IM	Interstitial Macrophage
IP	Immunoprecipitation
ISH	<i>In Situ</i> Hybridization
kDa	Kilo Dalton
LDCT	Low-Dose Computed Tomography
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
MF	Macrophage
MRI	Magnetic Resonance Imaging
MS	Mass Spectrometry
MSBI	Multiple Sclerosis Brain Isolate
n	Nano
NCT	National Center for Tumor Disease
Neu5Ac	N-Acetylneuraminic Acid
Neu5Gc	N-Glycolylneuraminic Acid
NGS	Next-Generation Sequencing
NK	Natural Killer
NSAID	Nonsteroidal Anti-Inflammatory Drugs
NSCLC	Non-Small Cell Lung Cancer
ORF	Open Reading Frame
PBS	Phosphate-Buffered Saline
PCAWG	Pan-Cancer Analysis of Whole Genomes
PDAC	Pancreatic Ductal Adenocarcinoma
PET	Positron Emission Tomography
PY	Pack Year
qRT-PCR	Quantitative Real Time Polymerase Chain Reaction
RCR	Rolling-Circle Replication
Rep	Replication Protein
RIPA	Radioimmunoprecipitation assay buffer

RNA	Ribonucleic acid
RNS	Reactive Nitrogen Species
ROS	Reactive Oxygen Species
SABR	Stereotactic Ablative Body Radiotherapy
SCLC	Small Cell Lung Cancer
scRNA-seq	Single Cell RNA Sequencing
SDS-PAGE	Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis
TAM	Tumor-Associated Macrophage
TBS	Tris-Buffered Saline
TME	Tumor Microenvironment
TRM	Tissue-Resident Macrophage
tSNE	T-distributed Stochastic Neighbour Embedding
WB	Western Blot
WH	Winged Helix

1. Introduction

1.1. Bovine Meat and Milk Factors (BMMF)

1.1.1. Identification and source of BMMF

When referring to cancers attributable to infections, it is important to distinguish between direct carcinogenesis and indirect carcinogenesis. In brief, direct carcinogenesis implies the persistence of nucleic acid of the infectious agents inside the cancer cells and the malignant growth depends on the expression of specific genes (e.g. viral genes). An example is Papillomavirus, where the tumor development is induced by the expression of the viral genes E6 and E7 [1]. Indirect carcinogenesis implies the role of the chronic inflammation – with production of oxygen radicals inducing random mutations in the host DNA - and does not depend on the persistence of the viral DNA in the cancer cells throughout time [2]. An example is *Helicobacter pylori* role in stomach cancer development. *H. pylori* infections, in fact, initially cause gastritis, but in persistent infections, this will lead to an increased risk of adenocarcinoma [3].

Several specific types of cancer have been linked to diet as a major risk factor. For instance, consuming red meat (which includes beef, lamb, and pork) has been identified as a risk factor for developing colorectal and lung cancer [4], [5], [6], [7]. Studies have found that long-term intake of dairy products may increase the risk of breast, prostate, and liver cancer [8], [9], [10], [11], [12]. Moreover, high childhood dairy intake (defined as 2+ glasses of milk per day or equivalent) was associated with 2.90-fold increased odds of colorectal cancer in adulthood, compared to low dairy intake (1/2 glass of milk or less per day) [13]. Initially, the increased cancer risk associated with high red meat intake was attributed to chemical byproducts like aromatic hydrocarbons and nitrosamines occurring during preparation or metabolism [14], [15]. But studies have not found any significantly elevated cancer risk with white meat consumption, such as poultry and fish, even though they produce the same aromatic compounds during preparation [7], [9], [16].

According to global epidemiology, there appears to be an association between the incidence of colon cancer worldwide and the distribution of dairy cattle

derived from the Eurasian aurochs (*Bos taurus*) [9], [17]. Therefore, zur Hausen proposed a model linking colon cancer incidence and specific nutrition products from bovine species. Evidence of species-specificity is demonstrated by the low incidence of colon cancer in Mongolia (6.6/6.1 per 100,000 in men/women) [18], [19], [20], despite the high consumption of beef, but where different type of bovine species (Yaks and Zebu-related Chinese Yellow cattle) is consumed [9], [17]. Similarly, Bolivia also shows low levels of colon cancer cases (6.3/5.1 per 100,000 in men/women) [18], [19], [20], which may be attributed to the consumption of a different cattle population such as Zebu [10]. These findings have led to investigation of a possible link between species-specific zoonotic infectious risk factors and cancer development [2], [21].

As a result, our group isolated over 200 circular single-stranded DNA sequences from a variety of sources, including dairy products, cattle sera, human sera, and brain tissue samples obtained from patients with multiple sclerosis [22], [23], [24], [25]. These agents were named, because of their origin, Bovine Meat and Milk Factors (BMMF) [21], [22], [23], [24], [25], [26] (Figure 1). BMMF were divided into four distinct families, labeled as BMMF1, BMMF2, BMMF3 and BMMF4, based on their similarities to previously known DNA elements (Sphinx 1.76, Sphinx 2.36, Gemycircular virus and *Psychrobacter* plasmid DNA) [21] [26]. BMMF1 and BMMF2 were defined because of their homology to the transmissible spongiform encephalopathy (TSE)-associated circular DNA isolates Sphinx 1.76 (GenBank HQ444404.1) and Sphinx 2.36 (GenBank HQ444405.1) [27], [28]. The genome size of the isolated BMMF ranges from approximately 600 to 3100 nucleotides [26], and both BMMF1 and BMMF2 are characterized by a conserved open reading frame (ORF) that encodes a replication initiation protein (Rep), which is different for both groups. Both BMMF1 and BMMF2 genomes share partial sequence identity with plasmids or bacteriophages of *Acinetobacter spp* [21], [26], [29]. H1MSB.1 - its genome isolated from the brain of a MS patient (GenBank LK931491.1) - is the most extensively studied BMMF so far [21], [26], [30] because of its high replication, transcription and translation in human cells [31]. Additionally, measurement

of seroreactivity against the replication protein (Rep) of H1MSB.1 revealed detection of ABs in humans implying BMMF exposure [31]. In addition, protein structure of the first domain of the H1MSB.1 Rep has been resolved through crystallography [31].

Recently, BMMF genomes have been found in the milk of domesticated water buffaloes (living in zoos), sheep, and goats in Germany [32], [33]. The authors have assigned several circular ssDNA to BMMF1 and BMMF2 groups, following the classification [32], [33].

1.1.2. Genomic features and bioactivity of BMMF

BMMF1 isolates contain conserved DNA sequences upstream of the Rep coding gene (Figure 1). These sections consist of numerous adenine and thymine nucleotides known as AT-stretches, as well as 3-4 22 nt-long, repetitive, iteron-like sequences [21]. These iteron-like sequences are well-known motifs found in bacterial plasmids that act as binding sites for Rep proteins, thereby regulating Rep transcription and initiating DNA replication [34], [35]. Moreover, a conserved 12 nucleotide palindromic structure that lies before the iteron repeats may function as a potential origin of replication [36]. The H1MSB.1 Rep has been demonstrated to bind to its genome specifically via the iterons through an electronic mobility shift assay (EMSA) [37]. BMMF1 isolates share approximately 70% nucleotide identity with plasmids present in *Acinetobacter baumannii* spp., which persist as a set of extrachromosomal plasmids at differing number transferred horizontally and vertically in addition to the much larger bacterial genome [21], [38].

The BMMF2 isolates (Figure 1) exhibit similarities with e.g. the plasmid p4ABAYE of *Acinetobacter baumannii* in terms of their genome size (2000-2800 nt), nucleotide composition, absence of iterons, and encoding of a Rep protein belonging to the Rep superfamily 1/2 [26]. Plasmid p4ABAYE has no iterons, is not mobile, does not participate in gene exchange with the *A. baumannii* genome but can still replicate [39], [40]. The analysis of BMMF2 ORFs revealed major ORFs coding for Rep protein and two additional ORFs (with sizes ranging from 101 to 138 amino acids and 96 to 148 amino acids [26]). Those encode for a protein that is characterized by its RNA-binding domain, which is essential for its role in the retrotransposition process, and a

protein that is composed of two main domains: a reverse transcriptase domain and an endonuclease domain [29]. Smaller ORFs are present on the sense strand as well as multiple ORFs on the anti-sense strand with sizes of up to 125 amino acids. However, the putative proteins of these small and reverse strand ORFs revealed minimal similarity to known proteins [26]. A poly(A) signal was observed in almost all putative transcripts, indicating transcription in eukaryotes [26].

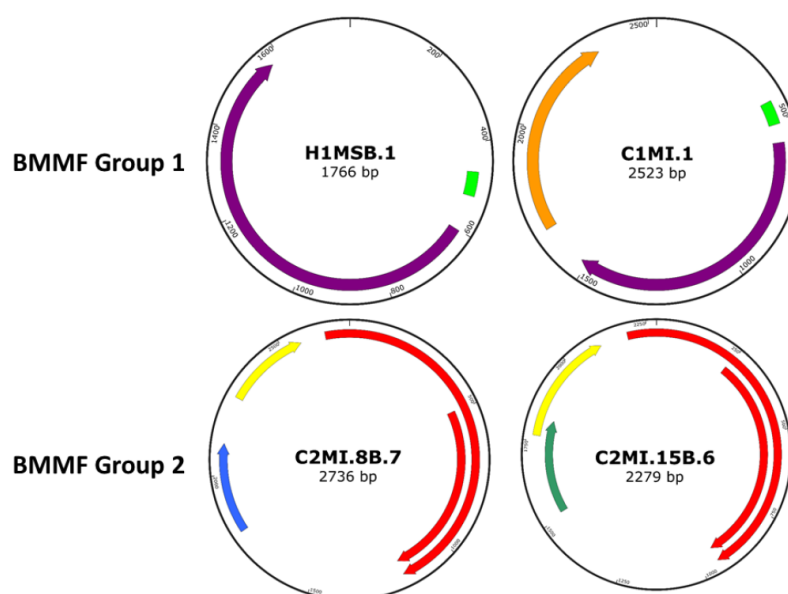


Figure 1. Representative Bovine Meat and ilk Factor (BMMF) isolates grouped according to their sequence homology into BMMF1 and BMMF2. Genomes were categorized based on their nucleotide sequence resemblance with Sphinx 1.76 and Sphinx 2.36. BMMF1 genomes, derived from either an MS brain isolate or cow milk (or bovine blood, not shown), are relevant to this study. MSB denotes a MS Brain isolate, CMI signifies Cow Milk isolates referring to BMMF1 (C1MI) or BMMF2 (C2MI). Rep encoding genes of BMMF1 are depicted in purple, with repetitive iteron regions upstream of BMMF1 Rep shown in light green. Both Rep-encoding ORFs, whether initiated with ATG or alternative start codons, are colored in red for BMMF2. Other open reading frames commonly identified in BMMF2 genomes are color-coded as blue, dark green, and yellow. BLASTP analyses of these putative proteins revealed minimal or no similarity to known proteins (Blue, Dark Green, Yellow). Arrows of the same color (red or yellow) denote ORFs encoding putative proteins displaying at least 60% amino acid sequence similarity between isolates. **Source:** Beck S. Production and characterization of novel BMMF2 Rep-specific antibodies and applications in cancer histology. PhD Thesis 2022.

Replication proteins (Rep) are encoded on bacterial plasmids, phage genomes, and viruses. They allow for autonomous replication of the circular plasmids without integrating into the host

chromosomes/pseudochromosomes of eukaryotes or bacteria [41]. Rep proteins have multiple functions such as DNA binding, helicase activity, or the ability to recruit host helicases. Rolling circle replication (RCR) is the method of genome replication mediated by these Rep proteins, which was first observed in the genome replication mechanism of bacteriophage ϕ X174 [42]. Rolling circle replication (RCR) is widely used for the replication of bacterial plasmids, viruses, and mobile genetic elements [35], [41], [43], [44]. RCR is characterized by asymmetric replication, where a leading and lagging DNA strand are separately synthesized [45], [46]. Initiation involves the Rep protein to induce a specific cleavage at the double-strand origin site (DSO) within a single-stranded region of the DNA. After the cleavage, the Rep protein remains covalently bound to the 5' end of the DNA at the cleavage site, while the newly generated 3' hydroxyl group serves as a primer for the elongation of the leading strand by the host's DNA polymerase. As DNA synthesis progresses, the 5' end of the original strand is displaced, and a second single-strand break occurs at the DSO. Subsequently ssDNA strands are ligated generating one double stranded plasmid and an additional ssDNA circle, which is in turn synthesized by host polymerases at the single-stranded origin (SSO). During the RCR, the leading strand is synthesized continuously by the host DNA polymerase, which may result in a higher relative abundance of this strand as replication progresses. The displaced strand, initially single-stranded, is expected to accumulate as replication continues [41], [47].

Using X-ray crystallography, the N-terminal domain of the Rep encoded by BMMF1 isolate H1MSB.1 (amino acids 2-133) was resolved with a resolution of 1.53 Å. The resulting structure showed a very high structure similarity to the bacterial RepA of *Pseudomonas syringae* despite low sequence identity (28%) supporting that at least parts of the BMMF1 Rep have replicating function [48]. Structure predictions based on the amino acid sequences of BMMF1 Rep proteins also suggest similarities with RepB in the Rep superfamily 3 [26]. Generally, RepB exhibits nicking-closing (topoisomerase I)-like activity and conducts a strand transfer reaction on single-stranded DNA (ssDNA) containing its specific target sequence [46]. Furthermore, Reps from BMMF1 show partial structural homology with

certain replication initiator proteins, including bacterial RepA or RepE, primarily within the central region of the Rep, although no additional conserved motifs were identified. Among BMMF group 1 Rep proteins, there is a notable amino acid sequence identity ranging from 80% to 98%.

The H1MSB.1 Rep ORF encodes a protein of 324 amino acids with a molecular weight of approximately 38 kDa. Based on structural homology modeling, a predicted structure of the BMMF1 Rep reveals two winged-helix (WH) domains, WH1 and WH2, supported by a common structural motif observed in replication proteins. DNA binding in Reps is facilitated by the pseudo symmetry of the protein structure, comprising two WH domains commonly found in transcriptional regulators [49], [50]. Winged helix domains are characterized by a helix-turn-helix (HtH) DNA-binding motif accompanied by one or two β -hairpin wings [51].

1.1.3. BMMF expression in CRC tissues and indirect carcinogenesis

Mouse monoclonal antibodies were created against BMMF1 H1MSB.1 Rep (or in general for BMMF1 Rep proteins) [30] and BMMF2 [52] (Figure 2 A). BMMF1 Rep-detecting antibodies were produced with two different approaches: one method involved two (synthetical) peptides derived from two conserved and immunogenic amino acid regions that were then used (individually) to immunize mice to generate consensus Abs. In a second approach, full length H1MSB.1 Rep was used for immunization. AB 3, 7, 10, 11 are specific for H1MSB.1, while AB 1, 2, 15 are consensus antibodies allowing for detection of overexpressed Rep proteins encoded on different BMMF1 isolates. The Rep monoclonal antibodies recognise different epitopes. AB3 and 10 detect an epitope at the C-terminal of the Rep protein, AB1 has its epitope in the central WH2 domain of Rep and AB2 and 15 detect an epitope at the N-terminal WH1 domain. AB11 stained positive for the WH1 region and detected the purified C-terminal, while AB 7 stained the full-length Rep protein, but was negative in the sub-localizations [30].

Immunohistochemical analyses using anti-Rep ABs (AB1, 2, 3, 7, 10 and 11) resulted in identification of Rep expression primarily in peritumor tissues of

colorectal cancer (CRC), particularly around the Lieberkühn crypts in the lamina propria (Figure 2 B) [30], [53]. Almost no Rep expression was observed in tumor sections of CRC. When comparing expression levels in the tumor-adjacent mucosa of CRC patient with age-matched healthy controls, significantly less Rep expression was observed in the healthy controls [30]. Immunostaining with inflammation-/immune cell-associated markers showed a congruent staining of CD68⁺ and BMMF1 Rep⁺ macrophages, indicating that Rep expression almost completely localizes in CD68⁺ cells, identifying CD68⁺ cells as the permissive cells for BMMF1 infection in the colon. Up to 40% of peritumor macrophages were Rep⁺, most of them CD68⁺CD163⁺ supporting their role as tumor-promoting M2-like macrophages [30]. A persistent increase in macrophage levels in the peritumor of CRC patients compared to healthy controls suggests persistent chronic inflammation [30]. Furthermore, BMMF1 DNA similar to H1MSB.1 was isolated specifically from peritumor tissue with high anti-Rep antibody staining via laser microdissection and not from sites localized in the tumor core [26], [30].

Based on the results, our group formulated a model for BMMF1 isolates as indirect causes of CRC and carcinogens [2], [21], [26], [30], [53]. The highest risk of BMMF infections occurs during the weaning period or when breastfeeding is absent. Therefore, the most critical time for BMMF infection is likely during the first years of life, particularly the first year after birth. During this period, feeding infants with dairy products contaminated by high levels of BMMF may lead to systemic infections, as the immune system is not yet fully developed. This can result in partial or complete immune tolerance to the pathogen's epitopes, as shown by a lack of T and B cell response in tissue areas with high Rep detection. In the following years and decades, a process of chronic inflammation occurs, during which BMMF molecules gradually accumulate until they reach detectable levels of protein expression without efficiently being cleared by the host. Various types of innate immune cells, including platelets, likely play a significant role in this process, particularly in initiating and sustaining chronic inflammation [30], [53]. Specific platelet activation, controlled by macrophages, in the liver known as CD68⁺ Kupffer cells, has been linked to chronic liver inflammation,

tissue damage, reactive oxygen species (ROS) production, and the formation of liver cancer. Platelets are direct targets for certain nonsteroidal anti-inflammatory drugs, which have been reported to reduce the incidence of liver, lung and colon cancer, among others. A diminished immune response against a widespread nutritionally acquired infection leads to the slow development of chronic inflammation, resulting in the generation of mutagenic radicals like ROS/RNA (as detected by surrogate markers, e.g. of oxidative DNA damage, such as 8-hydroxyguanosine antibody [30]). ROS/RNS can cause damage to macromolecules and initiate stress response mechanisms, promoting inflammation and potentially cell death [54]. This process eventually contributes to DNA mutations in replicating crypt cells (Ki67⁺ proliferative basal cells and early daughter cells located at the base of colon crypts) that leads to induction of dysplasia and adenoma and ultimately of CRC tumor. This transformation typically spans several decades [30], [53].

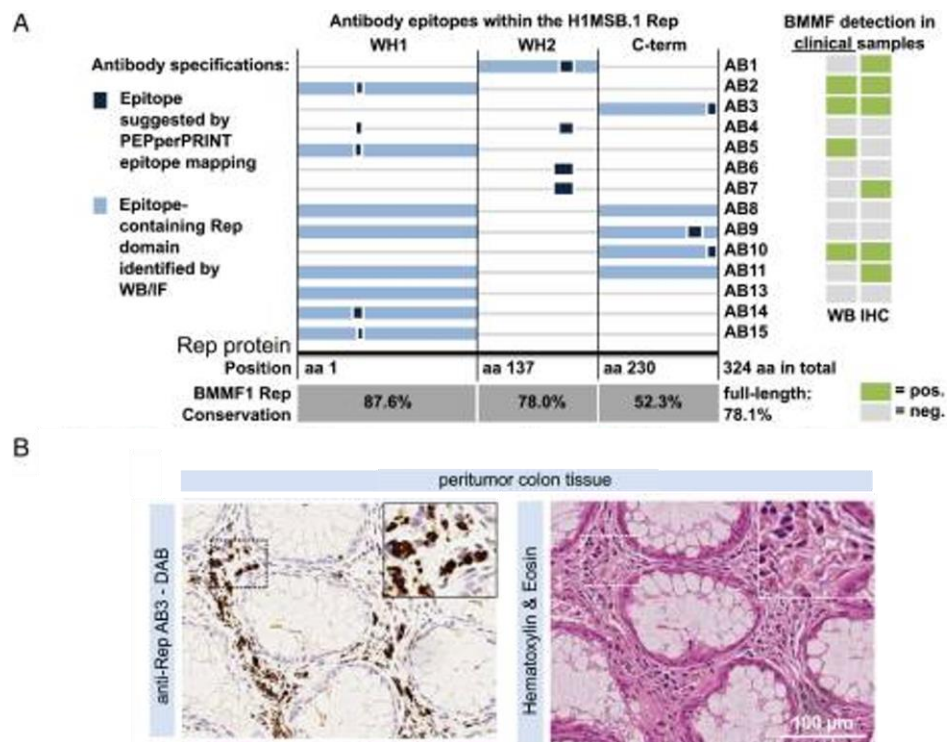


Figure 2. Localization of anti-BMMF1 Rep antibody epitopes together with clinical information of BMMF detection by IHC in CRC. **A:** Epitopes are shown for each anti-Rep antibody after PEPperPRINT epitope mapping (dark blue box) and WB/IF (light blue box) on a linear representation of the H1MSB.1 Rep with a summary of clinical staining results (WB and IHC). Sequence similarity of individual Rep domains is indicated for a set of BMMF1 Rep proteins (C1ML.1–4, H1MSB.1+2, C1HB.3–6, Sphinx1.76). **B:** IHC (DAB) staining of consecutive peritumor colon tissue sections from a CRC patient with Rep AB3 and H&E. (Magnification: E, Insets, 2x.). **Source:** Bund T. et al., Analysis of chronic inflammatory lesions of the colon for BMMF Rep antigen expression and CD68 macrophage interactions. PNAS. 2021.

The link between beef consumption and the development of CRC might be further explained by the sialic acid glycan N- glycolylneuraminic acid (Neu5Gc) [55], [56]. In humans, this sugar molecule cannot be produced due to a deletion in the CMAH gene, which is responsible for producing the enzyme that converts N-Acetylneuraminic acid (Neu5Ac) into Neu5Gc. As a result, humans acquire Neu5Gc e.g. from consuming red meat and dairy products [2], [21], [57], [58], [59], [60]. Neu5Gc plays a role in the infection from certain types of viruses, including polyomavirus type 9, parvoviruses, noroviruses, and rotaviruses [2], [21], [53], [61]. Neu5Gc might be crucial for BMMF entry, contributing to the development of CRC.

1.1.4. Case vs. control analysis in CRC and implications in prognosis

As mentioned above, Rep staining was mostly found in CD68⁺ macrophages localized in the peritumor area of CRC tissues and was increased when compared to healthy donors. The elevated Rep expression detected via BMMF Rep antibody staining suggests its potential use as a histologic marker for CRC. These findings also provide support for the model of a causal association between BMMF infection and CRC [30].

The majority of CD68⁺ macrophages that show positivity for Rep are M2-like macrophages. M2-like tumor-associated macrophages (TAMs) are a key component of the tumor microenvironment and they can promote cancer progression [62].

The presence of M2-like macrophages alone may not necessarily indicate a negative risk factor for CRC. However, evidence suggests an increase in Rep⁺ M2-like macrophages between cases and controls, particularly in tissues adjacent to adenomas and tumors. This suggests a causative function of BMMF already in the early stages of CRC development, also supported by the observation that the BMMF protein was detected in precancerous stages, indicating its potential as an early risk marker for colorectal cancer.

Additionally, high Rep expression in tumor-adjacent mucosa appears to correlate with increased CRC-specific death rates, suggesting a negative impact on patient prognosis even after tumor resection [63]. These findings underline the potential role of BMMF-induced inflammation in shaping the tumor microenvironment and influencing patient outcomes. It is possible that increased Rep expression modulates the tumor microenvironment unfavorably, promoting new tumors and metastasis as well as therapy resistance. Consequently, high Rep expression might shape the tumor-adjacent mucosa of cancer patients, even after tumor resection, in a way that induces an unfavorable immunological state with a worse prognosis, possibly by inducing higher levels of pro-tumorigenic M2-like macrophages. The high prevalence of BMMF Rep expression in CRC patients, along with its association with Rep⁺ M2-like CD68⁺ interstitial macrophages, suggests a potential causative involvement of BMMF in CRC development. Early

detection of BMMF Rep expression may offer insights into CRC risk and therapy response [63].

1.1.5. Pleomorphic vesicular BMMF structures detected by iEM in LUAD, PDAC and CRC

Conventional immunohistochemical staining demonstrated strong Rep expression in CD68⁺ macrophages in CRC but also LUAD and PDAC tissues. Co-immunofluorescence confirmed the localization of Rep within these macrophages [64]. Investigations of BMMF expression in a small set of CRC, LUAD and PDAC tissues using pre-embedding immunoelectron microscopy (IEM) on FFPE sections as well as structure retaining cryo-embedding, revealed pleomorphic, vesicular structures ranging from 50-200 nm in size, densely labeled with gold particles (Figure 3). These structures - identified predominantly in lamina propria macrophages - exhibited partly filamentous substructures and were consistent across all three cancer types studied. In contrast to pre-embedding IEM, cryo-IEM of fresh samples from CRC patients preserved cellular fine structure, facilitating the identification of the same target structures and the proximity to lipid membranes in control expressions of Rep in human cells.

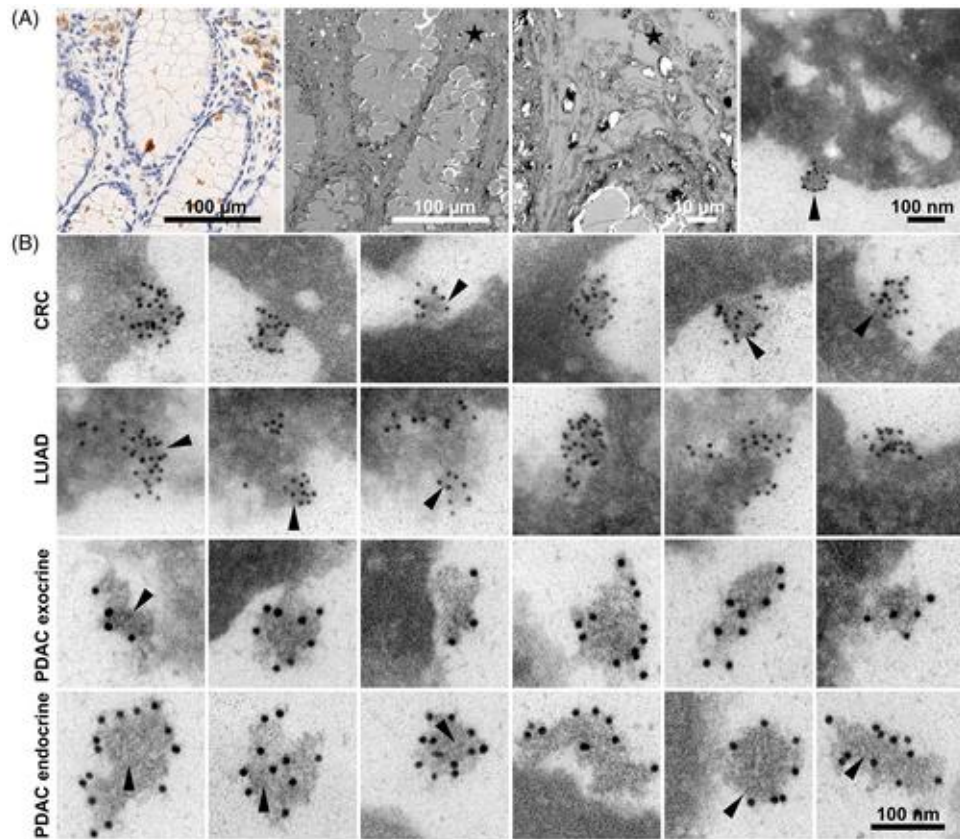


Figure 3. Ultrastructural representation of BMMF in tissues of CRC, LUAD and PDAC patients stained with anti-Rep AB3. **A:** Immunostaining of BMMF in human CRC, correlation of IHC with IEM in a representative peritumour source tissue of a CRC patient. Two consecutive sections depict the same region, one labeled with DAB (left) and the other with immunogold (magnification series). Small spherical bodies with a dense gold label (arrowhead) are observed in the DAB-positive tissue area. Asterisks denote the center of the following magnified sub image. Gold grain size: 5 nm. **B:** BMMF pleomorphic vesicular structures in human FFPE biopsies of CRC, LUAD, and PDAC exocrine and endocrine tissue regions. All three tumor tissues show small, 100 nm-sized bodies heavily stained with BMMF antibody. Although lacking a discernible structural motif, filamentous arrangements suggest higher-order organization (arrowheads). Gold grain size: 5 nm (CRC and LUAD) and 10 nm (PDAC). AB, antibody; BMMF, bovine meat and milk factor. **Source:** Nikitina E. et al., Structural expression of bovine milk and meat factors in tissues of colorectal, lung and pancreatic cancer patients. IJC. 2022.

1.1.6. Indications of BMMF involvement in lung cancer development

Infectious pathogens such as viruses, bacteria or parasites are involved in the development of certain types of cancer [65]. In 2018, 2.2 million of cancers attributable to infections were diagnosed worldwide representing 14-16% of the global cancer incidence [65].

Lung cancer remains the leading cause of cancer mortality in men and women worldwide [66]. About 90% of lung cancer cases are caused by smoking and the use of tobacco products, but there are cases reported for people who never smoked. Chronic inflammation not related to smoke can also play a significant role in the tumour development through the release of ROS/RNS, as well as TNF- α . Consequently, chronic inflammation can facilitate tumour growth via activation of specific pathways that can also lead to suppression of immune responses [66].

Recent research has shown that there is a comprehensive microbiome present in the lungs of both healthy individuals and those with respiratory diseases, contrary to the long-held belief that the area below the vocal cords in healthy lungs was sterile [67], [68]. Some factors that may contribute to the development of lung cancer include chronic inflammation, abnormalities in immune cells, scar tissue formation, and gene mutations caused by *Mycobacterium tuberculosis*, but no definitive link has been established yet [69]. Studies have found unique bacterial profiles in lung cancer-free patients, as well as differences in bacterial compositions between small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) [67]. In terms of the virome, Anelloviridae have been identified as the dominant viruses. Occasionally detected were herpesviruses, papillomaviruses, retroviruses, and other respiratory viruses. Respiratory viruses, including Anelloviridae, can infect and replicate in different lung cell types [67]. These can include epithelial cells lining the airways and alveoli, as well as immune cells such as macrophages and dendritic cells present in the lung tissue. It is worth noting that Anelloviridae are widespread in blood and tissues among adults [70], and they are typically under immune control and not linked to any physiological effects or disease [71]. Recent research has also identified members of the Redondoviridae family as the second most common DNA virus in ill patients. Redondoviridae belong to a group of viruses called circular replicase-encoding single stranded DNA (CRESS) viruses [72]. Overall, while the lung microbiome appears to be resident and persistent in individuals with respiratory diseases, whether it is the cause, or the result of the disease is still unclear [68].

Epidemiological patterns show an increased incidence of lung cancer in countries with high bovine meat or dairy product consumption [73]. Moreover, an increased incidence of lung cancer was observed for butchers, meat workers and animal farmers [4], [74], [75] and there is evidence that a low consumption of milk and other dairy products had decreased risks of lung cancer, particularly observed in people with a lactose intolerance [76]. The role of inflammation as a risk factor for lung cancer development is also suggested by the data showing a reduction of 36% risk of lung cancer among individuals who regularly take non-steroidal anti-inflammatory drugs (NSAIDs), such as aspirin and ibuprofen [2], [77], [78].

1.2. Lung cancer

1.2.1. Macrophages in pulmonary microenvironment

In the lungs, tissue-resident macrophages (TRMs) are integral to maintaining organ homeostasis, as they sense and integrate local and systemic information to provide specialized tissue cells with growth factors, nutrient recycling, and waste removal, which are critical for tissue growth, homeostasis, and repair. They contribute to organogenesis, promote tissue regeneration following damage, and defend against infectious diseases, making them essential for tissue metabolism and defense [79].

There are notable distinctions among these TRMs concerning their transcriptional patterns, development, characteristics, positioning, and roles. Additionally, in inflammatory conditions, monocytes can be recruited to the lung and assume roles within the macrophage population. They are identified as recruited macrophages (monocyte-derived macrophages) and exhibit many differences, both in terms of their phenotypic features and metabolic characteristics compared to the pre-existing TRMs. These distinctions are particularly evident during the initial days following the onset of infection or injury [80], [81].

TRMs and monocyte-derived macrophages arise from two different lineages. TRMs self-renew locally, independent of adult hematopoiesis, while monocyte-derived macrophages are produced in the bone marrow and arise

from adult hematopoietic stem cells and accumulate mostly in inflamed lesions [82].

Regardless of the origin of the macrophages, the macrophage colony-stimulating factor 1 receptor (CSF1R, *Csf1r* in mice) is the main regulator and important factor for differentiation. This receptor, a class III transmembrane tyrosine kinase, is expressed across most mononuclear phagocytic cells. Macrophages also express the surface markers CD14 and CD68 (F4/80 in mice) [83].

There are two different types of TRMs: interstitial macrophages (IM) and alveolar macrophages (AM) [84], [85]. AMs are TRMs specific to the lungs and exhibit a distinct phenotype and transcriptional fingerprint shaped by the pulmonary microenvironment. Positioned within the luminal region of the alveolar niche, AMs are surrounded by type I and type II alveolar epithelial cells, capillary endothelial cells, and alveolar interstitial fibroblasts, that provide cytokines essential for AM sustenance [81].

AMs exhibit the transcription factor peroxisome proliferator-activated receptor γ (PPAR γ), crucial in governing lipid metabolism, induced by the cytokine granulocyte-macrophage colony-stimulating factor (GM-CSF). Studies in mice have also shown the necessity of transforming growth factor β (TGF- β) in AMs development and maintenance. Unlike GM-CSF, predominantly secreted by alveolar type II cells, TGF- β originates from alveolar macrophages themselves for their homeostasis. Moreover, AMs express diverse factors promoting immune tolerance. Examples are TGF- β and inhibitory receptors, that maintain the steady-state and restrict the pro-inflammatory activity [84].

IMs contribute to barrier immunity with the AMs and are important for antigen presentation and tissue remodeling within the lung environment [84]. They are primarily present in the interstitial spaces surrounding the broncho vascular bundle and they are adjacent to various nonhematopoietic cell types including neurons, lymphatic vessels, blood vessels and hematopoietic cells. Unlike AMs, IMs are distributed throughout the body and exhibit a comparable transcriptional profile across diverse organs, indicating a conserved, specialized role irrespective of their anatomical location [81].

A characteristic of IMs is their capacity to produce the immunosuppressive cytokine interleukin-10 (IL-10). Moreover, IMs were recently categorized into two primary lineages in mice: i) LYVE-1^{low} MHC Class II^{high} IMs, which are positioned close to neurons, exhibit a role in antigen presentation. ii) LYVE-1^{high} MHC Class II^{low} perivascular macrophages, which exhibit a role in wound healing and tissue repair (Figure 4) [84].

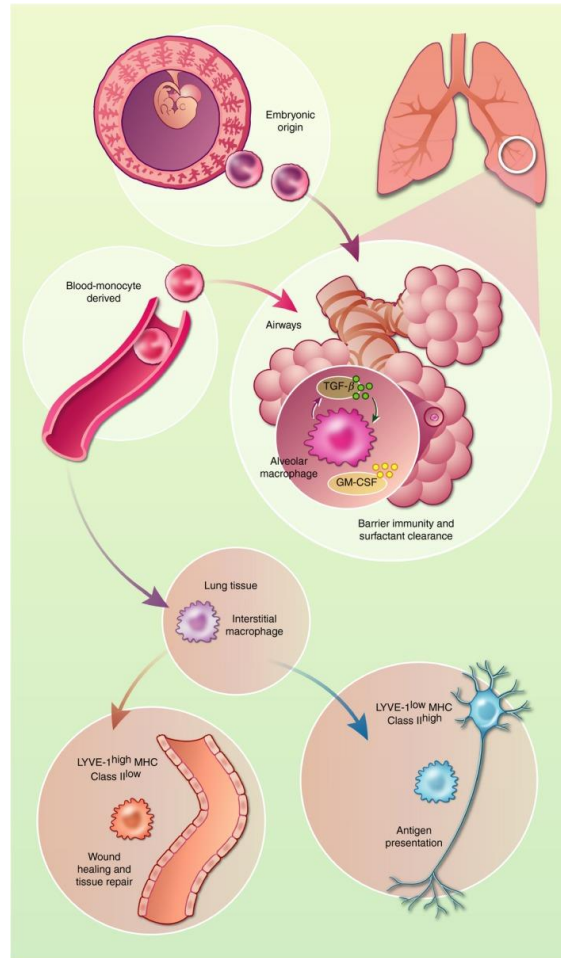


Figure 4. Heterogeneity of lung macrophages. Different locations within the lung correspond to specialized functions of lung macrophages. Alveolar macrophages primarily inhabit the airways, contributing to barrier immunity and the clearance of surfactants. These cells rely on specific cytokines such as granulocyte–macrophage colony-stimulating factor (GM-CSF) and transforming growth factor β (TGF- β). Interstitial macrophages are situated within the lung tissue. Studies in mice have identified two distinct subsets of interstitial macrophages: (i) LYVE-1^{high} MHC Class II^{low}, positioned near blood vessels, which participate in wound healing and tissue repair; (ii) LYVE-1^{low} MHC Class II^{high}, located in proximity to neurons, specializing in antigen presentation. Lung macrophages can originate from either embryonic precursors or circulating blood monocytes. **Source:** Evren E. at al., Origin and ontogeny of lung macrophages: from mice to humans. Immunology. 2020.

TRMs have a major role in shaping the tumor microenvironment (TME). The TME includes tumor cells, macrophages, fibroblasts, endothelial cells, pericytes, extracellular matrix (ECM) and soluble signals. Evidence showed that TRMs accumulate significantly in the tumor microenvironment, specifically close to tumor cells [62], [82].

These macrophages, named tumor-associated macrophages (TAMs), are conditioned by tumor cells to assist tumor growth by promoting tumor immune escape, angiogenesis, tumor growth, and metastasis. TAMs mostly show the pro-tumorigenic lineage M2 [85]. Macrophages of every tissue can, in fact, polarize towards various phenotypes, like the M1 and M2 macrophage subtypes (Table 1). For example, the lipopolysaccharide (LPS) from microbial sources can drive macrophage polarization towards the M1 phenotype, in charge of orchestrating the inflammatory response by the secretion of cytokines such as IL-6, IL-12, and tumor necrosis factor (TNF). Interleukin 4 (IL-4) promotes polarization towards the M2 type, that displays anti-inflammatory characteristics and contributes to tissue repair mechanisms. During infection, macrophages initially assume a proinflammatory M1 phenotype to combat pathogens. Subsequently, they transition towards an anti-inflammatory M2 phenotype, facilitating tissue repair processes in the aftermath of infection-induced damage [62].

Table 1. Macrophages polarization markers in humans and mice. While most macrophage markers are conserved between humans and mice, some differences exist. Murine M1 macrophages express Marco, Cxcl9, Cxcl10, Cxcl11, Nos2, and Socs1, and secrete Il-6, Il-12, Il-1 α , and TNF- α , like human M1 macrophages. However, they lack CD86 and CD64 surface markers. Murine M2 macrophages express Cd206, Cd163, TGF- β , Ppar γ , Ccl14, Ccl22, and Arg-1, resembling human M2 macrophages, but uniquely possess Chil3. **Source:** Boutilier AJ, Elsawa SF. Macrophage Polarization States in the Tumor Microenvironment. Int J Mol Sci. 2021

Species	M0	M1	M2
Mouse	Csf1r, F4/80, CD11b	Marco, Cxcl9, Cxcl10, Cxcl11, Nos2, Socs1	Cd206, Tgm2, Fizz1, Chil3, Arg1, Ccl22, Cd163
Human	CSF1R, CD14, CD68, CD11B	CD86, MARCO, CXCL9, CXCL10, CXCL11, NOS2, SOCS1, CD64	TGM2, CD23, ARG1, CCL22, CD163, CD206

It is important to note that macrophage polarization toward M1 and M2 lineage is not a binary polarization but a spectrum instead, with the M1 and M2 lineages being at the opposite extremes. Moreover, M1 macrophages can

switch, at least in part, to M2 and the other way around, according to the microenvironment stimuli [83]. The M1 lineage has the ability to lysing and phagocytosing tumor cells as well as promoting the proliferation of immune cells like CD8⁺ T and NK cells, by the secretion of IL-6, IL-12 and TNF. The M2 lineage, in contrast, assists the tumor growth and prevent the immune response [83].

In the TME, there is the formation of a positive feedback loop between tumor and TAMs: i) Tumor Secretion of Factors. Tumors release various factors such as interleukin-4 (IL-4), interleukin-6 (IL-6), interleukin-10 (IL-10), macrophage colony-stimulating factor (M-CSF or MDF), transforming growth factor-beta 1 (TGF- β 1), and prostaglandin E2 (PGE2). These factors promote the polarization of macrophages towards the M2 phenotype, which is characterized by an immunosuppressive and pro-tumoral phenotype. ii) Induction of Tumor Cell Proliferation and Metastasis. Once polarized towards the M2 phenotype, TAMs secrete various growth factors such as platelet-derived growth factor (PDGF), transforming growth factor-beta (TGF- β), hepatocyte growth factor (HGF), and basic fibroblast growth factor (bFGF). These factors stimulate tumor cell proliferation, survival, and metastasis, thereby contributing to tumor progression. iii) Necrotic Avascular Area and Angiogenesis. In regions of the tumor characterized by necrosis and low oxygen levels (hypoxia), high levels of hypoxia-inducible factor 1-alpha (HIF-1 α) are present. Hypoxia, along with factors like vascular endothelial growth factor (VEGF), tumor necrosis factor-alpha (TNF- α), interleukin-8 (IL-8), basic fibroblast growth factor (bFGF), matrix metalloproteinases (MMPs), and cyclooxygenase-2 (COX-2), promote angiogenesis. This process leads to the formation of new blood vessels, which supply oxygen and nutrients to the tumor, facilitating its growth and metastasis. iv) Impact on Cancer-Associated Fibroblasts (CAFs) and Stroma Remodeling. M2 macrophages also influence the differentiation and function of cancer-associated fibroblasts (CAFs) within the tumor stroma. Additionally, they contribute to the breakdown of bone marrow, participate in collagen degradation and rearrangement, and facilitate the formation of tumor neo-

vessels. This remodeling of the tumor microenvironment enhances tumor invasion capabilities and further supports tumor growth and progression [83]. In conclusion, TAMs, mostly M2-like, drive a tissue remodeling program and promote tumor cell invasiveness during the initial stages of tumor progression, while also promoting a T_{reg} cell response that protects tumor cells from $CD8^+$ T cells. For this evidence, TRMs are identified as a potential target for the prevention or treatment of early lung cancer lesions [82].

1.2.2. Lung cancer epidemiology

In 2018, GLOBOCAN data revealed a global count of 2,094,000 newly diagnosed cases of lung cancer, positioning it as the most prevalent cancer type in terms of incidence worldwide, developing nations having the highest incidence (Figure 5). Among these cases, an estimated 1,369,000 were identified in men (65.4 %), making lung cancer the second most frequent cancer in males, following prostate cancer. Correspondingly, 725,000 cases were reported in women, establishing lung cancer as the second most prevalent cancer among females, following breast cancer. The age-adjusted cumulative lifetime risk for developing lung cancer stands at 3.8% for men and 1.77% for women [20], [73], [86].

Lung cancer is the most diagnosed cancer in men across 37 countries, like Russia, China, and significant portions of Eastern Europe, the Middle East, and Southeast Asia. North Korea and Hungary hold the top positions for lung cancer incidence in men and in women, respectively. North America and Northern and Western Europe display the globe's highest rates. Conversely, the regions of western, central, and eastern Africa demonstrate the lowest incidences for both men and women [20], [73], [86].

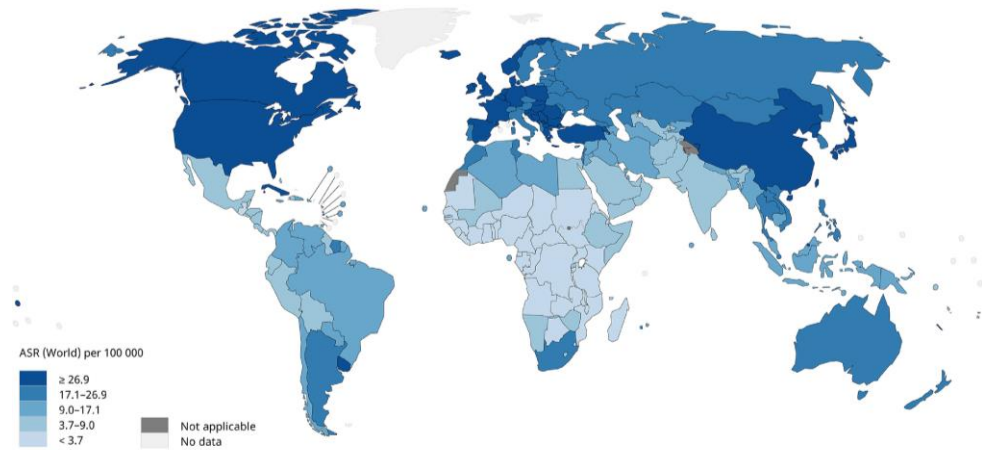


Figure 5. Estimated age-standardized incidence rates (world) in 2020, lung, both sexes, all ages. The map was produced by the International Agency for Research on Cancer (IARC) and World Health Organization (WHO). **Source:** GLOBOCAN

As reported by the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program, the United States witnessed approximately 229,000 fresh instances of lung cancer in 2020, constituting 12.7% of the total cancer diagnoses. The present rate of occurrence, standing at 45.6 cases per 100,000, has decreased from its apex of 69.5 cases per 100,000 back in 1992. This decline can be predominantly attributed to the cessation of smoking habits [86], [87].

Worldwide, lung cancer is the leading cause of cancer death in men and women (Figure 6). According to the SEER database, the most recent 5-year survival rate for lung cancer in the United States, covering the period from 2010 to 2016, stood at 20.5%. This reflects a notable increase from the initial recorded rate of 11.5% back in 1975 [86], [87]. These advancements in survival rates can likely be attributed to enhanced early detection practices, such as the recommended use of computed tomography (CT) screening for individuals with substantial smoking histories. Additionally, the introduction of targeted therapies like EGFR, ALK, ROS, and BRAF inhibitors, along with immune-based treatments like PD-1, PDL-1, and CTLA-4 inhibitors, has contributed to improved outcomes. Progress in survival rates is further evident in the context of disease stage at diagnosis. For individuals in the US diagnosed with localized lung cancer (stage I-II), the 5-year survival rate reaches 59.0%. However, this rate decreases to 31.7% for those with stage III

and significantly drops to 5.8% for those with metastatic disease (stage IV). It is important to note that 57% of lung cancer cases in the US are detected after they have already reached the metastatic stage, emphasizing the ongoing need for early detection and expanded screening initiatives. Furthermore, access to the latest diagnostic and treatment resources tends to be better in higher-income nations, which contributes to the more favorable lung cancer survival rates observed in these regions [86], [87], [88].

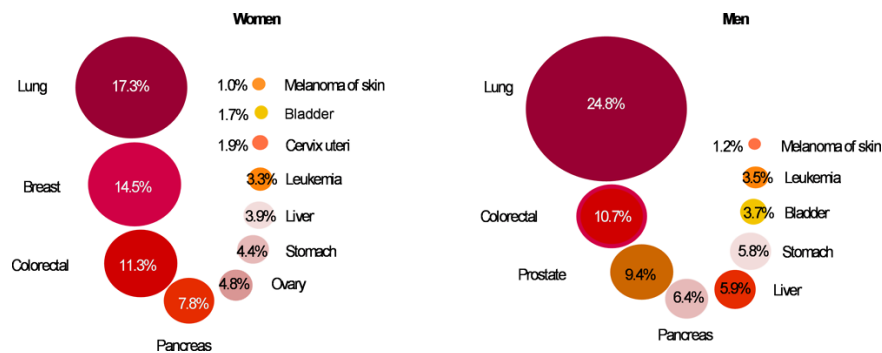


Figure 6. Main causes of cancer mortality across OECD (Organization for Economic Co-operation and Development) countries, by sex, 2017. Proportion of the sums of cancer-related deaths across OECD countries, by sex. Source: OECD Health Statistics 2019.

1.2.3. Risk factors for lung cancer development

Tobacco smoking is the major cause of lung cancer. It contains more than 7000 chemicals, including at least 69 well-established carcinogens and additional harmful substances linked as well to various serious diseases. Even though only approximately 15% of smokers ultimately develop lung cancer, tobacco smoking accounts for 80 to 90% of diagnosed cases of lung cancer in the United States.

The indirect contact with smoke among non-smokers drastically decreased in the last forty years due to tobacco control efforts and smoke-free laws and policies in workplaces and public places. Cancer-causing agents identified in secondhand smoke comprise polycyclic aromatic hydrocarbons, nitrosamines, and aromatic amines. Studies indicate the presence of nicotine and its metabolite cotinine, along with DNA adducts resulting from tobacco carcinogens, in the urine of individuals who have never smoked but are exposed to secondhand smoke.

Propylene glycol and vegetable glycerin are the primary solvents in electronic cigarettes. Research has demonstrated that vapors produced from these solvents contain toxic and carcinogenic carbonyl compounds such as formaldehyde, acetaldehyde, acetone, and acrolein. Additionally, studies indicate that the use of e-cigarettes is linked to an increased level of oxidative stress, potentially leading to inflammation, cytotoxicity, and increased permeability of endothelial cells.

While cigarettes continue to be the predominant form of tobacco consumption, other tobacco products such as pipes, cigars, and water pipes remain prevalent and are linked to elevated risks and mortality rates associated with lung cancer. Lung cancer risk is lower for individuals using these products compared to those who smoke cigarettes.

Smoking of cannabis releases numerous chemical toxins and carcinogens similar to those found in tobacco smoke, like acetaldehyde, acrolein, ammonia, carbon monoxide, formaldehyde, phenols, nitrosamines, and polycyclic aromatic hydrocarbons. Despite this, conclusive evidence establishing a direct association between cannabis smoking and increased incidence of lung cancer is still unclear.

Radon is an invisible, odorless, and tasteless radioactive gas originating from soil (e.g. enriching in open-soil basements), naturally generated during the radioactive decay of thorium and uranium. Everyone, despite significant geographic disparities, is exposed to this gas and 3% to 14% of lung cancer cases are attributed to radon exposure worldwide.

Asbestos is a collective term used to describe naturally occurring mineral silicate fibers, including amphiboles such as crocidolite, amosite, tremolite, anthophyllite, and actinolite, as well as chrysotile. While the precise mechanisms underlying asbestos-related diseases are intricate and not entirely elucidated, it is well-supported that both direct and indirect cellular and molecular effects contribute to the etiology of lung cancer by oxidative stress, chronic inflammation, genetic and epigenetic alterations, as well as cellular toxicity and fibrosis.

Chronic obstructive pulmonary disease (COPD) is a condition which includes emphysema and chronic bronchitis, represents an irreversible chronic

inflammatory state characterized by permanent constriction of small airways and degradation of alveolar walls. The persistent inflammatory response within the bronchi perpetuates a cycle of continual injury and repair, potentially serving as a key factor in the development of lung cancer. Tobacco smoking stands out as the primary risk factor for COPD.

Asthma, a common childhood illness affecting around 300 million individuals globally, is distinguished by persistent lung inflammation leading to symptoms such as airway hyper-reactivity, excessive mucus production, and respiratory blockage. The relationship between asthma and lung cancer risk has been under investigation, as inflammation plays a central role in the pathogenesis of both conditions.

Infectious respiratory diseases can also increase the risk for lung cancer. Pneumococcal disease is a term indicating several syndromes triggered by various organisms, leading to diverse manifestations and outcomes. The primary cause of pneumococcal disease is typically the *Streptococcus pneumoniae* bacterium, which can result in infections such as pneumonia (lung infection), bacteremia (bloodstream infection), and meningitis (infection of the tissues and fluids surrounding the brain and spinal cord). Pneumonia is considered a potential risk factor for lung cancer, possibly through multiple mechanisms involving chronic local inflammation mediators such as elevated reactive oxygen species, which can induce DNA damage and somatic mutations, anti-apoptotic signaling, and increased angiogenesis. *Chlamydia pneumoniae* is the predominant intracellular bacterial pathogen, commonly causing sinusitis, pharyngitis, and pneumonia. Transmission typically occurs through respiratory secretions. There is evidence that *Chlamydia pneumoniae* infection might increase the risk of lung cancer via inflammatory mediators, similarly to those implicated in pneumonia infection.

Individuals infected with HIV face an elevated risk of developing numerous cancers. The reasons are attributable to HIV-related immunosuppression, which compromises the body's ability to control oncogenic viral infections, as well as inflammatory mediators and co-infection with oncogenic viruses like Hepatitis B and C (Figure 8) [89].

Individuals with a family history of lung cancer are at a higher risk, suggesting a potential genetic predisposition. Additionally, specific genetic mutations, such as those in the EGFR gene, have been linked to an increased susceptibility to lung cancer.

Lung cancer in never-smokers is diagnosed most frequently in women suggesting a potential influence of hormones. Estrogen receptors (ER α and β) are found in both normal lung tissue and tumors. Studies indicate higher expression of ER- β in women, never smokers, and certain adenocarcinomas. Activation of ERs can enhance the expression of genes related to cell proliferation and tumor growth. ER expression is linked to prognosis, with ER- α indicating poorer prognosis and ER- β associated with a better outlook [90].

Table 2. Established and putative risk factors for lung cancer. A summary of causative and putative lung cancer risk factors. **Source:** Schabath MB, Cote ML. Cancer Progress and Priorities: Lung Cancer. Cancer Epidemiol Biomarkers Prev. 2019.

Risk factor	Magnitude of association
Tobacco smoking	20-fold increased risk vs. never smoker
Secondhand smoke	25% to 28% increased risk vs. never smoker
Electronic cigarettes	Presently unknown
Other tobacco use (cigars, pipes, water pipes)	1.9 to 4.6-fold increased risk
Smoked cannabis	Presently no known risk
Radon	14% to 29% increased risk
Asbestos	12% to 24% increased risk
History of COPD, emphysema, or chronic bronchitis	2 to 3-fold increased risk
History of asthma	28% to 44% increased risk
History of pneumonia	30% to 57% increased risk
History of <i>Chlamydia pneumoniae</i>	1.2 to 2.4-fold increased risk
History of tuberculosis	48% to 76% increased risk
HIV	2-fold increased risk

1.2.4. ROS and inflammation in lung cancer development

Oxygen is essential for the metabolic pathways of all aerobic organisms. Reactive Oxygen Species (ROS) as well as Reactive Nitrogen Species (RNS) are normally formed as a by-product and they are necessary for several cellular events like signal transduction, enzyme activation and gene

expression [91], [92], [93]. ROS and RNS in low concentration have a role in:

- 1) activation of transcription factors such as nuclear factor kappa B (NF- κ B), activator protein-1 (AP-1).
- 2) induction of apoptosis or necrosis.
- 3) altering mitogenic signals that affect cell growth and cell signaling cascades [94].

ROS are the most abundant products of cell metabolism. Examples of ROS include compounds like superoxide anion ($O_2^{\cdot-}$), hydrogen peroxide (H_2O_2), hydroxyl radical ($HO\cdot$), singlet oxygen ($^1O_2$) and ozone (O_3). They are mainly produced by mitochondria, cytochrome p450 and peroxisomes by electron transfer reactions through enzymatic and non-enzymatic processes [93], [95]. Although they act as intracellular signaling molecules in many biological processes, free radicals and ROS at higher concentrations are dangerous for living organisms and can cause oxidative damage to all major cellular constituents (membrane lipids, proteins, enzymes, DNA), leading to inflammation and cancer [93], [94].

Any type of lung injury (biological, physical or chemical) leads to an inflammatory response in order to initiate a protective mechanism to reestablish the physiological healthy condition [94]. Inflammation can be categorized as either acute or chronic, depending on various factors such as clinical symptoms and the nature of the injury. Acute inflammation is the immediate response that typically lasts for a short duration. PMNs (polymorphonuclear neutrophils) are the first cells to be recruited, followed by monocytes that differentiate in macrophages in the tissues. The latter are involved in ROS, growth factors and cytokines production which could lead to disorganization of lung microenvironment. Chronic inflammation originates from a failure to regulate acute inflammation and it is characterized by persistent inflammation and tissue damage that could also lead to cancer development [94], [96], [97].

The functional role of chronic inflammation in cancer is not novel. Markedly increased proliferation potential of cells is achieved in an environment that is rich in inflammatory cells, producing abundant ROS/RNS promoting

unremitting DNA damage, inactivation of apoptosis, upregulation of growth factors, cytokines, and activation of growth supporting genes [98].

The development of lung tumors is a complex process that involves several stages, including initiation, promotion, and progression. The initiation stage involves DNA damage by, for example, genotoxic carcinogens. Inflammation in the lungs can also have indirect effects through the activation of enzymes that produce ROS, leading to DNA and protein damage. If DNA damage is not repaired, it can result in point mutations which are common in adenocarcinoma cases. These mutations produce oncogenes with dominant gain-of-function and tumor suppressor genes with recessive loss-of-function. Chromosomal abnormalities, DNA adduct formation, methylation, and acetylation are also characteristic of lung cancer [94].

The second stage of cancer development is known as promotion, where initiated cells undergo clonal expansion under persistent oxidative stress conditions, resulting in the formation of focal lesions that may eventually progress to invasive cancers. The final stage, progression, involves the formation of completely malignant cells from early neoplastic clones through genetic and epigenetic mechanisms, resulting in uncontrolled autonomous proliferation. Epidermal growth factor receptor (EGFR) plays a significant role in promoting abnormal cell proliferation in lung tumors, with various alterations such as mutations, gene amplification, overexpression of receptor ligands, and regulatory mechanism loss or gain. In squamous-cell carcinomas of the lung, EGFR is overexpressed in up to 80% of cases. Genome instability is also crucial in cancer initiation and progression, with microsatellite instability in some lung cancers resulting from a lack of mismatch at the nucleotide level. Inflammation-induced oxidative stress, hyperoxia, radiation, particulate exposure, and chemical carcinogens are all significant sources of endogenous and exogenous oxidative stress in lung carcinogenesis [94].

Inflammation can be triggered by infection or cellular damage, causing the release and activation of certain molecules that induce the recruitment of leukocytes to the affected area. The recruited leukocytes then produce reactive oxygen and nitrogen species (ROS/RNS) that can kill invading organisms, but excessive production of these molecules can also lead to tissue

damage and chronic inflammation. The accumulation of p53, which is stimulated by local ROS/RNS production, can inhibit the apoptotic mechanisms that normally end the acute phase of inflammation and transition it into the chronic phase. When cell division is not balanced with cell repair in damaged cells, it increases the chances of inducing mutations, which can lead to cancer. Exposure to activated phagocytes can cause oxidative damage to DNA, proteins, and lipids, which increases the risk of mutagenesis and the development of cancer. ROS/RNS produced by inflammatory cells also stimulate oncogenes, such as jun and fos, which are directly associated with lung cancer. Oxidative modification of lipids induces mutagenesis by forming genotoxic by-products that react with DNA, while protein oxidation can indirectly promote mutagenesis by affecting DNA repair enzymes [94].

1.2.5. Pathology of lung cancer

In general, lung carcinoma is often associated with the presence of pulmonary nodules (< 3 cm) at an early stage. However, it is important to note that lung nodules are not always indicative of cancer, and most nodules are benign and results of infections, scar tissue, or other causes. Following the detection of a lung nodule, the typical course of action involves scheduling a repeat CT scan to monitor any changes in size over time. If subsequent scans reveal growth or if the nodule presents other concerning characteristics, a biopsy will be conducted. If cancer is diagnosed, additional pathological tests will be conducted to determine the specific type of cancer [99].

Small cell lung cancer and non-small cell lung cancer are the two main types of lung cancer (Figure 7). A third less common type of lung cancer is called carcinoid [100].

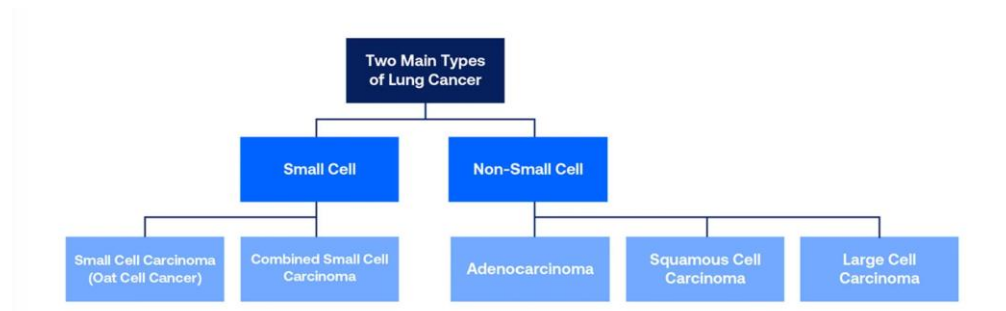


Figure 7. Pathological differentiation of lung cancer. The two main subtypes of lung cancer, small-cell lung cancer and non-small cell lung cancer, can be respectively divided into small-cell carcinoma, combined small-cell carcinoma and adenocarcinoma, squamous cell carcinoma and large cell carcinoma, based on histological features. **Source:** American Lung Association.

Small-cell lung cancer (SCLC) (Figures 7, 8) is a high-grade neuroendocrine carcinoma primarily found in current or former smokers, and it is associated with an exceptionally poor prognosis. SCLC constitutes approximately 15% of all lung cancer cases. Patients with SCLC typically exhibit respiratory symptoms such as cough, dyspnea (difficulty breathing), or hemoptysis (coughing up blood), and imaging often reveals a centrally located lung mass along with extensive involvement of thoracic lymph nodes. About two-thirds of patients have metastatic disease at the time of initial diagnosis, with common sites of metastasis including the contralateral lung, brain, liver, adrenal glands, and bone. Reflecting its high metastatic propensity, SCLC exhibits one of the highest concentrations of circulating tumor cells (CTCs) among all solid tumors. In most SCLC cases, there is concurrent inactivation of two tumor suppressors, p53 and RB (encoded by TP53 and RB1, respectively). This simultaneous inactivation of tumor suppressors differs from the primary oncogenic drivers observed in many other solid tumors, particularly non-small-cell lung cancers (NSCLC), where activating oncogenic mutations appear to be crucial for tumorigenesis. The amplification of MYC family genes (MYC, MYCL, and MYCN) in a subset of SCLC tumors was also reported in recent studies. According to the WHO pathological classification, SCLC can be divided in two subtypes: (pure) SCLC, which accounts for approximately 80% of cases, and combined SCLC, which comprises around 20% of cases. Combined SCLC contains a mixture

of small cell carcinoma cells and cells of another histological type, such as adenocarcinoma, squamous cell carcinoma, or large cell carcinoma [101].

The name non-small cell lung cancer (NSCLC) (Figures 7, 8) refers generally to three types of lung cancers: adenocarcinoma (LUAD), squamous cell carcinoma (LUSC), and large cell carcinoma. LUAD, constituting half of all lung cancer cases, is the most prevalent subtype within NSCLC. It is frequently located in the outer regions of the lung and it arises from epithelial cells, which line body cavities and surfaces and form glands. LUSC, previously the most common lung cancer type, typically originates at the tracheobronchial tree but is increasingly diagnosed in the lung periphery. Large cell carcinoma is diagnosed when other subtypes are excluded. It is poorly differentiated and cannot be further classified by immunohistochemistry or electron microscopy, although 90% of cases display squamous, glandular, or neuroendocrine features. It can occur in any part of the lung and tends to grow and spread faster than adenocarcinoma or squamous cell carcinoma.

Additionally, NSCLC includes other subsets of lung cancer with diverse categories and terminology, such as adenosquamous carcinoma, sarcomatoid carcinoma, and non-small cell neuroendocrine tumors. The clinical manifestations of NSCLC may include cough, hemoptysis, chest pain, dyspnea, or hoarseness. LUSC can lead to Pancoast Syndrome, characterized by shoulder pain (which may extend to the forearm, scapula, or fingers), Horner Syndrome, atrophy of hand muscles, or bony destruction. Approximately 20% of NSCLC cases show bone metastases initially. LUAD, on the other hand, has a higher frequency of brain metastases, which may manifest as headache, vomiting, visual field deficits, seizures, or focal neurologic deficits. [102].

Common mutations in NSCLC are found in: EGFR, KRAS, FGFR1, ALK, MET, PIK3CA, BRAF, ROS1, NTRK, RET and HER2. The frequency of mutations in EGFR, MET, ROS1 and HER2 are significantly higher in nonsmokers than in smokers [103].

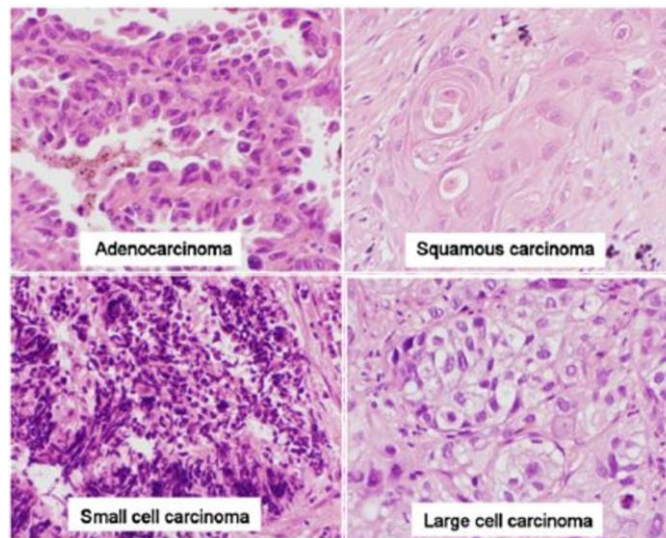


Figure 8. Histological profile of carcinomas of the lung and pleura. Lung adenocarcinoma is characterized by glandular or acinar structures with cuboidal or columnar cells. It often shows mucin production and may have lepidic growth along the alveolar walls. Squamous Cell Carcinoma is composed of stratified layers of squamous epithelial cells. Typically, it exhibits keratinization and intercellular bridges. Small Cell Carcinoma features small, round cells with high nuclear/cytoplasmic ratio and scant cytoplasm. It often shows extensive necrosis and frequent mitotic figures. Large Cell Carcinoma comprises large, pleomorphic cells with prominent nucleoli and abundant cytoplasm. It lacks distinctive histological features seen in other types. Stain: Haematoxylin and Eosin. Magnification x20 for all images. **Source:** adapted from L. R. Motadi et al., Molecular genetics and mechanisms of apoptosis in carcinomas of the lung and pleura: Therapeutic targets. International Immunopharmacology. 2008.

1.2.6. Advances and limitations in lung cancer screening and therapy

Despite considerable advancements in diagnosing and treating lung cancer, progress in improving survival rates has been slow. Lung cancer care now emphasizes personalized treatment, alongside notable progress in immunotherapy, radiotherapy, and surgical techniques. However, survival rates remain low, largely because most lung cancer cases are diagnosed at an advanced stage. The five-year survival rate stands at 13%, with significant discrepancies observed across different regions of Europe. This prognosis is primarily linked to the delayed detection of the disease, as most cases are identified at an advanced stage, while only a small percentage are caught in early stage [104].

Screening for lung cancer involves imaging tests to detect the presence of the disease before symptoms manifest. The primary aim of lung cancer screening

is to identify the condition at an early stage when treatment options are more effective. In fact, if lung cancer is detected at an early stage, 68% to 92% of people can survive more than five years [104].

In the past, chest X-rays (CXR) were commonly used for lung cancer screening and they are still the first-line investigation for lung cancer in several healthcare systems. However, studies have shown that CXR has limited sensitivity for detecting early-stage lung cancer, leading to many cases being missing. [105].

Sputum cytology is another dated method involving the analysis of sputum samples under a microscope to look for cancer cells [106].

Low-dose computed tomography (LDCT) is the current prevailing method utilized for lung cancer screening. A low-dose CT scan utilizes a specialized X-ray technique, capturing multiple images while the individual lies on a movable table within the machine. Subsequently, a computer integrates these images to generate a comprehensive depiction of the lungs [100]. LDCT is a non-invasive procedure, the screening test is painless, typically lasting less than 10 minutes and it is relatively cost-effective compared, making it a more accessible option for patients. Negative aspects include the detection of false positives. Non-cancerous nodules or tumors may be detected, leading to unnecessary invasive procedures like biopsies, bronchoscopy or lung surgery [104].

Bronchoscopy is a procedure used to examine the airways leading to the lungs. While it is not typically used as a primary screening tool for lung cancer, it plays a crucial role in the diagnosis and staging of lung cancer, as well as in evaluating other lung conditions.

Bronchoscopy is often performed when there are symptoms or findings suggestive of lung cancer, such as persistent cough, hemoptysis (coughing up blood), or abnormal lung nodules detected on imaging studies like chest X-ray or LDCT scan [107].

If lung cancer is diagnosed, additional tests are conducted to determine its extent of spread within the lungs, lymph nodes, and the rest of the body. This

procedure is referred to as staging. Treatment for lung cancer varies depending on the type of lung cancer and its stage of progression [100].

Surgery is an option for some patients depending on the type, location and stage of their lung cancer and other medical conditions. Removing the tumor is considered the best option when the cancer is localized and unlikely to have spread [100].

Chemotherapy uses cytotoxic drugs to destroy cancer cells. The drugs circulate throughout the body in the bloodstream. It can be prescribed before or after surgery: i) neoadjuvant chemotherapy is given before surgery with the aim of shrinking the tumor. ii) Adjuvant chemotherapy is given after surgery to kill any remaining cancer cells that may not have been removed during surgery or may have spread to nearby tissues [107].

Chemotherapy is the primary treatment for SCLC due to its effectiveness. Given that SCLC is typically diagnosed after it has spread with metastasis beyond the lung, chemotherapy is particularly crucial in addressing the spread of the disease. [107].

Radiotherapy technologies are advancing rapidly, leading to more precise and expedited treatments with reduced side effects. For example, new technologies involve exceptionally high doses of radiation precisely targeting the tumor minimizing radiation exposure to surrounding healthy tissues and organs [100], [108]. Another method for radiotherapy treatment uses a radiation source that is placed inside the body, enabling a concentrated dose of radiation to target the specific area where the tumor is located or was situated before surgery [109]. Side effects of radiotherapy may vary based on where radiation is located. Common side effects include skin reactions, nausea, tiredness, painful swallowing [100].

Immunotherapy exhibits promising efficacy regardless of lung cancer histology and driver mutations, often leading to sustained remission. The goal is to stimulate an immune response, particularly T-cell-mediated cytotoxicity, against tumor-specific antigens enabling selective tumor destruction. Unfortunately, in recent years, immunotherapy has faced challenges in the context of lung cancer due to limited immune responses. Lung cancer cells can, in fact, evade immunosurveillance, and decrease the general

immunological responses by modulating T-cell-mediated cytotoxicity, secreting immune-suppressive cytokines, and downregulating major histocompatibility complex expression. Nonetheless, recent technological advancements led to the evolution of therapeutic vaccines, immune modulators, autologous cellular therapies, and monoclonal antibodies targeting checkpoint inhibitor signals associated with activated T-cells and/or cancer cells [110].

Targeted therapy is a treatment option for patients with specific alterations detected in their tumors through biomarker testing, like mutations, additions, deletions, or rearrangements. The treatment directly addresses these mutations present in cancer cells rather than healthy cells, resulting in fewer side effects. Based on the genetics of the tumor, not all individuals are eligible for targeted therapies for lung cancer [100].

2. Aims of the Study

Infectious pathogens play an important role in cancer development. According to de Martel et al. [111], the year 2018 witnessed 2.2 million cancer diagnoses linked to infections, accounting for approximately 14-16% of the overall cancer incidence. While up to 90% of lung cancer cases are attributed to smoking and tobacco use, there are also cases in non-smokers and where diffuse states of chronic inflammation unrelated to smoking seems to play a significant role in tumor development [66]. Moreover, in the last decades, there have been a trend for an increase of lung cancer cases in young adults who never smoked (mostly women) [112].

Given the increased incidence of lung cancer observed in butchers and animal farmers, with the evidence linking lower lung cancer risk to reduced consumption of milk and dairy products, it is likely that additional infectious and/or dietary elements such as BMMF might also contribute to lung cancer. So far, BMMF1 genomes and proteins have been analyzed in colorectal cancer tissues [30], [63]; supporting the model of BMMF-induced indirect carcinogenesis upon chronic inflammation. However, additional exploration is required in other types of tumors associated with inflammation-derived carcinogenesis.

This PhD project represents the first combined wet lab and multi-omics study of BMMF1 detection in lung cancer, aiming to understand their putative role and contribution in associated chronic inflammation processes.

If the assumption of a contribution of BMMF to lung cancer is true, BMMF should be identified in lung cancer tissues on protein level (IHC, WB etc.) and/or on nucleic acid level (qRT-PCR, sequencing). If BMMF is involved in the initiation of a chronic inflammatory state, they should be associated with specific immune cell populations.

To study these assumptions, in this study, I conducted multi-omics BMMF-specific characterization of lung cancer and age-matched controls tissues by application of systematic immunohistochemistry (IHC) tissue staining including also new anti-BMMF antibodies on a broader set of samples to study BMMF as possible lung cancer biomarker. This could be the starting point for the development of non-invasive early diagnostic tests, such as

serum-based assays for protein or antibody markers, ultimately contributing to improving patient outcomes and therapeutic interventions.

Research objectives:

- (I) Immunohistochemical characterization of BMMF Rep expression in lung cancer tissues based on IHC (immunoblotting, mass spectroscopy) to validate general detection of BMMF (Rep) protein as possible risk factor in lung cancer.
- (II) Study of the association between BMMF and cancer-promoting immune cell populations, which might fuel specific patterns of tumor-promoting chronic inflammation by multi-channel IF and *in situ* hybridization.
- (III) Tumor-peritumor and case-control comparison of BMMF detection to confirm BMMF (Rep) protein detection as putative biomarker by histochemical and cell-based quantification methods.
- (IV) Investigating BMMF's mode of action towards indirect carcinogenesis by assessing e.g. BMMF's possible impact on macrophage polarization toward the "M2-like" lineage by IHC and scRNA-seq/bulk RNA-seq.

3. Materials and Methods

3.1. Materials

3.1.1. Cohort description

We obtained the lung tissue samples used in this analysis from ThoraxKlinik (NCT) Heidelberg. They are grouped into three projects (Table 3).

- Project 2750 includes FFPE sections from 21 lung cancer patients (paired tumor and peritumor tissues) and 2 FFPE sections of unpaired cancer tissues.
- Project 3359 includes FFPE sections as well as cryosections. For this project, 6 positive lung cancer patients (paired tumor and peritumor tissues) were chosen between the patients of project 2750 to then obtain more FFPE and frozen tissues to deepen the analysis. To include case vs. control analysis, FFPE and cryosections from 5 non-cancer individuals were also obtained within this project. Specifically, we used (tissues adjacent) to hamartoma as controls. Pulmonary hamartoma is a benign lung tumor; usually asymptomatic and usually not linked to inflammation. It contains mainly avascular cartilage, but it could also contain some bronchial epithelium, which has a vascular supply and connection to some bronchi. For this reason, it may get inflamed, though this is not a common event [113].
- Project 3694 includes FFPE and cryosections from 14 lung cancer patients (peritumor and tumor tissues paired) and from 14 non-cancer individuals (hamartoma patients).

Moreover, lung cancer fresh peritumor tissues were also obtained from ThoraxKlinik (NCT) Heidelberg after surgeries of 8 patients. Clinical information of patients is summarized in Table 4.

Table 3. Overview of cohorts and methods used for the analysis. N/A indicates that the analysis was not applicable for the specific cohort. ✓ indicates that the analysis was done. X indicates that the analysis was not performed. Preliminary indicates that just one sample of the cohort was used for the analysis.

		DAB staining	IF staining	Cell-based quant.	WB	RT-qPCR	FACS	ISH
NCT Project 2750								
21x peritumor 21x (+2) tumor	Paired	FFPE	✓	✓	✓	N/A	N/A	N/A
NCT Project 3359								
6x peritumor 6x tumor 5x hamartoma	Paired	FFPE + Cryosections	✓	✓	✓	✓	N/A	X
NCT Project 3694								
14x peritumor 14x tumor 14x hamartoma	Paired	FFPE + Cryosections	✓	✓	✓	✓	N/A	X
Fresh tissues after surgeries								
8x peritumor		✓	X	X	✓	✓	✓	Preliminary

Table 4. Clinical information of patients included in the analysis. Patients are grouped by colour code in the different ThoraxKlinik (NCT) project. NA indicates that the information is not available. / indicates that the information is not applicable in that specific contest.

Pseudonym	Sex	Age	Histology	Smoking Status	PY
IIX6KF	Female	78	NSCLC	Never	/
16PS7Y	Female	51	NSCLC	Smoker/Ex-Smoker?	NA
1LYLZF	Male	63	NSCLC	Ex-Smoker	80
1FAVEM	Male	65	NSCLC	NA	NA
1DVRTK	Male	71	NSCLC	Ex-Smoker	10
ORJWVD	Female	52	SCLC	Smoker	60
1PZHGD	Male	69	SCLC	Ex-Smoker	70
14CIFM	Female	49	NSCLC	Ex-Smoker	50
UJEQ8N	Male	61	NSCLC	Smoker	80
OVHCKC	Male	70	SCLC	Ex-Smoker	30
1IK3RV	Male	77	NSCLC	Never	/
HHU68Y	Female	64	NSCLC	Ex-Smoker	90
NK1IKR	Male	65	NSCLC	Smoker	45
6LBF64	Male	63	SCLC	Ex-Smoker	40
XNB56E	Female	50	SCLC	Smoker	70
1SRX8R	Male	63	SCLC	Ex-Smoker	25
ZCZ8WH	Male	69	NSCLC	Never	/
MO57SD	Male	78	NSCLC	Ex-Smoker	45
O6QVVN	Male	68	LCNEC	Ex-Smoker	NA
1NZV3L	Female	64	LANED	Ex-Smoker	NA
2V6GGQ	Male	45	LCNEC	Smoker	10
12SI9R	Female	56	NSCLC	Smoker	40
DTN9K4	NA	NA	NA	NA	NA
9DFQ5K	Male	55	Hamartoma	Smoker	45
13DRL6	Female	47	Hamartoma	Ex-Smoker	5
WPPSS6	Female	61	Hamartoma	Never	/
123UWA	Male	59	Hamartoma	Ex-Smoker	5
1CVVKA	Female	67	Hamartoma	Never	/

1RGUL5	Female	71	Hamartoma	Never	/
IUDT9Z	Male	60	Hamartoma	Smoker	40
4E3TIC	Female	76	Hamartoma	Ex-Smoker	20
1A3RXW	Female	57	Hamartoma	Ex-Smoker	30
1XFDVM	Male	48	Hamartoma	NA	NA
TXZBEE	Male	44	Hamartoma	Ex-Smoker	10
NC2CFC	Male	60	Hamartoma	Ex-Smoker	30
10X1PE	Male	48	Hamartoma	Ex-Smoker	30
1G944Y	Male	55	Hamartoma	Smoker	80
J0ZIYM	Male	56	Hamartoma	Smoker	15
9SSVPV	Male	59	Hamartoma	Ex-Smoker	40
3N4U9G	Male	68	Hamartoma	NA	NA
1OE7G1	Female	78	Hamartoma	Ex-Smoker	5
PJ841U	Male	51	Hamartoma	Ex-Smoker	45
11F3TZ	Female	68	NSCLC-SCC	Ex-Smoker	60
1W2LHA	Male	62	NSCLC-AD	Smoker	42
RGN43J	Male	54	NSCLC-SCC	Ex-Smoker	30
1TA5ZI	Female	76	NSCLC-SCC	NA	NA
JGS1F3	Male	72	NSCLC-SCC	Ex-Smoker	20
1GCLAF	Female	80	NSCLC-AD	Never	/
57D3UO	Male	62	NSCLC-AD	Smoker	NA
B3X2KD	Male	60	NSCLC-AD	Never	/
TJHWQ0	Female	75	NSCLC-AD	Never	/
14ZOKY	Female	74	NSCLC-AD	Never	/
4B8HD0	Male	74	NSCLC-AD	Smoker	15
E6B6PL	Male	64	NSCLC-AD	Ex-Smoker	20
NKPU2X	Female	80	NSCLC-AD	NA	NA
HC87BC	Male	50	NSCLC-AD	Ex-Smoker	15
15183	Female	59	NSCLC-AD	Smoker	40
15184	Male	65	NSCLC-AD	Smoker	50
15188	Male	72	NSCLC-AD	Smoker	30
15192	Female	62	NSCLC-AD	Smoker	50
15193	Female	75	NSCLC-AD	Never	/
15194	Female	58	NSCLC-AD	Smoker	50
15195	Female	51	NSCLC-AD	Ex-Smoker	20
15196	Male	68	NSCLC	Ex-Smoker	25

NCT – Thorax Klinik Project 2750

NCT – Thorax Klinik Project 3359

NCT – Thorax Klinik Project 3694

NCT – Thorax Klinik Fresh Tissues

3.1.2. Lists of reagents, buffers, devices and software

Table 5. Antibodies

Antibodies	Company	Identifier
Alexa 488 goat-Anti Rabbit IgG	Invitrogen	A11034
Alexa 594 Goat -Anti Mouse highly cross adsorbed	Invitrogen	A11032
Alexa 647 donkey-Anti Mouse IgG H&L	Invitrogen	A31571
Anti-CD163 antibody - monoclonal mouse	Novusbio	NB110-40686
Anti-CD68 antibody - monoclonal rabbit	Cell signaling	76437
Anti- β -Actin - polyclonal rabbit	Invitrogen	PA5-85271
APC anti-human CD48 - mouse IgG2b	Bio Legend	333809
BMMF1 Rep antibodies - mouse, rabbit	Work of Dr. Timo Bund	
FITC anti-human CD326 (Ep-CAM) - mouse IgG2b	Bio Legend	324203
Mouse IgG1 Isotype Control (MOPC-21)	Invitrogen	MA110406
PE anti-human CD45 - mouse IgG1	Bio Legend	304007
PE/Cyanine7 anti-human CD3 - mouse IgG1	Bio Legend	344815
Peroxidase AffiniPure Goat Anti-Mouse IgG (H+L) - goat	Jackson ImmunoResearch	115-035-062
Peroxidase AffiniPure Goat Anti-Rabbit IgG (H+L) - goat	Jackson ImmunoResearch	111-035-144

Table 6. Probes and primers

Probe	Company	Target region
RNAscope® Negative Control Probe – DapB	ACDbio	/
RNAscope® Positive Control Probe – PPIB	ACDbio	/
RNAscope® HCB14.296	ACDbio - Designed by Sumen Siqin	331-1606 bp
RNAscope® MSB12.176	ACDbio - Designed by Sumen Siqin	580-1601 bp
RNAscope® C1MI.3M.1	ACDbio - Designed by Sumen Siqin	273-1352 bp
RNAscope® MSB11.176	ACDbio - Designed by Sumen Siqin	461-1565 bp
Primers	Company	Target region
CMI5.240 Fw172	Sigma-Aldrich - Designed by Sumen Siqin	GAACATAGAGCGAGCCGTGT
CMI5.240 Rev172	Sigma-Aldrich - Designed by Sumen Siqin	AAACCACGCCAAACTCCTCC
GAPDH Fw	Sigma-Aldrich	GAAGGTGAAGGTCGGAGTC
GAPDH Rev	Sigma-Aldrich	GAAGATGGTGATGGGATTTC

Table 7. Chemicals, enzymes, kits, equipment

Name	Company	Identifier
10x Tris/Glycine/SDS buffer	Bio-rad	1610772
2-Mercaptoethanol	Merck	S1863906
AllPrep DNA/RNA/Protein mini kit	Qiagen	80004
Blocking solution (IF)	Zytomed Systems	ZUC007-100
Clarity western ECL substrate	Bio-rad	170-5060
DAB Substrate Kit High Contrast (BF)	Zytomed Systems	DAB500plus
Dako pen (IF)	Agilent	S200230-2
DAPI (IF)	Appllichem	A1001,0010
Dnase I recombinant, Rnase-free	Roche	04 716 728 001
ECL select WB detection reagent	Cytiva	RPN2235
EDTA Buffer ph 8,5 10x Antigen Retrieval	Sigma-Aldrich	E1161
Einbettmedium Prolong Gold (IF)	Invitrogen	P36930
FACS buffer (1% BSA, 5% FBS, 5mM EDTA, 0.1% sodium azide in 1X PBS)	Self-prepared	
FcR blocking reagent human	Miltenyi Biotec	130-059-901
Hematoxylin, Mayer's (Lillie's Modification)	Zytomed Systems	ZUC087-500
Histogel (BF)	Histoprime	E6000
Lämmli buffer (250mM Tris/HCl pH6.8, 8% SDS, 30% glycerol, 0.2% bromophenol blue)	Self-prepared	
MACS® Tissue Storage Solution	Miltenyi Biotec	130-100-008
Nail Polish Quick & Shine (IF)	Astor	

Novex 4-20%, Tris-Glycine Plus WedgeWell 1.0mm X 12	Invitrogen	XP04202BOX
Novex 4-20%, Tris-Glycine Plus WedgeWell 1.0mm X 15	Invitrogen	XP04205BOX
OneComp eBeads	Invitrogen	01-1111-42
Page Ruler Prestained Protein Ladder	Thermo Scientific	26617
PBS pH7.2 10X	Gibco	70013-016
Peroxide Block	Zytomed Systems	ZUC019-500
PhosSTOP	Roche	04 906 845 001
Ponceau S solution	Genaxxon Bioscience	M3350.0100
PowerUp SYBER Green master mix	Applied Biosystem	A25742
Protease inhibitor cocktail	Sigma-Aldrich	P8340-5ml
Protein A/G PLUS-Agarose	Santa Cruz	sc-2003
Protein assay	Bio-rad	500-0006
Restore PLUS western blot stripping buffer	Thermo Scientific	46430
RevertAid H Minus First Strand cDNA Synthesis Kit	Thermo Scientific	K1632
RIPA buffer (1M Tris/HCl pH8.0, 5M NaCl, 0.5M EDTA, 1% NP40, 20% SDS)	Self-prepared	
RNAscope® 2.0 HD Detection Kit – BROWN	ACDbio	322300
RNeasy plus micro kit (50)	Qiagen	74034
ROTI®Histofix 4,5%	Roth	2213
Skim milk powder	Gerbu	1602.05
Trans-blot turbo mini-size nitrocellulose	Bio-rad	L002049 B
Trans-blot turbo mini-size transfer stacks	Bio-rad	L002043 B
Tumor Dissociation Kit, human	Miltenyi biotec	130-095-929
Tween 20	Sigma-Aldrich	P1379-500ml
Zombie Aqua dye	Bio Legend	77143
Zytomed Chem-Plus HRP Polymer-Kit	Zytomed Systems	POLHRP-100

Table 8. Devices

Device	Name	Company
Automated IHC stainer	BOND MAX	Leica
Benchtop centrifuge	Centrifuge 5424	Eppendorf
Centrifuge	2K15C	Sigma
Dissociator	GentleMACS Octo Dissociator with Heaters	Miltenyi Biotec
Electrophoresis chamber	XCell SureLock	Thermo Fisher
Electrophoresis power supply	PowerPac 300	Bio-Rad
Fluorometer	Qubit 4	Invitrogen
Imaging system	Chemidoc Touch Imaging System	Bio-Rad
Laboratory oven	/	Bachofer
Microtome	Rotationsmikrotom	Microm
Real-Time PCR system	QuantStudio 5	Thermo Fisher
Shacker	Duomax 2030	Heidolph
Slide scanner	Nano Zoomer S60	Hamamatsu
Spectrophotometer	NanoDrop2000	Thermo Fisher
Spectrophotometer	Ultrospec 3000 pro	Pharmacia
Thermocycler	Flex Cycler2	Analytik Jena
Tissue homogenizer	Precellys Evolution	Bertin
Transfer blot System	TransBlot Turbo Transfer System	Bio-Rad

Table 9. Softwares

Software	Version/Link
BLAST (NCBI)	https://blast.ncbi.nlm.nih.gov/Blast.cgi
GraphPad Prism	9
ImageJ/Fiji	1.53T
ImageLab	4.2.1
NDP	Hamamatsu v2.7.25+
QuantStudio™ Design & Analysis Software	Thermo Fisher v1.4.3
RStudio	4.2.2 (2022-10-31 ucrt)
SnapGene	v6.0.2

3.2. Methods

3.2.1. FFPE sections preparation

We received fresh lung tissues after surgery from ThoraxKlinik (Heidelberg) in MACS tissue storage solution and they were cut into small pieces for downstream analysis (western blot, immunohistochemistry DAB staining, quantitative real time PCR, FACS sorting, *in situ* hybridization). With the help of Claudia Ernst (TA, D460/F200), we fixed tissue pieces for IHC DAB and ISH in 4.5% Histofix solution and then embedded them in paraffin using the STP 420 D dehydration/infiltration unit and the HistoCore Arcadia H embedding station (Leica). This process involved stepwise dehydration through a graded alcohol series, followed by transfer to xylene, and then paraffin infiltration. After embedding, tissues were cut with a microtome into 3 μm (IHC DAB) or 5 μm (ISH) thick sections. Cuts were placed on Superfrost Plus slides and they were dried overnight at 60° in a laboratory oven.

3.2.2. Immunohistochemistry staining

I performed the immunohistochemistry staining together with Claudia Ernst (TA, D460/F200). We obtained FFPE tissues sections with thickness of 3-5 μm from ThoraxKlinik, or we prepared FFPE tissue sections on Superfrost Plus slides from fresh tissues. These sections were then subjected to:

- Brightfield DAB (diaminobenzidine) staining, testing up to 8 Rep specific antibodies and anti-CD68 antibody.
- Immunofluorescence (IF) triple staining with anti-Rep, anti-CD68 and anti-CD163 antibodies.

We performed epitope retrieval by heating the sections in EDTA buffer at 90°C for 30 minutes. After staining, we digitized the results using a Hamamatsu scanner S60. To ensure antibody specificity, we incubated the FFPE tissue sections with an antibody solution either with or without a 40-fold molar excess of purified Rep protein for 30 minutes at room temperature, followed by standard DAB detection.

3.2.2.1. DAB positive area quantification

I performed the quantification of the DAB positive area using a custom script developed for ImageJ/Fiji (v1.53q). The script processes RGB images to segment nuclei stained with haematoxylin and cells stained positively with 3,3'-diaminobenzidine (DAB). The script performs the following functions:

- Segments and counts all nuclei stained with hematoxylin.
- Identifies and counts nuclei in proximity to DAB-positive cells.
- Measures the total tissue area and the area stained positively with DAB.

I then used the Mann-Whitney test to analyse the differences between normal DAB staining and staining in which we preincubated the IHC reactions with purified H1MSB.1 Rep protein.

3.2.2.2. Cell-based quantification

I performed cell-based quantification of Rep⁺, CD68⁺ and CD163⁺ cells using a custom script developed for ImageJ/Fiji (v1.53q). This script was based on

DAPI nuclear staining, with approximately 4000 nuclei per tissue analyzed in digitized scans. To analyze differences between tumor, peritumor or tissues-adjacent to hamartoma tissues, I used the ANOVA test in GraphPad Prism 8.

3.2.2.3. Analysis of smoking status and clinical data of patients in correlation to Rep expression

The clinical information of patients was obtained from ThoraxKlinik (NCT) Heidelberg. Those include the sex and age of the patient, the smoking status and the histology of lung cancer. The smoking status was indicated as “non-smokers”, “ex-smokers” and “smokers” and for some ex-smokers and smokers, the PY (pack-year, a clinical quantification of cigarette smoking used to measure a person's exposure to tobacco) was available. I used this data to further divide the smokers and ex-smokers in high and low dose smoking. The goal was to achieve an approximately equal division of ~50% low and ~50% high (ex) smokers. For ex-smokers, those with 30 PY or more were classified as high ex-smokers. However, many patients had exactly 30 PY, leading to an unbalanced division.

3.2.3. Immunoblot

3.2.3.1. Tissue lysate preparation

I used frozen tissues obtained from ThoraxKlinik (tumor, peritumor, hamartoma) or fresh peritumor tissues obtained after surgeries of about 100 mg. I transferred them to a complete RIPA lysis buffer containing protease and phosphatase inhibitors. I then lysed the tissues using a tissue homogenizer following the following protocol: 3 cycles of 20 seconds each at a hard setting. After homogenization, I cleared the lysate by two 10-minutes steps centrifugation at 22,000 g and 4 °C.

3.2.3.2. Bradford protein assay

I performed protein quantification using the Biorad Protein staining dye based on the Bradford protein assay. To do this, I added 5 µl of the protein solution to 795 µl of water and mixed it with 200 µl of the ready-to-use Bradford

reagent. After incubating the mixture for 10 minutes at room temperature, I measured the optical density at 595 nm. I then calculated the protein concentration using a calibration curve obtained with BSA.

3.2.3.3. SDS-PAGE and western blot

I used commercially available gradient 4-20% acrylamide SDS-PAGE gels to separate proteins based on their molecular weight in denaturing conditions. To prepare for SDS-PAGE, I mixed the lung tissue lysates with Lämmli buffer and heated them at 95 °C for 10 minutes after the addition of 2.5% β -mercapthoethanol. As positive control, I used lysate of HEK293TT cells that were transfected with pcDNA3.1-H1MSB.1 DNA (codon optimized, HIS-tagged) and for negative control, I used lysate of untransfected HEK293TT cells. I determined the molecular weight using the Page Ruler Prestained Protein Ladder, which ranges from 10 to 170 kDa. To run the gel, I used running buffer containing tris/glycin. After the SDS-PAGE, I performed transfer blotting to transfer proteins onto nitrocellulose membranes, using the Trans-Blot Turbo Transfer System for purchased mini-gels. To check the total protein loading, I incubated the blotted membrane in Ponceau S, a reversible staining technique that allows for further immunological detection. I then washed the membrane with water and 1x PBS and I blocked the membranes for 1 hour at room temperature on a shaker with 5% milk in 1x PBS.

3.2.3.4. Immunodetection

To detect proteins in the nitrocellulose membranes, I utilized horseradish peroxidase (HRP)-coupled secondary antibodies that targeted the host-species of BMMF Rep monoclonal primary antibodies. All antibodies were diluted in 5% milk in 1x PBS. I applied the primary antibodies to the membranes overnight at 4 °C and the secondary for 1 hour at room temperature.

Between primary and secondary antibody incubation and just before protein detection, I washed the membranes 3 times with 1x PBS containing 0.1% Tween 20. I then detected target proteins using a chemiluminescent substrate

(ECL Select or Clarity Western ECL Substrate, depending on the required sensitivity).

I conducted protein detection on a Chemidoc Touch Imaging System and analyzed the pictures using Image Lab version 6.0.1. In cases where antibody removal was necessary to facilitate immunodetection with additional antibodies, I incubated the membranes with Restore PLUS Western Blot Stripping Buffer for 10 to 15 minutes at 37°C. After washing, I blocked again the membranes and I repeated the immunodetection procedure.

3.2.4. Immunoprecipitation

For the IP experiment, I used one peritumor tissue and one hamartoma tissue. I prepared the protein lysate by diluting it to contain 20 µg of protein in 1 ml of complete RIPA buffer. To establish the input control, I took 5% of this lysate and mixed it with Lämmli buffer in a 1:1 ratio. The remaining lysate was then incubated with 2 µg of primary Rep antibodies (AB2 and AB10) for 1.5 hours at 4°C on a rotating device.

Following antibody incubation, I added Protein A/G plus-Agarose suspension and incubated it for another 1.5 hours at 4°C on the rotating device. After this incubation period, I collected the pellet by centrifugation at 600xg for 3 minutes at 4°C. Then, I washed the pellet three times with complete RIPA buffer, each time repeating the centrifugation step mentioned earlier.

To evaluate the efficacy of the immunoprecipitation, I performed Western blot analyses with Rep AB3, after boiling the samples in Lämmli buffer for 10 minutes.

3.2.5. Mass spectrometry (proteomics DKFZ core facility)

To further investigate WB positive bands obtained after the immunoprecipitation, the same tissue lysates underwent mass spectrometry analysis at the Proteomics Core Facility, led by Dr. Dominic Helm and Martin Schneider.

The protein samples were applied to 1D-SDS-PAGE and fractionated. Gel pieces corresponding to four regions of interest (60-100 kDa, 45-60 kDa, 35-45 kDa and 20-30 kDa) were excised. Cysteines were reduced by adding DTT

and then carbamidomethylated using iodoacetamide, followed by a 4-hour digestion with Trypsin. The resulting peptides were loaded onto a cartridge trap column packed with Acclaim PepMap300 C18, separated via a gradient from 3% to 40% ACN on a nanoEase MZ Peptide analytical column, and analyzed by an online coupled Orbitrap Exploris 480 mass spectrometer. Data analysis was performed using MaxQuant (version 1.6.14.0).

Quantification was done using a label-free quantification approach based on the MaxLFQ algorithm. The peptide assignment was conducted using an up-to-date BMMF library containing possible peptides from all ORFs in all six frames.

3.2.6. *In situ* hybridization

RNA in situ hybridization was performed using the RNAscope® 2.5 HD Detection Reagent – BROWN Kit (Advanced Cell Diagnostics, Inc., Newark, CA, USA) according to the manufacturer's instructions. Briefly, FFPE tissue sections were deparaffinized using xylene and rehydrated through a graded series of ethanol solutions (99%, 96%, 70%) and TBS. Endogenous peroxidase activity was blocked by incubating the sections with hydrogen peroxide for 10 minutes at room temperature. Antigen retrieval was performed by boiling the slides in RNAscope® Target Retrieval Reagents for 15 minutes, followed by protease treatment using RNAscope® Protease Plus for 30 minutes at 40°C.

The tissue sections were hybridized with the target-specific RNAscope® Probe for 2 hours at 40°C in a HybEZ™ Hybridization System. After hybridization, the slides were washed with RNAscope® Wash Buffer and incubated with RNAscope® Signal Amplification reagents. The hybridized RNA was detected using the RNAscope® 2.5 HD Detection Reagent – BROWN, which produces a brown chromogenic signal. Finally, the slides were counterstained with Gill's hematoxylin, dehydrated, and mounted with Eukitt Mounting Medium. The stained slides were scanned using a Hamamatsu scanner S60 at 40x magnification. Digital images were analyzed using NDP software.

Probe	Rep region	Target region
HCBI4.296	480-1406 bp	331-1606 bp
MSBI2.176	603-1562 bp	580-1601 bp
C1MI.3M.1	292-1260 bp	273-1352 bp
MSBI1.176	602-1576 bp	461-1565 bp

3.2.7. Quantitative PCR

3.2.7.1. RNA extraction

I homogenized the frozen or fresh tissues with a tissue homogenizer (3 cycles of 20 seconds each at a hard setting) in RLT Plus buffer. Then, I isolated the RNA with the kit AllPrep DNA/RNA (Qiagen) according to manufacturer's instruction and I eluted the RNA in nuclease-free H₂O and stored it at -80°C. I determined the concentration on the RNA using a NanoDrop.

3.2.7.2. cDNA preparation

The RNA was then used to make cDNA using the kit ReverAid H Minus First Strand cDNA Synthesis Kit. I mixed a total of 1 µg of RNA with 1µl of random primers (stock 500 µM) and water to achieve a final volume of 17.5 µl. The mixture was then incubated at 70°C for 10 minutes and subsequently placed on ice for 2 minutes. Following this, I mixed the samples with a solution containing dNTPs, the reverse transcriptase enzyme and 5x reaction buffer to reach a final volume of 25 µl. The thermocycler protocol I used was 42°C for 90 minutes and 70°C for 10 minutes. I then diluted the cDNA 1:10 and stored at 4°C.

3.2.7.3. RT-qPCR

I performed the quantitative PCR using Power SYBR Green master mix on QuantStudio 5 System instrument. For each sample, I mixed the cDNA, with 10 µM of forward and reverse primer, water and SYBR Green Master Mix for a total volume of 10 µl. The thermal-cycling condition for this master mix was used according to the manufacturer's instruction, with an initial enzyme activation step at 95 °C for 10 min, followed by 40 cycles of 15 sec denaturing

at 95 °C and annealing for 60 sec at 60 °C. Melting curves were recorded simultaneously. I quantified β -actin or GAPDH in each plate and each reaction was measured in triplicate to test reproducibility. QuantStudio Design & Analysis Software automatically determined baseline and threshold. I considered negative every CT values greater than or equal to 35 and calculated the Δ CT as the difference in CT values of my gene of interest and endogenous control (GAPDH or β -Actin).

3.2.8. Fluorescence-Activated Cell Sorting (FACS)

3.2.8.1. Fresh tissues dissociation

Lung fresh tissues were cleaned from fat, fibrous and necrotic areas, cut into small pieces of 4-5 mm and dissociated using a gentleMACS Dissociator with heathers after being transfer into gentleMACS C tubes containing RPMI with an enzyme mix, according to the manufacturer's instructions. I used the program Medium 37C_h_TDK_2. If after the termination of the program there were still some larger pieces of tissues, I transferred the supernatants to other tubes and the pieces were dissociated again after adding more RPMI with the program m_imptumor_01. After this, the samples were resuspended and applied to a MACS SmartStrainer (20 or 70 μ m) placed on a 50 ml tube. I then washed the strainers with 20 ml of RPMI. I centrifuged the cell suspensions at 300xg for 7 minutes. If the samples appeared viscous, a DNase treatment was performed. After the supernatants were completely removed, samples were resuspended in 1x PBS and the cells were counted.

3.2.8.2. Fluorescent antibody labeling

For compensation we used one tube for unstained tissues, one tube for ZombieAqua viability staining and 4 tubes with beads for single staining. ZombieAqua was used to discriminate between living and dead cells. The sample was centrifuged at 350xg for 5 minutes, supernatant was fully removed and cells were incubated with ZombieAqua diluted in 1x PBS, according to the manufacturer's instructions. The tube was incubated for 20 minutes in the dark at room temperature.

Fc block step was performed with FcR Blocking Reagent, incubated for 10 minutes at room temperature.

Cell surface markers I used:

	Isotype control	Antibodies
FITC	IgG	EpCAM
PE	IgG	CD45
APC	IgG	CD68
Pe_Cy7	IgG	CD3
PO	IgG	ZombieAqua

I incubated each staining mix with the cells for 20 minutes in the dark. Then I centrifuged the samples at 350xg for 5 minutes, supernatants were discarded and cells were washed with PBS before being resuspended in 500 µl FACS buffer.

3.2.8.3. Sorting (DKFZ Flow Cytometry – FACS Facility)

Fluorescence-activated cell sorting (FACS) was performed at the DKFZ Flow Cytometry Core Facility using a BD FACS Aria Fusion, with the assistance of Dr. Marcus Eich who determined compensation and gating strategies. After gating, macrophages (CD45⁺, CD68⁺), T-cells CD45⁺, CD3⁺) and epithelial cells (CD45⁻, EpCAM⁺) were sorted and collected in 1x PBS.

3.2.9. RNA sequencing

3.2.9.1. RNA extraction from sorted cells

Sorted cells collected in 1x PBS were immediately processed for RNA extraction using the kit RNeasy Plus Micro Kit, according to the manufacturer's instructions. RNAs were eluted in 14 µl of RNAase-free water. RNA concentration was assessed with a Qubit, with the help of Ing. Sabine Schmidt (D300).

3.2.9.2. Bulk RNA-seq (DKFZ NGS Core Facility)

Total RNA was subjected to an initial quality and quantity check using the BioAnalyzer by Agilent Technologies (RNA Pico Kit). Starting with 0,26-10 ng of input material. Total RNA was converted to cDNA using the SMART-Seq® v4 Ultra® Low Input RNA Kit for Sequencing (Takara Bio USA, Inc.) according to the manufacturer's instructions. During first-strand synthesis the product was primed by the 3' SMART-Seq CDS Primer II A and the SMART-Seq v4 Oligonucleotide was used for template switching at the 5' end of the transcript. The cDNA was then amplified by long distance PCR using 9 or 12 PCR cycles (if the input amount was below 1ng). The BioAnalyzer was used to check for successful cDNA synthesis and amplification (High Sensitivity DNA Kit). On the Covaris LE220 the amplified double-stranded cDNA was sheared to 150-250bp fragments. Successful fragmentation was verified on the TapeStation by Agilent Technologies (HS DNA 1000 Kit). For subsequent library preparation the NEBNext® ChIP-Seq Library Prep Master Mix Kit was used. After end repair, dA-Tailing and adaptor-ligation the cDNA was subjected to PCR amplification introducing i5 and i7 primers using 12 PCR cycles. Library quality was assessed on the TapeStation (DNA 1000 Kit). Sequencing was performed on the NovaSeq 6000 platform (Illumina).

3.2.9.3. Analysis of Bulk RNA-seq results

The raw screening for BMMF hits was done by Lisa Häfele (PhD, D460/F200) using the D-ViSioN pipeline (Detection of Integrated Viral Sequences by Singletons) that she adapted during her master thesis [114] and that was provided with a library of BMMF sequences. D-ViSioN uses two subsequent alignment steps performed by BWA mem and BLAST. The fast BWA mem alignment is performed first, and the reads extracted from the BWA mem output are submitted as input to the more sensitive second alignment step conducted via BLAST [114]. The sample in which BMMF1 hits were found was used to generate a gene count table. Gene expression levels were quantified based on two predefined gene lists representing M1- and M2-like macrophages [115]. To determine the differences in gene

expression between M1- and M2-like macrophages, a Mann-Whitney U test was employed, given the non-normal distribution of the data as confirmed by the Shapiro-Wilk test.

3.2.10. *In silico* analysis

3.2.10.1. Pan-Cancer Analysis of Whole Genomes (PCAWG) screening

The raw screening of the Pan-Cancer Analysis of Whole Genomes (PCAWG) was done by Lisa Häfele (PhD, D460/F200) using the D-ViSioN pipeline (see above).

After that, I analyzed the results to evaluate the presence of BMMF positivity in two types of lung cancer: adenocarcinoma and squamous cell carcinoma. For each cancer type, I calculated the percentage of BMMF-positive patients using data from both whole-genome sequencing (WGS) and RNA sequencing (RNA-seq).

3.2.10.2. Single Cell RNA Sequencing (scRNA-seq)

We utilized a publicly available scRNA-seq dataset (from 49 clinical biopsies taken from 30 patients with metastatic lung cancer) to perform a screening of BMMF reads [115].

The raw screening of the FASTq files was done by Lisa Häfele (PhD, D460/F200) using the D-ViSioN pipeline (see above). After this, I first created a subset of the immune cells from the original Seurat object, focusing on cells annotated as macrophages. The subset was further modified in order to include only cells derived from lung biopsy samples (without including metastasis). Based on the raw screening from Lisa Häfele, I created a list on an Excel file indicating BMMF positivity for each cell and I incorporated this list into the Seurat object as an additional metadata column. To classify macrophages into M1- and M2-like phenotypes, I utilized a gene signature approach using two lists of marker genes for the of M1- and M2-like phenotypes (Table S3) from the authors. These gene sets were used to calculate module scores for each cell using the “AddModuleScore” function

in Seurat, which resulted in the assignment of M1 and M2 scores to each macrophage.

The expression patterns of the M1- and M2-like gene signatures were visualized using `FeaturePlot`, displaying the distribution of M1 and M2 scores across BMMF-positive and BMMF-negative cells. Additionally, violin plots were generated to compare the distribution of M1 and M2 scores between BMMF-positive and BMMF-negative cells. These plots were supplemented with boxplots to highlight the central tendency and variability within each group. Finally, with the collaboration of Dr. Silvia Calderazzo (C060), robust linear regression models were fitted to assess the relationship between BMMF positivity and M1 or M2 scores, while controlling for the sample origin avoiding the confounding effect. The models were fitted using the `lmrob` function from the `robustbase` package, which provides robust estimates that are resistant to outliers and violations of model assumptions.

4. Results

4.1. Cytoplasmic staining of specific lung cells detected by IHC

To comprehensively test BMMF1 detection in lung cancer bio samples and identify a possible causal contribution of BMMF1 to lung cancer development, I applied state-of the art multi omics readouts including IHC staining, immunodetection, IP, mass spectrometry analysis, RT-qPCR, *in situ* hybridization, FACS analysis and *in silico* approaches like BMMF reads detection in PCAWG data set or scRNA-seq analysis. First-line experimental analyses were based on immunohistochemical characterization of lung (cancer) tissues. For that, we analyzed a cohort of histopathologically characterized lung cancer tissues for the presence of Rep in both tumor and peritumor tissue parts. With the collaboration of Claudia Ernst (TA, D460/F200), we stained FFPE sections with a set of eight monoclonal antibodies against the Rep protein (AB 1, 2, 3, 7, 10, 11, 15, 21) (Table S1). Preliminary staining based on 6 peritumor tissues was also conducted with anti-Rep antibodies AB 22 and AB 23, which were generated as a consequence of positive *in silico* BMMF screening results in different cancer tissues (see below, Figure S1). Tissue staining was observed in speckled foci in the cytoplasm of a specific population of cells based on immunohistochemical staining with DAB (Figure 9).

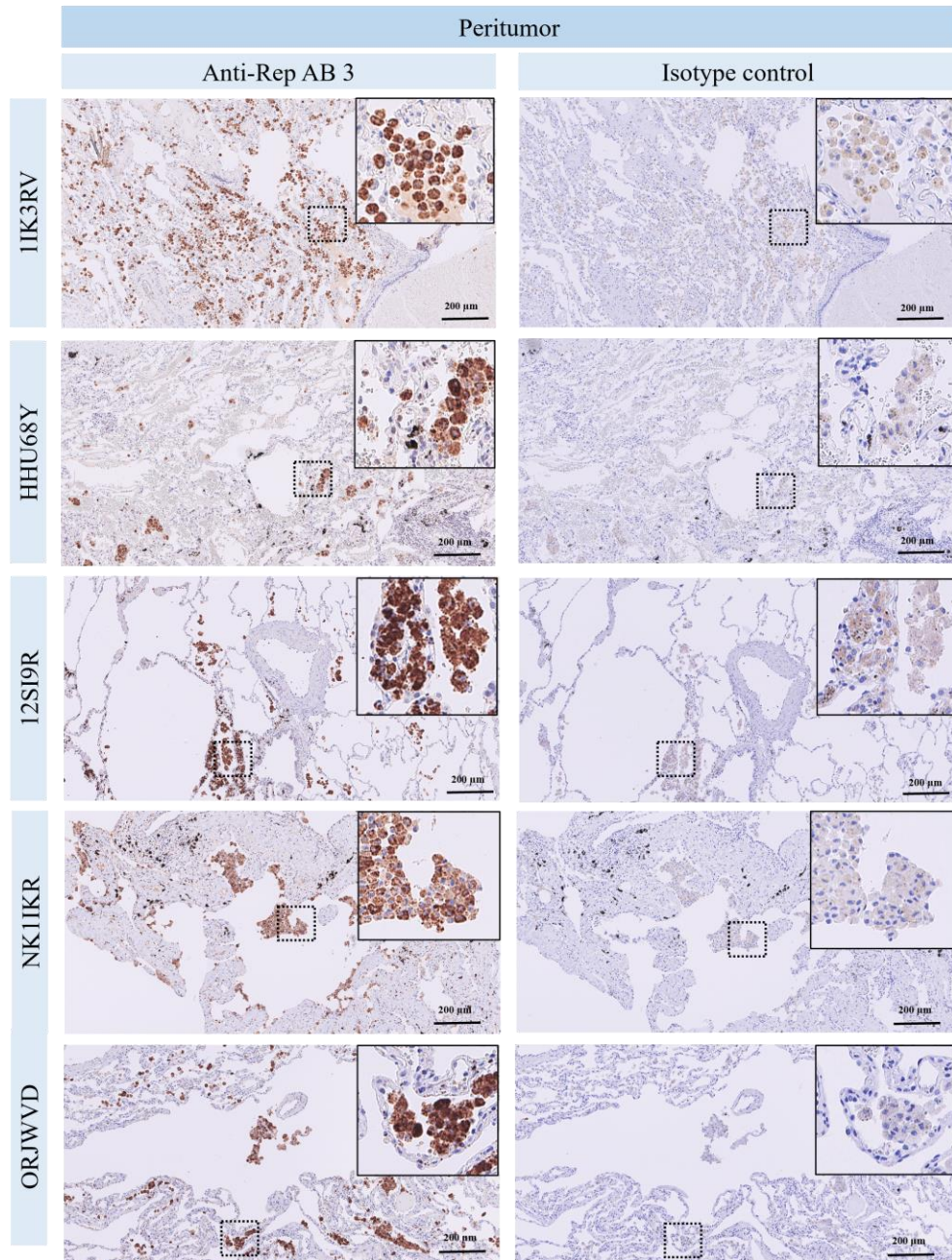


Figure 9. Immunohistochemical detection of BMMF Rep in peritumor tissues. Consecutive peritumor tissue sections from lung cancer patients were stained with Rep AB 3 or isotype control antibody (IgG1). Stainings of five representative patients are shown. The staining is localized in speckled cytoplasmic foci of a specific population of cells.

4.2. BMMF Rep staining localizes primarily in peritumor tissues

We obtained positive staining in all peritumor tissues tested (n=42) while positive staining in tumor tissues was observed in 28% of tumor tissues tested (n=37). In general, 8 out of 8 antibodies tested showed staining in the lung

tissues. We observed specific and reproducible staining with 5 out of 8 antibodies tested (Ab 1, 2, 3, 10, 21). Weaker staining was observed with the remaining 3 antibodies (Ab 7, 11, 15). We obtained the strongest signal with AB 3 and 10, which share the same epitopes (Table 10, Figure 10). AB 3 showed staining in 66 out of 98 tissues: 11 out 36 tumor, 42 out of 42 peritumor, 13 out of 19 controls. AB 10 showed staining in 47 out of 93 tissues: 6 out 37 tumor, 34 out of 38 peritumor, 7 out of 18 controls. Additional antibodies with staining in a high number of tissues were AB 21 (46 out of 94 tissues, 4 out of 34 tumor, 33 out of 41 peritumor, 9 out of 19 controls), AB 1 (45 out of 97 tissues: 3 out 36 tumor, 31 out of 41 peritumor, 11 out of 19 controls) and AB 2 (40 out of 92 tissues, 4 out of 34 tumor, 32 out of 38 peritumor, 4 out of 15 controls)

In peritumor tissues, stained cells were distributed over the entire tissue section. The *foci* may reflect Rep aggregation in the cytoplasm of specific population of cells localized on the surface/the inner space of lung alveoli. In tumor tissues, when present, the staining was mostly found in the tumor stroma while the solid tumor was mostly negative (Figure S2).

Despite staining with 8 different antibodies (with 5 different epitopes) and detecting no staining or a much less intense, greyish staining with isotype controls indicated specific staining, we additionally performed immunohistochemical competition experiments with purified H1MSB.1 Rep to confirm specificity of the tissue staining. Upon pre-incubation with purified H1MSB.1 Rep, we observed a decrease of 55%, indicating effective blocking of the antibody interaction with Rep AB 3 when compared to the control (Figure 11).

Table 10. Overview of IHC DAB staining results. Tissues with strong antibody staining in the full section are indicated in dark green color. Regionally stained tissues and tissues with medium or low intensity staining are represented in green. Tissues with staining in less than 5 tissue regions were indicated as negative, together with tissues with too high background and/or no detectable staining. White cells represent non-stained tissues.

Patient	Tissue	AB 3	AB 10	AB 1	AB 7	AB 11	AB 15	AB 2	AB 21	CD 68
1IX6KF	T									
	NG									
16P87Y	T									
	NG	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1LYLZF	T									
	NG									
1FAVEM	T									
	NG									
1DVWTK	T									
	NG									
0RFPWD	T									
	NG									
1PZBGD	T									
	NG									
14CFEM	T									
	NG	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
UJQJQN	T									
	NG									
OVBCKC	T									
	NG									
1HK3RV	T									
	NG									
1HUC8Y	T									
	NG									
NK1HR	T									
	NG									
6LBF64	T									
	NG									
XNB50E	T									
	NG									
1SRX3R	T									
	NG									
ZCZ8WH	T									
	NG									
MO87SD	T									
	NG									
0RQVFN	T									
	NG									
1NZV3L	T									
	NG									
2V6GGQ	T									
	NG									
12SIDR	T									
	NG									
DTN9K4	T									
	NG									
11F3TZ	T									
	NG									
1GCCLAF	T									
	NG									
RGN4XJ	T									
	NG									
1TASZJ	T									
	NG									

Patient	Tissue	AB 3	AB 10	AB 1	AB 7	AB 11	AB 15	AB 2	AB 21	CD 68
JGS1F3	T									
	NG									
57DSUO	T									
	NG	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1W2LHA	T									
	NG									
RUX2KD	T									
	NG									
TJHWQ9	T									
	NG									
14ZOKY	T									
	NG									
4BRHD9	T									
	NG									
EAB4PL	T									
	NG									
NKPUCX	T									
	NG									
HCK7BC	T									
	NG									
1RGCLS	H									
	H									
4E3TC	H									
	H									
1XFDVM	H									
	H									
TXZBEZ	H									
	H									
NC2CFC	H									
	H									
10X1PE	H									
	H									
1G944Y	H									
	H									
10ZPCM	H									
	H									
9SSVPV	H									
	H									
1N4CNG	H									
	H									
10E7GJ	H									
	H									
PJB41U	H									
	H									
90FQ9K	H									
	H									
13DRLE	H									
	H									
WPP5S6	H									
	H									
123UWA	H									
	H									
1GVYKA	H									
	H									
15183	PT									
	PT									
15184	PT									
	PT									
15188	PT									
	PT									
15192	PT									
	PT									
15193	PT									
	PT									
15194	PT									
	PT									
15195	PT									
	PT									
15196	PT									
	PT									

NCT – Thorax Klinik Project 2750
NCT – Thorax Klinik Project 3359
NCT – Thorax Klinik Project 3694
NCT – Thorax Klinik Fresh Tissues

Strongly Positive
Positive
Negative/Inconclusive
Not Stained

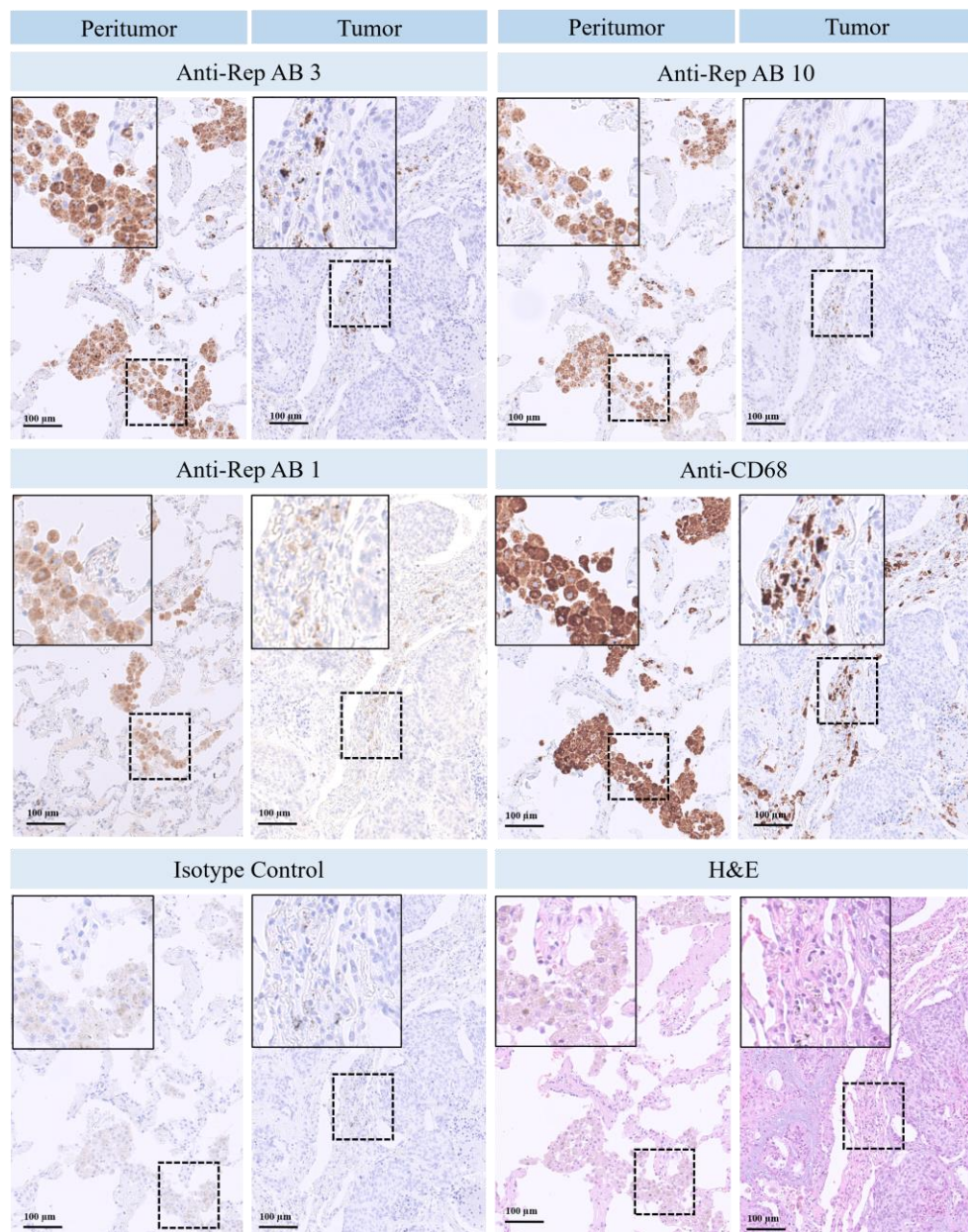
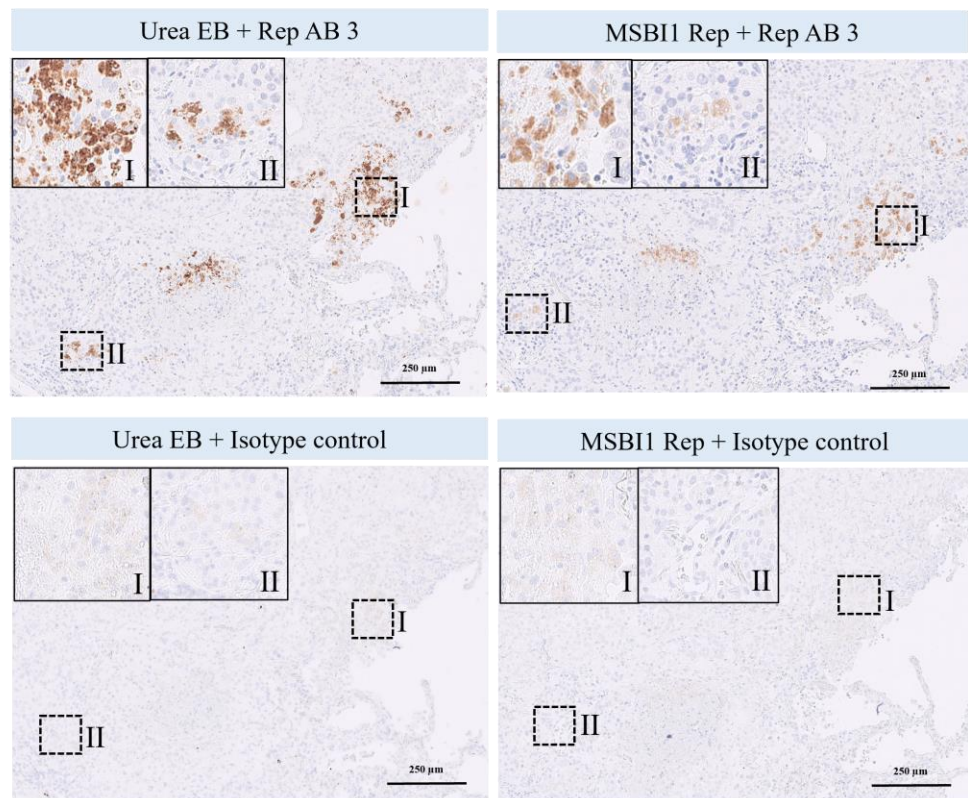


Figure 10. Immunohistochemical detection of BMMF Rep and CD68 in peritumor and tumor tissues.

Consecutive peritumor and tumor tissue sections from lung cancer patients (35 patients + 2 tumor tissues unpaired + 7 peritumor tissues unpaired) were stained with 8 Rep antibodies (AB 3, AB 10, AB 1 shown), anti-CD68 or isotype control antibodies. The intensity of the staining depends on the Rep antibodies used (Rep AB 3 and AB 10 showing the strongest staining pattern). In tumor tissues, when present, the staining is mostly found in the tumor stroma.

A



B

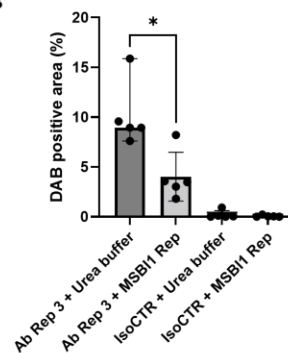


Figure 11. Blocking of anti-Rep antibody staining by pre-incubation with Rep in DAB IHC. Consecutive lung cancer peritumor tissue sections were immunohistologically stained with anti-Rep AB3 antibody after pre-incubation with purified H1MSB.1 protein or with the elution buffer (urea buffer). **A:** For better orientation, identical tissue areas are marked with I-II and enlarged (xxx-fold?). **B:** Quantification of the DAB positive area (%). Pre-incubation with Rep decreases the staining by 55% (median first group 8.9, median second group 4.0) when compared to the control (pre-incubation with urea buffer). Mann-Whitney test applied. $P = 0.016$.

4.3. Rep expression is predominantly observed in tissue parts with CD68⁺ (alveolar) macrophages

We then stained consecutive tissues with anti-CD68 antibody, widely used as a pan-macrophage marker [116]. We observed that the staining pattern of CD68 was congruent with Rep⁺ cells in specific areas of the tissues, while in other areas, CD68⁺ cells did not exhibit Rep staining. These observations confirmed the morphological characteristics of tissue staining described in a previous and smaller pilot study [117]. In the previous and in this study, it is presumed that Rep⁺ CD68⁺ macrophages correspond to alveolar macrophages, which are larger and round shaped and localized specifically to the surface of lung alveoli. Conversely, the non-Rep-stained Rep⁻ CD68⁺ macrophages may represent a population of interstitial macrophages, characterized by a smaller and less rounded shape and location in the middle of the interstitium, which was supported by the DAB stainings (Figure 12). The general overlap of staining between Rep and CD68 in consecutive tissue cuts of patient tissues was observed with AB 3, 10, 1. DAB staining with anti-Rep antibodies showed a more speckled cytoplasmic staining pattern when compared with the more homogenous staining with anti-CD68 antibodies.

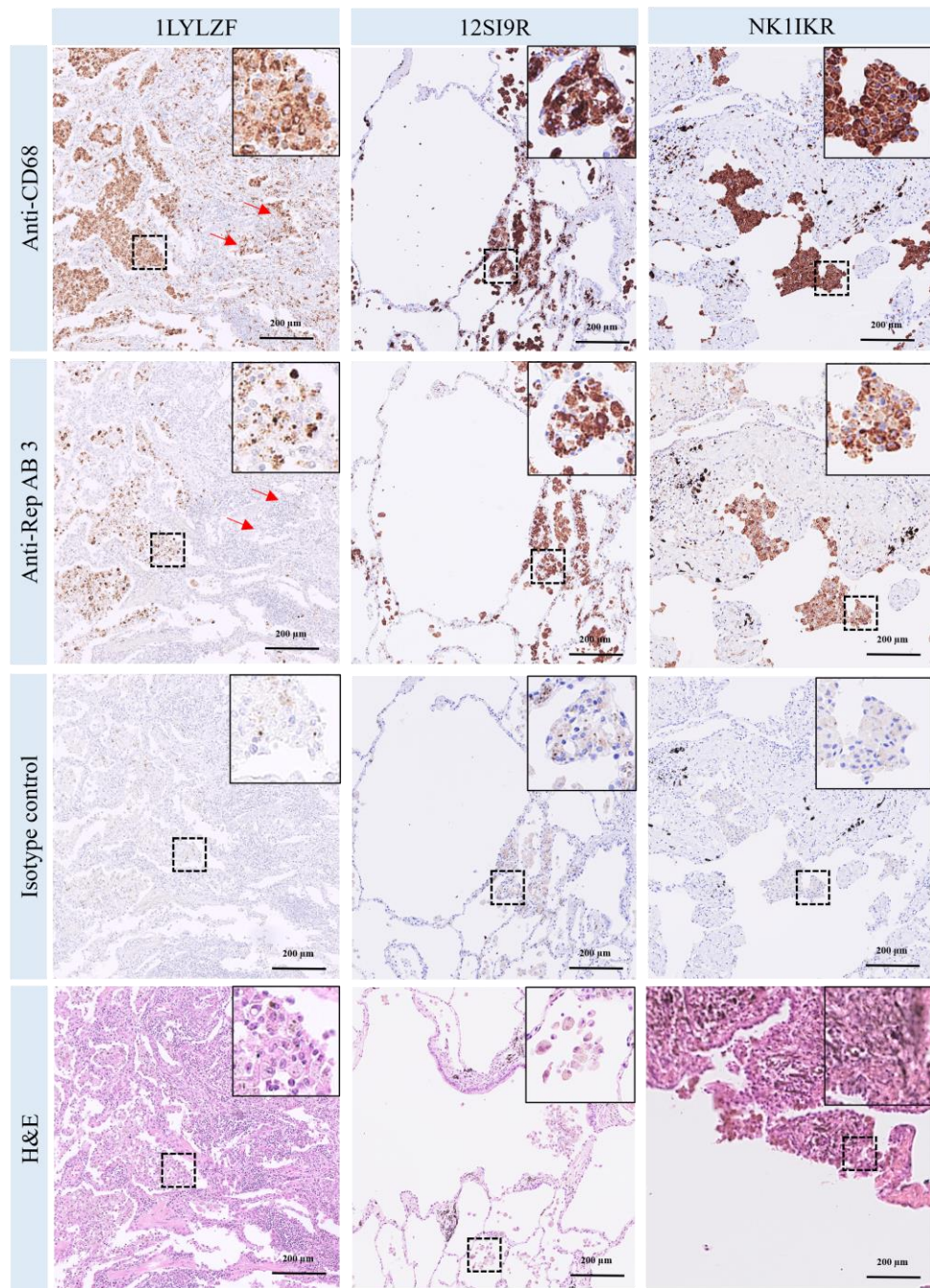


Figure 12. Immunohistochemical detection of BMMF Rep and CD68 in peritumor tissues. Consecutive peritumor tissue sections from lung cancer patients were stained with Rep AB 3, anti-CD68 or isotype control antibody. Tissues of 3 representative patients are shown (n=42). The staining pattern of CD68 generally overlaps with the staining pattern of Rep but in some areas, CD68⁺ cells do not exhibit Rep staining (red arrow).

4.4. Increased Rep expression in peritumor compared to tumor and control tissues

To validate the congruency of staining of Rep and CD68 cells, with the collaboration of Claudia Ernst (TA, D460/F200), we performed an IHC IF

staining using a mix 1:1 of Rep AB3 and AB10 (which performed best in IHC DAB staining) and anti-CD68 antibodies. 35 lung cancer patients (paired tumor and peritumor tissues) and 2 unpaired tumor tissues were included in the analysis. Again, like in DAB staining, anti-Rep tissue staining was most intense in speckled *foci* in the cytoplasm of CD68⁺ cells, confirming the IHC data.

To test disease association between BMMF and lung cancer, we performed a cell-based, semi-automated case vs. control comparison of Rep and CD68 expression of non-cancer and peritumor tissues. For that, we included lung tissues from 19 non-cancer individuals in the analysis.

Semi-automated quantification showed a significant increase of Rep⁺ cells in peritumor (median: 10.4% Rep-positive cells from all cells) compared to tissues-adjacent to hamartomas (6.5%, P=0.009) and to tumor tissues (0.03%, P<0.0001) (Figure 14 A).

The distribution of Rep⁺ cells generally revealed strong heterogeneity within the peritumor cohort with high and low responders ranging from 0.009 to 52.3%.

4.5. Rep expression correlates with the detection of (M2-like) macrophages

To test an association of Rep with macrophage detection, we analysed the tissue sections (35 peritumor, 37 tumor and 19 non-cancer individuals) by triple Co-IF staining and cell-based quantification of Rep⁺ and CD68⁺ (and CD163⁺, see below) cells.

Quantification of CD68⁺ cells, alone, did follow the same trend of Rep⁺ cells and was increased in peritumor (24.4% CD68⁺ cell of all cells, median) when compared to tumor (0.5%, P<0.0001) and tissues-adjacent to hamartoma (15.1%, P=0.021) (Figure 14 B). Importantly, while numbers of Rep⁺CD68⁺ macrophages again revealed a significant differentiation between peritumor and tissues-adjacent to hamartoma (peritumor: 10.1%, tumor: 0.0%, tissues-adjacent to hamartoma: 5.3%, P<0.0001 & 0.009, respectively) comparable to Rep⁺ cells alone (Figure 14 C), numbers of Rep⁻ macrophages did not show

a significant difference between peritumor and tissues-adjacent to hamartoma (Figure 14 D).

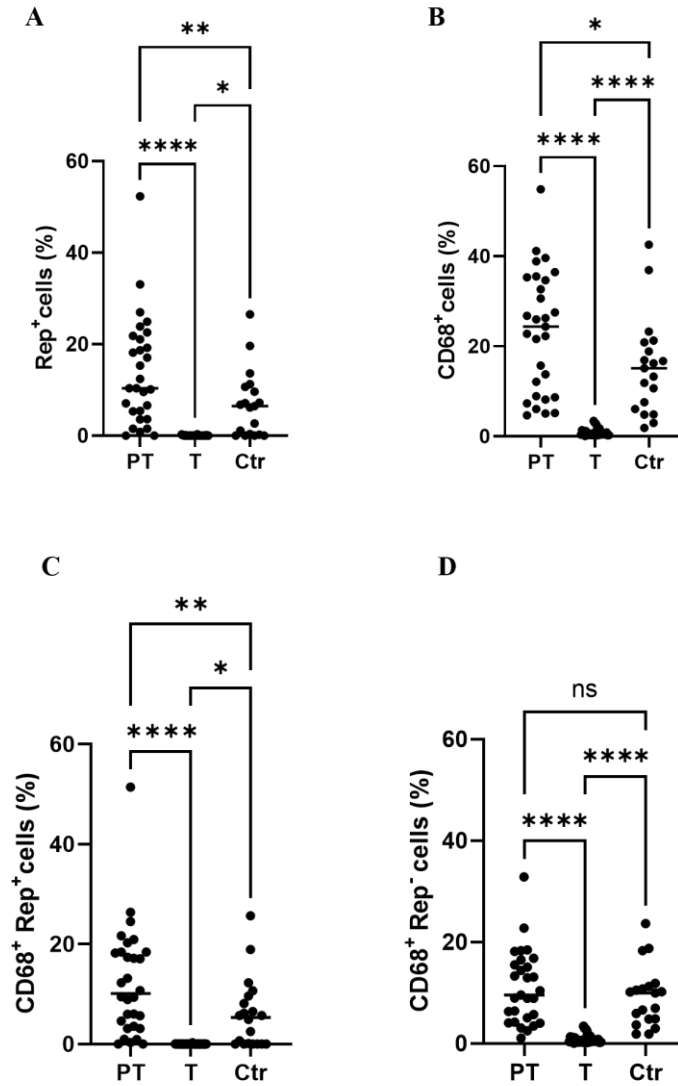


Figure 13. Quantification of Rep⁺ and CD68⁺ cells based on co-immunofluorescence microscopy. Semi-automated quantification of Rep⁺ (A), CD68⁺ (B), CD68⁺Rep⁺ (C) and CD68⁺Rep⁻ (D) cells is given in % of all cells. Peritumor tissues showed the highest amount of CD68⁺ cells, followed by non-cancer individual and solid tumor tissues, the last showing very low level of macrophages (B). Rep positivity (A) is mostly found in CD68⁺ cells (C). High level of Rep⁺CD68⁺ cells were observed in peritumor tissues. Tissues from non-cancer individuals also showed high level of Rep⁺ macrophages. Rep negative macrophages quantification (D) did not show significant difference between peritumor and non-cancer individual tissues. Level of CD68⁺Rep⁻ cells remained low in tumor tissues. One-way ANOVA test was used for the statistical analysis. Significance: *P<0.05, **P< 0.01, ***P<0.001.

We next performed quantification based on the full triple IHC IF staining setup based on Rep antibodies (AB3 and AB10 in a 1:1 mix), anti-CD68

antibody and the M2-like macrophages marker anti-CD163 antibody. The population of macrophages was further differentiated into M1-like (CD68⁺CD163⁻) or M2-like (CD68⁺CD163⁺) macrophages to test an association of the sub-populations with Rep detection (Figure 15).

Numbers of M2-like macrophages were significantly increased in the peritumor (6.3%) compared to the tumor (0.01%, $P<0.0001$) and tissues-adjacent to hamartoma (3.3%, $P=0.048$) (Figure 16 A). Almost identical trend was observed for the number of Rep⁺ M2-like macrophages (peritumor: 4.6%, tumor: 0.0%, tissues-adjacent to hamartoma: 2.6%, $P<0.0001$ and 0.044, respectively) (Figure 16 B), while populations of Rep⁻ M2-like macrophages did not show a significant difference between peritumor (1.3%) and tissues-adjacent to hamartoma (0.8%, $P=0.551$) (Figure 16 C). Numbers of M1-like macrophages (total, Rep⁺ and Rep⁻) did not show a significant increase in peritumor compared to controls (Figure S3).

Although there was only a slight and non-significant increase in the numbers of Rep-positive M2-like macrophages (4.5%) compared to Rep-positive M1-like macrophages (4.0%, $P=0.814$) in the peritumor tissues (Figure 16 D), a significant increase was observed for the number of Rep-negative M1-like macrophages (8.2%) compared to Rep-negative M2-like macrophages (1.3%, $P<0.0001$) (Figure 16 E). This finding suggests that Rep is primarily associated with M2-like macrophage populations.

This is further supported by a stronger and highly significant correlation between the number of Rep⁺ cells and CD68⁺CD163⁺ macrophages ($r=0.88$, $P<0.0001$) (Figure 16 F) compared to the correlation between Rep⁺ cells and CD68⁺CD163⁻ macrophages ($r=0.43$, $P=0.01$) (Figure 16 G).

Considering only control samples, the correlation between Rep and M2-like macrophages is slightly weaker ($r=0.81$, $P<0.0001$), while it is now also more pronounced for M1-like macrophages ($r=0.65$, $P=0.002$) (Figure S4).

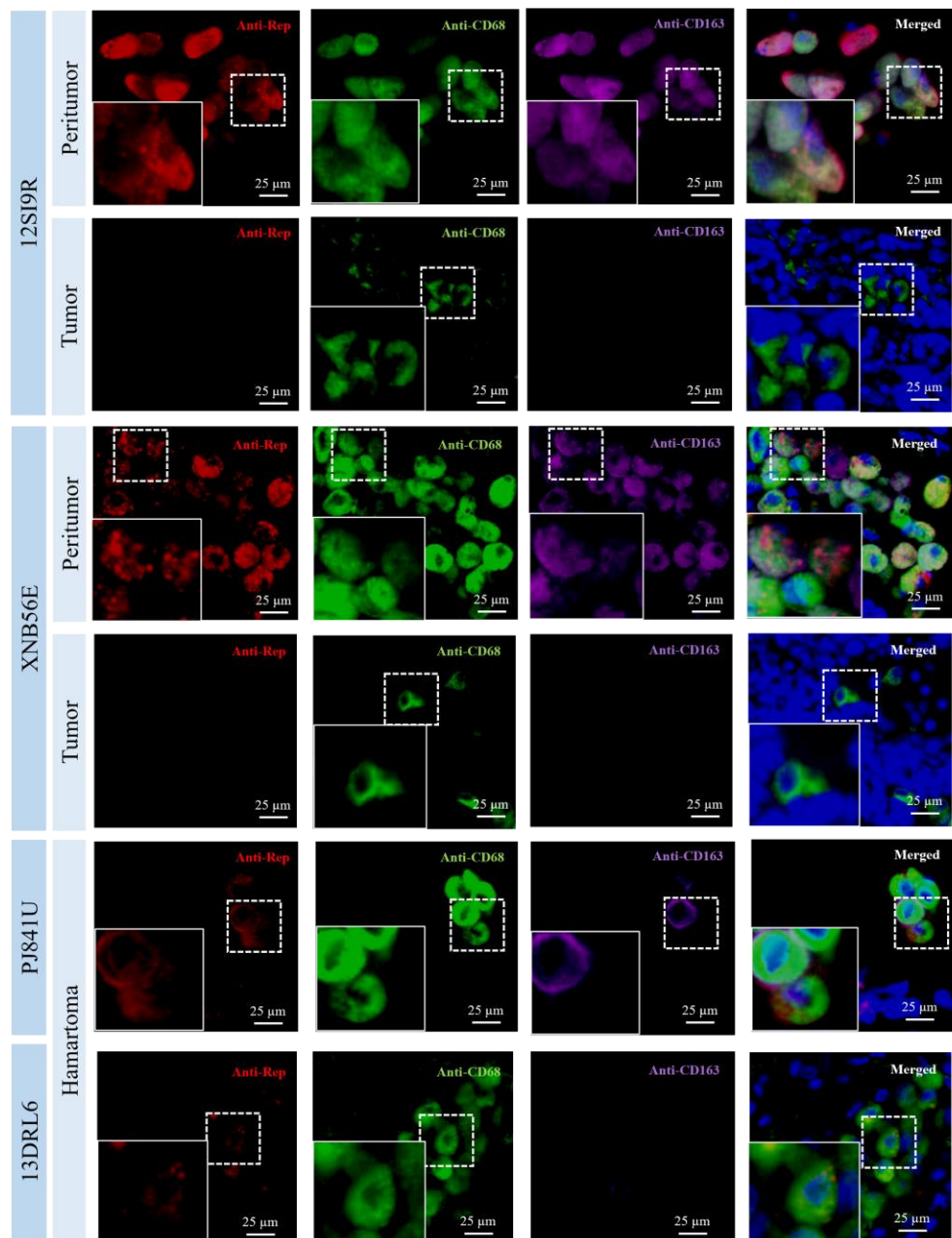


Figure 14. Co-immunodetection of Rep, anti-CD68 and anti-CD163 expression. Tissue sections from lung cancer patients (n=35 [paired] + 2 tumor tissues unpaired) and controls (n=19) were stained in triple IHC IF with Rep AB 3, anti-CD68 and anti-CD163. Scale bar: 25 μ m, DAPI staining of nuclei in blue.

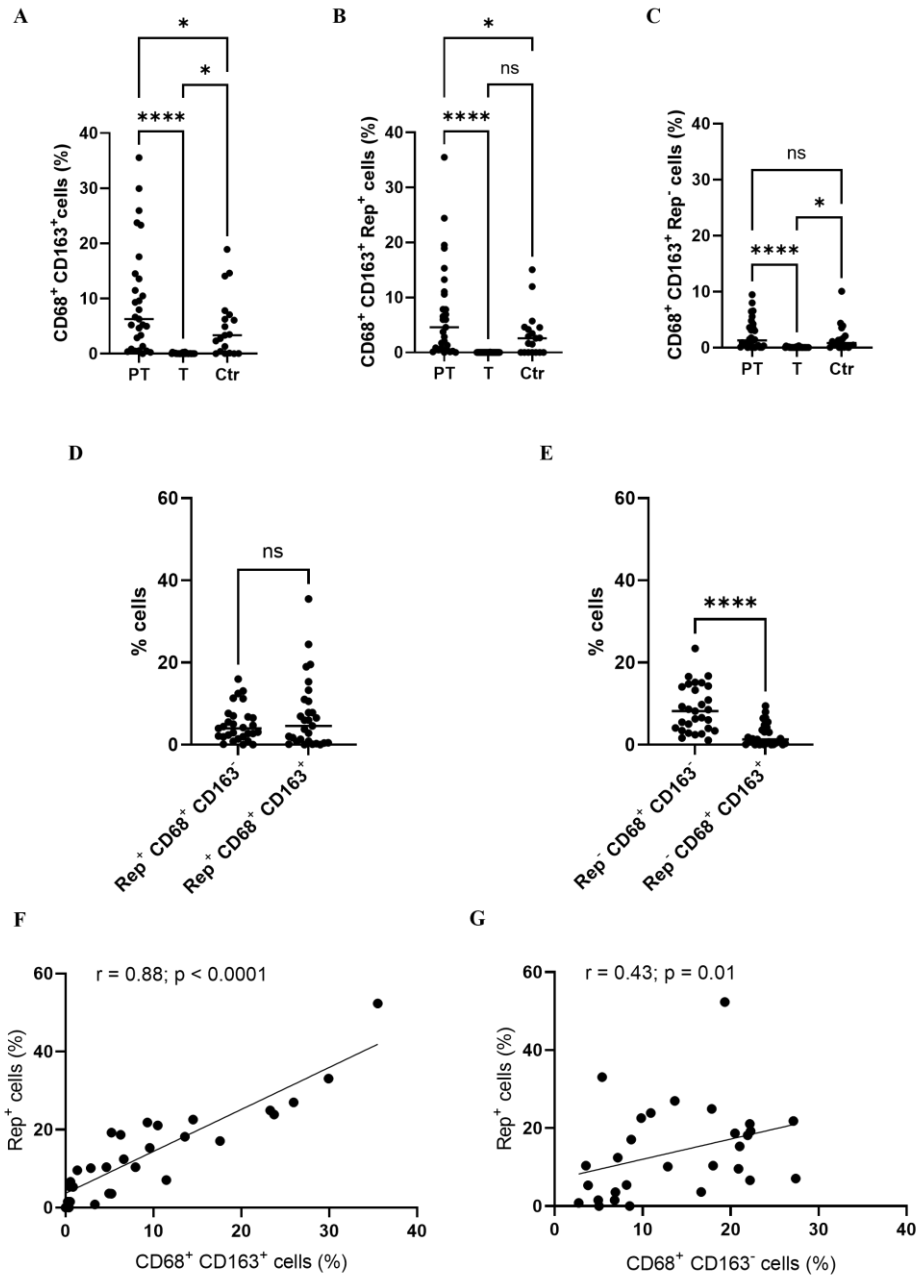


Figure 15. Quantification of Rep⁺, CD68⁺ and CD163⁺ cells based on co-immunofluorescence microscopy and correlation of Rep⁺ cells with M1 or M2-like macrophages. Cell quantification based on tripe Co-IF given in % of all quantified cells. Quantification of CD68⁺CD163⁺ (A), CD68⁺CD163⁺Rep⁺ (B) and CD68⁺CD163⁺Rep⁻ (C) cells. One-way ANOVA test. Significance: *P<0.05, **P< 0.01, ***P<0.001. D: Quantification of Rep negative M1- and M2-like macrophages in peritumor tissues. E: Quantification of Rep positive M1- and M2-like macrophages in peritumor tissues. Mann-Whitney test. ***P value<0.001. Correlation plots for Rep⁺ and CD68⁺CD163⁺ (F) and Rep⁺ and CD68⁺CD163⁻ cell populations (G): test by simple linear regression and Spearman correlation for statistical analysis.

4.6. BMMF expression is evident in low dose smokers

Because it is known that smoking can significantly affect inflammatory patterns in the lung, I analysed whether the smoking habits of the individual donors could be associated with changes of BMMF expression in lung tissues. For that, each patient's smoking status was categorized as “no smoker,” “ex-smoker,” or “smoker”.

In general, comparison of stained cell populations for non-smokers, ex-smokers and smokers including hamartoma patients did not result in any significant differences between the groups (Figure 17 A-J, Figure S6 A-N). All cohorts revealed a very heterogenic distribution of data points with strong variation between minimum and maximum cell numbers (with the exception of the more uniform groups of smokers). Yet, median numbers of Rep⁺, CD68⁺CD163⁻ and CD68⁺CD163⁺ cells were increased in non-smokers (10.4%, 18.0%, 8.0%) and smokers (10.0%, 13.4%, 5.2%) compared to ex-smokers (6.6%, 10.1%, 5.0%) except for Rep⁺CD68⁺CD163⁺ which were only increased in non-smokers (6.5%) and on a low level in both ex-smokers (2.6%) and smokers (3.1%) (Figure S6 A-D).

To more directly associate tissue staining with smoking status, the groups of ex-smokers and smokers were further stratified into either a ‘high’ dose or ‘low’ dose smoking category based on tobacco consumption (information on packages per year, PY). Patients with unclear smoking histories were excluded from the analysis, as were individuals without specified PY data. Despite the stratification did not generate a more homogenous data distribution and still no significant differences were observed between the groups, two observations stood out.

First, Rep⁺ cells were more frequently observed in high dose ex-smokers (12.2%) compared to low dose ex-smokers (2.7%) (Figure S6 E), while numbers of M1-like macrophages were on an equal level (3.5% compared to 4.8%) for the same groups. Second, low dose smokers seem to possess slightly higher numbers of M1-like macrophages (11.6%) compared to high dose smokers (3.5%) (Figure S6 F).

We next tested correlation of the number of stained cells with the smoking intensity indicated numerically by packages per year (PY) and by individuum.

For smokers, a non-significant trend for increased detection of Rep⁺ cells with low smoking doses (PY) was observed ($r=0.380$, $P=0.246$) (Figure 17 A). A similar trend was detected for M1- and M2-like macrophages: smokers with lower PY had higher amounts of both M1- (Figure 17 B) and M2-like macrophages (Figure 17 C). ($r=0.559$, $P=0.077$ and $r=0.697$, $P=0.02$, respectively).

Further stratification of cell populations into Rep⁺ M2-like macrophages decreased the P-value for anti-correlation with smoking status even more when compared to Rep⁺ cells, alone ($r=0.486$, $P=0.131$) (Figure 17 E).

In contrast, ex-smokers displayed greater variability. Especially for patients with low PY, the number of Rep, M1- or M2-like cell populations show a pronounced heterogeneity with both low and high responders. This is especially true for quantifications of Rep⁺ cells ($r=0.379$, $P=0.09$), where the gap between high and low responders is the largest (Figure 17 F). M1- and M2-like cells exhibited a more horizontal trendline ($r=0.151$, $P=0.512$ and $r=0.173$, $P=0.451$, respectively) (Figure 17 G, H). Differentiation into Rep⁺ M2-like and Rep⁺ M1-like revealed that Rep⁺ M2-like macrophages showed a more neutral incline compared to Rep⁺ M1-like macrophages in the correlation plots ($r=0.500$, $P=0.02$ and 0.233 , $P=0.309$, respectively) indicating that low PY is more associated with Rep⁺ M2-like MFs (Figure 17 I-J).

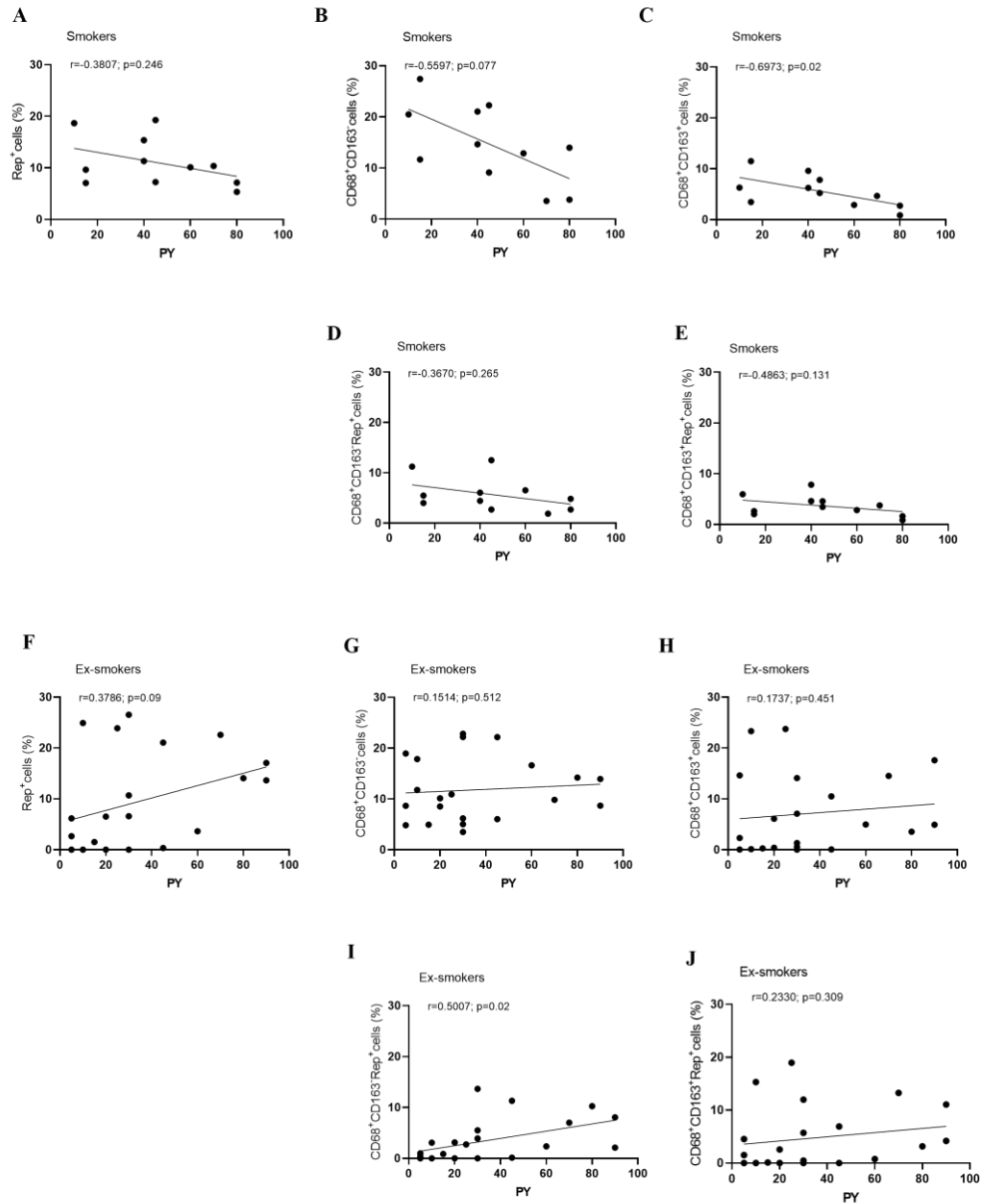


Figure 16. Rep expression in non-smokers, ex-smokers and smokers. A-J. Ex-smokers and smokers were stratified into either 'low' or 'high'-dose based on consumption levels given as packages per year (PY). Non-smokers included in the analysis: 7; ex-smokers: 21 (9 low and 12 high-dose); smokers: 11 (5 low and 6 high-dose). Test by simple linear regression and Spearman correlation was used for statistical analysis. Quantifications of Rep⁺, CD68⁺ and CD163⁺ cells are given in % of all cells.

4.7. Characterization of tissue lysates by immunoblotting

Rep AB 3 and AB 10, which showed the strongest signal in IHC DAB staining, were also used for detection of BMMF antigen by immunoblotting. I performed this analysis with cryosections from 20 lung cancer patients (tumor and peritumor paired) and 19 non-cancer individuals as well as 7 fresh peritumor after surgeries (which were also tested by IHC). As positive control for WB, I used cell lysate from HEK293TT cells transfected with H1MSB.1 plasmid. WB analysis showed bands at around 37 kDa, which is in line with the expected size of the H1MSB.1 Rep, and a second band at around 45 kDa. I observed a stronger WB detection in peritumor samples compared to tumor samples that showed weaker bands or no bands at all (Figure 18 A, Figure S7).

To confirm which of the two bands detected in the lung lysate corresponded to the band of the positive control, I spiked purified H1MSB.1 Rep into the lysate. This resulted in an overlapping of the bands at around 37 kDa (Figure 18 B).

Moreover, to assess the specificity of the primary antibodies used, one membrane was incubated with Rep AB 10 together with purified H1MSB.1 Rep which resulted in a lower band intensity of the positive control and loss of detection of the bands in the tissue lysates (Figure 18 C).

4.8. Immunoprecipitation with tissue lysates results in increased WB band detection

To enrich Rep protein for immunodetection and mass spectroscopy readouts, lung tissue lysates were subjected to immunoprecipitation. I incubated two tissue lysates (one from tissues adjacent to hamartoma and one from peritumor tissue) with Rep AB10 and AB2 and Protein-A/G beads, allowing immobilization of putative Rep proteins in the lysate by forming antibody-antigen complexes. IP samples were analyzed by WB with Rep AB 3. IP with Rep AB2 revealed a band at the target size which was stronger when compared to the IP with an isotype control antibody (and the input control)

indicating effective immunoprecipitation of the putative Rep target (Figure 18 D).

4.9. Validation of BMMF Rep peptide through mass spectrometry analysis

To validate any BMMF Rep targets in the WB bands after IP, still, both IP samples were tested for presence of Rep protein fragments based on mass spectroscopy (DKFZ Proteomics Core Facility, Dominic Helm, Martin Schneider). In total, 45 identified fragments have been assigned to peptides of the group of BMMF1 isolates. After data processing and additional detailed spectral analysis for specific validation of peptide targets, one BMMF1 peptide was successfully validated (after AB2 IP). The identified and validated peptide sequence LEEQFTQYDIEQISGLSSAYAVR matches to the Rep sequence encoded by the BMMF1 isolates C1MI.2, C1MI.3M.1, C1MI.4, C1MI.5.1, C1MI.5.2, H1MSB.1, Sphinx1.76 and 25BAMI. The sequence is located between the Rep winged helix WH1 and WH2 domain at amino acid position 124-147 in the H1MSB.1 Rep (Figure 18 E).

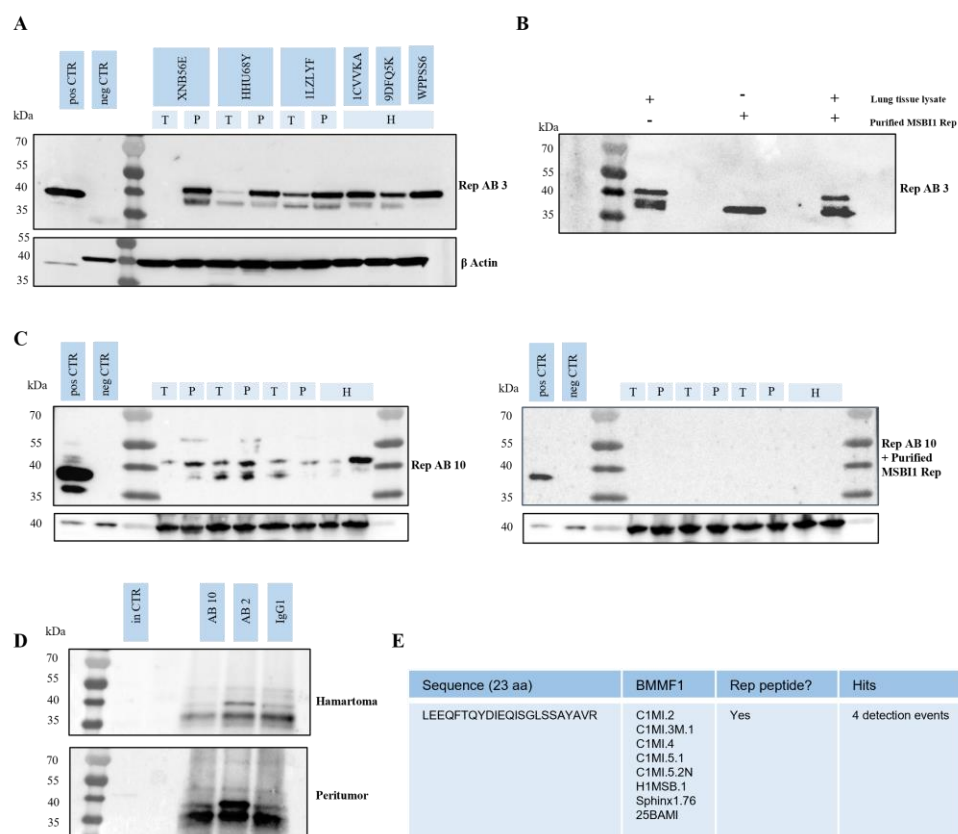


Figure 17. Detection of BMMF protein by immunoblot and mass spectrometry analysis. **A:** WB results for Rep AB 3 for peritumor and tumor tissues from 3 lung cancer patients (n=20) and 3 control tissues from 3 non-cancer individuals (n=19). Frozen tissues used for lysate preparation. Bands are detected at ~37 kDa and ~45 kDa. **B:** Positive control spiked in lung tissue lysate. The band of positive control overlaps the lower band at ~37 kDa of the tissue lysate. Rep AB3 used for detection. **C:** Blocking of anti-Rep antibody detection by pre-incubation of WB membrane with Rep. The pre-incubation resulted in a lower intensity of positive control detection and loss of detection bands. **D:** Detection of WB bands in one hamartoma and one peritumor after immunoprecipitation with Rep AB 10 and 2. Both resulted in bands with an increase of intensity compared to the very weak band of the input control (in CTR). **E:** Mass-spectrometry-validated BMMF1 Rep peptide identified in peritumor sample after immunoprecipitation with anti-Rep antibodies AB2 and AB.

4.10. BMMF reads detected by *in silico* profiling

After identification of BMMF in lung cancer tissues on the level of protein, *in silico* analysis of the PCAWG-WGS (Pan-Cancer Analysis of Whole Genomes – Whole Genome Sequencing) database was done in collaboration with Lisa Häfele (PhD student, D460/F200) who performed the raw data analyses (identification and annotation of BMMF reads).

In the WGS data of 42 Lung Adenocarcinoma (LUAD) patients (42 tumor tissues and 42 distant tissues), BMMF reads were detected in 8 out of 42 patients (19.0%). BMMF hits were found in 5 out of 42 distant tissues

(12.0%) and in 4 out of 42 tumor tissues (9.5%). For the Lung Squamous Cell Carcinoma (LUSC), 48 donors were analyzed (48 tumor tissues and 48 distant tissues), of which 21 (43.7%) showed DNA positivity. BMMF hits were detected in 15 out of 48 (31.2%) distant tissues and 20 out of 48 (41.6%) tumor tissues (Table 19 A). Out of 48 detected reads in total, 40 reads were attributable to the subgroup of BMMF1. The BMMF1 isolate C1MI.3 was the sequence with the highest number of reads, followed by C1MI.2, C1MI.3M.1, H1MSB.2 and C1MI.5.1 (Table S1 A).

BMMF screening in the RNA-seq dataset The Cancer Genome Atlas (TCGA, US) database showed the presence of BMMF reads in 2 patients out of 48 analyzed (4.17%) for LUAD and 5 patients out of 50 (10%) for the Lung Squamous Cell Carcinoma (LUSC) (Table 19 B). No differentiation for tumor and distant tissues was available in this case. CIHB.4 is the template with the highest number of reads detected, followed by C1MI.3M.1 (Table S1 B).

Table 11. *In silico* BMMF screening in lung cancer sequencing archives. **A:** Detection of BMMF reads in lung adenocarcinoma and lung squamous cell carcinoma from the WGS/DNA PCAWG dataset. **B:** Detection of BMMF reads in lung adenocarcinoma and lung squamous cell carcinoma from the RNA-seq PCAWG dataset

A

Lung Adenocarcinoma (LUAD) – WGS/DNA	
8 patients out of 42 (19%)	4 tumour tissue out of 42 (9.5%)
	5 distant tissues out of 42 (12%)
Lung Squamous Cell Carcinoma (LUSC) – WGS/DNA	
21 patients out of 48 (43.7%)	20 tumour tissues out of 48 (31.2%)
	15 distant tissues out of 48 (41.6%)

B

Lung Adenocarcinoma (LUAD) – RNA-seq
2 files out of 48 (4.17%)
Lung Squamous Cell Carcinoma (LUSC) – RNA-seq
5 files out of 50 (10%)

To validate the detection of BMMF hits in RNA-seq screening, I then performed RT-qPCR using frozen tissues from lung cancer patients (n=14) and non-cancer individuals (n=14) with the BMMF1 primer CMI5.240_172. 28 out of 42 samples showed borderline CT detection levels (average CT value=32.7); samples with CT value ≥ 35 were considered negative. Δ CT was calculated as the difference in CT values of my gene of interest and endogenous control. No significant changes were detected between Δ CTs of control, peritumor and tumor (Figure 19).

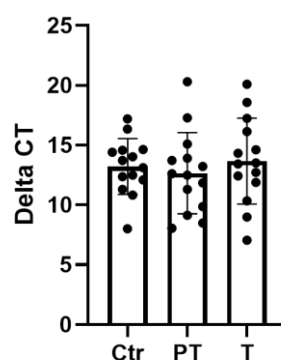


Figure 18. Detection of BMMF1 based on RT-qPCR. RT-qPCR was performed based on tumor and peritumor frozen tissues from lung cancer patients (n=14) and frozen tissues from non-cancer individuals (n=14) with BMMF1 primer CMI5.240_172. Detection of the house-keeping gene GAPDH was used as a control. Average CT value = 32.7; samples with CT value ≥ 35 were considered negative. Delta CTs were calculated by subtracting the CT values of the CMI5.240_172 gene from the CT values of the house-keeping gene. One-way ANOVA test was used for the statistical analysis, no significant changes found.

To increase sensitivity of readout and enrich putative BMMF-containing targets/cells, FACS sorting of cells was performed based on fresh lung tissues homogenates. The sorting strategy was focusing on sorting macrophages (CD45⁺CD68⁺), T-cells (CD45⁺CD3⁺) and epithelial cells (CD45⁻EpCAM⁺). We extracted RNA from sorted cells and the RNAs were sent for bulk sequencing at the DKFZ WGS Core Facility. Screening for BMMF hits was done in collaboration with Lisa Häfele (PhD student, D460/F200). One hit for C1HB.4 was specifically detected in RNA from CD45⁺CD68⁺ sorted cells. No BMMF hits were observed in T cells (CD45⁺CD3⁺) and epithelial cells (CD45⁻EpCAM⁺) (Table 13).

A comparison of gene expression levels based on CD45⁺CD68⁺ sorted cells revealed that the sorted BMMF-positive macrophages were more associated

with a M2-like macrophage phenotype (P=0.03) when compared to BMMF-negative MFs (Figure 20).

Table 12. BMMF hits in sorted cells. CD45⁺CD68⁺ sorted cells were used for RNA extraction. One hit for C1HB.4 was specifically detected after Bulk RNA-seq.

	HCBI8.215	C1HB.4	C2MI.1A.1	H2HB1	MSSI1.16
CD45 ⁺ CD68 ⁺	4	1	1	1	0

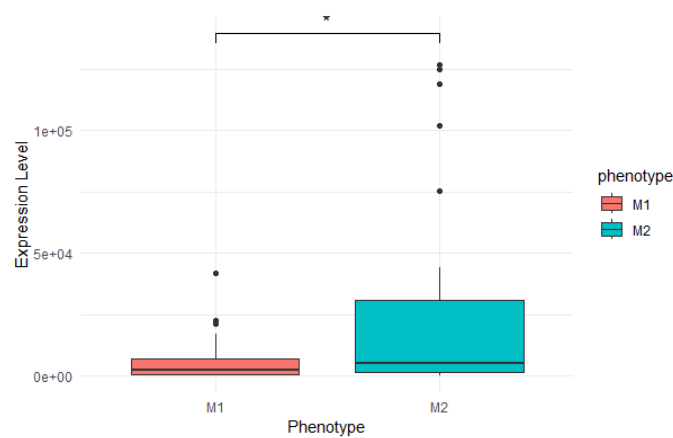


Figure 19. Gene expression levels for M1- and M2-like genes. Comparison of gene expression levels between M1- and M2-like macrophages based on CD45⁺CD68⁺ sorted cells that gave BMMF positivity in Bulk RNA-seq. The data were analyzed using gene counts from a sorted cell sample, with gene expression levels calculated according to predefined gene lists for M1- and M2-like macrophages. Mann-Whitney U test was used for the statistical analysis, P=0.03.

We next performed an *in situ* hybridization preliminary experiment based on MSBI1, MSBI2, C1MI.3M.1 and C1HB.4 sense and anti-sense RNA probes on a strongly IHC-positive peritumor tissue section from fresh lung tissue. MSBI1 and MSBI2 sense and antisense showed staining (Figure 20) localized in condensed *foci* in the cytoplasm close to the nuclei, resembling ribosomes. The staining was in general stronger with the reverse probe compared to the forward probes. Negative (dapB) and isotype control did not show any staining, while the positive control (PPIB) showed speckled foci close to the nucleus in almost all cells. The probes C1HB.4 and C1MI.3M.1 either showed high background or were negative.

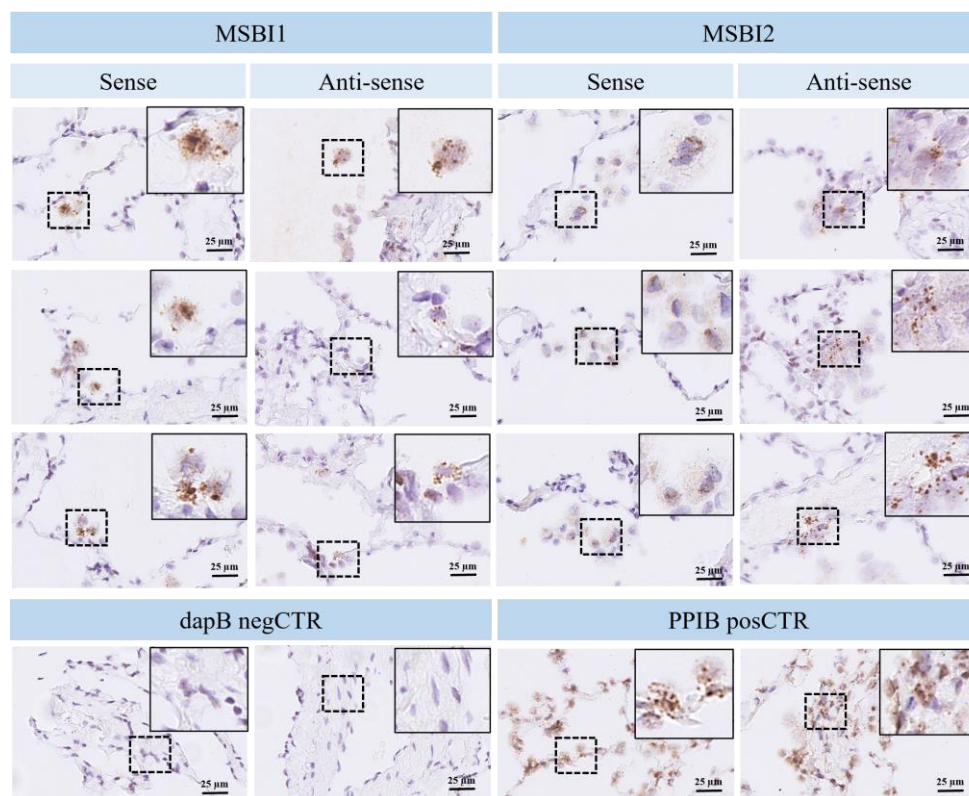


Figure 20. Histologic detection of BMMF RNA by in situ hybridization in peritumor of lung cancer patients. Peritumor lung tissues (n=6) were stained with RNA sense and anti-sense probes based on MSBI1, MSBI2, C1MI.3M.1 and C1HB.4. MSBI1 and MSBI2 probes showed specific staining while C1MI.3M.1 and C1HB.4 showed too high background or were negative. dapB negative and PPIB positive control were also included as well as one isotype control.

4.11. BMMF reads detected in publicly available scRNA-seq dataset

In immunohistochemistry (DAB and IF) analysis of lung cancer tissues, Rep positivity was predominantly observed in the peritumor tissues and specifically in CD68⁺ CD163⁺ M2-like AMs. To further support these findings, we utilized a publicly available scRNA-seq dataset from a study by Maynard et al. [115] to investigate the presence of BMMF(1) RNA in various cell types prepared from the tissue (homogenates) of lung cancer patients. This was done in collaboration with Lisa Häfele (PhD student, D460/F200) who performed the raw data analyses.

The scRNA-seq data revealed BMMF-positive cells in 25 out of 30 patients (83%) and 33 out of 49 biopsies (including metastasis). Among the different cell types, T-cells (24) and macrophages (23) showed the highest number of

BMMF hits (Figure 20 A, Figure S8) but when considering only hits for BMMF1, macrophages become the cells with most hits (14 macrophages, 10 T-cells).

Most of the BMMF hits were attributed to the BMMF1 group (Table S2), particularly the BMMF1 representative C1HB.4 (Figure 20 B) and C1ML3M.1 (Figure 20 C). Since, based on the IF triple staining and cell-based quantification, Rep positivity seemed to be present specifically in CD163⁺ macrophages, I clustered specifically M1-like (Figure 20 D) and M2-like (Figure 20 E) macrophages based on a list of 54 and 40 genes respectively (Table S3) [115] and searched for BMMF hits in the two groups (Figure 20 F). Robust linear regression models showed a trend ($P=0.062$) for BMMF hits specifically in M2-like MFs (Figure 20 G).

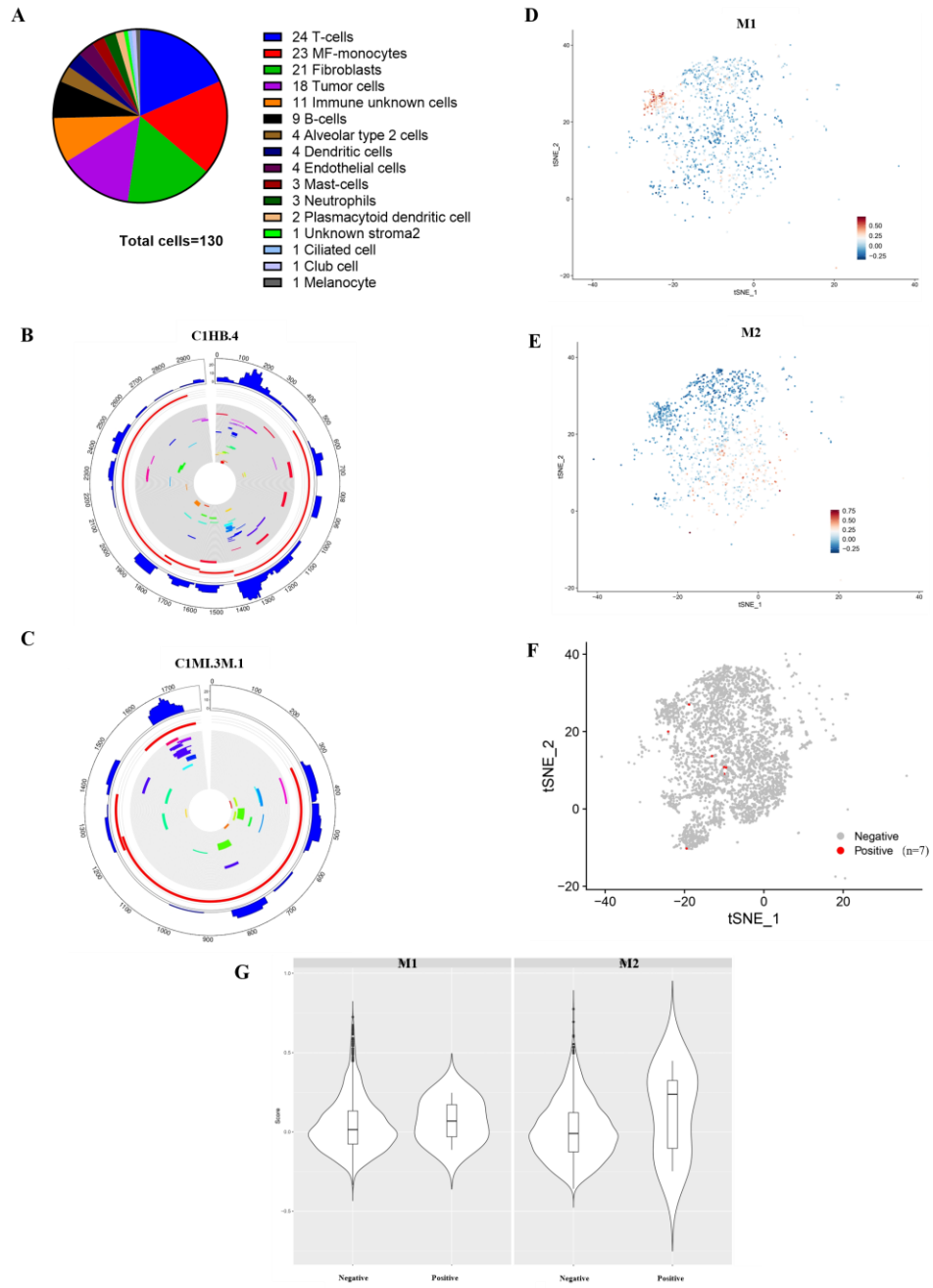


Figure 21. BMMF reads detected in publicly available scRNA-seq dataset. **A:** Out of 23261 cells, 130 resulted positive for BMMF with T-cells and macrophages as the most positive cells. **B:** Circular plot for the BMMF1 representative C1HB.4. ORFs marked in red outer circles. Individual cells with BMMF reads are represented by different colors in the inner circle. **C:** Circular plot for BMMF1 representative C1MI.3M.1. ORFs marked in red outer circles. Individual cells with BMMF reads are represented by different colors in the inner circle. **D, E:** The differentiation into M1- or M2-like macrophages was done using a list of marker genes (Table S3) from the authors of the paper [115]. **F:** tSNE plot showing BMMF1 positive macrophages. **G:** Robust linear regression models were fitted to assess the relationship between BMMF positivity and M1 or M2 scores. P value=0.062.

5. Discussion

5.1. Rep expression is associated with alveolar macrophages

Infectious pathogens play an important role in cancer development. According to de Martel et al., the year 2018 witnessed a global diagnosis of 2.2 million cancers linked to infections, accounting for approximately 14-16% of the overall cancer incidence [65]. When discussing infections as a cause of cancer, it is important to distinguish between direct carcinogenesis and indirect carcinogenesis. In direct carcinogenesis, the persistence of infectious agents' nucleic acid within cancer cells and the expression of specific viral genes contribute to the malignant growth. On the other hand, indirect carcinogenesis involves chronic inflammation, which leads to the production of oxygen/nitrogen radicals able to induce random mutations in the host DNA. In this case, the presence of viral DNA in cancer cells over time is not a requirement [2].

Our group postulated a specific association of several types of cancer with zoonotic-linked infections by meat and milk consumption from specific Aurochs-derived Eurasian dairy cattle (*Bos Taurus*) [17]. In fact, the global epidemiological patterns of colon and breast, but also lung cancer incidence, points to a relation with the consumption of products originating from these specific breeds of cattle. This led to the isolation of over 200 single-stranded circular DNA genomes, named Bovine Meat and Milk Factors (BMMF) [26]. The presence of BMMF was already detected in peritumor M2-like macrophages from CRC patients at both DNA and protein levels and was increased in CRC peritumor when compared to tumor and healthy donors. Strong nuclear staining of 8-Oxo-2'-deoxyguanosine (8-OHdG), a product of DNA oxidation by ROS and RNS was observed in areas with higher Rep presence. This finding supports the model that BMMF contribute to indirect carcinogenesis by chronic inflammation and DNA mutation. Moreover, a correlation between higher levels of Rep expression in tumor-adjacent tissue and an increased risk of CRC-specific death was found [30], [63]. In LUAD tissues, BMMF detection by IEM led to the identification of the same pleomorphic vesicular structures observed in peritumor tissue of CRC and

PDAC patients. Still, it was so far only tested in a small number of patients and no systematic analysis on BMMF detection was performed in the context of lung cancer.

Lung cancer remains the leading cause of cancer-related mortality globally. While approximately 90% of lung cancer cases are attributed to smoking and tobacco use, there are high numbers of cases in non-smokers [66], [73]. Chronic inflammation unrelated to smoking seems to play a significant role in tumor development, mediated by the release of ROS/RNS and TNF- α and can facilitate tumor growth by activating specific pathways and suppressing immune responses [66]. Epidemiological patterns have revealed an increased incidence of lung cancer in countries with high consumption of bovine meat or dairy products [73]. Occupations such as butchers, meat workers, and animal farmers have been associated with an elevated risk while a lower consumption of milk and dairy products (for examples people with lactose intolerance) is associated with a decreased risk [1], [4], [74], [75], [76]. A positive association of high childhood milk uptake with lung cancer risk was confounded by smoking behavior during adulthood [119]. The role of inflammation as a risk factor for lung cancer development is also supported by data showing a 36% reduction in cancer risk among individuals regularly taking non-steroidal anti-inflammatory drugs like aspirin and ibuprofen for at least 6 months [1], [77], [78].

Considering epidemiological patterns of lung cancer incidence and bovine meat and milk consumption as well as the indication of the presence of an additional risk factor for lung cancer related to inflammation, we assessed whether BMMF could be involved in the development of lung cancer and tested for BMMF positivity of different lung cancer sample collections.

Rep IHC (DAB and IF) staining showed Rep expression in all peritumor tissue sections. 7 out of 8 antibodies generated to detected BMMF1 Rep protein (Rep AB 1, 2, 3, 7, 10, 11, 15) have been produced upon immunization with H1MSB.1 Rep protein or peptide thereof. Both antibodies detecting H1MSB.1 Rep (AB 3 and 10) as well as antibodies detecting H1MSB.1 Rep and at the same time additional Reps encoded by other BMMF1-

representatives (consensus antibodies - AB 1, 2, 15) and AB 7 and 11 showed tissues staining. Also, antibodies targeting the Reps of the BMMF-1 isolates H1CB.4 and H1MSB.2, C1MI.3M.1, C1MI.9M.1 revealed staining (AB 21 and 22, 23). Detection with ABs 3 and 10 suggests that indeed H1MSB.1 Rep protein is present in the tissues. Detection with consensus ABs confirm this finding, but also suggests that Rep of additional types of BMMF-1-like isolated might be present in the tissues, which might be exemplified by the finding that competition experiment with purified H1MSB.1 Rep during staining did not result in a complete neutralization of staining but in reduction of around 55%. In total, Rep antibodies which showed tissue staining cover five different epitopes suggesting, altogether, a reliable detection of Rep protein in the tissue: AB3 and 10 detect an epitope at the C-terminal of the Rep protein, AB1 has its epitope in the central WH2 domain of Rep and AB2 and 15 detect an epitope at the N-terminal WH1 domain.

The staining was localized in the cytoplasm of a specific population of cells and appeared organized in distinct, speckled *foci*. This is in line with the observations of speckled structures in the peritumor tissues of LUAD patients detected by IEM [64].

Isotype control staining showed no or only a significantly reduced staining, which did not show the typical brownish DAB colour but a more greyish colour. To additionally confirm specific detection via IHC, consecutive lung cancer peritumor tissue sections were immunohistologically stained with Rep AB 3 pre-incubated with purified H1MSB.1 protein, showing a 55% of reduction of the staining.

If BMMFs may contribute to cancer-promoting chronic inflammation by interaction with macrophages also in lung cancer, a correlation between Rep and CD68⁺ macrophages could exist.

On consecutive tissue cuts in DAB staining, the tissue regions showing high Rep expression were indeed overlapping with the regions with high CD68 expression and, in co-immunofluorescence analysis, a congruent CD68 and Rep staining was observed in the cytoplasmic compartment of Rep⁺CD68⁺ double-positive cells. Comparison of the numbers of double positive

CD68⁺Rep⁺ cells to Rep⁺ cells, which were on a comparable level, demonstrated that the Rep antigen is almost exclusively localized in CD68⁺ cells, identifying CD68⁺ cells as the putative BMMF target cells in the lung. The number of CD68⁺ cells was generally higher when compared to the Rep⁺ cells in all peritumor tissues sections, which is explainable in large part by a generally high number of alveolar macrophages in the alveoli and lung parenchyma [120].

Based on morphological characterization, we assume that Rep⁺CD68⁺ macrophages correspond to alveolar macrophages, which are larger and round shaped and specifically localize to the surface of lung alveoli. Conversely, Rep⁻CD68⁺ macrophages may represent a population of interstitial macrophages, characterized by a smaller and less rounded shape [118]. Alveolar macrophages (or dust cells) are a common type of myeloid immune cells with phagocytic activity, found in the airways and at the level of the alveoli in the lungs [120]. AMs are primarily responsible for the first line of defense in the alveoli and airways. They play a crucial role in scavenging pathogens such as bacteria, viruses, and environmental particles through phagocytosis [80]. This process is initiated by the interaction of toll-like receptors on AMs with pathogen-associated molecular patterns on microbes, leading to the engulfment of pathogens and subsequent secretion of pro-inflammatory cytokines to enhance the immune response. IMs serve as gatekeepers of the vasculature and lung interstitium. They have a distinct role compared to AMs, focusing on maintaining homeostasis and regulating immune responses in the lung tissue. While they also possess phagocytic capabilities, their primary functions include tissue repair and modulation of inflammation rather than direct pathogen clearance [80]. Since AMs are the ones mostly involved in phagocytosis activity, they might be predisposed for a first encounter with BMMF. Moreover, IMs (derived from bone marrow-derived monocytes) are continuously replenished by circulating monocytes, while AMs (tissue-resident macrophages that originate primarily from foetal liver monocytes) have generally a very long half-life ranging from months to years with minimal replenishment from circulating monocytes [121]. This can result in longer BMMF exposure for AMs. Also, the extended lifespan of

AMs enables them to undergo significant reprogramming in response to environmental cues, such as viral infections. For example, studies have shown that AMs can be "trained" by respiratory viral infections like influenza, leading to long-lasting changes in their function, including altered responses to pathogens [122].

5.2. Rep expression is increased in case versus control

While all peritumor sections showed Rep expression in IHC, the paired tumor sections were mostly negative. Only 28% of them showed Rep signal and even in this case, the intensity was in general weaker compared to peritumor sections. This finding was also supported by the general appearance and quantification of bands by WB, where the tumor samples exhibited weaker bands compared to the paired peritumor samples.

Control tissues showed an intermediate level of Rep positivity in IHC and strong bands in WB analysis. Control individuals used in this project were prepared from non-cancer hamartoma patients, a benign tumor that is usually asymptomatic and rarely linked to inflammation [114]. Hamartomas originate from mesenchymal cells and consist of varying amounts of mature cartilage, fat, respiratory epithelial cells, osteoid, and smooth muscle tissues. Although pulmonary hamartomas are benign lesions, the risk of lung cancer in lung hamartoma patients increases at diagnosis or at follow-up [123], and this might explain the high presence of BMMF in hamartoma patients. Only cells from tissue tissue-adjacent to hamartoma were quantified in IHC analysis.

Nevertheless, cell-based quantification showed a significant increase in the number of both Rep⁺ cells and CD68⁺Rep⁺ cells in peritumor compared to control tissues. CD68⁺ cells were significantly higher in peritumor compared to controls tissues while CD68⁺Rep⁻ cells are similarly high in peritumor and controls suggesting a possible usage of Rep⁺CD68⁺ cell populations as possible biomarker

5.3. Are Rep⁺ macrophages biomarkers for lung cancer?

Despite advancements in diagnosis and therapy over the past 25 years, the prognosis for lung cancer patients remains unsatisfactory. A work highlighted

the important role of alveolar macrophages in tumor progression and their role as targets in early treatment for lung cancer [84]. Alveolar macrophages accumulate near tumor cells during the early stages of tumor development (tumor associated macrophages – TAM) and induce an epithelial-mesenchymal transition (EMT) program, promoting cell invasiveness. Additionally, alveolar macrophages stimulate a Treg cell response, protecting tumor cells from CD8⁺ T cell-mediated killing [84]. TAMs have been shown to express an M2-like phenotype [86]. Moreover, Clodronate liposomes have been proved to effectively reduce TAMs and inhibit tumor growth in mouse models [124]. In fact, Clodronate treatment can lead to significant tumor growth inhibition by depleting macrophages that support tumor progression [124].

In line with previous findings [125], we observed an increase of the number of M2-like macrophages in peritumor tissues compared to age-matched non-cancer individuals. A non-significant trend for increase of M1-like macrophages was also observed, still in line with the general increase of M2- and M1-like macrophages in cancer patients compared to non-cancer patients [125]. Most of the M2-like MFs were Rep positive (4.6% median), suggesting that indeed a large proportion of M2-like cells is exposed to BMMF or might have undergone transition toward a M2-like differentiation. In general, a significant shift towards an M2-like MF phenotype was observed in CD68⁺Rep⁺ cells in peritumor tissues. Importantly, for Rep⁻ M2-like macrophages, no differences were observed between controls and peritumor tissues. The number of Rep⁻ M1-like macrophages is significantly higher in peritumor tissues, compared to the number of Rep⁻ M2-like macrophages, which may suggest an active immune response against BMMF clearing BMMF in M1-like MFs. In general, high heterogeneity is observed within the cohorts. This is most likely because of high variation in the number of MFs (both M1- and M2-like) in the different patients due to cancer (sub-)type and stage, patient age and genetic factors as well as tumor microenvironment [126]. Our findings are still in line with existing literature, where also heterogeneity in the number of macrophages in different patients is describes.

This general situation of case vs. control sensitive BMMF–macrophage cell populations would generally support the proposed model of inflammatory-driven, indirect cancer induction [21]. The process would involve mechanisms affecting large areas of normal tissue regions over several decades, where the infectious agent does not persist in the cells and where the malignant growth propagates independently of the infectious agent but coincides with chronic inflammatory reaction in the affected tissues [2], [21], [53]. In line the previously proposed model of BMMF-driven indirect carcinogenesis [30], [63], the infectious risk factors are not present and do not persist in the tumor cells. Instead, upon initial and ongoing infection, they induce pathogenic modulations over long-latency periods including chronic inflammation and induction of radicals diffusing over long distances up to 10-20 micrometers or even longer depending on tissue composition. The radicals eventually also reach replication-competent stem and early daughter cells (basal cells in the larger airways and alveolar type II cells in the alveolar regions) - which themselves are BMMF negative and were never in contact with BMMF (no tissue staining observed in this type of cells) - where they cause random DNA mutations that might lead to cancer development over the years.

5.4. Evaluation of possible association of BMMF expression with sex, age and pathology of the tumor

Possible associations of Rep expression with age, sex and pathological status of the tumor were investigated (Figure S5). Stratification of quantified stained cell populations for males and females did not result in any significant changes, neither for (BMMF-negative) immune cell populations nor for BMMF-positive cell fractions between the sample groups. This might indicate that the exposure/uptake of BMMF is independent of the sex (hormones etc.) and that dietary/consumption habits which might be different for males and females do not translate into a significant change of tissue expression patterns. Lung cancer incidence is reported to be 27% higher among males compared to females within the recent years [127] – so this discrepancy is not followed up or attributable to BMMF expression, in

general. The cohort analyzed in this thesis reflect this general trend including higher number of men compared to women (31 males compared to 16 females).

A slight but non-significant trend for increased detection of Rep⁺ cells was observed in older patients (especially above the age of 65), which did not depend on sex. This could be explained by the observation that, in some cases, the pathogen load could increase over time, especially if the function of the immune system is compromised or if the individual is exposed to additional sources of (synergistic) infections [128].

When investigating the possible association between Rep expression and lung pathology, the two major types of lung cancer were considered: NSCLC and SCLC. Between the two groups, no significant differences of Rep expression were observed, but there is a small trend for Rep expression being increased in SCLC, while NSCLC presented more heterogeneity and a stronger discrepancy of high and low responders. This might indicate possible subclasses in NSCLC which might show different association levels with BMMF.

In fact, NSCLC is a very broad and diverse group that includes any type of epithelial lung cancer other than SCLC. The most common types of NSCLC are LUSC, LCC, and LUAD, but there are several other types that occur less frequently, and all types can occur in unusual histological variants [129]. Even if NSCLC patients are the majority of our cohort (n=19), for several patients we did not have a specification about the type of NSCLC, so we could not further stratify the analysis in LUAD and LUSC. In general, about 10% of NSCLC cases are unrelated to smoking [130]. Between the subgroups, LUSC is the subgroup which is strongest linked to smoking, while most of non-smokers, specifically women, are diagnosed with LUAD (only 5 cases of LUAD were included in the cell-based quantification, 1 of which is female), which is supported by literature indicating that NSCLC and especially LUAD are more likely to be linked to inflammation-derived carcinogenesis [130].

SCLC (12-15% of all lung cancer cases) in light/never smokers is rare and comprises approximately 2.5% of cases [131]. These patients are more likely

to be female of 35 to 49 years or ≥ 80 years. The incidence of SCLC appears to be rising in never-smokers, especially in Asian populations [131]. Never-smoker SCLC patients have some unique clinical and molecular characteristics compared to their smoking counterparts [132]. Risk factors for SCLC in light/never smokers are not well understood but occupational exposures (like exposures to asbestos, silica dust, chromium, nickel, diesel exhaust and polycyclic aromatic hydrocarbons) are listed among the contributing factors [133]. It was already shown that the risk for lung cancer is increased in butchers and animal farmers [74], [75] pointing towards BMMF to be one of the risk factors. Moreover, the histologic transformation of SCLC from NSCLC is becoming more frequent, with an incidence of up to 15%, so early NSCLC formations can develop further into the more aggressive SCLC [133].

5.5. Heterogeneity of Rep expression in low-dose smokers

Cigarette smoking is the major risk factor for development of lung cancer and the leading cause of chronic obstructive pulmonary disease [134]. The disease process involves oxidative stress, inflammation, proteinase/anti-proteinase imbalance, tissue destruction and inadequate repair; and AMs play an important role in these processes [135]. Smoking induces an increased number of immune cells (mostly neutrophils and MFs) in the lung tissues, at least until a certain smoking intensity [134]. There is in fact a state of immune exhaustion or tolerance that can be built in heavy smokers [136]. However, the relationship between the number of myeloid cells (and macrophages in particular) and tobacco consumption or the number of cigarettes smoked is not well understood [134]. A work based on COPD patients showed that the total number of macrophages may decrease with higher smoking exposure [137]. Our findings show that there is indeed a trend for an increase of M2-like and especially M1-like macrophages in low-dose smokers compared to high dose smokers.

For ex-smokers the situation is less clear due to the high heterogeneity observed in our analysis. This heterogeneity can be explained by the findings that the effects and responses of smoking on adaptive immunity persist for

about 10-15 years after quitting smoking [138] and the unavailability of the information on the time passed since patients quit smoking and timepoint of sampling in our study, which might obscure any associations.

In smokers, there is a trend for an inverse correlation between the amount of Rep⁺ cells and smoking intensity (indicated as packages/year, PY) suggesting a higher number of Rep⁺ cells in low-dose smokers. The situation is different in ex-smokers, where we observed more heterogeneity, in general, and where two different populations are distinguishable: i) PY^{low}Rep^{low} and ii) PY^{low}Rep^{high}. The first situation (PY^{low}Rep^{low}) might either reflect the immune system that is returning from exhaustion to a more normal state after quitting smoking (and in some cases it might start clearing up infections [BMMF] again after a period of exhaustion with re-increasing numbers of M1-Mfs), or it might be the consequence of general life and/or dietary adjustments of the individuals to improve the general fitness/health levels in parallel to quitting smoking. In contrast, other individuals might still quit smoking but continue with a less healthy lifestyle including e.g. high consumption of red meat and milk, which might still result in continuous and high BMMF exposure (PY^{low}Rep^{high}).

Although no significant differences of Rep (Rep-MF) populations were observed between different smoking levels, the trends for anti-correlation of BMMF-positive cells with smoking intensity support the speculation that for low-dose smokers (and ex-smokers) (PY^{low}Rep^{high}), BMMF infections might contribute to carcinogenesis and increase the cancer risk (even after quitting smoking).

5.6. Detection of BMMF hits by immunodetection and mass spectroscopy and by *in silico* BMMF profiling

To support histological tissue detection of BMMF Rep protein in lung cancer tissues, we performed western blot detection based on cryosamples from a subset of the same patients included in the IHC study (for 14 patients cryosamples were not available).

WB analysis led to the detection of two bands: one at the expected size of H1MSB.1 Rep (at around 37 kDa) and a second band at around 45 kDa.

Spiking in purified H1MSB.1 Rep into the samples, the band from the purified protein aligned with the 37 kDa band, confirming that the 37 kDa band corresponds to H1MSB.1 Rep in the lysate sample matrix. Moreover, co-incubation of the antibody solution with purified H1MSB.1 Rep led to a reduction of the intensity of the positive control signal and the loss of both the 37 kDa and 45 kDa bands. This suggests that the antibody specifically recognizes both bands, indicating that the 45 kDa band is also related to H1MSB.1 Rep. The second band at 45 kDa could be the result of a post-translational modification of BMMF Rep, such as glycosylation, phosphorylation, SUMOylation or ubiquitination, which increases the molecular weight. Another possibility is that the second band reflects an isoform of the BMMF Rep due to alternative splicing or expression of protein isoforms that differ from those found in *in vitro* systems or a different (larger) type of Rep. Last, although the monoclonal antibodies used (AB 3 and 10) detect H1MSB.1 Rep, it is possible that they cross-react with similar proteins encoded by other BMMF isolate not yet identified. The running difference might also be explained by the presence of SDS-resistant (secondary structure) parts of the protein affecting running properties.

In general, the bands intensities reflected the IHC findings: stronger bands were detected in peritumor compared to the paired tumor tissues, but the differences were smaller which might be due to the less clear status of the tissue (tumor vs. peritumor content) which is more easily identifiable by histology. Tumor samples exhibited faint bands or no bands while hamartoma samples showed detection of bands of similar intensity of the peritumor, still suggesting that hamartomas might not be the adequate controls and that there is the need to acquire tissue samples from healthy individuals. [139]. It is plausible that the BMMF isolates found in hamartoma might be low risk isolates and not involved in lung cancer development. This situation could also reflect what is already known for HPV virus, in which different variants exist but only some of them are related to cancer development [139].

To validate any BMMF target detected by WB, MS analysis was conducted with one peritumor and one hamartoma sample, after enriching BMMF Rep protein by IP. This led to the validation of one BMMF Rep peptide

specifically in the peritumor sample. No BMMF hits were found in the hamartoma sample. The validated peptide, LEEQFTQYDIEQISGLSSAYAVR, is localized between the WH1 and WH2 domain of the Rep protein and matches the Rep sequence of the BMMF1 isolates C1MI.2, C1MI.3M.1, C1MI.4, C1MI.5.1, C1MI.5.2, H1MSB.1, Sphinx1.76 and 25BAMI, further supporting the presence of not only H1MSB.1 Rep, but also other BMMF1 Reps like already showed by IHC. The identified peptide sequence does not contain any epitope of our produced monoclonal anti-Rep antibodies.

To confirm wet-lab results with *in silico* BMMF detection, we screened for BMMF reads in different cancer sequencing archives. The *in silico* screening of the lung cancer American cohorts (LUAD and LUSC) from the PCAWG project showed detection of BMMF reads in both WGS (LUSC=43.7% of patients with BMMF reads, LUAD=19%) and RNA-seq (LUSC=10%, LUAD=4.17%) datasets. In general, BMMF1 reads were mostly found in LUSC, where a total of 40 reads were identified, against only 1 BMMF2 read. In particular, C1MI.3, C1MI.2 and C1MI.3M.1 were the BMMF1 isolates for which the highest number of reads was identified in WGS. In LUAD there is a lower detection of BMMF reads, with only a detection of 4 BMMF1 and 4 BMMF2 reads. The situation is similar in RNA-seq, where again BMMF reads were mostly identified in LUSC (14 BMMF1 reads, no reads for BMMF2) while in LUAD only 2 reads for BMMF1 and 4 for BMMF2 were detected. CIHB.4 and C1MI.3M.1 were the BMMF isolates for which the highest number of reads was identified. Interestingly, reads associated with CIHB.4 were also found in bulk RNA-seq upon sorting of CD68⁺ cells from fresh lung peritumor tissue. While the screening of both cohorts resulted in detection of BMMF reads, the majority of reads were detected in WGS dataset. WGS captures all DNA sequences, regardless of any (bioactive) gene expression status. The higher positivity can be explained by the fact that the WGS cohort contains also sequences that are present on the level of DNA but not necessarily transcribed into RNA. In contrast, RNA-seq data only reflects sequences that are actively transcribed into RNA. RNA molecules which did

not originate from the host are unlikely to persist so long and endure infection. Thus, the limited but available detection of BMMF RNA reads might indicate limited but present BMMF transcription in a small set of BMMF permissive cells.

RT-qPCR showed only borderline BMMF detection and no significant differences between case vs. control and peritumor vs. tumor were detected but only a small trend of stronger BMMF1 detection in peritumor, followed by hamartomas with the weakest detection observed in tumor. This result might reflect the impossibility of knowing the exact tumor-peritumor content of the cryosections used for the analysis. Another explanation for the discrepancy between the protein expression levels observed in IHC and WB analyses and the minimal changes detected at the mRNA level by RT-qPCR is post-transcriptional regulation, where mRNA levels do not directly correlate with protein levels due to various regulatory processes such as RNA stability, translational efficiency, or post-translational modifications. Proteins often undergo different rates of degradation, stabilization, or modifications that can significantly alter their levels independent of mRNA expression. Alternatively, BMMF protein originating from non-lung tissues (cells) or which might be already present during infection (as nucleoprotein particles/vesicles) might be responsible for BMMF protein detection in the lung tissues – without any countable RNA transcription.

In situ hybridization was used to gain more information about detection of a possible microscopic presence of BMMF RNA in the tissues. The positive signals observed in the RNA *in situ* hybridizations in alveolar MFs indicate the presence and localization of BMMF RNA transcripts within the same IHC-stained tissue samples and cell types suggesting active transcription in these cells. This positivity was observed for MSBI1 and MSBI2 sense and antisense probes, with the antisense probe resulting in strongest detections. This result supports the understanding that a reverse probe generally detects the sense RNA strands, which are more abundant and represent the mRNA that is translated into protein. Moreover, all the probes used for the ISH were designed to include the Rep sequence, so the RNA that is detected in the

tissues is the one specifically translating for Rep protein. The observed DAB staining upon *in situ* hybridization suggests the expected localization of the putative BMMF RNA in the cytoplasm and close to nucleus, most likely in ribosomes. The staining partly resembles the speckled-like staining of the classical anti-Rep protein IHC, which might indicate BMMF RNA at the site of the cytoplasmic protein speckles and in ribosomes. *In situ* hybridization with negative controls (DapB) and additional BMMF sense and anti-sense probes (HCBI.4 and C1MI.3M.1) did not result in any detection, while the staining with positive control (PPIB) resulted in a strong and speckled detection in the entire tissue section in the cytoplasm of cells, supporting successful hybridization and staining, in general. The probes have generally been designed to hybridize with a large (final) part of the encoded Rep gene (HCBI.4 and MSBI1) or with the entire Rep gene (MSBI2 and C1MI.3M.1). Thus, tissue staining with such a probe suggests presence of this crucial, potential transcript highlighting again the importance of the BMMF Rep, in general, and a possible bioactive transcription.

Despite the lack of HCBI.4 detection by ISH, HCBI4 RNA was detected in CD68⁺ sorted cells through bulk RNA-seq and also in the RNA-seq dataset of the PCWAG cohort. This unexpected result could be attributed to, for example, post-transcriptional modifications, RNA stability or degradation, suboptimal probe design (e.g., low specificity due to secondary structures or homology with non-target regions), and technical issues like inadequate probe penetration, suboptimal hybridization conditions, or low RNA integrity could have reduced the RNA levels detectable by RNA-ISH, resulting in a weak or absent signal.

In general, BMMF DNA detection is observed by the screening of PCWAG WGS dataset, while RNA is identified in several cases by RT-qPCR, *in situ* hybridization, scRNA-seq (see below) suggesting that both types of nucleic acid molecules are identifiable supporting transcriptional activity based on cells. In particular, RNA detection would indicate active production of BMMF components and lower risk of putative contamination of other sources of DNA/RNA.

5.7. Trend for association of Rep⁺ cells with “M2-like” macrophages in scRNA-seq

The obvious localization of Rep expression especially in peritumor (alveolar) macrophages prompted intensifying characterization of the differentiation of any BMMF-positive immune cells for identification of putative pathogenic (immune) regulation patterns by scRNA-seq [116]. *In silico* BMMF profiling in scRNA-seq revealed BMMF positivity in 130 cells, in total, including metastasis sites and all BMMF groups (BMMF1-4). The cell type with the highest amount of BMMF reads were T-cells (n=24) and macrophages (n=23).

When refining the analysis focusing only on macrophages of the primary lung cancer tumor (excluding metastasis) and only on BMMF1, cells with highest number of BMMF1 hits were macrophages (n=14). In addition, there is a trend for BMMF1 hits to be specifically present in M2-like macrophages. This trend was not evident when considering both primary tumor and metastasis samples and all BMMF groups (Figure S9), suggesting a more specific association between BMMF1 detection with M2-like macrophages specifically in the lung cancer context. This observation may indicate a selective enrichment of BMMF1 in M2-like macrophages. Still, the number of positive cells is inadequate for statistical analysis and further investigations are needed with other scRNA-seq dataset to validate these findings and explore the differences in gene expression between BMMF1 positive and negative (M2-like) macrophages to understand whether BMMF1 influence macrophages polarization, or whether BMMF preferentially infect (already differentiated) M2-like macrophages. Also, the localization of BMMF-positive cells in tSNE plots was not within the clusters of the better defined M2- (or M1- like) MFs but somewhere in between these groups, indicating that a completely new type of differentiation pattern might be associated with BMMF.

It is known that AMs can potentially be reprogrammed into an M2-like phenotype by viruses or pathogens [140]. For example, certain viruses can manipulate the immune response, including the polarization of macrophages.

If BMMF are involved in macrophages reprogramming, they could induce the expression of cytokines such as IL-4 or IL-13, which are known to promote M2 polarization in macrophages. This is a strategy adopted by several viruses and could also push a macrophage with M1-like phenotype to switch to a M2-like phenotype. This reprogramming has significant implications for immune evasion by pathogens but also for the tumorigenic potential of M2-like macrophages [83], [140].

In conclusion, BMMF1 positivity was detected by antibody staining (IHC) in CD68⁺(CD163⁺) macrophages of lung peritumour tissues and was enriched in case vs. control suggesting use as a cancer biomarker and a possible causal contribution. These findings were also supported by WB detection and MS analysis. While BMMF DNA was only found in *in silico* screening, BMMF RNA was also detected by ISH, Bulk RNA-seq and, with borderline positivity, in RT-qPCR. The identification of Rep⁺ macrophages in scRNA-seq, with the addition of further scRNA data, might enable to study expression changes induced by BMMF.

5.8. Limitations of the study

Despite identification of BMMF as putative cancer risk factor and biomarker based on multi-omics analyses, the most obvious limitation of this thesis is the lack of healthy individuals to use as adequate controls due to lack of availability. Although hamartomas, which were used as controls in this study, are benign tumors, the risk of developing lung cancer in hamartoma patients can be increased [123]. Future studies might and should therefore include e.g. blood or exhaled breath condensate (EBC) samples from healthy individuals as an alternative with easier access. Analysis of blood samples from lung cancer patients and healthy individuals is, in fact, still lacking, even though BMMF is found in diverse organs, and this might indicate a presence in the blood stream as well. Collection of EBC samples is a noninvasive method for obtaining samples from the lungs [141] as opposed, for examples, to bronchoalveolar lavage (BAL) that are samples collected with a method perceived as unpleasant in which a saline solution is put through the

bronchoscope to wash the airways and a fluid sample is captured [142]. Because of that, this method is not well accepted by patients. Collection of EBC could be repeated several times with short intervals between sampling and the collection device can be portable, assuring a longitudinal sampling. Moreover, this study did not consider the possible presence of COPD in patients. This is an inflammatory condition that could alter the number of immune cells (macrophages) in patients. Even though COPD should be taken into account in future analysis, its prevalence among smokers is not as high as one might expect, with only 15% of smokers developing the COPD [143]. The geographic origin of the cohort was also limited; all the patients included in the analysis were from Germany (Heidelberg, *in silico* analysis were performed with American cohorts) and future studies should include a more diversified cohort to determine if the current Rep antibodies are effective everywhere. This is especially important because there are over 20 known isolates in the same group as H1MSB.1 [24], [26], and these isolates might have different distributions worldwide and could contain different Reps, which might not all be detected by the current antibodies.

The limitation of the size of the lung (cancer) cohort used for this thesis did not allow for sub-differentiations and more specific analysis of Rep association with clinical parameters (e.g. smoker women vs. ex-smoker women vs. non-smoker women, LUAD vs. LUSC, etc.), also, the PY definition (especially for ex-smokers) is not well defined. Moreover, for some patients, the PY information was lacking, not allowing the inclusion of such patients in the low-high dose smoking discrimination. The impact of smoking in lung cancer is of course very high, but in the last years the diagnosis of lung cancer in non-smokers has increased, mostly in young women, and this seems to point towards another lung cancer risk factor apart from smoking [113]. Still, most of the lung cancer patients are either smokers or ex-smokers and this might complicate the investigation to test BMMF as an additional risks factor for lung cancer.

Another limitation of this study is the inability to isolate BMMF DNA from lung (cancer) tissues, even though DNA sequences were identified in *in silico* analysis (PCAWG WGS). One reason could be that the DNA may be present

in very low quantities, below the detection limit of our extraction and amplification techniques. Second, the DNA might be fragmented or modified in a way that makes it difficult to isolate.

The amount of BMMF positive cells detected with scRNA-seq is not adequate to perform a statistical analysis about the association between BMMF1 positivity and specific macrophage subtypes. Also, the limited cell numbers can increase variability and potential bias in the results. If the sequencing depth is insufficient, some positive cells may be lost.

6. Conclusions and outlook

BMMF represent a group of plasmid-like DNA molecules frequently identified in bovine milk and blood which have been described as zoonotic-linked infectious agents and cancer risk factors associated with induction of chronic inflammation and, consequently, the formation of cancer upon indirect carcinogenesis. BMMF group 1 representatives encode for a conserved protein (Rep). Antibody staining against Rep identified increased Rep expression in the peritumor of CRC patient tissues when compared to the mucosa of healthy controls suggesting a role as (causal) cancer biomarker in CRC. In this work, I studied the presence of BMMF1 and their possible contribution on the development of lung cancer by multi-omics approaches. IHC DAB staining showed Rep expression in the peritumor tissues of all lung cancer patients tested (n=42). The staining was concentrated on cytoplasmic speckles in the cytoplasm of alveolar macrophages and was almost absent or weaker in coupled tumor tissues. Association of Rep expression with CD68⁺ MFs was confirmed based on DAB staining of consecutive tissue cuts and by co-immunofluorescence microscopy, where a congruent staining of Rep and CD68 localized in the alveolar surface-associated, bigger and round-shaped AMs but not in the IMs. Rep⁺ (and CD68⁺Rep⁺) cell numbers were significantly increased in the peritumor of lung cancer patients when compared to tissues taken adjacent to hamartoma (benign lung tumor, usually asymptomatic and rarely linked to inflammation), used as controls in our analysis. This case vs. control difference and correlation of Rep with CD68 positive cells suggests importance of Rep-positive cells (MFs) as possible cancer biomarkers and a possible causal involvement of BMMF on carcinogenesis upon induction of tissue inflammation.

Triple IF staining (anti-Rep/-CD68/-CD163) confirmed that Rep expression is present in a higher proportion in CD68⁺CD163⁺ alveolar macrophages associated with an M2-like phenotype. Alveolar macrophages expressing the M2-like phenotype have been already identified as targets in early lung cancer treatment, as they accumulate near tumor cells and promote cell invasiveness. Cell-based quantification confirmed that Rep⁺ cells are mostly CD68⁺CD163⁺ M2-like macrophages and that their number is significantly increased in

peritumor compared to controls and tumor. We observed a correlation between the number of Rep⁺ cells and M2-like macrophages, that is not present when considering M1-like macrophages

When taking into account the smoking status of patients, all cohorts tested (non-smokers, ex-smokers and smokers) revealed a very heterogenic distribution of Rep expression with strong variation between minimum and maximum cell numbers (except for smokers). Yet, there was a trend pointing towards increased numbers of Rep⁺ cells in non-smokers and low-dose smokers suggesting an anti-correlation between Rep expression and cigarette smoking.

IHC detection of BMMF protein was supported by analysis of cryosections and fresh tissue lysates by immunoblotting, where bands were identified at the expected size (around 37 kDa). IP-enriched samples were analyzed by mass spectroscopy identifying one BMMF1 Rep peptide in a peritumor sample matching to the Rep sequence encoded on the BMMF1 isolates H1MSB.1, C1MI.2, C1MI.3M.1, C1MI.4, C1MI.5.1, C1MI.5.2, which is also detected by the anti-Rep Abs 3 and 10 used for IHC and WB..

To confirm wet-lab BMMF detection, we additionally performed BMMF *in silico* profiling in different sequencing archives of lung cancer patients. BMMF detection in the Pan-Cancer Analysis of Whole Genomes (PCAWG) dataset revealed BMMF DNA reads in 19% of lung adenocarcinoma patients and 43.7% of lung squamous cell carcinoma and BMMF RNA-seq hits in 4.17% of lung adenocarcinoma patient and 10% of lung squamous cell carcinoma patients. The highest number of reads were detected for C1MI.3, C1MI.2, C1MI.3M.1 and H1MSB.2 on PCAWG WG (LUSC) and CIHB.4 in PCAWG RNA-seq (LUSC), which have not been in the focus of histochemical and immunodetection readouts, before, and have therefore been included in the analyses. Thus, upon *in silico* identification, new antibodies have been generated resulting in strong and specific detection levels in IHC, supporting BMMF role as possible, new disease markers. Rep AB 21, one of the ABs that showed high IHC positivity, was raised against CIHB.4 Rep protein (and none of the other known Reps), while Rep AB 22 and 23 were raised against Reps from BMMF1 isolates C1MI.3M.1 and

C1M1.9M.1 from MSBI2 cluster and also showed tissue staining. Detection of the MSBI2 isolate was also observed by *in situ* hybridization where sense and especially anti-sense probes directed against MSBI1 and MSBI2 showed specific speckled like tissue staining in cells with an alveolar macrophage shape localized on the surface of peritumor lung alveoli.

RNA was extracted from frozen and fresh tissues and used for a quantitative real time PCR that led to BMMF detection with BMMF specific primer. Detection of BMMF reads in RT-qPCR, Bulk RNA-seq and scRNA-seq generally support bioactive transcription of BMMF arguing against e.g. systematic bacterial contaminations.

BMMF detection in scRNA-seq resulted in detection of BMMF-positive cells, especially T cells and MFs and positive MFs were associated more closely to a M2-like phenotype.

In general, this PhD project represents the first combined multi-omics study on investigation of the role of BMMF in lung cancer. BMMF proteins were found in tissues of lung cancer patients, specifically (alveolar) macrophages of peritumor tissues which are markers of chronic inflammation. Increased detection of BMMF expression in lung cancer peritumor tissues versus controls supports that BMMF, especially in the cases of non- or low-smoking associated lung cancer, might be involved in indirect carcinogenesis by inducing chronic inflammation leading to oxidative stress and DNA mutation in nearby replicating cells (basal cells in the larger airways and alveolar type II cells in the alveolar regions). Differential detection of BMMF protein by IHC in tumor versus peritumor and case versus control could be the first step towards the development of a method for (early) detection of lung cancer based on BMMF biomarkers. Detection of BMMF during long latency periods prior to symptoms developing would be essential for specific preventive and early therapeutic measures. This finding would gain significance if BMMF could be consistently detected in the bloodstream, potentially enabling the development of a blood test for early lung cancer detection and improving patient outcomes through earlier intervention.

Future *in vitro* experiments may include an uptake assay, to investigate whether macrophages specifically internalize BMMF. This can be done by labeling BMMF with fluorescent dyes or markers, co-culturing with macrophages, and using imaging techniques like confocal microscopy or flow cytometry for readout.

Another approach could involve a co-culture experiment to determine if BMMF preferentially influences macrophage polarization towards an M2-like phenotype, by assessing macrophage polarization markers after BMMF exposure, or a gene expression profiling using RNA-seq or RT-qPCR. Lastly, a co-culture experiment with macrophages, BMMF Rep, and lung cancer cells could explain how BMMF affect interactions between macrophages and tumor cells, and their impact on tumor cell behavior or proliferation.

Apart from *in vitro* experiments, prospective clinical monitoring of pre-cancer individuals would be important to assess BMMF as (predictive) putative biomarker of lung cancer development and progression. Moreover, this can facilitate the early detection of cancer, improving patient outcomes. Future steps will also involve improving and optimizing a DNA extraction protocol to isolate BMMF DNA more effectively, aiming to reduce the gap between *in silico* findings and experimental data. Since IEM on FFPE sections revealed pleomorphic structures ranging from 50-200 nm in size, a starting point would be using methods such as ultracentrifugation, polyethylene glycol and size exclusion chromatography to isolate vesicles. Including additional immune cell markers, particularly for myeloid lineage cells such as macrophages, and performing cytokine profiling will help distinguish between pro-tumorigenic and anti-tumorigenic immune responses.

Establishing a mouse model would be important to test *in vivo* the effect of macrophage inhibition. Clodronate, a bisphosphonate drug that is primarily used for the depletion of macrophages, could be used for that purpose. It is often encapsulated in liposomes to enhance its delivery specifically to macrophages, where it induces apoptosis through phagocytosis.

Finally, testing the BMMF expression through IHC and correlating it with patient survival would be essential in determining the putative role as a

prognostic biomarker. If a significant correlation is found between BMMF expression levels and patient survival, BMMF could serve as a prognostic marker to stratify patients based on their risk of disease progression or recurrence. Moreover, if high BMMF expression is associated with poor prognosis, targeted therapies could be developed specifically against BMMF to improve patient outcomes.

7. Supporting information

7.1. Figures

Figure S1

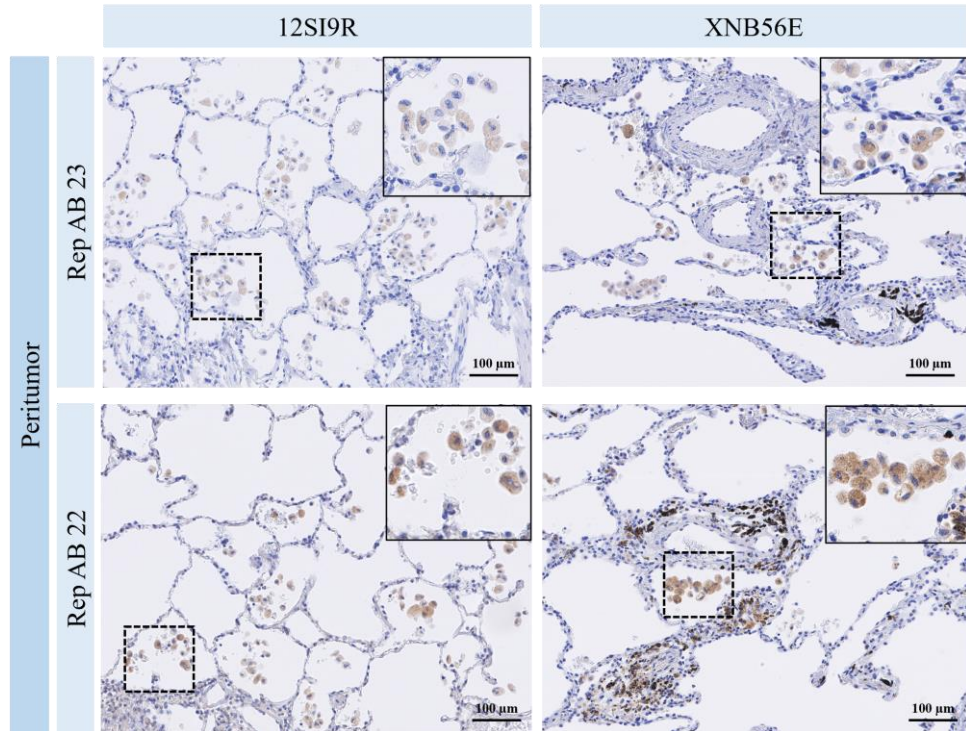


Figure S1. Immunohistochemical detection of BMMF MSBI2 Rep in peritumor tissues. Preliminary staining of lung peritumor tissues with Rep AB 22 and AB 23 from MSBI2 cluster. In total, 6 peritumor tissues were stained with both antibodies.

Figure S2

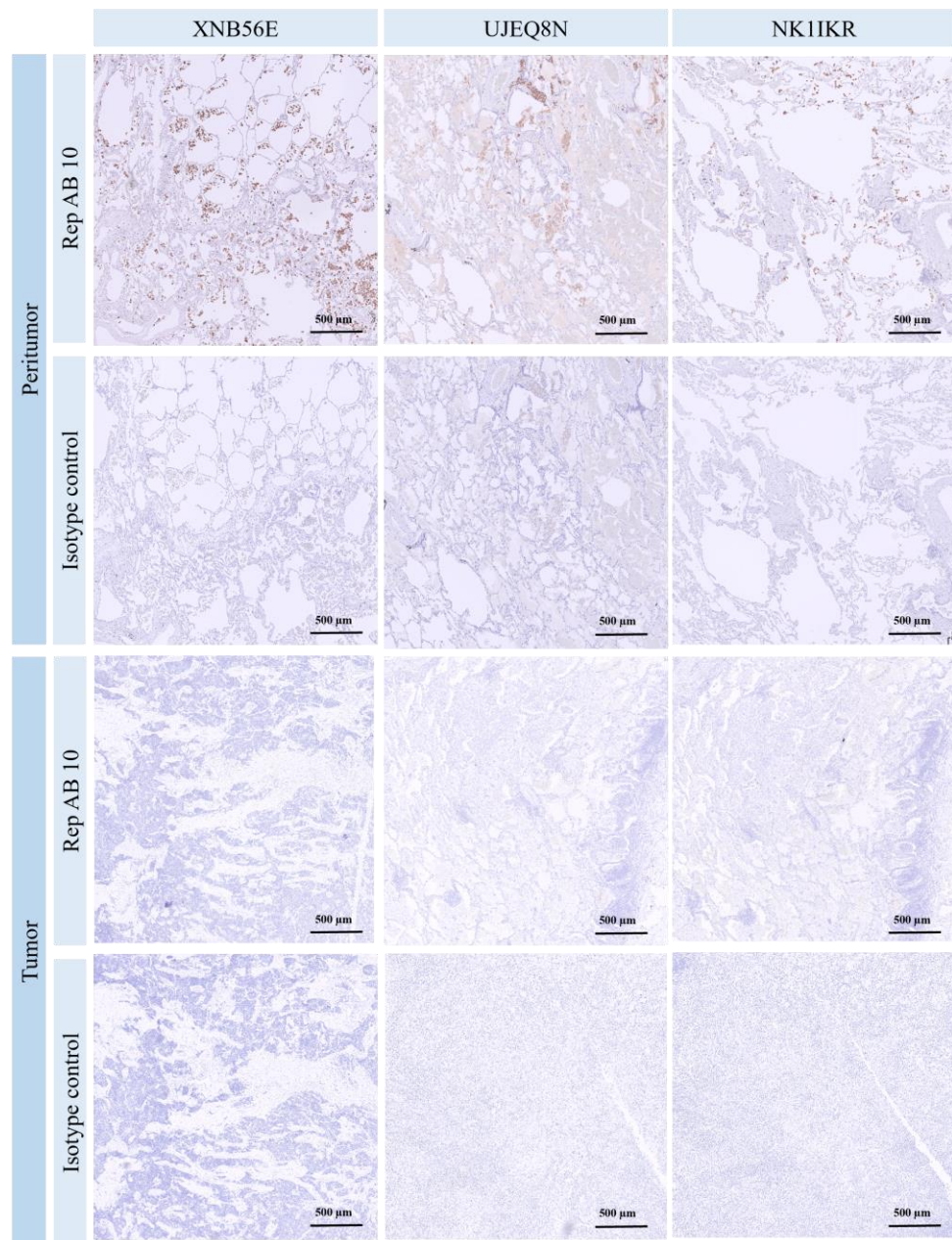


Figure S2. Immunohistochemical detection of BMMF Rep in peritumor and tumor tissues. Consecutive peritumor and tumor tissue sections from lung cancer patients (n=35 + 2 tumor tissues unpaired + 7 peritumor tissues unpaired) were stained with 8 Rep antibodies (AB 10 shown) and isotype control. Rep staining in peritumor tissues is distributed in a homogeneous way in the entire section while the staining is absent in the solid part of the tumor.

Figure S3

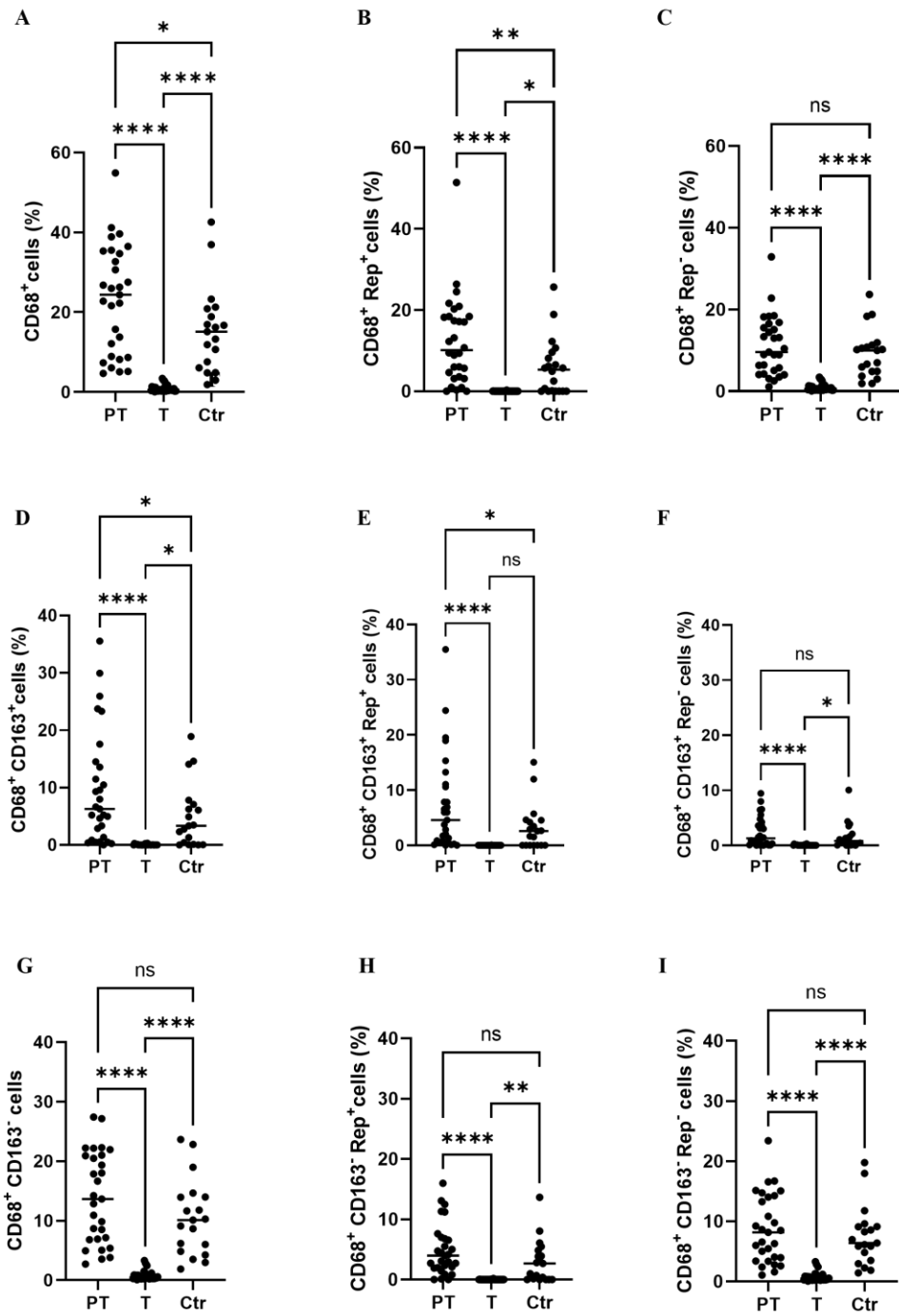


Figure S3. Quantification of Rep⁺, CD68⁺ and CD163⁺ cells based on co-immunofluorescence microscopy. Quantifications of Rep⁺, CD68⁺ and CD163⁺ cells are given in % of all cells. **A.** Quantification of total number of macrophages (CD68⁺ cells) with differentiation in Rep⁺ cells (**B**) and Rep⁻ cells (**C**). **D.** Macrophages have been differentiated into CD68⁺CD163⁺ M2-like cells with further subgroup into Rep⁺ M2-like macrophages (**E**) and Rep⁻ M2-like macrophages (**F**). **G.** Macrophages have been differentiated into CD68⁺CD163⁻ M1-like cells with further subgroup into Rep⁺ M1-like macrophages (**H**) and Rep⁻ M1-like macrophages (**I**). One-way ANOVA test was used for the statistical analysis. Significance: *P<0.05, **P<0.01, ***P<0.001.

Figure S4

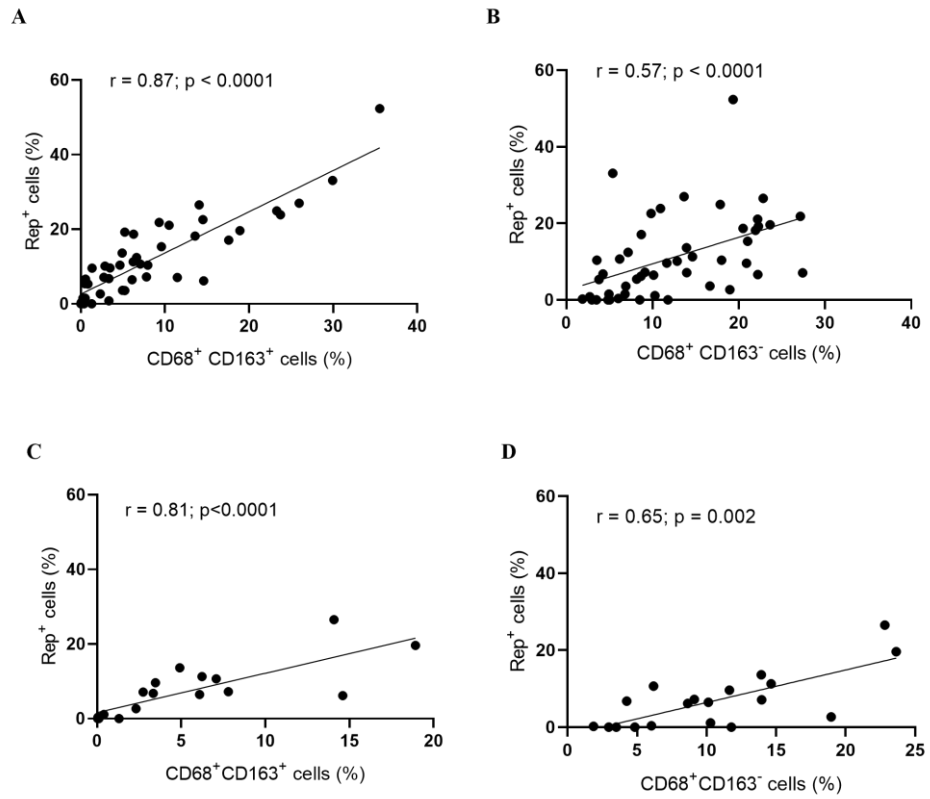


Figure S4. Correlation of Rep⁺ cells with M1 and M2-like macrophages. Simple linear regression of Rep⁺ cells and M2-like macrophages (**A**) and M1-like macrophages (**B**) in lung peritumor and control tissues. Simple linear regression of Rep⁺ cells and M2-like macrophages (**C**) and M1-like macrophages (**D**) in only control tissues (tissues-adjacent to hamartoma). Spearman correlation was used for statistical analysis.

Figure S5

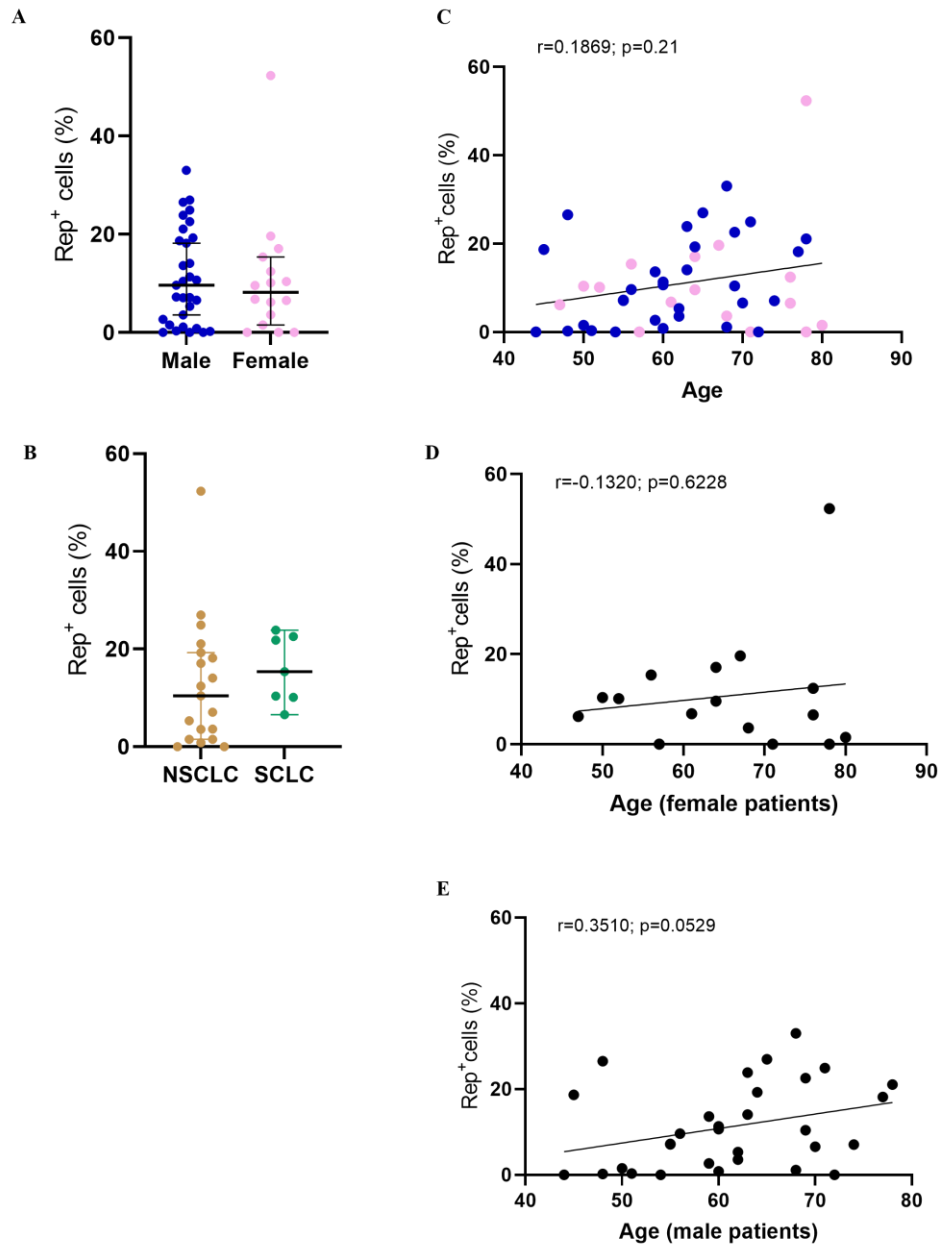


Figure S5. Rep correlation with clinical-epidemiological parameters. **A.** Scatter plot showing Rep expression grouped by sex. **B.** Scatter plot showing Rep expression split by type of lung cancer. Mann-Whitney test was used for the statistical analysis. **C.** Simple linear regression of Rep expression and age of patients. Pink dots represent female patients while blue dots represent male patients. **D.** Simple linear regression of Rep expression and age of female patients. **E.** Simple linear regression of Rep expression and age of male patients. Spearman correlation was used for statistical analysis.

Figure S6

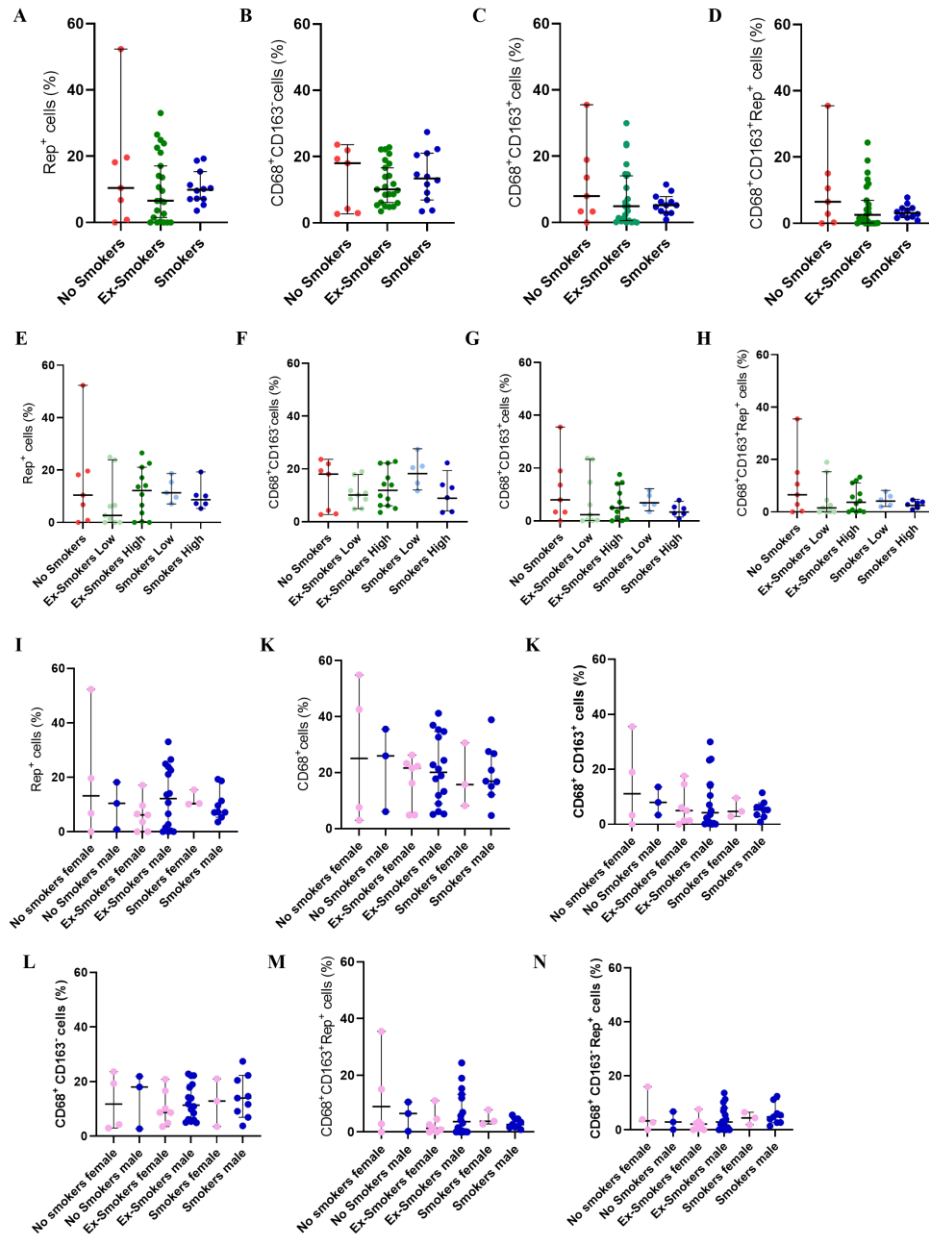


Figure S6. Rep expression in no smokers, ex-smokers and smokers with further differentiation by sex. Scatter plots of Rep⁺ cells (**A**), CD68⁺CD163⁻ M1-like cells (**B**), CD68⁺CD163⁺ M2-like cells (**C**) and Rep⁺CD68⁺CD163⁺ cells (**D**) in no smokers (n=7), ex-smokers (n=23) and smokers (n=12). **E-H:** Ex-smokers and smokers were divided into low and high-dose based on PY (Packages per Year - a clinical measure of tobacco exposure). The division was done to approximately achieve ~50% low and ~50% high (ex) smokers. For ex-smokers, those with 30 PY or more were classified as high ex-smokers. However, many patients had exactly 30 PY, leading to an unbalanced division. No smokers included in the analysis: 7; ex-smokers: 21 (9 low and 12 high-dose); smokers: 11 (5 low and 6 high-dose). **I-N:** Scatter plots of Rep⁺ cells (**I**), CD68⁺ cells (**J**), CD68⁺CD163⁺ M2-like cells (**K**), CD68⁺CD163⁻ M1-like cells (**L**), Rep⁺CD68⁺CD163⁺ cells (**M**) and Rep⁺CD68⁺CD163⁻ cells (**N**) in no smokers, ex-smokers and smokers with further differentiation by sex. One-way ANOVA test was used for the statistical analysis. No significant differences detected.

Figure S7

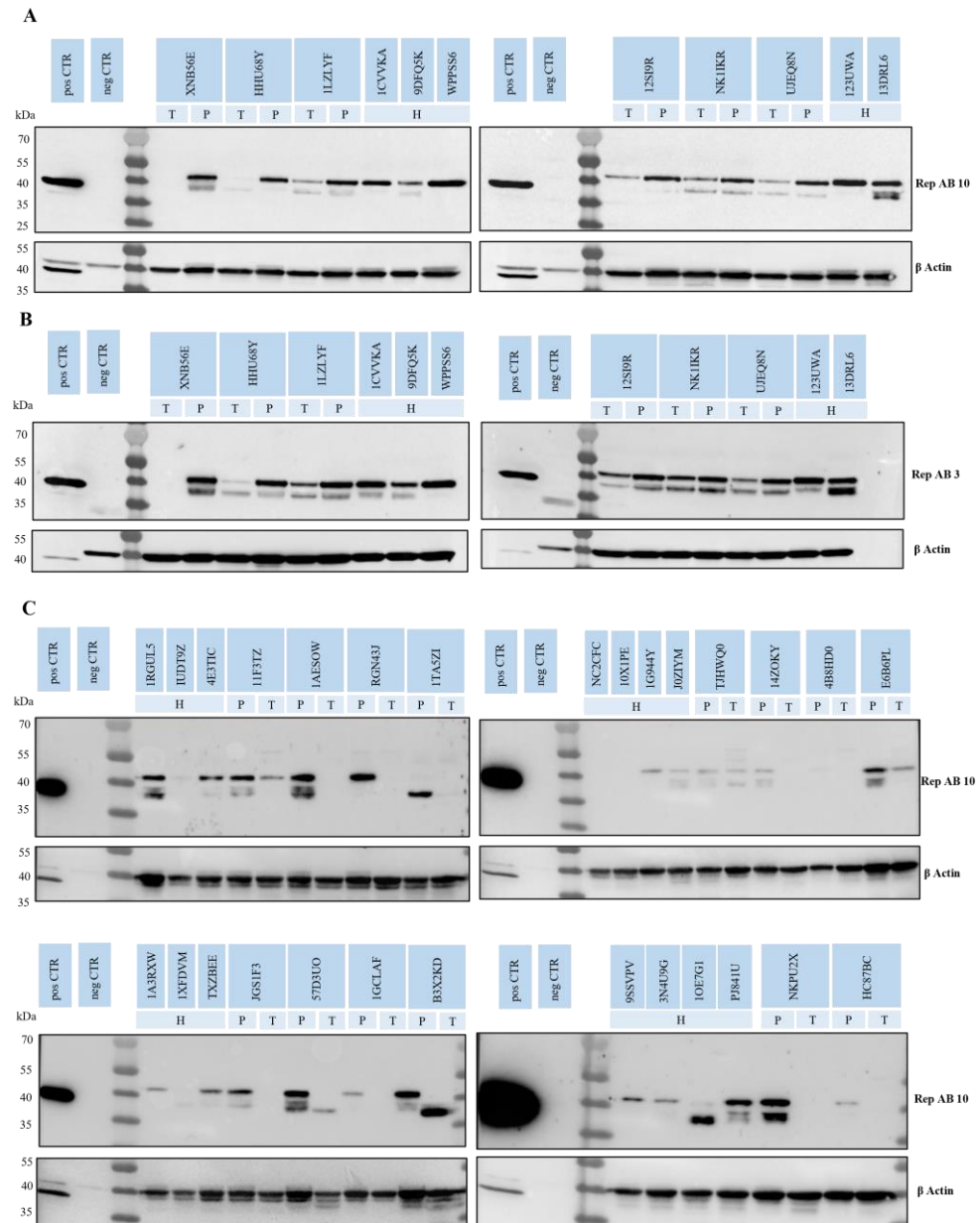
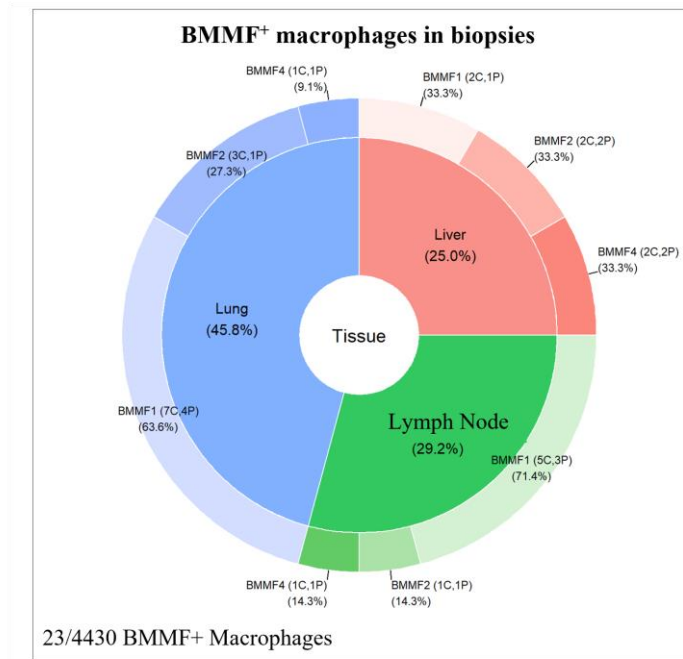


Figure S7. Detection of BMMF protein through immunoblot. A. Immunoblot of frozen lung samples from ThoraxKlinik project 3359 with Rep AB 10 and β actin. **B.** Immunoblot of frozen lung samples from ThoraxKlinik project 3359 with Rep AB 3 and β actin. **C.** Immunoblot of frozen lung samples from ThoraxKlinik project 3694 with Rep AB 10 and β actin. 20 cancer patients and 19 non-cancer individuals tested in total.

Figure S8

A



B

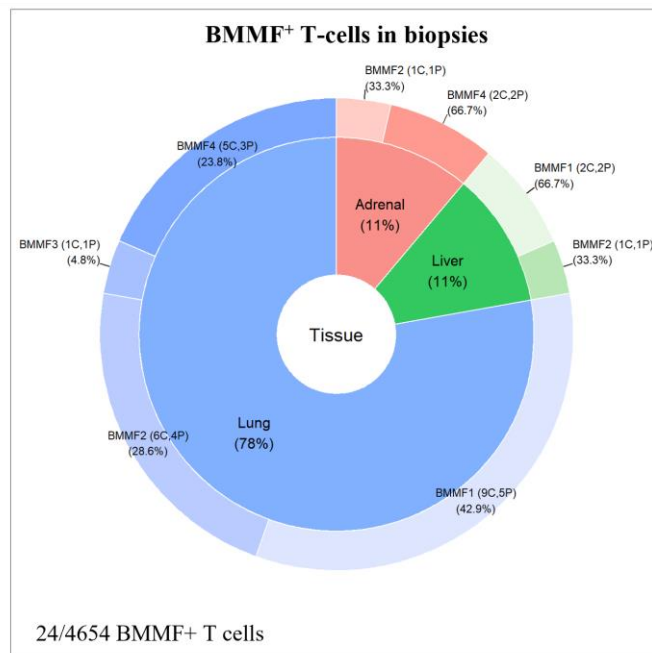


Figure S8. BMMF hits in macrophages and T-cells. Donut plots showing BMMF hits in macrophages (A) and T-cells (B) coming from biopsies of primary lung tumor and metastasis. C = cells, P = patients.

Figure S9

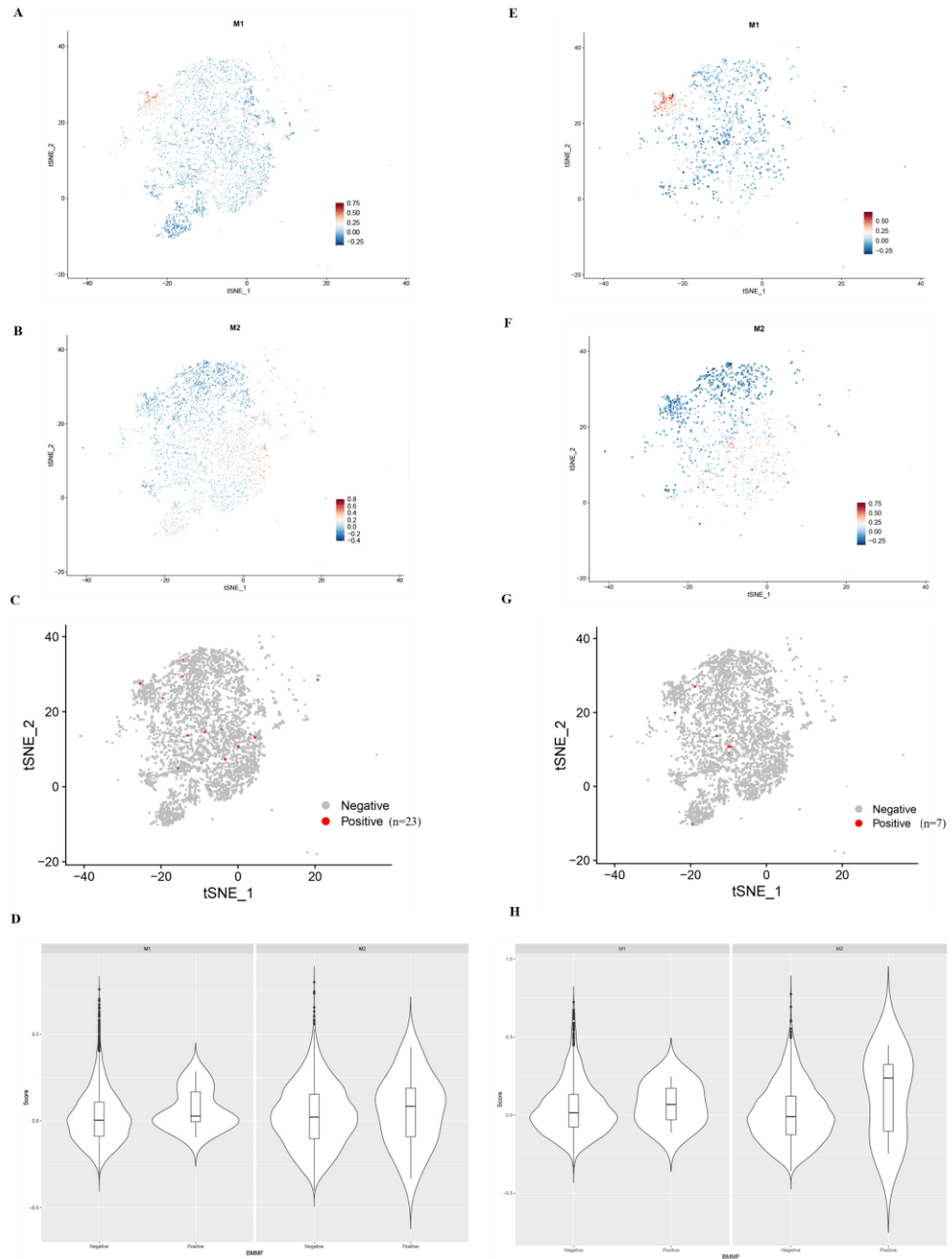


Figure S9. BMMF hits in scRNA-seq data set. A-D. Metastasis and all BMMF included in the analysis. **E-H.** Only primary lung tumor and BMMF1 were included in the analysis. Number of positive cells is reduced but trend pointing towards M2-like macrophages being the lineage with most of the hits for BMMF1 is increased.

7.2. Tables

Table S1

Table S1. Summary table of IHC DAB stainings of clinical lung tissues. Peritumor and tumor tissues from lung cancer patients as well as tissues from non-cancer individuals (tissues adjacent to hamartomas) were stained with eight Rep monoclonal antibodies (Rep AB 1, 2, 3, 7, 10, 11, 15 and 21) and anti-CD68 antibody. N/A indicates that the FFPE sections were missing for that specific patient. X indicates that the staining was not performed, or it was performed but not included in the analysis due to poor quality of the section. Several FFPE sections from project 3694 presented not adequate quality and had to be excluded. ✓ indicates that the staining was performed. “T” indicates tumor, “NG” indicates peritumor and “Hama” indicates hamartoma.

Patient	Tissue	AB 3	AB 10	AB 1	AB 7	AB 11	AB 15	AB 2	AB 21	CD68
11X6KF	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
16P87Y	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1LYLZF	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
1FAVEM	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
1DVRTK	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
ORJWVD	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
1PZHGD	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
14CIFM	T	✓	✓	✓	✓	✓	X	X	X	✓
	NG	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
UJEQ8N	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
OVHCKC	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
1IK3RV	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
HHU68Y	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
NK1IKR	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
6LBF64	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
XNB56E	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
ISR88R	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
ZCZ8WH	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
MOS7SD	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
O6QVYN	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	X	X	✓
1NZV3L	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
2V6GGQ	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
12SI9R	T	✓	✓	✓	✓	✓	✓	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
DTN9K4	T	✓	✓	✓	✓	✓	X	X	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓

Patient	Tissue	AB 3	AB 10	AB 1	AB 7	AB 11	AB 15	AB 2	AB 21	CD 68
11F3TZ	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
1GCLAF	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
RGN43J	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
1TA5ZI	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
JGS1F3	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	X	✓	✓	X	X	✓	✓	✓
57D3UO	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
1W2LHA	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
B3X2KD	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
TJHWQ0	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	X	✓	X	✓	X	✓	✓	✓
14ZOKY	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
4B8HD0	T	X	X	X	X	X	X	✓	✓	✓
	NG	✓	✓	✓	✓	✓	X	✓	✓	✓
E6B6PL	T	✓	✓	✓	✓	X	X	✓	✓	✓
	NG	✓	✓	X	X	✓	X	✓	X	✓
NKPU2X	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	X	✓	✓	X	X	✓	✓	✓
HC87BC	T	✓	✓	✓	✓	✓	X	✓	✓	✓
	NG	✓	X	✓	✓	X	X	✓	✓	✓
1RGUL5	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
IUDT9Z	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
4E3TIC	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
1A3RXW	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
1XFDVM	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
TXZBEE	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
NC2CFC	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
10X1PE	Hama.	✓	X	✓	X	✓	X	✓	✓	✓
1G944Y	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
J0ZIYM	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
9SSVPV	Hama.	✓	✓	✓	✓	✓	X	X	✓	✓
3N4U9G	Hama.	✓	✓	✓	✓	✓	X	X	✓	✓
1OE7GI	Hama.	✓	✓	✓	✓	✓	X	X	✓	✓
PJ841U	Hama.	✓	✓	✓	✓	✓	X	X	✓	✓

Patient	Tissue	AB 3	AB 10	AB 1	AB 7	AB 11	AB 15	AB 2	AB 21	CD 68
9DFQ5K	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
13DRL6	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
WPPSS6	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
123UWA	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
1CVVKA	Hama.	✓	✓	✓	✓	✓	X	✓	✓	✓
15183	PT	✓	✓	✓	✓	✓	✓	✓	✓	✓
15184	PT	✓	✓	✓	✓	✓	✓	✓	✓	✓
15188	PT	✓	✓	✓	✓	✓	✓	✓	✓	✓
15192	PT	✓	✓	✓	✓	✓	✓	✓	✓	✓
15193	PT	✓	✓	✓	✓	✓	✓	✓	✓	✓
15194	PT	✓	✓	✓	✓	✓	✓	✓	✓	✓
15195	PT	✓	✓	✓	✓	✓	✓	✓	✓	✓
15196	PT	✓	✓	✓	✓	✓	✓	✓	✓	✓

NCT – Thorax Klinik Project 2750
 NCT – Thorax Klinik Project 3359
 NCT – Thorax Klinik Project 3694
 NCT – Thorax Klinik Fresh Tissues

Table S2

Table S2. Redundant BMMF hits detected in PCAWG dataset. Summary of BMMF redundant hits detected in LUAD and LUSC tissues by D-ViSioN pipeline in PCAWG WGS (A) and PCAWG-RNA-seq (A) and dataset. Blue: BMMF1 hits. Orange: BMMF2 hits.

A

LUAD – WGS	
Template	Hits
C2MI.15B.16	2
C1HB.6.1	1
C1MI.5.1	1
C1MI.9M.1	1
C2MI.15B.17	1
C2MI.8B.4	1
H1MSB.1	1

LUSC – WGS	
Template	Hits
C1MI.3	11
C1MI.2	5
C1MI.3M.1	5
H1MSB.2	4
C1HB.3	3
C1MI.5.1	3
pHD4	3
Sphinx2.36	3
C1MI.4	2
C1MI.9M.1	2
coD99N.LD.No.1	2
5.12	2
MSSI1.162	2
C1HB.6.1	1
C1MI.9M.2	1
C2MI.7B.17	1
coD08N.LD.Nn.1	1
5.5	1

B

LUAD – RNA-seq	
Template	Hits
C1HB.4	2
C2MI.16B.1	1
C2MI.16B.10	1
C2MI.16B.2	1
C2MI.16B.8	1
MSSI1.162	1
pRGRH0636	1

LUSC – RNA-seq	
Template	Hits
C1HB.4	9
pHD4	2
C1MI.3M.1	1
C1MI.9M.2	1
C1MI.9M.1	1
C1MI.15M.2	1
C1MI.15M.1	1

Table S3

Table S3. Redundant BMMF hits detected in scRNA-seq dataset. Summary of BMMF redundant hits detected in LUAD tissues and metastasis by D-ViSioN pipeline in scRNA-seq data set. Blue: BMMF1 hits. Orange: BMMF2 hits.

Template	Hits		
MSSI1.162	1057	C2MI.15B.2	54
C1HB.4	177	C2MI.9B.14	51
C2MI.8B.6	137	C2MI.5B.7	51
C2MI.8B.5	137	C2MI.5B.6	51
C2MI.8A.3	137	C2MI.7A.5	49
C2MI.8A.2	137	C2MI.7A.4	49
C2MI.8A.1	137	C1MI.9M.2	48
pHD4	121	C1MI.9M.1	48
C1MI.3M.1	116	pRGRH0103	47
C2MI.1A.4	98	C2MI.15B.10	46
C2MI.7A.10	95	C2MI.15B.9	46
C2MI.7A.8	95	C2MI.15B.17	45
C2MI.13B.2	91	C2MI.5A.4	43
C2MI.13B.1	91	C2MI.5A.2	43
C2MI.15B.16	91	HCB19.212	43
C2MI.9B.6	91	coD99N.LD.No.17.10	41
C2MI.9B.7	89	C2MI.15B.12	41
C2MI.9B.4	89	C2MI.15B.11	41
C1HB.3	89	H2HB7	41
C2MI.9B.8	88	C1HB.6.1	41
C2MI.16B.11	80	C1MI.5.2	41
C2MI.16B.9	80	C1MI.4	40
C2MI.15B.18	80	coD08N.LD.Nn.15.8	39
C2MI.15B.14	80	C2MI.7B.17	38
C2MI.7A.9	80	C2MI.7B.15	38
C2MI.15B.4	77	C2MI.7B.11	38
C2MI.15B.3	71	C2MI.5B.9	37
C1MI.3	67	C1HB.6.2	36
C2MI.1A.2	65	C2MI.8B.7	35
C2MI.9B.11	64	C2MI.5B.10	35
C1MI.2	64	coD09N.LD.No.17.12	34
C1MI.15M.2	63	coD09N.LD.No.17.11	34
C1MI.15M.1	63	coD09N.LD.No.17.4	34
H1MSB.2	62	coD09N.LD.No.17.4	34
pRGRH0636	58	coD09N.LD.No.17.2	34
C2MI.8B.4	58	coD99N.LD.No.17.11	34
C2MI.10A.2	57	coD99N.LD.No.17.7	34
C2MI.13B.3	56	coD99N.LD.No.17.3	34
Sphinx2.36	55	coD99N.LD.No.15.10	34
C2MI.9B.9	55	coD99N.LD.No.15.8	34
C2MI.10A.1	54	C1MI.1	34
		H1MSB.1	34

coD08N.LD.Nn.15.12	33
coD08N.LD.Nn.15.9	33
coD08N.LD.Nn.15.4	33
coD09N.LD.No.17.7	33
coD09N.LD.No.17.6	33
coD99N.LD.No.17.4	33
coD99N.LD.No.17.2	33
C2MI.16B.10	33
C2MI.16B.1	33
coD99N.LD.No.15.5	32
coD99N.LD.No.15.2	32
C2MI.7B.14	32
C2MI.7B.12	32
C2MI.7A.7	32
C2MI.7A.3	32
C2MI.7A.2	32
C2MI.7A.1	32
coD09N.LD.No.17.10	31
coD99N.LD.No.15.7	31
C2MI.16B.8	31
coD09N.LD.Nn.17.1	30
C2MI.15B.8	30
coD09N.LD.No.17.3	29
coD09N.LD.No.17.1	29
coD09N.LD.No.17.9	28
coD99N.LD.No.15.3	28
C2MI.15B.1	28
C2MI.5B.8	28
C2MI.5B.5	28
C2MI.5A.3	28
C2MI.5A.1	28
coD08N.LD.Nn.15.5	27
coD09N.LD.No.17.8	27
coD99N.LD.No.17.9	27
coD99N.LD.No.17.5	27
coD99N.LD.No.15.6	27
coD99N.LD.No.15.4	27
coD99N.LD.No.15.12	26
C2MI.15B.13	26
C2MI.15B.7	26
C2MI.15B.6	26
C2MI.15B.5	26
C2MI.5B.13	26

coD09N.LD.No.17.5	25
coD99N.LD.No.17.8	25
C2MI.16B.2	25
C2MI.4B.3	23
C2MI.4A.2	23
C2MI.4A.1	23
C2MI.9B.13	22
C2MI.9B.12	22
C2MI.9A.3	22
C2MI.9A.2	22
C2MI.5B.12	22
C1MI.5.1	21
C2MI.3A.2	20
C2MI.3A.1	20
C2MI.16B.6	20
Sphinx1.76	20
pTS236	19
C2MI.16B.7	19
C2MI.15B.15	19
C2MI.9B.10	19
C2MI.9B.5	19
C2MI.7B.16	19
C2MI.7A.6	19
C2MI.1A.3	17
C2MI.1A.1	17
C2MI.16B.12	17
C2MI.16B.5	17
C2MI.9A.1	17
C2HB2	17
C2MI.16B.4	16
C2MI.16B.3	11
MSSI2.225	10
C1MIs.3M.1	10
4ABAYE	6
C2MI.7B.18	3
C2MI.5B.11	3
H2HB1	3
C1HB.5	3
pA85	2
C2MI.9B.15	1
C2MI.7B.13	1

Table S4

Table S4. List of genes used for M1/M2 cells differentiation. A. List of genes used for differentiating macrophages in M1-like cells. **B.** List of genes used for differentiating macrophages in M2-like cells.

A

CCR7	PLA1A	TRAIL	CSPG2	IL12B	CCL5
IL2RA	OASL	IL6	APOL3	IL15RA	CCL15
IL15RA	CHI3L2	CCL20	IGFBP4	SLC31A2	PSMB9
IL7R	HSD11B1	PBEF1	APOL1	OAS2	PSMA2
CXCL11	AK3	ECGF1	PDGFA	PTX3	HSXIAPAF1
CCL19	SPHK1	BCL2A1	EDN1	INDO	SLC7A5
CXCL10	PFKFB3	FAS	APOL2	SLC21A15	HESX1
CXCL9	PSME2	BIRC3	INHBA	SLC2A6	IRF1
TNF	PFKP	GADD45G	APOL6	ATF3	IRF7

B

GPR86	P2RY14	SLC21A9	CERK	FN1	DCL-1
P2RY5	DCSIGN	SLC4A7	HS3ST2	FGL2	MSR1
TGFB2	CLECSF13	SLC38A6	LTA4H	GAS7	CXCR4
HRH1	MS4A4A	CTSC	CA2	EGR2	DECTIN1
TLR5	MRC1	HEXB	ALOX15	MAF	LIPA
HS3ST1	TGFBI	SEPP1	CHN2	IGF1	ADK
CCL13	CCL18	TPST2	HNMT	CCL23	

8. Literature

- [1] E.-M. de Villiers and H. zur Hausen, "Bovine Meat and Milk Factors (BMMFs): Their Proposed Role in Common Human Cancers and Type 2 Diabetes Mellitus," *Cancers (Basel)*, vol. 13, no. 21, p. 5407, Oct. 2021, doi: 10.3390/cancers13215407.
- [2] E.-M. de Villiers and H. zur Hausen, "Bovine Meat and Milk Factors (BMMFs): Their Proposed Role in Common Human Cancers and Type 2 Diabetes Mellitus," *Cancers (Basel)*, vol. 13, no. 21, p. 5407, Oct. 2021, doi: 10.3390/cancers13215407.
- [3] S. Ishaq and L. Nunn, "Helicobacter pylori and gastric cancer: a state of the art review.," *Gastroenterol Hepatol Bed Bench*, vol. 8, no. Suppl 1, pp. S6–S14, 2015.
- [4] X.-J. Xue, Q. Gao, J.-H. Qiao, J. Zhang, C.-P. Xu, and J. Liu, "Red and processed meat consumption and the risk of lung cancer: a dose-response meta-analysis of 33 published studies.," *Int J Clin Exp Med*, vol. 7, no. 6, pp. 1542–53, 2014.
- [5] P. Gnagnarella, S. Caini, P. Maisonneuve, and S. Gandini, "Carcinogenicity of High Consumption of Meat and Lung Cancer Risk Among Non-Smokers: A Comprehensive Meta-Analysis," *Nutr Cancer*, vol. 70, no. 1, pp. 1–13, Jan. 2018, doi: 10.1080/01635581.2017.1374420.
- [6] D. S. M. Chan *et al.*, "Red and Processed Meat and Colorectal Cancer Incidence: Meta-Analysis of Prospective Studies," *PLoS One*, vol. 6, no. 6, p. e20456, Jun. 2011, doi: 10.1371/journal.pone.0020456.
- [7] R. R. Huxley, A. Ansary-Moghaddam, P. Clifton, S. Czernichow, C. L. Parr, and M. Woodward, "The impact of dietary and lifestyle risk factors on risk of colorectal cancer: A quantitative overview of the epidemiological evidence," *Int J Cancer*, vol. 125, no. 1, pp. 171–180, Jul. 2009, doi: 10.1002/ijc.24343.
- [8] W. Yang *et al.*, "A prospective study of dairy product intake and the risk of hepatocellular carcinoma in U.S. men and women," *Int J Cancer*, vol. 146, no. 5, pp. 1241–1249, Mar. 2020, doi: 10.1002/ijc.32423.
- [9] H. zur Hausen, "Red meat consumption and cancer: Reasons to suspect involvement of bovine infectious factors in colorectal cancer," *Int J Cancer*, vol. 130, no. 11, pp. 2475–2483, Jun. 2012, doi: 10.1002/ijc.27413.
- [10] H. zur Hausen, "What do breast and CRC cancers and MS have in common?," *Nat Rev Clin Oncol*, vol. 12, no. 10, pp. 569–570, Oct. 2015, doi: 10.1038/nrclinonc.2015.154.
- [11] T. Duarte-Salles *et al.*, "Dairy products and risk of hepatocellular carcinoma: The European Prospective Investigation into Cancer and Nutrition," *Int J Cancer*, vol. 135, no. 7, pp. 1662–1672, Oct. 2014, doi: 10.1002/ijc.28812.

- [12] V. Pala *et al.*, "Meat, eggs, dairy products, and risk of breast cancer in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort," *Am J Clin Nutr*, vol. 90, no. 3, pp. 602–612, Sep. 2009, doi: 10.3945/ajcn.2008.27173.
- [13] J. C. van der Pols, C. Bain, D. Gunnell, G. D. Smith, C. Frobisher, and R. M. Martin, "Childhood dairy intake and adult cancer risk: 65-y follow-up of the Boyd Orr cohort," *Am J Clin Nutr*, vol. 86, no. 6, pp. 1722–9, Dec. 2007, doi: 10.1093/ajcn/86.5.1722.
- [14] T. Sugimura, K. Wakabayashi, H. Nakagama, and M. Nagao, "Heterocyclic amines: Mutagens/carcinogens produced during cooking of meat and fish," *Cancer Sci*, vol. 95, no. 4, pp. 290–299, Apr. 2004, doi: 10.1111/j.1349-7006.2004.tb03205.x.
- [15] M. Oostindjer *et al.*, "The role of red and processed meat in colorectal cancer development: a perspective," *Meat Sci*, vol. 97, no. 4, pp. 583–596, Aug. 2014, doi: 10.1016/j.meatsci.2014.02.011.
- [16] Aicr and WCRF, "Diet, nutrition, physical activity and colorectal cancer."
- [17] H. zur Hausen and E.-M. de Villiers, "Dairy cattle serum and milk factors contributing to the risk of colon and breast cancers," *Int J Cancer*, vol. 137, no. 4, pp. 959–967, Aug. 2015, doi: 10.1002/ijc.29466.
- [18] J. Ferlay *et al.*, "Cancer statistics for the year 2020: An overview," *Int J Cancer*, vol. 149, no. 4, pp. 778–789, Aug. 2021, doi: 10.1002/ijc.33588.
- [19] E. M. , L. F. C. M. M. L. P. M. Z. A. S. I. B. F. Ferlay J, "Global Cancer Observatory: Cancer Today," Lyon, France: International Agency for Research on Cancer. Available from: <https://gco.iarc.fr/today>, accessed 26 February 2022. 2020.
- [20] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA Cancer J Clin*, vol. 71, no. 3, pp. 209–249, May 2021, doi: 10.3322/caac.21660.
- [21] H. zur Hausen, T. Bund, and E.-M. de Villiers, "Infectious Agents in Bovine Red Meat and Milk and Their Potential Role in Cancer and Other Chronic Diseases," 2017, pp. 83–116. doi: 10.1007/82_2017_3.
- [22] K. Falida, S. Eilebrecht, K. Gunst, H. zur Hausen, and E.-M. de Villiers, "Isolation of Two Virus-Like Circular DNAs from Commercially Available Milk Samples," *Genome Announc*, vol. 5, no. 17, Apr. 2017, doi: 10.1128/genomeA.00266-17.
- [23] I. Lamberto, K. Gunst, H. Müller, H. zur Hausen, and E.-M. de Villiers, "Mycovirus-Like DNA Virus Sequences from Cattle Serum and Human Brain and Serum Samples from Multiple Sclerosis Patients," *Genome Announc*, vol. 2, no. 4, Aug. 2014, doi: 10.1128/genomeA.00848-14.

- [24] C. Whitley, K. Gunst, H. Müller, M. Funk, H. zur Hausen, and E.-M. de Villiers, "Novel Replication-Competent Circular DNA Molecules from Healthy Cattle Serum and Milk and Multiple Sclerosis-Affected Human Brain Tissue," *Genome Announc*, vol. 2, no. 4, Aug. 2014, doi: 10.1128/genomeA.00849-14.
- [25] M. Funk, K. Gunst, V. Lucansky, H. Müller, H. zur Hausen, and E.-M. de Villiers, "Isolation of Protein-Associated Circular DNA from Healthy Cattle Serum," *Genome Announc*, vol. 2, no. 4, Aug. 2014, doi: 10.1128/genomeA.00846-14.
- [26] E.-M. de Villiers, K. Gunst, D. Chakraborty, C. Ernst, T. Bund, and H. zur Hausen, "A specific class of infectious agents isolated from bovine serum and dairy products and peritumoral colon cancer tissue," *Emerg Microbes Infect*, vol. 8, no. 1, pp. 1205–1218, Jan. 2019, doi: 10.1080/22221751.2019.1651620.
- [27] L. Manuelidis, "Infectious particles, stress, and induced prion amyloids," *Virulence*, vol. 4, no. 5, pp. 373–383, Jul. 2013, doi: 10.4161/viru.24838.
- [28] L. Manuelidis, "Nuclease resistant circular DNAs copurify with infectivity in scrapie and CJD," *J Neurovirol*, vol. 17, no. 2, pp. 131–145, Apr. 2011, doi: 10.1007/s13365-010-0007-0.
- [29] T. Longkumer *et al.*, "Acinetobacter phage genome is similar to Sphinx 2.36, the circular DNA copurified with TSE infected particles," *Sci Rep*, vol. 3, no. 1, p. 2240, Jul. 2013, doi: 10.1038/srep02240.
- [30] T. Bund *et al.*, "Analysis of chronic inflammatory lesions of the colon for BMMF Rep antigen expression and CD68 macrophage interactions," *Proceedings of the National Academy of Sciences*, vol. 118, no. 12, Mar. 2021, doi: 10.1073/pnas.2025830118.
- [31] S. Eilebrecht *et al.*, "Expression and replication of virus-like circular DNA in human cells," *Sci Rep*, vol. 8, no. 1, p. 2851, Dec. 2018, doi: 10.1038/s41598-018-21317-w.
- [32] M.-T. König, R. Fux, E. Link, G. Sutter, E. Märtlbauer, and A. Didier, "Circular Rep-Encoding Single-Stranded DNA Sequences in Milk from Water Buffaloes (*Bubalus arnee f. bubalis*)," *Viruses*, vol. 13, no. 6, p. 1088, Jun. 2021, doi: 10.3390/v13061088.
- [33] M.-T. König, R. Fux, E. Link, G. Sutter, E. Märtlbauer, and A. Didier, "Identification and Characterization of Circular Single-Stranded DNA Genomes in Sheep and Goat Milk," *Viruses*, vol. 13, no. 11, p. 2176, Oct. 2021, doi: 10.3390/v13112176.
- [34] D. K. Chatteraj, "Control of plasmid DNA replication by iterons: no longer paradoxical," *Mol Microbiol*, vol. 37, no. 3, pp. 467–476, Jan. 2002, doi: 10.1046/j.1365-2958.2000.01986.x.

- [35] G. del Solar, R. Giraldo, M. J. Ruiz-Echevarría, M. Espinosa, and R. Díaz-Orejas, "Replication and Control of Circular Bacterial Plasmids," *Microbiology and Molecular Biology Reviews*, vol. 62, no. 2, pp. 434–464, Jun. 1998, doi: 10.1128/MMBR.62.2.434-464.1998.
- [36] L. Dziewit *et al.*, "Plasmid diversity in arctic strains of *Psychrobacter* spp.," *Extremophiles*, vol. 17, no. 3, pp. 433–444, May 2013, doi: 10.1007/s00792-013-0521-0.
- [37] A. Burk-Koerner, "Bovine Milk and Meat Factors: Functional Characterization and Association with Colorectal Cancer," Combined Faculty of Natural Sciences and Mathematics. Ruperto Carola University Heidelberg, Heidelberg, Germany, 2019.
- [38] I. Artuso *et al.*, "Genome diversity of domesticated *Acinetobacter baumannii* ATCC 19606T strains," *Microb Genom*, vol. 8, no. 1, Jan. 2022, doi: 10.1099/mgen.0.000749.
- [39] M. Fondi *et al.*, "Exploring the evolutionary dynamics of plasmids: the *Acinetobacter* pan-plasmidome," *BMC Evol Biol*, vol. 10, no. 1, p. 59, Dec. 2010, doi: 10.1186/1471-2148-10-59.
- [40] S. S. Lean and C. C. Yeo, "Small, Enigmatic Plasmids of the Nosocomial Pathogen, *Acinetobacter baumannii*: Good, Bad, Who Knows?," *Front Microbiol*, vol. 8, Aug. 2017, doi: 10.3389/fmicb.2017.01547.
- [41] P. Wawrzyniak, G. Płucienniczak, and D. Bartosik, "The Different Faces of Rolling-Circle Replication and Its Multifunctional Initiator Proteins," *Front Microbiol*, vol. 8, Nov. 2017, doi: 10.3389/fmicb.2017.02353.
- [42] W. Gilbert and D. Dressler, "DNA Replication: The Rolling Circle Model," *Cold Spring Harb Symp Quant Biol*, vol. 33, no. 0, pp. 473–484, Jan. 1968, doi: 10.1101/SQB.1968.033.01.055.
- [43] S. A. Khan, "Rolling-circle replication of bacterial plasmids," *Microbiology and Molecular Biology Reviews*, vol. 61, no. 4, pp. 442–455, Dec. 1997, doi: 10.1128/mnbr.61.4.442-455.1997.
- [44] U. M. Pinto, K. M. Pappas, and S. C. Winans, "The ABCs of plasmid replication and segregation," *Nat Rev Microbiol*, vol. 10, no. 11, pp. 755–765, Nov. 2012, doi: 10.1038/nrmicro2882.
- [45] J. A. Ruiz-Masó, C. Machón, L. Bordanaba-Ruiseco, M. Espinosa, M. Coll, and G. Del Solar, "Plasmid Rolling-Circle Replication," *Microbiol Spectr*, vol. 3, no. 1, Feb. 2015, doi: 10.1128/microbiolspec.PLAS-0035-2014.
- [46] M. Moscoso, R. Eritja, and M. Espinosa, "Initiation of replication of plasmid pMV158: mechanisms of DNA strand-transfer reactions mediated by the initiator RepB protein," *J Mol Biol*, vol. 268, no. 5, pp. 840–856, May 1997, doi: 10.1006/jmbi.1997.1012.

- [47] E. Tarasova, S. Dhindwal, M. Popp, S. Hussain, and R. Khayat, "Mechanism of DNA Interaction and Translocation by the Replicase of a Circular Rep-Encoding Single-Stranded DNA Virus," *mBio*, vol. 12, no. 4, Aug. 2021, doi: 10.1128/mBio.00763-21.
- [48] T. Kilic *et al.*, "Structural analysis of a replication protein encoded by a plasmid isolated from a multiple sclerosis patient," *Acta Crystallogr D Struct Biol*, vol. 75, no. 5, pp. 498–504, May 2019, doi: 10.1107/S2059798319003991.
- [49] H. Komori, "Crystal structure of a prokaryotic replication initiator protein bound to DNA at 2.6 Å resolution," *EMBO J*, vol. 18, no. 17, pp. 4597–4607, Sep. 1999, doi: 10.1093/emboj/18.17.4597.
- [50] R. Giraldo, C. Fernández-Tornero, P. R. Evans, R. Díaz-Orejas, and A. Romero, "A conformational switch between transcriptional repression and replication initiation in the RepA dimerization domain," *Nat Struct Mol Biol*, vol. 10, no. 7, pp. 565–571, Jul. 2003, doi: 10.1038/nsb937.
- [51] K. S. Gajiwala and S. K. Burley, "Winged helix proteins," *Curr Opin Struct Biol*, vol. 10, no. 1, pp. 110–116, Feb. 2000, doi: 10.1016/S0959-440X(99)00057-3.
- [52] S. Beck, "Production and characterization of novel BMMF2 Rep-specific antibodies and application in cancer histology," Combined Faculty of Natural Sciences and Mathematics. Ruperto Carola University, Heidelberg, Germany, 2022.
- [53] H. zur Hausen, T. Bund, and E. Villiers, "Specific nutritional infections early in life as risk factors for human colon and breast cancers several decades later," *Int J Cancer*, vol. 144, no. 7, pp. 1574–1583, Apr. 2019, doi: 10.1002/ijc.31882.
- [54] M. Canton *et al.*, "Reactive Oxygen Species in Macrophages: Sources and Targets," *Front Immunol*, vol. 12, Sep. 2021, doi: 10.3389/fimmu.2021.734229.
- [55] A. N. Samraj, H. Läubli, N. Varki, and A. Varki, "Involvement of a Non-Human Sialic Acid in Human Cancer," *Front Oncol*, vol. 4, 2014, doi: 10.3389/fonc.2014.00033.
- [56] P. Tangvoranuntakul *et al.*, "Human uptake and incorporation of an immunogenic nonhuman dietary sialic acid," *Proceedings of the National Academy of Sciences*, vol. 100, no. 21, pp. 12045–12050, Oct. 2003, doi: 10.1073/pnas.2131556100.
- [57] F. Alisson-Silva, K. Kawanishi, and A. Varki, "Human risk of diseases associated with red meat intake: Analysis of current theories and proposed role for metabolic incorporation of a non-human sialic acid," *Mol Aspects Med*, vol. 51, pp. 16–30, Oct. 2016, doi: 10.1016/j.mam.2016.07.002.

- [58] A. N. Samraj *et al.*, "A red meat-derived glycan promotes inflammation and cancer progression," *Proceedings of the National Academy of Sciences*, vol. 112, no. 2, pp. 542–547, Jan. 2015, doi: 10.1073/pnas.1417508112.
- [59] V. Padler-Karavani *et al.*, "Diversity in specificity, abundance, and composition of anti-Neu5Gc antibodies in normal humans: Potential implications for disease," *Glycobiology*, vol. 18, no. 10, pp. 818–830, Jun. 2008, doi: 10.1093/glycob/cwn072.
- [60] C. Dhar, A. Sasmal, and A. Varki, "From 'Serum Sickness' to 'Xenosialitis': Past, Present, and Future Significance of the Non-human Sialic Acid Neu5Gc," *Front Immunol*, vol. 10, Apr. 2019, doi: 10.3389/fimmu.2019.00807.
- [61] Z. M. Khan *et al.*, "Crystallographic and Glycan Microarray Analysis of Human Polyomavirus 9 VP1 Identifies N -Glycolyl Neuraminic Acid as a Receptor Candidate," *J Virol*, vol. 88, no. 11, pp. 6100–6111, Jun. 2014, doi: 10.1128/JVI.03455-13.
- [62] C. Yunna, H. Mengru, W. Lei, and C. Weidong, "Macrophage M1/M2 polarization," *Eur J Pharmacol*, vol. 877, p. 173090, Jun. 2020, doi: 10.1016/j.ejphar.2020.173090.
- [63] E. Nikitina *et al.*, "Bovine meat and milk factor protein expression in tumor-free mucosa of colorectal cancer patients coincides with macrophages and might interfere with patient survival," *Mol Oncol*, Feb. 2023, doi: 10.1002/1878-0261.13390.
- [64] E. Nikitina *et al.*, "Structural expression of bovine milk and meat factors in tissues of colorectal, lung and pancreatic cancer patients," *Int J Cancer*, vol. 153, no. 1, pp. 173–182, Jul. 2023, doi: 10.1002/ijc.34374.
- [65] C. de Martel, D. Georges, F. Bray, J. Ferlay, and G. M. Clifford, "Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis," *Lancet Glob Health*, vol. 8, no. 2, pp. e180–e190, Feb. 2020, doi: 10.1016/S2214-109X(19)30488-7.
- [66] H. Lemjabbar-Alaoui, O. U. Hassan, Y.-W. Yang, and P. Buchanan, "Lung cancer: Biology and treatment options," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1856, no. 2, pp. 189–210, Dec. 2015, doi: 10.1016/j.bbcan.2015.08.002.
- [67] S. A. Whiteside, J. E. McGinniss, and R. G. Collman, "The lung microbiome: progress and promise," *Journal of Clinical Investigation*, vol. 131, no. 15, Aug. 2021, doi: 10.1172/JCI150473.
- [68] K. F. Budden *et al.*, "Emerging pathogenic links between microbiota and the gut–lung axis," *Nat Rev Microbiol*, vol. 15, no. 1, pp. 55–63, Jan. 2017, doi: 10.1038/nrmicro.2016.142.
- [69] K. Xiong, W. Sun, Y. He, and L. Fan, "Advances in molecular mechanisms of interaction between Mycobacterium tuberculosis and lung cancer: a

narrative review.," *Transl Lung Cancer Res*, vol. 10, no. 10, pp. 4012–4026, Oct. 2021, doi: 10.21037/tlcr-21-465.

- [70] D. Focosi, G. Antonelli, M. Pistello, and F. Maggi, "Torquetenovirus: the human virome from bench to bedside," *Clinical Microbiology and Infection*, vol. 22, no. 7, pp. 589–593, Jul. 2016, doi: 10.1016/j.cmi.2016.04.007.
- [71] G. Freer, F. Maggi, M. Pifferi, M. E. Di Cicco, D. G. Peroni, and M. Pistello, "The Virome and Its Major Component, Anellovirus, a Convoluting System Molding Human Immune Defenses and Possibly Affecting the Development of Asthma and Respiratory Diseases in Childhood," *Front Microbiol*, vol. 9, Apr. 2018, doi: 10.3389/fmicb.2018.00686.
- [72] A. A. Abbas *et al.*, "Redondoviridae, a Family of Small, Circular DNA Viruses of the Human Oro-Respiratory Tract Associated with Periodontitis and Critical Illness," *Cell Host Microbe*, vol. 25, no. 5, pp. 719–729.e4, May 2019, doi: 10.1016/j.chom.2019.04.001.
- [73] "GLOBOCAN 2020 (IARC)."
- [74] R. Durusoy *et al.*, "Lung cancer risk and occupational exposure to meat and live animals," *Int J Cancer*, vol. 118, no. 10, pp. 2543–2547, May 2006, doi: 10.1002/ijc.21690.
- [75] H. Besson, R. Banks, and P. Boffetta, "Cancer Mortality Among Butchers: A 24-State Death Certificate Study," *J Occup Environ Med*, vol. 48, no. 3, pp. 289–293, Mar. 2006, doi: 10.1097/01.jom.0000184867.83288.d0.
- [76] J. Ji, J. Sundquist, and K. Sundquist, "Lactose intolerance and risk of lung, breast and ovarian cancers: aetiological clues from a population-based study in Sweden," *Br J Cancer*, vol. 112, no. 1, pp. 149–152, Jan. 2015, doi: 10.1038/bjc.2014.544.
- [77] R. E. Harris, "Ibuprofen in the Prevention and Therapy of Cancer," in *Ibuprofen*, Chichester, UK: John Wiley & Sons, Ltd, 2015, pp. 518–546. doi: 10.1002/9781118743614.ch13.
- [78] A. Akhmedkhanov, P. Toniolo, A. Zeleniuch-Jacquotte, K. L. Koenig, and R. E. Shore, "Aspirin and lung cancer in women," *Br J Cancer*, vol. 87, no. 1, pp. 49–53, Jul. 2002, doi: 10.1038/sj.bjc.6600370.
- [79] T. Lazarov, S. Juarez-Carreño, N. Cox, and F. Geissmann, "Physiology and diseases of tissue-resident macrophages," *Nature*, vol. 618, no. 7966, pp. 698–707, Jun. 2023, doi: 10.1038/s41586-023-06002-x.
- [80] F. Hou, K. Xiao, L. Tang, and L. Xie, "Diversity of Macrophages in Lung Homeostasis and Diseases," *Front Immunol*, vol. 12, Sep. 2021, doi: 10.3389/fimmu.2021.753940.
- [81] H. Aegerter, B. N. Lambrecht, and C. V. Jakubzick, "Biology of lung macrophages in health and disease.," *Immunity*, vol. 55, no. 9, pp. 1564–1580, Sep. 2022, doi: 10.1016/j.immuni.2022.08.010.

- [82] M. Casanova-Acebes *et al.*, "Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells," *Nature*, vol. 595, no. 7868, pp. 578–584, Jul. 2021, doi: 10.1038/s41586-021-03651-8.
- [83] A. J. Boutilier and S. F. Elswa, "Macrophage Polarization States in the Tumor Microenvironment," *Int J Mol Sci*, vol. 22, no. 13, p. 6995, Jun. 2021, doi: 10.3390/ijms22136995.
- [84] E. Evren, E. Ringqvist, and T. Willinger, "Origin and ontogeny of lung macrophages: from mice to humans.," *Immunology*, vol. 160, no. 2, pp. 126–138, Jun. 2020, doi: 10.1111/imm.13154.
- [85] C. Lee *et al.*, "Targeting of M2-like tumor-associated macrophages with a melittin-based pro-apoptotic peptide," *J Immunother Cancer*, vol. 7, no. 1, p. 147, Dec. 2019, doi: 10.1186/s40425-019-0610-4.
- [86] K. Chaitanya Thandra, A. Barsouk, K. Saginala, J. Sukumar Aluru, and A. Barsouk, "Epidemiology of lung cancer," *Współczesna Onkologia*, vol. 25, no. 1, pp. 45–52, 2021, doi: 10.5114/wo.2021.103829.
- [87] N. A. K. M. M. D. B. A. Y. M. R. J. T. Z. M. A. L. D. C. H. F. E. C. K. (eds). Howlader N, "SEER Cancer Statistics Review, 1975-2017, National Cancer Institute," Bethesda, MD, https://seer.cancer.gov/csr/1975_2017/, based on November 2019 SEER data submission, posted to the SEER web site, April 2020.
- [88] J. A. Barta, C. A. Powell, and J. P. Wisnivesky, "Global Epidemiology of Lung Cancer," *Ann Glob Health*, vol. 85, no. 1, Jan. 2019, doi: 10.5334/aogh.2419.
- [89] M. B. Schabath and M. L. Cote, "Cancer Progress and Priorities: Lung Cancer.," *Cancer Epidemiol Biomarkers Prev*, vol. 28, no. 10, pp. 1563–1579, Oct. 2019, doi: 10.1158/1055-9965.EPI-19-0221.
- [90] A. G. Pallis and K. N. Syrigos, "Lung cancer in never smokers: Disease characteristics and risk factors," *Crit Rev Oncol Hematol*, vol. 88, no. 3, pp. 494–503, Dec. 2013, doi: 10.1016/j.critrevonc.2013.06.011.
- [91] M. Yu, "The Role of Interleukin-6 in Pulmonary Inflammation and Injury Induced by Exposure to Environmental Air Pollutants," *Toxicological Sciences*, vol. 68, no. 2, pp. 488–497, Aug. 2002, doi: 10.1093/toxsci/68.2.488.
- [92] W. Dröge, "Free Radicals in the Physiological Control of Cell Function," *Physiol Rev*, vol. 82, no. 1, pp. 47–95, Jan. 2002, doi: 10.1152/physrev.00018.2001.
- [93] A. Valavanidis, T. Vlachogianni, K. Fiotakis, and S. Loidas, "Pulmonary Oxidative Stress, Inflammation and Cancer: Respirable Particulate Matter, Fibrous Dusts and Ozone as Major Causes of Lung Carcinogenesis through Reactive Oxygen Species Mechanisms," *Int J Environ Res Public Health*, vol. 10, no. 9, pp. 3886–3907, Aug. 2013, doi: 10.3390/ijerph10093886.

- [94] N. Azad, Y. Rojanasakul, and V. Vallyathan, "Inflammation and Lung Cancer: Roles of Reactive Oxygen/Nitrogen Species," *Journal of Toxicology and Environmental Health, Part B*, vol. 11, no. 1, pp. 1–15, Jan. 2008, doi: 10.1080/10937400701436460.
- [95] H. Bartsch and J. Nair, "Chronic inflammation and oxidative stress in the genesis and perpetuation of cancer: role of lipid peroxidation, DNA damage, and repair," *Langenbecks Arch Surg*, vol. 391, no. 5, pp. 499–510, Sep. 2006, doi: 10.1007/s00423-006-0073-1.
- [96] J. D. Lang, P. J. McArdle, P. J. O'Reilly, and S. Matalon, "Oxidant-Antioxidant Balance in Acute Lung Injury*," *Chest*, vol. 122, no. 6, pp. 314S-320S, Dec. 2002, doi: 10.1378/chest.122.6_suppl.314S.
- [97] J. C. Marshall, "Inflammation, coagulopathy, and the pathogenesis of multiple organ dysfunction syndrome," *Crit Care Med*, vol. 29, pp. S99–S106, Jul. 2001, doi: 10.1097/00003246-200107001-00032.
- [98] J. A. Cook, D. Gius, D. A. Wink, M. C. Krishna, A. Russo, and J. B. Mitchell, "Oxidative stress, redox, and the tumor microenvironment," *Semin Radiat Oncol*, vol. 14, no. 3, pp. 259–266, Jul. 2004, doi: 10.1016/j.semradonc.2004.04.001.
- [99] R.-L. Xie *et al.*, "Lung nodule pre-diagnosis and insertion path planning for chest CT images," *BMC Med Imaging*, vol. 23, no. 1, p. 22, Feb. 2023, doi: 10.1186/s12880-023-00973-z.
- [100] "lung.org - American Lung Association."
- [101] C. M. Rudin, E. Brambilla, C. Faivre-Finn, and J. Sage, "Small-cell lung cancer," *Nat Rev Dis Primers*, vol. 7, no. 1, p. 3, Jan. 2021, doi: 10.1038/s41572-020-00235-0.
- [102] S. B. Clark and S. Alsubait, *Non–Small Cell Lung Cancer*. 2024.
- [103] "webmd.com - WebMD."
- [104] D. Amicizia *et al.*, "Systematic Review of Lung Cancer Screening: Advancements and Strategies for Implementation," *Healthcare*, vol. 11, no. 14, p. 2085, Jul. 2023, doi: 10.3390/healthcare11142085.
- [105] E. J. van Beek, "Lung cancer screening: Computed tomography or chest radiographs?," *World J Radiol*, vol. 7, no. 8, p. 189, 2015, doi: 10.4329/wjr.v7.i8.189.
- [106] A. Ammanagi, V. Dombale, A. Miskin, G. Dandagi, and S. Sangolli, "Sputum cytology in suspected cases of carcinoma of lung (Sputum cytology a poor man's bronchoscopy!)," *Lung India*, vol. 29, no. 1, p. 19, 2012, doi: 10.4103/0970-2113.92356.
- [107] "cancer.org - American Cancer Society."

- [108] S. K. Vinod and E. Hau, "Radiotherapy treatment for lung cancer: Current status and future directions," *Respirology*, vol. 25, no. S2, pp. 61–71, Nov. 2020, doi: 10.1111/resp.13870.
- [109] J. Skowronek, "Brachytherapy in the treatment of lung cancer – a valuable solution," *J Contemp Brachytherapy*, vol. 4, pp. 297–311, 2015, doi: 10.5114/jcb.2015.54038.
- [110] A. Lahiri *et al.*, "Lung cancer immunotherapy: progress, pitfalls, and promises," *Mol Cancer*, vol. 22, no. 1, p. 40, Feb. 2023, doi: 10.1186/s12943-023-01740-y.
- [111] C. de Martel, D. Georges, F. Bray, J. Ferlay, and G. M. Clifford, "Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis," *Lancet Glob Health*, vol. 8, no. 2, pp. e180–e190, Feb. 2020, doi: 10.1016/S2214-109X(19)30488-7.
- [112] "cancercenter.com."
- [113] O. R. Brook, L.-A. Best, and L. Guralnik, "Inflamed pulmonary hamartoma," *European Journal of Radiology Extra*, vol. 51, no. 3, pp. 107–108, Sep. 2004, doi: 10.1016/j.ejrex.2004.06.005.
- [114] L. Häfele, "Detection of Bovine Meat and Milk Factors in colorectal and breast cancer patients using D-VISION.," Ruprecht-Karls-Universität Heidelberg, 2020.
- [115] A. Maynard *et al.*, "Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing," *Cell*, vol. 182, no. 5, pp. 1232–1251.e22, Sep. 2020, doi: 10.1016/j.cell.2020.07.017.
- [116] C. L. Holness and D. L. Simmons, "Molecular cloning of CD68, a human macrophage marker related to lysosomal glycoproteins," *Blood*, vol. 81, no. 6, pp. 1607–13, Mar. 1993.
- [117] M. Fathi, A. Johansson, M. Lundborg, L. Orre, C. M. Sköld, and P. Camner, "Functional and Morphological Differences between Human Alveolar and Interstitial Macrophages," *Exp Mol Pathol*, vol. 70, no. 2, pp. 77–82, Apr. 2001, doi: 10.1006/exmp.2000.2344.
- [118] J. C. van der Pols, C. Bain, D. Gunnell, G. D. Smith, C. Frobisher, and R. M. Martin, "Childhood dairy intake and adult cancer risk: 65-y follow-up of the Boyd Orr cohort.," *Am J Clin Nutr*, vol. 86, no. 6, pp. 1722–9, Dec. 2007, doi: 10.1093/ajcn/86.5.1722.
- [119] B. A. C. J. Mandel. Steven E. Weinberger, *Principles of pulmonary medicine*. 2019.
- [120] H. Aegerter, B. N. Lambrecht, and C. V Jakubzick, "Biology of lung macrophages in health and disease.," *Immunity*, vol. 55, no. 9, pp. 1564–1580, Sep. 2022, doi: 10.1016/j.immuni.2022.08.010.

- [121] A. V. Misharin, L. Morales-Nebreda, G. M. Mutlu, G. R. S. Budinger, and H. Perlman, "Flow Cytometric Analysis of Macrophages and Dendritic Cell Subsets in the Mouse Lung," *Am J Respir Cell Mol Biol*, vol. 49, no. 4, pp. 503–510, Oct. 2013, doi: 10.1165/rcmb.2013-0086MA.
- [122] T. Wang *et al.*, "Influenza-trained mucosal-resident alveolar macrophages confer long-term antitumor immunity in the lungs," *Nat Immunol*, vol. 24, no. 3, pp. 423–438, Mar. 2023, doi: 10.1038/s41590-023-01428-x.
- [123] V. Erdogan, "Does Pulmonary Hamartoma Increase the Risk of Lung Cancer? Outcomes of 38 Pulmonary Hamartoma Cases," *Sisli Etfal Hastanesi Tip Bulteni / The Medical Bulletin of Sisli Hospital*, 2020, doi: 10.14744/SEMB.2020.06936.
- [124] S. M. Zeisberger, B. Odermatt, C. Marty, A. H. M. Zehnder-Fjällman, K. Ballmer-Hofer, and R. A. Schwendener, "Clodronate-liposome-mediated depletion of tumour-associated macrophages: a new and highly effective antiangiogenic therapy approach," *Br J Cancer*, vol. 95, no. 3, pp. 272–281, Aug. 2006, doi: 10.1038/sj.bjc.6603240.
- [125] J. Jackute *et al.*, "Distribution of M1 and M2 macrophages in tumor islets and stroma in relation to prognosis of non-small cell lung cancer," *BMC Immunol*, vol. 19, no. 1, p. 3, Dec. 2018, doi: 10.1186/s12865-018-0241-4.
- [126] S. S. Sedighzadeh, A. P. Khoshbin, S. Razi, M. Keshavarz-Fathi, and N. Rezaei, "A narrative review of tumor-associated macrophages in lung cancer: regulation of macrophage polarization and therapeutic implications," *Transl Lung Cancer Res*, vol. 10, no. 4, pp. 1889–1916, Apr. 2021, doi: 10.21037/tlcr-20-1241.
- [127] L. May, K. Shows, P. Nana-Sinkam, H. Li, and J. W. Landry, "Sex Differences in Lung Cancer," *Cancers (Basel)*, vol. 15, no. 12, p. 3111, Jun. 2023, doi: 10.3390/cancers15123111.
- [128] C. E. Cressler, W. A. Nelson, T. Day, and E. McCauley, "Disentangling the interaction among host resources, the immune system and pathogens," *Ecol Lett*, vol. 17, no. 3, pp. 284–293, Mar. 2014, doi: 10.1111/ele.12229.
- [129] "National Cancer Institute NIH."
- [130] "Yale Medicine," yalemedicine.org.
- [131] F. Zhou and C. Zhou, "Lung cancer in never smokers-the East Asian experience," 2018, *AME Publishing Company*. doi: 10.21037/tlcr.2018.05.14.
- [132] G. T. Moffat, T. Wang, and A. G. Robinson, "Small Cell Lung Cancer in Light/Never Smokers – A Role for Molecular Testing?," *Journal of the National Comprehensive Cancer Network*, vol. 21, no. 4, pp. 336–339, Apr. 2023, doi: 10.6004/jnccn.2022.7089.

- [133] S. T. Lugg, A. Scott, D. Parekh, B. Naidu, and D. R. Thickett, "Cigarette smoke exposure and alveolar macrophages: mechanisms for lung disease," *Thorax*, vol. 77, no. 1, pp. 94–101, Jan. 2022, doi: 10.1136/thoraxjnl-2020-216296.
- [134] S. T. Lugg, A. Scott, D. Parekh, B. Naidu, and D. R. Thickett, "Cigarette smoke exposure and alveolar macrophages: mechanisms for lung disease," *Thorax*, vol. 77, no. 1, pp. 94–101, Jan. 2022, doi: 10.1136/thoraxjnl-2020-216296.
- [135] H. W. Grievink *et al.*, "Cardiovascular risk factors: The effects of ageing and smoking on the immune system, an observational clinical study," *Front Immunol*, vol. 13, Sep. 2022, doi: 10.3389/fimmu.2022.968815.
- [136] L. I. Kunz *et al.*, "Smoking status and anti-inflammatory macrophages in bronchoalveolar lavage and induced sputum in COPD," *Respir Res*, vol. 12, no. 1, p. 34, Dec. 2011, doi: 10.1186/1465-9921-12-34.
- [137] V. Saint-André *et al.*, "Smoking changes adaptive immunity with persistent effects," *Nature*, vol. 626, no. 8000, pp. 827–835, Feb. 2024, doi: 10.1038/s41586-023-06968-8.
- [138] "cancer.org - American Cancer Society."
- [139] C.-M. Su, Y. Du, R. R. R. Rowland, Q. Wang, and D. Yoo, "Reprogramming viral immune evasion for a rational design of next-generation vaccines for RNA viruses," *Front Immunol*, vol. 14, Apr. 2023, doi: 10.3389/fimmu.2023.1172000.
- [140] I. Horváth, J. Hunt, and P. J. Barnes, "Exhaled breath condensate: methodological recommendations and unresolved questions," *European Respiratory Journal*, vol. 26, no. 3, pp. 523–548, Sep. 2005, doi: 10.1183/09031936.05.00029705.
- [141] "medlineplus.gov."
- [142] A. G. Wheaton *et al.*, "Chronic Obstructive Pulmonary Disease and Smoking Status — United States, 2017," *MMWR Morb Mortal Wkly Rep*, vol. 68, no. 24, pp. 533–538, Jun. 2019, doi: 10.15585/mmwr.mm6824a1.

Acknowledgments

Thank you to my PI, Dr. Timo Bund, for your mentorship, patience, and for always pushing me to reach my full potential. I am grateful to have had the opportunity to work in your lab, where I was able to grow both as a scientist and as a person.

Thank you to my TAC members and to my defense committee, Prof. Bartenschlager, Prof. Heikenwälder, Prof. Fricker, Prof. Papavasiliou, and Prof. Lohmann, for your guidance, insightful feedback and support. I greatly appreciate the time and effort you have taken to evaluate my work.

Thank you to my fellow PhD students, Gunjan, Paul and Simon, for always listening to all the “it did not work”, for teaming up with me. The lab was a better place with you.

Thank you to Sumen, Kristina, Kate, Imke, Karin, Claudia, Caroline and Lisa for your invaluable assistance and support. I have learned so much from all of you.

Thank you to the late Professor zur Hausen, for your invaluable advice and the privilege of working under your guidance.

Thank you to Joao, Francesca, Daniel, Geraldine and all the friends from the E041 department. You made me feel at home from the very beginning.

Thank you, Daria, for being the best friend I could have hoped to find in this grey country.

Thank you, Dario, Vale, Chiara, Franci and Laura for all our ~~therapy sessions~~ dinners together. For always listening to my problems or complaints. For the wine – lots of wine.

Thank you, Ryan, Olga, Beata, Manu for all the trips and get-together we shared, for adding some enjoyable and much-needed breaks to this journey.

Thank you to Lucrezia, Miryam, Chiara, and Jacopo for proving that friendships do not care about kilometers.

Thank you to my parents and Lorenzo, for the support and encouragement you have shown for all my decisions, even when they led me far from home.

Thank you, Lukas, for sharing your life with me and for being always at my side, for accepting my frustrations and for helping with my RStudio mental breakdowns.

Thank you, Micky, for all the purrs on my lap while I was writing this thesis.
Thank you to the nameless woman at the Erasmus office of Tuscia University
for pushing me to accept the grant.

Finally, thank you to me, for taking the risks and the doubts of starting a new
life in Germany. For not giving up all the time I thought “I am not enough”,
for having the courage to reply “Yes” to the woman at the Erasmus office.