A Precision Medicine Perspective on Disease Risk Assessment Driven by Metabolic Phenotypes

Von der Medizinischen Fakultät
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades
eines Doktors der Medizin
genehmigte Dissertation
vorgelegt von

Feng Cao

aus

Anqing (China)

Berichter: Priv.-Doz. Dr. med. Franziska Meister
Jun.-Prof. Dr. med. Carolin Schneider

Tag der mündlichen Prüfung : 25. 08. 2026

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Abstract (English)

Metabolic dysregulation is increasingly recognized as a fundamental driver of disease development and progression. MASLD may represent its hepatic manifestation, whereas cancer may reflect a major long-term consequence across multiple organs. Rather than being isolated entities, MASLD and cancer are likely to share common pathobiological foundations. However, conventional clinical phenotypes alone cannot fully capture the complex intrinsic features linking these conditions. With the advancement of precision medicine, disease characterization is increasingly shifting toward endotypes that emphasize mechanistic heterogeneity.

Accordingly, this study integrated lipidomics, metabolomics, proteomics, and regional body composition to systematically investigate the biological basis linking MASLD and cancer from both circulating molecular and host structural perspectives, Using large-scale population-based cohort and GWAS data from the UK Biobank and FinnGen. Logistic regression, Cox regression, machine learning, multi-omics factor analysis (MOFA), and enrichment analysis were applied according to the specific research questions and outcomes.

The results showed that, at the circulating molecular level, ketone bodies and large HDL-related lipidomic components distinguished MASLD from MetALD with relatively high accuracy, indicating diagnostic potential. At the host structural level, regional body composition was significantly associated with mortality risk in cholangiocarcinoma and its subtypes, with distinct site-specific patterns. Further analyses suggested that these associations may extend to pan-cancer risk. In addition, biologically interpretable body composition patterns were identified for specific cancers, potentially involving pathways related to amino acid transport, oxidative detoxification, and neurotransmitter-related processes.

In summary, this study suggests that metabolic dysregulation may link MASLD and cancer through a continuum of measurable metabolic endotypes across disease stages and organ systems. As two complementary readouts, circulating multi-omic features and regional body composition deepen the understanding of MASLD heterogeneity and cancer risk variation, while offering new perspectives for precise disease characterization, risk stratification, and mechanistic investigation.

Abstrakt (Deutsch)

Die metabolische Dysregulation gilt zunehmend als zentraler Treiber der Krankheitsentstehung und -progression. MASLD kann als hepatische Manifestation, Krebs als wesentliche Langzeitfolge ihrer kumulativen Auswirkungen verstanden werden. Beide Erkrankungen beruhen wahrscheinlich auf gemeinsamen pathobiologischen Grundlagen. Da konventionelle klinische Phänotypen die komplexen intrinsischen Merkmale dieser Verbindung nur unzureichend erfassen, gewinnt in der Präzisionsmedizin die Charakterisierung von Endotypen zur Abbildung mechanistischer Heterogenität zunehmend an Bedeutung.

Vor diesem Hintergrund integrierte die vorliegende Arbeit Lipidomik, Metabolomik, Proteomik und regionale Körperzusammensetzung, um die biologische Grundlage der Verbindung zwischen MASLD und Krebs systematisch aus molekularer und struktureller Perspektive zu untersuchen. Grundlage bildeten populationsbasierte Kohorten- und GWAS-Daten aus der UK Biobank und FinnGen. Je nach Fragestellung und Endpunkt wurden logistische Regression, Cox-Regression, maschinelles Lernen, MOFA und Enrichment-Analysen eingesetzt.

Die Ergebnisse zeigten, dass Ketonkörper und large-HDL-assoziierte lipidomische Komponenten MASLD mit relativ hoher Genauigkeit von MetALD unterscheiden konnten. Zudem war die regionale Körperzusammensetzung signifikant mit dem Mortalitätsrisiko bei Cholangiokarzinomen und deren Subtypen assoziiert und zeigte lokalisationsspezifische Muster. Weiterführende Analysen deuteten auf eine mögliche Relevanz auch für das pan-kanzeröse Risiko hin. Darüber hinaus wurden für spezifische Tumorentitäten biologisch interpretierbare Muster der Körperzusammensetzung identifiziert, denen potenziell Signalwege des Aminosäuretransports, der oxidativen Detoxifikation und neurotransmitterassoziierter Prozesse zugrunde liegen.

Insgesamt deuten die Befunde darauf hin, dass metabolische Dysregulation MASLD und Krebs über messbare metabolische Endotypen verbindet. Zirkulierende Multi-Omics-Merkmale und regionale Körperzusammensetzung bieten neue Ansätze für präzisere Krankheitscharakterisierung, Risikostratifizierung und mechanistische Forschung.

Contents

1. Introduction 1

1.1. Steatotic Liver Disease and Cancer within Global Public Health 1

1.2. The Definition and Epidemiological Burden of MASLD 2

1.2.1. The Challenges in the Clinical Management of MASLD 3

1.2.2. The Phenotypic Heterogeneity of MASLD and MetALD 5

1.3. The Global Burden of Cancer 7

1.3.1. The Diagnostic and Therapeutic Challenges in Cancer 8

1.4. Metabolic Endotypes and Measurable Readouts 10

1.4.1. NMR-Based Lipidomics for MASLD Endotyping 11

1.4.2. Regional Body Composition for Cancer Risk 13

1.4.3. Integrative Metabolomics and Proteomics for Molecular Endotype Mapping in Cancer 14

2. Research objective 17

3. Materials and methods 18

3.1.1. Data Sources and Study Populations 18

3.1.2. UK Biobank 18

3.1.3. FinnGen Consortium 18

3.1.4. Inclusion and Exclusion Criteria 18

3.2. Body Composition Measurements 20

3.3. NMR-Based Metabolomic Profiling 20

3.4. Proteomic Profiling 21

3.5. Outcomes 21

3.6. Covariates 23

3.7. Analytical Approaches 23

3.7.1. Logistic and Cox Regression Analyses 23

3.7.2. Mendelian Randomization Analyses 24

3.7.3. Machine Learning Models 25

3.7.4. Multi-Omics Factor Analysis 25

3.7.5. Functional and Pathway Enrichment Analyses 26

4. Results 27

4.1. Discriminative Metabolomics Signatures for MASLD and MetALD 27

4.1.1. Baseline Characteristics of MASLD and MetALD patients 27

4.1.2. Discriminative Metabolomics Markers Between MASLD and MetALD 28

4.1.3. Genetic Validation of Key Discriminative Metabolites 32

4.2. Associations and Prognostic Models Between regional BC and CCA Mortality 33
4.2.1. Baseline Characteristics of the CCA Cohort 33
4.2.2. Associations Between Body Composition and CCA Mortality 35
4.2.3. Performance and Comparative Evaluation of Prediction Models 36
4.2.4. Shapley-Based Interpretation of Model Contributions 38
4.3. Multi Omics Analysis between regional BC and Pan-Cancer 40
4.3.1. Baseline Characteristics of Pan-Cancer Cohort 40
4.3.2. Individual-Level Associations Between regional BC and Cancer Risk 41
4.3.3. Genetic Causal Associations Between regional BC and Cancer Risk 44
4.3.4. MOFA based Signatures and Associations between BC and Cancer 46
4.3.5. Enrichment Pathway of molecular metabolites and protein 51
4.3.6. The BC-Cancer Risk Association Atlas 53
5. Discussion 55
5.1. Discriminative NMR Metabolites for MASLD and MetALD 56
5.2. Regional BC and CCA Mortality 59
5.3. Pan-Cancer Atlas of BC with Multi-Omics Integration 62
5.4. Integrative Interpretation Within the Metabolic Endotype Framework 66
5.5. Limitations 68
5.6. Future Directions 69
6. Conclusion 71
7. Reference 72
8. Appendix 95
8.1. List of Tables 95
8.2. List of Figures 95
8.3. Acknowledgment 99
8.4. Affidavit according to § 5 (1) for Data Retention 101
8.5. Affidavit according to § 5 (1) and § 11 (3) 12. of the doctoral regulations 102

Abbreviation

NCDs	Non-communicable diseases
T2DM	Type 2 diabetes mellitus
IR	Insulin resistance
SLD	Steatotic liver disease
MASLD	Metabolic dysfunction-associated steatotic liver disease
HCC	Hepatocellular carcinoma
CCA	Cholangiocarcinoma
CRC	Colorectal cancer
MetALD	Metabolic dysfunction with alcohol-related liver disease
CMRF	Cardiometabolic risk factor
ALD	Alcohol-related liver disease
MASH	Metabolic dysfunction–associated steatohepatitis
CVD	Cardiovascular disease
CKD	Chronic kidney disease
IARC	International Agency for Research on Cancer
GCO	Global Cancer Observatory
ICC	Intrahepatic cholangiocarcinoma
ECC	Extrahepatic cholangiocarcinoma
VCTE	Vibration-controlled transient elastography
PDFF	Proton density fat fraction
CDT	Carbohydrate-deficient transferrin
NMR	Nuclear magnetic resonance
UKB	UK Biobank
GlycA	Glycoprotein acetylation composite signal
ICIs	Immune checkpoint inhibitors
BMI	Body mass index
VAT	Visceral adipose tissue
SAT	Subcutaneous adipose tissue
BC	Body composition
CT	Computed tomography
FM	Fat mass
FFM	Fat-free mass
BIA	Bioelectrical impedance analysis
BCAA	Branched-chain amino acids

2-HG	2-hydroxyglutarate
CEA	Carcinoembryonic antigen
CA19-9	Carbohydrate antigen 19-9
cfDNA	Cell-free DNA
CPTAC	Clinical Proteomic Tumor Analysis Consortium
MR	Mendelian randomisation
ML	Machine learning
SHAP	SHapley Additive exPlanations
PheWAS	Phenome-wide association study
MOFA	Multi-omics factor analysis
GWAS	Genome-wide association study
PEA	Proximity Extension Assay
NPX	Normalized Protein eXpression
ICD-10	International Classification of Diseases, 10th Revision
BH	Benjamini–Hochberg
FDR	False discovery rate
CI	Confidence intervals
OR	Odds ratio
HR	Hazard ratios
RCS	Restricted cubic splines
SNPs	Single-nucleotide polymorphisms
LD	Linkage disequilibrium
IVW	Inverse-variance weighted
MR-PRESSO	MR pleiotropy residual sum and outlier
RFE	Recursive feature elimination
SMOTE	Synthetic minority over-sampling technique
SVM	Support vector machine
GBM	Gradient boosting machine
XGB	EXtreme Gradient Boosting
NN	Neural network
MLP	Multilayer perceptron
RBF	Radial basis function
AUC	Area under curve
ELBO	Evidence lower bound
GO	Gene Ontology

KEGG	Kyoto Encyclopedia of Genes and Genomes
ALT	Alanine aminotransferases
AST	Aspartate aminotransferases
GGT	Gamma-glutamyl transferase
FP	Fat percentage
GML	Global metabolic and lean pattern
HDL	High-density lipoprotein
LFLB	Lower-body fat–lean balance pattern
DHA	Docosahexaenoic acid
LDL	Low-density lipoprotein
IDL	Intermediate-density lipoprotein
ULD	Upper-body lean dominant pattern
VLDL	Very low-density lipoprotein
GFLB	Global fat–lean balance pattern
GAD	Generalised adiposity-dominant pattern
TA–LLM	Trunk adiposity with limb lean mass pattern
CAD	Central adiposity-dominant pattern
SLC	Solute carrier
GABA	Gamma-aminobutyric acid
cAMP	Cyclic adenosine monophosphate
RCT	Reverse cholesterol transport
ApoA1	Apolipoprotein A1
CETP	Cholesteryl ester transfer protein
OS	Overall survival
IGF	Insulin-like growth factor
FFAs	Free fatty acids
FFMI	Fat-free mass index
BFMI	Body fat mass index
CLS-B	Crown-like structures of the breast
PGE2	Prostaglandin E2
ER	Oestrogen receptor
IL-6	Interleukin-6
TRL	Triglyceride-rich lipoprotein
RC	Remnant cholesterol
mTORC1	Mechanistic target of rapamycin complex 1

TLR	Toll-like receptor
ROS	Reactive oxygen species
CSCs	Cancer stem cells
R5P	Ribose-5-phosphate
NADPH	Nicotinamide adenine dinucleotide phosphate
PPP	Pentose phosphate pathway
NK	Natural killer

1. Introduction

1.1. Steatotic Liver Disease and Cancer within Global Public Health

Over the past half-century, the global disease landscape has undergone a structural transition, in which the dominant contributors to disease burden have progressively shifted from communicable diseases to non-communicable diseases (NCDs) typified by metabolic syndrome, including obesity, type 2 diabetes mellitus (T2DM) and dyslipidaemia [1]. Meanwhile, the global burden of cancer has continued to rise and shows substantial overlap with these metabolic risk factors [2,3]. At the pathophysiological level, this transition highlights the central role of metabolic dysregulation as a key systemic derangement [4]. Increasing evidence indicates that metabolic dysregulation constitutes a key part of chronic disease risks across multiple organs, and its contribution to multisystem disease aggregation has become increasingly evident [4,5].

Metabolic dysregulation refers to a systemic disturbance of energy homeostasis across key processes, including nutrient intake, storage, and oxidative utilization [6]. Hallmark features of metabolic dysrefulation include excessive and ectopic adipose accumulation, insulin resistance (IR), and disturbances in glucose and lipid metabolism [7,8]. These features are frequently accompanied by low-grade chronic inflammation, heightened oxidative stress, and mitochondrial dysfunction. Within this continuum, the liver is highly susceptible to metabolic stress because it serves as a central hub that integrates glucose and lipid fluxes, coordinates insulin signalling, and regulates whole-body energy allocation [9]. Metabolic dysfunction-associated steatotic liver disease (MASLD), can be considered a representative hepatic phenotype of metabolic dysregulation and is characterised by fat deposition in the liver as well as cardiometabolic dysfunction [10]. It is closely linked to the long-term accumulation of both intrahepatic end-stage outcomes like cirrhosis and extrahepatic complications [11,12]. Moreover, steatotic liver disease (SLD) reflects not only local hepatic lipid accumulation and chronic liver injury, but also, to some extent, the persistent presence of systemic metabolic dysregulation and its multi-organ consequences [13].

However, SLD is not the only liver-related outcome of metabolic dysregulation. The impact of metabolic abnormalities extends further to cancer development and the associated mortality burden [14]. In recent years, the sustained rise in the cancer burden exhibits an epidemiological pattern that is closely intertwined with phenotypes of

metabolic dysregulation [15,16]. Prior population-based studies have shown that a substantial proportion of incident cancers and cancer-related deaths occur among individuals with metabolic dysregulation phenotypes, including obesity, T2DM, and MASLD [17,18]. Moreover, accumulating evidence suggests that excessive adiposity and aberrant fat distribution, IR, and disordered lipid metabolism are associated with an elevated risk of multiple malignancies, including hepatocellular carcinoma (HCC), cholangiocarcinoma (CCA), colorectal cancer (CRC), and female breast cancer [19–21]. Collectively, these findings further support that metabolic abnormalities are not only strongly linked to cardiovascular and liver-related outcomes but are also importantly associated with the risk of cancer development and the burden of cancer-related mortality. This further suggests that metabolic dysfunction may constitute an important shared pathological basis linking SLD and cancer.

1.2. The Definition and Epidemiological Burden of MASLD

A multi-society Delphi consensus has updated the nomenclature and stratification framework for SLD, stating that MASLD and the phenotype of metabolic dysfunction with alcohol-related liver disease (MetALD) constitute two major subtypes within the SLD spectrum [10]. Both entities share core features of hepatocellular lipid accumulation and accompanying metabolic remodelling, and their development is driven by complex interactions among genetic susceptibility, lifestyle factors, and environmental exposures [22,23]. MASLD is defined by evidence of hepatic steatosis in the presence of at least one cardiometabolic risk factor (CMRF), while excluding other specific causes of steatosis [24]. With respect to alcohol exposure, MASLD requires alcohol intake below established thresholds (women <140 g/week; men <210 g/week), thereby emphasising the predominant role of metabolic abnormalities in disease initiation. MetALD, in turn, describes individuals who meet the metabolic criteria for MASLD but have moderate alcohol exposure, defined as alcohol intake exceeding the above thresholds yet remaining below the higher levels typically associated with alcohol-related liver disease (ALD) (women 140–350 g/week; men 210–420 g/week) [24].

In recent years, MASLD has become a major contributor to the global burden of NCDs, characterised by both high prevalence and a sustained upward epidemiological trend [25]. Contemporary estimates suggest that the global prevalence of MASLD is approximately 38%, with substantial regional heterogeneity. The estimated prevalence is higher in Latin America (approximately 44.4%) and comparatively lower in Western

Europe (approximately 25.1%) [26,27]. Notably, the burden of MASLD has increased markedly over the past decade. By 2021, approximately 1.27 billion individuals worldwide were affected, and modelling studies project that the prevalence may further rise to 55.4% by 2040 [24,28]. Importantly, MASLD is not a static phenotype of hepatic fat accumulation. Instead, it may progress along a clinical spectrum to metabolic dysfunction–associated steatohepatitis (MASH), fibrosis, and cirrhosis, thereby increasing the risk of HCC [24]. Previous studies have reported that approximately 12%–40% of individuals with simple steatosis may progress to MASH within 8–13 years. Among individuals with MASH and established fibrosis, around 15%–25% may further develop cirrhosis over a similar follow-up interval [29]. In addition, among patients with MASLD-related compensated cirrhosis, approximately 7% may develop HCC within 10 years, and a considerable proportion ultimately undergo liver transplantation or die from liver-related events [30]. Beyond intrahepatic progression, MASLD is also associated with an increased risk of extrahepatic outcomes, including cardiovascular disease (CVD), T2DM, and chronic kidney disease (CKD) [18]. Lee et al.[31,32] reported that compared with individuals without SLD and without CMRFs, patients with MASLD had approximately 38% and 70% higher risks of CVD and CKD events, respectively.

Therefore, MASLD is characterized not only by a high prevalence and a continuously progressive natural course, but also by close associations with a wide range of adverse intrahepatic and extrahepatic outcomes. These features, including its broad clinical burden, marked heterogeneity, and progressive nature, pose substantial challenges to early identification, precise diagnosis, and risk stratification [33].

1.2.1. The Challenges in the Clinical Management of MASLD

Current evidence suggests that intrahepatic and extrahepatic outcomes in MASLD are not determined solely by hepatic phenotypes but are also influenced by CMRFs [34]. These outcomes are jointly shaped by coupled processes of intrahepatic disease progression and extrahepatic risk [35,36]. Multiple studies consistently indicate that fibrosis stage is a key predictor of MASLD progression and prognosis, and its prognostic value often exceeds that of steatosis severity or inflammatory activity alone [37]. The severity of metabolic abnormalities, particularly T2DM, also substantially drives fibrosis progression and the burden of multisystem adverse outcomes [38,39]. However, current clinical practice often remains focused on detecting fatty liver, while early identification of individuals at high risk of fibrosis and systematic management of CMRFs are still

insufficient [13]. This gap contributes to under-recognition of individuals at truly elevated risk and delays in referral. In addition, existing HCC surveillance strategies in MASLD may miss a subset of patients [40]. Some evidence suggests that MASLD-related HCC does not invariably follow the canonical pathway of cirrhosis preceding HCC, and that a proportion of cases can arise in the absence of cirrhosis [41]. In a study by Souvik and colleagues, 20%–50% of HCC cases occurred among individuals with MASLD without cirrhosis [42]. Nevertheless, most current HCC surveillance strategies primarily target patients with cirrhosis, and more precise risk stratification approaches for non-cirrhotic MASLD populations remain limited [43].

In terms of diagnosis and risk assessment, abdominal ultrasonography has traditionally been used as the first-line tool to detect hepatic steatosis [44]. However, ultrasound has limited sensitivity for mild steatosis and its performance is strongly influenced by operator dependence and body habitus [45]. Clinical guidance also notes that ultrasound has inadequate detection capability when hepatic fat content is below 20%, and therefore a negative ultrasound result cannot reliably exclude MASLD [45]. Given the strong association between fibrosis stage and clinical outcomes, the EASL–EASD–EASO guideline recommends prioritising blood-based risk scores such as FIB-4 for initial triage, followed by further risk stratification using liver elastography, including vibration-controlled transient elastography (VCTE), LiverMultiScan, or other non-invasive tests[13]. Nevertheless, real-world implementation of non-invasive assessment pathways faces multiple constraints. Thresholds for risk scores are difficult to standardise across settings and are affected by age, the degree of obesity, inflammatory status, and comorbidities [46,47]. VCTE may be unreliable in adults with class II obesity, and the limited availability and high cost of LiverMultiScan restrict broader adoption [13,48]. In addition, although proton density fat fraction (PDFF) is considered a non-invasive reference standard for quantitative liver fat assessment in MASLD, its primary strength lies in fat quantification and it provides comparatively limited information for fibrosis staging [49]. Collectively, these limitations indicate that existing non-invasive tools remain suboptimal with respect to precision, consistency, and the stability of decision thresholds across clinical contexts.

Notably, the current challenges in the diagnosis and risk assessment of MASLD arise not only from the limited performance of existing diagnostic tools but are also closely related to the intrinsic phenotypic heterogeneity of SLD itself [13]. Although different SLD subtypes may share the common feature of hepatic steatosis, they are not

uniform in their pathogenic drivers, the relative contributions of metabolic dysfunction and alcohol exposure, patterns of disease progression, or the risks of intrahepatic and extrahepatic outcomes [24]. Therefore, a deeper investigation into the phenotypic heterogeneity of SLD and its implications for precise clinical diagnosis and risk stratification is of considerable theoretical and practical importance.

1.2.2. The Phenotypic Heterogeneity of MASLD and MetALD

With the updated SLD nomenclature, MetALD should not be regarded as a simple extension of MASLD. Instead, it represents an overlapping phenotype jointly driven by metabolic abnormalities and alcohol exposure [50]. Importantly, risk differences between MetALD and MASLD are not determined solely by average alcohol intake [10]. Drinking patterns and variations in the exposure time can also modify the cumulative burden of liver injury, thereby contributing to heterogeneity across the exposure spectrum [51]. In MetALD, metabolic load and alcohol exposure may act synergistically, accelerating inflammatory and fibrotic processes [50]. As a result, even when comparable levels of hepatic fat accumulation are present, MetALD may be more likely than MASLD to progress to adverse liver outcomes [36]. Evidence further indicates that, compared with individuals without SLD, the risks of cirrhosis and HCC are elevated across SLD subtypes [52]. A gradient of risk has been reported, with the highest risk in ALD, followed by MetALD and then MASLD, which suggests that alcohol exposure may further amplify adverse liver outcomes in the context of metabolic abnormalities [53]. In addition, a recent systematic review and meta-analysis reported that MetALD is associated with a higher risk of liver-related events than MASLD, whereas differences in cardiovascular events and all-cause mortality were not statistically significant [54]. Taken together, these findings imply that metabolic burden may be a dominant driver of cardiovascular risk, while alcohol exposure may more prominently intensify progression to end-stage liver outcomes and hepatobiliary carcinogenesis.

In real-world clinical practice, standardised management of SLD still faces challenges in accurate classification and differential diagnosis [55]. At present, distinguishing MASLD from MetALD relies largely on patient self-reported alcohol consumption [10]. However, the reliability and reproducibility of self-reported alcohol intake have been questioned because of recall bias, social desirability bias, and stigma related to excessive drinking [56,57]. In addition, some patients have limited awareness of their actual alcohol intake, particularly when consuming mixed beverages with

relatively high alcohol content. Beyond total intake, drinking patterns, beverage type, drinking speed, and cumulative duration of alcohol use further increase the complexity of phenotyping [51,58]. One study reported that among individuals initially diagnosed with MASLD, up to 29% had positive alcohol biomarkers, which suggests potential misclassification due to inaccurate reporting of alcohol consumption [59]. Moreover, it is difficult in clinical practice to determine the predominant driver and relative contribution of metabolic factors versus alcohol exposure based solely on clinical phenotype and non-invasive assessments such as ultrasound, FIB-4, or PDFF [13]. Although alcohol-related biomarkers, particularly phosphatidylethanol (PEth), ethyl glucuronide/ethyl sulfate (EtG/EtS), and carbohydrate-deficient transferrin (CDT), can improve the detection of covert alcohol consumption [60,61], emerging evidence suggests that PEth, as an objective marker of recent alcohol intake, has relatively strong potential for clinical application and may facilitate more accurate identification of SLD subcategories, thereby reducing diagnostic misclassification [62]. However, the clinical interpretation of these biomarkers remains influenced by the detection window, the choice of cutoff values, and patterns of alcohol consumption [63,64]. In addition, although liver biopsy allows direct assessment of histological changes such as steatosis, hepatocellular ballooning, and fibrosis, and has long been regarded as an important reference standard for characterizing the histological spectrum of steatotic liver disease [65], its invasiveness, sampling variability, and limited feasibility for repeated assessment substantially restrict its use in routine clinical practice and large-scale population studies [66]. Accordingly, further work is needed to develop more objective, quantifiable, and scalable biomarkers or integrated assessment frameworks to improve the accuracy of differentiating MASLD from MetALD.

In summary, MASLD is not only a representative clinical manifestation of metabolic dysregulation at the hepatic level, but also a major public health concern because of its high prevalence, substantial clinical heterogeneity, and the complexity of risk stratification. Moreover, the health consequences of metabolic dysregulation are not limited to hepatic lipid accumulation and chronic liver injury but further extend to adverse outcomes involving multiple organs and systems, among which malignant neoplasms warrant particular attention [12]. Given that cancer remains one of the leading contributors to the global burden of disease, further clarification of its epidemiological burden and the challenges of clinical management will not only help to delineate the overall landscape of the evolving spectrum of metabolic dysfunction-related diseases

from a public health perspective, but also provide essential background for subsequent investigations into the role of metabolic factors in MASLD and cancer, as well as their potential underlying mechanisms.

1.3. The Global Burden of Cancer

Cancer remains one of the most burdensome disease categories worldwide in terms of morbidity and mortality, and its epidemiological burden continues to increase [67]. According to summary estimates from the International Agency for Research on Cancer (IARC) Global Cancer Observatory (GCO) based on GLOBOCAN 2022, there were nearly 20 million new cancer cases and approximately 9.7 million deaths globally in 2022 [67]. These figures correspond to an estimated lifetime risk of about one in five individuals developing cancer. The same report predicted that by 2050 the annual number of cancer cases will rise to approximately 35.3 million (an increase of around 76.6% relative to 2022), and annual deaths will reach about 18.5 million (an increase of around 89.7%) [67]. These projections indicate that global cancer prevention and control will face increasingly severe long-term challenges. With respect to cancer type composition, lung cancer remains among the most frequently diagnosed cancers worldwide, followed by female breast cancer and CRC [67]. In terms of mortality, lung cancer also ranks first globally. Liver cancer (liver and intrahepatic bile ducts) represents the third leading cause of cancer death worldwide, and a meaningful proportion has been attributed to intrahepatic cholangiocarcinoma (ICC), which is estimated at approximately 15% in prior reports [67,68]. Together, these observations underscore the substantial contribution of hepatobiliary malignancies to the global cancer burden.

Notably, although CCA has a relatively low incidence in the general population, it is highly aggressive and is generally associated with poor prognosis. In recent years, its epidemiological burden has shown a sustained upward trend across multiple regions [69]. Available evidence indicates marked geographic heterogeneity in CCA incidence, with an overall higher burden reported in mainland Asia [69,70]. At the same time, mortality trends also differ across regions and have increased in several European countries over the past decade, with a more pronounced rise for ICC. Using the Baltic states as an example, the average annual change in ICC mortality was approximately 18.14% and 23.74% among men in Latvia and Lithuania, respectively, and 64.69% and 18.87% among women, respectively [69,71]. In addition, countries with particularly notable

increases in mortality from extrahepatic cholangiocarcinoma (ECC) include Hungary, Norway, and Spain [69,71]. However, the seriousness of cancer is reflected not only in its continuously increasing epidemiological burden, but also in the considerable complexity of its clinical diagnosis and management. For many cancers, including CCA, difficulty in early detection, marked disease heterogeneity, limited therapeutic options, and insufficient improvement in prognosis remain core challenges in current clinical practice [72]. Therefore, on the basis of clarifying its disease burden, it is necessary to further examine the diagnostic and therapeutic challenges associated with cancer.

1.3.1. The Diagnostic and Therapeutic Challenges in Cancer

Beyond the epidemiological burden described above, cancer prevention and control continue to face multiple challenges even in an era of rapidly evolving diagnostic and therapeutic technologies. One major reason is the pronounced heterogeneity of cancer [73]. Tumours arising from different organ systems differ fundamentally in their carcinogenic exposure spectra, tissue origins, and treatment sensitivities. Pan-cancer genomic studies have identified multiple mutational signatures associated with exogenous exposures, with smoking-related signatures being more prominent in lung cancer and ultraviolet-related signatures being more characteristic of skin cancer [74]. Importantly, substantial heterogeneity also exists within the same cancer type, where molecular subtypes and the tumour microenvironment display marked diversity [75]. For example, molecular subtypes of breast cancer, including Luminal A/B, HER2-enriched, and Basal-like tumours, are associated with distinct biological behaviours, treatment responses, and clinical outcomes[76]. Cancer prevention and control are further constrained by limited capacity for early detection and prognostic prediction. Current evidence-based screening strategies primarily cover a small number of cancers, such as breast cancer, cervical cancer, CRC, and lung cancer in selected high-risk populations, indicating that broadly applicable tools for early detection and risk stratification remain insufficient from a pan-cancer perspective [77]. Pancreatic cancer provides a representative example, as studies have reported that more than half of patients have already developed distant metastases at the time of diagnosis [78]. CCA presents even greater clinical challenges. Its early manifestations often lack specificity, and its complex anatomical location can constrain diagnostic pathways. Curative opportunities are limited, recurrence rates are high, and overall prognosis remains poor [79,80]. Although surgical resection remains the primary curative-intent treatment, only

approximately 20%–30% of newly diagnosed patients are eligible for upfront curative resection [81]. Even after curative-intent resection, recurrence after ICC is common, with reported recurrence rates of approximately 60%–70% [81]. A substantial proportion of recurrences occur early after surgery, which suggests that the disease burden may be underestimated in routine clinical practice [82].

Although cancer research has made substantial advances in circulating biomarkers, transcriptomics, and epigenetics in recent years, the role of host physiological and metabolic status in cancer development and outcomes has received comparatively less attention [83,84]. Cancer biology has increasingly recognised metabolism as a key pathophysiological hallmark, suggesting that metabolic processes are not merely passive correlates of malignancy but can actively contribute to tumour growth and inter-individual differences in clinical outcomes [85]. At the individual level, the metabolic background may influence susceptibility to carcinogenic exposures as well as tumour progression and prognosis [86]. A meta-analysis of patients treated with immune checkpoint inhibitors (ICIs) reported that individuals in higher body mass index (BMI) categories had better overall survival than those in lower BMI categories [87]. Dose–response analyses further suggested that each 1 kg/m² increase in BMI was associated with an approximately 3.6% lower risk of death, indicating that metabolism-related host factors may contribute to variability in immunotherapy response [87]. In addition, a study by Bea and colleagues reported that visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) were each associated with increased breast cancer risk [88]. However, metabolic status is readily influenced by medication use, comorbidities, and fluctuations in organ function. Many metabolism-related indicators show limited reproducibility and generalisability across cohorts and treatment contexts, which complicates stable quantification and hampers translation into reliable tools for risk stratification and clinical decision-making [89]. This challenge may be particularly pronounced in CCA, where underlying hepatobiliary disease, cholestasis, and recurrent inflammation can further perturb metabolic signals and exacerbate confounding, thereby increasing the difficulty of early detection and prognostic assessment [90].

In summary, MASLD and cancer not only jointly constitute major global public health burdens but also exhibit substantial complexity in clinical management. This complexity arises not merely from their apparent clinical heterogeneity but is also closely related to differences in the degree and pattern of the underlying metabolic dysregulation [13]. However, traditional phenotypic frameworks and assessments based on

conventional clinical indicators are often insufficient to fully capture these deeper biological differences [91]. In recent years, precision medicine research has increasingly emphasized that disease characterization should move beyond phenotype to endotype, that is, from observable clinical manifestations to disease subtypes defined by specific pathobiological mechanisms [92]. Therefore, advancing the understanding of disease heterogeneity at the level of metabolic endotypes, together with systematic characterization using multidimensional and measurable biological readouts, may facilitate more precise characterization and stratification of MASLD and cancer.

1.4. Metabolic Endotypes and Measurable Readouts

Metabolic endotypes refer to metabolically distinct states that arise within a background of similar clinical phenotypes, are driven by specific pathobiological mechanisms, and can be identified by a set of reproducible and interpretable metabolic molecules, proteins, and phenotypic indicators [93]. Unlike traditional clinical phenotypes, metabolic endotypes place greater emphasis on the underlying mechanistic basis of disease and primarily involve multiple biological processes, including impaired lipid handling, lipoprotein remodeling, insulin resistance, disordered amino acid metabolism, and disrupted muscle adipose tissue interaction [92,94]. With the rapid development of metabolomics, lipidomics, and integrative multi-omics analytical approaches, metabolic endotypes have gradually emerged as an important entry point for disease risk stratification, mechanistic investigation, and precision medicine research [92]. Existing evidence generally suggests that multiple metabolism related diseases, including MASLD, CVD, and cancer, are not isolated entities, but are more likely to represent continuous manifestations across different organs after the long-term accumulation and interconnection of distinct metabolic driving axes [95–97].

Metabolic endotypes can be approximated through a series of measurable readouts. First, at the level of circulating molecules, metabolomics and proteomics provide an important window for identifying metabolism related driving axes [98]. Plasma lipid subclasses and fatty acid composition not only reflect hepatic lipid export and patterns of energy substrate utilization, but also reveal the state of inflammation and metabolic coupling [99]. Previous studies have shown that the circulating abnormalities associated with MASLD are not limited to elevated triglycerides, but involve broader remodeling of metabolic networks, including alterations in fatty acid composition, abnormalities in lipoprotein subclasses, and disturbances in amino acid metabolism [100]. A study based on

the UK Biobank further showed that lipid signals centered on high density lipoprotein (HDL) related markers could, to some extent, distinguish MASLD from MetALD, suggesting that relatively stable and identifiable circulating metabolic fingerprints may underline different SLD subtypes [101]. In addition, at the level of host structure, regional BC represents another key category of readout for metabolic endotypes. Compared with traditional anthropometric measures such as BMI or waist circumference, regional BC can more precisely distinguish the amount and distribution of fat and muscle across different anatomical sites, and therefore more closely reflects the actual metabolic burden and host reserve status [102]. Studies have shown that individuals with the same BMI may have markedly different fat burden and muscle reserve, and that these differences are strongly associated with insulin resistance, lipotoxicity, and inflammatory immune responses [103]. For example, Christakoudi, et al. [104,105] showed that increased visceral adiposity and muscle fat infiltration were associated with a higher risk of cancer, and that these associations were independent of conventional obesity indicators.

The circulating multi-omic features and regional BC are neither independent nor mutually substitutable indicators, but rather two complementary readouts of the same underlying metabolic endotype [106]. The former primarily reflects dynamic molecular perturbations in disease related metabolic networks, whereas the latter captures the long-term cumulative remodeling of metabolic endotypes at the anatomical and physiological levels [92,107]. Therefore, metabolic endotypes can be systematically characterized from both circulating molecular features and host structural features, thereby providing a theoretical foundation for subsequent analyses from the perspectives of lipidomics, body composition, and multi-omics integration.

1.4.1. NMR-Based Lipidomics for MASLD Endotyping

At the level of circulating molecular features, lipidomics provides an important entry point for characterizing MASLD related metabolic endotypes. Given that the biological heterogeneity of MASLD and MetALD is closely linked to lipid metabolic remodeling, molecular features captured by lipidomics, including lipoprotein profiles and fatty acid composition, are of substantial importance for the precise identification of different subtypes and for prognostic assessment [101]. Among available platforms, nuclear magnetic resonance (NMR) metabolomics offers high throughput, good reproducibility, and the capability for absolute quantification [108]. It is particularly well

suited for robust measurement of lipoprotein particles and their subclasses, fatty acid composition, and metabolites related to glycolysis [108]. Large population-based cohorts, exemplified by the UK Biobank (UKB), have released NMR metabolomics datasets covering approximately 249 metabolic measures. These data provide a foundation for biomarker discovery, risk prediction, and causal inference, and they have facilitated the application of NMR lipidomics in disease research [109,110]. For example, Zhang and colleagues used UKB NMR metabolomics data to derive a metabolomic signature of biological ageing and reported that a metabolomic ageing score may have potential value in predicting mortality and multi-disease risk [111]. Julkunen and colleagues similarly leveraged UKB NMR metabolomic profiles to conduct a systematic association analysis across more than 700 disease outcomes, highlighting the potential of metabolomics in public health and precision medicine research [112].

In MASLD-related research, lipoprotein profiles, particularly the spectrum of triglyceride-rich lipoproteins, can reflect the dynamic balance between hepatic lipid export and peripheral lipid return. These profiles have been considered informative indicators for characterising the coupling between hepatic steatosis and cardiometabolic risk [113,114]. Previous NMR-based studies have reported associations between signals related to triglyceride transfer protein and acute-phase glycoproteins and the risk of developing SLD [115]. Fatty acid composition and ratios may serve, to some extent, as objective surrogates of long-term dietary patterns and metabolic status, and they have been linked to adverse outcomes in chronic liver disease, including risk of HCC [116]. In addition, NMR enables quantification of the glycoprotein acetylation composite signal (GlycA), which captures the integrated inflammatory burden of multiple acute-phase glycoproteins. Compared with single inflammatory markers, GlycA is more stable and may better characterise the low-grade chronic inflammatory phenotype observed in MASLD [117].

Although existing studies suggest that NMR lipidomics holds considerable promise for the stratification and prognostic assessment of SLD, metabolomics research focusing on biomarkers specific to MASLD and MetALD remains relatively limited [101]. This is particularly the case in the context of the coexistence of different patterns of alcohol exposure and variations in metabolic burden, where the precise discrimination of distinct subtypes and accurate risk prediction still require further investigation. At the same time, circulating lipidomic features mainly reflect dynamic molecular perturbations within metabolic networks and are less able to adequately capture the

cumulative remodeling of long-term metabolic burden at the host level [118]. Particularly in cancer risk research, structural differences such as fat mass and fat distribution often constitute important intermediates linking metabolic dysregulation to tumor susceptibility [119]. Therefore, a systematic characterization of metabolic endotypes still requires complementary structural readouts at the host level. Compared with circulating molecular features, regional BC is better able to reflect the cumulative effects of long-term metabolic burden and thus provides another important perspective for further understanding the metabolic heterogeneity underlying cancer risk.

1.4.2. Regional Body Composition for Cancer Risk

Body composition (BC), as an important host-level readout of physiological status and metabolic endotypes, can to some extent reflect energy reserves, metabolic burden, and functional reserve, and is of considerable value in the stratification of cancer occurrence and prognosis [120,121]. Prior studies have shown that excess adiposity is associated with increased risk of multiple cancers, including breast cancer, CRC, and pancreatic cancer [122–124]. A cohort study including 427,929 participants reported that fat-free mass was associated with an increased risk of gastric cancer [125]. In addition, the IARC has reported causal evidence linking excess body fatness to the risk of cancers at multiple anatomical sites [126]. Sanchez further reported that exercise and nutritional interventions targeting BC can improve outcomes among patients with bladder cancer, which suggests that BC is not only associated with cancer risk but may also be modifiable with translational relevance [127]. Moreover, BC-related phenotypes such as sarcopenia have been associated with adverse cancer outcomes, increased treatment toxicity, and shorter survival [128,129]. Watanabe, et al.[130] reported that BMI and sarcopenia were associated with poorer survival across multiple cancer types. In the context of cholangiocarcinoma, studies have similarly suggested that patients with concomitant sarcopenia experience significantly reduced overall survival [131].

However, important methodological limitations remain in existing literature. Many studies still rely on relatively coarse indicators, such as BMI, waist circumference, or sarcopenia, which are insufficient to capture more granular metabolic features, including fat distribution, muscle mass, and regional heterogeneity [132,133]. With advances in imaging and automated analytic approaches, BC assessment is shifting from conventional anthropometric indices towards more refined measurements. Recent studies have suggested that computed tomography (CT)-derived muscle and adipose

parameters can serve as opportunistic imaging biomarkers with potential utility for identifying cancer-related vulnerability and for prognostic assessment [134,135]. BC is defined not only by total quantity but also by anatomical location and tissue composition. Evidence indicates that fat mass (FM) and fat-free mass (FFM) in the limbs and trunk exhibit distinct metabolic characteristics and may exert differential effects on cancer development and progression [136,137]. In a CCA-related analysis, Wan and colleagues reported inverse associations between FM and FFM and the risk of CCA [138]. Another study suggested that leg FM may be more strongly associated with bladder cancer risk than arm or trunk FM [139]. Traditional measures such as BMI and waist circumference do not accurately reflect FM and FFM or their regional distribution, which limits the investigation of site-specific BC phenotypes in relation to cancer risk. In contrast, bioelectrical impedance analysis (BIA) provides a relatively convenient approach for estimating segmental BC by assessing body water distribution and soft tissue composition [140]. However, regional BC assessed by BIA remains insufficient to fully elucidate the molecular pathway networks underlying these differences and their links to cancer initiation and progression. Therefore, on the basis of host phenotypic characterization, the further integration of multi-omics information at the molecular level is of considerable importance for the more systematic identification of cancer-related metabolic endotypes.

1.4.3. Integrative Metabolomics and Proteomics for Molecular Endotype Mapping in Cancer

The integration of metabolomics and proteomics provides a research approach with greater mechanistic interpretability for the characterization of cancer-related metabolic endotypes. In particular, metabolomics can capture dysregulated signals related to glucose and lipid metabolism, amino acid metabolism, and redox homeostasis, thereby creating opportunities for biomarker discovery, risk stratification, and mechanistic investigations in cancer [141,142]. For example, Mayers and colleagues, in analyses across multiple cohorts, reported that elevated branched-chain amino acids (BCAA) were associated with a substantially increased risk of pancreatic cancer [143]. Dang, et al.[144,145] demonstrated that IDH1 mutations can drive tumour cells to produce and accumulate 2-hydroxyglutarate (2-HG), an oncometabolite that not only serves as a hallmark of metabolic reprogramming but can also be detected non-invasively using techniques such as NMR. These findings highlight the potential value of metabolomics for clinical

risk signal detection and diagnostic support. In addition, research targeting metabolic pathways has shown considerable promise for cancer therapy. Evidence indicates that ivosidenib, a targeted inhibitor for IDH1-mutant tumours, suppresses mutant enzyme activity and reduces 2-HG–related metabolic abnormalities [146]. In the phase III ClarIDHy clinical trial, ivosidenib was reported to reduce the risk of disease progression or death by 63% among patients with advanced CCA [146,147].

In contrast, proteomics more directly reflects pathway activity and the status of molecular signalling. In clinical practice, established circulating tumour markers such as carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) have long been used for adjunctive assessment and follow-up surveillance in CRC and pancreatic cancer [148,149]. CancerSEEK, proposed by Cohen et al., combines circulating protein markers with circulating cell-free DNA (cfDNA) mutation profiling. In eight cancer types, the reported median sensitivity was 70% and the specificity reached 99%, supporting the feasibility of incorporating protein biomarkers into multi-analyte cancer detection and indicating potential incremental gains [150]. Importantly, proteomics also offers irreplaceable advantages for mechanistic studies and drug development. Proteogenomic efforts such as those led by the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC) integrate genomic data with proteomic and phosphoproteomic profiles to elucidate the functional consequences of somatic alterations. These studies have helped prioritise candidate driver genes within recurrent copy number–altered regions and have shown, in cancers such as breast cancer, that protein and phosphorylation layers can more accurately represent targetable pathway activity [151]. Nevertheless, although the value of metabolomics and proteomics for biomarker discovery, risk stratification, and mechanistic interpretation has become increasingly recognised, key gaps remain in cancer risk research frameworks that start from inter-individual differences in host metabolic phenotypes. Specifically, it remains unclear which metabolites and proteins can form robust and scalable intermediate molecular phenotypes, and which pathway networks primarily mediate the links between host metabolic status and tumour initiation and progression.

In summary, MASLD and cancer are not isolated disease entities but are more likely to share a common foundation of metabolic dysregulation driven by lipid metabolic imbalance, inflammatory activation, and energy metabolic reprogramming [7,10]. Compared with traditional disease classification approaches based on external clinical phenotypes, the concept of metabolic endotypes places greater emphasis on the

mechanistic basis underlying disease heterogeneity, thereby providing a more integrative theoretical perspective for explaining the potential biological continuity between MASLD and cancer [92,94]. Within this framework, circulating multi-omics features and regional body composition can serve as complementary approaches for characterizing underlying metabolic endotypes at the molecular and structural levels, respectively. The former helps capture dynamic molecular perturbations involving lipid metabolism, amino acid metabolism, and inflammatory responses, whereas the latter reflects the cumulative burden of long-term metabolic stress, the remodeling of fat and muscle distribution, and the status of functional reserve [92,107]. Their integration not only facilitates a more precise delineation of disease-related metabolic heterogeneity and associated differences in risk but also provides an important foundation for further mechanistic investigation. On this basis, the present thesis proposes the following hypothesis: metabolic endotypes captured by circulating multi-omic features or BC constitute an important biological bridge linking MASLD and cancer, particularly hepatobiliary malignancies. These measurable metabolic endotypes may not only enable the precise identification of disease heterogeneity and risk differentials but may also further reveal key pathways through which metabolic dysregulation influences tumor initiation and progression.

2. Research objective

Based on the aforementioned theoretical framework, the study takes metabolic endotypes as its central conceptual axis and integrates circulating multi-omic features with regional body composition measures to precisely identify and characterize the key metabolic endotypes underlying disease heterogeneity and risk differences from both the circulating molecular and host phenotypic perspectives. It further aims to explore their potential pathways and biological significance in metabolic dysfunction-associated steatotic liver disease (MASLD) and cancer, thereby providing a foundation for the future optimization of clinical risk management and the development of targeted intervention strategies.

The specific objectives of this thesis are as follows.

First, to characterize the metabolic heterogeneity related to MASLD on the basis of circulating lipidomic features and to explore lipidomic characteristics of metabolic endotypes that may enable the precise identification of MASLD.

Second, to evaluate the associations between host metabolic phenotypes and cancer risk from the perspective of regional body composition, and to identify key body composition features associated with tumor development.

Third, to integrate metabolomic and proteomic information to delineate cancer-related molecular metabolic endotypes and to explore their potential biological pathways.

3. Materials and methods

3.1.1. Data Sources and Study Populations

3.1.2. UK Biobank

The UKB is a large prospective population-based cohort study. Between 2006 and 2010, more than 500,000 adults aged 37 to 73 years were recruited from 22 assessment centres across England, Wales, and Scotland [152]. At baseline, participants completed standardised questionnaires and underwent physical measurements and biospecimen collection. UKB also conducts ongoing linkage to national health record systems, which provides access to hospital admission data, cancer registry records, death registrations, and other healthcare information. This linkage supports long-term follow-up analyses of incident disease and mortality outcomes. All participants provided written informed consent at enrolment. UKB data collection and use are conducted under ethical approval and in accordance with established data governance procedures. The present study was performed under UKB application numbers [245437 and 71300].

3.1.3. FinnGen Consortium

FinnGen is a national research initiative jointly advanced by multiple Finnish academic institutions, hospital-based biobanks, and industry partners. The project leverages Finland's biobank infrastructure to collect biospecimens and perform genotyping and genome-wide imputation. It also conducts continuous linkage to nationwide health registries in Finland to systematically capture phenotypic information, including inpatient and outpatient diagnostic records, prescription data, cancer registry records, and death registrations [153]. Based on this integrated data resource, FinnGen provides high-quality genome-wide association study (GWAS) summary statistics across a wide range of disease endpoints, together with reproducible phenotype definitions. The overarching aim is to combine large-scale population genomic data with long-term follow-up from national health registries to support genetic discovery and translational medical research.

3.1.4. Inclusion and Exclusion Criteria

The study population was derived from baseline characteristics and follow-up data from approximately 500,000 UKB participants. Project-specific inclusion and exclusion criteria were defined according to each research question. For all analyses, the baseline assessment date was used as the start of follow-up. Participants were required to have the core information necessary to define the main exposures and outcomes. Individuals

with missing data on key exposure variables, outcome ascertainment, or follow-up information were excluded to minimise information bias and to improve reproducibility. For MASLD-related analyses, we included only participants with both liver MRI measurements and metabolomics data. We further excluded individuals for whom SLD subtype classification could not be established using clinical records and baseline alcohol intake information. For the CCA outcome analyses, we included participants with complete regional BC measurements and valid follow-up information. CCA mortality was defined using the primary cause of death from death registry records according to prespecified coding rules. For the pan-cancer incidence analyses, we included participants with complete regional BC measurements and follow-up data who had no history of cancer at baseline. Incident cancer outcomes were ascertained from cancer registry records and covered common cancers across multiple organ systems. For multi-omics integration, we further restricted analyses to individuals with matched regional BC, metabolomics, and proteomics measurements to ensure individual-level alignment across omics layers. Genetic analyses were conducted using a two-sample MR framework. For exposures, we used UKB GWAS summary statistics for regional BC traits. For outcomes, we selected pan-cancer GWAS data from the external, independent FinnGen cohort. This design was intended to minimise sample overlaps and strengthen the robustness of causal inference (Figure 1).

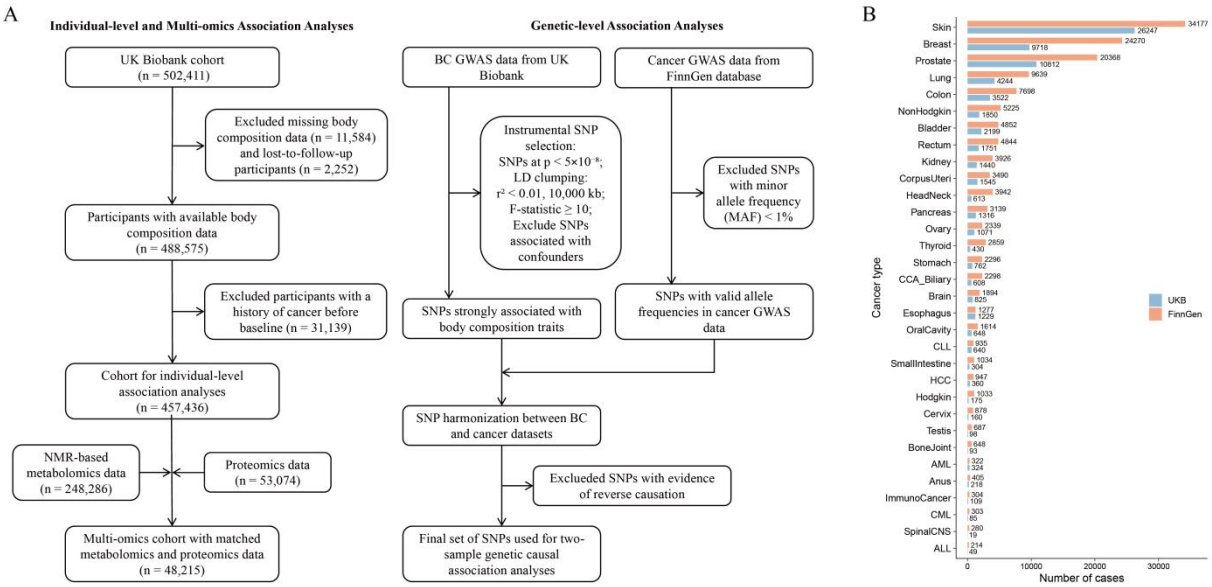


Figure 1. Study design and workflow. (A) Inclusion and exclusion flowchart for participant-level and genetic-level data. (B) Number of cancer cases in the UK Biobank and FinnGen datasets.

3.2. Body Composition Measurements

Regional BC data were obtained from baseline physical assessments in the UKB using the Tanita BC418MA bioelectrical impedance analyser, which provides whole-body and segmental bioimpedance measures. Under standardised protocols, this device generates multidimensional body composition outputs, including basal metabolic rate, FM, body fat percentage, FFM, total body water, predicted muscle mass, and impedance values for the limbs, trunk, and whole body [154]. Based on these measurements, we derived regional BC phenotypes that primarily capture differences in the distribution of adipose and fat-free tissues across the limbs and trunks. These phenotypes were used in two main research settings. First, in the CCA analyses, we evaluated associations between regional BC and mortality risk for CCA and its subtypes, and we incorporated regional BC as candidate predictors in prognostic modelling. Second, in the pan-cancer analyses, we systematically assessed association patterns between regional BC and pan-cancer incidence risk. In multi-omics integration analyses, regional BC was included as a host metabolic phenotype layer together with metabolomics and proteomics data to explore molecular phenotype axes related to regional BC and their potential pathway networks.

3.3. NMR-Based Metabolomic Profiling

NMR metabolomics data were obtained from the UKB high-throughput NMR platform, which performed quantitative profiling using plasma samples collected at baseline. This platform provides a multidimensional metabolite spectrum covering lipid and energy metabolism, including particle concentrations and lipid composition of lipoprotein subclasses, triglyceride- and cholesterol-related measures, total fatty acids and their compositional proportions, as well as amino acids and other energy metabolites. In this thesis, these data were primarily used in two research settings. First, in the SLD project, NMR lipidomics measures served as candidate phenotyping biomarkers to discriminate MASLD from MetALD and to quantify the strength of associations with risk of these phenotypes at the individual level. In addition, key lipid traits were further treated as exposures in genetic validation analyses to evaluate potential causal relationships with alcohol consumption using Mendelian randomisation. Second, in the pan-cancer multi-omics project, NMR metabolic profiles were incorporated as an intermediate molecular phenotype layer together with proteomics data within an integrative modelling framework. This approach was used to derive

latent factors capturing cross-omics shared variation and omics-specific variation, followed by downstream pathway enrichment and functional annotation. These analyses were intended to support molecular-level interpretation of the associations between regional body composition and cancer risk.

3.4. Proteomic Profiling

Proteomics data was obtained from the UKB Olink plasma proteomics programme. This programme used Olink's Proximity Extension Assay (PEA) technology to perform high-throughput protein quantification in 54,306 plasma samples, covering 2,923 protein assays. Protein measurements were provided as Normalized Protein eXpression (NPX) values. Using the encoding index (encoding 143), assay identifiers can be linked to UniProt annotations, thereby facilitating protein functional annotation and downstream pathway enrichment analyses. Before release, these data were evaluated and pre-filtered by the Pharma Proteomics Project consortium. UKB further harmonised a subset of validation assays that appeared across multiple panels to improve comparability. Within the analytical framework of this thesis, plasma proteomics were primarily used in the pan-cancer multi-omics integration project. Proteomics data were incorporated as a key intermediate molecular phenotype layer together with NMR metabolomics to derive latent factors capturing cross-omics shared variation and omics-specific variation. We then integrated protein annotation information to perform functional enrichment and pathway analyses, thereby supporting molecular-level interpretation of associations between regional body composition phenotypes and prioritised cancer risks.

3.5. Outcomes

Outcomes in this study comprised three categories: SLD-related phenotypic classification, CCA mortality, and pan-cancer incidence. For SLD phenotyping, hepatic steatosis was defined using an imaging threshold of PDFF $\geq$ 5%. Based on this criterion, MASLD and MetALD were classified according to the updated definition framework, integrating UKB clinical records and baseline alcohol intake information. CCA mortality was determined using the primary cause of death recorded in death registry data. Pan-cancer incidence outcomes were ascertained primarily from cancer registry records and encompassed cancers across multiple organ systems, including the respiratory, digestive, genitourinary, skin, and haematological systems. All outcomes were defined according to prespecified rules. Follow-up started at the baseline assessment date. Follow-up ended at the earliest of the first occurrence of the relevant

outcome, death, loss to follow-up, or the administrative censoring date. All cancer outcomes were defined using International Classification of Diseases, 10th Revision (ICD-10) codes (Table 1).

Table 1 ICD-10 Cancer Endpoint Definitions and Corresponding Codes

Cancer	ICD-10 codes
Esophagus	C15.0, C15.1, C15.2, C15.3, C15.4, C15.5, C15.8, C15.9
Stomach	C16.0, C16.1, C16.2, C16.3, C16.4, C16.5, C16.6, C16.8, C16.9
SmallIntestine	C17.0, C17.1, C17.2, C17.3, C17.8, C17.9
Colon	C18.2, C18.3, C18.4, C18.5, C18.6, C18.7, C18.8, C18.9
Rectum	C20
Anus	C21.0, C21.1, C21.8
HCC	C22.0
CCA_Biliary	C22.1, C23, C24.0, C24.1, C24.8, C24.9
Pancreas	C25.0, C25.1, C25.2, C25.3, C25.4, C25.7, C25.8, C25.9
Lung	C34.0, C34.1, C34.2, C34.3, C34.8, C34.9
Kidney	C64
Bladder	C67.0, C67.1, C67.2, C67.3, C67.4, C67.5, C67.6, C67.7, C67.8, C67.9
Prostate	C61
Testis	C62.0, C62.1, C62.9
Cervix	C53.0, C53.1, C53.8, C53.9
CorpusUteri	C54.0, C54.1, C54.2, C54.3, C54.8, C54.9, C55
Ovary	C56
Breast	C50.0, C50.1, C50.2, C50.3, C50.4, C50.5, C50.6, C50.8, C50.9
Brain	C71.0, C71.1, C71.2, C71.3, C71.4, C71.5, C71.6, C71.7, C71.8, C71.9
SpinalCNS	C72.0, C72.1, C72.2, C72.3, C72.4, C72.5, C72.8, C72.9
HeadNeck	C76.0, C07, C08.0, C08.1, C08.8, C08.9, C09.0, C09.1, C09.8, C09.9, C10.0, C10.2, C10.3, C10.4, C10.8, C10.9, C11.0, C11.1, C11.2, C11.3, C11.8, C11.9, C12, C13.0, C13.1, C13.2, C13.8, C13.9, C14.0, C14.8
OralCavity	C00.0, C00.1, C00.2, C00.3, C00.4, C00.5, C00.6, C00.9, C01, C02.0, C02.1, C02.2, C02.3, C02.4, C02.8, C02.9, C03.0, C03.1, C04.0, C04.1, C04.8, C04.9, C05.0, C05.1, C05.2, C05.8, C05.9, C06.0, C06.1, C06.2, C06.8, C06.9
Skin	C43.0, C43.1, C43.2, C43.3, C43.4, C43.5, C43.6, C43.7, C43.8, C43.9, C44.0, C44.1, C44.2, C44.3, C44.4, C44.5, C44.6, C44.7, C44.8, C44.9
Thyroid	C73
BoneJoint	C40.0, C40.1, C40.2, C40.3, C40.8, C40.9, C41.0, C41.1, C41.2, C41.3, C41.4, C41.9
ImmunoCancer	C88.0, C88.2, C88.3, C88.4, C88.7, C88.9
ALL	C91.0
CLL	C91.1
AML	C92.0
CML	C92.1
Hodgkin	C81.0, C81.1, C81.2, C81.3, C81.4, C81.7, C81.9
NonHodgkin	C82.0, C82.1, C82.2, C82.3, C82.4, C82.5, C82.6, C82.7, C82.9, C83.0, C83.1, C83.2, C83.3, C83.4, C83.5, C83.6, C83.7, C83.8, C83.9, C85.7, C85.9

Notes: HCC, hepatocellular carcinoma; CCA_Biliary, cholangiocarcinoma; CorpusUteri, cancer of the uterine corpus; SpinalCNS, spinal cord and other central nervous system cancers; ALL, acute lymphoblastic leukemia; CLL, chronic lymphocytic leukemia; AML, acute myeloid leukemia; CML, chronic myeloid leukemia; Hodgkin, Hodgkin lymphoma; NonHodgkin, non-Hodgkin lymphoma.

3.6. Covariates

All covariates were derived from standardised questionnaires and physical measurements collected at the UKB baseline assessment, covering key domains including sociodemographic characteristics, lifestyle factors (smoking and alcohol consumption), socioeconomic position, and medical history. Sociodemographic variables included age (years), sex (female/male), and ethnicity. Adiposity-related status was captured using baseline BMI (kg/m²), which was included as a continuous covariate. Indicators of socioeconomic position included educational attainment, total household income before tax, and the Townsend deprivation index, which reflects area-level socioeconomic deprivation. For pre-existing conditions, we constructed indicator variables for digestive system disease and circulatory system disease to capture major comorbidity burden and to reduce confounding related to comorbidities. For categorical covariates with missing values, an "Unknown" category was used in accordance with common UKB practice to retain sample size and to mitigate the risk of selection bias.

3.7. Analytical Approaches

Baseline characteristics were summarised and compared between groups using standard statistical methods. Continuous variables were presented as means with standard deviations and, where the normality assumption was met, between-group differences were assessed using independent-samples t tests. Categorical variables were summarised as counts and percentages and compared using chi-square tests. Before analysis, each regional BC measure and metabolite variable was standardised, resulting in a dimensionless scale that ensured comparability across markers while preserving relative differences. To control the type I error rate arising from multiple comparisons, we applied Bonferroni correction or the Benjamini–Hochberg (BH) procedure to control the false discovery rate (FDR), depending on the analytical objective. Unless otherwise specified, statistical significance was defined as a two-sided $P < 0.05$.

3.7.1. Logistic and Cox Regression Analyses

The logistic regression and Cox proportional hazards regression models were used to systematically evaluate exposure–outcome associations at the individual level, with the modelling framework selected according to outcome type. Effect estimates were reported using standard measures from the corresponding models, together with 95% confidence interval (CI) to reflect uncertainty. Unless otherwise specified, all models were adjusted for the prespecified covariates included in this study. For MASLD-related

analyses, differences in NMR lipidomic markers between MASLD and MetALD were evaluated using logistic regression models, and effect sizes were expressed as odds ratio (OR) per marker. Additionally, for PheWAS analyses, a logistic regression framework was applied to model each binary disease phenotype separately, estimating associations between regional BC measures and disease occurrence across the phenome. Multiple testing was addressed in the parallel analyses using appropriate error control procedures. Finally, Cox proportional hazards models were used to assess associations between regional BC measures and mortality risk from CCA and its subtypes, as well as associations with pan-cancer incidence. Results were reported as hazard ratios (HR). Potential non-linear relationships between regional BC measures and cancer outcomes were examined using restricted cubic splines (RCS).

3.7.2. Mendelian Randomization Analyses

Instrumental variables for the exposures were selected as single-nucleotide polymorphisms (SNPs) that reached genome-wide significance ($P < 5 \times 10^{-8}$). Linkage disequilibrium (LD) clumping was performed using an LD threshold of $r^2 < 0.01$ and a clumping window of 10,000 kb. Instrument strength was evaluated using the F statistics. Exposure and outcome GWAS datasets were harmonised to align effect alleles. SNPs that could not be aligned because of allele mismatches were excluded. In addition, SNPs that were significantly associated with potential confounders, such as sex, BMI, smoking, and alcohol consumption, were removed. We further applied the Steiger directionality test to assess causal direction and excluded SNPs that explained more variance in the outcome than in the exposure. This step was implemented to reduce bias due to reverse causation or horizontal pleiotropy. The overall procedure adhered to the three core Mendelian randomization assumptions, namely the relevance assumption, the independence assumption, and the exclusion restriction assumption.

The above MR framework was applied to evaluate potential causal relationships between alcohol-related exposures and differential metabolites, as well as between regional BC traits and pan-cancer incidence risk. For causal effect estimation, the primary analysis used the inverse-variance weighted (IVW) method, and MR-Egger, the weighted median, the simple mode, and the weighted mode methods were implemented as sensitivity analyses to examine robustness and potential pleiotropic effects. Cochran's Q test and the MR-Egger regression intercept were used to assess heterogeneity and horizontal pleiotropy, with evidence considered present at $P < 0.05$.

In addition, we conducted MR pleiotropy residual sum and outlier (MR-PRESSO) analysis to identify outliers using a threshold of $P < 0.05$ [155]. When outliers were detected, they were removed and MR causal estimates were re-evaluated.

3.7.3. Machine Learning Models

During prognostic model development, we first applied recursive feature elimination (RFE) to iteratively select an optimal feature subset, thereby evaluating the information gain and relative importance of each regional BC variable and determining the most informative predictors [156]. To address class imbalance in CCA outcomes during model development and optimisation, we used the synthetic minority over-sampling technique (SMOTE) [157]. The dataset was then randomly split into training and testing sets at a 7:3 ratio. Based on the selected features, we developed and compared ML models, including support vector machine (SVM), gradient boosting machine (GBM), eXtreme Gradient Boosting (XGB), neural network (NN), and multilayer perceptron (MLP). Model training used 10-fold cross-validation for internal validation and hyperparameter tuning. For SVM, we tuned the kernel function and penalty parameter, with emphasis on optimising parameters for the radial basis function (RBF) kernel. For GBM and XGB, key hyperparameters included the number and depth of trees, the learning rate (shrinkage), and regularisation-related parameters. For NN and MLP, we adjusted network architecture, activation functions, learning rates, and regularisation strategies to reduce overfitting and improve generalisation performance. Model performance was primarily evaluated using the area under the receiver operating characteristic curve (AUC). To enhance interpretability, after iterative optimisation of the XGB model, we applied the SHAP method for post hoc interpretation. SHAP, based on the Shapley value framework, quantifies the marginal contribution of each feature to model output, thereby characterising both the direction and magnitude of feature effects on predictions.

3.7.4. Multi-Omics Factor Analysis

MOFA was used for multi-omics integration in the pan-cancer project. MOFA is an unsupervised multi-view latent variable model that learns a set of low-dimensional latent factors across multiple omics views. By doing so, it decomposes and characterises both cross-omics shared variation and view-specific variation components, thereby reducing noise and redundancy in high-dimensional omics data while preserving interpretability [158]. The model was implemented using the default model specification

and training options in the MOFA2 package. Under a variational inference framework, model parameters were iteratively updated by optimising the evidence lower bound (ELBO), and training was terminated when the maximum number of iterations was reached or when a convergence criterion based on incremental ELBO improvement was met [159]. To determine the number of latent factors (K), we trained candidate models with K = 10, 15, 20, and 30 and calculated the overall variance explained reported by MOFA2 for each model. We then plotted the change in overall variance explained across K values and applied the elbow method to select the K value corresponding to the point at which gains in overall variance explained began to diminish substantially, thereby balancing model complexity and explanatory performance. After finalizing the model, we obtained participant-level factor scores, the variance explained by each factor within each omics view, and feature loadings for each omics layer for downstream analyses.

3.7.5. Functional and Pathway Enrichment Analyses

To elucidate the biological processes represented by MOFA latent factors and to enhance molecular-level interpretability, we performed functional and pathway enrichment analyses for metabolite and protein features most strongly associated with each latent factor. Specifically, within each omics view, features were ranked according to the magnitude of their factor loadings, and the top 10 metabolites and top 10 proteins with the largest absolute loading values were selected as representative features. Functional and pathway enrichment analyses were then conducted using the MetaboAnalyst 6.0 online platform and Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) resources [160]. Multiple testing was controlled using the BH procedure.

4. Results

4.1. Discriminative Metabolomics Signatures for MASLD and MetALD

4.1.1. Baseline Characteristics of MASLD and MetALD patients

Of the SLD-diagnosed participants with metabolomics and lipidomics data, 5426 MASLD participants and 644 MetALD participants were first analysed to see if their lipidomic profiles at baseline could distinguish MASLD from MetALD. The first lipidomic measurements as well as the standard serum parameters were taken around 6 years prior to the liver MRI. As we believe that the liver disease is slowly progressing and therefore quite stable over time, we first investigated differences in baseline serum profiles between MASLD and MetALD patients (Table 2). For liver enzymes, MetALD patients exhibited higher levels of serum alanine aminotransferases (ALT), aspartate aminotransferases (AST) and gamma-glutamyl transferase (GGT) compared to the MASLD cohort ($P < 0.05$, Table 2). A similar trend was observed for Apolipoprotein A and HDL cholesterol, whereby MetALD participants had significantly higher levels compared to MASLD (HDL: 1.46 ± 0.36mmol/L vs. 1.28 ± 0.3mmol/L, $P < 0.001$, Table 2).

Table 2: Summary statistics of patients with MetALD and MASLD

Section	Metabolite (Unit)	MetALD (Mean ± SD) n=644	MASLD (Mean ± SD) n=5426	Univariate *P*
Demographics at time of MRI				
	AGE (years)	54.15 ± 7.35	55.47 ± 7.31	0.001
	BMI (kg/m²)	28.70 ± 3.92	29.20 ± 4.16	0.009
	Liver lipid content on MRI (%)	11.81 ± 7	10.48 ±5.81	8E-05
Liver Tests at baseline				
	Alanine aminotransferase (U/L)	31.19 ± 17.13	29.37 ± 18.01	0.033
	Albumin (g/L)	45.81 ± 2.45	45.52 ± 2.52	0.022
	Alkaline phosphatase (U/L)	79.66 ± 22.74	84.06 ± 24.86	0.0001
	Aspartate aminotransferase (U/L)	29.23 ± 10.90	27.75 ± 12.23	0.007
	Direct bilirubin (µmol/L)	1.94 ± 0.74	1.85 ± 0.83	0.018
	Total bilirubin (µmol/L)	9.79 ± 4.04	9.42 ± 4.82	0.07
	Gamma glutamyltransferase (U/L)	57.73 ± 63.38	41.97 ± 36.13	3E-07
Lipid Panel at baseline				
	Apolipoprotein A (g/L)	1.58 ± 0.26	1.44 ± 0.23	8E-21
	Apolipoprotein B (g/L)	1.09 ± 0.23	1.08 ± 0.24	0.34
	Cholesterol (mmol/L)	5.93 ± 1.04	5.75 ± 1.13	5E-04
	HDL cholesterol (mmol/L)	1.46 ± 0.36	1.28 ± 0.3	4E-19
	LDL direct (mmol/L)	3.71 ± 0.81	3.68 ± 0.85	0.39
	Lipoprotein A (mg/dL)	43.16 ± 49.21	48.67 ± 48.67	0.92
	Triglycerides (mmol/L)	2.04 ± 1.14	2.15 ± 1.09	0.043
General Biochemistry at baseline				
	Calcium (mmol/L)	2.38 ± 0.09	2.38 ± 0.09	0.48
	Creatinine (µmol/L)	74.54 ± 13.33	74.96 ± 14.07	0.52

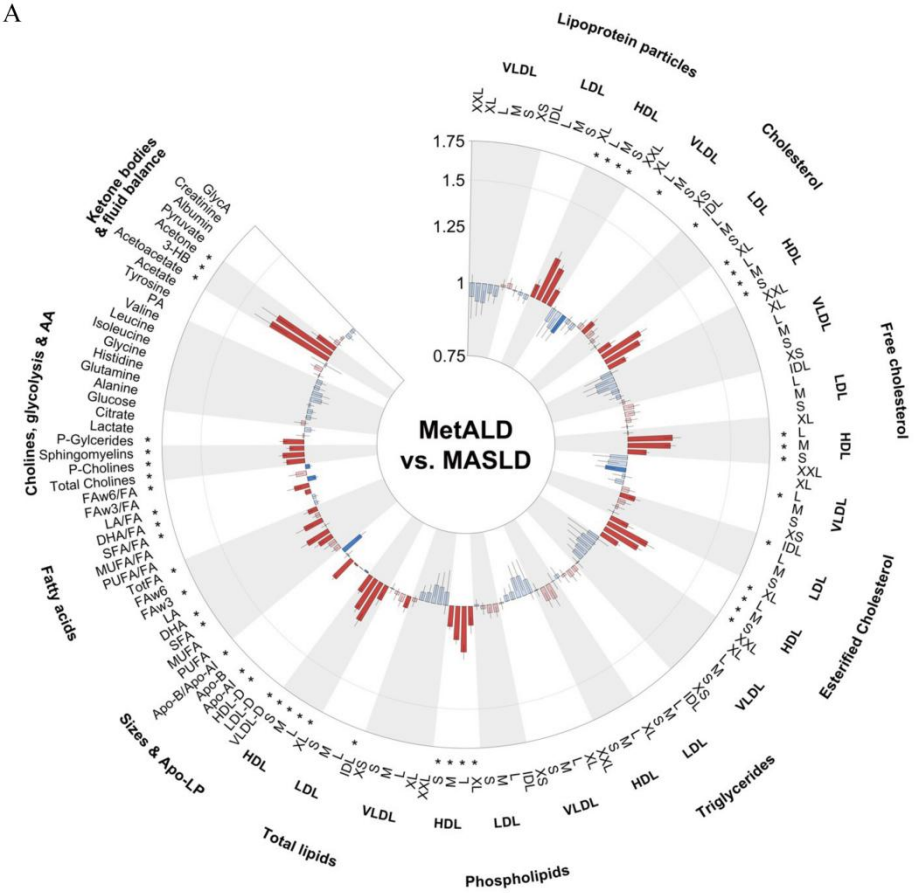
Cystatin C (mg/L)	0.88 ± 0.12	0.91 ± 0.13	2E-06
Glucose (mmol/L)	5.16 ± 1.01	5.14 ± 1.18	0.62
HbA1c (mmol/mol)	35.01 ± 5.27	36.24 ± 6.30	5E-06
IGF-1 (ng/mL)	20.71± 4.89	21.64 ± 5.51	2E-04
Phosphate (mmol/L)	1.13 ± 0.17	1.14 ± 0.17	0.09
Urate (µmol/L)	356.78 ± 81.94	339.50 ± 75.02	2E-05
Urea (mmol/L)	5.28 ± 1.21	5.42 ± 1.23	0.02
Vitamin D (nmol/L)	48.59 ± 22.48	46.39 ± 19.61	0.04
C-reactive protein (mg/L)	2.33 ± 2.89	2.77 ± 3.73	0.003
Oestradiol (pmol/L)	387.03 ± 239.28	409.89 ± 329.26	0.44
Rheumatoid factor (IU/mL)	21.11 ± 14.35	22.95 ± 17.96	0.45
SHBG (nmol/L)	38.94 ± 18.66	38.37 ± 19.16	0.55
Testosterone (nmol/L)	8.38 ± 5.71	7.66 ± 5.56	0.012

Comparisons were made between MASLD (allowing an alcohol consumption below 140 g/week for females and 210 g/week for males) and MetALD (MASLD + alcohol consumption between 140 to 350 g/week for females and 210 to 420 g/week for males). Continuous variables were expressed as means and standard deviation, while categorical variables were expressed as number and frequency (%). Abbreviations: MASLD, Metabolic dysfunction-associated steatotic liver disease; MetALD, Metabolic dysfunction and alcohol-related liver disease; LDL, low density lipoprotein; HDL, high density lipoprotein; IGF-1, Insulin growth factor-1; SHBG, Sex-hormone binding globulin.

4.1.2. Discriminative Metabolomics Markers Between MASLD and MetALD

The unbiased analysis included 249 lipidomic and metabolomic parameters, which were evaluated for their potential to discriminate between MASLD and MetALD. After adjusting for age, sex and BMI, we identified a significant difference in the lipidomic profiles of MASLD and MetALD. There are 107 differentially expressed lipidomic markers when comparing MetALD to MASLD patients (Figure 2A). Of lipid parameters, large-sized VLDL lipoprotein was significantly lower in MetALD participants. Consistently, total HDL lipids were significantly higher, with the largest effect ize distinguishing MetALD and MASLD among lipidomic parameters (Table 3). Additionally, Lipidomic profiling revealed 32 metabolites that were differentially expressed in ALD participants compared to MASLD (Figure 2B). Of these, large-, medium- and small-sized particles of HDL as well as HDL cholesterol and phospholipids were significantly higher in the ALD cohort. Apolipoprotein A1, saturated fatty acids and total choline were also overrepresented in ALD participants.

A


MetALD
vs. MASLD
Lipoprotein particles
Cholesterol
Free cholesterol
Esterified Cholesterol
Triglycerides
Phospholipids
Total lipids
Sizes & Apo-LP
Fatty acids
Cholines, glycolysis & AA
Ketone bodies & fluid balance
1.75
1.5
1.25
1
0.75

B

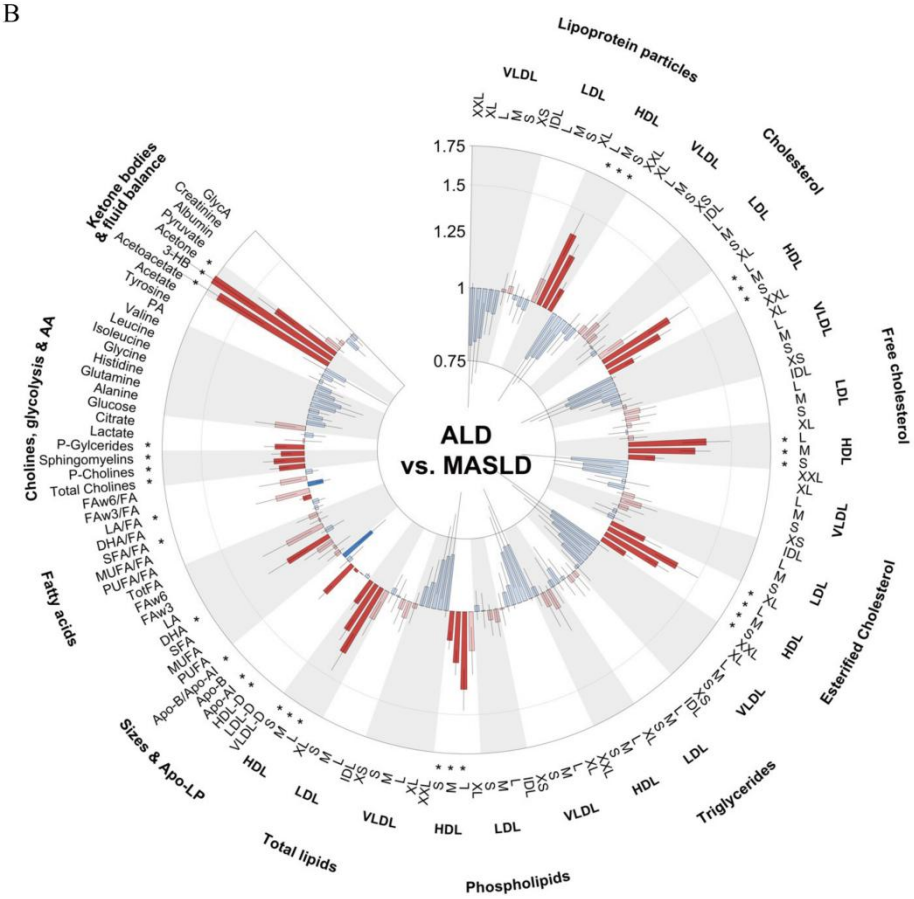

ALD
vs. MASLD
Lipoprotein particles
Cholesterol
Free cholesterol
Esterified Cholesterol
Triglycerides
Phospholipids
Total lipids
Sizes & Apo-LP
Fatty acids
Cholines, glycolysis & AA
Ketone bodies & fluid balance
1.75
1.5
1.25
1
0.75

Figure 2 Circle plots for lipidomic analysis. (A) Circle plot for lipidomic analysis for MetALD versus MASLD UKB participants. Red=Higher in MetALD, blue=higher in MASLD. (B) Circle plot for lipidomic analysis for ALD versus MASLD UKB participants. Red=Higher in ALD, blue=higher in MASLD. Baseline Lipidomic parameters were measured through NMR spectroscopy. Hazard ratios (with 95% confidence intervals) are presented per 1-SD higher metabolic biomarker on the natural log scale, stratified by age, sex, body mass index and Townsend deprivation index. *Original code by Diego J. Aguilar-Ramirez. Apo-A1, apolipoprotein A1, Apo-B, apolipoprotein B, FA, fatty acids, FAw3, omega-3 fatty acids, FAw6, omega-6 fatty acids, LA, linoleic acid, MUFA, monounsaturated fatty acids, PUFA, polyunsaturated fatty acids, SFA, saturated fatty acids, TG, triglycerides, 3-HB, 3-Hydroxybutyrate, DHA, docosahexaenoic acid.

The top metabolites with the highest discriminatory power were identified (Table 3). These included: Acetoacetate (OR: 1.30, 95% CI: 1.21–1.41; *P* = 1.48E-11), 3-Hydroxybutyrate (OR:1.29, 95% CI: 1.20–1.39; *P* = 1.72E-11), Phospholipids in Large HDL (OR: 1.20, 95% CI: 1.16–1.23; *P* = 1.85E-32), Concentration of large HDL particles, free cholesterol in large HDL, cholesterol in large HDL, cholesteryl esters in large HDL, Total lipids in large HDL, free cholesterol in medium HDL and cholesterol in medium HDL. All of these were significantly higher in the MetALD group compared to the MASLD group. Strikingly, the metabolites with the greatest effect sizes to differentiate MASLD and MetALD are associated with large-sized HDL particles such as total lipids, cholesterol and phospholipids in large HDL.

Table 3: Top 30 discriminatory metabolites between MetALD and MASLD.

Metabolite	OR	95%CI	P-Value
Acetoacetate	1.30	1.21-1.41	1.48E-11
3-Hydroxybutyrate	1.29	1.20-1.39	1.72E-11
Phospholipids in Large HDL	1.20	1.16-1.23	1.85E-32
Concentration of Large HDL Particles	1.20	1.16-1.24	6.76E-26
Free Cholesterol in Large HDL	1.19	1.15-1.23	1.86E-24
Cholesterol in Large HDL	1.19	1.15-1.23	7.54E-24
Cholesteryl Esters in Large HDL	1.19	1.15-1.23	2.38E-23
Total Lipids in Large HDL	1.19	1.15-1.22	1.77E-28
Free Cholesterol in Medium HDL	1.18	1.16-1.21	3.08E-50
Cholesterol in Medium HDL	1.17	1.14-1.19	3.08E-52
Cholesteryl Esters in Medium HDL	1.16	1.14-1.18	7.23E-52
Concentration of Medium HDL Particles	1.16	1.14-1.18	9.25E-51
Total Lipids in Medium HDL	1.15	1.13-1.17	6.12E-51
Phospholipids in Medium HDL	1.14	1.12-1.16	3.88E-49
Apolipoprotein A1	1.10	1.09-1.12	3.82E-49
Phosphatidylcholines	1.09	1.07-1.10	5.53E-27
Phospholipids in Small HDL	1.09	1.07-1.10	1.86E-36
Phosphoglycerides	1.08	1.07-1.10	3.09E-25
Total Lipids in Small HDL	1.08	1.06-1.09	4.9E-33

Cholesteryl Esters in Small HDL	1.08	1.06-1.09	4.16E-29
Cholesterol in Small HDL	1.08	1.06-1.09	5.74E-31
Free Cholesterol in Small HDL	1.08	1.06-1.09	7.56E-28
Total Cholines	1.07	1.06-1.09	5.93E-24
Concentration of Small HDL Particles	1.07	1.06-1.08	3.22E-28
Sphingomyelins	1.06	1.04-1.07	1.02E-15
Saturated Fatty Acids to Total Fatty Acids percentage	1.02	1.02-1.03	8.09E-15
Average Diameter for HDL Particles	1.01	1.00-1.01	6.15E-16
Linoleic Acid to Total Fatty Acids percentage	0.97	0.96-0.98	3.91E-10
Apolipoprotein B to Apolipoprotein A1 ratio	0.91	0.89-0.94	2.62E-12

Note: Differential expression of baseline lipidomic markers between MASLD and MetALD was determined using logistic regression models corrected for age, sex and BMI. Metabolites with the highest discriminatory power between the two conditions were identified based on their OR, p-values, and 95% confidence intervals. A rank list of top 30 metabolites with lowest p-values was generated for further investigation and sorted by OR. Abbreviations: MASLD, Metabolic dysfunction-associated steatotic liver disease; MetALD, Metabolic dysfunction-associated alcohol-related liver disease; HDL, high density lipoprotein.

Similar to the top discriminatory metabolites identified in MetALD, participants with ALD also exhibited the highest discriminatory power with phospholipids, total lipids, concentration and cholesterol in large HDL, followed by apolipoprotein A1 (Table 4). Intriguingly, the effect size was highest for 3-hydroxybutyrate (OR= 1.76, 95% CI: 1.44–2.15; *P* = 4.01E-08) and acetoacetate (OR= 1.66, 95% CI: 1.35–2.06; *P* = 2.2E-06). For both ALD and MetALD, large HDL components (Cholesterol, Cholesteryl Esters, Phospholipids and Total Lipids) show elevated levels. However, ALD shows even higher ORs (ranging from 1.34 to 1.36) compared to MetALD (ranging from 1.19 to 1.20, Table 4). The significance is profound in both groups, but the effect sizes in ALD are higher. Regarding Apolipoprotein A1, ALD shows a much higher odds ratio (OR= 1.16, 95% CI: 1.12–1.21; *P* = 2.79E-16, Table 4) compared to MetALD (OR= 1.10, 95% CI: 1.09–1.12; *P* = 3.82E-49, Table 3), again suggesting that apolipoprotein A1 alterations are more pronounced in ALD.

Table 4: Top 30 discriminatory metabolites between ALD and MASLD.

Metabolite	**OR**	**95%CI**	**P-Value**
3-Hydroxybutyrate	1.76	1.44- 2.15	4.01E-08
Acetoacetate	1.66	1.35- 2.06	2.2E-06
Concentration of Large HDL Particles	1.36	1.24- 1.49	7.89E-11
Cholesteryl Esters in Large HDL	1.36	1.24- 1.50	3.84E-10
Cholesterol in Large HDL	1.36	1.24- 1.49	2.55E-10
Free Cholesterol in Large HDL	1.36	1.24- 1.49	1.85E-10
Phospholipids in Large HDL	1.36	1.25- 1.47	4.47E-13
Total Lipids in Large HDL	1.34	1.23- 1.46	8.87E-12

Acetone	1.33	1.22- 1.46	1.33E-10
Free Cholesterol in Medium HDL	1.30	1.22- 1.38	5.38E-17
Cholesterol in Medium HDL	1.27	1.21- 1.35	4.03E-18
Cholesteryl Esters in Medium HDL	1.27	1.20- 1.34	4.22E-18
Concentration of Medium HDL Particles	1.26	1.19- 1.33	5.32E-17
Total Lipids in Medium HDL	1.24	1.18- 1.30	6.42E-17
Phospholipids in Medium HDL	1.22	1.17- 1.28	5.83E-16
Docosahexaenoic Acid	1.21	1.10- 1.33	4.99E-05
Apolipoprotein A1	1.16	1.12- 1.21	2.79E-16
Phosphatidylcholines	1.13	1.08- 1.18	5.08E-08
Phospholipids in Small HDL	1.12	1.08- 1.16	2.14E-10
Phosphoglycerides	1.12	1.07- 1.17	2.01E-07
Cholesteryl Esters in Small HDL	1.11	1.07- 1.16	9.3E-09
Cholesterol in Small HDL	1.11	1.07- 1.15	3.34E-09
Total Lipids in Small HDL	1.11	1.07- 1.15	2.39E-09
Free Cholesterol in Small HDL	1.11	1.07- 1.15	3.49E-08
Total Cholines	1.11	1.06- 1.15	3.72E-07
Concentration of Small HDL Particles	1.10	1.06- 1.14	4.22E-08
Sphingomyelins	1.10	1.05- 1.14	4.65E-06
Average Diameter for HDL Particles	1.01	1.1- 1.01	2.42E-07
Linoleic Acid to Total Fatty Acids percentage	0.94	0.91- 0.97	5.01E-05
Apolipoprotein B to Apolipoprotein A1 ratio	0.86	0.80- 0.93	7.24E-05

Note: Differential expression of baseline lipidomic markers between MASLD and MetALD was determined using logistic regression models corrected for age, sex and BMI. Metabolites with the highest discriminatory power between the two conditions were identified based on their OR, p-values and 95% confidence intervals. A rank list of the top 30 metabolites with the lowest p-values was generated for further investigation and sorted by OR. Abbreviations: HDL, high-density lipoprotein; MASLD, metabolic dysfunction-associated steatotic liver disease; MetALD, metabolic dysfunction and alcohol-related
liver disease.

4.1.3. Genetic Validation of Key Discriminative Metabolites

To validate the association of HDL with alcohol-related liver phenotypes, we used MR analysis. In the MR analysis, genetically predicted alcohol consumption was associated with increased levels of HDL-related metabolites (Figure 3). These included Acetoacetate levels (OR_{IVW}: 1.23, 95% CI: 1.09-1.4; $P = 0.001$), Concentration of Large HDL Particles (OR_{IVW}: 1.2, 95% CI: 1.06–1.36; $P = 0.005$), Total Lipids in Medium HDL (OR_{IVW}: 1.69, 95% CI: 1.52-1.87; $P < 0.001$) and Free Cholesterol in medium HDL (OR_{IVW}: 1.56, 95% CI: 1.39-1.75; $P < 0.001$).

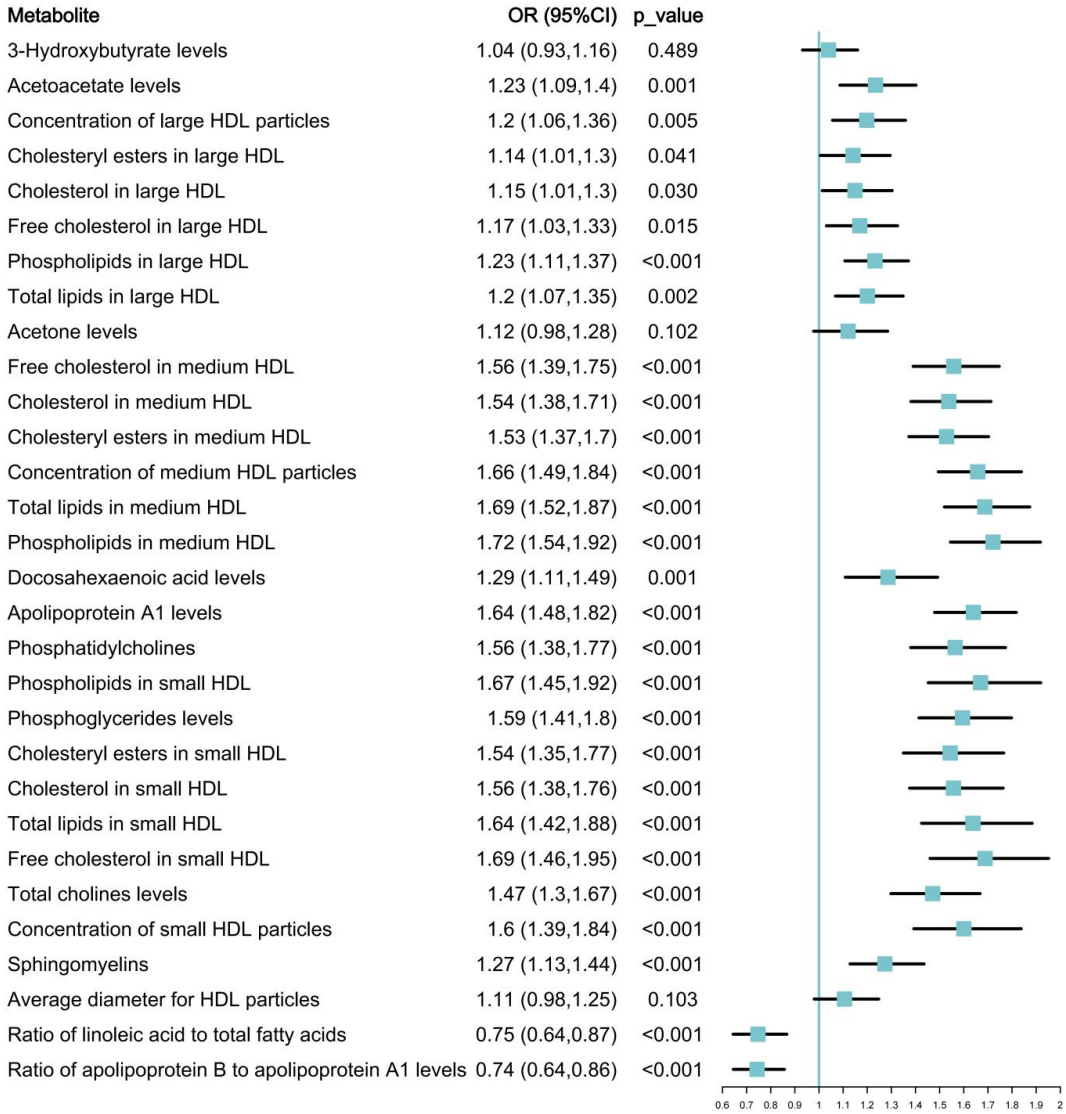


Figure 3 Forest Plot of Mendelian randomisation analysis between alcohol consumption and the top 30 discriminatory metabolites using Inverse variance weighting. OR: odds ratio; 95% CI: 95% confidence interval.

4.2. Associations and Prognostic Models Between regional BC and CCA Mortality

4.2.1. Baseline Characteristics of the CCA Cohort

A total of 423 participants were registered as dying from CCA during the follow-up. Compared to the control group, the CCA mortality group had a higher mean age (60.95 vs 56.50 years, $P < 0.001$) and BMI (28.20 kg/m^2 vs 27.43 kg/m^2, $P = 0.001$). No significant differences were found in sex or ethnicity distribution between the two groups. However, the proportion of smokers was higher, and the frequency of drinking alcohol was lower in the mortality group than in the control group ($P < 0.05$).

Additionally, the mortality group had a lower level of education and annual income and were more likely to have circulatory and digestive diseases ($P < 0.05$) (Table 5).

Table 5 Baseline characteristics of participants CCA cohort.

Characteristics	Control (n=485312)	Mortality (n=423)	*P*
Age	56.50 ± 8.08	60.95 ± 6.31	**<0.001**
Sex			1.000
Female	265111 (54.6%)	231 (54.6%)	
Male	220201 (45.4%)	192 (45.4%)	
Ethnicity			0.264
White	458043 (94.4%)	409 (96.7%)	
Mixed	2841 (0.6%)	1 (0.2%)	
Asian or Asian British	10861 (2.2%)	5 (1.2%)	
Black or Black British	7632 (1.6%)	3 (0.7%)	
Others	5935 (1.2%)	5 (1.2%)	
Body mass index (BMI)	27.43 ± 4.76	28.20 ± 5.06	**0.001**
Townsend deprivation index	-1.32 ± 3.07	-1.36 ± 3.18	0.792
Smoking status			**0.001**
Never	266411 (54.9%)	195 (46.1%)	
Previous	168047 (34.6%)	181 (42.8%)	
Current	50854 (10.5%)	47 (11.1%)	
Alcohol frequency			**0.005**
Regularly	211278 (43.5%)	176 (41.6%)	
Occasionally	179739 (37.0%)	139 (32.9%)	
Rarely	94295 (19.4%)	108 (25.5%)	
Education			**<0.001**
< 10 years	81220 (16.7%)	116 (27.4%)	
10-14 years	182332 (37.6%)	130 (30.7%)	
≥15 years	213226 (43.9%)	168 (39.7%)	
Unknown	8534 (1.8%)	9 (2.1%)	
Annual household income before tax, £			**<0.001**
<18,000	72582 (15.0%)	66 (15.6%)	
18,000-30,999	92808 (19.1%)	115 (27.2%)	
31 000-51,999	105032 (21.6%)	116 (27.4%)	
52,000-100,000	108119 (22.3%)	67 (15.8%)	
>100,000	84411 (17.4%)	46 (10.9%)	
Unknown	22360 (4.6%)	13 (3.1%)	
Circulation system disease			**<0.001**
No	319140 (65.8%)	235 (55.6%)	
Yes	166172 (34.2%)	188 (44.4%)	
Digestive system disease			**0.020**
No	398003 (82.0%)	328 (77.5%)	
Yes	87309 (18.0%)	95 (22.5%)	

4.2.2. Associations Between Body Composition and CCA Mortality

In all Cox regression models, FM in limbs, trunk and whole body were positively associated with CCA mortality. After full adjustment for confounders, the HRs for left and right arm FM were 1.32 (95% CI: 1.07-1.62, $P = 0.008$) and 1.28 (95% CI: 1.06-1.56, $P = 0.012$), respectively. Whole body FM (HR= 1.62, 95% CI: 1.25-2.11, $P < 0.001$) and leg FM (HR-left= 1.62, $P = 0.001$, HR-right= 1.44, $P = 0.006$) were also significantly associated with increased CCA mortality (Table 6). Additionally, whole body (HR= 1.22, 95%CI: 1.02-1.44, $P = 0.018$) and left arm impedance (HR= 1.2, 95%CI: 1.02-1.40, $P = 0.024$) emerged as risk factors for CCA mortality after full adjustment. Unexpectedly, leg FFM (HR-left= 1.29, HR-right = 1.27, $P < 0.05$) was also positively associated with an increased risk of CCA mortality. In contrast, FFM in other regions showed no significant associations (Table 6).

To further examine the association between BC and specific CCA subtypes, ICC and ECC were analyzed separately. The results demonstrated that increased FM in arm (HR-left = 1.32, HR-right= 1.29; $P < 0.05$) and leg (HR-left= 1.75, HR-right= 1.54; $P < 0.05$) were associated with a higher risk of ICC mortality. In addition, higher leg impedance (HR-left = 1.19; HR-right= 1.18; $P < 0.05$) emerged as a risk factor for adverse outcomes (Table 6). For ECC, a one standard deviation (SD) increases in trunk FM and FP, and whole-body FFM increased ECC mortality by 120% (HR= 2.2, 95% CI: 1.29-3.75, $P = 0.004$), 82% (HR= 1.82, 95% CI: 1.11-2.99, $P = 0.017$), and 90% (HR= 1.9, 95% CI: 1.01-3.54, $P = 0.045$), respectively (Table 6). However, FM in the limbs were not associated with ECC mortality, whereas leg FFM (HR-left= 2.33, HR-right= 2.24, $P < 0.005$) exhibited a greater effect on ECC mortality compared to other risk factor (Table 6).

Table 6 The association between regional BC and the mortality of CCA.

Variables	CCA		ICC		ECC	
	HR (95%CI)	*P*	HR (95%CI)	*P*	HR (95%CI)	*P*
Arm fat percentage (left)	1.26 (0.93, 1.7)	0.129	1.17 (0.85, 1.61)	0.336	2.08 (0.94, 4.6)	0.072
Arm fat percentage (right)	1.2 (0.9, 1.6)	0.223	1.13 (0.83, 1.55)	0.434	1.77 (0.8, 3.92)	0.157
Arm fat mass (left)	1.32 (1.07, 1.62)	0.008	1.32 (1.05, 1.66)	0.016	1.3 (0.8, 2.12)	0.283
Arm fat mass (right)	1.28 (1.06, 1.56)	0.012	1.29 (1.04, 1.59)	0.018	1.24 (0.76, 2.05)	0.39
Arm fat-free mass (right)	1.18 (0.94, 1.5)	0.159	1.14 (0.88, 1.46)	0.322	1.53 (0.82, 2.89)	0.185
Arm fat-free mass (left)	1.18 (0.93, 1.48)	0.165	1.14 (0.89,	0.305	1.46 (0.78, 2.71)	0.236

			1.46)			
Arm predicted mass (left)	1.2 (0.95, 1.51)	0.131	1.16 (0.9, 1.49)	0.245	1.45 (0.78, 2.73)	0.243
Arm predicted mass (right)	1.21 (0.95, 1.53)	0.122	1.16 (0.9, 1.5)	0.251	1.53 (0.81, 2.9)	0.192
Trunk fat percentage	1.17 (0.99, 1.38)	0.07	1.1 (0.92, 1.32)	0.295	1.82 (1.11, 2.99)	0.017
Trunk fat mass	1.35 (1.11, 1.63)	0.002	1.25 (1.02, 1.54)	0.03	2.2 (1.29, 3.75)	0.004
Trunk fat-free mass	1.16 (0.94, 1.43)	0.173	1.12 (0.89, 1.4)	0.343	1.48 (0.83, 2.63)	0.182
Trunk predicted mass	1.16 (0.94, 1.44)	0.172	1.12 (0.89, 1.41)	0.338	1.48 (0.83, 2.65)	0.189
Basal metabolic rate	1.29 (1.03, 1.62)	0.028	1.2 (0.94, 1.54)	0.136	1.97 (1.08, 3.56)	0.026
Body fat percentage	1.24 (0.97, 1.58)	0.081	1.19 (0.92, 1.55)	0.183	1.65 (0.82, 3.33)	0.159
Whole body fat mass	1.62 (1.25, 2.11)	<0.001	1.55 (1.17, 2.05)	0.002	2.22 (1.07, 4.61)	0.032
Whole body fat-free mass	1.24 (0.98, 1.56)	0.075	1.16 (0.9, 1.49)	0.257	1.9 (1.01, 3.54)	0.045
Whole body water mass	1.23 (0.97, 1.55)	0.086	1.14 (0.89, 1.47)	0.296	1.92 (1.03, 3.58)	0.04
Leg fat percentage (left)	1.27 (0.88, 1.82)	0.196	1.48 (1.01, 2.18)	0.045	0.45 (0.17, 1.17)	0.1
Leg fat percentage (right)	1.18 (0.85, 1.63)	0.331	1.35 (0.95, 1.91)	0.093	0.47 (0.2, 1.13)	0.092
Leg fat mass (left)	1.62 (1.23, 2.13)	0.001	1.75 (1.32, 2.33)	<0.001	0.82 (0.33, 2.05)	0.665
Leg fat mass (right)	1.44 (1.11, 1.86)	0.006	1.54 (1.19, 1.99)	0.001	0.78 (0.32, 1.9)	0.581
Leg fat-free mass (left)	1.29 (1.03, 1.63)	0.027	1.17 (0.91, 1.5)	0.223	2.33 (1.36, 3.98)	0.002
Leg fat-free mass (right)	1.27 (1.02, 1.58)	0.03	1.15 (0.9, 1.45)	0.26	2.24 (1.38, 3.65)	0.001
Leg predicted mass (left)	1.3 (1.03, 1.64)	0.027	1.17 (0.91, 1.51)	0.223	2.36 (1.37, 4.06)	0.002
Leg predicted mass (right)	1.27 (1.02, 1.58)	0.035	1.14 (0.9, 1.46)	0.277	2.24 (1.37, 3.67)	0.001
Impedance of whole body	1.22 (1.04, 1.44)	0.018	1.23 (1.04, 1.47)	0.018	1.11 (0.69, 1.78)	0.659
Impedance of arm (left)	1.2 (1.02, 1.4)	0.024	1.17 (0.99, 1.39)	0.066	1.37 (0.93, 2.02)	0.112
Impedance of arm (right)	1.15 (0.99, 1.33)	0.066	1.13 (0.96, 1.32)	0.134	1.26 (0.88, 1.8)	0.205
Impedance of leg (left)	1.12 (0.98, 1.27)	0.087	1.19 (1.04, 1.36)	0.013	0.71 (0.48, 1.05)	0.083
Impedance of leg (right)	1.11 (0.98, 1.26)	0.109	1.18 (1.03, 1.35)	0.015	0.69 (0.47, 1.02)	0.06

Note: CCA: Cholangiocarcinoma; ICC: intrahepatic cholangiocarcinoma; ECC: extrahepatic cholangiocarcinoma; Model: adjusted for ethnicity, age, sex, BMI, education, annual household income before tax, smoking status, alcohol frequency, circulatory and digestive system disease.

4.2.3. Performance and Comparative Evaluation of Prediction Models

All BC variables were included as candidate predictors and ranked by importance using various algorithms. The intersecting variables from the top 20 variables were then

selected as initial predictors. However, further analysis revealed that the predictive performance of the model was significantly improved by adding age and BMI variables. Consequently, age, BMI and the 10 most important intersecting variables (left arm and trunk FM, left and right arm FFM, left and right leg FP, impedance of left and right arm, left leg and whole-body) were selected as the final predictors of the prediction model.

Five exploratory ML models were developed and internally validated to examine whether regional BC variables could discriminate mortality risk for CCA, ICC, and ECC. Among the internally validated models, XGB showed the highest discriminatory performance for CCA and ICC mortality, with AUCs of 0.913 and 0.851 for CCA and 0.893 and 0.805 for ICC in the training and test cohorts, respectively (Figure 4A-B). The GBM and MLP yielded comparable but slightly lower AUCs, whereas SVM and NN also showed moderate discrimination in the test cohort. For ECC mortality, the NN model yielded the highest AUC within the present internal validation framework (0.904 in the training cohort and 0.807 in the test cohort), although this result should be interpreted cautiously given the exploration nature of the analysis (Figure 4C).

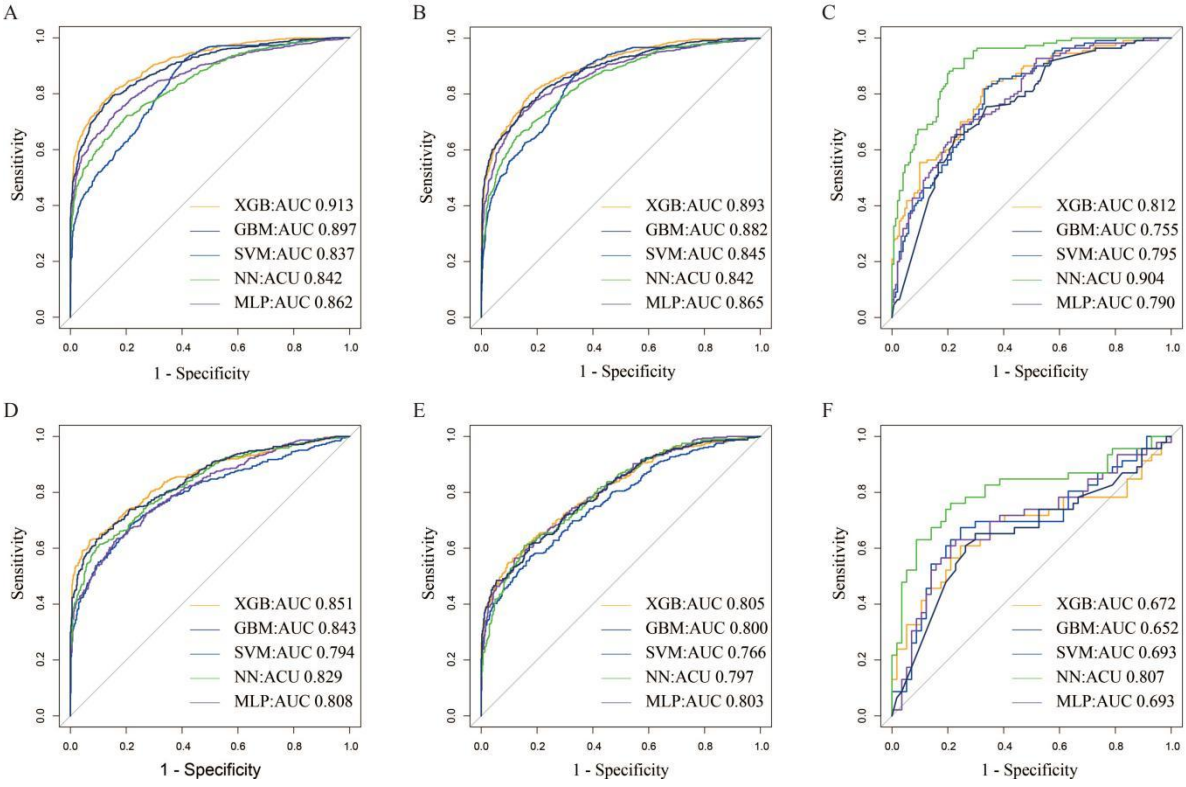


Figure 4. Prognostic performance of prediction models for cholangiocarcinoma (CCA) and its subtypes. (A-C) Training cohort for overall CCA, ICC and ECC model; (D-F) Test cohort for CCA, ICC and ECC model. Note: AUC: Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve; XGB: eXtreme Gradient Boosting; GBM: Gradient Boosting Machine; SVM: Support Vector Machine; NN: Neural Network; MLP: Multilayer Perceptron.

4.2.4. Shapley-Based Interpretation of Model Contributions

Machine learning models are often considered "black box" models, making predictive results interpretation challenging. To enhance interpretability, we applied the SHAP method based on the XGB model to quantify the contribution of each variable to predicting CCA and its subtypes' mortality, providing both global and individual-level explanations. At the global level, mean SHAP values were used to evaluate the magnitude and direction of the contribution of the predictors, ranking them by the degree of contribution. The results indicated that age showed the largest contribution to prediction across overall CCA and for both subtypes, ICC and ECC. Among regional BC variables, the contribution patterns differed by outcome: left upper-limb FM ranked highest for CCA and ICC, whereas left upper-limb bioelectrical impedance contributed most prominently in the ECC model (Figure 5).

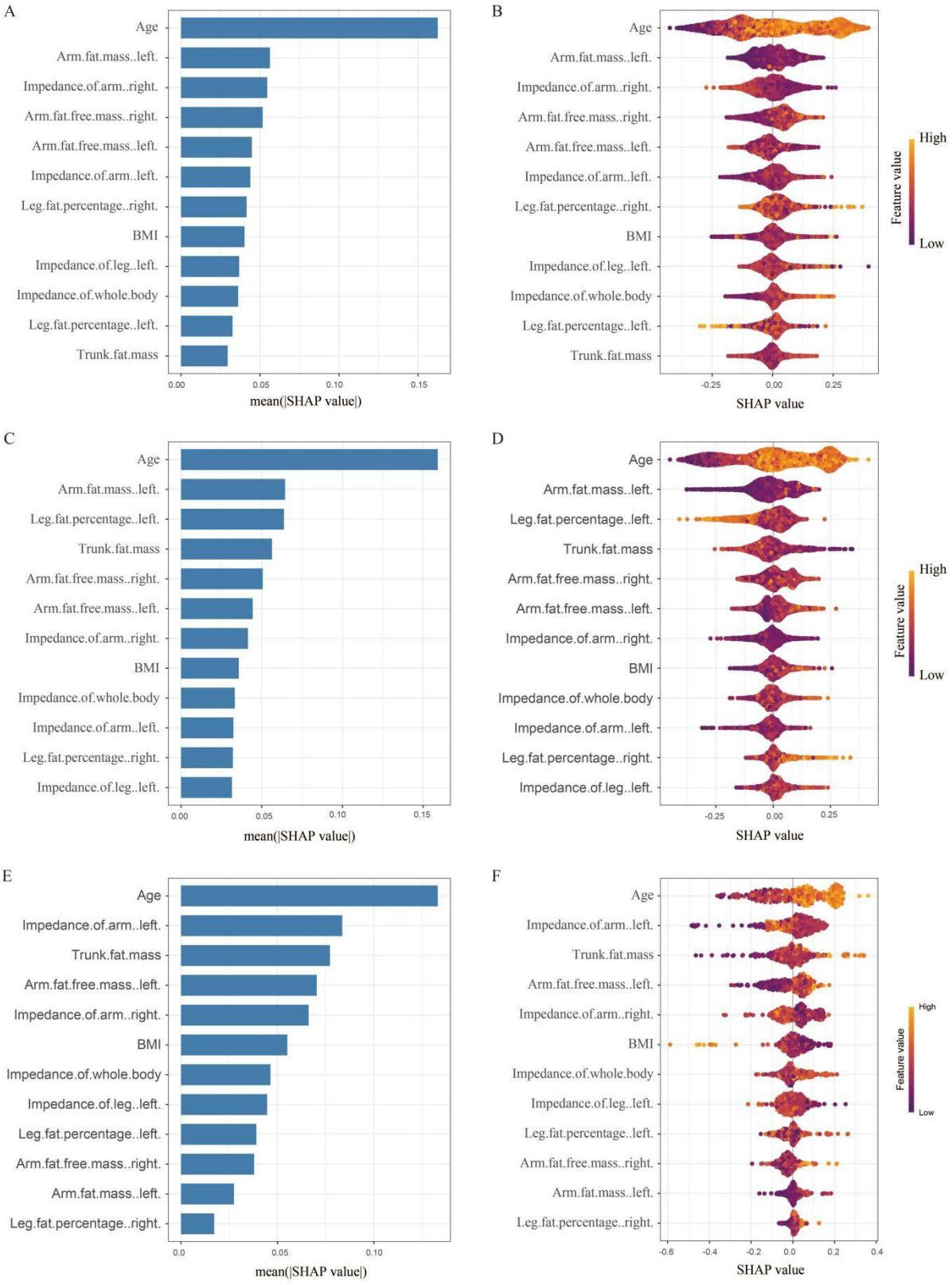


Figure 5 Global feature explanations using the SHAP method. (A) SHAP summary bar plot for overall CCA; (B) SHAP summary dot plot for overall CCA; (C) SHAP summary bar plot for ICC; (D) SHAP summary dot plot for ICC; (E) SHAP summary bar plot for ECC; (F) SHAP summary dot plot for ECC.

4.3. Multi Omics Analysis between regional BC and Pan-Cancer

4.3.1. Baseline Characteristics of Pan-Cancer Cohort

At the individual level, 457,436 participants were included in the analysis of associations between regional BC measures and pan-cancer incidence risk, among whom 70,012 developed cancers during follow-up. The most common cancer types were skin cancer (26,247 cases), prostate cancer (10,812 cases), female breast cancer (9,718 cases), and lung cancer (4,244 cases) (Figure 1). The mean age of the overall population was 56.25 ± 8.09 years. Compared with the cancer-free group, participants who developed cancer were significantly older (59.80 ± 6.99 vs 55.60 ± 8.11 years, $P <$ 0.001). Sex distribution differed significantly between groups ($P < 0.001$). BMI also differed statistically between groups, although the difference in mean values was small (27.55 ± 4.64 vs 27.42 ± 4.79 kg/m², $P < 0.001$). In addition, ethnic composition differed significantly, with a higher proportion of White participants in the cancer group (96.9% vs 93.49%) and relatively lower proportions of Mixed, Asian, and Black participants ($P < 0.001$). With respect to lifestyle factors, the cancer group had higher proportions of former smokers and frequent alcohol consumers (both $P < 0.001$). Socioeconomic characteristics also differed significantly. The cancer group had a lower overall educational attainment, a lower Townsend deprivation index (−1.49 ± 3.02 vs −1.27 ± 3.10, $P < 0.001$), and a higher proportion of low-income participants (Table 7).

Table 7. Baseline characteristics of the participants of cancer cohort.

Characteristics	Total (n=457,436)	Cancer-free (n=387,424)	Cancer (n=70,012)	P-value
Age, mean (SD), years	56.25 ± 8.09	55.60 ± 8.11	59.80 ± 6.99	<0.001
Sex, n (%)				<0.001
Female	248,161 (54.3)	214,726 (55.42)	33,435 (47.76)	
Male	209,275 (45.7)	172,698 (44.58)	36,577 (52.24)	
BMI, mean (SD), kg/m2	27.44 ± 4.77	27.42 ± 4.79	27.55 ± 4.64	<0.001
Ethnicity, n (%)				<0.001
White	430,064 (94.0)	362,220 (93.49)	67,844 (96.9)	
Mixed	2,753 (0.6)	2525 (0.65)	228 (0.33)	
Asian	10756 (2.4)	10066 (2.6)	690 (0.99)	
Black	7473 (1.6)	6886 (1.78)	587 (0.84)	
Unknown	6390 (1.4)	5727 (1.48)	663 (0.95)	
Smoking status, n (%)				<0.001
Never	250665 (54.8)	216203 (55.81)	34462 (49.22)	
Previous	156128 (34.1)	128871 (33.26)	27257 (38.93)	
Current	48367 (10.6)	40460 (10.44)	7907 (11.29)	
Unknown	2276 (0.5)	1890 (0.49)	386 (0.55)	
Alcohol frequency, n (%)				<0.001

Frequent drinkers	198675 (43.4)	165624 (42.75)	33050 (47.21)	
Moderate/Occasional drinkers	169226 (37.0)	144900 (37.4)	24326 (34.75)	
Rare/Non-drinkers	88536 (19.4)	76019 (19.62)	12517 (17.88)	
Unknown	1000 (0.2)	881 (0.23)	119 (0.17)	
Education, n (%)				<0.001
Higher or Professional Education	200896 (43.9)	171042 (44.15)	29854 (42.64)	
Secondary Education	172052 (37.6)	147810 (38.15)	24242 (34.63)	
No Formal Qualification	75737 (16.6)	61298 (15.82)	14439 (20.62)	
Unknown	8751 (1.9)	7274 (1.88)	1477 (2.1)	
Townsend deprivation index	-1.30 ± 3.09	-1.27 ± 3.10	-1.49 ± 3.02	<0.001
Average total household income before tax (£), n (%)				<0.001
Less than 18,000	86299 (18.9)	71271 (18.4)	15028 (21.46)	
18,000 to 30,999	97994 (21.4)	81590 (21.06)	16404 (23.43)	
31,000 to 51,999	102312 (22.4)	87756 (22.65)	14556 (20.79)	
52,000 to 100,000	80688 (17.6)	70426 (18.18)	10262 (14.66)	
Greater than 100,000	21391 (4.7)	18724 (4.83)	2667 (3.81)	
Unknown	68752 (15.0)	57657 (14.88)	11095 (15.85)	

Abbreviations: SD, standard deviation; BMI, body mass index.

4.3.2. Individual-Level Associations Between regional BC and Cancer Risk

Firstly, the PheWAS results indicated that regional BC measures were broadly and significantly associated with a wide spectrum of multisystem diseases, primarily involving cardiovascular, respiratory, and neoplasm-related phenotypes (Figure 6A). Within the neoplasm category, regional BC showed significant associations with several common malignancies, including lung cancer, skin cancer, breast cancer, and CRC. These findings provided a rationale for subsequent in-depth analyses focused on cancer outcomes (Figure 6B).

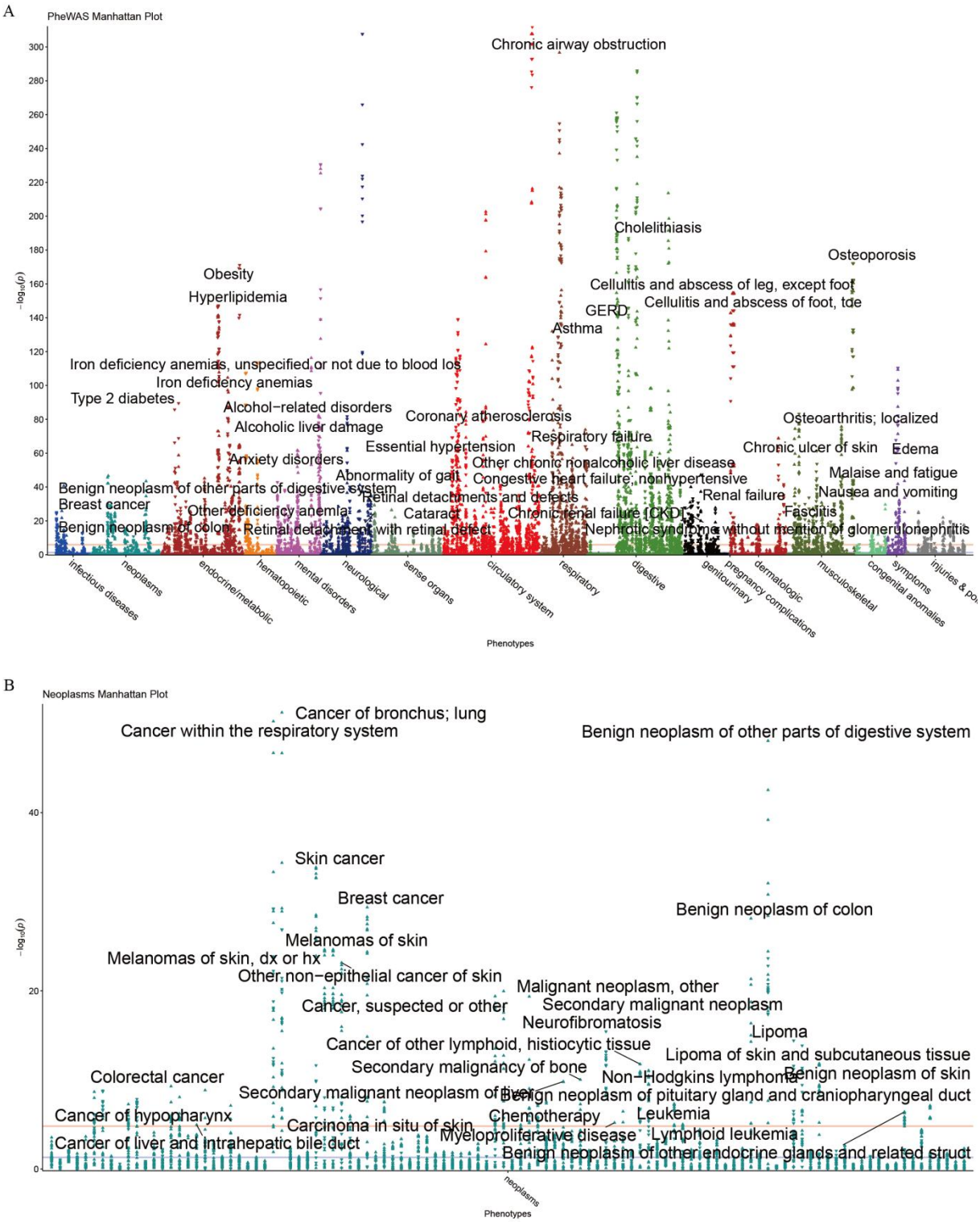


Figure 6. Manhattan plot of phenome-wide association study (PheWAS) results. (A) All system phenotypes; (B) Neoplasm categories. Each point represents a phenotype tested for association with regional body composition (BC). Phenotypes are displayed along the x-axis and grouped by disease category; the y-axis indicates the strength of evidence as −log10(P) from models adjusted for sex, age, BMI, and ancestry. The horizontal pink line denotes the phenome-wide significance threshold after Benjamini-Hochberg (BH) multiple-testing correction. Upward- and downward-pointing triangles indicate positive and negative associations, respectively (effect estimate $\beta > 0$ and $\beta < 0$).

Further analyses incorporating follow-up time showed that multiple regional BC features were also significantly associated with the incidence risk of various cancers (Figure 7). Arm FM was primarily associated with increased risks of pancreatic cancer, lung cancer, female breast cancer, and skin cancer, but with a lower risk of prostate cancer. In contrast, arm FFM was associated with higher risks of colon cancer, kidney cancer, bladder cancer, and prostate cancer, and with lower risks of HCC and head and neck cancers. For the trunk region, trunk FM and trunk FFM were positively associated with risks of female breast cancer and skin cancer. Notably, trunk FM was positively associated with lung cancer risk and inversely associated with prostate cancer risk, whereas trunk FFM showed the opposite pattern, being inversely associated with lung cancer risk and positively associated with prostate cancer risk. Similarly, leg BC features exhibited cancer-specific association patterns. Leg fat percentage (FP) and leg FM were mainly associated with increased risks of lung cancer and female breast cancer. Leg FFM was positively associated with risks of colon cancer, kidney cancer, pancreatic cancer, female breast cancer, and Hodgkin lymphoma, but inversely associated with lung cancer risk. In addition, regional impedance measures were primarily associated with higher risks of HCC and lung cancer and with lower risks of prostate cancer and non-Hodgkin lymphoma.

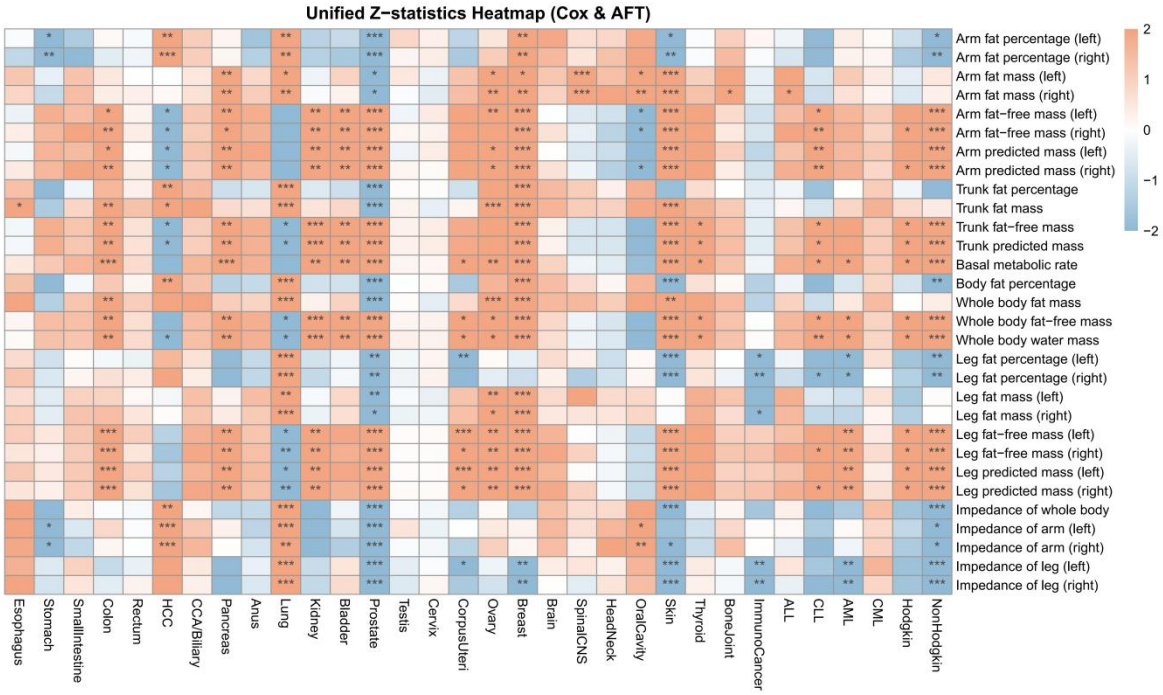


Figure 7. Heatmap of Cox proportional hazards regression results. Rows represent regional body composition features and columns represent cancer endpoints. Each cell shows the hazard ratio (HR)

estimated from multivariable Cox models, with color intensity reflecting the magnitude and direction of the association (HR > 1 indicates increased risk; HR < 1 indicates decreased risk). Statistically significant associations after Benjamini-Hochberg (BH) multiple-testing correction are marked (e.g., asterisks). All models were adjusted for prespecified covariates (e.g., age, sex, BMI, ancestry, and other relevant confounders).

4.3.3. Genetic Causal Associations Between regional BC and Cancer Risk

At the genetic level, we further evaluated associations between regional BC traits and cancer risk. Arm FP was associated with lower risks of female breast cancer and skin cancer. Arm FM was associated with a higher risk of lung cancer, but with lower risks of HCC and female breast cancer. Arm FFM was associated with higher risks of gastric cancer, colon cancer, lung cancer, and bladder cancer. For trunk traits, both trunk FM and trunk FFM were associated with higher risks of lung cancer and non-Hodgkin lymphoma. In addition, trunk FM was associated with a higher risk of bladder cancer and with lower risks of female breast cancer and skin cancer, whereas trunk FFM was associated with a higher risk of female breast cancer. For lower-limb traits, leg FM was primarily associated with higher risks of oesophageal cancer and bladder cancer, and leg FFM was associated with higher risks of colon cancer, lung cancer, and non-Hodgkin lymphoma. BIA-derived impedance features indicated that upper-limb impedance was associated with a higher risk of HCC, and that lower-limb impedance was associated with a higher risk of prostate cancer. In addition, lower-limb impedance was associated with a lower risk of lung cancer (Figure 8).

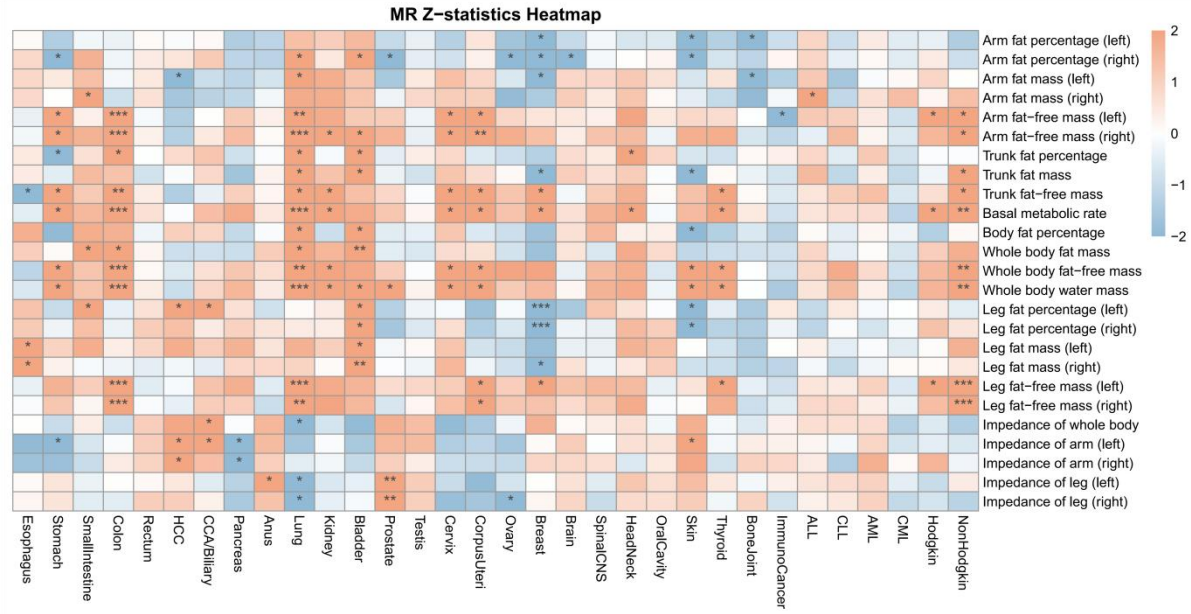


Figure 8. Heatmap of MR analysis results. Rows represent regional body composition features and columns represent cancer endpoints. Each cell shows the odds ratio (OR) estimated from MR analysis

results, with color intensity reflecting the magnitude and direction of the association (OR > 1 indicates increased risk; OR < 1 indicates decreased risk). Statistically significant associations after Benjamini-Hochberg (BH) multiple-testing correction are marked (e.g., asterisks).

Taken together, findings from both the individual-level and genetic-level analyses showed concordant directions for several regional BC traits. Specifically, arm and leg FP demonstrated consistent protective-direction associations with skin cancer risk, and both arm and leg FFM were associated with increased risks of colon cancer and non-Hodgkin lymphoma. Notably, however, arm FP and trunk and leg FM and FFM showed discordant directions of association for lung cancer, female breast cancer, and prostate cancer. To further explore potential explanations for these inconsistencies, we used RCS to test for non-linear associations. The results suggested non-linear relationships between leg impedance and risks of lung cancer and prostate cancer. Non-linear associations were also observed between female breast cancer risk and fat mass in the arm, trunk, and right leg. In contrast, no clear evidence of non-linearity was observed for associations between trunk, leg, and whole-body FFM and lung cancer risk (Figure 9).

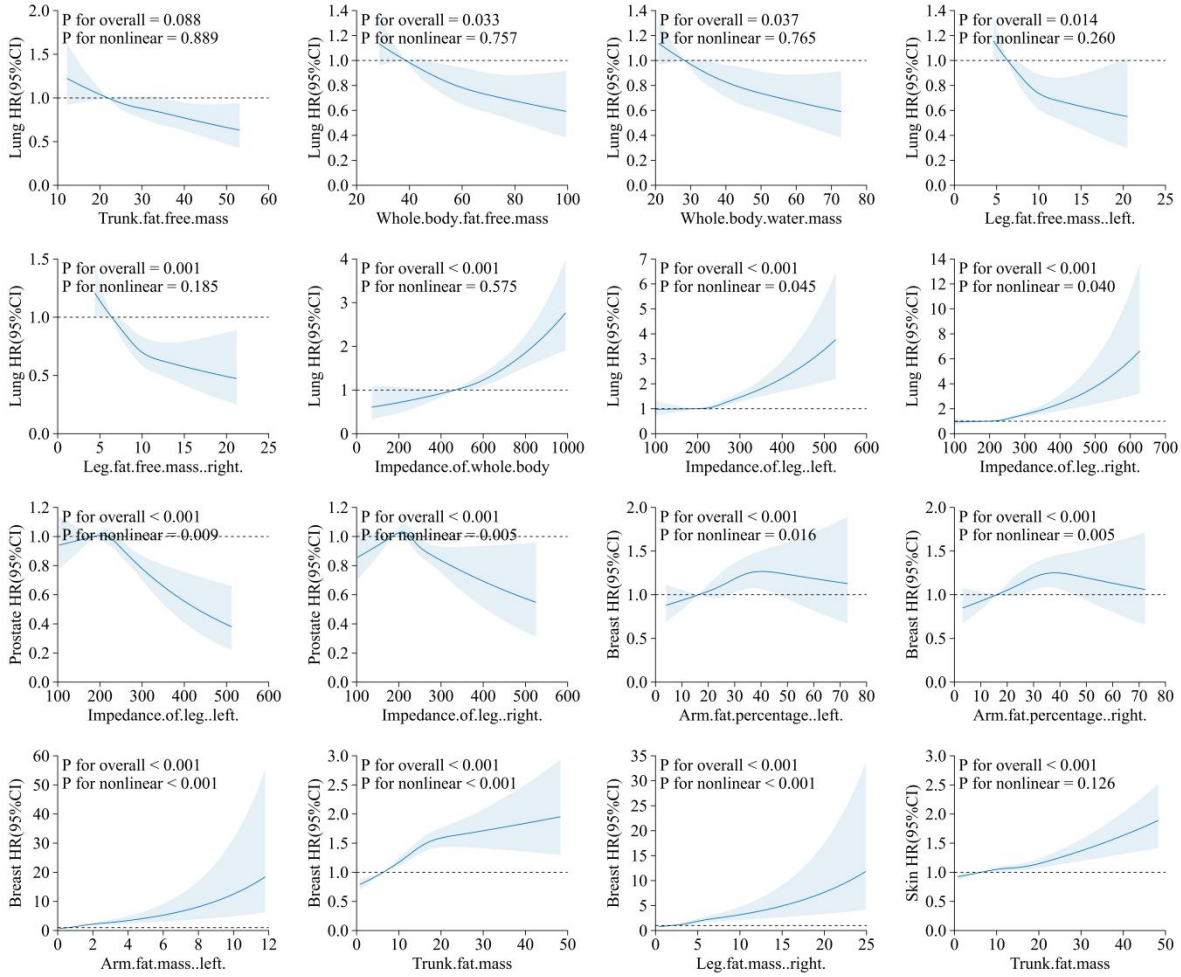


Figure 9 Restricted cubic spline (RCS) analysis illustrating discordant associations between regional BC and cancer risk across the individual-level and genetic-level analyses. The solid line depicts the adjusted hazard ratio (HR) estimated by Cox proportional hazards models incorporating RCS terms for regional BC, and the shaded area indicates the 95% confidence interval. Models were adjusted for prespecified covariates (e.g., age, sex, BMI, ethnicity, and other relevant confounders). P values for the overall association and for nonlinearity were obtained from Wald tests comparing models with and without spline terms.

4.3.4. MOFA based Signatures and Associations between BC and Cancer

We applied MOFA to perform unsupervised integration of multi-omics data from 48,215 participants to characterise latent variation structures that were shared across data layers or specific to individual layers (Figure 10A). By comparing models with different numbers of latent factors, we observed that gains in overall variance explained began to plateau beyond 15 factors, suggesting that this level of complexity captured the major cross-view signals adequately (Figure 10B). In the final model, variance explained differed substantially across views. Composite Metabolic Traits showed the highest variance explained (87.7%), followed by Body Composition (76.2%).

Proteomics explained a moderate proportion of variance (35.6%), whereas Molecular Metabolites explained a smaller proportion (17.2%). These results indicate that the latent factors captured structural variation more effectively for composite metabolic traits and body composition than for the other views (Figure 10C). At the factor level, contributions to variance explained also showed clear view-specific patterns. Factor 1 and Factor 3 accounted for the largest proportions of explained variance in the Body Composition layer, suggesting that they primarily captured dominant variation structures related to body composition. Factor 2 explained the largest proportion of variance in the Composite Metabolic Traits layer, consistent with a shared pattern of variation in metabolic phenotypes. In addition, Factor 4 explained the largest proportion of variance in the Proteomics layer, indicating that it may represent one of the major sources of proteomic variation (Figure 10D).

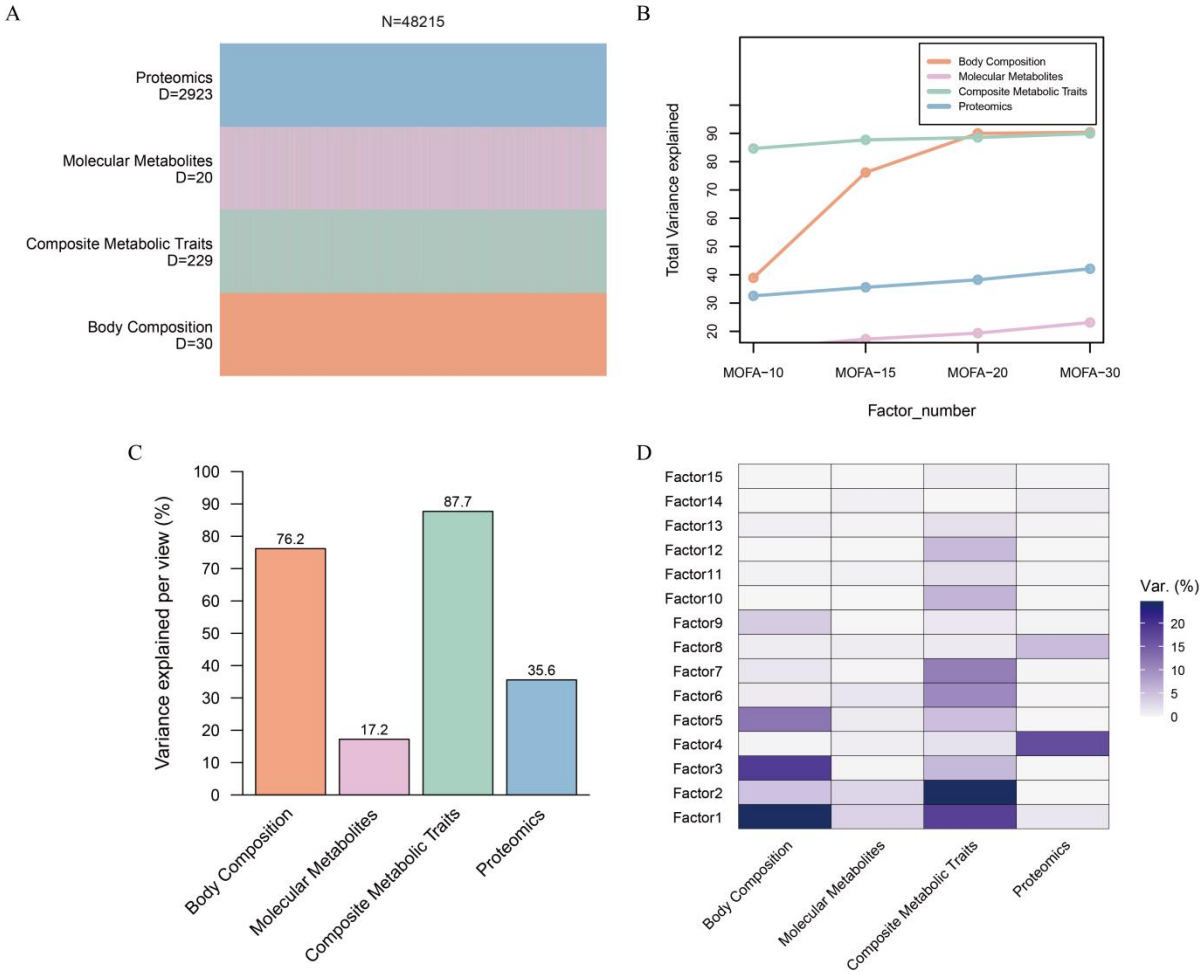


Figure 10. Construction of the MOFA-based multi-omics integration model and assessment of variance explained. (A) Overview of the total sample size and the omics views included in the MOFA framework. (B) Total variance explained by each view across models fitted with different numbers of

latent factors. (C) Variance explained by each view in the final MOFA model with 15 latent factors. (D) Heatmap showing the variance explained by individual latent factors within each omics view, illustrating view-specific contributions and between-factor heterogeneity.

We next evaluated associations between factor scores and risks of multiple cancer outcomes. The results indicated that signals were relatively concentrated for lung cancer, female breast cancer, HCC, skin cancer, and several haematological malignancies (Figure 11). For lung cancer, Factor 1, Factor 2, and Factor 6 showed significant inverse associations with risk, whereas Factor 3 and Factor 5 were associated with increased risk. For hepatobiliary malignancies, Factor 8 showed a strong positive association with HCC risk, while Factor 10 showed a strong inverse association. Associations were also observed for biliary tract tumors, although these signals appeared more outcome-specific and were generally weaker in statistical strength. For skin cancer, Factor 2 and Factor 3 were primarily associated with lower risk. Among haematological malignancies, Factor 1 and Factor 2 were associated with reduced risk of non-Hodgkin lymphoma. In addition, Factor 1 and Factor 2 were also associated with lower risks of colon cancer, kidney cancer, and female breast cancer.

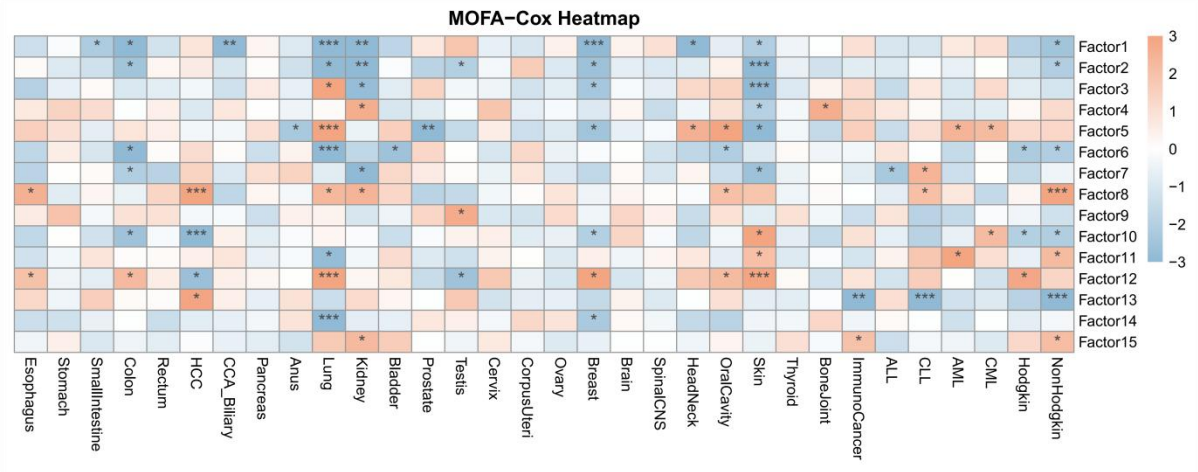


Figure 11. Association heatmap between MOFA latent factors and cancer outcomes. Heatmap summarizing the associations of MOFA-derived latent factors with cancer incidence outcomes estimated using Cox proportional hazards models. Rows represent cancer endpoints and columns represent latent factors (Factor 1–Factor 15). Each cell shows the hazard ratio (HR) estimated from multivariable Cox models, with color intensity reflecting the magnitude and direction of the association (HR > 1 indicates increased risk; HR < 1 indicates decreased risk). Statistically significant associations after Benjamini-Hochberg (BH) multiple-testing correction are marked (e.g., asterisks). All models were adjusted for prespecified covariates (e.g., age, sex, BMI, ancestry, and other relevant confounders).

To further characterise the biological features of MOFA latent factors, we extracted and examined the top 10 feature loadings across the four omics views for the three

factors that explained the largest proportion of variance (Figure 12). For Factor 1, the Body Composition view was dominated by basal metabolic rate and measures related to limb and whole-body FFM. These features showed largely concordant negative loadings, suggesting that this factor captured a global metabolic and lean pattern (GML). Within the Molecular Metabolites view, glycine was one of the major positively loading features, whereas leucine, isoleucine, tyrosine, and glutamine showed predominantly negative loadings. In the Composite Metabolic Traits view, the main contributions were driven by cholesterol and phospholipid proportions within high-density lipoprotein (HDL) subclasses. In the Proteomics view, ALDH1A1, RBKS, and DDT were among the key proteins contributing to Factor 1 (Figure 12A).

For Factor 2, the Body Composition view was characterised primarily by lower-body regional BC features, including leg FP and trunk and leg FFM, which is consistent with a lower-body fat–lean balance pattern (LFLB). In the Molecular Metabolites view, linoleic acid and docosahexaenoic acid (DHA) were prominent contributors. The Composite Metabolic Traits view was mainly driven by triglyceride burden related to low-density lipoprotein (LDL) and intermediate-density lipoprotein (IDL) subclasses, with predominantly positive loadings. In the Proteomics view, CLEC2L, PALM3, and MORF4L2 were important contributors to the variance explained by Factor 2 (Figure 12B).

In contrast, Factor 3 was dominated in the BC view by measures related to bilateral upper-limb and trunk FFM, consistent with an upper-body lean dominant pattern (ULD), highlighting the contribution of trunk and upper-limb lean tissue phenotypes. In the Molecular Metabolites view, valine, creatinine, and glutamine showed overall negative loadings. The Composite Metabolic Traits view was primarily driven by cholesterol and triglyceride proportions within very-low-density lipoprotein (VLDL) and IDL subclasses. The Proteomics view highlighted PKLR, BLVRB, and PRDX2 as key contributing proteins (Figure 12C).

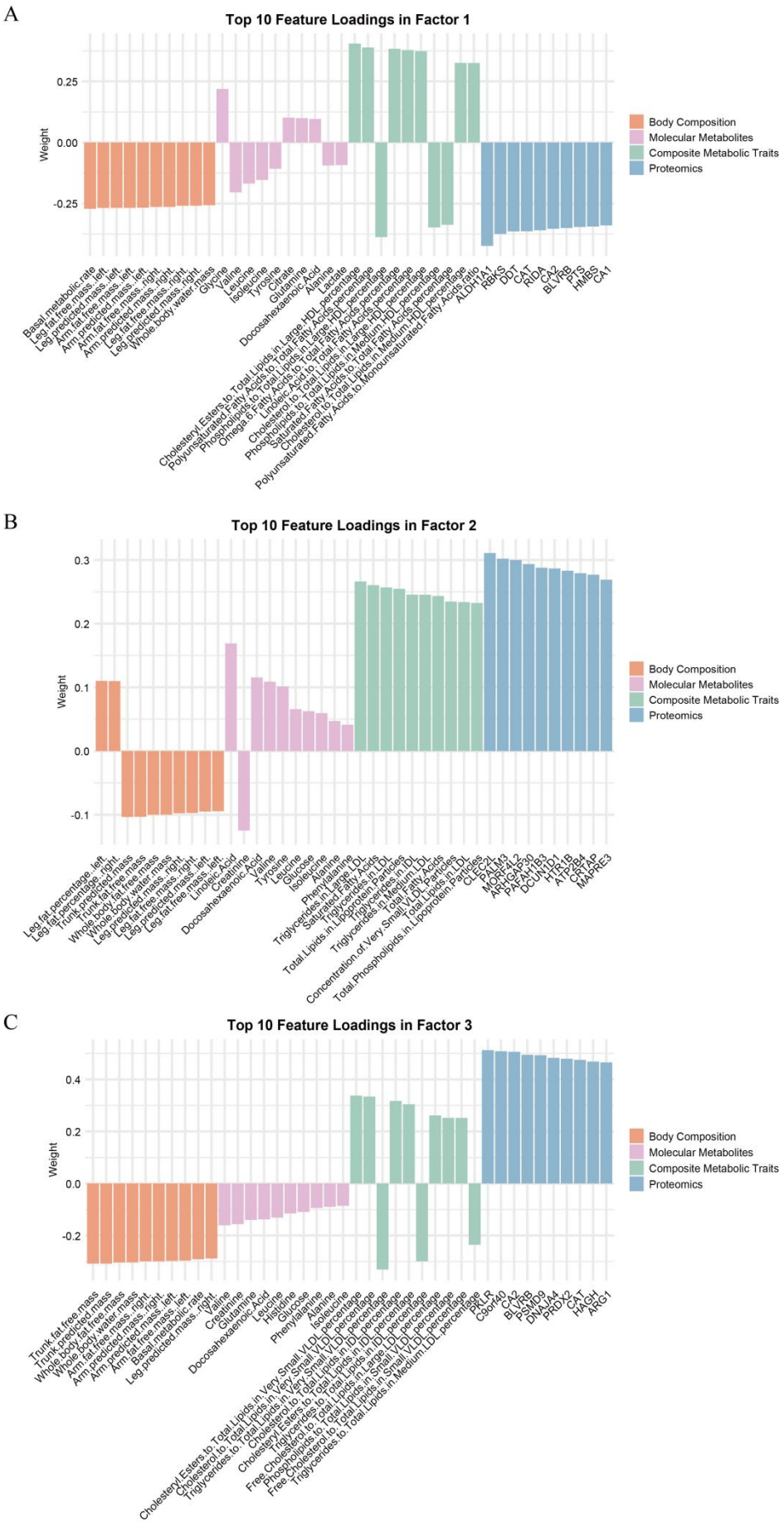

A
Top 10 Feature Loadings in Factor 1
Weight
0.25
0.00
-0.25
Basal.metabolic.rate
Leg.fat.free.mass..left.
Leg.predicted.mass..left.
Arm.fat.free.mass..left.
Arm.predicted.mass..left.
Arm.fat.free.mass..right.
Arm.predicted.mass..right.
Leg.fat.free.mass..right.
Leg.predicted.mass..right.
Whole.body.water.mass
Glycine
Valine
Leucine
Isoleucine
Tyrosine
Citrate
Glutamine
Docosahexaenoic.Acid
Alanine
Lactate
Cholesteryl.Esters.to.Total.Lipids.in.Large.HDL.percentage
Polyunsaturated.Fatty.Acids.to.Total.Lipids.in.Large.HDL.percentage
Phospholipids.to.Total.Lipids.in.Large.HDL.percentage
Omega.6.Fatty.Acids.to.Total.Fatty.Acids.percentage
Linoleic.Acid.to.Total.Fatty.Acids.percentage
Cholesterol.to.Total.Lipids.in.Large.HDL.percentage
Phospholipids.to.Total.Lipids.in.Medium.HDL.percentage
Saturated.Fatty.Acids.to.Total.Fatty.Acids.percentage
Cholesterol.to.Total.Lipids.in.Medium.HDL.percentage
Polyunsaturated.Fatty.Acids.to.Monounsaturated.Fatty.Acids.ratio
ALDH1A1
RBKS
DDT
CAT
RIDA
CA2
BLVRB
PTS
HMBS
CA1
Body Composition
Molecular Metabolites
Composite Metabolic Traits
Proteomics
B
Top 10 Feature Loadings in Factor 2
Weight
0.3
0.2
0.1
0.0
-0.1
Leg.fat.percentage...left.
Leg.fat.percentage...right.
Trunk.predicted.mass
Trunk.fat.free.mass
Whole.body.fat.free.mass
Whole.body.water.mass
Leg.predicted.mass..right.
Leg.fat.free.mass..right.
Leg.predicted.mass..left.
Leg.fat.free.mass..left.
Linoleic.Acid
Creatinine
Docosahexaenoic.Acid
Valine
Tyrosine
Leucine
Glucose
Isoleucine
Alanine
Phenylalanine
Triglycerides.in.Large.LDL
Saturated.Fatty.Acids
Triglycerides.in.LDL
Total.Lipids.in.Lipoprotein.Particles
Triglycerides.in.IDL
Triglycerides.in.Medium.LDL
Total.Fatty.Acids
Triglycerides.in.Very.Small.VLDL
Total.Lipids.in.LDL
Concentration.of.Very.Small.VLDL.Particles
Total.Phospholipids.in.Lipoprotein.Particles
CLEC2L
PALM3
MOFEL2
ARHGAP30
PAFAH1B3
DCUN1D1
HTR1B
ATP2B4
CRTAP
MAPRE3
Body Composition
Molecular Metabolites
Composite Metabolic Traits
Proteomics
C
Top 10 Feature Loadings in Factor 3
Weight
0.4
0.2
0.0
-0.2
Trunk.fat.free.mass
Trunk.predicted.mass
Whole.body.fat.free.mass
Whole.body.water.mass
Arm.fat.free.mass..right.
Arm.predicted.mass..right.
Arm.fat.free.mass..left.
Arm.predicted.mass..left.
Basal.metabolic.rate
Leg.predicted.mass..right.
Valine
Creatinine
Glutamine
Docosahexaenoic.Acid
Leucine
Histidine
Glucose
Phenylalanine
Alanine
Isoleucine
Cholesteryl.Esters.to.Total.Lipids.in.Very.Small.VLDL.percentage
Cholesterol.to.Total.Lipids.in.Very.Small.VLDL.percentage
Triglycerides.to.Total.Lipids.in.Very.Small.VLDL.percentage
Cholesterol.to.Total.Lipids.in.IDL.percentage
Cholesteryl.Esters.to.Total.Lipids.in.IDL.percentage
Triglycerides.to.Total.Lipids.in.IDL.percentage
Free.Cholesterol.to.Total.Lipids.in.Large.LDL.percentage
Phospholipids.to.Total.Lipids.in.Small.VLDL.percentage
Cholesterol.to.Total.Lipids.in.Small.VLDL.percentage
Free.Cholesterol.to.Total.Lipids.in.Small.VLDL.percentage
Triglycerides.to.Total.Lipids.in.Medium.LDL.percentage
PKLR
C9orf40
CA2
BLVRB
PSMD9
DNAJA4
PRDX2
CAT
HAGH
ARG1
Body Composition
Molecular Metabolites
Composite Metabolic Traits
Proteomics

Figure 12. The feature loadings across four omics views for MOFA Factors 1-3. Bar plots display the top 10 features ranked by the largest absolute loadings for each omics view. Loadings quantify both the direction and magnitude of each feature's contribution to the corresponding latent factor; positive and negative values indicate concordant and discordant contributions to factor scores, respectively. The features are ordered by absolute loading within each view.

In addition to Factors 1–3, we further performed phenotypic summarisation and interpretation for Factors 5–8 based on the top 10 loading features within the BC view for each factor. Factor 5 represented a global fat–lean balance pattern (GFLB), characterised by coordinated variation in leg and trunk FP and FM measures. This factor reflected relative balance differences between FM and FFM, with an emphasis on proportional change rather than absolute fat accumulation. Factor 6 represented a generalised adiposity-dominant pattern (GAD), characterised by loadings related to whole-body adiposity and regional FM and FP measures in the trunk, upper limbs, and lower limbs, which together captured overall adiposity levels. Factor 7 represented a trunk adiposity with limb lean mass pattern (TA–LLM), defined by prominent contributions from trunk and upper-body FM measures along with limb FFM measures, suggesting a mixed BC phenotype in which trunk adiposity predominates while limb lean mass varies in parallel. Factor 8 represented a central adiposity-dominant pattern (CAD), with particularly strong contributions from trunk and whole-body FM measures, consistent with a phenotype dominated by central fat distribution and overall adiposity burden. This pattern was also associated with a higher tendency towards metabolic risk.

4.3.5. Enrichment Pathway of molecular metabolites and protein

Metabolite pathway enrichment analyses indicated that the Molecular Metabolites signals for Factor 1, Factor 2, and Factor 3 were jointly and significantly enriched in pathways related to solute carrier (SLC)–mediated amino acid and neurotransmitter transport, as well as protein translation and tRNA aminoacylation. In addition to these shared pathways, Factors 1 and 3 also showed enrichment in bile acid transport and organic anion transport pathways. By contrast, Factor 2 exhibited factor-specific enrichment in amino acid metabolism pathways, primarily involving branches related to tryptophan, glutamine, and biogenic amines (Figure 13A–C). Furthermore, GO and KEGG enrichment analyses based on the Proteomics view suggested distinct functional profiles across factors. For Factor 1, enrichment patterns mainly pointed to oxidative metabolism and detoxification-related processes, including amino acid and carbohydrate catabolism, haem and porphyrin metabolism, and cellular detoxification.

Subcellular localisation terms were predominantly related to the peroxisome (Figure 13D). For Factor 2, enrichment was centred on neurotransmitter signalling and calcium homeostasis. GO terms indicated involvement in negative regulation of serotonin and gamma-aminobutyric acid (GABA) secretion, regulation of cytosolic calcium concentration, and biological processes related to suppression of myocardial contraction. KEGG analyses further suggested a significant association with the cyclic adenosine monophosphate (cAMP) signalling pathway (Figure 13E). For Factor 3, enrichment patterns highlighted potential roles in oxidative detoxification, nitrogen metabolism, and immune regulation. GO and KEGG results indicated significant involvement in hydrogen peroxide detoxification, pyruvate metabolism, and key metabolic pathways such as the urea cycle (Figure 13F).

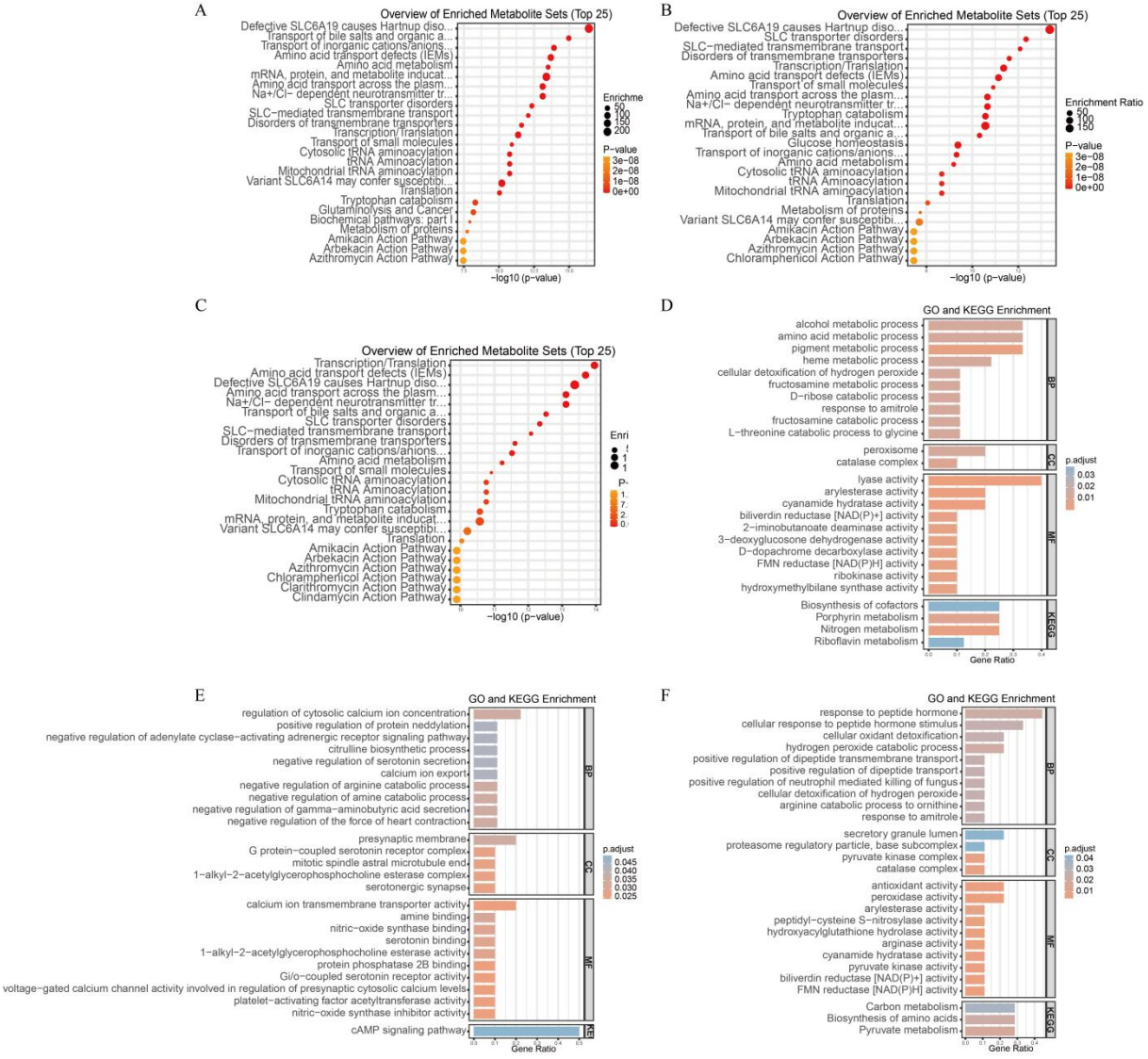


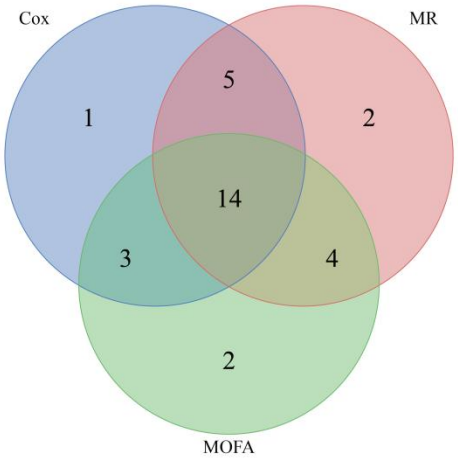
Figure 13. Metabolite pathway enrichment and gene functional enrichment results. (A-C) Bubble plots show pathway enrichment results for molecular metabolites associated with MOFA Factors 1-3,

respectively. (D-F) Bar charts show Gene Ontology (GO) term and KEGG pathway enrichment results for genes corresponding to MOFA Factors 1-3, respectively, and the enrichment results were adjusted for multiple testing using the Benjamini–Hochberg (BH) multiple-testing correction.

4.3.6. The BC-Cancer Risk Association Atlas

Integrating results from the individual-level analyses, genetic-level analyses, and the MOFA latent factor layer, we identified 14 cancer outcomes that reached statistical significance across all three evidence levels (Figure 14A). We then evaluated consistency of association patterns using RCS and excluded cancer outcomes for which the directions of association were inconsistent between the individual-level and genetic-level analyses. This process yielded a final set of 11 cancers as the terminal outcomes for subsequent integrative analyses (Figure 14B). Accordingly, Figure 14C summarizes the MOFA latent factors that were significantly associated with these 11 cancers, indicating selective and heterogeneous relationships between different cancer types and latent phenotypic dimensions (Figure 14C). Building on these findings, we combined the BC view features of the significant latent factors with their effect estimates for cancer outcomes and visualized the association structure between BC feature patterns and cancer endpoints as an integrative risk phenotype map to summarise the multi-omics results (Figure 14D).

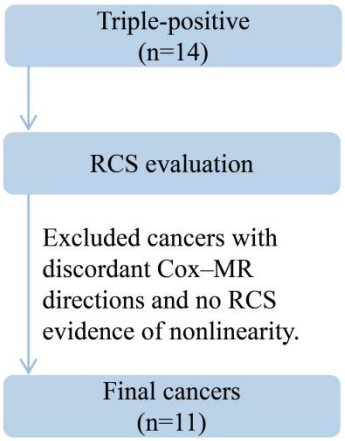
A



C

Cancer	MOFA factor(s)
Lung	1, 2, 3, 5, 6, 8, 11, 12, 14
Breast	1, 2, 3, 5, 10, 12, 14
Skin	1, 2, 3, 5, 7, 10, 11, 12
HCC	8, 10, 12, 13
Colon	1, 2, 6, 7, 10, 12
Kidney	1, 2, 3, 4, 7, 8, 15
Prostate	5
Bladder	6
ALL	7
Hodgkin	6, 10, 12
NonHodgkin	1, 2, 6, 8, 10, 11, 13, 15

B



D

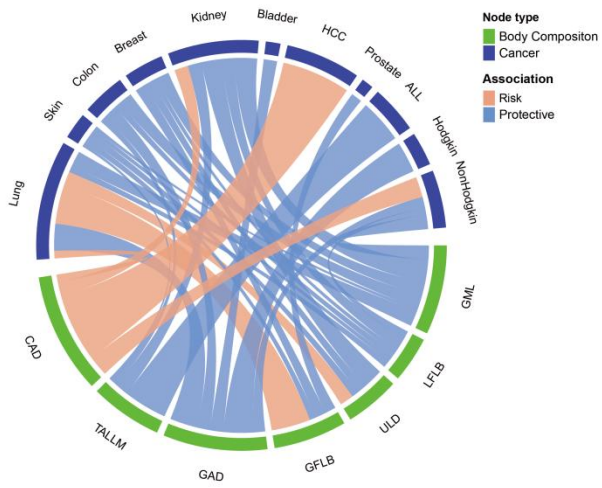


Figure 14. Construction of the final body composition–cancer risk association atlas. (A) Venn diagram showing the number of cancer endpoints that were statistically significant across three evidence layers, including individual-level analyses, genetic-level analyses, and MOFA latent factor–based analyses. (B) Flowchart illustrates the exclusion of cancer endpoints with inconsistent directions of association between the individual-level and genetic-level analyses. (C) Summary of the final set of cancer outcomes retained for downstream analyses and the corresponding significantly associated MOFA latent factors. (D) Chord diagram depicting the atlas of risk associations linking body composition (BC) patterns represented by latent factors to the final cancer outcomes.

5. Discussion

The thesis established a unified research framework in which metabolic endotypes may constitute an important biological link connecting MASLD and cancer, particularly hepatobiliary malignancies. As shown in Figure 15, circulating multi-omic features and measurable host-level representations, particularly regional BC, provide complementary characterization of metabolic endotypes from both molecular and structural perspectives, and thereby offer a basis for disease identification, risk stratification, and the interpretation of potential mechanisms (Figure 15). Within this framework, the individual chapters of this thesis effectively examined different aspects of the overarching model. The findings showed that circulating lipid metabolic features could identify SLD-related metabolic heterogeneity more sensitively than conventional phenotypic classification and provide objective and quantifiable molecular evidence for the refined stratification of MASLD and MetALD. In addition, regional BC was associated with the prognosis of CCA, highlighting the potential value of host structural metabolic phenotypes in tumor progression, risk stratification, and prognostic assessment. Moreover, this association further extended to pan-cancer risk, and integrative multi-omics analyses identified BC patterns with clear biological interpretability, together with potentially relevant molecular pathways. Based on these findings, the following discussion systematically addresses the results of each chapter of this thesis, with a particular focus on the role of lipidomic features in identifying MASLD heterogeneity, the association between regional BC and CCA prognosis, and the significance of regional BC in pan-cancer risk stratification and its potential molecular mechanisms.

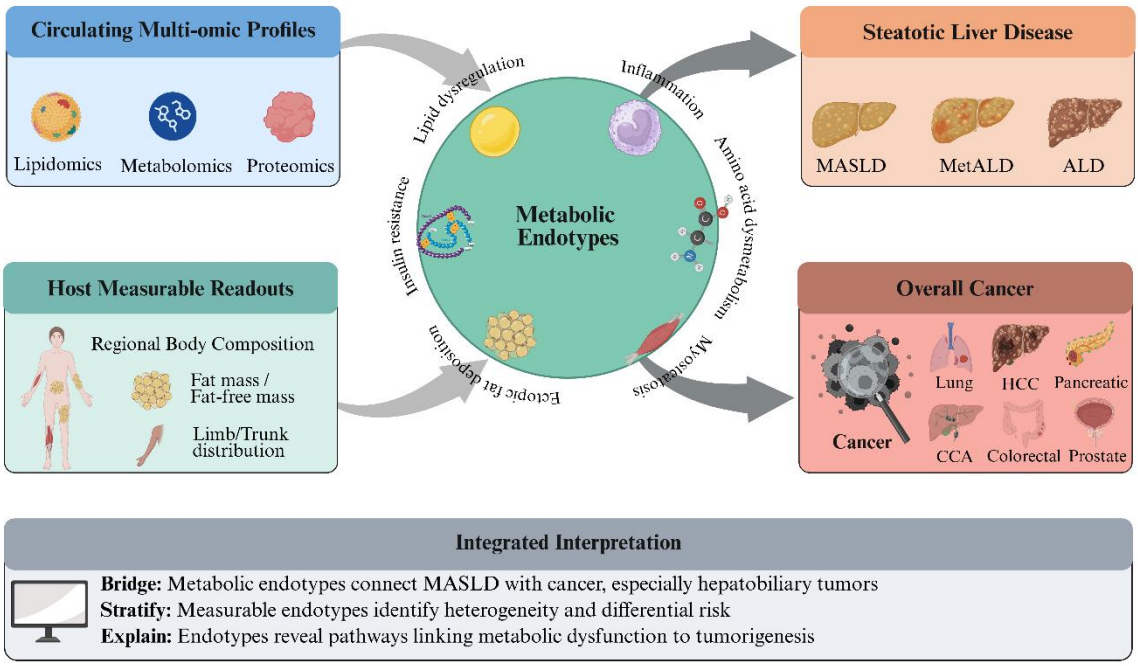


Figure 15. Conceptual framework of metabolic endotypes in the link between MASLD and cancer. MASLD, metabolic dysfunction-associated steatotic liver disease; MetALD, metabolic dysfunction-associated alcohol-related liver disease; ALD, alcohol-related liver disease; HCC, hepatocellular carcinoma; CCA, cholangiocarcinoma.

5.1. Discriminative NMR Metabolites for MASLD and MetALD

Using NMR metabolomics, this project identified multiple lipid biomarkers that showed stable and discriminative differences between MASLD and MetALD. Compared with MASLD, ketone bodies such as acetoacetate and 3-hydroxybutyrate, as well as HDL-related molecular components within large HDL particles, including phospholipids, free cholesterol, and cholesteryl esters, were significantly higher in MetALD and ALD cases and were significantly associated with increased risk, indicating strong discriminatory potential. In addition, genetic evidence provided indirect support for these findings, suggesting clear causal relationships between alcohol consumption and these discriminative lipid biomarkers.

Previous studies have suggested that modifying hepatic lipid fluxes through lifestyle modification and pharmacological interventions may represent a potential strategy to mitigate the steatotic liver disease spectrum and its tumour-related outcomes [161]. This possibility may be related to ethanol metabolism, in which alcohol dehydrogenase–mediated oxidation and the generation of acetaldehyde can disrupt lipoprotein metabolic homeostasis and alter HDL concentration and composition [162,163]. Although the molecular mechanisms by which alcohol increases HDL and its

components have not been fully elucidated, available evidence suggests a link to the central role of HDL in reverse cholesterol transport (RCT) [164]. The association between moderate alcohol consumption and atherosclerosis risk has been extensively investigated [165–167]. Alcohol intake may increase HDL cholesterol levels by upregulating the transport rates of apolipoprotein A1 and apolipoprotein A2 [168]. Cellular cholesterol efflux is a key early step in RCT and depends on phospholipid-rich acceptor particles and the capacity of apolipoprotein A1 (ApoA1) to accept cholesterol [169].

ApoA1, the principal protein component of HDL, exhibited elevated levels in MetALD. The hypothesis is that the RCT system is potentially upregulated in MetALD participants as well, a phenomenon that may be influenced by the interference of ethanol with apolipoprotein metabolism. Phospholipids in various HDL subfractions were also noted as discriminative features between MASLD and MetALD. These findings are supported by prior studies indicating that plasma HDL phospholipid levels are major determinants of cholesterol efflux and that cholesterol efflux increases after alcohol consumption [164,170]. Given the key role of alcohol in lipoprotein structure and cholesterol transport, the observed changes may be attributable to alcohol intake and support a contribution of alcohol exposure to the development of the MetALD phenotype, particularly in terms of the number and composition of HDL subfractions [171]. In addition, MetALD and ALD phenotypes showed higher levels of large and medium HDL particles, which is broadly consistent with previous evidence that increasing alcohol intake is associated with larger HDL particle size and a dose-dependent increase in larger HDL particles [172,173]. This pattern may reflect alcohol-induced changes in lipoprotein lipase activity, together with reduced hepatic lipase activity and decreased cholesteryl ester transfer protein (CETP) activity. These alterations may collectively increase HDL particle size and remodel HDL lipid cargo [174–176]. Nevertheless, the ultimate phenotype of HDL particle size and composition is likely determined by the combined regulation of catabolic pathways and apolipoprotein production. Further validation using more direct functional assays and mechanistic experiments is warranted.

Compared with previous work, our findings further support the rationale for considering MASLD and MetALD as distinct phenotypic categories. Prior studies aiming to differentiate MASLD, MetALD, and ALD have similarly reported more pronounced alterations in HDL particle composition in alcohol-related phenotypes [177,178]. In our study, both MetALD and ALD showed systematic differences in the composition of medium, small, and large HDL particles compared with MASLD, with larger effect

sizes observed for ALD. This pattern suggests that lipid metabolic dysregulation is more pronounced under higher alcohol exposure [179]. These molecular phenotype differences provide support for the MetALD construct. Specifically, the metabolomic profile of MetALD appeared closer to ALD overall, while remaining distinguishable from typical ALD, indicating that MetALD is not a simple subset of MASLD or a noise-driven classification.

From a clinical perspective, our findings highlight the importance of considering disease-specific metabolic landscapes when developing diagnostic approaches and potential therapeutic strategies. This is particularly relevant because MASLD, MetALD, and ALD share substantial pathological overlaps in lipid metabolism [10]. De novo lipogenesis, fatty acid oxidation, and cholesterol transport systems play key roles in MASLD pathogenesis, with particular relevance to hepatocellular LDL and VLDL production [180,181]. VLDL assembly and secretion are tightly coupled with phospholipid homeostasis, and phosphatidylcholine is a critical structural component of VLDL particles [182,183]. In parallel, experimental models suggest that alcohol exposure can exacerbate lipid abnormalities through mechanisms involving AMP-activated protein kinase, which may contribute to the larger effect sizes of lipid dysregulation observed in the metabolomic profiles of MetALD and ALD [184]. Recent lipidomic analyses of ALD and MASLD further indicate that alcohol may promote hepatic uptake of free fatty acids and lysophosphatidylcholine, thereby inducing lipotoxicity and accelerating liver injury [185]. A key distinction of our findings is that we emphasised large HDL subfractions rather than VLDL, LDL, or fatty acid parameters as discriminatory signals. Although the largest effects were observed for ALD, even moderate alcohol consumption has been shown to increase HDL levels [162]. For example, a study in middle-aged women reported that long-term alcohol intake was associated with significant changes in serum HDL cholesterol and ApoA1 [186]. Previous work also supports an approximately linear association between alcohol intake and HDL across heavy drinkers and non-drinkers, suggesting that multiple HDL subfractions remain informative for distinguishing MASLD from MetALD even in the presence of insulin resistance [187]. Changes in HDL size and composition, and therefore function, have been implicated in multiple aspects of metabolic disease pathophysiology [188].

However, circulating lipid metabolic features primarily reveal dynamic perturbations of metabolic endotypes at the molecular level, whereas the long-term effects of metabolic dysregulation on disease outcomes may also be manifested in

remodeling at the host structural level, including changes in regional fat deposition, lean mass distribution, and functional reserve [189,190]. Compared with circulating molecular indicators, regional BC can, to some extent, reflect the cumulative burden of long-term metabolic stress and may be more directly associated with disease progression and survival outcomes in patients with cancer [104,122,125]. Therefore, in the second part of this thesis, we further focused on the relationship between regional BC and the risk of CCA mortality, with the aim of examining the potential role of metabolic endotypes in tumor prognosis from the perspective of host structural phenotypes.

5.2. Regional BC and CCA Mortality

This project evaluated associations between regional BC measures and mortality risk from CCA and its subtypes, and further conducted exploratory, internally validated ML modelling to assess whether BC-centered variables carried prognostic information. We found that FM in the limbs, trunk, and whole body were associated with higher CCA mortality risk. In addition, leg FFM was positively associated with CCA mortality, whereas FFM at other sites did not show consistent significant associations. This heterogeneity was further supported in subtype-specific analyses. Mortality risk from ICC was primarily associated with higher arm and leg FM, whereas mortality risk from ECC was more strongly associated with higher trunk FM and trunk FP, with no significant associations observed for limb FM. In parallel, the internally validated ML analyses suggested that BC-based variables may provide useful discriminatory information for CCA and ICC mortality, with relatively better performance observed for XGB in these outcomes and for NN in ECC. In the model interpretation analyses, age consistently contributed most to prediction, while the relative contribution of regional BC features varied by outcome. For CCA and ICC, left upper-limb FM ranked prominently among BC-related predictors, whereas left upper-limb impedance showed greater contribution in ECC. Together, these findings further underscore the relevance of BC phenotypes to CCA mortality risk and highlight their value for prognostic modelling.

Previous studies have preliminarily revealed potential associations between BC and risk of various cancers. A prospective study found that lean mass was strongly associated with lower cancer risk [121,191]. Similar findings were reported in specific cancer types. Namely, Hagens et al. found that preoperative body fat accumulation and muscle loss negatively affected treatment outcomes in esophageal and gastric cancers,

shortening postoperative overall survival (OS) [192,193]. However, research on CCA remains limited and controversial. A study from Japan showed that BMI and fat accumulation were not associated with survival outcomes in patients with advanced CCA [194]. However, another study found that myosteastosis significantly reduced OS in CCA patients [195]. This discrepancy reflects the complexity of BC function. A longitudinal aging study found that both leg muscle mass and FM increased risk of mortality in older men, while in older women, only leg FM was linked to higher risk of mortality [196]. This finding aligns with our results, where FM in different regions was significantly associated with CCA mortality and varied across subgroups. Similarly, Wan et al. further clarified this phenomenon in bladder cancer through Mendelian randomization analysis, demonstrating that leg FM had a greater impact on mortality risk compared to arm and trunk [139].

Trunk FM contains mainly visceral adipose, which surrounds organs, is highly metabolically active, and is involved in the secretion of a variety of mitogenic factors, such as insulin, insulin-like growth factor and pregnancy-associated plasma protein-A [197,198]. However, this type of fat is insensitive to the antilipolytic effects of insulin but more sensitive to the lipolytic effects of catecholamines, leading more readily to elevated levels of free fatty acids in cancer patients, triggering insulin resistance, which inhibits apoptosis and stimulates tumor development [199,200]. In contrast, limb fat, mainly subcutaneous adipose, has relatively lower metabolic activity. Although it also secretes some hormones and cytokines, it is typically associated with more secretion of anti-inflammatory factors such as adiponectin and IL-10 [201,202]. Moreover, increased limb FM often accompanies decreased muscle mass, the phenomenon known as myosteatosis, which is particularly pronounced in cancer patients [203,204]. The presence of myosteatosis not only indicates abnormal accumulation of fat within the muscle tissue but also reflects decreased in overall endurance and metabolic imbalance [204]. This is consistent with our study, which indicates that the increased mortality risk in CCA patients is primarily linked to fat accumulation rather than changes in FFM. Studies have shown that ectopic fat deposition in the muscle promotes the release of cytokines and inflammatory mediators, leading to thrombocytosis, immunosuppression, and cachexia, thus severely affecting the prognosis of patients [205,206]. Additionally, myosteatosis is associated with increased chemotherapy toxicity and early treatment interruption, resulting in early mortality outcome [207,208].

With the development of artificial intelligence, numerous studies have created ML models to predict disease and cancer outcomes. For instance, Hu et al.[209] constructed an interpretable predictive model for acute kidney injury in children using gender, BMI, and blood indicators. Li et al.[210] developed a diagnostic model for gastric cancer based on tongue images. Although these ML models have demonstrated potential in cancers, they usually rely on a single algorithm and ignore the complexity of patients' physiological and metabolic characteristics. Moreover, there are currently no prognosis models for CCA. To address this gap, we employed multiple ML algorithms and incorporated BC variables to develop a prognostic prediction model for CCA and its subtypes. The application of BC not only compensates for the limitations of traditional models in terms of individual specificity, but also has the advantages of being cost-effective and noninvasive. It is now widely used in the prediction of various diseases, including hypertension, nonalcoholic fatty liver disease, and osteoporosis [211–213]. In addition, the application and comparison of multiple ML algorithms more comprehensively reveal data characteristics and associations, mitigates possible biases of single models and improves prediction accuracy and stability. Nonetheless, However, because these ML analyses in this study were exploratory and only internally validated, the models should not yet be considered clinically applicable and require external validation in independent populations.

Although the present study demonstrated a significant association between regional BC and mortality risk in patients with CCA, and further suggested the potential value of host structural metabolic phenotypes in prognostic stratification, the implications of this finding are not confined to CCA as a single cancer type. In fact, regional fat deposition, reduced muscle reserve, and the long-term metabolic stress reflected by these features are not unique to CCA, but may instead constitute a common host basis underlying the occurrence, progression, and adverse clinical outcomes of multiple cancers. Moreover, the molecular basis and potential biological mechanisms corresponding to regional BC across multiple cancers, including CCA, remain inadequately and insufficiently characterized in a systematic manner. Therefore, to further clarify the overall pattern of associations between regional BC and the risk of multiple cancers, and to deepen the mechanistic understanding of its biological significance, the present study further integrated metabolomic and proteomic data to systematically investigate the potential biological pathways and molecular basis underlying differences in cancer risk related to distinct patterns of regional BC.

5.3. Pan-Cancer Atlas of BC with Multi-Omics Integration

To further address the scientific question of how regional BC relates to risk across different types of cancer, we established an analytical framework in which three complementary evidence streams were jointly evaluated, including individual-level association analyses, genetic causal inference, and unsupervised multi-omics integration. Using this triangulated framework, we explored potential molecular mechanisms and ultimately constructed a BC–cancer risk map. At the individual level, upper-body FM was mainly associated with higher risks of lung cancer, female breast cancer, and skin cancer. In contrast, lower-limb FFM showed stronger associations with higher risks of colon cancer, kidney cancer, and pancreatic cancer. These patterns were also supported by genetic causal inference. In the subsequent MOFA-based multi-omics integration, we identified latent factors closely related to regional BC. Several BC patterns, including GML, LFLB, and ULD, were associated with risk signals for lung cancer, female breast cancer, HCC, and skin cancer. The underlying molecular pathways implicated by these factors primarily involved amino acid transport, oxidative detoxification, lipoprotein metabolism, and neuroendocrine processes. By integrating evidence across the three analytical layers, we further prioritised cancer outcomes with consistent directions and robust signals, highlighting the heterogeneity of regional BC influences across cancer types and providing a conceptual basis for future precision risk stratification and mechanistic investigations.

Compared with conventional indices such as BMI and waist circumference, regional BC measures may more sensitively capture the decoupling between fat deposition patterns and skeletal muscle reserves, thereby characterising metabolic load and metabolic buffering in a spatially resolved manner [214]. Previous studies have shown that, even at the same BMI, differences in BC distribution can be associated with markedly different metabolic risk profiles for colon cancer [215]. The central metabolic load represented by visceral fat and upper-body fat deposition has been reported to influence cancer risk independently of BMI and waist circumference [105,215]. Trunk adiposity more directly reflects central metabolic load, and VAT has higher metabolic and inflammatory activity [216]. VAT can promote chronic low-grade inflammation and insulin resistance, amplify insulin and insulin-like growth factor (IGF) axis signalling, and induce adipokine imbalance, thereby facilitating cellular proliferation, angiogenesis, and the development of an immunosuppressive microenvironment [217,218]. In contrast, arm adiposity largely represents upper-body SAT, which is more lipolytically active and

contributes more to circulating free fatty acids (FFAs). Upper-body SAT often co-occurs with trunk fat accumulation and is generally associated with a more adverse metabolic profile [219,220]. This pattern may help explain why arm and trunk FM showed concordant risk associations across several cancers, including lung cancer, female breast cancer, skin cancer and prostate cancer in our study. Conversely, lower-body adiposity is more likely to function as a metabolic sink. Because lipolysis is relatively constrained, lower-body fat may reduce fatty acid spillover and ectopic deposition in the liver and skeletal muscle, thereby mitigating lipotoxicity and systemic metabolic stress [221,222]. In addition, lean tissue or fat-free mass may reflect metabolic buffering capacity and can influence cancer development and host susceptibility through pathways such as glucose disposal, secretion of anti-inflammatory myokines, and stress tolerance [223]. However, this buffering effect is not uniformly protective. For example, in our study, both trunk FM and trunk FFM were positively associated with risks of female breast cancer and skin cancer. This is consistent with previous evidence showing that the fat-free mass index (FFMI) is significantly associated with higher postmenopausal breast cancer risk, with an association stronger than that observed for the body fat mass index (BFMI) [190]. The underlying explanation may not be a direct carcinogenic effect of healthy skeletal muscle per se. Instead, it may reflect larger visceral organ volume, increased metabolic load, and myosteatosis, which can further enhance insulin–IGF signalling and chronic inflammation [224,225].

At the cancer-type level, metabolically sensitive cancers such as colon cancer, HCC, and kidney cancer are more likely to be driven by insulin resistance, chronic inflammation, and lipid metabolic reprogramming mediated by central fat accumulation [226–228]. HCC represents a prototypical example. The portal vein effect of VAT allows lipolytic products and pro-inflammatory mediators to reach the liver more directly, thereby promoting hepatic insulin resistance and lipid metabolic dysregulation, and potentially facilitating hepatocarcinogenesis during the progression of MASLD and MASH [229]. By contrast, hormone-sensitive cancers, such as female breast cancer and prostate cancer, may show associations not only with increased upper-body adiposity but also with trunk FFM. The underlying mechanisms may extend beyond metabolic and inflammatory pathways and involve additional contributions from hormonal axes and cytokine signalling. Obesity-related inflammation in breast adipose tissue can lead to the formation of crown-like structures of the breast (CLS-B) [230]. Through activation of inflammatory signalling, including the COX-2 and prostaglandin E2 (PGE2) pathway,

CLS-B can upregulate aromatase expression in adipose stromal cells, thereby increasing local oestrogen production and enhancing oestrogen receptor (ER) signalling [230,231]. In addition, adipokines such as leptin and interleukin-6 (IL-6) may act in concert with the insulin–IGF axis, strengthening the biological coupling between adiposity and breast carcinogenesis [232,233]. Moreover, for metabolic–environment interaction tumours, including lung cancer, skin cancer, and lymphomas, exogenous exposures such as smoking, ultraviolet radiation, and infection or immune dysregulation are often dominant drivers. However, regional BC may modify these effects by influencing host immunometabolic status [234]. Prior evidence suggests that reduced muscle reserve is associated with impaired mitochondrial function, elevated systemic inflammation, and decreased tolerance to immune and metabolic stress, which may increase host vulnerability to smoking-related DNA damage and oxidative stress [235]. This may create synergistic effects between regional FFM and smoking exposure [235,236].

Beyond single-site effects, we observed that several regional BC traits showed consistent co-variation signals for specific cancers, suggesting that their biological effects are not independent. Based on the MOFA model, the GML pattern and LFLB pattern captured metabolic buffering phenotypes dominated by whole-body and lower-body lean mass, respectively, and were associated with lower risks across multiple cancers, including lung cancer, female breast cancer, skin cancer, and colon cancer. This protective-direction association may be partially mediated by the proportions of cholesterol and phospholipids within HDL subclasses. HDL promotes cholesterol efflux via ApoA1, thereby limiting clustering of receptor tyrosine kinases on tumour cell membranes and attenuating downstream signal amplification [237,238]. ApoA1 may also suppress inflammasome-related signalling and reduce pro-inflammatory cytokine production [239]. In addition, greater lean mass may reduce hepatic triglyceride flux and the burden of remnant particles, which could further alleviate systemic inflammatory pressure [240]. Notably, the ULD pattern behaved as a risk-related pattern for lung cancer, indicating that an upper-body lean predominance does not necessarily translate into metabolic protection. When the burden of triglyceride-rich lipoprotein (TRL) remnants or remnant cholesterol (RC) is high, inflammatory activity may increase and offset potential metabolic benefits [241]. Prior studies have also suggested that RC may act synergistically with inflammatory markers to increase cancer risk, with a more pronounced elevation for lung cancer [242,243]. Consistent with this interpretation, the CAD pattern was associated with higher risks of lung cancer, HCC, kidney cancer, and

non-Hodgkin lymphoma. In addition, within the GML, LFLB, and ULD patterns, lower availability of BCAAs and glutamine may reduce intracellular amino acid availability, thereby attenuating SLC transporter-mediated mechanistic target of rapamycin complex 1 (mTORC1) activation and downstream protein synthesis [244]. This may ultimately constrain tumour biosynthetic capacity and stress adaptation [245,246]. Enrichment of glycine may also reflect greater metabolic buffering capacity. Ramos-Jiménez, et al. [247] reported that glycine can promote muscle growth and regeneration, improve muscle function, and exert anti-inflammatory and anti-catabolic effects. Moreover, DHA may modulate membrane lipid platforms and Toll-like receptor (TLR) signalling, thereby reducing chronic inflammation and an immunosuppressive tendency, which could act synergistically with protective BC phenotypes [248,249].

At the pathway level, the GML pattern was accompanied by features consistent with higher mitochondrial activity and oxidative phosphorylation, which may sustain reactive oxygen species (ROS) generation and lipid peroxidation stress. Proteins such as ALDH1A1, RBKS, and DDT were mainly enriched in pathways related to oxidative metabolism and detoxification, suggesting that this pattern may mediate protective effects by strengthening metabolic homeostasis and enhancing clearance of toxic by-products. In particular, ALDH, as a member of the aldehyde dehydrogenase family, can detoxify reactive aldehydes generated during lipid peroxidation, thereby attenuating oxidative stress-driven inflammatory cascades and tissue injury and limiting the formation of a pro-tumorigenic microenvironment [250,251]. However, within tumour cells, ALDH1 is also frequently used as a marker of cancer stem cells (CSCs) and drug resistance and has been associated with tumour progression and poor prognosis [252]. Additionally, the LFLB pattern was primarily linked to neurotransmitter signalling, calcium homeostasis, and cAMP signalling, suggesting that it may influence cancer susceptibility through neuroendocrine-immune pathways. Because lower-limb skeletal muscle constitutes a large muscle mass and is closely related to habitual physical activity, exercise-induced sympathetic activation and catecholamine release can rapidly mobilise and redistribute immune cells, including natural killer (NK) cells, into tumour tissue [253]. Evidence indicates that voluntary running can suppress tumour growth through adrenaline- and IL-6-dependent mobilisation of NK cells [254]. In addition, exercise-induced myokine secretion can modulate metabolism, inflammatory responses, and immune cell function in distant tissues and within the tumour microenvironment [255,256]. Hojman, et al.[257,258] further reported that plasma muscle-derived factors increase

substantially during exercise, and that oncostatin M can exert direct anti-proliferative effects on hormone-sensitive breast cancer cells. Relative to LFLB, the ULD pattern may be accompanied by lower levels of activity-related metabolic stimulation and reduced myokine release, which could weaken the neuroendocrine-immune modulation conferred by exercise.

5.4. Integrative Interpretation Within the Metabolic Endotype Framework

Taken together, the findings across the chapters of this thesis converge on a common biological framework, namely that circulating molecular features and host structural phenotypes can serve as complementary readouts of underlying metabolic endotypes and characterize, from different dimensions, the key biological features linking MASLD and cancer, particularly hepatobiliary malignancies. These findings not only provide supporting evidence for precise disease discrimination and risk stratification but also lay a foundation for further elucidation of the underlying mechanistic pathways.

First, at the level of circulating molecular readouts, this study found that lipid components related to large HDL particles could relatively stable distinguish MASLD from MetALD, suggesting that lipidomic features have good discriminatory potential in the differentiation of SLD subtypes. It should be noted that although MASLD and MetALD both belong to the spectrum of SLD, they do not necessarily share identical metabolic driving mechanisms, but may instead correspond to distinct metabolic endotypes shaped by different combinations and relative contributions of underlying factors [10,23]. Specifically, MASLD is more likely to represent an endogenous state of metabolic imbalance predominantly driven by metabolic dysfunction, whereas MetALD may better reflect a more complex metabolic disturbance arising from the combined effects of metabolic burden and alcohol exposure [10]. Moreover, previous studies have shown that alcohol intake can influence HDL levels and the composition of lipoprotein subclasses, suggesting that HDL-related lipid features may reflect alcohol-related lipoprotein metabolic remodeling rather than simply representing traditionally protective lipid changes [259]. This further supports the interpretation that HDL-related lipid components are more likely to capture integrated molecular features involving lipoprotein remodeling, shifts in energy substrate utilization, and alcohol-related metabolic abnormalities. Therefore, within the framework of this study, lipidomics not only provides molecular-level support for the subtype differentiation of MASLD and

MetALD, but also offers measurable circulating molecular readouts for understanding the potential differences in their underlying metabolic drivers.

In addition, the findings of this study regarding regional BC may be understood as a further characterization of underlying metabolic endotypes at the level of host structural readouts. Previous studies have shown that abnormalities in BC, particularly sarcopenia, abnormal fat distribution, and reduced functional reserve, are closely associated with adverse outcomes in CCA and multiple other cancer types [19,123,131]. Building on this foundation, the present study found that multiple regional BC features were consistently associated with mortality risk in CCA and its subtypes, and that these associations could be extended to the level of pan-cancer risk. These results suggest that regional BC should not be regarded merely as a set of local anatomical indicators, but rather as a cumulative host-level manifestation of long-term metabolic burden, patterns of energy allocation, and functional reserve. Moreover, such associations are unlikely to be unique to any single cancer type, but may instead correspond to a host metabolic susceptibility state that exists across cancers. More importantly, by further integrating multi-omics analyses, this study identified biologically interpretable BC patterns, suggesting that the observed associations may involve molecular pathways related to amino acid transport, protein translation, oxidative detoxification, and neurotransmitter-related processes. These findings provide a deeper molecular explanation for the links between regional BC and cancer risk and prognosis. Previous studies have shown that tumor cells and their host microenvironment are highly dependent on amino acid uptake and transport, protein synthesis, and redox homeostasis, and that these processes are also closely related to muscle wasting, nutritional redistribution, metabolic reprogramming, and systemic inflammation in cancer-associated cachexia [260–262]. Therefore, the regional BC patterns identified in this study no longer merely reflect differences in fat mass alone, but more likely represent a systemic metabolic state that is closely related to nutrient substrate supply, protein synthesis demands, adaptation to oxidative stress, and tumor-host interactions. Taken together, regional BC not only constitutes a host structural readout of metabolic endotypes, but also extends into potential molecular mechanism networks associated with tumor initiation and progression, thereby enhancing its biological interpretability.

In summary, within the metabolic endotype framework proposed in this study, lipid components related to large HDL particles and regional BC represent two complementary categories of readout at the circulating molecular level and the host

structural level, respectively. The former reflects dynamic molecular changes such as lipoprotein remodeling and alcohol-related metabolic perturbations, whereas the latter captures host structural remodeling and differences in functional reserve under the cumulative effects of long-term metabolic burden. Together, these findings support the central hypothesis of this thesis, namely that metabolic dysregulation does not act separately and independently on MASLD, CCA, and other cancers, but may instead be continuously manifested through a series of measurable metabolic endotypes across different stages of disease and different organ systems, thereby further influencing disease stratification, prognostic assessment, and the identification of potential mechanistic pathways.

5.5. Limitations

This study has several limitations that should be acknowledged. First, all individual-level data were derived from the UK Biobank, which predominantly includes participants of European ancestry, potentially limiting the generalisability of our findings to other populations. Second, in the MASLD-related analyses, a key constraint was the unavailability of metabolomic measurements on the same day as the liver MRI assessment. Metabolomics data were collected at baseline and at follow-up 1, whereas the MRI was performed at follow-up 2, which limits our ability to evaluate the temporal relationship between ongoing alcohol intake and the observed metabolomic differences. In addition, in the Mendelian randomisation (MR) analyses, alcohol consumption was based on self-reported data, which may be subject to reporting bias and may not adequately capture heterogeneity in drinking patterns (for example, frequency, intensity, and episodic drinking). Although we selected SNPs associated with alcohol consumption from GWAS, these variants may not fully reflect biologically grounded variation in alcohol exposure. Third, in the BC-related analyses, regional BC was primarily assessed using BIA. BIA is sensitive to hydration status, measurement conditions, and short-term physiological fluctuations, which may introduce non-differential measurement error and reduce the precision of effect estimates. Moreover, BC is a dynamic exposure, whereas most analyses relied on a single baseline measurement, which may lead to regression dilution bias and may not adequately capture long-term exposure levels or temporal changes. Fourth, although we adjusted for a broad set of known covariates in the PheWAS, Cox regression, and downstream modelling, residual confounding cannot be fully excluded in an observational design.

Fifth, in the pan-cancer risk analyses, substantial heterogeneity exists across cancer types. For some cancers, limited case numbers and complex histological and molecular subtypes may result in unstable effect estimates. Therefore, findings for rare cancers and highly heterogeneous cancers should be interpreted with caution. Sixth, in the multi-omics integration analyses, NMR metabolomics and proteomics data were available only in the subset of UK Biobank participants. The resulting sample size was substantially smaller than that of the full regional BC dataset, and differences in batch effects, missingness mechanisms, and inclusion criteria may further increase the risk of selection bias. Finally, the association patterns and prediction models identified in this study were not validated in independent external cohorts. Their generalisability and clinical utility therefore require further evaluation across diverse populations, platforms, and multi-centre datasets.

5.6. Future Directions

Building on UKB data, this study established a systematic framework spanning metabolic biomarker prioritisation, regional BC phenotype association analyses, pan-cancer risk profiling, and multi-omics mechanistic interrogation. To further enhance the generalisability, robustness, and translational value of our conclusions, several future directions warrant pursuit. First, multi-ancestry replication studies are needed to reproduce our findings in non-European cohorts and to evaluate consistency in effect directions and magnitudes. External validation in independent biobanks or multi-centre clinical cohorts is also essential to assess reproducibility of associations and to test the calibration and transportability of the prediction models. Second, future work could leverage high-resolution imaging modalities, including MRI, CT, and dual-energy X-ray absorptiometry, to quantify fat distribution, muscle mass, and myosteatosis with greater precision. Using more accurate phenotyping, subsequent studies can more comprehensively characterise associations between BC and cancer incidence and prognosis. In addition, a multimodal integration strategy should be advanced by jointly modelling imaging-derived body composition, physical activity, sleep, diet, and clinical biomarkers, thereby enabling a more interpretable and clinically actionable personalised risk stratification system. At the multi-omics level, future studies should expand sample sizes for UKB metabolomics and proteomics, and incorporate cross-platform validation using complementary technologies, such as LC–MS-based metabolomics or lipidomics and transcriptomics, to strengthen the stability and reproducibility of latent multi-omics

factors. Finally, to achieve a closed loop from population-level associations to mechanisms and translational applications, candidate lipid biomarkers should be validated in cellular and animal models to establish biological plausibility and diagnostic potential for distinguishing MASLD from MetALD. Functional validation of key pathways and candidate targets linking regional BC to cancer risk and prognosis will also be necessary to identify intervention-relevant targets and to inform future clinical risk management and targeted prevention or treatment strategies.

6. Conclusion

The present study indicates that measurable metabolic endotypes reflected by circulating multi-omic features and regional BC may constitute an important biological basis for MASLD and cancer, and may contribute to the development of SLD heterogeneity, tumor progression, and differences in cancer risk. At the circulating molecular level, the results showed that ketone bodies and lipidomic components related to HDL particles could distinguish MASLD from MetALD with relatively high accuracy and demonstrated promising diagnostic potential, suggesting that circulating lipidomic features may provide important complementary information for the identification of SLD subtypes. In addition, at the host structural level, regional BC was associated with mortality risk of CCA and its subtypes. Specifically, limb fat-related features were more strongly associated with ICC, whereas trunk fat-related features were more strongly associated with ECC. Moreover, these associations could be further extended to pan-cancer risk. Through integrative multi-omics analyses, biologically interpretable BC patterns associated with specific cancers were further identified, and these associations may involve molecular pathways related to amino acid transport, protein translation, oxidative detoxification, and neurotransmitter-related processes. In summary, the present study suggests that metabolic dysregulation does not act separately and in isolation on MASLD and cancer, but is more likely to be continuously manifested through a series of measurable metabolic endotypes across different stages of disease and different organ systems. As two complementary categories of readout, circulating multi-omic features and regional BC not only help deepen the understanding of MASLD heterogeneity and differences in cancer risk, but also provide new perspectives for precise disease identification, risk stratification, and the investigation of potential mechanisms.

7. Reference

1. Saklayen MG. The Global Epidemic of the Metabolic Syndrome. *Curr Hypertens Rep*. 2018;20(2):12. doi:10.1007/s11906-018-0812-z

2. Noubiap JJ, Nansseu JR, Lontchi-Yimagou E, et al. Global, regional, and country estimates of metabolic syndrome burden in children and adolescents in 2020: a systematic review and modelling analysis. *The Lancet Child & Adolescent Health*. 2022;6(3):158-170. doi:10.1016/S2352-4642(21)00374-6

3. Madan K, Paliwal S, Sharma S, Kesar S, Chauhan N, Madan M. Metabolic Syndrome: The Constellation of Co-morbidities, A Global Threat. *EMIDDT*. 2023;23(12):1491-1504. doi:10.2174/1871530323666230309144825

4. Samson SL, Garber AJ. Metabolic Syndrome. *Endocrinology and Metabolism Clinics of North America*. 2014;43(1):1-23. doi:10.1016/j.ecl.2013.09.009

5. Saltiel AR, Olefsky JM. Inflammatory mechanisms linking obesity and metabolic disease. *Journal of Clinical Investigation*. 2017;127(1):1-4. doi:10.1172/JCI92035

6. Smith RL, Soeters MR, Wüst RCI, Houtkooper RH. Metabolic Flexibility as an Adaptation to Energy Resources and Requirements in Health and Disease. *Endocr Rev*. 2018;39(4):489-517. doi:10.1210/er.2017-00211

7. Chouchani ET, Kajimura S. Metabolic adaptation and maladaptation in adipose tissue. *Nat Metab*. 2019;1(2):189-200. doi:10.1038/s42255-018-0021-8

8. Saadati S, Godini R, Reddy A, Teede H, Mousa A. Metabolic crossroads in insulin resistance: exploring lipid dysregulation and inflammation. *Front Immunol*. 16:1692742. doi:10.3389/fimmu.2025.1692742

9. Uehara K, Santoleri D, Whitlock AEG, Titchenell PM. Insulin Regulation of Hepatic Lipid Homeostasis. In: Prakash YS, ed. *Comprehensive Physiology*. 1st ed. Wiley; 2023:4785-4809. doi:10.1002/cphy.c220015

10. Rinella ME, Lazarus JV, Ratziu V, et al. A multisociety Delphi consensus statement on new fatty liver disease nomenclature. *Journal of Hepatology*. 2023;79(6):1542-1556. doi:10.1016/j.jhep.2023.06.003

11. Jurek JM, Zablocka-Sowinska K, Clavero Mestres H, Reyes Gutiérrez L, Camaron J, Auguet T. The Impact of Dietary Interventions on Metabolic Outcomes in Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) and Comorbid Conditions, Including Obesity and Type 2 Diabetes. *Nutrients*. 2025;17(7):1257. doi:10.3390/nu17071257

12. Byrne CD, Armandi A, Pellegrinelli V, Vidal-Puig A, Bugianesi E. Metabolic dysfunction-associated steatotic liver disease: a condition of heterogeneous metabolic

risk factors, mechanisms and comorbidities requiring holistic treatment. *Nat Rev Gastroenterol Hepatol*. 2025;22(5):314-328. doi:10.1038/s41575-025-01045-z

13. Tacke F, Horn P, Wai-Sun Wong V, et al. EASL–EASD–EASO Clinical Practice Guidelines on the management of metabolic dysfunction-associated steatotic liver disease (MASLD). *Journal of Hepatology*. 2024;81(3):492-542. doi:10.1016/j.jhep.2024.04.031

14. Pîrvu BF, Clenciu D, Beldie LA, et al. The burden of cancer in metabolic dysfunction-associated steatotic liver disease. *Rom J Morphol Embryol*. 2024;65(4):627-635. doi:10.47162/RJME.65.4.09

15. Lin J, Rao D, Zhang M, Gao Q. Metabolic reprogramming in the tumor microenvironment of liver cancer. *J Hematol Oncol*. 2024;17(1):6. doi:10.1186/s13045-024-01527-8

16. Gyamfi J, Kim J, Choi J. Cancer as a Metabolic Disorder. *IJMS*. 2022;23(3):1155. doi:10.3390/ijms23031155

17. Liu X, Ju W, Huo C, Zhang S, Wang X, Huang K. Overweight and Obesity as Independent Factors for Increased Risk of Hepatocellular Cancer-Related Mortality: A Meta-Analysis. *Journal of the American College of Nutrition*. 2021;40(3):287-293. doi:10.1080/07315724.2020.1751007

18. Kwak M, Kim H seok, Jiang ZG, et al. MASLD/MetALD and mortality in individuals with any cardio-metabolic risk factor: A population-based study with 26.7 years of follow-up. *Hepatology*. 2025;81(1):228-237. doi:10.1097/HEP.0000000000000925

19. Shi S, Yi H, Zheng Y, Zhao Y, Yu D. Adipose distribution patterns as prognostic factors in patients with HCC: A systematic review and meta-analysis. *European Journal of Radiology*. 2023;167:111025. doi:10.1016/j.ejrad.2023.111025

20. Wang S, Cheng L, Dou L, et al. Geriatric nutritional risk index and body composition dictate the prognosis of elderly patients with intrahepatic cholangiocarcinoma. *Front Nutr*. 2025;12:1565317. doi:10.3389/fnut.2025.1565317

21. Li W, Liu T, Qian L, et al. Insulin resistance and inflammation mediate the association of abdominal obesity with colorectal cancer risk. *Front Endocrinol*. 2022;13:983160. doi:10.3389/fendo.2022.983160

22. Jonas W, Schürmann A. Genetic and epigenetic factors determining NAFLD risk. *Molecular Metabolism*. 2021;50:101111. doi:10.1016/j.molmet.2020.101111

23. Loomba R, Wong VW. Implications of the new nomenclature of steatotic liver disease and definition of metabolic dysfunction-associated steatotic liver disease. *Aliment Pharmacol Ther*. 2024;59(2):150-156. doi:10.1111/apt.17846

24. Ali SMJ, Lai M. Metabolic Dysfunction–Associated Steatotic Liver Disease. *Ann Intern Med*. 2025;178(1):ITC1-ITC16. doi:10.7326/ANNALS-24-02933

25. Feng G, Targher G, Byrne CD, et al. Global burden of metabolic dysfunction-associated steatotic liver disease, 2010 to 2021. *JHEP Rep*. 2024;7(3):101271. doi:10.1016/j.jhepr.2024.101271

26. Younossi ZM, Kalligeros M, Henry L. Epidemiology of metabolic dysfunction-associated steatotic liver disease. *Clin Mol Hepatol*. 2025;31(Suppl):S32-S50. doi:10.3350/cmh.2024.0431

27. Younossi ZM, Golabi P, Paik JM, Henry A, Van Dongen C, Henry L. The global epidemiology of nonalcoholic fatty liver disease (NAFLD) and nonalcoholic steatohepatitis (NASH): a systematic review. *Hepatology*. 2023;77(4):1335-1347. doi:10.1097/HEP.0000000000000004

28. Danpanichkul P, Suparan K, Diaz LA, et al. The Rising Global Burden of MASLD and Other Metabolic Diseases (2000–2021). *UEG Journal*. 2025;13(7):1141-1154. doi:10.1002/ueg2.70072

29. Miller DM, McCauley KF, Dunham-Snary KJ. Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD): Mechanisms, Clinical Implications and Therapeutic Advances. *Endocrino Diabet & Metabol*. 2025;8(6):e70132. doi:10.1002/edm2.70132

30. Lekakis V, Papatheodoridis GV. Natural history of metabolic dysfunction-associated steatotic liver disease. *European Journal of Internal Medicine*. 2024;122:3-10. doi:10.1016/j.ejim.2023.11.005

31. Lee HH, Lee HA, Kim EJ, et al. Metabolic dysfunction-associated steatotic liver disease and risk of cardiovascular disease. *Gut*. 2024;73(3):533-540. doi:10.1136/gutjnl-2023-331003

32. Kim DW, Son M, Lee HJ, et al. Chronic Kidney Disease Risk Associated With Metabolic Dysfunction-Associated Steatotic Liver Disease: A Nationwide Cohort Study in Korea. *Hepatology Research*. 2025;55(9):1239-1250. doi:10.1111/hepr.14226

33. Hagström H, Shang Y, Hegmar H, Nasr P. Natural history and progression of metabolic dysfunction-associated steatotic liver disease. *The Lancet Gastroenterology & Hepatology*. 2024;9(10):944-956. doi:10.1016/S2468-1253(24)00193-6

34. Park H, Cheuk-Fung Yip T, Yoon EL, et al. Impact of cardiometabolic risk factors on hepatic fibrosis and clinical outcomes in MASLD: A population-based multi-cohort study. *JHEP Rep*. 2025;7(6):101388. doi:10.1016/j.jhepr.2025.101388

35. Kirik A, Sumbul HE, Koca N, et al. Prevalence of MASLD and Fibrosis Risk in Turkish Adults with Cardiometabolic Risk Factors: A Nationwide Multicenter Study (DAHUDER MASLD Study). *J Clin Med*. 2025;14(19):7098. doi:10.3390/jcm14197098

36. Ochoa-Allemant P, Hubbard RA, Kaplan DE, Serper M. Adverse Liver Outcomes, Cardiovascular Events, and Mortality in Steatotic Liver Disease. *JAMA Intern Med*. 2025;185(8):986. doi:10.1001/jamainternmed.2025.1809

37. Minetti ET, Hamburg NM, Matsui R. Drivers of cardiovascular disease in metabolic dysfunction-associated steatotic liver disease: the threats of oxidative stress. *Front Cardiovasc Med*. 2024;11:1469492. doi:10.3389/fcvm.2024.1469492

38. Su J, Guan H, Fan X, et al. Associations of serum aminotransferase and the risk of all-cause and cause-specific mortality in Chinese type 2 diabetes: a community-based cohort study. *BMJ Open*. 2023;13(7):e068160. doi:10.1136/bmjopen-2022-068160

39. Bae JC. DPP-4 Inhibitor in Type 2 Diabetes Mellitus Patient with Non-Alcoholic Fatty Liver Disease: Achieving Two Goals at Once? *Endocrinol Metab*. 2022;37(6):858-860. doi:10.3803/EnM.2022.605

40. Rodriguez LA, Schmittdiel JA, Liu L, et al. Hepatocellular Carcinoma in Metabolic Dysfunction-Associated Steatotic Liver Disease. *JAMA Netw Open*. 2024;7(7):e2421019. doi:10.1001/jamanetworkopen.2024.21019

41. de Mattos AZ, Bombassaro IZ, Vogel A, Debes JD. Hepatocellular carcinoma-the role of the underlying liver disease in clinical practice. *World J Gastroenterol*. 2024;30(19):2488-2495. doi:10.3748/wjg.v30.i19.2488

42. Sarkar S, Alurwar A, Ly C, et al. A Machine Learning Model to Predict Risk for Hepatocellular Carcinoma in Patients With Metabolic Dysfunction-Associated Steatotic Liver Disease. *Gastro Hep Advances*. 2024;3(4):498-505. doi:10.1016/j.gastha.2024.01.007

43. Mi K, Ye T, Zhu L, Pan CQ. Risk-stratified hepatocellular carcinoma surveillance in non-cirrhotic patients with MASLD. *Gastroenterology Report*. 2025;13:goaf018. doi:10.1093/gastro/goaf018

44. Archer AJ, Belfield KJ, Orr JG, Gordon FH, Abeysekera KW. EASL clinical practice guidelines: non-invasive liver tests for evaluation of liver disease severity and prognosis. *Frontline Gastroenterol*. 2022;13(5):436-439. doi:10.1136/flgastro-2021-102064

45. Jin X, Yip TCF, Wong GLH, Wong VWS, Lai JCT. The new definition of metabolic dysfunction-associated steatotic liver disease: the role of ultrasound and elastography. *Ultrasonography*. 2025;44(3):189-201. doi:10.14366/usg.24219

46. McPherson S, Hardy T, Dufour JF, et al. Age as a Confounding Factor for the Accurate Non-Invasive Diagnosis of Advanced NAFLD Fibrosis. *American Journal of Gastroenterology*. 2017;112(5):740-751. doi:10.1038/ajg.2016.453

47. Vallet-Pichard A, Mallet V, Nalpas B, et al. FIB-4: An inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology*. 2007;46(1):32-36. doi:10.1002/hep.21669

48. Kwok R, Tse Y -K., Wong GL -H., et al. Systematic review with meta-analysis: non-invasive assessment of non-alcoholic fatty liver disease – the role of transient elastography and plasma cytokeratin-18 fragments. *Aliment Pharmacol Ther*. 2014;39(3):254-269. doi:10.1111/apt.12569

49. Castera L, Friedrich-Rust M, Loomba R. Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2019;156(5):1264-1281.e4. doi:10.1053/j.gastro.2018.12.036

50. Ayares G, Diaz LA, Idalsoaga F, et al. METALD : New Perspectives on an Old Overlooked Disease. *Liver International*. 2025;45(5):e70017. doi:10.1111/liv.70017

51. Arab JP, Díaz LA, Rehm J, et al. Metabolic dysfunction and alcohol-related liver disease (MetALD): Position statement by an expert panel on alcohol-related liver disease. *J Hepatol*. 2025;82(4):744-756. doi:10.1016/j.jhep.2024.11.028

52. Kim GA, Jeong S, Jang H, Lee DH, Joo SK, Kim W. Metabolic Dysfunction-Associated Steatotic Liver Disease and Metabolic Dysfunction-Associated Steatotic Liver Disease with Increased Alcohol Intake Increase the Risk of Developing Hepatocellular Carcinoma and Incident or Decompensated Cirrhosis: A Korean Nationwide Study. *Liver Cancer*. 2023;13(4):426-437. doi:10.1159/000535943

53. Chen YT, Chen TI, Yang TH, et al. Long-term Risks of Cirrhosis and Hepatocellular Carcinoma Across Steatotic Liver Disease Subtypes. *Am J Gastroenterol*. 2024;119(11):2241-2250. doi:10.14309/ajg.0000000000002778

54. Celsa C, Pennisi G, Tulone A, et al. Risk of hepatic and extrahepatic outcomes associated with metabolic dysfunction-associated steatotic liver disease and metabolic dysfunction and alcohol-associated steatotic liver disease: a systematic review and meta-analysis. *The Lancet Gastroenterology & Hepatology*. 2025;10(11):998-1012. doi:10.1016/S2468-1253(25)00188-8

55. Canivet CM, Boursier J, Loomba R. New Nomenclature for Nonalcoholic Fatty Liver Disease: Understanding Metabolic Dysfunction-Associated Steatotic Liver Disease, Metabolic Dysfunction- and Alcohol-Associated Liver Disease, and Their Implications in Clinical Practice. *Semin Liver Dis*. 2024;44(01):035-042. doi:10.1055/s-0044-1785196

56. Davis CG, Thake J, Vilhena N. Social desirability biases in self-reported alcohol consumption and harms. *Addictive Behaviors*. 2010;35(4):302-311. doi:10.1016/j.addbeh.2009.11.001

57. Thursz M, Gual A, Lackner C, et al. EASL Clinical Practice Guidelines: Management of alcohol-related liver disease. *Journal of Hepatology*. 2018;69(1):154-181. doi:10.1016/j.jhep.2018.03.018

58. Acuña P, Albhaisi O, Idalsoaga F, et al. Research and clinical trial design in patients with metabolic dysfunction and alcohol-associated steatotic liver disease

(MetALD): a narrative review. *Metab Target Organ Damage*. 2025;5(4). doi:10.20517/mtod.2025.61

59. Staufer K, Huber-Schönauer U, Strebinger G, et al. Ethyl glucuronide in hair detects a high rate of harmful alcohol consumption in presumed non-alcoholic fatty liver disease. *Journal of Hepatology*. 2022;77(4):918-930. doi:10.1016/j.jhep.2022.04.040

60. Harris JC, Leggio L, Farokhnia M. Blood Biomarkers of Alcohol Use: A Scoping Review. *Curr Addict Rep*. 2021;8(4):500-508. doi:10.1007/s40429-021-00402-7

61. Torp N, Israelsen M, Thiele M, Rinella ME, Krag A. Phosphatidylethanol in steatotic liver disease. *Journal of Hepatology*. 2025;83(5):1189-1203. doi:10.1016/j.jhep.2025.07.019

62. Tavaglione F, Díaz LA, Ajmera V, et al. Clinical utility of phosphatidylethanol to detect underreported alcohol use and enhance steatotic liver disease subclassification. *Journal of Hepatology*. 2025;83(5):1023-1034. doi:10.1016/j.jhep.2025.05.030

63. Perilli M, Toselli F, Franceschetto L, et al. Phosphatidylethanol (PEth) in Blood as a Marker of Unhealthy Alcohol Use: A Systematic Review with Novel Molecular Insights. *Int J Mol Sci*. 2023;24(15):12175. doi:10.3390/ijms241512175

64. Hahn JA, Murnane PM, Vittinghoff E, et al. Factors associated with phosphatidylethanol (PEth) sensitivity for detecting unhealthy alcohol use: An individual patient data meta-analysis. *Alcohol Clin Exp Res*. 2021;45(6):1166-1187. doi:10.1111/acer.14611

65. Rinella ME, Neuschwander-Tetri BA, Siddiqui MS, et al. AASLD Practice Guidance on the clinical assessment and management of nonalcoholic fatty liver disease. *Hepatology*. 2023;77(5):1797-1835. doi:10.1097/HEP.0000000000000323

66. Long MT, Noureddin M, Lim JK. AGA Clinical Practice Update: Diagnosis and Management of Non-Alcoholic Fatty Liver Disease (NAFLD) in Lean Individuals: Expert Review. *Gastroenterology*. 2022;163(3):764-774.e1. doi:10.1053/j.gastro.2022.06.023

67. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J Clinicians*. 2024;74(3):229-263. doi:10.3322/caac.21834

68. Zhang H, Yang T, Wu M, Shen F. Intrahepatic cholangiocarcinoma: Epidemiology, risk factors, diagnosis and surgical management. *Cancer Letters*. 2016;379(2):198-205. doi:10.1016/j.canlet.2015.09.008

69. Qurashi M, Vithayathil M, Khan SA. Epidemiology of cholangiocarcinoma. *European Journal of Surgical Oncology*. 2025;51(2):107064. doi:10.1016/j.ejso.2023.107064

70. Bertuccio P, Malvezzi M, Carioli G, et al. Global trends in mortality from intrahepatic and extrahepatic cholangiocarcinoma. *Journal of Hepatology*. 2019;71(1):104-114. doi:10.1016/j.jhep.2019.03.013

71. Vithayathil M, Khan SA. Current epidemiology of cholangiocarcinoma in Western countries. *Journal of Hepatology*. 2022;77(6):1690-1698. doi:10.1016/j.jhep.2022.07.022

72. Crosby D, Bhatia S, Brindle KM, et al. Early detection of cancer. *Science*. 2022;375(6586):eaay9040. doi:10.1126/science.aay9040

73. El-Sayes N, Vito A, Mossman K. Tumor Heterogeneity: A Great Barrier in the Age of Cancer Immunotherapy. *Cancers (Basel)*. 2021;13(4):806. doi:10.3390/cancers13040806

74. Alexandrov LB, Kim J, Haradhvala NJ, et al. The repertoire of mutational signatures in human cancer. *Nature*. 2020;578(7793):94-101. doi:10.1038/s41586-020-1943-3

75. Luo X, Yin S, Yang L, et al. Development and Validation of a Pathology Image Analysis-based Predictive Model for Lung Adenocarcinoma Prognosis - A Multi-cohort Study. *Sci Rep*. 2019;9(1):6886. doi:10.1038/s41598-019-42845-z

76. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869-10874. doi:10.1073/pnas.191367098

77. Loud JT, Murphy J. Cancer Screening and Early Detection in the 21st Century. *Seminars in Oncology Nursing*. 2017;33(2):121-128. doi:10.1016/j.soncn.2017.02.002

78. Miller BF, Petrykowska HM, Elnitski L. Assessing ZNF154 methylation in patient plasma as a multicancer marker in liquid biopsies from colon, liver, ovarian and pancreatic cancer patients. *Sci Rep*. 2021;11(1):221. doi:10.1038/s41598-020-80345-7

79. Ilyas SI, Khan SA, Hallemeier CL, Kelley RK, Gores GJ. Cholangiocarcinoma — evolving concepts and therapeutic strategies. *Nat Rev Clin Oncol*. 2018;15(2):95-111. doi:10.1038/nrclinonc.2017.157

80. Zeng D, Wang S, Cheng N, Liu G, Li B. Survival Outcomes of Preoperative Serum Biomarkers in Patients With Surgically Treated Intrahepatic Cholangiocarcinoma. *Clin Transl Gastroenterol*. 2025;16(6):e00845. doi:10.14309/ctg.0000000000000845

81. Moris D, Palta M, Kim C, Allen PJ, Morse MA, Lidsky ME. Advances in the treatment of intrahepatic cholangiocarcinoma: An overview of the current and future therapeutic landscape for clinicians. *CA A Cancer J Clinicians*. 2023;73(2):198-222. doi:10.3322/caac.21759

82. Bekki Y, Von Ahrens D, Takahashi H, Schwartz M, Gunasekaran G. Recurrent Intrahepatic Cholangiocarcinoma – Review. *Front Oncol*. 2021;11:776863. doi:10.3389/fonc.2021.776863

83. Ma L, Guo H, Zhao Y, et al. Liquid biopsy in cancer: current status, challenges and future prospects. *Signal Transduct Target Ther*. 2024;9:336. doi:10.1038/s41392-024-02021-w

84. Paul D, Nedelcu AM. The underexplored links between cancer and the internal body climate: Implications for cancer prevention and treatment. *Front Oncol*. 2022;12:1040034. doi:10.3389/fonc.2022.1040034

85. Hanahan D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery*. 2022;12(1):31-46. doi:10.1158/2159-8290.CD-21-1059

86. Altea-Manzano P, Decker-Farrell A, Janowitz T, Erez A. Metabolic interplays between the tumor and the host shape the tumor macroenvironment. *Nat Rev Cancer*. 2025;25(4):274-292. doi:10.1038/s41568-024-00786-4

87. Xu H, Cao D, He A, Ge W. The prognostic role of obesity is independent of sex in cancer patients treated with immune checkpoint inhibitors: A pooled analysis of 4090 cancer patients. *International Immunopharmacology*. 2019;74:105745. doi:10.1016/j.intimp.2019.105745

88. Bea JW, Ochs-Balcom HM, Valencia CI, et al. Abdominal visceral and subcutaneous adipose tissue associations with postmenopausal breast cancer incidence. *JNCI Cancer Spectr*. 2025;9(1):pkaf007. doi:10.1093/jncics/pkaf007

89. Gu XY, Yang JL, Lai R, et al. Impact of lactate on immune cell function in the tumor microenvironment: mechanisms and therapeutic perspectives. *Front Immunol*. 2025;16:1563303. doi:10.3389/fimmu.2025.1563303

90. Labib PL, Goodchild G, Pereira SP. Molecular Pathogenesis of Cholangiocarcinoma. *BMC Cancer*. 2019;19:185. doi:10.1186/s12885-019-5391-0

91. Thiele M, Villesen IF, Niu L, et al. Opportunities and barriers in omics-based biomarker discovery for steatotic liver diseases. *Journal of Hepatology*. 2024;81(2):345-359. doi:10.1016/j.jhep.2024.03.035

92. Agache I, Akdis CA. Precision medicine and phenotypes, endotypes, genotypes, regiotypes, and theratypes of allergic diseases. *Journal of Clinical Investigation*. 2019;129(4):1493-1503. doi:10.1172/JCI124611

93. Lim AMW, Lim EU, Chen PL, Fann CSJ. Unsupervised clustering identified clinically relevant metabolic syndrome endotypes in UK and Taiwan Biobanks. *iScience*. 2024;27(7):109815. doi:10.1016/j.isci.2024.109815

94. Karsdal MA, Hallsworth K, Scragg J, et al. Serum levels of fibrogenesis biomarkers reveal distinct endotypes predictive of response to weight loss in advanced

nonalcoholic fatty liver disease. *Hepatol Commun*. 2023;7(10):e0254. doi:10.1097/HC9.0000000000000254

95. Targher G, Byrne CD, Tilg H. MASLD: a systemic metabolic disorder with cardiovascular and malignant complications. *Gut*. 2024;73(4):691-702. doi:10.1136/gutjnl-2023-330595

96. Pipitone RM, Ciccioli C, Infantino G, et al. MAFLD: a multisystem disease. *Ther Adv Endocrinol Metab*. 2023;14:20420188221145549. doi:10.1177/20420188221145549

97. Wilcox NS, Amit U, Reibel JB, Berlin E, Howell K, Ky B. Cardiovascular disease and cancer: shared risk factors and mechanisms. *Nat Rev Cardiol*. 2024;21(9):617-631. doi:10.1038/s41569-024-01017-x

98. Beger RD, Dunn W, Schmidt MA, et al. Metabolomics enables precision medicine: "A White Paper, Community Perspective." *Metabolomics*. 2016;12(9):149. doi:10.1007/s11306-016-1094-6

99. Hornburg D, Wu S, Moqri M, et al. Dynamic lipidome alterations associated with human health, disease and ageing. *Nat Metab*. 2023;5(9):1578-1594. doi:10.1038/s42255-023-00880-1

100. Tan EY, Muthiah MD, Sanyal AJ. Metabolomics at the cutting edge of risk prediction of MASLD. *Cell Rep Med*. 2024;5(12):101853. doi:10.1016/j.xcrm.2024.101853

101. Schneider KM, Cao F, Huang HYR, et al. The Lipidomic Profile Discriminates Between MASLD and MetALD. *Aliment Pharmacol Ther*. 2025;61(8):1357-1371. doi:10.1111/apt.70012

102. Sweatt K, Garvey WT, Martins C. Strengths and Limitations of BMI in the Diagnosis of Obesity: What is the Path Forward? *Curr Obes Rep*. 2024;13(3):584-595. doi:10.1007/s13679-024-00580-1

103. Kim K, Park SM. Association of muscle mass and fat mass with insulin resistance and the prevalence of metabolic syndrome in Korean adults: a cross-sectional study. *Sci Rep*. 2018;8:2703. doi:10.1038/s41598-018-21168-5

104. Christakoudi S, Tsilidis KK, Evangelou E, Riboli E. Association of body-shape phenotypes with imaging measures of body composition in the UK Biobank cohort: relevance to colon cancer risk. *BMC Cancer*. 2021;21:1106. doi:10.1186/s12885-021-08820-6

105. Lu Y, Zhao YC, Liu K, et al. A validated estimate of visceral adipose tissue volume in relation to cancer risk. *J Natl Cancer Inst*. 2024;116(12):1942-1951. doi:10.1093/jnci/djae193

106. Tarabeih N, Kalinkovich A, Ashkenazi S, Cherny SS, Shalata A, Livshits G. Relationships between Circulating Biomarkers and Body Composition Parameters in Patients with Metabolic Syndrome: A Community-Based Study. *Int J Mol Sci*. 2024;25(2):881. doi:10.3390/ijms25020881

107. Goossens GH. The Metabolic Phenotype in Obesity: Fat Mass, Body Fat Distribution, and Adipose Tissue Function. *Obes Facts*. 2017;10(3):207-215. doi:10.1159/000471488

108. Nagana Gowda GA, Raftery D. NMR-Based Metabolomics. In: Hu S, ed. *Cancer Metabolomics*. Vol 1280. Advances in Experimental Medicine and Biology. Springer International Publishing; 2021:19-37. doi:10.1007/978-3-030-51652-9_2

109. Jin T, Wu Y, Zhang S, et al. The association between metabolomic profiles of lifestyle and the latent phase of incident chronic kidney disease in the UK Population. *Sci Rep*. 2025;15(1):2299. doi:10.1038/s41598-025-86030-x

110. Yang S, Liu R, Xin Z, et al. Plasma Metabolomics Identifies Key Metabolites and Improves Prediction of Diabetic Retinopathy. *Ophthalmology*. 2024;131(12):1436-1446. doi:10.1016/j.ophtha.2024.07.004

111. Zhang S, Wang Z, Wang Y, et al. A metabolomic profile of biological aging in 250,341 individuals from the UK Biobank. *Nat Commun*. 2024;15(1):8081. doi:10.1038/s41467-024-52310-9

112. Julkunen H, Cichońska A, Tiainen M, et al. Atlas of plasma NMR biomarkers for health and disease in 118,461 individuals from the UK Biobank. *Nat Commun*. 2023;14(1):604. doi:10.1038/s41467-023-36231-7

113. Moreno-Vedia J, Rosales R, Ozcariz E, et al. Triglyceride-Rich Lipoproteins and Glycoprotein A and B Assessed by 1H-NMR in Metabolic-Associated Fatty Liver Disease. *Front Endocrinol*. 2022;12:775677. doi:10.3389/fendo.2021.775677

114. Heeren J, Scheja L. Metabolic-associated fatty liver disease and lipoprotein metabolism. *Molecular Metabolism*. 2021;50:101238. doi:10.1016/j.molmet.2021.101238

115. Schneider CV, Hehl L, Creasy KT, et al. A coding variant in the microsomal triglyceride transfer protein reduces both hepatic steatosis and plasma lipids. *Aliment Pharmacol Ther*. 2023;58(2):238-249. doi:10.1111/apt.17566

116. Liang J, Liang Y, Chen L, et al. Cardiovascular-kidney-metabolic progression associated with major adverse liver outcomes: mediating roles of plasma metabolites. *Front Nutr*. 2025;12:1675899. doi:10.3389/fnut.2025.1675899

117. Dierckx T, Verstockt B, Vermeire S, Van Weyenbergh J. GlycA, a Nuclear Magnetic Resonance Spectroscopy Measure for Protein Glycosylation, is a Viable Biomarker for Disease Activity in IBD. *Journal of Crohn's and Colitis*. 2019;13(3):389-394. doi:10.1093/ecco-jcc/jjy162

118. Smilowitz JT, Wiest MM, Watkins SM, et al. Lipid Metabolism Predicts Changes in Body Composition during Energy Restriction in Overweight Humans. *The Journal of Nutrition*. 2009;139(2):222-229. doi:10.3945/jn.108.095364

119. Larsson SC, Spyrou N, Mantzoros CS. Body fatness associations with cancer: evidence from recent epidemiological studies and future directions. *Metabolism*. 2022;137:155326. doi:10.1016/j.metabol.2022.155326

120. Tran TXM, Chang Y, Choi HR, et al. Adiposity, Body Composition Measures, and Breast Cancer Risk in Korean Premenopausal Women. *JAMA Netw Open*. 2024;7(4):e245423. doi:10.1001/jamanetworkopen.2024.5423

121. Liu C, Zhang Q, Liu T, et al. Predicted lean body mass trajectories, and cancer risk and cancer-specific and all-cause mortality: A prospective cohort study. *J Cachexia Sarcopenia Muscle*. 2023;14(6):2916-2924. doi:10.1002/jcsm.13370

122. Wu Y, Li X, Li Q, Cheng C, Zheng L. Adipose tissue-to-breast cancer crosstalk: Comprehensive insights. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*. 2022;1877(5):188800. doi:10.1016/j.bbcan.2022.188800

123. Ortega LS, Bradbury KE, Cross AJ, Morris JS, Gunter MJ, Murphy N. A Prospective Investigation of Body Size, Body Fat Composition and Colorectal Cancer Risk in the UK Biobank. *Sci Rep*. 2017;7:17807. doi:10.1038/s41598-017-17997-5

124. Srilatha M, Malla R, Adem MP, Foote JB, Nagaraju GP. Obesity associated pancreatic ductal adenocarcinoma: Therapeutic challenges. *Seminars in Cancer Biology*. 2023;97:12-20. doi:10.1016/j.semcancer.2023.11.002

125. Liu A, He Q, Wu W, et al. Body composition and risk of gastric cancer: A population-based prospective cohort study. *Cancer Med*. 2021;10(6):2164-2174. doi:10.1002/cam4.3808

126. Lauby-Secretan B, Scoccianti C, Loomis D, Grosse Y, Bianchini F, Straif K. Body Fatness and Cancer — Viewpoint of the IARC Working Group. *N Engl J Med*. 2016;375(8):794-798. doi:10.1056/NEJMsr1606602

127. Sanchez A, Kissel S, Coletta A, Scott J, Furberg H. Impact of body size and body composition on bladder cancer outcomes: risk stratification and opportunity for novel interventions. *Urol Oncol*. 2020;38(9):713-718. doi:10.1016/j.urolonc.2020.03.017

128. Fang P, Zhou J, Xiao X, et al. The prognostic value of sarcopenia in oesophageal cancer: A systematic review and meta-analysis. *J Cachexia Sarcopenia Muscle*. 2022;14(1):3-16. doi:10.1002/jcsm.13126

129. Moreira-Pais A, Ferreira R, Oliveira PA, Duarte JA. Sarcopenia versus cancer cachexia: the muscle wasting continuum in healthy and diseased aging. *Biogerontology*. 2021;22(5):459-477. doi:10.1007/s10522-021-09932-z

130. Watanabe J, Matsui R, Sasanuma H, et al. Body composition assessment and sarcopenia in patients with biliary tract cancer: A systematic review and meta-analysis. *Clinical Nutrition*. 2022;41(2):321-328. doi:10.1016/j.clnu.2021.12.005

131. Zhang L, Wang K, Liu R, et al. Body composition as a prognostic factor in cholangiocarcinoma: a meta-analysis. *Nutr J*. 2024;23(1):145. doi:10.1186/s12937-024-01037-w

132. Sweatt K, Garvey WT, Martins C. Strengths and Limitations of BMI in the Diagnosis of Obesity: What is the Path Forward? *Curr Obes Rep*. 2024;13(3):584-595. doi:10.1007/s13679-024-00580-1

133. Cichosz SL, Rasmussen NH, Vestergaard P, Hejlesen O. Is predicted body-composition and relative fat mass an alternative to body-mass index and waist circumference for disease risk estimation? *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*. 2022;16(9):102590. doi:10.1016/j.dsx.2022.102590

134. Koh K, Scott R, Cespedes Feliciano EM, et al. Cancer-Associated Cachexia: Bridging Clinical Findings with Mechanistic Insights in Human Studies. *Cancer Discov*. 2025;15(8):1543-1568. doi:10.1158/2159-8290.CD-25-0293

135. Bates DDB, Pickhardt PJ. CT-Derived Body Composition Assessment as a Prognostic Tool in Oncologic Patients: From Opportunistic Research to Artificial Intelligence–Based Clinical Implementation. *American Journal of Roentgenology*. 2022;219(4):671-680. doi:10.2214/AJR.22.27749

136. Stefan N. Causes, consequences, and treatment of metabolically unhealthy fat distribution. *The Lancet Diabetes & Endocrinology*. 2020;8(7):616-627. doi:10.1016/S2213-8587(20)30110-8

137. Tchkonia T, Thomou T, Zhu Y, et al. Mechanisms and Metabolic Implications of Regional Differences among Fat Depots. *Cell Metabolism*. 2013;17(5):644-656. doi:10.1016/j.cmet.2013.03.008

138. Best TD, Mercaldo SF, Bryan DS, et al. Multilevel Body Composition Analysis on Chest Computed Tomography Predicts Hospital Length of Stay and Complications After Lobectomy for Lung Cancer: A Multicenter Study. *Annals of Surgery*. 2022;275(5):e708-e715. doi:10.1097/SLA.0000000000004040

139. Wan B, Ma N, Lu W. Mendelian randomization investigation identified the causal relationship between body fat indexes and the risk of bladder cancer. *PeerJ*. 2023;11:e14739. doi:10.7717/peerj.14739

140. Nwosu AC, Mayland CR, Mason S, et al. Bioelectrical impedance vector analysis (BIVA) as a method to compare body composition differences according to cancer stage and type. *Clinical Nutrition ESPEN*. 2019;30:59-66. doi:10.1016/j.clnesp.2019.02.006

141. Liang C, Lin JY, Liao LH, et al. Plasma metabolic landscape unveils key regulators of leukemia subtype progression. *Future Science OA*. 2025;11(1):2527015. doi:10.1080/20565623.2025.2527015

142. Sun Y, Zhang X, Hang D, et al. Integrative plasma and fecal metabolomics identify functional metabolites in adenoma-colorectal cancer progression and as early diagnostic biomarkers. *Cancer Cell*. 2024;42(8):1386-1400.e8. doi:10.1016/j.ccell.2024.07.005

143. Mayers JR, Wu C, Clish CB, et al. Elevated circulating branched chain amino acids are an early event in pancreatic adenocarcinoma development. *Nat Med*. 2014;20(10):1193-1198. doi:10.1038/nm.3686

144. Dang L, White DW, Gross S, et al. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*. 2009;462(7274):739-744. doi:10.1038/nature08617

145. Choi C, Ganji SK, DeBerardinis RJ, et al. 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. *Nat Med*. 2012;18(4):624-629. doi:10.1038/nm.2682

146. Zhu AX, Macarulla T, Javle MM, et al. Final Overall Survival Efficacy Results of Ivosidenib for Patients With Advanced Cholangiocarcinoma With *IDH1* Mutation: The Phase 3 Randomized Clinical ClarIDHy Trial. *JAMA Oncol*. 2021;7(11):1669. doi:10.1001/jamaoncol.2021.3836

147. Frampton JE. Ivosidenib: A Review in Advanced Cholangiocarcinoma. *Targ Oncol*. 2023;18(6):973-980. doi:10.1007/s11523-023-01002-3

148. Duffy MJ. Carcinoembryonic Antigen as a Marker for Colorectal Cancer: Is It Clinically Useful? *Clinical Chemistry*. 2001;47(4):624-630. doi:10.1093/clinchem/47.4.624

149. Chen J, Wang H, Zhou L, Liu Z, Tan X. A combination of circulating tumor cells and CA199 improves the diagnosis of pancreatic cancer. *Clinical Laboratory Analysis*. 2022;36(5):e24341. doi:10.1002/jcla.24341

150. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018;359(6378):926-930. doi:10.1126/science.aar3247

151. Mertins P, Mani DR, Ruggles KV, et al. Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature*. 2016;534(7605):55-62. doi:10.1038/nature18003

152. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med*. 2015;12(3):e1001779. doi:10.1371/journal.pmed.1001779

153. Kurki MI, Karjalainen J, Palta P, et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature*. 2023;613(7944):508-518. doi:10.1038/s41586-022-05473-8

154. Feng Q, Bešević J, Conroy M, Omiyale W, Lacey B, Allen N. Comparison of body composition measures assessed by bioelectrical impedance analysis versus dual-energy X-ray absorptiometry in the United Kingdom Biobank. *Clinical Nutrition ESPEN*. 2024;63:214-225. doi:10.1016/j.clnesp.2024.06.040

155. Liu Z, Suo C, Fan H, Zhang T, Jin L, Chen X. Dissecting causal relationships between nonalcoholic fatty liver disease proxied by chronically elevated alanine transaminase levels and 34 extrahepatic diseases. *Metabolism*. 2022;135:155270. doi:10.1016/j.metabol.2022.155270

156. Ding H, Gao Y, Gao Y, et al. Exploring novel molecular mechanisms underlying recurrent pregnancy loss in decidual tissues. *Sci Rep*. 2025;15(1):25460. doi:10.1038/s41598-025-10604-y

157. Xu Y, Park Y, Park JD, Sun B. Predicting Nurse Turnover for Highly Imbalanced Data Using the Synthetic Minority Over-Sampling Technique and Machine Learning Algorithms. *Healthcare*. 2023;11(24):3173. doi:10.3390/healthcare11243173

158. Argelaguet R, Velten B, Arnol D, et al. Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol Syst Biol*. 2018;14(6):e8124. doi:10.15252/msb.20178124

159. Argelaguet R, Arnol D, Bredikhin D, et al. MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol*. 2020;21:111. doi:10.1186/s13059-020-02015-1

160. Pang Z, Lu Y, Zhou G, et al. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Research*. 2024;52(W1):W398-W406. doi:10.1093/nar/gkae253

161. Misceo D, Mocciaro G, D'Amore S, Vacca M. Diverting hepatic lipid fluxes with lifestyles revision and pharmacological interventions as a strategy to tackle steatotic liver disease (SLD) and hepatocellular carcinoma (HCC). *Nutr Metab (Lond)*. 2024;21(1):112. doi:10.1186/s12986-024-00871-3

162. Badia RR, Pradhan RV, Ayers CR, Chandra A, Rohatgi A. The Relationship of Alcohol Consumption and HDL Metabolism in the Multiethnic Dallas Heart Study. *Journal of Clinical Lipidology*. 2023;17(1):124-130. doi:10.1016/j.jacl.2022.10.008

163. Khodja Y, Samuels ME. Ethanol-mediated upregulation of APOA1 gene expression in HepG2 cells is independent of de novo lipid biosynthesis. *Lipids Health Dis*. 2020;19(1):144. doi:10.1186/s12944-020-01309-4

164. Van Der Gaag MS, Van Tol A, Vermunt SHF, Scheek LM, Schaafsma G, Hendriks HFJ. Alcohol consumption stimulates early steps in reverse cholesterol

transport. *Journal of Lipid Research*. 2001;42(12):2077-2083. doi:10.1016/S0022-2275(20)31537-6

165. Zyriax BC, Lau K, Klähn T, Boeing H, Völzke H, Windler E. Association between alcohol consumption and carotid intima–media thickness in a healthy population: data of the STRATEGY study (Stress, Atherosclerosis and ECG Study). *Eur J Clin Nutr*. 2010;64(10):1199-1206. doi:10.1038/ejcn.2010.144

166. Huang Y, Li Y, Zheng S, Yang X, Wang T, Zeng J. Moderate alcohol consumption and atherosclerosis: Meta-analysis of effects on lipids and inflammation. *Wien Klin Wochenschr*. 2017;129(21-22):835-843. doi:10.1007/s00508-017-1235-6

167. Daya NR, Rebholz CM, Appel LJ, Selvin E, Lazo M. Alcohol Consumption and Risk of Hospitalizations and Mortality in the Atherosclerosis Risk in Communities Study. *Alcoholism Clin & Exp Res*. 2020;44(8):1646-1657. doi:10.1111/acer.14393

168. De Oliveira E Silva ER, Foster D, McGee Harper M, et al. Alcohol Consumption Raises HDL Cholesterol Levels by Increasing the Transport Rate of Apolipoproteins A-I and A-II. *Circulation*. 2000;102(19):2347-2352. doi:10.1161/01.CIR.102.19.2347

169. Ouimet M, Barrett TJ, Fisher EA. HDL and Reverse Cholesterol Transport: Basic Mechanisms and their Roles in Vascular Health and Disease. *Circ Res*. 2019;124(10):1505-1518. doi:10.1161/CIRCRESAHA.119.312617

170. Fournier N, Paul JL, Atger V, et al. HDL Phospholipid Content and Composition as a Major Factor Determining Cholesterol Efflux Capacity From Fu5AH Cells to Human Serum. *ATVB*. 1997;17(11):2685-2691. doi:10.1161/01.ATV.17.11.2685

171. Gaziano JM, Buring JE, Breslow JL, et al. Moderate Alcohol Intake, Increased Levels of High-Density Lipoprotein and Its Subfractions, and Decreased Risk of Myocardial Infarction. *N Engl J Med*. 1993;329(25):1829-1834. doi:10.1056/NEJM199312163292501

172. Zaid M, Miura K, Okayama A, et al. Associations of High-Density Lipoprotein Particle and High-Density Lipoprotein Cholesterol With Alcohol Intake, Smoking, and Body Mass Index ― The INTERLIPID Study ―. *Circ J*. 2018;82(10):2557-2565. doi:10.1253/circj.CJ-18-0341

173. Mukamal KJ, Mackey RH, Kuller LH, et al. Alcohol Consumption and Lipoprotein Subclasses in Older Adults. *The Journal of Clinical Endocrinology & Metabolism*. 2007;92(7):2559-2566. doi:10.1210/jc.2006-2422

174. Morton RE, Liu Y. The lipid transfer properties of CETP define the concentration and composition of plasma lipoproteins. *Journal of Lipid Research*. 2020;61(8):1168-1179. doi:10.1194/jlr.RA120000691

175. Brien SE, Ronksley PE, Turner BJ, Mukamal KJ, Ghali WA. Effect of alcohol consumption on biological markers associated with risk of coronary heart disease:

systematic review and meta-analysis of interventional studies. *BMJ*. 2011;342(feb22 1):d636-d636. doi:10.1136/bmj.d636

176. Siri-Tarino PW. Effects of Diet on High-Density Lipoprotein Cholesterol. *Curr Atheroscler Rep*. 2011;13(6):453-460. doi:10.1007/s11883-011-0207-y

177. Abubakr M, Abrar Z, Haider MJ, Barqaat SA. Letter on 'Head-to-Head Comparison Between Phosphatidylethanol Versus Indirect Alcohol Biomarkers for Diagnosis of METALD Versus MASLD : A Prospective Study.' *Aliment Pharmacol Ther*. 2025;61(6):1082-1083. doi:10.1111/apt.70008

178. Mathur K, Vilar-Gomez E, Connelly MA, et al. Circulating high density lipoprotein distinguishes alcoholic hepatitis from heavy drinkers and predicts 90-day outcome. *Journal of Clinical Lipidology*. 2021;15(6):805-813. doi:10.1016/j.jacl.2021.10.002

179. Chen N, Luo J, Zhou T, et al. Lysine β-hydroxybutyrylation promotes lipid accumulation in alcoholic liver disease. *Biochemical Pharmacology*. 2024;228:115936. doi:10.1016/j.bcp.2023.115936

180. Lambert JE, Ramos–Roman MA, Browning JD, Parks EJ. Increased De Novo Lipogenesis Is a Distinct Characteristic of Individuals With Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2014;146(3):726-735. doi:10.1053/j.gastro.2013.11.049

181. Wei Y, Jiang Y, Zhu J, et al. CD36-mediated uptake of oxidized LDL induces double-negative regulatory T cell ferroptosis in metabolic dysfunction-associated steatotic liver disease. *Metabolism*. 2025;164:156127. doi:10.1016/j.metabol.2024.156127

182. Van Der Veen JN, Kennelly JP, Wan S, Vance JE, Vance DE, Jacobs RL. The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 2017;1859(9):1558-1572. doi:10.1016/j.bbamem.2017.04.006

183. A. Elmihi K, Leonard KA, Nelson R, Thiesen A, Clugston RD, Jacobs RL. The emerging role of ethanolamine phosphate phospholyase in regulating hepatic phosphatidylethanolamine and plasma lipoprotein metabolism in mice. *The FASEB Journal*. 2024;38(18):e70063. doi:10.1096/fj.202401321R

184. Mihaylova MM, Shaw RJ. The AMPK signalling pathway coordinates cell growth, autophagy and metabolism. *Nat Cell Biol*. 2011;13(9):1016-1023. doi:10.1038/ncb2329

185. Mota M, Banini BA, Cazanave SC, Sanyal AJ. Molecular mechanisms of lipotoxicity and glucotoxicity in nonalcoholic fatty liver disease. *Metabolism*. 2016;65(8):1049-1061. doi:10.1016/j.metabol.2016.02.014

186. Cho KH, Nam HS, Kang DJ, Park MH, Kim JH. Long-Term Alcohol Consumption Caused a Significant Decrease in Serum High-Density Lipoprotein

(HDL)-Cholesterol and Apolipoprotein A-I with the Atherogenic Changes of HDL in Middle-Aged Korean Women. *IJMS*. 2022;23(15):8623. doi:10.3390/ijms23158623

187. Shimomura T, Wakabayashi I. Inverse associations between light-to-moderate alcohol intake and lipid-related indices in patients with diabetes. *Cardiovasc Diabetol*. 2013;12(1):104. doi:10.1186/1475-2840-12-104

188. Davidson WS, Cooke AL, Swertfeger DK, Shah AS. The Difference Between High Density Lipoprotein Subfractions and Subspecies: an Evolving Model in Cardiovascular Disease and Diabetes. *Curr Atheroscler Rep*. 2021;23(6):23. doi:10.1007/s11883-021-00925-4

189. Belhaj MR, Lawler NG, Hoffman NJ. Metabolomics and Lipidomics: Expanding the Molecular Landscape of Exercise Biology. *Metabolites*. 2021;11(3):151. doi:10.3390/metabo11030151

190. Gram MA, Olsen A, Andersen ZJ, Tjønneland A, Mellemkjær L. Associations of body fat mass and fat-free mass with breast cancer incidence in postmenopausal women: A Danish prospective cohort study. *Annals of Epidemiology*. 2023;80:30-36. doi:10.1016/j.annepidem.2023.01.014

191. Jeong SM, Lee DH, Giovannucci EL. Predicted lean body mass, fat mass and risk of lung cancer: prospective US cohort study. *Eur J Epidemiol*. 2019;34(12):1151-1160. doi:10.1007/s10654-019-00587-2

192. Hagens ERC, Feenstra ML, Van Egmond MA, et al. Influence of body composition and muscle strength on outcomes after multimodal oesophageal cancer treatment. *J cachexia sarcopenia muscle*. 2020;11(3):756-767. doi:10.1002/jcsm.12540

193. Kamarajah SK, Bundred J, Tan BHL. Body composition assessment and sarcopenia in patients with gastric cancer: a systematic review and meta-analysis. *Gastric Cancer*. 2019;22(1):10-22. doi:10.1007/s10120-018-0882-2

194. Takaoka S, Hamada T, Takahara N, et al. Body mass index and survival among patients with advanced biliary tract cancer: a single-institutional study with nationwide data-based validation. *J Gastroenterol*. 2024;59(8):732-743. doi:10.1007/s00535-024-02124-9

195. Wang G, Otto CC, Heij LR, et al. Impact of Altered Body Composition on Clinical and Oncological Outcomes in Intrahepatic Cholangiocarcinoma. *JCM*. 2023;12(24):7747. doi:10.3390/jcm12247747

196. Wijnhoven HAH, Snijder MB, Van Bokhorst-de Van Der Schueren MAE, Deeg DJH, Visser M. Region-Specific Fat Mass and Muscle Mass and Mortality in Community-Dwelling Older Men and Women. *Gerontology*. 2012;58(1):32-40. doi:10.1159/000324027

197. Zimmet PZ. Hyperinsulinemia—How Innocent a Bystander? *Diabetes Care*. 1993;16(Supplement_3):56-70. doi:10.2337/diacare.16.3.56

198. Yu H. Role of the Insulin-Like Growth Factor Family in Cancer Development and Progression. *Journal of the National Cancer Institute*. 2000;92(18):1472-1489. doi:10.1093/jnci/92.18.1472

199. Wajchenberg BLE. Subcutaneous and Visceral Adipose Tissue: Their Relation to the Metabolic Syndrome.

200. Wackerhage H, Christensen JF, Ilmer M, Von Luettichau I, Renz BW, Schönfelder M. Cancer catecholamine conundrum. *Trends in Cancer*. 2022;8(2):110-122. doi:10.1016/j.trecan.2021.10.005

201. Ibrahim MM. Subcutaneous and visceral adipose tissue: structural and functional differences. *Obesity Reviews*. 2010;11(1):11-18. doi:10.1111/j.1467-789X.2009.00623.x

202. Kuchler JC, Siqueira BS, Ceglarek VM, et al. The Vagus Nerve and Spleen: Influence on White Adipose Mass and Histology of Obese and Non-obese Rats. *Front Physiol*. 2021;12:672027. doi:10.3389/fphys.2021.672027

203. Kumari M, Khanna A. Sarcopenic Obesity and Its Impact on Quality of Life: A Concise Review. *CAS*. 2023;16(3):194-198. doi:10.2174/1874609816666230525161020

204. Aleixo GFP, Shachar SS, Nyrop KA, Muss HB, Malpica L, Williams GR. Myosteatosis and prognosis in cancer: Systematic review and meta-analysis. *Critical Reviews in Oncology/Hematology*. 2020;145:102839. doi:10.1016/j.critrevonc.2019.102839

205. Madeddu C, Mantovani G, Gramignano G, Astara G, Macciò A. Muscle Wasting as Main Evidence of Energy Impairment in Cancer Cachexia: Future Therapeutic Approaches. *Future Oncol*. 2015;11(19):2697-2710. doi:10.2217/fon.15.195

206. Maccio A, Sanna E, Neri M, Oppi S, Madeddu C. Cachexia as Evidence of the Mechanisms of Resistance and Tolerance during the Evolution of Cancer Disease. *IJMS*. 2021;22(6):2890. doi:10.3390/ijms22062890

207. Aleixo GFP, Yu H, Chen YT, et al. Myosteatosis evaluation using erector spinae and psoas muscles to predict adverse events during adjuvant chemotherapy for breast cancer. *Breast Cancer Res Treat*. 2021;186(2):487-495. doi:10.1007/s10549-020-06061-y

208. Hong S, Kim KW, Park HJ, et al. Impact of Baseline Muscle Mass and Myosteatosis on the Development of Early Toxicity During First-Line Chemotherapy in Patients With Initially Metastatic Pancreatic Cancer. *Front Oncol*. 2022;12:878472. doi:10.3389/fonc.2022.878472

209. Hu J, Xu J, Li M, et al. Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. *eClinicalMedicine*. 2024;68:102409. doi:10.1016/j.eclinm.2023.102409

210. Yuan L, Yang L, Zhang S, et al. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study. *eClinicalMedicine*. 2023;57:101834. doi:10.1016/j.eclinm.2023.101834

211. Nematollahi MA, Jahangiri S, Asadollahi A, et al. Body composition predicts hypertension using machine learning methods: a cohort study. *Sci Rep*. 2023;13(1):6885. doi:10.1038/s41598-023-34127-6

212. Razmpour F, Daryabeygi-Khotbehsara R, Soleimani D, et al. Application of machine learning in predicting non-alcoholic fatty liver disease using anthropometric and body composition indices. *Sci Rep*. 2023;13(1):4942. doi:10.1038/s41598-023-32129-y

213. Zhao F, Li C, Wang W, et al. Machine learning predicts the risk of osteoporosis in patients with breast cancer and healthy women. *J Cancer Res Clin Oncol*. 2024;150(2):102. doi:10.1007/s00432-024-05622-8

214. Wells JCK. Using Body Composition Assessment to Evaluate the Double Burden of Malnutrition. *Ann Nutr Metab*. 2019;75(2):103-108. doi:10.1159/000503666

215. Christakoudi S, Tsilidis KK, Evangelou E, Riboli E. Association of body-shape phenotypes with imaging measures of body composition in the UK Biobank cohort: relevance to colon cancer risk. *BMC Cancer*. 2021;21:1106. doi:10.1186/s12885-021-08820-6

216. Kolb H. Obese visceral fat tissue inflammation: from protective to detrimental? *BMC Med*. 2022;20:494. doi:10.1186/s12916-022-02672-y

217. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nat Rev Cancer*. 2004;4(8):579-591. doi:10.1038/nrc1408

218. Jensen MD. Role of Body Fat Distribution and the Metabolic Complications of Obesity. *The Journal of Clinical Endocrinology & Metabolism*. 2008;93(11_supplement_1):s57-s63. doi:10.1210/jc.2008-1585

219. Koutsari C, Jensen MD. Thematic review series: Patient-Oriented Research. Free fatty acid metabolism in human obesity. *Journal of Lipid Research*. 2006;47(8):1643-1650. doi:10.1194/jlr.R600011-JLR200

220. Santosa S, Jensen MD. Why are we shaped differently, and why does it matter? *American Journal of Physiology-Endocrinology and Metabolism*. 2008;295(3):E531-E535. doi:10.1152/ajpendo.90357.2008

221. McQuaid SE, Humphreys SM, Hodson L, Fielding BA, Karpe F, Frayn KN. Femoral Adipose Tissue May Accumulate the Fat That Has Been Recycled as VLDL and Nonesterified Fatty Acids. *Diabetes*. 2010;59(10):2465-2473. doi:10.2337/db10-0678

222. Manolopoulos KN, Karpe F, Frayn KN. Gluteofemoral body fat as a determinant of metabolic health. *Int J Obes*. 2010;34(6):949-959. doi:10.1038/ijo.2009.286

223. Sylow L, Tokarz VL, Richter EA, Klip A. The many actions of insulin in skeletal muscle, the paramount tissue determining glycemia. *Cell Metabolism*. 2021;33(4):758-780. doi:10.1016/j.cmet.2021.03.020

224. Nunney L. Size matters: height, cell number and a person's risk of cancer. *Proc Biol Sci*. 2018;285(1889):20181743. doi:10.1098/rspb.2018.1743

225. Grimberg A. Mechanisms by which IGF-I May Promote Cancer. *Cancer Biol Ther*. 2003;2(6):630-635.

226. Frühbeck G, Criado S, Gómez-Ambrosi J, et al. Increased IL- 36γ in visceral adipose tissue as a key mediator of obesity-driven inflammation in colon cancer. *J Mol Med*. 2025;103(6):699-711. doi:10.1007/s00109-025-02546-9

227. Chen Z, Ding C, Chen K, Gu Y, Qiu X, Li Q. Investigating the causal association between obesity and risk of hepatocellular carcinoma and underlying mechanisms. *Sci Rep*. 2024;14(1):15717. doi:10.1038/s41598-024-66414-1

228. Vrieling A, Kampman E, Knijnenburg NC, et al. Body Composition in Relation to Clinical Outcomes in Renal Cell Cancer: A Systematic Review and Meta-analysis. *European Urology Focus*. 2018;4(3):420-434. doi:10.1016/j.euf.2016.11.009

229. Rytka JM, Wueest S, Schoenle EJ, Konrad D. The Portal Theory Supported by Venous Drainage–Selective Fat Transplantation. *Diabetes*. 2011;60(1):56-63. doi:10.2337/db10-0697

230. Bowers LW, Brenner AJ, Hursting SD, Tekmal RR, deGraffenried LA. Obesity-associated systemic interleukin-6 promotes pre-adipocyte aromatase expression via increased breast cancer cell prostaglandin E2 production. *Breast Cancer Res Treat*. 2015;149(1):49-57. doi:10.1007/s10549-014-3223-0

231. Iyengar NM, Zhou XK, Gucalp A, et al. Systemic Correlates of White Adipose Tissue Inflammation in Early-Stage Breast Cancer. *Clin Cancer Res*. 2016;22(9):2283-2289. doi:10.1158/1078-0432.CCR-15-2239

232. Shin E, Koo JS. The Role of Adipokines and Bone Marrow Adipocytes in Breast Cancer Bone Metastasis. *Int J Mol Sci*. 2020;21(14):4967. doi:10.3390/ijms21144967

233. Schmidt S, Monk JM, Robinson LE, Mourtzakis M. The integrative role of leptin, oestrogen and the insulin family in obesity-associated breast cancer: potential effects of exercise. *Obes Rev*. 2015;16(6):473-487. doi:10.1111/obr.12281

234. Wu S, Zhu W, Thompson P, Hannun YA. Evaluating intrinsic and non-intrinsic cancer risk factors. *Nat Commun*. 2018;9:3490. doi:10.1038/s41467-018-05467-z

235. Chen X, Ji Y, Liu R, et al. Mitochondrial dysfunction: roles in skeletal muscle atrophy. *J Transl Med*. 2023;21:503. doi:10.1186/s12967-023-04369-z

236. Nelke C, Dziewas R, Minnerup J, Meuth SG, Ruck T. Skeletal muscle as potential central link between sarcopenia and immune senescence. *eBioMedicine*. 2019;49:381-388. doi:10.1016/j.ebiom.2019.10.034

237. Xiao M, Xu J, Wang W, et al. Functional significance of cholesterol metabolism in cancer: from threat to treatment. *Exp Mol Med*. 2023;55(9):1982-1995. doi:10.1038/s12276-023-01079-w

238. Imaizumi S, Navab M, Morgantini C, et al. Dysfunctional High-Density Lipoprotein and the Potential of Apolipoprotein A-1 Mimetic Peptides to Normalize the Composition and Function of Lipoproteins. *Circ J*. 2011;75(7):1533-1538. doi:10.1253/circj.cj-11-0460

239. Georgila K, Vyrla D, Drakos E. Apolipoprotein A-I (ApoA-I), Immunity, Inflammation and Cancer. *Cancers (Basel)*. 2019;11(8):1097. doi:10.3390/cancers11081097

240. Ginsberg HN, Packard CJ, Chapman MJ, et al. Triglyceride-rich lipoproteins and their remnants: metabolic insights, role in atherosclerotic cardiovascular disease, and emerging therapeutic strategies—a consensus statement from the European Atherosclerosis Society. *Eur Heart J*. 2021;42(47):4791-4806. doi:10.1093/eurheartj/ehab551

241. Varbo A, Benn M, Tybjærg-Hansen A, Nordestgaard BG. Elevated Remnant Cholesterol Causes Both Low-Grade Inflammation and Ischemic Heart Disease, Whereas Elevated Low-Density Lipoprotein Cholesterol Causes Ischemic Heart Disease Without Inflammation. *Circulation*. 2013;128(12):1298-1309. doi:10.1161/CIRCULATIONAHA.113.003008

242. Qie R, Liu H, Liu Y, et al. Remnant Cholesterol, C-Reactive Protein, and Risks of Lung Cancer Morbidity and Mortality Among Women: A Prospective Cohort Study. *American Journal of Preventive Medicine*. 2026;70(1). doi:10.1016/j.amepre.2025.108140

243. Xu H, Liu P, Chen S, Wang G, Wu S, Zhang X. Remnant cholesterol and systemic inflammation as synergistic predictors of cancer risk: A 16-year prospective cohort study. *Cancer*. 2025;131(19):e70096. doi:10.1002/cncr.70096

244. Nicklin P, Bergman P, Zhang B, et al. Bidirectional Transport of Amino Acids Regulates mTOR and Autophagy. *Cell*. 2009;136(3):521-534. doi:10.1016/j.cell.2008.11.044

245. Kanai Y. Amino acid transporter LAT1 (SLC7A5) as a molecular target for cancer diagnosis and therapeutics. *Pharmacology & Therapeutics*. 2022;230:107964. doi:10.1016/j.pharmthera.2021.107964

246. Wang Q, Holst J. L-type amino acid transport and cancer: targeting the mTORC1 pathway to inhibit neoplasia. *Am J Cancer Res*. 2015;5(4):1281-1294.

247. Ramos-Jiménez A, Hernández-Torres RP, Hernández-Ontiveros DA, et al. An Update of the Promise of Glycine Supplementation for Enhancing Physical Performance and Recovery. *Sports (Basel)*. 2024;12(10):265. doi:10.3390/sports12100265

248. Wong SW, Kwon MJ, Choi AMK, Kim HP, Nakahira K, Hwang DH. Fatty Acids Modulate Toll-like Receptor 4 Activation through Regulation of Receptor Dimerization and Recruitment into Lipid Rafts in a Reactive Oxygen Species-dependent Manner. *J Biol Chem*. 2009;284(40):27384-27392. doi:10.1074/jbc.M109.044065

249. Luo J, Peng S, Jiang Z, et al. Roles and therapeutic opportunities of ω-3 long-chain polyunsaturated fatty acids in lung cancer. *iScience*. 2024;28(1):111601. doi:10.1016/j.isci.2024.111601

250. Singh S, Brocker C, Koppaka V, et al. Aldehyde Dehydrogenases in Cellular Responses to Oxidative/electrophilic Stress. *Free Radic Biol Med*. 2013;56:89-101. doi:10.1016/j.freeradbiomed.2012.11.010

251. Zhao W, Xia Y, Gao Z, Chen J, Zhang E. Multiple roles of ALDH1 in health and disease. *Front Physiol*. 2025;16:1627164. doi:10.3389/fphys.2025.1627164

252. Duan X, Hu H, Wang L, Chen L. Aldehyde dehydrogenase 1 family: A potential molecule target for diseases. *Cell Biology International*. 2024;48(7):909-922. doi:10.1002/cbin.12188

253. Idorn M, Hojman P. Exercise-Dependent Regulation of NK Cells in Cancer Protection. *Trends in Molecular Medicine*. 2016;22(7):565-577. doi:10.1016/j.molmed.2016.05.007

254. Pedersen L, Idorn M, Olofsson GH, et al. Voluntary Running Suppresses Tumor Growth through Epinephrine- and IL-6-Dependent NK Cell Mobilization and Redistribution. *Cell Metabolism*. 2016;23(3):554-562. doi:10.1016/j.cmet.2016.01.011

255. Severinsen MCK, Pedersen BK. Muscle–Organ Crosstalk: The Emerging Roles of Myokines. *Endocr Rev*. 2020;41(4):594-609. doi:10.1210/endrev/bnaa016

256. Park SY, Hwang BO, Song NY. The role of myokines in cancer: crosstalk between skeletal muscle and tumor. *BMB Rep*. 2023;56(7):365-373. doi:10.5483/BMBRep.2023-0064

257. Hojman P, Dethlefsen C, Brandt C, Hansen J, Pedersen L, Pedersen BK. Exercise-induced muscle-derived cytokines inhibit mammary cancer cell growth. *Endocrinol Metab*. 2011;301.

258. Catoire M, Kersten S. The search for exercise factors in humans. *The FASEB Journal*. 2015;29(5):1615-1628. doi:10.1096/fj.14-263699

259. Huang S, Li J, Shearer GC, et al. Longitudinal study of alcohol consumption and HDL concentrations: a community-based study12. *Am J Clin Nutr*. 2017;105(4):905-912. doi:10.3945/ajcn.116.144832

260. Bröer S. Amino Acid Transporters as Targets for Cancer Therapy: Why, Where, When, and How. *Int J Mol Sci*. 2020;21(17):6156. doi:10.3390/ijms21176156

261. Kovalski JR, Kuzuoglu-Ozturk D, Ruggero D. Protein synthesis control in cancer: selectivity and therapeutic targeting. *EMBO J*. 2022;41(8):e109823. doi:10.15252/embj.2021109823

262. Durham WJ, Dillon EL, Sheffield-Moore M. Inflammatory burden and amino acid metabolism in cancer cachexia. *Curr Opin Clin Nutr Metab Care*. 2009;12(1):72-77. doi:10.1097/MCO.0b013e32831cef61

8. Appendix

8.1. List of Tables

Table 1. ICD-10 Cancer Endpoint Definitions and Corresponding Codes

Table 2. Summary statistics of patients with MetALD and MASLD

Table 3. Top 30 discriminatory metabolites between MetALD and MASLD.

Table 4: Top 30 discriminatory metabolites between ALD and MASLD.

Table 5. Baseline characteristics of participants CCA cohort.

Table 6. The association between regional BC and the mortality of CCA.

Table 7. Baseline characteristics of the participants of cancer cohort.

8.2. List of Figures

Figure 1. Study design and workflow. (A) Inclusion and exclusion flowchart for participant-level and genetic-level data. (B) Number of cancer cases in the UK Biobank and FinnGen datasets.

Figure 2 Circle plots for lipidomic analysis. (A) Circle plot for lipidomic analysis for MetALD versus MASLD UKB participants. Red=Higher in MetALD, blue=higher in MASLD. (B) Circle plot for lipidomic analysis for ALD versus MASLD UKB participants. Red=Higher in ALD, blue=higher in MASLD. Baseline Lipidomic parameters were measured through NMR spectroscopy. Hazard ratios (with 95% confidence intervals) are presented per 1-SD higher metabolic biomarker on the natural log scale, stratified by age, sex, body mass index and Townsend deprivation index. *Original code by Diego J. Aguilar-Ramirez. Apo-A1, apolipoprotein A1, Apo-B, apolipoprotein B, FA, fatty acids, FAw3, omega-3 fatty acids, FAw6, omega-6 fatty acids, LA, linoleic acid, MUFA, monounsaturated fatty acids, PUFA, polyunsaturated fatty acids, SFA, saturated fatty acids, TG, triglycerides, 3-HB, 3-Hydroxybutyrate, DHA, docosahexaenoic acid.

Figure 3 Forest Plot of Mendelian randomisation analysis between alcohol consumption and the top 30 discriminatory metabolites using Inverse variance weighting. OR: odds ratio; 95% CI: 95% confidence interval.

Figure 4. Prognostic performance of prediction models for cholangiocarcinoma (CCA) and its subtypes. (A-C) Training cohort for overall CCA, ICC and ECC model; (D-F) Test cohort for CCA, ICC and ECC model. Note: AUC: Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve; XGB: eXtreme Gradient

Boosting; GBM: Gradient Boosting Machine; SVM: Support Vector Machine; NN: Neural Network; MLP: Multilayer Perceptron.

Figure 5 Global feature explanations using the SHAP method. (A) SHAP summary bar plot for overall CCA; (B) SHAP summary dot plot for overall CCA; (C) SHAP summary bar plot for ICC; (D) SHAP summary dot plot for ICC; (E) SHAP summary bar plot for ECC; (F) SHAP summary dot plot for ECC.

Figure 6. Manhattan plot of phenome-wide association study (PheWAS) results. (A) All system phenotypes; (B) Neoplasm categories. Each point represents a phenotype tested for association with regional body composition (BC). Phenotypes are displayed along the x-axis and grouped by disease category; the y-axis indicates the strength of evidence as −log10(P) from models adjusted for sex, age, BMI, and ancestry. The horizontal pink line denotes the phenome-wide significance threshold after Benjamini-Hochberg (BH) multiple-testing correction. Upward- and downward-pointing triangles indicate positive and negative associations, respectively (effect estimate $\beta > 0$ and $\beta < 0$).

Figure 7. Heatmap of Cox proportional hazards regression results. Rows represent regional body composition features and columns represent cancer endpoints. Each cell shows the hazard ratio (HR) estimated from multivariable Cox models, with color intensity reflecting the magnitude and direction of the association (HR > 1 indicates increased risk; HR < 1 indicates decreased risk). Statistically significant associations after Benjamini-Hochberg (BH) multiple-testing correction are marked (e.g., asterisks). All models were adjusted for prespecified covariates (e.g., age, sex, BMI, ancestry, and other relevant confounders).

Figure 8. Heatmap of MR analysis results. Rows represent regional body composition features and columns represent cancer endpoints. Each cell shows the odds ratio (OR) estimated from MR analysis results, with color intensity reflecting the magnitude and direction of the association (OR > 1 indicates increased risk; OR < 1 indicates decreased risk). Statistically significant associations after Benjamini-Hochberg (BH) multiple-testing correction are marked (e.g., asterisks).

Figure 9 Restricted cubic spline (RCS) analysis illustrating discordant associations between regional BC and cancer risk across the individual-level and genetic-level analyses. The solid line depicts the adjusted hazard ratio (HR) estimated by Cox proportional hazards models incorporating RCS terms for regional BC, and the shaded area indicates the 95% confidence interval. Models were adjusted for prespecified

covariates (e.g., age, sex, BMI, ethnicity, and other relevant confounders). P values for the overall association and for nonlinearity were obtained from Wald tests comparing models with and without spline terms.

Figure 10. Construction of the MOFA-based multi-omics integration model and assessment of variance explained. (A) Overview of the total sample size and the omics views included in the MOFA framework. (B) Total variance explained by each view across models fitted with different numbers of latent factors. (C) Variance explained by each view in the final MOFA model with 15 latent factors. (D) Heatmap showing the variance explained by individual latent factors within each omics view, illustrating view-specific contributions and between-factor heterogeneity.

Figure 11. Association heatmap between MOFA latent factors and cancer outcomes. Heatmap summarizing the associations of MOFA-derived latent factors with cancer incidence outcomes estimated using Cox proportional hazards models. Rows represent cancer endpoints and columns represent latent factors (Factor 1–Factor 15). Each cell shows the hazard ratio (HR) estimated from multivariable Cox models, with color intensity reflecting the magnitude and direction of the association (HR > 1 indicates increased risk; HR < 1 indicates decreased risk). Statistically significant associations after Benjamini-Hochberg (BH) multiple-testing correction are marked (e.g., asterisks). All models were adjusted for prespecified covariates (e.g., age, sex, BMI, ancestry, and other relevant confounders).

Figure 12. The feature loadings across four omics views for MOFA Factors 1-3. Bar plots display the top 10 features ranked by the largest absolute loadings for each omics view. Loadings quantify both the direction and magnitude of each feature's contribution to the corresponding latent factor; positive and negative values indicate concordant and discordant contributions to factor scores, respectively. The features are ordered by absolute loading within each view.

Figure 13. Metabolite pathway enrichment and gene functional enrichment results. (A-C) Bubble plots show pathway enrichment results for molecular metabolites associated with MOFA Factors 1-3, respectively. (D-F) Bar charts show Gene Ontology (GO) term and KEGG pathway enrichment results for genes corresponding to MOFA Factors 1-3, respectively, and the enrichment results were adjusted for multiple testing using the Benjamini–Hochberg (BH) multiple-testing correction.

Figure 14. Construction of the final body composition–cancer risk association atlas. (A) Venn diagram showing the number of cancer endpoints that were statistically

significant across three evidence layers, including individual-level analyses, genetic-level analyses, and MOFA latent factor–based analyses. (B) Flowchart illustrates the exclusion of cancer endpoints with inconsistent directions of association between the individual-level and genetic-level analyses. (C) Summary of the final set of cancer outcomes retained for downstream analyses and the corresponding significantly associated MOFA latent factors. (D) Chord diagram depicting the atlas of risk associations linking body composition (BC) patterns represented by latent factors to the final cancer outcomes.

8.3. Acknowledgment

First and foremost, I would like to express my sincere gratitude to my supervisor, PD Dr. Med. Franziska Alexandra Meister, for her dedicated guidance and support throughout my doctoral studies. I am deeply grateful for the opportunity to learn from her and for the way she introduced me to the rigorous and inspiring world of medical research.

I would also like to extend my special thanks to my second supervisor, Prof. Dr. Carolin Victoria Schneider. I am profoundly grateful for her continuous care, support, and encouragement in both my professional and personal life. I also sincerely appreciate her invaluable guidance and commitment to all of my research projects. Whether in advancing my academic work or supporting my personal development, she has always offered generous encouragement and unwavering support.

My sincere thanks also go to PD Dr. Med. Katharina Jöchle for her kind help and support at the beginning of my doctoral journey. Her assistance enabled me to start my studies and research more smoothly and laid an important foundation for my subsequent work.

In addition, I would like to thank Prof. Dr. Kai Markus Schneider for his guidance and support during my doctoral oral examination.

I am also deeply grateful to all collaborators who contributed to the publication of my work and to the related research projects: Kai Markus Schneider, Helen Ye Rim Huang, Lanlan Chen, Yazhou Chen, Anastasia Raptis, Kate Townsend Creasy, Jan Clusmann, Felix van Haag, Paul-Henry Koop, Adrien Guillot, Tom Luedde, Rohit Loomba, Sven Francque, and Carolin Victoria Schneider. I sincerely thank them for their important support and contributions to study design, manuscript development, academic exchange, and collaboration.

Furthermore, I would like to express my heartfelt appreciation to the colleagues and friends who supported me in my studies, work, and daily life in Germany: Benjamin Laevens, Yazhou Chen, Yuanyuan Liu, Corinna Meeßen, Tobias Seibel, Niharika Jakhar, Thriveni Basavanapura Raju, Holger Wilms, and Abdallah Abouabdallah. I am deeply thankful for their kindness, help, and support as I adapted to a new academic environment, research setting, and everyday life in Germany. Their generosity and companionship made me feel warmth and belonging far from home, and gave me strength to move forward with confidence.

I would also like to express my sincere gratitude to Prof. Dr. Franziska Alexandra Meister and Prof. Dr. Carolin Victoria Schneider for serving as reviewers of this dissertation, and to Prof. Daniel Truhn for chairing the examination committee. I greatly appreciate the time, effort, and care they devoted to the evaluation of this thesis and the defense process.

Finally, my deepest gratitude goes to my family and friends. Thank you for your unwavering support, understanding, and companionship throughout my time studying in Aachen. Your love, patience, and encouragement have sustained me through every challenge, and I would not have come this far without you.

8.4. Affidavit according to § 5 (1) for Data Retention

I hereby declare that the original data forming the basis of this doctoral thesis are stored by me personally, Feng Cao, Kühlwetterstraße 8, 52072, Aachen, deposited.

8.5. Affidavit according to § 5 (1) and § 11 (3) 12. of the doctoral regulations

I, Feng Cao, hereby declare on oath that I independently prepared and analyzed the following results portrayed in the dissertation "A Precision Medicine Perspective on Disease Risk Assessment Driven by Metabolic Phenotypes". I had the following assistance with completing the dissertation. These are listed in the acknowledgments.

Names	Candidate Feng Cao	Practical supervisor Prof. Dr. Carolin V Schneider	Yazhou Chen	Doctoral supervisor Prof. Dr. Franziska Alexandra Meister	Sum (%)
Study supervision	20	40	10	30	100
Study design/conception		50		50	100
Provision of samples		100			100
Data management	50	50			100
Data analysis	100				100
Results interpretation	40	30	10	20	100
Dissertation drafting	100				100
Dissertation revision	40	30	10	20	100

_____

Signature of the doctoral candidate

As the supervisor of the above dissertation, I confirm the statements of Feng Cao

_____

Signature of the doctoral supervisor