Impact of age, sex and obesity on cancer biology

Masters in Bioinformatics - Spring 2026

Sebastian Dyhr
202104408
Msc. Student Bioinformatics,
Aarhus University

Nicolai Birkbak
Supervisor
Professor, Molekylær medicinsk afdeling - MOMA,
Aarhus University

12.06.2026

Abbreviations

ASCAT: allele-specific copy number analysis of tumours

BH: Benjamini-Hochberg

BLCA: bladder urothelial carcinoma

BMI: body mass index

BRCA: breast invasive carcinoma

CD4: cluster of differentiation 4

CD8: cluster of differentiation 8

CESC: cervical squamous cell carcinoma

CN_HIGH: copy number high

COAD: colon adenocarcinoma

COSMIC: catalogue of somatic mutations in cancer

DEG: differentially expressed gene

DLBC: diffuse large B-cell lymphoma

DMP: differentially methylated position

EMT: epithelial-mesenchymal transition

ESCA: esophageal carcinoma

FDR: false discovery rate

GDC: Genomic Data Commons

GSEA: gene set enrichment analysis

HNSC: head and neck squamous cell carcinoma

IFN: interferon

JAK2: Janus kinase 2

KIRC: kidney renal clear cell carcinoma

LIHC: liver hepatocellular carcinoma

LUAD: lung adenocarcinoma

MAF: minor allele frequency

MSI-H: microsatellite instability high

MSigDB: Molecular Signature Database

NES: normalised enrichment score

NK: natural killer cell

ORA: overrepresentation analysis

OV: ovarian serous cystadenocarcinoma

PKA: protein kinase A

POLE: DNA polymerase epsilon

PRR: pattern recognition receptor

READ: rectum adenocarcinoma

SARC: sarcoma

SASP: Senescence-Associated Secretory Phenotype

SCNA: somatic copy number alteration

SKCM: skin cutaneous melanoma

SNV: single nucleotide variant

STAT3: signal transducer and activator of transcription 3

TCGA: The Cancer Genome Atlas

TCRA: T cell receptor alpha

THCA: thyroid carcinoma

THYM: thymoma

TLR7: Toll-like receptor 7

TME: tumour microenvironment

TREC: T cell receptor excision circle

TREG: regulatory T cell

UCEC: uterine corpus endometrial carcinoma

UCS: uterine carcinosarcoma

UCSC: University of California Santa Cruz

WES: whole exome sequencing

fgsea: fast gene set enrichment analysis

logFC: log fold change

wGII: weighted genome integrity index

Contents

1 Abstract

Age, sex, and BMI are established risk factors for cancer, yet their molecular impact on tumour biology across cancer types is not completely understood. In this thesis I investigate the association between these clinical variables and tumour biology across 33 cancer types from The Cancer Genome Atlas (TCGA), using gene expression, DNA methylation, genomic instability, somatic driver mutations, and immune infiltration data. Genomic instability was assessed using the weighted genome integrity index (wGII). Gene expression and DNA methylation associations with BMI and age were assessed using Spearman correlation, while sex differences were identified using differential expression and methylation analyses. Pathway enrichment was assessed using fgsea for age and BMI variables, and over-representation analysis (ORA) for sex. Somatic driver mutation patterns were investigated using driver weight correlation, total mutation burden comparison, and logistic regression. Immune infiltration was characterised using TcellExTRECT and CIBERSORTx. Age showed the most consistent associations with genomic instability across cancer types. Immune-related pathways emerged as a central theme, with complement and allograft rejection recurrently associated with BMI, interferon and hormonal signalling most prominent in sex differences, and inflammatory gene sets broadly associated with age. T cell infiltration was not associated with BMI or sex, but was negatively associated with age in THYM, consistent with thymic involution. Driver mutation analyses revealed largely cancer-type specific patterns, with SCF complex components recurring in BMI-associated cancers and PI3K/AKT/mTOR pathway genes prominent in younger patients. These findings underscore the importance of considering age, sex, and BMI as biological variables in cancer research, with immune-related pathways emerging as a shared theme across all three variables.