Title

Determining the Mechanisms by which Exercise Exerts its Effect on Cancer

Author

Tsai, Brandon L

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Determining the Mechanisms by which Exercise Exerts its Effect on Cancer

A dissertation submitted in partial satisfaction of the requirements

for the degree Doctor of Philosophy in Human Genetics

by

Brandon Liang Tsai

2024

ABSTRACT OF THE DISSERATION

Determining the Mechanisms

by which Exercise

Exerts its Effect on Cancer

by

Brandon Liang Tsai

Doctor of Philosophy in Human Genetics

University of California, Los Angeles, 2024

Professor Paul Christopher Boutros, Chair

Exercise is broadly beneficial for human health and is one of the most potent modifiable lifestyle risk factors for many diseases, including cancer. Cancer is a highly prevalent disease and is one of the leading causes of death worldwide. While cancer incidence is a stochastic process, many factors influence the probability that an individual will develop cancer, including genetics, environmental exposures and lifestyle factors. An estimated two-thirds of cancer deaths in the U.S. can be attributed to modifiable risk factors such as smoking, diet and exercise. Exercise is the strongest positive modifiable risk factor and has been linked to almost all cancer types and stages of disease progression. Individuals who exercise more have reduced risk of developing cancer and improved clinical outcomes. However, even within a specific cancer type, tumors appear to respond differently to exercise. Indeed, the molecular mechanisms by which exercise exerts its

effect on cancer outcomes are almost entirely unclear. My dissertation aims to fill this fundamental gap in our understanding of cancer etiology, uncovering how exercise affects diverse host and tumor molecular landscapes, as well as clinical outcomes. To enhance our molecular understanding of exercise oncology, I have separated my research into three chapters. These chapters cover a variety of study designs, including large cross-sectional patient cohorts, prospective longitudinal clinical trials, and experimental mouse studies. Each study design provides its own unique advantages, complementing each other and coming together to reveal novel insights into exercise oncology.

In Chapter 1, I present the study of a large cross-sectional cohort of 5,150 patients with linked tumor genomic sequencing from 38 different cancer types and clinical annotation of post-diagnosis exercise dose and other important covariate lifestyle behaviors such as smoking, alcohol, and diet. Leveraging the large sample size and diversity of cancer types, we investigated both pan-cancer and cancer type-specific exercise-associated modulation of the cancer genome. Tumors differed in mutation burden, mutational signatures, and specific driver mutations in an exercise dose-dependent manner. The direction and magnitude of these associations varied across cancer types, yet exercisers had reduced hazard of all-cause mortality for all cancers combined and for multiple individual cancer types. Our data show exercise promotes genome maintenance, which may have broad implications for understanding how exercise suppresses tumor pathogenesis and potentially other common age and lifestyle-related diseases. To our knowledge, this study is the first to characterize the link between exercise and human tumor genomic profiling.

In Chapter 2, I present the study of a decentralized, digital prospective longitudinal clinical trial of 13 patients with prescribed exercise therapy. The prescribed exercise regimen between cancer diagnosis and surgical resection allows for the control of exercise dose in humans linked to high-

quality clinical data. We performed longitudinal multiparametric profiling of host physiology, plasma, gut microbial composition, and tumor tissue before, during, and after exercise therapy intervention. Time-series analyses revealed hundreds of host molecular changes in the plasma proteome and metabolome and gut microbiome involved in a diverse-array of biological processes. System-wide changes were paralleled by modulation of core tumor gene expression pathways notably tumor cell cycle regulation, stress response, and metabolism. Integrative network analyses revealed the complexity of the host–tumor interaction under exercise therapy regulation, elucidating novel mechanistic insights. Variability at baseline and in response to treatment emphasized highly personalized responses to uniform exercise therapy. Our study provides an example of the application of a digital approach to generate a longitudinal high-definition dataset providing a framework of the integrative effects of exercise therapy with considerable translational and discovery potential. To our knowledge, this study is the first deep longitudinal host-tissue molecular characterization of patient response to exercise therapy.

In Chapter 3, I present a mouse study of tumor xenografts from seven human breast cancer cell lines and syngeneic grafts from one mouse breast cancer cell line, with the cell lines representing a range of breast cancer subtypes. Tumors derived from different cell lines displayed differential growth phenotypes, with some tumors growing faster and others growing slower in response to exercise treatment. The tumors also had distinct genomic, transcriptomic, and proteomic changes with exercise. These molecular changes pointed to perturbations in common biological pathways, including DNA repair. The integration of exercise-associated molecular alterations and growth phenotypes across multiple breast cancer subtypes provides further evidence that, indeed, the effect of exercise on cancer is context dependent, not only varying by cancer type, but also by subtype, adding another layer of complexity to the field of exercise oncology.

The dissertation of Brandon Liang Tsai is approved.

Alex Anh-Tuan Bui

David Wayne Dawson

Owen N. Witte

Paul Christopher Boutros, Committee Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

LIST OF TABLES

Chapter 1

Chapter 3

ACKNOWLEDGEMENTS

Chapter 1 is a version of the manuscript Tsai, B. L., Arbet, J., Eng, S. E., Lee, C. P., Underwood, W. P., Liu, L. Y., Fickera, G., Chun, S., Moffitt, A., Walsh, J., Harrison, J., Bliss, J. W., Tammela, T., Boutros, P. C., Jones, L. W., Exercise-associated modulation of the cancer genome. All authors directly participated in manuscript planning and development.

Chapter 2 is a version of the manuscript Tsai, B. L., Liu, L. Y., Eng, S. E., Underwood, W. P., Gardner, G., Mueller, J., Ehdaie, B., Laudone, V. P., Eastham, J. A., Arbet, J., Wang, N. K., Oh, J., Tao, S., Weigelt, B., Vahdatinia, M., Stylianou, A., Bose, S., Locasale, J. W., Boutros, P. C., Jones, L. W. Personalized Dynamic Mapping of Integrative Response to Exercise Therapy in Cancer. All authors directly participated in manuscript planning and development.

Chapter 3 is a version of a manuscript still in preparation.

Thank you to the UCLA-Caltech Medical Scientist Training Program (MSTP) and the National Institute of Health (NIH) National Institute of General Medical Sciences (NIGMS) Institutional National Research Service Award (T32) project numbers 2T32GM8042-36A1 and 5T32GM8042-37 for funding and support.

VITA

EDUCATION

University of California, Los Angeles

PhD, Human Genetics	2021 –
MD, David Geffen School of Medicine	2019 –
BS, Microbiology, Immunology, and Molecular Genetics	2013-2017

PUBLICATIONS

Google Scholar: https://scholar.google.com/citations?user=ppQYjV4AAAAJ&hl=en

RESEARCH

Dr. Paul Boutros, UCLA	April 2021 –
Dr. Daniel Geschwind, UCLA	June – August 2019
Dr. Owen Witte, UCLA	February 2017 – May 2019
Dr. Xia Yang, UCLA	January 2014 – February 2017
Dr. Hong Wu, Peking University	June – August 2016
Dr. Utpal Banerjee and Dr. Cory Evans, UCLA	October – December 2014
Dr. Bruce Blumberg, UC Irvine	June – August 2013, 2014

AWARDS AND HONORS

UCLA JCCC Graduate Student 3rd Place Poster Award	2024
University of California-Wide Grad Slam First Place Award	2023
UCLA Grad Slam First Place Award	2023
UCLA Grad Slam Audience Choice Award	2023
UCLA Grad Slam Finalist	2022
UCLA Med Mentor of the Year	2021
NIH Medical Scientist Training Program Fellowship	2019 –
UCLA Cum Laude	2017
UCLA Vice Provost Award for Excellence in Research	2017
UCLA Chancellor's Student Achievement Award	2015-2017
UCLA-PKU Joint Research Institute Program	2016
UCLA Amgen Scholars Program	2015
UCLA Undergraduate Research Scholars Program	2014-2017
UCLA Regents Scholarship (top 1% of incoming class)	2013-2017
UCLA Dean's Honors List	2013-2017
UCI Cancer Research Institute Youth Science Fellowship Program	2013

PRESENTATIONS

UCLA Jonsson Comprehensive Cancer Center	May 2024
UCLA Medical Scientist Training Program	January 2024
UCLA Development	January 2024
UCLA Life Science Dean's Advisory Board	May 2023
University of California-Wide Grad Slam	May 2023
UCLA Grad Slam	March 2023

American Association for Cancer Research	April 2023
UCLA Jonsson Comprehensive Cancer Center	May 2023
American Association for Cancer Research	April 2023
UCLA Institute for Quantitative and Computational Biosciences	March 2023
UCLA Institute for Quantitative and Computational Biosciences	September 2022
UCLA Jonsson Comprehensive Cancer Center	April 2022
American Association for Cancer Research	April 2022
American Association for Cancer Research	April 2021
Engineered Immunity and Parker Institute for Cancer Immunotherapy	May 2018
UCLA Broad Stem Cell Research Center	February 2018
UCLA Molecular and Medical Pharmacology	November 2017
UCLA Research Poster Day	May 2017
UCLA Regents Scholars Society	November 2016
UCLA-Peking University Joint Research Institute	August 2016
UCLA Research Poster Day	May 2016
UCLA Microbiology, Immunology, and Molecular Genetics	December 2016
UCLA Regents Scholars Society	October 2015
UCLA Amgen Scholars Program	August 2015
UCLA Summer Programs for Undergraduate Research	August 2015
UCLA Research Poster Day	May 2015
UCLA Regents Scholars Society	October 2014
UCI Cancer Research Institute Youth Fellowship Program	August 2013

PEER REVIEW

Blood Advances	March 2023

LEADERSHIP, MENTORSHIP, AND EDUCATIONAL OUTREACH

Med Mentors Executive Board Member	February 2020 –
UCLA MSTP Admission Committee Member	September 2019

CHAPTER 1:

Exercise-associated modulation of the cancer genome

Abstract

Exercise influences the primary incidence of several types of human cancer, but the underlying molecular mechanisms remain enigmatic. We linked the genomic landscapes of 5,150 patient tumors from 38 different cancer types with clinical annotation of post-diagnosis exercise dose. Tumors differed in mutation burden, mutational signatures and specific driver mutations in an exercise dose-dependent manner. The direction and magnitude of these associations varied across cancer types, yet exercisers had reduced hazard of all-cause mortality for all cancers combined and for multiple individual cancer types. Our data show exercise promotes genome maintenance, which may have broad implications for understanding how exercise suppresses tumor pathogenesis and potentially other common age and lifestyle-related diseases.

An individual's lifetime risk of cancer is influenced by both intrinsic and extrinsic factors. Intrinsic processes include stochastic errors in DNA replication, while extrinsic ones relate to environmental exposures, including lifestyle behaviors (1-5). Genotoxic lifestyle factors such as smoking, alcohol consumption and ultraviolet (UV) radiation exposure are well-characterized to elevate lifetime cancer risk (6). Correlative and mechanistic studies show that most genotoxic exposures, including certain unhealthy lifestyle factors, (7-9) influence cancer risk by promoting specific mutational processes, thus shaping mutational architecture and tumor evolution.

To date, molecular epidemiology studies have focused exclusively on factors that elevate cancer risk. By contrast, our understanding of how lifestyle factors that lower cancer risk influence tumor genomic landscapes is nascent. One of the most important lifestyle factors in this context is exercise. Adults reporting regular exercise have lower risk of non-communicable diseases (10),

including reduced primary incidence of multiple cancers (11, 12). Paradoxically, exercise is associated with elevated incidence of melanoma and prostate cancer (12). Despite a wealth of studies linking exercise to cancer incidence, we do not understand how exercise influences mutational processes and shapes tumor evolution (13).

To fill this gap, we linked DNA sequencing of 5,150 human cancers to detailed clinical quantification of exercise (Figure 1A, Supplemental Table 1). Self-reported, post-diagnosis exercise was evaluated using validated surveys and quantified as moderate to vigorous metabolic equivalents of task hours per week (MET-hours/week), consistent with international guidelines (14). This proxy for long-term exercise status (15, 16) enabled interrogation of the dose-dependent impact of exercise on the cancer genome. Patients represented a broad range of clinico-epidemiologic characteristics (Figures 1B-C, Supplemental Figure 1). Tumor and matched normal DNA were sequenced to identify somatic single nucleotide variants (SNVs), copy number alterations (CNAs) and structural variants (SVs) (17, 18) across a total of 38 cancer types, including 11 individual cancer types with at least 75 patients (Figure 1D).

We first examined the impact of exercise dose on genome instability. Three complementary metrics were assessed: fraction genome altered (FGA; fraction of genome affected by gains or losses), microsatellite instability (MSI; percentage of somatic microsatellite sites) and tumor mutational burden (TMB; SNVs per Mbp) (19, 20). Exercise was associated with lower FGA for all cancers combined ($p = 0.035$, adjusted linear regression; ALR) as well as for three individual cancer types: breast cancer, melanoma and ovarian cancer (Figure 1E). These and subsequent analyses adjusted for clinico-epidemiologic features appropriate to the specific cancer type being studied (e.g., age, sex, smoking; Supplemental Figure 2, Supplemental Table 2). Similarly, exercise was also associated with lower rates of MSI in the same three cancer types. By contrast,

exercise was marginally linked to higher TMB for all cancers combined (p = 0.10, ALR), although no individual cancer type reached statistical significance. Sensitivity analyses (Supplementary Methods) showed exercise associations were directionally consistent across clinico-epidemiologic features (Supplemental Figure 3). Thus, higher exercise lowers genome instability in some cancer types, including several without prior epidemiologic support (i.e. melanoma, ovarian), but not all.

Different mutagens and defects in DNA damage repair genes can induce specific patterns of genome instability (8). To determine whether the associations of exercise with genome instability were rooted in specific mutational processes, we modeled the relationship between exercise dose and trinucleotide mutational signatures (8). Of the mutational signatures detected across all cancers combined (Supplementary Methods), none were associated with exercise (Figure 1F). However, several individual cancer types showed strong associations between exercise dose and mutational signatures. For example, the clock-like SBS1 signature was more frequently detected in breast cancers and sarcomas arising in patients reporting higher exercise (pbreast=0.016, psarcoma=0.041, adjusted Firth's logistic regression; AFLR). By contrast, the SBS3 signature of defective homologous recombination DNA damage repair was less common in renal cell carcinomas arising in patients with higher exercise (p = 0.026, AFLR; Figure 1F).

To assess whether differences in genome instability and mutational processes might represent individual driver events, we tested whether exercise dose influenced gene-specific alterations in tumor mutational architecture. Across all cancers combined, exercise was linked to fewer somatic mutations in cell cycle regulators CCND1 and CDKN2A ($p < 0.08$ for each gene, AFLR) (Supplemental Figure 4). Across the 11 individual cancer types studied, mutations in 18 genes were statistically significantly associated with exercise dose ($p < 0.05$; Figure 1G).

Most cancer types studied were limited in statistical power due to small sample size, however detailed analyses were possible in breast cancer (n=1,018; Supplemental Table 3) and non-small cell lung cancer (NSCLC; n=811, Supplemental Table 4). Higher exercise dose was associated with fewer mutations in the cell division regulating kinase AURKA in breast cancer ($p = 0.0057$, AFLR; Figure 2A). In NSCLC, higher exercise dose was associated with fewer mutations in cell cycle regulator CDKN2B ($p = 0.026$, AFLR) and tumor suppressor ARID1A ($p = 0.044$, AFLR) but more mutations in oncogene MET ($p = 0.036$, AFLR; Figure 2B). These results were adjusted for potential confounders (Supplemental Table 2). Stratification analyses further confirmed these results, for example showing that AURKA in breast cancer (Figure 2C) and ARID1A in NSCLC are associated with exercise dose but not age or body mass index (BMI) (Figure 2D).

We next evaluated whether the impact of exercise might vary across cancer subtypes. We focused on breast cancer, stratifying patients by receptor status (Supplemental Table 3). Sensitivity analyses showed that exercise dose influenced the cancer genome similarly in all subtypes (Figure 2E). Similarly in NSCLC, the histological subtypes of squamous cell carcinoma and adenocarcinoma showed directional consistency for all exercise-associated gene mutations (Figure 2F). Intriguingly, not all exercise associations were subtype-independent. Higher exercise dose was strongly associated with more PAK1 mutations only in HR-negative disease ($p = 0.031$), but not in any other subgroups (Figure 2E). Thus, exercise appears to primarily shape tumor evolution in a site-specific manner, with modest subtype-specific effects.

Increased genome instability is associated with more aggressive tumor biology and poorer patient survival (21, 22). Given the associations between exercise and genome instability, we reasoned exercise dose might influence clinical outcome. Patients were classified as exercisers (>0 MET-hours/week; n=3,365; 65%) and non-exercisers (0 MET-hours/week; n=1,785; 35%), and overall

survival was compared between these groups. Median 3.9-year follow-up was available. Exercisers had better overall survival for all cancers combined (hazard ratio (HR) = 0.64, 95% confidence interval (CI), 0.55-0.74, adjusted Cox proportional hazards regression; ACPHR), and for breast cancer (HR=0.60, 95% CI, 0.42-0.88, ACPHR), NSCLC (HR=0.59, 95% CI, 0.42-0.83, ACPHR) and soft tissue sarcoma (HR=0.31, 95% CI, 0.12-0.78, ACPHR) compared with non-exercisers and controlling for stage (Figure 2G).

Exercise is posited to influence cancer risk and progression in a complex, type-dependent manner (11, 12). Our results suggest that exercise may influence tumor evolution, or the rate at which different tumor molecular features evolve from pre-malignant states. However, our analyses relied on self-report of exercise post-diagnosis, which is standard in epidemiological studies, but prone to misclassification and reverse-causation bias. To minimize the latter, we employed both multivariable analyses and stratified analyses. Several of our findings are consistent with a large body of previous literature, further supporting the validity and reliability of our exercise data. For example, we observed exercise-associated genomic features preferentially in cancer types with strong epidemiological evidence that exercise lowers risk (e.g., breast, NSCLC), but not in those where exercise appears to have little impact (e.g., pancreas) or may even increase risk (e.g., prostate cancer) (11, 12). Melanoma provides a particularly illustrative example of this consistency. It has been widely hypothesized that the higher incidence of melanoma in exercisers is caused by their increased incidental sun exposure (23). Consistent with this hypothesis, melanoma patients with higher exercise had tumors that tended to have increased detection of SBS7a and SBS7b (UV light exposure).

It is unclear how exercise influences mutational processes, independent of age, BMI and other lifestyle factors, but two broad non-mutually exclusive explanations appear plausible. First,

exercise may directly impact cell-intrinsic mutagenesis. Exercise lowers both circulating and tissue bioavailability of effectors that either directly induce DNA damage (reactive oxygen species) or promote mutagenic processes like cell division. Thus, DNA damage and/or chances of acquiring oncogenic mutations might be lower in exercisers, decreasing genome instability either directly or via epigenetic mechanisms (24). Second, exercise may influence cancer evolution via the tumor microenvironment by enhanced tumor immunity (25, 26) or modulation of growth factor, oxygen and nutrient availability (so called "public goods") (27). These effects might alter cell-environment fitness dynamics, selecting for clones better adapted to the level of exercise-conditioning in an individual patient's tissue. While the mechanism remains uncertain, it is clear that exercise modulates tumor evolution, with broad potential implications for clinical and public health guidelines. Understanding the mechanism by which exercise suppresses tumor pathogenesis may provide insight into the many other common age and lifestyle-related diseases for which exercise is the strongest known protective factor.

Figure 1

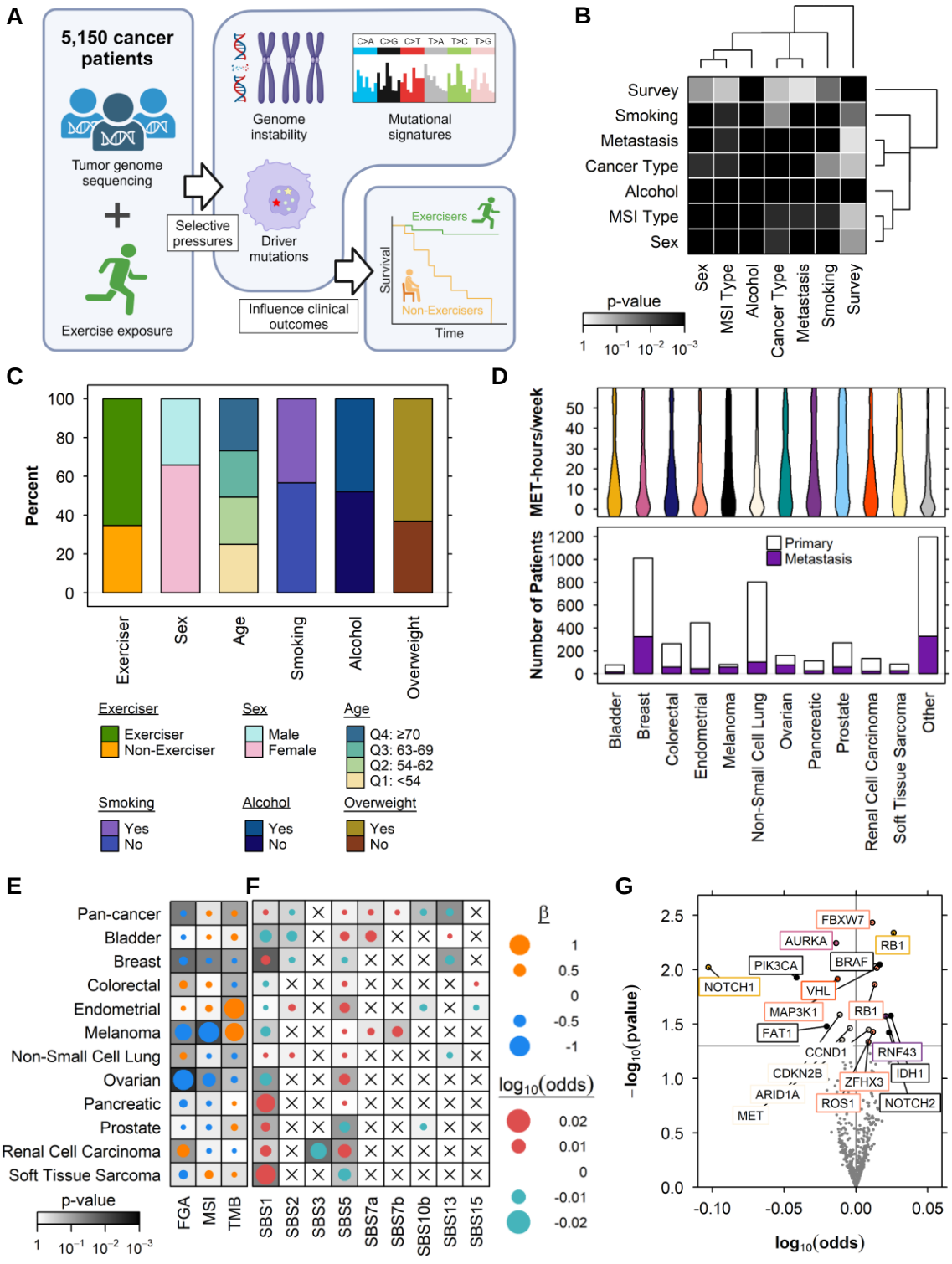
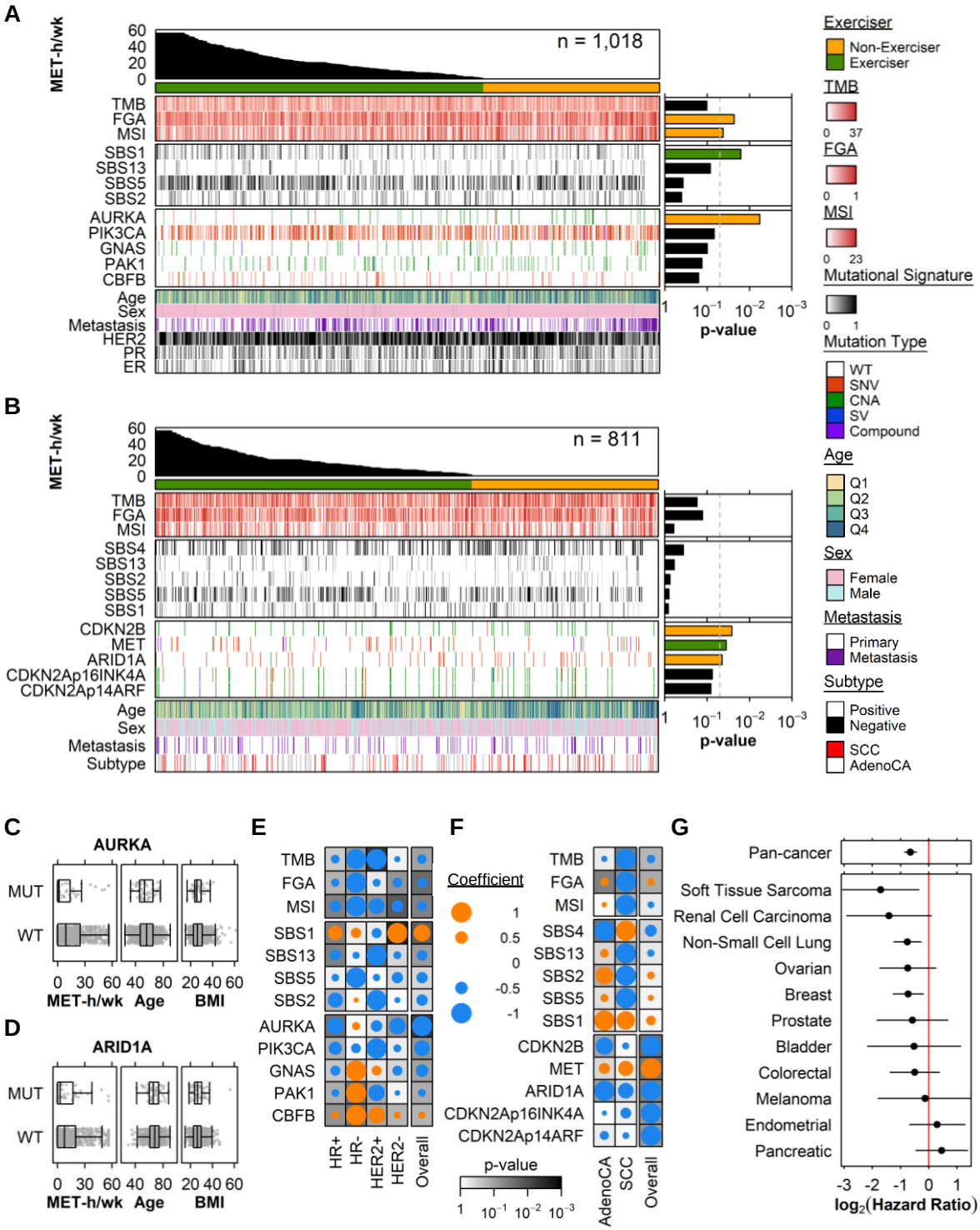
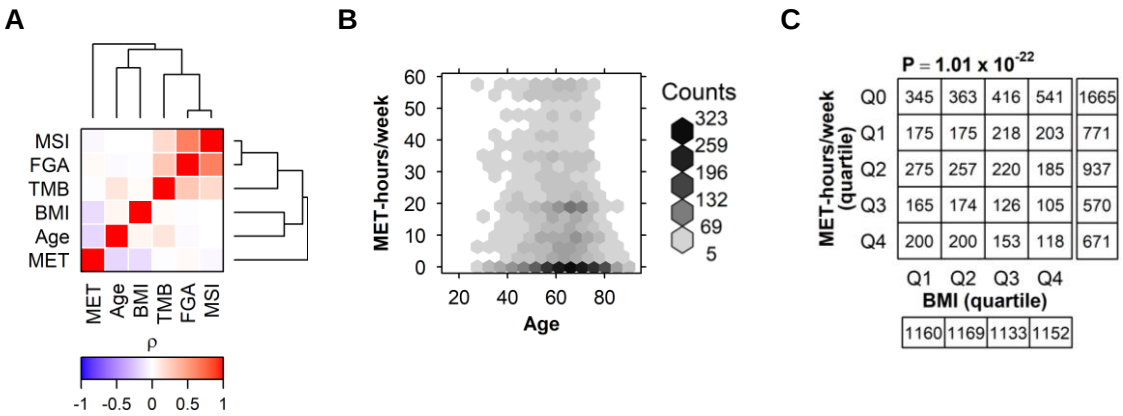

A
5,150 cancer patients
Tumor genome sequencing
Exercise exposure
Selective pressures
Genome instability
Mutational signatures
Driver mutations
Influence clinical outcomes
Exercisers
Non-Exercisers
Survival
Time
B
Survey
Smoking
Metastasis
Cancer Type
Alcohol
MSI Type
Sex
p-value
C
Percent
Exerciser
Sex
Age
Smoking
Alcohol
Overweight
Exerciser
Exerciser
Non-Exerciser
Sex
Male
Female
Age
Q4: ≥70
Q3: 63-69
Q2: 54-62
Q1: <54
Smoking
Yes
No
Alcohol
Yes
No
Overweight
Yes
No
D
MET-hours/week
Number of Patients
Primary
Metastasis
Bladder
Breast
Colorectal
Endometrial
Melanoma
Non-Small Cell Lung
Ovarian
Pancreatic
Prostate
Renal Cell Carcinoma
Soft Tissue Sarcoma
Other
E
Pan-cancer
Bladder
Breast
Colorectal
Endometrial
Melanoma
Non-Small Cell Lung
Ovarian
Pancreatic
Prostate
Renal Cell Carcinoma
Soft Tissue Sarcoma
FGA
MSI
TMB
p-value
F
SBS1
SBS2
SBS3
SBS5
SBS7a
SBS7b
SBS10b
SBS13
SBS15
β
log10(odds)
G
−log10(pvalue)
log10(odds)
FBXW7
AURKA
RB1
PIK3CA
BRAF
NOTCH1
VHL
MAP3K1
RB1
FAT1
CCND1
RNF43
CDKN2B
ZFHX3
IDH1
ARID1A
ROS1
NOTCH2
MET

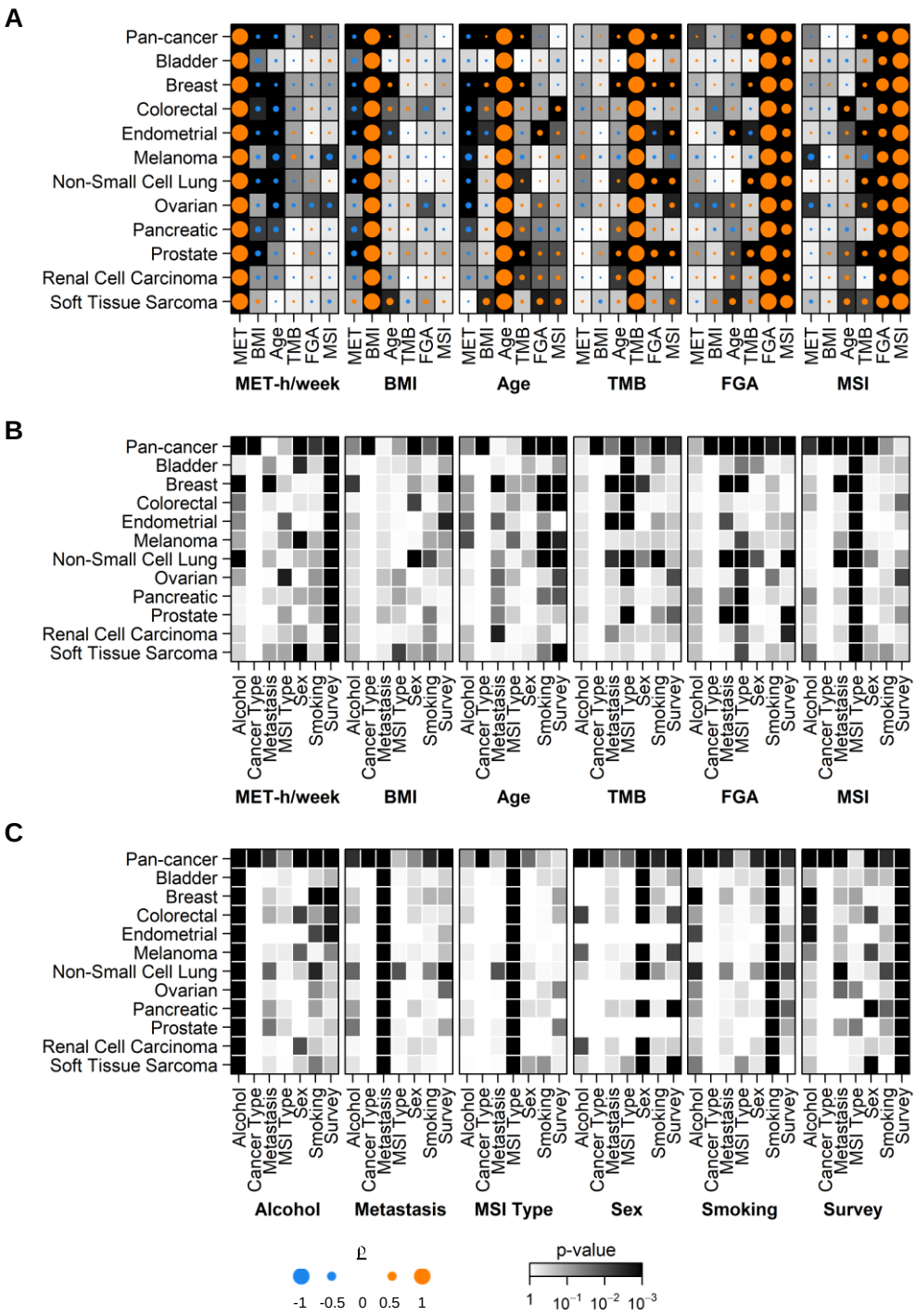
Figure 2


A
MET-h/wk
n = 1,018
TMB
FGA
MSI
SBS1
SBS13
SBS5
SBS2
AURKA
PIK3CA
GNAS
PAK1
CBFB
Age
Sex
Metastasis
HER2
PR
ER
p-value
B
n = 811
SBS4
SBS13
SBS2
SBS5
SBS1
CDKN2B
MET
ARID1A
CDKN2Ap16INK4A
CDKN2Ap14ARF
Subtype
Exerciser
Non-Exerciser
Exerciser
Mutational Signature
Mutation Type
WT
SNV
CNA
SV
Compound
Q1
Q2
Q3
Q4
Female
Male
Primary
Metastasis
Positive
Negative
SCC
AdenoCA
C
AURKA
MUT
WT
MET-h/wk
Age
BMI
D
ARID1A
E
HR+
HR-
HER2+
HER2-
Overall
F
Coefficient
p-value
AdenoCA
SCC
Overall
G
Pan-cancer
Soft Tissue Sarcoma
Renal Cell Carcinoma
Non-Small Cell Lung
Ovarian
Breast
Prostate
Bladder
Colorectal
Melanoma
Endometrial
Pancreatic
log2(Hazard Ratio)

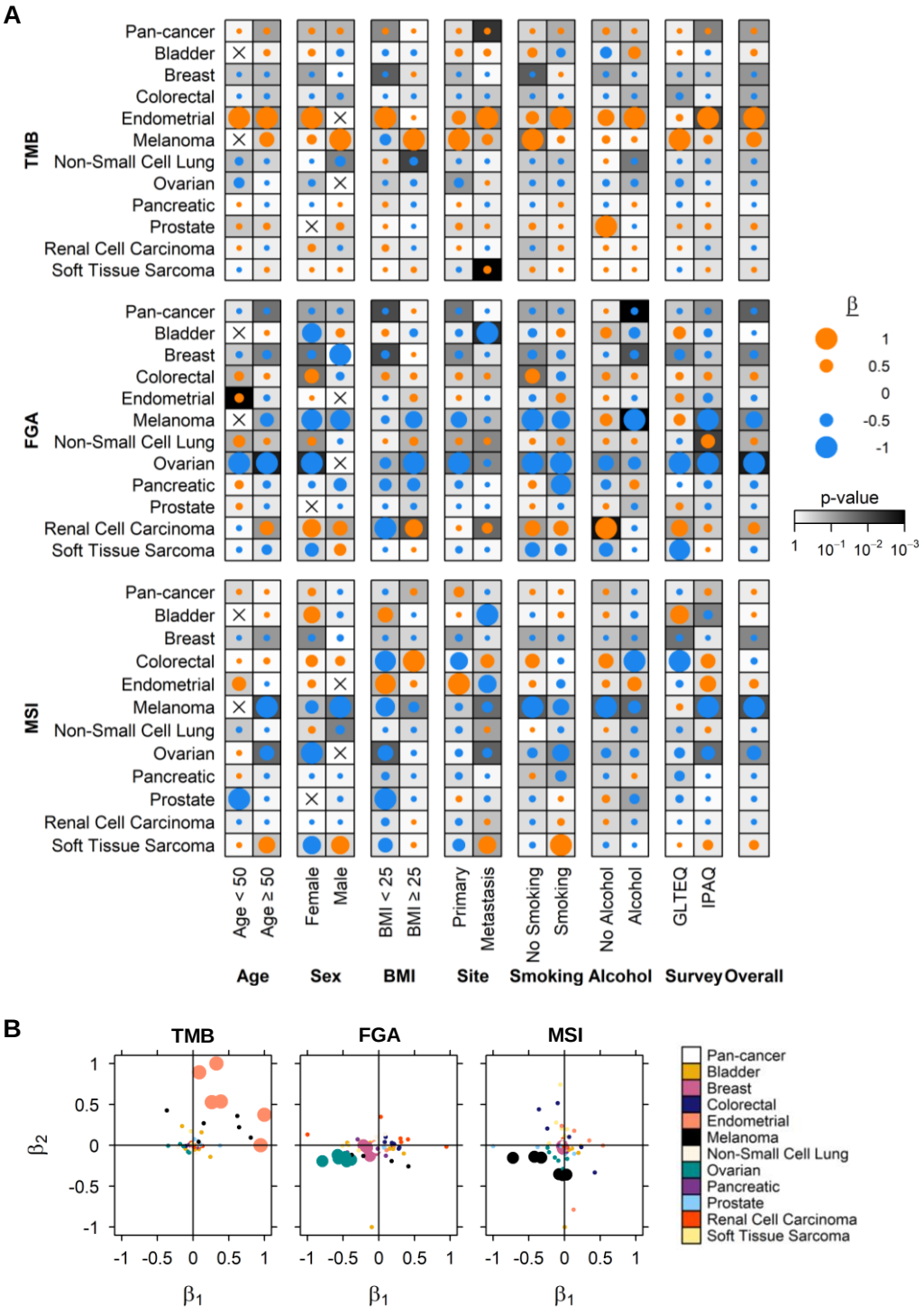
Supplemental Figure 1



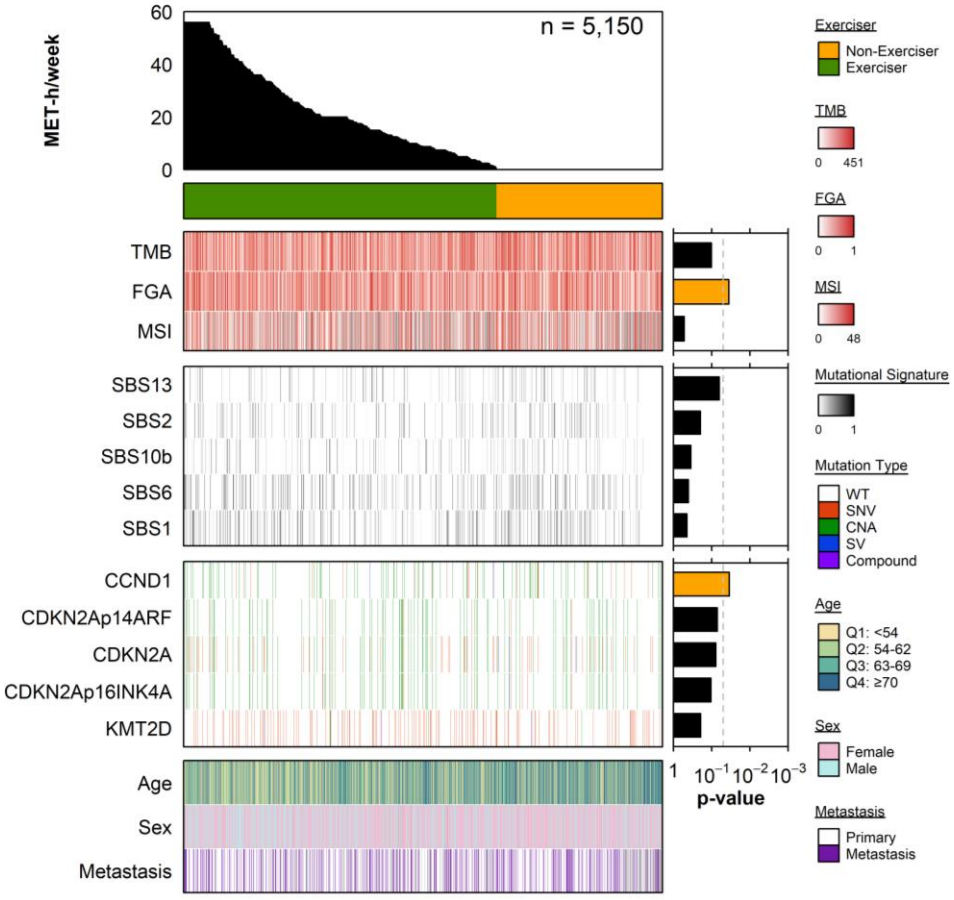
Supplemental Figure 2



Supplemental Figure 3



Supplemental Figure 4



Supplemental Table 1

Cancer – no. (%)	
Breast Cancer	1,018 (20)
Non-Small Cell Lung Cancer	811 (16)
Mature B-Cell Neoplasms	453 (9)
Endometrial Cancer	450 (9)
Prostate Cancer	272 (5)
Colorectal Cancer	265 (5)
Leukemia	186 (4)
Ovarian Cancer	160 (3)
Renal Cell Carcinoma	133 (3)
Other	121 (2)
Pancreatic Cancer	115 (2)
Soft Tissue Sarcoma	86 (2)
Myelodysplastic Workup	84 (2)
Melanoma	83 (2)
Myelodysplastic Syndromes	82 (2)
Bladder Cancer	77 (1)
Mature T and NK Neoplasms	77 (1)
Esophagogastric Cancer	71 (1)
Cancer of Unknown Primary	67 (1)
Myeloproliferative Neoplasms	56 (1)
B-Lymphoblastic Leukemia/Lymphoma	49 (1)
Myeloproliferative Workup	40 (1)
Head and Neck Cancer	33 (1)
Hepatobiliary Cancer	33 (1)
Thyroid Cancer	32 (1)
Appendiceal Cancer	31 (1)
Gastrointestinal Stromal Tumor	27 (1)
Glioma	26 (1)
Hodgkin Lymphoma	26 (1)
Cervical Cancer	25 (0)
Myelodysplastic/Myeloproliferative Neoplasms	23 (0)
Uterine Sarcoma	20 (0)
Salivary Gland Cancer	19 (0)
Skin Cancer, Non-Melanoma	19 (0)
Germ Cell Tumor	17 (0)
Blood Cancer, NOS	16 (0)
Small Cell Lung Cancer	13 (0)
Small Bowel Cancer	12 (0)
Bone Cancer	11 (0)
Mesothelioma	11 (0)

Site of tumor sample – no. (%)	
Primary	3,497 (68)
Metastasis	1,140 (22)
Body mass index – mean (SD)	27.6 (± 6.02)
Smoking history – no. (%)	2,217 (43)
Alcohol consumption – no. (%)	2,251 (44)

[1]: metabolic equivalent of task (hours per week)

[2]: standard deviation

Supplemental Table 2

	All adjusted models		Survival	Cancer type-specific covariates			
	Age	Survey	Tumor Site	Sex	Smoking	Alcohol	Cancer Type
Pan-cancer	+	+	+	+	+	+	+
Bladder	+	+	+	+	+	-	-
Breast	+	+	+	-	-	+	-
Colorectal	+	+	+	+	+	+	-
Endometrial	+	+	+	-	-	-	-
Melanoma	+	+	+	+	-	-	-
Non-Small Cell Lung	+	+	+	+	+	-	-
Ovarian	+	+	+	-	-	-	-
Pancreatic	+	+	+	+	+	-	-
Prostate	+	+	+	-	-	-	-
Renal Cell Carcinoma	+	+	+	+	+	-	-
Soft Tissue Sarcoma	+	+	+	+	-	-	-

Supplemental Table 3

Characteristic	n=1,018
MET-h/week[1] – mean (SD[2])	15.3 (± 17.5)
Age – mean (SD)	55.9 (± 12.4)
ER Status – no. (%)	
Negative	121 (12)
Positive	704 (69)
PR Status – no. (%)	
Negative	201 (20)
Positive	623 (61)
HER2 Status – no. (%)	
Negative	614 (60)
Positive	104 (10)
Site of tumor sample – no. (%)	
Primary	684 (67)
Metastasis	325 (32)
Body mass index – mean (SD)	27.2 (± 6.16)
Smoking history – no. (%)	340 (33)
Alcohol consumption – no. (%)	2,251 (44)

[1]: metabolic equivalent of task (hours per week)

[2]: standard deviation

Supplemental Table 4

Characteristic	n=811
MET-h/week[1] – mean (SD[2])	13.2 (± 15.8)
Age – mean (SD)	67.0 (± 9.59)
Female – no. (%)	549 (68)
Subtype – no. (%)	
Adenocarcinoma	661 (82)
Squamous Cell Carcinoma	70 (9)
Site of tumor sample – no. (%)	
Primary	699 (86)
Metastasis	102 (13)
Body mass index – mean (SD)	26.9 (± 5.39)
Smoking history – no. (%)	587 (72)
Alcohol consumption – no. (%)	444 (55)

[1]: metabolic equivalent of task (hours per week)

[2]: standard deviation

Figure Captions

Figure 1. Study overview and tumor genomic landscapes by exercise dose

(A) Schematic of study. (B) Heatmap of pairwise Pearson's X2 test of independence p-values between categorical clinico-epidemiologic features. (C) Epidemiologic distribution of cohort. (D) Distribution of cancer type and exercise dose. (E) Dotmap of associations between TMB/FGA/MSI and exercise dose, evaluated by adjusted linear regression. Dot size indicates the magnitude of the exercise β coefficient. β coefficients were scaled from -1 to 1. Dot color indicates the direction of association. Background shading indicates statistical significance of association. (F) Dotmap of associations between mutational signatures and exercise dose, evaluated by adjusted Firth's logistic regression. Dot size indicates the magnitude of the exercise log10(odds) coefficient. Dot color indicates the direction of association. Background shading indicates statistical significance of association. (G) Volcano plot of associations between gene mutations and exercise dose, evaluated by adjusted Firth's logistic regression. Significant genes ($p < 0.05$) are labeled and colored by the cancer type they were associated in.

Abbreviations: MET – metabolic equivalents of task; FGA – fraction of genome altered; MSI – microsatellite instability; TMB – tumor mutation burden; Q – quartile; SBS – single base substitution

Figure 2. Exercise dose and tumor genomic landscapes in breast and non-small cell lung cancer

(A) Tumor genomic landscape in breast cancer. (B) Tumor genomic landscape in non-small cell lung cancer (NSCLC). From top to bottom: exercise dose, exercise status, markers of genome instability, mutational signatures, gene mutation status and clinico-epidemiologic features. Columns represent patients sorted by decreasing exercise. Markers of genome instability heatmap

shows TMB/FGA/MSI values. Mutational signatures heatmap shows fraction of single base substitutions attributed to the mutational signature. Gene mutation status heatmap shows type of mutation present. Barplots to the right show p-values and are colored in the direction of significant ($p < 0.05$) associations. Orange bars indicate the feature was significantly associated with decreasing exercise, and green bars indicate the feature was significantly associated with increasing exercise. (C) Distribution of exercise dose, age and BMI by AURKA mutation status in breast cancer. (D) Distribution of exercise dose, age and BMI by ARID1A mutation status in NSCLC. (E) Sensitivity analysis by receptor status in breast cancer. (F) Sensitivity analysis by histology in NSCLC. Dot size indicates the magnitude of the exercise coefficient. Coefficients were scaled from -1 to 1. Dot color indicates the direction of association. Background shading indicates statistical significance of association. (G) Forest plot of hazard ratios of all-cause mortality and 95% confidence intervals between exercisers and non-exercisers by cancer type, evaluated by adjusted Cox proportional hazards regression.

Abbreviations: MET-h/wk – metabolic equivalents of task hours per week; TMB – tumor mutation burden; FGA – fraction of genome altered; MSI – microsatellite instability; SBS – single base substitution; HER2 – human epidermal growth factor receptor 2; PR – progesterone receptor; ER – estrogen receptor; WT – wildtype; SNV – single nucleotide variant; CNA – copy number alteration; SV – structural variant; Q – quartile; SCC – squamous cell carcinoma; AdenoCA – adenocarcinoma; MUT – mutant; BMI – body mass index; HR – hormone receptor

Supplementary Figure Captions

Supplemental Figure 1: Pan-cancer univariate analyses of clinico-epidemiologic features association with exercise dose in 5,150 tumors

(A) Heatmap of pairwise Spearman's correlation ρ between continuous clinico-epidemiologic features. (B) Hexbin plot of exercise dose vs. age. Shading of hexbin indicates density. (C) Contingency table of exercise dose and BMI quartiles. P-value calculated by Pearson's X2 test of independence.

Abbreviations: MET – metabolic equivalents of task hours per week; BMI – body mass index; TMB – tumor mutation burden; FGA – fraction of genome altered; MSI – microsatellite instability; MET-hours/week – metabolic equivalents of task hours per week; Q – quartile

Supplemental Figure 2: Per-cancer type univariate analyses of clinico-epidemiologic feature associations

(A) Dotmaps of Spearman's correlation ρ between continuous clinico-epidemiologic features. Dot size indicates the magnitude of correlation. Dot color indicates the direction of correlation. Background shading indicates statistical significance of correlation. (B) Heatmaps of Welch's two-sided t-test p-values between continuous and categorical clinico-epidemiologic features. (C) Heatmaps of Pearson's X2 test of independence p-values between categorical clinico-epidemiologic features.

Abbreviations: MET-h/week – metabolic equivalents of task hours per week; BMI – body mass index; TMB – tumor mutation burden; FGA – fraction of genome altered; MSI – microsatellite instability

Supplemental Figure 3: Sensitivity analyses of association between exercise dose and markers of genome instability by clinico-epidemiologic features

(A) Dotmaps of associations between TMB (top), FGA (middle) and MSI (bottom) and exercise dose, evaluated by adjusted linear regression. Each cohort is dichotomized into two groups for each clinico-epidemiologic feature. Un-dichotomized overall associations are on the right. Dot size indicates the magnitude of the exercise β coefficient. β coefficients were scaled from -1 to 1. Dot color indicates the direction of association. Background shading indicates statistical significance of association. (B) Scatterplots of association between TMB/FGA/MSI and exercise dose, evaluated by adjusted linear regression. Each dot represents the association for a dichotomized clinico-epidemiologic feature. β1 represents the exercise β coefficient for the first strata in each dichotomy (i.e. age<50, female, BMI<25, primary, no smoking, no alcohol, GLTEQ), and β2 represents the coefficient for the second strata (i.e. age ≥50, male, BMI ≥25, metastasis, smoking, alcohol, IPAQ). β coefficients were scaled from -1 to 1. Larger dots indicate the un-stratified overall association was significant ($p < 0.05$). Large dots are primarily in quadrants I and IV, showing significant exercise-associations demonstrate directional consistency when stratified by clinico-epidemiologic features.

Abbreviations: TMB – tumor mutation burden; FGA – fraction of genome altered; MSI – microsatellite instability; BMI – body mass index; GLTEQ - Godin Leisure-Time Exercise Questionnaire; IPAQ –International Physical Activity Questionnaire

Supplemental Figure 4: Pan-cancer genomic landscape and exercise dose

From top to bottom: exercise dose, exercise status, markers of genome instability, mutational signatures, gene mutation status and clinico-epidemiologic features. Columns represent patients

sorted by decreasing exercise dose. Markers of genome instability heatmap shows TMB/FGA/MSI values. Mutational signatures heatmap shows fraction of single base substitutions attributed to the mutational signature. Gene mutation status heatmap shows type of mutation present. Barplots to the right show p-values and are colored in the direction of significant ($p < 0.05$) associations. Orange bars indicate the feature was significantly associated with decreasing exercise, and green bars indicate the feature was significantly associated with increasing exercise.

Abbreviations: MET-h/week – metabolic equivalents of task hours per week; TMB – tumor mutation burden; FGA – fraction of genome altered; MSI – microsatellite instability; SBS – single base substitution; WT – wildtype; SNV – single nucleotide variant; CNA – copy number alteration; SV – structural variant; Q – quartile

Supplementary Table Legends

Supplemental Table 1 – Cohort characteristics

Supplemental Table 2 – Covariates included in adjusted models by cancer type

Supplemental Table 3 – Breast cancer cohort characteristics

Supplemental Table 4 – NSCLC cohort characteristics

Supplemental Table 5 – Statistical tables

Supplementary Methods

Patients and clinical data extraction

A total of 5,150 patients with 38 different cancer types sequenced at Memorial Sloan Kettering Cancer Center from March 2001 to May 2023 were included in this study. All tumors were profiled using the Memorial Sloan Kettering Integrated Molecular Profiling of Actionable Cancer Targets (MSK-IMPACT) clinical sequencing assay, a hybridization capture-based, next-generation sequencing platform (28). This set included samples sequenced with the IMPACT-HEME-400 (n=875), IMPACT-HEME-468 (n=218), IMPACT341 (n=41), IMPACT410 (n=659), IMPACT468 (n=2748) and IMPACT505 (n=606) panels. While some patients may have multiple cancer diagnoses (i.e. recurrence or metastasis), only the samples profiled by the MSK-IMPACT were considered for this study. Clinical data were retrieved from the institutional electronic health records (EHR) database. Metastatic events were extracted from the pathology report of the sequenced samples and patients' electronic health records. The EHR includes International Classification of Diseases billing codes which classify a comprehensive list of diseases, disorders, injuries and other health conditions. Fraction genome altered (FGA) was calculated by taking the fraction of the genome affected by gains or losses identified by GATK DepthOfCoverage (29). Microsatellite instability (MSI) was calculated by taking the percentage of unstable somatic microsatellite sites from the total number of microsatellite sites identified by MSIsensor (18). Tumor mutation burden (TMB) was calculated by dividing the number of mutations by the total bases sequenced by the panel used, per Mbp. These annotations have been widely used on MSK-IMPACT data previously (17, 18).

Exercise dose

Patients reported post-diagnosis exercise as weekly minutes of moderate and vigorous activity using the Godin Leisure-Time Exercise Questionnaire (GLTEQ) (30) or International Physical Activity Questionnaire (IPAQ) (31). Both surveys are extensively validated. In scenarios in which patients completed more than one survey, we used the survey closest to the date of pathology (median=399 days from pathology). The date of diagnosis is equivalent to the date of pathology for the sample profiled by the MSK-IMPACT. Self-reported exercise is stable from several years prior diagnosis to several years post diagnosis and treatment. Therefore, assessment of post-diagnosis exercise is a good proxy of pre-diagnosis exercise, the period of cancer development. Minutes of weekly moderate and vigorous activity were capped at 240 minutes (4 hours) per week each. Exercise intensity was then weighted by an estimate of the 5 metabolic equivalent of task (MET) for moderate exercise and 9 MET for vigorous exercise (32). The weighted estimates were summed to yield a maximum total of 56 MET hours per week (MET-hours/week).

Data processing and filtering

Only patients in the MSK-IMPACT dataset with exercise annotation were included in our cohort (n=5,150). Tumor-type-specific analyses were only performed on solid cancer types with at least 75 patients. Hypermutated (TMB>100) samples were excluded from analyses not involving TMB. The only model in which hypermutated samples were included was the adjusted linear regression of TMB as a function of exercise dose. Similarly, MSI unstable (MSI>10) samples were excluded from analyses involving not involving MSI. The only model in which MSI unstable samples were included was the adjusted linear regression of MSI as a function of exercise dose. TMB, FGA and MSI were treated as continuous variables. Mutational signature calls were treated as a categorical variable (detected/undetected). Gene mutation status was treated as a categorical variable (mutant/wildtype). Exercise dose was treated as a continuous variable (MET-hours/week) for all

analyses except for survival analyses, for which exercise dose was modeled as a categorical variable (exercisers: >0 MET-hours/week, non-exercisers: 0 MET-hours/week). Modeling of the association between gene mutations and exercise dose was only performed for genes mutated in at least 5% of the samples. Gene-level SNV, CNA and SV calls were obtained directly through the cBioPortal for Cancer Genomics. Genes were considered to have a compound mutation if they had any combination of two or more mutation types (SNV, CNA, SV).

Univariate analysis

Correlations between continuous variables were evaluated using Spearman's correlation. Comparisons between continuous and categorical variables were evaluated using Welch's two-sided t-tests. Cancer type was the only categorical variable with more than two categories, thus we used analysis of variance to evaluate comparisons between continuous variables and cancer type. Independence between categorical variables was evaluated using Pearson's X2 test of independence.

Multiple variable analysis

Models of the association between exercise and tumor genomic features were adjusted for age, survey and cancer type-specific covariates (Supplemental Table 2). For pan-cancer analyses, we also adjusted for cancer type, aggregating any cancer types with n<10 as "Other".

Association between TMB/FGA/MSI and exercise dose were evaluated using adjusted linear regression (ALR) based on the model:

TMB/FGA/MSI ~ exercise dose + age + survey + cancer type-specific covariates

Association between mutational signatures and exercise dose were evaluated using adjusted Firth's logistic regression (AFLR) based on the model:

Detection of mutational signature ~ exercise dose + age + survey + cancer type-specific covariates

Association between gene mutations and exercise dose were evaluated using adjusted Firth's logistic regression (AFLR) based on the model:

Gene mutation status ~ exercise dose + age + survey + cancer type-specific covariates

ALR exercise dose coefficient estimates were scaled from -1 to 1 independently for each outcome for visualization. AFLR exercise dose coefficients were transformed to log10(odds) for visualization.

Mutational signature calling

Data for calling mutational signatures was processed using SigProfilerMatrixGeneratorR (v1.2.13) (33). Mutational signatures were then called using SigProfilerExtractor (v1.1.20) using the following parameters: reference_genome='GRCh37', maximum_signatures=10, nmf_replicates=100, min_nmf_iterations=1000, max_nmf_iterations=200000, nmf_test_conv=1000, nmf_tolerance=1e-8 (34). Only a subset of 9 mutational signatures called in more than one cancer type or significantly affected by exercise dose were visualized. Statistical results for all other mutational signature calls are provided in Supplemental Table 5.

Sensitivity analysis

Although our regression models were already adjusted for clinico-epidemiologic features, we wanted to evaluate the consistency of our overall results between clinico-epidemiologic strata. Analyses of genomic features were followed up with sensitivity analyses. Each cancer type cohort

was dichotomized into two groups for each clinico-epidemiologic feature. Age was dichotomized at age 50, sex was dichotomized into female and male, BMI was dichotomized at BMI 25, tumor site was dichotomized into primary and metastatic, smoking was dichotomized into no smoking history and smoking history, alcohol was dichotomized into no alcohol use and alcohol use and survey was dichotomized into GLTEQ and IPAQ. For evaluating cancer subtypes, the breast cancer cohort was stratified by hormone receptor (HR) status and human epidermal growth factor receptor 2 (HER2) receptor status. HR-positive tumors were defined as tumors positive for either estrogen receptor (ER) or progesterone receptor (PR). The NSCLC cohort was stratified by adenocarcinoma (AdenoCA) and squamous cell carcinoma (SCC). Adjusted regression was performed on each strata following the multiple variable analysis framework aforementioned. Whenever the feature being stratified was also included in the overall multiple variable model, we removed the covariate from the model to avoid perfect multicollinearity. For example, the overall model for evaluating the association between FGA and exercise dose in breast cancer was: FGA ~ MET-hours/week + age + survey + alcohol history. When performing sensitivity analysis by stratifying the breast cancer cohort by age, the age covariate was removed, and the model was: FGA ~ MET-hours/week + survey + alcohol history. When performing sensitivity analysis by stratifying the breast cancer cohort by receptor status, the model remained unchanged. Exercise dose coefficient estimates were scaled from -1 to 1 for visualization.

Survival analysis

Difference in overall survival between exercisers and non-exercisers was evaluated using both univariate and adjusted Cox proportional hazards regression. Overall survival was defined as time from date of pathology to date of death or last follow up. Exercisers were defined as >0 MET-hours/week, and non-exercisers as 0 MET-hours/week. The univariate model was:

Survival ~ exercise classification

The adjusted regression was fit as:

Survival ~ exercise classification + age + survey + tumor site + cancer type-specific covariates

Hazard ratios were transformed to log2(HR) for visualization.

Statistical analysis and data visualization software

All statistical analyses were performed using R (v4.1.1). Linear regression was performed using stats (v4.1.1). Firth's logistic regression was performed using logistf (v1.24.1). Survival analysis was performed using survival (v3.2.11). Data visualization was performed using BPG (v5.3.4) (35).

Data availability

The MSK-IMPACT Clinical Sequencing Cohort dataset can be accessed through the cBioPortal for Cancer Genomics (https://www.cbioportal.org/study/summary?id=msk_impact_2017). All statistical results are provided in Supplemental Table 5.

Source data are provided with this paper.

References

1. V. Perduca et al., Stem cell replication, somatic mutations and role of randomness in the development of cancer. Eur J Epidemiol 34, 439-445 (2019).
2. R. Noble, O. Kaltz, L. Nunney, M. E. Hochberg, Overestimating the Role of Environment in Cancers. Cancer Prev Res (Phila) 9, 773-776 (2016).
3. S. Wu, S. Powers, W. Zhu, Y. A. Hannun, Substantial contribution of extrinsic risk factors to cancer development. Nature 529, 43-47 (2016).
4. N. A. Ashford et al., Cancer risk: role of environment. Science 347, 727 (2015).
5. C. Tomasetti, B. Vogelstein, Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science 347, 78-81 (2015).
6. M. Song, E. Giovannucci, Preventable Incidence and Mortality of Carcinoma Associated With Lifestyle Factors Among White Adults in the United States. JAMA Oncol 2, 1154-1161 (2016).
7. L. B. Alexandrov et al., Mutational signatures associated with tobacco smoking in human cancer. Science 354, 618-622 (2016).
8. L. B. Alexandrov et al., The repertoire of mutational signatures in human cancer. Nature 578, 94-101 (2020).
9. H. K. Seitz, F. Stickel, Molecular mechanisms of alcohol-mediated carcinogenesis. Nat Rev Cancer 7, 599-612 (2007).
10. C. P. Wen et al., Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. Lancet 378, 1244-1253 (2011).
11. J. A. Lavery et al., Association of exercise with pan-cancer incidence and overall survival. Cancer Cell, (2023).

12. S. C. Moore et al., Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults. JAMA Intern Med 176, 816-825 (2016).
13. G. J. Koelwyn, D. F. Quail, X. Zhang, R. M. White, L. W. Jones, Exercise-dependent regulation of the tumour microenvironment. Nat Rev Cancer 17, 620-632 (2017).
14. K. L. Campbell et al., Exercise Guidelines for Cancer Survivors: Consensus Statement from International Multidisciplinary Roundtable. Med Sci Sports Exerc 51, 2375-2390 (2019).
15. K. B. Borch, T. Braaten, E. Lund, E. Weiderpass, Physical activity before and after breast cancer diagnosis and survival - the Norwegian women and cancer cohort study. BMC Cancer 15, 967 (2015).
16. C. Mason et al., Long-term physical activity trends in breast cancer survivors. Cancer Epidemiol Biomarkers Prev 22, 1153-1161 (2013).
17. B. Nguyen et al., Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. Cell 185, 563-575 e511 (2022).
18. A. Zehir et al., Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 23, 703-713 (2017).
19. S. Negrini, V. G. Gorgoulis, T. D. Halazonetis, Genomic instability--an evolving hallmark of cancer. Nat Rev Mol Cell Biol 11, 220-228 (2010).
20. B. Niu et al., MSIsensor: microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics 30, 1015-1016 (2014).
21. C. Valero et al., The association between tumor mutational burden and prognosis is dependent on treatment context. Nat Genet 53, 11-15 (2021).

22. R. M. Samstein et al., Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet 51, 202-206 (2019).

23. E. Schwitzer et al., No association between prediagnosis exercise and survival in patients with high-risk primary melanoma: A population-based study. Pigment Cell Melanoma Res 30, 424-427 (2017).

24. M. Ferioli et al., Role of physical exercise in the regulation of epigenetic mechanisms in inflammation, cancer, neurodegenerative diseases, and aging process. J Cell Physiol 234, 14852-14864 (2019).

25. I. L. Gomes-Santos et al., Exercise Training Improves Tumor Control by Increasing CD8(+) T-cell Infiltration via CXCR3 Signaling and Sensitizes Breast Cancer to Immune Checkpoint Blockade. Cancer Immunol Res 9, 765-778 (2021).

26. H. Rundqvist et al., Cytotoxic T-cells mediate exercise-induced reductions in tumor growth. Elife 9, (2020).

27. C. C. Maley et al., Classifying the evolutionary and ecological features of neoplasms. Nat Rev Cancer 17, 605-619 (2017).

28. D. T. Cheng et al., Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): A Hybridization Capture-Based Next-Generation Sequencing Clinical Assay for Solid Tumor Molecular Oncology. J Mol Diagn 17, 251-264 (2015).

29. A. McKenna et al., The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20, 1297-1303 (2010).

30. G. Godin, R. J. Shephard, A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 10, 141-146 (1985).

31. C. L. Craig et al., International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 35, 1381-1395 (2003).

32. B. E. Ainsworth et al., Compendium of physical activities: an update of activity codes and MET intensities. Med Sci Sports Exerc 32, S498-504 (2000).

33. E. N. Bergstrom et al., SigProfilerMatrixGenerator: a tool for visualizing and exploring patterns of small mutational events. BMC Genomics 20, 685 (2019).

34. S. M. A. Islam et al., Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genom 2, None (2022).

35. C. P'ng et al., BPG: Seamless, automated and interactive visualization of scientific data. BMC Bioinformatics 20, 42 (2019).

CHAPTER 2:

Personalized Dynamic Mapping of Integrative Response to Exercise Therapy in Cancer

Abstract

Regular exercise confers significant health benefit but full characterization of the system-wide and tissue-specific responses in humans have not been performed. We leveraged a decentralized, digital platform approach to perform longitudinal, multiparametric profiling of host physiology (exercise capacity, body composition, continuous glucose), plasma (metabolome, proteome), gut microbial composition (metagenomics), and tumor tissue (transcriptome) in 13 patients with solid tumors, before, during, and after high-controlled exercise therapy in a preoperative "window" study. Time-series analyses revealed hundreds of host molecular changes in the plasma proteome and metabolome and gut microbiome involved in a diverse array of biological processes. System-wide changes were paralleled by modulation of core tumor gene expression pathways notably tumor cell cycle regulation, stress response, and metabolism. Integrative network analyses revealed the complexity of the host–tumor interaction under exercise therapy regulation, elucidating novel mechanistic insights. Variability at baseline and in response to treatment emphasized highly personalized responses to uniform exercise therapy. Our study provides an example of the application of a digital approach to generate a longitudinal high-definition dataset providing a framework of the integrative effects of exercise therapy with considerable translational and discovery potential.

Introduction

Exercise is a holistic strategy that influences multiple processes at the organismal (host), tissue and cellular levels (1). It is postulated that exercise-induced cumulative host-level adaptations, the product of highly integrated cellular and tissue-specific alterations, establish a higher homeostatic

"set point," enhancing performance (healthspan), tolerance to system perturbation (resilience), and longevity (2). Consequently, regular exercise associates with reduced risk of a wide variety of non-communicable diseases, and improves clinical outcomes after diagnosis of such conditions (3).

Over the past decade, preclinical studies have increased our understanding of the molecular and cellular mechanisms underpinning the beneficial effects of exercise in health and aging,4-6 and in the context of disease states such as cancer (7, 8), depression (9), Alzheimer's (10), and cardiovascular disease (11). Comparable investigation in humans is limited. Most "mechanistic" investigation of exercise in humans is confined to host-level analysis via peripheral blood, typically with measurement of a limited number of factors within an isolated molecular process, such as metabolism or inflammation (12-14). Analysis is also conducted at low sampling frequency, typically two timepoints (e.g. pre- and post-intervention). Tissue-level interrogation is almost exclusively confined to skeletal muscle (2, 14). Whether and how host-level alterations link with regulation of distal tissues not centrally involved in the acute cardiovascular response to exercise, like the breast, prostate and lung remains enigmatic (2, 14). Overall, we have limited understanding of host, tissue, and cellular alterations, and their interaction, that occur in response to chronic exercise therapy in humans.

We performed time-series, multiparametric profiling of host-tissue (tumor) molecular response to short-term exercise therapy in patients with cancer.

Results

Study design, patients, and exercise therapy

Thirteen inactive (<90 minutes of moderate or vigorous exercise per week) patients with treatment-naïve breast, endometrial, or prostate cancer scheduled for surgical resection were enrolled into a pre-operative "window of opportunity" single-arm (non-randomized), prospective study (Table S1). The pre-operative setting permits longitudinal interrogation of exercise therapy on tumor tissue in the absence of any other form of concurrent anticancer therapy (15).

A detailed description of the study methods has been reported previously and provided in the Supplemental Materials (16). In brief, this was a fully decentralized clinical study: all study procedures were conducted remotely in patients' homes. All patients were shipped a study "kit" containing an etablet, treadmill and several Bluetooth-enabled health devices. Health devices included a smartwatch (mobility and heart rate measurement every 10 minutes, 24/7), a blood pressure monitor (daily measurement), a body composition scale (daily measurement), and continuous glucose monitoring (CGM) (interstitial glucose measurement every 15 minutes, 24/7). CGM was paired with oral glucose tolerance testing at pre- and post-intervention. Dietary intake was evaluated for 1-3 days at the beginning and end of the intervention using real-time monitoring via a dietary mobile application. Finally, a remotely-administered exercise tolerance test was performed in patients' homes with real-time monitoring at pre- and post-intervention (Figure 1A-B). The final dataset contained a total of 69,606 lifestyle state measures and 100,447 physiologic measures, or an average of 342 measures per day per patient.

Exercise therapy comprised of treadmill walking sessions five times weekly for 3 to 12 consecutive weeks, depending on the available pre-surgical window (16). All sessions were conducted at ≈70% of the individual patient's measured submaximal exercise capacity for ≈30 minutes per session to yield the recommended 150 minutes per week dose (17). Mean length of exercise therapy was 5.5

weeks (range: 3 to 12 weeks) or mean of 23 (range: 14 to 48) unique exercise therapy sessions per patient. Mean compliance was 87% (293 sessions attended of 336 planned).

All sessions were performed in the patients' primary residence with remote, real-time, one-on-one supervision and monitoring by study exercise physiologists using two-way video conferencing. Serial plasma collection was performed weekly via a mobile phlebotomy solution. Stool was also sampled weekly and returned via a pre-paid courier service. Finally, we acquired cancer tissues before (from the diagnostic biopsy) and after (from surgical resection) exercise therapy for tumor molecular characterization (Figure 1A-B).

Lifestyle and physiological changes

We first evaluated alterations in lifestyle and physiological outcomes during a representative 24-hour period in patient 010 on day 3 of study participation. Study devices profiled diurnal and nocturnal patterns as well as accurately captured physiological changes in response to daily activities or perturbations over a 24-hr period (Figure 1C). For example, in Figure 1C, CGM showed higher interstitial glucose during a glucose challenge (mean 140 mg/dL, peak 180 mg/dL; vertical green shaded area, hours ≈7 to 10) and heart rate during an exercise therapy session (mean 96 bpm, max 114 bpm; vertical red shaded area, hour ≈15), compared to rest (mean 95 mg/dL and mean 67 bpm). Evaluation of daily lifestyle states enabled time-series profiling across the entire study period for all patients. Representative lifestyle "heatmap" for patient 010 is presented in Figure 1D.

Patient diurnal and nocturnal patterns were characterized by substantial day-to-day and between-patient variability. Variance analysis indicated that between-patient variability accounted for 30%, 46%, 23% of the total variance for daily active, sedentary, and sleep time, respectively. There was

no significant within-patient temporal variability; no cohort-level trend was detected for any state as a function of exercise therapy length ($p < 0.05$; Figure 2A). No significant changes in total caloric intake or macronutrient intake was observed at the cohort level during the intervention (Figure 2B). Dynamic, time series analyses revealed patient physiological response was characterized by day-to-day variability, with cohort-level linear decreases for fat mass ($p < 0.001$, daily effect size = –0.015 standard deviations), total mass ($p < 0.001$, daily effect size -0.012 standard deviations), and linear increases for muscle mass ($p = 0.028$, daily effect size = 0.007) and resting heart rate ($p = 0.002$, daily effect size = 0.009 standard deviations) (Figure 2C). Exercise capacity increased from pre- to post-intervention (Figure 2D). No cohort-level changes were observed for OGTT (Figure 2E). Overall, alterations in covariate lifestyle states during exercise therapy were minimal but physiological improvements were observed in several outcomes; this study therefore provides an appropriate context to interrogate host and tissue molecular response to exercise therapy.

Time-series plasma molecular alterations during exercise therapy

To evaluate longitudinal molecular changes in circulating plasma analytes, we performed proteomic and metabolomic profiling on weekly plasma samples. After filtering and data processing, the final dataset contained a total of 5,811 circulating analytes (5,667 proteins and 144 metabolites) per timepoint per patient (Supplemental Figure 1). The total number of data points measured was 383,526; the mean total data points per patient was 31,960.5 (mean of 5.5 timepoints per patient). We evaluated longitudinal alterations during exercise therapy using linear modeling to determine whether circulating analyte trajectories differed from baseline. Cohort-level modeling identified 132 proteins (2.33%; Figure 3A, Supplemental Figure 2A) and 5 metabolites (3.47%; Figure 3B, Supplemental Figure 2B) with significantly altered trajectories. Overall, cohort-level

alterations were modest in comparison with weekly variability of analyte abundance and between-patient heterogeneity (Supplemental Figures 3-4). We therefore used Fisher's method of combining p-values to prioritize patient-level proteins (Figure 3D) and metabolites (Figure 3E) significantly altered in at least one patient (Fisher's p-value < 0.05). Top-hit proteins were involved in regulation of numerous biological processes including iron storage and metabolism, DNA recognition and regulation of apoptosis, neuronal migration, interleukin-6 cytokine signaling, and muscle contraction. Top-hit metabolites were involved in neurotransmission and thymine breakdown. Of note, several proteins shared similar trajectory patterns across all patients. For instance, UNC-45A, CDK5, OSM, and TPM2 decreased during exercise therapy, whereas HLA, in general, increased. Similar response trajectories were generally not observed for any top-hit metabolites. As a consequence of between-patient heterogeneity, the number of significantly altered analytes varied, ranging from 74 to 308 proteins and 5 to 22 metabolites.

We next evaluated longitudinal trajectory clusters to evaluate cross-talk between the plasma proteome and metabolome using fuzzy c-means clustering. Proteins and metabolites co-clustered, distributing across five clusters with distinct trajectory patterns. As expected, cohort-level clustering revealed modest (static) directional trajectories relative to patient-level clustering (Figures 3E-F, Supplemental Figure 5). We then performed pathway enrichment analysis on each cluster. Major pathways perturbed in response to exercise therapy included immune response, metabolism, and signal transduction (Supplemental Figure 6). Again, considerable between-patient heterogeneity was apparent. Several patients had multiple clusters enriched with the same pathway; for instance, patient 012 had three clusters enriched for adaptive immunity; cluster 1 was static, cluster 2 had a positive trajectory, and cluster 3 had a negative trajectory. Single clusters were also enriched for multiple pathways, such as the inverted U-shaped cluster 4 in patient 018.

While multiple clusters, both within and between patients, may have been enriched for a shared pathway, the clusters did not necessarily have similar trajectory patterns, demonstrated by the various patterns in clusters enriched for post-translational modification. Taken together, our high-resolution, longitudinal sampling revealed highly personalized system-wide host molecular responses to short-term exercise therapy.

Gastrointestinal microbiome changes during exercise therapy

Eight patients consented to conduct optional weekly stool collection, with a median four timepoints per patient (range: 3 to 9). We performed stool microbiome profiling via 16S-sequencing, with a median of 29,936 informative reads per sample (range: 10,984 to 361,007; Supplemental Figure 7A). We observed 684 Amplicon Sequence Variants, a median of 133 per sample (range: 60 to 313), after agglomerating to the rank of Species (Figure 4A). Considerable variability between patients at baseline was observed, with 55% of Species being shared in no more than two patients (Supplemental Figure 7B). Alpha-diversity, or within-sample diversity, as measured by the Shannon index varied across samples with a mean of 3.62 (standard deviation 0.37), with no significant linear change as a function of exercise therapy at the cohort level (Supplemental Figure 8A). Beta-diversity, or between-sample diversity, as reflected by double principal coordinates analysis showed within-patient microbial composition remained stable across different timepoints during the exercise therapy intervention (Supplemental Figure 8B), and differences between patients were stable due to taxa from the Phyla Firmicutes and Bacteroidetes (Supplemental Figure 8C). As a consequence of between-patient variability, no taxa were significantly altered by exercise therapy at the cohort level. Conversely, patient-specific linear abundance changes in microbial taxa were observed, with a median of 1 taxa (range: 0 to 11) altered per patient (Figure 4B-C). Multiple patients shared significant hits from the Families

Lachnospiraceae and Ruminococcaceae, including increases in known commensal gut microbes such as Blautia faecis (18) (006), Monoglobus pectinilyticus (19) (008), and Faecalibacterium prausnitzii (20) (017), as well as Roseburia intestinalis (21) (Fisher's FDR < 0.05) (Figure 4B-C). Overall, short-term exercise therapy induces highly individualized alterations in the gut microbiota among patients with cancer.

Tumor and microenvironmental evolution during exercise therapy

To identify changes indicative of a "pharmacodynamic response" to exercise therapy, we performed transcriptomics on pre- and post-intervention tumor tissue, identifying 743 differentially abundant genes (296 upregulated, 447 downregulated) (FDR < 0.05) (Figure 5A). Gene ontology revealed downregulated genes were not enriched for any pathways; upregulated genes were preferentially involved in GTPase activity, transcription factor AP-1 complex and regulation of cell growth (FDR < 0.05). Gene set enrichment analysis revealed upregulated genes were also involved in cell growth and energy metabolism: oxidative phosphorylation, adipogenesis and myogenesis, as well as epithelial mesenchymal transition (EMT) (FDR < 0.25, Figure 5B). Interestingly, tumors from both patients with endometrial cancer exhibited lower hypoxia in response to exercise therapy, but this was not observed in prostate cancer (Figure 5C). Taken together, our data highlights the highly personalized, multi-faceted response of cancer to exercise therapy, which could be influenced by tumor location and cancer cell histology.

Integrative (systems) network analysis

We applied information-theory-based mutual information network analysis, a systems-based approach, to our high-sampling frequency, integrated dataset to characterize exercise therapy-dependent regulation of the host-tumor interaction. Highly connected, conserved co-regulation

patterns were identified across patients (Figure 6A). As expected, the plasma proteome and metabolome displayed highly correlated patterns (FDR < 0.01). Pyruvic acid, a metabolite established to increase after exercise therapy, (22) was among the most interconnected plasma metabolites in our cohort-level network. The detection of cytosine as another highly connected metabolite suggests epigenetics, specifically DNA methylation, may be an important coordinator of exercise therapy response (23). The most interconnected plasma proteins were myosin-binding protein C2, a protein highly expressed in cardiac muscle, (24) and proenkephalin, peptides active in the endocrine and nervous systems (25). Gut microbial Amplicon Sequence Variants abundance was highly correlated with the host plasma metabolome and proteome (FDR < 0.01). Intriguingly, patterns involving Akkermansia muciniphila, a common gut microbial species demonstrated to potentiate the antitumor efficacy of chemotherapy and immunotherapy, (26, 27) was prominent, as was Phascolarctobacterium faecium, another species implicated in immunotherapy response (28). The tumor molecular feature most highly inter-connected was epithelial-to-mesenchymal transition (EMT) (FDR < 0.01). Patient-specific network analysis revealed highly personalized interplay between host and tumor features (Figure 6B). A pronounced interconnected network was apparent for 008. Changes in plasma levels of SERPINA12, acetylcholine, and tumor EMT were associated with exercise therapy dose (compliance) and alterations in sedentary time, resting heart rate, and, most predominantly, fat mass. This highlights a scenario in which fat mass alterations appear a patient-specific “central hub” of host-tumor regulation during exercise therapy. Our findings reveal the complex, reveal highly-integrated and personalized cross-talk between physiology, lifestyle states, host molecular alterations and tumor cell phenotypes in response to short-term exercise therapy.

Discussion

Cancer evolution is regulated by interplay between tumor cell-intrinsic, cell-extrinsic (TME), and systemic (host) features. These components, and their interactions, are in turn regulated by host-level perturbations such as aging, age-related comorbidities, and lifestyle (29-31). In essence, cancer is an ecosystem. Hence, understanding how a whole-body (holistic) strategy such as exercise therapy regulates the cancer ecosystem requires a broad integrative approach to samples longitudinally collected before, during, and after treatment intervention at the host, TME, and tumor cell intrinsic levels.

We leveraged digitized clinical trial methodology to perform dense, personalized, longitudinal phenotyping of the integrative response to high-fidelity exercise therapy in patients with cancer. High-frequency sampling of the plasma proteome and metabolome together with parallel sequencing of gut microbial composition revealed novel insights into the magnitude and nature of host (system)-wide molecular alterations occurring during exercise therapy. We also showed system-wide changes were paralleled by regulation of key tumor cell-intrinsic biological processes including tumor cell cycle regulation, stress response, and metabolism. Integrative network analyses revealed the complexity of the host-tumor interaction under chronic exercise therapy, elucidating novel mechanistic insights. Finally, between-patient variability at baseline and in response to treatment underscored that integrative response to a uniform exercise therapy regimen is highly personalized. Our study provides an example of leveraging digital technology solutions to conduct a fully decentralized trial integrating multiparametric passive data collection and remote biospecimen collection at high resolution to generate a comprehensive, longitudinal, physiological and molecular map of response to exercise therapy in a clinical population, providing a framework for future studies. It is noteworthy, however, that we investigated the effects of short-term exercise

therapy in a small, and likely biased, cohort of patients mostly diagnosed with localized prostate cancer. Therefore, our findings should be considered exploratory/hypothesis-generating and must be complemented by future work investigating exercise therapy over the longer-term in larger patient cohorts.

Better understanding of the molecular mechanisms how exercise therapy influences disease pathogenesis will ultimately require direct interrogation of effects at the tissue/organ site of disease manifestation. This is a significant challenge. Certain oncology settings, however, enable longitudinal sampling of tumor tissue and possibly non-malignant tissue via biopsy procedures performed as part of standard of care. The neoadjuvant/pre-operative context is one such setting, permitting investigation of exercise therapy as "single-agent" (i.e., without concurrent anticancer therapy) but also presents its own challenges due to the short treatment window and high patient burden (32). Indeed, to our knowledge, only one other trial has leveraged this setting to evaluate exercise therapy regulation of tumor biology. Ligibel et al (33). reported that exercise therapy (target dose of 220 minutes per week of resistance and endurance exercise) did not impact tumor cell proliferation or apoptosis compared with control in primary breast cancer; whole-tumor transcriptomics revealed exercise therapy enriched several gene expression pathways with many being immune related (e.g., NF-kappa ß signaling, natural killer cell mediated cytotoxicity, and T cell receptor signaling). Secondary analysis using established deconvolution approaches confirmed exercise increased intratumoral content of macrophages and B cells (34). Interestingly, the proinflammatory-immune cell infiltration effect of exercise was recapitulated in animal models of breast cancer; exercise-inhibition of breast cancer growth was CD8+ T-cell dependent and sensitized immunotherapy-refractory breast cancers to treatment (8). These data are distinct to our own findings showing exercise modulated pathways involved in diverse tumor biological

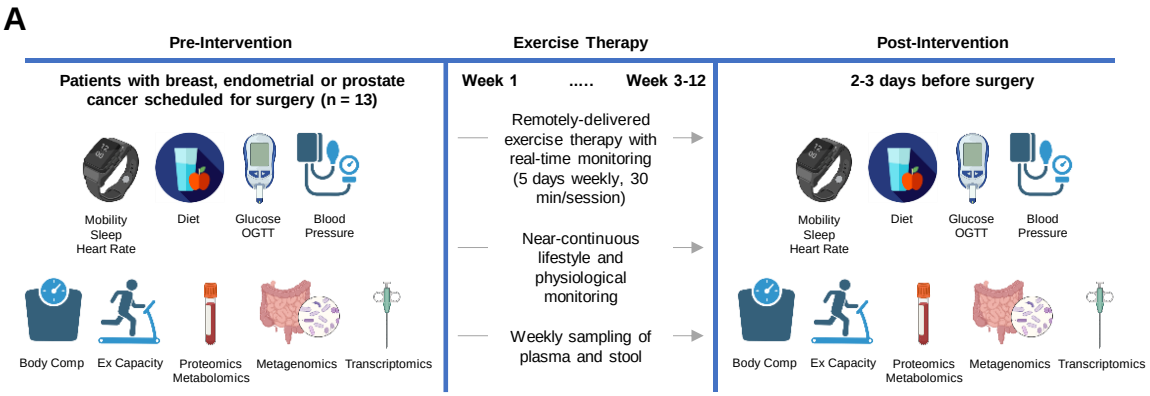
processes, but limited effects on inflammation and immunity. Our findings, together with those of Ligibel et al. (33), provide initial evidence that short-term exercise therapy confers marked alterations in the biology of several common solid tumors, but effects may be influenced by tumor location and histological cancer type.

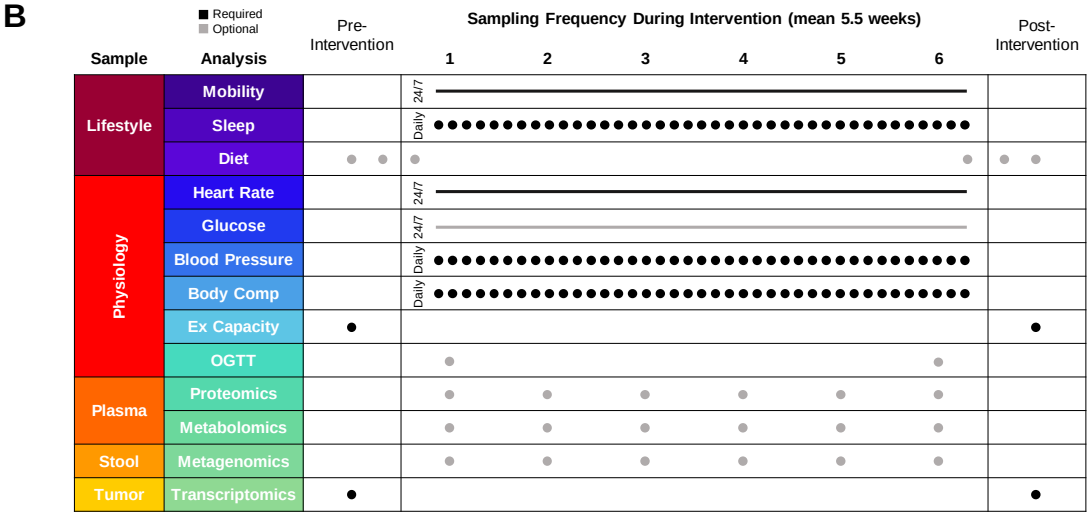
Optimal harnessing of "ecological" strategies, such as exercise therapy, as cancer preventive or treatment approaches will require, by definition, understanding of how such strategies modulate the interconnection of cancer ecosystem components. Application systems network analytics to our integrative sampling scheme aggregating longitudinal multi-parametric data at the host, tissue, and molecular level was designed specifically to facilitate preliminary investigation of this critical knowledge gap (2). Such approaches may identify the trajectory and hierarchy higher-level "nodes" in the network (31) providing mechanistic insights as well as potentially identifying key targets that can be therapeutically modulated by exercise, other therapies, or exercise-drug combination approaches. Discovery of a connection between alterations in select plasma proteins and metabolites with gut microbial species shown to mediate the antitumor efficacy of chemotherapy and immunotherapy, (26, 27) and tumor features was particularly notable. These initial findings support further testing of the hypothesis that tumor biological activity of exercise therapy is dependent on alterations in host factors. A related discovery of interest was exercise-induced alterations in host factors primarily converged on upregulation of tumor EMT. EMT is a complex development program, and partial activation of EMT via EMT-transcription factors (EMT-TFs) enhances cancer cell motility, promoting invasion and dissemination (35, 36). Intriguingly, EMT-TFs maintain stemness properties and regulate related biological processes enabling cancer cells to adapt to the selective pressures/stress conditions (e.g., physical constraints, hypoxia, inflammation) encountered during development and progression (37). Elegant preclinical

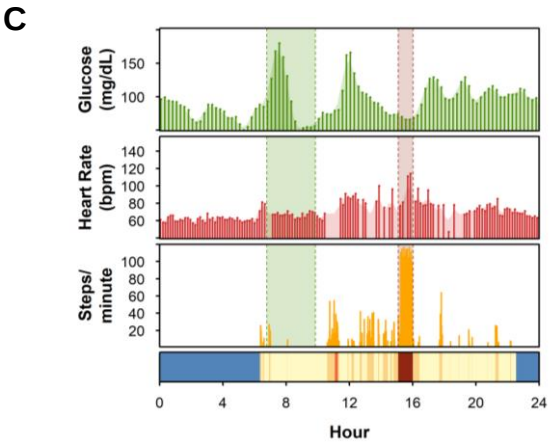
studies demonstrate exercise promotes stemness in multiple non-malignant cell types (e.g., skeletal muscle, brain, bone marrow hematopoietic stem and progenitor cells) (5) as well as modulates intratumoral vascularization and hypoxia in various solid tumors (8, 38, 39). We speculate that exercise therapy as a whole-body perturbation simultaneously alters the strength and nature of selective pressures on distal tissue niches harboring malignant and non-malignant cells, triggering highly conserved stress-response programs in an attempt to maintain/restore homeostasis. Aberrant reactivation of EMT programming generally promotes promalignant phenotypes, thus understanding how exercise therapy-induced regulation of EMT and related phenotypes influences cancer evolution will be instructive. Given the heterogeneity in the number, type, and nature of host-tumor connections in response to highly uniform exercise therapy observed in the present study, understanding the interaction between patient-specific baseline characteristics (e.g., demographics, host physiological and molecular status), exercise dose, and tumor response should be an important objective of future work.

In conclusion, our study provides an in-depth, personalized mapping of the integrative response to exercise therapy in patients with cancer. Such studies might facilitate personalized exercise prescriptions, development of exercise-mimetic pharmacologic strategies, and rational exercise – pharmacologic combination approaches to optimally harness the therapeutic potential of exercise therapy in health and disease.

Figure 1


A
Pre-Intervention
Exercise Therapy
Post-Intervention
Patients with breast, endometrial or prostate cancer scheduled for surgery (n = 13)
Week 1 ..... Week 3-12
2-3 days before surgery
Remotely-delivered exercise therapy with real-time monitoring (5 days weekly, 30 min/session)
Near-continuous lifestyle and physiological monitoring
Weekly sampling of plasma and stool
Mobility
Sleep
Heart Rate
Diet
Glucose OGTT
Blood Pressure
Body Comp
Ex Capacity
Proteomics
Metabolomics
Metagenomics
Transcriptomics


B
Required
Optional
Sample
Analysis
Pre-Intervention
Sampling Frequency During Intervention (mean 5.5 weeks)
Post-Intervention
1
2
3
4
5
6
Lifestyle
Mobility
Sleep
Diet
Physiology
Heart Rate
Glucose
Blood Pressure
Body Comp
Ex Capacity
OGTT
Plasma
Proteomics
Metabolomics
Stool
Metagenomics
Tumor
Transcriptomics
24/7
Daily


C
Glucose (mg/dL)
Heart Rate (bpm)
Steps/ minute
Hour

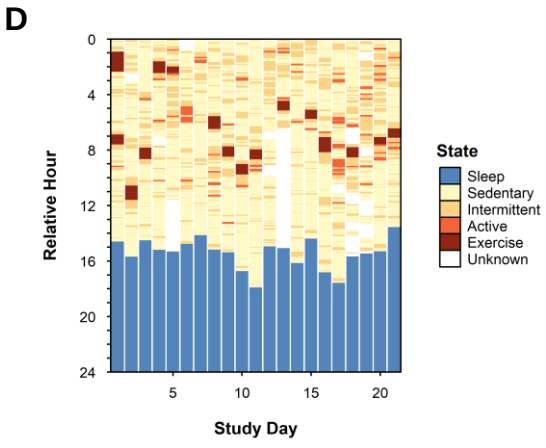

D
Relative Hour
Study Day
State
Sleep
Sedentary
Intermittent
Active
Exercise
Unknown

Figure 2

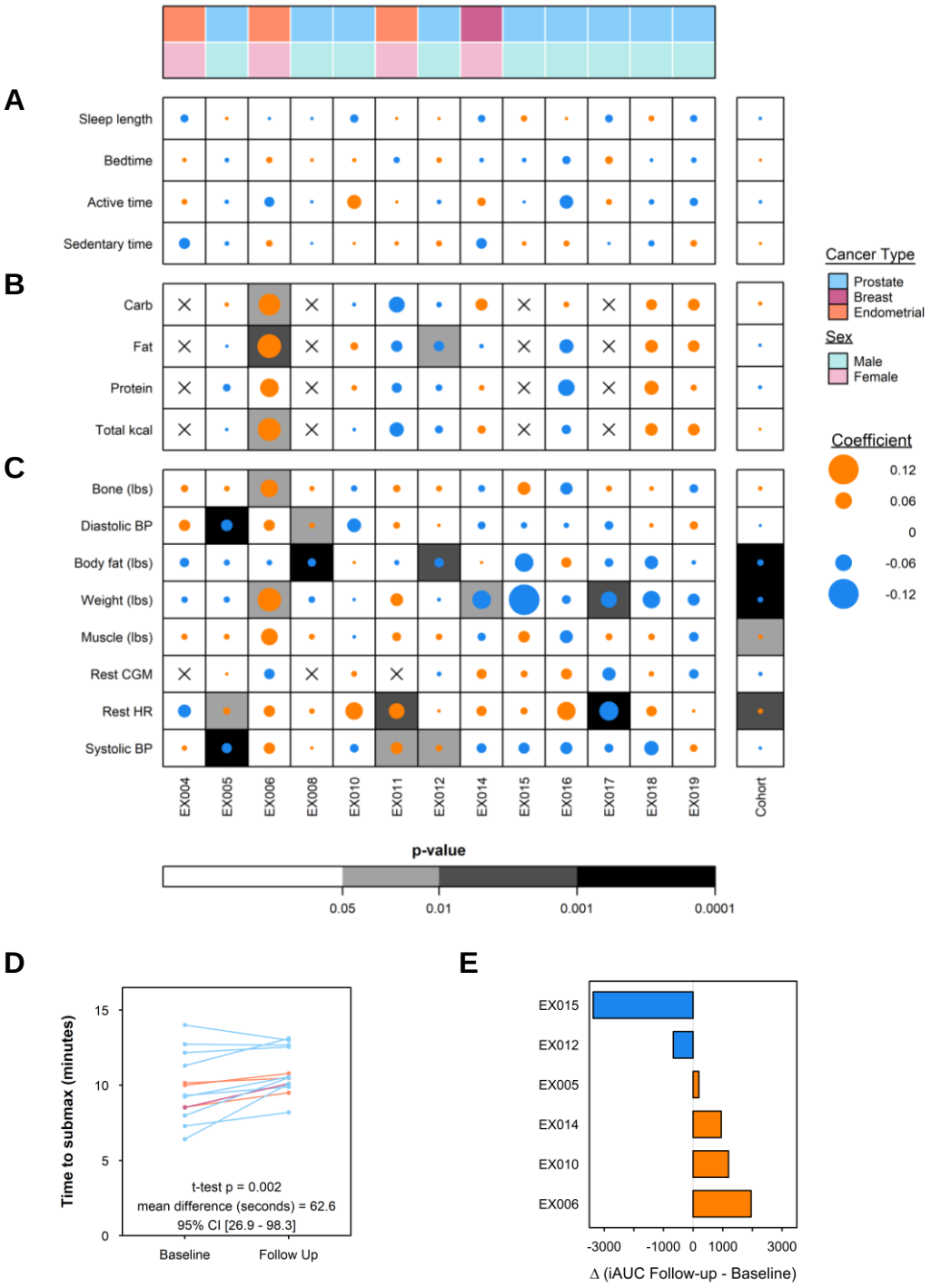

A
Sleep length
Bedtime
Active time
Sedentary time
B
Carb
Fat
Protein
Total kcal
C
Bone (lbs)
Diastolic BP
Body fat (lbs)
Weight (lbs)
Muscle (lbs)
Rest CGM
Rest HR
Systolic BP
EX004
EX005
EX006
EX008
EX010
EX011
EX012
EX014
EX015
EX016
EX017
EX018
EX019
Cohort
Cancer Type
Prostate
Breast
Endometrial
Sex
Male
Female
Coefficient
0.12
0.06
0
-0.06
-0.12
p-value
0.05
0.01
0.001
0.0001
D
Time to submax (minutes)
15
10
5
0
t-test p = 0.002
mean difference (seconds) = 62.6
95% CI [26.9 - 98.3]
Baseline
Follow Up
E
EX015
EX012
EX005
EX014
EX010
EX006
-3000
-1000
0
1000
3000
Δ (iAUC Follow-up - Baseline)

Figure 3

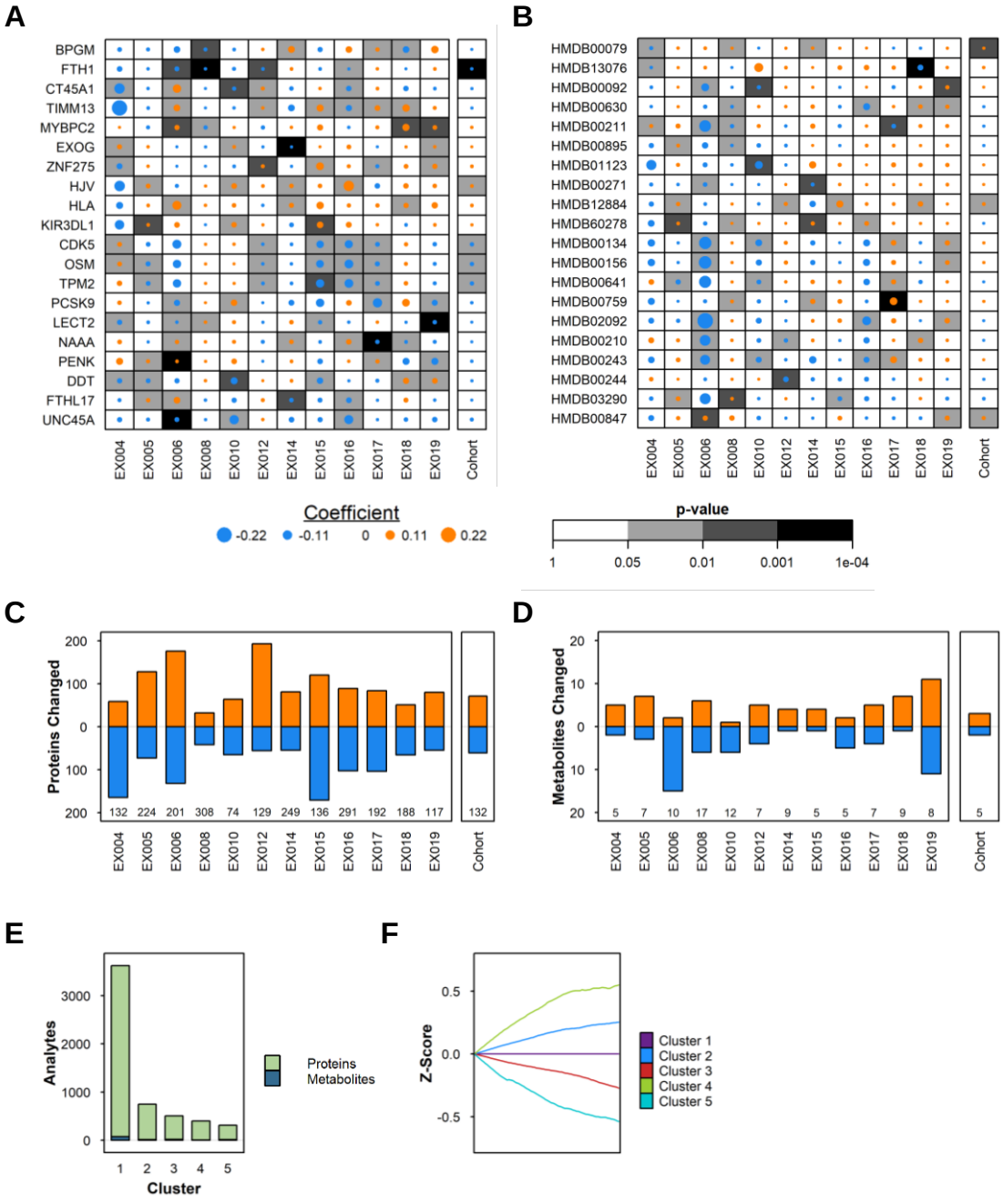

A
BPGM
FTH1
CT45A1
TIMM13
MYBPC2
EXOG
ZNF275
HJV
HLA
KIR3DL1
CDK5
OSM
TPM2
PCSK9
LECT2
NAAA
PENK
DDT
FTHL17
UNC45A
EX004 EX005 EX006 EX008 EX010 EX012 EX014 EX015 EX016 EX017 EX018 EX019 Cohort
Coefficient
-0.22 -0.11 0 0.11 0.22
B
HMDB00079
HMDB13076
HMDB00092
HMDB00630
HMDB00211
HMDB00895
HMDB01123
HMDB00271
HMDB12884
HMDB60278
HMDB00134
HMDB00156
HMDB00641
HMDB00759
HMDB02092
HMDB00210
HMDB00243
HMDB00244
HMDB03290
HMDB00847
EX004 EX005 EX006 EX008 EX010 EX012 EX014 EX015 EX016 EX017 EX018 EX019 Cohort
p-value
1 0.05 0.01 0.001 1e-04
C
Proteins Changed
200 100 0 100 200
132 224 201 308 74 129 249 136 291 192 188 117 132
EX004 EX005 EX006 EX008 EX010 EX012 EX014 EX015 EX016 EX017 EX018 EX019 Cohort
D
Metabolites Changed
20 10 0 10 20
5 7 10 17 12 7 9 5 5 7 9 8 5
EX004 EX005 EX006 EX008 EX010 EX012 EX014 EX015 EX016 EX017 EX018 EX019 Cohort
E
Analytes
3000 2000 1000 0
1 2 3 4 5
Cluster
Proteins
Metabolites
F
Z-Score
0.5 0.0 -0.5
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Cluster 5

Figure 4

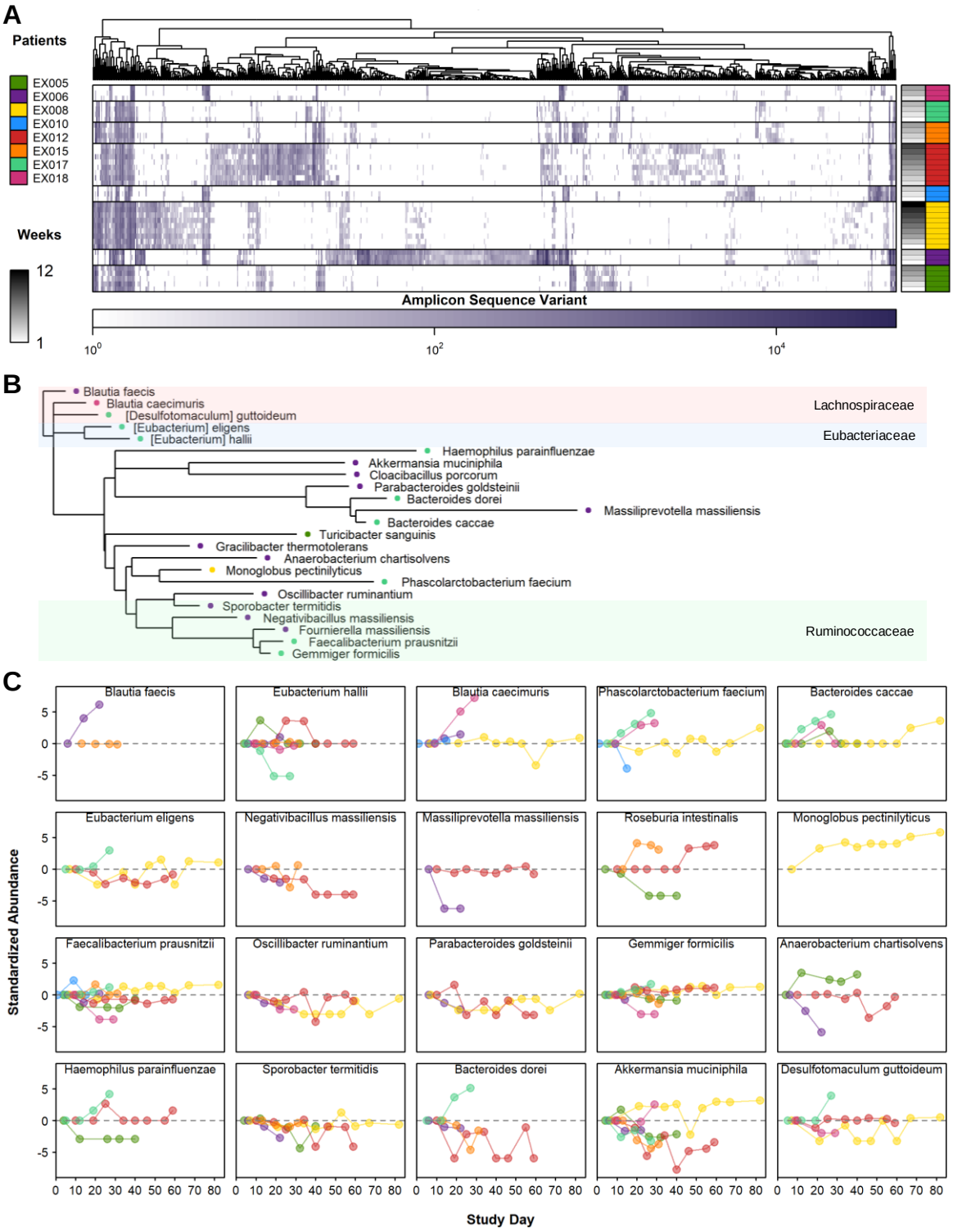

A
Patients
EX005
EX006
EX008
EX010
EX012
EX015
EX017
EX018
Weeks
12
1
Amplicon Sequence Variant
10^0
10^2
10^4
B
Blautia faecis
Blautia caecimuris
[Desulfotomaculum] guttoideum
[Eubacterium] eligens
[Eubacterium] hallii
Lachnospiraceae
Eubacteriaceae
Haemophilus parainfluenzae
Akkermansia muciniphila
Cloacibacillus porcorum
Parabacteroides goldsteinii
Bacteroides dorei
Massiliprevotella massiliensis
Bacteroides caccae
Turicibacter sanguinis
Gracilibacter thermotolerans
Anaerobacterium chartisolvens
Monoglobus pectinilyticus
Phascolarctobacterium faecium
Oscillibacter ruminantium
Sporobacter termitidis
Negativibacillus massiliensis
Fournierella massiliensis
Faecalibacterium prausnitzii
Gemmiger formicilis
Ruminococcaceae
C
Blautia faecis
Eubacterium hallii
Blautia caecimuris
Phascolarctobacterium faecium
Bacteroides caccae
Eubacterium eligens
Negativibacillus massiliensis
Massiliprevotella massiliensis
Roseburia intestinalis
Monoglobus pectinilyticus
Faecalibacterium prausnitzii
Oscillibacter ruminantium
Parabacteroides goldsteinii
Gemmiger formicilis
Anaerobacterium chartisolvens
Haemophilus parainfluenzae
Sporobacter termitidis
Bacteroides dorei
Akkermansia muciniphila
Desulfotomaculum guttoideum
Standardized Abundance
5
0
-5
0 10 20 30 40 50 60 70 80
Study Day

Figure 5

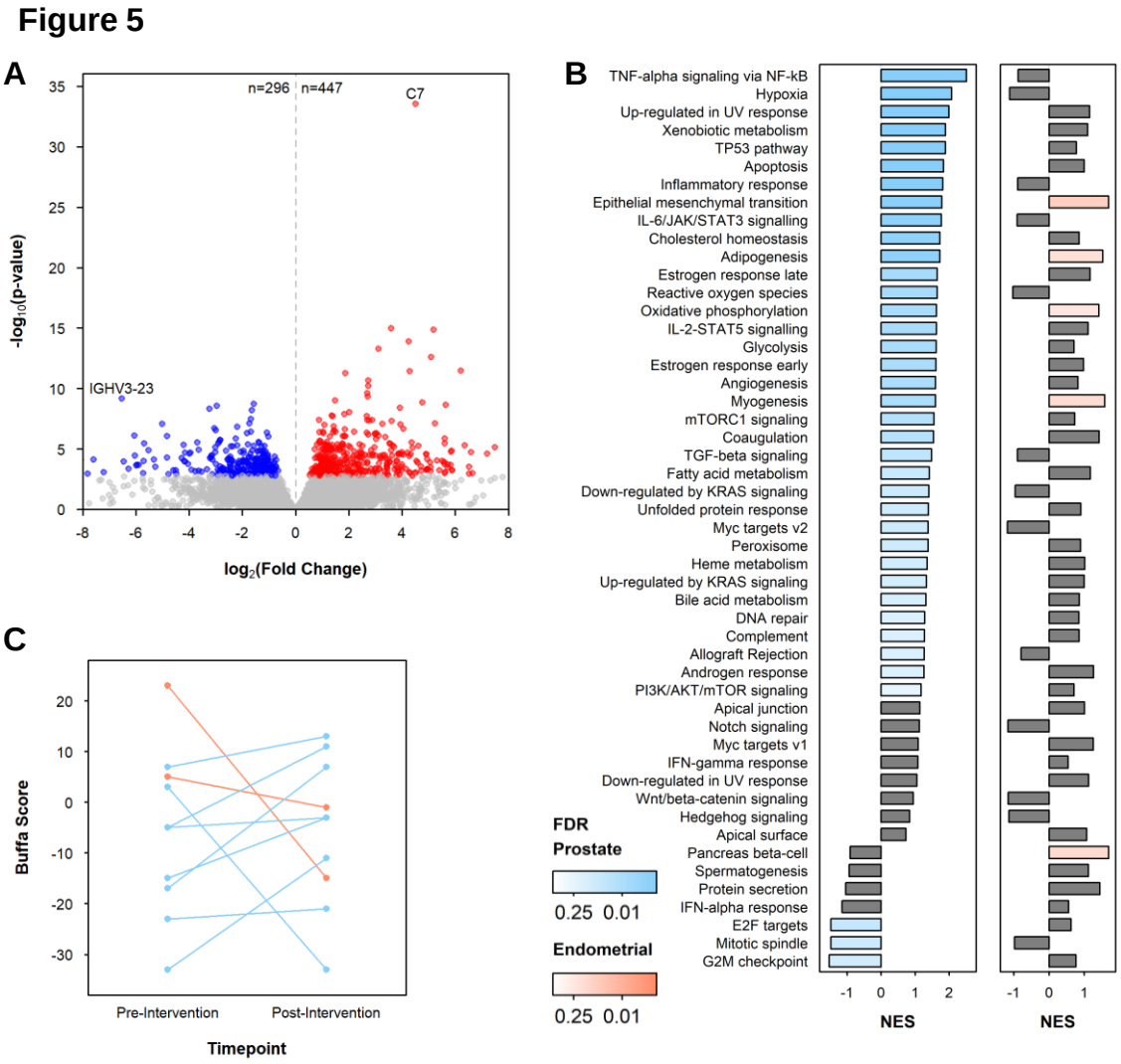
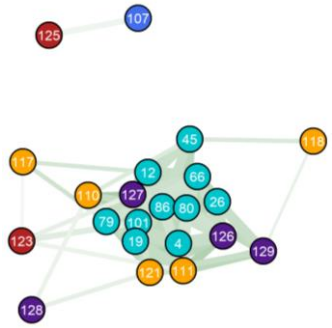


Figure 5
A
n=296
n=447
C7
IGHV3-23
-log10(p-value)
log2(Fold Change)
B
TNF-alpha signaling via NF-kB
Hypoxia
Up-regulated in UV response
Xenobiotic metabolism
TP53 pathway
Apoptosis
Inflammatory response
Epithelial mesenchymal transition
IL-6/JAK/STAT3 signalling
Cholesterol homeostasis
Adipogenesis
Estrogen response late
Reactive oxygen species
Oxidative phosphorylation
IL-2-STAT5 signalling
Glycolysis
Estrogen response early
Angiogenesis
Myogenesis
mTORC1 signaling
Coagulation
TGF-beta signaling
Fatty acid metabolism
Down-regulated by KRAS signaling
Unfolded protein response
Myc targets v2
Peroxisome
Heme metabolism
Up-regulated by KRAS signaling
Bile acid metabolism
DNA repair
Complement
Allograft Rejection
Androgen response
PI3K/AKT/mTOR signaling
Apical junction
Notch signaling
Myc targets v1
IFN-gamma response
Down-regulated in UV response
Wnt/beta-catenin signaling
Hedgehog signaling
Apical surface
Pancreas beta-cell
Spermatogenesis
Protein secretion
IFN-alpha response
E2F targets
Mitotic spindle
G2M checkpoint
FDR
Prostate
0.25 0.01
Endometrial
0.25 0.01
NES
NES
C
Buffa Score
Pre-Intervention
Post-Intervention
Timepoint

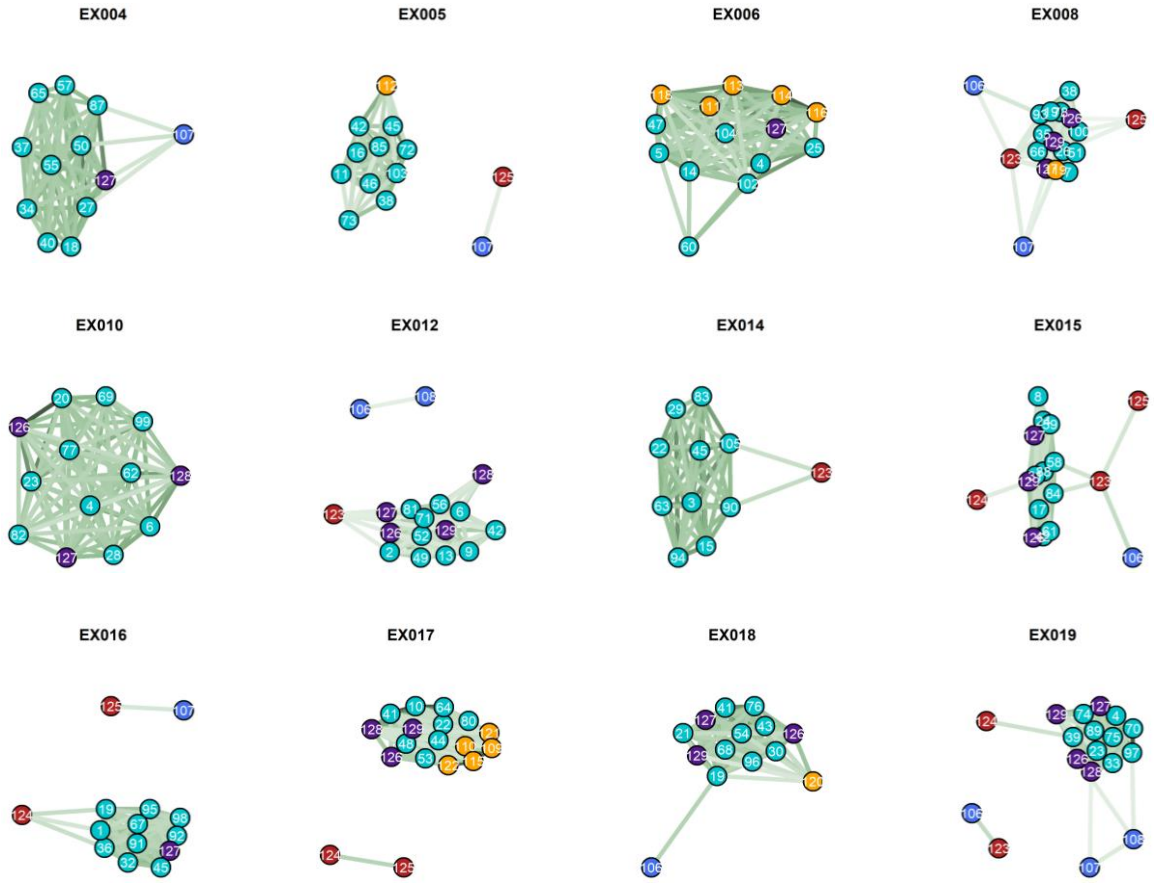
Figure 6

A

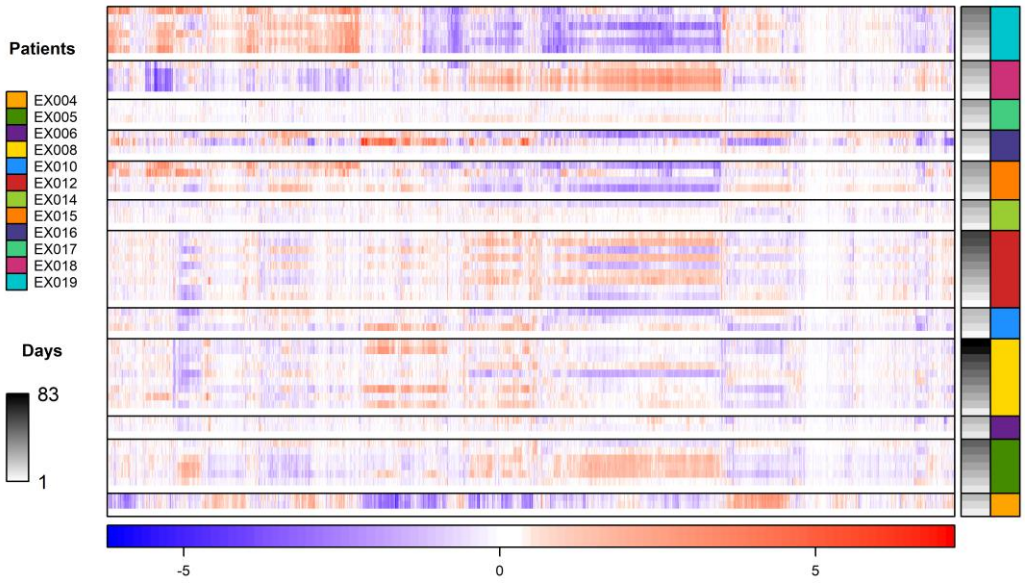


1. HMDB00017	27. HMDB00925	53. CAPS	79. MYBPC2	105. VSIG8
2. HMDB00045	28. HMDB01123	54. CBX7	80. NAAA	106. Session RDI
3. HMDB00079	29. HMDB01385	55. CCDC115	81. NETO1	107. Sedentary
4. HMDB00092	30. HMDB01422	56. CKB	82. NME1	108. Sleep
5. HMDB00156	31. HMDB01870	57. CRIP2	83. NME8	109. ASV 1057
6. HMDB00157	32. HMDB02092	58. CST7	84. PC	110. ASV 120
7. HMDB00191	33. HMDB02250	59. CYREN	85. PDGFC	111. ASV 15
8. HMDB00192	34. HMDB02259	60. DDX58	86. PENK	112. ASV 170
9. HMDB00195	35. HMDB02266	61. EGF	87. PLEKHO2	113. ASV 324
10. HMDB00211	36. HMDB02820	62. EPHA3	88. POLR1C	114. ASV 338
11. HMDB00226	37. HMDB02925	63. EXOG	89. PROC	115. ASV 354
12. HMDB00243	38. HMDB03290	64. FATE1	90. PTH	116. ASV 397
13. HMDB00244	39. HMDB06210	65. FMO3	91. PTK7	117. ASV 403
14. HMDB00254	40. HMDB06464	66. FTH1	92. RAPGEF1	118. ASV 442
15. HMDB00271	41. HMDB06471	67. GALNS	93. SERPINA12	119. ASV 451
16. HMDB00296	42. HMDB12884	68. ICAM4	94. SLC26A5	120. ASV 63
17. HMDB00491	43. HMDB13076	69. KLRK1	95. SMAD2	121. ASV 68
18. HMDB00517	44. HMDB28691	70. LECT2	96. SSR2	122. ASV 758
19. HMDB00630	45. HMDB60278	71. LGI3	97. STAP1	123. Fat
20. HMDB00715	46. ADH1C	72 LMAN2	98. SULT2B1	124. Rest CGM sleep (mean)
21. HMDB00742	47. ALDH1A3	73. LRRK2	99. SYTL4	125. Rest HR sleep (mean)
22. HMDB00759	48. ARHGEF7	74. LTA4H	100. TAPBPL	126. Adipogenesis
23. HMDB00791	49. ARL5A	75. MICOS10	101. TIMM13	127. Epithelial mesenchymal transition
24. HMDB00795	50. BMERB1	76. MIF	102. TYRO3	128. Myogenesis
25. HMDB00847	51. BPGM	77. MTSS2	103. UBE2D3	129. Oxidative phosphorylation
26. HMDB00895	52. C17orf78	78. MVD	104. UNC45A	

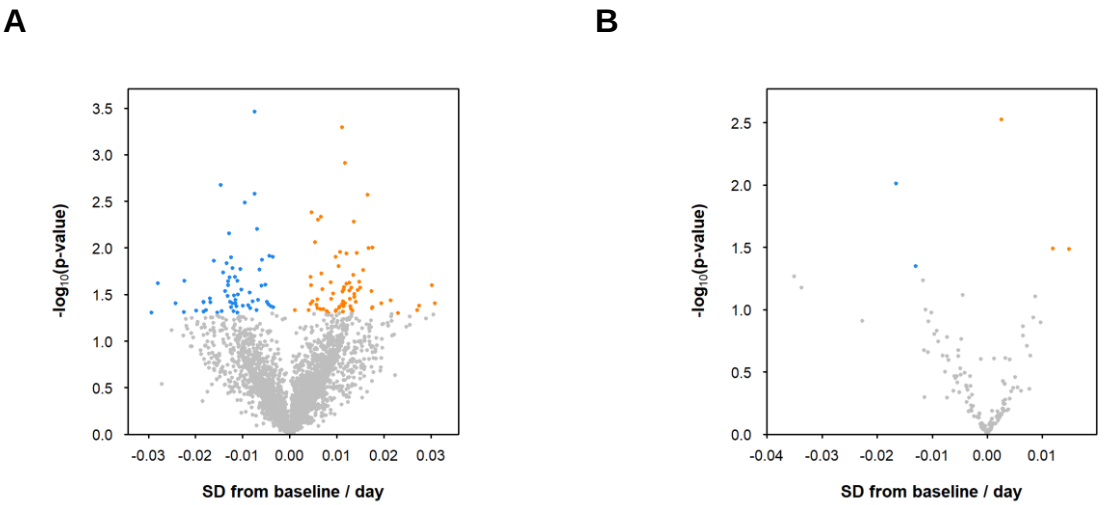
B



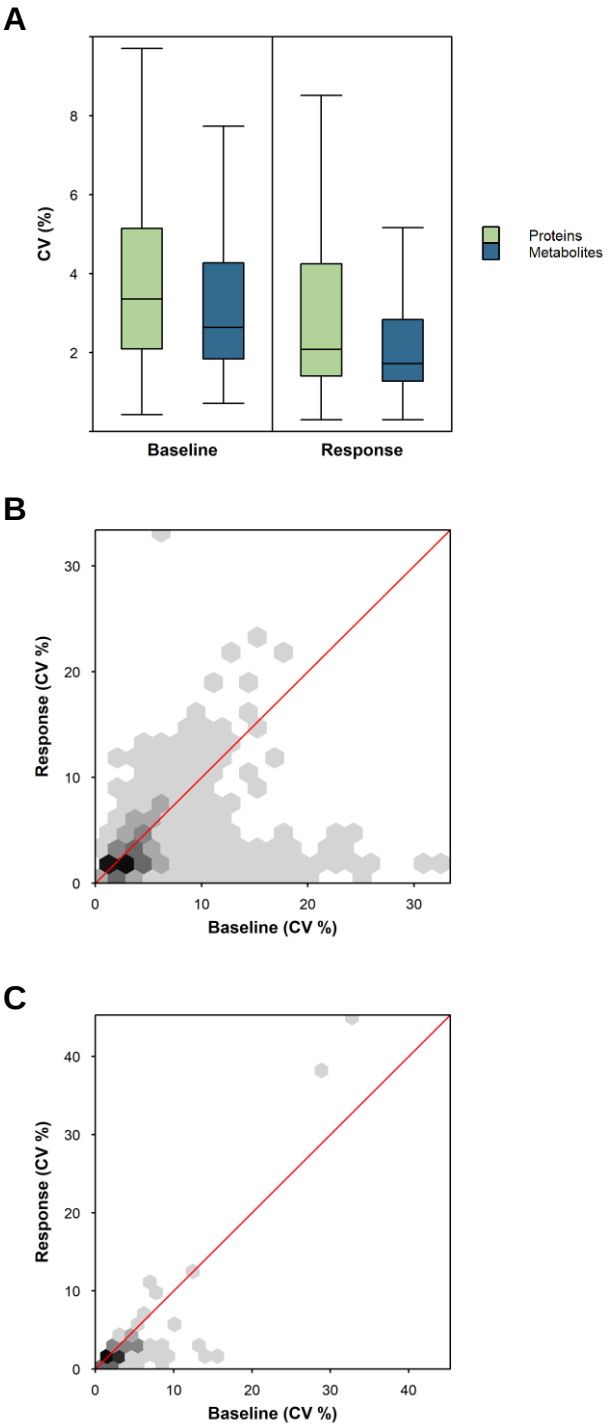
Supplemental Figure 1



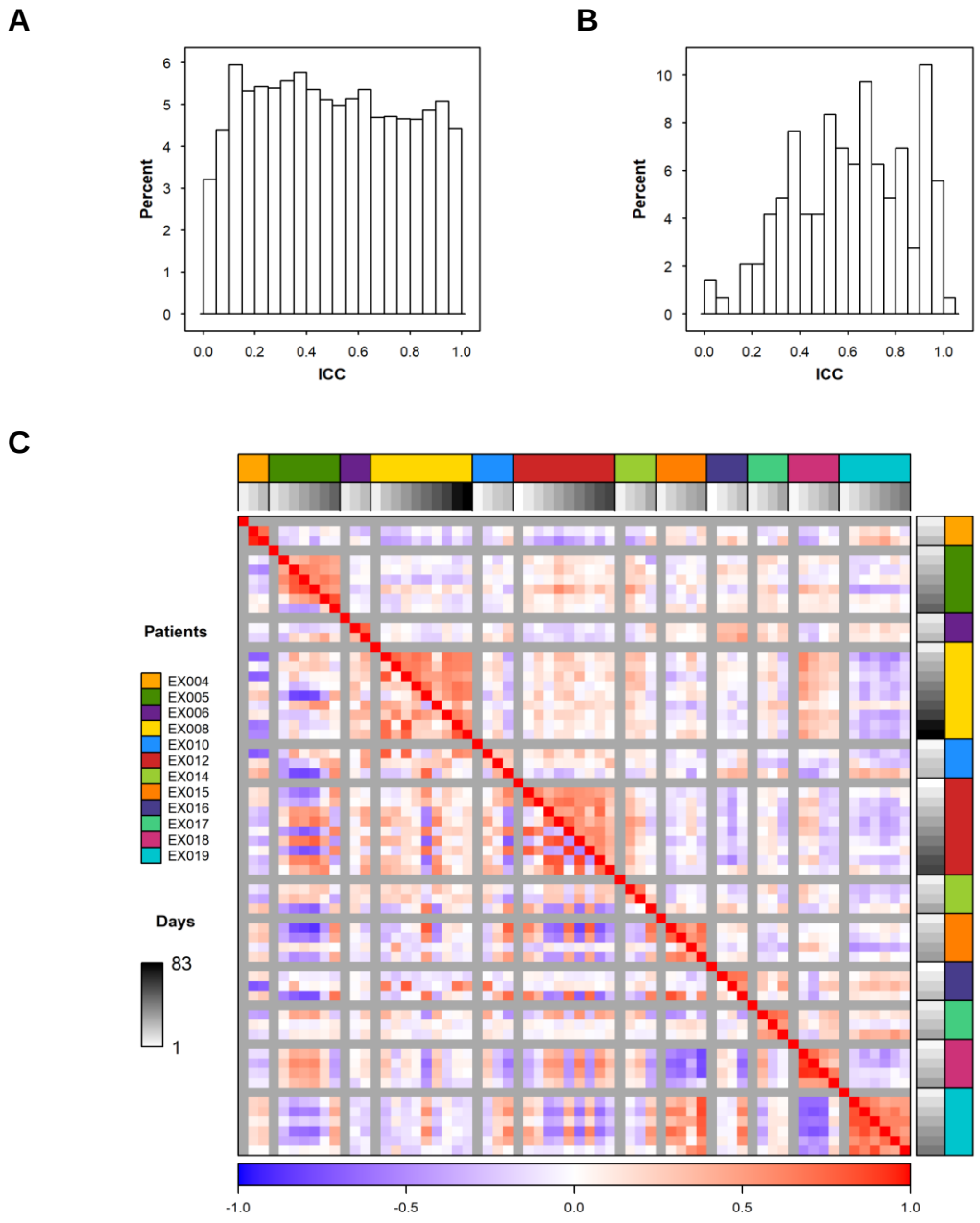
Supplemental Figure 2



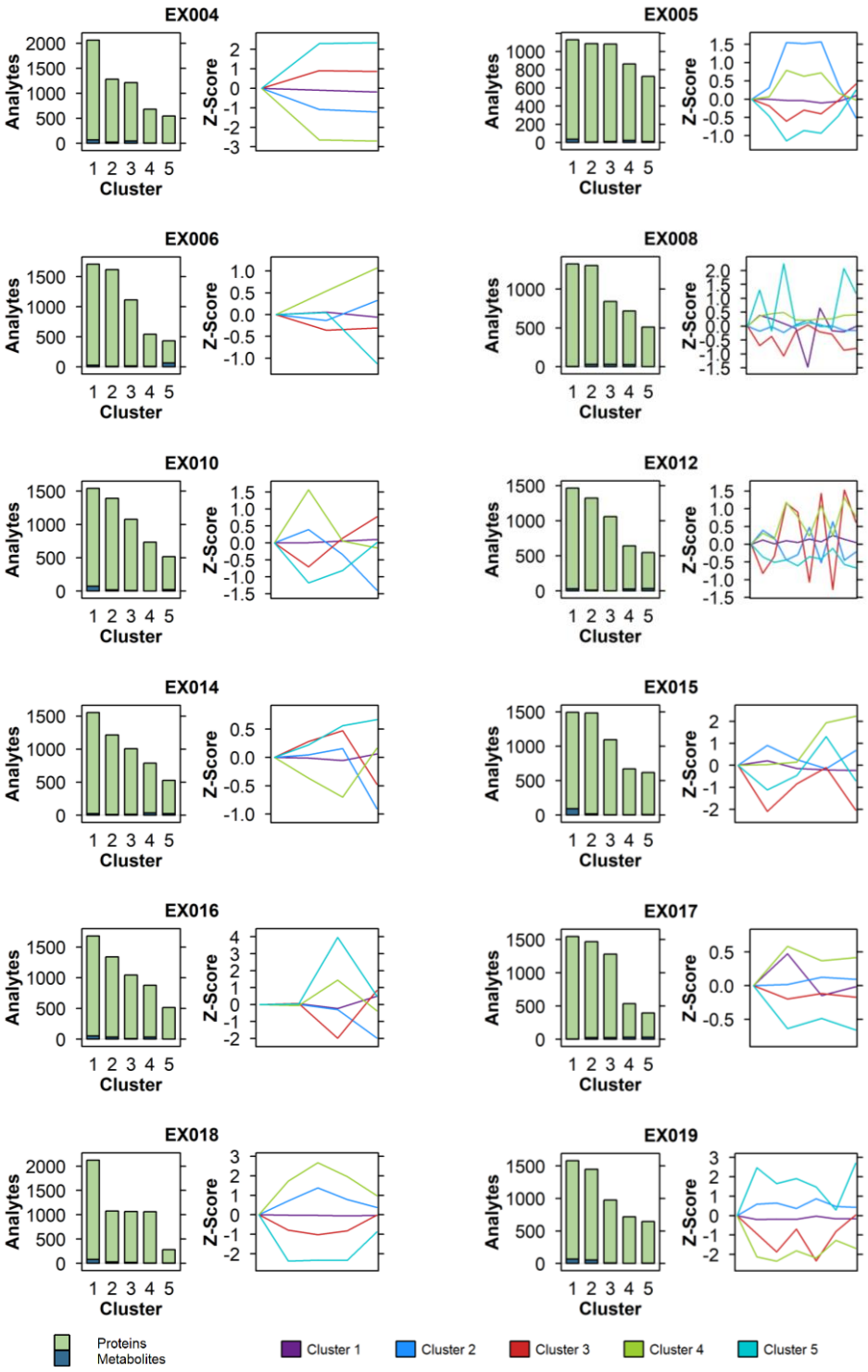
Supplemental Figure 3



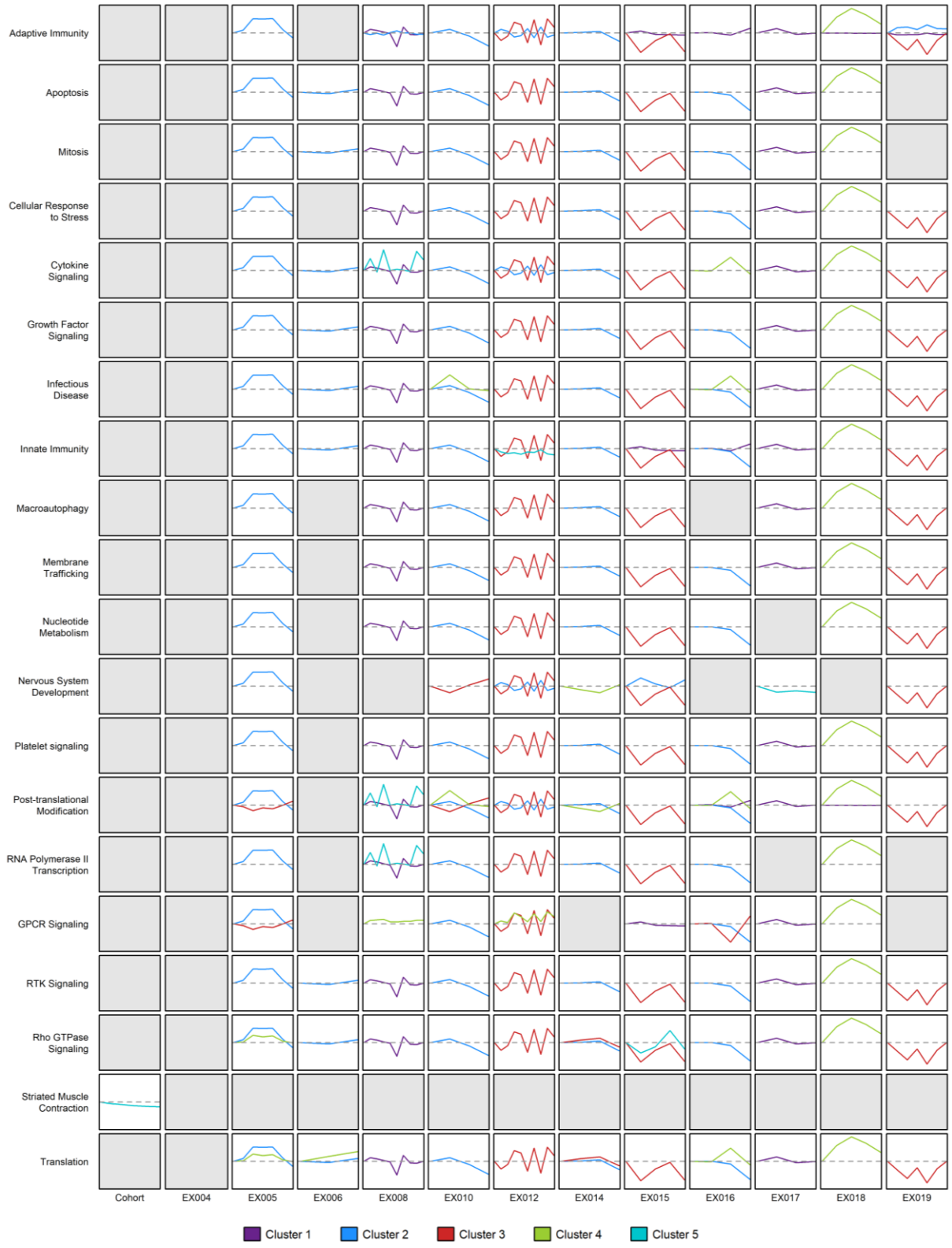
Supplemental Figure 4



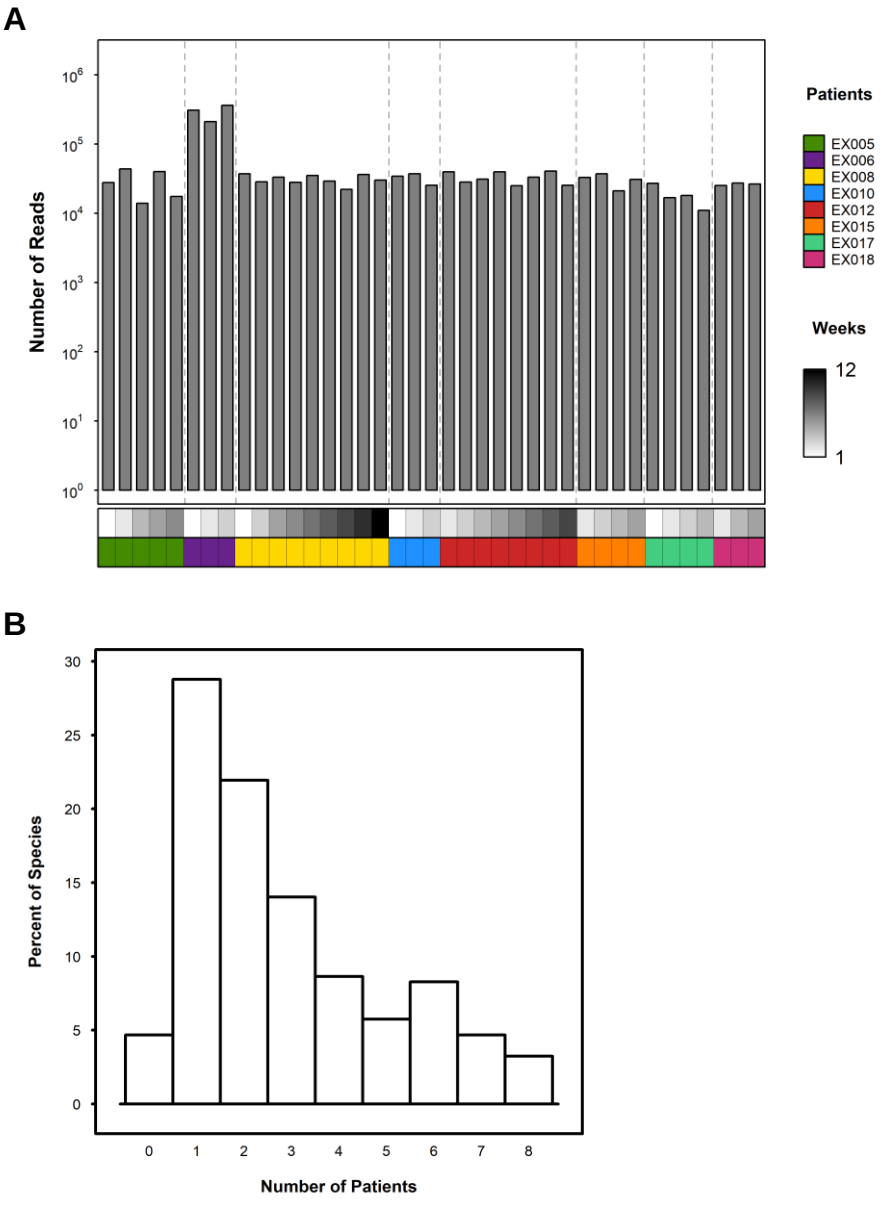
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



Supplemental Figure 8

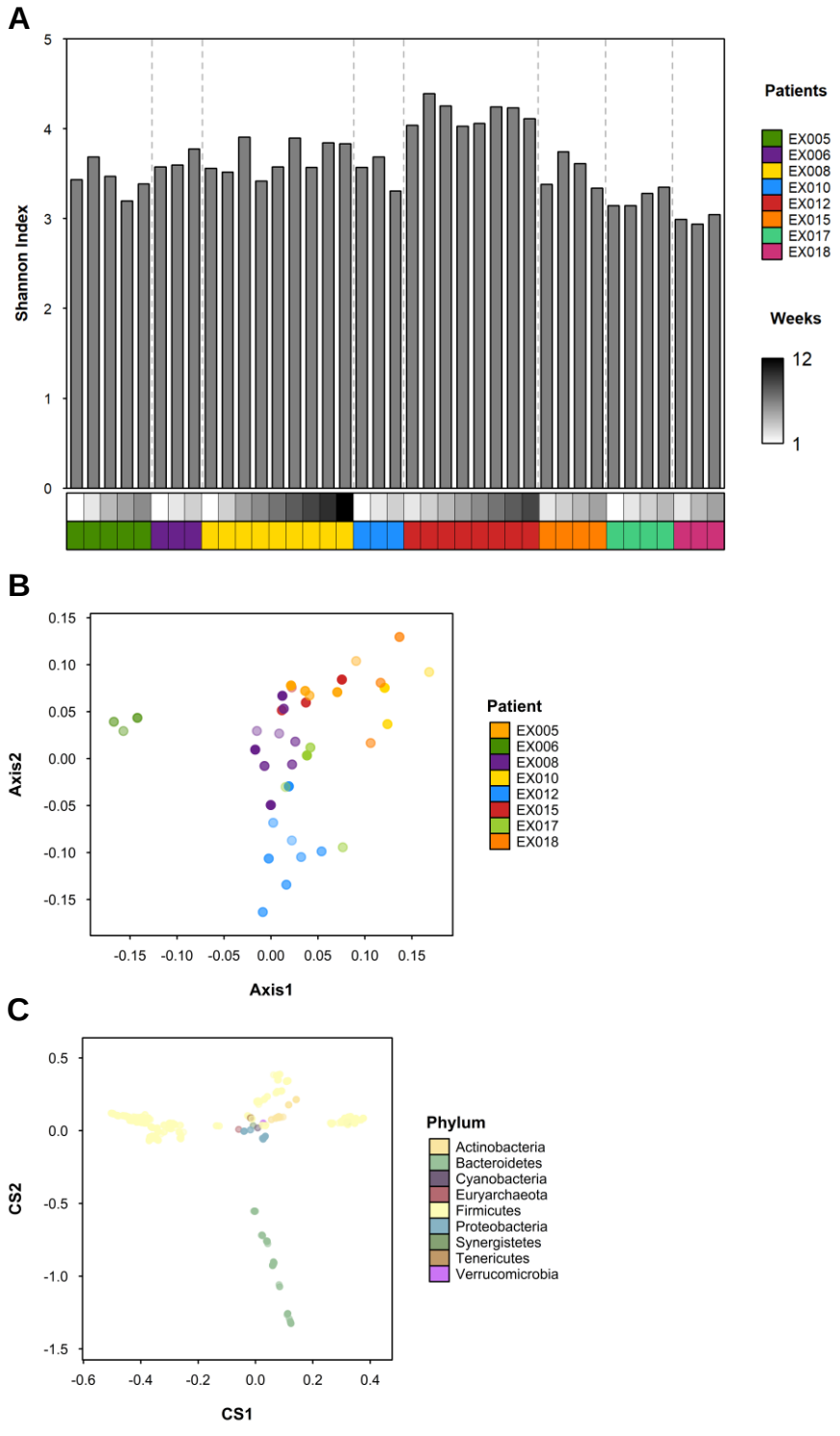


Figure Captions

Figure 1. Study overview.

(A) Schematic of study. (B) Types of data and collection timeline. (C) Continuous lifestyle and physiologic tracking on day 3 in patient 010. Shaded green bar indicates glucose challenge and shaded red bar indicates exercise therapy session. (D) Summary of states in patient 10 over the course of the study. Relative hour 0 indicates when the patient awakes for the day.

Abbreviations: OGTT – oral glucose tolerance test; Comp – composition; Ex – exercise; bpm – beats per minute

Figure 2. Longitudinal alterations in host lifestyle and physiology with exercise therapy.

(A) Lifestyle, (B) dietary, and (C) physiologic changes. Dot size indicates the magnitude of the linear model coefficient, measuring the daily effect size in standard deviations from the mean. Dot color indicates the direction of the model coefficient. Background shading indicates statistical significance of the coefficient. (D) Changes in fitness, measured by time to submaximal exercise capacity, from baseline to follow up. (E) Changes in glucose tolerance, measured by iAUC, from baseline to follow up.

Abbreviations: Carb – carbohydrate; kcal – Calories; BP – blood pressure; HR – heart rate; submax – submaximal exercise capacity; iAUC – incremental area under the curve

Figure 3. Longitudinal alterations in host circulating analytes with exercise therapy.

(A) Proteins and (B) metabolites (Fisher's $p < 0.05$, top 20 shown). Dot size indicates magnitude of the linear model coefficient, measuring the daily effect size in standard deviations relative to baseline. Dot color indicates direction of the model coefficient. Background shading indicates

statistical significance of the coefficient. (C) Number of proteins and (D) metabolites significantly altered from linear modeling ($p < 0.05$). Total number of analytes changed indicated by the numbers below. Orange indicates analytes with increasing trajectories. Blue indicates analytes with decreasing trajectories. (E) Cohort-level analyte distribution among five fuzzy c-means clusters. (F) Median analyte trajectory relative to baseline over time for each cluster.

Figure 4. Longitudinal alterations in host gut microbiome with exercise therapy.

(A) Amplicon sequence variant abundance heatmap. (B). Significantly altered Species. Dot color indicates which patient the Species was altered in. (C). Species abundance trajectories, showing standard deviations relative to baseline over time.

Figure 5. Alterations in tumor transcriptome with exercise therapy.

(A) Differentially abundant mRNA (FDR < 0.05). (B) Pathways enriched by differentially abundant mRNA, separated by cancer type. Colored shading indicates statistical significance of enrichment. (C) Changes in hypoxia, measured by Buffa score, from pre- to post-intervention.

Figure 6. Mutual information networks of all alterations with exercise therapy.

(A) Cohort-level and (B) patient-specific networks. Node color indicates the data type. Edge width indicates number of significant connections. Edge shading indicates normalized mutual information, measured as the proportion of shared information relative to total information.

Supplementary Figure Captions

Supplemental Figure 1. Host circulating analyte abundance heatmap. Protein and metabolite abundance, measured in standard deviations relative to baseline.

Supplemental Figure 2. Circulating analytes were longitudinally altered with exercise therapy. (A) 132 proteins and (B) 5 metabolites were significantly altered from the cohort-level linear modeling ($p < 0.05$).

Upward trajectories in orange and downward trajectories in blue.

Abbreviations: SD – standard deviation

Supplemental Figure 3. Exercise therapy reduced variability in both protein and metabolite levels. There was less variability in metabolite levels than protein levels. (A) Distribution of coefficient of variation of circulating analytes at baseline and in response to exercise therapy across patients. (B) Coefficient of variation of each protein and (C) metabolite at baseline and in response to exercise therapy across patients.

Abbreviations: CV – coefficient of variation

Supplemental Figure 4. Metabolites were more susceptible to inter-patient variability than proteins.

(A) Distribution of intra-class correlation of proteins and (B) metabolites across patients from cohort-level linear modeling. (C) Spearman correlation matrix of analyte levels relative to baseline between samples. Protein correlation in lower left triangle. Metabolite correlation in upper right triangle.

Abbreviations: ICC – intra-class correlation

Supplemental Figure 5. Analyte distribution and median trajectory pattern per fuzzy c-means cluster.

Supplemental Figure 6. Pathway enrichment of analyte clusters.

Supplemental Figure 7. Gut microbiome sequencing.

(A) 16S-sequencing read distribution. (B) Species distribution.

Supplemental Figure 8. Gut microbiome diversity.

(A) Alpha-diversity, measured by Shannon index. (B-C) Beta-diversity, measured by double principal coordinate analysis

Supplementary Methods

Exercise therapy clinical trial

Details for the clinical trial (NCT03813615) are discussed in the manuscript "A Digital, Decentralized Trial of Exercise Therapy in Patients with Cancer: Rationale, Methods and Feasibility Evaluation", which is under review at npj Digital Medicine. Briefly, we leverage digital methods enabling: (1) all study procedures to be digitized and conducted remotely in patients' homes, and (2) longitudinal mapping of near-continuous physiological response (e.g., heart rate, continuous glucose monitoring). Exercise therapy comprised of treadmill walking three sessions weekly, ≈30 minutes/session at 70% of measured exercise capacity for 3-11 weeks.

Patient state imputation

Sleep data was imputed using a linear mixed model predicting the sleep time and the sleep length. A variety of covariates were included such as previous night's sleep, time since last event and the sleep weekday with a random effect per patient. We used a linear mixed effects model adjusted for age across all patients for hidden state data (percentage of time in active, rest, sleep states) and physiological data (resting heart rate, blood pressure, weight and body fat). Patient effects were compared with individual linear models for each of the Z-score scaled physiological end points. The coefficients for study day were compared across patients and results were DIANA clustered by patient and end point by the p-values and effect sizes.

Plasma proteomics and metabolomics

Weekly plasma proteomic and metabolomic profiling were performed to measure and identify circulating analytes that change over the course of the exercise regimen. Proteins were quantified

by aptamer-based SomaScan. Metabolites were quantified using mass spectrometry approaches. Circulating analyte abundances were standardized relative to baseline. Longitudinal trajectories were measured using linear modeling. Trajectories were then clustered using fuzzy c-means clustering with c = 5 clusters, such that each patient had 5 clusters with distinct trajectory patterns. We use gProfileR to perform pathway analysis on the analytes within each cluster.

Gut microbiome metagenomics

Weekly stool samples underwent 16S sequencing. Alpha-diversity was calculated using Shannon index, and beta-diversity was calculated using double principal coordinate analysis. Differentially abundant Species were identified using Phyloseq.

Tumor transcriptomics

RNA sequencing was performed on pre- and post-intervention tumor specimens. Differentially abundant mRNAs were identified using DESeq. Pathway analysis was performed using gProfileR. We performed immune deconvolution analysis using Cibersortx.

Mutual information networks

Mutual information network analysis was performed to integrate diverse data types. Networks were visualized using iGraph.

References

1. Hawley, J.A., Hargreaves, M., Joyner, M.J. & Zierath, J.R. Integrative biology of exercise. Cell 159, 738-749 (2014).
2. Koelwyn, G.J., Quail, D.F., Zhang, X., White, R.M. & Jones, L.W. Exercise-dependent regulation of the tumour microenvironment. Nature reviews. Cancer 17, 620-632 (2017).
3. Warburton, D.E., Nicol, C.W. & Bredin, S.S. Health benefits of physical activity: the evidence. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne 174, 801-809 (2006).
4. Buckley, M.T., et al. Cell-type-specific aging clocks to quantify aging and rejuvenation in neurogenic regions of the brain. Nat Aging 3, 121-137 (2023).
5. Liu, L., et al. Exercise reprograms the inflammatory landscape of multiple stem cell compartments during mammalian aging. Cell Stem Cell 30, 689-705 e684 (2023).
6. De Miguel, Z., et al. Exercise plasma boosts memory and dampens brain inflammation via clusterin. Nature 600, 494-499 (2021).
7. Kurz, E., et al. Exercise-induced engagement of the IL-15/IL-15Ralpha axis promotes anti-tumor immunity in pancreatic cancer. Cancer cell 40, 720-737 e725 (2022).
8. Gomes-Santos, I.L., et al. Exercise training improves tumor control by increasing CD8+ T-cell infiltration via CXCR3 signaling and sensitizes breast cancer to immune checkpoint blockade. Cancer Immunol Res (2021).
9. Hunsberger, J.G., et al. Antidepressant actions of the exercise-regulated gene VGF. Nature medicine 13, 1476-1482 (2007).

10. Lourenco, M.V., et al. Exercise-linked FNDC5/irisin rescues synaptic plasticity and memory defects in Alzheimer's models. Nature medicine 25, 165-175 (2019).
11. Frodermann, V., et al. Exercise reduces inflammatory cell production and cardiovascular inflammation via instruction of hematopoietic progenitor cells. Nature medicine 25, 1761-1771 (2019).
12. Ashcroft, S.P., Stocks, B., Egan, B. & Zierath, J.R. Exercise induces tissue-specific adaptations to enhance cardiometabolic health. Cell Metab (2023).
13. Murphy, R.M., Watt, M.J. & Febbraio, M.A. Metabolic communication during exercise. Nat Metab 2, 805-816 (2020).
14. Neufer, P.D., et al. Understanding the Cellular and Molecular Mechanisms of Physical Activity-Induced Health Benefits. Cell Metab 22, 4-11 (2015).
15. Marron, T.U., et al. Neoadjuvant clinical trials provide a window of opportunity for cancer drug discovery. Nat Med 28, 626-629 (2022).
16. Underwood WP, L., LY, Eng S, Michalski MG, Lee CP, Moskowitz CS, Gardner G, Mueller J, Dang CT, Behfar E, Laudone VP, Eastham JA, Scott JM, Tsai B, Boutros PC, Jones LW. A digital, decentralized trial of exercise therapy in patients with cancer: Rationale, methods, and feasibility evaluation NPJ Digit Med (in press).
17. Campbell, K.L., et al. Exercise Guidelines for Cancer Survivors: Consensus Statement from International Multidisciplinary Roundtable. Medicine and science in sports and exercise 51, 2375-2390 (2019).
18. Liu, X., et al. Blautia-a new functional genus with potential probiotic properties? Gut Microbes 13, 1-21 (2021).

19. Kim, C.C., et al. Genomic insights from Monoglobus pectinilyticus: a pectin-degrading specialist bacterium in the human colon. ISME J 13, 1437-1456 (2019).

20. Sokol, H., et al. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. Proceedings of the National Academy of Sciences of the United States of America 105, 16731-16736 (2008).

21. Nie, K., et al. Roseburia intestinalis: A Beneficial Gut Organism From the Discoveries in Genus and Species. Front Cell Infect Microbiol 11, 757718 (2021).

22. Naimark, A., Jones, N.L. & Lal, S. The Effect of Hypoxia on Gas Exchange and Arterial Lactate and Pyruvate Concentration during Moderate Exercise in Man. Clin Sci 28, 1-13 (1965).

23. Swiatowy, W.J., et al. Physical Activity and DNA Methylation in Humans. Int J Mol Sci 22(2021).

24. Winegrad, S. Myosin-binding protein C (MyBP-C) in cardiac muscle and contractility. Advances in experimental medicine and biology 538, 31-40; discussion 40-31 (2003).

25. Lewis, R.V. & Stern, A.S. Biosynthesis of the enkephalins and enkephalin-containing polypeptides. Annu Rev Pharmacol Toxicol 23, 353-372 (1983).

26. Routy, B., et al. Gut microbiome influences efficacy of PD-1-based immunotherapy against epithelial tumors. Science 359, 91-97 (2018).

27. Lee, K.A., et al. Cross-cohort gut microbiome associations with immune checkpoint inhibitor response in advanced melanoma. Nature medicine 28, 535-544 (2022).

28. Gopalakrishnan, V., et al. Gut microbiome modulates response to anti-PD-1 immunotherapy in melanoma patients. Science 359, 97-103 (2018).

29. de Visser, K.E. & Joyce, J.A. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. Cancer cell 41, 374-403 (2023).

30. Greaves, M. & Maley, C.C. Clonal evolution in cancer. Nature 481, 306-313 (2012).

31. Kroemer, G., McQuade, J.L., Merad, M., Andre, F. & Zitvogel, L. Bodywide ecological interventions on cancer. Nature medicine 29, 59-74 (2023).

32. Koelwyn, G.J. & Jones, L.W. Exercise as a Candidate Antitumor Strategy: A Window into the Future. Clinical cancer research : an official journal of the American Association for Cancer Research 25, 5179-5181 (2019).

33. Ligibel, J.A., et al. Impact of a Pre-Operative Exercise Intervention on Breast Cancer Proliferation and Gene Expression: Results from the Pre-Operative Health and Body (PreHAB) Study. Clin Cancer Res 25, 5398-5406 (2019).

34. Sims, A.H., Leggate, M. & Campbell, A. Exercise Window Trial in Newly Diagnosed Breast Cancer-Letter. Clinical cancer research : an official journal of the American Association for Cancer Research 25, 7609-7610 (2019).

35. Nieto, M.A., Huang, R.Y., Jackson, R.A. & Thiery, J.P. Emt: 2016. Cell 166, 21-45 (2016).

36. Brabletz, T., Kalluri, R., Nieto, M.A. & Weinberg, R.A. EMT in cancer. Nature reviews. Cancer 18, 128-134 (2018).

37. Puisieux, A., Brabletz, T. & Caramel, J. Oncogenic roles of EMT-inducing transcription factors. Nature cell biology 16, 488-494 (2014).

38. Betof, A.S., et al. Modulation of murine breast tumor vascularity, hypoxia and chemotherapeutic response by exercise. Journal of the National Cancer Institute 107(2015).

39. Savage, H., et al. Aerobic exercise alters the melanoma microenvironment and modulates ERK5 S496 phosphorylation. Cancer Immunol Res (2023).

CHAPTER 3:

Differential Molecular Responses to Exercise Among Breast Cancer Subtypes in Mice

Abstract

Exercise is broadly considered beneficial for many diseases, including cancer. However, cancer is a heterogeneous disease with different tumor types and subtypes. To study the differential effect of exercise on breast cancer subtypes, we examined tumor xenografts from seven human breast cancer cell lines and syngeneic grafts from one mouse breast cancer cell line in mice, with the cell lines representing a range of breast cancer subtypes. Tumors derived from different cell lines displayed different growth phenotypes, with some tumors growing faster and others growing slower with exercise treatment. The tumors also had distinct genomic, transcriptomic, and proteomic changes in response to exercise. These molecular changes pointed to perturbations in common biological pathways, including DNA repair. Together, we can link the molecular alterations to the growth phenotypes across the different breast cancer subtypes to gain insight into the mechanisms by which exercise exerts its effect on cancer, particularly in different subtype contexts.

Introduction

While exercise has been linked to reduced risk and improved clinical outcomes of cancer, the effect of these health benefits varies by cancer type (1, 2). To broadly claim that exercise is beneficial for all cancers is incorrect and neglects the heterogeneity of the disease. Furthermore, specific cancer types, like breast cancer, can be categorized into subtypes that display dramatically different molecular and clinical profiles (3). Studies of exercise oncology comparing multiple breast cancer subtypes have yet to be conducted, but would provide valuable insight into how definitive molecular profiles may mediate the effect of exercise on tumors. In this study, we perform

molecular profiling on tumors derived from eight different breast cancer cell lines representing various breast cancer subtypes to investigate the relationships between exercise-associated perturbations and cancer subtypes.

Results

Seven human breast cancer cell lines (HCC1937, MCF7, MDAMB468, HCC38, JIMT1, BT474, KPL1) and one mouse breast cancer cell line (4T1) were subcutaneously injected into BALB/c mice (4, 5). Given that tumors, even within a specific cancer type, appear to respond differently to exercise, the cell lines were chosen to reflect a full range of breast cancer subtypes (**Table 1**). The mice were separated into high dose exercise (120 minutes/day), low dose exercise (30 minutes/day) and sham control groups. These doses were chosen to mimic high and low dose exercise levels in humans. At the end of 16 weeks of treatment, tumors were measured for size and subjected to genomic, transcriptomic and proteomic profiling to quantify molecular differences between tumors in exercise and control mice (**Figure 1A-B**). Tumors originating from different cell lines of origin displayed differential growth patterns in response to exercise. Tumors from HCC1937 had an accelerated growth phenotype, while tumors from MCF7, MDAMB468, and HCC38 showed no growth change. Tumors from JIMT1, BT474, KPL1, and 4T1 had a suppressed growth phenotype (**Figure 1C**).

Tumors from four of the human cell lines of origin were selected for transcriptomic profiling, representing the spectrum of growth responses: HCC1937 to represent tumors that grew with exercise, MCF7 to represent no difference between treatment groups and JIMT1 and BT474 to represent tumors that shrunk with exercise. Tumors from the 4T1 mouse cell line of origin also

underwent transcriptomic profiling. We performed tumor transcriptomic analysis to measure and identify mRNAs that responded to exercise. Among the tumors derived from the human cell lines, we identified 12 differently abundant mRNAs (Fisher's FDR < 0.1; **Figure 2A**). The mRNA abundances of these hits were variably correlated to tumor volume across the cell lines. For example, ORAI2, a gene downregulated in tumors from both BT474 and JIMT1, was also correlated with tumor volume in those tumors (**Figure 2B**). We next used the mRNA abundance to calculate Hallmark signature (6, 7) scores for each sample, then correlated those scores with the mRNA abundance (**Figure 2C**). The top two globally correlated signatures were DNA repair and interferon alpha response, showing inverse correlations with tumor volume (**Figures 2D-E**). In both the correlation of mRNA abundance and Hallmark signature score with tumor volume, few hits showed consistent directionality of correlation. The heterogeneity is consistent with the varied tumor growth patterns in response to exercise and provides further evidence that breast cancer subtypes respond differently to exercise. In the tumors derived from the 4T1, differentially abundant mRNAs (FDR < 0.1) were enriched for DNA damage response, signal transduction by p53 class mediator, amino acid biosynthesis and metabolism pathways (**Figures 3A-B**). Once again, we observed perturbation of DNA damage related pathways among tumors that were suppressed by exercise.

We performed proteomics analysis on tumors derived from all seven human and one mouse breast cancer cell lines. Differentially abundant proteins were identified in every tumor type (**Figure 4**).

We performed whole-exome sequencing on the tumors derived from 4T1 to elucidate whether the phenotypic and/or molecular differences observed in response to exercise are attributable to changes in the tumor genome.

Discussion

Exercise is linked with lower risk and better outcomes in breast cancer, but not in some other cancers, showing a cancer type-specific effect (1, 2, 8). With some of our tumors demonstrating an accelerated growth phenotype with exercise, we can determine that exercise may also exert a subtype-specific effect and is not broadly beneficial in breast cancer. There were no immediately obvious subtype similarities between the cell lines that shared growth phenotypes. However, the cell types that had a suppressed growth phenotype tended to have larger tumors overall, regardless of exercise treatment. There were very few differentially abundant mRNAs shared between each tumor; in fact, the only shared hit was the downregulation of ORAI2 in BT474 and JIMT1, both of which are cell lines with the suppressed growth phenotype. Interestingly, both of these cell lines also had a positive correlation between tumor volume and the Hallmark DNA repair pathway, suggesting that larger tumors may have activated additional DNA repair mechanisms. A similar result can be found in the tumors derived from 4T1, which showed downregulation of genes involved in DNA damage response. As above, 4T1 was another cell line that had a suppressed growth phenotype with exercise. The commonalities between BT474, JIMT1, and 4T1 point to possible shared mechanisms affected by exercise that suppress tumor growth despite them not coming from the same canonical subtype classification. Overall, our study provides evidence for investing additional attention toward the study of cancer subtypes in future exercise oncology studies.

Figure 1

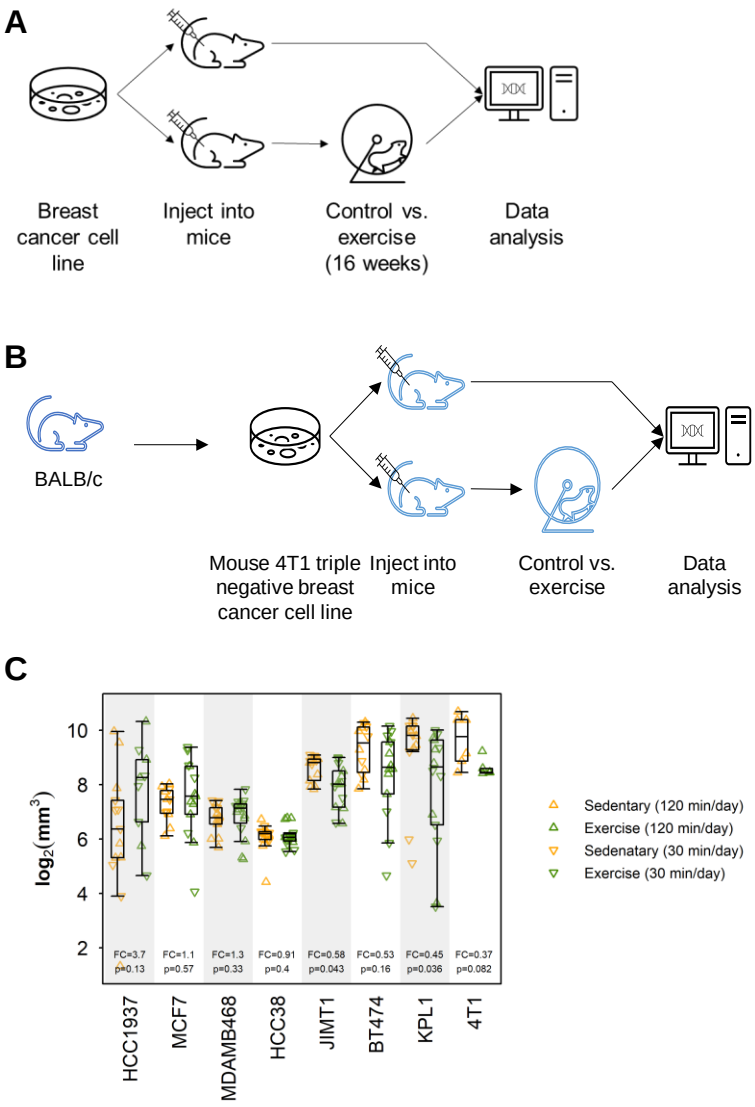

A
Breast cancer cell line
Inject into mice
Control vs. exercise (16 weeks)
Data analysis
B
BALB/c
Mouse 4T1 triple negative breast cancer cell line
Inject into mice
Control vs. exercise
Data analysis
C
log2(mm3)
10
8
6
4
2
FC=3.7 p=0.13
FC=1.1 p=0.57
FC=1.3 p=0.33
FC=0.91 p=0.4
FC=0.58 p=0.043
FC=0.53 p=0.16
FC=0.45 p=0.036
FC=0.37 p=0.082
HCC1937
MCF7
MDAMB468
HCC38
JIMT1
BT474
KPL1
4T1
Sedentary (120 min/day)
Exercise (120 min/day)
Sedenatary (30 min/day)
Exercise (30 min/day)

Figure 2

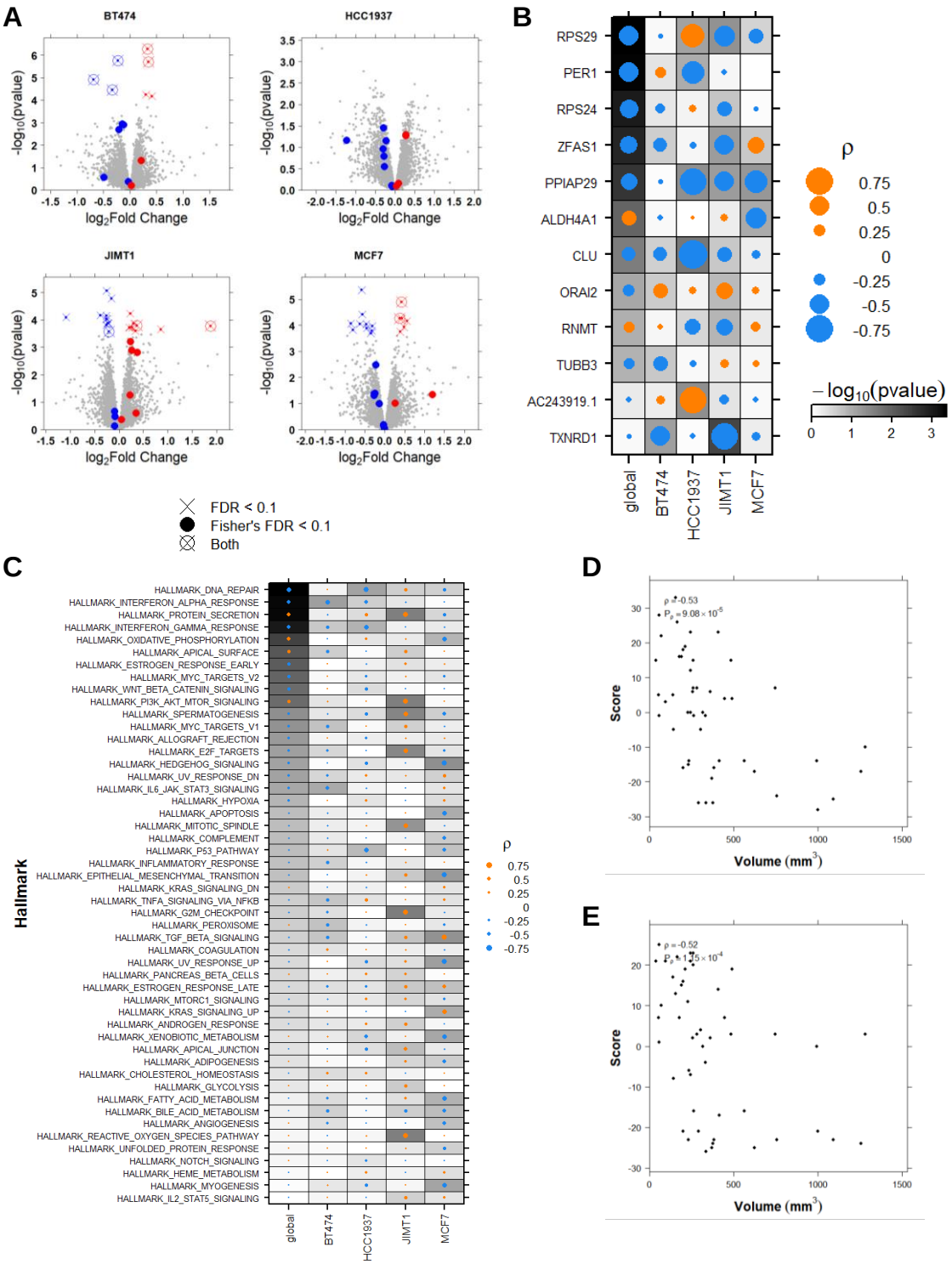

A
BT474
HCC1937
JIMT1
MCF7
-log10(pvalue)
log2Fold Change
FDR < 0.1
Fisher's FDR < 0.1
Both
B
RPS29
PER1
RPS24
ZFAS1
PPIAP29
ALDH4A1
CLU
ORAI2
RNMT
TUBB3
AC243919.1
TXNRD1
global
BT474
HCC1937
JIMT1
MCF7
ρ
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
−log10(pvalue)
C
Hallmark
HALLMARK_DNA_REPAIR
HALLMARK_INTERFERON_ALPHA_RESPONSE
HALLMARK_PROTEIN_SECRETION
HALLMARK_INTERFERON_GAMMA_RESPONSE
HALLMARK_OXIDATIVE_PHOSPHORYLATION
HALLMARK_APICAL_SURFACE
HALLMARK_ESTROGEN_RESPONSE_EARLY
HALLMARK_MYC_TARGETS_V2
HALLMARK_WNT_BETA_CATENIN_SIGNALING
HALLMARK_PI3K_AKT_MTOR_SIGNALING
HALLMARK_SPERMATOGENESIS
HALLMARK_MYC_TARGETS_V1
HALLMARK_ALLOGRAFT_REJECTION
HALLMARK_E2F_TARGETS
HALLMARK_HEDGEHOG_SIGNALING
HALLMARK_UV_RESPONSE_DN
HALLMARK_IL6_JAK_STAT3_SIGNALING
HALLMARK_HYPOXIA
HALLMARK_APOPTOSIS
HALLMARK_MITOTIC_SPINDLE
HALLMARK_COMPLEMENT
HALLMARK_P53_PATHWAY
HALLMARK_INFLAMMATORY_RESPONSE
HALLMARK_EPITHELIAL_MESENCHYMAL_TRANSITION
HALLMARK_KRAS_SIGNALING_DN
HALLMARK_TNFA_SIGNALING_VIA_NFKB
HALLMARK_G2M_CHECKPOINT
HALLMARK_PEROXISOME
HALLMARK_TGF_BETA_SIGNALING
HALLMARK_COAGULATION
HALLMARK_UV_RESPONSE_UP
HALLMARK_PANCREAS_BETA_CELLS
HALLMARK_ESTROGEN_RESPONSE_LATE
HALLMARK_MTORC1_SIGNALING
HALLMARK_KRAS_SIGNALING_UP
HALLMARK_ANDROGEN_RESPONSE
HALLMARK_XENOBIOTIC_METABOLISM
HALLMARK_APICAL_JUNCTION
HALLMARK_ADIPOGENESIS
HALLMARK_CHOLESTEROL_HOMEOSTASIS
HALLMARK_GLYCOLYSIS
HALLMARK_FATTY_ACID_METABOLISM
HALLMARK_BILE_ACID_METABOLISM
HALLMARK_ANGIOGENESIS
HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY
HALLMARK_UNFOLDED_PROTEIN_RESPONSE
HALLMARK_NOTCH_SIGNALING
HALLMARK_HEME_METABOLISM
HALLMARK_MYOGENESIS
HALLMARK_IL2_STAT5_SIGNALING
global
BT474
HCC1937
JIMT1
MCF7
ρ
0.75
0.5
0.25
0
-0.25
-0.5
-0.75
D
ρ = -0.53
Score
Volume (mm3)
E
ρ = -0.52
Score
Volume (mm3)

Figure 3

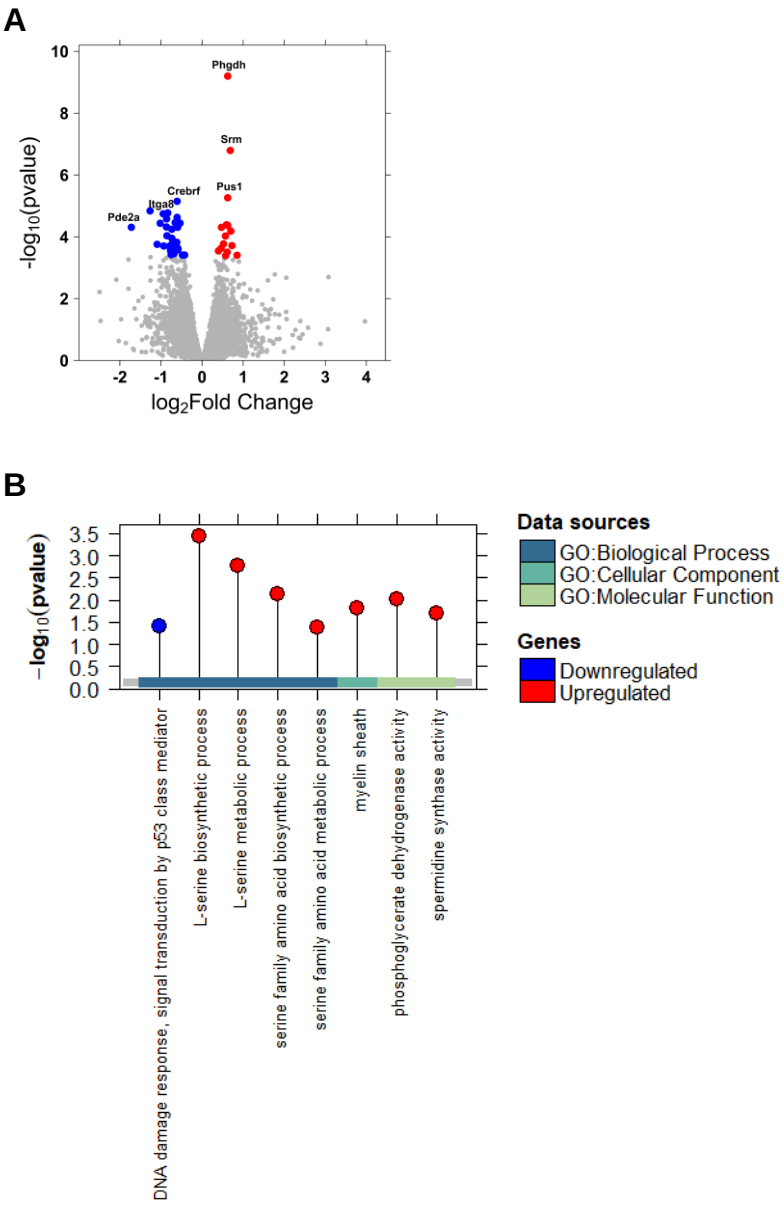

A
Phgdh
Srm
Pus1
Crebrf
Itgal
Pde2a
-log10(pvalue)
log2Fold Change
B
−log10(pvalue)
DNA damage response, signal transduction by p53 class mediator
L-serine biosynthetic process
L-serine metabolic process
serine family amino acid biosynthetic process
serine family amino acid metabolic process
myelin sheath
phosphoglycerate dehydrogenase activity
spermidine synthase activity
Data sources
GO:Biological Process
GO:Cellular Component
GO:Molecular Function
Genes
Downregulated
Upregulated

Figure 4

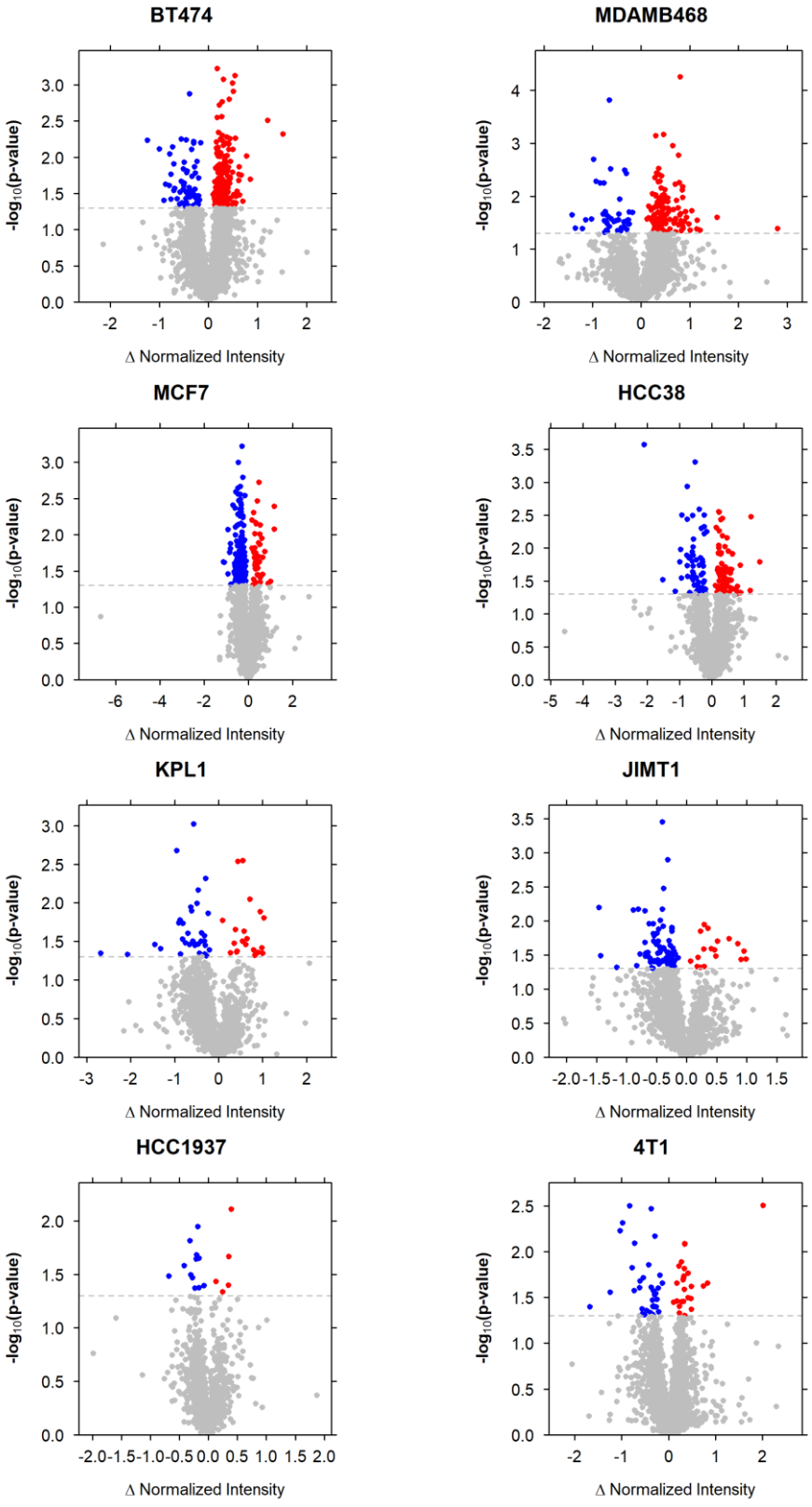

BT474
MDAMB468
MCF7
HCC38
KPL1
JIMT1
HCC1937
4T1
-log10(p-value)
Δ Normalized Intensity

Table 1

Cell Line	**Origin**	**Metastatic Potential**	**ER Status**	**PR Status**	**HER2 Status**	**Subtype**	**p53 Status**	**BRCA1 Status**
HCC1937	Primary	Low	Negative	Negative	Negative	Basal-like	Mutant	Mutant
MCF7	Primary	Low	Positive	Positive	Negative	Luminal A	Wild-type	Wild-type
MDA-MB-468	Primary	Moderate	Negative	Negative	Negative	Basal-like	Mutant	Wild-type
HCC38	Primary	Moderate	Negative	Negative	Negative	Basal-like	Mutant	Wild-type
JIMT1	Metastatic	High	Negative	Negative	Positive	HER2-enriched	Mutant	Wild-type
BT-474	Primary	Low	Positive	Positive	Positive	Luminal B	Wild-type	Wild-type
KPL-1	Primary	Low	Positive	Positive	Positive	Luminal B	Wild-type	Wild-type
4T1	Primary	High	Negative	Negative	Negative	Basal-like	Mutant	Wild-type

Figure Captions

Figure 1. Study overview.

(A) Study design for human and (B) mouse breast cancer cell lines. (C) Volume of tumors derived from each cell line.

Figure 2. Transcriptomic analysis of tumors derived from human breast cancer cell lines

(A) Differentially abundant mRNAs from tumors derived from four cell lines. (B) Spearman correlation between sample-matched differential mRNA abundance (Fisher's FDR < 0.1) and tumor volume. (C) Spearman correlation between sample-matched Hallmark signature scores and tumor volume. (D) DNA repair and (E) interferon alpha response signature scores vs. tumor volume.

Figure 3. Transcriptomic analysis of tumors derived from the 4T1 mouse cell line

(A) Differentially abundant mRNAs (FDR < 0.1). (B) Pathways enriched for differentially abundant genes.

Figure 4. Differentially abundant proteins

Table Legends

Table 1. Human breast cancer cell line characteristics

Supplementary Methods

Mouse experiment

Description of procedures

Female athymic Nude-nu mice (~6-8 weeks of age). Rooms were maintained at 21°C with 35-40% relative humidity and light (light-dark cycle 12:12h) controlled room. All animals were fed a modified 'western diet' (Research Diets, catalog number #D12079B) composed of 17% protein, 43% carbohydrates and 41% dairy-based fat. Experimental animals were purchased at ~6 weeks of age and allowed to acclimatize for 10 days prior to the commencement of study procedures. At 7 weeks, cells were subcutaneously implanted into the right flank of recipient animals. One week following implantation, animals were then stratified by tumor size (as appropriate) and randomly allocated to one of two experimental groups: (1) exercise training or (2) sham control. Treatment continued for 16 weeks, or until the mean tumor volume in the sham group of a particular cell line reached 800 millimeters cubed, whichever came first.

Exercise training

Animals were trained on a treadmill with 6 chambers, allowing 6 animals to be trained each session. The exercise groups were progressively trained to run to ~22 meters/minute at 0% grade 5 days/week for a maximum of 16 weeks. The animals were continuously monitored for the entire duration of exercise. This training intensity corresponds to approximately 70-75% of maximal exercise tolerance. Exercise training began at 10 meters/minute, 0% grade, for 10 minutes for 5 days in the week following xenograft implantation (to familiarize the animals with the treadmill). Following randomization, the training dose was systematically increased until the desired training

protocol is achieved. A puff of air was used to encourage animals to exercise. Electric shock was not used as negative reinforcement.

To ensure that the physical and social environments are similar, a second treadmill was used as a sham-exercise for the non-exercising groups, as previously described. Sham animals were placed on a stationary treadmill for the same amount of time for 5 days/week at 0 m/min at 0% grade for the length of the experiment. As such, each animal received the same handling and treadmill containment as the experimental groups with the exception of exercise training.

Minimization of pain and distress

Euthanasia was performed in compliance with American Veterinary Medical Association (AVMA) guidelines.

Transcriptomic data processing and analysis

FASTQ files were processed using UCLA-CDS pipelines to align and quantify RNA sequencing data. Pipeline-align-RNA v6.2.2 performs quality control with FastQC, trims reads with FASTP v0.21.0, aligns with STAR v2.7.6, marks duplicates with Picard tools MarkDuplicates Spark v4.1.4.1, check duplication rate with dupRadar v1.24.0 and outputs sorted BAM files. Reads were aligned to human genome GRCh38.p13. Pipeline-quantitate-RNA v2.0.0 performs isoform and gene quantitation with RSEM v1.3.3. Pipeline-quantitate-SpliceIsoform v2.0.5 quantitates the relative usage of splice isoforms with rMATS v4.1.0. Pipeline-call-RNAEditingSite v5.6.0 calls RNA editing events with REDItools2 v1.0.0. Pipeline-call-FusionTranscript calls gene fusion events with FusionCatcher v1.33.

Differential mRNA abundance analysis was performed using DESeq. We used Fisher's method to combine p-values and adjusted for multiple hypothesis testing using the false discovery rate (FDR) method. Sample-matched mRNA abundance was correlated with tumor volume using Spearman's correlation. For each gene in the 50 Hallmarks gene sets from the Molecular Signatures Database (MSigDB), we dichotomized the samples by the median mRNA abundance value. Samples with mRNA abundance greater than the median were given a +1 score, while samples with mRNA abundance less than the median were given a -1 score. A Hallmark score was calculated for each sample by taking the sum of the scores for each gene in the gene set. Sample-match Hallmark scores were correlated with tumor volume using Spearman's correlation. Differentially abundant mRNAs were separated into up- and down-regulated genes, and pathway analysis was performed using gProfileR.

Proteomic data processing and analysis

Mass spectrometry RAW files were processed using UCLA-CDS pipelines. Pipeline-search-ProteomicMSMS v1.0.0 performs library searching on tandem mass spectrometry data for proteomic studies. We used Comet to map to a combined human-mouse proteome library from UniProt. Pipeline-validate-ProteomicPSM takes the output of library search engines, performs validation by calculating statistical confidence of FDR and PEP (posterior error probability), and filters out PSMs with low confidence. Pipeline-quantitate-ProteomicTMT takes the mass spectrometry data from isobaric labeled studies (TMT or iTraq), maps it with the library search results, and quantitate peptide abundance. Pipeline-quantitate-ProteomicProtein multiple consensusXML files with quantitated PSM and performs protein inference and identification.

Peptides and proteins abundance levels were normalized across plexes using MAD normalization. Only proteins mapping exclusively to human proteins were analyzed in tumors derived from human cell lines, and only proteins mapping exclusive to mouse proteins were analyzed in tumors derived from the mouse cell line. Differential protein abundance analysis was performed using unpaired, two-sided t-test.

Statistical analyses and data visualization were performed in the R statistical environment (v4.0.2) using the BPG (v6.0.1) package.

References

1. Lavery, J. A., et al. (2024). "Association of exercise with pan-cancer incidence and overall survival." Cancer Cell 42(2): 169-171.
2. Moore, S. C., et al. (2016). "Association of Leisure-Time Physical Activity With Risk of 26 Types of Cancer in 1.44 Million Adults." Jama Internal Medicine 176(6): 816-825.
3. Koboldt, D. C., et al. (2012). "Comprehensive molecular portraits of human breast tumours." Nature 490(7418): 61-70.
4. Barretina, J., et al. (2012). "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity (vol 483, pg 603, 2012)." Nature 492(7428): 290-290.
5. Stransky, N., et al. (2015). "Pharmacogenomic agreement between two cancer cell line data sets." Nature 528(7580): 84-+.
6. Liberzon, A., et al. (2015). "The Molecular Signatures Database Hallmark Gene Set Collection." Cell Systems 1(6): 417-425.
7. Subramanian, A., et al. (2005). "Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles." Proceedings of the National Academy of Sciences of the United States of America 102(43): 15545-15550.
8. Holmes, M. D., et al. (2005). "Physical activity and survival after breast cancer diagnosis." Jama-Journal of the American Medical Association 293(20): 2479-2486.

CHAPTER 4:

Discussion

Discussion

Exercise is the strongest positive modifiable risk factor for cancer and therefore belongs at the forefront of cancer prevention and treatment (1). Exercise can be safe, tolerable, and accessible, giving it immense public health potential (2). Oncologists are already routinely recommending exercise as part of their clinical practice based on the evidence from many epidemiological and observational studies (3, 4). However, how and why exercise exerts its anti-tumor effects remain understudied. My dissertation tackles this question and revealed that the anti-tumor effects of exercise are host-, dose-, and cancer-specific. In Chapter 1, the study of a large cross-sectional cohort revealed that exercise dose plays a variable role exerting evolutionary pressures on the tumor, thus shaping the tumor genome across different cancer types. Ultimately, exercise led to improved overall survival for all cancers combined. This study design allowed us to study a large enough sample size to power the first ever exploratory analysis linking human cancer genomes with exercise. However, given the cross-sectional design of this study, we rely on the assumption that post-diagnosis exercise is representative of long-term exercise habits. In Chapter 2, the study of a prospective exercise clinical trial addresses this shortcoming by controlling the exercise regimen. In addition to pre- and post-intervention tumor transcriptomic profiling, we also collected near-continuous lifestyle, physiologic, plasma molecular, and stool microbiome data throughout the duration of the trial, providing a rich dataset to interrogate the longitudinal effects of exercise on not just the tumor, but the host as well. We observed many molecular changes within each patient, but underlying patient variability made it difficult to find many cohort-level conclusions despite a uniform exercise regimen. In Chapter 3, the study of breast cancer xenografts and syngeneic grafts provided a highly controlled experimental design in which we could control host variability, which was a point of emphasis in Xhapter 2. Furthermore, we also controlled for

exercise dose and important environmental factors like diet. Only the tumors (grafted cell lines) were variable, but even then, had well-characterized molecular features and were carefully selected to represent the range of breast cancer subtypes. In this highly controlled context, we uncovered common exercise-associated molecular signals shared across multiple subtypes. Overall, the research presented in this dissertation has elucidated novel insights into the molecular underpinnings of exercise oncology. While no single study perfectly captures the entirety of the story, each provides advantages where others fall short, coming together to paint a more comprehensive picture.

There is still much work to be done, which include the study of other large cohorts, larger and more streamlined prospective clinical trials, and further comprehensive experiments, to continue determining the mechanisms by which exercise exerts its effect on cancer. Other large cohorts that can be studied include the Prostate, Lung, Colorectal and Ovarian (PLCO) screening trial and the UK Biobank (5, 6). The PLCO cohort consists of over 150,000 healthy individuals with germline genomic microarray data, some of whom developed cancer during or after the trial. The UK Biobank (UKBB) cohort consists of 500,000 healthy individuals and cancer patients with germline whole-genome sequencing. Much like the MSK-IMPACT dataset, both of these datasets implement cross-sectional exercise surveys; the PLCO evaluated exercise in 63,000 patients and the UKBB evaluated exercise in all participants at enrollment. The UKBB also features seven days of wearable device exercise data for 100,000 participants as well. In the case of PLCO, all of these exercise assessments were performed pre-diagnosis, and in the case of UKBB, healthy individuals were assessed pre-diagnosis and cancer patients were assessed post-diagnosis. Regardless of timing of exercise evaluation, exercise surveys are subject to recall bias. Post-diagnosis survey analyses are especially subject to reverse causation when studying the effect of exercise on cancer.

Resources permitting, the ideal cohort would have at least two exercise assessments: pre- and post-diagnosis, and would be even better with wearable activity monitors instead of exercise surveys. For future prospective exercise clinical trials, improvements can be made by increasing the sample size to improve power, as well as decreasing patient variability to better identify cohort-level effects. One approach to decreasing patient variability is to limit the inclusion criteria to a single cancer type, such as prostate cancer, and doing so would also remove other covariates like sex. As such, we've conducted a Phase 1 trial of 50 prostate cancer patients and are in the process of analyzing the data. Our mouse studies can be improved by including additional breast cancer cell lines and doing full genomic, transcriptomic, and proteomic analyses on every cell line to better identify shared molecular patterns between cell lines. Furthermore, resources permitting, whole-genome sequencing is preferred over whole-exome sequencing to allow better detection of additional genomic events such as structural variation. While we were able to use TMT-Integrator's virtual reference method to compare protein levels across different plexes in the setting of a missing control reference channel, future experiments should include a control reference channel to avoid this work-around.

References

1. Patel, A. V., et al. (2019). "American College of Sports Medicine Roundtable Report on Physical Activity, Sedentary Behavior, and Cancer Prevention and Control." Medicine & Science in Sports & Exercise 51(11): 2391-2402.
2. Jones, L. W. (2015). "Precision Oncology Framework for Investigation of Exercise As Treatment for Cancer." Journal of Clinical Oncology 33(35): 4134-+.
3. Schmitz, K. H., et al. (2019). "Exercise is medicine in oncology: Engaging clinicians to help patients move through cancer." Ca-a Cancer Journal for Clinicians 69(6): 468-484.
4. Stout, N. L., et al. (2021). "A systematic review of rehabilitation and exercise recommendations in oncology guidelines." Ca-a Cancer Journal for Clinicians 71(2): 149-175.
5. Gohagan, J. K., et al. (2000). "The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: History, organization, and status." Controlled Clinical Trials 21(6): 251s-272s.
6. Sudlow, C., et al. (2015). "UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age." Plos Medicine 12(3).