nature genetics

Brief Communication

https://doi.org/10.1038/s41588-024-01969-3

Obesity-dependent selection of driver mutations in cancer

Received: 9 January 2024

Accepted: 1 October 2024

Published online: 28 October 2024

Check for updates

Cerise Tang[1], Venise Jan Castillon[1], Michele Waters[1], Chris Fong[1], Tricia Park[1], Sonia Boscenco[1], Susie Kim[1], Kelly Pekala[1], Jian Carrot-Zhang[1], A. Ari Hakimi[2], Nikolaus Schultz[1], Irina Ostrovnaya[3], Alexander Gusev[4,5,6], Justin Jee[1,7] ✉ & Ed Reznik[1] ✉

Obesity is a risk factor for cancer, but whether obesity is linked to specific genomic subtypes of cancer is unknown. We examined the relationship between obesity and tumor genotype in two clinicogenomic corpora. Obesity was associated with specific driver mutations in lung adenocarcinoma, endometrial carcinoma and cancers of unknown primaries, independent of clinical covariates, demographic factors and genetic ancestry. Obesity is therefore a driver of etiological heterogeneity in some cancers.

The physiological and environmental forces leading to cancers with specific genotypes are largely unknown. Even as oncogenic mutations emerge in healthy tissue, a confluence of physiological, immunological and epigenetic changes ultimately elicits tumorigenesis[1]. Obesity is a risk factor for the development of numerous cancers[2,3] and growing evidence suggests that medical interventions to reduce obesity can reduce cancer risk[4]. Obesity itself is associated with changes in systemic immune surveillance, metabolism and inflammation[5–7]. In composite, these changes have the potential to shape the selective pressure for specific driver mutations in cancer, connecting the evolution of tumor-intrinsic genotypes to aspects of systemic health.

Testing the hypothesis that body mass index (BMI, a quantitative surrogate of obesity status) associates with specific tumor genotypes at population scale requires jointly collected clinical and genomic data, which is largely absent in large-scale cancer genomics datasets to date. We systematically extracted information on BMI and other demographic factors in 34,274 patients profiled as part of their routine clinical care by the Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) clinical sequencing platform[8] (Extended Data Fig. 1). We assessed the statistical association between BMI and tumor genotype by modeling the incidence of mutations in 341 cancer-associated genes as a function of BMI, focusing on gene–cancer type pairs with sufficient numbers of mutant patients for adequate statistical power (Methods). Considering only oncogenic mutations as annotated by a Food and Drug Administration-recognized database[9], we identified six genes across three separate cancer types demonstrating statistically significant enrichment with BMI in specific cancer types ($q < 0.05$) (Fig. 1a and Supplementary Table 1). In lung adenocarcinoma, three genes, *KRAS* ($q = 2.6 \times 10^{-5}$, estimate = 0.03), *SETD2* (a tumor-suppressive histone methyltransferase, $q = 1.31 \times 10^{-2}$, estimate = 0.06) and *PPP2R1A* (a protein phosphatase known to regulate a variety of pathways, including phosphoinositide 3-kinase signaling, $q = 1.31 \times 10^{-2}$, estimate = 0.16), were mutated at a higher frequency in patients with obesity and one (*EGFR* $q = 3.0 \times 10^{-10}$, estimate = −0.05) was mutated at a lower frequency in the same patient group, although the number of patients with *SETD2* and *PPP2R1A* mutations was small (Table 1). In addition, we found *BAP1* in cancers of unknown primaries to be positively associated with BMI ($q = 3.5 \times 10^{-2}$, estimate = 0.15) and *POLE* in uterine endometrioid carcinoma to be negatively associated with BMI ($q = 1.6 \times 10^{-2}$, estimate = −0.08). *POLE* was similarly associated with a low BMI in a cohort restricted to only patients with endometrioid carcinoma ($P = 6 \times 10^{-3}$, estimate = −0.09). In contrast, we found no statistically significant associations between the presence of silent mutations (under putatively neutral selection) and BMI at the gene–cancer type level ($q < 0.05$) (Extended Data Fig. 2). We observed no statistically significant associations between copy number alterations and BMI.

Driver mutations are known to accumulate at different rates in patients of distinct genetic ancestries, raising the possibility that an

[1]Computational Oncology Service, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [3]Biostatistics Service, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [4]Division of Population Sciences, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA, USA. [5]Division of Genetics, Brigham & Women's Hospital, Boston, MA, USA. [6]The Broad Institute, Cambridge, MA, USA. [7]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. ✉e-mail: jeej@mskcc.org; reznike@mskcc.org

https://doi.org/10.1038/s41588-024-01969-3

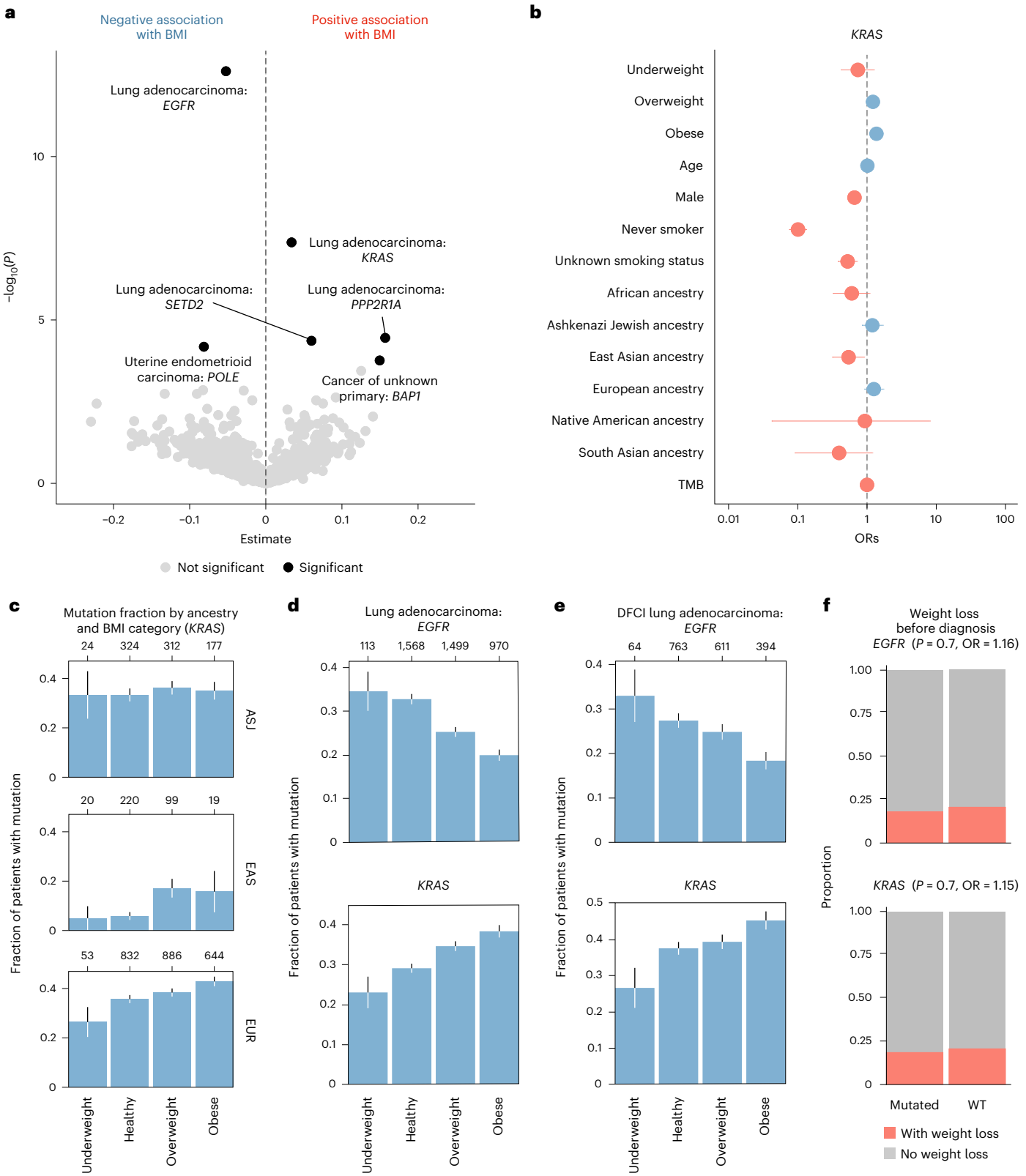


Fig. 1 | Oncogenic mutations are associated with BMI. a, Statistical association between continuous BMI and genotype across gene–cancer type pairs. The $-\log_{10}(P$ values) and estimated sizes from univariate logistic regression are on the *y* and *x* axes, respectively. Statistically significant pairs are in black. **b**, Multivariate regression demonstrating that BMI categories (underweight, BMI < 18.5 kg m^{-2}; healthy, 18.5 ≤ BMI < 25 kg m^{-2}; overweight, 25 ≤ BMI < 30 kg m^{-2}; obese, BMI ≥ 30 kg m^{-2}) are associated with *KRAS* mutations independent of other clinical factors. Results for multivariate regression with BMI as a continuous variable are shown in Supplementary Table 3. Error bars represent the 95% confidence interval (CI). **c**, *KRAS* mutation frequency in patients with lung adenocarcinoma categorized by BMI and genetic ancestry. ASJ, Ashkenazi Jewish; EAS, East Asian; EUR, European. Error bars represent the s.e. **d**, *EGFR* (top) and *KRAS* (bottom) mutation frequency in BMI categories in the MSKCC cohort. Error bars represent the s.e. **e**, *EGFR* (top) and *KRAS* (bottom) mutation frequency in BMI categories in the DFCI cohort. Error bars represent the s.e. **f**, *EGFR* (top) and *KRAS* (bottom) are not associated with weight loss before cancer diagnosis in lung adenocarcinoma using the χ^2 test.

https://doi.org/10.1038/s41588-024-01969-3

Table 1 | Patient distribution by BMI and genotype with relative risk

	EGFR LUAD	*KRAS* LUAD	*SETD2* LUAD	*PPP2R1A* LUAD	All LUAD	*BAP1* CUP	All CUP	*POLE* ENDO	All ENDO	All patients
Underweight	39 (1.08)	26 (0.73)	3 (1.02)	0 (NA)	113	0 (NA)	14	2 (2.84)	8	1,259
Healthy	513	456	41	1	1,568	2	168	17	162	11,957
Overweight	378 (0.69)	519 (1.29)	34 (0.86)	5 (5.24)	1,499	4 (2.46)	139	23 (1.08)	205	11,564
Obese	193 (0.51)	372 (1.52)	54 (2.20)	5 (8.12)	970	6 (5.19)	102	20 (0.41)	433	9,494

Distribution of patients by BMI and genotype, with relative risk in brackets. All CUP, number of patients with cancer of unknown primary tumors; All ENDO, number of patients with uterine endometrial carcinoma tumors; All LUAD, number of patients with lung adenocarcinoma tumors; All patients, total number of patients in each BMI category; *BAP1* CUP, number of patients with cancers that are unknown primaries and *BAP1*-mutant tumors; *EGFR* LUAD, number of patients with lung adenocarcinoma and *EGFR*-mutant tumors; *KRAS* LUAD, number of patients with lung adenocarcinoma and *KRAS*-mutant tumors; *POLE* ENDO, number of patients with uterine endometrial carcinoma and *POLE*-mutant tumors; *PPP2RA* LUAD, number of patients with lung adenocarcinoma and *PPP2R1A*-mutant tumors; *SETD2* LUAD, number of patients with lung adenocarcinoma and *SEDT2*-mutant tumors. Underweight (with BMI in $kg\,m^{-2}$), BMI < 18.5; healthy, 18.5 ≤ BMI < 25; overweight, 25 ≤ BMI < 30; obese, BMI ≥ 30. NA, not applicable.

association between tumor genotypes and obesity status could emerge indirectly[10]. To control for the contributions of genetic ancestry, we obtained ancestry information for each patient from a previously reported computational analysis of ancestry from germline sequencing data[11]. After correcting for age, sex, genetic ancestry and mutational burden, all six significant univariate associations between BMI and genotype remained statistically significant in the same direction, although the effect in some individual ancestries (for example, patients of Ashkenazic ancestry harboring *KRAS* mutations) was attenuated (Fig. 1c and Supplementary Table 1). Thus, elevated BMI is associated with specific cancer genotypes independent of genetic ancestry, sex, age and tumor mutational burden (TMB).

We further considered the possibility that other demographic factors, etiologies or exposures might confound an association between BMI and tumor genotypes. To directly examine this possibility, we focused on lung adenocarcinoma, where there is an established association between smoking status and elevated frequency of *KRAS* mutations/reduced frequency of *EGFR* mutations[12]. We modeled the presence of driver mutations in *KRAS* and *EGFR* as a function of BMI, clinical smoking history, TMB, ancestry, age and sex, for 4,150 patients with lung adenocarcinoma with adequate genomic and clinical data. The association between BMI and *EGFR* ($P = 7.4 \times 10^{-5}$, estimate = −0.04) and BMI and *KRAS* ($P = 1 \times 10^{-3}$, estimate = 0.03) remained statistically significant after controlling for all covariates (Fig. 1b–d and Supplementary Table 3). We also controlled for smoking quantitatively using pack-years and genomically using a previously published, computationally inferred, smoking signature[13] and associations between BMI and *EGFR* and *KRAS* remained statistically significant (pack-years: *EGFR* = 0.008, *KRAS* = 0.02; smoking signature: *EGFR* = 7.4×10^{-5}, *KRAS* = 0.001). As smoking is associated with C/G > A/T substitutions common among *KRAS* drivers, we repeated the above analysis in *KRAS* separately for C/G > A/T mutations and non-C/G > A/T mutations. We found a consistent positive association in *KRAS* across both subsetted groups (C > A estimate = 0.02, $P = 0.02$, non-C > A estimate = 0.03, $P = 4 \times 10^{-3}$). We also considered the possibility that socioeconomic status or exposure to PM2.5 air pollutants could potentially explain the association between genotype and obesity. All six statistically significant genotype–obesity associations remained statistically significant after controlling for socioeconomic status (via the Yost index), air pollution (https://www.breezometer.com/air-quality-map/) or both (Supplementary Table 4).

To corroborate these findings in an independent cohort, we obtained somatic mutation calls and BMI information from 2,727 patients with lung adenocarcinoma at a separate institution. Analysis of this dataset confirmed that *EGFR* mutations were negatively associated with BMI ($P = 5 \times 10^{-2}$, estimate = −0.02) and *KRAS* and *SETD2* mutations were positively associated with BMI ($P = 9 \times 10^{-4}$, estimate = 0.03; $P = 0.01$, estimate = 0.04, respectively) (Fig. 1e and Supplementary Table 5), after controlling for age, sex and ancestry. *PPR21A* mutation was not significantly associated with obesity although the number of patients with this mutation was small ($P = 0.84$, estimate = 0.01, $n = 22$). Furthermore, after controlling for pack-years and clinical smoking status in this validation cohort, our results remained statistically significant (Supplementary Table 5). Thus, in lung adenocarcinoma, obesity predisposes patients to a higher frequency of *KRAS* and *SETD2* mutations and a lower frequency of *EGFR* mutations.

Finally, we considered the possibility that associations between BMI measurements and genomic alterations were indirectly induced by cancer-associated weight changes at cancer diagnosis (for example, cachexia). To evaluate this possibility in lung cancer, we reviewed initial diagnosis medical notes from a subset of patients with treatment-naive lung adenocarcinoma treated at Memorial Sloan Kettering (MSK) and classified them according to whether they exhibited signs of cachexia, anorexia or other unexplained weight loss in the 6 months before diagnosis (Methods). There was no association between clinically notable weight loss before diagnosis and the presence of *EGFR* ($P = 0.70$, odds ratio (OR) = 1.16) or *KRAS* ($P = 0.69$, OR = 1.15) mutations (Fig. 1f), indicating that the association between BMI and these mutations is not confounded by pre-diagnosis weight loss.

Previous work has demonstrated the modulatory effects of obesity on the emergence of cancer, which is complex and results in higher risk of certain cancers, such as liver and colorectal, and lower risk of others, such as lung. Our results suggest that, after controlling for numerous host and environmental factors, obesity also predisposes patients to cancers with particular somatic genotypes.

Our data suggest that *EGFR*-mutant lung adenocarcinoma has a lower frequency in patients with elevated BMI and, conversely, that *KRAS*-mutant lung adenocarcinoma is more frequent in patients with obesity. These findings have treatment implications: both genes have molecularly targeted therapies[14] and higher proportions of immune-sensitive *KRAS* drivers may explain why obese patients respond better to immunotherapy[15]. More fundamentally, few modifiable risk factors of *EGFR/KRAS*-mutant lung adenocarcinoma have been characterized. A recent study[6] suggests that air pollution may trigger microenvironmental changes, allowing for proliferation of cells with activated *EGFR*. This is one of the first studies to suggest that obesity is a similar modifier of *EGFR*-mutant lung adenocarcinoma risk. There are many plausible mechanisms by which obesity might affect selection for certain molecular drivers, including modulation of lipid metabolism in cancer cells via *PIK3CA*/mammalian target of rapamycin pathways[7]. Obesity itself may also inhibit the function of $CD8^+$ T cells[5], which can affect the selective pressure for immunogenic driver alterations. As *KRAS*-mutant lung adenocarcinoma is generally more sensitive to immune checkpoint blockade than *EGFR*-mutant lung adenocarcinoma[16], it seems plausible that changes to the immune microenvironment downstream of obesity may disproportionately allow for proliferation of *KRAS*-mutant lung adenocarcinoma in patients with obesity. Although real-world histopathological imaging

data increasingly excel at identifying tumor-infiltrating lymphocytes and other markers of immune system activity[17], accompanying BMI data are not routinely published. Such data are necessary to confirm any causal link between obesity and immunological mechanisms of selection for particular drivers.

Our study has several limitations. We were underpowered to detect associations between BMI and genomic alterations in rare cancer types and infrequently mutated genes. There may be other genomic changes, including structural rearrangements, associated with BMI, which we did not consider in this analysis. BMI itself is only one of many metrics for studying obesity and we did not have access to other metrics, such as skinfold thickness, in the clinical record[3]. Genomic studies do not fully capture the molecular landscape of a tumor, and broader studies that combine genomic, transcriptomic and metabolomic analyses would provide a more well-rounded view of the effect of obesity on cancer tumor phenotypes[18]. Finally, and most importantly, prospective studies will have to be done to determine whether modifications to BMI, through diet, exercise or medical interventions, can causally modify the incidence of particular cancer genotypes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-01969-3.

References

1. Yokoyama, A. et al. Age-related remodelling of oesophageal epithelia by mutated cancer drivers. *Nature* **565**, 312–317 (2019).
2. Lauby-Secretan, B. et al. Body fatness and cancer—viewpoint of the IARC Working Group. *N. Engl. J. Med.* **375**, 794–798 (2016).
3. Rask-Andersen, M. et al. Adiposity and sex-specific cancer risk. *Cancer Cell* **41**, 1186–1197.e4 (2023).
4. Aminian, A. et al. Association of bariatric surgery with cancer risk and mortality in adults with obesity. *JAMA* **327**, 2423–2433 (2022).
5. Ringel, A. E. et al. Obesity shapes metabolism in the tumor microenvironment to suppress anti-tumor immunity. *Cell* **183**, 1848–1866.e26 (2020).
6. Hill, W. et al. Lung adenocarcinoma promotion by air pollutants. *Nature* **616**, 159–167 (2023).
7. Koundouros, N. & Poulogiannis, G. Reprogramming of fatty acid metabolism in cancer. *Br. J. Cancer* **122**, 4–22 (2020).
8. Cheng, D. T. 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).
9. Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol*. https://doi.org/10.1200/PO.17.00011 (2017).
10. Jing, L., Su, L. & Ring, B. Z. Ethnic background and genetic variation in the evaluation of cancer risk: a systematic review. *PLoS ONE* **9**, e97522 (2014).
11. Arora, K. et al. Genetic ancestry correlates with somatic differences in a real-world clinical cancer sequencing cohort. *Cancer Discov.* **12**, 2552–2565 (2022).
12. Riely, G. J. et al. Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma. *Clin. Cancer Res.* **14**, 5731–5734 (2008).
13. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
14. Suehnholz, S. P. et al. Quantifying the expanding landscape of clinical actionability for patients with cancer. *Cancer Discov.* **14**, 49–65 (2024).
15. Kichenadasse, G. et al. Association between body mass index and overall survival with immune checkpoint inhibitor therapy for advanced non-small cell lung cancer. *JAMA Oncol.* **6**, 512–518 (2020).
16. Lisberg, A. et al. A phase II study of pembrolizumab in *EGFR*-mutant, PD-L1+, tyrosine kinase inhibitor naïve patients with advanced NSCLC. *J. Thorac. Oncol.* **13**, 1138–1145 (2018).
17. Rakaee, M. et al. Association of machine learning-based assessment of tumor-infiltrating lymphocytes on standard histologic images with outcomes of immunotherapy in patients with NSCLC. *JAMA Oncol.* **9**, 51–60 (2023).
18. Nguyen, H.-L. et al. Obesity-associated changes in molecular biology of primary breast cancer. *Nat. Commun.* **14**, 4418 (2023).

Methods

Patient cohort

The present study utilized data from MSK-IMPACT (no. NCT01775072), a prospective observational cohort study of tumor evolution. The present study received full ethical approval by the institutional review board at MSK Cancer Center (MSKCC). For the present study, patients provided written informed consent for the use of their genomic data for research. Participants in the present study were not compensated for their participation. Tumors were sequenced using the MSK-IMPACT assay through to 23 March 2023. In patients with multiple samples, only one sample (the earliest sampled primary tumor) was included in the final cohort. Clinical characteristics were annotated per the standard MSK-IMPACT workflow. The total cohort constitutes 34,274 samples spanning 102 cancer types. Only cancer–gene pairs with >50 samples in the cancer type and >10 mutations of the given gene in the cancer type were tested for genotype–BMI associations. For each patient, we identified the BMI measurement collected at the date nearest to the date of sample acquisition leading to genomic sequencing, that is, the earliest possible date for which a given genomic alteration could be confirmed in a patient's tumor. BMI distribution of cancer types is shown in Extended Data Fig. 1. Patients without a BMI measurement within 30 d of tumor acquisition were excluded. Patients at the Dana-Farber Cancer Institute (DFCI) had tumor genomic profiling with OncoPanel[19], a targeted sequencing platform analogous to MSK-IMPACT.

Targeted DNA sequencing with MSK-IMPACT

DNA sequencing was performed using the MSK-IMPACT sequencing panel, which is a hybridization capture-based, next-generation sequencing assay, in a Clinical Laboratory Improvement Amendments-certified molecular laboratory. Genomic DNA from formalin-fixed, paraffin-embedded primary or metastatic tumors and matched normal samples was extracted and sheared. Customized probes were then synthesized for targeted sequencing of all exons and select introns of 341, 410, 468 or 505 genes. Illumina HiSeq 2500 was used to capture pooled libraries containing captured DNA fragments to high, uniform coverage (>500 median coverage). All classes of genomic alterations including substitutions, insertions/deletions (indels), copy number alterations and rearrangements were determined and called against the patient's matched normal sample. The computational pipelines used for variant calling are based on standard best practices and used a combination of open-source and custom-written scripts and programs.

The OncoKB precision oncology knowledge base was used to annotate genomic alterations. OncoKB identifies functionally relevant cancer variants and their potential clinical actionability. Only alterations classified as oncogenic by OncoKB were used in this analysis. Reported alteration frequencies were adjusted to account for the specific set of genes included in each version of the MSK-IMPACT panel used by dividing the number of gene-specific alterations by the number of samples for which a given gene was sequenced.

Logistic regression model

Logistic regression was performed where BMI (as a continuous variable) was used to predict somatic mutation status ('Gene') in a particular cancer type. Only detailed cancer types with >50 patients and the 341 genes in the initial MSK-IMPACT panel were included in the present study. Covariates for the logistic regression included age, sex, ancestry and TMB. Smoking status was included in models with lung adenocarcinoma. The glm function from the stats R package was used to conduct the modeling. *P* values were Benjamini–Hochberg corrected.

The following equations were used:

(1) Gene ~ BMI
(2) Gene ~ BMI + age + sex + ancestry + TMB
(3) Gene ~ BMI + age + sex + ancestry + smoking status + TMB.

Pre-diagnosis weight loss identification and analysis

To identify pre-diagnosis weight loss in patients, 800 initial consultation notes were randomly selected from patients with treatment-naive lung adenocarcinoma and manually categorized. Records with mentions of the patient reporting weight loss, anorexia or decreased appetite were labeled as having pre-diagnosis weight loss. Records with no mention of weight loss, anorexia or decreased appetite were labeled as having no pre-diagnosis weight loss.

Statistics and reproducibility

No statistical method was used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment because no subjective judgements were used.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All clinical and genomic sequencing data described in this paper have been deposited in the cBioPortal for Cancer Genomics (https://www.cbioportal.org) and are available for online browsing and download there. Data are also available at https://doi.org/10.5281/zenodo.11075026 (ref. 20). Although raw sequencing data are restricted to protect patient privacy in accordance with federal and state law, de-identified data are available. De-identified data can be requested for research use from the corresponding authors. Data will be shared for a span of 2 years within 2 weeks of execution of a data transfer agreement with MSKCC, which retains all title and rights to the data and results from their use.

Code availability

Code for reproducing the results in Fig. 1 and associated Supplementary Tables is available at https://github.com/reznik-lab/bmi_genomics. Please note that patients with ages ≥90 years have their age listed as 90 to protect protected health information. However, their actual ages were used in the analyses.

References

19. Garcia, E. P. et al. Validation of OncoPanel: a targeted next-generation sequencing assay for the selection of aomatic variants in cancer. *Arch. Pathol. Lab. Med.* **141**, 751–758 (2017).
20. Tang, C. et al. Obesity-dependent selection of driver mutations in cancer. *Zenodo* https://doi.org/10.5281/zenodo.11075026 (2024).

Acknowledgements

We thank the members of E.R., J.J. and N.S.'s laboratories for discussion and support. This work was supported by grant nos. P50 CA247749-01 and P30 CA008748, the Molecular Diagnostics Service in the Department of Pathology and the Marie-Josee and Henry R. Kravis Center for Molecular Oncology. J.J. was supported by grant no. K12 CA184746 and a Lung Cancer Research Foundation—AstraZeneca Research Grant. E.R. was supported by the National Institutes of Health (grant no. R37 CA276200) and Department of Defense (grant no. HT9425-23-1-0995), as well as the Society of Memorial Sloan Kettering Cancer Center.

Author contributions

E.R. and J.J. conceived the study. C.T., V.J.C., M.W., C.F., T.P., S.B., S.K., K.P., J.C.-Z, A.A.H., N.S., I.O. and A.G. assisted with genomic data collection and analytical methodology development. C.T., V.J.C., A.G., J.J. and E.R. designed and performed the experiments. C.T., V.J.C., J.J. and E.R. wrote the paper with input from all authors.

Competing interests
The authors declare no competing interests.

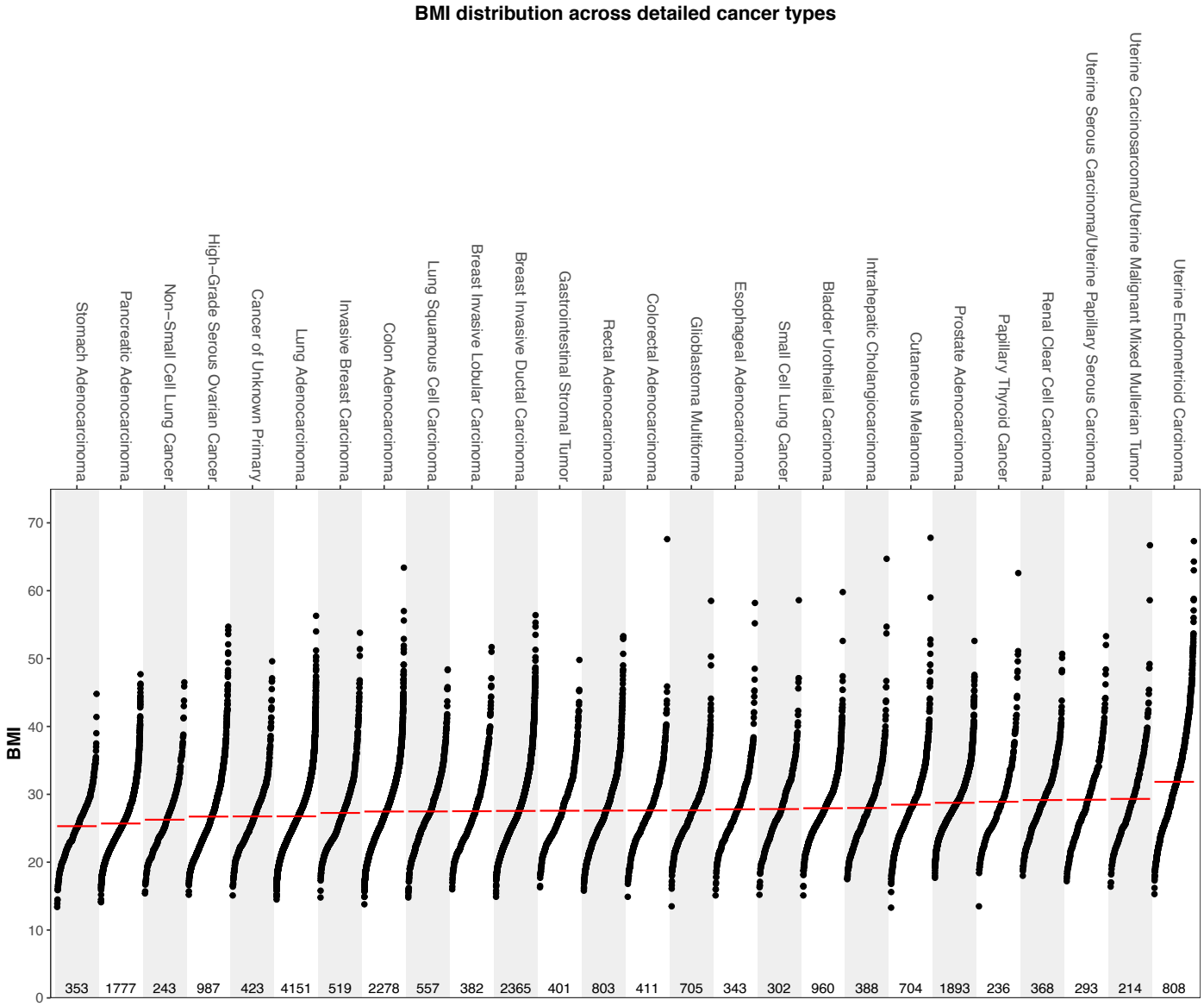
Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41588-024-01969-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-01969-3.

Correspondence and requests for materials should be addressed to Justin Jee or Ed Reznik.

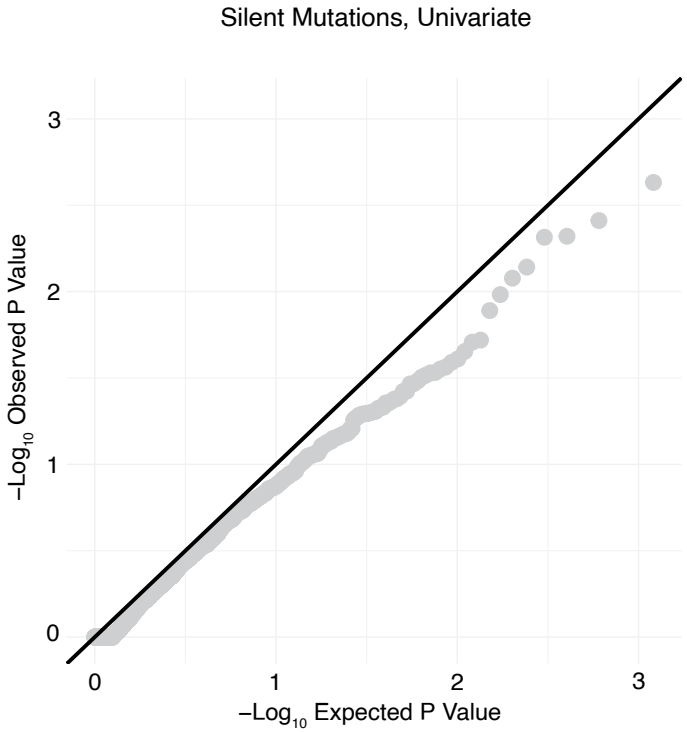
Peer review information *Nature Genetics* thanks Kristy Brown and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41588-024-01969-3



Extended Data Fig. 1 | BMI Distribution Across Cancer Types. BMI values across cancer types with over 200 samples. Each dot corresponds to a patient, with the y-axis representing the patient's BMI. Cancer types are ordered by median BMI, with the lowest median BMI on the left and highest BMI on the right.

https://doi.org/10.1038/s41588-024-01969-3



Extended Data Fig. 2 | Silent mutations are not associated with body mass index. Scatterplot comparing -$\log_{10}$ expected p-values vs. -$\log_{10}$ observed p-values in univariate logistic regression models for silent mutations.

nature portfolio

Corresponding author(s): Ed Reznik
Justin Jee

Last updated by author(s): Sep 3, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect this data.
Data analysis	All analyses were performed in R (version 4.3.2). All code for analyses is publicly available at https://github.com/reznik-lab/bmi_genomics.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All clinical and genomic sequencing data described in this manuscript has been deposited in the cBioPortal for Cancer Genomics and are available for online browsing and download there. Data is also available at https://zenodo.org/records/11075026. While raw sequencing data is restricted to protect patient privacy in accordance with federal and state law, de-identified data are available. De-identified data can be requested for research use from the corresponding authors. Data

will be shared for a span of 2 years within 2 weeks of execution of a data transfer agreement with MSKCC, which retains all title and rights to the data and results from their use.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We report sex in the publicly available data and it is used as a covariate in our multivariate analyses.
Reporting on race, ethnicity, or other socially relevant groupings	We report genetic ancestry in the publicly available data and it is used as a covariate in our multivariate analyses.
Population characteristics	Our study population is 55% female, 45% male. The median age is 62. All study participants have been diagnosed with cancer.
Recruitment	NA
Ethics oversight	MSK-IMPACT (NCT01775072), a prospective observational cohort study of tumor evolutions, was approved by the Institutional Review Board at Memorial Sloan Kettering Cancer Center.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Tumors were sequenced using the MSK-IMPACT assay through 23/03/2023. In patients with multiple samples, only one sample (the earliest sampled primary tumor) was included in the final cohort. The total cohort constitutes 34,274 samples spanning 102 cancer types. No statistical analysis was performed to pre-determine sample size.
Data exclusions	Only cancer-gene pairs with over 50 samples in the cancer type and over 10 mutations of the given gene in the cancer type were tested.
Replication	Findings were corroborated in an external cohort of 2727 lung adenocarcinoma samples sequenced at Dana Farber Cancer Institute. Statistical analyses were performed independently.
Randomization	Participants were not randomized into groups. Covariates were controlled using multivariate logistic regressions.
Blinding	Blinding was not performed as there were no subjective judgments involved.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT01775072
Study protocol	https://clinicaltrials.gov/study/NCT01775072?tab=history&a=48
Data collection	Tumors were sequenced using the MSK-IMPACT assay through 23/03/2023. Samples were collected from study locations listed in the clinical trial.
Outcomes	No outcomes were included in this study.

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA