FACTORS ASSOCIATED WITH GENOMIC ALTERATIONS IN TUMOR SAMPLES

Group 10: Amanda Carrico, Yuxiang Ren Faculty Advisor/Client: Dr. Xi Kathy Zhou (Prof. of Research in PHS)

Weill Cornell Medicine

INTRODUCTION

Genomic alteration measurements have been shown to be diagnostic and increasingly used in treatment decision making. Understanding the association between patient characteristics and genomic alteration measurements can help advance personalized treatment decision making. The relationships between these measurements, Tumor Mutation Burden and Fraction Genome Altered, and clinical factors, are explored through this study. This study aims to answer the question "Are there any associations between Sex/Race/BMI/Smoking History an TMB/FGA in cancer patients?"

METHODS

- **Data Source:**
  - 28 studies subset by Factor
  - Non-missing Smoking dataset best optimizes non-missing data for significant factors
  - Final post imputation dataset contains data from “Bladder Urothelial Carcinoma Carcinoma”, “Cervical Squamous Cell Carcinoma and Endocervical”, “Colorectal Cancer”, “Glioblastoma”, “Intrahepatic Cholangiocarcinoma”, “Kidney Renal Papillary Cell Carcinoma”, “MSK-IMPACT Clinical Sequencing Cohort”, and “Pan-Lung Cancer” studies
- **Data Combination:**
  - Smoking indicator values standardized using the National Cancer Institute (NCI) Common Data Element (CDE)
  - Race grouped into "White", "Black", "Asian", and "Other"
  - Cancer type and tumor site category combinations based on sites of cancer in body as needed

  Other combinations to aid with redundancies
- **Missing Data Imputation:**
  - Every obs. has missing values due to different variables across different studies
  - Impute categorical missing data with level "unknown"
  - Impute numeric missing data with the Predictive Mean Matching (PMM) method
- **Descriptive Statistics and Univariate Analysis:**
  - Descriptive Statistics table was developed using the table1 package
  - Individual factor analysis utilized Boxplots, T-tests, Analysis of Variance, Kruskal-Wallis, and Man-Whitney U tests via the R package BKTR
- **Mixed Effect Models:**
  - Squared root transformation of FGA & Log transformation of TM
  - Random effect: Patient ID, Cancer Type, and Study
  - Fixed effect: age, sex, smoking history, race, tumor purity
  - Model diagnosis with AIC, BIC, conditional R-Square, and residuals distribution
  - Compare between different hierarchical structures with Anova
  - Smoking history effect in different cancer type

RESULTS

$$\log_2(\text{TMB}+1) = \beta_0 + \beta_{(1\text{AGE})} + \beta_{(2\text{SEX})} + \beta_{(3\text{RACE})} + \beta_{(4\text{TUMOR\_PURITY})} + \beta_{(5\text{SMOKING\_HISTORY})} + u_{\text{PATIENT\_ID}} + u_{\text{Study}} + u_{\text{CANCER\_TYPE}} + \varepsilon$$

$$\sqrt{FGA_{i,j,k}} = \beta_0 + \beta_{(1\text{AGE})} + \beta_{(2\text{SEX})} + \beta_{(3\text{RACE})} + \beta_{(4\text{TUMOR\_PURITY})} + \beta_{(5\text{SMOKING\_HISTORY})} + u_{\text{PATIENT\_ID}} + u_{\text{Study}} + u_{\text{CANCER\_TYPE}} + \epsilon_{i,j,k}$$

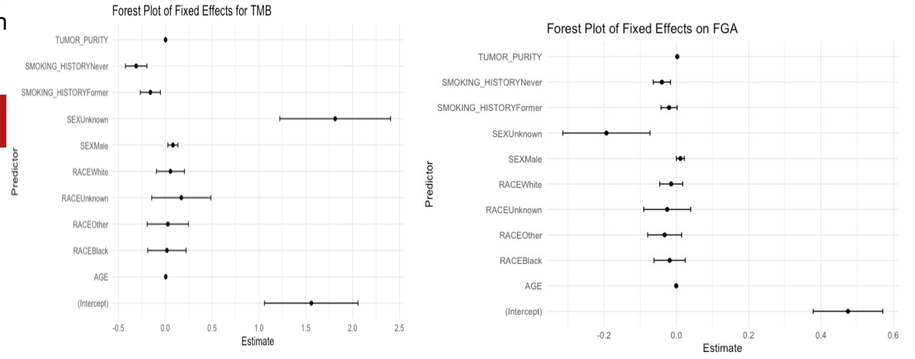


Fig1: Fixed Effect

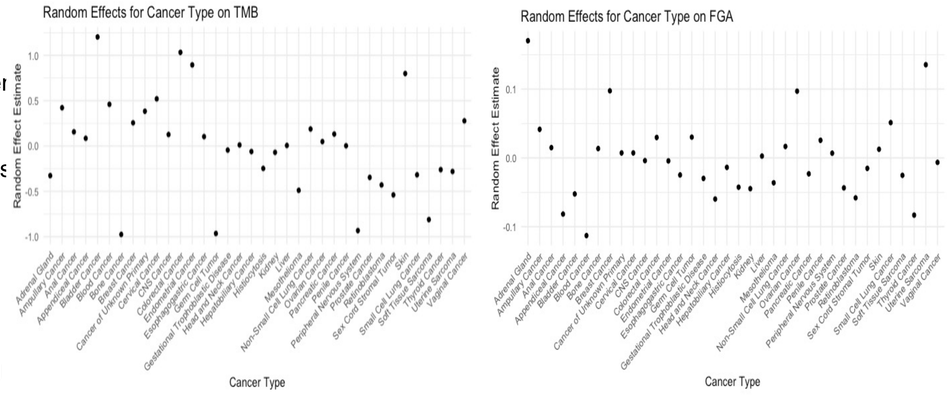


Fig2: Random Effect

$$\log_2(\text{TMB}+1) \sim \beta_0 + \beta_1 \cdot \text{AGE} + \beta_2 \cdot \text{SEX} + \beta_3 \cdot \text{RACE} + \beta_4 \cdot \text{TUMOR\_PURITY} + \beta_5 \cdot \text{SMOKING\_HISTORY} + (1|\text{PATIENT\_ID}) + (1|\text{Study}) + (1 + \text{SMOKING\_HISTORY}|\text{CANCER\_TYPE})$$

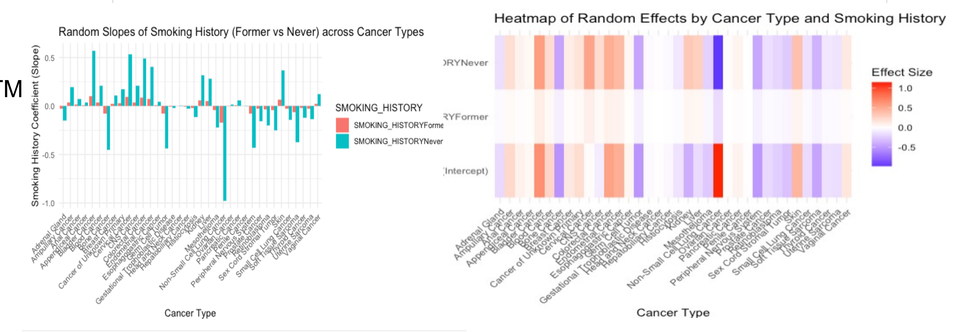
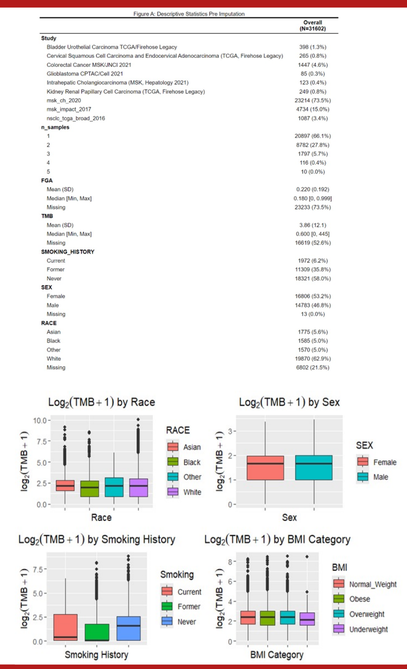


Fig3: Effect of Smoking History over Canver Types

Figure A: Descriptive Statistics Pre Imputation

	Overall (N=31602)
Study	
Bladder Urothelial Carcinoma TCGA/Firehose Legacy	398 (1.3%)
Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (TCGA, Firehose Legacy)	265 (0.8%)
Colorectal Cancer MSKCANCER 2021	1447 (4.6%)
Glioblastoma CPTAC/Cell 2021	85 (0.3%)
Intrahepatic Cholangiocarcinoma (MSK, Hepatology 2021)	123 (0.4%)
Kidney Renal Papillary Cell Carcinoma (TCGA, Firehose Legacy)	249 (0.8%)
msk_ch_2020	23214 (73.5%)
msk_impact_2017	4734 (15.0%)
nsclc_tcga_broad_2016	1087 (3.4%)
n_samples	
1	20897 (66.1%)
2	8792 (27.8%)
3	1797 (5.7%)
4	116 (0.4%)
5	10 (0.0%)
FGA	
Mean (SD)	0.220 (0.192)
Median [Min, Max]	0.180 [0, 0.999]
Missing	23233 (73.5%)
TMB	
Mean (SD)	3.86 (12.1)
Median [Min, Max]	0.600 [0, 443]
Missing	16619 (52.6%)
SMOKING_HISTORY	
Current	1972 (6.2%)
Former	11309 (35.8%)
Never	18321 (58.0%)
SEX	
Female	16806 (53.2%)
Male	14783 (46.8%)
Missing	13 (0.0%)
RACE	
Asian	1775 (5.6%)
Black	1585 (5.0%)
Other	1570 (5.0%)
White	19670 (62.9%)
Missing	6802 (21.5%)



CANCER_TYPE	
Adrenal Gland	75 (0.2%)
Ampullary Cancer	64 (0.2%)
Anal Cancer	70 (0.2%)
Appendiceal Cancer	200 (0.6%)
Bladder Cancer	1220 (3.9%)
Blood Cancer	154 (0.5%)
Bone Cancer	340 (1.1%)
Breast Cancer	4003 (12.7%)
Cancer of Unknown Primary	726 (2.3%)
Cervical Cancer	400 (1.3%)
CNS Cancer	1816 (5.7%)
Colorectal Cancer	4164 (13.2%)
Endometrial Cancer	940 (3.0%)
Esophagogastric Cancer	1298 (4.0%)
Germ Cell Tumor	475 (1.5%)
Gestational Trophoblastic Disease	14 (0.0%)
Head and Neck Cancer	663 (2.1%)
Hepatobiliary Cancer	947 (3.0%)
Histiocytosis	2 (0.0%)
Kidney	996 (3.1%)
Mesothelioma	249 (0.8%)
Non-Small Cell Lung Cancer	4942 (15.6%)
Ovarian Cancer	771 (2.4%)
Pancreatic Cancer	1612 (5.1%)
Penile Cancer	11 (0.0%)
Peripheral Nervous System	133 (0.4%)
Prostate Cancer	1794 (5.7%)
Retinoblastoma	44 (0.1%)
Sex Cord Stromal Tumor	42 (0.1%)
Skin	1280 (4.1%)
Small Cell Lung Cancer	214 (0.7%)
Soft Tissue Sarcoma	1140 (3.6%)
Thyroid Cancer	537 (1.7%)
Uterine Sarcoma	208 (0.7%)
Vaginal Cancer	16 (0.1%)
Missing	102 (0.3%)
TUMOR_PURITY	
Mean (SD)	44.2 (21.6)
Median [Min, Max]	40.0 [10.0, 100]
Missing	25671 (81.2%)
AGE	
Mean (SD)	59.0 (15.4)
Median [Min, Max]	61.0 [0.120, 98.7]
Missing	5625 (18.4%)

DISCUSSION/CONCLUSION

- Possible that small effect despite significance in individual factor analysis due to the large datasets used, needs further exploration
- There was little data available on certain levels in factors, not reflective of true populations - future studies should include data from more diverse population to have greater generalizability
- Patient ID, study, and cancer type significantly contributed to variability in genomic alterations.
- Age, male gender, tumor purity, and current smoking status were significant predictors of genomic alterations.
- Cancer type moderated the effect of smoking history on TMB, with varying magnitude and direction across different cancer types.
- The model with crossed random effects provided the best fit based on AIC and BIC values.
- Large sample sizes revealed significant individual differences, highlighting the need for personalized treatment.

REFERENCES

Bladder Urothelial Carcinoma (TCGA, Firehose Legacy). (2020, September 15). Mskcc.org. https://datacatalog.mskcc.org/dataset/10466

Boerner, T., Drill, E., Pak, L. M., Nguyen, B., Sigel, C. S., Alexandre Doussot, Shin, P., Goldman, D. A., Mithat
Gönen, Allen, P. J., Balachandran, V. P., Cercek, A., Harding, J. J., Solit, D. B., Schultz, N., Kundra, R.,
Walch, H., D'Angelica, M. I., DeMatteo, R. P., & Drebin, J. A. (2021). Genetic Determinants of Outcome
in Intrahepatic Cholangiocarcinoma. Hepatology, 74(3), 1429–1444. https://doi.org/10.1002/hep.31829

cBioPortal for Cancer Genomics. (n.d.). Www.cbioportal.org. https://www.cbioportal.org/

Cercek, A., Chatila, W. K., Yaeger, R., Walch, H., Dos, G., Krishnan, A., Lerie Palmaira, Maio, A., Kemel, Y.,
Srinivasan, P., Chaitanya Bandlamudi, Salo-Mullen, E. E., Prince Rainier Tejada, Kimeisha Belanfanti,
Galle, J., Vijai, J., Segal, N. H., Varghese, A. M., Diane Lauren Reidy, & Shia, J. (2021). A Comprehensive
Comparison of Early-Onset and Average-Onset Colorectal Cancers. 113(12), 1683–1692. https://doi.org/10.1093/jnci/djab124

Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma. (2020, October 20). Mskcc.org. https://datacatalog.mskcc.org/dataset/10476

Kidney Renal Papillary Cell Carcinoma (TCGA, Firehose Legacy). (2020, October 27). Mskcc.org. https://datacatalog.mskcc.org/dataset/10480

Seldon, C., Karthik Meiyappan, Hoffman, H. I., Guo, J. A., Goel, N., Hwang, W. L., Nguyen, P. L.,
Mahal, B. A., & Alshalalfa, M. (2022). Genomic alterations predictive of poor clinical outcomes in pan- cancer. Oncotarget, 13(1), 1069–1077. https://doi.org/10.18632/oncotarget.28276

Smith, J., Parl, F. F., & Dupont, W. D. (2023). Mutation Burden Independently Predicts Survival
in the Pan-Cancer Atlas. 7. https://doi.org/10.1200/po.22.00571

The Pan-Cancer Atlas. (2013). Cell.com. https://www.cell.com/pbassets/consortium/PanCancerAtlas/PanCani3/index.html

Uterine Corpus Endometrial Carcinoma (TCGA, Firehose Legacy). (2020, November 19). Mskcc.org. https://datacatalog.mskcc.org/dataset/10494

Wang, L.-B., Karpova, A., Gritsenko, M. A., Kyle, J. E., Cao, S., Li, Y., Rykunov, D., Colaprico, A., Rothstein, J. H.,
Hong, R., Stathias, V., Cornwell, M., Petralia, F., Wu, Y., Reva, B., Krug, K., Pugliese, P., Kawaler, E.,
Olsen, L. K., & Liang, W.-W. (2021). Proteogenomic and metabolomic characterization of human glioblastoma. Cancer Cell, 39(4), 509-528.e20. https://doi.org/10.1016/j.ccell.2021.01.006

Wan, L., Wang, Z., Xue, J., Yang, H., & Zhu, Y. (2020). Tumor mutation burden predicts response and survival to immune checkpoint inhibitors: a meta-analysis. Translational Cancer Research, 9(9), 5437–5449. https://doi.org/10.21037/tcr-20-1131

PanCanAtlas Publications | NCI Genomic Data Commons. (n.d.). Gdc.cancer.gov. https://gdc.cancer.gov/about-data/publications/pancanatlas

Samstein, R. M., Lee, C.-H., Shoushtari, A. N., Hellmann, M. D., Shen, R., Janjigian, Y. Y., Barron, D. A., Zehir,
A., Jordan, E. J., Omuro, A., Kaley, T. J., Kendall, S. M., Motzer, R. J., Hakimi, A. A., Voss, M. H., Russo, P.,
Rosenberg, J., Iyer, G., Bochner, B. H., & Bajorin, D. F. (2019). Tumor mutational load predicts survival
after immunotherapy across multiple cancer types. Nature Genetics, 51(2), 202–206. https://doi.org/10.1038/s41588-018-0312-8