Genetic prediction of colitis in non-small cell lung cancer patients on immune checkpoint inhibitor therapy

Pooja Middha\textsuperscript{1}, Rohit Thummalapalli\textsuperscript{2}, Michael J. Betti\textsuperscript{3}, Zoe Quandt\textsuperscript{4,5}, Karmugi Balaratnam\textsuperscript{6}, Eduardo Cardenas\textsuperscript{1}, Christina J. Falcon\textsuperscript{7}, David M. Faleck\textsuperscript{8}, Princess Margaret Lung Group\textsuperscript{6}, Matthew A. Gubens\textsuperscript{9,10}, Scott Huntsman\textsuperscript{1}, Linda Kachuri\textsuperscript{11,12}, Khaleeq Khan\textsuperscript{6}, Min Li\textsuperscript{1}, Christine M. Lovly\textsuperscript{13}, Megan H. Murray\textsuperscript{14}, Devalben Patel\textsuperscript{6}, Kristin Werking\textsuperscript{15}, Luna Jia Zhan\textsuperscript{6}, Geoffrey Liu\textsuperscript{16}, Melinda C. Aldrich\textsuperscript{3}, Adam J. Schoenfeld\textsuperscript{17}, Elad Ziv\textsuperscript{1,18}

\textsuperscript{1} Department of Medicine, University of California San Francisco, San Francisco, CA, USA
\textsuperscript{2} Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY
\textsuperscript{3} Department of Medicine, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
\textsuperscript{4} Division of Endocrinology and Metabolism, Department of Medicine, University of California San Francisco, San Francisco, CA, USA
\textsuperscript{5} Diabetes Center, University of California San Francisco, San Francisco, CA, USA
\textsuperscript{6} Princess Margaret Cancer Centre, Toronto, ON, Canada
\textsuperscript{7} Fiona and Stanley Druckenmiller Center for Lung Cancer Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA
\textsuperscript{8} Gastroenterology, Hepatology & Nutrition Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA
\textsuperscript{9} Medical Oncology, University of California San Francisco, San Francisco, CA, USA
\textsuperscript{10} Department of Medicine, Weill Cornell Medical Center, New York, NY, USA
\textsuperscript{11} Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, CA, USA
\textsuperscript{12} Stanford Cancer Institute, Stanford University of Medicine, Stanford, CA, USA
\textsuperscript{13} Department of Medicine, Division of Hematology and Oncology, Vanderbilt University Medical Center and Vanderbilt Ingram cancer center, Nashville, TN, USA
\textsuperscript{14} Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN
\textsuperscript{15} Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN, USA
\textsuperscript{16} Princess Margaret Cancer Centre, Temerty School of Medicine, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
\textsuperscript{17} Thoracic Oncology Service, Memorial Sloan Kettering Cancer Center, New York, New York
\textsuperscript{18} Helen Diller Family Comprehensive Cancer Center, Center for Genes, Environment and Health and Institute for Human Genetics, University of California San Francisco, San Francisco, California.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract
Immune checkpoint inhibitors (ICIs) are a remarkable advancement in cancer therapeutics; however, a substantial proportion of patients develop severe immune-related adverse events (irAEs). Understanding and predicting irAEs is the key to advancing precision immuno-oncology. Immune-mediated colitis (IMC) is a significant complication from ICI and can have life-threatening consequences. Genetic susceptibility to Crohn’s disease (CD) and ulcerative colitis (UC) may predispose to IMC, but the link is poorly understood. We developed and validated polygenic risk scores for CD (PRS_CD) and UC (PRS_UC) in cancer-free individuals and assessed the role of each of these PRSs on IMC in a cohort of 1,316 patients with non-small cell lung cancer (NSCLC) who received ICIs. Prevalence of all-grade IMC in our cohort is 4% (55 cases), and for severe IMC, 2.5% (32 cases). The PRS_UC predicted the development of all-grade IMC (HR=1.34 per standard deviation [SD], 95% CI=1.02-1.76, P=0.04) and severe IMC (HR=1.62 per SD, 95% CI=1.12-2.35, P=0.01). PRS_CD was not associated with IMC or severe IMC. This is a first study to demonstrate the potential clinical utility of a PRS for ulcerative colitis in identifying NSCLC patients receiving ICI at high risk of developing IMC, where risk reduction and close monitoring strategies could help improve overall patient outcomes.

Introduction
Immunotherapy with immune checkpoint inhibitors (ICI) has substantially improved clinical outcomes in advanced cancers such as melanoma, non-small cell lung cancer (NSCLC), bladder, renal, breast, and other cancers1–6. ICIs block the ability of malignant cells to escape detection through immune checkpoints such as programmed cell death protein 1/ programmed cell death ligand 1 (PD-1/PD-L1) or cytotoxic T-lymphocyte associated protein 4 (CTLA-4). Blockade of these checkpoints restores host immunosurveillance in some tumors by stimulating cytotoxic T-cells to induce cancer cell apoptosis7–9.

Despite ICIs being a paradigm-shifting breakthrough in cancer treatment, enhanced activation of the immune system can lead to immune-related adverse events (irAEs) that can result in permanent discontinuation of ICIs, severe morbidity, and even patient
The most severe irAEs include hypophysitis, diabetes, colitis, hepatitis, and pneumonitis, with other common irAEs including rash and thyroiditis. The incidence of immune checkpoint inhibitor-mediated colitis (IMC) ranges from 1%-25% and varies by ICI therapy. Endoscopic and histological findings suggest that the presentation of IMC mimics autoimmune colitis such as ulcerative colitis (UC), a form of inflammatory bowel disease (IBD). Severe IMC can lead to hospitalization and treatment cessation in up to 15–30% of patients receiving ICIs. Despite the phenotypic similarities between IBD and IMC, it is unclear if the underlying mechanism is shared or distinct.

We sought to characterize the relationship between genetic predisposition to common types of autoimmune colitis (ulcerative colitis (UC) and Crohn’s disease (CD)), and IMC in a cohort of NSCLC patients receiving ICI treatment. We first developed polygenic risk scores (PRS) for UC and CD using individuals not diagnosed with cancer at baseline in UK Biobank (UKB) and validated these PRSs in an independent dataset of cancer-free participants in Vanderbilt University Medical Center biobank (BioVU). We then evaluated the association between each of these PRS and the development of IMC in NSCLC patients on ICI therapy. We further investigated the association between human leukocyte antigen (HLA) alleles known to affect UC risk with IMC. Additionally, we examined the role of IMC and PRS for UC, and CD, respectively, on progression free survival (PFS) and overall survival (OS).

Results
Patient Characteristics

We analyzed data from 1,316 study participants included in the GeRI cohort, which included four sites (Table 1 and See Methods). The cohort comprised approximately 50% men and the mean age at diagnosis was 65 years (+/-10.3). A small proportion (9%) received the combination therapy of anti PD-1/PD-L1 and anti CTLA-4 inhibitors and the remainder received either anti PD-1 or PD-L1 inhibitors (91%). The cumulative incidence of IMC was ~4% (55 events); it was ~2% (32 events) for severe IMC. The rates were similar across all study sites. The analytic strategy of our study is illustrated in Figure 1.
Development and validation of PRS for UC and CD

We used 70% of the cancer-free UKB dataset to tune parameters for PRS using LDpred\textsuperscript{20}. We then obtained effect estimates for the PRS for CD and UC in the remaining 30% (testing data). In the UKB testing data, the area under the receiver operating curve (AUROC) for the PRS\textsubscript{UC} was 0.66 (95% CI = 0.64-0.68), and the AUROC for PRS\textsubscript{CD} was 0.72 (95% CI = 0.69-0.74) (Supplementary Figure 1). The adjusted PRS\textsubscript{UC} was strongly associated with UC with an odds ratio (OR) of 1.84 per standard deviation (SD) (95% CI = 1.76-1.93, p<1.0x10\textsuperscript{-12}). Similarly, PRS\textsubscript{CD} was positively associated with CD with OR of 1.83 per SD (95% CI = 1.72-1.95, p<1.0x10\textsuperscript{-12}).

The two PRSs were validated in another sample of cancer-free individuals from BioVU\textsuperscript{21}. Similar to the UKB results, the individual PRS\textsubscript{CD} and PRS\textsubscript{UC} were also strongly associated with CD and UC in BioVU. We observed an OR of 2.18 per SD (95% CI = 2.05-2.32, p<1.0x10\textsuperscript{-12}) for PRS\textsubscript{CD}, and an OR of 1.75 per SD (95% CI = 1.59-1.92, p<1.0x10\textsuperscript{-12}) for PRS\textsubscript{UC}. The AUROC for PRS\textsubscript{CD} and PRS\textsubscript{UC} were 0.72 (95% CI = 0.70-0.73) and 0.65 (95% CI = 0.62-0.68), respectively (Supplementary Figure 2).

PRS of autoimmune colitis as a predictor of IMC

The mean PRS\textsubscript{UC} was significantly higher in patients who developed IMC (Supplementary Figure 3). We examined the cumulative incidence of IMC (all-grade and severe) in the top 10\textsuperscript{th} percentile (high genetic risk), 10-90\textsuperscript{th} percentile (average genetic risk), and lowest 10\textsuperscript{th} percentile (low genetic risk) of the PRS\textsubscript{UC}. Individuals in the top 10\textsuperscript{th} percentile of the PRS\textsubscript{UC} had higher rates of IMC (all-grade: p=0.01 and severe: p=0.03) compared to other two categories (Figure 2). Using Cox proportional hazards model and adjusting for genetic ancestry, recruiting site, age, sex, cancer histology, and type of therapy, we observed that the PRS\textsubscript{UC} was significantly associated with any diagnosis of IMC in the GeRI cohort with a hazard ratio (HR) of 1.34 per SD (95% CI = 1.02-1.76, p=0.04). For a diagnosis of severe IMC, the HR per SD was 1.62 (95% CI = 1.12-2.35, p=0.01) (Table 2). We found no significant association between PRS\textsubscript{CD} and IMC or severe IMC (Table 2).
Additionally, we conducted stratified analysis by type of therapy and histology of lung cancer to further characterize the association between PRS\textsubscript{UC} and IMC (all-grade and severe). For all-grade IMC, the results showed little attenuation and nominal significance when stratified by type of therapy (Table 2). However, for severe IMC we observed a HR per SD of 1.51 (95% CI = 1.01-2.27, p=0.04) in patients receiving anti-PD1/anti PD-L1 monotherapy versus a HR per SD of 4.31 (95% CI = 1.08-17.24, p=0.03) in those patients receiving a combined therapy. Patients with adenocarcinoma had a HR per SD of 1.43 (95% CI = 1.06-1.93, p=0.02) for all-grade IMC and a HR per SD of 2.12 (95% CI = 1.37-3.26, p=6x10\textsuperscript{-04}) for severe IMC.

**Role of known UC-HLA associations on IMC in GeRI cohort**

We assessed the association between all-grade IMC and HLA markers known to be associated with ulcerative colitis\textsuperscript{22,23} (Supplementary Figure 4). Out of 12 known UC-associated HLA markers, we observed an OR of 2.63 (95% CI = 1.08-6.40, p=0.03) for HLA-DRB1*12:01 and all-grade IMC. However, at false discovery rate (FDR)<0.05 none of the known HLA markers were associated with all-grade IMC in the GeRI cohort.

**IMC and PRS of autoimmune colitis as a predictor of PFS and OS**

To assess the role of IMC on clinical outcomes, we first evaluated the association between IMC (all-grade and severe) and PFS in the GeRI cohort. An IMC event was significantly associated with improved PFS (HR = 0.54, 95% CI = 0.38-0.77, p=6.5x10\textsuperscript{-04}). This association was stronger with a HR of 0.43 (95% CI = 0.25-0.73, p=1.6x10\textsuperscript{-03}) with severe IMC. However, the 90-day landmark analysis showed attenuation of association for both all-grade IMC and severe IMC (Table 3 and Supplementary Figure 5). We observed the effect of all-grade IMC on OS with a HR of 0.37 (95% CI = 0.23-0.61, p=1.0x10\textsuperscript{-04}) and of severe IMC on OS with a HR of 0.21 (95% CI = 0.09-0.50, p=4.8x10\textsuperscript{-04}). Moreover, the association between OS and IMC remained significant in the 90-day landmark sensitivity analyses (Table 3 and Figure 4).

Despite the association between PRS\textsubscript{UC} and IMC, PRS\textsubscript{UC} was not associated with PFS (HR per SD = 1.00, 95% CI = 0.94-1.07, p=0.95) and OS (HR per SD = 1.01, 95% CI =
0.93-1.09, p=0.91) in our cohort. Similarly, we observed no association between PRS_{CD} and PFS (HR per SD = 0.97, 95% CI = 0.91-1.05, p=0.47) and OS (HR per SD = 1.02, 95% CI = 0.93-1.11, p=0.68), respectively.

**Discussion**

Immune checkpoint inhibitors are part of standard regimens to treat advanced unresectable NSCLC and in the adjuvant and neoadjuvant settings in early-stage NSCLC\textsuperscript{5,24–31}. Immune-related adverse events are common complications from ICI, and there are few predictors of irAEs\textsuperscript{32–36}. We sought to identify genetic predictors of immune-mediated colitis which frequently results in hospitalization and ICI discontinuation and can occasionally lead to death\textsuperscript{15,16}. Specifically, we evaluated the relationship between genetic predisposition for autoimmune colitis (UC, CD) and IMC, and found that the PRS_{UC} performed well in predicting IMC. The association was stronger when analyses were restricted to those individuals who had severe IMC. This is important since the most important clinical cases to identify were best predicted by PRS_{UC}. Furthermore, we assessed the role of HLA markers associated with UC on development of IMC and found no significant associations after adjusting for multiple hypothesis testing, suggesting that the association between UC and IMC genetics is not entirely driven by HLA.

Our findings have implications for the clinical understanding of IMC and may also impact management of patients treated with ICIs. First, we demonstrate that IMC has some genetic overlap with UC but we found no evidence for overlap with CD. This is consistent with clinical reports in which the most frequent phenotype of IMC resembles UC most closely\textsuperscript{17,18}. Our results also suggest that as genetic risk prediction of UC improves, genetic risk of IMC may also be improved. In particular, rare variants in certain genes substantially increase the risk of UC and we hypothesize may also affect IMC risk\textsuperscript{37–40}. Prior reports on ICI-induced hypothyroidism\textsuperscript{34,35} and rash\textsuperscript{36} demonstrated that PRS for autoimmune disorders predict irAEs, suggesting that ICI may unmask autoimmune syndromes in some genetically predisposed individuals.
We also found that individuals who developed IMC had improved survival outcomes when compared to those who did not develop IMC, including in a landmark sensitivity analysis, which is concordant with previously published literature\textsuperscript{41–46}. However, PRS\textsubscript{UC} and PRS\textsubscript{CD} were not associated with PFS or OS, suggesting that the genetic basis of autoimmune disease susceptibility is distinct from genetic factors influencing survival outcomes. It has been postulated that both anti-tumor responses to ICIs, and development of irAEs are representative of a robust immune response; however, one possible explanation for our finding is that the genetic contributions captured in the autoimmune PRSs are probably capturing the cross-presentation of shared antigens which may not be associated with clinical outcomes. This suggests there could be other genetic and environmental factors driving the association between IMC and overall survival.

Our study has implications that may impact future clinical trials. For example, our results could be used to select patients at high risk of IMC in a clinical trial setting to assess the role of preventative measures such as the commencement of concurrent anti-TNFα therapies or anti-integrin α4β7 antibodies\textsuperscript{47,48} along with ICI treatment in early stages of NSCLC treatment. Additionally, these findings could facilitate clinical decision-making. Several studies and trials have shown the benefit of combination immunotherapies but with substantial increase in the risk of irAEs\textsuperscript{44,49–53}. Our stratified analysis by type of therapy showed that association between PRS\textsubscript{UC} and severe IMC remains consistent in individuals receiving anti PD-1/PD-L1 and anti CTLA4 combination therapy versus those receiving anti PD-1/PD-L1 monotherapy. Therefore, among patients who may be candidates for combination immunotherapies but have high genetic risk based on PRS\textsubscript{UC}, oncologists may consider monotherapy, particularly if the benefits of dual therapy are modest. Conversely, patients who are at relatively low risk based on PRS\textsubscript{UC} may be better candidates for combination therapy.

Although our study has important clinical implications, it also has a few limitations. First, our cohort only included NSCLC. IMC is a common complication of treatment with ICI in many tumor types and we suspect that PRS\textsubscript{UC} will predict IMC in all patients on ICI, but more data are needed to validate this hypothesis. Second, most of our patients were
treated with PD-1 or PD-L1 monotherapy. Patients on combination therapy have a much higher risk of IMC. Our stratified analyses demonstrated a consistent effect of PRS<sub>UC</sub> in the dual therapy group, but the sample size of that analysis was small and larger studies in dual therapy treated patients are required to confirm the effect. Similarly, we found a strong effect of PRS<sub>UC</sub> when analyses were restricted to adenocarcinoma. However, we observed no effect of PRS<sub>UC</sub> on IMC risk among patients with squamous cell lung cancer; however, due to small sample size, our study may be underpowered to detect associations in squamous cell carcinoma patients and larger studies will be required to validate the association. We developed PRSs in a predominantly European ancestry cohort (UK Biobank) and the GeRI cohort was also predominantly of European ancestry; more work is needed to generalize these results to other ancestries. Our study comprises advanced NSCLC patients receiving ICI therapy; therefore, another possible limitation could be survivor bias<sup>54,55</sup> where patients who respond to therapy and are on therapy longer are at an increased risk of developing irAEs. The impact of the survivor bias in our study was controlled for by using a Cox model with 90-day landmark (landmark analysis)<sup>56</sup> for both PFS and OS.

Overall, our findings suggest a shared genetic basis between ulcerative colitis and immune-mediated colitis among patients undergoing ICI treatment. Prediction of IMC using genetic information should create new opportunities for better risk stratification and ultimately for better management and possibly prevention of this common and important side effect from immunotherapy.

**Methods**

**Study Population**

Genetics of immune-related adverse events and Response to Immunotherapy (GeRI) cohort is comprised of 1,316 advanced Stage IIIB/IV NSCLC patients who received ICI therapy (PD-1 or PD-L1 inhibitors as monotherapy or in combination with either CTLA-4 inhibitors and/or chemotherapy) and were recruited from four different institutions: Memorial Sloan Kettering Cancer Center (MSKCC), Vanderbilt University Medical Center
(VUMC), Princess Margaret Cancer Center (PM), and University of California, San Francisco (UCSF).

A total of 752 individuals were treated with ICIs at MSKCC between 2011 and 2018 and had an available blood sample. Clinical data were extracted from a manual review of medical and pharmacy records for demographics, lung cancer histology, and ICI treatment history, including detailed information on immune-related adverse events (irAEs). The VUMC cohort is composed of 267 patients who received ICI therapy at the medical center between 2009 and 2019. Patients participated in BioVU, Vanderbilt’s biomedical repository of DNA that is linked to de-identified health records. Treatment dates and irAEs were extracted using manual chart review by a trained thoracic oncology nurse. The PM cohort included 266 advanced NSCLC patients who received ICI therapy between 2011 and 2022; all provided a blood sample and completed a questionnaire. Clinical data were manually extracted by trained abstractors, supplemented by the PM Cancer Registry. From UCSF, 31 patients who had received ICIs were identified by thoracic oncologists between 2019 and 2021 and provided either a blood or saliva sample after informed consent. Clinical data including, demographics, history of lung cancer and ICI treatment, and irAEs were extracted after manual review of electronic health records. Institutional Review Board approvals were obtained at each site individually and written informed consent was acquired from all study participants prior to inclusion in the study.

Immune checkpoint inhibitor-mediated colitis (IMC)

After the initiation of ICI therapy, immune checkpoint inhibitor-mediated colitis (IMC) was defined based on clinical chart review and documentation of IMC by the primary oncologist, gastroenterologist, and/or other clinicians treating the patient based on clinical features and/or radiologic/histologic evidence suggesting colitis due to ICI. Participants who were diagnosed with infectious causes of colitis including *Clostridium difficile*, or a pathogen on a gastrointestinal pathogen panel or ova and parasite test were excluded. To assess the severity of IMC, we used 2 metrics based on NCI Common Terminology Criteria for Adverse Events Version 5 (NCI-CTCAE) that captures grade 3 IMC or above:
(i) hospitalization for management of IMC and/or (ii) permanent cessation of ICI therapy due to the adverse event.

IMC was coded as a dichotomous variable (1: all IMC, 0: no IMC) and time-to-IMC was assessed from start of the ICI therapy to the date of onset of IMC or date of ICI discontinuation due to IMC. Patients who did not experience IMC were censored either at the end of treatment due to any reason or last follow-up date if the treatment was ongoing. Based on the severity criteria, severe IMCs were also coded as binary variables (1: severe IMC, 0: no IMC).

**Ascertainment of clinical outcomes**

Progression free survival (PFS) and overall survival (OS) were evaluated from the date of initiation of ICI therapy to date of progression and death, respectively, at MSK, PM and UCSF sites. At VUMC, time-to-discontinuation of therapy due to progression from therapy initiation was used as a surrogate. If the treatment was ongoing, patients were censored at the date of last follow-up. The VUMC cohort is de-identified and not linked to the National Death Index; therefore, all-cause mortality (overall survival) information is unavailable for VUMC participants (n=267).

**Quality control, genotyping, and imputation of GeRI cohort**

DNA from blood or saliva was extracted and genotyped using Affymetrix Axiom Precision Medicine Diversity Array. Samples with a call rate <95% were excluded from the analysis and SNPs with missing rates >5% were also excluded from the analysis. Genetic ancestry was calculated using principal component analysis in PLINK after linkage disequilibrium pruning ($R^2$<0.1). Imputation was performed using the Michigan Imputation Server with the 1000 Genomes phase3 v5 reference panel. Standard genotyping and quality control procedures were implemented. Variants with minor allele frequency <0.01 were excluded from the analysis.

**Development and validation of polygenic risk score (PRS) for autoimmune colitis**
We used the LDpred2\textsuperscript{20} method to develop PRS of CD and UC. LDpred2 estimates the posterior effect sizes based on summary statistics from genome-wide association study while taking into account the linkage disequilibrium between variants and assuming a prior on the markers. To derive PRS, summary statistics were obtained from previously published genome-wide association study of CD, and UC\textsuperscript{57}. We developed PRS for CD (1,312 CD cases and 16,303 controls) and UC (2,814 UC cases and 16,303 controls), separately using UK Biobank (UKB) data, where we divided the data into two parts: 70% for hyperparameter tuning and 30% of the remaining data for testing the PRS. Genetic data from both the UKB Affymetrix Axiom array (89%) or the UK BiLEVE array (11%)\textsuperscript{58} which have been imputed using the Haplotype Reference Consortium and the UK10K and 1000 Genomes phase 3 reference panels\textsuperscript{58} were utilized in the analysis. Information from both self-report and ICD9/10 codes were used to capture CD (1,312 cases) and UC (2,814 cases) phenotype in UKB. Analyses were restricted to European ancestry individuals based on self-reported White ethnicity and genetic ancestry PCs within five standard deviations of the population mean. PRS was constructed using the formula: 

\begin{equation}
\text{PRS} = \beta_1 \times \text{SNP}_1 + \beta_2 \times \text{SNP}_2 + \ldots + \beta_n \times \text{SNP}_n,
\end{equation}

where \( \beta \) was estimated using LDpred2-auto function. Each PRS was standardized to have a mean of zero and standard deviation of 1. The association of \( \text{PRS}_{\text{CD}} \) and \( \text{PRS}_{\text{UC}} \) with each respective target phenotype was assessed using logistic regression models, adjusted for age at diagnosis for cases and age at enrollment for controls, sex, genotyping array, and the top 10 genetic ancestry principal components (PCs). Area under the receiver operating characteristic (AUROC) curves were calculated in the testing dataset and used to assess the overall prediction accuracy of each the PRS in UKB.

We validated the two PRSs in a sample of cancer-free individuals (1,420 CD cases, 459 UC cases, and 20,876 controls in the VUMC BioVU\textsuperscript{21}). All analyses were restricted to individuals of European ancestry and adjusted for age, sex, and ten principal components. AUROC curves were used to estimate the prediction of the PRSs.

\textit{Assessment of autoimmune colitis PRS to predict IMC in GeRI cohort}
Using the weights generated from LDpred2 for CD, and UC, we separately calculated two weighted PRSs (PRS\textsubscript{CD}, PRS\textsubscript{UC}) for the GeRI participants. Cumulative incidence of IMC (all-grade and severe) was assessed by categories of PRS percentiles. Individuals in the top 10% of the PRS distribution (PRS > 90th percentile) were classified as having high genetic risk, those in the bottom 10% (PRS ≤ 10th percentile) were classified as low risk, and the middle category (> 10th to ≤ 90th percentile) classified as average genetic risk. Additionally, to evaluate the performance of each potential PRS on either time-to-IMC or time-to-severe-IMC, we used Cox proportional hazards models, adjusted for age at diagnosis, sex, lung cancer histology, type of therapy, recruiting site, and the first 5 genetic ancestry PCs. To further understand the differential effects of type of therapy and histology on the association between PRS\textsubscript{UC} and IMC, we conducted stratified analysis by type of ICI therapy and histology of lung cancer.

Role of HLA markers associated with UC and CD on IMC in GeRI cohort

To elucidate the role of known UC-associated HLA markers on IMC, we performed HLA imputation using CookHLA\textsuperscript{59} and HATK\textsuperscript{60}. HLA alleles were imputed at 2-digit resolution against the Type 1 Diabetes Genetics Consortium reference panel\textsuperscript{61} and using the nomenclature from IPD-IMGT/HLA database v3.51. Association analysis with all-grade IMC was conducted using logistic regression models adjusted for age at diagnosis, sex, lung cancer histology, type of therapy, recruiting site, and 5 PCs. Analyses were restricted to common HLA alleles (frequency ≥ 0.01) known to be associated with ulcerative colitis\textsuperscript{23}.

Impact of IMC and PRS of autoimmune colitis on PFS and OS in GeRI cohort

Association of IMC (all-grade and severe) on PFS and OS was examined using the Cox proportional hazards model. All models were adjusted for age at diagnosis, sex, lung cancer histology, type of therapy, and 5 PCs. Survival curves and rates were estimated using Kaplan-Meier method. As a sensitivity analysis, we conducted a landmark analysis by examining only the patients who had PFS and OS longer than 90 days. To investigate the association between PRS\textsubscript{CD}, PRS\textsubscript{UC} on PFS and OS, we conducted Cox proportional hazards models, adjusted for age at diagnosis, sex, histology, type of therapy, and 5 PCs.
All $P$ values are two-sided, and analyses were conducted using Plink2, R v4.2.2 (R foundation for Statistical Computing) with RStudio v2022.12.0.353.

References


32. Zhang Y, Zhang X, Li W, Du Y, Hu W, Zhao J. Biomarkers and risk factors for the early...


**Funding**

This work was supported by the National Institutes of Health R01-CA227466 and K24-CA169004 to E.Ziv; C.M.Lovly was supported in part by NIH NCI UG1CA233259, P01CA129243, and P30CA068485; R.Thummalapalli was supported by T32-CA009207; The Lusi Wong Fund, Posluns Fund, Alan Brown Chair in Molecular Genomics, Princess Margaret Cancer Foundation were awarded to G. Liu for this work; M.C. Aldrich was supported in part by R01-CA227466, U01CA253560, R01CA251758 and the Vanderbilt Institute for Clinical and Translational Research (UL1TR002243); Z. Quandt was supported by the American Diabetes Association Grant (1–19-PDF-131; A.J.Schoenfeld, D.Faleck were supported by the Memorial Sloan Kettering Cancer Center Support Grant/Core (P30-CA008748), the Druckenmiller Center for Lung Cancer Research at Memorial Sloan Kettering Cancer Center.

**Acknowledgements**

Princess Margaret Lung Group: Natasha B. Leighl, Penelope A. Bradbury, Frances A. Shepherd, Adrian G. Sacher, Lawson Eng. Megan H. Murray’s work on this project was completed in August 2022 while she was working at Vanderbilt University Medical Center.
Figure 1: Overview of the analytical pipeline
Development and validation of the polygenic risk scores (PRSs): LDPred2 method was used to tune the parameters for the PRS for ulcerative colitis and Crohn’s disease (PRS\textsubscript{UC}, PRS\textsubscript{CD}) in 70% of the UK Biobank, using the summary statistics from the largest genome-wide association study of UC and CD. The PRSs were then tested in the remaining 30% of the UK Biobank and validated in BioVU. In the next step, the role of PRS\textsubscript{UC} and PRS\textsubscript{CD} on all-grade and severe immune-mediated colitis (IMC) was evaluated in a cohort of 1,316 non-small cell lung cancer patients who received at least one dose of immune checkpoint inhibitor therapy. Finally, associations of all-grade and severe IMC along with PRS\textsubscript{UC} and PRS\textsubscript{CD} on progression-free survival (PFS) and overall survival (OS) were assessed.
Polygenic risk score (PRS) for Ulcerative Colitis (UC) and Crohn’s Disease (CD)

GWAS Summary Statistics (de Lange et al., 2017)
(12194 CD cases, 12366 UC cases and 28072 controls)

PRS as a Predictor of Immune-Mediated Colitis (IMC)

LDPred2

Parameter Tuning

Testing

Validation

UKB (70%)

UKB (30%)

BioVU

918 CD cases
1970 UC cases
11412 controls

394 CD cases
844 UC cases
4891 controls

1420 CD cases
459 UC cases
20876 controls

IMC and PRS as a Predictor of Overall Survival (OS) and Progression-Free Survival (PFS)

Genetics of immune-related adverse events and Response to Immunotherapy (GeRI) cohort

1316 NSCLC patients

All Grade IMC (55 cases)

Severe IMC (32 cases)

HLA imputation and analysis

Individuals

AATAACACCA
AGCACTGATT
TGTCACTAGA
AGTAGTTTAA

CookHLA

Reference panel

IPD-IMGT database

PRS analysis

HLA Alleles

B*03:01
DQB1*15:01
DPB1*13:01
A*35:02

Density

PRS

All Grade IMC

Severe IMC

Immunotherapy (GeRI) cohort

1316 NSCLC patients

Progression (999 cases)

Deceased (639 cases)

PRS as predictor of PFS and OS

PRS as predictor of PFS and OS
Figure 2: Cumulative incidence curves of (i) All-grade immune-mediated colitis (IMC) and (ii) Severe IMC by categories of polygenic risk score of ulcerative colitis (PRSUC) in the entire GeRI cohort. PRSUC is categorized as ≤10th percentile (low genetic risk), 10-90th percentile (average genetic risk), and >90th percentile (high genetic risk).
Figure 3: Immune-mediated colitis (IMC) as a predictor of overall survival (OS) in the entire GeRI cohort (i) and (ii) All-grade IMC, (iii) and (iv) Severe IMC. Kaplan-Meier survival curves are unadjusted and compare those who had an IMC (all-grade or severe) with those who did not have an IMC (No IMC). The p-values in the graph represent the log-rank p-values and the dotted line represents median survival time. Graphs (i) and (iii) are Cox models using all data while graphs (ii) and (iv) are Cox models with 90-day landmark.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>GeRI (n=1316)</th>
<th>MSKCC (n=752)</th>
<th>PM (n=266)</th>
<th>UCSF (n=31)</th>
<th>VUMC (n=267)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age at diagnosis (SD)</td>
<td>65.24 (10.26)</td>
<td>66.13 (10.52)</td>
<td>64.60 (10.48)</td>
<td>65.30 (9.71)</td>
<td>63.33 (9.01)</td>
</tr>
<tr>
<td>Sex n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>659 (50.1)</td>
<td>353 (46.9)</td>
<td>142 (53.4)</td>
<td>12 (38.7)</td>
<td>152 (56.9)</td>
</tr>
<tr>
<td>Female</td>
<td>657 (49.9)</td>
<td>399 (53.1)</td>
<td>124 (46.6)</td>
<td>19 (61.3)</td>
<td>115 (43.1)</td>
</tr>
<tr>
<td>Histology n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td>949 (72.1)</td>
<td>580 (77.1)</td>
<td>193 (72.6)</td>
<td>24 (77.4)</td>
<td>152 (56.9)</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>221 (16.8)</td>
<td>111 (14.8)</td>
<td>46 (17.3)</td>
<td>5 (16.1)</td>
<td>59 (22.1)</td>
</tr>
<tr>
<td>Other</td>
<td>146 (11.1)</td>
<td>61 (8.1)</td>
<td>27 (10.2)</td>
<td>2 (6.5)</td>
<td>56 (21.0)</td>
</tr>
<tr>
<td>Type of therapy n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti PD-1/PD-L1 therapy</td>
<td>1198 (91.0)</td>
<td>671 (89.2)</td>
<td>257 (96.6)</td>
<td>31 (100.0)</td>
<td>239 (89.5)</td>
</tr>
<tr>
<td>Anti PD-1/PD-L1 + Anti CTLA4 therapy</td>
<td>118 (9.0)</td>
<td>81 (10.8)</td>
<td>9 (3.4)</td>
<td>0 (0.0)</td>
<td>28 (10.5)</td>
</tr>
<tr>
<td></td>
<td>IMC n (%)</td>
<td>Mean time in months to IMC (SD)</td>
<td>Severe IMC* n (%)</td>
<td>Progression n (%)</td>
<td>Overall survival n (%)</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------</td>
<td>--------------------------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>55 (4.2)</td>
<td>9.85 (13.37)</td>
<td>32 (2.4)</td>
<td>999 (75.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>32 (4.3)</td>
<td>9.78 (13.42)</td>
<td>15 (2.0)</td>
<td>640 (85.1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13 (4.9)</td>
<td>11.07 (13.75)</td>
<td>10 (3.8)</td>
<td>177 (66.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 (3.2)</td>
<td>(-)</td>
<td>1 (3.2)</td>
<td>22 (71.0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9 (3.4)</td>
<td>8.19 (11.77)</td>
<td>6 (2.2)</td>
<td>160 (59.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1261 (95.8)</td>
<td></td>
<td>720 (95.7)</td>
<td>315 (23.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>253 (95.1)</td>
<td></td>
<td>253 (95.1)</td>
<td>112 (14.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>30 (96.8)</td>
<td></td>
<td></td>
<td>89 (33.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>258 (96.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Missing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2 (0.2)</td>
<td></td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0 (0.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (3.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 (0.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean time in months to progression (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.02 (13.44)</td>
<td></td>
<td>10.11 (13.68)</td>
<td>9.67 (11.13)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>27.07 (23.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>8.18 (11.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overall survival n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Status</td>
<td>Alive at last follow-up</td>
<td>Deceased</td>
<td>Missing</td>
<td>Mean time in months to death (SD)</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>----------</td>
<td>----------------------------------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>410 (31.2)</td>
<td>639 (48.6)</td>
<td>267 (20.3)</td>
<td>22.09 (17.81)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>262 (34.8)</td>
<td>490 (65.2)</td>
<td>0 (0.0)</td>
<td>21.71 (16.99)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>131 (49.2)</td>
<td>135 (50.8)</td>
<td>0 (0.0)</td>
<td>20.84 (18.57)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17 (54.8)</td>
<td>14 (45.2)</td>
<td>0 (0.0)</td>
<td>43.37 (18.88)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0 (0.0)</td>
<td>0 (0.0)</td>
<td>267 (100.0)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The combined percentage does not add to 100 because mild-to-moderate grade IMC were excluded from this grouping. MSKCC, Memorial Sloan Kettering Cancer Center; PM, Princess Margaret Cancer Centre; UCSF, University of California San Francisco; VUMC, Vanderbilt University Medical Center; SD: Standard Deviation; IMC: Immune-Mediated Colitis
Table 2: Polygenic risk score (PRS) of ulcerative colitis (UC) and Crohn’s disease (CD) as a predictor of time to development of all-grade and severe immune-mediated colitis (IMC) in the entire GeRI cohort, using Cox proportional hazards models and stratified analysis assessing the association between PRS\textsubscript{UC} and all-grade/severe IMC by type of therapy and lung cancer histology

<table>
<thead>
<tr>
<th>PRS\textsuperscript{a}</th>
<th>All-grade IMC</th>
<th>Severe IMC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR per SD</td>
<td>95% CI</td>
</tr>
<tr>
<td>PRS\textsubscript{UC}</td>
<td>1.34</td>
<td>1.02-1.76</td>
</tr>
<tr>
<td>PRS\textsubscript{CD}</td>
<td>0.97</td>
<td>0.72-1.32</td>
</tr>
</tbody>
</table>

**Stratified analysis restricted to PRS\textsubscript{UC} and All-grade and Severe IMC**

<table>
<thead>
<tr>
<th>Therapy\textsuperscript{b}</th>
<th>All-grade IMC</th>
<th>Severe IMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anti-PD1/Anti-PD-L1 monotherapy</td>
<td>1.33</td>
<td>0.99-1.78</td>
</tr>
<tr>
<td>Anti-PD1/Anti-PD-L1 + Anti-CTLA4</td>
<td>1.64</td>
<td>0.67-4.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Histology\textsuperscript{c}</th>
<th>All-grade IMC</th>
<th>Severe IMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma</td>
<td>1.43</td>
<td>1.06-1.93</td>
</tr>
<tr>
<td>Squamous cell carcinoma</td>
<td>0.79</td>
<td>0.16-3.78</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Models are adjusted for age at diagnosis, sex, histology, type of therapy, recruiting site, and 5 principal components. \textsuperscript{b} Models are adjusted for age at diagnosis, sex, histology, recruiting site, and 5 principal components. \textsuperscript{c} Models are adjusted for age at diagnosis, sex, type of therapy, recruiting site, and 5 principal components. HR: Hazard ratio, SD: Standard Deviation, CI: Confidence interval, IMC: Immune-mediated colitis, PRS: Polygenic risk score
Table 3: All-grade and severe Immune-Mediated Colitis (IMC) as predictors of Progression-free survival (PFS) and Overall Survival (OS) in the entire GeRi cohort, using Cox proportional models (all and with 90-day landmark)

<table>
<thead>
<tr>
<th>All-grade IMC</th>
<th>PFS</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models</td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Cox models</td>
<td>0.54</td>
<td>0.38-0.77</td>
</tr>
<tr>
<td>Cox models with 90-day landmark</td>
<td>0.80</td>
<td>0.55-1.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Severe IMC</th>
<th>PFS</th>
<th>OS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Models</td>
<td>HR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Cox models</td>
<td>0.43</td>
<td>0.25-0.72</td>
</tr>
<tr>
<td>Cox models with 90-day landmark</td>
<td>0.61</td>
<td>0.34-1.09</td>
</tr>
</tbody>
</table>

All models are adjusted for age at diagnosis, sex, histology, type of therapy, recruiting site, and 9 ancestry-informative principal components. HR: Hazards ratio, SD: Standard Deviation, CI: Confidence interval, irAEs: Immune-related adverse events, PFS: Progression free survival, OS: Overall survival.