

Peer Review File

MHC class II genotypes are independent predictors of anti-PD1 immunotherapy response in melanoma

This manuscript has been previously reviewed at another Nature Portfolio journal. This document only contains reviewer comments and rebuttal letters for versions considered at *Communications Medicine*.



Open Access This file is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. In the cases where the authors are anonymous, such as is the case for the reports of anonymous peer reviewers, author attribution should be to 'Anonymous Referee' followed by a clear attribution to the source work. The images or other third party material in this file are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

In "MHC class I and class II genotypes are independent predictors 1 of anti-PD1 2 immunotherapy response in melanoma," Claeys and Van den Eynden describe the utility of MHC genotypes as predictive biomarkers for treatment response to immune checkpoint blockade (ICB) therapies. I think the authors do a good job of emphasizing the benefit of a TMB-independent biomarker, which could be particularly useful, with a unique way to evaluate the capability of antigen presentation; however, I do think the conclusions are strongly worded, with some disregard for other relevant papers in the field, and the method is misrepresented in the abstract.

One of my major concerns is the statements regarding the novelty of the approach and the references utilized in the abstract and introduction. Phrases like "tumor mutation burden is currently the best available biomarker" (in the abstract and reiterated in the second paragraph of the main text) and that the cohort of 144 melanoma patients treated with nivolumab is the "largest ICB dataset with available genomic and clinical follow-up data." TMB has been demonstrated to be an inconsistent biomarker, and there have been many published melanoma IO datasets published. More of these could be leveraged, as well as other cancer types, to validate the approach. There have been a number of meta-analyses and individual studies that should be better cited throughout the text.

The method emphasizes the role of TMB, with its interpretation that it leads to "high expected neoantigen load," is inferior to the HLA genotype. I think the authors do a good job of highlighting the challenges related to statistical power in identifying significant associations with response. I think the utility of only MHC genotype is an interesting idea for a tumor-agnostic biomarker; however, I struggle with it for the same reasons I struggle with TMB. We consistently see that it can come down to a handful of neoantigens identified to contribute to a specific antitumor response, and while predicted neoantigen burden is correlative with TMB, the validated amount is not. Second, it has been shown that immunodominant antigens can also lead to antigen presentation loss. The authors did not address this very significant event that can occur to mediate tumor-immune evolution (noting that melanoma would be one of the poorer cancer types to consider this, given the low rates of LOH in HLA machinery). That being said, I don't think the authors clearly delineate their approach in the abstract (the method implies that the genotype itself is used as a biomarker in the study), since it is the *in silico* inferred antigenicity that is used, not the genotype itself.

The authors find that MGBS-II score was associated with worse ICB, citing that "tumoral MHC-II expression has been associated with ICB responses." The *in silico* prediction of how likely an MHC-II genotype is to express antigens (my understanding of the MGBS scoring system) may be completely different than the class II expression on the tumor cells, and may be associated with very different stages of immune-mediated elimination of the tumor cells. Then, given divergence in association with survival by MGBS-I and MGBS-II, it only makes sense that the difference between these features augments that association.

Note that the online methods do not mention how TMB was calculated and which samples MHC genotyping was performed on (I'm assuming the normal PBMC WES).

Reviewer #2 (Remarks to the Author):

The authors studied the role of MHC class I and II genotypes as tumor-agnostic and independent predictors of IO response in melanoma. They integrate this through MGBS (MHC-I and -II Genotype Binding Scores) scores derived through *in silico* peptide-binding affinity in a cohort of 144 genomically-characterized patients with melanoma treated with nivolumab. The MGBS-II score showed significance when dichotomized into stronger vs. weaker binding and associated with overall survival across the study cohort, which was sustained when combined with MGBS-I into the MGBS-d metric. The MGBS-II score was also more associated with improved survival in

ipilimumab-pretreated patients. TMB and low MGBS-II were associated with interferon-gamma response gene expression signatures.

Overall, the study attempted to simplify immunotherapy response prediction to HLA genotypes, independent of tumor features (e.g., TMB, neoantigens), but has limitations in review of context, novelty, and impact.

Major comments:

1. Context: The study is a re-analysis of data from Liu et al [15] as noted by the authors. This prior study, where TMB and MHC expression, and MHC-II HLA score specifically in ipilimumab-pretreated patients, were predictive of outcomes with the overall finding that HLA genotype influenced immunotherapy response. Although this study attempts to generalize these findings to a scoring metric based on MHC class I and II genotype alone, while also finding a strong association with MHC class II genotype similar to the parent study in ipilimumab-pretreated patients, these findings need to be better contextualized to those found in the prior study. Also, there have been associations in melanoma regarding HLA heterozygosity and response prediction (Anagnostou et al, Cell Rep Med, 2020), which are not contextualized here. Across other tumor types, HLA genotypes and TMB were prior associated with outcomes to immunotherapy (e.g., Goodman et al, Genomic Med, 2020); these and likely others should be referenced for context.
2. MHC binding prediction: The in silico prediction of peptide-binding affinities used random 9-mer and 15-mers, although MHC class I and II molecules bind different sizes. I anticipate there are potential benefits and limitations to the size-restricted approach, but it remains unclear if this is an ideal strategy for prediction. Also, the use of randomly-generated peptides – versus some derivation of known/expressed proteins – is straightforward, but it is also unclear if this is an ideal strategy. There is limited discussion as to why this overall strategy was taken.
3. MGBS-I, MGBS-II, and MGBS-d metrics: The opposing associations with overall survival for the dichotomized higher MGBS-I and lower MGBS-II was statistically significant for the only latter, which held for the integrative MGBS-d. For Figure 2, was MGBS-d differentially associated with ipilimumab-pretreated patients or only MGBS-II? Because MGBS-d had associations with some of the gene expression analysis in ipilimumab-pretreated patients, it is unclear why these comparisons were not presented.
4. Validation: The study is limited by lack of external validation in a second cohort, even if limited by available data for melanoma as the authors note. However, the referenced work from Chowell et al, Science, seems to include a potential additional validation set as well as other studies from Rizvi [8] and Anagnostou [noted prior] if possible. Given the established role of MHC and immunotherapy response, this would clarify the generalizability of an HLA genotype-only biomarker given prospective validation remains to be seen.
5. Discussion of MHC genotype: Although agreeably a “static variable” and therefore simple to assess in the host, there is external evidence that differences in MHC expression – including tumor-specific, including relevance of other markers like B2M – are relevant for immunotherapy outcome. Given the prevalence of somatic sequencing of tumors en masse for clinical and research purposes, and ability to derive HLA genotypes therewithin, I feel the generalizability of this strategy might be better discussed given host HLA haplotyping is likely less common.

Minor comments:

1. Although TMB is a relevant biomarker, it is complicated to call it the “best available biomarker” across cancer types. The clinically-used biomarker of PD-L1 – even with its own limitations – is used in lung, esophageal, and other malignancies. Although TMB is used disease agnostically, it remains as potentially limited as PD-L1 status.
2. On lines 93-94, the low and high risk groups are defined parenthetically as the same, which should be clarified.

Point-by-point response

Manuscript COMMSMED-23-0552-T

“MHC class II genotypes are independent predictors of anti-PD1 immunotherapy response in melanoma”

Dear editor, dear reviewers,

Below you can find our point-by-point response to the reviewers' comments on our manuscript, now entitled “*MHC class II genotypes are independent predictors of anti-PD1 immunotherapy response in melanoma*”.

We are most grateful to the reviewers for providing appropriate and constructive comments on our manuscript and we sincerely thank everyone for their time and effort. The points have been addressed to the best of our abilities, and we believe that this has led to the study being considerably improved.

Main changes/additions:

- We have now **contextualized** our study in light of prior work and added several analyses to explicitly compare findings with other studies.
- Interestingly, a **new multivariate logistic regression approach** points to an independent role of our MGBS-II metric and MHC-II expression in predicting anti-PD1 clinical responses (Fig. 3 and Fig. S6).
- We validated our MGBS-II results on **2 independent anti-PD1 treated melanoma datasets** (Fig. S4).
- As we found the strongest and most specific signals for MHC-II, both using the original data (including the newly added logistic regression approach) and the validation data sets, we adapted our title (stronger focus on MHC-II related findings, which is also the main novelty of our study).
- The technical approach is now explained in more detail in Methodology. We have also deposited the source code on GitHub https://github.com/CCGGLab/mhc_immunotherapy.

We checked the submission checklist, reformatted to *Communications Medicine* guidelines and completed the reporting summary and editorial policy, as requested.

For clarity, reviewer's remarks are shown below in black while our responses are in blue.

Reviewer #1

In “MHC class I and class II genotypes are independent predictors 1 of anti-PD1 2 immunotherapy response in melanoma,” Claeys and Van den Eynden describe the utility of MHC genotypes as predictive biomarkers for treatment response to immune checkpoint blockade (ICB) therapies. I think the authors do a good job of emphasizing the benefit of a TMB-independent biomarker, which could be particularly useful, with a unique way to evaluate the capability of antigen presentation; however, I do think the conclusions are strongly worded, with some disregard for other relevant papers in the field, and the method is misrepresented in the abstract.

We thank the reviewer for the positive comments. We have now better positioned our work and results in light of previous studies and provided additional clarifications in the abstract to avoid wrong interpretation.

Positioning:

- Lines 65-57: *While some associations have been described between more frequent MHC-I allele groups (e.g., B44 and B62) and ICB responses¹⁶, statistical power is lacking for the majority of (infrequent) alleles.*
- New paragraph on lines 239-259 describes a newly added multivariate logistic regression approach (Fig. 3 and Fig. S6-S7) and explicitly compares the findings with other studies.
- Analyses of data from 5 independent studies were added (Fig. S4) and described in lines 205-214.

Abstract now contains an explicit methods section following journal recommendations:

Lines 17-20: ... We developed an overall MHC binding capacity score, solely based on the germline MHC class I (MHC-I) and class II (MHC-II) genotypes. These scores were then correlated to survival and clinical responses following anti-PD1 immunotherapy in a previously published dataset of 144 melanoma patients. ...

One of my major concerns is the statements regarding the novelty of the approach and the references utilized in the abstract and introduction. Phrases like “tumor mutation burden is currently the best available biomarker” (in the abstract and reiterated in the second paragraph of the main text) and that the cohort of 144 melanoma patients treated with nivolumab is the “largest ICB dataset with available genomic and clinical follow-up data.” TMB has been demonstrated to be an inconsistent biomarker, and there have been many published melanoma IO datasets published.

We thank the reviewer for raising this point. We fully agree that TMB is an inconsistent biomarker and believe that our findings shed new light on putative covariables that explain the reported inconsistencies. A high TMB will result in an elevated MHC-I neoantigen burden as well as an elevated MHC-II neoantigen burden, which, according to our findings would have an opposite result on ICB treatment responses. The latter effect would be masked by methods

focusing only on the neoantigens. To make this clear to the reader, we added an additional analysis on MHC-I and MHC-II neoantigen burdens (Fig. S5) and discuss this in a newly added paragraph.

Lines 215-226: As an alternative strategy to MGBS, we also explored the predictive value of absolute MHC-I and MHC-II neoantigen burden (NeoAgB). Both metrics showed a strong positive association with survival ($P = 3.9e-3$ and $P = 1.3e-3$ for MHC-I and MHC-II respectively), in line with the findings from previous studies²⁹, but opposite to what we observed using MGBS-II (Supplementary Fig. 5). The strong correlation with TMB (Supplementary Fig. 3) suggests that these associations are secondary to TMB, resulting in a masking of the negative MHC-II association. Indeed, after normalizing NeoAgB to TMB (referred to as the normalized NeoAgB), similar but weaker, non-significant associations with survival were observed as for MGBS (Supplementary Fig. 5). We speculate that the reason for these weaker signals is related to the relatively few mutations (median 250) that are used to calculate NeoAgB on the one hand (i.e., relatively high statistical uncertainty) and the tumor evolutionary selection processes that could have shaped the mutational landscape on the other hand.

The TMB and PD-L1 expression are currently the only FDA-approved biomarkers for ICB responses in solid cancers (apart from MMR deficiency for some specific cancers, not including melanoma). We added some relevant citations, rephrased the sentences and have also added analysis on PD-L1 expression. Interestingly, PD-L1 expression was identified as the only independent variable to our MHC-II genotype score (MGBS-II) in a Cox multivariate regression analysis of ipilimumab pretreated patients (Fig. S7).

Lines 54-56: Despite these successes, ICB responses are hard to predict, and the TMB is currently one of the few FDA-approved biomarkers in different cancer types (the other being PD-L1 expression and microsatellite instability)⁸⁻¹⁵

The main novelty of our study lies in the role of the MHC-II genotype in (negatively) predicting ICB responses. While previous studies (which we now cite more explicitly) identified MHC-II gene expression as a putative predictor, this has not been shown on the genotype level before. In our revised manuscript we stressed this more explicit in title, abstract and main text.

Lines 262-265: While MHC-II expression has been suggested as a biomarker for anti-PD1 outcome previously^{19,29,30}, our study is, to our knowledge, the first to demonstrate a role of the MHC-II genotype-determined HLA binding potential..

More of these could be leveraged, as well as other cancer types, to validate the approach. There have been a number of meta-analyses and individual studies that should be better cited throughout the text.

We have added analyses of 2 independent anti-PD1 treated melanoma studies (*Hugo et al, 2016* and *Gide et al, 2019*) to validate our results. Interestingly, while these studies confirm our

MHC-II findings (both when evaluating responses using survival data and using RECIST criteria; Fig. S4), MHC-I findings are more variable. We have also added analysis of 2 anti-CTLA4 treated melanoma studies, where the MHC-II effect was absent, suggesting it is anti-PD1 specific. Interestingly, a final analysis on anti-PD1 treated lung cancer patients (Rizvi et al., 2015) showed similar trends.

Lines 205-214: We validated our findings on 2 independent anti-PD1 treated melanoma datasets with available clinical and molecular data, one from Hugo et al.²¹ (n=37), the other from Gide et al.²² (n=36). Despite the relatively small sample sizes, both studies confirmed the association between lower-than-median MGBS-II and better survival, although this was only significant for the Hugo study ($P = 0.014$) and not the Gide study ($P = 0.10$; Supplementary Fig. 4). Interestingly, RECIST clinical responses were significantly worse ($P = 0.0076$, Chi-square test) for patients with higher MGBS scores for the latter study. While similar, though non-significant patterns were observed in anti-PD1 lung cancer patients⁸, no correlation was observed between outcome and MGBS in 2 studies of anti-CTLA4 treated patients^{9,10} (Supplementary Fig. 4).

The method emphasizes the role of TMB, with its interpretation that it leads to “high expected neoantigen load,” is inferior to the HLA genotype. I think the authors do a good job of highlighting the challenges related to statistical power in identifying significant associations with response. I think the utility of only MHC genotype is an interesting idea for a tumor-agnostic biomarker; however, I struggle with it for the same reasons I struggle with TMB. We consistently see that it can come down to a handful of neoantigens identified to contribute to a specific antitumor response, and while predicted neoantigen burden is correlative with TMB, the validated amount is not.

We fully agree that with the low reported accuracy of currently available neoantigen prediction methods, it is nearly impossible to predict neoantigenicity of individual mutations. In this regard, we believe that neoantigen burdens are nothing more than a surrogate biomarker for TMB. We added 2 alternative approaches to demonstrate this: 1 focusing on the absolute neoantigen burden (neoAgB), the other focusing on the TMB-normalized neoAgB (Fig. S5). As expected, TMB is highly correlated to neoAgB (both for MHC-I and MHC-II; Fig. S3) and the absolute neoAgB-outcome associations are almost identical to TMB. When we then normalize neoAgB to the TMB, an approach that is in essence similar to our main MGBS approach, we do see a weak signal of higher normalized MHC-II neoAgB being associated with worse survival. We believe that the reason this signal is weaker is the high statistical uncertainty (based on relatively few mutations/peptides as compared to 10^6 peptides for MGBS) and prior selection processes that could have influenced the mutational landscape. As indicated higher, we elaborated on this in a newly added paragraph (lines 215-226).

Second, it has been shown that immunodominant antigens can also lead to antigen presentation loss. The authors did not address this very significant event that can occur to mediate tumor-immune evolution (noting that melanoma would be one of the poorer cancer types to consider this, given the low rates of LOH in HLA machinery). That being said, I don't

think the authors clearly delineate their approach in the abstract (the method implies that the genotype itself is used as a biomarker in the study), since it is the in silico inferred antigenicity that is used, not the genotype itself.

As part of the multivariate Cox (Fig. S7) and logistic regression analysis (Fig. 3 and Fig. S6) we added on previously reported biomarkers, we also evaluated LOH of the HLA region and the *B2M* gene. Interestingly, *B2M* loss is one of the only 4 variables that contributed significantly to our multivariate logistic regression model in ipilimumab-pretreated patients.

*Lines 246-250: Strikingly, responses of the pretreated patient group were predicted highly accurately (mean AUC = 0.89, calculated from 100 iterations of a 5-fold cross-validation approach) by a combination of high MHC II expression ($P = 0.0092$), low MGBS-II scores ($P = 0.014$), *B2M* LOH ($P = 0.047$) and tumor heterogeneity ($P = 0.028$; Fig. 3a-b and Supplementary Fig. 6).*

We restructured and rephrases the abstract to clarify that we indeed focused on the antigenicity, as derived solely from the MHC genotype.

Lines 17-20: ... We developed an overall MHC binding capacity score, solely based on the germline MHC class I (MHC-I) and class II (MHC-II) genotypes. These scores were then correlated to survival and clinical responses following anti-PD1 immunotherapy in a previously published dataset of 144 melanoma patients. ...

The authors find that MGBS-II score was associated with worse ICB, citing that “tumoral MHC-II expression has been associated with ICB responses.” The in silico prediction of how likely an MHC-II genotype is to express antigens (my understanding of the MGBS scoring system) may be completely different than the class II expression on the tumor cells, and may be associated with very different stages of immune-mediated elimination of the tumor cells. Then, given divergence in association with survival by MGBS-I and MGBS-II, it only makes sense that the difference between these features augments that association.

We agree with the reviewer’s interpretation: our score (MGBS-II) indeed quantifies how likely it is that an MHC-II genotype will result in presentable neoantigens, independent of MHC-II expression itself. As indicated higher, we have now added several new uni- and multivariate analyses (Fig. 3 and Fig. S6-S7) to distinguish MGBS-II and MHC-II expression for their predictive capacity. Interestingly, our multivariate logistic regression analysis suggests that MHC-II expression and MGBS-II are independent, complementary predictors in the former subgroup. These findings are described in a newly added paragraph:

Lines 239-259: Our findings shed new light on previous studies suggesting an influence of prior anti-CTLA4 treatment on the predictive features of anti-PD1 responses^{19,20}. In this regard, Liu et al. developed a logistic regression model that predicted clinical responses in patients that received prior anti-CTLA4 therapy based on MHC-II expression, serum LDH levels and lymph node metastases¹⁹. As MHC genotypes were

not directly considered in the Liu study, we redeveloped this regression model, using an unbiased feed forward-selection approach (see Methods) that included several previously described biomarkers (e.g., PD-L1 expression, B2M LOH, MHC zygosity) and the 3 MGBS metrics (Fig. 3 and Supplementary Fig. 6). Strikingly, responses of the pretreated patient group were predicted highly accurately (mean AUC = 0.89, calculated from 100 repetitions of a 5-fold cross-validation approach) by a combination of high MHC II expression ($P = 0.0092$), low MGBS-II scores ($P = 0.014$), B2M LOH ($P = 0.047$) and tumor heterogeneity ($P = 0.028$; Fig. 3a-b and Supplementary Fig. 6). Notably, MGBS-II had the highest predictive value of all variables in a univariate setting (mean AUC = 0.78) and exclusion of MGBS-II from the multivariate model reduced the mean AUC from 0.89 to 0.81, similar to what was reported by Liu et al. when MHC genotypes were not considered¹⁹ (Fig. 3b). When patients were subsequently stratified using model predictions, large survival differences were observed between predicted responders and non-responders in the ipilimumab pretreated subgroup ($P < 1.0e-4$; Fig. 3c). Using a similar feed-forward multivariate Cox regression approach on the subgroup of ipilimumab pretreated patients, survival was independently associated with PD-L1 expression ($HR = 0.38$, $P = 0.031$) and MGBS-II ($HR = 3.4$, $P = 0.0071$; Supplementary Fig. 7).

Note that the online methods do not mention how TMB was calculated and which samples MHC genotyping was performed on (I'm assuming the normal PBMC WES).

We thank the reviewer for noting this. Genotypes were derived from normal, non-tumoral DNA sequencing data indeed. We added this information to methodology.

Reviewer #2

The authors studied the role of MHC class I and II genotypes as tumor-agnostic and independent predictors of IO response in melanoma. They integrate this through MGBS (MHC-I and -II Genotype Binding Scores) scores derived through in silico peptide-binding affinity in a cohort of 144 genomically-characterized patients with melanoma treated with nivolumab. The MGBS-II score showed significance when dichotomized into stronger vs. weaker binding and associated with overall survival across the study cohort, which was sustained when combined with MGBS-I into the MGBS-d metric. The MGBS-II score was also more associated with improved survival in ipilimumab-pretreated patients. TMB and low MGBS-II were associated with interferon-gamma response gene expression signatures.

We thank the reviewer for this accurate summary.

Overall, the study attempted to simplify immunotherapy response prediction to HLA genotypes, independent of tumor features (e.g., TMB, neoantigens), but has limitations in review of context, novelty, and impact.

Major comments:

1. Context: The study is a re-analysis of data from Liu et al [15] as noted by the authors. This prior study, where TMB and MHC expression, and MHC-II HLA score specifically in ipilimumab-pretreated patients, were predictive of outcomes with the overall finding that HLA genotype influenced immunotherapy response. Although this study attempts to generalize these findings to a scoring metric based on MHC class I and II genotype alone, while also finding a strong association with MHC class II genotype similar to the parent study in ipilimumab-pretreated patients, ...

Our study is indeed primarily based on the data from *Liu et al.* However, our focus on MHC genotypes makes our study highly complementary, so we strongly disagree it is solely a reanalysis. **The Liu study demonstrated a role of MHC II expression, but did not evaluate MHC genotypes** as such. Our score (MGBS-II) quantifies how likely it is that any MHC-II genotype will result in presentable neoantigens, independent of MHC-II expression itself

We elaborated on the differences between our study and the Liu study in a newly added paragraph. Interestingly, when combining all putative predictors in a multiple logistic regression model (Fig. 3 and Fig. S6), we demonstrate that both MHC-II expression and MGBS-II independently predict clinical response to ICB in ipilimumab pretreated patients. This underscores the potential of MGBS-II as a putative biomarker and sheds new light on the role of MHC II in modulating ICB responses.

Lines 239-265: Our findings shed new light on previous studies suggesting an influence of prior anti-CTLA4 treatment on the predictive features of anti-PD1 responses^{19,20}. In this regard, Liu et al. developed a logistic regression model that predicted clinical responses

in patients that received prior anti-CTLA4 therapy based on MHC-II expression, serum LDH levels and lymph node metastases¹⁹. As MHC genotypes were not directly considered in the Liu study, we redeveloped this regression model, using an unbiased feed forward-selection approach (see Methods) that included several previously described biomarkers (e.g., PD-L1 expression, B2M LOH, MHC zygosity) and the 3 MGBS metrics (Fig. 3 and Supplementary Fig. 6). Strikingly, responses of the pretreated patient group were predicted highly accurately (mean AUC = 0.89, calculated from 100 repetitions of a 5-fold cross-validation approach) by a combination of high MHC II expression ($P = 0.0092$), low MGBS-II scores ($P = 0.014$), B2M LOH ($P = 0.047$) and tumor heterogeneity ($P = 0.028$; Fig. 3a-b and Supplementary Fig. 6). Notably, MGBS-II had the highest predictive value of all variables in a univariate setting (mean AUC = 0.78) and exclusion of MGBS-II from the multivariate model reduced the mean AUC from 0.89 to 0.81, similar to what was reported by Liu et al. when MHC genotypes were not considered¹⁹ (Fig. 3b). When patients were subsequently stratified using model predictions, large survival differences were observed between predicted responders and non-responders in the ipilimumab pretreated subgroup ($P < 1.0e-4$; Fig. 3c). Using a similar feed-forward multivariate Cox regression approach on the subgroup of ipilimumab pretreated patients, survival was independently associated with PD-L1 expression (HR = 0.38, $P = 0.031$) and MGBS-II (HR = 3.4, $P = 0.0071$; Supplementary Fig. 7).

The independent association between MHC-II expression and the MHC-II genotype (quantified using MGBS-II) with patient responses suggests an important role of the MHC-II complex in modulating responses to anti-PD1 immunotherapy. While MHC-II expression has been suggested as a biomarker for anti-PD1 outcome previously^{19,29,30}, our study is, to our knowledge, the first to demonstrate a role of the MHC-II genotype-determined HLA binding potential.

We noted there could be some semantic confusion regarding “scores”. While our MGBS-II score is solely based on the **MHC genotype**, Liu refers to the **MHC II expression**-based logistic regression score when referring to “MHC-II HLA score”.

these findings need to be better contextualized to those found in the prior study. Also, there have been associations in melanoma regarding HLA heterozygosity and response prediction (Anagnostou et al, Cell Rep Med, 2020), which are not contextualized here. Across other tumor types, HLA genotypes and TMB were prior associated with outcomes to immunotherapy (e.g., Goodman et al, Genomic Med, 2020); these and likely others should be referenced for context.

We thank the reviewer for bringing these studies to our attention.

One of the main findings of the **Anagnostou study** is the correlation between MHC-II expression and ICB response. We refer to our previous answer and newly added paragraph on this topic (note that citation 26 refers to the Anagnostou study). We have also added MHC-I and MHC-II heterozygosity to our multivariate logistic (Fig. S6) and Cox (Fig. S7) regression approaches but did not find any correlation to outcome.

Notably, the Anagnostou study also demonstrates that MHC-I and MHC-II neoantigen loads/burdens (NeoAgB) are higher in responding tumors than non-responding tumors, and we strongly believe this is related to the TMB being the main factor determining the overall NeoAgB. We added several new analyses to demonstrate a strong correlation between TMB and the NeoAgB (Fig. S3) and subsequently show how the MHC-II NeoAgB is positively associated with ICB outcome (Fig. S5), opposite to what we see using MGBS-II scores. This TMB correlation can be removed by normalizing NeoAgB to TMB, resulting in similar outcome associations as with the MGBS approach. We elaborated on the differences between the (TMB-normalized) neoantigen burdens and MGBS in a newly added paragraph.

Lines 215-226: As an alternative strategy to MGBS, we also explored the predictive value of absolute MHC-I and MHC-II neoantigen burden (NeoAgB). Both metrics showed a strong positive association with survival ($P = 3.9e-3$ and $P = 1.3e-3$ for MHC-I and MHC-II respectively), in line with the findings from previous studies²⁹, but opposite to what we observed using MGBS-II (Supplementary Fig. 5). The strong correlation with TMB (Supplementary Fig. 3) suggests that these associations are secondary to TMB, resulting in a masking of the negative MHC-II association. Indeed, after normalizing NeoAgB to TMB (referred to as the normalized NeoAgB), similar but weaker, non-significant associations with survival were observed as for MGBS (Supplementary Fig. 5). We speculate that the reason for these weaker signals is related to the relatively few mutations (median 250) that are used to calculate NeoAgB on the one hand (i.e., relatively high statistical uncertainty) and the tumor evolutionary selection processes that could have shaped the mutational landscape on the other hand.

The **Goodman study** focused on MHC-I presentability of a set of cancer driver mutations, using a metric called PHBR. Similar to our study, they conclude that poor MHC-I binding genotypes result in failure to respond to ICB. The main differences with our study is the focus on the overall MHC binding properties (not just restricted to peptides translated from driver mutations), and, most importantly, the focus on the MHC-II genotype and the finding that it oppositely affects ICB outcomes (the Goodman study was restricted to the MHC-I genotype). We contextualized our MGBS-I findings in light of the Goodman study:

Lines 184-186: These findings are in agreement with a previously suggested association between poor ICB therapy outcome and weaker MHC-I presentability of peptides translated from a set of cancer driver mutations²⁸.

2. MHC binding prediction: The in silico prediction of peptide-binding affinities used random 9-mer and 15-mers, although MHC class I and II molecules bind different sizes. I anticipate there are potential benefits and limitations to the size-restricted approach, but it remains unclear if this is an ideal strategy for prediction. Also, the use of randomly-generated peptides – versus some derivation of known/expressed proteins – is straightforward, but it is also unclear if this is an ideal strategy. There is limited discussion as to why this overall strategy was taken.

The most obvious alternative approach would be to derive the MHC-I and MHC-II neoAgB directly from the genomic data (i.e., mutated peptides). In reference to our previous answer, we added a new paragraph to discuss differences between the approaches in lines 215-226.

The reviewer raises a valid point that MHC-I/II can bind peptides of different lengths. Our peptide length choices are the default values in netMHCpan, the algorithm that was used to calculate MHC binding affinities (see *Methods*). Additionally, for MHC-I, 9-mers are also most immunogenic (i.e., the majority of known immunogenic peptides in IEDB are 9-mers; see also *Jurtz et al. 2017* for discussion). Our aim is to determine the overall germline-specific MHC binding capacity and we assume this will not be different for other peptide lengths. To demonstrate that this is a valid assumption, we calculated MGBS scores for peptides varying in length from 8-mers to 11-mers for MGBS-I and 13-mers to 17-mers for MGBS-II. As expected, this did not affect the results from our analysis (see fig. R1, below). Relatedly, also deriving MGBS scores from real peptides (data derived from EnsemblDB) did not affect our conclusions (see fig. R1, below)

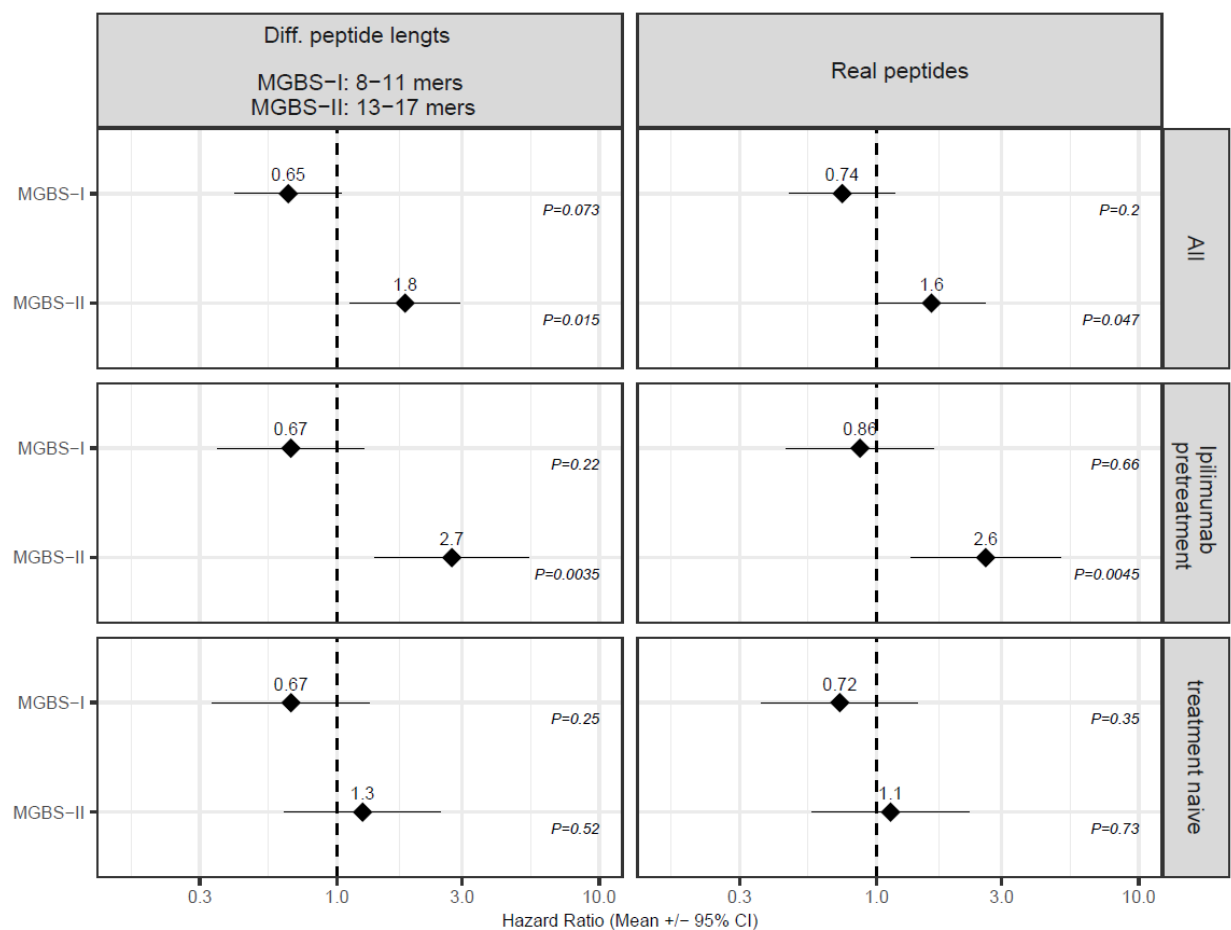


Figure R1 Forest plots indicating hazard ratio +/- 95% confidence intervals as determined from a univariate Cox regression analysis. MGBS-I and MGBS-II was calculated from 10,000 random peptides with length 8-11 (MGBS-I) and 13-17 (MGBS-II) or from 10,000 real peptides, randomly derived from EnsemblDB.

3. MGBS-I, MGBS-II, and MGBS-d metrics: The opposing associations with overall survival for the dichotomized higher MGBS-I and lower MGBS-II was statistically significant for the only latter, which held for the integrative MGBS-d.

We agree the MGBS-I signal is borderline significant ($P = 0.055$) and rephrased.

While we did observe some significant associations between MGBS-I and outcome (e.g., fig. 1d P75 analysis, fig. 1h and fig. S6 treatment-naïve), these associations were generally weaker than for MGBS-II and could also not be confirmed in independent validation datasets (Fig. S4). Together with the new evidence we found for MGBS-II (i.e., highest predictive value in a logistic regression model, fig. 3 and fig. S6; significantly lower scores in responding patients, fig. 1i), we decided to rephrase the manuscript's title and abstract more towards MHC II.

For Figure 2, was MGBS-d differentially associated with ipilimumab-pretreated patients or only MGBS-II? Because MGBS-d had associations with some of the gene expression analysis in ipilimumab-pretreated patients, it is unclear why these comparisons were not presented.

We thank the reviewer for noting that MGBS-d was missing in figure 2 and have now added this analysis, which confirms a significant survival difference in both ipilimumab pretreated ($P = 0.0026$) and treatment naïve groups ($P = 0.027$).

4. Validation: The study is limited by lack of external validation in a second cohort, even if limited by available data for melanoma as the authors note. However, the referenced work from Chowell et al, Science, seems to include a potential additional validation set as well as other studies from Rizvi [8] and Anagnostou [noted prior] if possible. Given the established role of MHC and immunotherapy response, this would clarify the generalizability of an HLA genotype-only biomarker given prospective validation remains to be seen.

We have now extended our analysis to 5 other, independent datasets (Fig. S4): 2 anti-PD1 treated melanoma datasets (*Hugo et al, 2016* and *Gide et al., 2019*), 1 anti-PD1 treated lung cancer dataset (*Rizvi et al., 2015*) and 2 anti-CTLA4 treated melanoma datasets (*Snyder et al, 2014* and *Van Allen et al., 2015*). These data largely confirmed our MGBS-II findings in anti-PD1 treated melanoma patients (significant for survival in *Hugo et al., P = 0.014*; and RECIST in *Gide et al., P = 0.0076*). This is remarkable given the relatively small size of these studies and hence limited statistical power.

Interestingly, we also noted a similar, though non-significant, survival and RECIST response in the *Rizvi et al.* anti-PD1 treated lung cancer data. The MGBS-outcome associations were not present in the anti-CTLA4 treated melanoma patients.

The study of *Chowell et al.*, is a meta-analysis of different ICB datasets (different cancer types and treatments), several of which were included in our study. The main findings in the manuscript (outcome associations with HLA supertypes B44 and B62) were related to anti-CTLA4 treated melanoma patients (*Snyder* and *Van Allen*). We added reference to this work.

5. Discussion of MHC genotype: Although agreeably a “static variable” and therefore simple to assess in the host, there is external evidence that differences in MHC expression – including tumor-specific, including relevance of other markers like B2M – are relevant for immunotherapy outcome. Given the prevalence of somatic sequencing of tumors en masse for clinical and research purposes, and ability to derive HLA genotypes therewithin, I feel the generalizability of this strategy might be better discussed given host HLA haplotyping is likely less common.

The reviewer raised a highly interesting point. Several, more recent studies have indeed suggested an association between high MHC-II expression and ICB outcome. However, to our knowledge, no associations have been described yet with the MHC-II genotype.

Referring to one of our previous answers, we added an integrative analysis where we associated all previously described biomarkers for which information was available (e.g., MHC-II expression, LOH B2M) with outcome (both survival and clinical response) in Fig. 3 and Fig. S6-S7 and contextualized our findings in a newly added paragraph. Interestingly, an unbiased, feed-forward multivariate logistic regression analysis (clinical response as outcome variable) resulted in a highly accurate prediction model (mean AUC = 0.89, 5-fold cross validation) in ipilimumab-pretreated patients, consisting of 4 variables (MGBS-II, MHC-II expression, LOH B2M and tumor heterogeneity). Strikingly, MGBS-II had the highest prediction accuracy of all variables in a univariate setting (mean AUC = 0.78) and the mean AUC decreased from 0.89 to 0.81 when MGBS scores were left out of the model. Notably, these AUC values are highly similar to the ones reported by the original authors (who did not directly consider the MHC genotypes). The independent contribution of both MGBS-II and MHC-II expression suggests an important immunomodulatory role of the MHC-II complex in the subgroup of ipilimumab-pretreated patients.

Minor comments:

1. Although TMB is a relevant biomarker, it is complicated to call it the “best available biomarker” across cancer types. The clinically-used biomarker of PD-L1 – even with its own limitations – is used in lung, esophageal, and other malignancies. Although TMB is used disease agnostically, it remains as potentially limited as PD-L1 status.

We agree and rephrased, also including PD-L1 expression as an established and FDA-approved biomarker.

Lines 38-40: ... and the TMB is currently one of the few FDA-approved biomarkers in different cancer types (the other being PD-L1 expression and microsatellite instability)⁵⁴⁻

⁵⁶.

Additionally, we have added PD-L1 (CD274) expression as a variable in the multivariate analyses that were added (Fig. 3 and Fig. S6-S7). Interestingly, in ipilimumab-pretreated patients we found that MGBS-II and PD-L1 expression were the only variables that were independently associated with survival in a multivariate Cox regression analysis (Fig. S7).

2. On lines 93-94, the low and high risk groups are defined parenthetically as the same, which should be clarified.

We thank the reviewer for noting this error. In light of the other (supplementary) analyses that were added we felt the risk group analysis had little added value so we decided to remove from the current version of the manuscript.

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

I want to congratulate the authors on their attention to the feedback from both reviewers. I have no further comments.

Reviewer #2 (Remarks to the Author):

We appreciate that the authors responses to the reviewers' critiques, including several additional analyses. The updated manuscript now focuses on the independent role of MHC class II genotype binding scores (MGBS-II) with anti-PD1 response in melanoma.

Comments by specific updates:

- The inclusion of external validation of MGBS-II with two additional melanoma cohorts -- one significantly associated with outcomes, the other with tumor response (RECIST) -- provides more evidence of generalizability of this strategy. I appreciate these efforts.

- The feed forward multivariate regression analysis integrates both MHC class II genotype binding scores and expression, which clarifies the independence of genotype and phenotype: both are similarly predictive of outcomes alone, with the combination better in a cross-validation strategy. A few follow-up questions given this new model:

1. Given the model has four parameters performing best - MGBSII, MHC class II expression, LOH B2M, and tumor heterogeneity - that may be available outside of Liu et al, can any external validation be done (e.g., with Riaz et al as a source)?

2. Amongst the four components, do the two MGBSII and MHC class II expression parameters together provide as robust a prediction as the four parameters? I may be underappreciating the feed forward process, but this would isolate the effect to MHC class II alone, which is in line with the manuscripts conclusions.

3. Does this four parameter model generalize to treatment naive patients, which one might expect not?

- The additional context and references improved clarity of findings. However, the abstract uses several different ways of describing MGBS including "MHC binding capacity score", "MHC binding genotype score" (MBGS vs. MGBS?), and "MHC genotype" (alone) which could be streamlined. Although derived directly from MHC genotype, the MGBS is the metric used throughout subsequent analyses. Perhaps the abstract conclusion can simply reiterate it is calculable from MHC genotypes alone?

Overall, the manuscript effectively analyzes and describes the additional relevance of MHC class II genotype, beyond phenotype, on immunotherapy response.

Point-by-point response
Manuscript COMMSMED-23-0552B
“MHC class II genotypes are independent predictors of anti-PD1 immunotherapy response in melanoma”

Dear editor, dear reviewers,

Below you can find our point-by-point response to the reviewers' comments on our manuscript, entitled “*MHC class II genotypes are independent predictors of anti-PD1 immunotherapy response in melanoma*”.

The main addition is a **multivariate analysis on the validation data** (Suppl. Fig. 8), mainly focusing on an MHC-II 2-parameter model that included the MGBS-II metric and MHC-II expression.

To make data comparable between the different studies, we reprocessed all gene expression data using the same pipeline (see *Methods*), resulting in small numeric differences, obviously without altering our study's main conclusions.

For clarity, reviewer's remarks are shown below in black while our responses are in blue.

Reviewer #1

I want to congratulate the authors on their attention to the feedback from both reviewers. I have no further comments.

Thank you.

Reviewer #2

We appreciate that the authors responses to the reviewers' critiques, including several additional analyses. The updated manuscript now focuses on the independent role of MHC class II genotype binding scores (MGBS-II) with anti-PD1 response in melanoma.

Comments by specific updates:

- The inclusion of external validation of MGBS-II with two additional melanoma cohorts -- one significantly associated with outcomes, the other with tumor response (RECIST) -- provides more evidence of generalizability of this strategy. I appreciate these efforts.
- The feed forward multivariate regression analysis integrates both MHC class II genotype binding scores and expression, which clarifies the independence of genotype and phenotype: both are similarly predictive of outcomes alone, with the combination better in a cross-validation strategy. A few follow-up questions given this new model:

1. Given the model has four parameters performing best - MGBSII, MHC class II expression, LOH B2M, and tumor heterogeneity - that may be available outside of Liu et al, can any external validation be done (e.g., with Riaz et al as a source)? & 2. Amongst the four components, do the two MGBSII and MHC class II expression parameters together provide as robust a prediction as the four parameters? I may be underappreciating the feed forward process, but this would isolate the effect to MHC class II alone, which is in line with the manuscripts conclusions.

We thank the reviewer for these suggestions and performed the required analyses as part of a newly added **supplementary figure 8**.

Related to question 2, we agree on the added value of an analysis focusing on both MHC-II parameters (MGBS-II and MHC-II expression). We noted that this 2-parameter model is less performant than the 4-parameter model (AUC = 0.78 versus AUC = 0.87 respectively), but still amongst the better performing models evaluated, in line with our study's conclusions. This observation was added to fig. 3b. We then applied this 2-parameter logistic regression model to **independently predict survival in the anti-PD1 treated melanoma validation datasets** (Suppl. Fig. 8a). This MGBS-II and MHC-II expression-based prediction resulted in clear survival differences between predicted responders and progressors in both the Hugo and the Gide dataset.

Lines 285-292: In light of this putative MHC-II immunomodulatory role, we evaluated the predictive capacity of the combined MGBS-II and MHC-II expression values, using a logistic regression model developed on the Liu ipilimumab pretreated patient data. As expected, predicted responders had better survival than predicted progressors in the Liu data ($P = 9.7e-04$ and $P = 0.0098$ in the overall and pretreated groups, respectively). Interestingly, when this model was applied on the 3 independent anti-PD1 melanoma validation studies, significant survival differences were now observed in the Gide data ($P = 0.043$; 22/35 predicted responders). A similar trend was observed in the Hugo data ($P = 0.19$; 14/26 predicted responders) ...

Note that an extension to the 4-parameter model was not possible as the variables LOH B2M and tumor heterogeneity were not available. To clarify how each of these predictors compare between different studies on a univariate level, we added a direct comparison of all studies' available biomarkers and their association with survival in suppl. fig. 8b.

Following the reviewer's suggestion we also performed an analysis on the data from Riaz et al. (2017). Surprisingly, no outcome significant associations were found with the MGBS-II metric. However, we noted that the outcome associations with several other biomarkers were also different than what would be expected, specifically in a subgroup of patients that were reported to be progressive on ipilimumab (see suppl. fig. 4b and suppl. fig. 8b). Most striking was the negative (!) association with tumor mutation burden in this group. Additionally, no correlation was observed with MHC-II expression, contrary to other studies. These results suggest that the MHC-II modulatory signal that we and others observed in the other examined anti-PD1

melanoma studies is largely absent in the Riaz data. We speculate that this is related to differences in patient inclusion. In this regard it is striking to note significant differences in tumor heterogeneity in the Riaz pretreatment data, with higher values mainly observed in the ipilimumab progressive group (suppl. Fig. 8c).

Lines 292-306: ... while the differences between predicted responders and progressors were rather marginally in the Riaz data ($P = 0.52$; Supplementary Fig. 8a). To better understand these differences, we compared the Cox HRs of the evaluated biomarkers between all 4 studies. A remarkable difference was noted between the pretreated patients of the Liu and the Riaz data (Supplementary Fig. 4b and 8b). Indeed, while the HRs in the treatment naïve groups were strongly correlated (Pearson's $r = 0.87$, $P = 1.1e-04$), no correlation was observed in the ipilimumab pretreated group. Most remarkable, a lack of survival association was observed with MHC-II expression, PD-L1 expression and TMB, with the latter even trending towards an opposite association than what is generally expected ($HR = 2.3$, $P = 0.10$). These results suggest that the Riaz patients lack the MHC-II immunomodulation seen in the other studies. The discrepancies between the ipilimumab pretreated cohorts from both studies point towards differences in patient selection procedures. This conclusion that is further supported by a remarkable difference in tumor heterogeneity between the Liu and the Riaz study ($P = 1.1e-07$, Kruskal-Wallis rank sum test; Supplementary 8c).

3. Does this four parameter model generalize to treatment naïve patients, which one might expect not?

This is indeed not the case. It was already suggested in the initial study from *Liu et al (2019)* that the biomarkers predicting response in ipilimumab treated and treatment naïve patients are different. This was also confirmed in the recent study from *Campbell et al (2023)*, which focusses specifically on differences between both groups.

For clarity we performed the 4-parameter analysis on the treatment-naïve group (suppl. fig. 6) and found an AUC that was lower than solely heterogeneity (0.65 vs. 0.69), suggesting no/little added value of the MHC-II metrics (MGBS-II & MHC-II expression). This is also confirmed when solely focusing on MGBS-II & MHC-II expression in this group (AUC = 0.61). Overall, these results confirm that MHC-II plays an important role in ipilimumab pretreated patients and not in treatment-naïve patients.

The additional context and references improved clarity of findings. However, the abstract uses several different ways of describing MGBS including "MHC binding capacity score", "MHC binding genotype score" (MBGS vs. MGBS?), and "MHC genotype" (alone) which could be streamlined. Although derived directly from MHC genotype, the MGBS is the metric used throughout subsequent analyses. Perhaps the abstract conclusion can simply reiterate it is calculable from MHC genotypes alone?

We followed the reviewer's suggestion and rephrased the abstract.

Lines 17-19: We developed an overall MHC genotype binding capacity score (MGBS), indicative of a patient's MHC class I (MHC-I) and class II (MHC-II) neoantigen binding capacity and solely based on the germline MHC-I (MGBS-I) and MHC-II (MGBS-II) genotypes.

...

Lines 29-31: Our results suggest that MGBS, calculated solely from the MHC genotype, has clinical potential as a non-invasive and tumor-independent biomarker to guide anti-cancer immunotherapy in melanoma.

Overall, the manuscript effectively analyzes and describes the additional relevance of MHC class II genotype, beyond phenotype, on immunotherapy response.

We thank the reviewer for the time and effort.

REVIEWERS' COMMENTS:

Reviewer #2 (Remarks to the Author):

I thank the authors for their thorough responses, including additional analyses and commentaries, which enabled nuance discussion of these findings in the context of prior datasets/manuscripts (e.g., Riaz et al versus others).

I appreciate the extra work and time taken by the authors.