

Cell Reports Medicine, Volume 4

Supplemental information

**Membrane-localized neoantigens predict
the efficacy of cancer immunotherapy**

Zoe Goldberger, Sylvie Hauert, Kevin Chang, Trevin Kurtanich, Aaron T. Alpar, Grégoire Repond, Yue Wang, Suzana Gomes, Raga Krishnakumar, Prabha Siddarth, Melody A. Swartz, Jeffrey A. Hubbell, and Priscilla S. Briquez

Supplemental Information

Membrane-Localized Neoantigens Predict the Efficacy of Cancer Immunotherapy

Zoe Goldberger^{1,2,*}, Sylvie Hauert^{1,*}, Kevin Chang¹, Trevin Kurtanich¹, Aaron T. Alpar¹, Grégoire Repond¹, Yue Wang¹, Suzana Gomes¹, Raga Krishnakumar³, Prabha Siddarth⁴, Melody A. Swartz^{1,5,6,7}, Jeffrey A. Hubbell^{1,6,7+}, Priscilla S. Briquez^{1,8+}

* These authors contributed equally to this work

⁺ Correspondence: jhubbell@uchicago.edu, priscilla.briquez@uniklinik-freiburg.de

Affiliations:

¹ Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL, USA

² Department of Bioengineering, McGill University, Montreal, QC, Canada

³ Sandia National Laboratories, Livermore, CA, USA

⁴ Semel Institute for Neuroscience & Human Behavior, UCLA, Los Angeles, CA, USA

⁵ Ben May Department of Cancer Research, University of Chicago, Chicago, IL, USA

⁶ Committee on Immunology, University of Chicago, Chicago, IL, USA

⁷ Committee on Cancer Biology, University of Chicago, Chicago, IL, USA

⁸ Department of General and Visceral Surgery, Medical Center – University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

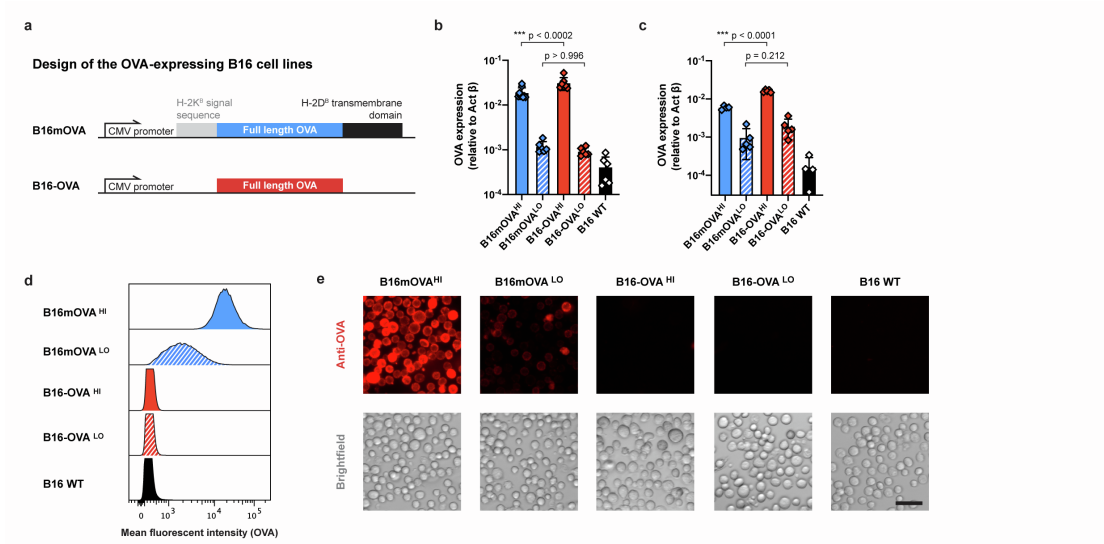


Figure S1. Characterization of OVA-expressing B16-F10 melanoma cell lines and tumors, related to Figures 1 and 2. **a**, Design of the different OVA-expressing B16-F10 cell lines, expressing membrane OVA (mOVA) or soluble OVA. **b**, OVA expression in the modified B16 cell lines in culture *in vitro*, assessed by qPCR (N≥6, mean ± SD, ANOVA with Sidak's post-test). **c**, OVA expression in the modified B16 tumors *in vivo*, assessed by qPCR (N≥4, mean ± SD, ANOVA with Sidak's post-test). **d**, Cell-surface staining of OVA quantified by flow cytometry via the mean fluorescence intensity. **e**, Detection of cell plasma membrane-bound OVA on the different OVA-expressing B16 cell lines assessed by microscopy (red: anti-OVA; scale bar = 50 μm).

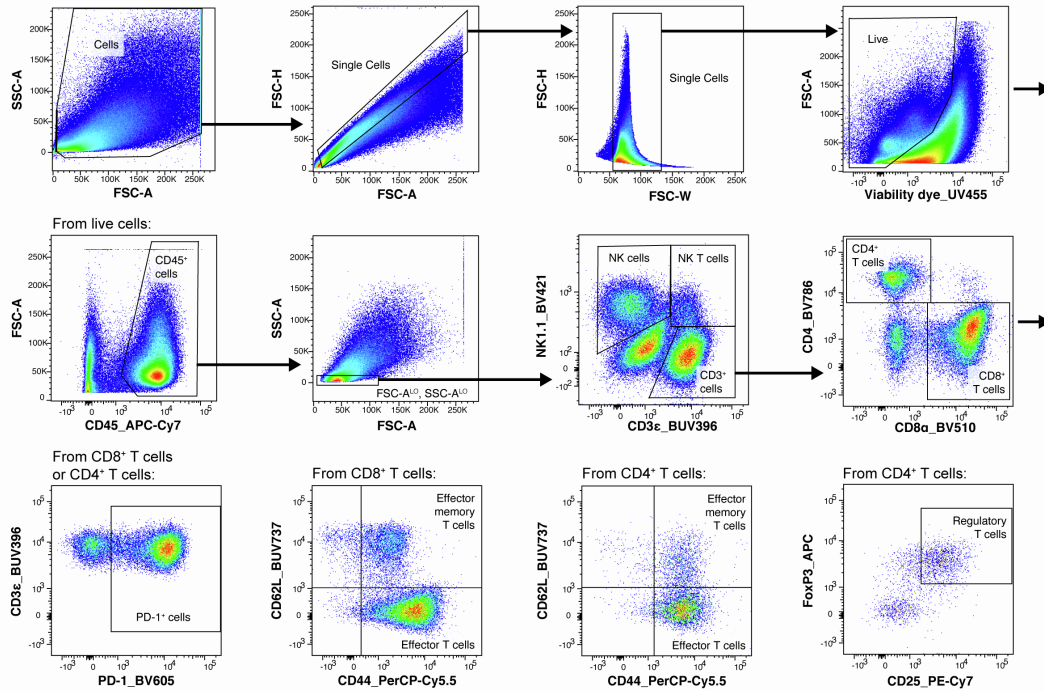


Figure S2. Gating strategy for the characterization of T and NK cells, related to Figure 1. Multi-colored flow cytometry was used to analyze the subsets of T and NK cells in the tumors at day 10 post-injection. Subset of immune cells were defined using the following markers: NK cells (FSC^{LO} , SSC^{LO} , $CD45^+$, $NK1.1^+$, $CD3\epsilon^+$), NK T cells (FSC^{LO} , SSC^{LO} , $CD45^+$, $CD3\epsilon^+$, $NK1.1^+$), $CD8^+$ T cells (FSC^{LO} , SSC^{LO} , $CD45^+$, $NK1.1^-$, $CD3\epsilon^+$, $CD8^+$), $CD4^+$ T cells (FSC^{LO} , SSC^{LO} , $CD45^+$, $NK1.1^-$, $CD3\epsilon^+$, $CD8^+$), effector T cells (same markers than T cells yet with $CD44^+$, $CD62L^-$), effector memory T cells (same markers than T cells yet with $CD44^+$, $CD62L^+$), regulatory T cells (same as $CD4^+$ T cells with $CD25^+$, $FoxP3^+$).

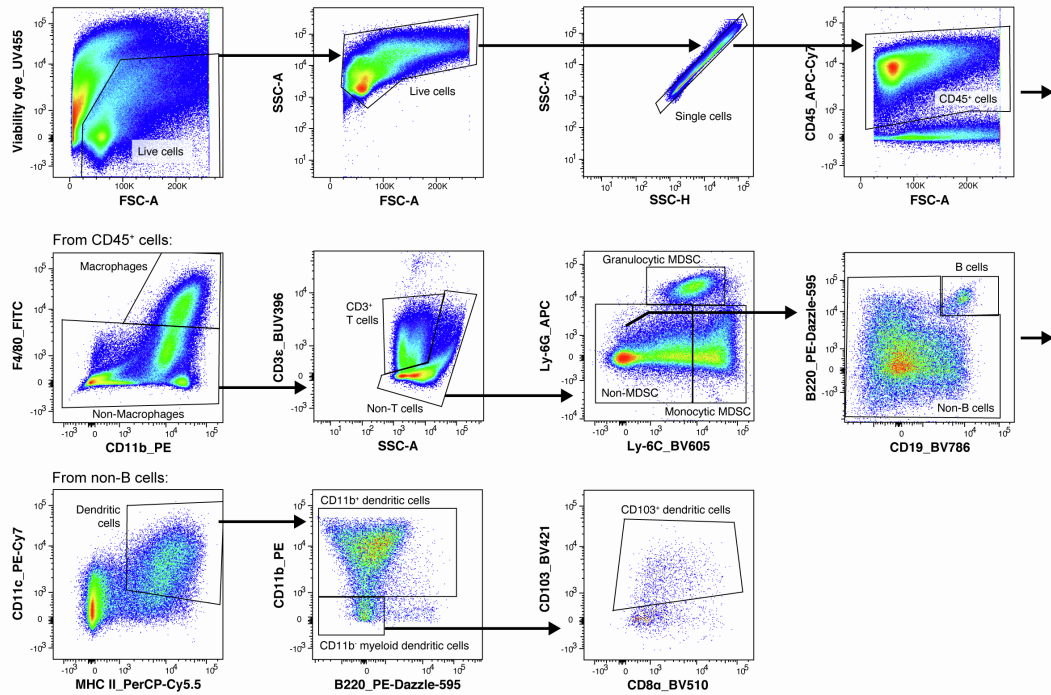


Figure S3. Gating strategy for the characterization of B cells and myeloid cell subsets, related to Figure 1. Multi-colored flow cytometry was used to analyze the subsets of B cells and myeloid cells in the tumors at day 10 post-injection. Subset of immune cells were defined using the following markers: Macrophages ($CD45^+$, $F4/80^+$, $CD11b^+$), Granylocytic myeloid-derived suppressor cells (MDSC) ($CD45^+$, $F4/80^-$, $CD3\epsilon^-$, $Ly6G^+$, $Ly6C^{MID/HI}$), Monocytic MDSC ($CD45^+$, $F4/80^-$, $CD3\epsilon^-$, $Ly6G^-$, $Ly6C^{HI}$), B cells ($CD45^+$, $F4/80^-$, $CD3\epsilon^-$, $Ly6G^-$, $Ly6C^{LO/MID}$, $CD19^+$, $B220^+$), dendritic cells (DCs) ($CD45^+$, $F4/80^-$, $CD3\epsilon^-$, $Ly6G^-$, $Ly6C^{LO/MID}$, $CD11c^+$, $MHCII^+$), $CD11b^+$ DCs (same than DCs with $CD11b^+$), $CD103^+$ DCs (same than DCs with $CD11b^-$, $B220^-$, $CD103^+$).

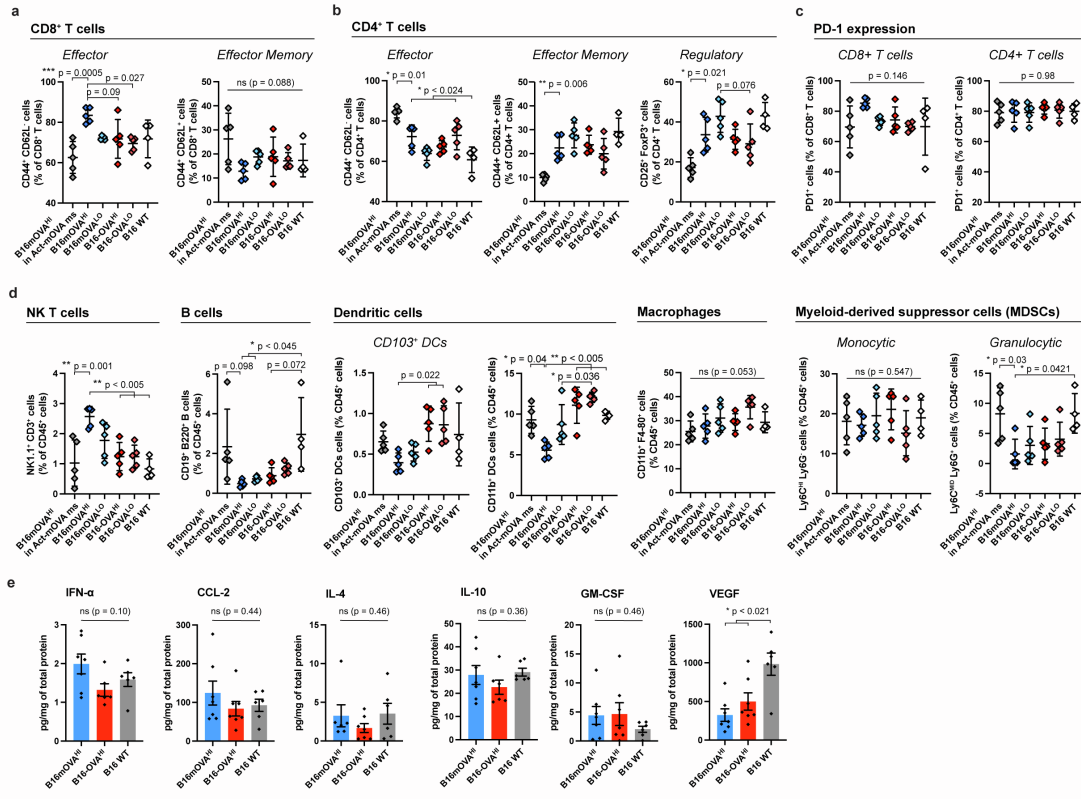


Figure S4. Comparison of B16mOVA and B16-OVA melanoma tumor immunogenicity in mice, related to Figure 1. Flow cytometry analysis of immune cells infiltrated in tumors 10 days post-injection ($N \geq 4$, mean \pm SD, ANOVA with Tukey's post-test and Brown-Forsythe correction when needed). **a**, CD8⁺ and **b**, CD4⁺ effector and effector memory T cells subsets in the different tumors. **c**, Proportion of PD-1 expressing CD8⁺ and CD4⁺ T cells. **d**, Proportion of NK T cells, B cells, dendritic cells, macrophages and myeloid-derived suppressor cells relative to the total CD45⁺ immune cell populations. **e**, Intratumoral cytokine quantification via Legendplex *in vivo* in tumors at day 10.

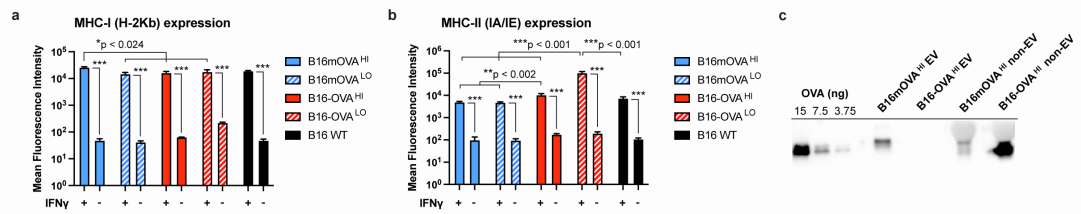


Figure S5. Expression of MHC by the different modified B16-F10 melanoma cells lines and presence of OVA on their secreted extracellular vesicles, related to Figure 1. **a**, MHC-I (H-2Kb) (**a**) and MHC-II (IA/IE) (**b**) expression on B16-mOVA and B16-OVA cell lines measured via flow cytometry. **c**, Western blot analysis for OVA detection in the extracellular vesicles (EV) produced *in vitro* by B16mOVA^{HI} or B16-OVA^{HI} cell lines or in the non-EV fraction (black = positive detection of OVA).

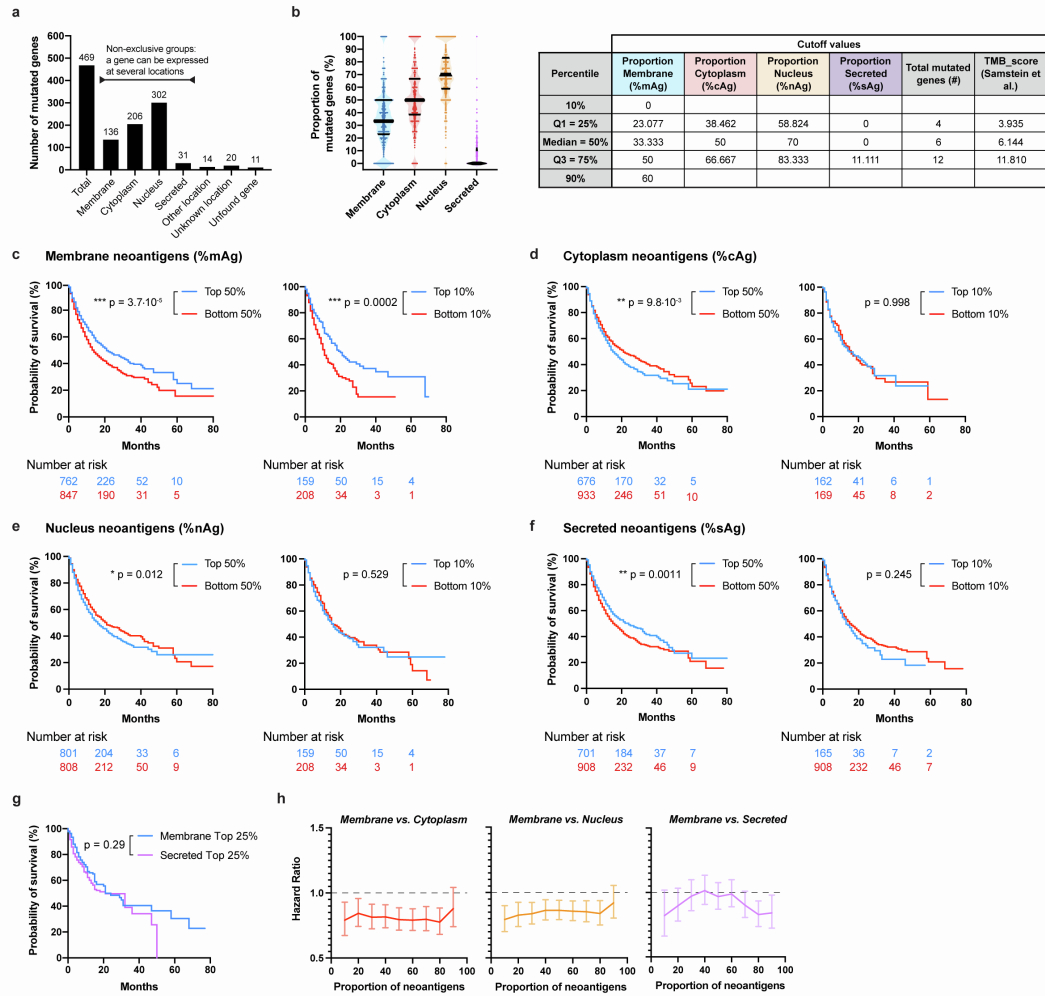


Figure S6. Analysis of the subcellular localizations of the tumor mutated genes and their subsequent impact on patient survival, related to Figure 3. **a**, Number of tumor mutated genes associated with each subcellular location among the 469 genes sequenced by MSK-IMPACT method. **b**, Proportion of tumor mutated genes per subcellular location in patients treated with immunotherapy in the pan-cancer group, and corresponding percentile cutoff values used for the analysis in Fig. 3. **c-f**, Survival of ICI-treated patients harboring high (Top 50% or 10%) or low (Bottom 50% or 10%) proportions of mAg (**c**) of cAg (**d**) of nAg (**e**) and of sAg (**f**) (log-rank tests). **g**, Survival of ICI-treated patients with high (Top 25%) proportion of mAg or sAg (log-rank test). **h**, HR of survival at different proportion of mAg vs cAg, nAg, or sAg (HR < 1 indicate significant extension of survival by mAg).

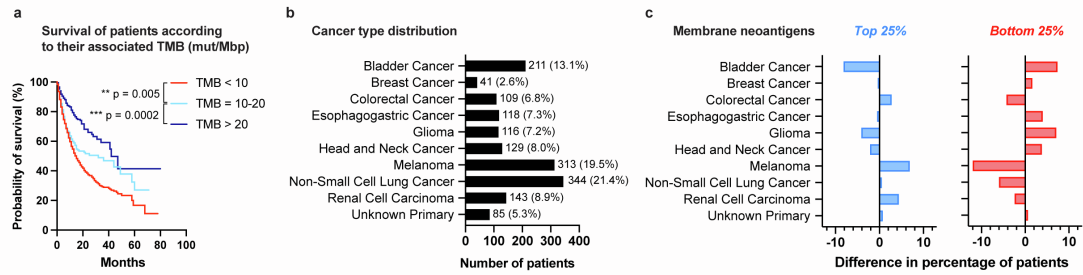


Figure S7. Survival of patients according to their TMB level, distributions of patients by cancer types and according to their mAg proportion, related to Figures 4 and 5. **a**, Survival of patients with TMB <10 mut/Mbp (69.7% of the cohort), TMB between 10 and 20 mut/Mbp (15.5% of the cohort) or TMB >20 mut/Mbp (14.6% of the cohort). No difference in survival were further observed for patients within the group with TMB > 20 mut/Mbp. **b**, Distribution of the ICI-treated patients per cancer type included in the pan-cancer analysis. **c**, Differences in patient distribution per cancer type for the groups with high (Top 25%) or low (Bottom 25%) proportion of mAg, as compared to the distribution of the entire ICI-treated cohort as in panel b.

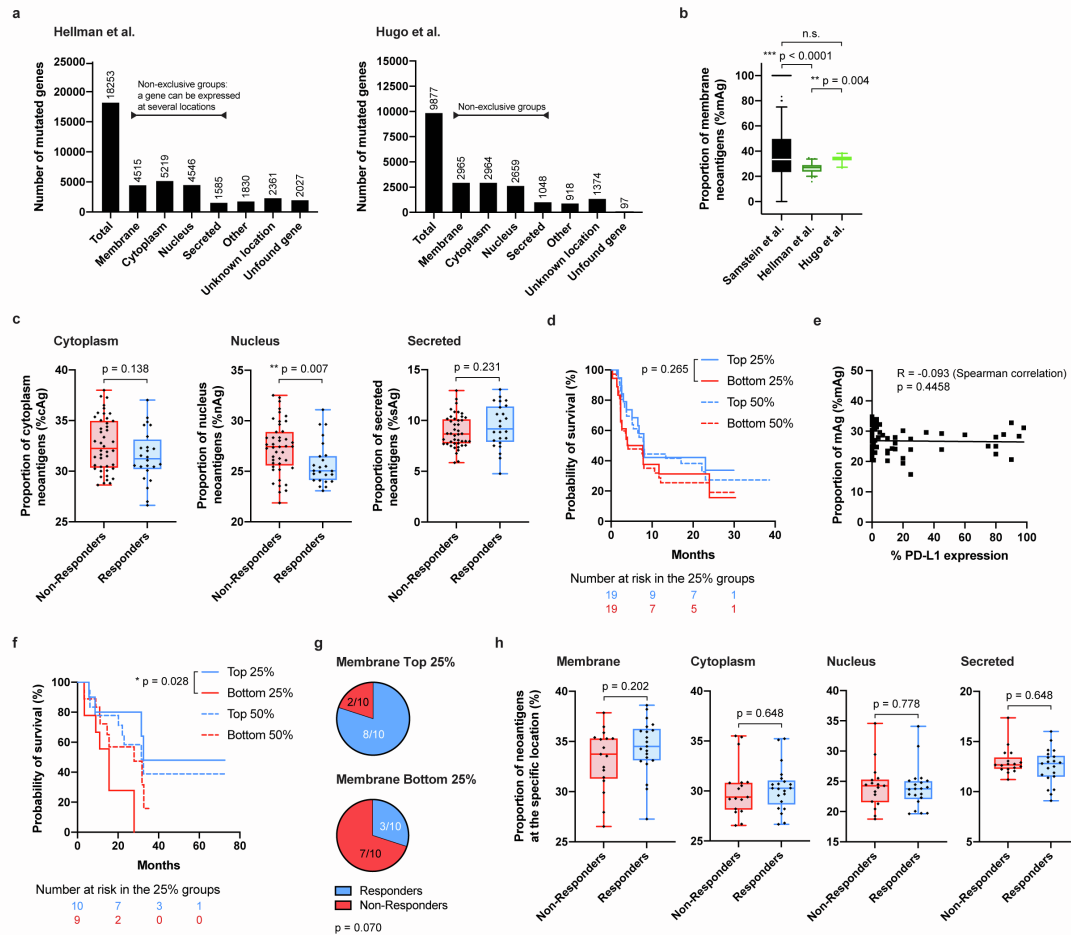


Figure S8. Response to immunotherapy based on the proportion of neoantigens at specific subcellular localizations, related to Figure 6. Patients (N=75) with non-small cell lung cancer were treated with anti-PD-1 + anti-CTLA-4 in the cohort from Hellman *et al.*⁸, and patients (N=38) with advanced melanoma cancer were treated with anti-PD-1 in the cohort from Hugo *et al.*⁹. In both studies, tumor mutated genes were sequenced by the WES method. **a**, Number of tumor mutated genes detected across all patients in the Hellman *et al.* and Hugo *et al.* studies, respectively, and their associated subcellular locations. **b**, Comparison of the mAg proportion detected in ICI-treated patient cohorts from the studies by Samstein *et al.*⁷, Hellman *et al.* and Hugo *et al.* **c**, Proportion of mutated genes per subcellular location in patients that responded or not to immunotherapy in the Hellman *et al.* cohort (Mann-Whitney test). **d**, Survival of patients with high (Top 25% and 50%) or low (Bottom 25% and 50%) proportion of mAg in the Hellman *et al.* cohort (log-rank test). **e**, Correlation between the proportion of mAg and the expression of PD-L1 in the tumor (Spearman correlation). **f**, Same as in panel d, but with the patient cohort from Hugo *et al.* **g**, Proportion of responders or non-responders to anti-PD-1 among patients that have high (Top 25%) or low (Bottom 25%) proportion of mAg in the Hugo *et al.* cohort (Fisher's exact test). **h**, Same as in panel c, but with the patient cohort from Hugo *et al.*

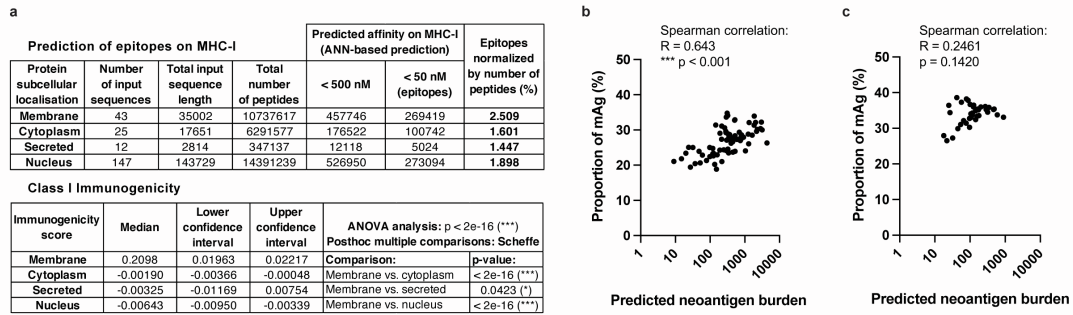


Figure S9. Mutated genes encoding for membrane proteins act as neoantigens, related to Figure 6. **a**, Prediction of peptides (all, and affinity < 500 nM) and epitopes (affinity < 50 nM) derived from membrane, cytoplasmic, nuclear and secreted proteins on MHC-I and comparison of their immunogenicity score using tools from IEDB³⁶. **b**, Correlations between the predicted neoantigen burden and the proportion of mAg in the dataset from Hellman *et al.* **c**, Same as panel b in the dataset from Hugo *et al.*

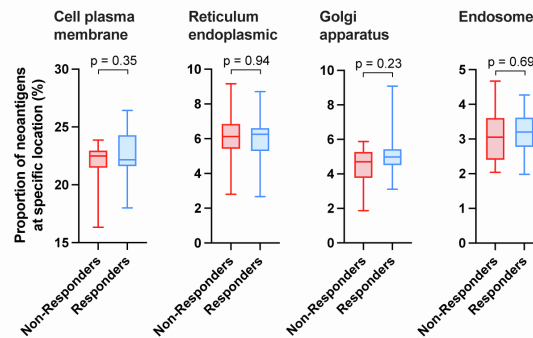


Figure S10. Proportion of mAg at the different membranes, related to Figure 6. Proportion of mutated genes encoding proteins at the cell plasma membrane or in other specific membrane-containing cell organelles in responders and non-responders to immunotherapy from the Hugo *et al.*

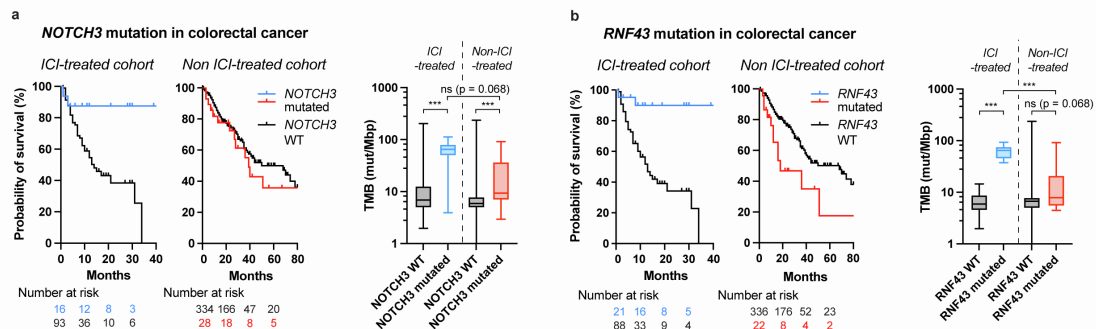


Figure S11. Specific membrane-associated mutated genes as predictive clinical biomarkers for extended survival upon ICI, related to Figure 7. Data analyzed from Samstein *et al.*⁷ using the ICI-treated or non-ICI-treated cohorts. **a**, **b**, Survival curves of ICI and non-ICI treated patients bearing *NOTCH3* (**a**) or *RNF43* (**b**) mutations in colorectal cancer and associated TMB burden.

Methods S1. Original codes used for the data processing and analysis

Three original codes were generated for processing using RStudio, version 1.4.1103 (2009-2021 Rstudio, PBC). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>

The same codes were used for the processing of the datasets of Samstein *et al.*, Hellman *et al.* and Hugo *et al.*; below are the codes used for the processing of Samstein *et al.* dataset.

Code n°1: Directory.

This code finds the subcellular location for each mutated genes appearing in the Samstein *et al.* cohort.

Input files:

- SamsteinData.csv (list of mutated genes per patient).
- HomoSapiens gene protein localisation (Data S1).

Output files:

- AllGeneNames.csv (subcellular location for all mutated genes).
- Samstein_directory-AllLocal.csv (classification of mutated genes in the categories: membrane, cytosol, nucleus and secreted).

Raw code:

```
library(corrplot)
library(tidyverse)
library(biomaRt)
library(data.table)
library(BiocManager)
library(UniProt.ws)
library(org.Hs.eg.db)
library(UniprotR)
#install.packages('openssl')
library(openssl)
UniProt.ws(taxId=9606)

rm(list=ls())

#load the directory from other patients
partial.directory <- read.csv("HomoSapiens Gene Protein Localization.csv", header = TRUE, sep = ";", stringsAsFactors = F)

#filter duplicates
Hdata <- read.csv("SamsteinData.csv", header = TRUE, sep = ";", stringsAsFactors = F)
names <- Hdata[,2]
names <- as.data.frame(names)
all.gene.names <- distinct(names)
col <- c(1:469)*0 # 469 is the total number of genes mutated in Samstein et al. patients.
all.gene.names <- cbind(all.gene.names, col)
all.gene.names <- as.data.frame(all.gene.names)
all.gene.names[,2] <- lapply(all.gene.names[,2], as.character)

#Go through all gene names and find localization
for(row in 1:469){
  gene.name <- all.gene.names[row,1]
  if(!isEmpty(gene.name[like(gene.name,";")])){
    ind<-str_locate(gene.name,";")
    pos<-ind[1]-1
    gene.name<-substr(gene.name,1,pos)
  }
}
#Locate the index where gene.name is found in directory
```

```

row.number <- grep(gene.name, partial.directory[,5], ignore.case = FALSE, perl = F, value = FALSE,
  fixed = FALSE, useBytes = FALSE, invert = FALSE)

final.row.number<-integer()

if(length(row.number) != 0){

  # Ensure that the gene's name is not a substring of another gene name
  for (ligne in 1:length(row.number)){
    list.names<-str_split(partial.directory[row.number[ligne],5], " ", simplify=TRUE)
    if(gene.name %in% list.names){
      final.row.number<-rbind(final.row.number, row.number[ligne])
    }
  }
  final.row.number<-as.integer(final.row.number)
}else {final.row.number<-row.number}

local<-character()

# Find the gene subcellular location

if(length(final.row.number)==0){
all.gene.names[row,2]<-as.character("GENEnotFOUND")
count.missed.genes<-count.missed.genes+1
missed.genes <- rbind(missed.genes,data.frame(gene.name))
}else {
for (l in final.row.number){#1:length(final.row.number)}
  local<-paste(local,partial.directory[l,9],sep="") # 7 au lieu de 9 si on clean partial directory
}

if (local!="") {
  #get localization info
  all.gene.names[row,2]<-local
  #print("Sublocalization filled")
}else {
  #if not found in partial directory then get it from website
  new.localization <- GetSubcellular_location(partial.directory[final.row.number[1],1], directorypath = NULL)
  if(is.na(new.localization[1,1])){
    # print("Sublocalization unknown")
    all.gene.names[row,2]<-as.character("Unknown")
    unknown.localization<-rbind(unknown.localization,data.frame(gene.name,partial.directory[row.number,1]))
    count.unknown.localization<-count.unknown.localization+1
  }else{
    #print("Sublocalization FOUND")
    count.found.localization<-count.found.localization+1
    newlocal<-as.character(new.localization[1,1])
    all.gene.names[row,2]<-newlocal
  }
}
}
}
}
write.csv(all.gene.names, file = "AllGeneNames.csv", row.names= FALSE, sep = ",")

all.gene.names<-read.csv("AllGeneNames.csv", header = TRUE, sep = ",", stringsAsFactors = F)
directory<-all.gene.names

#Now make the localization in different columns
for (i in 1:469){
  protein<-directory[i,2]

  directory[i,3]<-0
  directory[i,3]<-!isEmpty(protein[like(protein, "Nucleus")])
  if (directory[i,3] == "TRUE"){
    directory[i,3] <- 1
  }
  directory[i,4]<-0
}

```

```

directory[i,4]<-!isEmpty(protein[like(protein, "Cytoplasm")])
if (directory[i,4] == "TRUE"){
  directory[i,4] <- 1
}
directory[i,5]<-0
directory[i,5]<-!isEmpty(protein[like(protein, "Cell membrane")|like(protein,"Membrane")])
if (directory[i,5] == "TRUE"){
  directory[i,5] <- 1
}
directory[i,6]<-0
directory[i,6]<-!isEmpty(protein[like(protein, "Secreted")])
if (directory[i,6] == "TRUE"){
  directory[i,6] <- 1
}
directory[i,7]<-0
directory[i,7]<-!isEmpty(protein[like(protein, "Cell membrane")])
if (directory[i,7] == "TRUE"){
  directory[i,7] <- 1
}
directory[i,8]<-0
directory[i,8]<-!isEmpty(protein[like(protein, "Membrane")])
if (directory[i,8] == "TRUE"){
  directory[i,8] <- 1
}
directory[i,9]<-0
directory[i,9]<-!isEmpty(protein[like(protein, "reticulum")|like(protein,"Reticulum")])
if (directory[i,9] == "TRUE"){
  directory[i,9] <- 1
}
directory[i,10]<-0
directory[i,10]<-!isEmpty(protein[like(protein, "golgi")|like(protein,"Golgi")])
if (directory[i,10] == "TRUE"){
  directory[i,10] <- 1
}
directory[i,11]<-0
directory[i,11]<-!isEmpty(protein [like(protein, "endosome") | like(protein,"Endosome")| like(protein,"Endosomal") |
like(protein,"endosomal") ] )
if (directory[i,11] == "TRUE"){
  directory[i,11] <- 1
}
directory[i,12]<-0
directory[i,12]<-!isEmpty(protein[like(protein, "Mitochondrion")])
if (directory[i,12] == "TRUE"){
  directory[i,12] <- 1
}
}
colnames(directory)<-c("GeneName", "Localization", "Nucleus", "Cytoplasm", "CellMembraneBoth", "Secreted", "CellMem",
"Membrane", "Reticulum", "Golgi", "Endosome", "Mitochondrion")
write.csv(directory, file = "Samstein_directory-AllLocal.csv", row.names= FALSE, sep = ",")

```

Code n°2: Patients.

This code lists the subcellular location of mutated genes for each patients of Samstein et al.

Input files:

- SamsteinData.csv (list of mutated genes per patient).
- Samstein_directory-AllLocal.csv (from Code n°1).

Output files:

- AllPatientsLocalization_Samstein-AllLocal.csv (list of patients' mutated genes with the corresponding subcellular locations)

Raw code:

```
library(corrplot)
library(tidyverse)
library(biomaRt)
library(data.table)
library(BiocManager)
library(UniProt.ws)
library(org.Hs.eg.db)
library(UniprotR)
library(openssl)
UniProt.ws(taxId=9606)

rm(list=ls())

AllPatients <- read.csv("SamsteinData.csv", header = TRUE, sep = ";", stringsAsFactors = F)
AllPatients <- distinct(AllPatients) # -- No if snSNV

SamsteinDirectory <- read.csv("Samstein_directory-AllLocal.csv", header = TRUE, sep = ";", stringsAsFactors = F)

row.number <- which(AllPatients[,2] == "")

for (i in row.number){
  AllPatients[i,1]="X"
}

for (i in 1:nrow(AllPatients)){

  row.number <- which(SamsteinDirectory$GeneName == AllPatients[i,2])

  if(length(row.number)!=0){
    AllPatients[i,3:12] <- SamsteinDirectory[row.number,3:12]
  }
}

write.csv(AllPatients, file = "AllPatientsLocalization_Samstein-AllLocal.csv", row.names= FALSE, sep = ";")
```

Code n°3: SplitPatientsMutations.

This code finds the subcellular location for each mutated genes in the patients cohort of Samstein et al.

Input files:

- AllPatientsLocalization_Samstein-AllLocal. (from Code n°2).

Output files:

- SummaryMutationsPatients-AllLocal.csv (mutation counts per localisation per patient)

Raw code:

```
library(corrplot)
library(tidyverse)
library(biomaRt)
library(data.table)
library(BiocManager)
library(UniProt.ws)
library(org.Hs.eg.db)
library(UniprotR)
library(openssl)
UniProt.ws(taxId=9606)

AllPatients <- read.csv("AllPatientsLocalization_Samstein-AllLocal.csv", header = TRUE, sep = ",", stringsAsFactors = F)

NumberPatients <- AllPatients[,1]
NumberPatients<-as.data.frame(NumberPatients)
NumberPatients<-distinct(NumberPatients)

for (i in 1:length(NumberPatients[,1])){

  row.patients<-which(AllPatients[,1]==NumberPatients[i,1])

  TotalMut<-length(row.patients)
  MutNucleus<-0
  MutCytoplasm<-0
  MutMembraneBoth<-0
  MutSecreted<-0
  MutCell<-0
  MutMem<-0
  MutReti<-0
  MutGolgi<-0
  MutEndos<-0
  MutMitoch<-0

  for (p in row.patients)
  {
    MutNucleus<-MutNucleus+AllPatients[p,3]
    MutCytoplasm<-MutCytoplasm+AllPatients[p,4]
    MutMembraneBoth<-MutMembraneBoth+AllPatients[p,5]
    MutSecreted<-MutSecreted+AllPatients[p,6]
    MutCell<-MutCell+AllPatients[p,7]
    MutMem<-MutMem+AllPatients[p,8]
    MutReti<-MutReti+AllPatients[p,9]
    MutGolgi<-MutGolgi+AllPatients[p,10]
    MutEndos<-MutEndos+AllPatients[p,11]
    MutMitoch<-MutMitoch+AllPatients[p,12]

    # Somme de toutes les colonnes sauf CellMem and Membrane
    res<-
    AllPatients[p,3]+AllPatients[p,4]+AllPatients[p,5]+AllPatients[p,6]+AllPatients[p,9]+AllPatients[p,10]+AllPatients[p,11]+AllPatients[p,12]
```

```

resMem<-
AllPatients[p,3]+AllPatients[p,4]+AllPatients[p,6]+AllPatients[p,7]+AllPatients[p,8]+AllPatients[p,9]+AllPatients[p,10]+AllPatients[p,11]+AllPatients[p,12]
}

NumberPatients[i,2]<-TotalMut
NumberPatients[i,3]<-MutNucleus
NumberPatients[i,4]<-MutCytoplasm
NumberPatients[i,5]<-MutMembraneBoth
NumberPatients[i,6]<-MutSecreted
NumberPatients[i,7]<-MutCell
NumberPatients[i,8]<-MutMem
NumberPatients[i,9]<-MutReti
NumberPatients[i,10]<-MutGolgi
NumberPatients[i,11]<-MutEndos
NumberPatients[i,12]<-MutMitoch
}

colnames(NumberPatients)<-c("Patients", "TotalMutations", "TotalNucleusGenes", "TotalCytoplasmGenes",
"TotalMembraneBothGenes", "TotalSecretedGenes", "TotalCellOnlyGenes", "TotalMemOnlyGenes", "TotalReticulumGenes",
"TotalGolgiGenes", "TotalEndosomalGenes", "TotalMitochondriumGenes")
#colnames(NumberPatients)<-c("Patients", "TotalMutations", "TotalNucleusMutations", "TotalCytoplasmMutations",
"TotalMembraneBothMutations", "TotalSecretedMutations", "TotalCellOnlyMutations", "TotalMemOnlyMutations",
"TotalReticulumMutations", "TotalGolgiMutations", "TotalEndosomalMutations", "TotalMitochondriumMutations")
write.csv(NumberPatients, file = "SummaryMutationsPatients-AllLocal-verif.csv", row.names= FALSE, sep = ",")

```