Table S5. The percentage of randomly selected 8-genes models predicting PFS and OS based on Cox proportional hazard models at $P \leq 0.05$.

	Training	Internal Validation	External Validation
No. of patients	123	82	111
PFS			
random 8 gene panels	9.7%	4.2%	4.9%
non-GMS random 8 gene panels	7.9%	3.1%	4.1%
OS			
random 8 gene panels	6.3%	3.7%	NA
non-GMS random 8 gene panels	5.3%	2.4%	NA

Abbreviations: GMS, genomic mutation signature; PFS, progression free survival; OS, overall survival.

Table S6. The percentage of randomly selected 8-genes models which outperform GMS based on Cox proportional hazard models predicting PFS and OS

	Training	Internal Validation	External Validation
No. of patients	123	82	111
PFS	$\boldsymbol{P_{GMS}}$**=0.0001**	$\boldsymbol{P_{GMS}}$**=0.0102**	$\boldsymbol{P_{GMS}}$**=0.0100**
random 8 gene panels	0.0%	0.7%	1.1%
non-GMS random 8 gene panels	0.0%	0.5%	0.8%
OS	$\boldsymbol{P_{GMS}}$**=0.0129**	$\boldsymbol{P_{GMS}}$**=0.0047**	NA
random 8 gene panels	2.6%	0.4%	NA
non-GMS random 8 gene panels	1.9%	0.2%	NA

Abbreviations: GMS, gene mutation-based signature; PFS, progression free survival; OS, overall survival.

Sensitivity Analysis: Clinical Outcomes of Random Gene Panels and non-GMS Random Gene Panels

METHORDS

To prove that genes in our GMS model are nonrandom and irreplaceable, sensitivity analysis was

performed in training cohort, internal validation cohort and external validation cohort in turn. Two separate analyses were performed in each cohort:

1. Clinical Outcomes of Random Gene Panels

For training cohort and internal validation cohort, 1000 random panels containing 8 genes were randomly generated from MSK-IMPACT gene panels. Cox proportional hazard model was performed for each random panel to examine for association with PFS and OS, after which the P value of the model was calculated. The random panels which were significant and more significant than GMS were counted respectively.

For external validation cohort, included Checkmate-012 and Keynote/SU2C cohorts, which were profiled by whole exome sequencing (WES). 8 genes from WES gene list were randomly selected to generate random panels. Following steps were the same as those in training cohort and internal validation cohort.

2. Clinical Outcomes of non-GMS Random Gene Panels

For training cohort and internal validation cohort, 8 genes of GMS were first excluded from the MSK-IMPACT gene list. Then 1000 random panels containing 8 genes were randomly generated from this non-GMS MSK-IMPACT gene list. Cox proportional hazard model was performed for each random panel to examine for association with PFS and OS, after which the P value of the model was calculated. The random panels which were significant and more significant than GMS were counted respectively.

For external validation cohort which have mutation information of WES gene list, 8 genes of GMS were first excluded from human gene list. 1000 random panels containing 8 genes were randomly generated from this non-GMS WES gene list. Following steps are the same as those in training cohort and internal validation cohort.

RESULTS

1. Random Gene Panels

1) PFS

For training cohort, 9.7% random gene panels were significant, while none was more significant than GMS in predicting PFS at P = 0.0001.

For internal validation cohort, 4.2% random gene panels were significant, while only 0.7% were more significant than GMS in predicting PFS at P = 0.0102.

For external validation cohort, 4.9% random gene panels were significant, while only 1.1% were more significant than GMS in predicting PFS at P = 0.0100.

2) OS

For training cohort, 6.3% random gene panels were significant, while only 2.6% were more significant than GMS in predicting OS at P = 0.0129.

For internal validation cohort, 3.7% random gene panels were significant, while only 0.4% were more significant than GMS in predicting OS at P = 0.0047.

Note that external validation cohort lacks the information of OS.

2. non-GMS Random Gene Panels

3) PFS

For training cohort, 7.9% non-GMS random gene panels were significant, while none was more significant than GMS in predicting PFS at P = 0.0001.

For internal validation cohort, 3.1% non-GMS random gene panels were significant, while only 0.5% were

more significant than GMS in predicting PFS at P = 0.0102.
For external validation cohort, 4.1% non-GMS random gene panels were significant, while only 0.8% were more significant than GMS in predicting PFS at P = 0.0100.
4) OS
For training cohort, 5.3% non-GMS random gene panels were significant, while only 1.9% were more significant than GMS in predicting OS at P = 0.0129.
For internal validation cohort, 2.4% non-GMS random gene panels were significant, while only 0.2% were more significant than GMS in predicting OS at P = 0.0047.
Note that external validation cohort lacks the information of OS.

DISSCUSSION

The low percentage of random gene panels which were significant for PFS and OS demonstrated that our GMS model is nonrandom. In addition, compared with 1000 random gene panels, 1000 non-GMS random gene panels had a lower percentage of significance, which means the 8 genes in GMS play crucial roles in predicting PFS and OS.
The result of sensitivity analysis also demonstrated that our GMS model is irreplaceable for the fact that the percentage of random gene panels and non-GMS random panels which were more significant than GMS were both close to zero.