Supplemental Information

Figures S1-S8

Table S8

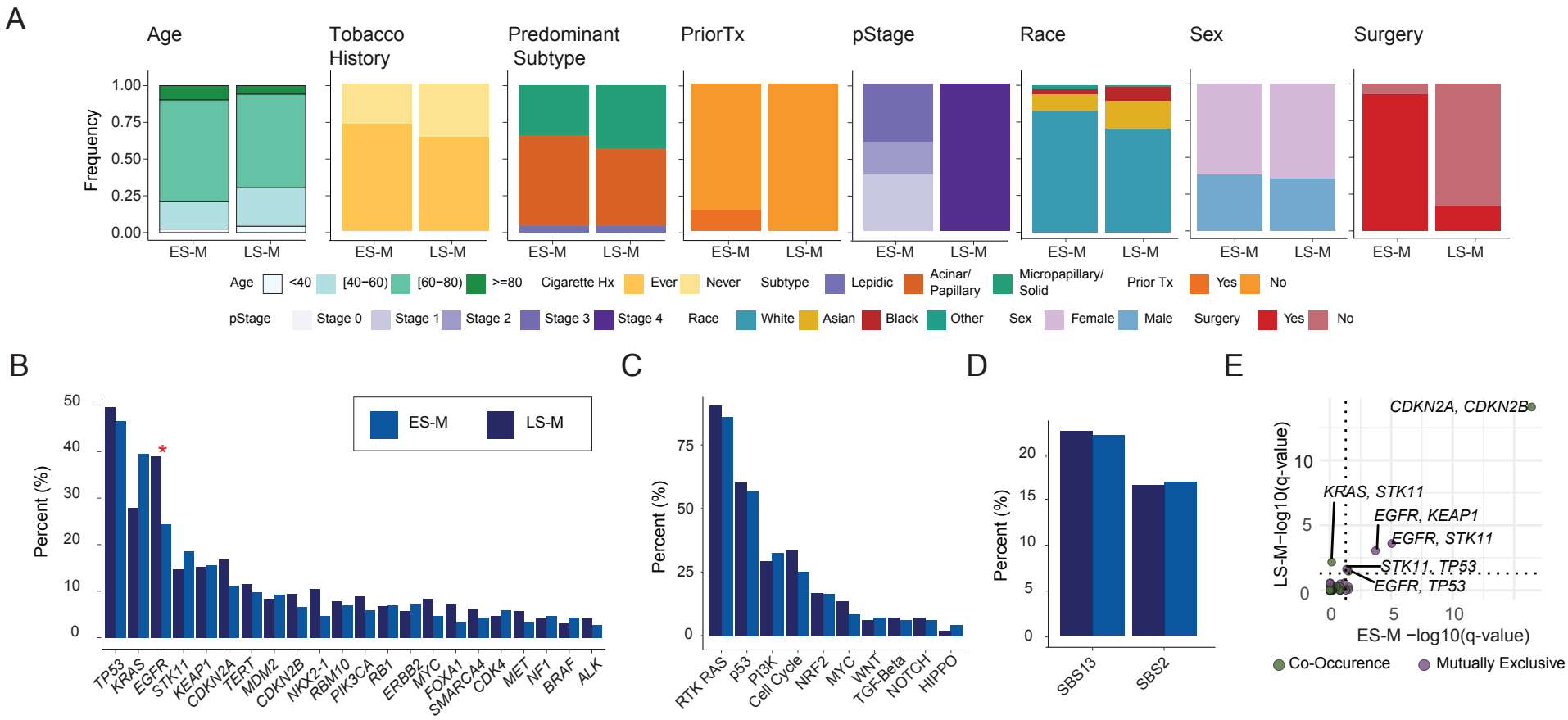

A
Age
Tobacco History
Predominant Subtype
PriorTx
pStage
Race
Sex
Surgery
Frequency
1.00
0.75
0.50
0.25
0.00
ES-M
LS-M
Age <40 [40−60) [60−80) >=80
Cigarette Hx Ever Never
Subtype Lepidic Acinar/ Papillary Micropapillary/ Solid
Prior Tx Yes No
pStage Stage 0 Stage 1 Stage 2 Stage 3 Stage 4
Race White Asian Black Other
Sex Female Male
Surgery Yes No
B
Percent (%)
ES-M
LS-M
TP53 KRAS EGFR STK11 KEAP1 CDKN2A TERT MDM2 CDKN2B NKX2-1 RBM10 PIK3CA RB1 ERBB2 MYC FOXA1 SMARCA4 CDK4 MET NF1 BRAF ALK
C
Percent (%)
RTK RAS p53 PI3K Cell Cycle NRF2 MYC WNT TGF-Beta NOTCH HIPPO
D
Percent (%)
SBS13 SBS2
E
CDKN2A, CDKN2B
KRAS, STK11
EGFR, KEAP1
EGFR, STK11
STK11, TP53
EGFR, TP53
LS-M−log10(q-value)
ES-M −log10(q-value)
Co-Occurence
Mutually Exclusive

Figure S1

Figure S1. Clinicopathologic and genomic differences between early-stage metastatic (ES-M) and late-stage metastatic (LS-M) primary tumors, Related to Figure 1. (A) Bar plots comparing clinicopathologic features between ES-M and LS-M groups. (B) Bar plots showing the percentage of samples with gene alterations. (C) Bar plots depicting the percentage of samples with pathway alterations. (D) Bar plots displaying the percentage of samples with APOBEC signatures present. (E) Scatter plot showing *q*-value of gene pairs for co-occurrence and mutual exclusivity in ES-M and LS-M tumors. Dotted lines represent $q=0.05$; significant gene pairs labeled. Statistical analyses: (A-E) Fisher's exact test. *$q<0.05$, adjusted for false-discovery rate (FDR). Hx, history; pStage, pathologic stage; Tx, treatment.

Figure S2

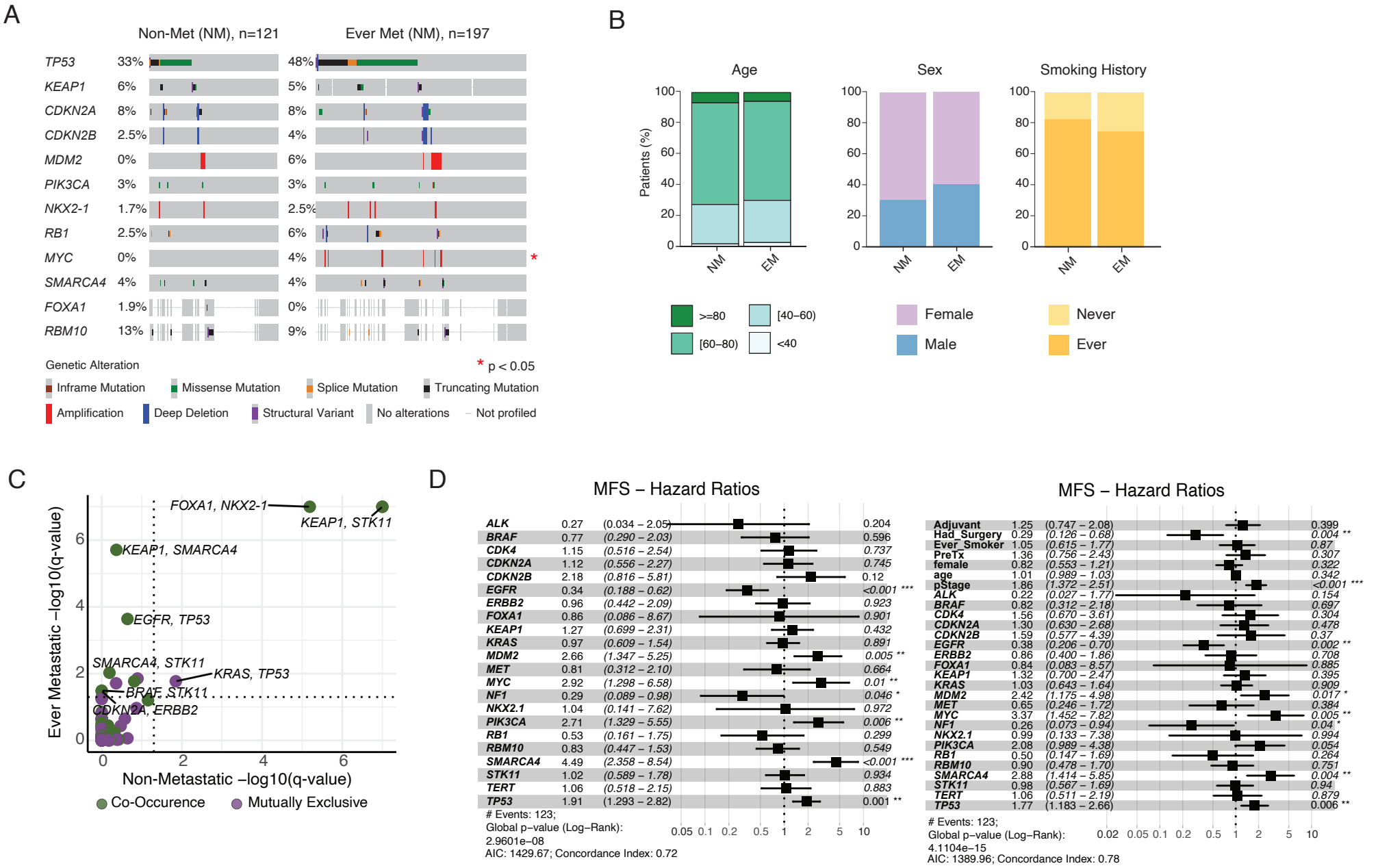

A
Non-Met (NM), n=121
Ever Met (NM), n=197
TP53 33% 48%
KEAP1 6% 5%
CDKN2A 8% 8%
CDKN2B 2.5% 4%
MDM2 0% 6%
PIK3CA 3% 3%
NKX2-1 1.7% 2.5%
RB1 2.5% 6%
MYC 0% 4% *
SMARCA4 4% 4%
FOXA1 1.9% 0%
RBM10 13% 9%
Genetic Alteration
* p < 0.05
Inframe Mutation
Missense Mutation
Splice Mutation
Truncating Mutation
Amplification
Deep Deletion
Structural Variant
No alterations
Not profiled
B
Age
Sex
Smoking History
Patients (%)
NM
EM
>=80
[60–80)
[40–60)
<40
Female
Male
Never
Ever
C
Ever Metastatic –log10(q-value)
Non-Metastatic –log10(q-value)
FOXA1, NKX2-1
KEAP1, STK11
KEAP1, SMARCA4
EGFR, TP53
SMARCA4, STK11
KRAS, TP53
BRAF, STK11
CDKN2A, ERBB2
Co-Occurence
Mutually Exclusive
D
MFS – Hazard Ratios
ALK 0.27 (0.034 – 2.05) 0.204
BRAF 0.77 (0.290 – 2.03) 0.596
CDK4 1.15 (0.516 – 2.54) 0.737
CDKN2A 1.12 (0.556 – 2.27) 0.745
CDKN2B 2.18 (0.816 – 5.81) 0.12
EGFR 0.34 (0.188 – 0.62) <0.001 ***
ERBB2 0.96 (0.442 – 2.09) 0.923
FOXA1 0.86 (0.086 – 8.67) 0.901
KEAP1 1.27 (0.699 – 2.31) 0.432
KRAS 0.97 (0.609 – 1.54) 0.891
MDM2 2.66 (1.347 – 5.25) 0.005 **
MET 0.81 (0.312 – 2.10) 0.664
MYC 2.92 (1.298 – 6.58) 0.01 **
NF1 0.29 (0.089 – 0.98) 0.046 *
NKX2.1 1.04 (0.141 – 7.62) 0.972
PIK3CA 2.71 (1.329 – 5.55) 0.006 **
RB1 0.53 (0.161 – 1.75) 0.299
RBM10 0.83 (0.447 – 1.53) 0.549
SMARCA4 4.49 (2.358 – 8.54) <0.001 ***
STK11 1.02 (0.589 – 1.78) 0.934
TERT 1.06 (0.518 – 2.15) 0.883
TP53 1.91 (1.293 – 2.82) 0.001 **
Events: 123;
Global p–value (Log–Rank): 2.9601e–08
AIC: 1429.67; Concordance Index: 0.72
0.05 0.1 0.2 0.5 1 2 5 10
MFS – Hazard Ratios
Adjuvant 1.25 (0.747 – 2.08) 0.399
Had_Surgery 0.29 (0.126 – 0.68) 0.004 **
Ever_Smoker 1.05 (0.615 – 1.77) 0.87
PreTx 1.36 (0.756 – 2.43) 0.307
female 0.82 (0.553 – 1.21) 0.322
age 1.01 (0.989 – 1.03) 0.342
pStage 1.86 (1.372 – 2.51) <0.001 ***
ALK 0.22 (0.027 – 1.77) 0.154
BRAF 0.82 (0.312 – 2.18) 0.697
CDK4 1.56 (0.670 – 3.61) 0.304
CDKN2A 1.30 (0.630 – 2.68) 0.478
CDKN2B 1.59 (0.577 – 4.39) 0.37
EGFR 0.38 (0.206 – 0.70) 0.002 **
ERBB2 0.86 (0.400 – 1.86) 0.708
FOXA1 0.84 (0.083 – 8.57) 0.885
KEAP1 1.32 (0.700 – 2.47) 0.395
KRAS 1.03 (0.643 – 1.64) 0.909
MDM2 2.42 (1.175 – 4.98) 0.017 *
MET 0.65 (0.246 – 1.72) 0.384
MYC 3.37 (1.452 – 7.82) 0.005 **
NF1 0.26 (0.073 – 0.94) 0.04 *
NKX2.1 0.99 (0.133 – 7.38) 0.994
PIK3CA 2.08 (0.989 – 4.38) 0.054
RB1 0.50 (0.147 – 1.69) 0.264
RBM10 0.90 (0.478 – 1.70) 0.751
SMARCA4 2.88 (1.414 – 5.85) 0.004 **
STK11 0.98 (0.567 – 1.69) 0.94
TERT 1.06 (0.511 – 2.19) 0.879
TP53 1.77 (1.183 – 2.66) 0.006 **
Events: 123;
Global p–value (Log–Rank): 4.1104e–15
AIC: 1389.96; Concordance Index: 0.78
0.02 0.05 0.1 0.2 0.5 1 2 5 10

Figure S2. Clinicopathologic and genomic differences between nonmetastatic (NM) and ever-metastatic (EM) primary tumors, Related to Figure 2. (A) Oncoprint of NM and EM primary tumors from Dana-Farber Cancer Institute in the GENIE-Biopharma Collaborative data set. All genes altered at significantly different frequencies between the two groups in our cohort are displayed. (B) Comparisons of clinicopathologic features between nonmetastatic and metastatic patients in the GENIE validation cohort. (C) Scatter plot of the *q*-values of gene pairs for co-occurrence and mutual exclusivity between NM and EM tumors. The indicated gene pairs are those shared between EM and NM tumors and those private to NM tumors. Dotted lines represent q=0.05. (D) Forrest plot showing metastasis-free survival (MFS) for all genes altered in at least 3% of samples (*left*) and genes plus clinicopathologic features (*right*). Statistical analyses: (A-C) Fisher's exact test. (D) Cox-proportional hazards, log-rank test. *p*-values are as indicated. Squares represent hazard ratio (HR) and whiskers display 95% confidence interval (CI).

Figure S3

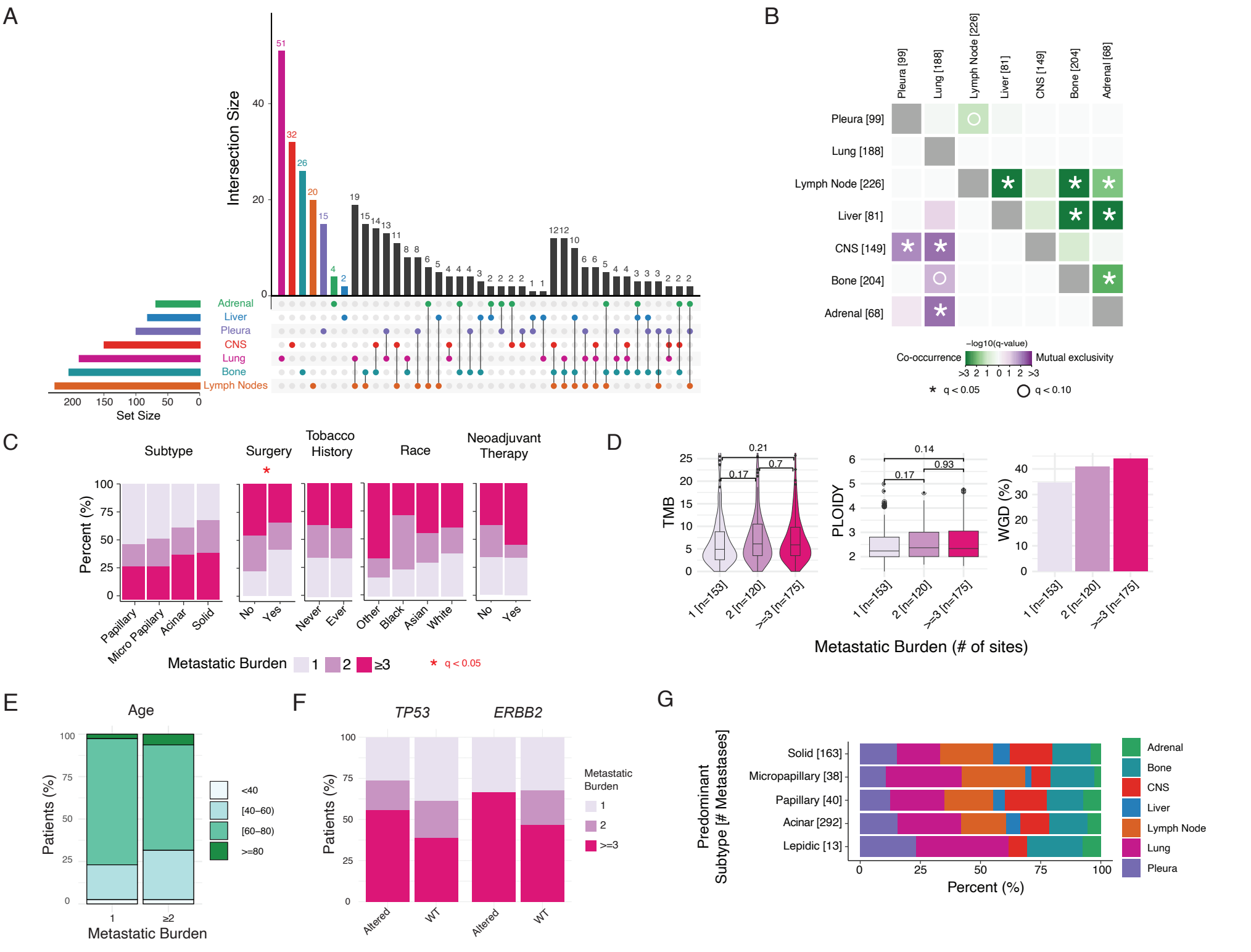

A
Intersection Size
Set Size
Adrenal
Liver
Pleura
CNS
Lung
Bone
Lymph Nodes
B
Pleura [99]
Lung [188]
Lymph Node [226]
Liver [81]
CNS [149]
Bone [204]
Adrenal [68]
−log10(q-value)
Co-occurrence
Mutual exclusivity
* q < 0.05
q < 0.10
C
Subtype
Surgery
Tobacco History
Race
Neoadjuvant Therapy
Percent (%)
Papillary
Micro Papilary
Acinar
Solid
No
Yes
Never
Ever
Other
Black
Asian
White
Metastatic Burden 1 2 ≥3
* q < 0.05
D
TMB
PLOIDY
WGD (%)
1 [n=153]
2 [n=120]
>=3 [n=175]
Metastatic Burden (# of sites)
E
Age
Patients (%)
<40
[40–60)
[60–80)
>=80
Metastatic Burden
F
TP53
ERBB2
Patients (%)
Metastatic Burden
Altered
WT
G
Predominant Subtype [# Metastases]
Solid [163]
Micropapillary [38]
Papillary [40]
Acinar [292]
Lepidic [13]
Percent (%)
Adrenal
Bone
CNS
Liver
Lymph Node
Lung
Pleura

Figure S3. Comparisons of site-specific non- and ever-metastatic tumors and metastatic burden, Related to Figure 3. (A). Upset plot illustrating abundance of metastatic patterns across patients with at most 3 metastatic sites. (B) Heat map demonstrating co-occurrence and mutual exclusivity between metastatic sites. (C) Bar plots showing the distribution of metastatic burden for clinicopathologic features. (D) Violin plot of the distribution of tumor mutational burden (TMB) and boxplot showing ploidy for tumors with 1, 2, or ≥3 distinct metastatic sites. Boxplots display median values, interquartile range (IQR) boxes, and whiskers demonstrating 1.5 x IQR. *p*-values for pairwise comparisons between groups noted. Bar plots demonstrating proportion of patients with whole-genome duplication (WGD) by number of metastatic sites. (E) Bar plot showing the breakdown of age in patients, stratified by metastatic burden, for GENIE validation cohort. (F) Bar plots displaying the breakdown of metastatic burden for TP53 and ERBB2 altered and wild-type (WT) tumors for GENIE validation cohort. (G) Bar plots showing proportion of primary tumors with metastasis to a given site stratified by predominant histologic subtype. Statistical analyses: (B-C, E-G) Fisher's exact test. *q*-values correct for multiple comparisons using the false-discovery rate (FDR). (D) Wilcoxon rank-sum test for TMB and ploidy; Fisher's exact test for WGD. CNS, central nervous system; Met, metastatic.

Figure S4

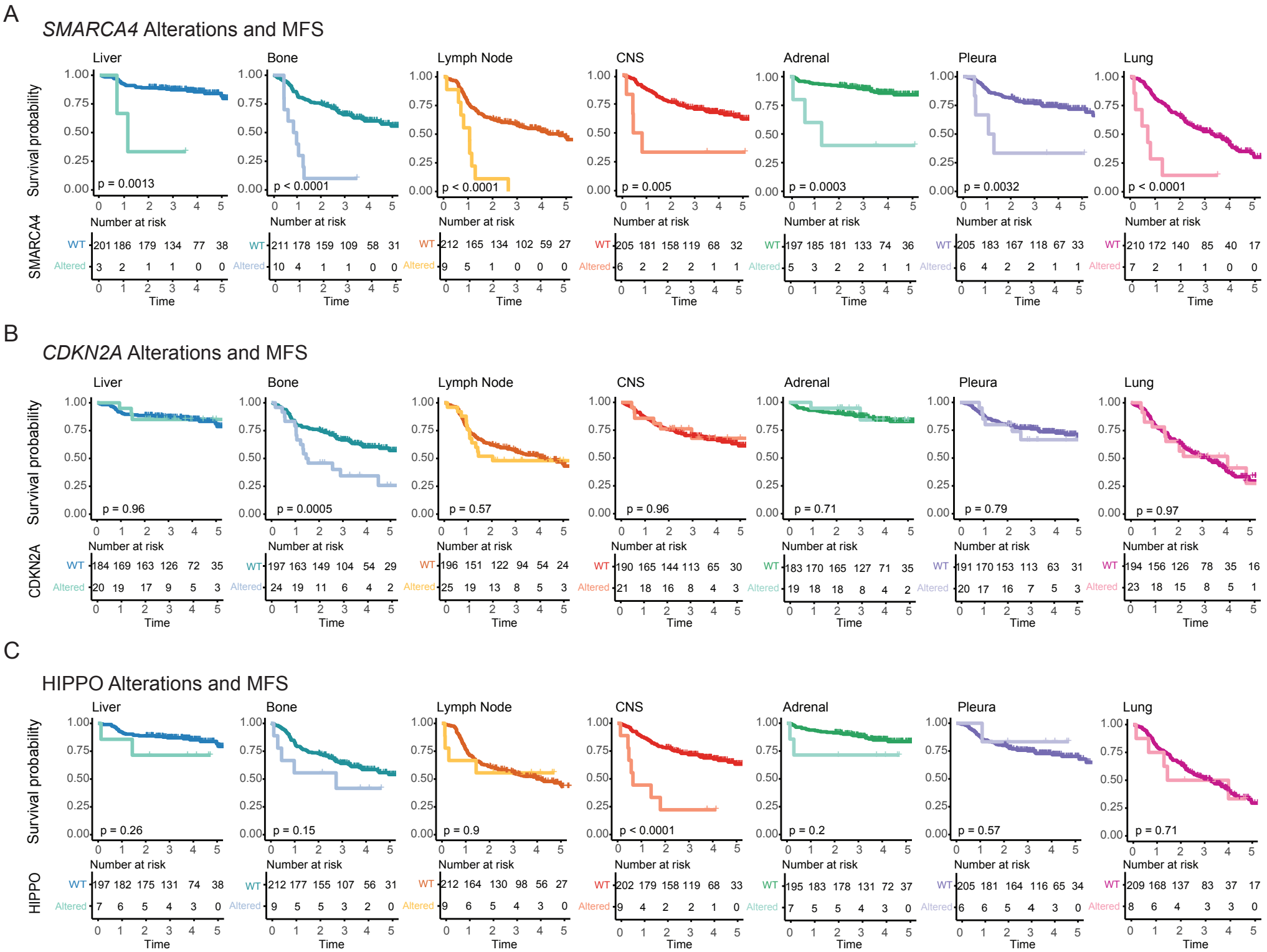
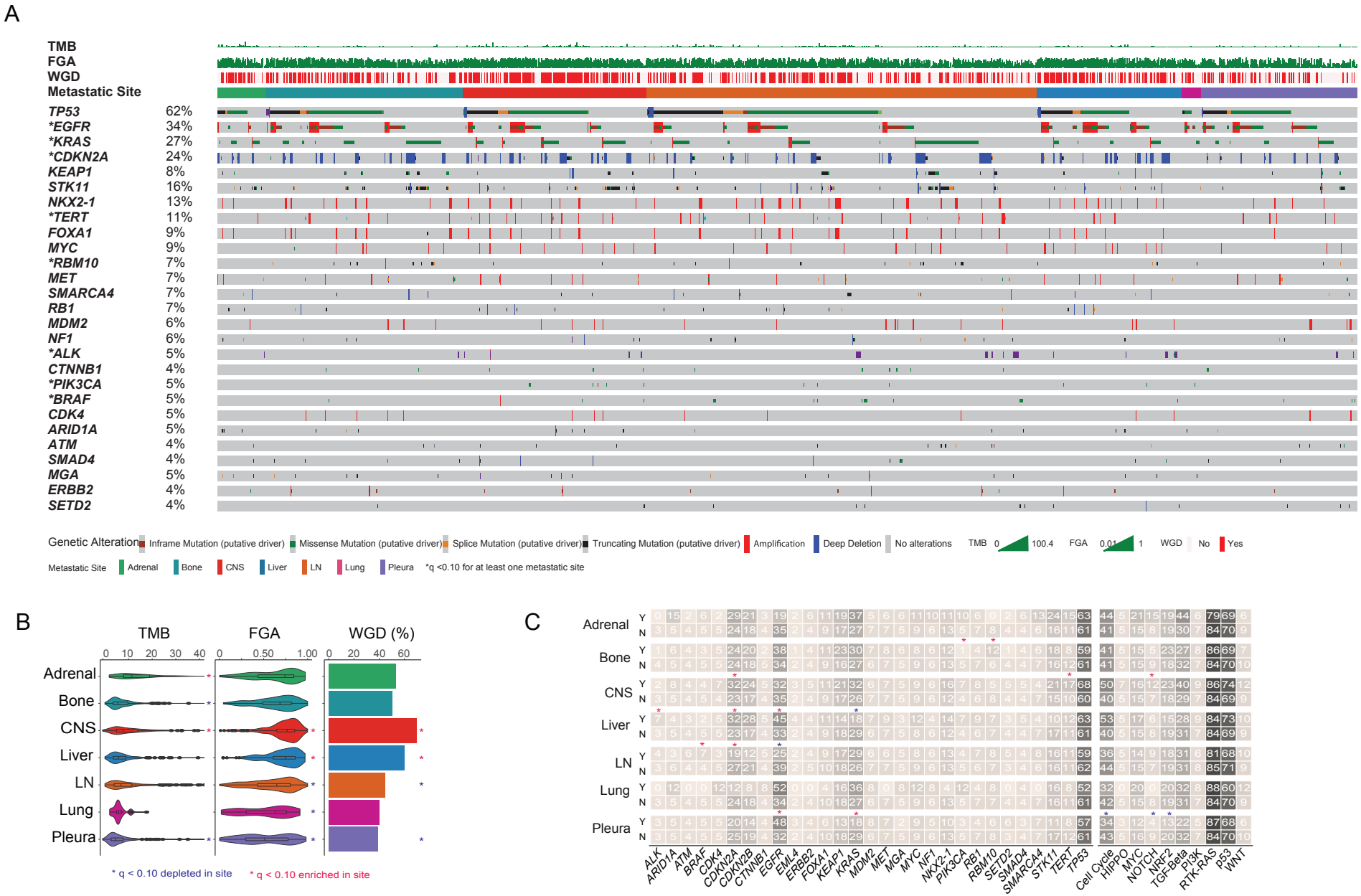

A
SMARCA4 Alterations and MFS
Liver
Bone
Lymph Node
CNS
Adrenal
Pleura
Lung
Survival probability
p = 0.0013
p < 0.0001
p < 0.0001
p = 0.005
p = 0.0003
p = 0.0032
p < 0.0001
SMARCA4
Number at risk
WT 201 186 179 134 77 38
Altered 3 2 1 1 0 0
WT 211 178 159 109 58 31
Altered 10 4 1 1 0 0
WT 212 165 134 102 59 27
Altered 9 5 1 0 0 0
WT 205 181 158 119 68 32
Altered 6 2 2 2 1 1
WT 197 185 181 133 74 36
Altered 5 3 2 2 1 1
WT 205 183 167 118 67 33
Altered 6 4 2 2 1 1
WT 210 172 140 85 40 17
Altered 7 2 1 1 0 0
Time
B
CDKN2A Alterations and MFS
Liver
Bone
Lymph Node
CNS
Adrenal
Pleura
Lung
Survival probability
p = 0.96
p = 0.0005
p = 0.57
p = 0.96
p = 0.71
p = 0.79
p = 0.97
CDKN2A
Number at risk
WT 184 169 163 126 72 35
Altered 20 19 17 9 5 3
WT 197 163 149 104 54 29
Altered 24 19 11 6 4 2
WT 196 151 122 94 54 24
Altered 25 19 13 8 5 3
WT 190 165 144 113 65 30
Altered 21 18 16 8 4 3
WT 183 170 165 127 71 35
Altered 19 18 18 8 4 2
WT 191 170 153 113 63 31
Altered 20 17 16 7 5 3
WT 194 156 126 78 35 16
Altered 23 18 15 8 5 1
Time
C
HIPPO Alterations and MFS
Liver
Bone
Lymph Node
CNS
Adrenal
Pleura
Lung
Survival probability
p = 0.26
p = 0.15
p = 0.9
p < 0.0001
p = 0.2
p = 0.57
p = 0.71
HIPPO
Number at risk
WT 197 182 175 131 74 38
Altered 7 6 5 4 3 0
WT 212 177 155 107 56 31
Altered 9 5 5 3 2 0
WT 212 164 130 98 56 27
Altered 9 6 5 4 3 0
WT 202 179 158 119 68 33
Altered 9 4 2 2 1 0
WT 195 183 178 131 72 37
Altered 7 5 5 4 3 0
WT 205 181 164 116 65 34
Altered 6 6 5 4 3 0
WT 209 168 137 83 37 17
Altered 8 6 4 3 3 0
Time

Figure S4. Kaplan-Meier Curves for Metastasis-Free Survival (MFS), Related to Figure 3.

(A) Kaplan-Meier curves for MFS for each site stratified by SMARCA4 status. (B) Kaplan-Meier curves for MFS for each site stratified by CDKN2A status. (C) Kaplan-Meier curves for MFS for each site stratified by Hippo pathway status. CNS, central nervous system; WT, wild-type.

Figure S5


A
TMB
FGA
WGD
Metastatic Site
TP53 62%
*EGFR 34%
*KRAS 27%
*CDKN2A 24%
KEAP1 8%
STK11 16%
NKX2-1 13%
*TERT 11%
FOXA1 9%
MYC 9%
*RBM10 7%
MET 7%
SMARCA4 7%
RB1 7%
MDM2 6%
NF1 6%
*ALK 5%
CTNNB1 4%
*PIK3CA 5%
*BRAF 5%
CDK4 5%
ARID1A 5%
ATM 4%
SMAD4 4%
MGA 5%
ERBB2 4%
SETD2 4%
Genetic Alteration
Inframe Mutation (putative driver)
Missense Mutation (putative driver)
Splice Mutation (putative driver)
Truncating Mutation (putative driver)
Amplification
Deep Deletion
No alterations
TMB 0 100.4
FGA 0.01 1
WGD No Yes
Metastatic Site
Adrenal
Bone
CNS
Liver
LN
Lung
Pleura
*q <0.10 for at least one metastatic site

B
TMB
FGA
WGD (%)
Adrenal
Bone
CNS
Liver
LN
Lung
Pleura
* q < 0.10 depleted in site
* q < 0.10 enriched in site

C
Adrenal
Bone
CNS
Liver
LN
Lung
Pleura
ALK
ARID1A
ATM
BRAF
CDK4
CDKN2A
CDKN2B
CTNNB1
EGFR
EML4
ERBB2
FOXA1
KEAP1
KRAS
MDM2
MET
MGA
MYC
NF1
NKX2-1
PIK3CA
RB1
RBM10
SETD2
SMAD4
SMARCA4
STK11
TERT
TP53
Cell Cycle
HIPPO
MYC
NOTCH
NRF2
TGF-Beta
PI3K
RTK-RAS
p53
WNT

Figure S5. Genomic comparisons between metastatic lesions from different sites, Related to Figure 4. (A) Oncoprint of metastatic lesions (ML) stratified by anatomic site displaying genomic features and all genes altered in ≥3% of cohort. *q<0.10 for at least one metastatic site. (B) Violin plots of the distribution of tumor mutational burden (TMB) and fraction of genome altered (FGA) for seven metastatic sites. Boxplots display median values, interquartile range (IQR) boxes, and whiskers demonstrating 1.5 x IQR. *Right*: Bar plots showing frequency of whole genome duplication (WGD) across each site. (C) Heat map listing percentage of samples with a given gene or pathway alteration, stratified by metastatic site. *q<0.10, red * indicates alteration enriched in metastatic site, blue * indicates alteration depleted in metastatic site. Y, yes - metastatic lesions from indicated anatomic site; N, no - metastatic lesions from all other anatomic sites. Statistical analyses: (A, C) Fisher's exact test comparing site of interest to all other sites. *q*-values correct for multiple comparisons using the false-discovery rate (FDR). **q*<0.10. (B) Wilcoxon rank-sum test for TMB and FGA; Fisher's exact test for WGD. CNS, central nervous system; LN, lymph node.

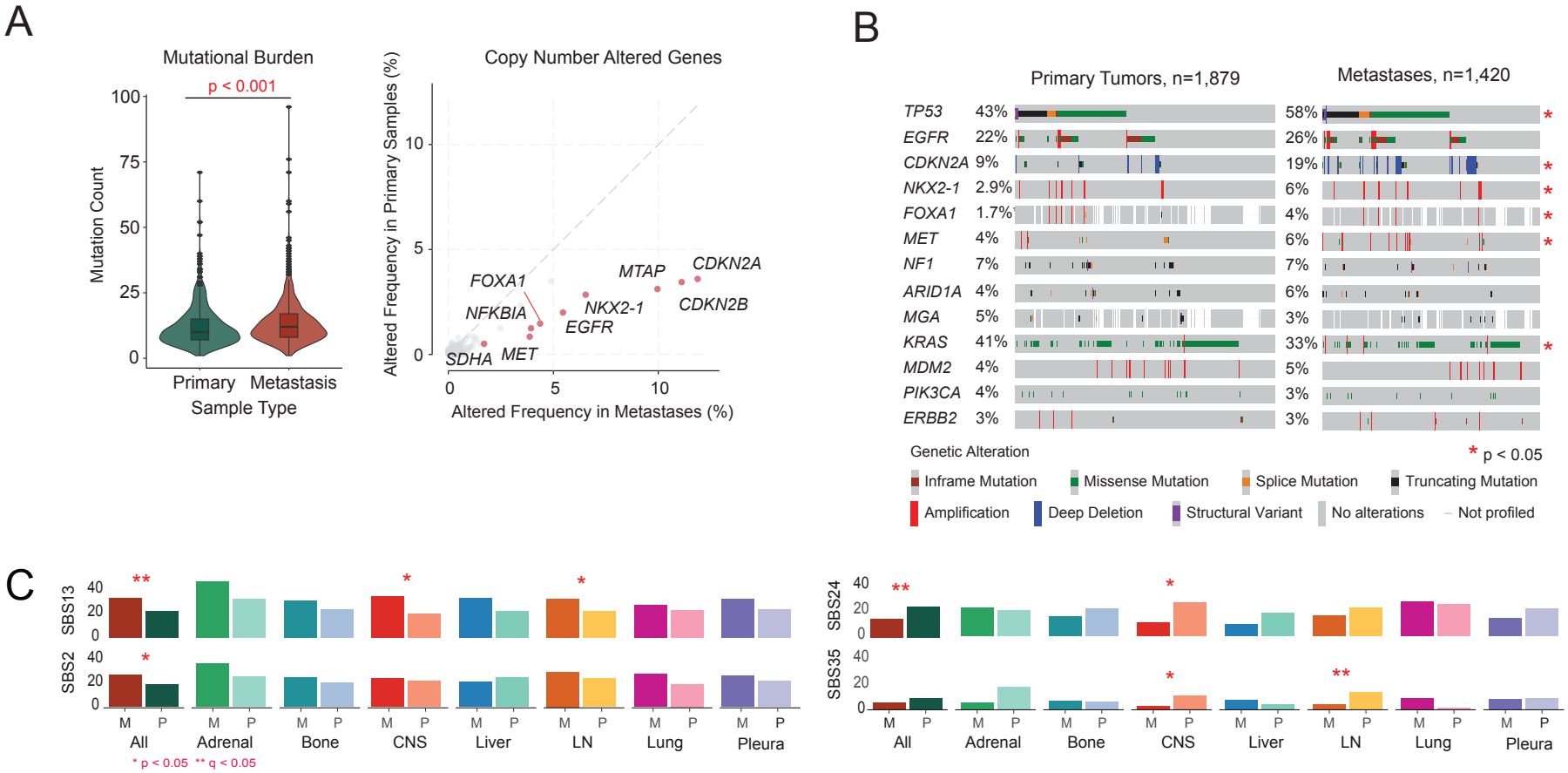

A
Mutational Burden
p < 0.001
Mutation Count
100
75
50
25
0
Primary
Metastasis
Sample Type
Copy Number Altered Genes
Altered Frequency in Primary Samples (%)
Altered Frequency in Metastases (%)
FOXA1
NFKBIA
SDHA
MET
EGFR
NKX2-1
MTAP
CDKN2A
CDKN2B
B
Primary Tumors, n=1,879
Metastases, n=1,420
TP53 43% 58%
EGFR 22% 26%
CDKN2A 9% 19%
NKX2-1 2.9% 6%
FOXA1 1.7% 4%
MET 4% 6%
NF1 7% 7%
ARID1A 4% 6%
MGA 5% 3%
KRAS 41% 33%
MDM2 4% 5%
PIK3CA 4% 3%
ERBB2 3% 3%
Genetic Alteration
* p < 0.05
Inframe Mutation
Missense Mutation
Splice Mutation
Truncating Mutation
Amplification
Deep Deletion
Structural Variant
No alterations
Not profiled
C
SBS13
SBS2
SBS24
SBS35
M P
All
Adrenal
Bone
CNS
Liver
LN
Lung
Pleura
* p < 0.05 ** q < 0.05

Figure S6

Figure S6. Comparison of alterations between primaries and metastases in GENIE validation cohort and mutation signatures for MSK-IMPACT samples, Related to Figure 4.

(A) GENIE validation cohort. *Left*: Violin plot showing differences in mutational burden between primary and metastatic samples. Boxplots display median values, interquartile range (IQR) boxes, and whiskers demonstrating 1.5 x IQR. *Right*: Dot plot comparing differences in gene-level copy number alterations called between primary and metastases. (B) Oncoprint of primary and metastatic tumors in GENIE validation cohort. (C) Frequencies of IMPACT samples with APOBEC-related signatures present according to metastatic site. Statistical analyses: (A) Wilcoxon rank-sum test. (B-C) Fisher's exact test. *q*-values account for multiple tests through false-discover rate (FDR) correction.

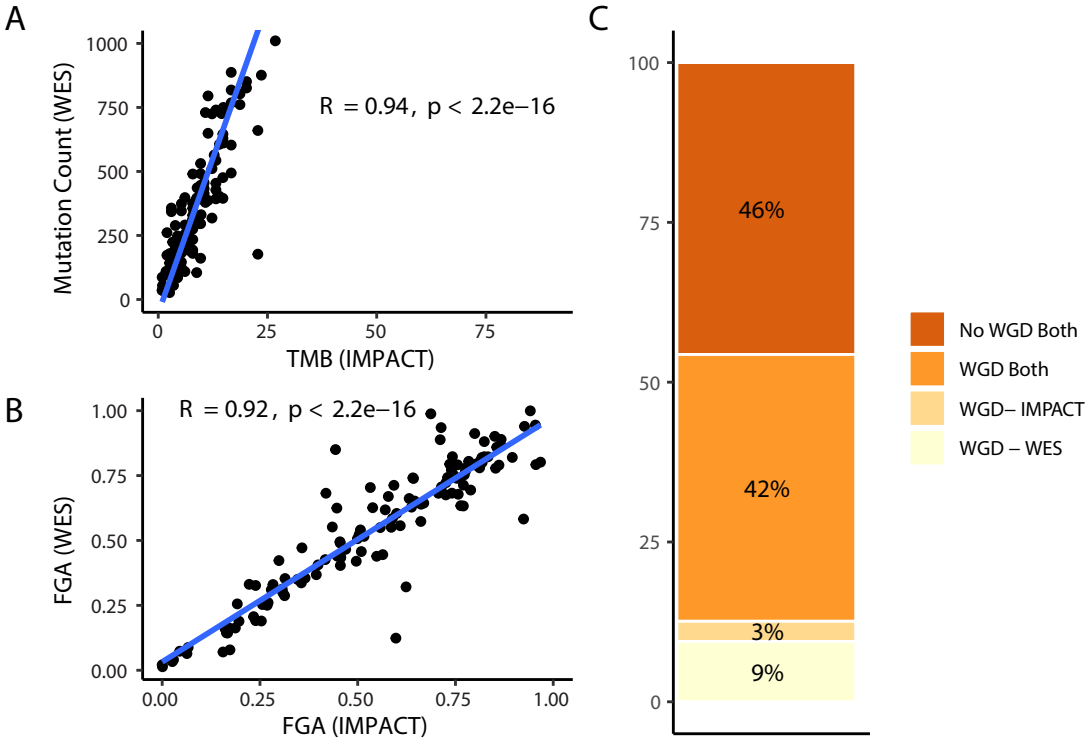
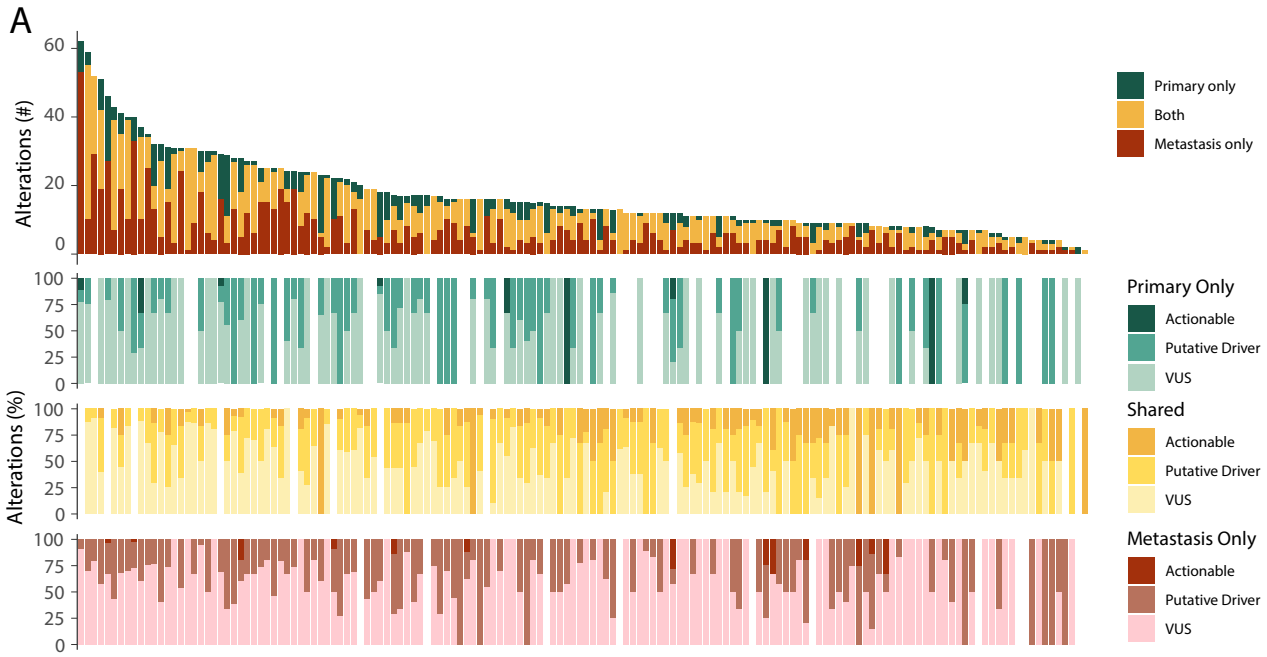
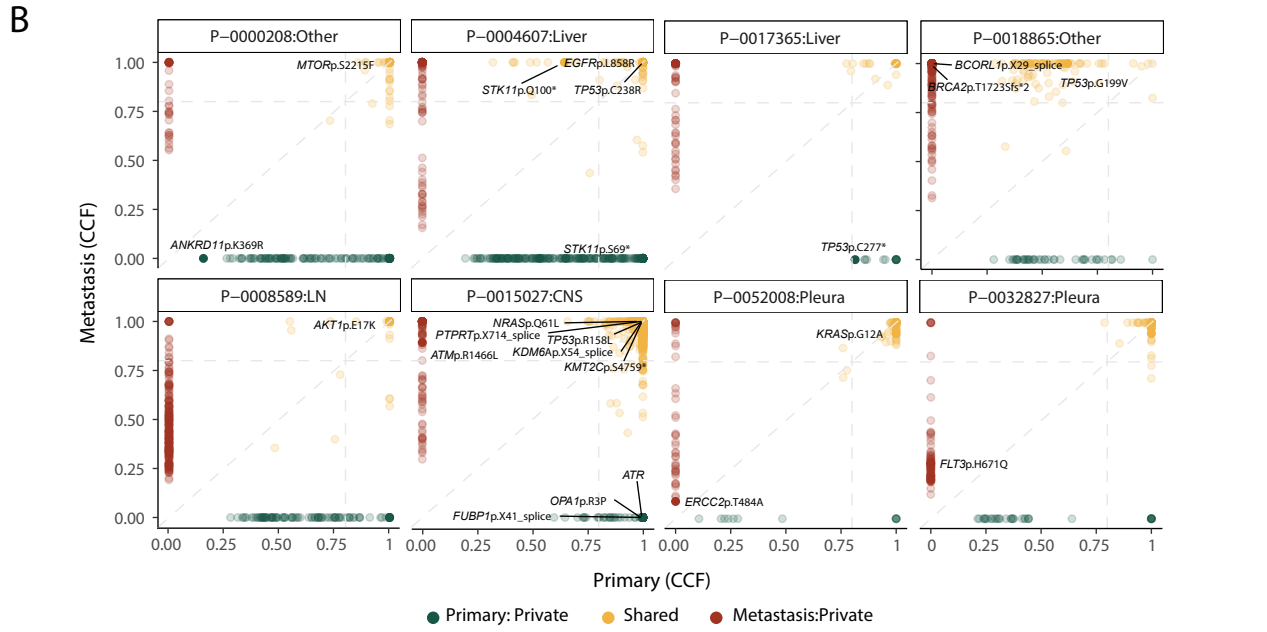

A
Mutation Count (WES)
1000
750
500
250
0
0
25
50
75
TMB (IMPACT)
R = 0.94, p < 2.2e−16
B
FGA (WES)
1.00
0.75
0.50
0.25
0.00
0.00
0.25
0.50
0.75
1.00
FGA (IMPACT)
R = 0.92, p < 2.2e−16
C
100
75
50
25
0
46%
42%
3%
9%
No WGD Both
WGD Both
WGD– IMPACT
WGD – WES

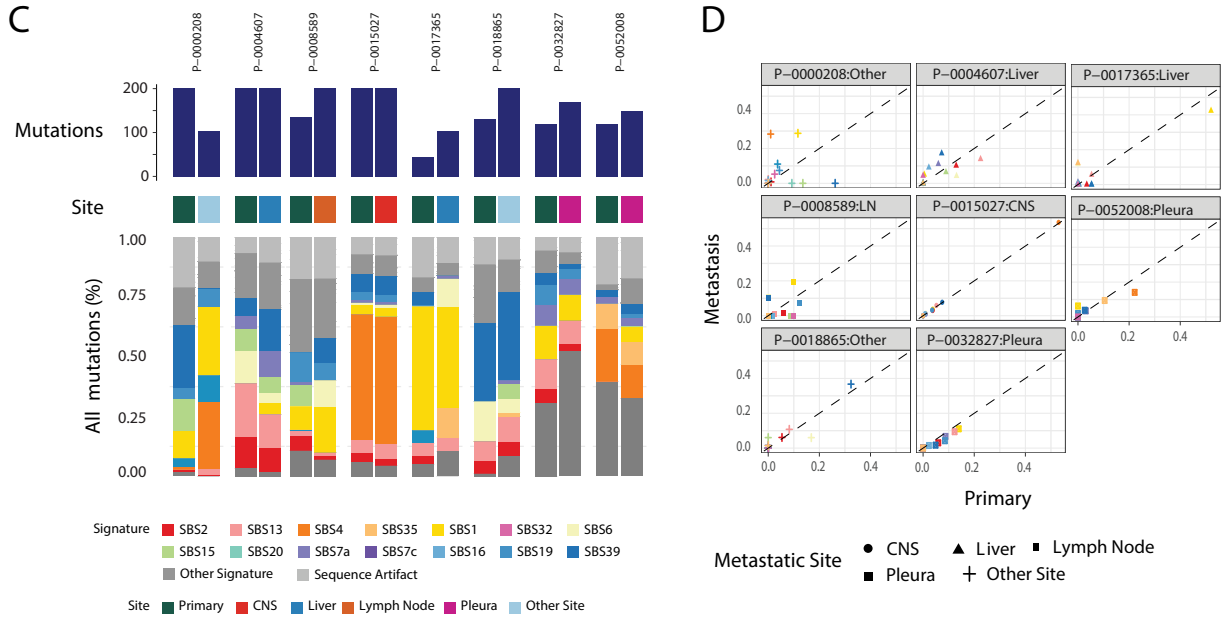
Figure S7

Figure S7. Comparison of sample-level genomic features for samples sequenced with MSK-IMPACT and whole-exome sequencing (WES), Related to Figure 4. (A) Scatter plot highlighting correlation between tumor mutational burden (TMB) reported by MSK-IMPACT (x-axis) and number of mutations detected by WES (y-axis). (B) Fraction of genome altered (FGA) estimated from MSK-IMPACT (x-axis) versus FGA estimated from WES (y-axis). (C) Percentage of samples with similar or differing calls for whole-genome duplication (WGD) based on data from MSK-IMPACT and WES. In our cohort, 121 samples were sequenced by next-generation sequencing (MSK-IMPACT) and WES. We observed agreement between the two methods in ~88% of whole-genome duplication calls. Statistical analyses: (A-B) Pearson correlation with correlation coefficients and p-values as indicated.

Figure S8


A
Alterations (#)
Primary only
Both
Metastasis only
Alterations (%)
Primary Only
Actionable
Putative Driver
VUS
Shared
Actionable
Putative Driver
VUS
Metastasis Only
Actionable
Putative Driver
VUS


B
P–0000208:Other
P–0004607:Liver
P–0017365:Liver
P–0018865:Other
P–0008589:LN
P–0015027:CNS
P–0052008:Pleura
P–0032827:Pleura
Metastasis (CCF)
Primary (CCF)
Primary: Private
Shared
Metastasis:Private


C
Mutations
Site
All mutations (%)
Signature
SBS2
SBS13
SBS4
SBS35
SBS1
SBS32
SBS6
SBS15
SBS20
SBS7a
SBS7c
SBS16
SBS19
SBS39
Other Signature
Sequence Artifact
Site
Primary
CNS
Liver
Lymph Node
Pleura
Other Site

D
Metastasis
Primary
Metastatic Site
CNS
Liver
Lymph Node
Pleura
Other Site

Figure S8. Patient-level comparisons in paired primary and metastatic samples in IMPACT samples and whole-exome sequencing (WES), Related to Figure 5. (A) *Top:* Bar plot illustrating fraction of alterations shared and private between matched primary and metastases. *Bottom:* Bar plots showing the distribution of actionable, oncogenic, and variants of unknown significance (VUS) separated by alterations private to primary, shared alterations, and alterations private to metastasis. (B) Comparison of cancer cell fraction (CCF) for individual mutations in paired WES primary and metastatic samples. Each point represents an individual mutation. Mutations considered oncogenic per OncoKB are highlighted by displaying name of their associated gene. (C) Comparison of mutational signatures for paired WES primary and metastatic samples. (D) Quantification of differences in selected mutational signatures from panel C. CNS, central nervous system.

Table S8. Overview of cohorts used in specific analyses, Related to STAR Methods

	Cohort					
Analysis	**NM**	**ES-M**	**LS-M**	**ML**	**MP-M**	**WES**
1. Features associated with metastasis in primary tumors	X	X	X			
2. Features associated with MFS in surgical patients	X	X				
3. Patterns of metastasis and metastatic burden		X	X			
4. Features associated with site-specific metastasis		X	X			
5. Time-to-event site-specific metastasis		X				
6. Comparisons of metastases across sites				X		
7. Analysis of mutational signatures	X	X	X	X		X
8. Comparisons of unmatched metastases and primary tumors		X	X	X		
9. Comparisons of matched primary tumors and metastases					X	X

ES-M, early-stage metastatic; LS-M, late-stage metastatic; MFS, metastasis-free survival; ML, metastatic lesion; MP-M, matched primary-metastasis; NM, nonmetastatic; WES, whole-exome sequencing.