Supplementary Figures

Figure S1

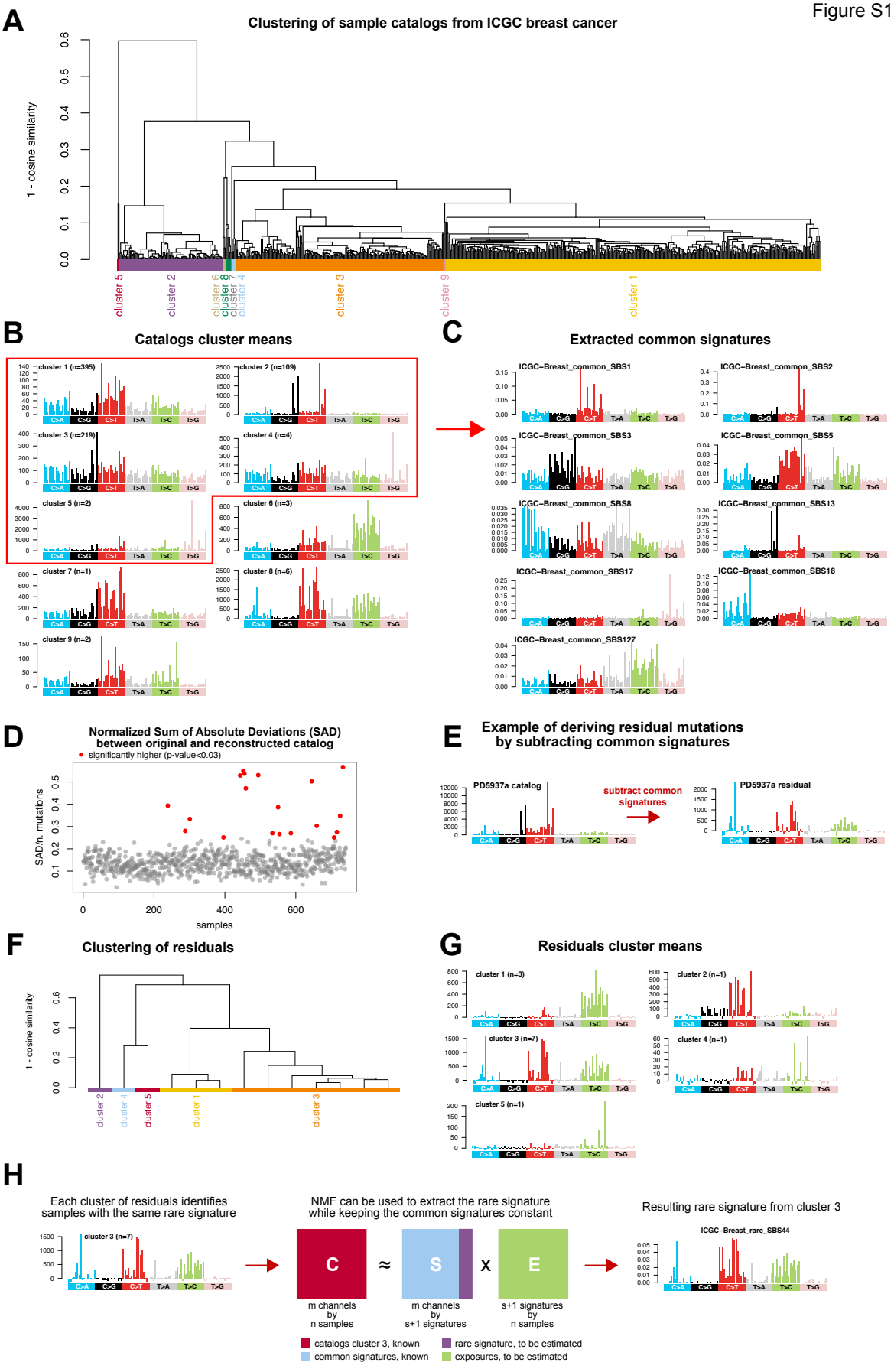

A
Clustering of sample catalogs from ICGC breast cancer
1 - cosine similarity
cluster 5
cluster 2
cluster 6
cluster 8
cluster 7
cluster 4
cluster 3
cluster 9
cluster 1
B
Catalogs cluster means
cluster 1 (n=395)
cluster 2 (n=109)
cluster 3 (n=219)
cluster 4 (n=4)
cluster 5 (n=2)
cluster 6 (n=3)
cluster 7 (n=1)
cluster 8 (n=6)
cluster 9 (n=2)
C
Extracted common signatures
ICGC–Breast_common_SBS1
ICGC–Breast_common_SBS2
ICGC–Breast_common_SBS3
ICGC–Breast_common_SBS5
ICGC–Breast_common_SBS8
ICGC–Breast_common_SBS13
ICGC–Breast_common_SBS17
ICGC–Breast_common_SBS18
ICGC–Breast_common_SBS127
D
Normalized Sum of Absolute Deviations (SAD)
between original and reconstructed catalog
significantly higher (p-value<0.03)
SAD/n. mutations
samples
E
Example of deriving residual mutations
by subtracting common signatures
PD5937a catalog
subtract common signatures
PD5937a residual
F
Clustering of residuals
1 - cosine similarity
cluster 2
cluster 4
cluster 5
cluster 1
cluster 3
G
Residuals cluster means
cluster 1 (n=3)
cluster 2 (n=1)
cluster 3 (n=7)
cluster 4 (n=1)
cluster 5 (n=1)
H
Each cluster of residuals identifies
samples with the same rare signature
cluster 3 (n=7)
NMF can be used to extract the rare signature
while keeping the common signatures constant
C
≈
S
X
E
m channels
by
n samples
m channels
by
s+1 signatures
s+1 signatures
by
n samples
Resulting rare signature from cluster 3
ICGC–Breast_rare_SBS44
catalogs cluster 3, known
rare signature, to be estimated
common signatures, known
exposures, to be estimated

Fig. S1. Extraction of common and rare signatures in the ICGC Breast cancer cohort, as an example of the methodological steps. (**A**) Mutational catalogs from the ICGC-Breast cohort were clustered using hierarchical clustering with average linkage and 1-cosine similarity as distance. (**B**) The averaged profiles of the catalogs in each of the nine clusters identified in (A), with n indicating the number of catalogs in each cluster. The red box highlights the clusters that were used to extract common signatures. These include the three largest clusters and two smaller clusters that appear to contain the well-reported SBS17. Samples in clusters 6-9 were excluded from the extraction of common signatures. (**C**) Common signatures obtained using our extraction framework from the samples highlighted in (B). (**D**) Distribution of the error between the original catalogs and the reconstructed catalogs modelled using the extracted common signatures. The error is given by the sum of absolute deviations between the original and reconstructed catalogs, divided by the number of mutations in the samples. The error was modelled using a normal distribution and red dots show significantly greater error values (p-value less than 0.03). (**E**) Residual mutations of example sample PD5937a, obtained as the difference between the original catalog and a reconstructed catalog modelled using the common signatures. The exposures of the reconstructed catalog were obtained using non-negative least squares with the additional constraint that the residual should be mostly positive (Materials and Methods). (**F**) Residuals are clustered using hierarchical clustering with average linkage and 1-cosine similarity as distance. (**G**) Average pattern of residual clusters in (F). (**H**) For each cluster of residuals in (G), a rare signature was extracted using a variant of NMF where a signature can be estimated while holding the common signatures as constant.

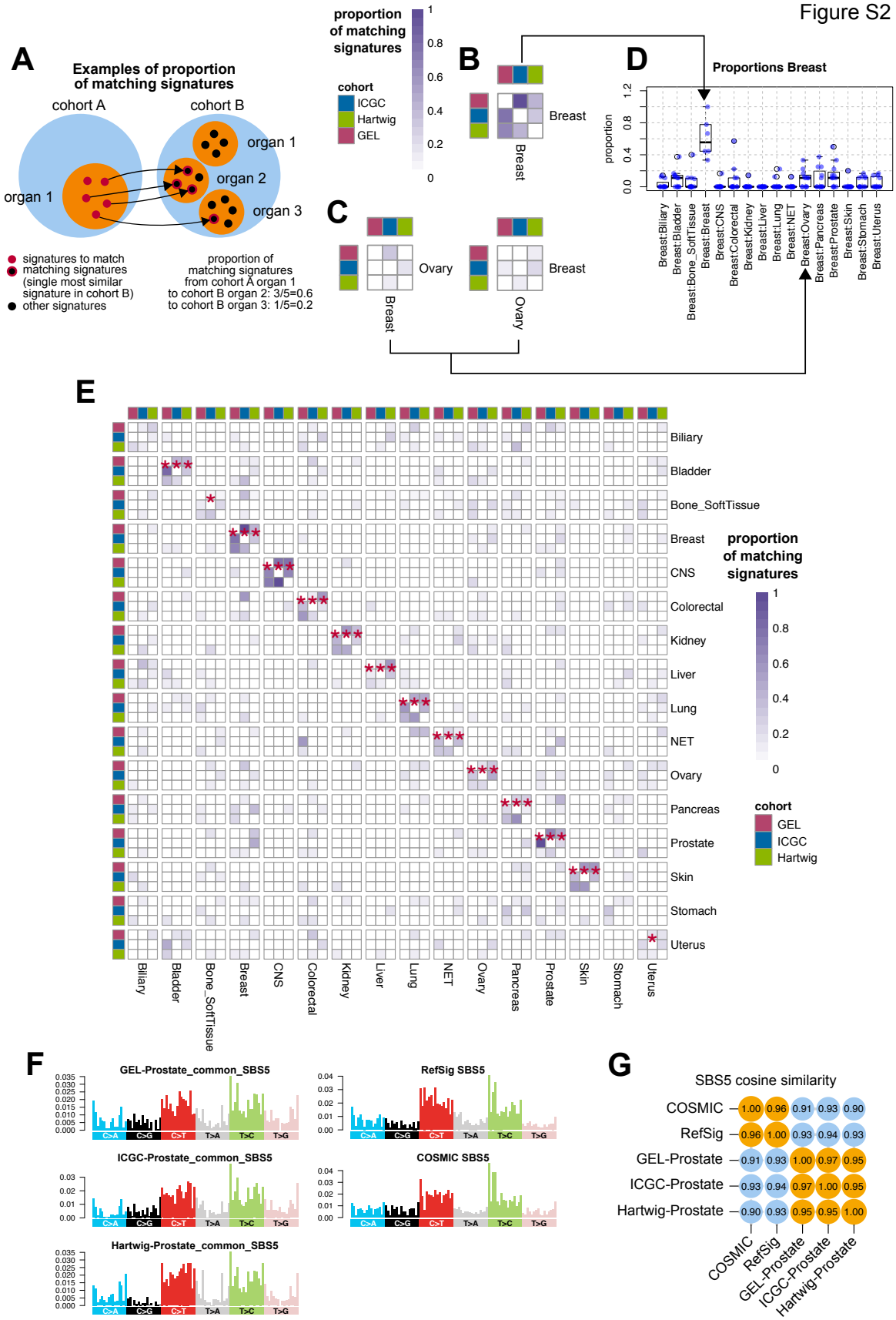


Fig. S2. Evidence of organ-specificity of signatures across three cohorts. (**A**) Example of how to seek 'proportion of matching signatures'. For each common organ-specific signature in

an organ (organ 1) in a first cohort (cohort A), the single most similar common organ-specific signature in the second cohort (cohort B) is sought. The proportion of matching signatures is the proportion of signatures in the organ in the first cohort that have the most similar signatures in a given organ in the second cohort. For example, in panel (A), the proportion of matching signatures from cohort A organ 1 to cohort B organ 2 is 3/5=0.6, while from cohort A organ 1 to cohort B organ 3 is 1/5=0.2. Some signatures may be unmatched if there is no signature in cohort B with a cosine similarity of at least 0.85. (**B**) Example of a comparison of signatures between the same tissue-type (breast vs breast) where there are six possible comparison outcomes (GEL-ICGC, GEL-Hartwig, ICGC-Hartwig, ICGC-GEL, Hartwig-GEL, Hartwig-ICGC), and thus six values of proportion of matching signatures. (**C**) Example of a comparison between different organs (breast vs ovary), where there are 12 possible comparisons. (**D**) After considering all possible cohort and organ combinations, a Tukey test is used to determine if the same organ comparison exhibits the greatest similarity than each of the other 15 inter-organ comparisons (confidence level 0.95, p-value threshold 0.05). (**E**) Agnostic three-way comparison of the common organ-specific signatures across all three cohorts in 16 organs. We use *** to indicate that the proportion of matching signatures for the same organ comparison (for example Breast to Breast) is significantly higher than all 15 other organ comparisons, and * higher than 7, according to the Tukey test in (D). (**F**) Prostate-specific SBS5 signatures from the 3 three cohorts are shown next to Reference Signature (RefSig) SBS5 and COSMIC SBS5. RefSig SBS5 is obtained as an average of cohort-organ signatures, see text. (**G**) Cosine similarity of signatures shown in (F). Colors are used to highlight groups of signatures, with orange used if the cosine similarity is greater than or equal to 0.95.

Figure S3

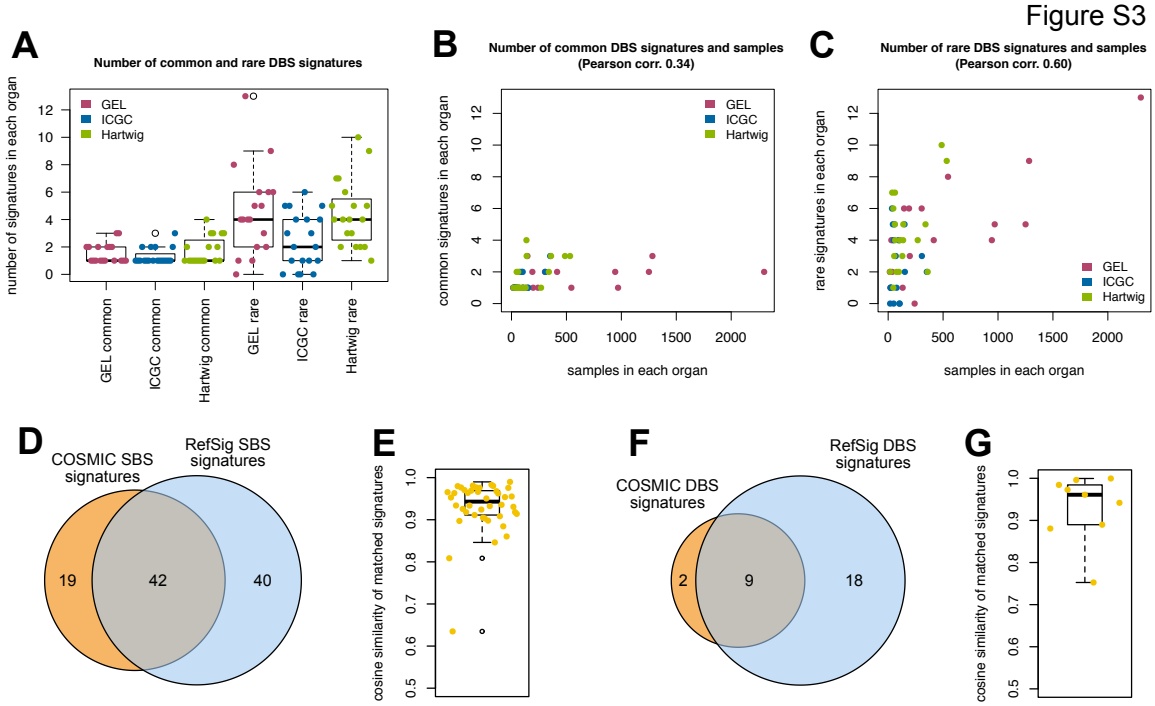


Fig S3. Identifying organ-specific and reference signatures and comparison to literature. (**A**) Number of common and rare DBS signatures in each cohort. (**B**) Common DBS signatures as a function of number of samples analyzed. (**C**) Rare DBS signatures as a function of number of samples analyzed. (**D**) Venn diagram comparing SBS COSMIC signatures version 3.2 and the SBS reference signatures identified in this study. (**E**) Cosine similarity of the 42 SBS signatures in the Venn diagram intersection in (D). (**F**) Venn diagram comparing DBS COSMIC signatures version 3.2 and the DBS reference signatures identified in this study. (**G**) Cosine similarity of the 9 DBSs in the Venn diagram intersection in (F).

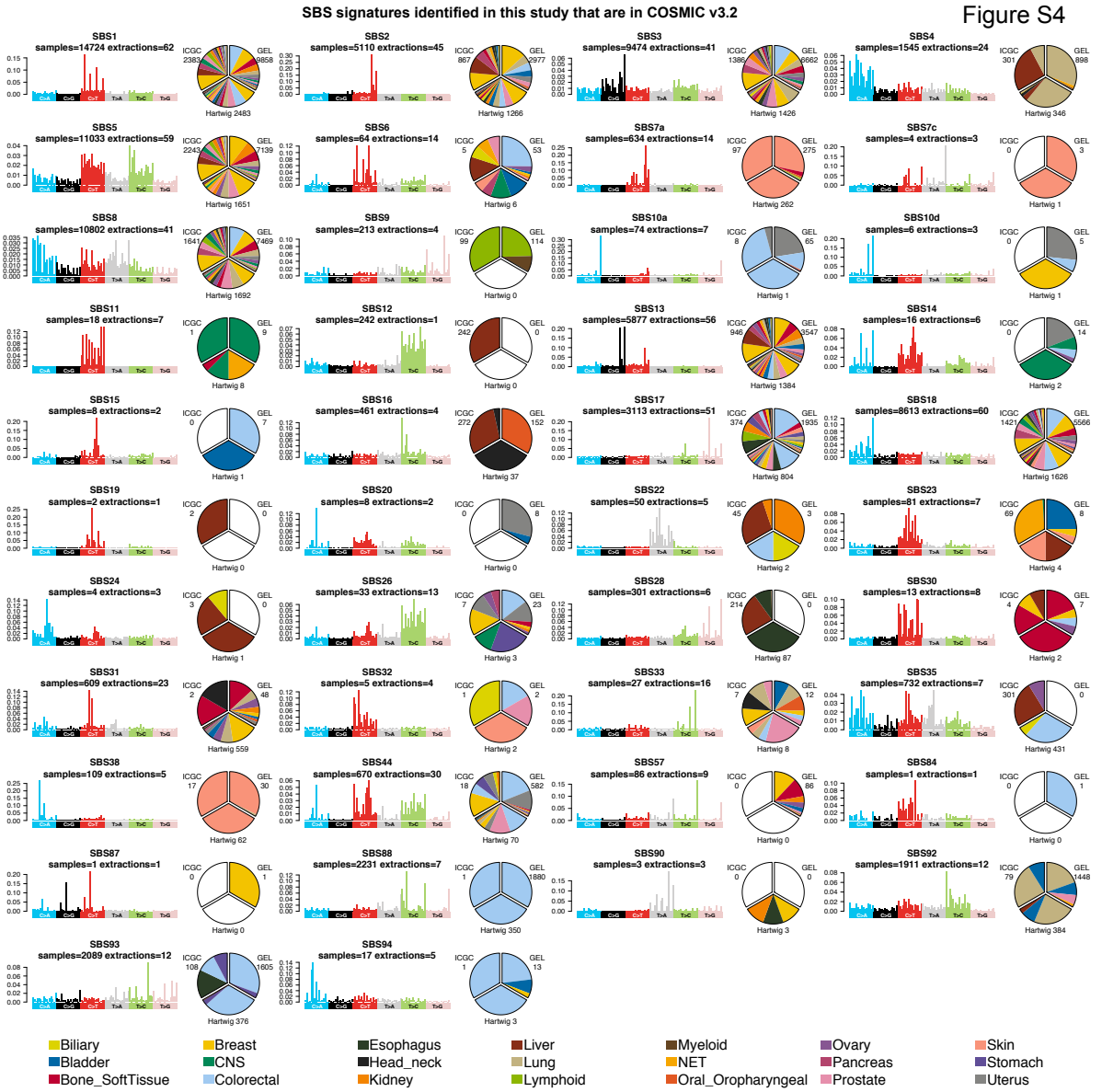


Fig. S4. SBS reference signatures that have been previously reported (COSMIC v3.2). Previously reported SBS reference signatures identified in this study are depicted, where the corresponding pie-charts report the prevalence of the respective signatures in the samples across all tumor types, in the three cohorts (GEL, ICGC and Hartwig). “Samples” indicates total number of samples across the three cohorts with the signature. “Extractions” indicates the number of independent extractions that the signature was found in (note that extractions are performed for every organ independently in the three cohorts).

Figure S5

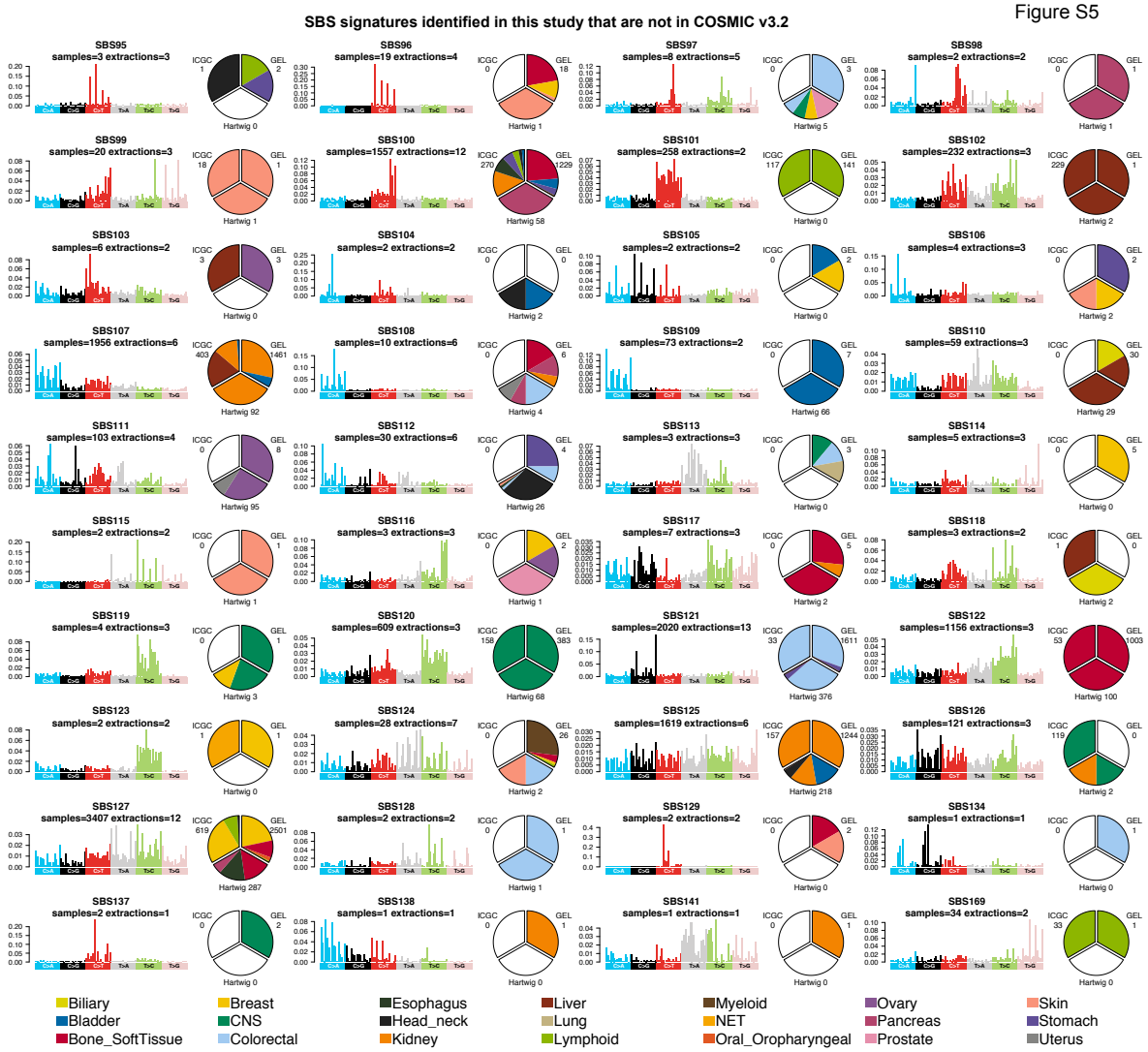


Fig. S5. Previously unreported SBS reference signatures identified in this study. Previously unreported SBS reference signatures identified in this study and their distribution in the samples across cohorts and organs. “Samples” indicates total number of samples across the three cohorts with the signature. “Extractions” indicates the number of independent extractions that the signature was found in (note that extractions are performed for every organ of three cohorts).

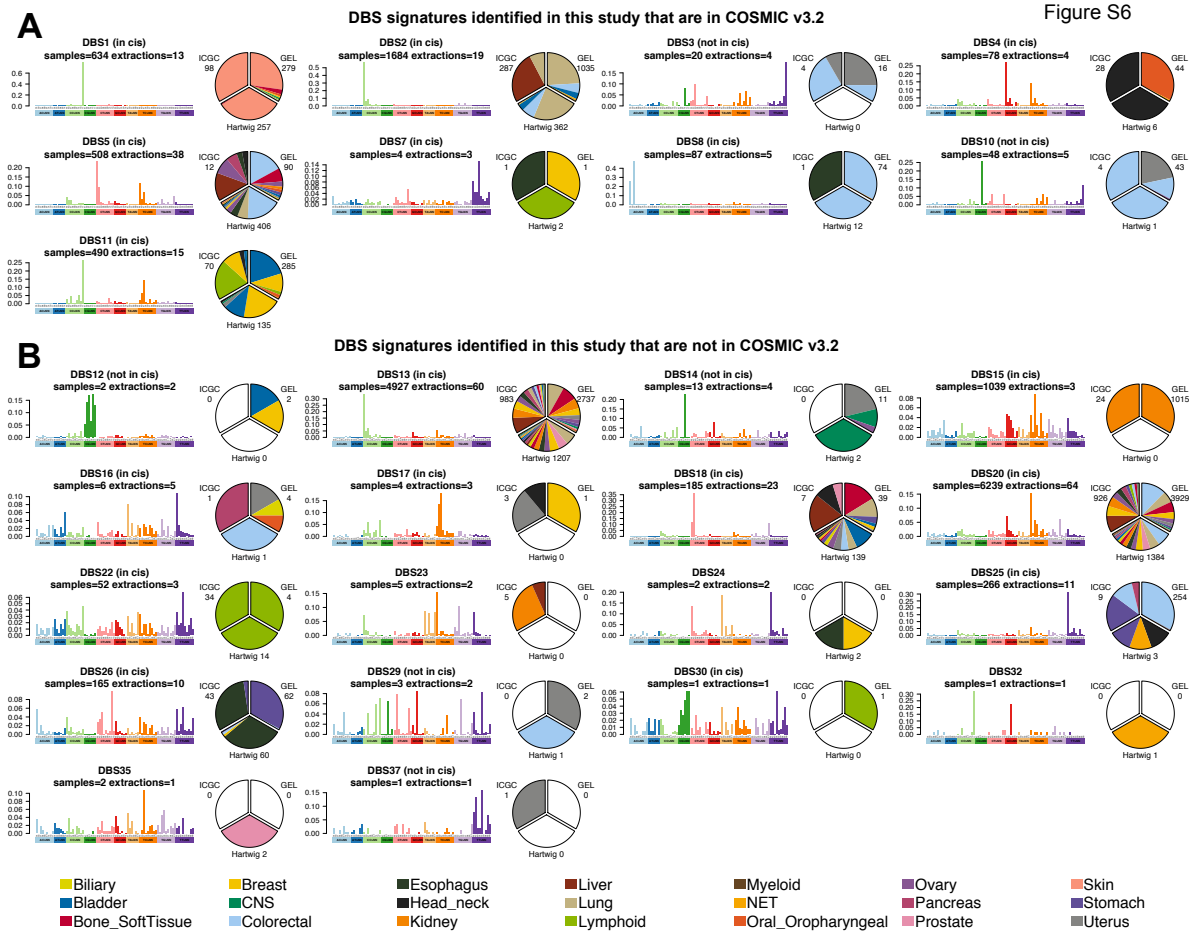


Fig. S6. DBS reference signatures. Previously known (COSMIC v3.2) (**A**) and previously unreported (**B**) DBS signatures identified in this study and their distribution across cohorts and organs. "Samples" indicates total number of samples across the three cohorts with the signature. "Extractions" indicates the number of independent extractions that the signature was found in (note that extractions are performed for every organ of three cohorts). Where alignment data were available, next to the name we indicate whether the double substitutions were observed to be in *cis* or otherwise, the latter would suggest that the DBS signature is a false positive signature.

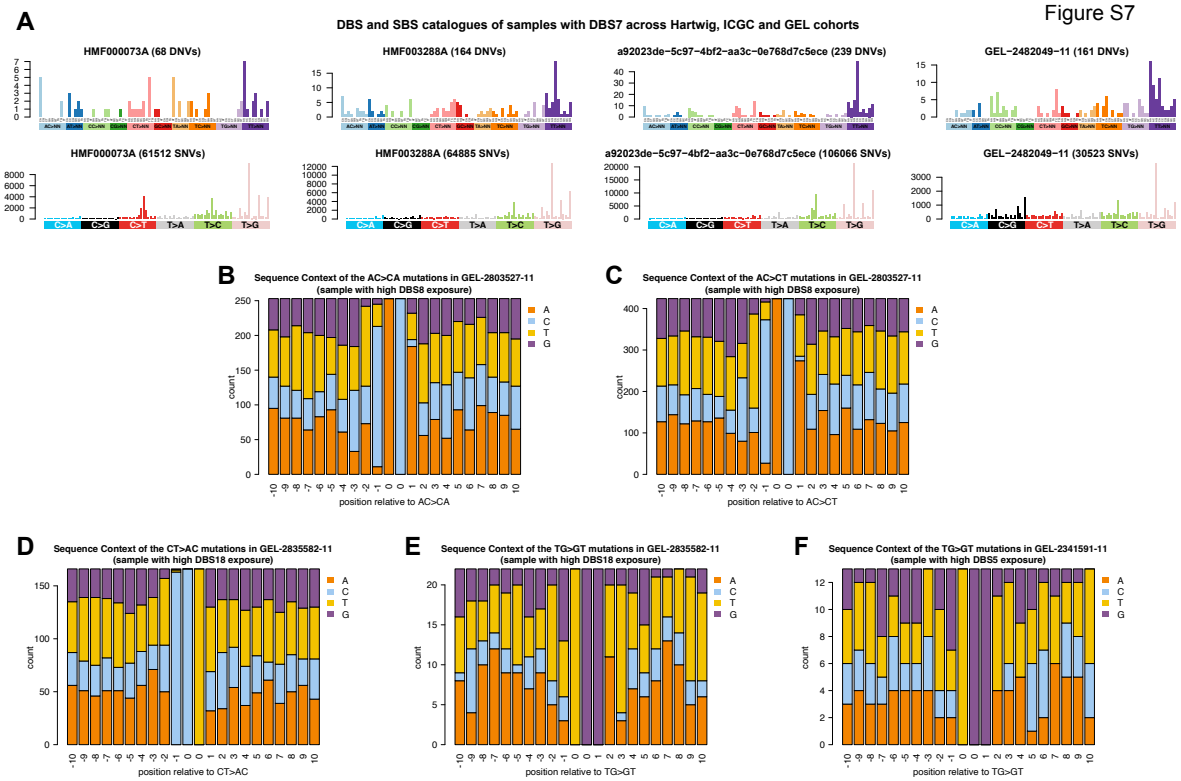


Fig. S7. Exploring flanking sequence context for DBS signatures. (**A**) DBS and SBS catalogs of four samples with DBS7. (**B**) Flanking sequence context of AC>CA dinucleotide variants from a sample with high exposure of DBS8, indicating a preference for a CpACpA context. (**C**) Similar to (B), considering AC>CT variants. (**D**) Flanking sequence context of dinucleotide variants from samples with high exposure of DBS18, indicating a preference for a CpCTpN context in the case of CT>AC mutations. (**E**) Similar to (D), considering TG>GT mutations in a DBS18 sample, indicating a preference for a NpTGpG context. (**F**) Similar to (E), TG>GT mutations in a DBS5 sample, also indicating a preference for a NpTGpG context.

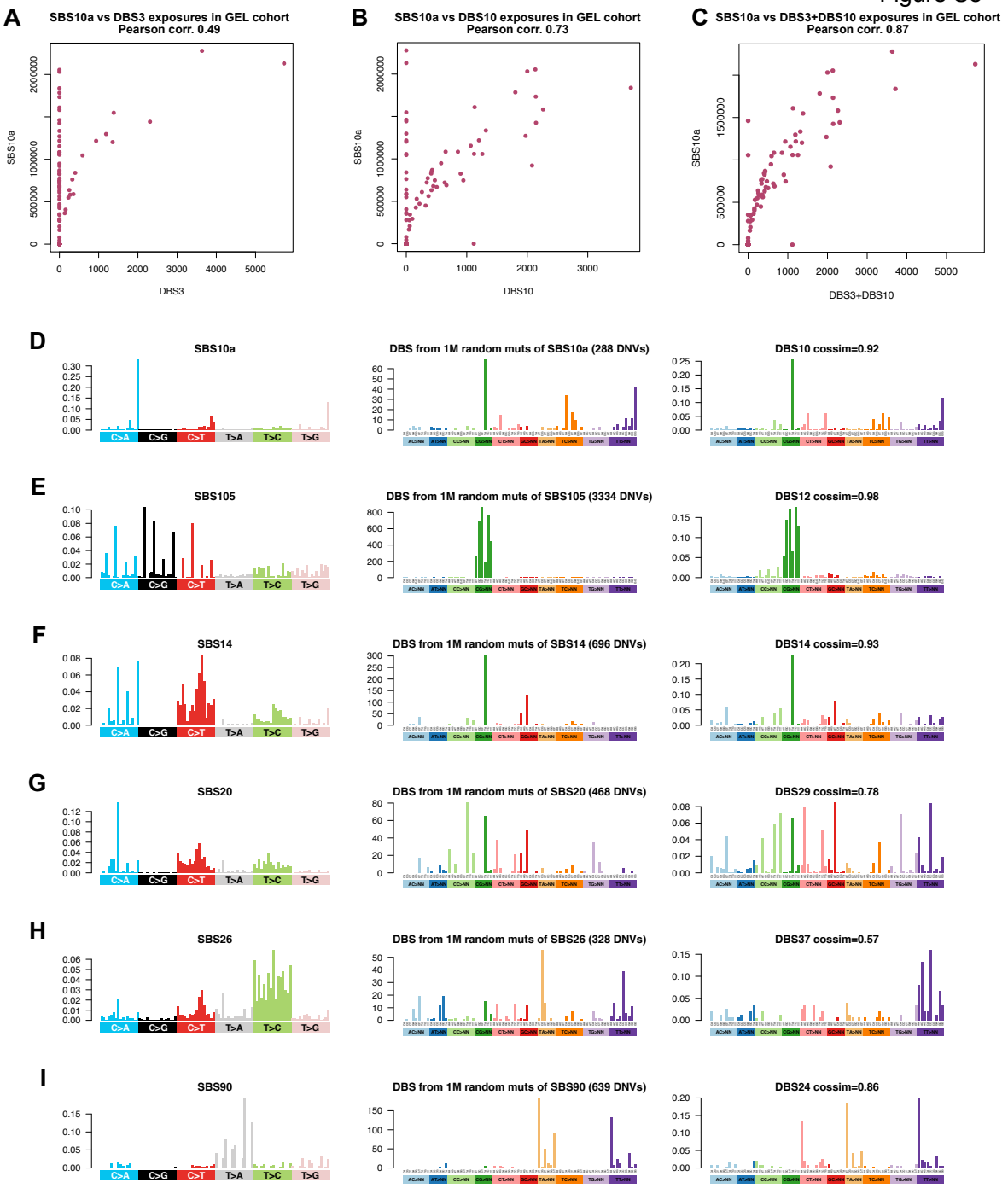


Fig. S8. Relationships between SBS and DBS signatures: correlations and simulations of expected patterns. (**A**) Correlation between the exposures of SBS10a and DBS3 across all samples in the GEL cohort. **(B)** Correlation between the exposures of SBS10a and DBS10 across all samples in the GEL cohort. **(C)** Correlation between the exposures of SBS10a and the sum of DBS3 and DBS10 across all samples in the GEL cohort. The highest correlation is obtained when the exposures of SBS10a are compared to the sum of the exposures of DBS3 and DBS10. (**D**) Simulation study to determine whether SBS10a can produce DBS signatures due to random adjacent single nucleotide variants. One million mutations have been randomly sampled and placed across the genome according the probability defined by the SBS10a pattern (signature on the left). Variants that are adjacent are then considered as double substitutions and a DBS mutational catalog is generated accordingly (signature at the center). The most

similar DBS signature is shown on the right, DBS10, displaying the cosine similarity between the DBS signatures and the sampled profile. **(E)** Similar to (D), sampling SBS105 and DBS12 as the most similar DBS signature. **(F)** Similar to (D), sampling SBS14 and DBS14 as the most similar DBS signature. **(G)** Similar to (D), sampling SBS20 and DBS29 as the most similar DBS signature. **(H)** Similar to (D), sampling SBS26 and DBS37 as the most similar DBS signature. **(I)** Similar to (D), sampling SBS90 and DBS24 as the most similar DBS signature.

Figure S9

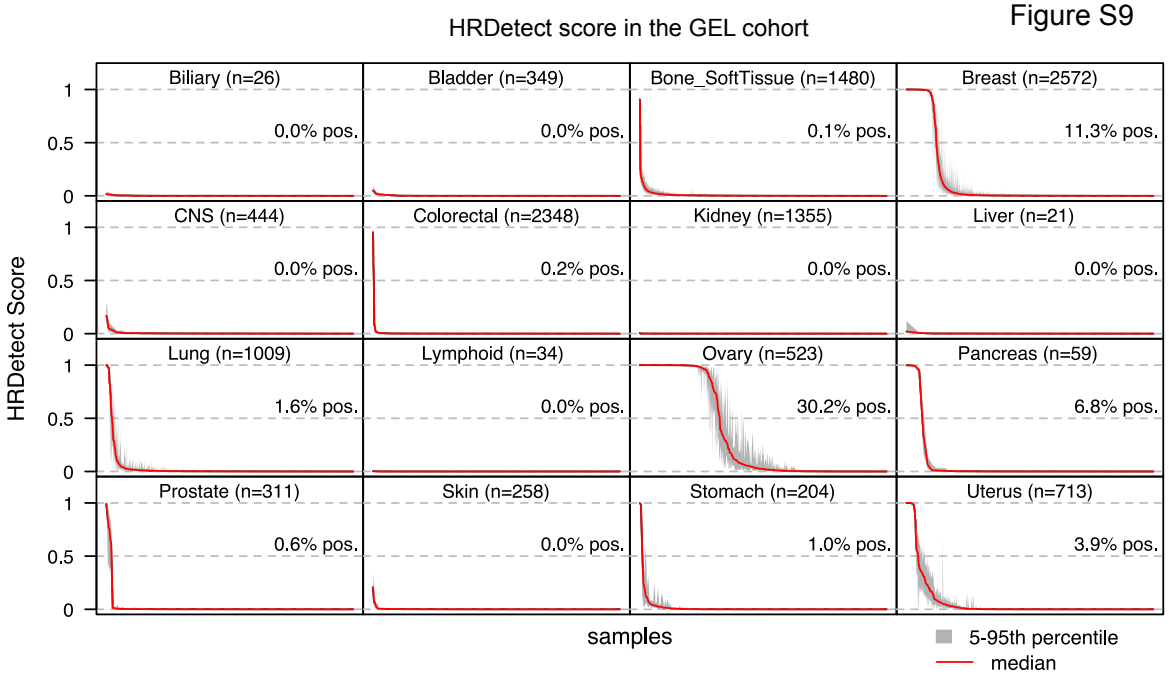


Fig. S9. HRDetect scores across 16 organs. HRDetect score computed in 16 organs in the GEL cohort. For each sample, we computed 1000 bootstrap scores by perturbing the HRDetect input features and provided a median and 5-95th percentile confidence interval (CI). A sample was considered HRDetect score positive, indicating likely homologous recombination deficiency, if all the CI was above 0.5. For each organ we indicated the number of samples (n) for which the score was calculated, and we indicated the percentage of positive samples (pos.).

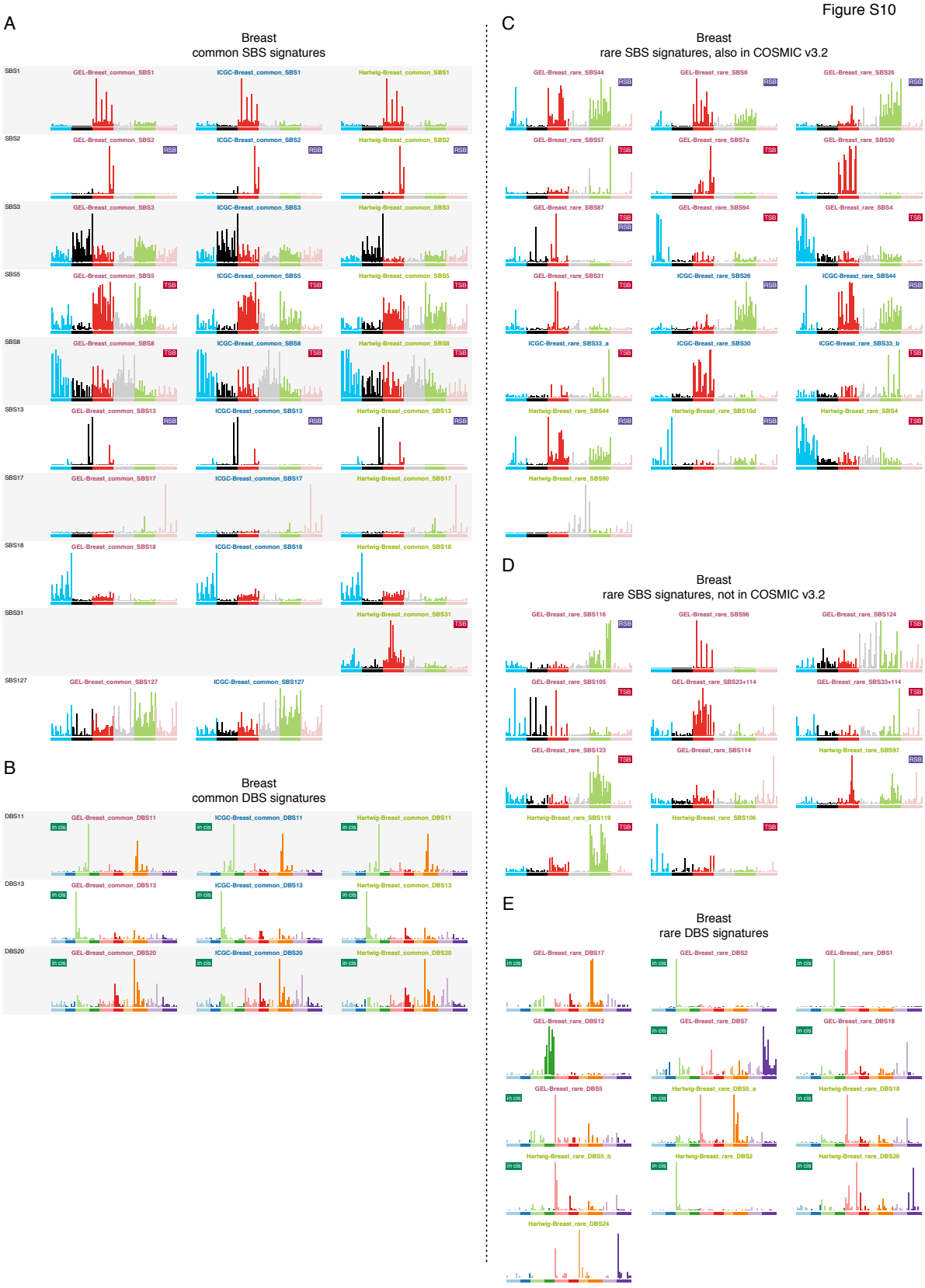


Fig. S10. Summary of breast cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S11

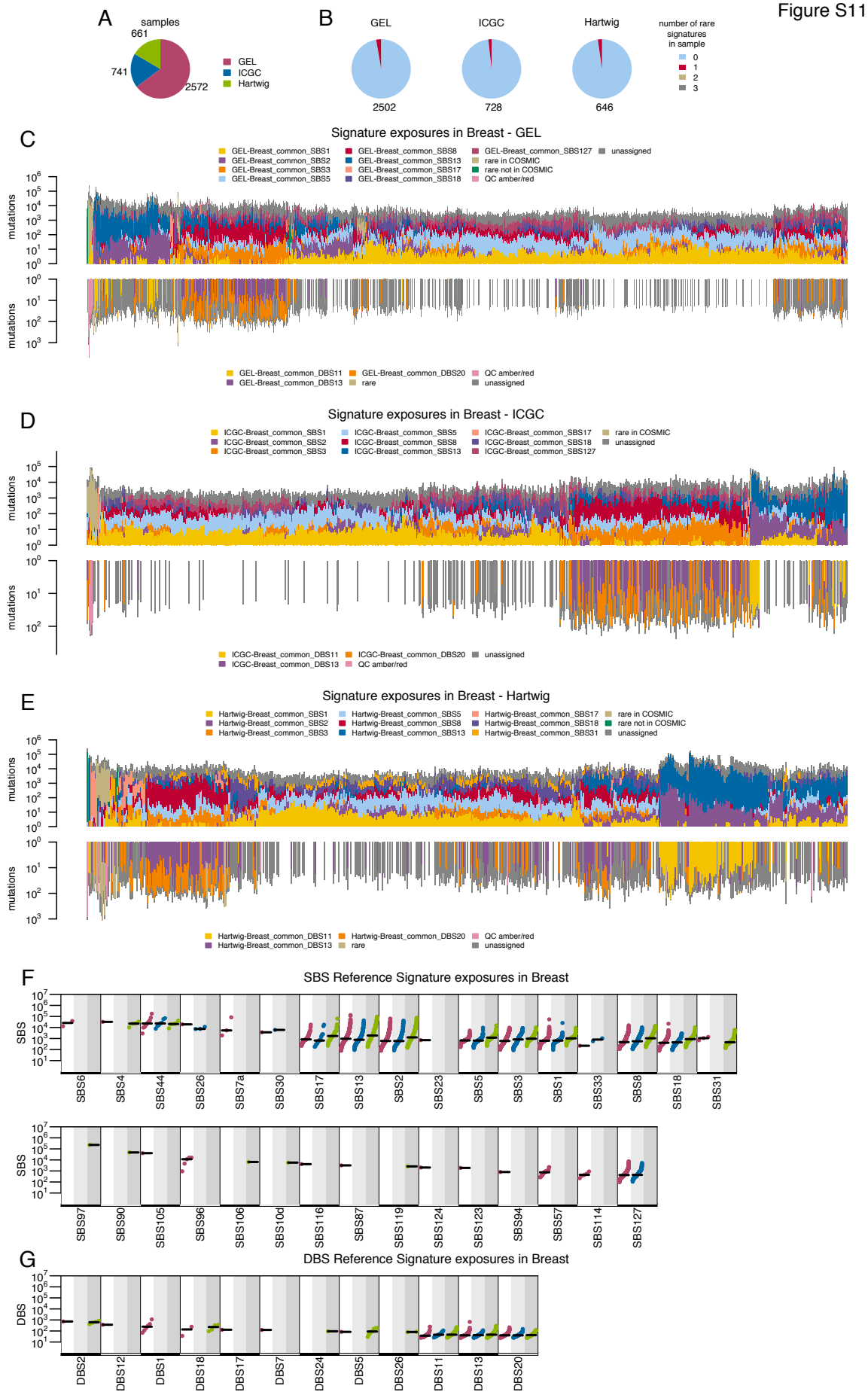

A
samples
661
741
2572
GEL
ICGC
Hartwig
B
GEL
ICGC
Hartwig
2502
728
646
number of rare signatures in sample
0
1
2
3
C
Signature exposures in Breast - GEL
GEL-Breast_common_SBS1
GEL-Breast_common_SBS2
GEL-Breast_common_SBS3
GEL-Breast_common_SBS5
GEL-Breast_common_SBS8
GEL-Breast_common_SBS13
GEL-Breast_common_SBS17
GEL-Breast_common_SBS18
GEL-Breast_common_SBS127
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
GEL-Breast_common_DBS11
GEL-Breast_common_DBS13
GEL-Breast_common_DBS20
rare
D
Signature exposures in Breast - ICGC
ICGC-Breast_common_SBS1
ICGC-Breast_common_SBS2
ICGC-Breast_common_SBS3
ICGC-Breast_common_SBS5
ICGC-Breast_common_SBS8
ICGC-Breast_common_SBS13
ICGC-Breast_common_SBS17
ICGC-Breast_common_SBS18
ICGC-Breast_common_SBS127
rare in COSMIC
unassigned
ICGC-Breast_common_DBS11
ICGC-Breast_common_DBS13
ICGC-Breast_common_DBS20
E
Signature exposures in Breast - Hartwig
Hartwig-Breast_common_SBS1
Hartwig-Breast_common_SBS2
Hartwig-Breast_common_SBS3
Hartwig-Breast_common_SBS5
Hartwig-Breast_common_SBS8
Hartwig-Breast_common_SBS13
Hartwig-Breast_common_SBS17
Hartwig-Breast_common_SBS18
Hartwig-Breast_common_SBS31
rare in COSMIC
rare not in COSMIC
Hartwig-Breast_common_DBS11
Hartwig-Breast_common_DBS13
Hartwig-Breast_common_DBS20
F
SBS Reference Signature exposures in Breast
SBS
SBS6
SBS4
SBS44
SBS26
SBS7a
SBS30
SBS17
SBS13
SBS2
SBS23
SBS5
SBS3
SBS1
SBS33
SBS8
SBS18
SBS31
SBS97
SBS90
SBS105
SBS96
SBS106
SBS10d
SBS116
SBS87
SBS119
SBS124
SBS123
SBS94
SBS57
SBS114
SBS127
G
DBS Reference Signature exposures in Breast
DBS
DBS2
DBS12
DBS1
DBS18
DBS17
DBS7
DBS24
DBS5
DBS26
DBS11
DBS13
DBS20

Fig. S11. Summary of signatures distribution in breast cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S12. Summary of ovarian cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S13

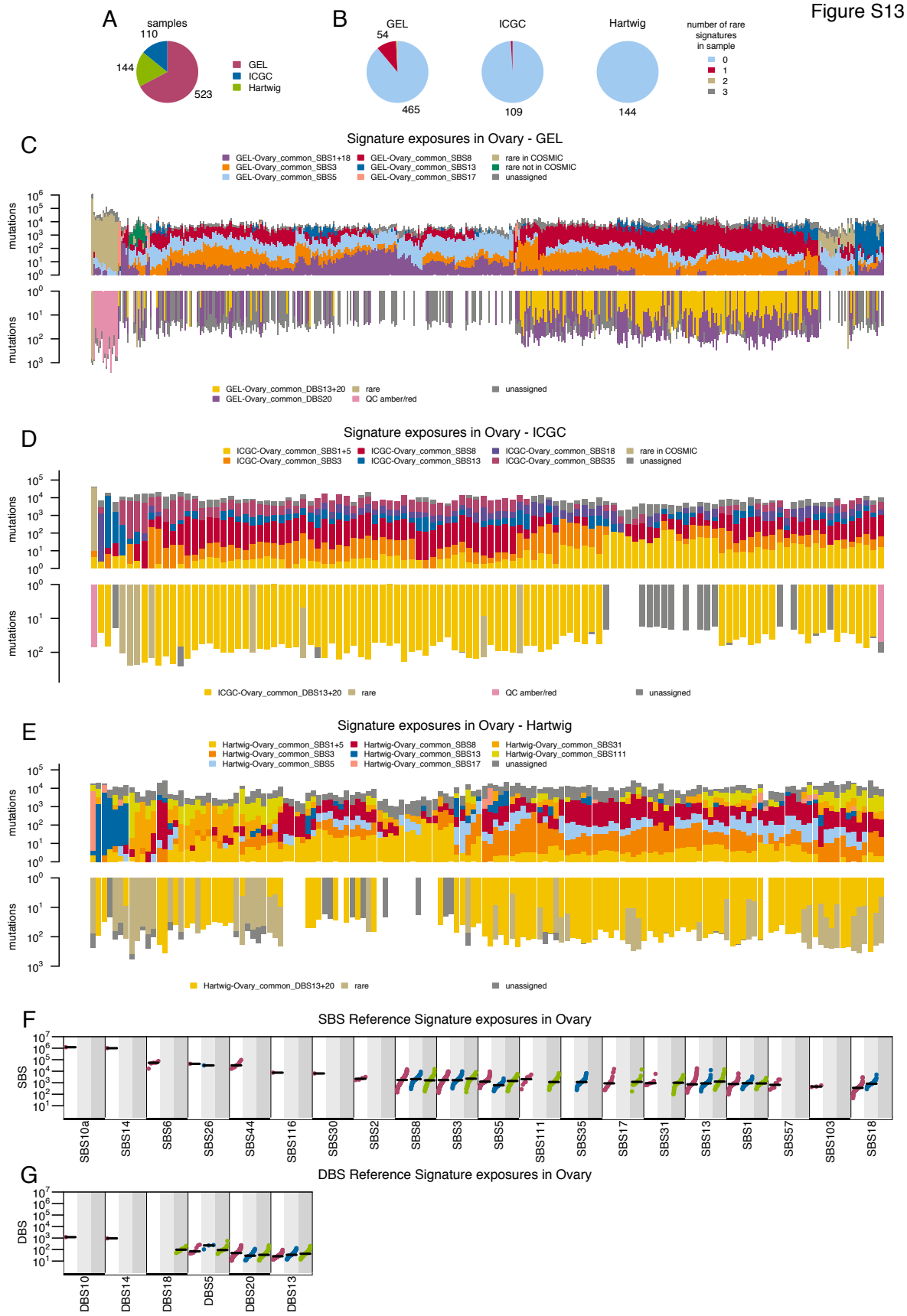


Fig. S13. Summary of signatures distribution in ovarian cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)**

Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S14. Summary of kidney cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S15

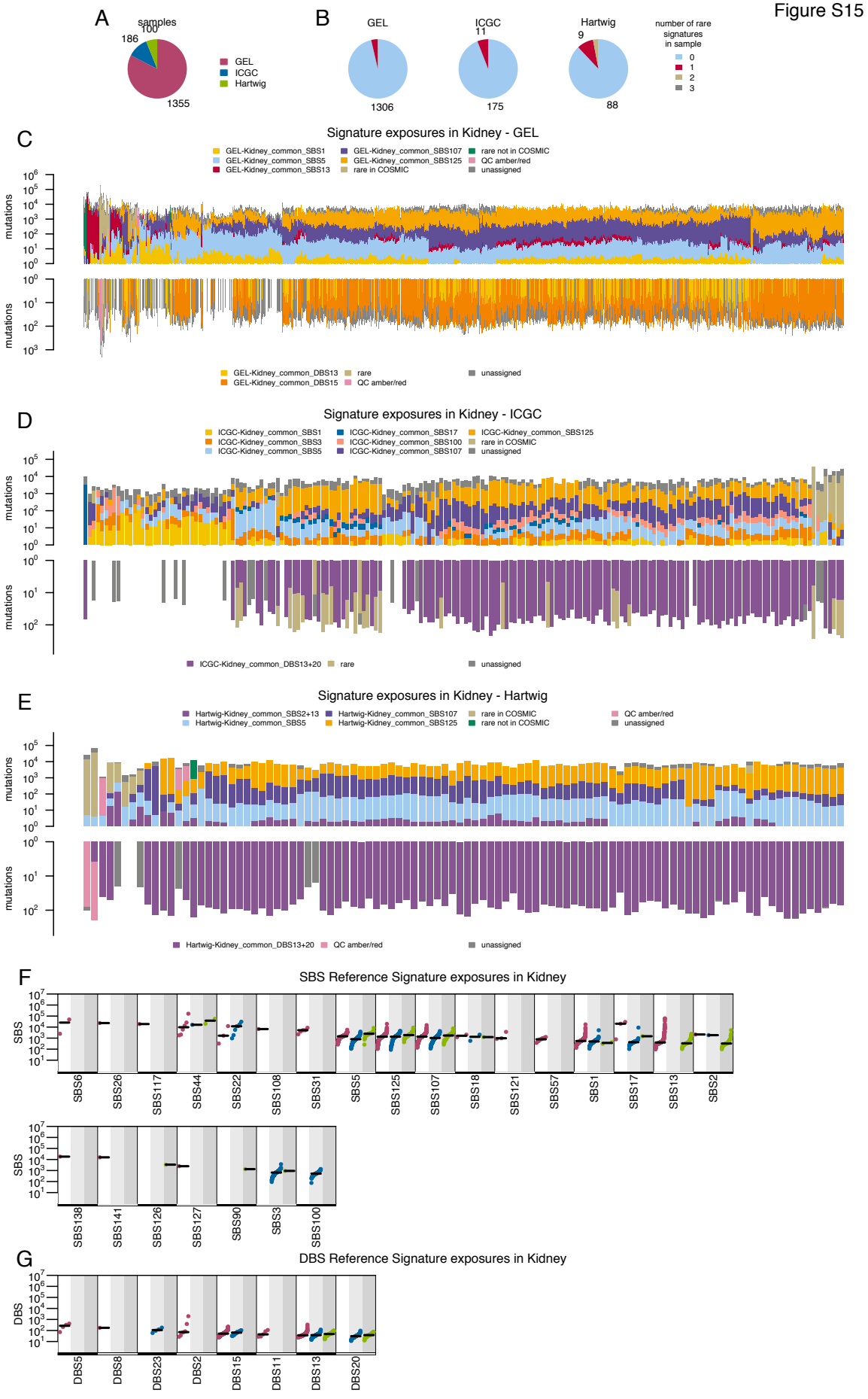
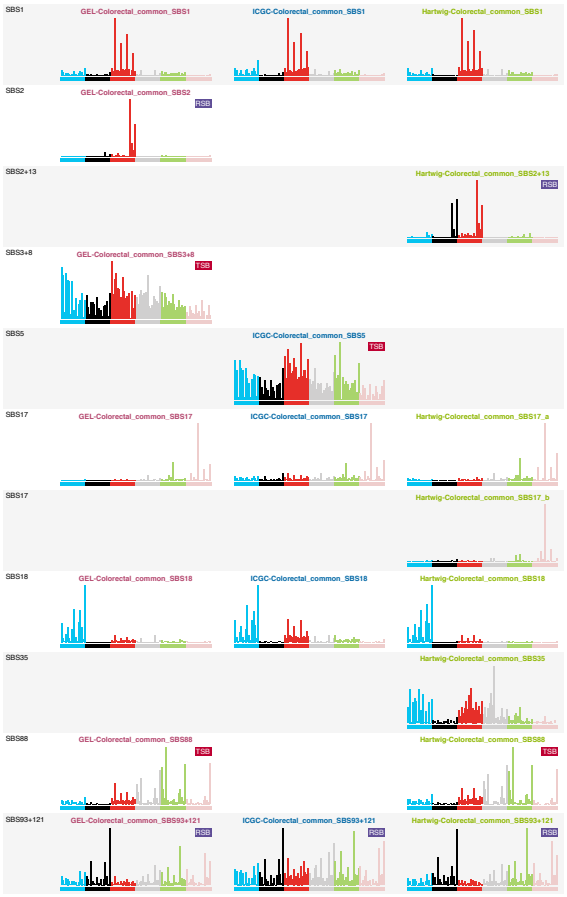

A
samples
186
100
1355
GEL
ICGC
Hartwig
B
GEL
1306
ICGC
11
175
Hartwig
9
88
number of rare signatures in sample
0
1
2
3
C
Signature exposures in Kidney - GEL
GEL-Kidney_common_SBS1
GEL-Kidney_common_SBS5
GEL-Kidney_common_SBS13
GEL-Kidney_common_SBS107
GEL-Kidney_common_SBS125
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
GEL-Kidney_common_DBS13
GEL-Kidney_common_DBS15
rare
QC amber/red
unassigned
D
Signature exposures in Kidney - ICGC
ICGC-Kidney_common_SBS1
ICGC-Kidney_common_SBS3
ICGC-Kidney_common_SBS5
ICGC-Kidney_common_SBS17
ICGC-Kidney_common_SBS100
ICGC-Kidney_common_SBS107
ICGC-Kidney_common_SBS125
rare in COSMIC
unassigned
mutations
ICGC-Kidney_common_DBS13+20
rare
unassigned
E
Signature exposures in Kidney - Hartwig
Hartwig-Kidney_common_SBS2+13
Hartwig-Kidney_common_SBS5
Hartwig-Kidney_common_SBS107
Hartwig-Kidney_common_SBS125
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
Hartwig-Kidney_common_DBS13+20
QC amber/red
unassigned
F
SBS Reference Signature exposures in Kidney
SBS
SBS6
SBS26
SBS117
SBS44
SBS22
SBS108
SBS31
SBS5
SBS125
SBS107
SBS18
SBS121
SBS57
SBS1
SBS17
SBS13
SBS2
SBS
SBS138
SBS141
SBS126
SBS127
SBS90
SBS3
SBS100
G
DBS Reference Signature exposures in Kidney
DBS
DBS5
DBS8
DBS23
DBS2
DBS15
DBS11
DBS13
DBS20

Fig. S15. Summary of signatures distribution in kidney cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Figure S16

A

Colorectal
common SBS signatures



B

Colorectal
common DBS signatures

C

Colorectal
rare SBS signatures, also in COSMIC v3.2

D

Colorectal
rare SBS signatures, not in COSMIC v3.2

E

Colorectal
rare DBS signatures

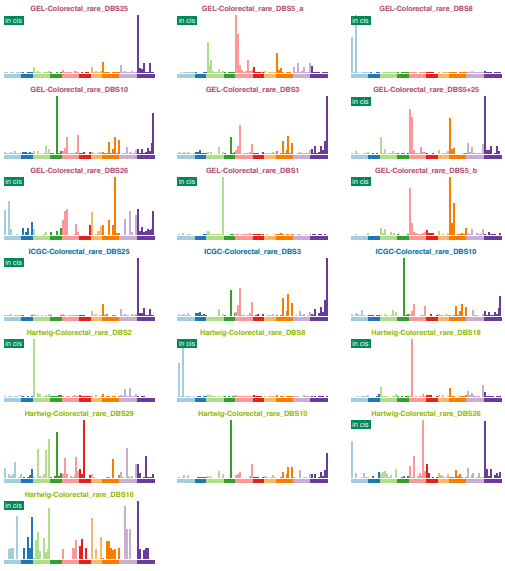


Fig. S16. Summary of colorectal cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S17

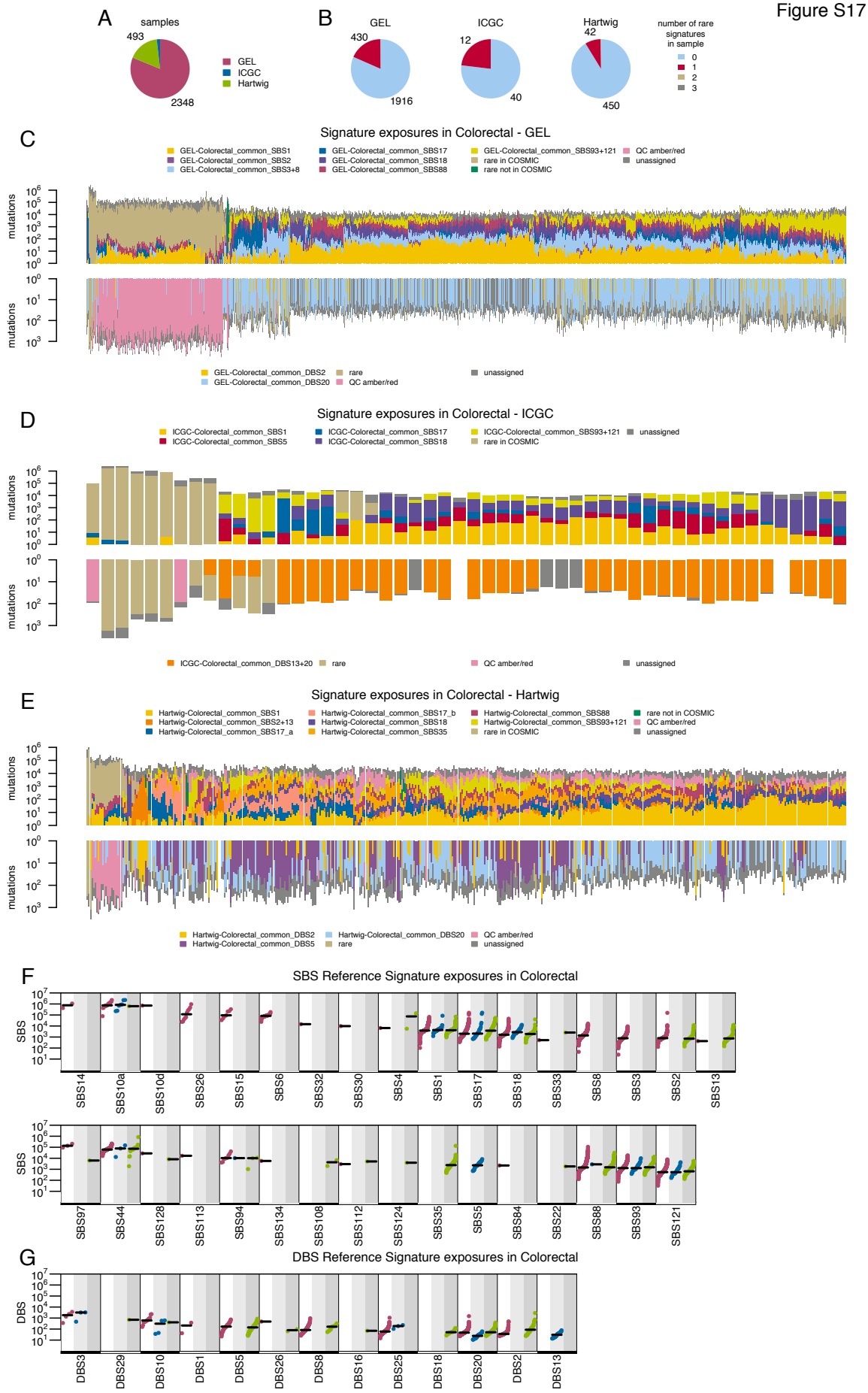

A
samples
493
2348
GEL
ICGC
Hartwig
B
GEL
430
1916
ICGC
12
40
Hartwig
42
450
number of rare signatures in sample
0
1
2
3
C
Signature exposures in Colorectal - GEL
GEL-Colorectal_common_SBS1
GEL-Colorectal_common_SBS2
GEL-Colorectal_common_SBS3+8
GEL-Colorectal_common_SBS17
GEL-Colorectal_common_SBS18
GEL-Colorectal_common_SBS88
GEL-Colorectal_common_SBS93+121
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
GEL-Colorectal_common_DBS2
GEL-Colorectal_common_DBS20
rare
QC amber/red
unassigned
D
Signature exposures in Colorectal - ICGC
ICGC-Colorectal_common_SBS1
ICGC-Colorectal_common_SBS5
ICGC-Colorectal_common_SBS17
ICGC-Colorectal_common_SBS18
ICGC-Colorectal_common_SBS93+121
rare in COSMIC
unassigned
mutations
ICGC-Colorectal_common_DBS13+20
rare
QC amber/red
unassigned
E
Signature exposures in Colorectal - Hartwig
Hartwig-Colorectal_common_SBS1
Hartwig-Colorectal_common_SBS2+13
Hartwig-Colorectal_common_SBS17_a
Hartwig-Colorectal_common_SBS17_b
Hartwig-Colorectal_common_SBS18
Hartwig-Colorectal_common_SBS35
Hartwig-Colorectal_common_SBS88
Hartwig-Colorectal_common_SBS93+121
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
Hartwig-Colorectal_common_DBS2
Hartwig-Colorectal_common_DBS5
Hartwig-Colorectal_common_DBS20
rare
QC amber/red
unassigned
F
SBS Reference Signature exposures in Colorectal
SBS
SBS14
SBS10a
SBS10d
SBS26
SBS15
SBS6
SBS32
SBS30
SBS4
SBS1
SBS17
SBS18
SBS33
SBS8
SBS3
SBS2
SBS13
SBS97
SBS44
SBS128
SBS113
SBS94
SBS134
SBS108
SBS112
SBS124
SBS35
SBS5
SBS84
SBS22
SBS88
SBS93
SBS121
G
DBS Reference Signature exposures in Colorectal
DBS
DBS3
DBS29
DBS10
DBS1
DBS5
DBS26
DBS8
DBS16
DBS25
DBS18
DBS20
DBS2
DBS13

Fig. S17. Summary of signatures distribution in colorectal cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S18. Summary of bone and soft tissue cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S19

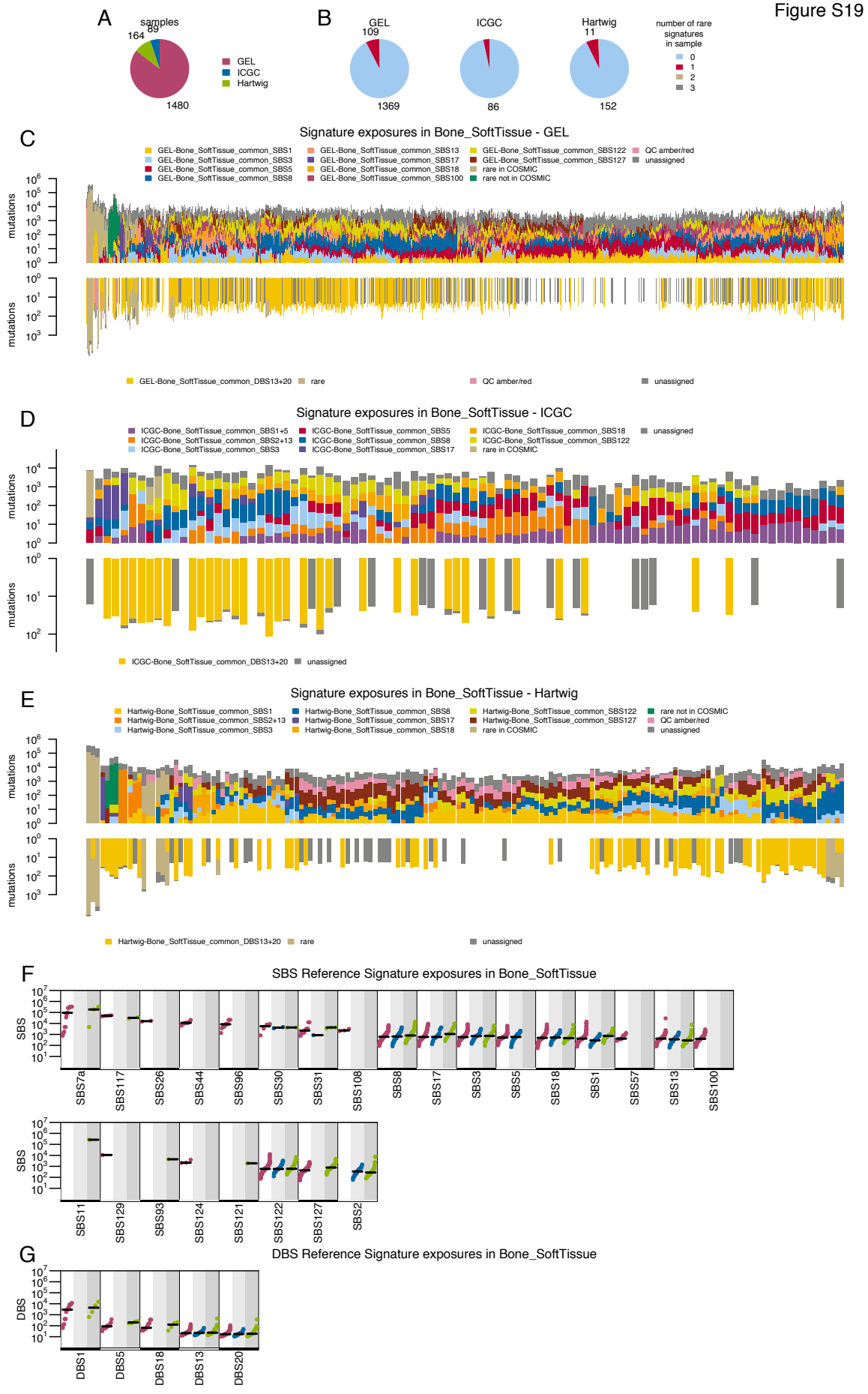

A
samples
164
89
1480
GEL
ICGC
Hartwig
B
GEL
109
1369
ICGC
86
Hartwig
11
152
number of rare signatures in sample
0
1
2
3
C
Signature exposures in Bone_SoftTissue - GEL
GEL-Bone_SoftTissue_common_SBS1
GEL-Bone_SoftTissue_common_SBS3
GEL-Bone_SoftTissue_common_SBS5
GEL-Bone_SoftTissue_common_SBS8
GEL-Bone_SoftTissue_common_SBS13
GEL-Bone_SoftTissue_common_SBS17
GEL-Bone_SoftTissue_common_SBS18
GEL-Bone_SoftTissue_common_SBS100
GEL-Bone_SoftTissue_common_SBS122
GEL-Bone_SoftTissue_common_SBS127
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
GEL-Bone_SoftTissue_common_DBS13+20
rare
QC amber/red
unassigned
D
Signature exposures in Bone_SoftTissue - ICGC
ICGC-Bone_SoftTissue_common_SBS1+5
ICGC-Bone_SoftTissue_common_SBS2+13
ICGC-Bone_SoftTissue_common_SBS3
ICGC-Bone_SoftTissue_common_SBS5
ICGC-Bone_SoftTissue_common_SBS8
ICGC-Bone_SoftTissue_common_SBS17
ICGC-Bone_SoftTissue_common_SBS18
ICGC-Bone_SoftTissue_common_SBS122
rare in COSMIC
unassigned
mutations
ICGC-Bone_SoftTissue_common_DBS13+20
unassigned
E
Signature exposures in Bone_SoftTissue - Hartwig
Hartwig-Bone_SoftTissue_common_SBS1
Hartwig-Bone_SoftTissue_common_SBS2+13
Hartwig-Bone_SoftTissue_common_SBS3
Hartwig-Bone_SoftTissue_common_SBS8
Hartwig-Bone_SoftTissue_common_SBS17
Hartwig-Bone_SoftTissue_common_SBS18
Hartwig-Bone_SoftTissue_common_SBS122
Hartwig-Bone_SoftTissue_common_SBS127
rare not in COSMIC
rare in COSMIC
QC amber/red
unassigned
mutations
Hartwig-Bone_SoftTissue_common_DBS13+20
rare
unassigned
F
SBS Reference Signature exposures in Bone_SoftTissue
SBS
SBS7a
SBS117
SBS26
SBS44
SBS96
SBS30
SBS31
SBS108
SBS8
SBS17
SBS3
SBS5
SBS18
SBS1
SBS57
SBS13
SBS100
SBS11
SBS129
SBS93
SBS124
SBS121
SBS122
SBS127
SBS2
G
DBS Reference Signature exposures in Bone_SoftTissue
DBS
DBS1
DBS5
DBS18
DBS13
DBS20

Fig. S19. Summary of signatures distribution in bone and soft tissue cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S20. Summary of lung cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S21

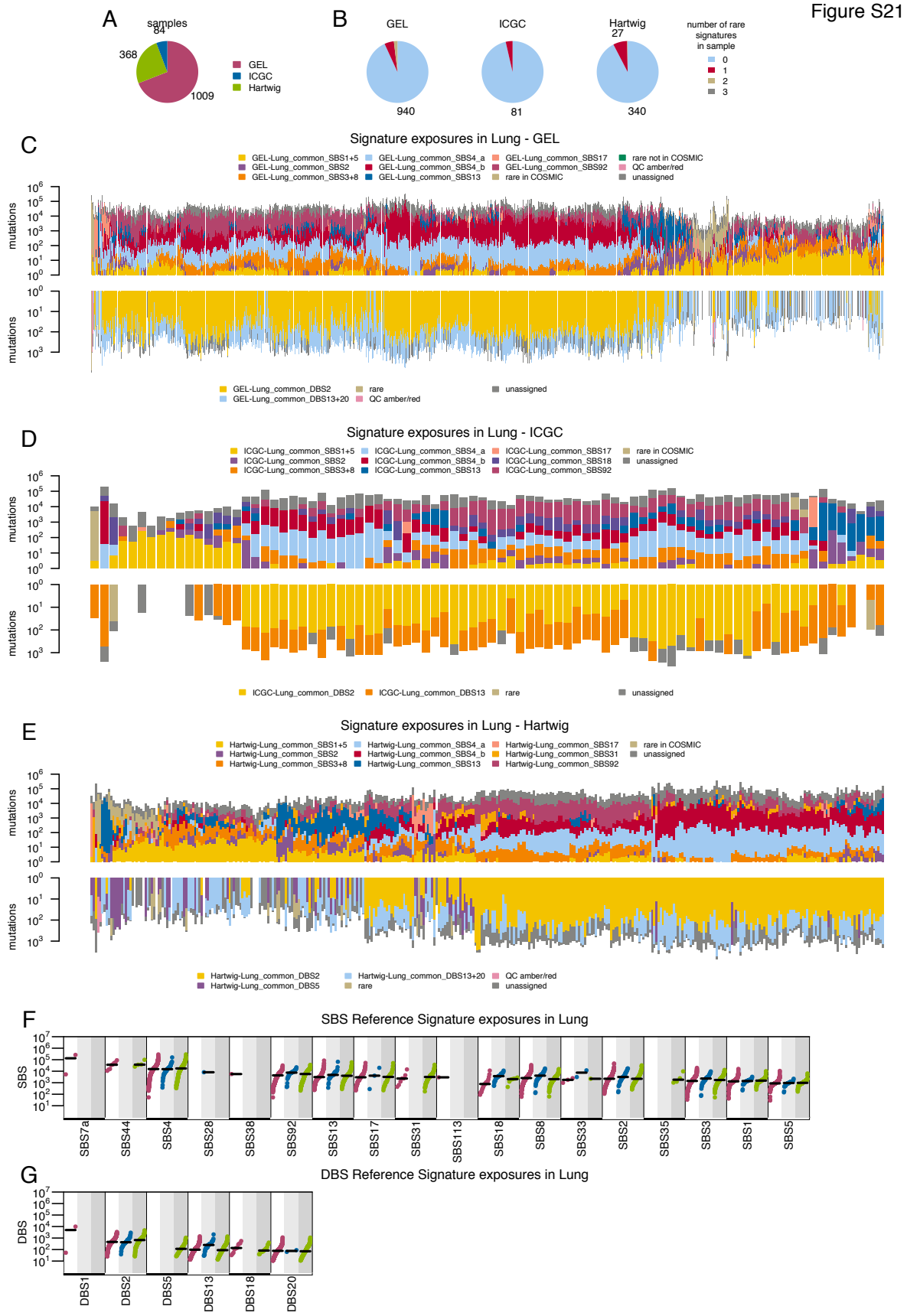


Fig. S21. Summary of signatures distribution in lung cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)**

Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S22. Summary of uterine cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S23

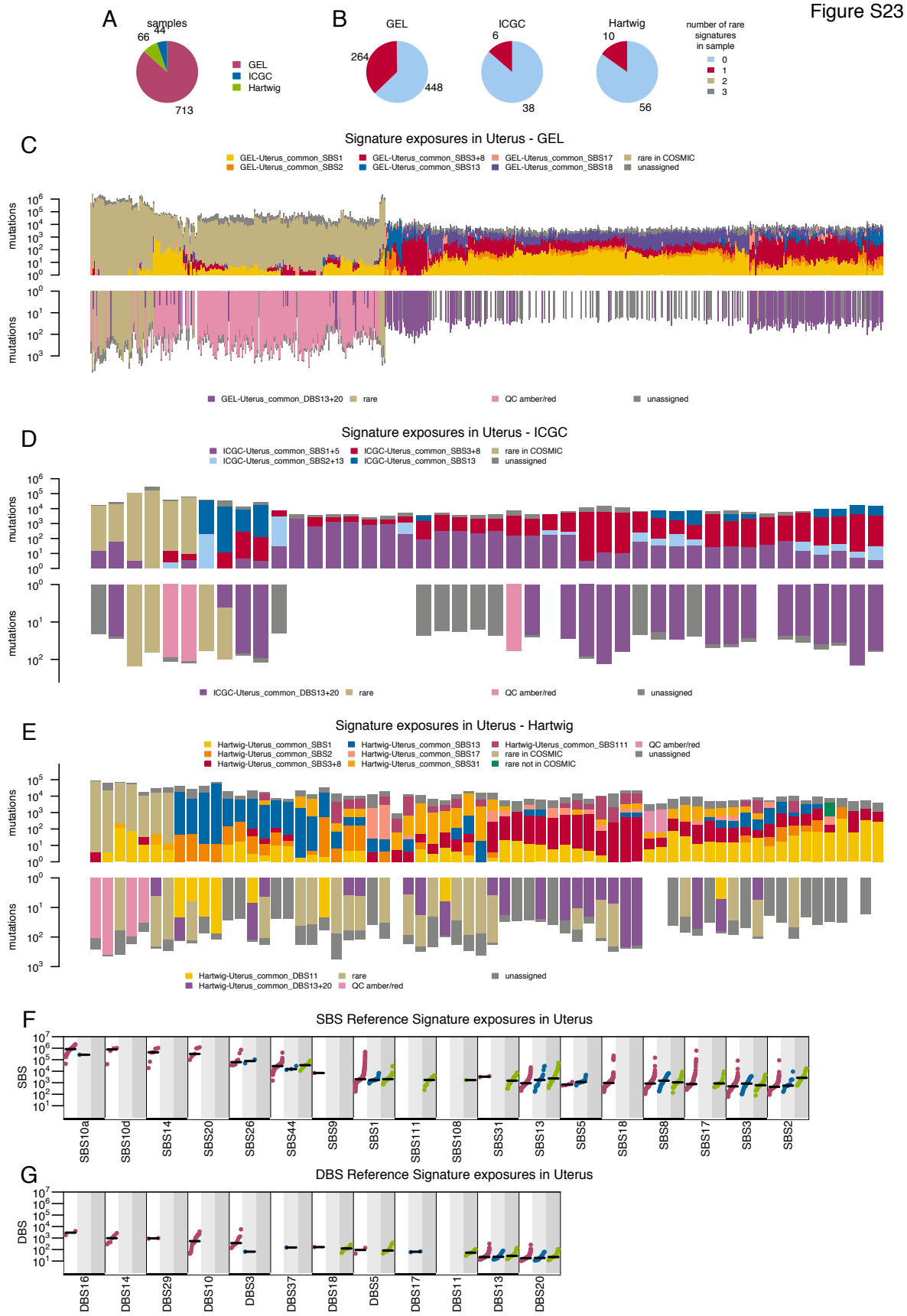


Fig. S23. Summary of signatures distribution in uterine cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)**

Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S24. Summary of central nervous system (CNS) cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S25

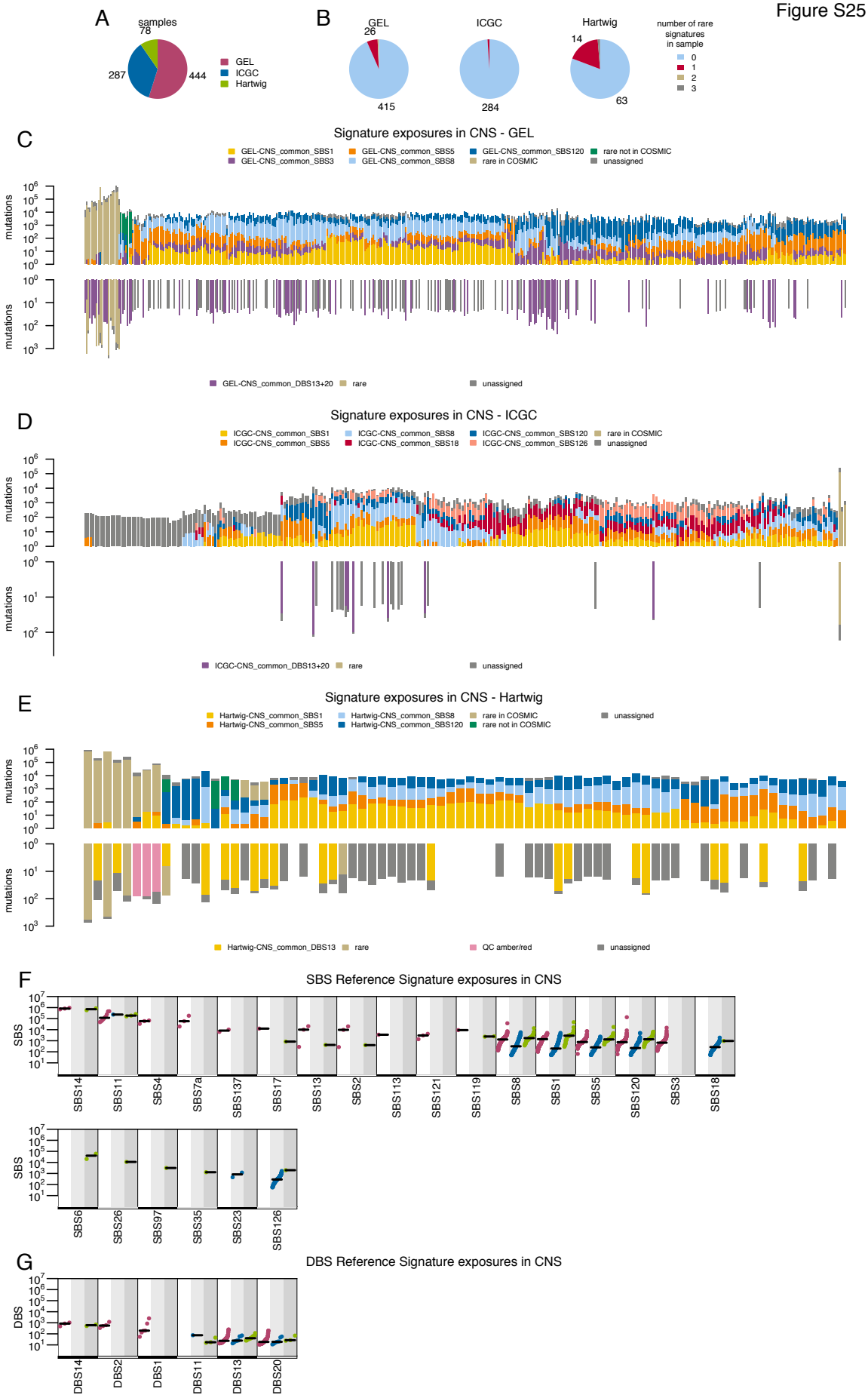

A
samples
78
287
444
GEL
ICGC
Hartwig
B
GEL
26
415
ICGC
284
Hartwig
14
63
number of rare signatures in sample
0
1
2
3
C
Signature exposures in CNS - GEL
GEL-CNS_common_SBS1
GEL-CNS_common_SBS3
GEL-CNS_common_SBS5
GEL-CNS_common_SBS8
GEL-CNS_common_SBS120
rare in COSMIC
rare not in COSMIC
unassigned
mutations
GEL-CNS_common_DBS13+20
rare
unassigned
D
Signature exposures in CNS - ICGC
ICGC-CNS_common_SBS1
ICGC-CNS_common_SBS5
ICGC-CNS_common_SBS8
ICGC-CNS_common_SBS18
ICGC-CNS_common_SBS120
ICGC-CNS_common_SBS126
rare in COSMIC
unassigned
mutations
ICGC-CNS_common_DBS13+20
rare
unassigned
E
Signature exposures in CNS - Hartwig
Hartwig-CNS_common_SBS1
Hartwig-CNS_common_SBS5
Hartwig-CNS_common_SBS8
Hartwig-CNS_common_SBS120
rare in COSMIC
rare not in COSMIC
unassigned
mutations
Hartwig-CNS_common_DBS13
rare
QC amber/red
unassigned
F
SBS Reference Signature exposures in CNS
SBS
SBS14
SBS11
SBS4
SBS7a
SBS137
SBS17
SBS13
SBS2
SBS113
SBS121
SBS119
SBS8
SBS1
SBS5
SBS120
SBS3
SBS18
SBS6
SBS26
SBS97
SBS35
SBS23
SBS126
G
DBS Reference Signature exposures in CNS
DBS
DBS14
DBS2
DBS1
DBS11
DBS13
DBS20

Fig. S25. Summary of signatures distribution in central nervous system (CNS) cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

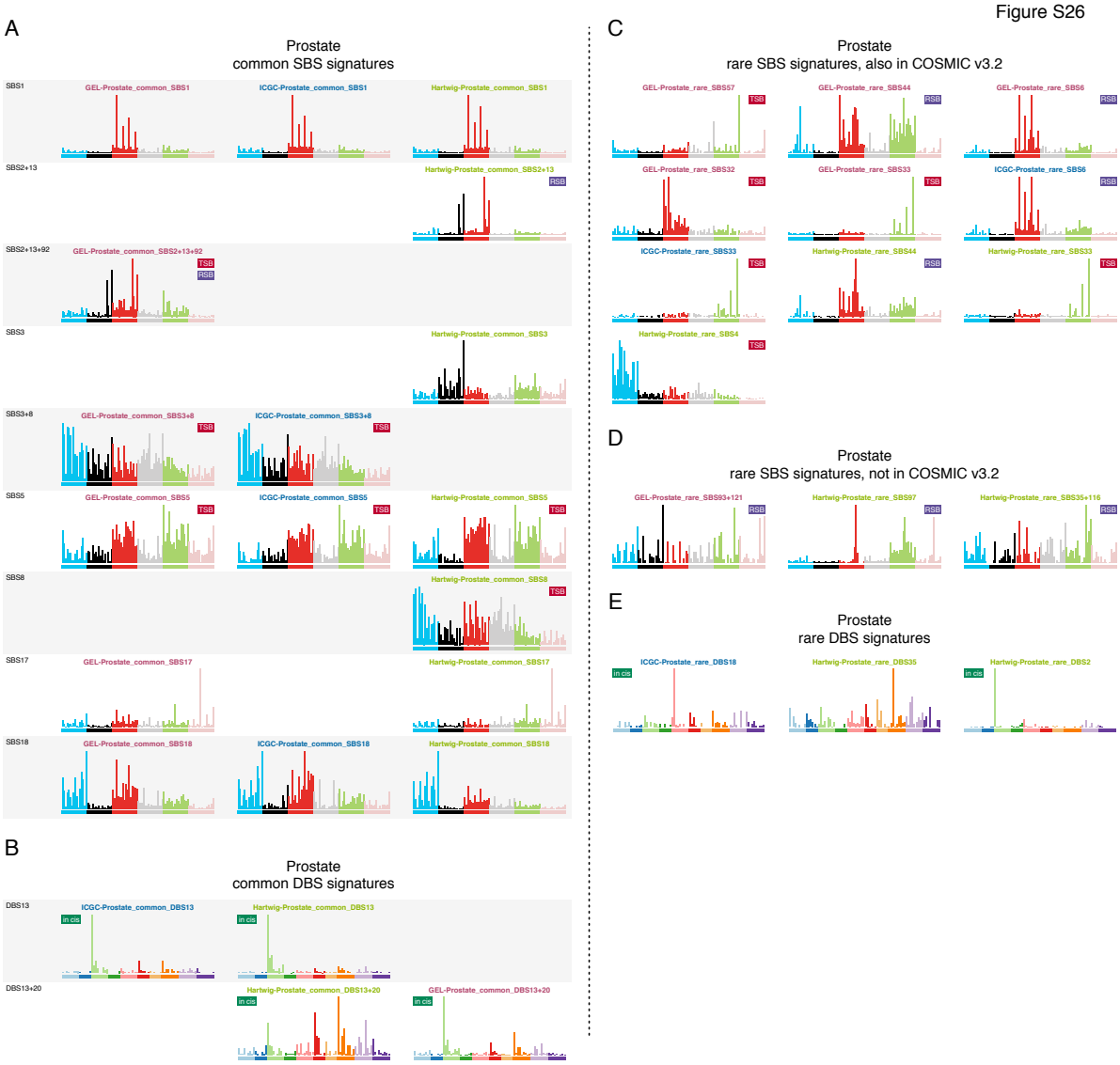


Fig. S26. Summary of prostate cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S27

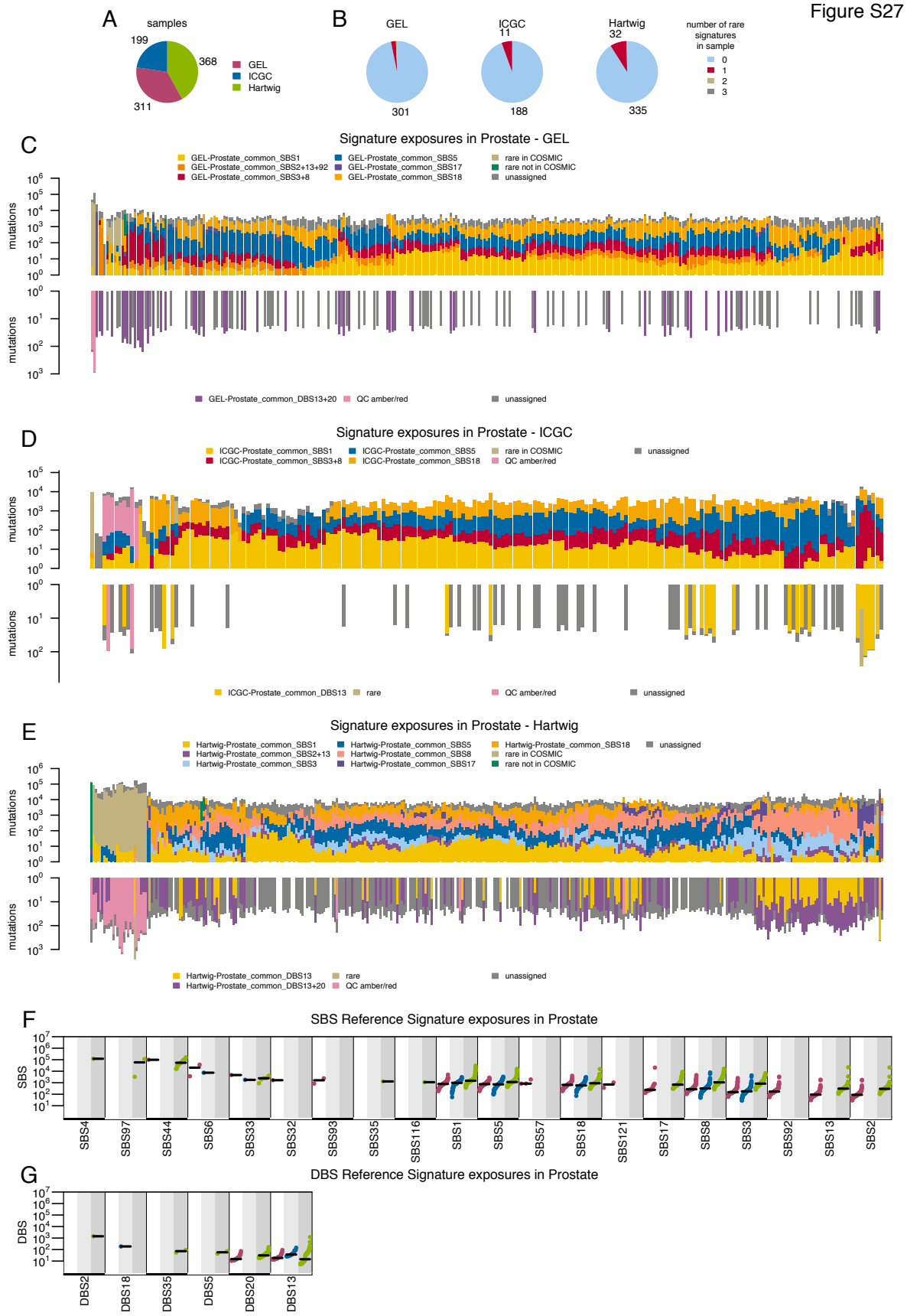


Fig. S27. Summary of signatures distribution in prostate cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)**

Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S28. Summary of bladder cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S29

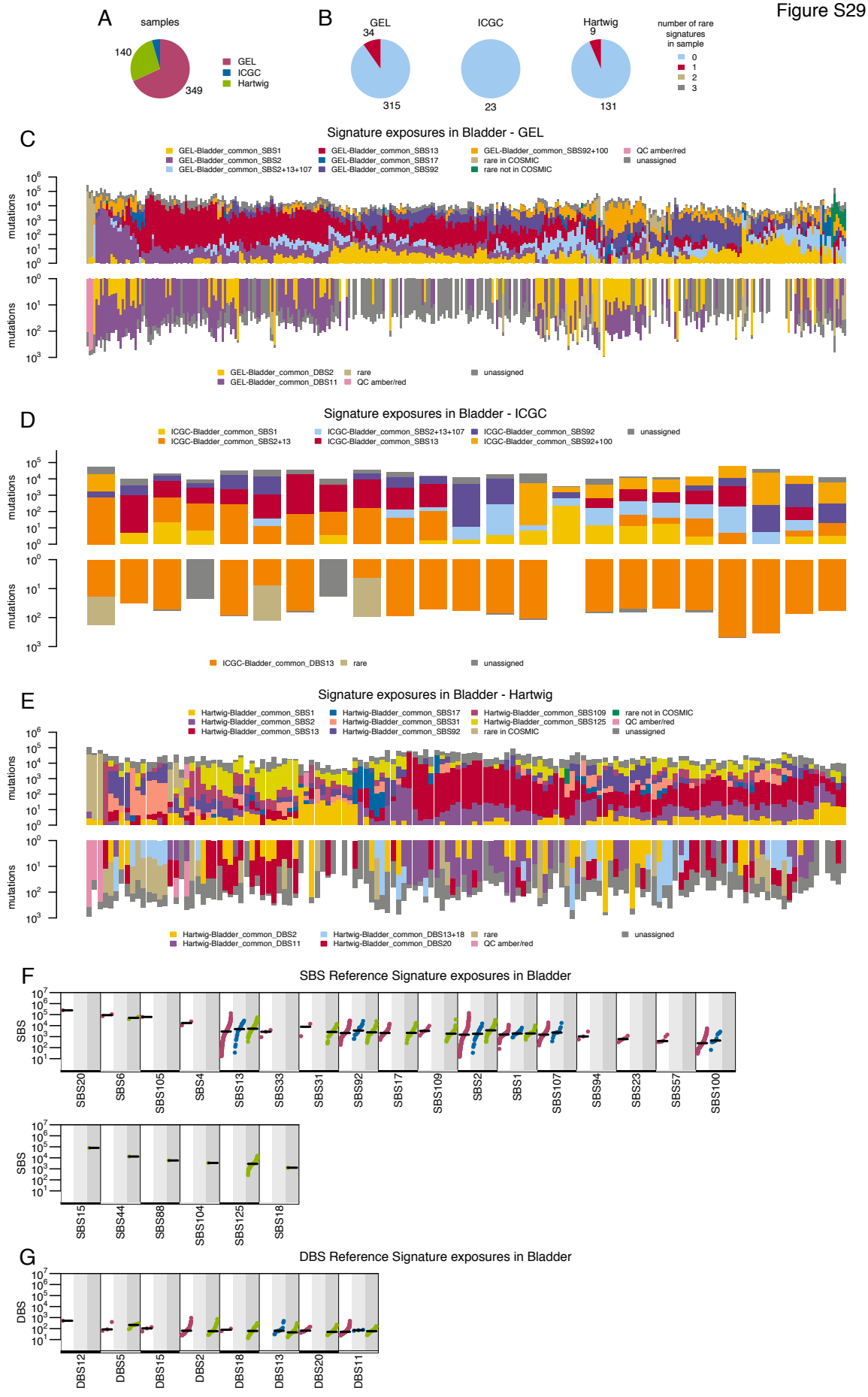

A
samples
140
349
GEL
ICGC
Hartwig
B
GEL
34
315
ICGC
23
Hartwig
9
131
number of rare signatures in sample
0
1
2
3
C
Signature exposures in Bladder - GEL
GEL-Bladder_common_SBS1
GEL-Bladder_common_SBS2
GEL-Bladder_common_SBS2+13+107
GEL-Bladder_common_SBS13
GEL-Bladder_common_SBS17
GEL-Bladder_common_SBS92
GEL-Bladder_common_SBS92+100
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
GEL-Bladder_common_DBS2
GEL-Bladder_common_DBS11
rare
QC amber/red
unassigned
D
Signature exposures in Bladder - ICGC
ICGC-Bladder_common_SBS1
ICGC-Bladder_common_SBS2+13
ICGC-Bladder_common_SBS2+13+107
ICGC-Bladder_common_SBS13
ICGC-Bladder_common_SBS92
ICGC-Bladder_common_SBS92+100
unassigned
mutations
ICGC-Bladder_common_DBS13
rare
unassigned
E
Signature exposures in Bladder - Hartwig
Hartwig-Bladder_common_SBS1
Hartwig-Bladder_common_SBS2
Hartwig-Bladder_common_SBS13
Hartwig-Bladder_common_SBS17
Hartwig-Bladder_common_SBS31
Hartwig-Bladder_common_SBS92
Hartwig-Bladder_common_SBS109
Hartwig-Bladder_common_SBS125
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
Hartwig-Bladder_common_DBS2
Hartwig-Bladder_common_DBS11
Hartwig-Bladder_common_DBS13+18
Hartwig-Bladder_common_DBS20
rare
QC amber/red
unassigned
F
SBS Reference Signature exposures in Bladder
SBS
SBS20
SBS6
SBS105
SBS4
SBS13
SBS33
SBS31
SBS92
SBS17
SBS109
SBS2
SBS1
SBS107
SBS94
SBS23
SBS57
SBS100
SBS15
SBS44
SBS88
SBS104
SBS125
SBS18
G
DBS Reference Signature exposures in Bladder
DBS
DBS12
DBS5
DBS15
DBS2
DBS18
DBS13
DBS20
DBS11

Fig. S29. Summary of signatures distribution in bladder cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

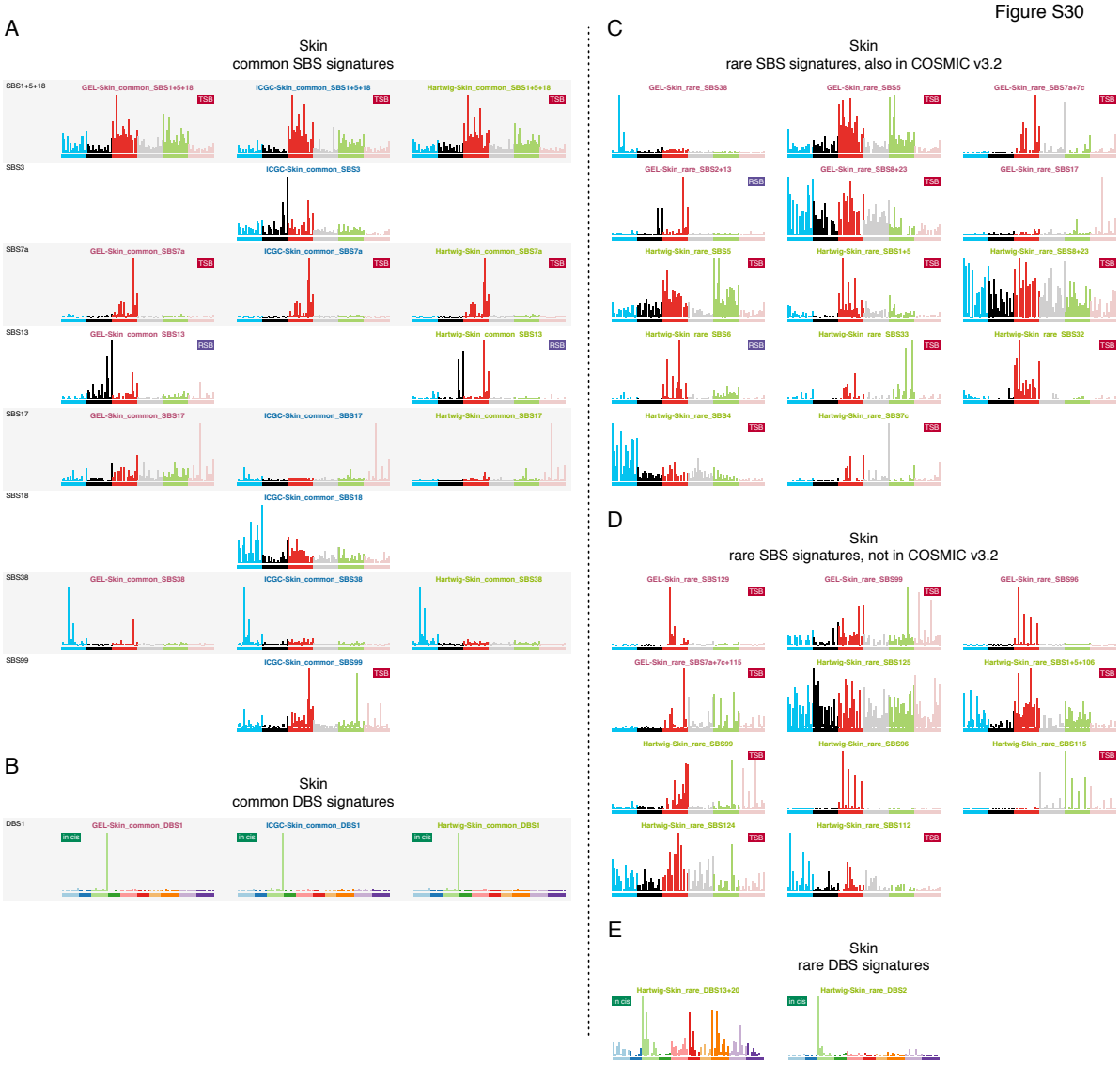


Fig. S30. Summary of skin cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S31

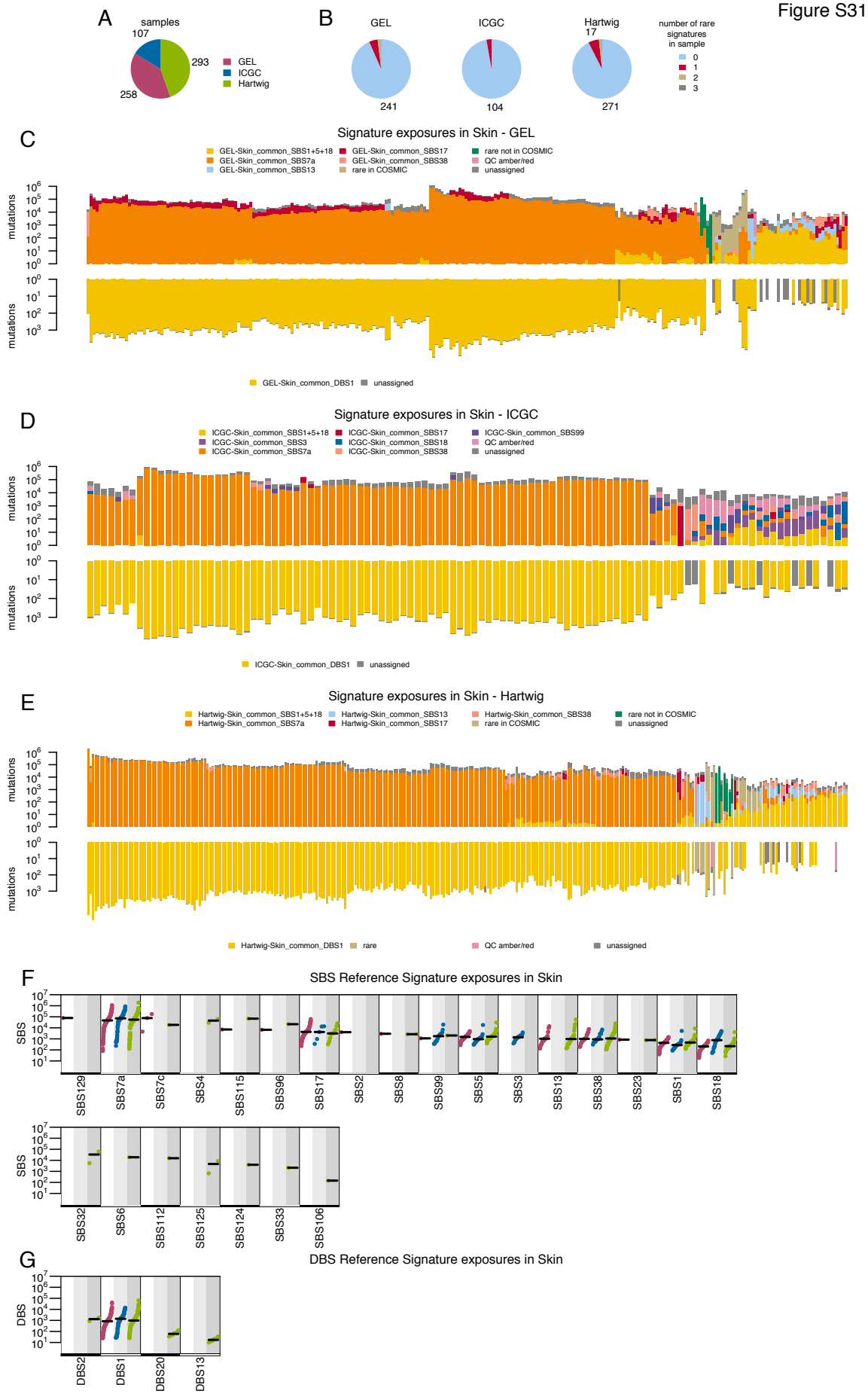

A
samples
107
293
258
GEL
ICGC
Hartwig
B
GEL
ICGC
Hartwig
17
241
104
271
number of rare signatures in sample
0
1
2
3
C
Signature exposures in Skin - GEL
GEL-Skin_common_SBS1+5+18
GEL-Skin_common_SBS17
rare not in COSMIC
GEL-Skin_common_SBS7a
GEL-Skin_common_SBS38
QC amber/red
GEL-Skin_common_SBS13
rare in COSMIC
unassigned
mutations
GEL-Skin_common_DBS1
unassigned
D
Signature exposures in Skin - ICGC
ICGC-Skin_common_SBS1+5+18
ICGC-Skin_common_SBS17
ICGC-Skin_common_SBS99
ICGC-Skin_common_SBS3
ICGC-Skin_common_SBS18
QC amber/red
ICGC-Skin_common_SBS7a
ICGC-Skin_common_SBS38
unassigned
mutations
ICGC-Skin_common_DBS1
unassigned
E
Signature exposures in Skin - Hartwig
Hartwig-Skin_common_SBS1+5+18
Hartwig-Skin_common_SBS13
Hartwig-Skin_common_SBS38
rare not in COSMIC
Hartwig-Skin_common_SBS7a
Hartwig-Skin_common_SBS17
rare in COSMIC
unassigned
mutations
Hartwig-Skin_common_DBS1
rare
QC amber/red
unassigned
F
SBS Reference Signature exposures in Skin
SBS
SBS129
SBS7a
SBS7c
SBS4
SBS115
SBS96
SBS17
SBS2
SBS8
SBS99
SBS5
SBS3
SBS13
SBS38
SBS23
SBS1
SBS18
SBS32
SBS6
SBS112
SBS125
SBS124
SBS33
SBS106
G
DBS Reference Signature exposures in Skin
DBS
DBS2
DBS1
DBS20
DBS13

Fig. S31. Summary of signatures distribution in skin cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S32. Summary of stomach cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S33

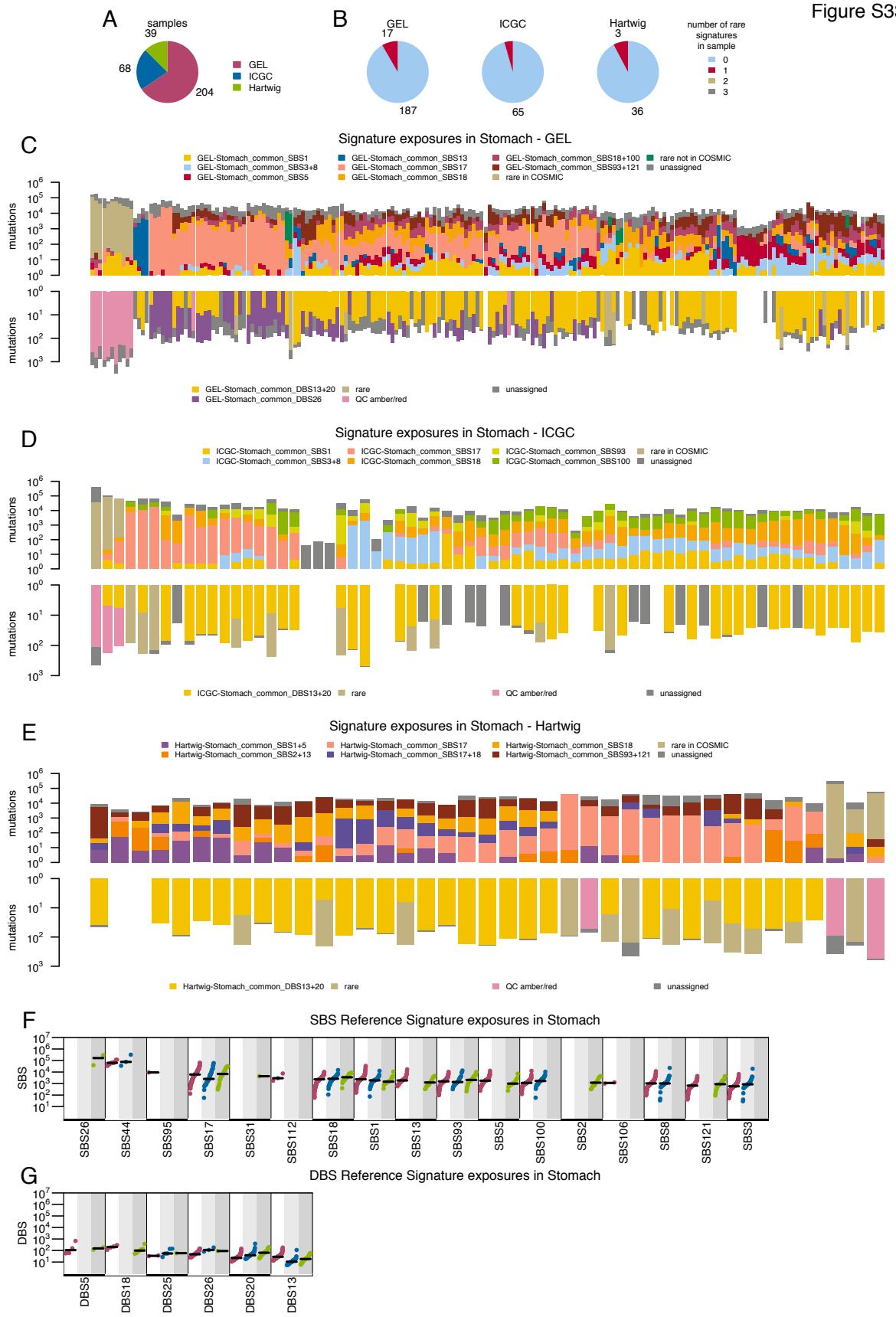


Fig. S33. Summary of signatures distribution in stomach cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)**

Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S34. Summary of neuroendocrine cancer (NET) signatures extracted in this study. **(A)** Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S35

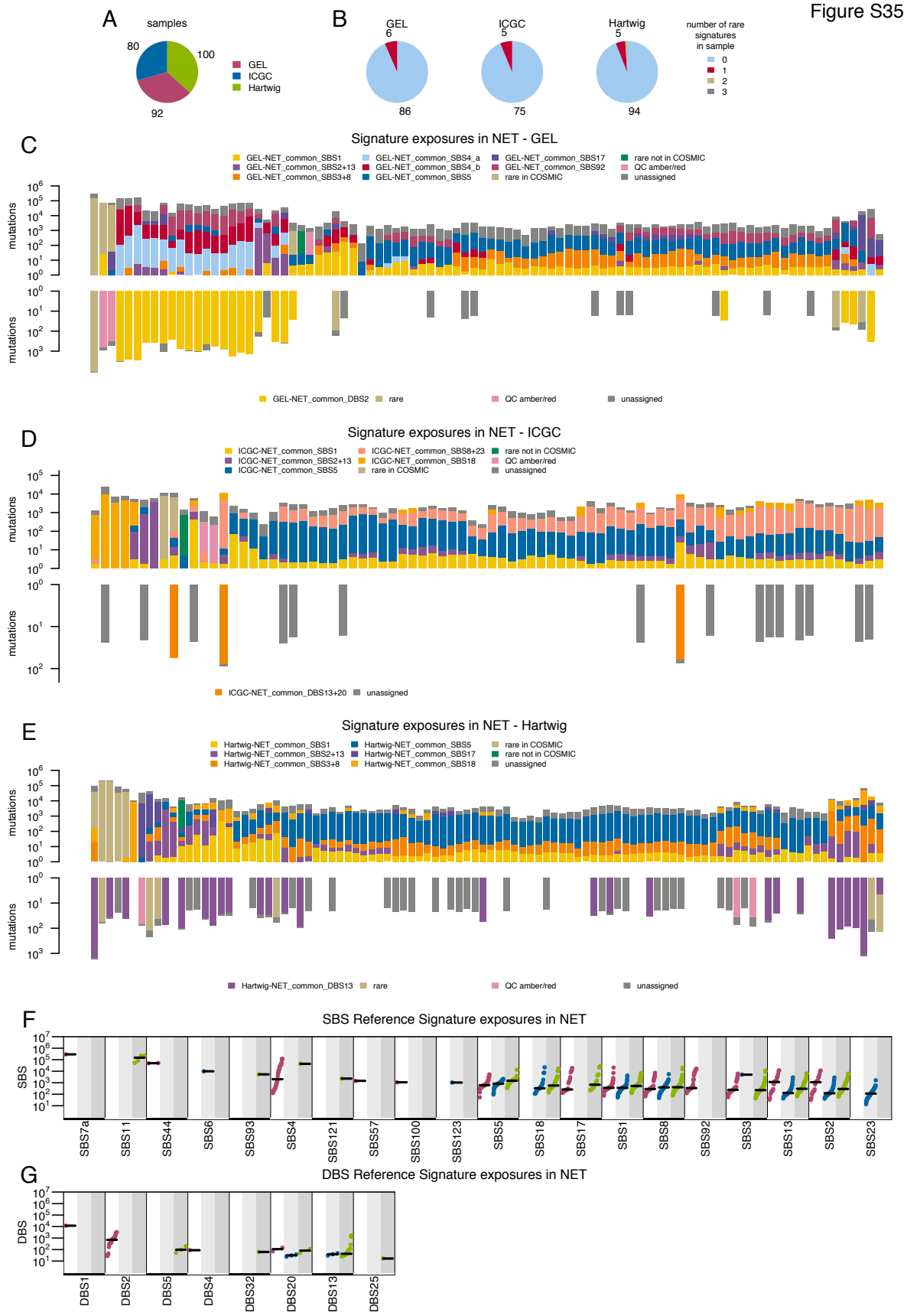


Fig. S35. Summary of signatures distribution in neuroendocrine cancer (NET) samples. **(A)** Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample

in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S36. Summary of pancreatic cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S37

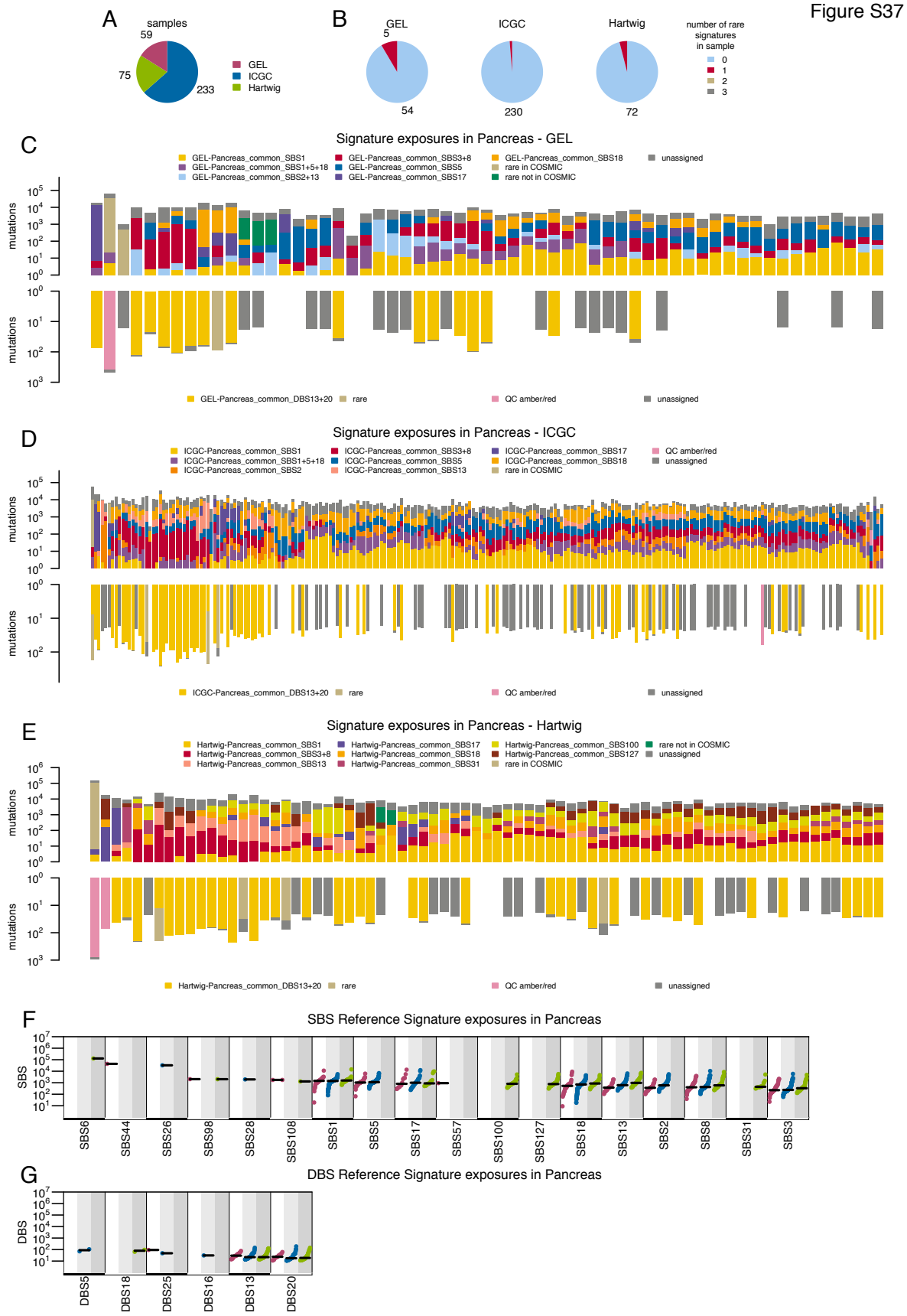


Fig. S37. Summary of signatures distribution in pancreatic cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)**

Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S38. Summary of biliary cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S39

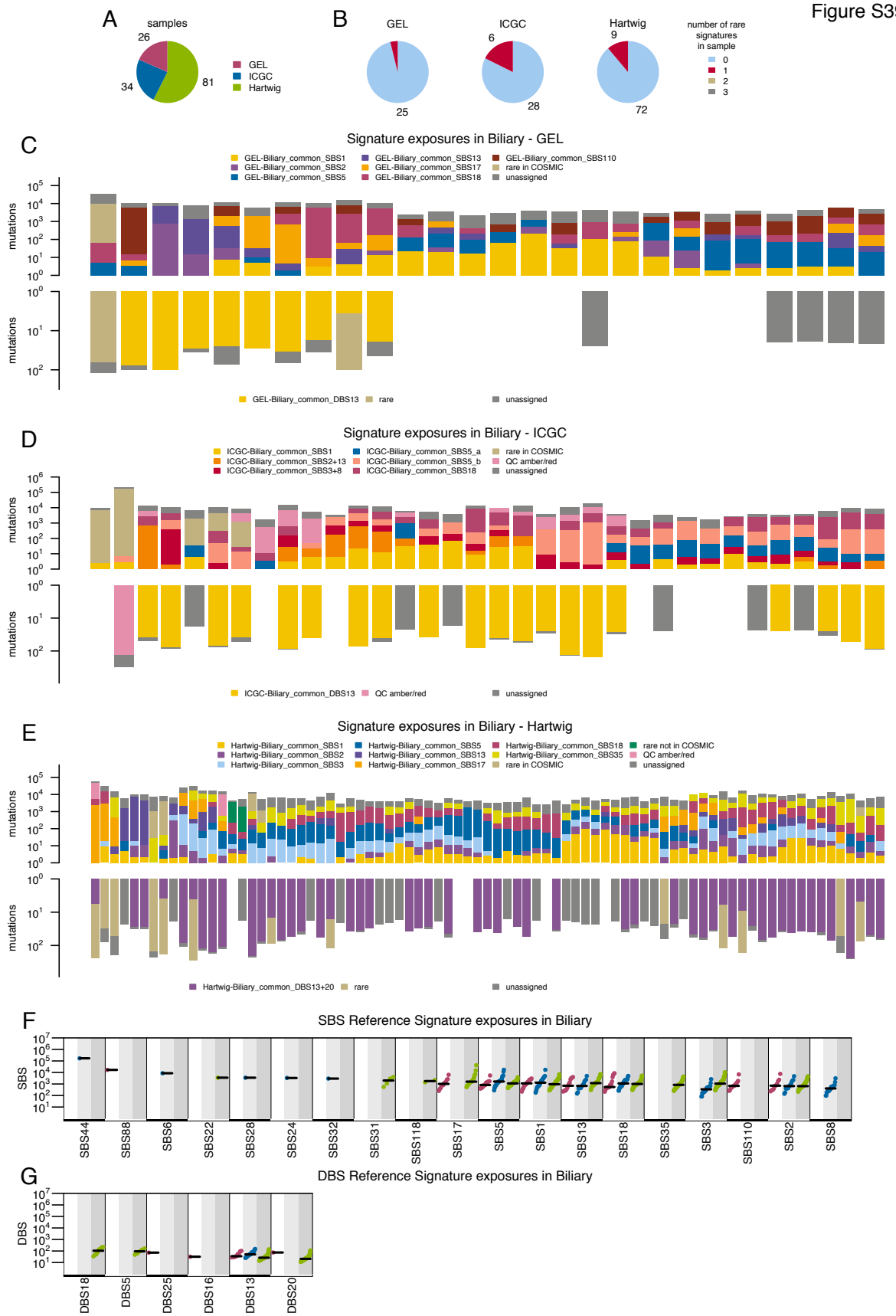


Fig. S39. Summary of signatures distribution in biliary cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)**

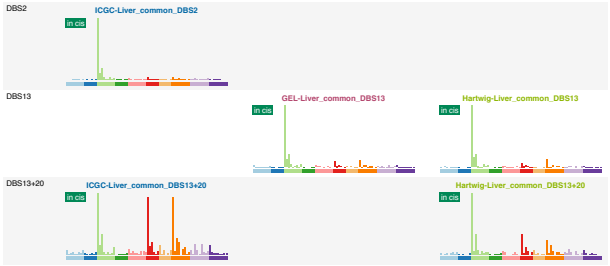
Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

A

Liver
common SBS signatures

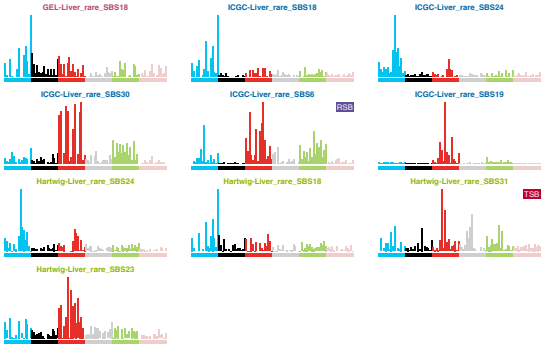
B

Liver
common DBS signatures



C

Liver
rare SBS signatures, also in COSMIC v3.2



D

Liver
rare SBS signatures, not in COSMIC v3.2

E

Liver
rare DBS signatures

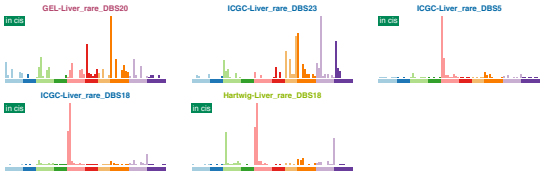


Fig. S40. Summary of liver cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S41

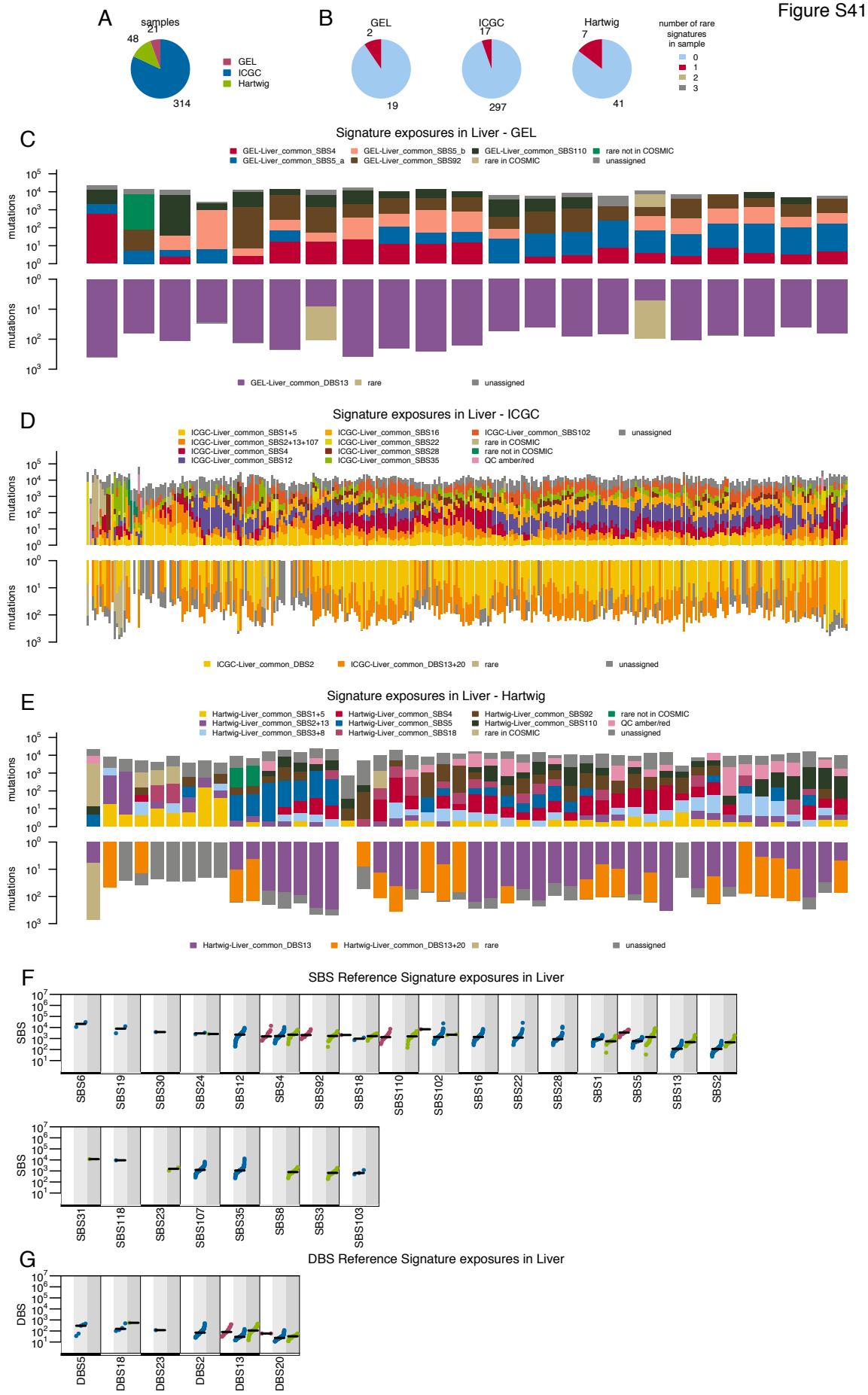

A
samples
48
21
314
GEL
ICGC
Hartwig
B
GEL
2
19
ICGC
17
297
Hartwig
7
41
number of rare signatures in sample
0
1
2
3
C
Signature exposures in Liver - GEL
GEL-Liver_common_SBS4
GEL-Liver_common_SBS5_a
GEL-Liver_common_SBS5_b
GEL-Liver_common_SBS92
GEL-Liver_common_SBS110
rare in COSMIC
rare not in COSMIC
unassigned
mutations
GEL-Liver_common_DBS13
rare
unassigned
D
Signature exposures in Liver - ICGC
ICGC-Liver_common_SBS1+5
ICGC-Liver_common_SBS2+13+107
ICGC-Liver_common_SBS4
ICGC-Liver_common_SBS12
ICGC-Liver_common_SBS16
ICGC-Liver_common_SBS22
ICGC-Liver_common_SBS28
ICGC-Liver_common_SBS35
ICGC-Liver_common_SBS102
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
ICGC-Liver_common_DBS2
ICGC-Liver_common_DBS13+20
rare
unassigned
E
Signature exposures in Liver - Hartwig
Hartwig-Liver_common_SBS1+5
Hartwig-Liver_common_SBS2+13
Hartwig-Liver_common_SBS3+8
Hartwig-Liver_common_SBS4
Hartwig-Liver_common_SBS5
Hartwig-Liver_common_SBS18
Hartwig-Liver_common_SBS92
Hartwig-Liver_common_SBS110
rare in COSMIC
rare not in COSMIC
QC amber/red
unassigned
mutations
Hartwig-Liver_common_DBS13
Hartwig-Liver_common_DBS13+20
rare
unassigned
F
SBS Reference Signature exposures in Liver
SBS
SBS6
SBS19
SBS30
SBS24
SBS12
SBS4
SBS92
SBS18
SBS110
SBS102
SBS16
SBS22
SBS28
SBS1
SBS5
SBS13
SBS2
SBS31
SBS118
SBS23
SBS107
SBS35
SBS8
SBS3
SBS103
G
DBS Reference Signature exposures in Liver
DBS
DBS5
DBS18
DBS23
DBS2
DBS13
DBS20

Fig. S41. Summary of signatures distribution in liver cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

Fig. S42. Summary of lymphoid cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S43

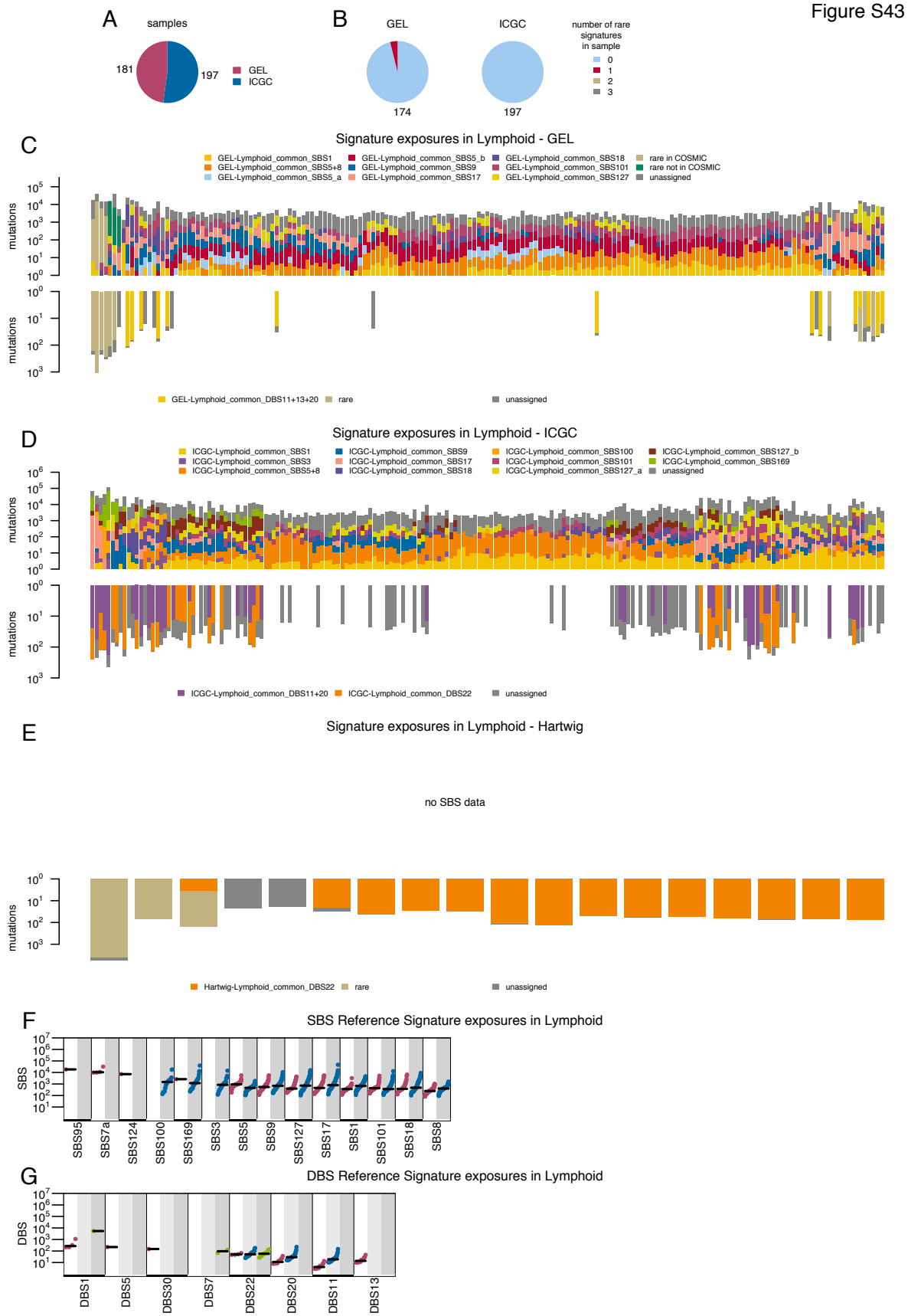


Fig. S43. Summary of signatures distribution in lymphoid cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)**

Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(E)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(F)** Number of mutations associated with each SBS reference signature across GEL, ICGC and Hartwig samples. **(G)** Number of mutations associated with each DBS reference signature across GEL, ICGC and Hartwig samples.

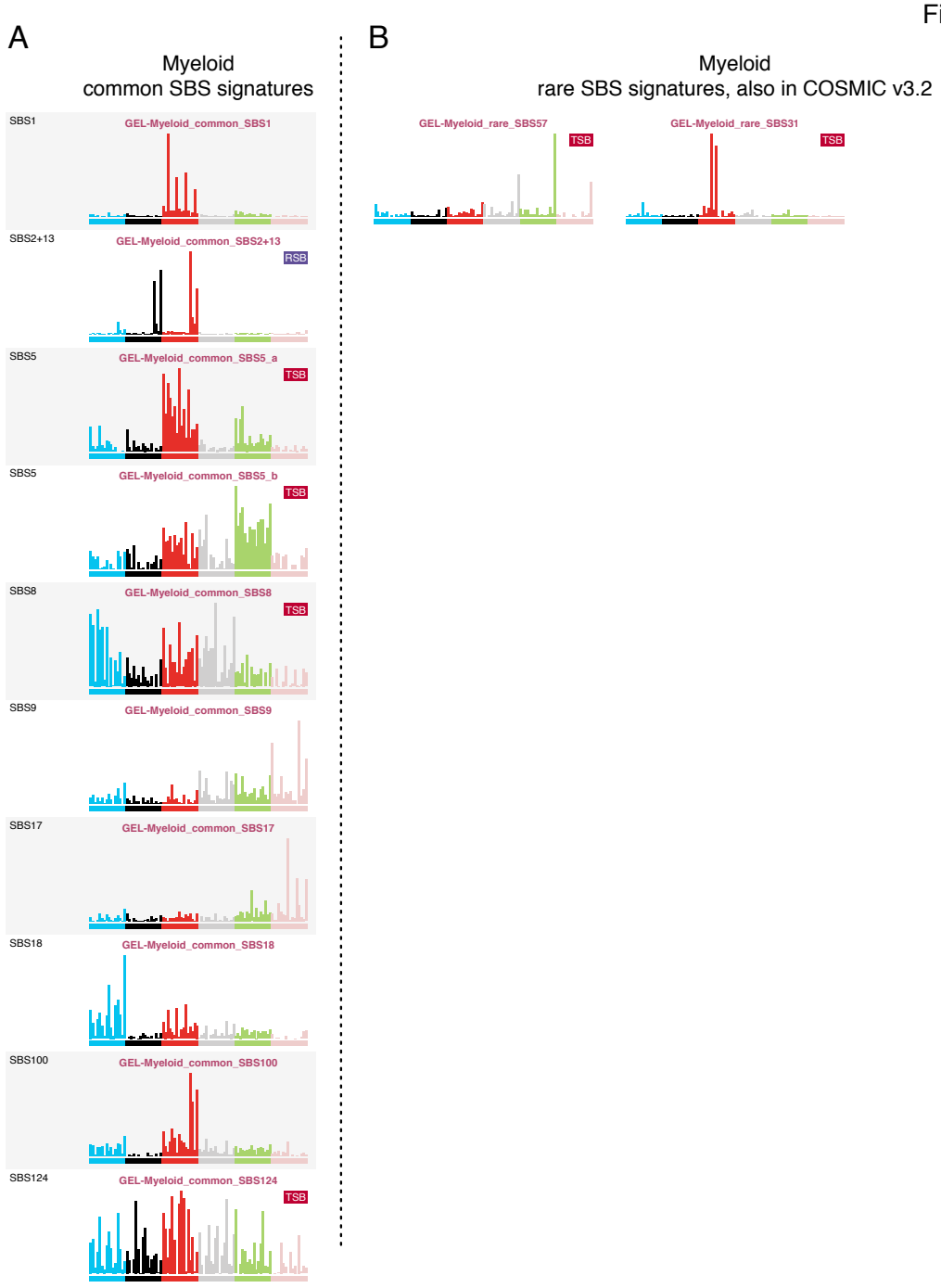


Fig. S44. Summary of myeloid cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2.

Figure S45

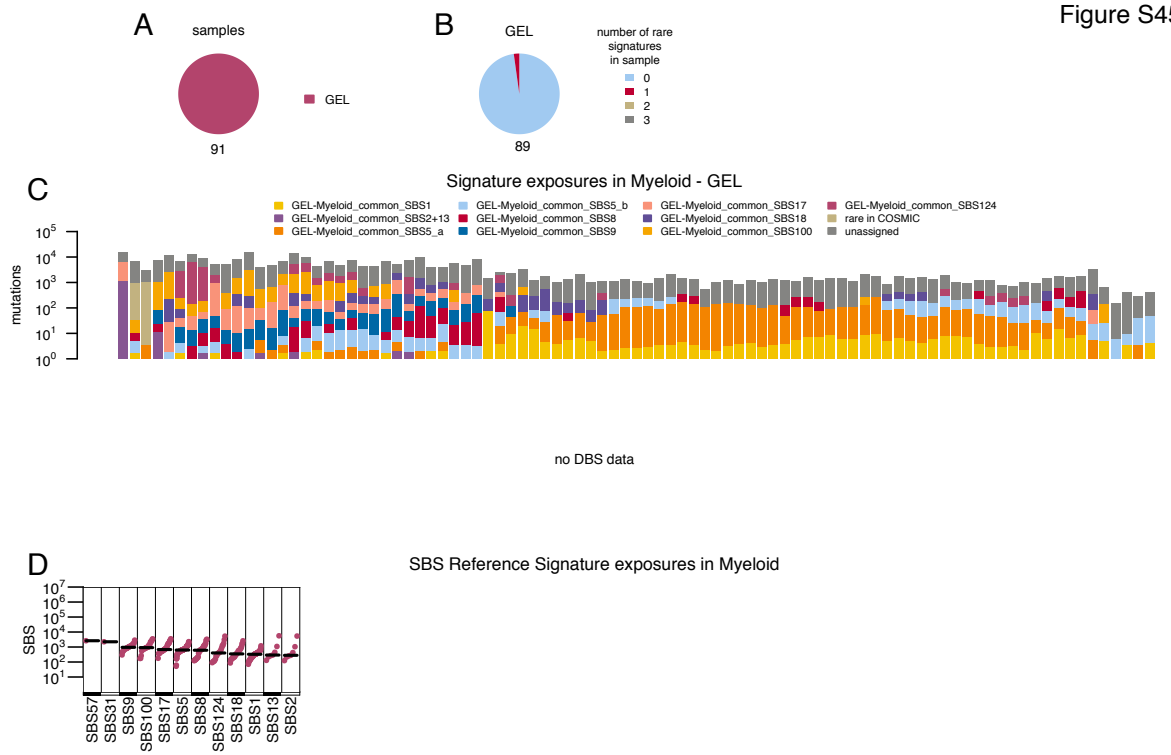


Fig. S45. Summary of signatures distribution in myeloid cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS reference signature across GEL samples.

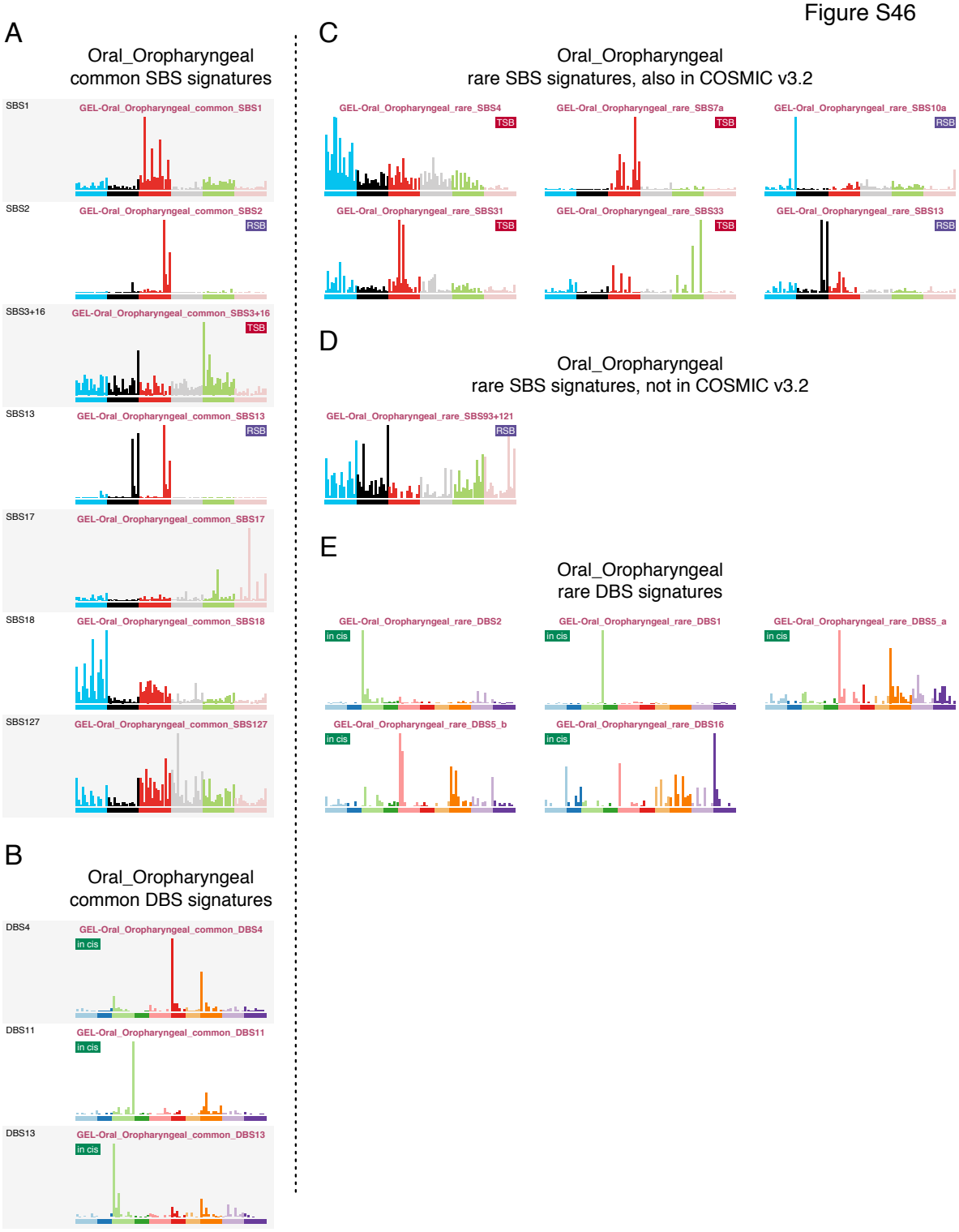


Fig. S46. Summary of oral and oropharyngeal cancer signatures extracted in this study. **(A)** Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

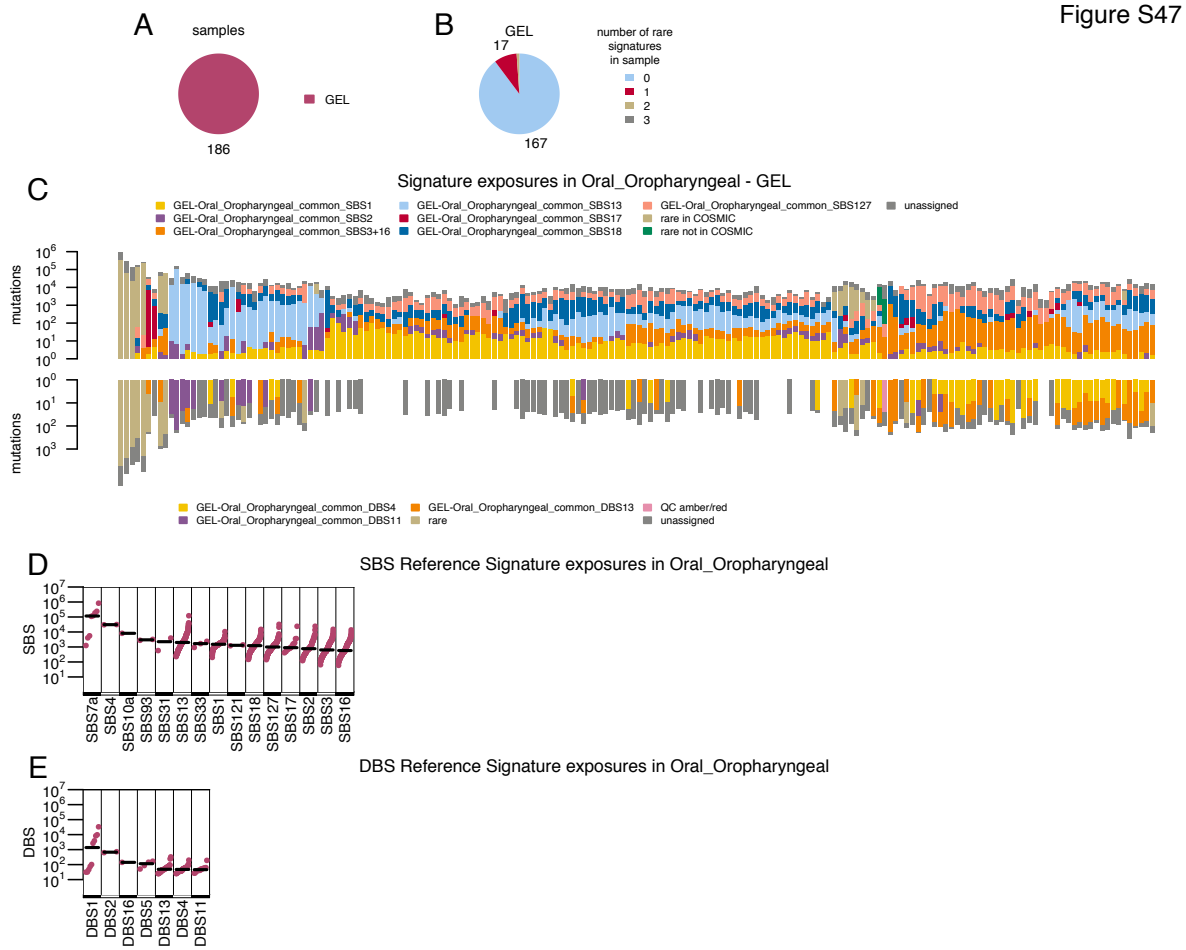


Fig. S47. Summary of signatures distribution in oral and oropharyngeal cancer samples. **(A)** Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in GEL. **(D)** Number of mutations associated with each SBS reference signature across GEL samples. **(E)** Number of mutations associated with each DBS reference signature across GEL samples.

Fig. S48. Summary of esophageal cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in

the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S49

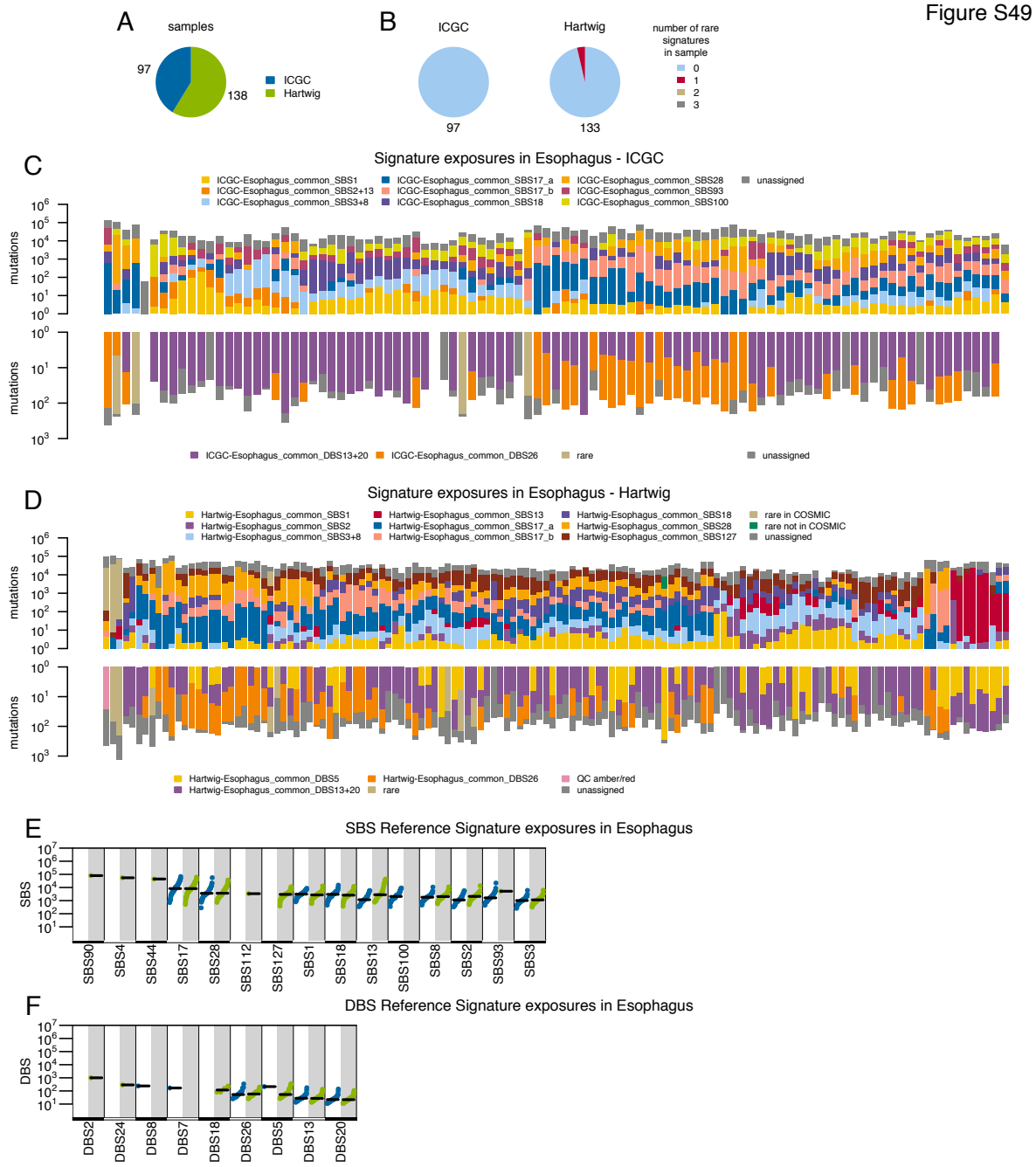


Fig. S49. Summary of signatures distribution in esophageal cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(E)** Number of mutations associated with each SBS reference signature across ICGC and Hartwig samples. **(F)** Number of mutations associated with each DBS reference signature across ICGC and Hartwig samples.

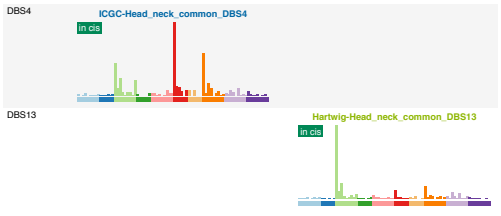
Figure S50

A

Head_neck
common SBS signatures

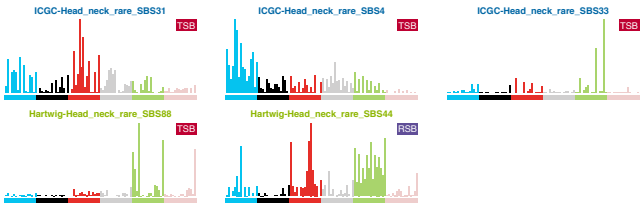
B

Head_neck
common DBS signatures



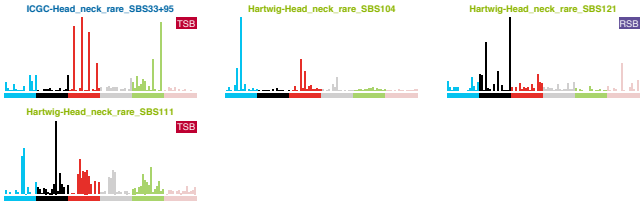
C

Head_neck
rare SBS signatures, also in COSMIC v3.2



D

Head_neck
rare SBS signatures, not in COSMIC v3.2



E

Head_neck
rare DBS signatures

Fig. S50. Summary of head and neck cancer signatures extracted in this study. (A) Common SBS signatures, **(B)** common DBS signatures, **(C)** rare SBS signatures that are also present in the COSMIC signatures dataset version 3.2, **(D)** rare SBS signatures not reported in the COSMIC signatures dataset version 3.2, **(E)** rare DBS signatures.

Figure S51

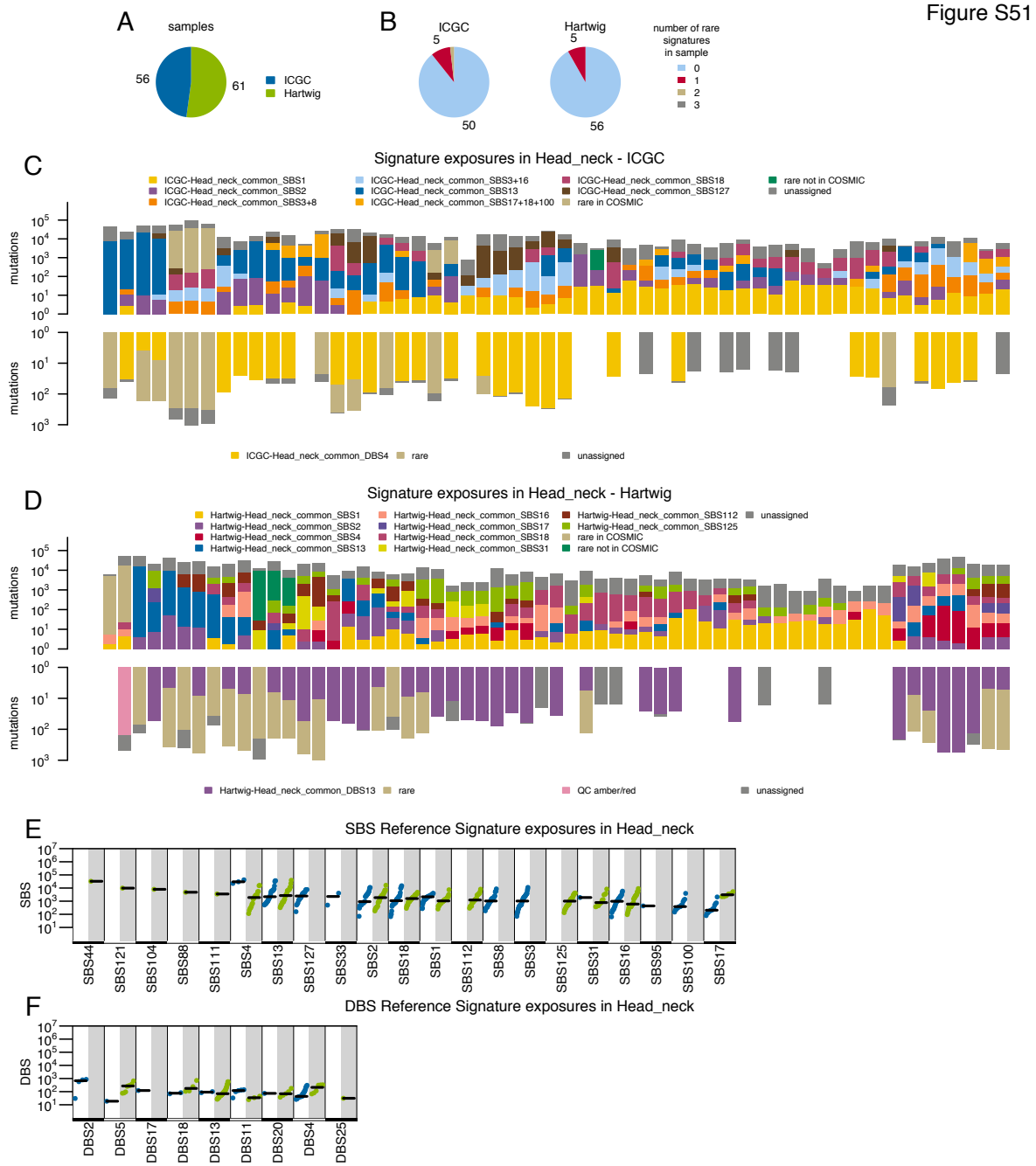


Fig. S51. Summary of signatures distribution in head and neck cancer samples. (A) Number of samples analyzed. **(B)** Proportion of samples with and without rare signatures. **(C)** Number of mutations associated with each SBS and DBS organ signature in each sample in ICGC. **(D)** Number of mutations associated with each SBS and DBS organ signature in each sample in Hartwig. **(E)** Number of mutations associated with each SBS reference signature across ICGC and Hartwig samples. **(F)** Number of mutations associated with each DBS reference signature across ICGC and Hartwig samples.

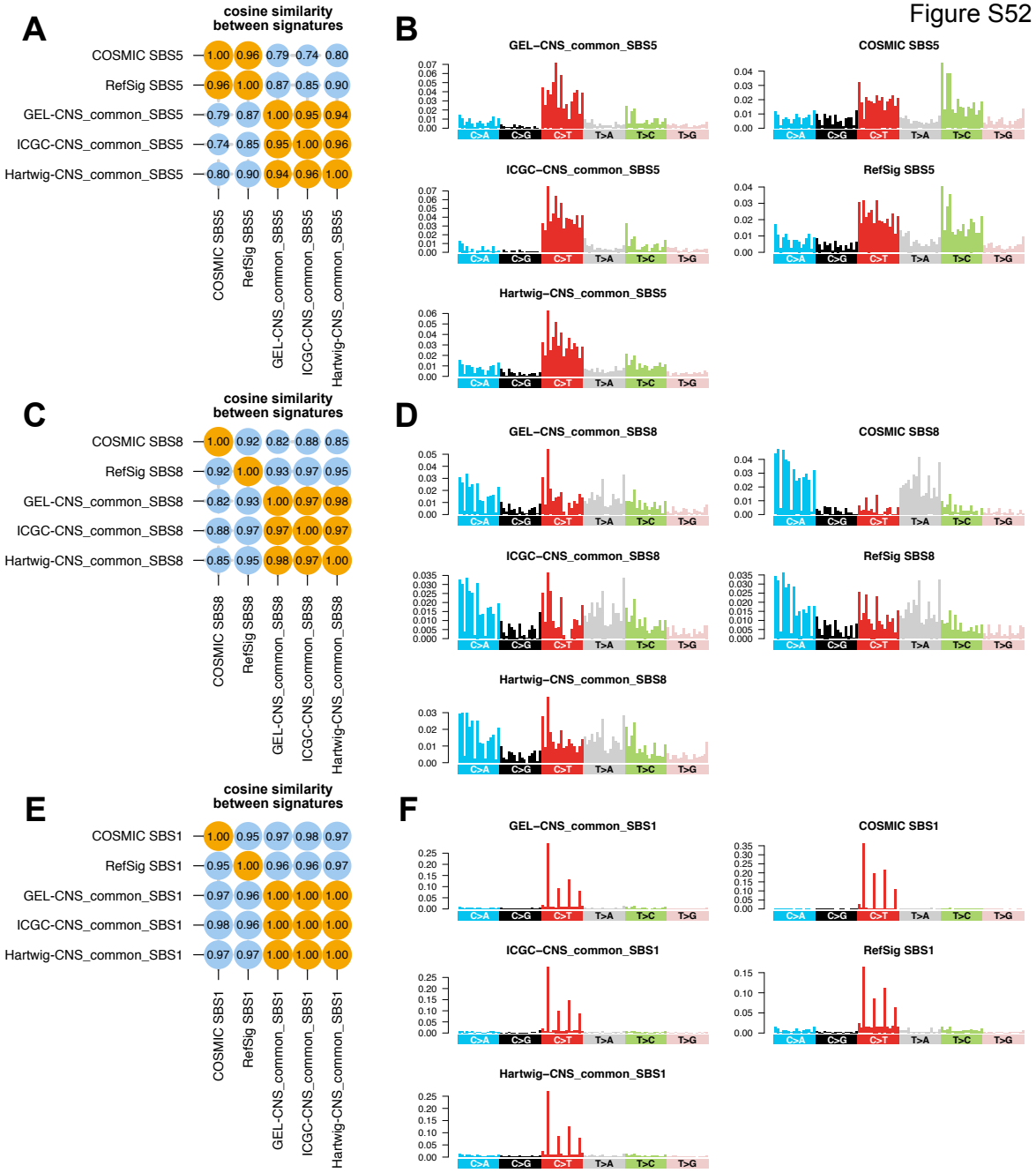


Fig. S52. Exploring SBS signatures in CNS tumors. Cosine similarities of comparisons between COSMIC SBS, RefSig SBS and tissue-specific SBS signatures are shown for SBS5 **(A)** and SBS5-associated signatures (**B**), SBS8 **(C)** and SBS8-associated signatures (**D**) and SBS1 **(E)** and SBS1-associated signatures (**F**). Signatures extracted in CNS in the three cohorts (GEL, ICGC and Hartwig) are more similar to one another than they are to the respective Reference Signatures and COSMIC version 3.2 signatures. Reference signatures and COSMIC signatures tend to be more similar to one another, as they are both averages of signatures extracted across many cohorts and organs.

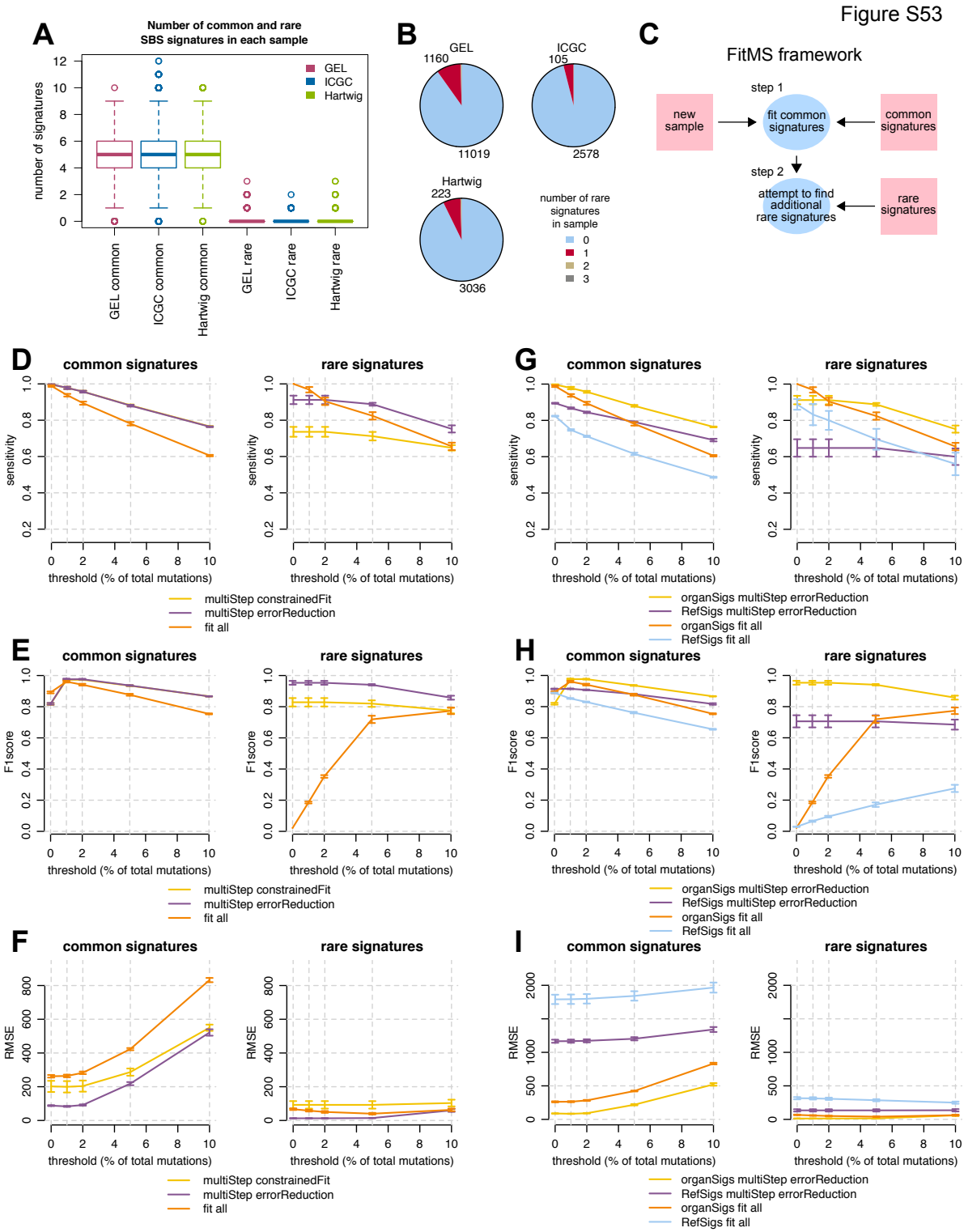


Fig. S53. Fitting mutational signatures: how to use common and rare signatures for signature assignments. Mutational signature fit with FitMS. (**A**) Numbers of common and rare SBS signatures found in each sample across the three cohorts. (**B**) Pie charts indicating the proportion of samples with one or more SBS rare signatures. (**C**) FitMS can be applied to individual new samples as a two-step approach. In the first step, common organ-specific signatures that were obtained from the same tissue of origin as the new sample are used. In a second step, FitMS attempts to identify additional rare signatures that may be present. (**D**) Simulation study sensitivity of two FitMS implementations (constrainedFit and errorReduction) and a simple "fit all" approach. **(E)** Same simulations as (D), F1 score. **(F)**

Same simulations as (D), root mean squared error (RMSE). (**G**) Simulation study sensitivity of the best FitMS strategy (errorReduction) and "fit all", using either the organ-specific signatures (organSigs) or the reference signatures (RefSigs) as common signatures. **(H)** Same simulations as (G), F1 score. **(I)** Same simulations as (G), RMSE. Error bars are standard error of the mean, number of repeats is n=5.

Table legends

Table S1. Count of somatic mutations in GEL, ICGC and Hartwig cohorts. Overall count of single nucleotide variants (SNV), double nucleotide variants (DNV) small insertion and deletions (Indels), and rearrangements (Rearr), across the GEL, ICGC and Hartwig cohorts, along with the number of samples used. For the GEL cohort, we indicate the proportion of samples that had PCR cycles in the library preparation.

Table S2. GEL high quality filters for samples. Filters used to select high quality samples in the GEL cohort.

Table S3. Number of samples used for SBS analysis in each cohort and in each organ.

Table S4. Number of samples used for DBS analysis in each cohort and in each organ.

Table S5. Mapping of organ names used in this study to organ/tissue/study organization of the GEL, ICGC and Hartwig cohorts.

Table S6. Full list of samples from GEL, ICGC and Hartwig. Summary of all samples used in this study with the corresponding organ. We also indicate which samples were used for the common signature extraction of SBSs and DBSs.

Table S7. SBS mutational catalogs. SBS mutational catalogs of all samples analyzed in this study.

Table S8. SBS mutational catalogs. DBS mutational catalogs of all samples analyzed in this study.

Table S9. Organ-specific SBS signatures.

Table S10. Organ-specific DBS signatures.

Table S11. Number of common and rare SBS signatures extracted in each cohort and each organ.

Table S12. Number of common and rare DBS signatures extracted in each cohort and each organ.

Table S13. Clustering of all organ-specific SBS signatures.

Table S14. Clustering of all organ-specific DBSs signatures.

Table S15. SBS distinct patterns. Obtained as the average of the clusters obtained clustering all organ-specific SBS signatures (table S13).

Table S16. DBS distinct patterns. Obtained as the average of the clusters obtained clustering all organ-specific DBS signatures (table S14).

Table S17. SBS distinct patterns info table. Annotation of SBS distinct patterns into recurrent, mixed and singleton, with corresponding reference signature name.

Table S18. DBS distinct patterns info table. Annotation of DBS distinct patterns into recurrent, mixed and singleton, with corresponding reference signature name.

Table S19. SBS reference signatures info table. Summary table of all SBS reference signatures identified in this study. We report QC status, proposed etiology, transcription and replication strand bias, number of samples with the signatures and other annotations.

Table S20. DBS reference signatures info table. Summary table of all DBS reference signatures identified in this study. We report QC status, proposed etiology, number of samples with the signatures and other annotations.

Table S21. SBS reference signatures.

Table S22. DBS reference signatures.

Table S23. SBS reference signatures exposures. Number of mutations associated with each reference signature in all samples analyzed in this study.

Table S24. DBS reference signatures exposures. Number of mutations associated with each reference signature in all samples analyzed in this study.

Table S25. SBS conversion matrix. Conversion matrix mapping organ-specific SBS signatures into reference signatures.

Table S26. DBS conversion matrix. Conversion matrix mapping organ-specific DBS signatures into reference signatures.

Table S27. Examples of trinucleotide mutational catalogs from GEL, ICGC and Hartwig cohorts.

Table S28. MBD4 and OGG1 G308E driver mutations occurring in samples in GEL.

Table S29. Substitution and indels driver mutations for selected signatures and genes

Table S30. Rearrangements driver mutations for selected signatures and genes

Table S31. HRDetect input features and scores for the GEL cohort.

Table S32. Transcription and Replication strand bias of SBS reference signatures in the GEL cohort.

Table S33. SBS common and rare signatures to be used with FitMS. Common and rare signatures are provided for each organ analyzed in this study.