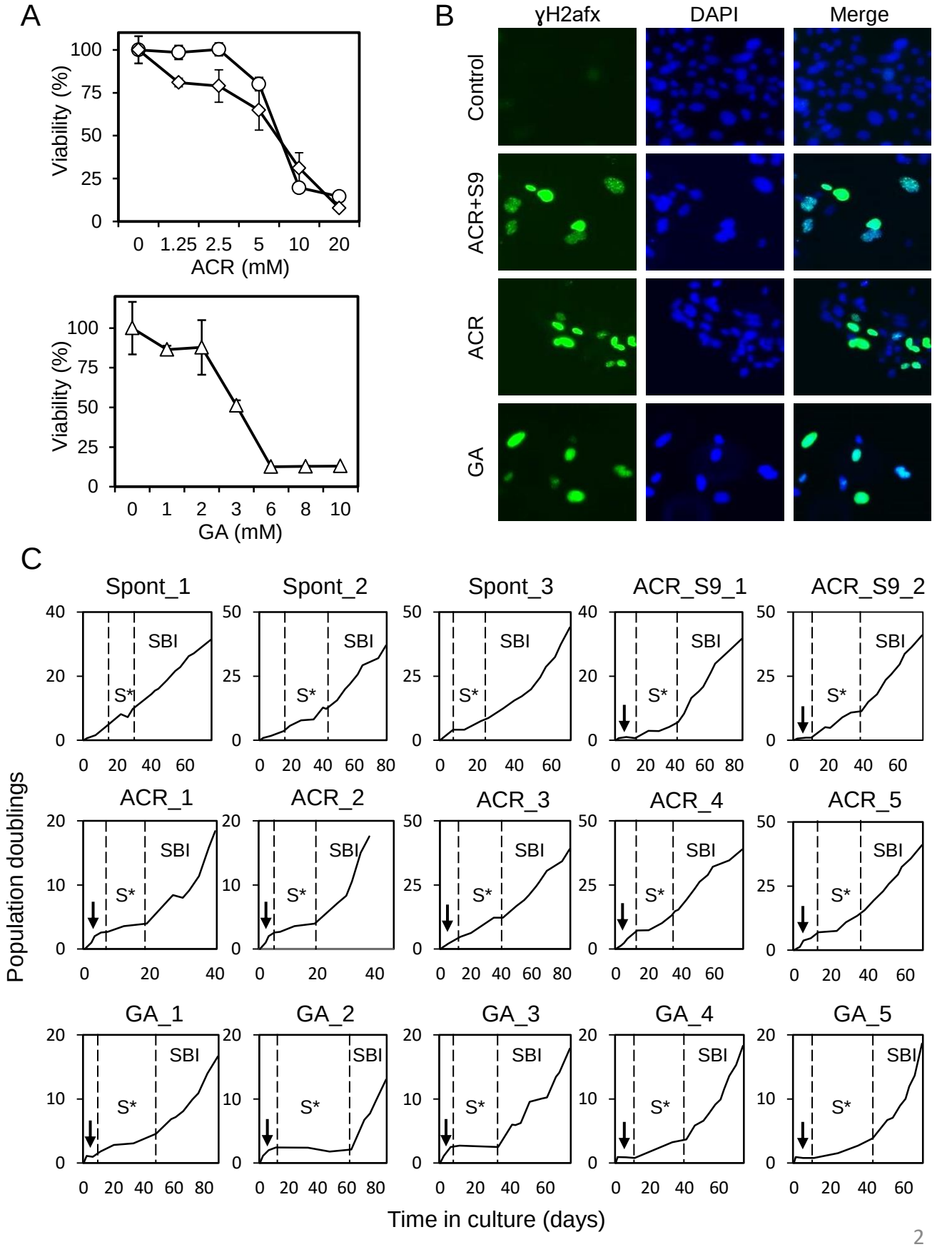
Supplemental Figures S1 to S11

"Experimental and pan-cancer genome analyses reveal widespread contribution of acrylamide exposure to carcinogenesis in humans"

Zhivagui M. *et al*, 2019

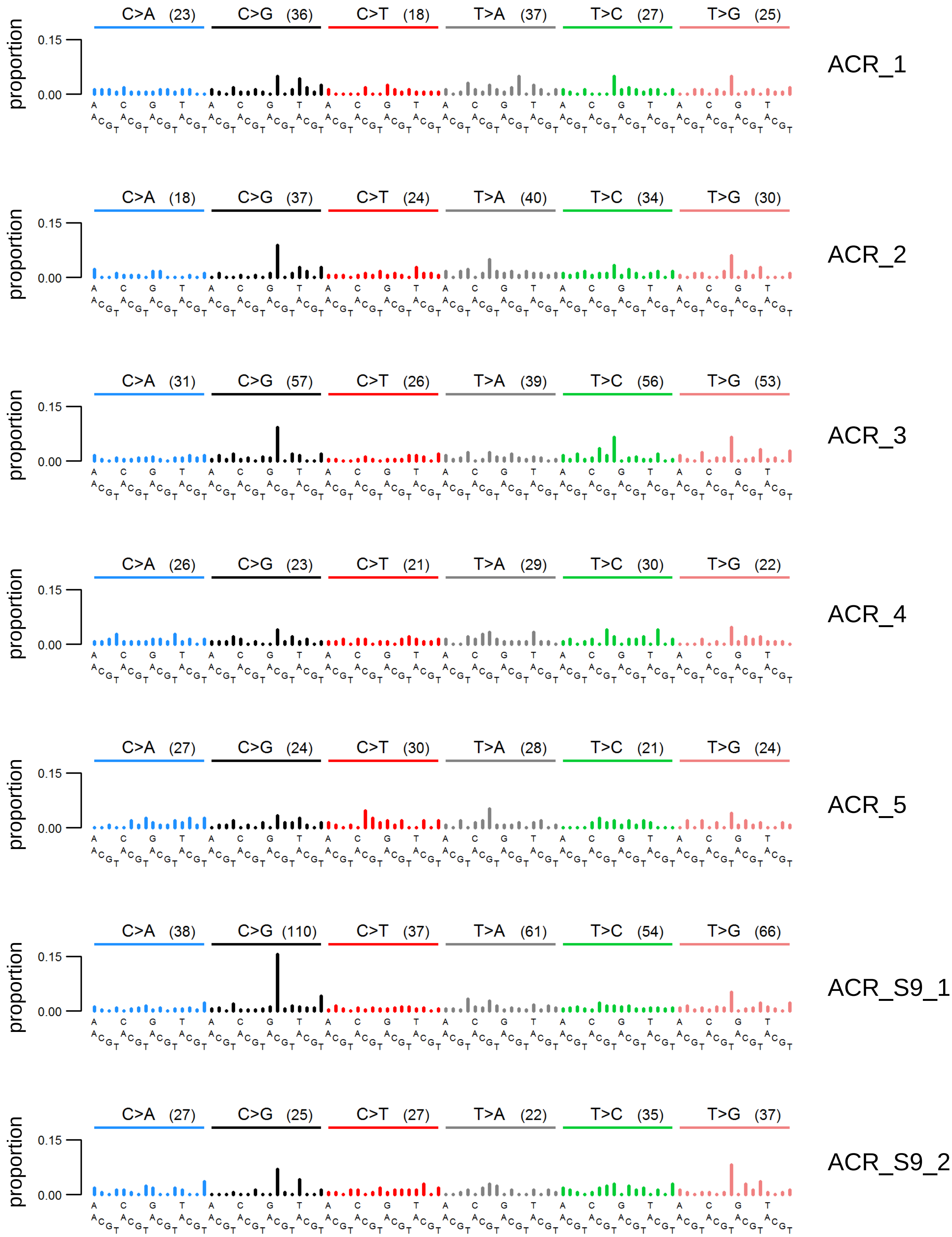
Suppl. Fig. S1


A
Viability (%)
ACR (mM)
GA (mM)
B
γH2afx
DAPI
Merge
Control
ACR+S9
ACR
GA
C
Spont_1
Spont_2
Spont_3
ACR_S9_1
ACR_S9_2
ACR_1
ACR_2
ACR_3
ACR_4
ACR_5
GA_1
GA_2
GA_3
GA_4
GA_5
SBI
S*
Population doublings
Time in culture (days)

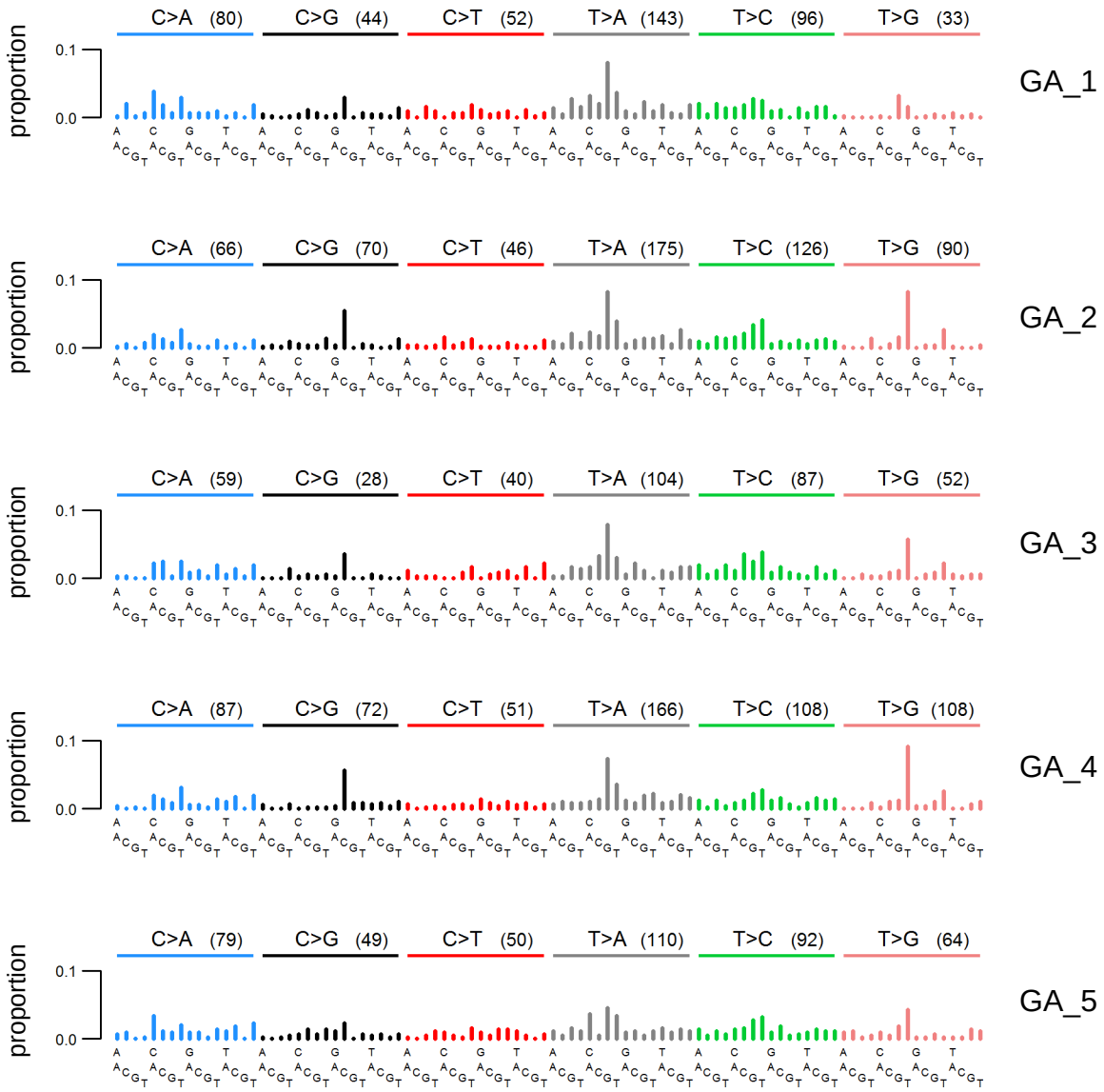
Supplemental Fig. S1: Acrylamide- and glycidamide-induced cytotoxicity and genotoxicity and *in vitro* immortalization. (**A**) Cell viability, following 24-hour treatment of primary MEFs with the indicated concentrations of acrylamide (top panel), in the absence (diamonds) and presence (circles) of human S9 fraction, and glycidamide (bottom panel), as determined by MTT assay. Absorbance was measured 48 hours after treatment cessation and was normalized to untreated cells. The results are expressed as mean percent ±SD of three replicates. (**B**) DNA damage assessment by immunofluorescence with an antibody specific for Ser139-phosphorylated histone H2Ax (ɣH2Ax). Primary MEFs were treated with acrylamide or glycidamide for 24 hours prior to immunofluorescence. Compound concentrations used were based on 20-70% viability reduction in the MTT assay: 10 mM acrylamide, 5 mM acrylamide in the presence of S9 fraction and 3 mM glycidamide. ACR: acrylamide; GA: glycidamide. (**C**) Primary cells were either left untreated (Spont) or were exposed to acrylamide (ACR±S9) or glycidamide (GA). X-axis represents days in culture. Y-axis represents the cumulative doubling populations. The dashed vertical lines demarcate the senescence stage. Arrow: compound exposure; S*: senescence; SBI: senescence bypass/immortalization.

Suppl. Fig. S2

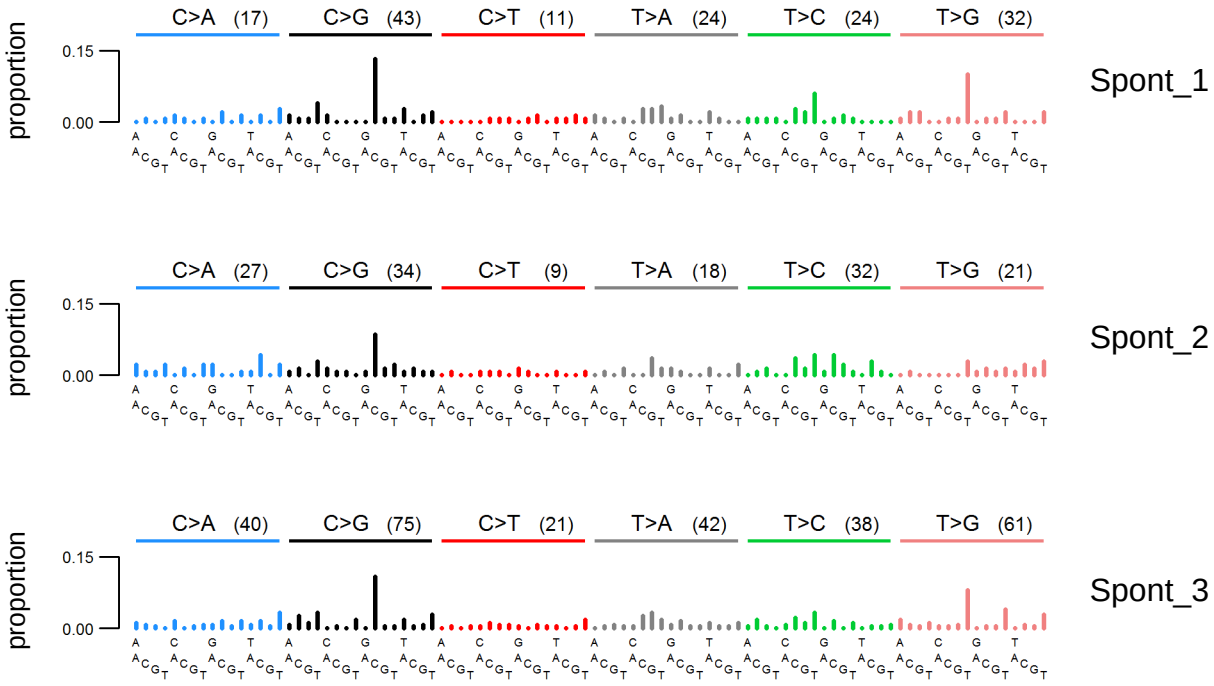
A



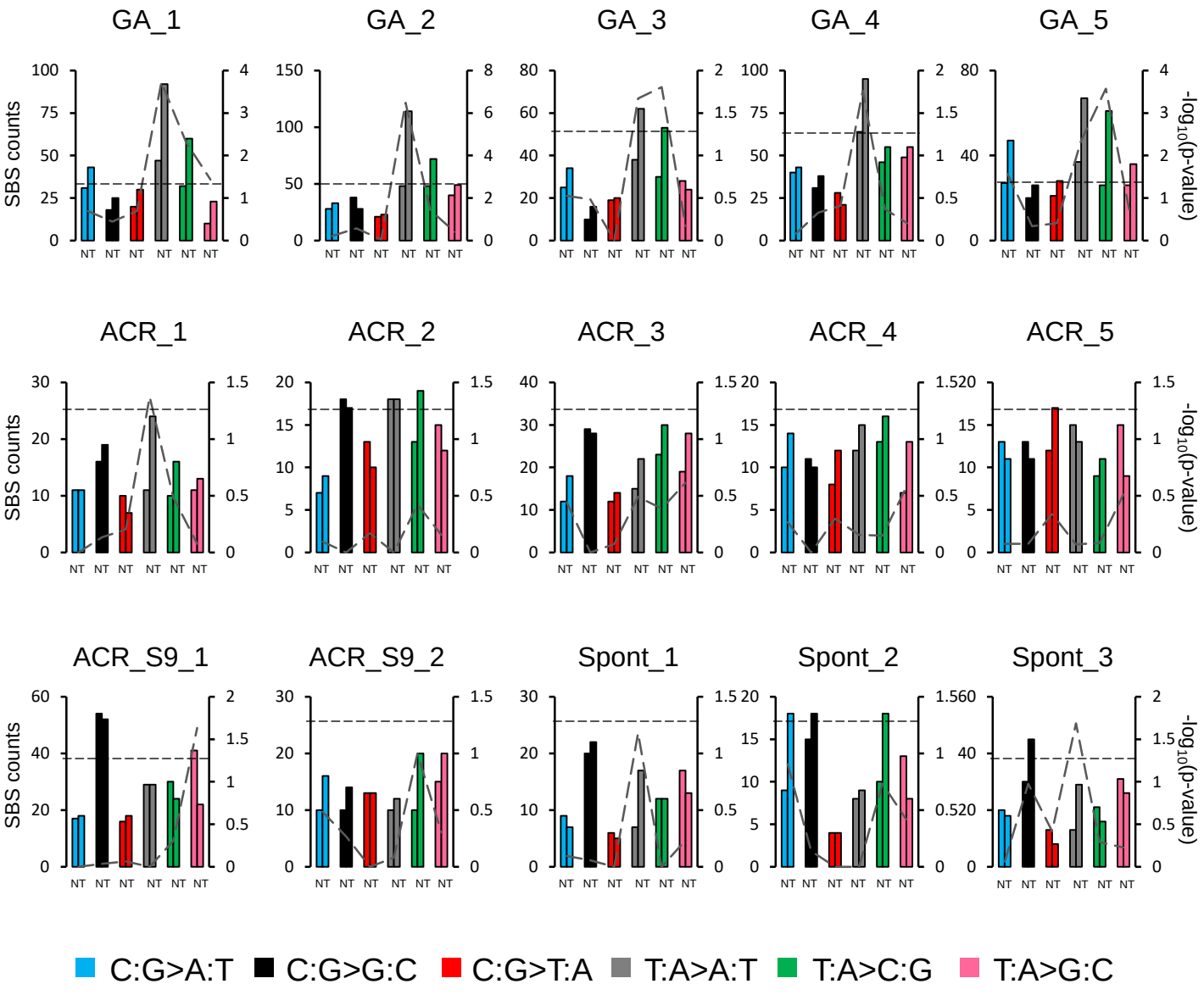
B


C>A (80)
C>G (44)
C>T (52)
T>A (143)
T>C (96)
T>G (33)
proportion
0.1
0.0
GA_1
C>A (66)
C>G (70)
C>T (46)
T>A (175)
T>C (126)
T>G (90)
proportion
0.1
0.0
GA_2
C>A (59)
C>G (28)
C>T (40)
T>A (104)
T>C (87)
T>G (52)
proportion
0.1
0.0
GA_3
C>A (87)
C>G (72)
C>T (51)
T>A (166)
T>C (108)
T>G (108)
proportion
0.1
0.0
GA_4
C>A (79)
C>G (49)
C>T (50)
T>A (110)
T>C (92)
T>G (64)
proportion
0.1
0.0
GA_5

C

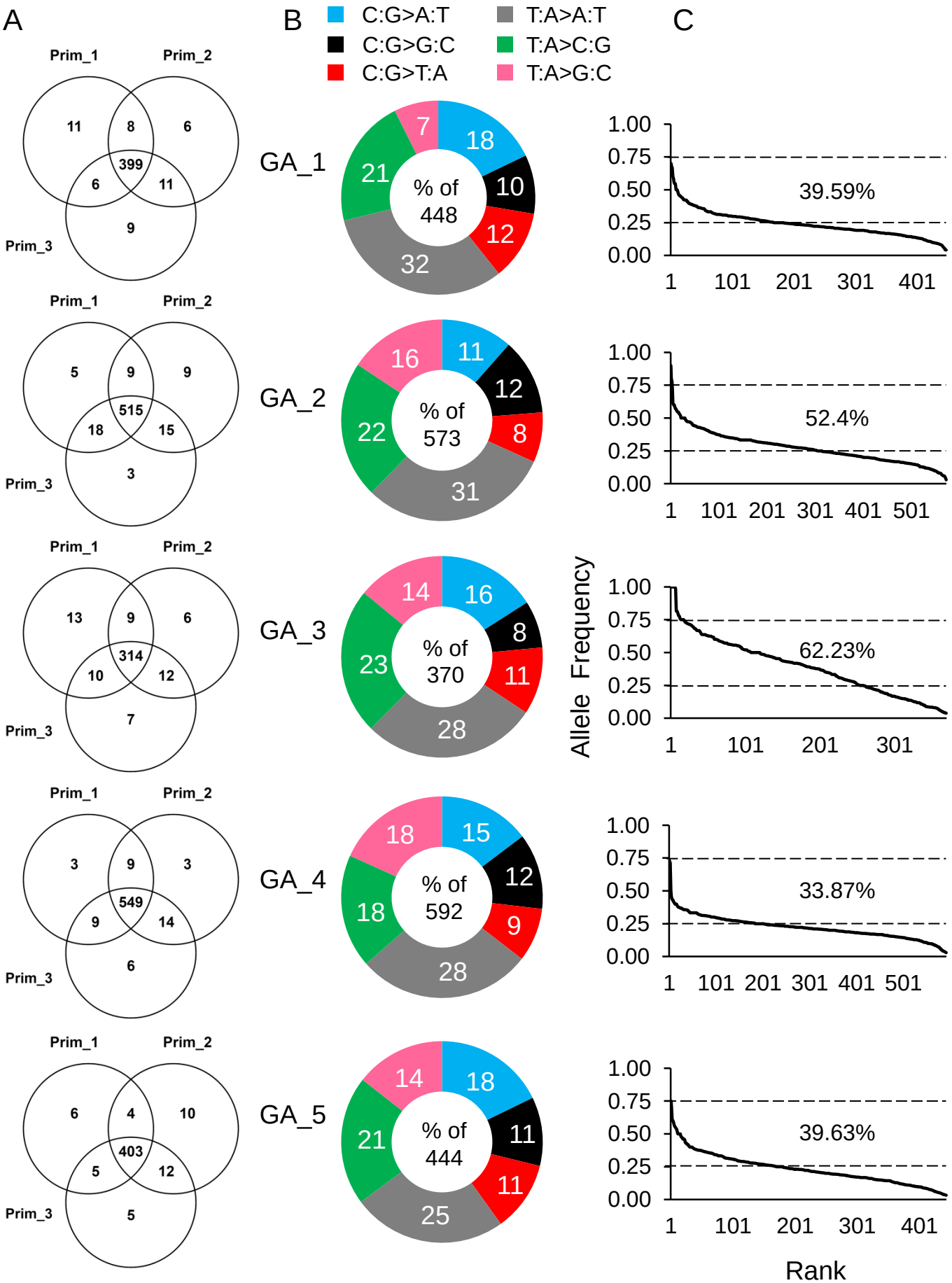


Supplemental Fig. S2: Individual sample mutation spectra derived from exome sequencing data from immortalized Hupki MEF clones derived from exposure to (**A**) acrylamide (ACR) or (**B**) glycidamide (GA), or (**C**) by spontaneous immortalization (Spont). X-axis represents the trinucleotide sequence context. Y-axis represents the frequency distribution of the mutations in each context.



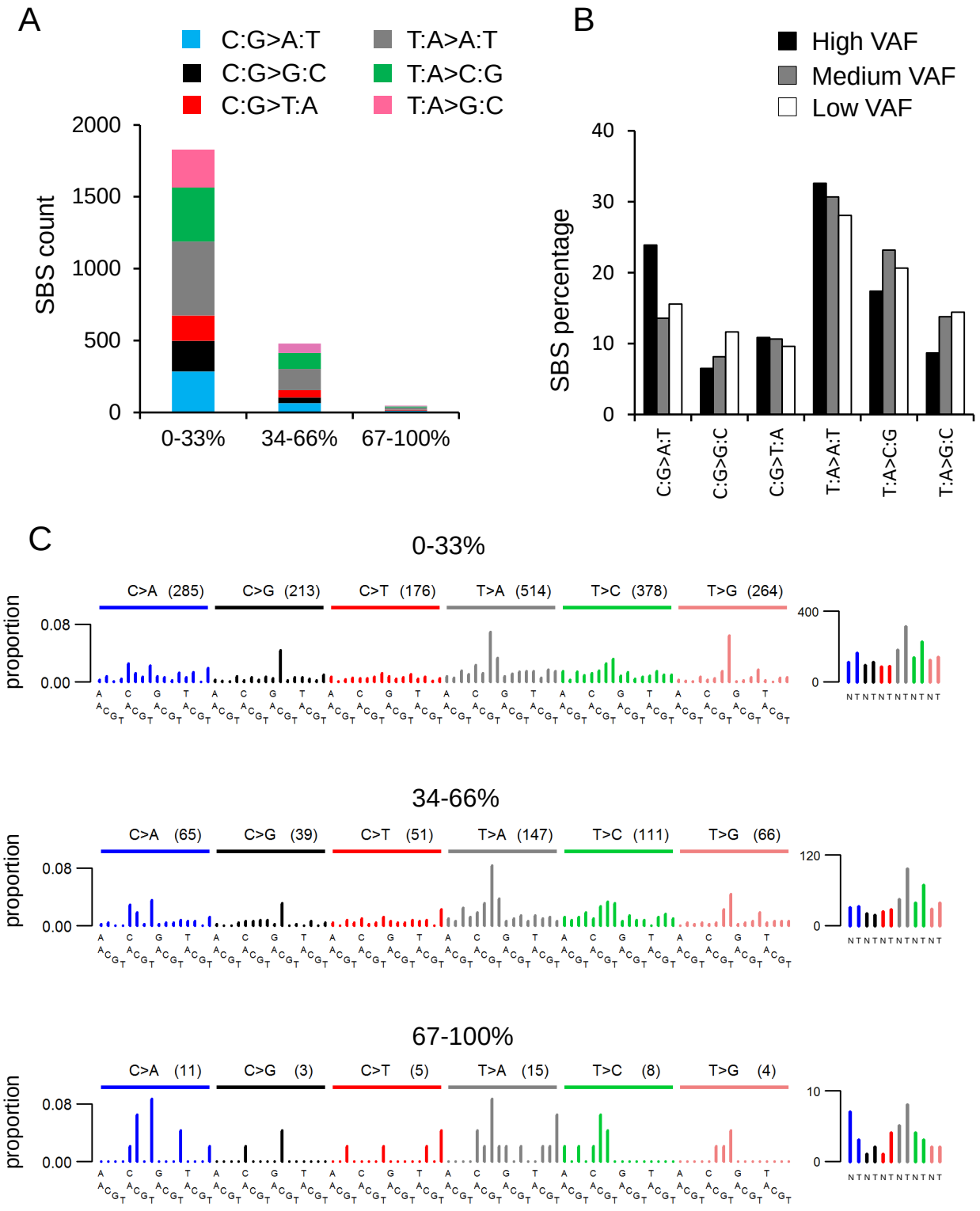
Supplemental Fig. S3: Illustration of the transcription strand bias derived from the analysis of exome sequencing data from immortalized Hupki MEF cell lines. GA: glycidamide-derived clones; ACR: acrylamide-derived clones; Spont: spontaneously immortalized clones. The six mutation types are represented by different colors. For each mutation type, the number of mutations occurring on the transcribed (T) and non-transcribed (N) strand, as well as the p-values for strand bias is shown on the y-axes. The dashed grey line in each graph indicates the p-values for strand bias for each mutation type. The horizontal, dashed black line represents a significance threshold of $p < 0.05$.

Suppl. Fig. S4


A
B
C
C:G>A:T
C:G>G:C
C:G>T:A
T:A>A:T
T:A>C:G
T:A>G:C
Prim_1
Prim_2
Prim_3
11
8
6
399
6
11
9
GA_1
% of 448
7
18
10
12
32
21
39.59%
5
9
9
515
18
15
3
GA_2
% of 573
16
11
12
8
31
22
52.4%
13
9
6
314
10
12
7
GA_3
% of 370
14
16
8
11
28
23
62.23%
3
9
3
549
9
14
6
GA_4
% of 592
18
15
12
9
28
18
33.87%
6
4
10
403
5
12
5
GA_5
% of 444
14
18
11
11
25
21
39.63%
Allele Frequency
1.00
0.75
0.50
0.25
0.00
1
101
201
301
401
501
Rank

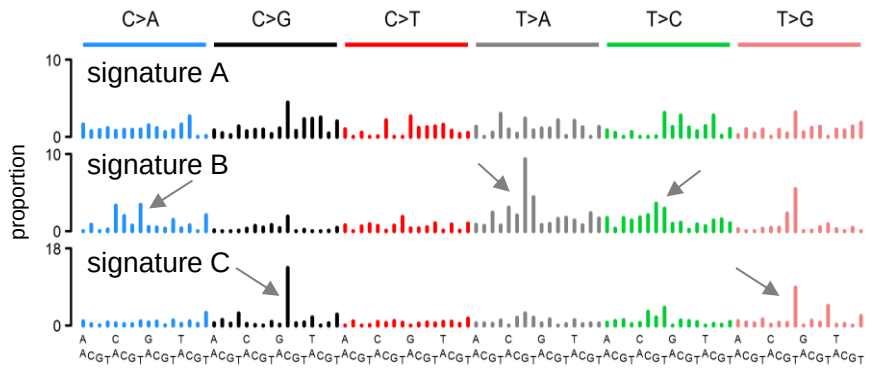
Supplemental Fig. S4: Single-base substitution characteristics summary of the GA clones. (**A**) Venn diagrams show the overlap of variants called in glycidamide (GA)-derived clones after normalization to three different batches of primary Hupki MEF cells (Prim_1, Prim_2, and Prim_3). (**B**) Percentages of the six mutation types, color-coded, among all mutations identified in GA clones. The overall mutation number for each sample is indicated in the centre of the pie chart. (**C**) Distribution of mutations based on their allelic frequencies in the five glycidamide (GA)-derived clones. Mutations in individual cell lines were ranked and plotted based on decreasing allelic frequency. Percentage of mutations with allelic frequency between 25% and 75% is indicated.

Suppl. Fig. S5

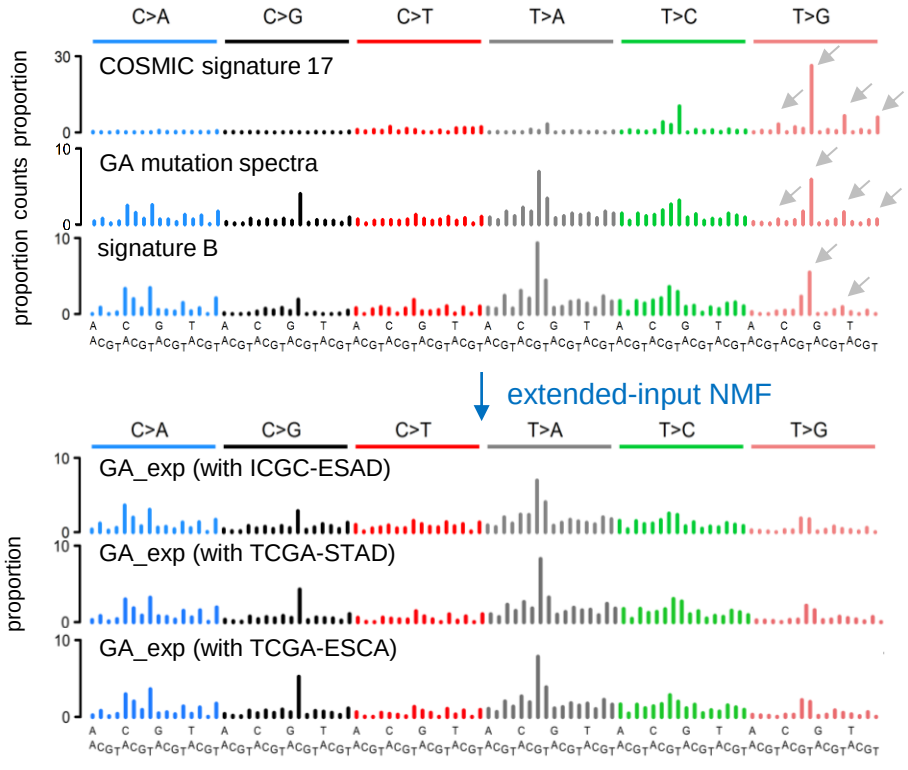

A
C:G>A:T
C:G>G:C
C:G>T:A
T:A>A:T
T:A>C:G
T:A>G:C
SBS count
2000
1500
1000
500
0
0-33%
34-66%
67-100%
B
High VAF
Medium VAF
Low VAF
SBS percentage
40
30
20
10
0
C:G>A:T
C:G>G:C
C:G>T:A
T:A>A:T
T:A>C:G
T:A>G:C
C
0-33%
C>A (285)
C>G (213)
C>T (176)
T>A (514)
T>C (378)
T>G (264)
proportion
0.08
0.00
400
0
NTNTNTNTNTNT
34-66%
C>A (65)
C>G (39)
C>T (51)
T>A (147)
T>C (111)
T>G (66)
proportion
0.08
0.00
120
0
NTNTNTNTNTNT
67-100%
C>A (11)
C>G (3)
C>T (5)
T>A (15)
T>C (8)
T>G (4)
proportion
0.08
0.00
10
0
NTNTNTNTNTNT

Supplemental Fig. S5: Mutation type and mutation spectra analysis with respect to variant allele frequency (VAF). The analysis was carried out using exome sequencing data from immortalized Hupki MEF clones derived from exposure to glycidamide. (**A**) Mutation counts were stratified into three VAF bins ([0-33% = low VAF]; [34-66% = medium VAF]; [67-100% = high VAF]). (**B**) The relative contribution of the six mutation types to the overall number of mutations in each VAF bin is shown on the y-axis. (**C**) Mutation spectra (left) and strand bias (right) analysis for the different VAF bins. Mutation spectra analysis: X-axis represents the trinucleotide sequence context. Y-axis represents the frequency distribution of the mutations. The counts for each mutation type are indicated in parentheses. Strand bias analysis: For each mutation type, the number of mutations occurring on the transcribed and non-transcribed strand is shown on the y-axis. T: transcribed strand; N: non-transcribed strand.

A



B

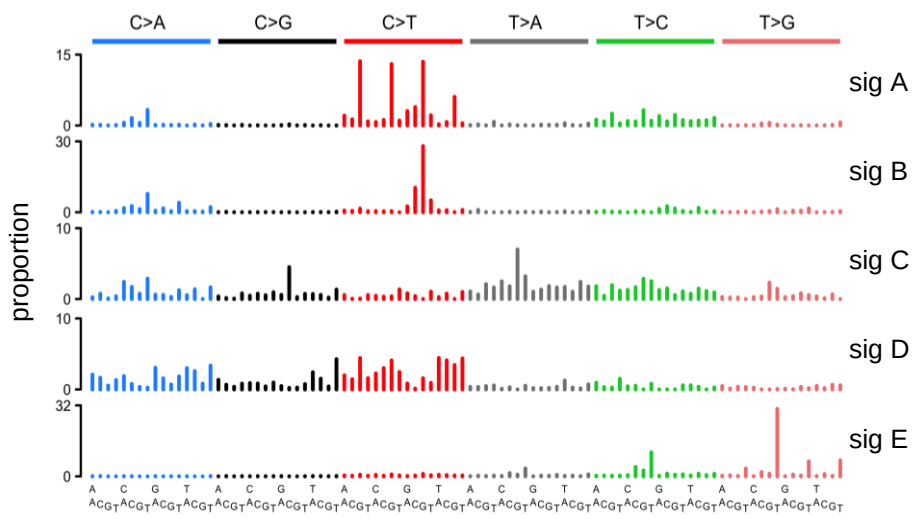


C

	Sig. 17 peaks	with ICGC-ESAD	with TCGA-STAD	with TCGA-ESCA
C[T>A]C	0	12.42	7.01
C[T>A]G	25.7	11.52	16.13
C[T>A]T	8	26.81	12.45
C[T>C]C	22.4	18.92	19.88
C[T>C]G	29.7	15.35	20.96
C[T>C]T	20	6.66	31.65
A[T>G]T	100	96.33	100
C[T>G]C	42.9	21.78	34.3
C[T>G]G	19.1	5.85	4.09
C[T>G]T	68.2	71.54	64.38
G[T>G]T	52.7	41.4	42.41
T[T>G]T	0	0	0

Supplemental Fig. S6: Extraction of the glycidamide (GA) mutational signature. (**A**) Mutational signatures identified by non-negative matrix factorization (NMF) in the 15 Hupki MEF-derived clones (signature A, signature B, and signature C, corresponding to enrichment bar-graphs in Fig. 1B). X-axis represents the trinucleotide sequence context. Y-axis represents the frequency distribution of the mutations. The predominant trinucleotide context for T:A > A:T mutations is indicated in sig B (5'-CTG-3'), alongside two other predominant mutation types (T:A>C:G, C:G>A:T) induced by GA treatment. The trinucleotide contexts for the background C:G > G:C (5'-GCC-3') and T:A > G:C mutations (5'-NTT-3') are highlighted in sig C. (**B**) The background signature 17 (top track) marked by the arrows is observed in GA mutation spectra (Suppl. Fig. S3) as well as in signature B. The extended NMF input included either 17-rich ICGC ESAD data, or TCGA ESCA (enriched for or lacking signature 17) or TCGA STAD (enriched for or lacking signature 17) and gave rise to the corresponding GA_exp signatures (lower right). The ICGC and TCGA public sample sets are listed in the Methods section. (**C**) Quantification of the efficiency of reducing the signature 17 background during glycidamide-mutational signature extraction (see Methods). The heat-map table indicates the final proportionate reduction of the signature 17-specific peaks after conducting NMF with input extended with the ICGC or TCGA data sets.

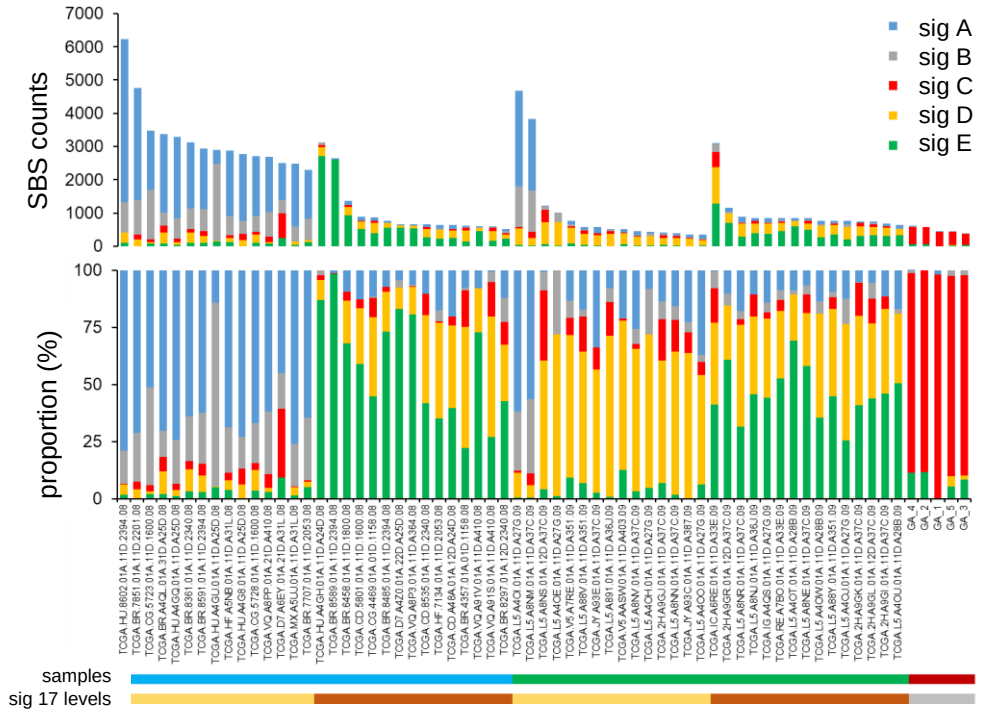
A



B

	A	B	C	D	E
SBS1	0.90	0.44	0.03	0.44	0.05
SBS2	0.04	0.04	0.06	0.45	0.03
SBS3	0.25	0.23	0.71	0.64	0.23
SBS4	0.16	0.24	0.57	0.49	0.05
SBS5	0.44	0.29	0.56	0.68	0.22
SBS6	0.89	0.72	0.10	0.44	0.06
SBS7a	0.10	0.06	0.10	0.54	0.04
SBS7b	0.17	0.07	0.14	0.54	0.04
SBS7c	0.09	0.05	0.31	0.18	0.14
SBS7d	0.11	0.06	0.22	0.19	0.09
SBS8	0.14	0.19	0.65	0.48	0.10
SBS9	0.21	0.16	0.46	0.36	0.57
SBS10a	0.04	0.11	0.14	0.27	0.02
SBS10b	0.27	0.08	0.07	0.38	0.07
SBS11	0.18	0.20	0.13	0.46	0.03
SBS12	0.22	0.09	0.42	0.18	0.17
SBS13	0.01	0.02	0.13	0.39	0.01
SBS14	0.21	0.30	0.32	0.27	0.03
SBS15	0.73	0.94	0.13	0.31	0.06
SBS16	0.12	0.06	0.31	0.27	0.07
SBS17a	0.09	0.02	0.32	0.08	0.38
SBS17b	0.02	0.06	0.13	0.02	0.94
SBS18	0.20	0.26	0.34	0.60	0.06
SBS19	0.18	0.14	0.17	0.42	0.05
SBS20	0.34	0.41	0.37	0.21	0.06
SBS21	0.20	0.14	0.25	0.09	0.08
SBS22	0.04	0.02	0.68	0.10	0.07
SBS23	0.25	0.27	0.13	0.38	0.04
SBS24	0.21	0.36	0.31	0.48	0.07
SBS25	0.36	0.23	0.82	0.53	0.21
SBS26	0.28	0.13	0.43	0.15	0.22
SBS27	0.07	0.07	0.28	0.14	0.03
SBS28	0.04	0.03	0.05	0.07	0.41
SBS29	0.23	0.29	0.31	0.54	0.03
SBS30	0.24	0.17	0.14	0.58	0.06
SBS31	0.19	0.11	0.35	0.47	0.09
SBS32	0.29	0.20	0.18	0.51	0.05
SBS33	0.16	0.07	0.20	0.10	0.06
SBS34	0.07	0.05	0.24	0.12	0.02
SBS35	0.16	0.23	0.63	0.43	0.10
SBS36	0.10	0.20	0.29	0.50	0.02
SBS37	0.22	0.11	0.39	0.26	0.26
SBS38	0.12	0.20	0.31	0.21	0.03
SBS39	0.19	0.14	0.50	0.54	0.14
SBS40	0.23	0.22	0.59	0.77	0.29
SBS41	0.22	0.16	0.38	0.36	0.23
SBS42	0.27	0.41	0.34	0.39	0.07
SBS43	0.04	0.07	0.22	0.09	0.14
SBS44	0.51	0.55	0.36	0.28	0.09
SBS45	0.14	0.23	0.34	0.26	0.01
SBS46	0.26	0.14	0.48	0.30	0.29
SBS47	0.07	0.08	0.28	0.32	0.10
SBS48	0.01	0.02	0.02	0.09	0.00
SBS49	0.05	0.09	0.15	0.11	0.00
SBS50	0.11	0.11	0.24	0.49	0.07
SBS51	0.15	0.08	0.29	0.34	0.09
SBS52	0.03	0.06	0.11	0.29	0.01
SBS53	0.11	0.11	0.33	0.23	0.04
SBS54	0.29	0.12	0.34	0.09	0.12
SBS55	0.06	0.07	0.20	0.12	0.06
SBS56	0.06	0.12	0.19	0.33	0.02
SBS57	0.21	0.11	0.27	0.31	0.22
SBS58	0.18	0.11	0.30	0.51	0.12
SBS59	0.02	0.15	0.14	0.21	0.05
SBS60	0.02	0.04	0.07	0.04	0.04
Hupki_AA	0.12	0.10	0.81	0.22	0.18
Hupki_AID	0.36	0.44	0.27	0.36	0.11
Hupki_BaP	0.28	0.25	0.64	0.54	0.29
Hupki_MNNG	0.22	0.25	0.17	0.50	0.05
Mouse_DMBA	0.07	0.07	0.74	0.15	0.11
Mouse_MNU	0.19	0.19	0.15	0.50	0.04
Mouse_Urethane	0.29	0.20	0.73	0.36	0.23

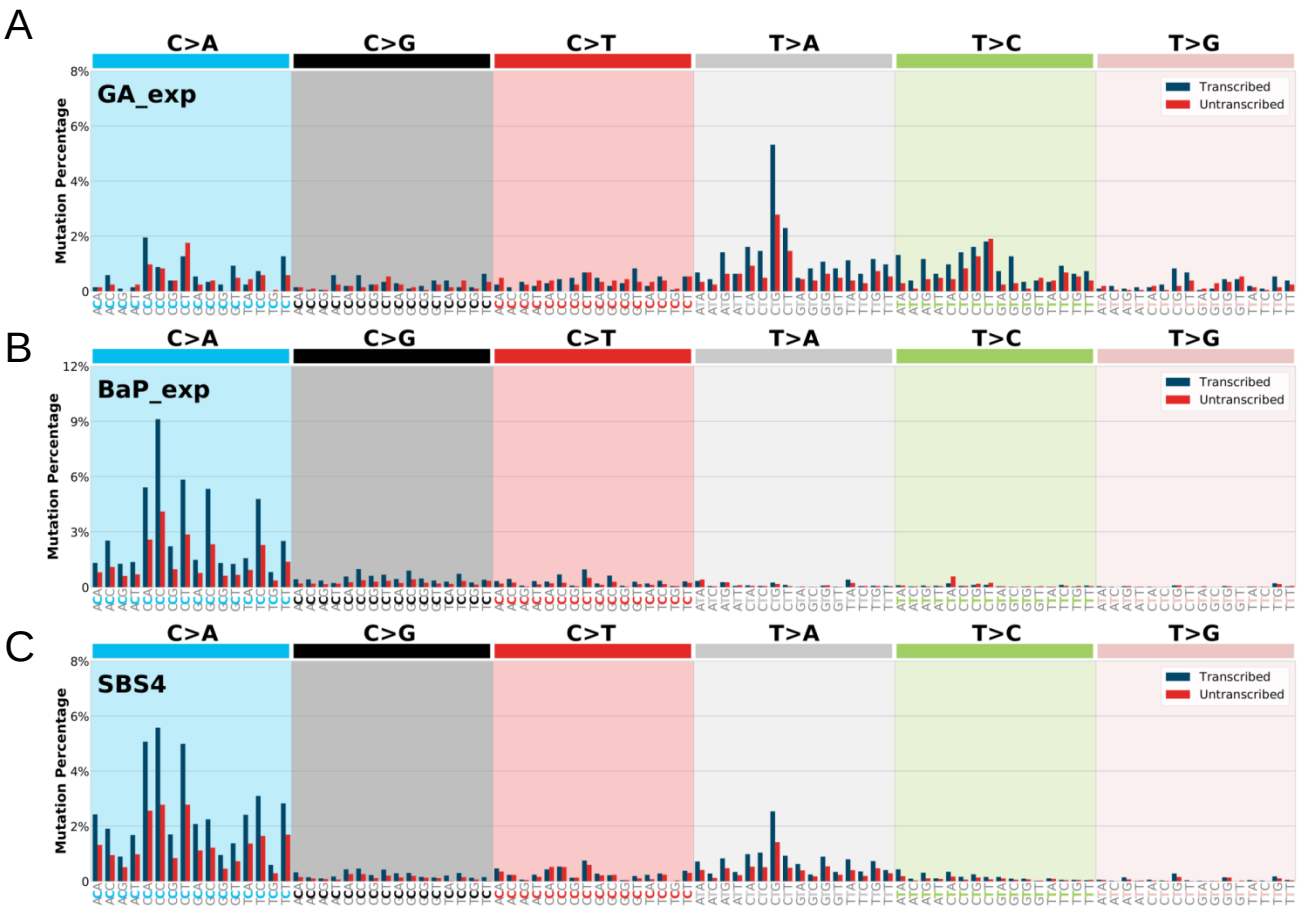
C



samples
- GA clones
- TCGA ESCA_EAC
- TCGA STAD

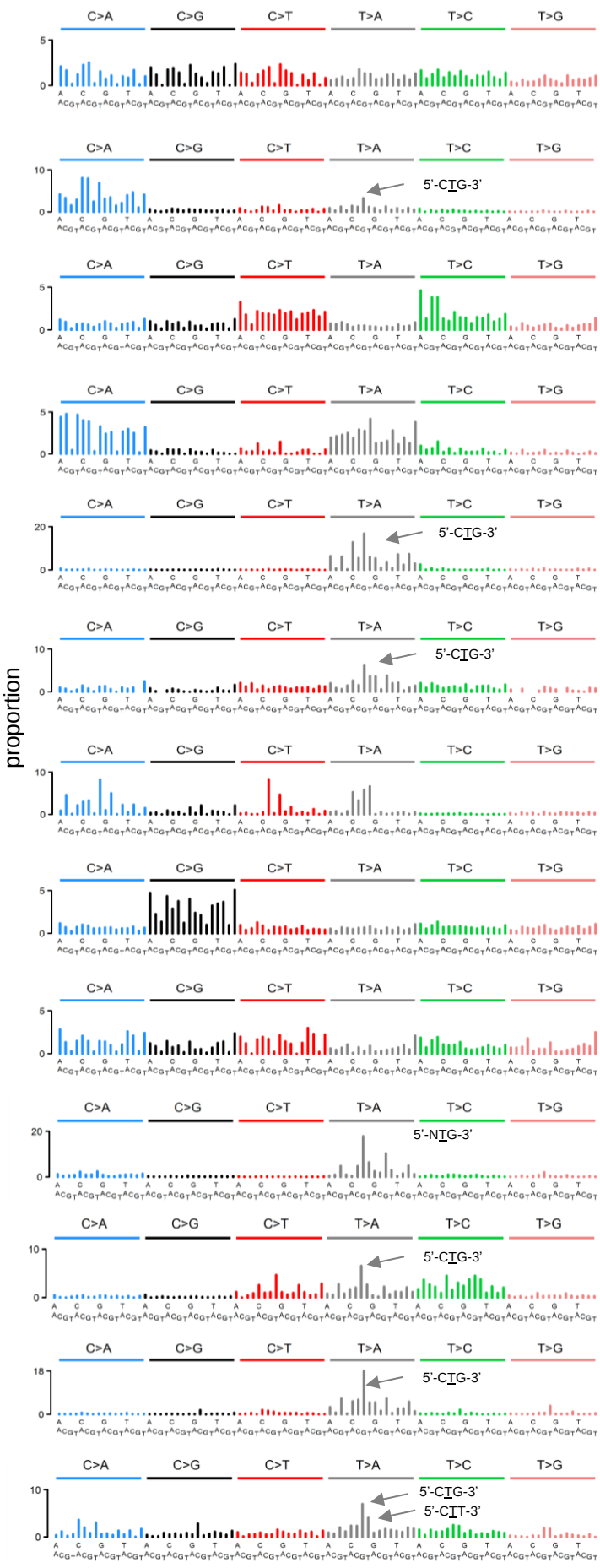
sig 17 levels
- High sig 17
- Low sig 17
- background sig 17 contents

Supplemental Fig. S7: Extraction of experimental mutational signature of GA in MEF cells using NMF, with signature 17-positive and signature 17-negative TCGA ESCA (EAC) and TCGA STAD datasets added to the input (see Methods and Supplemental Methods for details). (**A**) Principal signatures extracted from the mixed input containing GA-treated, clonally immortalized MEF cells and 60 TCGA samples. (**B**) Analysis of cosine similarity of signatures obtained in (A) with PCAWG-7 SBS signatures and select experimental data. (**C**) An enrichment (as SBS counts, upper bar graph, or as relative contents, lower bar graph) of the mutational signatures shown in (A) across all analyzed samples, with signature C being predominantly enriched in the MEF GA clones.



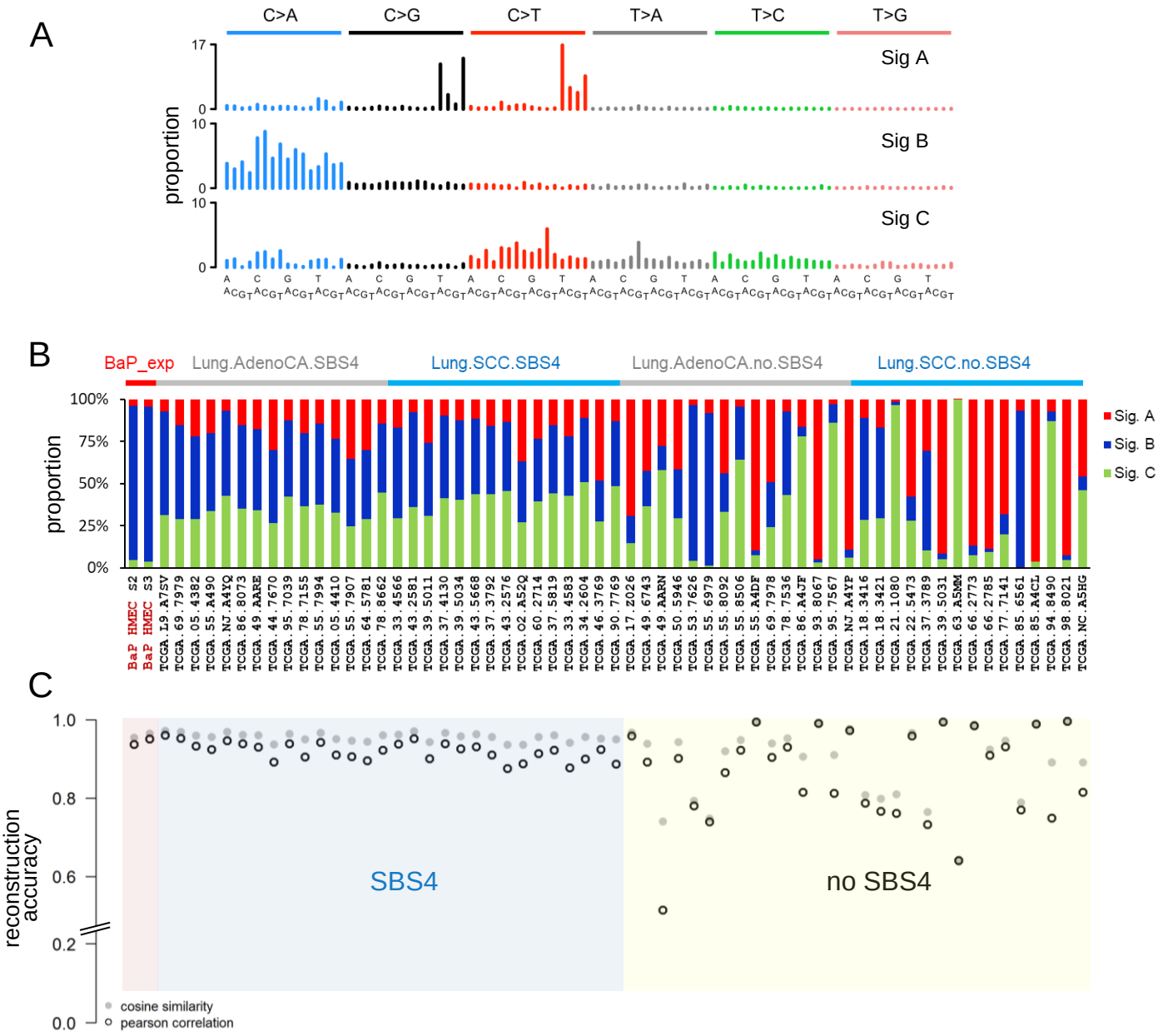
Supplemental Fig. S8: Transcription strand-biased (TSB) patterns of GA_exp signature from MEF cells (**A**). BaP_exp signature from HMEC cells (**B**). PCAWG tobacco smoking SBS4 signature (**C**). The enrichment of mutations on transcribed strand reflects the pyrimidine convention used to describe the SBS classes (C>A, C>G, C>T, T>A, T>C, T>G). The underlying matrices for each TSB signature are listed in Supplemental Table S4.

Suppl. Fig. S9



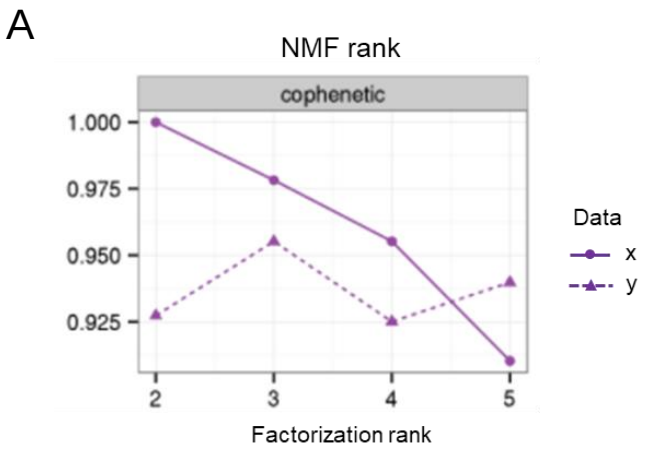
Etiology	Sample origin	SBS signature	Strand Bias
BRCA1 and BRCA2 mutations	Breast, ovarian and pancreatic cancer	SBS3	No SB
Smoking	Lung cancer	SBS4	C>A; C>G; T>A; T>C
Clock-like	All cancer types	SBS5	T>C
Unknown	Breast and brain cancer	SBS8	C>A
AA	Urothelial carcinoma	SBS22	T>A
Unknown	Hodgkin's lymphomas	SBS25	T>A
Unknown	-	SBS35	C>A
Unknown	-	SBS39	No SB
Unknown	-	SBS40	T>A
DMBA	Skin cancer (mouse)	-	T>A
Urethane	Lung cancer (mouse)	-	C>T; T>A; T>C
AA	Hupki MEF	-	T>A
Glycidamide	Hupki MEF	-	C>A; T>A; T>C

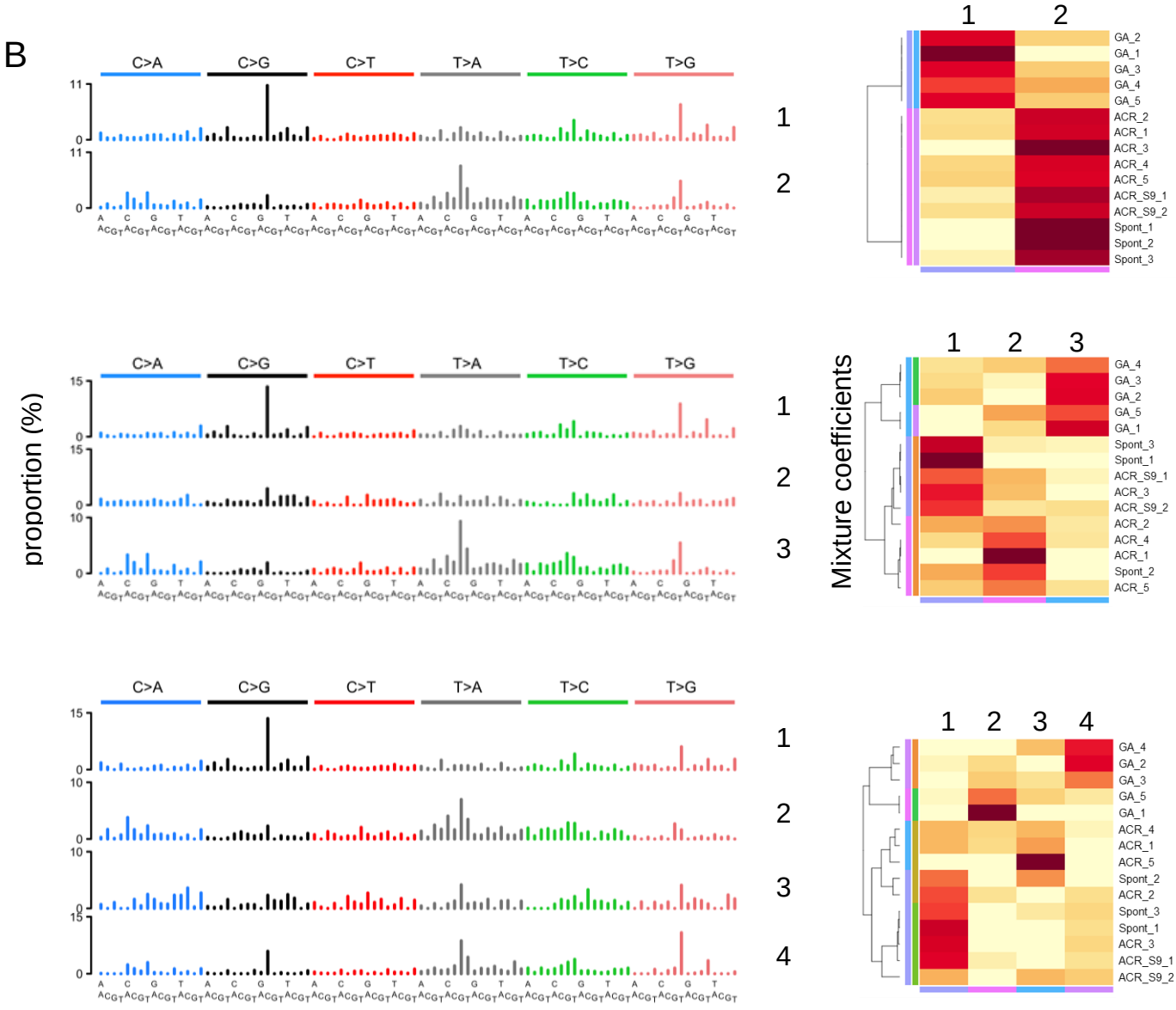
Supplemental Fig. S9: T:A>A:T enriched mutational signatures used for cosine similarity analysis (see also Fig. 2A). The individual signatures were originally derived from human cancer sequencing data or experimental models (animal bioassays, cell lines) of carcinogen exposure. X-axis represents the trinucleotide sequence context. Y-axis represents the frequency distribution of the mutations. The predominant trinucleotide contexts for T:A>A:T mutations are indicated by arrows. AA: aristolochic acid; DMBA: 7,12-dimethylbenz[a]anthracene.



Supplemental Fig. S10: Extraction of experimental mutational signature of B[a]P modelled in human HMEC cells and extracted by NMF by using SBS4-positive and SBS4 negative lung tumor datasets. (**A**) Three main signatures extracted from the mixed input containing B[a]P-treated, clonally immortalized HMEC cells and 60 TCGA lung cancer samples (see Methods for details). (**B**) Proportionate enrichment of the signatures shown in (A) across all analyzed samples, with signature B being essentially the single signature observed in the B[a]P HMEC clones (S2, S3). (**C**) Reconstruction accuracy analysis of the identified mutational signatures in individual samples is shown in the dot plot (Y-axis value of 1 = 100% accuracy). See Methods for additional details.

Suppl. Fig. S11





Supplemental Fig. S11: Estimation of mutational signature number for NMF analysis of the MEF clones. (**A**) The cophenetic correlation coefficient from factorization rank analysis indicates a minimum of 2 signatures should be extracted. (**B**) The setting and results for 2, 3 and 4 signatures to be extracted are shown, as signatures (histograms) and the separation of samples based on corresponding signature enrichment (heat maps, hierarchical clustering based on mixture coefficient). Refer to the Supplemental Methods file for further details.