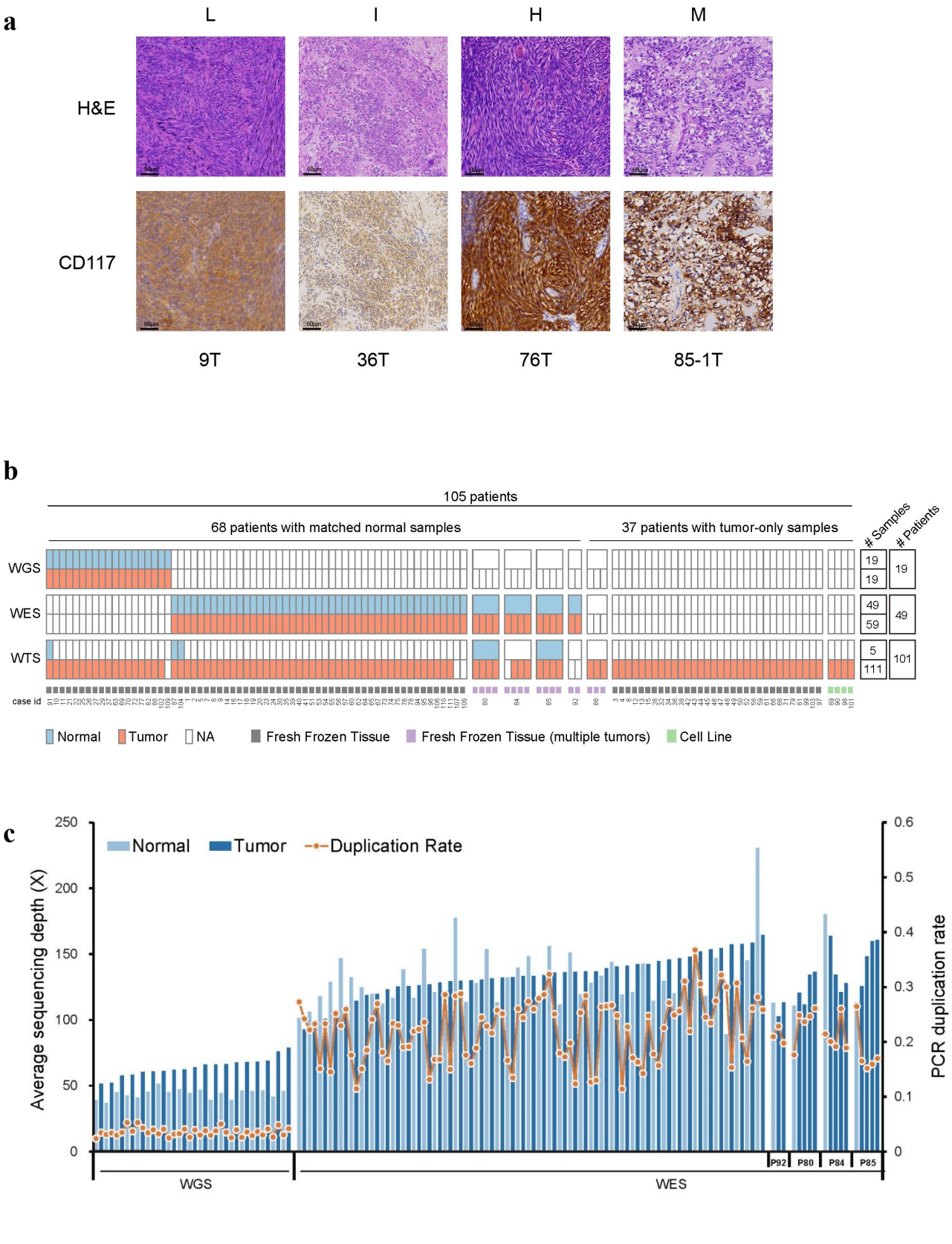
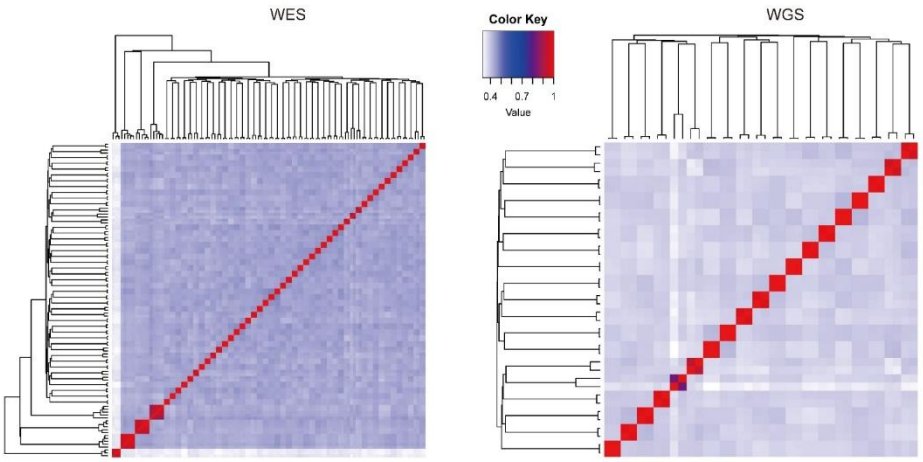
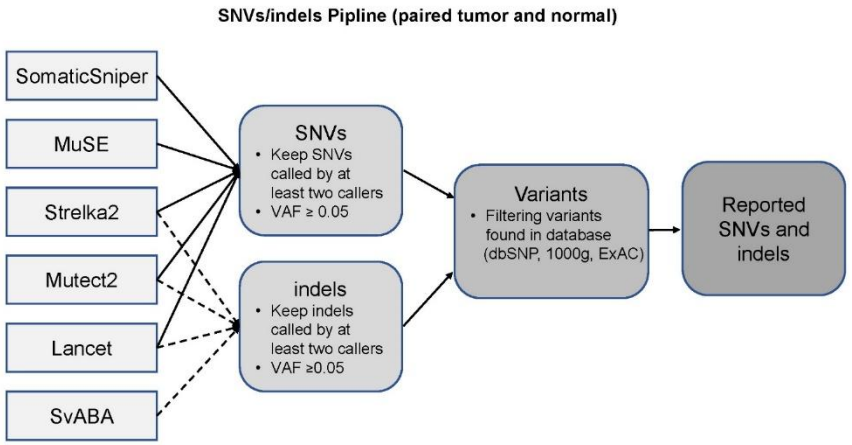
Figure S1


a
L
I
H
M
H&E
CD117
9T
36T
76T
85-1T
b
105 patients
68 patients with matched normal samples
37 patients with tumor-only samples
# Samples
# Patients
WGS
WES
WTS
19
19
19
49
59
49
5
111
101
case id
Normal
Tumor
NA
Fresh Frozen Tissue
Fresh Frozen Tissue (multiple tumors)
Cell Line
c
Normal
Tumor
Duplication Rate
Average sequencing depth (X)
PCR duplication rate
250
200
150
100
50
0
0.6
0.5
0.4
0.3
0.2
0.1
0
P92
P80
P84
P85
WGS
WES

**d**


WES
Color Key
0.4 0.7 1
Value
WGS

**e**


SNVs/indels Pipline (paired tumor and normal)
SomaticSniper
MuSE
Strelka2
Mutect2
Lancet
SvABA
SNVs
• Keep SNVs called by at least two callers
• VAF ≥ 0.05
indels
• Keep indels called by at least two callers
• VAF ≥0.05
Variants
• Filtering variants found in database (dbSNP, 1000g, ExAC)
Reported SNVs and indels

**f**

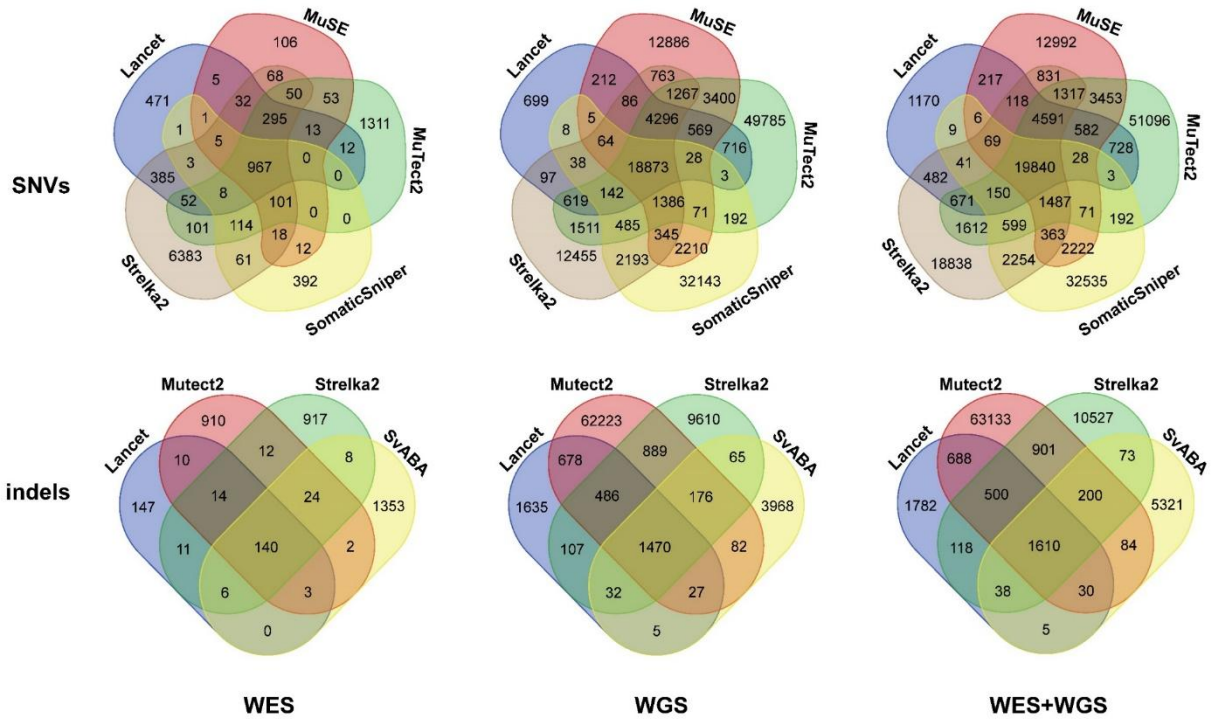

SNVs
indels
MuSE
Lancet
MuTect2
Strelka2
SomaticSniper
Mutect2
Strelka2
SvABA
Lancet
WES
WGS
WES+WGS

**Figure S1. Histopathology of GIST specimens, overview of three next generation sequencing cohorts, quality control of WGS and WES data, and analytical workflow.**

**(a)** The amount of viable tumor tissue and morphologic appearance for each GIST specimen used in the WGS, WES and/or RNA-seq. To determine the amount of viable tumor tissue and morphologic appearance, frozen sections were prepared from each cryopreserved sample with sufficient material for analysis. Cryosections of the same tumor specimen used in WGS, WES and/or RNA-seq were evaluated. Sections were mounted on precoated glass slides and stained with H&E frozen section kit (top) and CD117 immunohistochemistry staining (bottom). These validation studies showed that each specimen had a spindle or epithelioid cell morphology, and contained 25% or less nonneoplastic elements. L denotes low-risk, I denotes intermediate-risk, H denotes high-risk, M denotes metastatic.

**(b)** Composition of three GIST next generation sequencing cohorts in this study. The number of cases and tumor samples per cohort is labeled in parentheses.

**(c)** Quality estimation of WGS and WES data. Distribution of depth, and PCR duplicates for all samples analyzed by WGS and WES.

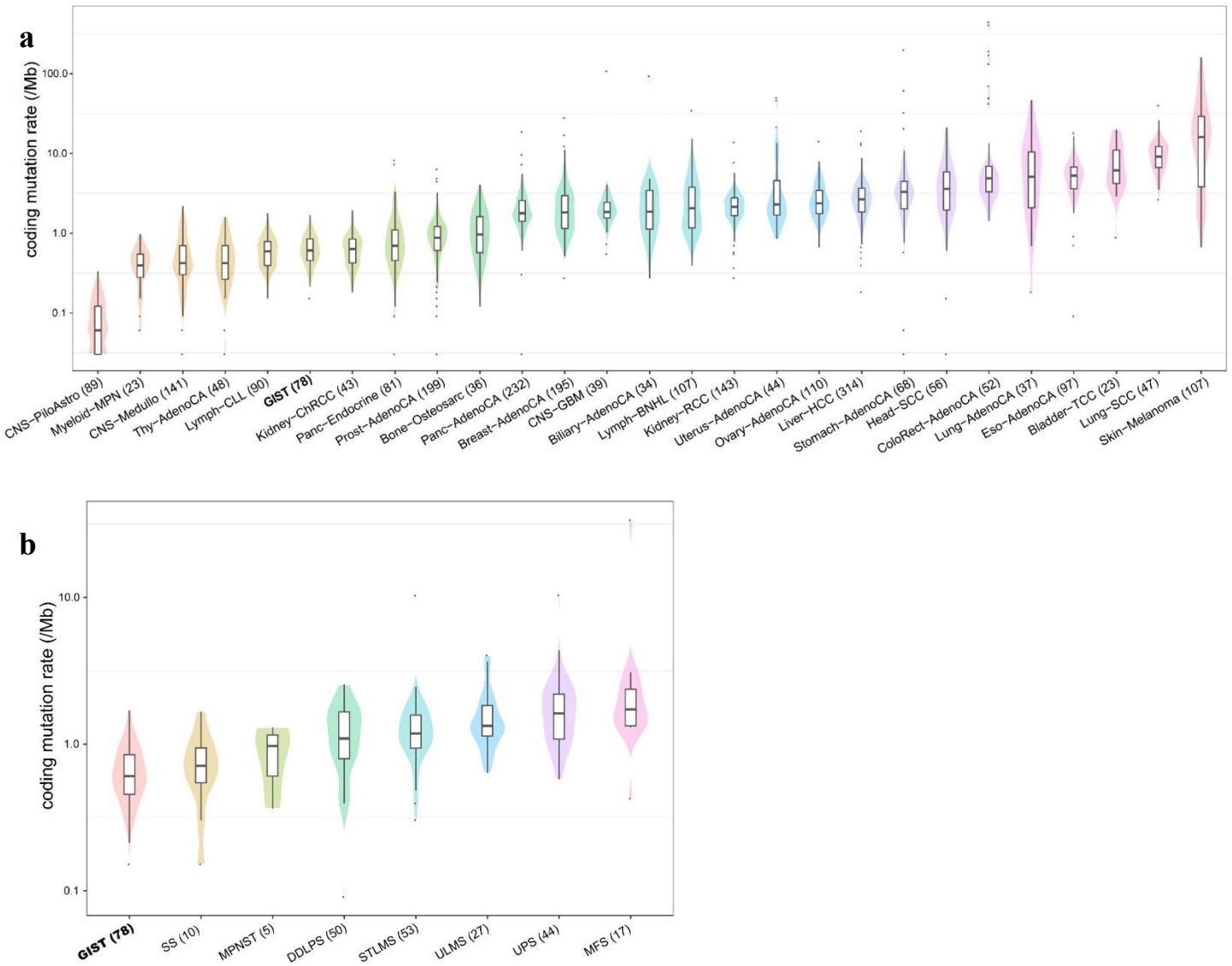
**(d)** BAM-matcher analysis for paired tumor and normal samples. Heatmap showing the fraction of common SNPs to determine whether two BAM files represent samples from the same patient.

**(e)** Analytical workflow. Variant calling pipelines paired tumor-normal samples.

**(f)** Venn plots showing the number of somatic SNVs (top) and indels (bottom) identified by different algorithms.

Source data are provided as a Source Data file.

**Figure S2**


a
coding mutation rate (/Mb)
100.0
10.0
1.0
0.1
CNS−PiloAstro (89)
Myeloid−MPN (23)
CNS−Medullo (141)
Thy−AdenoCA (48)
Lymph−CLL (90)
GIST (78)
Kidney−ChRCC (43)
Panc−Endocrine (81)
Prost−AdenoCA (199)
Bone−Osteosarc (36)
Panc−AdenoCA (232)
Breast−AdenoCA (195)
CNS−GBM (39)
Biliary−AdenoCA (34)
Lymph−BNHL (107)
Kidney−RCC (143)
Uterus−AdenoCA (44)
Ovary−AdenoCA (110)
Liver−HCC (314)
Stomach−AdenoCA (68)
Head−SCC (56)
ColoRect−AdenoCA (52)
Lung−AdenoCA (37)
Eso−AdenoCA (97)
Bladder−TCC (23)
Lung−SCC (47)
Skin−Melanoma (107)


b
coding mutation rate (/Mb)
10.0
1.0
0.1
GIST (78)
SS (10)
MPNST (5)
DDLPS (50)
STLMS (53)
ULMS (27)
UPS (44)
MFS (17)

**Figure S2. Somatic coding mutation rates of GIST and other human cancers.** The numbers of cases in each tumor type are shown next to the labels in parentheses. Each dot represents one tumor sample. Tumor types are ordered by the coding mutation rate.

**(a)** Comparison of TMB between GIST and 26 PCAWG (Pan-Cancer Analysis of Whole Genomes) human cancers. Mutation data of PCAWG were retrieved from published paper[1]. Only tumor types with more than 20 samples in PCAWG are shown. AdenoCA, adenocarcinoma; BNHL, B-cell non-Hodgkin lymphoma; ChRCC, chromophobe renal cell carcinoma; CLL, chronic lymphocytic leukemia; CNS, central nervous system; ColoRect, colorectal; Eso, esophageal; GBM, glioblastoma; HCC, hepatocellular carcinoma; Medullo, medulloblastoma; MH, microhomology; MPN, myeloproliferative neoplasm; Osteosarc, osteosarcoma; Panc, pancreatic; PiloAstro, pilocytic astrocytoma; Prost; prostate; RCC, renal cell carcinoma; SCC, squamous cell carcinoma; TCC, transitional cell carcinoma; Thy, thyroid.

**(b)** Comparison of TMB between GIST and 6 non-GIST sarcomas. Mutation data of 6 non-GIST sarcomas were retrieved from published paper[2]. SS, synovial sarcoma; MPNST, malignant peripheral nerve sheath tumor; DDLPS, dedifferentiated liposarcoma; STLMS, soft tissue leiomyosarcoma; ULMS, gynecologic leiomyosarcoma; UPS, undifferentiated pleomorphic sarcoma; MFS, myxofibrosarcoma.

In this figure, the low bound, centerline, and upper bound of boxplot represent the first quartile, the median, and the third quartile of data, respectively; the upper and lower whiskers extend to the largest and smallest value within 1.5 times of the interquartile range (IQR) and the points greater than $\pm 1.5 \times$ IQR are plotted as outlier dots.

Source data are provided as a Source Data file.

**Figure S3**

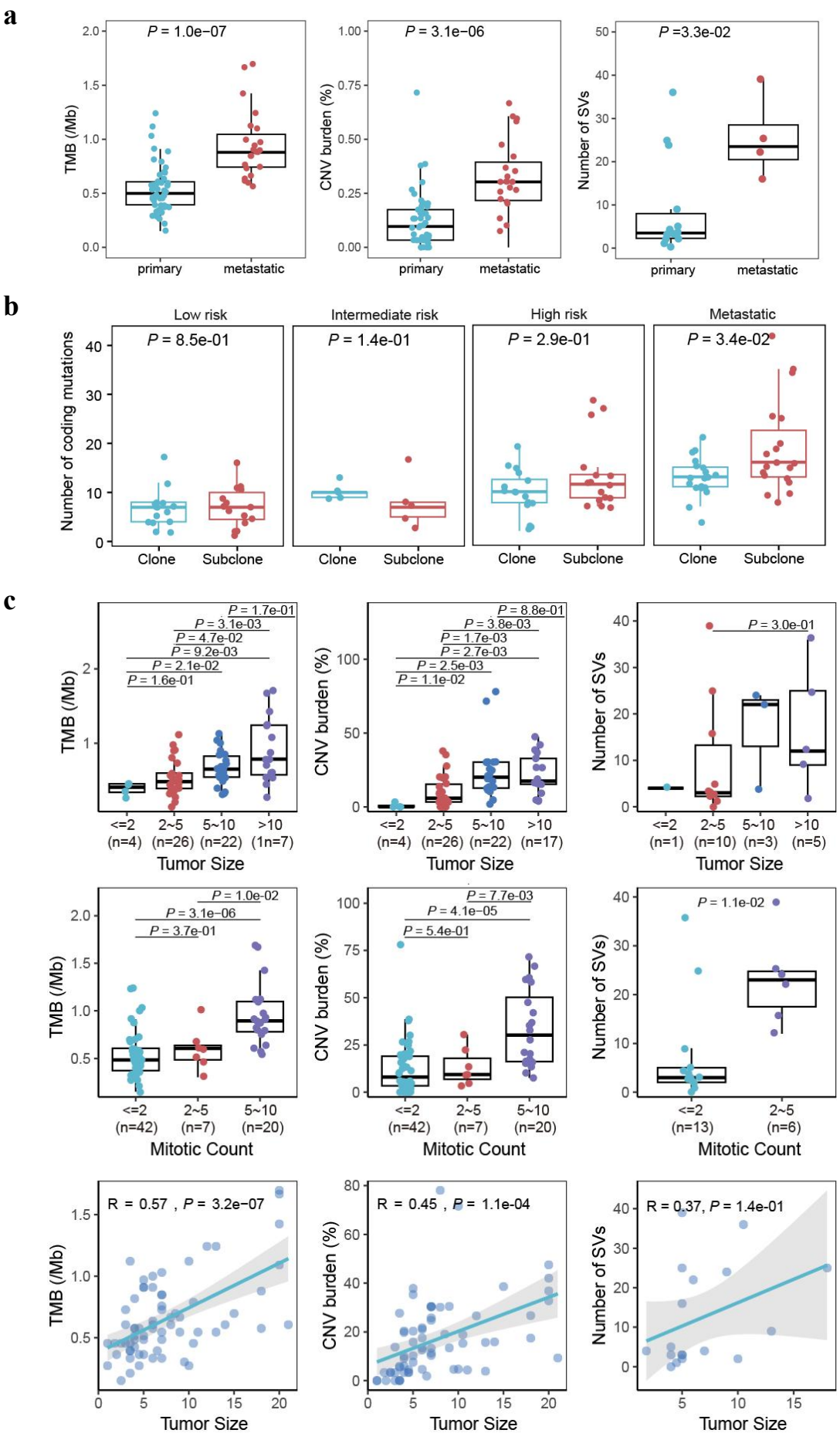

a
P = 1.0e−07
TMB (/Mb)
primary
metastatic
P = 3.1e−06
CNV burden (%)
P =3.3e-02
Number of SVs
b
Low risk
Intermediate risk
High risk
Metastatic
P = 8.5e-01
P = 1.4e-01
P = 2.9e-01
P = 3.4e-02
Number of coding mutations
Clone
Subclone
c
P = 1.7e-01
P = 3.1e-03
P = 4.7e-02
P = 9.2e-03
P = 2.1e-02
P = 1.6e-01
P = 8.8e-01
P = 3.8e-03
P = 1.7e-03
P = 2.7e-03
P = 2.5e-03
P = 1.1e-02
P = 3.0e-01
<=2 (n=4)
2~5 (n=26)
5~10 (n=22)
>10 (1n=7)
>10 (n=17)
<=2 (n=1)
2~5 (n=10)
5~10 (n=3)
>10 (n=5)
Tumor Size
P = 1.0e-02
P = 3.1e−06
P = 3.7e-01
P = 7.7e-03
P = 4.1e−05
P = 5.4e-01
P = 1.1e-02
<=2 (n=42)
2~5 (n=7)
5~10 (n=20)
<=2 (n=13)
2~5 (n=6)
Mitotic Count
R = 0.57 , P = 3.2e−07
R = 0.45 , P = 1.1e-04
R = 0.37, P = 1.4e-01

**Figure S3. Metastatic GISTs harbor more genomic aberrations.**

**(a)** Boxplots showing the alternation burdens in primary and metastatic GISTs.

**(b)** Number of clone and subclone mutations in coding region among different risk stratification.

**(c)** TMB and CNV burden are positively associated with tumor size and mitotic count.

In this figure, *P* values were calculated by the two-sided wilcoxon rank sum test. The low bound, centerline, and upper bound of boxplot represent the first quartile, the median, and the third quartile of data, respectively; the upper and lower whiskers extend to the largest and smallest value within 1.5 times of the interquartile range (IQR) and the points greater than $\pm 1.5 \times$ IQR are plotted as outlier dots. Source data are provided as a Source Data file.

**Figure S4**

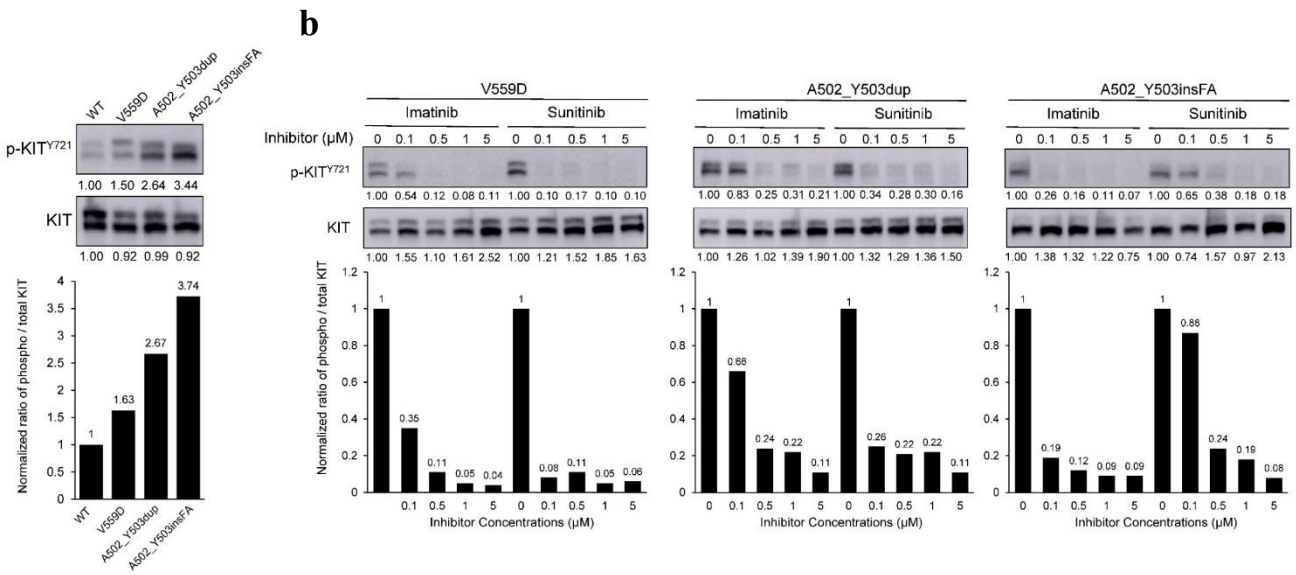

a
b
WT
V559D
A502_Y503dup
A502_Y503insFA
p-KIT Y721
1.00 1.50 2.64 3.44
KIT
1.00 0.92 0.99 0.92
Normalized ratio of phospho / total KIT
1
1.63
2.67
3.74
V559D
A502_Y503dup
A502_Y503insFA
Imatinib
Sunitinib
Inhibitor (μM)
0 0.1 0.5 1 5 0 0.1 0.5 1 5
p-KIT Y721
1.00 0.54 0.12 0.08 0.11 1.00 0.10 0.17 0.10 0.10
KIT
1.00 1.55 1.10 1.61 2.52 1.00 1.21 1.52 1.85 1.63
Normalized ratio of phospho / total KIT
1 0.35 0.11 0.05 0.04 1 0.08 0.11 0.05 0.06
Inhibitor Concentrations (μM)
1.00 0.83 0.25 0.31 0.21 1.00 0.34 0.28 0.30 0.16
1.00 1.26 1.02 1.39 1.90 1.00 1.32 1.29 1.36 1.50
1 0.66 0.24 0.22 0.11 1 0.26 0.22 0.22 0.11
1.00 0.26 0.16 0.11 0.07 1.00 0.65 0.38 0.18 0.18
1.00 1.38 1.32 1.22 0.75 1.00 0.74 1.57 0.97 2.13
1 0.19 0.12 0.09 0.09 1 0.88 0.24 0.19 0.08

**Figure S4. Constitutive activation of the novel *KIT* mutant (A502_Y503insFA) identified in GIST and its biochemical response to first-line and second-line TKIs (imatinib and sunitinib).**

**(a)** Constitutive activation of the novel KIT mutant (A502_Y503insFA). Mutant A502_Y503insFA cDNA was transfected into HEK293 cells. Well-established oncogenic KIT mutants (exon 9, A502_Y503dup; exon 11, V559D) were also transfected as controls.

**(b)** Comparative biochemical activity of TKIs for inhibiting KIT activity in transiently transfected HEK293 cells. HEK293 cells were transiently transfected, and protein lysates from transfected cells were prepared and subjected to western blotting. Representative results for the comparative activity of imatinib or sunitinib for inhibiting the novel *KIT* exon 9 mutant (A502_Y503insFA). The *KIT* exon 9 mutant (A502_Y503dup) and *KIT* exon 11 mutant (V559D) serve as positive controls. All experiments were replicated independently three times.

Source data are provided as a Source Data file.

**Figure S5**

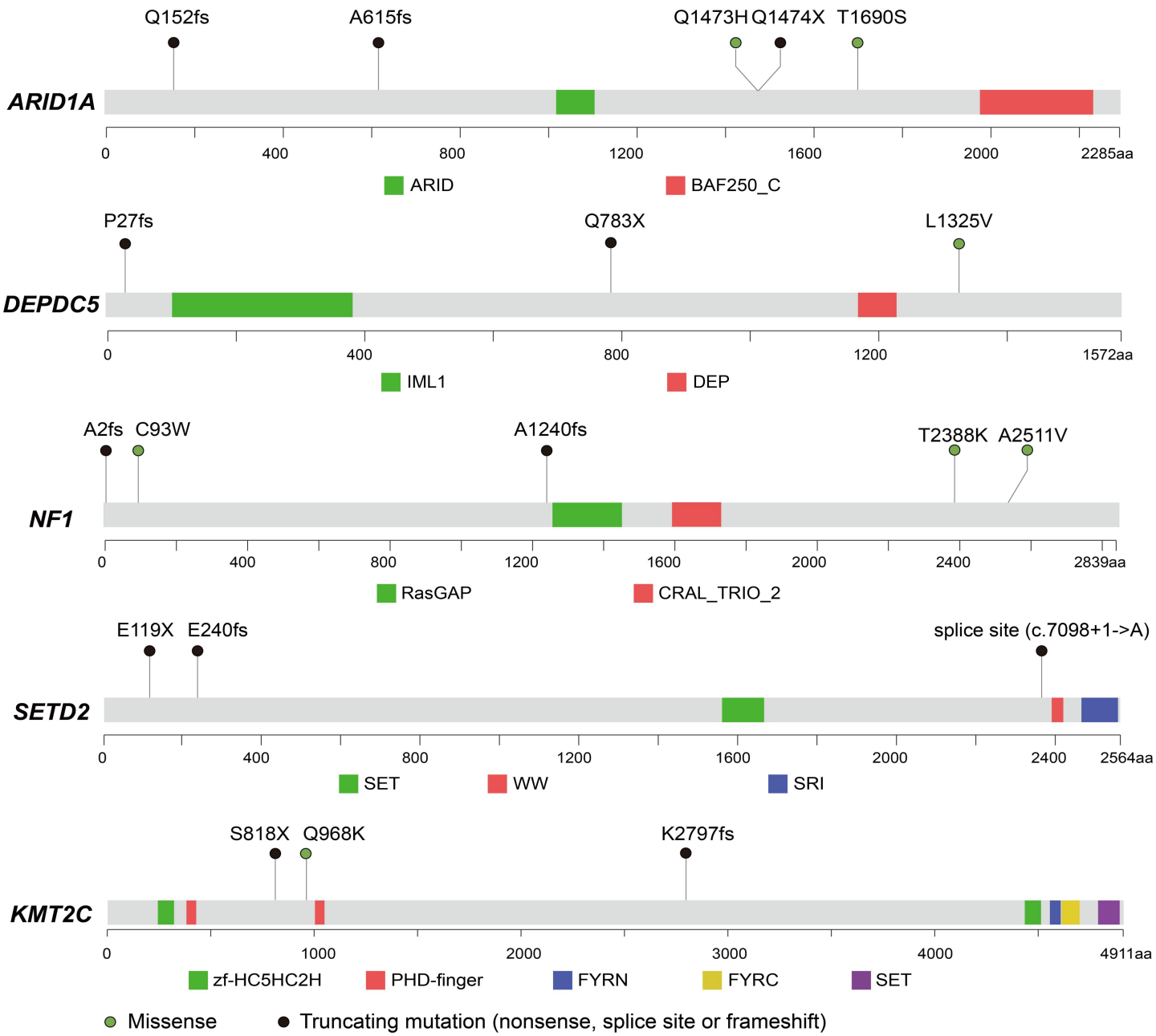
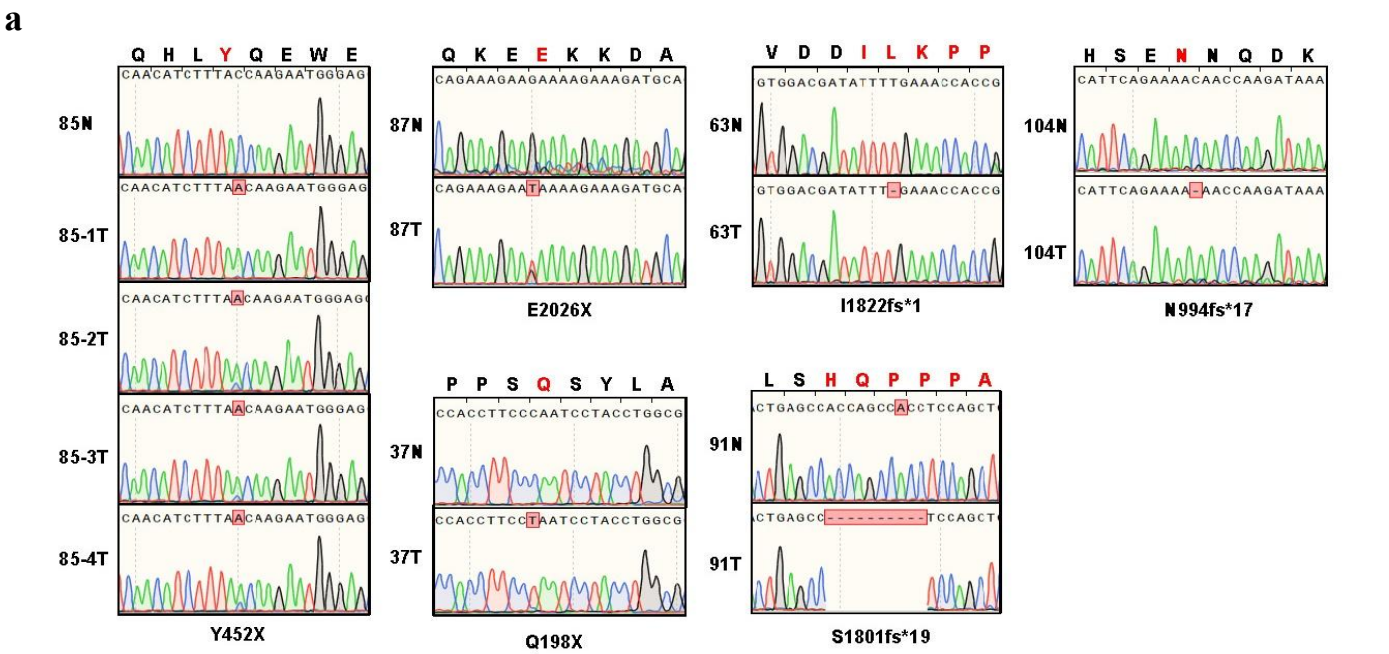
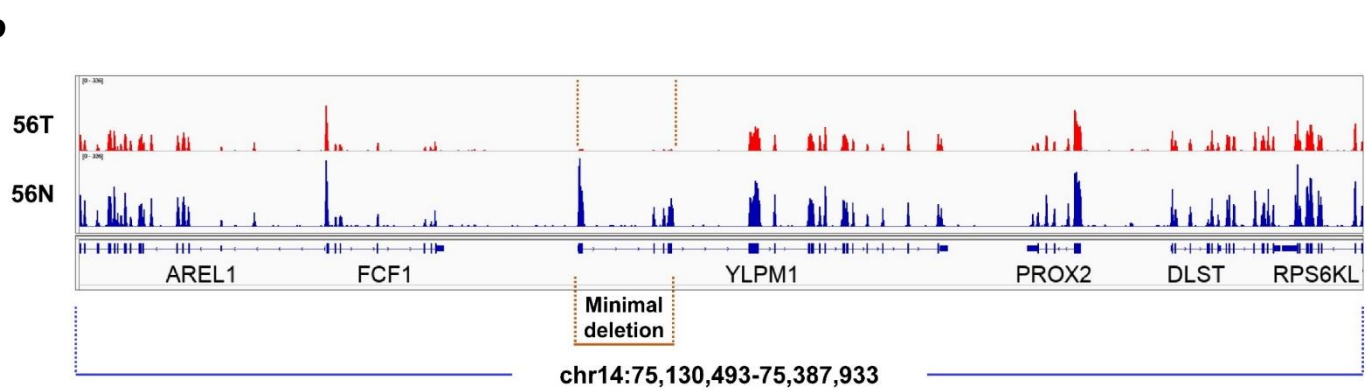

Q152fs
A615fs
Q1473H Q1474X
T1690S
ARID1A
0
400
800
1200
1600
2000
2285aa
ARID
BAF250_C
P27fs
Q783X
L1325V
DEPDC5
0
400
800
1200
1572aa
IML1
DEP
A2fs
C93W
A1240fs
T2388K
A2511V
NF1
0
400
800
1200
1600
2000
2400
2839aa
RasGAP
CRAL_TRIO_2
E119X
E240fs
splice site (c.7098+1->A)
SETD2
0
400
800
1200
1600
2000
2400
2564aa
SET
WW
SRI
S818X
Q968K
K2797fs
KMT2C
0
1000
2000
3000
4000
4911aa
zf-HC5HC2H
PHD-finger
FYRN
FYRC
SET
Missense
Truncating mutation (nonsense, splice site or frameshift)

**Figure S5. Lollipop plots showing the distribution of all non-silent mutations in *ARID1A*, *DEPDC5*, *NF1* and *SETD2.*** The protein domains of each gene are indicated in the figure. The scale bars represent the length (amino acids) of the protein sequence and the protein domains are indicated by colors.

**Figure S6**


a
85N
85-1T
85-2T
85-3T
85-4T
Y452X
87N
87T
E2026X
63N
63T
I1822fs*1
104N
104T
N994fs*17
37N
37T
Q198X
91N
91T
S1801fs*19


b
56T
56N
AREL1
FCF1
YLPM1
PROX2
DLST
RPS6KL
Minimal deletion
chr14:75,130,493-75,387,933

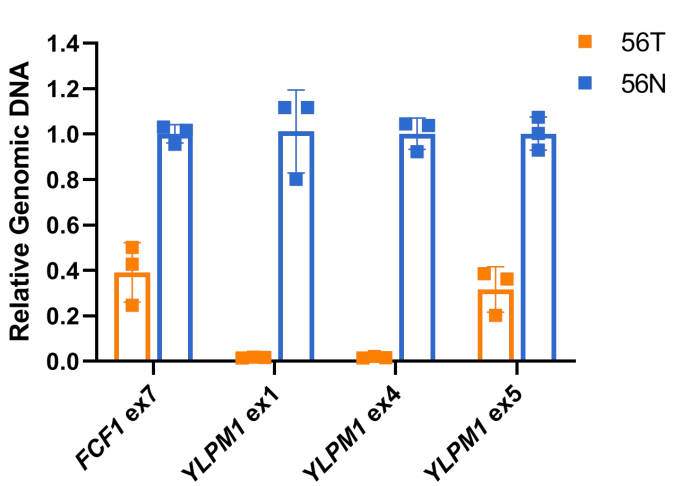

c
56T
56N
Relative Genomic DNA
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
FCF1 ex7
YLPM1 ex1
YLPM1 ex4
YLPM1 ex5

**Figure S6. Validation of *YLPM1* genomic aberrations in GISTs.**
**(a)** Sanger sequencing confirms the somatic homozygous mutations of *YLPM1*. Identical *YLPM1* nonsense mutation in multiple anatomically distinct metastases from the same patients (case 85).
**(b-c)** Validation of homozygous *YLPM1* deletions in GISTs. (**b**) Integrative genome viewer images of part of chromosome 14q from matched tumor and nonneoplastic cell DNAs from the same patient (case 56), demonstrating intragenic *YLPM1* deletion (exons 1–4 homozygous deletion) and the tumor-restricted nature of the *YLPM1* alteration. (**c**) Genomic qPCR analysis validation of the homozygous deletion of *YLPM1* in case 56. Source data are provided as a Source Data file. Data are presented as mean values ± s.d. n=3.

**Figure S7**

**a**

| | *YLPM1* aberrations | Normal *YLPM1* | Total | *YLPM1* aberrations frequency |
|---|---|---|---|---|
| **GISTs** | 9 | 59 | 68 | 13.24% |
| **Non-GIST sarcomas** (TCGA, Pan-Cancer Atlas) | 8* | 243 | 253 | 3.16% |
| **All TCGA Pan-Cancer** | 364 | 10589 | 10953 | 3.32% |

**b**

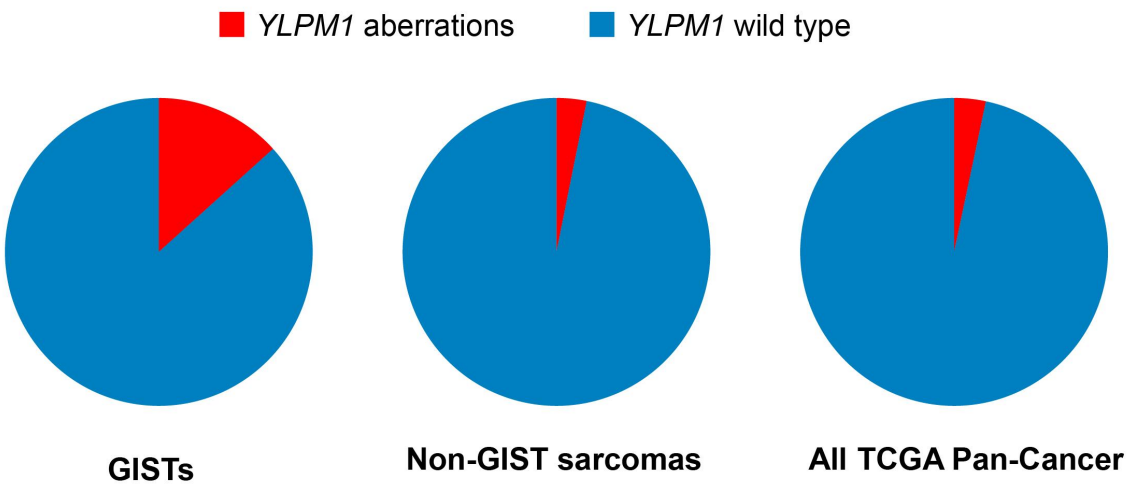


**Figure S7. Genomic *YLPM1* aberrations in 339 human sarcomas, including 86 GISTs and 253 non-GIST sarcomas.**
**(a)** The 253 non-GIST sarcomas (TCGA Pan-Cancer Atlas program, Hoadley KA *et al*, Cell-of-origin patterns dominate the molecular classification of 10,000 tumors from 33 types of cancer, ***Cell***, 2018) include 99 leiomyosarcomas, 59 dedifferentiated liposarcomas, 50 undifferentiated pleomorphic sarcomas/malignant fibrous histiocytoma/high-grade spindle cell sarcomas, 24 myxofibrosarcomas, 10 synovial sarcomas, 9 malignant peripheral nerve sheath tumors and 2 desmoid/aggressive fibromatosis. Genomic *YLPM1* aberrations are significantly more frequent in GISTs compared with non-GIST sarcomas ($p<0.0001$, two-tailed Fisher's test). ***** The 8 *YLPM1* aberrations are: 5 amplifications (3 in myxofibrosarcomas, 1 in a dedifferentiated liposarcoma, and 1 in a malignant peripheral nerve sheath tumor), *YLPM1-KHDRBS2* fusion in a malignant peripheral nerve sheath tumor, E263D in a myxofibrosarcoma, and R1357X in a malignant peripheral nerve sheath tumor.
**(b)** The pie chart indicates the frequency of *YLPM1* aberrations in GIST and other cancer types.

**Figure S8**

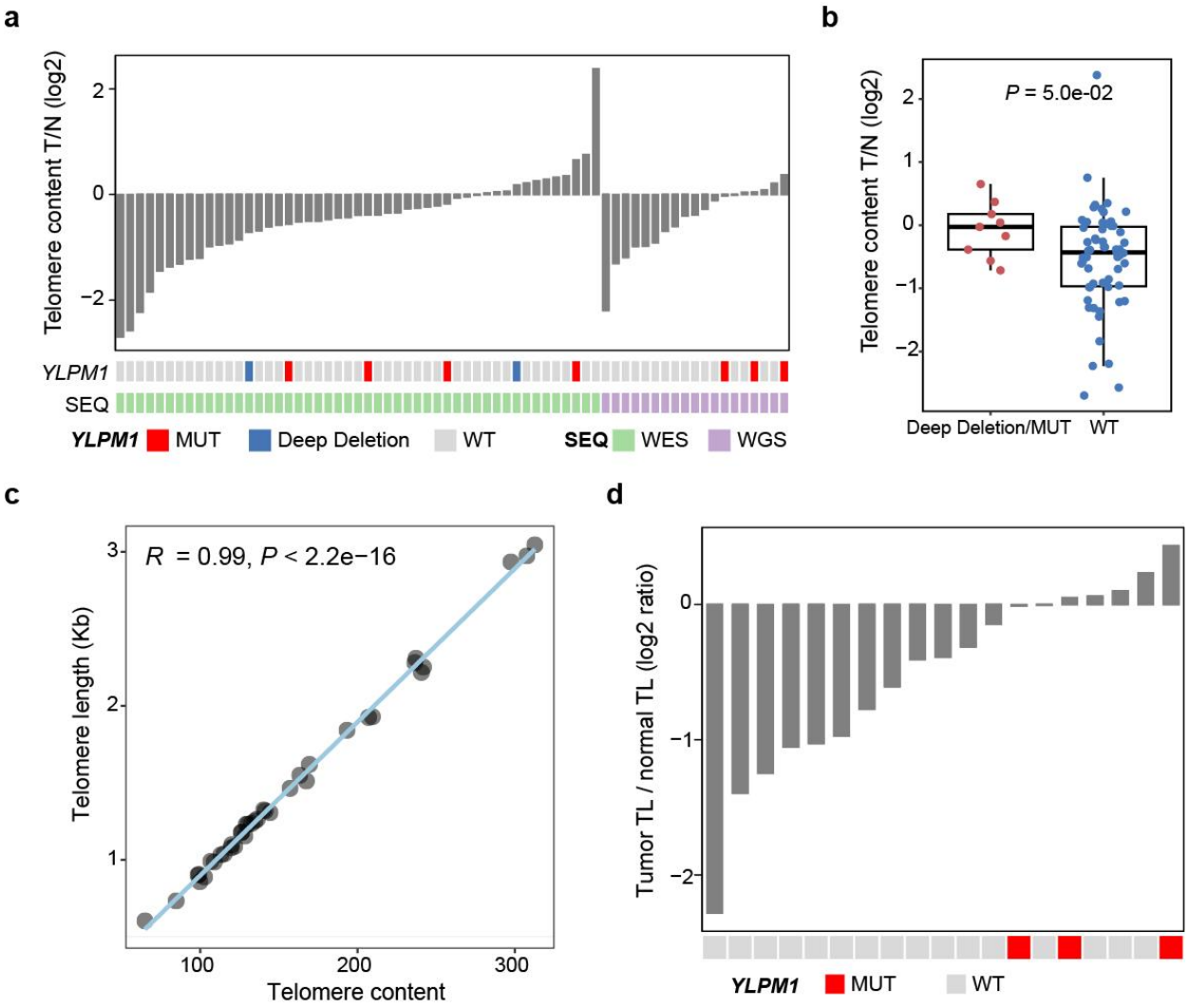

a
Telomere content T/N (log2)
2
0
−2
YLPM1
SEQ
YLPM1 MUT
Deep Deletion
WT
SEQ WES
WGS
b
Telomere content T/N (log2)
P = 5.0e-02
2
1
0
−1
−2
Deep Deletion/MUT
WT
c
R = 0.99, P < 2.2e−16
Telomere length (Kb)
3
2
1
100
200
300
Telomere content
d
Tumor TL / normal TL (log2 ratio)
0
−1
−2
YLPM1
MUT
WT

**Figure S8. Genomic alteration of *YLPM1* is correlated with telomere length in human GISTs.**
**(a)** Bar plot showing the log2 ratio of telomere content in tumors with *YLPM1* mutations.
**(b)** Box plot showing the log2 ratio of telomere content in tumors with *YLPM1* mutations. *P* value is determined by two-sided Wilcoxon rank-sum test. The low bound, centerline, and upper bound of boxplot represent the first quartile, the median, and the third quartile of data, respectively; the upper and lower whiskers extend to the largest and smallest value within 1.5 times of the interquartile range (IQR) and the points greater than ±1.5 × IQR are plotted as outlier dots.
**(c)** Scatter plot showing the correlation between telomere content and telomere length detected on WGS data.
**(d)** Bar plot showing the log2 ratio of telomere length in tumors with *YLPM1* mutations based on WGS data.
Source data are provided as a Source Data file.

**Figure S9**

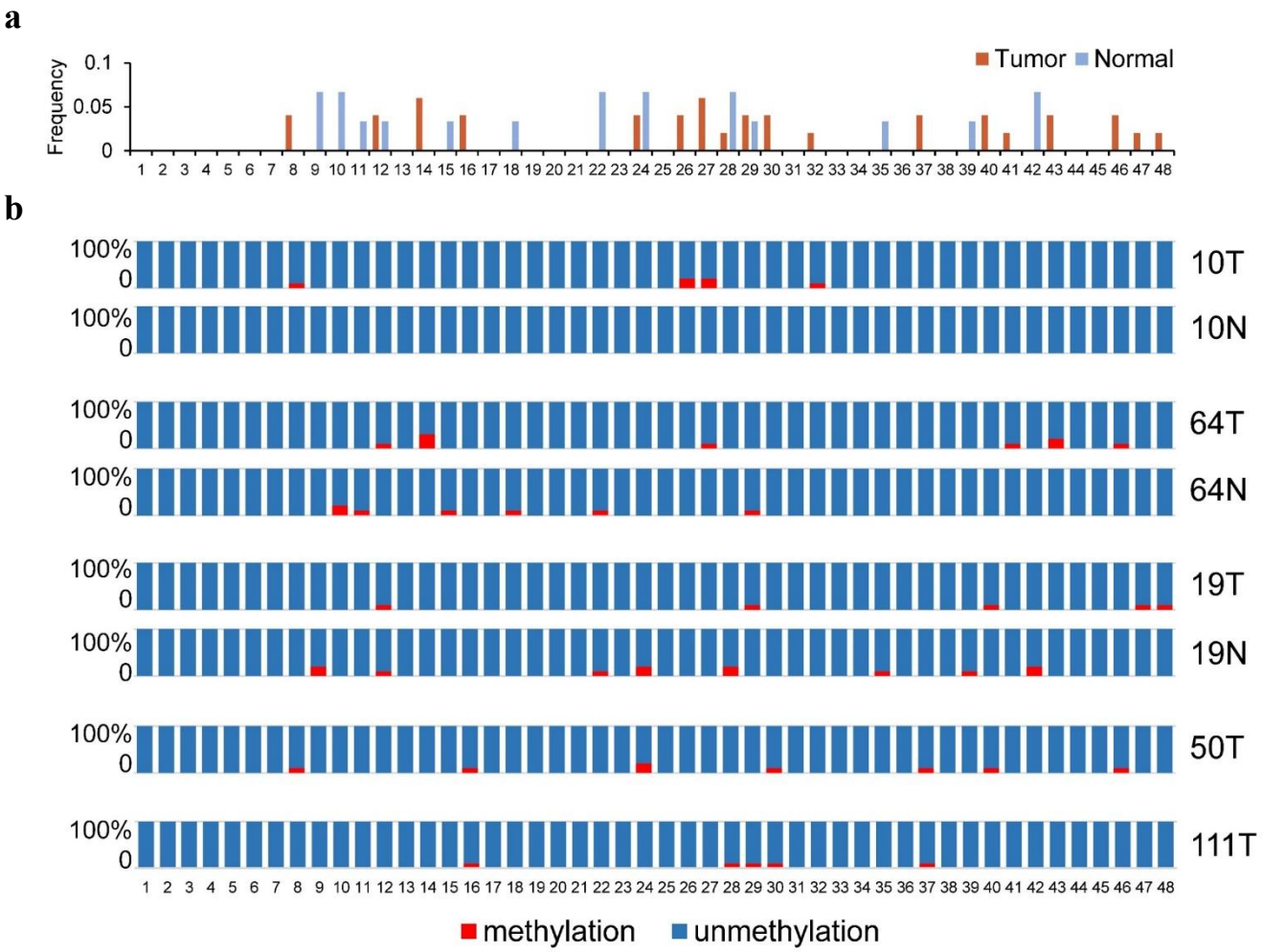
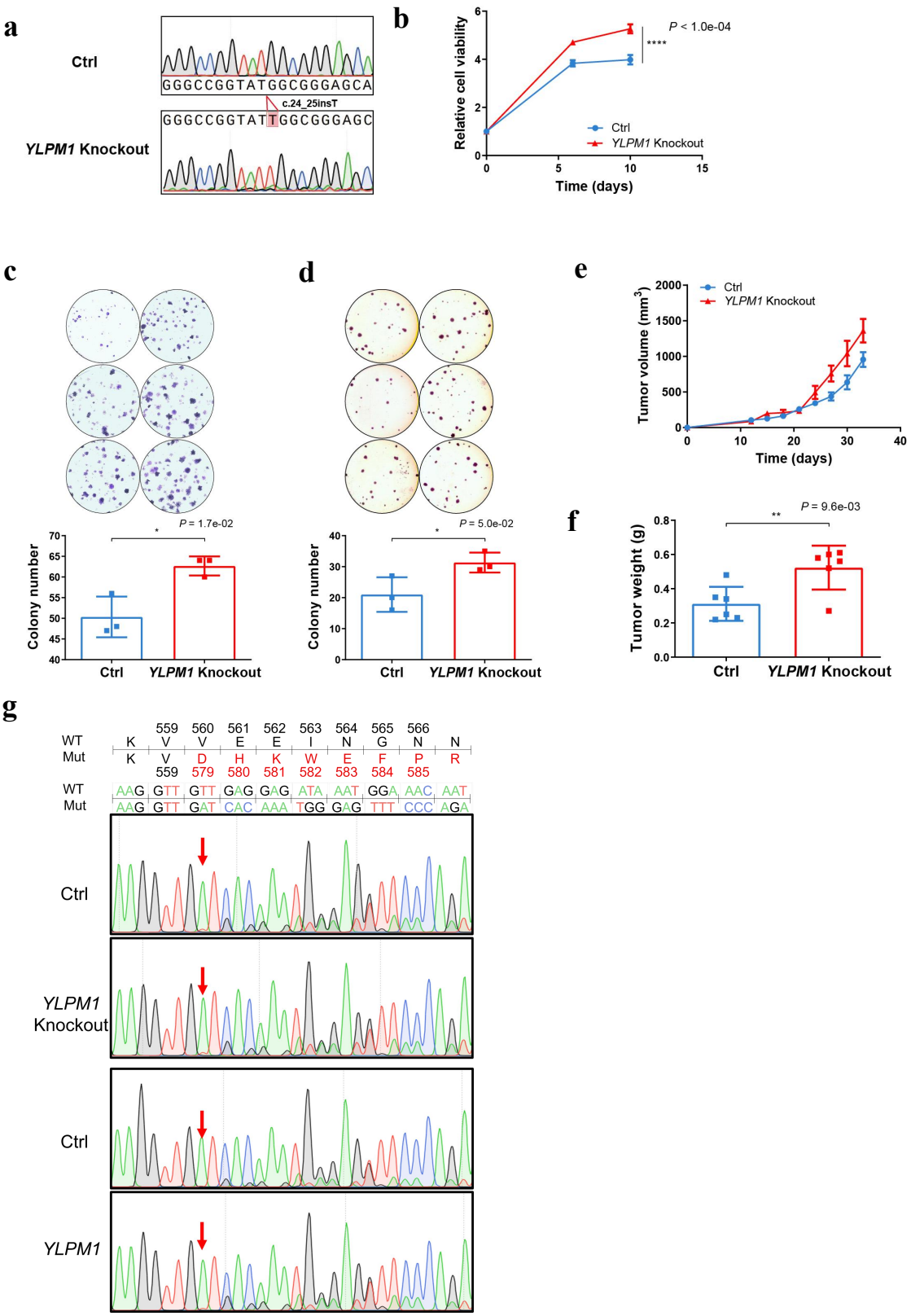

a
Frequency
0.1
0.05
0
Tumor
Normal
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
b
100%
0
10T
10N
64T
64N
19T
19N
50T
111T
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
methylation
unmethylation

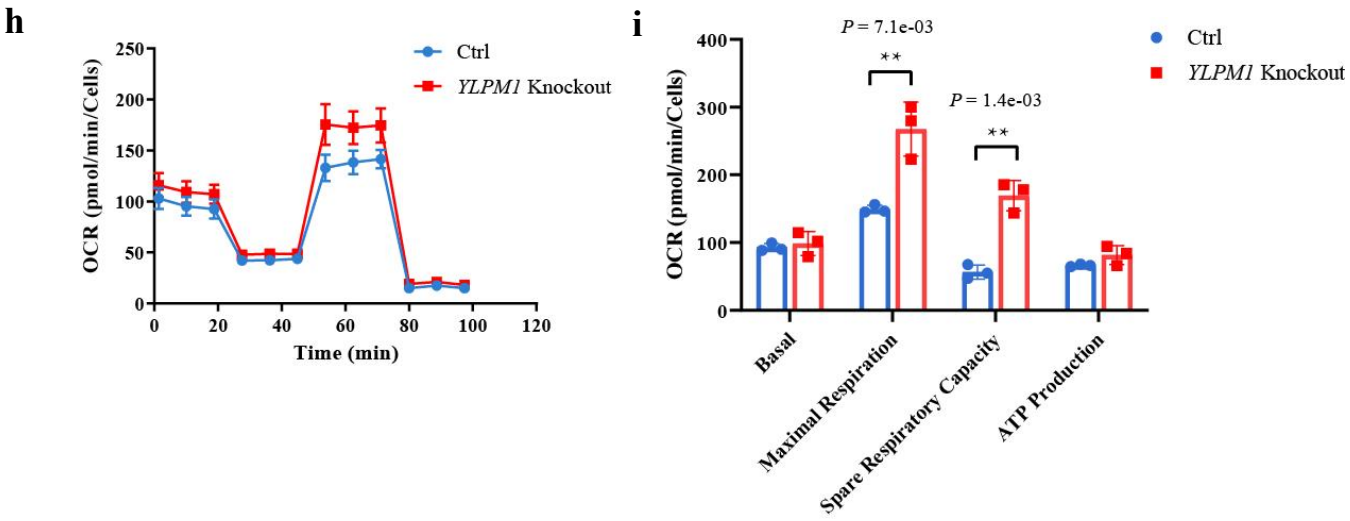
**Figure S9. Evaluation of *YLPM1* CpG island methylation status in GISTs.**

**(a)** Total frequency of methylated and unmethylated status of 48 CpGs in the *YLPM1* promoter region of the tumors with low *YLPM1* FPKM. The matched normal samples were also checked if available. The red bar indicates the frequency of methylated clones in tumors, and the blue bar indicates methylated clones in paired normal samples.

**(b)** Number of methylated and unmethylated clones at each CpG site. Colored bars represent the aggregated methylation status as assessed by bisulfite genomic sequencing at each position: red, methylated; blue, unmethylated.

**Figure S10**


a
Ctrl
GGGCCGGTATGGCGGGAGCA
c.24_25insT
YLPM1 Knockout
GGGCCGGTATTGGCGGGAGC
b
Relative cell viability
P < 1.0e-04
****
Ctrl
YLPM1 Knockout
Time (days)
c
Colony number
P = 1.7e-02
Ctrl
YLPM1 Knockout
d
Colony number
P = 5.0e-02
Ctrl
YLPM1 Knockout
e
Tumor volume (mm3)
Ctrl
YLPM1 Knockout
Time (days)
f
Tumor weight (g)
P = 9.6e-03
Ctrl
YLPM1 Knockout
g
WT K V V E E I N G N N
Mut K V D H K W E F P R
559 560 561 562 563 564 565 566
559 579 580 581 582 583 584 585
WT AAG GTT GTT GAG GAG ATA AAT GGA AAC AAT
Mut AAG GTT GAT CAC AAA TGG GAG TTT CCC AGA
Ctrl
YLPM1 Knockout
Ctrl
YLPM1


h
OCR (pmol/min/Cells)
Time (min)
Ctrl
YLPM1 Knockout
i
P = 7.1e-03
P = 1.4e-03
**
**
OCR (pmol/min/Cells)
Basal
Maximal Respiration
Spare Respiratory Capacity
ATP Production
Ctrl
YLPM1 Knockout

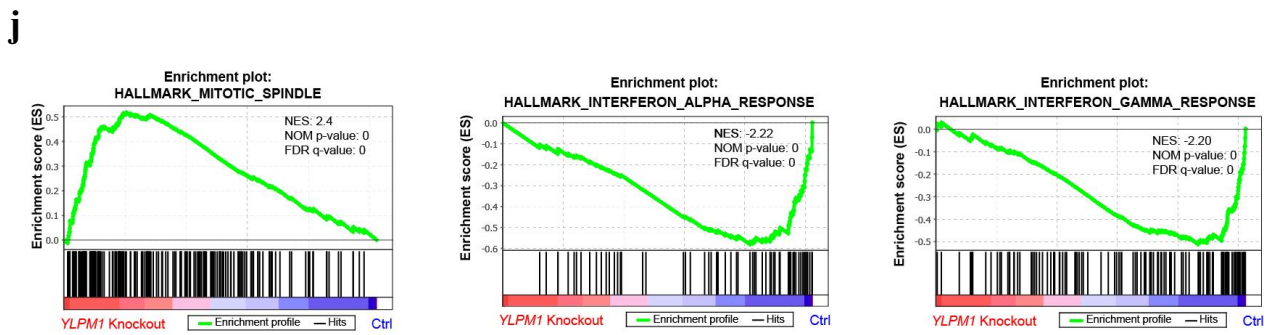

j
Enrichment plot:
HALLMARK_MITOTIC_SPINDLE
NES: 2.4
NOM p-value: 0
FDR q-value: 0
Enrichment score (ES)
YLPM1 Knockout
Enrichment profile
Hits
Ctrl
Enrichment plot:
HALLMARK_INTERFERON_ALPHA_RESPONSE
NES: -2.22
NOM p-value: 0
FDR q-value: 0
Enrichment score (ES)
YLPM1 Knockout
Enrichment profile
Hits
Ctrl
Enrichment plot:
HALLMARK_INTERFERON_GAMMA_RESPONSE
NES: -2.20
NOM p-value: 0
FDR q-value: 0
Enrichment score (ES)
YLPM1 Knockout
Enrichment profile
Hits
Ctrl

**Figure S10. *YLPM1* knockout promotes tumor proliferation and oxidative phosphorylation in GIST.**

**(a)** Sanger sequencing confirmation of the *YLPM1* knockout (c.24_25 insT; p.G9fs*11) in GIST-T1 cells.

**(b)** *YLPM1* knockout increases the viability of GIST-T1 cells, as assessed by the CellTiter-Glo viability assay. n=3. ***$P$ < 0.001. The $P$ value are calculated using the two-sided Student's t test.

**(c)** Crystal violet staining assay shows that *YLPM1* knockout promotes GIST-T1 cell proliferation. Representative plates (top) and mean colony numbers (bottom) are shown (± standard error of the mean). n=3. *$P$ < 0.05. The $P$ value are calculated using the two-sided Student's t test.

**(d)** *YLPM1* knockout promotes anchorage-independent growth of GIST-T1 cells. Representative plates (top) and mean colony numbers (bottom) are shown (±s.e.m. of the mean). n=3. *$P$ < 0.05. The $P$ value are calculated using the two-sided Student's t test.

**(e-f)** *YLPM1* knockout accelerates the growth of GIST-T1 xenografts in nude mice. Growth curves (e), and tumor weights (f) of transplanted tumors. Error bars are the mean ± s.e.m. of 6 replicates.

**(g)** Sanger sequencing confirms *KIT* exon 11 mutation (c.1678_1734del, p.Val560_Tyr578del) in *YLPM1*-KO tumor and *YLPM1* overexpressed tumor.

**(h)** The OCR profile following the addition of mitochondrial inhibitors (1 μM oligomycin, 1 μM FCCP, 0.5 μM rotenone/antimycin A) was determined at various time points by XFe24 and normalized with cell numbers.

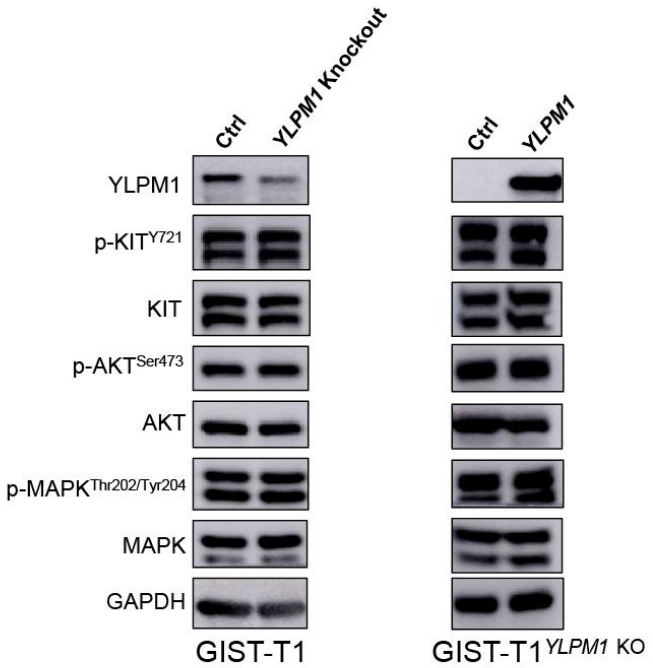
**(i)** Parameters of mitochondrial function are calculated from data shown in (h). Significance was assessed by Student's *t*-test. Data are shown as mean ± s.e.m., **$P$ < 0.01. Basal respiration was calculated as the difference between basal OCR and rotenone/antimycin-inhibited OCR. Maximal respiration was calculated as the difference between FCCP stimulated OCR and rotenone/antimycin-inhibited OCR. The $P$ values are calculated using the two-sided Student's t test.

**(j)** GSEA showing that genes involved in hallmark mitotic spindle gene sets are significantly upregulated in YLPM1 inactivated GIST-T1 cells, whereas immune gene sets are significantly downregulated in YLPM1 inactivated GIST-T1 cells. NES, normalized enrichment score.

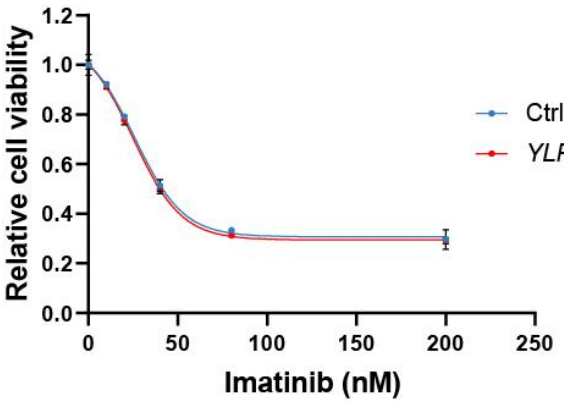
Source data are provided as a Source Data file.

**Figure S11**

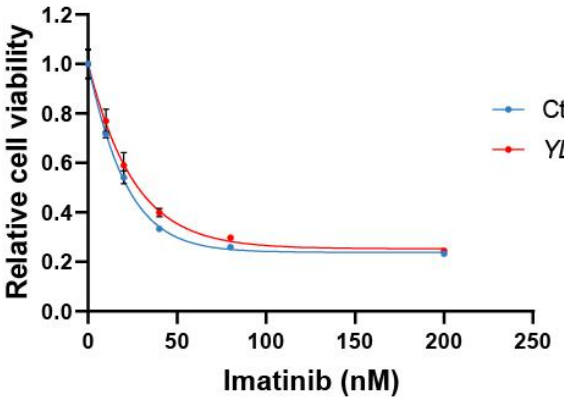
**a**




**b**




| Group | $IC_{50}$ (nM) |
|---|---|
| Ctrl | 30.86 |
| *YLPM1* Knockout | 30.26 |




| Group | $IC_{50}$ (nM) |
|---|---|
| Ctrl | 21.47 |
| *YLPM1* | 23.92 |

**Figure S11. YLPM1 does not regulate the KIT oncogenic pathway and the sensitivity of GIST cells to imatinib.**
**(a)** (Left) *YLPM1* knockout does not regulate the KIT oncogenic pathways in GIST-T1 cells. (Right) *YLPM1* restoration does not regulate the KIT oncogenic pathways in GIST-T1$^{YLPM1\ KO}$ cells.
**(b)** (Top) *YLPM1* knockout does not change the sensitivity of GIST-T1 cells to a KIT inhibitor (imatinib). $IC_{50}$ values of GIST-T1 with or without the *YLPM1* knockout are shown. (Bottom) *YLPM1* restoration does not change the sensitivity of GIST-T1$^{YLPM1\ KO}$ cells to the KIT inhibitor (imatinib). $IC_{50}$ values of GIST-T1$^{YLPM1\ KO}$ with or without *YLPM1* restoration are shown. Data are presented as mean values ± s.d. n=3.
Source data are provided as a Source Data file.

**Figure S12**

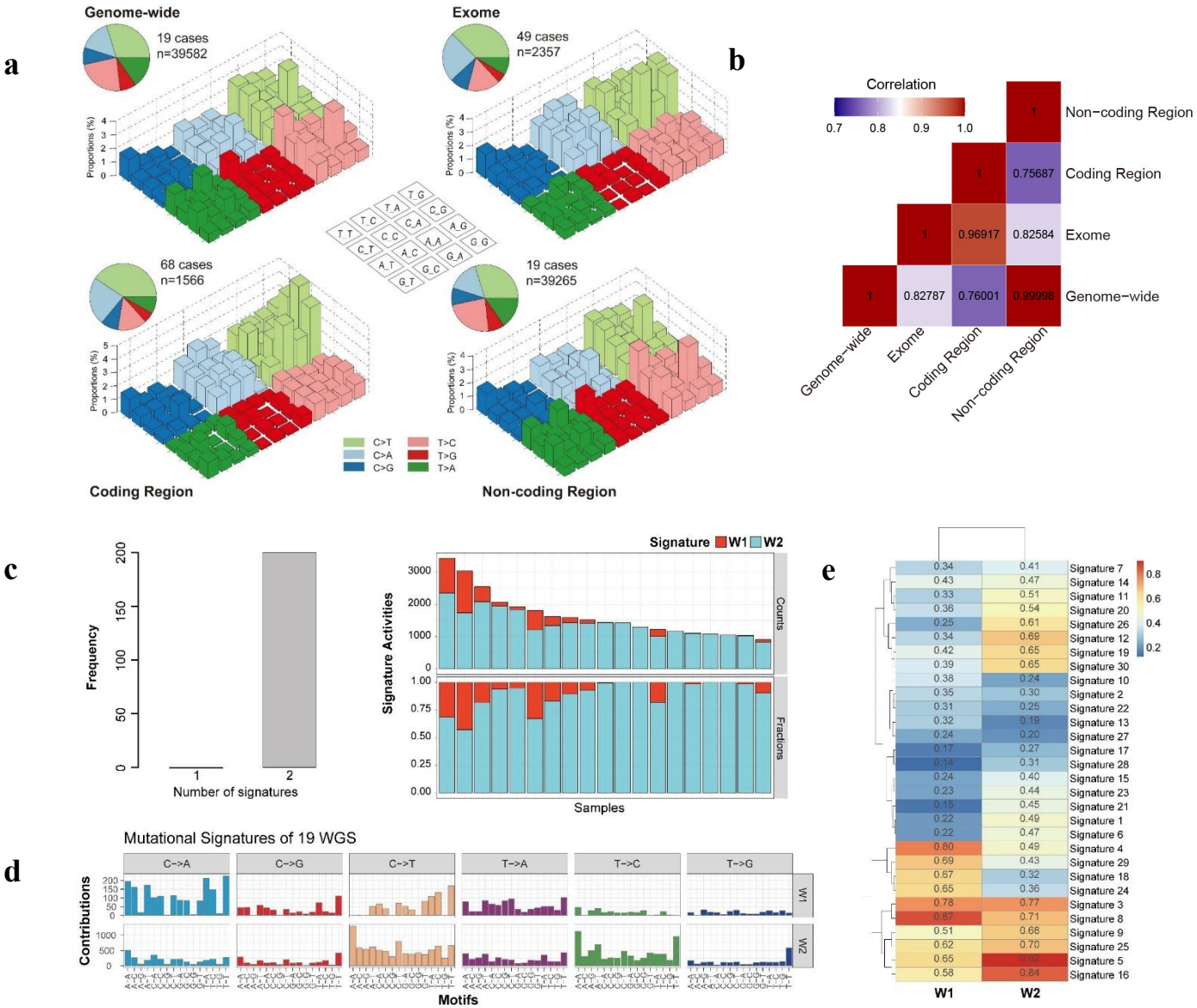

a
Genome-wide
19 cases
n=39582
Exome
49 cases
n=2357
68 cases
n=1566
19 cases
n=39265
Proportions (%)
C>T
C>A
C>G
T>C
T>G
T>A
Coding Region
Non-coding Region
b
Correlation
0.7
0.8
0.9
1.0
1
0.75687
0.96917
0.82584
0.82787
0.76001
0.99998
Non−coding Region
Coding Region
Exome
Genome−wide
c
Frequency
Number of signatures
Signature W1 W2
Signature Activities
Counts
Fractions
Samples
d
Mutational Signatures of 19 WGS
C−>A
C−>G
C−>T
T−>A
T−>C
T−>G
W1
W2
Contributions
Motifs
e
Signature 7
Signature 14
Signature 11
Signature 20
Signature 26
Signature 12
Signature 19
Signature 30
Signature 10
Signature 2
Signature 22
Signature 13
Signature 27
Signature 17
Signature 28
Signature 15
Signature 23
Signature 21
Signature 1
Signature 6
Signature 4
Signature 29
Signature 18
Signature 24
Signature 3
Signature 8
Signature 9
Signature 25
Signature 5
Signature 16
W1
W2

**Figure S12. Mutational signature of GISTs.**

**(a)** Lego plots of 96 mutation profiles in tumor-normal paired GISTs. SNVs are divided into six categories with 16 flanking trinucleotide contexts. Inset pie charts display the proportion of 6 categories of mutations. Genome-wide includes all somatic SNVs detected in the 19 WGS cases, exome includes all somatic SNVs in target region of the 49 WES cases, coding region includes all somatic SNVs in both WES and WGS cases, non-coding region includes all somatic non-coding SNVs in 19 WGS cases.

**(b)** Correlation between any two groups in Figure S12a.

**(c)** De novo signatures analysis of GISTs using bayesNMF based on 19 WGS data. The bar plot shows the frequencies of one and two signatures obtained through signature analysis; 200 independent analyses were performed (left). Bar plot of contributions for two signatures extracted by bayesNMF in 19 WGS (right).

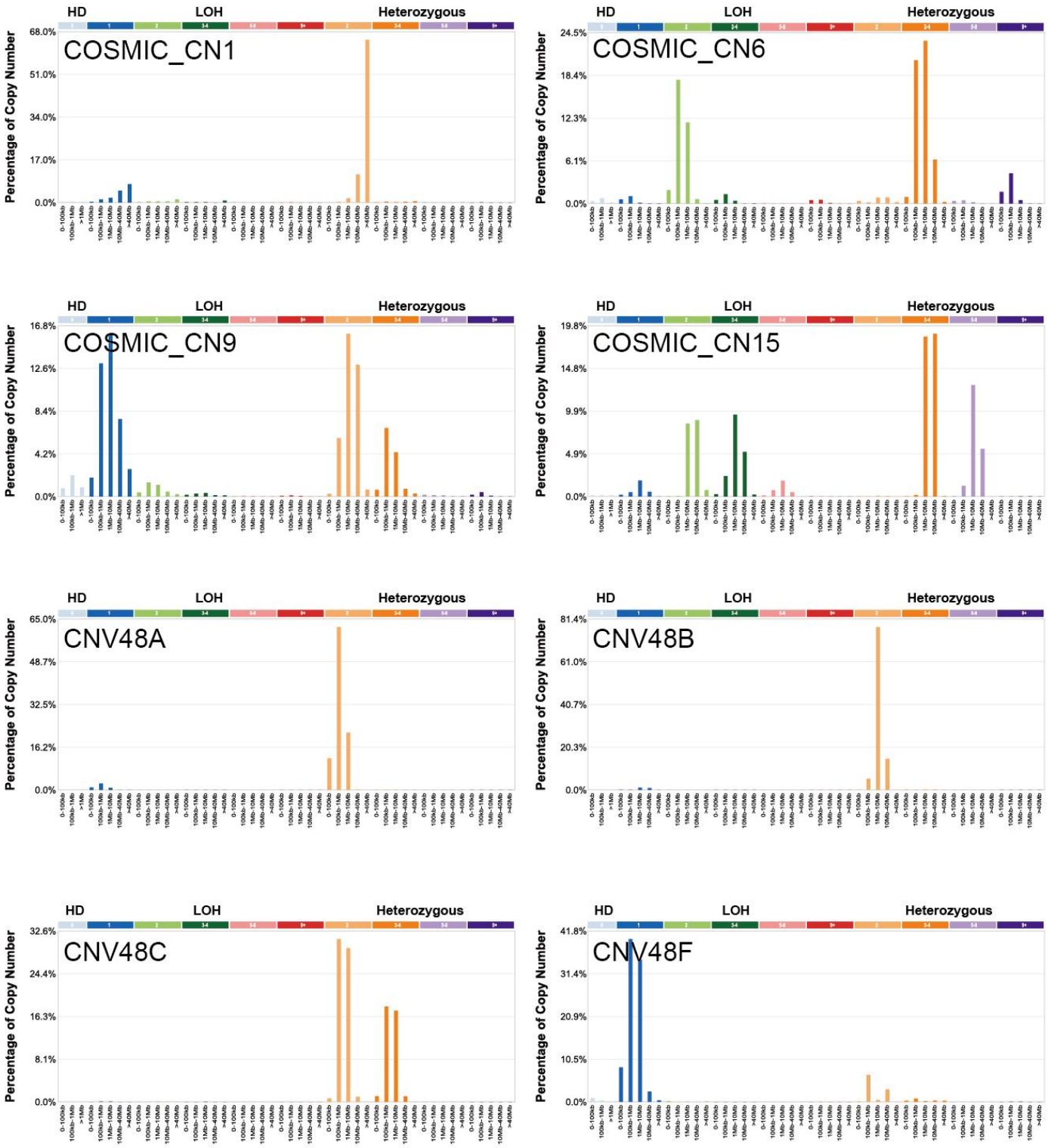
**(d)** 96-substitution classification of 2 signatures (W1 and W2) extracted by bayesNMF in 19 WGS. Each color refers to one of the six base substitutions, with x-axis showed 96 mutation types and y-axis showed the estimated mutations of each mutation type.

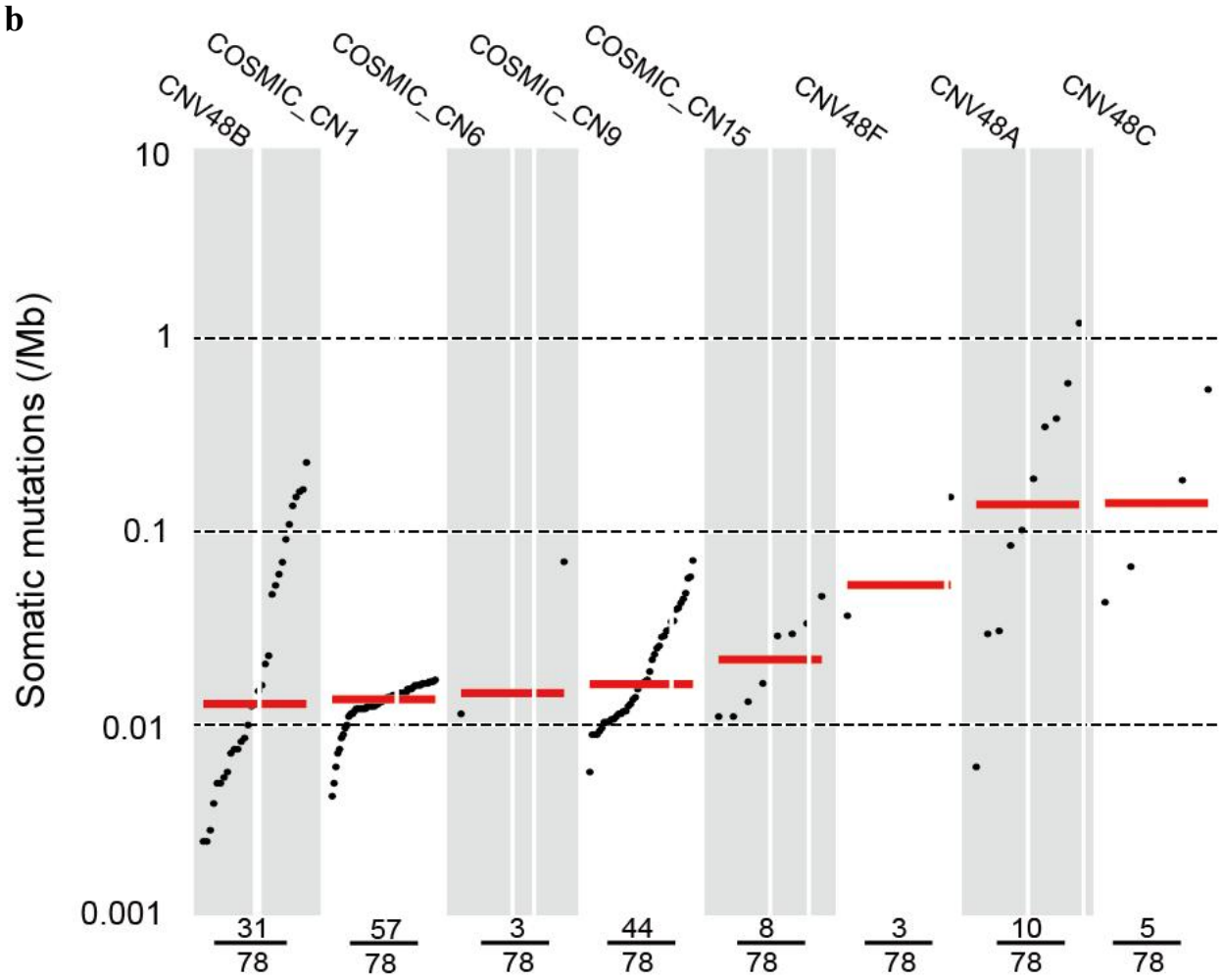
**(e)** Heatmap showing the cosine similarity between mutational signatures (W1 and W2) identified in 19 WGS using bayesNMF and COSMIC database (v2).

Source data are provided as a Source Data file.

**Figure S13**

**a**


HD
LOH
Heterozygous
Percentage of Copy Number
COSMIC_CN1
COSMIC_CN6
COSMIC_CN9
COSMIC_CN15
CNV48A
CNV48B
CNV48C
CNV48F


b
CNV48B
COSMIC_CN1
COSMIC_CN6
COSMIC_CN9
COSMIC_CN15
CNV48F
CNV48A
CNV48C
Somatic mutations (/Mb)
10
1
0.1
0.01
0.001
31/78
57/78
3/78
44/78
8/78
3/78
10/78
5/78

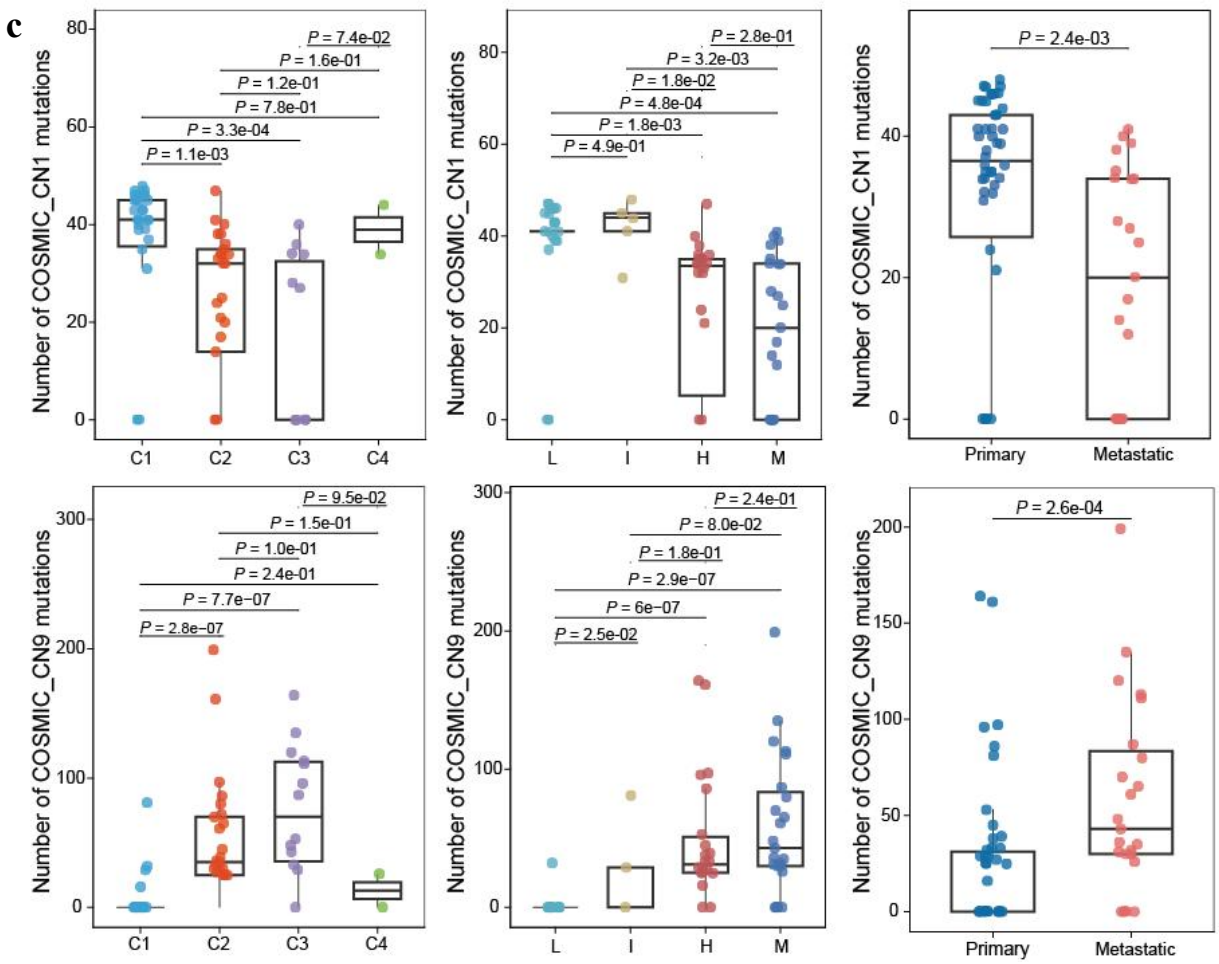

c
Number of COSMIC_CN1 mutations
P = 7.4e-02
P = 1.6e-01
P = 1.2e-01
P = 7.8e-01
P = 3.3e-04
P = 1.1e-03
C1
C2
C3
C4
P = 2.8e-01
P = 3.2e-03
P = 1.8e-02
P = 4.8e-04
P = 1.8e-03
P = 4.9e-01
L
I
H
M
P = 2.4e-03
Primary
Metastatic
Number of COSMIC_CN9 mutations
P = 9.5e-02
P = 1.5e-01
P = 1.0e-01
P = 2.4e-01
P = 7.7e-07
P = 2.8e-07
P = 2.4e-01
P = 8.0e-02
P = 1.8e-01
P = 2.9e-07
P = 6e-07
P = 2.5e-02
P = 2.6e-04

**Figure S13. CNV signatures in GISTs.**
**(a)** Total profiles of CNV signatures in 78 GISTs form 68 patients.
**(b)** Mutation burden of CNV signatures. Mutation burden per megabase of the CNV signatures was sorted by median (red line) with each dot representing one tumor and the number of tumors with signature indicated below.
**(c)** The correlation of CNV signature intensities with subtypes and clinicopathologic classifications. Boxplots showing the number of COSMIC_CN1/CN9 mutations in mRNA subtypes and clinicopathologic classifications, respectively. *P* values are determined by two-sided Wilcoxon rank-sum test. L, low-risk; I, intermediate-risk; H, high-risk and M, metastatic. The low bound, centerline, and upper bound of boxplot represent the first quartile, the median, and the third quartile of data, respectively; the upper and lower whiskers extend to the largest and smallest value within 1.5 times of the interquartile range (IQR) and the points greater than ±1.5 × IQR are plotted as outlier dots.
Source data are provided as a Source Data file.

**Figure S14**

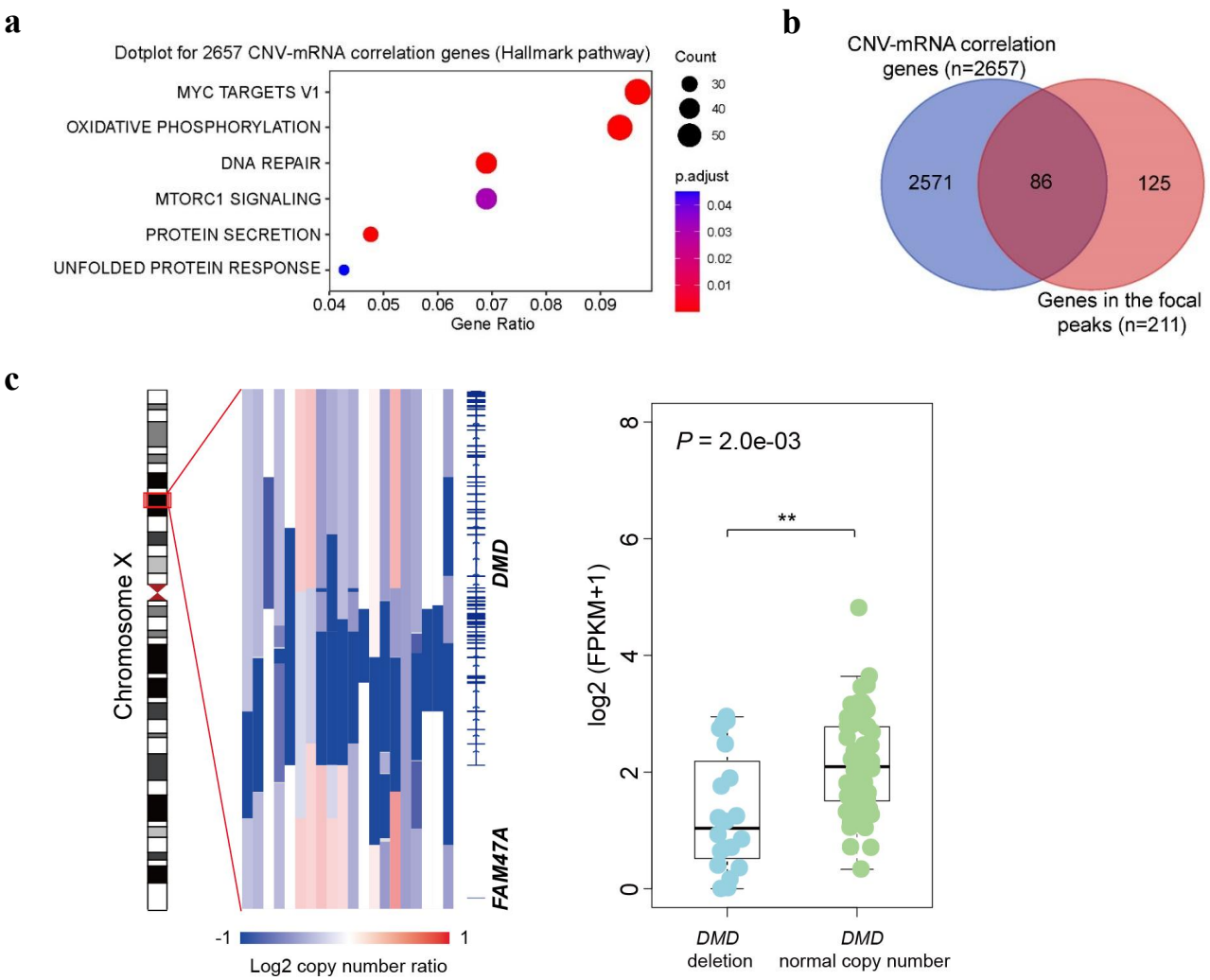
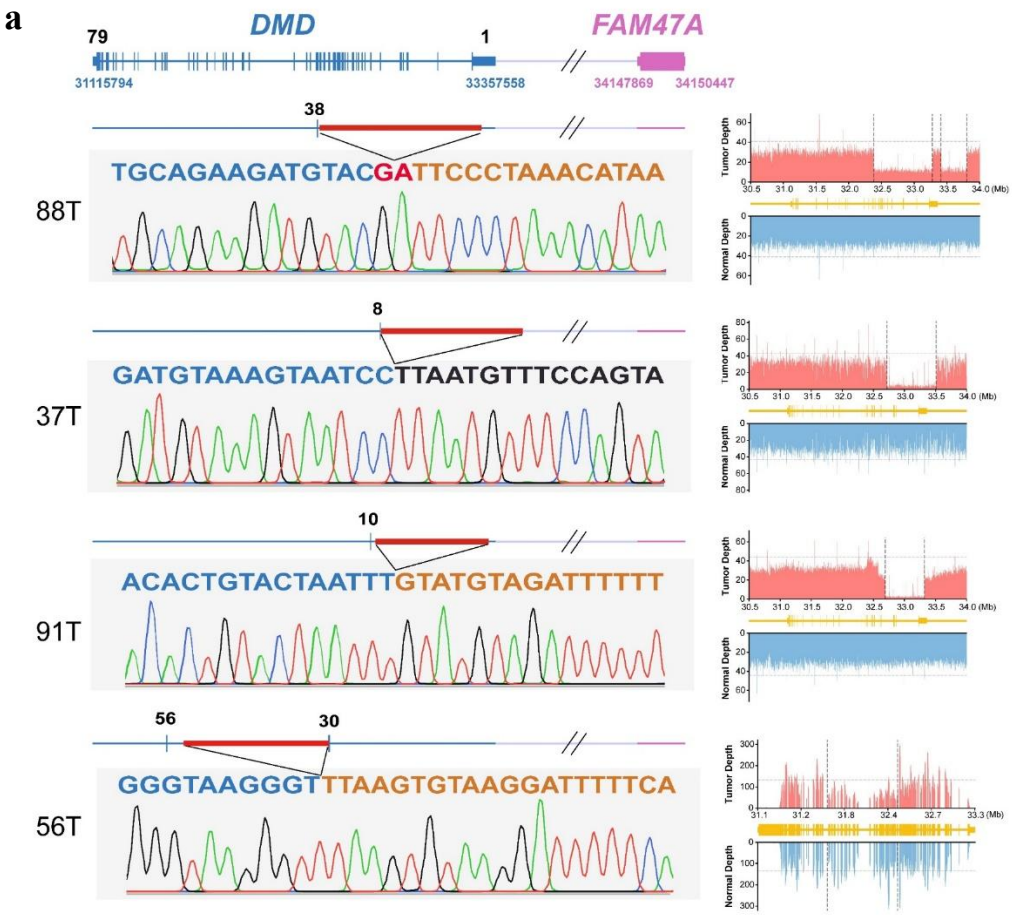
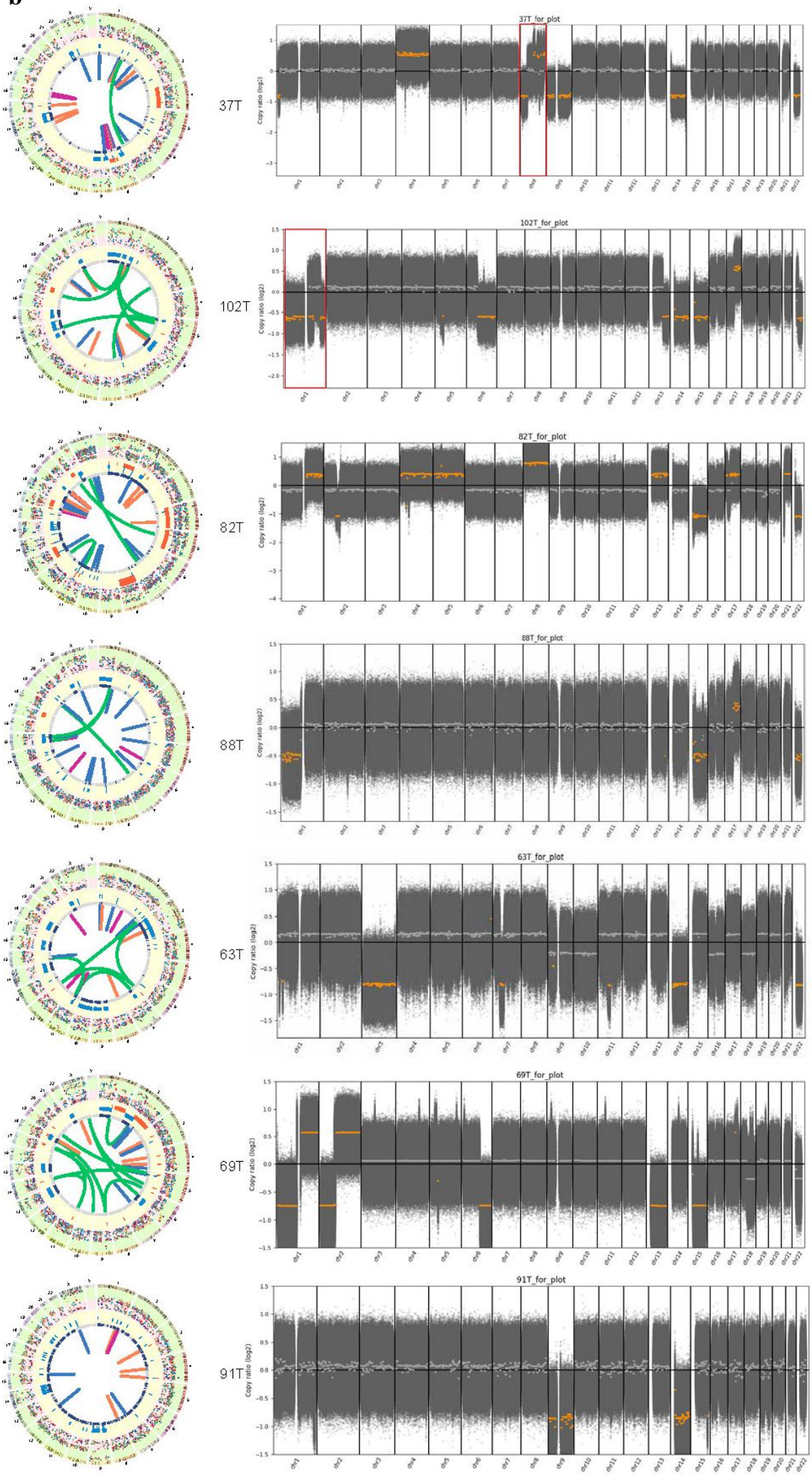

a
Dotplot for 2657 CNV-mRNA correlation genes (Hallmark pathway)
MYC TARGETS V1
OXIDATIVE PHOSPHORYLATION
DNA REPAIR
MTORC1 SIGNALING
PROTEIN SECRETION
UNFOLDED PROTEIN RESPONSE
0.04 0.05 0.06 0.07 0.08 0.09
Gene Ratio
Count
30
40
50
p.adjust
0.04
0.03
0.02
0.01
b
CNV-mRNA correlation genes (n=2657)
2571
86
125
Genes in the focal peaks (n=211)
c
Chromosome X
DMD
FAM47A
-1
1
Log2 copy number ratio
P = 2.0e-03
**
log2 (FPKM+1)
0
2
4
6
8
DMD deletion
DMD normal copy number

**Figure S14. CN/GE correlations, *DMD* copy number variations and CN-LOH of *KIT* in GISTs.**
**(a)** Hallmark pathways enrichment.
**(b)** Venn diagrams of CNV-mRNA correlation genes and genes in the focal peaks.
**(c)** Copy number variations of *DMD*. (Left) Schematic diagram displaying *DMD* CNV segments with the corresponding log2 ratio. (Right) Tumors with *DMD* deletions showing decreased mRNA expression than those with normal *DMD* copy number. *P* value is determined by two-sided Wilcoxon rank-sum test. The low bound, centerline, and upper bound of boxplot represent the first quartile, the median, and the third quartile of data, respectively; the upper and lower whiskers extend to the largest and smallest value within 1.5 times of the interquartile range (IQR) and the points greater than ±1.5 × IQR are plotted as outlier dots.

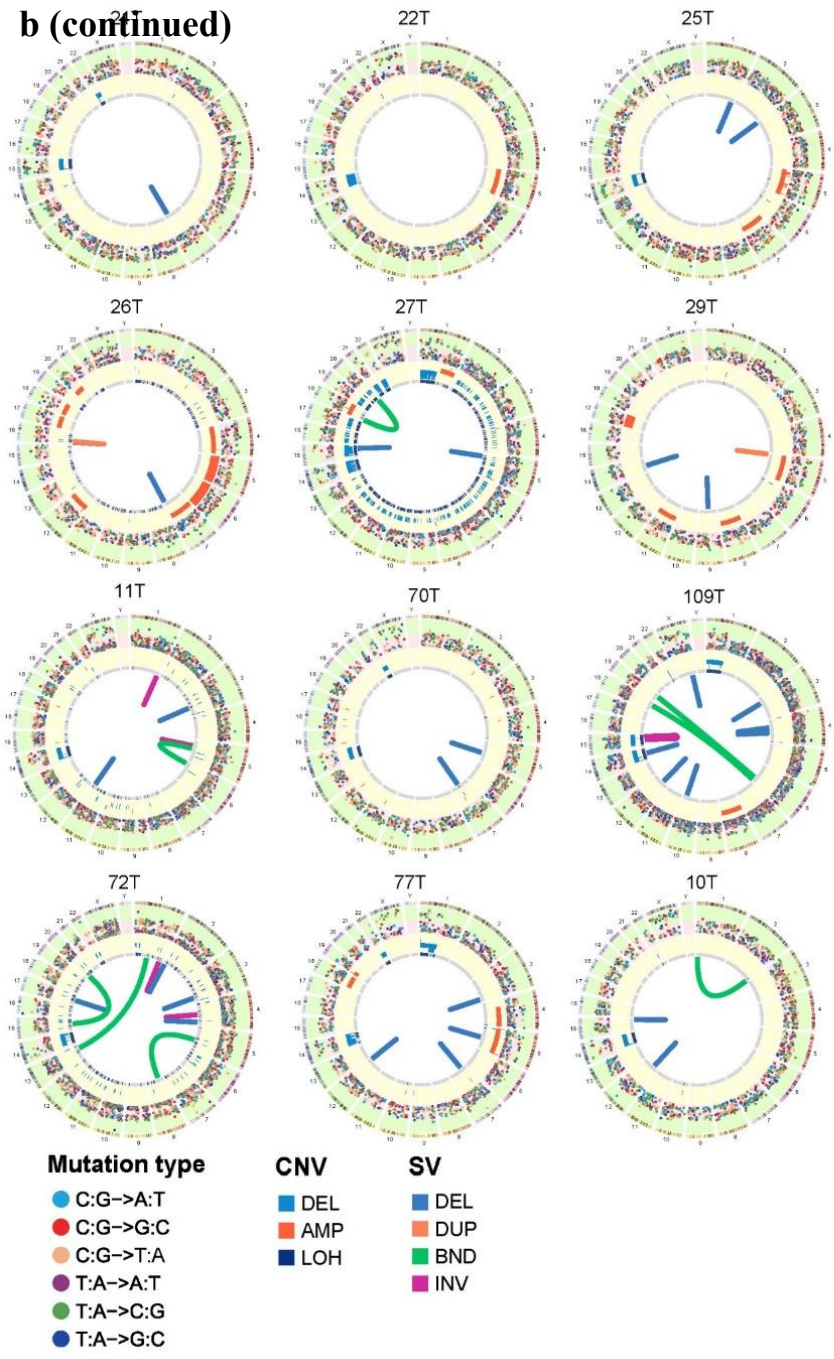
**Figure S15**

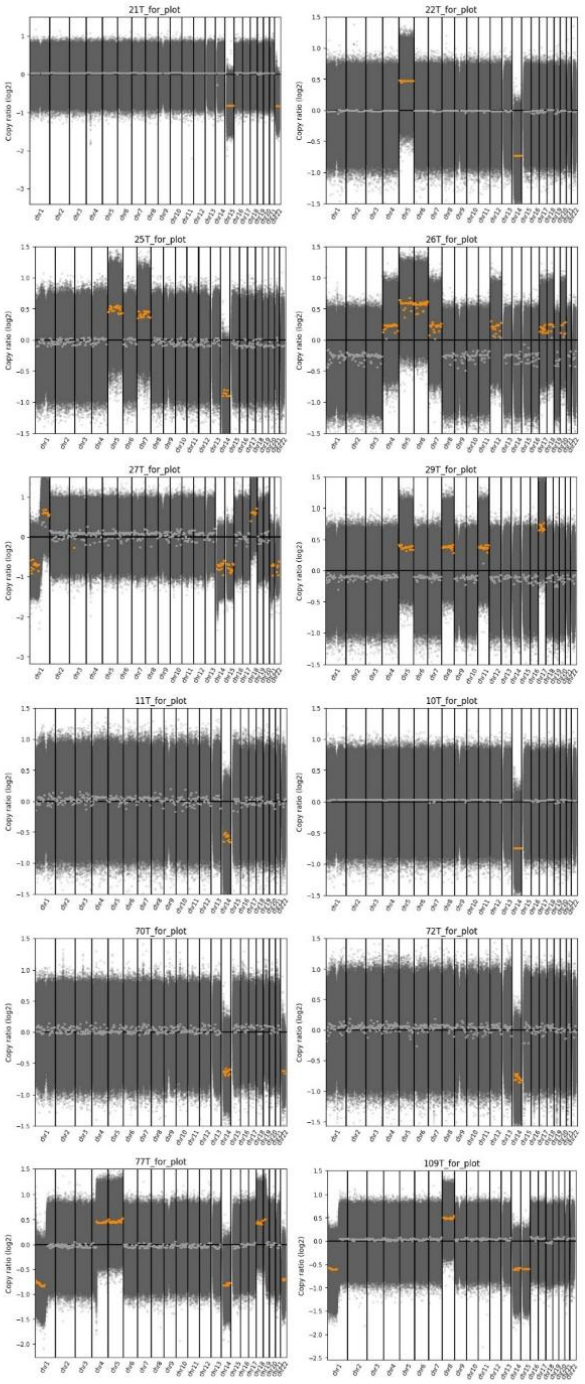

a
DMD
FAM47A
79
1
31115794
33357558
34147869
34150447
38
TGCAGAAGATGTACGATTCCCTAAACATAA
88T
8
GATGTAAAGTAATCCTTAATGTTTCCAGTA
37T
10
ACACTGTACTAATTTGTATGTAGATTTTTT
91T
56
30
GGGTAAGGGTTTAAGTGTAAGGATTTTTCA
56T
Tumor Depth
Normal Depth
(Mb)

**b**


37T
37T_for_plot
102T
102T_for_plot
82T
82T_for_plot
88T
88T_for_plot
63T
63T_for_plot
69T
69T_for_plot
91T
91T_for_plot
Copy ratio (log2)

**b (continued)**


21T
22T
25T
26T
27T
29T
11T
70T
109T
72T
77T
10T
Mutation type
C:G->A:T
C:G->G:C
C:G->T:A
T:A->A:T
T:A->C:G
T:A->G:C
CNV
DEL
AMP
LOH
SV
DEL
DUP
BND
INV


21T_for_plot
22T_for_plot
25T_for_plot
26T_for_plot
27T_for_plot
29T_for_plot
11T_for_plot
10T_for_plot
70T_for_plot
72T_for_plot
77T_for_plot
109T_for_plot
Copy ratio (log2)

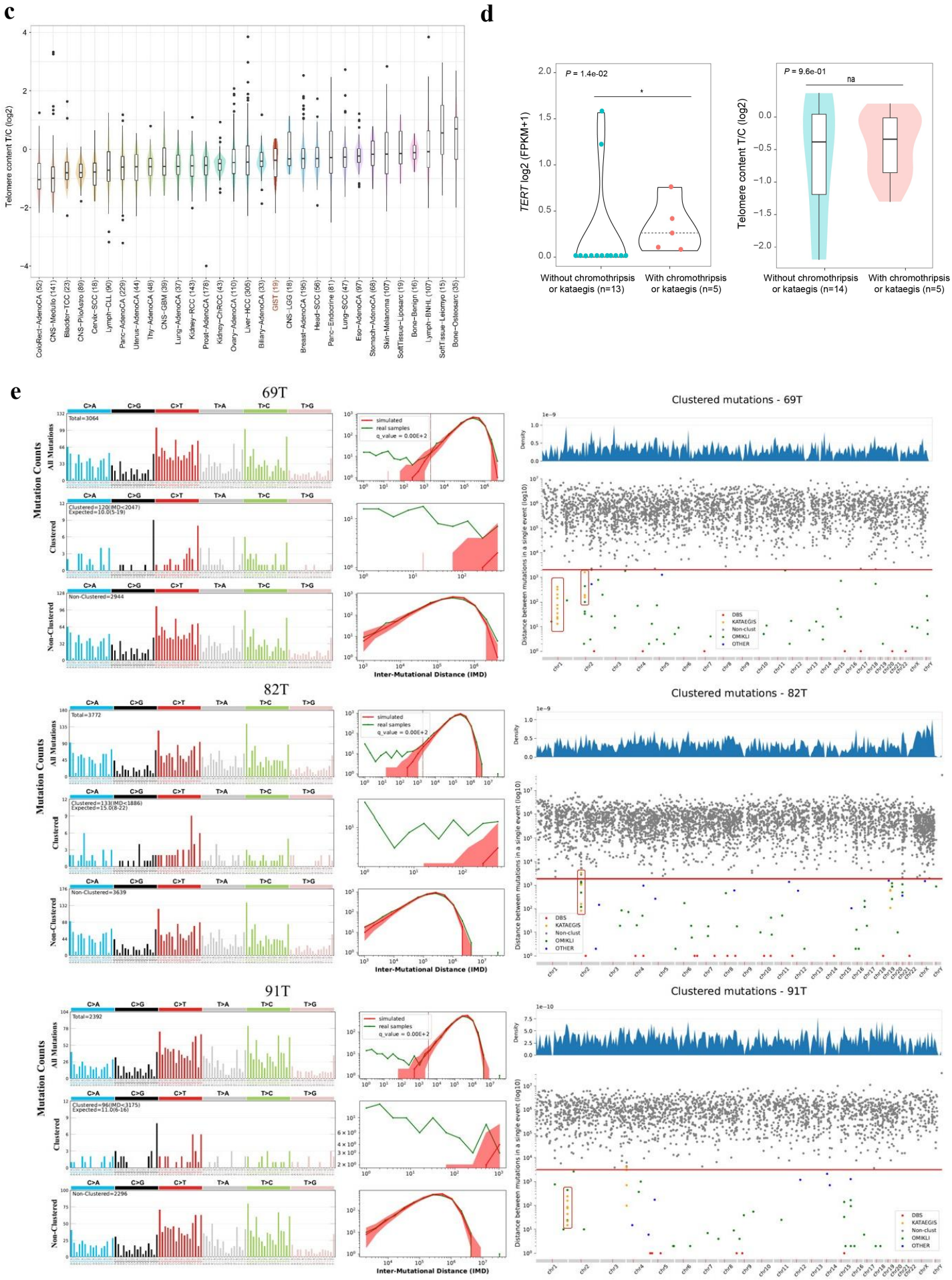

c
Telomere content T/C (log2)
ColoRect−AdenoCA (52)
CNS−Medullo (141)
Bladder−TCC (23)
CNS−PiloAstro (89)
Cervix−SCC (18)
Lymph−CLL (90)
Panc−AdenoCA (229)
Uterus−AdenoCA (44)
Thy−AdenoCA (48)
CNS−GBM (39)
Lung−AdenoCA (37)
Kidney−RCC (143)
Prost−AdenoCA (178)
Kidney−ChRCC (43)
Ovary−AdenoCA (110)
Liver−HCC (305)
Biliary−AdenoCA (33)
GIST (19)
CNS−LGG (18)
Breast−AdenoCA (195)
Head−SCC (56)
Panc−Endocrine (81)
Lung−SCC (47)
Eso−AdenoCA (97)
Stomach−AdenoCA (68)
Skin−Melanoma (107)
SoftTissue−Liposarc (19)
Bone−Benign (16)
Lymph−BNHL (107)
SoftTissue−Leiomyo (15)
Bone−Osteosarc (35)
d
TERT log2 (FPKM+1)
P = 1.4e-02
Without chromothripsis or kataegis (n=13)
With chromothripsis or kataegis (n=5)
Telomere content T/C (log2)
P = 9.6e-01
na
Without chromothripsis or kataegis (n=14)
With chromothripsis or kataegis (n=5)
e
69T
82T
91T
Mutation Counts
All Mutations
Clustered
Non-Clustered
C>A
C>G
C>T
T>A
T>C
T>G
Total=3064
Clustered=120(IMD<2047)
Expected=10.0(5-19)
Non-Clustered=2944
Total=3772
Clustered=133(IMD<1886)
Expected=15.0(8-22)
Non-Clustered=3639
Total=2392
Clustered=96(IMD<3175)
Expected=11.0(6-16)
Non-Clustered=2296
simulated
real samples
q_value = 0.00E+2
Inter-Mutational Distance (IMD)
Clustered mutations - 69T
Clustered mutations - 82T
Clustered mutations - 91T
Density
Distance between mutations in a single event (log10)
DBS
KATAEGIS
Non-clust
OMIKLI
OTHER

**Figure S15. Sanger validation of the 4 *DMD* deletions, circos plots and log2 copy ratio plots， telomere content and *TERT* expression in GISTs.**

**(a)** Sanger validation of the 4 *DMD* SVs identified by Manta. (Left) Sanger sequencing chromatograms showing *DMD* deletion regions; (Right) Comparison of the normalized *DMD* depth in tumor and paired normal samples. For WGS samples, the read depths of the tumor and paired normal tissues were calculated in a 100-bp window in the *DMD* gene and its adjacent region. For WES samples, only the read depths of the exon region were calculated in the *DMD* gene, while the intron region was scaled down by 200 times. To avoid the influence of sequencing depth on the calculation of read depth in the *DMD* gene, the read depth of the normal samples was normalized by the sequencing depth ratio between tumor and paired normal samples.

(**b**) Circos plots (left) and log2 copy ratio plots (right) for all analyzed WGS samples. The first (outer) circle shows the chromosomal positions. The second circle shows all SNVs and indels detected in the sample, with the position of each point indicating the VAF. The third circle shows the inferred copy number of each CNV segment, with amplification depicted in orange and deletions depicted in blue. The fourth circle shows all LOH events. The fifth (inner) circle shows all SVs. Red boxes on the copy ratio plots show chromothripsis detected by ShatterSeek.

**(c-d)** Telomere content and *TERT* expression in GISTs. The low bound, centerline, and upper bound of boxplot represent the first quartile, the median, and the third quartile of data, respectively; the upper and lower whiskers extend to the largest and smallest value within 1.5 times of the interquartile range (IQR) and the points greater than ±1.5 × IQR are plotted as outlier dots. (**c**) Comparison of telomere content in GIST and other human cancers. (**d**) TERT mRNA increases in GISTs with chromothripsis or kataegis (Left). Comparison of telomere content in GISTs with chromothripsis or kataegis versus without group (Right). Source data are provided as a Source Data file.

(**e**) The mutational spectra for different mutational pattern (left), rainfall plot illustrating the distribution of IMDs (right) for all analyzed WGS samples. Red boxes show the kataegis detected by SeqKat.

**Figure S16**

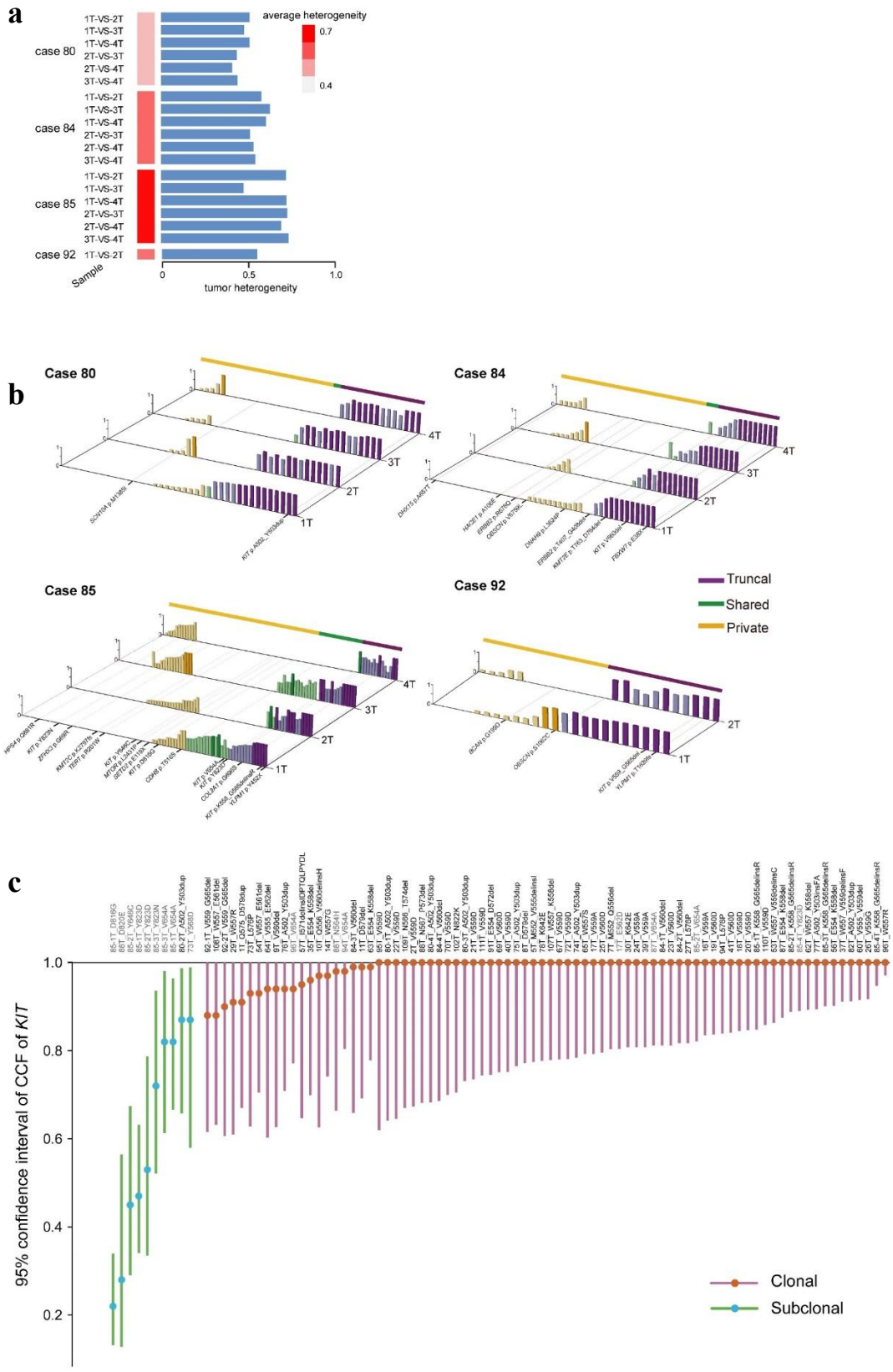

a
average heterogeneity
0.7
0.4
case 80
case 84
case 85
case 92
1T-VS-2T
1T-VS-3T
1T-VS-4T
2T-VS-3T
2T-VS-4T
3T-VS-4T
Sample
0
0.5
1.0
tumor heterogeneity
b
Case 80
Case 84
Case 85
Case 92
Truncal
Shared
Private
1T
2T
3T
4T
c
95% confidence interval of CCF of KIT
1.0
0.8
0.6
0.4
0.2
Clonal
Subclonal

**Figure S16. Genomic heterogeneity among spatially and temporally separated GISTs from individual patients and clonal analysis of *KIT* in GISTs.**

**(a)** Bar graphs showing the tumor genomic heterogeneity of 4 cases with multiple GISTs. Each bar represents a single tumor lesion.

**(b)** The cancer cell fraction (CCF) of the somatic mutations in the 4 cases with multiple GISTs. Individual mutations are shown on the horizontal axis, and CCF is plotted on the vertical axis. Selected cancer-associated genes are labeled on the horizontal axis. Purple = mutations present in all samples; green = mutations shared by partial samples, yellow = private mutations, light color bars indicate subclones and dark color bars indicate clones.

**(c)** The CCF of the somatic *KIT* mutations in GISTs. The probability distributions of the CCF are showed in red (Clonal) and blue (Subclonal). The primary mutations are highlighted in bold.

Source data are provided as a Source Data file.

**Figure S17**

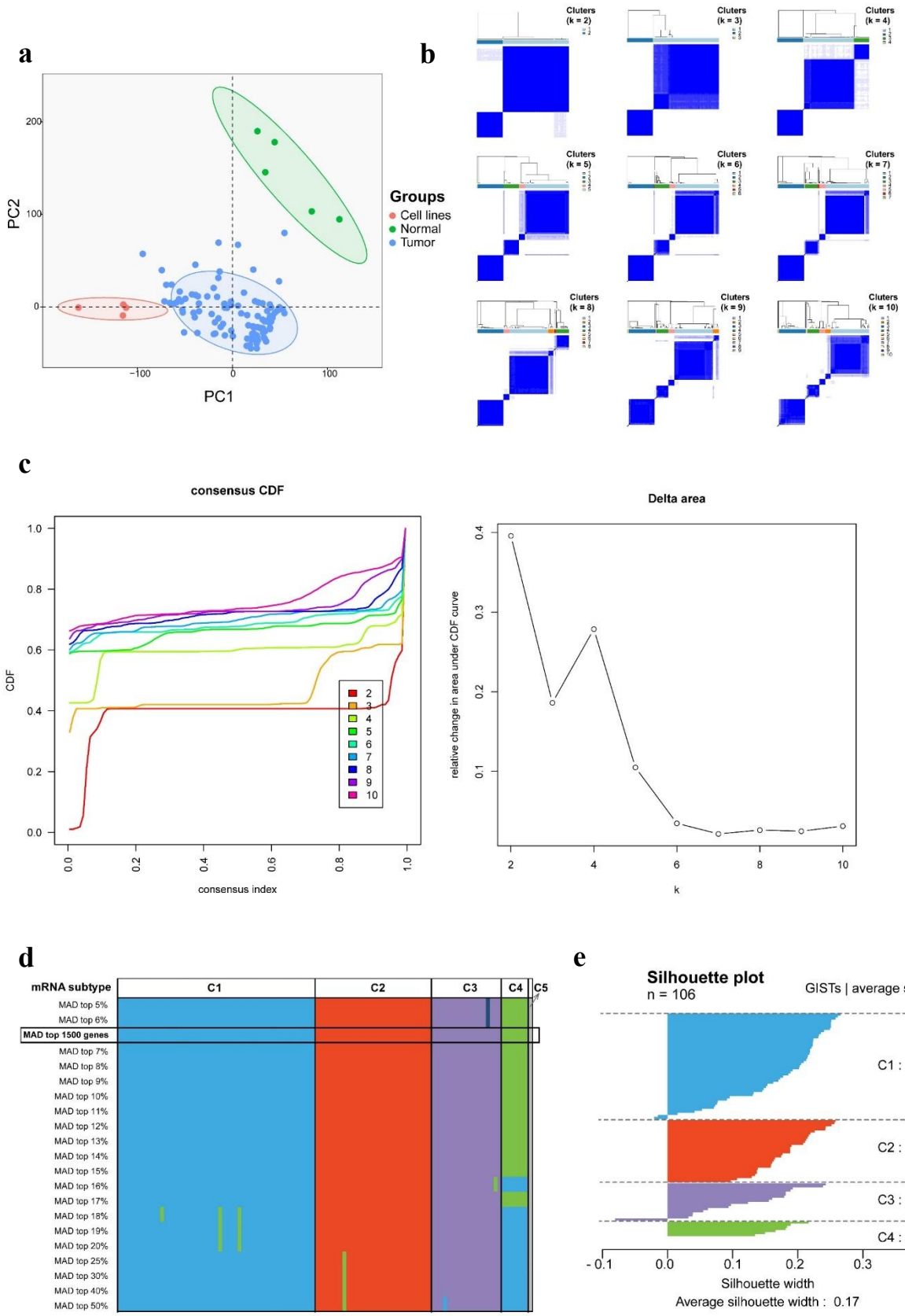

a
PC2
PC1
Groups
Cell lines
Normal
Tumor
b
Cluters (k = 2)
Cluters (k = 3)
Cluters (k = 4)
Cluters (k = 5)
Cluters (k = 6)
Cluters (k = 7)
Cluters (k = 8)
Cluters (k = 9)
Cluters (k = 10)
c
consensus CDF
CDF
consensus index
Delta area
relative change in area under CDF curve
k
d
mRNA subtype
C1
C2
C3
C4
C5
MAD top 5%
MAD top 6%
MAD top 1500 genes
MAD top 7%
MAD top 8%
MAD top 9%
MAD top 10%
MAD top 11%
MAD top 12%
MAD top 13%
MAD top 14%
MAD top 15%
MAD top 16%
MAD top 17%
MAD top 18%
MAD top 19%
MAD top 20%
MAD top 25%
MAD top 30%
MAD top 40%
MAD top 50%
e
Silhouette plot
n = 106
GISTs | average silhouette width
C1 : n = 51 | 0.17
C2 : n = 30 | 0.19
C3 : n = 18 | 0.12
C4 : n = 7 | 0.17
Silhouette width
Average silhouette width : 0.17

**Figure S17, related to Figure 7. Processing and validation of mRNA clustering.**

**(a)** Principal component analysis divides the samples into 3 groups (green, 5 normal tissues; red, 4 GIST cell lines; blue, 107 frozen GISTs).

**(b)** Heatmap shows consensus matrices of the 107 GIST samples with mRNA data for k = 2 to k = 10 resulting from the consensus clustering approach.

**(c)** Determination of the optimal cluster number of 107 GISTs. (Right) Consensus empirical cumulative distribution function (CDF) with different k values (k = 2 to k = 10); (Left) delta area changes with the number of clusters. The x-axis represents the number of k and the y-axis represents the relative change in area under CDF curve. The optimal cluster number was five, but cluster 5 only included a WT GIST and were excluded from the following analysis, so we chose four as the optimal cluster number.

**(d)** Number of selected genes (based on the median absolute deviation, MAD) in the clustering results.

**(e)** Silhouette plot of clustering results (4 clusters) in 106 GISTs using the top 1500 genes (MAD).

Source data are provided as a Source Data file.

**Figure S18**

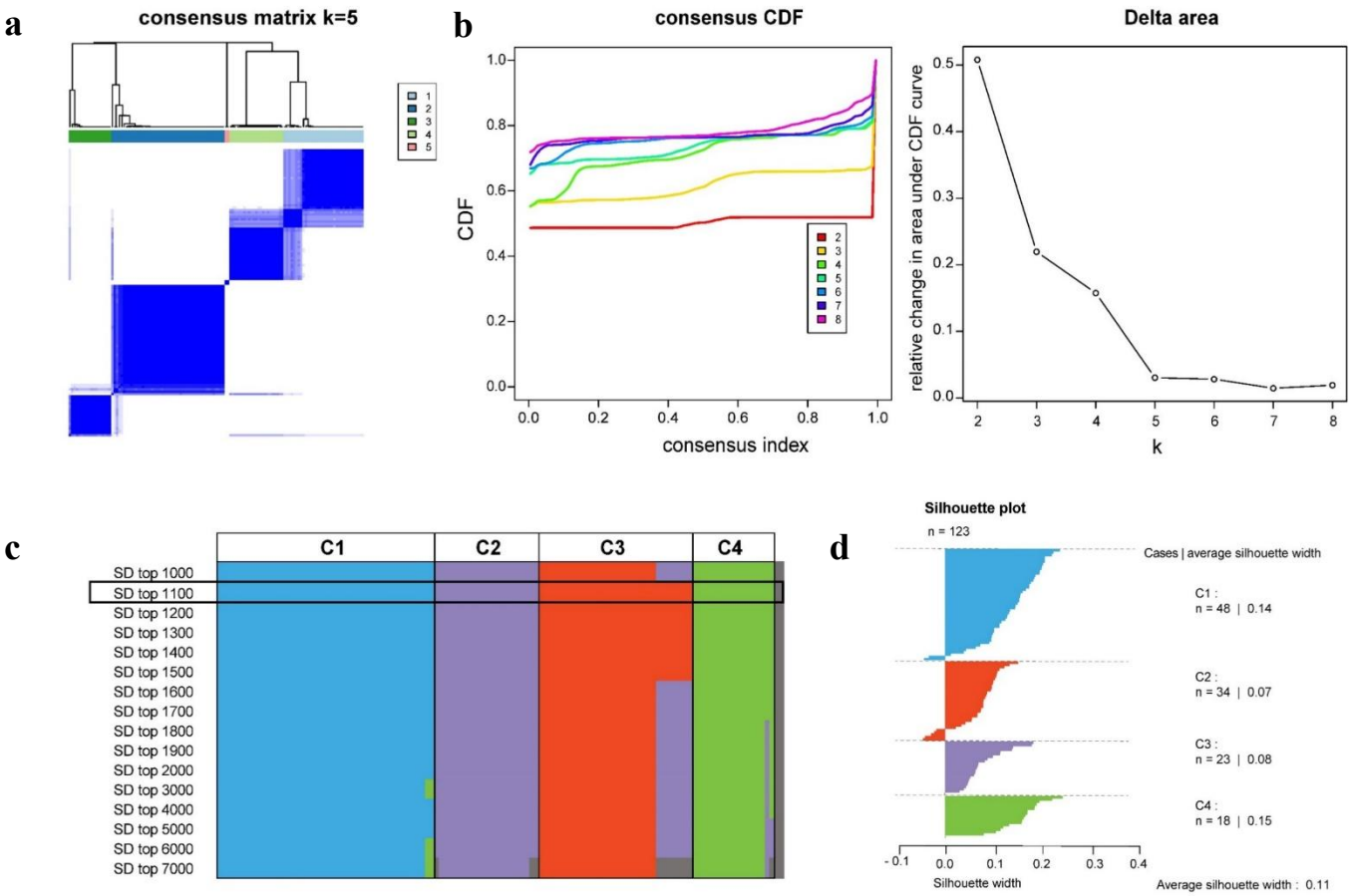

a
consensus matrix k=5
1
2
3
4
5
b
consensus CDF
CDF
consensus index
2
3
4
5
6
7
8
Delta area
relative change in area under CDF curve
k
c
C1
C2
C3
C4
SD top 1000
SD top 1100
SD top 1200
SD top 1300
SD top 1400
SD top 1500
SD top 1600
SD top 1700
SD top 1800
SD top 1900
SD top 2000
SD top 3000
SD top 4000
SD top 5000
SD top 6000
SD top 7000
d
Silhouette plot
n = 123
Cases | average silhouette width
C1 :
n = 48 | 0.14
C2 :
n = 34 | 0.07
C3 :
n = 23 | 0.08
C4 :
n = 18 | 0.15
Silhouette width
Average silhouette width : 0.11

**Figure S18. Validation of mRNA-based subtypes of GISTs in Japanese (n = 65) and Complexity Index in Sarcomas (CINSARC) cohorts (n = 60).**

**(a)** Consensus clustering matrices of 125 GISTs in validation cohorts with mRNA data for k = 2 to k = 10. Heatmap shows consensus matrices of the 125 GISTs with five clusters (k = 5).

**(b)** Determination of the optimal cluster number of 125 GISTs form validation cohorts. (Left) Consensus empirical cumulative distribution function (CDF) with different k values (k = 2 to k = 10); (Right) delta area changes with the number of clusters. The x-axis represents the number of k and the y-axis represents the relative change in area under CDF curve. The optimal cluster number was five, but cluster 5 only included 2 KIT/PDGFRA wild type GISTs and were excluded from the following analysis, so we chose four as the optimal cluster number

**(c)** Number of selected genes (based on standard deviation, SD) in the clustering results.

**(d)** Silhouette plot of clustering results (4 clusters) in 123 GISTs using the top 1100 genes (SD).

Source data are provided as a Source Data file.

**Figure S19**

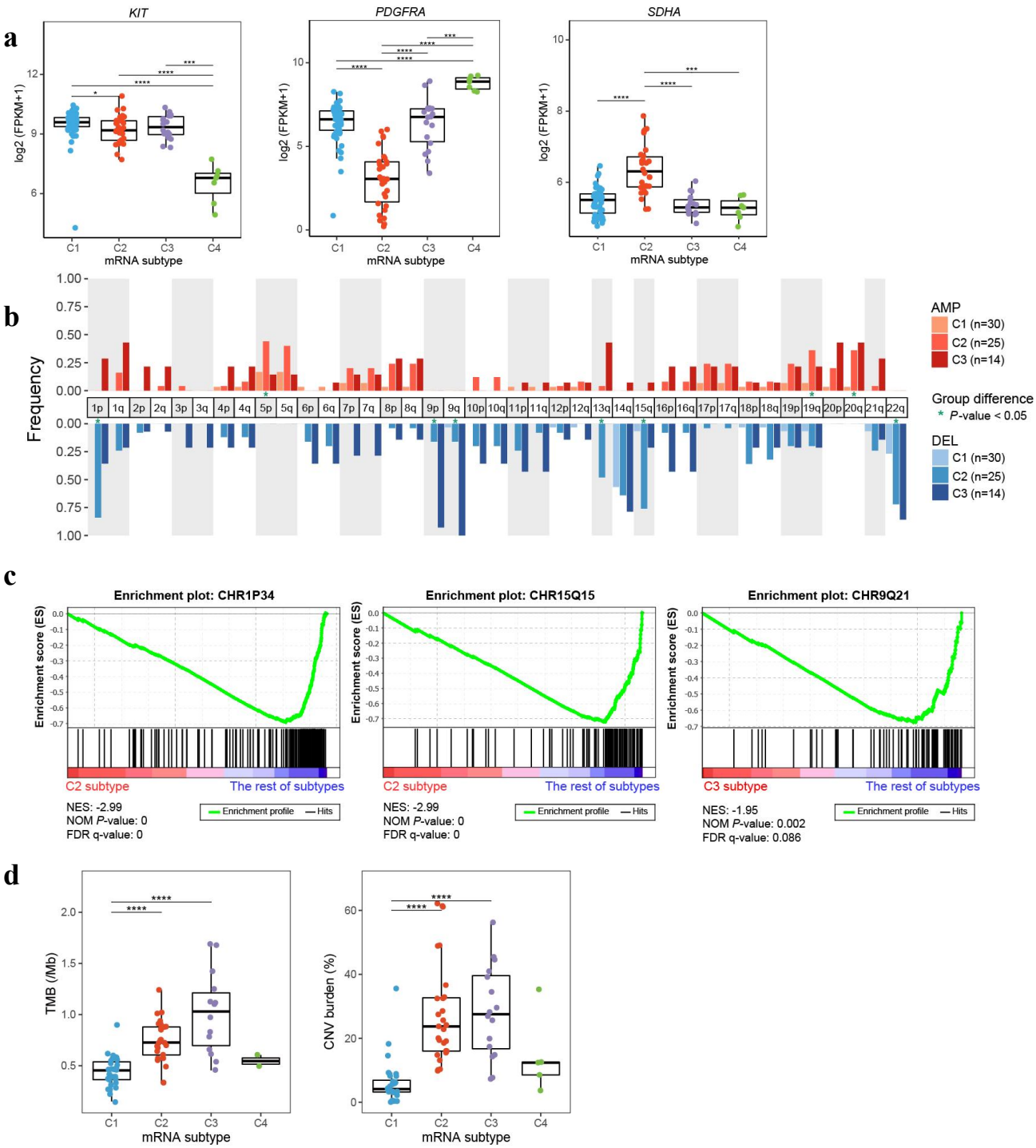

a
KIT
PDGFRA
SDHA
log2 (FPKM+1)
mRNA subtype
C1
C2
C3
C4
b
Frequency
AMP
C1 (n=30)
C2 (n=25)
C3 (n=14)
Group difference
* P-value < 0.05
DEL
C1 (n=30)
C2 (n=25)
C3 (n=14)
1p 1q 2p 2q 3p 3q 4p 4q 5p 5q 6p 6q 7p 7q 8p 8q 9p 9q 10p 10q 11p 11q 12p 12q 13q 14q 15q 16p 16q 17p 17q 18p 18q 19p 19q 20p 20q 21q 22q
c
Enrichment plot: CHR1P34
Enrichment plot: CHR15Q15
Enrichment plot: CHR9Q21
Enrichment score (ES)
C2 subtype
The rest of subtypes
C3 subtype
NES: -2.99
NOM P-value: 0
FDR q-value: 0
NES: -1.95
NOM P-value: 0.002
FDR q-value: 0.086
Enrichment profile
Hits
d
TMB (/Mb)
CNV burden (%)
mRNA subtype

**Figure S19. Molecular features in GIST mRNA subtypes.**

**(a)** Boxplots of differentially expressed driver genes among the 4 mRNA subtypes.

**(b)** Chromosome arm-level CNV frequencies across 4 mRNA subtype of 69 GISTs. Dark red, red and light red represent the amplification (AMP) frequencies of GISTs in C1 (n=30), C2 (n=25) and C3 (n=14) subtype, respectively. Dark blue, blue and light blue represent the deletion (DEL) frequencies of GISTs in C1 (n=30), C2 (n=25) and C3 (n=14) subtype, respectively. Arms with significant group differences are denoted by green asterisks (both the GISTIC q value and chi-square test p value less than 0.05)

**(c)** Gene set enrichment analysis (GSEA) of expression data for 4 mRNA subtypes, utilizing MSigDB positional gene sets (MSigDB v7.0, collection C1). NES, normalized enrichment score.

**(d)** Comparisons of tumor mutational burdens (somatic coding mutation and CNVs) among the 4 mRNA subtypes.

In (a) and (d), *P* values were calculated by the two-sided wilcoxon rank sum test. The low bound, centerline, and upper bound of boxplot represent the first quartile, the median, and the third quartile of data, respectively; the upper and lower whiskers extend to the largest and smallest value within 1.5 times of the interquartile range (IQR) and the points greater than $\pm 1.5 \times$ IQR are plotted as outlier dots.

**Figure S20**

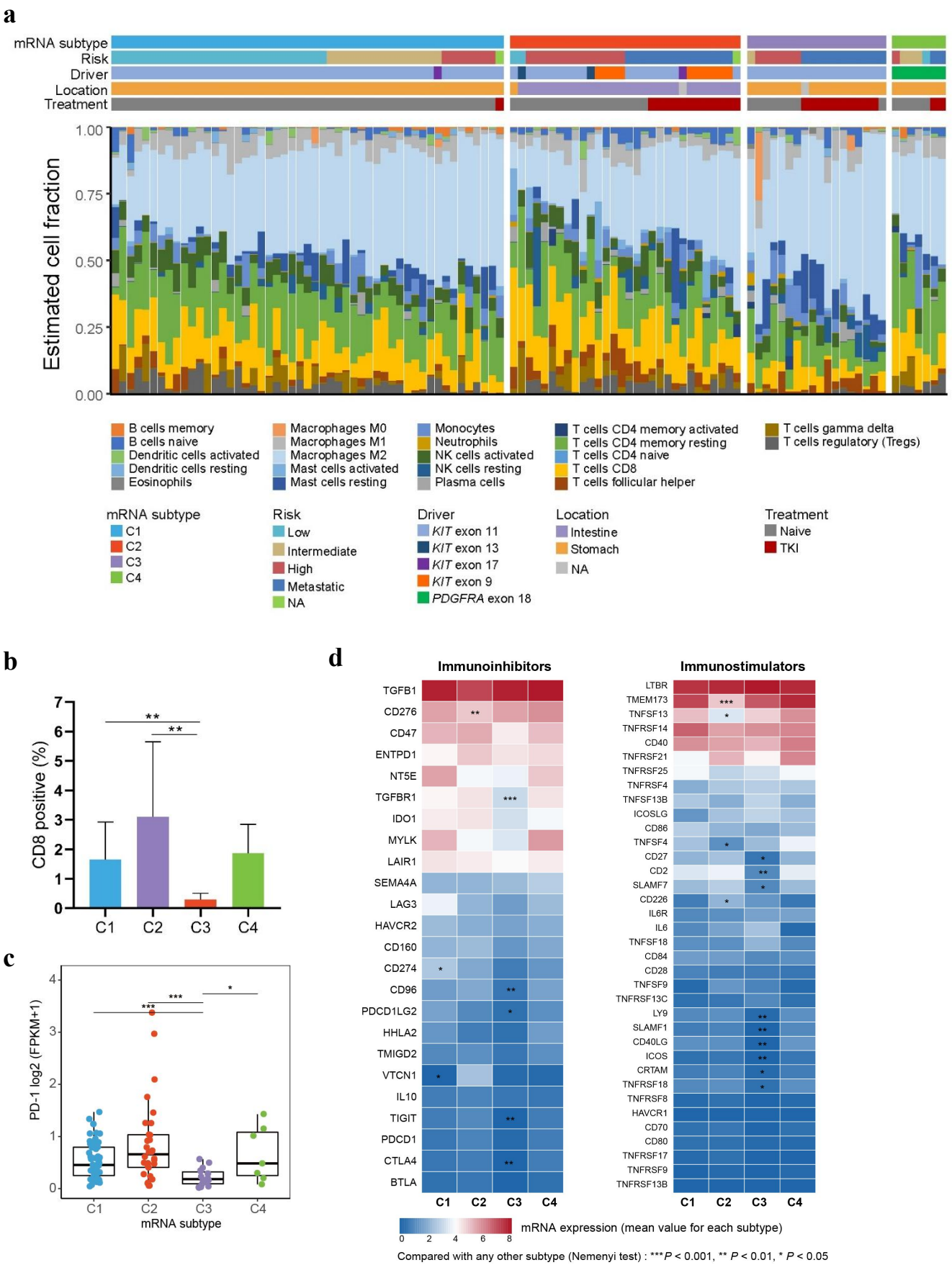

a
mRNA subtype
Risk
Driver
Location
Treatment
Estimated cell fraction
1.00
0.75
0.50
0.25
0.00
B cells memory
B cells naive
Dendritic cells activated
Dendritic cells resting
Eosinophils
Macrophages M0
Macrophages M1
Macrophages M2
Mast cells activated
Mast cells resting
Monocytes
Neutrophils
NK cells activated
NK cells resting
Plasma cells
T cells CD4 memory activated
T cells CD4 memory resting
T cells CD4 naive
T cells CD8
T cells follicular helper
T cells gamma delta
T cells regulatory (Tregs)
mRNA subtype
C1
C2
C3
C4
Risk
Low
Intermediate
High
Metastatic
NA
Driver
KIT exon 11
KIT exon 13
KIT exon 17
KIT exon 9
PDGFRA exon 18
Location
Intestine
Stomach
NA
Treatment
Naive
TKI
b
CD8 positive (%)
C1
C2
C3
C4
c
PD-1 log2 (FPKM+1)
mRNA subtype
d
Immunoinhibitors
TGFB1
CD276
CD47
ENTPD1
NT5E
TGFBR1
IDO1
MYLK
LAIR1
SEMA4A
LAG3
HAVCR2
CD160
CD274
CD96
PDCD1LG2
HHLA2
TMIGD2
VTCN1
IL10
TIGIT
PDCD1
CTLA4
BTLA
Immunostimulators
LTBR
TMEM173
TNFSF13
TNFRSF14
CD40
TNFRSF21
TNFRSF25
TNFRSF4
TNFSF13B
ICOSLG
CD86
TNFSF4
CD27
CD2
SLAMF7
CD226
IL6R
IL6
TNFSF18
CD84
CD28
TNFSF9
TNFRSF13C
LY9
SLAMF1
CD40LG
ICOS
CRTAM
TNFRSF18
TNFRSF8
HAVCR1
CD70
CD80
TNFRSF17
TNFRSF9
TNFRSF13B
mRNA expression (mean value for each subtype)
Compared with any other subtype (Nemenyi test) : ***P < 0.001, ** P < 0.01, * P < 0.05

**Figure S20. Immune characterization in GIST mRNA subtypes.**

**(a)** Bar plots showing the percentage of 22 infiltrating immune cell types in the 4 mRNA subtypes (calculated using the CIBERSORT algorithm, 547 signature genes).

**(b)** Bar plots showing the proportion of CD8 positive cells in 30 GISTs with IHC. The percentage of positive was quantitatively evaluated using QuPath's built-in 'Positive cell detection'. Data are presented as mean values ± s.d. n=5. **, $P$<0.01. The $P$ values are calculated using the two-sided Student's t test.

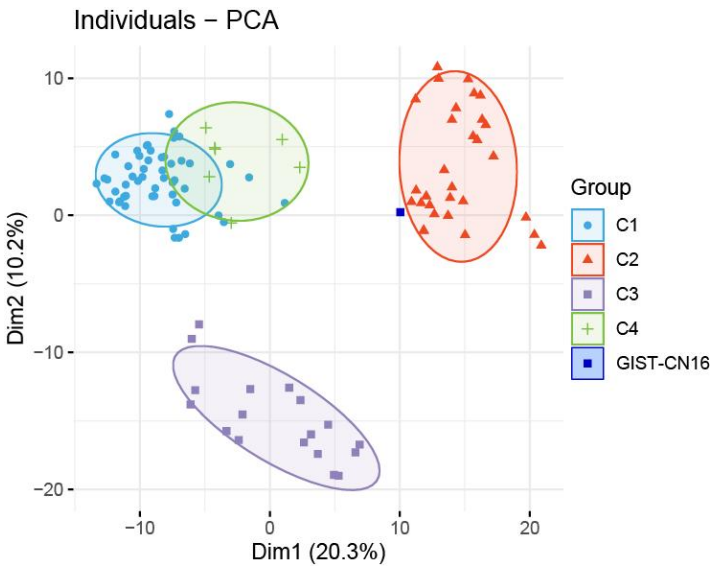
**(c)** Boxplot of gene expression for PD-1. The low bound, centerline, and upper bound of boxplot represent the first quartile, the median, and the third quartile of data, respectively; the upper and lower whiskers extend to the largest and smallest value within 1.5 times of the interquartile range (IQR) and the points greater than $\pm 1.5 \times$ IQR are plotted as outlier dots. $P$ values were calculated by the two-sided wilcoxon rank sum test.

**(d)** The expression of immunoinhibitory (left) and immunostimulatory molecules (right) in the 4 mRNA subtypes. The Nemenyi test was conducted and annotated as follows (compared with any other subtypes): ***, $P < 0.001$, **, $P < 0.01$, *, $P < 0.05$.

Source data are provided as a Source Data file.

**Figure S21**

**a**




**b**

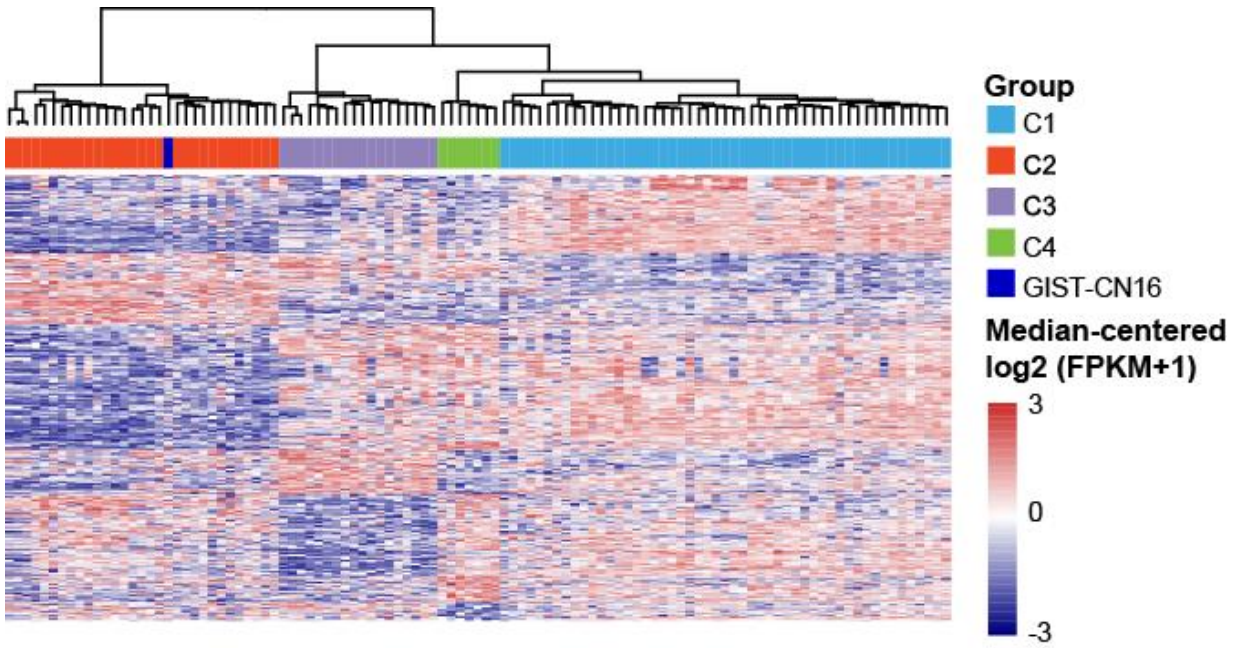


**Figure S21. GIST-CN16 primary cells belong to the C2 subtype.**
(a) Principal component analysis identifies that GIST-CN16 belongs to the C2 subtype.
(b) The transcriptome heatmap shows that GIST-CN16 belongs to the C2 subtype.
Source data are provided as a Source Data file.

**Figure S22**

**Fig. 3C**

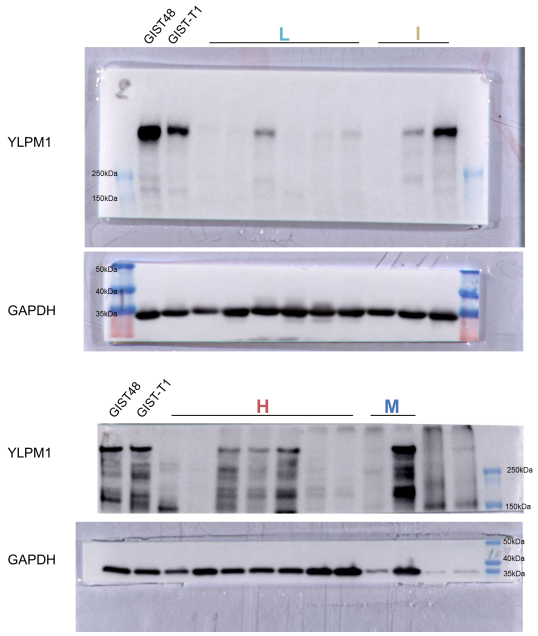



**Fig. S4a**

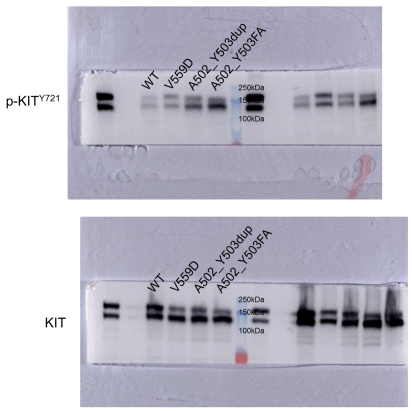



**Fig. S4b**

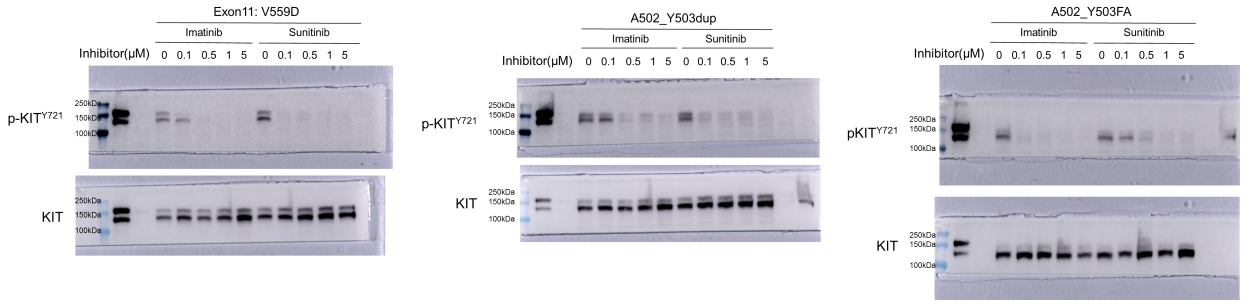
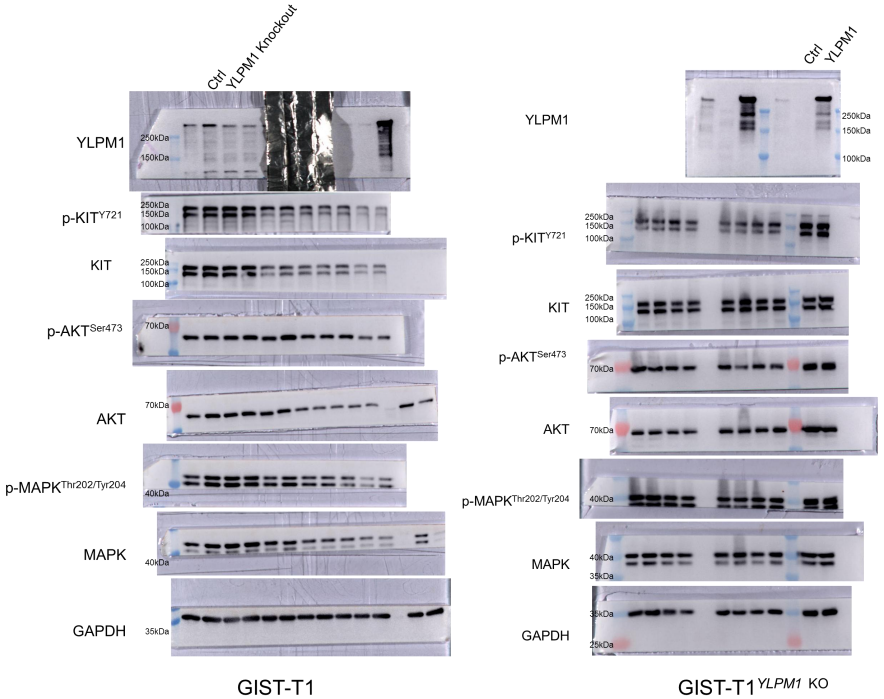



**Fig.S11a**




GIST-T1 GIST-T1$^{YLPM1\ KO}$

**Figure S22. Uncropped blots of Fig. 3c, Fig. S4a, Fig. S4b and Fig. S11a.**

## References

1. Alexandrov, L.B. *et al.* The repertoire of mutational signatures in human cancer. *Nature* **578**, 94-101 (2020).

2. Cancer Genome Atlas Research Network. Comprehensive and integrated genomic characterization of adult soft tissue sarcomas. *Cell* **171**, 950-965 e28 (2017).