
a
Austria
Center IDH wt/mut patients
Linz 13/0
St.Pölten
Vienna 117/11
10/0
14/1 Salzburg
14/0
Wiener Neustadt
21/2 Innsbruck
16/0 Graz
Klagenfurt
12/0
b
Patient status
alive
dead
lost
Glioblastoma validation cohort
Median age at diagnosis: 65 years
Median time to progression: 8.4 months
Median overall survival: 8.8 months
Male/Female ratio: 62 (59%) / 43 (41%)
c
Frozen prog: r=0.751 (N=14)
Frozen val: r=0.785 (N=44)
FFPE prog: r=0.14 (N=320)
FFPE val: r=0.425 (N=106)
RCL: r=0.08 (N=15)
Unique CpGs (million)
Raw reads (million)
RCL: r=−0.823 (N=15)
FFPE val: r=0.139 (N=106)
FFPE prog: r=−0.086 (N=320)
Frozen prog: r=0.115 (N=14)
Frozen val: r=0.107 (N=44)
Randomly fragmented reads (%)
Surgery date
RCL: r=−0.191 (N=15)
FFPE prog: r=0.131 (N=274)
FFPE val: r=−0.012 (N=99)
Frozen prog: r=0.129 (N=12)
Frozen val: r=0.293 (N=343)
Relative share necrosis
d
Progression cohort
Validation cohort
DNA methylation (%)
Methylated spike-in control
Unmethylated spike-in control
FFPE
Frozen
RCL
e
Progression cohort
Validation cohort
Promoters (1.5 kb)
Tiling regions (5 kb)
Enhancers
DNA methylation (%)
FFPE N=2.7e6
Frozen N=8.4e4
RCL N=7.8e4
FFPE N=1.6e7
Frozen N=9.4e5
RCL N=2.6e5
FFPE N=1.7e6
Frozen N=1.2e5
RCL N=1.9e4
FFPE N=1.1e6
Frozen N=5.8e5
FFPE N=6.0e6
Frozen N=5.4e6
FFPE N=6.6e5
Frozen N=6.8e5
Sample type
f
Progression cohort
Validation cohort
Promoters (1.5 kb)
Tiling regions (5 kb)
Enhancers
DNA methylation (%)
N=5.5e5
N=1.1e6
N=1.0e6
N=2.2e5
N=4.7e6
N=6.9e6
N=4.7e6
N=7.3e5
N=5.8e5
N=7.7e5
N=4.7e5
N=6.2e4
N=7.4e5
N=3.5e5
N=5.6e5
N=3.1e4
N=6.6e6
N=2.1e6
N=2.6e6
N=1.0e5
N=8.2e5
N=2.4e5
N=2.6e5
N=8.3e3
Quality tier
g
MDS1
MDS2
Surgery date
2015
2010
2005
2000
Sequencing date
2018−01
2017−07
2017−01
2016−07
2016−01
Cohort
Progression
Validation
Center
Graz
Innsbruck
Klagenfurt
Linz
Vienna
Salzburg
St. Poelten
Wr. Neustadt
Sample material
FFPE
Frozen
RCL
Diagnosis
Glioblastoma
Glioma
Epilepsy (control)
h
FFPE
Frozen
RCL
r=0.943 N=344666
r=0.941 N=44357
r=0.944 N=7309
DNA methylation recurring tumor (%)
DNA methylation primary tumor (%)
Density
high
low

Supplementary Figure 1

RRBS profiling in a national glioblastoma cohort

(a) Overview of the clinical centers in Austria that contributed samples and data to this study. The number of *IDH*-wildtype and *IDH*-mutated patients are indicated for each center.

(b) Overview of the validation cohort, summarizing the disease courses of 105 primary glioblastoma patients with *IDH*-wildtype status and only one tumor resection, ordered by time of first surgery.

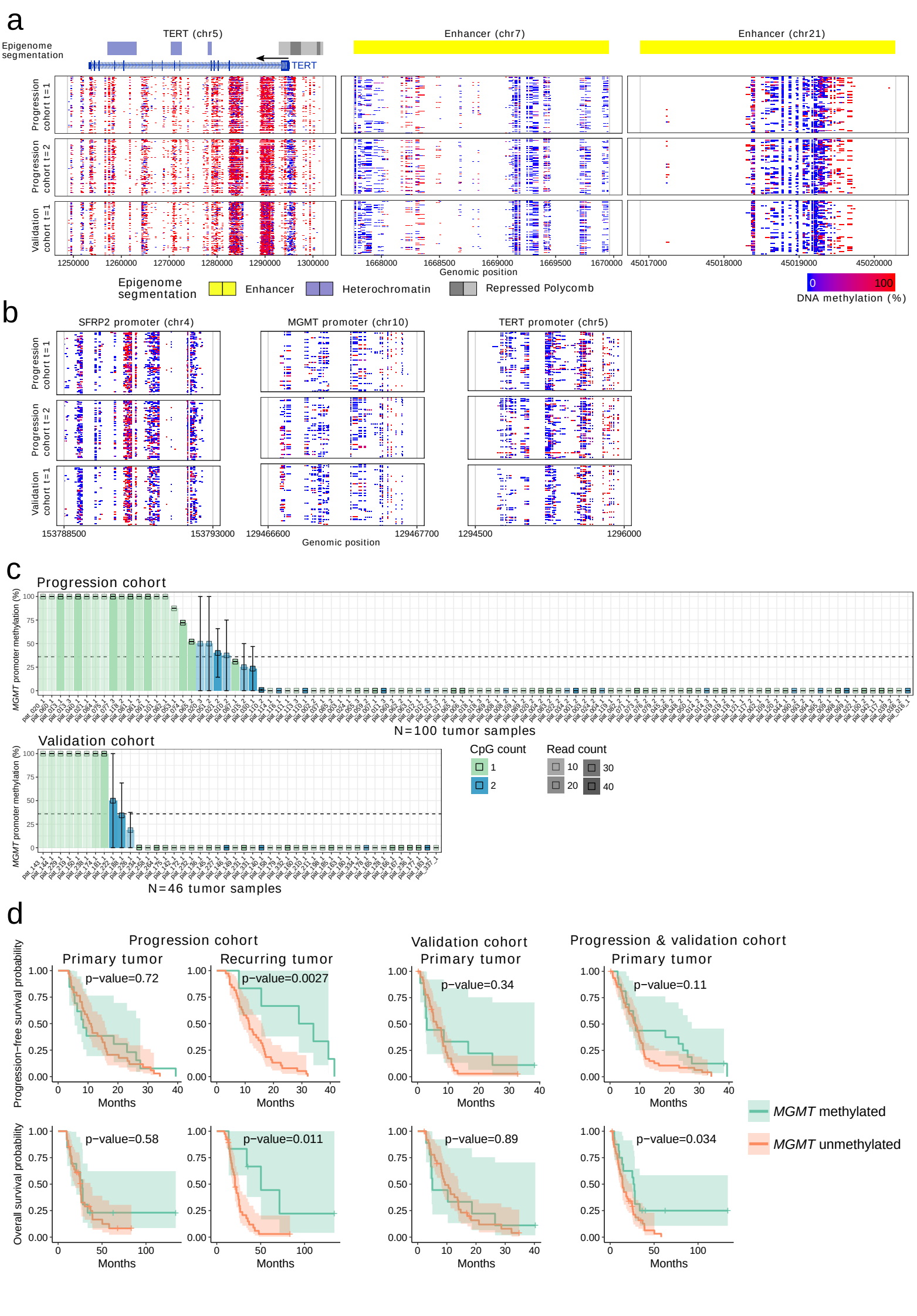
(c) Scatterplots summarizing the RRBS profiling data. A high number of unique CpGs and a low proportion of randomly fragmented reads are indicative of high-quality samples with limited pre-fragmentation of the input DNA. The cohorts (prog: progression, val: validation) and sample types (FFPE: formalin-fixed paraffin-embedded; Frozen: fresh-frozen, RCL: ethanol-based conservation) are indicated by color. r: Pearson correlation. N denotes the number of samples in each category.

(d) DNA methylation levels of methylated and unmethylated synthetic spike-in control DNA. Dashed lines indicate DNA methylation levels of 5% and 95%.

(e,f) Distribution of DNA methylation levels across different types of genomic regions (including all regions covered on average by more than 10 reads per CpG) for different sample types (e) and sample quality tiers (f). The quality tiers (Supplementary Table 2) were defined by the number of unique CpGs detected in each RRBS library (tier 1: more than 3 million; tier 2: between 2 and 3 million; tier 3: between 1 and 2 million; tier 4: below 1 million). Enhancer regions: Astrocyte enhancers as defined by the Roadmap Epigenomics Project[43]. N denotes the number of data points (samples × genomic regions).

(g) Multidimensional scaling plots depicting the relative similarity of all samples included in this study based on all 5-kilobase tiling regions. Colors indicate potential sources of batch effects (surgery date, sequencing date, cohort, center) as well as factors that are known or expected to influence the DNA methylation profiles (sample material, diagnosis).

(h) Scatterplots depicting the relationship of DNA methylation levels in 5-kilobase tiling regions (containing more than 25 CpGs and covered on average by more than 10 reads per CpG) between primary and recurring tumors for the three different sample types. r: Pearson correlation. N denotes the number of data points (samples × genomic regions).


a
TERT (chr5)
Enhancer (chr7)
Enhancer (chr21)
Epigenome segmentation
TERT
Progression cohort t=1
Progression cohort t=2
Validation cohort t=1
1250000
1260000
1270000
1280000
1290000
1300000
1668000
1668500
1669000
1669500
1670000
45017000
45018000
45019000
45020000
Genomic position
Epigenome segmentation
Enhancer
Heterochromatin
Repressed Polycomb
0
100
DNA methylation (%)
b
SFRP2 promoter (chr4)
MGMT promoter (chr10)
TERT promoter (chr5)
153788500
153793000
129466600
129467700
1294500
1296000
Genomic position
c
Progression cohort
MGMT promoter methylation (%)
N=100 tumor samples
Validation cohort
N=46 tumor samples
CpG count
1
2
Read count
10
20
30
40
d
Progression cohort
Primary tumor
Recurring tumor
Validation cohort
Primary tumor
Progression & validation cohort
Primary tumor
p−value=0.72
p−value=0.0027
p−value=0.34
p−value=0.11
p−value=0.58
p−value=0.011
p−value=0.89
p−value=0.034
Progression−free survival probability
Overall survival probability
Months
MGMT methylated
MGMT unmethylated

Supplementary Figure 2

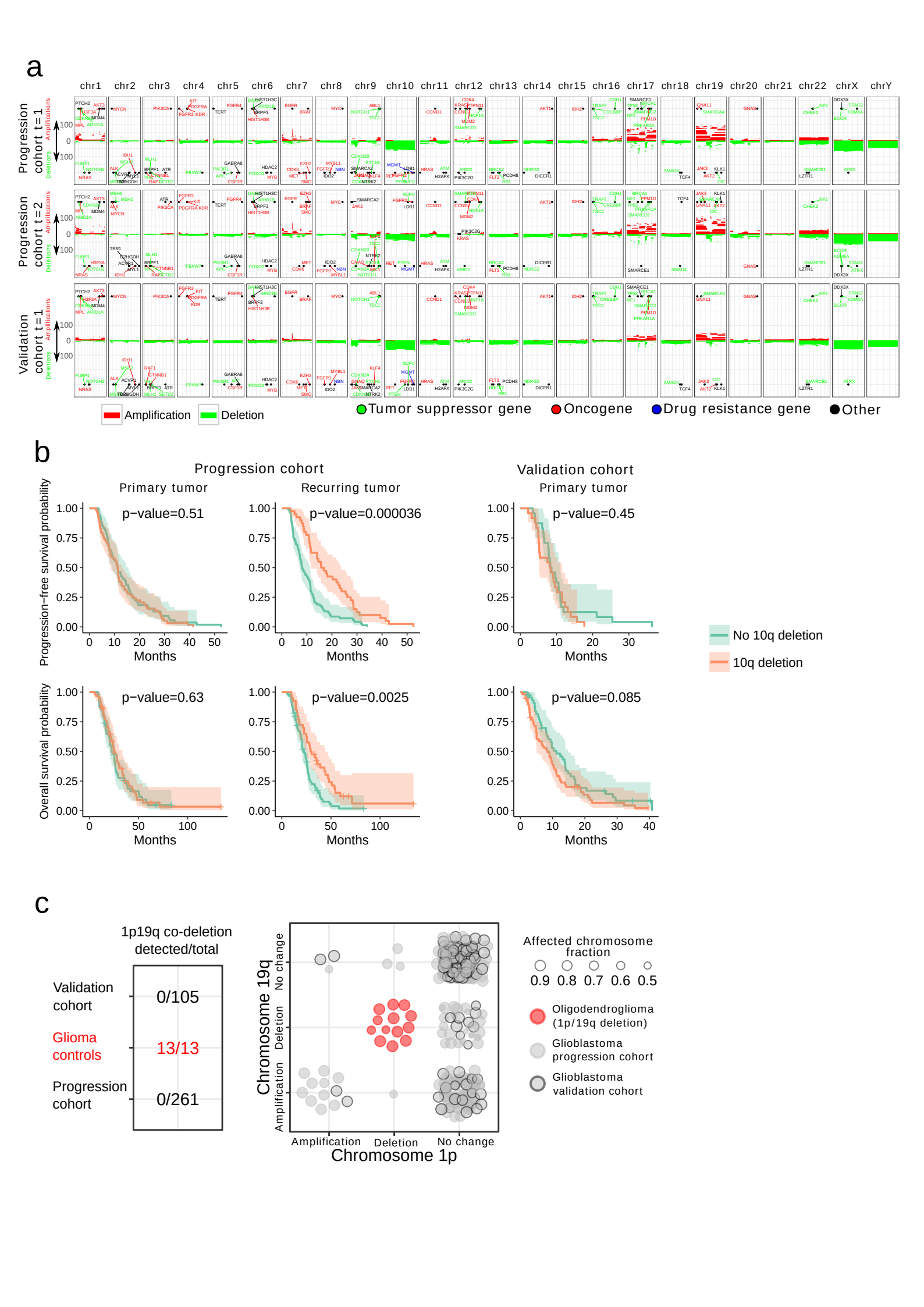
DNA methylation heterogeneity at selected genomic loci

(a) DNA methylation profiles for primary and recurring tumors at one glioblastoma-linked gene locus (*TERT*) and two enhancer loci. Gene annotations and epigenome segmentations for astrocytes were obtained from the WashU EpiGenome Browser (session URL: http://epigenome-segmentation.computational-epigenetics.org).

(b) Zoomed-in DNA methylation profiles for promoter regions depicted in Fig. 1c and Supplementary Fig. 2a.

(c) Mean *MGMT* promotor methylation levels averaged across two CpGs (cg12434587 and cg12981137) that have been described as highly informative for identifying methylated *MGMT* promoters (Bady, P., Delorenzi, M. & Hegi, M.E. Sensitivity Analysis of the MGMT-STP27 Model and Impact of Genetic and Epigenetic Context to Predict the MGMT Methylation Status in Gliomas and Other Tumors. J Mol Diagn 18, 350-361 (2016)). Error bars indicate the maximum and minimum detected DNA methylation levels in each sample. The dashed line indicates the threshold of 36% (derived from Bady et al.), below which samples are considered unmethylated.

(d) Kaplan-Meier plots showing progression-free and overall survival stratified by *MGMT* promoter methylation status as depicted in panel c. For the validation cohort, recorded progression-free survival time was available only for a subset of patients, while it was estimated to be two months less than the overall survival time for those patients with missing data. In the joint analysis of progression and validation cohort, only CpGs covered by more than five reads were considered, and a mean methylation value above 50% was required to consider the *MGMT* promoter methylated. The number of patients (N) is provided in Supplementary Table 3.


a
chr1 chr2 chr3 chr4 chr5 chr6 chr7 chr8 chr9 chr10 chr11 chr12 chr13 chr14 chr15 chr16 chr17 chr18 chr19 chr20 chr21 chr22 chrX chrY
Progression cohort t=1
Progression cohort t=2
Validation cohort t=1
Amplifications
Deletions
Amplification
Deletion
Tumor suppressor gene
Oncogene
Drug resistance gene
Other
b
Progression cohort
Validation cohort
Primary tumor
Recurring tumor
Primary tumor
Progression−free survival probability
Overall survival probability
p−value=0.51
p−value=0.000036
p−value=0.45
p−value=0.63
p−value=0.0025
p−value=0.085
Months
No 10q deletion
10q deletion
c
1p19q co-deletion detected/total
Validation cohort
Glioma controls
Progression cohort
0/105
13/13
0/261
Chromosome 19q
Amplification
Deletion
No change
Chromosome 1p
Affected chromosome fraction
0.9 0.8 0.7 0.6 0.5
Oligodendroglioma (1p/19q deletion)
Glioblastoma progression cohort
Glioblastoma validation cohort

Supplementary Figure 3

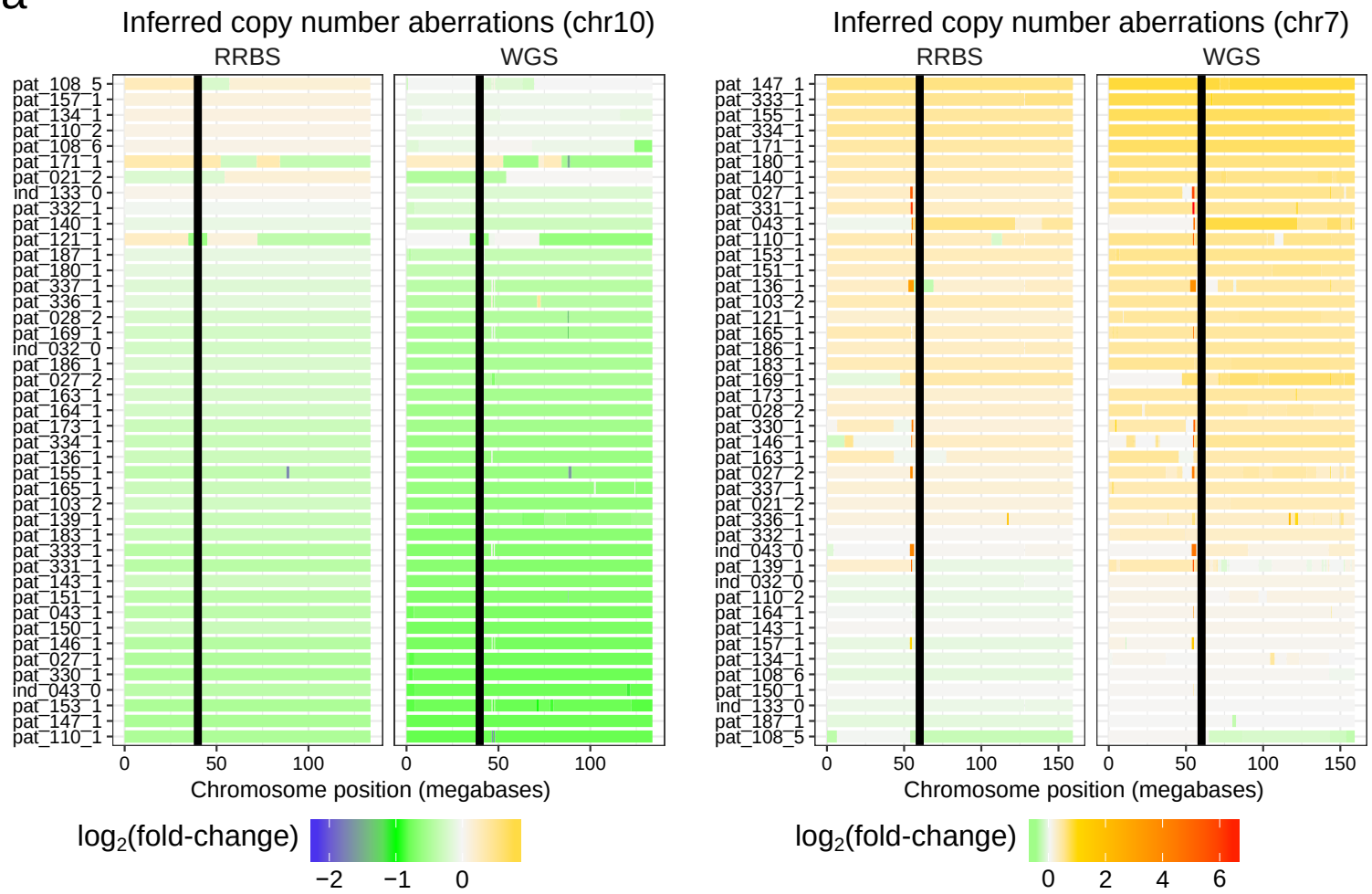
Inference of copy number aberrations from RRBS data

(a) Overview of the RRBS-based inference of copy number aberrations. Horizontal lines denote inferred amplifications (red) and deletions (green) in individual patients. The height of each stack reflects the total number of aberrations at the respective chromosomal position. Genes with reported relevance in glioblastoma are shown in colors that indicate their putative role as tumor suppressor genes, oncogenes, or drug resistance genes.

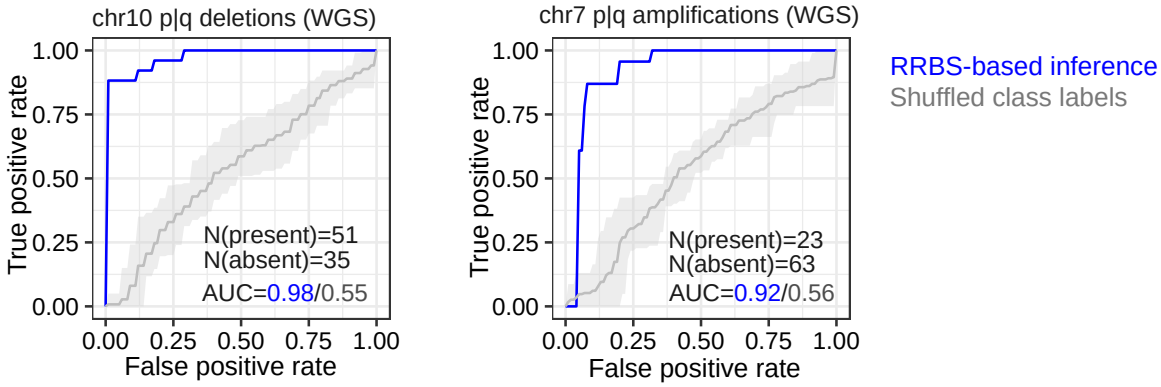
(b) Kaplan-Meier plots showing progression-free and overall survival probabilities stratified by chromosome 10q deletion status inferred from the RRBS data. The number of patients (N) is provided in Supplementary Table 3.

(c) Assessment of 1p/19q co-deletion status in the *IDH*-wildtype primary glioblastoma samples and, for comparison, in oligodendroglioma samples with known 1p/19q co-deletion. The size of the bubbles represents the mean fraction of the respective chromosome arms that are affected by the indicated copy number aberrations.

a


Inferred copy number aberrations (chr10)
RRBS
WGS
pat_108_5
pat_157_1
pat_134_1
pat_110_2
pat_108_6
pat_171_1
pat_021_2
ind_133_0
pat_332_1
pat_140_1
pat_121_1
pat_187_1
pat_180_1
pat_337_1
pat_336_1
pat_028_2
pat_169_1
ind_032_0
pat_186_1
pat_027_2
pat_163_1
pat_164_1
pat_173_1
pat_334_1
pat_136_1
pat_155_1
pat_165_1
pat_103_2
pat_139_1
pat_183_1
pat_333_1
pat_331_1
pat_143_1
pat_151_1
pat_043_1
pat_150_1
pat_146_1
pat_027_1
pat_330_1
ind_043_0
pat_153_1
pat_147_1
pat_110_1
0
50
100
Chromosome position (megabases)
log2(fold-change)
−2
−1
0
Inferred copy number aberrations (chr7)
RRBS
WGS
pat_147_1
pat_333_1
pat_155_1
pat_334_1
pat_171_1
pat_180_1
pat_140_1
pat_027_1
pat_331_1
pat_043_1
pat_110_1
pat_153_1
pat_151_1
pat_136_1
pat_103_2
pat_121_1
pat_165_1
pat_186_1
pat_183_1
pat_169_1
pat_173_1
pat_028_2
pat_330_1
pat_146_1
pat_163_1
pat_027_2
pat_337_1
pat_021_2
pat_336_1
pat_332_1
ind_043_0
pat_139_1
ind_032_0
pat_110_2
pat_164_1
pat_143_1
pat_157_1
pat_134_1
pat_108_6
pat_150_1
ind_133_0
pat_187_1
pat_108_5
0
50
100
150
Chromosome position (megabases)
log2(fold-change)
0
2
4
6

b

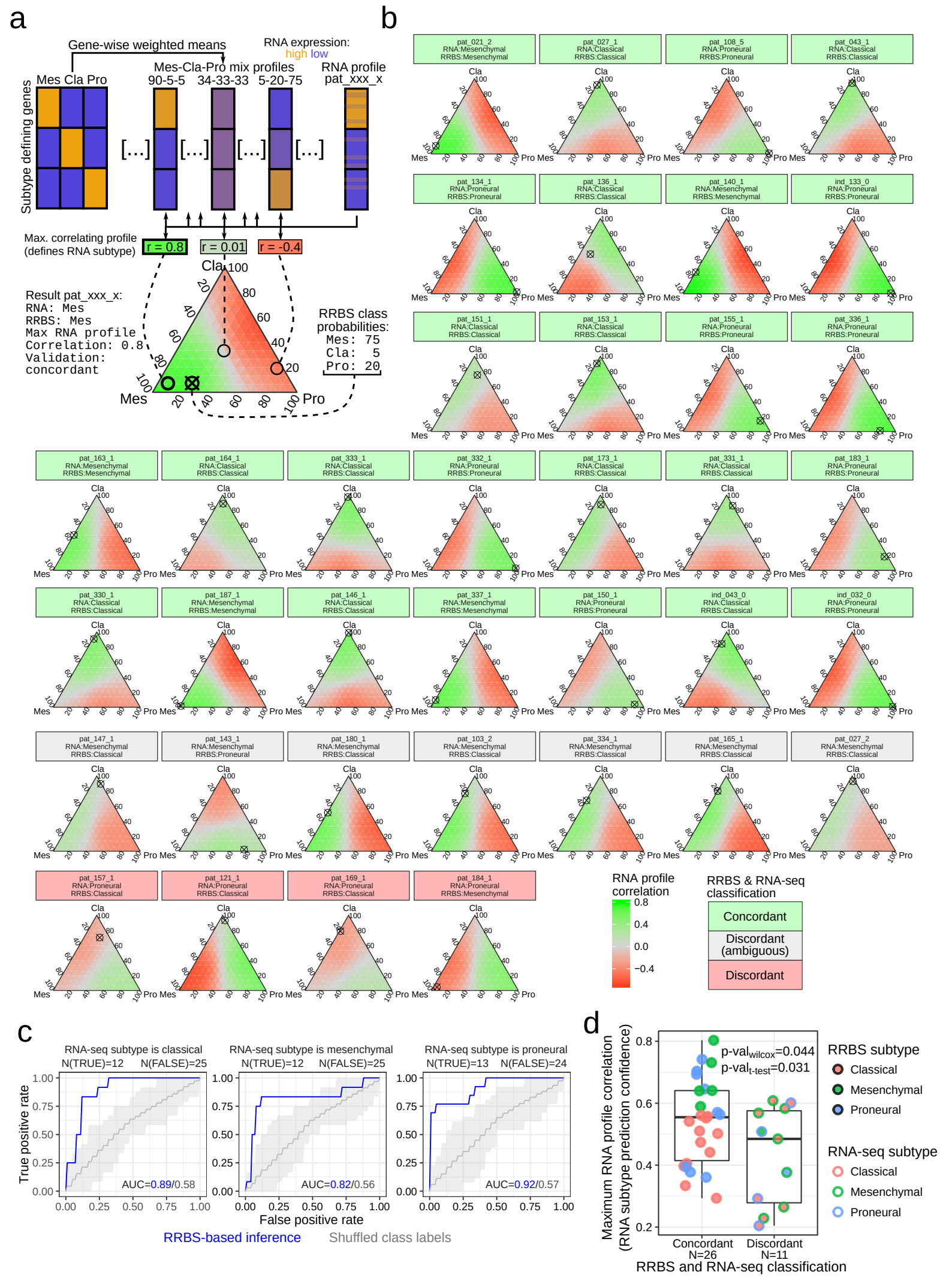

chr10 p|q deletions (WGS)
True positive rate
False positive rate
0.00
0.25
0.50
0.75
1.00
N(present)=51
N(absent)=35
AUC=0.98/0.55
chr7 p|q amplifications (WGS)
True positive rate
False positive rate
0.00
0.25
0.50
0.75
1.00
N(present)=23
N(absent)=63
AUC=0.92/0.56
RRBS-based inference
Shuffled class labels

Supplementary Figure 4

Validation of copy number aberrations inferred from RRBS data

(a) Heatmap depicting copy number aberrations on chromosome 10 and chromosome 7 inferred from RRBS data (left) and from low-coverage whole genome sequencing data (right) in 43 glioblastoma samples. The centromere is marked by a vertical black line.

(b) ROC curves assessing the performance of RRBS-based inference of chromosome 10 deletions (left) and chromosome 7 amplifications (right), evaluated against the low-coverage whole genome sequencing data (absolute log2 fold change of 0.5). The number of samples with (N present) and without (N absent) the respective copy number aberration are shown.


a
Gene-wise weighted means
RNA expression:
high low
Mes-Cla-Pro mix profiles
RNA profile
pat_xxx_x
Mes Cla Pro
90-5-5
34-33-33
5-20-75
Subtype defining genes
Max. correlating profile
(defines RNA subtype)
r = 0.8
r = 0.01
r = -0.4
Result pat_xxx_x:
RNA: Mes
RRBS: Mes
Max RNA profile
Correlation: 0.8
Validation:
concordant
RRBS class
probabilities:
Mes: 75
Cla: 5
Pro: 20
b
RNA profile
correlation
0.8
0.4
0.0
−0.4
RRBS & RNA-seq
classification
Concordant
Discordant
(ambiguous)
Discordant
c
RNA-seq subtype is classical
N(TRUE)=12 N(FALSE)=25
RNA-seq subtype is mesenchymal
N(TRUE)=12 N(FALSE)=25
RNA-seq subtype is proneural
N(TRUE)=13 N(FALSE)=24
True positive rate
False positive rate
AUC=0.89/0.58
AUC=0.82/0.56
AUC=0.92/0.57
RRBS-based inference
Shuffled class labels
d
Maximum RNA profile correlation
(RNA subtype prediction confidence)
p-val_wilcox=0.044
p-val_t-test=0.031
Concordant
N=26
Discordant
N=11
RRBS and RNA-seq classification
RRBS subtype
Classical
Mesenchymal
Proneural
RNA-seq subtype
Classical
Mesenchymal
Proneural

Supplementary Figure 5

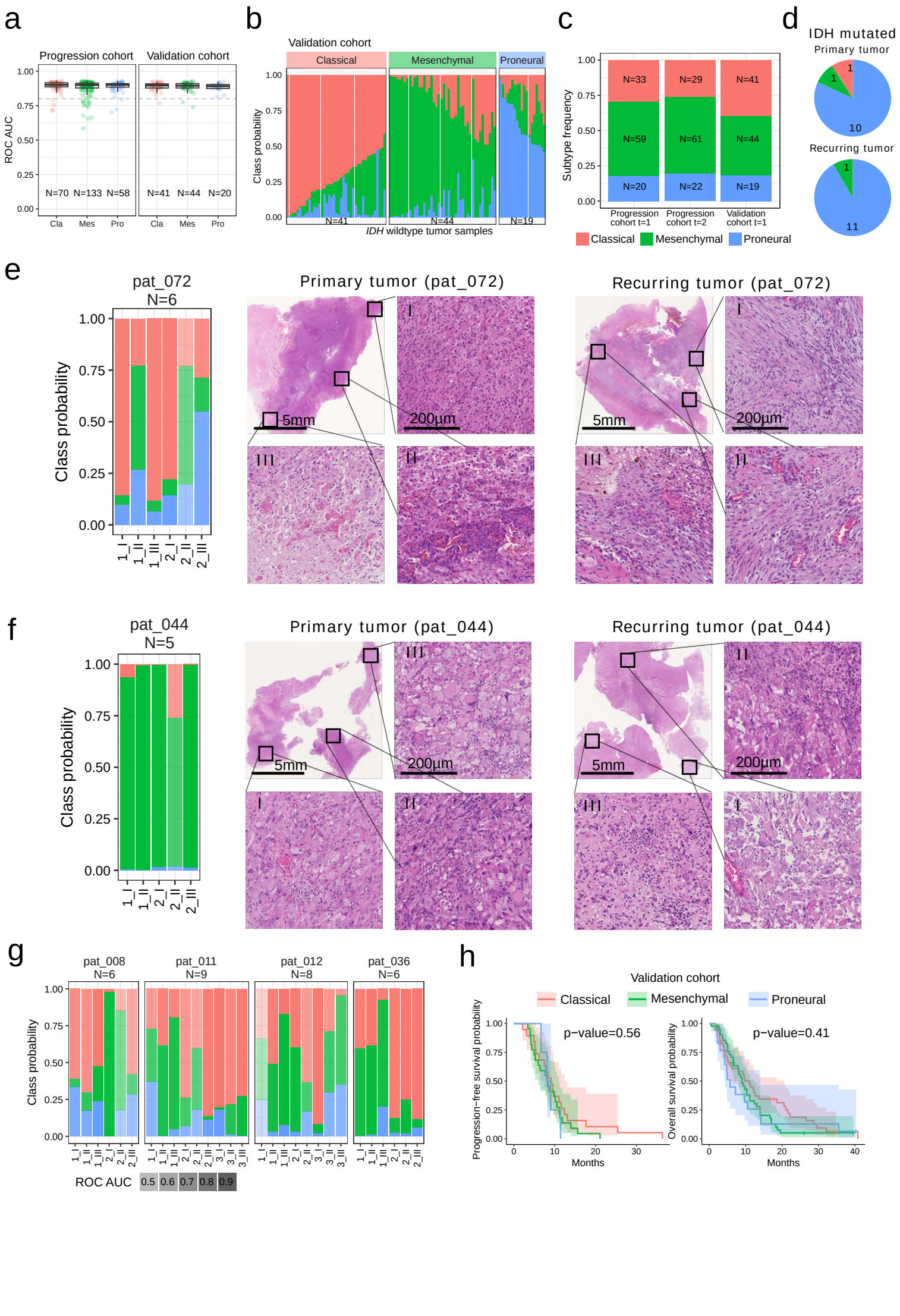
Validation of the RRBS-based transcriptional subtype prediction

(a) Illustration of the approach for validating the RRBS-based prediction of glioblastoma transcriptional subtypes (Mes: Mesenchymal, Cla: Classical, Pro: Proneural) using matched RNA-seq data. Correlation coefficients (green: positive correlation, grey: no correlation, red: negative correlation) with pure and mix profiles based on the RNA-seq data are shown as triangular diagrams. The position of the RRBS-based transcriptional subtype prediction is marked according to the class probabilities obtained from the machine learning classifier.

(b) RNA-seq based and RRBS-based transcriptional subtype predictions for each of the 37 validation samples, plotted as described in panel a. The color of the header bar indicates the validation status for each sample.

(c) ROC curves for RRBS-based prediction of RNA-seq based transcriptional subtype assignments. The number of validation samples assigned to the same subtypes (N[TRUE]) and those that are assigned to a different subtype (N[FALSE]) is shown.

(d) Boxplots showing lower confidence of RNA-seq based subtype assignments for discordantly classified samples compared to concordantly classified samples. $p\text{-val}_{wilcox}$: P-value based on one-sided Wilcoxon rank sum test; $p\text{-val}_{t\text{-test}}$: P-value based on one-sided t-test. N denotes the number of tumor samples.


a
Progression cohort
Validation cohort
ROC AUC
N=70 N=133 N=58
N=41 N=44 N=20
Cla Mes Pro
b
Validation cohort
Classical
Mesenchymal
Proneural
Class probability
N=41
N=44
N=19
IDH wildtype tumor samples
c
Subtype frequency
N=33 N=29 N=41
N=59 N=61 N=44
N=20 N=22 N=19
Progression cohort t=1
Progression cohort t=2
Validation cohort t=1
Classical Mesenchymal Proneural
d
IDH mutated
Primary tumor
1 1 10
Recurring tumor
1 11
e
pat_072
N=6
Class probability
Primary tumor (pat_072)
Recurring tumor (pat_072)
5mm
200µm
I
II
III
f
pat_044
N=5
Class probability
Primary tumor (pat_044)
Recurring tumor (pat_044)
5mm
200µm
g
pat_008 N=6
pat_011 N=9
pat_012 N=8
pat_036 N=6
Class probability
ROC AUC 0.5 0.6 0.7 0.8 0.9
h
Validation cohort
Classical Mesenchymal Proneural
Progression−free survival probability
p−value=0.56
Months
Overall survival probability
p−value=0.41
Months

Supplementary Figure 6

Prediction of glioblastoma transcriptional subtypes from RRBS data

(a) Sample-wise ROC AUC values on the TCGA training data, providing a confidence measure for the RRBS-based transcriptional subtype predictions. Dashed line: ROC AUC = 0.8

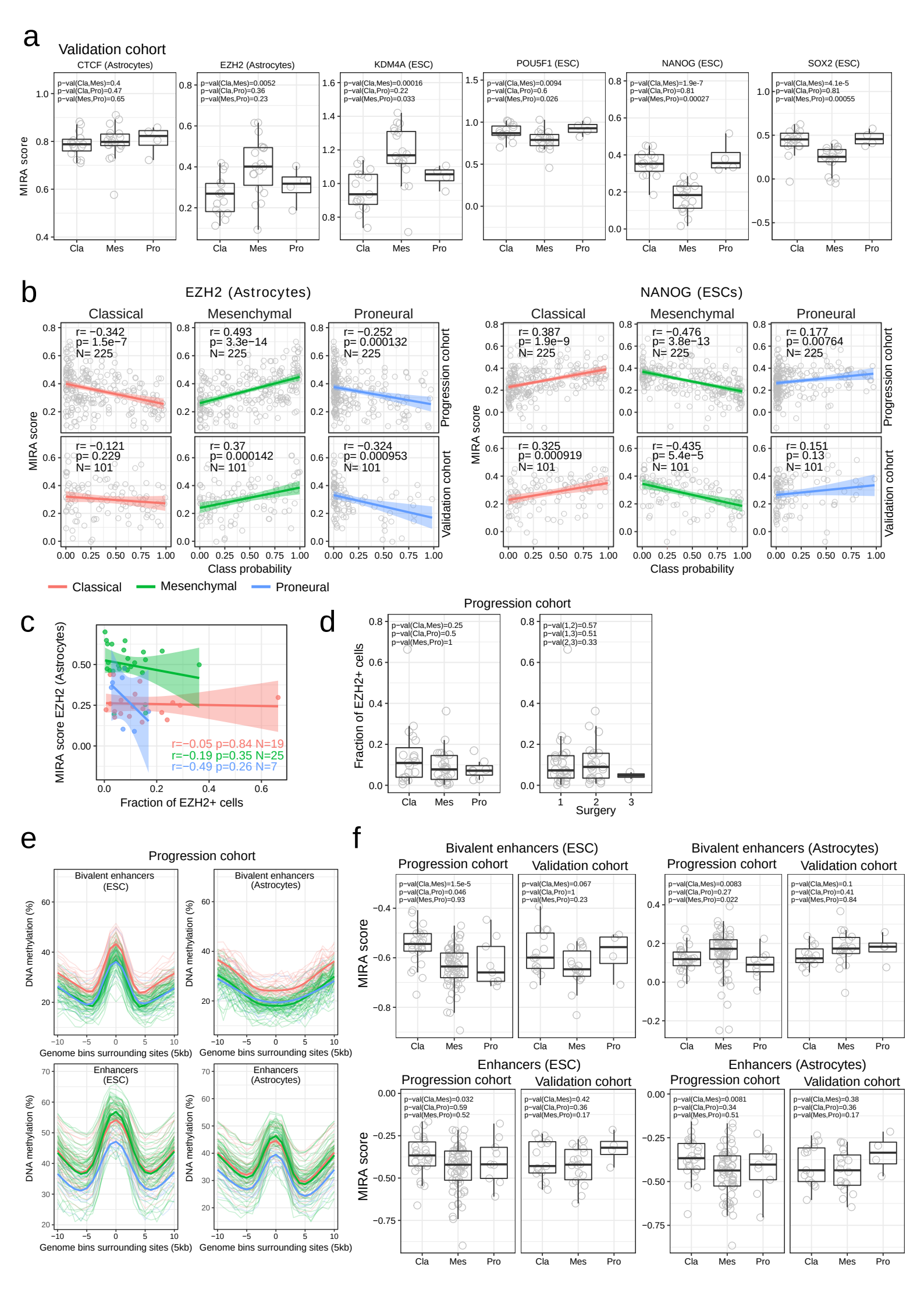
(b) Transcriptional subtype heterogeneity in the validation cohort samples, as indicated by class probabilities of the RRBS-based subtype classifier. Samples are grouped and ordered by their dominant subtype.

(c) Distribution of RRBS-based transcriptional subtype predictions for *IDH*-wildtype samples in the progression and validation cohort. N denotes the number of tumor samples.

(d) Distribution of RRBS-based transcriptional subtype predictions for *IDH*-mutated samples.

(e, f, g) Distribution of class probabilities across different regions of the same tumor (indicated by Roman numbers) and across different surgeries (indicated by Arabic numbers). For two of the six patients with multisector samples, matched H&E stains are shown for the tumor regions from which the multi-sector samples originated (section III of the primary tumor of patient 044 was excluded from subtype prediction due to low CpG coverage). N denotes the number of tumor samples displayed for each patient.

(h) Kaplan-Meier plots showing progression-free and overall survival probabilities for patients in the validation cohort stratified by RRBS-based transcriptional subtype predictions. Only patients for which the sample-specific subtype classifier achieved high accuracy (ROC AUC > 0.8) were included in the analyses. The number of patients (N) is provided in Supplementary Table 3.


a
Validation cohort
CTCF (Astrocytes)
p−val(Cla,Mes)=0.4
p−val(Cla,Pro)=0.47
p−val(Mes,Pro)=0.65
EZH2 (Astrocytes)
p−val(Cla,Mes)=0.0052
p−val(Cla,Pro)=0.36
p−val(Mes,Pro)=0.23
KDM4A (ESC)
p−val(Cla,Mes)=0.00016
p−val(Cla,Pro)=0.22
p−val(Mes,Pro)=0.033
POU5F1 (ESC)
p−val(Cla,Mes)=0.0094
p−val(Cla,Pro)=0.6
p−val(Mes,Pro)=0.026
NANOG (ESC)
p−val(Cla,Mes)=1.9e−7
p−val(Cla,Pro)=0.81
p−val(Mes,Pro)=0.00027
SOX2 (ESC)
p−val(Cla,Mes)=4.1e−5
p−val(Cla,Pro)=0.81
p−val(Mes,Pro)=0.00055
MIRA score
Cla
Mes
Pro
b
EZH2 (Astrocytes)
NANOG (ESCs)
Classical
Mesenchymal
Proneural
r= −0.342
p= 1.5e−7
N= 225
r= 0.493
p= 3.3e−14
N= 225
r= −0.252
p= 0.000132
N= 225
r= −0.121
p= 0.229
N= 101
r= 0.37
p= 0.000142
N= 101
r= −0.324
p= 0.000953
N= 101
r= 0.387
p= 1.9e−9
N= 225
r= −0.476
p= 3.8e−13
N= 225
r= 0.177
p= 0.00764
N= 225
r= 0.325
p= 0.000919
N= 101
r= −0.435
p= 5.4e−5
N= 101
r= 0.151
p= 0.13
N= 101
Progression cohort
Validation cohort
MIRA score
Class probability
Classical
Mesenchymal
Proneural
c
MIRA score EZH2 (Astrocytes)
Fraction of EZH2+ cells
r=−0.05 p=0.84 N=19
r=−0.19 p=0.35 N=25
r=−0.49 p=0.26 N=7
d
Progression cohort
p−val(Cla,Mes)=0.25
p−val(Cla,Pro)=0.5
p−val(Mes,Pro)=1
p−val(1,2)=0.57
p−val(1,3)=0.51
p−val(2,3)=0.33
Fraction of EZH2+ cells
Surgery
e
Progression cohort
Bivalent enhancers (ESC)
Bivalent enhancers (Astrocytes)
Enhancers (ESC)
Enhancers (Astrocytes)
DNA methylation (%)
Genome bins surrounding sites (5kb)
f
Bivalent enhancers (ESC)
Bivalent enhancers (Astrocytes)
Enhancers (ESC)
Enhancers (Astrocytes)
Progression cohort
Validation cohort
p−val(Cla,Mes)=1.5e-5
p−val(Cla,Pro)=0.046
p−val(Mes,Pro)=0.93
p−val(Cla,Mes)=0.067
p−val(Cla,Pro)=1
p−val(Mes,Pro)=0.23
p−val(Cla,Mes)=0.0083
p−val(Cla,Pro)=0.27
p−val(Mes,Pro)=0.022
p−val(Cla,Mes)=0.1
p−val(Cla,Pro)=0.41
p−val(Mes,Pro)=0.84
p−val(Cla,Mes)=0.032
p−val(Cla,Pro)=0.59
p−val(Mes,Pro)=0.52
p−val(Cla,Mes)=0.42
p−val(Cla,Pro)=0.36
p−val(Mes,Pro)=0.17
p−val(Cla,Mes)=0.0081
p−val(Cla,Pro)=0.34
p−val(Mes,Pro)=0.51
p−val(Cla,Mes)=0.38
p−val(Cla,Pro)=0.36
p−val(Mes,Pro)=0.17
MIRA score
Cla
Mes
Pro

Supplementary Figure 7

Association between transcriptional subtypes and (epi)genome regulation

(a) MIRA scores in the validation cohort for three sets of transcription factor binding sites that were hypomethylated in the mesenchymal subtype of the progression cohort (CTCF, EZH2, KDM4A) and for the binding sites of three key regulators of pluripotency (POU5F1(OCT4), NANOG, SOX2).

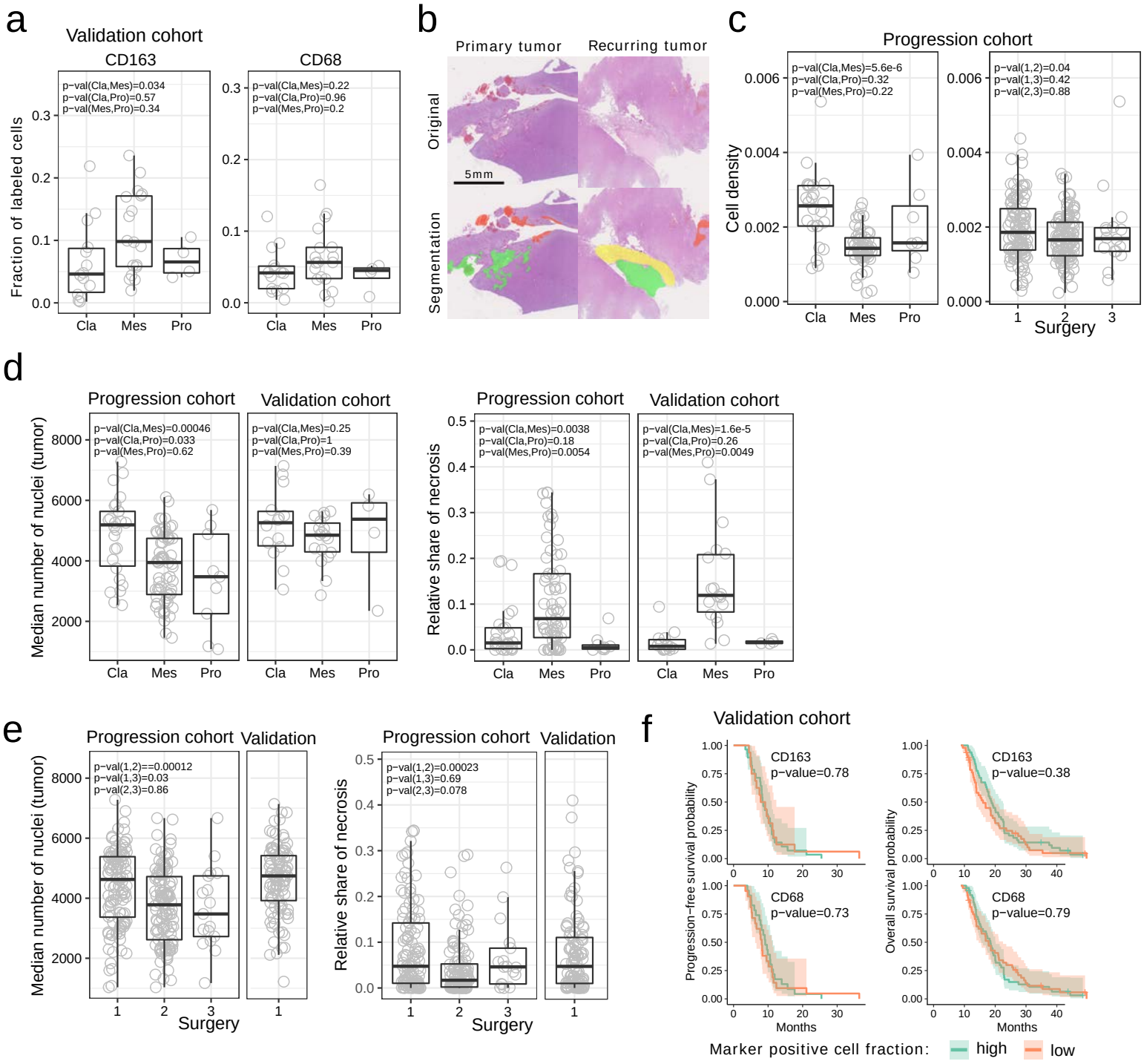
(b) Scatterplots displaying the relationship between MIRA scores for the indicated transcription factors (EZH2, NANOG) and class probabilities for the indicated transcriptional subtypes. r: Pearson correlation, p: p-value.

(c) Scatterplot displaying the relationship between MIRA scores for EZH2 in astrocytes and the fraction of EZH2 positive cells as determined by quantitative immunohistochemistry. r: Pearson correlation, p: p-value.

(d) Fraction of EZH2 positive cells as measured by quantitative immunohistochemistry across transcriptional subtypes (left) and across tumor resections (right).

(e, f) DNA methylation profiles (e) and corresponding MIRA scores (f) for four sets of genomic regulatory regions (enhancers and bivalent enhancers in embryonic stem cells as well as astrocytes, as defined by the Roadmap Epigenomics Project[43]), compared across the three transcriptional subtypes.

Only samples for which the sample-specific subtype classifier achieved high accuracy (ROC AUC > 0.8) were included in the analysis for panels b, e, and f. In the scatterplots (panels b, c), N denotes the number of tumor samples in each analysis. The number of tumor samples (N) in the association tests (panels a, d, f) is provided in Supplementary Table 4.


a
Validation cohort
CD163
CD68
p−val(Cla,Mes)=0.034
p−val(Cla,Pro)=0.57
p−val(Mes,Pro)=0.34
p−val(Cla,Mes)=0.22
p−val(Cla,Pro)=0.96
p−val(Mes,Pro)=0.2
Fraction of labeled cells
Cla
Mes
Pro
b
Primary tumor
Recurring tumor
Original
Segmentation
5mm
c
Progression cohort
p−val(Cla,Mes)=5.6e-6
p−val(Cla,Pro)=0.32
p−val(Mes,Pro)=0.22
p−val(1,2)=0.04
p−val(1,3)=0.42
p−val(2,3)=0.88
Cell density
Surgery
d
Progression cohort
Validation cohort
p−val(Cla,Mes)=0.00046
p−val(Cla,Pro)=0.033
p−val(Mes,Pro)=0.62
p−val(Cla,Mes)=0.25
p−val(Cla,Pro)=1
p−val(Mes,Pro)=0.39
Median number of nuclei (tumor)
p−val(Cla,Mes)=0.0038
p−val(Cla,Pro)=0.18
p−val(Mes,Pro)=0.0054
p−val(Cla,Mes)=1.6e-5
p−val(Cla,Pro)=0.26
p−val(Mes,Pro)=0.0049
Relative share of necrosis
e
Progression cohort
Validation
p−val(1,2)==0.00012
p−val(1,3)=0.03
p−val(2,3)=0.86
p−val(1,2)=0.00023
p−val(1,3)=0.69
p−val(2,3)=0.078
Surgery
f
Validation cohort
CD163
p−value=0.78
p−value=0.38
CD68
p−value=0.73
p−value=0.79
Progression−free survival probability
Overall survival probability
Months
Marker positive cell fraction:
high
low

Supplementary Figure 8

Association of histopathological tumor properties with transcriptional subtypes and disease progression

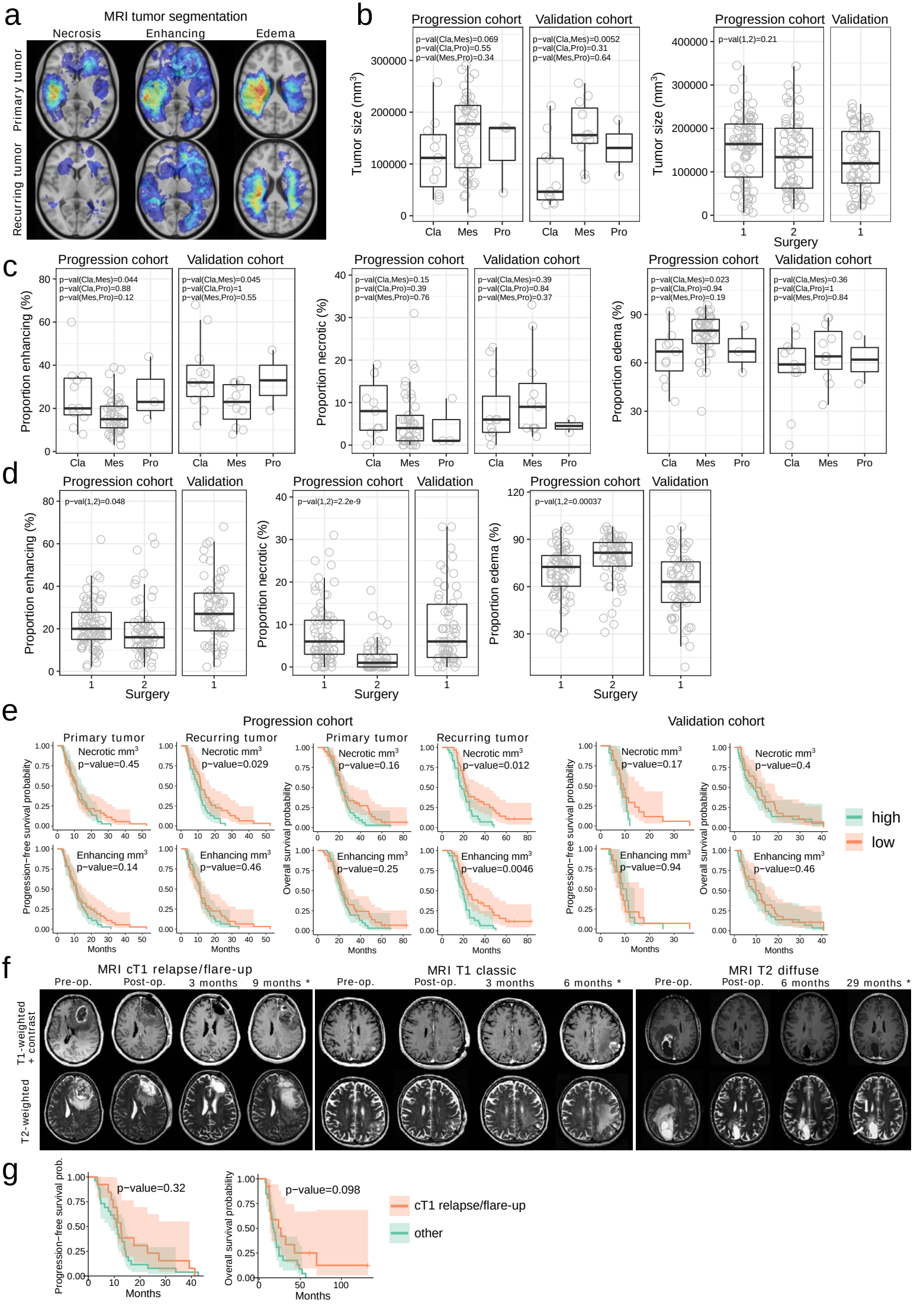
(a) Level of tumor-infiltrating immune cells (CD163, CD68) as measured by quantitative immunohistochemistry in the validation cohort, compared across transcriptional subtypes.

(b) Segmented H&E stains illustrating the quantification of histopathological tumor properties (red: hemorrhage, green: necrosis, yellow: meningeal scarring).

(c, d, e) Histopathological tumor properties compared across transcriptional subtypes and across tumor resections (1: primary tumor, 2: recurring tumor, 3: second recurrence).

(f) Kaplan-Meier plots showing progression-free and overall survival for patients in the validation cohort stratified by the level of CD163 positive and CD68 positive immune cell infiltration. The number of patients (N) is provided in Supplementary Table 3.

The number of tumor samples (N) for the association tests (panels a, c, d, e) is provided in Supplementary Table 4.


a
MRI tumor segmentation
Necrosis
Enhancing
Edema
Primary tumor
Recurring tumor
b
Progression cohort
Validation cohort
p−val(Cla,Mes)=0.069
p−val(Cla,Pro)=0.55
p−val(Mes,Pro)=0.34
p−val(Cla,Mes)=0.0052
p−val(Cla,Pro)=0.31
p−val(Mes,Pro)=0.64
Tumor size (mm3)
Cla
Mes
Pro
p−val(1,2)=0.21
Surgery
c
Proportion enhancing (%)
p−val(Cla,Mes)=0.044
p−val(Cla,Pro)=0.88
p−val(Mes,Pro)=0.12
p−val(Cla,Mes)=0.045
p−val(Cla,Pro)=1
p−val(Mes,Pro)=0.55
Proportion necrotic (%)
p−val(Cla,Mes)=0.15
p−val(Cla,Pro)=0.39
p−val(Mes,Pro)=0.76
p−val(Cla,Mes)=0.39
p−val(Cla,Pro)=0.84
p−val(Mes,Pro)=0.37
Proportion edema (%)
p−val(Cla,Mes)=0.023
p−val(Cla,Pro)=0.94
p−val(Mes,Pro)=0.19
p−val(Cla,Mes)=0.36
p−val(Cla,Pro)=1
p−val(Mes,Pro)=0.84
d
p−val(1,2)=0.048
p−val(1,2)=2.2e-9
p−val(1,2=0.00037
Surgery
e
Progression cohort
Validation cohort
Primary tumor
Recurring tumor
Necrotic mm3
p−value=0.45
p−value=0.029
p−value=0.16
p−value=0.012
p−value=0.17
p−value=0.4
Enhancing mm3
p−value=0.14
p−value=0.46
p−value=0.25
p−value=0.0046
p−value=0.94
p−value=0.46
Progression−free survival probability
Overall survival probability
Months
high
low
f
MRI cT1 relapse/flare-up
MRI T1 classic
MRI T2 diffuse
Pre-op.
Post-op.
3 months
9 months *
6 months *
6 months
29 months *
T1-weighted + contrast
T2-weighted
g
p−value=0.32
p−value=0.098
Progression−free survival prob.
Overall survival probability
Months
cT1 relapse/flare-up
other

Supplementary Figure 9

Association of MRI-derived tumor properties with transcriptional subtypes and disease progression

(a) Segmented MRI illustrating the quantification of MRI-derived tumor properties. Heatmap intensity overlays indicate the extent of necrosis, contrast-enhancing (active) tumor volume, and edema – averaged across the progression cohort and normalized to a standard brain.

(b, c, d) Comparison of MRI-derived tumor properties across transcriptional subtypes and across tumor resections (1: primary tumor, 2: recurring tumor, 3: second recurrence).

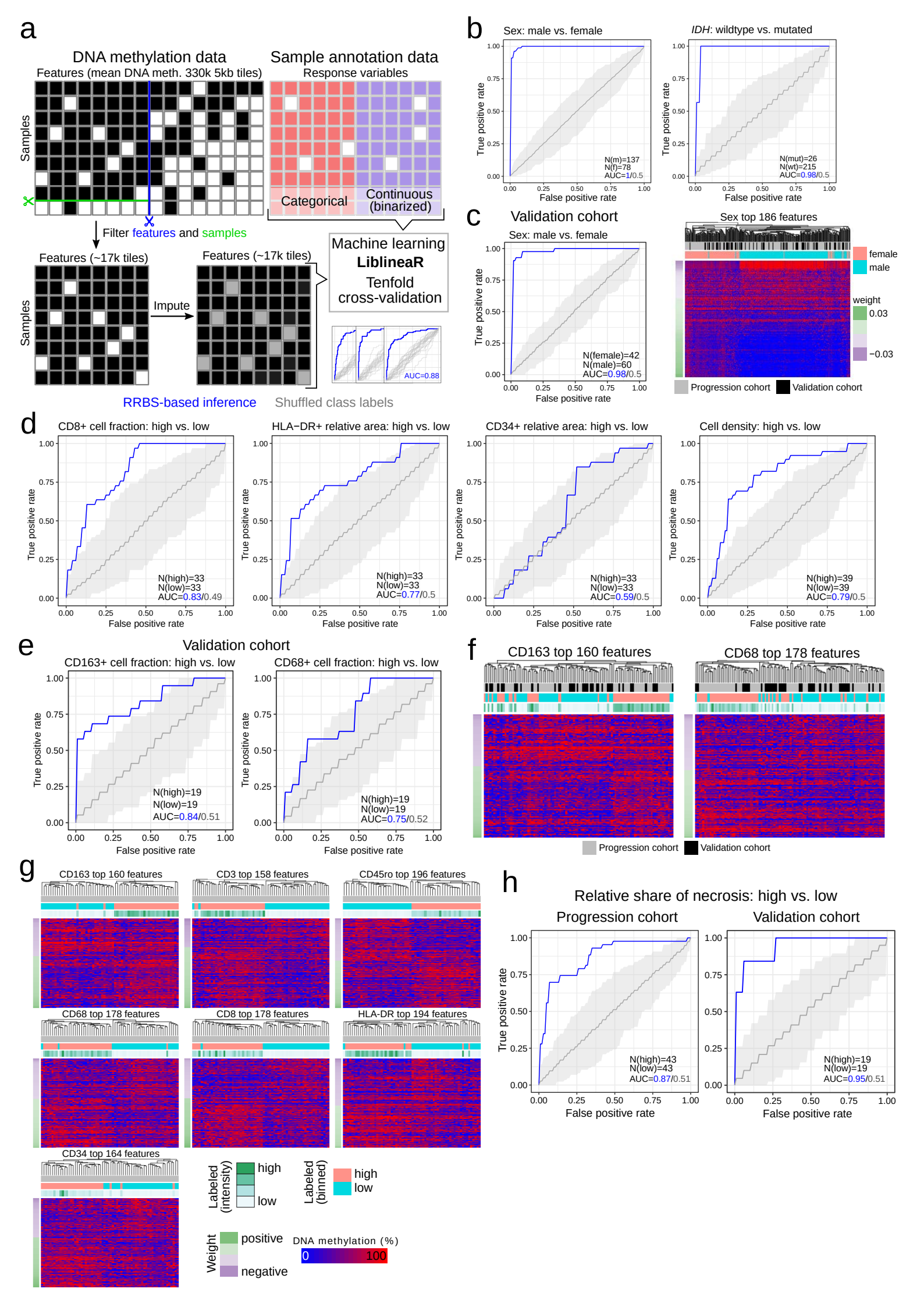
(e) Kaplan-Meier plots showing progression-free and overall survival stratified by the level of necrotic and contrast-enhancing (active) tumor volume based on MRI, respectively.

(f) Time series of T1-contrast enhanced and T2/FLAIR MRI sequences, illustrating the three MRI progression types (cT1 relapse/flare-up, classic T1, T2 diffuse). '*': tumor recurrence.

(g) Kaplan-Meier plots showing progression-free and overall survival probabilities for patients in the progression cohort stratified by their MRI-derived progression types.

The number of tumor samples (N) for the association tests (panels b, c, d) is provided in Supplementary Table 4.

The number of patients (N) in the Kaplan-Meier analysis (panels e, g) is provided in Supplementary Table 3.


a
DNA methylation data
Features (mean DNA meth. 330k 5kb tiles)
Samples
Sample annotation data
Response variables
Categorical
Continuous (binarized)
Filter features and samples
Features (~17k tiles)
Impute
Features (~17k tiles)
Samples
Machine learning
LiblineaR
Tenfold cross-validation
AUC=0.88
RRBS-based inference
Shuffled class labels
b
Sex: male vs. female
True positive rate
False positive rate
N(m)=137
N(f)=78
AUC=1/0.5
IDH: wildtype vs. mutated
N(mut)=26
N(wt)=215
AUC=0.98/0.5
c
Validation cohort
Sex: male vs. female
N(female)=42
N(male)=60
AUC=0.98/0.5
Sex top 186 features
female
male
weight
0.03
−0.03
Progression cohort
Validation cohort
d
CD8+ cell fraction: high vs. low
N(high)=33
N(low)=33
AUC=0.83/0.49
HLA−DR+ relative area: high vs. low
N(high)=33
N(low)=33
AUC=0.77/0.5
CD34+ relative area: high vs. low
N(high)=33
N(low)=33
AUC=0.59/0.5
Cell density: high vs. low
N(high)=39
N(low)=39
AUC=0.79/0.5
e
Validation cohort
CD163+ cell fraction: high vs. low
N(high)=19
N(low)=19
AUC=0.84/0.51
CD68+ cell fraction: high vs. low
N(high)=19
N(low)=19
AUC=0.75/0.52
f
CD163 top 160 features
CD68 top 178 features
Progression cohort
Validation cohort
g
CD163 top 160 features
CD3 top 158 features
CD45ro top 196 features
CD68 top 178 features
CD8 top 178 features
HLA-DR top 194 features
CD34 top 164 features
Labeled (intensity)
high
low
Labeled (binned)
high
low
Weight
positive
negative
DNA methylation (%)
0
100
h
Relative share of necrosis: high vs. low
Progression cohort
N(high)=43
N(low)=43
AUC=0.87/0.51
Validation cohort
N(high)=19
N(low)=19
AUC=0.95/0.51

Supplementary Figure 10

DNA methylation based prediction of the tumor microenvironment

(a) Schematic illustrating the machine learning approach used to assess the predictability of various tumor properties based on the RRBS data. Black squares indicate DNA methylation values; white squares indicate missing values; grey squares indicate imputed values; and colored squares indicate annotation values.

(b) ROC curves for RRBS-based predictions of two features that are known to be predictable based on DNA methylation data (patient sex, *IDH* mutation status).

(c) ROC curve and clustered heatmap for DNA methylation based prediction of patient sex trained in the progression cohort (panel b) and applied without re-training in the validation cohort.

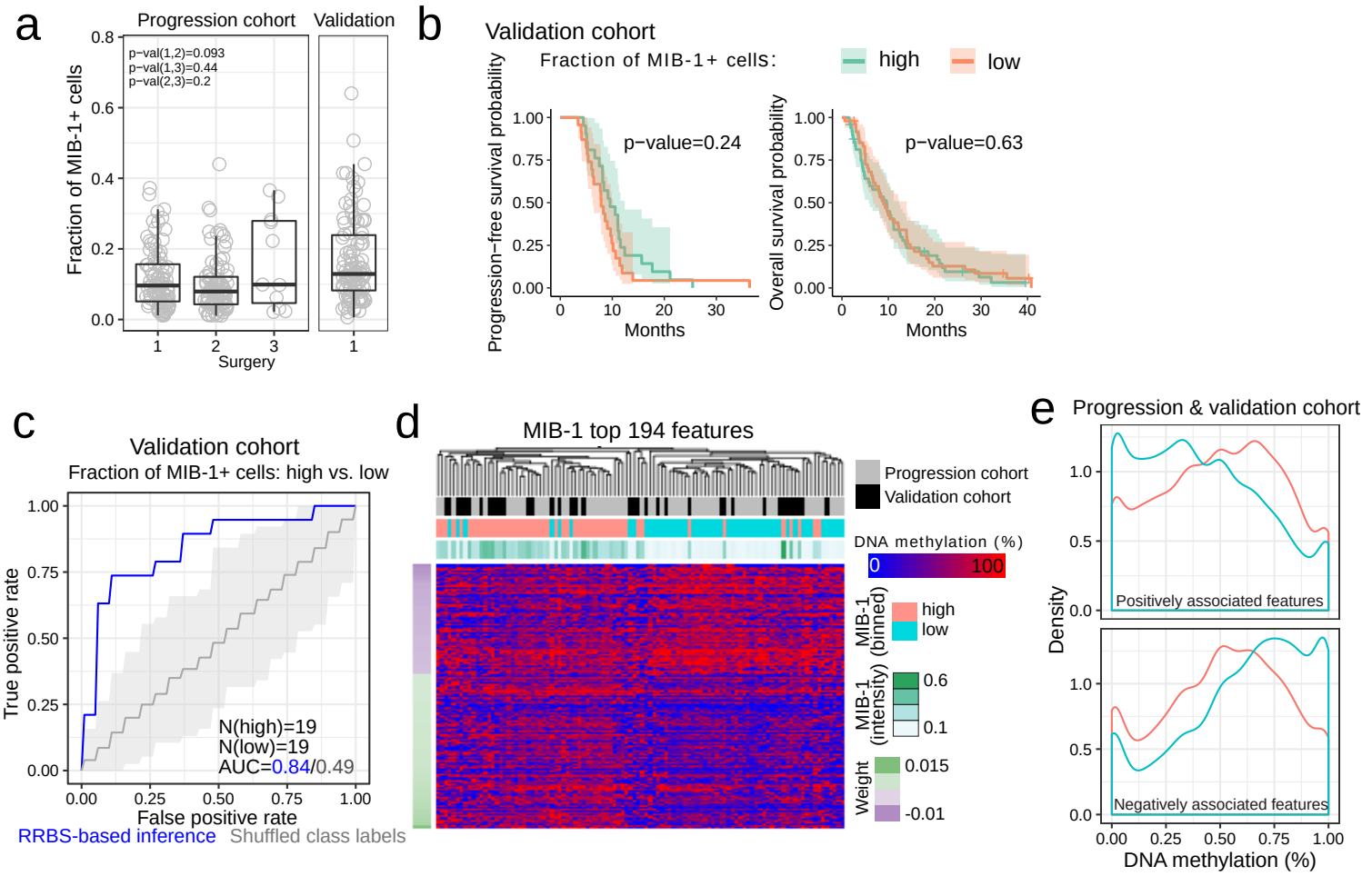
(d) ROC curves for the DNA methylation based prediction of several histopathologic tumor properties.

(e, f) ROC curves (e) and clustered heatmaps (f) for DNA methylation based prediction of immune cell infiltration levels (CD163, CD68) trained in the progression cohort and applied without re-training in the validation cohort.

(g) Clustered heatmap for the column-scaled DNA methylation levels of the most predictive genomic regions (5-kilobase tiling regions) from the classifiers predicting immune cell infiltration levels (CD163, CD68, CD45ro, CD3, CD8) and tumor cell characteristics (CD34, HLA-DR).

(h) ROC curves for RRBS-based prediction of the abundance of necrotic tissue in the progression cohort (left) and the validation cohort (right). For validation purposes, the classifier trained in the progression cohort was applied and evaluated in the validation cohort “as is”, without any additional training.

For the ROC curves, N denotes the number of tumor samples in each group.


a
Progression cohort
Validation
Fraction of MIB-1+ cells
p−val(1,2)=0.093
p−val(1,3)=0.44
p−val(2,3)=0.2
Surgery
b
Validation cohort
Fraction of MIB-1+ cells:
high
low
Progression−free survival probability
p−value=0.24
Overall survival probability
p−value=0.63
Months
c
Validation cohort
Fraction of MIB-1+ cells: high vs. low
True positive rate
False positive rate
N(high)=19
N(low)=19
AUC=0.84/0.49
RRBS-based inference
Shuffled class labels
d
MIB-1 top 194 features
Progression cohort
Validation cohort
DNA methylation (%)
MIB-1 (binned)
high
low
MIB-1 (intensity)
Weight
e
Progression & validation cohort
Density
Positively associated features
Negatively associated features
DNA methylation (%)
f
Progression cohort
Validation
Median AVG eccentricity (tumor)
p−val(1,2)=0.15
p−val(1,3)=0.21
p−val(2,3)=0.64
Surgery
Median COV eccentricity (tumor)
p−val(1,2)=0.13
p−val(1,3)=0.18
p−val(2,3)=0.71

Supplementary Figure 11

DNA methylation based prediction of tumor cell intrinsic properties

(a) Fraction of proliferating (MIB-1 positive) cells across tumor resections (1: primary tumor, 2: recurring tumor, 3: second recurrence) in the progression and validation cohort.

(b) Kaplan-Meier plots showing progression-free and overall survival for patients in the validation cohort stratified by the fraction of proliferating (MIB-1 positive) cells. The number of patients (N) is provided in Supplementary Table 3.

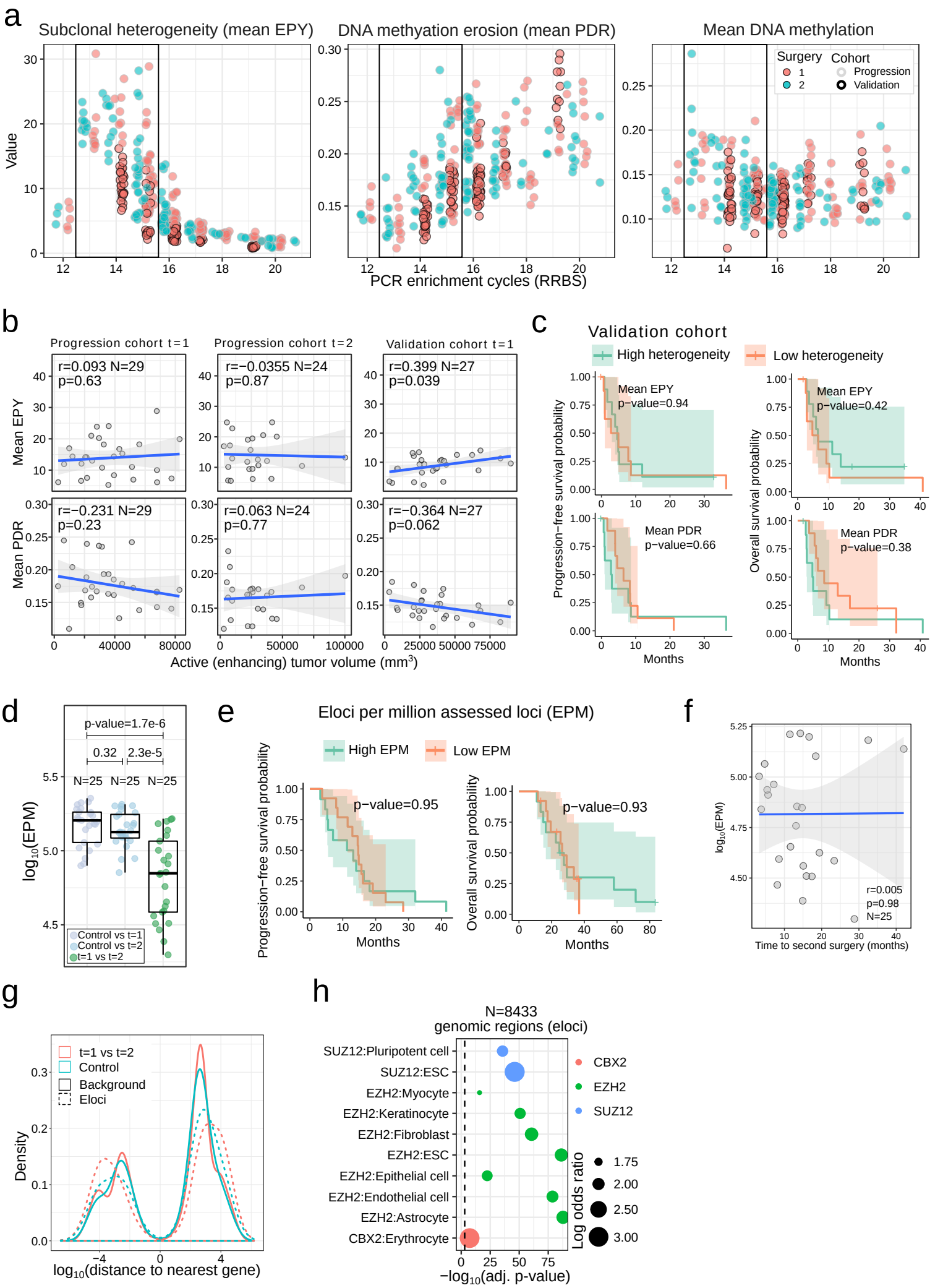
(c) ROC curves for the RRBS-based prediction of levels of proliferating (MIB-1 positive) cells trained in the progression cohort and applied without re-training in the validation cohort. N denotes the number of tumor samples in each group.

(d) Clustered heatmaps for the column-scaled DNA methylation levels of the most predictive genomic regions (5-kilobase tiling regions) from the classifier predicting the fraction of proliferating (MIB-1 positive) cells in the progression cohort as well as the validation cohort.

(e) Histogram showing the distribution of DNA methylation levels across the most predictive genomic regions from the classifier predicting the fraction of proliferating (MIB-1 positive) cells, displayed separately for samples with high (red) or low (blue) fractions of MIB-1 positive cells and for positively (top) or negatively (bottom) associated features as defined by the classifier-assigned weights.

(f) Average nuclear eccentricity (AVG) and its coefficient of variation (COV) across tumor resections (1: primary tumor, 2: recurring tumor, 3: second recurrence).

The number of tumor samples (N) for the association tests (panels a, f) is provided in Supplementary Table 4.


a
Subclonal heterogeneity (mean EPY)
DNA methyation erosion (mean PDR)
Mean DNA methylation
Value
PCR enrichment cycles (RRBS)
Surgery
1
2
Cohort
Progression
Validation
b
Progression cohort t=1
Progression cohort t=2
Validation cohort t=1
r=0.093 N=29
p=0.63
r=−0.0355 N=24
p=0.87
r=0.399 N=27
p=0.039
r=−0.231 N=29
p=0.23
r=0.063 N=24
p=0.77
r=−0.364 N=27
p=0.062
Mean EPY
Mean PDR
Active (enhancing) tumor volume (mm3)
c
Validation cohort
High heterogeneity
Low heterogeneity
Mean EPY
p−value=0.94
Mean EPY
p−value=0.42
Mean PDR
p−value=0.66
Mean PDR
p−value=0.38
Progression−free survival probability
Overall survival probability
Months
d
p-value=1.7e-6
0.32
2.3e-5
N=25 N=25 N=25
log10(EPM)
Control vs t=1
Control vs t=2
t=1 vs t=2
e
Eloci per million assessed loci (EPM)
High EPM
Low EPM
p−value=0.95
p−value=0.93
Progression−free survival probability
Overall survival probability
Months
f
log10(EPM)
r=0.005
p=0.98
N=25
Time to second surgery (months)
g
t=1 vs t=2
Control
Background
Eloci
Density
log10(distance to nearest gene)
h
N=8433
genomic regions (eloci)
SUZ12:Pluripotent cell
SUZ12:ESC
EZH2:Myocyte
EZH2:Keratinocyte
EZH2:Fibroblast
EZH2:ESC
EZH2:Epithelial cell
EZH2:Endothelial cell
EZH2:Astrocyte
CBX2:Erythrocyte
CBX2
EZH2
SUZ12
Log odds ratio
1.75
2.00
2.50
3.00
−log10(adj. p-value)

Supplementary Figure 12

Analysis of DNA methylation heterogeneity in primary and recurring tumors

(a) Scatterplots displaying the association between PCR enrichment cycles during RRBS library preparation and DNA methylation heterogeneity (as measured by PDR or EPY) as well as mean DNA methylation levels.

(b) Scatterplots displaying the association between DNA methylation heterogeneity (PDR, EPY) and tumor size as determined by MRI-based enhancing (active) tumor mass. r: Pearson correlation, p: p-value, t=1: primary tumor, t=2: recurring tumor.

(c) Kaplan-Meier plots showing progression-free and overall survival for patients in the validation cohort stratified by PDR and EPY scores, respectively. For the validation cohort, recorded progression-free survival time was available only for a subset of patients, while it was estimated to be two months less than the overall survival time for those patients with missing data.

(d) Epiallele shift between control samples (non-cancerous brain tissue), primary tumors (t=1), and recurring tumors (t=2), measured by the fraction of covered loci that show significant changes in epiallele composition between two samples (EPM: eloci per million assessed loci).

(e) Kaplan-Meier plots showing progression-free and overall survival probabilities stratified by the degree of epiallele shift (as measured by the EPM score) between primary and recurring tumors.

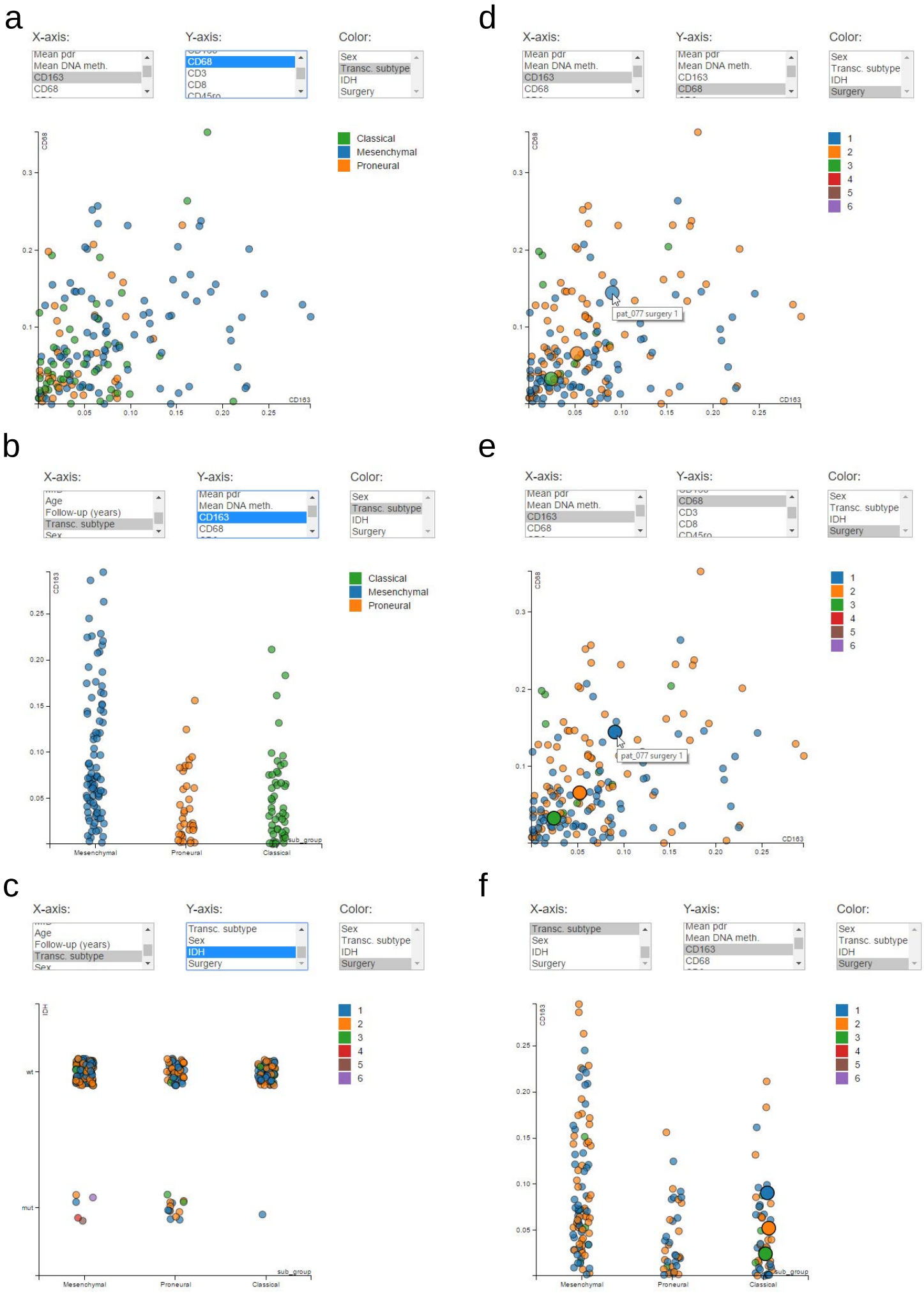
(f) Scatterplot displaying the association between epiallele shift during progression (as measured by the EPM score) and the time between first and second surgery. r: Pearson correlation, p: p-value.

(g) Histogram showing the distribution of the distance to the nearest gene for loci assessed in the epiallele entropy comparison between control samples (non-cancerous brain tissue), primary tumors (t=1), and recurring tumors (t=2), stratified by whether or not they were identified as eloci.

(h) LOLA region set enrichment analysis for eloci identified in the comparison of primary and recurring tumors. Region sets with an adjusted p-value below 0.001 are shown.

The number of patients (N) in the Kaplan-Meier analysis (panels c, e) is provided in Supplementary Table 3.

In the association plots (panels b, d, f), N denotes the number of tumor samples.


a
X-axis:
Mean pdr
Mean DNA meth.
CD163
CD68
Y-axis:
CD68
CD3
CD8
Color:
Sex
Transc. subtype
IDH
Surgery
Classical
Mesenchymal
Proneural
CD68
CD163
d
pat_077 surgery 1
1
2
3
4
5
6
b
Age
Follow-up (years)
sub_group
Mesenchymal
Proneural
Classical
e
c
IDH
wt
mut
f

Supplementary Figure 13

Illustration of the graphical data explorer on the Supplementary Website (http://glioblastoma-progression.computational-epigenetics.org/)

(a) Comparison between two continuous variables.

(b) Comparison between one continuous and one categorical variable.

(c) Comparison between two categorical variables.

(d) Hovering over an individual data point shows information about the specific data point. It also highlights all matched samples from the same patient.

(e, f) Clicking on a data point locks the highlighting (e), allowing the user to follow the selected data point through additional analyses (f).

Supplementary Table 1. Patient summary table

Clinical annotation data for all patients and samples included in this study.

Supplementary Table 2. RRBS summary table

Sequencing statistics for the RRBS profiling of all samples included in this study.

Supplementary Table 3. Survival analysis summary table

Summary of Cox regression (sheet 1) and Kaplan-Meier survival analyses (sheet 2), assessing the association of the specified variables with overall survival (OS) and progression-free survival (PFS). The number of patients and events (N) are shown for each analysis, and 95% confidence intervals are indicated. P-values were calculated using two-sided Wald tests (for Cox regression) and two-sided log-rank tests (for Kaplan-Meier survival analysis) without adjustment for multiple testing.

Supplementary Table 4. Association analysis summary table

Summary of the statistical tests for association among categorical and continuous variables. The number of tumor samples (N) in each analysis as well as raw and adjusted p-values (BH: Benjamini-Hochberg, BF: Bonferroni) are indicated in the tables. P-values were calculated with a two-sided Wilcoxon rank sum test.