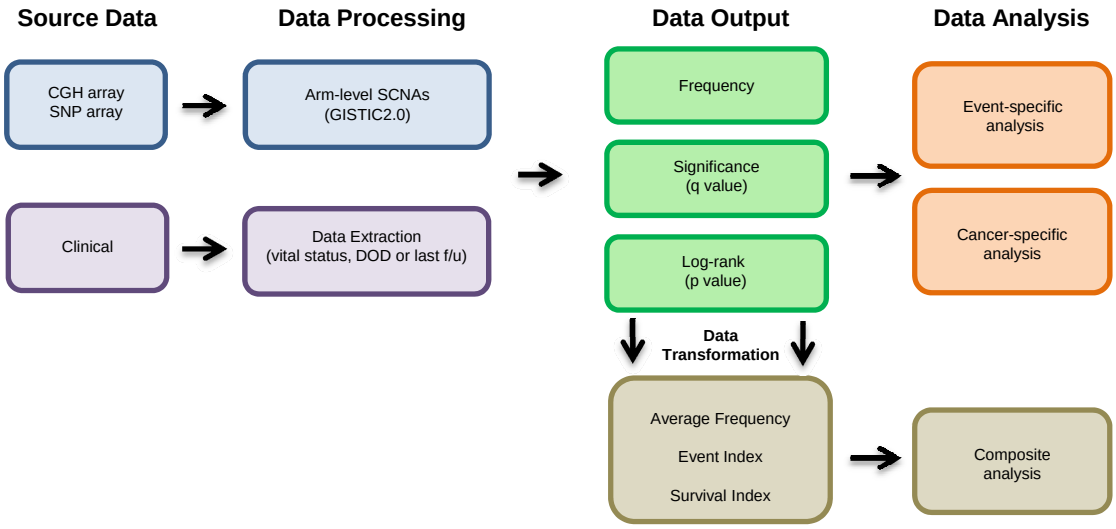
Supplemental Data

A



B

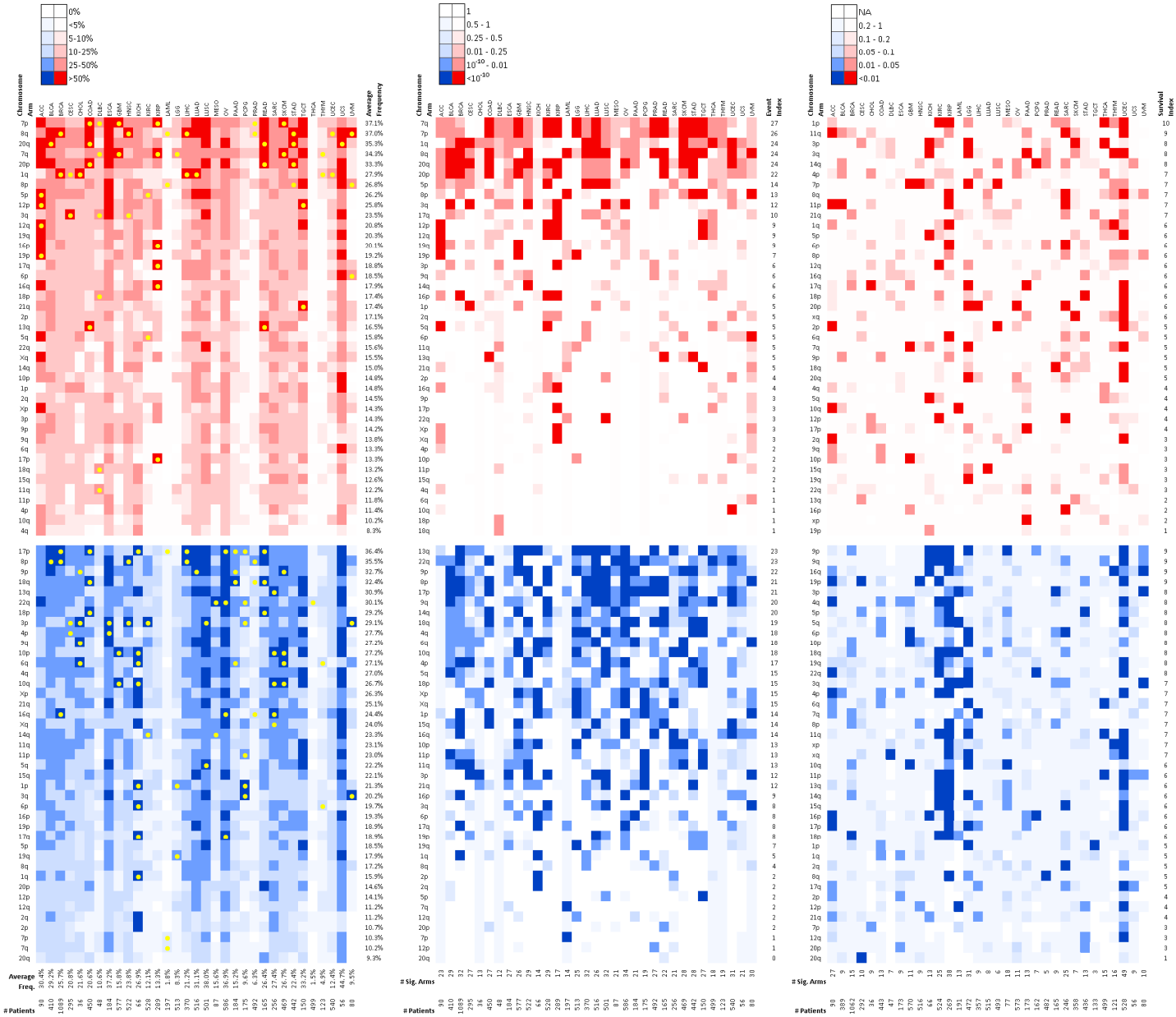


Figure S1, related to Figure 1. Landscape of SCNA alterations and prognostic associations across cancer.

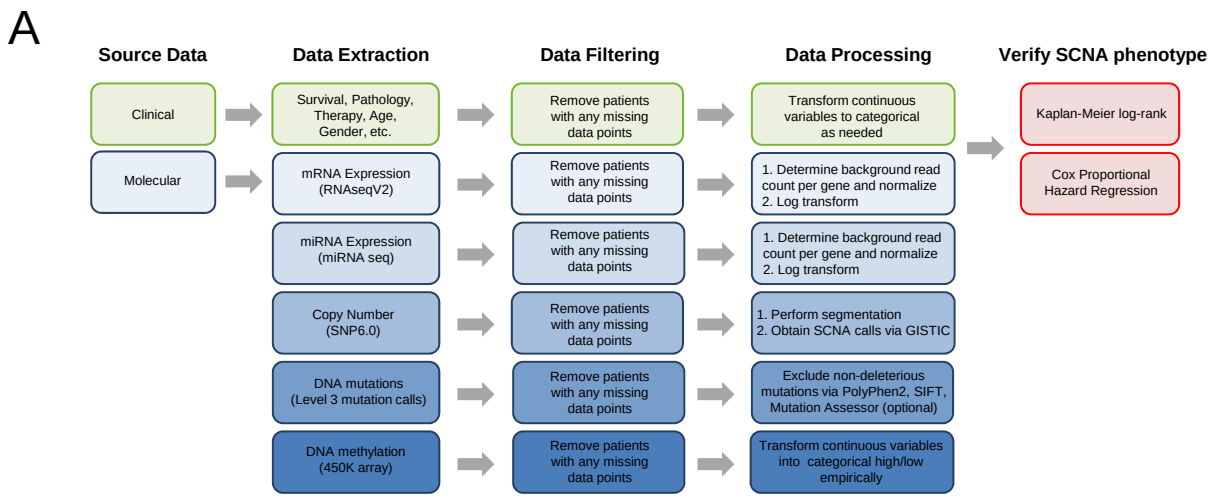
(A) Data acquisition, processing, and analytical workflow for pan-cancer characterization of arm-level SCNA prognostic impact.

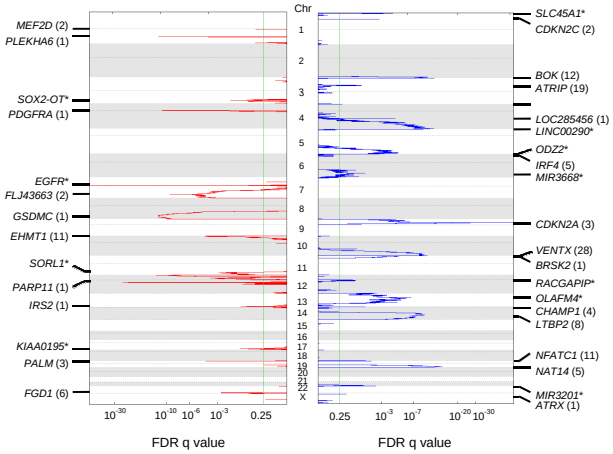
(B) Heatmaps depicting arm-level SCNA frequencies (left panel), GISTIC event q value (middle panel), and log-rank survival p value (right panel) in 33 TCGA datasets. Tumor types labeled along the top. Abbreviations established by TCGA (tcga-data.nci.nih.gov/tcga). Yellow dot denotes individual arm-level SCNA frequency above background (i.e. frequency >2 standard deviations above total SCNA average frequency within each tumor type). Event Index is the number of datasets where SCNA q<0.25. Survival Index is the number of datasets where SCNA p<0.1.

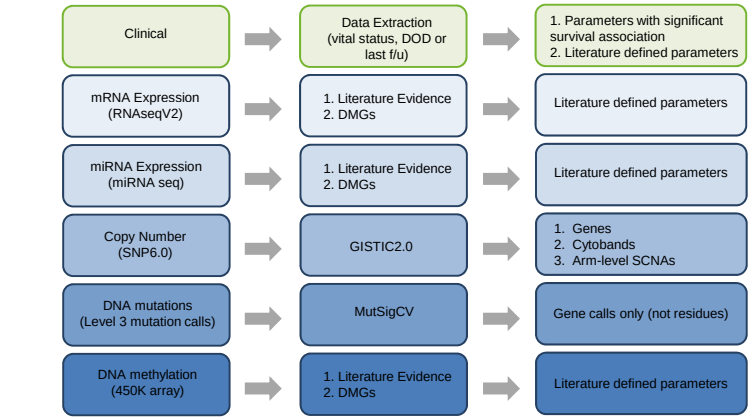
(C) Two-dimensional representation of SCNA frequency, GISTIC significance -$\log_{10}$(q value), and log-rank overall survival association -$\log_{10}$(p value) in TCGA lower grade glioma (LGG).

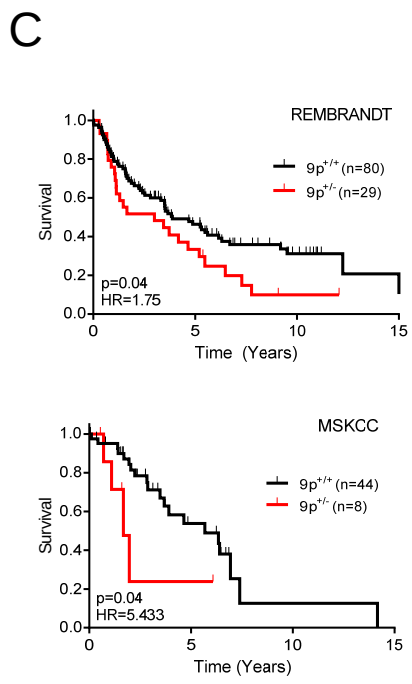
SCNA, somatic copy number alteration; DOD, date of death; f/u, follow up; Freq, frequency; Sig, significant; NA, not available. SCNA, somatic copy number alteration. High/Low scale corresponds to the 3 parameters individually.

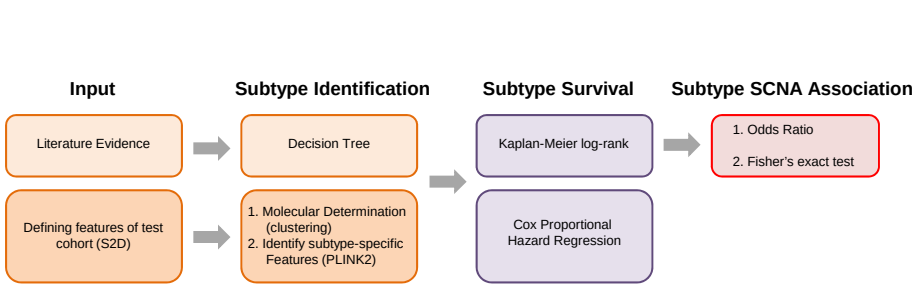
Table S1, related to Figure 1. Arm-level somatic copy number alterations (SCNAs) in 33 TCGA cancer types. Provided as an Excel file.

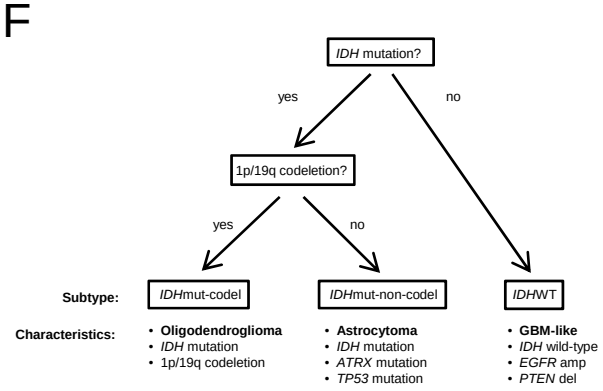

A
Source Data
Data Extraction
Data Filtering
Data Processing
Verify SCNA phenotype
Clinical
Molecular
Survival, Pathology, Therapy, Age, Gender, etc.
mRNA Expression (RNAseqV2)
miRNA Expression (miRNA seq)
Copy Number (SNP6.0)
DNA mutations (Level 3 mutation calls)
DNA methylation (450K array)
Remove patients with any missing data points
Transform continuous variables to categorical as needed
1. Determine background read count per gene and normalize
2. Log transform
1. Perform segmentation
2. Obtain SCNA calls via GISTIC
Exclude non-deleterious mutations via PolyPhen2, SIFT, Mutation Assessor (optional)
Transform continuous variables into categorical high/low empirically
Kaplan-Meier log-rank
Cox Proportional Hazard Regression

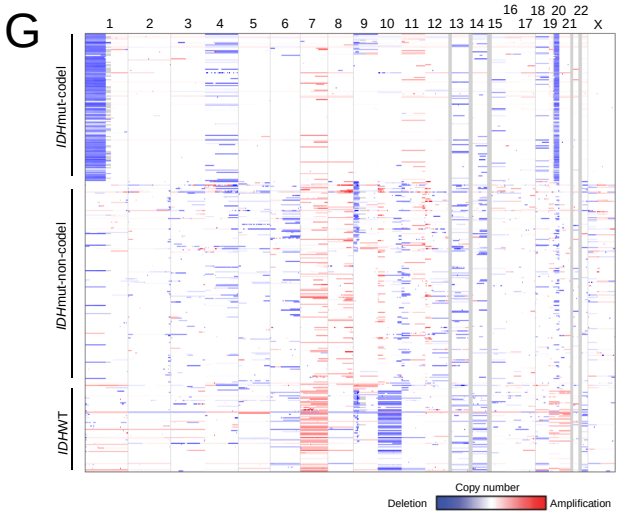

B
MEF2D (2)
PLEKHA6 (1)
SOX2-OT*
PDGFRA (1)
EGFR*
FLJ43663 (2)
GSDMC (1)
EHMT1 (11)
SORL1*
PARP11 (1)
IRS2 (1)
KIAA0195*
PALM (3)
FGD1 (6)
Chr
SLC45A1*
CDKN2C (2)
BOK (12)
ATRIP (19)
LOC285456 (1)
LINC00290*
ODZ2*
IRF4 (5)
MIR3668*
CDKN2A (3)
VENTX (28)
BRSK2 (1)
RACGAPIP*
OLAFM4*
CHAMP1 (4)
LTBP2 (8)
NFATC1 (11)
NAT14 (5)
MIR3201*
ATRX (1)
FDR q value
FDR q value

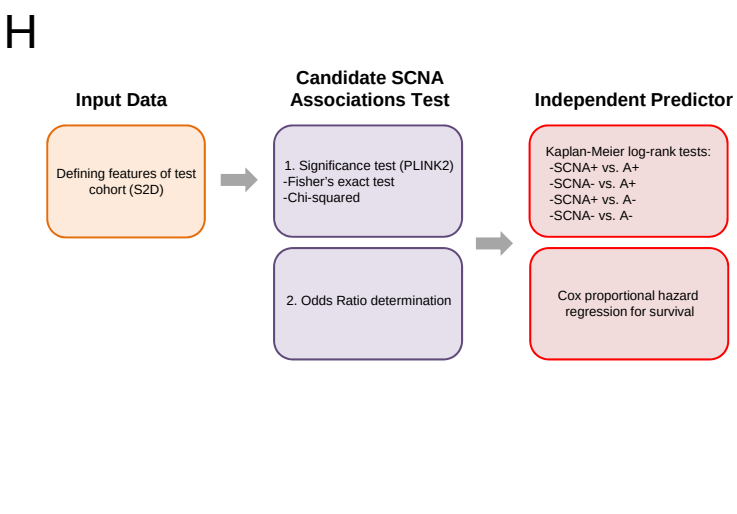

D
Input Data
Data Processing
Data Output
Clinical
mRNA Expression (RNAseqV2)
miRNA Expression (miRNA seq)
Copy Number (SNP6.0)
DNA mutations (Level 3 mutation calls)
DNA methylation (450K array)
Data Extraction (vital status, DOD or last f/u)
1. Literature Evidence
2. DMGs
GISTIC2.0
MutSigCV
1. Parameters with significant survival association
2. Literature defined parameters
Literature defined parameters
1. Genes
2. Cytobands
3. Arm-level SCNAs
Gene calls only (not residues)

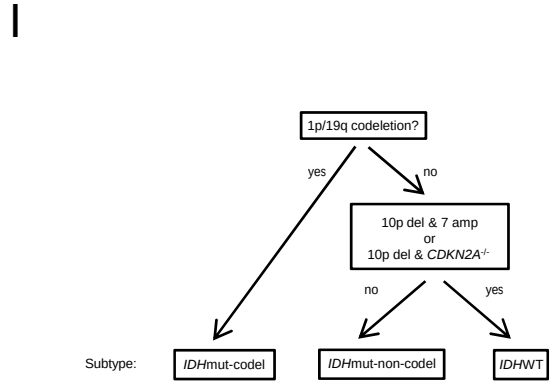

C
REMBRANDT
9p+/+ (n=80)
9p+/- (n=29)
Survival
p=0.04
HR=1.75
Time (Years)
MSKCC
9p+/+ (n=44)
9p+/- (n=8)
Survival
p=0.04
HR=5.433
Time (Years)

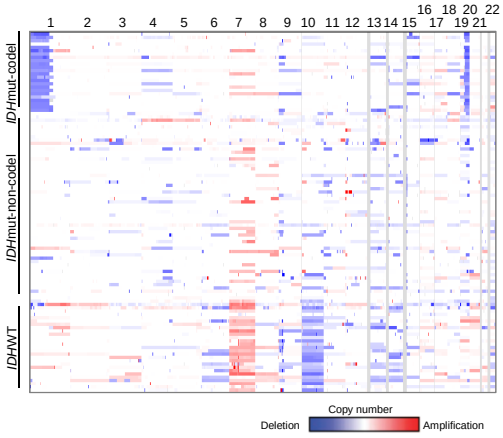

E
Input
Subtype Identification
Subtype Survival
Subtype SCNA Association
Literature Evidence
Defining features of test cohort (S2D)
Decision Tree
1. Molecular Determination (clustering)
2. Identify subtype-specific Features (PLINK2)
Kaplan-Meier log-rank
Cox Proportional Hazard Regression
1. Odds Ratio
2. Fisher's exact test

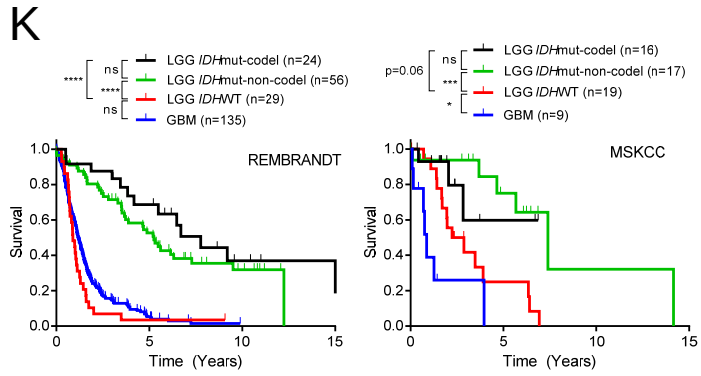

F
IDH mutation?
yes
no
1p/19q codeletion?
yes
no
Subtype:
IDHmut-codel
IDHmut-non-codel
IDHWT
Characteristics:
Oligodendroglioma
IDH mutation
1p/19q codeletion
Astrocytoma
IDH mutation
ATRX mutation
TP53 mutation
GBM-like
IDH wild-type
EGFR amp
PTEN del


G
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X
IDHmut-codel
IDHmut-non-codel
IDHWT
Copy number
Deletion
Amplification


H
Input Data
Candidate SCNA Associations Test
Independent Predictor
Defining features of test cohort (S2D)
1. Significance test (PLINK2)
-Fisher's exact test
-Chi-squared
2. Odds Ratio determination
Kaplan-Meier log-rank tests:
-SCNA+ vs. A+
-SCNA- vs. A+
-SCNA+ vs. A-
-SCNA- vs. A-
Cox proportional hazard regression for survival


I
1p/19q codeletion?
yes
no
10p del & 7 amp
or
10p del & CDKN2A-/-
no
yes
Subtype:
IDHmut-codel
IDHmut-non-codel
IDHWT


J
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
IDHmut-codel
IDHmut-non-codel
IDHWT
Copy number
Deletion
Amplification


K
****
ns
****
ns
LGG IDHmut-codel (n=24)
LGG IDHmut-non-codel (n=56)
LGG IDHWT (n=29)
GBM (n=135)
REMBRANDT
p=0.06
ns
***
*
LGG IDHmut-codel (n=16)
LGG IDHmut-non-codel (n=17)
LGG IDHWT (n=19)
GBM (n=9)
MSKCC
Survival
Time (Years)

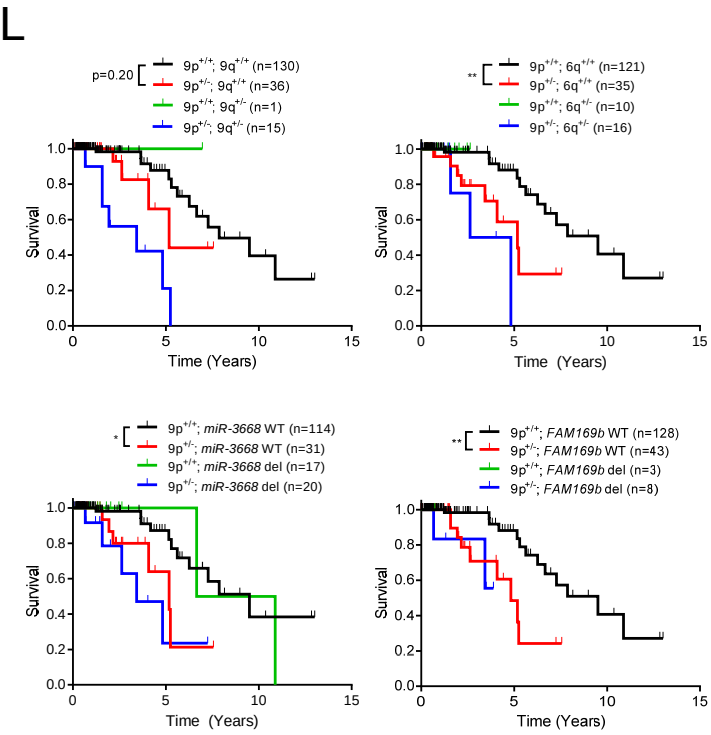

L
p=0.20
9p+/+; 9q+/+ (n=130)
9p+/-; 9q+/+ (n=36)
9p+/+; 9q+/- (n=1)
9p+/-; 9q+/- (n=15)
**
9p+/+; 6q+/+ (n=121)
9p+/-; 6q+/+ (n=35)
9p+/+; 6q+/- (n=10)
9p+/-; 6q+/- (n=16)
*
9p+/+; miR-3668 WT (n=114)
9p+/-; miR-3668 WT (n=31)
9p+/+; miR-3668 del (n=17)
9p+/-; miR-3668 del (n=20)
**
9p+/+; FAM169b WT (n=128)
9p+/-; FAM169b WT (n=43)
9p+/+; FAM169b del (n=3)
9p+/-; FAM169b del (n=8)
Survival
Time (Years)

Figure S2, related to Figure 2. An analysis of 9p loss by subtype in LGG.

(A) Data acquisition, processing, and analytical workflow for data collection and establishment of test cohorts**.**

(B) Amplification (left, red) and deletion (right, blue) peaks from across the genome (y-axis) in TCGA LGG cohort via GISTIC. Known or putative gene targets within the peak regions are indicated for significant peaks (Asterisk indicates nearest gene for peaks not containing any genes). Values in parentheses represent the number of genes in the peak region. GISTIC false discovery rate (FDR) q values (x-axis) are shown. Green line denotes significance threshold ($q<0.25$).

(C) Kaplan-Meier curves showing survival outcome for 9p loss in REMBRANDT and MSKCC LGG cohorts.

(D) Data acquisition, processing, and workflow for the identification of disease-defining clinical and molecular features.

(E) Workflow for defining molecular subtypes of the larger disease cohort and testing for survival differences.

(F) Schematic of LGG subtypes according to known molecular criteria and previously published characteristics of each group (Yan et al., 2009).

(G) Segmentation map of all SCNAs in the TCGA LGG cohort clustered by subtype and 9p deletion status.

(H) Strategy for identification and elimination of potential confounding variables.

(I) Schematic of LGG subtypes according to copy number alteration only in TCGA LGG discovery and validation datasets.

(J) Segmentation map of all SCNAs in the REMBRANDT LGG cohort clustered by subtype.

(K) Kaplan-Meier curve showing survival outcome in REMBRANDT and MSKCC cohorts for GBM and LGG by subtype.

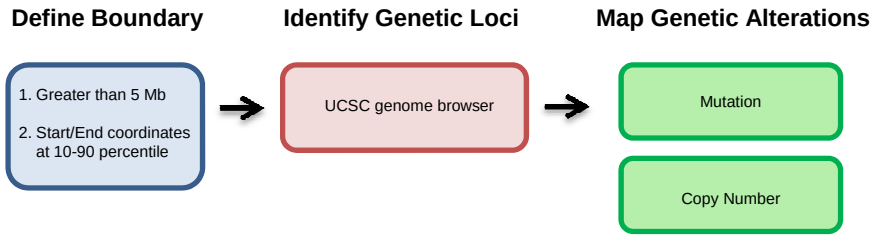
(L) Kaplan-Meier curve showing survival outcome of 9p and/or 9q, 6q, *miR-3668*, and *FAM169b* loss in *IDH*mut-non-codel subtype in the TCGA LGG cohort.

SCNA, somatic copy number alteration; LGG, lower grade glioma; GBM, glioblastoma; HR, hazard ratio; amp, amplification; del, deletion; mut, mutation; codel, codeletion; WT, wild-type; DMGs, differentially methylated genes; A, any alteration; +, present; -, absent; DEGs, differentially expressed genes. *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$; ****, $p<0.0001$; ns, not significant.

Table S2, related to Figure 2. Pairwise associations between significant covariates in the TCGA LGG cohort. Provided as an Excel file.

Table S3, related to Figure 2. Pairwise associations between 9p loss and significant covariates in the TCGA LGG cohort. Provided as an Excel file.

A



B

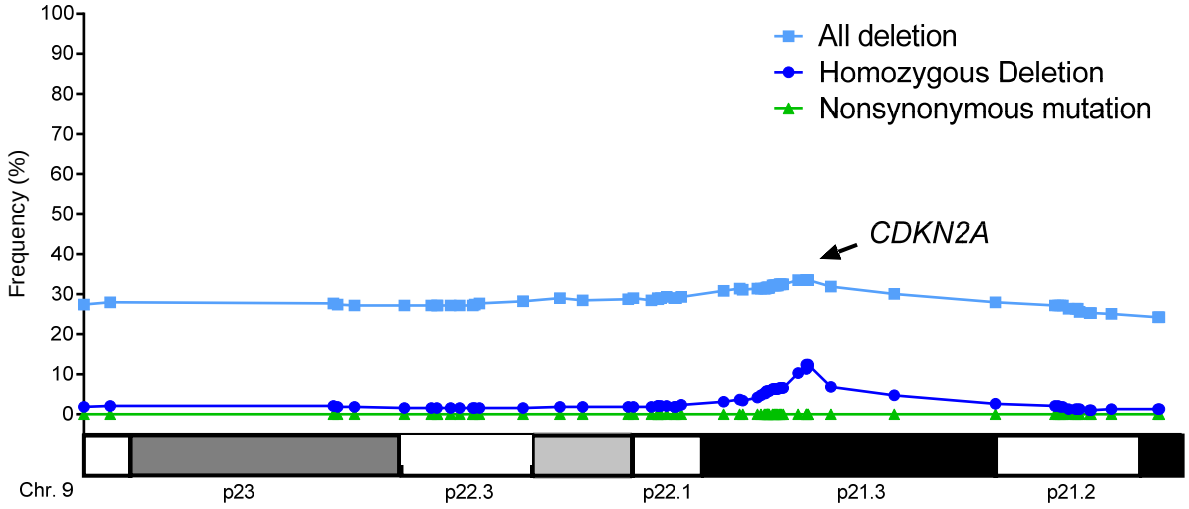


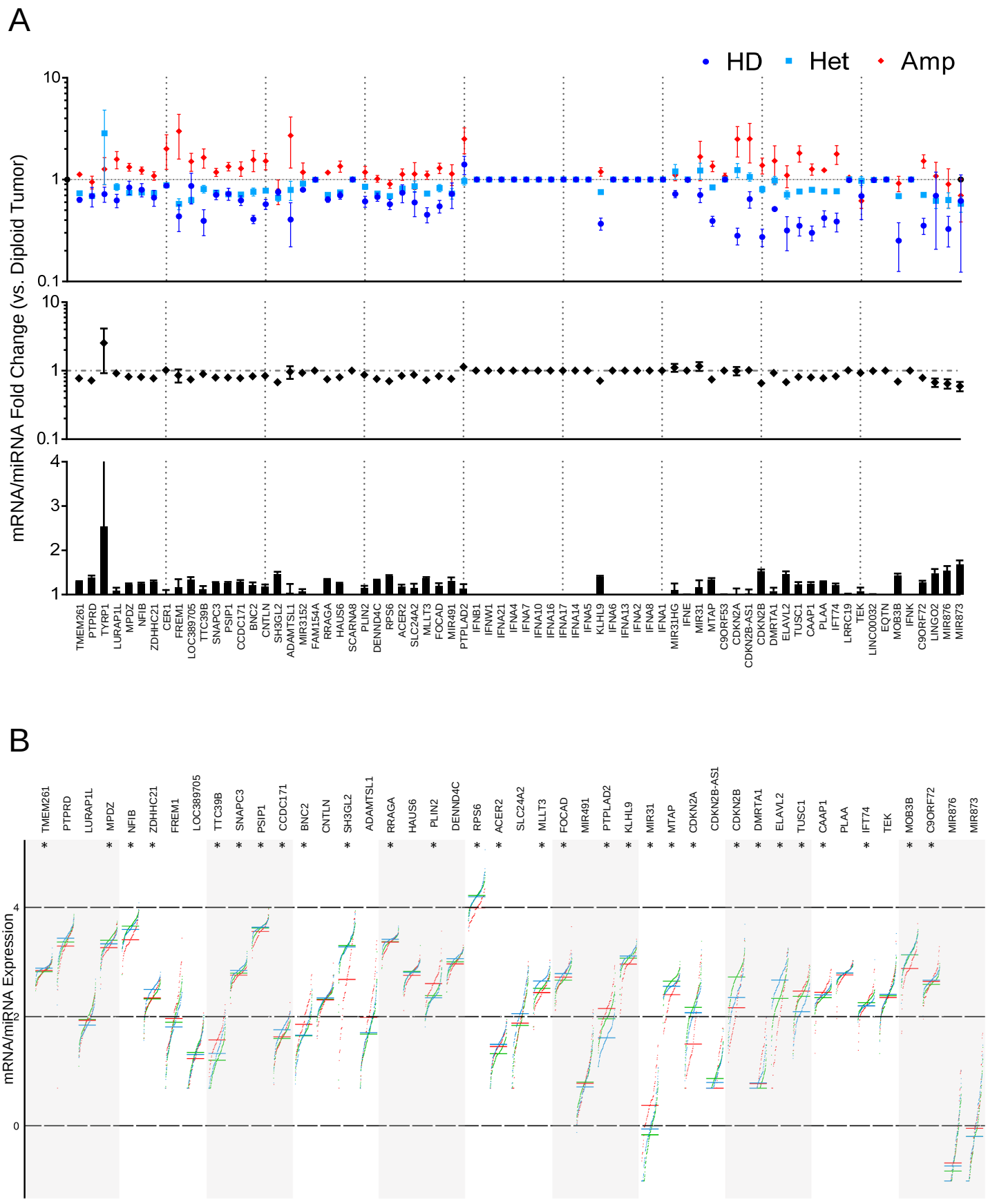
Figure S3 related to Figure 3. Genetic alterations by locus across the 9p commonly deleted region.

(A) Workflow for the characterization of individual arm-level SCNAs and their genetic features.

(B) Frequency of all deletion events, homozygous deletions, and nonsynonymous mutations at 72 loci within the 9p deletion in the TCGA LGG cohort.

Chr, chromosome; Mb; Megabase.

Table S4, related to Figure 3. Genetic loci contained within the 9p commonly deleted region in LGG. Provided as an Excel file.


A
HD
Het
Amp
mRNA/miRNA Fold Change (vs. Diploid Tumor)
10
1
0.1
4
3
2
1
TMEM261
PTPRD
TYRP1
LURAP1L
MPDZ
NFIB
ZDHHC21
CER1
FREM1
LOC389705
TTC39B
SNAPC3
PSIP1
CCDC171
BNC2
CNTLN
SH3GL2
ADAMTSL1
MIR3152
FAM154A
RRAGA
HAUS6
SCARNA8
PLIN2
DENND4C
RPS6
ACER2
SLC24A2
MLLT3
FOCAD
MIR491
PTPLAD2
IFNB1
IFNW1
IFNA21
IFNA4
IFNA7
IFNA10
IFNA16
IFNA17
IFNA14
IFNA5
KLHL9
IFNA6
IFNA13
IFNA2
IFNA8
IFNA1
MIR31HG
IFNE
MIR31
MTAP
C9ORF53
CDKN2A
CDKN2B-AS1
CDKN2B
DMRTA1
ELAVL2
TUSC1
CAAP1
PLAA
IFT74
LRRC19
TEK
LINC00032
EQTN
MOB3B
IFNK
C9ORF72
LINGO2
MIR876
MIR873
B
mRNA/miRNA Expression
4
2
0

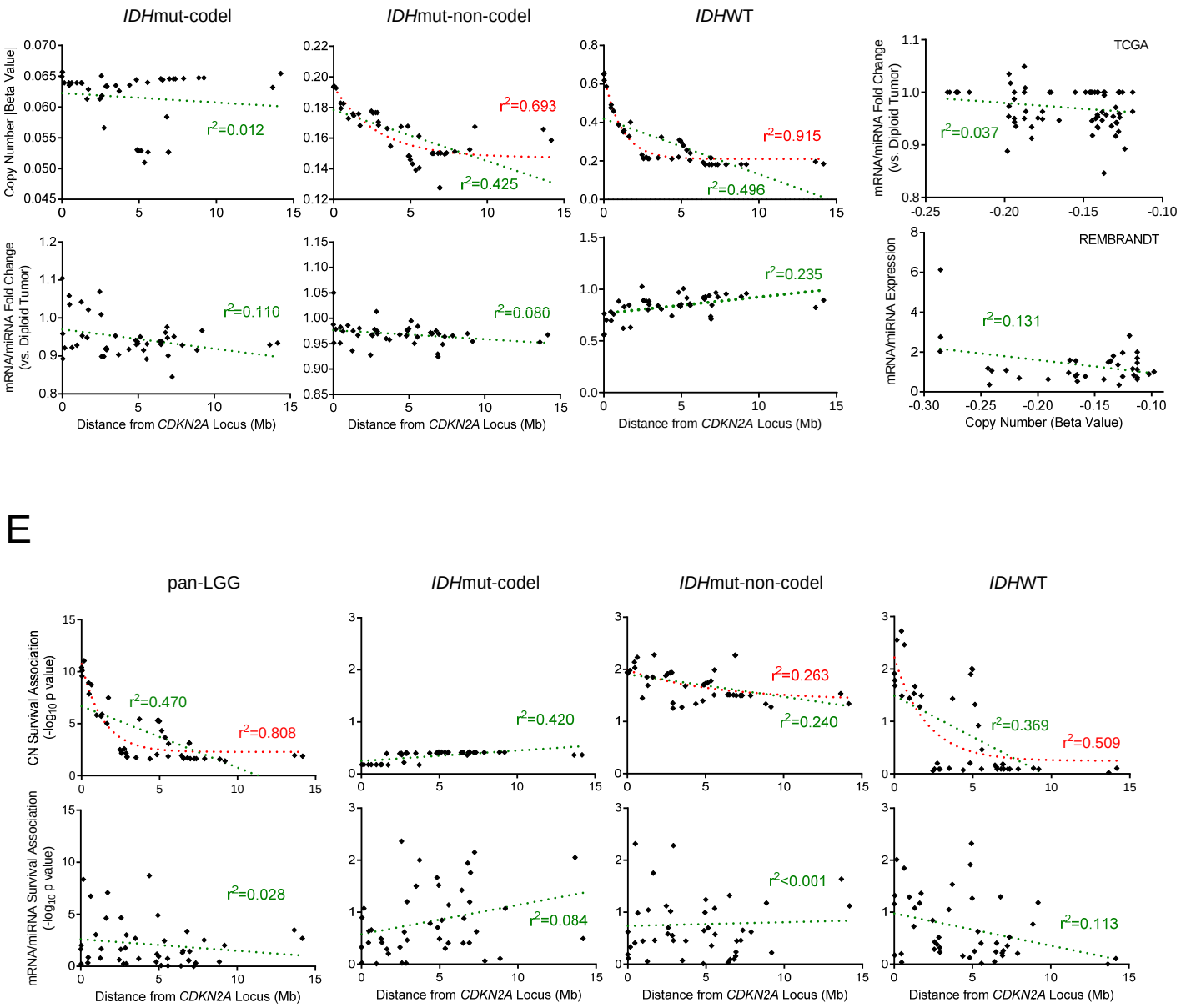

C
IDHmut-codel
IDHmut-non-codel
IDHWT
Copy Number |Beta Value|
mRNA/miRNA Fold Change (vs. Diploid Tumor)
r2=0.012
r2=0.693
r2=0.425
r2=0.915
r2=0.496
r2=0.110
r2=0.080
r2=0.235
Distance from CDKN2A Locus (Mb)
D
TCGA
REMBRANDT
mRNA/miRNA Fold Change (vs. Diploid Tumor)
mRNA/miRNA Expression
r2=0.037
r2=0.131
Copy Number (Beta Value)
E
pan-LGG
IDHmut-codel
IDHmut-non-codel
IDHWT
CN Survival Association (-log10 p value)
mRNA/miRNA Survival Association (-log10 p value)
r2=0.470
r2=0.808
r2=0.420
r2=0.263
r2=0.240
r2=0.369
r2=0.509
r2=0.028
r2=0.084
r2<0.001
r2=0.113
Distance from CDKN2A Locus (Mb)

Figure S4, related to Figure 4. Analysis of copy number status and expression by gene within the 9p commonly deleted region.

(A) mRNA/miRNA expression fold change vs. diploid tumor shown based on zygosity (top panel), average fold change (middle panel), and absolute fold change (lower panel) at 72 loci within the 9p commonly deleted region in the TCGA LGG cohort. Error bars, ± SEM.

(B) mRNA/miRNA expression (log read counts) in the TCGA LGG cohort by individual LGG molecular subtype. Red, *IDH*WT; Blue, *IDH*mut-codel; Green, *IDH*mut-non-codel. Colored bars represent median for each group. *, $p<0.01$ (ANOVA).

(C) Linear (green) and logarithmic (red) regression plots of copy number (top) or mRNA/miRNA fold change (bottom) vs. distance from the *CDKN2A* locus in the TCGA LGG cohort.

(D) Linear regression plot of mRNA/miRNA expression vs. copy number in the TCGA and REMBRANDT LGG cohorts.

(E) Linear (green) and logarithmic (red) regression plots of $-\log_{10}$(p value) following Cox regression proportional hazard test for overall survival at 44 loci within the 9p commonly deleted region according to copy number beta value (top) or mRNA/miRNA expression (bottom) vs. distance from *CDKN2A* locus in the TCGA LGG cohort.

CN, copy number; HD, homozygous deletion; Het, heterozygous deletion; Amp, amplification; Mb, megabase; LGG, lower grade glioma.

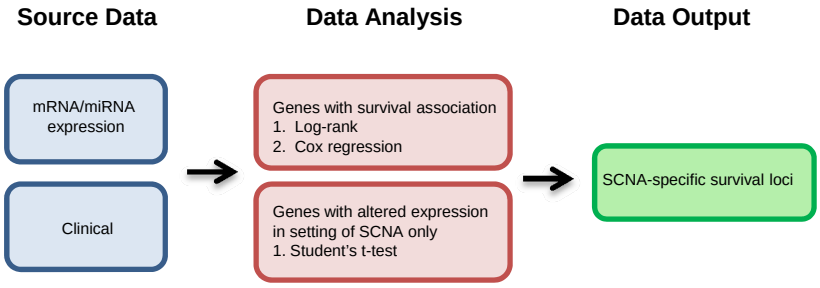
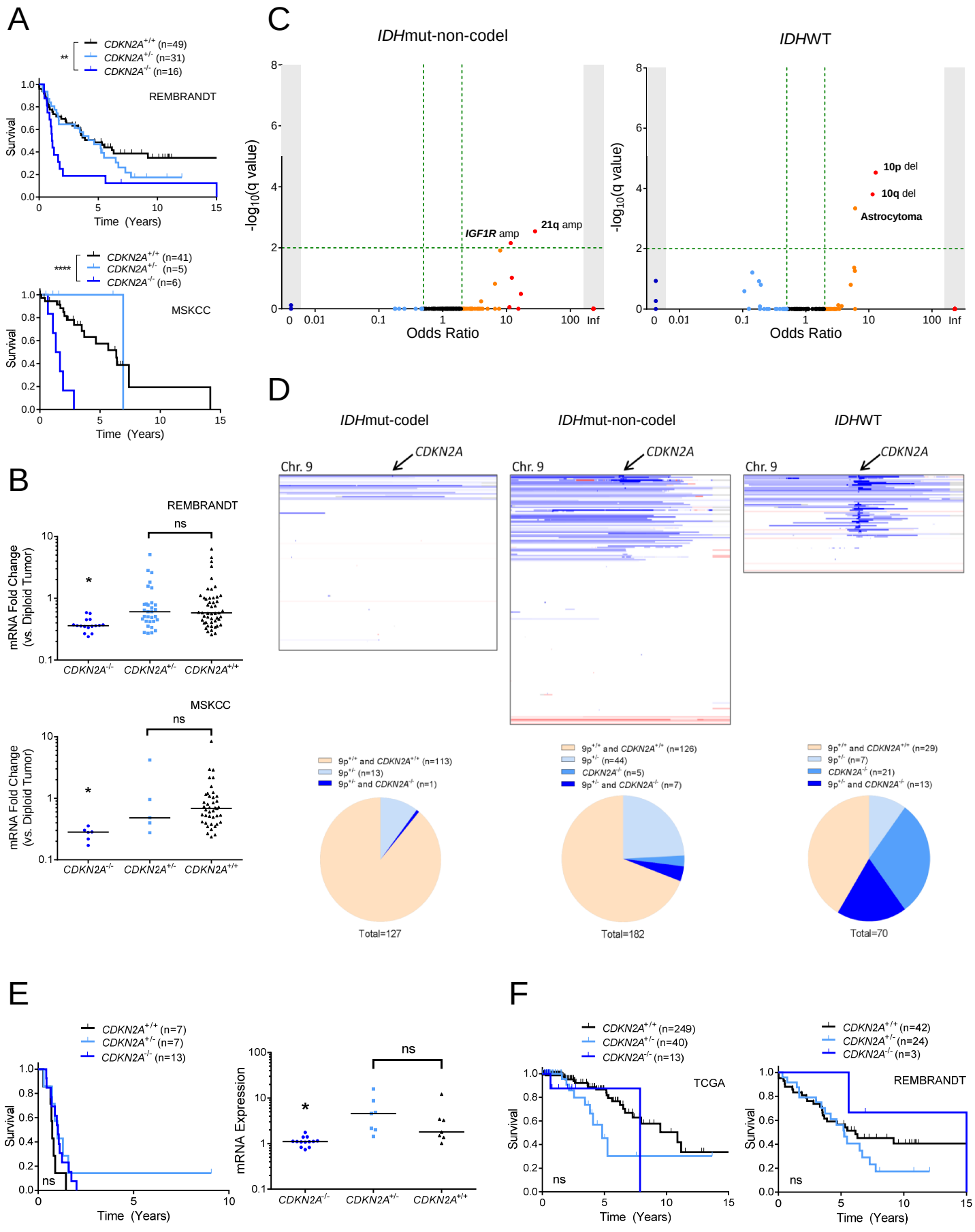


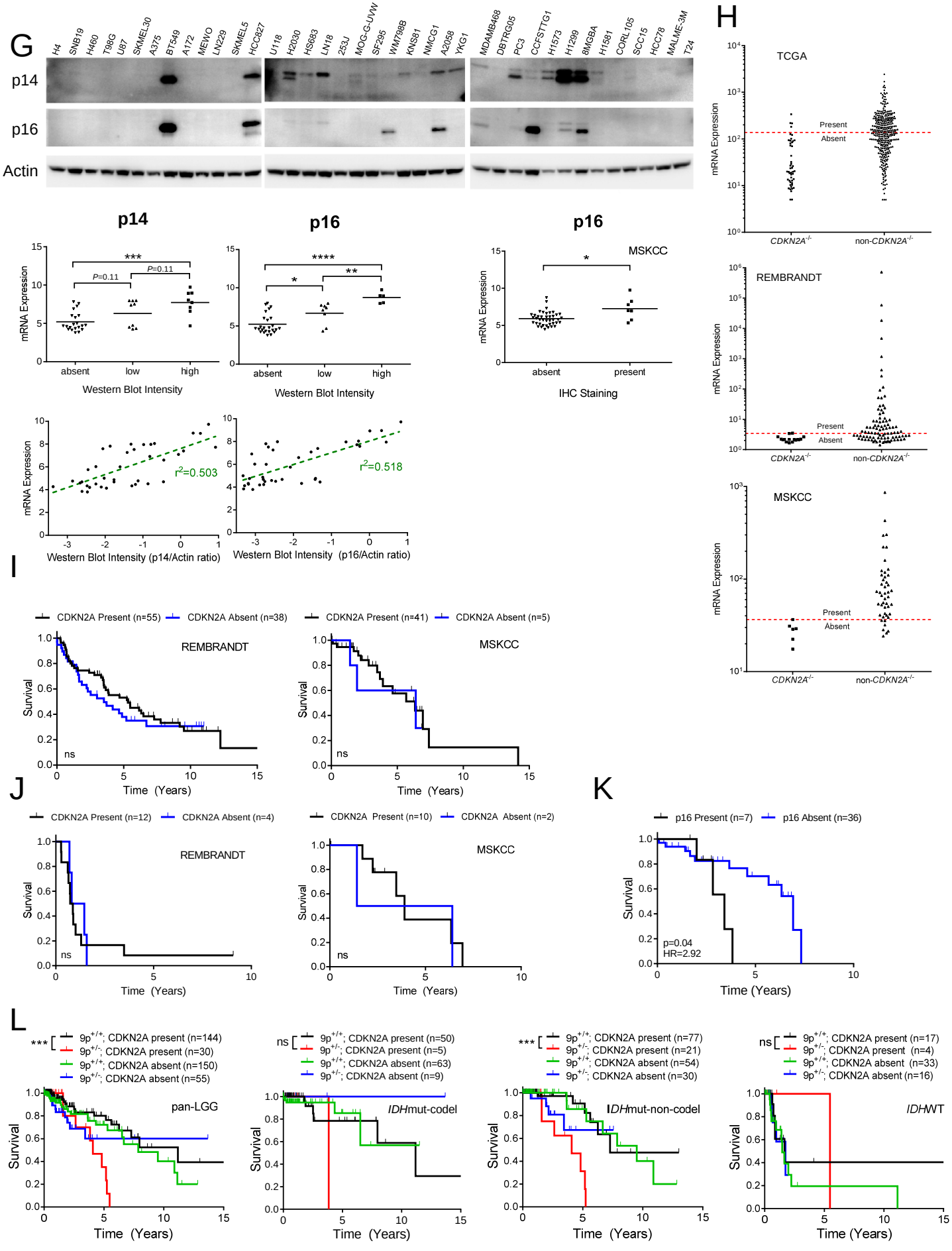
Figure S5, related to Figure 5. Pinpointing genetic loci that are SCNA event-specific prognostic markers.

Workflow for identifying genetic loci that are SCNA event-specific prognostic markers.

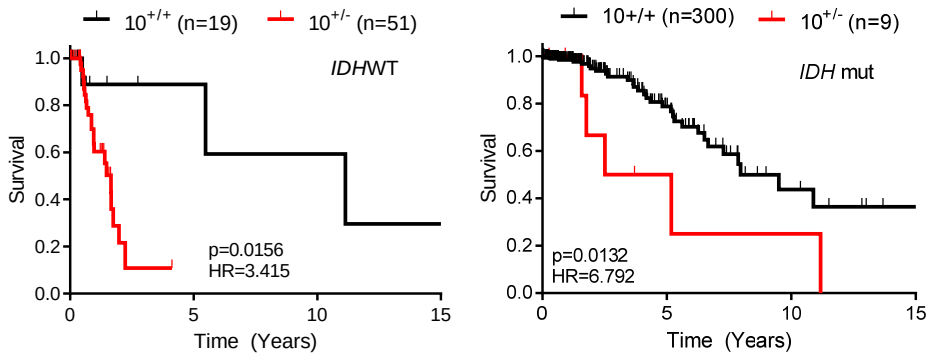
SCNA, somatic copy number alteration.

Table S5, related to Figure 5. Univariate survival analysis of genes targeted by 9p loss. Provided as an Excel file.


A
CDKN2A+/+ (n=49)
CDKN2A+/- (n=31)
CDKN2A-/- (n=16)
**
REMBRANDT
Survival
Time (Years)
CDKN2A+/+ (n=41)
CDKN2A+/- (n=5)
CDKN2A-/- (n=6)
****
MSKCC
Survival
Time (Years)
B
REMBRANDT
ns
mRNA Fold Change (vs. Diploid Tumor)
CDKN2A-/-
CDKN2A+/-
CDKN2A+/+
MSKCC
ns
mRNA Fold Change (vs. Diploid Tumor)
CDKN2A-/-
CDKN2A+/-
CDKN2A+/+
C
IDHmut-non-codel
-log10(q value)
IGF1R amp
21q amp
Odds Ratio
IDHWT
-log10(q value)
10p del
10q del
Astrocytoma
Odds Ratio
D
IDHmut-codel
IDHmut-non-codel
IDHWT
Chr. 9
CDKN2A
9p+/+ and CDKN2A+/+ (n=113)
9p+/- (n=13)
9p+/- and CDKN2A-/- (n=1)
Total=127
9p+/+ and CDKN2A+/+ (n=126)
9p+/- (n=44)
CDKN2A-/- (n=5)
9p+/- and CDKN2A-/- (n=7)
Total=182
9p+/+ and CDKN2A+/+ (n=29)
9p+/- (n=7)
CDKN2A-/- (n=21)
9p+/- and CDKN2A-/- (n=13)
Total=70
E
CDKN2A+/+ (n=7)
CDKN2A+/- (n=7)
CDKN2A-/- (n=13)
Survival
ns
Time (Years)
mRNA Expression
ns
CDKN2A-/-
CDKN2A+/-
CDKN2A+/+
F
CDKN2A+/+ (n=249)
CDKN2A+/- (n=40)
CDKN2A-/- (n=13)
TCGA
Survival
ns
Time (Years)
CDKN2A+/+ (n=42)
CDKN2A+/- (n=24)
CDKN2A-/- (n=3)
REMBRANDT
Survival
ns
Time (Years)


G
H4 SNB19 H460 T98G U87 SKMEL30 A375 BT549 A172 MEWO LN229 SKMEL5 HCC827 U118 H2030 HS683 LN18 253J MOG-G-UVW SF295 WM798B KNS81 NMCG1 A2058 YKG1 MDAMB468 DBTRG05 PC3 CCFSTTG1 H1573 H1299 8MGBA H1581 CORL105 SCC15 HCC78 MALME-3M T24
p14
p16
Actin
p14
p16
p16
mRNA Expression
absent low high
Western Blot Intensity
P=0.11
P=0.11
***
*
**
****
r2=0.503
r2=0.518
Western Blot Intensity (p14/Actin ratio)
Western Blot Intensity (p16/Actin ratio)
MSKCC
absent present
IHC Staining
H
TCGA
REMBRANDT
MSKCC
Present
Absent
CDKN2A-/-
non-CDKN2A-/-
I
CDKN2A Present (n=55)
CDKN2A Absent (n=38)
CDKN2A Present (n=41)
CDKN2A Absent (n=5)
REMBRANDT
MSKCC
Survival
Time (Years)
ns
J
CDKN2A Present (n=12)
CDKN2A Absent (n=4)
CDKN2A Present (n=10)
CDKN2A Absent (n=2)
REMBRANDT
MSKCC
K
p16 Present (n=7)
p16 Absent (n=36)
p=0.04
HR=2.92
L
***
ns
9p+/+; CDKN2A present (n=144)
9p+/-; CDKN2A present (n=30)
9p+/+; CDKN2A absent (n=150)
9p+/-; CDKN2A absent (n=55)
9p+/+; CDKN2A present (n=50)
9p+/-; CDKN2A present (n=5)
9p+/+; CDKN2A absent (n=63)
9p+/-; CDKN2A absent (n=9)
9p+/+; CDKN2A present (n=77)
9p+/-; CDKN2A present (n=21)
9p+/+; CDKN2A absent (n=54)
9p+/-; CDKN2A absent (n=30)
9p+/+; CDKN2A present (n=17)
9p+/-; CDKN2A present (n=4)
9p+/+; CDKN2A absent (n=33)
9p+/-; CDKN2A absent (n=16)
pan-LGG
IDHmut-codel
IDHmut-non-codel
IDHWT
Survival
Time (Years)

M


10+/+ (n=19)
10+/- (n=51)
IDHWT
p=0.0156
HR=3.415
10+/+ (n=300)
10+/- (n=9)
IDH mut
p=0.0132
HR=6.792
Survival
Time (Years)

N

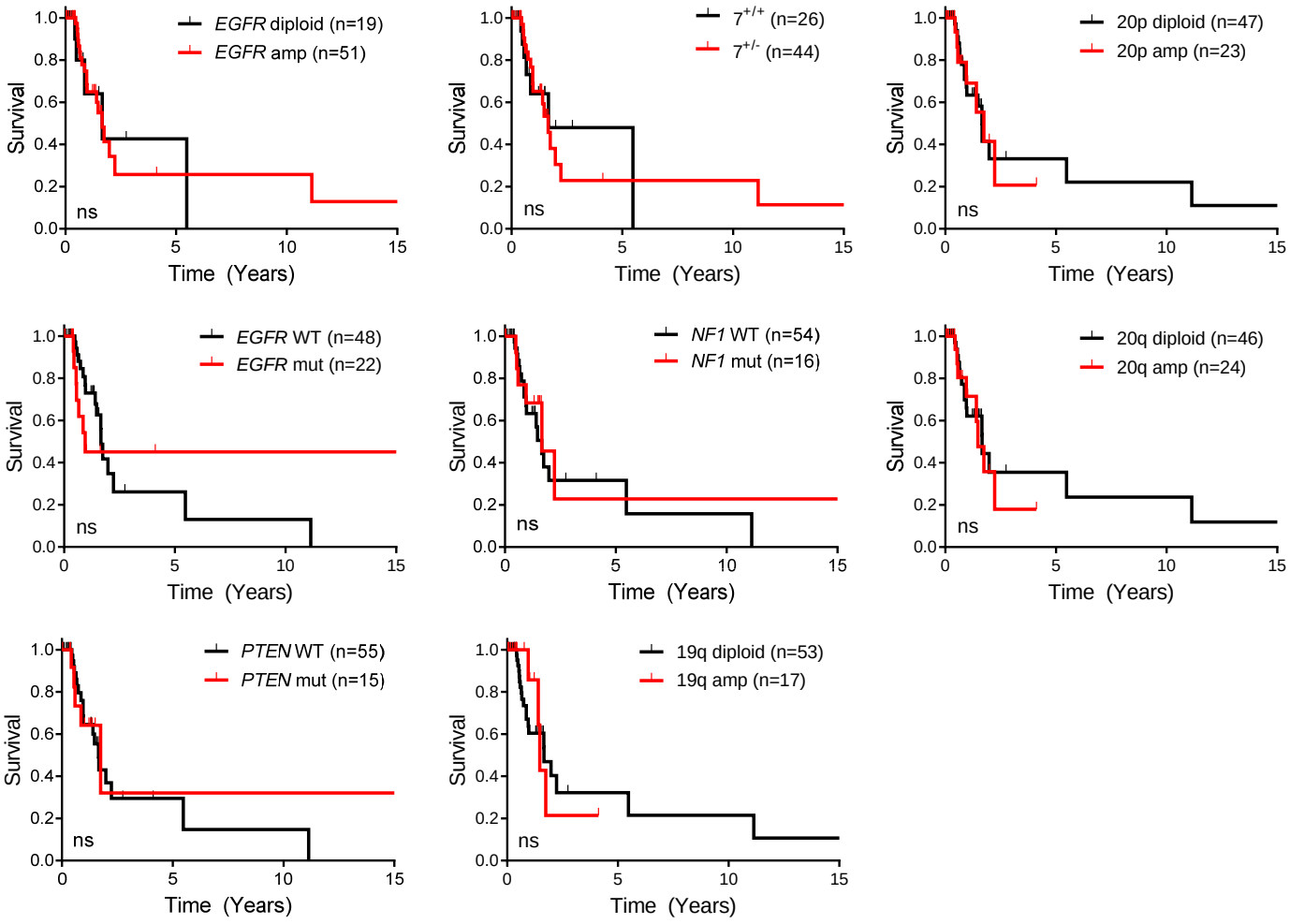

EGFR diploid (n=19)
EGFR amp (n=51)
7+/+ (n=26)
7+/- (n=44)
20p diploid (n=47)
20p amp (n=23)
EGFR WT (n=48)
EGFR mut (n=22)
NF1 WT (n=54)
NF1 mut (n=16)
20q diploid (n=46)
20q amp (n=24)
PTEN WT (n=55)
PTEN mut (n=15)
19q diploid (n=53)
19q amp (n=17)
ns
Survival
Time (Years)

Figure S6, related to Figure 6. Clarifying the role of *CDKN2A* in LGG.

(A and B) Kaplan-Meier curve showing survival outcomes (A) and mRNA fold change (B) in both REMBRANDT and MSKCC LGG cohorts based on *CDKN2A* copy number status. Horizontal bars in (B) depict the median.

(C) Scatterplot showing top clinical and molecular associations with *CDKN2A* homozygous deletion in TCGA LGG subtypes *IDH*mut-non-codel and *IDH*WT. Fisher's exact test $-\log_{10}$(q value) (y-axis) and odds ratio (x-axis) for each association are shown.

(D) Segmentation map of SCNAs in the 9p commonly deleted region (top panel) and pie chart showing *CDKN2A* and 9p deletion frequencies (bottom panel) by LGG subtype in the TCGA cohort.

(E) Kaplan-Meier curve for survival outcomes (left panel) and mRNA expression (right panel) in REMBRANDT LGG *IDH*WT tumors based on *CDKN2A* deletion status. Horizontal bar depicts the median.

(F) Kaplan-Meier curve for overall survival (OS) in TCGA and REMBRANDT *IDH* mutant tumors depending on *CDKN2A* copy number status.

(G) Western blot of p14 and p16 in 39 cell lines (top) and corresponding association between protein and mRNA expression in cell lines (bottom left) and MSKCC patient tumors (bottom right). Horizontal bar depicts the median. Expression microarray data for cell lines was obtained from the Cancer Cell Line Encyclopedia (CCLE).

(H) Scatterplots showing CDKN2A mRNA expression in TCGA, REMBRANDT, and MSKCC LGG cohorts based on the absence or presence of *CDKN2A* homozygous deletion. Red dotted line indicates the 90th percentile for CDKN2A mRNA expression in *CDKN2A*$^{-/-}$ tumors. Expression above and below this line is considered present and absent, respectively.

(I and J) Kaplan-Meier curves for OS based on CDKN2A mRNA expression status in patients without *CDKN2A* homozygous deletion in the entire REMBRANDT and MSKCC cohorts (I) and *IDH*WT alone (J).

(K) Kaplan-Meier curve for OS based on p16 status in patients without *CDKN2A* homozygous deletion in the MSKCC cohort.

(L) Kaplan-Meier plot for OS based on 9p loss and CDKN2A mRNA expression.

(M) Kaplan-Meier curves showing survival outcome in either *IDH*WT (left panel) or *IDH* mutant patients (right panel) according to chromosome 10 copy number status in TCGA LGG cohort.

(N) Kaplan-Meier curves showing survival outcomes for patients with genetic alterations in TCGA LGG *IDH*WT subtype.

LGG, lower grade glioma; ns, not significant; SCNA, somatic copy number alteration; OS, overall survival; amp, amplification; del, deletion; Inf, infinite; HR, hazard ratio; WT, wild-type; mut, mutant; IHC, immunohistochemistry. *, p<0.05; **, p<0.01; ***, p<0.001; ****, p<0.0001; ns, not significant (Kaplan-Meier, log-rank test; CDKN2A mRNA expression scatterplots, Student's t-test).

Table S6, related to Figure 6. Pairwise associations between *CDKN2A* homozygous deletion and significant covariates. Provided as an Excel file.

Table S7, related to Figure 6. CDKN2A mRNA and p14/p16 protein expression in cancer cell lines from the Cancer Cell Line Encyclopedia. Provided as an Excel file.