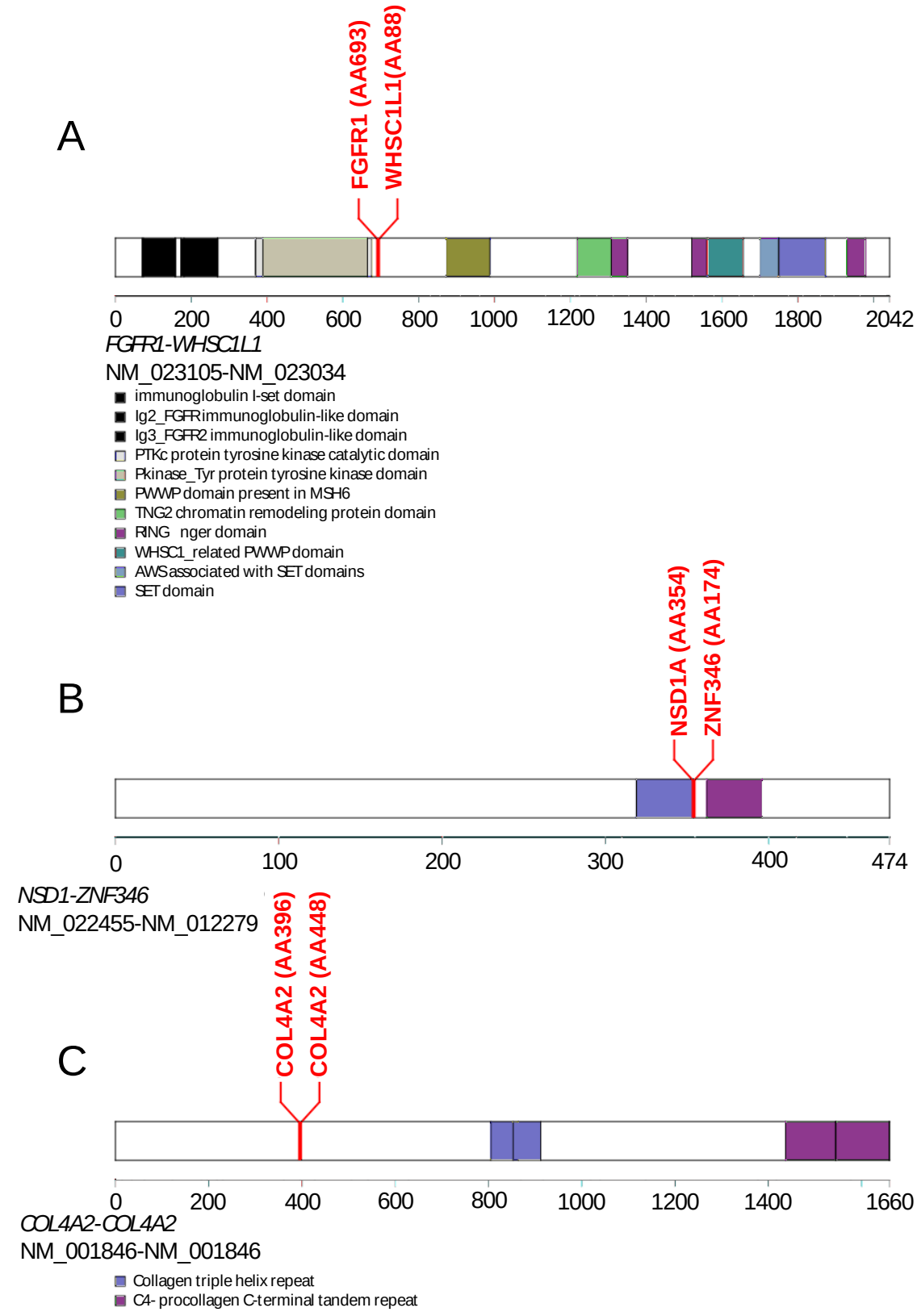
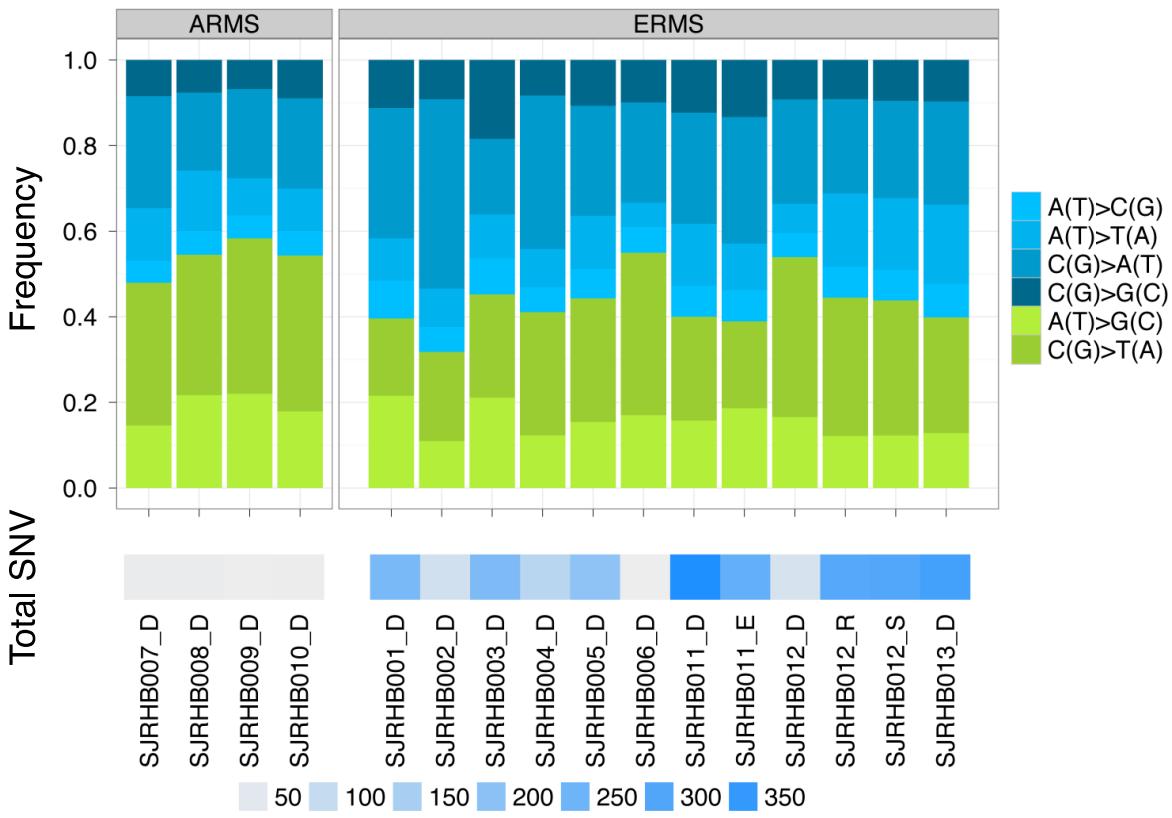
SUPPLEMENTAL DATA



D


ARMS
ERMS
Frequency
1.0
0.8
0.6
0.4
0.2
0.0
A(T)>C(G)
A(T)>T(A)
C(G)>A(T)
C(G)>G(C)
A(T)>G(C)
C(G)>T(A)
Total SNV
SJRHB007_D
SJRHB008_D
SJRHB009_D
SJRHB010_D
SJRHB001_D
SJRHB002_D
SJRHB003_D
SJRHB004_D
SJRHB005_D
SJRHB006_D
SJRHB011_D
SJRHB011_E
SJRHB012_D
SJRHB012_R
SJRHB012_S
SJRHB013_D
50
100
150
200
250
300
350

E

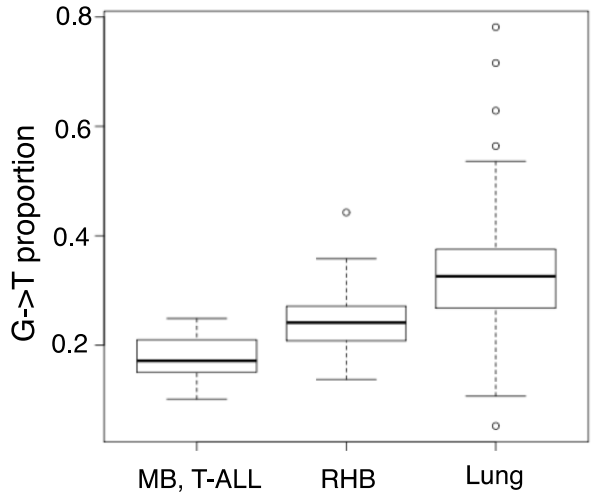

G->T proportion
0.8
0.6
0.4
0.2
MB, T-ALL
RHB
Lung

Figure S1 related to Figure 1. Map of 3 fusion genes that were found to be expressed in the discovery cohort for rhabdomyosarcoma tumors and mutation spectrum.

(A) The breakpoint identified in SJRHB001 for FGFR1-WHSC1L1 is shown in red. (**B)** The breakpoint identified in SJRHB011 for the NSD1-ZNF346 fusion is shown in red. (**C)** The COL4A2 has a deletion that leads to skipping of an exon in SJRHB002. All data were presented had concordance between WGS of the genomic DNA and transriptome sequencing of the corresponding RNA for each sample. (**D)** The mutation spectrum of rhabdomyosarcoma normalized to the total for each sample in the discovery cohort shows elevated G→T transversions. (**E)** Comparison of G→T transversions in rhabdomyosarcoma (n=19; mean = 0.249) to lung cancer (n=172; mean = 0.327), pediatric MB and T-ALL combined (n=48; mean 0.178). The rhabdomyosarcoma G→T transversion proportion was significantly higher than for pediatric medulloblastoma + T-ALL (p=0.0003, unpaired t-test).

Table S1 related to Figure 1. Clinical features of discovery and recurrency cohorts.

Provided as separate file.

Table S2 related to Figure 1. Coverage data for DNA and RNA sequencing.

Provided as separate file.

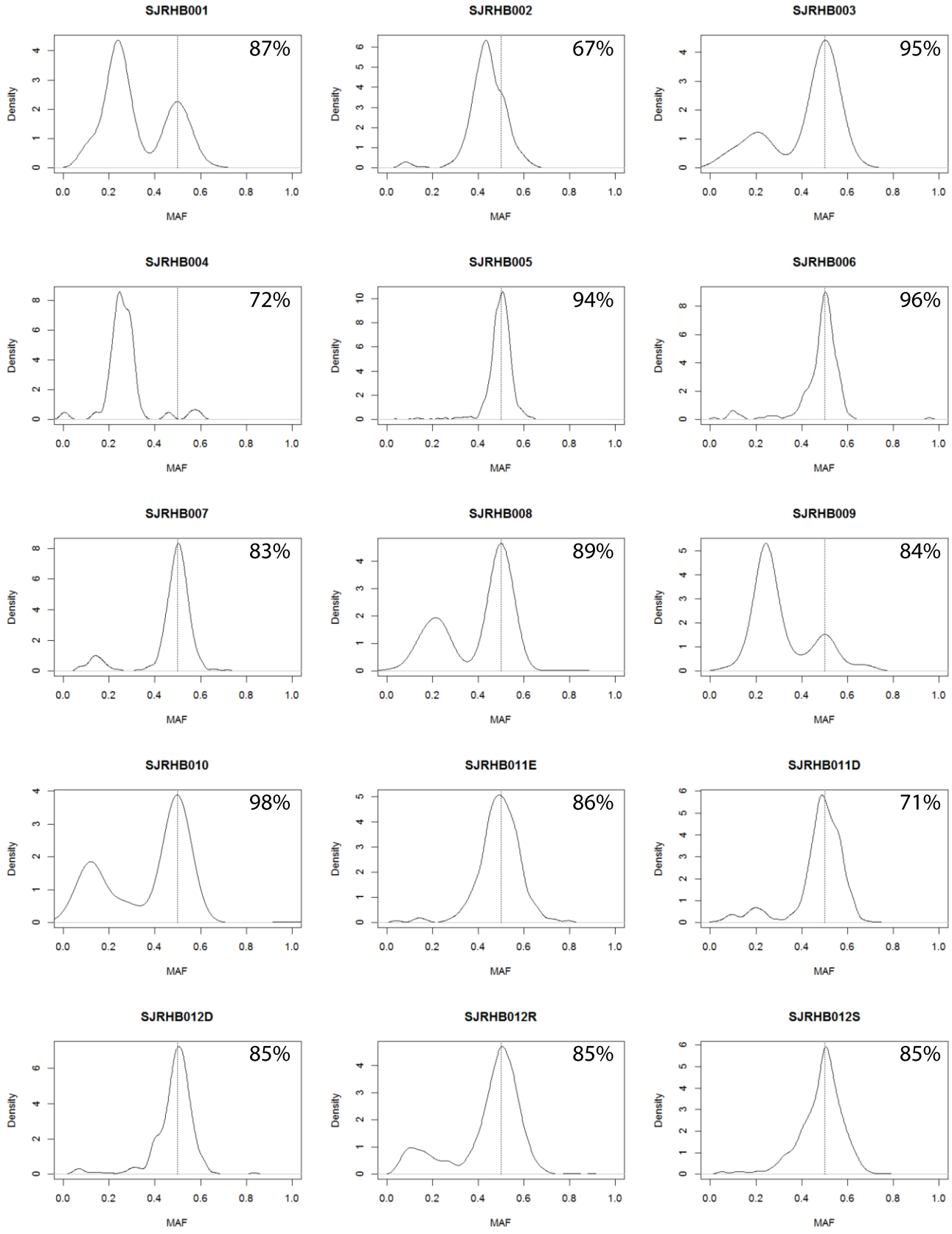
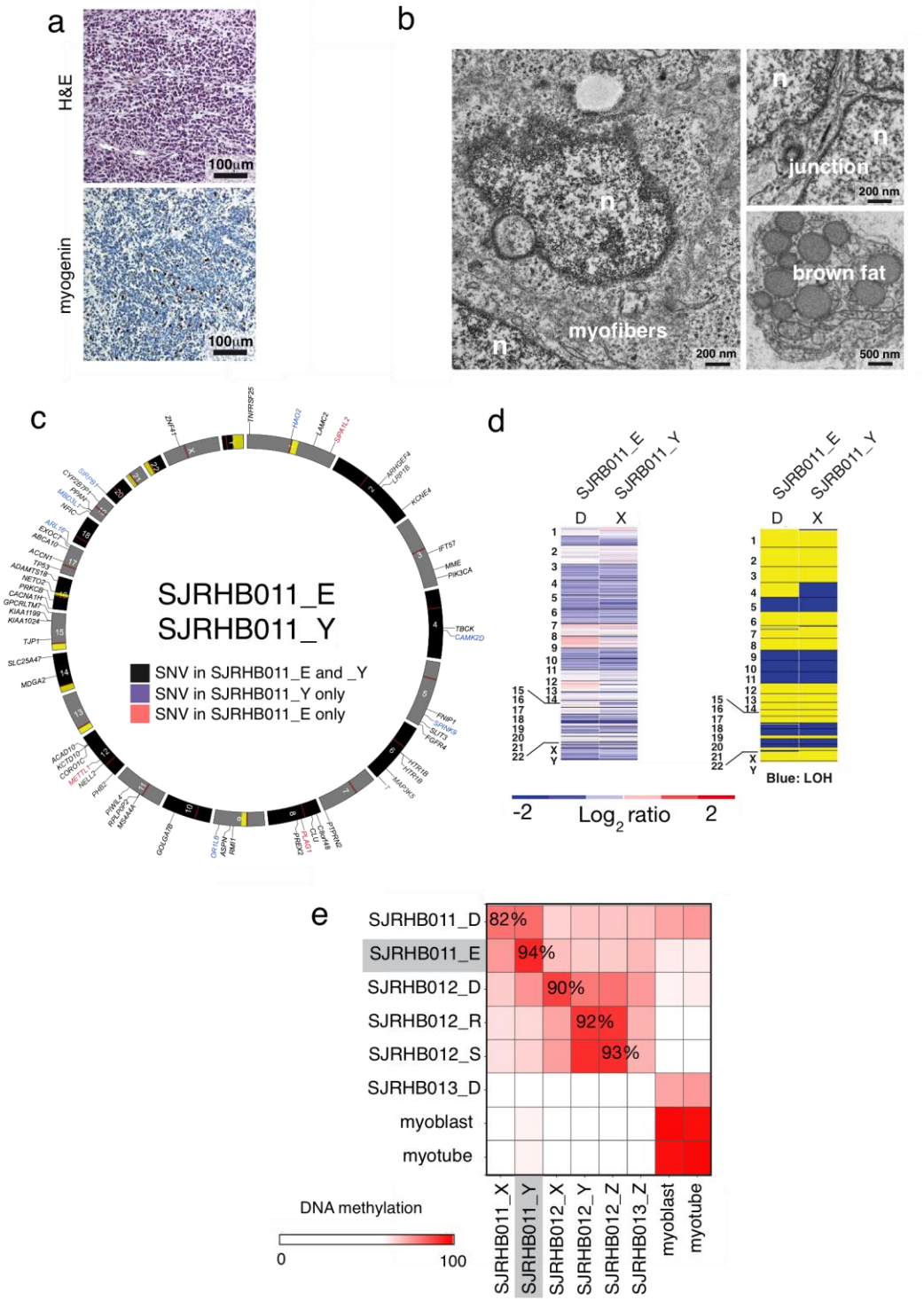
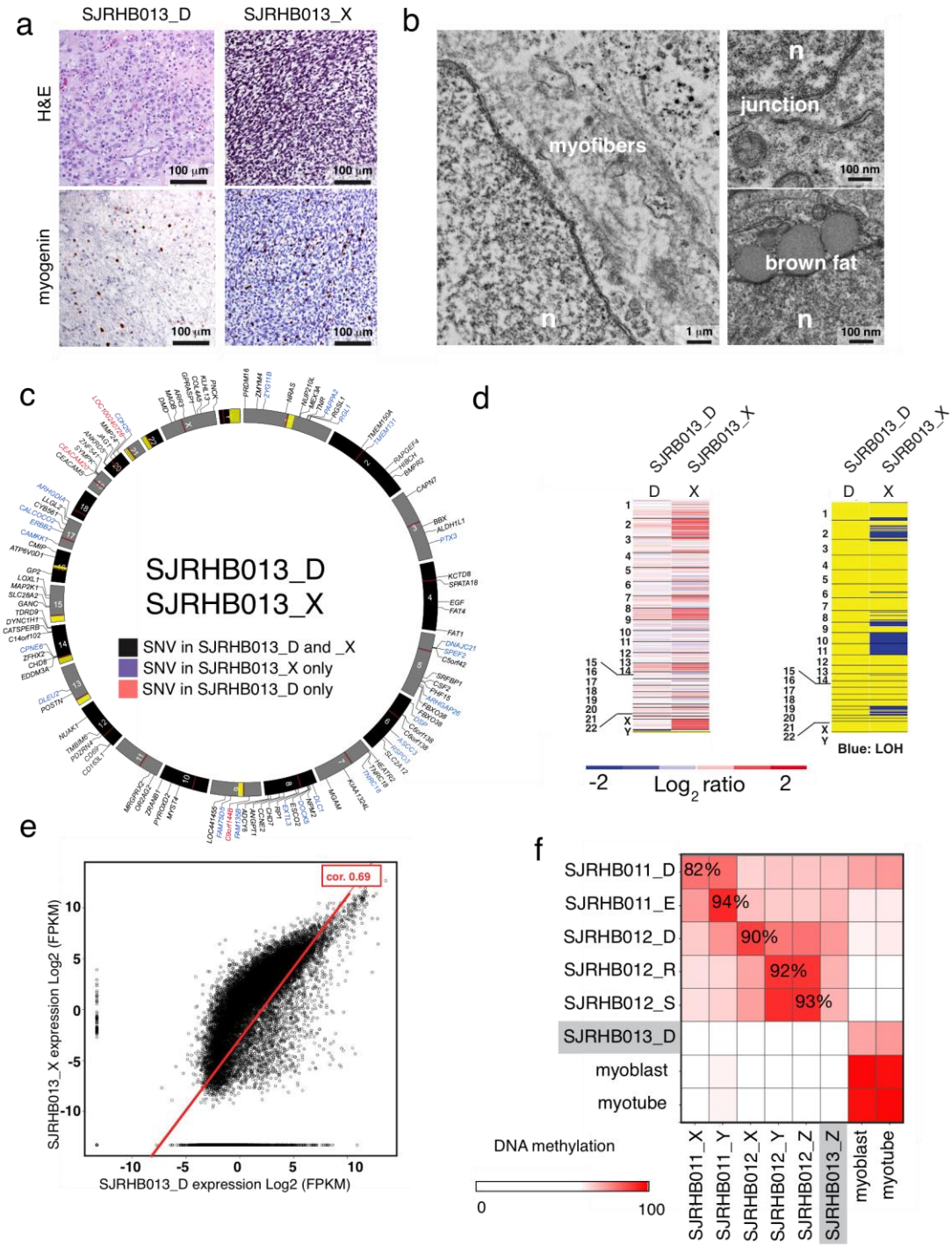

SJRHB001
87%
SJRHB002
67%
SJRHB003
95%
SJRHB004
72%
SJRHB005
94%
SJRHB006
96%
SJRHB007
83%
SJRHB008
89%
SJRHB009
84%
SJRHB010
98%
SJRHB011E
86%
SJRHB011D
71%
SJRHB012D
85%
SJRHB012R
85%
SJRHB012S
85%
Density
MAF

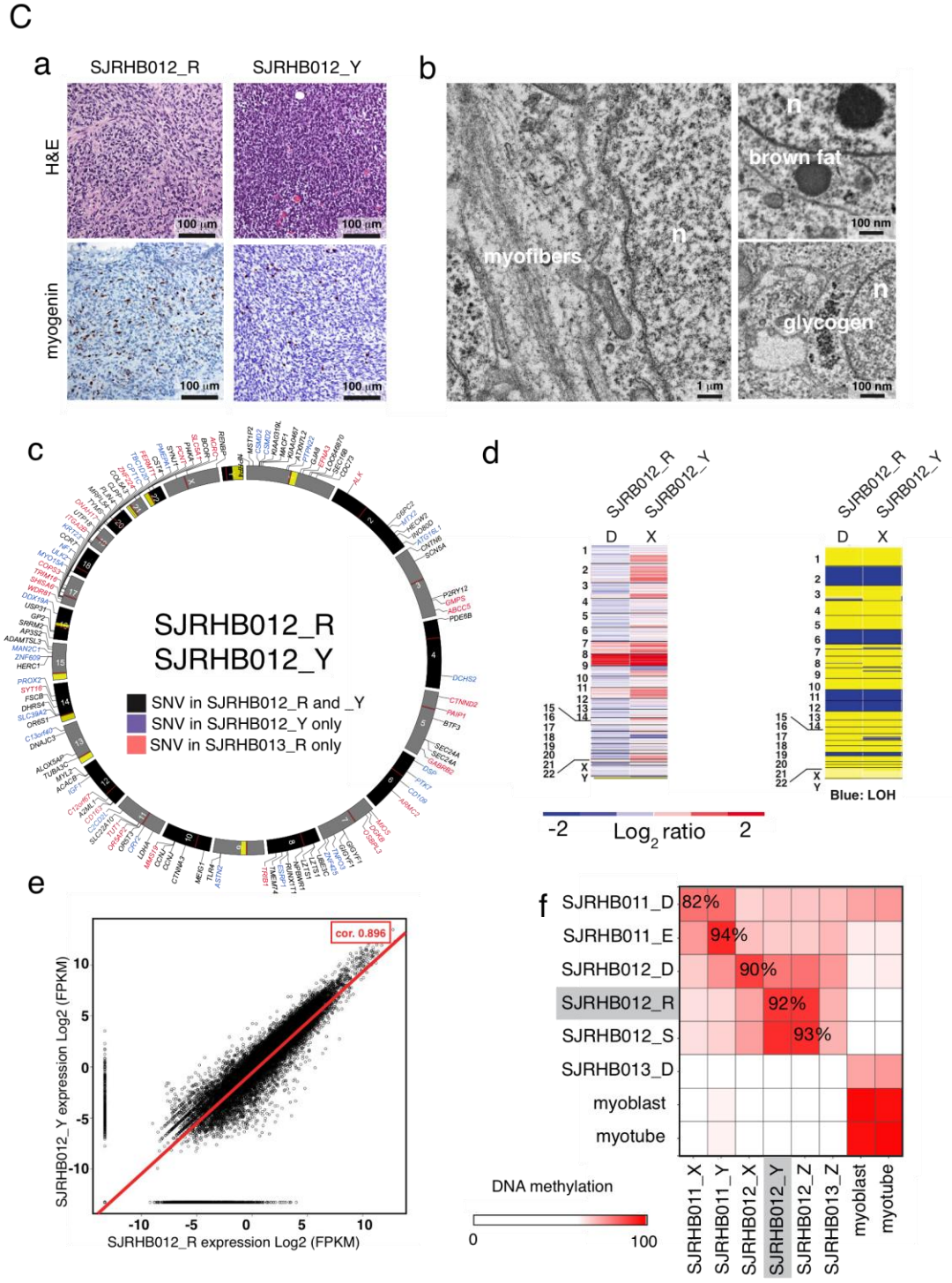
Figure S2 related to Figures 2. Tumor purity and tumor heterogeneity of rhabdomyosarcoma. Tumor purity adjusted mutant allele frequency (MAF) for 15 of the 16 samples in our discovery cohort. SJRHB013 was excluded from this analysis due to low tumor purity. The actual tumor purity that was used to adjust the MAF is shown as a percentage for each tumor.

A

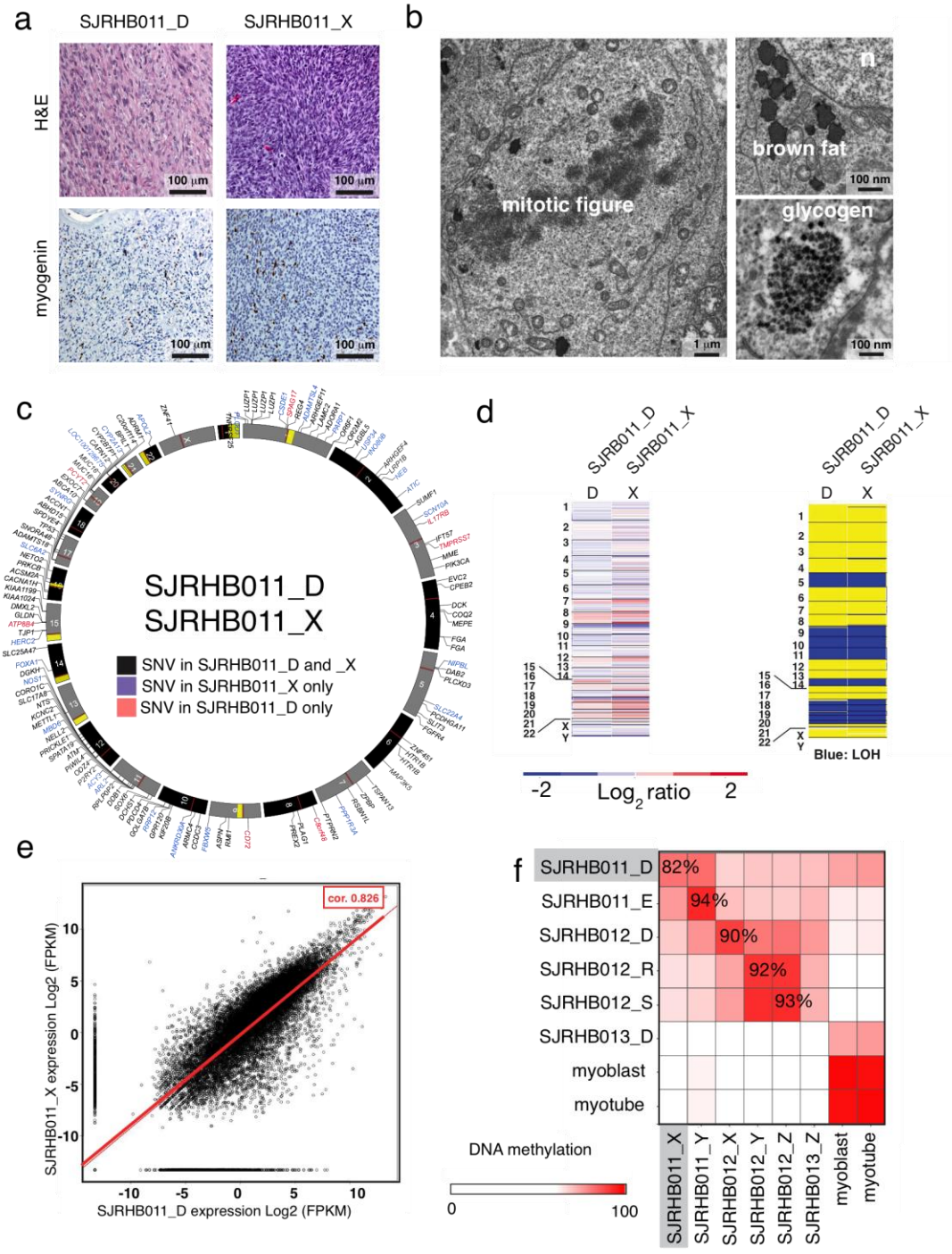

a
H&E
100μm
myogenin
100μm
b
n
myofibers
200 nm
n
n
junction
200 nm
brown fat
500 nm
c
SJRHB011_E
SJRHB011_Y
SNV in SJRHB011_E and _Y
SNV in SJRHB011_Y only
SNV in SJRHB011_E only
d
SJRB011_E
SJRB011_Y
D
X
Blue: LOH
-2
Log2 ratio
2
e
SJRHB011_D
SJRHB011_E
SJRHB012_D
SJRHB012_R
SJRHB012_S
SJRHB013_D
myoblast
myotube
82%
94%
90%
92%
93%
SJRHB011_X
SJRHB011_Y
SJRHB012_X
SJRHB012_Y
SJRHB012_Z
SJRHB013_Z
myoblast
myotube
DNA methylation
0
100

B

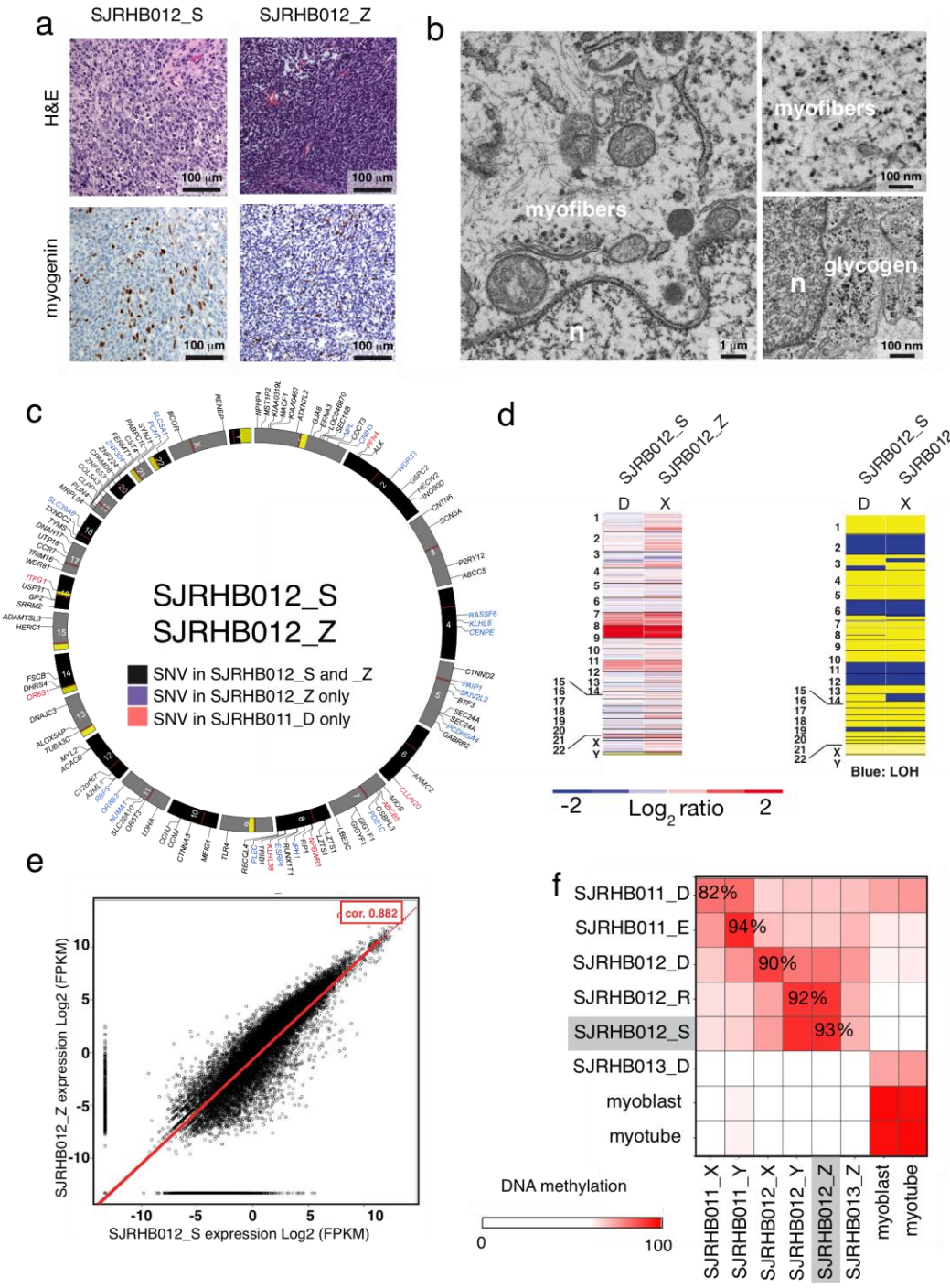

a
SJRHB013_D
SJRHB013_X
H&E
myogenin
100 μm
b
myofibers
n
junction
100 nm
brown fat
1 μm
c
SJRHB013_D
SJRHB013_X
SNV in SJRHB013_D and _X
SNV in SJRHB013_X only
SNV in SJRHB013_D only
d
SJRHB013_D
SJRHB013_X
D
X
Blue: LOH
-2
Log2 ratio
2
e
cor. 0.69
SJRHB013_X expression Log2 (FPKM)
SJRHB013_D expression Log2 (FPKM)
f
SJRHB011_D
SJRHB011_E
SJRHB012_D
SJRHB012_R
SJRHB012_S
SJRHB013_D
myoblast
myotube
82%
94%
90%
92%
93%
DNA methylation
0
100
SJRHB011_X
SJRHB011_Y
SJRHB012_X
SJRHB012_Y
SJRHB012_Z
SJRHB013_Z
myoblast
myotube


C
a
SJRHB012_R
SJRHB012_Y
H&E
myogenin
100 μm
b
myofibers
n
brown fat
glycogen
100 nm
1 μm
c
SJRHB012_R
SJRHB012_Y
SNV in SJRHB012_R and _Y
SNV in SJRHB012_Y only
SNV in SJRHB013_R only
d
D
X
-2
Log2 ratio
2
Blue: LOH
e
cor. 0.896
SJRHB012_Y expression Log2 (FPKM)
SJRHB012_R expression Log2 (FPKM)
f
SJRHB011_D
SJRHB011_E
SJRHB012_D
SJRHB012_R
SJRHB012_S
SJRHB013_D
myoblast
myotube
82%
94%
90%
92%
93%
DNA methylation
0
100
SJRHB011_X
SJRHB011_Y
SJRHB012_X
SJRHB012_Y
SJRHB012_Z
SJRHB013_Z

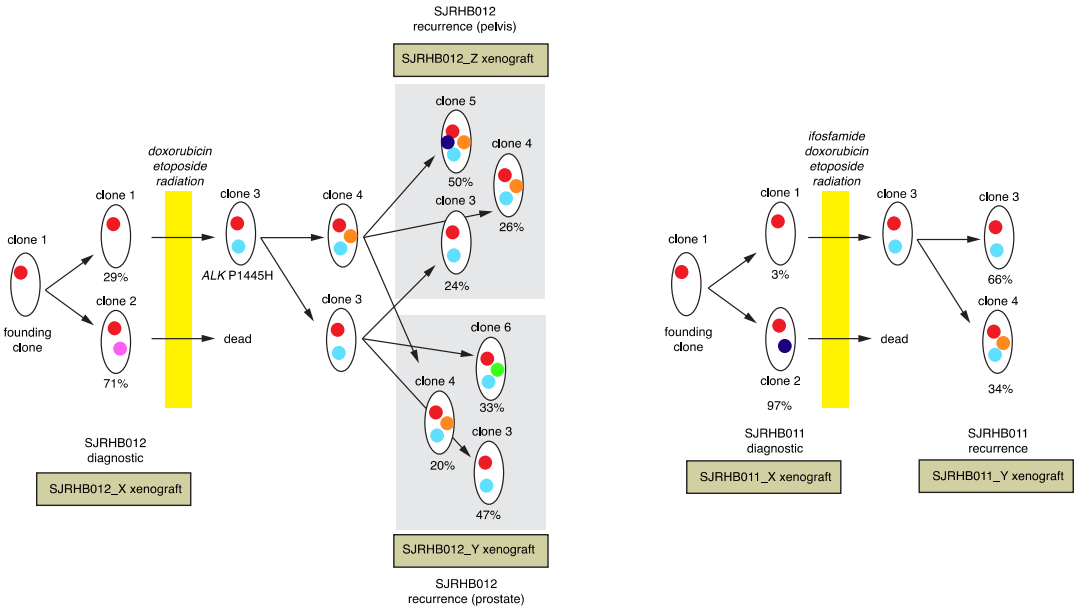
D


a
SJRHB011_D
SJRHB011_X
H&E
myogenin
100 μm
b
mitotic figure
1 μm
n
brown fat
100 nm
glycogen
100 nm
c
SJRHB011_D
SJRHB011_X
SNV in SJRHB011_D and _X
SNV in SJRHB011_X only
SNV in SJRHB011_D only
d
SJRHB011_D
SJRHB011_X
D
X
-2
Log2 ratio
2
Blue: LOH
e
cor. 0.826
SJRHB011_X expression Log2 (FPKM)
SJRHB011_D expression Log2 (FPKM)
f
SJRHB011_D
SJRHB011_E
SJRHB012_D
SJRHB012_R
SJRHB012_S
SJRHB013_D
myoblast
myotube
82%
94%
90%
92%
93%
DNA methylation
0
100
SJRHB011_X
SJRHB011_Y
SJRHB012_X
SJRHB012_Y
SJRHB012_Z
SJRHB013_Z
myoblast
myotube

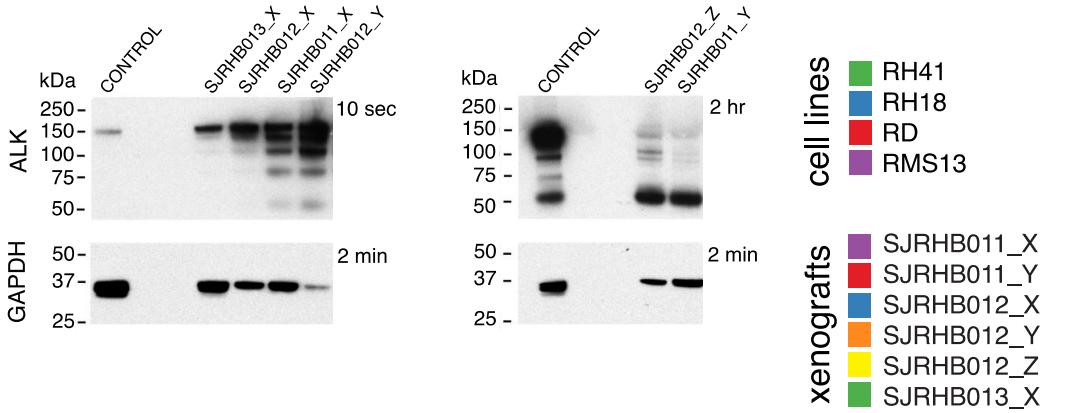
E


a
SJRHB012_S
SJRHB012_Z
H&E
myogenin
100 μm
b
myofibers
n
1 μm
100 nm
glycogen
c
SJRHB012_S
SJRHB012_Z
SNV in SJRHB012_S and _Z
SNV in SJRHB012_Z only
SNV in SJRHB011_D only
d
D
X
-2
Log2 ratio
2
Blue: LOH
e
cor. 0.882
SJRHB012_Z expression Log2 (FPKM)
SJRHB012_S expression Log2 (FPKM)
f
SJRHB011_D
SJRHB011_E
SJRHB012_D
SJRHB012_R
SJRHB012_S
SJRHB013_D
myoblast
myotube
82%
94%
90%
92%
93%
DNA methylation
0
100

F


SJRHB012
recurrence (pelvis)
SJRHB012_Z xenograft
clone 5
clone 4
50%
clone 3
26%
24%
doxorubicin
etoposide
radiation
clone 1
clone 1
29%
clone 2
71%
founding
clone
clone 3
ALK P1445H
clone 4
clone 3
dead
clone 6
clone 4
33%
clone 3
20%
47%
SJRHB012
diagnostic
SJRHB012_X xenograft
SJRHB012_Y xenograft
SJRHB012
recurrence (prostate)
ifosfamide
doxorubicin
etoposide
radiation
clone 1
clone 1
3%
clone 2
97%
founding
clone
clone 3
dead
clone 3
66%
clone 4
34%
SJRHB011
diagnostic
SJRHB011_X xenograft
SJRHB011
recurrence
SJRHB011_Y xenograft

G


kDa
CONTROL
SJRHB013_X
SJRHB012_X
SJRHB011_X
SJRHB012_Y
ALK
250
150
100
75
50
10 sec
GAPDH
50
37
25
2 min
kDa
CONTROL
SJRHB012_Z
SJRHB011_Y
250
150
100
75
50
2 hr
50
37
25
2 min
cell lines
RH41
RH18
RD
RMS13
xenografts
SJRHB011_X
SJRHB011_Y
SJRHB012_X
SJRHB012_Y
SJRHB012_Z
SJRHB013_X

H

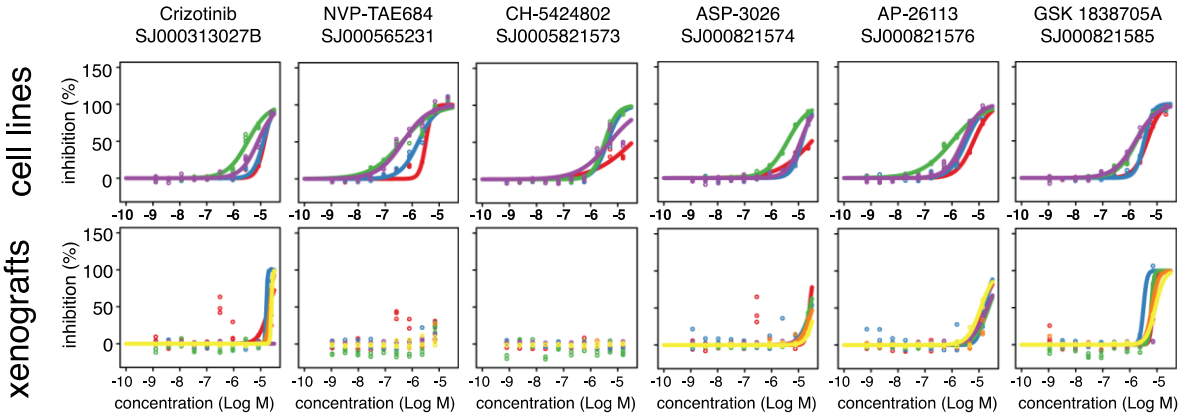
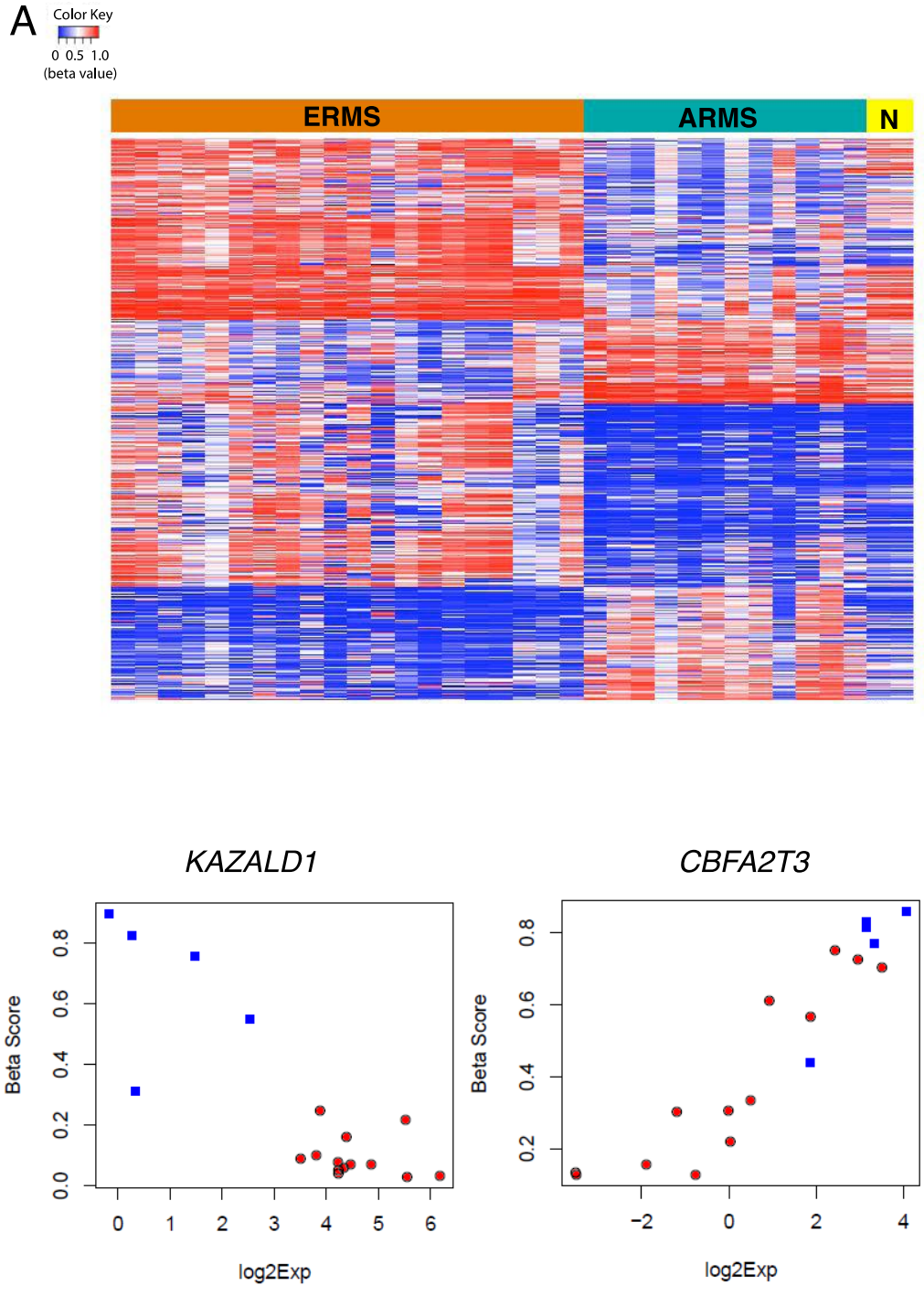

Crizotinib
SJ000313027B
NVP-TAE684
SJ000565231
CH-5424802
SJ0005821573
ASP-3026
SJ000821574
AP-26113
SJ000821576
GSK 1838705A
SJ000821585
cell lines
xenografts
inhibition (%)
150
100
50
0
-10 -9 -8 -7 -6 -5
concentration (Log M)

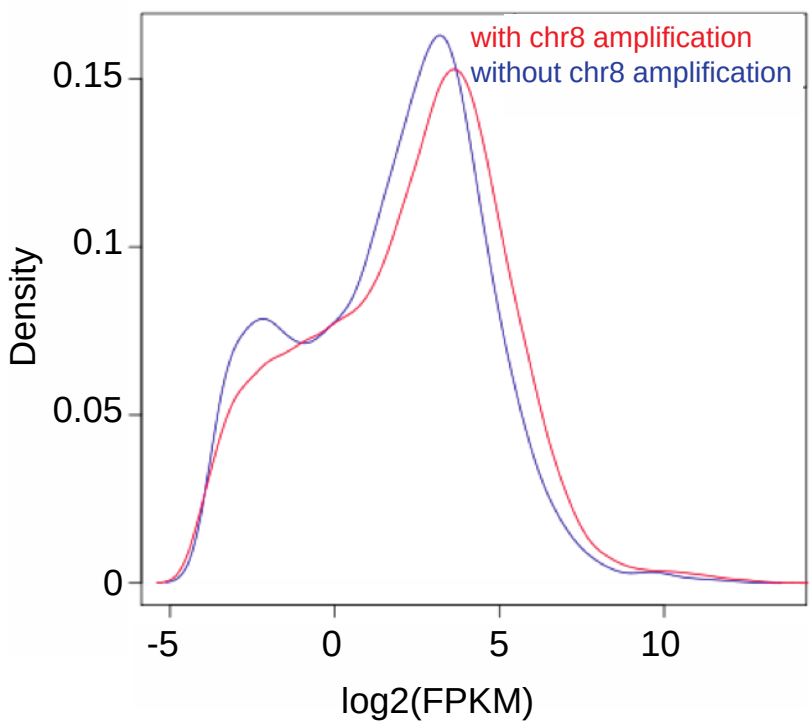
Figure S3 related to Figure 4. Analysis of orthotopic xenografts. (A) Characterization of xenograft SJRHB011_Y from sample SJRHB011_E. (**a)** Representative H&E and myogenin immunohistochemistry for the xenografts. (**b)** Transmission electron micrographs of the xenografts showing features of rhabdomyosarcoma including myofibers, junctions and brown fat deposits. (**c)** Circos plot of exonic SNVs for the xenograft and diagnostic tumor pairs. Gene names in black contain SNVs found in the primary and xenograft samples and those shown in blue are unique to the xenograft and those shown in red are lost in the xenograft. (**d)** SNP 6.0 analysis showing copy number changes and LOH for the matched primary and xenograft samples with red showing gain and blue showing loss for copy number and blue showing LOH for the lower panel. (**e)** Heatmap of DNA methylation analysis for the matched diagnostic and xenograft pairs using the Illumina Infinium Human Methylation 450 BeadChip platform. (**B)** Characterization of xenograft SJRHB013_X from sample SJRHB013_D. (**a)** Representative H&E and myogenin immunohistochemistry for the xenograft and the primary tumor. (**b)** Transmission electron micrographs of the xenografts showing features of rhabdomyosarcoma including myofibers, junctions and brown fat deposits. (**c)** Circos plot of exonic SNVs for the xenograft and diagnostic tumor pairs. Gene names in black contain SNVs found in the primary and xenograft samples and those shown in blue are unique to the xenograft and those shown in red are lost in the xenograft. (**d)** SNP 6.0 analysis showing copy number changes and LOH for the matched primary and xenograft samples with red showing gain and blue showing loss for copy number and blue showing LOH for the lower panel. (**e)** Plot of gene expression data from the xenograft and primary tumor from the RNA-Seq data. The correlation line and coefficient is indicated in red. (**f)** Heatmap of DNA methylation analysis for the matched diagnostic and xenograft pairs using the Illumina Infinium Human Methylation 450 BeadChip platform. (**C)**

Characterization of xenograft SJRHB012_Y from sample SJRHB012_R. (**a)** Representative H&E and myogenin immunohistochemistry for the xenograft and the primary tumor. (**b)** Transmission electron micrographs of the xenografts showing features of rhabdomyosarcoma including myofibers, glycogen and brown fat deposits. (**c)** Circos plot of exonic SNVs for the xenograft and diagnostic tumor pairs. Gene names in black contain SNVs found in the primary and xenograft samples and those shown in blue are unique to the xenograft and those shown in red are lost in the xenograft. (**d)** SNP 6.0 analysis showing copy number changes and LOH for the matched primary and xenograft samples with red showing gain and blue showing loss for copy number and blue showing LOH for the lower panel. (**e)** Plot of gene expression data from the xenograft and primary tumor from the RNA-Seq data. The correlation line and coefficient is indicated in red. (**f)** Heatmap of DNA methylation analysis for the matched diagnostic and xenograft pairs using the Illumina Infinium Human Methylation 450 BeadChip platform. (**D)** Characterization of xenograft SJRHB011_X from sample SJRHB011_D. (**a)** Representative H&E and myogenin immunohistochemistry for the xenograft and the primary tumor. (**b)** Transmission electron micrographs of the xenografts showing features of rhabdomyosarcoma including mitotic figure, glycogen and brown fat deposits. (**c)** Circos plot of exonic SNVs for the xenograft and diagnostic tumor pairs. Gene names in black contain SNVs found in the primary and xenograft samples and those shown in blue are unique to the xenograft and those shown in red are lost in the xenograft. (**d)** SNP 6.0 analysis showing copy number changes and LOH for the matched primary and xenograft samples with red showing gain and blue showing loss for copy number and blue showing LOH for the lower panel. (**e)** Plot of gene expression data from the xenograft and primary tumor from the RNA-Seq data. The correlation line and coefficient is indicated in red. (**f)** Heatmap of DNA methylation analysis for the matched diagnostic and

xenograft pairs using the Illumina Infinium Human Methylation 450 BeadChip platform. (**E)** Characterization of xenograft SJRHB012_Z from sample SJRHB012_S. (**a)** Representative H&E and myogenin immunohistochemistry for the xenograft and the primary tumor. (**b)** Transmission electron micrographs of the xenografts showing features of rhabdomyosarcoma including myofibers and glycogen. (**c)** Circos plot of exonic SNVs for the xenograft and diagnostic tumor pairs. Gene names in black contain SNVs found in the primary and xenograft samples and those shown in blue are unique to the xenograft and those shown in red are lost in the xenograft. (**d)** SNP 6.0 analysis showing copy number changes and LOH for the matched primary and xenograft samples with red showing gain and blue showing loss for copy number and blue showing LOH for the lower panel. (**e)** Plot of gene expression data from the xenograft and primary tumor from the RNA-Seq data. The correlation line and coefficient is indicated in red. (**f)** Heatmap of DNA methylation analysis for the matched diagnostic and xenograft pairs using the Illumina Infinium Human Methylation 450 BeadChip platform. (**F)** Clonal evolution of SJRHB012 and xenografts generated from the diagnostic and two recurrent sites. (**G)** Immunoblot of ALK protein expression from a short exposure (10 seconds) and a long exposure (2 hours). (**H)** Sensitivity of orthotopic xenografts and rhabdomyosarcoma cell lines to 6 different ALK inhibitors in triplicate.


A
Color Key
0 0.5 1.0
(beta value)
ERMS
ARMS
N
KAZALD1
CBFA2T3
Beta Score
log2Exp
0.0
0.2
0.4
0.6
0.8
0
1
2
3
4
5
6
−2
0
2
4

B


with chr8 amplification
without chr8 amplification
Density
0.15
0.1
0.05
0
-5
0
5
10
log2(FPKM)

C

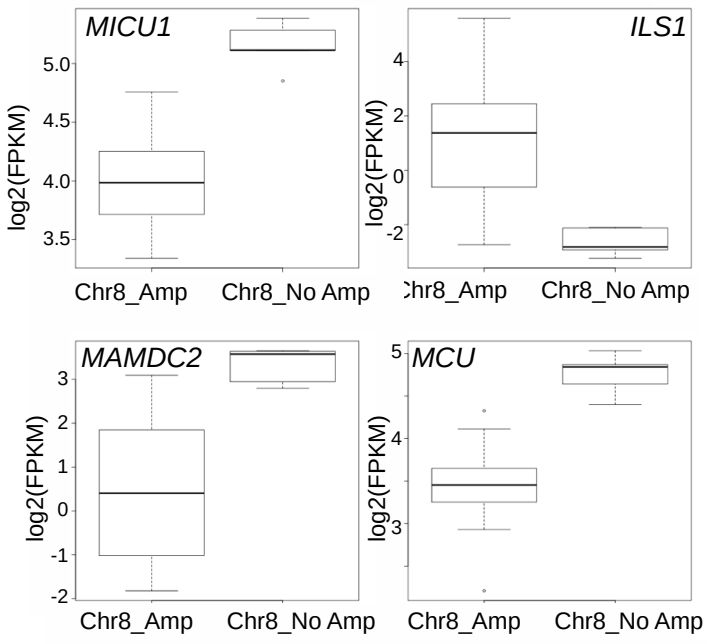

MICU1
ILS1
MAMDC2
MCU
log2(FPKM)
Chr8_Amp
Chr8_No Amp

Figure S4 related to Figure 5. DNA methylation and RNA-seq analysis of rhabdomyosarcoma. (A) Heatmap of beta-values for DNA methylation analysis showing distinct patterns of DNA methylation for ERMS, ARMS and normal myotubes and myoblasts. Integration of changes in DNA methylation (beta-value) and gene expression (RNA-seq) identified a series of genes with significant correlations between gene expression and DNA methylation. Two representative genes are shown in the lower portion of panel (A). The red data points are ERMS and the blue data points are ARMS. (**B)** Chromosome 8 gene expression measured with RNA-seq analysis of ERMS tumors with chromosome 8 gain (red line) was compared to those without chromosome 8 gain (blue line). (**C)** A whole genome gene-wise comparison of expression was performed for the two groups. Boxplots of gene expression for 4 genes with at least 2-fold difference between the two groups are shown (chromosome 8 gain and chromosome 8 normal).

Table S3 related to Figure 5. Integrated RNA and DNA methylation analysis.

Provided as separate file.

Table S4 related to Figure 5. Hh and WNT pathway analysis.

Provided as separate file.

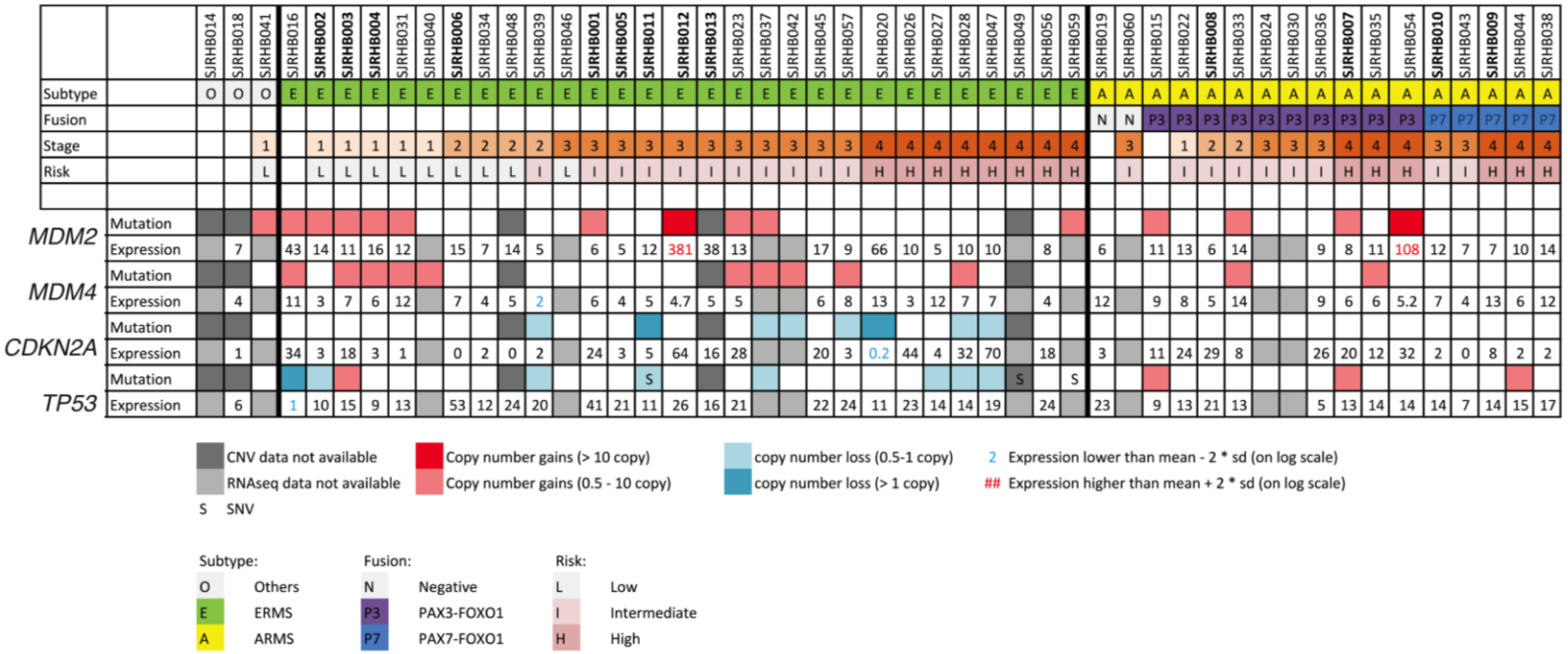
Table S5 related to Figure 5. Validation of genome sequence data.

Provided as separate file.

Table S6 related to Figure 5. Analysis of muscle enriched genes.

Provided as separate file.

A


SJRHB014 SJRHB018 SJRHB041 SJRHB016 SJRHB002 SJRHB003 SJRHB004 SJRHB031 SJRHB040 SJRHB006 SJRHB034 SJRHB048 SJRHB039 SJRHB046 SJRHB001 SJRHB005 SJRHB011 SJRHB012 SJRHB013 SJRHB023 SJRHB037 SJRHB042 SJRHB045 SJRHB057 SJRHB020 SJRHB026 SJRHB027 SJRHB028 SJRHB047 SJRHB049 SJRHB056 SJRHB059 SJRHB019 SJRHB060 SJRHB015 SJRHB022 SJRHB008 SJRHB033 SJRHB024 SJRHB030 SJRHB036 SJRHB007 SJRHB035 SJRHB054 SJRHB010 SJRHB043 SJRHB009 SJRHB044 SJRHB038
Subtype
Fusion
Stage
Risk
MDM2
Mutation
Expression
MDM4
Mutation
Expression
CDKN2A
Mutation
Expression
TP53
Mutation
Expression
CNV data not available
RNAseq data not available
S SNV
Copy number gains (> 10 copy)
Copy number gains (0.5 - 10 copy)
copy number loss (0.5-1 copy)
copy number loss (> 1 copy)
2 Expression lower than mean - 2 * sd (on log scale)
Expression higher than mean + 2 * sd (on log scale)
Subtype:
O Others
E ERMS
A ARMS
Fusion:
N Negative
P3 PAX3-FOXO1
P7 PAX7-FOXO1
Risk:
L Low
I Intermediate
H High

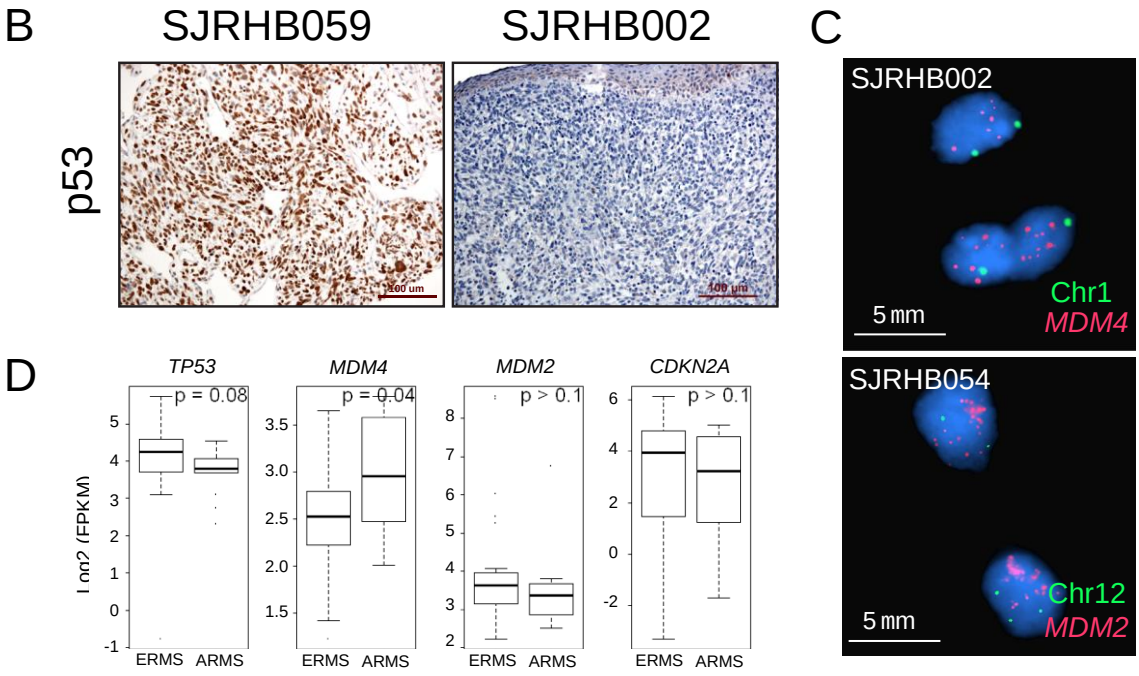


Figure S5, related to Figure 6. Analysis of p53 pathway in rhabdomyosarcoma. (A) Heatmap of p53 pathway including expression data and mutation data (SNV, CNV, indel, SV, LOH). Samples SJRHB003, SJRHB011, SJRHB049 and SJRHB059 have missense mutations in *TP53* and nuclear accumulation of p53 protein. SJRHB016 is likely to have a homozygous deletion of TP53 with reduced expression. SJRHB012 and SJRHB054 have amplification of MDM2 with increased expression and SJRHB020 has deletion of *CDKN2A* with reduced expression. (**B)** Representative p53 IHC showing nuclear accumulation in SJRHB059 with a missense mutation compared to a control sample (SJRHB002) with copy number gains in *MDM4*. (**C)** Representative FISH images of tumors with *MDM2* amplification and *MDM4* gain. (**D)** Boxplots of expression of *TP53, MDM4, MDM2* and *CDKN2A* as measured by RNA-seq in ERMS vs. ARMS.

Table S7 related to Figure 6. Mutation and expression of p53 pathway genes.

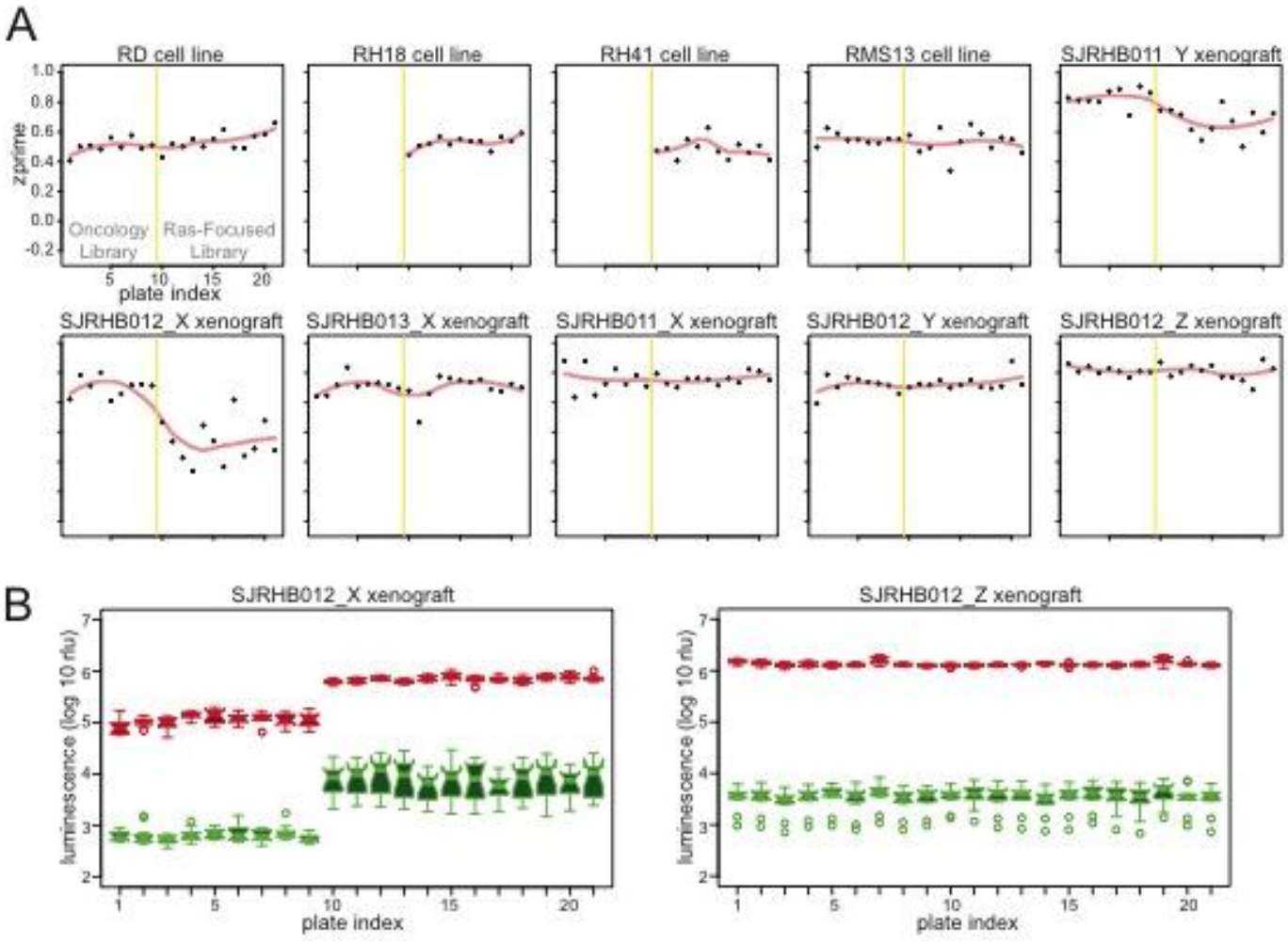
Sample	TP53		MDM4		MDM2		CDKN2A	
	CNA	FPKM	CNA	FPKM	CNA	FPKM	CNA	FPKM
SJRHB001	-	40.9	-	5.79	0.7-copy gain	6.20	-	24.2
SJRHB002	0.7-copy loss	9.87	-	3.03	0.6-copy gain	14.3	-	3.19
SJRHB003	2-copy gain	15.4	0.9-copy gain	6.65	1.5-copy gain	10.7	-	18.4
SJRHB004	-	8.63	1.1-copy gain	5.94	0.8-copy gain	15.7	-	3.10
SJRHB005	-	20.9	-	3.99	-	5.37	-	2.78
SJRHB006	-	53.2	-	6.93	-	15.0	-	0.10
SJRHB007	0.8-copy gain	13.2	-	6.40	0.7-copy gain	8.38	-	19.5
SJRHB008	-	21.4	-	5.02	-	5.74	0.4 –copy loss	28.9
SJRHB009	-	18.3	-	12.6	-	7.35	-	8.30
SJRHB010	-	13.7	-	7.20	-	12.2	-	2.34
SJRHB011_E	1-copy loss	NA	-	NA	1.5-copy gain	NA	*1-copy loss*	NA
SJRHB011_D	0.8-copy loss	11.4	-	5.40	-	12.4	*1.5-copy loss*	4.94
SJRHB012_D	-	19.0	-	6.96	*>100-copy gain*	12.5	-	15.3
SJRHB012_R	-	30.5	-	4.70	*100-copy gain*	365	-	27.8
SJRHB012_S	-	25.8	-	4.67	*100-copy gain*	381	-	63.8
SJRHB013	Low purity	15.7	Low purity	5.07	Low purity	38.4	Low purity	16.1
SJRHB015	0.6-copy gain	8.56	-	8.70	0.5-copy gain	10.7	0.45-copy loss	10.5
SJRHB016	*1.3-copy loss*	0.58	1.2-copy gain	10.5	6-copy gain	43.0	-	34.4
SJRHB018	-	6.21	-	4.32	-	7.47	-	0.61
SJRHB019	-	23.2	-	12.0	-	6.10	-	2.77
SJRHB020	-	10.5	-	12.6	-	65.6	*1.8-copy loss*	0.18
SJRHB022	-	12.8	-	8.36	-	12.8	-	23.5
SJRHB023	-	20.5	0.7-copy gain	4.85	1.6-copy gain	13.2	-	28.1
SJRHB024	-	NA	-	NA	-	NA	-	NA
SJRHB026	-	22.5	-	2.68	0.4-copy gain	9.58	-	44.1
SJRHB027	0.9-copy loss	13.6	-	11.7	-	4.66	0.4-copy loss	3.63
SJRHB028	0.5-copy loss	13.6	0.8-copy gain	6.91	-	10.2	0.5-copy loss	31.6
SJRHB030	-	NA	-	NA	-	NA	-	NA
SJRHB031	-	13.0	2.5-copy gain	11.9	0.9-copy gain	11.9	-	1.05
SJRHB033	-	12.8	0.9-copy gain	14.0	0.6-copy gain	14.0	-	7.52
SJRHB034	-	11.7	-	3.59	-	7.25	-	1.56
SJRHB035	-	14.0	1-copy gain	5.88	-	11.1	-	12.0
SJRHB036	-	4.99	-	8.82	-	8.88	-	26.1
SJRHB037	0.5-copy loss	NA	1.8-copy gain	NA	1.2-copy gain	NA	0.6-copy loss	NA
SJRHB039	0.6-copy loss	20.1	-	2.33	-	5.11	0.7-copy loss	2.11
SJRHB040	-	NA	0.7-copy gain	NA	-	NA	-	NA
SJRHB041	-	NA	0.35-copy loss	NA	0.5-copy gain	NA	0.35-copy loss	NA
SJRHB042	-	NA	5-copy gain	NA	-	NA	*1-copy loss*	NA
SJRHB043	-	6.64	-	4.02	-	7.08	-	0.31

SJRHB044	0.5-copy gain	14.8	-	5.56	-	10.0	-	1.93
SJRHB045	-	22.2	-	6.36	-	16.9	-	20.3
SJRHB046	-	NA	-	NA	-	NA	-	NA
SJRHB047	0.5-copy loss	18.8	-	7.09	-	9.83	0.5-copy loss	70.3
SJRHB054	-	14.2	-	5.18	*14-copy gain*	108	-	32.3
SJRHB056	-	23.9	-	3.63	-	8.18	-	18.0
SJRHB057	-	24.4	1.1-copy gain	7.51	-	8.88	*1-copy loss*	3.45
SJRHB059	-	NA	-	NA	1.1-copy gain	NA	-	NA
SJRHB060	-	NA	-	NA	-	NA	-	NA

ARMS
ERMS
Failed QC
FOCAL CNA

Table S8 related to Figure 6. Metagene analysis.

Provided as separate file.


A
RD cell line
RH18 cell line
RH41 cell line
RMS13 cell line
SJRHB011_Y xenograft
SJRHB012_X xenograft
SJRHB013_X xenograft
SJRHB011_X xenograft
SJRHB012_Y xenograft
SJRHB012_Z xenograft
zprime
Oncology Library
Ras-Focused Library
plate index
B
SJRHB012_X xenograft
SJRHB012_Z xenograft
luminescence (log 10 rlu)
plate index

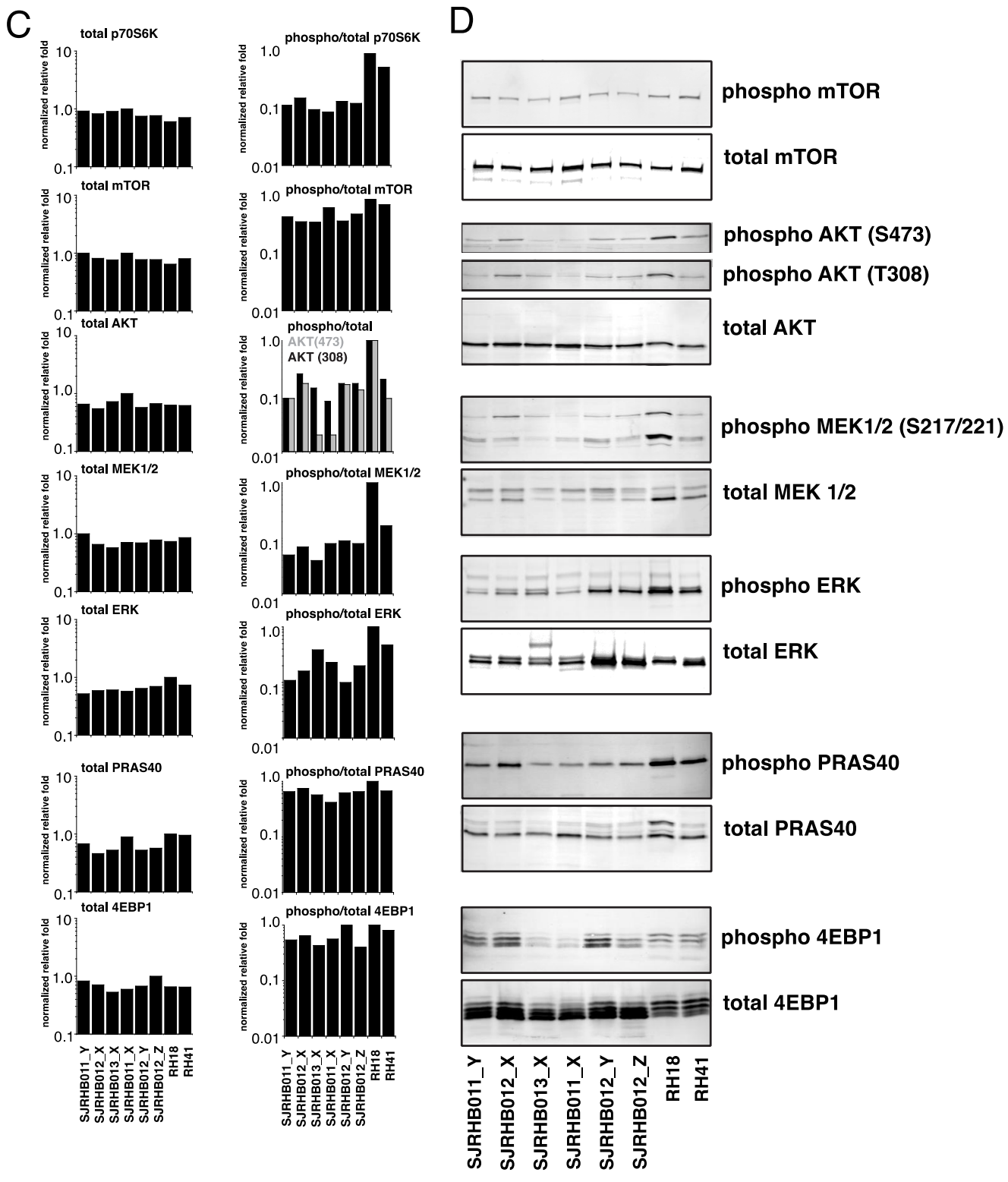

C
total p70S6K
phospho/total p70S6K
total mTOR
phospho/total mTOR
total AKT
phospho/total
AKT(473)
AKT (308)
total MEK1/2
phospho/total MEK1/2
total ERK
phospho/total ERK
total PRAS40
phospho/total PRAS40
total 4EBP1
phospho/total 4EBP1
normalized relative fold
SJRHB011_Y
SJRHB012_X
SJRHB013_X
SJRHB011_X
SJRHB012_Y
SJRHB012_Z
RH18
RH41
D
phospho mTOR
total mTOR
phospho AKT (S473)
phospho AKT (T308)
total AKT
phospho MEK1/2 (S217/221)
total MEK 1/2
phospho ERK
total ERK
phospho PRAS40
total PRAS40
phospho 4EBP1
total 4EBP1

Figure S6 related to Figure 7. High throughput screening using xenograft cells. (A) Plot of z' for each run. (**B)** Plot of luminescence for Cell-TiterGlo readout for the screening showing good separation between the positive (green) and negative controls (red) for two of the xenograft samples. The error bars represent mean +/- standard deviation. (**C)** Histograms of quantitation of total protein and phospho/total protein for components of the Ras/MEK/ERK and PI3K pathways. (**D)** Representative immunoblots for the data shown in (C).

Table S9 related to Figure 7. Drugs and drug response for rhabdomyosarcoma xenografts.

Provided as separate file.

Table S10 related to Figure 7. Quality control for drug screening.

Provided as separate file.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Whole Genome Sequencing and Transcriptome Sequencing

Whole-genome sequencing (WGS), transcriptome sequencing (mRNA-seq), and SNP or gene expression profiling by array were performed as previously described(Zhang et al., 2012a; Zhang et al., 2012b). For both WGS and mRNA-seq, paired-end sequencing was performed using the Illumina GAIIx or HighSeq platform with 100bp read length. The WGS data are deposited at the European Bioinformatics Institute (EBI) with accession number EGAS00001000256.

WGS mapping, coverage and quality assessment, single nucleotide variation (SNV) / indel detection, tier annotation for sequence mutations, prediction of deleterious effects of missense mutations, and identification of loss of heterozygosity (LOH) have been described previously(Zhang et al., 2012b) Structural variations (SVs) were analyzed using CREST and annotated as previously described(Wang et al., 2011; Zhang et al., 2012b). The reference human genome assembly NCBI Build 36 was used for mapping 5 samples (SJRHB001 – 005) and Build 37 was used for mapping other samples. Genomic coordinates for variations in samples mapped to Build 36 were lifted over to Build 37. Copy number variations (CNVs) were identified by evaluating the difference of read depth for each tumour and its matching normal using the algorithm CONSERTING (COpy Number SEgmentation by Regression Tree In Next-Gen sequencing).

SNVs were classified into the following three tiers, as previously described (1):

Tier 1: coding synonymous, nonsynonymous, splice-site, and non-coding RNA variants

Tier 2: conserved variants (cutoff: conservation score $\geq$ 500, based on either the

phastConsElements28way table or the phastConsElements17way table from the UCSC genome browser, and variants in regulatory regions annotated by UCSC annotation (Regulatory annotations included are targetScanS, ORegAnno, tfbsConsSites, vistaEnhancers,eponine, firstEF, L1 TAF1 Valid, Poly(A), switchDbTss, encodeUViennaRnaz, laminB1, cpgIslandExt)

Tier 3: variants in non-repeat masked regions.

Tier 4: All other SNVs.

Paired-end reads from mRNA-seq were aligned to the following 4 database files using BWA (0.5.5) aligner (4): (i) human NCBI Build 37 reference sequence, (ii) RefSeq, (iii) a sequence file that represents all possible combinations of non-sequential pairs in RefSeq exons, and (iv) AceView flat file downloaded from UCSC and representing transcripts constructed from human EST. The final BAM file was constructed by selecting the best alignment in the four databases. SV detection was carried out using CREST (1) and deFuse (5) as well as an algorithm that searched for the predicted junction breakpoints from detected SVs in matching WGS samples.

Exome Sequencing

RHB DNA libraries were prepared from 1 ug of WGA material from matched samples using the Illumnia TruSeq DNA library prep kit following the recommended manufacturer's protocol. Libraries were analyzed on an Agilent Bioanalyzer to inspect quality of each library construction. Germline and diagnostic library samples were independently pooled and applied for exome capture using the Illumina TruSeq Exome Enrichment kit as described by the manufacturer. Captured libraries were then clustered on the Illumina c-bot and were sequenced on an Illumina

HiSeq 2000 platform with 100 base pair end multiplexed reads at an equivalent of 3 samples per lane.

Validation of SNVs, SVs and Indels

For enrichment of the regions containing putative alterations, genomic coordinates of the putative WGS targets were used to order either Nimbelgen Seqcap EZ solution bait sets (Roche), or Nimblegen Seqcap 2.1M arrays (Roche). The library construction and target enrichment was performed per manufacturer's instructions (Roche) using repli-G (Qiagen) WGA DNA. Enriched targets were sequenced on the illumina platform using paired end 100 cycle sequencing. The resulting data was converted to FASTQ files using CASAVA 1.8.2 (Illumina), and mapped with BWA prior to pipeline analysis.

Putative SNVs from exon sequencing were validated by NGS amplicon sequencing. Briefly, primers were designed to genomic regions (hg19) flanking the SNV no closer than 100 base pairs to the detected SNV. PCR was performed using 20 to 30 ng of whole genome amplified (WGA) DNA from each patient sample. DNA from the tumor (diagnostic sample) and a matched germline sample were used for each primer set to confirm presence of the SNV in the diagnostic sample. Standard PCR was performed in 25 ul reactions using Accuprime GC-rich DNA polymerase (Invitrogen) with the following parameters: 95°C for 3 minutes, 35 cycles of 95°C for 30 seconds, 65°C for 30 seconds, 72°C for 1 minute followed by a 72°C 10 minute extension with cooling to 4°C. All PCR amplicons were checked on a 2% E-gel (Invitrogen) to ensure single amplified products.

PCR amplified products were collected into diagnostic or germline pools and purified using a Qiagen PCR purification kit. One nanogram of pooled sample was treated for sequencing using

the Nextera XT kit (Illumina) following the manufacturer's protocol. Samples were normalized and denatured prior to the MiSeq run using the 2x150 base pair run (version 1) reagent kit.

34 Metagene Analysis

Metascore of the 34 genes, derived from previous RMS studies(Davicioni et al., 2010), was calculated for each patient, as a weighted sum of the gene expression value, with the weights being the signed square root of the Cox c2 test statistic. The weights were downloaded from the original paper(Davicioni et al., 2010), and gene expression was estimated as follows. First RPKM (Reads Per Kilobase of exon per Million reads mapped) values for all SJRHB samples were calculated at both exon level and transcript level. Then, the gene expression was estimated as either the maximum normalized log transformed FPKM value of all transcripts or maximum normalized log transformed FPKM value for the highest expressed exon. The metascore is the weighted sum plus 125, to eliminate negative values. Since the original weights for each of the 34 genes were derived originally from microarray data, we used average rank of exon-based expression and transcript-based expression for the transcriptome data.

Statistical Analysis

The statistical significance in Figure 1a was evaluated using Wilcoxon rank sum test. All statistical analyses were performed using R (http://www.R-project.org, version 2.11.1, with basic packages). The statistical analysis of risk group association of RAS pathway mutations was evaluated using exact Chi-square test in Cytel Studio version 9.

Muscle Signature Gene Identification

Data from the ENCODE Affymetrix Human Exon 1.0 ST Array study (GSE15805) containing myoblast and myotube lines was used for this analysis. Samples that were from over-represented groups were trimmed leaving 32 samples of primary cells and 32 samples of transformed cell lines. Gene level unequal variance t-tests were performed to compare myoblast samples to all others excluding myotube samples and comparing myotube samples to all others, excluding myoblast samples. Genes that were expressed at higher levels by at least a logratio of 2 in myotubes or myoblasts as compared to the other cell types were included in the muscle signature gene list. Next, data from an Affymetrix Human Genome U133 Plus 2.0 Array study of adult tissue (GSE3526) was downloaded and RMA was summarized in Partek Genomics Suite 6.6. Skeletal muscle was then compared by unequal variance t-test to all other samples. As for the cell line analysis, those genes that were expressed at higher levels by at least a logratio of 2 were included in the muscle signature gene list.

Tumor Purity Estimations

For germline heterogeneous SNPs, loss of heterozygosity (LOH) measures the absolute difference between the mutant allele fraction in tumor and that in germline sample (0.5). LOH is the result of copy number alterations and/or copy neutral-LOH in tumor cells. Compared to copy number gains (a single copy gain in 100% tumor results in a LOH value of 0.167), regions with copy number loss showed stronger LOH (a single copy loss in 100% tumor result in a LOH value of 0.5). Consequently, we used LOH signals in copy neutral or heterozygous copy number loss regions (CNA value between [-1, 0]) to estimate tumor purity for all WGS samples. Briefly, a single copy loss in x% tumor cells resulted in an estimated CNA value of $-\frac{x}{100}$ and a LOH value

of $\frac{x}{400-2x}$. Assuming the remaining LOH signal came from CN-LOH (CN-LOH in x% tumor cell resulted in a LOH value of $\frac{x}{200}$), the tumor content in a region could be estimated as the sum of the fraction with copy number loss and the fraction with CN-LOH by: $-CNA + 2 * \left(LOH - \frac{-CNA}{4-2CNA}\right)$. Using tumor content estimates from various regions within the genome, we performed an unsupervised clustering analysis using the *mclust* package (version 3.4.8) in R (version 2.11.1). The tumor purity of the sample was defined as the highest cluster center value among all clusters.

Purity Adjusted Mutant Allele Fraction (MAF) Estimation

MAF for validated SNVs was estimated as $\frac{\#Mutant\ reads}{(\#Total\ reads)\times(tumor\ purity)}$ using deep sequencing data.

Tumor Heterogeneity Estimation

We used all validated autosomal SNVs satisfying the following criteria in heterogeneity analysis:

1) In copy neutral region (Log2ration between (-0.1, 0.1) in CNV analysis).
2) Not in regions with LOH (LOH value < 0.1).
3) With MAF > 0.05 or mutant allele count > 2.

We drew the kernel density estimate plot for MAFs of the qualifying SNVs using the *density* function in the *stat* package in R. We also estimated the number of significant peaks and the relative MAF component for each peak (peaks with less than 5 SNVs, peaks with less than 1% SNVs, and peaks with excessive variance were ignored). A sample with heterogeneity shows density peaks at a MAF smaller than 0.5 (the expected MAF assuming heterogeneous SNVs).

Tumor Evolution Analysis

For SJRHB011 or SJRHB012, we designed a single chip to capture predicted mutations in all tumors for the patient. For initial assessment, we collected deep sequencing data from SNVs (Tiers 1-3) satisfying the following 2 criteria:

1) Validated in at least one tumor
2) Good coverage in all tumors (> 20X)

A subset of the validated SNVs satisfying additional criteria were used for the quantitative evolution analysis:

3) In diploid regions among all tumors
4) In regions without LOH
5) MAF was set to 0 if MAF < 0.05 and mutant allele count < 3

Mutations unique to a single tumor were clustered using the MAF distribution by the *mclust* package in R. Mutations shared by at least 2 tumors were clustered using the joint MAF distribution by *mclust*. Clusters with excessive variance were defined as outliers. The number of clusters was determined by merging clusters with similar centers (with different covariance matrices) and ignoring clusters with less than 1% of SNVs. The clustering analysis was finalized using the *mclust* package with the determined number of clusters.

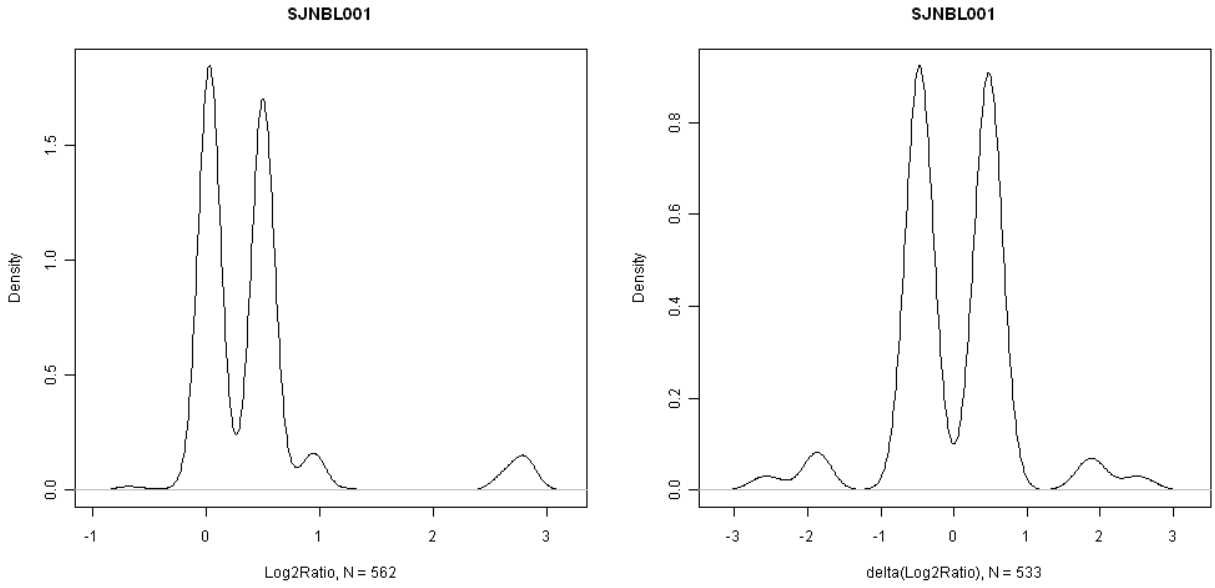
In SJRHB012, we performed PCA analysis on the joint MAF mutation using the *prcomp* function in the *stat* package of R. PC1 and PC2 explained 61.8% and 32.1% of variance. Consequently, we used PC1 and PC2 in the clustering analysis.

Quantitative Analysis of Chromothrypsis

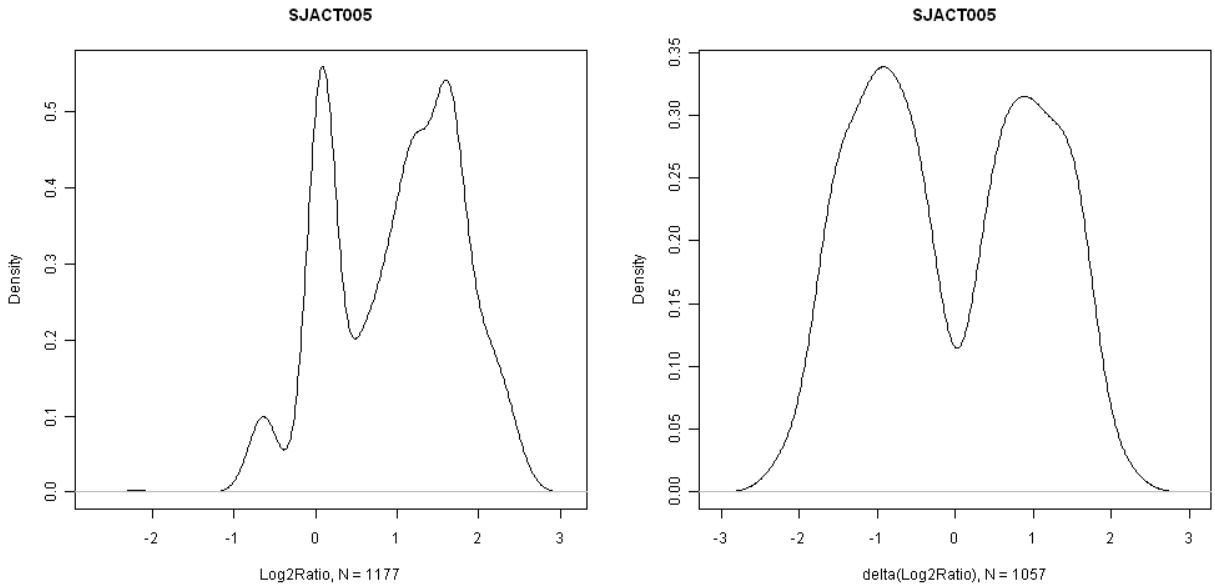
A quantitative score was developed to measure the fitness of the chromothripsis hypothesis in a certain sample (a real chromothripsis sample has a score significantly higher than 0 while samples without chromothripsis have scores close to 0) and significance was measured empirically.

There were three criteria used for analysis of chromothripsis based on the assumption that chromosomes with chromothripsis will have a significant number of SV supported CNV segments (in the analysis, chromosomes with less than 20 SV supported CNV segments were excluded). The first criteria is that the log2 ratio of the CNV segments oscillating among a small number of copy number states for chromothripsis. Therefore, there will be distinct peaks in density plot for log2 ratio of SV supported CNV fragments with at least 2 major peaks corresponding to the states in chromothripsis. The second criteria for chromothripsis is the presence of distinct peaks in the density plot for Δ(log2 ratio) at the SV breakpoints. The third criteria is that for the 2 major log2 ratio peaks by chromothripsis (with a difference of d between the peak centers), there will 2 major peaks in the Δ(log2 ratio) density plot, centered around both d and $-d$.

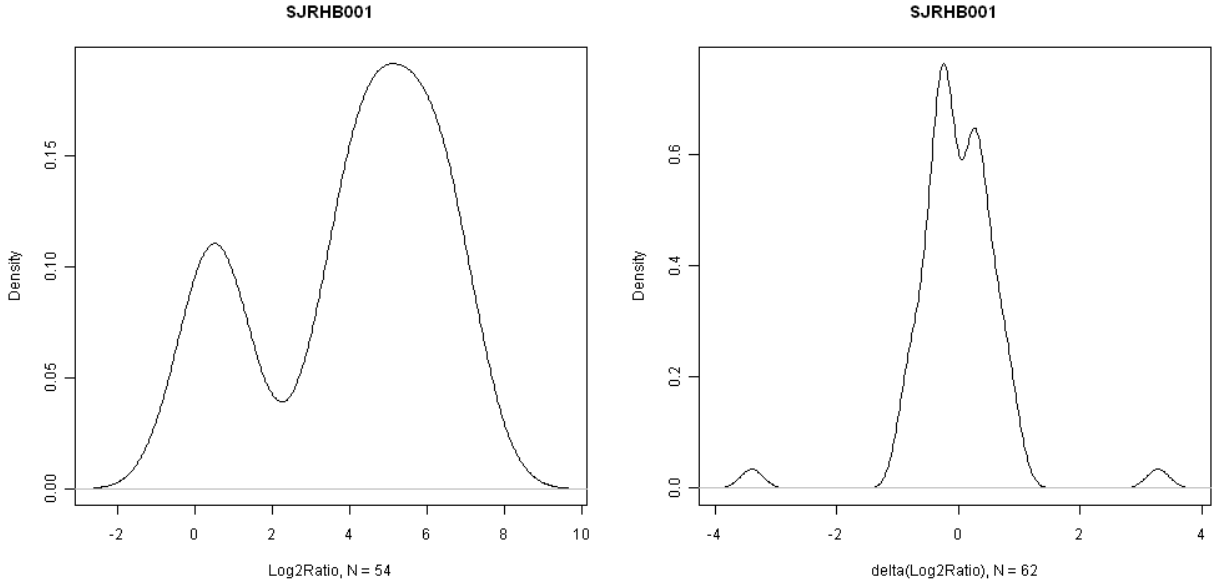
Among the 16 RHB samples, 9 samples do not have any chromosome satisfying Criteria 1 (above). The log2 ratio density plot and Δ(log2 ratio) density plot were shown here for a few remaining examples, together with 2 positive controls: SJNBL001 (2-state chromothripsis) and SJACT005 (multi-state chromothripsis). For SJNBL001 (distinct peaks in both plots), the peak position in the Δ(log2 ratio) plot matches the distance between the 2 major peaks in the log2 ratio plot consistent with 2-state chromothripsis.



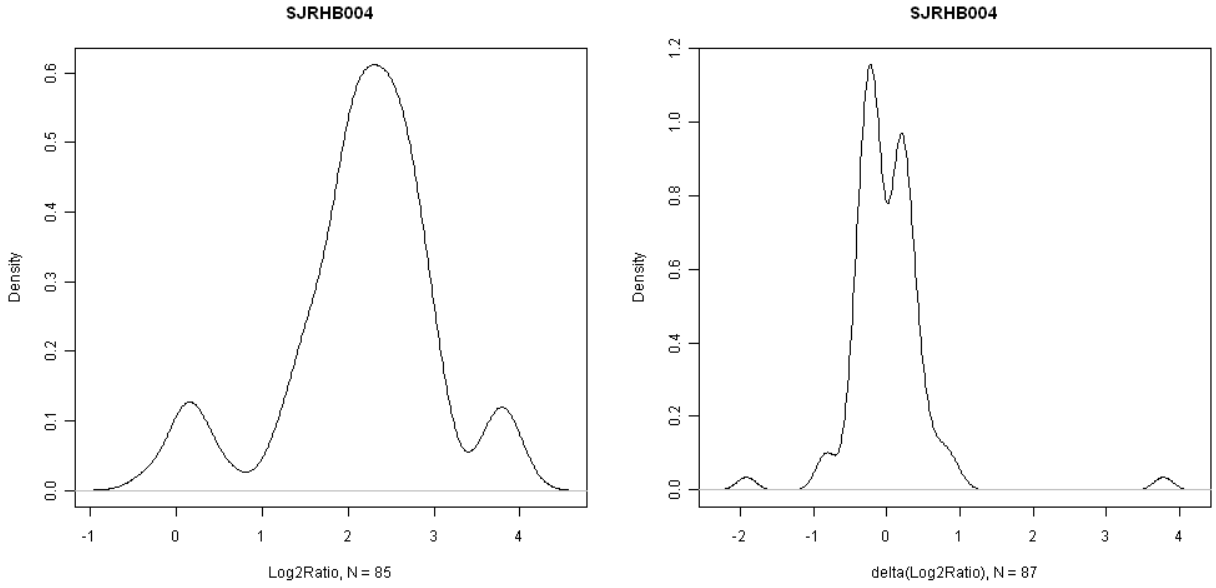
1. SJACT005 (multi-state chromothripsis)



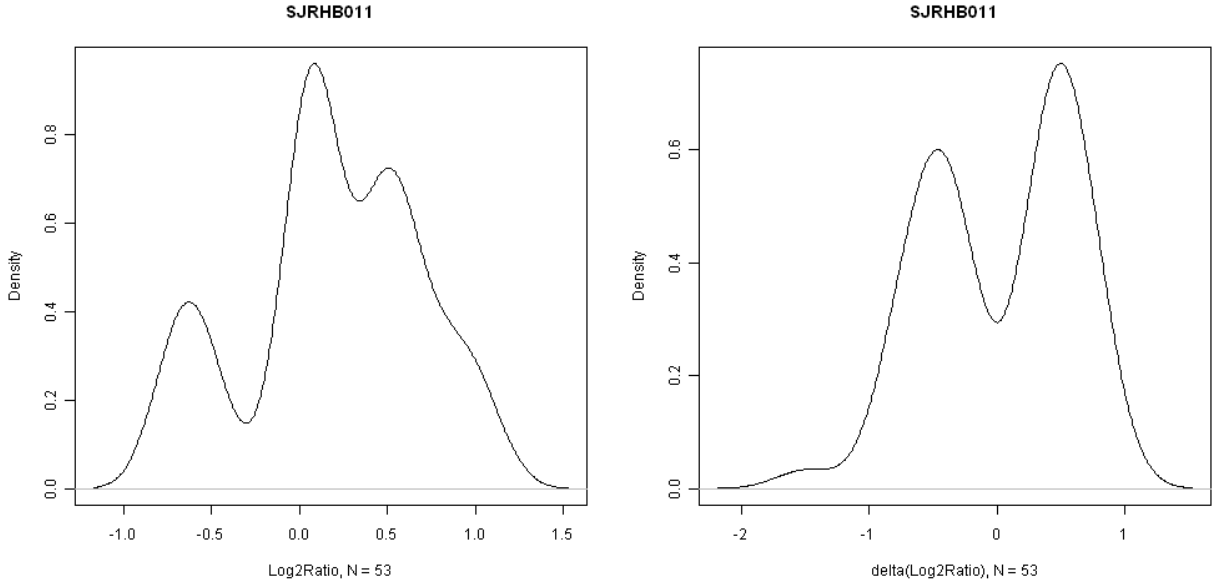
2. SJRHB001 (2 peaks in log2 ratio plot, but a single peak in Δ(log2 ratio) plot)



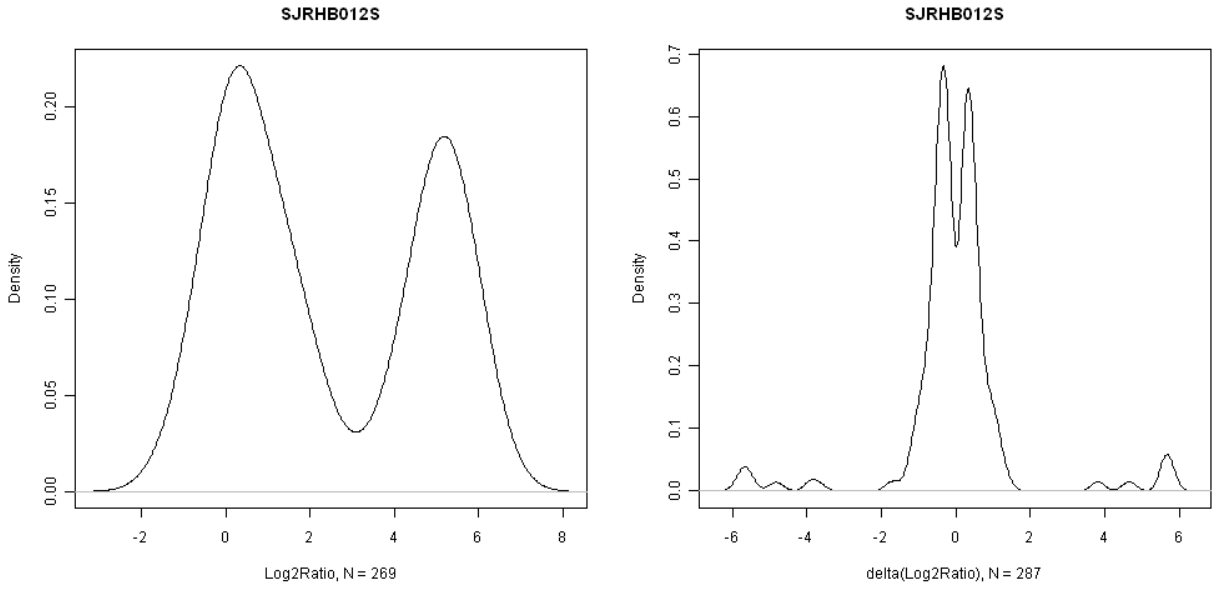
3. SJRHB004 (a single major peak in both plots)



4. SJRHB011 (too many peaks in log2 ratio peaks, which is unlikely derived from chromothripsis for a imited number of CNV segments)



5. SJRHB012S (distinct peaks in both log2 plots. However, the peak position in the Δ(log2 ratio) plot does not match the distance between the 2 major peaks in the log2 ratio plot).



The density peaks were estimated using the *mclust* package in R. The maximum number of allowed peaks is determined by $max\left(2, min\left(10, ceil\left(\sqrt{\frac{\#CNV\ segments}{50}}\right)\right)\right)$.

A chromothripsis score was developed to measure the fitness of the chromothripsis hypothesis in a sample by the following formula:

$$chromothripsis\ score$$
$$= (peak\ separation\ score\ in\ log2\ ratio) \times (peak\ separation\ score\ in\ \Delta log2\ ratio)$$
$$\times (peak\ matching\ scores)$$

where the peak separation score in a density plot is measured as

$$2 \times (min(peak\ area_1, peak\ area_2) - shared\ area\ between\ peak1\ and\ peak2)$$

and the peak matching score is measured as

$$max\left(prob\left(peak\ center_{\Delta(log2\ ratio)}\ from\ N\left(peak\ difference_{log2\ ratio}, \frac{\substack{pooled\ Variance_{in-cluster} in\ log2\ ratio + \\ pooled\ Variance_{in-cluster} in\ \Delta(log2\ ratio)}}{2}\right)\right)\right) \times$$

$$max\left(prob\left(peak\ center_{\Delta(log2\ ratio)}\ from\ N\left(-peak\ difference_{log2\ ratio}, \frac{\substack{pooled\ Variance_{in-cluster} in\ log2\ ratio + \\ pooled\ Variance_{in-cluster} in\ \Delta(log2\ ratio)}}{2}\right)\right)\right)$$

The chromothripsis scores for the set of samples are listed below (0 for samples that failed the first criteria). Empirical p values were estimated from an empirical null distribution by permuting the log2 ratio values in the SJNBL001 and SJACT005 samples (1000 times each).

Chromothripsis Scores							
SJACT005	4.45E-2 p = 0.001	SJNBL001	0.280 p < 0.0005				
SJRHB001	0 p = 1	SJRHB002	0 p = 1	SJRHB003	0 p = 1	SJRHB004	0 p = 1
SJRHB005	0 p = 1	SJRHB006	0 p = 1	SJRHB007	0 p = 1	SJRHB008	0 p = 1
SJRHB009	0 p = 1	SJRHB010	0 p = 1	SJRHB011E	0 p = 1	SJRHB011D	0 p = 1

SJRHB012D	0 p = 1	SJRHB012R	0 p = 1	SJRHB012S	8.21E-41 p = 0.424	SJRHB013	0 p = 1

Chr8 Analysis

Using CNV analysis from WGS or SNP6, ERMS samples were classified into 2 groups: one with gross amplification on chr8 and the other without gross amplification on chr8. Samples without CNV data were removed from this analysis. Density of expression levels (FPKM from RNAseq) of all genes on chr8 was estimated and plotted for each group (on log2 scale). Differential gene expression between the 2 groups was also investigated across the genome. A gene is considered to be differentially expressed if the FDR q-value <= 0.05 (t-test on log2 scale) and the fold change >= 2.

Rhabdomyosarcoma Xenografts

General Drug Screening Information. All xenograft lines are assayed for optimal growth conditions, including plating density, DMSO-sensitivity, and positive control compounds before being validated and assayed for drug sensitivities All RMS xenograft lines are plated in vitro in SkBM-2 cell culture medium (Lonza Cat#CC-3246) and supplemented with SingleQuots supplements (Lonza Cat#CC-3244).

Plating Density and DMSO-Sensitivity. On a flat-bottomed, white 384-well plate (Corning Costar Cat#8804BC), disaggregated rhabdomyosarcoma xenograft cells are plated in eight densities beginning at 5,000 cells/well in 25 µl and diluted serially 1:2, resulting in 96 wells per

cell density. Twenty four hours after plating, all cell densities are treated with varying DMSO concentrations of 0.083% DMSO, 0.197% DMSO, 0.443% DMSO, and no DMSO, resulting in twelve wells per cell density at a given DMSO concentration. Seventy two hours following the addition of DMSO, CellTiter-Glo (Promega Cat#G7573) is added at 25µl/well to determine a relative light unit (RLU) signal with a PerkinElmer EnVision plate reader. RLU data is then interpreted to determine the optimal cell density to achieve logarithmic growth at 96h post-plating and whether a particular xenograft line is sensitive to DMSO.

Optimal Cell-plating Densities

Xenograft Line	Cells/well
SJRHB011_Y	5,000
SJRHB011_X	5,000
SJRHB012_X	1,250
SJRHB012_Y	5,000
SJRHB012_Z	2,500
SJRHB013_X	5,000

Positive Control Compound Selection. Disaggregated rhabdomyosarcoma xenograft cells are plated at optimal density in a 384-well plate as described. Twenty four hours post-plating, cells are drugged with six compounds (doxorubicin HCl, staurosporine, etoposide, SN-38, bortezomib, and cycloheximide) at both relatively high concentration as well as 1:3 serially-diluted dose-response curves. Seventy two hours post-drugging, RLUs are generated by the CellTiter-Glo/EnVision system. Data is interpreted to determine the best suited positive control compound for future assays which achieves complete or nearly complete cell death at high concentration and provides a distinct IC_{50} midpoint concentration.

Assay Validation. Validation of the assay is performed to demonstrate reproducibility of results. On two separate days, disaggregated rhabdomyosarcoma xenograft cells are plated at optimal density in triplicate on 384-well plates. Twenty four hours post-plating, each validation plate is drugged with the previously determined positive control compound. Each plate is drugged in a pattern of alternating columns of high concentration, IC_{50} concentration, and DMSO. Seventy two hours post-drugging, RLUs are determined by the CellTiter-Glo/EnVision system. The assay will be validated if the endpoint data shows clearly distinct patterns of cell death corresponding to the positive control compound concentration on both days of the validation.

Drug Screening. After the assay conditions have been validated, the xenograft cell lines may be screened against compounds for activity. Disaggregated rhabdomyosarcoma xenograft cells are plated at previously determined optimal density in 384-well plates. Twenty four hours post-plating, assay plates are drugged in triplicate with compounds of interest. For single-point drugging, each compound is drugged against one well per assay plate; for dose-response drugging, each compound is drugged against multiple wells per plate at different concentrations per well. Seventy two hours post-drugging, RLUs are determined by the CellTiter-Glo/EnVision system. Data is then be analyzed to determine active compounds in the case of single-point drugging as well as dose-response curves.

Immunohistochemistry

Immunohistochemical studies were performed on 4 µm thick sections of formalin-fixed paraffin-embedded tissue blocks using commercially available antibodies against myogenin (clone F5D; DAKO; 1:500) and beta-catenin (clone 14; BD Biosciences; 1:600). Myogenin immunostaining was carried out using the heat-induced epitope retrieval procedure with CC1-Standard and the

Ventana IVIEW DAB Detection Kit on the Ventana Benchmark XT automated stainers (Ventana Systems, Tucson AZ). Beta-catenin immunostaining was conducted using the heat-induced epitope retrieval pretreatment with ER1 for 20 minutes and the Refine Polymer DAB detection kit on the Leica BOND-MAX and Bond-III automated instruments (Leica Biosystems, Bannockburn, IL). Myogenin immunostaining was considered positive if there was distinct nuclear staining in any tumor cells and further specified as patchy or diffuse based on a heterogeneous or homogenous staining pattern, respectively. Beta-catenin immunostaining was scored on a scale of negative to 3+ according to the percentage of cells with positive nuclear staining (negative, no nuclear staining; rare, <1-4%; 1+, 5-25%; 2+, 26-50%; 3+, >50%).

TP53, MDM4, CDKN2A and MDM2 Fluorescence In Situ Hybridization Studies.

The fluorescence in situ hybridization (FISH) probe sets were designed using bacterial artificial chromosome (BAC) clones according to the UCSC Genome Bioinformatics database (http://genome.ucsc.edu). DNA was isolated from BAC clones (BACPAC Resources, Oakland, CA) according to a modified Qiagen (Valencia, CA) extraction protocol. The probes were labeled by nick translation using a modification of the manufacturer's protocol (Life Technologies, Inc., Carlsbad, CA). FISH analysis was performed on 4 micron-thick FFPE tissue sections using the previously published methods(Bahrami et al., 2012). Hybridization signals were evaluated in 200 interphase nuclei of each sample. FISH images were captured and processed as previously described(Bahrami et al., 2012).

TP53 Immunostaining

The corresponding formalin-fixed, paraffin-embedded (FFPE) tissue blocks for each specimen were cut at 4 micron thickness. Immunohistochemical staining was performed using an antibody directed against p53 protein (DO-7, DAKO, 1:50) and processed with standard heat-induced

epitope retrieval (Ventana CC1) and the Ventana IVIEW detection systems. p53 nuclear staining was scored using a previously published scoring system(Papai et al., 1997) with a minor modification as follows: samples with no staining cells were scored as negative; samples with <5% p53 immunopositive cells were scored as rare; samples with 5-25% immunopositive cells were scored as 1+; samples with 26-50% immunopositive cells were scored as 2+; samples with >50% immunopositive cells were scored as 3+.

ALK Immunoblotting

Xenograft extracts were prepared by grinding tissues under liquid nitrogen, vortexing in lysis buffer containing protease and phosphatase inhibitors (Cell Signaling, Danvers, MA; cat# 9803, 5872, and 8553, respectively), and separation on QIAshredder spin columns (Qiagen, Germantown, MD). The eluant was then sonicated on ice and after centrifugation, the supernatant was subjected to separation using 4-12% BisTris SDS-PAGE (Life Technologies, Carlsbad, CA). Following electrophoresis, proteins were transferred to Immobilon-P membrane (EMD Millipore, Billerica, MA). Immunoreactivity to ALK and GAPDH was undertaken using standard protocols with antibodies obtained from Cell Signaling (ALK Ab – mouse monoclonal, cat# 3791; GAPDH Ab – rabbit HRP conjugated monoclonal, cat# 8884). Detection was accomplished using ECL reagents (GE Healthcare Biosciences, Pittsburgh, PA).

Clonal Heterogeneity and Clonal Evolution

In the discovery cohort, we analyzed tumor purity and intratumor heterogeneity. Tumor purity was estimated from copy number alterations (CNAs) and loss of heterogeneity (LOH) data from WGS and tumor heterogeneity was estimated using the purity adjusted mutant allele fraction (MAF) derived from deep sequencing of all SNV mutations by liquid phase capture and Illumina sequencing (Supplementary method).

Genome instability (CNAs, and copy neutral LOH, CN-LOH) is frequently observed in solid tumor including rhabdomyosarcoma. To avoid potential complications from CNAs and CN-LOH, we focused on validated SNVs in diploid regions without CN-LOH in the tumor colonal heterogeneity analysis.

Sup. Fig. S2 showed the MAF density plot and RHB Sup. Table A listed the number of significant peaks and their corresponding MAF estimates for 15 of the 16 tumors in the discovery cohort (SJRHB013, a sample with low tumor content, was excluded). Thirteen of the 15 samples showed a peak corresponding to MAF of 0.5, suggesting that the tumor purity estimate from WGS CNAs and LOH analysis is accurate. SJRHB002 showed a MAF peak at 0.45 and a close inspection of the density plot in Sup. Fig. 2A indicated this peak has a heavy right tail and it is likely to have another (secondary) peak around MAF 0.5. SJRHB004 had a single peak around 0.26 (with 40 SNVs). An inspection of the 30 validated SNVs on chrX (male, no CNAs) showed 2 distinct peaks with MAFs of 1.01 and 0.53, respectively (data not shown), corresponding to MAFs of 0.50 and 0.26 in diploid regions, confirming that the initial tumor purity estimation was correct. This result suggested that there are relatively few mutations in the founding clone in SJRHB004 with a subclone in 50% of the tumor containing majority of detected mutations. We found a missense mutation in ARHGAP1 (G128C, MAF = 0.30) on chr11 (Copy number, CN=4

with LOH). ARHGAP1 is a negative regulator of Cdc42 and deficiency in this protein leads to genome instability(Wang et al., 2007). Potentially G128C in ARHGAP1 led into the accumulation of other mutations in the same subclone. Similar to SJRHB004, majority of mutations detected in SJRHB001 and SJRHB009 were also presented in subclones. Overall, 10 of the 15 tumors had peaks with significant lower MAF than 0.5 and 9 showed multiple peaks (7 with 2, 1 with 3 and 1 with 4 peaks), which is similar to the estimate in acute myeloid leukaemia (4/7 showed multiple clusters)(Ding et al., 2012).

Although NGS technique facilitated tumor evolutionary analysis in leukemia tumors(Ding et al., 2012; Jan et al., 2012; Janssens et al., 2012; Walter et al., 2012), large scale genome instability in solid tumors poses intrinsic challenges in evolutionary analysis. In our discovery cohort we analyzed tumors from 2 patients that had recurrent disease (Sup. Table 1). In one case (SJRHB011), a recurrence (SJRHB011_D) was compared to the diagnostic tumor isolated 15 months earlier (SJRHB011_E) (Fig. 2a). In the second case (SJRHB012), recurrence tumors from two distinct sites (SJRHB012_R , Recurrence 1, prostate, and SJRHB012_S, Recurrence 2, pelvis) were compared to the primary tumor from the same patient (SJRHB012_D, Primary) that had been resected 14 months earlier (Fig. 2b). In both cases, the patient received chemotherapy prior to surgical resection and whole genome sequence analysis. This provides us with the opportunity to begin to explore clonal evolution in rhabdomyosarcoma. We designed a single capture chip for each patient that included predicted mutations in primary as well as recurrence tumors. Using this design, deep sequencing quantify the MAFs for the same set of mutations in all tumors, and can distinguish true *de novo* mutations in recurrence tumors from the mutations present in the primary tumor at very low frequencies, which could be missed in the primary tumor analysis due to limited coverage in WGS.

In SJRHB011, the median coverage of targeted sites in the primary tumor (SJRHB011_E, Infratemporal fossa) was 116X while the median coverage in recurrence tumor (SJRHB011_D, Neck) was 662X. A total of 3524 autosomal SNVs (148 in coding regions) had good coverage (>20X) in both the primary tumor and the recurrence tumor in deep sequencing, including 1599 (45.4%) shared mutations with 65 (43.9%) in coding regions, 354 (10.0%) mutations unique to the primary tumor with 8 (5.4%) in coding regions, and 1571 (44.6%) mutation unique to recurrence tumor with 75 (50.7%) in coding regions. TP53 was the only cancer consensus gene with functional mutation in this patient. It was homozygously mutated (C176F coupled with loss of the WT copy) in both primary (MAF = 0.92) and recurrence (MAF = 0.91) tumors. To quantitatively estimate the evolutionary path, we focused on 841 heterozygous mutations (MAF < 0.6) in diploid regions (Copy Number, CN=2) without LOH (Fig. 3a). After ignoring a small outlier cluster (in black, 15 shared mutations), there were 4 distinct clusters of mutations. Cluster A (in red, 346 mutations) were found at major clones in both primary (median MAF = 0.49) and recurrence (median MAF = 0.53) tumors, representing the mutations in the founding clone. Cluster B (in sky blue, 353 mutations) were observed in the major clones in recurrence tumor (median MAF = 0.46) but absent from the primary tumor. Cluster C (in orange, 70 mutations) were unique to the recurrence tumor but only in a subclone (median MAF = 0.17). These 2 clusters represented newly acquired mutation after the surgery and chemotherapy. Cluster D (in dark blue, 57 mutations) represented the mutation found in the primary tumor (median MAF = 0.48) only and cells with these mutation did not survive the treatment. A possible evolutionary path was shown in Fig. 3b. At diagnostic, the majority (~97%) of tumors cells were A^+D^+ while a small fraction (3%) of cells was derived directly from the founding clones (A^+D^-). The initial treatment destroyed the dominant clone in the primary tumor. However, some cells from the

A^+D^- clone either evaded the treatment or gained resistance to the treatment through acquiring additional mutations ($A^+ B^+D^-$) and expanded to the recurrence tumor. A fraction (34%) of the $A^+ B^+D^-$ cells further acquired Cluster C mutations and became $A^+ B^+ C^+D^-$. In consistent with this hypothesis, a fraction of mutations in the outlier cluster (in black) displayed very low MAF (0.005 – 0.02) in the primary tumor and high MAF (>0.4) in the recurrence tumor.

Similarly, in SJRHB012, the median coverage in the primary (SJRHB012_D, prostate), the recurrence 1 (SJRHB012_R, prostate) and the recurrence 2 (SJRHB012_S, pelvis) tumors is 335X, 130X and 145X, respectively. A total of 3580 autosomal SNVs (163 in coding regions) had good coverage among 3 tumors in deep sequencing, including 634 (17.7%) shared mutations among all tumors with 30 (18.4%) in coding regions, 2372 (66.2%) shared mutations between the 2 recurrence tumors but absent from the primary tumor with 104 (63.8%) in coding regions, 129 (3.6%) mutation unique to the primary tumor with 3 (1.8%) in coding regions, 358 (10.0%) mutation unique to the recurrence 1 tumor with 25 (15.3%) in coding regions, and 80 (2.2%) mutation unique to the recurrence 2 tumor with 1 (0.6%) in coding regions. The remaining 7 mutations (0.2%, none in coding regions) were shared between the primary and recurrence 2 tumor but absent from the recurrence 1 tumor. However, the MAFs in the recurrence 2 tumor was small (median 0.03, range 0.01 – 0.09) and these mutations were exclude from further investigation. We did not detect any mutation in cancer consensus genes in the primary tumor. An ALK mutation (P1445H) was detected at the major clones in both recurrence tumors (MAF = 0.46 in recurrent 1 and 0.43 in recurrent 2, both in diploid region). Both point mutations and CNA in ALK has been reported in rhabdomyosarcoma(Van Gaal et al., 2012) and potentially acquiring this mutation helped the recurrent tumor to survivor the initial treatment and developed into recurrence tumors. A second GMPS mutation (Q188L) unique to a subclone of recurrence 1

tumor (MAF = 0.12, CN=3) was identified. A RECQL4 missense mutation (G120R) was found in the recurrence 2 tumor (MAF = 0.29, CN estimated around 3-4) but not in the primary tumor. However, its status in the recurrence 1 tumor was uncertain due to insufficient coverage. Next, we focused on a subset of 1049 heterozygous mutations in diploid regions without LOH among all 3 tumors to infer the evolutionary path. The unsupervised clustering analysis showed 6 distinct clusters of mutations (Fig. 3c). Cluster A (in red, 174 mutations) represented mutations in the founding clone, which were present in all cells of the 3 tumors (median MAF = 0.48, 0.50 and 0.50 in the primary, recurrence 1 and recurrence 2, respectively). Cluster B (in sky blue, 567) contains the largest number of mutations, which were present in almost all tumor cells for both recurrence tumors (median MAF = 0.49 and 0.50 in the recurrence 1 and recurrence 2 tumors, respectively) but absent from the primary tumor. These mutations were likely acquired after the initial treatment and helped the carrying cells to gain resistance. As these cells were found in all recurrence cells, they were acquired before the divergence of the 2 recurrence tumor. The ALK mutation matches the characteristics of mutations in this cluster. Cluster C (in orange, 117 mutations).were detected in the dominant clone of the recurrence 2 tumor (median MAF = 0.38), a small fraction of the recurrence 1 tumor (median MAF = 0.10) and absent in the primary tumor. These mutations were acquired shortly before the divergence of the recurrence tumors. Combined with the unique mutations in each recurrence tumor, Cluster C mutations could explain the different grow potential for various subclones at distinct sites (pelvis for Recurrence 2 and prostate for recurrence 1). Cluster D (in dark blue, 28 mutations) were unique to a subclone in the recurrence 2 tumor (median MAF = 0.25). Cluster E (in green, 144 mutations) were unique to a small fraction of the recurrent 1 tumor (median MAF = 0.165). Cluster F (C6, in pink, 17 mutations) represented the mutations acquired in the dominant clone in the primary tumor

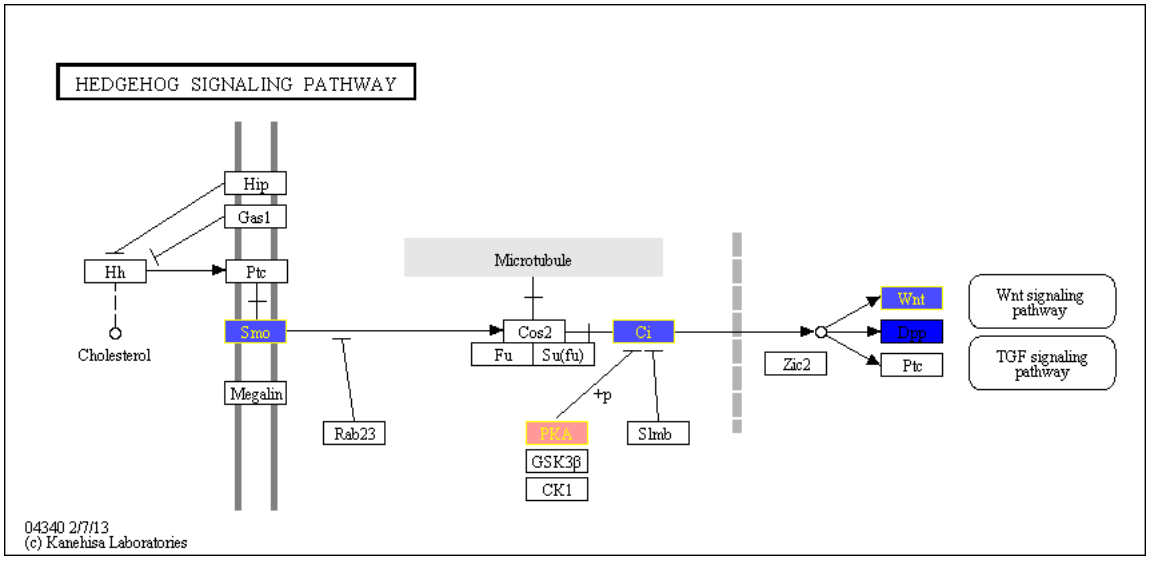
(median MAF = 0.35) and cells in this clone did not survive the initial treatment. A possible evolution path was displayed in Fig. 3d. It suggested that although single sample analysis did not detect multiple clones in the primary tumor (Sup. Fig. 2), multiple time point analysis revealed at least 2 different subclones. The dominant clone (A^+F^+) acquired a small number of mutations compared to the founding clone (A^+F^-). Similar to SJRHB011, the recurrence tumors in SJRHB012 were derived from the founding clone after acquiring additional mutations, including the ALK mutation ($A^+ B^+F^-$). Then a fraction of the $A^+ B^+F^-$ clone acquired additional mutation and became $A^+ B^+ C^+F^-$. Both clones ($A^+ B^+ C^-F^-$ and $A^+ B^+ C^+F^-$) seeded the 2 recurrence tumors.

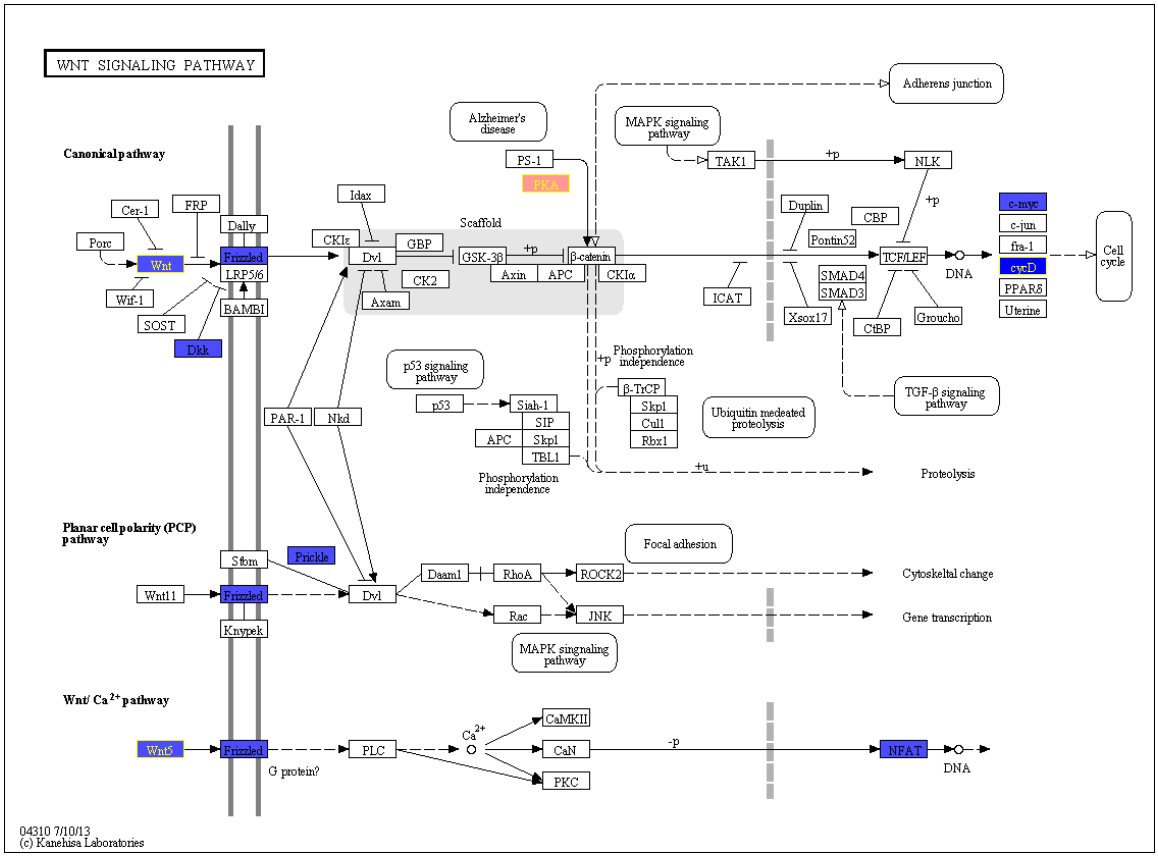
Comparison of SNV Mutation Spectrum Between Diagnostic and Recurrent Samples

We compared the mutation spectrum between the mutations detected in the diagnostic tumors with those acquired in recurrence for SJRHB011 and SJRHB012. A SNV was classified as "newly acquired in recurrence" if it was validated in recurrence samples but not detected in the diagnostic sample. The mutation signatures in the two diagnostic tumors were significantly different. Transversions were the major mutation type in SJRHB011 (61%), transitions was the major type in SJRHB012 (53%). However, the mutation spectrum was very similar among the mutations acquired in the recurrent tumor samples, consistent with the fact that both patients received similar treatment regimens (Sup. Fig. 2 and Sup. Table 9).

Gene Expression Network Analysis and Analysis of Developmental Pathways in Rhabdomyosarcoma.

For the network analysis, the gene to gene distance was computed based on the pair-wise correlation of gene (probeset) expression levels. More precisely, the distance (1-correlation.coefficient)^beta was used, where beta is a value that makes the network structure satisfying scale-free topology. In this analysis beta was set to 7. This was done to remove the weaker correlations from network construction by retaining strong correlations between genes. Then, based on the distance, all genes/probesets were clustered. Combinations of a few parameters (tree height, module size=25, whether to do further deep split in the branches) were selected so that clusters of genes (i.e. modules) have at least 25 probesets in each module. For each module (gene set), a principal component analysis was performed to get the top component, then those values were associate with the trait of samples, in order to identify which modules (gene sets) are correlated (positively or negatively) with the phenotype (traits). For the significant modules selected, a network was constructed based on whether it has connections (strong correlations) between two genes. For each module, a gene set enrichment analysis was run against KEGG pathway definition to find the significant pathways (FDR adjusted p-value <0.05). The WNT and SHH pathways are presented here with colored genes being deregulated and those with yellow highlight are also differentially methylated:


HEDGEHOG SIGNALING PATHWAY
Hip
Gas1
Hh
Ptc
Cholesterol
Smo
Megalin
Rab23
Microtubule
Cos2
Fu
Su(fu)
Ci
+p
PKA
GSK3β
CK1
Slmb
Zic2
Wnt
Dpp
Ptc
Wnt signaling pathway
TGF signaling pathway
04340 2/7/13
(c) Kanehisa Laboratories


WNT SIGNALING PATHWAY
Canonical pathway
Cer-1
FRP
Dally
Porc
Wnt
Frizzled
LRP5/6
Wif-1
SOST
BAMBI
Dkk
Idax
CKIε
Dvl
GBP
CK2
Axam
Scaffold
GSK-3β
+p
β-catenin
Axin
APC
CKIα
Alzheimer's disease
PS-1
PKA
MAPK signaling pathway
TAK1
NLK
Duplin
CBP
Pontin52
TCF/LEF
DNA
SMAD4
SMAD3
Xsox17
CtBP
Groucho
ICAT
c-myc
c-jun
fra-1
cycD
PPARδ
Uterine
Cell cycle
Adherens junction
TGF-β signaling pathway
p53 signaling pathway
p53
Siah-1
SIP
APC
Skp1
TBL1
β-TrCP
Skp1
Cul1
Rbx1
Ubiquitin medeated proteolysis
Phosphorylation independence
Proteolysis
PAR-1
Nkd
Planar cell polarity (PCP) pathway
Sfrm
Prickle
Wnt11
Frizzled
Knypek
Dvl
Daam1
RhoA
ROCK2
Rac
JNK
Focal adhesion
Cytoskeltal change
Gene transcription
MAPK signaling pathway
Wnt/ Ca2+ pathway
Wnt5
Frizzled
G protein?
PLC
Ca2+
CaMKII
CaN
PKC
-p
NFAT
DNA
04310 7/10/13
(c) Kanehisa Laboratories

SUPPLEMENTAL REFERENCES

Bahrami, A., Dalton, J. D., Krane, J. F., and Fletcher, C. D. (2012). A subset of cutaneous and soft tissue mixed tumors are genetically linked to their salivary gland counterpart. Genes, chromosomes & cancer *51*, 140-148.

Ding, L., Ley, T. J., Larson, D. E., Miller, C. A., Koboldt, D. C., Welch, J. S., Ritchey, J. K., Young, M. A., Lamprecht, T., McLellan, M. D.*, et al.* (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature *481*, 506-510.

Jan, M., Snyder, T. M., Corces-Zimmerman, M. R., Vyas, P., Weissman, I. L., Quake, S. R., and Majeti, R. (2012). Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Science translational medicine *4*, 149ra118.

Janssens, A., Van Roy, N., Poppe, B., Noens, L., Philippe, J., Speleman, F., and Offner, F. (2012). High-risk clonal evolution in chronic B-lymphocytic leukemia: single-center interphase fluorescence in situ hybridization study and review of the literature. European journal of haematology *89*, 72-80.

Papai, Z., Feja, C. N., Hanna, E. N., Sztan, M., Olah, E., and Szendroi, M. (1997). P53 Overexpression as an Indicator of Overall Survival and Response to Treatment in Osteosarcomas. Pathology oncology research : POR *3*, 15-19.

Van Gaal, J. C., Van Der Graaf, W. T., Rikhof, B., Van Hoesel, Q. G., Teerenstra, S., Suurmeijer, A. J., Flucke, U. E., Loeffen, J. L., Sleijfer, S., and De Bont, E. S. (2012). The impact of age on outcome of embryonal and alveolar rhabdomyosarcoma patients. A multicenter study. Anticancer research *32*, 4485-4497.

Walter, M. J., Shen, D., Ding, L., Shao, J., Koboldt, D. C., Chen, K., Larson, D. E., McLellan, M. D., Dooling, D., Abbott, R.*, et al.* (2012). Clonal architecture of secondary acute myeloid leukemia. The New England journal of medicine *366*, 1090-1098.

Wang, J., Mullighan, C. G., Easton, J., Roberts, S., Heatley, S. L., Ma, J., Rusch, M. C., Chen, K., Harris, C. C., Ding, L.*, et al.* (2011). CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nature methods *8*, 652-654.

Wang, L., Yang, L., Debidda, M., Witte, D., and Zheng, Y. (2007). Cdc42 GTPase-activating protein deficiency promotes genomic instability and premature aging-like phenotypes. Proceedings of the National Academy of Sciences of the United States of America *104*, 1248-1253.

Zhang, J., Benavente, C. A., McEvoy, J., Flores-Otero, J., Ding, L., Chen, X., Ulyanov, A., Wu, G., Wilson, M., Wang, J.*, et al.* (2012a). A novel retinoblastoma therapy from genomic and epigenetic analyses. Nature *481*, 329-334.