The American Journal of Human Genetics, Volume *87*

Supplemental Data

Mechanisms of Genomic Instabilities Underlying Two Common Fragile-Site-Associated Loci, *PARK2* and *DMD*, in Germ Cell and Cancer Cell Lines

Jun Mitsui, Yuji Takahashi, Jun Goto, Hiroyuki Tomiyama, Shunpei Ishikawa, Hiroyo Yoshino, Narihiro Minami, David I. Smith, Suzanne Lesage, Hiroyuki Aburatani, Ichizo Nishino, Alexis Brice, Nobutaka Hattori, and Shoji Tsuji

Figure S1. Six Independent Deletions Occurring Simultaneously in One Cancer Cell Line

Scan data of array CGH analysis of one cancer cell line with 6 independent deletions occurring simultaneously. The horizontal axis represents the nucleotide position. The vertical axis represents $\log_2$ (case/reference signal intensities) on array CGH. Dots of $\log_2$ (case/reference signal intensities) of more than 0 are shown in red and those of less than 0 are shown in green. Black double arrows represent the length of each deletion.

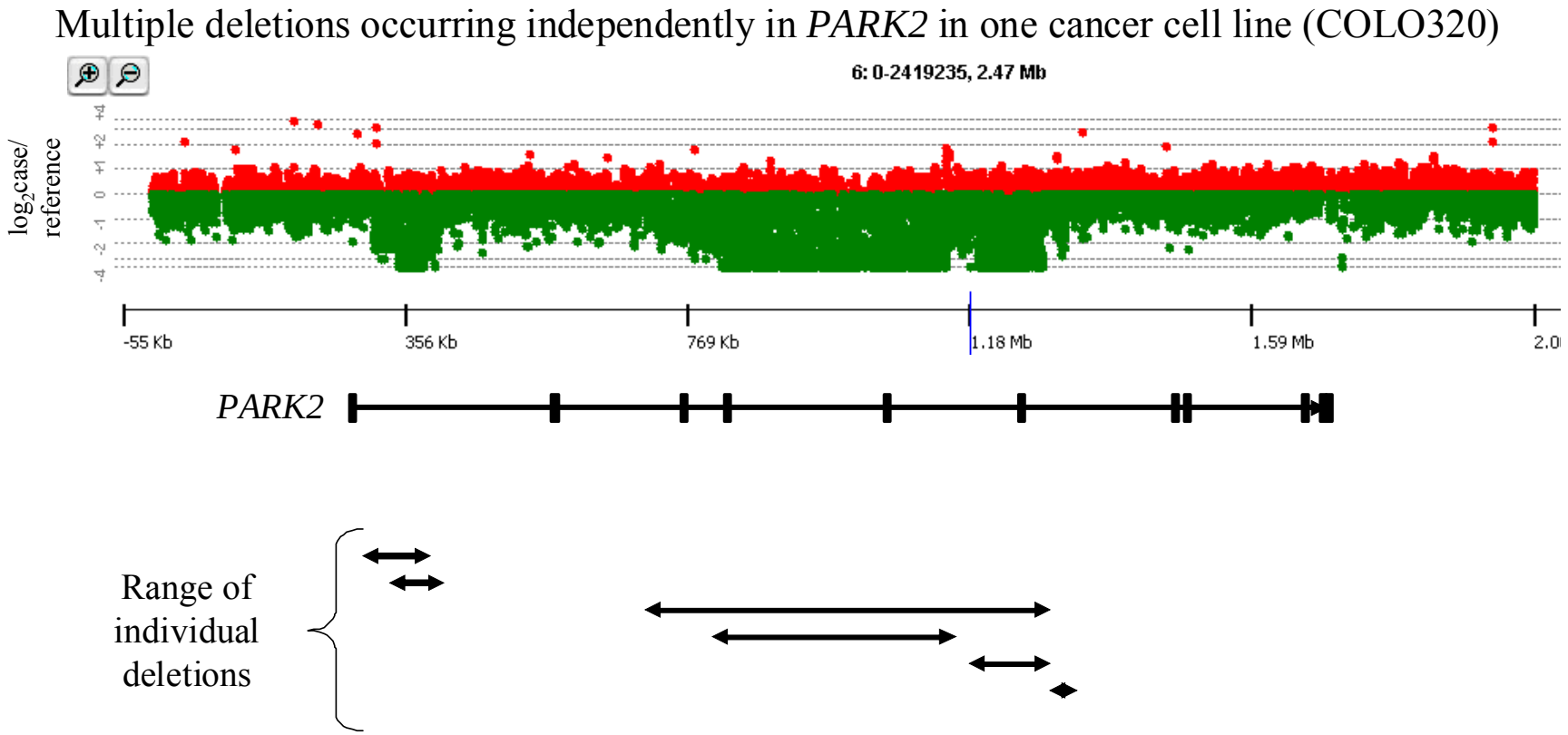
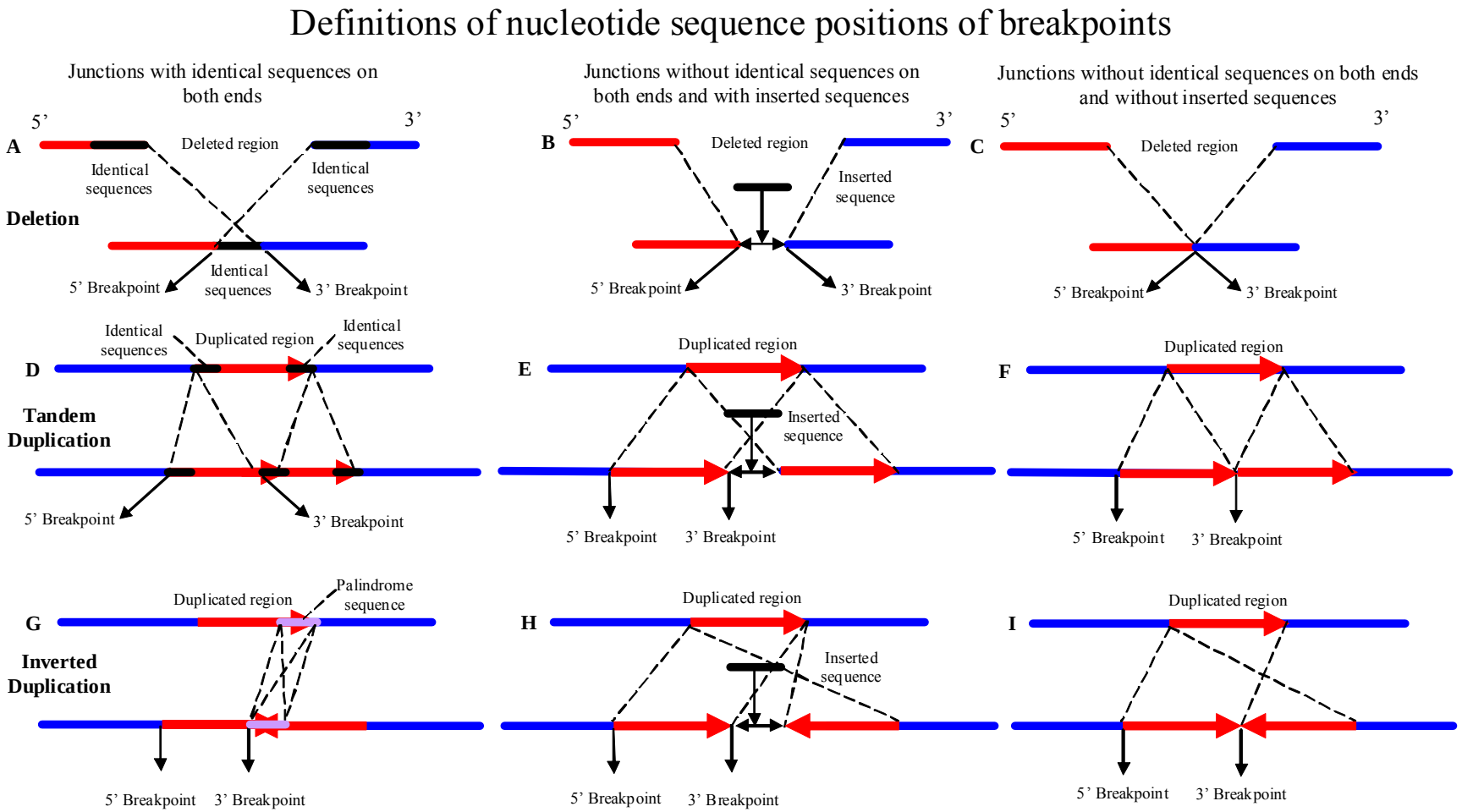


Figure S2. Definitions of Nucleotide Positions of Breakpoints and Schemes of Three Types of Junction

Scheme 1: junctions with identical sequences on both ends. Scheme 2: junctions without identical sequences on both ends and with inserted sequences. Scheme 3: junctions without identical sequences on both ends and without inserted sequences. In each scheme, nucleotide positions of breakpoints are indicated. The numbers of breakpoints in patients with AR-JP, patients with DMD/BMD, and cancer cell lines are shown below.

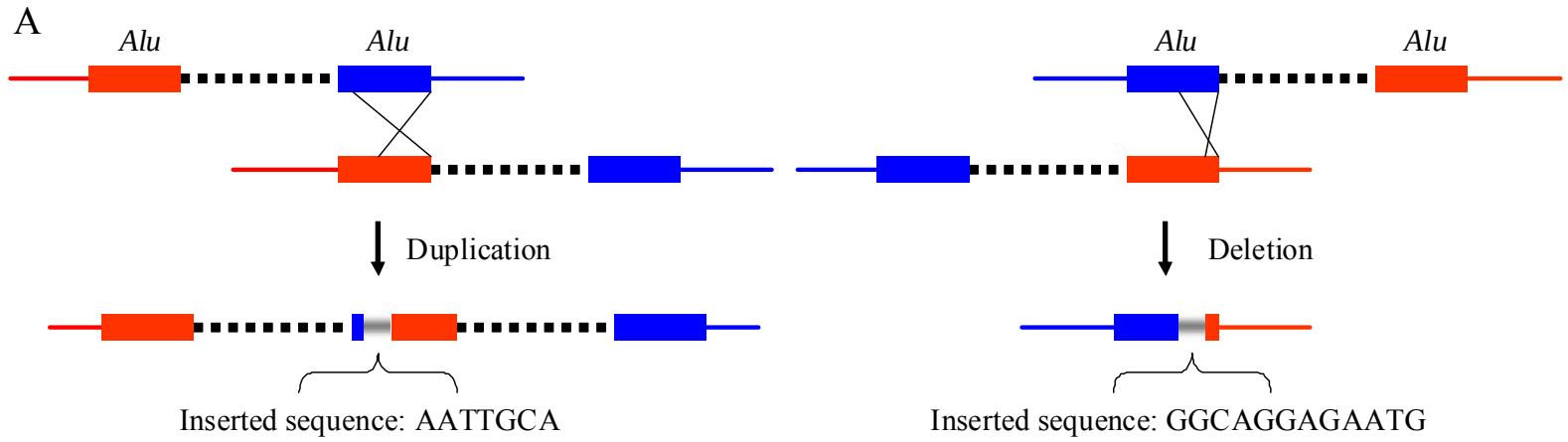


Sample sources	A	B	C	D	E	F	G	H	I	Total
PARK2 (Patients with AR-JP)	94	47	7	9	5	0	0	0	0	162
PARK2 (Cancer cell lines)	20	9	2	0	1	0	0	0	0	32
DMD (Patients with DMD/BMD)	120	48	15	9	5	0	0	0	0	197
DMD (Cancer cell lines)	3	2	1	0	0	0	0	0	0	6

Figure S3. Scheme of Junctions Forming Partial Chimeric *Alu*/*Alu* with Inserted Sequences of 7 bp and 12 bp at Their Junctions

A. Both ends of breakpoints are embedded in *Alu* sequences and partial chimeric *Alu*/*Alu* sequences are formed at the breakpoint junctions. The blue and red boxes represent *Alu* sequences of the left and right sides of junctions, respectively. Light gray boxes are deleted *Alu* sequences from chimeric *Alu*/*Alu* junction segments. There are inserted sequences in the regions of light gray boxes.

B. Specific sequences are shown below for junctions forming partial chimeric *Alu*/*Alu* sequences. The nucleotide sequences of the segments are in the same colors as those in part A. Double arrows denote connections of sequences of both sides.



B

Duplication of "Chromosome 6: 162,253,158 - 162,123,917" with inserted sequence "AATTGCA" at its junction

```
TTTATTTTTTTGAGATGGAGTCTTCGCTGTCGCCCAGGCTGGAGTGCAGTGGCATGATC
||| |||||||  <-------  ||| ||| AATTGCA ---------->  ||| |||||
TTTTTTTTTTTAGGACAGATTC-TCCTCTGTCGTCCAGGCTGGAGTGCAGTGGTGTGATC

TCGGCTCACTGCAACCTTGGCCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCCAA
||  ||||||||||||     ||||||| |||||||||  |||||||||||||||||||||||||
TCCACTCACTGCAACTGCTGCCTCCTGGGTTCAAGCAATTCTCCTGCCTCAGCCTCCCAA

GTAGCTGGGACTACAGGTGCATGCCACCATGCCCTGCTAATTTTTGTATTTTTAGTAGAG
||||  |||||  |||||||||    ||||  ||   |||  |  ||||||||||||||||||||||||
GTAGTTGGGATTACAGGTGCCCGCCATCACACCCGGGTAATTTTTGTATTTTTAGTAGAG

ATGGGGTTTCGCCATGTTAACCAGGCTGGTCTCTAACTCCTGACCTCAGGTGATCCGCCC
|  ||  |||||  ||||||||  ||||||||||||  |  ||||||||||||||||||||||  |  ||
ACGGTGTTTCACCATGTTGACCAGGCTGGTCACGAACTCCTGACCTCAGGTGATCTGTCC

ACCTCAGCCTCCCAAGGTGCTGGGATTACAGGCATGAGCCACTGCACCTGGCC
 ||||  |||||||||  |||||||||||||||||||||||||  ||     |||||||
GCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCTACCATGCCTGGCC
```

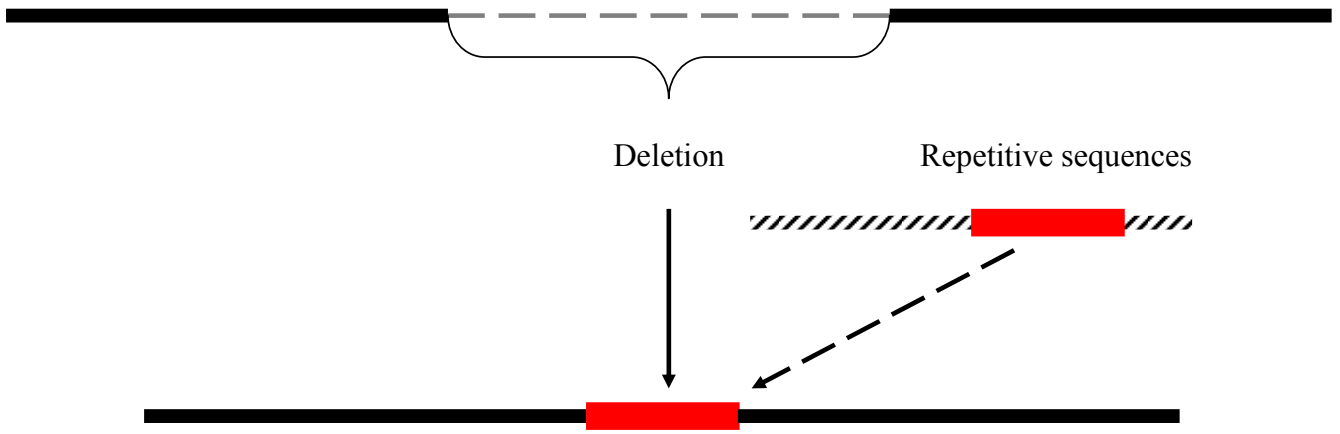
Deletion of "Chromosome X: 32,076,013 - 31,762,546" with inserted sequence "GGCAGGAGAATG" at its junction

```
GGCCGGGTGCGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCTGA
|||  |||   |||||||||  ||||  ||||||||||||||||  ||||||||||||  |||  ||
GGCTGGGCATGGTGGCTCATGCCTATAATCCCAGCACTTTGCGAGGCCGAGGCAGGCGGA

TCA--CGAGGTCAGGAGATCGAGACCATCCTGGCGAACACGGTGAAACCCCGTCTCTACT
|||   ||||||||||| ||||||||  |||||||  ||||  ||||||||||  |||||||||
TCACCTGAGGTCAGGAGTTCGAGACCTGCCTGGCCAACATGGTGAAACCCTGTCTCTACT

AAAAATATAAAAAATTAGCCTGGCGTGGTAGTGGGTGCCTGTAGTCCCAGCTACTCGGGA
|||||||  |||||  ||||  |  |||||  |  ||  |||||||  ||||||||||||  ||||
AAAAATAT-GAAAATGAGCCAGCTGTGGTGGCAGGCGCCTGTAATCCCAGCTACTTGGGA

GGCTGAGGCAGGAGAATGGCGTG------GGCAGGAGAATG------GAGATCGC
||||||||||||||||  |  ||<----                ---->||||||||
GGCTGAGGCAGGAGAATCACTTG------------------------GAGATCGC

GCCACTGCACTCCAGCCTGGGTGACA-GAGCGAGACTCCATCTCAAAA
||   ||||||||||||||||  ||||  ||||||  |||||  ||||||||
GCTGTTGCACTCCAGCCTGGGCGACACGAGCGAAACTCCGTCTCAAAA
```

Figure S4. Inserted Sequences at Deletion Breakpoint Junctions Originated from Repetitive Sequences

A. A part of repetitive sequences (shown in red box) is inserted into breakpoint junction.
B. The specific inserted sequences are shown in blue and the corresponding repetitive sequences (2 *Alu* and 1 THE1B) are shown in red.

A



B

```
  1 GGTCAGGAGATGGAGACCATCCTG-TTAACACGGTGAAACCCTGTC 45
    |||||||||||||||||||||||  ||||||||||||||||| |||
 67 GGTCAGGAGATCGAGACCATCCTGGCTAACACGGTGAAACCCCGTC 112
```

67 to 112 bp of *Alu*

```
  1 TAGTGGCTCACGTCTGTAAT 20
    ||||||||||  |||||||
 10 CGGTGGCTCACGCCTGTAAT 30
```

10 to 30 bp of *Alu*

```
  1 TTGGTGGGAGGTGATTGAATCATGGGG-CGGTTCC--CCCATACTGTTCTCG 49
    ||||||||||  ||||||||||||||| |||  ||   |||  | |||||||||
 83 --GGTGGGAGGTAATTGAATCATGGGGGCGGGTCTTTCCCGTGCTGTTCTCG 13

 50 TGGTA-----TAAGTCCCACGAGATCCGATGTGAAAGTTTTAT---GAGG 91
    ||  ||     ||||||  |||||||||  ||||     |||||||   | ||
133 TGATAGTGAATAAGTCTCACGAGATCTGATG-----GTTTTATAAAGGGG 177

 92 GTTTCTCCTTTTGCTTGGTCTCATTCTCTCCTGC-TGCTGCCATGTAAGA 140
      |||  |||        | |  |||  |||||  |||  |||  ||||||||||
178 AGTTCCCCT--------G-CACAWGCTCTCTTGCCTGCCGCCATGTAAGA 218

141 CGTGCCTTT----------CACCTTCCACCATGATCGCAGGGCCTCCCCA 180
    |||||||||          | ||||||  |||||||  |   ||||||||||
219 CGTGMCTTTGCTCCTCCTTCGCCTTCCGCCATGATTGTGAGGCCTCCCCA 268

181 GCCATGTGGAACTG--AATCCATTCAACCTCTT--CTTTATAAATTACCC 226
    ||||||||||||||  | ||||||  ||||||||  |||||||||||||||
269 GCCATGTGGAACTGTGAGTCCATTAAACCTCTTTCCTTTATAAATTACCC 318

227 AGTCTCAGGTATGT 240
    ||||||  |||||||
319 AGTCTCGGGTATGT 332
```

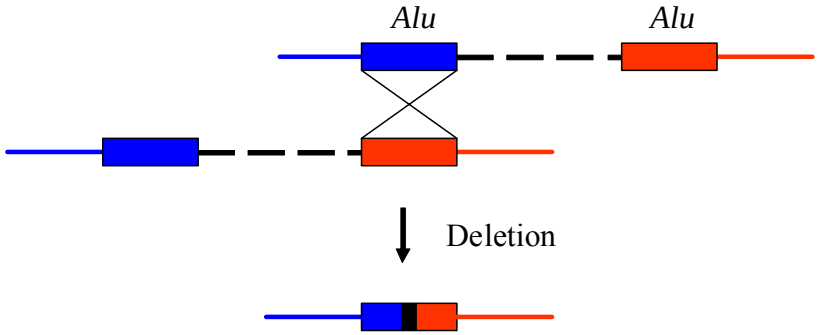
83 to 332 bp of THE1B

Figure S5. Scheme of Junction Sequences of Recurrently Observed Breakpoints #1 and #14

A. Both ends of the breakpoints are embedded in *Alu* sequences and fully chimeric *Alu*/*Alu* sequences are formed at the breakpoint junctions. The blue and red lines denote sequences of the left and right side, respectively. The blue and red boxes represent *Alu* sequences, and the black box represents identical sequences at the junction.

B. Specific sequences are shown below for recurrently observed breakpoints #1 (left) and #14 (right). The nucleotide sequences of segments are the same colors as those in part A

A



B

Deletion of "Chromosome 6: 162,506,819 – 162,503,760" with identical sequence "GATTACAGGCATGAGCCACC" at its junction

```
TCTTTTTTTTTTTTTTTTTTTTTTTTTGAGATAGAGTCTCGCTCTGTTGCCCAGGCTGGAG
| || ||||| ||| ||| |||||||||||| ||| |||||||||||| ||||||||||||||
TTTTCTTTTTATTTATTTATTTTTTTGAGACAGAATCTCGCTCTGTCGCCCAGGCTGGAG

TGCAGTGGCATGACCCTGGCTCACTGCAACCCTTGCCACCTGGGTTCAAGCAGTTCTCCT
| |||||||| || |  ||||||||||||||  |||| || |||||||||  |||| ||
TACAGTGGCACGATCTCGGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGTGATTCTTCT

GTTTCAGCCTCCTGAGTAGCTGAGACTACAGGTGCCCACCACCACACCCGGCTAATTTTT
|  ||||||||| ||||||||| |||||||||||||||||||| ||| |||  || ||||||
GCCTCAGCCTCCCGAGTAGCTGGGACTACAGGTGCCCACCATCACGCCCCACTGATTTTT

GTA-TTTTAGTAGAGAAGGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACCCCTGACC
||| |||||||||||| |||||||||||||| |||||||||||||||| |||| ||| |||
GTATTTTTAGTAGAGATGGGGTTTCACCATATTGGCCAGGCTGGTCTCGAACTCCTTACC

TAAGGGGATCCGCCTGCCTCGGCCTCCCAAAGTGCTGAGATTACAGGCATGAGCCACCAC
|   | ||||||||||||||||||||||||||||||||| |||||||||||||||||||| |
T--TGTGATCCGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGC

ACTCGGCC
|| |||||
ACCCGGCC
```

Deletion of "Chromosome 6: 162,461,205 – 162,365,391" with identical sequence "AAAAATA" at its junction

```
GGCCAAGCACAGTGGCTCACGCCTATAATCCCAATGCTTTGGGAGGATGAAGCAGGAGAA
|||   || | ||||||||||||||||||||||   ||||| |||| || || || | |
GGCTGGGCGCTGTGGCTCACGCCTATAATCCCAGCACTTTGAGAGGCCGAGGCCGGTGGA

TTGCTTGAGCCCAGGAGTTCCATACCAGCCTTAGCAACAGGGTGAAACCTTGTCTCTGCA
|  | |||   ||||||||  | |||||||| | |||||  |  ||||| ||||||| |
TCACCTGAAGTCAGGAGTTTGAGACCAGCCTGACCAACATAGCAAAACCCTGTCTCTAC-

AAAAAAATAAAAAAATTAGCTGGGTGTGCTTG----TG------GTCCCAGTTA-CT---C
 |||||||  |||||||||||||||  || | |    ||        ||||| || ||
TAAAAAATACAAAAATTAGCTGGGCATGTTGGCACATGCAATTACCCCCAGCTACCTGGGG

GTCGATGTGGGAGGATCACTGGAGCCTGAGAAGTTGAGGCTGCAGTGAGCCACAGTCTCG
|  || |  |||| ||  || || |||| || |  |||| |||||||| |||   |   |
GCTGAAGCAGGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAACCAAGATTGTG

CCACTGAGCTCCAGTCTGGATGACAGAGCAAGACCCT
| | ||  ||| || ||||   ||||||  | || ||
CTATTGCACTCTAGCCTGGGCAACAGAGTGAAACTCT
```

Figure S6. Schematic Representation of Physical Relationships between Genes Located in CFSs and large LADs

Physical maps of genes located in CFSs (*FHIT*, *WWOX*, *GRID2*, and *LARGE*) and chromosomal R/G bands are shown. Box filled with diagonal lines denotes position of LADs. The genes located in CFSs are embedded in large LADs spanning several Mb (approximately 1.7 to 4.7 Mb)

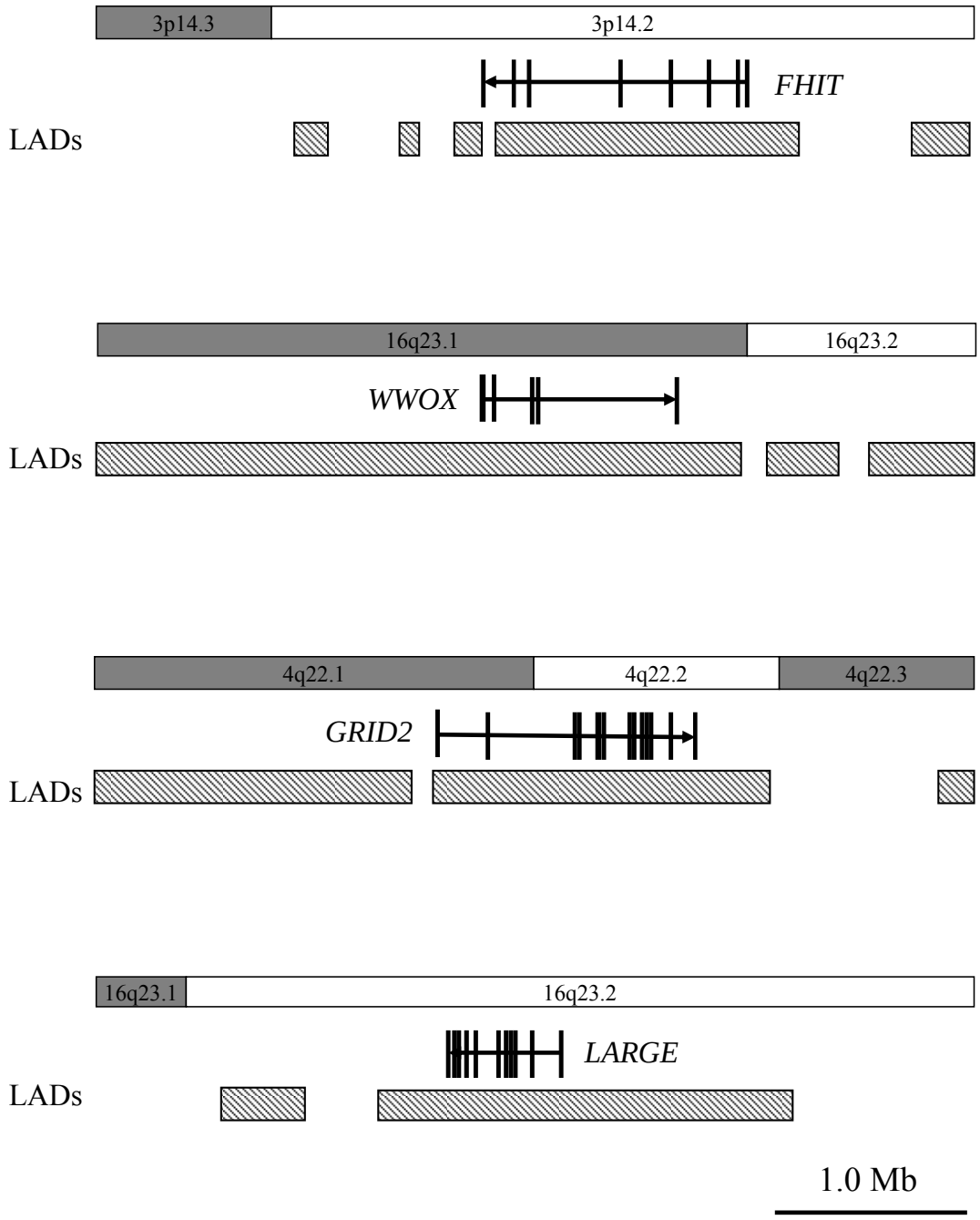


Table S1. List of Cancer Cell Lines

Cell line	Histology
22Rv1	Prostate cancer
A2058	Melanoma
A431	Skin cancer
A549	Lung cancer
AGS	Gastric cancer
Alex	Liver cancer
AZ-521	Gastric cancer
BT-474	Breast cancer
BT-483	Breast cancer
C33A	Cervical cancer
CaCo-2	Colon cancer
CAKI2	Kidney cancer
CAPAN-1	Pancreatic cancer
COLO320	Colon cancer
DLD1	Colon cancer
DU145	Prostate cancer
G-361	Melanoma
H1299	Lung cancer
H157	Lung cancer
H2009	Lung cancer
H209	Lung cancer
H2122	Lung cancer
H2347	Lung cancer
H292	Lung cancer
H460	Lung cancer
H522	Lung cancer
HCC1008	Breast cancer
HCC1143	Breast cancer
HCC1187	Breast cancer
HCC1395	Breast cancer
HCC1599	Breast cancer
HCC1937	Breast cancer
HCC1954	Breast cancer
HCC202	Breast cancer
HCC2157	Breast cancer
HCC2218	Breast cancer
HCC38	Breast cancer
HCT116	Colon cancer
HepG2	Liver cancer
HL60	Leukemia
HLC-1	Lung cancer
HLF	Liver cancer
Hs578T	Breast cancer
HT1080	Fibrosarcoma
HT1197	Bladder cancer
HT1376	Bladder cancer
HT17	Liver cancer
HT29	Colon cancer
HTB49	Kidney cancer
HuH28	Bile duct cancer
HuH6	Liver cancer
HuH7	Liver cancer
J82	Bladder cancer
JMSU3	Bladder cancer
K-562	Leukemia
KATOIII	Gastric cancer
KLM-1	Pancreatic cancer
LNCaP clone FGC	Prostate cancer
LOVO	Colon cancer
MCF7	Breast cancer
MDA PCA 2b	Prostate cancer
MDA468	Cervical cancer
MDA-MB-175-VII	Breast cancer
MDA-MB-231	Breast cancer
MDA-MB-361	Breast cancer
MDA-MB-415	Breast cancer
MDA-MB-435S	Breast cancer
ME180	Uterus cancer
MKN-1	Gastric cancer
MKN-45	Gastric cancer
MKN-45	Gastric cancer
NCI-H1395	Lung cancer
NCI-H2126	Lung cancer
NCI-H358	Lung cancer
NCI-H526	Lung cancer
NCI-H69	Lung cancer
NMCG1	Glioblastoma
NUGC3	Gastric cancer
NUGC4	Gastric cancer
OCUM-2MD3	Gastric cancer
OV167	Ovarian cancer
OV202	Ovarian cancer
OVCAR5	Ovarian cancer
PANC-1	Pancreatic cancer
PC-3	Prostate cancer
PK-0	Pancreatic cancer
PK-1	Pancreatic cancer
PK-45H	Pancreatic cancer
PK-45P	Pancreatic cancer
PK-59	Pancreatic cancer
PK-8	Pancreatic cancer
RPMI8226	Myeloma
RT4	Bladder cancer
Saos-2	Osteosarcoma
SCH	Choriocarcinoma
SF188	Glioblastoma
SF763	Glioblastoma
SJSA-1	Osteosarcoma
SK-BR-3	Breast cancer
SK-MEL-28	Melanoma
SKOV8	Ovarian cancer
SW480	Colon cancer
T-47D	Breast cancer
T84	Colon cancer
T98G	Glioblastoma
TCC-SUP	Bladder cancer
TCS	Uterus cancer
TE10	Esophagus cancer
TE11	Esophagus cancer
TE14	Esophagus cancer
TE4	Esophagus cancer
TE6	Esophagus cancer
TE8	Esophagus cancer
TE9	Esophagus cancer
TGW	Neuroblastoma
THP-1	Leukemia
TOV-112D	Ovarian cancer
U343	Glioblastoma
U87	Glioblastoma
UACC812	Breast cancer
UACC893	Breast cancer
VMRC-RCZ	Kidney cancer
WM-115	Melanoma
ZR-75-1	Breast cancer
ZR-75-30	Breast cancer

Table S2. Results of Sequence Motif Analyses

Motif name	Motif sequence	Breakpoint regions				Control sequences	
		PARK2 (n=384)	%	*DMD* (n=406)	%	(n=5,000)	%
X-element *E.coli*	GCTGGTGG	5	1.3	2	0.5	36	0.7
Ade6-M26	ATGACGT	2	0.5	5	1.2	26	0.5
ARS consensus *S. cerevisiae*	WTTTATRTTTW	7	1.8	9	2.2	96	1.9
ARS consensus *S. pombe*	WRTTTATTTAW	4	1.0	2	0.5	100	2.0
Consensus scaffold attachment region 1	AATAAAYAAA	7	1.8	8	2.0	198	4.0
Consensus scaffold attachment region 2	TTWTWTTWTT	207	53.9	249	61.3	3362	67.2
Consensus scaffold attachment region 3	WADAWAYAWW	181	47.1	236	58.1	2786	55.7
Consensus scaffold attachment region 4	TWWTDTTWWW	393	102.3	469	115.5	6292	125.8
Deletion hotspot consensus	TGRRKM	728	189.6	783	192.9	9626	192.5
DNA polymerase arrest site	WGGAG	458	119.3	349	86.0	5962	119.2
DNA polymerase a frameshift hotspot 1	TCCCCC	38	9.9	30	7.4	308	6.2
DNA polymerase a frameshift hotspot 2	CTGGCG	9	2.3	6	1.5	68	1.4
DNA polymerase b frameshift hotspot 1	ACCCWR	173	45.1	146	36.0	2030	40.6
DNA polymerase a/b frameshift hotspot 1	ACCCCA	28	7.3	49	12.1	476	9.5
DNA polymerase a/b frameshift hotspot 2	TGGNGT	156	40.6	145	35.7	2244	44.9
Drosophila topoisomerase II consensus	GTNWAYATTNATNNR	4	1.0	2	0.5	20	0.4
Heptamer recombination signal	CACAGTG	20	5.2	15	3.7	268	5.4
Human hypervariable minisatellites sequence 1	GGAGGTGGGCAGGARG	0	0.0	0	0.0	0	0.0
Human hypervariable minisatellites sequence 2	AGAGGTGGGCAGGTGG	0	0.0	0	0.0	0	0.0
Human minisatellites core sequence	GGGCAGGARG	2	0.5	0	0.0	8	0.2
Human replication origin consensus	WAWTTDDWWWDHWGWHMAWTT	0	0.0	0	0.0	8	0.2
Human minisatellites conserved sequence	GCWGGWGG	12	3.1	10	2.5	252	5.0
Ig heavy chain class switch repeat 1	GAGCT	144	37.5	127	31.3	1702	34.0
Ig heavy chain class switch repeat 2	GGGCT	101	26.3	77	19.0	1420	28.4
Ig heavy chain class switch repeat 3	GGGGT	95	24.7	90	22.2	1238	24.8
Ig heavy chain class switch repeat 4	TGGGG	144	37.5	150	36.9	2108	42.2
Ig heavy chain class switch repeat 5	TGAGC	173	45.1	143	35.2	2252	45.0
LTR-IS motif	TGGAAATCCCC	0	0.0	0	0.0	2	0.0
Mariner transposon-like element	GAAAATGAAGCTATTTACCCAGGA	0	0.0	0	0.0	0	0.0
Murine MHC recombination hotspot	CAGRCAGR	18	4.7	16	3.9	288	5.8
Murine parvovirus recombination hotspot	CTWTTY	334	87.0	405	99.8	4894	97.9
Nonamer recombination signal	ACAAAAACC	1	0.3	1	0.2	6	0.1
Pur binding site	GGNNGAGGGAGARRRR	0	0.0	1	0.2	4	0.1
Recombination hotspot	CCNCCNTNNCCNC	10	2.6	4	1.0	110	2.2
Retrotransposon	TCATACACCACGCAGGGGTAGAGGACT	0	0.0	0	0.0	0	0.0
Translin binding site 1	ATGCAG	42	10.9	48	11.8	660	13.2
Translin binding site 2	GCCCWSSW	29	7.6	25	6.2	458	9.2
Vaccinia topoisomerase I consensus	YCCTT	368	95.8	419	103.2	5186	103.7
Vertebrate topoisomerase II consensus	RNYNNCNNGYNGKTNYNY	3	0.8	3	0.7	82	1.6
XY32 homopurine-pyrimidine H-palindrome motif	AAGGGAGAARGGGTATAGGGRAAGAGGGAA	0	0.0	0	0.0	0	0.0

Table S3. Result of the Analyses of AT Content and Flexibility Peak

Region of interest	Position	AT content (%)	Average twist angle (degree)	Flexibility peak (peaks/Mb) [a]	Unified peak (peaks/Mb) [b]	Cluster of peaks (peaks/Mb) [c]
	chromosome 6: 159,870,000-160,370,000	56.70	10.66	40	26	4
	chromosome 6: 160,370,000-160,870,000	56.90	10.70	32	18	4
	chromosome 6: 160,870,000-161,370,000	59.43	10.81	44	36	4
	chromosome 6: 161,370,000-161,870,000	56.98	10.68	52	32	8
	chromosome 6: 161,870,000-162,370,000	59.16	10.85	74	44	10
Breakpoint clustering region	chromosome 6: 162,370,000-162,870,000	60.66	10.90	50	24	0
	chromosome 6: 162,870,000-163,370,000	61.86	10.96	64	32	8
	chromosome 6: 163,370,000-163,870,000	58.61	10.74	32	24	4
	chromosome 6: 163,870,000-164,370,000	57.59	10.72	36	26	4
	chromosome 6: 164,370,000-164,870,000	60.67	10.87	104	46	12
	chromosome 6: 164,870,000-165,370,000	62.64	11.03	88	50	12
	chromosome X: 29,000,000-29,500,000	63.70	11.08	116	56	18
	chromosome X: 29,500,000-30,000,000	63.22	11.07	72	50	10
	chromosome X: 30,000,000-30,500,000	61.32	10.92	36	24	4
	chromosome X: 30,500,000-31,000,000	59.65	10.81	56	38	6
	chromosome X: 31,000,000-31,500,000	61.19	10.89	34	30	0
Breakpoint clustering region	chromosome X: 31,500,000-32,000,000	63.96	11.07	52	40	2
	chromosome X: 32,000,000-32,500,000	64.69	11.15	100	62	14
	chromosome X: 32,500,000-33,000,000	63.76	11.12	80	56	8
	chromosome X: 33,000,000-33,500,000	64.18	11.13	100	66	18
	chromosome X: 33,500,000-34,000,000	64.44	11.11	78	42	12
	chromosome X: 34,000,000-34,500,000	63.77	11.08	82	50	12

[a] Average twist angles ≥ 13.7 were considered.

[b] Distance between adjacent peaks ≤ 100 bp was unified.

[c] Equal or more than 3 unified peaks whose distance between adjacent unified peaks are smaller than 5,000 bp.