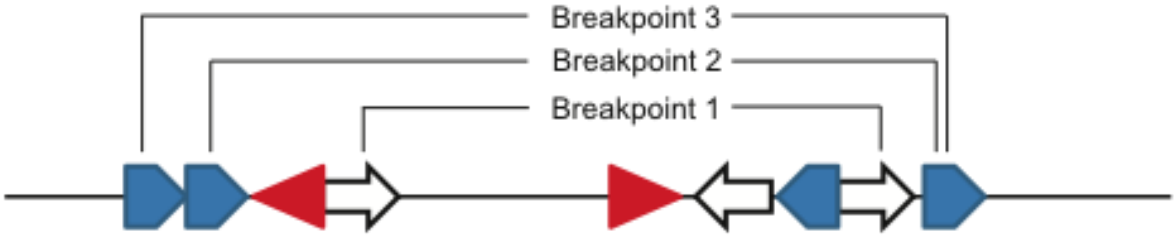
The American Journal of Human Genetics, Volume 90

Supplemental Data

Resolving the Breakpoints of the 17q21.31 Microdeletion Syndrome with Next-Generation Sequencing

Andy Itsara, Lisenka E.L.M. Vissers, Karyn Meltz Steinberg, Kevin J. Meyer, Michael C. Zody, David A Koolen, Joep de Ligt, Edwin Cuppen, Carl Baker, Choli Lee, Tina Graves, Richard K. Wilson, Robert B. Jenkins, Joris A. Veltman, and Evan E. Eichler

A



Relative CN after NAHR-mediate deletion

SD	Total copies, no deletion	Breakpoint 1	Breakpoint 2	Breakpoint 3
	3	1/3	0	0
	4	3/4	2/4	1/4
	2	1/2	0	0

B

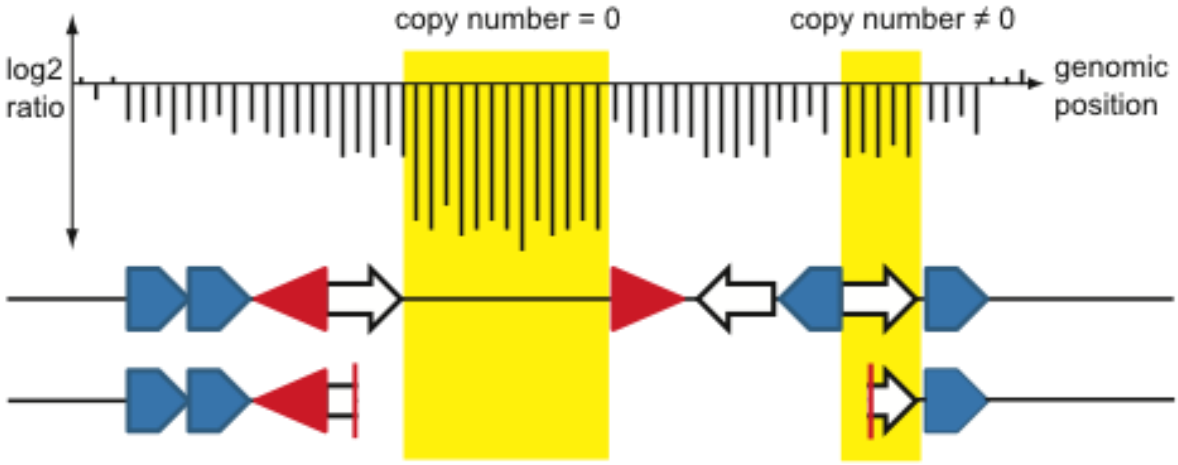


Figure S1 Overview of analysis to delineate breakpoints within SDs using array CGH in somatic cell hybrids.

(a) A hypothetical region is shown in which SDs (red, white, and blue arrows) may mediate deletions by NAHR. Because NAHR-mediated deletions require SDs in direct orientation, breakpoints occur in one of three pairs of SDs. Prior to deletion, there are initially three copies of the white SD, four copies of the blue SD, and two copies of the red SD (see figure table). Deletions mediated by NAHR breakpoints 1, 2, or 3 result in distinct patterns of relative copy number (CN). (b) Simple visualization of array CGH data fails to localize the CNV breakpoints. However, array CGH signal at the white SD does not appear as copy number 0, implying that the array CGH data is most consistent with breakpoint 1.

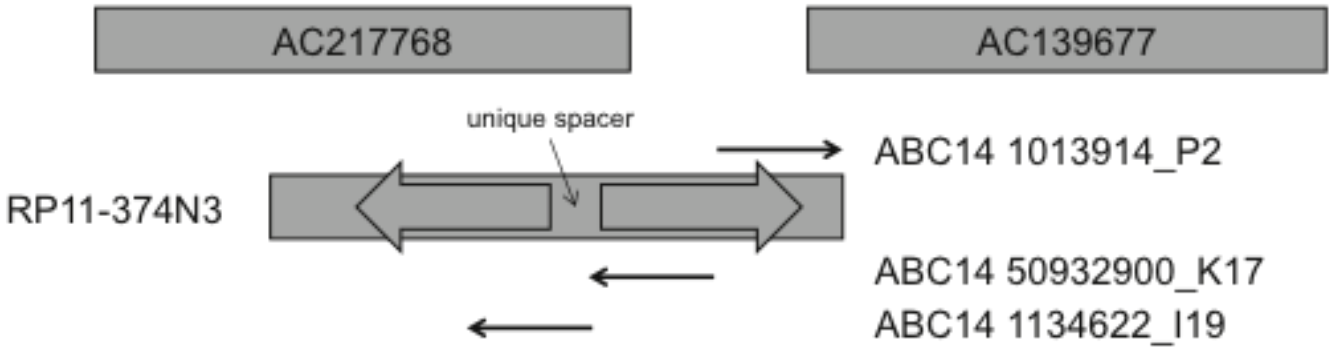


Figure S2 Strategy for closing the proximal gap in the H2 assembly.
The proximal gap (Gap 1) in the H2 assembly between AC217768 and AC139677 was spanned by RP11-374N3, but RP11-374N3 could not be assembled by shotgun sequencing alone. This was due to two arms of oppositely oriented, highly identical sequence separated by a ~12 kbp spacer of sequence unique within the clone. In order to finish RP11-374N3 and close the gap in the H2 assembly, we chose three fosmids from the ABC14 library (derived from an H1/H2 individual of European descent, NA12156). Two fosmids have end-sequences in the spacer sequence and extend either proximally or distally (1134622_I19 and 50932900_K17, respectively). A third fosmid (1013914_P2) extended proximally from AC139677 into the sequence gap and overlapped the distal end of RP11-374N3. Arrows roughly indicate fosmid size and the direction of clone insert amplification by the pCC2FOS forward primer.

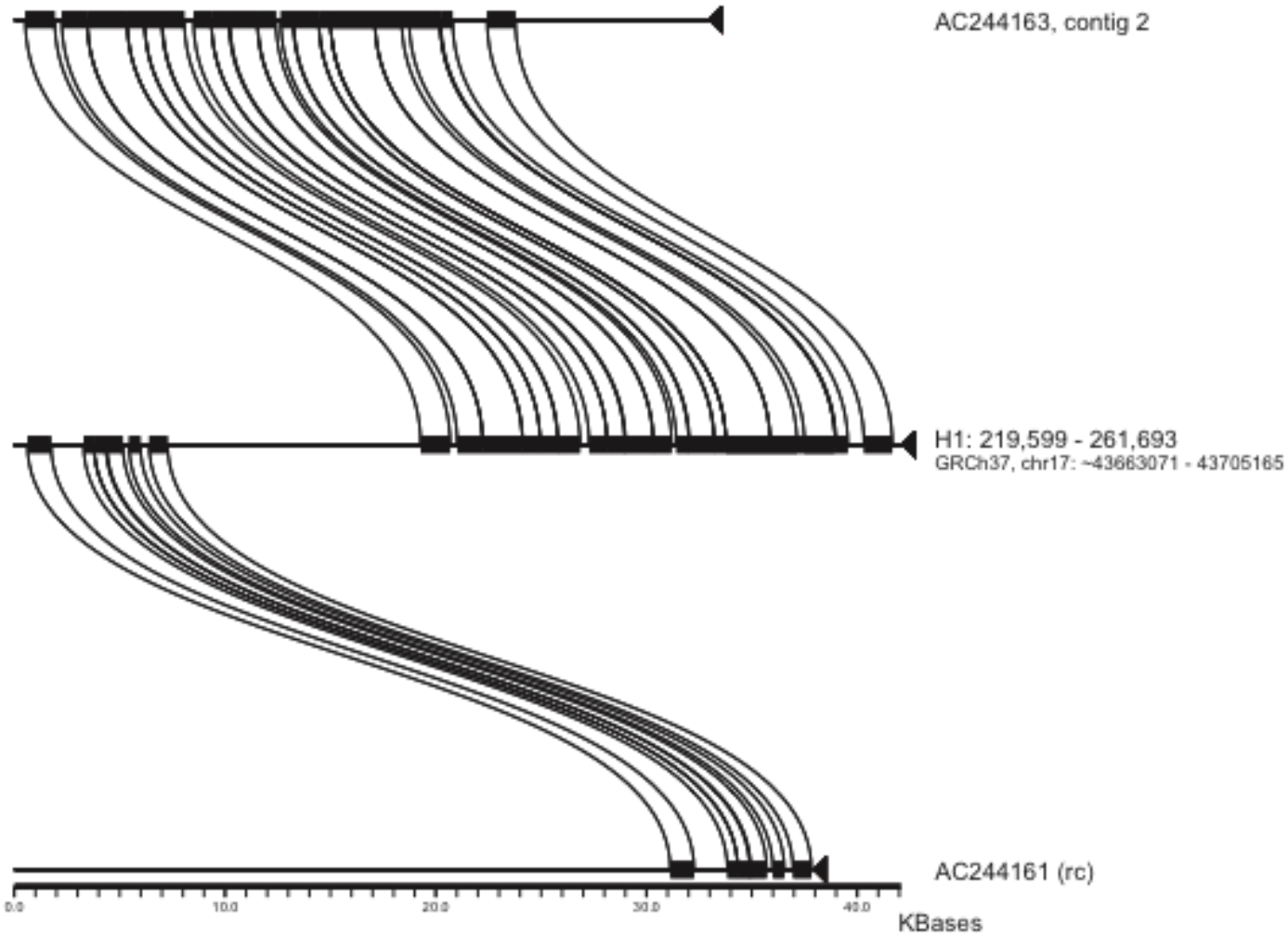


Figure S3 Alignment of Gap 1 clones to paralogous sequence on the H1 haplotype.
50932900_K17 (AC244161) and 1013914_P2 (AC244163) were inferred to overlap a ~40 kbp region of paralogous sequence to H1:219599-261693 (GRCh37, chr17:~43663071-43705165) based on end-sequence alignments. Consistent with this, miropeats alignments (threshold = 400 [1]) of draft sequence demonstrate contiguous alignments of AC244161 and AC244163 to the proximal and distal ends of the inferred region of paralogy, respectively. rc: reverse-complement.

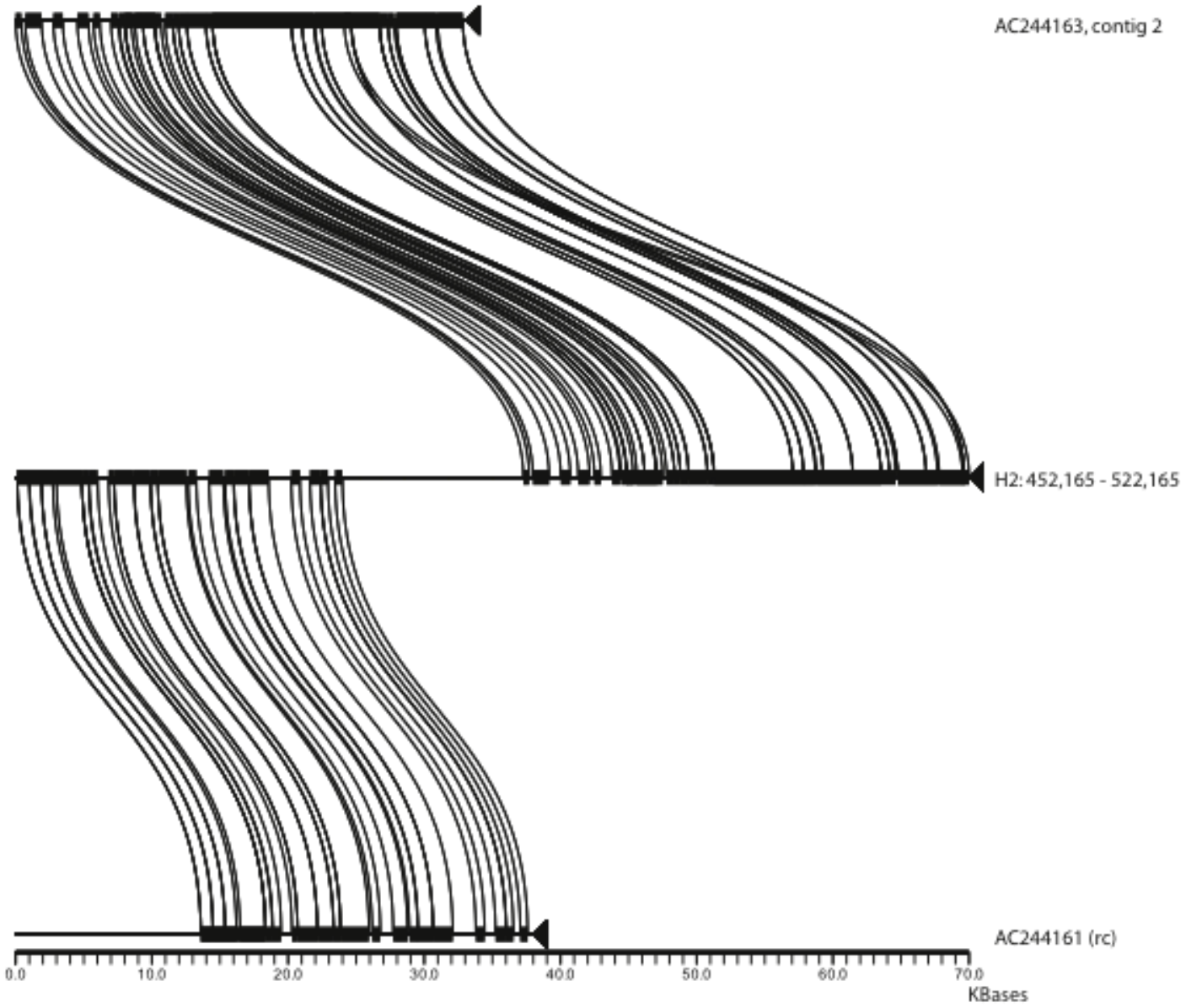


Figure S4 Alignment of Gap 1 clones to paralogous sequence on the H2 haplotype.
50932900_K17 (AC244161) and 1013914_P2 (AC244163) were inferred to overlap a ~70 kbp region of paralogous sequence to H2:452165-522165 based on end-sequence alignments. Consistent with this, miropeats alignments (threshold = 400) of draft sequence demonstrate contiguous alignments of AC244161 and AC244163 to the proximal and distal ends of the inferred region of paralogy, respectively. rc: reverse complement.

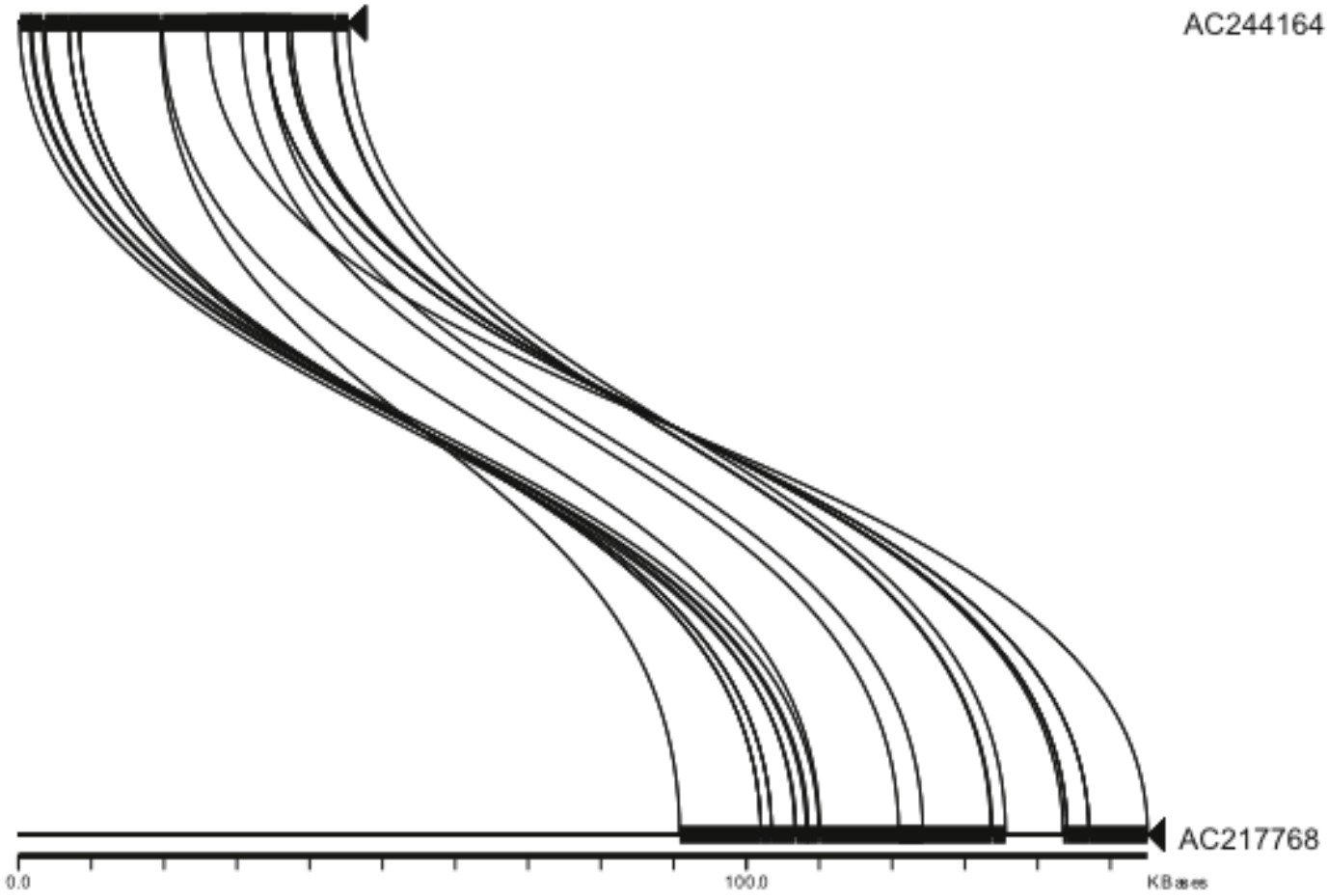


Figure S5 Alignment of AC244164 to paralogous sequence on the H2 haplotype.
Based on end-sequence alignments, 1134622_I19 (AC244164) was predicted to map to finished sequence entirely contained within AC217768 on the H2 haplotype assembly. Consistent with this, miropeats alignment (threshold = 400) for all of AC244164 mapped to AC217768.

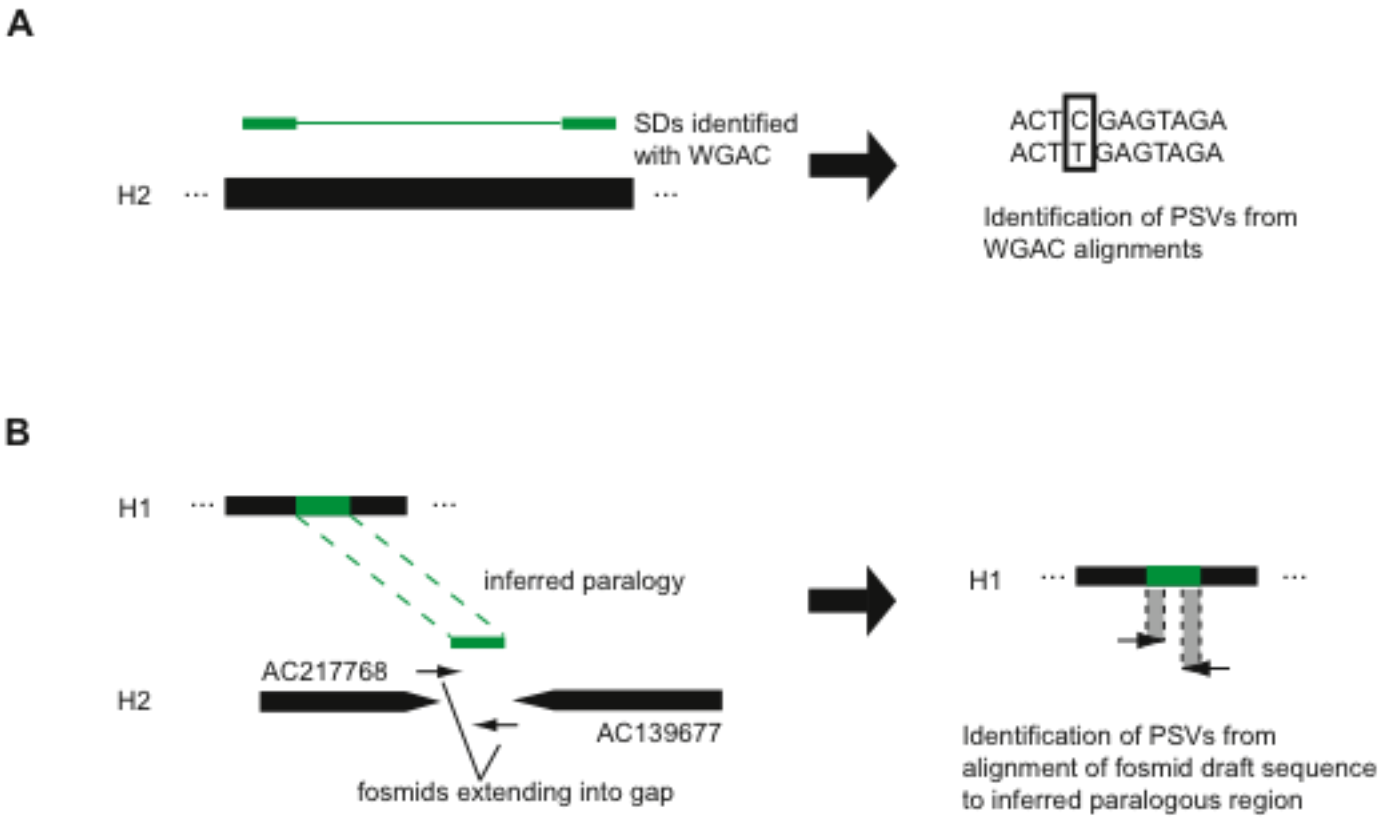


Figure S6 Schematic of strategies to identify paralogous sequence variation within SDs in the H2 assembly.
(a) For regions of the H2 assembly for which reference sequences exist, identification of paralogous sequence variants (PSVs) may be derived from whole-genome assembly comparison (WGAC) alignments. (b) In Gap 1 of the H2 assembly, PSVs were identified from draft assemblies of fosmid clones mapping into the region aligned to regions of inferred, directly oriented paralogy on the H1 and H2 haplotypes.

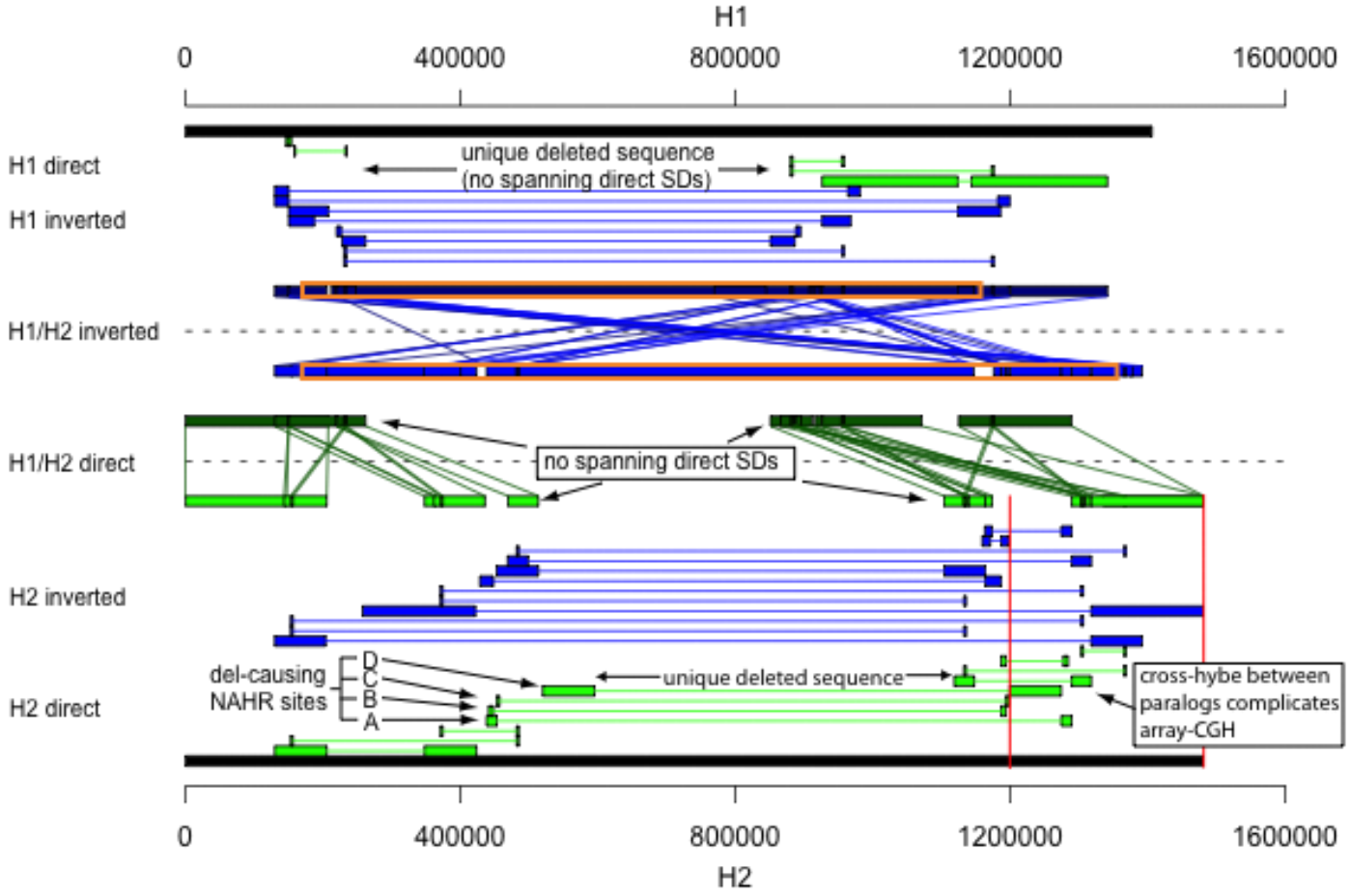


Figure S7 Segmental duplication structure of the H1 and H2 haplotypes.
The relative extent (black bars) of the current H1 and H2 assemblies reveals the H2 haplotype is slightly larger (~80 kbp) than the H1 in addition to containing two sequence gaps (red lines) owing to additional sequences of SDs. Intrachromosomal SDs in direct orientation (green blocks connected by green lines) and inverted orientation (blue blocks connected by blue lines) have been plotted on separate lines, highlighting potential sites of NAHR. Unlike the H1 haplotype, there are SDs in direct orientation on the H2 haplotype flanking the ~540 kbp of unique sequence lost in recurrent 17q21.31 deletions (labeled breakpoints A-D). H1/H2 interchromosomal alignments in inverted orientation (dark blue boxes connected to blue boxes by lines) illustrate the ~970 kbp inversion (orange boxes) between the two haplotypes. Examining H1/H2 interchromosomal alignments in direct orientation (dark green boxes connected to green boxes by lines), the current H2 assembly does not predict sequence alignments in direct orientation spanning the ~540 kbp of unique sequence. An example region has been labeled illustrating how simple breakpoint delineation by array CGH on the H2 haplotype is complicated by cross-hybridization between sequence paralogs.

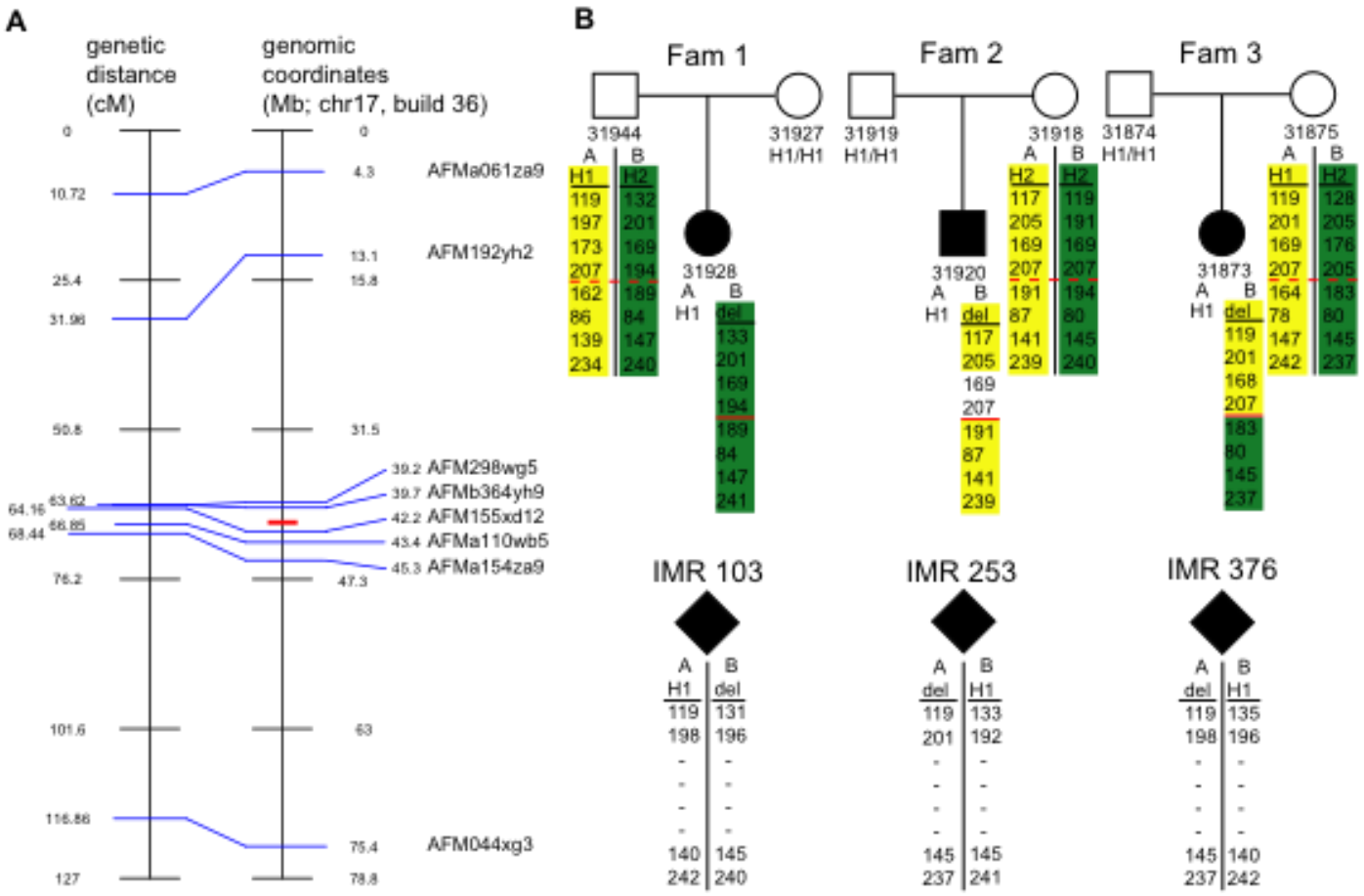


Figure S8 Microsatellite genotyping of somatic cell hybrids.
For each of three probands and the transmitting parent, two hybrid cell lines (A and B) were generated for each chromosome 17 homologue. (a) Comparison of genetic (Marshfield Clinic) and physical distance (build 36) of genotyped microsatellites on chromosome 17. Genotyped loci include those spread throughout the chromosome confirming separation of intact chromosome homologues during the creation of somatic cell hybrids as well as loci immediately flanking the deletion locus (red bar). (b) H1/H2 genotypes as determined by a 238 bp deletion in intron 9 of *MAPT* are shown along with microsatellite genotypes ordered as in (a). In two of three families examined (Fam 1 and Fam 2), the *de novo* 17q21 deletion (relative location indicated by dotted/solid red line in parents/probands) arises on the H2 haplotype and is likely a result of intrachromosomal NAHR. In Fam 3, crossover between the H1 and H2 haplotypes and the 17q21.31 deletion co-occur within a genetic distance of less than 0.54 cM (~1.32 cM as determined by HapMap), suggesting potential interchromosomal NAHR between the H1 and H2 haplotypes.

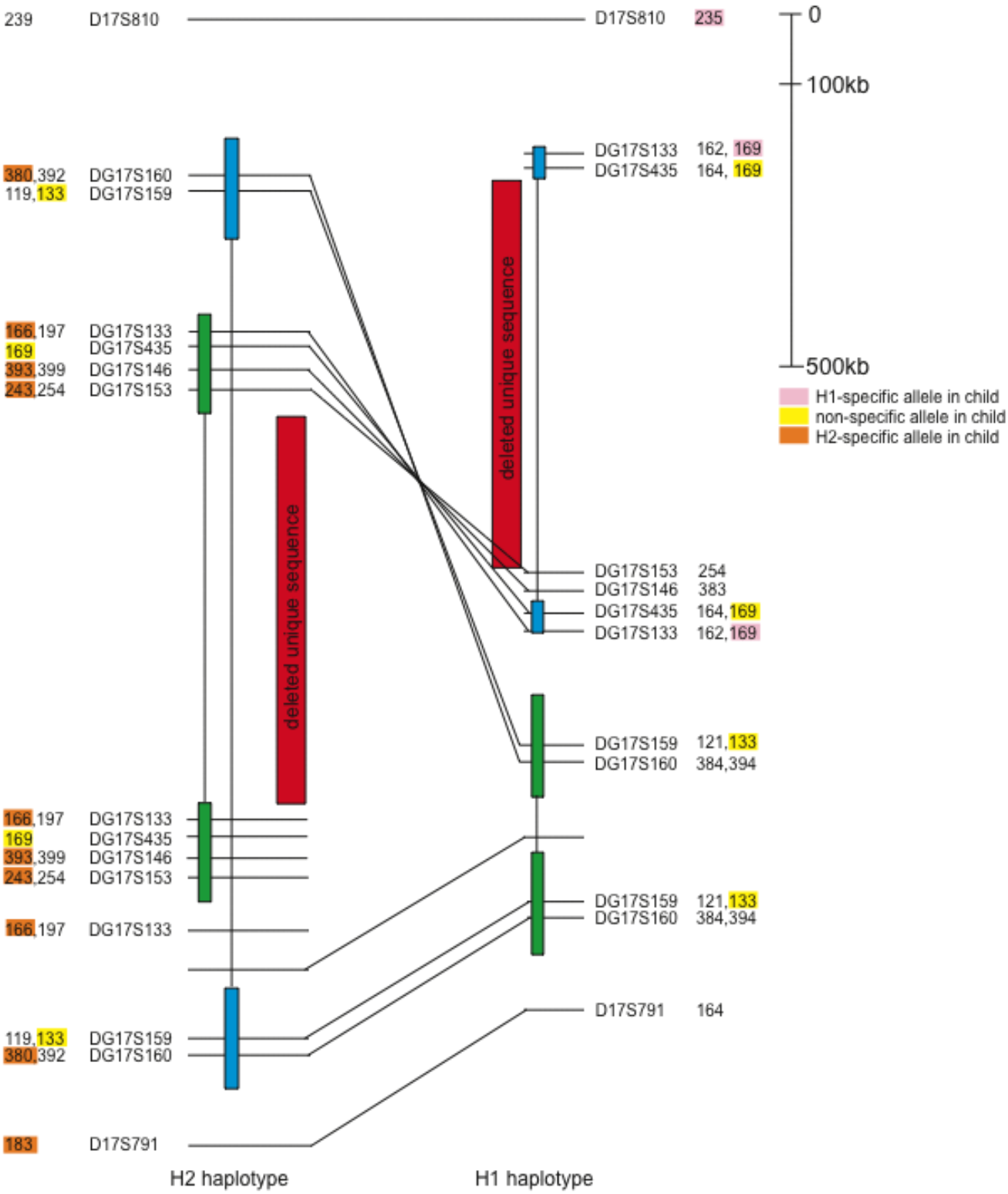


Figure S9 Microsatellite genotyping within SDs flanking 17q21.31 deletions in Fam 3 are consistent with H1/H2 NAHR and not H2/H2 NAHR.

Microsatellites within the 17q21.31 deletion locus, including those within SDs, were genotyped in hybrid cell lines of the H1 chromosome in Fam 3 mother, H2 chromosome in Fam 3 mother, and the H2-deletion

chromosome in Fam 3 proband. The relative position of each of the markers on the H1 and H2 haplotypes are shown based on positions previously reported by Stefansson *et al.*[2] with SDs in direct (green) and inverted (blue) orientation depicted as blocks connected with lines. On the H2 haplotype, intrachromosomal H2/H2 NAHR is hypothetically due to the directly oriented SDs flanking the unique deleted sequence (red block). Lines have been drawn between the two haplotypes demonstrating the inversion between the two sequences. Marker genotypes of H1 and H2 chromosomes from Fam 3 mother are shown next to their corresponding loci. For SDs, multiple genotypes are listed due to the inability to distinguish between paralogs. Marker genotypes observed in the 17q21.31 deletion-bearing chromosome of Fam 3 proband have been highlighted indicating H1-specific (pink), H2-specific (orange), or nonhaplotype-specific alleles (yellow). The H1-specific allele at DG17S133, just proximal to deleted unique sequence, is consistent with H1/H2 NAHR. Conversely, we observe a loss of heterozygosity at DG17S159 and DG17S160, although sequences containing these markers are not predicted to be lost during H2/H2 NAHR.

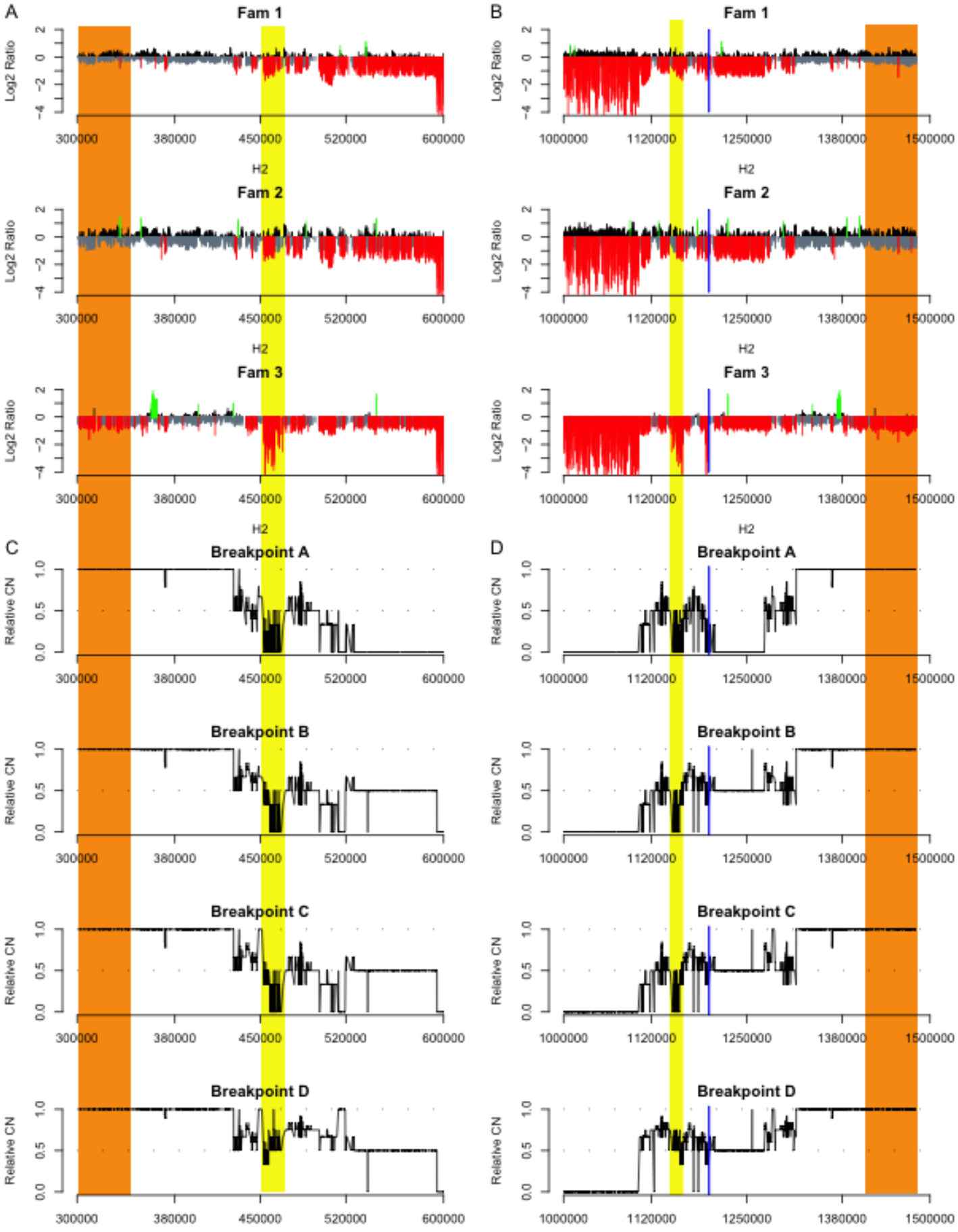

A
Fam 1
Log2 Ratio
H2
Fam 2
Fam 3
B
C
Breakpoint A
Relative CN
Breakpoint B
Breakpoint C
Breakpoint D
D
300000
380000
450000
520000
600000
1000000
1120000
1250000
1380000
1500000

Figure S10 Inference of specific SDs mediating NAHR through modeling of expected changes in relative copy number.

Relative gain (black), loss (gray), and gains and losses >3 standard deviations beyond the chromosome 17 mean (green and red, respectively) are plotted versus H2 contig position. The relative location of Gap 1 is indicated by a vertical blue line. Observed $\log_2$ ratios at the proximal (a) and distal (b) breakpoints of 17q21.31 deletions are shown versus expected changes in relative copy number ((c) and (d), respectively). As indicated by the orange highlighted regions, the deletion observed in Fam 3 has signal loss not observed in Fam 1 or Fam 2 and not predicted assuming NAHR at breakpoints A-D, demonstrating that Fam 3 has distinct breakpoints from that of Fam 1 and Fam 2 and may not be explained by intrachromosomal NAHR in the current H2 assembly. As indicated by the yellow highlighted regions, observed $\log_2$ ratios for Fam 1 and Fam 2 are most consistent with NAHR occurring at breakpoint D. CN: copy number.

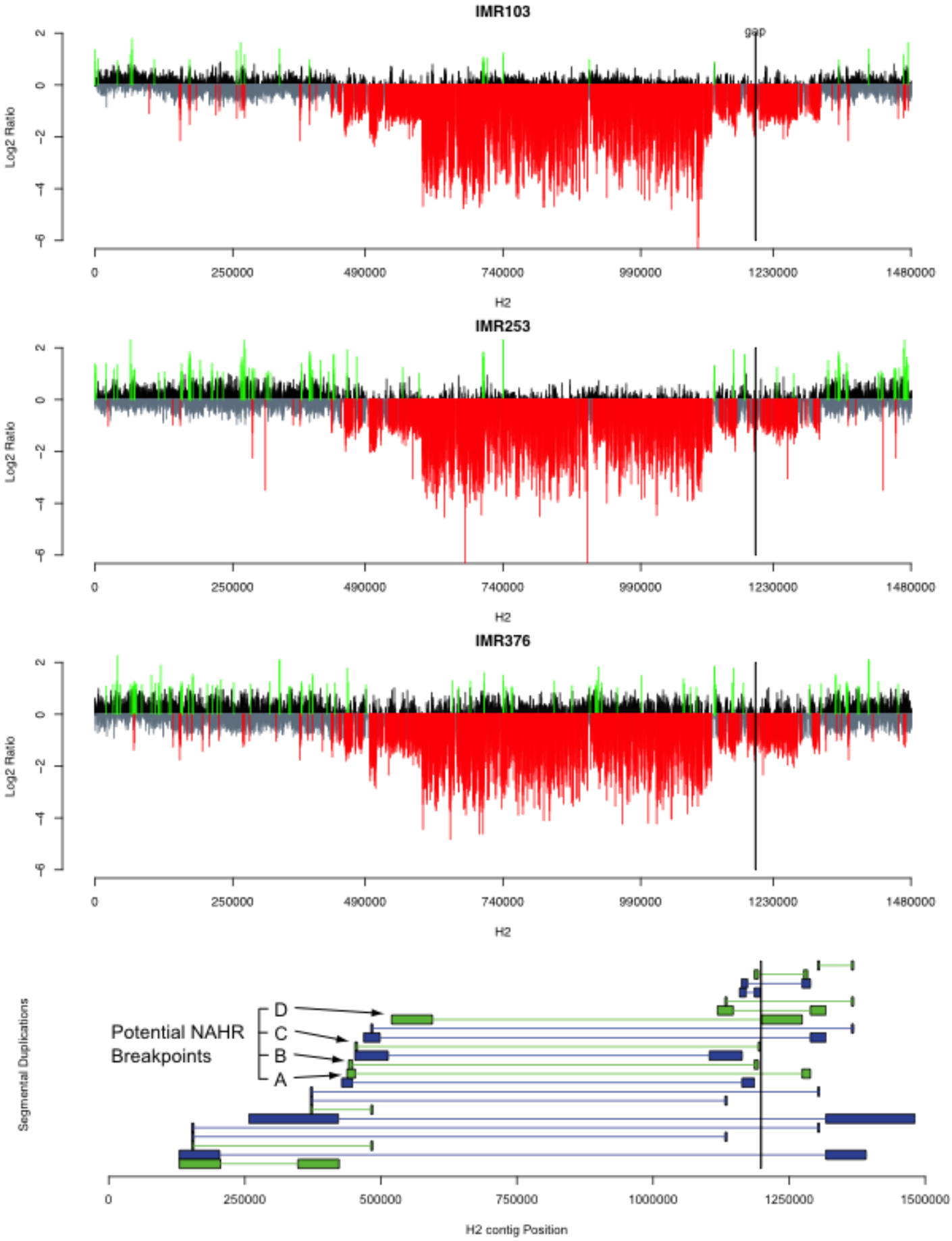

IMR103
gap
Log2 Ratio
2
0
-2
-4
-6
0
250000
490000
740000
990000
1230000
1480000
H2
IMR253
Log2 Ratio
H2
IMR376
Log2 Ratio
H2
Segmental Duplications
Potential NAHR
Breakpoints
D
C
B
A
0
250000
500000
750000
1000000
1250000
1500000
H2 contig Position

Figure S11 Array CGH of hybrid cell lines of unrelated 17q.21.31 deletion probands versus an H2 chromosome.

Array CGH of somatic cell hybrids of 17q21.31 deletion chromosomes versus an H2 chromosome (Fam 2 mother line A, H8) are shown with relative gain (black), loss (gray), and gains and losses >3 SDs beyond the chromosome 17 mean (green and red, respectively) plotted against genomic position on the H2 assembly. Pairs of SDs in direct (green) or inverted (blue) orientation as determined by WGAC are shown as blocks connected by lines. Array CGH from these individuals appears similar to that of Fam 1 and Fam 2, suggesting that deletions as observed in Fam 3 are relatively uncommon (1/6 events).

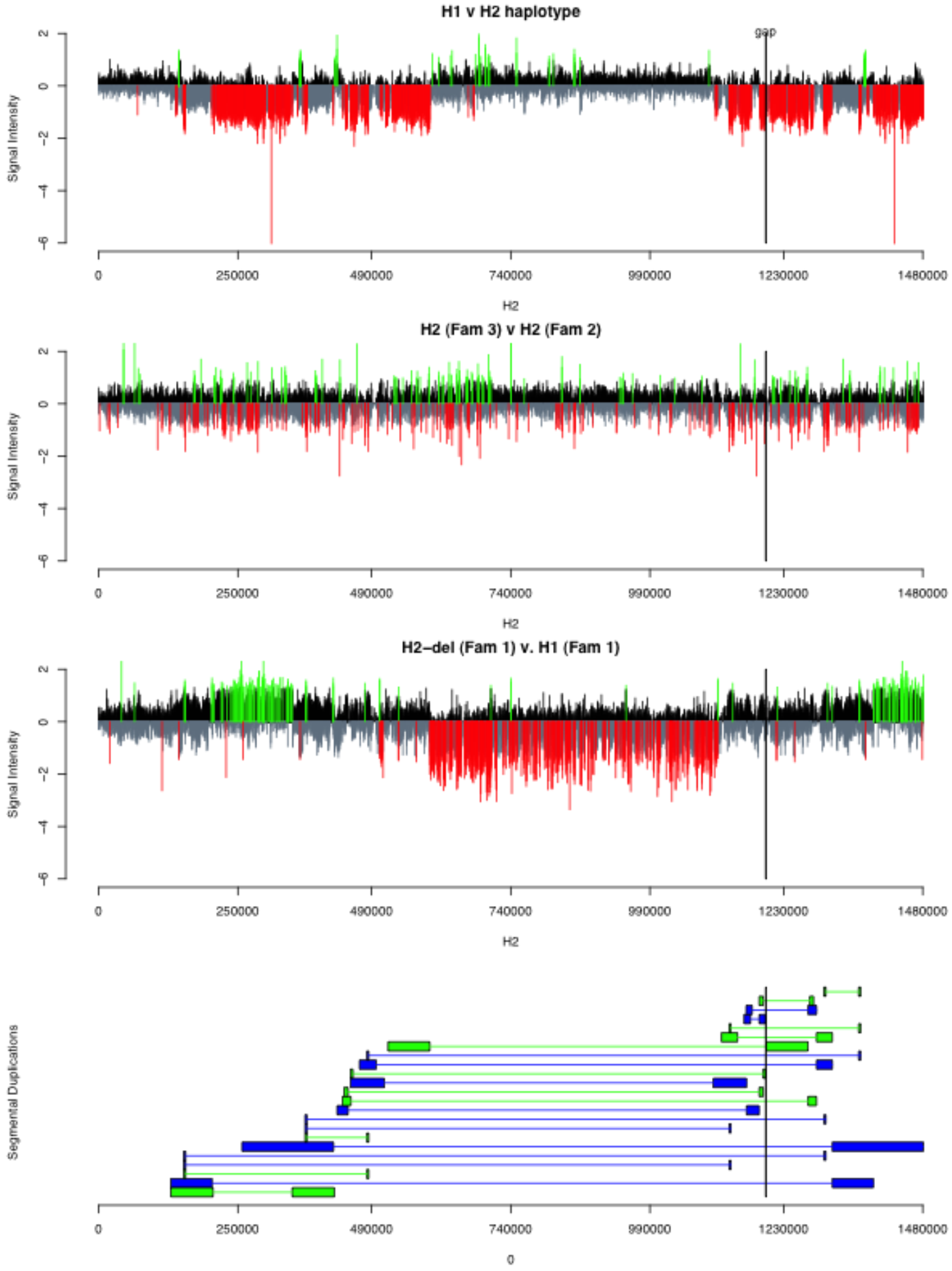

H1 v H2 haplotype
gap
Signal Intensity
2
0
−2
−4
−6
0
250000
490000
740000
990000
1230000
1480000
H2
H2 (Fam 3) v H2 (Fam 2)
Signal Intensity
H2
H2−del (Fam 1) v. H1 (Fam 1)
Signal Intensity
H2
Segmental Duplications
0

Figure S12 Control array CGH hybridizations of hybrid cell lines.
Relative gain (black), loss (gray), and gains and losses >3 SDs beyond the chromosome 17 mean (green and red, respectively) are highlighted. SDs in direct (green) and inverted (blue) orientation are shown as blocks connected by lines. The H1/H2 hybridization (Fam 1 father line A, E8 versus Fam 1 father line B, B4) demonstrates the relative depletion of the H1 for SDs versus the H2 haplotype. Hybridization between the H2 chromosomes Fam 3 mother line B, F2 versus Fam 2 mother line A, E3 demonstrates no evidence of relative copy-number gain or loss, indicating that the H2 chromosome involved in the Fam 3 proband deletion appears to be similar in sequence structure to that of other H2 chromosomes. Finally, hybridization of an H2-deletion chromosome (Fam 1 proband line A, D7) versus an H1 chromosome (Fam 1 father line A, E8) demonstrates how hybridization of differing haplotypes may confound deletion breakpoint detection within SDs.

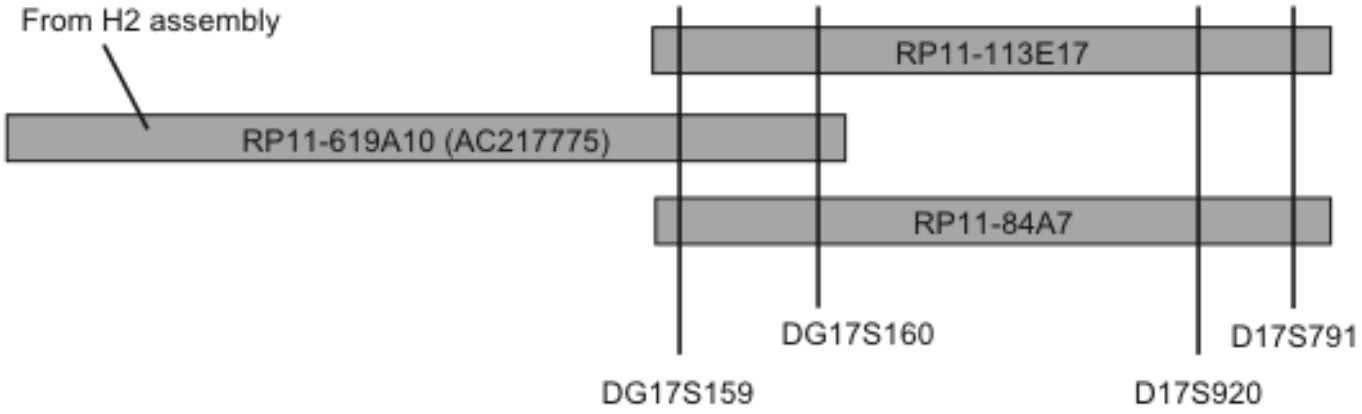


Figure S13 Markers used in confirming placement of RP11-84A7 to close the distal gap in the H2 assembly.
The last clone in the H2 sequence assembly (RP11-619A10) by Zody *et al.*[3] was compared to two BACs originally reported in Stefansson *et al.*[2] to join either the H1 or H2 haplotypes to the reference genome. End-sequencing of RP11-84A7 confirmed placement of the T7 end on RP11-619A10 and the SP6 end on AC019319 but was nonspecific for the H1 or H2 haplotypes. Microsatellite genotypes at DG17S159 and DG17S160 confirmed overlap with RP11-619A10 and genotypes at D17S920 and D17S791 confirmed a distinct haplotype from negative control RP11-113E17 (Table S5).

REFERENCES

1. Parsons, J.D. (1995). Miropeats: graphical DNA sequence comparisons. Comput Appl Biosci 11, 615-619.
2. Stefansson, H., Helgason, A., Thorleifsson, G., Steinthorsdottir, V., Masson, G., Barnard, J., Baker, A., Jonasdottir, A., Ingason, A., Gudnadottir, V.G., et al. (2005). A common inversion under selection in Europeans. Nat Genet 37, 129-137.
3. Zody, M.C., Jiang, Z., Fung, H.C., Antonacci, F., Hillier, L.W., Cardone, M.F., Graves, T.A., Kidd, J.M., Cheng, Z., Abouelleil, A., et al. (2008). Evolutionary toggling of the MAPT 17q21.31 inversion region. Nat Genet.

Marker Name	Forward (5'->3')	Reverse (5'->3')	Purpose*	Note
MAPT Intron 9 del	GGAAGACGTTCTCACTGATCTG	AAGAGTCTGGCTTCAGTCTCTC	1	
AFMa061za9	AGCTGCTTCTGCAAAGATG	TACAAGTCCTGGGCCAC	2,3	D17S1876
AFM192yh2	GCCGTCAGTTATTTCACTTG	ATGTTATGGAGACAGTGCTAGG	2,3	D17S799
AFMa154za9	AGTGCCAGAGATATACCGTG	GTCTGCAAGGCAAGTTGTC	2,3	D17S1795
AFM044xg3	GAGTCTCCTAAATGCTGGGG	AGCTCCTGCACAGTTCTTAAATA	2,3	D17S784
AFM298wg5	[6-FAM]GGCCTCCCAAACTGCTT	TCTACCCCGATGAGCCA	3	
AFMb364yh9	[5HEX]GCGCAATCTCAGGTCA	ACCACCGTGTCTGGCTA	3	
AFM155xd12	[6-FAM]GTTTTCTCCAGTTATTCCCC	GCTCGTCCTTTGGAAGAGTT	3	
AFMa110wb5	[6-FAM]GACACTTCAGAACCACCTCC	GTCCTCANACTGCCTTTC	3	
D17S810	[5HEX]CCCTGTACCCGACTCAGAA	GCCTGGTTGCATAACCC	3	
DG17S133	[6-FAM]GCCTGGGAAACAAGAGTGAG	CCAGAGGTCCTAGTGCCAAA	3	
DG17S146	[5HEX]TGCAAGCAGTTTAAGCAGGA	CCACAACTTGGCTGCTCAC	3	
DG17S153	[6-FAM]AGATCACGCCACTGCACTC	ACCAACACAGGATGGTAGGC	3	
DG17S159	[5HEX]TCATTCTGGCTCCCTGTCTT	TCAGGTTCCTAAATGAGGAGTAGG	3	
DG17S160	[5HEX]GCAATCATACTCCGTTGCAG	ATGACGGGACTTGTGTGGAT	3	
DG17S435	[5HEX]GCAGGGTCACCACCTTATTC	GGAGGTGAGCCATGATTGAG	3	
D17S920	[6-FAM]CTTCTGCATGGATCAGAAAA	ACTCACCCAGATGATTGCTT	4	

*1 - determine H1/H2 genotype. 2 - check integrity of chr17 in hybrid cell lines. 3 - determine chromosome(s) of origin in 17q21.31 deletion. 4 - confirm clone placement in H2 assembly.

Table S1 Somatic cell hybrid genotyping markers.

Probe Category	Purpose	Similarity Filter	number of probes	genome	notes
17q21.31 human	region of interest	No Filter	24797	hg18	chr17:40250000-42750000
interchromosomal SDs with paralogous sequence in 17q21	additional probes within region of interest	No Filter	7940	hg18	intervals from various chromosomes
human chr17, excluding 17q21	array normalization	Similarity Score Filter	60000	hg18	chr17:1-40000000; chr17:43000000-81195210
human genome probe backbone	if needed, assay other human chromosomes present	Similarity Score Filter	30000	hg18	all human chromosomes
11qE1 mouse	if needed, assess human-mouse cross-hybridization	Similarity Score Filter	14339	mm9	chr11:102669307-104539951
mouse chr11 (contains all human chr17 synteny)	if needed, use to assess cross-hybridization	Similarity Score Filter	40000	mm9	chr11:1-121798632
mouse genome	if needed, normalization to this set allows for relative level of human chr17 to be assayed	Similarity Score Filter	29998	mm9	all mouse chromosomes
Agilent mouse normalization probes	if needed, normalization to this set allows for relative level of human chr17 to be assayed	NA	11492	mm9	
Agilent mouse replicate probes	assess replicability of signal on array	NA	5000	mm9	
Agilent Control Probes	present on all Agilent arrays	NA	5099	various	

Table S2 Design specifications of the custom Agilent array to interrogate 17q21.31 in somatic cell hybrids.

Sample	chr17 homologue	cell line	H1/H2 genotype
Fam 1 Father	A	E8	H1
Fam 1 Father	A	C2	H1
Fam 1 Father	B	B4	H2
Fam 1 Father	B	D4	H2
Fam 1 Proband	A	B5	H1
Fam 1 Proband	A	D4	H1
Fam 1 Proband	B	D7	-
Fam 1 Proband	B	G9	-
Fam 2 Mother	A	E3	H2
Fam 2 Mother	A	H8	H2
Fam 2 Mother	B	E5	H2
Fam 2 Mother	B	G2	H2
Fam 2 Proband	A	E5	H1
Fam 2 Proband	A	H10	H1
Fam 2 Proband	B	C1	-
Fam 2 Proband	B	C8	-
Fam 3 Mother	A	B8	H1
Fam 3 Mother	A	C11	H1
Fam 3 Mother	B	F2	H2
Fam 3 Mother	B	G4	(H2?)
Fam 3 Proband	A	B7	H1
Fam 3 Proband	A	C7	H1
Fam 3 Proband	B	A7	-
Fam 3 Proband	B	G3	-
IMR 103	A	F5	H1
IMR 103	A	G7	(H1?)
IMR 103	B	F6	-
IMR 103	B	F10	-
IMR 253	A	E10	-
IMR 253	A	F10	-
IMR 253	B	D6	H1
IMR 253	B	F11	H1
IMR 376	A	B6	-
IMR 376	A	B7	-
IMR 376	B	C4	H1
IMR 376	B	H3	H1

Table S3 Hybrid cell lines generated.
Originally, H1/H2 genotypes were inferred using a 238 bp deletion in intron 9 of *MAPT*. In 17q21.31 deletions, this gene is lost entirely and no amplification is observed. For H1/H2 genotypes listed in parentheses, the original genotyping failed but was later inferred from microsatellite data.

QNAME	QB	QE	QLEN	SNAME	SB	SE	per_SE	SE_sim	K2p	SE_kimura
H1	129729	150021	1405709	H1	981035	963600	0.9817	NA	0.0186	0.0010
H1	129729	150021	1405 09	H1	1198970	1181462	0.9823	NA	0.0179	0.0010
H1	146032	150021	1405709	H1	50025	154014	0.9990	0.0005	0.0010	0.0005
H1	150025	207448	1405709	H1	1185446	1123304	0.9835	NA	0.0167	0.0005
H1	150025	187318	1405709	H1	967593	925941	0.9804	NA	0.0199	0.0007
H1	158243	159594	1405709	H1	232620	233843	0.9010	0.0087	0.1071	0.0102
H1	219599	226634	1405709	H1	894677	887698	0.9835	0.0015	0.0167	0.0016
H1	227453	261693	1405709	H1	886258	850934	0.9885	0.0006	0.0116	0.0006
H1	230482	233843	1405709	H1	958106	954620	0.9096	0.0050	0.0970	0.0058
H1	230482	233843	1405709	H1	1175961	1172475	0.9099	0.0050	0.0966	0.0058
H1	879862	883229	1405709	H1	954620	958106	0.9111	0.0050	0.0953	0.0057
H1	879862	883229	1405709	H1	1172475	1175961	0.9114	0.0050	0.0949	0.0057
H1	925941	1123304	1405709	H1	1143518	1341017	0.9981	NA	0.0019	0.0001
H1	1	150021	1405709	H2	1	147152	0.9963	NA	0.0037	0.0002
H1	129729	150021	1405709	H2	347549	365072	0.9819	NA	0.0184	0.0010
H1	129729	150021	1405709	H2	1391378	1373864	0.9820	NA	0.0182	0.0010
H1	150025	207448	1405709	H2	143159	205298	0.9834	NA	0.0168	0.0005
H1	150025	219717	1405709	H2	361079	435489	0.9841	NA	0.0160	0.0005
H1	150025	205626	1405709	H2	1377857	1317537	0.9826	NA	0.0176	0.0006
H1	212314	219747	1405709	H2	1186182	1175766	0.9865	0.0013	0.0136	0.0014
H1	219599	261693	1405709	H2	469149	513195	0.9873	0.0005	0.0128	0.0006
H1	219599	926076	1405709	H2	1146505	437775	0.9948	NA	0.0052	0.0001
H1	219599	247714	1405709	H2	1316783	1289117	0.9875	0.0007	0.0126	0.0007
H1	230482	233843	1405709	H2	370571	374065	0.9108	0.0050	0.0956	0.0057
H1	230482	233843	1405709	H2	152652	156148	0.9086	0.0050	0.0980	0.0058
H1	230482	233843	1405709	H2	1368369	1364872	0.9099	0.0050	0.0966	0.0058
H1	769304	844580	1405709	H2	1273881	1198880	0.9956	NA	0.0045	0.0002
H1	850934	886258	1405709	H2	1103477	1138618	0.9903	0.0005	0.0098	0.0005
H1	865792	886258	1405709	H2	1289117	1308890	0.9929	0.0006	0.0071	0.0006
H1	879862	883229	1405709	H2	374065	370571	0.9129	0.0049	0.0932	0.0056
H1	879862	883229	1405709	H2	156148	152652	0.9111	0.0050	0.0953	0.0057
H1	879862	883229	1405709	H2	1364872	1368369	0.9123	0.0049	0.0939	0.0057
H1	887698	911684	1405709	H2	1139437	1163543	0.9866	0.0007	0.0135	0.0008
H1	887698	895435	1405709	H2	1309709	1317540	0.9831	0.0015	0.0171	0.0015
H1	907113	911684	1405709	H2	1197879	1193288	0.9901	0.0015	0.0099	0.0015
H1	910905	926076	1405709	H2	1289120	1273938	0.9849	0.0010	0.0153	0.0010
H1	916037	926076	1405709	H2	1163440	1173502	0.9902	0.0010	0.0098	0.0010
H1	916037	923242	1405709	H2	1193391	1186179	0.9896	0.0012	0.0105	0.0012
H1	925941	1198970	1405709	H2	403035	129631	0.9978	NA	0.0022	0.0001
H1	925941	1071136	1405709	H2	1335915	1481052	0.9974	0.0001	0.0026	0.0001
H1	954620	958106	1405709	H2	484902	481538	0.9093	0.0050	0.0973	0.0058
H1	954620	958106	1405709	H2	1132239	1135600	0.9096	0.0050	0.0970	0.0058
H1	954620	958106	1405709	H2	1302509	1305872	0.9099	0.0050	0.0966	0.0058
H1	1123304	1146564	1405709	H2	423242	400003	0.9979	0.0003	0.0021	0.0003

H1	1124829	1288708	1405709	H2	1317451	1481052	0.9976	NA	0.0024	0.0001
H1	1172475	1175961	1405709	H2	484902	481538	0.9096	0.0050	0.0970	0.0058
H1	1172475	1175961	1405709	H2	1132239	1135600	0.9099	0.0050	0.0967	0.0058
H1	1172475	1175961	1405709	H2	1302509	1305872	0.9102	0.0050	0.0963	0.0058
H1	1200001	1341017	1405709	H2	346518	205298	0.9983	0.0001	0.0018	0.0001
H2	129631	205298	1481053	H2	347549	423242	0.9989	NA	0.0011	0.0001
H2	129631	203770	1481053	H2	1391378	1317451	0.9975	0.0002	0.0025	0.0002
H2	152652	156148	1481053	H2	481538	484902	0.9090	0.0050	0.0977	0.0058
H2	152652	156148	1481053	H2	1135600	1132239	0.9093	0.0050	0.0974	0.0058
H2	152652	156148	1481053	H2	1305872	1302509	0.9096	0.0050	0.0970	0.0058
H2	257472	421714	1481053	H2	1481053	1317451	0.9985	NA	0.0015	0.0001
H2	370571	374065	1481053	H2	481538	484902	0.9111	0.0050	0.0953	0.0057
H2	370571	374065	1481053	H2	1135600	1132239	0.9114	0.0050	0.0949	0.0057
H2	370571	374065	1481053	H2	1305872	1302509	0.9117	0.0050	0.0946	0.0057
H2	428107	447822	1481053	H2	1186182	1163440	0.9908	0.0007	0.0092	0.0007
H2	437714	452954	1481053	H2	1273877	1289120	0.9860	0.0010	0.0142	0.0010
H2	440620	447822	1481053	H2	1186179	1193391	0.9922	0.0010	0.0078	0.0010
H2	452165	513195	1481053	H2	1163543	1103477	0.9902	0.0004	0.0099	0.0004
H2	452165	456744	1481053	H2	1193288	1197879	0.9926	0.0013	0.0075	0.0013
H2	468394	498328	1481053	H2	1317540	1289117	0.9916	0.0005	0.0085	0.0006
H2	481538	484902	1481053	H2	1368369	1364872	0.9103	0.0050	0.0962	0.0058
H2	519560	594627	1481053	H2	1198880	1273881	0.9975	NA	0.0025	0.0002
H2	1118360	1147260	1481053	H2	1289117	1317540	0.9937	0.0005	0.0063	0.0005
H2	1132239	1135600	1481053	H2	1364872	1368369	0.9105	0.0050	0.0960	0.0058
H2	1158952	1170649	1481053	H2	1197879	1186179	0.9991	0.0003	0.0009	0.0003
H2	1162742	1173563	1481053	H2	1289120	1273877	0.9871	0.0011	0.0130	0.0011
H2	1186179	1193391	1481053	H2	1276778	1283979	0.9850	0.0014	0.0152	0.0015
H2	1302509	1305872	1481053	H2	1364872	1368369	0.9108	0.0050	0.0956	0.0057

QNAME - query name; QB - query begin; QE - query end; QLEN - query sequence length; SNAME - sequence name; SB - sequence start; SE - sequence end; per_SE - percent sequence identity; K2p: Kimura 2 parameter distance; SE_sim - standard error of similarity (NA for merged alignments); SE_kimura - standard error of kimura.

Table S4 Pairwise sequence alignments (≥90% identity, ≥1 kbp) between the H1 and H2 haplotype assemblies (Zody et al.).

Clone	DG17S159	DG17S160	DG17S920	D17S791
RP11-113E17	119	384	98	173
RP11-619A10	131	380	NA	NA
RP11-84A7	131	380	94	185

Table S5 Microsatellite genotyping to confirm placement of RP11-84A7 on the H2 haplotype.

Table S6 Locus-specific 36-mers for identifying 17q21.31 recurrent deletion breakpoints.
Contig: contig of the PSV (H1, H2, or Gap 1 of H2). contigPos: position of PSV in the corresponding contig (NA for gap sequence). Alignment: paralogous sequence pair to which the k-mer belongs (H2 - breakpoint D non-gap sequence; H2infer - Gap 1 vs. proximal H2 sequence; H1infer - Gap 1 vs. proximal H1 sequence). Member: whether the k-mer belongs to the proximal or distal member of the paralogous sequence pair. alignPos: relative position of the PSV in the paralogous sequence alignment. Offset: Position at which the PSV of interest is located in the k-mer. Inform: NA for k-mers outside gap 1. For k-mers in Gap 1, classification of whether the k-mer is a SUNK, H1-informative or H2-informative. k-mer: 36-mer of interest. A given 36-mer may correspond to multiple PSVs and therefore may be listed more than once.

<LARGE TABLE - SEE EXCEL FILE>