nature | **methods**

BreakDancer: an algorithm for high-resolution mapping of genomic structural variation

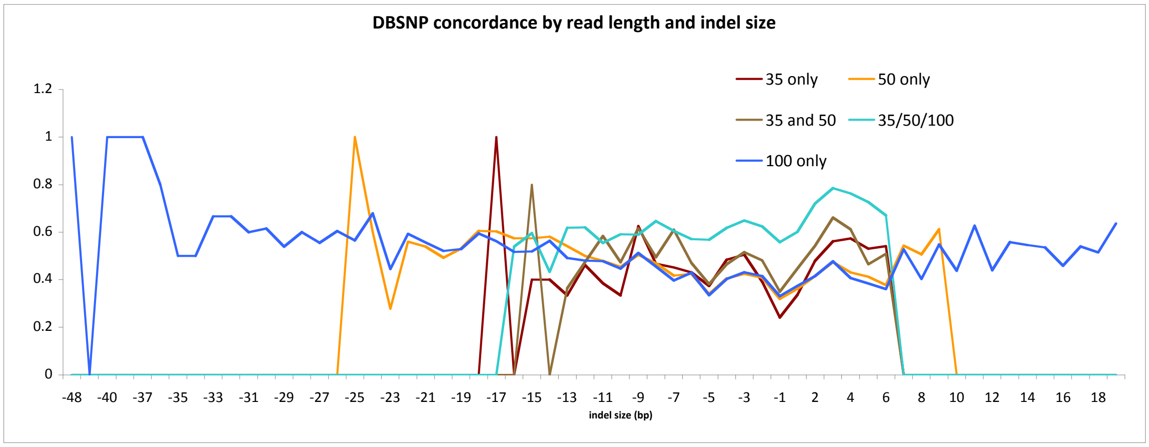
Ken Chen, John W Wallis, Michael D McLellan, David E Larson, Joelle M Kalicki, Craig S Pohl, Sean D McGrath, Michael C Wendl, Qunyuan Zhang, Devin P Locke, Xiaoqi Shi, Robert S Fulton, Timothy J Ley, Richard K Wilson, Li Ding & Elaine R Mardis

Supplementary figures and text:

Note: Supplementary Tables 1 and 3 as well as Supplementary Software are available on the Nature Methods website.

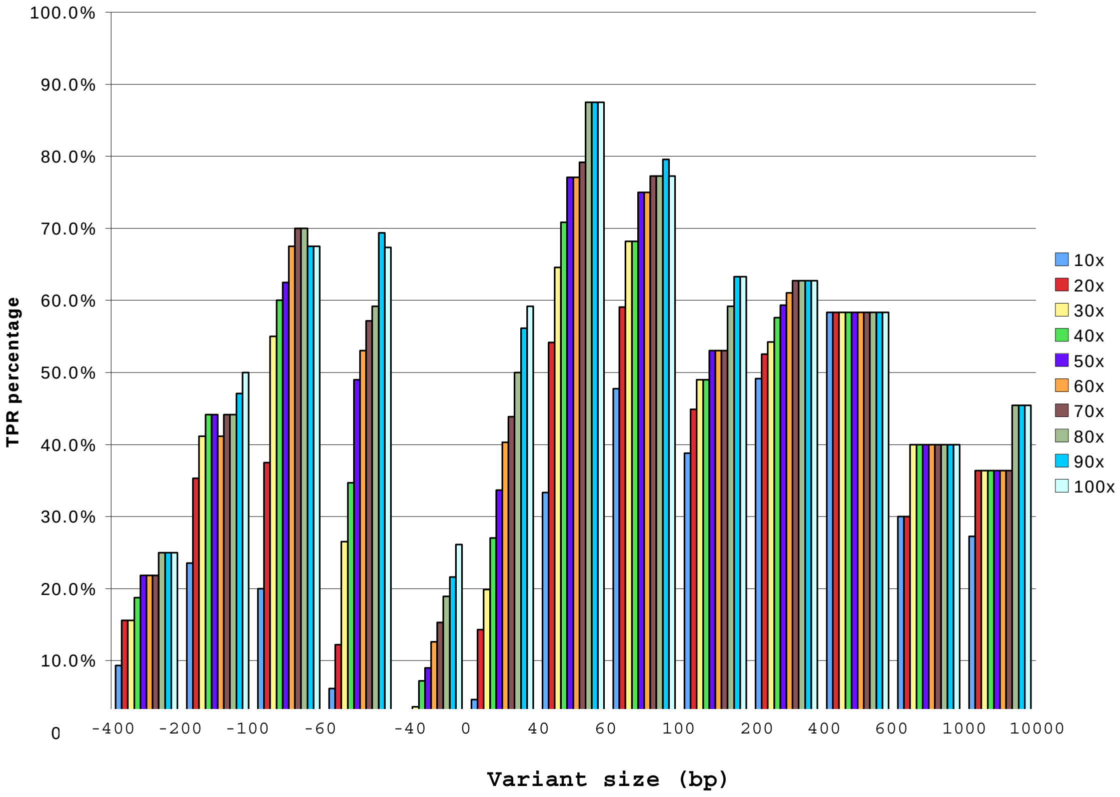
Supplementary Figure 1 The size of indels detectable by MAQ paired end Smith-Waterman alignment algorithm[1].

Minus (-) represents deletions and plus (+) insertions. These results were obtained from one of our genome sequencing projects (data not shown).



Supplementary Figure 2 The true positive rate (TPR) w.r.t. coverages in different size bins.

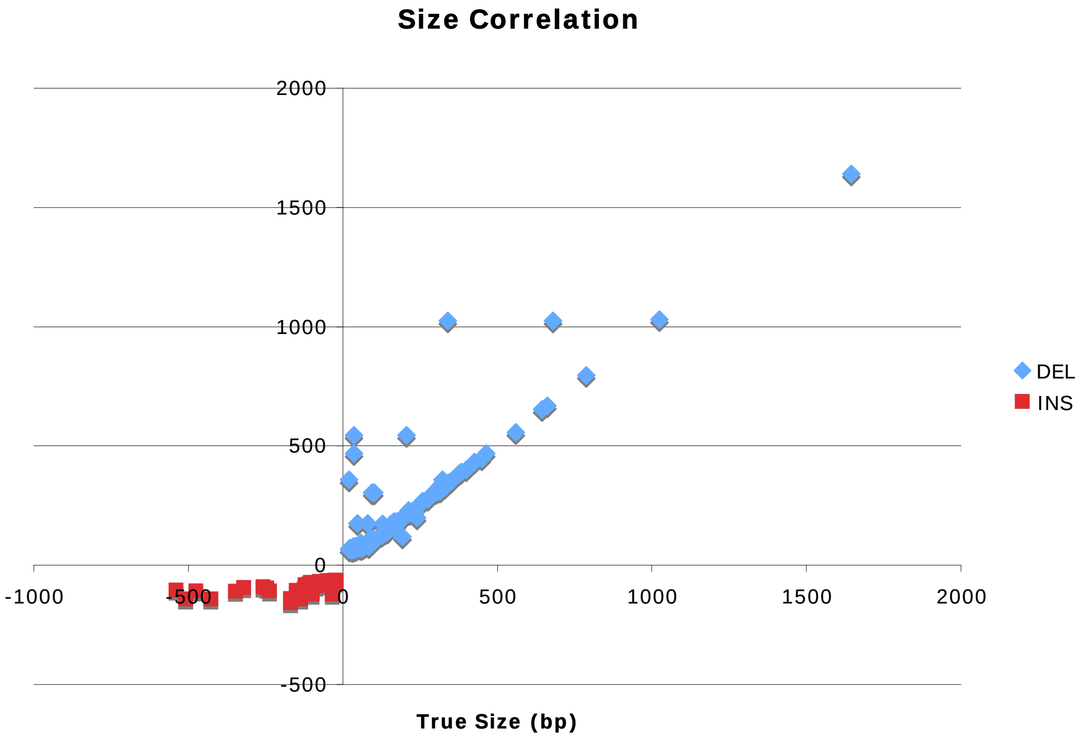
Results were obtained from simulation using 50 bp paired end reads with a 200 bp mean and 20 bp s.d insert size randomly produced from J. Craig Venter's chromosome 17 nucleotide sequence by MAQ-0.7.1 (- represents insertions and + deletions).



Nature Methods: doi: 10.1038/nmeth.1363

Supplementary Figure 3 Size and Breakpoint Accuracy of BreakDancerMax in simulation

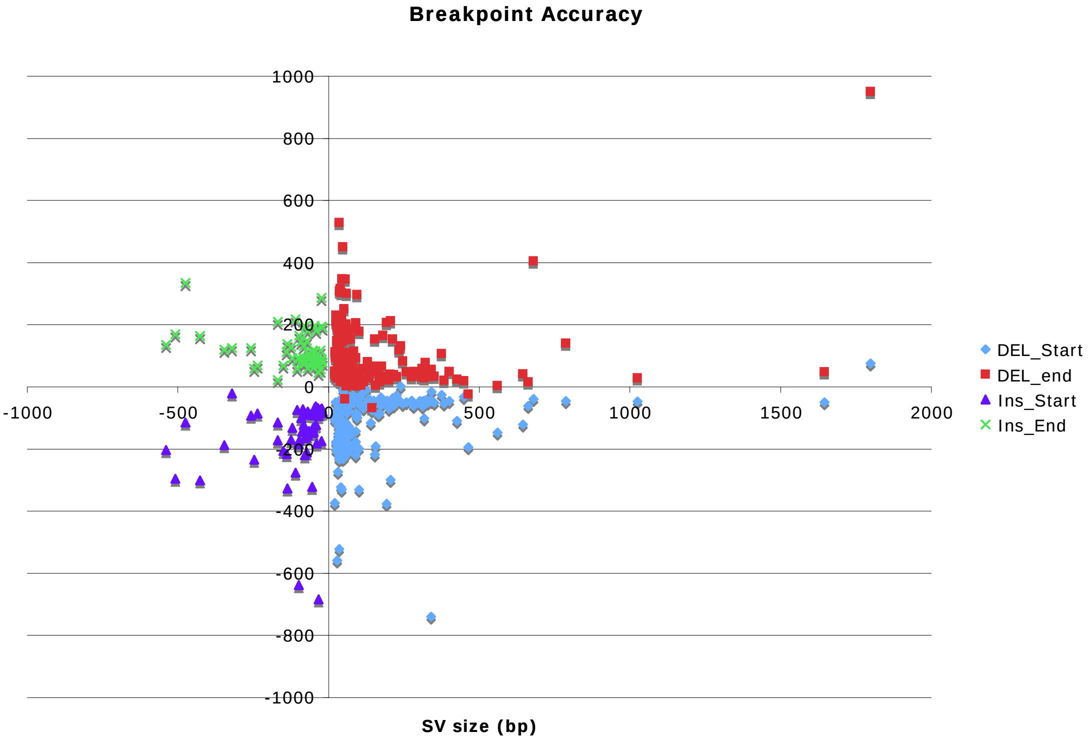
(**a**) Correlation between the predicted and the true variant size.



Supplementary Figure 3

(**b**) Accuracy of the predicted variant boundaries.

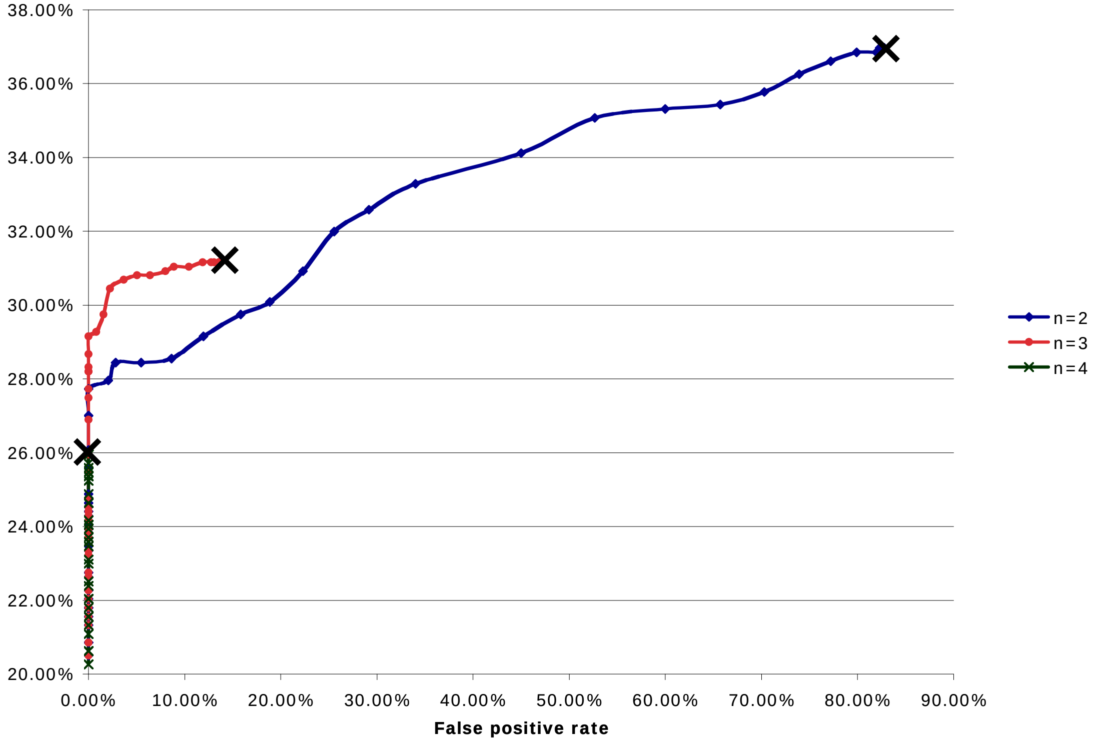
The x axis represents the size of SV (+ for the deletions, - for the insertions). The y axis is the distance (in bp) between the predicted SV boundaries and the corresponding true breakpoints (- for the start, + for the end).



Nature Methods: doi: 10.1038/nmeth.1363

Supplementary Figure 4 Receiver operator characteristics (ROC) w.r.t. the number of anomalous read pairs (n) and the confidence score (Q).

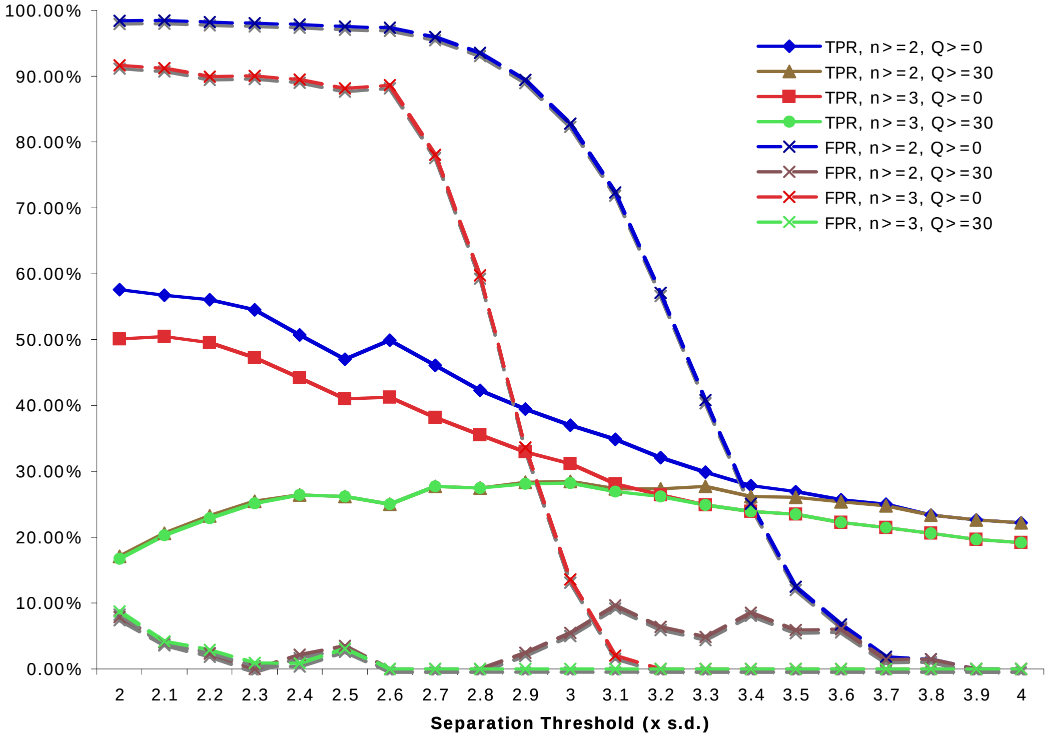
The x axis and the y axis are the false positive rate (TPR) and the true positive rate (FPR) in simulated structural variation detection. The crosses represent the discrete tradeoffs achieved at n ≥ 2, 3, and 4 and Q ≥ 0. The colored lines represent the contiguous tradeoff achieved from applying additional confidence score threshold from Q ≥ 0 to Q ≥ 90 with n ≥ 2.



Nature Methods: doi: 10.1038/nmeth.1363

Supplementary Figure 5 The true positive rate (TPR) and the false positive rate (FPR) given various separation thresholds.

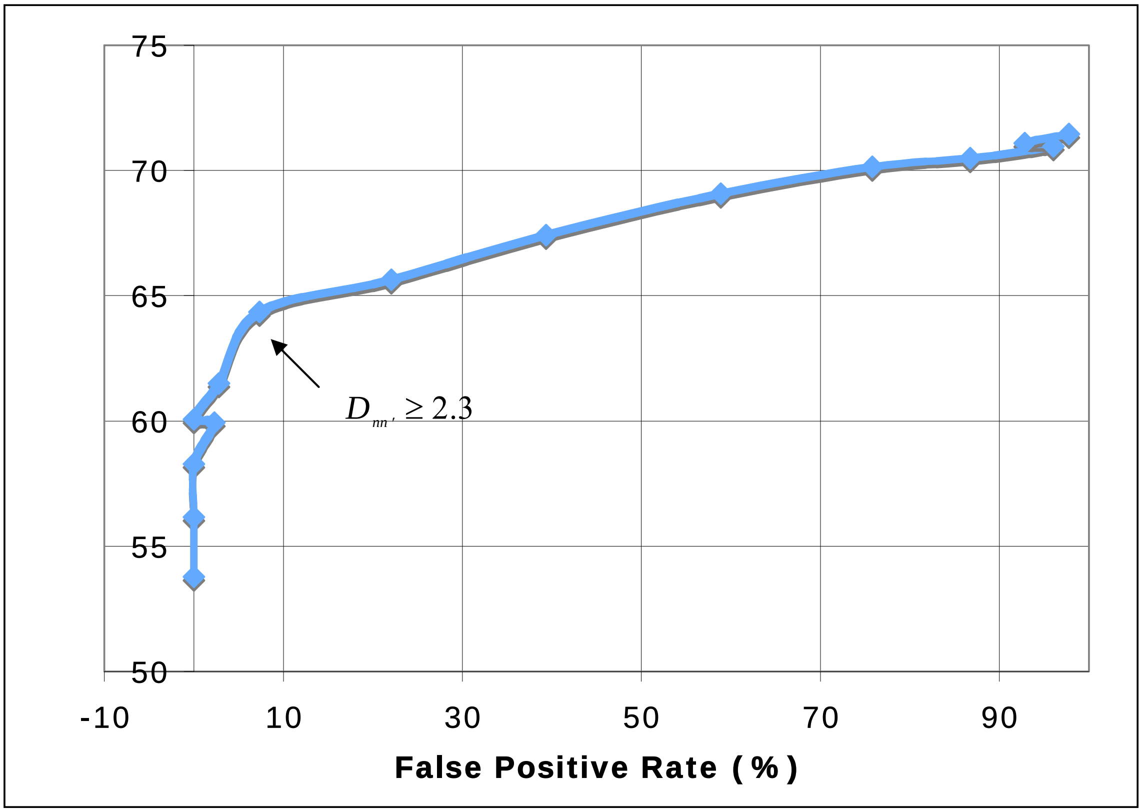
The x axis represents the separation distance threshold in unit of the standard deviation insert size. The y axis is the TPR and the FPR obtained under various conditions defined by the number of anomalous read pairs (n) and the confidence scores (Q). For thresholds ranging from two to four s.d. in insert size, culling based on confidence scores (Q ≥ 30) resulted in relatively steady TPRs and FPRs, while ARP-based culling (Q ≥ 0) led to substantially variable results.



Nature Methods: doi: 10.1038/nmeth.1363

Supplementary Figure 6 The Receiver Characteristic (ROC) of BreakDancerMini at different threshold of $D_{nn'}$.

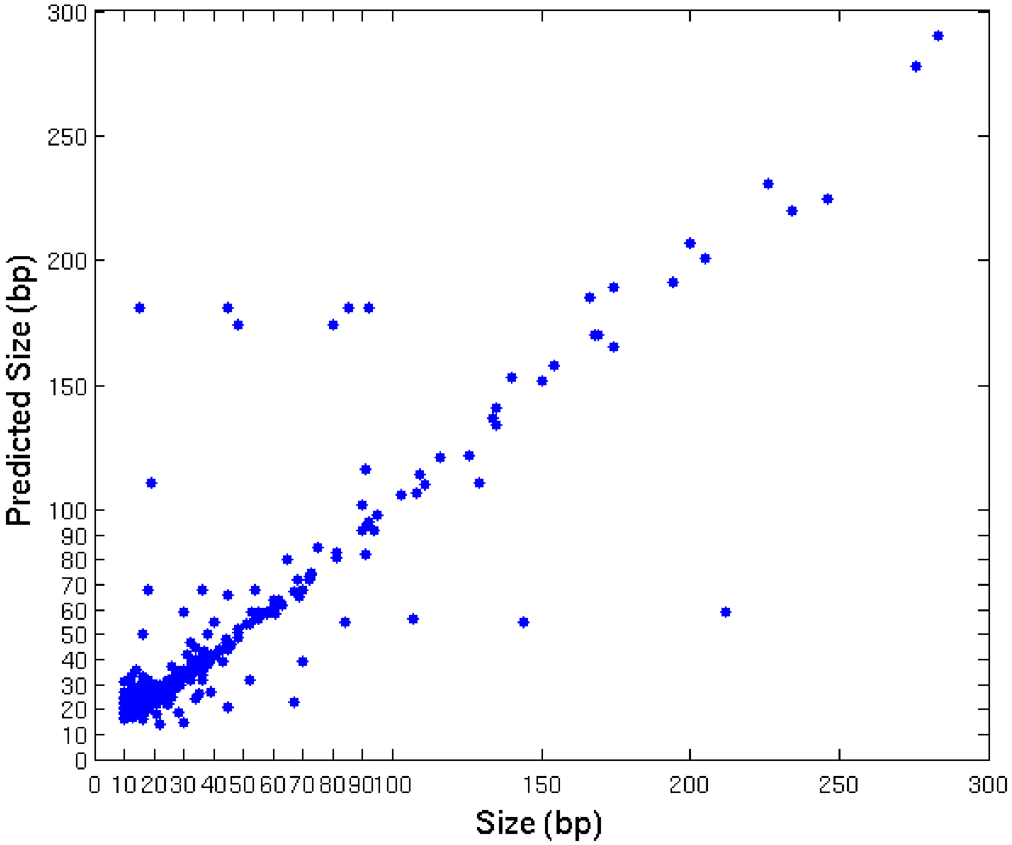
This curve was obtained via simulated structural variant detection using our synthetic Chromosome 17 data (MAQ mapping quality > 10). The false positive rate reduced to below 10% at $D_{nn'} \geq 2.3$ before true positive rate started to drop quickly.



Nature Methods: doi: 10.1038/nmeth.1363

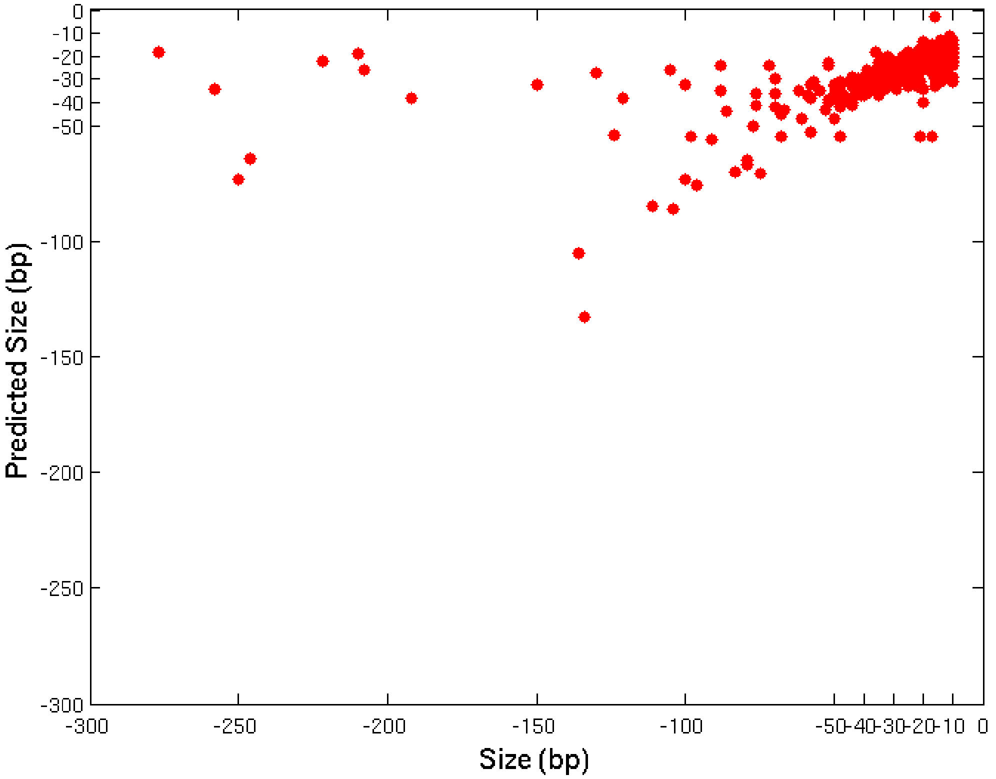
Supplementary Figure 7 Accuracy of variant size estimated by BreakDancerMini

(**a**) Correlation between the BreakDancerMini predicted size and the actual size for deletions between 10 to 300 bp on Venter's Chromosome 17 (Pearson's r=0.7812).



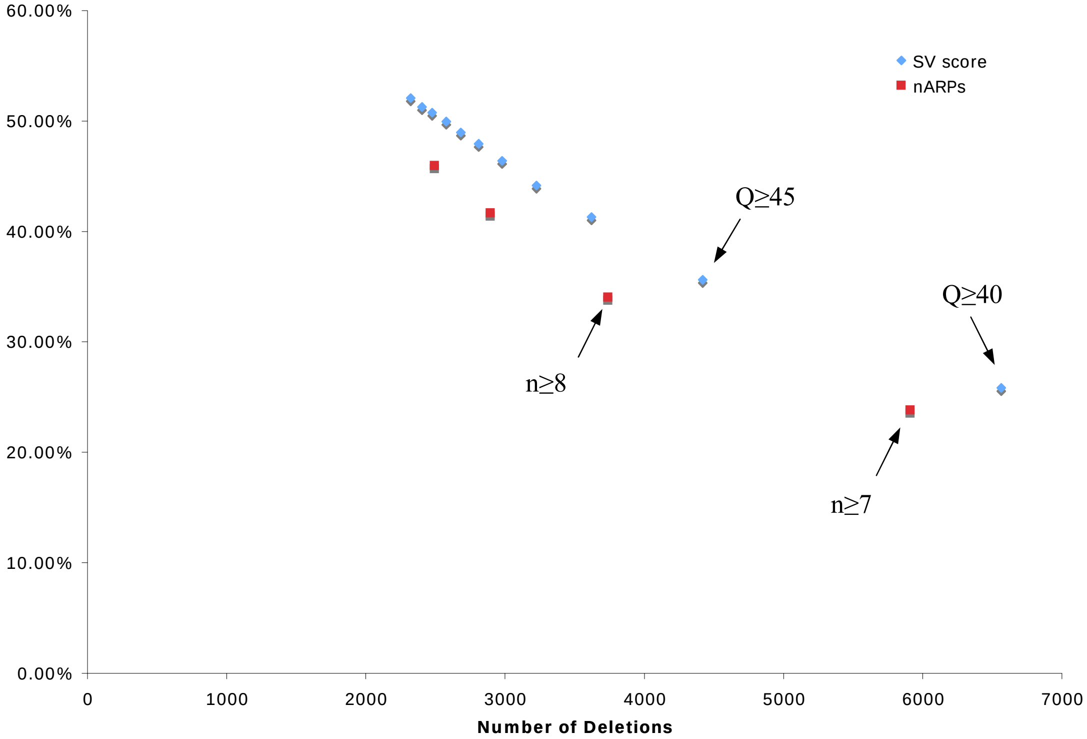
Supplementary Figure 7

(**b**) Correlation between the BreakDancerMini predicted size and the actual size for insertions between 10bp to 300bp on Venter's Chr 17 (Pearson's r=0.5897).



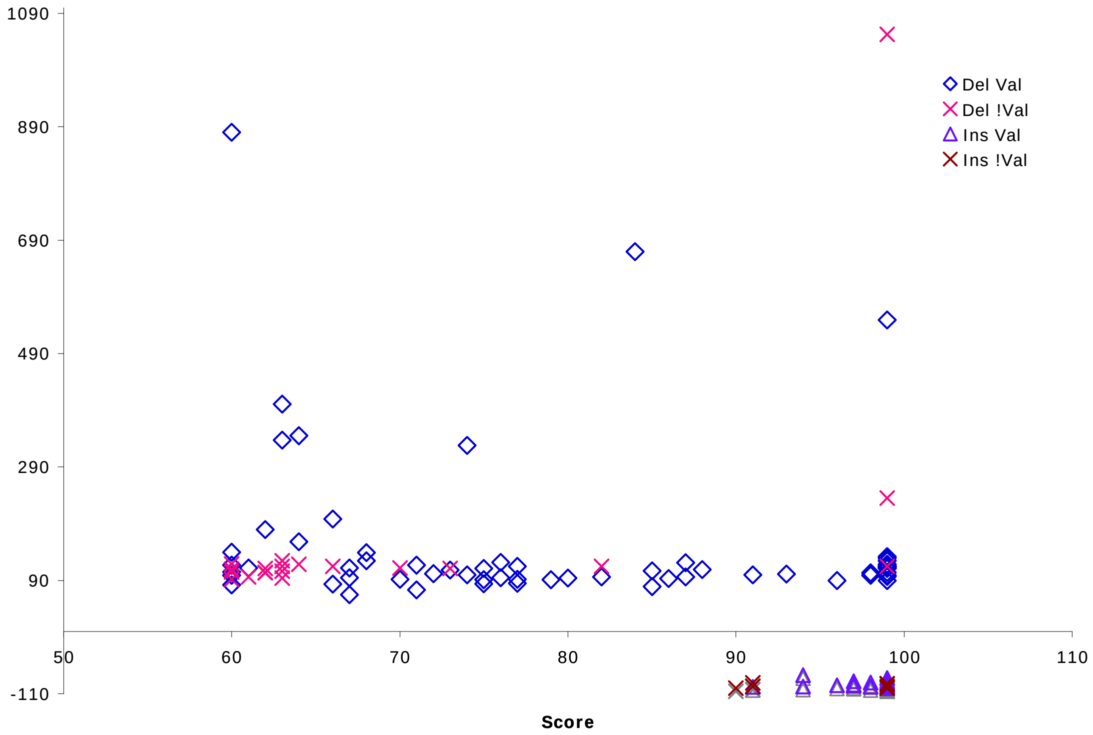
Supplementary Figure 8 Percent overlap between the AML structural variants and the DGV (v5) as functions of the confidence score or number of anomalous read pairs.

As we increased the confidence threshold, the number of putative variants dropped quickly and the percentage of deletions that overlap by 50% or more with any entry in the database of genomic variants (DGV, v. 5.0: http://projects.tcag.ca/variation/) increased rapidly.



Supplementary Figure 9 AML validation result w.r.t variant size and confidence score

Del Val are deletions validated, Del !Val deletions not validated, Ins Val insertions validated, and Ins !Val insertions not validated.

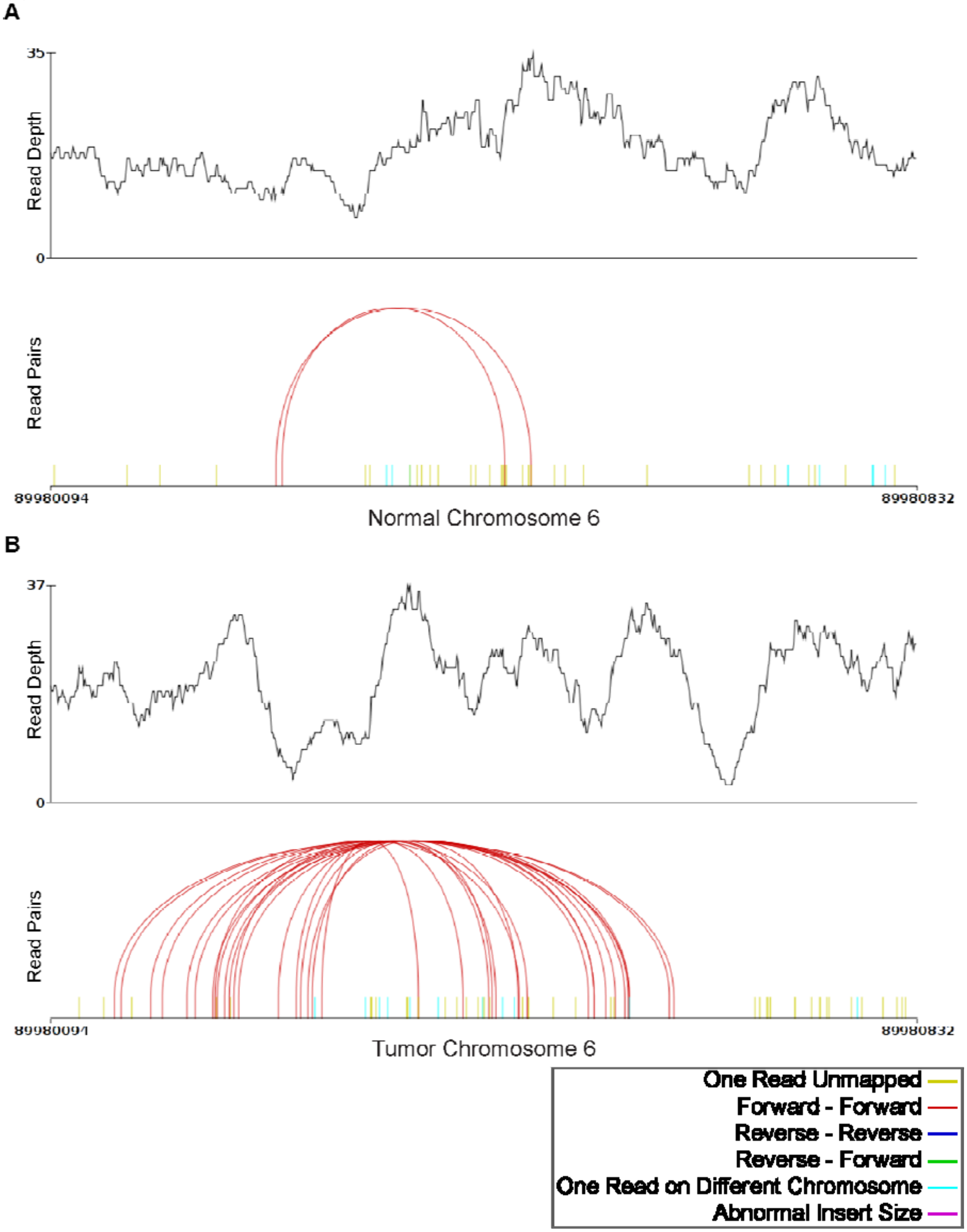


Nature Methods: doi: 10.1038/nmeth.1363

Supplementary Figure 10 Plots of identified inversions and intra-chromosomal translocations in the AML genome

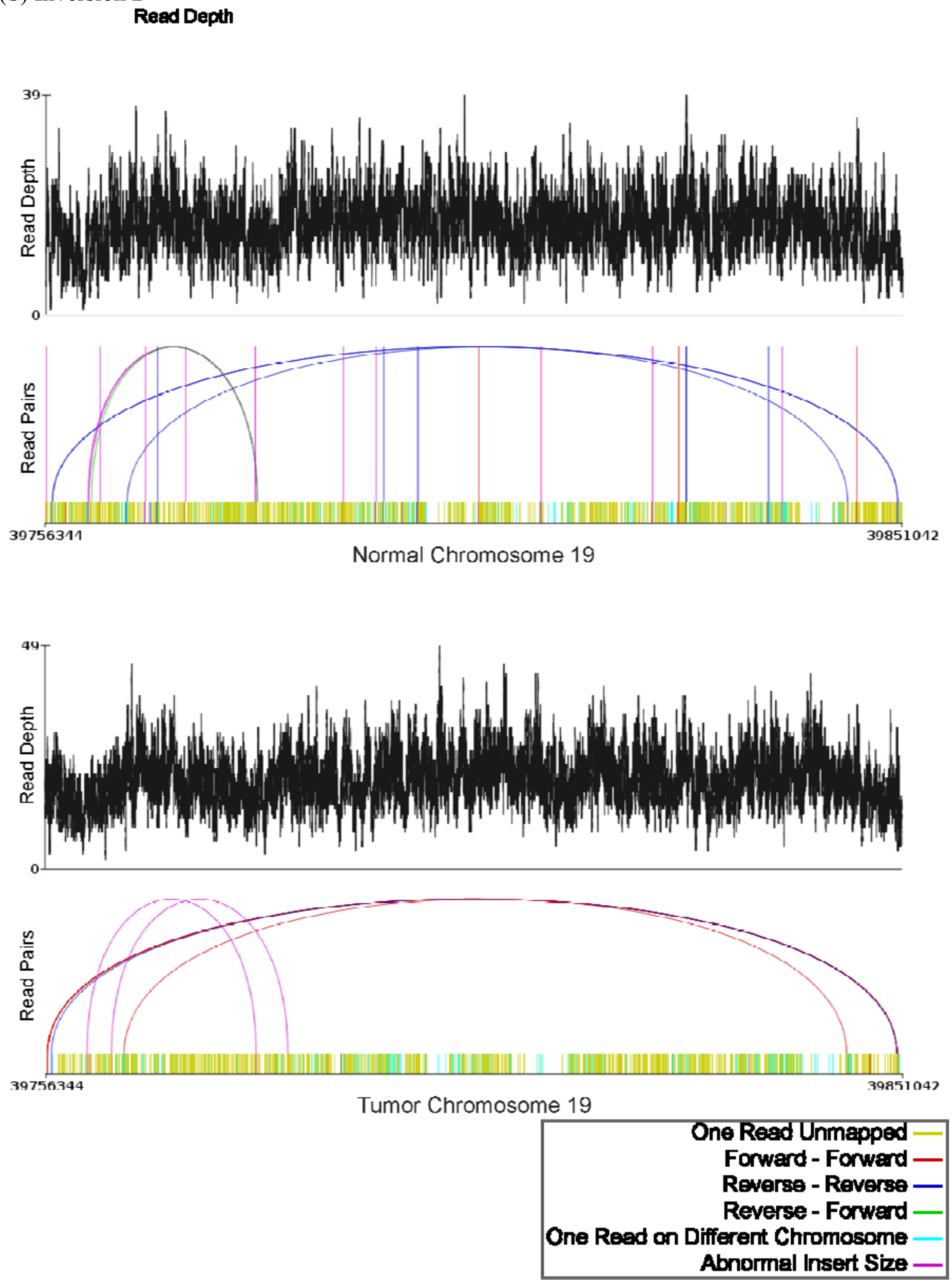
Plots were produced by Yenta, a prototype structural variant visualization tool developed in house at Washington University Genome Center by David E. Larson et al.

(**a**) Inversion 1



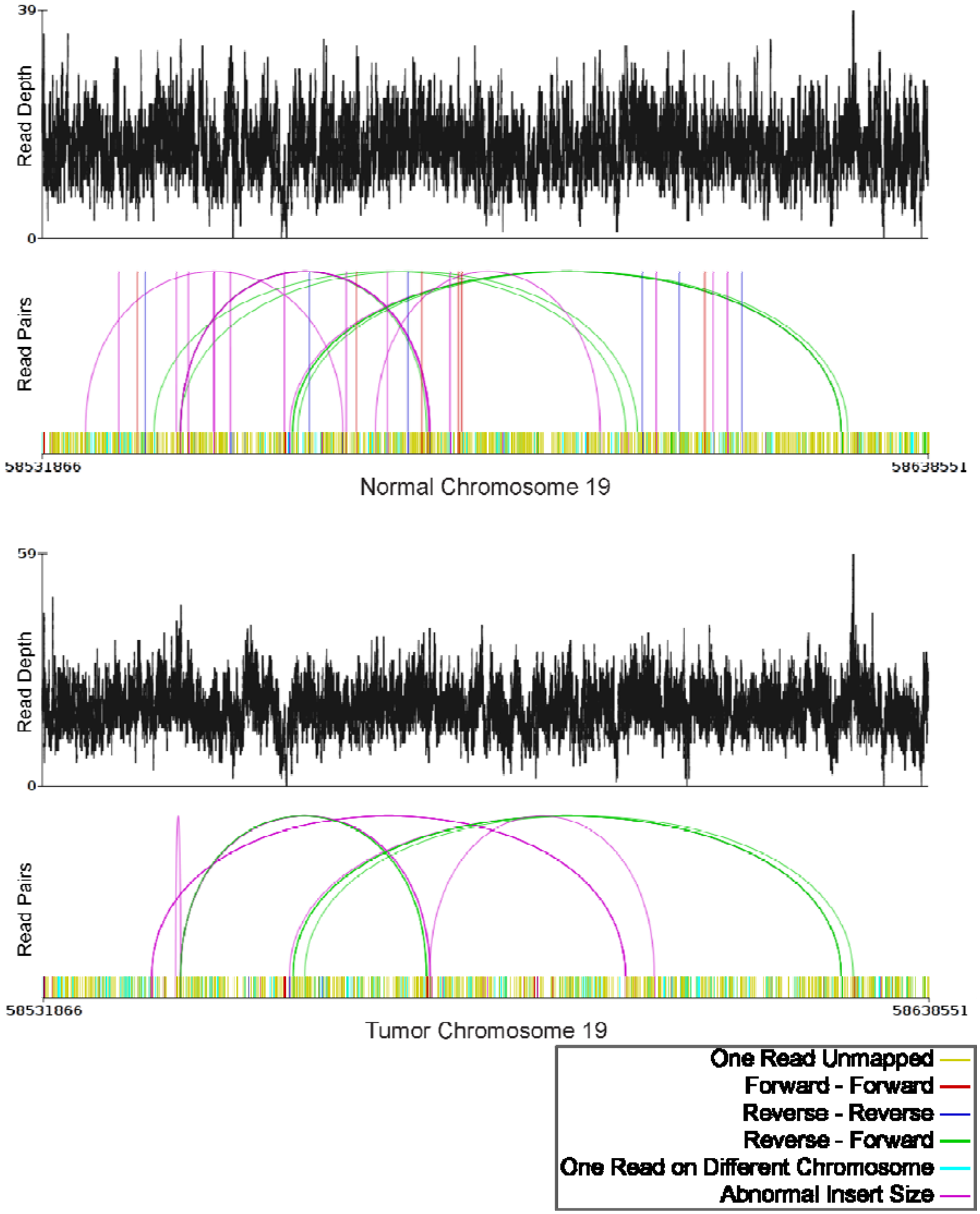
Supplementary Figure 10

(**b**) Inversion 2



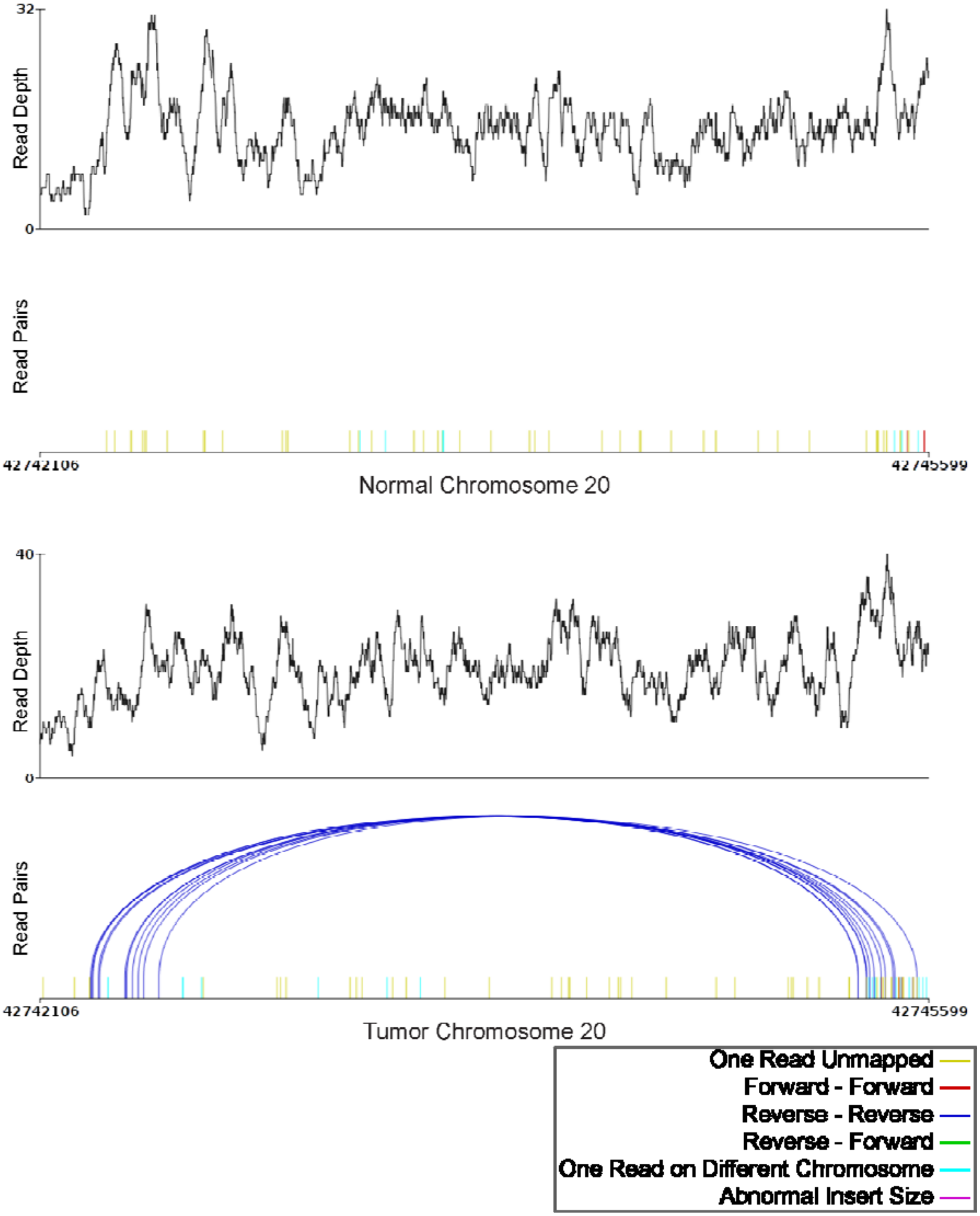
Supplementary Figure 10

(**c**) Inversion 3



Supplementary Figure 10

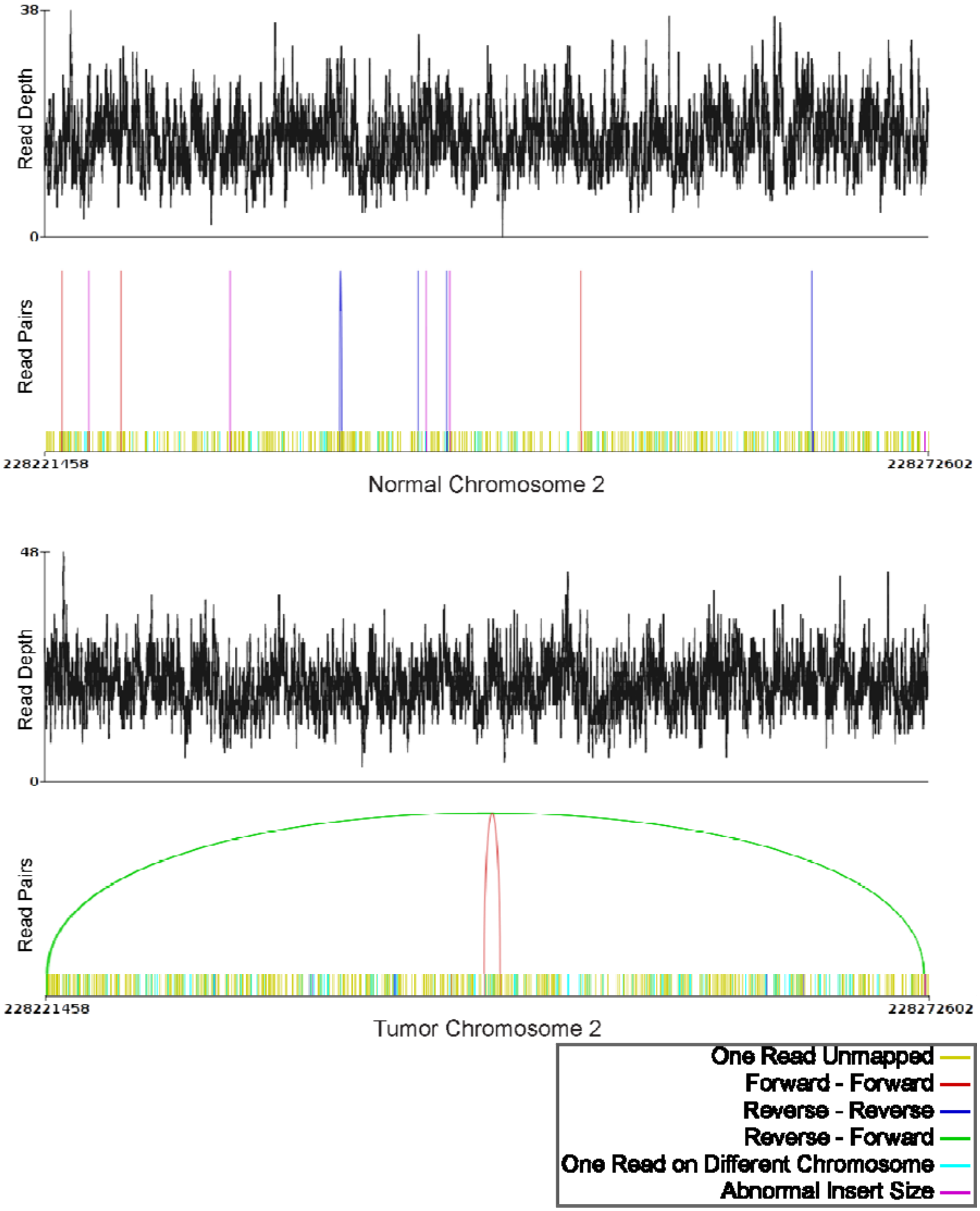
(**d**) Inversion 4



Nature Methods: doi: 10.1038/nmeth.1363

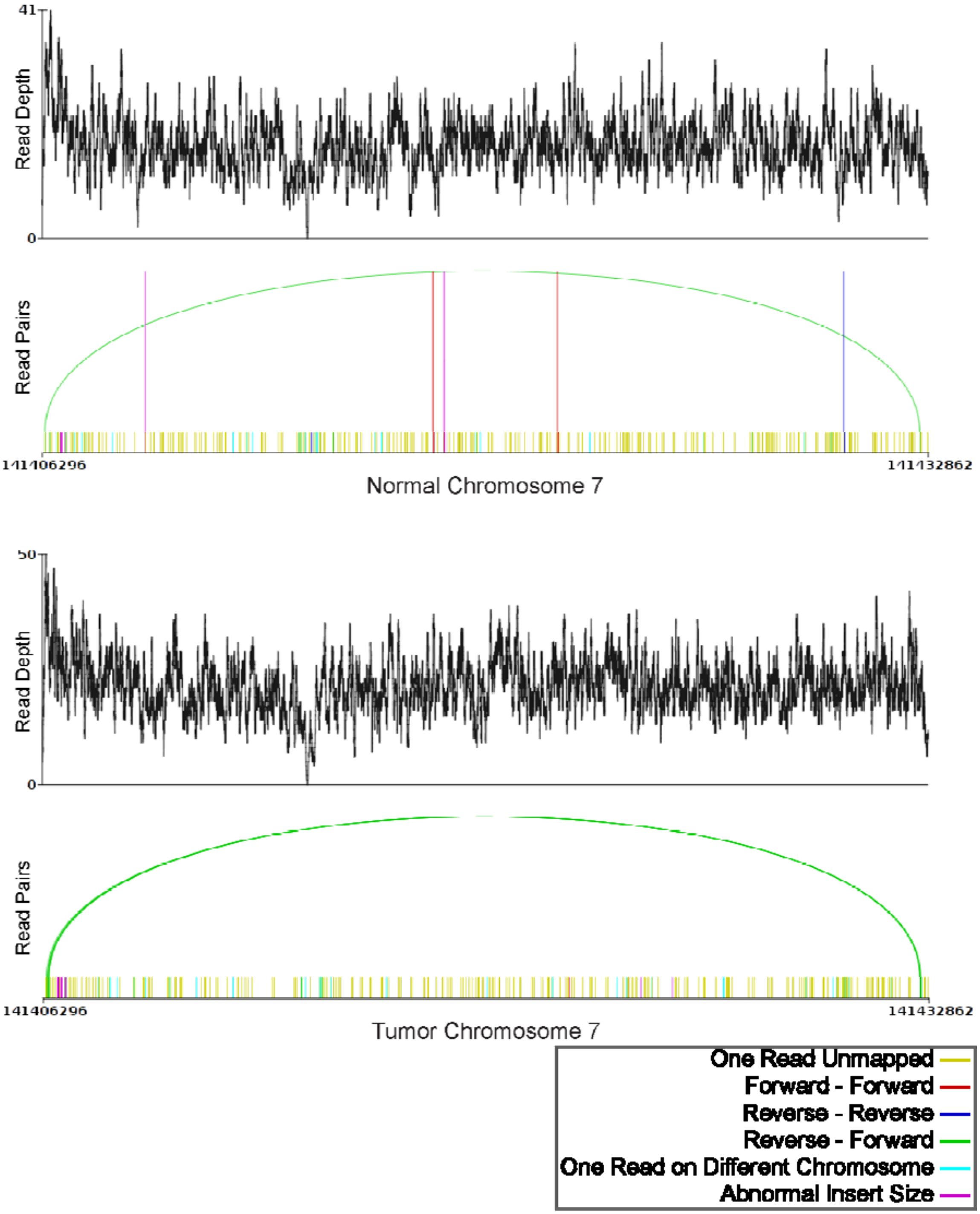
Supplementary Figure 10

(**e**) Intra-chromosomal translocation 1



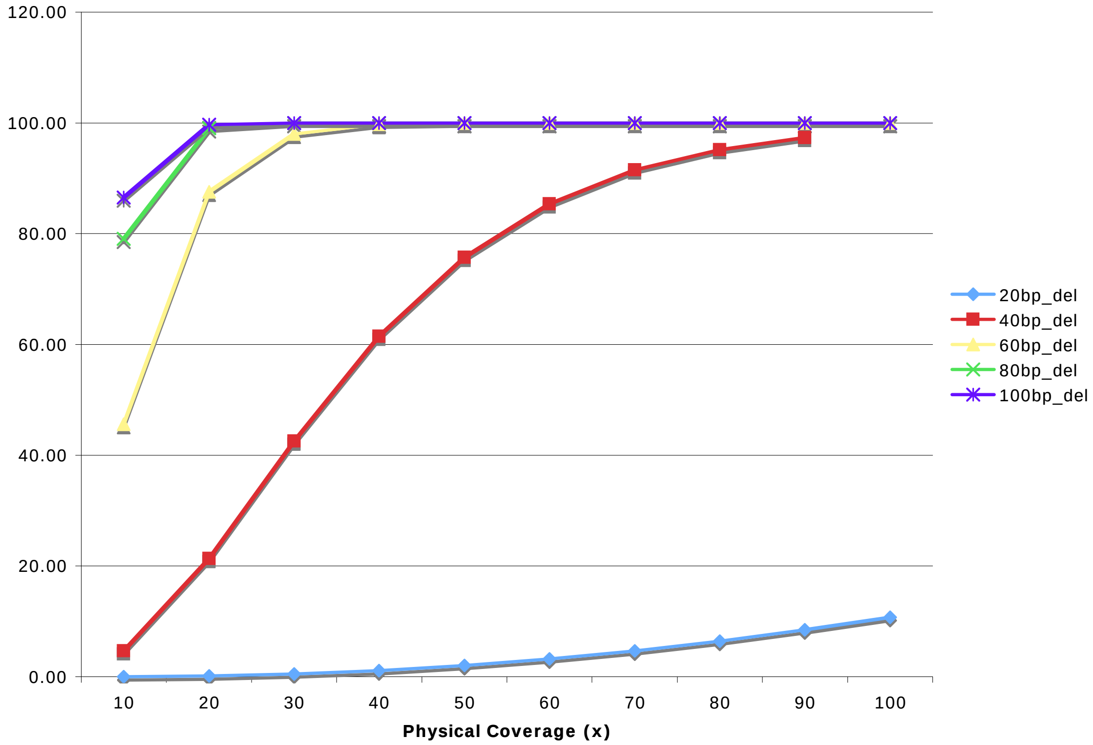
Supplementary Figure 10

(f) Intra-chromosomal translocation 2



Supplementary Figure 11 The analytic true positive rate w.r.t. variant size and coverage

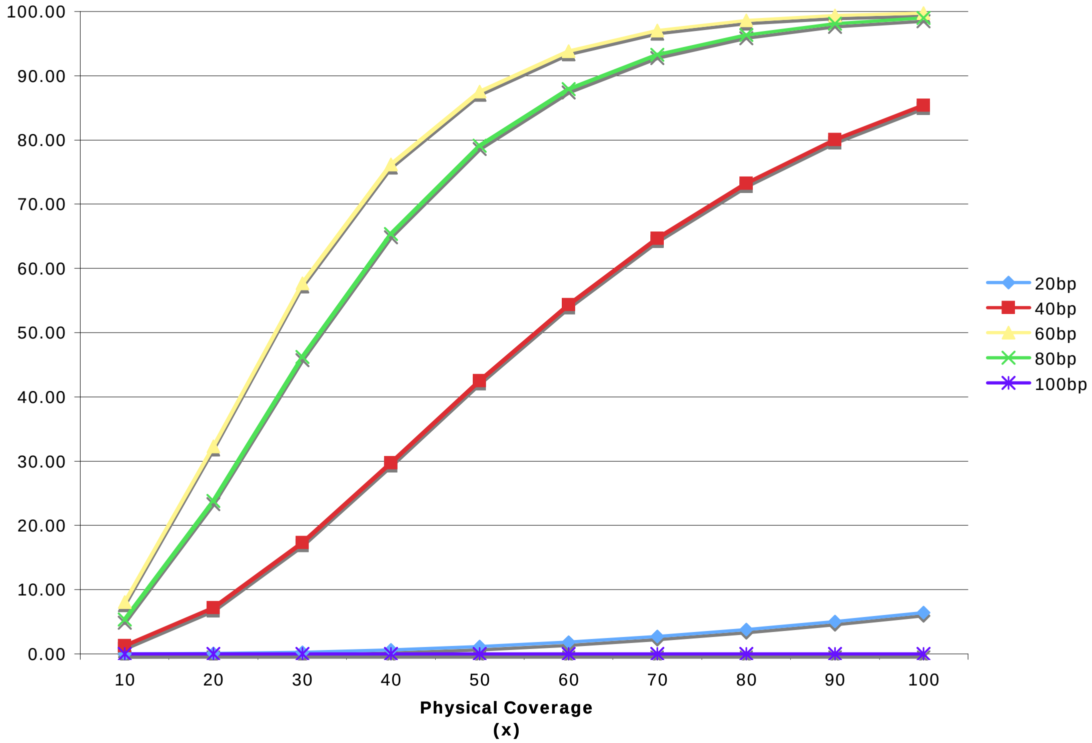
(**a**) Analytic true positive rates for the 20, 40, 60, 80, and 100 bp deletions using a 200 bp insert library (s.d. 20 bp and read length 50 bp) from 10 to 100 × physical coverage



Nature Methods: doi: 10.1038/nmeth.1363

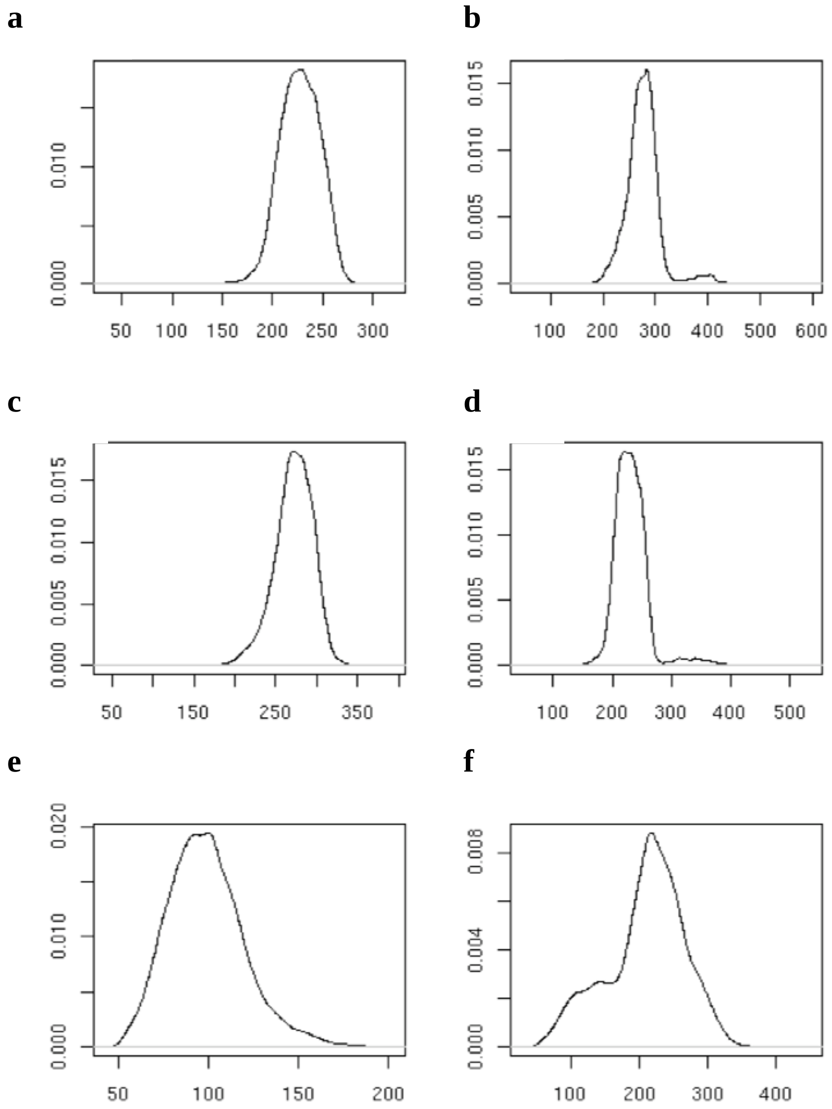
Supplementary Figure 11

(**b**) Analytic true positive rates for the 20, 40, 60, 80, and 100 bp insertions using a 200 bp insert library (s.d. 20 bp and read length 50 bp) from 10 to 100 × physical coverage



Supplementary Figure 12 The insert size distribution of the libraries in the AML project.

Plotted are insert size distributions from four tumor libraries (**a-d**) and two normal libraries (**e-f**). The x axis represents the insert size in bp, the y axis the probability density.



Supplementary Table 2 Number and type of structural variants detected by BreakDancerMax in the tumor-normal paired AML genome.

BreakDancerMax was run at a threshold of 3 s.d. and MAQ mapping quality > 35.

Score	# SV	# DEL	Perc_DGV5	# INS	# INV	# IRX
0	1078733	1052246	0.43%	15979	9378	1130
10	898114	871627	0.50%	15979	9378	1130
20	218800	196387	1.84%	12200	9083	1130
30	46060	29832	8.71%	9952	5169	1107
40	18762	7267	25.85%	5904	4510	1081
50	10986	3934	41.29%	2836	3158	1058
60	7087	3170	46.42%	1570	1382	965
70	5499	2835	48.96%	1212	644	808
80	4890	2608	50.77%	1052	508	722
90	4470	2437	52.07%	935	438	660

Supplementary Table 4 The paired end libraries used for the structural variation detection of the 1000 genomes trio individuals.

Library	Mean (bp)	SD (bp)	AvgReadLen (bp)	Sample ID	Sequence Depth (x)	Physical Coverage (x)
NA12878_libSC_1.1.map	130.669466	30.69132	36	NA12878	2.956	5.364707521
NA12878_libSC_2.1.map	173.370426	42.05891	36	NA12878	4.309	10.37573841
NA12891_libSC_1.1.map	127.291342	26.62561	36	NA12891	5.048	8.924537422
NA12891_libSC_2.1.map	143.931114	39.65158	35.880778	NA12891	2.407	4.827685055
NA12892_libSC_1.1.map	139.137373	38.50202	35.838347	NA12892	3.415	6.629130088
NA12892_libSC_2.1.map	150.54879	37.79393	35.888298	NA12892	4.331	9.084114403
H_IJ-NA19240-071608a.rmdup.map	273.12	22.79	35.26	NA19240	5.53	21.41737947
H_IJ-NA19240-042108a.rmdup.map	264.85	41.75	37.03	NA19240	13.372	47.82033756
H_IJ-NA19239-071608a.rmdup.map	300	30	35.49	NA19239	7.854	33.19526627
H_IJ-NA19239-042108a.rmdup.map	230	30	35.57	NA19239	9.547	30.8660388
H_IJ-NA19238-071608a.rmdup.map	274.76	50.89	35.38	NA19238	5.391	20.93317072
H_IJ-NA19238-042108a.rmdup.map	240.85	38.63	35.54	NA19238	9.036	30.61790377

Supplementary Table 5 Deletions detected by BreakDancerMax on chromosome 5 of NA12878.

Deletions predicted by BreakDancerMax (BD) on chromosome 5 of NA12878 were compared (50% overlap in interval) with those predicted in other samples in the CEU trio (NA12891 and NA12892), in the YRI trio (NA19240, NA19238, and NA19239), or identified using fosmid ESP[2], array CGH, and SNP arrays[3]. Results in the upper panel were obtained from the individual analysis, those in the lower panel from the pooled analysis, and those in the four columns using various s.d. and mapping quality (q) thresholds.

		4SD, q>=35					4SD, q>=10					3SD, q>=35					3SD, q>=10				
Independent		overlap			BD		overlap			BD		overlap			BD		overlap			BD	
	NA12891	88	120	73.33%	125	70.40%	115	160	71.88%	171	67.25%	98	140	70.00%	140	70.00%	126	179	70.39%	200	63.00%
	NA12892	98	133	73.68%	125	78.40%	126	163	77.30%	171	73.68%	103	144	71.53%	140	73.57%	124	171	72.51%	200	62.00%
	NA19240	88	246	35.77%	125	70.40%	114	308	37.01%	171	66.67%	97	309	31.39%	140	69.29%	120	375	32.00%	200	60.00%
	NA19238	87	164	53.05%	125	69.60%	108	208	51.92%	171	63.16%	90	201	44.78%	140	64.29%	112	249	44.98%	200	56.00%
	NA19239	86	235	36.60%	125	68.80%	115	309	37.22%	171	67.25%	98	309	31.72%	140	70.00%	130	403	32.26%	200	65.00%
	FOSMID_ESP	2	7	28.57%	125	1.60%	3	7	42.86%	171	1.75%	2	7	28.57%	140	1.43%	3	7	42.86%	200	1.50%
	SeqDel	1	4	25.00%	125	0.80%	2	4	50.00%	171	1.17%	1	4	25.00%	140	0.71%	2	4	50.00%	200	1.00%
	DGV	79	722	10.94%	125	63.20%	93	722	12.88%	171	54.39%	81	722	11.22%	140	57.86%	99	722	13.71%	200	49.50%
	WTSIaCGH	6	25	24.00%	125	4.80%	9	25	36.00%	171	5.26%	6	25	24.00%	140	4.29%	9	25	36.00%	200	4.50%
	Affy6	4	14	28.57%	125	3.20%	6	14	42.86%	171	3.51%	4	14	28.57%	140	2.86%	6	14	42.86%	200	3.00%
	Nimblegen	8	23	34.78%	125	6.40%	10	23	43.48%	171	5.85%	8	23	34.78%	140	5.71%	10	23	43.48%	200	5.00%
Pooled																					
	NA12891	122	146	83.56%	161	75.78%	171	209	81.82%	220	77.73%	137	170	80.59%	184	74.46%	210	258	81.40%	270	77.78%
	NA12892	132	152	86.84%	161	81.99%	175	199	87.94%	220	79.55%	146	180	81.11%	184	79.35%	202	241	83.82%	270	74.81%
	NA19240	110	302	36.42%	161	68.32%	142	376	37.77%	220	64.55%	126	427	29.51%	184	68.48%	165	533	30.96%	270	61.11%
	NA19238	108	220	49.09%	161	67.08%	133	286	46.50%	220	60.45%	121	277	43.68%	184	65.76%	151	365	41.37%	270	55.93%
	NA19239	117	323	36.22%	161	72.67%	148	408	36.27%	220	67.27%	134	430	31.16%	184	72.83%	173	545	31.74%	270	64.07%
	FOSMID_ESP	3	7	42.86%	161	1.86%	4	7	57.14%	220	1.82%	3	7	42.86%	184	1.63%	4	7	57.14%	270	1.48%
	SeqDel	2	4	50.00%	161	1.24%	3	4	75.00%	220	1.36%	2	4	50.00%	184	1.09%	3	4	75.00%	270	1.11%
	DGV	97	722	13.43%	161	60.25%	110	722	15.24%	220	50.00%	102	722	14.13%	184	55.43%	119	722	16.48%	270	44.07%
	WTSIaCGH	9	25	36.00%	161	5.59%	10	25	40.00%	220	4.55%	9	25	36.00%	184	4.89%	10	25	40.00%	270	3.70%
	Affy6	5	14	35.71%	161	3.11%	7	14	50.00%	220	3.18%	5	14	35.71%	184	2.72%	7	14	50.00%	270	2.59%
	Nimblegen	9	23	39.13%	161	5.59%	11	23	47.83%	220	5.00%	9	23	39.13%	184	4.89%	11	23	47.83%	270	4.07%

FOSMID ESP, validated deletions discovered through FOSMID end sequence profiling, Kidd et al. Nature 08
SeqDel, completely sequenced deletions identified from FOSMID_ESP
DGV, database of Genomic Variants
WTSIaCGH, CNV calls made by WTSI based on 2.1M Nimblegen array CGH, half of the calls are from the CNVs of the reference
Affy6, CNV calls made by Mccarroll et al. 2008 using Array 6.0
Nimblegen, deletion calls made by Sebat group using Nimblegen arrays.

Supplementary Table 6 Deletions detected by BreakDancerMax on chromosome 5 of NA19240.

Deletions predicted by BreakDancerMax (BD) on chromosome 5 of NA19240 were compared with those predicted in other samples in the YRI trio (NA19238 and NA19239), in the CEU trio (NA12878, NA12891, and NA12892), or identified using fosmid ESP[2], array CGH, and SNP arrays[3]. Results in the upper panel were obtained from the individual analysis, those in the lower panel from the pooled analysis, and those in the four columns using various s.d. and mapping quality (q) thresholds.

		4SD, q>35					4SD, q>10					3SD, q>35					3SD, q>10				
Independent		overlap			BD		overlap			BD		overlap			BD		overlap			BD	
	NA19238	126	164	76.83%	246	51.22%	161	208	77.40%	308	52.27%	149	201	74.13%	309	48.22%	190	249	76.31%	375	50.67%
	NA19239	168	235	71.49%	246	68.29%	212	309	68.61%	308	68.83%	207	309	66.99%	309	66.99%	258	403	64.02%	375	68.80%
	NA12878	88	125	70.40%	246	35.77%	114	171	66.67%	308	37.01%	97	140	69.29%	309	31.39%	120	200	60.00%	375	32.00%
	NA12891	78	120	65.00%	246	31.71%	102	160	63.75%	308	33.12%	88	140	62.86%	309	28.48%	110	179	61.45%	375	29.33%
	NA12892	92	133	69.17%	246	37.40%	113	163	69.33%	308	36.69%	94	144	65.28%	309	30.42%	113	171	66.08%	375	30.13%
	FOSMID_s8v	8	11	72.73%	246	3.25%	8	11	72.73%	308	2.60%	8	11	72.73%	309	2.59%	8	11	72.73%	375	2.13%
	DGV	123	722	17.04%	246	50.00%	142	722	19.67%	308	46.10%	130	722	18.01%	309	42.07%	144	722	19.94%	375	38.40%
	aCGH_WTSI	1	1	100.00%	246	0.41%	1	1	100.00%	308	0.32%	1	1	100.00%	309	0.32%	1	1	100.00%	375	0.27%
	Affy6	5	10	50.00%	246	2.03%	5	10	50.00%	308	1.62%	5	10	50.00%	309	1.62%	5	10	50.00%	375	1.33%
Pooled																					
	NA19238	177	220	80.45%	302	58.61%	235	286	82.17%	376	62.50%	228	277	82.31%	427	53.40%	309	365	84.66%	533	57.97%
	NA19239	259	323	80.19%	302	85.76%	323	408	79.17%	376	85.90%	339	430	78.84%	427	79.39%	441	545	80.92%	533	82.74%
	NA12878	111	161	68.94%	302	36.75%	142	220	64.55%	376	37.77%	126	184	68.48%	427	29.51%	165	270	61.11%	533	30.96%
	NA12891	99	146	67.81%	302	32.78%	132	209	63.16%	376	35.11%	114	170	67.06%	427	26.70%	158	258	61.24%	533	29.64%
	NA12892	105	152	69.08%	302	34.77%	129	199	64.82%	376	34.31%	114	180	63.33%	427	26.70%	143	241	59.34%	533	26.83%
	FOSMID_s8v	8	11	72.73%	302	2.65%	8	11	72.73%	376	2.13%	8	11	72.73%	427	1.87%	8	11	72.73%	533	1.50%
	DGV	138	722	19.11%	302	45.70%	152	722	21.05%	376	40.43%	147	722	20.36%	427	34.43%	168	722	23.27%	533	31.52%
	aCGH_WTSI	1	1	100.00%	302	0.33%	1	1	100.00%	376	0.27%	1	1	100.00%	427	0.23%	1	1	100.00%	533	0.19%
	Affy6	5	10	50.00%	302	1.66%	5	10	50.00%	376	1.33%	5	10	50.00%	427	1.17%	5	10	50.00%	533	0.94%

FOSMID ESP, validated deletions discovered through FOSMID end sequence profiling, Kidd et al. Nature 08
DGV, database of Genomic Variants
WTSIaCGH, CNV calls made by WTSI based on 2.1M Nimblegen array CGH, half of the calls are from the CNVs of the reference
Affy6.0, CNV calls made by Mccarroll et al. 2008 using Array 6.0

Supplementary Table 7 Analytic true positive rate in simulated structural variant detection using a 200 bp insert library.
Inserts are normally distributed with a s.d. of 20 bp. Read length equals to 50 bp.

Phy_Cov	Seq_Cov	deletion	total_del	percent	insertion	total_ins	percent	indels	total_indels	percent
10	5	191	425	44.94%	0	415	0.00%	191	840	22.74%
20	10	213	425	50.12%	15	415	3.61%	228	840	27.14%
30	15	237	425	55.76%	56	415	13.49%	293	840	34.88%
40	20	256	425	60.24%	77	415	18.55%	333	840	39.64%
50	25	278	425	65.41%	94	415	22.65%	372	840	44.29%
60	30	294	425	69.18%	114	415	27.47%	408	840	48.57%
70	35	302	425	71.06%	126	415	30.36%	428	840	50.95%
80	40	314	425	73.88%	144	415	34.70%	458	840	54.52%
90	45	325	425	76.47%	153	415	36.87%	478	840	56.90%
100	50	335	425	78.82%	158	415	38.07%	493	840	58.69%
110	55	335	425	78.82%	174	415	41.93%	509	840	60.60%
120	60	352	425	82.82%	182	415	43.86%	534	840	63.57%
130	65	364	425	85.65%	194	415	46.75%	558	840	66.43%
140	70	384	425	90.35%	202	415	48.67%	586	840	69.76%
150	75	384	425	90.35%	231	415	55.66%	615	840	73.21%
160	80	399	425	93.88%	231	415	55.66%	630	840	75.00%
170	85	399	425	93.88%	238	415	57.35%	637	840	75.83%
180	90	425	425	100.00%	265	415	63.86%	690	840	82.14%
190	95	425	425	100.00%	265	415	63.86%	690	840	82.14%
200	100	425	425	100.00%	278	415	66.99%	703	840	83.69%
210	105	425	425	100.00%	281	415	67.71%	706	840	84.05%
220	110	425	425	100.00%	307	415	73.98%	732	840	87.14%
230	115	425	425	100.00%	307	415	73.98%	732	840	87.14%

Supplementary Table 8 Analytic true positive rate in simulated structural variant detection using a 400 bp insert library.

Inserts are normally distributed with a s.d. of 40 bp. Read length equals to 50 bp.

Physical coverage	Sequence_Coverage	deletion	total_del		insertion	total_ins		indels	total_indels	
10	2.5	138	425	32.47%	29	415	6.99%	167	840	19.88%
20	5	156	425	36.71%	58	415	13.98%	214	840	25.48%
30	7.5	169	425	39.76%	74	415	17.83%	243	840	28.93%
40	10	181	425	42.59%	86	415	20.72%	267	840	31.79%
50	12.5	187	425	44.00%	91	415	21.93%	278	840	33.10%
60	15	198	425	46.59%	98	415	23.61%	296	840	35.24%
70	17.5	199	425	46.82%	102	415	24.58%	301	840	35.83%
80	20	204	425	48.00%	110	415	26.51%	314	840	37.38%
90	22.5	210	425	49.41%	120	415	28.92%	330	840	39.29%
100	25	219	425	51.53%	123	415	29.64%	342	840	40.71%
110	27.5	221	425	52.00%	131	415	31.57%	352	840	41.90%
120	30	229	425	53.88%	133	415	32.05%	362	840	43.10%
130	32.5	234	425	55.06%	140	415	33.73%	374	840	44.52%
140	35	237	425	55.76%	146	415	35.18%	383	840	45.60%
150	37.5	252	425	59.29%	150	415	36.14%	402	840	47.86%
160	40	256	425	60.24%	156	415	37.59%	412	840	49.05%
170	42.5	267	425	62.82%	171	415	41.20%	438	840	52.14%
180	45	271	425	63.76%	176	415	42.41%	447	840	53.21%
190	47.5	271	425	63.76%	176	415	42.41%	447	840	53.21%
200	50	278	425	65.41%	183	415	44.10%	461	840	54.88%
210	52.5	281	425	66.12%	188	415	45.30%	469	840	55.83%
220	55	294	425	69.18%	201	415	48.43%	495	840	58.93%
230	57.5	294	425	69.18%	206	415	49.64%	500	840	59.52%
240	60	302	425	71.06%	206	415	49.64%	508	840	60.48%
250	62.5	314	425	73.88%	215	415	51.81%	529	840	62.98%
260	65	314	425	73.88%	220	415	53.01%	534	840	63.57%
270	67.5	318	425	74.82%	220	415	53.01%	538	840	64.05%
280	70	318	425	74.82%	233	415	56.14%	551	840	65.60%
290	72.5	325	425	76.47%	233	415	56.14%	558	840	66.43%
300	75	325	425	76.47%	241	415	58.07%	566	840	67.38%
310	77.5	335	425	78.82%	241	415	58.07%	576	840	68.57%
320	80	335	425	78.82%	253	415	60.96%	588	840	70.00%
330	82.5	352	425	82.82%	253	415	60.96%	605	840	72.02%
340	85	352	425	82.82%	261	415	62.89%	613	840	72.98%
350	87.5	364	425	85.65%	261	415	62.89%	625	840	74.40%
360	90	364	425	85.65%	290	415	69.88%	654	840	77.86%
370	92.5	364	425	85.65%	290	415	69.88%	654	840	77.86%
380	95	384	425	90.35%	297	415	71.57%	681	840	81.07%
390	97.5	384	425	90.35%	297	415	71.57%	681	840	81.07%
400	100	384	425	90.35%	297	415	71.57%	681	840	81.07%
410	102.5	399	425	93.88%	324	415	78.07%	723	840	86.07%
420	105	399	425	93.88%	324	415	78.07%	723	840	86.07%
430	107.5	425	425	100.00%	324	415	78.07%	749	840	89.17%
440	110	425	425	100.00%	337	415	81.20%	762	840	90.71%
450	112.5	425	425	100.00%	337	415	81.20%	762	840	90.71%
460	115	425	425	100.00%	337	415	81.20%	762	840	90.71%
470	117.5	425	425	100.00%	363	415	87.47%	788	840	93.81%
480	120	425	425	100.00%	363	415	87.47%	788	840	93.81%
490	122.5	425	425	100.00%	363	415	87.47%	788	840	93.81%
500	125	425	425	100.00%	363	415	87.47%	788	840	93.81%

Supplementary Notes Additional results in simulation

Variant assembly from the simulated data

We have tried to assemble the 840 known regions on chromosome 17 of Venter's genome using our simulation data at 100 ×. Of the 838 regions examined, 408 structural variants were called. 376 (92%) are of the correct type, 243 (65%) of which also have the exact size. 46/376 (12.2%) of the correct assembly calls were not called by BreakDancerMax while 291/621 (47%) of the correct BreakDancerMax structural variant calls were not called by assembly. The primary difficulty is that this set contain many simple/tandem repeats that are difficult to assemble.

The Effect of MAQ mapping quality threshold in simulation

The performance is generally similar but less sensitive with mapping quality > 35 than with mapping quality > 10. For example, at 100 × coverage, BreakDancerMax achieved a 423/844 (50.12%) true positive rate and 36/460 (8.04%) false positive rate with mapping quality > 35.

References

1. Li, H., Ruan, J. & Durbin, R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**, 1851-8 (2008).
2. Kidd, J.M. et al. Mapping and sequencing of structural variation from eight human genomes. *Nature* **453**, 56-64 (2008).
3. McCarroll, S.A. et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat Genet* **40**, 1166-74 (2008).