Journal of Human Genetics (2014) 59, 649–654

doi:10.1038/jhg.2014.88

Journal of human genetics (http://www.nature.com/jhg/index.html)

Original Article

Precise detection of chromosomal translocation or inversion breakpoints by whole-genome sequencing

Toshifumi Suzuki[1,2], Yoshinori Tsurusaki[1], Mitsuko Nakashima[1], Noriko Miyake[1], Hirotomo Saitsu[1], Satoru Takeda[2], Naomichi Matsumoto[1]

1. Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan

2. Department of Obstetrics and Gynecology, Juntendo University Faculty of Medicine, Tokyo 113-8421, Japan

Correspondence to:

Naomichi Matsumoto

Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan

Tel.: +81-45-787-2606

Fax: +81-45-786-5219

E-mail: naomat@yokohama-cu.ac.jp (N. Matsumoto)

Running title

Whole genome sequencing of structural variations

Keywords (5 keywords < 8 keywords)

Structural variations, breakpoints, whole-genome sequencing, BreakDancer, Integrative Genomics Viewer

Abstract

Structural variations (SVs), including translocations, inversions, deletions, and duplications, are potentially associated with Mendelian diseases and contiguous gene syndromes. Determination of SV-related breakpoints at the nucleotide level is important to reveal genetic causes for diseases. Whole -genome sequencing (WGS) by next-generation sequencers is expected to determine structural abnormalities more directly and efficiently than conventional methods. In this study, 14 SVs (nine balanced translocations, one inversion and four microdeletions) in nine patients were analyzed by WGS with a shallow (5×) to moderate read coverage (20×). Among 28 breakpoints (as each SV has two breakpoints.), 19 SV breakpoints had been determined previously at the nucleotide level by any other methods and nine were uncharacterized. BreakDancer and Integrative Genomics Viewer determined 20 breakpoints (16 translocation, two inversion and two deletion breakpoints), but did not detect eight breakpoints (two translocation and six deletion breakpoints). These data indicate the efficacy of WGS for the precise determination of translocation and inversion breakpoints.

Introduction

Structural variations (SVs), including translocations, inversions, deletions, and duplications, potentially lead to human genetic diseases arising from disruption and dosage changes of functionally important genes [1]. In particular, apparently balanced chromosomal rearrangements have been frequently associated with human diseases, such as premature ovarian failure, Sotos syndrome, Peters anomaly, testicular atrophy, Mowat-Wilson syndrome, developmental delay, and intellectual disability[2-6]. The incidence of apparently balanced chromosomal rearrangements is in the range of from 1/500–1/625 [7]. Precise structural analysis of SVs and their breakpoints may lead to identification of the genetic causes of such diseases. The conventional methods to determine SV breakpoints including fluorescence *in situ* hybridization (FISH) using bacterial artificial chromosome clones; Southern blot hybridization; and inverse PCR or long-range PCR, which are laborious, time consuming, and have limited success rates[8]. Recently, whole genome sequencing (WGS) using next-generation sequencers has provided new avenue for SV analysis[2, 7-9]. However, accurate detection of SV breakpoints using WGS has not been fully established. In this study, WGS was used to analyze nine patients having 14 SVs. As each SV has two breakpoints, 28 SV breakpoints were analyzed. Among them, 19 SVs had already been determined by conventional methods in our previous studies[6, 10-13] were used as a training set and nine other uncharacterized breakpoints were analyzed. The purpose of this study is to investigate the chromosomal breakpoint of the patients in whom G-banded karyotyping was already performed. The results of WGS analysis of these patients are presented.

Material and methods

Subjects

Nine patients, including eight who were reported previously[6, 10-14] were included in this study. G-banded karyotyping was performed for all patients (Table 1). The nine patients possess a total of 14 SVs (nine translocations, one inversion and four microdeletions) (Tables 1 and 2). As each SV event involves two breakpoints, a total of 28 SV breakpoints are the targets of this study. Among 28 SV breakpoints, 19 were

previously determined at the nucleotide level by conventional methods[6, 10-14] and used as a training set (Tables 1 and 2). Peripheral blood samples were collected from all patients after obtaining written informed consent. Genomic DNA was extracted from leukocytes using the QuickGene-610L DNA extraction system (Fujifilm, Tokyo, Japan) was used to extract genomic DNA from leukocytes, according to the manufacturer's instruction. The institutional review board of Yokohama City University School of Medicine approved the study.

Whole genome sequencing

Briefly, 1 μg of genomic DNA with each sample was shared using the Covaris model S2 system (Covaris, Woburn, MA, USA). The target size was 350 bp. DNA was prepared using the TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA, USA) or the TruSeq DNA PCR-Free Sample Prep Kit (Illumina). The HiSeq 2000 or 2500 platform (Illumina) performed WGS with 101-bp paired-end reads. Sequence-control, software real-time analysis and CASAVA software v1.8.2 (Illumina) performed image analysis and base calling.

Structural variation breakpoint analysis

The analytical flow chart is illustrated in Figure 1A. Burrows-Wheeler Aligner (BWA-MEM) v0.7.1 [15] with default parameters was used to map the data to the hg19 human genome reference sequence from UCSC Genome Browser. BreakDancerMax (BD) ver.1.4.4 with the default setting was used to validate breakpoints of SVs, including translocations, inversions, and deletions at the nucleotide level using the WGS data (Binary Alignment/Map format). A Poisson model [16] was used to calculate the confidence score for each candidate variant. BD is able to identify inter-chromosomal translocation (CTX), inversion (INV), and deletion (DEL). We focused on variant reads adjacent to chromosomal breakpoint positions from the information of G-banded karyotyping. Aligned reads adjacent to SV breakpoints were visualized and carefully evaluated using Integrative Genomics Viewer (IGV)[17]. In IGV, chimeric read pairs that mapped to different chromosomes at each end were predicted to cover translocation breakpoints (Figure 1B). Discordant read pairs that mapped to the reference genome with abnormal distance and/or

orientation were predicted to cover breakpoints of inversion and insertion or deletion. Soft-clipped reads consisting of two different sequences (within a single read) mapped to discontinuous parts of chromosome(s) that potentially covered SV breakpoints (Figure 1B).

Validation of chromosomal breakpoint positions

PCR and Sanger sequencing confirmed all potential SV breakpoints. Primer3Plus (http://primer3plus.com/) was used to design the primer sequences. PCR was performed using KOD FX Neo polymerase (TOYOBO, Osaka, Japan). Primer sequences and PCR conditions are available on request. PCR products were electrophoresed through a 1.0 % agarose gel and sequenced by Sanger sequencing on an ABI3500xl sequencer (Applied Biosystems, Foster City, CA, USA).

Results

We analyzed 28 SVs in nine patients. The analytical workflow of the respective patients is shown in Figure 2. Mean read depth of WGS was in the range of 5.95–21.92× (Table 1). Initially, genomic DNA of each patient was sequenced using TruSeq DNA Sample Prep Kit. However, the read coverage did not reach the expected level because of high PCR duplication rates (Supplementary Table 1). Therefore, we switched the kit to the PCR-Free Sample Prep Kit and successfully attained the expected read coverage (Supplementary Table 1). We were able to detect 18 SV breakpoints of 28 (64.3 %) using BD (Tables 1 and 2, and Figure 2).

For translocations and an inversion, the numbers of CTX and INV read by BD through a whole genome had a range of 61–4,698 (Supplementary Table 2). We then focused on those related to the involved chromosome(s) by translocation or inversion, and found that 1–39 CTX or INV reads remained as candidates (Supplementary Table 2). Among the data, 12–31 chimeric read pairs and 28 discordant read pairs were carefully evaluated, which may have spanned SV breakpoints by IGV in patients 2, 3, 4, 6, 7, 8, and 9 (Table 1, Figure 3, and Supplementary Figure 1). In patients 1, 3, 4, 7, 8, and 9, one to six soft-clipped reads in IGV covered the SV breakpoints accurately (Table 1, Figure 3, and Supplementary

Table 3). In patient 1, one CTX read by BD was a false-positive (Table S2), however, one soft-clipped read by IGV covered the 1q32 breakpoint (Table 1, Supplementary Table 3, and Figures 2 and 3). The 9q13 breakpoint region was undetected by either BD or IGV (Table 1, and Figures 2 and 3), because the genomic sequences of the region around centromeric 9q13 are unavailable. In combination with BD and IGV, 16 out of 18 translocation (88.9%) and two out of two inversion breakpoints (100%) were successfully determined.

Deletions in patient 3 (a 4,192-bp deletion in the X-chromosome and a 7,029-bp deletion in chromosome 4) and patient 4 (a 806,297-bp deletion in chromosome 7 and an approximately 4.6-Mb deletion in chromosome 15) were determined previously by conventional methods[12, 14]. A total of 1,943–1,945 DEL reads were called by BD and 51–159 DEL reads related to the involved chromosomes remained as candidates; however, only one DEL read accurately covered the deletion breakpoint in chromosome 7. Therefore, we were able to detect the deletion breakpoints by BD in two of eight deletion breakpoints (25%) (Table 2 and Supplementary Table 4). Using IGV, nine discordant read pairs accurately covered the deletion breakpoints in patient 4 (Table 2) Furthermore, two soft-clipped reads in IGV accurately covered the breakpoint in patient 4 (Table 2 and Supplementary Table 3). Of note, in patient 3, the deletions were adjacent to translocation breakpoints (Supplementary Figure 2).

Discussion

In this study, 20 out of 28 SV breakpoints were successfully determined by WGS (71.4%). A relatively shallow (5×) read coverage enabled us to determine the translocation breakpoints (Table 1). Translocation and inversion breakpoints were highly detected by our method (88.9-100%), athough the detection rate of deletion breakpoints was relatively low (25%). The false negative rates by BD solely and BD combined with IGV were 10 out of 28 (35.7%) and 8 out of 28 (28.6%), respectively. The total number of called reads by BD including CTX, INV, and DEL were quite different among samples (61-4,698) (Supplementary Table 2 and 4). The estimation of the false positive rate (FPR) was difficult, because large and varying numbers of reads were called by BD. Therefore, FPR was unknown in the present study.

In patient 1, it was expected to be difficult or impossible to determine the 9q13 breakpoint because the genomic sequence data of the 9q13 centromeric region were unavailable. However, although no chimeric read pairs covering the breakpoints were obtained, one soft-clipped read accurately determined the der(1) breakpoint at the nucleotide level (Figures 2 and 3, Supplementary Table 3). The sequence with unknown origin in the soft-clipped read should be derived from the centromeric region at 9q13, as shown in a previous study[10].

In patient 3, two CTX reads were called by BD (Supplementary Table 2). Interestingly, deletions existed adjacent to the reciprocal translocation in both chromosomes X and 4 (Supplementary Figure 2). However, BD did not call any DEL presumably because the sequences either side of the deletion breakpoints are connected to different chromosomes.

In patients 4 (for t(9;14)), 8, and 9, translocation or inversion breakpoints had not been determined previously at the nucleotide level by any conventional method. We were able to determine the breakpoint positions of these patients by BD (Table 1, and Figures 2 and 3). A total of 14–31 chimeric read pairs or 28 discordant read pairs covered the SV breakpoints (Table 1 and Figure 3). Among the soft-clipped reads, 2–6 reads also covered the precise breakpoints, including eight or nine-nucleotide insertions of unknown origin (Table1, and Supplementary Table 3).

In patient 5, chromosomal breakpoints could not be detected by our method (Table 1, Supplementary Table 2). Breakpoint sequences were determined in the previous study, and no repetitive sequences and structural abnormalities were found around the breakpoints regardless of the relatively reasonable read coverage at the breakpoints (17 reads or 22 reads at Xq22.3 and 2p14, respectively). The reason for detection failure remains elusive.

The reason for the low detection rate of deletion breakpoints is that BD can only detect deletions with the sizes of <1 Mb. One 4.6-Mb deletion in which we were unable to determine deletion breakpoints was far beyond the size of the detection limit of BD. In addition, two deletions were adjacent to the translocation breakpoints in patient 3. Therefore, the two deletions were complicated. Each end of the two deletions

and two translocation breakpoints are in the same location in patient 3. The only deletion in which we could determine breakpoints was the only simple 806-kb deletion within a single chromosome.

In conclusion, our approach, using shallow-to-moderate WGS data, enabled us to determine accurately the breakpoints of SVs especially for chromosomal translocations and inversions. Conventional karyotyping, as well as the approximate localization of the SV breakpoints by FISH, was absolutely important for our WGS-based breakpoint detection. WGS analysis should be first considered for determination of SV breakpoints in NGS era.

Conflict of interest: The authors declare no conflicts of interest.

Acknowledgements

We would like to thank all the patients and their families for their participation in this study. We also thank Nobuko Watanabe for her technical assistance. This work was supported by the Ministry of Health, Labour and Welfare of Japan; the Japan Society for the Promotion of Science (a Grant-in-Aid for Scientific Research (B), and a Grant-in-Aid for Scientific Research (A)); the Takeda Science Foundation; the fund for Creation of Innovation Centers for Advanced Interdisciplinary Research Areas Program in the Project for Developing Innovation Systems; the Strategic Research Program for Brain Sciences; and a Grant-in-Aid for Scientific Research on Innovative Areas (Transcription Cycle) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Figure Legends

Figure 1. Schematic presentation of the analytical flow and graphical presentation of chimeric read pairs, discordant read pairs, and a soft-clipped read. A) Flow chart of structural variation (SV) breakpoint analysis using whole genome sequencing (WGS) data. Burrows-Wheeler Aligner (BWA-MEM) v0.7.1 was used to map WGS data to human genome hg19. Inter-chromosomal translocation (CTX), inversion (INV), and deletion (DEL) were predicted using BreakDancerMax (BD). CTX, INV, and DEL calls were selected only from the involved chromosomes. Furthermore, we focused on chimeric read pairs, discordant read pairs, and soft-clipped reads using Integrative Genomics Viewer (IGV). Finally, we confirmed SV breakpoints by Sanger sequencing. B) Graphical images of chimeric read pairs, discordant read pairs, and a soft-clipped read. Upper; illustration of chimeric read pairs covering a translocation breakpoint, t(A;B). A soft-clipped read covers the breakpoint. Middle; intrachromosomal inversion. P and Q are marked to show the orientation. Inversion may lead to discordant read pairs. Lower; discordant read pairs detect intrachromosomal deletion

Figure 2. Flow chart of the analysis in nine patients. Left and right panels are shown in the detection of translocation/inversion and deletion breakpoints, respectively. BD was first used to detect SV breakpoints. We then carefully evaluated the chimeric read pairs, discordant read pairs and soft-clipped reads using IGV. Validation was performed by the Sanger method

Figure 3. Breakpoint junction sequences in nine patients. Upper, middle, and lower sequences indicate reference sequences of one end of an SV, derivative/deleted chromosome, and the other end of the SV. Breakpoint positions are marked with short longitudinal lines. Numbers are based on the nucleotide position in the UCSC genome browser coordinates, February 2009 version (hg19). Bold sequences are novel sequences that have never been deposited to any databases. Boxes indicate nucleotide insertions. Total numbers of chimeric read pairs and soft-clipped reads are described. cen: centromere, tel: telomere

References

1. Feuk, L., Carson, A.R. & Scherer, S.W. Structural variation in the human genome. *Nature reviews. Genetics.* **7**, 85-97 (2006).

2. Utami, K.H., Hillmer, A.M., Aksoy, I., Chew, E.G., Teo, A.S., Zhang, Z. et al. Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. *PloS one.* **9**, e90852 (2014).

3. Vandeweyer, G. & Kooy, R.F. Balanced translocations in mental retardation. *Human genetics.* **126**, 133-147 (2009).

4. Fantes, J.A., Boland, E., Ramsay, J., Donnai, D., Splitt, M., Goodship, J.A. et al. FISH mapping of de novo apparently balanced chromosome rearrangements identifies characteristics associated with phenotypic abnormality. *American journal of human genetics.* **82**, 916-926 (2008).

5. Rizzolio, F., Bione, S., Sala, C., Goegan, M., Gentile, M., Gregato, G. et al. Chromosomal rearrangements in Xq and premature ovarian failure: mapping of 25 new cases and review of the literature. *Human reproduction.* **21**, 1477-1483 (2006).

6. Imaizumi, K., Kimura, J., Matsuo, M., Kurosawa, K., Masuno, M., Niikawa, N. et al. Sotos syndrome associated with a de novo balanced reciprocal translocation t(5;8)(q35;q24.1). *American journal of medical genetics.* **107**, 58-60 (2002).

7. Dong, Z., Jiang, L., Yang, C., Hu, H., Wang, X., Chen, H. et al. A Robust Approach for Blind Detection of Balanced Chromosomal Rearrangements with Whole-Genome Low-Coverage

Sequencing. *Human mutation.* (2014).

8. Schluth-Bolard, C., Labalme, A., Cordier, M.P., Till, M., Nadeau, G., Tevissen, H. et al. Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *Journal of medical genetics.* **50**, 144-150 (2013).

9. Abel, H.J. & Duncavage, E.J. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. *Cancer genetics.* **206**, 432-440 (2013).

10. Saitsu, H., Osaka, H., Sugiyama, S., Kurosawa, K., Mizuguchi, T., Nishiyama, K. et al. Early infantile epileptic encephalopathy associated with the disrupted gene encoding Slit-Robo Rho GTPase activating protein 2 (SRGAP2). *American journal of medical genetics. Part A.* **158A**, 199-205 (2012).

11. Saitsu, H., Igarashi, N., Kato, M., Okada, I., Kosho, T., Shimokawa, O. et al. De novo 5q14.3 translocation 121.5-kb upstream of MEF2C in a patient with severe intellectual disability and early-onset epileptic encephalopathy. *American journal of medical genetics. Part A.* **155A**, 2879-2884 (2011).

12. Nishimura-Tadaki, A., Wada, T., Bano, G., Gough, K., Warner, J., Kosho, T. et al. Breakpoint determination of X;autosome balanced translocations in four patients with premature ovarian failure. *Journal of human genetics.* **56**, 156-160 (2011).

13. Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T. et al. Haploinsufficiency of NSD1 causes Sotos syndrome. *Nature genetics.* **30**, 365-366 (2002).

14. Saitsu, H., Kurosawa, K., Kawara, H., Eguchi, M., Mizuguchi, T., Harada, N. et al. Characterization of the complex 7q21.3 rearrangement in a patient with bilateral split-foot malformation and hearing loss. *American journal of medical genetics. Part A.* **149A**, 1224-1230 (2009).

15. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England).* **25**, 1754-1760 (2009).

16. Chen, K., Wallis, J.W., McLellan, M.D., Larson, D.E., Kalicki, J.M., Pohl, C.S. et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nature methods.* **6**, 677-681 (2009).

17. Thorvaldsdottir, H., Robinson, J.T. & Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics.* **14**, 178-192 (2013).

Table 1 Detection of chromosomal breakpoints by BreakDancer (Translocation or inversion)

Patient	Karyotype	Breakpoint detection by other methods (hg19)	Mean coverage (x)	preparation	PCR duplication (%)	BreakDancer detection	IGV			References
							Chimeric read pair counts	Discordant read pair counts	Soft-clipped read counts	
1	46,XX,t(1;9)(q32;q13)	der(1): 206574512 der(9): **undetermined**	15.23	PCR-free	1.6	der(1): **negative** der(9): **negative**	der(1):**0** der(9):**0**	-	der(1): 1 der(9): **0**	Saitsu et al. (2012) [10]
2	46,XX,t(5;15)(q13.3;q26.1)	der(5): 88300578 der(15): 98166676	9.18	PCR	7.0	der(5): positive der(15): positive	der(5):14 der(15):14	-	der(5): **0** der(15): **0**	Saitsu et al. (2011) [11]
3	46,X,t(X;4)(q21.3;p15.2)	der(X): 107436209 der(4): 12244290	21.92	PCR-free	1.9	der(X): positive der(4): positive	der(X):14 der(4):12	-	der(X): 5 der(4): 2	Nishimura-Tadaki et al. (2011) [12]
4	46,XX,t(7;15)(q21;q15) 46,XX,t(9;14)(q21;q11.2)	der(7): 96377471 der(15): 45376277 der(9): **undetermined** der(14): **undetermined**	18.49	PCR-free	1.8	der(7): positive der(15): positive der(9): positive der(14): positive	der(7):18 der(15):18 der(9):31 der(14):31	- -	der(7): **0** der(15): 4 chr(9): 3 der(14): 5	Saitsu et al. 2009 [14]
5	46,X,t(X;2)(q22;p13)	der(X): 107946031 der(2): 66226306	20.35	PCR-free	2.1	der(X): **negative** der(2): **negative**	der(X):**0** der(2):**0**	-	der(X): **0** der(2): **0**	Nishimura-Tadaki et al. (2011) [12]
6	46,X,t(X;4)(q22.1;q12)	der(X): 101594056 der(4): 57969312	11.69	PCR	45.8	der(X): positive der(4): positive	der(X):20 der(4):20	-	der(X) :**0** der(4): **0**	Nishimura-Tadaki et al. (2011) [12]
7	46,X,t(X;14)(q24;q32.1)	der(X): 117318221 der(14):91614197	12.93	PCR	21.3	der(X): positive der(14): positive	der(X):24 der(14):24	-	der(X) 2 der(14): **0**	Nishimura-Tadaki et al. (2011) [12]
8	46,XX,t(5;8)(q35;q24.1)	**undetermined**	5.95	PCR-free	1.2	der(5): positive der(8): positive	der(5):14 der(8):14	-	der(5): 2 der(8): 2	Imaizumi et al. (2002) [6] Kurotaki et al. (2002) [13]
9	Inv(5)(p15q13)	**undetermined**	14.37	PCR-free	1.8	5p15:positive 5q13:positive	-	chr5:28 (inversion)	5p15: 3 5q13:6	This study

Abbreviation: IGV, Integrative Genomics Viewer.

Bold numbers and text indicate the detection failure.

Table 2 Detection of chromosomal breakpoints by BreakDancer (Deletion)

Patient	Breakpoint detection by other methods (hg19)	BreakDancer detection	IGV	
			Discordant read pair counts	Soft-clipped read counts
3	chrX:107,436,210-107,440,401 chr4: 12,237,261-12,244,289	chrX: **negative** chr4: **negative**	chrX:**0** chr4:**0**	chrX: **0** chr4: **0**
4	chr7:96,949,067-97,755,363 chr15: **undetermined**	chr7: positive chr15: **negative**	chr7:9 chr15:**0**	chr7: 2 chr15: **0**

Abbreviation: IGV, Integrative Genomics Viewer.

Supplementary Information

Precise detection of chromosomal translocation or inversion breakpoints by whole genome sequencing

Toshifumi Suzuki・Yoshinori Tsurusaki・Mitsuko Nakashima・Noriko Miyake・Hirotomo Saitsu・Satoru Takeda・Naomichi Matsumoto

Correspondence to:

Naomichi Matsumoto

Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan

Tel.: +81-45-787-2606

Fax: +81-45-786-5219

E-mail: naomat@yokohama-cu.ac.jp (N. Matsumoto)

Supplementary Information

- **Supplementary Tables 1–4**
- **Supplementary Figures 1-2**

Supplementary Table 1 PCR duplication in whole genome sequencing

Patient	Mean coverage (x)		PCR duplication (%)	
	TruSeq DNA Sample Prep Kit	TruSeq DNA PCR-Free Sample Prep Kit	TruSeq DNA Sample Prep Kit	TruSeq DNA PCR-Free Sample Prep Kit
Patient 1	6.35	6.82	56.7	1.4
Patient 3	10.32	10.13	10.7	1.6
Patient 4	5.94	7.68	50.5	1.5
Patient 5	3.25	9.55	71.8	1.7
Patient 9	1.77	5.06	80.7	0.5

Supplementary Table 2 Numbers of inter-chromosomal translocation or inversion reads

Patient	1	2	3	4	5	6	7	8	9
Numbers of CTX or INV reads	126	204	133	134	137	4,698	1,080	61	389
Numbers of reads involving targeted chromosome(s)	chr1/chr9: 1	chr5/chr15: 1	chrX/chr4: 2	chr7/15: 8 chr9/14: 1	chrX/chr2: 1	chrX/chr4: 39	chrX/chr14: 3	chr5/chr8: 1	chr5: 24
Numbers of genuine read covering SV breakpoints	**0**	chr5/chr15: 1	chrX/chr4: 2	chr7/15: 1 chr9/14: 1	**0**	chrX/chr4: 1	chrX/chr14: 1	chr5/chr8: 1	chr5:1

CTX : inter-chromosomal translocation, INV : inversion

Supplementary Table 3-1 List of soft-clipped reads covering SV breakpoints

Supplementary Table 3-1 List of soft-clipped reads covering SV breakpoints

patient	soft-clipped reads*	derivative chromosome	read_start (bps)	read_end (bps)	start position	end position	unknown sequence
1	1	der(1)	1	88	chr1: 206574425	chr1: 206574512	-
			89	101	-	-	TATTCGATTGGAA
3	1	der(X)	1	88	chrX: 107436122	chrX: 107436209	-
			89	93	-	-	TACAG
			94	101	chr4:12237260	chr4:12237267	-
	2	der(X)	1	90	chrX: 107436120	chrX: 107436209	-
			91	95	-	-	TACAG
			96	101	chr4:12237260	chr4:12237265	-
	3	der(X)	1	54	chr4:12237207	chr4:12237260	
			55	59	-	-	CTGTA (-strand)
			60	101	chrX:107436209	chrX:107436168	
	4	der(X)	1	85	chr4:12237176	chr4:12237260	
			86	90			CTGTA (-strand)
			91	101	chrX:107436209	chrX:107436219	
	5	der(X)	1	66	chr4:12237195	chr4:12237260	
			67	71	-	-	CTGTA (-strand)
			72	101	chrX:107436209	chrX:107436180	
	1	der(4)	1	12	chrX:107440391	chrX:107440402	-
			13	15	-	-	GAG
			16	101	chr4: 12244290	chr4: 12244375	-
	2	der(4)	1	13	chrX:107440390	chrX:107440402	-
			14	16	-	-	GAG
			17	101	chr4: 12244290	chr4: 12244374	-
4	1	der(15)	1	54	chr15: 45376224	chr15: 45376277	-
			53	101	chr7: 96377469	chr7: 96377517	-
	2	der(15)	1	90	chr15: 45376188	chr15: 45376277	-
			89	101	chr7: 96377469	chr7: 96377481	-
	3	der(15)	1	76	chr15: 45376202	chr15: 45376277	-
			75	101	chr7: 96377469	chr7: 96377495	-
	4	der(15)	1	64	chr15: 45376214	chr15: 45376277	-
			63	101	chr7: 96377469	chr7: 96377507	-
4	1	der(9)	1	97	chr9: 80735757	chr9: 80735853	-
			98	101	chr14:23270388	chr14:23270391	-
	2	der(9)	1	90	chr9:80735764	chr9:80735853	-
			91	101	chr14:23270388	chr14:23270398	-
	3	der(9)	1	68	chr9:80735786	chr9:80735853	-
			69	101	chr14:23270388	chr14:23270420	-
	1	der(14)	1	84	chr14: 23270320	chr14: 23270383	-
			84	101	chr9:80735853	chr9:80735870	-
	2	der(14)	1	86	chr14: 23270318	chr14: 23270383	-
			86	101	chr9:80735853	chr9:80735868	-
	3	der(14)	1	88	chr14: 23270316	chr14: 23270383	-
			88	101	chr9: 80735853	chr9:80735866	-
	4	der(14)	1	11	-	-	TATTTGTTATA
			12	63	chr14: 23270332	chr14: 23270383	-
			63	101	chr9: 80735853	chr9: 80735891	-
	5	der(14)	1	89	chr14: 23270315	chr14: 23270383	-
			89	101	chr9:80735853	chr9:8735865	-
4	1(DEL)	chr7	1	54	chr7: 96949012	chr7: 96949065	-
			55	101	chr7: 97755361	chr7: 97755407	-
	2(DEL)	chr7	1	53	chr7:96949013	chr7: 96949065	-
			54	101	chr7: 97755361	chr7: 97755408	-

***: each read is numbered. Red characters indicate exact breakpoint positions and sequences.**

Supplementary Table 3-2 List of soft-clipped reads covering SV breakpoints

Supplementary Table 3-2 List of soft-clipped reads covering SV breakpoints

patient	soft-clipped reads*	derivative chromosome	read_start (bps)	read_end (bps)	start position	end position	unknown sequence
7	1	der(X)	1	87	chrX: 117318137	chrX: 117318223	-
			86	101	chr14: 91614203	chr14: 91614218	-
	2	der(X)	1	79	chrX: 117318145	chrX: 117318223	-
			78	101	chr14: 91614203	chr14: 91614226	-
8	1	der(5)	1	84	chr5: 176562234	chr5: 176562317	-
			85	93	-	-	TCACATTGG
			94	101	chr8: 117943670	chr8: 117943677	-
	2	der(5)	1	73	chr5: 176562245	chr5: 176562317	-
			74	83	-	-	TCACATTGG
			84	101	chr8: 117943670	chr8: 117943687	-
	1	der(8)	1	68	chr8: 117943605	chr8: 117943672	-
			69	76	-	-	AACAAGAT
			77	101	chr5: 176562322	chr5: 176562346	-
	2	der(8)	1	72	chr8: 117943601	chr8: 117943672	-
			73	80	-	-	AACAAGAT
			81	101	chr5: 176562322	chr5: 176562342	-
9	1	5parm	1	76	chr5: 8008547	chr5: 8008622	-
			76	101	chr5: 68258651	chr5: 68258626	-
	2	5parm	1	74	chr5: 8008549	chr5: 8008622	-
			74	101	chr5: 68258651	chr5: 68258624	-
	3	5parm	1	38	chr5: 68258688	chr5: 68258651	-
			39	101	chr5: 8008622	chr5: 8008684	-
	1	5qarm	1	16	chr5:8008743	chr5:8008758	-
			17	101	chr5: 68258651	chr5: 68258735	-
	2	5qarm	1	62	chr5: 68258590	chr5: 68258651	-
			62	101	chr5: 8008622	chr5: 8008583	-
	3	5qarm	1	-	-	-	C
			2	22	chr5: 8008642	chr5: 8008622	-
			23	101	chr5: 68258651	chr5: 68258729	-
	4	5qarm	1	71	chr5: 68258581	chr5: 68258651	-
			71	101	chr5: 8008622	chr5: 8008593	-
	5	5qarm	1	62	chr5: 68258590	chr5: 68258651	-
			62	101	chr5: 8008622	chr5: 8008583	-
	6	5qarm	1	22	chr5: 8008643	chr5: 8008622	-
			23	101	chr5: 68258651	chr5: 68258729	-

***: each read is numbered. Red characters indicate exact breakpoint positions and sequences.**

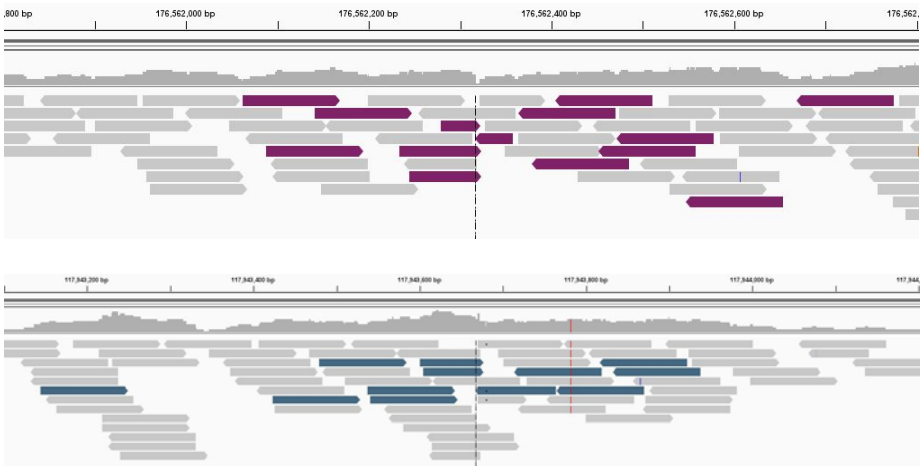
Supplementary Table 4 Numbers of deletion reads by BreakDancer.

Supplementary Table 4 Numbers of deletion reads by BreakDancer.

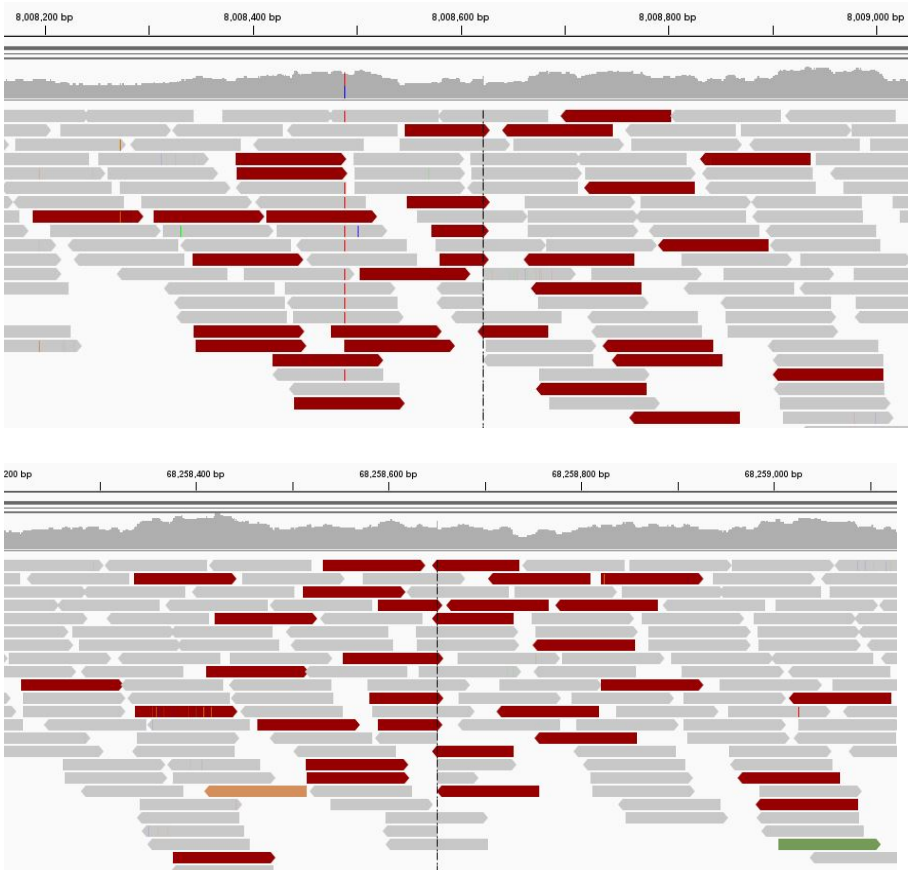
Patient	3		4	
Chromosome	**4**	**X**	**7**	**15**
Deletion position	chr4:12,237,261-12,244,289	chrX:107,436,210-107,440,401	chr7:96,949,067-97,755,363	Undetermined
Deletion size (bp)	7,029	4,192	806,297	Undetermined
Number of called reads for DEL	1,945	1,945	1,943	1,943
Targeted chromosome	159	51	128	58
Number of accurate DEL reads	0	0	1	0

DEL : deletion

A

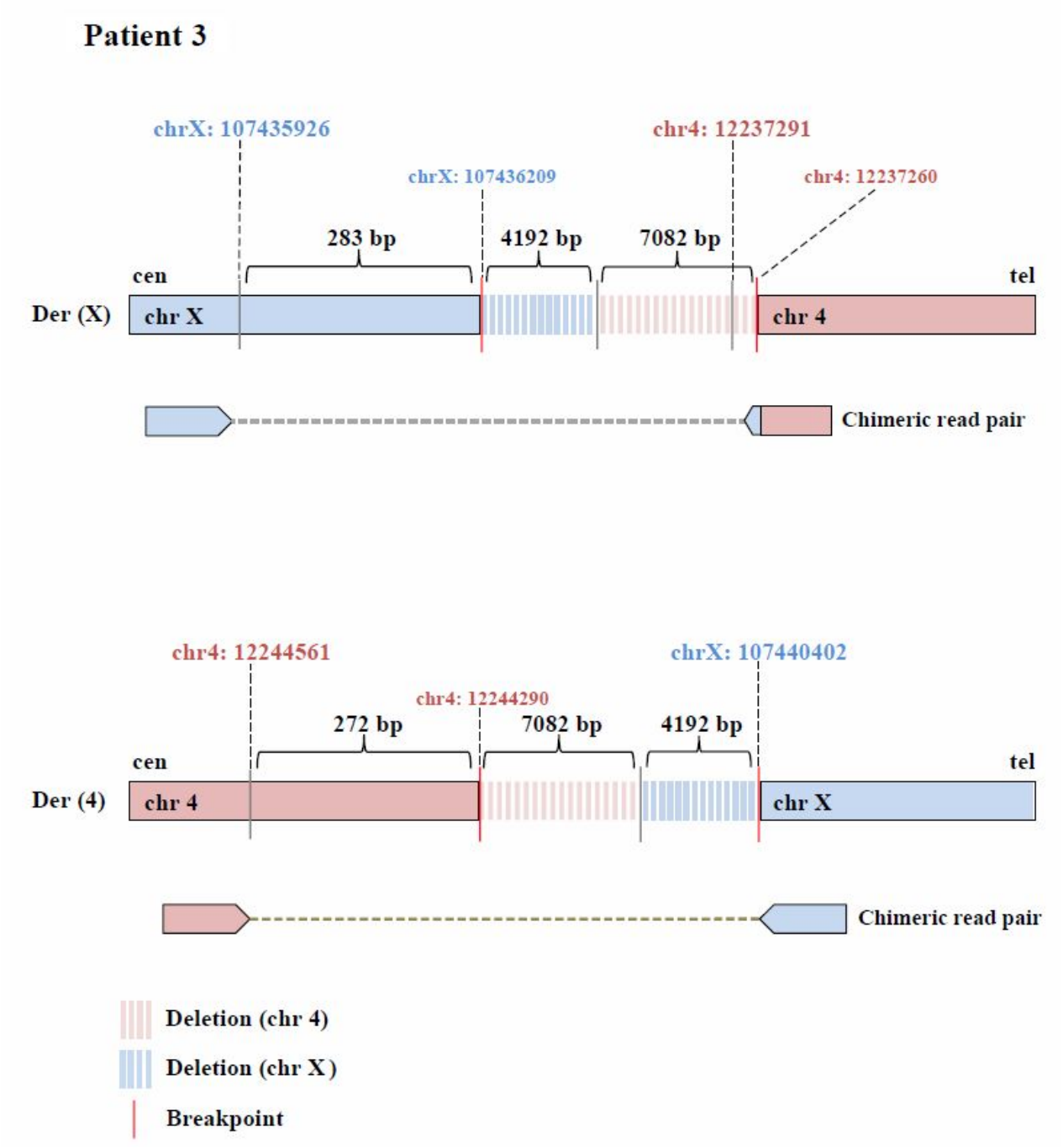

176,562,000 bp
176,562,200 bp
176,562,400 bp
176,562,600 bp

B


8,008,200 bp
8,008,400 bp
8,008,600 bp
8,008,800 bp
8,009,000 bp
68,258,400 bp
68,258,600 bp
68,258,800 bp
68,259,000 bp

Supplementary Figure 1. IGV images at around breakpoints.

A. Patient 8. Purple and dark green directed lines were chimeric read pairs covering translocation breakpoints. Chromosomes 5 (upper) and 8 (lower) are shown. Black vertical dot lines indicate translocation breakpoints. B. Patient 9. Maroon directed lines were discordant read pairs spanning inversion breakpoint. Upper panel shows 5p inversion breakpoint and lower indicate 5q inversion breakpoint. Black vertical dot lines indicate breakpoint locations. Orange and green directed lines in the lowest panel were chimeric read pairs whose other-end reads were mapped to chromosomes 14 (orange) and 3 (green).


Patient 3
chrX: 107435926
chr4: 12237291
chrX: 107436209
chr4: 12237260
283 bp
4192 bp
7082 bp
cen
tel
Der (X)
chr X
chr 4
Chimeric read pair
chr4: 12244561
chrX: 107440402
chr4: 12244290
272 bp
7082 bp
4192 bp
cen
tel
Der (4)
chr 4
chr X
Chimeric read pair
Deletion (chr 4)
Deletion (chr X)
Breakpoint

Supplementary Figure 2. Schematic presentation of the chromosomal translocation and deletion in Patient 3.

Blue and red boxes indicate chromosomes X and 4. The upper panel indicates derivative X and the lower shows derivative 4 chromosome. Red and blue horizontal thick lines represent deletions of chromosome 4 and X. Red vertical lines indicate chromosomal breakpoints. Two chimeric read pairs were called by BreakDancerMax.

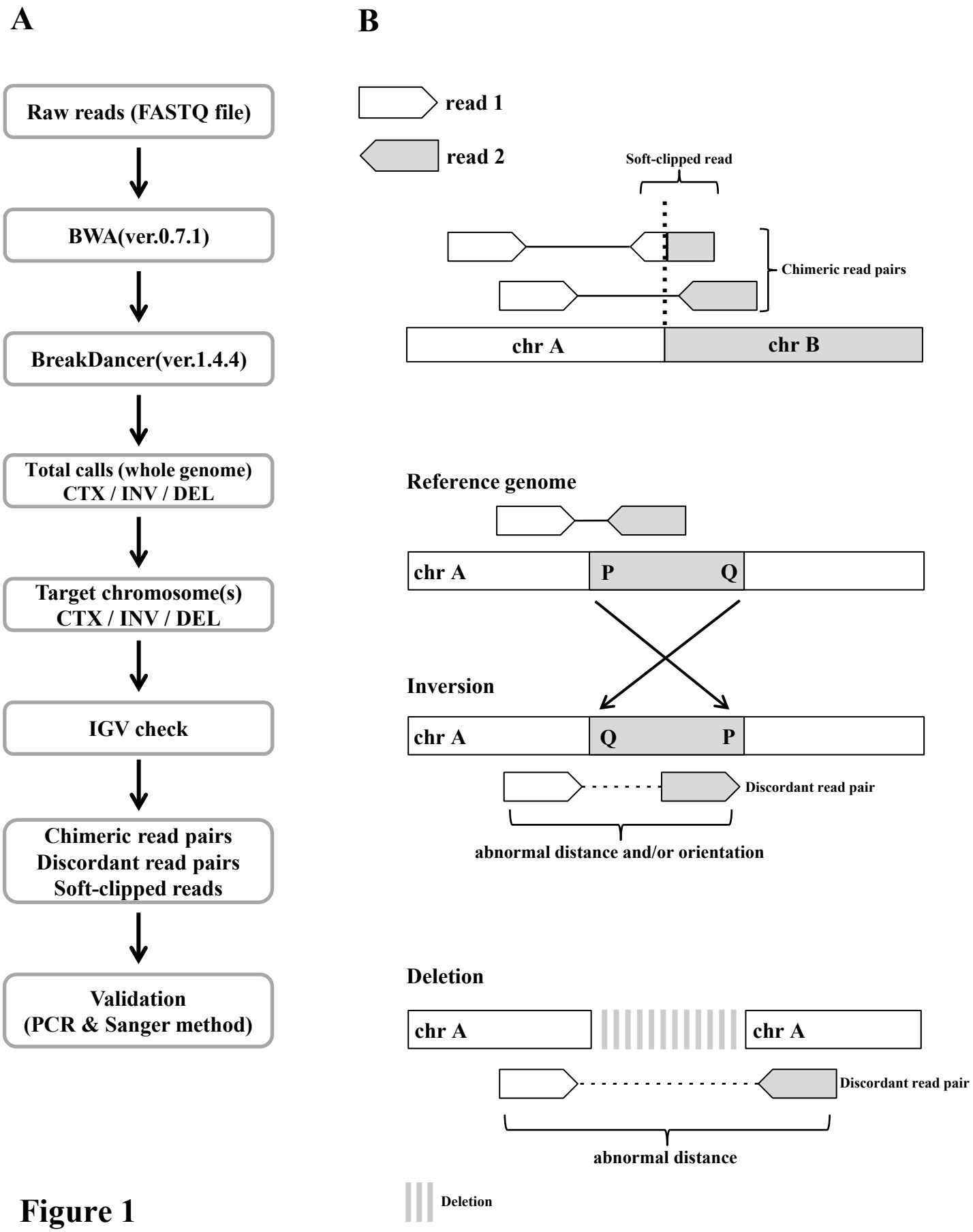

A
Raw reads (FASTQ file)
BWA(ver.0.7.1)
BreakDancer(ver.1.4.4)
Total calls (whole genome)
CTX / INV / DEL
Target chromosome(s)
CTX / INV / DEL
IGV check
Chimeric read pairs
Discordant read pairs
Soft-clipped reads
Validation
(PCR & Sanger method)
B
read 1
read 2
Soft-clipped read
Chimeric read pairs
chr A
chr B
Reference genome
chr A
P
Q
Inversion
chr A
Q
P
Discordant read pair
abnormal distance and/or orientation
Deletion
chr A
chr A
Discordant read pair
abnormal distance
Deletion

Figure 1

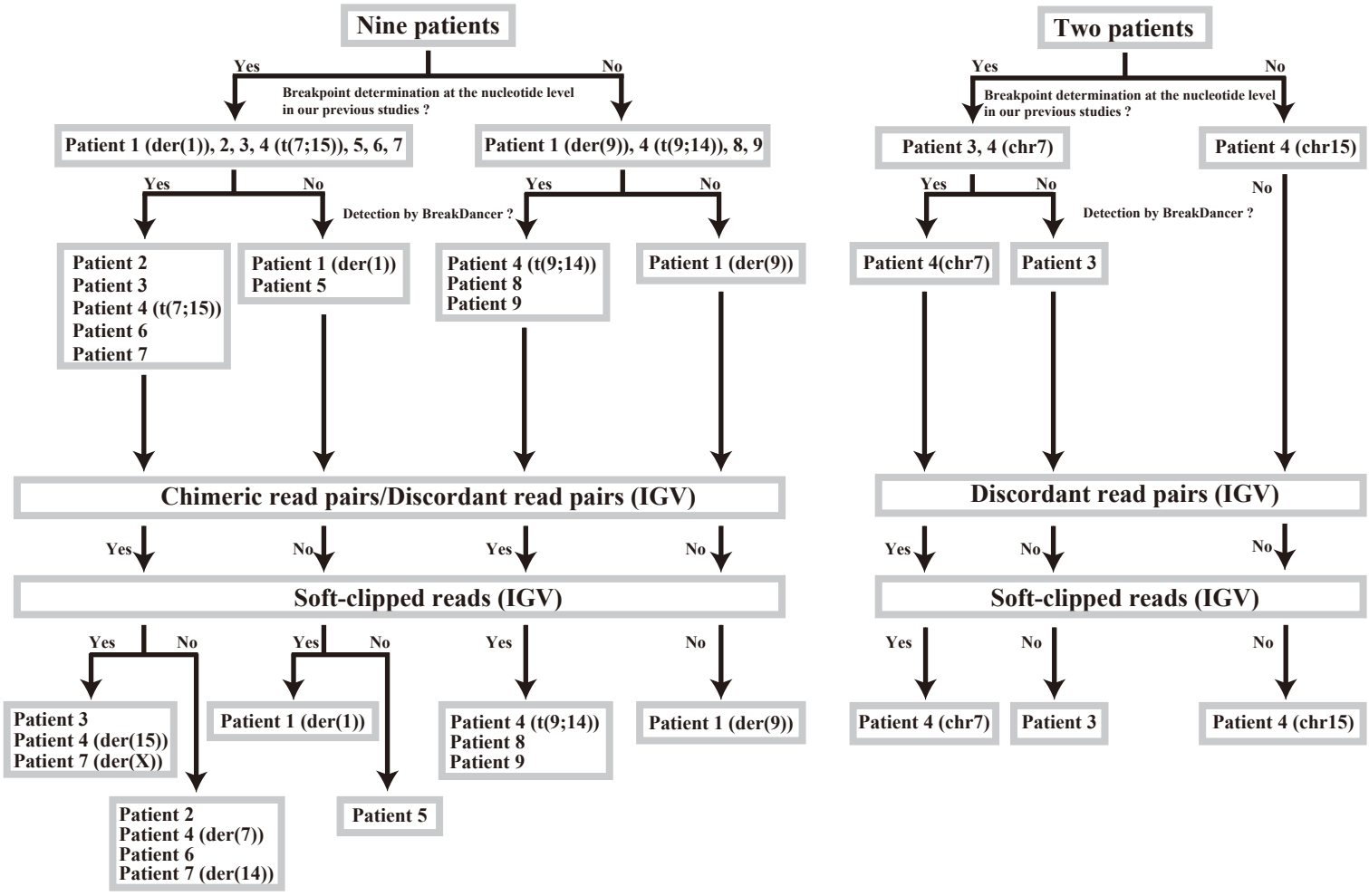

Translocation or inversion breakpoints
Deletion breakpoints
Nine patients
Yes
No
Breakpoint determination at the nucleotide level in our previous studies ?
Patient 1 (der(1)), 2, 3, 4 (t(7;15)), 5, 6, 7
Patient 1 (der(9)), 4 (t(9;14)), 8, 9
Yes
No
Yes
No
Detection by BreakDancer ?
Patient 2
Patient 3
Patient 4 (t(7;15))
Patient 6
Patient 7
Patient 1 (der(1))
Patient 5
Patient 4 (t(9;14))
Patient 8
Patient 9
Patient 1 (der(9))
Chimeric read pairs/Discordant read pairs (IGV)
Yes
No
Yes
No
Soft-clipped reads (IGV)
Yes
No
Yes
No
Yes
No
Patient 3
Patient 4 (der(15))
Patient 7 (der(X))
Patient 1 (der(1))
Patient 4 (t(9;14))
Patient 8
Patient 9
Patient 1 (der(9))
Patient 2
Patient 4 (der(7))
Patient 6
Patient 7 (der(14))
Patient 5
Two patients
Yes
No
Breakpoint determination at the nucleotide level in our previous studies ?
Patient 3, 4 (chr7)
Patient 4 (chr15)
Yes
No
No
Detection by BreakDancer ?
Patient 4(chr7)
Patient 3
Discordant read pairs (IGV)
Yes
No
No
Soft-clipped reads (IGV)
Yes
No
No
Patient 4 (chr7)
Patient 3
Patient 4 (chr15)

Figure 2

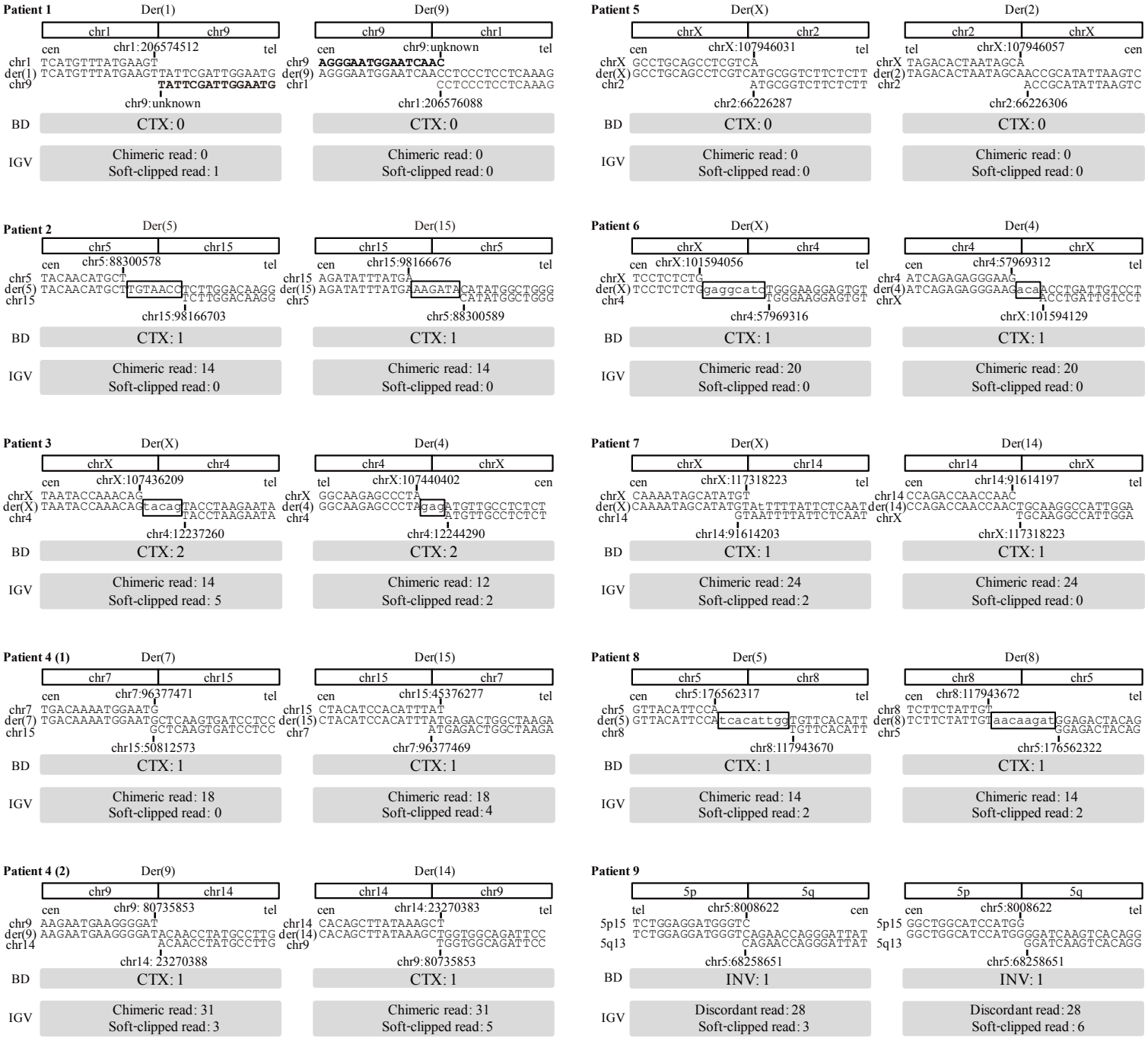

Patient 1
Der(1)
Der(9)
chr1
chr9
cen
chr1:206574512
tel
chr9:unknown
chr1 TCATGTTTATGAAGT
der(1) TCATGTTTATGAAGTTATTCGATTGGAATG
chr9 TATTCGATTGGAATG
chr9 AGGGAATGGAATCAAC
der(9) AGGGAATGGAATCAACCTCCCTCCTCAAAG
chr1 CCTCCCTCCTCAAAG
chr9:unknown
chr1:206576088
BD
CTX: 0
CTX: 0
IGV
Chimeric read: 0
Soft-clipped read: 1
Chimeric read: 0
Soft-clipped read: 0
Patient 2
Der(5)
Der(15)
chr5
chr15
chr5:88300578
chr15:98166676
chr5 TACAACATGCT
der(5) TACAACATGCTTGTAACCTCTTGGACAAGG
chr15 TCTTGGACAAGG
chr15 AGATATTTATGA
der(15) AGATATTTATGAAAGATACATATGGCTGGG
chr5 CATATGGCTGGG
chr15:98166703
chr5:88300589
BD
CTX: 1
CTX: 1
IGV
Chimeric read: 14
Soft-clipped read: 0
Chimeric read: 14
Soft-clipped read: 0
Patient 3
Der(X)
Der(4)
chrX
chr4
chrX:107436209
chrX:107440402
chrX TAATACCAAACAG
der(X) TAATACCAAACAGcacagTACCTAAGAATA
chr4 TACCTAAGAATA
chrX GGCAAGAGCCCTA
der(4) GGCAAGAGCCCTAgagATGTTGCCTCTCT
chr4 ATGTTGCCTCTCT
chr4:12237260
chr4:12244290
BD
CTX: 2
CTX: 2
IGV
Chimeric read: 14
Soft-clipped read: 5
Chimeric read: 12
Soft-clipped read: 2
Patient 4 (1)
Der(7)
Der(15)
chr7
chr15
chr7:96377471
chr15:45376277
chr7 TGACAAAATGGAATG
der(7) TGACAAAATGGAATGCTCAAGTGATCCTCC
chr15 GCTCAAGTGATCCTCC
chr15 CTACATCCACATTTAT
der(15) CTACATCCACATTTATGAGACTGGCTAAGA
chr7 ATGAGACTGGCTAAGA
chr15:50812573
chr7:96377469
BD
CTX: 1
CTX: 1
IGV
Chimeric read: 18
Soft-clipped read: 0
Chimeric read: 18
Soft-clipped read: 4
Patient 4 (2)
Der(9)
Der(14)
chr9
chr14
chr9: 80735853
chr14:23270383
chr9 AAGAATGAAGGGGAT
der(9) AAGAATGAAGGGGATACAACCTATGCCTTG
chr14 ACAACCTATGCCTTG
chr14 CACAGCTTATAAAGCT
der(14) CACAGCTTATAAAGCTGGTGGCAGATTCC
chr9 TGGTGGCAGATTCC
chr14: 23270388
chr9:80735853
BD
CTX: 1
CTX: 1
IGV
Chimeric read: 31
Soft-clipped read: 3
Chimeric read: 31
Soft-clipped read: 5
Patient 5
Der(X)
Der(2)
chrX
chr2
chrX:107946031
chrX:107946057
chrX GCCTGCAGCCTCGTCA
der(X) GCCTGCAGCCTCGTCATGCGGTCTTCTCTT
chr2 ATGCGGTCTTCTCTT
chrX TAGACACTAATAGCA
der(2) TAGACACTAATAGCAACCGCATATTAAGTC
chr2 ACCGCATATTAAGTC
chr2:66226287
chr2:66226306
BD
CTX: 0
CTX: 0
IGV
Chimeric read: 0
Soft-clipped read: 0
Chimeric read: 0
Soft-clipped read: 0
Patient 6
Der(X)
Der(4)
chrX
chr4
chrX:101594056
chr4:57969312
chrX TCCTCTCTG
der(X) TCCTCTCTGgaggcatcTGGGAAGGAGTGT
chr4 TGGGAAGGAGTGT
chr4 ATCAGAGAGGGAAG
der(4) ATCAGAGAGGGAAGacaACCTGATTGTCCT
chrX ACCTGATTGTCCT
chr4:57969316
chrX:101594129
BD
CTX: 1
CTX: 1
IGV
Chimeric read: 20
Soft-clipped read: 0
Chimeric read: 20
Soft-clipped read: 0
Patient 7
Der(X)
Der(14)
chrX
chr14
chrX:117318223
chr14:91614197
chrX CAAAATAGCATATGT
der(X) CAAAATAGCATATGTAtTTTTATTCTCAAT
chr14 GTAATTTTATTCTCAAT
chr14 CCAGACCAACCAAC
der(14) CCAGACCAACCAACTGCAAGGCCATTGGA
chrX TGCAAGGCCATTGGA
chr14:91614203
chrX:117318223
BD
CTX: 1
CTX: 1
IGV
Chimeric read: 24
Soft-clipped read: 2
Chimeric read: 24
Soft-clipped read: 0
Patient 8
Der(5)
Der(8)
chr5
chr8
chr5:176562317
chr8:117943672
chr5 GTTACATTCCA
der(5) GTTACATTCCAtcacattggTGTTCACATT
chr8 TGTTCACATT
chr8 TCTTCTATTGT
der(8) TCTTCTATTGTaacaagatGGAGACTACAG
chr5 GGAGACTACAG
chr8:117943670
chr5:176562322
BD
CTX: 1
CTX: 1
IGV
Chimeric read: 14
Soft-clipped read: 2
Chimeric read: 14
Soft-clipped read: 2
Patient 9
5p
5q
chr5:8008622
chr5:8008622
5p15 TCTGGAGGATGGGTC
5q13 TCTGGAGGATGGGTCAGAACCAGGGATTAT
CAGAACCAGGGATTAT
5p15 GGCTGGCATCCATGG
5q13 GGCTGGCATCCATGGGGATCAAGTCACAGG
GGATCAAGTCACAGG
chr5:68258651
chr5:68258651
BD
INV: 1
INV: 1
IGV
Discordant read: 28
Soft-clipped read: 3
Discordant read: 28
Soft-clipped read: 6

Figure 3

Table 1 Detection of chromosomal breakpoints by BreakDancer (Translocation or inversion)

Patient	Karyotype	Breakpoint detection by other methods (hg19)	Mean coverage	preparation	PCR duplication (%)	BreakDancer detection	IGV			References
							Chimeric read pair counts	Discordant read pair counts	Soft-clipped read counts	
1	46,XX,t(1;9)(q32;q13)	der(1): 206574512 der(9): **undetermined**	15.23	PCR-free	1.6	der(1): **negative** der(9): **negative**	der(1):**0** der(9):**0**	-	der(1): 1 der(9): **0**	Saitsu et al. (2012) [10]
2	46,XX,t(5;15)(q13.3;q26.1)	der(5): 88300578 der(15): 98166676	9.18	PCR	7.0	der(5): positive der(15): positive	der(5):14 der(15):14	-	der(5): **0** der(15): **0**	Saitsu et al. (2011) [11]
3	46,X,t(X;4)(q21.3;p15.2)	der(X): 107436209 der(4): 12244290	21.92	PCR-free	1.9	der(X): positive der(4): positive	der(X):14 der(4):12	-	der(X): 5 der(4): 2	Nishimura-Tadaki et al. (2011) [12]
4	46,XX,t(7;15)(q21;q15) 46,XX,t(9;14)(q21;q11.2)	der(7): 96377471 der(15): 45376277 der(9): **undetermined** der(14): **undetermined**	18.49	PCR-free	1.8	der(7): positive der(15): positive der(9): positive der(14): positive	der(7):18 der(15):18 der(9):31 der(14):31	- -	der(7): **0** der(15): 4 chr(9): 3 der(14): 5	Saitsu et al. 2009 [14]
5	46,X,t(X;2)(q22;p13)	der(X): 107946031 der(2): 66226306	20.35	PCR-free	2.1	der(X): **negative** der(2): **negative**	der(X):**0** der(2):**0**	-	der(X): **0** der(2): **0**	Nishimura-Tadaki et al. (2011) [12]
6	46,X,t(X;4)(q22.1;q12)	der(X): 101594056 der(4): 57969312	11.69	PCR	45.8	der(X): positive der(4): positive	der(X):20 der(4):20	-	der(X) :**0** der(4): **0**	Nishimura-Tadaki et al. (2011) [12]
7	46,X,t(X;14)(q24;q32.1)	der(X): 117318221 der(14):91614197	12.93	PCR	21.3	der(X): positive der(14): positive	der(X):24 der(14):24	-	der(X) 2 der(14): **0**	Nishimura-Tadaki et al. (2011) [12]
8	46,XX,t(5;8)(q35;q24.1)	**undetermined**	5.95	PCR-free	1.2	der(5): positive der(8): positive	der(5):14 der(8):14	-	der(5): 2 der(8): 2	Imaizumi et al. (2002) [6] Kurotaki et al. (2002) [13]
9	Inv(5)(p15q13)	**undetermined**	14.37	PCR-free	1.8	5p15:positive 5q13:positive	-	chr5:28 (inversion)	5p15: 3 5q13:6	This study

Abbreviation: IGV, Integrative Genomics Viewer.

Bold numbers and text indicate the detection failure.

Table 2 Detection of chromosomal breakpoints by BreakDancer (Deletion)

Patient	Breakpoint detection by other methods (hg19)	BreakDancer detection	IGV	
			Discordant read pair counts	Soft-clipped read counts
3	chrX:107,436,210-107,440,401	chrX: **negative**	chrX:**0**	chrX: **0**
	chr4: 12,237,261-12,244,289	chr4: **negative**	chr4:**0**	chr4: **0**
4	chr7:96,949,067-97,755,363	chr7: positive	chr7:9	chr7: 2
	chr15: **undetermined**	chr15: **negative**	chr15:**0**	chr15: **0**

Abbreviation: IGV, Integrative Genomics Viewer.