Published in final edited form as:

Curr Protoc Hum Genet. ; 96: 8.18.1–8.18.16. doi:10.1002/cphg.51.

Balanced chromosomal rearrangements detection by low-pass whole-genome sequencing

Zirui Dong^{1,2,3,*}, Lingfei Ye^{3,4,*}, Zhenjun Yang^{3,4,*}, Haixiao Chen^{3,4}, Jianying Yuan^{3,4}, Huilin Wang^{1,2,5}, Xiaosen Guo^{3,4}, Yun Li^{3,4}, Jun Wang^{3,4}, Fang Chen^{3,4}, Sau Wai Cheung^{6,7}, Cynthia C. Morton^{8,9,10,11,12}, Hui Jiang^{3,4,#}, and Kwong Wai Choy^{1,2,6,#}

¹Department of Obstetrics and Gynaecology, The Chinese University of Hong Kong, Hong Kong, China

²Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China

³BGI-Shenzhen, Shenzhen 518083, China

⁴China National Genebank-Shenzhen, BGI-Shenzhen, Shenzhen 518083, China

⁵Department of Central Laboratory, Bao'an Maternity and Child Healthcare Hospital Affiliated to Jinan University School of Medicine, Key Laboratory of Birth Defects Research, Birth Defects Prevention Research and Transformation Team, Shenzhen, China

⁶The Chinese University of Hong Kong-Baylor College of Medicine Joint Center for Medical Genetics, Hong Kong, China

⁷Department of Molecular and Human Genetics, Baylor College of Medicine Houston, TX 77030, USA

⁸Department of Obstetrics and Gynecology, Brigham and Women's Hospital, Boston, MA 02115, USA

⁹Harvard Medical School, Boston, MA 02115, USA

¹⁰Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Boston, MA 02142, USA

¹¹Department of Pathology, Brigham and Women's Hospital, Boston, MA 02115, USA

¹²Division of Evolution and Genomic Sciences, School of Biological Sciences, University of Manchester, Manchester Academic Health Science Center, Manchester M13 9PL, UK

Abstract

Balanced chromosomal rearrangement (or balanced chromosome abnormality, BCA) is one of the common chromosomal structural variants. Emerging studies have demonstrated the feasibility of using whole-genome sequencing (WGS) in detection of BCA-associated breakpoints but the a priori requirement of knowing the rearranged regions from prior G-banded chromosome analysis limits its application. Here, we describe a protocol based on low-pass WGS for detecting BCA

#Contact information: Kwong Wai CHOY, richardchoy@cuhk.edu.hk; Hui JIANG, jianghui@genomics.cn.

*These authors contributed equally

events independent from chromosome analysis, which has been validated in genomic data from the 1000 Genomes Project. A non-size-selected mate-pair library (3~8 kb) with 2~3 µg DNA as input is adopted and only 30 million read-pairs (50bp), which is equivalent to 1-fold base-coverage is required for each sample. The complete procedure of this testing takes 13 days and the total cost was estimated to be less than 600 USD per sample.

Keywords

Low-pass whole-genome sequencing; chromosomal structural rearrangements; balanced translocations; inversions

Introduction

Balanced chromosomal rearrangements (or abnormalities, BCAs), including translocations, insertions, and inversions are currently only detectable by G-banded chromosome analysis as no cytogenetically apparent gain or loss of chromatin is involved. The incidence of BCAs has been estimated to be approximately 0.2-0.5% in the general population (Dong et al., 2016; Jonas et al., 2014; Kearney et al., 2011; Tang & Amon, 2013) and phenotypically normal carriers themselves are at risk for reproductive problems such as recurrent miscarriage. Recent studies show that BCAs also cause clinical anomalies due to gene disruption, cryptic imbalances involving the breakpoints or dysregulation from a positional effect (Jonas et al., 2014; Kearney et al., 2011; Redin et al., 2017; Tang & Amon, 2013).

Utilizing paired-end whole-genome sequencing (WGS) data, the breakpoints of BCAs can be localized at base-pair resolution, providing precision for clinical interpretation. However, testing and validation was largely restricted to DNA samples harboring previously recognized BCAs. One of the difficulties in detecting BCAs without knowing the involved bands from prior cytogenetic methods is false alignment of the paired-end read-pairs. Our pilot study provided an approach with a dataset of false positives for detecting BCAs in a blinded fashion (Dong et al., 2016; Jonas et al., 2014; Kearney et al., 2011; Tang & Amon, 2013), and by applying this analytical tool to the existing WGS data (with >8.25-fold physical coverage) released by the 1000 Genomes Project without cytogenetic results available (M.Oliver-Bonet et al., 2002; Mackie Ogilvie & Scriven, 2002; Redin et al., 2017; Van Dyke et al., 1983), we identified eight BCAs including four balanced translocations and four inversions (De Gregori et al., 2007; Redin et al., 2017; Talkowski, Rosenfeld, et al., 2012; Warburton, 1991). Particularly, BCAs such as the identified t(16;17)(q23.1;q24.2) would not be reliably detected by G-banded chromosome analysis due to translocated segments with similar banding patterns and the fragment sizes despite both sizes (i.e., >5 Mb) considered to be potentially visible at the microscopic level (Dong et al., 2014).

In order to reduce the read-pairs amount required and the quantity of input DNA, as compared to the size-selected mate-pair library construction, in this study, we describe an approach (Figure 1) with a non-size-selected mate-pair library (3~8 kb) with 2~3 µg DNA as input for BCA detection with low-pass WGS (Figure 1). The minimum base coverage requirement for each sample by using this protocol is as low as 1-fold genome-wide.

Basic Protocol 1: DNA Fragmentation and Library Construction

In this protocol, we aimed to create at least four DNA mate-pair libraries for paired-end sequencing on a HiSeq 2000™ platform (Illumina Inc., San Diego, CA) in one lane (eight lanes per slide). Mate-pair library construction is usually used in *de novo* assembly of a genome (i.e., mammal) (Sudmant et al., 2015), and a narrow size of DNA templates is required for construction of contigs and the scaffold. Therefore, a size selection process via gel electrophoresis is commonly carried out before DNA circularization. However, 20 µg of DNA is recommended for each sample as input (Talkowski et al., 2011), although subsequently reduced to 5~10 µg (Hanscom & Talkowski, 2014), which may limit application in the clinical setting. For identification of BCAs, a larger range of insert size is acceptable as we can only select read-pairs aligned with distances larger than the largest insert size for event clustering. In addition, larger fragment size enables identifying uniquely aligned reads in both ends, particularly when the BCAs are likely mediated by repetitive elements (Carvalho & Lupski, 2016). By using a 2-kb mate-pair library (Talkowski, Ordulu, et al., 2012), there is still an approximately 9% failure in identification of BCA breakpoints (Redin et al., 2017).

In our study, we chose non-size selection mate-pair library with 2~3 µg DNA as input. In general, human genomic DNA is sheared to a range of 3 to 8 kb, after end repair, DNA fragments are biotin-labeled and subsequently blunt-endly circularized. Circular DNA is recovered after digestion of linear DNAs and further fragmented to a small size (i.e., 400~600 bp). After enrichment of biotin-labeled DNAs, barcoded adapter ligation is performed following repair of the ends of double strand DNAs and addition of adenine to the 3' ends. The barcoded adapters are listed in our previous protocol (Dong, Xie, et al., 2017; see UNIT 8.17).

Materials

- DNA sample
- 1× TE (pH8.0) (Ambion, Thermo Fisher Scientific, Waltham, MA)
- 0.2M HCl (Ambion)
- 0.2M NaOH (Ambion)
- dNTP Mix (Invitrogen, Thermo Fisher Scientific, Waltham, MA)
- Biotin-dNTP Mix (Enzymatics, Qiagen Beverly, Beverly, MA)
- T4 DNA Polymerase (Enzymatics)
- Klenow DNA Polymerase (Enzymatics)
- Molecular Biology Agarose (Bio-Rad Laboratories, Hercules, CA)
- T4 PNK (Enzymatics)
- 10×T4 DNA Ligase (Enzymatics)
- 10×T4 Ligase Buffer (Enzymatics)

25mM ATP (Enzymatics)
10×Blue Buffer (Enzymatics)
1 mM dATP (Enzymatics)
Klenow3'-5'(exo) (Enzymatics)
2×Rapid Buffer (Enzymatics)
PE Adapter Oligo Mix (Invitrogen)
PE index Adapter Oligo (Invitrogen)
EB buffer (Qiagen, Hilden, Germany)
Milli-Q water
Anhydrous ethanol
Plasmid-safe 10×Buffer (Epicentre Biotechnologies, Illumina, Madison, WI)
Plasmid-safe™ DNase (Epicentre Biotechnologies)
Exonuclease I (NEB)
0.5 M EDTA (Ambion)
DL2000 marker (Tiangen Biotech Co., LTD., Beijing, China)
λ-HindIII marker (TaKaRa Biotechnology (Dalian) Co., Ltd., Dalian, Liaoning, China)
1 kb DNA Ladder (Tiangen)
50 bp DNA Ladder (NEB, Ipswich, MA)
Agencourt AMPure beads (Beckman Coulter, Inc., Brea, CA. cat. no. A29152)
Streptavidin Magnetic Beads (Invitrogen)
1.5-ml microcentrifuge tubes (Axygen, Corning, Axxygen Scientific Inc, Union City, CA)
50 ml microcentrifuge tubes (BD Biosciences, San Jose, CA)
1.5 ml Non-stick RNase-Free Microfuge Tube (Ambion)
Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific)

Equipment

Hydroshear (Genemachine, San Carlos, CA)
Covaris S2 (Covaris, Woburn, MA)
NanoDrop 2000™ (Thermo Fisher Scientific)
Vortex-5 (Haimen Kylin-Bell Lab Instruments Co., Ltd, Haimen, Jiangsu, China)
Magnetic separator (Dexter Magnetic Technologies, Elk Grove Village, IL)

Electrophoresis System (Thermo Fisher Scientific)

ThermoMixer (Eppendorf, Hamburg, Germany)

Microcentrifuge (Eppendorf)

Qubit (Thermo Fisher Scientific)

Sample testing

1. Verify DNA integrity with 0.6% agarose gel electrophoresis at 120 Voltage for 40 min, using the D2000 DNA marker and λ -HindIII as size marker (Dong, Xie, et al., 2017).
2. Measure the purity of DNA by NanoDrop™ 2000™ ($O.D_{260}/O.D_{280} > 1.8$; $O.D_{260}/O.D_{230} > 1.5$).
3. Obtain an approximate concentration using the Qubit HS.

Fragment genomic DNAs

4. Dilute 3 μ g DNA to 150 μ l with TE (pH 8.0) in a 1.5 ml tube, incubate for 30 min at 37 °C, vortex every 10 min. Centrifuge at 14,000 rpm for 15 min, and transfer the DNA solution to a clean tube if a pellet is formed.

The minimum DNA input requirement is 2 μ g whereas the recommended amount is 3 μ g.

5. Use 0.2 μ m needle filters for 0.2M HCl, 0.2M NaOH, MillQ-H₂O and TE Buffer (pH 8.0), respectively and store each of these four solutions in an independent 50 ml tube (BD).

6. Run HydroShear with a large syringe adapter, using the settings given below:

Volume (μ l):	150
Number of cycles:	20
Speed Code:	5

7. Click “Edit Wash Scheme” and choose the times for different buffers as below:

0.2M HCl:	6
0.2M NaOH:	6
MillQ-H ₂ O:	3
TE (pH 8.0):	3
Discharge air:	2

8. Hit “Start” and click “OK” to perform a wash. Turn the handle to “input” and aspirate the sample. Then eject the sample to the syringe immediately after turning the handle to “output”. Wipe the “input” pipe by using 75% alcohol to prevent cross contamination.

9. Expel air and wipe the “input” pipe with 75% alcohol before starting the program.

10. Put the “output” pipe in a 1.5 ml tube to collect the samples.

11. Click on the software interface in “MANUAL OPERATION” button and change the volume to 450 μ l to expel the remained sample manually. Repeat step 7 to step 11 for the fragmentation of next sample.

12. After finishing fragmentation for all samples, wash the shearing assemblies with buffers and wash times in the order as below:

- (1) 0.2M HCl: 6
- (2) 0.2M NaOH: 6
- (3) MillQ-H₂O: 6

At last, moisturize the pistons with 200 μ l MillQ-H₂O.

13. Verify the fragment size with 0.6% agarose gel electrophoresis at 120 Voltage for 50 min, using the 1 kb DNA Ladder as size marker. DNA fragments range from 5 to 8 kb with smear ranging from 3 to 10 kb (Figure 2).

Perform end-repair

14. Prepare the following reaction mix:

Fragmented DNA sample (from step 11)	75 μ l
10 $\times$ PNK Buffer	10 μ l
UltraPure TM Water	2.4 μ l
Natural dNTP Mix (25mM)	1.6 μ l
T4 DNA Polymerase	5 μ l
T4 PNK	5 μ l
Klenow DNA Polymerase	1 μ l
Total volume =	100 μ L
Mix well and spin down.	

15. Incubate 30 min at 20°C.

16. Purify with an AMPure beads, following the instructions below and elute in 71 μ l of EB buffer.

- a. Mix the bottle of AMPure beads by gentle shaking. To generate a 100 μ l end-repair reaction volume, add 100 μ l of bead suspension in a 1.5-ml micro-centrifuge tube.
- b. Mix by vortexing for 30 sec to achieve a homogeneous suspension and leave at room temperature (20 °C) for 5 min.
- c. Spin down briefly and fix the tube into a magnetic separator. Leave until the solution is clear (~10 min).
- d. Remove the solution by pipetting without disturbing the beads.
- e. Add 200 μ l of 70% ethanol to the beads without disturbing the beads, leave for 30 sec at room temperature, remove the ethanol by pipetting, and discard the ethanol. Repeat this washing step one more time.

- f. Leave the lid of the tube open to allow the beads to dry. This can take longer than the recommended 20 min, but depends upon the volume of ethanol remaining after the previous step. The bead pellet will have a cracked appearance when it is dry.
- g. Remove the tube from the magnet, add 71 μ l water and mix thoroughly by pipetting. Be sure to resuspend the beads fully.
- h. Return the tube to the magnet and leave for approximately 10 min.
- i. Collect and retain the liquid (as this now contains the DNAs), and discard the beads.

Label with biotin

17. Prepare the following reaction mix:

End-repaired DNA sample (from step 16)	69 μ l
10 $\times$ PNK Buffer	10 μ l
UltraPure TM Water	0 μ l
Biotin dNTP Mix	10 μ l
T4 DNA Polymerase	5 μ l
T4 PNK	5 μ l
Klenow DNA Polymerase	1 μ l
Total volume = 100 μ l	
Mix well and spin down.	

18. Incubate 30 min at 20°C.

19. Purify on AMPure beads as described above (steps 16a to 16i), and elute in 67 μ l of EB buffer.

Circularize DNA

20. Obtain the DNA concentration (from step 19) based on Qubit HS.

21. Prepare the following reaction mix:

Biotin labeled DNA sample (from step 19)	800 ng
10 $\times$ T4 Ligase Buffer	40 μ l
UltraPure Water	x μ l*
10 $\times$ T4 DNA Ligase	20 μ l
25 mM ATP	5 μ l
Total volume = 400 μ l	

22. Incubate 16 h at 16°C.

Digest linear DNAs

23. Prepare the following reaction mix:

Circular DNA sample (from step 22)	100 μ l
Plasmid-safe 10 $\times$ Buffer	12 μ l

Plasmid-safe™ DNase	4 µl
Exonuclease I	2 µl
25mM ATP	4 µl
Total volume = 122 µl	

24. Incubate 30 min at 37 °C.

25. Incubate 20 min at 75°C.

Fragment circular DNAs

26. Transfer DNA sample (from step 25) to a 13×65 mm Covaris vial.

27. Add 4 µl 0.5 M EDTA and 20 µl glass beads into the vial. Mix well and spin down. Add TE buffer to a total volume of 600 µl.

28. Insert the sample vial into the holder, run the Covaris with the settings given below to generate the fragment sizes ranging from 350 to 400 bp:

Duty/cycle(%):	10
Intensity:	10
Cycles per burst:	1000
Time(s):	60
Cycle:	4

29. Transfer DNA fragments into a 1.5-ml tube, 13000 rpm, 4 min.

30. Take supernatant, recover the DNA sample with the mini QIAGEN quick PCR Purification Kit, and elute in 50 µl of EB buffer.

Purify biotin-labeled DNA

31. Take 20 µl Streptavidin Magnetic Beads into a Non-stick RNase-Free 1.5-ml Microfuge Tube.

32. Wash the Streptavidin Magnetic Beads with 50 µl Bead Binding Buffer, twice. Use 50 µl Bead Binding Buffer to resuspend the Streptavidin Magnetic Beads.

33. Add the beads (from step 32) into the DNA sample (from step 30).

34. Incubate 15 min at 20°C.

35. Remove the supernatant by pipetting without disturbing the beads.

36. Wash the beads with 200 µl Bead Wash Buffer, 3 times.

37. Remove the supernatant by pipetting, resuspend the beads with 200 µl EB buffer.

38. Remove EB Buffer. Resuspend the beads with 50 µl EB buffer. Collect and retain the magnetic beads suspension, transfer it into a new Non-stick RNase-Free 1.5-ml microfuge tube.

Perform end-repair

39. Prepare the following reaction mix:

Fragmented DNA sample (from step 4)	50 µl
10×T4 PNK Buffer	10 µl
UltraPure™ Water	27.4 µl
25mM dNTP Mix	1.6 µl
T4 DNA Polymerase	5 µl
T4 PNK	5 µl
Klenow DNA Polymerase	1 µl
Total volume = 100 µl	
Mix well and spin down.	

40. Incubate 30 min at 20°C.

41. Remove the supernatant by pipetting without disturbing the beads.

42. Wash the beads as described above (steps 35 to 38), and elute in 32 µl of EB buffer.

Create an A-overhang

43. Prepare the following reaction mix:

End-repaired DNA sample (from step 4)	32 µl
10× Blue Buffer	5 µl
1 mM dATP	10 µl
Klenow 3′-5′(exo)	3 µl
Total volume = 50 µl	
Mix well and spin down.	

44. Incubate 30 min at 37°C.

45. Remove the solution by pipetting, taking care not to disturb the beads.

46. Wash the beads as described above (steps 35 to 38), and elute in 19 µl of EB buffer.

Ligate adapter

47. Prepare the following reaction mix:

End-repaired, A-tailed DNA sample (from step 46)	19 µl
2×Rapid Buffer	25 µl
PE Adapter Oligo Mix	1 µl
10×T4 DNA Ligase	5 µl
Total volume = 50 µl	
Mix well and spin down.	

48. Incubate 15 min at 20°C.

49. Remove the supernatant by pipetting, taking care not to disturb the beads.

50. Wash the beads as described above (steps 35 to 37).

51. Remove EB buffer. Resuspend the beads to the next reaction system.

Basic Protocol 2: PCR Amplification, Size-Selection and Measurement

After amplification of DNA templates by PCR with 18 cycles, DNAs with a size range of 450 to 650 bp (with adaptors) are selected by gel electroporation and extraction. Further, we used the same methods for DNA measurement and quantitation of library concentration as described in our previously published protocol (Dong, Xie, et al., 2017). Follow the detailed method, including the materials and equipment in that protocol (UNIT 8.17).

In terms of the evaluation from our pilot study, the minimum requirement of read-pairs for the mate-pair library for our analysis is only 30 million read-pairs 50 bp in size, which is equivalent to 1× whole-genome base coverage. In our protocol, four samples with different barcodes with equal molality are mixed into a pooled library for paired-end sequencing in a single lane on a HiSeq 2000™ platform (Illumina) (Dong et al., 2014).

Materials

DNA library (see above)

Phusion DNA Polymerase (NEB)

Pfx polymerase (Invitrogen, cat.no. 11708-013)

10× pfx buffer (Invitrogen, cat.no. 11708-013)

2.5 mM dNTP mix (Invitrogen, cat. no. R72501)

PCR Primers [see Table 8.17.1 Genomic DNA Oligonucleotide Sequences in UNIT 8.17 (Dong, Xie, et al., 2017)]

EB buffer (Qiagen)

10 mM Tris·Cl, pH 8.5(Ambion)

0.2- and 1.5-ml microcentrifuge tubes (Ambion)

1.5 ml Non-stick RNase-Free Microfuge Tube (Ambion)

Molecular Biology Agarose (Bio-Rad)

Low Range Ultra Agarose (Bio-Rad)

Probe [10 μM, see Table 8.17.1 Genomic DNA Oligonucleotide Sequences (Dong, Xie, et al., 2017)]

Equipment

Vortex-5 (Haimen Kylin-Bell Lab Instruments Co., Ltd)

Bioanalyzer 2100 (Agilent)

Thermal cycler (Qiagen)

Magnetic separator (Dexter Magnetic Technologies)

Microcentrifuge (Eppendorf)

Perform PCR

1. Prepare the following reaction mix:

Phusion DNA Polymerase	25 µl
PCR primer 1.0 (10 pmol/µl)	1 µl
PCR primer 2.0 (10 pmol/µl)	1 µl
UltraPure™ Water	23 µl
Total volume = 50 µl	

2. The beads (from step 50) were resuspended by the PCR reaction mix.
3. Mix well and spin down.
4. Carry out the amplification using the following cycling conditions in a thermal cycler:

1 cycle:	30 s 98°C (initial denaturation)
18 cycles:	10 sec 98°C (denaturation)
	30 sec 65°C (annealing)
	30 sec 72°C (extension)
1 cycle:	5 min 72°C (final extension)
	Indefinitely 4°C (hold)
5. Resuspend the beads after finishing the reactions. Transfer the liquid beads to a new Non-stick RNase-Free 1.5-ml Microfuge Tube.
6. Discard the beads, collect and retain the supernatant. Perform gel electrophoresis with the supernatant.

Select DNA size

7. Verify the integrity of DNA with 2% agarose gel electrophoresis at 100 Voltage for 2h, using the NEB 50 bp DNA Ladder as size marker.
8. Cut the gel in a range of 450 to 650 bp.
9. Recover the DNA sample with QIAquick PCR Purification Kit, and elute in 20 µl EB buffer.

Measure DNA and quantitate library concentration

10. Perform DNA measurement with Bioanalyzer 2100 (Agilent) and quantitate the library concentration with qPCR assay described in UNIT 8.17 (Dong, Xie, et al., 2017).

Pool four libraries into a library

11. Four libraries are mixed with equal molecular amounts (pM) into a library, using the indexed adapter tubes from a TruSeq LT Kit (Dong, Xie, et al., 2017). Each pooled library is subsequently subjected to 50-bp-end multiplex sequencing (PE 50) on the HiSeq 2000™ (Illumina) following the standard protocol.

Basic Protocol 3: Bioinformatic Pipeline

The analysis pipeline and the detailed algorithm are described in our pilot study (Dong, Wang, et al., 2017). Briefly, the detection pipeline includes five steps: (1) event clustering of chimeric read-pairs; (2) filtering of systematic errors with a dataset of false positives in a general population, such as the existing genomic data from the 1000 Genomes Project (Cao et al., 2015; Maretty et al., 2017); (3) filtering of random errors with a cluster property matrix (i.e., the number of supporting read-pairs); (4) Aligned orientation of each event based on the q/p arm genetic exchange; and (5) molecular delineation of the composition of the derivative chromosome(s) with assistance of CNV analysis (Dong et al., 2016). Each detailed procedure is described below and the required computational resources are mentioned accordingly.

Materials

Linux based command system

Alignment file [via Burrows-Wheeler Aligner, bwa (Li & Durbin, 2009)] from paired-end WGS data

BCA.tar.gz

samtools (<http://samtools.sourceforge.net>)

picard (<http://broadinstitute.github.io/picard/>)

Software download and decompression

The programs and the updated modules can be downloaded from <https://sourceforge.net/projects/bca-analysis/files/>. Run the following commands from the command line:

1. Decompressing the source file:

```
tar xzvf BCA.tar.gz
```

2. Change to the source directory:

```
cd BCA
```

In this folder, the file structure is:

- Scripts and programs:
 - ./physical_depth_chimeric: a script for extracting the chimeric read-pairs and reporting the physical coverage of the file input
 - ./raw_fil: a script to eliminate the unmapped or multiple aligned reads or reads with more than 4% mismatches
 - ./difchr_pe: a script for making pairs of the chimeric read-pairs, which aligned to different chromosomes

- ./samechr_pe: a script for making pairs of the chimeric read-pairs aligned to the same chromosomes with larger distance than expected insert size (> 10 kb)
- ./CRA/Difchr/difchr.sh: a shell and a set of scripts in the same folder for identification of inter-chromosomal rearrangements
- ./CRA/Samechr/samechr.sh: a shell and a set of scripts in the same folder for identification of intra-chromosomal rearrangements
- ./CRA/check_difchr_cluster: a script for identification of balanced translocations
- ./CRA/check_samechr_cluster: a script for identification of inversions

- Folders and files:

- ./Cohort/Build_control_1/difchr.control.final.1: a dataset of the number of linkages from a bin to the locations in other chromosomes, data format as below:

```
chr1  0  145  0:70:55:6:14
```

Information is provided as chromosome, bin (10 kb in size), the number of linkages from this bin to the locations in other chromosomes and the number of linkages in different ethnic populations [AFR (African), AMR (American), ASN (East Asian), EUR (European) or SAN (South Asian)(Sudmant et al., 2015)]

- ./Cohort/Build_control_1/samechr.control.final.1: a dataset of the number of linkages from a bin to the other locations in the same chromosome, data format as below:

```
chr1  0  3  0:1:1:0:1
```

Information is provided as chromosome, bin (10 kb in size), the number of linkages from this bin to the other locations in the same chromosome and the number of linkages in different ethnic populations

- ./Cohort/Build_control_2/difchr.control.final.2: a dataset of the number of linkages between a bin and another bin from the other chromosomes, data format as below:

```
chr1  chr2  0  18214  1  0:1:0:0:0
```

Information is provided as:

- ◆ Column 1 and 3: chromosome and location from one bin (10 kb in size)

- ◆ Column 2 and 4: chromosome and the location of the other bin in the other chromosome
- ◆ Column 5 and Column 6: the number of linkages between these two bins and the number of linkages in different ethnic populations
- ./Cohort/Build_control_2/samechr.control.final.2: a dataset of the number of linkages between a bin and another bin from the other location in the same chromosome, data format as below:

```
chr1  chr1  0  24923  1  0:1:0:0:0
```

Information is provided as:

- ◆ Column 1 and 2: chromosome
- ◆ Column 3 and 4: locations of one bin and another
- ◆ Column 5 and Column 6: the number of linkages between these two bins and the number of linkages in different ethnic populations
- ./ DataBase/hg19.cytoBand.txt: a file of the locations of each band region in hg19
- ./ DataBase/numgap.list: a file of the unidentified region (N) in hg19
- ./ DataBase/refGene/refGene.txt: a file of the locations of each transcript and the corresponding gene

Data processing and QC analysis

3. extract chimeric reads from sorted (by coordinate) bam file

```
./physical_depth_chimeric ./BAM ./difchr.gz ./samechr.gz <alignment method: aln or mem> <M|N: -M option for mem> <path for samtools (use absolute address)>
```

This script is used for extracting the chimeric read-pairs, which aligned to different chromosomes or to the same chromosome but with a distance larger than expected (10 kb), and it will report the physical coverage of the file input after finishing run. It should be used for the sorted bam (by coordinate via samtools or picard). And as “mem” alignment method will report the shorter split hits, please remember to specify whether the shorter split hits have been marked as secondary hits (-M option in bwa). And the absolute path for samtools should be also provided. If there are more than one alignment files for a sample, run the script independently and merge the files together (via zcat) if the total physical coverage is larger than 8.25-fold. The memory required for each running depends on the amount of reads for each file, it takes 3 to 6 G memory typically.

4. filter out the read with low quality

```
./raw_fil ./difchr.gz ./difchr.fil.gz <alignment method: aln or mem> && ./raw_fil ./
samechr.gz ./samechr.fil.gz <alignment method: aln or mem>
```

This script aims to eliminate the unmapped or multiple aligned reads or reads with more than 4% mismatches, and to prepare the input files for making pairs. Only 1 G memory is required.

5. sort the reads to mate pair

```
zless -NS ./difchr.fil.gz |sort -k 1 |gzip -9 > ./difchr.fil.sort.gz && zless -NS ./
samechr.fil.gz |sort -k 1 |gzip -9 > ./samechr.fil.sort.gz
```

sort each file by the read name

6. make pairs of the reads

```
./difchr_pe ./difchr.fil.sort.gz ./difchr.fil.sort.final.gz && ./samechr_pe ./
samechr.fil.sort.gz ./samechr.fil.sort.final.gz
```

Identify BCAs

7. inter-chromosomal rearrangement detectionsh

```
./CRA/Difchr/difchr.sh ./difchr.fil.sort.final.gz <sample name or ID> <ethnic
pop:AFR,AMR,ASN,EUR,SAN> <500|5000>
```

This shell and the scripts in the same folder provide the detection of inter-chromosomal rearrangements. The detection steps follow by sorting the read-pairs, event clustering by read 1, event clustering by read 2 and event filtering. Chimeric read-pairs are shown in *.dif.fil.final.gz.sort.ratio1.2 (Figure 3A) and the result of inter-chromosomal rearrangements is shown in *.dif.fil.final.gz.sort.result.clean.1, respectively.

8. intra-chromosomal rearrangement detectionsh

```
./CRA/Samechr/samechr.sh ./samechr.fil.sort.final.gz <sample name or ID> <ethnic
pop:AFR,AMR,ASN,EUR,SAN> <500|5000>
```

This shell and the scripts in the same folder provide the detection of intra-chromosomal rearrangements. The detection steps are similar to the detection of inter-chromosomal rearrangements described in step 7. Chimeric read-pairs are shown in *.same.fil.final.gz.sort.ratio1.2 (Figure 3B) and the result of intra-chromosomal rearrangements is shown in *.same.fil.final.gz.sort.result.clean.1, respectively.

Commentary**Background Information**

Balanced chromosomal rearrangements (or abnormalities, BCAs), including translocations, insertions, and inversions, are well known to cause reproductive problems such as recurrent miscarriage due to the high risk of unbalanced segregation (Sugiura-Ogasawara et al., 2004).

In addition, studies of BCAs with clinical anomalies such as in the Developmental Genome Anatomy Project (DGAP) (Ordulu et al., 2016; Redin et al., 2017; Talkowski, Ordulu, et al., 2012; Talkowski, Rosenfeld, et al., 2012) among others (De Gregori et al., 2007; Nilsson et al., 2017) have a high yield in identification of genetic disorders due to gene disruption or imbalances involving the BCA breakpoints or dysregulation by a positional effect. Therefore, facile and accurate identification of BCAs is urgently required so that future informed treatment strategies might be formulated.

Since the 1970s, the detection of BCAs has relied on G-banded chromosome analysis as the current high-resolution techniques, such as chromosomal microarray analysis, do not typically identify such kinds of abnormalities. However, detection based on G-banded chromosome analysis reflects banding resolution, and studies using subtelomeric fluorescence in situ hybridization (FISH) demonstrated that cryptic balanced translocations may lead to mental retardation in the offspring of carriers (Ravnan et al., 2006).

In the past decade since its development, Next-generation sequencing (NGS) has been proven as a state-of-the-art technology for molecular delineation of BCA breakpoints subsequent to knowledge of the involved chromosomal regions from prior cytogenetic analyses. As cytogenetic results are unavailable in some cases [such as in the 1000 Genomes Project (Sudmant et al., 2015)], building on this knowledge, we developed a method for identification of BCAs independently from G-banded chromosome analysis (Dong et al., 2014). We further demonstrated the feasibility of using this approach in an existing dataset of genomic data (from the 1000 Genomes Project), which are without karyotypes, for identification of BCAs. Particularly, we identified some BCAs, such as t(16;17)(q23.1;q24.2) in NA18612, which would not be reliably detected by G-banded chromosome analysis due to translocated segments of similar banding patterns and similar sizes (Dong, Wang, et al., 2017).

In this protocol, we describe this approach for BCA detection in a clinical setting. The resolution of intra-chromosomal rearrangements (i.e., inversion) is 10 kb according to the insert size range of the fragmented DNAs. Compared to small-insert (i.e., 300~500bp) library, the usage of mate-pair library in this protocol not only reduces the read-pairs amount (or base coverage) required, thus, reducing the sequencing cost per test, but also increasing the detection of BCAs which might be mediated by repetitive elements (Carvalho & Lupski, 2016; Dong, Wang, et al., 2017). The complete procedure of this testing takes 13 days and the total cost was estimated to be less than 600 USD per sample on a HiSeq 2000™ platform (Illumina) described in our pilot study (ref). As the total cost and turnaround time are reduced the protocol can be optimized based on the new technique and platform (i.e., HiSeq 2500™).

Critical Parameters

In the basic protocol for DNA fragmentation, library construction, PCR amplification and sequencing:

- a. HydroShear is used for the DNA fragmentation because of the narrow size of DNA fragments and the higher recovery rate. An alternative method would be

using miniTUBE (red) for 5 kb shearing from Covaris with AFA technology (Hanscom & Talkowski, 2014).

- b.** DNA fragment size is one of the important QC criteria, re-sampling DNA for fragmentation is required if the fragment size is much smaller than expected (Figure 2).

In the protocol for BCA analysis, several critical parameters are:

- a.** The minimum physical coverage, which is estimated by counting aligned distances from the non-chimeric and uniquely aligned read-pairs, was estimated to be 8.25-fold in terms of our study (Dong, Wang, et al., 2017). This QC criteria is required before BCA analysis.
- b.** As genomic imbalances are commonly found involving the breakpoints of BCAs, molecular delineation of BCA events have to be performed with consideration of CNV analysis (Dong et al., 2014; Redin et al., 2017), particularly in the region(s) with the putative breakpoint(s).
- c.** As the read length used in our detection is as short as 50 bp, we can fine map the breakpoints to base pair resolution with assistance of Sanger sequencing results.

Troubleshooting

In this study, the dataset of linkages was collected and built up by using data from the 1000 Genome Project (<http://www.1000genomes.org/>), in which the majority of the subjects were sequenced with small-insert libraries (ref). It is useful to build up an in-house dataset, because:

- a.** The same ethnic group as samples being tested is advised, particularly for identification of intra-chromosomal rearrangements (i.e., inversions), as it would be helpful to consider whether it is a population specific event if present in more than one sample.
- b.** The same sequencing platform, library construction and alignment method (i.e., bwa) are required as systematic errors may be introduced by the specific platform, library construction or alignment method. Of note, the pipeline for BCA detection was developed based on data from sequencing by synthesis (Goodwin et al., 2016) (i.e., HiSeq, MiSeq or NextSeq); it would not be feasible for data generated by another sequencing method.

Anticipated Results

The incidence of balanced translocations is estimated as 1 in 500 to 1 in 625 in the general population, and it is increased in disease groups. Although some BCAs cannot be reliably detected by G-banded chromosome analysis, the incidence of balanced translocations in a certain study group can be estimated based on the current studies and would not be dramatically higher than expected. Although some inversions are expected to be detectable at the microscopic level, they might be cryptic by G-band chromosome analysis because the inverted segments have a similar or identical banding patterns to the normal chromosome (e.g., inversion wholly within a light or dark band). But there are limited publications

discussing the incidence of microscopic inversions by using a WGS-based approach (Collins et al., 2017).

If the protocols are followed strictly and a local dataset of linkages is generated using approximate ethnic-matched samples and on the same platform, a reasonable number of BCAs are expected to be observed.

Time Considerations

The turnaround time for each step was estimated in our pilot study (Dong et al., 2014) and in total, this testing takes 13 days using a HiSeq 2000™ platform (Illumina). Turnaround time may be shortened if a rapid sequencing platform is used (i.e., HiSeq 2500™).

Acknowledgments

This project is supported by the National Natural Science Foundation of China (No.81370715, 81300075), the Natural Science Foundation of Guangdong Province (No. 2014A030313795), the Guangdong Enterprise Key Laboratory of Human Disease Genomics (No.2011A060906007), the Shenzhen Municipal Government of China (No.CXZZ20150330171838997), Shenzhen Engineering Laboratory for Birth Defects Screening (DRC-SZ [2011]861), Shenzhen Key Laboratory of genomics (No.CXB200903110066A) and Health and Medical Research Fund (HMRP, Project 04152666 and 6904471). CCM acknowledges the NIH/NIGMS GM061354 for support.

Literature Cited

- Cao H, Wu H, Luo R, Huang S, Sun Y, Tong X, Xie Y, Liu B, Yang H, Zheng H, Li J, Li B, Wang Y, Yang F, Sun P, Liu S, Gao P, Huang H, Sun J, Chen D, He G, Huang W, Huang Z, Li Y, Tellier LC, Liu X, Feng Q, Xu X, Zhang X, Bolund L, Krogh A, Kristiansen K, Drmanac R, Drmanac S, Nielsen R, Li S, Wang J, Yang H, Li Y, Wong GK, Wang J. De novo assembly of a haplotype-resolved human genome. *Nat Biotechnol.* 2015; 33(6):617–622. [PubMed: 26006006]
- Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. *Nat Rev Genet.* 2016; 17(4):224–238. [PubMed: 26924765]
- Collins RL, Brand H, Redin CE, Hanscom C, Antolik C, Stone MR, Glessner JT, Mason T, Pregno G, Dorrani N, Mandrile G, Giachino D, Perrin D, Walsh C, Cipicchio M, Costello M, Stortchevoi A, An JY, Currall BB, Seabra CM, Ragavendran A, Margolin L, Martinez-Agosto JA, Lucente D, Levy B, Sanders SJ, Wapner RJ, Quintero-Rivera F, Kloosterman W, Talkowski ME. Defining the diverse spectrum of inversions, complex structural variation, and chromothripsis in the morbid human genome. *Genome Biol.* 2017; 18(1):36. [PubMed: 28260531]
- De Gregori M, Ciccone R, Magini P, Pramparo T, Gimelli S, Messa J, Novara F, Vetro A, Rossi E, Maraschio P, Bonaglia MC, Anichini C, Ferrero GB, Silengo M, Fazzi E, Zatterale A, Fischetto R, Previdere C, Belli S, Turci A, Calabrese G, Bernardi F, Meneghelli E, Riegel M, Rocchi M, Gueneri S, Lalatta F, Zelante L, Romano C, Fichera M, Mattina T, Arrigo G, Zollino M, Giglio S, Lonardo F, Bonfante A, Ferlini A, Cifuentes F, Van Esch H, Backx L, Schinzel A, Vermeesch JR, Zuffardi O. Cryptic deletions are a common finding in “balanced” reciprocal and complex chromosome rearrangements: a study of 59 patients. *J Med Genet.* 2007; 44(12):750–762. [PubMed: 17766364]
- Dong Z, Jiang L, Yang C, Hu H, Wang X, Chen H, Choy KW, Hu H, Dong Y, Hu B, Xu J, Long Y, Cao S, Chen H, Wang WJ, Jiang H, Xu F, Yao H, Xu X, Liang Z. A robust approach for blind detection of balanced chromosomal rearrangements with whole-genome low-coverage sequencing. *Hum Mutat.* 2014; 35(5):625–636. [PubMed: 24610732]
- Dong Z, Wang H, Chen H, Jiang H, Yuan J, Yang Z, Wang WJ, Xu F, Guo X, Cao Y, Zhu Z, Geng C, Cheung WC, Kwok YK, Yang H, Leung TY, Morton CC, Cheung SW, Choy KW. Identification of Balanced Chromosomal Rearrangements Previously Unknown Among Participants in the 1000 Genomes Project: Implications for Interpretation of Structural Variation in Genomes and the Future of Clinical Cytogenetics. *Genet Med*, Accepted. 2017

- Dong Z, Xie W, Chen H, Xu J, Wang H, Li Y, Wang J, Chen F, Choy KW, Jiang H. Copy-Number Variants Detection by Low-Pass Whole-Genome Sequencing. *Curr Protoc Hum Genet*. 2017; 94:8 17 11–18 17 16. [PubMed: 28696555]
- Dong Z, Zhang J, Hu P, Chen H, Xu J, Tian Q, Meng L, Ye Y, Wang J, Zhang M, Li Y, Wang H, Yu S, Chen F, Xie J, Jiang H, Wang W, Choy KW, Xu Z. Low-pass whole-genome sequencing in clinical cytogenetics: a validated approach. *Genet Med*. 2016; 18(9):940–948. [PubMed: 26820068]
- Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet*. 2016; 17(6):333–351. [PubMed: 27184599]
- Hanscom C, Talkowski M. Design of large-insert jumping libraries for structural variant detection using Illumina sequencing. *Curr Protoc Hum Genet*. 2014; 80:7 22 21–29.
- Jonas RK, Montojo CA, Bearden CE. The 22q11.2 deletion syndrome as a window into complex neuropsychiatric disorders over the lifespan. *Biol Psychiatry*. 2014; 75(5):351–360. [PubMed: 23992925]
- Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST. Working Group of the American College of Medical Genetics Laboratory Quality Assurance, C. American College of Medical Genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. *Genet Med*. 2011; 13(7):680–685. [PubMed: 21681106]
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25(14):1754–1760. [PubMed: 19451168]
- Oliver-Bonet M, Navarro J, Carrera M, Egozcue J, Benet J. Aneuploid and unbalanced sperm in two translocation carriers evaluation of the genetic risk. *Mol Hum Reprod*. 2002; 8(10):958–963. [PubMed: 12356948]
- Mackie Ogilvie C, Scriven PN. Meiotic outcomes in reciprocal translocation carriers ascertained in 3-day human embryos. *European journal of human genetics : EJHG*. 2002; 10(12):801–806. [PubMed: 12461686]
- Marett L, Jensen JM, Petersen B, Sibbesen JA, Liu S, Villesen P, Skov L, Belling K, Theil Have C, Izarzugaza JMG, Grosjean M, Bork-Jensen J, Grove J, Als TD, Huang S, Chang Y, Xu R, Ye W, Rao J, Guo X, Sun J, Cao H, Ye C, van Beusekom J, Espeseth T, Flindt E, Friberg RM, Halager AE, Le Hellard S, Hultman CM, Lescai F, Li S, Lund O, Longren P, Mailund T, Matey-Hernandez ML, Mors O, Pedersen CNS, Sicheritz-Ponten T, Sullivan P, Syed A, Westergaard D, Yadav R, Li N, Xu X, Hansen T, Krogh A, Bolund L, Sorensen TIA, Pedersen O, Gupta R, Rasmussen S, Besenbacher S, Borglum AD, Wang J, Eiberg H, Kristiansen K, Brunak S, Schierup MH. Sequencing and de novo assembly of 150 genomes from Denmark as a population reference. *Nature*. 2017; 548(7665):87–91. [PubMed: 28746312]
- Nilsson D, Pettersson M, Gustavsson P, Forster A, Hofmeister W, Wincent J, Zachariadis V, Anderlid BM, Nordgren A, Makitie O, Wirta V, Kaller M, Vezzi F, Lupski JR, Nordenskjold M, Syk Lundberg E, Carvalho CM, Lindstrand A. Whole-Genome Sequencing of Cytogenetically Balanced Chromosome Translocations Identifies Potentially Pathological Gene Disruptions and Highlights the Importance of Microhomology in the Mechanism of Formation. *Hum Mutat*. 2017; 38(2):180–192. [PubMed: 27862604]
- Ordulu Z, Kammin T, Brand H, Pillalamarri V, Redin CE, Collins RL, Blumenthal I, Hanscom C, Pereira S, Bradley I, Crandall BF, Gerrol P, Hayden MA, Hussain N, Kanengisser-Pines B, Kantarci S, Levy B, Macera MJ, Quintero-Rivera F, Spiegel E, Stevens B, Ulm JE, Warburton D, Wilkins-Haug LE, Yachelevich N, Gusella JF, Talkowski ME, Morton CC. Structural Chromosomal Rearrangements Require Nucleotide-Level Resolution: Lessons from Next-Generation Sequencing in Prenatal Diagnosis. *Am J Hum Genet*. 2016; 99(5):1015–1033. [PubMed: 27745839]
- Ravnan JB, Tepperberg JH, Papenhausen P, Lamb AN, Hedrick J, Eash D, Ledbetter DH, Martin CL. Subtelomere FISH analysis of 11 688 cases: an evaluation of the frequency and pattern of subtelomere rearrangements in individuals with developmental disabilities. *J Med Genet*. 2006; 43(6):478–489. [PubMed: 16199540]
- Redin C, Brand H, Collins RL, Kammin T, Mitchell E, Hodge JC, Hanscom C, Pillalamarri V, Seabra CM, Abbott MA, Abdul-Rahman OA, Aberg E, Adley R, Alcaraz-Estrada SL, Alkuraya FS, An Y, Anderson MA, Antolik C, Anyane-Yeboah K, Atkin JF, Bartell T, Bernstein JA, Beyer E, Blumenthal I, Bongers EM, Brilstra EH, Brown CW, Bruggenwirth HT, Callewaert B, Chiang C,

Corning K, Cox H, Cuppen E, Currall BB, Cushing T, David D, Deardorff MA, Dheedene A, D'Hooghe M, de Vries BB, Earl DL, Ferguson HL, Fisher H, FitzPatrick DR, Gerrol P, Giachino D, Glessner JT, Gliem T, Grady M, Graham BH, Griffis C, Gripp KW, Gropman AL, Hanson-Kahn A, Harris DJ, Hayden MA, Hill R, Hochstenbach R, Hoffman JD, Hopkin RJ, Hubshman MW, Innes AM, Irons M, Irving M, Jacobsen JC, Janssens S, Jewett T, Johnson JP, Jongmans MC, Kahler SG, Koolen DA, Korzelius J, Kroisel PM, Lacassie Y, Lawless W, Lemyre E, Leppig K, Levin AV, Li H, Li H, Liao EC, Lim C, Lose EJ, Lucente D, Macera MJ, Manavalan P, Mandrile G, Marcelis CL, Margolin L, Mason T, Masser-Frye D, McClellan MW, Mendoza CJ, Menten B, Middelkamp S, Mikami LR, Moe E, Mohammed S, Mononen T, Mortenson ME, Moya G, Nieuwint AW, Ordulu Z, Parkash S, Pauker SP, Pereira S, Perrin D, Phelan K, Aguilar RE, Poddighe PJ, Pregno G, Raskin S, Reis L, Rhead W, Rita D, Renkens I, Roelens F, Ruliera J, Rump P, Schilit SL, Shaheen R, Sparkes R, Spiegel E, Stevens B, Stone MR, Tagoe J, Thakuria JV, van Bon BW, van de Kamp J, van Der Burgt I, van Essen T, van Ravenswaaij-Arts CM, van Roosmalen MJ, Vergult S, Volker-Touw CM, Warburton DP, Waterman MJ, Wiley S, Wilson A, Yerena-de Vega MC, Zori RT, Levy B, Brunner HG, de Leeuw N, Kloosterman WP, Thorland EC, Morton CC, Gusella JF, Talkowski ME. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat Genet.* 2017; 49(1):36–45. [PubMed: 27841880]

Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Hsi-Yang Fritz M, Konkel MK, Malhotra A, Stutz AM, Shi X, Paolo Casale F, Chen J, Hormozdiari F, Dayama G, Chen K, Malig M, Chaisson MJ, Walter K, Meiers S, Kashin S, Garrison E, Auton A, Lam HY, Jasmine Mu X, Alkan C, Antaki D, Bae T, Cerveira E, Chines P, Chong Z, Clarke L, Dal E, Ding L, Emery S, Fan X, Gujral M, Kahveci F, Kidd JM, Kong Y, Lameijer EW, McCarthy S, Flicek P, Gibbs RA, Marth G, Mason CE, Menelaou A, Muzny DM, Nelson BJ, Noor A, Parrish NF, Pendleton M, Quitadamo A, Raeder B, Schadt EE, Romanovitch M, Schlattl A, Sebra R, Shabalina AA, Untergasser A, Walker JA, Wang M, Yu F, Zhang C, Zhang J, Zheng-Bradley X, Zhou W, Zichner T, Sebat J, Batzer MA, McCarroll SA, Genomes Project, C. Mills RE, Gerstein MB, Bashir A, Stegle O, Devine SE, Lee C, Eichler EE, Korbel JO. An integrated map of structural variation in 2,504 human genomes. *Nature.* 2015; 526(7571):75–81. [PubMed: 26432246]

Sugiura-Ogasawara M, Ozaki Y, Sato T, Suzumori N, Suzumori K. Poor prognosis of recurrent aborters with either maternal or paternal reciprocal translocations. *Fertil Steril.* 2004; 81(2):367–373. [PubMed: 14967375]

Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi TK, Shen Y, Borowsky M, Daly MJ, Morton CC, Gusella JF. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *Am J Hum Genet.* 2011; 88(4):469–481. [PubMed: 21473983]

Talkowski ME, Ordulu Z, Pillalamarri V, Benson CB, Blumenthal I, Connolly S, Hanscom C, Hussain N, Pereira S, Picker J, Rosenfeld JA, Shaffer LG, Wilkins-Haug LE, Gusella JF, Morton CC. Clinical diagnosis by whole-genome sequencing of a prenatal sample. *N Engl J Med.* 2012; 367(23):2226–2232. [PubMed: 23215558]

Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, Pereira S, Ruderfer D, Kirby A, Ripke S, Harris DJ, Lee JH, Ha K, Kim HG, Solomon BD, Gropman AL, Lucente D, Sims K, Ohsumi TK, Borowsky ML, Loranger S, Quade B, Lage K, Miles J, Wu BL, Shen Y, Neale B, Shaffer LG, Daly MJ, Morton CC, Gusella JF. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell.* 2012; 149(3):525–537. [PubMed: 22521361]

Tang YC, Amon A. Gene copy-number alterations: a cost-benefit analysis. *Cell.* 2013; 152(3):394–405. [PubMed: 23374337]

Van Dyke DL, Weiss L, Roberson JR, Babu VR. The frequency and mutation rate of balanced autosomal rearrangements in man estimated from prenatal genetic studies for advanced maternal age. *Am J Hum Genet.* 1983; 35(2):301–308. [PubMed: 6837576]

Warburton D. De novo balanced chromosome rearrangements and extra marker chromosomes identified at prenatal diagnosis: clinical significance and distribution of breakpoints. *Am J Hum Genet.* 1991; 49(5):995–1013. [PubMed: 1928105]

Figure 1. Workflow of BCA detection by low-pass whole-genome sequencing
The workflow includes three basic protocols: (1) DNA preparation and library construction, (2) PCR amplification, size selection and DNA measurement, (3) Analysis QC and (4) BCA analysis. Detailed procedures are described in each protocol.

Figure 2. Verification of DNA size after fragmentation

Determine the size range of DNA fragments by gel electrophoresis after fragmentation. Samples are shown in each column and the markers (1 kb ladder) are shown in the first and the last column, respectively. Sizes of four markers are indicated with arrows accordingly. The QC failed sample (sample 5 and 6), both of which are with smaller size range than expected indicated by red arrows.

A

chr17	chr16	64950291	75337992	-	+	0
chr17	chr16	64950982	75336323	-	+	0
chr17	chr16	64951015	75339305	-	+	0
chr17	chr16	64951503	75336998	-	+	3
chr17	chr16	64952323	75337688	-	+	0
chr17	chr16	64952363	75337677	-	+	0
chr17	chr16	64952465	75340772	-	+	1
chr17	chr16	64953109	75333026	+	-	2
chr17	chr16	64953171	75333670	+	-	1
chr17	chr16	64953323	75334246	+	-	1
chr17	chr16	64953359	75331180	+	-	0
chr17	chr16	64953566	75331697	+	-	0
chr17	chr16	64953663	75334649	+	-	3
chr17	chr16	64953748	75331615	+	-	0
chr17	chr16	64953797	75331703	+	-	1

B

chr4	chr4	26815085	26842112	-	-	0
chr4	chr4	26815191	26842254	-	-	1
chr4	chr4	26815192	26842467	-	-	1
chr4	chr4	26815230	26842681	-	-	3
chr4	chr4	26815339	26845546	-	-	0
chr4	chr4	26815358	26842351	-	-	2
chr4	chr4	26815419	26845459	-	-	4
chr4	chr4	26815728	26853344	+	+	3
chr4	chr4	26815952	26850009	+	+	0
chr4	chr4	26816105	26853014	+	+	0
chr4	chr4	26816192	26853502	+	+	0
chr4	chr4	26816399	26852750	+	+	2
chr4	chr4	26816441	26852188	+	+	0
chr4	chr4	26816494	26852930	+	+	2
chr4	chr4	26816506	26849525	+	+	1

Figure 3. Distribution of chimeric reads-pairs supporting an inter- and intra-chromosomal rearrangement
Distribution of read-pairs supporting a *cis*-joining type of translocation t(16;17) (q23.1;q24.2) in NA18612 shown in figure (A) and a *trans*-joining type of inversion inv(14) (p15.2) in HG00698 in figure (B). All the sequencing data with mate-pair library were downloaded from The European Nucleotide Archive (ENA) with BioProject accession PRJEB11005. In figure (A), each column represents the derivative chromosomes, locations of read 1 from derivative chromosome 1 and read 2 from derivative chromosome 2, aligned orientations of read 1 and read 2 and the number of mismatches in this read-pair. Two derivative chromosomes were separately by the different aligned orientations shown by a red line. In figure (B), each column represents the affected chromosome, locations of read 1 and read 2 with distance larger than expected (10 kb), aligned orientations of read 1 and read 2 and the total number of mismatch in each read-pair. The two independent linkages were separately by the different aligned orientations also shown by a red line. The detailed descriptions were discussed in our pilot study (Dong et al., 2014).