CLINICAL GENETICS WILEY

Evaluation of *SHOX* defects in the era of next-generation sequencing

Mariana F.A. Funari[1] | **Juliana S. de Barros**[2] | **Lucas S. Santana**[3] |
Antonio M. Lerario[1,3,4] | **Bruna L. Freire**[1] | **Thais K. Homma**[1,3] |
Gabriela A. Vasques[1,3] | **Berenice B. Mendonca**[1] | **Mirian Y. Nishi**[1] |
Alexander A.L. Jorge[1,3]

[1]Unidade de Endocrinologia do Desenvolvimento, Laboratório de Hormônios e Genética Molecular/LIM42, Hospital das Clínicas, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

[2]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil

[3]Unidade de Endocrinologia Genética/LIM25, Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil

[4]Department of Internal Medicine, Division of Metabolism, Endocrinology and Diabetes, University of Michigan, Ann Arbor, Michigan

Correspondence
Alexander A. L. Jorge, Av. Dr. Arnaldo, 455, Sao Paulo-SP, CEP 01246-903, Brazil.
Email: alexj@usp.br

Peer Review
The peer review history for this article is available at https://publons.com/publon/10.1111/cge.13587/

Abstract

Short stature homeobox (*SHOX*) haploinsufficiency is a frequent cause of short stature. Despite advances in sequencing technologies, the identification of *SHOX* mutations continues to be performed using standard methods, including multiplex ligation-dependent probe amplification (MLPA) followed by Sanger sequencing. We designed a targeted panel of genes associated with growth impairment, including *SHOX* genomic and enhancer regions, to improve the resolution of next-generation sequencing for *SHOX* analysis. We used two software packages, CONTRA and Nexus Copy Number, in addition to visual analysis to investigate the presence of copy number variants (CNVs). We evaluated 15 patients with previously known *SHOX* defects, including point mutations, deletions and a duplication, and 77 patients with idiopathic short stature (ISS). The panel was able to confirm all known defects in the validation analysis. During the prospective evaluation, we identified two new partial *SHOX* deletions (one detected only by visual analysis), including an intragenic deletion not detected by MLPA. Additionally, we were able to determine the breakpoints in four cases. Our results show that the designed panel can be used for the molecular investigation of patients with ISS, and it may even detect CNVs in *SHOX* and its enhancers, which may be present in a significant fraction of patients.

KEYWORDS
copy number variants, growth disorder, next-generation sequencing, short stature, *SHOX* gene, targeted panel sequencing

1 | INTRODUCTION

The short stature homeobox (*SHOX*) gene, located inside the pseudoautosomal region 1 of sex chromosomes, encodes a transcriptional activator essential for growth plate physiology. *SHOX* haploinsufficiency is considered one of the main monogenic causes of short stature and is a well-established indication for growth hormone therapy to improve adult height.[1,2] Approximately 2% to 15% of children classified as idiopathic short stature (ISS) and 70% to 90% of patients with Léri-Weill dyschondrosteosis (LWD) have *SHOX* haploinsufficiency.[1] Despite recent advances in sequencing technologies, the molecular investigation of *SHOX* mutations continues to be

performed by standard methods, such as multiplex ligation-dependent probe amplification (MLPA) analysis followed by gene sequencing.[3]

The massively parallel sequencing technologies have increased the diagnostic yields in several genetic disorders. The molecular investigation of children with short stature of unknown cause has followed this same trend, with the simultaneous analysis of multiple genes using exome or targeted panel strategies.[4,5] Considering this scenario, the study of *SHOX* is a major challenge because of the large number of copy number variants (CNVs) and the importance of investigating regulatory regions.[6]

We designed a targeted panel for next-generation sequencing (NGS) to evaluate a cohort of patients with ISS. One of the main challenges of this panel would be identifying *SHOX* defects. For this reason, we included the entire genomic *SHOX* region and its enhancers to improve the resolution of NGS for the analysis of this gene. First, we validated this panel by analyzing samples from patients with known *SHOX* CNVs and point mutations. Additionally, we prospectively screened *SHOX* defects in a cohort of patients (n = 77) with ISS.

2 | MATERIALS AND METHODS

2.1 | Patients

This study was approved by the local Ethics Committee. Patients or guardians gave their informed written consent. Fifteen patients with known *SHOX* defects were selected to evaluate the efficiency of the panel in identifying those defects: two patients with point mutations (p.Tyr35* and p.Arg147His), nine with different sized heterozygous deletions involving the gene, one with an intragenic duplication, and three with deletions located downstream of *SHOX* (Table 1, Figure S1). Another 77 ISS patients were included for a prospective genetic evaluation. None of these patients had apparent skeletal disease or evidence of LWD. Thirty-one (31/77; 40%) patients had had a normal *SHOX* evaluation by MLPA and Sanger sequencing.

2.2 | Molecular genetic analysis

Genomic DNA of the patients was analyzed by a customized panel of targeted sequencing based on the Agilent SureSelect[XT] capture system. In this panel, we included the entire genomic region of *SHOX* and selected up- and downstream enhancer regions. DNA libraries were sequenced using Illumina platform.

The raw data were aligned to the reference genome (GRCh37/hg19) with BWA tools. Variant calling was performed with Freebayes and annotated with ANNOVAR. The variants were filtered as previously described.[5] CNV analyses were performed using two software packages able to call CNVs for target region based on the normalized depth of coverage: COpy Number Targeted Resequencing Analysis (CONTRA)[7] and Nexus Copy Number (BioDiscovery, Inc.).[8] We also visually inspected *SHOX* coverage using Integrative Genomics Viewer (IGV) software.

MLPA analysis was carried out using the commercial kit P018-*SHOX*-G1 (MRC Holland). Sanger sequencing was used to confirm and determine the exact breakpoints observed by panel sequencing.

More details of the Methods are available in the Appendix S1 and Table S1.

3 | RESULTS

The mean coverage depths for the genomic and *SHOX* coding regions were 357× and 387×, respectively. Among 92 samples (validation:15; prospective analysis:77), all coding regions of the main *SHOX* transcript (NM_000451.3) had more than a 20-fold depth of coverage, and 81% of *SHOX* genomic region had more than a 10-fold depth of coverage (Tables S2-S3, Figure S2).

3.1 | Validation

Using the aforementioned association of analyzes, our panel was able to identify all 16 previously known *SHOX* mutations. The variant call format (VCF) files filtration identified both point mutations. CONTRA analyses were able to detect 13 (13/14; 93%) known CNVs, and Nexus detected 12 (12/14, 86%). CONTRA missed the intragenic duplication (Case 12, Figure 1). Nexus was able to detect this copy number gain (Figure 1); however, it did not identify the deletion located downstream of *SHOX* in Cases 5 and 15, both detected by CONTRA (Figure S3).

Through direct visualization of the *SHOX* region, we confirmed all 14 CNVs. IGV visualization also enabled the identification of deletion breakpoints in some cases (Figure 2): Case 11 had a previously reported intragenic deletion involving exons 4, 5, and 6a; however, its breakpoints had not been defined.[9] After the IGV visualization, we designed specific primers and sequenced the region across the deletion breakpoints. The deletion was defined in 11 722 bp, with 4 bp overlapping in both breakpoints. We could also determine the exact breakpoints of two probands with deletions in the downstream enhancer region (Cases 13 and 14). They have the same 47 543 bp recurrent deletion previously identified in several individuals with ISS and LWD.[10]

3.2 | Prospective evaluation

We did not identify pathogenic point mutations or small indels in *SHOX* in our cohort of patients with ISS. However, we identified two (2/77; 2.6%) new partial *SHOX* deletions. The first one is a new intragenic deletion encompassing exon 1 identified in a patient with disproportionate short stature that has not been previously screened for *SHOX* mutations (Case 16). This deletion was detected by both CNV software and was also identified in the patients' mother, who also exhibited body disproportion.

The panel further identified another intragenic deletion involving the last 20 nucleotides from exon 3 and part of intron 3 (Figure 2) in a patient with disproportionate short stature in whose previous *SHOX* study had been negative (Case 17). This deletion was identified by the visual inspection of *SHOX* but was missed in CONTRA and Nexus analyses. The deletion was defined in 679 bp; it segregated with the phenotype in the family and was also identified in the patient's father and sister, both with altered body proportions.

TABLE 1 Results of the CNV analyses of *SHOX* gene region of 15 validation patients and the two patients with deletions identified prospectively

| Case | Sex | Cohort | Mutation | MLPA | Predicted deletion interval[a] | | CONTRA | Nexus | IGV | Sanger |
					Minimum	Maximum				
1	F	Validation	Point mutation	−	−	−	−	−	+	c.440G > A (p.Arg147His)
2	M	Validation	Point mutation	−	−	−	−	−	+	c.105C > A (p.Tyr35*)
3	F	Validation	Large deletion	+	6.9 Mb	8.2 Mb[b]	+	+	+	NP
4	F	Validation	Large deletion	+	1.5 Mb	2.8 Mb[b]	+	+	+	NP
5	F	Validation	Large deletion	+	1.5 Mb	5.9 Mb[b]	+	+	+	NP
			Downstream deletion	+	113.2 Kb	194.4 Kb	+	−	+	NP
6	M	Validation	Large deletion	+	1.2 Mb	1.5 Mb[b]	+	+	+	NP
7	M	Validation	Large deletion	+	1.2 Mb	1.5 Mb[b]	+	+	+	NP
8	F	Validation	Large deletion	+	652.0 Kb	1.1 Mb[b]	+	+	+	NP
9	M	Validation	Large deletion	+	243.4 Kb	265.4 Kb	+	+	+	NP
10	M	Validation	Large deletion	+	444.2 Kb	501.8 Kb[b]	+	+	+	NP
11	M	Validation	Intragenic deletion	+	3.8 Kb	16.3 Kb	+	+	+	11 722 bp; ChrX:596149-607 872[c]
12	F	Validation	Intragenic duplication	+	15.0 Kb	294.1 Kb	−	+	+	NP
13	F	Validation	Downstream deletion	+	36.6 Kb	76.8 Kb	+	+	+	47 543 bp; ChrX:780549-828 093[c]
14	F	Validation	Downstream deletion	+	36.6 Kb	76.8 Kb	+	+	+	47 543 bp; ChrX:780549-828 093[c]
15	F	Validation	Downstream deletion	+	113.2 Kb	194.4 Kb	+	−	+	NP
16	F	Prospective analysis	Intragenic deletion	+	64 bp	11.3 Kb	+	+	+	NP
17	F	Prospective analysis	Intragenic deletion	−	−	−	−	−	+	679 bp; ChrX:595541-596-221[c]

Abbreviations: bp: base pair; CNV, copy number variant; F, female; IGV, Integrative Genomics Viewer, Kb: kilobases; MLPA, multiplex ligation-dependent probe amplification; M, male; Mb: megabases; +, positive result (the methodology was able to identify the CNV); −, negative result (the methodology was not able to identify the CNV). NP: not performed.

[a]Minimum and maximum approximated deletion interval predicted by MLPA results. MLPA was performed with commercial kits available at the time (P018-*SHOX*-C1, D1, E1, G1) (MRC Holland, Amsterdam, Netherlands).

[b]This maximum interval must be larger because there is no conserved probe upstream to the first deleted probe.

[c]Size of the deletions and breakpoint positions determined by Sanger sequencing of the cases that had their breakpoints identified. The approximate coordinates are according to chromosome X, GRCh37/hg19.

4 | DISCUSSION

The molecular genetic investigation of children with growth disorders has gradually gained importance. Because *SHOX* defects have a relatively high frequency and have important clinical implications, proper analysis of this gene is essential in this era of multi-gene sequencing technology. Currently, the analysis of *SHOX* is still performed by MLPA followed by Sanger sequencing of the coding region.[1] This candidate gene approach is feasible for a small number of patients with a high probability of harboring a *SHOX* defect. However, patients with *SHOX* defects frequently lack specific phenotype, and the use of NGS is a potential strategy to investigate a large number of patients classified as ISS.[4,5]

We evaluated the performance of a targeted panel that can be useful for ISS diagnosis without the need for the previous implementation of MLPA for *SHOX* analysis. Our panel presented good coverage of *SHOX* coding and enhancer regions. The data analysis with CONTRA and Nexus software combined with the visual inspection of *SHOX* resulted in an efficiency of CNV detection similar to those reached by standard methods. The panel was able to detect all previously known *SHOX* mutations (two point mutations and 14 CNVs), including deletions located downstream of *SHOX* in the enhancer region.[10] It was also able to show two novel partial *SHOX* deletions in the prospective analysis, including one deletion (Case 17) previously missed in MLPA analysis.[11]
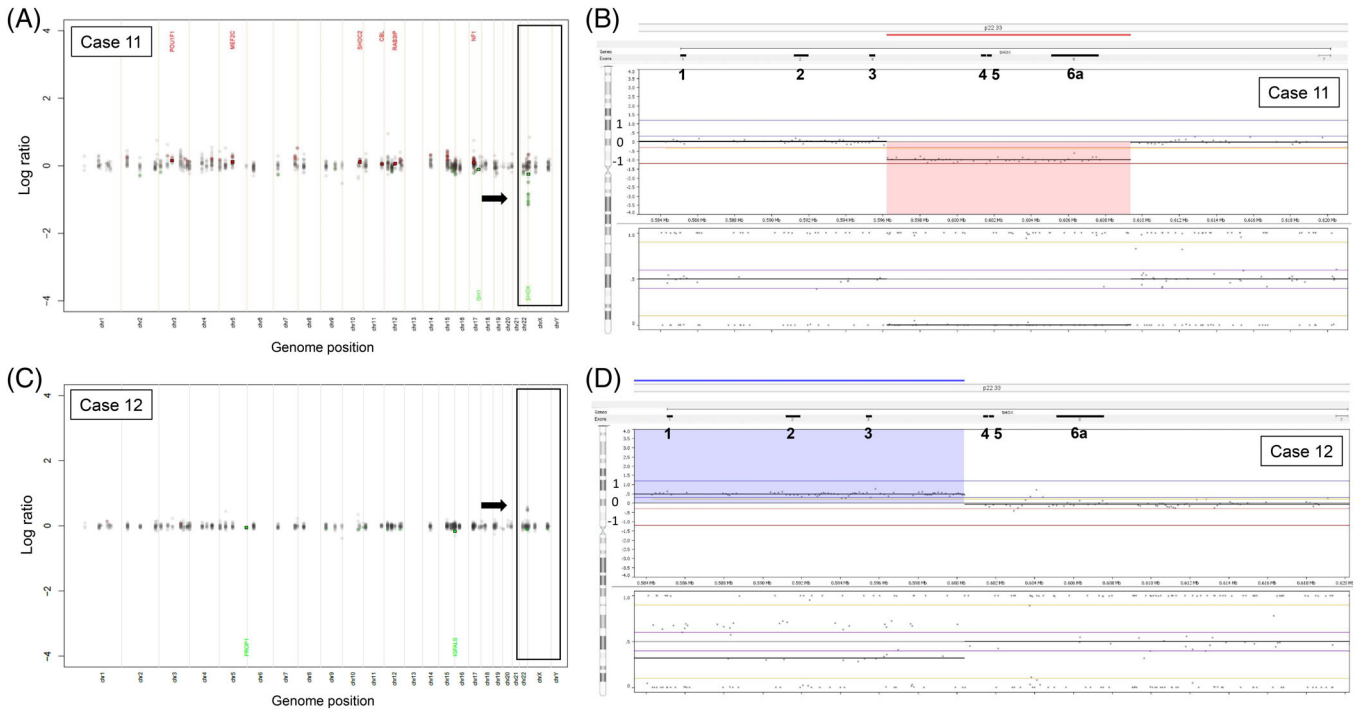
FIGURE 1 CONTRA and Nexus analyses of Cases 11 and 12. CONTRA A, and Nexus B, plots of Case 11, which has a heterozygous deletion involving exons 4, 5 and 6a of *SHOX*. The plots show dots with log ratios near −1.0 in *SHOX* region, indicating the deletion. In CONTRA plot (A), the dots are indicated by the black arrow, while in Nexus (B), they are inside the red region. CONTRA (C) and Nexus (D) plots of Case 12, which has a duplication of exons 1, 2 and 3 in *SHOX*. The CONTRA plot (C) in this case shows dots suggestive of a duplication; however, the log ratios of these regions were not higher than +0.7 (used as filtration criteria). The Nexus plot (D), in turn, identified the duplication, showing dots with log ratios near +1.0 in regions that correspond to these exons (blue region). The black bars and numbers located at the top of Nexus plots (B, D) indicate the *SHOX* exons [Colour figure can be viewed at wileyonlinelibrary.com]
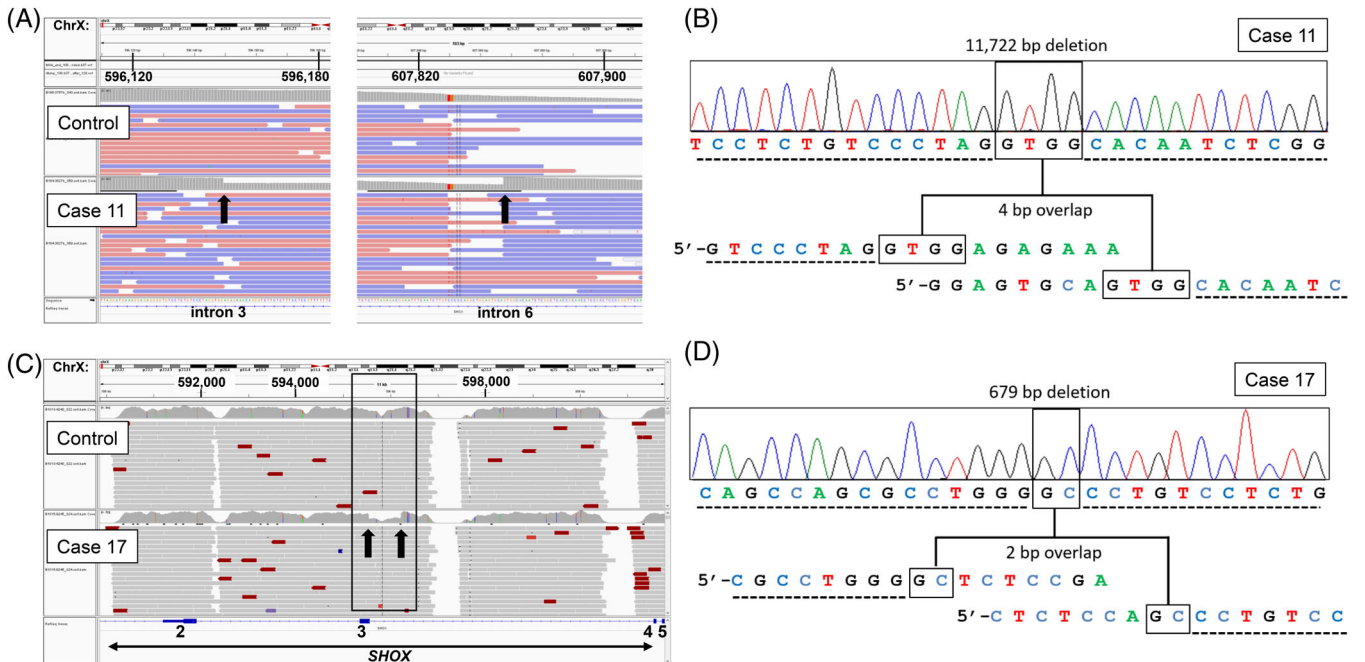
FIGURE 2 Integrative Genomics Viewer (IGV) and Sanger sequencing of breakpoint regions in Cases 11 (A, B) and 17 (C, D), both with intragenic deletions. In IGV images, the black arrows indicate the breakpoints (left arrow indicates the 5′ breakpoints, and right arrow, the 3′). The bars and numbers located at the bottom of IGV image (C) indicate *SHOX* exons. Coordinates are according to chromosome X (GRCh37/hg19). Sanger sequencing (B, D) defined the exact size of the deletions [Colour figure can be viewed at wileyonlinelibrary.com]

CONTRA software presented a sensitivity of 87.5% (14/16) in the detection of CNVs. A sensitivity of 68% was described by Li et al for CONTRA for deletions of 50 to 200 bp and of 96.4% for full exon deletions.[7] The sensitivity of Nexus was 81.2% (13/16). The estimated sensitivity for the exome-based discovery of rare CNVs containing three or more exons is approximately 76%.[12] Based on this assumption, CONTRA and Nexus analyses proved to be satisfactory in detecting CNVs. We believe that the pipeline of CNV analyses could be improved by the inclusion of other strategies beyond the depth of coverage, such as read-pair and split-read approaches.[13]

Visual inspection was essential for the identification of some CNVs and for the characterization of breakpoints. Although it does not seem to be the ideal method for the analysis of large numbers of samples, it can be useful until we have more sensitive software for the detection of CNVs available. We also believe that the inclusion of larger regions and other *SHOX* neighboring genes in next versions of the panel may improve the analysis of the CNVs, helping to estimate their range sizes.

Our results show that this customized targeted panel associated with automated CNV analysis and the direct visualization of *SHOX* sequencing is efficient for diagnosing *SHOX* defects. Using a robust panel that aimed to capture *SHOX* genomic and enhancer regions, we improved the resolution of NGS for the study of *SHOX*. This NGS targeted panel can be useful in the molecular investigation of patients with ISS, allowing the simultaneous analysis of several genes involved in growth impairment without missing *SHOX* CNVs.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Mariana F.A. Funari https://orcid.org/0000-0003-4316-4999
Alexander A.L. Jorge https://orcid.org/0000-0003-2567-7360

REFERENCES

1. Marchini A, Ogata T, Rappold GA. A Track Record on SHOX: From Basic Research to Complex Models and Therapy. *Endocr Rev*. 2016;37 (4):417-448.
2. Blum WF, Ross JL, Zimmermann AG, et al. GH treatment to final height produces similar height gains in patients with SHOX deficiency and Turner syndrome: results of a multicenter trial. *J Clin Endocrinol Metab*. 2013;98(8):E1383-E1392.
3. Binder G, Rappold GA. SHOX Deficiency Disorders. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *Genetic Reviews®*. Seattle (WA) University of Washington, Seattle; 1993–2019; 2015 [updated 2018 Jun 28].
4. Hauer NN, Popp B, Schoeller E, et al. Clinical relevance of systematic phenotyping and exome sequencing in patients with short stature. *Genet Med*. 2018;20(6):630-638.
5. Freire BL, Homma TK, Funari MFA, et al. Multigene sequencing analysis of children born small for gestational age with isolated short stature. *J Clin Endocrinol Metab*. 2019;104:2023-2030.
6. Benito-Sanz S, Belinchon-Martínez A, Aza-Carmona M, et al. Identification of 15 novel partial SHOX deletions and 13 partial duplications, and a review of the literature reveals intron 3 to be a hotspot region. *J Hum Genet*. 2017;62(2):229-234.
7. Li J, Lupat R, Amarasinghe KC, et al. CONTRA: copy number analysis for targeted resequencing. *Bioinformatics*. 2012;28(10):1307-1313.
8. Darvishi K. Application of Nexus copy number software for CNV detection and analysis. *Curr Protoc Hum Genet*. 2010;Chapter 4;65(1):4.14.1-4.14.28. https://doi.org/10.1002/0471142905.hg0414s65.
9. Funari MFA, Jorge AAL, Pinto EM, Arnhold IJP, Mendonca BB, Nishi MY. Cryptic intragenic deletion of the SHOX gene in a family with Léri-Weill dyschondrosteosis detected by Multiplex Ligation-Dependent Probe Amplification (MLPA). *Arq Bras Endocrinol Metabol*. 2008;52(8):1382-1387.
10. Benito-Sanz S, Royo JL, Barroso E, et al. Identification of the first recurrent PAR1 deletion in Léri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. *J Med Genet*. 2012;49(7):442-450.
11. Funari MFA, Jorge AAL, Souza SCAL, et al. Usefulness of MLPA in the detection of SHOX deletions. *Eur J Med Genet*. 2010;53(5): 234-238.
12. Krumm N, Sudmant PH, Ko A, et al. Copy number variation detection and genotyping from exome sequence data. *Genome Res*. 2012;22(8): 1525-1532.
13. Tattini L, D'Aurizio R, Magi A. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Front Bioeng Biotechnol*. 2015;3:92.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.