

Identifying NAHR mechanism between two distinct Alu elements through breakpoint junction mapping by NGS

Gil Monteiro Novo Filho

Universidade de Sao Paulo Faculdade de Medicina

Gleyson Francisco da Silva Carvalho

Universidade de Sao Paulo Faculdade de Medicina

Amom Mendes Nascimento

Universidade de Sao Paulo Faculdade de Medicina

Marilia Moreira Montenegro

Universidade de Sao Paulo Faculdade de Medicina

Julian Gabriel Damasceno

Universidade de Sao Paulo Faculdade de Medicina

Évelin Aline Zanardo

Universidade de Sao Paulo Faculdade de Medicina

Samar Nasser Chehimi

Universidade de Sao Paulo Faculdade de Medicina

Yanca Gasparini Oliveira

Universidade de Sao Paulo Faculdade de Medicina

Alexandre Torchio Dias

Universidade de Sao Paulo Faculdade de Medicina

Chong Ae Kim

Universidade de Sao Paulo Faculdade de Medicina

Leslie Domenici Kulikowski (✉ lesliekulik@usp.br)



Universidade de Sao Paulo Faculdade de Medicina <https://orcid.org/0000-0003-2236-3956>

Short report

Keywords: NAHR, Alu elements, NGS

Posted Date: January 8th, 2020

DOI: <https://doi.org/10.21203/rs.2.20329/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Identifying NAHR mechanism between two distinct *Alu* elements through breakpoint junction mapping by NGS

Gil M Novo-Filho¹, Gleyson F S Carvalho¹, Amom M Nascimento¹, Marilia M Montenegro¹, Jullian G Damasceno¹, Évelin A Zanardo¹, Samar N Chehimi¹, Yanca G Oliveira¹, Alexandre T Dias¹, Chong A Kim², Leslie D Kulikowski¹.

¹Laboratório de Citogenômica, Departamento de Patologia, LIM 03 Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, SP, BR.

*Av. Dr. Enéas Carvalho de Aguiar, 155 - 2º andar – Bloco 12
Cerqueira César, São Paulo – SP
01246-100*

²Unidade de Genética, Departamento de Pediatria, Instituto da Criança, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, BR

*Av. Dr. Enéas Carvalho de Aguiar, 647 - 7º andar
Cerqueira César, São Paulo - SP
05403-000*

Corresponding author

*Leslie Domenici Kulikowski, Ph.D., D.Sc.
Coordinator of the Laboratory of Cytogenomics and Molecular Pathology
Department of Pathology, Faculty of Medicine of the University of São Paulo,
Brazil
Phone: +55 11 26619506*

Abstract

Background:

Genomic rearrangements encompass deletions, duplications, inversions, insertions and translocations and may be the cause of several genetic diseases. One of the most frequent mechanisms that generates these rearrangements is the Non-Allelic Homologous Recombination (NAHR). They are caused by a misalignment between regions of high level of similarity, like Low Copy Repeats (LCRs) and *Alu* sequences. We aimed to sequence the breakpoint of a patient with a single deletion on chromosome 22q13.2 in order to understand the genomic structure of the region involved as well as elucidate the mechanism behind this rearrangement. Investigating breakpoints are of the utmost importance in the understanding the influence of the genomic architecture in clinical assays.

Results:

We flanked the breakpoint detected by array and then we captured the regions using Illumina Nextera Rapid Capture Custom to sequence with Illumina MiSeq. We found a chimeric read on Chr22:41,026,090, setting a 624,688 bp deletion on Chr22:41,026,112-41,650,780 (hg19).

This deletion merges the intronic region of *MKL1* and *RANGAP1* genes, on two different *Alu* sequences (*AluSx* and *AluY*, respectively).

Conclusions: The sequence of the breakpoint reveals that *Alu* elements are an important characteristic of the human genome on generating rearrangements.

Keywords: NAHR, *Alu* elements, NGS

Findings

Genomic rearrangements are structural changes on DNA molecule, whose length can range from hundreds to millions base pairs [1]. These changes encompass deletions, duplications, inversions, insertions and translocations and are responsible for great part of differences between human genomes. Furthermore, they may be the cause of several genetic diseases [1,2,3]. Furthermore, they may be the cause of several genetic diseases. These rearrangements can be simple, with only one breakpoint, or complex, with simple rearrangements of 2 or more breakpoints [4,5,6].

The mechanisms that cause these rearrangements are classified as replicative and non-replicative. Moreover, the human genome features can result in a susceptibility of a specific region to rearrangements, leading to a genomic instability [5,6].

There recurrent rearrangements are caused mainly by low copy repeats (LCR) regions [7,8]. The most frequent recurrent mechanism known is the non-allelic homologous recombination (NAHR). NAHR mainly occur due to the presence of LCRs. These sequences, because of their high similarity, may eventually misalign during either mitosis or meiosis, between either homologous or non-homologous chromosomes. These misalignment leads to deletions, duplications, inversions and translocations [1,9].

Additionally to LCR regions, other repetitive elements can lead to NAHR, as short interspersed nuclear elements (SINEs) [10]. SINEs are retrotransposons, sequences that amplify themselves throughout the genome through RNA intermediates, originated probably due to parasitic DNA [11,12]. Alu elements are the most frequent SINE known on the human genome and they are present only in primates [13, 14]. They have ~300 bp long comprehending more than 10% of the human genome, being classified in several subfamilies [14].

Searching for a breakpoint formation mechanism demands to sequence the breakpoint at base pair level. Therefore, it is possible detect its exact position, analyze the genomic architecture of the loci involved and identify the nucleotides sequence of the breakpoint

[15,16,17,18,19]. Only with this information, it is possible to understand the molecular event that triggered the rearrangement.

Thus, this study aimed to sequence the breakpoint of a patient with a single deletion on chromosome 22q13.2 in order to understand the genomic structure of the region involved as well as elucidate the mechanism behind this rearrangement.

This patient presented a SNP array (Illumina Cyto-SNP-12) result with a deletion encompassing chr22:41,032,863-41,640,297 (hg19). We aligned at the breakpoint there is a read searched for chimeric reads in which we could find the breakpoint and the junction sequence that connect the two loci. With this approach, we find the exact coordinate of the entire deletion: Chr22:41,026,112-41,650,780 (hg19), which sets a 624,688 bp deletion.

The deletion occurred in introns of two distinct genes, merging the MKL1 and RANGAP1 genes. The investigation of the region revealed that both locus are part of two distinct SINEs Alu sequences. Chr22:41,026,112 region is inside an AluSx sequence, and Chr22:41,650,780 region is inside an AluY sequence. The analysis of both SINEs showed a high level of similarity of both sequences. At the breakpoint there is a region of total homology and, subsequently, the sequence of the strand after the deletion (Figure 1).

The sequence shows two regions with a high similarity level and that the breakpoint occurs without micro-homology, although there is some homology after the breakpoint. After this point, the strand follows the other sequence.

| | | | |
|----------------------------|---------------------------|--------------|----------------|
| Reference Chr22:41,026,090 | ...GCGGATCACCTGAGGTCAGGAG | TTT GAGACCAG | CCTGGCCAACA... |
| Read Chr22:41,026,090 | ...GCGGATCACCTGAGGTCAGGAG | ATC GAGACCAG | CCTGGCTAACG... |
| Reference Chr22:41,650,751 | ...TGGGTGGATCATGAGGTCAGGA | ATC GAGACCAT | CCTGGCTAACG... |

Figure 1: Sequence of the breakpoint: on the first and last lines are the hg19 reference sequence of the designated regions. On the second line is the sequence of the patient found by breakpoint sequencing. Highlighted in yellow the breakpoint.

Using a previous array result to determine a region to be sequenced is an effective and inexpensive way to flank a breakpoint. Since arrays are comparative tests that aim to assess copy number imbalances in the genome and uses a logarithmic scale to reckon gains and loss regions, it is not very precise in determine a breakpoint [20,21,22]. The same DNA sample that undergoes for the same test can present different breakpoint positions. To solve this limitation of the technique, we considered a range of three SNP array probes that preceded and three SNP array probes that succeeded each breakpoint region previously detected. This approach proved to be a success, since we were able to directly reach and sequence the breakpoint.

Sanger sequencing of this region would be a challenge, it would demand a great number of primers for several regions of repetitive elements. Once the whole regions to be sequenced encompass both genic and intergenic sequences, it was required that we built a customized NGS panel. However, a panel aiming to sequence a small region would be unviable. Thus, we took advantage to insert this region in other custom panel, drastically reducing the expense of the assay.

Since we reach the exact breakpoint, we could assess the region involved. The fact that the breakpoint is inserted in two different Alu and the absence of micro-homology inside the breakpoint, lead us to consider NAHR mechanism driven by Alu. AluY is a subfamily of the most recent Alu elements in the human genome, and the only category that encompass active elements, that is, Alu elements that still replicates themselves throughout the genome. AluSx belongs to the AluS subfamily, which is not active anymore.

The fact of the high similarity between these two Alu elements can be explain by the fact that AluY subfamily is derived of the AluS subfamily, and it was relocated to a different type of subfamily because of its activity [20,23,24,25,26,27,28,29]. It is the first time that a fusion between *MKL1* and *RANGAP1* genes are sequenced.

MKL1 gene is known to cause leukemia due to a fusion with *RBM15* [30,31]. However, its fusion with *RANGAP1* gene remains unclear. None of these two genes are related with any phenotypic

features of the patient and it is unlikely that the fusion of these genes were the responsible for these characteristics.

The phenotype is better explained by the deletion of *EP300* gene, causing the rare autosomal dominant Rubinstein-Taybi Syndrome 2, a mild form of Rubinstein-Taybi Syndrome, which fits with the clinical signs observed [32,33].

Flanking breakpoints by using array platform to capture and sequencing by using NGS platform is an inexpensive and effective way to reveal the sequence of a breakpoint junction. Alu elements are an important SINE regarding recurrent non replicative mechanisms, like NAHR, for genomic imbalances on the human genome.

Delineate and study breakpoints is of the utmost importance in the understanding of the influence of the genomic architecture in clinical assays.

Methods

We assessed the DNA sample of a male patient, 20 years old, double aortic arch, esophageal stricture, seizures, inguinal hernia DI, cryptorchidism, eyelid crack down, speech disorder, polydactyly, microcephaly, floating nose. Karyotype and FISH analysis for 22q13.2 were normal. Array study revealed a 607 kb deletion in chromosome 22q13.2 (Illumina Cyto-SNP-12). In order to precisely define the region of the breakpoint, we reanalyzed the result on BlueFuse Multi software. On chr22:41,640,297 locus, we defined the 3 probes that preceded the last probe identified as deleted and three probes that succeeded the first probe as not deleted. Then we reached two regions for sequencing: Chr22:41,025,934-41,036,329 and Chr22:41,640,297-41,651,034.

In order to capture this region, we designed the probes for Nextera Rapid Capture Custom on Illumina Design Studio. The experiments were performed according to the manufacturer instructions and sequenced on Illumina MiSeq.

DNA sequence was aligned by Agilent Sure Call and the bam file was analyzed by IGV.

Availability of data and materials

All data in this article is included in the published work.

Abbreviations

LINE-1: Long Interspersed Element 1

LCR: Low copy repeats

NAHR: Non-Allelic Homologous Recombination

SINEs: short interspersed nuclear elements

References

1. Gu, W., Zhang, F., & Lupski, J. R. (2008). Mechanisms for human genomic rearrangements. *Pathogenetics*, 1(1), 4. doi: 10.1186/1755-8417-
2. Chen, J. M., Cooper, D. N., Ferec, C., Kehrer-Sawatzki, H., & Patrinos, G. P. (2010). Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol*, 20(4), 222-233. doi: 10.1016/j.semcancer.2010.05.007
3. Onishi-Seebacher, M., & Korb, J. O. (2011). Challenges in studying genomic structural variant formation mechanisms: the short-read dilemma and beyond. *Bioessays*, 33(11), 840-850. doi: 10.1002/bies.201100075
4. Liu, P., Carvalho, C. M., Hastings, P. J., & Lupski, J. R. (2012). Mechanisms for recurrent and complex human genomic rearrangements. *Curr Opin Genet Dev*, 22(3), 211-220. doi: 10.1016/j.gde.2012.02.012
5. Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nat Rev Genet*, 10(8), 551-564. doi: 10.1038/nrg2593

6. Lupski, J. R., & Stankiewicz, P. (2005). Genomic disorders: molecular mechanisms for rearrangements and conveyed phenotypes. *PLoS Genet*, 1(6), e49. doi: 10.1371/journal.pgen.0010049
7. Arlt, M. F., Wilson, T. E., & Glover, T. W. (2012). Replication stress and mechanisms of CNV formation. *Curr Opin Genet Dev*, 22(3), 204-210. doi: 10.1016/j.gde.2012.01.009
8. Stankiewicz, P., & Lupski, J. R. (2002). Genome architecture, rearrangements and genomic disorders. *Trends Genet*, 18(2), 74-82.
9. Verdin, H., D'Haene, B., Beysen, D., Novikova, Y., Menten, B., Sante, T., . . . De Baere, E. (2013). Microhomology-mediated mechanisms underlie non-recurrent disease-causing microdeletions of the FOXL2 gene or its regulatory domain. *PLoS Genet*, 9(3), e1003358. doi: 10.1371/journal.pgen.1003358
10. Lee, J. A., Carvalho, C. M., & Lupski, J. R. (2007). A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell*, 131(7), 1235-1247. doi: 10.1016/j.cell.2007.11.037
11. Roy-Engel, A. M., Carroll, M. L., Vogel, E., Garber, R. K., Nguyen, S. V., Salem, A. H., . . . Deininger, P. L. (2001). Alu insertion polymorphisms for the study of human genomic diversity. *Genetics*, 159(1), 279-290
12. Schmid, C. W., & Deininger, P. L. (1975). Sequence organization of the human genome. *Cell*, 6(3), 345-358.
13. Jurka, J., & Smith, T. (1988). A fundamental division in the Alu family of repeated sequences. *Proc Natl Acad Sci U S A*, 85(13), 4775-4778.

14. Lebedev, Y. B., Amosova, A. L., Mamedov, I. Z., Fisunov, G. Y., & Sverdlov, E. D. (2007). Most recent AluY insertions in human gene introns reduce the content of the primary transcripts in a cell type specific manner. *Gene*, 390(1-2), 122-129. doi: 10.1016/j.gene.2006.09.031
15. Chen, J. M., Cooper, D. N., Ferec, C., Kehrer-Sawatzki, H., & Patrinos, G. P. (2010). Genomic rearrangements in inherited disease and cancer. *Semin Cancer Biol*, 20(4), 222-233. doi: 10.1016/j.semcancer.2010.05.007
16. Hermetz, K. E., Newman, S., Conneely, K. N., Martin, C. L., Ballif, B. C., Shaffer, L. G., . . . Rudd, M. K. (2014). Large inverted duplications in the human genome form via a fold-back mechanism. *PLoS Genet*, 10(1), e1004139. doi: 10.1371/journal.pgen.1004139
17. Schluth-Bolard, C., Labalme, A., Cordier, M. P., Till, M., Nadeau, G., Tevissen, H., . . . Sanlaville, D. (2013). Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. *J Med Genet*, 50(3), 144-150. doi: 10.1136/jmedgenet-2012-101351
18. Sobreira, N. L., Gnanakkan, V., Walsh, M., Marosy, B., Wohler, E., Thomas, G., . . . Valle, D. (2011). Characterization of complex chromosomal rearrangements by targeted capture and next-generation sequencing. *Genome Res*, 21(10), 1720-1727. doi: 10.1101/gr.122986.111
19. Utami, K. H., Hillmer, A. M., Aksoy, I., Chew, E. G., Teo, A. S., Zhang, Z., . . . Cacheux, V. (2014). Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. *PLoS One*, 9(6), e90852. doi: 10.1371/journal.pone.0090852

20. Manning, M., Hudgins, L., Professional, P., & Guidelines, C. (2010). Array-based technology and recommendations for utilization in medical genetics practice for detection of chromosomal abnormalities. *Genet Med*, 12(11), 742-745. doi: 10.1097/GIM.0b013e3181f8baad
21. Schaaf, C. P., Wiszniewska, J., & Beaudet, A. L. (2011). Copy number and SNP arrays in clinical diagnostics. *Annu Rev Genomics Hum Genet*, 12, 25-51. doi: 10.1146/annurev-genom-092010-110715
22. Bennett, E. A., Keller, H., Mills, R. E., Schmidt, S., Moran, J. V., Weichenrieder, O., & Devine, S. E. (2008). Active Alu retrotransposons in the human genome. *Genome Res*, 18(12), 1875-1883. doi: 10.1101/gr.081737.108
23. Costa, E., Duque, F., Oliveira, J., Garcia, P., Goncalves, I., Diogo, L., & Santos, R. (2007). Identification of a novel AluSx-mediated deletion of exon 3 in the SBDS gene in a patient with Shwachman-Diamond syndrome. *Blood Cells Mol Dis*, 39(1), 96-101. doi: 10.1016/j.bcmd.2007.02.002
24. Elliott, B., Richardson, C., & Jasin, M. (2005). Chromosomal translocation mechanisms at intronic alu elements in mammalian cells. *Mol Cell*, 17(6), 885-894. doi: 10.1016/j.molcel.2005.02.028
25. Jjingo, D., Conley, A. B., Wang, J., Marino-Ramirez, L., Lunyak, V. V., & Jordan, I. K. (2014). Mammalian-wide interspersed repeat (MIR)-derived enhancers and the regulation of human gene expression. *Mob DNA*, 5, 14. doi: 10.1186/1759-8753-5-14
26. Kurosaki, T., Ueda, S., Ishida, T., Abe, K., Ohno, K., & Matsuura, T. (2012). The unstable CCTG repeat responsible for myotonic dystrophy type 2 originates from an AluSx element insertion into an early primate genome. *PLoS One*, 7(6), e38379. doi: 10.1371/journal.pone.0038379

- 27.Nozu, K., Iijima, K., Ohtsuka, Y., Fu, X. J., Kaito, H., Nakanishi, K., & Vorechovsky, I. (2014). Alport syndrome caused by a COL4A5 deletion and exonization of an adjacent AluY. *Mol Genet Genomic Med*, 2(5), 451-453. doi: 10.1002/mgg3.89
- 28.Roy-Engel, A. M., Carroll, M. L., Vogel, E., Garber, R. K., Nguyen, S. V., Salem, A. H., . . . Deininger, P. L. (2001). Alu insertion polymorphisms for the study of human genomic diversity. *Genetics*, 159(1), 279-290.
- 29.Ma, Z., Morris, S. W., Valentine, V., Li, M., Herbrick, J. A., Cui, X., Hitzler, J. K. (2001). Fusion of two novel genes, RBM15 and MKL1, in the t(1;22)(p13;q13) of acute megakaryoblastic leukemia. *Nat Genet*, 28(3), 220-221. doi: 10.1038/90054
- 30.Mercher, T., Coniat, M. B., Monni, R., Mauchauffe, M., Nguyen Khac, F., Gressin, L., . . . Bernard, O. A. (2001). Involvement of a human gene related to the *Drosophila* spen gene in the recurrent t(1;22) translocation of acute megakaryocytic leukemia. *Proc Natl Acad Sci U S A*, 98(10), 5776-5779. doi: 10.1073/pnas.101001498
- 31.Bartsch, O., Labonte, J., Albrecht, B., Wieczorek, D., Lechno, S., Zechner, U., & Haaf, T. (2010). Two patients with EP300 mutations and facial dysmorphism different from the classic Rubinstein-Taybi syndrome. *Am J Med Genet A*, 152A(1), 181-184. doi: 10.1002/ajmg.a.33153
- 32.Zimmermann, N., Acosta, A. M., Kohlhasse, J., & Bartsch, O. (2007). Confirmation of EP300 gene mutations as a rare cause of Rubinstein-Taybi syndrome. *Eur J Hum Genet*, 15(8), 837-842. doi: 10.1038/sj.ejhg.5201791

Funding

This work was supported by FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) grants number 2016/09452-0 and CNPQ (Conselho Nacional de Desenvolvimento Científico e Tecnológico) grants number 446470/2014-7

Author information

Affiliations

Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina FMUSP, Universidade de São Paulo, São Paulo, SP, BR.

Unidade de Genética, Departamento de Pediatria, Instituto da Criança, Hospital das Clínicas HCFMUSP, Faculdade de Medicina, Universidade de São Paulo, São Paulo, SP, BR

Contributions

Novo-Filho GM and Carvalho GFS performed the experiments and NGS analysis and wrote the paper; Nascimento AM prepared the samples and performed DNA extraction; Zanardo EA, Chehimi SN and Dias AT performed the genomic array; Oliveira YG and Montenegro MM performed classical cytogenetic analysis; Kim CA provided the sample and clinically assessed the patient; Kulikowski LD designed and coordinated the study. All authors read and approved the final manuscript.

Corresponding author

Correspondence to Leslie Domenici Kulikowski

Ethics declarations and Consent for publication

The Research Ethics Committee of the Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo (HC-FMUSP) approved this study, (CAPPesq n° 0718/11) and written informed consent for publication was obtained from the parent of the patient (CAPPesq n° 881.661).

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Not applicable

Figures



Figure 1

Sequence of the breakpoint: on the first and last lines are the hg19 reference sequence of the designated regions. On the second line is the sequence of the patient found by breakpoint sequencing. Highlighted in yellow the breakpoint.