Journal of Genetics (2021)100:2
https://doi.org/10.1007/s12041-020-01250-6

PERSPECTIVES

Homozygosity stretches around homozygous mutations in autosomal recessive disorders: patients from nonconsanguineous Indian families

SHUBHA R. PHADKE*, PRIYANKA SRIVASTAVA, PANKAJ SHARMA, ARCHANA RAI and SUZENA MASIH

Department of Medical Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow 226 014, India

*For correspondence. E-mail: shubharaophadke@gmail.com.

Received 4 May 2020; revised 19 July 2020; accepted 24 August 2020

Abstract. India has a large heterogeneous population with its unique social and genetic characteristics. Tradition of marriage between specific caste groups have produced unique characteristics to the mutation spectrum of genetic disorders and may be a higher prevalence of autosomal recessive (AR) disorders in some communities. We observed that in many nonconsanguineous families with rare autosomal disorders, maternally and paternally inherited mutations are same, indicating common ancestor. In this era of genomic techniques, finding homozygous regions have become easy. It was seen that the patients with AR disorders, who were homozygous for the disease causing pathogenic / likely pathogenic variations, have large stretches (0.6–188 Mb) of homozygosity around the causative sequence variations. SNP microarray data of patients from consanguineous and nonconsanguineous families also showed that even patients from nonconsanguineous families had 3–49 Mb size regions of homozygosity. Long stretches of homozygosity around homozygous rare pathogenic variants in nonconsanguineous families with rare AR disorders supports the notion that these couples may have a common ancestor for more than six generations and the system of marriages between same groups. Hence, using the strategy of homozygosity by descent even in nonconsanguineous families can be fruitful in identifying the novel pathogenic variations and novel genes.

Keywords. India; homozygosity; autosomal recessive inheritance; regions of loss of heterozygosity; consanguinity.

Introduction

In India, the custom of marriage among the same caste groups has been followed for ages. We have seen that for many rare autosomal recessive (AR) disorders, the affected individuals are homozygous for rare disease causing pathogenic variations, suggesting effects of inbreeding (Dalal *et al.* 2012; Bidchol *et al.* 2014; Ankala *et al.* 2015; Bhavani *et al.* 2016; Mandal *et al.* 2016). Hence, we analysed the exome sequencing data of cases with pathogenic / likely pathogenic sequence variations for AR disorders in homozygous regions around the causative sequence variations. Consanguinity may lead to slightly increased possibility of AR disorders in the offspring and is more frequent in families with rare genetic disorders (table 1). In the era of genomic tools, consanguinity is a major asset for diagnostics and research of AR disorders. Presence of causative disease, causing genetic variations in homozygous forms in nonconsanguineous families suggest that the homozygosity-based strategy can be suitable for analysis of next-generation sequencing-based testing even in nonconsanguineous families of AR disorders.

Materials and methods

The published literature on mutation data of Indian patients with rare AR disorders was searched. The Morquio disease, Hurler syndrome (mucopolysaccharidosis I), pycnodysostosis, progressive pseudorheumatoid arthritis of childhood (PPAC) and multinodular osteolysis with nodular arthropathy (MONA) were chosen for assessing the magnitude of homozygous disease-causing sequence variations in probands from nonconsanguineous families. These five disorders were chosen to represent the magnitude of homozygous disease-causing sequence variations in rare AR disorders. The reason behind choosing these five disorders is that these are large series of mutation data in Indian patients with AR disorders.

The data of exome sequencing performed for AR disorders was reanalysed and the regions of homozygosity around

Published online: 09 February 2021

Table 1. Comparison of consanguineous parents of a common disorder, beta thalassaemia major as compared to that of other AR disorders.

	Thalassaemia major (%)	Other AR disorders (%)
Consanguineous	10.7	26.6
Nonconsanguineous	89.3	73.4

the homozygous causative pathogenic sequence variations from consanguineous and nonconsanguineous families were studied for their sizes. Similarly, the data of another 100 cases of SNP microarray (done for patient diagnosis) was analysed. This included 50 cases each from consanguineous and nonconsanguineous families. The regions of homozygosity (ROH) were noted for all. The data were compiled and compared.

Results

In a large series of Morquio disease, 58 of the 68 probands were homozygous, while consanguinity was present only in 40 cases. Similarly, for the pycnodysostosis, 11 of the 12 probands from nonconsanguineous families were homozygous. A similar situation was for progressive pseudorheumatoid arthritis of childhood (PPAC) and MONA, respectively, i.e. eight of nine cases of nonconsanguineous families were homozygous for disease causing variations. The representative data of homozygous pathogenic variations of rare AR disorders in Indian patients is shown in table 2.

The data from exome sequencing performed for AR disorders was serched for identification of pathogenic or likely pathogenic sequence variations. All the 24 cases with AR disorders from consanguineous families were homozygous for the disease-causing pathogenic or likely pathogenic variations (12 of 24 being novel variations) and had large stretches of homozygosity (average, 77.2 Mb; range, 5–271 Mb) around the disease-causing pathogenic or likely pathogenic variations (table 3). For AR disorders from nonconsanguineous families, the disease-causing variations were homozygous in 13 (nine being novel) of 19 cases and six were compound heterozygous. In the cases with homozygous pathogenic variations from nonconsanguineous families, there were stretches of homozygosity around the causative sequence variations (average, 27.9 Mb; range 0.6–188 Mb).

Table 3. Stretches of homozygosity around the homozygous pathogenic variations.

Consanguinity, in families with homozygous pathogenic variation		Pathogenic variation identified		Stretches of homozygosity around the homozygous pathogenic variation	
Consanguinity	Number	Known	Novel	Average (Mb)	Range (Mb)
Yes	24	12	12	77.2	5–271
No	13	4	9	27.9	0.6–188

We also reviewed our data of SNP microarray of cases from 50 consanguineous and 50 nonconsanguineous families (table 4). In cases born to consanguineous parents, the sizes of ROH were 28–770 Mb (average 251.44 Mb, i.e. 8.75% of genome per case; range from 0.97 to 26.78% of total genome). Although the third degree consanguinity is the common form of consanguinity, many families have consanguineous marriages in previous generations as well leading to homozygosity of more than 10% of the genome. In cases born to nonconsanguineous parents, the sizes of ROH regions varied from 3 to 49 Mb (0.10–1.7% of total genome; average 0.74%). Twenty-six cases had runs of homozygosity stretching more than 5 Mb, while 24 cases did not have any region of homozygosity in more than 5 Mb. Although 56% cases had ROH regions more than 5 Mb, eight of 26 had more than 1% genome with ROH (up to 1.7%). It is noteworthy that 17 of 50 probands have ROH more than 0.78% of the genome. ROH of 0.78% correspond to consanguineous marriages of second cousins once

Table 2. Representative data from Indian patients with pathogenic and likely pathogenic sequence variations (mutations) in AR disorders.

AR disorder	Probands with mutations (homozygous mutations)	Consanguineous		Percentage of homozygous mutations in probands from nonconsanguineous families	References
		Yes	No		
Morquio disease - MPS IV A	68 (56)	28 (41.2%)	40	28/40 (70%)	Bidchol *et al.* (2014)
Mucopolysaccharidosis I	30 (25)	19 (63.3%)	11	6/11 (54.6%)	Uttarilli *et al.* (2016)
PPAC	25 (24)	15 (60%)	9	8/9 (88.9%)	Dalal *et al.* (2012)
Pycnodysostosis	22 (21)	10 (45.5%)	12	11/12 (91.7%)	Mandal *et al.* (2016) and present data
MONA	11 (11)	7 (63.6%)	4	4/4 (100%)	Mandal *et al.* (2016)

Table 4. ROH in SNP microarray data of individuals born to consanguineous and nonconsanguineous parents.

Consanguinity (number)	Total runs of homozygosity (range)			Cases with ROH regions of sizes more than 5 Mb	
	Average (Mb)	Percentage of genome*	Total ROH (range) (Mb)	No. of cases	No. of stretches more than 5 Mb
Yes (50)	251.44	8.5% (0.97–26.78%)	28–770	50	11.59 (1–25)
No (50)	21.52	0.74% (0.10–1.7%)	3–49	28	2.33 (1–15)

*Taking genome size as $3x10^9$bp.

removed. Two of the 50 cases from nonconsanguineous families had ROH more than 1.58% of genome, which is equivalent to inbreeding coefficient of marriages between second cousins.

Discussion

Table 2 shows the data of consanguinity and disease-causing sequence variations in series of Indian patients with five rare AR diseases. The percentage of probands from consanguineous families vary from 41.2 to 63.6. Published literature for various AR disorders from India show that even in nonconsanguineous families, the homozygosity is very high. For example, in case of Mucopolysaccharidosis I, 54.6% probands were from nonconsanguineous families (Uttarilli *et al.* 2016) and for MONA, 100% probands were from nonconsanguineous families as reported by Mandal *et al.* 2016. This lead us to look at the exome sequencing data of the cases with rare AR disorders. Of the 43 cases with AR disorders in whom the causative sequence variation was detected by exome sequencing, 24 were from consanguineous families. In 19 probands of nonconsanguineous families, 68.4% (13/19) were homozygous for the disease-causing sequence variation and six were compound heterozygotes. The average length of stretches of homozygosity around the homozygous pathogenic variations in these probands from nonconsanguineous families was 27.9 Mb as against 77.2 Mb in consanguineous families. This suggests that the pathogenic sequence variations inherited from nonconsanguineous parents had also probably originated from a founder ancestor. Likely, this could be explained as the most of the marriages in India are among specific caste groups and this leads to inbreeding. This suggests that even in many nonconsanguineous marriages, the partners share a significant part of their genomes. To look at this issue, we looked at the SNP data of SNP microarray performed in 100 different samples. The ROH in probands from consanguineous and nonconsanguineous families were between 251.44 Mb and 21.52 Mb, respectively. This means that the 0.74% of genome (0.10–1.7%) sharing was present in nonconsanguineous families. Second cousins once removed have 0.78% of their genomes common. In our data, 17 of 50 probands from nonconsanguineous families had ROH more than 22.4 Mb. This is equivalent to 0.78% of the genome and is seen in marriages between second cousins once removed. In two of 50 probands from nonconsanguineous families, ROH region was 1.56%, which is equivalent to inbreeding coefficient of second cousins. Although the possibility of some families being unaware of consanguinity or have not purposefully shared the information cannot be totally ruled out; these results support the presence of inbreeding over generations, does lead to a lot of sharing of genome and homozygosity for mutations inherited from common ancestors.

In conclusion, the study documents long stretches of homozygosity around homozygous rare pathogenic variants in probands from nonconsanguineous families. The SNP microarray data also shows large ROH suggestive of common ancestors even in nonconsanguineous families. It supports the notion that due to the system of marriages between closed groups (castes) many disease causing genetic variations are passed over generations from the common ancestors and manifest in homozygous form in rare AR disorders in nonconsanguineous families in India. This supports the notion that using the strategy of homozygosity by descent even in nonconsanguineous families can be fruitful in identifying the novel pathogenic variations and novel genes.

Acknowledgements

We acknowledge the Indian Council of Medical Research (63/8/2010-BMS), Department of Biotechnology (BT/PR9635/MED/97/198/2013) for their financial support.

References

Ankala A., Tamhankar P. M., Valencia C. A., Rayam K. K., Kumar M. M. and Hegde M. R. 2015 Clinical applications and implications of common and founder mutations in Indian subpopulations. *Hum. Mutat.* **36**, 1–10.

Bhavani G. S., Shah H., Shukla A., Gupta N., Gowrishankar K., Rao A. P. *et al.* 2016 Clinical and mutation profile of multicentric osteolysis nodulosis and arthropathy. *Am. J. Med. Genet. A.* **170A**, 410–417.

Bidchol A. M., Dalal A., Shah H., Nampoothiri S., Kabra M., Gupta N. *et al.* 2014 GALNS mutations in Indian patients with mucopolysaccharidosis IVA. *Am. J. Med. Genet. A.* **164A**, 2793–2801.

Dalal A., Bhavani G. S. L., Togarrati P. P., Bierhals T., Nandineni M. R., Danda S. *et al.* 2012 Analysis of the WISP3 gene in Indian families with progressive pseudorheumatoid dysplasia. *Am. J. Med. Genet. A.* **158A**, 2820–2828.

Mandal K., Ray S., Saxena D., Srivastava P., Moirangthem A., Ranganath P. *et al.* 2016 Pycnodysostosis: mutation spectrum in five unrelated Indian children. *Clin. Dysmorphol*. **25,** 113–120.

Uttarilli A., Ranganath P., Matta D., Md Nurul Jain J., Prasad K., Babu A. S. *et al.* 2016 Identification and characterization of 20 novel pathogenic variants in 60 unrelated Indian patients with mucopolysaccharidoses type I and type II. *Clin. Genet.* **90**, 496–508.

Corresponding editor: S. GANESH