

Original Article

STUDY ON IMPLICATIONS OF COPY NUMBER VARIATIONS (CNVs) IN HUMAN POPULATION

ANUSHKA YADAV¹, POOJA SINGH¹, PRIYA RANJAN KUMAR^{1*}, SARIKA SRIVASTAVA², SANTOSH KUMAR MISHRA¹

¹Department of Biotechnology, IMS Engineering College, NH24, Adhyatmik Nagar, Ghaziabad, UP, ²Department of Biosciences, IMS Ghaziabad (University Courses Campus), NH24, Adhyatmik Nagar, Ghaziabad, UP
Email: priyaranjan.biet07@gmail.com

Received: 08 Apr 2017 Revised and Accepted: 09 May 2017

ABSTRACT

Objective: To investigate the role and implications of copy number variations (CNVs) in different diseases found in the human population using various computational tools and databases.

Methods: Five different diseases were taken into consideration namely Autism, Type-II Diabetes, Rheumatoid Arthritis, Breast Cancer, and Psoriasis. To validate the CNV's associated with various human diseases different tools and databases were used such as CNV annotator, DECIPHER, Database of Genomic Variants (DGV), CNVD, CNV Workshop, CNV Web store. Finally, the results were analysed to identify the extent of CNVs association in selected diseases.

Results: Among all the selected diseases, the maximum numbers of CNV's were found in the case of breast cancer which in total 3851 at chromosome number one. Among all the selected diseases, minimum numbers of CNV's were found in the case of psoriasis, and a significant amount of CNVs are present in all the selected diseases.

Conclusion: CNVs constitutes a substantial fraction of total genetic variability and it has the importance in modulating human diseases. This study has shown a significant presence of CNVs in all the selected diseases. Hence it can be concluded that CNVs can be major causing factors in many other life threatening diseases as well and a specific study designed to identify these variations can open a new dimension in the development of novel therapy for those diseases.

Keywords: Copy number variations, Array comparative genomic hybridisation, Single nucleotide polymorphism, End-sequence profiling, Genomic database

© 2017 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)
DOI: <http://dx.doi.org/10.22159/ijpps.2017v9i6.19029>

INTRODUCTION

Copy number variations (CNVs) are a form of structural variation; it is defined as the alterations of the DNA of a genome that results in the cell having a normal or abnormal variation in copy number of either single part or multiple parts of the DNA. It belongs to a comparatively larger part of the genome which has been either duplicated or deleted on certain chromosomes [1]. The human genome has 3 billion pairs of bases or nucleotides in DNA, which are packed into 23 pairs of chromosomes, one set of each pair inherited from each parent. The DNA encodes nearly 30,000 genes. It is a general conception that only two copies of these genes are present in a genome. However, it has been revealed by recent discoveries that large chunk of DNA, varying in size from a few thousand to millions of DNA bases, can be present in multiple copy numbers. Such variations can encompass genes leading to dosage imbalances. For example, it has always been thought, that genes always present in only two copies per genome have now been found that sometimes it can be found in one, three, or more than three copies. Most traits, including susceptibility to disease, can be influenced by these changes. From previous studies, It has been discovered that 12% of the human genome were copy number variable in the 270 DNA test samples. About 2900 genes or 10% of those known is encompassed by these CNVs. Some CNVs found in the general population can be millions of bases in size, affect CNVs are comprised of the insertion, deletion, and duplication of DNA fragments with lengths ranging from one kilobase to five megabases. It has been found in recent studies that CNVs are extensively related to various diseases such as cancer and neuropsychiatric disorders [2].

Most CNVs are usually neutral and do not directly involve in diseases, but there are several instances also where CNVs that affect critical developmental genes do cause diseases. CNVs introduce huge genetic variations on genes dosage and their expression levels [3]. Due to their impact on human disease CNVs can be used in both the

diagnosis and treatment of diseases such as autism, psoriasis, rheumatoid arthritis, breast cancer, psoriasis. Autism is a neurodevelopmental disorder characterized by impairments in reciprocal social interaction, communication deficits and repetitive and restricted patterns of behavior. With the use of comparative genome hybridization screening technique, the association between Autism susceptibility and CNV have been revealed. In this case, many CNV hotspots have been found for example 16p11.2 and 22q11.2 of the human genome [4].

Rheumatoid Arthritis (RA) is a systemic inflammatory disease. On the basis of blood tests for Rheumatoid Factor (RF) or autoantibodies against citrullinated proteins (ACPA), RA segregates into seropositive and seronegative patient populations [5]. RF and anti-CCP antibodies are expressed in seropositive patients. Studies have revealed the association of low copy number of the Fc gamma receptor 3B (FCG, R3B) gene in systemic autoimmune disease [6]. This receptor for IgG is present almost exclusively on neutrophils and plays a role in their interaction with immune complexes [7].

In breast cancer, the role of Copy number variations (CNVs) in the genome has been fully explored [8]. It is estimated that 5-10% of these families harbor germline mutations or complex genomic changes that render inactive one of four high percentage genes BRAC A1, BRAC A2, PTEN or TP53 or moderate penetrance genes CHEK2, PALB2, ATM and BRIP1.

Accordingly, a very common inflammatory skin disorder is Psoriasis. In humans, DEFB genes secrete β -defensins, which are small antimicrobial peptides. DEFB genes have three main gene clusters, one on 8p23.1 and two on chromosome 20. Red-scaling, elevated plaques, commonly on the trunk, knees and elbows are the main characteristics of Psoriasis [9].

T2DM (Type-2 diabetes mellitus) is a type of metabolic disorder which can be characterized by insulin resistance and hyperglycemia. As per the

experiments of Bae *et al.* in 2011, it has been found that more than 170 million people worldwide are affected by this disease in which a significant role is played by genetic factors [10]. The risk of diabetes is around 3-4 times higher in the children of those having T2DM as compared to those children having no familial history of diabetes [11]. In the present study different bioinformatics based approach were applied to better understand the role of CNVs in these different diseases.

In this study, we used various computational tools and databases to identify the extent of CNVs in these selected diseases and also found the reason of these variations on human chromosomes.

MATERIALS AND METHODS

Different tools and database to diagnose genes, sequences and chromosomes related to CNVs which causes diseases have been used. These tools and database include Copy number variation disease (CNVD) database, CNV annotator, DECIPHER, CNV-WebStore, CNV Workshop, and Database of genomic variations (DGV).

CNVD

It has CNV related information for around 792 diseases spread over 22 different species from diversified experiments. It ensures CNVs and disease relationship with very high confidence and comprehensive representations. Various types of query modes are also available in this database. In addition, it also provides graphical representations of results. The user-friendly interface along with the integrated information of various CNVs related to multiple diseases, it offers an accurate and comprehensive platform which can be used for the study of chromosomal structural variations related to different diseases [12]. The CNVD interface can be accessed from <http://bioinfo.hrbmu.edu.cn/CNVD>.

CNV annotator

CNVannotator is a web based server used for CNV analysis from a human genomic position input set in a tabular format. It can create the genomic overlaps from the input coordinates by using its numerous functional features. It includes a list of 356,817 common reported CNVs in which there are 181,261 disease-associated CNVs. It also has around 140,342 SNPs reported from various genome-wide association studies [2]. CNVannotator also has 2,211,468 genomic features which include segmental duplication, encode regulatory elements, genome fragile site, cytoband, pseudogene, promoter, CpG island, enhancers, and methylation site. CNVannotator can also apply various types of search filters to find a subgroup of CNVs reported in many oncogenes and tumor suppressor genes, helping the researchers belong to cancer research community. 5,277,234 unique genomic coordinates in total, with various functional features, are available. It can be used to generate output in a simple plain text format and can be downloaded freely. The CNVannotator server and all of its obtained results can be found at <http://bioinfo.mc.vanderbilt.edu/CNVannotator/>.

Decipher

Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources (DECIPHER) is an interactive online database having a suite of tools designed to interpret the submicroscopic

chromosomal imbalance, translocations and inversions [13]. DECIPHER can be used to find common CNVs in general populations and hence, by exclusion, it can help to identify changes which are novel and potentially pathogenic. It can also help in genetic counselling by fetching relevant information from various types of bioinformatics resources. Gene information which are already known or predicted are available in the patient report segment of DECIPHER. It has also highlighted and prioritized all the clinically recognized importance genes. Clinical scientists of the entire world can maintain records of their patient's phenotype and chromosome rearrangement information on the DECIPHER server and with patient's informed consent, they can also share this information with other clinical researchers using Ensembl genome browser display feature. This helps in identifying the common phenotype and structural rearrangements in the clusters of rare cases which also leads to the identification of new syndromes and understanding of related gene function.

DGV

Database of Genomic Variants (DGV) has a publicly accessible and comprehensive catalogue of genomic structural variation (SV) found in worldwide human populations. In DGV, all the new updates and features can be found which can be used by both the basic researchers and clinical diagnostic communities. With the recent update, DGV currently has 55 published studies which include more than 2.5 million entries from more than 22300 genomes. The information available in DGV is the curated information obtained from the accessioned data sets of the archival structural variations databases DGVA (EBI) and dbVar (NCBI). It also has a visualization tool naming gbrowse which has been recently upgraded and now include an additional feature to facilitate data comparison and analysis. To provide interactive and flexible access to the data, a new query tool has also been developed. The DGV contents are regularly submitted into other genome reference databases which can be used in new database and product development using standard data resource specifically for the testing of copy number variations in clinical labs. The cataloging of accurate information related to structural variations in DGV help to enable more research in the area of genome sequencing and medical genetics [14].

RESULTS AND DISCUSSION

As per this study using different computational tools and databases, the role of CNVs in selected diseases was found to be very different.

Autism

Autism is neuro-developmental disorder characterized by impairments in reciprocal social interaction, communication deficits and repetitive and restricted patterns of behavior and interest. Approximately 10% of the autism spectrum disorder population is thought to have large chromosomal rearrangements. Genome structure can be altered by CNV By either through duplication of any chromosomal region or deletion. By the use of comparative genomic hybridization methods for screening CNV, its relation with autism susceptibility has become clearer. The result of the disease based search in CNVD database for the identification of a number of CNVs associated with Autism has been shown in fig. 1:

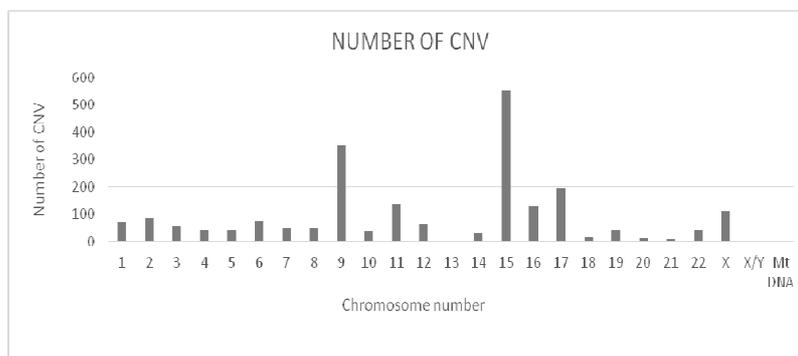


Fig. 1: Number of CNVs found on different chromosome number in case of Autism

Fig. 1 indicates that more or less all chromosomes have Autism associated CNVs but maximum numbers of CNVs are located on chromosome number 15 which is 552 in number and minimum was

observed on chromosome number 13. The frequency percentages of different types of CNVs on all chromosomes were also obtained as shown in table 1.

Table 1: Table showing the different types of CNV frequency in case of Autism

S. No.	Type of CNV	Frequency
1.	Gain	35.84%
2.	Loss	47.16%
3.	Copy number	16.98%

From table 1, it can be concluded that the major reason of variation in copy number in case of Autism is the loss and gain of nucleotides on the different chromosomes. This result is in correlation with the findings of Marshall and Scherer who have detected and characterized copy number variation in autism disorder [15].

Psoriasis

Psoriasis vulgaris (PsV) is a common inflammatory skin disorder characterized by epidermal hyperproliferation, altered keratinocyte differentiation, and inflammation. In PsV patients below the age of 40 y, the major risk allele identified to be an HLA-C allele and its

variants with strong linkage disequilibrium (LD). The most evidentiary psoriasis susceptibility locus PSORS1 has been identified using this finding. Besides this risk factor and PSORS4, replicated associations with candidate genes RAPTOR and SLC12A8 have been reported so far at PSORS2 and PSORS5. It has also been reported that CNV on chromosome 8p23.1 genomic segment has a cluster of a gene which encodes a small antimicrobial peptide of β defensins, which was reported to have an association with psoriasis in German and Dutch case-control cohort. The result of the disease based search in CNVD database for the identification of a total number of CNVs associated with Psoriasis has been shown in fig. 2:

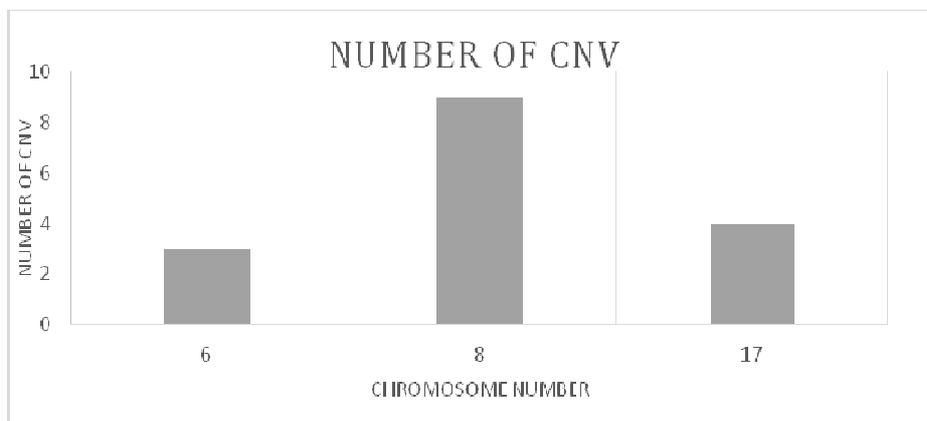


Fig. 2: Number of CNVs found on different chromosome number in case of psoriasis

Fig. 2 indicates that CNV contributes less in the case of Psoriasis. Only three chromosomes have Psoriasis associated CNVs among which maximum CNVs are located on chromosome number 8 which

is 9 in number and minimum was observed on chromosome number 6. The frequency percentages of different types of CNVs on these chromosomes were also obtained as shown in table 2.

Table 2: Table showing the different types of CNV frequency in case of psoriasis

S. No.	Type of CNV	Frequency
1.	Gain	75%
2.	Loss	0%
3.	Copy number	25%

From table 2, it can be concluded that the major reason of variation in copy number in case of Psoriasis is the gain of nucleotides on the different chromosomes. Similar findings have also been reported by Duffin *et al.*, in 2010. They have reported the presence of some significant Psoriasis causing genetic variation on chromosome number 6 [16].

Rheumatoid arthritis

It is the most common form of autoimmune arthritis, similar to Fibromyalgia which is a common chronic pain disorder affecting an estimated 2% of the general population. American College of Rheumatology (ACR) has defined Fibromyalgia as at least 3 mo of widespread pain and pain on palpation at nearly 18 tender point

sites. The result of the disease based search in CNVD database for the identification of a total number of CNVs associated with Rheumatoid Arthritis has been shown in fig. 3:

Fig. 2 indicates that the total number of CNVs vary from one to five in nine different chromosomes in the case of Rheumatoid Arthritis. Among these, maximum numbers of variations are found on chromosome number 6 whereas minimum was observed on chromosome number 2, 7 and 13. The frequency percentages of different types of CNVs on these chromosomes were also obtained as shown in table 3. Chen *et al.*, have also reported the strong association of few CNVs with systemic lupus erythematosus and rheumatoid arthritis which is in correlation with this finding [17].

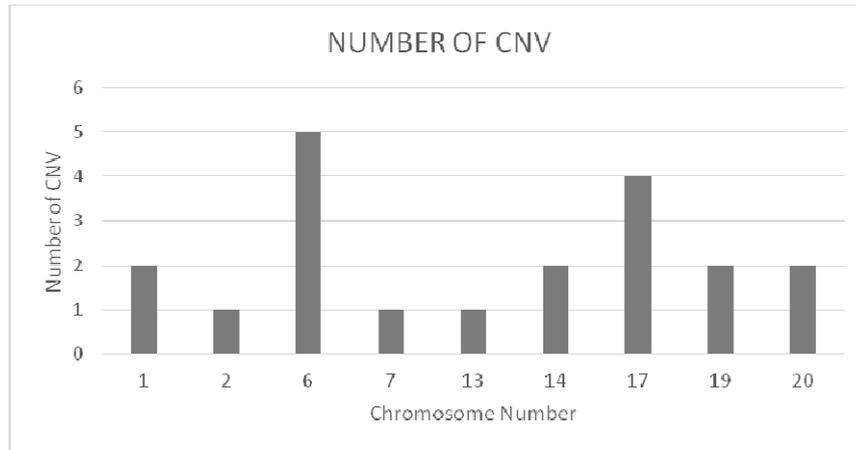


Fig. 3: Number of CNVs found on different chromosome number in case of rheumatoid Arthritis

Table 3: table showing the different types of CNV frequency in case of rheumatoid arthritis

S. No.	Type of CNV	Frequency
1.	Gain	0
2.	Loss	0%
3.	Copy number	100%

From table 3, it can be concluded that the reason of variation in copy number in case of Rheumatoid Arthritis is the gene copy number on different chromosomes.

Type 2 diabetes

Type 2 diabetes mellitus (T2DM) is a serious metabolic disorder that is characterized by insulin resistance and hyperglycemia. This disease affects more than 170 million people worldwide and is, therefore, a major health problem. The familial and twin studies

have suggested the higher prevalence of T2DM in a certain population and hence it is suspected by scientists that in the development of this disease, a significant role is played by genetic factors. The result of the disease based search in CNVD database for the identification of a total number of CNVs associated with type-2 Diabetes has been shown in fig. 4:

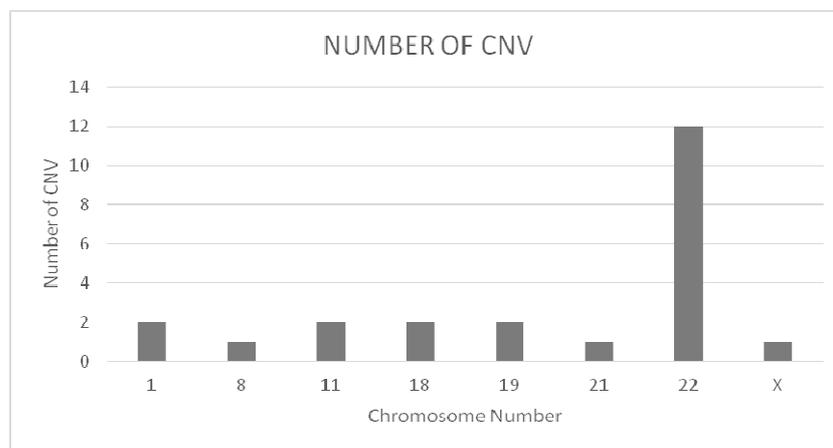


Fig. 4: Number of CNVs found on different chromosome number in case of type 2 diabetes

Fig. 4 indicates that the total number of CNVs vary from one in chromosome number 8, 21 and X to maximum twelve in chromosomes number 22 in the case of T2DM. The frequency percentages of different types of CNVs on these chromosomes

were also obtained as shown in table 4. Prabhanjan *et al.*, have also reported the hotspot for the T2DM on chromosome number 22, 12, 6, 19 and 11 which is completely in accordance with our findings [18].

Table 4: Table showing the different types of CNV frequency in case of type 2 diabetes

S. No.	Type of CNV	Frequency
1.	Gain	0%
2.	Loss	0%
3.	Copy number	100%

From the table 4, it can be concluded that the reason of variation in copy numbers in case of T2DM is the gene copy number on different chromosomes.

Breast cancer

It is the cancer of breast tissues which includes lumps formation in those tissues or change in the shape of breast and skin patches etc. It is common among family members having a history of this disease. In last two decades various genes have been reported which are unequivocally related to Breast cancer risk but still, there are a high

proportion of families that cannot be accounted for by these genes. CNVs contributions in Breast cancer are yet to be explored completely. To identify novel genes and loci which are associated with the risk of this disease, CNV analysis can be an important medium. The result of the disease based search in CNVD database for the identification of a total number of CNVs associated with Breast cancer has been shown in fig. 5:

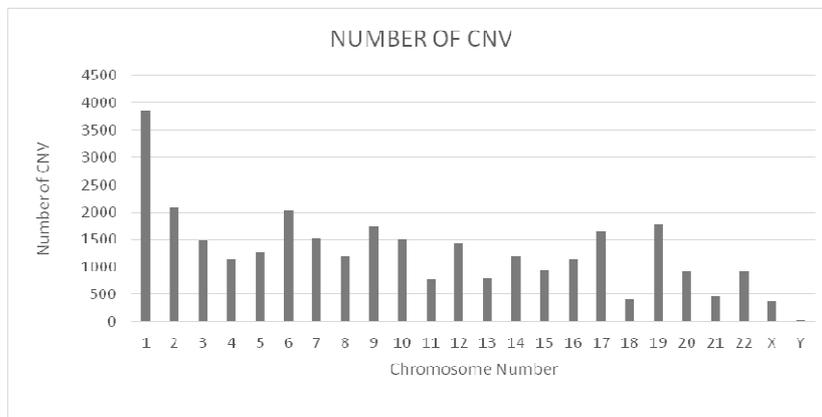


Fig. 5: Number of CNVs found on different chromosome number in case of breast cancer

Fig. 5 indicates that the numbers of variations in copies of various genes or loci in the human genome is very high in case of Breast cancer. Very large numbers of CNVs are found in almost all chromosomes. The maximum 3851 CNVs are present on

chromosomes number one and minimum variations are present on chromosome number Y. The frequency percentages of different types of CNVs on these chromosomes were also obtained as shown in table 5.

Table 5: table showing the different types of CNV frequency in case of breast cancer on chromosome number 1

S. No.	Type of CNV	Frequency
1.	Gain	3.58%
2.	Loss	90.98%
3.	Copy number	5.97%

From the table 5, it can be concluded that in the case of Breast cancer, the major reason of variation in copy numbers on the chromosome number

one is the loss of nucleotide. The frequencies of different types of CNV on chromosome number six were also calculated as given in table 6.

Table 6: table showing the different types of CNV frequency in case of breast cancer on chromosome number 6

S. No.	Type of CNV	Frequency
1.	Gain	53.34%
2.	Loss	46.35%
3.	Copy number	0.48%

From the table 6, it can be concluded that in case of Breast cancer, gain and loss of nucleotide contributes almost equally on the chromosome number six for effective variation in copy numbers. These results are also in accordance with the findings of Walker *et al.* and Shlien and Malkin. They have reported that there exist very strong correlations with CNVs and breast cancer. They have also concluded that amplitude of CNVs in cancer is very high which is similar to our findings as well [19, 20].

CONCLUSION

It can be briefly summarized that structural genetic variation, including copy number variations (CNVs), constitutes a substantial fraction of total genetic variability and the importance of structural genetic variants in modulating human disease is increasingly being recognized. Initial success in the identification of diseased associated variations in copy number using a candidate gene approach stress that in future also any disease associated studies must include structural genetic variations. Currently, disease-

associated genetic variation studies have become an evolving field in the area of medical research. But still, the lack of advance technology, computational databases and tools for statistical studies in this area is the major limitation. Only a few databases and tools are available to analyse the experimental results. Hence, in addition to the development of better CNV genotyping computational platforms, we also stress that more rigorous attention must be given for the study design and statistical analysis for the early detection of the disease association with genetic variations.

In this study, minimum CNVs were found in the case of psoriasis in which the major reason of variations was found to be the gain of new nucleotides mainly on chromosome number eight. The maximum CNV's were found in the case of breast cancer specifically on chromosome number one which is 3851. In this case gain and loss of nucleotide on the genome were and the main cause of these variations. Variations in copy numbers were also found in other diseases as well. Hence, it can be concluded that CNVs can be major causing factors in many life threatening diseases and a specific study

designed to identify these variations can open a new dimension in the development of novel therapy for these diseases.

AUTHOR CONTRIBUTION

All persons listed as authors meet the authorship criteria and certify that they have sufficiently participated in the work for various contents, including the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication. In this work, conception and design of the study were prepared by Yadav A, Singh P, Kumar PR, Srivastava S and Mishra SK. Data acquisition was done by Yadav A and Singh P followed by analysis and interpretation of data. Manuscript was drafted by Yadav A, Singh P and Kumar PR whereas Srivastava S and Mishra SK have reviewed and revised the manuscript critically for any error and important intellectual contents. Further, the final version of the manuscript was approved by all the authors of this paper.

CONFLICT OF INTERESTS

Declared none

REFERENCES

1. Yue GH. Recent advances of genome mapping and marker-assisted selection in aquaculture. Fish and Fisheries, Blackwell Publishing Ltd. Singapore; 2012.
2. Zhao M, Zhao Z. CN Vannotator: A comprehensive annotation server for copy number variation in the human genome. Plos One 2013;8. <https://doi.org/10.1371/journal.pone.0080170>
3. Hollox EJ, Barber JCK, Brookes AJ, Armour JAL. Defensins and the dynamic Genome: what we can learn from structural variation at human chromosome band 8p23.1. Genome Res 2016;18:1686-97.
4. Shishido E, Aleksic B, Ozaki N. Copy number variations in the pathogenesis of autism spectrum disorder. Psychiatry Clin Neurosci 2013;68:456-64.
5. Jonah DH, Logan RB, Maynard R, Jennifer EA, Brian CB, David JR, et al. VISA-vector integration site analysis server: a web-based server to rapidly identify retroviral integration sites from next-generation sequencing. BMC Bioinf 2015;16:212.
6. Almal S, Padh H. Frequency distribution of autoimmunity associated fcgr3b gene copy number in indian population. Immunogenetics 2016;1:2.
7. Willer CJ, Speliotes EK, Loos RJ. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet 2009;41:25-34.
8. Masson LA, Talseth-Palmer BA, Evans EJ, Grice MD, Hannan NG, Scott RJ. Expanding the genetics basis of copy number variation in familial breast cancer. Hered Cancer Clin Practice 2014;12:15-25.
9. Hollox EJ, Huffmeier U, Zeeuwen PLJM, Raquel P, Lascorz J, Olthuis DR, et al. Psoriasis is associated with increased beta-defensin genomic copy number. Nat Genet 2008;40:23-5.
10. Bae JS, Cheong SH, Kim JH, Park BL, Kim HJ, Park JT, et al. The genetic effect of copy number variations on the risk of type 2 diabetes in a Korean population. Plos One 2016;22:e19091.
11. Zilina O, Koltzina M, Raid R, Kurg A, Tonisson N, Salumets A. Somatic mosaicism for copy-neutral loss of heterozygosity and DNA copy number variations in the human genome. BMC Genomics 2015;16:703.
12. Qiu F, Xu Y, Li K, Li Z, Liu Y, DuanMu H, et al. CNVD: text mining-based copy number variation in disease database. Hum Mutat 2012;33:2375-81.
13. Helen VF, Shola MR, Bevan AP, Stephen C, Manuel C, Diana R, et al. Decipher: a database of chromosomal imbalance and phenotype in humans using ensembl resources. Am J Hum Genet 2016;84:524-33.
14. MacDonald JR, Robert Z, Ryan KCY, Lars F, Stephen WS. The database of genomic variants: a curated collection of structural variation in the human genome. Nucleic Acids Res 2014;42:986-92.
15. Marshall CR, Scherer SW. Detection and characterization of copy number variation in autism spectrum disorder. Methods Mol Biol 2012;838:115-35.
16. Duffin KC, Woodcock J, Krueger GG. Genetic variations associated with psoriasis and psoriatic arthritis found by genome-wide association. Dermatol Ther 2010;23:101-13.
17. Chen JY, Wang CM, Chang SW, Cheng CH, Wu YJ, Lin JC, et al. Association of FCGR3A and FCGR3B copy number variations with systemic lupus erythematosus and rheumatoid arthritis in Taiwanese patients. Arthritis Rheumatol 2014;66:3113-21.
18. Prabhajan M, Suresh RV, Murthy MN, Ramachandra NB. Type 2 diabetes mellitus disease risk genes identified by genome wide copy number variation scan in normal populations. Diabetes Res Clin Pract 2016;113:160-70.
19. Walker LC, Wiggins GAR, Pearson JF. The role of constitutional copy number variants in breast cancer. Microarray 2015;4:407-23.
20. Shlien A, Malkin D. Copy number variations and cancer. Genome Med 2009;1:62.

How to cite this article

- Anushka Yadav, Pooja Singh, Priya Ranjan Kumar, Sarika Srivastava, Santosh Kumar Mishra. Study on implications of copy number variations (CNVs) in the human population. Int J Pharm Pharm Sci 2017;9(6):205-210.