Review Article

J Genet Med 2015;12(1):1-5
http://dx.doi.org/10.5734/JGM.2015.12.1.1
ISSN 1226-1769 (Print) 2383-8442 (Online)

JGM Journal of Genetic Medicine

The role of *de novo* variants in complex and rare diseases pathogenesis

Mahir Rahman[1,2], Woohyung Lee[3], and Murim Choi[2,*]

[1]Yale College, Yale University, New Haven, CT, United States
[2]Department of Biomedical Science, Seoul National University College of Medicine, Seoul, Korea
[3]Hong Kong University of Science and Technology, Hong Kong, China

De novo variants (DNVs) can arise during parental germ cell formation, fertilization, and the processes of embryogenesis. It is estimated that each individual carries 60-100 such spontaneous variants in the genome, most of them benign. However, a number of recent studies suggested that DNVs contribute to the pathogenesis of a variety of human diseases. Applications of DNVs include aiding in clinical diagnosis and identifying disease-causing genetic factors in patients with atypical symptoms. Therefore, understanding the roles of DNVs in a trio, with healthy parents and an affected offspring, would be crucial in elucidating the genetic mechanism of disease pathogenesis in a personalized manner.

Key words: *De novo* variants, Autistic disorder, Congenital heart disease, Schizophrenia, Rare diseases.

Introduction

De novo variants (DNVs) can occur during germ cell formation or at any point during the development of a fertilized egg. Such phenomena originate in the offspring and are not present in either parent. Predominantly found as single-nucleotide variants (SNVs), DNVs also appear as short insertions or deletions (indels) or copy number variants (CNVs). Indels involve the insertion or deletion of one to a thousand base pairs, whereas CNVs can involve lengths of 1 kb or more. Application of improved next-generation sequencing (NGS) technique-based whole genome sequencing (WGS) and whole exome sequencing (WES) can yield valuable information regarding the association of DNVs with the disease, particularly in family-based trios with one affected proband and an unaffected sibling. The successful detection of true DNVs, as opposed to false positives, is crucial in understanding the origin of various common and complex disorders [1].

Nature of *De Novo* Variants

On an average, approximately 74 *de novo* SNVs occur per genome per generation. The frequency of *de novo* SNVs in the population suggests the benign nature of most of these DNVs. The extensive cell divisions that occur during spermatogenesis play a significant role in DNV frequency. With increased paternal age, there is a corresponding increase in mutational load for male-germline cells [1].

One crucial question is whether DNVs occur in a random fashion or if there are DNV "hot spots" in the genome. In a study involving monozygotic twins, Michaelson et al. [2] reconceptualized the idea of gene locality, or location along the

Received: 5 December 2014, Revised: 18 March 2015, Accepted: 17 April 2015, Published: 30 June 2015
*Corresponding author: Murim Choi, Ph.D.
Department of Biomedical Science, Seoul National University College of Medicine, 103 Daehak-ro, Jongro-gu, Seoul 110-799, Korea.
Tel: +82-2-740-8912, Fax: +82-2-3673-2167, E-mail: murimchoi@snu.ac.kr
Conflict of interest: We declare that we do not have any conflicts of interest.

gene, to "hot spots", hypermutable regions in the genome. The mutation rate was not uniform along the genome in any case; "hot spots" may possess up to a 100-fold increased tendency to give rise to DNVs.

It is important to note that the biological effects of DNVs are heavily influenced by their relative location on a gene and that DNV presence alone may not lead to a disease phenotype. DNVs with complete penetrance may impose negative selection on an individual and the DNV won't survive in a population, whereas DNVs that partially contribute to risk or are benign will be under incomplete selection and will have a high population frequency [1]. In the case of autism, risk-conferring DNVs are frequently found in brain-related genes [3]. A higher frequency of highly damaging variants in disease-related genes notably lowers false discovery rates and strongly denotes a risk association. Although the false discovery rate increases with sample size, such an increase also raises the probability of finding highly damaging risk-conferring DNVs, particularly CNVs that can span several genes. Numerous studies agree that DNVs play a role in disease phenotype; however, due to technical variations in methodology, locus heterogeneity and patient diagnosis and sampling, a consensus on the extent to which DNVs play a role in disease has not been reached [4].

Complex Disease

Genetic heterogeneity may complicate the interpretation of the functional roles of rare DNVs in cases of common complex disease [1], but risk-conferring DNVs can help elucidate disease-relevant genes and pathways for complex diseases, particularly congenital disorders that may affect reproductive success during human evolution. However, multiple observations of damaging mutations in such genes are necessary to confirm their risk-conferring abilities. Furthermore, the extent of DNVs' role in complex disease must be compared to the extent of other predominant risk factors such as heredity [4]. To elucidate the extent of the role of DNVs, this paper will examine the effects of such mutations in autism spectrum disorder (ASD), schizophrenia, and congenital heart disease (CHD). Overall, the results of the studies cited demonstrate that DNVs play a large role in the risk of presenting with complex disease phenotype.

1. Autism spectrum disorder

Particularly in ASD cases, DNVs found in in brain-expressed genes increase risk of presenting with the disease phenotype. In a WES study by Sanders et al. [3], from 200 quartets with an affected proband and an unaffected sibling, 15 out of 125 non-synonymous *de novo* SNVs in probands were presumed to be highly damaging (i.e., nonsense frameshift and splice-site DNVs), while only 5 out of 85 such SNVs were found to be highly damaging in the unaffected siblings. The study concluded that two or more highly damaging DNVs in the same gene were unlikely to occur by chance; *SCN2A* was the only gene to have two nonsense *de novo* SNVs, but further whole-exome analysis allowed the authors to posit *KATNAL2* and *CHD8* also as ASD risk-conferring candidate genes. An accompanying study by Neale et al. [5], reaching similar conclusions, found 18 genes with two functional DNVs from 175 trios and matched those genes to loss-of-function variants (nonsense, splice-site, and frameshift); *SCN2A*, *KATNAL2*, and *CHD8* variants were found unlikely to occur by chance. After further evaluation of 935 probands and 870 controls, three additional loss-of-function variants were found at *KATNAL2* and *CHD8* in probands but none in controls. All three genes are associated with nervous system development or function [3], and Neale et al. [5] observed that protein-encoding genes with missense and nonsense *de novo* SNVs were functionally connected to each other and other ASD-associated genes. Such results implicate ASD risk may be strongly associated with specific functional genes, predominantly brain-related genes.

In a similar setup, O'Roak et al. [6] found 49 out of 126 damaging DNVs mapped to an extensive protein network. Genes of the networks are involved in β-catenin pathways, p53 signaling, chromatin remodeling, ubiquitination, and neuronal development. Serving various critical roles in functional development, many of the genes in the network have been implicated as autism candidate genes; *KATNAL* and *CHD8* are among them. Such findings build upon the findings in Neale et al. [5] and the ASD risk association with protein-encoding genes and specific gene localities. The positive correlation found between mutation rate and conservation among both probands and controls suggests that mutability may be coupled with functionality; hypermutability was found to be positively correlated with highly conserved sequences and low genetic diversity. Likewise, mutability was significantly elevated for genes preferentially expressed in the brain and implicated in ASD [2]. Such findings support the functionality of DNVs in a gene and the gene's functionality influences the size of their effects.

2. Schizophrenia

The relation of DNVs to gene locality seems to be a recurring phenomenon in neuropsychiatric disorders. Fromer et al.

[7] recently conducted a large exome sequencing study on schizophrenia. The study's analysis validated 482 non-synonymous DNVs in 617 probands. Proteins with non-synonymous DNVs demonstrated great functional connectivity to each other and other synaptic proteins, with directly interacting proteins involved in processes such as regulating synaptic plasticity, kainate receptor trafficking, and regulation of actin dynamics. Many of the affected synapses are involved with activity-regulated cytoskeleton-associated proteins and *N*-methyl-*D*-aspartate receptor complexes, most of which have also been implicated by the fragile X mental retardation protein. Genes carrying non-synonymous DNVs in schizophrenia cases overlap with such genes associated with ASD and intellectual disability, lending support to the idea of a common molecular mechanism underlying various traits among neuropsychiatric disorders. Like the brain in ASD cases and the heart in CHD cases, the DNV locality in synaptic-related genes provides support to the association between gene locality and DNV effects.

3. Congenital heart disease

Similar to neuropsychiatric cases, the presence of DNVs in heart-related genes can increase the risk of presenting with CHD phenotypes. Zaidi et al. [8] compared 362 CHD proband trios with 264 unaffected siblings of ASD probands as controls, using a Bayesian inference-based quality score assessment of the DNV calling algorithm. The study found a significant increase in the rate of protein altering DNVs in genes highly expressed in the heart in CHD probands relative to controls. When the date were partitioned to include only deleterious protein altering mutations, the odds ratio of developing CHD was increased to 7.50. Similar to the findings of Michaelson et al. [2], the *H3K4* and *H3K27* methylation pathway genes were the affected most heavily, implying a critical role in cardiovascular development. Eight genes with DNVs from CHD probands are part of the *H3K4*me-*H3K27*me pathway, and patients with DNVs in the pathway presented diverse cardiac phenotypes [8]. Such findings reveal gene loci function relates to the effect DNVs can express.

Rare Disease

1. Pseudohypoaldosteronism type II and Cornelia de Lange syndrome

The role of mutated *WNK1* and *WNK4* (WNK lysine deficient protein kinase 1 and 4, respectively) in electrolyte flux pathways have been revealed in studies on pseudohypoaldosteronism type II (PHAII). PHAII patients constitutively increase salt reabsorption regardless of cell volume status and inhibit K^+ secretion despite hyperkalaemia. However, only a small fraction of these PHAII patients (~10%) are affected by *WNK1* or *WNK4* mutations. By using WES, researchers have identified two mutated genes that cause PHAII from 41 unrelated families: *KLHL3* and *CUL3*. They form an ubiquitin ligase complex, which uses WNK1 and WNK4 as substrates. Otherwise, the precise molecular mechanism is not completely understood [9]. The genetic investigation of the PHAII patients revealed that *KLHL3* variants were mostly inherited, whereas *CUL3* variants were mostly *de novo*. The clinical data analysis showed that *CUL3*-mutation carrying patients displayed more severe phenotypes than *KLHL3*-mutation carriers; therefore, one can conclude that severe DNVs pose a large negative selection force on the individual that ultimately affects reproduction efficiency [9].

Another example of the contribution of DNVs to rare disease is Cornelia de Lange syndrome (CdLS). Typically caused by a dominant mutation in cohesion-related genes, such as *NIPBL* (~6%), *SMC1A* (~5%), and *SMC3* (<1%), the patients display systemic defects in intellectual, skeletal, and facial development. Most recently, an X-linked gene *HDCA8* was identified by a genetic screening, which confirmed that the encoded protein functions as an SMC3 deacetylase. It is notable that the six CdLS patients carrying *HDAC8* loss-of-function mutations all acquired the mutations as *de novo* [10].

2. A family with idiopathic autoimmune diseases

A case study involving a father and an infant son who both presented with previously unreported autoimmune symptoms demonstrated that a DNV on the *NLRC4* gene is responsible for autoimmune symptoms. The infant, who presented with diarrhea and fever at two days after birth, suffered severe autoimmune features and died at 23 days after birth due to diffuse alveolar hemorrhage. Shortly following the death of the infant, the father, who presented with high levels of systemic inflammation and apparent mental and physical stress, was also hospitalized. Hospital screenings returned high levels of ferritin, C-reactive protein, and interleukin-18, without any infectious agent isolated. Past hospital records revealed that the patient had suffered the same symptoms as his deceased son without a specific diagnosis, and had clinical features suggesting hemophagocytic lymphohistiocytosis, which led to an examination of genes implicated in this syndrome. WES allowed the identification of 34 new protein-altering variants (absent in dbSNP, 1,000 Genomes Project, NHLBI, and Yale exome

databases) shared by the index case and his father, including six positions that were invariant among orthologs. Among the six, the effects of the SNP in *NLRC4* demonstrated consistency with the clinical syndrome and relevance to increased inflammations that the patients experienced. Subsequent experiments demonstrated that the NLRC4 p.Val341Ala variant increased NLRC4-mediated innate immune signaling, which in turn enhanced downstream caspase activation and cytokine production. Overall, the study provided strong evidence that the *NLRC4* mutation, which was *de novo* in the father and inherited by his son, is the cause of this syndrome [11].

Conclusion

This review primarily focused on the implication of DNVs in an increasing number of diseases with a spectrum of clinical features. These DNV associations were discovered due to recent research advances, including NGS-based genome analyses. We selected several recent studies on common complex diseases, rare diseases, and atypical symptoms that are not classified by conventional clinical criteria, which implicate DNVs to various extents (Table 1).

One of the fundamental questions to be addressed in future research is the quantified contribution of DNVs in specific disease models. Assuming 100 individuals of certain clinical characteristics, how many of them will show the characteristic due to DNVs? The answer will depend on many aspects, such as the inheritance pattern, severity, prevalence and heterogeneity of the disease. Large-scale, case-control WES-based DNV analysis allows a rough estimation of the quantified disease contribution of DNVs. It is interesting to note that both ASD and CHD studies have concluded that the DNVs for these diseases account for about 10% of the patient cohort [8,12]. Since about 80% of the patients carry functional DNVs in coding regions, only about 12% of the DNVs carried by the patients would be directly causal to the disorders, which leads to another critical problem of selecting the correct 12% of the DNVs. The initial large-scale studies depended on choosing the variants with more severe effects on protein sequences, on recurrence, and on genes with similar functions. While preparing this review, large-scale follow-up ASD studies were published. In a WES study, the researchers recruited about 2,000 trios of healthy parents with ASD offspring [13,14]. To select genes that were the most likely to be functional, they developed an algorithm to assess functionality of the variants based on the background variant burden in a healthy population, and selected the 107 genes with the highest likelihood of having functionality [15].

Therefore, although not the major factor, DNVs contribute to a substantial fraction of congenital complex disorders. Further technical advances will improve detection and analysis of the risk-conferring DNVs, and will help in constructing the complete molecular map of disease pathogenesis mechanisms.

On the other hand, the involvement of DNVs in rare diseases is more profound (Table 1). If a disease is inherited in a dominant manner and its phenotype is so severe that the mutation directly impairs the survival or reproductivity of the patient, most of the mutations cannot be inherited but will be caused *de novo*. Among these phenomena, we chose to show PHAII *CUL3*-mutated and CdLS *HDAC8*-mutated patients, suffering from different but similarly life-threatening clinical conditions. In the final example introduced, the patient displayed milder symptoms when than most NIPBL cases, but it was still a severe clinical case that was deciphered by the identification of a risk-conferring DNV. For previously genetically uncharacterizable and therefore untreatable diseases, recent advances in WGS aid researchers and physicians in efficiently identifying disease-causing DNVs and in relating them with the patient phenotype, allowing active and personalized intervention for the specific

Table 1. Summary of the disease models used in this review

Disease	Feature	Prevalence	Genes involved with the disease (n)	DNV functionality	References
ASD	Congenital/complex	~1%	>500	~10%	3, 5, 6
Schizophrenia	Congenital/complex	~1%	>200	NA[a]	7
CHD	Congenital/complex	~1%	>300	~10%	8
PHAII	Rare/dominant	Unknown	~5	100% for *CUL3* variants	9
CdLS	Rare/dominant	~0.001%	~5	100% if phenotype is severe	10
A family with *NRLC4*-mutation	Single/dominant /undiagnosed	Single case	1	100%	11

DNV, *de novo* variant; ASD, autism spectrum disorder; CHD, congenital heart disease; PHAII, pseudohypoaldosteronism type II; CdLS, Cornelia de Lange syndrome; NA, not available.
[a]Such data was not presented in the sourced paper.

clinical symptoms that the patients suffer.

Acknowledgements

This work was supported by the Research Resettlement Fund for the new faculty of Seoul National University. Mahir Rahman is a Richard U. Light Fellowship recipient.

References

1. Veltman JA, Brunner HG. De novo mutations in human genetic disease. Nat Rev Genet 2012;13:565-75.
2. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, Jin X, et al. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. Cell 2012;151:1431-42.
3. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, et al. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 2012;485:237-41.
4. Gratten J, Visscher PM, Mowry BJ, Wray NR. Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. Nat Genet 2013;45:234-8.
5. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, Sabo A, et al. Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 2012;485:242-5.
6. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, Coe BP, et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 2012;485:246-50.
7. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, Gormley P, et al. De novo mutations in schizophrenia implicate synaptic networks. Nature 2014;506:179-84.
8. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, et al. De novo mutations in histone-modifying genes in congenital heart disease. Nature 2013;498:220-3.
9. Boyden LM, Choi M, Choate KA, Nelson-Williams CJ, Farhi A, Toka HR, et al. Mutations in kelch-like 3 and cullin 3 cause hypertension and electrolyte abnormalities. Nature 2012;482:98-102.
10. Deardorff MA, Bando M, Nakato R, Watrin E, Itoh T, Minamino M, et al. HDAC8 mutations in Cornelia de Lange syndrome affect the cohesin acetylation cycle. Nature 2012;489:313-7.
11. Romberg N, Al Moussawi K, Nelson-Williams C, Stiegler AL, Loring E, Choi M, et al. Mutation of NLRC4 causes a syndrome of enterocolitis and autoinflammation. Nat Genet 2014;46:1135-9.
12. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, et al. De novo gene disruptions in children on the autistic spectrum. Neuron 2012;74:285-99.
13. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of de novo coding mutations to autism spectrum disorder. Nature 2014;515:216-21.
14. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 2014;515:209-15.
15. Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, et al. A framework for the interpretation of de novo mutation in human disease. Nat Genet 2014;46:944-50.