Received: 18 Apr 2024 | **Accepted:** 26 Apr 2024 **Genel Tıp Derg. Volume** 34/**Issue** 2 (April), 266-272
DOI: 10.54005/geneltip.1463733

ORIGINAL ARTICLE

Investigation of the Frequency and Characteristic Features of De Novo Mutations in Clinical Exome Sequence Trio Samples

Klinik Ekzom Sekans Trio Örneklerinde De Novo Mutasyonların Sıklığı ve Karakteristik Özelliklerinin Araştırılması

[1]Nadir Koçak, Ali Torabi, Batuhan Şanlıtürk, Özkan Bağcı, Ebru Marzioğlu Özdemir, Tülün Çora

[1]Selçuk Üniversitesi Tıp Fakültesi, Tıbbi Genetik ABD, Konya, Türkiye

Correspondence

Nadir Koçak, Selçuk Üniversitesi Tıp Fakültesi, Tıbbi Genetik ABD, Konya, Türkiye

E-Mail: nadirkocak@yahoo.com

How to cite ?

Koçak N, Torabi A, Şanlıtürk B, Bagci O, Marzioğlu-özdemir E, Çora T. Klinik Ekzom Sekans Trio Örneklerinde De Novo Mutasyonların Sıklığı ve Karakteristik Özlelliklerinin Araştırılması. Genel Tıp Derg. 34(2): 266-272.

ABSTRACT

Advanced genome sequencing technologies have provided us with the opportunity to deeply understand the mechanisms underlying conditions associated with the genome. There has been significant interest recently in understanding the characteristics of de novo mutations, which are genetic changes that arise in reproductive cells and are not present in parents, as well as the mechanisms involved in their occurrence. These mutations can be transmitted to subsequent generations and have the potential to influence genetic diversity and susceptibility to diseases, making this topic important. Due to limited studies in this area, the formation mechanisms and characteristic features of such mutations have not yet been fully understood.
In this study, we aimed to conduct a comprehensive analysis of de novo mutations in families undergoing trio clinical exome sequencing analysis. The objectives of the study were to investigate the relationship between parental ages and the frequency of de novo mutations, the distribution, prevalence, relationships, and molecular characteristics of de novo mutations. A total of 69 families who underwent Trio Clinical Exome Sequencing (CES) analysis at the Department of Medical Genetics, Faculty of Medicine, Selçuk University, between January 1, 2017, and December 31, 2023, were included in the study. DNA samples extracted from peripheral venous blood of individuals were sequenced using the Roche CES kit and DNBSEQ-G400™ sequencing device, and a total of 3892 genes were analyzed using the Seq Platform.
Correlation analysis revealed no significant relationship between parental age and the number of de novo mutations, and regression analysis showed that age was not a significant parameter in determining the number of de novo mutations. After analysis, 407 de novo variants were identified, with the majority being variants of unknown significance (55.28%). When examining the base change profile, the most common changes were found to be C -> G, G -> A, A -> G. The most commonly mutated genes were found to be DSPP, HPS4, VCL, and BMP4 genes.

Keywords: Mutation, Exome Sequencing, Genom

ÖZ

Gelişmiş genom dizileme teknolojileri genomla ilişkili durumların temelinde yatan mekanizmaları derinlemesine anlamaıza fırsat sağladı. De novo mutasyonlar olarak adlandırılan, ebeveynlerde bulunmayan ve üreme hücrelerinde ortaya çıkan genetik değişikliklerin karakteristiklerinin ve ortaya çıkmasında rol oynayan mekanizmaların anlaşılması son dönemde büyük ilgi görmektedir. Bu tür mutasyonlar, sonraki nesillere aktarılabilir ve genetik çeşitliliği ve hastalık yatkınlığını etkileyebilme potansiyeli bu konuyu önemli hale getirmektedir. Bu konuda çalışmalar sınırlı olduğu için henüz bu tür mutasyonların oluşum mekanizmaları ve karakteristik özellikleri tam olarak anlaşılamamıştır.
Bu araştırmada, trio klinik exome dizileme analizi yapılan ailelerde de novo mutasyonların kapsamlı bir analizi gerçekleştirmeyi amaçladık. Araştırmanın amaçları olarak, ebeveynlerin yaşları ile de novo mutasyonların meydana gelme sıklığı arasındaki ilişkiyi, de novo mutasyonların dağılımını, prevalansını, birbirleriyle ilişkilerini ve moleküler karakteristiklerini araştırmak olarak belirledik.
Araştırmada, 1 Ocak 2017 ile 31 Aralık 2023 tarihleri arasında Selçuk Üniversitesi Tıp Fakültesi Tıbbi Genetik Anabilim Dalı'nda Trio Klinik Exome Dizileme (CES) analizi yapılan 69 aile incelenmiştir. Araştırmada bireylerin periferik venöz kanlardan ekstrakte edilen DNA örnekleri Roche CES kitini kullanarak ekstrakte edilmiş ve DNBSEQ-G400™ dizileme cihazıyla dizilenmiş ve Seq Platformu kullanılarak toplamda 3892 gen analiz edilmiştir.
Korelasyon analizi sonucunda, annenin ve babanın yaşının de novo mutasyon sayısı ile anlamlı bir ilişkisinin olmadığı saptanmış ve regresyon analiz sonrasında da yaşın de novo mutasyon sayısını belirlemede anlamlılık bir parametre oluşturmadığı görülmüştür
Analiz sonrası 407 de novo varyant tespit edilmiş ve bunların çoğununun anlamı bilinmeyen varyantlar (%55.28) olduğu saptanmıştır. Baz değişim profili incelendiğinde en sık rastlanan değişikliklerin C -> G, G -> A, A -> G olduğu belirlenmiştir. En sık de novo mutasyon rastlanan genlerin DSPP, HPS4, VCL, BMP4 genleri olduğu saptanmıştır.

Anahtar Kelimeler: Mutasyon, ekzom sekans, Genome, Gen

Background/Aims

Advancements in advanced genome sequencing technologies have revolutionized our understanding of genetic disorders and enabled researchers to gain deeper insights into the underlying mechanisms of hereditary conditions. One particularly intriguing area of study is the investigation of de novo mutations, which arise spontaneously in the reproductive cells of parents and are subsequently passed on to subsequent generations. Understanding the nature of de novo mutations in the human genome forms the basis for unraveling phenomena such as genetic inheritance, genetic diversity, and susceptibility to disease. While a typical human genome exhibits millions of genetic variations, a subset of these variants - rare and novel

Peer-Review: Double anonymized - Two External
Plagiarism Checks: Yes - intihal.net
Complaints: geneltip@selcuk.edu.tr

Selcuk University Press
Genel Tıp Dergisi | e-ISSN: 2602-3741
https://dergipark.org.tr/tr/pub/geneltip
https://yayinevi.selcuk.edu.tr/

mutations - arises as de novo events, contributing to genetic diversity and disease susceptibility. However, the mechanisms underlying the formation and distribution of de novo mutations are not yet fully understood. DNA replication errors, particularly in CpG dinucleotides, inadequate repair mechanisms, and exposure to endogenous or exogenous mutagens, all contribute to the occurrence of de novo mutations. These mutations do not occur randomly throughout the genome; instead, specific genomic regions exhibit higher mutability due to intrinsic properties such as sequence composition, replication timing, and transcriptional activity. Additionally, recent studies have identified mutational clusters and hotspots, which denote specific genomic regions prone to de novo mutations (1, 2).

Advancements in next-generation sequencing (NGS) technologies have revolutionized studies on de novo mutations, enabling researchers to comprehensively characterize these genetic alterations. Through genome-wide studies conducted in parent-child trios, researchers have scaled the rate of de novo mutations and elucidated their parental origins, distribution, and impact on disease susceptibility (1). However, despite these advancements, many questions remain unanswered regarding the precise mechanisms of de novo mutation formation, factors influencing their distribution across the genome, and their contributions to human diseases. In this study, we conducted a comprehensive analysis of de novo mutations in families undergoing clinical Exome Sequencing (CES) analysis for various reasons. Our primary objectives in this research were to investigate the relationship between parental ages at the start of pregnancy and the occurrence of de novo mutations, the frequencies and distributions of de novo mutations, their relationships with each other, and their molecular characteristics.

Methods

In this study, a total of 69 families who underwent trio Clinical Exome Sequencing (CES) analysis at the Department of Medical Genetics, Selçuk University Faculty of Medicine, due to various complaints and suspected genetic origins between January 1, 2017, and December 31, 2023, were included. The demographic information and clinical characteristics of the patients were obtained retrospectively through data recorded during examinations and the hospital's electronic medical record system. Based on this data, the ages of the parents at the beginning of pregnancy were determined. Family trees of all cases included in the study were drawn, their histories and clinical information were collected in detail, and informed consents with wet signatures were obtained from the patients and/or legal guardians. After samples were collected in the wet laboratory, DNA extraction was performed using the Roche CES kit followed by sequencing using the DNBSEQ-G400™ sequencing platform (MGI Tech Co., Ltd.). A total of 3892 genes were analyzed using the mentioned kit. Variant classification and analysis were performed using the Seq Platform (Genomize Inc.). The Seq Platform utilized advanced artificial intelligence algorithms to expedite variant prioritization workflow and generate a variant list. FASTQ files were uploaded to the Seq Platform, and reads were aligned to the human reference genome GRCh37 (Hg19) using the Burrows-Wheeler Aligner (BWA) platform. The aligned reads were then used for variant calling with FreeBayes. The obtained variants were subsequently annotated using VEP v102. ACMG pathogenicity classification was conducted according to the guidelines published by Richards et al. Variants were filtered in various modes: exonic (coding) regions, within 20 base pairs of exon-intron junctions, and with a frequency of less than 1% in the healthy population (according to gnomAD, the 1000 Genomes Project (1000G), and the Exome Aggregation Consortium (ExAC)). Additionally, ACMG, ClinVar, and Franklin (genoox) data were used for pathogenicity determination. Trio analysis was conducted using the trio analysis tool of the Seq platform.

Statistical analysis was performed using IBM SPSS Statistics Version 26 (IBM Corp, USA). The primary dependent variable was the number of de novo mutations, and the independent variables included maternal age, paternal age, and gender. Pearson and Spearman's rank correlation coefficients were calculated to evaluate linear and monotonic relationships, respectively. Additionally, multiple linear regression was used to investigate the combined effects of maternal and paternal ages on the number of de novo mutations. The relationship between gender and the number of de novo mutations was examined using independent samples t-test and Mann-Whitney U test. While the t-test assessed mean differences, the Mann-Whitney U test provided a non-parametric perspective on distribution inequalities. These analyses aimed to provide robust findings considering both statistical and practical significance.

Results

The analyses revealed that the average age of mothers at the beginning of pregnancy was 25.95. Among them, there were 20 individuals aged 30 and above during this period, while the remaining 53 individuals were in the group under 30 years old. The average age of mothers aged 30 and above was calculated as 34.05, whereas it was 22.84 for those aged 30. The average age of fathers at the beginning of pregnancy was determined as 29.34. Among them, there were 29 individuals aged 30 and above during this period, while the remaining 44 individuals were under 30. After performing Correlation Analysis (Spearman's Rank Correlation), it was found that the mother's age had a non-significant and negligible monotonic relationship with the number of de novo mutations (rho = -0.013, p = 0.922). Similarly, it was determined that the father's age had a weak and non-significant monotonic relationship with the number of de novo mutations (rho = 0.012, p = 0.932). Regression analysis was conducted to investigate whether a regression model including maternal and paternal ages was formed to predict the number of de novo mutations. After the analysis,

it was found that the model including the age of the mother and father explained only a small portion (1.2%) of the variance in the number of de novo mutations. An ANOVA analysis was conducted to determine the overall significance level of the total model, which was found to be statistically non-significant (p = 0.724). This indicates that the combination of the mother's and father's ages does not significantly predict the number of de novo mutations. As a general conclusion from these analyses, it was concluded that the number of de novo mutations showed a weak correlation with age variables and that de novo mutations cannot be attributed solely to either maternal or paternal age.

A total of 407 de novo variants were detected in the genes included in CES in the children of the families. The distribution of these variants revealed that 225 were variants of unknown significance (VUS) (55.28%), 93 were Likely Benign (LB) (22.85%), 80 were Benign (B) (19.66%), 8 were Likely Pathogenic (LP) (1.97%), and 1 was Pathogenic (P) (0.25%) (Figure 1). The variants were distributed as 79 (19.4%) missense, 13 (3.19%) frameshift, 32 (7.86%) splicing region, 3 (0.74%) start codon loss, 3 (0.74%) start codon gain, 8 (1.97%) inframe deletion, 13 (3.19%) inframe insertion, 55 (13.5%) synonymous, 165 (40.54%) 3'UTR, and 36 (8.84%) 5'UTR variants (Figure 2). When classified as deletion, insertion, IN/DEL, and substitution, 100 deletion, 103 insertion, 11 IN/DEL, and 193 substitution type variants were found. Looking at the base change profile, the most common changes were as follows: C -> G (12.87%) (85), 68 G -> A (10.3%), 43 A -> G (6.51%), 40 T -> C (6.06%), 26 A -> C (3.9%), 26 C -> A (3.9%), 21 CT -> C (3.2%), 21 T -> TA (3.2%), and 21 TA -> T (3.2%) (Figure 3). Of these changes, 151 were transitions (C <-> T, A <-> G) (37.1%), and 256 were transversions (other changes) (62%) (Figure 4).

When looking at the most commonly observed genes for de novo mutations, they were distributed as follows: DSPP in 8 individuals (14.28571%), HPS4 in 8 individuals (14.28571%), VCL in 8 individuals (14.28571%), BMP4 in 7 individuals (12.5%), ACVR2B in 6 individuals (10.71%), TPK1 in 6 individuals (10.71429%), FLG in 5 individuals (8.92%), FYCO1 in 5 individuals (8.92%), MTPAP in 5 individuals (8.92%), PTEN in 5 individuals (8.92%), ANKRD1 in 4 individuals (7.1%), CACNB2 in 4 individuals (7.1%), EYA4 in 4 individuals (7.1%), LYZ in 4 individuals (7.1%), MUC5B in 4 individuals (7.1%), PDE4D in 4 individuals (7.14%), and RP1L1 in 4 individuals (7.1%) (Figure 5).

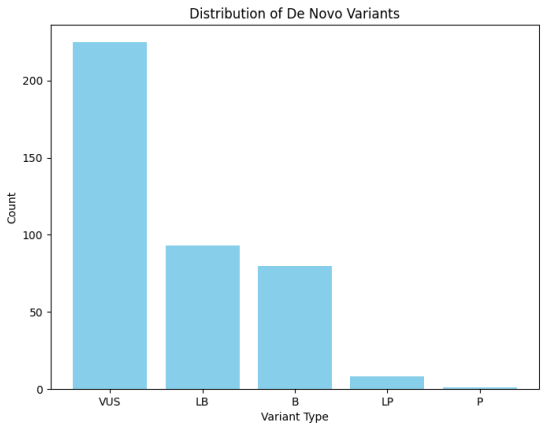

Figure 1. Clinical Significance Distribution

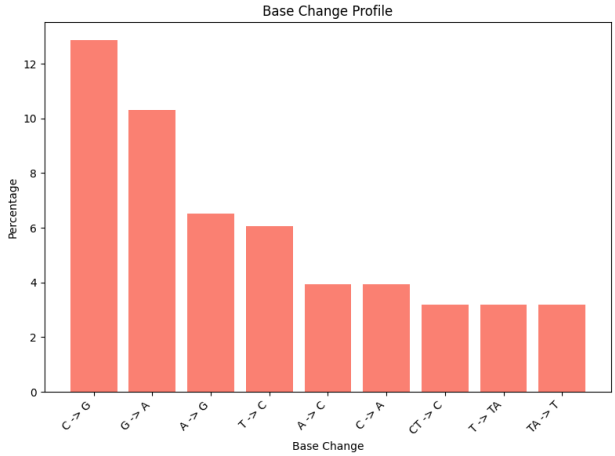

Figure 3. Base change profile of population under study

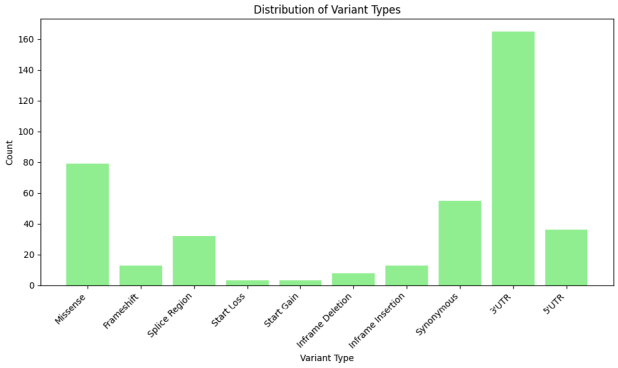

Figure 2. Mutation Type Distribution, considering the 3UTR and 5UTR classification

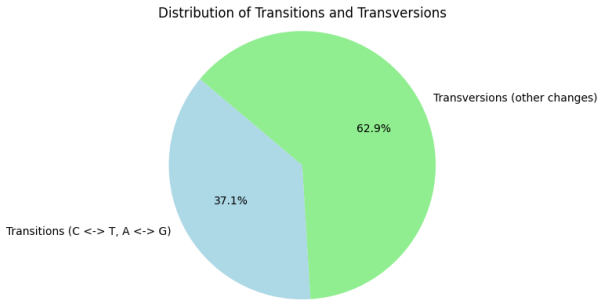

Figure 4. Distribution of transition and transversion mutations

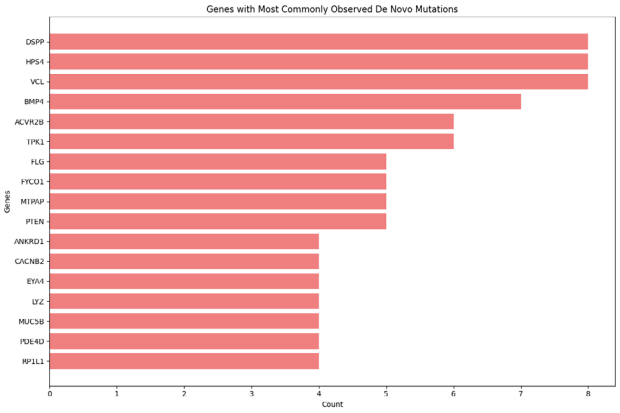

Figure 5. Genes with most commonly Observed De Novo Mutations

Conclusions

Human genome-wide Next Generation Sequencing (NGS) studies provide a range of de novo mutation rates for single nucleotide variations (SNVs) in germ cells, typically ranging from 1.0 to 1.8 × 10-8 [3, 4]. Various recent studies have been conducted to explore the reasons behind this phenomenon, leading to different interpretations. In vitro studies have shown that DNA polymerases ε and δ involved in DNA replication can perform unique base pairings during replication at rates of 10-4 to 10-5 (5, 6). Additionally, it has been observed that replication timing influences the occurrence of these errors. Regions replicating late in the cell cycle exhibit higher mutation rates compared to early replicating regions (7). This phenomenon is thought to be explained by a decrease in dNTP and protein pools contributing to replication during late replication periods [8, 9]. Furthermore, recent whole genome sequencing (WGS) studies have emphasized the clustering of mutations in specific regions of the genome, highlighting the formation of mutation hotspots in these areas (10). Currently, there is no consensus on whether de novo mutations coincide with these hotspots (10). Mutations can be found in regions ranging from 10 to 100 kb within mutation clusters (3, 10). Additionally, the analysis revealed that the rate of transition mutations is higher than that of transversion mutations. This is often associated with base changes in CpG islands abundant throughout the genome. These regions are believed to be particularly prone to errors during replication due to their repetitive nature and sensitivity to methylation during replication (11, 12). The prevalence of transversion mutations in regions harboring mutation clusters is notable. It has been suggested that this could result from dysfunctional replication forks and errors in DNA repair processes (10, 13).

Pioneering trio studies suggest that the majority of de novo mutations identified in the germline originate from the father, and the mutation rate is associated with paternal age (14). Moreover, it has been reported that approximately 80% of the de novo mutations identified in studies originate from the father, and this is correlated with paternal age (14, 15). Limited studies have been conducted to determine the relationship between maternal age and de novo mutations. While some studies suggest a mild relationship between maternal age and an increase in de novo mutations, others have not supported these findings. However, the limited number of studies on this topic complicates interpretation (16, 3, 14). Although limited studies have reported on the influence of maternal de novo mutation formation, the dominant utilization of proteins and enzymes from the mother's cytoplasm during early embryonic development, especially during initial cell divisions, suggests that this maternal factor should potentially be considered in the formation of replication errors at these early stages. Therefore, further research is needed to clarify the possible factors and origins of de novo mutation formation.

The observed distributions of variant pathogenicity in our study indicate that the majority of variants identified as de novo mutations are of unknown significance (VUS) [2]. The presence of benign (B) and likely benign (LB) mutations alongside VUSs suggests that de novo mutations may arise as new variants with no definitively established significant pathogenic importance as determined by previous clinical association studies. However, the clinical significance of these mutations should not be underestimated. Even when categorized at the lowest levels of importance, such mutations may contribute to the diversification of probands' clinical presentations or susceptibilities. Therefore, reporting and archiving these mutations should be considered for future follow-ups on evidence levels and potential revisions. Additionally, these mutations may harbor digenic or polygenic profiles and require further investigation. Indeed, the process of determining variant pathogenicity requires comprehensive evaluation of evidence from the literature, variant databases, and population studies, although these are still insufficient today. While large-scale population studies provide valuable data, the occurrence of variants in these databases should not be the sole criterion for considering variants benign. Public variant databases may contain outdated or conflicting data, underscoring the importance of primarily consulting the literature for variant classification. When assessing variant frequency in the general population for Mendelian diseases, factors such as disease inheritance models, prevalence, and penetrance should be considered. Therefore, high allele frequency in the general population does not exclude pathogenicity in variants. Although gene databases are valuable resources, they should not be the sole criterion for determining pathogenicity when it comes to de novo mutations. A comprehensive evaluation of statistical and functional evidence together is important for accurately assessing the clinical significance of a variant (2, 18).

In this study, the most commonly encountered type of mutation is 3'UTR variants, constituting 40.54% of mutations (17). In the 5'UTR region, this rate is 8.84%, and variants in this region can affect post-transcriptional regulation, mRNA stability, and translation efficiency

(18-20). Disruptions in UTRs can contribute to irregular gene expression and susceptibility or progression of diseases. Following 3'UTR variants, missense mutations account for 19.4% of the identified mutations. These variants can alter protein structure and function, leading to various phenotypic outcomes. They often play a role in Mendelian and complex diseases, highlighting the importance of such variants in disease etiology and therapeutic approaches. Synonymous variants occur with a frequency of 13.5%. Although they do not change the amino acid sequence, synonymous variants play a regulatory role in gene expression and protein function, affecting mRNA stability, splicing efficiency, and translation kinetics. Splicing site variants represent 7.86% of the identified mutations, representing disruptions in splicing consensus sequences. These variants can lead to abnormal splicing patterns and the production of dysfunctional protein isoforms, playing roles in various diseases and emphasizing their importance in understanding disease mechanisms and developing targeted therapies. Variants causing reading frame shifts have a frequency of 3.19%. These mutations often lead to early stop codon formation and subsequently short, non-functional protein products, associating frameshift variants with severe forms of diseases. Loss and gain of start codon variants are each detected at a frequency of 0.74%. These variants affect translation initiation sites and result in the complete loss of protein length. Therefore, these variants have highly destructive effects on the respective genes, resulting in an increase in disease phenotype. De novo mutations occurring in protein-coding genes are classified into three classes in the literature based on the aforementioned effects: 1) likely gene-disrupting SNVs (LGD-SNV) (stop codon, frameshift, splice donor, and acceptor), 2) missense, and 3) synonymous mutations. The impact of such mutations has been extensively studied in various types of diseases, such as neurodevelopmental disorders (NDDs); LGD and missense mutations are more frequently encountered in patients with NDDs (21). On the other hand, synonymous mutations, which play a role in regulating gene expression, are associated with both NDDs and more broadly with neuropsychiatric disorders (20, 21).

Interestingly, the frequency of de novo mutations observed in genes located on chromosomes 10 and 3 highlights a potential co-regulation or functional relationship due to their proximity. Chromosome 10 hosts a cluster of genes such as VCL (Vinculin), which is located at cytoband 10q22.2 and is involved in cell-cell adhesion and cell-matrix interactions. MTPAP, located at cytoband 10p12.31, is vital for polyadenylation of mitochondrial RNA transcripts. PTEN, located at cytoband 10q23.31, functions as a tumor suppressor gene regulating cell growth and survival. ANKRD1, located at cytoband 10q23.33, plays a role in muscle function and cardiovascular development. CACNB2, located at cytoband 10p12.33, also plays a role in calcium channel regulation. Chromosome 3 also harbors a pair of genes including ACVR2B (Activin A Receptor Type IIB) and FYCO1 (FYVE and Coiled-Coil Domain Autophagy Adaptor 1). ACVR2B functions as a receptor for activin and influences various cellular processes, while FYCO1 regulates autophagosome traffic.

Interestingly, out of the 31 DNM mutations detected in genes on chromosome 10, 28 (90%) are located in UTR regions (23 in 3'UTR and 5 in 5'UTR), and three are in splicing-associated regions. Among these, 25 are insertion/deletion and the rest are SNVs. The grouped genes on chromosome 3, all containing 3'UTR, comprise 11 mutations associated with insertion/deletion mechanisms. These two chromosome groups collectively contain 42 DNMs, the majority of which (85.7%) are of uncertain significance (VUS). Interestingly, exceptions are noted in MTPAP (5 out of 6) and CACNB2 (1 out of 6), where the density of B and LB variants stands out. Intriguingly, all 6 LB and B variants in these genes come from regions other than 3'UTR. These findings underscore the importance of 3'UTR regions in determining pathogenicity and warrant further investigation into the mechanisms of de novo mutations.

However, without considering the classification of 3'UTR and 5'UTR regions, out of 408, 203 are associated with CNV (<50bp), and 193 are associated with SNV mechanisms. These findings contrast with some other studies showing the predominance of SNVs (1 bp), but the limitations and statistical biases of these studies should be considered (21).

Nucleotide changes in DNA sequences are described as underlying molecular events, categorized as transitions and transversions; transitions occur more frequently than transversions and lead to a higher transition/transversion ratio across the genome. Transitions are often attributed to the variability of CpG dinucleotides. Methylation of cytosine in CpG dinucleotides forms 5-methylcytosine (5-mC), which is chemically unstable and prone to deamination, leading to G:T mismatches. CpG dinucleotides exhibit significantly higher mutation susceptibility compared to other dinucleotides. Interestingly, the mutation susceptibility of CpG dinucleotides varies across genomic regions. Contrary to expectations, CpG-rich regions exhibit a lower mutation rate compared to the rest of the genome. This difference is attributed to factors such as lower methylation levels, selective pressures associated with gene regulation, or physical prevention of spontaneous deamination due to stronger DNA binding. Understanding mutational signatures associated with specific mutational processes is crucial for determining the underlying mechanisms leading to genetic variations. Mutational signatures characterized by different mutation patterns have been identified in somatic cells, and correlations between these signatures and de novo mutations have been observed. Mutational signatures representing a significant portion of germline de novo mutations, signatures 1 and 5, are associated with high rates of C -> T transitions and A -> G transitions in CpG dinucleotides, respectively. Although the exact mechanisms underlying these signatures are unclear,

they likely involve processes such as deamination of methylated cytosine and spontaneous deamination of adenine. The presence of these mutational signatures has potential implications for genetic variations in both somatic and germ cells, necessitating further investigation into these mechanistic bases. However, our study indicates the predominance of transition mutations over transversion mutations. This finding may be a result of including the CNV mechanism in our population. CpG island analysis for our patients was also conducted using the UCSC online database, but no correlation was found. On the other hand, some mutations, such as those associated with HPS4, DSPP, and PTEN, were found in regions with a GC content of over 50%. This may contribute to a higher mutation rate in these regions, but the situation is different for other genes with the highest DNM counts, such as VCL, which exhibit high GC content in the mutation region.

Tandem Repeats (TRs) were also checked using the UCSC online database platform. Interestingly, some genes with a high number of DNMs were exactly located at TR sites. Genes such as CACNB2, ACVR2B, PTEN, HPS4, ANKRD1, and DSPP had variants found at TR sites. However, variants in the VCL gene were located outside of this region (22).

Overall, we identified de novo mutation (DNM) variants in a total of 242 genes. Pathway analysis for these genes was conducted using the Reactome online platform. The obtained analysis was transferred to a CSV file, filtered based on p-value < 0.05, and sorted according to the filters defined in the pathway analysis. The signaling transduction pathway (R-HSA-162582) was identified as the most common pathway among the identified genes (Figure 6). Signaling transduction is a critical cellular process where external signals lead to changes in cellular behaviors. Transmembrane receptors including receptor tyrosine kinases (RTKs) and TGF-beta receptors perceive these signals and initiate downstream cascades affecting cellular functions such as cell proliferation and survival. While RTKs activate pathways involving RAF/MAP kinases and AKT, TGF-beta receptors phosphorylate SMAD proteins, regulating gene expression. WNT receptors initially classified as G-protein coupled receptors utilize beta-catenin to regulate gene transcription. Integrins activated by extracellular matrix components influence cell adhesion and shape through cytosolic kinases. Rho GTPases respond to signals by altering cytoskeletal organization, affecting cell polarity and connections. These mechanisms enable cells to dynamically respond to their environments. Our analysis shows that our genes primarily contribute to the Hedgehog and Tumor Growth Factor Beta (TGF-BETA) families in this pathway. Studies indicate that Hedgehog signaling activates a mammalian heterochronic gene regulatory network controlling differentiation timing among cell lineages of different origins. Moreover, these genes exhibit a high tendency for de novo mutation rates, supporting our findings (23, 24, 15, 16). However, it is important to validate this hypothesis with further studies.

Declaration of Conflicting Interests: The authors declare that they have no conflict of interest.

Funding: No funding.

Ethical aspects of the research: This study was conducted with the approval of the ethics committee of Selcuk University Medical Faculty Non-Interventional Research Ethics Committee dated 16.04.2024 and numbered 2024/198. All procedures in the study were performed according to the World Medical Association Declaration of Helsinki.

Author contributions: Conception: Nadir Kocak , Ali Torabi , Batuhan Sanlıturk , Özkan Bagcı , Ebru Marzioglu Ozdemir , Tulun Cora. Design: Nadir Kocak , Ali Torabi , Batuhan Sanlıturk. Supervision: Nadir Kocak. Resource: Ozkan Bagcı , Ebru Marzioglu Ozdemir , Tulun Cora. Materials: Nadir Kocak , Ali Torabi , Batuhan Sanlıturk. Data Collection and/or Processing: Ozkan Bagcı, Ebru Marzioglu Ozdemir , Tulun Cora. Analysis and/or Interpretation: Nadir Kocak , Ali Torabi , Batuhan Sanlıturk. Literature Review : Nadir Kocak , Ali Torabi , Batuhan Sanlıturk. Writer: Nadir Kocak Critical Review: Nadir Kocak.

References

1.Rocio Acuna-Hidalgo, Joris A. Veltman, and Alexander Hoischen. New insights into the generation and role of de novo mutations in health and disease. Genome Biology. 2016; 17: 241

2.H Duzkale, J Shen, H McLaughlin, A Alfares, M A Kelly, T J Pugh, B H Funke, H L Rehm, M S Lebo. A systematic approach to assessing the clinical significance of genetic variants. Clin Genet. 2013 Nov;84(5):453-63.

3.Goldmann JM, Wong WSW, Pinelli M, Farrah T, Bodian D, Stittrich AB, et al. Parentof-origin-specific signatures of de novo mutations. Nat Genet. 2016; 48: 935–9.

4.Conrad DF, Keebler JEM, DePristo MA, Lindsay SJ, Zhang Y, Casals F, et al. Variation in genome-wide mutation rates within and between human families. Nat Genet. 2011; 43: 712–4

5.Korona DA, LeCompte KG, Pursell ZF. The high fidelity and unique error signature of human DNA polymerase. Nucleic Acids Res. 2011; 39: 1763–73.

6.Schmitt MW, Matsumoto Y, Loeb LA. High fidelity and lesion bypass capability of human DNA polymerase δ. Biochimie. 2009; 91:1163–72

7.Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. Human mutation rate associated with DNA replication timing. Nat Genet. 2009; 41:393–5.

8.Chen CL, Rappailles A, Duquenne L, Huvet M, Guilbaud G, Farinelli L, et al. Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. Genome Res. 2010; 20: 447–57.

9.Koren A, Polak P, Nemesh J, Michaelson JJ, Sebat J, Sunyaev SR, et al. Differential relationship of DNA replication timing to different forms of human mutation and variation. Am J Hum Genet. 2012; 91: 1033–40.

10.Chan K, Gordenin DA. Clusters of multiple mutations: incidence and molecular mechanisms. Annu Rev Genet. 2015; 49: 243–67.

11.Gao Z, Wyman MJ, Sella G, Przeworski M. Interpreting the dependence of mutation rates on age and time. PLoS Biol. 2016;14:e1002355.

12.Makova KD, Hardison RC. The effects of chromatin organization on variation in mutation rates in the genome. Nat Rev Genet. 2015; 16: 213–23.

13.Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer.

Nature. 2013;500:415–21

14.Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, et al. Rate of de novo mutations and the importance of father's age to disease risk. Nature. 2012;488:471–5.

15.Rahbari R, Wuster A, Lindsay SJ, Hardwick RJ, Alexandrov LB, Al Turki S, et al. Timing, rates and spectra of human germline mutation. Nat Genet. 2015; 48: 126–33

16.Francioli LC, Polak PP, Koren A, Menelaou A, Chun S, Renkens I, et al. Genome-wide patterns and properties of de novo mutations in humans. Nat Genet. 2015; 47: 822–6

17.Stella Marie Reamon-Buettner, Si-Hyen Cho and Juergen Borlak. Mutations in the 3'-untranslated region of GATA4 as molecular hotspots for congenital heart disease (CHD). BMC Medical Genetics 2007, 8: 38

18.Samantha L. Schuster, Sonali Arora, Cynthia L. Wladyka, Pushpa Itagi, Lukas Corey, Dave Young, Bethany L. Stackhouse, Lori Kollath, Qian V. Wu, Eva Corey, Lawrence D. True, Gavin Ha, Patrick J. Paddison, Andrew C. Hsieh. Multi-level functional genomics reveals molecular and cellular oncogenicity of patient-based 3' untranslated region mutations. Cell Rep. 2023 Aug 29;42(8):112840

19.Samantha L Schuster, Andrew C Hsieh. The Untranslated Regions of mRNAs in Cancer. Trends Cancer. 2019;5(4):245-262.

20.Tychele N Turner, Evan E Eichler. The Role of De Novo Noncoding Regulatory Mutations in Neurodevelopmental Disorders. Trends Neurosci. 2019;42(2):115-127

21.Amy B Wilfert, Arvis Sulovari, Tychele N Turner, Bradley P Coe, Evan E Eichler. Recurrent de novo mutations in neurodevelopmental disorders: properties and clinical implications. Genome Med. 2017;9(1):101

22.Cody J Steely, W Scott Watkins, Lisa Baird, Lynn B Jorde. The mutational dynamics of short tandem repeats in large, multigenerational families. Genome Biol. 2022 Dec 12;23(1):253

23.Reactome Pathway Database. https://reactome.org/

24.Ivanova, Maria, "Evolutionary conservation and times of action of heterochronic genes" (2023). Thesesand Dissertations. 3169. https://rdw.rowan.edu/etd/3169