Identification of pathogenic gene variants in carpal tunnel syndrome using bioinformatics approaches

Dwi Tiva Widyanti S. Humolungo¹, Ratna Anjani², and Lalu Muhammad Irham¹*, Nanik Sulistyani¹, Muhammad Ma'ruf¹, Danang Prasetyaning Amukti¹, Wirawan Adikusuma², Made Ary Sarasmita³, Suhail Khairi⁴, Barkah Djaka Purwanto³, Suatmim Suatmim³, Lalu Muhammad Harmain Siswanto³, Rahmat Darni Satria⁵, Satriya Pranata⁶, Rockie Chong¹¹

¹Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta 55164, Indonesia
²Department of Pharmacy, University of Muhammadiyah Mataram, Mataram 83127, Indonesia
³Department of Clinical Pharmacy, College of Pharmacy, Taipei Medical University, Taipei 110, Taiwan
⁴School of Nursing, College of Nursing, Taipei Medical University, Taipei 11031, Taiwan
⁵Faculty of Medicine, Universitas Ahmad Dahlan, Yogyakarta 55191, Indonesia 7 PKU Muhammadiyah Bantul Hospital, Bantul, Yogyakarta 55711, Indonesia
⁶Department of Histology, Faculty of Medicine, Universitas Sebelas Maret, Surakarta 57126, Indonesia
⁷Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong 16911, Indonesia 10 Mataram
⁸Department of Clinical Pathology and Laboratory Medicine, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
⁹Clinical Laboratory Installation, Dr. Sardjito Central General Hospital, Yogyakarta 55281, Indonesia
¹⁰Faculty of Nursing and Health Sciences, Universitas Muhammadiyah Semarang, Semarang City, Central Java, Indonesia
¹¹Department of Chemistry and Biochemistry, University of California, Los Angeles, Los Angeles, CA 90095, USA

Abstract. Carpal Tunnel Syndrome (CTS) is one of the disorders that occur in the upper extremities caused by a narrowing in the carpal tunnel so that there is pressure on the median nerve located in the wrist. In this study, pathogenic variants associated with Carpal Tunnel Syndrome (CTS) were prioritized using bioinformatics and genetic data in populations. The study used GWAS data to identify single nucleotide polymorphisms to look for genomic variants associated with CTS. The data obtained is then continued using HaploReg and GTEx portal for analysis with ensembles. Furthermore, the results of the GTEx portal identified genetic variants with gene expression throughout human tissue. The results obtained obtained two gene variants, namely rs61749613 encoded by VCAN and rs62621197 encoded by ADAMTS10. Of the two variants, the gene as a whole can be expressed in the aortic tissue. The allele frequency distribution of the two gene variants obtained different results from each continent.

1 Introduction

Carpal Tunnel Syndrome (CTS) is one of the disorders that occur in the upper extremities caused by narrowing in the carpal tunnel so that there is pressure on the median nerve located in the wrist. Symptoms that often appear in CTS are pain, numbness, paresthesia, and weakness in the median nerve.

The National Health Interview Study (NHIS) in 2010 estimated that the prevalence of carpal tunnel syndrome (CTS) reported among the adult general population with an incidence of 329 cases per 100,000 people per year was 3.9% (8.3 thousand). CTS is more common in women with a percentage (4.5%) compared to men (1.9%) with the highest prevalence at the age of 45-64 years [1]. The International Labour Organization (ILO) report shows that Carpal Tunnel Syndrome (CTS) is almost every year found in every case of occupational diseases in several countries, even in China in 2010 there was an increase in the number of cases of occupational CTS by approximately 30% compared to 2001 (ILO, 2013). While in the United States the incidence of CTS reaches 1-3 cases per 1000 people per year, with a prevalence rate of 50 cases per 1000 people. The incidence percentage is 5%, lower than in the UK, the incidence of CTS reaches 7%-16%.

In Indonesia, the incidence of Carpal Tunnel Syndrome is not known with certainty due to the lack of incident reports [2]. However, based on research conducted by Napasa et al, (2019) it was found that 38 (70.4%) of 54 people were positive for Carpal Tunnel Syndrome (CTS) with the most common symptoms found were 54% experiencing pain complaints, 44% tingling complaints, 37% nocturnal complaints, 26% numbness complaints, and 22% with complaints of grip weakness.

* Corresponding author: lalu.irham@pharm.uad.ac.id

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).
2 Method

Genomic information can not only be used to identify disease related gene variants but can also be utilized for knowledge about the disease. CTS is a step disease that can cause damage to the carpal caused by There is an increase in pressure in the carpal tunnel structure which causes compression of the median caused by work. In this study, we used a bioinformatics-based approach to prioritize pathogenic variants that have the potential to trigger CTS. Complete information about the research design can be seen in Figure 1. We used the keyword "Carpal Tunnel Syndrome" to obtain CTS using the genome-wide (GWAS) National Human Geographic database Catalog nome Research Institute (NHGRI) (https://www.ebi.ac.uk/gwas) CTS associated with single nucleotide polymorphism (SNP) obtained 83 then further analysis using HaploReag (version 4.2).

The results of the subsequent analysis resulted in a total of 12 SNPs with significant p-values < 10^-8. This value can be used to identify the relationship between genetic variants and gene expression. then identification is carried out to evaluate genetic variation and gene expression using the GTex portal (https://www.gtexportal.org/home). Results are obtained through gene expression from various tissues. VCAN and ADAMTS10 are present in genetic variants present in human skin and tissue exposed to sunlight obtained from the portal GTex database. Then confirmation of the variant using ensembl (https://asias.ensembl.org/index.html) CTS-related genetic variants were evaluated on several continents including Africa, the Americas, East Asia, Europe and Southeast Asia.

![Fig. 1. Bioinformatics Workflow for identification of genetic variation bound to CTS. GWAS, Genome-wide association studies, SNPs, single nucleotide polymorphisms](image)

3 Results and discussion

SNPs associated with CTS were identified from the GWAS database, and resulted in 83 SNPs tied to CTS. After duplicating all SNPs, the result is obtained 12 unique SNPs associated with CTS, as shown in table 1. Based on the number of SNPs obtained, SNPs are further restricted and prioritized using HaploReg version 4.2, with p values < 10^-8. Based on the findings presented in table 2, we obtained two genomic variants of the same gene that qualify as biological risk SNPs for CTS from this study.

<table>
<thead>
<tr>
<th>Allele variants and risks</th>
<th>p-value</th>
<th>Gene code</th>
<th>Allele Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>RS5758007</td>
<td>9 x 10^-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS57822625</td>
<td>9 x 10^-10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS604168</td>
<td>1 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS6117726</td>
<td>1 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS61749613</td>
<td>1 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS62422907</td>
<td>2 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS62621197</td>
<td>2 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS66525731</td>
<td>2 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS6712473</td>
<td>3 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS67245807</td>
<td>3 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS6843953</td>
<td>4 x 10^-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RS6977081</td>
<td>4 x 10^-9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2. CTS risk variants and prioritized SNP allele encoding

Through an integrative bioinformatics approach, we obtained two gene variants: rs61749613 which codes for the VCAN gene and rs62621197 which codes for the ADAMTS10 gene obtained as a biological risk SNP for CTS. Carpal tunnel syndrome (CTS) is a symptomatic compression neuropathy of the median nerve at wrist level characterized by hand pain, numbness, and tingling in the distribution of the median nerve (thumb, index, middle finger, and radial side of the ring finger) as well as decreased grip strength and hand function. The severity of symptoms can be clinically categorized into mild, moderate, and severe CTS caused by pinching of the median nerve as high as the carpal tunnel, bounded by carpal bone and by transverse carpal ligament. Physiological evidence suggests increased pressure within the carpal tunnel, and hence decreased median nerve function at such levels. Symptoms of CTS can include pain, paresthesia and numbness in the distal distribution of the median nerve which includes the thumb, index, middle finger and radial side of the ring finger as well as reduced grip strength on the affected hand. The pain felt is like burning, stabbing, or numbness [3–5].

VCAN (Versican) is a protein-coding gene where Versican is a large proteoglycan chondroitin sulfate (CSPG) found in the extracellular matrix (ECM) of most soft tissues [6]. Usually, it is present in low amounts, but increases dramatically when the tissue is inflamed. VCAN is first isolated from cultured fibroblast media and then cloned from the placenta. fibroblasts and are named versican in recognition of their domain structure and versatility as highly interactive molecules. shows the importance and highly conserved properties of this proteoglycan In normal tissues, versican levels are low but still high [7]. VCAN disease increases dramatically, as seen in early intima vascular lesions typical of the development of human atherosclerosis [8]. In addition,
VCAN accumulates in some cancers and many other cancers or diseases as well. VCAN is negatively charged due to its glycosaminoglycan (GAG) chain and attract water, contributing to the viscoelasticity of the pericellular microenvironment [9]. In addition, versicans interact with a number of nearby ECM components on the cell surface including HA, tenascin-R and -C, thrombospondin 1, fibronectin, and fibrillin to create mechanically active biopolymers around the cell that affect the cell's ability to deform, attach, multiply, migrate, and assemble other ECM components and survive. Versican molecules, and the ECM that binds to VCAN, can change mechanical rigidity around cells contributing to mechanotransduction changes affecting cell behavior and phenotype [11].

ADAMTS10 protein is a member of the ADAMTS family of secreted metalloproteinases that contribute to the formation and turnover of the extracellular matrix (ECM). This protein has been shown to have an important role in the storage and activation of TGF β, known as a regulator of collagen turnover, suggesting that ADAMTS10 mutations can cause changes in the composition, structure, and mechanical properties of ECM [12]. The ADAMTS family includes nineteen ADAMTS proteases and seven ADAMTS-like proteins, which are superfamilies of glycoproteins in the ECM or residues associated with the cell surface after secretion. ADAMTS proteases have a highly homologous N-terminal protease domain consisting of signal peptides to target them for secretion, pro-peptides that are normally eliminated by pro-protein furin/PACE convertase to activate ADAMTS proteases, the metalloproteinase catalytic domain itself, and a disintegrin-like domain. Their C-terminal auxiliary domain has a core region consisting of type 1 disintegrin-like domain. Their C-terminal auxiliary domain requires substrate recognition, binding, specificity, and cell surface tethering or ECM (Silou, 2007). They also undergo C-terminal processing that affects substrate specificity and ECM binding [13]. The domain requires Fbn-1 microfibrils for localization in the ECM. It regulates the assembly of microfibrils and accelerates the assembly of microfibrils through direct interaction with fibrillin-13. The protein ADAMTS10 consists of a pro-peptide domain that is expected to be cleaved as it exits the cell and (marked by a black arrow-furin processing site), followed by a catalytic domain and a disintegrin-like domain with thrombospondin repeats that can interact with ECM19. Patient variation is localized in the pro-domain of the ADAMTS10 protein and replaces evolutionarily conserved amino acids within the conserved region [12].

3.1 VCAN gene expression and ADAMTS10 in human tissue

To evaluate VCAN and ADAMTS10 expression within human tissues, the GTEx Portal database (https://www.gtexportal.org/) is used, which contains gene expression levels in various tissues as shown in Figure 2, where the eQTL notation comprises the most pronounced functional consequences of genetic variants. The VCAN gene is expressed in the adult brain, cardiovascular system, skin and musculoskeletal tissue [14]. VCAN genes exist primarily as large aggregates with hyaluronan (HA) glycosaminoglycans [6]. Similar traits to other members of the hyalectan (or lectican) family, which include aggrecan, neurocan, and brevican. In contrast to the versican, members of this family have a limited distribution, with aggregan mainly confined to cartilage, and neurocan and brevican selectively present in the central nervous system. VCAN genes have significance in biology, especially during embryogenesis and pathogenesis of some human diseases. VCAN is essential for myocardial and valve development, and of course, blood-deprived VCAN tixus will die mid-pregnancy from heart anomalies. VCAN is involved in the migration of nerve crest cells, and is necessary for musculoskeletal development [15]. VCAN mutations that alter the normal state of exon splicing VCAN in humans give rise to a rare eye disorder called Wagner syndrome, and a related condition called erosive vitreoretinopathy [16]. VCAN has been extensively investigated in the context of acquired abnormalities in humans, specifically cardiovascular disorders such as atherosclerosis and arterial stenosis and more recently, in various types of cancer, the presence of which is usually associated with increased tumor malignancy [12].

ADAMTS10 expressed in the skin, chondrocytes of the fetus, and adults [17]. The manifestation of ADAMTS10 gene abnormalities in humans suggests that they participate in the structural and regulatory role of microfibrils in extracellular matrix (ECM) function and eye, heart, and skeletal development [18]. ADAMTS10 gene regulates microphytophilic biogenesis, fibrillin can bind to Fbn1 with high specificity and affinity at two sites and collocate with Fbn1 in human tissue. When added to fetal bovine ligament cells or fibroblast cultures, it accelerates biogenesis and microfibril assembly [19]. The tissue microfibrils formed by Fbn1 are important structural components of ECM. Fbn1 also regulates TGFβ signaling and biogenesis as well as the balance of elastic fibers, which support the skin, ligaments, and blood vessels. Fibrillin microfibrils are the core structure of the lenticular zonula, which is disrupted in WMS, and causes lens dislocation [20]. Other genes mutated in WMS code for ECM proteins localized to the Fbn1 microfibrils in the ECM or may be involved in the formation of microfibrils. Heart abnormalities seen in patients include ductus arteriosus patent, pulmonary stenosis, aortic stenosis, and mitral valve prolapse. Thoracic aortic aneurysms and cervical artery dissection were reported in three-generation families with FBN1-associated WMS [21]. Systematic electrocardiograms show prolonged QT in individuals with WMS [22].
3.2 Frequency of candidate alleles of carpal tunnel syndrome (CTS) disease variants on different continents

After identifying candidate variants related to VCAN and ADAMTS10 gene expression, we determined allele frequencies in all populations across continents, as shown in Table 3. The quantity-free alleles for both variants were evaluated in different populations including African, American, East Asian, European, and Southeast Asian populations. In the rs61749613 allele variant, samples from each region consisted of 1 individual (Africa), 22 individuals (Americas), 1 individual (East Asia), 49 individuals (Europe), and 27 individuals (Southeast Asia). While the variant allele rs62621197, samples from each region consisted of 4 individuals (Africa), 24 individuals (America), 1 individual (East Asia), 53 individuals (Europe), and 4 individuals (Southeast Asia). We extracted allele frequencies in Africa, the Americas, East Asia, Europe, and Southeast Asia from the Ensemble Genome Browser (http://www.ensembl.org). Allele frequencies between populations are different for each population of VCAN and ADAMTS10 variants. Table 3 and Figure 3 show gene expression levels at higher frequencies of the associated allele (G) rs61749613 and the associated allele (T) rs62621197. At the population frequency of the rs61749613 allele (G), the populations of East Asia and Africa are expressed at a much lower level than the populations, Americas, Southeast Asians, and Europeans. The frequency of the T allele allele “rs62621197” in African and Asian populations (East Asia and Southeast Asia) is expressed at a much lower level than in American and European populations. Overall, the allele frequencies of the variants “rs61749613” and “rs62621197” indicate a contribution of differential variant prevalence to VCAN and ADAMTS10 gene expression. Associations with various VCANs and gene ADAMTS10 can then be analyzed using publicly available databases. We use publicly available databases such as GTEx Portal and Ensembl.

Allele frequencies in all populations are different for each SNP, as shown in Figure 3. In conclusion, by utilizing a bioinformatic-based approach, pathogenic variants potentially associated with CTS were revealed. The gene variants of this study were identified using in silico methods so that they must be validated first for clinical implementation (Irham et al., 2023). We propose that this variant could be used for further research to identify diagnostic biomarkers of CTS as well as prognosis. However, we acknowledge that there are limitations to the bioinformatic-based approach used to investigate genetic variants associated with CTS. One limitation is that not all variants have genes that encode them (variants that do not code), and even if they do, those genes or genetic variants may not be suitable drug targets. Nonetheless, clinical validation is needed as a next step to confirm our findings and gain a better understanding of the underlying etiology and functional effects of CTS disease.

Table 3. Allele frequency analysis for VCAN genes and ADAMTS10 from SNP variant annotations

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Position (hg38)</th>
<th>Gene Symbol</th>
<th>Lokasi</th>
<th>Allele</th>
<th>Allele Frequency (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr 5:83519351</td>
<td>VCAN Missense</td>
<td>A</td>
<td>G:</td>
<td>0.003</td>
<td>1.00</td>
</tr>
<tr>
<td>Chr 19:8605262</td>
<td>ADAMTS10 Missense</td>
<td>C</td>
<td>T:</td>
<td>0.003</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Fig. 2. CTS-related gene expression in multiple human tissues from GTEx Portal
Table 3 and Figure 3 show the presence of gene expression levels at higher frequencies in rs62621197 with associated allele codes (T), rs61749613 and associated allele codes (G). The T code allele with SNP at rs62621197, in African and East Asian populations (0), is therefore expressed lower compared to populations in the Americas, Southeast Asia and Europe. Allele frequencies in G codes with rs61749613 and rs62621197 showed the highest expression followed by America, Southeast Asia, East Asia and Europe. Allele frequencies in VCAN variants rs61749613 and ADAMTS10 rs62621197 which are genomic variants associated with CTS. Therefore, it can be assumed that Europe, American, and Southeast Asian populations have a higher potential to develop CTS with increased expression of VCAN gene variants rs61749613 and ADAMTS10 rs62621197 compared to Africa and East Asia.

4 Conclusion

In this study, we conducted a comprehensive bioinformatic analysis of CTS from the genome database, from the results obtained there were differences in tissue expression in the VCAN gene rs61749613 and ADAMTS10 rs62621197 in human tissue. VCAN rs61749613 can be expressed in aorta, lung, and lymphocyte cells while ADAMTS10 gene rs62621197 is expressed in aorta and cervical cancer. Overall, alleles for rs61749613 and rs62621197 have the lowest frequency in Africa (rs61749613) and East Asia (rs62621197) than in American, Southeast Asian and European populations.

References

5. D. Z. Farqihan and A. Taufik, J. Kesehat. Qamarul Huda 10, 177 (2022)