國立臺灣大學醫學院基因體暨蛋白體醫學研究所
博士論文
Graduate Institute of Medical Genomics and Proteomics
College of Medicine
National Taiwan University
Doctoral Dissertation

# 胸主動脈瘤與剝離症侯群的系統性基因分析及結構變異檢測算法在單一或組合策略下的性能比較

# Systemic Genetic Analysis for Thoracic Aortic Aneurysm and Dissection (TAAD) and Comparison of Performances of Structural Variants Detection Algorithms in Solitary or Combination Strategy

段德敏
De-Min Duan

指導教授：陳沛隆 博士
Advisor: Pei-Lung Chen, MD, PhD

中華民國 114 年 6 月
June 2025

doi:10.6342/NTU202500995

# 國立臺灣大學博士學位論文
# 口試委員會審定書
## DOCTORAL DISSERTATION ACCEPTANCE CERTIFICATE
## NATIONAL TAIWAN UNIVERSITY

胸主動脈瘤與剝離症侯群 (TAAD) 的系統性基因分析及
結構變異檢測算法在單一或組合策略下的性能比較

Systemic Genetic Analysis for Thoracic Aortic Aneurysm and Dissection (TAAD) and
Comparison of Performances of Structural Variants Detection Algorithms in Solitary or Combination Strategy

本論文係 段德敏（姓名） D06455002（學號）在國立臺灣大學 基因體暨蛋白體醫學研究所完成之博士學位論文，於民國 114 年 05 月 26 日承下列考試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Graduate Institute of Medical Genomics and Proteomics on 26 (date) May (month) 2025 (year) have examined a Doctoral Dissertation entitled above presented by De-Min Duan (name) D06455002 (student ID) candidate and hereby certify that it is worthy of acceptance.

口試委員 Oral examination committee:

陳洋陞　楊銘堯　柯俐

（指導教授 Advisor）

許秉毓　游沛節　邱馨慧

系（所、學位學程）主管 Director: 陳洋陞

# Acknowledgement

I sincerely express my deepest gratitude to my advisor, Professor Pei-Lung Chen, whose invaluable guidance, unwavering support, and insightful expertise have been instrumental in shaping this study. His encouragement and dedication have profoundly influenced my research journey. Additionally, I am truly appreciative of the support and assistance provided by Dr. Yu-lin Ko, Chief of the Department of Research at Taipei Tzu Chi Hospital. His guidance and the resources made available through his leadership have significantly contributed to the successful completion of this work. Lastly, I extend my heartfelt thanks to my family for their unconditional love, patience, and unwavering belief in me. Their constant encouragement has been a driving force behind my perseverance and determination throughout this endeavor.



doi:10.6342/NTU202500995

# 摘要

胸主動脈瘤與剝離症候群 (TAAD) 是一種隱蔽且潛在致命的疾病。在本研究中，針對來自 NTU TAAD cohort 的 125 位受試者，使用基於次世代定序技術的 NTU TAAD 基因組 (29 genes) 進行篩檢，得出了 46 個基因診斷結果。此基因組篩檢在證實早期診斷及對已知主動脈擴張患者進行剝離風險分層方面展現出高效性。基因組的特異性為 82.4%，陽性預測值 (PPV) 為 87%。然而，其靈敏度為 54.8%，在此世代研究中的檢測率為 43%。值得注意的是，皮膚擴張紋可能作為初步篩查指標，適合在進行進一步、更全面的檢查之前使用。

此外，在我們對結構變異檢測算法的比較研究中，評估了五個先進的結構變異檢測算法和商業軟體 DRAGEN 的性能，使用來自 GIAB v0.6 Tier 1 基準集以及著名的 HGSVC2 基準集的短讀序全基因組數據。每種算法均獨立測試，並且在多樣協議 (multiple agreement) 和聯合 (union) 策略下進行了多種組合測試。聯合策略達到了更高的召回率 (recall rate)，而多樣協議策略則展現出更優秀的精確度 (precision)。我們的分析顯示，每種算法具有內在的優勢和劣勢，導致所檢測到的結構變異類型和大小有所不同。組合策略有效協調了這些差異，提升了總體性能，並達到與商業軟體 DRAGEN 相當的 F1 分數。

在世代研究中，未找到致病基因的 83 位前驅患者中，34 位臨床特徵為主動脈瘤或剝離篩選為 NTU TAAD WGS cohort 的試驗者並接受了全基因組定序和基因變異篩選，過程中使用兩組虛擬基因組篩選變異，分別為專注型 (focused panel) 和廣譜型 (broad spectrum panel)。在這 34 位前驅患者中，經過專注型虛擬基因資組過濾後，在 5 位患者中發現了致病變異點，包括在 *FBN1* 基因中發現的一個單核苷酸變異點 (SNV)、兩個剪接位點變異 (splice site variant) 和兩個結構變異 (structural variant)。廣譜虛擬基因組篩選後每位患者大約識別出 260 個候選變異。為了偵測全新的致病基因，針對剩餘的 29 位患者的廣譜基因組篩選出來之候選變異進行了基因層級的交集分析，此方法對於發現潛在導致胸主動脈瘤和剝離病理的新基因至關重要。

關鍵字：次世代基因定序、胸主動脈瘤與剝離徵候群、基因組篩選、結構變異檢測算法、全基因體定序



doi:10.6342/NTU202500995

# Abstract

Thoracic aortic aneurysm and dissection (TAAD) is a silent yet potentially severe condition. In the NTU TAAD cohort study, 125 probands were screened using a next-generation sequencing NTU TAAD panel comprising 29 genes, leading to 46 genetic diagnoses. The panel effectively confirmed early-stage cases and stratified dissection risk, showing a specificity of 82.4% and a positive predictive value (PPV) of 87%. However, sensitivity was 54.8%, with a detection rate of 43%. Notably, skin striae distensae may serve as an early screening indicator before comprehensive examinations.

Our comparative study evaluated five advanced SV detection algorithms alongside the commercial software DRAGEN using short-read whole-genome sequencing data from the GIAB v0.6 Tier 1 and HGSVC2 benchmark sets. The union strategy improved recall, while the multiple agreement method enhanced precision. Each algorithm exhibited strengths and limitations, affecting SV detection variability. Combining approaches optimized performance, yielding F1 scores similar to those of DRAGEN.

Among 83 probands previously testing negative for disease-causing genes, 34 underwent whole-genome sequencing and variant prioritization in the NTU TAAD WGS cohort. We applied focused (169 genes) and broad-spectrum (3197 genes) virtual panels, identifying disease-causing variants in 5 probands, including 1 SNV, 2 splice site variants, and 2 structural variants in *FBN1*. The broad-spectrum panel yielded roughly 260 candidate variants per proband, and gene-level intersection analysis among the remaining 29 probands provided insights into potential novel genetic contributors to TAAD pathology.

Keywords: next generation sequencing, thoracic aortic aneurysm and dissection, gene panel analysis, structural variants detection, whole genome sequencing



doi:10.6342/NTU202500995

# Table of Contents





doi:10.6342/NTU202500995

doi:10.6342/NTU202500995

doi:10.6342/NTU202500995

# List of figures





doi:10.6342/NTU202500995

# List of tables





doi:10.6342/NTU202500995

# Supplementary





doi:10.6342/NTU202500995

# 1. Introduction

## 1.1 Anatomy of aorta

The aorta, the largest artery in the human body, consists of several segments: the aortic root, ascending aorta, aortic arch, and descending aorta (Figure 1). It is composed of three layers: the intima, media, and adventitia (Figure 2). The intima is the innermost layer, consisting of endothelial cells supported by a connective tissue matrix, playing a crucial role in maintaining vascular health by providing a smooth lining for blood flow and preventing thrombosis. The media, the middle layer, is composed primarily of smooth muscle cells, elastic fibers, collagen proteins, and polysaccharides, organized into over 50 concentric layers known as elastic lamellae, providing the necessary strength and elasticity to withstand and regulate the high-pressure blood flow from the heart. The outer layer, the adventitia, consists of connective tissue, fibroblasts, nerves, and the vasa vasorum, small blood vessels that supply nutrients to the larger artery, providing structural support and housing nerve fibers that regulate vascular function. This multi-layered structure of the aorta ensures it can manage the dynamic and high-pressure environment of the cardiovascular system, maintaining efficient blood distribution throughout the body[(1)].



doi:10.6342/NTU202500995

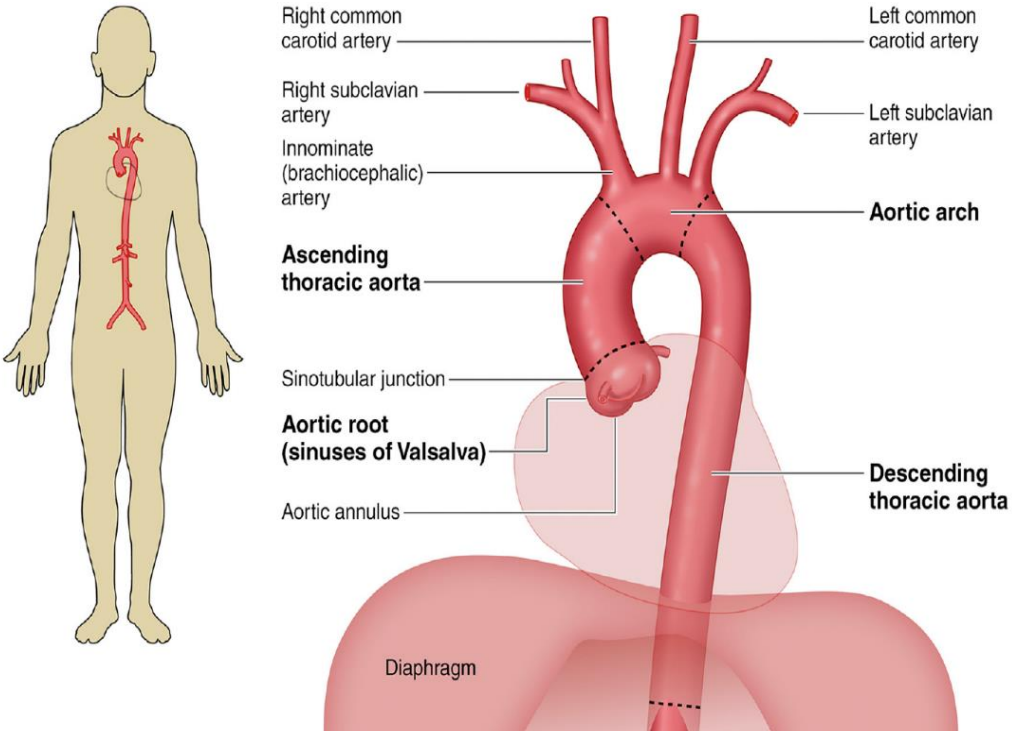



**Figure 1. Normal anatomy of thoracic aorta**[1]

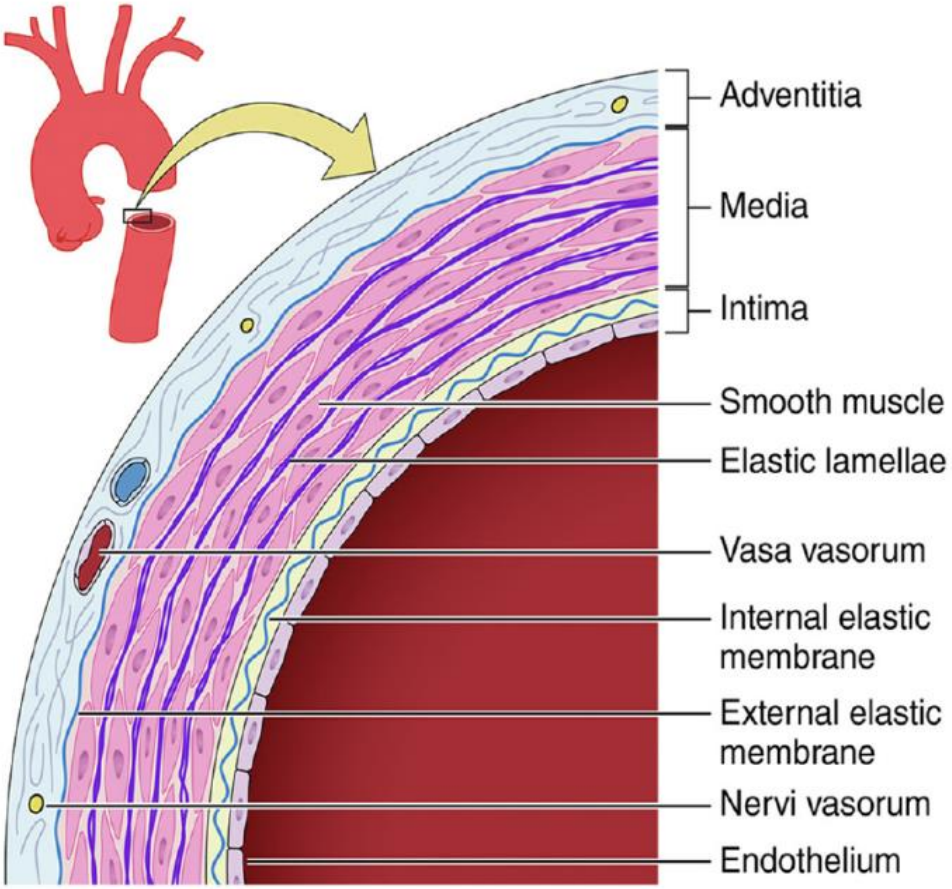



**Figure 2. An illustration of the histologic composition of the aortic wall** [1]



doi:10.6342/NTU202500995

## 1.2 Normal aortic size

Normal aortic size in children varies with age, body size, and sex, much like in adults (Figure 3). Standardized growth charts and Z scores (the numbers of SDs above or below the predicted mean normal diameter) are used to assess and monitor aortic dimensions in pediatric populations. As children grow, their aortic size increases in proportion to their body surface area (BSA), which is why regular monitoring and are crucial[2, 3].

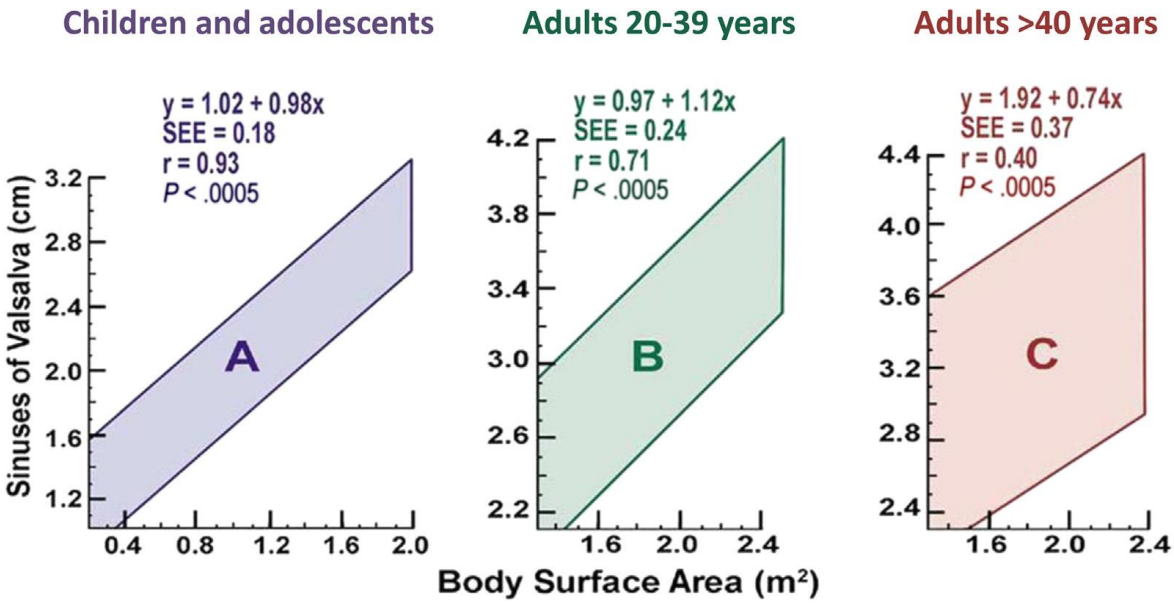



**Figure 3. Sinuses of Valsalva measurements by age and body surface area:** in children and adolescents (A), adults aged 20 to 39 years (B), and adults aged ≥40 years (C)[2, 3]

## 1.3 Definition of aortic aneurysm and dissection

The definition of an aortic aneurysm has evolved over time. Traditionally, an arterial aneurysm is defined as any artery that is dilated to at least 1.5 times its expected normal diameter[4]. However, a more current and specific definition applies to the aorta, where an aortic diameter is considered dilated when the Z score is equal to or greater than 2[5]. An aortic dissection, on the other hand, is a serious condition characterized by a tear in the inner layer of the aorta, the large blood vessel branching off the heart. This tear causes



doi:10.6342/NTU202500995

blood to surge through the tear, creating a separation (dissection) between the aorta's inner and middle layers. If the blood-filled channel ruptures through the outside aortic wall, aortic dissection can often be fatal. Standford system classifies aortic dissections into two types: Type A, which involves the ascending aorta and requires immediate surgical intervention, and Type B, which involves the descending aorta and may be managed medically or surgically depending on the severity of the condition(Figure 4)[6, 7].

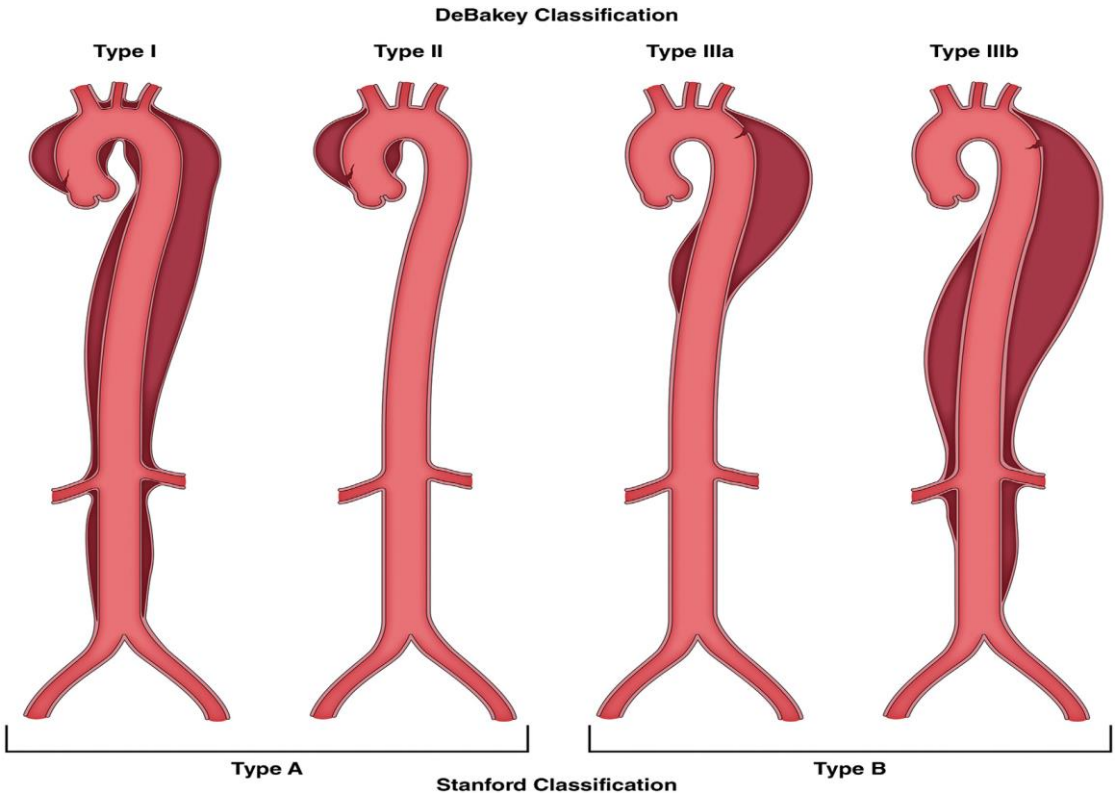



**Figure 4. Standford classification of aortic dissections**[1]

## 1.4 Revised Ghent criteria for diagnosis of Marfan syndrome

Given that the diagnosis of Marfan syndrome is predominantly linked to the risk of aortic aneurysm or dissection, it is crucial to avoid diagnosing MFS in patients without substantial evidence of such risk. An international expert panel has established a revised Ghent nosology, which assigns greater importance to cardiovascular manifestations, with aortic root aneurysm and ectopia lentis being the principal clinical features, presented as



doi:10.6342/NTU202500995

follows[8]:

In the absence of family history:

(1) Aortic aneurysm (Z ≥2) AND ectopia lentis

(2) Aortic aneurysm (Z ≥2) AND disease-causing *FBN1* variant

(3) Aortic aneurysm (Z ≥2) AND Systemic features (≥7pts)

(4) Ectopia lentis AND disease-causing *FBN1* variant with known Aortic aneurysm

Ectopia lentis with or without Systemic features AND with uncertain significant *FBN1* variant or no *FBN1* variant = Ectopia lentis syndrome

Aortic size (Z <2) AND Systemic features (≥5pts with at least one skeletal feature) without Ectopia lentis = MASS phenotype

Mitral valve prolapse AND Aortic size (Z <2) AND Systemic features (<5pts) without EL = Mitral valve prolapse syndrome

In the presence of family history:

(5) Ectopia lentis AND Family history of Marfan syndrome

(6) Systemic features (≥7 pts) AND Family history of Marfan syndrome

(7) Aortic aneurysm (Z≥2 above 20 years old, ≥3 below 20 years) + Family history of Marfan syndrome

Scoring of systemic features

1. Wrist AND thumb sign = 3 (wrist OR thumb sign = 1)
2. Pectus carinatum deformity = 2 (pectus excavatum or chest asymmetry = 1)
3. Hindfoot deformity = 2 (plain pes planus = 1)
4. Pneumothorax = 2



doi:10.6342/NTU202500995

5. Dural ectasia = 2
6. Protrusio acetabuli = 2
7. Reduced upper segment/lower segment ratio AND increased arm/height AND no severe scoliosis = 1
8. Scoliosis or thoracolumbar kyphosis =1
9. Reduced elbow extension = 1
10. Facial features (3/5) = 1 (dolichocephaly, enophthalmos, downslanting palpebral fissures, malar hypoplasia, retrognathia)
11. Skin striae = 1
12. Myopia > 3 diopters = 1
13. Mitral valve prolapse (all types) = 1

Maximum total: 20 points; score ≥7 indicates systemic involvement

## 1.5 Causes of aortic aneurysm and dissection

Etiologies of aortic aneurysm can be broadly categorized into congenital, non-congenital, and other contributing factors such as hypertension, trauma, inflammation, infection, and degenerative causes[9-11].

Congenital causes of aortic aneurysm often involve genetic disorders that affect the connective tissue. These include:

- Marfan syndrome: A genetic disorder affecting the body's connective tissue, leading to cardiovascular complications, caused by variants in the *FBN1* gene.
- Loeys-Dietz syndrome: This syndrome involves variants in the *TGFBR1* or *TGFBR2* genes, resulting in aggressive vascular manifestations.
- Vascular Ehlers-Danlos syndrome: Caused by variants in the *COL3A1* gene, it leads to fragile blood vessels and a higher risk of aneurysm and dissection.



doi:10.6342/NTU202500995

- Smooth muscle dysfunction syndrome: Characterized by variants in the *ACTA2* gene, impacting the aorta's structural integrity.
- Other syndromes: Conditions attributable to pathogenic variants in genes such as *FLNA*, *BGN*, and *LOX*.
- Congenital bicuspid aortic valve: A common congenital heart defect where the aortic valve has only two leaflets instead of three, increasing the risk of aortic aneurysm.
- Turner syndrome: A genetic disorder affecting females, characterized by the partial or complete absence of one X chromosome, and associated with cardiovascular abnormalities, including aortic aneurysm.
- Williams syndrome: A rare genetic disorder caused by the deletion of genes on chromosome 7, leading to cardiovascular issues, including supravalvular aortic stenosis and potential aortic aneurysm.

Non-congenital causes of aortic aneurysm often involve acquired factors or environmental influences. These include:

- Hypertension: Chronic high blood pressure can weaken the aortic wall, leading to aneurysm formation.
- Trauma: Direct injury to the aorta from accidents or surgical procedures can result in aneurysm formation.
- Inflammation: Inflammatory aortitis is a condition characterized by inflammation of the aorta, which can weaken the aortic wall and lead to the formation of aneurysms. This inflammation can be caused by various underlying conditions, including:
    - Giant cell arteritis: Also known as temporal arteritis, this condition



doi:10.6342/NTU202500995

primarily affects the large and medium arteries of the head and neck but can also involve the aorta.

- Takayasu arteritis: A rare form of vasculitis that causes inflammation of the aorta and its major branches, leading to narrowing, occlusion, or aneurysm formation.
- Rheumatoid arthritis: A chronic inflammatory disorder that primarily affects the joints but can also cause systemic vasculitis, including the aortic wall.
- Systemic lupus erythematosus (SLE): An autoimmune disease that can cause widespread inflammation, including in the aortic wall, leading to aortitis.
- IgG4-related disease: A condition characterized by the infiltration of IgG4-positive plasma cells into various organs, including the aorta, causing inflammatory aortitis.

- Infection: Infections like syphilis or bacterial endocarditis can damage the aorta and cause aneurysms.
- Degeneration: Degenerative causes of aortic aneurysm are often related to the aging process and the gradual breakdown of the aortic wall's structural integrity. These include:
  - Atherosclerosis: The buildup of plaques in the arterial walls can lead to weakening and aneurysm formation.
  - Degenerative changes: Age-related degeneration of the elastic and collagen fibers in the aortic wall can result in aneurysm development.
  - Calcification: The deposition of calcium in the aortic wall can reduce its elasticity and increase the risk of aneurysm[(1)].



doi:10.6342/NTU202500995

## 1.6 Genetic diagnosis for Mendelian disease

Genetic diagnosis of Mendelian diseases involves a series of methodical steps to accurately identify pathogenic variants in an individual's genome. Initially, a detailed family history and clinical evaluation are conducted to ascertain the likelihood of a Mendelian inheritance pattern. Following this, genomic DNA is extracted from the patient's blood or tissue samples. Next-generation sequencing (NGS) techniques, such as whole-exome sequencing (WES), whole genome sequencing (WGS), or targeted gene sequencing (TGS), are employed to detect single nucleotide variants (SNVs), small insertions and deletions (indels), and structural variants (SVs) that may underlie the disease[12]. Bioinformatics tools are then utilized to analyze the sequencing data, aligning it to reference genomes and identifying potential pathogenic variants. Confirmatory tests, such as Sanger sequencing, may be performed to validate the findings. Additionally, functional assays and segregation analysis in family members can provide further evidence of the variant's pathogenicity. The results are interpreted in the context of clinical findings and family history, enabling a precise genetic diagnosis and informing personalized treatment and management strategies for the patient.

## 1.7 Distinct types of genetic variants

### 1.7.1 Single nucleotide variants (SNVs)

Single Nucleotide Variants (SNVs) are defined as alterations in the DNA sequence where a single nucleotide base (A, T, C, or G) is changed to another (Figure 5). SNVs can be found approximately every 1,000 nucleotides, resulting in 4 to 5 million SNVs per genome. They can occur within coding regions of genes, potentially altering protein synthesis, as well as in non-coding regions, influencing gene expression and regulation.



doi:10.6342/NTU202500995

Although often neutral, some SNVs can have significant consequences on gene function and health.

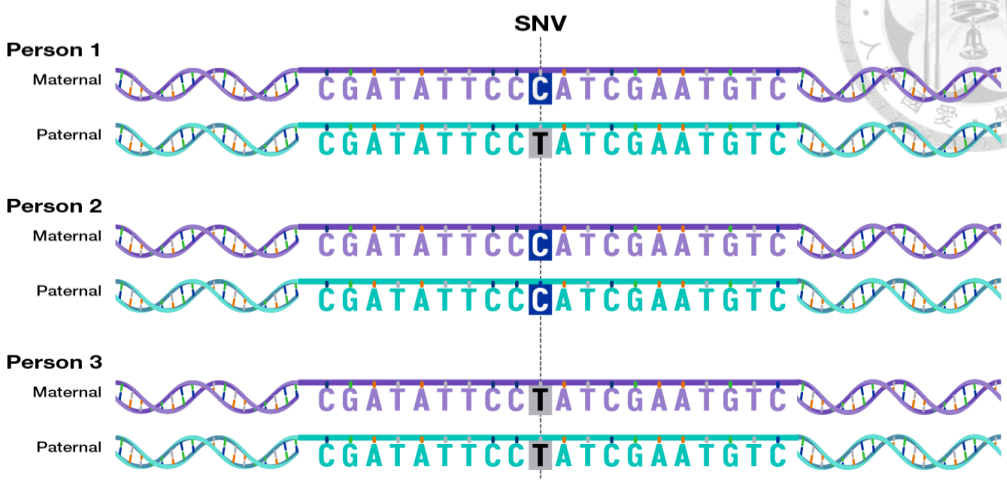



**Figure 5. Schematic illustration of a single nucleotide variant** (downloaded from website of National Human Genome Research institute, https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genomic-variation)

### 1.7.2 Small indels

Small insertions and deletions, commonly referred to as indels, play a significant role in genetic variation and evolution. Indels are defined as the insertion or deletion of small DNA segments, typically ranging from one to a few hundred base pairs (Figure 6). They can occur in coding regions of the genome, potentially leading to frameshift variants that alter the protein-coding sequence, thereby affecting protein structure and function. Alternatively, indels in non-coding regions can influence gene regulation and expression. Despite their smaller size compared to larger genomic alterations, indels can have profound impacts on an organism's phenotype and contribute to various genetic disorders, such as cystic fibrosis and certain forms of cancer. The mechanisms leading to indels include errors during DNA replication or repair processes, and the presence of repetitive DNA sequences that can cause misalignment and recombination.



doi:10.6342/NTU202500995


**Figure 6. Schematic illustration of small indels** (downloaded from website of National Human Genome Research institute, https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genomic-variation)

### 1.7.3 Structural variants (SVs)

Structural variants (SVs) are large-scale alterations in the genome involving segments of DNA typically larger than 50 base pairs, contributing significantly to genomic function, phenotypic diversity, and disease susceptibility. Unlike SNVs, SVs affect larger DNA stretches and can profoundly impact gene function and regulation. Types of SVs include deletions (loss of DNA segments), duplications (copying and inserting DNA segments), inversions (reversal of DNA segments), insertions (addition of new DNA segments), and translocations (exchange of DNA segments between non-homologous chromosomes) (Figure 7). These alterations arise through mechanisms such as Non-Allelic Homologous Recombination (NAHR), Non-Homologous End Joining (NHEJ), and Fork Stalling and Template Switching (FoSTeS). SVs have significant implications for human health, playing a crucial role in genetic disorders (e.g., DiGeorge syndrome, Charcot-Marie-Tooth disease), cancer (e.g., BCR-ABL fusion in leukemia, MYC amplification), and complex diseases (e.g., autism, schizophrenia, cardiovascular diseases). Advances in genomic technologies have improved our ability to detect and



doi:10.6342/NTU202500995

characterize SVs, providing insights into their mechanisms and effects, which is essential for understanding their full impact on health and developing targeted therapies and interventions[13].

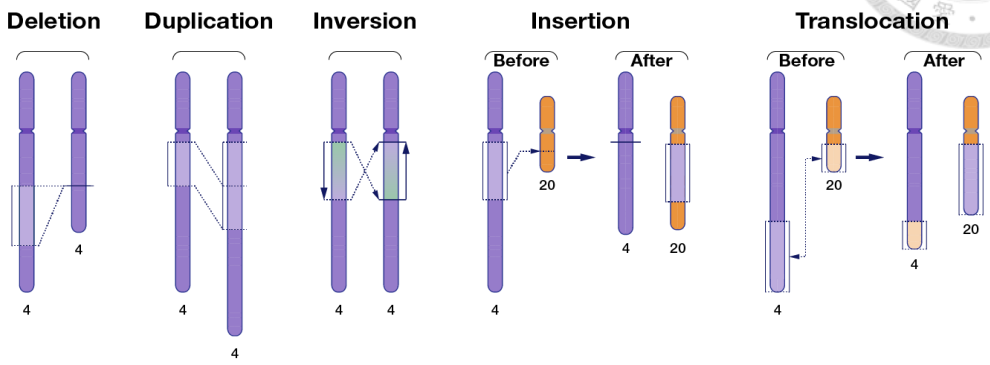



**Figure 7. Schematic illustration of structural variants** (downloaded from website of National Human Genome Research institute, https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genomic-variation)

### 1.7.4 Mobile element insertions (MEIs)

Mobile element insertions (MEIs) represent a fascinating and dynamic aspect of genomic variation. These are segments of DNA that can move from one location to another within the genome, a process facilitated by transposons and retrotransposons. MEIs contribute significantly to genetic diversity, evolution, and can play a role in disease. By definition, MEIs involve the insertion of mobile genetic elements, such as Alu, LINE-1 (L1), and SVA elements, into new genomic locations (Figure 8). These insertions can disrupt gene function or regulation, leading to various phenotypic effects. Alu elements, part of the short interspersed nuclear elements (SINEs) family, are the most abundant MEIs in the human genome, with over a million copies. Although typically considered non-coding, Alu insertions can influence gene expression by inserting into exons, introns, or regulatory regions. L1 elements, belonging to the long interspersed nuclear elements (LINEs) family, are autonomous transposons capable of copying themselves and



doi:10.6342/NTU202500995

reinserting into the genome. This self-propagating mechanism is facilitated by the L1 encoded proteins, which reverse transcribe RNA back into DNA and integrate it into new locations. SVA elements are composite transposons formed from SINE, VNTR (variable number tandem repeat), and Alu sequences. Although less numerous than Alu and L1 elements, SVA insertions can still significantly impact the genome. The mechanisms driving MEIs include retrotransposition, where an RNA intermediate is reverse transcribed into DNA and inserted into the genome, and DNA transposition, where DNA elements move directly from one position to another. These processes can be influenced by factors such as mobile element activity, host regulatory mechanisms, and environmental stimuli. MEIs are not merely genomic parasites; they also play a role in shaping the genome. They contribute to genetic innovation by creating new genes, regulatory elements, and facilitating recombination events. This genomic plasticity is crucial for adaptation and evolution, as it introduces variability that can be acted upon by natural selection. However, the insertion of mobile elements can also be detrimental. Pathogenic insertions can lead to genetic disorders, cancer, and other diseases by disrupting gene function or regulatory networks. Understanding the balance between the beneficial and harmful effects of MEIs is essential for appreciating their role in human health and disease[14].

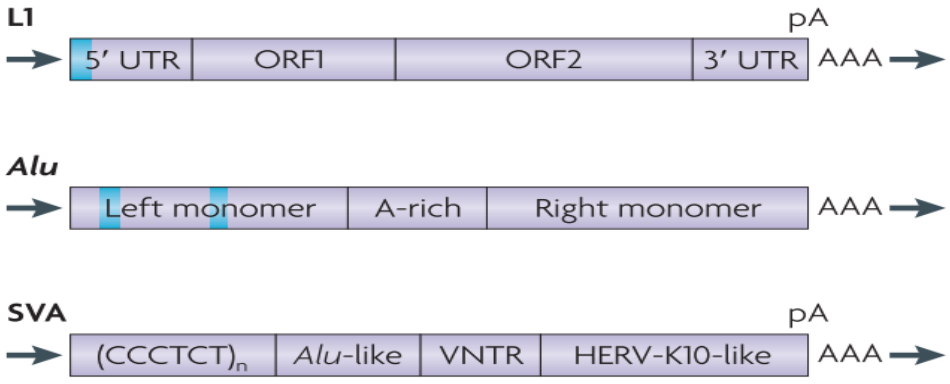



**Figure 8. Schematic illustration of mobile element insertions (MEIs).** LINE-1 (L1), Alu and SVA elements[14].



doi:10.6342/NTU202500995

## 1.8 Next generation sequencing for monogenic disease

Next-generation sequencing (NGS) for monogenic diseases enables rapid and high-accuracy detection of pathogenic variants, replacing traditional linkage analysis and candidate gene sequencing. It allows targeted gene panels, WES, and WGS to identify variants associated with inherited disorders. NGS has increased diagnostic efficiency, reduced costs, and improved early detection, facilitating precision medicine and personalized treatment strategies. Emerging technologies, such as long-read sequencing and AI-driven variant interpretation, continue to enhance its clinical applications. In summary, the evolution of NGS from a research tool to a cornerstone of clinical diagnostics has significantly impacted the management of monogenic diseases. Continued advancements in sequencing technologies and data analysis will further enhance our ability to diagnose, understand, and treat these conditions, paving the way for a future where precision medicine is the standard of care[15].



doi:10.6342/NTU202500995

# 2. Gene panel analysis of NTU TAAD cohort

## 2.1 Background

Using data from Taiwan's National Health Insurance claims database, our recent study found that the overall prevalence of Marfan syndrome (MFS) is 10.2 per 100,000 people in Taiwan. Acute aortic dissection occurred in 10% of Marfan patients at an average age of 36.6 ±10.7 years. The annual mortality rate of MFS in Taiwan is 0.23%[16], compared to aortic aneurysm and dissection causing 1-2% of deaths in Western countries[16, 17]. Population-based studies showed an annual incidence ranging from 6 per 100,000 in the UK to 9.1 per 100,000 in women and 16.3 per 100,000 in men in Sweden[18]. Aortic aneurysm and dissection cause significant morbidity and mortality, leading to over 10,000 deaths and contributing to more than 17,000 deaths annually in the United States[19]. Genetic factors are recognized as a significant component in assessing an individual's risk for aneurysms, and genetic testing has become a standard part of clinical practice. NGS has become more efficient and cheaper, successfully identifying pathogenic variants in TAAD patients and their families for earlier diagnosis.

## 2.2 Aims

The aim of the prospective study of TAAD was to systematically screen and manage participants recruited at a tertiary medical center in Taiwan. By utilizing a 29-gene next-generation sequencing panel (NTU TAAD panel)[11, 20], the study sought to provide comprehensive genetic testing alongside meticulous clinical evaluations conducted by a multidisciplinary team. This rigorous approach aimed to ensure precise diagnoses and effective management of TAAD, ultimately improving patient outcomes through the integration of genetic insights and tailored treatment strategies.



doi:10.6342/NTU202500995

## 2.3 Materials and Methods

### 2.3.1 Enrollment of participants

From April 2016 to March 2020, participants were consecutively recruited for screening and management of TAAD at a tertiary medical center in Taiwan, recorded in Research Electronic Data Capture (REDCap), a browser-based web application[21]. Each participant underwent a comprehensive evaluation of clinical manifestations and genetic testing using a NTU TAAD panel. The comprehensive evaluation included assessments by a multidisciplinary team comprising a trained research assistant, a cardiologist, an ophthalmologist, a dentist, and a radiologist. Clinical information was meticulously collected and analyzed to ensure accurate diagnosis and effective management of TAAD. Imaging techniques such as echocardiography and computed tomography (CT) scans were employed to diagnose dilated aorta. Z scores of the sinuses of Valsalva were calculated using established equations to provide precise measurements crucial for diagnosing Marfan syndrome and related conditions. Additionally, participants underwent plain films of the thoracolumbar spine, hip, and feet, along with spine MRIs, to thoroughly investigate musculoskeletal features associated with the disease. This rigorous approach ensured a holistic understanding of each patient's condition, paving the way for tailored treatment strategies and improved patient outcomes.

### 2.3.2 Establishment of NTU TAAD panel (29 genes)

For the whole transcript, the captured segments encompass regions from 10 kilobases upstream of the 5' untranslated region (UTR) to 5 kilobases downstream of the 3' UTR. This approach ensures thorough coverage of regulatory and coding regions in genes such as *ACTA2*, *ADAMTS10*, *CBS*, *COL1A1*, *COL1A2*, *COL3A1*, *COL4A5*, *COL5A1*, *COL5A2*, *EFEMP2*, *ELN*, *FBN1*, *FBN2*, *FLNA*, *GATA5*, *MAT2A*, *MFAP5*,



doi:10.6342/NTU202500995

*MYH11*, *MYLK*, *NOTCH1*, *PLOD1*, *SKI*, *SLC2A10*, *SMAD3*, *TGFB2*, *TGFB3*, *TGFBR1*, *TGFBR2*. For targeted exons, regions were captured extending 50 base pairs upstream and 50 base pairs downstream of each exon, specifically focusing on genes like *PRKG1* (Figure 9). This comprehensive capture strategy is designed to include essential genomic elements, facilitating the accurate identification and characterization of genetic variants within the gene panel.

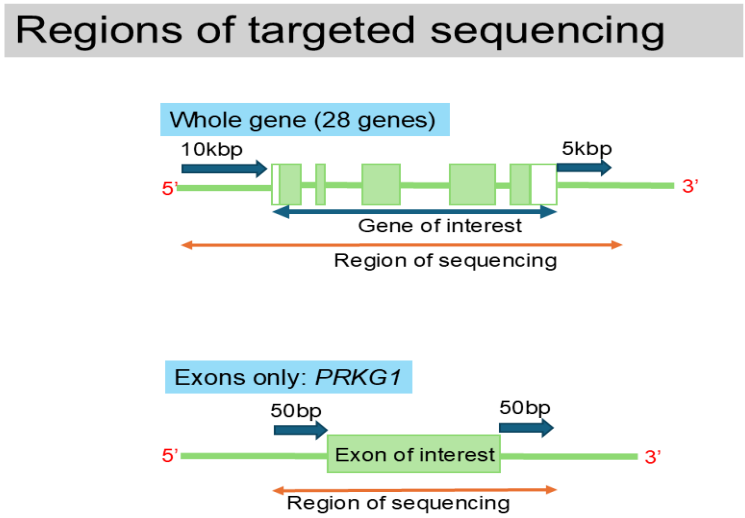



**Figure 9. Schematic illustration of targeted genes in the NTU TAAD panel**

### 2.3.3 Library construction and next generation sequencing

Genomic DNA was isolated from peripheral blood mononuclear cells collected in EDTA tubes using the Gentra Puregene Blood Kit (QIAGEN, Venlo, the Netherlands), following the manufacturer's recommended protocol to ensure optimal yield and purity. The DNA concentration was quantified by measuring absorbance at 260 nm, while its purity was assessed by calculating the 260/280 nm absorbance ratio, a standard metric for evaluating potential protein contamination. To verify the integrity of the isolated genomic DNA, a 1% agarose gel electrophoresis was performed, ensuring the presence of high-molecular-weight intact DNA.



doi:10.6342/NTU202500995

The genomic DNA was sheared by sonication (Covaris M220, Woburn, MA, U.S.), with the fragmentation process optimized to achieve a peak fragment length of approximately 800 base pairs, facilitating efficient downstream library preparation. The size distribution and concentration of the sheared DNA fragments were analyzed using the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, U.S.), while precise quantification was performed using Qubit fluorometric measurements (Thermo Scientific, Waltham, MA, U.S.), ensuring consistency in library input concentration.

Library preparation was conducted using the TruSeq Library Preparation Kit (Illumina, San Diego, CA, U.S.), generating sequencing-ready libraries with high fidelity and minimal bias. To further refine specificity, probe-based target enrichment was performed utilizing the SeqCap EZ Hybridization and Wash Kit (Roche NimbleGen, Madison, WI, U.S.), capturing genomic regions corresponding to the NTU TAAD gene panel. The enriched libraries were processed according to the NimbleGen SeqCap EZ Library Kit version 4.1 protocol, ensuring optimal hybridization efficiency and specificity.

Following enrichment, paired-end sequencing was conducted on the Illumina MiSeq platform using the Illumina MiSeq Reagent Kit v3, enabling high-throughput sequencing with enhanced accuracy. The workflow, from DNA extraction to sequencing, was carefully optimized to ensure high-quality data acquisition, supporting reliable variant identification and downstream analyses.

### 2.3.4 Computational resources

All bioinformatic analysis pipelines were executed on the Taiwania 3 computational platform, the most powerful CPU high-performance computing server in Taiwan, provided by the National Center for High-Performance Computing (NCHC). This advanced computational resource facilitated the efficient and accurate processing of



doi:10.6342/NTU202500995

large-scale genomic data. A Python script was meticulously developed to filter and prioritize the variants. This ensured a comprehensive analysis of potential genetic contributors to thoracic aortic aneurysm and dissection, underscoring the rigorous and methodical approach taken in this study to ensure the validity and reliability of the results.

### 2.3.5 Variant calling and annotation

Following paired-end sequencing on the Illumina MiSeq platform, raw reads were aligned to the human reference genome (GRCh37/hg19) using BWA-MEM v0.7.12 from the Wellcome Trust Sanger Institute, following the GATK best practices workflow. Preprocessing steps[22] began with bcl2fastq conversion from BCL to FASTQ format, followed by read alignment via BWA-MEM, generating BAM files, which were subsequently sorted using Picard's SortSam tool. Duplicate reads were marked with Picard's MarkDuplicates to prevent biases in downstream analyses. Indels were locally realigned using GATK's RealignerTargetCreator and IndelRealigner, correcting misalignments caused by insertions or deletions.

Variant discovery was performed using GATK's HaplotypeCaller in VCF mode, followed by variant annotation using ANNOVAR (dbSNP build 137, February 2016) to determine functional significance. ANNOVAR mapped variants to genomic coordinates, distinguishing between protein-coding, regulatory, and disease-associated regions, and classified their locations as exonic, intronic, intergenic, or within splice signal regions. Functional effects were further characterized by predicting synonymous/non-synonymous changes, stop-gain/loss mutations, frameshift, and non-frameshift insertions/deletions, relative to the protein positions[23]. The damaging effect of variant in coding region on protein function was predicted using in silico predictive algorithms, including PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/)[24], SIFT (Bioinformatics



doi:10.6342/NTU202500995

Institute, Matrix, Singapore)[25] and MutationTaster (http://www.mutationtaster.org)[26]. For variants located in noncoding regions, the impact on splicing elements is predicted using Human Splicing Finder version 3.1 (https://www.genomnis.com/hsf, Bioinformatics & Genetics Team, Aix Marseille Université, Aix-en-Provence, France)[27], MaxEntScan method[28] and SpliceAI[29]. Moreover, allele frequencies from various databases, including the 1000 Genomes Project (https://www.internationalgenome.org), the Exome Aggregation Consortium projects (http://exac.broadinstitute.org), the Genome Aggregation Database (https://gnomad.broadinstitute.org/), and the Taiwan Biobank (https://www.twbiobank.org.tw/), are incorporated to provide population-specific insights. Clinical pathogenicity is assessed through the ClinVar database[30] (https://www.ncbi.nlm.nih.gov/clinvar/), VarSome (https://varsome.com/) and Human Gene Mutation Database (https://www.hgmd.cf.ac.uk/ac/index.php) offering a comprehensive understanding of the variants' potential clinical relevance. This meticulous and systematic approach ensures the accuracy and reliability of the sequencing data, thereby facilitating the precise identification of genetic variants associated with various disorders, including thoracic aortic aneurysm and dissection. The integration of advanced computational pipelines and extensive variant annotation underscores the rigorous and methodical efforts undertaken to achieve high-quality genomic analyses.

### 2.3.6 Prioritization and interpretation of variants

Following genetic variant annotation, those with a minor allele frequency (MAF) below 1% were prioritized as potential disease-causing candidates. A systematic filtering process was applied, considering variant type, predicted functional impact, population frequency, inheritance mode, database searches, and literature reviews to refine selection.



doi:10.6342/NTU202500995

For single nucleotide variants (SNVs), candidates met at least two of three criteria: a SIFT-pred score < 0.05, PolyPhen-2 score > 0.85, or a MutationTaster classification indicating a damaging effect. Variants with SpliceAI scores > 0.5, as well as predictions from MaxEntScan and Human Splicing Finder, were included if splicing alterations were detected. Classification followed the American College of Medical Genetics and Genomics (ACMG) guidelines 2015[31] , categorizing variants as pathogenic, likely pathogenic, uncertain significance, likely benign, or benign, with only pathogenic or likely pathogenic variants reported as disease-causing. To confirm a genetic diagnosis, the variant's mode of inheritance—dominant, recessive, or X-linked—was assessed for consistency with the associated Mendelian disease, ensuring clinically relevant findings.

### 2.3.7 Confirmation of identified variants

The confirmation of identified variants was strengthened through visualization using the Integrative Genomics Viewer (IGV)[32, 33] (https://igv.org) , which facilitates detailed inspection of sequence alignments and variant calls. By loading aligned BAM files into IGV, we could evaluate read coverage, variant frequency, and supporting reads at specific genomic locations, enabling precise validation of single nucleotide variants (SNVs) and small insertions/deletions (indels). This visual verification enhances confidence in the accuracy of bioinformatic variant calls, ensuring robustness in the sequencing analysis. Additionally, Sanger sequencing[34] was employed for further validation using the ABI 3730 DNA Analyzer and the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit. Primers were carefully designed with Geneious Software to specifically amplify DNA segments harboring the suspected variants. Using 20 ng of genomic DNA, amplification was performed with GoTaq PCR reagents (Promega GoTaq Green Master Mix) on an API Verti Thermal Cycler (Applied Biosystems). Following amplification, PCR products



doi:10.6342/NTU202500995

were purified and subjected to sequencing via the BigDye Terminator Ready Reaction Mix (Applied Biosystems). The resulting sequences were rigorously analyzed to confirm the presence of the variants, ensuring high precision and reliability in genetic variant detection. This two-step confirmation approach—IGV visualization and Sanger sequencing—provides a robust framework for validating sequencing results, reinforcing the accuracy of detected variants in genomic studies.

### 2.3.8 Statistics

Continuous variables were presented as median values with interquartile ranges (25%-75%), while categorical variables were expressed as frequencies and percentages. For comparing continuous variables across two or more groups, Mann-Whitney U tests and Kruskal-Wallis tests were used, respectively. Comparisons of categorical variables were conducted using Chi-Square or Fisher's exact tests. Logistic Regression Analysis, adjusting for age and gender, was reported with odds ratios (OR) and 95% confidence intervals (CI). The Cox proportional hazard model was employed to assess factors influencing the lifetime risk of aortic aneurysm and dissection (AoAD), presented as hazard ratios (HR) with 95% CI. All statistical analyses were performed using IBM SPSS Statistics V21, with significance determined at a two-sided p-value of $\leq 0.05$.

## 2.4 Results

### 2.4.1 Baseline characteristics of NTU TAAD cohort

This registry commenced in April 2016 and continued until the end of March 2020, enrolling a total of 113 patients. After excluding 6 patients lost to follow-up, 107 patients were included in the analysis. Table 1 details the baseline characteristics of the cohort stratified by the reasons for referral: Known AoAD (n = 57), Appearance (n = 36), Family



doi:10.6342/NTU202500995

History (n = 11), and Ectopic Lens (n = 3).

Patients referred for Known AoAD had a median age of 30 years [17–43], which was significantly higher than the Appearance group with a median age of 18 years [6–30] ($p < 0.001$). The Family History group had a median age of 25 years [14–36], and the Ectopic Lens group had a median age of 35 years. Despite similar proportions of males across groups (63.2% in Known AoAD, 61.1% in Appearance, 63.6% in Family History, and 66.7% in Ectopic Lens; $p = 1.000$) and comparable body heights (BH: 175 cm [157–193] in Known AoAD, 178 cm [160–196] in Appearance, 175 cm [164–186] in Family History, and 162 cm in Ectopic Lens; $p = 0.527$), body weight (BW) and body mass index (BMI) differed significantly. The Known AoAD group had a higher BW of 67.5 kg [40.7–94.3] compared to 54.0 kg [31.5–76.5] in the Appearance group ($p < 0.001$) and a higher BMI of 21.0 kg/m² [15.1–26.9] versus 17.0 kg/m² [12.1–21.9] ($p < 0.001$). The ratios of arm length to BH and upper to lower segments were similar, with p-values of 0.401 and 0.8804, respectively.

In terms of comorbidities, hypertension (HTN) was present in 8 patients in the Known AoAD group, while it was absent in the other groups ($p = 0.038$). The prevalence of dyslipidemia (DLP) and diabetes mellitus (DM) did not show significant intergroup differences ($p = 0.438$ and $p = 1.000$, respectively). Although myopia was common across groups (with 25, 27, 5, and 1 cases respectively; $p = 0.083$), the presence of ectopic lens differed significantly (11 in Known AoAD, 6 in Appearance, 0 in Family History, and 3 in Ectopic Lens; $p = 0.006$). Finally, both the systemic score and the proportion of patients with a systemic score ≥7 were similar among the groups ($p = 0.974$ and $p = 0.978$, respectively).

As shown in Figure 10A, a pie chart of the reasons for referral indicates that more than half of the patients were referred due to known dilated aorta, while other referral



doi:10.6342/NTU202500995

categories included appearance, family history of TAAD, and ectopic lens. Notably, patients referred for “appearance” were significantly younger and had lower body weight and BMI compared to other groups (Table 1). Among the cohort, AoAD was confirmed in 73 patients (68.2%); within this subset, 68 patients had ascending aortic dilatation, 2 had descending aortic dilatation, and 3 exhibited involvement of both segments. Furthermore, 17 dissections were identified among patients with aortic dilatation (11 ascending, 5 descending, and 1 spanning both segments). Figure 10B illustrates both the absolute numbers and incidence of AoAD stratified by referral reason, revealing that the detection of AoAD was highest in patients referred for known TAAD, followed by those referred for ectopic lens, family history, and appearance. Lastly, Figure 10C displays the distribution of systemic scores between patients with and without AoAD, suggesting that higher systemic scores may be associated with aortic pathology. Collectively, these results highlight the clinical heterogeneity of the TAAD cohort and emphasize the importance of incorporating diverse patient referral reasons and phenotypic assessments in the diagnostic process.



doi:10.6342/NTU202500995

**Table 1. The baseline characteristics grouped by "the reasons of referral"**

| | Known AoAD (57) | Appearance (36) | Family History (11) | Ectopic Lens (3) | P value* |
|---|---|---|---|---|---|
| Age | 30 [17-43] | 18 [6-30] | 25 [14-36] | 35 | <0.001 |
| Sex (male) | 36 (63.2%) | 22 (61.1%) | 7 (63.6%) | 2 (66.7%) | 1.000 |
| BH (cm) | 175 [157-193] | 178 [160-196] | 175 [164-186] | 162 | 0.527 |
| BW (kg) | 67.5 [40.7-94.3] | 54.0 [31.5-76.5] | 58.0 [46-70] | 58 | <0.001 |
| BMI ($kg/m^2$) | 21.0 [15.1-26.9] | 17.0 [12.1-21.9] | 20.7 [18-23.4] | 22.7 | <0.001 |
| Arm/BH | 1.02 [0.97-1.07] | 1.00 [0.97-1.03] | 1.02 [0.96-1.08] | 0.98 | 0.401 |
| Ratio of Upper/Lower segments | 0.72 [0.58-0.86] | 0.70 [0.54-0.86] | 0.80 [0.63-0.97] | 0.71 | 0.8804 |
| HTN | 8 | 0 | 0 | 0 | 0.038 |
| DLP | 5 | 1 | 0 | 0 | 0.438 |
| DM | 1 | 0 | 0 | 0 | 1.000 |
| Myopia | 25 | 27 | 5 | 1 | 0.083 |
| Ectopic Lens | 11 | 6 | 0 | 3 | 0.006 |
| Systemic Score | 6 [0-12] | 7 [1-13] | 7 [0-14] | 5 | 0.974 |
| Systemic score ≥7 | 18 | 14 | 3 | 1 | 0.978 |

Continuous variables were presented as median [25%-75% quartiles]. Categorical variables were presented as frequency (percentage).

*For comparison between groups, Kruskal-Wallis test and Fisher's exact test were used for continuous and categorical variables, respectively.

Abbreviation: AoAD, dilated or dissected aorta; BH, body height; BMI, body mass index; BW body weight; DLP, dyslipidemia; DM, diabetes mellitus; HTN, hypertension



doi:10.6342/NTU202500995

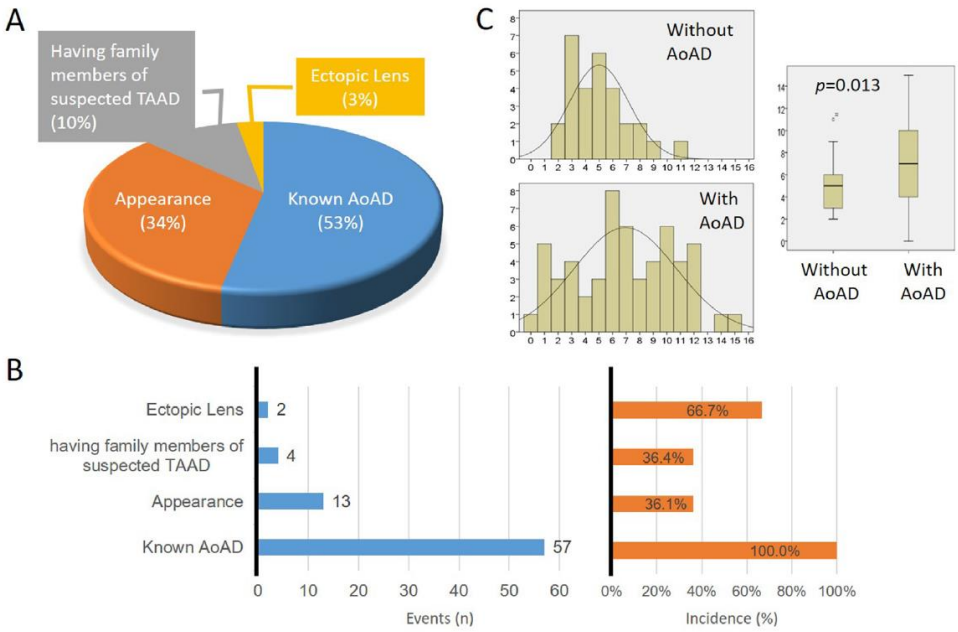



**Figure 10. Baseline characteristics of NTU TAAD cohort.** (A) The pie chart of the patient population according to "the reasons of referral". (B) The absolute numbers and the incidence of AoAD among the patients according to "the reasons of referral". (C) The distribution of systemic scores among patents with and without AoAD.

### 2.4.2 Physical appearance as a potential prediction marker

In our analysis of potential markers for aortic dilatation or dissection (AoAD), we compared 73 patients with AoAD to 34 without AoAD (Table 2). Although the median age and sex distribution were similar between the two groups (age: 25 [9–41] versus 24.5 [8.5–40.5] years, $p = 0.018$; male: 65.8% versus 55.9%, $p = 0.392$), significant differences emerged in other parameters. Specifically, patients with AoAD exhibited a higher body weight (68.0 kg [46.5–89.5] versus 54.0 kg [47.1–60.9], $p < 0.001$, adjusted $p = 0.016$) and a greater body mass index (21.2 kg/m² [16.3–26.1] versus 17.7 kg/m² [12.9–22.5], $p = 0.001$, although the adjusted $p = 0.114$). Other anthropometric measures such as body height, the ratio of arm length to body height, and the upper-to-lower segment ratio did not differ significantly between groups. Importantly, the systemic score—a composite measure reflecting features associated with connective tissue disorder—was markedly



doi:10.6342/NTU202500995

higher in the AoAD group (median 8 [3–13] compared to 4.5 [0.5–8.5], $p = 0.013$, adjusted $p = 0.003$), with a significantly greater percentage of AoAD patients having a systemic score ≥7 (53.6% versus 20.7%, $p = 0.004$, adjusted $p = 0.001$). Ocular findings also contributed to the distinction between groups, as an ectopic lens was observed in 26.6% of AoAD patients compared to only 9.1% in those without AoAD ($p = 0.044$, adjusted $p = 0.029$). Furthermore, genetic evaluations revealed that positive findings on the TAAD panel were significantly more frequent in the AoAD group (54.8% versus 17.6%, $p < 0.001$) and variants in *FBN1* were identified in 43.8% of these patients compared to 11.8% of non-AoAD patients ($p = 0.001$, adjusted $p < 0.001$). These results underscore that increased body weight, higher systemic scores, the presence of ectopic lens, and specific genetic markers—particularly *FBN1* variants—may serve as potential clinical markers for AoAD in this patient cohort.

Table 3 summarizes potential markers of AoAD among systemic features in our cohort by comparing patients with AoAD (n = 73) to those without AoAD (n = 34). There were no significant differences between the groups in the prevalence of wrist or thumb signs (77.8% vs. 88.9%, p = 0.225) or chest wall abnormalities (54.5% vs. 33.3%, p = 0.100). Notably, pectus carinatum was significantly more frequent in the AoAD group compared to the non-AoAD group (29.1% vs. 7.4%, p = 0.026). Similarly, the presence of skin striae distensae was markedly higher among patients with AoAD (56.4% vs. 24%, p = 0.007), remaining significant even after adjustment for age and sex. In addition, although mitral valve prolapse was observed in 69.8% of patients with AoAD compared to 50.0% in those without ($p = 0.079$), adjustment by age and gender yielded a significant difference (adjusted p = 0.023). In contrast, other systemic features—including pectus excavatum, hind foot abnormality, pes planus, pneumothorax, dural ectasia, hip abnormality, scoliosis, elbow extension, facial features, severe myopia, and the



doi:10.6342/NTU202500995

upper/lower ratio—showed no statistically significant differences between the two groups.

The features commonly relied upon by primary physicians, such as being tall and thin or having loose joints, were not significantly associated with AoAD. In contrast, a detailed review of systemic scores based on the revised Ghent nosology revealed a significant association between skin striae distensae and AoAD, with 56.4% of patients with AoAD exhibiting skin striae distensae compared to 24% in those without ($p = 0.008$). This association remained robust after adjusting for age and sex (OR 4.506, 95% CI 1.504–13.497, $p = 0.007$) and persisted following further adjustments for body weight (OR 3.658, 95% CI 1.176–11.380, $p = 0.025$) or BMI (OR 4.794, 95% CI 1.566–14.679, $p = 0.006$). Although individual physical features did not differ significantly between patients with and without AoAD, the overall systemic score was notably higher in patients with AoAD (median 8 vs. 4.5, $p = 0.013$). Figure 10C further illustrates that 53.6% of patients with AoAD had a systemic score ≥7, compared to only 20.7% of patients without AoAD ($p = 0.004$, and $p = 0.001$ after adjustment for age and sex), underscoring the potential of systemic scores as a marker for AoAD.



doi:10.6342/NTU202500995

**Table 2. Potential markers of aortic aneurysm or dissection**

| | AoAD (73) | No AoAD (34) | P value* | P value adjusted by age and sex† |
|---|---|---|---|---|
| Age | 25 [9-41] | 24.5 [8.5-40.5] | 0.018 | |
| Sex (male) | 48 (65.8%) | 19 (55.9%) | 0.392 | |
| BH | 176 [157-194] | 176 [163-188] | 0.105 | |
| BW | 68.0 [46.5-89.5] | 54.0 [47.1-60.9] | <0.001 | 0.016 |
| BMI | 21.2 [16.3-26.1] | 17.7 [12.9-22.5] | 0.001 | 0.114 |
| Arm/BH | 1.02 [0.97-1.07] | 1.00 [0.96-1.04] | 0.090 | |
| Ratio of Upper/Lower segments | 0.71 [0.56-0.86] | 0.74 [0.58-0.80] | 0.357 | |
| HTN | 7 (13.0%) | 1 (3.8%) | 0.264 | |
| DLP | 6 (11.5%) | 0 | 0.168 | |
| DM | 1 (1.9%) | 0 | 1.000 | |
| Systemic Score | 8 [3-13] | 4.5 [0.5-8.5] | 0.013 | 0.003 |
| Systemic score >=7 | 30 (53.6%) | 6 (20.7%) | 0.004 | 0.001 |
| Ectopic Lens | 17 (26.6%) | 3 (9.1%) | 0.044 | 0.029 |
| TAAD panel | 40 (54.8%) | 6 (17.6%) | <0.001 | <0.001 |
| *FBN1* | 32 (43.8%) | 4 (11.8%) | 0.001 | <0.001 |

Continuous variables were presented as median [25%-75% quartiles]. Categorical variables were presented as frequency (percentage). *For comparison between groups, Mann-Whitney U test and Fisher's exact test were used for continuous and categorical variables, respectively. †Logistic Regression Analysis was used for age and gender adjustment. Abbreviation: AoAD, dilated or dissected aorta; BH, body height; BMI, body mass index; BW body weight; DLP, dyslipidemia; DM, diabetes mellitus; HTN, hypertension; TAAD, thoracic aortic aneurysm and dissection.



doi:10.6342/NTU202500995

**Table 3. Potential markers of AoAD among systemic features**

| | AoAD (73) | No AoAD (34) | P value* | P value adjusted by age and sex† |
|---|---|---|---|---|
| Wrist or Thumb signs | 42 (77.8%) | 24 (88.9%) | 0.225 | |
| Chest wall abnormality | 30 (54.5%) | 9 (33.3) | 0.100 | |
| Pectus Carinatum | 16 (29.1%) | 2 (7.4%) | 0.026 | |
| Pectus Excavatum | 14 (25.5%) | 7 (25.9%) | 0.963 | |
| Hind foot abnormality | 3 (5.6%) | 0 | 0.548 | |
| Pes planus | 26 (48.4%) | 9 (32.1%) | 0.165 | |
| Pneumothorax | 6 (11.3%) | 3 (11.5%) | 1.000 | |
| Dural Ectasia | 15 (68.2%) | 6 (46.2%) | 0.199 | |
| Hip abnormality | 7 (53.8%) | 1 (10.0%) | 0.074 | 0.999 |
| Scoliosis | 26 (47.3%) | 12 (44.4%) | 0.809 | |
| Elbow Extension | 1 (2.1%) | 1 (4.8%) | 0.519 | |
| Facial features | 9 (18.4%) | 3 (14.3%) | 1.000 | |
| Skin striae distensae | 31 (56.4%) | 6 (24%) | 0.007 | 0.007 |
| Myopia severe | 36 (62.1%) | 22 (71%) | 0.401 | |
| Mitral valve prolapse | 37 (69.8%) | 14 (50.0%) | 0.079 | 0.023 |
| Upper/Lower Ratio | 5 (11.4%) | 0 | 0.315 | |

Continuous variables were presented as median [25%-75% quartiles]. Categorical variables were presented as frequency (percentage). *For comparison between groups, Mann-Whitney U test and Fisher's exact test were used for continuous and categorical variables, respectively. †Logistic Regression Analysis was used for age and gender adjustment.



doi:10.6342/NTU202500995

### 2.4.3 Prognostic utility of the NTU TAAD panel

The Kaplan–Meier survival analyses demonstrated that patients with these pathogenic variants exhibited significantly reduced event-free survival compared to those without, reinforcing the prognostic value of genetic screening. Figure 11 illustrates the Kaplan–Meier survival analyses evaluating event-free survival based on the NTU TAAD multi-genetic NGS panel stratification into patients with (gTAAD+) versus without (gTAAD−) disease-causing variants. For the composite endpoint of aortic aneurysm or dissection (Panel A), gTAAD+ patients exhibited a significantly elevated risk, with an unadjusted hazard ratio (HR) of 3.346 (95% CI: 2.058–5.442, $p < 0.001$) and an adjusted HR of 3.685 (95% CI: 1.543–7.307, $p = 0.003$) after controlling for age, sex, hypertension, dyslipidemia, and diabetes mellitus. In Panel B, the occurrence of dissection alone was markedly higher in gTAAD+ patients, yielding an unadjusted HR of 5.274 (95% CI: 1.914–14.532, $p = 0.001$) and an adjusted HR of 7.512 (95% CI: 1.703–33.135, $p = 0.008$). In a subset analysis of patients with confirmed aneurysm (Panel C), the unadjusted HR for developing dissection was 4.952 (95% CI: 1.578–13.135, $p = 0.005$); however, this association was attenuated after adjustment (adjusted HR of 3.619, 95% CI: 0.794–16.490, $p = 0.097$). Overall, these findings suggest that the presence of disease-causing variants, as identified by the NTU TAAD panel, is significantly associated with an increased risk of adverse aortic events, particularly dissection, underscoring the prognostic utility of genetic screening in this patient population.



doi:10.6342/NTU202500995

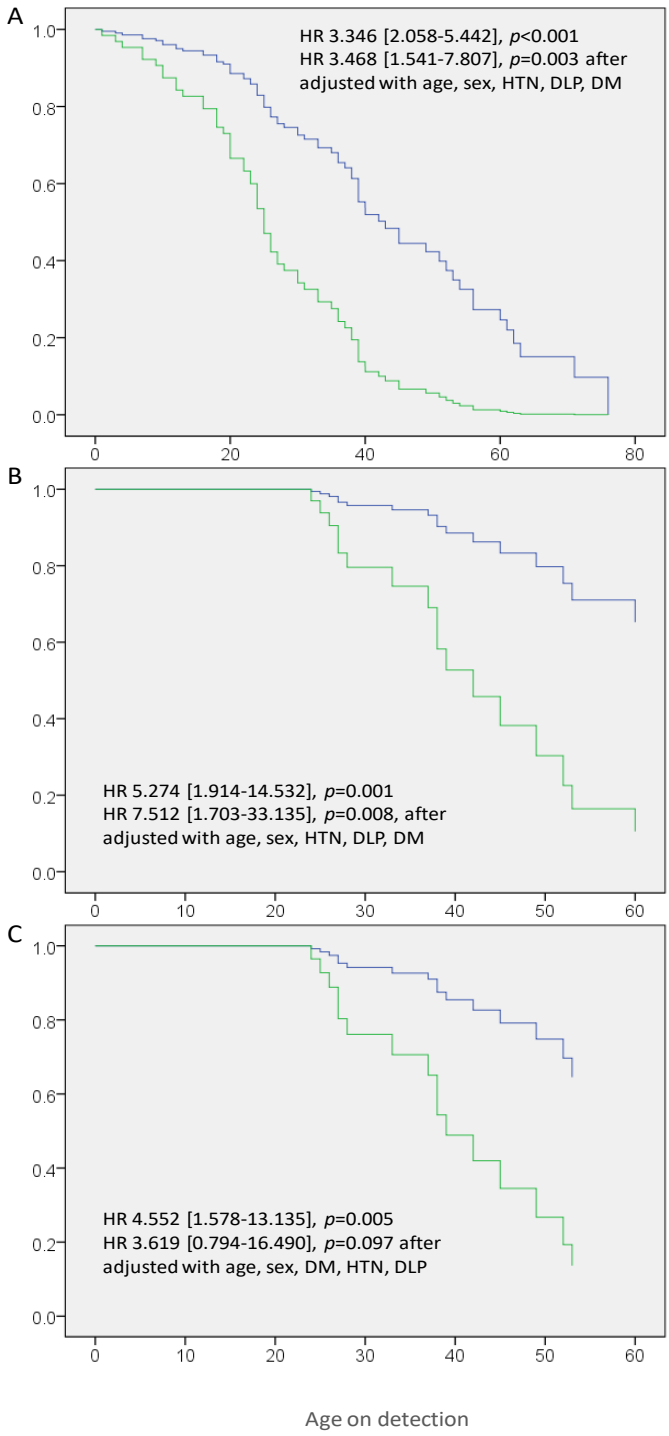



**Figure 11. Kaplan-Meier curves depicting Event-Free survival based on the NTU TAAD panel** (A) The occurrence of aortic aneurysm or dissection in patients with (gTAAD+, green line) or without (gTAAD-, blue line) disease causing variants. (B) The occurrence of dissection in gTAAD+ (green line) or gTAAD- (blue line) patients. (C) The occurrence of dissection in gTAAD+ (green line) or gTAAD-(blue line) patients with already confirmed aneurysm.



doi:10.6342/NTU202500995

### 2.4.4 Multi-gene NGS panel outperforms *FBN1*-only analysis in predictive diagnosis

We evaluated the performance of the NTU TAAD multi-genetic NGS panel in predicting aortic aneurysm or dissection (AoAD), and notably, 46 individuals in the cohort were found to harbor disease-causing genes—resulting in a detection rate of 43.0%. This finding underscores the panel's utility in identifying patients with a genetic predisposition to AoAD. Table 4 summarizes the predictive power of the NTU TAAD panel and a focused *FBN1* analysis in our cohort of 107 patients (73 with AoAD and 34 without). Using the NTU TAAD panel, 46 patients were found to be genotype positive (gTAAD+), of whom 40 had AoAD and 6 did not, whereas 61 patients were genotype negative (gTAAD–), including 33 with AoAD and 28 without. This corresponds to a sensitivity of 54.8%, a specificity of 82.4%, a positive predictive value (PPV) of 87.0%, and a negative predictive value (NPV) of 45.9%. In a parallel analysis considering only *FBN1*, 36 patients were *FBN1*-positive, with 32 having AoAD and 4 without AoAD, while 71 were *FBN1*-negative, including 41 with AoAD and 30 without, yielding a sensitivity of 43.8%, specificity of 88.2%, PPV of 88.9%, and NPV of 42.3%. Notably, among the 46 gTAAD+ patients, pathogenic variants were identified in 24 individuals and likely pathogenic variants in 22. Within this group, the disease-causing genes were predominantly *FBN1* (78.3%), followed by *TGFBR1* (10.9%), *TGFBR2* (6.5%), and *FBN2* (4.3%). The types of variants observed were diverse, consisting of 24 missense (52.2%), 10 nonsense (21.7%), 4 frameshift (8.7%), and 8 splice site variants (17.4%); moreover, 29 variants (63.0%) were novel, having not been previously reported in the dbSNP database (Table S1). Overall, the NTU TAAD panel demonstrated a higher sensitivity for detecting AoAD compared to *FBN1* testing alone, while specificity, PPV, and NPV were similar between the two approaches, underscoring the panel's clinical



doi:10.6342/NTU202500995

utility in risk stratification and management of AoAD.

**Table 4. The prediction power of TAAD panel**

| TAAD panel | AoAD | | Total | |
|---|---|---|---|---|
| | AoAD | No AoAD | | |
| A | | | | |
| + | 40 | 6 | 46 | PPV=87.0% |
| - | 33 | 28 | 61 | NPV=45.9% |
| Total | 73 | 34 | 107 | |
| | Sen.=54.8% | Spe.=82.4% | | |
| *FBN1* | AoAD | | Total | |
| | AoAD | No AoAD | | |
| B | | | | |
| + | 32 | 4 | 36 | PPV=88.9% |
| - | 41 | 30 | 71 | NPV=42.3% |
| Total | 34 | 73 | 107 | |
| | Sen.=43.8% | Spe.=88.2% | | |

2-by-2 contingency table reporting counts of genotypes for AoAD cases vs. controls by TAAD panel(table A) or *FBN1* (table B).

Abbreviation: AoAD, dilated or dissected aorta; NPV, negative predictive value; PPV, positive predictive value; Sen., sensitivity; Spe., specificity.

### 2.4.5 Genetic characteristics of TAAD patients

In the TAAD gene panel–positive group (n = 46), disease causing variants were predominantly identified in *FBN1*, accounting for 78.3% (36 patients), while *TGFBR1*, *TGFBR2*, and *FBN2* were implicated in 10.9% (5 patients), 6.5% (3 patients), and 4.3% (2 patients) of cases, respectively. The spectrum of variant types included missense mutations (52.2%, 24 variants), nonsense mutations (21.7%, 10 variants), frameshift mutations (8.7%, 4 variants), and splice site mutations (17.4%, 8 variants). Notably, 63.0% (29 variants) of these alterations were novel, as they had not been recorded in the dbSNP database (Table S1).

In the group with negative TAAD gene panel test results (gTAAD−), 16 variants of uncertain significance (VUS) were identified in 14 patients. Notably, one patient harbored



doi:10.6342/NTU202500995

VUS in both *MYH11* and *PLOD1*, while another had VUS in both *COL1A2* and *MYLK*; no suspicious variant was found in the remaining 47 patients. These 16 VUS comprised 5 variants in *MYH11*, 4 in *COL1A2*, 3 in *COL4A5*, and 1 each in *COL5A1*, *ELN*, *FBN2*, and *PLOD1*. All of the variants were missense mutations, and only three of them had not been previously reported in the dbSNP database (not shown).

### 2.4.6 Disease-causing variants associate with early onset and enhanced systemic manifestations in TAAD patients

The clinical characteristics of patients with (gTAAD+, n = 46) and without (gTAAD−, n = 61) disease-causing variants identified by the NTU TAAD panel. Patients in the gTAAD+ group were significantly younger (median 23 years [7–39]) than those in the gTAAD− group (median 29 years [12–46], p = 0.008) (Table S2). Although the proportion of males was lower in the gTAAD+ group (52.2% vs. 70.5%, p = 0.053), body height was comparable between groups (175 cm vs. 177 cm, p = 0.495). Body weight was marginally higher in the gTAAD+ group (61.8 kg vs. 61.5 kg, p = 0.039), while BMI did not differ significantly (20.25 vs. 20.35 kg/m², p = 0.154). With respect to systemic features, patients with disease-causing variants had a higher median systemic score (8 vs. 5, p = 0.020, adjusted p = 0.017) and a greater proportion had a systemic score ≥7 (60.5% vs. 23.8%, p = 0.001, adjusted p = 0.003). Ocular findings also differed, as an ectopic lens was more frequently observed in the gTAAD+ group (30.2% vs. 13.0%, p = 0.037), although this difference did not remain significant after adjustment (adjusted p = 0.083). Regarding skeletal features, the gTAAD+ group demonstrated a significantly higher prevalence of chest wall abnormalities (61.9% vs. 32.5%, p = 0.008, adjusted p = 0.008) and pectus excavatum (35.7% vs. 15.0%, p = 0.032, adjusted p = 0.024), as well as a higher incidence of pes planus (56.1% vs. 29.3%, p = 0.014, adjusted p = 0.026) and dural



doi:10.6342/NTU202500995

ectasia (83.3% vs. 35.3%, p = 0.004, adjusted p = 0.007). Other parameters, including wrist or thumb signs (85.7% vs. 76.9%, p = 0.0309), pectus carinatum (26.2% vs. 17.5%, p = 0.0342), hind foot abnormalities, pneumothorax, scoliosis, elbow extension, facial features, skin striae, severe myopia, mitral valve prolapse, and the upper/lower segment ratio did not differ significantly between the groups. Collectively, these findings suggest that patients harboring disease-causing variants tend to be younger and exhibit more pronounced systemic manifestations, such as higher systemic scores, chest wall anomalies, pes planus, and dural ectasia, compared to those without such genetic findings.

## 2.5 Discussion

In this study, we examined the clinical manifestations and findings from NTUH TAAD panel in TAAD patients at an integrated clinic. Our results indicated that a higher systemic score, rather than just tall and thin features, was associated with occult aortic disease. Notably, the presence of skin striae distensae was significantly linked to a diseased aorta, even after adjusting for age and gender, and could serve as an accessible indicator for further investigation. The gene panel demonstrated high predictive power but low sensitivity in screening patients. A positive genetic test result indicated a higher risk of aortic dissection, suggesting the need for earlier intervention in patients with known aortic dilatation.

### 2.5.1 Skin striae distensae as a marker for AoAD screening

Early detection of AoAD is crucial before the aorta risks rupture. While echocardiography is minimally invasive, it is also time-consuming and costly, making it impractical for general population screening. Patients are often referred due to tall and thin features, which are not significantly associated with AoAD. Conversely, some



doi:10.6342/NTU202500995

patients with acute dissection exhibit a regular appearance, complicating detection. Although the Ghent nosology systemic score correlates with AoAD, acquiring this score requires further imaging studies, which are impractical for clinics or small hospitals. Skin striae distensae, however, presents a convenient marker. These are caused by rapid connective tissue expansion resulting in dermal atrophy and epidermal thinning[35]. Striae are also seen in pregnancy or rapid weight gain, but in connective tissue disease patients, normal growth may cause dermal damage, leading to striae. This finding aligns with former case-control study, which reported that skin striae were significantly more frequent in MFS patients (92%) compared to controls (61%) ($p = 0.0001$)[36]. Striae in unusual locations had high specificity (84%) with notable sensitivity (66% vs. 16% in controls, $p < 0.0001$). Further studies with larger groups are warranted to explore the distribution and characteristics of striae distensae.

### 2.5.2 The gene panel can confirm early-stage AoAD

The gene panel can confirm early-stage AoAD, boasting high specificity and positive predictive value. It is particularly valuable for stratifying dissection risks among patients with aortic dilatation, thereby aiding in the determination of follow-up frequency and type.

In our cohort, 68% of patients had a diseased aorta, but only 43% had pathogenic or likely pathogenic variants detected with the panel. This panel demonstrated a specificity of 82.4% and a positive predictive value (PPV) of 87%. Its sensitivity, however, was 54.8% (Table 4A). Limiting the panel to the *FBN1* gene alone increased specificity and PPV to 88.2% and 88.9%, respectively, but reduced sensitivity to 43.8% (Table 4B). The panel is beneficial for precision medicine, particularly for conditions like thoracic aortic aneurysm, MFS, Loeys-Dietz syndrome, and familial TAAD syndrome. This conditions



doi:10.6342/NTU202500995

are among the “actionable” phenotypes highlighted by the American College of Medical Genetics and Genomics[37]. Positive genetic testing aids in patient and family counseling, cascade genetic screening, prenatal testing, and clinical evaluation.

Among patients with aortic aneurysm, those with positive genetic findings had a higher risk of dissection than those without, similar to observations by other authors who recommended different aortic dimension cutpoints for surgical intervention based on genotype-phenotype correlations[20, 38].

Detection rates vary due to different cohorts, inclusion criteria, panel sizes, and sequencing methods. Our cohort's detection rate of 43% was higher, likely due to inclusive age ranges, strict criteria, and advanced sequencing methods. Comparatively, a combined Yale and UK cohort identified 49 variants in 47 cases out of 1025 with a targeted 15-gene panel, and a Chinese cohort found 92 variants in 248 probands (37.1%)[39, 40].

### 2.5.3 The gene panel can help stratify the risk of dissection

Among patients with aortic aneurysm, those with positive genetic findings demonstrated a higher risk of dissection compared to those without such findings. This observation aligns with the research conducted by Wolford et al., who noted that individuals harboring pathogenic variants experienced aortic dissection at a significantly younger age (41 years) compared to those without these variants (57 years)[38]. Moreover, Brownstein et al. provided recommendations for varying aortic dimension cutpoints for surgical intervention based on genotype-phenotype correlations[20]. Specifically, their study suggested that patients with certain genetic variants, such as those in the *FBN1* gene, should undergo surgical intervention at smaller aortic dimensions due to the increased risk of dissection associated with these variants. The recommended aortic dimension



doi:10.6342/NTU202500995

cutoffs were lower for individuals with pathogenic variants, reflecting the higher risk of early dissection[20]. However, it is important to note that our study was constrained by the limited number of cases and disease-causing genes, which precluded a comprehensive analysis of these associations.

### 2.5.4 High detection rate due to strict inclusion criteria

The detection rate varied among different cohorts due to differing patient populations, inclusion criteria, and the size of the gene panel used in the studies. Our cohort's detection rate of 43.0% was higher than that of other contemporary cohorts. This higher rate can be attributed to the inclusive age range of our patient population, which spans from infants to adults, the younger mean and median ages, strict inclusion criteria, advanced sequencing methods, and the implementation of new prediction tools and online aggregated databases for variant classification. In a combined study of Yale and UK cohorts, the targeted PCR and NGS-based panel of 15 genes in 1,025 unrelated TAAD cases identified 49 variants in 47 cases. This cohort had a higher mean age of 60 years[39]. Conversely, a Chinese cohort that included 248 probands with aortic disease or MFS found 92 variants, resulting in a detection rate of 37.1% using the NGS-based 15-gene panel[40]. For cohorts that targeted only the *FBN1* gene, 5.8% of the sporadic TAAD cases were found to have pathogenic variants, while up to 40% of the 157 TAAD families were detected with these variants[41, 42].

## 2.6 Conclusions

TAAD is a silent and potentially devastating disease. Skin striae distensae may serve as a preliminary screening indicator prior to further examinations. A multi-gene NGS panel can confirm early-stage diagnosis and help stratify dissection risk among patients



doi:10.6342/NTU202500995

with a known dilated aorta.

## 2.7 Limitations

However, the study had several limitations. First, although the patient enrollment was prospective, the retrospective review of the age of AoAD detection could introduce recall bias, potentially skewing the results. Second, some patients' aortic dissections were replaced with artificial grafts, making it impossible to determine the original diameter before rupture. Third, the highly selective patient enrollment may not reflect the general population. Fourth, incomplete clinical manifestations in some patients could affect the observations. Finally, this study was limited to a Taiwanese cohort, predominantly Han people, which may limit the generalizability of the findings.



doi:10.6342/NTU202500995

# 3. Comparisons of performances of structural variants detection algorithms in solitary or combination strategy

## 3.1 Background

Despite substantial advancements in various structural variant (SV) detection methods, their effectiveness in analyzing short-read sequences remains limited. Current sequencing-based methods employ several approaches to derive information about SVs: read-pair (RP), read-depth (RD), split-read (SR), and assembly (AS) approaches. However, the performance of these methods in short-read sequences continues to be constrained due to a high rate of SV miscalling, irrespective of the strategy employed. Over the years, numerous SV detection algorithms have been developed and published, contributing significantly to the field of genomics. Many projects utilizing popular SV detection algorithms have demonstrated relatively high accuracy in identifying structural variants. Nevertheless, no single algorithm can accurately and sensitively detect all types and sizes of SVs. This limitation has been highlighted in several studies that evaluated the individual performance of various callers.

In recent years, the scientific community has made considerable progress in improving short-read-based general-purpose SV detection. New algorithms and combination strategies have been proposed to enhance the accuracy and sensitivity of SV identification. For instance, the integration of multiple agreement strategies, which require the consensus of more than two tools, and union strategies, which combine the outputs of all selected algorithms, have shown promising results. These approaches have achieved higher recall rates compared to single algorithms and have attained similar recall and F1 scores to commercial software such as DRAGEN.

Furthermore, advancements in machine learning and artificial intelligence have opened new avenues for SV detection. By leveraging these technologies, researchers have



doi:10.6342/NTU202500995

developed sophisticated models that can better distinguish between true SVs and false positives. Additionally, the incorporation of long-read sequencing data, which provides more comprehensive coverage of the genome, has further improved the detection of complex structural variants.

In conclusion, while substantial advancements have been made in SV detection methods, challenges remain in accurately analyzing short-read sequences. The continued development of novel algorithms, combination strategies, and the integration of advanced technologies hold immense potential for overcoming these limitations and advancing the field of precision medicine.

## 3.2 Aims

We propose that combination strategies, which include multiple agreement strategies requiring the consensus of more than two tools, as well as union strategies that integrate the outputs of all selected algorithms, offer superior accuracy and sensitivity in detecting structural variants compared to using single algorithms. By employing these combination strategies, researchers can achieve higher recall rates, and similar recall and F1 scores to commercial software such as DRAGEN. These approaches allow for more comprehensive identification of structural variants, thereby enhancing the robustness of genomic analysis.

## 3.3 Study design

Our study reviews six state-of-the-art SV detection algorithms to establish effective SV detection strategies. These algorithms include LUMPY, which integrates signals from paired-end, split-read, and read-depth approaches to detect a wide range of structural variants[43]; Manta, known for its high sensitivity and precision in leveraging paired-end



doi:10.6342/NTU202500995

and split-read data[44]; Delly, which uses paired-end and split-read evidence to detect structural variants like deletions, duplications, inversions, and translocations[45]; GRIDSS, a hybrid SV caller that integrates assembly-based and read-pair based SV calling techniques[46]; SvABA, an assembly-based tool that calls structural variants by locally assembling reads around candidate breakpoints[47]; and Illumina DRAGEN[48], a commercial caller known for its superior speed and accuracy using hardware acceleration. By evaluating these six algorithms, we aim to identify the best-performing strategies for accurately and sensitively detecting structural variants in short-read sequences.

## 3.4 Materials and Methods

### 3.4.1 Truth set

The GIAB v0.6 Tier 1 benchmark set[49] was obtained from ftp://ftp-trace.ncbi.nlm.nih.gov/ReferenceSamples/giab/data/AshkenazimTrio/analysis/NIST_SVs_Integration_v0.6/. The HG00154, HG00733, and NA19240 benchmark set was obtained from HGSVC2[50] (https://www.internationalgenome.org/data-portal/data-collection/hgsvc2). The distribution of "DELs" and "INSs" of SVs in the truth set and each sample were detailed in Table S7. To assess the performances of variant calling tools, deletions (DELs) and insertions (INSs) were analyzed separately, allowing for independent performance evaluation of these two SV types.

### 3.4.2 SV detection algorithms

We selected structural variant (SV) detection algorithms capable of handling short-read whole-genome sequencing (WGS) through various methodologies. Based on the literature, five publicly available SV detection algorithms were selected: DELLY, LUMPY, Manta, GRIDSS, and SvABA. Additionally, the commercial SV detection tool,



doi:10.6342/NTU202500995

Illumina's DRAGEN (Dynamic Read Analysis for GENomics), was included for comparative performance analysis, both as individual algorithms and in conjunction with different algorithms. DELLY integrates read-pair (RP) and split-read (SR) methods for SV detection, utilizing paired-end short reads data from various insert sizes. LUMPY employs a probabilistic representation of SV breakpoints, leveraging RP and SR analyses while clustering overlapping breakpoints to accurately identify SVs. Manta operates with a two-phase workflow to achieve high parallelization, constructing a graph of break-end associations for variant hypothesis generation, assembly, scoring, and VCF reporting. GRIDSS detects SVs of any size using assembly (AS), RP, and SR methods, creating break-end contigs and applying a probabilistic model to infer SVs. SvABA efficiently detects SVs within local 25k base pairs assembly windows with 2k base pairs overlaps, demonstrating low resource requirements for short-read sequencing data analysis. The DRAGEN pipeline employs advanced computational techniques, including read-pair, read-depth, split-read, and assembly approaches, to effectively detect SVs in genomic data.

### 3.4.3 Calling and refinement of structural variations

We obtained the whole-genome paired-end short reads of HG002 (NA24385) from the National Institute of Standards and Technology (https://www.nist.gov), and HG00514, HG00733, and NA19240 from HGSVC2 (https://www.internationalgenome.org/data-portal/data-collection/hgsvc2). These raw reads were aligned to the GRCh37 (for HG002) or GRCh38 (for HG00514, HG00733, and NA19240) versions of the human reference genome using BWA-MEM, as the SV benchmark set was based on the references. Subsequently, the resulting BAM files were processed through various SV detection algorithms, utilizing their default parameters as recommended by the authors for calling



doi:10.6342/NTU202500995

SVs.

For the refinement, filtering, and deduplication, the variants underwent a systematic four-step process: 1) SV Annotation: we employed the R package[51] to annotate SV types and lengths. 2) Size Exclusion: SVs smaller than 50 bp were excluded. 3) Record Deduplication: In cases where some SV callers (GRIDSS, SvABA, LUMPY, and Manta) generated output files containing separate records for the start and end of the same SV, we removed the end position record. 4) Final Filtering: Only variants marked 'PASS' in the 'FILTER' field were retained for further analysis.

### 3.4.4 Combination strategies of multiple algorithms

Structural variants (SVs) were detected using multiple algorithms to enhance sensitivity and specificity, and a dual approach was employed to combine neighboring SV calls from different callers. In the Multiple-agreement method, SVs were merged if they satisfied the following criteria: an SV length greater than 49 bp, start positions within ±500 base pairs, and matching SV types, with support from at least two of the variant callers (Manta, DELLY, GRIDSS, LUMPY, or SvABA). In contrast, the Union method required a call from at least one of the variant callers (Manta, DELLY, GRIDSS, LUMPY, or SvABA) and used less stringent criteria. This combinatorial strategy ensured robust and reliable integration of SV calls across multiple detection platforms, thereby improving the overall accuracy of SV identification in our genomic analyses.

Individual SVs were annotated with the name of their respective SV detection algorithms in the 'INFO' field (e.g., caller=DELLY). For grouping, one or more variants from different SV callers were considered a neighbor group if they existed within the range of +/- 500 bp of 'POS' (start position) and shared the same 'SVTYPE'. When referring to candidate SVs within a neighbor group, we retained the variant with the



doi:10.6342/NTU202500995

smallest start position as the representative.

Merging variants were conducted in the following ways: a combination of three algorithms, DELLY, Manta, and GRIDSS, formed the three SV (III) combination group, while a combination of five algorithms, Manta, DELLY, GRIDSS, LUMPY, and SvABA, formed the five SV (V) combination group. The “multiple agreement” strategy defined a 'true-called variant' if the members of a neighboring SV group were detected by at least two or more algorithms. Conversely, the union strategy defined a “true-called variant” if at least one member of a neighbor SV was detected by at least one algorithm, as demonstrated in (Figure 12).

| **Combinaiton of Multiple Callers** | |
|---|---|
| **Multiple-agreement** | **Union** |
| Meet all of the following criteria:<br>1. 49 < SVLEN and<br>2. “Start position” +/- 500 base pairs and<br>3. Same SV type<br>4. Callers ≥ 2 | Meet all of the following criteria:<br>1. 49 < SVLEN and<br>2. “Start position” +/- 500 base pairs and<br>3. Same SV type<br>4. Callers ≥ 1 |
| III: Manta, DELLY, GRIDSS<br>V: Manta, DELLY, GRIDSS, LUMPY, SvABA | III: Manta, DELLY, GRIDSS<br>V: Manta, DELLY, GRIDSS, LUMPY, SvABA |

**Figure 12. Combination of neighbor SV from multiple callers**

### 3.4.5 Evaluation of performance for single algorithm and combination strategies

The recall, precision, and F1 score for different single algorithms and combinations of multiple algorithms, either in “multiple agreement” strategy or union strategy, were calculated. We focused on the SV types 'DEL' and 'INS' to analyze the performance of different single algorithms and various combination strategies, because the truth sets only provide benchmarks for these types of SVs [49]. Variants detected by single algorithms



doi:10.6342/NTU202500995

and combination strategies (referred to as 'TEST' sets) were compared against the truth sets.

For evaluation, the true positive (TP) was defined if a variant in the truth set was located within the range of +/-500 bp of the 'POS' (start position) of a variant in the 'TEST' set and shared the same SV type. False positives (FP) were calculated as the total variant count of the 'TEST' set minus TP, and false negatives (FN) were calculated as the total variant count of the 'TRUTH' set minus TP. Precision, recall, and F1 score were then calculated using the following formulas.

$$\text{precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{recall} = \frac{TP}{TP+FN} \quad (2)$$

$$F1 = 2 \times \frac{\text{precison} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

To determine the rank of overall performance of each single caller and combination strategies, combined precision score (cPr), combined recall score (cRc), and combined F1 score (cF1), in which the values for all four data sets were integrated, and were calculated by micro average as follows:

$$cPr = \frac{TP_1+TP_2+TP_3+TP_4}{TP_1+TP_2+TP_3+TP_4 + FP_1+FP_2+FP_3+FP_4} \quad (4)$$

$$cRc = \frac{TP_1+TP_2+TP_3+TP_4}{TP_1+TP_2+TP_3+TP_4+FN_1+FN_2+FN_3+FN_4} \quad (5)$$

$$cF1 = 2 \times \frac{cPr \times cRc}{cPr + cRc} \quad (6)$$



doi:10.6342/NTU202500995

### 3.4.6 Computational resources

All bioinformatic computations were conducted on the Taiwania 3 computational platform of the Taiwan National Center for High-Performance Computing (NCHC), recognized as the most powerful CPU high-performance computing server available for open- service applications in Taiwan. A Python script was developed to refine, filter, merge, intersect, and union the variants, enabling the calculation of recall, precision, and F1 score for different single algorithms and various combination strategies.

## 3.5 Results

### 3.5.1 SV detection methods included in this study

We began by surveying the current literature on short-read structural variant (SV) detection tools, evaluating them based on performance, popularity, active maintenance, the quality of their software and documentation, and their compatibility with our computational setup. Based on these criteria, along with recent benchmarking studies[52-55], we selected five high-performing SV detection algorithms—DELLY, LUMPY, Manta, GRIDSS, and SvABA—for our analysis. To offer a more comprehensive evaluation, we also included the commercial SV detection tool DRAGEN. All selected callers were run using their recommended settings, and their outputs were carefully recorded. These tools are designed to analyze short-read whole genome sequencing (WGS) data by processing BAM files of aligned reads, each employing distinct methodological approaches. Additionally, most of these algorithms are implemented in multiple programming languages and are optimized for multi-core architecture. Table 5 summarizes their dependencies as well as their respective capabilities in detecting specific types of SVs.



doi:10.6342/NTU202500995

**Table 5. Overview of SV detection methods and the types of SVs identified by different tools**

| SV caller | Version | Algorithm | Published year | First author | Latest updated year | Tools webpage |
|---|---|---|---|---|---|---|
| Manta | v1.6.0 | RP-SR-AS | 2016 | Xiaoyu Chen | 2019 | https://github.com/Illumina/manta |
| DELLY | v1.1.5 | RP-SR | 2012 | Tobias Rausch | 2024 | https://github.com/DELLYtools/DELLY |
| GRIDSS | v2.13.2 | RP-SR-AS | 2017 | Daniel L Cameron | 2023 | https://github.com/PapenfussLab/GRIDSS |
| LUMPY | v0.3.1 | RP-SR-RD | 2014 | Ryan M Layer | 2024 | https://github.com/brentp/smoove |
| SvABA | v1.1.0 | RP-SR-AS | 2018 | Jeremiah A Wala | 2024 | https://github.com/walaj/svaba |

Input: BAM; Output: VCF (Variant calling format). RP: read-pair; SR: split read; RD: read depth; AS: assembly.



doi:10.6342/NTU202500995

### 3.5.2 Evaluation of SV detection methods based on the GIAB benchmark set and well-referenced cell lines

We evaluated the performance of various SV detection tools using the HG002 sample, which comes with a high-confidence SV call set mapped to GRCh37. HG002, part of a well-consented Ashkenazim trio from the Genome in a Bottle Consortium (GIAB), provided whole-genome paired-end short read data obtained via the National Institute of Standards and Technology (https://www.nist.gov). To further enhance sample diversity and enrich our assessment, we incorporated benchmark datasets from HGSVC2[50], which offers a catalog of SVs—including insertions (INSs) and deletions (DELs)—across over 3,000 genomes from the 1000 Genomes Project. In addition, three widely recognized cell lines, namely HG00514, HG00733, and NA19240, were also obtained from HGSVC2 for our analysis.

The raw reads were aligned to the respective human reference genomes—GRCh37 for HG002 and hg38 for HG00514, HG00733, and NA19240—using BWA-MEM, following settings consistent with the truth sets (Figure 13 illustrates our workflow). The resulting BAM files then served as input for a series of SV detection algorithms, each run with default parameters as per the creators' recommendations. In total, six different SV callers, employing distinct analytical approaches, successfully identified all five SV categories present in the truth set: deletions (DELs), insertions (INSs), duplications (DUPs), inversions (INVs), and translocations (CTXs) (Figure 14 and Supplemental Tables S3–S6).



doi:10.6342/NTU202500995

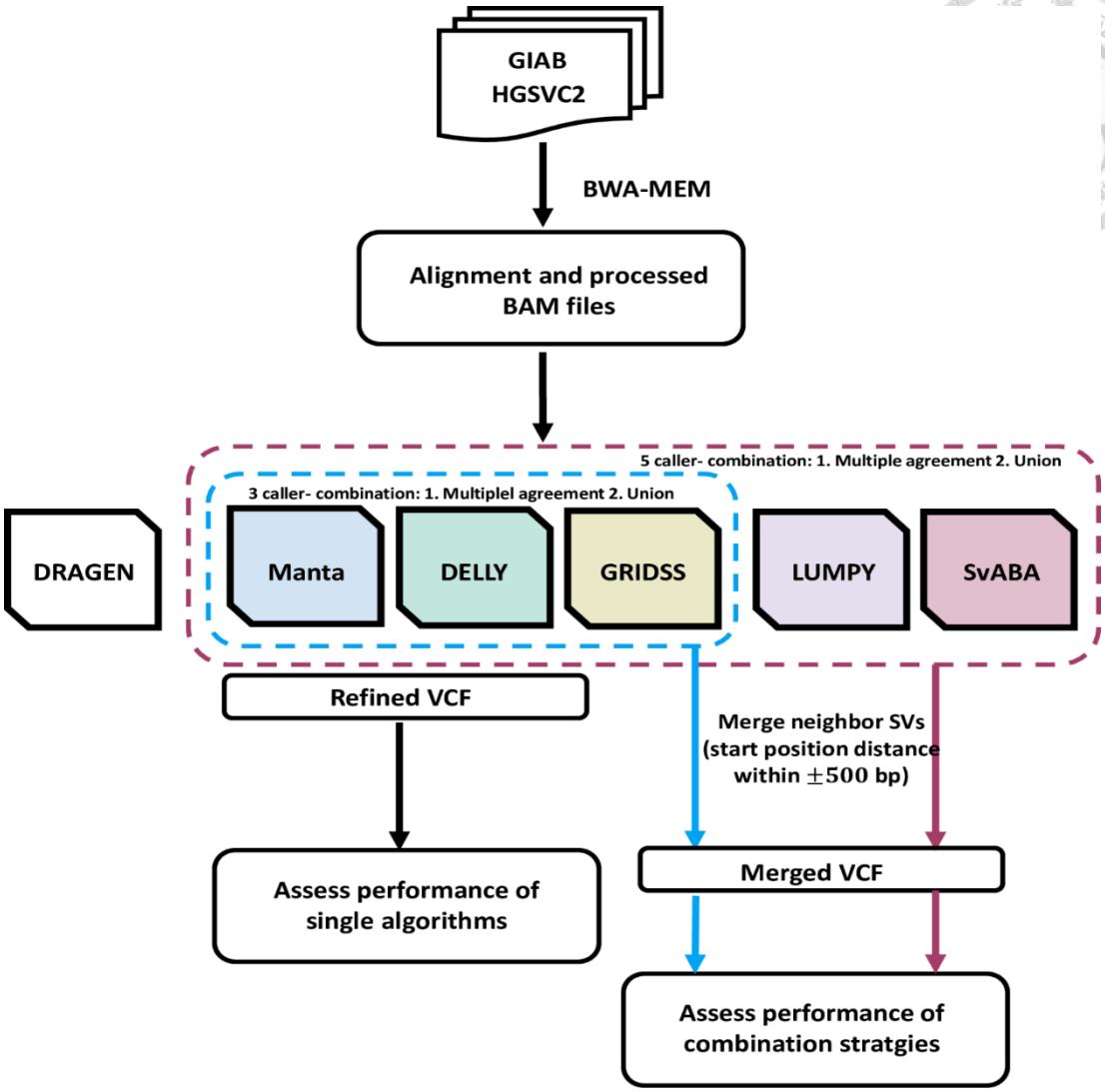



**Figure 13. SV detection workflow.**

Benchmark sets from GIAB and HGSVC2 are aligned with BWA-MEM to create BAM files. These BAM files are then processed by various SV callers (DRAGEN, Manta, DELLY, LUMPY, GRIDSS, SvABA) to generate VCF files, which are evaluated based on recall, precision, and F1 score using both single tool and combination strategies.



doi:10.6342/NTU202500995

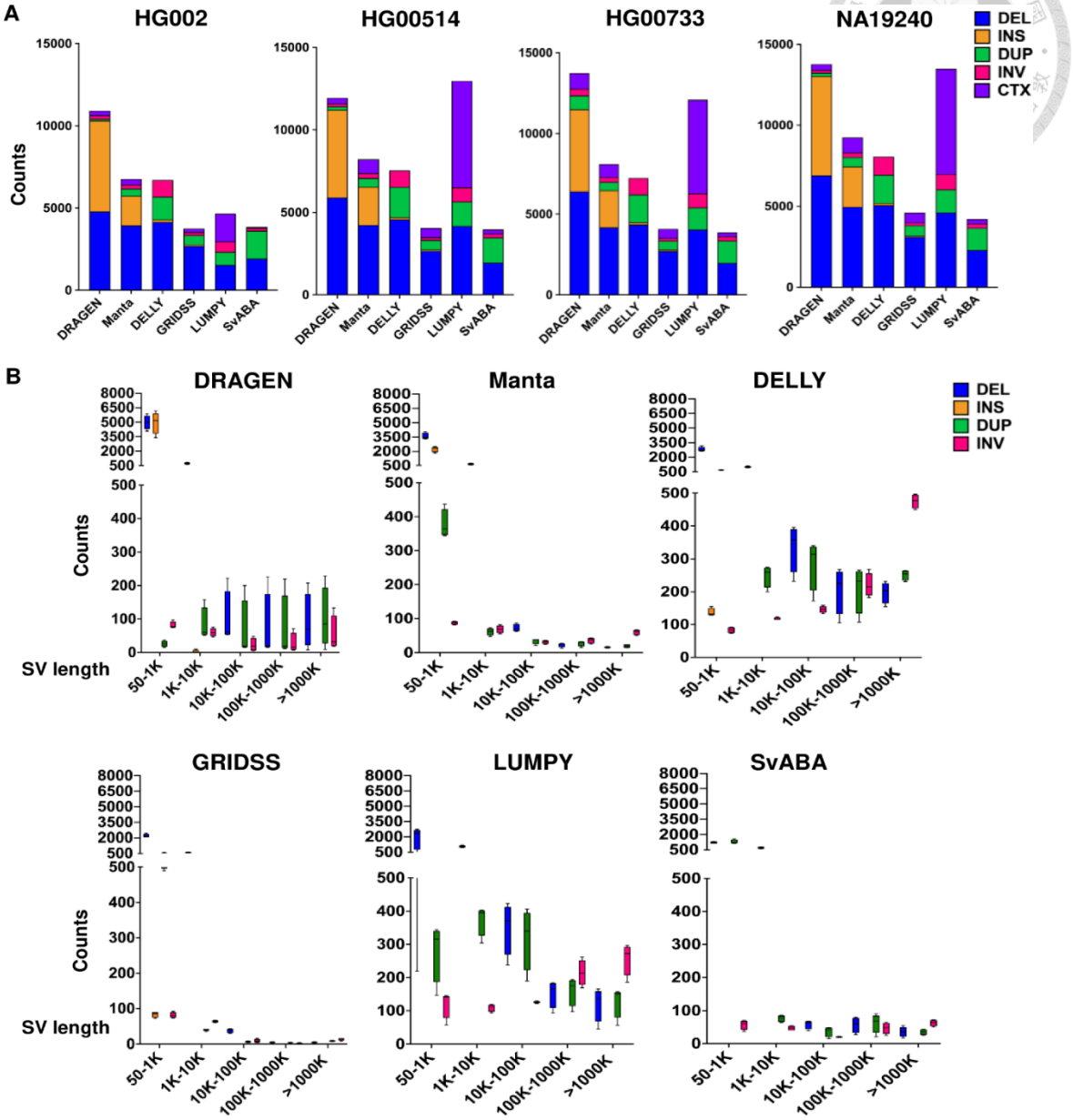



**Figure 14. The SVs detected by six SV callers.** (A) SVs of sizes ≥ 50 bp and CTX detected by individual SV callers in samples, HG002, HG00514, HG00733, and NA19240. (B) Length distribution of different variants for all samples detected by individual SV callers. The maximum, minimum, and median are based on the integrated values from all sample sets. DEL: deletion, INS: insertion, DUP: duplication, INV: inversion, and CTX: complex translocation.



doi:10.6342/NTU202500995

Figure 14 and Tables S3–S6 highlight the varied structural variants of the six SV callers. In the HG002 dataset, the DRAGEN algorithm produced the highest number of SV calls, totaling 10,912, followed by Manta with 7,758 SVs; the remaining tools—DELLY, LUMPY, SvABA, and GRIDSS—detected fewer variants. A similar detection trend was observed in the three additional datasets from HGSVC2, where LUMPY notably identified more complex translocations (CTX) than the other algorithms (Figure 14A).

We also examined the SV size distributions across the different tools, revealing significant differences (Figure 14B and Table S3–S6). For instance, GRIDSS uncovered a large number of small indels, while DELLY and LUMPY, despite both relying on split-read and read-pair approaches, differed in the sizes and types of SVs they reported. In contrast, DRAGEN, Manta, and SvABA primarily detected SVs ranging from 50 to 10,000 base pairs. All algorithms consistently identified deletions, duplications, and inversions, yet although every tool detected insertions, LUMPY and SvABA showed particularly low sensitivity for these variants. Notably, all algorithms, except DELLY, were able to detect CTXs. Overall, the distribution of SV types remained consistent across the datasets analyzed.

### 3.5.3 The individual performance of each algorithm

To assess the recall, precision, and F1 score for each SV detection algorithm, we employed the GIAB v0.6 Tier 1 benchmark dataset, which spans approximately 2.51 billion base pairs of the genome and represents high-confidence regions where every variant has been confirmed by at least one diploid assembly. In addition, we obtained benchmark datasets for HG00514, HG00733, and NA19240 from HGSVC2[50]. These benchmarks primarily encompass SVs ranging in size from 50 to 10,000 base pairs. We



doi:10.6342/NTU202500995

analyzed the size distribution of insertions (INSs) and deletions (DELs) (Figure 15A). Notably, because LUMPY and SvABA did not detect any INSs larger than 50 base pairs (Figure 14 and Table S3–S6), our overall performance evaluation was based solely on deletions, with LUMPY and SvABA being excluded from the INS analysis.

To assess each algorithm's performance, we defined matching SVs as those located within ± 500 base pairs of the truth set's start position and exhibiting the same SV type. In the HG002 dataset, for DELs, DRAGEN outperformed the other algorithms by identifying 3,425 true positives (TPs), with Manta, DELLY, GRIDSS, and SvABA trailing behind, and LUMPY detecting the fewest TPs. In terms of total errors (false positives plus false negatives), DRAGEN and Manta recorded the lowest error rates, while LUMPY showed the highest. A similar pattern of performance was observed across the additional HGSVC2 datasets. For INSs, DRAGEN and Manta achieved the highest counts of true positives, while DELLY and GRIDSS, despite yielding fewer false positives, suffered from excessively high false negatives across all evaluated datasets (Figure 15 and Table 6).



doi:10.6342/NTU202500995

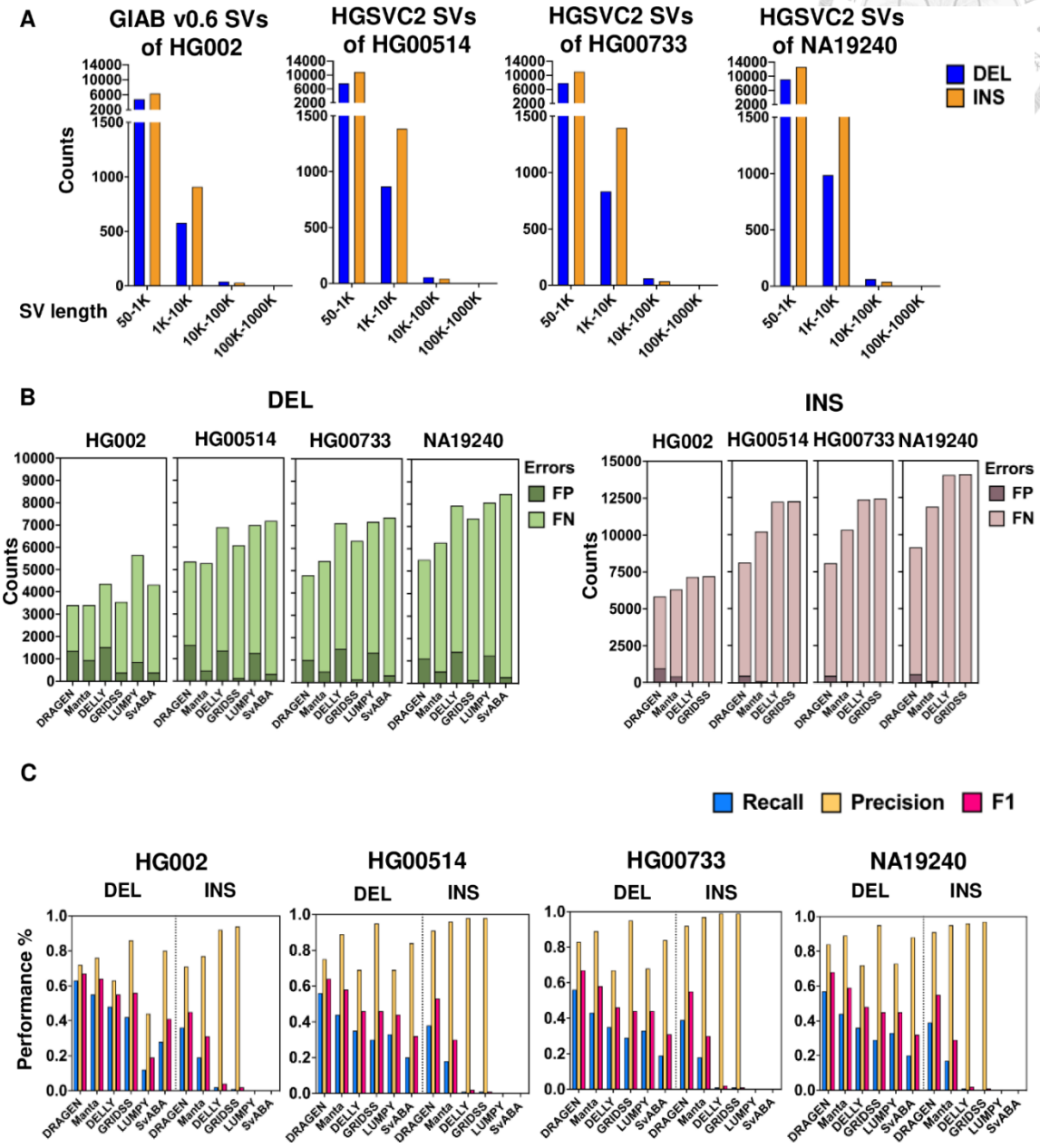



**Figure 15. The distribution of SVs in truth sets and the performance of individual algorithms.** (A) The sizes and counts of "DELs" and "INSs" in each samples. (B) The comparison of False negative (FN) and False positive (FP) numbers among individual algorithms. (C) The precision, recall, and F1 score of each individual algorithm in detecting "DELs" and "INSs" with the size ≥ 50 bp. DEL: deletion, INS: insertion.



doi:10.6342/NTU202500995

**Table 6. Performance of single algorithm in detecting DELs and INSs.**

| HG002 | | | | | | | |
|---|---|---|---|---|---|---|---|
| **SV caller** | **DEL** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 4,787 | **3,425** | 1,362 | **2,039** | **0.63** | 0.72 | **0.67** |
| Manta | 3,935 | 2,996 | 939 | 2,468 | 0.55 | 0.76 | 0.64 |
| DELLY | 4,142 | 2,620 | 1,522 | 2,844 | 0.48 | 0.63 | 0.55 |
| GRIDSS | 2,681 | 2,299 | **382** | 3,165 | 0.42 | **0.86** | 0.56 |
| LUMPY | 1,528 | 668 | 860 | 4,796 | 0.12 | 0.44 | 0.19 |
| SvABA | 1,910 | 1,526 | 384 | 3,938 | 0.28 | 0.80 | 0.41 |
| | | | | | | | |
| **SV Caller** | **INS** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 3,403 | **2,426** | 977 | **4,855** | **0.36** | 0.71 | **0.45** |
| Manta | 1,800 | 1,387 | 413 | 5,894 | 0.19 | 0.77 | 0.31 |
| DELLY | 155 | 142 | 13 | 7,139 | 0.02 | 0.92 | 0.04 |
| GRIDSS | 90 | 85 | **5** | 7,196 | 0.01 | **0.94** | 0.02 |
| **HG00514** | | | | | | | |
| **SV caller** | **DEL** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 6,403 | **4,790** | 1,613 | **3,739** | **0.56** | 0.75 | **0.64** |
| Manta | 4,183 | 3,710 | 473 | 4,819 | 0.44 | 0.89 | 0.58 |
| DELLY | 4,359 | 2,996 | 1,363 | 5,533 | 0.35 | 0.69 | 0.46 |
| GRIDSS | 2,718 | 2,582 | **136** | 5,947 | 0.30 | **0.95** | 0.46 |
| LUMPY | 4,053 | 2,798 | 1,255 | 5,731 | 0.33 | 0.69 | 0.44 |
| SvABA | 1,981 | 1,668 | 313 | 6,861 | 0.20 | 0.84 | 0.32 |
| | | | | | | | |
| **SV Caller** | **INS** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 5,100 | **4,654** | 446 | **7,679** | **0.38** | 0.91 | **0.53** |
| Manta | 2,291 | 2,204 | 87 | 10,129 | 0.18 | 0.96 | 0.30 |
| DELLY | 128 | 126 | 2 | 12,207 | 0.01 | **0.98** | 0.02 |
| GRIDSS | 90 | 88 | **2** | 12,245 | 0.01 | **0.98** | 0.01 |
| **HG00733** | | | | | | | |
| **SV caller** | **DEL** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 5,895 | **4,896** | 999 | **3,798** | **0.56** | 0.83 | **0.67** |
| Manta | 4,232 | 3,749 | 483 | 4,945 | 0.43 | 0.89 | 0.58 |
| DELLY | 4,574 | 3,075 | 1,499 | 5,619 | 0.35 | 0.67 | 0.46 |
| GRIDSS | 2,655 | 2,512 | **143** | 6,182 | 0.29 | **0.95** | 0.44 |
| LUMPY | 4,170 | 2,843 | 1,327 | 5,851 | 0.33 | 0.68 | 0.44 |
| SvABA | 1,944 | 1,634 | 310 | 7,060 | 0.19 | 0.84 | 0.31 |
| | | | | | | | |
| **SV Caller** | **INS** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 5,313 | **4,868** | 445 | **7,627** | **0.39** | 0.92 | **0.55** |
| Manta | 2,312 | 2,236 | 76 | 10,259 | 0.18 | 0.97 | 0.30 |
| DELLY | 130 | 129 | 1 | 12,366 | 0.01 | **0.99** | 0.02 |
| GRIDSS | 81 | 80 | **1** | 12,415 | 0.01 | **0.99** | 0.01 |
| **NA19240** | | | | | | | |
| **SV caller** | **DEL** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |



doi:10.6342/NTU202500995

| DRAGEN | 6,897 | **5,775** | 1,122 | **4,390** | **0.57** | 0.84 | **0.68** |
|---|---|---|---|---|---|---|---|
| Manta | 4,955 | 4,423 | 532 | 5,742 | 0.44 | 0.89 | 0.59 |
| DELLY | 5,061 | 3,649 | 1,412 | 6,516 | 0.36 | 0.72 | 0.48 |
| GRIDSS | 3,098 | 2,957 | **141** | 7,208 | 0.29 | **0.95** | 0.45 |
| LUMPY | 4,608 | 3,354 | 1,254 | 6,811 | 0.33 | 0.73 | 0.45 |
| SvABA | 2,283 | 2,001 | 282 | 8,164 | 0.20 | 0.88 | 0.32 |
| | | | | | | | |
| **SV Caller** | **INS** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 6,144 | **5,574** | 570 | **8,579** | **0.39** | 0.91 | **0.55** |
| Manta | 2,488 | 2,375 | 113 | 11,778 | 0.17 | 0.95 | 0.29 |
| DELLY | 134 | 129 | 5 | 14,024 | 0.01 | 0.96 | 0.02 |
| GRIDSS | 72 | 70 | **2** | 14,083 | 0.00 | **0.97** | 0.01 |

We computed recall, precision, and F1 scores for each algorithm on deletions (DELs). In the HG002 dataset, DRAGEN registered the highest F1 scores (ranging from 0.64 to 0.68) across all benchmark sets, with Manta, DELLY, and GRIDSS following closely (Figure 15C and Table 6). In contrast, LUMPY recorded the lowest F1 score (0.19) for HG002, and although SvABA performed moderately in HG002, its results were poorer in the other datasets. Among the non-commercial tools, Manta showed the highest recall (0.43–0.55%), whereas GRIDSS, while maintaining high precision levels (0.86–0.95%), exhibited lower recall.

When assessing INSs specifically, most of the individual algorithms demonstrated disappointing performance. Across all datasets, DRAGEN produced the highest F1 scores for insertions (0.45–0.55), with Manta performing nearly as well. Conversely, DELLY (F1: 0.02–0.04) and GRIDSS (F1: 0.01–0.02) performed poorly. Among non-commercial software, Manta again achieved the highest recall for INSs (0.16–0.19) but with only moderate precision (0.77–0.96), while both GRIDSS and DELLY displayed excellent precision (0.91–0.99) yet extremely low recall (0.00–0.02%), which ultimately compromised their overall performance. This overall low recall for insertions is clearly depicted in Figure 15C and summarized in Table 6.

In addition to accuracy metrics, we evaluated the computational efficiency of each



doi:10.6342/NTU202500995

SV caller using three datasets from HGSVC2. The analysis considered both the total CPU time, which measures overall CPU consumption, and the maximum memory usage (Figure S1). It is important to note that wall time—the actual elapsed time—can be considerably shorter than the total CPU time when efficient parallel processing is employed. Among the non-commercial tools, Manta consistently demonstrated the shortest runtime and minimal memory usage, while GRIDSS incurred the highest computational demands in terms of both processing time and memory consumption.

### 3.5.4 Systematic identification of the performances of combination strategies

Due to the unsatisfactory performance of individual algorithms, we adopted a strategy—commonly applied in previous studies[50, 52, 56]—that favors structural variants (SVs) identified by multiple callers to improve precision. To enhance SV calling accuracy, we evaluated various combination strategies by calculating accuracy, precision, recall, and F1 scores under two frameworks: the “multiple agreement” strategy and the union strategy, each applied to either three or five algorithms. For the three-caller combination, we selected Manta, DELLY, and GRIDSS, as they consistently showed superior performance. For the five-caller approach, we incorporated all tools except the commercial DRAGEN product. Each SV was annotated with its calling algorithm in the 'INFO' field.

Following consensus from the GIAB benchmark study—which excluded potentially complex variants if two or more supported variants (≥50 bp) were within 1000 bp—we set a ±500 bp overlap threshold to group SVs. Variants from different callers were grouped as “neighboring SVs” if they overlapped within ±500 bp at their start positions and shared the same SV type. In the “multiple agreement” strategy, a variant was



doi:10.6342/NTU202500995

considered true-positive only if at least two algorithms in the neighboring group detected it, while the union strategy treated a variant as true-called if at least one caller identified it.

We further analyzed the distance distribution between variant positions within neighboring SV groups and summarized the counts and the number of contributing callers (Table S8). All grouped variants fell within a 0–500 bp range. Specifically, the median distance was 4.0–5.0 bp for the three-caller multiple agreement strategy (III-multiple agreement), 7.0–10.0 bp for the five-caller multiple agreement strategy (V-multiple agreement), and 0 bp for both union strategies. The interquartile ranges (IQR) were 7.0 bp for three-caller multiple agreement, 17.0–23.0 bp for five-caller multiple agreement, 1.0–2.0 bp for three-caller union (III-union), and 7.0–10.0 bp for five-caller union (V-union), indicating that most neighboring SVs were detected in close proximity (Figure S2).

Moreover, the majority of neighboring SV groups in the three-caller set comprised fewer than three members, and similarly, those in the five-caller set comprised fewer than five (Table S8). This reflects that only a small fraction of distinct SVs were grouped under the ±500 bp merge condition. The degree of concordance among the algorithms is visually depicted in the Venn diagram (Figure 16), while the overall SV distribution under the different combination strategies (Figure 17A and Table S9).

For DELs in HG002, DRAGEN achieved the highest true positive count at 3,425, followed by the V-union, III-union, and V-multiple agreement strategies, with the III-multiple agreement approach yielding the lowest number of true positives at 2,554. In contrast, the false positive count was lowest for III-multiple agreement (533) and highest for V-union (2,199). A similar pattern was observed in the HGSVC2 datasets. When total errors (the sum of false positives and false negatives) were considered, combination



doi:10.6342/NTU202500995

strategies employing all five algorithms generally resulted in higher error rates compared to those using three algorithms. For insertions (INSs), DRAGEN again recorded the highest true positive counts, while the multiple agreement strategies incurred the highest overall error rates. Overall, despite the use of combination strategies, the high false negative rate remained a significant concern (Figure 17B and Table 7). We also computed recall, precision, and the F1 score. For deletions across the samples, DRAGEN achieved the highest F1 score (ranging between 0.64 and 0.68), followed by the union strategies (both III-union and V-union), while the multiple agreement strategies (III-multiple and V-multiple agreement) had lower F1 scores (Figure 17C and Table 7). DRAGEN also showed the highest recall, with the union groups trailing closely (recall values between 0.50 and 0.62), whereas the multiple agreement groups yielded lower recall (from 0.33 to 0.48). Conversely, precision was notably higher in the multiple agreement groups (ranging from 0.74 to 0.93) compared to the union groups, which achieved precision levels between 0.61 and 0.76.



doi:10.6342/NTU202500995

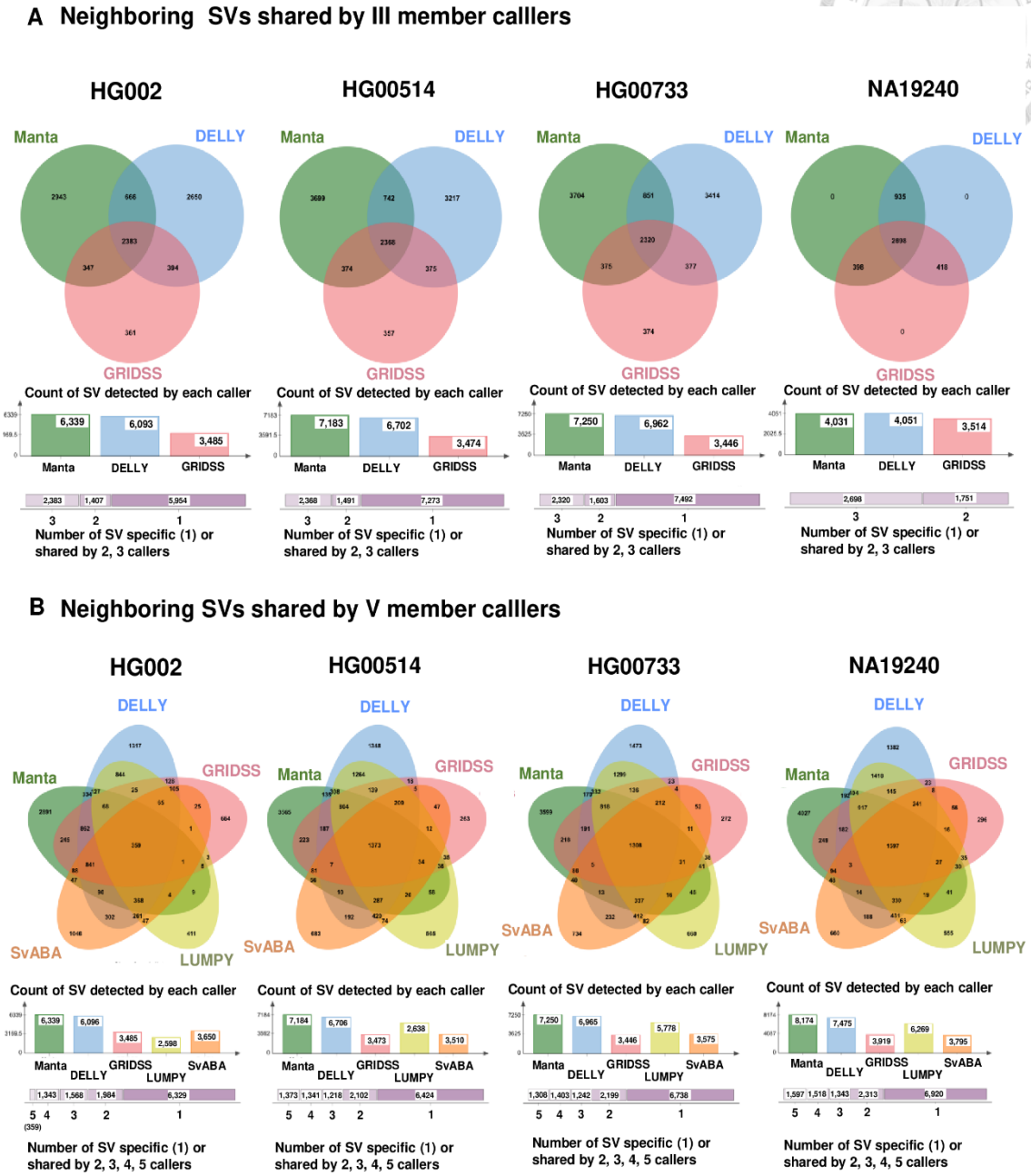



**Figure 16. Concordance of neighbor SVs detected by member callers in different combination strategies.** (A) The agreement among three SV detection tools (Manta, DELLY, and GRIDSS). (B) The agreement among five SV detection tools (Manta, DELLY, GRIDSS, LUMPY, and SvABA).



doi:10.6342/NTU202500995

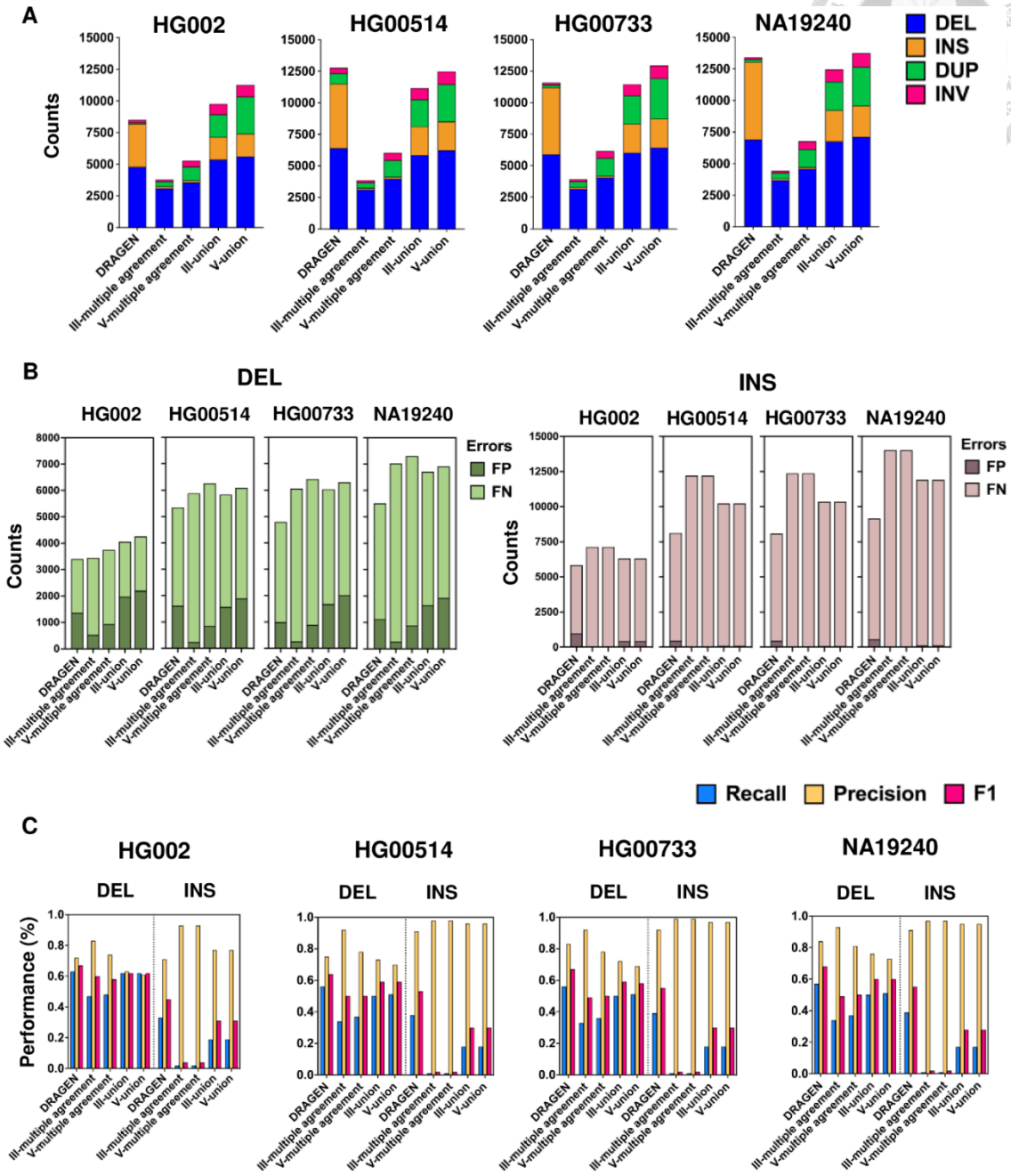



**Figure 17. Distribution and performance of combination strategies.** (A) The distribution of SVs in different combination strategies. (B) The comparison of False negative (FN) and False positive (FP) numbers among individual algorithms. (C) The precision, recall, and F1 score of combination strategies in the detection of DELs and INSs. DEL: deletion, INS: insertions, DUP: duplication, INV: inversion.



doi:10.6342/NTU202500995

**Table 7. Performance of different combination strategies in detecting DELs and INSs.**

| HG002 | | | | | | | |
|---|---|---|---|---|---|---|---|
| **SV caller** | **DEL** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 4,787 | **3,425** | 1,362 | **2,039** | **0.63** | 0.72 | **0.67** |
| III-multiple agreement | 3,087 | 2,554 | **533** | 2,910 | 0.47 | **0.83** | 0.60 |
| V-multiple agreement | 3,565 | 2,638 | 927 | 2,826 | 0.48 | 0.74 | 0.58 |
| III-union | 5,353 | 3,380 | 1,973 | 2,084 | 0.62 | 0.63 | 0.62 |
| V-union | 5,605 | 3,406 | 2,199 | 2,058 | 0.62 | 0.61 | 0.62 |
| | | | | | | | |
| **SV Caller** | **INS** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 3,403 | **2,426** | 977 | **4,855** | **0.33** | 0.71 | **0.45** |
| III-multiple agreement | 162 | 150 | **12** | 7,131 | 0.02 | **0.93** | 0.04 |
| V-multiple agreement | 162 | 150 | **12** | 7,131 | 0.02 | **0.93** | 0.04 |
| III-union | 1,799 | 1,387 | 412 | 5,894 | 0.19 | 0.77 | 0.31 |
| V-union | 1,799 | 1,387 | 412 | 5,894 | 0.19 | 0.77 | 0.31 |
| **HG00514** | | | | | | | |
| **SV caller** | **DEL** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 6,403 | **4,790** | 1,613 | **3,739** | **0.56** | 0.75 | **0.64** |
| III-multiple agreement | 3,125 | 2,886 | **239** | 5,643 | 0.34 | **0.92** | 0.50 |
| V-multiple agreement | 3,988 | 3,130 | 858 | 5,399 | 0.37 | 0.78 | 0.50 |
| III-union | 5,842 | 4,269 | 1,573 | 4,260 | 0.50 | 0.73 | 0.59 |
| V-union | 6,234 | 4,340 | 1,894 | 4,189 | 0.51 | 0.70 | 0.59 |
| | | | | | | | |
| **SV Caller** | **INS** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 5,100 | **4,654** | 446 | **7,679** | **0.38** | 0.91 | **0.53** |
| III-multiple agreement | 132 | 129 | **3** | 12,204 | 0.01 | **0.98** | 0.02 |
| V-multiple agreement | 132 | 129 | **3** | 12,204 | 0.01 | **0.98** | 0.02 |
| III-union | 2,277 | 2190 | 87 | 10,143 | 0.18 | 0.96 | 0.30 |
| V-union | 2,277 | 2190 | 87 | 10,143 | 0.18 | 0.96 | 0.30 |
| **HG00733** | | | | | | | |
| **SV caller** | **DEL** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 5,895 | **4,896** | 999 | **3,798** | **0.56** | 0.83 | **0.67** |
| III-multiple agreement | 3,154 | 2,894 | **260** | 5,800 | 0.33 | **0.92** | 0.49 |
| V-multiple agreement | 4,043 | 3,158 | 885 | 5,536 | 0.36 | 0.78 | 0.50 |
| III-union | 6,013 | 4,336 | 1,677 | 4,358 | 0.50 | 0.72 | 0.59 |
| V-union | 6,426 | 4,413 | 2,013 | 4,281 | 0.51 | 0.69 | 0.58 |
| | | | | | | | |
| **SV Caller** | **INS** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 5,313 | **4,868** | 445 | **7,627** | **0.39** | 0.92 | **0.55** |



doi:10.6342/NTU202500995

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| III-multiple agreement | 136 | 135 | **1** | 12,360 | 0.01 | **0.99** | 0.02 |
| V-multiple agreement | 136 | 135 | **1** | 12,360 | 0.01 | **0.99** | 0.02 |
| III-union | 2,293 | 2,218 | 75 | 10,277 | 0.18 | 0.97 | 0.30 |
| V-union | 2,293 | 2,218 | 75 | 10,277 | 0.18 | 0.97 | 0.30 |
| **NA19240** | | | | | | | |
| **SV caller** | **DEL** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 6,897 | **5,775** | 1,122 | **4,390** | **0.57** | 0.84 | **0.68** |
| III-multiple agreement | 3,694 | 3,418 | **276** | 6,747 | 0.34 | **0.93** | 0.49 |
| V-multiple agreement | 4,594 | 3,723 | 871 | 6,442 | 0.37 | 0.81 | 0.50 |
| III-union | 6,751 | 5,102 | 1,649 | 5,063 | 0.50 | 0.76 | 0.60 |
| V-union | 7,100 | 5,176 | 1,924 | 4,989 | 0.51 | 0.73 | 0.60 |
| | | | | | | | |
| **SV Caller** | **INS** | **TP** | **FP** | **FN** | **Recall** | **Precision** | **F1** |
| DRAGEN | 6,144 | **5,574** | 570 | **8,579** | **0.39** | 0.91 | **0.55** |
| III-multiple agreement | 135 | 131 | **4** | 14,022 | 0.01 | **0.97** | 0.02 |
| V-multiple agreement | 135 | 131 | **4** | 14,022 | 0.01 | **0.97** | 0.02 |
| III-union | 2,481 | 2,366 | 115 | 11,787 | 0.17 | 0.95 | 0.28 |
| V-union | 2,481 | 2,366 | 115 | 11,787 | 0.17 | 0.95 | 0.28 |

For INSs, the combination strategies also yielded disappointing results across all benchmark sets (Figure 17 and Table 7). DRAGEN attained the highest F1 scores, ranging from 0.45 to 0.55, followed by the union strategies, which achieved F1 scores between 0.28 and 0.31. In contrast, the multiple agreement strategies performed the worst, with F1 scores as low as 0.02–0.04. Notably, increasing the number of tools from three to five did not enhance performance, regardless of whether the multiple agreement or union strategies were implemented.

To establish an overall performance ranking of each individual caller and combination strategy in detecting DELs and INSs, we calculated aggregated metrics: the combined precision score (cPr), combined recall score (cRc), and combined F1 score (cF1). These metrics were derived by integrating the values from all four datasets using a micro average (Table 8). Figure 18 further illustrates the comprehensive performance across all callers based on the integrated values from all sample sets.



doi:10.6342/NTU202500995

**Table 8. Combined performance of single caller and different combination strategies in detecting DELs and INSs.**

| SV caller | DEL | TP | FP | FN | Recall | Precision | F1 |
|---|---|---|---|---|---|---|---|
| DRAGEN | 23,982 | **18,886** | 5,096 | 13,966 | **0.57** | 0.79 | **0.66** |
| Manta | 17,305 | 14,878 | 2,427 | 17,974 | 0.45 | 0.86 | 0.59 |
| DELLY | 18,136 | 12,340 | 5,796 | 20,512 | 0.38 | 0.68 | 0.48 |
| GRIDSS | 11,152 | 10,350 | **802** | 22,502 | 0.32 | **0.93** | 0.47 |
| LUMPY | 14,359 | 9,663 | 4,696 | 23,189 | 0.29 | 0.67 | 0.41 |
| SvABA | 8,118 | 6,829 | 1,289 | **26,023** | 0.21 | 0.84 | 0.33 |
| III-multiple agreement | 13,060 | 11,752 | 1,308 | 21,100 | 0.36 | 0.90 | 0.51 |
| V-multiple agreement | 16,190 | 12,649 | 3,541 | 20,203 | 0.39 | 0.78 | 0.52 |
| III-union | 23,959 | 17,087 | 6,872 | 15,765 | 0.52 | 0.71 | 0.60 |
| V-union | 25,365 | 17,335 | 8,030 | 15,517 | 0.53 | 0.68 | 0.60 |

| SV caller | INS | TP | FP | FN | Recall | Precision | F1 |
|---|---|---|---|---|---|---|---|
| DRAGEN | 19,960 | **17,522** | 2,438 | **28,740** | **0.38** | 0.88 | **0.53** |
| Manta | 8,891 | 8,202 | 689 | 38,060 | 0.18 | 0.92 | 0.30 |
| DELLY | 547 | 526 | 21 | 45,736 | 0.01 | 0.96 | 0.02 |
| GRIDSS | 333 | 323 | **10** | 45,939 | 0.01 | **0.97** | 0.01 |
| III-multiple agreement | 565 | 545 | 20 | 45,717 | 0.01 | 0.96 | 0.02 |
| V-multiple agreement | 565 | 545 | 20 | 45,717 | 0.01 | 0.96 | 0.02 |
| III-union | 8,850 | 8,161 | 689 | 38,101 | 0.18 | 0.92 | 0.30 |
| V-union | 8,850 | 8,161 | 689 | 38,101 | 0.18 | 0.92 | 0.30 |



doi:10.6342/NTU202500995

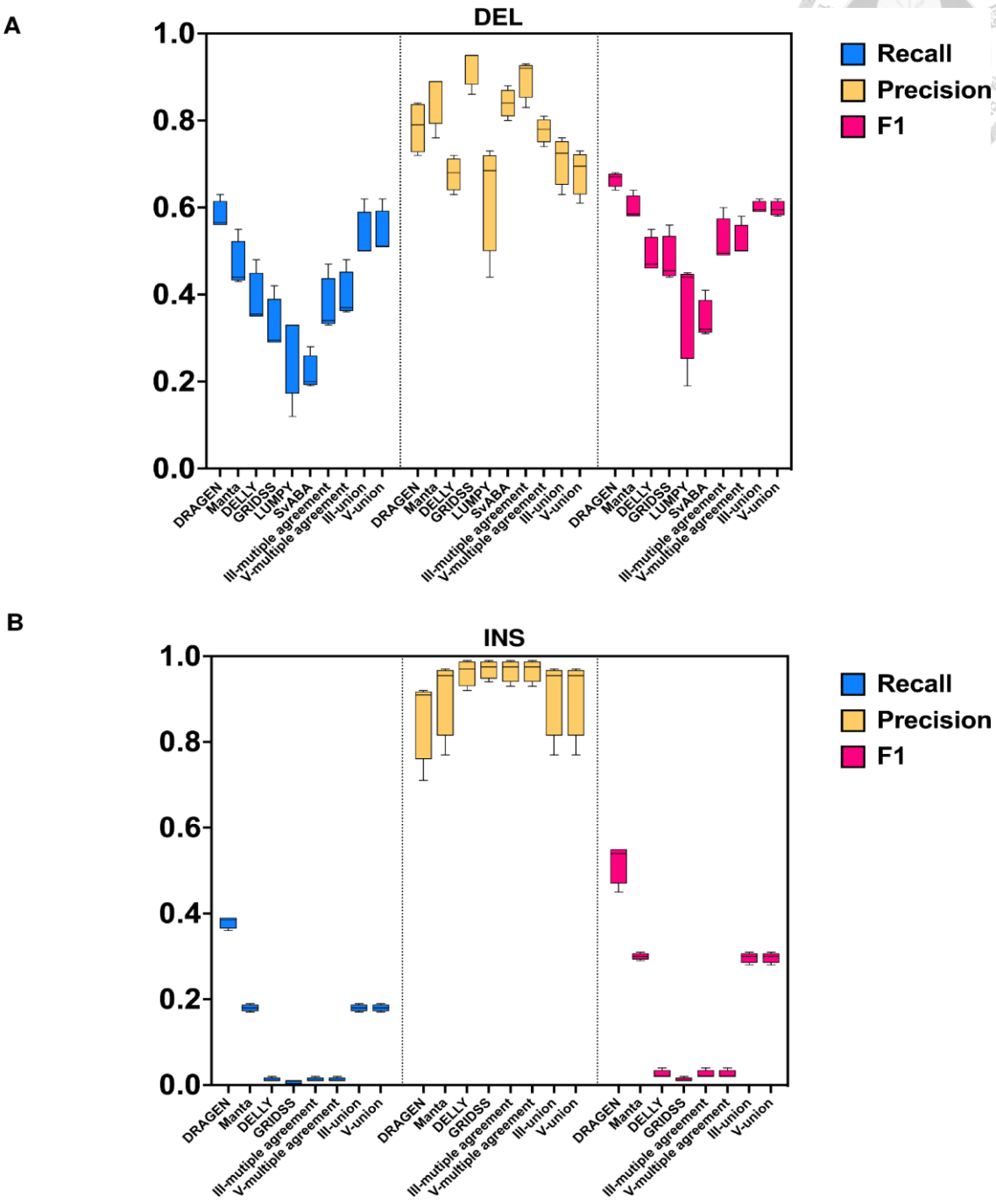



**Figure 18. Combination performance of each single caller and combination strategy in detecting DELs and INSs**. The maximum, minimum, and macro average of recall, precision, and F1 score are based on the integrated values from all sample sets.



doi:10.6342/NTU202500995

For DELs, the union strategy outperformed most individual tools—except for Manta—and also surpassed the multiple agreement combinations in terms of the F1 score, achieving values comparable to those of the commercial DRAGEN software (Figure 18 and Table 8). However, while the multiple agreement methods delivered higher precision than both the union strategy and DRAGEN, their recall rates were notably lacking.

For INSs, the performance was generally disappointing across all individual callers and combination methods, with F1 scores ranging from 0.01 to 0.30 versus 0.53 for DRAGEN. Although every single caller and combination approach registered higher precision compared to DRAGEN, their recall was insufficient. Among the non-commercial tools, Manta achieved the highest F1 score (0.30), ranking only behind DRAGEN, and within the combination strategies, the union approach produced F1 scores similar to those of Manta. Despite demonstrating moderate to high precision for INS detection, the overall results were undermined by the markedly low recall rates (Figure 18 and Table 8).

Overall, expanding the integration from three to five tools did not result in performance improvements, regardless of whether a multiple agreement or union strategy was applied. Nevertheless, employing combination strategies did enhance performance compared to using most single callers individually.

## 3.6 Discussion

Structural variants (SVs) can significantly influence phenotype by disrupting gene function, altering gene regulation, or modifying gene dosage. Their relevance in molecular biology and medicine has been increasingly recognized in recent studies[56], suggesting that SV screening might become a routine practice in clinical diagnostics.

Our study highlights the limitations of relying on a single SV detection algorithm, as



doi:10.6342/NTU202500995

this approach often restricts the detectable SV size and type while yielding low F1 scores. By contrast, the multiple agreement strategy—based on consensus among algorithms—demonstrates higher precision but at the expense of recall. A union-based approach improves recall and F1 scores for deletions (DELs) compared to single algorithms and the multiple agreement strategy. However, insertion (INS) detection remains challenging across most tools due to low recall rates. Specifically, DELLY and GRIDSS exhibited poor F1 scores (0.02 and 0.01, respectively), while Manta achieved the highest performance among non-commercial tools with an F1 score of 0.30. Incorporating Manta into a union strategy resulted in comparable INS detection accuracy, as shown in Figure 18. These findings underscore the need to select appropriate algorithms based on SV type and size for different clinical applications.

False positive (FP) SVs pose another concern, as they can be mistakenly linked to a phenotype. To address this, we examined FPs that were consistently detected by multiple callers in the HG002 dataset, comparing them with DRAGEN results. We categorized FPs into three groups: those identified solely by the multi-agreement strategy (M), those detected exclusively by DRAGEN (D), and those called by both methods (Both). Analysis of FP SV sizes, including median and interquartile ranges (Figure S3), revealed that most FPs were smaller than 1000 base pairs. Among deletion FPs, the M group had a median size of 326 bp, whereas the D group had a median of 120 bp. Due to low recall rates for insertions, the number of FP INS variants was limited, with a median SV length of 81 bp in the D group. Given that short-read sequencing generally covers fragments between 300 and 500 bp, these methods should theoretically be capable of accurately detecting small variants. Different SV callers apply unique detection principles, often incorporating hard filters—such as quality thresholds and read depth—to minimize FPs. However, it is crucial to balance filtering strategies to avoid discarding true positives. In clinical practice,



doi:10.6342/NTU202500995

additional validation steps, such as gene filtering using virtual panels or confirmation via the Integrative Genomics Viewer (IGV) and PCR, are commonly implemented. Considering these factors, prioritizing high recall to minimize false negatives (FN) enhances the clinical utility of SV detection.

Research has demonstrated that several factors affect FP rates in short-read sequencing-based SV detection. For instance, Cameron et al. evaluated the impact of sequence context and SV size on precision across 10 different callers using simulated data and well-characterized cell lines. Their findings identified key contributors to FP occurrence: 1) The presence of SNVs or indels near SV breakpoints increases FP rates. 2) SVs located in low-complexity regions, simple tandem repeats, or specific genomic elements like LTRs, LINEs, and SINEs exhibit reduced precision, resulting in higher FP counts. 3) SV size affects detection accuracy—variants under 100 bp are particularly challenging, whereas deletions between 300 and 500 bp tend to have higher precision.

The field of SV detection using short-read sequencing has evolved significantly. A recent advancement, PanSVR[57], enhances SV calling by integrating a pan-genome reference and read re-alignment, improving accuracy in VNTR and STR regions. Benchmarks indicate that PanSVR outperforms Manta and DELLY in detecting both DELs and INSs. Similarly, INSurVeyor[58], developed by Rajaby et al., specializes in INS detection, surpassing Manta in sensitivity. These findings emphasize the potential of combining well-performing tools, such as PanSVR and INSurVeyor, with union-based strategies to enhance SV detection.

Machine learning (ML)-based approaches have also gained traction in SV identification. Several classification-based tools, including SVM$^2$[59], forestSV[60], Wham[61], and svclassify[62] , utilize RP signal-derived features to refine SV predictions. Recent ML-driven frameworks have demonstrated notable improvements in performance.



doi:10.6342/NTU202500995

For example, MPRClassify[63] employs a multi-part read alignment strategy using three independent random forest classifiers tailored to specific SV types, achieving results comparable to DELLY, Manta, and Softsv[64]. Cue[65] transforms sequence alignments into image-based SV signatures, leveraging convolutional neural networks for classification. Evaluations show Cue surpasses Manta, DELLY, LUMPY, and SvABA, particularly in detecting various SV types such as deletions, duplications, and inversions. In another recent study, Guiwu Zhuang et al.[55] established an expert-reviewed tumor-specific clinically relevant SV call set, evaluating several algorithms for SV detection, and developed a random-forest-based decision model to improve the precision. Further exploration of ML-integrated union strategies could enhance SV detection accuracy, particularly in clinical applications.

Long-read sequencing (LRS) plays an essential role in SV detection, offering superior accuracy compared to short-read methods, especially for complex or long variants. For applications that require identifying very long or complex SVs, performing a localized long-read de novo assembly is more effective than conventional short-read mapping approaches. Recent benchmarking studies[66-68] reinforce LRS's advantages in sensitivity and specificity across multiple tools. Moreover, sequencing technologies from Oxford Nanopore and PacBio continue to improve in cost and accessibility.Existing LRS-based SV detection approaches primarily rely on alignment-based methods with heuristic filtering tailored to sequencing platform characteristics[13, 67, 69-72]. Examples include Svvalidator[72] and Kled[71] , which analyze CIGAR string patterns and split-read signatures to validate SVs. Kled further refines genotype clustering using a multi-threaded Omni Merging Algorithm. While manually designed heuristics may introduce noise, ML-based alternatives, such as SVDF[70], apply adaptive clustering techniques to filter artifacts and improve precision. Combining short-read and long-read



doi:10.6342/NTU202500995

sequencingmethodologies offers a promising path for developing robust SV detection strategies suited to diverse research and clinical needs.

Although our study primarily examined DEL and INS detection—representing the most abundant yet relatively simple SV types in the human genome—methods that fail to identify these variants may also struggle with more complex SV types. By assessing the effectiveness of different detection strategies, our findings provide valuable insights into optimizing SV calling performance and shaping future advancements in genomic research.

## 3.7 Conclusions

This study assessed the performance of five advanced structural variant (SV) detection algorithms alongside the commercial software DRAGEN, using short-read whole-genome sequencing data from the GIAB v0.6 Tier 1 benchmark set and the well-characterized HGSVC2 reference dataset. We evaluated each algorithm individually as well as in various combination strategies, applying both multiple agreement and union approaches. The union strategy demonstrated higher recall, while the multiple agreement method provided superior precision. Our findings highlight the distinct strengths and limitations of each algorithm, influencing the types and sizes of SVs they detect. By integrating their capabilities, union-based combination strategies effectively harmonized these differences, improving overall performance and achieving F1 scores comparable to those of DRAGEN.

## 3.8 Acknowledgments

We express our gratitude for the technical support provided by the Taiwan Computing Cloud (TWCC) at the Taiwan National Center for High-Performance



doi:10.6342/NTU202500995

Computing (NCHC). The GPT-3.5 architecture, developed by OpenAI, polished the grammar, punctuation, sentence structure, and overall coherence



doi:10.6342/NTU202500995

# 4. NTU WGS cohort

## 4.1 Background

Whole genome sequencing (WGS) has revolutionized the field of medical genetics by providing comprehensive insights into the genetic factors underlying various diseases, including thoracic aortic aneurysm and dissection (TAAD). TAAD is a life-threatening condition characterized by the enlargement and rupture of the aorta, and its early detection and management are critical. The genetic basis of TAAD is complex and involves multiple genes that contribute to the structural integrity and function of the aortic wall[11, 18-20].

Recent advancements in WGS technology have enabled researchers to identify genetic variants associated with TAAD with greater accuracy and depth. Unlike targeted sequencing approaches, such as whole exome sequencing (WES) or gene panels, WGS provides a comprehensive view of the entire genome, enabling the detection of a wide range of genetic variations, including single nucleotide variants (SNVs), structural variants (SVs), copy number variants (CNVs), and mitochondrial DNA variants[73, 74]. This breadth of coverage makes WGS particularly valuable for diagnosing rare genetic disorders, identifying actionable variants in cancer, and uncovering novel disease mechanisms[73, 75].

The diagnostic yield of WGS has been shown to surpass that of other molecular tests, especially in cases where traditional methods fail to provide conclusive results[74]. For example, in pediatric patients with suspected genetic disorders, WGS has demonstrated higher diagnostic accuracy compared to WES and conventional care[74]. Furthermore, the ability to archive WGS data for future reanalysis offers a lifelong resource for patients, enabling updates as new genetic discoveries are made[73].



doi:10.6342/NTU202500995

## 4.2 Aims

The aim of the WGS cohort of TAAD is to utilize whole genome sequencing to provide comprehensive insights into the genetic factors underlying thoracic aortic aneurysm and dissection. By sequencing the entire genome, we aim to detect various genetic variants, including single nucleotide variants (SNVs), insertions (INSs), deletions (DELs), structural variants (SVs) and mobile element insertions (MEIs) that may predispose individuals to TAAD. This approach surpasses traditional gene panel tests by covering the entire genome, enabling the discovery of novel and unexpected variants in genes not previously associated with TAAD. Ultimately, the goal is to enhance diagnostic accuracy, inform personalized treatment strategies, and facilitate the discovery of new therapeutic targets for this life-threatening condition.

## 4.3 Study design

The study design of the WGS cohort of TAAD incorporates several pivotal aspects to ensure comprehensive and detailed genetic analysis. The study involved the enrollment of probands diagnosed with thoracic aortic aneurysm and dissection, selected based on clinical symptoms and family history. Participants were chosen from the previous NTU TAAD cohort where no disease-causing variants were identified, but who presented with the phenotype of aortic aneurysm or dissection. Genomic DNA samples were extracted from whole blood using the Gentra Puregene Blood Kit, with quality checks performed for purity and integrity. The DNA underwent whole genome sequencing on a high-throughput NGS platform, generating comprehensive genomic data. Post-sequencing, rigorous bioinformatic analysis identified and annotated genetic variants, including SNVs, INSs, DELs, SVs, and MEIs, with advanced tools ensuring accuracy. Variants were evaluated for pathogenicity, with clinical correlation established by comparing genetic



doi:10.6342/NTU202500995

and clinical data. The study adhered to ethical guidelines, with informed consent and confidentiality maintained, and ethical approval obtained. Outcomes included the identification of TAAD-associated variants, improved diagnostic accuracy, and the potential for personalized treatment strategies, paving the way for future research into the genetic mechanisms of TAAD and novel therapeutic targets.

## 4.4 Materials and Methods

### 4.4.1 Enrollment of participants

Among the 125 probands who underwent screening with the NTU TAAD panel, which includes 29 genes associated with aortic aneurysm and/or dissection, 42 probands (33.6%) were identified to have disease-causing variants, with 22 classified as pathogenic and 20 as likely pathogenic. From the remaining 83 probands, 34 (40.9%) who exhibited symptoms of thoracic aortic aneurysm and/or dissection were selected as the study population. Within this group, 4 probands were found to possess variants of uncertain significance, specifically 3 in the *MYH11* gene and 1 in the *ELN* gene.

### 4.4.2 Establishment of virtual panels

The Human Phenotype Ontology (HPO)[76] currently encompasses over 13,000 terms with more than 156,000 annotations pertinent to hereditary diseases. We utilized the 2023-09-01 Release version sourced from the project homepage (http://www.human-phenotype-ontology.org). Our selection process involved identifying specific HPO terms related to a variety of aortic and arterial diseases.

This selection procedure was designed to create a targeted "focused virtual panel" for genetic analysis, focusing on the genetic determinants associated with aortic and arterial conditions[77, 78]. The selected Human Phenotype Ontology (HPO) terms relevant to these



doi:10.6342/NTU202500995

conditions include "Aortic root aneurysm," "Aortic aneurysm," "Ascending tubular aortic aneurysm," "Descending aortic aneurysm," "Aortic arch aneurysm," "Thoracic aortic aneurysm," "Aortic dissection," "Ascending aortic dissection," "Descending aortic dissection," "Arterial dissection," and "Coronary artery dissection." By utilizing these HPO terms, a total of 169 distinct genes were identified and selected.

For the "broad-spectrum virtual panel," additional Human Phenotype Ontology (HPO) terms were incorporated to complement the previously mentioned terms. These added terms encompass a broader spectrum of clinical features and conditions. The included HPO terms are "Kyphosis," "Cervical kyphosis," "Thoracolumbar scoliosis," "Thoracic scoliosis," "Thoracolumbar kyphoscoliosis," "Brachydactyly," "Arachnodactyly," "Hyperextensibility of the finger joints," "Interphalangeal joint contracture of the finger," "Slender finger," "Craniosynostosis," "Joint dislocation," "Joint hypermobility," "Tall stature," "Pes planus," "Abnormal rib cage morphology," "Abnormal hip joint morphology," "Abnormal joint morphology," "Intellectual disability," "Hypotonia," "Hyporeflexia," "Joint stiffness," "Umbilical hernia," "Ventricular septal defect," "Pulmonic stenosis," "Patent ductus arteriosus," "Bicuspid aortic valve," "Camptodactyly of the finger," "Dural ectasia," and "Long fingers." As a result, a total of 3,197 genes were selected for inclusion in this expanded virtual panel, reflecting a comprehensive range of genetic factors associated with the diverse clinical features and conditions identified by the included HPO terms[77, 78].

### 4.4.3 Library construction and whole genome sequencing

Genomic DNA was extracted and prepared as described for the panel test. Whole-genome libraries were meticulously prepared following the NovaSeq Library Preparation protocol, ensuring the highest standards of quality and accuracy. Sequencing was



doi:10.6342/NTU202500995

performed on an Illumina NovaSeq 6000 System using a paired-end 150 base pair read length configuration with S2 flow cells. The specific steps involved in the NovaSeq Library Preparation protocol included DNA fragmentation, end repair, A-tailing, adapter ligation, and PCR amplification. These steps are critical for generating high-quality libraries that ensure comprehensive genomic coverage. Each library was quantified using qPCR to ensure an optimal amount of DNA for sequencing, and quality was assessed using an Agilent Bioanalyzer. Sequencing was conducted by AllBio Corporation, a leading provider of genomic services, which adheres to stringent protocols to ensure the generation of high-quality data.

### 4.4.4 Computational resources

All bioinformatic analysis pipelines were executed on the Taiwania 3 computational platform, the most powerful CPU high-performance computing server in Taiwan, provided by the National Center for High-Performance Computing (NCHC). This advanced computational resource facilitated the efficient and accurate processing of large-scale genomic data. A Python script was meticulously developed to refine, filter, merge, and prioritize the variants identified through sequencing, enabling the intersection of candidate genes among probands.

### 4.4.5 Calling and annotation of SNVs and small indels

Single nucleotide variants (SNVs) and small insertions and deletions (indels) were identified using methodologies analogous to those employed in the panel test, albeit with updated tools and databases[22, 79]. Pathogenicity information was obtained from ClinVar[30], a comprehensive resource for variant classification, with reference to the archive version released on January 7, 2024, which provided the latest variant



doi:10.6342/NTU202500995

classifications. For variants located in noncoding regions, their impact on splicing elements was predicted using SpliceAI[29]. Additionally, to enhance the analysis of small indels, MutPred-Indel[80] was employed. This tool predicts the pathogenicity of indels by analyzing sequence and structural features, providing a pathogenicity score that quantifies each indel's potential impact on protein function.

### 4.4.6 Calling and annotation of structural variants

To detect larger structural variations (SVs) such as deletions, duplications, insertions, inversions, and translocations, the Illumina DRAGEN[48] pipeline was employed. Renowned for its accuracy and speed, the DRAGEN pipeline uses hardware acceleration for efficient genomic data analysis. Post-alignment, preprocessing steps include marking duplicate reads and base quality score recalibration (BQSR), enhancing the reliability of variant calls. The pipeline's algorithms precisely identify and characterize SVs, offering insights into genetic disorders. AnnotSV[81] then annotates these SVs with functional information, integrating data from multiple genomic databases to detail their impact on protein-coding regions, non-coding regions, and regulatory elements.

### 4.4.7 Calling and annotation of mobile element insertions

Mobile element insertions (MEIs) were detected using algorithms including MELT (Mobile Element Locator Tool)[82], SCRAMble[83], xTEA (x-Transposable element analyzer)[84], and Mobster[85]. MELT identifies MEIs by leveraging paired-end sequencing data to locate discordant read pairs and split reads, which indicate the presence of mobile elements. SCRAMble uses split-read alignments to directly detect mobile element insertion sites. xTEA utilizes a combination of split-read, discordant read-pair, and clipped alignment signatures to accurately identify transposable element



doi:10.6342/NTU202500995

insertions. Mobster employs a probabilistic framework to analyze split-read and discordant read-pair data to detect MEI events. These advanced tools enhance the accuracy and reliability of MEI detection. We retained a variant as a reliable variant only if it was called by at least two MEI tools. The MEIs undergo a meticulous annotation process using AnnotSV which integrates data from various genomic databases, providing detailed functional annotations for each variant. Following this, the variants located in exonic or intronic regions are annotated according to the GENCODE Version 19 gene annotations[86], ensuring comprehensive and precise mapping of genomic variants.

### 4.4.8 Variant prioritization pathway 1: by Clinvar database

Following the annotation of genetic variants, those with a minor allele frequency of less than 1% were selected for further analysis as potential disease-causing candidates. A systematic filtering process prioritized variants most likely to be biologically relevant or pathogenic. This involved key criteria such as variant type, predicted functional impact, and frequency in population databases, as well as mode of inheritance, database searches, and literature reviews. These variants were classified using a five-tier system according to the American College of Medical Genetics and Genomics (ACMG) guidelines from 2015[31]. Only variants labeled as pathogenic or likely pathogenic were reported as disease-causing. Confirming a genetic diagnosis required consistency between the gene's mode of inheritance—dominant, recessive, or X-linked—and the related Mendelian disease.

The SNVs and small indels undergo filtering through records available in the ClinVar database (Figure 19). ClinVar, a freely accessible, public archive of reports of relationships among human variations and phenotypes, provides essential clinical annotations and evidence for the clinical significance of these variants. Published article



doi:10.6342/NTU202500995

emphasize the importance of ClinVar as a repository that supports the interpretation of genetic variants by clinicians and researchers alike[30]. By leveraging ClinVar data, we prioritize variants with established clinical relevance, focusing on those classified as "pathogenic" or "likely pathogenic." This process ensures that only the most relevant and impactful genetic variations are considered for further analysis, enhancing the accuracy of our genetic assessments and the potential for identifying disease-causing variants.

### 4.4.9 Variant prioritization pathway 2: by prediction score

For single nucleotide variants (SNVs), those meeting at least two out of the following three criteria were considered candidate variants: a SIFT-pred score less than 0.05, a PolyPhen-2 score greater than 0.85, and a CADD-phred score[87] greater than 20. In the case of non-frameshifting insertions/deletions (indels), a score exceeding a threshold of 0.50, interpreted as a probability, was indicative of potential pathogenicity. Furthermore, variants with a SpliceAI score greater than 0.5 were also deemed candidate variants due to their predicted splice-altering effects (Figure 19).

### 4.4.10 Variant prioritization pathway 3: by ACMG classification and 4: by AnnotSV ranking score

Structural variants (SVs) and mobile element insertions (MEIs) were subjected to an additional layer of scrutiny. Variants meeting one of the following exclusion criteria were eliminated from consideration: an AnnotSV ranking score less than -0.89 or an ACMG score lower than 3. Variants located exclusively within introns were also excluded, as their impact on the protein-coding regions is typically minimal (Figure 20).



doi:10.6342/NTU202500995

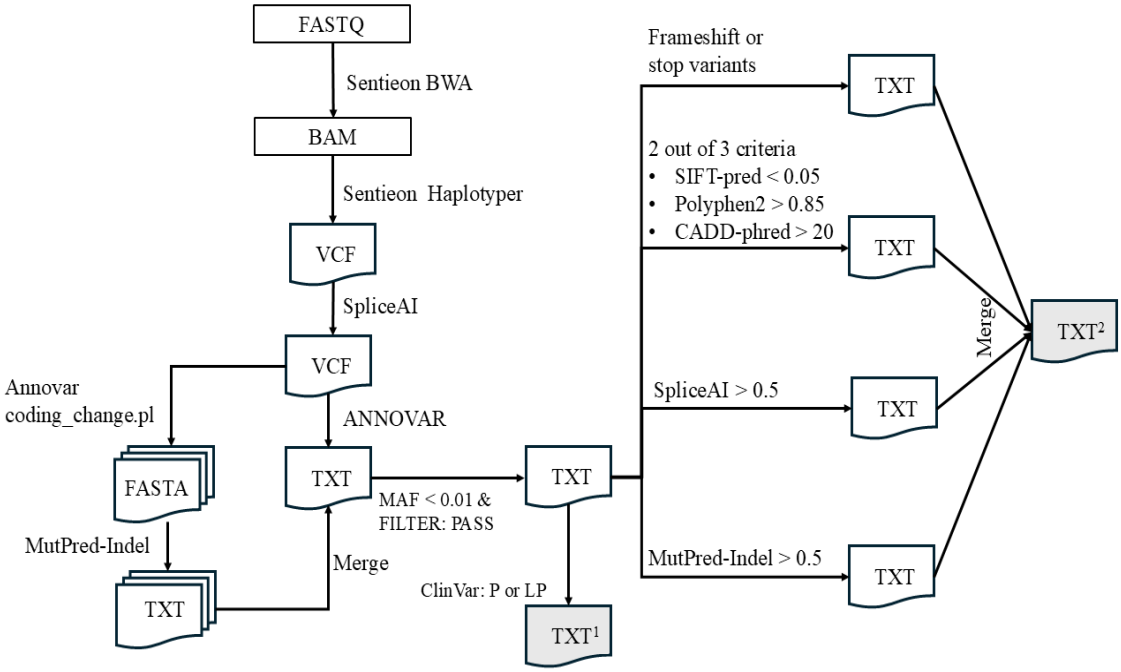



**Figure 19. Illustration of prioritization algorithm for SNVs and small indels**

[1] Prioritization pathway 1: by clinical significance level in Clinvar database

[2] Prioritization pathway 2: by prediction score

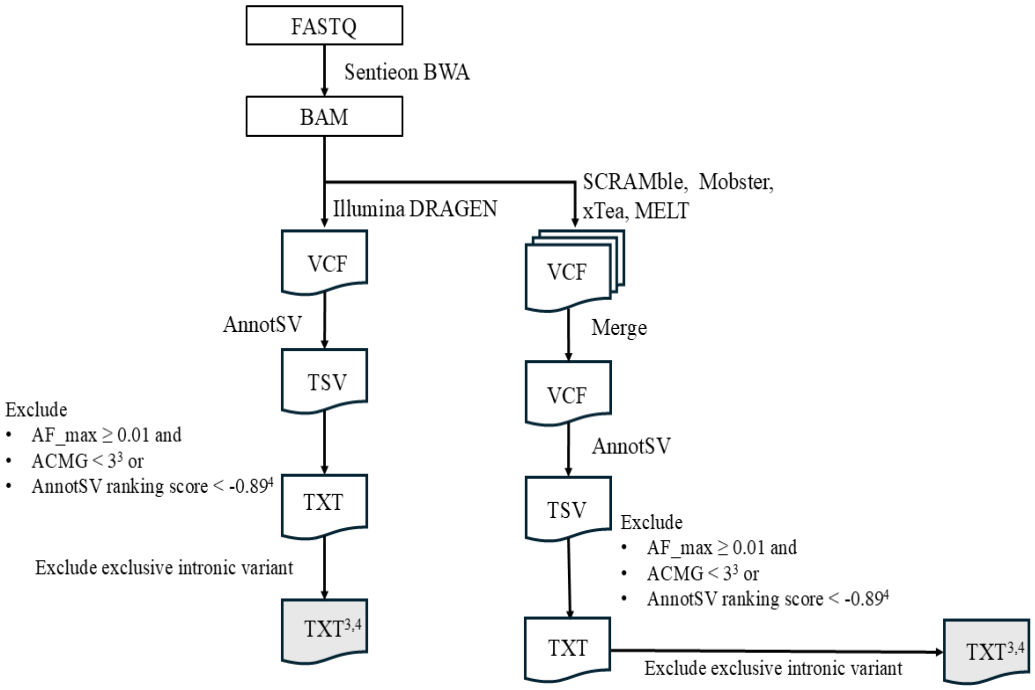



**Figure 20. Illustration of prioritization algorithm for SVs and MEIs**

[3] Prioritization pathway 3: by ACMG classification

[4] Prioritization pathway 4: by AnnotSV ranking score



doi:10.6342/NTU202500995

### 4.4.11 Classification and interpretation of variants

The variants were classified using a five-tier system according to the American College of Medical Genetics and Genomics (ACMG) guidelines from 2015[31]. Only variants labeled as pathogenic or likely pathogenic were reported as disease-causing. Confirming a genetic diagnosis required consistency between the gene's mode of inheritance—dominant, recessive, or X-linked—and the related Mendelian disease. The mode of inheritance of genes was referenced from the latest version updated on December 04, 2023, of the Clinical Genomic Database (https://research.nhgri.nih.gov/CGD/).

### 4.4.12 Identification of novel disease-causing gene through gene-level intersection analysis

The process of identifying de novo disease-causing genes involves comparing candidate genes among probands. This method aims to find commonality among genes from different individuals, particularly focusing on those genes that are not present in a larger population dataset. By intersecting the candidate genes carried by unrelated probands, we can pinpoint potential novel genes responsible for the disease. This approach involves a detailed bioinformatics analysis to ensure accuracy and reliability.

Our analysis posits that common novel disease-causing genes among unrelated probands can be identified by intersecting the genetic data of 24 probands, excluding those with known pathogenic variants. This method focuses on pinpointing overlapping genes and employs detailed bioinformatics analysis for accuracy. Identified genes are further validated through literature review and database comparisons (ClinVar, UMD, Varsome).



doi:10.6342/NTU202500995

### 4.4.13 Confirmation of identified variants

The confirmation methods for identified variants are detailed in Section 2.3.7.

## 4.5 Results

### 4.5.1 Baseline characteristics

The study cohort consisted of 34 unrelated individuals with a mean age of 41.21 years (±16.44). Among them, 26 (76.5%) were male, and 4 (14.8%) had a family history of related conditions. Cardiovascular manifestations were prevalent, with aortic aneurysms observed in 32 individuals (94.1%) and aortic dissection in 7 individuals (20.6%). Musculoskeletal involvement was also notable, with 14 individuals (41.2%) exhibiting kyphosis or scoliosis. Ocular abnormalities, including ectopia lentis or myopia, were identified in 19 individuals (59.3%). The systemic score varied across participants, with 6 individuals (17.6%) presenting a score of 7 or higher, indicating more severe systemic manifestations. The mean systemic score was 3.53 (±4.33), with an interquartile range of 1.50 (0–5.75), reflecting variability in disease expression within the cohort. (Table 9 and Table S10).



doi:10.6342/NTU202500995

**Table 9. Baseline characteristics of unrelated probands**

| Total population (34 unrelated individuals) | |
|---|---|
| Age (mean ± SD), yrs | 41.21 ± 16.44 |
| Male | 26 (76.5%) |
| Family history | 4 (14.8%) |
| Aortic aneurysm | 32 (94.1%) |
| Aortic dissection | 7 (20.6%) |
| Ectopia lentis or Myopia | 19 (59.3%) |
| Kyphosis or Scoliosis | 14 (41.2%) |
| Systemic score ≥ 7 | 6 (17.6%) |
| Systemic score (mean ± SD) | 3.53 ± 4.33 |
| Systemic score (IQR) | (0 – 5.75) |

### 4.5.2 Participant selection and genetic variants prioritization

The study involved 125 probands, among whom 42 received a genetic diagnosis using the NTUH TAAD panel (29 genes). The remaining 83 probands underwent further analysis. After excluding 49 individuals who did not present with aortic aneurysm or dissection, 34 probands were selected for whole genome sequencing and genetic variant prioritization (Figure 21 and Table S10).

Whole genome sequencing revealed approximately 4.8 million single nucleotide variants (SNVs) and small indels, 9,000 structural variants (SVs), and 1,500 mobile element insertions (MEIs) per individual (Table S11). These variants were prioritized using multiple filtration pathways, yielding around 250 candidate variants per proband. To refine variant selection, filtration was performed using both focused and broad-spectrum virtual panels. The focused virtual panel (169 genes) identified five probands with reasonably supported disease-causing variants in *FBN1*, including SNVs, splice site variants, and SVs (Table 10). The broad-spectrum virtual panel (3,197 genes) yielded



doi:10.6342/NTU202500995

approximately 30 supported variants per proband (Table S12). For the remaining 29 probands, gene-level intersection analysis was employed to identify novel disease-associated genes, excluding the five probands with previously supported variants. The distribution of variants and the algorithms used for disease-causing variant detection and novel gene identification in the cohort are illustrated in Figure 21. This approach highlights the effectiveness of integrating virtual panel filtration with intersection analysis in rare disease research, enabling the discovery of novel genetic contributors to disease pathogenesis.

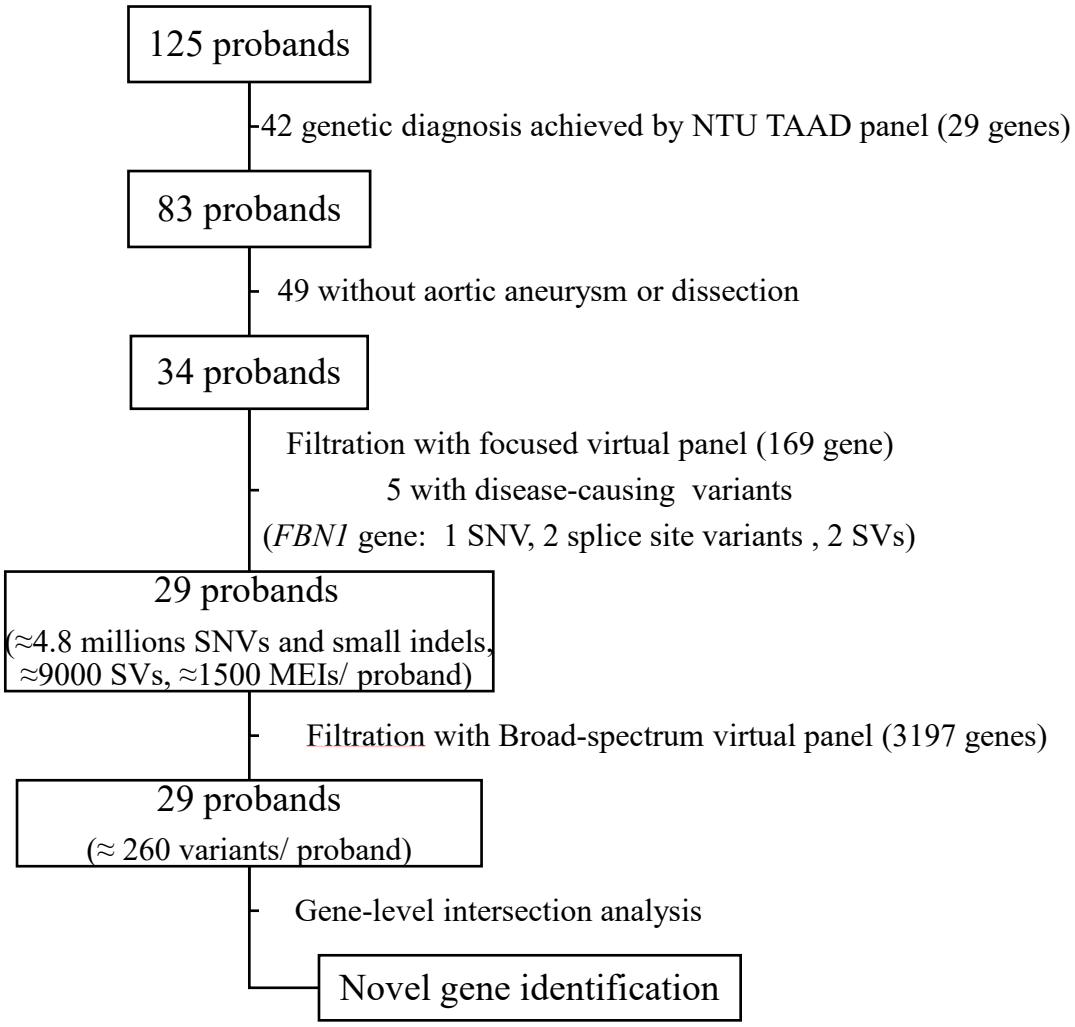



**Figure 21. Participant selection and genetic variants of the cohort**



doi:10.6342/NTU202500995

### 4.5.3 Supported variants filtered through focused virtual panel

Among the 34 probands analyzed by whole genome sequencing and variant prioritization, our integrated filtering strategy—which combined four prioritization pathways (clinical significance by ClinVar, variant impact prediction scores, ACMG classification criteria, and AnnotSV ranking scores)—uncovered a series of variants that had been missed by the former NTU TAAD gene panel. In this context, the reasons listed in the table represent why certain variants were not detected during the initial gene panel testing. For instance, variants in genes not included in the NTU TAAD panel (reason 1), intronic variants (reason 2), structural variants (reason 3), and synonymous variants (reason 4) were typically overlooked in the panel design, as in Table 10.

In the subset of reasonably supported variants, five probands exhibited alterations in *FBN1*. For instance, proband TWGS029 carried a synonymous variant (c.6354C>T, p.Ile2118Ile; rs112989722) with an allele frequency of 0.0006. Despite being synonymous, which is typically filtered out in panel testing, this variant met multiple ACMG criteria (PVS1, PM2, PP4, PP5) and was retrospectively designated with reason 4—indicating that its silent nature contributed to its non-detection in the original panel. Proband TWGS010 carried a novel intronic variant (c.5917+5G>A) and TWGS011 harbored another novel intronic change (c.1837+5G>T); both were heterozygous and exhibited notable SpliceAI scores of 0.74 and 0.87, respectively. These intronic variants were not detected in the previous gene panel tests because such tests typically focused on coding regions and canonical splice sites, leaving intronic alterations largely unexamined. Furthermore, the prediction tools available during that era were limited in their ability to assess the functional impact of intronic changes, contributing to the oversight of these variants in the initial screening process. Moreover, structural deletions were identified in TWGS018, which had an 8,243 bp deletion confined to *FBN1*, and in TWGS006, where



doi:10.6342/NTU202500995

a similar deletion spanned both *FBN1* and *CEP152*, underscoring the limitations of targeted sequencing in capturing large structural variants (reason 3).

In addition to these *FBN1* findings, nine probands exhibited marginally supported variants in other genes, including *PKD1*, *IFH1*, *TPM2*, *SMAD2*, *RIN2*, *NEDD4L*, *TGFBR1*, *BAZ1B*, *TGFBR2*, *DNMT3A*, and *ERMARD*. For example, TWGS0002 showed a nonsense variant in *PKD1* (c.3520C>T, p.Gln1174Ter; rs1057518899) and TWGS025 an intronic splice site variant in *IFH1* (c.2807+1G>A, SpliceAI score 0.98); these variants were often flagged with combinations such as reason 1 and 2, indicating that either the gene was not included in the NTU TAAD panel or the variant type (e.g., intronic) rendered it undetectable by the panel. Similarly, a variant in *TPM2* (c.773-14T>A) was assigned reason 1 + 2, while structural and splice-altering events in *SMAD2*, *RIN2*, *NEDD4L*, and *TGFBR1* were designated reasons 1, 3, or combinations thereof. Variants in *BAZ1B*, *TGFBR2*, *DNMT3A*, and *ERMARD* were generally prioritized via pathway 2 and flagged with reason 1, underscoring that their detection was beyond the scope of the original gene panel test.

Collectively, these findings not only illustrate the genetic heterogeneity underlying TAAD, but also underscore the limitations inherent in targeted gene panel testing. The reasons for failed detection provided in the table—specifically, that variants in genes absent from the panel (reason 1), intronic variants (reason 2), structural variants (reason 3), and synonymous variants (reason 4)—highlight the enhanced sensitivity of whole genome sequencing coupled with an integrated multi-pathway prioritization strategy in uncovering a broader spectrum of disease-associated variants.



doi:10.6342/NTU202500995

**Table 10. Supported variants filtered through focused virtual panel**

| Proband | Gene | Transcript no. | c.DNA | p.Aminoacid | rs no. | AF | Zyogsity | ACMG | Clinvar | SpliceAI | Prioritization | Reasons |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Reasonably supported variants | | | | | | | | | | | | |
| TWGS029 | *FBN1* | NM_000138 | c.6354C>T | pIle2118Ile | rs112989722 | 0.0006 | Hetero | P (PVS1, PM2, PP4, PP5) | P | - | 1 | 4 |
| TWGS010 | *FBN1* | NM_000138 | c.5917+5G>A | - | novel | NIL | Hetero | LP (PM1, PM2, PP3, PP4) | NIL | 0.74 | 2 | 2 |
| TWGS011 | *FBN1* | NM_000138 | c.1837+5G>T | - | novel | NIL | Hetero | LP (PM1, PM2, PP3, PP4) | NIL | 0.87 | 2 | 2 |
| TWGS018 | *FBN1* | NM_000138 | SVTYPE=DEL; SVLEN=-8243 | - | novel | - | Hetero | - | NIL | - | 3 + 4 | 3 |
| TWGS006 | *FBN1, CEP152* | - | SVTYPE=DEL; SVLEN=-8243 | - | novel | - | Hetero | - | NIL | - | 3 + 4 | 3 |
| Marginally supported variants | | | | | | | | | | | | |
| TWGS002 | *PKD1* | NM_000296 | c.3520C>T | p.Gln1174Ter | rs1057518899 | NIL | Hetero | P (PVS1, PM2, PP4, PP5) | NIL | - | 1 + 2 | 1 |
| TWGS025 | *IFH1* | NM_022168 | c.2807+1G>A | - | rs35732034 | 0.0006 | Hetero | P (PVS1, PM2, PP4, PP5) | NIL | 0.98 | 1 + 2 | 1 |
| TWGS003 | *TPM2* | NM_003289 | c.773-14T>A | - | novel | NIL | Hetero | VUS (PM2, PP4) | NIL | 0.75 | 2 | 1 + 2 |
| TWGS007 | *SMAD2* | NM_001135937 | c.1174_1177delCCCA | p.Trp392Glufs*9 | novel | NIL | Hetero | P (PVS1, PM2, PP4) | NIL | - | 2 | 1 |
| TWGS012 | *RIN2* | NM_018993 | c.916_917insGC | p.Ser306Cysfs*42 | novel | NIL | Hetero | P (PVS1, PM2, PP4) | NIL | - | 2 | 1 |
| TWGS017 | *NEDD4L* | - | SVTYPE=DEL; SVLEN=-13507 | - | novel | - | - | - | NIL | - | 3 + 4 | 1 + 3 |



doi:10.6342/NTU202500995

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TWGS024 | *TGFBR1* | NM_004612 | c.704_705ins | p.Leu236Argfs*15 | novel | NIL | Hetero | P (PVS1, PM2, PP4) | NIL | - | 2 | 3 |
| | *PKD1* | NM_000296 | c.1849+6_c.1849+10ACTCT>- | - | novel | NIL | Hetero | VUS (PM2, PP4) | NIL | 0.74 | 2 | 1 + 2 |
| TWGS026 | *BAZ1B* | NM_032408 | c.3072-1116A>G | - | novel | NIL | Hetero | VUS (PM2, PP4) | NIL | 0.69 | 2 | 1 + 2 |
| TWGS027 | *TGFBR2* | NM_001024847 | c.94+167T>G | - | novel | 0.0013 | Hetero | VUS (PM2, PP4) | NIL | 0.56 | 2 | 2 |
| TWGS032 | *DNMT3A* | NM_001320893 | c.534G>A | p.Trp178Ter | novel | NIL | Hetero | VUS (PM2, PP4) | NIL | - | 2 | 1 |
| TWGS034 | *ERMARD* | NM_001278532 | c.925C>T | p.Gln309Ter | novel | NIL | Hetero | VUS (PM2, PP4) | NIL | - | 2 | 1 |

Prioritization pathway 1: by clinical significance level in Clinvar database; Prioritization pathway 2: by prediction score; Prioritization pathway 3: by ACMG classification; Prioritization pathway 4: by AnnotSV ranking score

reason 1: genes not involved in NTU TAAD panel; reason 2: intronic variant; reason 3: structural variant; reason 4: synonymous variant



doi:10.6342/NTU202500995

### 4.5.4 Recurrent gene identified via gene-level intersection analysis within a broad-spectrum panel

The supported variants filtered through the broad-spectrum virtual panel were pooled, resulting in a total of 880 variants (data not shown). To refine the analysis, these pooled variants underwent gene-level intersection among 29 probands, which allowed for the identification of recurrent genes within the cohort. In total, 67 recurrent genes were identified in at least two probands, 21 recurrent genes in at least three probands, 8 recurrent genes in at least four probands, and 4 recurrent genes were consistently detected in five probands (Table 11 and data not shown). This recurrent gene set represents potential candidates for further investigation, as their presence across multiple individuals suggests possible biological or clinical relevance. For instance, *ABCA4* was detected in TWGS001, TWGS022, and TWGS030, and *COL6A3* recurred in TWGS002, TWGS017, and TWGS028. Additional genes such as *USF3*, *CUX1, MACF1*, *DCHS1, LRP5, ASXL3*, *FLNB*, *SREBF1*, *VWF*, *FGFRL1*, *FSCN2*, *TGM6*, *JMJD1C*, *RYR1*, *RREB1*, *PCLO*, *DYNC2H1*, *PIEZO1*, and *GJC2* were each identified in three probands. Notably, *SETBP1* and *PKD1* were recurrent across four probands—with *SETBP1* found in TWGS001, TWGS004, TWGS020, and TWGS030, and *PKD1* in TWGS002, TWGS004, TWGS022, and TWGS024—and *ADGRV1* and *CLCN7* also appeared in four probands. A combined



doi:10.6342/NTU202500995

genetic cluster encompassing *GJA8*, *PEX11B*, *GJA5*, *HJV*, *POLR3GL*, *NOTCH2NLC*, and *RBM8A* was identified in TWGS009, TWGS023, TWGS025, and TWGS030, while *CAMSAP1*, *RELN*, and *PCDH15* recurred in four or more probands. Collectively, these recurrent findings, summarized in Table 11, not only illustrate the genetic heterogeneity underlying TAAD but also highlight novel candidate genes that warrant further functional validation to elucidate their pathogenic significance.



doi:10.6342/NTU202500995

**Table 11. Recurrent gene identified via gene-level intersection analysis within broad spectrum panel**

| Recurrent genes | Probands |
|---|---|
| *ABCA4* | TWGS001, TWGS022, TWGS030 |
| *COL6A3* | TWGS002, TWGS017, TWGS028 |
| *USF3* | TWGS002, TWGS023, TWGS025 |
| *CUX1* | TWGS002, TWGS017, TWGS032 |
| *MACF1* | TWGS003, TWGS005, TWGS013 |
| *DCHS1* | TWGS003, TWGS019, TWGS022 |
| *LRP5* | TWGS003, TWGS025, TWGS034 |
| *ASXL3* | TWGS003, TWGS019, TWGS024 |
| *FLNB* | TWGS004, TWGS013, TWGS014 |
| *SREBF1* | TWGS005, TWGS031, TWGS032 |
| *VWF* | TWGS008, TWGS022, TWGS033 |
| *FGFRL1* | TWGS008, TWGS015, TWGS020 |
| *FSCN2* | TWGS013, TWGS024, TWGS025 |
| *TGM6* | TWGS014, TWGS028, TWGS034 |
| *JMJD1C* | TWGS017, TWGS024, TWGS034 |
| *RYR1* | TWGS017, TWGS025, TWGS030 |
| *RREB1* | TWGS019, TWGS022, TWGS025 |
| *PCLO* | TWGS019, TWGS030, TWGS031 |
| *DYNC2H1* | TWGS020, TWGS022, TWGS032 |
| *PIEZO1* | TWGS021, TWGS022, TWGS034 |
| *GJC2* | TWGS023, TWGS025, TWGS031 |
| *SETBP1* | TWGS001, TWGS004, TWGS020, TWGS030 |
| *PKD1* | TWGS002, TWGS004, TWGS022, TWGS024 |
| *ADGRV1* | TWGS002, TWGS017, TWGS023, TWGS024 |
| *CLCN7* | TWGS003, TWGS019, TWGS023, TWGS027 |
| *CAMSAP1* | TWGS003, TWGS007, TWGS008, TWGS033 |
| *RELN* | TWGS005, TWGS008, TWGS026, TWGS034 |
| *PCDH15* | TWGS007, TWGS012, TWGS022, TWGS026 |
| *GJA8;PEX11B;GJA5;HJV;POLR3GL;NOTCH2NLC; RBM8A* | TWGS009, TWGS023, TWGS025, TWGS030 |



doi:10.6342/NTU202500995

## 4.6 Discussion

### 4.6.1 Maximizing variant detection through whole genome sequencing

Our cohort analysis utilizing whole genome sequencing (WGS) revealed a higher number of supported variants compared to gene panel testing, underscoring the enhanced sensitivity of WGS in variant detection. Unlike targeted gene panels, which are constrained by predefined selections, WGS provides a comprehensive assessment of coding and non-coding regions, enabling the identification of rare and structural variants that might otherwise be overlooked. This increased detection rate is particularly valuable in precision medicine, as it facilitates more informed variant interpretation and potentially uncovers novel disease-associated genes. Integrating findings with reference databases such as ClinVar and Taiwan Biobank further strengthens the clinical relevance of detected variants, allowing for more accurate pathogenicity assessments. The integration of deep phenotyping and comprehensive variant data—showcasing automated tools that improved interpretation in large-scale genomic initiatives such as the 1,000 Genomes Project—reinforces the importance of refined filtration strategies in uncovering causative variants, thereby aligning with the methodological framework of this study[88]. Our results suggest that WGS is a powerful tool in improving diagnostic yield, particularly for



doi:10.6342/NTU202500995

conditions where gene panel testing has previously yielded inconclusive results.

### 4.6.2 Different filtration pathways prevent missed detection of variants

In our cohort, implementing multiple filtration pathways significantly improved the detection of germline variants beyond the limitations of gene panel testing. During the initial filtration process, synonymous variants were excluded, leading to their omission in gene panel testing. However, when cross-referencing with the ClinVar database, we identified a synonymous variant that had been previously filtered out, demonstrating the value of multi-tiered variant evaluation. This finding underscores the importance of integrating curated variant databases into filtration workflows to minimize missed detections. While standard filtration criteria help refine the dataset, incorporating clinically relevant annotations ensures that variants with potential significance are retained for further assessment. Our approach highlights the necessity of flexible filtration strategies that adapt to evolving variant classification methodologies, ultimately enhancing genomic analysis accuracy and clinical interpretation. By integrating ClinVar annotations, prediction scores, AnnotSV ranking scores, and ACMG criteria, we systematically refined variant selection, uncovering clinically relevant variants that were



doi:10.6342/NTU202500995

previously missed. This approach not only identified overlooked variants but also revealed structural variations that targeted sequencing methods often fail to detect. Recent study highlights the need for robust computational pipelines, aligning with our approach in maximizing sensitivity and confidence in variant identification[89]. By integrating different filtration strategies, we successfully identified novel candidate genes and variants, expanding potential genetic contributors to disease pathology. Our findings demonstrate that a diversified filtration framework enhances sensitivity and improves genomic insights, ultimately strengthening the precision of molecular diagnosis and advancing the field of genetic medicine.

### 4.6.3 Balancing sensitivity and specificity in broad-spectrum virtual panel

In our cohort, we designed two types of virtual gene panels—focused and broad-spectrum—to assess their effectiveness in variant detection. The broad-spectrum panels demonstrated a higher detection rate, identifying a greater number of variants compared to the focused panels. However, this increased sensitivity came at the cost of reduced specificity in phenotype correlation, highlighting the trade-off between comprehensive variant identification and clinical relevance.



doi:10.6342/NTU202500995

The study supports the advantages of virtual panels over static gene lists, emphasizing the refinement of virtual panels to improve diagnostic accuracy, particularly in specialized disease contexts and their flexibility in adapting to novel gene-disease associations[90].

Our findings align with these studies, demonstrating that while broad-spectrum panels maximize variant detection, focused panels provide greater specificity in phenotype correlation. This underscores the need for a balanced approach in virtual panel design, ensuring that diagnostic sensitivity does not compromise clinical interpretability. Future refinements should aim to integrate adaptive filtering strategies that optimize both variant yield and phenotype relevance, ultimately enhancing the precision of genomic diagnostics.

### 4.6.4 Intersection analysis between probands give an opportunity for detection of novel gene

In our study, we performed gene-level intersection analysis on a cohort of 29 probands using a broad-spectrum panel, which initially yielded 880 variants. These variants were then systematically filtered to identify recurrent genes, revealing 66 genes in at least two probands, 21 genes in at least three, 8 genes in at least four, and 4 genes in



doi:10.6342/NTU202500995

at least five probands (Table 11). This systematic filtration and intersection approach not only enabled us to uncover a wide array of recurrent genes but also highlighted a notable pattern: recurrent genes tended to be more commonly shared among smaller subsets of probands, while only a limited number of genes recurred across larger groups. This pattern suggests that rare genetic contributors may exhibit higher specificity within smaller patient subsets, whereas genes that recur broadly across multiple individuals might represent more established disease-associated candidates.

For instance, several genes such as *ABCA4*, *COL6A3*, *USF3*, and *CUX1* were observed across multiple probands, supporting their potential contributory role in TAAD pathogenesis. Even more striking was the recurrent detection of genes like *SETBP1* and *PKD1* in four probands, and a genetic cluster comprising *GJA8*, *PEX11B*, *GJA5*, *HJV*, *POLR3GL*, *NOTCH2NLC*, and *RBM8A* in multiple cases, thereby suggesting a possible shared pathogenic mechanism that may have been overlooked using traditional gene panel approaches.

Our findings are in line with previous work, which demonstrated that network-based prioritization can greatly enhance the identification of novel disease genes[91]. Moreover, research on recurrent genomic disorder deletions has underscored the significance of



doi:10.6342/NTU202500995

segmental haploid genetics in characterizing genes for autosomal recessive diseases, further reinforcing the importance of recurrent gene analysis in rare disease research. Collectively, the recurrent gene set presented in Table 11 not only emphasizes the genetic heterogeneity underlying TAAD but also provides a compelling foundation for further functional validation, which may enhance our understanding of the molecular mechanisms involved and improve diagnostic strategies for this complex condition.

## 4.7 Conclusions

Whole genome sequencing (WGS) enhances variant detection beyond the limitations of gene panel testing by providing a more comprehensive assessment of genetic alterations, including rare and structural variants. The use of multiple filtration pathways—such as ClinVar annotations, prediction scores, AnnotSV ranking scores, and ACMG criteria—refines variant selection and minimizes missed detections. Broad-spectrum virtual panels detect a higher number of variants but at the expense of phenotype specificity, whereas focused panels maintain greater clinical relevance. Additionally, intersecting supported genes among probands reveals recurrent gene patterns, aiding in novel gene discovery. Genes shared by fewer individuals tend to be more specific to certain phenotypes, while recurrent genes found in multiple probands may represent well-



doi:10.6342/NTU202500995

characterized disease associations. These findings underscore the importance of balanced filtration strategies, tailored panel designs, and intersection analysis in genomic diagnostics, ultimately improving sensitivity, specificity, and the discovery of novel disease-related genes.



doi:10.6342/NTU202500995

# 5. References


1. Writing Committee M, Isselbacher EM, Preventza O, Hamilton Black Iii J, Augoustides JG, Beck AW, et al. 2022 ACC/AHA Guideline for the Diagnosis and Management of Aortic Disease: A Report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol. 2022;80(24):e223-e393.
2. Roman MJ, Devereux RB, Kramer-Fox R, O'Loughlin J. Two-dimensional echocardiographic aortic root dimensions in normal children and adults. The American journal of cardiology. 1989;64(8):507-12.
3. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr. 2015;28(1):1-39 e14.
4. Hellmich B, Agueda A, Monti S, Buttgereit F, de Boysson H, Brouwer E, et al. 2018 Update of the EULAR recommendations for the management of large vessel vasculitis. Ann Rheum Dis. 2020;79(1):19-30.
5. van Kimmenade RR, Kempers M, de Boer MJ, Loeys BL, Timmermans J. A clinical appraisal of different Z-score equations for aortic root assessment in the diagnostic evaluation of Marfan syndrome. Genetics in medicine : official journal of the American College of Medical Genetics. 2013;15(7):528-32.
6. Czerny M, Schmidli J, Adler S, van den Berg JC, Bertoglio L, Carrel T, et al. Current options and recommendations for the treatment of thoracic aortic pathologies involving the aortic arch: an expert consensus document of the European Association for Cardio-Thoracic surgery (EACTS) and the European Society for Vascular Surgery (ESVS). Eur J Cardiothorac Surg. 2019;55(1):133-62.
7. Tsai TT, Nienaber CA, Eagle KA. Acute aortic syndromes. Circulation. 2005;112(24):3802-13.
8. Loeys BL, Dietz HC, Braverman AC, Callewaert BL, De Backer J, Devereux RB, et al. The revised Ghent nosology for the Marfan syndrome. Journal of Medical Genetics. 2010;47(7):476-85.
9. Amemiya K, Ishibashi-Ueda H, Mousseaux E, Achouh P, Ochiai M, Bruneval P. Comparison of the damage to aorta wall in aortitis versus noninflammatory degenerative aortic diseases. Cardiovasc Pathol. 2021;52:107329.




doi:10.6342/NTU202500995

10. Vapnik JS, Kim JB, Isselbacher EM, Ghoshhajra BB, Cheng Y, Sundt TM, 3rd, et al. Characteristics and Outcomes of Ascending Versus Descending Thoracic Aortic Aneurysms. The American journal of cardiology. 2016;117(10):1683-90.
11. Pinard A, Jones GT, Milewicz DM. Genetics of Thoracic and Abdominal Aortic Diseases. Circ Res. 2019;124(4):588-606.
12. Metzker ML. Sequencing technologies — the next generation. Nature Reviews Genetics. 2010;11(1):31-46.
13. Mahmoud M, Gobet N, Cruz-Davalos DI, Mounier N, Dessimoz C, Sedlazeck FJ. Structural variant calling: the long and the short of it. Genome Biol. 2019;20(1):246.
14. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nature Reviews Genetics. 2009;10(10):691-703.
15. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: Promises and challenges. Cancer Letters. 2013;340(2):284-95.
16. Chiu H-H, Wu M-H, Chen H-C, Kao F-Y, Huang S-K. Epidemiological Profile of Marfan Syndrome in a General Population: A National Database Study. Mayo Clinic Proceedings. 2014;89(1):34-42.
17. Verstraeten A, Luyckx I, Loeys B. Aetiology and management of hereditary aortopathy. Nature reviews Cardiology. 2017;14(4):197-208.
18. Goldfinger JZ, Halperin JL, Marin ML, Stewart AS, Eagle KA, Fuster V. Thoracic aortic aneurysm and dissection. J Am Coll Cardiol. 2014;64(16):1725-39.
19. Isselbacher EM, Lino Cardenas CL, Lindsay ME. Hereditary Influence in Thoracic Aortic Aneurysm and Dissection. Circulation. 2016;133(24):2516-28.
20. Brownstein AJ, Ziganshin BA, Kuivaniemi H, Body SC, Bale AE, Elefteriades JA. Genes Associated with Thoracic Aortic Aneurysm and Dissection: An Update and Clinical Implications. Aorta (Stamford, Conn). 2017;5(1):11-20.
21. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42(2):377-81.
22. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11 0 1- 0 33.
23. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res.



doi:10.6342/NTU202500995

2010;38(16):e164.
24. adzhubei Ia. A method and server for predicting damaging missense mutations. 2010;7(4)
25. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Research. 2003;31(37):3812-3814.
26. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. Nature Methods. 2014;11(4):361-2.
27. Desmet FO, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67.
28. GENE YEO, CHRISTOPHER B. BURGE. MaxEntscan: Maximum Entropy Modeling of Short Sequence Motifs with Applications to RNA Splicing Signals. JOURNAL OF COMPUTATIONAL BIOLOGY. 2004;11(2-3):377-394.
29. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, et al. SpliceAI: Predicting Splicing from Primary Sequence with Deep Learning. Cell. 2019;176(3):535-48 e24.
30. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-d7.
31. Richards S. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. 2015, 17 (5): 405-423.
32. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature Biotechnology. 2011;29(1):24-6.
33. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013;14(2):178-92.
34. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences. 1977;74(12):5463-7.
35. Al-Himdani S, Ud-Din S, Gilmore S, Bayat A. Striae distensae: a comprehensive review and evidence-based evaluation of prophylaxis and treatment. British Journal of Dermatology. 2014;170(3):527-47.
36. Ledoux M, Beauchet A, Fermanian C, Boileau C, Jondeau G, Saiag P. A case-control study of cutaneous signs in adult patients with Marfan disease: Diagnostic value of striae. Journal of the American Academy of Dermatology. 2011;64(2):290-5.



doi:10.6342/NTU202500995

37. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genetics in Medicine. 2017;19(2):249-55.
38. Wolford BN, Hornsby WE, Guo D, Zhou W, Lin M, Farhat L, et al. Clinical Implications of Identifying Pathogenic Variants in Individuals With Thoracic Aortic Dissection. Circ Genom Precis Med. 2019;12(6):e002476.
39. Weerakkody R, Ross D, Parry DA, Ziganshin B, Vandrovcova J, Gampawar P, et al. Targeted genetic analysis in a large cohort of familial and sporadic cases of aneurysm or dissection of the thoracic aorta. Genetics in medicine : official journal of the American College of Medical Genetics. 2018;20(11):1414-22.
40. Yang H, Luo M, Fu Y, Cao Y, Yin K, Li W, et al. Genetic testing of 248 Chinese aortopathy patients using a panel assay. Sci Rep. 2016;6:33002.
41. Hung CC, Lin SY, Lee CN, Cheng HY, Lin SP, Chen MR, et al. Mutation spectrum of the fibrillin-1 (FBN1) gene in Taiwanese patients with Marfan syndrome. Ann Hum Genet. 2009;73(Pt 6):559-67.
42. Guo J, Cai L, Jia L, Li X, Xi X, Zheng S, et al. Wide mutation spectrum and frequent variant Ala27Thr of FBN1 identified in a large cohort of Chinese patients with sporadic TAAD. Sci Rep. 2015;5:13115.
43. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. Genome Biology. 2014; 15(6):R84.
44. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Kallberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016;32(8):1220-2.
45. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333-i9.
46. Cameron DL, Schroder J, Penington JS, Do H, Molania R, Dobrovic A, et al. GRIDSS: sensitive and specific genomic rearrangement detection using positional de Bruijn graph assembly. Genome Res. 2017;27(12):2050-60.
47. Wala JA, Bandopadhayay P, Greenwald NF, O'Rourke R, Sharpe T, Stewart C, et al. SvABA: genome-wide detection of structural variants and indels by local assembly. Genome Res. 2018;28(4):581-91.
48. Behera S, Catreux S, Rossi M, Truong S, Huang Z, Ruehle M, et al. Comprehensive



doi:10.6342/NTU202500995

genome analysis and variant detection at scale using DRAGEN. Nature Biotechnology. 2024.
49. Zook JM, Hansen NF, Olson ND, Chapman L, Mullikin JC, Xiao C, et al. A robust benchmark for detection of germline large deletions and insertions. Nat Biotechnol. 2020;38(11):1347-55.
50. Porubsky D, Ebert P, Audano PA, Vollger MR, Harvey WT, Marijon P, et al. Fully phased human genome assembly without parental data using single-cell strand sequencing and long reads. Nat Biotechnol. 2021;39(3):302-8.
51. Cameron DL, Dong R, Papenfuss AT. StructuralVariantAnnotation: a R/Bioconductor foundation for a caller-agnostic structural variant software ecosystem. Bioinformatics. 2022;38(7):2046-8.
52. Kosugi S, Momozawa Y, Liu X, Terao C, Kubo M, Kamatani Y. Comprehensive evaluation of structural variation detection algorithms for whole genome sequencing. Genome Biol. 2019;20(1):117.
53. Gong T, Hayes VM, Chan EKF. Detection of somatic structural variants from short-read next-generation sequencing data. Brief Bioinform. 2021;22(3).
54. Joe S, Park JL, Kim J, Kim S, Park JH, Yeo MK, et al. Comparison of structural variant callers for massive whole-genome sequence data. BMC Genomics. 2024;25(1):318.
55. Zhuang G, Zhang X, Du W, Xu L, Ma J, Luo H, et al. A benchmarking framework for the accurate and cost-effective detection of clinically-relevant structural variants for cancer target identification and diagnosis. Journal of translational medicine. 2024;22(1):65.
56. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526(7571):75-81.
57. Li G, Jiang T, Li J, Wang Y. PanSVR: Pan-Genome Augmented Short Read Realignment for Sensitive Detection of Structural Variations. Front Genet. 2021;12:731515.
58. Rajaby R, Liu DX, Au CH, Cheung YT, Lau AYT, Yang QY, et al. INSurVeyor: improving insertion calling from short read sequencing data. Nat Commun. 2023;14(1):3243.
59. Chiara M, Pesole G, Horner DS. SVM²: an improved paired-end-based tool for the detection of small genomic structural variations using high-throughput single-genome resequencing data. Nucleic Acids Res. 2012;40(18):e145.



doi:10.6342/NTU202500995

60. Michaelson JJ, Sebat J. forestSV: structural variant discovery through statistical learning. Nat Methods. 2012;9(8):819-21.
61. Kronenberg ZN, Osborne EJ, Cone KR, Kennedy BJ, Domyan ET, Shapiro MD, et al. Wham: Identifying Structural Variants of Biological Consequence. PLoS Comput Biol. 2015;11(12):e1004572.
62. Parikh H, Mohiyuddin M, Lam HY, Iyer H, Chen D, Pratt M, et al. svclassify: a method to establish benchmark structural variant calls. BMC Genomics. 2016;17:64.
63. Alzaid EA, Allali AE, Aboalsamh H. A Classification Approach for Genome Structural Variations Detection. Journal of Proteomics & Bioinformatics. 2018;11(12).
64. Bartenhagen C, Dugas M. Robust and exact structural variation detection with paired-end and soft-clipped alignments: SoftSV compared with eight algorithms. Briefings in Bioinformatics. 2015;17(1):51-62.
65. Popic V, Rohlicek C, Cunial F, Hajirasouliha I, Meleshko D, Garimella K, et al. Cue: a deep-learning framework for structural variant discovery and genotyping. Nat Methods. 2023;20(4):559-68.
66. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods. 2018;15(6):461-8.
67. Jiang T, Liu Y, Jiang Y, Li J, Gao Y, Cui Z, et al. Long-read-based human genomic structural variation detection with cuteSV. Genome Biol. 2020;21(1):189.
68. Gaitán N, Duitama J. A graph clustering algorithm for detection and genotyping of structural variants from long reads. Gigascience. 2024;13.
69. Heller D, Vingron M. SVIM: structural variant identification using mapped long reads. Bioinformatics. 2019;35(17):2907-15.
70. Hu H, Gao R, Gao W, Gao B, Jiang Z, Zhou M, et al. SVDF: enhancing structural variation detect from long-read sequencing via automatic filtering strategies. Brief Bioinform. 2024;25(4).
71. Zhang Z, Jiang T, Li G, Cao S, Liu Y, Liu B, et al. Kled: an ultra-fast and sensitive structural variant detection tool for long-read sequencing data. Brief Bioinform. 2024;25(2).
72. Zheng Y, Shang X. SVvalidation: A long-read-based validation method for genomic structural variation. PloS one. 2024;19(1):e0291741.
73. Bagger FO, Borgwardt L, Jespersen AS, Hansen AR, Bertelsen B, Kodama M, et al. Whole genome sequencing in clinical practice. BMC Medical Genomics.



doi:10.6342/NTU202500995

2024;17(1):39.
74. Nurchis MC, Altamura G, Riccardi MT, Radio FC, Chillemi G, Bertini ES, et al. Whole genome sequencing diagnostic yield for paediatric patients with suspected genetic disorders: systematic review, meta-analysis, and GRADE assessment. Archives of Public Health. 2023;81(1):93.
75. Austin-Tse CA, Jobanputra V, Perry DL, Bick D, Taft RJ, Venner E, et al. Best practices for the interpretation and reporting of clinical whole genome sequencing. NPJ Genom Med. 2022;7(1):27.
76. Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The Human Phenotype Ontology in 2017. Nucleic Acids Res. 2017;45(D1):D865-d76.
77. Molina-Ramirez LP, Kyle C, Ellingford JM, Wright R, Taylor A, Bhaskar SS, et al. Personalised virtual gene panels reduce interpretation workload and maintain diagnostic rates of proband-only clinical exome sequencing for rare disorders. J Med Genet. 2022 Apr;59(4):393-398.
78. Schon U, Holzer A, Laner A, Kleinle S, Scharf F, Benet-Pages A, et al. HPO-driven virtual gene panel: a new efficient approach in molecular autopsy of sudden unexplained death. BMC Med Genomics. 2021;14(1):94.
79. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2018:201178.
80. Pagel KA, Antaki D, Lian A, Mort M, Cooper DN, Sebat J, et al. Pathogenicity and functional impact of non-frameshifting insertion/deletion variation in the human genome. PLoS Comput Biol. 2019;15(6):e1007112.
81. Geoffroy V, Herenger Y, Kress A, Stoetzel C, Piton A, Dollfus H, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics. 2018;34(20):3572-4.
82. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, et al. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res. 2017;27(11):1916-29.
83. Torene RI, Galens K, Liu S, Arvai K, Borroto C, Scuffins J, et al. Mobile element insertion detection in 89,874 clinical exomes. Genetics in medicine : official journal of the American College of Medical Genetics. 2020;22(5):974-8.
84. Chu C, Borges-Monroy R, Viswanadham VV, Lee S, Li H, Lee EA, et al. Comprehensive identification of transposable element insertions using multiple sequencing technologies. Nat Commun. 2021;12(1):3836.



doi:10.6342/NTU202500995

85. Djie Tjwan Thung, Joep de Ligt, Lisenka EM Vissers, Marloes Steehouwer, Mark Kroon, Petra de Vries, Eline P Slagboom, Kai Ye, Joris A Veltman and Jayne Y Hehir-Kwa. Mobster: accurate detection of mobile element insertions in next generation sequencing data. Genome Biology (2014);15:488.
86. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22(9):1760-74.
87. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res. 2019;47(D1):D886-D94.
88. Jacobsen JOB, Kelly C, Cipriani V, Research Consortium GE, Mungall CJ, Reese J, et al. Phenotype-driven approaches to enhance variant prioritization and diagnosis of rare disease. Hum Mutat. 2022;43(8):1071-81.
89. Barbitoff YA, Ushakov MO, Lazareva TE, Nasykhova YA, Glotov AS, Predeus AV. Bioinformatics of germline variant discovery for rare disease diagnostics: current approaches and remaining challenges. Brief Bioinform. 2024;25(2).
90. Sheikh Hassani M, Jain R, Ramaswamy S, Sinha S, El Naofal M, Halabi N, et al. Virtual Gene Panels Have a Superior Diagnostic Yield for Inherited Rare Diseases Relative to Static Panels. Clin Chem. 2025;71(1):169-84.
91. Schluter A, Velez-Santamaria V, Verdura E, Rodriguez-Palmero A, Ruiz M, Fourcade S, et al. ClinPrior: an algorithm for diagnosis and novel gene discovery by network-based prioritization. Genome Med. 2023;15(1):68.



doi:10.6342/NTU202500995

# 6. Appendices

**Table S1. Detailed genetic variants of NTU TAAD panel positive patients.**

| TAAD no. | Age | Sex | Aneurysm | Dissection | EL | FH | MFS score | cTAAD | gTAAD | Gene | Transcript no. | c.DNA | p.Aminoacid | Type | Zygosity | rs no. | SIFT | Polyphen-2 | AF | ACMG |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TAAD0006 | 29 | F | + | + | - | - | 16 | + | + | FBN1 | NM_000138 | c.6580G>T | p.Glu2194Ter | nonsense | Hetero | NIL | NA | NA | NA | P |
| TAAD0022 | 8 | F | + | - | - | - | 9 | + | + | FBN1 | NM_000138 | c.8034G>C | p.Tyr2678Ter | nonsense | Hetero | NIL | NA | NA | NA | P |
| TAAD0048 | 23 | M | + | - | - | - | 8 | + | + | FBN1 | NM_000138 | c.679C>T | p.Gln227Ter | nonsense | Hetero | NIL | NA | NA | NA | P |
| TAAD0049 | 36 | F | + | - | NA | - | 4 | + | + | FBN1 | NM_000138 | c.2901C>A | p.Cys967Ter | nonsense | Hetero | NIL | NA | NA | NA | P |
| TAAD0062 | 19 | M | + | - | + | - | 12 | + | + | FBN1 | NM_000138 | c.284C>A | p.Ser95Ter | nonsense | Hetero | NIL | NA | NA | NA | P |
| TAAD0097 | 26 | F | + | - | - | - | 6 | + | + | FBN1 | NM_000138 | c.2393G>T | p.Gly795Ter | nonsense | Hetero | NIL | NA | NA | NA | P |
| TAAD0100 | 25 | F | + | - | NA | + | 10 | + | + | FBN1 | NM_000138 | c.4621C>T | p.Arg1541Ter | nonsense | Hetero | rs794728228 | NA | NA | NA | P |
| TAAD0013 | 17 | M | + | - | - | + | 11 | + | + | FBN1 | NM_000138 | c.4816+1G>A | NA | splice site | Hetero | NIL | NA | NA | NA | P |
| TAAD0025 | 25 | M | + | - | + | - | 11 | + | + | FBN1 | NM_000138 | c.5918-1G>C | NA | splice site | Hetero | NIL | NA | NA | NA | P |
| TAAD0052 | 23 | M | + | - | + | - | 7 | + | + | FBN1 | NM_000138 | c.4460-2A>C | NA | splice site | Hetero | NIL | NA | NA | NA | P |
| TAAD0056 | 8 | M | + | - | + | - | 10 | + | + | FBN1 | NM_000138 | c.5296+2T>C | NA | splice site | Hetero | rs113963967 | NA | NA | NA | P |
| TAAD0126 | 35 | M | + | + | - | - | 5 | + | + | FBN1 | NM_000138 | c.5671+1G>- | NA | splice site | Hetero | NIL | NA | NA | NA | P |
| TAAD0067 | 20 | M | + | - | - | - | 12 | + | + | FBN1 | NM_000138 | c.4376dupG | p.Gly1459fs | frameshift | Hetero | NIL | NA | NA | NA | P |
| TAAD0087 | 27 | F | + | + | - | - | 12 | + | + | FBN1 | NM_000138 | c.7039_7040delAT | p.Met2347fs | frameshift | Hetero | rs794728319 | NA | NA | NA | P |
| TAAD0101 | 19 | M | + | - | + | - | 8 | + | + | FBN1 | NM_000138 | c.6281_6282insGA | p.Asp2094fs | frameshift | Hetero | NIL | NA | NA | NA | P |
| TAAD0021 | 30 | M | + | - | + | + | 3 | + | + | FBN1 | NM_000138 | c.640G>A | p.Gly214Ser | missense | Hetero | rs113111224 | 0.002 | 1 | NA | P |
| TAAD0037 | 27 | F | - | - | - | + | 3 | - | + | FBN1 | NM_000138 | c.2860C>T | p.Arg954Cys | missense | Hetero | NIL | 0 | 1 | NA | P |
| TAAD0081 | 12 | M | + | - | - | - | 10 | + | + | FBN1 | NM_000138 | c.6388G>A | p.Glu2130Lys | missense | Hetero | rs794728334 | 0.002 | 0.999 | 0 | P |



doi:10.6342/NTU202500995

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TAAD0031 | 9 | F | - | - | - | - | 5 | - | + | FBN2 | NM_001999 | c.1850-1G>A | NA | splice site | Hetero | NIL | NA | NA | NA | P |
| TAAD0024 | 6 | F | - | - | - | - | 5 | - | + | FBN2 | NM_001999 | c.3802C>T | p.Cys1268Arg | missense | Hetero | NIL | 0.001 | 0.999 | NA | P |
| TAAD0055 | 39 | F | + | + | - | + | 5 | + | + | TGFBR2 | NM_003242 | c.1483C>T | p.Arg495Ter | nonsense | Hetero | rs104893819 | NA | NA | NA | P |
| TAAD0080 | 5 | M | + | - | - | - | 8 | + | + | TGFBR2 | NM_003242 | c.1582C>T | p.Arg528RCys | missense | Hetero | rs104893810 | 0 | 1 | NA | P |
| TAAD0027 | 6 | F | + | - | - | - | 9 | + | + | TGFBR1 | NM_004612 | c.722C>T | p.Ser241Leu | missense | Hetero | rs111854391 | 0 | 1 | NA | P |
| TAAD0030 | 26 | F | + | - | - | + | 1 | + | + | TGFBR1 | NM_004612 | c.1459C>T | p.Arg487Trp | missense | Hetero | rs111426349 | 0 | 1 | 0 | P |
| TAAD0001 | 26 | M | + | + | NA | - | NA | + | + | FBN1 | NM_000138 | c.3012C>G | p.Tyr1004Ter | nonsense | Hetero | rs397515784 | NA | NA | NA | LP |
| TAAD0115 | 20 | F | + | - | - | - | 9 | + | + | FBN1 | NM_000138 | c.3408C>G | p.Tyr1136Ter | nonsense | Hetero | NIL | NA | NA | NA | LP |
| TAAD0044 | 31 | F | + | - | + | - | 9 | + | + | FBN1 | NM_000138 | c.5788+5G>A | NA | splice site | Hetero | rs193922219 | NA | NA | 0 | LP |
| TAAD0109 | 7 | F | - | - | NA | - | 7 | + | + | FBN1 | NM_000138 | c.5788+5G>A | NA | splice site | Hetero | rs193922219 | NA | NA | 0 | LP |
| TAAD0050 | 25 | F | - | - | NA | - | 9 | - | + | FBN1 | NM_000138 | c.3952delA | p.Thr1318fs | frameshift | Hetero | NIL | NA | NA | NA | LP |
| TAAD0039 | 45 | F | + | + | + | NA | 4 | + | + | FBN1 | NM_000138 | c.2627G>T | p.Cys876Phe | missense | Hetero | NIL | 0.001 | 1 | NA | LP |
| TAAD0045 | 23 | M | + | - | + | - | 8 | + | + | FBN1 | NM_000138 | c.4532G>A | p.Cys1511Tyr | missense | Hetero | rs1060501074 | 0 | 1 | NA | LP |
| TAAD0054 | 34 | M | + | + | - | + | 5 | + | + | FBN1 | NM_000138 | c.6182G>C | p.Cys2061Ser | missense | Hetero | NIL | 0.003 | 0.932 | NA | LP |
| TAAD0058 | 34 | F | + | + | - | - | 7 | + | + | FBN1 | NM_000138 | c.1766A>G | p.Asn589Ser | missense | Hetero | NIL | 0 | 0.993 | NA | LP |
| TAAD0059 | 69 | F | + | - | - | + | 5 | + | + | FBN1 | NM_000138 | c.6332G>A | p.Cys2111Tyr | missense | Hetero | rs1131691467 | 0 | 0.981 | NA | LP |
| TAAD0060 | 11 | M | + | - | + | - | 11 | + | + | FBN1 | NM_000138 | c.3545G>A | p.Cys1182Tyr | missense | Hetero | NIL | 0.001 | 0.997 | NA | LP |
| TAAD0061 | 35 | M | + | - | + | + | 6 | + | + | FBN1 | NM_000138 | c.7493T>G | p.Phe2498Cys | missense | Hetero | NIL | 0.069 | 0.999 | NA | LP |
| TAAD0065 | 6 | M | - | - | + | - | 3 | - | + | FBN1 | NM_000138 | c.2306G>C | p.Cys769Ser | missense | Hetero | NIL | 0 | 0.997 | NA | LP |
| TAAD0076 | 40 | F | + | - | - | - | 9 | + | + | FBN1 | NM_000138 | c.6113G>A | p.Cys2038Tyr | missense | Hetero | rs363804 | 0 | 0.979 | NA | LP |
| TAAD0111 | 23 | M | + | - | - | + | 7 | + | + | FBN1 | NM_000138 | c.4258T>A | p.Cys1420Ser | missense | Hetero | NIL | 0 | 0.991 | NA | LP |
| TAAD0120 | 3 | M | + | - | + | - | 4 | + | + | FBN1 | NM_000138 | c.5950T>C | pCys1984Arg | missense | Hetero | NIL | 0 | 0.999 | NA | LP |



doi:10.6342/NTU202500995

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TAAD0124 | 33 | F | - | - | + | + | 1 | - | + | FBN1 | NM_000138 | c.4532G>A | pCys1511Tyr | missense | Hetero | rs1060501074 | 0 | 1 | NA | LP |
| TAAD0130 | 2 | F | + | - | + | + | 5 | - | + | FBN1 | NM_000138 | c.6331T>C | p.C2111R | missense | Hetero | rs363815 | 0 | 0.944 | NA | LP |
| TAAD0023 | 38 | M | + | - | - | + | 6 | + | + | TGFBR1 | NM_004612 | c.688G>C | p.Ala230Pro | missense | Hetero | NIL | 0 | 1 | NA | LP |
| TAAD0038 | 16 | M | + | - | - | - | 11 | + | + | TGFBR1 | NM_004612 | c.827T>C | p.Leu276Pro | missense | Hetero | NIL | 0 | 1 | NA | LP |
| TAAD0103 | 47 | M | + | + | - | NA | 2 | + | + | TGFBR1 | NM_004612 | c.848A>G | p.His283Arg | missense | Hetero | NIL | 0 | 1 | NA | LP |
| TAAD0075 | 59 | M | + | + | - | + | 3 | + | + | TGFBR2 | NM_003242 | c.1344C>A | p.Thr445Asn | missense | Hetero | rs886038936 | 0.005 | 0.99 | NA | LP |

EL: ectopia lentis; FH: family history of TAAD; S-score: systemic score; Hetero: heterozygous; MAF: minor allele frequency of east Asian in genome aggregation database (gnomAD); P: pathogenic; LP: likely pathogenic; ACMG: American College of Medical Genetics



doi:10.6342/NTU202500995

**Table S2. Clinical characteristics of TAAD patients stratified by NTU TAAD panel genetic findings.**

| | gTAAD+ (46) | gTAAD- (61) | P value | P value adjusted by age and sex |
|---|---|---|---|---|
| Age | 23 [7-39] | 29 [12-46] | 0.008 | |
| Sex (male) | 24 (52.2%) | 43 (70.5%) | 0.053 | |
| BH (cm) | 175 [151-196] | 177 [164-189] | 0.495 | |
| BW (kg) | 61.8 [38.4-85.2] | 61.5 [44.7-78.3] | 0.0390 | |
| BMI (kg/m$^2$) | 20.25 [15.35-25.15] | 20.35 [14.75-25.95] | 0.154 | |
| Arm/BH | 1.02 [0.97-1.07] | 1.01 [0.98-1.04] | 0.620 | |
| Ratio of Upper/Lower segments | 0.71 [0.57-0.85] | 0.74 [0.57-0.91] | 0.172 | |
| HTN | 2 (5.3%) | 6 (14.3%) | 0.179 | |
| DLP | 2 (5.3%) | 4 (10.5%) | 0.674 | |
| DM | 0 (0.0%) | 1 (2.6%) | 1.000 | |
| Systemic score | 8 [3-13] | 5.0 [1-11] | 0.020 | 0.017 |
| Systemic score ≥7 | 26 (60.5%) | 10 (23.8%) | 0.001 | 0.003 |
| Ectopic lens | 13 (30.2%) | 7 (13.0%) | 0.037 | 0.083 |
| Wrist or Thumb signs | 36 (85.7%) | 30 (76.9%) | 0.0309 | |
| Chest wall abnormality | 26 (61.9%) | 13 (32.5%) | 0.008 | 0.008 |
| Pectus carinatum | 11 (26.2%) | 7 (17.5%) | 0.0342 | |
| Pectus excavatum | 15 (35.7%) | 6 (15.0%) | 0.032 | 0.024 |
| Hind foot abnormality | 2 (4.9%) | 1 (2.6%) | 1.000 | |
| Pes planus | 23 (56.1%) | 12 (29.3%) | 0.014 | 0.026 |
| Pneumothorax | 3 (7.5%) | 6 (15.4%) | 0.311 | |
| Dural ectasia | 15 (83.3%) | 6 (35.3%) | 0.004 | 0.007 |
| Hip abnormality | 6 (66.7%) | 2 (14.3%) | 0.023 | |
| Scoliosis | 19 (45.2%) | 19 (47.5%) | 0.837 | |
| Elbow extension | 0 (0.0%) | 2 (6.1%) | 0.225 | |
| Facial features | 7 (18.4%) | 5 (15.6%) | 0.757 | |
| Skin striae | 21 (50%) | 16 (42.1%) | 0.479 | |
| Severe myopia (>3D) | 31 (73.8%) | 27 (57.4%) | 0.106 | |
| Mitral valve prolapse | 30 (73.2%) | 21 (52.5%) | 0.054 | 0.167 |
| Upper/Lower Ratio | 2 (5.9%) | 3 (12.0%) | 0.641 | |

Abbreviation: BH, body height; BMI, body mass index; BW body weight; DLP, dyslipidemia; DM, diabetes mellitus; gTAAD+, disease-causing gene found on TAAD panel; gTAAD-, disease-causing gene not found on TAAD panel; HTN, hypertension; TAAD, thoracic aortic aneurysm and dissection.



doi:10.6342/NTU202500995

**Table S3. Types and sizes of all SVs detected by each individual algorithm in HG002**

| SV callers | SV sizes | SV types | | | | |
|---|---|---|---|---|---|---|
| | | DEL | INS | DUP | INV | CXT |
| Manta | [0, 50) | 40 | 518 | 1 | 0 | - |
| | [50, 1K) | 3,297 | 1,800 | 344 | 85 | - |
| | [1K, 10K) | 544 | 0 | 47 | 55 | - |
| | [10K, 100K) | 62 | 0 | 21 | 31 | - |
| | [100K, 1000K) | 19 | 0 | 15 | 25 | - |
| | ≥1000K | 13 | 0 | 14 | 50 | - |
| Raw (Total) | 7,317 | 3,976 | 2,318 | 442 | 246 | 335 |
| **Filtered (≥50 bp)** | **7,758** | **3,936** | **1,800** | **441** | **246** | **335** |
| DELLY | [0, 50) | 5,229 | 1,719 | 0 | 0 | - |
| | [50, 1K) | 2,739 | 155 | 677 | 73 | - |
| | [1K, 10K) | 912 | 0 | 200 | 115 | - |
| | [10K, 100K) | 232 | 0 | 172 | 158 | - |
| | [100K, 1000K) | 105 | 0 | 107 | 182 | - |
| | ≥1000K | 154 | 0 | 231 | 490 | - |
| Raw (Total) | 13,650 | 9,371 | 1,874 | 1,387 | 1,018 | 0 |
| **Filtered (≥50 bp)** | **6,702** | **4,142** | **155** | **1,387** | **1,018** | **0** |
| GRIDSS | [0, 50) | 24,045 | 23206 | 51 | 6 | - |
| | [50, 1K) | 2,097 | 90 | 514 | 92 | - |
| | [1K, 10K) | 527 | 0 | 36 | 60 | - |
| | [10K, 100K) | 43 | 0 | 9 | 16 | - |
| | [100K, 1000K) | 7 | 0 | 4 | 4 | - |
| | ≥1000K | 7 | 0 | 11 | 16 | - |
| Raw (Total) | 51,056 | 26,726 | 23,296 | 628 | 194 | 212 |
| **Filtered (≥50 bp)** | **3,748** | **2,681** | **90** | **577** | **188** | **212** |
| LUMPY | [0, 50) | 0 | 0 | 0 | 0 | - |
| | [50, 1K) | 219 | 0 | 147 | 57 | - |
| | [1K, 10K) | 933 | 0 | 304 | 93 | - |
| | [10K, 100K) | 238 | 0 | 190 | 126 | - |
| | [100K, 1000K) | 93 | 0 | 97 | 169 | - |
| | ≥1000K | 45 | 0 | 56 | 186 | - |
| Raw (Total) | 4,656 | 1,528 | 0 | 794 | 631 | 1,703 |
| **Filtered (≥50 bp)** | **4,656** | **1,528** | **0** | **794** | **631** | **1,703** |
| SvABA | [0, 50) | 3 | 1 | 4 | 0 | - |
| | [50, 1K) | 1,217 | 0 | 1,585 | 37 | - |



doi:10.6342/NTU202500995

| | | | | | | |
|---|---|---|---|---|---|---|
| | [1K, 10K) | 611 | 0 | 64 | 40 | - |
| | [10K, 100K) | 38 | 0 | 15 | 21 | - |
| | [100K, 1000K) | 27 | 0 | 21 | 25 | - |
| | ≥1000K | 17 | 0 | 25 | 50 | - |
| Raw (Total) | 3,849 | 1,913 | 1 | 1,714 | 173 | 48 |
| **Filtered (≥50 bp)** | **3,841** | **1,910** | **0** | **1,710** | **173** | **48** |
| **DRAGEN** | [0, 50) | 41 | 2126 | 0 | 0 | - |
| | [50, 1K) | 4,091 | 3,403 | 15 | 78 | - |
| | [1K, 10K) | 620 | 0 | 51 | 46 | - |
| | [10K, 100K) | 53 | 0 | 15 | 28 | - |
| | [100K, 1000K) | 15 | 0 | 11 | 21 | - |
| | ≥1000K | 8 | 0 | 9 | 40 | - |
| Raw (Total) | 10,953 | 4,828 | 5,529 | 101 | 213 | 282 |
| **Filtered (≥50 bp)** | **10,912** | **4,787** | **5,529** | **101** | **213** | **282** |

Raw: total number of detected structural variants (SVs); Filtered: number of detected SVs ≥50 bp.

**Table S4. Types and sizes of all SVs detected by each individual algorithm in HG00514**

| SV callers | SV sizes | SV types | | | | |
|---|---|---|---|---|---|---|
| | | DEL | INS | DUP | INV | CXT |
| **Manta** | [0, 50) | 31 | 495 | 0 | 0 | - |
| | [50, 1K) | 3,439 | 2,297 | 353 | 92 | - |
| | [1K, 10K) | 637 | 0 | 73 | 66 | - |
| | [10K, 100K) | 67 | 0 | 39 | 34 | - |
| | [100K, 1000K) | 25 | 0 | 31 | 42 | - |
| | ≥1000K | 15 | 0 | 24 | 67 | - |
| Raw (Total) | 8,632 | 4,214 | 2,786 | 520 | 301 | 811 |
| **Filtered (≥50 bp)** | **8,106** | **4,183** | **2,291** | **520** | **301** | **811** |
| **DELLY** | [0, 50) | 4,977 | 1,471 | 0 | 0 | - |
| | [50, 1K) | 2,616 | 128 | 645 | 88 | - |
| | [1K, 10K) | 962 | 0 | 267 | 116 | - |
| | [10K, 100K) | 343 | 0 | 300 | 133 | - |
| | [100K, 1000K) | 240 | 0 | 251 | 221 | - |
| | ≥1000K | 198 | 0 | 265 | 464 | - |
| Raw (Total) | 13,685 | 9,336 | 1,599 | 1,728 | 1,022 | 0 |
| **Filtered (≥50 bp)** | **7,237** | **4,359** | **128** | **1,728** | **1,022** | **0** |
| **GRIDSS** | [0, 50) | 23,091 | 23,185 | 66 | 13 | - |
| | [50, 1K) | 2,120 | 90 | 490 | 78 | - |



doi:10.6342/NTU202500995

| | | | | | | |
|---|---|---|---|---|---|---|
| | [1K, 10K) | 559 | 0 | 42 | 63 | - |
| | [10K, 100K) | 30 | 0 | 7 | 7 | - |
| | [100K, 1000K) | 5 | 0 | 4 | 1 | - |
| | ≥1000K | 4 | 0 | 7 | 16 | - |
| Raw (Total) | 50,444 | 25,809 | 23,275 | 616 | 178 | 566 |
| **Filtered (≥50 bp)** | **4,089** | **2,718** | **90** | **550** | **165** | **566** |
| **LUMPY** | [0, 50) | 150 | 0 | 0 | 7 | - |
| | [50, 1K) | 2,302 | 0 | 305 | 142 | - |
| | [1K, 10K) | 1,073 | 0 | 391 | 110 | - |
| | [10K, 100K) | 361 | 0 | 319 | 122 | - |
| | [100K, 1000K) | 180 | 0 | 186 | 221 | - |
| | ≥1000K | 137 | 0 | 153 | 277 | - |
| Raw (Total) | 12,245 | 4,203 | 0 | 1,354 | 879 | 5,809 |
| **Filtered (≥50 bp)** | **12,088** | **4,053** | **0** | **1,354** | **872** | **5,809** |
| **SvABA** | [0, 50) | 3 | 0 | 8 | 0 | - |
| | [50, 1K) | 1,113 | 0 | 1,136 | 66 | - |
| | [1K, 10K) | 684 | 0 | 82 | 54 | - |
| | [10K, 100K) | 68 | 0 | 47 | 18 | - |
| | [100K, 1000K) | 80 | 0 | 70 | 65 | - |
| | ≥1000K | 36 | 0 | 42 | 69 | - |
| **Raw (Total)** | **3,883** | **1,984** | **0** | **1,385** | **272** | **242** |
| **Filtered (≥50 bp)** | **3,872** | **1,981** | **0** | **1,377** | **272** | **242** |
| **DRAGEN** | [0, 50) | 39 | 2,296 | 0 | 0 | - |
| | [50, 1K) | 4,922 | 5,090 | 37 | 97 | - |
| | [1K, 10K) | 825 | 10 | 158 | 76 | - |
| | [10K, 100K) | 222 | 0 | 200 | 48 | - |
| | [100K, 1000K) | 226 | 0 | 220 | 71 | - |
| | ≥1000K | 208 | 0 | 229 | 133 | - |
| Raw (Total) | 16,072 | 6,442 | 7,396 | 844 | 425 | 965 |
| **Filtered (≥50 bp)** | **13,737** | **6,403** | **5,100** | **844** | **425** | **965** |

Raw: total number of detected structural variants (SVs); Filtered: number of detected SVs ≥50 bp.

**Table S5. Types and sizes of all SVs detected by each individual algorithm in HG00733**

| SV callers | SV sizes | SV types | | | | |
|---|---|---|---|---|---|---|
| | | DEL | INS | DUP | INV | CXT |
| **Manta** | [0, 50) | 30 | 513 | 0 | 0 | - |
| | [50, 1K) | 3,499 | 2,312 | 376 | 91 | - |



doi:10.6342/NTU202500995

| | | | | | | |
|---|---|---|---|---|---|---|
| | [1K, 10K) | 603 | 0 | 62 | 82 | - |
| | [10K, 100K) | 87 | 0 | 39 | 36 | - |
| | [100K, 1000K) | 25 | 0 | 31 | 33 | - |
| | ≥1000K | 18 | 0 | 24 | 58 | - |
| Raw (Total) | 8,747 | 4,262 | 2,825 | 532 | 300 | 828 |
| **Filtered (≥50 bp)** | **8,204** | **4,232** | **2,312** | **532** | **300** | **828** |
| **DELLY** | [0, 50) | 4,939 | 1,497 | 0 | 0 | - |
| | [50, 1K) | 2,734 | 130 | 712 | 78 | - |
| | [1K, 10K) | 944 | 0 | 253 | 115 | - |
| | [10K, 100K) | 396 | 0 | 340 | 145 | - |
| | [100K, 1000K) | 268 | 0 | 266 | 210 | - |
| | ≥1000K | 232 | 0 | 263 | 450 | - |
| Raw (Total) | 13,972 | 9,513 | 1,627 | 1,834 | 998 | 0 |
| **Filtered (≥50 bp)** | **7,536** | **4,574** | **130** | **1,834** | **998** | **0** |
| **GRIDSS** | [0, 50) | 22,605 | 23,244 | 61 | 7 | - |
| | [50, 1K) | 2,102 | 81 | 523 | 73 | - |
| | [1K, 10K) | 504 | 0 | 41 | 67 | - |
| | [10K, 100K) | 42 | 0 | 4 | 9 | - |
| | [100K, 1000K) | 3 | 0 | 5 | 3 | - |
| | ≥1000K | 4 | 0 | 10 | 15 | - |
| Raw (Total) | 49,972 | 25,260 | 23,325 | 644 | 174 | 569 |
| **Filtered (≥50 bp)** | **4,055** | **2,655** | **81** | **583** | **167** | **569** |
| **LUMPY** | [0, 50) | 165 | 0 | 0 | 4 | - |
| | [50, 1K) | 2,351 | 0 | 325 | 145 | - |
| | [1K, 10K) | 1,046 | 0 | 403 | 114 | - |
| | [10K, 100K) | 423 | 0 | 406 | 129 | - |
| | [100K, 1000K) | 184 | 0 | 194 | 207 | - |
| | ≥1000K | 166 | 0 | 157 | 268 | - |
| Raw (Total) | 13,134 | 4,335 | 0 | 1,485 | 867 | 6,447 |
| **Filtered (≥50 bp)** | **12,965** | **4,170** | **0** | **1,485** | **863** | **6,447** |
| **SvABA** | [0, 50) | 0 | 2 | 6 | 2 | - |
| | [50, 1K) | 1,113 | 0 | 1,256 | 56 | - |
| | [1K, 10K) | 644 | 0 | 86 | 52 | - |
| | [10K, 100K) | 61 | 0 | 48 | 20 | - |
| | [100K, 1000K) | 72 | 0 | 90 | 43 | - |
| | ≥1000K | 54 | 0 | 43 | 60 | - |
| Raw (Total) | 3,976 | 1,944 | 2 | 1,529 | 233 | 268 |



doi:10.6342/NTU202500995

| | | | | | | |
|---|---|---|---|---|---|---|
| **Filtered (≥50 bp)** | **3,966** | **1,944** | **0** | **1,523** | **231** | **268** |
| **DRAGEN** | [0, 50) | 41 | 2,029 | 0 | 0 | - |
| | [50, 1K) | 5,072 | 5,312 | 27 | 76 | - |
| | [1K, 10K) | 675 | 1 | 62 | 60 | - |
| | [10K, 100K) | 67 | 0 | 17 | 9 | - |
| | [100K, 1000K) | 19 | 0 | 18 | 7 | - |
| | ≥1000K | 62 | 0 | 81 | 24 | - |
| Raw (Total) | 14,000 | 5,936 | 7,342 | 205 | 176 | 341 |
| **Filtered (≥50 bp)** | **11,903** | **5,895** | **5,313** | **205** | **176** | **341** |

Raw: total number of detected structural variants (SVs); Filtered: number of detected SVs ≥50 bp.

**Table S6. Types and sizes of all SVs detected by each individual algorithm in NA19240**

| SV callers | SV sizes | SV types | | | | |
|---|---|---|---|---|---|---|
| | | **DEL** | **INS** | **DUP** | **INV** | **CXT** |
| **Manta** | [0, 50) | 35 | 599 | 0 | 0 | - |
| | [50, 1K) | 4,083 | 2,488 | 437 | 81 | - |
| | [1K, 10K) | 754 | 0 | 63 | 71 | - |
| | [10K, 100K) | 74 | 0 | 36 | 24 | - |
| | [100K, 1000K) | 27 | 0 | 32 | 43 | - |
| | ≥1000K | 17 | 0 | 15 | 65 | - |
| **Raw (Total)** | **9,867** | **4,990** | **3,087** | **583** | **284** | **923** |
| **Filtered (≥50 bp)** | **9,233** | **4,955** | **2,488** | **583** | **284** | **923** |
| **DELLY** | [0, 50) | 5,825 | 1,681 | 0 | 0 | - |
| | [50, 1K) | 3,161 | 134 | 674 | 92 | - |
| | [1K, 10K) | 1107 | 0 | 275 | 123 | - |
| | [10K, 100K) | 372 | 0 | 328 | 146 | - |
| | [100K, 1000K) | 213 | 0 | 214 | 268 | - |
| | ≥1000K | 208 | 0 | 246 | 497 | - |
| **Raw (Total)** | **15,564** | **10,886** | **1,815** | **1,737** | **1,126** | **0** |
| **Filtered (≥50 bp)** | **8,058** | **5,061** | **134** | **1,737** | **1,126** | **0** |
| **GRIDSS** | [0, 50) | 27,252 | 26,535 | 61 | 0 | - |
| | [50, 1K) | 2,411 | 72 | 577 | 81 | - |
| | [1K, 10K) | 642 | 0 | 42 | 65 | - |
| | [10K, 100K) | 36 | 0 | 6 | 4 | - |
| | [100K, 1000K) | 6 | 0 | 3 | 2 | - |
| | ≥1000K | 3 | 0 | 9 | 9 | - |
| **Raw (Total)** | **58,446** | **30,350** | **26,607** | **698** | **171** | **620** |



doi:10.6342/NTU202500995

| | | | | | | |
|---|---|---|---|---|---|---|
| **Filtered (≥50 bp)** | **4,598** | **3,098** | **72** | **637** | **171** | **620** |
| **LUMPY** | [0, 50) | 179 | 0 | 0 | 5 | - |
| | [50, 1K) | 2,741 | 0 | 345 | 142 | - |
| | [1K, 10K) | 1,196 | 0 | 400 | 119 | - |
| | [10K, 100K) | 382 | 0 | 362 | 127 | - |
| | [100K, 1000K) | 154 | 0 | 165 | 262 | - |
| | ≥1000K | 135 | 0 | 149 | 297 | - |
| **Raw (Total)** | **13,667** | **4,787** | **0** | **1,421** | **952** | **6,507** |
| **Filtered (≥50 bp)** | **13,483** | **4,608** | **0** | **1,421** | **947** | **6,507** |
| **SvABA** | [0, 50) | 5 | 1 | 10 | 0 | - |
| | [50, 1K) | 1,331 | 0 | 1,161 | 70 | - |
| | [1K, 10K) | 802 | 0 | 73 | 40 | - |
| | [10K, 100K) | 66 | 0 | 36 | 22 | - |
| | [100K, 1000K) | 49 | 0 | 68 | 55 | - |
| | ≥1000K | 35 | 0 | 31 | 72 | - |
| **Raw (Total)** | **4,213** | **2,288** | **1** | **1,379** | **259** | **286** |
| **Filtered (≥50 bp)** | **4,197** | **2,283** | **0** | **1,369** | **259** | **286** |
| **DRAGEN** | [0, 50) | 49 | 2,375 | 0 | 0 | - |
| | [50, 1K) | 5,905 | 6,144 | 24 | 74 | - |
| | [1K, 10K) | 840 | 0 | 58 | 57 | - |
| | [10K, 100K) | 55 | 0 | 20 | 5 | - |
| | [100K, 1000K) | 21 | 0 | 19 | 11 | - |
| | ≥1000K | 76 | 0 | 88 | 18 | - |
| **Raw (Total)** | **16,188** | **6,946** | **8,501** | **209** | **165** | **367** |
| **Filtered (≥50 bp)** | **13,764** | **6,897** | **6,126** | **209** | **165** | **367** |

Raw: total number of detected structural variants (SVs); Filtered: number of detected SVs ≥50 bp.



doi:10.6342/NTU202500995

**Table S7. Sizes and counts of "DELs" and "INSs" in truth set**

| SV truth sets | SV sizes | SV types | |
|---|---|---|---|
| | | DEL | INS |
| GIAB v.6 Tier1 (HG002) | [50, 1K) | **4,846** | **6,341** |
| | [1K, 10K) | **578** | **909** |
| | [10K, 100K) | 39 | 29 |
| | [100K, 1000K) | 1 | 2 |
| Raw (Total) | 12,745 | **5,464** | **7,281** |
| HGSVC2 (HG00514) | [50, 1K) | **7,604** | **10,906** |
| | [1K, 10K) | 867 | 1,385 |
| | [10K, 100K) | 55 | 42 |
| | [100K, 1000K) | 3 | 0 |
| Raw (Total) | 20,862 | 8,529 | 12,333 |
| HGSVC2 (HG00733) | [50, 1K) | **7,799** | **11,064** |
| | [1K, 10K) | **832** | **1,396** |
| | [10K, 100K) | 61 | 35 |
| | [100K, 1000K) | 2 | 0 |
| Raw (Total) | 21,189 | 8,694 | 12,495 |
| HGSVC2 (NA19240) | [50, 1K) | **9,112** | **12,612** |
| | [1K, 10K) | **986** | **1,502** |
| | [10K, 100K) | 62 | 39 |
| | [100K, 1000K) | 5 | 0 |
| Raw (Total) | 24,318 | 10,165 | 14,153 |

Raw: total number of SVs in truth set.



doi:10.6342/NTU202500995

**Table S8. Count of neighbor SVs.**

| HG002 | | |
|---|---|---|
| | **3-SV callers** | **5-SV callers** |
| Total SV | 16,660 | 23,406 |
| Total Neighbor SV | 9,771 | 11,255 |
| Neighbor SV (include 1 member variant) | **5,762** | **5,746** |
| Neighbor SV (include 2 member variants) | **1,535** | **1,964** |
| Neighbor SV (include 3 member variants) | **2,383** | **1,585** |
| Neighbor SV (include 4 member variants) | 58 | **1,277** |
| Neighbor SV (include 5 member variants) | 13 | **589** |
| Neighbor SV (include >5 member variants) | 20 | 94 |
| Multiple agreement (≥2 callers detected the variants) | 3,789 | 5,280 |
| Union (≥1 callers detected the variants) | 9,771 | 9,771 |
| **HG00514** | | |
| | **3-SV callers** | **5-SV callers** |
| Total SV | 18,055 | 27,964 |
| Total Neighbor SV | 11,164 | 12,498 |
| Neighbor SV (include 1 member variant) | **7,016** | **6,251** |
| Neighbor SV (include 2 member variants) | **1,680** | **2,010** |
| Neighbor SV (include 3 member variants) | **2,386** | **1,253** |
| Neighbor SV (include 4 member variants) | 56 | **1,432** |
| Neighbor SV (include 5 member variants) | 15 | **1,409** |
| Neighbor SV (include >5 member variants) | 11 | 143 |
| Multiple agreement (≥2 callers detected the variants) | 3,858 | 6,034 |
| Union (≥1 callers detected the variants) | 11,164 | 12,498 |
| **HG00733** | | |
| | **3-SV callers** | **5-SV callers** |
| Total SV | 15,383 | 22,975 |
| Total Neighbor SV | 9,252 | 10,382 |
| Neighbor SV (include 1 member variant) | **5,509** | **5,476** |
| Neighbor SV (include 2 member variants) | **1,429** | **1,324** |
| Neighbor SV (include 3 member variants) | **2,259** | **966** |
| Neighbor SV (include 4 member variants) | 40 | **1,267** |
| Neighbor SV (include 5 member variants) | 11 | **1,271** |
| Neighbor SV (include >5 member variants) | 4 | 78 |
| Multiple l agreement (≥2 callers detected the variants) | 3,640 | 4,756 |
| Union (≥1 callers detected the variants) | 9,252 | 10,382 |
| **NA19240** | | |
| | **3-SV callers** | **5-SV callers** |
| Total SV | 18,398 | 28,614 |
| Total Neighbor SV | 11,454 | 12,944 |
| Neighbor SV (include 1 member variant) | **7,210** | **6,531** |
| Neighbor SV (include 2 member variants) | **1,816** | **2,110** |
| Neighbor SV (include 3 member variants) | **2,356** | **1,320** |
| Neighbor SV (include 4 member variants) | 46 | **1,474** |
| Neighbor SV (include 5 member variants) | 16 | **1,372** |
| Neighbor SV (include >5 member variants) | 10 | 137 |
| Multiple agreement (≥2 callers detected the variants) | 3,922 | 6,154 |
| Union (≥1 callers detected the variants) | 11,454 | 12,944 |



doi:10.6342/NTU202500995

**Table S9. Distribution of SVs detected by different combination strategies.**

| HG002 | | | | | |
|---|---|---|---|---|---|
| **Strategy** | **Total** | **DEL** | **INS** | **DUP** | **INV** |
| DRAGEN | 8,504 | 4,787 | 3,403 | 101 | 213 |
| III-multiple agreement | 3,789 | 3,087 | 162 | 382 | 158 |
| V-multiple agreement | 5,280 | 3,565 | 162 | 1,079 | 474 |
| III-union | 9,771 | 5,353 | 1,799 | 1,786 | 833 |
| V-union | 11,255 | 5,605 | 1,799 | 2,933 | 918 |
| **HG00514** | | | | | |
| **Strategy** | **Total** | **DEL** | **INS** | **DUP** | **INV** |
| DRAGEN | 12,772 | 6,403 | 5,100 | 844 | 425 |
| III-multiple agreement | 3,858 | 3,125 | 132 | 427 | 174 |
| V-multiple agreement | 6,034 | 3,988 | 132 | 1,321 | 593 |
| III-union | 11,164 | 5,842 | 2,277 | 2,140 | 905 |
| V-union | 12,498 | 6,234 | 2,277 | 2,964 | 1,023 |
| **HG00733** | | | | | |
| **Strategy** | **Total** | **DEL** | **INS** | **DUP** | **INV** |
| DRAGEN | 11,589 | 5,895 | 5,313 | 205 | 176 |
| III-multiple agreement | 3,922 | 3,154 | 136 | 448 | 184 |
| V-multiple agreement | 6,154 | 4,043 | 136 | 1,422 | 553 |
| III-union | 11,454 | 6,013 | 2,293 | 2,251 | 897 |
| V-union | 12,944 | 6,426 | 2,293 | 3,195 | 1,030 |
| **NA19240** | | | | | |
| **Strategy** | **Total** | **DEL** | **INS** | **DUP** | **INV** |
| DRAGEN | 13,415 | 6,897 | 6,144 | 209 | 165 |
| III-multiple agreement | 4,449 | 3,694 | 135 | 457 | 163 |
| V-multiple agreement | 6,767 | 4,594 | 135 | 1,404 | 634 |
| III-union | 12,461 | 6,751 | 2,481 | 2,257 | 972 |
| V-union | 13,755 | 7,100 | 2,481 | 3,088 | 1,086 |



doi:10.6342/NTU202500995

**Table S10. Detailed clinical characteristics of the cohort**

| No. | TWGS No. | Age | Sex | FHx | AoA | AoD | MVP | BAV | EL | Myopia | Kyphosis | Scoliosis | Dolicho | Arachno | Sx score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | TWGS001 | 38 | M | NA | + | - | NA | NA | - | + | NA | NA | NA | NA | 1 |
| 2 | TWGS002 | 39 | M | - | + | + | + | + | - | + | - | - | - | + | 4 |
| 3 | TWGS003 | 37 | M | - | + | - | + | - | NA | NA | - | - | - | - | 1 |
| 4 | TWGS004 | 66 | M | + | + | + | NA | NA | - | - | NA | NA | NA | NA | 0 |
| 5 | TWGS005 | 76 | M | - | + | - | - | - | - | - | NA | NA | NA | NA | 0 |
| 6 | TWGS006 | 23 | F | - | + | - | + | - | + | + | - | + | + | + | 15 |
| 7 | TWGS007 | 37 | M | - | + | - | + | - | NA | NA | NA | NA | NA | NA | 1 |
| 8 | TWGS008 | 22 | F | - | + | - | - | - | NA | NA | NA | NA | NA | NA | 0 |
| 9 | TWGS009 | 16 | M | - | + | - | NA | NA | - | + | NA | + | - | - | 6 |
| 10 | TWGS010 | 42 | F | - | + | - | + | - | + | + | - | + | - | + | 11 |
| 11 | TWGS011 | 20 | M | - | + | - | + | - | - | + | - | - | - | + | 7 |
| 12 | TWGS012 | 55 | M | - | + | - | - | - | - | + | - | + | + | - | 6 |
| 13 | TWGS013 | 61 | M |  | + | - | - | + | NA | NA | NA | NA | NA | NA | 0 |
| 14 | TWGS014 | 21 | M | - | + | - | - | - | - | + | - | + | - | - | 14 |
| 15 | TWGS015 | 62 | M | - | + | - | - | - | NA | NA | NA | NA | NA | NA | 0 |
| 16 | TWGS016 | 34 | M | - | + | - | - | + | - | + | + | - | - | - | 9 |
| 17 | TWGS017 | 18 | F | - | + | - | - | - | - | + | - | - | - | - | 3 |
| 18 | TWGS018 | 24 | F | + | + | - | + | - | + | + | + | + | - | + | 12 |
| 19 | TWGS019 | 18 | F |  | + | - | - | - | + | + | + | NA | NA | NA | 5 |
| 20 | TWGS020 | 28 | M | + | + | + | - | - | - | + | + | - | - | - | 1 |
| 21 | TWGS021 | 44 | M | NA | + | - | - | - | NA | NA | NA | NA | NA | NA | 0 |
| 22 | TWGS022 | 54 | M | - | + | - | - | - | NA | NA | NA | NA | NA | NA | 0 |
| 23 | TWGS023 | 56 | M | NA | + | - | - | - | NA | NA | NA | NA | NA | NA | 0 |
| 24 | TWGS024 | 39 | M | - | - | + | - | - | - | + | - | - | + | - | 5 |
| 25 | TWGS025 | 53 | M | + | + | + | - | - | NA | NA | - | + | - | - | 3 |
| 26 | TWGS026 | 42 | F | - | + | - | + | - | - | + | NA | NA | NA | - | 1 |
| 27 | TWGS027 | 40 | M | + | + | - | - | - | NA | NA | + | - | - | - | 2 |
| 28 | TWGS028 | 60 | M | NA | - | + | - | - | NA | NA | NA | - | NA | NA | 0 |
| 29 | TWGS029 | 34 | M | - | + | - | - | - | + | + | - | + | - | + | 6 |
| 30 | TWGS030 | 51 | M | NA | + | - | - | - | NA | NA | NA | NA | NA | NA | 0 |
| 31 | TWGS031 | 26 | M | - | + | - | + | + | NA | NA | + | + | - | - | 4 |
| 32 | TWGS032 | 65 | F | - | + | - | - | - | - | + | - | + | - | - | 3 |
| 33 | TWGS033 | 61 | M | - | + | - | - | - | - | + | - | - | - | - | 0 |
| 34 | TWGS034 | 39 | M | - | + | + | - | - | - | + | - | - | - | - | 0 |

FHx: family history; AoA: aortic aneurysm; AoD: aortic dissection; MVP: mitral valve prolapse; BAV: biscuipid aortic valve; EL: ectopic lentis; Dolicho: dolichostenomelia ; Arachno: arachnodactyly; Sx score: systemic score



doi:10.6342/NTU202500995

## Table S11. Genetic variants identified in each proband

| No. | TWGS No. | SNVs and small indels | | | | SVs | | | MEIs | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | all | rare† | path 1 | path 2 | all | rare | path 3 | all | rare | path 4 |
| 1 | TWGS001 | 4875172 | 364485 | 8 | 275 | 9388 | 857 | 27 | 1441 | 475 | 8 |
| 2 | TWGS002 | 4892659 | 363794 | 11 | 293 | 9334 | 874 | 26 | 1415 | 489 | 9 |
| 3 | TWGS003 | 4906171 | 369540 | 6 | 269 | 9824 | 1138 | 28 | 1468 | 535 | 8 |
| 4 | TWGS004 | 4900520 | 372071 | 5 | 268 | 10562 | 1168 | 36 | 1541 | 549 | 12 |
| 5 | TWGS005 | 4961424 | 383496 | 11 | 294 | 10959 | 1043 | 35 | 1538 | 360 | 4 |
| 6 | TWGS006 | 4891939 | 361471 | 5 | 260 | 9494 | 923 | 41 | 1462 | 350 | 8 |
| 7 | TWGS007 | 4848760 | 352927 | 7 | 264 | 10191 | 1232 | 43 | 1455 | 325 | 3 |
| 8 | TWGS008 | 4917008 | 361433 | 9 | 242 | 9325 | 845 | 31 | 1404 | 422 | 7 |
| 9 | TWGS009 | 4886046 | 363861 | 7 | 263 | 9678 | 1140 | 44 | 1478 | 432 | 10 |
| 10 | TWGS010 | 4887132 | 363053 | 6 | 241 | 9160 | 1145 | 40 | 1425 | 334 | 7 |
| 11 | TWGS011 | 4856438 | 366417 | 7 | 275 | 9014 | 818 | 33 | 1423 | 509 | 12 |
| 12 | TWGS012 | 4907374 | 369615 | 5 | 268 | 9517 | 909 | 29 | 1455 | 457 | 10 |
| 13 | TWGS013 | 4910376 | 384613 | 10 | 263 | 9960 | 924 | 31 | 1435 | 481 | 7 |
| 14 | TWGS014 | 4863455 | 364475 | 6 | 273 | 8431 | 796 | 20 | 1334 | 336 | 6 |
| 15 | TWGS015 | 4860732 | 368270 | 7 | 284 | 8033 | 712 | 16 | 1368 | 366 | 4 |
| 16 | TWGS016 | 4869552 | 360981 | 5 | 326 | 9965 | 891 | 24 | 1427 | 512 | 7 |
| 17 | TWGS017 | 4887754 | 347232 | 2 | 250 | 9443 | 795 | 24 | 1388 | 325 | 6 |
| 18 | TWGS018 | 4909976 | 368572 | 3 | 246 | 10224 | 1277 | 48 | 1471 | 512 | 8 |
| 19 | TWGS019 | 4916762 | 368749 | 5 | 310 | 8888 | 840 | 28 | 1376 | 479 | 11 |
| 20 | TWGS020 | 4878850 | 381615 | 4 | 296 | 8684 | 826 | 34 | 1385 | 431 | 7 |
| 21 | TWGS021 | 4863907 | 360959 | 5 | 257 | 8871 | 797 | 24 | 1369 | 503 | 6 |
| 22 | TWGS022 | 4874908 | 360753 | 9 | 256 | 8862 | 1041 | 44 | 1418 | 342 | 6 |
| 23 | TWGS023 | 4886478 | 360243 | 10 | 282 | 9954 | 900 | 33 | 1465 | 520 | 9 |
| 24 | TWGS024 | 4930192 | 371558 | 5 | 320 | 10541 | 1253 | 42 | 1536 | 366 | 6 |
| 25 | TWGS025 | 4827412 | 360076 | 8 | 260 | 8375 | 951 | 35 | 1384 | 315 | 5 |
| 26 | TWGS026 | 4816914 | 349233 | 5 | 247 | 7430 | 791 | 29 | 1340 | 452 | 5 |
| 27 | TWGS027 | 4853994 | 362907 | 5 | 236 | 8204 | 723 | 16 | 1368 | 468 | 4 |
| 28 | TWGS028 | 4848947 | 360852 | 8 | 257 | 8560 | 989 | 32 | 1451 | 336 | 3 |
| 29 | TWGS029 | 486148 | 358742 | 3 | 249 | 9095 | 845 | 28 | 1437 | 348 | 9 |
| 30 | TWGS030 | 4872452 | 364148 | 5 | 278 | 9068 | 886 | 28 | 1458 | 505 | 6 |
| 31 | TWGS031 | 4857124 | 366860 | 11 | 264 | 9210 | 879 | 30 | 1454 | 325 | 4 |
| 32 | TWGS032 | 4952868 | 366155 | 5 | 288 | 9712 | 1173 | 34 | 1460 | 475 | 5 |
| 33 | TWGS033 | 4892976 | 363098 | 4 | 272 | 10792 | 956 | 33 | 1525 | 473 | 10 |
| 34 | TWGS034 | 4872963 | 360265 | 6 | 260 | 10034 | 899 | 39 | 1471 | 503 | 12 |

†Allele frequency less than 1%; Prioritization pathway 1: by clinical significance level in Clinvar database; Prioritization pathway 2: by prediction score; Prioritization pathway 3: by ACMG classification; Prioritization pathway 4: by AnnotSV ranking score



doi:10.6342/NTU202500995

**Table S12. Supported variants filtered through various panels**

| No. | TWGS No. | Supported variants | Focused panel | Broad spectrum panel |
|---|---|---|---|---|
| 1 | TWGS001 | 266 | 0 | 25 |
| 2 | TWGS002 | 282 | 3 | 46 |
| 3 | TWGS003 | 272 | 2 | 32 |
| 4 | TWGS004 | 282 | 5 | 35 |
| 5 | TWGS005 | 292 | 1 | 45 |
| 6 | TWGS006 | 268 | 4 | 31 |
| 7 | TWGS007 | 265 | 3 | 30 |
| 8 | TWGS008 | 239 | 0 | 23 |
| 9 | TWGS009 | 278 | 2 | 25 |
| 10 | TWGS010 | 257 | 4 | 32 |
| 11 | TWGS011 | 283 | 3 | 36 |
| 12 | TWGS012 | 265 | 5 | 23 |
| 13 | TWGS013 | 275 | 4 | 27 |
| 14 | TWGS014 | 256 | 3 | 27 |
| 15 | TWGS015 | 272 | 1 | 23 |
| 16 | TWGS016 | 311 | 0 | 24 |
| 17 | TWGS017 | 251 | 5 | 25 |
| 18 | TWGS018 | 262 | 3 | 33 |
| 19 | TWGS019 | 304 | 2 | 32 |
| 20 | TWGS020 | 301 | 1 | 37 |
| 21 | TWGS021 | 250 | 0 | 17 |
| 22 | TWGS022 | 267 | 4 | 40 |
| 23 | TWGS023 | 284 | 6 | 40 |
| 24 | TWGS024 | 329 | 5 | 34 |
| 25 | TWGS025 | 270 | 5 | 37 |
| 26 | TWGS026 | 253 | 3 | 22 |
| 27 | TWGS027 | 225 | 2 | 25 |
| 28 | TWGS028 | 263 | 3 | 32 |
| 29 | TWGS029 | 257 | 0 | 29 |
| 30 | TWGS030 | 280 | 5 | 32 |
| 31 | TWGS031 | 261 | 2 | 30 |
| 32 | TWGS032 | 283 | 2 | 36 |
| 33 | TWGS033 | 278 | 4 | 27 |
| 34 | TWGS034 | 275 | 3 | 29 |



doi:10.6342/NTU202500995

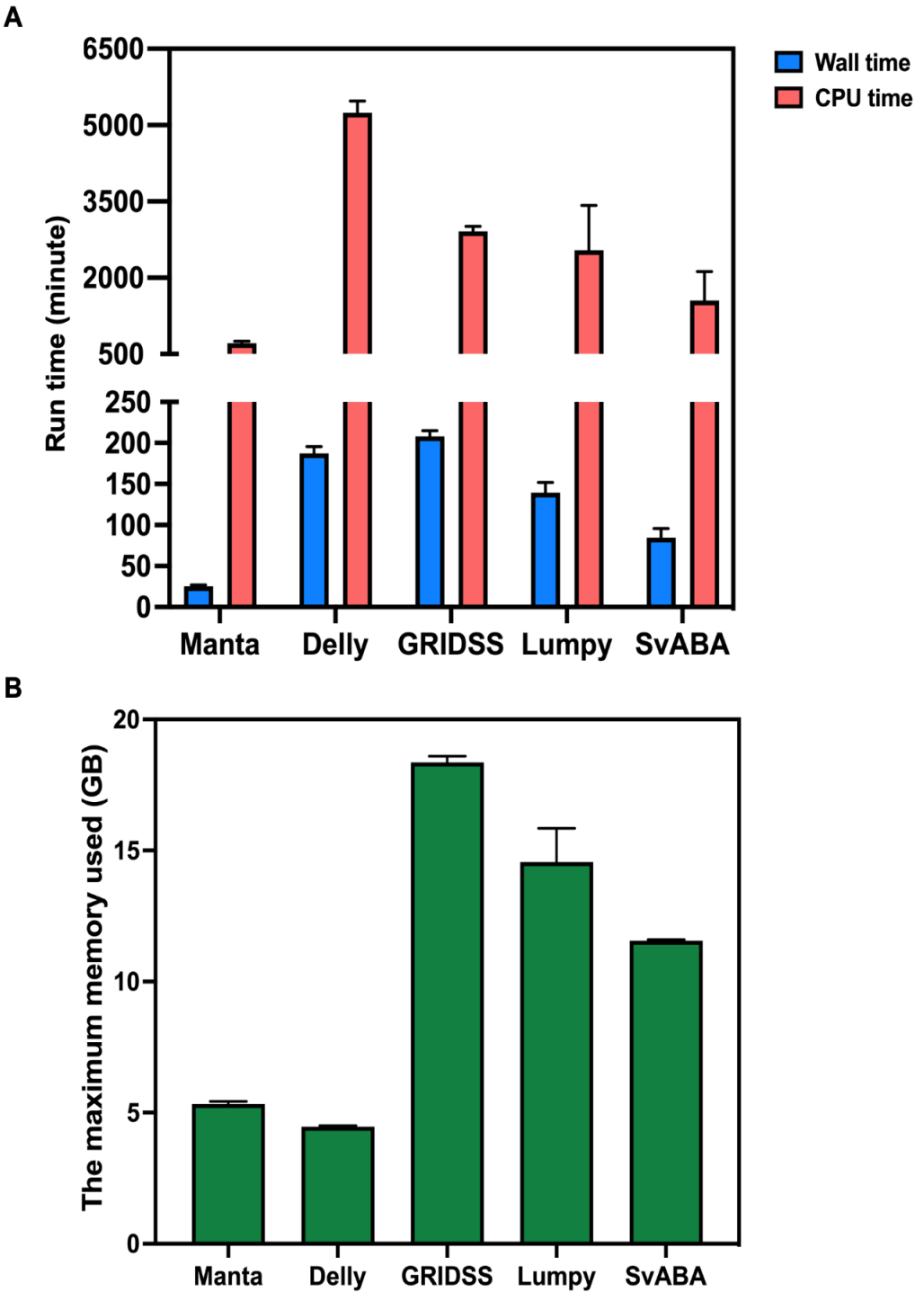
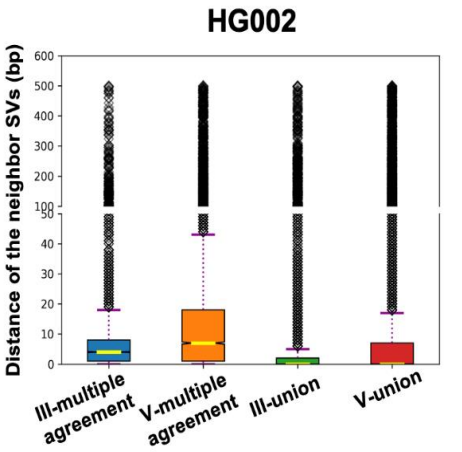


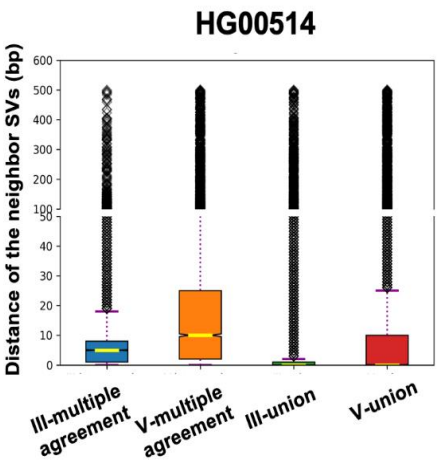

**Figure S1. Runtime performance of SV callers**



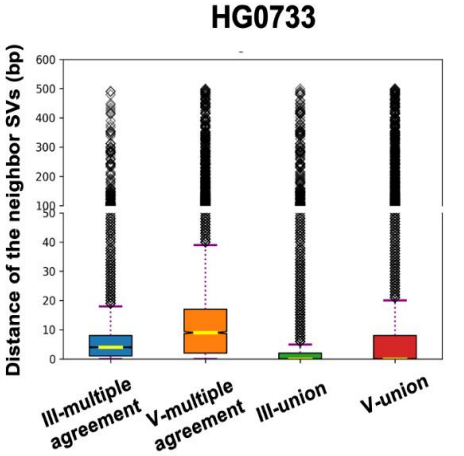
doi:10.6342/NTU202500995



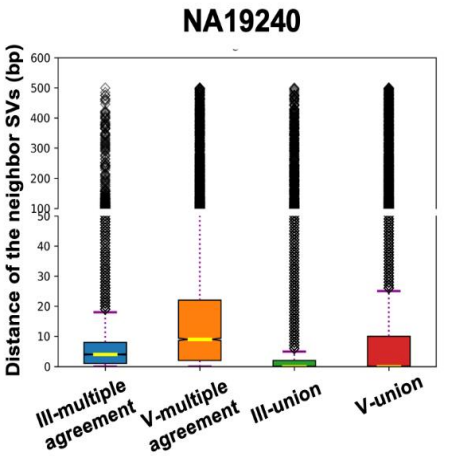

| Strategy | Q1 | Median | Q3 | IQR |
|---|---|---|---|---|
| III-multiple agreement | 1 | 4.0 | 8.0 | 7.0 |
| V-multiple agreement | 1 | 7.0 | 18.0 | 17.0 |
| III-union | 0 | 0 | 2.0 | 2.0 |
| V-union | 0 | 0 | 7.0 | 7.0 |




| Strategy | Q1 | Median | Q3 | IQR |
|---|---|---|---|---|
| III-multiple agreement | 1 | 5.0 | 8.0 | 7.0 |
| V-multiple agreement | 2 | 10.0 | 25.0 | 23.0 |
| III-union | 0 | 0 | 1.0 | 1.0 |
| V-union | 0 | 0 | 10.0 | 10.0 |




| Strategy | Q1 | Median | Q3 | IQR |
|---|---|---|---|---|
| III-multiple agreement | 1 | 5.0 | 8.0 | 7.0 |
| V-multiple agreement | 2 | 9.0 | 25.0 | 23.0 |
| III-union | 0 | 0 | 2.0 | 2.0 |
| V-union | 0 | 0 | 10.0 | 10.0 |




| Strategy | Q1 | Median | Q3 | IQR |
|---|---|---|---|---|
| III-multiple agreement | 1 | 4.0 | 8.0 | 7.0 |
| V-multiple agreement | 2 | 9.0 | 22.0 | 20.0 |
| III-union | 0 | 0 | 2.0 | 2.0 |
| V-union | 0 | 0 | 10.0 | 10.0 |

**Figure S2. Distance of neighboring SVs in combination strategies**



doi:10.6342/NTU202500995

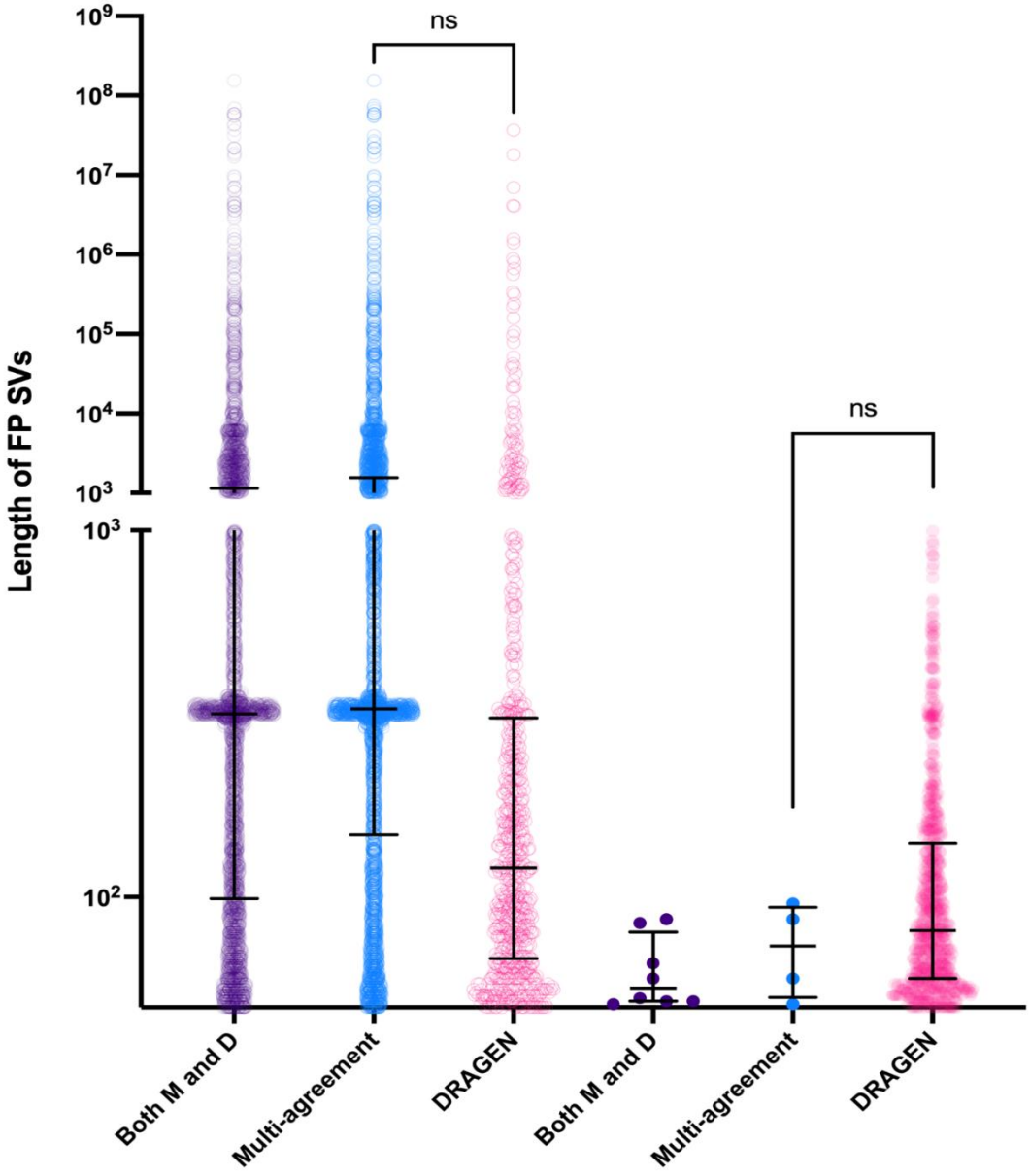



**Figure S3. Length distribution of FP SVs**



doi:10.6342/NTU202500995

## Declaration of authorship and use of published work

I confirm that this doctoral dissertation, titled “Systemic Genetic Analysis for Thoracic Aortic Aneurysm and Dissection (TAAD) and Comparison of Performances of Structural Variants Detection Algorithms in Solitary or Combination Strategy”, incorporates sections of my previously published work and unpublished work. These publications have contributed to the development of the research presented herein. I certify that, except where indicated, this dissertation is my original work and has not been submitted for any other degree or qualification at this or any other institution.

**Published work**

1. Duan DM, Chiu HH, Chen PL, Yeh PT, Yu CW, Yang KC, Yu CC. Clinical manifestations and genetic characteristics in the Taiwan thoracic aortic aneurysm and dissection cohort - a prospective cohort study. J Formos Med Assoc. 2022 Jun;121(6):1093-1101.

Referenced in section 2: Gene panel analysis of NTU TAAD cohort

2. Duan DM, Cheng C, Huang YS, Chung AK, Chen PX, Chen YA, Hsu JS, Chen PL. Comparisons of performances of structural variants detection algorithms in solitary or combination strategy. PLoS One. 2025 Feb 6;20(2):e0314982.

Referenced in section 3: Comparisons of performances of structural variants detection algorithms in solitary or combination strategy

**Unpublished work**

Section 4: NTU TAAD WGS cohort



doi:10.6342/NTU202500995