# Rare Non-coding Variants in Enhancer Regions Contribute to Bicuspid Aortic Valve Pathology

Grado en Biotecnología

Autor: Álvaro Esteban Pérez

Tutora: Ana Isabel Jiménez Belenguer

Tutora externa: Pelin Sahlén

Centro de Investigación: Science for Life Laboratory (SciLifeLab)

Curso: 2022/2023

Estocolmo, Junio de 2023

## RESUMEN

La válvula aórtica bicúspide (BAV) es un defecto genético de la válvula cardíaca asociado con múltiples complicaciones cardíacas. Una BAV es una válvula aórtica que contiene solo dos cúspides en lugar de tres (TAV). BAV es una enfermedad rara con una alta heredabilidad. Sin embargo, en la mayoría de los casos no se han identificado las variantes genéticas específicas responsables de la enfermedad. Debido a la importancia de la genética en BAV, nuestro objetivo fue investigar el papel de las variantes raras no codificantes en el desarrollo de esta enfermedad. En este estudio se han analizado los genomas, transcriptomas y epigenomas de células endoteliales aórticas de individuos con válvula aórtica bicúspide y tricúspide. Se han identificado variantes raras no codificantes en regiones potenciadoras del genoma en muestras de BAV. Las variantes raras no codificantes más prometedoras serán seleccionadas para su validación mediante ensayos de reporteros. Con estos ensayos, pretendemos demostrar que las variantes raras no codificantes tienen un alto impacto en el desarrollo de esta patología.

Este trabajo se relaciona con el siguiente ODS de la Agenda 2030: 3 (Salud y Bienestar).

## PALABRAS CLAVE

VAB; VAT; Captura Hi-C (Hi-Cap); Variantes raras; Potenciadores; Pez zebra; WGS.

**Título:** Variantes raras no codificantes en regiones potenciadoras contribuyen a la patología de la válvula aórtica bicúspide

**Alumno:** D. Álvaro Esteban Pérez

**Tutora:** Prof. Dña. Ana Isabel Jiménez Belenguer

**Tutora externa:** Prof. Dña. Pelin Sahlén

**Titulación:** Grado en Biotecnología

**Localidad y fecha:** Estocolmo, Junio de 2023

## ABSTRACT

Bicuspid aortic valve (BAV) is a genetic defect of the heart valve associated with multiple heart complications. A BAV is an aortic valve that contains only two cusps instead of three (TAV). BAV is a rare disease with a high heritability. However, the specific genetic variants responsible for the disease have not been identified in most cases. Because of the importance of genetics in BAV, we aimed to investigate the role of rare non-coding variants in the development of this disease. In this study, the genomes, transcriptomes, and epigenomes of aortic endothelial cells from individuals with either bicuspid or tricuspid aortic valve have been analyzed. Rare non-coding variants in enhancer regions of the genome have been identified in BAV samples. The most promising rare non-coding variants will be selected for validation using reporter assays. With these assays we aim to prove that rare non-coding variants have a high impact in the development of this pathology.

This work is related to the Sustainable Development Goal (SDG) 3: Good health and Wellbeing.

## KEY WORDS

BAV; TAV; Capture Hi-C (Hi-Cap); Rare variants; Enhancers; Zebrafish; WGS.

**Título:** Rare Non-coding Variants in Enhancer Regions Contribute to Bicuspid Aortic Valve Pathology

**Student:** D. Álvaro Esteban Pérez

**Tutor:** Prof. Dña. Ana Isabel Jiménez Belenguer

**External Tutor:** Prof. Dña. Pelin Sahlén

**Degree:** Degree in Biotechnology

**Location and date:** Stockholm, June 2023

## INDEX

# 1. INTRODUCTION

## 1.1. BAV PATHOLOGY

Bicuspid aortic valve (BAV) is a genetic defect of the heart valve associated with multiple heart complications. A BAV is an aortic valve that contains only two cusps instead of three (TAV). It is the most common congenital heart defect, affecting approximately 2% of the population (PENNMEDICINE, 2023).

In adulthood, BAV has been linked with various complications, like aortic valve stenosis, endocarditis, aortic aneurysm formation, and aortic dissection (Siu & Silversides, 2010). Despite its prevalence, our understanding of BAV disease remains incomplete.

Several observations suggest a genetic component in the development of BAV disease, such as familial clustering of the condition. A study made by Glick and Roberts reported a 24% prevalence of aortic valve disease in families with multiple affected individuals (Glick & Roberts, 1994). However, unraveling the genetic basis of BAV is a complex task that poses challenges.

Improving our understanding of the genetics of BAV disease will not only enhance our knowledge of the condition itself but also provide insights into the broader field of cardiovascular genetics. It may enable the development of more accurate diagnostic tools, risk stratification models, and targeted therapeutic interventions for individuals affected by BAV and related cardiovascular disorders.

## 1.2. COMPLEX DISEASES

Previous studies investigating the genetic basis of BAV have primarily focused on familial linkage-based approaches, which have identified protein-coding mutations in genes such as *NOTCH1* (Garg et al., 2005), *KCNJ2* (Andelfinger et al., 2002), and *ROBO4* (Gould et al., 2019). However, these protein-coding mutations are responsible for only a small fraction (<1%) of BAV cases (Gould et al., 2019), indicating that the genetic architecture of BAV is more complex than initially anticipated. This approach disregards the

significant portion of the genetic material that is non-coding and yet plays a vital biological role.

To gain a broader understanding of the genetic factors contributing to BAV, large-scale population-based studies have utilized genome-wide association (GWAS) methods. These studies have analysed the entire genome to identify common genetic variants associated with BAV. However, in a cohort of 480 BAV patients, no specific variants reached genome-wide significance, highlighting the lack of common variation role for the condition (Fulmet et al., 2019).

These findings suggest that the genetic underpinnings of BAV are complex, involving both protein-coding and non-coding variants. To address the possible involvement of regulatory variation in BAV pathology we searched for regulatory mutations impacting valve development.

## 1.3. ENHANCER VARIANTS IN COMPLEX DISEASES

Enhancers are short stretches of DNA (between 100 and 1000bp) capable of modulating the expression levels of genes over long genomic distances (Panigrahi & O'Malley, 2021). However, enhancer identification has been proven to be very challenging for various reasons (Pennacchio et al., 2013).

Firstly, enhancers are dispersed across the non-coding part of the genome (approximately 98%) composed of roughly 3.136 billion base pairs (Pennacchio et al., 2013).

Secondly, the location of enhancers relative to their target gene can vary deeply. Enhancers can be found upstream or downstream of genes, they can also reside within introns or influence genes located at a considerable distance. Some enhancers even have the capacity to regulate various genes (Mohrs et al., 2001). Thirdly, while the sequence code of protein-coding genes is well-defined and understood, the general sequence code of enhancers, if it even exists, is still not well understood. Consequently, it is challenging to computationally identify enhancers solely based on DNA sequence information (Pennacchio et al., 2013).

Lastly, the activity of enhancers can be highly specific, being restricted to particular tissues or cell types, a specific point in time, or specific physiological, pathological or environmental conditions (Pennacchio et al., 2013).

Enhancer variants have been implicated in the pathogenesis of various complex diseases. Genome-wide association studies (GWAS) have identified enhancers associated with diseases such as cancer (Herz et al., 2014), cardiovascular disorders (Smemo et al., 2012), neurodegenerative diseases (Carullo et al., 2019), and autoimmune conditions (Dey et al., 2022). These enhancers can harbour genetic variations that affect their regulatory activity, leading to dysregulation of target genes and contributing to disease susceptibility or progression. Exploring the role of enhancer variants in complex diseases can provide valuable insights into disease mechanisms and potential therapeutic targets.

## 1.4. CHROMATIN STRUCTURE STUDY

The three-dimensional organization of the genome plays a crucial role in gene regulation. Genomes are organized three-dimensionally thanks to the folding of DNA into various structures like chromatin fibers, chromosome domains, and complete chromosomes. In the interphase, chromosomes exist as chromosome territories. These territories are defined as the nuclear space occupied by the DNA of each chromosome. Although the internal structure of chromosome territories is not fully understood, it is believed to consist of a network of interconnected channels between looping chromatin fibers. This relatively open structure allows gene regulatory factors to access the interior of chromosome territories. The higher-order organization of the genome into chromatin fibers and chromosomes is known as a critical factor contributing to gene regulation (Misteli, 2010).

Chromatin structure, including the folding and looping of DNA, influences the interaction between enhancers and target genes. Techniques such as chromosome conformation capture (3C) and Hi-C have been developed to study the three-dimensional genome architecture. These methods allow researchers to capture the physical interactions between distal genomic regions and identify long-range enhancer-

gene interactions. By investigating the three-dimensional organization of the genome, we can unravel the spatial relationships between enhancers and their target genes, providing insights into the regulatory networks that govern gene expression and their perturbations in disease.

### 1.4.1. CHROMOSOME CONFORMATION CAPTURE (3C)

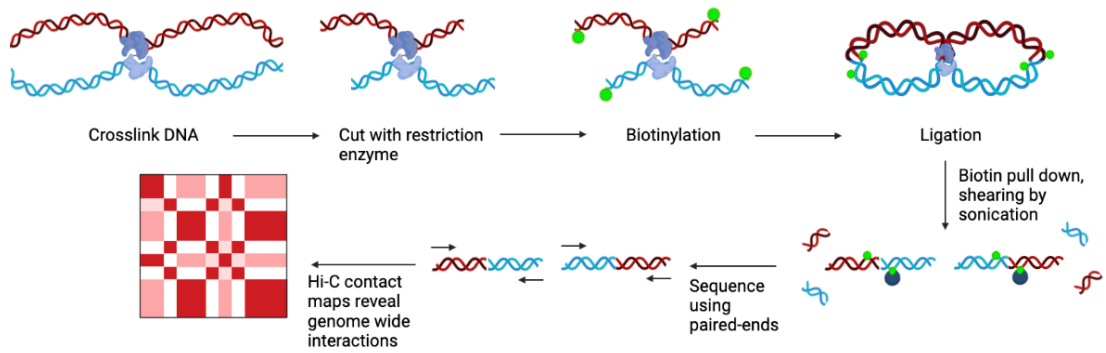
To study the structure of chromatin a snapshot of the genomic packaging of DNA is taken by treating the sample with formaldehyde. Formaldehyde links the molecules that are physically close together with a covalent bond in a process known as crosslinking.

In Chromosome conformation capture (3C) DNA is first crosslinked with formaldehyde. The crosslinked DNA is then digested and re-ligated resulting in the formation of ligation products. These ligation products contain information about the physical location of DNA fragments in the nucleus (Belton et al., 2012).

### 1.4.2. HI-C

However, it wasn't possible to combine 3C with next-generation sequencing due to large portions of 3C material coming from unligated DNA (genomic DNA). In Hi-C, a biotin-labelled nucleotide is introduced at the ligation junction. This biotin label allows for the specific enrichment of chimeric DNA ligation junctions as shown in Figure 1 (Belton et al., 2012).

**Figure 1.** Scheme of the Hi-C technique.



### 1.4.3. CAPTURE HI-C (Hi-Cap, CHi-C)

Hi-C provides a comprehensive view of genomic contacts, but it suffers from low resolution. Hi-C requires deep sequencing to accurately identify specific interactions. To overcome this challenge and focus on specific regulatory loops, researchers have developed library enrichment strategies like Capture Hi-C.

Capture Hi-C (Hi-Cap) is a specialized technique that enhances the resolution and specificity of Hi-C. Hi-Cap employs RNA baits that are designed to target and select predefined regions of interest from the pool of ligated Hi-C contacts before sequencing, for instance, promoter enhancer regions (Aljogol et al., 2022).

This targeted approach increases the effective sequencing depth for the selected regions, allowing for a more detailed examination of chromatin structures at a finer scale. By capturing promoter interactions with regulatory elements, Hi-Cap has been widely utilized to study the spatial organization of the genome and the disrupted interactions associated with disease risk loci (Furlan-Magaril et al., 2021).

## 1.5. OTHER HEART DISEASES CAUSED BY ENHANCER MUTATIONS

Several human Mendelian disorders have been associated with genetic variation found in remote enhancer regions. Holt-Oram syndrome (HOS) is a genetic disorder inherited in an autosomal dominant pattern. It is characterized by the presence of congenital heart defects and abnormalities of the radial bones in the forearm. These manifestations of HOS are primarily caused by variants in the *TBX5* gene (Vanlerberghe et al., 2019).

However, mutations in the enhancers regulating the cardiac expression of TBX5 have been shown to result in cardiac malformations but not in forelimb malformations, decoupling the heart-limb phenotype previously associated with *TBX5* coding mutations, showing that this enhancer is only responsible for the cardiac expression of *TBX5* (Pennacchio & O'Malley, 2021).

### 1.6. ZEBRAFISH ENHANCER REPORTER ASSAY

After the selection of the possible enhancers, we need to test if they truly regulate gene expression. We will use zebrafish as a model organism to validate the activity of putative enhancer regions identified in individuals with BAV pathology.

Until now, the majority of discovery and validation procedures have been conducted using mice as a model organism (Visel et al., 2007). However, the use of mice is more suitable for small to medium-scale experiments and has certain limitations.

In order to overcome these limitations and facilitate rapid and large-scale genomic screening, zebrafish (*Danio rerio*) has emerged as a valuable alternative. Zebrafish has been demonstrated to be an effective model for detecting enhancer activity in vivo, making it a favorable choice for efficient and high-throughput studies (de la Calle-Mustienes et al., 2005).

Zebrafish shares a high degree of genetic and physiological similarity with humans, its embryos are transparent, allowing for direct visualization of developmental processes and facilitating the assessment of enhancer activity. Furthermore, zebrafish embryos develop rapidly, with a highly conserved vertebrate body.

By introducing the enhancer sequences into zebrafish embryos and monitoring the resulting changes in gene expression, we can directly assess the functionality of enhancers and their impact on target genes.

## 2. OBJECTIVES

Considering the potential role of enhancers in BAV development, our study aimed to achieve the following objectives:

1. Investigate the role of mutations in non-coding enhancer regions of the genome in the development of BAV.

2. Analyze the genomes, transcriptomes, and epigenomes of aortic endothelial cells from individuals with both bicuspid and tricuspid aortic valves to identify candidate enhancer variants.

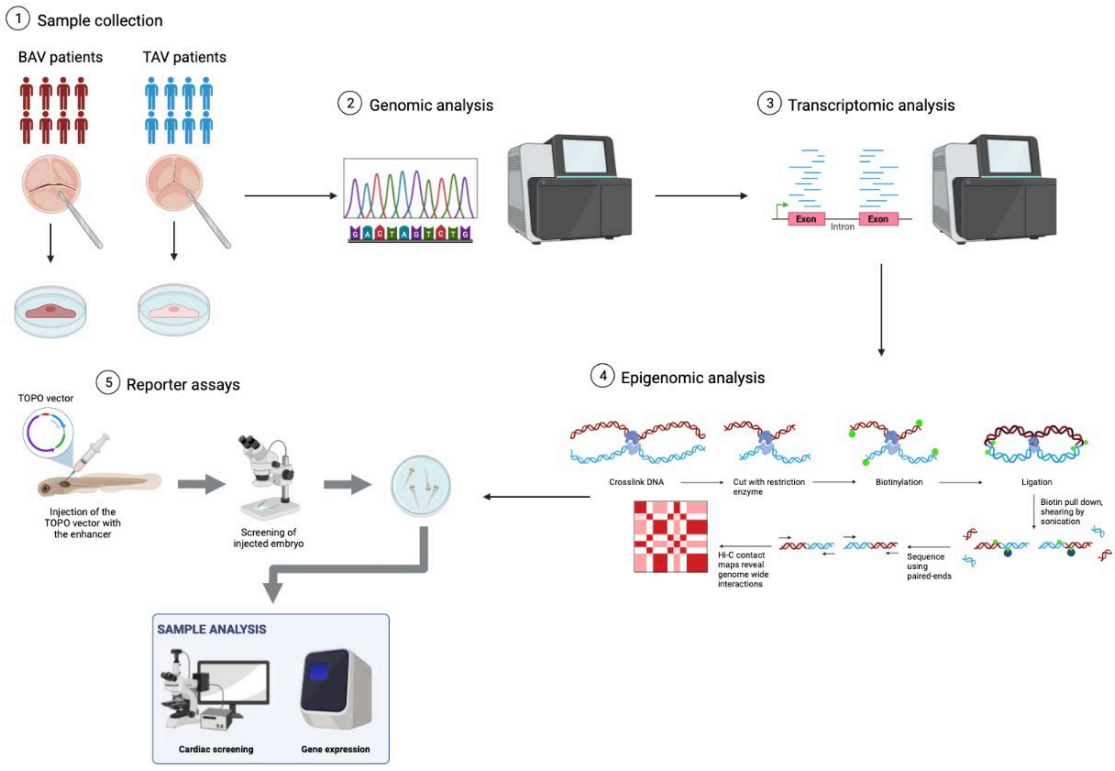
3. Select a subset of promising rare non-coding variants for validation using zebrafish reporter assays.

4. Study the impact of these validated enhancer variants on the regulation of target genes and their association with BAV pathology.

## 3. MATERIAL AND METHODS

The material and methodology used to evaluate the role of enhancers in BAV pathology are detailed below. The entire experimental procedure follows the scheme represented in Figure 2.

**Figure 2.** Scheme of the study design of the project. (BAV; Bicuspid Aortic Valve, TAV; Tricuspid Aortic Valve).



## 3.1. SAMPLE COLLECTION

Human aortic endothelial cells (AEC) were obtained from the ascending aorta of 16 individuals (8 with BAV phenotype and 8 with TAV phenotype) undergoing open-heart surgery at the Cardiothoracic Surgery Unit, Karolinska University Hospital, Stockholm, Sweden. Patients with significant coronary artery disease and with Marfan syndrome were excluded from this study. Ethical approval was obtained from the Human Research Ethics Committee at Karolinska Institutet, and informed consent was obtained from all participants following the guidelines of the Declaration of Helsinki.

The isolation process involved washing the collected biopsies in a solution of PBS with Calcium chloride and Magnesium chloride. The media and adventitia layers were then separated, and the tissues were placed in a non-tissue culture-treated Petri dish. Enzymatic digestion was carried out using a solution containing collagenase A and dispase, allowing for the release of endothelial cells. The endothelial side of the tissue was carefully scraped, and the collected solution containing endothelial cells was centrifuged and filtered to obtain a purified cell suspension. The cells were then resuspended in a basal medium supplemented with the EGM-2 BulletKit and seeded into gelatin-coated flasks. The culture medium was replenished regularly, and the cells were passaged or frozen for future use when reaching confluence.

## 3.2. GENOMIC ANALYSIS

The samples were processed using the TruSeq PCRfree DNA Kit and sequenced achieving a sequencing depth of 30X. To analyze the data, the raw reads were initially aligned to the human reference genome, GRCh37/hg19, using the BWA aligner (version 0.7.8). Duplicate reads were removed using the Picard tool, and SAMtools (version 0.1.19) was employed to filter out reads that were not primarily aligned or not in proper pairs. Variant calling was performed using the Genome Analysis Toolkit (GATK). The Haplotype module of GATK pipelines was utilized for variant calling, and the GenotypeGVCFs tool was applied for joint variant calling.

To identify private variants in individual samples, we utilized the SelectVariants module from GATK pipelines, which specifically selected heterozygous or homozygous variants

in a given sample while requiring other samples to have a normal reference at that position. We obtained combined genotyped VCF data from the SweGen variant frequency dataset (version 20180409). Variant frequency data in our sample cohort were then imputed using the VariantAnnotator module from GATK pipelines, and rare variants with a minor allele frequency (MAF) of less than 0.5% were identified. These rare variants were overlapped with the private interactions observed in our samples, and occurrences where rare variants and private interactions coincided were referred to as concordant allele-specific interactions (Pradhananga, 2020).

## 3.3. TRANSCRIPTOMIC ANALYSIS

RNA-seq was conducted on aortic endothelial cells using the following protocol, with each experiment performed in duplicates. RNA extraction was performed using the RiboCop Kit (Lexogen), and rRNA depletion was carried out according to the manufacturer's protocol using 900 ng of total RNA per sample. Libraries were prepared from 8 uL of the depleted RNA, following the manufacturer's protocol. The libraries were sequenced on the Illumina NextSeq platform using single-end sequencing with a read length of 100 bp.

To align the reads, we utilized the GENCODE transcript annotation (gencode.v26lift37) obtained from the UCSC Table Browser. The STAR aligner was employed to map the reads to the human hg19 genome assembly. Count and RPKM (Reads per Kilobase Million) values were calculated for each transcript and gene using StringTie (version 2.1.4) software. Subsequently, principal component analysis (PCA) was performed to separate the transcripts based on the read count data. Transcripts expressed in all samples were selected for further analysis (Pradhananga, 2020).

## 3.4. EPIGENOMIC ANALYSIS

Hi-Cap was used to identify genome-wide interactions. Hi-Cap consists of two processes previously explained. Firstly, Hi-C was used to select the DNA fragments that are

physically close together. Then, RNA baits were employed to capture the sequences of interest from the Hi-C library.

### 3.4.1 Hi-C

Firstly, cells with intact nuclei were lysed. Next, the DNA was crosslinked using 1% formaldehyde and then digested using the restriction enzyme MboI. The sticky ends resulting from the digestion were then labeled with biotin-14-dATP and ligated thanks to the use of T4 DNA ligase. Once the ligation process was finished, the crosslinked DNA molecules were subjected to heat treatment and purified. The purified DNA was then treated with a DNA polymerase enzyme that has exonuclease activity to remove any unligated ends, ensuring that only ligated fragments remain. To generate smaller DNA fragments, the DNA was fragmented using sonication. Fragments containing a ligation junction were isolated using streptavidin-coated beads, which have a high affinity for the biotin-labeled DNA. The unbound fragments were washed away, leaving behind the desired fragments for further analysis. Finally, the isolated DNA fragments were ready for subsequent steps in the experiment (Pradhananga, 2020).

### 3.4.2. Sequence capture

A probe panel was designed targeting genomic regions of interest (promoters) to enrich the Hi-C library. A set of targets with no known regulatory potential was also designed to be used as negative controls to call significant interactions (HiCapTools). The probe panels were then hybridized to the Hi-C library under stringent conditions for 24 hours. The library was then washed to remove unhybridized DNA. Finally, the resulting enriched DNA was purified and sequenced using Illumina single index paired-end sequencing (Pradhananga, 2020).

### 3.4.3. Hi-Cap interaction calling

The next step consists of identifying and characterizing interactions between paired genomic regions in the Hi-Cap experiment. The interaction calling process reveals two

types of interactions: promoter-distal (P-D) interactions, which involve interactions between a promoter region and a distal regulatory element, and promoter-promoter (P-P) interactions, which involve interactions between two promoter regions.

Mapped reads from the HiCap data are examined, and interactions between genomic regions are detected. To assess the significance of these interactions, a p-value is assigned to each interaction. This p-value is determined by comparing the observed interactions with a background distribution generated from negative control interactions. By assigning p-values, interaction calling helps distinguish statistically significant interactions from random background noise, enabling us to focus on the most relevant and meaningful interactions for further analysis and interpretation. Bowtie2 was used to map raw reads to the human genome (hg19) and HiCUP was used to filter the reads. The filtered aligned files were then processed using HiCaptools proximity detector to detect proximities and significant interactions (Pradhananga, 2020).

## 3.5. REPORTER ASSAYS

In order to validate the enhancers, we focused on the 8 most promising mutations shown in table 1.

**Table 1.** Variants chosen for the enhancer assays. The enhancers affected by the mutation, the genes regulated by the enhancers and the patients in which each mutation is found are also shown.

| Interactor Gene | Var_ID | TF | Patient |
|---|---|---|---|
| S1OOA4 | chr1:153584370:C:A | HMGA2 | BAV2714 |
| FERMT2 | chr14:52296991:A:T | FOXC1 | BAV2714 |
| SMAD7 | chr18:49184756:C:T | MSX1; MSX2 | BAV2716 |
| TGFBR2 | chr3:30575454:T:C | FOXA1 | BAV2375 |
| TGFBR2 | chr3:31939664:A:AGACCACACTTGACCTGGC | NKX2-1 | BAV2930 |
| CTNNB1 | chr3:41240775:T:TTAAAG | SOX9 | BAV2716 |
| PHLDB2 | chr3:111922275:T:C | RBPJ | BAV3004 |
| EDN1 | chr6:12570360:G:T | TBX20 | BAV2375 |

### 3.5.1. CLONING INTO PLASMIDS

To proceed with the cloning process, the constructs containing the desired variants will be amplified directly from patient and control samples. This will involve generating a total of 16 vectors to be used for further analysis. Additionally, for the heterozygous variants, an extra step of sequencing will be performed to identify the vector carrying the alternative variant.

Our aim was to create amplicons of around 500bp centered around the variant of interest. To facilitate the cloning process, the PCRS Topo cloning kit (K59120) was used, which involves the use of a plasmid enabling restriction digestion with EcoRI.

16 primers were designed to amplify the chosen variants. Some of the primers hybridized in zones of repetitive DNA shown in Table 2 in lower cases. For those primers a BLAST against the human genome was performed to ensure that they only hybridized in the desired location.

**Table 2.** Primers chosen to amplify each mutation.

| TF | Left Primer | Right Primer |
|---|---|---|
| HMGA2 | AGGCAGCTTTGCAGGAAAAA | gtgagggcaacgtggtgaaa |
| FOXC1 | CCCAGAGCAGCTCCATCAAT | TCGCCTCTGCTTCATATAAAGA |
| MSX1; MSX2 | GCCACAACTCACTCCCAACT | GGGATGGGGGATGGTTCTTG |
| FOXA1 | gaacacattcgaaccaACGCA | TCGCTATGGGATCAACCAAGG |
| NKX2-1 | GCTCCATTCTGGACACCCTC | aggaaggtgacgaagaAGGA |
| SOX9 | TGGACAGTTTACCAGTTGCCT | TCACTGCACTGTCTCGTGTG |
| RBPJ | GTACGTTCATTCTGCCCCCA | ggtgagggtgcagaaagact |
| TBX20 | TCAGAACTTGTTTGCTGCCC | TGGGACTAGAAGGTAGGAACTT |

Each SNP will be amplified using patient and control DNA, resulting in two PCR reactions per SNP. For heterozygous SNPs, multiple colonies will be selected, and colony PCR will be performed to identify the construct containing the alternative allele. These constructs will then be cloned into TOPO vectors for further analysis.

Sanger sequencing will be conducted to verify the integrity and accuracy of the cloned constructs.

### 3.5.2. ZEBRAFISH ENHANCER REPORTER ASSAY

Finally, to validate the enhancers, a two-step approach using zebrafish as a model system will be employed. Initially, the enhancers will be cloned into Multisite Gateway-based zebrafish transgenesis vectors and injection-based testing will be performed as a preliminary assessment. Subsequently, the ability to regulate cardiac expression of the most interesting candidates will be evaluated in zebrafish embryos.

# 4. RESULTS

We performed whole-genome sequencing of germ-line cells from all sixteen samples and used the GATK pipeline to jointly call variants. In total, we identified 9.2 million variants, including 7.7 million single nucleotide polymorphisms (SNPs) and 1.5 million insertion-deletion (INDEL) variants. To enhance our variant analysis, allele frequency data from the SweGen population variant database was incorporated, which contains variant frequencies from 1000 Swedish individuals. This allowed us to identify 1.8 million rare variants with minor allele frequencies (MAF) less than 0.5% in our sample cohort.

Regarding transcriptome profiling, we were unable to distinguish between the BAV and TAV samples. However, our Capture Hi-C (HiCap) libraries showed a strong correlation within the BAV and TAV groups. Through principal component analysis, we observed a clear separation between the BAV and TAV samples, indicating distinct regulatory landscapes. Using HiCaptools, we identified numerous promoter-distal (P-D) interactions and promoter-promoter (P-P) interactions. On average, we detected 226,000 P-D interactions involving 196,890 promoter-interacting regions (PIRs) and 14,343 promoters, as well as 14,912 P-P interactions. We applied stringent criteria, including a supporting pair threshold and a false discovery rate (FDR) corrected p-value threshold of less than 0.05 and excluded interactions spanning distances larger than 5 megabases. The analysis of promoter interactions successfully differentiated the BAV and TAV samples. These findings suggest that the regulatory landscape of aortic cells in BAV and TAV provides valuable information related to the pathology, which is not captured by bulk RNA profiling alone.

Overall, our preliminary analysis of the genomes, transcriptomes, and epigenomes of aortic endothelial cells from individuals with BAV and TAV has uncovered rare non-coding variants in enhancer regions that may be associated with BAV pathology. Additionally, differential gene expression analysis has revealed genes with altered expression patterns in BAV samples, indicating their potential involvement in the disease.

Validation of the identified enhancer variants using reporter assays in zebrafish is currently underway. The results obtained from these functional experiments will provide valuable insights into the impact of these variants on enhancer activity and their potential role in the regulation of target genes associated with BAV.

## 5. CONCLUSION

Our study aims to investigate the contribution of rare non-coding variants in enhancer regions to the development of BAV. By analyzing the genomes, transcriptomes, and epigenomes of aortic endothelial cells, we have identified candidate variants that may play a role in BAV pathology. The ongoing validation experiments will provide further evidence regarding the functional impact of these enhancer variants.

The use of zebrafish as a model organism for functional validation of enhancer regions is effective and efficient, enabling high-throughput studies. Zebrafish's genetic and physiological similarity to humans, along with its transparent embryos and rapid development, make it the best choice for studying enhancer activity in vivo.

Understanding the role of enhancer variants in BAV not only enhances our knowledge of the disease's genetic basis but also opens avenues for the development of targeted therapies and precision medicine approaches. By elucidating the regulatory mechanisms underlying BAV, we can potentially identify novel therapeutic targets.

One drawback of our study is the use of control samples coming from patients that had undergone heart surgery. These individuals may have had pre-existing cardiovascular conditions that led to the need for surgery. Consequently, the control group may not accurately represent the general population. Another drawback is the small sample size

consisting of only 8 BAV individuals. Furthermore, all these samples were obtained from the same surgery unit. The samples may not represent the broader population with BAV and TAV.

Until now the majority of the studies investigating the underlying molecular mechanisms of disease have focused primarily on mutations on coding sequences, neglecting the impact of the rest of the genome that does not code for proteins. The study of the impact of gene regulation on disease appears to be an exciting field for the years to come.

## 6. REFERENCES


1. Aljogol, D., Thompson, I. R., Osborne, C. S., & Mifsud, B. (2022). Comparison of Capture Hi-C Analytical Pipelines. *Frontiers in genetics*, 13, 786501. https://doi.org/10.3389/fgene.2022.786501

2. Andelfinger, G., Tapper, A. R., Welch, R. C., Vanoye, C. G., George, A. L., Jr, & Benson, D. W. (2002). KCNJ2 mutation results in Andersen syndrome with sex-specific cardiac and skeletal muscle phenotypes. *American journal of human genetics*, 71(3), 663–668. https://doi.org/10.1086/342360

3. Belton, J. M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods (San Diego, Calif.)*, *58*(3), 268–276. https://doi.org/10.1016/j.ymeth.2012.05.001

4. *Bicuspid Aortic Valve* (BAV). Pennmedicine.org. https://www.pennmedicine.org/for-patients-and-visitors/patient-information/conditions-treated-a-to-z/bicuspid-aortic-valve

5. Carullo, N. V. N., & Day, J. J. (2019). Genomic Enhancers in Brain Health and Disease. *Genes*, *10*(1), 43. https://doi.org/10.3390/genes10010043

6. de la Calle-Mustienes, E., Feijóo, C. G., Manzanares, M., Tena, J. J., Rodríguez-Seguel, E., Letizia, A., Allende, M. L., & Gómez-Skarmeta, J. L. (2005). A functional survey of the enhancer activity of conserved non-coding sequences from vertebrate Iroquois cluster gene deserts. Genome research, 15(8), 1061–1072. https://doi.org/10.1101/gr.4004805

7. Dey, K. K., Gazal, S., van de Geijn, B., Kim, S. S., Nasser, J., Engreitz, J. M., & Price, A. L. (2022). SNP-to-gene linking strategies reveal contributions of enhancer-related and candidate master-regulator genes to autoimmune disease. *Cell genomics*, 2(7), 100145. https://doi.org/10.1016/j.xgen.2022.100145

8. Fulmer, D., Toomer, K., Guo, L., Moore, K., Glover, J., Moore, R., Stairley, R., Lobo, G., Zuo, X., Dang, Y., Su, Y., Fogelgren, B., Gerard, P., Chung, D., Heydarpour, M., Mukherjee, R., Body, S. C., Norris, R. A., & Lipschutz, J. H. (2019). Defects in the Exocyst-Cilia Machinery Cause Bicuspid Aortic Valve Disease and Aortic Stenosis. *Circulation*, 140(16), 1331–1341. https://doi.org/10.1161/CIRCULATIONAHA.119.038376

9. Furlan-Magaril, M., Ando-Kuri, M., Arzate-Mejía, R. G., Morf, J., Cairns, J., Román-Figueroa, A., Tenorio-Hernández, L., Poot-Hernández, A. C., Andrews, S., Várnai, C., Virk, B., Wingett, S. W., & Fraser, P. (2021). The global and promoter-centric 3D genome organization temporally resolved during a circadian cycle. *Genome biology*, 22(1), 162. https://doi.org/10.1186/s13059-021-02374-3

10. Garg, V., Muth, A. N., Ransom, J. F., Schluterman, M. K., Barnes, R., King, I. N., Grossfeld, P. D., & Srivastava, D. (2005). Mutations in NOTCH1 cause aortic valve disease. *Nature*, 437(7056), 270–274. https://doi.org/10.1038/nature03940

11. Glick, B. N., & Roberts, W. C. (1994). Congenitally bicuspid aortic valve in multiple family members. *The American journal of cardiology*, 73(5), 400–404. https://doi.org/10.1016/0002-9149(94)90018-3

12. Gould, R. A., Aziz, H., Woods, C. E., Seman-Senderos, M. A., Sparks, E., Preuss, C., Wünnemann, F., Bedja, D., Moats, C. R., McClymont, S. A., Rose, R., Sobreira, N., Ling, H., MacCarrick, G., Kumar, A. A., Luyckx, I., Cannaerts, E., Verstraeten, A., Björk, H. M., Lehsau, A. C., … Dietz, H. C. (2019). ROBO4 variants predispose individuals to bicuspid aortic valve and thoracic aortic aneurysm. *Nature genetics*, 51(1), 42–50. https://doi.org/10.1038/s41588-018-0265-y

13. Herz, H. M., Hu, D., & Shilatifard, A. (2014). Enhancer malfunction in cancer. *Molecular cell*, *53*(6), 859–866. https://doi.org/10.1016/j.molcel.2014.02.033

14. Misteli T. (2010). Higher-order genome organization in human disease. *Cold Spring Harbor perspectives in biology*, *2*(8), a000794. https://doi.org/10.1101/cshperspect.a000794

15. Mohrs, M., Blankespoor, C. M., Wang, Z. E., Loots, G. G., Afzal, V., Hadeiba, H., Shinkai, K., Rubin, E. M., & Locksley, R. M. (2001). Deletion of a coordinate regulator of type 2 cytokine expression in mice. *Nature immunology*, *2*(9), 842–847. https://doi.org/10.1038/ni0901-842

16. Panigrahi, A., O'Malley, B.W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol* 22, 108 (2021). https://doi.org/10.1186/s13059-021-02322-1

17. Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A., & Bejerano, G. (2013). Enhancers: five essential questions. *Nature reviews. Genetics*, *14*(4), 288–295. https://doi.org/10.1038/nrg3458

18. Pradhananga, S. (2020). Towards understanding of complex disease etiology using sequence Capture Hi-C (HiCap) and associated high throughput functional assays (PhD dissertation, KTH Royal Institute of Technology). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-286086

19. Siu, S. C., & Silversides, C. K. (2010). Bicuspid aortic valve disease. *Journal of the American College of Cardiology*, *55*(25), 2789–2800. https://doi.org/10.1016/j.jacc.2009.12.068

20. Smemo, S., Campos, L. C., Moskowitz, I. P., Krieger, J. E., Pereira, A. C., & Nobrega, M. A. (2012). Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Human molecular genetics*, *21*(14), 3255–3263. https://doi.org/10.1093/hmg/dds165

21. Vanlerberghe, C., Jourdain, A. S., Ghoumid, J., Frenois, F., Mezel, A., Vaksmann, G., Lenne, B., Delobel, B., Porchet, N., Cormier-Daire, V., Smol, T., Escande, F., Manouvrier-Hanu, S., & Petit, F. (2019). Holt-Oram syndrome: clinical and molecular description of 78 patients with TBX5 variants. *European journal of human genetics : EJHG*, *27*(3), 360–368. https://doi.org/10.1038/s41431-018-0303-3

22. Visel, A., Minovitsky, S., Dubchak, I., & Pennacchio, L. A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic acids research*, *35*(Database issue), D88–D92. https://doi.org/10.1093/nar/gkl822