UTILISING SNP-SV CORRELATIONS TO FIND STRUCTURAL VARIANTS ASSOCIATED WITH ALZHEIMER'S DISEASE

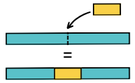
1 – BACKGROUND

Single Nucleotide Polymorphism

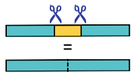
A single base pair change

Structural Variant

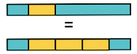
Insertion



Deletion



Tandem Repeat (TR)



- **Genome Wide Association Studies (GWAS):** A study which aims to identify genetic components related to a trait by testing for differences in the frequency of the genetic components between test and control cases.
- **Quantitative Trait Loci (QTL):** Positions in the genome (called loci), that are associated with the variation of a quantitative trait, such as gene expression.
- **Gene Set Enrichment Analysis (GSEA):** An analysis which aims to find biological functions which are statistically over represented (enriched) in the associated functions of a set of genes.
- **Fine-mapping:** The process of identifying a representative set of genetic components by clustering interrelated genetic components around the most statistically significant genetic components
- **Regulatory Element:** Regions outside genes which regulate the expression of nearby genes

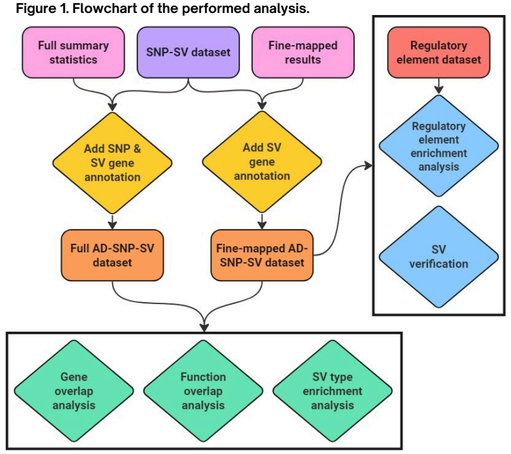
2 – Introduction

Alzheimer's disease (AD) is a neurodegenerative disease affecting roughly 40 million people. 70% of the heritability of AD is expected to be explained by Structural Variants (SVs), however these have been scarcely studied in the context of AD. One reason for this is that SVs have been difficult to detect accurately in the past, however new technologies referred to as Third Generation Sequencing have overcome these limitations. Using these technologies, the SNP-SV dataset consisting of SNPs and their correlation with SVs was created using sequence data from 214 individuals.

This study aims to employ the SNP-SV dataset to identify new correlations between SVs and Alzheimer's disease, and to investigate the properties of these correlated SVs, the associated genes, and their functions.

3 – METHODOLOGY

Figure 1. Flowchart of the performed analysis.



Datasets are obtained from a large Alzheimer's GWAS [1]

Gene annotation is done using by selecting the nearest gene

Gene and function overlap is calculated relative to known Alzheimer's genes and functions obtained from the fine-mapped results. Functions are obtained using GSEA of the genes

SV type enrichment is calculated relative to all SVs in the SNP-SV dataset. SV types refer to TRs, insertions and deletions

The dataset is from a study which partitions the genome into areas associated with various regulatory elements and other mechanisms of gene expression [2]

Regulatory element enrichment is calculated relative to all regulatory elements in the regulatory element dataset

SV verification is performed by comparing the SVs found with 21 SVs found in a previous study by Wang et al. [3]

CONTACT INFORMATION

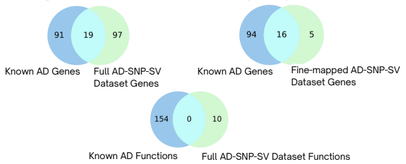
❖ Joris Belder — Author — j.belder-1@student.tudelft.nl ❖ Niccoló Tesi — Supervisor ❖ Marcel Reinders — Responsible Professor

4 – RESULTS

Table 1. Summary statistics of the datasets

Dataset	Total Size	#Unique SNPs	#Unique SVs	#TRs	#Insertions	#Deletions
SNP-SV dataset	7,209,765	3,535,362	22,590	3,323,889	1,993,380	1,892,496
Full summary statistics	12,634	12,634	N/A	N/A	N/A	N/A
Full AD-SNP-SV dataset	11,841	5,248	307	6,329	3,379	2,133
Fine-mapped results	89	89	N/A	N/A	N/A	N/A
Fine-mapped AD-SNP-SV dataset	85	39	85	52	21	12

Figure 2. Gene and function overlap analysis results.



- **Table 1** shows statistics regarding the size and the number of elements in the different datasets.
- **Gene and function overlap analysis** resulted in poor overlap for both genes and functions for the full AD-SNP-SV datasets. No functions were found for the fine-mapped AD-SNP-SV dataset and the gene overlap was deficient. Results are shown in **Figure 2.**
- **SV type enrichment analysis** found a significant enrichment of TRs and a significant depletion of deletions for both datasets analysed.
- **Regulatory element enrichment analysis** resulted in a significant enrichment of heterochromatin areas and a significant depletion of both weak enhancers and transcription elongation areas. The results are shown in **Table 2.**
- **SV verification** found that 5 out of the 85 unique SVs were also found in the previous study. The results are shown in **Table 3.**

Table 2. Regulatory element counts in the fine-mapped AD-SNP-SV dataset compared to the genome

Regulatory element	Weak Enhancer	Weak Txn[1]	Heterochromatin	Strong Enhancer	Weak Promoter	Insulator
Genome	178,463	82,187	74,863	64,078	35,021	33,214
Dataset	2	16	54	2	4	1
Regulatory element	**Txn[1] Elongation**	**Polycomb-repressed**	**Txn[1] Transition**	**Active Promoter**	**Repetitive/CNV[2]**	**Poised Promoter**
Genome	26,479	2,435	16,215	15,256	14,087	5,253
Dataset	9	5	4	0	0	1

[1]Txn = Transcription, [2]CNV = Copy Number Variation

Table 3. Verified SVs

Verified SV	Known SV
chr2:203034349:DEL	chr2:203034369-203039560:DEL
chr6:40959080:INS	chr6:40959079-40959079:INS
chr7:12242078:DEL	chr7:12242077-12242399:DEL
chr10:122457364:DEL	chr10:122457302-122457747:DEL
chr12:113245307:DEL	chr12:113245316-113245625:DEL

5 – CONCLUSION

- 307 unique SVs were present in the full AD-SNP-SV dataset.
- 85 SVs were also present in the fine-mapped AD-SNP-SV dataset, of which 5 were found in previous work as well.
- The gene and consequently the function overlap was poor for both datasets. Most likely due to the simplistic gene annotation approach used.
- An enrichment of Tandem Repeats and a depletion of deletions was found for both the datasets.
- An enrichment of heterochromatin areas and a depletion of both weak enhancers and transcription elongation areas was found for the fine-mapped AD-SNP-SV dataset.

6 – LIMITATIONS AND FUTURE WORK

- The SNP-SV dataset uses sequencing data only from the Netherlands. A more ethnically diverse set of samples may reveal additional associations
- A more comprehensive gene annotation approach of the SVs was infeasible due to a lack of available data mapping SVs to genes. QTL studies involving SVs would be needed to more accurately determine the genes associated with SVs
- Only the types of SVs were tested for enrichment, but other properties may be tested as well
- The regulatory element dataset was based on an outdated reference genome and had to be realigned. This process is not able to realign all areas of the genome, however this did not affect the results as no SV overlapped these areas. A new version of the dataset would improve the reliability of the results.
- Only 21 SVs associated with Alzheimer's presented with their exact genomic location and type were found in previous studies. Future work should expand the verification set as more SVs associated with Alzheimer's are identified.

REFERENCES

[1] Bellenguez,C., K¨u,c¨ukali,F., Jansen,I.E., Kleineidam,L., Moreno-Grau,S., Amin,N., Naj,A.C., Campos-Martin,R., Grenier-Boley,B., Andrade,V., et al. (2022) New insights into the genetic etiology of Alzheimer's disease and related dementias. Nat Genet, 54, 412–436.

[2] Ernst,J., Kheradpour,P., Mikkelsen,T.S., Shoresh,N., Ward,L.D., Epstein,C.B., Zhang,X., Wang,L., Issner,R., Coyne,M., et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature, 473, 43–49.

[3] Wang,H., Dombroski,B.A., Cheng,P.-L., Tucci,A., Si,Y.- Q., Farrell,J.J., Tzeng,J.-Y., Leung,Y.Y., Malamon,J.S., Wang,L.-S., et al. (2023) Structural Variation Detection and Association Analysis of Whole-Genome-Sequence Data from 16,905 Alzheimer's Diseases Sequencing Project Subjects. medRxiv, 10.1101/2023.09.13.23295505, 13 September 2023, pre-print: not peer-reviewed.

TU Delft