Navigating Genetics of Dementia in the Indian Subcontinent

Karan Gujarati[1], Disha Awasthy[1], Shrutee Jakhanwal[1], Ajay Kumar K S[1], Sobha George[1], Pallavi Ghana[1], Aparna Ganapathy[1], Aditi Chatterjee[1], Sameer Phalke[1], Sanjeev Jain[2], Meera Purushottam[2], Biju Viswanath[2], Chinu Patra[2], Ebin Joseph[2], Radhakrishna Bettadapura[1], *Ramesh Hariharan[1]

Strand Life Sciences, Bengaluru, India[1]
Molecular Genetics Laboratory, Department of Psychiatry, NIMHANS, Bengaluru, India[2]

VISIT OUR **WEBSITE**

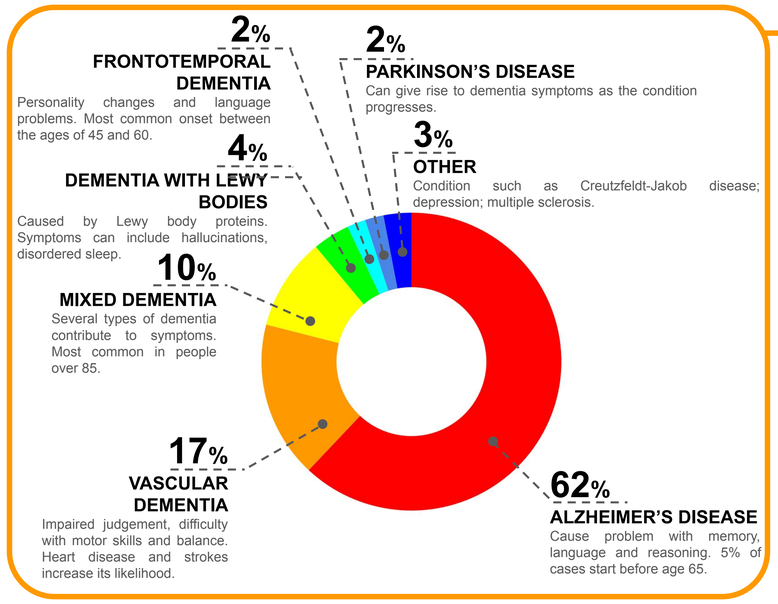
strand
WE SOLVE PROBLEMS IN GENOMICS.

The Global Cell & Gene Therapy Summit 2024
July 8-10, 2024
Boston, MA, USA

Contact
Shrutee Jakhanwal
Product Manager
Ph: +91 9650040413
Email: shrutee.jakhanwal@strandls.com

Introduction

Affecting over 55 million people globally and adding 10 million new cases each year, dementia is the seventh leading cause of death and poses a significant burden. The economic impact is significant, with informal care accounting for half of the costs. This highlights the urgent need for effective interventions and drug targets identified through GWAS studies, particularly in high-prevalence regions such as India.



Objective

1. **Identifying Genetic Risk Factors**
   GWAS seeks to pinpoint common genetic variations in dementia patients, thereby shedding light on disease susceptibility.
2. **Understanding Disease Mechanisms**
   By linking genetic variants to dementia, GWAS offers insights into the biological pathways driving the disease, aiding in drug discovery.
3. **Early Detection and Diagnosis**
   Genetic markers from GWAS may serve as biomarkers, enabling early identification of dementia risk and timely interventions.
4. **Personalized Medicine**
   GWAS results allow tailoring treatments based on individuals' genetic profiles, enhancing precision and efficacy.
5. **Risk Assessment and Prevention**
   Genetic risk scores derived from GWAS inform preventive measures, empowering individuals to reduce their risk through lifestyle changes and interventions.

Highlights

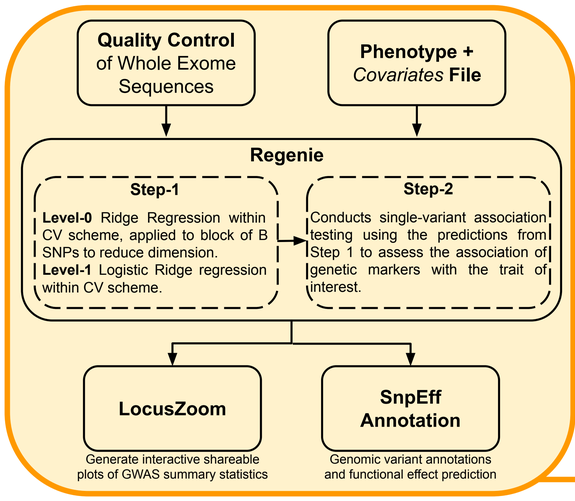
A GWAS pipeline was developed and applied to dementia patients using the UK Biobank dataset, revealing SNP associations predominantly observed populations with European ancestry.

Limited exploration exists regarding these SNP associations in populations with complex genetic substructure like **South Asians**. An additional analysis was conducted on an **Indian** cohort, showcasing population-specific genetic associations distinct from those in European populations.

The well-known **APOE** and **TOMM40** genes show a strong association with AD and dementia in the European cohort.

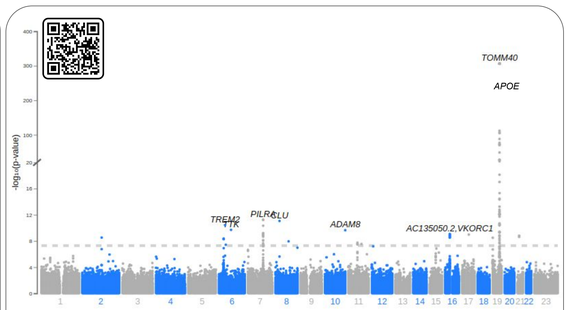
Materials and methods

Our GWAS pipeline analyzed dementia patients using both the UK Biobank dataset and an Indian dataset.
Quality control in exome sequencing is crucial to ensure the reliability and accuracy of genetic variant.
We applied QC criteria: minimum allele count of 10, minor allele frequency of 0.0001, and missing genotype and phenotype rate of 10%, with a Hardy-Weinberg equilibrium threshold of 1e-15. The Hardy-Weinberg equilibrium describes allele and genotype frequencies remaining constant in a population under no evolutionary influences.
In **Regenie step-1** a subset of genetic markers are used to fit a whole genome regression model that captures a good fraction of the phenotypic variance attributable to genetic effects using a leave one chromosome out scheme, that avoids proximal contamination.
In **Regenie step-2**, a larger set of genetic markers are tested for association with the phenotype conditional upon the prediction from the regression model in Step 1.

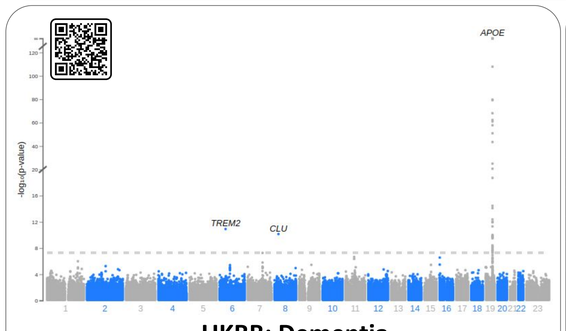


Results

Gene	ID	Odd Ratio	-LOG 10P	Conse-quence	Cases - 56,431			Controls - 285,446			Freq. gnomAD
					Allele Freq.	Aa	aa	Allele Freq.	Aa	aa	
APOE	rs429358	1.676	306.6	Missense	0.218	19799	2427	0.144	70472	5810	0.1476
APOE	rs7412	0.795	73.4	Missense	0.068	7162	234	0.083	43566	2038	0.07385
BCAM	rs28399653	0.862	14.4	Missense	0.030	3253	50	0.034	18928	334	0.02815
BCAM	rs28399654	0.863	14.3	Missense	0.030	3253	50	0.034	18910	337	0.02743
CBLC	rs3208856	0.878	12.1	Missense	0.032	3507	61	0.036	20064	380	0.03072
EXOC3L2	rs10411314	1.121	12.	Missense	0.046	4992	118	0.041	22873	510	0.05209
PILRA	rs1859788	1.050	11.3	Missense	0.689	24137	26840	0.679	124143	131755	0.6768
TREM2	rs75932628	1.435	10.4	Missense	0.004	479	0	0.003	1704	1	0.002855
TTK	rs1391204580	1.422	9.7	Missense	0.004	463	0	0.003	1651	0	0.1707



UKBB: Alzheimer's Disease (AD) (with proxy)*

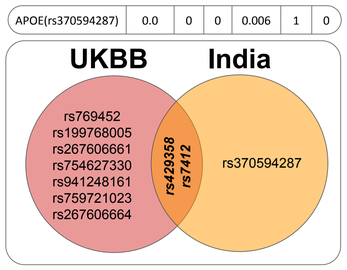


UKBB: Dementia

Gene	ID	Odd Ratio	-LOG 10P	Conse-quence	Cases - 6,760			Controls - 340,979			Freq. gnomAD
					Allele Freq.	Aa	aa	Allele Freq.	Aa	aa	
APOE	rs429358	2.796	613.1	Missense	0.328	2962	738	0.144	84453	6976	0.1476
APOE	rs7412	0.574	51.0	Missense	0.049	623	23	0.083	51995	2382	0.07385
BCAM	rs28399654	0.648	14.5	Missense	0.022	292	6	0.034	22467	397	0.02743
BCAM	rs28399653	0.652	14.1	Missense	0.023	294	6	0.034	22477	395	0.02815
CBLC	rs3208856	0.689	11.9	Missense	0.026	325	10	0.036	23884	445	0.03072
TREM2	rs75932628	2.334	10.9	Missense	0.007	89	1	0.003	2041	2	0.002855
NPW	rs11248906	1.114	6.6	Missense	0.242	2460	404	0.222	117710	16793	0.1968
APOE	rs769452	2.499	6.4	Missense	0.003	43	0	0.001	976	1	0.002026
EXOC3L2	rs189063316	1.352	6.0	Missense	0.023	310	3	0.017	11669	107	0.01673

Intersection of Variants in Indian Cohort and UKBB Dementia Study

Gene (ID)	Cases - 30			Controls - 85		
	Allele Freq.	Aa	aa	Allele Freq.	Aa	aa
APOE (rs429358)	0.100	6	0	0.065	11	0
APOE (rs7412)	0.183	5	3	0.112	19	0
NPW (rs11248906)	0.067	2	1	0.024	4	0
EXOC3L2 (rs189063316)	0.033	2	0	0.024	4	0
APOE(rs370594287)	0.0	0	0	0.006	1	0



The well-known **APOE** and **TOMM40** genes show a strong association with **AD** and **dementia** in the European cohort.

APOE gene, specifically the **ε4** allele (**rs429358**) has a well documented association as a risk factor for Alzheimer disease and other neurodegenerative diseases. **TOMM40** gene is another gene located 5′-upstream of the APOE gene and several studies have shown that the variants TOMM40 gene may contribute to AD risk.

This table lists the intersection of the top missense and variants identified in the UK Biobank (UKBB) Dementia study with those found in the Indian cohort. The accompanying Venn diagram illustrates the overlap of APOE variants between the UKBB and Indian cohorts.

APOE	TREM2	EXOC3L2
rs370594287 ***rs429358*** ***rs7412***	rs75272959 rs143332484 rs753325601	***rs189063316*** rs775521544 rs143444923 rs10411314 rs569002007 rs538391435
BCAM		
rs550392791 rs759591873 rs28399656 rs28399626 rs117737673 rs148391498 rs1135062 rs28399659 rs750550926 rs199854072		
	CBLC	**NPW**
	rs377440313 rs35106910 rs114569424 rs149074838 rs66944506 rs116023028	rs375002657 rs543718628 rs2286471 ***rs11248906*** rs2286472

This table lists the variants found in the Indian cohort that have the same variant effect as the top variants identified in the UKBB Dementia study.

Conclusion

In this study, we investigated the genetics of dementia through GWAS on both UK Biobank and Indian cohorts. Consistent with findings by Goldberg et al. in "*Association of APOE ε2 genotype with Alzheimer's and non-Alzheimer's neurodegenerative pathologies*" (i.e. OR for ε2=0.54 and ε4=3.6) our results highlight the APOE ε4 allele (rs429358) as a significant genetic risk factor for AD with an odds ratio (OR) of 2.8, and the ε2 allele (rs7412) as a protective factor against AD with an OR of 0.57. This is a preliminary study, and further validation is planned as more data becomes available (approximately 400 cases).

Acknowledgement

We gratefully acknowledge the **UK Biobank** and **NIMHANS** Hospital for providing the invaluable data that made this research possible.

***Definition of AD by proxy:** To determine a participant's Alzheimer's disease (AD) risk, assign a maximum risk value of 2 if the participant has been diagnosed with AD. Otherwise, sum the risk from each biological parent: a diagnosed parent contributes 1 each, with a total of 2 if both are diagnosed. The participant's final risk is the highest value from these steps. Participants with a final risk of 1 or above are cases, while those below 1 are controls.